LINEARE ALGEBRA I & II

F. Kappel

Studienjahr 2005/06

Inhaltsverzeichnis

[1] Gauß, Carl Friedrich, 30. 4. 1777 (Braunschweig) – 23. 2. 1855 (Göttingen), neben Euler, Newton und Archimedes einer der bedeutendsten Mathematiker aller Zeiten, “princeps mathematicorum”(de.wikipedia.org/wiki/Carl_Friedrich_Gau%C3%9F).

Kapitel 1

Einiges aus der Mengenlehre

1.1 Mengen und Mengenoperationen

Grundlegend für den Aufbau der Mengenlehre ist die Relation „$\in$". Für zwei Objekte a, M gilt entweder $a \in M$ oder $a \notin M$ (in Worten: a ist Element von M bzw. a ist nicht Element von M). Steht für jedes a fest, ob $a \in M$ oder $a \notin M$ gilt, so nennen wir M eine Menge.

M sei eine Menge. Ist P ein Prädikat (= Eigenschaft), das genau dann für ein Objekt zutrifft, wenn $a \in M$ gilt, so schreiben wir

$$M = \{a \mid P \text{ ist für } a \text{ richtig}\},$$

d.h. M ist die Menge aller Objekte mit der Eigenschaft P. Man beachte, dass M bereits als Menge angenommen wurde. Es ist naheliegend, für ein beliebiges Prädikat Q anzunehmen, dass die Menge aller Objekte, für die das Prädikat Q zutrifft, existiert. Dass dies falsch ist, zeigt das folgende Beispiel (bekannt unter dem Namen „Russell'sche[2] Antinomie"):

> „a besitzt die Eigenschaft Q" sei definiert als $a \notin a$. Angenommen, es existiere die Menge
>
> $$M = \{a \mid Q \text{ ist für } a \text{ richtig}\}.$$
>
> Für M gilt entweder $M \notin M$ oder $M \in M$ (d.h. M besitzt die Eigenschaft Q oder M besitzt die Eigenschaft Q nicht). Angenommen, es gelte $M \notin M$, d.h. Q gilt für M. Dann müßte aber (wegen der Definition von M) $M \in M$ gelten. Gilt andererseits $M \in M$, so gilt Q für M nicht, woraus $M \notin M$ folgt. Diese Überlegungen zeigen, dass für M weder $M \in M$ noch $M \notin M$ gelten kann. Dieser Widerspruch zeigt, dass M nicht existieren kann.

Die Russell'sche Antinomie zeigt, dass man der Bildung von Mengen Einschränkungen auferlegen muß, und dass die folgende von Georg Cantor[3] gegebene Definition des Begriffes „Menge" zu Widersprüchen führt:

> Unter einer „Menge" verstehen wir jede Zusammenfassung M von bestimmten wohl unterschiedenen Objekten unserer Anschauung oder unseres Denkens (welche die „Elemente" von M genannt werden) zu einem Ganzen. In Zeichen drücken wir dies so aus: $M = \{m\}$.

Um den mit einer inhaltlichen Definition des Mengenbegriffes unvermeidlichen Schwierigkeiten aus dem Wege zu gehen, wird in der axiomatischen Mengenlehre keine inhaltliche

[2] Russell, Bertrand Arthur William, 18. 5. 1872 – 2. 2. 1970, 1950 Nobelpreis für Literatur, verfolgte als Philosoph das Programm, die Mathematik auf Grundlage der Logik aufzubauen, Verfasser sozialkritischer Schriften, (en.wikipedia.ord/wiki/Bertrand_Russell).

[3] Cantor, Georg Ferdinand Ludwig Phillip, 3. 3. 1845 (St. Petersburg) – 6. 1. 1918 (Halle/Saale), Schöpfer der Mengenlehre (en.wikipedia.org/wiki/Georg_Cantor).

Definition des Mengenbegriffes gegeben, sondern es werden Axiome für die Relation „$a \in M$" und für die Bildung neuer Mengen aus gegebenen Mengen bereitgestellt. Im folgenden sind $A, B, \ldots$ meist Symbole für Mengen und $a, b, c, \ldots$ Symbole für deren Elemente. Dabei ist jedoch zu beachten, dass eine Menge selbst Element einer anderen Menge sein kann. Wir können hier nicht auf eines der existierenden Axiomensysteme der Mengenlehre eingehen und verweisen den Leser auf die angegebene Literatur. Im folgenden werden wir es immer mit „wohldefinierten" Mengen zu tun haben, d.h. es wird stets entweder „$a \in A$" oder „$a \notin A$" wahr sein. Ebenso werden die angegebenen Operationen, nach welchen aus vorgegebenen Mengen neue Mengen gebildet werden, zu keinen Widersprüchen führen.

Gilt „$x \in A$" genau für $x = a, b, c, \ldots$, so schreiben wir $A = \{a, b, c, \ldots\}$. Durch $\{1, 2, 3\}$, $\{2, 1, 3\}$, $\{2, 2, 1, 3\}$ beispielsweise wird ein und dieselbe Menge definiert.

Definition 1.1. *A und B seien Mengen.*
a) *A heißt* ***Teilmenge*** *von B, $A \subset B$, genau dann, wenn ein Element von A auch immer Element von B ist. B heißt dann* ***Obermenge*** *von A. Die Menge A heißt* ***echte Teilmenge*** *von B, $A \subsetneqq B$, genau dann, wenn $A \subset B$ gilt und mindestens ein Element von B nicht Element von A ist.*
b) *A ist gleich B, $A = B$, genau dann, wenn $A \subset B$ und $B \subset A$ gilt, d.h. wenn „$x \in A$" äquivalent zu „$x \in B$" ist (oder, was dasselbe ist, „$x \notin A$" äquivalent zu „$x \notin B$" ist).*

Die in Teil b) gegebene Definition der Gleichheit zweier Mengen tritt in der axiomatischen Mengenlehre als so genanntes Extensionalitätsaxiom auf und bedeutet, dass eine Menge alleine dadurch bestimmt ist, welche Objekte sie als Elemente enthält. Auch die folgende Behauptung ist bei axiomatischem Aufbau der Mengenlehre entweder ein Axiom oder ein Satz:

> Ist M eine Menge und ist P ein Prädikat, so existiert eine Teilmenge A von M, deren Elemente genau jene $x \in M$ sind, welche die Eigenschaft P besitzen. Wir schreiben dafür:
>
> $$A = \{x \in M \mid x \text{ besitzt die Eigenschaft } P\}.$$

Man beachte den Unterschied zur Cantor'schen Mengendefinition, in der für ein beliebiges Prädikat P die Existenz einer Menge M behauptet wird, für die „$x \in M$" genau dann wahr ist, wenn x die Eigenschaft P besitzt.

Insbesondere existiert zu einer Menge M stets die Teilmenge $\emptyset_M = \{x \in M \mid x \neq x\}$, die ***leere Teilmenge von*** M. Für beliebige Objekte x gilt $x \notin \emptyset_M$: Ist $x \notin M$, so gilt auch $x \notin \emptyset_M$ (weil $\emptyset_M \subset M$). Ist $x \in M$, so ist $x \neq x$ falsch, d.h. es gilt $x \notin \emptyset_M$.

Es seien nun $\emptyset_{M_1}$ und $\emptyset_{M_2}$ die leeren Teilmengen von zwei Mengen M_1 und M_2. Für beliebige Objekte x gilt sowohl $x \notin \emptyset_{M_1}$ als auch $x \notin \emptyset_{M_2}$. Dies bedeutet insbesondere $x \notin \emptyset_{M_1}$ ist äquivalent zu $x \notin \emptyset_{M_2}$, also $\emptyset_{M_1} = \emptyset_{M_2}$ nach Definition 1.1, b). Die folgende Definition ist daher sinnvoll:

Definition 1.2. *Die Menge $\emptyset$ mit $x \notin \emptyset$ für alle x heißt die* ***leere Menge****.*

Die obigen Überlegungen zeigen $\emptyset \subset M$ für beliebige Mengen M.

Definition 1.3. a) *$A \cap B := \{x \mid x \in A$ und $x \in B\}$ heißt der* ***Durchschnitt*** *von A und B. Die Mengen A und B heißen genau dann* ***disjunkt*** *oder* ***elementefremd****, wenn $A \cap B = \emptyset$ ist.*
b) *$A \cup B := \{x \mid x \in A$ oder $x \in B\}$ heißt die* ***Vereinigungsmenge*** *von A und B. (Die Existenz von $A \cup B$ muß in der axiomatischen Mengenlehre durch ein Axiom oder einen Satz gesichert werden.)*
c) *$A \setminus B := \{x \mid x \in A$ und $x \notin B\}$ heißt die* ***Differenzmenge*** *von A und B. Die Menge $(A \setminus B) \cup (B \setminus A)$ heißt die* ***symmetrische Differenz*** *von A und B.*

d) *Es sei $B \subset A$. Die Menge $\complement_A B := A \setminus B$ heißt das **Komplement** von B in A.*

Es gelten die folgenden Regeln:

$$A \cap A = A, \quad A \cup A = A \quad \text{(Idempotenz)},$$

$$(A \cap B) \cap C = A \cap (B \cap C),$$
$$(A \cup B) \cup C = A \cup (B \cup C) \quad \text{(Assoziativität)},$$

$$A \cap B = B \cap A, \quad A \cup B = B \cup A \quad \text{(Kommutativität)},$$

$$A \cap (A \cup B) = A, \quad A \cup (A \cap B) = A \quad \text{(Adjunktivität)},$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{(Distributivität)},$$

$$A \cap \emptyset = \emptyset, \quad A \cup \emptyset = A,$$

$$A \subset B \iff A \cap B = A \iff A \cup B = B.$$

In den folgenden Regeln sei $A \subset C$ und $B \subset C$:

$$\complement_C(A \cap B) = \complement_C A \cup \complement_C B,$$
$$\complement_C(A \cup B) = \complement_C A \cap \complement_C B \quad \text{(De Morgansche[4] Regeln)},$$

$$A \subset B \iff \complement_C B \subset \complement_C A,$$

$$\complement_C(\complement_C A) = A.$$

Definition 1.4. *Die Menge $\mathcal{P}(A) := \{B \mid B \subset A\}$ heißt die **Potenzmenge** von A.*

In der axiomatischen Mengenlehre wird die Existenz der Potenzmenge als Axiom postuliert.

Definition 1.5. a) *A_i, $i = 1, \dots, n$, seien nicht-leere Mengen. Für $a_i \in A_i$, $i = 1, \dots, n$, bezeichnet $(a_1, a_2, \dots, a_n)$ das **geordnete n-Tupel** mit a_i als i-tem Element. Für $n = 2$ heißt (a_1, a_2) das **geordnete Paar** mit a_1 als erstem und a_2 als zweitem Element.*
b) *$(a_1, \dots, a_m)$ sei ein geordnetes m-Tupel und $(b_1, \dots, b_n)$ ein geordnetes n-Tupel. Wir definieren:*

$$(a_1, \dots, a_m) = (b_1, \dots, b_n) \overset{\text{Df}}{\iff} m = n \quad \textit{und} \quad a_i = b_i,\ i = 1, \dots, n.$$

Man beachte den Unterschied zwischen dem geordneten n-Tupel $(a_1, \dots, a_n)$ und der Menge $\{a_1, \dots, a_n\}$.

Definition 1.6. *Gegeben seien die Mengen A_i, $i = 1, \dots, n$. Die Menge*

$$A_1 \times \cdots \times A_n := \{(a_1, \dots, a_n) \mid a_i \in A_i,\ i = 1, \dots, n\}$$

*heißt das **cartesische Produkt** von $A_1, \dots, A_n$. Gilt $A_i = A$, $i = 1, \dots, n$, so schreiben wir auch A^n an Stelle von $\underbrace{A \times \cdots \times A}_{n\text{-mal}}$.*

Die oben gegebene Definition des geordneten n-Tupels erweckt den Eindruck, dass es sich hierbei um einen neuen Begriff außerhalb der Mengenlehre handelt. Dass dies nicht der Fall ist, zeigen die folgenden Überlegungen, die wir der Einfachheit halber nur für $n = 2$

[4] De Morgan, Augustus, 27. 6. 1806 (Madura) – 18. 3. 1871 (London), Arbeiten zur Algebra und Logik (en.wikipedia.org/wiki/Augustus_De_Morgan).

durchführen. Es seien A_1, A_2 nicht-leere Mengen. Für $a_1 \in A_1$ und $a_2 \in A_2$ definieren wir die Menge
$$(a_1, a_2) := \big\{\{a_1\}, \{a_1, a_2\}\big\}$$
und nennen sie das geordnete Paar mit a_1 als erstem und a_2 als zweitem Element. Die Gleichheit geordneter Paare muß dann nicht definiert werden, sondern folgt aus der Gleichheit von Mengen:

Es gilt
$$\big\{\{a_1\}, \{a_1, a_2\}\big\} = \big\{\{b_1\}, \{b_1, b_2\}\big\} \tag{1.1}$$
genau dann, wenn
$$a_1 = b_1 \quad \text{und} \quad a_2 = b_2. \tag{1.2}$$

Beweis. Dass (1.1) aus (1.2) folgt, ist trivial. Es gelte (1.1).
Fall 1: $a_1 \neq a_2$.
In diesem Fall gilt $\{a_1\} \neq \{a_1, a_2\}$. Wegen (1.1) muß dann auch $\{b_1\} \neq \{b_1, b_2\}$ sein. Wäre $\{b_1\} = \{b_1, b_2\}$, so müßte $b_1 = b_2$ sein. Aus (1.1) würde $\{a_1, a_2\} \in \big\{\{b_1\}, \{b_1, b_2\}\big\} = \big\{\{b_1\}\big\}$ folgen, d.h. $\{a_1, a_2\} = \{b_1\}$. Dies wiederum würde $a_1 = a_2 = b_1$ in Widerspruch zu $a_1 \neq a_2$ ergeben. Wegen (1.1) muß $\{a_1\} = \{b_1\}$ und $\{a_1, a_2\} = \{b_1, b_2\}$ gelten, d.h. $a_1 = b_1$ und $a_2 = b_2$.
Fall 2: $a_1 = a_2$.
In diesem Fall gilt $\big\{\{a_1\}, \{a_1, a_2\}\big\} = \big\{\{a_1\}\big\}$. Wegen (1.1) folgt daraus $\{b_1\} = \{b_1, b_2\} = \{a_1\}$, d.h. $b_1 = b_2 = a_1$. $\square$

1.2 Relationen

Es seien zwei Mengen A und B gegeben. Unter einer ***Relation*** zwischen A und B verstehen wir ein Prädikat R, das jedem Paar $(a, b) \in A \times B$ zukommt oder nicht. Besitzt (a, b) die Eigenschaft R, so schreiben wir $a\,R\,b$ (in Worten: a steht zu b in der Relation R). Ist $A = B$, so sprechen wir von einer Relation auf A. Jeder Relation zwischen zwei Mengen A und B kann man eine Teilmenge $G_R \subset A \times B$ zuordnen:
$$G_R := \big\{(a, b) \in A \times B \mid a\,R\,b\big\}.$$
Umgekehrt entspricht jeder Teilmenge $G \subset A \times B$ eine Relation R_G zwischen A und B:

Für $a \in A$ und $b \in B$ gilt $a\,R_G\,b$ genau dann, wenn $(a, b) \in G$.

Eine spezielle Klasse von Relationen ist durch die so genannten Äquivalenzrelationen auf einer Menge A gegeben.

Definition 1.7. *Eine Relation R auf einer Menge A heißt **Äquivalenzrelation** auf A, wenn gilt:*

(i) *Für jedes $a \in A$ gilt $a\,R\,a$ (Reflexivität).*

(ii) *Für beliebige $a, b \in A$ folgt aus $a\,R\,b$ auch $b\,R\,a$ (Symmetrie).*

(iii) *Für beliebige $a, b, c \in A$ folgt aus $a\,R\,b$ und $b\,R\,c$ auch $a\,R\,c$ (Transitivität).*

Im Falle einer Äquivalenzrelation R schreibt man oft „$a \sim b$“ an Stelle von „$a\,R\,b$“ und spricht von der Äquivalenzrelation „ $\sim$ “.

Beispiel 1.8. $A = \mathbb{N}$ und $a \sim b$ bedeute, $a - b$ ist durch 2 teilbar. An Stelle von $a \sim b$ schreibt man $a \equiv b (\mathrm{mod} 2)$. $\Diamond$

Beispiel 1.9. A sei die Menge der gerichteten Strecken $\overrightarrow{p\,q}$ im $\mathbb{R}^2$ bzw. im $\mathbb{R}^3$. $\overrightarrow{p_1\,q_1} \sim \overrightarrow{p_2\,q_2}$ bedeute, $\overrightarrow{p_1\,q_1}$ kann durch eine Parallelverschiebung in $\overrightarrow{p_2\,q_2}$ übergeführt werden. $\Diamond$

Beispiel 1.10. A sei die Menge aller Dreiecke Δ im $\mathbb{R}^2$. $\Delta_1 \sim \Delta_2$ bedeute, Δ_1 und Δ_2 sind ähnlich. $\Diamond$

Es sei R eine Äquivalenzrelation auf A. Für beliebiges $a \in A$ heißt die Teilmenge $K(a) := \{b \in A \mid a\,R\,b\}$ die ***Äquivalenzklasse*** von a. Man überzeugt sich leicht davon, dass für Äquivalenzklssen die folgenden Aussagen gelten:

$$a \in K(a) \quad \text{für alle } a \in A. \tag{1.3}$$

$$a\,R\,b \iff K(a) = K(b). \tag{1.4}$$

$$K(a) \cap K(b) \neq \emptyset \iff K(a) = K(b). \tag{1.5}$$

Beweis. Die Aussage (1.3) folgt unmittelbar aus der Reflexivität von R. Bezüglich der Aussage (1.4) bemerken wir zunächst, dass aus $K(a) = K(b)$ vermittels (1.3) $b \in K(a)$ folgt, d.h., es gilt $a\,R\,b$. Gilt umgekehrt $a\,R\,b$ und ist $c \in K(a)$, so folgt zunächst $a\,R\,c$. Wegen der Transitivität folgt weiter $b\,R\,c$ (beachte, dass wegen der Symmetrie $b\,R\,a$ gilt), d.h., $c \in K(b)$. Damit ist $K(a) \subset K(b)$ gezeigt. Analog zeigt man $K(b) \subset K(a)$. Also gilt $K(a) = K(b)$. Damit ist (1.4) bewiesen.

Bei der Aussage (1.5) ist die Implikation „$\Longleftarrow$" trivial. Um die Implikation „$\Longrightarrow$" zu beweisen, nehmen wir an, dass $K(a) \cap K(b) \neq \emptyset$ sei. Es gibt also ein $c \in A$ mit $a\,R\,c$ und $b\,R\,c$. Auf Grund der Symmetrie und Transitivität von R folgt $a\,R\,b$. Wegen (1.4) bedeutet dies $K(a) = K(b)$. $\square$

Wir bezeichnen mit $\mathcal{K}_R := \{K(a) \mid a \in A\}$ die Menge aller aller Äquivalenzklassen von R. Aus den Aussagen (1.3) – (1.5) folgen unmittelbar die folgenden Eigenschaften von $\mathcal{K}_R$:

Ist $A \neq \emptyset$ und R eine Äquivalenzrelation auf A, dann gilt:

(i) $K \neq \emptyset$ für alle $K \in \mathcal{K}_R$.

(ii) $K_1 \neq K_2 \iff K_1 \cap K_2 = \emptyset$ für $K_1, K_2 \in \mathcal{K}_R$.

(iii) $A = \bigcup_{K \in \mathcal{K}_R} K$[5]

Definition 1.11. *Es sei $A \neq \emptyset$ und $\mathcal{K} \subset \mathcal{P}(A)$. $\mathcal{K}$ heißt eine* ***Klasseneinteilung*** *oder* ***Partition*** *von A, wenn für $\mathcal{K}$ die Eigenschaften* (i) – (iii) *von oben gelten. Für $K \in \mathcal{K}$ heißt jedes Element $a \in K$ ein* ***Repräsentant*** *der Klasse K. Eine Teilmenge V von A heißt ein* ***vollständiges Repräsentantensystem*** *für $\mathcal{K}$, wenn V zu jedem $K \in \mathcal{K}$ genau ein Element $a \in K$ enthält und sonst keine Elemente von A.*

Wie wir gesehen haben, bilden die Äquivalenzklassen einer Äquivalenzrelation R auf einer nichtleeren Menge eine Klasseneinteilung dieser Menge. Wenn umgekehrt eine Klasseneinteilung $\mathcal{K}$ von $A \neq \emptyset$ gegeben ist, so wird durch

$$a \sim b \overset{\text{Df}}{\iff} a, b \in K \quad \text{für ein } K \in \mathcal{K}$$

eine Äquivalenzrelation „ $\sim$ " auf A definiert (Beweis als Übung).

Eine weitere wichtige Klasse von Relationen ist durch die Ordnungsrelationen gegeben.

Definition 1.12. *R sei ein Relation auf der Menge A. R heißt genau dann ein* ***Ordnungsrelation*** *auf A, wenn gilt:*

[5] $\bigcup_{K \in \mathcal{K}_R} K = \{a \mid \text{Es existiert ein } K \in \mathcal{K}_R \text{ mit } a \in K\}$.

(i) *Es gilt* aRa *für alle* $a \in A$ *(Reflexivität).*

(ii) *Für alle* $a, b \in A$ *gilt:* aRb *und* $bRa \iff a = b$.

(iii) *Für alle* $a, b, c \in A$ *gilt:* aRb *und* $bRc \iff aRc$ *(Transitivität).*

*R heißt eine **Totalordnung** oder **lineare Ordnung** auf A, wenn zusätzlich gilt:*

(iv) *Für alle* $a, b \in A$ *gilt* aRb *oder* bRa.

*Die Menge A heißt eine **geordnete Menge** bzw. **total geordnete Menge**, wenn auf A eine Ordnungsrelation bzw. eine Totalordnung definiert ist.*

Beispiel 1.13. Es sei $A = \mathbb{R}$ und aRb bedeute $a \leq b$. Dann ist R eine Totalordnung auf $\mathbb{R}$. $\Diamond$

Beispiel 1.14. Es sei M eine Menge. Für beliebige Teilmengen A, B von M erklären wir

$$A\,R\,B \quad \overset{\text{Df}}{\iff} \quad A \subset B.$$

Dann ist R eine Ordnungsrelation auf der Potenzmenge P von M, aber im allgemeinen keine Totalordnung. $\Diamond$

1.3 Abbildungen

Es seien X und Y nicht-leere Mengen. Unter einer ***Abbildung*** von X nach Y versteht man eine Vorschrift f, die jedem $x \in X$ genau ein $y \in Y$ als ***Bild*** zuordnet, $y = f(x)$. Man verwendet die Schreibweise

$$\begin{aligned} f :& X \to Y, \\ & x \mapsto f(x) \in Y. \end{aligned}$$

Die Menge X heißt der ***Definitionsbereich*** von f, während Y der ***Bildbereich*** oder ***Wertebereich*** genannt wird.

Jeder Abbildung $f : X \to Y$ kann eine Teilmenge $G_f \subset X \times Y$ zugeordnet werden:

$$G_f = \big\{(x, y) \in X \times Y \mid y = f(x)\big\}.$$

Die Menge G_f heißt der ***Graph*** der Abbildung f. Man überzeugt sich leicht davon, dass eine Teilmenge $G \subset X \times Y$ genau dann Graph einer Abbildung f_G ist, wenn gilt:

1. Für alle $x \in X$ existiert ein $y \in Y$ mit $(x, y) \in G$.
2. Aus $(x, y_1) \in G$ und $(x, y_2) \in G$ folgt $y_1 = y_2$.

Die Abbildung f_G ist wie folgt definiert: $f_G(x) = y$, wobei y das eindeutig bestimmte Element in Y ist mit $(x, y) \in G$. Der Graph von f_G ist durch G gegeben. Umgekehrt ist die durch G_f definierte Abbildung wiederum f:

$$f_{G_f} = f \quad \text{und} \quad G_{f_G} = G.$$

Es besteht somit eine ein-eindeutige Zuordnung zwischen Abbildungen von X nach Y und jenen Teilmengen von $X \times Y$, welche die Eigenschaften 1. und 2. von oben besitzen. Man kann daher Abbildungen $X \to Y$ einfach als spezielle Teilmengen von $X \times Y$ auffassen.

Definition 1.15. *Es seien die Abbildungen* $f : X \to Y$ *und* $g : \tilde{X} \to \tilde{Y}$ *gegeben. Die Abbildungen* f *und* g *sind genau dann gleich,* $f = g$*, wenn gilt:*

(i) $X = \tilde{X}$ *und* $Y = \tilde{Y}$.

(ii) *Für alle* $x \in X = \tilde{X}$ *gilt* $f(x) = g(x)$.

Man beachte, dass man neben den naheliegenden Forderungen nach Gleichheit der Definitionsbereiche und der Zuordnungsvorschrift auch die Gleichheit der Bildbereiche fordert.

Definition 1.16. *Die nicht-leeren Mengen* X, Y *und die Abbildung* $f : X \to Y$ *seien gegeben.*
a) *Die Abbildung* f *heißt **injektiv**, wenn aus* $f(x_1) = f(x_2)$ *stets* $x_1 = x_2$ *folgt.*
b) *Die Abbildung* f *heißt **surjektiv**, wenn es für alle* $y \in Y$ *ein* $x \in X$ *gibt mit* $y = f(x)$.
c) *Die Abbildung* f *heißt **bijektiv**, wenn* f *injektiv und surjektiv ist.*

Statt "Die Abbildung f ist injektiv" sagt man auch "Die Abbildung f ist ein-eindeutig". Anstelle von "f ist surjektiv" verwendet man auch "f ist eine Abbildung auf Y".

Beispiel 1.17. X und Y seien endliche, nicht-leere Mengen, d.h. Mengen mit endliche vielen Elementen und es sei $f : X \to Y$ gegeben. Ist die Anzahl der Elemente in X und Y gleich, so gilt:

$$f \text{ injektiv} \quad \Longleftrightarrow \quad f \text{ surjektiv}$$

$\diamondsuit$

Beispiel 1.18. X sei eine nicht-leere Menge. Die durch $x \mapsto x$, $x \in X$ definierte Abbildung $id_X : X \to X$ heißt die ***identische Abbildung*** auf X. $\diamondsuit$

Definition 1.19. *Die nicht-leeren Mengen* X, Y, *die Abbildung* $f : X \to Y$ *und die Mengen* $A \subset X$, $B \subset Y$ *seien gegeben.*
a) *Die Menge*

$$f(A) = \big\{ f(x) \mid x \in A \big\} \subset Y$$

*heißt die **Bildmenge** von* A *unter der Abbildung* f.
b) *Die Menge*

$$f^{-1}(B) = \big\{ x \in X \mid f(x) \in B \big\} \subset X$$

*heißt die **Urbildmenge** von* B *unter der Abbildung* f.

Ist $B = \{y\}$ für ein $y \in Y$, so ist $f^{-1}(\{y\}) = \{x \in X \mid f(x) = y\}$. Für eine Menge $B \subset Y$ mit $B \cap f(X) = \emptyset$ gilt $f^{-1}(B) = \emptyset$. Wir haben die zwei folgenden Aussagen, deren Beweis eine einfache Übung ist:

$$\begin{aligned} f \text{ injektiv} \quad &\Longleftrightarrow \quad \text{Für alle } y \in f(X) \text{ ist } f^{-1}(\{y\}) \text{ eine Menge mit nur einem Element.} \\ f \text{ surjektiv} \quad &\Longleftrightarrow \quad f(X) = Y. \end{aligned}$$

Die Abbildung $f : X \to Y$ sei bijektiv. Wir definieren eine Abbildung $g : Y \to X$ durch die Vorschrift: Für $y \in Y$ definieren wir

$$g(y) = x, \quad \text{wobei } x \in X \text{ ein Element mit } f(x) = y \text{ ist.}$$

Die Abbildung g ist wohldefiniert. Denn wegen der Surjektivität von f existiert zu jedem $y \in Y$ ein $x \in X$ mit $f(x) = y$ und wegen der Injektivität von f ist dieses x eindeutig bestimmt. Die Abbildung g heißt die zu f ***inverse Abbildung***, $g = f^{-1}$.

Beispiel 1.20. Es sei $f : X \in Y$ injektiv, aber nicht surjektiv, d.h. $f(X) \subsetneq Y$. Wir definieren eine Abbildung $\tilde{f} : X \to f(X)$ durch $\tilde{f}(x) = f(x)$ für alle $x \in X$. Dann ist $\tilde{f}$ bijektiv und besitzt somit eine inverse Abbildung $\tilde{f}^{-1}$, während zu f keine inverse Abbildung existiert. Dies zeigt, dass für eine Abbildung f nicht nur die Angabe des Urbildbereiches X und der Zuordnungsvorschrift $x \mapsto f(x)$ wichtig ist sondern auch die Angabe des Bildbereiches Y. $\diamondsuit$

Es seien die Abbildung $f : X \to Y$ und eine Teilmenge $A \subset X$ gegeben. Unter der ***Einschränkung*** von f auf A verstehen wir die Abbildung $f\big|_A : A \to Y$, die durch $f\big|_A (x) = f(x)$, $x \in A$, definiert ist.

Es ist wichtig zwischen der Bildung des Urbildes $f^{-1}(\cdot)$ und dem Bild $f^{-1}(\cdot)$ eines Elementes unter der inversen Abbildung zu f zu unterscheiden. Wenn das Argument von f^{-1} eine Teilmenge $B \subset Y$ ist, bedeutet $f^{-1}(B)$ die Urbildmenge von B. Wenn das Argument von f^{-1} ein einzelnes Element $y \in Y$ ist, setzen wir voraus, dass die inverse Abbildung zu f existiert, und verstehen unter $x = f^{-1}(y)$ das Bild von y unter der Abbildung f^{-1}. In diesem Fall ist dies von $f^{-1}(\{y\}) = \{x\}$ zu unterscheiden.

Definition 1.21. *Es seien X, Y und Z nicht-leere Mengen und $f : X \to Y$, $g : Y \to Z$ Abbildungen. Die durch $h(x) = g(f(x))$, $x \in X$, definierte Abbildung $X \to Z$ heißt das* ***Kompositum*** *der Abbildungen f und g, $h = g \circ f$*

Satz 1.22. *Es seien die nicht-leeren Mengen X, Y, Z, W und die Abbildungen $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$ gegeben. Dann gilt:*

a) *Es ist $h \circ (g \circ f) = (h \circ g) \circ f$.*

b) *Es gilt $f \circ id_X = f$ und $id_Y \circ f = f$.*

c) *Ist f invertierbar, so gilt*

$$f^{-1} \circ f = id_X \quad \textit{und} \quad f \circ f^{-1} = id_Y.$$

Beweis. Die Aussage a) folgt unmittelbar aus der Definition des Kompositums: Für alle $x \in X$ ist $\big(h \circ (g \circ f)\big)(x) = h\big((g \circ f)(x)\big) = h(g(f(x)))$ und $\big((h \circ g) \circ f\big)(x) = (h \circ g)(f(x)) = h(g(f(x)))$.

Die Aussagen unter b) sind unmittelbar einsichtig. Die Aussagen unter c) folgen direkt aus der Definition der inversen Abbildung. □

Eine Charakterisierung von Injektivität und Surjektivität (und damit auch Bijektivität) gibt der folgende Satz:

Satz 1.23. *Es seien die nicht-leeren Mengen X, Y, Z und die Abbildungen $f : X \to Y$, $g : Y \to Z$ gegeben. Dann gilt:*

a) *f ist genau dann injektiv, wenn es eine Abbildung $h : Y \to X$ gibt mit $h \circ f = id_X$.*

b) *Sind f und g injektiv, so gilt dies auch für $g \circ f$.*

c) *f ist genau dann surjektiv, wenn es eine Abbildung $h : Y \to X$ gibt mit $f \circ h = id_Y$.*

d) *Sind f und g surjektiv, so gilt dies auch für $g \circ f$.*

e) *f ist genau dann bijektiv, wenn es eine Abbildung $h : Y \to X$ gibt mit $h \circ f = id_X$ und $f \circ h = id_Y$. Falls diese Bedingung erfüllt ist, gilt $h = f^{-1}$.*

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

g) *Sind f und g invertierbar, so auch $g \circ f$ und es gilt*

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}.$$

Beweis. a) Es sei $f(x_1) = f(x_2)$ für $x_1, x_2 \in X$. Dann gilt $x_1 = (h \circ f)(x_1) = h(f(x_1))$ und $x_2 = (h \circ f)(x_2) = h(f(x_2))$. Andererseits folgt aus $f(x_1) = f(x_2)$ sofort $h(f(x_1)) = h(f(x_2))$ und somit $x_1 = x_2$, d.h. f ist injektiv.

Es sei nun f injektiv. Mit einem beliebigen Element $x_0 \in X$ setzen wir

$$h(y) = \begin{cases} x_0, & \text{falls } y \neq f(X), \\ x, & \text{wobei } x \in X \text{ gewählt wird mit } y = f(x), \text{ falls } y \in f(X). \end{cases}$$

Man beachte, dass für $y \in f(X)$ das Element $x \in X$ mit $f(x) = y$ eindeutig bestimmt ist. Man prüft leicht nach, dass $h \circ f = id_X$ gilt.

b) Es seien f und g injektiv. Nach a) existieren Abbildungen $h_1 : Y \to X$ und $h_2 : Z \to Y$ mit $h_1 \circ f = id_X$ und $h_2 \circ g = id_Y$. Dann gilt, für $h = h_1 \circ h_2 : Z \to X$,

$$h \circ (g \circ f) = h_1 \circ ((h_2 \circ g) \circ g) = h_1 \circ (id_Y \circ f) = h_1 \circ f = id_X,$$

d.h. die Abbildung $g \circ f$ ist injektiv.

c) Es sei $y \in Y$ gegeben. Dann ist $y = f(g(y))$, d.h. $y = f(x)$ mit $x = g(y)$. f ist somit surjektiv.

Es nun umgekehrt f surjektiv. Daraus folgt, dass $f^{-1}(\{y\}) \neq \emptyset$ für alle $y \in Y$. Für jedes $y \in Y$ wählen wir ein Element $h(y) \in f^{-1}(\{y\})$. Die durch $y \mapsto h(y)$ definierte Abbildung $Y \to X$ erfüllt $f \circ h = id_Y$.

d) Der Beweis für die Aussage d) des Satzes ist analog dem für die Aussage b) gegebenen Beweis.

e) Existiert eine Abbildung $h : Y \to X$ mit $h \circ f = id_X$ und $f \circ h = id_Y$, so ist f wegen der bereits bewiesenen Aussagen a) und c) sowohl injektiv als auch surjektiv und daher bijektiv. Ist f bijektiv so existieren wegen a) und c) Abbildungen $h_1, h_2 : Y \to X$ mit

$$h_1 \circ f = id_X \quad \text{und} \quad f \circ h_2 = id_Y. \tag{1.6}$$

Da f bijektiv ist existiert die inverse Abbildung $f^{-1} : Y \to X$. Aus (1.6) folgt daher $h_1 = (h_1 \circ f) \circ f^{-1} = id_X \circ f^{-1} = f^{-1}$. Analog sieht man $h_2 = f^{-1}$. Somit ist $h_1 = h_2 = f^{-1}$.

f) Ist f invertierbar, so folgt aus Satz 1.22

$$f^{-1} \circ f = id_X \quad \text{und} \quad f \circ f^{-1} = id_Y.$$

Aus Aussage e) mit $h = f$ folgt, dass f^{-1} invertierbar ist und $(f^{-1})^{-1} = f$ gilt.

g) Setzen wir $h = f^{-1} \circ g^{-1}$, so folgt

$$(g \circ f) \circ h = id_X \quad \text{und} \quad h \circ (g \circ f) = id_Z.$$

Aus Aussage e) für folgt, dass die Abbildung $g \circ f$ bijektiv und somit invertierbar ist. Ferner gilt $(g \circ f)^{-1} = h = f^{-1} \circ g^{-1}$. $\square$

Abbildungen spielen auch eine wichtige Rolle, wenn man Mengen in Hinblick auf die Anzahl ihrer Elemente vergleichen will.

Definition 1.24. *A und B seien nicht-leere Mengen.*

a) *A und B heißen* ***gleich-mächtig*** *genau dann, wenn es eine bijektive Abbildung $f : A \to B$ gibt.*

b) *Die Menge A heißt* ***endlich*** *genau dann, wenn für ein $n \in \mathbb{N}$ die Mengen A und $\{1, \dots, n\}$ gleich-mächtig sind. Die Menge A heißt* ***unendlich*** *genau dann, wenn sie nicht endlich ist.*

c) *Die Menge A heißt* ***abzählbar unendlich*** *genau dann, wenn A und $\mathbb{N}$ gleich-mächtig sind.*

d) *Die Menge A heißt* ***abzählbar*** *genau dann, wenn A endlich oder abzählbar unendlich ist.*

Ist $\mathcal{S}$ eine Menge, deren Elemente selbst Mengen sind, so ist die Gleichmächtigkeit eine Äquivalenzrelation auf $\mathcal{S}$.

Beispiel 1.25. Die Menge $\mathbb{N}_0 = \mathbb{N} \cup 0$ ist abzählbar unendlich. Die durch $f(n) = n+1$, $n \in \mathbb{N}_0$ definierte Abbildung $f : N_0 \to \mathbb{N}$ ist bijektiv. $\Diamond$

Beispiel 1.26. Die Menge $\mathbb{Z}$ der ganzen Zahlen ist abzählbar unendlich. Eine bijektive Abbildung $g : \mathbb{Z} \to \mathbb{N}_0$ wird durch

$$g(k) = \begin{cases} 2k, & \text{für } k \geq 0, \\ -2k-1, & \text{für } k < 0. \end{cases}$$

definiert. Mit f aus Beispiel 1.25 ist $f \circ g$ eine bijektive Abbildung $\mathbb{Z} \to \mathbb{N}$. $\Diamond$

Beispiel 1.27. Ist $A \neq \emptyset$ Teilmenge einer abzählbaren Menge M, so ist A ebenfalls abzählbar. Laut Annahme über M existiert eine bijektive Abbildung $M \to \mathbb{N}$.
Fall 1: $f(A)$ ist endlich. Dann ist auch A endlich, da $f\,|_A$ eine bijektive Abbildung $A \to f(A)$ ist.
Fall 2: $f(A)$ ist unendlich. Auch hier ist $f\,|_A$ eine bijektive Abbildung $A \to f(A)$. Als Teilmenge von $\mathbb{N}$ besitzt $f(A)$ ein kleinstes Element $n_1 \in \mathbb{N}$. Die Menge $f(A) \setminus \{n_1\}$ ist nicht leer und besitzt ein kleinstes Element $n_2 > n_1$ in $\mathbb{N}$. Auf diese Weise fortfahrend erhält man Elemente $n_1 < n_2 < \cdots$ in $\mathbb{N}$. Für jedes n_k ist die Menge $f(A) \setminus \{n_1, \dots, n_k\}$ nicht-leer, da andernfalls $f(A)$ endlich wäre. Somit existiert stets ein $n_{k+1} > n_k$. Es ist

$$f(A) = \big\{ n_k \mid k \in \mathbb{N} \big\}.$$

Auf Grund der Konstruktion ist klar, dass $\big\{ n_k \mid k \in \mathbb{N} \big\} \subset f(A)$ gilt. Es sei $m \in f(A)$ gewählt. Da stets $n_k < n_{k+1}$ gilt existiert ein k_0 mit $m < n_{k_0}$. Auf Grund der Konstruktion der Menge $\big\{ n_k \mid k \in \mathbb{N} \big\}$ muss $m = n_k$ für ein $k < k_0$ gelten.

Die durch $k \mapsto n_k$ definierte Abbildung $g : \mathbb{N} \to f(A)$ ist bijektiv. Somit ist $g^{-1} \circ f\,|_A$ eine bijektive Abbildung $A \to \mathbb{N}$, d.h. A ist abzählbar unendlich. $\Diamond$

Später werden wir das folgende Resultat benötigen:

Satz 1.28. *Es seien $A_1, \dots, A_n$ abzählbare Mengen. Dann ist auch $A_1 \times \cdots \times A_n$ abzählbar.*

Beweis. Wir beweisen die Aussage des Satzes zunächst für $n = 2$. Auf Grund der Voraussetzung über die Mengen A_1 und A_2 ist

$$\begin{aligned} A_1 &= \{a_1, \dots, a_\ell\} \quad \text{oder} \quad A_1 = \{a_1, a_2, \dots\} \\ A_2 &= \{b_1, \dots, b_m\} \quad \text{oder} \quad A_2 = \{b_1, b_2, \dots\}. \end{aligned}$$

Wir betrachten im Folgenden nur den Fall, dass A_1 und A_2 unendlich sind, und ordnen die Elemente von $A_1 \times A_2$ in einem zweidimensionalen Schema an:

$$\begin{tikzcd}
(a_1,b_1) \arrow[r] & (a_1,b_2) \arrow[dl] & (a_1,b_3) \arrow[r] & (a_1,b_4) \arrow[dl] & (a_1,b_5) \arrow[r] & \cdots \\
(a_2,b_1) \arrow[d] & (a_2,b_2) \arrow[ur] & (a_2,b_3) \arrow[dl] & (a_2,b_4) \arrow[ur] & (a_2,b_5) & \\
(a_3,b_1) \arrow[ur] & (a_3,b_2) \arrow[dl] & (a_3,b_3) \arrow[ur] & (a_3,b_4) & (a_3,b_5) & \\
(a_4,b_1) \arrow[d] & (a_4,b_2) \arrow[ur] & (a_4,b_3) & (a_4,b_4) & (a_4,b_5) & \\
(a_5,b_1) \arrow[ur] & (a_5,b_2) & (a_5,b_3) & (a_5,b_4) & (a_5,b_5) &
\end{tikzcd}$$

Die Abbildung $f : \mathbb{N} \to A_1 \times A_2$ ist in dem oben dargestellten Schema durch Pfeile angedeutet, d.h., $f(1) = (a_1, b_1)$, $f(2) = (a_1, b_2)$, $f(3) = (a_2, b_1)$, $f(4) = (a_3, b_1)$, $\ldots$. Es ist klar, dass f bijektiv ist. Der Fall A_1 endlich oder A_2 endlich sollte nun keine Schwierigkeiten bereiten.

Der Fall $n = 3$ wird auf den Fall $n = 2$ zurückgeführt, indem man beobachtet, dass ein geordnetes Tripel $(a_1, a_2, a_3) \in A_1 \times A_2 \times A_3$ auch als Paar $(a_1, (a_2, a_3)) \in A_1 \times (A_2 \times A_3)$ aufgefaßt werden kann. Auf analoge Weise kann man allgemein den Fall n auf den Fall $n - 1$ zurückführen. $\square$

Ein Begriff, der später immer wieder verwendet wird, ist der Begriff der „Familie“ von Elementen einer Menge. Es seien J und A nicht-leere Mengen. Wir nennen eine Abbildung $J \to A$ mit $\iota \in J \mapsto a_\iota \in A$ eine ***Familie*** von Elementen aus A mit ***Indexmenge*** J. Eine kompakte Schreibweise dafür ist: $\big(a_\iota\big)_{\iota \in J} \subset A$.

Eine Familie mit Indexmenge $J = \mathbb{N}$ nennt man eine ***Folge*** in A: $\big(a_n\big)_{n\in\mathbb{N}}$ oder $\big(a_n\big)_{n=1,2,\ldots}$.

Literatur:

Halmos, P.: „*Naive Set Theory*“, Van Nostrand, Princeton 1960.
Meschkowski, H.: „*Hundert Jahre Mengenlehre*“, dtv, München 1973.

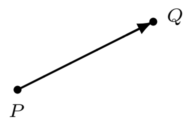
Kapitel 2

Vektorrechnung in der Ebene und im Raum

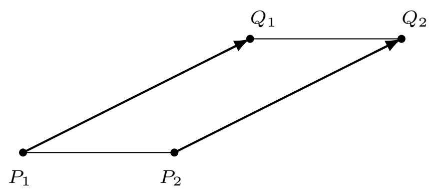
In diesem einleitenden Kapitel werden einige grundlegende Konzepte der linearen Algebra und analytischen Geometrie auf sehr anschauliche Weise eingeführt und diskutiert. Vielfach wird in heuristischer Form argumentiert und Bezug auf geometrische Anschauung genommen. Es wird beispielsweise in diesem Kapitel nicht versucht, eine Definition der euklidischen Räume $\mathbb{R}^2$ oder $\mathbb{R}^3$ zu geben. Auf diese Weise soll dem Anfänger der Einstieg in die abstrakte Behandlung des Stoffes, wie sie notwendigerweise in den späteren Kapiteln erfolgen muß, erleichtert werden. Dem Leser wird dringend empfohlen, während des Studiums der späteren Kapitel immer wieder die Darstellung des Stoffes dort mit der Darstellung in diesem Kapitel zu vergleichen.

2.1 Vektoren, Vektoraddition und skalare Multiplikation

Im folgenden bezeichnen $P, Q, \ldots$ Punkte im $\mathbb{R}^2$ bzw. $\mathbb{R}^3$. Unter der ***gerichteten Strecke*** von P nach Q verstehen wir das geordnete Paar (P, Q). Geometrisch werden wir die gerichtete Strecke (P, Q) durch einen von P nach Q führenden Pfeil veranschaulichen:



Zwei gerichtete Strecken (P_1, Q_1) und (P_2, Q_2) heißen ***verschiebungsgleich***, $(P_1, Q_1) \sim (P_2, Q_2)$, wenn es eine Parallelverschiebung gibt, die P_2 in P_1 und Q_2 in Q_1 überführt.



Man prüft leicht nach, dass „verschiebungsgleich" eine Äquivalenzrelation auf der Menge der gerichteten Strecken ist. Zu dieser Äquivalenzrelation gibt es eine Klasseneinteilung der Menge der gerichteten Strecken in Äquivalenzklassen.

Definition 2.1. *Unter einem* ***Vektor*** *(im $\mathbb{R}^2$ oder $\mathbb{R}^3$) verstehen wir eine Äquivalenzklasse gerichteter Strecken. Wir bezeichnen Vektoren meist mit Kleinbuchstaben. Ist a ein Vektor und (P, Q) ein Repräsentant von a, so schreiben wir $a = \overrightarrow{PQ}$.*

Unmittelbar einzusehen ist, dass den Vektoren im $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ eindeutig die Parallelverschiebungen des $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ entsprechen.

Als ***Norm*** eines Vektors a definieren wir die Länge eines beliebigen seiner Repräsentanten:

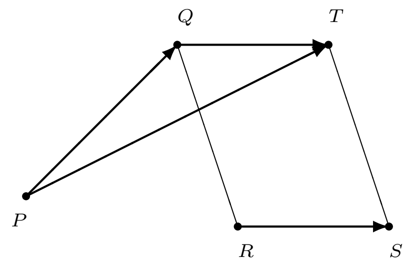
$$\|a\| = \text{ Länge von } (P, Q), \text{ wobei } a = \overrightarrow{PQ}.$$

Diese Definition ist sinnvoll, da offensichtlich alle Repräsentanten des Vektors a dieselbe Länge haben.

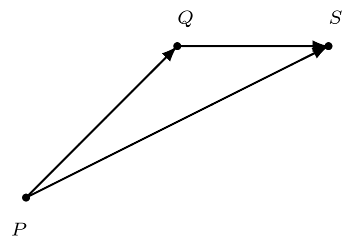
In der folgenden Definition erklären wir eine algebraische Operation in der Menge der Vektoren (im $\mathbb{R}^2$ bzw. $\mathbb{R}^3$) , die wir als Addition auffassen:

Definition 2.2. *a und b seien Vektoren im $\mathbb{R}^2$ bzw. $\mathbb{R}^3$. Unter der* ***Summe*** *$c = a + b$ von a und b verstehen wir jenen Vektor c, der wie folgt konstruiert wird:*

Sind (P, Q) und (R, S) Repräsentanten von a bzw. b, so ist $c = \overrightarrow{PT}$, wobei in der Zeichnung (Q, T) jene gerichtete Strecke ist, die man erhält, indem man die Strecke (R, S) so parallelverschiebt, dass der Punkt R in den Punkt Q übergeführt wird.



Man macht sich leicht klar, dass die oben gegebene Definition der Summe zweier Vektoren unabhängig von der speziellen Wahl der Repräsentanten ist. Wir werden daher meist Repräsentanten der Form (P, Q) und (Q, S) für a bzw. b wählen. (P, S) ist dann ein Repräsentant des Summenvektors.



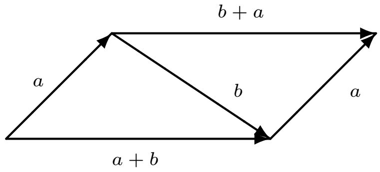
Auf Grund der geometrischen Konfiguration sagt man, dass die Vektoraddition entsprechend der ***Parallelogrammregel*** erfolgt.

Die Vektoraddition ist ***kommutativ****, d.h. für je zwei Vektoren gilt:*

$$a + b = b + a.$$

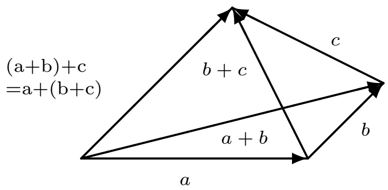
Dies entnimmt man sofort der folgenden Zeichnung[6]:

[6]Im folgenden werden wir häufig in den Zeichnungen Repräsentanten als Vektoren bezeichnen.



*Die Vektoraddition ist **assoziativ**, d.h. für je drei Vektoren gilt:*

$$a + (b + c) = (a + b) + c.$$



Der durch einen beliebigen Punkt P bestimmte Vektor $o = \overrightarrow{PP}$ heißt ***Nullvektor***.

Für beliebige Vektoren a gilt offensichtlich:

$$a + o = o + a = a. \tag{2.1}$$

Es gibt keinen anderen, von o verschiedenen Vektor o' mit dieser Eigenschaft, d.h. o ist eindeutig durch (2.1) bestimmt. Dies kann man bereits beweisen, ohne die Definition von Vektoren zu benutzen:

$$\text{Aus} \quad o + o' = o' \quad \text{und} \quad o + o' = o \quad \text{folgt} \quad o = o'.$$

Ist $a = \overrightarrow{PQ}$ und $b = \overrightarrow{QP}$, so gilt offensichtlich $a + b = b + a = o$. Damit haben wir gezeigt:

Zu jedem Vektor a existiert ein Vektor $-a$ mit der Eigenschaft

$$a + (-a) = (-a) + a = o. \tag{2.2}$$

$-a$ heißt der zu a ***negative Vektor***. Statt $b + (-a)$ schreiben wir meist $b - a$. Der Vektor $-a$ ist eindeutig durch (2.2) bestimmt. (Es sei b ein zweiter Vektor mit der Eigenschaft (2.2). Dann ist $a + b = o$ und daher $-a = (-a) + (a + b) = (-a + a) + b = o + b = b$. Hier wurden (2.1) und das Assoziativgesetz benutzt.) Aus (2.2) folgt unmittelbar

$$-(-a) = a.$$

Zusammenfassend können wir feststellen, dass die Vektoraddition kommutativ und assoziativ ist. Ferner gibt es ein neutrales Element, den Nullvektor, und zu jedem Vektor a den negativen Vektor $-a$.

Diesen Sachverhalt kann man auch kurz wie folgt ausdrücken (siehe Abschnitt 4.1):

*Die Menge der Vektoren im $\mathbb{R}^2$ bzw. im $\mathbb{R}^3$ bildet eine **abelsche**[7] **Gruppe**.*

[7] Abel, Niels Henrik, 5. 8. 1802 (Finnøy bei Stavanger) – 6. 4. 1829 (Eisenwerk Froland bei Arendal), Beiträge zur Algebra, algebraische Funktionen, elliptische Funktionen, Integrale (en.wikipedia.org/wiki/Niels_Henrik_Abel).

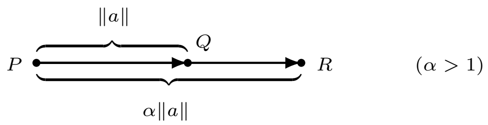
Auf Grund der Assoziativität und Kommutativität gilt beispielsweise:

$$-(a+b) = (-a) + (-b). \tag{2.3}$$

Dies folgt aus $((-a)+(-b))+(a+b) = ((-b)+((-a)+a)+b) = (-b)+o+b = (-b)+b = o$ und der eindeutigen Bestimmtheit des negativen Elementes.

Ist α eine reelle Zahl und a ein Vektor, so können wir den Vektor αa wie folgt definieren:

a) *Es sei $\alpha > 0$. Ist $a = \overrightarrow{PQ}$, so ist $\alpha a = \overrightarrow{PR}$, wobei $\|\alpha a\| = \alpha\|a\|$ ist und (P,Q) und (P,R) dieselbe Richtung haben.*



b) *Für $\alpha = 0$ setzen wir*

$$\alpha a = o.$$

c) *Für $\alpha < 0$ definieren wir (man beachte, dass $-\alpha > 0$ ist)*

$$\alpha a = (-\alpha)(-a). \tag{2.4}$$

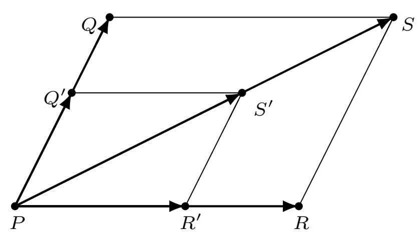
Damit haben wir in eindeutiger Weise jedem Paar bestehend aus einer reellen Zahl α und einem Vektor a einen weiteren Vektor zugeordnet, den wir als Produkt αa schreiben. Diese algebraische Operation heißt ***skalare Multiplikation***. Die reellen Zahlen werden hier zum Unterschied zu den Vektoren ***Skalare*** genannt.

Für $\alpha > 0$ ist $-\alpha < 0$ und wegen (2.4) muß $(-\alpha)(-a) = \alpha(-(-a)) = \alpha a$ gelten. Dies zeigt, dass (2.4) für beliebige $\alpha \in \mathbb{R}$ und beliebige Vektoren a gilt.

Aus dem Strahlensatz der Geometrie folgt sofort, dass für beliebige Vektoren a, b und beliebige positive reelle Zahlen α

$$\alpha(a+b) = \alpha a + \alpha b \tag{2.5}$$

gilt.



$a = \overrightarrow{PQ}$, $\quad \alpha a = \overrightarrow{PQ'}$,
$b = \overrightarrow{PR}$, $\quad \alpha b = \overrightarrow{PR'}$,
$a + b = \overrightarrow{PS}$, $\quad \alpha(a+b) = \overrightarrow{PS'}$.

Die Beziehung (2.5) gilt auch für $\alpha = 0$. Für $\alpha < 0$ muß man nicht mehr die geometrische Anschauung zu Hilfe nehmen. Man erhält mit Hilfe von (2.3) und (2.4):

$$\begin{aligned}\alpha(a+b) &= (-\alpha)(-(a+b)) = (-\alpha)((-a)+(-b)) \\ &= (-\alpha)(-a) + (-\alpha)(-b) = \alpha a + \alpha b.\end{aligned}$$

Also gilt (2.5) für beliebige reelle Zahlen, d.h. es gilt das ***erste Distributivgesetz*** für die skalare Multiplikation und Vektoraddition:

(V1) *Für beliebige* $\alpha \in \mathbb{R}$ *und beliebige Vektoren* a, b *gilt:*

$$\alpha(a+b) = \alpha a + \alpha b$$

Mit Hilfe von (V1) und (2.4) können wir beispielsweise folgende Aussage über $-(\alpha a)$ beweisen:

$$-(\alpha a) = (-\alpha)a = \alpha(-a). \tag{2.6}$$

Beweis. 1. $(-\alpha)a = \alpha(-a)$ wegen (2.4).
2. $\alpha a + (-\alpha)a = \alpha a + \alpha(-a) = \alpha(a-a) = \alpha o = o$ (hier wurde auch (2.4) benutzt), d.h. $(-\alpha)a = -(\alpha a)$. □

Es seien nun α, β nicht-negative reelle Zahlen und a ein Vektor. In diesem Falle haben die Vektoren $\alpha a, \beta a$ und $(\alpha+\beta)a$ dieselbe Richtung. Daher hat $\alpha a + \beta a$ die Norm $\alpha\|a\| + \beta\|a\| = (\alpha+\beta)\|a\|$. Daraus folgt

$$\alpha a + \beta a = (\alpha+\beta)a. \tag{2.7}$$

Ist $\alpha > 0$, $\beta < 0$ und $\alpha + \beta > 0$, so gilt wegen (2.7) und (2.6)

$$\begin{aligned} -(\beta a) + (\alpha+\beta)a &= (-\beta)a + (\alpha+\beta)a = (-\beta+\alpha+\beta)a = \alpha a, \text{ d.h.} \\ (\alpha+\beta)a &= \alpha a + \beta a. \end{aligned}$$

Für $\alpha \geq 0$, $\beta < 0$ und $\alpha + \beta < 0$ gilt

$$\alpha a - (\alpha+\beta)a = \alpha a + (-\alpha-\beta)a = (-\beta)a = -(\beta a),$$

woraus wieder

$$\alpha a + \beta a = (\alpha+\beta)a$$

folgt.

Den Fall $\alpha < 0$ und $\beta \geq 0$ führt man auf den vorhergehenden Fall zurück, indem man die Rollen von α und β vertauscht. (Dies ist wegen der Kommutativität der Vektoraddition möglich.)

Für die skalare Multiplikation und die Vektoraddition gilt somit auch das ***zweite Distributivgesetz***:

(V2) *Für beliebige reelle Zahlen* α, β *und beliebige Vektoren* a *gilt:*

$$\alpha a + \beta a = (\alpha+\beta)a.$$

Ferner gilt für die skalare Multiplikation das ***Assoziativgesetz***:

(V3) *Für beliebige* $\alpha, \beta \in \mathbb{R}$ *und beliebige Vektoren* a *gilt*

$$\alpha(\beta a) = (\alpha\beta)a.$$

Beweis. Für nichtnegative Zahlen α, β sind die Vektoren $(\alpha\beta)a, \beta a$ und $\alpha(\beta a)$ gleichgerichtet und für deren Normen gilt $\|(\alpha\beta)a\| = (\alpha\beta)\|a\| = \alpha\|\beta a\| = \alpha(\beta\|a\|)$. Daraus folgt $\alpha(\beta a) = (\alpha\beta)a$.

Für alle $\alpha < 0$ und $\beta \geq 0$ benutzen wir (2.6) und erhalten $\alpha(\beta a) = (-\alpha)(-(\beta a)) = (-\alpha)(\beta(-a)) = (-\alpha\beta)(-a) = (\alpha\beta)a$. Analog erhält man im Fall $\alpha \geq 0, \beta < 0$

$$\alpha(\beta a) = \alpha((-\beta)(-a)) = (-\alpha\beta)(-a) = (\alpha\beta)a.$$

Ist $\alpha \leq 0$ und $\beta < 0$, so gilt

$$\alpha(\beta a) = (-\alpha)(-(\beta a)) = (-\alpha)((-\beta)a) = ((-\alpha)(-\beta))a = (\alpha\beta)a.$$

Damit ist das Assoziativgesetz bewiesen. □

Unmittelbar einsichtig ist die Regel

(V4) $1 \cdot a = a$ *für alle Vektoren* a.

Mit Hilfe von (V1) – (V4) können weitere Regeln bewiesen werden. Beispielsweise gilt:

$$\textit{Ist } \alpha a = o, \textit{ so gilt } \alpha = 0 \textit{ oder } a = o. \tag{2.8}$$

Beweis. Es sei $\alpha a = o$. Ist $\alpha = 0$ so ist nichts mehr zu beweisen. Für $\alpha \neq 0$ folgt aus $\alpha a = o$ auch

$$\frac{1}{\alpha}(\alpha a) = \frac{1}{\alpha} o = o.$$

Andererseits ist $\frac{1}{\alpha}(\alpha a) = (\frac{1}{\alpha}\alpha)a = 1 \cdot a = a$, d.h. $a = o$. □

Man hätte (2.8) natürlich auch direkt mit Hilfe der Definition der skalaren Multiplikation beweisen können. Der oben gegebene Beweis für (2.8) zeigt jedoch, dass (2.8) immer dann gilt, wenn für die Operationen „+“ und „·“ die Regeln (V1) – (V4) gelten.

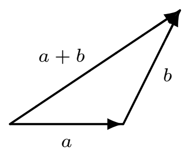
Für die Norm eines Vektors gelten die folgenden Aussagen:

(N_1) $\|a\| \geq 0$ *für alle* a *und* $\|a\| = 0$ *genau für* $a = o$ (*positive Definitheit*).

(N_2) $\|\alpha a\| = |\alpha|\|a\|$ *für alle* $\alpha \in \mathbb{R}$ *und alle Vektoren* a (*Homogenität*).

(N_3) $\|a + b\| \leq \|a\| + \|b\|$ *für alle* a, b (*Dreiecksungleichung*).
„$=$“ *steht genau dann, wenn* $a = \alpha b$ *oder* $b = \beta a$ *mit* $\alpha \geq 0$, $\beta \geq 0$.

Beweis. (N_1) folgt unmittelbar aus der Definition der Norm. (N_2) ist für $\alpha \geq 0$ klar. Für $\alpha < 0$ gilt $\|\alpha a\| = \|(-\alpha)(-a)\| = (-\alpha)\| - a\| = |\alpha|\|a\|$. Die Dreiecksungleichung gilt, da in einem Dreieck die Summe zweier Seiten nicht kleiner sein kann als die dritte Seite. □



Aus der Dreiecksungleichung folgt:

$$\|a - b\| \geq \Big| \|a\| - \|b\| \Big| \quad \textit{für alle } a, b. \tag{2.9}$$

Beweis. Mit Hilfe der Dreiecksungleichung erhält man

$$\|a\| = \|a - b + b\| \leq \|a - b\| + \|b\|,$$

woraus

$$\|a\| - \|b\| \leq \|a - b\|$$

folgt. Analog erhält man

$$\|b\| - \|a\| \leq \|b - a\| = \|a - b\|$$

und

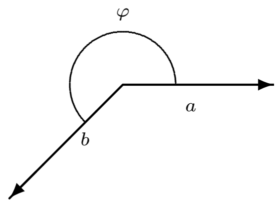
$$\|a - b\| \geq \Big| \|a\| - \|b\| \Big|.$$

□

Der Beweis von (2.9) zeigt, dass diese Aussage immer dann gilt, wenn für $\| \cdot \|$ die Dreiecksungleichung gilt. Man muß nicht mehr auf die Definition der Norm zurückgreifen.

2.2 Inneres Produkt

Je einem Paar von Vektoren $a \neq o$ und $b \neq o$ ist in eindeutiger Weise ein Winkel φ mit $0 \leq \varphi \leq 2\pi$ zugeordnet (der Winkel von a nach b).

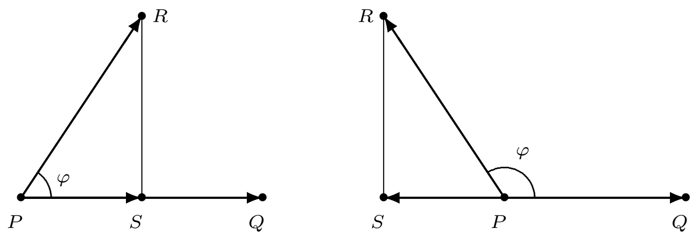


Wir können nun je zwei Vektoren a, b eine reelle Zahl $\langle a, b\rangle$ gemäß der folgenden Definition zuordnen:

$$\begin{aligned} &\langle a, b\rangle = 0, \quad \text{falls } a = o \text{ oder } b = o, \\ &\langle a, b\rangle = \|a\|\,\|b\| \cos\varphi,\ \ \varphi \text{ der Winkel von } a \text{ nach } b, \quad \text{falls } a \neq o \text{ und } b \neq o. \end{aligned}$$

$\langle a, b\rangle$ heißt das ***innere Produkt*** (oder ***Skalarprodukt***) von a und b.

Der folgenden Zeichnung entnimmt man, dass $\langle a, b\rangle$ das Produkt der Norm von a und der Norm der Projektion von b auf a ist, falls $0 \leq \varphi \leq \frac{\pi}{2}$ und $\frac{3\pi}{2} \leq \varphi \leq 2\pi$ ist. Falls $\frac{\pi}{2} < \varphi \leq \frac{3\pi}{2}$ ist, wird noch mit -1 multipliziert.



$a = \overrightarrow{PQ}$, $b = \overrightarrow{PR}$, $\overrightarrow{PS}$ = Projektion von b auf a.

Unmittelbar aus der Definition folgt:

$$\langle a, a\rangle = \|a\|^2 \quad \textit{für alle } a. \tag{2.10}$$

Daraus erhält man sofort (wegen der Definitheit der Norm)

(I1) $\langle a, a\rangle \geq 0$ *und* $\langle a, a\rangle = 0$ *genau dann,wenn* $a = o$ (*Definitheit des inneren Produktes*).

Die Definition des inneren Produktes liefert auch sofort:

(I2) $\langle a, b\rangle = \langle b, a\rangle$ (*Symmetrie*).

Für beliebige reelle Zahlen α und Vektoren a, b gilt:

(I3) $\langle \alpha a, b\rangle = \alpha\langle a, b\rangle$ (*Homogenität*).

Beweis. Der Fall $\alpha = 0$ ist trivial. Für $\alpha > 0$ ist der Winkel φ von a nach b derselbe wie jener von αa nach b. Daher gilt

$$\begin{aligned} \langle \alpha a, b\rangle &= \|\alpha a\|\,\|b\| \cos\varphi = |\alpha|\,\|a\|\,\|b\| \cos\varphi \\ &= \alpha\|a\|\,\|b\| \cos\varphi = \alpha\langle a, b\rangle. \end{aligned}$$

Für $\alpha < 0$ ist der Winkel ψ von αa nach b durch $\pi - \varphi$ für $0 \leq \varphi < \pi$ und durch $\varphi - \pi$ für $\pi \leq \varphi < 2\pi$ gegeben, wobei φ wieder den Winkel von a nach b bezeichnet. In beiden Fällen ist $\cos\psi = -\cos\varphi$. Daher gilt:

$$\begin{aligned}\langle \alpha a, b\rangle &= \|\alpha a\|\,\|b\|\cos(\psi) = |\alpha|\,\|a\|\,\|b\|(-\cos\varphi)\\ &= (-\alpha)\|a\|\,\|b\|(-\cos\varphi) = \alpha\|a\|\,\|b\|\cos\varphi = \alpha\langle a,b\rangle.\end{aligned}$$

□

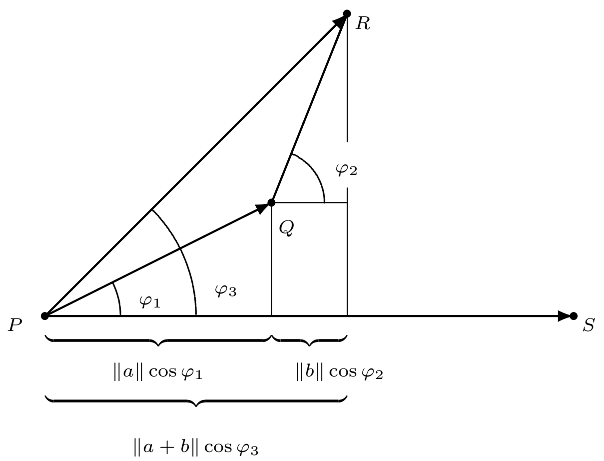
Eine weitere wichtige Eigenschaft gilt für je drei Vektoren a, b, c:

(I4) $\langle a+b, c\rangle = \langle a,c\rangle + \langle b,c\rangle$ (*Additivität*).

Beweis. Aus der untenstehenden Zeichnung entnimmt man sofort

$$\|a\|\cos\varphi_1 + \|b\|\cos\varphi_2 = \|a+b\|\cos\varphi_3,$$

wobei φ_1, φ_2 bzw. φ_3 der Winkel von a nach c, von b nach c und von $a+b$ nach c ist. Daraus folgt sofort das Ergebnis. □



$a = \overrightarrow{PQ}$, $b = \overrightarrow{QR}$, $a+b = \overrightarrow{PR}$, $c = \overrightarrow{PS}$.

Die Eigenschaften (I3) und (I4) können zusammengefaßt werden:
Für beliebige reelle Zahlen α, β und Vektoren a, b, c gilt:

$$\begin{aligned}\langle \alpha a + \beta b, c\rangle &= \alpha\langle a,c\rangle + \beta\langle b,c\rangle,\\ \langle c, \alpha a + \beta b\rangle &= \alpha\langle c,a\rangle + \beta\langle c,b\rangle \quad (\textit{Bilinearität}).\end{aligned}$$

Definition 2.3. a) *Zwei Vektoren a, b heißen **orthogonal**, $a \perp b$, wenn $\langle a,b\rangle = 0$ gilt.*
b) *Vektoren $a_1, \ldots, a_k$ heißen orthogonal, wenn $\langle a_i, a_j\rangle = 0$ für $i \neq j$ gilt. Ist zusätzlich $\|a_i\| = 1$, $i = 1, \ldots, k$, so heißen $a_1, \ldots, a_k$ **orthonormiert**.*

Ist $a = o$ oder $b = o$, so gilt trivialerweise $a \perp b$. Unmittelbar aus der Definition folgt wegen $|\cos\varphi| \leq 1$:

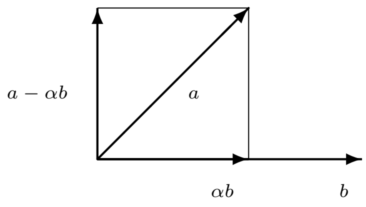
$$|\langle a,b\rangle| \leq \|a\|\,\|b\| \quad (\textit{Schwarzsche}^{8}\ \textit{Ungleichung}^{9}).$$

Das Gleichheitszeichen gilt dann und nur dann, wenn es reelle Zahlen α, β gibt mit $\alpha a + \beta b = o$ und $(\alpha, \beta) \neq (0, 0)$.

Für den Beweis der Schwarzschen Ungleichung müssen wir jedoch nicht mehr auf die Definition des inneren Produktes zurückgreifen. Für $b = o$ ist alles klar. Insbesondere gilt das Gleichheitszeichen und $0a + 1b = 0$. Es seien daher a, b Vektoren mit $b \neq o$. Wir suchen zunächst eine Zahl α derart, dass

$$a = \alpha b + (a - \alpha b),$$

wobei $\alpha b \perp a - \alpha b$.



Geometrisch bedeutet dies, dass wir eine Darstellung von a als Summe eines Vektors parallel zu b und eines Vektors orthogonal zu b suchen

$$b \perp (a - \alpha b) \quad \text{bedeutet} \quad 0 = \langle b, a - \alpha b \rangle = \langle a, b \rangle - \alpha \|b\|^2.$$

Wegen $b \neq o$ erhalten wir daraus $\alpha = \langle a, b \rangle / \|b\|^2$.

Aus $a = \alpha b + (a - \alpha b)$ und $\alpha b \perp (a - \alpha b)$ folgt

$$\|a\|^2 = \langle \alpha b + (a - \alpha b), \alpha b + (a - \alpha b) \rangle = \alpha^2 \|b\|^2 + \|a - \alpha b\|^2$$

und daraus weiter

$$\|a\|^2 \geq \alpha^2 \|b\|^2 = \frac{\langle a, b \rangle^2}{\|b\|^2}.$$

Schließlich erhält man $\langle a, b \rangle^2 \leq \|a\|^2 \, \|b\|^2$ bzw.

$$|\langle a, b \rangle| \leq \|a\| \, \|b\|.$$

In den oben durchgeführten Umformungen gilt genau dann überall das Gleichheitszeichen, wenn $\|a - \alpha b\|^2 = 0$ ist, d.h. $a = \alpha b$ gilt.

Man beachte, dass dieser Beweis nur (2.10) und die Eigenschaften des inneren Produktes benutzt.

Im Beweis der Schwarzschen Ungleichung wurde die folgende Aussage mitbewiesen:

Es seien a, b Vektoren mit $b \neq o$. Dann besitzt a eine eindeutige Darstellung

$$a = a_1 + a_2$$

*mit $a_2 \perp b$, wobei $a_1 = \big(\langle a, b \rangle / \|b\|^2\big) b$ und $a_2 = a - a_1$. Der Vektor a_2 heißt die **Normalkomponente** von a bezüglich b.*

2.3 Basis, Koordinaten

Definition 2.4. a) *$a_1, \ldots, a_n$ seien Vektoren. Ein Vektor c heißt **Linearkombination** von $a_1, \ldots, a_n$, wenn*

$$c = \alpha_1 a_1 + \cdots + \alpha_n a_n = \sum_{j=1}^{n} \alpha_j a_j$$

für reelle Zahlen $\alpha_1, \ldots, \alpha_n$ *gilt. Die Linearkombination heißt nicht-trivial, falls mindestens ein* $\alpha_i \neq 0$ *ist.*
b) *Die Vektoren* $a_1, \ldots, a_n$ *heißen* ***linear abhängig****, wenn der Nullvektor eine nicht-triviale Linearkombination von* $a_1, \ldots, a_n$ *ist, d.h. es gilt*

$$o = \alpha_1 a_1 + \cdots + \alpha_n a_n$$

und $\alpha_i \neq 0$ *für mindestens ein* i*. Die Vektoren* $a_1, \ldots, a_n$ *heißen* ***linear unabhängig****, wenn sie nicht linear abhängig sind.*

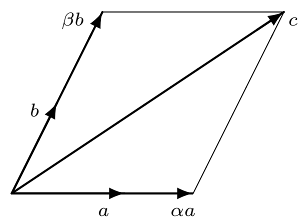
Die Vektoren $a_1, \ldots, a_n$ sind somit genau dann linear unabhängig, wenn die einzige Linearkombination von $a_1, \ldots, a_n$, welche o ergibt, die triviale ist, d.h. aus $\alpha_1 a_1 + \cdots + \alpha_n a_n = o$ folgt $\alpha_1 = \cdots = \alpha_n = 0$. Ist einer der Vektoren $a_i = o$, so sind die Vektoren $a_1, \ldots, a_n$ linear abhängig. Dies ist klar wegen $o = 0 \cdot a_1 + \cdots + 0 \cdot a_{i-1} + 1 \cdot a_i + 0 \cdot a_{i+1} + \cdots + 0 \cdot a_n$.

Satz 2.5. a) *Je drei Vektoren im* $\mathbb{R}^2$ *sind linear abhängig. Es gibt jedoch Vektoren* e_1, e_2 *im* $\mathbb{R}^2$*, die linear unabhängig sind.*
b) *Je vier Vektoren im* $\mathbb{R}^3$ *sind linear abhängig. Es gibt jedoch Vektoren* e_1, e_2, e_3 *im* $\mathbb{R}^3$*, die linear unabhängig sind.*

Beweis. a) Es seien a, b, c Vektoren im $\mathbb{R}^2$. Nur der Fall, wo keiner der Vektoren der Nullvektor ist, muß noch untersucht werden. Auf Grund der unten stehenden Zeichnung sieht man sofort dass

$$c = \alpha a + \beta b$$

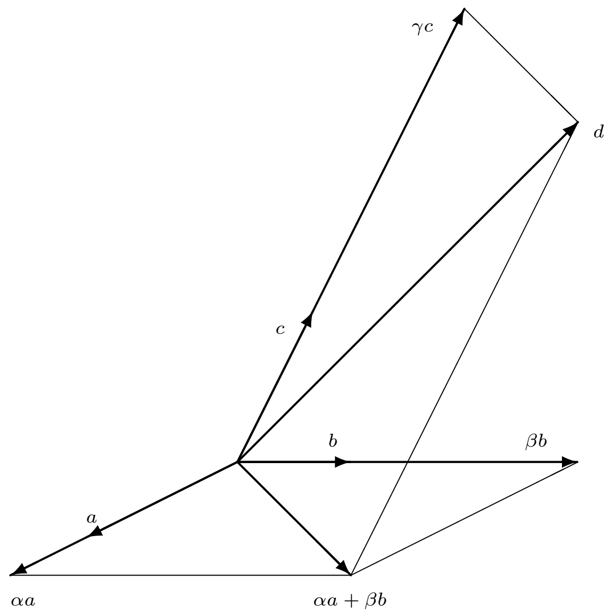
mit geeignet gewählten reellen Zahlen α, β gilt. Andererseits ist auch klar, dass etwa die Vektoren a, b linear unabhängig sind.



b) Es seien a, b, c, d Vektoren im $\mathbb{R}^3$. Wir setzen wieder voraus, dass keiner der Vektoren der Nullvektor ist. Auch hier sieht man, dass für geeignete α, β, γ

$$d = \alpha a + \beta b + \gamma c$$

gilt und dass ferner die Vektoren a, b, c linear unabhängig sind. $\square$



Es seien e_1, e_2 linear unabhängige Vektoren im $\mathbb{R}^2$ und a sei ein weiterer Vektor im $\mathbb{R}^2$. Dann sind a, e_1 und e_2 linear abhängig (Satz 2.5) und es gilt mit Zahlen $\tilde{\alpha}$, β_1 und β_2, die nicht sämtliche Null sind,

$$\tilde{\alpha}a + \beta_1 e_1 + \beta_2 e_2 = o.$$

Es muß $\tilde{\alpha} \neq 0$ sein, da sonst $\beta_1 e_1 + \beta_2 e_2 = o$ sein müßte, woraus wegen der linearen Unabhängigkeit von e_1, e_2 auch $\beta_1 = \beta_2 = 0$ folgen würde. Es folgt daher

$$a = \alpha_1 e_1 + \alpha_2 e_2$$

mit $\alpha_1 = -\beta_1/\tilde{\alpha}$, $\alpha_2 = -\beta_2/\tilde{\alpha}$. Die Zahlen α_1, α_2 sind eindeutig bestimmt. (Wäre $a = \beta_1 e_1 + \beta_2 e_2$, so müßte $o = (\alpha_1 - \beta_1)e_1 + (\alpha_2 - \beta_2)e_2$ gelten. Wegen der linearen Unabhängigkeit von e_1, e_2 folgt daraus aber $\alpha_1 = \beta_1$ und $\alpha_2 = \beta_2$.) Wir nennen jedes Paar (e_1, e_2) linear unabhängiger Vektoren im $\mathbb{R}^2$ eine ***Basis*** für den Raum der Vektoren im $\mathbb{R}^2$. Die eindeutig bestimmten Zahlen α_1, α_2 heißen ***Koordinaten*** von a (bezüglich der Basis (e_1, e_2)).

Eine analoge Situation liegt für Vektoren im $\mathbb{R}^3$ vor. Je drei linear unabhängige Vektoren e_1, e_2, e_3 im $\mathbb{R}^3$ bilden eine Basis und für jeden Vektor a gibt es eindeutig bestimmte Zahlen $\alpha_1, \alpha_2, \alpha_3$ mit

$$a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3.$$

Die Zahlen $\alpha_1, \alpha_2, \alpha_3$ heißen wieder die Koordinaten von a (bezüglich der Basis (e_1, e_2, e_3)).

Die Vektoren $\alpha_i e_i$ werden vielfach die ***Komponenten*** von a bezüglich der Basis (e_1, e_2) bzw. (e_1, e_2, e_3) genannt.

Sind a, b, c (bzw. a, b) vom Nullvektor verschiedene orthogonale Vektoren im $\mathbb{R}^3$ (bzw. $\mathbb{R}^2$), so sind sie linear unabhängig und bilden eine Basis für die Vektoren im $\mathbb{R}^3$ (bzw. $\mathbb{R}^2$). Gilt zusätzlich $\|a\| = \|b\| = \|c\| = 1$, so nennen wir (a, b, c) (bzw. (a, b)) eine ***Orthonormalbasis***.

Wir führen nur den Beweis der linearen Unabhängigkeit von a, b, c. Es sei

$$\alpha a + \beta b + \gamma c = o.$$

Wir bilden das innere Produkt mit a, b bzw. c und erhalten wegen der Orthogonalität von a, b, c

$$\alpha\|a\|^2 = \beta\|b\|^2 = \gamma\|c\|^2 = 0$$

woraus $\alpha = \beta = \gamma = 0$ folgt (wegen $a \neq o$, $b \neq o$, $c \neq o$).

Es sei (e_1, e_2, e_3) eine Basis für die Vektoren im $\mathbb{R}^3$. Wie wir gesehen haben, entspricht jedem Vektor a im $\mathbb{R}^3$ genau ein Tripel $(\alpha_1, \alpha_2, \alpha_3)$ von reellen Zahlen mit $a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$. Umgekehrt bestimmt jedes Tripel reeller Zahlen vermöge dieser Beziehung einen Vektor im $\mathbb{R}^3$. Es besteht daher eine eindeutige Zuordnung zwischen den Vektoren im $\mathbb{R}^3$ und den Tripeln reeller Zahlen. An Stelle von $(\alpha_1, \alpha_2, \alpha_3)$ schreiben wir

$$\mathfrak{a} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

und nennen $\mathfrak{a}$ den ***Koordinatenvektor*** von a (bzgl. (e_1, e_2, e_3)).

Für $a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$ und $b = \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3$ gilt

$$a + b = (\alpha_1 + \beta_1)e_1 + (\alpha_2 + \beta_2)e_2 + (\alpha_3 + \beta_3)e_3$$

und

$$\mu a = (\mu\alpha_1)e_1 + (\mu\alpha_2)e_2 + (\mu\alpha_3)e_3, \quad \mu \in \mathbb{R}.$$

Die Koordinatenvektoren von $a + b$ und μa sind somit gegeben durch

$$\begin{pmatrix} \alpha_1 + \beta_1 \\ \alpha_2 + \beta_2 \\ \alpha_3 + \beta_3 \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} \mu\alpha_1 \\ \mu\alpha_2 \\ \mu\alpha_3 \end{pmatrix}.$$

Es ist daher sinnvoll für $\mathfrak{a} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$ und $\mathfrak{b} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$

$$\mathfrak{a} + \mathfrak{b} := \begin{pmatrix} \alpha_1 + \beta_1 \\ \alpha_2 + \beta_2 \\ \alpha_3 + \beta_3 \end{pmatrix} \quad \text{und} \quad \mu\mathfrak{a} := \begin{pmatrix} \mu\alpha_1 \\ \mu\alpha_2 \\ \mu\alpha_3 \end{pmatrix}, \quad \mu \in \mathbb{R},$$

zu definieren. Damit haben wir für Spaltenvektoren eine Addition und skalare Multiplikation erklärt. Man sieht sofort, dass die so erklärte Addition kommutativ und assoziativ ist. Ferner ist $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ das Nullelement und zu $\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$ ist $\begin{pmatrix} -\alpha_1 \\ -\alpha_2 \\ -\alpha_3 \end{pmatrix}$ das negative Element.

Für die oben erklärte Addition und skalare Multiplikation gelten darüber hinaus auch (V1) – (V4) .

Für das innere Produkt erhält man eine besonders einfache Darstellung, wenn man eine Orthonormalbasis (e_1, e_2, e_3) wählt: Mit Hilfe von (I1) – (I4) erhält man für $a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$, $b = \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3$ (man beachte $\langle e_i, e_j \rangle = 0$ für $i \neq j$ und $\langle e_i, e_i \rangle = 1$ für $i, j = 1, 2, 3$):

$$\langle a, b \rangle = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3. \tag{2.11}$$

Wegen (2.10) erhält man damit auch

$$\|a\| = (\alpha_1^2 + \alpha_2^2 + \alpha_3^2)^{1/2}. \tag{2.12}$$

Die Formel (2.11) legt nahe, das innere Produkt zweier Koordinatenvektoren $\mathfrak{a} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$ und $\mathfrak{b} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$ durch

$$\langle \mathfrak{a}, \mathfrak{b} \rangle := \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3$$

zu definieren.

Ist (e_1, e_2, e_3) eine beliebige Basis für die Vektoren im $\mathbb{R}^3$, so können wir umgekehrt durch (2.11) ein inneres Produkt definieren,

$$\langle a, b\rangle_1 = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3, \tag{2.13}$$

welches dann im allgemeinen nicht mit dem Produkt $\langle\cdot,\cdot\rangle$ übereinstimmt. Die Vektoren e_1, e_2, e_3 sind bezüglich des inneren Produktes $\langle\cdot,\cdot\rangle_1$ orthonormiert. Die Koordinatenvektoren von e_1, e_2, e_3 bzgl. der Basis (e_1, e_2, e_3) sind

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

Der Beweis, dass $\langle\cdot,\cdot\rangle_1$ die Eigenschaften (I1) – (I4) hat, wird als nützliche Übung empfohlen.

Wir definieren $\|a\|_1 := (\langle a, a\rangle_1)^{1/2}$. Die Schwarzsche Ungleichung gilt ebenfalls für $\langle\cdot,\cdot\rangle_1$ bzw. $\|\cdot\|_1$:

$$|\langle a, b\rangle_1| \leq \|a\|_1\|b\|_1.$$

Der seinerzeit geführte Beweis hat nur die Eigenschaften (I1) – (I4) des inneren Produktes und (2.10) benutzt.

> $\|\cdot\|_1$ *hat die Eigenschaften* (N_1) – (N_3), *welche wir weiter oben* (*Abschnitt 2.1*) *für* $\|\cdot\|$ *vermerkt haben.*

Beweis. 1. Aus (I1) folgt sofort $\|a\|_1 \geq 0$ sowie die Äquivalenz von $\|a\|_1 = 0$ und $a = o$.
2. Aus (I3) und (I2) erhält man $\|\alpha a\|_1^2 = \langle\alpha a, \alpha a\rangle_1 = \alpha^2\langle a, a\rangle_1 = \alpha^2\|a\|_1^2$, d.h. $\|\alpha a\|_1 = |\alpha|\,\|a\|_1$.
3. Mit Hilfe der Schwarzschen Ungleichung erhalten wir

$$\begin{aligned}\|a+b\|_1^2 &= \langle a+b, a+b\rangle_1\\ &= \|a\|_1^2 + 2\langle a, b\rangle_1 + \|b\|_1^2\\ &\leq \|a\|_1^2 + 2\|a\|_1\|b\|_1 + \|b\|_1^2 = (\|a\|_1 + \|b\|_1)^2,\end{aligned}$$

d.h. es gilt

$$\|a+b\|_1 \leq \|a\|_1 + \|b\|_1.$$

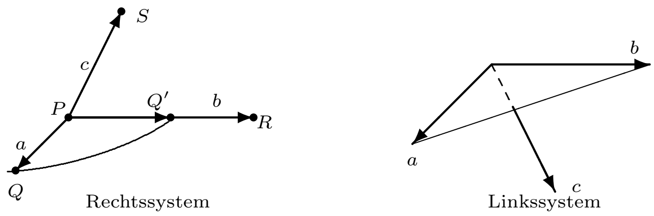
□

Analoge Überlegungen gelten für Vektoren im $\mathbb{R}^2$ bzw. Spaltenvektoren der Form $\begin{pmatrix}\alpha_1\\\alpha_2\end{pmatrix}$.

2.4 Äußeres Produkt, Spatprodukt*

Für Vektoren des $\mathbb{R}^3$ gibt es eine weitere sehr wichtige Operation, welche zwei Vektoren einen dritten zuordnet. Wir benötigen dazu die folgende Begriffsbildung:

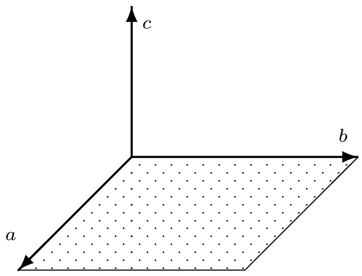
a, b, c seien linear unabhängige Vektoren. Diese Vektoren bilden genau dann ein ***Rechtssystem***, wenn gilt: Wählen wir Repräsentanten mit $a = \overrightarrow{PQ}$, $b = \overrightarrow{PR}$ und $c = \overrightarrow{PS}$, und blickt man vom Punkt S auf die Ebene, welche von den Punkten P, Q, R aufgespannt wird, so kann die Strecke (P, Q) durch eine Drehung um einen Winkel $0 < \varphi < \pi$ gegen den Uhrzeigersinn in eine Strecke (P, Q') übergeführt werden, sodass $\overrightarrow{PQ'} = \alpha b$ mit $\alpha > 0$ gilt. Bilden die Vektoren a, b, c kein Rechtssystem, so bilden sie ein ***Linkssystem***.



Definition 2.6. *a, b, c seien Vektoren im $\mathbb{R}^3$. Der Vektor c ist das* ***äußere Produkt*** *von a, b, $c = a \times b$, wenn c folgenden Bedingungen genügt:*

1. *$c = o$, falls $a = o$ oder $b = o$.*
2. *Für $a \neq o$ und $b \neq o$ gilt:*
   (a) *$c \perp a$ und $c \perp b$.*
   (b) *$\|c\| =$ Flächeninhalt des von a und b aufgespannten Parallelogramms.*
   (c) *a, b, c bilden ein Rechtssystem.*

Die geometrische Bedeutung ist in der folgenden Zeichnung dargestellt.



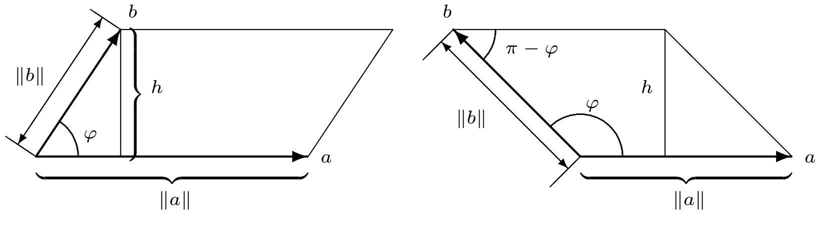
Es ist klar, dass der Vektor c durch die in Definition 2.6 angegebenen Bedingungen eindeutig bestimmt ist. Für die Norm des äußeren Produktes zweier Vektoren gilt:

$$\|a \times b\|^2 = \|a\|^2 \|b\|^2 - \langle a, b \rangle^2. \tag{2.14}$$

Beweis. Für $a = o$ oder $b = o$ ist nichts zu beweisen. Sind a und b vom Nullvektor verschieden, so gilt für den Flächeninhalt des von a und b aufgespannten Parallelogramms

$$F = \|a\| h,$$

wobei $h = \|b\| \sin\varphi$ für $0 \leq \varphi \leq \pi/2$ und $h = \|b\| \sin(\pi - \varphi) = \|b\| \sin\varphi$ für $\pi/2 < \varphi \leq \pi$. φ ist der Winkel von a nach b.



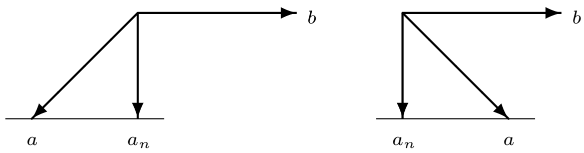
Somit gilt $\|c\|^2 = \|a\|^2 h^2 = \|a\|^2\|b\|^2 \sin^2\varphi = \|a\|^2\|b\|^2(1 - \cos^2\varphi)$, woraus wegen $\cos\varphi = \langle a, b\rangle/(\|a\|\,\|b\|)$ die Behauptung folgt. □

Um später das Distributivgesetz für das äußere Produkt beweisen zu können, benötigen wir

Lemma 2.7. *a, b seien Vektoren mit $b \neq o$. Die Normalkomponente von a bezüglich b wird mit a_n bezeichnet. Dann gilt:*

$$a \times b = a_n \times b.$$

Beweis. Bilden a, b, c ein Rechtssystem, so auch a_n, b, c, d.h. $a \times b$ und $a_n \times b$ haben dieselbe „Richtung".



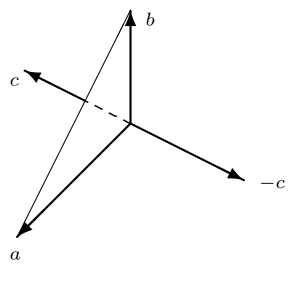
Aus der Zeichnung entnimmt man, dass der Flächeninhalt der von a und b bzw. von a_n und b aufgespannten Parallelogramme übereinstimmt, d.h. $\|a \times b\| = \|a_n \times b\|$. Damit ist das Lemma bewiesen. □

Für das äußere Produkt gilt:

Satz 2.8. *a, b und c seien Vektoren im $\mathbb{R}^3$ und λ sei eine reelle Zahl. Dann gilt:*

1. $a \times b = -(b \times a)$ (*Antikommutativität*),
2. $(\lambda a) \times b = a \times (\lambda b) = \lambda(a \times b)$ (*Assoziativität bezgl. der skalaren Multiplikation*),
3. $(a + b) \times c = a \times c + b \times c$ (*Distributivität*).
4. $a \times b = 0$ *dann und nur dann, wenn a und b linear abhängig sind.*

Beweis. 1. Bilden a, b, c ein Rechtssystem, so auch $b, a, -c$.



2. Für $\lambda = 0$ ist die Assoziativität trivial. Es sei $\lambda > 0$. Zunächst ist klar, dass $c_1 = (\lambda a) \times b$, $c_2 = a \times (\lambda b)$ und $c = a \times b$ dieselbe Richtung haben. Ferner ist $\|c_1\|^2 = \|\lambda a\|^2\|b\|^2 - \langle \lambda a, b\rangle^2 = \lambda^2(\|a\|^2\|b\|^2 - \langle a, b\rangle^2) = \lambda^2\|c\|^2$. Analog sieht man $\|c_2\|^2 = \lambda^2\|c\|^2$. Daraus folgt aber $c_1 = c_2 = \lambda c$.

Für $\lambda < 0$ gilt $(\lambda a) \times b = ((-\lambda)(-a)) \times b) = (-\lambda)((-a) \times b) = (-\lambda)(-(a \times b)) = \lambda(a \times b)$. Hier wurde $(-a) \times b = -(a \times b)$ benutzt. Analog zeigt man $a \times (\lambda b) = \lambda(a \times b)$ für $\lambda < 0$.

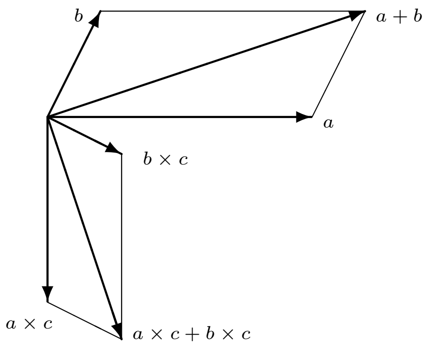
3. Für $c = o$ ist alles klar. Es sei daher $c \neq o$. Auf Grund von Hilfsatz 2.7 genügt es zu zeigen, dass

$$(a+b)_n \times c = a_n \times c + b_n \times c$$

gilt, wobei $(a+b)_n$, a_n und b_n die Normalkomponenten von $a+b$, a bzw. b bezüglich c sind. Man sieht sofort, dass $(a+b)_n = a_n + b_n$ ist. Dazu muß man nur beachten, dass

$$\begin{aligned} a_n &= a - \frac{\langle a, c\rangle}{\|c\|^2} c, \\ b_n &= b - \frac{\langle b, c\rangle}{\|c\|^2} c \quad \text{und} \\ (a+b)_n &= a + b - \frac{\langle a+b, c\rangle}{\|c\|^2} c. \end{aligned}$$

Es genügt daher, im Falle $c \neq o$ das Distributivgesetz für Vektoren a, b, c mit $a \perp c$ und $b \perp c$ (und damit auch $a+b \perp c$) zu beweisen. Wir wählen für die Vektoren Repräsentanten, die denselben Anfangspunkt haben. Dann liegen die Repräsentanten von $a, b, a+b, a\times c, b\times c$ und $(a+b)\times c$ in derselben Ebene senkrecht zum Repräsentanten von c. (In der untenstehenden Zeichnung ist der Repräsentant von c senkrecht zur Papierebene in Richtung des Lesers zu denken).



Da $a \times c \perp a$ und $b \times c \perp b$ gilt, ist der Winkel von a nach b bzw. von $a \times c$ nach $b \times c$ gleich. Darüberhinaus ist (wegen $a \perp c$ und $b \perp c$)

$$\|a \times c\| = \|a\| \cdot \|c\| \quad \text{und} \quad \|b \times c\| = \|b\| \cdot \|c\|. \tag{2.15}$$

Dies bedeutet, dass die Parallelogramme, welche von a und b bzw. $a \times c$ und $b \times c$ aufgespannt werden, ähnlich sind. Ferner gilt, dass die Vektoren $a+b$ und $a \times c + b \times c$ orthogonal sind. Wegen der Ähnlichkeit der zwei Parallelogramme und wegen (2.15) gilt ferner

$$\|a \times c + b \times c\| = \|a+b\|\,\|c\|.$$

Der Vektor $(a+b) \times c$ hat nun dieselbe Richtung wie $a \times c + b \times c$ und es ist (wegen $a+b \perp c$) $\|(a+b) \times c\| = \|a+b\|\,\|c\|$. Daraus folgt aber

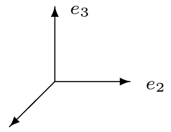
$$a \times c + b \times c = (a+b) \times c.$$

4. Ist $a = o$ oder $b = o$, so ist nichts zu beweisen. Es sei daher $a \neq o$ und $b \neq o$. Ist $a \times b = o$, so folgt aus (2.14), dass $|\langle a, b\rangle| = \|a\|\,\|b\|$ gilt, d.h. für den Winkel zwischen a und b gilt $\varphi = 0$ oder $\varphi = \pi$. Dann ist aber $a = \lambda a$ mit $\lambda > 0$ oder $\lambda < 0$, d.h. a und b sind linear abhängig. Es sei nun andererseits $b = \lambda a$ mit einem $\lambda \in \mathbb{R}$. Dann folgt $\|a \times b\| = 0$ aus (2.14). Damit ist der Satz 2.8 vollständig bewiesen. $\square$

Das äußere Produkt ist nicht assoziativ. Um dies zu sehen, wählen wir vom Nullvektor verschiedene Vektoren a, b mit $a \perp b$. Dann ist $a \times b \perp a$, $\|a \times b\| = \|a\|\,\|b\|$ und $\|a \times (a \times b)\| = \|a\| \cdot \|a \times b\| = \|a\|^2\|b\| \neq 0$, d.h. $a \times (a \times b) \neq o$. Andererseits ist $a \times a = o$ und daher $(a \times a) \times b = o$.

Es sei nun $\mathcal{B} = (e_1, e_2, e_3)$ eine Orthonormalbasis für die Vektoren im $\mathbb{R}^3$ und ein Rechtssystem, d.h. insbesondere

$$e_1 \times e_2 = e_3, e_2 \times e_3 = e_1, e_3 \times e_1 = e_2, \quad e_i \times e_i = 0,\ i = 1, 2, 3. \tag{2.16}$$



a und b seien Vektoren mit den Koordinatenvektoren $\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$ bzw. $\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}$ bezüglich $\mathcal{B}$. Dann erhält man unter Benutzung von (2.16)

$$\begin{aligned} a \times b &= (\alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3) \times (\beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3) \\ &= (\alpha_1\beta_2 - \alpha_2\beta_1)(e_1 \times e_2) + (\alpha_1\beta_3 - \alpha_3\beta_1)(e_1 \times e_3) \\ &\quad + (\alpha_2\beta_3 - \alpha_3\beta_2)(e_2 \times e_3) \\ &= (\alpha_2\beta_3 - \alpha_3\beta_2)e_1 - (\alpha_1\beta_3 - \alpha_3\beta_1)e_2 + (\alpha_1\beta_2 - \alpha_2\beta_1)e_3. \end{aligned}$$

Der Koordinatenvektor von $a \times b$ ist somit durch

$$\begin{pmatrix} \alpha_2\beta_3 - \alpha_3\beta_2 \\ -(\alpha_1\beta_3 - \alpha_3\beta_1) \\ \alpha_1\beta_2 - \alpha_2\beta_1 \end{pmatrix} \tag{2.17}$$

gegeben.

Es sei nun (e_1, e_2, e_3) eine beliebige Basis für die Vektoren im $\mathbb{R}^3$. Für Vektoren $a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$, $b = \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3$ definieren wir ein äußeres Produkt „$*$" durch (2.17), d.h. wir setzen

$$a * b := (\alpha_2\beta_3 - \alpha_3\beta_2)e_1 - (\alpha_1\beta_3 - \alpha_3\beta_1)e_2 + (\alpha_1\beta_2 - \alpha_2\beta_1)e_3. \tag{2.18}$$

Ferner definieren wir das innere Produkt $\langle \cdot, \cdot \rangle_1$ durch (2.13) und die Norm $\| \cdot \|_1$ durch $\|a\|_1^2 := \langle a, a \rangle_1$.

Es ist leicht zu sehen, dass (2.16) für „$*$" gilt. Ebenso einfach verifiziert man, dass die Punkte 1 – 3 aus Satz 2.8 auch für „$*$" gelten. Punkt 4 aus Satz 2.8 beweist man für „$*$" wie folgt: Es seien zunächst a und b linear abhängig. Ist $a = o$, so folgt aus (2.18) unmittelbar $a * b = o$. Im Falle $a \neq o$ muß $b = \lambda a$ für ein $\lambda \in \mathbb{R}$ gelten. Auch in diesem Fall sieht man an Hand von (2.18) sofort, dass $a * b = o$ gilt. Es sei nun umgekehrt $a * b = o$. Aus (2.18) folgt dann

$$\alpha_2\beta_3 = \alpha_3\beta_2, \quad \alpha_1\beta_3 = \alpha_3\beta_1, \quad \alpha_1\beta_2 = \alpha_2\beta_1. \tag{2.19}$$

Im Falle $\alpha_1 = \alpha_2 = \alpha_3 = 0$ gilt $a = o$ und a, b sind trivialerweise linear abhängig. Es sei ein $\alpha_i \neq 0$, etwa $\alpha_2 \neq 0$. Dann gilt $\beta_2 = \lambda\alpha_2$ mit einem $\lambda \in \mathbb{R}$ und aus den Gleichungen (2.19) folgt $\beta_1 = \big(\alpha_1/\alpha_2\big)\beta_2 = \lambda\alpha_1$ und $\beta_3 = \big(\alpha_3/\alpha_2\big)\beta_2 = \lambda\alpha_3$, d.h. $b = \lambda a$.

Die zu (2.14) analoge Beziehung

$$\|a * b\|_1^2 = \|a\|_1^2\|b\|_1^2 - \langle a, b \rangle_1^2$$

wird durch Ausrechnen verifiziert. Ebenso verifiziert man $a \perp a * b$ und $b \perp a * b$, wobei Orthogonalität im Sinne des inneren Produktes $\langle \cdot, \cdot \rangle_1$ zu verstehen ist.

Definition 2.9. *Gegeben seien die Vektoren a, b, c im $\mathbb{R}^3$. Die reelle Zahl*

$$\langle\langle a, b, c\rangle\rangle := \langle a, b \times c\rangle$$

heißt das ***Spatprodukt*** **(*Raumprodukt*)** *von a, b, c.*

Die wichtigsten Eigenschaften des Spatproduktes fassen wir zusammen in

Satz 2.10. $a, b, c, \ldots$ *seien Vektoren im $\mathbb{R}^3$ und es seien $\alpha_i, \beta_i, \gamma_i \in \mathbb{R}$, $i = 1, 2$. Dann gilt:*

a) $\langle\langle \alpha_1 a_1 + \alpha_2 a_2, b, c\rangle\rangle = \alpha_1\langle\langle a_1, b, c\rangle\rangle + \alpha_2\langle\langle a_2, b, c\rangle\rangle$,
$\langle\langle a, \beta_1 b_1 + \beta_2 b_2, c\rangle\rangle = \beta_1\langle\langle a, b_1, c\rangle\rangle + \beta_2\langle\langle a, b_2, c\rangle\rangle$,
$\langle\langle a, b, \gamma_1 c_1 + \gamma_2 c_2\rangle\rangle = \gamma_1\langle\langle a, b, c_1\rangle\rangle + \gamma_2\langle\langle a, b, c_2\rangle\rangle$ *(Linearität in allen Argumenten).*

b) *Die Vektoren a, b, c bilden genau dann ein Rechtssystem, wenn $\langle\langle a, b, c\rangle\rangle > 0$ ist.*

c) *Die Vektoren a, b, c sind genau dann linear abhängig, wenn $\langle\langle a, b, c\rangle\rangle = 0$ ist.*

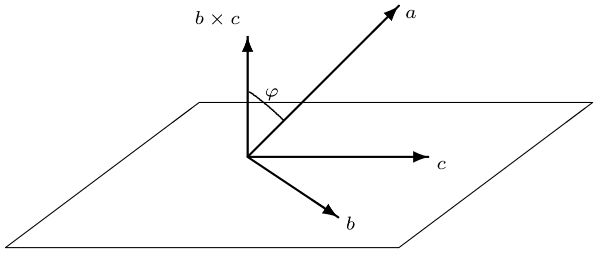
d) *Die Vektoren a, b, c seien linear unabhängig und V bezeichne das Volumen des von a, b, c aufgespannten Parallelflachs. Dann gilt*

$$\langle\langle a, b, c\rangle\rangle = \begin{cases} V, & \text{falls } a, b, c \text{ ein Rechtssystem bilden,} \\ -V, & \text{falls } a, b, c \text{ ein Linkssystem bilden.} \end{cases}$$

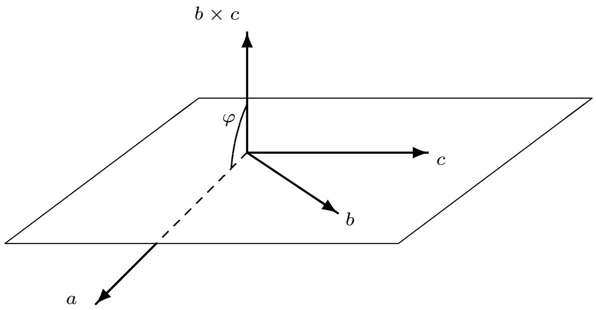
e) $\langle\langle a, b, c\rangle\rangle = \langle\langle b, c, a\rangle\rangle = \langle\langle c, a, b\rangle\rangle = -\langle\langle b, a, c\rangle\rangle = -\langle\langle c, b, a\rangle\rangle = -\langle\langle a, c, b\rangle\rangle$.

f) $\langle\langle a + \beta b, b, c\rangle\rangle = \langle\langle a, b, c\rangle\rangle$.

Beweis. a) Die Linearität in den Argumenten folgt unmittelbar aus der entsprechenden Eigenschaft des inneren Produktes bzw. des äußeren Produktes.
b) Bilden a, b, c ein Rechtssystem und wählt man für a, b, c und $b \times c$ Repräsentanten, die alle vom selben Punkt P ausgehen, so liegen die Repräsentanten von a und $b \times c$ auf derselben Seite der von den Repräsentanten von b und c aufgespannten Ebene. Daher ist der Winkel zwischen a und $b \times c$ kleiner als $\pi/2$, d.h. es ist $\langle a, b \times c\rangle > 0$.



Bilden a, b, c ein Linkssystem, so liegt die in der unterstehenden Zeichnung charakterisierte Situation vor, d.h. der Winkel zwischen a und $b \times c$ ist größer als $\pi/2$ und daher $\langle a, b \times c\rangle < 0$.

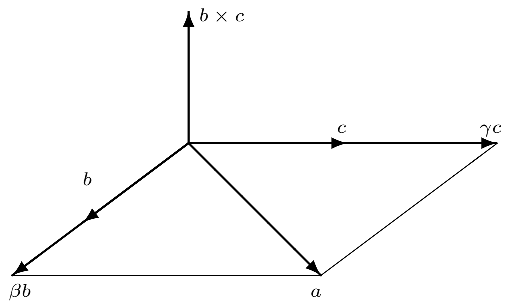


c) Es seien a, b, c linear abhängig. Dann existieren reelle Zahlen α, β, γ, die nicht alle Null sind, mit $\alpha a + \beta b + \gamma c = o$. Es sei etwa $\beta \neq 0$. Dann folgt

$$b = \tilde{\alpha} a + \tilde{\gamma} c$$

(mit $\tilde{\alpha} = -\frac{\alpha}{\beta}$, $\tilde{\gamma} = -\frac{\gamma}{\beta}$) und $b \times c = \tilde{\alpha}(a \times c) + \tilde{\gamma}(c \times c) = \tilde{\alpha}(a \times c)$. Da $a \times c$ und damit auch $\tilde{\alpha}(a \times c)$ orthogonal zu a ist, erhalten wir schließlich $\langle a, b \times c \rangle = \langle a, \tilde{\alpha}(a \times c) \rangle = 0$.

Es sei $\langle\langle a, b, c \rangle\rangle = 0$. Dann ist $a = o$ oder $b \times c = o$ oder $a \perp b \times c$. Im Falle $a = o$ sind a, b, c trivialerweise linear abhängig. Im Falle $b \times c = o$ sind bereits b, c linear abhängig und umsomehr a, b, c. Im Falle $a \perp b \times c$ liegt die in der untenstehenden Zeichnung charakterisierte Situation vor (falls zusätzlich $a \neq o$ und $b \times c \neq o$):

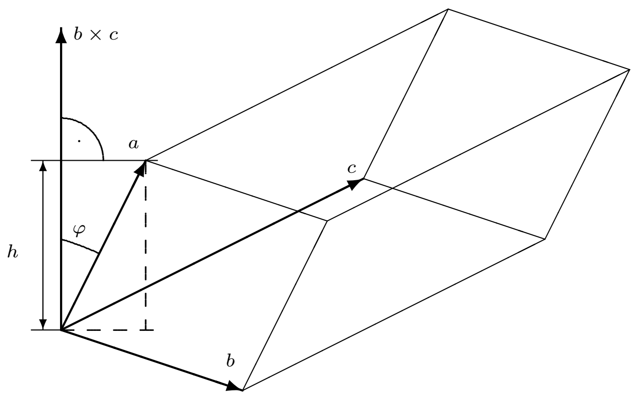


Daher gilt $a = \beta b + \gamma c$ mit zwei reellen Zahlen β, γ, d.h. a, b, c sind linear abhängig.

d) Die Vektoren a, b, c bilden ein Rechtssystem. Das Volumen des von a, b, c aufgespannten Parallelflachs ist gegeben durch

$$V = Fh,$$

wobei F die Fläche des durch die Vektoren b, c bestimmten Parallelogramms und h die Höhe des Parallelflachs ist.



F und h sind durch

$$F = \|b \times c\|, \quad h = \|a\| \cos \varphi,$$

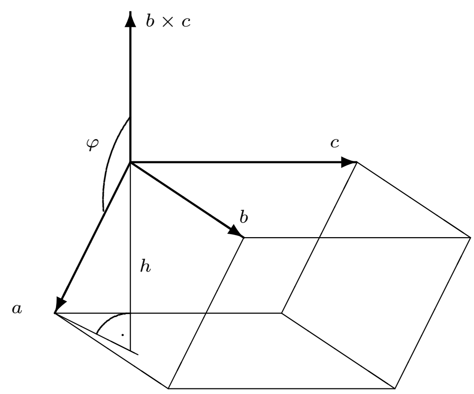
gegeben. Daraus folgt $V = Fh = \|b \times c\| \cdot \|a\| \cos \varphi = \langle a, b \times c \rangle$.

Die Vektoren a, b, c bilden ein Linkssystem. Dann gilt

$$F = \|b \times c\|, \quad h = \|a\| \cos(\pi - \varphi) = -\|a\| \cos \varphi$$

und daher

$$V = Fh = -\|a\| \, \|b \times c\| \cos \varphi = -\langle a, b \times c \rangle.$$



e) Falls die Vektoren a, b, c linear abhängig sind, ist das Spatprodukt bei beliebiger Anordnung dieser Vektoren Null. Es seien a, b, c linear unabhängig. Bilden a, b, c ein Rechts- bzw. Linkssystem, so auch b, c, a und c, a, b. Das jeweils aufgespannte Parallelflach hat dasselbe Volumen V. Daher gilt

$$\langle\!\langle a, b, c \rangle\!\rangle = \langle\!\langle b, c, a \rangle\!\rangle = \langle\!\langle c, a, b \rangle\!\rangle (= V \text{ oder } = -V).$$

Bilden a, b, c ein Rechts- bzw. Linkssystem, so bilden b, a, c, c, b, a und a, c, b ein Links- bzw. Rechtssystem. Da das Volumen des aufgespannten Parallelflach gleich bleibt, gilt

$$\langle\!\langle b, a, c \rangle\!\rangle = \langle\!\langle c, b, a \rangle\!\rangle = \langle\!\langle a, c, b \rangle\!\rangle = -\langle\!\langle a, b, c \rangle\!\rangle.$$

f) Wegen der Linearität des Spatproduktes gilt

$$\langle\!\langle a + \lambda b, b, c \rangle\!\rangle = \langle\!\langle a, b, c \rangle\!\rangle + \lambda \langle\!\langle b, b, c \rangle\!\rangle = \langle\!\langle a, b, c \rangle\!\rangle$$

(beachte $\langle\!\langle b, b, c\rangle\!\rangle = 0$). □

Für eine koordinatenmäßige Darstellung des Spatproduktes wählen wir wie im Falle des äußeren Produktes eine Orthonormalbasis (e_1, e_2, e_3), deren Vektoren ein Rechtssystem bilden. Sind α_i, β_i bzw. γ_i, $i = 1, 2, 3$, die Koordinaten von a, b bzw. c bezüglich der gewählten Basis, so gilt

$$\begin{aligned}\langle\!\langle a, b, c\rangle\!\rangle = \alpha_1\beta_2\gamma_3 + \alpha_2\beta_3\gamma_1 + \alpha_3\beta_1\gamma_2 \\ - \alpha_1\beta_3\gamma_2 - \alpha_2\beta_1\gamma_3 - \alpha_3\beta_2\gamma_1.\end{aligned}$$

Wir wählen nun eine beliebige Basis (e_1, e_2, e_3) für die Vektoren im $\mathbb{R}^3$ und definieren

$$\begin{aligned}\langle\!\langle a, b, c\rangle\!\rangle_1 := \alpha_1\beta_2\gamma_3 + \alpha_2\beta_3\gamma_1 + \alpha_3\beta_1\gamma_2 \\ - \alpha_1\beta_3\gamma_2 - \alpha_2\beta_1\gamma_3 - \alpha_3\beta_2\gamma_1,\end{aligned} \tag{2.20}$$

wobei $a = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$, $b = \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3$, $c = \gamma_1 e_1 + \gamma_2 e_2 + \gamma_3 e_3$. Mit Hilfe von (2.20) zeigt man leicht, dass gilt:

(i) $\langle\!\langle a, b, c\rangle\!\rangle_1 = -\langle\!\langle b, a, c\rangle\!\rangle_1 = -\langle\!\langle c, b, a\rangle\!\rangle_1 = -\langle\!\langle a, c, b\rangle\!\rangle_1$,

(ii) $\langle\!\langle \lambda_1 a_1 + \lambda_2 a_2, b, c\rangle\!\rangle_1 = \lambda_1\langle\!\langle a_1, b, c\rangle\!\rangle_1 + \lambda_2\langle\!\langle a_2, b, c\rangle\!\rangle_1$,

(iii) $\langle\!\langle e_1, e_2, e_3\rangle\!\rangle_1 = 1$.

(iv) $\langle\!\langle \cdot, \cdot, \cdot\rangle\!\rangle_1$ ist das einzige Produkt mit den Eigenschaften (i) – (iii).

***Beweis für* (iv).** Es sei $\langle\!\langle \cdot, \cdot, \cdot\rangle\!\rangle^*$ irgendein Produkt mit den Eigenschaften (i) – (iii). Dann folgt aus (i) für Vektoren b, c

$$\langle\!\langle b, b, c\rangle\!\rangle^* = -\langle\!\langle b, b, c\rangle\!\rangle^*,$$

d.h. $\langle\!\langle b, b, c\rangle\!\rangle^* = 0$. Damit folgt weiter wie oben für $\langle\!\langle \cdot, \cdot, \cdot\rangle\!\rangle$

$$\langle\!\langle a + \lambda b, b, c\rangle\!\rangle^* = \langle\!\langle a, b, c\rangle\!\rangle^*.$$

α_i, β_i bzw. γ_i, $i = 1, 2, 3$, seien die Koordinaten von a, b bzw. c bezüglich der Basis (e_1, e_2, e_3). Dann ist wegen (ii)

$$\begin{aligned}\langle\!\langle a, b, c\rangle\!\rangle^* &= \alpha_1\beta_1\gamma_1\langle\!\langle e_1, e_1, e_1\rangle\!\rangle^* + \alpha_2\beta_1\gamma_1\langle\!\langle e_2, e_1, e_1\rangle\!\rangle^* \\ &\quad + \alpha_3\beta_1\gamma_1\langle\!\langle e_3, e_1, e_1\rangle\!\rangle^* + \cdots = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}\alpha_i\beta_j\gamma_k\langle\!\langle e_i, e_j, e_k\rangle\!\rangle^* \\ &= \alpha_1\beta_2\gamma_3 + \alpha_2\beta_3\gamma_1 + \alpha_3\beta_1\gamma_2 - \alpha_1\beta_3\gamma_2 - \alpha_2\beta_1\gamma_3 - \alpha_3\beta_2\gamma_1 \\ &= \langle\!\langle a, b, c\rangle\!\rangle_1.\end{aligned}$$

Hier wurde benutzt, dass $\langle\!\langle e_i, e_j, e_k\rangle\!\rangle^* = 0$ ist, falls zwei der Indizes übereinstimmen (dies ist eine Folgerung aus (i)) und $\langle\!\langle e_1, e_2, e_3\rangle\!\rangle^* = \langle\!\langle e_2, e_3, e_1\rangle\!\rangle^* = \langle\!\langle e_3, e_1, e_2\rangle\!\rangle^* = 1$, $\langle\!\langle e_1, e_3, e_2\rangle\!\rangle^* = \langle\!\langle e_2, e_1, e_3\rangle\!\rangle^* = \langle\!\langle e_3, e_2, e_1\rangle\!\rangle^* = -1$. (Folgerung aus (i) und (iii); z.B. $\langle\!\langle e_2, e_1, e_3\rangle\!\rangle^* = -\langle\!\langle e_1, e_2, e_3\rangle\!\rangle^* = -1$ und $\langle\!\langle e_3, e_1, e_2\rangle\!\rangle = -\langle\!\langle e_1, e_3, e_2\rangle\!\rangle = \langle\!\langle e_1, e_2, e_3\rangle\!\rangle$). □

Man kann nun definieren, dass a, b, c bezüglich $\langle\!\langle \cdot, \cdot, \cdot\rangle\!\rangle_1$ ein Rechtssystem genau dann bilden, wenn $\langle\!\langle a, b, c\rangle\!\rangle_1 > 0$ ist. Ebenso kann man das „Volumen“ V des durch a, b, c bestimmten Parallelflachs durch

$$V = \langle\!\langle a, b, c\rangle\!\rangle_1$$

definieren, falls a, b, c ein Rechtssystem bilden, und durch

$$V = -\langle\!\langle a, b, c\rangle\!\rangle_1,$$

falls a, b, c ein Linkssystem bilden. Durch ein Produkt mit den Eigenschaften (i) – (iii) wird somit eine Einteilung der Basen für die Vektoren im $\mathbb{R}^3$ in Rechtssysteme und Linkssysteme definiert. Ferner wird die Volumsmessung erklärt.

Abschließend wollen wir noch einige wichtige Identitäten zusammenfassen:

Satz 2.11. a) *Für je drei Vektoren a, b, c im $\mathbb{R}^3$ gilt:*

$$a \times (b \times c) = \langle a, c\rangle b - \langle a, b\rangle c \qquad (\textit{Grassmannscher}^{10}\textit{Entwicklungssatz}).$$

b) *Für je vier Vektoren a, b, c, d im $\mathbb{R}^3$ gilt:*

$$(a \times b) \times (c \times d) = \langle\!\langle a, b, d\rangle\!\rangle c - \langle\!\langle a, b, c\rangle\!\rangle d = \langle\!\langle a, c, d\rangle\!\rangle b - \langle\!\langle b, c, d\rangle\!\rangle a.$$

c) *Für je vier Vektoren a, b, c, d im $\mathbb{R}^3$ gilt:*

$$\langle a \times b, c \times d\rangle = \langle a, c\rangle\langle b, d\rangle - \langle a, d\rangle\langle b, c\rangle \quad (\textit{Identität von Lagrange}^{11}).$$

Beweis. a) Wir wählen zu den Vektoren $a, b, c, b \times c$ und $a \times (b \times c)$ Repräsentanten mit demselben Anfangspunkt. Sind die Vektoren b und c linear abhängig, so ist $b \times c = o$ und damit auch $a \times (b \times c) = o$. Für linear abhängige Vektoren b, c gilt $b = \lambda c$ oder $c = \mu b$ mit reellen Zahlen λ, μ. Es sei etwa $b = \lambda c$. Dann ist

$$\begin{aligned}\langle a, c\rangle b - \langle a, b\rangle c &= \langle a, c\rangle(\lambda c) - \langle a, \lambda c\rangle c\\ &= \lambda\langle a, c\rangle c - \lambda\langle a, c\rangle c = o.\end{aligned}$$

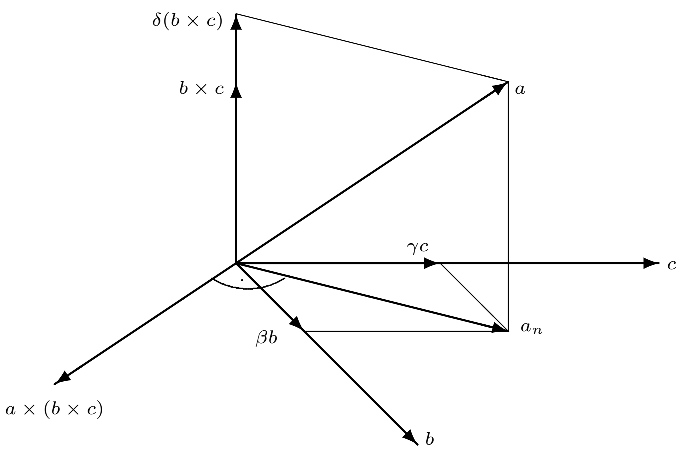
Die Formel stimmt also für linear abhängige b, c.

Es seien nun b und c linear unabhängig. Der Vektor $a \times (b \times c)$ ist orthogonal zu $b \times c$ und daher von der Form

$$a \times (b \times c) = \lambda b + \mu c, \qquad \lambda, \mu \in \mathbb{R}.$$

Es sind die Koeffizienten λ und μ zu bestimmen. Zu diesem Zweck werden wir Lemma 2.7 verwenden. a_n bezeichne die Normalkomponente von a bezüglich $b \times c$. Dann gilt mit reellen Zahlen β, γ

$$a_n = \beta b + \gamma c.$$



[10]Grassmann, Hermann Günther, 15. 4. 1809 (Stettin) – 26. 9. 1877 (Stettin). Die Bedeutung seiner Arbeiten zur Vektorrechnung (Ausdehnungslehre) wurde von den Mathematikern seiner Zeit nicht erkannt. Er wandte sich dann der Linguistik zu und fand durch seine Arbeiten über Sanskrit Anerkennung unter den Philologen (en.wikipedia.org/wiki/Herman_Grassmann).

[11]Lagrange, Joseph Louis, 25. 1. 1736 (Turin, als Guiseppe Luigi Lagrangia) – 10. 4. 1813 (Paris), einer der bedeutendsten Mathematiker seiner Zeit (en.wikipedia.org/wiki/Joseph_Louis_Lagrange).

Die Koeffizienten β und γ berechnet man wie folgt: Zunächst gilt

$$a = a_n + \delta(b \times c) = \beta b + \gamma c + \delta(b \times c), \quad \delta \in \mathbb{R}. \tag{2.21}$$

Durch Bildung des inneren Produktes mit b bzw. c erhalten wir daraus (beachte $b \perp b \times c$ und $c \perp b \times c$)

$$\begin{aligned} \langle a, b\rangle &= \beta\|b\|^2 + \gamma\langle b, c\rangle, \\ \langle a, c\rangle &= \beta\langle b, c\rangle + \gamma\|c\|^2. \end{aligned} \tag{2.22}$$

Aus diesen Gleichungen folgt

$$\begin{aligned} \beta &= \frac{\langle a, b\rangle\|c\|^2 - \langle a, c\rangle\langle b, c\rangle}{\|b\|^2\|c\|^2 - \langle b, c\rangle^2} = \frac{\langle a, b\rangle\|c\|^2 - \langle a, c\rangle\langle b, c\rangle}{\|b \times c\|^2} \\ \gamma &= \frac{\langle a, c\rangle\|b\|^2 - \langle a, b\rangle\langle b, c\rangle}{\|b\|^2\|c\|^2 - \langle b, c\rangle^2} = \frac{\langle a, c\rangle\|b\|^2 - \langle a, b\rangle\langle b, c\rangle}{\|b \times c\|^2}. \end{aligned}$$

Man beachte $\|b \times c\|^2 \neq 0$, weil b, c linear unabhängig sind.

Aus der Orthogonalität von $a \times (b \times c)$ zu a folgt auch $a_n \perp a \times (b \times c) = \lambda b + \mu c$, d.h. es muß

$$\langle \lambda b + \mu c, a_n\rangle = \langle \lambda b + \mu c, \beta b + \gamma c\rangle = 0$$

gelten. Dies ergibt die Gleichung

$$\lambda\beta\|b\|^2 + (\lambda\gamma + \mu\beta)\langle b, c\rangle + \mu\gamma\|c\|^2 = 0$$

bzw. (siehe (2.22))

$$\begin{aligned} 0 &= \lambda\big(\beta\|b\|^2 + \gamma\langle b, c\rangle\big) + \mu\big(\beta\langle b, c\rangle + \gamma\|c\|^2\big) \\ &= \lambda\langle a, b\rangle + \mu\langle a, c\rangle. \end{aligned} \tag{2.23}$$

Gilt $\langle a, b\rangle = \langle a, c\rangle = 0$, so ist $\beta = \gamma = 0$ und daher (siehe (2.21)) $a = \delta(b \times c)$, $\delta \in \mathbb{R}$. In diesem Fall stimmt der Grassmannsche Entwicklungssatz ebenfalls, da beide Seiten den Nullvektor ergeben.

Wir können daher annehmen, dass $\langle a, b\rangle \neq 0$ oder $\langle a, c\rangle \neq 0$ ist. Es sei etwa $\langle a, c\rangle \neq 0$. Aus Gleichung (2.23) folgt $\mu = -\big(\langle a, b\rangle/\langle a, c\rangle\big)\lambda$, $\lambda \in \mathbb{R}$ beliebig. Wir können daher $\lambda = \alpha\langle a, c\rangle$, $\alpha \in \mathbb{R}$ setzen, woraus $\mu = -\langle a, b\rangle\alpha$ folgt. Für den Vektor $a \times (b \times c)$ gilt daher

$$a \times (b \times c) = \alpha\big(\langle a, c\rangle b - \langle a, b\rangle c\big), \quad \alpha \in \mathbb{R}.$$

Es ist somit nur noch α zu bestimmen. Für die Norm von $a \times (b \times c)$ gilt

$$\begin{aligned} \|a \times (b \times c)\|^2 &= \Big\langle \alpha\Big(\langle a, c\rangle b - \langle a, b\rangle c\Big), \alpha\Big(\langle a, c\rangle b - \langle a, b\rangle c\Big)\Big\rangle \\ &= \alpha^2\Big(\langle a, c\rangle^2\|b\|^2 - 2\langle a, c\rangle\langle a, b\rangle\langle b, c\rangle + \langle a, b\rangle^2\|c\|^2\Big) \\ &= \alpha^2\Big(\langle a, c\rangle\big(\langle a, c\rangle\|b\|^2 - \langle a, b\rangle\langle b, c\rangle\big) \\ &\qquad + \langle a, b\rangle\big(\langle a, b\rangle\|c\|^2 - \langle a, c\rangle\langle b, c\rangle\big)\Big) \\ &= \alpha^2\Big(\gamma\langle a, c\rangle + \langle a, b\rangle\beta\Big)\|b \times c\|^2. \end{aligned} \tag{2.24}$$

Für die letzte Gleichung wurde die explizite Gestalt von β und γ benutzt.

Andererseits gilt, nach Lemma 2.7, $a \times (b \times c) = a_n \times (b \times c)$ und daher

$$\|a \times (b \times c)\|^2 = \|a_n \times (b \times c)\|^2 = \|a_n\|^2 \|b \times c\|^2. \tag{2.25}$$

Vergleicht man (2.24) und (2.25), so muß

$$\alpha^2\big(\gamma\langle a, c\rangle + \langle a, b\rangle\beta\big) = \|a_n\|^2$$

gelten. Für a_n gilt aber (siehe (2.22))

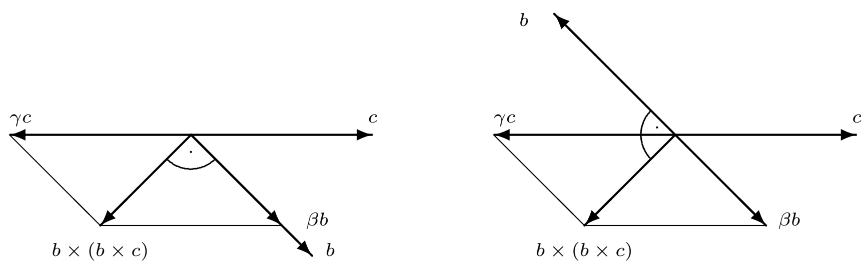
$$\begin{aligned}
\|a_n\|^2 &= \langle a_n, a_n\rangle = \langle \beta b + \gamma c, \beta b + \gamma c\rangle \\
&= \beta^2\|b\|^2 + 2\beta\gamma\langle b, c\rangle + \gamma^2\|c\|^2 \\
&= \beta\big(\beta\|b\|^2 + \gamma\langle b, c\rangle\big) + \gamma\big(\beta\langle b, c\rangle + \gamma\|c\|^2\big) \\
&= \beta\langle a, b\rangle + \gamma\langle a, c\rangle.
\end{aligned}$$

Daraus folgt $\alpha^2 = 1$, d.h.

$$\alpha = \pm 1.$$

Man beachte, dass $\|a_n\| \neq 0$ vorausgesetzt werden kann. Denn $a_n = o$ bedeutet $a = \delta(b \times c)$ mit $\delta \in \mathbb{R}$ und dieser Fall wurde bereits oben erledigt.

Wir haben also noch zu zeigen, dass $\alpha = 1$ ist. Dazu beweisen wir zunächst, dass $\alpha = 1$ im Spezialfall $a = b$ gilt, d.h. $b \times (b \times c) = \langle b, c\rangle b - \|b\|^2 c$. Hier genügt es zu zeigen, dass der Koeffizient von c negativ sein muß. Dies folgt aber sofort aus den unterstehenden Zeichnungen.



Dass $b \times (b \times c) = \langle b, c\rangle b - \|b\|^2 c$ gilt, kann auch folgendermaßen bewiesen werden: Die Vektoren $b, b \times c$ und $b \times (b \times c)$ müssen ein Rechtssystem bilden, d.h. es muß $\langle\!\langle b, b \times c, b \times (b \times c)\rangle\!\rangle > 0$ gelten. Für $b \times (b \times c) = \alpha(\langle b, c\rangle b - \|b\|^2 c)$, $\alpha = \pm 1$, erhalten wir

$$\begin{aligned}
\langle\!\langle b, b \times c, b \times (b \times c)\rangle\!\rangle &= \alpha\langle\!\langle b, b \times c, \langle b, c\rangle b - \|b\|^2 c\rangle\!\rangle \\
&= \alpha\langle b, c\rangle\langle\!\langle b, b \times c, b\rangle\!\rangle - \alpha\|b\|^2\langle\!\langle b, b \times c, c\rangle\!\rangle \\
&= -\alpha\|b\|^2\langle\!\langle b, b \times c, c\rangle\!\rangle = -\alpha\|b\|^2\langle\!\langle b \times c, c, b\rangle\!\rangle \\
&= \alpha\|b\|^2\langle\!\langle b \times c, b, c\rangle\!\rangle = \alpha\|b\|^2\langle b \times c, b \times c\rangle \\
&= \alpha\|b\|^2\|b \times c\|^2 > 0,
\end{aligned}$$

woraus $\alpha = 1$ folgt.

Um $\alpha = 1$ auch im allgemeinen Fall zu beweisen, beachten wir, dass die Vektoren $a, b \times c$ und $a \times (b \times c)$ ein Rechtssystem bilden müssen, d.h. es gilt

$$\delta := \langle\!\langle a, b \times c, a \times (b \times c)\rangle\!\rangle > 0.$$

Wir setzen $a \times (b \times c) = \alpha\big(\langle a, c\rangle b - \langle a, b\rangle c\big)$, $\alpha = \pm 1$, und erhalten

$$\begin{aligned}
\delta &= \alpha\Big(\langle a, c\rangle\langle\!\langle a, b \times c, b\rangle\!\rangle - \langle a, b\rangle\langle\!\langle a, b \times c, c\rangle\!\rangle\Big) \\
&= -\alpha\Big(\langle a, c\rangle\langle a, b \times (b \times c)\rangle + \langle a, b\rangle\langle a, c \times (c \times b)\rangle\Big).
\end{aligned}$$

Der Graßmannsche Entwicklungssatz ist für $b \times (b \times c)$ und $c \times (c \times b)$ bereits bewiesen. Daher gilt weiter

$$\begin{aligned}
\delta &= -\alpha\Big(\langle a,c\rangle\Big\langle a, \langle b,c\rangle b - \|b\|^2 c\Big\rangle + \langle a,b\rangle\Big\langle a, \langle b,c\rangle c - \|c\|^2 b\Big\rangle\Big) \\
&= \alpha\Big(\langle a,c\rangle^2\|b\|^2 - 2\langle a,b\rangle\langle a,c\rangle\langle b,c\rangle + \langle a,b\rangle^2\|c\|^2\Big) \\
&= \alpha\Big\langle \langle a,c\rangle b - \langle a,b\rangle c, \langle a,c\rangle b - \langle a,b\rangle c\Big\rangle.
\end{aligned}$$

Das innere Produkt ist nicht negativ und δ muß positiv sein. Daraus folgt $\alpha = 1$.
b) Mit Hilfe des Grassmannschen Entwicklungssatzes folgt sofort

$$\begin{aligned}
(a \times b) \times (c \times d) &= \langle a \times b, d\rangle c - \langle a \times b, c\rangle d \\
&= \langle d, a \times b\rangle c - \langle c, a \times b\rangle d = \langle\!\langle d,a,b\rangle\!\rangle c - \langle\!\langle c,a,b\rangle\!\rangle d \\
&= \langle\!\langle a,b,d\rangle\!\rangle c - \langle\!\langle a,b,c\rangle\!\rangle d.
\end{aligned}$$

Analog erhält man $(c \times d) \times (a \times b) = \langle\!\langle c,d,b\rangle\!\rangle a - \langle\!\langle c,d,a\rangle\!\rangle b = \langle\!\langle b,c,d\rangle\!\rangle a - \langle\!\langle a,c,d\rangle\!\rangle b$.
c) Auch hier wendet man den Grassmannschen Entwicklungssatz an:

$$\begin{aligned}
\langle a \times b, c \times d\rangle &= \langle\!\langle a \times b, c, d\rangle\!\rangle = \langle\!\langle c, d, a \times b\rangle\!\rangle \\
&= \langle c, d \times (a \times b)\rangle = \Big\langle c, \langle d,b\rangle a - \langle d,a\rangle b\Big\rangle \\
&= \langle a,c\rangle\langle b,d\rangle - \langle a,d\rangle\langle b,c\rangle.
\end{aligned}$$

$\square$

Kapitel 3

Analytische Geometrie im R^3

3.1 Affine und euklidische Räume

Bei der Einführung der Vektoren im $\mathbb{R}^3$ sind wir von den Punkten des $\mathbb{R}^3$ ausgegangen und haben schließlich Vektoren als Äquivalenzklassen gerichteter Strecken (das sind geordnete Punktepaare) eingeführt. Auf Grund dieser Einführung ist es klar, dass zwischen Punkten und Vektoren gewisse Beziehungen bestehen. Man mache sich klar, dass insbesondere die unten angegebenen Eigenschaften (a1) – (a3) erfüllt sind. Um später auch allgemeinere Situationen untersuchen zu können, wollen wir die wesentlichen dieser Beziehungen zwischen Punkten und Vektoren angeben.

Gegeben seien eine Menge $\mathcal{P} = \{P, Q, \dots\}$, deren Elemente ***Punkte*** genannt werden, und eine Menge $V = \{a, b, \dots\}$, deren Elemente ***Vektoren*** genannt werden. Für die Vektoren ist eine Addition und eine skalare Multiplikation erklärt. Die Addition ist assoziativ und kommutativ. Ferner existiert ein (eindeutig bestimmtes) Nullelement und zu jedem Vektor genau ein negatives Element. Für die Addition und skalare Multiplikation gelten die Regeln (V1) – (V4) . Wir nennen V einen ***reellen Vektorraum.*** Als Beispiele für reelle Vektorräume haben wir in Kapitel 2 die reellen Vektorräume der Vektoren im $\mathbb{R}^2$ bzw. im $\mathbb{R}^3$ kennengelernt.

Definition 3.1. *Es sei eine Menge $\mathcal{P}$ von Punkten und ein reeller Vektorraum V gegeben. Das Paar $(\mathcal{P}, V)$ heißt genau dann ein* ***affiner Raum****, wenn die folgenden Eigenschaften gelten:*

(a1) *Jedem geordneten Paar (P, Q) von Punkten aus $\mathcal{P}$ wird genau ein Vektor $a \in V$ zugeordnet, $a = \overrightarrow{PQ}$.*

(a2) *Ist P ein Punkt und a ein Vektor, so existiert genau ein Punkt $Q \in \mathcal{P}$ mit $a = \overrightarrow{PQ}$.*

(a3) *Sind P, Q, R Punkte und ist $a = \overrightarrow{PQ}$, $b = \overrightarrow{QR}$, so ist der Summenvektor $a + b$ dem Punktepaar (P, R) zugeordnet, $a + b = \overrightarrow{PR}$ (d.h. $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$).*

Ist zusätzlich auf V ein inneres Produkt $\langle \cdot, \cdot \rangle$ mit den Eigenschaften (I1) – (I4) *definiert, so heißt $(\mathcal{P}, V)$ ein* ***euklidischer***[12] ***Raum****.*

Einige Beziehungen, die in Kapitel 2 zum Teil mit heuristischen Argumenten begründet wurden, können jetzt mit Hilfe von (a1) – (a3) bewiesen werden. Beispielweise gilt:

1. $\overrightarrow{PP} = o$ *(Nullvektor)*.

[12] Euklid von Alexandria, ca. 365 v. Chr. – ca. 300 v. Chr., lehrte in Alexandria unter Ptolmäus I., berühmt durch die 13 Bände der “Elemente” (Stoicheia), Satz 20 im Buch 9: “Es gibt mehr Primzahlen als jede vorgelegte Anzahl von Primzahlen” (www.muehe.muc.kobic.de/awgeukli/Seite3.html).

2. $\overrightarrow{PQ} = -\overrightarrow{QP}$.

3. $\overrightarrow{P_1Q_1} = \overrightarrow{P_2Q_2} \Rightarrow \overrightarrow{P_1P_2} = \overrightarrow{Q_1Q_2}$ (*Parallelogrammregel*).

Beweis. 1. Für beliebige Punkte $P, R \in \mathcal{P}$ wählen wir $P = Q$ in (a3). Daraus folgt $\overrightarrow{PP} + \overrightarrow{PR} = \overrightarrow{PR}$, d.h. $\overrightarrow{PP} = o$.
2. Wählt man in (a3) $R = P$, so erhält man unter Verwendung von Punkt 1

$$\overrightarrow{PQ} + \overrightarrow{QP} = \overrightarrow{PP} = o, \text{ d.h. } \overrightarrow{PQ} = -\overrightarrow{QP}.$$

3. Aus (a3) folgt

$$\overrightarrow{P_1P_2} + \overrightarrow{P_2Q_2} = \overrightarrow{P_1Q_2}$$

und

$$\overrightarrow{P_1Q_1} + \overrightarrow{Q_1Q_2} = \overrightarrow{P_1Q_2}$$

bzw. $\overrightarrow{P_1P_2} = \overrightarrow{P_1Q_2} - \overrightarrow{P_2Q_2}$ und $\overrightarrow{Q_1Q_2} = \overrightarrow{P_1Q_2} - \overrightarrow{P_1Q_1}$. Durch Subtraktion der beiden Gleichungen folgt

$$\overrightarrow{P_1P_2} - \overrightarrow{Q_1Q_2} = \overrightarrow{P_1Q_1} - \overrightarrow{P_2Q_2} = o.$$

$\square$

Es seien $n \in \mathbb{N}$ und ein reeller Vektorraum V gegeben. Wenn in V je $n+1$ Vektoren linear abhängig sind, es aber n linear unabhängige Vektoren gibt, so sagen wir, dass V die ***Dimension*** n hat, $\dim V = n$. Ist V ein reeller Vektorraum mit $\dim V = n$, so nennen wir jedes n-Tupel $(e_1, \ldots, e_n)$ von linear unabhängigen Vektoren in V eine (geordnete) ***Basis*** von V. Ist $(\mathcal{P}, V)$ ein affiner Raum mit $\dim V = n$, so sagen wir, dass auch der affine Raum $(\mathcal{P}, V)$ die Dimension n hat, $\dim(\mathcal{P}, V) = \dim V$. Aus Gründen der Anschaulichkeit beschränken wir uns für den Rest dieses Kapitels auf affine Räume der Dimension 2 oder 3.

In $\mathcal{P}$ werde nun ein Punkt O fixiert. Dann bestimmt jeder Punkt P genau einen Vektor $a \in V$ (nämlich $a = \overrightarrow{OP}$) und umgekehrt bestimmt jeder Vektor $a \in V$ genau einen Punkt P mit $a = \overrightarrow{OP}$ (siehe (a2)). Dies führt auf die folgende Definition:

Definition 3.2. *Es sei $(\mathcal{P}, V)$ ein affiner Raum der Dimension 3, O ein Element aus $\mathcal{P}$ und (e_1, e_2, e_3) eine Basis für V. Dann heißt $(O; e_1, e_2, e_3)$ ein (affines)* ***Koordinatensystem*** *von $(\mathcal{P}, V)$. O heißt der* ***Koordinatenursprung****. Ist P ein Punkt aus $\mathcal{P}$, so heißt $\overrightarrow{OP}$ der* ***Ortsvektor*** *von P. Sind ξ_1, ξ_2, ξ_3 die Koordinaten des Vektors $\overrightarrow{OP}$ bezüglich der Basis (e_1, e_2, e_3), so heißen ξ_1, ξ_2, ξ_3 auch die* ***affinen Koordinaten*** *des Punktes P (bezüglich des Koordinatensystems $(O; e_1, e_2, e_3)$).*

Analoge Definitionen gelten für affine Räume der Dimension 2.

Das folgende Resultat wird in den folgenden Abschnitten wiederholt verwendet:

Lemma 3.3. *Es sei V ein 2-dimensionaler oder 3-dimensionaler reeller Vektorraum.*
a) *Es sei $a \in V$, $a \neq o$, gegeben. Dann ist*

$$U = \{\lambda a \mid \lambda \in \mathbb{R}\}$$

ein 1-dimensionaler reeller Vektorraum (mit der in V erklärten Vektoraddition und skalaren Multiplikation).
b) *Es sei $\dim V = 3$ und $a, b \in V$ seien linear unabhängig. Dann ist*

$$U = \left\{\lambda a + \mu b \mid \lambda, \mu \in \mathbb{R}\right\}$$

ein 2-dimensionaler reeller Vektorraum.

Beweis. 1. Wir zeigen zunächst, dass in beiden Fällen U ein reeller Vektorraum ist. Es seien $u \in U$ und $\alpha \in \mathbb{R}$ gegeben. Dann ist $u = \lambda a$ mit $\lambda \in \mathbb{R}$ bzw. $u = \lambda a + \mu b$ mit $\lambda, \mu \in \mathbb{R}$ und $\alpha u = \alpha(\lambda a) = (\alpha\lambda)a \in U$ bzw. $\alpha u = \alpha(\lambda a + \mu b) = (\alpha\lambda)a + (\alpha\mu)b \in U$. Dies beweist, dass in beiden Fällen, die sklalare Multiplikation $\mathbb{R} \times V \to V$ auch eine sklalare Multiplikation $\mathbb{R} \times U \to U$ definiert.

Für $u, v \in U$ gilt $u = \lambda a$, $v = \mu a$ mit $\lambda, \mu \in \mathbb{R}$ bzw. $u = \lambda_1 a + \mu_1 b$, $v = \lambda_2 a + \mu_2 b$ mit $\lambda_i, \mu_i \in \mathbb{R}$, $i = 1, 2,$. Daraus folgt $u+v = (\lambda+\mu)a \in U$ bzw. $u+v = (\lambda_1+\lambda_2)a+(\mu_1+\mu_2)b \in U$, d.h. die Vektoraddition $V \times V \to V$ definiert auch ein Vektoraddition $U \times U \to U$.

Die Vektoraddition und die skalare Multiplikation auf U erfüllen die für einen Vektorraum verlangten Eigenschaften, da diese auch für V gelten. Daher ist U in beiden Fällen ein reeller Vektorraum.

2. Wir beweisen nun, dass im Fall a) $\dim U = 1$ gilt. Da $a \in U$ und $a \neq o$ gilt, ist nur noch zu beweisen, dass je zwei Vektoren aus U linear abhängig sind. Es seien $u, v \in U$ gegeben. Dann ist $u = \alpha a$ und $v = \beta a$ mit $\alpha, \beta \in \mathbb{R}$. Ist $u = o$ oder $v = o$ so sind u, v linear abhängig (es gilt $0u + 1v = o$ oder $1u + 0v = o$). Wir müssen nur noch den Fall $u \neq o$ und $v \neq o$ untersuchen. In diesem Fall ist $\alpha \neq 0$ und $\beta \neq 0$. Für beliebige reelle Zahlen γ_1, γ_2 gilt $\gamma_1 u + \gamma_2 v = (\gamma_1\alpha + \gamma_2\beta)a$. Wegen $a \neq o$ ist $\gamma_1 u + \gamma_2 v = o$ daher gleichbedeutend mit

$$\alpha\gamma_1 + \beta\gamma_2 = 0.$$

Diese Gleichung hat die Lösung $\gamma_1 = -\beta/\alpha$, $\gamma_2 = 1 \neq 0$, d.h. u und v sind linear abhängig.

3. Um zu zeigen, dass im Falle b) $\dim U = 2$ gilt, ist nur noch zu beweise, dass je drei Vektoren aus U linear abhängig sind. Für Vektoren $u, v, w \in U$ gilt

$$u = \alpha_1 a + \beta_1 b, \quad v = \alpha_2 a + \beta_2 b, \quad w = \alpha_3 a + \beta_3 b$$

mit $\alpha_i, \beta_i \in \mathbb{R}$, $i = 1, 2, 3$. Für reelle Zahlen $\gamma_1, \gamma_2, \gamma_3$ gilt

$$\gamma_1 u + \gamma_2 v + \gamma_3 w = \big(\alpha_1\gamma_1 + \alpha_2\gamma_2 + \alpha_3\gamma_3\big)a + \big(\beta_1\gamma_1 + \beta_2\gamma_2 + \beta_3\gamma_3\big)b.$$

Die Gleichung $\gamma_1 u + \gamma_2 v + \gamma_3 w = o$ ist daher wegen der linearen Unabhängigkeit von a, b äquivalent zu

$$\begin{aligned} \alpha_1\gamma_1 + \alpha_2\gamma_2 + \alpha_3\gamma_3 &= 0, \\ \beta_1\gamma_1 + \beta_2\gamma_2 + \beta_3\gamma_3 &= 0. \end{aligned} \tag{3.1}$$

Ist einer der Vektoren u, v, w der Nullvektor, so sind die Vektoren u, v, w linear abhängig. Es bleibt daher der Fall $u \neq o$, $v \neq o$, $w \neq o$ zu untersuchen. Aus $u \neq o$ folgt, dass $\alpha_1 \neq 0$ oder $\beta_1 \neq 0$ gilt. Ohne Beeinträchtigung der Allgemeinheit nehmen wir $\alpha_1 \neq 0$ an (ansonsten vertauschen wir a und b). Aus der ersten Gleichung in (3.1) folgt

$$\gamma_1 = -\frac{\alpha_2}{\alpha_1}\gamma_2 - \frac{\alpha_3}{\alpha_1}\gamma_3. \tag{3.2}$$

Dies und die zweite Gleichung in (3.1) ergeben

$$\begin{aligned} &\tilde{\beta}_2\gamma_2 + \tilde{\beta}_3\gamma_3 = 0, \\ &\text{wobei} \quad \tilde{\beta}_2 = \beta_2 - \frac{\alpha_2}{\alpha_1}\beta_1, \quad \tilde{\beta}_3 = \beta_3 - \frac{\alpha_3}{\alpha_1}\beta_1. \end{aligned}$$

Diese Gleichung besitzt auc jeden Fall eine Lösung $(\gamma_2, \gamma_3) \neq (0, 0)$ (z.B. $(1, 0)$, falls $\tilde{\beta}_2 = 0$, und $(-\tilde{\beta}_3/\tilde{\beta}_2, 1)$, falls $\tilde{\beta}_2 \neq 0$). γ_1 kann dann nach (3.2) berechnet werden. Wir erhalten insgesamt eine Lösung $(\gamma_1, \gamma_2, \gamma_3) \neq (0, 0, 0)$, d.h. u, v, w sind linear abhängig. □

3.2 Gerade und Ebenen

In diesem Abschnitt sei $(\mathcal{P}, V)$ ein affiner Raum der Dimension 3. In der analytischen Geometrie interessieren wir uns besonders für spezielle Teilmengen von $\mathcal{P}$, den Geraden und Ebenen.

Definition 3.4. a) *Es sei $g \subset \mathcal{P}$. g ist genau dann eine **Gerade** in $\mathcal{P}$, wenn gilt:*

(i) *$V_g = \{\overrightarrow{PQ} \mid P, Q \in g\}$ ist ein reeller Vektorraum der Dimension 1.*

(ii) *Ist $P \in g$ und $\overrightarrow{PQ} \in V_g$ für ein $Q \in \mathcal{P}$, so gilt $Q \in g$.*

b) *Es sei $E \subset \mathcal{P}$. E ist genau dann eine **Ebene** in $\mathcal{P}$, wenn gilt:*

(i) *$V_E = \{\overrightarrow{PQ} \mid P, Q \in E\}$ ist ein reeller Vektorraum der Dimension 2.*

(ii) *Ist $P \in E$ und $\overrightarrow{PQ} \in V_E$ für ein $Q \in \mathcal{P}$, so gilt $Q \in E$.*

Bemerkung. Ist g eine Gerade in $\mathcal{P}$, so ist (g, V_g) ein affiner Raum der Dimension 1. Ist E eine Ebene in $\mathcal{P}$, so ist (E, V_E) ein affiner Raum der Dimension 2.

Wir beweisen nur die Aussage für Gerade in $\mathcal{P}$. Je zwei Punkten P, $Q \in g$ wird genau ein Vektor aus V_g zugeordnet, nämlich $\overrightarrow{PQ}$, d.h. (a1) ist erfüllt. Es sei nun $P \in g$ und $a \in V_g$. Dann gibt es wegen (a2) genau einen Punkt Q in $\mathcal{P}$ mit $a = \overrightarrow{PQ}$. Wegen (ii) in Definition 3.2, a), gilt $Q \in g$. Damit ist (a2) für (g, V_g) gezeigt. (a3) gilt in (g, V_g), weil es für $(\mathcal{P}, V)$ gilt.

Satz 3.5. a) *Es sei $P \in \mathcal{P}$ und V_1 ein eindimensionaler reeller Vektorraum von Vektoren in V. Dann gibt es genau eine Gerade g mit $P \in g$ und $V_g = V_1$, nämlich*

$$g = \{X \in \mathcal{P} \mid \overrightarrow{PX} \in V_1\}. \tag{3.3}$$

b) *Es sei $P \in \mathcal{P}$ und V_2 ein zweidimensionaler reeller Vektorraum von Vektoren in V. Dann gibt es genau eine Ebene E mit $P \in E$ und $V_E = V_2$, nämlich*

$$E = \{X \in \mathcal{P} \mid \overrightarrow{PX} \in V_2\}. \tag{3.4}$$

Beweis. Wir zeigen zunächst, dass durch (3.3) eine Gerade gegeben ist. Für $X_1, X_2 \in g$ gilt $\overrightarrow{PX_1} \in V_1$ und $\overrightarrow{PX_2} \in V_1$. Daraus folgt $\overrightarrow{X_1X_2} = \overrightarrow{X_1P} + \overrightarrow{PX_2} = \overrightarrow{PX_2} - \overrightarrow{PX_1} \in V_1$. Somit ist $V_g \subset V_1$. Ist andererseits $a \in V_1$, so gibt es genau einen Punkt $X \in \mathcal{P}$ mit $\overrightarrow{PX} = a$. Wegen (3.3) gilt dann $X \in g$ und daher $a = \overrightarrow{PX} \in V_g$. Damit ist auch $V_1 \subset V_g$ gezeigt, d.h. (i) aus Definition 3.2, a), gilt.

Es sei nun $X \in g$ und $Y \in \mathcal{P}$ mit $\overrightarrow{XY} \in V_g = V_1$. Dann ist $\overrightarrow{PY} = \overrightarrow{PX} + \overrightarrow{XY} \in V_1$, d.h. $Y \in g$. Damit ist auch (ii) aus Definition 3.2, a), bewiesen.

Ist nun $\tilde{g}$ ebenfalls eine Gerade mit $P \in \tilde{g}$ und $V_{\tilde{g}} = V_1$, so folgt aus $X \in g$ wegen (3.3) $\overrightarrow{PX} \in V_1$. Da $P \in \tilde{g}$ ist, folgt wegen Definition 3.2, a), (ii), sofort $X \in \tilde{g}$. Damit ist $g \subset \tilde{g}$ gezeigt. Analog folgt $\tilde{g} \subset g$, d.h. es gilt insgesamt $g = \tilde{g}$.

Der Beweis für die Aussage unter b) verläuft völlig analog. □

Korollar 3.6. a) *Es seien $P, Q \in \mathcal{P}$, $P \neq Q$ gegeben. Dann gibt es genau eine Gerade g mit $P \in g$ und $Q \in g$; g ist durch (3.3) gegeben, wobei $V_1 = \{\lambda\overrightarrow{PQ} \mid \lambda \in \mathbb{R}\}$ ist.*

b) *Es seien $P, Q, R \in \mathcal{P}$ derart gegeben, dass $\overrightarrow{PQ}$ und $\overrightarrow{PR}$ linear unabhängig sind. Dann gibt es genau eine Ebene E mit $P \in E$, $Q \in E$ und $R \in E$; E ist durch* (3.4) *gegeben, wobei $V_2 = \{\lambda\overrightarrow{PQ} + \mu\overrightarrow{PR} \mid \lambda, \mu \in \mathbb{R}\}$ ist.*

Beweis. a) V_1 ist wegen $\overrightarrow{PQ} \neq o$ ein eindimensionaler reeller Vektorraum (beachte Lemma 3.3). Nach Satz 3.5, a), existiert eine Gerade g mit $P \in g$ und $V_g = V_1$, welche durch (3.3) gegeben ist. Wegen $\overrightarrow{PQ} \in V_1$, folgt $Q \in g$. Ist $\tilde{g}$ eine weitere Gerade mit $P, Q \in \tilde{g}$, so ist $V_{\tilde{g}} = \{\lambda\overrightarrow{PQ} \mid \lambda \in \mathbb{R}\} = V_1$. Wegen der Eindeutigkeitsaussage von Satz 3.5, a), muß $\tilde{g} = g$ gelten.
b) Der Beweis für die Existenz einer Ebene E mit den verlangten Eigenschaften ist wieder analog. Ist $\tilde{E}$ eine weitere Ebene mit $P, Q, R \in \tilde{E}$, so gilt zunächst $\overrightarrow{PQ} \in V_{\tilde{E}}$ und $\overrightarrow{PR} \in V_{\tilde{E}}$. Daraus folgt $\lambda\overrightarrow{PQ} + \mu\overrightarrow{PR} \in V_{\tilde{E}}$ für $\lambda, \mu \in \mathbb{R}$, d.h. $V_2 \subset V_E$. Ist andererseits $a \in V_{\tilde{E}}$, so sind die Vektoren a, $\overrightarrow{PQ}$ und $\overrightarrow{PR}$ linear abhängig (wegen $\dim V_{\tilde{E}} = 2$). Daraus folgt, für geeignete reelle Zahlen $\lambda, \mu, a = \lambda\overrightarrow{PQ} + \mu\overrightarrow{PR} \in V_2$, d.h. $V_{\tilde{E}} = V_2$. Der Rest des Beweises verläuft wie im Falle der Geraden. □

Für eine Gerade g, welche mit einer Ebene E zwei verschiedene Punkte gemeinsam hat, gilt $g \subset E$.

Beweis. Es seien P, Q in $g \cap E$ mit $P \neq Q$. Nach Satz 3.5 gilt

$$g = \{X \mid \overrightarrow{PX} \in V_g\}$$

und

$$E = \{X \mid \overrightarrow{PX} \in V_E\}.$$

Nach Korollar 3.6, a), ist $V_g = \{\lambda\overrightarrow{PQ} \mid \lambda \in \mathbb{R}\}$. Da nun $\overrightarrow{PQ} \in V_E$ gilt, folgt unmittelbar $V_g \subset V_E$ und daher auch $g \subset E$. □

Es sei nun eine Gerade g gegeben. Ist $e \in V_g$, $e \neq o$, so können wir e als Basis von V_g wählen. Es sei $P \in g$. Nach Satz 3.5 gilt $g = \{X \mid \overrightarrow{PX} \in V_g\}$, d.h.

$$g = \{X \in \mathcal{P} \mid \overrightarrow{PX} = \lambda e \quad \text{mit} \quad \lambda \in \mathbb{R}\}.$$

Ist O der Koordinatenursprung des affinen Raums $(\mathcal{P}, V)$ so gilt $\overrightarrow{OX} = \overrightarrow{OP} + \overrightarrow{PX}$, d.h.

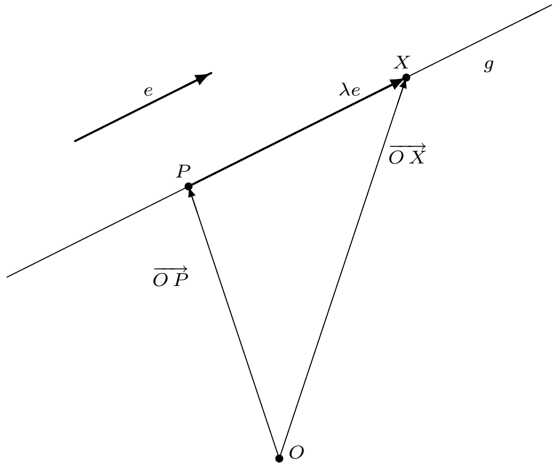
$$g = \{X \in \mathcal{P} \mid \overrightarrow{OX} = \overrightarrow{OP} + \lambda e, \quad \lambda \in \mathbb{R}\}. \tag{3.5}$$

Die hier gegebene Darstellung der Ortsvektoren der Punkte von g heißt die ***Parameterdarstellung von*** g.

Sind P, Q Punkte aus g, $P \neq Q$, so ist $\overrightarrow{PQ}$ eine Basis von V_g. Daher gilt

$$g = \{X \in \mathcal{P} \mid \overrightarrow{OX} = \overrightarrow{OP} + \lambda\overrightarrow{PQ} = \overrightarrow{OP} + \lambda(\overrightarrow{OQ} - \overrightarrow{OP}),\ \lambda \in \mathbb{R}\}. \tag{3.6}$$

Der Veranschaulichung möge die folgende Zeichnung dienen:



Es sei nun E eine Ebene. Wir wählen einen Punkt $P \in E$ und eine Basis e_1, e_2 von V_E. Nach Satz 3.5 gilt $E = \{X \mid \overrightarrow{PX} \in V_E\}$, d.h.

$$E = \{X \in \mathcal{P} \mid \overrightarrow{PX} = \lambda e_1 + \mu e_2 \text{ mit } \lambda, \mu \in \mathbb{R}\}.$$

Für $\overrightarrow{OX}$ gilt wieder $\overrightarrow{OX} = \overrightarrow{OP} + \overrightarrow{PX}$, d.h.

$$E = \{X \in \mathcal{P} \mid \overrightarrow{OX} = \overrightarrow{OP} + \lambda e_1 + \mu e_2,\ \lambda, \mu \in \mathbb{R}\}. \tag{3.7}$$

Dies ist die ***Parameterdarstellung der Ebene*** E.

Sind P, Q, R Punkte in E derart, dass $\overrightarrow{PQ}$ und $\overrightarrow{PR}$ linear unabhängig sind, so bilden $\overrightarrow{PQ}$ und $\overrightarrow{PR}$ eine Basis von V_E. Daher gilt

$$\begin{aligned} E &= \big\{X \in \mathcal{P} \mid \overrightarrow{OX} = \overrightarrow{OP} + \lambda\overrightarrow{PQ} + \mu\overrightarrow{PR}\big\} \\ &= \big\{X \in \mathcal{P} \mid \overrightarrow{OX} = \overrightarrow{OP} + \lambda(\overrightarrow{OQ} - \overrightarrow{OP}) \\ &\qquad\qquad + \mu(\overrightarrow{OR} - OP),\ \lambda, \mu \in \mathbb{R}\big\}. \end{aligned} \tag{3.8}$$

Definition 3.7. *g_1, g_2 seien Gerade und E_1, E_2 Ebenen in $(\mathcal{P}, V)$.*
a) *g_1 und g_2 bzw. E_1 und E_2 heißen genau dann* ***parallel****, $g_1 \| g_2$ bzw. $E_1 \| E_2$, wenn $V_{g_1} = V_{g_2}$ bzw. $V_{E_1} = V_{E_2}$ gilt.*
b) *g_1 und E_1 heißen genau dann parallel, $g_1 \| E_1$, wenn $V_{g_1} \subset V_{E_1}$ gilt.*

Satz 3.8. a) *g sei eine Gerade und E eine Ebene mit $g \nparallel E$. Dann ist $g \cap E = \{P\}$ für genau einen Punkt $P \in \mathcal{P}$. P heißt der* ***Schnittpunkt*** *von g und E.*
b) *E_1 und E_2 seien Ebenen mit $E_1 \nparallel E_2$. Dann ist $E_1 \cap E_2 = g$, g eine Gerade. g heißt die* ***Schnittgerade*** *von E_1 und E_2.*
c) *g_1 und g_2 seien Gerade mit $g_1 \nparallel g_2$ und $g_i \subset E$, $i = 1, 2$, E eine Ebene. Dann ist $g_1 \cap g_2 = \{P\}$ für genau ein $P \in \mathcal{P}$. P heißt der* ***Schnittpunkt*** *von g_1 und g_2.*

Beweis. a) e_1 sei eine Basis von V_g und e_2, e_3 sei eine Basis von V_E. Gilt $\gamma_1 e_1 + \gamma_2 e_2 + \gamma_3 e_3 = o$, so muß $\gamma_1 = 0$ sein (aus $\gamma_1 \neq 0$ würde $e_1 = -(\gamma_2/\gamma_1)e_2 - (\gamma_3/\gamma_1)e_3$ und damit $V_g \subset V_E$ folgen, d.h. $g \| E$). Wegen der linearen Unabhängigkeit von e_2, e_3 muß auch $\gamma_2 = \gamma_3 = 0$ gelten. Damit ist gezeigt, dass e_1, e_2, e_3 linear unabhängig sind und daher eine Basis von V bilden.

Für $X \in \mathcal{P}$ und Punkte $P \in g$, $Q \in E$ gilt nun (siehe (3.5) und (3.7)):

$$X \in g \cap E \text{ genau dann, wenn } \overrightarrow{OX} = \overrightarrow{OP} + \lambda e_1 = \overrightarrow{OQ} + \mu e_2 + \nu e_3$$

mit reellen Zahlen λ, μ, ν. Äquivalent dazu ist

$$\lambda e_1 - \mu e_2 - \nu e_3 = \overrightarrow{PQ}.$$

Da e_1, e_2, e_3 eine Basis von V bilden, gibt es genau ein Tripel λ, μ, ν, welches die letzte Gleichung löst und damit genau einen Punkt aus $g \cap E$.

b) e_1, e_2 bzw. f_1, f_2 sei eine Basis von V_{E_1} bzw. V_{E_2}. Dann sind die Vektoren e_1, e_2, f_1 oder die Vektoren e_1, e_2, f_2 linear unabhängig. Sind nämlich e_1, e_2, f_i linear abhängig, $i = 1, 2$, so würde mit reellen Zahlen α_i, β_i gelten $f_i = \alpha_i e_1 + \beta_i e_2$, $i = 1, 2$, d.h. $f_i \in V_{E_1}$, $i = 1, 2$, und damit auch $V_{E_1} = V_{E_2}$ in Widerspruch zur Voraussetzung $E_1 \nparallel E_2$. Es seien etwa e_1, e_2, f_1 linear unabhängig. Für $X \in \mathcal{P}$ und $P \in E_1$, $Q \in E_2$ gilt (wegen (3.7)):

$$X \in E_1 \cap E_2 \text{ genau dann, wenn } \overrightarrow{OX} = \overrightarrow{OP} + \lambda e_1 + \mu e_2 = \overrightarrow{OQ} + \sigma f_1 + \tau f_2$$

mit reellen Zahlen $\lambda, \mu, \sigma, \tau$. Da e_1, e_2, f_1 eine Basis von V bilden, gilt mit reellen Zahlen α, β, γ

$$f_2 = \alpha e_1 + \beta e_2 + \gamma f_1.$$

Es gilt somit $X \in E_1 \cap E_2$ genau dann, wenn es Zahlen λ, μ, σ und τ gibt mit

$$\begin{aligned}\overrightarrow{PQ} &= \lambda e_1 + \mu e_2 - \sigma f_1 - \tau(\alpha e_1 + \beta e_2 + \gamma f_1) \\ &= (\lambda - \alpha\tau)e_1 + (\mu - \beta\tau)e_2 - (\sigma + \gamma\tau)f_1.\end{aligned}$$

Die Gleichung $\gamma_1 e_1 + \gamma_2 e_2 + \gamma_3 f_1 = \overrightarrow{PQ}$ besitzt genau eine Lösung $\gamma_1, \gamma_2, \gamma_3$. Es gilt daher $X \in E_1 \cap E_2$ genau dann, wenn

$$\lambda - \alpha\tau = \gamma_1,\ \mu - \beta\tau = \gamma_2 \text{ und } \sigma + \gamma\tau = -\gamma_3,$$

d.h. $\lambda = \gamma_1 + \alpha\tau$, $\mu = \gamma_2 + \beta\tau$ und $\sigma = -\gamma_3 - \gamma\tau$. Die Punkte $X \in E_1 \cap E_2$ sind daher durch

$$\overrightarrow{OX} = \overrightarrow{OP} + \lambda e_1 + \mu e_2 = \overrightarrow{OP} + \gamma_1 e_1 + \gamma_2 e_2 + \tau(\alpha e_1 + \beta e_2) \tag{3.9}$$

charakterisiert (oder durch $\overrightarrow{OX} = \overrightarrow{OQ} - \gamma_3 f_1 + \tau(f_2 - \gamma f_1)$). Gemäß (a2) existiert zu P und $\gamma_1 e_1 + \gamma_2 e_2$ ein Punkt $P_1 \in \mathcal{P}$ mit $\overrightarrow{PP_1} = \gamma_1 e_1 + \gamma_2 e_2$, d.h. wir haben $\overrightarrow{OP} + \gamma_1 e_1 + \gamma_2 e_2 = \overrightarrow{OP_1}$. Setzen wir $e = \gamma_1 e_1 + \gamma_2 e_2$, so muß noch $e \neq o$ gezeigt werden. Aus $e = o$ würde wegen der linearen Unabhängigkeit von e_1, e_2 auch $\alpha = \beta = 0$ folgen. Dies würde $f_2 = \gamma f_1$ bedeuten, in Widerspruch zur linearen Unabhängigkeit von f_1, f_2. Aus (3.9) erhalten wir somit $\overrightarrow{OX} = \overrightarrow{OP_1} + \tau e$, $\tau \in \mathbb{R}$, d.h. $X \in E_1 \cap E_2$ genau dann, wenn $X \in g$ gilt mit $g = \{X \mid \overrightarrow{OX} = \overrightarrow{OP_1} + \tau e, \quad \tau \in \mathbb{R}\}$. Daraus folgt $E_1 \cap E_2 = g$.

c) Es sei e_i eine Basis von V_{g_i}, $i = 1, 2$. Die Vektoren e_1, e_2 sind linear unabhängig (weil $g_1 \nparallel g_2$) und bilden daher eine Basis von V_E (beachte $g_i \subset E$, $i = 1, 2$). Es sei $P \in g_1$ und $Q \in g_2$. Dann gilt $X \in g_1 \cap g_2$ genau dann, wenn

$$\overrightarrow{OX} = \overrightarrow{OP} + \lambda e_1 = \overrightarrow{OQ} + \mu e_2,\ \lambda, \mu \in \mathbb{R}.$$

Dies ist äquivalent zu

$$\lambda e_1 - \mu e_2 = \overrightarrow{QP},\ \lambda, \mu \in \mathbb{R}.$$

Wegen $Q, P \in E$ gilt $\overrightarrow{QP} \in V_E$, d.h. es gibt genau ein Paar reeller Zahlen λ, μ mit $\lambda e_1 - \mu e_2 = \overrightarrow{QP}$. Damit ist die Ausage bewiesen. □

Definition 3.9. a) *Punkte $P_i \in \mathcal{P}$, $i = 1, 2, 3$, heißen genau dann* ***kollinear****, wenn es eine Gerade g gibt mit $P_i \in g$, $i = 1, 2, 3$.*

b) *Punkte $P_i \in \mathcal{P}$, $i = 1, \ldots, 4$, heißen genau dann* **komplanar**, *wenn es eine Ebene E gibt mit $P_i \in E$, $i = 1, \ldots, 4$.*

Wir werden häufig die folgenden Kriterien für Kollinearität bzw. Komplanarität verwenden:

Satz 3.10. a) *Punkte $P_i \in \mathcal{P}$, $i = 1, 2, 3$, sind genau dann kollinear, wenn es reelle, nicht sämtlich verschwindende Zahlen α_i, $i = 1, 2, 3$, gibt mit*

$$\sum_{i=1}^{3} \alpha_i = 0 \quad \textit{und} \quad \sum_{i=1}^{3} \alpha_i \overrightarrow{O\,P_i} = o.$$

b) *Punkte $P_i \in \mathcal{P}$, $i = 1, \ldots, 4$, sind genau dann komplanar, wenn es reelle, nicht sämtlich verschwindende Zahlen α_i, $i = 1, \ldots, 4$, gibt mit*

$$\sum_{i=1}^{4} \alpha_i = 0 \quad \textit{und} \quad \sum_{i=1}^{4} \alpha_i \overrightarrow{O\,P_i} = o. \tag{3.10}$$

Beweis. Wir beweisen nur Teil b). Es seien die vier Punkte P_i komplanar. Dann gilt

$$\overrightarrow{O\,P_i} = \overrightarrow{O\,P_0} + \lambda_i e_1 + \mu_i e_2, \ i = 1, \ldots, 4, \tag{3.11}$$

mit reellen Zahlen λ_i, μ_i ($\overrightarrow{O\,X} = \overrightarrow{O\,P_0} + \lambda e_1 + \mu e_2$ ist die Parameterdarstellung der Ebene E mit $P_i \in E$). Gilt $P_k = P_j$ für ein Indexpaar $k \neq j$, so setzen wir $\alpha_k = 1$, $\alpha_j = -1$ und $\alpha_i = 0$ für $i \neq j$, $i \neq k$. Dann ist (3.10) erfüllt. Es genügt daher den Fall $P_k \neq P_j$ für $k \neq j$ zu betrachten. Mit Hilfe von (3.11) sehen wir, dass

$$\sum_{i=1}^{4} \alpha_i \overrightarrow{O\,P_i} = o \quad \text{und} \quad \sum_{i=1}^{4} \alpha_i = 0$$

äquivalent zu

$$\Big(\sum_{i=1}^{4} \alpha_i \lambda_i\Big) e_1 + \Big(\sum_{i=1}^{4} \alpha_i \mu_i\Big) e_2 = o \quad \text{und} \quad \sum_{i=1}^{4} \alpha_i = 0$$

ist. Dies wiederum ist äquivalent zum Gleichungssystem

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= 0, \\ \alpha_1\lambda_1 + \alpha_2\lambda_2 + \alpha_3\lambda_3 + \alpha_4\lambda_4 &= 0, \\ \alpha_1\mu_1 + \alpha_2\mu_2 + \alpha_3\mu_3 + \alpha_4\mu_4 &= 0 \end{aligned} \tag{S}$$

für $\alpha_1, \ldots, \alpha_4$. Die letzte Äquivalenz gilt, weil e_1, e_2 linear unabhängig sind. Das Gleichungssystem (S) ist äquivalent zu

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= 0, \\ (\lambda_2 - \lambda_1)\alpha_2 + (\lambda_3 - \lambda_1)\alpha_3 + (\lambda_4 - \lambda_1)\alpha_4 &= 0, \\ (\mu_2 - \mu_1)\alpha_2 + (\mu_3 - \mu_1)\alpha_3 + (\mu_4 - \mu_1)\alpha_4 &= 0. \end{aligned}$$

Wäre $\lambda_2 = \lambda_1$ und $\mu_2 = \mu_1$, so würde $P_1 = P_2$ gelten, in Widerspruch zur Annahme über die Punkte $P_1, \ldots, P_4$. Da $\lambda_2 \neq \lambda_1$ oder $\mu_2 - \mu_1 \neq 0$ ist, können wir (S) in das äquivalente

Gleichungssystem

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= 0, \\ \alpha_2 + \tilde{\lambda}_3\alpha_3 + \tilde{\lambda}_4\alpha_4 &= 0, \\ \tilde{\mu}_3\alpha_3 + \tilde{\mu}_4\alpha_4 &= 0 \end{aligned}$$

umformen (mit geeignet definierten Zahlen $\tilde{\lambda}_i, \tilde{\mu}_i$). Dieses Gleichungssystem besitzt immer eine nicht-triviale Lösung, etwa $\alpha_4 = 1$, $\alpha_3 = -\tilde{\mu}_4/\tilde{\mu}_3$, $\alpha_2 = \tilde{\lambda}_3\tilde{\mu}_4/\tilde{\mu}_3 - \tilde{\lambda}_4$ und $\alpha_1 = -\tilde{\lambda}_3\tilde{\mu}_4/\tilde{\mu}_3 + \tilde{\mu}_4/\tilde{\mu}_3 - 1$.

Es sei nun umgekehrt (3.10) erfüllt. Wir können ohne Einschränkung annehmen, dass $\alpha_1 \neq 0$ ist. Dann muß noch ein weiteres $\alpha_i \neq 0$ sein, etwa $\alpha_2 \neq 0$. Aus (3.10) folgt

$$\overrightarrow{O\,P_1} = -\tilde{\alpha}_2\overrightarrow{O\,P_2} - \tilde{\alpha}_3\overrightarrow{O\,P_3} - \tilde{\alpha}_4\overrightarrow{O\,P_4}, \quad \tilde{\alpha}_i = \frac{\alpha_i}{\alpha_1}, \quad i = 2, 3, 4, \tag{3.12}$$

wobei $\tilde{\alpha}_2 + \tilde{\alpha}_3 + \tilde{\alpha}_4 = -1$ und $\tilde{\alpha}_2 \neq 0$ ist. Die Punkte P_i sind komplanar, wenn die Vektoren $\overrightarrow{P_1\,P_i}$, $i = 2, 3, 4$, linear abhängig sind. Wegen $\overrightarrow{P_1\,P_i} = \overrightarrow{O\,P_i} - \overrightarrow{O\,P_1}$ und (3.12) ist $\lambda_2\overrightarrow{P_1\,P_2} + \lambda_3\overrightarrow{P_1\,P_3} + \lambda_4\overrightarrow{P_1\,P_4} = o$ gleichwertig mit

$$\Big(\lambda_2 + \tilde{\alpha}_2\sum_{i=2}^{4}\lambda_i\Big)\overrightarrow{O\,P_2} + \Big(\lambda_3 + \tilde{\alpha}_3\sum_{i=2}^{4}\lambda_i\Big)\overrightarrow{O\,P_3} + \Big(\lambda_4 + \tilde{\alpha}_4\sum_{i=2}^{4}\lambda_i\Big)\overrightarrow{O\,P_4} = o.$$

Diese Gleichung ist sicher erfüllt, wenn

$$\begin{aligned} \lambda_2 + \tilde{\alpha}_2\sum_{i=2}^{4}\lambda_i &= 0, \\ \lambda_3 + \tilde{\alpha}_3\sum_{i=2}^{4}\lambda_i &= 0, \\ \lambda_4 + \tilde{\alpha}_4\sum_{i=2}^{4}\lambda_i &= 0. \end{aligned} \tag{S$'$}$$

Wegen $\tilde{\alpha}_2 \neq 0$ ist $\sum_{i=2}^{4}\lambda_i = -\lambda_2/\tilde{\alpha}_2$. Damit folgt $\lambda_3 = (\tilde{\alpha}_3/\tilde{\alpha}_2)\lambda_2$, $\lambda_4 = (\tilde{\alpha}_4/\tilde{\alpha}_2)\lambda_2$. Das Gleichungssystem (S$'$) besitzt daher die nicht-triviale Lösung (beachte $\tilde{\alpha}_2 + \tilde{\alpha}_3 + \tilde{\alpha}_3 = -1$)

$$\lambda_2 = \tilde{\alpha}_2 \neq 0,\ \lambda_3 = \tilde{\alpha}_3,\ \lambda_4 = \tilde{\alpha}_4.$$

Die Vektoren $\overrightarrow{P_1\,P_2}$, $\overrightarrow{P_1\,P_3}$ und $\overrightarrow{P_1\,P_4}$ sind somit linear abhängig. □

3.3 Hessesche Normalform für Ebene und Gerade*

Als Vorbereitung beweisen wir

Lemma 3.11. *Es sei V ein reeller Vektorraum mit innerem Produkt $\langle\cdot,\cdot\rangle$. Es sei $\dim V = n$ mit $n = 2$ oder $n = 3$. Ferner sei $\mathfrak{n} \in V$, $\mathfrak{n} \neq o$, gegeben. Dann ist*

$$U = \{a \in V \mid \langle\mathfrak{n}, a\rangle = 0\}$$

ein reeller Vektorraum mit $\dim U = n - 1$.

Beweis. Man sieht sofort, dass für $a, b \in U$ auch $a + b$ und λa, $\lambda \in \mathbb{R}$, in U enthalten sind. Wie im Beweis zu Lemma 3.3 folgt daraus, dass U ein reeller Vektorraum ist. Es ist noch die Behauptung über die Dimension von U zu beweisen.

Sind für ein $a \in V$ die Vektoren $\mathfrak{n}$ und a linear abhängig, so ist $a = \alpha\mathfrak{n}$ mit einem $\alpha \in \mathbb{R}$ (wegen $\mathfrak{n} \neq o$). Wären die Vektoren $\mathfrak{n}$ und a für alle $a \in V$ linear abhängig, so würde dies

$$V = \{\lambda a \mid \lambda \in \mathbb{R}\}$$

bedeuten. Wegen Lemma 3.3 wäre dann $\dim V = 1$ in Widerspruch zu $dim V \geq 2$. Es existiert daher ein Vektor $a_0 \in V$ derart, dass $\mathfrak{n}$ und a_0 linear unabhängig sind.

Fall 1: $\dim V = 2$.

e sei die Normalkomponente von a_0 bezüglich $\mathfrak{n}$,

$$e = a_0 - \alpha\mathfrak{n}, \quad \alpha = \frac{\langle a_0, \mathfrak{n}\rangle}{\|\mathfrak{n}\|^2}.$$

Wegen $\langle e, \mathfrak{n}\rangle = 0$ ist $e \in U$. Ferner gilt $e \neq o$, denn $e = o$ würde $a_0 = \alpha\mathfrak{n}$ bedeuten, d.h. a_0 und $\mathfrak{n}$ wären linear abhängig.

Es seien nun $u, v \in U$ gegeben. Ist $u = o$ oder $v = o$, so sind u, v linear abhängig. Wir betrachten daher nur mehr den Fall $u \neq o$ und $v \neq o$. Aus $\gamma_0\mathfrak{n} + \gamma_1 u = o$ folgt – durch Bildung des inneren Produktes mit $\mathfrak{n}$ – $\gamma_0\|\mathfrak{n}\|^2 = 0$, d.h. $\gamma_0 = 0$. Dann gilt auch $\gamma_1 u = o$ und, wegen $u \neq o$, auch $\gamma_1 = 0$. Damit ist gezeigt, dass die Vektoren $\mathfrak{n}$ und u linear unabhängig sind. Wegen $\dim V = 2$ bilden $\mathfrak{n}$, u eine Basis von V. Es gibt daher $\alpha_0, \alpha_1 \in \mathbb{R}$ mit $v = \alpha_0\mathfrak{n} + \alpha_1 u$. Durch Bildung des inneren Produktes mit $\mathfrak{n}$ folgt hieraus $0 = \alpha_0\|\mathfrak{n}\|^2$, d.h., $\alpha_0 = 0$. Dann ist aber $v = \alpha_1 u$, d.h. die Vektoren u, v sind linear abhängig. Damit ist gezeigt, dass Vektoren $u, v \in U$ stets linear abhängig sind. Dies beweist $\dim U = 1$.

Fall 2: $\dim V = 3$.

Zusätzlich zu $\mathfrak{n}$ und a_0 wählen wir $b \in V$. Sind $\mathfrak{n}$, a_0 und b linear abhängig, so gilt $\nu\mathfrak{n} + \alpha a_0 + \beta b = o$ mit $(\nu, \alpha, \beta) \neq (0, 0, 0)$. Aus $\beta = 0$ würde $\nu\mathfrak{n} + \alpha a_0 = o$ folgen, d.h. die Vektoren $\mathfrak{n}$, a_0 wären linear abhängig. Dieser Widerspruch bedeutet $\beta \neq 0$. Dann folgt aus $\nu\mathfrak{n} + \alpha a_0 + \beta b = o$ sofort $b = \tilde{\nu} + \tilde{\alpha} a_0$ mit $\tilde{\nu}, \tilde{\alpha} \in \mathbb{R}$. Sind die Vektoren $\mathfrak{n}$, a_0 und b für jedes $b \in V$ linear abhängig, folgt daher

$$V = \{\nu\mathfrak{n} + \alpha a_0 \mid \nu, \alpha \in \mathbb{R}\}.$$

Nach Lemma 3.3 gilt $\dim V = 2$. Dieser Widerspruch beweist, dass ein $b_0 \in V$ existiert, sodass die Vektoren $\mathfrak{n}$, a_0 und b_0 linear unabhängig sind.

Es sei nun e bzw. f die Normalkomponente von a_0 bzw. b_0 bezüglich $\mathfrak{n}$:

$$e = a_0 - \alpha\mathfrak{n}, \quad \alpha = \frac{\langle a_0, \mathfrak{n}\rangle}{\|\mathfrak{n}\|^2},$$

$$f = a_0 - \beta\mathfrak{n}, \quad \beta = \frac{\langle b_0, \mathfrak{n}\rangle}{\|\mathfrak{n}\|^2}.$$

Es seien u, v und w beliebige Vektoren in U. Sind u, v linear abhängig, so auch u, v, w. Wir nehmen daher an, dass u und v linear unabhängig sind. Aus $\gamma_0\mathfrak{n} + \gamma_1 u + \gamma_2 v = o$ folgt – durch Bildung des inneren Produktes mit $\mathfrak{n}$ – $\gamma_0\|\mathfrak{n}\|^2 = 0$, d.h. $\gamma_0 = 0$. Dann gilt $\gamma_1 u + \gamma_2 v = o$, woraus – wegen der linearen Unabhängigkeit von u, v – $\gamma_1 = \gamma_2 = 0$ folgt. Damit ist gezeigt, dass die Vektoren $\mathfrak{n}$, u und v linear unabhängig sind und somit eine Basis von V bilden. Dann gilt $w = \nu\mathfrak{n} + \alpha u + \beta v$, $\nu, \alpha, \beta \in \mathbb{R}$. Durch Bildung des inneren Produktes mit $\mathfrak{n}$ folgt daraus $\nu\|\mathfrak{n}\|^2 = 0$, d.h. $\nu = 0$. Somit gilt $w = \alpha u + \beta v$, d.h. die Vektoren u, v, w sind linear abhängig. Damit ist $\dim U = 2$ gezeigt. □

Definition 3.12. a) $(\mathcal{P}, V)$ *sei ein zweidimensionaler euklidischer Raum und* g *sei eine Gerade in* $\mathcal{P}$. *Ein Vektor* $\mathfrak{n} \neq o$ *in* V *heißt genau dann ein* ***Normalenvektor*** *von* g, *wenn* $\langle \mathfrak{n}, a\rangle = 0$ *für alle* $a \in V_g$ *gilt.*

b) $(\mathcal{P}, V)$ *sei ein dreidimensionaler euklidischer Raum und* E *sei eine Ebene in* $\mathcal{P}$. *Ein Vektor* $\mathfrak{n} \neq o$ *in* V *heißt ein* ***Normalenvektor*** *von* E *genau dann, wenn* $\langle \mathfrak{n}, a \rangle = 0$ *für alle* $a \in V_E$ *gilt.*

Lemma 3.13. a) $(\mathcal{P}, V)$ *sei ein zweidimensionaler euklidischer Raum und* g *sei eine Gerade in* $\mathcal{P}$. *Es sei* $\mathfrak{N}$ *die Menge der Normalenvektoren von* g. *Dann ist* $\mathfrak{N} \cup \{o\}$ *ein eindimensionaler reeller Vektorraum.*

b) $(\mathcal{P}, V)$ *sei ein dreidimensionaler euklidischer Raum und* E *sei eine Ebene in* $\mathcal{P}$. *Es sei* $\mathfrak{N}$ *die Menge der Normalenvektoren von* E. *Dann ist* $\mathfrak{N} \cup \{o\}$ *ein eindimensionaler reeller Vektorraum.*

Beweis. Man sieht in beiden Fällen sofort, dass für zwei Vektoren $a, b \in \mathfrak{N} \cup \{o\}$ auch $a + b$ und λa, $\lambda \in \mathbb{R}$, in $\mathfrak{N} \cup \{o\}$ liegen. Daher ist $\mathfrak{N} \cup \{o\}$ ein reeller Vektorraum.

Im Falle einer Geraden g in einem zweidimensionalen euklidischen Raum $(\mathcal{P}, V)$ wählen wir $e \in V_g$, $e \neq o$. Wegen $\dim V_g = 1$ ist e eine Basis von V_g, d.h. jedes $a \in V_g$ ist von der Form $a = \lambda e$. Man sieht nun sofort, dass $\mathfrak{N} \cup \{o\} = \{b \in V \mid \langle b, a \rangle = 0 \text{ für alle } a \in V_g\} = \{b \in V \mid \langle b, e \rangle = 0\}$ gilt. Dann folgt aus Lemma 3.11 sofort $\dim \mathfrak{N} \cup \{o\} = 2 - 1 = 1$.

Im Falle einer Ebene E im dreidimensionalen euklidischen Raum $(\mathcal{P}, V)$ wählen wir eine Basis (e_1, e_2) von V_E. Analog wie oben sieht man, dass $\mathfrak{N} \cup \{o\} = \{b \in V \mid \langle b, e_1 \rangle = \langle b, e_2 \rangle = 0\}$ gilt. Es seien $b_1, b_2 \in \mathfrak{N} \cup \{o\}$ gewählt. Wir werden zeigen, dass b_1, b_2 linear abhängig sind. Dies ist trivial, wenn $b_1 = o$ ist. Es sei daher $b_1 \neq o$. Dann folgt aus $\lambda_1 e_1 + \lambda_2 e_2 + \mu b_1 = o$ durch Bildung des inneren Produktes mit b_1 sofort $\mu \|b_1\|^2 = 0$, d.h. $\mu = 0$. Wegen der linearen Unabhängigkeit von e_1, e_2 folgt auch $\lambda_1 = \lambda_2 = 0$. Die Vektoren e_1, e_2, b_1 sind daher linear unabhängig und bilden somit eine Basis von V. Daraus folgt

$$b_2 = \alpha_1 e_1 + \alpha_2 e_2 + \gamma b_1$$

mit reellen Zahlen α_1, α_2 und γ. Da die Vektoren b_1, b_2 orthogonal zu $\alpha_1 e_1 + \alpha_2 e_2$ sind, folgt weiter durch Bildung des inneren Produktes mit $\alpha_1 e_1 + \alpha_2 e_2$

$$\|\alpha_1 e_1 + \alpha_2 e_2\|^2 = 0 \quad \text{bzw.} \quad \alpha_1 e_1 + \alpha_2 e_2 = 0,$$

woraus $\alpha_1 = \alpha_2 = 0$ folgt. Es gilt somit $b_2 = \gamma b_1$. Je zwei Vektoren aus $\mathfrak{N} \cup \{o\}$ sind daher linear abhängig.

Um $\dim V = 1$ zu beweisen, genügt es zu zeigen, dass es in V vom Nullvektor verschiedene Vektoren gibt. Dies sieht man wie folgt. Zu den Vektoren e_1, e_2 existiert ein Vektor $a \in V$ derart, dass e_1, e_2, a eine Basis von V bilden. Wir suchen reelle Zahlen μ_1, μ_2 derart, dass

$$a_n = a - \mu_1 e_1 - \mu_2 e_2$$

orthogonal zu e_1, e_2 ist, d.h. $a_n \in \mathfrak{N}$ gilt (wegen der linearen Unabhängigkeit von e_1, e_2, a ist $a_n \neq o$). Es muß daher

$$\begin{aligned}
\|e_1\|^2 \mu_1 + \langle e_1, e_2 \rangle \mu_2 &= \langle a, e_1 \rangle, \\
\langle e_1, e_2 \rangle \mu_1 + \|e_2\|^2 \mu_2 &= \langle a, e_2 \rangle
\end{aligned}$$

gelten. Daraus folgt $\mu_1 = \|e_1\|^{-2}\big(\langle a, e_1 \rangle - \langle e_1, e_2 \rangle \mu_2\big)$ und

$$\big(\|e_1\|^2 \|e_2\|^2 - \langle e_1, e_2 \rangle^2\big)\mu_2 = \|e_1\|^2 \langle a, e_2 \rangle - \langle a, e_1 \rangle.$$

Da die Vektoren e_1, e_2 linear unabhängig sind, gilt $|\langle e_1, e_2 \rangle| < \|e_1\| \, \|e_2\|$ (siehe Abschnitt 2.2). Das Gleichungssystem für μ_1, μ_2 ist daher eindeutig lösbar. $\square$

Satz 3.14. a) $(\mathcal{P}, V)$ *sei ein zweidimensionaler euklidischer Raum und* g *sei eine Gerade in* $\mathcal{P}$ *mit* $P \in g$. *Ferner sei* $\mathfrak{n}$ *ein Normalenvektor von* g. *Dann gilt*

$$g = \{X \in \mathcal{P} \mid \langle \mathfrak{n}, \overrightarrow{OX} \rangle = \delta\}, \quad \textit{wobei } \delta = \langle \mathfrak{n}, \overrightarrow{OP} \rangle.$$

Darüberhinaus ist $V_g = \{a \in V \mid \langle \mathfrak{n}, a\rangle = 0\}$.

Ist umgekehrt $\mathfrak{n} \neq o$ *ein Vektor aus* V *und* $\delta \in \mathbb{R}$, *so ist*

$$g = \{X \in \mathcal{P} \mid \langle \mathfrak{n}, \overrightarrow{O\,X}\rangle = \delta\} \tag{3.13}$$

eine Gerade in $\mathcal{P}$ *(mit Normalenvektor* $\mathfrak{n}$*).*

b) $(\mathcal{P}, V)$ *sei ein dreidimensionaler euklidischer Raum und* E *sei eine Ebene in* $\mathcal{P}$ *mit* $P \in E$. *Ferner sei* $\mathfrak{n}$ *ein Normalenvektor von* E. *Dann gilt*

$$E = \{X \in \mathcal{P} \mid \langle \mathfrak{n}, \overrightarrow{O\,X}\rangle = \delta\}, \quad \textit{wobei } \delta = \langle \mathfrak{n}, \overrightarrow{O\,P}\rangle.$$

Darüberhinaus ist $V_E = \{a \in V \mid \langle \mathfrak{n}, a\rangle = 0\}$.

Ist umgekehrt $\mathfrak{n} \neq o$ *ein Vektor aus* V *und* $\delta \in \mathbb{R}$, *so ist*

$$E = \{X \in \mathcal{P} \mid \langle \mathfrak{n}, \overrightarrow{O\,X}\rangle = \delta\} \tag{3.14}$$

eine Ebene im $\mathcal{P}$ *(mit Normalenvektor* $\mathfrak{n}$*).*

Beweis. Wir führen den Beweis nur für Teil b). (e_1, e_2) sei eine Basis von V_E. Aus $\overrightarrow{O\,X} = \overrightarrow{O\,P} + \lambda e_1 + \mu e_2$ für $X \in E$ folgt $\langle \mathfrak{n}, \overrightarrow{O\,X}\rangle = \langle \mathfrak{n}, \overrightarrow{O\,P}\rangle =: \delta$ für alle $X \in E$. Damit ist

$$E \subset M := \{X \in \mathcal{P} \mid \langle \mathfrak{n}, \overrightarrow{O\,X}\rangle = \delta\}. \tag{3.15}$$

gezeigt. Bevor wir $E = M$ zeigen, beweisen wir, dass durch (3.14) eine Ebene definiert wird.

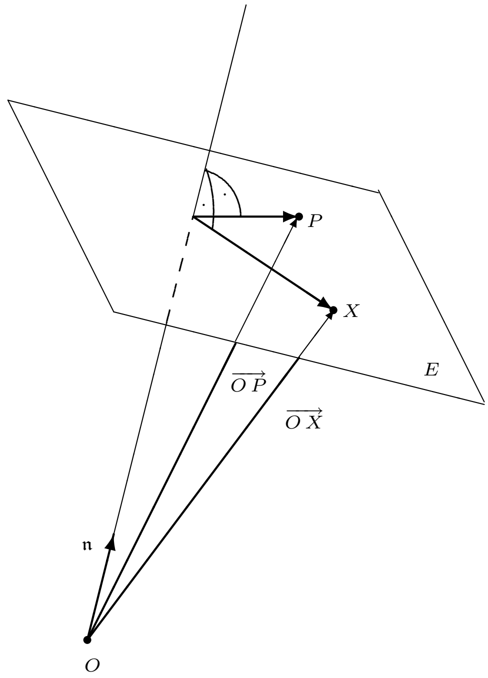
Es sei nun $\mathfrak{n} \neq o$ und $\delta \in \mathbb{R}$ gegeben. Wir setzen wieder

$$M = \{X \in \mathcal{P} \mid \langle \mathfrak{n}, \overrightarrow{O\,X}\rangle = \delta\}$$

und $V_M = \{\overrightarrow{X\,Y} \mid X, Y \in M\}$. Für $X, Y \in M$ gilt $\langle \mathfrak{n}, \overrightarrow{O\,X}\rangle = \langle \mathfrak{n}, \overrightarrow{O\,Y}\rangle$, d.h. $\langle \mathfrak{n}, \overrightarrow{X\,Y}\rangle = 0$. Gilt umgekehrt, für einen Vektor $a \in V$, $\langle \mathfrak{n}, a\rangle = 0$ und ist $X \in M$, so existiert ein $Y \in \mathcal{P}$ mit $\overrightarrow{X\,Y} = a$ (wegen (a1)). Daraus folgt $0 = \langle \mathfrak{n}, \overrightarrow{X\,Y}\rangle = \langle \mathfrak{n}, \overrightarrow{O\,Y}\rangle - \langle \mathfrak{n}, \overrightarrow{O\,X}\rangle$, d.h. $\langle \mathfrak{n}, \overrightarrow{O\,Y}\rangle = \langle \mathfrak{n}, \overrightarrow{O\,X}\rangle = \delta$ und somit $Y \in M$. Dann ist aber $a = \overrightarrow{X\,Y} \in V_M$. Wir haben damit

$$V_M = \{a \in V \mid \langle \mathfrak{n}, a\rangle = 0\}$$

bewiesen. Nach Lemma 3.11 ist V_M ein zweidimensionaler Vektorraum. Wir wählen $P \in M$ und $a \in V_M$. Wegen (a1) existiert ein Punkt $Q \in \mathcal{P}$ mit $\overrightarrow{P\,Q} = a$. Aus $\overrightarrow{O\,Q} = \overrightarrow{O\,P} + \overrightarrow{P\,Q}$ und $P \in M$ folgt $\langle \mathfrak{n}, \overrightarrow{O\,Q}\rangle = \langle \mathfrak{n}, \overrightarrow{O\,P}\rangle + \langle \mathfrak{n}, a\rangle = \delta + 0 = \delta$, d.h. $Q \in M$. Damit ist gezeigt, dass M eine Ebene ist. Dann ist aber $E \subset M$ gleichbedeutend mit $E = M$. $\square$



Es sei $(O; e_1, e_2)$ ein cartesisches Koordinatensystem des zweidimensionalen euklidischen Raumes $(\mathcal{P}, V)$. $\binom{\nu_1}{\nu_2}$ sei der Koordinatenvektor von $\mathfrak{n}$ (bzgl. e_1, e_2) und $\begin{smallmatrix}\xi_1\\ \xi_2\end{smallmatrix}$ sei der Koordinatenvektor von $\overrightarrow{OX}$, d.h. ξ_1, ξ_2 sind die affinen Koordinaten von X. Dann gilt genau dann $X \in g$, wenn

$$\nu_1\xi_1 + \nu_2\xi_2 = \delta. \tag{3.16}$$

Wählt man einen Normalenvektor $\mathfrak{n}$ von g mit $\|\mathfrak{n}\| = 1$, so nennt man $\langle \mathfrak{n}, \overrightarrow{OX}\rangle = \delta$ bzw. (3.16) eine ***Hessesche***[13] ***Normalform*** von g.

Ist $(O; e_1, e_2, e_3)$ ein cartesisches Koordinatensystem des dreidimensionalen euklidischen Raumes $(\mathcal{P}, V)$, so gilt genau dann $X \in E$, wenn

$$\nu_1\xi_1 + \nu_2\xi_2 + \nu_3\xi_3 = \delta, \tag{3.17}$$

wobei ν_1, ν_2, ν_3 bzw. ξ_1, ξ_2, ξ_3 die Koordinaten des Vektors $\mathfrak{n}$ bzw. $\overrightarrow{OX}$ sind. Die ξ_i, $i = 1, 2, 3$, sind gleichzeitig die affinen Koordinaten von X. Wenn $\mathfrak{n}$ mit $\|\mathfrak{n}\| = 1$ gewählt wurde, heißt $\langle \mathfrak{n}, \overrightarrow{OX}\rangle = \delta$ bzw. (3.17) eine ***Hessesche Normalform*** von E.

Satz 3.15. a) $(\mathcal{P}, V)$ *sei ein zweidimensionaler euklidischenr Raum und g sei eine Gerade in $\mathcal{P}$ mit der Hesseschen Normalform $\langle \mathfrak{n}, \overrightarrow{OX}\rangle = \delta > 0$. Dann ist δ der Abstand von g und O. Ist $\delta = 0$, so gilt $O \in g$.*

b) $(\mathcal{P}, V)$ *sei ein dreidimensionaler euklidischer Raum und E sei eine Ebene im $\mathcal{P}$ mit der Hesseschen Normalform*

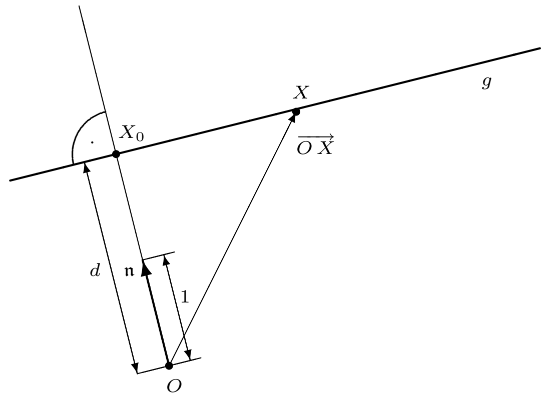
$$\langle \mathfrak{n}, \overrightarrow{OX}\rangle = \delta > 0.$$

Dann ist δ der Abstand von E und O. Im Falle $\delta = 0$ ist $O \in E$.

Beweis. a) Es gilt $\|\mathfrak{n}\| = 1$ und daher $\delta = \langle \mathfrak{n}, \overrightarrow{OX}\rangle = \|\overrightarrow{OX}\| \cos\varphi$, φ der Winkel zwischen $\mathfrak{n}$ und $\overrightarrow{OX}$. Für jeden Punkt $X \in g$ ist daher der Abstand zwischen X und O gegeben durch $\|\overrightarrow{OX}\| = \delta/\cos\varphi \geq \delta$. Für X_0 mit dem Ortsvektor $\overrightarrow{OX_0} = \delta\mathfrak{n}$ gilt $\langle \mathfrak{n}, \overrightarrow{OX_0}\rangle = \delta\|\mathfrak{n}\|^2 = \delta$ und $\|\overrightarrow{OX_0}\| = d$, d.h. $X_0 \in g$ und der Abstand zwischen X_0 und O ist δ.

[13] Hesse, Ludwig Otto, 22. 4. 1811 (Königsberg) – 4. 8. 1874 (München), Arbeiten zu algebraischen Invarianten (en.wikipedia.ord/wiki/Otto_Hesse).

b) Der Beweis ist völlig analog dem unter a) gegebenen Beweis. □



Satz 3.16. $(\mathcal{P}, V)$ *sei ein zwei- bzw. dreidimensionaler euklidischer Raum. Zwei Gerade in* $\mathcal{P}$ *bzw. zwei Ebenen in* $\mathcal{P}$ *sind genau dann parallel, wenn sie dieselben Normalenvektoren haben.*

Beweis. Es seien E_1 und E_2 Ebenen im $\mathcal{P}$ mit demselben Normalenvektor $\mathfrak{n}$. Dann folgt wegen Satz 3.14

$$V_{E_1} = V_{E_2} = \{a \in V \mid \langle \mathfrak{n}, a \rangle = 0\},$$

d.h. $E_1 \| E_2$.

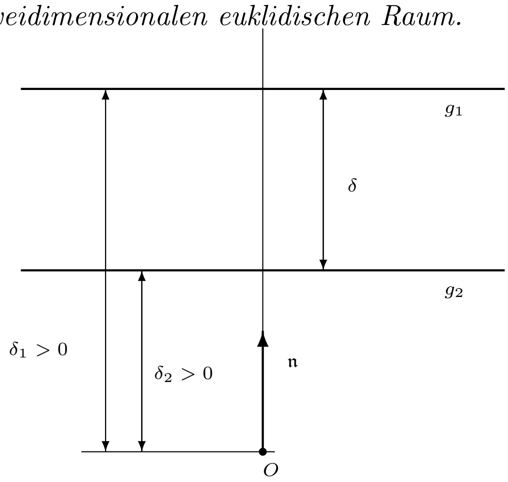
Es sei umgekehrt $E_1 \| E_2$, d.h. $V_{E_1} = V_{E_2}$. Dann ist jeder Normalenvektor von E_1 auch einer von E_2 und umgekehrt (siehe Definition 3.12). Der Beweis für Gerade in einem zweidimensionalen euklidischen Raum ist völlig analog. □

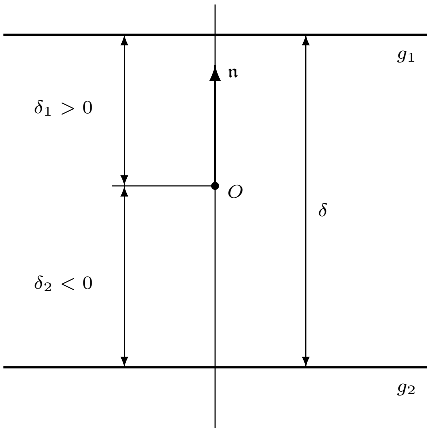
Unmittelbare Folgerungen der Sätze 3.15 und 3.16 sind:

Es sei $(\mathcal{P}, V)$ *ein dreidimensionaler euklidischer Raum und* E_i, $i = 1, 2$, *parallele Ebenen im* $\mathcal{P}$ *mit den Hesseschen Normalformen*

$$\langle \mathfrak{n}, \overrightarrow{OX} \rangle = \delta_i, \quad i = 1, 2.$$

Dann ist $\delta = |\delta_1 - \delta_2|$ *der Abstand zwischen* E_1 *und* E_2. *Der Koordinatenursprung* O *liegt genau dann zwischen* E_1 *und* E_2, *wenn* $\delta_1 \delta_2 < 0$ *ist. Analoges gilt für Gerade in einem zweidimensionalen euklidischen Raum.*

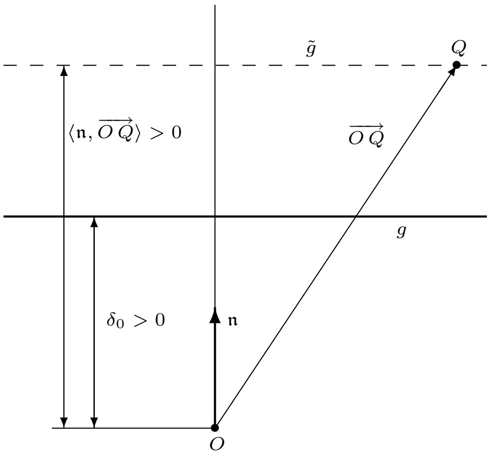




Es sei $(\mathcal{P}, V)$ ein dreidimensionaler euklidischer Raum und E eine Ebene im $\mathcal{P}$ mit der Hesseschen Normalform $\langle \mathfrak{n}, \overrightarrow{O\,X} \rangle = \delta_0$. Ist $Q \in \mathcal{P}$, so ist der Abstand von Q und E durch

$$\delta = |\delta_0 - \langle \mathfrak{n}, \overrightarrow{O\,Q} \rangle|$$

gegeben. Analoges gilt für Gerade in einem zweidimensionalen euklidischen Raum.



Zum Beweis beachte man, dass $\langle \mathfrak{n}, \overrightarrow{O\,X} \rangle = \langle \mathfrak{n}, \overrightarrow{O\,Q} \rangle$ eine Hessesche Normalform einer Ebene $\tilde{E}$ bzw. einer Geraden $\tilde{g}$ durch Q parallel zu E bzw. zu g ist.

3.4 Windschiefe Gerade*

Definition 3.17. *$(\mathcal{P}, V)$ sei ein dreidimensionaler euklidischer Raum und g_1, g_2 seien Gerade in $\mathcal{P}$. Die Geraden g_1 und g_2 heißen genau dann **windschief**, wenn*

(i) $g_1 \cap g_2 = \emptyset$,

(ii) $g_1 \nparallel g_2$

gilt.

Um den Abstand δ zweier windschiefer Geraden zu bestimmen, geht man wie folgt vor:

Wir wählen Basisvektoren e_1 von V_{g_1} und e_2 von V_{g_2}. Die Vektoren e_1, e_2 sind linear unabhängig (warum?). Es sei E eine Ebene durch einen Punkt $P \in g_1$ mit $V_E = \{\lambda e_1 + \mu e_2 \mid$

$\lambda, \mu \in \mathbb{R}\}$. Für diese Ebene gilt (Beweis!)

$$g_1 \subset E \text{ und } g_2 \| E.$$

δ ist gegeben durch den Abstand eines beliebigen Punktes $Q \in g_2$ von E. Als Endresultat dieses Lösungsweges erhält man die Formel (Beweis!)[14]

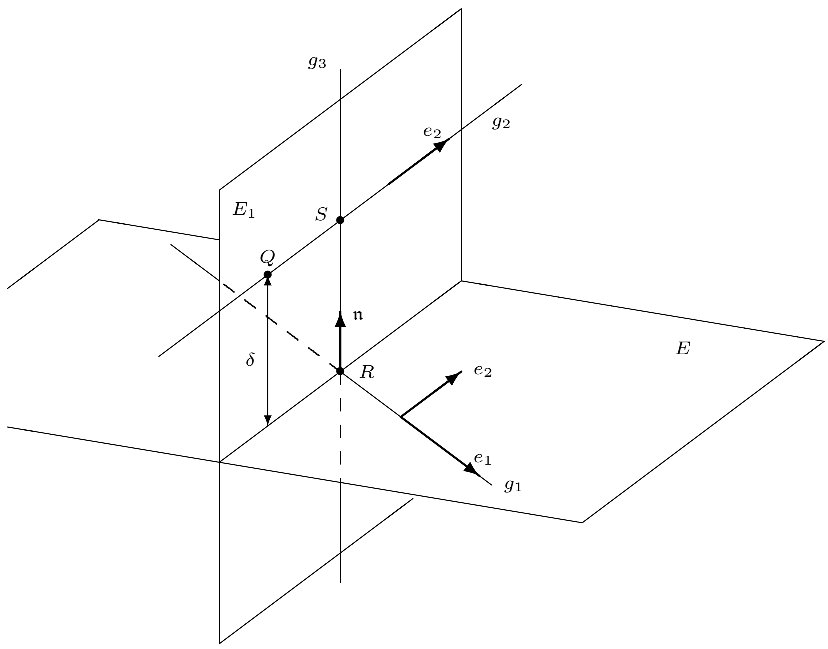
$$\delta = \frac{1}{\|e_1 \times e_2\|} |\langle e_1 \times e_2, \overrightarrow{OQ} - \overrightarrow{OP} \rangle| = \frac{|\langle e_1 \times e_2, \overrightarrow{PQ} \rangle|}{\|e_1 \times e_2\|}$$

Um die in der Zeichnung angegebene Gerade g_3 zu bestimmen, welche g_1 und g_2 unter einem rechten Winkel schneidet, bestimmt man zuerst eine Ebene E_1, welche g_2 enthält und orthogonal zu E ist. Ist $\mathfrak{n}$ ein Normalenvektor von E (etwa $\mathfrak{n} = e_1 \times e_2$) , so ist $\mathfrak{n} \times e_2$ ein Normalenvektor von E_1.

Der Schnittpunkt von E_1 und g_1 sei R (warum existiert R?). Dann ist durch

$$\overrightarrow{OY} = \overrightarrow{OR} + \lambda \mathfrak{n}$$

eine Parameterdarstellung von g_3 gegeben. Den Punkt S errechnet man als Schnitt von g_3 und g_2.



Eine alternative Möglichkeit ist die im Folgenden beschriebene. Mit den Vektoren e_1, e_2 und Punkten $P_1 \in g_1$, $P_2 \in g_2$ erhalten wir die folgenden Parameterdarstellungen von g_1 und g_2:

$$\overrightarrow{OX_1} = \overrightarrow{OP_1} + \lambda e_1, \quad \overrightarrow{OX_2} = \overrightarrow{OP_2} + \mu e_2,$$

wobei wir $\|e_1\| = \|e_2\| = 1$ voraussetzen können. Wegen der linearen Unabhängigkeit der Vektoren e_1, e_2 folgt aus der Schwarzschen Ungleichung (siehe Seite 20)

$$|\langle e_1, e_2 \rangle| < \|e_1\| \, \|e_2\| = 1. \tag{3.18}$$

Die Punkte $R \in g_2$ und $S \in g_1$ in der Zeichnung sind durch

$$\overrightarrow{RS} \perp e_1 \quad \text{und} \quad \overrightarrow{RS} \perp e_2 \tag{3.19}$$

[14] Das äußere Produkt „$\times$“ in einem dreidimensionalen euklidischen Raum wird durch (2.17) definiert, wobei als Basis eine Orthonormalbasis von V zugrunde gelegt wird.

charakterisiert. Mit Hilfe der Parameterdarstellungen von g_1 und g_2 erhält man

$$\overrightarrow{OR} = \overrightarrow{OP_2} + \mu_0 e_2, \quad \overrightarrow{OS} = \overrightarrow{OP_1} + \lambda_0 e_1$$

und

$$\overrightarrow{RS} = \overrightarrow{P_2P_1} + \lambda_0 e_1 - \mu_0 e_2.$$

Diese Darstellung von $\overrightarrow{RS}$ und (3.19) ergeben unter Beachtung von $\|e_1\| = \|e_2\| = 1$ die Gleichungen

$$\begin{aligned} \lambda_0 - \alpha\mu_0 &= \langle \overrightarrow{P_1P_2}, e_1 \rangle =: \beta_1, \\ \alpha\lambda_0 - \mu_0 &= \langle \overrightarrow{P_1P_2}, e_2 \rangle =: \beta_2. \end{aligned} \tag{3.20}$$

Unter Beachtung von (3.18) erhalten wir

$$\lambda_0 = \frac{\beta_1 - \alpha\beta_2}{1-\alpha^2}, \quad \mu_0 = \frac{\alpha\beta_1 - \beta_2}{1-\alpha^2}.$$

Damit können wir die Punkte R und S berechnen. Der Abstand von g_1 und g_2 ist durch $\|\overrightarrow{RS}\|$ gegeben.

Die folgenden Aufgaben sollten ohne große Schwierigkeiten gelösten werden:

1. $(\mathcal{P}, V)$ sei ein dreidimensionaler euklidischer Raum und g_1, g_2 seien windschiefe Gerade in $\mathcal{P}$ und P sei ein nicht auf g_1, g_2 liegender Punkt. Gesucht ist eine Gerade g mit $P \in g$ und $g \cap g_i \neq \emptyset$, $i = 1, 2$. Es ist insbesondere zu untersuchen, wann diese Aufgabe lösbar ist.

2. $(\mathcal{P}, V)$ sei ein dreidimensionaler euklidischer Raum und g_1, g_2 seien windschiefe Gerade in $\mathcal{P}$ und a sei ein Vektor in V. Gesucht ist eine Gerade g mit a als Basis für V_g und $g \cap g_i \neq \emptyset$, $i = 1, 2$.

3. $(\mathcal{P}, V)$ sei ein dreidimensionaler euklidischer Raum und g_i, $i = 1, 2, 3$, seien Gerade in der Ebene E mit $g_2 \nparallel g_3$. Dann gilt $g_1 \cap g_2 \neq \emptyset$ oder $g_1 \cap g_3 \neq \emptyset$. Man gebe eine vollständige Diskussion aller Möglichkeiten.

3.5 Dreiecksaufgaben*

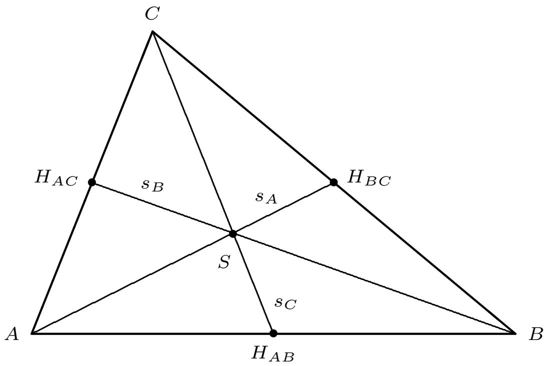
Im folgenden sei $(\mathcal{P}, V)$ stets ein zwei- bzw. dreidimensionaler euklidischer Raum. Zur Veranschaulichung kann man sich darunter die Punkte und Vektoren im $\mathbb{R}^2$ bzw. im $\mathbb{R}^3$ vorstellen (siehe Kapitel 2). Für die Eckpunkte eines Dreieckes wird stets vorausgesetzt, dass sie nicht kollinear sind.

3.5.1 Der Schwerpunkt

Es sei ein Dreieck mit den Eckpunkten A, B, C gegeben. Die ***Schwerlinie*** s_A sei jene Gerade, welche durch den Punkt A und den Halbierungspunkt H_{BC} der Seite $\overrightarrow{BC}$ bestimmt ist. Analog seien s_B und s_C definiert.

Satz 3.18. *Die Schwerlinien s_A, s_B und s_C schneiden sich alle in einem Punkt S, dem **Schwerpunkt** des Dreieckes. Dieser teilt die Strecke von A nach H_{BC} im Verhältnis $2:1$, d.h. $\overrightarrow{AS} = 2\overrightarrow{SH_{BC}}$ etc. Ferner gilt*

$$\overrightarrow{OS} = \frac{1}{3}(\overrightarrow{OA} + \vec{O}B + \overrightarrow{OC}).$$



Beweis. Zunächst gilt

$$\overrightarrow{OH_{BC}} = \frac{1}{2}(\overrightarrow{OB} + \overrightarrow{OC}),\ \overrightarrow{OH_{AC}} = \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OC}),\ \overrightarrow{OH_{AB}} = \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OB}).$$

Die Parameterdarstellungen von s_A, s_B, s_C sind dann:

$$\begin{aligned}
s_A: \quad \overrightarrow{OX} &= \frac{1}{2}(\overrightarrow{OB} + \overrightarrow{OC}) + \lambda(\overrightarrow{OA} - \frac{1}{2}(\overrightarrow{OB} + \overrightarrow{OC})),\\
s_B: \quad \overrightarrow{OX} &= \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OC}) + \mu(\overrightarrow{OB} - \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OC})),\\
s_C: \quad \overrightarrow{OX} &= \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OB}) + \nu(\overrightarrow{OC} - \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OB})),
\end{aligned}$$

bzw.

$$\begin{aligned}
s_A: \quad \overrightarrow{OX} &= \lambda\overrightarrow{OA} + \frac{1}{2}(1-\lambda)(\overrightarrow{OB} + \overrightarrow{OC}),\\
s_B: \quad \overrightarrow{OX} &= \mu\overrightarrow{OB} + \frac{1}{2}(1-\mu)(\overrightarrow{OA} + \overrightarrow{OC}),\\
s_C: \quad \overrightarrow{OX} &= \nu\overrightarrow{OC} + \frac{1}{2}(1-\nu)(\overrightarrow{OA} + \overrightarrow{OB}).
\end{aligned}$$

Man sieht, dass für $\lambda = \mu = \nu = \frac{1}{3}$ jeweils

$$\overrightarrow{OS} = \overrightarrow{OX} = \frac{1}{3}(\overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC}).$$

gilt. Als Übungsaufgaben beweise man, dass etwa s_A und s_B nicht parallel sind. Daher ist S der einzige Schnittpunkt.

Ferner gilt

$$\overrightarrow{AS} = \overrightarrow{OS} - \overrightarrow{OA} = -\frac{2}{3}\overrightarrow{OA} + \frac{1}{3}\overrightarrow{OB} + \frac{1}{3}\overrightarrow{OC}$$

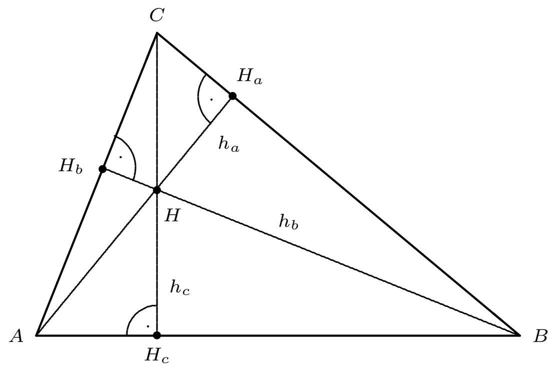
und

$$\begin{aligned}
\overrightarrow{SH_{BC}} &= \frac{1}{2}(\overrightarrow{OB} + \overrightarrow{OC}) - \frac{1}{3}(\overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC})\\
&= -\frac{1}{3}\overrightarrow{OA} + \frac{1}{6}\overrightarrow{OB} + \frac{1}{6}\overrightarrow{OC}.
\end{aligned}$$

□

3.5.2 Der Höhenschnittpunkt

Satz 3.19. *Die Höhenlinien eines Dreieckes schneiden sich in einem Punkt H, dem* ***Höhenschnittpunkt****.*



Beweis. Es sei H der Schnittpunkt von h_a und h_b, d.h. es gilt $\overrightarrow{AH}\perp\overrightarrow{BC}$ und $\overrightarrow{BH}\perp\overrightarrow{AC} = \overrightarrow{AB}+\overrightarrow{BC}$. Es genügt zu zeigen, dass $\overrightarrow{CH}\perp\overrightarrow{AB}$ ist.

An dem Dreieck ABH sieht man

$$\overrightarrow{BH} = \overrightarrow{AH} - \overrightarrow{AB}$$

und an Hand des Dreieckes BCH verifiziert man

$$\overrightarrow{CH} = \overrightarrow{BH} - \overrightarrow{BC}.$$

Dann gilt

$$\begin{aligned}\langle\overrightarrow{CH},\overrightarrow{AB}\rangle &= \langle\overrightarrow{BH}-\overrightarrow{BC},\overrightarrow{AC}-\overrightarrow{BC}\rangle\\ &= \langle\overrightarrow{BH},\overrightarrow{AC}\rangle - \langle\overrightarrow{BH},\overrightarrow{BC}\rangle - \langle\overrightarrow{BC},\overrightarrow{AC}-\overrightarrow{BC}\rangle\\ &= 0 - \langle\overrightarrow{BC},\overrightarrow{BH}+\overrightarrow{AC}-\overrightarrow{BC}\rangle = -\langle\overrightarrow{BC},\overrightarrow{AH}\rangle = 0.\end{aligned}$$

Den Punkt H berechnen wir als Schnitt von h_c und h_b:

$$\overrightarrow{OH} = \overrightarrow{OC} + \lambda\mathfrak{n}_c = \overrightarrow{OB} + \mu\mathfrak{n}_b, \text{ d.h. } \overrightarrow{BC} + \lambda\mathfrak{n}_c = \mu\mathfrak{n}_b,$$

wobei $\mathfrak{n}_b = \overrightarrow{H_bB}$ und $\mathfrak{n}_c = \overrightarrow{H_cC}$. Beachtet man $\langle\mathfrak{n}_c,\overrightarrow{AB}\rangle = 0$ und $\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle \neq 0$, so erhält man

$$\langle\overrightarrow{AB},\overrightarrow{BC}\rangle = \mu\langle\overrightarrow{AB},\mathfrak{n}_b\rangle, \text{ d.h. } \mu = \frac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle}$$

und $\overrightarrow{OH} = \overrightarrow{OB} + \dfrac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle}\mathfrak{n}_b$.

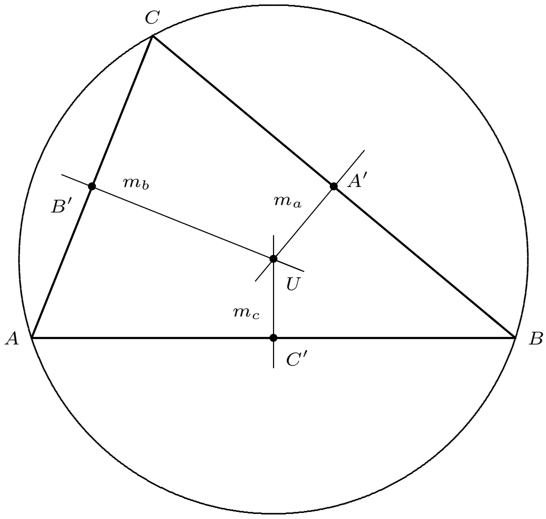
Beachtet man noch, dass $\mathfrak{n}_b$ die Normalkomponente von $\overrightarrow{CB}$ bezüglich $\overrightarrow{AC}$ ist, d.h. $\mathfrak{n}_b = -\overrightarrow{BC}+\langle\overrightarrow{BC},\overrightarrow{AC}\rangle/\|\overrightarrow{AC}\|^2\overrightarrow{AC}$, so erhält man nach einiger Rechnung unter Verwendung von Satz 2.11, a) und b), den Ortsvektor von H explizit:

$$\overrightarrow{OH} = \overrightarrow{OB} + \frac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{\langle\overrightarrow{AC}\times\overrightarrow{BC},\overrightarrow{AB}\times\overrightarrow{AC}\rangle}\Big(\overrightarrow{AC}\times(\overrightarrow{AC}\times\overrightarrow{BC})\Big).$$

□

3.5.3 Der Umkreismittelpunkt

Satz 3.20. *Die drei **Seitensymmetralen** eines Dreieckes schneiden sich in einem Punkt U, dem **Umkreismittelpunkt**.*



Beweis. A', B', C' seien die Halbierungspunkte der Dreieckseiten und U sei der Schnittpunkt der Seitensymmetralen m_c und m_a. Es genügt $\overrightarrow{B'U} \perp \overrightarrow{AC}$ zu zeigen.

Es gilt zunächst $\langle \overrightarrow{C'U}, \overrightarrow{AB} \rangle = 0 = \langle \overrightarrow{A'U}, \overrightarrow{BC} \rangle$ und $\overrightarrow{C'U} = \overrightarrow{AU} - \frac{1}{2}\overrightarrow{AB}$, $\overrightarrow{A'U} = \overrightarrow{BU} - \frac{1}{2}\overrightarrow{BC}$. Aus den letzten zwei Gleichungen folgt unter Beachtung von $\overrightarrow{BU} = \overrightarrow{AU} - \overrightarrow{AB}$

$$\overrightarrow{C'U} = \overrightarrow{A'U} + \frac{1}{2}\overrightarrow{AC}.$$

Ferner gilt $\overrightarrow{B'U} = \overrightarrow{AU} - \frac{1}{2}\overrightarrow{AC} = \overrightarrow{AU} - \frac{1}{2}\overrightarrow{AB} - \frac{1}{2}\overrightarrow{BC} = \overrightarrow{C'U} - \frac{1}{2}\overrightarrow{BC}$. Damit erhalten wir

$$\begin{aligned}
\langle \overrightarrow{B'U}, \overrightarrow{AC} \rangle &= \langle \overrightarrow{C'U} - \frac{1}{2}\overrightarrow{BC}, \overrightarrow{AB} + \overrightarrow{BC} \rangle \\
&= \langle \overrightarrow{C'U}, \overrightarrow{BC} \rangle - \frac{1}{2}\langle \overrightarrow{BC}, \overrightarrow{AC} \rangle \\
&= \langle \overrightarrow{BC}, \overrightarrow{C'U} - \frac{1}{2}\overrightarrow{AC} \rangle = \langle \overrightarrow{BC}, \overrightarrow{A'U} + \frac{1}{2}\overrightarrow{AC} - \frac{1}{2}\overrightarrow{AC} \rangle \\
&= \langle \overrightarrow{BC}, \overrightarrow{A'U} \rangle = 0.
\end{aligned}$$

Den Umkreismittelpunkt bestimmen wir als Schnittpunkt von m_c und m_b:

$$\overrightarrow{OC'} + \lambda \mathfrak{n}_c = \overrightarrow{OB'} + \mu \mathfrak{n}_b, \text{ d.h. } \overrightarrow{B'C'} + \lambda \mathfrak{n}_c = \mu \mathfrak{n}_b,$$

wobei $\mathfrak{n}_b = \overrightarrow{H_b B}$ und $\mathfrak{n}_c = \overrightarrow{H_c C}$ (siehe den vorhergehenden Abschnitt). Daraus folgt $\langle \overrightarrow{AB}, \overrightarrow{B'C'} \rangle = \mu \langle \mathfrak{n}_b, \overrightarrow{AB} \rangle$ und weiter

$$\mu = \frac{\langle \overrightarrow{AB}, \overrightarrow{B'C'} \rangle}{\langle \mathfrak{n}_b, \overrightarrow{AB} \rangle}.$$

Beachtet man noch $\overrightarrow{B'C'} = \overrightarrow{OC'} - \overrightarrow{OB'} = \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OB}) - \frac{1}{2}(\overrightarrow{OA} + \overrightarrow{OC}) = \frac{1}{2}\overrightarrow{CB} = -\frac{1}{2}\overrightarrow{BC}$, so ist

$$\overrightarrow{OU} = \overrightarrow{OB'} - \frac{1}{2}\frac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle}\mathfrak{n}_b = \frac{1}{2}\Big(\overrightarrow{OA} + \overrightarrow{OC} - \frac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle}\mathfrak{n}_b\Big).$$

□

3.5.4 Die Eulersche Gerade

Satz 3.21. *Die Punkte S, H und U liegen auf einer Geraden, der **Eulerschen**[15] **Geraden**. Ferner gilt*

$$\overrightarrow{SH} = -2\overrightarrow{SU}.$$

Beweis. Aus $\overrightarrow{OS} = \frac{1}{3}(\overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC})$ und der Darstellung von $\overrightarrow{OH}$ und $\overrightarrow{OU}$ folgt :

$$\begin{aligned}\overrightarrow{SH} = \overrightarrow{OH} - \overrightarrow{OS} &= \overrightarrow{OB} - \frac{1}{3}(\overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC}) + \frac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle}\mathfrak{n}_b \\ &= -\frac{1}{3}\overrightarrow{OA} + \frac{2}{3}\overrightarrow{OB} - \frac{1}{3}\overrightarrow{OC} + \frac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle}\mathfrak{n}_b\end{aligned}$$

und

$$\begin{aligned}-\overrightarrow{SU} = \overrightarrow{US} &= -\frac{1}{2}\overrightarrow{OA} - \frac{1}{2}\overrightarrow{OC} + \frac{1}{3}\big(\overrightarrow{OA} + \overrightarrow{OB} + \overrightarrow{OC}\big) + \frac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{2\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle}\mathfrak{n}_b \\ &= -\frac{1}{6}\overrightarrow{OA} + \frac{1}{3}\overrightarrow{OB} - \frac{1}{6}\overrightarrow{OC} + \frac{1}{2}\frac{\langle\overrightarrow{AB},\overrightarrow{BC}\rangle}{\langle\mathfrak{n}_b,\overrightarrow{AB}\rangle}\mathfrak{n}_b = \frac{1}{2}\overrightarrow{SH}.\end{aligned}$$

Die Parameterdarstellung der Eulerschen Geraden ist durch

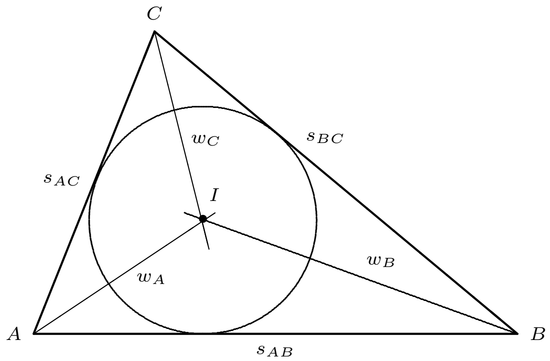
$$\overrightarrow{OX} = \overrightarrow{OS} + \lambda\overrightarrow{SH}$$

gegeben. □

3.5.5 Der Inkreismittelpunkt

Satz 3.22. *Die Winkelsymmetralen eines Dreieckes schneiden sich in einem Punkt I, dem **Inkreismittelpunkt** des Dreieckes.*

[15] Euler, Leonhard, 15. 4. 1707 (Riehen, Schweiz) – 18. 9. 1783 (St. Petersburg), einer der bedeutendsten Mathematiker aller Zeiten, 886 Publikationen (de.wikipedia.org/wiki/Leonhard_Euler).



Beweis. Jeder Punkt der Winkelsymmetralen w_A bzw. w_B hat von den Seiten s_{AB} und s_{AC} bzw. s_{BC} und s_{AB} denselben Abstand. Für den Schnittpunkt I von w_A und w_B gilt daher, dass dessen Abstand von s_{AB}, s_{AC} und s_{BC} gleich ist. Also liegt I auch auf w_c.

Für die Berechnung von I benutzt man die Parameterdarstellungen

$$\begin{aligned} w_A: \quad & \overrightarrow{OX} = \overrightarrow{OA} + \lambda\Big(\frac{\overrightarrow{AB}}{\|\overrightarrow{AB}\|} + \frac{\overrightarrow{AC}}{\|\overrightarrow{AC}\|}\Big), \\ w_B: \quad & \overrightarrow{OX} = \overrightarrow{OB} + \mu\Big(\frac{\overrightarrow{BA}}{\|\overrightarrow{BA}\|} + \frac{\overrightarrow{BC}}{\|\overrightarrow{BC}\|}\Big). \end{aligned}$$

Aus diesen Darstellungen erhält man

$$\begin{aligned} &\Big(1 - \frac{\lambda}{\|\overrightarrow{AB}\|} - \frac{\lambda}{\|\overrightarrow{AC}\|} - \frac{\mu}{\|\overrightarrow{AB}\|}\Big)\overrightarrow{OA} \\ &\quad + \Big(-1 + \frac{\lambda}{\|\overrightarrow{AB}\|} + \frac{\mu}{\|\overrightarrow{AB}\|} + \frac{\mu}{\|\overrightarrow{BC}\|}\Big)\overrightarrow{OB} + \Big(\frac{\lambda}{\|\overrightarrow{AC}\|} - \frac{\mu}{\|\overrightarrow{BC}\|}\Big)\overrightarrow{OC} = o. \end{aligned}$$

Die Summe der Koeffizienten ist Null. Da die Punkte A, B und C nicht kollinear sind, müssen die Koeffizienten alle Null sein (siehe Satz 3.10, a)). Daraus folgt

$$\lambda = \frac{\|\overrightarrow{AB}\|\,\|\overrightarrow{AC}\|}{\|\overrightarrow{AC}\| + \|\overrightarrow{AB}\| + \|\overrightarrow{BC}\|}$$

und damit

$$\begin{aligned} \overrightarrow{OI} = \overrightarrow{OA} &+ \frac{\|\overrightarrow{AC}\|}{\|\overrightarrow{AC}\| + \|\overrightarrow{AB}\| + \|\overrightarrow{BC}\|}\overrightarrow{AB} \\ &+ \frac{\|\overrightarrow{AB}\|}{\|\overrightarrow{AC}\| + \|\overrightarrow{AB}\| + \|\overrightarrow{BC}\|}\overrightarrow{AC}. \end{aligned}$$

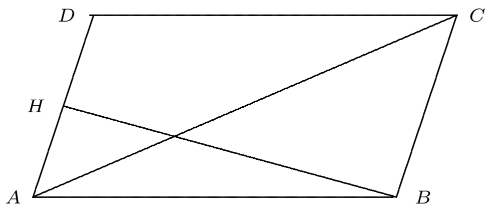
□

Übungsaufgaben.

1. Die Halbierungspunkte der Seiten eines ebenen Viereckes bilden ein Parallelogramm.
2. Ein ebenes Viereck ist genau dann ein Parallelogramm, wenn der Schnittpunkt der Diagonalen diese halbiert.

3. Ein Parallelogramm ist genau dann ein Rhombus (= Parallelogramm, in dem alle Seiten gleich lang sind), wenn die Diagonalen orthogonal sind.

4. Durch A, B, C, D sei ein Parallelogramm gegeben. Die durch den Halbierungspunkt H von A und D und den Punkt B führende Gerade teilt die Diagonale von A nach C im Verhältnis 1 : 2.



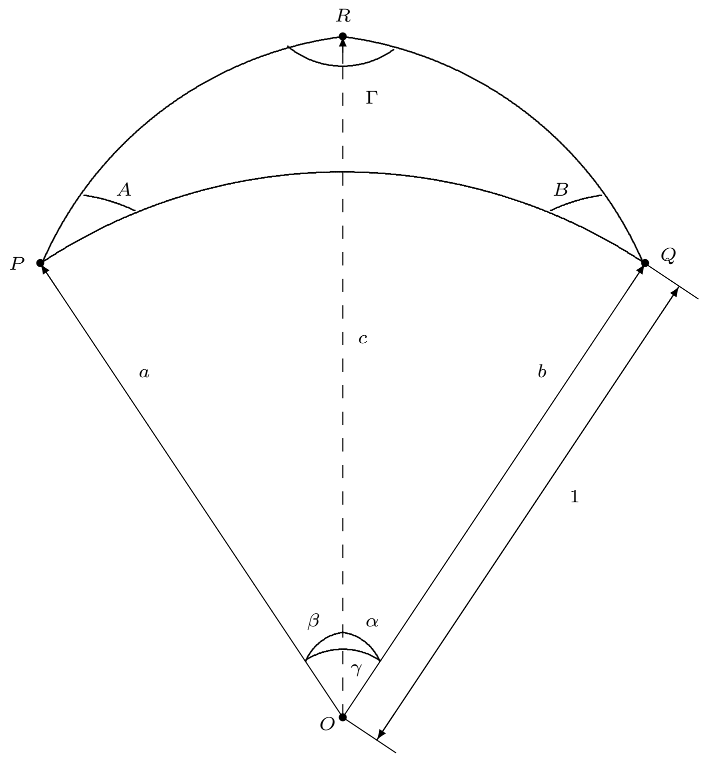
3.6 Geometrische Anwendungen des äußeren Produktes*

Der Einfachheit halber sei in diesem Abschnitt $(\mathcal{P}, V)$ der euklidische Raum der Punkte und Vektoren im $\mathbb{R}^3$.

3.6.1 Der Sinussatz der sphärischen Trigonometrie

Satz 3.23 (Sinussatz). *Auf einer Kugel mit Radius* 1 *und Mittelpunkt in* O *seien die Punkte* P, Q, R *derart gegeben, dass die zugehörigen Ortsvektoren ein Rechtssystem bilden. Für die in der untenstehenden Zeichnung eingetragenen Winkel gilt:*

$$\frac{\sin A}{\sin \alpha} = \frac{\sin B}{\sin \beta} = \frac{\sin \Gamma}{\sin \gamma}$$



Beweis. Wir setzen $a = \overrightarrow{OP}$, $b = \overrightarrow{OQ}$, $c = \overrightarrow{OR}$ und $V = \langle\langle a, b, c\rangle\rangle$. Dann gilt nach Satz 2.11, b),

$$(a \times b) \times (a \times c) = \langle\langle a, b, c\rangle\rangle a - \langle\langle a, b, a\rangle\rangle c = V a$$

und (man beachte $\|a\| = 1$)

$$\|(a \times b) \times (a \times c)\| = V.$$

Analog zeigt man

$$\|(b \times c) \times (b \times a)\| = \|(c \times a) \times (c \times b)\| = V.$$

Die Vektoren $a \times b$ bzw. $a \times c$ sind Normalenvektoren der durch die Punkte O, P, Q bzw. O, P, R aufgespannten Ebene. Der Winkel zwischen diesen Vektoren ist daher A. Es gilt somit

$$\|(a \times b) \times (a \times c)\| = \|a \times b\|\, \|a \times c\| \sin A = \sin\gamma \sin\beta \sin A.$$

Analog zeigt man $\|(b \times c) \times (b \times a)\| = \sin\alpha \sin\gamma \sin B$ und $\|(c \times a) \times (c \times b)\| = \sin\alpha \sin\beta \sin\Gamma$.
□

3.6.2 Der Kosinussatz der sphärischen Trigonometrie

Satz 3.24 (Kosinussatz). *Unter den Voraussetzungen des Sinussatzes gilt*

$$\cos\alpha = \cos\beta \cos\gamma + \sin\beta \sin\gamma \cos A.$$

Beweis. Es ist einerseits nach Satz 2.11, c),

$$\begin{aligned}\langle a \times b, a \times c\rangle &= \langle a, a\rangle\langle b, c\rangle - \langle a, c\rangle\langle a, b\rangle \\ &= \cos\alpha - \cos\beta\cos\gamma.\end{aligned}$$

Andererseits ist (A ist der Winkel zwischen $a \times b$ und $a \times c$)

$$\begin{aligned}\langle a \times b, a \times c \rangle &= \|a \times b\| \, \|a \times c\| \cos A \\ &= \sin\gamma \sin\beta \cos A.\end{aligned}$$

□

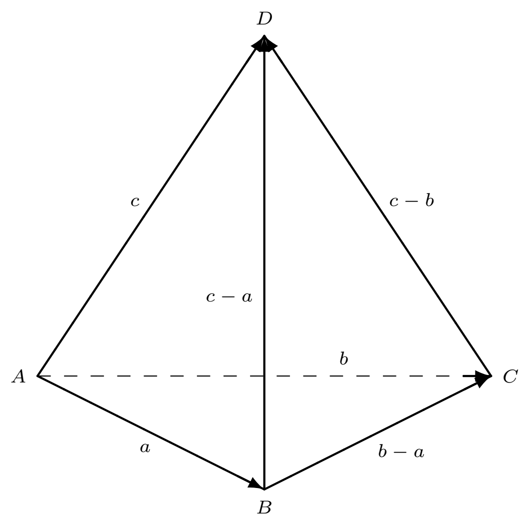
3.6.3 Ein Satz von Gauß[16]

Satz 3.25. *In einem Tetraeder ist die Summe der nach außen gerichteten Flächennormalen, deren Norm jeweils gleich dem Flächeninhalt der Seitenfläche ist, gleich dem Nullvektor o.*

Beweis. Wir setzen $a = \overrightarrow{AB}$, $b = \overrightarrow{AC}$ und $c = \overrightarrow{AD}$. Dann sind die Flächennormalen durch

$$x_1 = \frac{1}{2}(b \times a), \ x_2 = \frac{1}{2}(a \times c), \ x_3 = \frac{1}{2}(b - a) \times (c - a), \ x_4 = \frac{1}{2}(c \times b).$$

gegeben.



Daher ist

$$x_1 + x_2 + x_3 + x_4 = \frac{1}{2}\Big(b \times a + a \times c + b \times c - a \times c - b \times a + a \times a + c \times b\Big) = o.$$

□

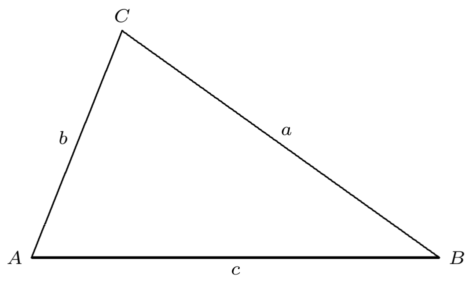
3.6.4 Die Heron'sche Flächenformel

Satz 3.26 (Heron'sche[17]Flächenformel). *Es seien a, b, c die Seitenlängen eines Dreieckes und $s = \frac{1}{2}(a + b + c)$. Dann gilt*

$$F_\Delta = \Big(s(s-a)(s-b)(s-c)\Big)^{1/2}.$$

[16] Gauß, Carl Friedrich, 30. 4. 1777 (Braunschweig) – 23. 2. 1855 (Göttingen), neben Euler, Newton und Archimedes einer der bedeutendsten Mathematiker aller Zeiten, "princeps mathematicorum"(de.wikipedia.org/wiki/Carl_Friedrich_Gau%C3%9F).

[17] Heron von Alexandria, ca. 10 (Alexandria?) – ca. 75, wahrscheinlich Lehrer am Museon in Alexandria, Werke über Mathematik, Mechanik, Pneumatik (www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Heron.html).



Beweis. Wir setzen $c = \|\overrightarrow{AB}\|$, $b = \|\overrightarrow{AC}\|$, $a = \|\overrightarrow{BC}\|$. Dann gilt (siehe Definition 2.6)

$$\begin{aligned}
F_\Delta^2 &= \frac{1}{4}\|\overrightarrow{AB} \times \overrightarrow{AC}\|^2 = \frac{1}{4}\Big\langle \overrightarrow{AB} \times \overrightarrow{AC}, \overrightarrow{AB} \times \overrightarrow{AC} \Big\rangle \\
&= \frac{1}{4}\Big(\|\overrightarrow{AB}\|^2\|\overrightarrow{AC}\|^2 - \langle \overrightarrow{AB}, \overrightarrow{AC}\rangle^2\Big) \\
&= \frac{1}{4}\Big(\|\overrightarrow{AB}\|\,\|\overrightarrow{AC}\| + \langle \overrightarrow{AB}, \overrightarrow{AC}\rangle\Big)\Big(\|\overrightarrow{AB}\|\,\|\overrightarrow{AC}\| - \langle \overrightarrow{AB}, \overrightarrow{AC}\rangle\Big).
\end{aligned}$$

Aus

$$\begin{aligned}
\|\overrightarrow{BC}\|^2 &= \|\overrightarrow{AB} - \overrightarrow{AC}\|^2 = \langle \overrightarrow{AB} - \overrightarrow{AC}, \overrightarrow{AB} - \overrightarrow{AC}\rangle \\
&= \|\overrightarrow{AB}\|^2 + \|\overrightarrow{AC}\|^2 - 2\langle \overrightarrow{AB}, \overrightarrow{AC}\rangle
\end{aligned}$$

folgt

$$\langle \overrightarrow{AB}, \overrightarrow{AC}\rangle = \frac{1}{2}\|\overrightarrow{AB}\|^2 + \frac{1}{2}\|\overrightarrow{AC}\|^2 - \frac{1}{2}\|\overrightarrow{BC}\|^2$$

und

$$\begin{aligned}
&\|\overrightarrow{AB}\|\,\|\overrightarrow{AC}\| \pm \langle \overrightarrow{AB}, \overrightarrow{AC}\rangle \\
&\qquad = \|\overrightarrow{AB}\|\,\|\overrightarrow{AC}\| \pm \frac{1}{2}\|\overrightarrow{AB}\|^2 \pm \frac{1}{2}\|\overrightarrow{AC}\|^2 \mp \frac{1}{2}\|\overrightarrow{BC}\|^2 \\
&\qquad = \pm\frac{1}{2}\Big(\|\overrightarrow{AB}\| \pm \|\overrightarrow{AC}\|\Big)^2 \mp \frac{1}{2}\|\overrightarrow{BC}\|^2 \\
&\qquad = \pm\frac{1}{2}(c \pm b)^2 \mp \frac{1}{2}a^2.
\end{aligned}$$

Damit gilt:

$$\begin{aligned}
F_\Delta^2 &= \frac{1}{16}\Big((c+b)^2 - a^2)(a^2 - (c-b)^2\Big) \\
&= \frac{1}{16}(c+b+a)(c+b-a)(a+c-b)(a-c+b) \\
&= \frac{1}{16}2s(2s-a-a)(2s-b-b)(2s-c-c) \\
&= s(s-a)(s-b)(s-c).
\end{aligned}$$

□

3.7 Teilverhältnisse, Doppelverhältnisse*

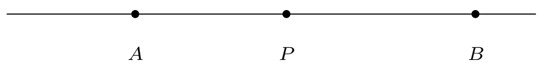
In diesem Abschnitt sei $(\mathcal{P}, V)$ ein zwei- bzw. dreidimensionaler affiner Raum.

Definition 3.27. (A, B) *sei ein Paar von Punkten in* $\mathcal{P}$ *und* P *sei ein von* A *und* B *verschiedener Punkt auf der durch* A *und* B *bestimmten Geraden. Die Zahl*

$$\lambda(P; A, B) = \frac{\lambda_{AP}}{\lambda_{BP}}$$

heißt das ***Teilverhältnis*** *des Punktes* P *in Bezug auf* A, B*. Hierbei sind* λ_{AP} *und* λ_{BP} *die eindeutig bestimmten Zahlen mit*

$$\overrightarrow{AP} = \lambda_{AP}\overrightarrow{AB}, \quad \overrightarrow{BP} = \lambda_{BP}\overrightarrow{AB}.$$



Ist $\overrightarrow{OX} = \overrightarrow{OA} + \lambda\overrightarrow{AB}$ eine Parameterdarstellung der Geraden durch A, B, so läßt sich diese auch in der Form

$$\overrightarrow{OX} = (1-\lambda)\overrightarrow{OA} + \lambda\overrightarrow{OB}$$

schreiben ($\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$). Gleichwertig damit ist

$$\overrightarrow{OX} = \alpha\overrightarrow{OA} + \beta\overrightarrow{OB}, \quad \alpha + \beta = 1,$$

d.h. insbesondere jeder Punkt der Geraden ist durch zwei reelle Zahlen α, β mit $\alpha + \beta = 1$ eindeutig bestimmt.

Für einen beliebigen Punkt P dieser Geraden gilt

$$\begin{aligned}\overrightarrow{AP} = \overrightarrow{OP} - \overrightarrow{OA} &= (\alpha - 1)\overrightarrow{OA} + \beta\overrightarrow{OB} \\ &= \beta(\overrightarrow{OB} - \overrightarrow{OA}) = \beta\overrightarrow{AB}\end{aligned}$$

und analog

$$\overrightarrow{BP} = -\alpha\overrightarrow{AB}.$$

Damit ist bewiesen:

Ist P *durch*

$$\overrightarrow{OP} = \alpha\overrightarrow{OA} + \beta\overrightarrow{OB} \text{ mit } \alpha + \beta = 1,\ \alpha \neq 0,$$

gegeben, so ist

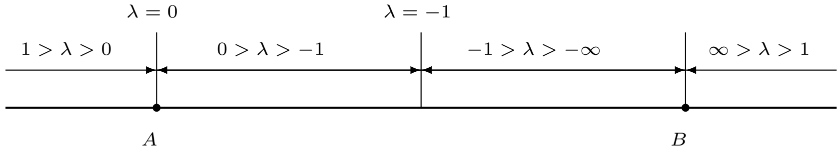
$$\lambda(P; A, B) = -\frac{\beta}{\alpha}.$$

Man sieht leicht, dass stets

$$\lambda(P; A, B) \neq 1$$

gilt ($\beta/\alpha = -1$ würde $\alpha = -\beta$ bedeuten, d.h. $\alpha + \beta = 0$).

Die folgende Zeichnung illustriert die Werte für $\lambda = \lambda(P; A, B)$ in Abhängigkeit von der Lage von P.



Durch Angabe von $\lambda = \lambda(P; A, B)$, $\lambda \neq 1$, ist P eindeutig bestimmt.

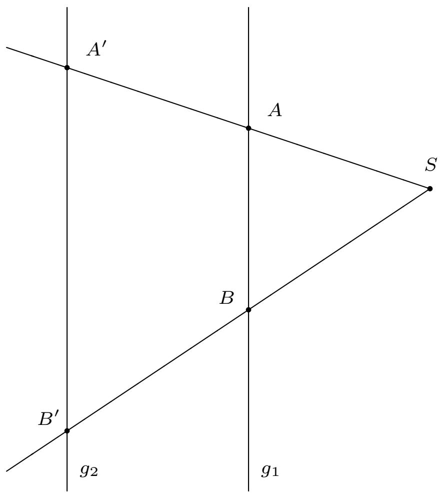
Beweis. Aus $-\beta/\alpha = \lambda$, $\alpha + \beta = 1$ folgt sofort $\alpha = 1/(1-\lambda)$ und $\beta = -\lambda/(1-\lambda)$. Der Ortsvektor von P ist dann durch

$$\overrightarrow{OP} = \alpha\overrightarrow{OA} + \beta\overrightarrow{OB} \tag{3.21}$$

gegeben. □

Satz 3.28 (Strahlensatz). *Es seien g_1, g_2 zwei verschiedene parallele Gerade in $\mathcal{P}$ und S ein Punkt in der durch g_1 und g_2 aufgespannten Ebene, der weder auf g_1 noch auf g_2 liegt. A, B seien verschiedene Punkte auf g_1 und A', B' seien die Projektionen von A und B auf g_2 mit dem Projektionszentrum S. Dann gilt:*

$$\lambda(S; B, B') = \lambda(S; A, A').$$



Beweis. Mit reellen Zahlen α, β gilt

$$\overrightarrow{OS} = (1-\alpha)\overrightarrow{OA} + \alpha\overrightarrow{OA'},$$
$$\overrightarrow{OS} = (1-\beta)\overrightarrow{OB} + \beta\overrightarrow{OB'}.$$

Wegen der Voraussetzung über S sind α und β von Null verschieden. Ferner ist

$$\lambda(S; A, A') = -\frac{\alpha}{1-\alpha}, \quad \lambda(S; B, B') = -\frac{\beta}{1-\beta}.$$

Wegen der Parallelität von g_1 und g_2 gilt

$$\overrightarrow{A'B'} = \lambda\overrightarrow{AB} = \lambda\overrightarrow{OB} - \lambda\overrightarrow{OA}, \quad \lambda \in \mathbb{R}.$$

Andererseits ist

$$\begin{aligned}
\overrightarrow{A'B'} &= \overrightarrow{OB'} - \overrightarrow{OA'} \\
&= \frac{1}{\beta}\overrightarrow{OS} - \frac{1-\beta}{\beta}\overrightarrow{OB} - \frac{1}{\alpha}\overrightarrow{OS} + \frac{1-\alpha}{\alpha}\overrightarrow{OS} \\
&= \Big(\frac{1}{\beta} - \frac{1}{\alpha}\Big)\overrightarrow{OS} + \frac{1-\alpha}{\alpha}\overrightarrow{OA} - \frac{1-\beta}{\beta}\overrightarrow{OB}.
\end{aligned}$$

Aus den zwei Darstellungen für $\overrightarrow{A'B'}$ folgt

$$\Big(\frac{1}{\beta}-\frac{1}{\alpha}\Big)\overrightarrow{OS}+\Big(\frac{1-\alpha}{\alpha}+\lambda\Big)\overrightarrow{OA}-\Big(\frac{1-\beta}{\beta}+\lambda\Big)\overrightarrow{OB}=o.$$

Da die Summe der Koeffizienten Null ist und die Punkte A, B, S nicht kollinear sind, folgt nach Satz 3.10, a),

$$\frac{1}{\beta}-\frac{1}{\alpha}=0,\quad \frac{1-\alpha}{\alpha}+\lambda=0,\quad \frac{1-\beta}{\beta}+\lambda=0$$

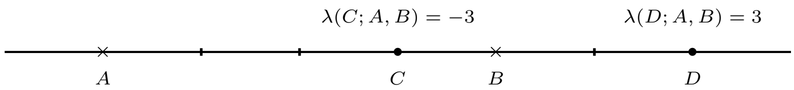
oder $\alpha=\beta$ (und $\lambda=-(1-\alpha)/\alpha$). □

Definition 3.29. *A, B, C, D seien vier paarweise verschiedene Punkte auf einer Geraden. Die Zahl*

$$\lambda(A,B;C,D)=\frac{\lambda(C;A,B)}{\lambda(D;A,B)}$$

*heißt das **Doppelverhältnis** dieser vier Punkte. Gilt $\lambda(A,B;C,D)=-1$, so heißt (A,B,C,D) ein **harmonisches Quadrupel**.*

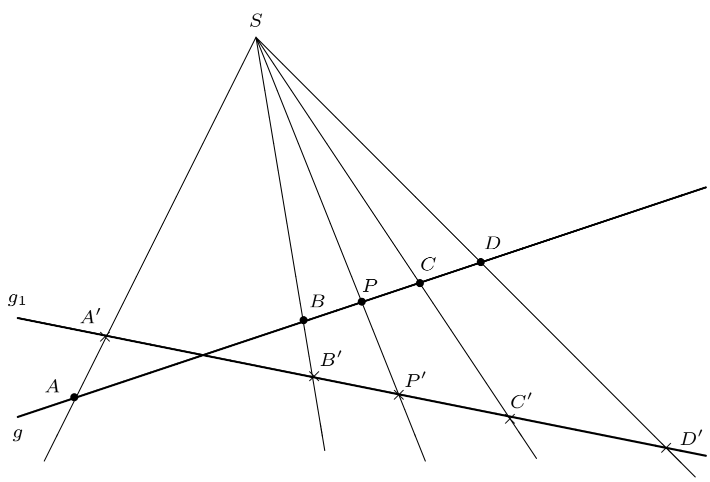
Für ein harmonisches Quadrupel liegt einer der Punkte C, D zwischen A und B, der andere außerhalb.



Satz 3.30. *A, B, C, D seien paarweise verschiedene Punkte auf einer Geraden g, S sei ein Punkt mit $S\notin g$. g_1 sei eine weitere Gerade mit $S\notin g_1$. A', B', C', D' bezeichne die Projektion der Punkte A, B, C, D auf die Gerade g_1 mit dem Projektionszentrum S. Dann gilt*

$$\lambda(A',B';C',D')=\lambda(A,B;C,D)$$

(*Invarianz des Doppelverhältnisses bei Projektion*).



Beweis. Wir untersuchen zunächst, wie sich das Teilverhältnis eines beliebigen Punktes P ($\neq B$) in Bezug auf A, B bei Projektion ändert.

Die Ortsvektoren der Punkte A', B' sind durch

$$\begin{aligned}\overrightarrow{OA'} &= (1-\alpha)\overrightarrow{OS} + \alpha\overrightarrow{OA},\\ \overrightarrow{OB'} &= (1-\beta)\overrightarrow{OS} + \beta\overrightarrow{OB}\end{aligned}$$

gegeben. Für den Punkt P gilt

$$\overrightarrow{OP} = (1-\mu)\overrightarrow{OA} + \mu\overrightarrow{OB}$$

und daher

$$\lambda(P; A, B) = -\frac{\mu}{1-\mu}.$$

Für P' gilt

$$\begin{aligned}\overrightarrow{OP'} &= (1-\gamma)\overrightarrow{OS} + \gamma\overrightarrow{OP}\\ &= (1-\gamma)\overrightarrow{OS} + \gamma(1-\mu)\overrightarrow{OA} + \gamma\mu\overrightarrow{OB}.\end{aligned} \tag{3.22}$$

P' liegt auf der Geraden durch A' und B', d.h.

$$\begin{aligned}\overrightarrow{OP'} &= (1-\lambda)\overrightarrow{OA'} + \lambda\overrightarrow{OB'}\\ &= \big((1-\lambda)(1-\alpha) + \lambda(1-\beta)\big)\overrightarrow{OS}\\ &\quad + (1-\lambda)\alpha\overrightarrow{OA} + \lambda\beta\overrightarrow{OB}.\end{aligned} \tag{3.23}$$

Die Zahlen λ, γ müssen noch bestimmt werden. Durch Gleichsetzen von (3.22) und (3.23) erhalten wir

$$(-\gamma + \alpha - \alpha\lambda + \beta\lambda)\overrightarrow{OS} + (\gamma - \alpha + \alpha\lambda - \mu\gamma)\overrightarrow{OA} + (\mu\gamma - \beta\lambda)\overrightarrow{OB} = o.$$

Die Summe der Koeffizienten ist gleich 0. Da die Punkte S, A, B nicht kollinear sind, folgt (siehe Satz 3.10, a)), dass die Koeffizienten sämtliche Null sein müssen. Dies ist gleichbedeutend, dass das folgende Gleichungssystem für γ und λ lösbar ist:

$$\begin{aligned}\gamma - \alpha + \alpha\lambda - \beta\lambda &= 0,\\ (1-\mu)\gamma + \alpha\lambda - \alpha &= 0,\\ \mu\gamma - \beta\lambda &= 0.\end{aligned}$$

Wegen $\mu \neq 1$ (beachte $P \neq B$) ist dieses System gleichwertig mit

$$\begin{aligned}\gamma + \frac{\alpha}{1-\mu}\lambda &= \frac{\alpha}{1-\mu},\\ \mu\gamma - \beta\lambda &= 0.\end{aligned} \tag{3.24}$$

Im Falle $\mu \neq 0$ subtrahieren wir das μ-fache der ersten Gleichung von der zweiten und erhalten das zu (3.24) äquivalente System

$$\begin{aligned}\gamma + \frac{\alpha}{1-\mu}\lambda &= \frac{\alpha}{1-\mu},\\ (\beta + \frac{\alpha\mu}{1-\mu})\lambda &= \frac{\alpha\mu}{1-\mu}.\end{aligned}$$

Da $\alpha \neq 0$ ist ($\alpha = 0$ würde $A' = S$ bedeuten, d.h. $S \in g_1$), muß $\beta + \frac{\alpha\mu}{1-\mu} \neq 0$ sein, damit das System lösbar ist. Als Lösung erhält man im Falle $\mu \neq 0$

$$\gamma = \frac{\alpha\beta}{\beta + \mu(\alpha - \beta)}, \quad \lambda = \frac{\mu\alpha}{\beta + \mu(\alpha - \beta)}.$$

Man erhält weiter $1 - \lambda = \frac{\beta(1-\mu)}{\beta+\mu(\alpha-\beta)}$ und

$$\lambda(P'; A', B') = -\frac{\lambda}{1-\lambda} = -\frac{\mu}{1-\mu}\frac{\alpha}{\beta} = \lambda(P; A, B)\frac{\alpha}{\beta}. \tag{3.25}$$

Ist $\mu = 0$ so hat (3.24) die Gestalt

$$\begin{aligned} \gamma + \alpha\lambda &= \alpha, \\ \beta\lambda &= 0. \end{aligned}$$

Wegen $\beta \neq 0$ (es ist $B' \neq S$ wegen $S \notin g_1$) erhält man $\lambda = 0$ (und $\gamma = \alpha$). Daher gilt in diesem Fall

$$\lambda(P'; A', B') = \lambda(P; A, B) = 0,$$

d.h. (3.25) gilt allgemein.

Die Aussage über das Doppelverhältnis folgt nun aus

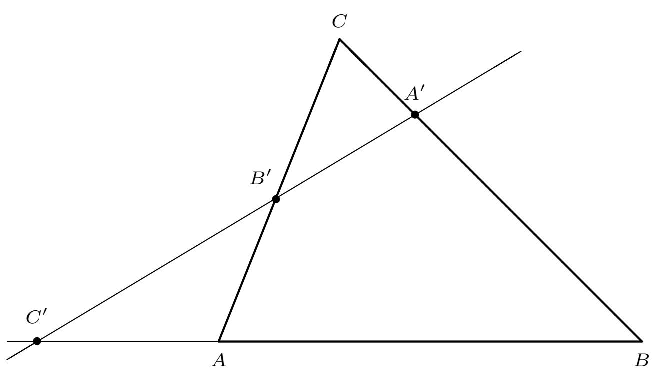
$$\begin{aligned} \lambda(A', B'; C', D') &= \frac{\lambda(C'; A', B')}{\lambda(D'; A', B')} = \frac{\lambda(C; A, B)\frac{\alpha}{\beta}}{\lambda(D; A, B)\frac{\alpha}{\beta}} \\ &= \lambda(A, B; C, D). \end{aligned}$$

□

Satz 3.31 (Menelaos[18]). *Durch die Punkte A, B, C sei ein Dreieck gegeben. Auf den durch A, B bzw. B, C bzw. C, A bestimmten Seiten des Dreieckes seien die Punkte C' bzw. A' bzw. B' gegeben, die sämtliche verschieden von A, B und C seien. Dann sind A', B', C' genau dann kollinear, wenn*

$$\lambda(C'; A, B)\lambda(A'; B, C)\lambda(B'; C, A) = 1$$

gilt.



[18]Menalaos von Alexandria, ca. 70 (Alexandria?) – ca. 130, machte am 14. 1. 98 astronomische Beobachtungen in Rom, von vielen Büchern die Menealos verfasste, ist nur “Sphaerica” überliefert, in dem sphärische Dreiecke behandelt werden (www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Menelaus.html).

Beweis. 1. Es seien A', B', C' kollinear. Mit geeigneten Konstanten $\alpha, \beta, \gamma, \mu$ gilt:

$$\begin{aligned}\overrightarrow{OA'} &= (1-\beta)\overrightarrow{OC} + \beta\overrightarrow{OB},\\ \overrightarrow{OB'} &= (1-\alpha)\overrightarrow{OC} + \alpha\overrightarrow{OA},\\ \overrightarrow{OC'} &= (1-\gamma)\overrightarrow{OA} + \gamma\overrightarrow{OB}\end{aligned} \tag{3.26}$$

und

$$\begin{aligned}\overrightarrow{OC'} &= (1-\mu)\overrightarrow{OA'} + \mu\overrightarrow{OB'}\\ &= (1-\beta+\mu\beta-\mu\alpha)\overrightarrow{OC} + (1-\mu)\beta\overrightarrow{OB} + \mu\alpha\overrightarrow{OA}.\end{aligned}$$

Man beachte, dass A', B', C' von den Punkten A, B, C verschieden sind. Daher gilt $\alpha, \beta, \gamma \neq 0$ und $\neq 1$.

Aus den Darstellungen für $\overrightarrow{OC'}$ erhalten wir

$$(1-\beta+\mu\beta-\mu\alpha)\overrightarrow{OC} + (\mu\alpha-1+\gamma)\overrightarrow{OA} + (\beta-\mu\beta-\gamma)\overrightarrow{OB} = o.$$

In dieser Beziehung ist die Summe der Koeffizienten Null. Die Punkte A, B, C sind nicht kollinear. Daher muß

$$1-\beta+\mu\beta-\mu\alpha = 0, \quad \mu\alpha-1+\gamma = 0, \quad \beta-\mu\beta-\gamma = 0$$

sein. Daraus folgt $\mu = \frac{1-\beta}{\alpha-\beta}$, $\gamma = \frac{\beta(\alpha-1)}{\alpha-\beta}$ und $1-\gamma = \frac{\alpha(1-\beta)}{\alpha-\beta}$. Es ist $\alpha \neq \beta$. Aus $\alpha = \beta$ würde, wegen (3.26), $\overrightarrow{A'B'} = -\alpha\overrightarrow{AB}$ folgen. Dann wäre die Gerade durch A', B' parallel zur Seite durch A, B und könnte diese nicht im Punkt C' schneiden. Daher gilt:

$$\begin{aligned}\lambda(C'; A, B)\lambda(A'; B, C)\lambda(B'; C, A) &= -\frac{\gamma}{1-\gamma} \cdot \frac{1-\beta}{\beta} \cdot \frac{\alpha}{1-\alpha}\\ &= -\frac{\beta(\alpha-1)}{\alpha(1-\beta)} \cdot \frac{1-\beta}{\beta} \cdot \frac{\alpha}{1-\alpha} = 1.\end{aligned}$$

2. Es sei für die durch (3.26) gegebenen Punkte

$$-\frac{\gamma}{1-\gamma} \cdot \frac{1-\beta}{\beta} \cdot \frac{\alpha}{1-\alpha} = 1.$$

Daraus folgt

$$\gamma = \frac{\beta(\alpha-1)}{\alpha-\beta}.$$

$\alpha = \beta$ ist nicht möglich, da sonst $\frac{\gamma}{1-\gamma} = -1$ sein müßte. Aus (3.26) erhalten wir damit

$$\overrightarrow{A'B'} = (\beta-\alpha)\overrightarrow{OC} + \alpha\overrightarrow{OA} - \beta\overrightarrow{OB}$$

und

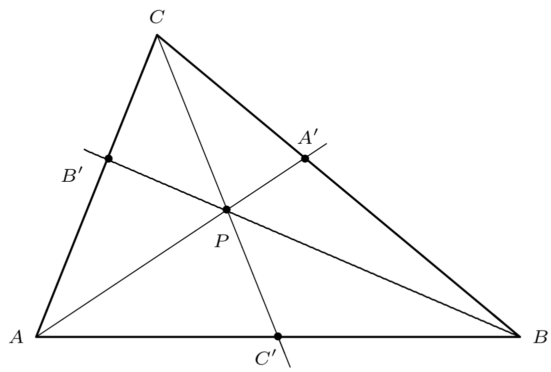
$$\begin{aligned}\overrightarrow{A'C'} = \overrightarrow{OC'} - \overrightarrow{OA'} &= -(1-\beta)\overrightarrow{OC} + (1-\gamma)\overrightarrow{OA} + (\gamma-\beta)\overrightarrow{OB}\\ &= -(1-\beta)\overrightarrow{OC} + \frac{\alpha(1-\beta)}{\alpha-\beta}\overrightarrow{OA} - \frac{\beta(1-\beta)}{\alpha-\beta}\overrightarrow{OB}\\ &= \frac{1-\beta}{\alpha-\beta}\overrightarrow{A'B'},\end{aligned}$$

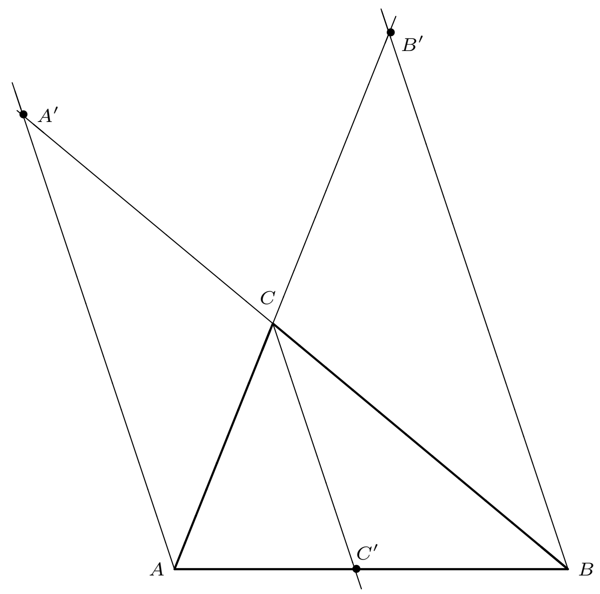
d.h. A', B', C' sind kollinear. □

Satz 3.32 (Ceva[19]). *Gegeben sei ein Dreieck mit den Eckpunkten A, B, C. Auf den Seiten seien die Punkte A', B', C' gegeben, die alle verschieden von A, B, C seien. Die Geraden durch A, A' bzw. B, B' bzw. C, C' schneiden sich dann und nur dann in einem Punkt P oder sind parallel, wenn*

$$\lambda(C'; A, B)\lambda(A'; B, C)\lambda(B'; C, A) = -1 \tag{3.27}$$

gilt.





Beweis. 1. Die Geraden durch A, A' bzw. B, B' bzw. C, C' schneiden sich im Punkt P. Durch A, B, C ist eine Ebene E bestimmt mit der Parameterdarstellung

$$\begin{aligned}\overrightarrow{OX} &= \overrightarrow{OA} + \lambda\overrightarrow{AB} + \mu\overrightarrow{AC} \\ &= (1-\lambda-\mu)\overrightarrow{OA} + \lambda\overrightarrow{OB} + \mu\overrightarrow{OC}.\end{aligned}$$

Dies ist gleichwertig mit

$$\overrightarrow{OX} = \alpha\overrightarrow{OA} + \beta\overrightarrow{OB} + \gamma\overrightarrow{OC}, \text{ wobei } \alpha+\beta+\gamma = 1. \tag{3.28}$$

[19] Ceva, Giovanni, 7. 12. 1647 (Mailand) – 15. 6. 1734 (Mantua), Hauptwerk im Bereich der Geometrie, seit 1686 an der Universität Mantua (www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Ceva_Giovanni.html).

Durch drei reelle Zahlen α, β, γ mit $\alpha + \beta + \gamma = 1$ ist vermöge (3.28) eindeutig ein Punkt X der Dreiecksebene bestimmt und umgekehrt. α, β, γ heißen auch die ***baryzentrischen Koordinaten*** von X.

Es seien $\alpha_0, \beta_0, \gamma_0$ die baryzentrischen Koordinaten von P, d.h. $\alpha_0 + \beta_0 + \gamma_0 = 1$ und

$$\overrightarrow{OP} = \alpha_0\overrightarrow{OA} + \beta_0\overrightarrow{OB} + \gamma_0\overrightarrow{OC}. \tag{3.29}$$

Es gilt $\alpha_0 \neq 0, 1$, $\beta_0 \neq 0, 1$ und $\gamma_0 \neq 0, 1$. Angenommen es sei $\alpha_0 = 1$. Dann gilt $\beta_0 + \gamma_0 = 0$ und somit $\overrightarrow{OP} = \overrightarrow{OA} + \beta_0(\overrightarrow{OB} - \overrightarrow{OC}) = \overrightarrow{OA} + \beta_0\overrightarrow{CB}$. Dies bedeutet, dass P auf der Geraden durch A und parallel zur Seite durch die Punkte B, C liegt. Die Gerade durch A und P kann daher die Seite durch B und C nicht im Punkt A' schneiden. Dieser Widerspruch beweist $\alpha_0 \neq 1$.

Wäre $\alpha_0 = 0$, so hätte man $\overrightarrow{OP} = \beta_0\overrightarrow{OB} + \gamma_0\overrightarrow{OC}$ mit $\beta_0 + \gamma_0 = 1$, d.h. P liegt auf der Seite durch B und C. Die Gerade durch B und P könnte daher die Seite durch A und C nur im Punkt C schneiden, d.h. es wäre $B' = C$ in Widerspruch zur Voraussetzung über die Punkte A', B' und C'.

Für die Punkte A', B', C' gilt

$$\begin{aligned} \overrightarrow{OA'} &= \lambda\overrightarrow{OB} + (1-\lambda)\overrightarrow{OC},\ \lambda \neq 0, 1, \\ \overrightarrow{OB'} &= \mu\overrightarrow{OC} + (1-\mu)\overrightarrow{OA},\ \mu \neq 0, 1, \\ \overrightarrow{OC'} &= \nu\overrightarrow{OA} + (1-\nu)\overrightarrow{OB},\ \nu \neq 0, 1. \end{aligned} \tag{3.30}$$

Andererseits liegen A', B' bzw. C' auf den Geraden durch A und P, B und P bzw. C und P. Daher ist

$$\begin{aligned} \overrightarrow{OA'} &= (1-\sigma)\overrightarrow{OA} + \sigma\overrightarrow{OP},\ \sigma \neq 1, \\ \overrightarrow{OB'} &= (1-\rho)\overrightarrow{OB} + \rho\overrightarrow{OP},\ \rho \neq 1, \\ \overrightarrow{OC'} &= (1-\tau)\overrightarrow{OC} + \tau\overrightarrow{OP},\ \tau \neq 1. \end{aligned} \tag{3.31}$$

Aus (3.30),(3.31) und (3.29) erhalten wir die Gleichungen

$$\begin{aligned} (1-\sigma(1-\alpha_0))\overrightarrow{OA} + (\sigma\beta_0 - \lambda)\overrightarrow{OB} + (\sigma\gamma_0 - 1 + \lambda)\overrightarrow{OC} &= o, \\ (\rho\alpha_0 - 1 + \mu)\overrightarrow{OA} + (1-\rho(1-\beta_0))\overrightarrow{OB} + (\rho\gamma_0 - \mu)\overrightarrow{OC} &= o, \\ (\tau\alpha_0 - \nu)\overrightarrow{OA} + (\tau\beta_0 - 1 + \nu)\overrightarrow{OB} + (1-\tau(1-\gamma_0))\overrightarrow{OC} &= o. \end{aligned}$$

Die Summe der Koeffizienten ist jeweils Null. Da die Punkte A, B, C nicht kollinear sind, müssen alle Koeffizienten Null sein. Das auf diese Weise entstehende Gleichungssystem von 9 Gleichungen für die Unbekannten λ, μ, ν, σ, ρ und τ ist daher eindeutig lösbar. Wir können uns daher auf die folgenden sechs Gleichungen beschränken:

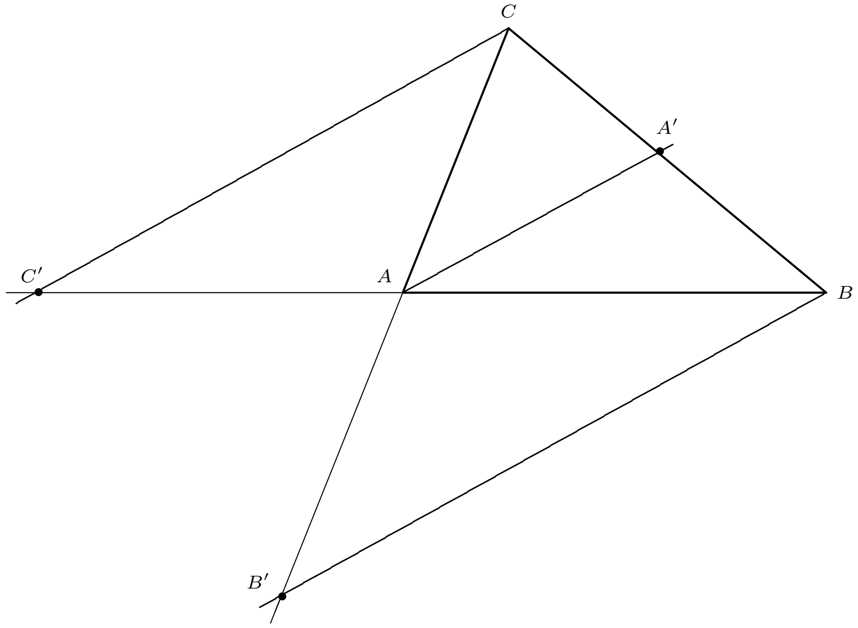
$$\begin{aligned} \sigma\beta_0 - \lambda &= 0, & \rho\gamma_0 - \mu &= 0, \\ 1-\sigma(1-\alpha_0) &= 0, & \tau\alpha_0 - \nu &= 0, \\ 1-\rho(1-\beta_0) &= 0, & 1-\tau(1-\gamma_0) &= 0. \end{aligned}$$

Die Lösung ist gegeben durch (wir interessieren uns nur für λ, μ, ν):

$$\lambda = \frac{\beta_0}{1-\alpha_0}, \qquad \mu = \frac{\gamma_0}{1-\beta_0}, \qquad \nu = \frac{\alpha_0}{1-\gamma_0}.$$

Daraus folgt $1-\lambda = \frac{1-\alpha_0-\beta_0}{1-\alpha_0} = \frac{\gamma_0}{1-\alpha_0}$, $1-\mu = \frac{\alpha_0}{1-\beta_0}$, $1-\nu = \frac{\beta_0}{1-\gamma_0}$ und $\lambda(A'; B, C) = -\frac{1-\lambda}{\lambda} = -\frac{\gamma_0}{\beta_0}$, $\lambda(B'; C, A) = -\frac{1-\mu}{\mu} = -\frac{\alpha_0}{\gamma_0}$, $\lambda(C'; A, B) = -\frac{1-\nu}{\nu} = -\frac{\beta_0}{\alpha_0}$. Dies ergibt (3.27).

2. Die Geraden durch A, A' bzw. B, B' bzw. C, C' sind parallel.



In diesem Fall gelten ebenso wie unter Punkt 1 die Darstellungen (3.30) für die Ortsvektoren $\overrightarrow{OA'}$, $\overrightarrow{OB'}$, $\overrightarrow{OC'}$. Aus der Parallelität der Geraden folgt auch

$$\overrightarrow{OB'} = \overrightarrow{OB} + \alpha\overrightarrow{AA'} = \overrightarrow{OB} + \alpha\overrightarrow{OA'} - \alpha\overrightarrow{OA},$$
$$\overrightarrow{OC'} = \overrightarrow{OC} + \beta\overrightarrow{AA'} = \overrightarrow{OC} + \beta\overrightarrow{OA'} - \beta\overrightarrow{OA}.$$

Offensichtlich muß (wegen der Voraussetzung über die Punkte A', B' und C') $\alpha \neq 0$ und $\beta \neq 0$ gelten. Daraus folgt mit Hilfe von (3.30)

$$(-\alpha - 1 + \mu)\overrightarrow{OA} + (1 + \alpha\lambda)\overrightarrow{OB} + (\alpha(1-\lambda) - \mu)\overrightarrow{OC} = o,$$
$$(-\beta - \nu)\overrightarrow{OA} + (\beta\lambda - 1 + \nu)\overrightarrow{OB} + (1 + \beta(1-\lambda))\overrightarrow{OC} = o.$$

Dies führt analog wie unter Teil 1 des Beweises zu den folgenden Gleichungen:

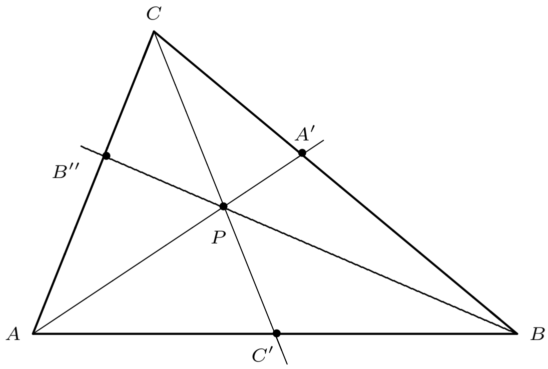
$$\alpha\lambda + 1 = 0, \quad \mu - \alpha - 1 = 0, \quad \beta + \nu = 0, \quad \beta\lambda - 1 + \nu = 0.$$

Daraus folgt

$$\lambda = -\frac{1}{\alpha}, \quad \mu = \alpha + 1, \quad \nu = \frac{\alpha}{\alpha + 1}$$

und $1 - \lambda = \frac{\alpha+1}{\alpha}$, $1 - \mu = -\alpha$, $1 - \nu = \frac{1}{\alpha+1}$. Wie in Teil 1 des Beweises gilt: $\lambda(A'; B, C) = -\frac{1-\lambda}{\lambda} = \alpha + 1$, $\lambda(B'; C, A) = -\frac{1-\mu}{\mu} = \frac{\alpha}{\alpha+1}$ und $\lambda(C'; A, B) = -\frac{1-\nu}{\nu} = -\frac{1}{\alpha}$, d.h. es gilt auch in diesem Fall (3.27).

3. Es sei (3.27) erfüllt. P sei der Schnittpunkt der Geraden durch die Punkte A, A' bzw. C, C'. B'' sei der Schnittpunkt der Geraden durch B, P mit der Seite durch A, C.



Dass B'' existiert zeigen wir in Teil 4 des Beweises. Nach Teil 1 des Beweises gilt

$$\lambda(A'; B, C)\lambda(B''; C, A)\lambda(C'; A, B) = -1.$$

Da auch $\lambda(A'; B, C)\lambda(B'; C, A)\lambda(C'; A, B) = -1$ gilt und die Teilverhältnisse jeweils $\neq 0$ sind (warum?), folgt sofort

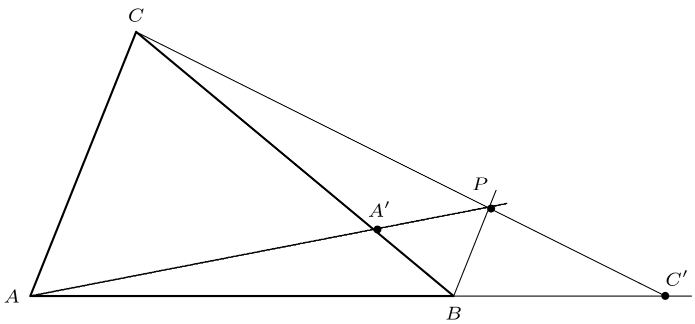
$$\lambda(B''; C, A) = \lambda(B'; C, A),$$

d.h. $B'' = B'$.

4. Der Punkt B'' existiert nur dann nicht, wenn die Gerade durch P, B parallel zur Seite durch A, C ist. In diesem Fall gilt wegen des Strahlensatzes (Satz 3.28)

$$\lambda(C'; A, B) = \lambda(C'; C, P). \tag{3.32}$$

Die Punkte P, C', C können als Projektionen der Punkte A', B, C mit dem Projektionszentrum A aufgefaßt werden.



Daher gilt (siehe (3.25))

$$\lambda(P; C', C) = \frac{\alpha}{\beta}\lambda(A'; B, C).$$

Die Konstanten α und β sind durch die Darstellungen

$$\overrightarrow{O\,C'} = (1-\alpha)\overrightarrow{O\,A} + \alpha\overrightarrow{O\,B},$$
$$\overrightarrow{O\,C} = (1-\beta)\overrightarrow{O\,A} + \beta\overrightarrow{O\,C}$$

bestimmt. Daraus folgt zunächst $\beta = 1$. Ist $\lambda_0 = \lambda(C'; A, B)$ so gilt $\overrightarrow{O\,C'} = \frac{1}{1-\lambda_0}\overrightarrow{O\,A} - \frac{\lambda_0}{1-\lambda_0}\overrightarrow{O\,B}$ (siehe (3.21)), d.h. $\alpha = -\frac{\lambda_0}{1-\lambda_0}$. Damit erhalten wir

$$\lambda(P; C', C) = -\frac{\lambda(C'; A, B)}{1 - \lambda(C'; A, B)}\lambda(A'; B, C). \tag{3.33}$$

Setzen wir $\mu_0 = \lambda(C'; C, P)$, so gilt

$$\overrightarrow{OC'} = \frac{1}{1-\mu_0}\overrightarrow{OC} - \frac{\mu_0}{1-\mu_0}\overrightarrow{OP}$$

und daher

$$\overrightarrow{OP} = -\frac{1-\mu_0}{\mu_0}\overrightarrow{OC'} + \frac{1}{\mu_0}\overrightarrow{OC}.$$

Daraus folgt

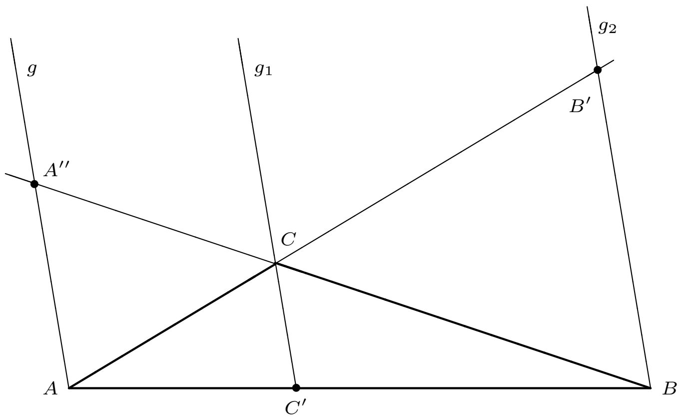
$$\lambda(P; C', C) = \frac{1}{1-\mu_0} = \frac{1}{1-\lambda(C'; C, P)} = \frac{1}{1-\lambda(C'; A, B)}.$$

Aus (3.33) folgt damit

$$\lambda(A'; B, C)\lambda(C'; A, B) = -1.$$

Wegen (3.27) müßte dann $\lambda(B'; C, A) = 1$ sein. Dieser Widerspruch zeigt, dass B'' existiert.

5. Es gelte nun (3.27) und die Geraden g_1 bzw. g_2 durch C, C' bzw. B, B' seien parallel. Wir legen durch A eine weitere zu g_1 und g_2 parallele Gerade g. g kann nicht zur Seite durch B, C parallel sein. Denn sonst müßte die Gerade g_2 mit der Seite durch B, C zusammenfallen, d.h. $B' = C$ sein, was ausgeschlossen ist. Die Gerade g schneidet somit die Seite durch B, C in einem Punkt A''.



Auf Grund von Teil 1 des Beweises gilt dann

$$\lambda(A''; B, C)\lambda(B'; C, A)\lambda(C'; A, B) = -1.$$

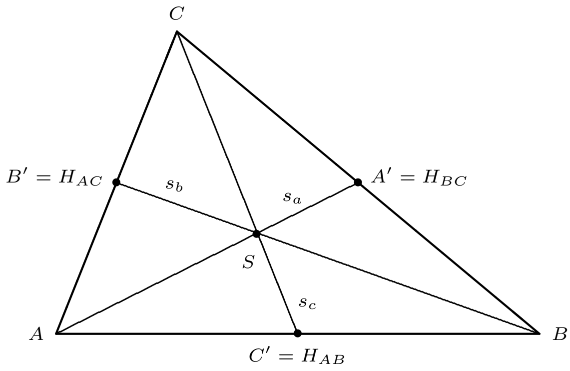
Wegen (3.27) folgt daraus

$$\lambda(A''; B, C) = \lambda(A'; B, C),$$

d.h. $A'' = A'$. Damit ist gezeigt, dass die Geraden durch A, A' bzw. B, B' bzw C, C' alle parallel zueinander sind. $\square$

Als Anwendungen des Satzes von Ceva beweisen wir nochmals die Sätze über den Schwerpunkt, den Höhenschnittpunkt bzw. Inkreismittelpunkt eines Dreieckes.

a) Schwerpunkt.
Hier sind A', B' und C' die Halbierungspunkte H_{BC}, H_{AC} und H_{AB}, d.h. $\lambda(A'; B, C) = \lambda(B'; C, A) = \lambda(C'; A, B) = -1$. Somit gilt (3.27) und die Schwerlinien schneiden sich in einem Punkt.



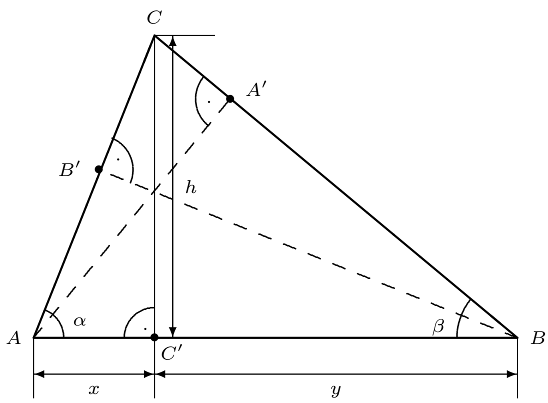
b) Höhenschnittpunkt.
Es ist $\tan\alpha = \frac{h}{x}$ und $\tan\beta = \frac{h}{y}$, woraus

$$\lambda(C'; A, B) = -\frac{\tan\beta}{\tan\alpha}$$

folgt. Analog zeigt man

$$\lambda(A'; B, C) = -\frac{\tan\gamma}{\tan\beta} \quad \text{und} \quad \lambda(B'; C, A) = -\frac{\tan\alpha}{\tan\gamma}.$$

Daraus folgt (3.27), d.h. die Höhen schneiden sich in einem Punkt.



c) Inkreismittelpunkt. Mit Hilfe des Sinussatzes der ebenen Trigonometrie erhält man (aus den Dreiecken A, B, A' bzw. A, A', C)

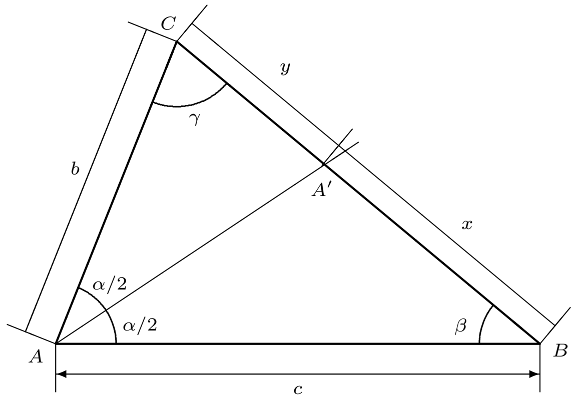
$$x = \frac{\sin\frac{\alpha}{2}}{\sin(\pi - \beta - \frac{\alpha}{2})} c = \frac{\sin\frac{\alpha}{2}}{\sin(\gamma + \frac{\alpha}{2})} c \quad (\text{wegen } \alpha + \beta + \gamma = \pi)$$

und

$$y = \frac{\sin\frac{\alpha}{2}}{\sin(\beta + \frac{\alpha}{2})} b.$$

Daraus folgt

$$\lambda(A'; B, C) = -\frac{x}{y} = -\frac{\sin(\beta + \frac{\alpha}{2})}{\sin(\gamma + \frac{\alpha}{2})} \frac{c}{b}.$$



In analoger Weise erhält man für die Schnittpunkte der anderen Winkelsymmetralen

$$\lambda(B';C,A) = -\frac{\sin(\gamma + \frac{\beta}{2})}{\sin(\alpha + \frac{\beta}{2})}\frac{a}{c},$$
$$\lambda(C';A,B) = -\frac{\sin(\alpha + \frac{\gamma}{2})}{\sin(\beta + \frac{\gamma}{2})}\frac{b}{a}.$$

Damit erhalten wir

$$\begin{aligned}&\lambda(A';B,C)\lambda(B';C,A)\lambda(C';A,B)\\ &\qquad = -\frac{\sin(\gamma + \frac{\beta}{2})\sin(\alpha + \frac{\gamma}{2})\sin(\beta + \frac{\alpha}{2})}{\sin(\alpha + \frac{\beta}{2})\sin(\beta + \frac{\gamma}{2})\sin(\gamma + \frac{\alpha}{2})} = -1,\end{aligned}$$

d.h. die drei Winkelsymmetralen schneiden sich in einem Punkt. Hier haben wir die Beziehungen

$$\begin{aligned}\sin\Big(\alpha + \frac{\beta}{2}\Big) &= \sin\Big(\pi - \beta - \gamma + \frac{\beta}{2}\Big) = \sin\Big(\pi - \Big(\gamma + \frac{\beta}{2}\Big)\Big) = \sin\Big(\gamma + \frac{\beta}{2}\Big),\\ \sin\Big(\beta + \frac{\gamma}{2}\Big) &= \sin\Big(\alpha + \frac{\gamma}{2}\Big),\\ \sin\Big(\gamma + \frac{\alpha}{2}\Big) &= \sin\Big(\beta + \frac{\alpha}{2}\Big)\end{aligned}$$

benutzt.

Literatur: P. Weiss: „*Lineare Algebra und Analytische Geometrie, Eine anwendungsbezogene Einführung*“, Trauner Verlag, Linz 1980.
J. Cigler: „*Einführung in die Lineare Algebra und Geometrie*“, Teil 1 und 2, Manz Verlag, Wien 1976/77.

Kapitel 4

Vektorräume

4.1 Gruppen und Körper

Eine ***innere Verknüpfung*** mit dem Verknüpfungszeichen „$*$" auf einer Menge M ist eine Abbildung $M \times M \to M$, wobei $a * b$ für das Bildelement von $(a, b) \in M \times M$ geschrieben wird. Meist nennen wir „$*$" selbst eine innere Verknüpfung auf M. $a*b$ heißt das Kompositum von a und b. Sehr oft wird als Verknüpfungszeichen „$+$" oder „$\cdot$" verwendet. In diesem Fall heißt $a + b$ bzw $a \cdot b$ die Summe bzw. das Produkt von a und b. Statt $a \cdot b$ schreibt man oft einfach ab.

Definition 4.1. *G sei eine nichtleere Menge mit innerer Verknüpfung $*$. $(G, *)$ heißt genau dann eine **Gruppe**, wenn gilt:*

(G1) *Für je drei Elemente a, b, c aus G gilt:*

$$a * (b * c) = (a * b) * c \quad (\textit{Assoziativgesetz}).$$

(G2) *Es existiert ein Element $e \in G$ mit*

$$e * a = a * e = a \tag{4.1}$$

für alle $a \in G$ (Existenz eines neutralen Elementes).

(G3) *Für jedes $a \in G$ existiert ein $a' \in G$ mit*

$$a * a' = a' * a = e \tag{4.2}$$

(Existenz des inversen Elementes).

Gilt zusätzlich

(G4) *Für je zwei Elemente $a, b \in G$ ist $a * b = b * a$ (Kommutativität),*

*so heißt $(G, *)$ eine **abelsche Gruppe**.*

Das Element e in (G2) ist eindeutig bestimmt. Wäre e' ebenfalls ein Element mit $e' * a = a * e' = a$ für alle $a \in G$, so folgt für $a = e$ sofort $e' * e = e$. Setzt man in (4.1) $a = e'$, so folgt $e' * e = e'$. Damit ist aber $e' = e$ gezeigt.

Auch das Element a' in (G3) ist eindeutig bestimmt. Wäre a'' ein Element mit $a * a'' = a'' * a = e$, so folgt

$$a' = a' * e = a' * (a * a'') = (a' * a) * a'' = e * a'' = a''.$$

Sehr oft wird für abelsche Gruppen das Verknüpfungszeichen „+“ verwendet. Das neutrale Element heißt dann Nullelement und man schreibt 0 statt e. Das zu a inverse Element wird mit $-a$ bezeichnet und heißt das zu a negative Element. Statt $a + (-b)$ schreibt man $a - b$. Ist die innere Verknüpfung durch „·“ bezeichnet, so heißt e das Einselement und an Stelle von a' schreibt man a^{-1}.

Beispiel 4.2. $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$ sind abelsche Gruppen; ebenso $(\mathbb{Q}\backslash\{0\}, \cdot)$ usw. $(\mathbb{N}, +)$ und $(\mathbb{R}, \cdot)$ sind beispielsweise keine Gruppen. Die Menge der Vektoren im $\mathbb{R}^2$ bzw. im $\mathbb{R}^3$ mit der Vektoraddition ist eine abelsche Gruppe. $\Diamond$

Beispiel 4.3. Ein Beispiel für eine im allgemeinen nicht abelsche Gruppe ist das folgende: M sei eine nicht-leere Menge, G sei die Menge der bijektiven Abbildungen $M \to M$. Als innere Verknüpfung definieren wir das Kompositum von Abbildungen, $f \circ g$ (siehe Definition 1.21). Neutrales Element ist die identische Abbildung von M und das zu f inverse Element ist die Umkehrabbildung f^{-1}. Als Übung beweise man, dass $(G, \circ)$ eine Gruppe ist. Ebenso beweise man, dass $(G, \circ)$ nicht abelsch ist, falls M mehr als drei Elemente hat. $\Diamond$

Es sei $(G, *)$ eine Gruppe. Dann sind für beliebige Elemente $a, b \in G$ die Gleichungen $a * x = b$ und $y * a = b$ eindeutig lösbar. Die Lösungen sind durch $x = a^{-1} * b$ bzw. $y = b * a^{-1}$ gegeben.

Beweis. Es sei $x \in G$ eine Lösung von $a * x = b$. Dann gilt $a^{-1} * (a * x) = a^{-1} * b$, woraus $x = a^{-1} * b$ folgt. Damit ist die Eindeutigkeit gezeigt. Man sieht sofort, dass $x = a^{-1} * b$ tatsächlich eine Lösung ist. Der Beweis für die Gleichung $y * a = b$ verläuft analog. □

Definition 4.4. *$\mathbb{K}$ sei eine Menge mit mindestens zwei Elementen. Auf $\mathbb{K}$ seien die inneren Verknüpfungen $+$ und $\cdot$ definiert. $(\mathbb{K}, +, \cdot)$ heißt genau dann ein* ***Körper****, wenn gilt:*

1. *$(\mathbb{K}, +)$ ist eine abelsche Gruppe.*
2. *Die Multiplikation „ · “ ist kommutativ und $(\mathbb{K}\backslash\{0\}, \cdot)$ ist eine Gruppe.*
3. *Für je drei Elemente $\alpha, \beta, \gamma \in \mathbb{K}$ gilt:*

$$\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma \quad (\textit{Distributivgesetz}).$$

In einem Körper gelten die üblichen Rechenregeln, wie wir sie von den reellen Zahlen etwa gewöhnt sind. Für das Element α^{-1} schreibt man häufig $\frac{1}{\alpha}$.
Als Übung beweise man:

Aus $\alpha\beta = 0$ folgt $\alpha = 0$ oder $\beta = 0$.

Es ist $\alpha \cdot 0 = 0$ für alle α.

$\alpha(-\beta) = (-\alpha)\beta = -(\alpha\beta)$.

$\alpha(\beta - \gamma) = \alpha\beta - \alpha\gamma$.

Wenn klar ist, welche Verknüpfungen als $+$ und $\cdot$ gewählt wurden, werden wir einfach $\mathbb{K}$ an Stelle von $(\mathbb{K}, +, \cdot)$ schreiben.

Beispiel 4.5. $\mathbb{Q}, \mathbb{R}$ und $\mathbb{C}$. $\Diamond$

Beispiel 4.6. $\mathbb{K} = \mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ mit der üblichen Addition und Multiplikation von reellen Zahlen. $\mathbb{K}$ ist ein Körper zwischen $\mathbb{Q}$ und $\mathbb{R}$. $\Diamond$

Beispiel 4.7. Es sei p eine Primzahl. Auf der Menge $\mathbb{Z}$ sei die folgende Äquivalenzrelation definiert:

$$n \equiv m(p) \Longleftrightarrow n - m \text{ ist durch } p \text{ teilbar.}$$

Sämtliche Äquivalenzklassen sind durch $\overline{0}, \overline{1}, \ldots, \overline{p-1}$ gegeben ($\overline{n} = \{n + kp \mid k \in \mathbb{Z}\}$). Wir setzen $\mathbb{Z}_p = \{\overline{0}, \ldots, \overline{p-1}\}$. Durch

$$\overline{n} + \overline{m} = \overline{n+m}, \quad \overline{n} \cdot \overline{m} = \overline{nm}$$

werden auf $\mathbb{Z}_p$ die inneren Verknüpfungen $+$ und $\cdot$ definiert.

$(\mathbb{Z}_p, +, \cdot)$ ist ein Körper bestehend aus p Elementen. Als Übung beweise man, dass $(\mathbb{Z}_p, +)$ eine abelsche Gruppe ist (beispielsweise $-\overline{n} = \overline{p-n}$) und dass die Multiplikation assoziativ und kommutativ ist. Ferner sieht man sofort, dass $\overline{1}$ das Einselement ist. Wir beweisen die Existenz des inversen Elementes für die Elemente $\overline{1}, \ldots, \overline{p-1}$:

Es seien n, m Elemente aus $\mathbb{Z}$ mit $1 \leq n,m \leq p-1$. Angenommen, es ist $\overline{n} \cdot \overline{m} = \overline{0}$. Dies bedeutet

$$nm = kp \text{ für ein } k \in \mathbb{Z},$$

d.h. p teilt nm. Da p Primzahl ist muß p ein Teiler von n oder m sein, was nicht möglich ist. Damit ist gezeigt:

$$\overline{n}, \overline{m} \in \mathbb{Z}_p \backslash \{\overline{0}\} \text{ impliziert } \overline{n} \cdot \overline{m} \neq \overline{0}$$

bzw.

$$\text{aus } \overline{n} \cdot \overline{m} = \overline{0} \text{ folgt } \overline{n} = \overline{0} \text{ oder } \overline{m} = \overline{0}.$$

Es sei nun $\overline{n} \in \mathbb{Z}_p$, $\overline{n} \neq \overline{0}$ gegeben. Gilt $\overline{n} \cdot \overline{m} = \overline{n} \cdot \overline{\ell}$ für $\overline{m}, \overline{\ell} \in \mathbb{Z}_p$, so folgt $\overline{m} = \overline{\ell}$ (denn aus $\overline{n} \cdot \overline{m} = \overline{n} \cdot \overline{\ell}$ folgt $\overline{n}(\overline{m} - \overline{\ell}) = \overline{0}$ und weiters $\overline{m} - \overline{\ell} = \overline{0}$ wegen $\overline{n} \neq \overline{0}$). Daher sind die Elemente

$$\overline{n} \cdot \overline{1}, \ \overline{n} \cdot \overline{2}, \ \ldots, \ \overline{n} \cdot \overline{p-1}$$

alle $\neq \overline{0}$ und paarweise verschieden. Dies bedeutet, dass für ein j, $1 \leq j \leq p-1$, $\overline{n} \cdot \overline{j} = \overline{1}$ gelten muß, d.h. $\overline{j} = (\overline{n})^{-1}$. $\Diamond$

Später werden wir noch den folgenden Begriff benötigen:

Definition 4.8. *$\mathbb{K}$ sei ein Körper, 1 das Einselement von $\mathbb{K}$ und $q \in \mathbb{N}$.*
*a) $\mathbb{K}$ hat die **Charakteristik** q genau dann, wenn gilt:*

(i) $\underbrace{1 + \cdots + 1}_{q\text{-mal}} = 0.$

(ii) $\underbrace{1 + \cdots + 1}_{k\text{-mal}} \neq 0$ *für alle* $k \in \mathbb{N}$ *mit* $1 \leq k < q$.

*b) $\mathbb{K}$ hat die **Charakteristik** 0, wenn $\underbrace{1 + \cdots + 1}_{k\text{-mal}} \neq 0$ ist für alle $k \in \mathbb{N}$.*

Die Körper $\mathbb{Q}, \mathbb{R}$ und $\mathbb{C}$ haben bei spielsweise die Charakteristik 0, die Körper $\mathbb{Z}_p$ die Charakteristik p.

4.2 Vektorraum, Unterraum, Basis

Definition 4.9. *V sei eine abelsche Gruppe mit additiv geschriebener Verknüpfung und $\mathbb{K}$ sei ein Körper. V heißt ein **Vektorraum** oder **linearer Raum** über $\mathbb{K}$ genau dann, wenn gilt:*

Es existiert eine Abbildung $\mathbb{K} \times V \to V$, welche jedem Paar (λ, a) mit $\lambda \in \mathbb{K}$ und $a \in V$ ein Element aus V zuordnet, welches als λa geschrieben wird (skalare Multiplikation) mit folgenden Eigenschaften:

(V1) $\lambda(a+b) = \lambda a + \lambda b$ *für alle* $\lambda \in \mathbb{K}$ *und* $a, b \in V$ *(1. Distributivgesetz).*

(V2) $(\lambda + \mu)a = \lambda a + \mu a$ *für alle* $\lambda, \mu \in \mathbb{K}$ *und* $a \in V$ *(2. Distributivgesetz).*

(V3) $\lambda(\mu a) = (\lambda\mu)a$ *für alle* $\lambda, \mu \in \mathbb{K}$ *und* $a \in V$ *(Assoziativgesetz).*

(V4) $1a = a$ *für alle* $a \in V$.

*Die Elemente aus V heißen **Vektoren**, jene aus $\mathbb{K}$ **Skalare**. V heißt reeller (komplexer) Vektorraum, wenn $\mathbb{K} = \mathbb{R}$ ($\mathbb{K} = \mathbb{C}$) gilt.*

Beispiel 4.10. $\mathbb{K}$ sei ein Körper. Wir definieren $V := \{(\lambda_1, \dots, \lambda_n) | \lambda_i \in \mathbb{K}\}$ und die Verknüpfungen

$$(\lambda_1, \dots, \lambda_n) + (\mu_1, \dots, \mu_n) := (\lambda_1 + \mu_1, \dots, \lambda_n + \mu_n),$$
$$\lambda(\lambda_1, \dots, \lambda_n) := (\lambda\lambda_1, \dots, \lambda\lambda_n)$$

für $(\lambda_1, \dots, \lambda_n), (\mu_1, \dots, \mu_n) \in V$ und $\lambda \in \mathbb{K}$. V ist ein linearer Raum über $\mathbb{K}$ und heißt der ***arithmetische Vektorraum*** der Dimension n über $\mathbb{K}$, $V = \mathbb{K}^n$. $\Diamond$

Beispiel 4.11. Es sei $\mathbb{K} = \mathbb{R}$ und $V = \{f \mid f : [0,1] \to \mathbb{R}\}$. Die Definition von $f + g$ für $f, g \in V$ ist:

$$(f+g)(x) = f(x) + g(x) \text{ für alle } x \in [0,1].$$

Definition von λg für $f \in V$ und $\lambda \in \mathbb{R}$:

$$(\lambda f)(x) = \lambda f(x) \text{ für alle } x \in [0,1].$$

V ist ein reeller Vektorraum. $\Diamond$

Beispiel 4.12. $\mathbb{K}[x]$ sei die Menge aller Polynome in einer Unbestimmten mit Koeffizienten aus dem Körper $\mathbb{K}$. Jedes $p \in \mathbb{K}[x]$ hat die Gestalt $p = \alpha_n x^n + \cdots + \alpha_0$ mit $n \in \mathbb{N}_0$[20] und $\alpha_i \in \mathbb{K}$. Wir definieren $p + q$ für $p = \alpha_n x^n + \cdots + \alpha_0$ und $q = \beta_m x^m + \cdots + \beta_0$ wie folgt: Für $k = \max(n, m)$ ist $p = \alpha_k x^k + \cdots + \alpha_0$ und $q = \beta_k x^k + \cdots + \beta_0$. Dabei ist $\alpha_j = 0$, $n < j \leq k$ und $\beta_j = 0$, $m < j \leq k$. Dann ist

$$p + q := (\alpha_k + \beta_k)x_k + \cdots + (\alpha_0 + \beta_0).$$

Für $\lambda \in \mathbb{K}$ definieren wir $\lambda p := (\lambda\alpha_n)x^n + \cdots + \lambda\alpha_0$. $\mathbb{K}[x]$ ist ein linearer Raum über $\mathbb{K}$. $\Diamond$

Beispiel 4.13. Die Vektoren im $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ bilden einen reellen Vektorraum (vgl. Abschnitt 2.1). $\Diamond$

Die folgenden Aussagen sind einfache Konsequenzen aus (V1) - (V4) und werden im folgenden, ohne dies immer zu erwähnen, verwendet (o ist das Nullelement in V):

(V4′) $\lambda a = o$ *gilt genau dann, wenn* $\lambda = 0$ *oder* $a = o$

und

$$(-\lambda)a = \lambda(-a) = -(\lambda a).$$

[20] $\mathbb{N}_0 = \{0, 1, 2, \dots\}$.

Beweis. a) Für $\lambda = 0$ gilt $0a = (0+0)a = 0a + 0a$, woraus $0a = o$ folgt (Addition von $-(0a)$ beiderseits). Analog beweist man $\lambda o = o$. Der Beweis für die umgekehrte Implikation ist derselbe wie für (2.8).

b) Die zweite Aussage folgt aus $\lambda a + (-\lambda)a = (\lambda - \lambda)a = 0 \cdot a = o$ und $\lambda a + \lambda(-a) = \lambda(a-a) = \lambda \cdot o = o$. □

In der Definition des Vektorraumes können wir (V4) durch (V4′) ersetzen, d.h. es gilt:

> *Es sei V eine abelsche Gruppe und $\mathbb{K}$ ein Körper. Ferner sei eine skalare Multiplikation $\mathbb{K} \times V \to V$ mit den Eigenschaften* (V1) – (V3) *gegeben. Dann gilt:*
>
> $$\text{(V4)} \quad \Longleftrightarrow \quad \text{(V4}'\text{)}.$$

Beweis. Die Implikation (V4) $\Longrightarrow$ (V4′) wurde bereits bewiesen. Es gelte nun (V4′). Für beliebiges $a \in V$ folgt mit Hilfe von (V1) – (V3):

$$1 \cdot (1 \cdot a - a) = 1 \cdot (1 \cdot a) - 1 \cdot a = (1 \cdot 1)a - 1 \cdot a = 1 \cdot a - 1 \cdot a = o.$$

Wegen (V4′) und $1 \neq 0$ folgt daraus

$$1 \cdot a - a = o,$$

d.h. $1 \cdot a = a$. □

Ist E eine Ebene im $\mathbb{R}^3$, so ist der Vektorraum V_E eine Teilmenge des Vektorraumes V_3 aller Vektoren im $\mathbb{R}^3$, wobei die algebraischen Operationen in V_E dieselben sind wie in V_3. Dies führt zu

Definition 4.14. *V sei ein linearer Raum über $\mathbb{K}$ und U sei eine nicht-leere Teilmenge von V. U heißt genau dann* (*linearer*) ***Unterraum*** *von V, wenn U bezüglich der in V gegebenen Addition und skalaren Multiplikation ein linearer Raum über $\mathbb{K}$ ist.*

Es ist nicht nötig, für U die Eigenschaften (V1) – (V4) nachzuweisen, wie der folgende Satz zeigt:

Satz 4.15. *Eine nichtleere Teilmenge U von V ist genau dann Unterraum von V, wenn gilt:*

(i) *Für alle $a, b \in U$ ist auch $a + b \in U$.*

(ii) *Für alle $\lambda \in \mathbb{K}$ und alle $a \in U$ ist auch $\lambda a \in U$.*

Beweis. 1. Falls U ein Unterraum von V ist, muß die Addition auf V auch eine innere Verknüpfung auf U sein, d.h. es muß (i) gelten. Ferner muß die skalare Multiplikation auf V auch auf U die skalare Multipliaktion ergeben, d.h. es muß (ii) gelten.

2. Es seien nun (i) und (ii) erfüllt. Die Verknüpfung „+“ ist auf U assoziativ und kommutativ, weil dies in V gilt. Da $U \neq \emptyset$ ist, existiert mindestens ein Element $a \in U$. Zu $a \in U$ enthält U auch $(-1)a = -(1a) = -a$ und daher auch $a + (-a) = o$. $(U, +)$ ist somit abelsche Gruppe. (V1) - (V4) gelten in U, weil sie in V gelten. Damit ist gezeigt, dass U ein Vektorraum über $\mathbb{K}$ ist.

Beispiel 4.16. V sei ein linearer Raum über $\mathbb{K}$. $U = \{o\}$ ist stets Unterraum von V, der sog. ***Nullraum***. ◊

Beispiel 4.17. g sei eine Gerade im $\mathbb{R}^3$, E eine Ebene im $\mathbb{R}^3$. V_g und V_E sind Unterräume von V_3 (= Vektorraum der Vektoren im $\mathbb{R}^3$). ◊

Beispiel 4.18. Es sei $V = \mathbb{K}^n$ und $U = \{(\lambda_1, \dots, \lambda_n) | \lambda_i \in \mathbb{K}, \lambda_1 = 0\}$. U ist Unterraum von $\mathbb{K}^n$. $\Diamond$

Beispiel 4.19. Es sei $V = \mathbb{R}^n$ und $U = \mathbb{Q}^n$. U ist kein Unterraum von V, weil z.B. $\sqrt{2}(1, 0, \dots, 0) = (\sqrt{2}, 0, \dots, 0) \notin \mathbb{Q}^n$ gilt. Es ist aber $U \subseteq V$. U ist natürlich ein linearer Raum über $\mathbb{Q}$. $\Diamond$

Beispiel 4.20. Es sei $V = \mathbb{K}[x]$. Unterräume sind:
$U_1 =$ Menge aller Polynome aus $\mathbb{K}[x]$ mit $\alpha_{2k} = 0$.
$U_2 =$ Menge aller Polynome aus $\mathbb{K}[x]$ mit $\alpha_{2k+1} = 0$.
$U_3 = \mathbb{K}_n[x] =$ Menge aller Polynome aus $\mathbb{K}[x]$ mit $\alpha_k = 0$ für $k > n$. $\Diamond$

Beispiel 4.21. V sei der reelle Vektorraum aller Abbildungen $[0, 1] \to \mathbb{R}$, $U_1 =$ Menge aller Abbildungen $f : [0, 1] \to \mathbb{R}$ mit $f(\frac{1}{2}) = 0$ und $U_2 =$ Menge aller Abbildungen aus V, welche auf $[0, 1]$ differenzierbar sind. Dann sind U_1, U_2 Unterräume von V. $\Diamond$

Definition 4.22. *V sei ein linearer Raum über $\mathbb{K}$ und b, $a_1, \dots, a_n$, $n \geq 1$, seien Vektoren aus V. b heißt genau dann eine **Linearkombination** von $a_1, \dots, a_n$, wenn für Skalare $\lambda_1, \dots, \lambda_n$ aus $\mathbb{K}$ gilt:*

$$b = \lambda_1 a_1 + \cdots + \lambda_n a_n.$$

Definition 4.23. *V sei ein Vektorraum über $\mathbb{K}$ und M eine nicht-leere Teilmenge von V. Die Menge*

$$[M] = \big\{a \in V \mid a \text{ ist Linearkombination endlich vieler Elemente aus } M\big\}$$

*heißt die **lineare Hülle** von M. Für eine endliche Menge $M = \{a_1, \dots, a_n\}$ schreiben wir $[a_1, \dots, a_n]$ an Stelle von $[M]$.*

Satz 4.24. *Unter den Voraussetzungen von Definition 4.23 gilt:*
a) *$[M]$ ist ein Unterraum von V.*
b) *Es sei U ein Unterraum von V. Dann ist $U = [M]$ genau dann, wenn gilt:*

(i) *$M \subset U$.*

(ii) *Ist U' Unterraum von V mit $M \subset U'$, so muß $U \subset U'$ sein.*

$[M]$ ist somit der (im Sinne der Inklusion) kleinste Unterraum von V, welcher M enthält.

Beweis. a) Sind a und b jeweils Linearkombinationen endlich vieler Elemente aus M, so auch $a + b$ und λa, $\lambda \in \mathbb{K}$. Das Ergebnis folgt dann aus Satz 4.15.
b) Es sei zunächst $U = [M]$. $M \subset [M]$ ist trivial. Ist U' ein Unterraum mit $M \subset U'$, so muß mit $a_1, \dots, a_n \in M$ auch $a = \alpha_1 a_1 + \cdots + \alpha_n a_n, \alpha_i \in \mathbb{K}$, in U' enthalten sein. Damit ist $[M] \subset U'$ gezeigt.

Es seien umgekehrt (i) und (ii) erfüllt. Dann folgt aus (i) wie oben $[M] \subset U$. Andererseits ist $[M]$ ein Unterraum von V mit $M \subset [M]$. Also gilt nach (ii) $U \subset [M]$, d.h. insgesamt $U = [M]$. $\square$

Satz 4.25. *V sei ein Vektorraum über $\mathbb{K}$. $\mathcal{S}$ sei ein Menge von Unterräumen von V. Dann gilt:*

$$D = \bigcap_{U \in \mathcal{S}} U \quad \text{ist ebenfalls ein Unterraum von } V.$$

*D heiß t der **Durchschnittsraum** der Unterräume $U \in \mathcal{S}$.*

Beweis. a) Aus $a, b \in D$ folgt $a \in U$ und $b \in U$ für alle $U \in \mathcal{S}$. Daraus folgt weiter $a + b \in U$ für alle $U \in \mathcal{S}$ und somit $a + b \in D$.
b) Aus $\lambda \in \mathbb{K}$, $b \in D$ folgt $\lambda b \in U$ für alle $U \in \mathcal{S}$ und daher auch $\lambda b \in D$. Nach Satz 4.15 ist D ein Unterraum von V. □

Korollar 4.26. *Unter den Voraussetzungen von Definition* 4.23 *gilt:*

$$[M] = \bigcap_{U \in \mathcal{S}} U, \text{ wobei } \mathcal{S} = \{U \mid U \text{ ist ein Unterraum von } V \text{ mit } M \subset U\}.$$

Definition 4.27. *V sei ein Vektorraum über $\mathbb{K}$ und $\mathcal{S}$ ein Menge von Unterräumen von V. Der Unterraum*

$$S = \sum_{U \in \mathcal{S}} U := [\bigcup_{U \in \mathcal{S}} U]$$

*heißt die **Summe** der Unterräume aus $\mathcal{S}$. Für $\mathcal{S} = \{U_1, \ldots, U_n\}$ schreiben wir $S = U_1 + \cdots + U_n$.*

Satz 4.28. *Der Summenraum $S = \sum_{U \in \mathcal{S}} U$ ist gegeben durch:*

$$S = \Big\{x \in V \mid x = \sum_{U \in \mathcal{S}} x_U, \text{ wobei } x_U \in U \text{ für alle } U \in \mathcal{S} \text{ und } x_U \neq o \text{ für höchstens endlich viele } U \in \mathcal{S}\Big\}.$$

Beweis. a) Jeder Vektor der Form $x = \sum_{U \in \mathcal{S}} x_U$ mit $x_U \in U$ für alle $U \in \mathcal{S}$ und $x_U \neq o$ für höchstens endlich viele $U \in \mathcal{S}$ ist eine Linearkombination von endlich vielen Vektoren aus $\bigcup_{U \in \mathcal{S}} U$ und daher in S enthalten.
b) Aus $a \in \sum_{U \in \mathcal{S}} U$ folgt $a = \lambda_1 u_1 + \cdots + \lambda_n u_n$ mit $\lambda_i \in \mathbb{K}$ und $u_i \in \bigcup_{U \in \mathcal{S}} U$. Faßt man in dieser Darstellung Vektoren zusammen, die in demselben Unterraum $U \in \mathcal{S}$ liegen, erhält man eine Darstellung von a, wie sie im Satz angegeben ist. □

Definition 4.29. *V sei ein Vektorraum über $\mathbb{K}$. S, U_1 und U_2 seien Unterräume von V. Der Unterraum S heißt genau dann **direkte Summe** von U_1 und U_2, $S = U_1 \oplus U_2$, wenn gilt:*

1. $S = U_1 + U_2$.
2. $U_1 \cap U_2 = \{o\}$.

Satz 4.30. *Unter den Vorausetzungen von Definition* 4.29 *gilt $S = U_1 \oplus U_2$ genau dann, wenn jedes Element $a \in S$ eine eindeutige Darstellung*

$$a = a_1 + a_2$$

mit $a_i \in U_i$, $i = 1, 2$, besitzt.

Beweis. a) Es sei $S = U_1 \oplus U_2$. Nach Satz 4.28 besitzt jedes $a \in S$ eine Darstellung der verlangten Form. Es sei nun $a = a_1 + a_2 = b_1 + b_2$ mit $a_i, b_i \in U_i$, $i = 1, 2$. Dann gilt $o = a_1 - b_1 + a_2 - b_2$ bzw. $a_1 - b_1 = b_2 - a_2 \in U_2$. Daraus folgt $a_1 - b_1 \in U_1 \cap U_2 = \{o\}$, d.h. $a_1 = b_1$ und weiter $a_2 = b_2$.
b) Wir beweisen die umgekehrte Implikation. Wegen Satz 4.28 folgt zunächst $S = U_1 + U_2$. Es sei nun $c \in U_1 \cap U_2$ und $a \in S$. Dann gilt mit $a_i \in U_i$, $i = 1, 2$,

$$a = a_1 + a_2 = (a_1 + c) + (a_2 - c).$$

Wegen $c \in U_1 \cap U_2$ ist $a_1 + c \in U_1$ und $a_2 - c \in U_2$. Die Darstellung von a muß aber eindeutig sein, d.h. $a_1 = a_1 + c$ und $a_2 = a_2 - c$. Daraus folgt $c = o$. □

Im Falle von mehr als zwei Unterräumen muß die Definition der direkten Summe etwas anders formuliert werden:

Definition 4.31. *V sei ein Vektorraum über $\mathbb{K}$ und $\mathcal{S}$ sei eine Menge von Unterräumen von V. Ein Unterraum S von V heißt genau dann* ***direkte Summe*** *von $\mathcal{S}$, $S = \bigoplus_{U \in \mathcal{S}} U$, wenn gilt:*

1. $S = \sum_{U \in \mathcal{S}} U$.
2. $U_0 \cap \left(\sum_{U \in \mathcal{S} \setminus \{U_0\}} U \right) = \{o\}$ *für alle* $U_0 \in \mathcal{S}$.

Es gilt der zu Satz 4.30 analoge Satz:

Satz 4.32. *Unter den Voraussetzungen von Definition 4.31 gilt: Es ist genau dann $S = \bigoplus_{U \in \mathcal{S}} U$, wenn für jedes Element $a \in S$ eindeutig bestimmte Elemente $a_U \in U$, $U \in \mathcal{S}$, existieren mit*

(i) $a_U \neq o$ *für höchstens endlich viele* $U \in \mathcal{S}$,

(ii) $a = \sum_{U \in \mathcal{S}} a_U$.[21]

Beweis. a) Es sei $S = \bigoplus_{U \in \mathcal{S}} U$. Nach Satz 4.28 existieren paarweise verschiedene $U_1, \ldots, U_r \in \mathcal{S}$ und $a_{U_i} \in U_i$, $i = 1, \ldots, r$, mit $a = a_{U_1} + \cdots + a_{U_r}$. Für $U \in \mathcal{S}$, $U \neq U_i$, $i = 1, \ldots, r$, setzen wir $a_U = o$. Dann ist $a = \sum_{U \in \mathcal{S}} a_U$ eine Darstellung von a mit den Eigenschaften (i) und (ii). Es sei $a = \sum_{U \in \mathcal{S}} b_U$ eine zweite derartige Darstellung von a. Dann gilt $\sum_{U \in \mathcal{S}} (a_U - b_U) = o$ (beachte, dass $a_U - b_U \neq o$ für höchstens endlich viele $U \in \mathcal{S}$ gilt). Für beliebiges $U_0 \in \mathcal{S}$ ist somit

$$a_{U_0} - b_{U_0} = - \sum_{U \in \mathcal{S} \setminus \{U_0\}} (a_U - b_U) \in \sum_{U \in \mathcal{S} \setminus \{U_0\}} U.$$

Wegen Punkt 2 aus Definition 4.31 ist $a_{U_0} = b_{U_0}$. Da $U_0 \in \mathcal{S}$ beliebig war, ist $a_U = b_U$ für alle $U \in \mathcal{S}$.

b) Es gelte (i) und (ii). Dann ist $S = \sum_{U \in \mathcal{S}} U$ wegen Satz 4.28. Jedes $a \in S$ besitzt eine eindeutige Darstellung $a = \sum_{U \in \mathcal{S}} a_U$ mit $a_U \in U$, $U \in \mathcal{S}$, und $a_U \neq o$ für höchstens endlich viele $U \in \mathcal{S}$.

Es sei nun $U_0 \in \mathcal{S}$ gewählt. Für $c \in U_0 \cap \sum_{U \in \mathcal{S} \setminus \{U_0\}} U$ ist $-c \in U_0$ und es existieren Elemente $c_U \in U$ für $U \in \mathcal{S} \setminus \{U_0\}$ mit $c = \sum_{U \in \mathcal{S} \setminus \{U_0\}} c_U$, wobei $c_U \neq o$ für höchstens endlich viele $U \in \mathcal{S} \setminus \{U_0\}$ gilt (Satz 4.28). Dann ist

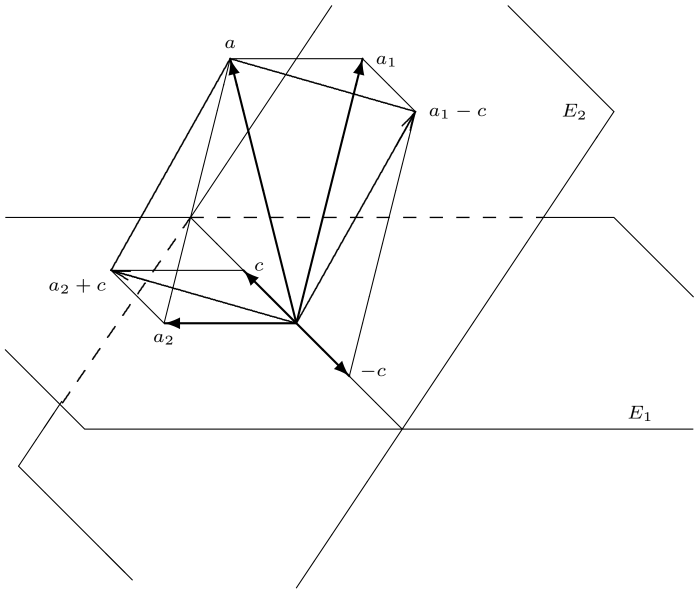
$$a = a_{U_0} + \sum_{U \in \mathcal{S} \setminus \{U_0\}} a_U = (a_{U_0} - c) + \sum_{U \in \mathcal{S} \setminus \{U_0\}} (a_U + c_U),$$

d.h. man hat zwei Darstellungen von a mit den Eigenschaften (i) und (ii). Wegen der Eindeutigkeit dieser Darstellungen muß $c = o$ gelten (und $c_U = o$ für $U \in \mathcal{S} \setminus \{U_0\}$). Damit ist $U_0 \cap \big(\sum_{U \in \mathcal{S} \setminus \{U_0\}} U = \{o\}$ bewiesen. □

Beispiel 4.33. V, U_1 und U_2 seien wie in Beispiel 4.20 gegeben. Es gilt $\mathbb{K}[x] = U_1 \oplus U_2$. ◊

[21] Beachte, dass die Summe wegen (i) endlich ist.

Beispiel 4.34. Es sei $V = V_3$ (= Vektoren im $\mathbb{R}^3$), E_1, E_2 seien Ebenen im $\mathbb{R}^3$ und $U_1 = V_{E_1}$, $U_2 = V_{E_2}$.
a) Es sei $E_1 \nparallel E_2$. Dann ist $U_1 \cap U_2 = V_g$, g die Schnittgerade von E_1 und E_2, und $U_1 + U_2 = V_3$. Die Summe ist jedoch nicht direkt. Jeder Vektor $a \in V_3$ hat die Darstellung $a = a_1 + a_2$ mit $a_1 \in U_1$ und $a_2 \in U_2$. Die Vektoren a_1 und a_2 sind jedoch nicht eindeutig.



b) Es sei $E_1 \| E_2$. Dann gilt $U_1 = U_2$ und damit $U_1 + U_2 = U_1 \cap U_2$. $\lozenge$

Beispiel 4.35. Es sei $V = V_2$ (Vektoren im $\mathbb{R}^2$). g_1, g_2 seien Gerade im $\mathbb{R}^2$, $g_1 \nparallel g_2$, $U_1 = V_{g_1}$, $U_2 = V_{g_2}$. Dann gilt $U_1 \oplus U_2 = V_2$. $\lozenge$

Definition 4.36. *V sei Vektorraum über $\mathbb{K}$.*
a) *$a_1, \dots, a_n$, $n \geq 1$, aus V heißen genau dann* ***linear unabhängig****, wenn aus $\lambda_1 a_1 + \cdots + \lambda_n a_n = o$ mit $\lambda_i \in \mathbb{K}$ stets*

$$\lambda_1 = \cdots = \lambda_n = 0$$

folgt.
b) *$a_1, \dots, a_n$ aus V heißen genau dann* ***linear abhängig****, wenn sie nicht linear unabhängig sind, d.h. wenn es ein n-Tupel $(\lambda_1, \dots, \lambda_n) \neq (0, \dots, 0)$ gibt mit $\lambda_1 a_1 + \cdots + \lambda_n a_n = o$.*
c) *Eine Teilmenge M von V heißt linear unabhängig, wenn je endlich viele, paarweise verschiedene Vektoren aus M linear unabhängig sind. Ist dies nicht der Fall, heißt M linear abhängig.*

Lemma 4.37. a) *Ein einziges Element $a \in V$ ist genau dann linear unabhängig, wenn $a \neq o$ ist.*
b) *$a_1, \dots, a_n \in V$, $n \geq 2$, sind genau dann linear abhängig, wenn für mindestens ein k, $1 \leq k \leq n$, gilt:*

$$a_k = \sum_{\substack{i=1 \\ i \neq k}}^{n} \lambda_i a_i \quad \textit{mit} \quad \lambda_i \in \mathbb{K}.$$

Beweis als Übung.

Definition 4.38. *V sei Vektorraum über $\mathbb{K}$, $V \neq \{o\}$ und B Teilmenge von V. B heißt genau dann **Basis** von V, wenn gilt:*

1. *B ist linear unabhängig.*
2. *$[B] = V$.*

Beispiel 4.39. *Die leere Menge $\emptyset$ ist eine linear unabhängige Teilmenge eines jeden Vektorraumes V.*

Beweis. Wir müssen zeigen, dass folgende Implikation wahr ist: Sind $a_1, \dots, a_n$ paarweise verschiendene Elemente aus $\emptyset$, so sind $a_1, \dots, a_n$ linear unabhängig. Diese Implikation ist wahr, da das Vorderglied stets falsch ist. □ ◊

$V = \{o\}$ hätte keine Basis im Sinne von Definition 4.38. Die einzigen Teilmengen sind $B = \{o\}$ und $B = \emptyset$. $B = \{o\}$ ist wegen Lemma 4.37, a) linear abhängig, also sicher keine Basis von V. Für $B = \emptyset$ gilt $[B] = \emptyset$. Daher ist auch $\emptyset$ keine Basis von V. Es ist aber sinnvoll für den Nullraum die leere Menge als Basis zu definieren.

Beispiel 4.40. Es sei $V = \mathbb{K}^3$, $a_1 = (1, 1, 0)$, $a_2 = (1, 0, 1)$, $a_3 = (0, 1, 1)$, 1 das Einselement von $\mathbb{K}$.

a) Die Charakteristik von $\mathbb{K}$ sei $\neq 2$ (etwa $\mathbb{K} = \mathbb{R}$, $\mathbb{C}$ oder $\mathbb{Q}$). Aus $\lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3 = o = (0, 0, 0)$ folgt $\lambda_1 + \lambda_2 = 0$, $\lambda_1 + \lambda_3 = 0$, $\lambda_2 + \lambda_3 = 0$. Dies impliziert $\lambda_2 = -\lambda_3$, $\lambda_1 = \lambda_3$ und $\lambda_1 = -\lambda_3$, woraus $\lambda_1 + \lambda_1 = (1 + 1)\lambda = 0$ folgt. Wegen $1 + 1 \neq 0$ ergibt dies $\lambda_1 = 0$ und $\lambda_2 = \lambda_3 = 0$, d.h. a_1, a_2, a_3 sind linear unabhängig.

Es sei $a = (\alpha_1, \alpha_2, \alpha_3)$ beliebig aus $\mathbb{K}^3$. Die Gleichung $a = \lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3$ ist äquivalent mit $\lambda_1 + \lambda_2 = \alpha_1, \lambda_1 + \lambda_3 = \alpha_2$, $\lambda_2 + \lambda_3 = \alpha_3$ bzw. mit $\lambda_1 = \frac{1}{2}(\alpha_1 + \alpha_2 - \alpha_3)$, $\lambda_2 = \frac{1}{2}(\alpha_1 + \alpha_3 - \alpha_2)$, $\lambda_3 = \frac{1}{2}(\alpha_2 + \alpha_3 - \alpha_1)$. Daraus folgt, dass a_1, a_2, a_3 eine Basis von $\mathbb{K}^3$ bilden.

b) Die Charakteristik von $\mathbb{K}$ sei 2. Dann gilt $a_1 + a_2 + a_3 = (1 + 1, 1 + 1, 1 + 1) = (0, 0, 0)$, d.h. a_1, a_2, a_3 sind linear abhängig. ◊

Beispiel 4.41. Es sei $V = \mathbb{K}^n$. Die Elemente

$$e_i = (0, \dots, 0, \underset{\substack{| \\ i\text{-te Stelle}}}{1}, 0, \dots, 0), \quad i = 1, \dots, n,$$

bilden eine Basis von $\mathbb{K}^n$, die sog. ***kanonische Basis***. ◊

Beispiel 4.42. Es sei $V = \mathbb{K}[x]$. Dann ist $B = \{x^n \mid n = 0, 1, 2, \dots\}$ eine Basis. ◊

Satz 4.43. *V sei Vektorraum über $\mathbb{K}$, $V \neq \{o\}$ und B Teilmenge von V. Folgende Aussagen sind gleichwertig:*

a) *B ist Basis von V (d.h. B ist linear unabhängig mit $[B] = V$).*

b) *$[B] = V$ und für beliebige Teilmengen C von V folgt aus $C \subsetneqq B$ stets $[C] \subsetneqq V$.*

c) *B ist linear unabhängig und aus $B \subsetneqq C \subset V$ folgt, dass C linear abhängig ist.*

d) *$[B] = V$ und die Darstellung der Elemente aus V als Linearkombination endlich vieler Elemente aus B ist eindeutig.*

Beweis. Wir führen den Beweis nach dem Schema a) ⇒ d) ⇒ b) ⇒ c) ⇒ a).

1. „a) $\Rightarrow$ d)“
$[B] = V$ ist klar. $a \in V$ besitze zwei Darstellungen als Linearkombination von je endlich vielen Elementen aus B:

$$a = \alpha_{i_1} b_{i_1} + \cdots + \alpha_{i_n} b_{i_n},\ \alpha_{i_k} \in \mathbb{K},\ b_{i_k} \in B,$$
$$a = \alpha_{j_1} b_{j_1} + \cdots + \alpha_{j_m} b_{j_m},\ \alpha_{j_\ell} \in \mathbb{K},\ b_{j_\ell} \in B.$$

Beide Linearkombinationen können als Linearkombination der Elemente $b_1, \ldots, b_q$ aufgefaßt werden (mit $\{b_1, \ldots, b_q\} = \{b_{i_1}, \ldots, b_{i_n}\} \cup \{b_{j_1}, \ldots, b_{j_m}\}$):

$$a = \alpha_1 b_1 + \cdots + \alpha_q b_q,$$
$$a = \beta_1 b_1 + \cdots + \beta_q b_q.$$

Aus diesen Darstellungen erhält man sofort

$$o = (\alpha_1 - \beta_1) b_1 + \cdots + (\alpha_q - \beta_q) b_q,$$

woraus wegen der linearen Unabhängigkeit von B (und damit auch der Elemente $b_1, \ldots, b_q$) sofort $\alpha_1 = \beta_1, \ldots, \alpha_q = \beta_q$ folgt.
2. „d) $\Rightarrow$ b)“
$[B] = V$ ist wieder klar. Angenommen C sei eine echte Teilmenge von B mit $[C] = V$. Wir wählen $b \in B$ mit $b \notin C$. Wegen $[C] = V$ gibt es Elemente $c_1, \ldots, c_k \in C$ und $\alpha_i \in \mathbb{K}$ mit

$$b = \alpha_1 c_1 + \cdots + \alpha_k c_k.$$

Andererseits ist $b = 1 \cdot b$. Man hätte damit zwei verschiedene Darstellungen von b als Linearkombination von Elementen aus B, in Widerspruch zu d). Es muß also $[C] \subsetneqq V$ gelten.
3. „b) $\Rightarrow$ c)“
Wir beweisen zunächst die lineare Unabhängigkeit von B. Angenommen B sei linear abhängig. Es gibt dann endlich viele linear abhängige Elemente $b_1, \ldots, b_n$ aus B.
Fall 1: $n = 1$. In diesem Fall muß $b_1 = o$ sein (Lemma 4.37, a)). Setzen wir $C = B \backslash \{b_1\}$, so gilt $[C] = [B] = V$. (Wegen $\alpha b_1 = o$ für alle $\alpha \in \mathbb{K}$, kann in jeder Linearkombination von Elementen aus B, in der b_1 vorkommt, b_1 gestrichen werden.) Da C echte Teilmenge von B ist, kann anderseits nicht $[C] = V$ sein (wegen b)). Fall 1 ist also nicht möglich.
Fall 2: $n \geq 2$. Nach Lemma 4.37, b), muß (bei geeigneter Numerierung)

$$b_1 = \alpha_2 b_2 + \cdots + \alpha_n b_n,\ \alpha_j \in \mathbb{K},$$

sein. Wir bilden wieder $C = B \backslash \{b_1\}$. Kommt in einer Linearkombination von Elementen aus B das Element b_1 vor, so kann man es durch die obige Darstellung ersetzen und erhält eine Linearkombination von Elementen aus C. Das heißt, auch in diesem Fall würde $[C] = V$ folgen in Widerspruch zu b).

B ist somit linear unabhängig. Es sei nun C echte Obermenge von B (natürlich $C \subset V$). Wir wählen $c \in C$ mit $c \notin B$. Wegen $[B] = V$ gibt es Vektoren $b_j \in B$ und Skalare $\alpha_j \in \mathbb{K}$, $j = 1, \ldots, n$, mit

$$c = \alpha_1 b_1 + \cdots + \alpha_n b_n.$$

Die Vektoren $c, b_1, \ldots, b_n$ und damit auch C sind nach Lemma 4.37, b) linear abhängig.
4. „c) $\Rightarrow$ a)“
Wegen c) ist B linear unabhängig. Wegen $B \subset [B]$ müssen wir nur $c \in V$ mit $c \notin B$ betrachten. Die Menge $C = B \cup \{c\}$ ist wegen c) linear abhängig, d.h. es gibt endlich viele Vektoren aus C, die linear abhängig sind. Unter diesen Vektoren muß c vorkommen (sonst hätte man endlich viele Vektoren aus B, die linear abhängig sind, d.h. B wäre linear abhängig); es seien c, $b_1, \ldots, b_n$ diese Vektoren. Für Skalare $(\alpha_0, \ldots, \alpha_n) \neq (0, \ldots, 0)$ gilt

$$\alpha_0 c + \alpha_1 b_1 + \cdots + \alpha_n b_n = o.$$

Es ist $\alpha_0 \neq 0$, da sonst die Vektoren $b_1, \dots, b_n$ und damit auch B linear abhängig wären. Dann folgt aber sofort

$$c = \Big(-\frac{\alpha_1}{\alpha_0}\Big)b_1 + \cdots + \Big(-\frac{\alpha_n}{\alpha_0}\Big)b_n \in [B].$$

Damit ist auch $[B] = V$ gezeigt. $\square$

Die Eigenschaft c) trifft im Falle $V = \{o\}$ für die leere Menge zu. Denn $\emptyset$ ist linear unabhängig (siehe Beispiel 4.39) und $C = \{o\}$, die einzige Obermenge von $\emptyset$ in V, ist linear abhängig. Die leere Menge als Basis des Nullraumes erfüllt allerdings die Eigenschaften a), b) und d) aus Satz 4.43 nicht.

Satz 4.44. *Es sei C eine linear unabhängige Teilmenge des Vektorraumes V. Dann gibt es mindestens eine Basis B von V mit $C \subset B$. Insbesondere besitzt jeder Vektorraum V eine Basis (man wähle etwa $C = \emptyset$).*

Der Beweis dieses Satzes erfordert für den allgemeinen Fall transfinite Methoden der Mengenlehre, die im ersten Semester im allgemeinen noch nicht zur Verfügung stehen. Wir geben jedoch hier den Beweis trotzdem, um die Möglichkeit zu geben, später den Beweis nachzulesen.

Wir benötigen folgende Begriffe. Es sei M eine Menge und $\mathcal{S} \subset \mathcal{P}(M)$. Die Inklusion definiert auf $\mathcal{S}$ eine Ordnung (siehe Definition 1.12). Eine Teilmenge $\mathcal{K}$ von $\mathcal{S}$ heißt eine ***Kette***, wenn für $K_1, K_2 \in \mathcal{K}$ stets $K_1 \subset K_2$ oder $K_2 \subset K_1$ gilt (man sagt auch, dass $\mathcal{K}$ bezüglich „$\subset$“ linear geordnet ist). Eine Menge $S \in \mathcal{S}$ heißt ***obere Schranke*** der Kette $\mathcal{K}$ in $\mathcal{S}$, wenn $K \subset S$ für alle $K \in \mathcal{K}$ gilt. Falls beispielsweise $S = \bigcup_{K \in \mathcal{K}} K \in \mathcal{S}$ gilt, ist S eine obere Schranke von $\mathcal{K}$. Diese einfache Situation liegt jedoch nicht immer vor.

Beispiel 4.45. Es sei $M = \mathbb{R}$ und $\mathcal{S}$ sei die Menge aller abgeschlossenen Intervalle in $\mathbb{R}$. Für $\mathcal{K} = \{[\frac{1}{n}, 1] \mid n = 1, 2, \dots\}$ gilt $\bigcup_{K \in \mathcal{K}} K = (0, 1] \notin \mathcal{S}$. $\mathcal{K}$ besitzt natürlich obere Schranken, etwa $S = [0, 1]$. $\Diamond$

Eine Menge $T \in \mathcal{S}$ heißt maximal in $\mathcal{S}$, wenn aus $S \supset T$, $S \in \mathcal{S}$, stets $S = T$ folgt (d.h. es gibt in $\mathcal{S}$ keine Mengen, die „größer“ als S sind). Es können in $\mathcal{S}$ mehrere maximale Elemente vorkommen oder auch keine.

Beispiel 4.46. a) Es sei $M = \{1, 2, 3, 4\}$ und $\mathcal{S}$ die Menge aller Teilmengen von M mit einer ungeraden Anzahl von Elementen. Dann sind $\{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 4\}$ und $\{2, 3, 4\}$ maximale Elemente in $\mathcal{S}$.

b) Es sei $M = \mathbb{N}$ und $\mathcal{S}$ die Menge aller endlichen Teilmengen von $\mathbb{N}$. $\mathcal{S}$ besitzt keine maximalen Elemente. $\Diamond$

Über die Existenz maximaler Elemente gibt das folgende Lemma Auskunft:

Zornsches[22] Lemma. *M sei eine Menge und $\mathcal{S} \subset \mathcal{P}(M)$. Besitzt jede Kette von $\mathcal{S}$ eine obere Schranke in $\mathcal{S}$, so besitzt $\mathcal{S}$ mindestens ein maximales Element.*

Für einen Beweis bzw. eine Diskussion der Bedeutung dieser Aussage für die Mengenlehre verweisen wir auf die Lehrbücher der Mengenlehre (vgl. Kapitel 1).

Beweis von Satz 4.44. Wir wählen $M = V$ und $\mathcal{S} = \{A \subset V \mid C \subset A$ und A linear unabhängig$\}$. $\mathcal{S}$ ist nicht leer, denn es ist $C \in \mathcal{S}$. $\mathcal{K}$ sei eine Kette in $\mathcal{S}$. Wir definieren $S = \bigcup_{K \in \mathcal{K}} K$. Es gilt trivialerweise $S \supset K$ für alle $K \in \mathcal{K}$. Wir müssen noch $S \in \mathcal{S}$ zeigen.

Zunächst ist $C \subset S$ weil $C \subset K$ für alle $K \in \mathcal{K}$. Es seien $b_1, \dots, b_n$ paarweise verschiedene Elemente aus S. Jedes b_i kommt in mindestens einer Menge $K_i \in \mathcal{K}$ vor, $i = 1, \dots, n$. Bei geeigneter Numerierung gilt $K_1 \subset K_2 \subset \cdots \subset K_n$, da $\mathcal{K}$ eine Kette ist. Dann gilt

[22] Zorn, Max August, 6. 6. 1906 (Krefeld) – 9. 3. 1993 (Bloomington, Indiana), Arbeiten zur Theorie unendlicher Mengen, Topologie und Algebra (www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Zorn.html).

$b_1, \dots, b_n \in K_n$. Die Menge K_n ist linear unabhängig und somit auch $b_1, \dots, b_n$. Damit ist gezeigt, dass S auch linear unabhängig ist, also $S \in \mathcal{S}$ gilt und somit S obere Schranke von $\mathcal{K}$ ist. Nach dem Zornschen Lemma besitzt $\mathcal{S}$ maximale Elemente. B sei ein solches. Für B gilt: $C \subset B$ und B ist linear unabhängig. Ist $C' \supsetneqq B$, so gilt $C' \notin \mathcal{S}$, weil B maximal ist. Ferner gilt $C \subset C'$. $C' \notin \mathcal{S}$ bedeutet daher, dass C' linear abhängig (weil $C \subset C'$). Nach Satz 4.43, c) ist B eine Basis von V. □

Lemma 4.47. *$a_1, \dots, a_m$ seien linear unabhängige Elemente des Vektorraumes V. Für ein Element $b \in V$ gelte $b = \beta_1 a_1 + \dots + \beta_m a_m$, $\beta_i \in \mathbb{K}$. Ist $\beta_k \neq 0$, so sind die Elemente $a_1, \dots, a_{k-1}, b, a_{k+1}, \dots, a_m$ ebenfalls linear unabhängig. Ferner gilt $[a_1, \dots, a_m] = [a_1, \dots, a_{k-1}, b, a_{k+1}, \dots, a_m]$.*

Beweis. Wir können o.B.d.A. $k = 1$ annehmen. Es sei $a \in [a_1, \dots, a_m]$, d.h. $a = \lambda_1 a_1 + \dots + \lambda_m a_m$, $\lambda_i \in \mathbb{K}$. Aus $\beta_1 \neq 0$ folgt $a_1 = \beta_1^{-1}(b - \beta_2 a_2 - \dots - \beta_m a_m)$ und weiter $a = \beta_1^{-1}\lambda_1 b + (\lambda_2 - \lambda_1\beta_1^{-1}\beta_2)a_2 + \dots + (\lambda_n - \lambda_1\beta_1^{-1}\beta_m)a_m$, d.h. $a \in [b, a_2, \dots, a_m]$.

Ist umgekehrt $a \in [b, a_2, \dots, a_m]$, d.h. $a = \lambda_1 b + \lambda_2 a_2 + \dots + \lambda_m a_m$, so folgt unmittelbar

$$a = \lambda_1\beta_1 a_1 + (\lambda_1\beta_2 + \lambda_2)a_2 + \dots + (\lambda_1\beta_m + \lambda_m)a_m,$$

d.h. es ist $a \in [a_1, \dots, a_m]$.

Um zu zeigen, dass $b, a_2, \dots, a_m$ linear unabhängig sind, nehmen wir an, es sei $\lambda b + \lambda_2 a_2 + \dots + \lambda_m a_m = o$. Daraus folgt $\lambda\beta_1 a_1 + (\lambda\beta_2 + \lambda_2)a_2 + \dots + (\lambda\beta_m + \lambda_m)a_m = o$ und $\lambda\beta_1 = \lambda\beta_2 + \lambda_2 = \dots = \lambda\beta_m + \lambda_m = 0$ (weil $a_1, \dots, a_m$ linear unabhängig sind). Wegen $\beta_1 \neq 0$ gilt $\lambda = 0$ und folglich auch $\lambda_2 = \dots = \lambda_n = 0$. □

Satz 4.48 (Austauschsatz von Steinitz[23]). *$\{a_1, \dots, a_n\}$, $n \geq 1$, sei eine Basis von V und $b_1, \dots, b_k \in V$ seien linear unabhängig. Dann gilt:*

1. *$k \leq n$.*
2. *Bei geeigneter Numerierung von $a_1, \dots, a_n$ ist $\{b_1, \dots, b_k, a_{k+1}, \dots, a_n\}$ ebenfalls eine Basis von V.*

Beweis. Der Beweis erfolgt mittels vollständiger Induktion nach k.

1. Es sei $k = 1$. In diesem Fall ist $k \leq n$ trivial. Da b linear unabhängig ist, muß nach Lemma 4.37, a), $b \neq o$ sein. Es gilt somit $b = \alpha_1 a_1 + \dots \alpha_n a_n$, wobei mindestens ein $\alpha_i \neq 0$ ist. Es sei etwa $\alpha_1 \neq 0$. Nach Lemma 4.47 kann a_1 durch b ersetzt werden.

2. Die Aussage sei für $k - 1$ richtig. Sind die Vektoren $b_1, \dots, b_k$ linear unabhängig, so auch $b_1, \dots, b_{k-1}$. Wegen der Induktionsvoraussetzung gilt $k-1 \leq n$ und $B^* = \{b_1, \dots, b_{k-1}, a_k, \dots, a_n\}$ ist eine Basis von V (bei geeigneter Numerierung der a_i). Angenommen es sei $k - 1 = n$. Dann ist $B^* = \{b_1, \dots, b_{k-1}\}$. Als Obermenge der Basis B^* ist $\{b_1, \dots, b_k\}$ linear abhängig (siehe Satz 4.43, c)). Dieser Widerspruch zur Voraussetzung beweist $k - 1 < n$, d.h. $k \leq n$. Da B^* Basis von V ist, gilt $b_k = \delta_1 b_1 + \dots + \delta_{k-1}b_{k-1} + \delta_k a_k + \dots + \delta_n a_n$. Die $\delta_k, \dots, \delta_n$ sind nicht alle 0, da sonst $b_1, \dots, b_k$ linear abhängig wären. Wir können $\delta_k \neq 0$ annehmen (bei geeigneter Numerierung der $a_k, \dots, a_n$). Nach Lemma 4.47 kann a_k in B^* durch b_k ersetzt werden, d.h. $\{b_1, \dots, b_k, a_{k+1}, \dots, a_n\}$ ist Basis von V. □

Korollar 4.49. *V sei Vektorraum über $\mathbb{K}$. Dann gilt:*

a) *Ist $\{a_1, \dots, a_n\}$ eine Basis von V, so besteht jede andere Basis von V ebenfalls aus n Elementen.*

[23] Steinitz, Ernst, 13. 6. 1871 (Laurahütte, Schlesien) – 29. 9. 1928 (Kiel), Beiträge zur Körpertheorie, "Algebraische Theorie der Körper" (1910) (www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Steinitz.html).

b) *Besitzt V eine Basis bestehend aus unendlich vielen Elementen, so enthalten auch alle anderen Basen unendlich viele Elemente.*

Beweis. 1. Es sei $\{b_1, \dots, b_m\}$ ebenfalls eine Basis von V. Die Vektoren $b_1, \dots, b_m$ sind linear unabhängig. Nach Satz 4.48 gilt $m \leq n$. Umgekehrt liefert Satz 4.48 für die Basis $\{b_1, \dots, b_m\}$ und die linear unabhängigen Elemente $a_1, \dots, a_n$ sofort $n \leq m$, d.h. es gilt insgesamt $m = n$.

Es sei nun $\{b_1, b_2, \dots\}$ eine Basis mit unendlich vielen Elementen. Dann sind die Elemente $b_1, \dots, b_{n+1}$ linear unabhängig. Nach Satz 4.48 müßte $n+1 \leq n$ gelten. Dieser Widerspruch zeigt, dass keine Basis von V mit unendlich vielen Elementen existiert.

2. V besitze eine Basis B mit unendlich vielen Elementen. Es sei $\{a_1, \dots, a_n\}$ ebenfalls Basis von V. Nach dem bereits bewiesenem Punkt a) müßte dann B ebenfalls n Elemente haben. □

Definition 4.50. *V sei Vektorraum über $\mathbb{K}$.*

a) *V besitzt die **Dimension** $n \in \mathbb{N}$, $\dim V = n$, genau dann, wenn V eine Basis aus n Elementen besitzt.*

b) *Es ist* $\dim V := \infty$ *genau dann, wenn V eine unendliche Basis besitzt.*

c) *Wir definieren* $\dim\{o\} = 0$.

Korollar 4.49 zeigt, dass diese Definition sinnvoll ist.

Beispiel 4.51. $\mathbb{K}^n$ besitzt die Dimension n (siehe Beispiel 4.41). ♢

Beispiel 4.52. V sei wie in Beispiel 4.11. $f_n \in V$, $n = 1, 2, \dots$, sei definiert durch $f_n(\frac{1}{n}) = 1$, $f_n(x) = 0$ für $x \neq \frac{1}{n}$, $x \in [0, 1]$. Dann ist $M = \{f_1, f_2, \dots\}$ eine linear unabhängige Teilmenge von V. Nach Satz 4.44 kann M zu einer Basis B von V erweitert werden, d.h. $\dim V = \infty$. ♢

Beispiel 4.53. Es seien $V = \mathbb{K}[x]$ und U_1, U_2, U_3 wie in Beispiel 4.20. Dann gilt:

$$\begin{aligned}
&\dim V = \dim U_1 = \dim U_2 = \infty \quad \text{(siehe Beispiel 4.42)},\\
&\dim U_3 = n + 1.
\end{aligned}$$

♢

Beispiel 4.54. Es ist $\dim V_2 = 2$, $\dim V_3 = 3$ (vgl. Satz 2.5). ♢

Beispiel 4.55. V_1 bezeichne $(\mathbb{R}, +)$ als linearen Raum über $\mathbb{K} = \mathbb{R}$ und V_2 bezeichne $(\mathbb{R}, +)$ als linearen Raum über $\mathbb{K} = \mathbb{Q}$.

Da $\{1\}$ eine Basis von V_1 ist, gilt $\dim V_1 = 1$. Angenommen es sei $\dim V_2 = n < \infty$. Dann gibt es $\mu_1, \dots, \mu_n \in \mathbb{R}$ mit $[\mu_1, \dots, \mu_n] = \{\alpha_1\mu_1 + \dots + \alpha_n\mu_n \mid \alpha_i \in \mathbb{Q}\} = V_2$. Ferner sind $\mu_1, \dots, \mu_n$ linear unabhängig in V_2. Den Elementen aus V_2 entsprechen somit ein-eindeutig die n-Tupel $(\alpha_1, \dots, \alpha_n) \in \mathbb{Q} \times \dots \times \mathbb{Q}$ (n-mal). Daraus folgt, dass V_2 und $\mathbb{Q} \times \dots \times \mathbb{Q}$ (n-mal) gleichmächtig sind. Da $\mathbb{Q}$ abzählbar unendlich ist, gilt dies gemäß Satz 1.28 für V_2. $\mathbb{R} = V_2$ ist nicht abzählbar. Dieser Widerspruch beweist $\dim V_2 = \infty$. ♢

Die Aussage a) von Korollar 4.49 kann auch wie folgt formuliert werden:

Bemerkung 4.56. V sei Vektorraum über $\mathbb{K}$ mit $\dim V = n < \infty$ und B sei eine linear unabhängige Teilmenge von V. Dann gilt:

B ist genau dann Basis von V, wenn B aus n Elementen besteht.

Beweis als Übung.

Weitere einfache Aussagen sind die folgenden:

Bemerkung 4.57. V sei Vektorraum über $\mathbb{K}$ und U Unterraum von V. Dann gilt:

a) $\dim U \leq \dim V$.

b) Es sei $\dim V < \infty$. Dann ist $\dim U = \dim V$ gleichbedeutend mit $U = V$.

Beweis. a) B^* sei Basis von U. Dann ist B^* linear unabhängig. Nach Satz 4.44 gibt es eine Basis $B \supset B^*$ von V. Daraus folgt $\dim V \geq \dim U$.

b) Übung. Man beachte Bemerkung 4.56. □

Satz 4.58. *V sei Vektoraum über $\mathbb{K}$ und U, W seien Unterräume von V. Dann gilt:*

$$\dim U + \dim W = \dim(U \cap W) + \dim(U + W),$$

(wobei man $\infty + \infty = \alpha + \infty = \infty$ für $\alpha \in \mathbb{N}$ setzt).

Beweis. 1. Der Fall $\dim U = \infty$ oder $\dim W = \infty$ ist trivial, weil dann auch $\dim(U+W) = \infty$ ist.

2. Es sei $\dim U < \infty$ und $\dim W < \infty$. Dann ist wegen Bemerkung 4.57 auch $\dim U \cap W < \infty$. $B_0 = \{c_1, \ldots, c_r\}$ sei eine Basis von $U \cap W$, falls $U \cap W \neq \{o\}$. Wir setzen $B_0 = \emptyset$, falls $\dim U \cap W = \{o\}$ ist. Nach Satz 4.48 gibt es Basen von U und von W der Form $B_0 \cup B_1$ bzw. $B_0 \cup B_2$, wobei $B_1 = \{a_1, \ldots, a_s\}$ und $B_2 = \{b_1, \ldots, b_t\}$. Es ist $B_1 = \emptyset$, falls $U \subset W$ (und damit $U \cap W = U$) ist, und $B_2 = \emptyset$, falls $W \subset U$ ist.

Als nächstes beweisen wir $B = B_0 \cup B_1 \cup B_2$ ist Basis von $U + W$.

a) Für $a \in U + W$ gilt wegen Satz 4.28 $a = u + w$ mit $u \in U$, $w \in W$. Daraus folgt $u \in [B_0 \cup B_1], w \in [B_0 \cup B_2]$ und somit $u + w \in [B]$. Damit ist $U + W \subset [B]$ gezeigt. Da andererseits jedes Element von B in $U + W$ enthalten ist, gilt $[B] \subset U + W$ (vgl. Satz 4.24). Somit gilt insgesamt $[B] = U + W$.

b) Es sei $\gamma_1 c_1 + \cdots + \gamma_r c_r + \alpha_1 a_1 + \cdots + \alpha_s a_s + \beta_1 b_1 + \cdots + \beta_t b_t = o$. Daraus folgt $u_1 := \gamma_1 c_1 + \cdots + \gamma_r c_r + \alpha_1 a_1 + \cdots + \alpha_s a_s = -\beta_1 b_1 - \cdots - \beta_t b_t =: w_1$ mit $u_1 \in U$ und $w_1 \in W$. Dies bedeutet $u_1 = w_1 \in U \cap W$. Wegen $w_1 \in U \cap W$ muß $w_1 = \lambda_1 c_1 + \cdots + \lambda_r c_r, \lambda_\rho \in \mathbb{K}$, gelten. Als Element von W besitzt somit w_1 die Basisdarstellungen

$$w_1 = \lambda_1 c_1 + \cdots + \lambda_r c_r + 0 \cdot b_1 + \cdots + 0 \cdot b_t$$

und

$$w_1 = 0 \cdot c_r + \cdots + 0 \cdot c_r - \beta_1 b_1 - \cdots - \beta_t b_t.$$

Nach Satz 4.43, (d), gilt daher $\lambda_1 = \cdots = \lambda_r = 0 = \beta_1 = \cdots = \beta_t$ und damit auch $w_1 = u_1 = o$. Da $B_0 \cup B_1$ Basis von U ist, folgt aus $u_1 = 0$

$$\gamma_1 = \cdots = \gamma_r = \alpha_1 = \cdots = \alpha_s = 0.$$

Damit ist gezeigt, dass B linear unabhängig ist, d.h. B ist Basis von $U + W$.

c) Die Beziehung zwischen den Dimensionen folgt aus $\dim(U + W) = r + s + t = (r + s) + (r + t) - r = \dim U + \dim W - \dim U \cap W$. □

Korollar 4.59. *V sei Vektoraum über $\mathbb{K}$ und U, W seien Unterräume von V mit $\dim U < \infty$ und $\dim W < \infty$. Dann gilt:*

Es ist $U + W = U \oplus W$ genau dann, wenn $\dim(U + W) = \dim U + \dim W$ gilt.

Beweis. Es gilt $U + W = U \oplus W$ genau dann, wenn $U \cap W = \{o\}$ gilt. Dies ist wiederum äquivalent zu $\dim U \cap W = 0$. Letzteres ist äquivalent zu $\dim(U + W) = \dim U + \dim W$. □

Beispiel 4.60. Es seien V, U_1 und U_2 wie in Beispiel 4.34. Dann gilt:

$$\dim U_1 = \dim U_2 = 2,$$
$$\dim(U_1 \cap U_2) = \begin{cases} 1 & \text{für } E_1 \nparallel E_2, \\ 2 & \text{für } E_1 \| E_2, \end{cases}$$
$$\dim(U_1 + U_2) = \begin{cases} 3 & \text{für } E_1 \nparallel E_2, \\ 2 & \text{für } E_1 \| E_2. \end{cases}$$

$\Diamond$

Es sei $\dim V = n$ und $B = \{a_1, \dots, a_n\}$ eine Basis von V. Dann kann jeder Vektor $a \in V$ in eindeutiger Weise dargestellt werden als

$$a = \alpha_1 a_1 + \cdots + \alpha_n a_n, \quad \alpha_i \in \mathbb{K},$$

(vgl. Satz 4.43, d)). Es ist nun naheliegend, dem Element a das n-Tupel $(\alpha_1, \dots, \alpha_n)$ zuzuordnen, da die α_i durch a eindeutig bestimmt sind. Dies setzt allerdings voraus, dass die Elemente der Basis geordnet sind. Für endlich dimensionale Räume werden wir daher die Definition des Begriffes „Basis" etwas modifizieren. Für unendlich dimensionale Räume spielen Basen im Sinne von Definition 4.38 (dann auch Hamel-Basen[24] genannt) eine sehr viel geringere Rolle.

Definition 4.61. *V sei ein Vektorraum über $\mathbb{K}$ mit $\dim V = n \neq 0$. Eine **geordnete Basis** $\mathcal{B}$ von V ist ein geordnetes n-Tupel von linear unabhängigen Vektoren aus V, $\mathcal{B} = (a_1, \dots, a_n)$, $a_i \in V$.*

Es ist klar, dass für eine geordnete Basis $\mathcal{B} = (a_1, \dots, a_n)$ die Menge $B = \{a_1, \dots, a_n\}$ eine Basis im Sinne von Definition 4.38 ist[25].

Definition 4.62. *$\mathcal{B} = (a_1, \dots, a_n)$ sei eine geordnete Basis von V, $\dim V = n$. Für $a \in V$ heißen die eindeutig bestimmten Elemente $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ mit*

$$a = \alpha_1 a_1 + \cdots + \alpha_n a_n$$

*die **Koordinaten** von a bezüglich der Basis $\mathcal{B}$. Der Vektor $(\alpha_1, \dots, \alpha_n) \in \mathbb{K}^n$ heißt der **Koordinatenvektor** von a bezüglich $\mathcal{B}$.*

Satz 4.63. *$\mathcal{B} = (a_1, \dots, a_n)$ sei eine geordnete Basis von V. Die Abbildung $\iota : V \to \mathbb{K}^n$ definiert durch*

$$\iota(a) = (\alpha_1, \dots, \alpha_n), \text{ wobei } a = \alpha_1 a_1 + \cdots + \alpha_n a_n \text{ gilt,}$$

ist bijektiv. Außerdem ist

$$\iota(\lambda a + \mu b) = \lambda \iota(a) + \mu \iota(b) \quad \text{für alle } a, b \in V \text{ und alle } \lambda, \mu \in \mathbb{K}.$$

Beweis. Dass ι bijektiv ist beweise man als Übung. Es sei $\iota(a) = (\alpha_1, \dots, \alpha_n)$ und $\iota(b) = (\beta_1, \dots, \beta_n)$, d.h. $a = \alpha_1 a_1 + \cdots + \alpha_n a_n, b = \beta_1 a_1 + \cdots + \beta_n a_n$. Daraus folgt $\lambda a + \mu b = (\lambda\alpha_1 + \mu\beta_1)a_1 + \cdots + (\lambda\alpha_n + \mu\beta_n)a_n$, d.h. $\iota(\lambda a + \mu b) = (\lambda\alpha_1 + \mu\beta_1, \dots, \lambda\alpha_n + \mu\beta_n) = \lambda(\alpha_1, \dots, \alpha_n) + \mu(\beta_1, \dots, \beta_n) = \lambda\iota(a) + \mu\iota(b)$. $\square$

Durch Satz 4.63 werden die algebraischen Operationen in V auf die entsprechenden Operationen im Raum $\mathbb{K}^n$ zurückgeführt, d.h. auf das Rechnen mit Koordinatenvektoren.

[24] Hamel, Georg Karl Wilhelm, 12. 9. 1877 (Düren) – 4. 10. 1954 (Landshut), Beiträge zur Mechanik, den Grundlagen der Mathematik und zur Funktionentheorie (www-groups.dcs.st-and.ac.uk/~history/Hamel.html).

[25] Im folgenden werden geordnete Basen durch Skriptbuchstaben bezeichnet.

Satz 4.64. *Es sei* $\dim V = n$. *Die Vektoren* $b_1, \dots, b_k \in V$ *sind genau dann linear unabhängig, wenn dies für die entsprechenden Koordinatenvektoren bezüglich einer festen Basis von* V *gilt.*

Beweis. Es sei $(\beta_1^i, \dots, \beta_n^i)$ der Koordinatenvektor von b_i bezüglich der gewählten Basis. Dann ist wegen Satz 4.63 $\lambda_1 a_1 + \cdots + \lambda_k a_k = o$ äquivalent zu $\lambda_1(\beta_1^1, \dots, \beta_n^1) + \cdots + \lambda_k(\beta_1^k, \dots, \beta_n^k) = (0, \dots, 0)$. □

Die Feststellung der linearen Abhängigkeit oder Unabhängigkeit von Vektoren im $\mathbb{K}^n$ führt auf die Diskussion von linearen Gleichungssystemen. Denn die Beziehung

$$\lambda_1(\beta_1^1, \dots, \beta_n^1) + \cdots + \lambda_k(\beta_1^k, \dots, \beta_n^k) = (0, \dots, 0)$$

bedeutet

$$(\lambda_1\beta_1^1 + \cdots + \lambda_k\beta_1^k, \dots, \lambda_1\beta_n^1 + \cdots + \lambda_k\beta_n^k) = (0, \dots, 0),$$

d.h. $\lambda_1, \dots, \lambda_k$ müssen Lösungen des folgenden Gleichungssystemes sein:

$$\begin{aligned} \beta_1^1\lambda_1 + \cdots + \beta_1^k\lambda_k &= 0, \\ \vdots \qquad\qquad \vdots \quad & \;\;\vdots \\ \beta_n^1\lambda_1 + \cdots + \beta_n^k\lambda_k &= 0. \end{aligned}$$

Kapitel 5

Lineare Abbildungen

5.1 Lineare Abbildungen

Definition 5.1. *X, Y seien Vektorräume über $\mathbb{K}$ und φ eine Abbildung $X \to Y$. φ heißt genau dann eine **lineare Abbildung** von X in Y, wenn gilt:*

$$\varphi(\alpha x + \beta y) = \alpha\varphi(x) + \beta\varphi(y) \text{ für alle } x, y \in X \text{ und alle } \alpha, \beta \in \mathbb{K}.$$

*Ist die lineare Abbildung φ bijektiv, so heißt sie ein **Isomorphismus** zwischen X und Y. X und Y heißen dann **isomorph**. Eine lineare Abbildung von X in X heißt (linearer) **Endomorphismus** von X, und ein bijektiver Endomorphismus von X heißt **Automorphismus**.*

Einige elementare Eigenschaften sind die folgenden:

φ sei eine lineare Abbildung von X in Y. Dann gilt:

1. *$\varphi(o) = o$.*
2. *$\varphi(-a) = -\varphi(a)$ für alle $a \in X$.*
3. *Sind $a_1, \ldots, a_k$ linear abhängig, so auch $\varphi(a_1), \ldots, \varphi(a_k)$ (oder gleichbedeutend: Sind $\varphi(a_1), \ldots, \varphi(a_k)$ linear unabhängig, so auch $a_1, \ldots, a_k$).*

Beweis. 1. Aus $\varphi(o) = \varphi(o + o) = \varphi(o) + \varphi(o)$ folgt $\varphi(o) = o$.
2. Aus $o = \varphi(o) = \varphi(a + (-a)) = \varphi(a) + \varphi(-a)$ folgt $\varphi(-a) = -\varphi(a)$.
3. $a_1, \ldots, a_k$ seien linear abhängig, d.h. für ein n-Tupel $(\alpha_1, \ldots, \alpha_k) \neq (0, \ldots, 0)$ gilt

$$\alpha_1 a_1 + \cdots + \alpha_k a_k = o.$$

Daraus folgt wegen 1. und der Linearität von φ die Gleichung $o = \varphi(\alpha_1 a_1 + \cdots + \alpha_k a_k) = \alpha_1\varphi(a_1) + \cdots + \alpha_k\varphi(a_k)$. Die Vektoren $\varphi(a_1), \ldots, \varphi(a_k)$ sind somit linear abhängig. □

Beispiel 5.2. $\varphi(x) = o \in Y$ für alle $x \in X$ definiert eine lineare Abbildung, die ***Nullabbildung*** $X \to Y$. ◊

Beispiel 5.3. Es sei $X = Y$. Die identische Abbildung (d.h. $\varphi(x) = x$ für alle $x \in X$) ist ein Automorphismus von X. ◊

Beispiel 5.4. Es sei $X = Y$ und $\lambda \neq 0$. $\varphi(x) = \lambda x$ für alle $x \in X$ definiert einen Automorphismus von X. ◊

Beispiel 5.5. Es sei $X = \mathbb{K}[x]$. $\varphi(p) = \alpha_0 + \alpha_1 x^2 + \alpha_2 x^4 + \cdots + \alpha_k x^{2k}$ für $p = \alpha_0 + \alpha_1 x + \cdots + \alpha_k x^k$, definiert einen injektiven, aber nicht surjektiven Endomorphismus. $\diamondsuit$

Beispiel 5.6. Es sei $X = \{f \mid f : (0,1) \to \mathbb{R},\ f$ beliebig oft differenzierbar$\}$ und $\varphi(f) = f'$ für alle $f \in X$. φ ist ein surjektiver Endomorphismus (zu $f \in X$ ist $\varphi(g) = f$ mit $g = \int f dx$). φ ist nicht injektiv, denn $\varphi(f + g) = \varphi(f)$ für $g \equiv c \in \mathbb{R}$. $\diamondsuit$

Beispiel 5.7. Die Abbildung ι aus Satz 4.63 ist ein Isomorphismus $V \to \mathbb{K}^n$. $\diamondsuit$

Satz 5.8. *X, Y seien Vektorräume über $\mathbb{K}$ und B eine Basis von X. Zu jeder Abbildung $\varphi^* : B \to Y$ existiert genau eine lineare Abbildung $\varphi : X \to Y$ mit $\varphi(a) = \varphi^*(a)$ für alle $a \in B$. Mit anderen Worten:*

> *Eine lineare Abbildung von X in Y ist schon durch ihre Werte auf einer Basis von X eindeutig bestimmt.*

Beweis. 1. φ sei eine lineare Abbildung mit $\varphi(a) = \varphi^*(a)$ für $a \in B$. Wir wählen $b \in X$. Der Vektor b hat eine Darstellung der Form $b = \lambda_1 a_1 + \cdots + \lambda_n a_n$ mit $\lambda_i \in \mathbb{K}$, $a_i \in B$. Wegen der Linearität von φ gilt

$$\varphi(b) = \lambda_1 \varphi(a_1) + \cdots + \lambda_n \varphi(a_n) = \lambda_1 \varphi^*(a_1) + \cdots + \lambda_n \varphi^*(a_n).$$

Damit ist gezeigt, dass es höchstens eine lineare Abbildung mit der verlangten Eigenschaft gibt.
2. Durch $\varphi(b) = \lambda_1 \varphi^*(a_1) + \cdots + \lambda_n \varphi^*(a_n)$, falls $b = \lambda_1 a_1 + \cdots + \lambda_n a_n$, wird eine lineare Abbildung φ von X in Y mit der verlangten Eigenschaft definiert. $\square$

Satz 5.9. *φ sei eine lineare Abbildung von X in Y, U Unterraum von X und V Unterraum von Y.*

1. *$\varphi(U)$ ist Unterraum von Y ($\varphi(U) := \{\varphi(x) \mid x \in U\}$, siehe Definition* 1.19, a)).
2. *$\varphi^{-1}(V)$ ist Unterraum von X ($\varphi^{-1}(V) := \{x \in X \mid \varphi(x) \in V\}$, siehe Definition* 1.19, b)).

Beweis. Wir müssen nur die Voraussetzungen aus Satz 4.15 verifizieren.
1. Für $a, b \in \varphi(U)$ gilt $a = \varphi(x)$, $b = \varphi(y)$ mit $x, y \in U$. Daraus folgt $a + b = \varphi(x) + \varphi(y) = \varphi(x + y) \in \varphi(U)$, weil $x + y \in U$. Für $\lambda \in \mathbb{K}, a \in \varphi(U)$ gilt $\lambda a = \lambda\varphi(x) = \varphi(\lambda x) \in \varphi(U)$, weil mit x auch $\lambda x \in U$ ist.
2. Aus $a, b \in \varphi^{-1}(V)$ folgt $\varphi(a), \varphi(b) \in V$. Dann ist $\varphi(a + b) = \varphi(a) + \varphi(b) \in V$ und somit $a + b \in \varphi^{-1}(V)$. Für $\lambda \in \mathbb{K}$, $a \in \varphi^{-1}(V)$ gilt $\varphi(a) \in V$ und $\varphi(\lambda a) = \lambda\varphi(a) \in V$, d.h. $\lambda a \in \varphi^{-1}(V)$. $\square$

Für $U = X$ erhält man den Unteraum $\varphi(X)$ von Y, für $V = \{o\}$ den Unterraum $\varphi^{-1}(\{o\}) = \{x \in X \mid \varphi(x) = o\}$. Der Unterraum $\ker\varphi = \varphi^{-1}(\{o\})$ heißt der ***Kern*** von φ.

Definition 5.10. *φ sei eine lineare Abbildung von X in Y.*
1. Die Zahl $\dim\varphi(X)$ *heißt der* ***Rang*** *von φ,* $\operatorname{rang}\varphi = \dim\varphi(X)$.
2. Die Zahl $\dim(\ker\varphi)$ *heißt der* ***Defekt*** *von φ,* $\operatorname{def}\varphi = \dim(\ker\varphi)$.

Satz 5.11. *φ sei eine lineare Abbildung von X in Y. Dann gilt:*
1. φ ist genau dann injektiv, wenn $\operatorname{def}\varphi = 0$*, d.h.* $\ker\varphi = \{o\}$.

2. Ist $\dim Y < \infty$*, so ist* φ *genau dann surjektiv, wenn* $\operatorname{rang}\varphi = \dim Y$.

Beweis. 1. Für beliebige $x_1, x_2 \in X$ ist $\varphi(x_1) = \varphi(x_2)$ äquivalent mit $\varphi(x_1 - x_2) = o$. φ injektiv ist somit äquivalent zur Aussage „$\varphi(x_1 - x_2) = o \Rightarrow x_1 - x_2 = o$“. Letzteres ist aber gleichbedeutend mit $\ker\varphi = \{o\}$.
2. φ surjektiv ist gleichbedeutend mit $\varphi(X) = Y$. Nach Bemerkung 4.57 ist dies wiederum äquivalent zu $\operatorname{rang}\varphi = \dim\varphi(X) = \dim Y$. □

Dass die Aussage 2 aus Satz 5.11 nicht mehr für $\dim Y = \infty$ gilt zeigt Beispiel 5.5. Der Zusammenhang zwischen $\operatorname{rang}\varphi$ und $\operatorname{def}\varphi$ ist durch den folgenden Satz gegeben:

Satz 5.12. *φ sei eine lineare Abbildung von X in Y. Dann gilt:*

$$\operatorname{rang}\varphi + \operatorname{def}\varphi = \dim X.$$

Beweis. B_0 sei eine Basis von $\ker\varphi$. Wir setzen $B_0 = \emptyset$, falls $\ker\varphi = \{o\}$. Nach Satz 4.44 kann B_0 zu einer Basis von X erweitert werden, d.h. es gibt eine Teilmenge B_1 mit $B_1 \cap B_0 = \emptyset$ von X derart, dass $B = B_0 \cup B_1$ Basis von X ist. Wir setzen $B_1 = \emptyset$, falls $\ker\varphi = X$ ist. Wir zeigen zunächst:

$$\varphi(B_1) \text{ ist eine Basis von } \varphi(X), \text{ falls } B_1 \neq \emptyset \text{ ist.}$$

$c_1, \dots, c_m$ seien m paarweise verschiedene Elemente aus $\varphi(B_1)$ und es sei $\alpha_1 c_1 + \cdots + \alpha_m c_m = o$. Es existieren Elemente $b_1, \dots, b_m$ aus B_1 mit $\varphi(b_j) = c_j$, $j = 1, \dots, m$. Dann gilt

$$o = \alpha_1 c_1 + \cdots + \alpha_m c_m = \varphi(\alpha_1 b_1 + \cdots + \alpha_m b_m),$$

d.h. $\alpha_1 b_1 + \cdots + \alpha_m b_m \in \ker\varphi$. Da B_0 eine Basis von $\ker\varphi$ ist muß es Elemente $b'_1, \dots, b'_k \in B_0$ geben mit

$$\beta_1 b'_1 + \cdots + \beta_k b'_k = \alpha_1 b_1 + \cdots + \alpha_m b_m.$$

Wegen Satz 4.43, d) muß $\beta_1 = \cdots = \beta_k = \alpha_1 = \cdots = \alpha_m = 0$ sein, d.h. $c_1, \dots, c_m$ sind linear unabhängig. Damit ist die lineare Unabhängigkeit von $\varphi(B_1)$ gezeigt. Es sei nun $y \in \varphi(X)$, d.h. $y = \varphi(x)$ für ein $x \in X$. Da $B_0 \cup B_1$ Basis von X ist, gilt

$$x = \alpha_1 b'_1 + \cdots + \alpha_m b'_m + \beta_1 b_1 + \cdots + \beta_n b_n$$

mit $b'_j \in B_0$, $b_i \in B_1$. Da die $b'_j \in \ker\varphi$ sind, folgt

$$\begin{aligned} y = \varphi(x) &= \sum_{j=1}^{m} \alpha_j \varphi(b'_j) + \sum_{i=1}^{n} \beta_i \varphi(b_i) \\ &= \sum_{i=1}^{n} \beta_i \varphi(b_i) \in [\varphi(B_1)]. \end{aligned}$$

Damit ist auch $\varphi(X) = [\varphi(B_1)]$ gezeigt. Ist $B_1 = \emptyset$ so ist $\varphi(X) = \{o\}$.

Die Aussage des Satzes folgt nun einfach aus der Tatsache, dass $\operatorname{def}\varphi$ die Anzahl der Elemente in B_0, $\operatorname{rang}\varphi$ die Anzahl der Elemente in $\varphi(B_1)$ und damit gleich der Anzahl der Elemente in B_1 ist und schließlich $\dim X$ durch die Anzahl der Elemente in $B_0 \cup B_1$ gegeben ist. □

Korollar 5.13. *Es sei* $\dim X = \dim Y = n < \infty$*. Dann ist φ genau dann injektiv, wenn φ surjektiv ist.*

Beweis. Injektivität von φ ist nach Satz 5.11 gleichbedeutend mit $\operatorname{def}\varphi = 0$. Dies ist aber nach Satz 5.12 äquivalent zu $\operatorname{rang}\varphi = \dim X = n$. $\operatorname{rang}\varphi = n = \dim Y$ ist aber gleichbedeutend mit Surjektivität von φ (Satz 5.11). □

5.2 Der lineare Raum $L(X, Y)$

Definition 5.14. *X, Y seien lineare Räume über $\mathbb{K}$.*

$$L(X, Y) := \{\varphi \mid \varphi \text{ ist lineare Abbildung } X \text{ in } Y\}.$$

Für $\varphi, \psi \in L(X, Y)$ und $\lambda \in \mathbb{K}$ definieren wir:

1. $(\varphi + \psi)(x) = \varphi(x) + \psi(x)$ *für alle* $x \in X$.
2. $(\lambda\varphi)(x) = \lambda\varphi(x)$ *für alle* $x \in X$.

Satz 5.15. *$L(X, Y)$ ist ein linearer Raum über $\mathbb{K}$.*

Beweis. 1. Aus $(\varphi+\psi)(\alpha x+\beta y) = \varphi(\alpha x+\beta y)+\psi(\alpha x+\beta y) = \alpha\varphi(x)+\beta\varphi(y)+\alpha\psi(x)+\beta\psi = \alpha(\varphi(x) + \psi(x)) + \beta(\varphi(y) + \psi(y)) = \alpha(\varphi + \psi)(x) + \beta(\varphi + \psi)(y)$ für alle $x, y \in X$, $\alpha, \beta \in \mathbb{K}$, folgt dass $\varphi + \psi$ eine lineare Abbildung von X in Y ist, d.h. die Abbildung $(\varphi, \psi) \to \varphi + \psi$ definiert eine innere Verknüpfung auf $L(X, Y)$. „+“ ist kommutativ und assoziativ (Beweis als Übung). Die Nullabbildung ist Nullelement. Zu φ ist die Abbildung $(-1)\varphi$ das negative Element in $L(X, Y)$, d.h. $-\varphi = (-1)\varphi$. $L(X, Y)$ ist somit eine abelsche Gruppe.
2. Es gilt $(\lambda\varphi)(\alpha x + \beta y) = \lambda\varphi(\alpha x + \beta y) = \lambda\alpha\varphi(x) + \lambda\beta\varphi(y) = \alpha\lambda\varphi(x) + \beta\lambda\varphi(y) = \alpha(\lambda\varphi)(x) + \beta(\lambda\varphi)(y)$ für alle $x, y \in X$ und $\alpha, \beta \in \mathbb{K}$ (hier wird erstmals die Kommutativität der Multiplikation im Körper $\mathbb{K}$ benötigt!). Daraus folgt, dass $\lambda\varphi$ eine lineare Abbildung ist, d.h. die Abbildung $(\lambda, \varphi) \to \lambda\varphi$ definiert eine skalare Multiplikation. Als Übung beweise man, dass (V1) - (V4) gelten. □

Es sei nun B eine Basis von X, B' eine Basis von Y. Nach Satz 5.8 existiert zu beliebig gewählten Elementen $a \in B$ und $a' \in B'$ genau eine lineare Abbildung $\omega_{a,a'}$ von X in Y mit $\omega_{a,a'}(a) = a'$ und $\omega_{a,a'}(b) = 0$ für alle $b \in B$, $b \neq a$.

Wir definieren

$$\Omega := \{\omega_{a,a'} \mid a \in B, a' \in B'\}$$

und $\Omega = \emptyset$, falls $\dim X = 0$ oder $\dim Y = 0$.

Satz 5.16. a) *Ω ist in $L(X, Y)$ linear unabhängig.*
b) *Es sei $\Omega \neq \emptyset$. Dann gilt:*

(i) *Es ist genau dann* $\dim X < \infty$, *wenn Ω Basis von $L(X, Y)$ ist.*

(ii) $\dim L(X, Y) = \dim X \cdot \dim Y$ (*hierbei vereinbart man* $\infty \cdot \infty = a \cdot \infty = \infty$ *für* $a \in \mathbb{N}$ *und* $\infty \cdot 0 = 0$).

Beweis. a) Wir wählen paarweise verschiedene $\omega_1, \ldots, \omega_r \in \Omega$. Für Elemente $a_i \in B$, $b_i \in B'$ gilt $\omega_i = \omega_{a_i,b_i}$. Da die ω_i paarweise verschieden sind, müssen die Paare (a_i, b_i), $i = 1, \ldots, r$, paarweise verschieden sein. Für $\lambda_i \in \mathbb{K}$ gelte

$$\lambda_1\omega_1 + \cdots + \lambda_r\omega_r = o$$

(hier bezeichnet o die Nullabbildung). Für a_k, $1 \le k \le r$, gilt nun

$$\omega_{a_i,b_i}(a_k) = \begin{cases} b_k & \text{für } i = k, \\ 0 & \text{für } i \ne k. \end{cases}$$

Damit erhalten wir

$$o = \sum_{i=1}^{r} \lambda_i \omega_{a_i,b_i}(a_k) = \lambda_k b_k + \lambda_{i_1} b_{i_1} + \cdots + \lambda_{i_m} b_{i_m}$$

wobei $i_1, \ldots, i_m$ jene Indices $\ne k$ sind, für welche $a_{i_1} = \cdots = a_{i_m} = a_k$ gilt. Da die Paare (a_i, b_i) paarweise verschieden sind, müssen $b_k, b_{i_1}, \ldots, b_{i_m}$ paarweise verschiedene Elemente aus B' sein. Da B' Basis ist, folgt aus der oben erhaltenen Beziehung

$$\lambda_k = \lambda_{i_1} = \cdots = \lambda_{i_m} = 0.$$

k war beliebig. Daher gilt $\lambda_1 = \cdots = \lambda_r = 0$, d.h. $\omega_1, \ldots, \omega_r$ sind linear unabhängig.

b1) Es sei $\dim X = n < \infty$ und $\varphi \in L(X,Y)$. $B = \{a_1, \ldots, a_n\}$ sei eine Basis von X, B' eine Basis von Y. Die Bilder $\varphi(a_i)$, $i = 1, \ldots, n$, sind Linearkombinationen je endlich vieler Elemente aus B'. Es kommen daher insgesamt nur endlich viele Elemente aus B' vor, etwa $b_1, \ldots, b_m$. Es gilt daher

$$\varphi(a_i) = \sum_{j=1}^{m} \alpha_{ji} b_j, \quad i = 1, \ldots, n.$$

Beachten wir die Definition von ω_{a_i,b_j}, so gilt für a_k, $1 \le k \le n$,

$$\begin{aligned} \Big(\sum_{j=1}^{m} \alpha_{ji} \omega_{a_i,b_j}\Big)(a_k) &= \sum_{j=1}^{m} \alpha_{ji} \omega_{a_i,b_j}(a_k) \\ &= \begin{cases} o & \text{für } i \ne k, \\ \sum_{j=1}^{m} \alpha_{jk} b_j = \varphi(a_k) & \text{für } i = k. \end{cases} \end{aligned}$$

Daraus folgt

$$\varphi(a_k) = \Big(\sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ji} \omega_{a_i,b_j}\Big)(a_k), \quad k = 1, \ldots, n,$$

d.h. $\varphi = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ji} \omega_{a_i,b_j}$ (man beachte Satz 5.8). Damit ist $\varphi \in [\Omega]$ gezeigt. Ω ist also Basis von $L(X,Y)$.

Wir müssen noch zeigen, dass Ω keine Basis ist, falls $\dim X = \infty$ gilt. B sei eine Basis von X und b sei fest aus B' (Basis von Y) gewählt. Es ist insbesondere $b \ne 0$. Nach Satz 5.8 existiert genau eine Abbildung $\varphi \in L(X,Y)$ mit $\varphi(a) = b$ für alle $a \in B$. Angenommen, es ist $\varphi = \alpha_1\omega_1 + \cdots + \alpha_m\omega_m$ mit $\omega_j \in \Omega$ und $\alpha_j \in \mathbb{K}$. Für jedes j existieren Elemente $a_j \in B$ und $b_j \in B'$ mit $\omega_j = \omega_{a_j,b_j}$ (d.h. $\omega_j(a_j) = b_j$ und $\omega_j(a) = o$ für $a \in B$, $a \ne a_j$). Es sei nun $a \in B \backslash \{a_1, \ldots, a_m\}$ gewählt. Dies ist möglich, da B unendlich ist. Dann gilt $\varphi(a) = b$ und andererseits $\varphi(a) = \alpha_1\omega_1(a) + \cdots + \alpha_m\omega_m(a) = o$. Dieser Widerspruch zeigt $\varphi \notin [\Omega]$, d.h. Ω ist keine Basis.

b2) Es sei $\dim X = 0$ oder $\dim Y = 0$. Dann ist $L(X,Y)$ der Nullraum, d.h. $\dim L(X,Y) = 0$.

Ist $0 < \dim X = n < \infty$ und $0 < \dim Y = r < \infty$, so folgt aus b), dass Ω eine Basis von $L(X,Y)$ ist. Da Ω nr Elemente besitzt, ist $\dim L(X,Y) = nr$.

Ist $\dim X = \infty$ oder $\dim Y = \infty$ und $\dim X \ne 0$, $\dim Y \ne 0$, so ist Ω unendlich und daher auch $\dim L(X,Y) = \infty$. $\square$

5.3 Matrizen

Für diesen Abschnitt wird $\dim X = n$, $0 < n < \infty$ und $\dim Y = r$, $0 < r < \infty$, vorausgesetzt. $\mathcal{B} = (a_1, \ldots, a_n)$ sei eine geordnete Basis von X und $\mathcal{B}' = (b_1, \ldots, b_r)$ eine geordnete Basis von Y. Statt ω_{a_i,b_j} schreiben wir kürzer ω_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, r$. $\mathcal{O} = (\omega_{11}, \ldots, \omega_{n1}, \omega_{12}, \ldots, \omega_{n2}, \ldots, \omega_{1r}, \ldots, \omega_{nr})$ heißt die zu $\mathcal{B}$ und $\mathcal{B}'$ gehörende kanonische Basis von $L(X, Y)$[26]. Zu jeder Abbildung $\varphi \in L(X, Y)$ existieren eindeutig bestimmte Koeffizienten α_{ji} mit

$$\varphi = \alpha_{11}\omega_{11} + \cdots + \alpha_{1n}\omega_{n1} + \alpha_{21}\omega_{12} + \cdots + \alpha_{2n}\omega_{n2} + \cdots + \alpha_{r1}\omega_{1r} + \cdots + \alpha_{rn}\omega_{nr}. \tag{5.1}$$

Die Koordinaten von φ bezüglich $\mathcal{O}$ faßt man nicht als nr-Tupel auf, sondern schreibt sie in ein Rechteckschema bestehend aus r Zeilen und n Spalten:

$$\begin{pmatrix} \alpha_{11} & \ldots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{r1} & \ldots & \alpha_{rn} \end{pmatrix}.$$

Man nennt ein solches Schema eine ***$r \times n$-Matrix***[27] mit Elementen aus $\mathbb{K}$. Die Zweckmäßigkeit dieser Vorgangsweise wird sehr bald klar werden. α_{ij}, $i = 1, \ldots, r$, $j = 1, \ldots, n$, ist das Element in der i-ten Zeile und j-ten Spalte der Matrix. $(\alpha_{i1}, \ldots, \alpha_{in})$ heißt der i-te ***Zeilenvektor*** der Matrix,

$$A = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{rj} \end{pmatrix}$$

der j-te ***Spaltenvektor***. Wir nennen A die der Abbildung φ bezüglich der geordneten Basen $\mathcal{B}$ und $\mathcal{B}'$ ***zugeordnete Matrix*** und schreiben $A = M(\varphi; \mathcal{B}, \mathcal{B}')$.

ξ sei der Koordinatenvektor von $x \in X$ bzgl. $\mathcal{B}$, η jener von $\varphi(x)$ bzgl. $\mathcal{B}'$ und es sei $A = M(\varphi; \mathcal{B}, \mathcal{B}')$. Es wäre wünschenswert, η mit Hilfe von A durch ξ auszudrücken. Zunächst beweisen wir das folgende Resultat über die Spalten von A:

Satz 5.17. *Es sei $\varphi \in L(X, Y)$ gegeben. $\mathcal{B} = (a_1, \ldots, a_n)$ bzw. $\mathcal{B}' = (b_1, \ldots, b_r)$ seien geordnete Basen von X bzw. Y. Eine Matrix $A = (\alpha_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, r$, ist genau dann bzgl. $\mathcal{B}$ und $\mathcal{B}'$ der Abbildung φ zugeordnet (d.h. die α_{ij} sind die Koordinaten von φ bzgl. der Basis $\mathcal{O}$ von $L(X, Y)$), wenn die k-te Spalte von A der Koordinatenvektor von $\varphi(a_k)$ bezgl. $\mathcal{B}'$ ist, $k = 1, \ldots, r$.*

Beweis. 1. Es sei $A = (\alpha_{ij}) = M(\varphi; \mathcal{B}, \mathcal{B}')$. Für $k = 1, \ldots, n$ gilt

$$\omega_{ij}(a_k) = \begin{cases} b_j, & \text{falls } i = k, \\ o & \text{sonst.} \end{cases}$$

Damit erhält man aus (5.1)

$$\varphi(a_k) = \alpha_{1k}b_1 + \cdots + \alpha_{rk}b_r,$$

d.h. $\begin{pmatrix} \alpha_{1k} \\ \vdots \\ \alpha_{rk} \end{pmatrix}$ ist der Koordinatenvektor von $\varphi(a_k)$ bzgl. $\mathcal{B}'$.

[26] Auf Grund der Resultate von Abschnitt 5.2 ist klar, dass $\mathcal{O}$ eine geordnete Basis von $L(X, Y)$ ist.

[27] Die Bezeichnung "Matrix" geht auf James Joseph Sylvester zurück (1850).

2. Wir bilden nun umgekehrt die Matrix A, indem wir in die k-te Spalte den Koordinatenvektor von $\varphi(a_k)$ bzgl. $\mathcal{B}'$ schreiben. Dann haben die Abbildungen φ und $\alpha_{11}\omega_{11}+\cdots+\alpha_{rn}\omega_{nr}$ dieselben Bilder für die Elemente der Basis $\mathcal{B}$. Nach Satz 5.8 gilt daher $\varphi = \alpha_{11}\omega_{11}+\cdots+\alpha_{rn}\omega_{nr}$, d.h. die α_{ij} sind die Koordinaten von φ bzgl. $\mathcal{O}$. □

Es sei nun $x \in X$ mit dem Koordinatenvektor ξ bezüglich der Basis $\mathcal{B}$ gegeben, d.h.

$$x = \xi_1 a_1 + \cdots + \xi_n a_n.$$

Dann gilt $\varphi(x) = \xi_1\varphi(a_1) + \cdots + \xi_n\varphi(a_n)$. Geht man zu den Koordinatenvektoren über, so erhält man unter Beachtung von Satz 4.63 und Satz 5.17

$$\begin{aligned}\eta &= \xi_1 \begin{pmatrix}\alpha_{11}\\ \vdots \\ \alpha_{r1}\end{pmatrix} + \cdots + \xi_n \begin{pmatrix}\alpha_{1n}\\ \vdots \\ \alpha_{rn}\end{pmatrix}\\ &= \begin{pmatrix}\alpha_{11}\xi_1 + \cdots + \alpha_{1n}\xi_n\\ \alpha_{21}\xi_1 + \cdots + \alpha_{2n}\xi_n\\ \vdots \\ \alpha_{r1}\xi_1 + \cdots + \alpha_{rn}\xi_n\end{pmatrix},\end{aligned} \tag{5.2}$$

wobei η der Koordinatenvektor von $\varphi(x)$ (bzgl. $\mathcal{B}'$) und $A = M(\varphi;\mathcal{B},\mathcal{B}')$ ist.

Ist $A = (\alpha_{ij})$ eine $r\times n$-Matrix und ξ ein n-dimensionaler Spaltenvektor, dann definieren wir das Produkt $\eta = A\xi$ durch (5.2), d.h. η ist ein r-dimensionaler Spaltenvektor, dessen i-te Koordinate η_i das innere Produkt (vgl. (2.11)) der i-ten Zeile von A mit ξ ist.

Satz 5.18. *Es seien $\varphi \in L(X,Y)$ sowie geordnete Basen $\mathcal{B} = (a_1,\ldots,a_n)$, $\mathcal{B}' = (b_1,\ldots,b_r)$ von X bzw. Y gegeben. Es sei $A = M(\varphi;\mathcal{B},\mathcal{B}')$. Dann gilt:*

$y = \varphi(x)$ gilt genau dann für ein $x \in X$, wenn der Koordinatenvektor η von y bezüglich $\mathcal{B}'$ durch $\eta = A\xi$ gegeben ist, wobei ξ der Koordinatenvektor von x bezüglich $\mathcal{B}$ ist.

Beweis. Es sei $\eta = A\xi$, d.h.

$$\eta = \xi_1 \begin{pmatrix}\alpha_{11}\\ \vdots \\ \alpha_{r1}\end{pmatrix} + \cdots + \xi_n \begin{pmatrix}\alpha_{1n}\\ \vdots \\ \alpha_{rn}\end{pmatrix}.$$

Nach Satz 5.17 gilt (vgl. auch Satz 4.63)

$$\begin{aligned}y &= \xi_1\varphi(a_1) + \cdots + \xi_n\varphi(a_n)\\ &= \varphi(\xi_1 a_1 + \cdots + \xi_n a_n) = \varphi(x).\end{aligned}$$

Die umgekehrte Aussage ist bereits weiter oben bewiesen. □

Korollar 5.19. *Es gilt genau dann $A = M(\varphi;\mathcal{B},\mathcal{B}')$, wenn $\eta = A\xi$ für alle $x \in X$ gilt, wobei ξ der Koordinatenvektor von x bzgl. $\mathcal{B}$ und η der Koordinatenvektor von $\varphi(x)$ bzgl. $\mathcal{B}'$ ist.*

Beweis. Es seien A und A' Matrizen mit $\eta = A\xi$ bzw. $\eta = A'\xi$ für alle $x \in X$. Wir wählen

$x = a_j$, a_j der j-te Vektor von $\mathcal{B}$. Dann ist der Koordinatenvektor ξ von x durch

$$\xi = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \\ \\ \\ - j\text{-te Stelle} \\ \\ \\ \\ \end{array}$$

gegeben und $A\xi$ bzw. $A'\xi$ ist die j-te Spalte von A bzw. A'. Wegen $\eta = A\xi = A'\xi$ stimmen daher die j-te Spalte von A und A' überein, $j = 1, \ldots, n$, d.h. $A = A'$. □

5.4 Rang einer Matrix

Für diesen Abschnitt seien X, Y Vektorräume über einem Körper $\mathbb{K}$ mit $\dim X = n < \infty$ und $\dim Y = r < \infty$.

Man kann die Zeilen bzw. Spalten einer $r \times n$-Matrix A als Elemente von $\mathbb{K}^n$ bzw. $\mathbb{K}^r$ auffassen. Wir definieren:

Zeilenrang von A := Maximalzahl linear unabhängiger Zeilen von A (als Elemente von $\mathbb{K}^n$),

Spaltenrang von A := Maximalzahl linear unabhängiger Spalten von A (als Elemente von $\mathbb{K}^r$).

Eines der wesentlichen Resultate dieses Abschnittes wird aussagen, dass stets „Zeilenrang von A = Spaltenrang von A“ gilt. Zunächst stellen wir einen Zusammenhang zwischen dem Rang einer linearen Abbildung und dem Spaltenrang der ihr zugeordneten Matrix her:

Satz 5.20. *φ sei eine lineare Abbildung $X \to Y$ und $A = \left(\alpha_{ij}\right)_{\substack{i=1,\ldots,r\\ j=1,\ldots,n}} = M(\varphi; \mathcal{B}, \mathcal{B}')$ für fest gewählte Basen $\mathcal{B}$ von X bzw. $\mathcal{B}'$ von Y. Dann gilt:*

$$\operatorname{rang} \varphi = \textit{Spaltenrang von } A.$$

Beweis. Nach Definition 5.10 ist $\operatorname{rang} \varphi = \dim \varphi(X)$. Es sei $(a_1, \ldots, a_n)$ die gewählte Basis für X. Dann gilt (Beweis als Übung)

$$\varphi(X) = [\varphi(a_1), \ldots, \varphi(a_n)].$$

Durch geeignete Numerierung der a_i kann man erreichen, dass die Vektoren $\varphi(a_1), \ldots, \varphi(a_k)$, $k \leq n$, linear unabhängig sind, während die $\varphi(a_i)$ mit $i = k+1, \ldots, n$, Linearkombinationen der $\varphi(a_1), \ldots, \varphi(a_k)$ sind. Dann ist $k = \dim \varphi(X) = \operatorname{rang} \varphi$. Da die Spalten von A die Koordinatenvektoren der $\varphi(a_i)$ bzgl. der in Y gewählten Basis sind, gilt: Die ersten k Spalten der Matrix A sind linear unabhängig, während die übrigen Spalten Linearkombinationen der ersten k Spalten sind (siehe Satz 4.64) . Daher gilt k = Spaltenrang von A. □

Ein wichtiges Hilfsmittel für die Untersuchung des Ranges einer Matrix sind gewisse einfache Umformungen.

Definition 5.21. *$\mathcal{B} = (a_1, \ldots, a_n)$ und $\mathcal{B}' = (a'_1, \ldots, a'_n)$ seien geordnete Basen von X. Die Basis $\mathcal{B}'$ ist genau dann durch* ***elementare Basistransformationen*** *aus $\mathcal{B}$ hervorgegangen, wenn $\mathcal{B}'$ aus $\mathcal{B}$ durch wiederholte Anwendung der folgenden Umformungen entstanden ist:*

(I) *Vertauschung von zwei Vektoren.*

(II) *Addition eines Vielfachen eines Vektors zu einem anderen Vektor.*

(III) *Multiplikation eines Vektors mit einem $\lambda \neq 0$ aus $\mathbb{K}$.*

Es ist klar, dass durch Umformungen des Typs (I) - (III) aus einer Basis immer eine Basis entsteht. Dies ist für den Typ (I) trivial. Für (II) und (III) folgt dies aus Lemma 4.47.

Wichtig ist es zu wissen, wie sich die Matrix einer Abbildung ändert, wenn die Basen von X und Y Umformungen des Typs (I) - (III) unterworfen werden.

Satz 5.22. *$\varphi \in L(X,Y)$ und geordnete Basen $\mathcal{B} = (a_1, \dots, a_n)$ bzw. $\mathcal{B}^* = (b_1, \dots, b_r)$ von X bzw. Y seien gegeben. Die Basen $\mathcal{B}_1$ bzw. $\mathcal{B}_1^*$ seien durch eine Umformung des Typs* (I) – (III) *aus den Basen $\mathcal{B}$ bzw. $\mathcal{B}^*$ entstanden. Ferner sei $A = M(\varphi; \mathcal{B}, \mathcal{B}^*)$ und $A_1 = M(\varphi; \mathcal{B}_1, \mathcal{B}_1^*)$.*

Dann entsprechen den Umformungen der Basen $\mathcal{B}$ bzw. $\mathcal{B}^*$ die in der folgenden Tabelle angegebenen Umformungen der Matrix A:

Basistransformation	*Matrixumformung*
Vertauschung von a_i und a_k	*Vertauschung der i-ten und k-ten Spalte*
Vertauschung von b_i und b_k	*Vertauschung der i-ten und k-ten Zeile*
Addition von λa_k, $\lambda \in \mathbb{K}$, $k \neq i$, zu a_i	*Addition der λ-fachen k-ten Spalte zur i-ten Spalte*
Addition von λb_k, $\lambda \in \mathbb{K}$, $k \neq i$, zu b_i	*Addition der $(-\lambda)$-fachen i-ten Zeile zur k-ten Zeile*
Multiplikation von a_i mit $\lambda \in \mathbb{K}$, $\lambda \neq 0$	*Multiplikation der i-ten Spalte mit λ*
Multiplikation von b_i mit $\lambda \in \mathbb{K}$, $\lambda \neq 0$	*Multiplikation der i-ten Zeile mit $1/\lambda$.*

Beweis. a) Umformungen des Typs (I) - (III) in der Basis $\mathcal{B}$ bewirken dieselben Umformungen in $(\varphi(a_1), \dots, \varphi(a_n))$. Da in den Spalten von A die Koordinatenvektoren der $\varphi(a_i)$ stehen, entsteht A_1 aus A jeweils durch Umformungen des Typs (I) - (III) in den Spaltenvektoren von A.

b) Wir betrachten zunächst eine Umformung des Typs (I) von $\mathcal{B}^*$, d.h. $\mathcal{B}_1^*$ entsteht aus $\mathcal{B}_1$ durch Vertauschung von b_i und b_k. Der Vektor $\varphi(a_j)$, $j = 1, \dots, n$, hat die Darstellungen

$$\varphi(a_j) = \alpha_{1j}b_1 + \cdots + \alpha_{ij}b_i + \cdots + \alpha_{kj}b_k + \cdots + \alpha_{rj}b_r$$

bzgl. $\mathcal{B}^*$ und

$$\varphi(a_j) = \alpha_{1j}b_1 + \cdots + \alpha_{kj}b_k + \cdots + \alpha_{ij}b_i + \cdots + \alpha_{rj}b_r$$

bzgl. $\mathcal{B}_1^*$. Daraus folgt sofort, dass A_1 aus A durch Vertauschung der i-ten und k-ten Zeile hervorgeht.

Es sei nun $\mathcal{B}_1^* = (b_1, \dots, b_i + \lambda b_k, \dots, b_r)$, $\lambda \in \mathbb{K}$, $k \neq i$. Dann folgt aus $\varphi(a_j) = \alpha_{1j}b_1 + \cdots + \alpha_{ij}b_i + \cdots + \alpha_{kj}b_k + \cdots + \alpha_{rj}b_r$

$$\varphi(a_j) = \alpha_{1j}b_1 + \cdots + \alpha_{ij}(b_i + \lambda b_k) + \cdots + (\alpha_{kj} - \lambda\alpha_{ij})b_k + \cdots + \alpha_{rj}b_r,$$

d.h. A_1 entsteht aus A durch Addition der $(-\lambda)$-fachen i-ten Zeile zur k-ten Zeile.

Gilt schließlich $\mathcal{B}_1^* = (b_1, \dots, \lambda b_i, \dots, b_n)$, $\lambda \neq 0$, so folgt aus $\varphi(a_j) = \alpha_{1j}b_1 + \cdots + \alpha_{ij}b_i + \cdots + \alpha_{rj}b_r$ auch

$$\varphi(a_j) = \alpha_{1j}b_1 + \cdots + \frac{1}{\lambda}\alpha_{ij}(\lambda b_i) + \cdots + \alpha_{rj}b_r,$$

d.h. A_1 entsteht aus A durch Multiplikation der i-ten Zeile mit $\frac{1}{\lambda}$.

Die auf der rechten Seite der Tabelle in Satz 5.22 angegebenen Umformungen einer Matrix werden wir kurz ***elementare Umformungen*** der Matrix A nennen.

Lemma 5.23. *A, B seien $r \times n$-Matrizen über $\mathbb{K}$.*

a) *Entsteht B aus A durch elementare Umformungen, so gilt*

Spaltenrang von B = Spaltenrang von A.

b) *Entsteht B aus A durch elementare Umformungen in den Zeilen, so ist*

Zeilenrang von B = Zeilenrang von A.

Beweis. a) Es sei $X = \mathbb{K}^n$, $Y = \mathbb{K}^r$. Wir wählen in beiden Räumen jeweils die kanonische Basis (siehe Beispiel 4.41). Dann ist A die Matrix jener linearen Abbildung $\varphi : \mathbb{K}^n \to \mathbb{K}^r$, die dem i-ten Vektor der kanonischen Basis von $\mathbb{K}^n$ den i-ten Spaltenvektor von A als Bild zuordnet. Da jeder elementaren Umformung von A gemäß Satz 5.22 eine elementare Basistransformation der Basis von $\mathbb{K}^n$ bzw. $\mathbb{K}^r$ entspricht, ist B die Matrix von φ bezüglich neuer Basen von $\mathbb{K}^n$ und $\mathbb{K}^r$, die jeweils aus der kanonischen Basis durch elementare Basistransformation hervorgegangen sind. Nach Satz 5.20 gilt:

$$\operatorname{rang}\varphi = \text{Spaltenrang von } A = \text{Spaltenrang von } B$$

b) Wir bezeichnen mit $\alpha_1, \dots, \alpha_r$ die Zeilen von A, d.h. $\alpha_i = (\alpha_{i1}, \dots, \alpha_{in})$, $i = 1, \dots, r$. Ist k der Zeilenrang von A, so gibt es k linear unabhängige Zeilen, während die restlichen Zeilen Linearkombinationen dieser k Zeilen sind. Eine Vertauschung zweier Zeilen ändert offensichtlich nichts an dieser Situation, d.h. der Zeilenrang bleibt bei Vertauschung von Zeilen unverändert. Wir können daher für das folgende annehmen, dass die Zeilen $\alpha_1, \dots, \alpha_k$ linear unabhängig sind, während die Zeilen $\alpha_{k+1}, \dots, \alpha_r$ als Linearkombinationen von $\alpha_1, \dots, \alpha_k$ dargestellt werden können, d.h. $\alpha_j \in [\alpha_1, \dots, \alpha_k]$, $j = k+1, \dots, r$.

Es werde nun α_i durch $\lambda\alpha_i$, $\lambda \neq 0$, ersetzt. Ist $1 \le i \le k$, so sind wegen Lemma 4.47 auch die Zeilen $\alpha_1, \dots, \lambda\alpha_i, \dots, \alpha_k$ linear unabhängig und $[\alpha_1, \dots, \alpha_{i-1}, \lambda\alpha_i, \alpha_{i+1}, \dots, \alpha_k] = [\alpha_1, \dots, \alpha_k]$. Der Zeilenrang der umgeformten Matrix ist wieder k. Den Fall $i > k$ untersuche man als Übung.

Es sei nun $1 \le i \le k$ und es werde α_i durch $\alpha_i + \lambda\alpha_j$, $i \neq j$, ersetzt. α_j besitzt eine Darstellung der Form

$$\alpha_j = \mu_1\alpha_1 + \cdots + \mu_k\alpha_k. \tag{5.3}$$

Daraus folgt

$$\alpha_i + \lambda\alpha_j = \lambda\mu_1\alpha_1 + \cdots + \lambda\mu_{i-1}\alpha_{i-1} + (1 + \lambda\mu_i)\alpha_i + \lambda\mu_{i+1}\alpha_{i+1} + \cdots + \lambda\mu_k\alpha_k.$$

Wir unterscheiden zwei Fälle:

Fall 1: $1 + \lambda\mu_i \neq 0$.

Nach Lemma 4.47 sind in diesem Fall die Vektoren

$$\alpha_1, \dots, \alpha_{i-1}, \alpha_i + \lambda\alpha_j, \alpha_{i+1}, \dots, \alpha_k$$

ebenfalls linear unabhängig und es gilt

$$[\alpha_1, \dots, \alpha_{i-1}, \alpha_i + \lambda\alpha_j, \alpha_{i+1}, \dots, \alpha_k] = [\alpha_1, \dots, \alpha_k],$$

d.h. jede Zeile, die eine Linearkombination von $\alpha_1, \dots, \alpha_k$ ist, ist auch eine von $\alpha_1, \dots, \alpha_{i-1}$, $\alpha_i + \lambda\alpha_j, \alpha_{i+1}, \dots, \alpha_k$. In diesem Fall besitzt die umgeformte Matrix ebenfalls den Zeilenrang k.

Fall 2: $1 + \lambda\mu_i = 0$.

In diesem Fall muß $\lambda \neq 0$ und $\mu_i \neq 0$ sein. Wegen $\mu_i \neq 0$ sind nach Hilfssatz 4.47 die Zeilen $\alpha_1, \dots, \alpha_{i-1}, \alpha_j, \alpha_{i+1}, \dots, \alpha_k$ linear unabhängig und es ist $[\alpha_1, \dots, \alpha_{i-1}, \alpha_j, \alpha_{i+1}, \dots, \alpha_k] =$

$[\alpha_1, \ldots, \alpha_k]$. Daraus folgt, dass jede Linearkombination der Zeilen $\alpha_1, \ldots, \alpha_k$ auch eine Linearkombination der Zeilen $\alpha_1, \ldots, \alpha_{i-1}, \alpha_j, \alpha_{i+1}, \ldots, \alpha_k$ ist. Auch in diesem Fall ist der Zeilenrang der umgeformten Matrix k.

Wir haben noch $i \geq k+1$ zu betrachten. Mit α_i ist aber auch $\alpha_i + \lambda\alpha_j$ eine Linearkombination von $\alpha_1, \ldots, \alpha_k$, woraus klar ist, dass der Zeilenrang der umgeformten Matrix ebenfalls k ist. □

Für die Bestimmung des Zeilen- bzw. Spaltenranges einer Matrix und später für das Lösen von linearen Gleichungssystemen ist der sog. ***Staffelalgorithmus*** wesentlich. Er besteht in der systematischen Hintereinanderausführung elementarer Zeilenumformungen.

Gegeben sei die $r \times n$-Matrix $A = (\alpha_{ij})$. Wir definieren zunächst die einzelnen Schritte des Staffelalgorithmus.

(s_j^i) Dies ist die Abfrage: Gibt es in der j-ten Spalte unter den Elementen $\alpha_{ij}, \alpha_{i+1,j}, \ldots, \alpha_{rj}$ ein Element $\neq 0$?

(v_j^i) Existiert in der j-ten Spalte unter den Elementen $\alpha_{ij}, \ldots, \alpha_{rj}$ ein Element $\neq 0$, so wähle eines aus – etwa α_{kj} –, dividiere die k-te Zeile durch α_{kj} und vertausche in der Matrix die i-te und k-te Zeile.

(n_j^i) Ist $\alpha_{ij} = 1$, so addiere das $(-\alpha_{mj})$-fache der i-ten Zeile zur m-ten Zeile, $m = 1, \ldots, r$, $m \neq i$.

Nach Durchführung von (v_j^i) ist in der umgeformten Matrix $\tilde{\alpha}_{ij} = 1$. Führt man zusätzlich (n_j^i) durch, so ist die j-te Spalte der umgeformten Matrix durch

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} - i\text{-te Stelle}$$

gegeben.

Der Staffelalgorithmus läuft folgendermaßen ab:

1. Beginne mit (s_1^1).

2. Ergibt (s_j^i) eine positive Antwort so führe (v_j^i) und danach (n_j^i) durch. Gehe dann zu (s_{j+1}^{i+1}).

3. Ergibt (s_j^i) eine negative Anwort, so gehe zu (s_{j+1}^i).

4. Ende des Verfahrens, falls $j + 1 = n + 1$.

Das folgende Ergebnis ist nun leicht einzusehen:

Satz 5.24. *Gegeben sei die $r \times n$-Matrix $A = (\alpha_{ij})$. Der Staffelalgorithmus bringt A in endlich* *vielen Schritten auf die Gestalt*

$$\tilde{A} = \left(\begin{array}{ccccccccccccccccccccc}
0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\
\vert & & \vert & 0 & 0 & \cdots & 0 & 1 & * & \cdots & & & \vert & \vert & \vert & & \vert & \vert & \vert & & \vert \\
\vert & & \vert & \vert & \vert & & \vert & 0 & 0 & \cdots & & & \vert & \vert & \vert & & \vert & \vert & \vert & & \vert \\
\vert & & \vert & \vert & \vert & & \vert & \vert & \vert & & & \cdots & * & 0 & \vert & & \vert & \vert & \vert & & \vert \\
\vert & & \vert & \vert & \vert & & \vert & \vert & \vert & & \cdots & \cdots & 0 & 1 & * & \cdots & * & 0 & \vert & & \vert \\
\vert & & \vert & \vert & \vert & & \vert & \vert & \vert & & & & \vert & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * \\
\vert & & \vert & \vert & \vert & & \vert & \vert & \vert & & & & \vert & \vert & \vert & & \vert & 0 & 0 & \cdots & 0 \\
\vert & & \vert & \vert & \vert & & \vert & \vert & \vert & & & & \vert & \vert & \vert & & \vert & \vert & \vert & & \vert \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0
\end{array}\right).$$

Spalten-Nr.: n_1 n_2 n_{k-1} n_k

Hierbei werden durch „∗“ Stellen bezeichnet, an welchen beliebige Elemente aus $\mathbb{K}$ *stehen können.*

Lemma 5.25. *Für die Matrix* $\tilde{A}$ *aus Satz* 5.24 *gilt:*

Spaltenrang von $\tilde{A}$ = *Zeilenrang von* $\tilde{A} = k$.

Beweis. Die ersten k Zeilen $\alpha_1, \dots, \alpha_k$ von $\tilde{A}$ sind linear unabhängig. Denn aus $\lambda_1\alpha_1 + \cdots + \lambda_k\alpha_k = o$ würde

$$(0, \dots, 0, \lambda_1, *, \dots, *, \lambda_2, *, \dots, *, \lambda_3, *, \dots, * \cdots *, \lambda_k, * \cdots *) = (0, \dots, 0)$$

folgen, d.h. $\lambda_1 = \cdots = \lambda_k = 0$. Der Zeilenrang von $\tilde{A}$ ist daher k.

Durch elementare Umformungen in den Spalten von $\tilde{A}$ kann $\tilde{A}$ in die $r \times n$-Matrix

$$I_k = \begin{pmatrix}
1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\
\vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\
0 & \cdots & 0 & 1 & \vdots & & \vdots \\
0 & \cdots & \cdots & \cdots & 0 & & \vdots \\
\vdots & & & & & & \vdots \\
0 & \cdots & & & & \cdots & 0
\end{pmatrix}, \tag{5.4}$$

übergeführt werden, wobei k Einselemente in der Hauptdiagonalen stehen. Der Spaltenrang von I_k ist offensichtlich k und wegen Lemma 5.23, a), gleich dem Spaltenrang von $\tilde{A}$. □

Satz 5.26. *Für jede Matrix ist der Spaltenrang gleich dem Zeilenrang.*

Beweis. Durch den Staffelalgorithmus (der nur elementare Umformungen in den Zeilen benutzt) wird weder Spaltenrang noch Zeilenrang geändert. (Lemma 5.23, a) und b)). Für die umgeformte Matrix gilt die Behauptung jedoch (Lemma 5.25). □

Auf Grund des eben bewiesenen Satzes ist die folgende Definition sinnvoll:

Definition 5.27. *A sei eine $r \times n$-Matrix. Der* ***Rang*** *von A ist definiert durch*

$$\operatorname{rang} A := \textit{Maximalzahl der linear unabhängigen Zeilen oder Spalten von } A.$$

Satz 5.20 kann nun neu formuliert werden:

Satz 5.28. *Die Voraussetzungen seien wie in Satz* 5.20*. Dann gilt*

$$\operatorname{rang} A = \operatorname{rang} \varphi.$$

Aus Satz 5.24 und dem Beweis von Lemma 5.25 folgt:

> *Ist A eine $r \times n$-Matrix mit $k = \operatorname{rang} A$, so läßt sie sich durch elementare Umformungen immer in die Matrix I_k umformen.*

Aus der Tatsache, dass die Zeilen bzw. Spalten einer $r \times n$-Matrix A Elemente von $\mathbb{K}^n$ bzw. $\mathbb{K}^r$ sind folgt die Abschätzung

$$0 \leq \operatorname{rang} A \leq \min(r, n).$$

Es ist klar, dass $\operatorname{rang} A = 0$ gleichbedeutend mit $A = 0$ ist.

Kapitel 6

Lineare Gleichungen

6.1 Lineare Gleichungen in Räumen beliebiger Dimension

Zahlreiche Problemstellungen innerhalb der Mathematik und in Anwendungsgebieten führen auf folgende Aufgabe:

Gegeben sind zwei Vektorräume X, Y über $\mathbb{K}$, eine lineare Abbildung $\varphi : X \to Y$ und ein Element $b \in Y$. Gesucht ist ein $x \in X$ mit

$$(\mathcal{L}_{\varphi,b}) \qquad \varphi(x) = b.$$

Wir nennen $(\mathcal{L}_{\varphi,b})$ eine ***lineare Gleichung***. Ist $b = o$, so heißt die Gleichung ***homogen***, sonst ***inhomogen***. Ein Element $x \in X$ mit $\varphi(x) = b$ heißt ***Lösung*** von $(\mathcal{L}_{\varphi,b})$.

Ist eine Gleichung $(\mathcal{L}_{\varphi,b})$ gegeben, so ist man an einer Beantwortung folgender Fragen interessiert:

1. *Existenz von Lösungen* für ein festes $b \in Y$ oder für alle $b \in Y$.

2. *Eindeutigkeit von Lösungen*, d.h. sind x_1, x_2 Lösungen von $(\mathcal{L}_{\varphi,b})$, so gilt $x_1 = x_2$. Eindeutigkeit bedeutet nicht, dass überhaupt Lösungen existieren.

3. *Berechnung der Lösungen.*

Die homogene Gleichung besitzt stets die Lösung $x = o$, die sogenannte ***triviale Lösung***. Die homogene Gleichung $(\mathcal{L}_{\varphi,o})$ heißt ***nicht-trivial lösbar***, wenn eine Lösung $x \neq o$ existiert. Aus der Linearität von φ folgt sofort

Satz 6.1 (Superpositionsprinzip). *Gegeben seien die Abbildung $\varphi \in L(X, Y)$, die Elemente $b_1, b_2 \in Y$ und $\alpha, \beta \in \mathbb{K}$. Dann gilt: Ist x_1 bzw. x_2 Lösung von $(\mathcal{L}_{\varphi,b_1})$ bzw. $(\mathcal{L}_{\varphi,b_2})$, so ist $\alpha x_1 + \beta x_2$ Lösung von $(\mathcal{L}_{\varphi,\alpha b_1+\beta b_2})$.*

Beweis. Aus $\varphi(x_1) = b_1$ und $\varphi(x_2) = b_2$ folgt sofort $\varphi(\alpha x_1 + \beta x_2) = \alpha\varphi(x_1) + \beta\varphi(x_2) = \alpha b_1 + \beta b_2$. □

Korollar 6.2. *Es sei $(\mathcal{L}_{\varphi,b})$ gegeben. Ferner sei x_0 eine Lösung von $(\mathcal{L}_{\varphi,b})$ und $x_1 \in X$. Dann gilt: x_1 ist dann und nur dann eine Lösung von $(\mathcal{L}_{\varphi,b})$, wenn $x_1 = x_0 + y$ gilt, wobei y eine Lösung der homogenen Gleichung $(\mathcal{L}_{\varphi,o})$ ist.*

Beweis. Ist x_1 eine Lösung von $(\mathcal{L}_{\varphi,b})$, so ist, nach Satz 6.1, $y = x_1 - x_0$ eine Lösung von $(\mathcal{L}_{\varphi,o})$. Ist y eine Lösung von $(\mathcal{L}_{\varphi,o})$, so ist $x_0 + y$ eine Lösung von $(\mathcal{L}_{\varphi,b})$ (wieder wegen Satz 6.1). □

Die Lösungsmenge der homogenen Gleichung ist der lineare Unterraum $\ker \varphi$ von Y. Die Aussage des obigen Korollares kann man auch wie folgt formulieren: Ist x_0 eine Lösung von $(\mathcal{L}_{\varphi,b})$, so ist die Lösungsmenge von $(\mathcal{L}_{\varphi,b})$ durch

$$x_0 + \ker \varphi = \{x_0 + y \mid y \in \ker \varphi\}$$

gegeben.

Wir kommen nun zur Beantwortung der oben gestellten Fragen 1 und 2:

Satz 6.3. *Gegeben sei* $(\mathcal{L}_{\varphi,b})$. *Dann gilt:*

a) $(\mathcal{L}_{\varphi,b})$ *besitzt genau dann eine Lösung, wenn* $b \in \varphi(X)$ *ist.*

b) $(\mathcal{L}_{\varphi,b})$ *besitzt genau dann für alle* $b \in Y$ *eine Lösung, wenn* φ *surjektiv ist.*

c) *Die folgenden drei Aussagen sind äquivalent:*

(i) *Die Lösungen von* $(\mathcal{L}_{\varphi,b})$ *sind eindeutig.*

(ii) φ *ist injektiv.*

(iii) $(\mathcal{L}_{\varphi,o})$ *besitzt nur die triviale Lösung.*

Beweis. a) und b) beweise man als Übung. c) ist klar wegen Korollar 6.2 und wegen Satz 5.11, 1. □

Das folgende Beispiel einer linearen Gleichung ist für Hörer im ersten Studienjahr nicht verständlich. Es wird jedoch empfohlen, sich dieses Beispiel später (wenn man etwa eine Vorlesung über Differentialgleichungen gehört hat) nochmals anzusehen.

Beispiel 6.4. Es sei $X = C^1(\alpha, \beta; \mathbb{R}^n)$ und $Y = C(\alpha, \beta; \mathbb{R}^n)$ (Raum der stetigen bzw. stetig differenzierbaren Funktionen, welche das offene Intervall (α, β) in den $\mathbb{R}^n$ abbilden). Die Abbildung $\psi : X \to Y$ ist durch

$$\psi(x)(t) = x'(t) - Ax(t), \quad t \in (\alpha, \beta),$$

gegeben, wobei A eine vorgegebene $n \times n$-Matrix ist. Die homogene Gleichung $\psi(x) = o$ ist nichts anderes als die homogene lineare Differentialgleichung

$$x'(t) = Ax(t), \quad t \in (\alpha, \beta),$$

während der inhomogenen Gleichung $(\mathcal{L}_{\psi,b})$, $b \in C(\alpha, \beta; \mathbb{R}^n)$, die inhomogene Differentialgleichung

$$x'(t) = Ax(t) + b(t), \quad t \in (\alpha, \beta),$$

entspricht. In der Theorie der gewöhnlichen Differentialgleichungen beweist man $\operatorname{def} \psi = n$, d.h. der Lösungsraum der homogenen Differentialgleichung hat die Dimension n. Die Eindeutigkeit der Lösungen ist daher nicht gegeben. Um Eindeutigkeit zu erzielen muß man bekanntlich Anfangswerte vorschreiben. Dies kann man wieder als lineare Gleichung formulieren: Wir wählen $X = C^1(\alpha, \beta; \mathbb{R}^n)$ und $Y = \mathbb{R}^n \times C(\alpha, \beta; \mathbb{R}^n)$. Für festes $t_0 \in (\alpha, \beta)$ definieren wir

$$\varphi(x) = \big(x(t_0), x'(\cdot) - Ax(\cdot)\big).$$

Dann ist $\varphi(x) = (x_0, b)$ gleichwertig mit

$$\begin{aligned} x'(t) &= Ax(t) + b(t), \quad t \in (\alpha, \beta), \\ x(t_0) &= x_0. \end{aligned} \tag{6.1}$$

Da $(x_0, b) = (x_0, o) + (o, b)$ gilt, folgt nach dem Superpositionsprinzip, dass die Lösung von (6.1) die Summe der Lösung von

$$x' = Ax, \quad x(t_0) = x_0,$$

und der Lösung von

$$x' = Ax + b(t), \quad x(t_0) = o$$

ist. Die sogenannte Variation-der-Konstanten-Formel

$$x(t) = e^{A(t-t_0)}x_0 + \int_{t_0}^{t} e^{A(t-s)}b(s)\,ds, \quad t \in (\alpha, \beta),$$

für die Lösung von (6.1) drückt gerade diesen Sachverhalt aus. $\Diamond$

6.2 Lineare Gleichungen in endlich-dimensionalen Räumen

In diesem Abschnitt werden wir die Ergebnisse des vorhergehenden Abschnittes spezialisieren und insbesondere auf die Fragestellung 3 (Berechnung der Lösungen) eingehen.

Es sei in diesem Abschnitt stets $\dim X = n$ und $\dim Y = r$. Wir wählen geordnete Basen $\mathcal{B} = (a_1, \ldots, a_n)$ und $\mathcal{B}^* = (b_1, \ldots, b_r)$ von X bzw. Y. A sei die Matrix von φ bzgl. $\mathcal{B}$ und $\mathcal{B}^*$, β sei der Koordinatenvektor von b bzgl. $\mathcal{B}^*$. Nach Satz 5.18 ist x_0 genau dann Lösung von $(\mathcal{L}_{\varphi,b})$, wenn für den Koordinatenvektor ξ_0 von x_0 bzgl. $\mathcal{B}$

$$A\xi_0 = \beta$$

gilt, d.h. ξ_0 Lösung von

$$(\mathcal{S}_{A,\beta}) \qquad A\xi = \beta$$

ist. Ausführlich geschrieben hat $(\mathcal{S}_{A,\beta})$ die Form

$$\begin{aligned} \alpha_{11}\xi_1 + \alpha_{12}\xi_2 + \cdots + \alpha_{1n}\xi_n &= \beta_1, \\ \vdots \qquad\qquad\qquad\quad \vdots \;\; &\;\; \vdots \\ \alpha_{r1}\xi_1 + \alpha_{r2}\xi_2 + \cdots + \alpha_{rn}\xi_n &= \beta_r, \end{aligned}$$

wobei $A = (\alpha_{ij})$, $\xi = (\xi_1, \ldots, \xi_n)^\mathsf{T}$ und $\beta = (\beta_1, \ldots, \beta_r)^\mathsf{T}$ gilt[28]. Man nennt $(\mathcal{S}_{A,\beta})$ ein ***lineares Gleichungssystem*** mit Systemmatrix A und rechter Seite β.

Ist andererseits A eine $r \times n$-Matrix und β ein Vektor aus $\mathbb{K}^r$, so können wir die Ergebnisse aus Abschnitt 6.1 sofort auf $(\mathcal{S}_{A,\beta})$ anwenden. Wir setzen $X = \mathbb{K}^n$, $Y = \mathbb{K}^r$ und definieren $\varphi_A \in L(\mathbb{K}^n, \mathbb{K}^r)$ durch

$$\varphi_A(\xi) = A\xi, \quad \xi \in \mathbb{K}^n. \tag{6.2}$$

Die Matrix A ist dann die Matrix der Abbildung φ_A bzgl. der kanonischen Basen von $\mathbb{K}^n$ und $\mathbb{K}^r$.

Satz 6.5. *Es sei $(\mathcal{S}_{A,\beta})$ vorgelegt. Dann gilt:*

a) *Es existiert genau dann mindestens eine Lösung von $(\mathcal{S}_{A,\beta})$, wenn*

$$\operatorname{rang} A = \operatorname{rang}(A, \beta)$$

gilt.

b) *$(\mathcal{S}_{A,\beta})$ ist genau dann für jedes $\beta \in \mathbb{K}^r$ lösbar, wenn* $\operatorname{rang} A = r$ *ist.*
Notwendigerweise muß dann $n \geq r$ sein.

[28] Mit $(\beta_1, \ldots, \beta_r)^\mathrm{T}$ bezeichnen wir den Spaltenvektor mit den Elementen $\beta_1, \ldots, \beta_r$.

c) *Die Lösungen von* $(\mathcal{S}_{A,\beta})$ *sind genau dann eindeutig, wenn* $\operatorname{rang} A = n$ *ist.*
Notwendigerweise muß dann $r \geq n$ *sein.*

d) *Die Dimension des Lösungsraumes von* $(\mathcal{S}_{A,o})$ *ist* $n - \operatorname{rang} A$.

Beweis. a) $(\mathcal{S}_{A,\beta})$ ist genau dann lösbar, wenn $\beta \in \varphi_A(\mathbb{K}^n)$ gilt (Satz 6.3). Es ist $\varphi_A(\mathbb{K}^n) = [\varphi_A(e_1), \ldots, \varphi_A(e_n)] = [Ae_1, \ldots, Ae_n]$, wobei durch e_i, $i = 1, \ldots, n$, die Vektoren der kanonischen Basis von $\mathbb{K}^n$ bezeichnet werden. Da Ae_i die i-te Spalte von A ist, gilt $\beta \in \varphi_A(\mathbb{K}^n)$ dann und nur dann, wenn β eine Linearkombination der Spalten von A ist. Dies bedeutet aber, dass die Maximalzahl linear unabhängiger Spalten von A und (A, β) gleich ist, d.h. $\operatorname{rang} A = \operatorname{rang}(A, \beta)$.

b) Nach der schon bewiesenen Aussage a) ist $(\mathcal{S}_{A,\beta})$ für jedes $\beta \in \mathbb{K}^r$ genau dann lösbar, wenn

$$\operatorname{rang} A = \operatorname{rang}(A, \beta) \quad \text{für alle } \beta \in \mathbb{K}^r \tag{6.3}$$

gilt.

Es sei nun $\operatorname{rang} A = r$. Dann muß $n \geq r$ und auch $\operatorname{rang}(A, \beta) \geq r$ für alle $\beta \in \mathbb{K}^r$ sein. Da die Spalten von (A, β) Vektoren in $\mathbb{K}^r$ sind und $\dim \mathbb{K}^r = r$ gilt, muß $\operatorname{rang}(A, \beta) \leq r$ für alle $\beta \in \mathbb{K}^r$ sein, d.h. es gilt (6.3).

Ist $k = \operatorname{rang} A < r$, so existieren k linear unabhängige Spalten $\alpha_1, \ldots, \alpha_k$ von A, während die übrigen Spalten von A Linearkombinationen dieser k Spalten sind. Zu den Spaltenvektoren $\alpha_1, \ldots, \alpha_k$ existiert ein $\beta \in \mathbb{K}^r$ derart, dass $\alpha_1, \ldots, \alpha_k, \beta$ linear unabhängig sind (siehe Satz 4.44 oder Satz 4.48). Daraus folgt $\operatorname{rang}(A, \beta) = k + 1 > \operatorname{rang} A$, d.h. (6.3) gilt nicht. Damit ist gezeigt, dass (6.3) äquivalent zu $\operatorname{rang} A = r$ ist.

c) Die Lösungen von $(\mathcal{S}_{A,\beta})$ sind genau dann eindeutig, wenn φ_A injektiv ist, was wiederum äquivalent zu $\operatorname{def} \varphi_A = 0$ bzw. zu $\operatorname{rang} \varphi_A = \operatorname{rang} A = n$ ist.

d) Dies folgt unmittelbar aus $\dim \ker \varphi_A = \operatorname{def} \varphi_A = n - \operatorname{rang} \varphi_A = n - \operatorname{rang} A$. □

Wir entwickeln nun im folgenden ein Lösungsverfahren für $(\mathcal{S}_{A,\beta})$ bei dem gleichzeitig die Kriterien aus Satz 6.5 überprüft werden. Grundlegend für dieses Verfahren wird der Staffelalgorithmus sein.

Lemma 6.6. *Gegeben seien die* $r \times n$*-Matrizen* A, $\tilde{A}$ *sowie die Vektoren* $\beta, \tilde{\beta} \in \mathbb{K}^r$. *Geht die Matrix* $(\tilde{A}, \tilde{\beta})$ *aus* (A, β) *durch elementare Umformungen in den Zeilen hervor, so besitzen die Gleichungssysteme* $(\mathcal{S}_{A,\beta})$ *und* $(\mathcal{S}_{\tilde{A},\tilde{\beta}})$ *dieselben Lösungen.*

Beweis. Eine Vertauschung von zwei Zeilen in (A, β) entspricht der Vertauschung von zwei Gleichungen, wodurch die Lösungsmenge offensichtlich nicht geändert wird. Für $\lambda \neq 0$ ist

$$\alpha_{i1}\xi_1 + \cdots + \alpha_{in}\xi_n = \beta_i$$

äquivalent zu

$$(\lambda\alpha_{i1})\xi_1 + \cdots + (\lambda\alpha_{in})\xi_n = \lambda\beta_i,$$

d.h. Multiplikation einer Zeile von (A, β) mit $\lambda \neq 0$ ändert nichts an der Lösungsmenge des zugehörigen Gleichungssystemes. Schließlich folgt aus

$$\alpha_{i1}\xi_1 + \cdots + \alpha_{in}\xi_n = \beta_i$$

und

$$\alpha_{j1}\xi_1 + \cdots + \alpha_{jn}\xi_n = \beta_j,$$

für $i \neq j$ auch

$$(\alpha_{i1} + \lambda\alpha_{j1})\xi_1 + \cdots + (\alpha_{in} + \lambda\alpha_{jn})\xi_n = \beta_i + \lambda\beta_j, \tag{6.4}$$

d.h. $(\xi_1, \ldots, \xi_n)^\mathsf{T}$ löst auch das System, in dem die i-te Gleichung durch (6.4) ersetzt wird. Löst andererseits $(\xi_1, \ldots, \xi_n)^\mathsf{T}$ das transformierte System, d.h. es gilt insbesondere

$$(\alpha_{i1} + \lambda\alpha_{j1})\xi_1 + \cdots + (\alpha_{in} + \lambda\alpha_{jn})\xi_n = \beta_i + \lambda\beta_j$$

und

$$\alpha_{j1}\xi_1 + \cdots + \alpha_{jn}\xi_n = \beta_j,$$

so folgt sofort

$$\alpha_{i1}\xi_1 + \cdots + \alpha_{in}\xi_n = \beta_i,$$

d.h. $(\xi_1, \ldots, \xi_n)^\mathsf{T}$ löst auch das ursprüngliche System. Damit ist das Lemma vollständig bewiesen. $\square$

Der Gaußsche Lösungsalgorithmus für $(\mathcal{S}_{A,\beta})$**:**

Wendet man auf (A, β) den Staffelalgorithmus an, so erhält man eine Matrix $(\tilde{A}, \tilde{\beta})$. Lemma 6.6 besagt, dass die Gleichungssysteme $(\mathcal{S}_{A,\beta})$ und $(\mathcal{S}_{\tilde{A},\tilde{\beta}})$ dieselben Lösungen besitzen. Die Matrix $(\tilde{A}, \tilde{\beta})$ hat folgende Gestalt (vgl. Satz 5.24):

$$\begin{array}{c}
\left(\begin{array}{ccccccccccccccccccc|c}
0 & \text{—} & 0 & 1 & * & \text{—} & * & 0 & * & \text{—} & & \text{—} & * & 0 & * & \text{—} & * & 0 & * \;\text{—}\; * & \tilde{\beta}_1 \\
\vdots & & \vdots & 0 & 0 & \text{—} & 0 & 1 & * & \text{—} & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \tilde{\beta}_2 \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & 0 & \text{—} & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & \text{—} & * & 0 & \vdots & & \vdots & \vdots & \vdots & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \cdots & \text{—} & 0 & 1 & * & \text{—} & * & 0 & \vdots & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & \vdots & 0 & 0 & \text{—} & 0 & 1 & * \;\text{—}\; * & \tilde{\beta}_k \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & \vdots & \vdots & \vdots & & \vdots & 0 & 0 \;\text{—}\; 0 & \tilde{\beta}_{k+1} \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\
0 & \text{—} & 0 & 0 & 0 & \text{—} & 0 & 0 & 0 & \text{—} & & \text{—} & 0 & 0 & 0 & \text{—} & 0 & 0 & 0 \;\text{—}\; 0 & \tilde{\beta}_r
\end{array}\right). \\
\text{Spalte: } n_1 \qquad\quad n_2 \qquad\qquad\qquad\qquad n_{k-1} \qquad\quad n_k
\end{array}$$

Hierbei ist $k = \operatorname{rang} A$. Die Lösbarkeitsbedingung $\operatorname{rang} A = \operatorname{rang}(A, \beta)$ ist wegen $\operatorname{rang} \tilde{A} = \operatorname{rang} A$ und $\operatorname{rang}(\tilde{A}, \tilde{\beta}) = \operatorname{rang}(A, \beta)$ genau dann erfüllt, wenn

$$\tilde{\beta}_{k+1} = \cdots = \tilde{\beta}_r = 0.$$

Ist diese Bedingung erfüllt, so erhalten wir

$$\xi_{n_i} = \tilde{\beta}_i - \sum_{j=n_i+1}^{n} \tilde{\alpha}_{ij}\xi_j, \quad i = 1, \ldots, k.$$

Es ist zu beachten, dass $\xi_{n_1}, \ldots, \xi_{n_k}$ wegen $\tilde{\alpha}_{i,n_j} = 0$ für $j \neq i$ auf der rechten Seite nicht vorkommen[29]. Die ξ_j mit $j \notin \{n_1, \ldots, n_k\}$ können beliebig gewählt werden. Den Lösungsvektor $\xi = (\xi_1, \ldots, \xi_n)^\mathsf{T}$ von $A\xi = \beta$ können wir in folgender Form schreiben:

$$\xi = \alpha_0 + \sum_{\substack{j=1 \\ j \notin \{n_1, \ldots, n_k\}}}^{n} \xi_j \alpha_j.$$

Für jede Wahl der ξ_j, $j \notin \{n_1, \ldots, n_k\}$, ist ξ eine Lösung. Hierbei sind die Spaltenvektoren α_j wie folgt definiert:

[29] $\tilde{A} = (\tilde{\alpha}_{ij})_{i,j=1,\ldots,n}$.

$$\alpha_0 = (0,\dots,0,\underset{\substack{|\\ n_1\text{-te}}}{\tilde\beta_1},0,\dots,0,\underset{\substack{|\\ n_2\text{-te}}}{\tilde\beta_2},0,\dots\dots,0,\underset{\substack{|\\ n_k\text{-te}}}{\tilde\beta_k},0,\dots,0)^{\mathsf{T}}, \quad \text{Stelle}$$

$$\alpha_j = (0,\dots,0,\underset{\substack{|\\ j\text{-te}}}{1},0,\dots\dots\dots,0)^{\mathsf{T}} \quad \text{Stelle} \quad \text{für } j = 1,\dots,n_1-1,$$

$$\alpha_j = (0,\dots,0,\underset{\substack{|\\ n_1\text{-te}}}{-\tilde\alpha_{1,j}},0,\dots,0,\underset{\substack{|\\ j\text{-te}}}{1},0,\dots,0)^{\mathsf{T}} \quad \text{Stelle} \quad \text{für } j = n_1+1,\dots,n_2-1,$$

$$\alpha_j = (0,\dots,0,\underset{\substack{|\\ n_1\text{-te}}}{-\tilde\alpha_{1,j}},0,\dots,0,\underset{\substack{|\\ n_2\text{-te}}}{-\tilde\alpha_{2,j}},0,\dots,0,\underset{\substack{|\\ j\text{-te}}}{1},0,\dots,0)^{\mathsf{T}} \quad \text{Stelle}$$
$$\text{für } j = n_2+1,\dots,n_3-1,$$

$$\vdots$$

$$\alpha_j = (0,\dots,0,\underset{\substack{|\\ n_1\text{-te}}}{-\tilde\alpha_{1,j}},0,\dots,0,\underset{\substack{|\\ n_2\text{-te}}}{-\tilde\alpha_{2,j}},0,\dots\dots,0,\underset{\substack{|\\ n_k\text{-te}}}{-\tilde\alpha_{k,j}},0,\dots,0,\underset{\substack{|\\ j\text{-te}}}{1},0,\dots,0)^{\mathsf{T}} \quad \text{Stelle}$$
$$\text{für } j = n_k+1,\dots,n.$$

α_0 ist jene Lösung von $A\xi = \beta$, welche man für $\xi_j = 0$, $j \notin \{n_1,\dots,n_k\}$, erhält, während die Spaltenvektoren

$$\alpha_1,\dots,\alpha_{n_1-1},\alpha_{n_1+1},\dots,\alpha_{n_2-1},\alpha_{n_2+1},\dots\dots\alpha_{n_k-1},\alpha_{n_k+1},\dots,\alpha_n$$

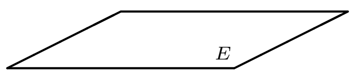
eine Basis des Lösungsraumes für die homogene Gleichung $A\xi = 0$ bilden. In der oben gegebenen Darstellung von ξ spiegelt sich das Resultat von Satz 6.1 wieder.

Abschließend sei noch eine geometrische Interpretation für den Fall $n = 3$ und $\mathbb{K} = \mathbb{R}$ gegeben. Eine Gleichung

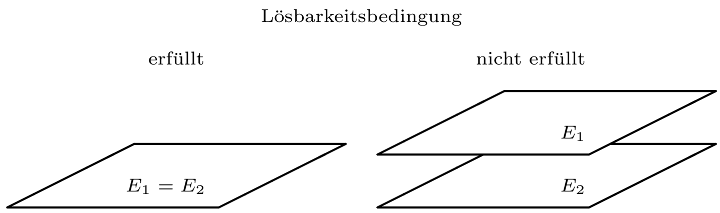
$$\alpha_{i,1}\xi_1 + \alpha_{i,2}\xi_2 + \alpha_{i,3}\xi_3 = \beta_i$$

können wir als Gleichung für die affinen Koordinaten der Punkte einer Ebene im $\mathbb{R}^3$ auffassen, falls $(\alpha_{i,1},\alpha_{i,2},\alpha_{i,3}) \neq (0,0,0)$ ist. Letzteres wollen wir für das weitere voraussetzen.

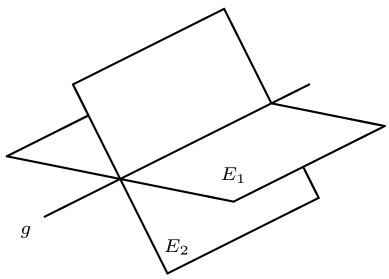
Fall 1: $r = 1$.
Das Gleichungssystem besteht in diesem Fall aus einer Gleichung, die eine Ebene E beschreibt. Die affinen Koordinaten eines jeden Punktes von E bilden eine Lösung.



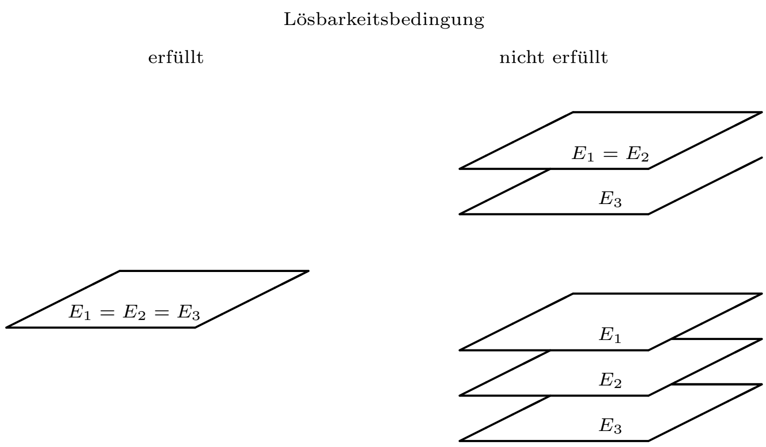
Fall 2: $r = 2$.
Die Gleichungen des Systems beschreiben zwei Ebenen E_1, E_2.
a) rang $A = 1$. Die Ebenen E_1, E_2 sind parallel. Ist die Lösbarkeitsbedingung erfüllt, so gilt $E_1 = E_2$ und jeder Punkt dieser Ebene gibt eine Lösung.



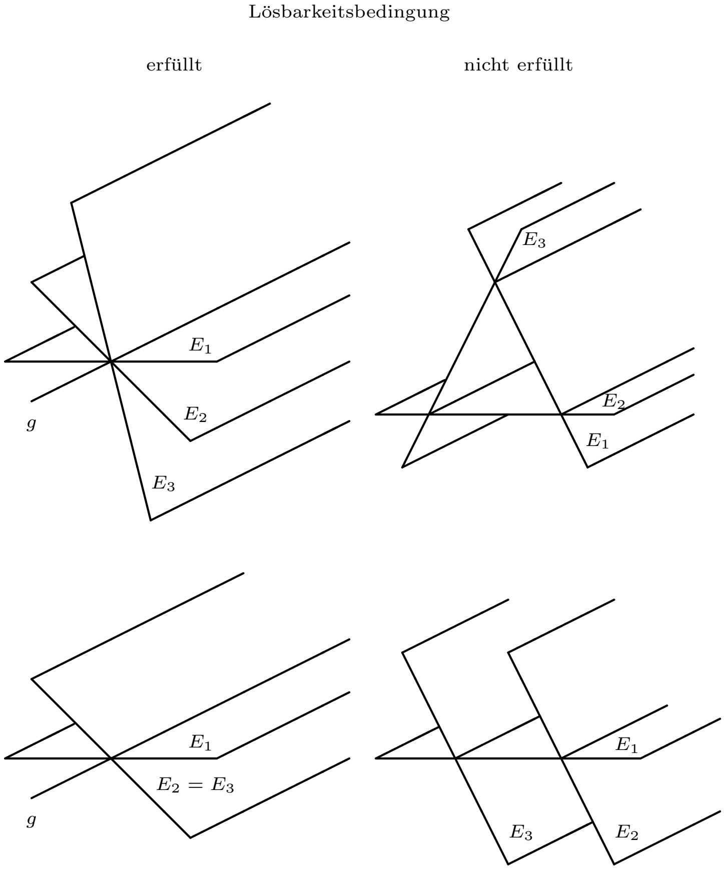
b) rang $A = 2$. Die Ebenen E_1, E_2 sind nicht parallel. Die affinen Koordinaten jedes Punktes der Schnittgeraden ergeben eine Lösung.



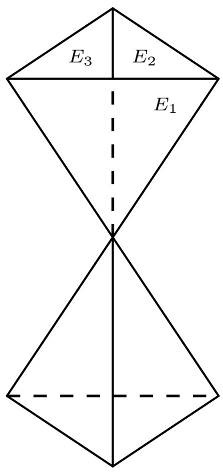
Fall 3: $r = 3$.
Die Gleichungen des Systems beschreiben drei Ebenen E_1, E_2 und E_3.
a) rang $A = 1$. Die Ebenen E_1, E_2, E_3 sind parallel. Ist die Lösbarkeitsbedingung erfüllt, so gilt $E_1 = E_2 = E_3$ und jeder Punkt dieser Ebene ergibt eine Lösung.



b) rang $A = 2$. Von den drei Ebenen sind mindestens zwei – etwa E_1 und E_2 – nicht parallel. Ist die Lösbarkeitsbedingung erfüllt, so schneiden sich die drei Ebenen längs einer Geraden. Jeder Punkt dieser Geraden ergibt eine Lösung.



c) $\operatorname{rang} A = 3$. Die drei Ebenen schneiden sich in einem Punkt. Die affinen Koordinaten dieses Punktes bilden die eindeutig bestimmte Lösung des Systems.



Als Übung diskutiere man den Fall $r = 4$.

Kapitel 7

Matrixalgebra

7.1 Die Algebra der Endomorphismen eines linearen Raumes

Es seien $X, Y, Z, \ldots$ im folgenden stets Vektorräume mit demselben Koeffizientenkörper $\mathbb{K}$. Wir wissen bereits, dass $L(X, Y)$ ein Vektorraum über $\mathbb{K}$ ist (Abschnitt 5.2). Für lineare Abbildungen $\varphi \in L(X, Y)$ und $\psi \in L(Y, Z)$ existiert – wie für beliebige Abbildungen – das Kompositum $\psi \circ \varphi : X \to Z$ (siehe Definition 1.21). Für lineare Abbildungen werden wir im Folgenden $\psi\varphi$ an Stelle von $\psi \circ \varphi$ schreiben. Aus $(\psi\varphi)(\alpha x + \beta y) = \psi(\varphi(\alpha x + \beta y)) = \psi(\alpha\varphi(x) + \beta\varphi(y)) = \alpha\psi(\varphi(x)) + \beta\psi(\varphi(y)) = \alpha(\psi\varphi)(x) + \beta(\psi\varphi)(y)$ für beliebige $x, y \in X$ und $\alpha, \beta \in \mathbb{K}$ folgt sofort $\psi\varphi \in L(X, Z)$.

Für die Komposition von linearen Abbildungen gelten die folgenden einfachen Aussagen:

Proposition 7.1. a) *Es seien $\varphi \in L(X, Y)$, $\psi \in L(Y, Z)$ gegeben. Dann gilt:*

$$\alpha(\psi\varphi) = (\alpha\psi)\varphi = \psi(\alpha\varphi) \quad \textit{für alle } \alpha \in \mathbb{K}.$$

b) *Für $\varphi \in L(X, Y)$, $\psi \in L(Y, Z)$ und $\chi \in L(Z, U)$ gilt*

$$\chi(\psi\varphi) = (\chi\psi)\varphi.$$

c) *Für $\varphi, \psi \in L(X, Y)$ und $\chi \in L(Y, Z)$ gilt*

$$\chi(\varphi + \psi) = \chi\varphi + \chi\psi.$$

d) *Für $\chi \in L(X, Y)$ und $\varphi, \psi \in L(Y, Z)$ gilt*

$$(\varphi + \psi)\chi = \varphi\chi + \psi\chi.$$

Beweis. a) Für den Beweis von Aussage a) wählen wir $x \in X$ beliebig. Dann gelten $\big((\alpha\psi)\varphi\big)(x) = (\alpha\psi)(\varphi(x)) = \alpha(\psi\varphi)(x)$ und $\big(\psi(\alpha\varphi)\big)(x) = \psi((\alpha\varphi(x)) = \psi(\alpha\varphi(x)) = \alpha(\psi(\varphi(x))) = \alpha(\psi\varphi(x))$

b) Die Aussage b) gilt für beliebige Abbildungen (siehe Satz 1.22, a)).

c) Für den Beweis von c) wählen wir $x \in X$ beliebig. Dann gilt:

$$\begin{aligned}\big(\chi(\varphi + \psi)\big)(x) &= \chi\big((\varphi + \psi)(x)\big) = \chi\big(\varphi(x) + \psi(x)\big)\\ &= \chi\big(\varphi(x)\big) + \chi\big(\psi(x)\big) = (\chi\varphi)(x) + (\chi\psi)(x)\\ &= (\chi\varphi + \chi\psi)(x),\end{aligned}$$

d.h. $\chi(\varphi + \psi) = \chi\varphi + \chi\psi$. Der Beweis für Aussage d) ist analog. $\square$

Wir erinnern daran, dass zu einer bijektiven Abbildung $\varphi : X \to Y$ die Umkehrabbildung $\varphi^{-1} : Y \to X$ existiert und eindeutig durch $\varphi\varphi^{-1} = id_Y$ und $\varphi^{-1}\varphi = id_X$ charakterisiert ist (siehe Satz 1.23, e)).

Proposition 7.2. *Es sei $\varphi \in L(X, Y)$ bijektiv. Dann gilt: $\varphi^{-1} \in L(Y, X)$.*

Beweis. Es seien $u, v \in Y$ und $\alpha, \beta \in \mathbb{K}$ gewählt. Wir setzen $x = \varphi^{-1}(u)$, $y = \varphi^{-1}(v)$, d.h. $u = \varphi(x)$ und $v = \varphi(y)$. Dann gilt $\varphi(\alpha x + \beta y) = \alpha\varphi(x) + \beta\varphi(y) = \alpha u + \beta v$. Dies bedeutet aber $\varphi^{-1}(\alpha u + \beta v) = \alpha x + \beta y = \alpha\varphi^{-1}(u) + \beta\varphi^{-1}(v)$. □

Bijektive Abbildungen zwischen endlich-dimensionalen Räumen können einfach charakterisiert werden:

Satz 7.3. *Es sei $\dim X < \infty$, $\dim Y < \infty$ und $\varphi \in L(X, Y)$. Dann ist φ genau dann bijektiv, wenn $\operatorname{rang} \varphi = \dim X = \dim Y$.*

Beweis. φ bijektiv ist gleichbedeutend mit φ surjektiv und φ injektiv. Letzteres ist äquivalent zu $\varphi(X) = Y$ und $\ker \varphi = \{0\}$, d.h. zu $\operatorname{rang} \varphi = \dim Y$ und $0 = \operatorname{def} \varphi = \dim X - \operatorname{rang} \varphi$. □

Eine wichtige Abschätzung für den Rang des Kompositums zweier linearer Abbildungen gibt der folgende Satz:

Satz 7.4 (Sylvester[30]). *Es seien X, Y, Z Vektorräume über $\mathbb{K}$ mit $\dim Y < \infty$. Für beliebige Abbildungen $\varphi \in L(X, Y)$, $\psi \in L(Y, Z)$ gilt:*

$$\operatorname{rang} \varphi + \operatorname{rang} \psi - \dim Y \le \operatorname{rang} \psi\varphi \le \min(\operatorname{rang} \varphi, \operatorname{rang} \psi).$$

Beweis. 1. Es ist $\operatorname{rang} \psi\varphi = \dim(\psi\varphi)(X) = \dim \psi(\varphi(X))$. Sind U, V lineare Räume und ist χ eine lineare Abbildung $U \to V$, so gilt $\operatorname{rang} \chi = \dim \chi(U) \le \dim U$ (Satz 5.12). Damit gilt weiter (wenn wir $U = \varphi(X)$, $V = Z$ und $\chi = \psi\,|_{\varphi(X)}$ wählen)

$$\operatorname{rang} \psi\varphi \le \dim \varphi(X) = \operatorname{rang} \varphi.$$

Aus $\varphi(X) \subset Y$ folgt $\psi(\varphi(X)) \subset \psi(Y)$ und weiter

$$\operatorname{rang} \psi\varphi = \dim \psi(\varphi(X)) \le \dim \psi(Y) = \operatorname{rang} \psi.$$

Damit ist für $\operatorname{rang} \psi\varphi$ die Abschätzung nach oben bewiesen.
2. Wir setzen $\tilde{\psi} = \psi\,|_{\varphi(X)}$, d.h. $\tilde{\psi} \in L(\varphi(X), Z)$ und $\tilde{\psi}(y) = \psi(y)$ für $y \in \varphi(X)$. Dann gilt

$$\ker \tilde{\psi} \subset \ker \psi \quad \text{und} \quad \tilde{\psi}(\varphi(X)) = \psi(\varphi(X)) = (\psi\varphi)(X),$$

woraus

$$\operatorname{def} \tilde{\psi} \le \operatorname{def} \psi \quad \text{und} \quad \operatorname{rang} \tilde{\psi} = \operatorname{rang} \psi\varphi$$

folgt. Nach Satz 5.12 gilt $\operatorname{rang} \tilde{\psi} + \operatorname{def} \tilde{\psi} = \dim \varphi(X)$, d.h.

$$\begin{aligned} \operatorname{rang} \psi\varphi &= \operatorname{rang} \tilde{\psi} = \dim \varphi(X) - \operatorname{def} \tilde{\psi} \ge \operatorname{rang} \varphi - \operatorname{def} \psi \\ &= \operatorname{rang} \varphi - (\dim Y - \operatorname{rang} \psi) = \operatorname{rang} \varphi + \operatorname{rang} \psi - \dim Y. \end{aligned}$$

[30] Sylvester, James Joseph, 3. 9. 1814 (London) – 15. 3. 1897 (London), Beiträge zur Matrizenrechnung (www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Sylvester.html).

Man beachte, dass bei den Umformungen in den letzten zwei Zeilen $\dim Y < \infty$ verwendet wurde ($\dim Y < \infty$ bedeutet insbesondere auch $\operatorname{def} \tilde{\psi} < \infty$ und $\operatorname{rang} \psi < \infty$). □

Für den Rest dieses Abschnittes betrachten wir Endomorphismen eines linearen Raumes X.

Definition 7.5. *Z sei ein linearer Raum über $\mathbb{K}$. Z heißt genau dann eine **Algebra** über $\mathbb{K}$, wenn auf Z zusätzlich zur skalaren Multiplikation und Vektoraddition eine weitere innere Verknüpfung (im folgenden multiplikativ geschrieben) mit den folgenden Eigenschaften erklärt ist:*

(i) $u(vw) = (uv)w$ *für alle* $u, v, w \in Z$,

(ii) $u(v + w) = uv + uw$ *und* $(u + v)w = uw + vw$ *für alle* $u, v, w \in Z$,

(iii) $\alpha(uv) = (\alpha u)v = u(\alpha v)$ *für alle* $u, v \in Z$ *und* $\alpha \in \mathbb{K}$.

*Z heißt **Algebra mit Einselement**, wenn zusätzlich gilt: Es existiert ein $e \in Z$ mit $eu = ue = u$ für alle $u \in Z$.*

Die zu Beginn dieses Abschnittes angegebenen Eigenschaften des Kompositums linearer Abbildungen führen sofort zu

Satz 7.6. *Der lineare Raum $L(X, X)$ der Endomorphismen von X ist eine Algebra mit Einselement über $\mathbb{K}$. Das Einselement in $L(X, X)$ ist die identische Abbildung ε auf X.*

Als Übung beweise man:

Satz 7.7. *Die Automorphismen von X bilden eine Gruppe (mit der Kompositumsbildung als innere Verknüpfung), die sog. **lineare Gruppe** $GL(X)$ auf X.*

Aus Satz 7.4 folgt für Endomorphismen sofort:

Satz 7.8. *Es sei $n = \dim X < \infty$ und $\varphi \in L(X, X)$ sei gegeben. Dann gilt: φ ist genau dann bijektiv, wenn ein $\psi \in L(X, X)$ mit $\psi\varphi = \varepsilon$ existiert. Falls ψ mit $\psi\varphi = \varepsilon$ existiert, so ist $\psi = \varphi^{-1}$.*

Beweis. Ist φ bijektiv, so hat $\psi = \varphi^{-1}$ die verlangten Eigenschaften. Andererseits folgt aus $\psi\varphi = \varepsilon$ und $n = \operatorname{rang} \varepsilon$ wegen Satz 7.4 $\operatorname{rang} \varphi = n$, d.h. φ ist bijektiv. Aus $\psi\varphi = \varepsilon$ folgt sofort $\psi = \psi(\varphi\varphi^{-1}) = (\psi\varphi)\varphi^{-1} = \varphi^{-1}$. □

Im nächsten Abschnitt werden wir sehr einfach sehen, dass es Endomorphismen φ, ψ gibt (im Falle $\dim X \geq 2$) mit $\varphi \neq 0$, $\psi \neq 0$ und $\varphi\psi = 0$, d.h. es gibt ***Nullteiler***. Man kann daher im allgemeinen nicht aus $\varphi\psi = \varphi\chi$ schließen, dass $\psi = \chi$ gilt. Ist jedoch $\operatorname{rang} \varphi = n$ ($= \dim X$), so folgt aus $\varphi\psi = \varphi\chi$ bzw. $\psi\varphi = \chi\varphi$ stets $\psi = \chi$. Wegen der Voraussetzung über φ existiert φ^{-1} und es folgt aus $\varphi\psi = \varphi\chi$ die Gleichung $\varphi^{-1}(\varphi\psi) = \varphi^{-1}(\varphi\chi)$, d.h. $\psi = \chi$.

7.2 Etwas Matrixalgebra

In Abschnitt 5.3 haben wir gesehen, wie im Falle endlich dimensionaler Räume X, Y einer linearen Abbildung $\varphi : X \to Y$ eine Matrix zugeordnet werden kann. Es sei $\dim X = n$, $\dim Y = r$ und $\mathcal{B} = (a_1, \ldots, a_n)$ bzw. $\mathcal{B}^* = (b_1, \ldots, b_r)$ seien geordnete Basen von X bzw. Y. Das Element α_{ij} der Matrix A von φ bzgl. $\mathcal{B}$ und $\mathcal{B}^*$ ist der Koeffizient von ω_{ij} in der Darstellung von φ als Linearkombination der Elemente von $\mathcal{O}$. Daraus folgt unmittelbar unter Beachtung von Satz 4.63:

Proposition 7.9. *$\varphi, \psi \in L(X,Y)$ seien gegeben. Ferner sei $A = (\alpha_{ij}) = M(\varphi; \mathcal{B}, \mathcal{B}^*)$ und $B = (\beta_{ij}) = M(\psi; \mathcal{B}, \mathcal{B}^*)$. Ist $C = (\gamma_{ij}) = M(\varphi + \psi; \mathcal{B}, \mathcal{B}^*)$, so gilt:*

$$\gamma_{ij} = \alpha_{ij} + \beta_{ij}, \quad i = 1, \ldots, r, \ j = 1, \ldots, n. \tag{7.1}$$

Ist $D = (\delta_{ij}) = M(\lambda\varphi; \mathcal{B}, \mathcal{B}^*)$, $\lambda \in \mathbb{K}$, so ist

$$\delta_{ij} = \lambda\alpha_{ij}, \quad i = 1, \ldots, r, \ j = 1, \ldots, n. \tag{7.2}$$

Das eben formulierte Resultat legt die folgende Definition nahe:

> *Sind $r \times n$-Matrizen $A = (\alpha_{ij})$ und $B = (\beta_{ij})$ gegeben, so definieren wir durch (7.1) bzw. (7.2) die Matrix $A + B$ bzw. λA, $\lambda \in \mathbb{K}$.*

Mit $M_{r,n}(\mathbb{K})$ bezeichnen wir die Menge aller $r \times n$-Matrizen über $\mathbb{K}$. Mit der eben definierten Addition und skalaren Multiplikation ist $M_{r,n}(\mathbb{K})$ ein linearer Raum über $\mathbb{K}$, der zu $L(X,Y)$ isomorph ist, falls $\dim X = n$, $\dim Y = r$ (Beweis als Übung).

Im folgenden werden wir häufig Ergebnisse für Matrizen aus Ergebnissen über lineare Abbildungen ableiten. Dazu ist das folgende Resultat wichtig:

> *Ist A eine $r \times n$-Matrix über $\mathbb{K}$, so ist A die Matrix der linearen Abbildung $\varphi_A : \mathbb{K}^n \to \mathbb{K}^r$ bezüglich der kanonischen Basen von $\mathbb{K}^n$ bzw. $\mathbb{K}^r$, welche durch*
>
> $$\varphi_A(\xi) = A\xi, \quad \xi \in \mathbb{K}^n,$$
>
> *definiert ist.*

Bezüglich des Kompositums linearer Abbildungen und der zugehörigen Matrizen gilt:

Satz 7.10. *Es sei* $\dim X = n$, $\dim Y = r$ *und* $\dim Z = s$. *Ferner seien in X, Y, Z je eine Basis $\mathcal{B}$, $\mathcal{B}^*$ bzw. $\mathcal{B}^{**}$ fixiert. Für lineare Abbildungen $\varphi \in L(X,Y)$, $\psi \in L(Y,Z)$ seien $A = (\alpha_{ij}) = M(\varphi; \mathcal{B}, \mathcal{B}^*)$ bzw. $B = (\beta_{ij}) = M(\psi; \mathcal{B}^*, \mathcal{B}^{**})$. Für die Matrix $C = (\gamma_{ij}) = M(\psi\varphi; \mathcal{B}, \mathcal{B}^{**})$ gilt:*

$$\gamma_{ij} = \sum_{k=1}^{r} \beta_{ik}\alpha_{kj}, \quad i = 1, \ldots, s, \ j = 1, \ldots, n. \tag{7.3}$$

Beweis. Es sei $\mathcal{B} = (a_1, \ldots, a_n)$, $\mathcal{B}^* = (b_1, \ldots, b_r)$ und $\mathcal{B}^{**} = (c_1, \ldots, c_s)$. Dann ist

$$(\psi\varphi)(a_j) = \gamma_{1j}c_1 + \gamma_{2j}c_2 + \cdots + \gamma_{sj}c_s,$$

$j = 1, \ldots, n$. Andererseits ist

$$\begin{aligned}
(\psi\varphi)(a_j) &= \psi(\varphi(a_j)) = \psi(\alpha_{1j}b_1 + \cdots + \alpha_{rj}b_r) \\
&= \alpha_{1j}\psi(b_1) + \cdots + \alpha_{rj}\psi(b_r) \\
&= \alpha_{1j}(\beta_{11}c_1 + \cdots + \beta_{s1}c_s) + \cdots + \alpha_{rj}(\beta_{1r}c_1 + \cdots + \beta_{sr}c_s) \\
&= (\beta_{11}\alpha_{1j} + \cdots + \beta_{1r}\alpha_{rj})c_1 + \cdots + (\beta_{s1}\alpha_{1j} + \cdots + \beta_{sr}\alpha_{rj})c_s.
\end{aligned}$$

Auf Grund der Eindeutigkeit der Basisdarstellung folgt

$$\gamma_{ij} = \sum_{k=1}^{r} \beta_{ik}\alpha_{kj}, \quad i = 1, \ldots, s, \ j = 1, \ldots, n.$$

□

In Hinblick auf den eben bewiesenen Satz definieren wir:

*Für eine $s \times r$-Matrix B und eine $r \times n$-Matrix A definieren wir die **Produktmatrix** $C = BA$ durch die Beziehung (7.3), d.h. das Element γ_{ij} in C ist das „Skalarprodukt" der i-ten Zeile von B mit der j-ten Spalte von A. C ist eine $s \times n$-Matrix.*

Die Definition des Produktes zweier Matrizen ist gerade so gewählt, dass dem Kompositum von linearen Abbildungen das Produkt der zugehörigen Matrizen (bezüglich fest gewählter Basen) entspricht. Es ist nun leicht, die in Abschnitt 7.1 erzielten Ergebnisse auf Matrizen zu übertragen (im Falle endlich dimensionaler Räume). Beispielsweise erhält man aus Satz 7.4:

Es sei B bzw. A eine $s \times r$- bzw. $r \times n$-Matrix über $\mathbb{K}$. Dann gilt die Sylvestersche Ungleichung:

$$\operatorname{rang} A + \operatorname{rang} B - r \leq \operatorname{rang} BA \leq \min(\operatorname{rang} A, \operatorname{rang} B). \tag{7.4}$$

Beweis. Wir können B bzw. A als Matrizen von linearen Abbildungen $\varphi_B : \mathbb{K}^r \to \mathbb{K}^s$ bzw. $\varphi_A : \mathbb{K}^n \to \mathbb{K}^r$ auffassen (jeweils bezüglich der kanonischen Basis in den einzelnen Räumen). Da $\operatorname{rang} A = \operatorname{rang} \varphi_A$ und $\operatorname{rang} B = \operatorname{rang} \varphi_B$ gilt (Satz 5.28), folgt das Resultat unmittelbar aus Satz 7.4. □

Wir überlassen es dem Leser, die in Proposition 7.1 unter a), (i), b), c) und d) angegebenen Resultate für Matrizen zu formulieren.

Satz 7.3 bedeutet, dass einer bijektiven linearen Abbildung immer eine $n \times n$-Matrix A mit $\operatorname{rang} A = n$ ($n = \dim X = \dim Y$) zugeordnet ist. Es sei $\varphi \in L(X,Y)$ bijektiv, $\dim X = \dim Y = n$ und A die Matrix von φ bzgl. einer Basis $\mathcal{B}$ von X und einer Basis $\mathcal{B}^*$ von Y. Ferner sei $A' = M(\varphi^{-1}; \mathcal{B}^*, \mathcal{B})$. Aus $\varphi\varphi^{-1} = id_Y$ und $\varphi^{-1}\varphi = id_X$ folgt sofort $AA' = E = A'A$. Hierbei ist E die Matrix von id_X bzw. id_Y bzgl. der Basis $\mathcal{B}$ von X bzw. der Basis $\mathcal{B}^*$ von Y. Man sieht sofort

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

E heißt die $n \times n$-***Einheitsmatrix***. Ist G eine beliebige $r \times n$-Matrix und F eine beliebige $n \times m$-Matrix, so gilt

$$GE = G \quad \text{und} \quad EF = F.$$

Statt A' schreiben wir A^{-1}. A^{-1} heißt die zu A ***inverse Matrix*** und ist durch

$$AA^{-1} = E$$

charakterisiert. Ist $\tilde{A}$ eine weitere $n \times n$-Matrix mit $A\tilde{A} = E$, so erhält man $A^{-1} = A^{-1}(A\tilde{A}) = (A^{-1}A)\tilde{A} = E\tilde{A} = \tilde{A}$.

Ist φ ein Endomorphismus auf X, so wird φ bei Wahl einer Basis $\mathcal{B}$ von X einer $n \times n$-Matrix A zugeordnet, $A = M(\varphi; \mathcal{B})$. Es ist zu beachten, dass sowohl für die Urbilder als auch für die Bilder dieselbe Basis $\mathcal{B}$ verwendet wird (d.h. in der j-ten Spalte von A steht der Koordinatenvektor von $\varphi(a_j)$ bzgl. $\mathcal{B}$, a_j der j-te Vektor von $\mathcal{B}$). Aus Satz 7.6 erhält man sofort: $M_{n,n}(\mathbb{K})$ mit der Matrixmultiplikation ist eine Algebra mit Einselement E über $\mathbb{K}$, welche zu $L(X,X)$ mit $\dim X = n$ isomorph ist. Wählt man eine Basis von X und definiert man die Abbildung $\Phi : L(X,X) \to M_{n,n}(\mathbb{K})$ durch $\Phi(\varphi) = M(\varphi; \mathcal{B})$, so ist Φ ein Isomorphismus (d.h. Φ ist bijektiv und es gilt $\Phi(\alpha\varphi + \beta\psi) = \alpha\Phi(\varphi) + \beta\Phi(\psi)$, $\Phi(\varphi\psi) = \Phi(\varphi)\Phi(\psi)$). Der Beweis wird als nützliche Übung dem Leser überlassen.

Da den Automorphismen auf X die invertierbaren $n \times n$-Matrizen entsprechen, erhält man aus Satz 7.7 sofort: Die invertierbaren $n \times n$-Matrizen über $\mathbb{K}$ bilden bzgl. der Matrixmultiplikation eine Gruppe. Der Beziehung $(\varphi\psi)^{-1} = \psi^{-1}\varphi^{-1}$ für Automorphismen entspricht

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

für invertierbare Matrizen $A, B \in M_{n,n}(\mathbb{K})$. Ferner gilt:

Eine $n \times n$-Matrix A ist genau dann invertierbar, wenn $\operatorname{rang} A = n$ *gilt.*

Beweis. A und A^{-1} sind Abbildungen φ_A bzw. $\varphi_{A^{-1}}$ von $\mathbb{K}^n \to \mathbb{K}^n$ zugeordnet. Die Beziehung $A^{-1}A = E = AA^{-1}$ ist äquivalent zu $\varphi_{A^{-1}}\varphi_A = \varepsilon = \varphi_{A^{-1}}\varphi_A$, d.h. φ_A ist bijektiv (mit $(\varphi_A)^{-1} = \varphi_{A^{-1}}$). Dies ist wiederum äquivalent zu $\operatorname{rang} A = \operatorname{rang} \varphi_A = n$. □

Eine $n \times n$-Matrix heißt ***regulär***, wenn sie invertierbar ist, sonst ***singulär***. Aus Satz 7.8 folgt:

Eine $n \times n$-Matrix A ist genau dann invertierbar, wenn eine $n \times n$-Matrix X mit

$$AX = E \tag{7.5}$$

existiert. Ist dies der Fall, so ist $X = A^{-1}$.

Die Gleichung (7.5) ist der Ausgangspunkt für einen effektiven Algorithmus zur Berechnung von A^{-1}. Es seien x_j, $j = 1, \ldots, n$, die Spalten der gesuchten Matrix X. Dann ist (7.5) äquivalent mit

$$Ax_j = e_j, \quad j = 1, \ldots, n, \tag{7.6}$$

wobei e_j die j-te Spalte von E bezeichnet. Wir haben somit n lineare Gleichungssysteme mit derselben Koeffizientenmatrix zu lösen. Da diese Systeme genau dann alle lösbar sind, wenn A invertierbar ist, d.h. $\operatorname{rang} A = n$ gilt, führt der Staffelalgorithmus die Matrix A in die Matrix E über. Führen wir daher den Staffelalgorithmus für die Matrix (A, e_j) durch, so erhalten wir (E, x_j), $j = 1, \ldots, n$. Damit erhalten wir folgendes Resultat:

Die Matrix A ist genau dann invertierbar, wenn der Staffelalgorithmus die Matrix (A, E) in eine Matrix (E, X) überführt. X ist dann die inverse Matrix zu A.

Abschließend seien noch einige Besonderheiten im Zusammenhang mit der Matrixmultiplikation erwähnt:

1. In $M_{n,n}(\mathbb{K})$ (und damit auch in $L(X, X)$) gibt es für $n \geq 2$ Nullteiler, d.h. es gibt Matrizen $A \neq 0$, $B \neq 0$ mit $AB = 0$. Ein Beispiel ist etwa

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Ist $AB = 0$ und $A \neq 0$, $B \neq 0$, so müssen A und B singulär sein. Denn wäre etwa A regulär, so hätte man $0 = A^{-1}(AB) = B$. Ist $AB = 0$ und A (bzw. B) regulär, so folgt $B = 0$ (bzw. $A = 0$).

2. In $M_{n,n}(\mathbb{K})$, $n \geq 2$, gibt es nilpotente Elemente. Eine Matrix A heißt ***nilpotent von der Ordnung*** k wenn $A^j \neq 0$, $j = 0, \ldots, k-1$, und $A^k = 0$ gilt. Die Matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \ldots & \ldots & \ldots & 0 \end{pmatrix} \in M_{n,n}(\mathbb{K})$$

ist beispielsweise nilpotent von der Ordnung n.

3. In $M_{n,n}(\mathbb{K})$, $n \geq 2$, gibt es Matrizen $A \neq E$ mit $A^2 = E$. A heißt dann ***involutorisch***. Gleichbedeutend ist $A = A^{-1}$. Ein Beispiel ist

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

4. Matrizen A, B in $M_{n,n}(\mathbb{K})$ mit $AB = BA$ heißen ***vertauschbar***. Je zwei Diagonalmatrizen $A = \operatorname{diag}(\alpha_1, \ldots, \alpha_n)$, $B = \operatorname{diag}(\beta_1, \ldots, \beta_n)$ sind vertauschbar, $AB = \operatorname{diag}(\alpha_1\beta_1, \ldots, \alpha_n\beta_n) = BA$. Hierbei bezeichnet $\operatorname{diag}(\gamma_1, \ldots, \gamma_n)$ die Matrix mit den Elementen $\gamma_1, \ldots, \gamma_n$ in der Hauptdiagonalen und Nullen an allen anderen Stellen.

Kapitel 8

Äquivalente und ähnliche Matrizen

In diesem Abschnitt befassen wir uns mit dem folgenden Problem: Gegeben sei $\varphi \in L(X,Y)$, X, Y endlich-dimensional, und geordnete Basen $\mathcal{B}, \mathcal{B}^*$ von X und Y. A sei die Matrix von φ bzgl. $\mathcal{B}, \mathcal{B}^*$. $\tilde{\mathcal{B}}$ bzw. $\tilde{\mathcal{B}}^*$ seien andere geordnete Basen von X bzw. Y und $\tilde{A}$ sei die Matrix von φ bzgl. $\tilde{\mathcal{B}}, \tilde{\mathcal{B}}^*$. Wie kann man den Übergang von $\mathcal{B}$ zu $\tilde{\mathcal{B}}$ (bzw. $\mathcal{B}^*$ zu $\tilde{\mathcal{B}}^*$) beschreiben und wie kann man dann $\tilde{A}$ aus A berechnen? Wir wollen zunächst den Übergang von einer Basis zu einer anderen untersuchen. Sämtliche Vektorräume in diesem Abschnitt werden endlich-dimensional vorausgesetzt.

Definition 8.1. *X sei ein Vektorraum über $\mathbb{K}$ mit $\dim X = n$, $\mathcal{B} = (a_1, \ldots, a_n)$, $\tilde{\mathcal{B}} = (\tilde{a}_1, \ldots, \tilde{a}_n)$ seien geordnete Basen von X und S eine $n \times n$-Matrix über $\mathbb{K}$.*

S heißt genau dann dem Paar $(\mathcal{B}, \tilde{\mathcal{B}})$ zugeordnet, $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$, wenn die j-te Spalte von S der Koordinatenvektor von $\tilde{a}_j$ bezüglich $\mathcal{B}$ ist, $j = 1, \ldots, n$.

Satz 8.2. *X sei ein Vektorraum über $\mathbb{K}$, $\dim X = n$.*

a) *Die einem Paar $(\mathcal{B}, \tilde{\mathcal{B}})$ von geordneten Basen von X zugeordnete Matrix $Tr(\mathcal{B}, \tilde{\mathcal{B}})$ ist stets regulär.*

b) *Es gilt genau dann $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$, wenn gilt: Für jeden Vektor $x \in X$ gilt $\xi = S\tilde{\xi}$, wobei ξ bzw. $\tilde{\xi}$ der Koordinatenvektor von x bzgl. $\mathcal{B}$ bzw. $\tilde{\mathcal{B}}$ ist.*

c) *Aus $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$ folgt $S^{-1} = Tr(\tilde{\mathcal{B}}, \mathcal{B})$.*

d) *Gilt $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$ und $T = Tr(\tilde{\mathcal{B}}, \hat{\mathcal{B}})$, so gilt $ST = Tr(\mathcal{B}, \hat{\mathcal{B}})$.*

e) *Zu jeder geordneten Basis $\mathcal{B}$ von X und jeder regulären $n \times n$-Matrix S über $\mathbb{K}$ existiert genau eine geordnete Basis $\tilde{\mathcal{B}}$ von X derart, dass $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$.*

Beweis. a) Die Spalten von S sind als Koordinatenvektoren der linear unabhängigen Vektoren $\tilde{a}_1, \ldots, \tilde{a}_n$ (bzgl. $\mathcal{B}$) ebenfalls linear unabhängig (Satz 4.64). Daher ist $\operatorname{rang} S = n$.

b) Es sei zunächst $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$. Wegen Definition 8.1 gilt $(S = (\sigma_{ij}))$

$$\tilde{a}_j = \sigma_{1j} a_1 + \cdots + \sigma_{nj} a_n, \quad j = 1, \ldots, n.$$

Daher gilt für beliebige $x \in X$:

$$\begin{aligned} x &= \tilde{\xi}_1 \tilde{a}_1 + \cdots + \tilde{\xi}_n \tilde{a}_n \\ &= \tilde{\xi}_1(\sigma_{11} a_1 + \cdots + \sigma_{n1} a_n) + \cdots + \tilde{\xi}_n(\sigma_{1n} a_1 + \cdots + \sigma_{nn} a_n) \\ &= (\sigma_{11}\tilde{\xi}_1 + \cdots + \sigma_{1n}\tilde{\xi}_n) a_1 + \cdots + (\sigma_{n1}\xi_1 + \cdots + \sigma_{nn}\tilde{\xi}_n) a_n. \end{aligned}$$

Da andererseits $x = \xi_1 a_1 + \cdots + \xi_n a_n$ gilt, folgt (Satz 4.43, d))

$$\xi_j = \sigma_{j1}\tilde{\xi}_1 + \cdots + \sigma_{jn}\tilde{\xi}_n, \quad j = 1, \ldots, n,$$

d.h. $\xi = S\tilde{\xi}$.

Es sei nun $\xi = S\tilde{\xi}$ für jedes $x \in X$. Setzen wir $x = \tilde{a}_j$, so ist

$$\tilde{\xi} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \\ \\ \\ -j\text{-te Stelle} \\ \\ \\ \\ \end{array}$$

und

$$\xi = S\tilde{\xi} = \begin{pmatrix} \sigma_{1j} \\ \vdots \\ \sigma_{nj} \end{pmatrix},$$

d.h. in der j-ten Spalte von S steht der Koordinatenvektor von $\tilde{a}_j$ bzgl. $\mathcal{B}$. Es gilt somit $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$.

c) Es sei $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$. Wegen b) bedeutet dies $\xi = S\tilde{\xi}$ für jedes $x \in X$ (ξ bzw. $\tilde{\xi}$ der Koordinatenvektor von x bzgl. $\mathcal{B}$ bzw. $\tilde{\mathcal{B}}$). Da S regulär ist, folgt $\tilde{\xi} = S^{-1}\xi$ für jedes $x \in X$. Wegen b) gilt $S^{-1} = Tr(\tilde{\mathcal{B}}, \mathcal{B})$.

d) Für $x \in X$ seien $\xi, \tilde{\xi}$ und $\hat{\xi}$ die Koordinatenvektoren bzgl. $\mathcal{B}, \tilde{\mathcal{B}}$ bzw. $\hat{\mathcal{B}}$. Dann gilt $\xi = S\tilde{\xi}$ und $\tilde{\xi} = T\hat{\xi}$, d.h. $\xi = ST\hat{\xi}$. Nach b) ist $ST = Tr(\mathcal{B}, \hat{\mathcal{B}})$.

e) Die Basis $\tilde{\mathcal{B}} = (\tilde{a}_1, \ldots, \tilde{a}_n)$ ist durch

$$\tilde{a}_j = \sigma_{1j}a_1 + \cdots + \sigma_{nj}a_n, \quad j = 1, \ldots, n,$$

gegeben. □

Wir können uns jetzt dem zu Beginn dieses Abschnittes gestellten Problem zuwenden:

Satz 8.3. *Es sei* $\dim X = n$ *und* $\dim Y = r$. $\mathcal{B}, \tilde{\mathcal{B}}$ *bzw.* $\mathcal{B}^*, \tilde{\mathcal{B}}^*$ *seien geordnete Basen von* X *bzw.* Y *und es gelte* $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$ *bzw.* $T = Tr(\mathcal{B}^*, \tilde{\mathcal{B}}^*)$. *Für eine lineare Abbildung* $\varphi \in L(X, Y)$ *sei* $A = M(\varphi; \mathcal{B}, \mathcal{B}^*)$ *und* $\tilde{A} = M(\varphi; \tilde{\mathcal{B}}, \tilde{\mathcal{B}}^*)$. *Dann gilt*

$$\tilde{A} = T^{-1}AS.$$

Beweis. Für $x \in X$ sei ξ bzw. $\tilde{\xi}$ der Koordinatenvektor von x bzgl. $\mathcal{B}$ bzw. $\tilde{\mathcal{B}}$ und η bzw. $\tilde{\eta}$ der Koordinatenvektor von $\varphi(x)$ bzgl. $\mathcal{B}^*$ bzw. $\tilde{\mathcal{B}}^*$. Wegen Satz 8.2 gilt

$$\xi = S\tilde{\xi} \quad \text{und} \quad \eta = T\tilde{\eta}.$$

Für die Matrizen A und $\tilde{A}$ gilt (Satz 5.18):

$$\eta = A\xi \quad \text{und} \quad \tilde{\eta} = \tilde{A}\tilde{\xi} \quad \text{für alle } x \in X.$$

Daraus folgt
$$\tilde{\eta} = T^{-1}AS\tilde{\xi} \quad \text{für alle } x \in X.$$
Daher muß $\tilde{A} = T^{-1}AS$ gelten (Korollar zu Satz 5.18). $\square$

Für Endomorphismen gilt analog

Satz 8.4. *Es sei* $\dim X = n$ *und* $\mathcal{B}$ *bzw.* $\tilde{\mathcal{B}}$ *seien geordnete Basen von* X. *Ferner sei* $S = Tr(\mathcal{B}, \tilde{\mathcal{B}})$. *Für* $\varphi \in L(X, X)$ *sei* $A = M(\varphi; \mathcal{B})$ *und* $\tilde{A} = M(\varphi; \tilde{\mathcal{B}})$. *Dann gilt*
$$\tilde{A} = S^{-1}AS.$$

Beweis. Die Aussage des Satzes folgt sofort aus Satz 8.3, wenn man $\mathcal{B} = \mathcal{B}^*$, $\tilde{\mathcal{B}} = \tilde{\mathcal{B}}^*$ beachtet, woraus $T = S$ folgt. $\square$

Der in den Sätzen 8.3 und 8.4 angegebene Zusammenhang zwischen den Matrizen A und $\tilde{A}$ führt zu folgender Definition:

Definition 8.5. a) *$A, \tilde{A}$ seien Matrizen aus $M_{r,n}(\mathbb{K})$. A und $\tilde{A}$ heißen genau dann* ***äquivalent****, wenn es reguläre Matrizen $R \in M_{r,r}(\mathbb{K})$, $S \in M_{n,n}(\mathbb{K})$ gibt mit*
$$\tilde{A} = RAS.$$
b) *A, $\tilde{A}$ seien Matrizen aus $M_{n,n}(\mathbb{K})$. A und $\tilde{A}$ heißen genau dann* ***ähnlich****, wenn es eine reguläre Matrix $S \in M_{n,n}(\mathbb{K})$ gibt mit*
$$\tilde{A} = S^{-1}AS.$$

Unmittelbar aus der Definition folgt, dass ähnliche Matrizen auch äquivalent sind.

Satz 8.6. a) *Die Äquivalenz von Matrizen ist eine Äquivalenzrelation auf $M_{r,n}(\mathbb{K})$.*
b) *Die Ähnlichkeit von Matrizen ist eine Äquivalenzrelation auf $M_{n,n}(\mathbb{K})$.*

Beweis. Es sei $A \in M_{r,n}(\mathbb{K})$. Dann gilt $A = RAS$ mit $R = E_r$ ($= r \times r$-Einheitsmatrix) und $S = E_n$ ($= n \times n$-Einheitsmatrix). Damit ist die Reflexivität gezeigt. Aus $\tilde{A} = RAS$ mit regulären Matrizen R, S folgt $A = R^{-1}\tilde{A}S^{-1}$, d.h. die Äquivalenz ist symmetrisch. Schließlich folgt aus $A_1 = R_1AS_1$ und $A_2 = R_2A_1S_2$ mit regulären Matrizen R_i, S_i, $i = 1, 2$, auch $A_2 = R_2R_1AS_1S_2$, d.h. A_2 und A sind äquivalent (R_2R_1 und S_1S_2 sind regulär). Damit ist auch die Transitivität nachgewiesen. Völlig analog verläuft der Beweis für die Ähnlichkeit. $\square$

Durch die Äquivalenz bzw. Ähnlichkeit wird eine Klasseneinteilung von $M_{r,n}(\mathbb{K})$ bzw. $M_{n,n}(\mathbb{K})$ gegeben (siehe Kapitel 1). Es ist nun wünschenswert, die Äquivalenzklassen genau zu charakterisieren und für jede Äquivalenzklasse einen möglichst einfachen Repräsentanten anzugeben. Für die Äquivalenz von Matrizen können wir dies mit den uns zur Verfügung stehenden Mitteln ohne Schwierigkeiten durchführen. Der Fall der Ähnlichkeit von Matrizen bereitet weitaus größere Schwierigkeiten und kann vorderhand noch nicht erledigt werden.

Satz 8.7. a) *A und B aus $M_{r,n}(\mathbb{K})$ sind genau dann äquivalent, wenn*
$$\operatorname{rang} A = \operatorname{rang} B$$

gilt.

b) *Es sei* $A \in M_{r,n}(\mathbb{K})$ *mit* $\operatorname{rang} A = k > 0$. *Dann ist* A *äquivalent zu* $I_k \in M_{r,n}(\mathbb{K})$ (*siehe* (5.4) *im Beweis zu Hilfssatz* 5.25). *Im Fall* $\operatorname{rang} A = 0$ *ist* $A = 0$ *und die Äquivalenzklasse von* A *enthält nur die Nullmatrix. Ein vollständiges Repräsentantensystem für die Äquivalenz von Matrizen auf* $M_{r,n}(\mathbb{K})$ *ist somit durch die Nullmatrix und die Matrizen* I_k, $k = 1, \ldots, \min(r, n)$, *gegeben.*

Beweis. Wir beweisen b) zuerst. Die Aussage für $k = 0$ ist klar. Es sei daher $k > 0$. Dann läßt sich A durch elementare Umformungen in die Matrix I_k überführen (siehe den Beweis zu Hilfssatz 5.25). Faßt man A als Matrix der linearen Abbildung $\varphi_A \in L(\mathbb{K}^n, \mathbb{K}^r)$ auf (siehe Abschnitt 7.2), so entsprechen den elementaren Umformungen in den Zeilen bzw. Spalten von A elementare Umformungen der Basen von $\mathbb{K}^r$ bzw. $\mathbb{K}^n$ (siehe Satz 5.22). Die Matrix I_k ist daher die Matrix von φ_A bezüglich anderer geordneter Basen von $\mathbb{K}^n$ und $\mathbb{K}^r$. Nach Satz 8.3 sind A und I_k äquivalent. Damit ist b) bewiesen.

Es sei nun $\operatorname{rang} A = \operatorname{rang} B = k$. Der Fall $k = 0$ ist wegen $A = B = 0$ trivial. Für $k > 0$ sind nach b) A und B äquivalent zu I_k und daher auch äquivalent zueinander. Sind andererseits die Matrizen $A, B \in M_{r,n}(\mathbb{K})$ äquivalent, so gilt $B = RAS$ mit regulären Matrizen $R \in M_{r,r}(\mathbb{K})$ und $S \in M_{n,n}(\mathbb{K})$. Aus der Sylvester'schen Ungleichung (siehe Satz 7.4 bzw. (7.4)) folgt unter Beachtung von $\operatorname{rang} S = n$ und $\min(\operatorname{rang} A, \operatorname{rang} S) = \operatorname{rang} A$

$$\operatorname{rang} AS = \operatorname{rang} A$$

und wegen $\operatorname{rang} R = r$ und $\min(\operatorname{rang} AS, \operatorname{rang} R) = \operatorname{rang} AS$ auch

$$\operatorname{rang} RAS = \operatorname{rang} AS = \operatorname{rang} A.$$

□

Da ähnliche Matrizen immer auch äquivalent sind, gilt: Sind A und B aus $M_{n,n}(\mathbb{K})$ ähnlich so ist $\operatorname{rang} A = \operatorname{rang} B$. Dass andererseits äquivalente $n \times n$-Matrizen nicht ähnlich sein müssen, zeigt das folgende Beispiel:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad B = E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Es ist $\operatorname{rang} A = \operatorname{rang} B = 2$. A und B sind somit äquivalent. Wäre $A = S^{-1}BS$ mit einer regulären 2×2-Matrix S, so hätte man wegen $B = E$ auch $A = E$.

Sind $A, B \in M_{r,n}(\mathbb{K})$ äquivalent, so gilt $B = RAS$ mit regulären Matrizen R, S. Wir wollen noch angeben, wie diese Matrizen berechnet werden können. Zunächst reicht es, reguläre Matrizen R_i, S_i zu berechnen mit ($k = \operatorname{rang} A = \operatorname{rang} B$)

$$I_k = R_1 A S_1, \quad I_k = R_2 B S_2.$$

Dann ist $B = R_2^{-1} R_1 A S_1 S_2^{-1}$, d.h. $R = R_2^{-1} R_1$ und $S = S_1 S_2^{-1}$.

Es sei also $A \in M_{r,n}(\mathbb{K})$ gegeben. Wir haben gesehen (siehe Beweis zu Hilfssatz 5.25), dass A durch elementare Zeilen- bzw. Spaltenumformungen in die Matrix I_k übergeführt werden kann. Diese elementaren Umformungen können wir durch Multiplikation von links bzw. von rechts mit geeigneten Matrizen darstellen. Wir definieren die folgenden $n \times n$-

Matrizen:

$$E_n(i,j) = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & 0 & \\ & & 1 & & & & & & \\ & & & 0 & \cdots & 0 & 1 & & \\ & & & \vdots & 1 & & 0 & & \\ & & & \vdots & & \ddots & \vdots & & \\ & & & 0 & & 1 & \vdots & & \\ & & & 1 & 0 & \cdots & 0 & & \\ & & & & & & & 1 & \\ & 0 & & & & & & & \ddots \\ & & & & & & & & 1 \end{pmatrix}, \quad i,j = 1,\ldots,n,$$

$$E_n(\lambda,i,j) = \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & 0 & \\ & & 1 & 0 & \cdots & 0 & \lambda & \\ & & & \ddots & & & 0 & \\ & & & & \ddots & & \vdots & \\ & & & & & \ddots & 0 & \\ & & & & & & 1 & \\ & 0 & & & & & & \ddots \\ & & & & & & & 1 \end{pmatrix}, \quad \lambda \in \mathbb{K},\ i,j = 1,\ldots,n,\ i \neq j,$$

$$E_n(\lambda,i) = \begin{pmatrix} 1 & & & & \\ & \ddots & & 0 & \\ & & 1 & & \\ & & & \lambda & \\ & & & & 1 & \\ & 0 & & & & \ddots \\ & & & & & & 1 \end{pmatrix}, \quad \lambda \in \mathbb{K},\ \lambda \neq 0,\ i = 1,\ldots,n.$$

In der Matrix $E_n(i,j)$ befinden sich die nicht in der Hauptdiagonalen stehenden Einsen in der i-ten Zeile und j-ten Spalte bzw. in der j-ten Zeile und i-ten Spalte, bei der Matrix $E_n(\lambda,i,j)$ steht λ in der i-ten Zeile und j-ten Spalte und in $E_n(\lambda,i)$ steht λ an der i-ten Stelle in der Hauptdiagonalen. Die Matrizen $E_n(i,j)$, $E_n(\lambda,i,j)$ und $E_n(\lambda,i)$, heißen ***Elementarmatrizen***.

Satz 8.8. a) *Es gilt:*

$$\begin{aligned} E_n(i,j)^{-1} &= E_n(i,j), \quad i,j = 1\ldots,n, \\ E_n(\lambda,i,j)^{-1} &= E_n(-\lambda,i,j), \quad \lambda \in \mathbb{K},\ i,j = 1,\ldots,n,\ i \neq j, \\ E_n(\lambda,i)^{-1} &= E_n(1/\lambda,i), \quad \lambda \in \mathbb{K},\ \lambda \neq 0,\ i = 1,\ldots,n. \end{aligned}$$

b) $A \in M_{r,n}(\mathbb{K})$ *sei gegeben. Das Ergebnis einer Multiplikation von A mit einer Elementarmatrix von links bzw. von rechts ist der folgenden Tabelle zu entnehmen:*

Die Matrix $A' =$	*entsteht aus* A *durch*
$E_r(i,j)A$	*Vertauschung der* i*-ten und* j*-ten Zeile*
$AE_n(i,j)$	*Vertauschung der* i*-ten und* j*-ten Spalte*
$E_r(\lambda,i,j)A$	*Addition des* λ*-fachen der* j*-ten Zeile zur* i*-ten Zeile*
$AE_n(\lambda,i,j)$	*Addition des* λ*-fachen der* i*-ten Spalte zur* j*-ten Spalte*
$E_r(\lambda,i)A$	*Multiplikation der* i*-ten Zeile mit* λ
$AE_n(\lambda,i)$	*Multiplikation der* i*-ten Spalte mit* λ.

Beweis. Man verifiziert die Behauptungen durch einfaches Ausrechnen. □

In Analogie zu Definition 5.21 nennen wir die Vertauschung von zwei Zeilen oder Spalten einer Matrix eine elementare Zeilenumformung oder Spaltenumformung des Typs (I), die Addition eines Vielfachen einer Zeile bzw. Spalte zu einer anderen Zeile bzw. Spalte eine elementare Zeilen- bzw. Spaltenumformung des Typs (II) und schlieẞs lich die Multiplikation einer Zeile bzw. einer Spalte mit einem Faktor $\lambda \in \mathbb{K}$ eine elementare Zeilen- bzw. Spaltenumformung des Typs (III).

Auf Grund von Satz 8.8 ist es einfach, die einzelnen Schritte bei der Umformung von A in I_k durch Multiplikation von links bzw. von rechts mit Elementarmatrizen darzustellen. Man erhält somit die Matrizen R_i, S_i als Produkte von Elementarmatrizen.

Eine nützliche Folgerung aus Satz 8.7, b), ist:

Korollar 8.9. *Jede reguläre* $n \times n$*-Matrix ist als Produkt endlich vieler Elementarmatrizen darstellbar.*

Beweis. Es sei $A \in M_{n,n}(\mathbb{K})$ mit $\operatorname{rang} A = n$ gegeben. Nach Satz 8.7, b), ist A zur Matrix $I_n = E$ äquivalent. Im Beweis zu Hilfssatz 5.25 wurde gezeigt, dass A durch elementare Zeilen- und Spaltenumformungen auf die Gestalt I_n transformiert werden kann. Da $\operatorname{rang} A = n$ ist, leistet dies der Staffelalgorithmus. Dieser benutzt nur elementare Zeilenumformungen, die gemäß Satz 8.8, b), durch Multiplikation von links mit Elementarmatrizen dargestellt werden können. Es gilt daher für Elementarmatrizen $E_1, \dots, E_m$

$$E_m \cdots E_1 A = E$$

woraus $A = E_1^{-1} \cdots E_m^{-1}$ folgt. Nach Satz 8.8, a), ist die inverse Matrix einer Elementarmatrix wieder eine Elementarmatrix. Damit ist der Satz bewiesen. □

Kapitel 9

Determinanten

9.1 Definition und grundlegende Eigenschaften

Für eine Reihe von Problemstellungen in der linearen Algebra spielen spezielle skalarwertige Funktionen der Zeilen (oder Spalten) von $n \times n$-Matrizen eine wichtige Rolle. Für eine Matrix $A \in M_{n,n}(\mathbb{K})$ bezeichnen wir die Zeilen mit $a_1, \dots, a_n$ und schreiben

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Die folgende Definition geht auf Weierstrass[31] zurück:

Definition 9.1. *Eine Abbildung* $\det : M_{n,n}(\mathbb{K}) \to \mathbb{K}$ *heißt* ***Determinante*** *auf* $M_{n,n}(\mathbb{K})$ *genau dann, wenn die folgenden Eigenschaften erfüllt sind:*

(D1) *det ist linear in den Zeilen, d.h. für beliebige* $i \in \{1, \dots, n\}$ *und alle* $\alpha, \beta \in \mathbb{K}$ *gilt*[32]

$$\det \begin{pmatrix} \vdots \\ \alpha a_i + \beta b_i \\ \vdots \end{pmatrix} = \alpha \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix} + \beta \det \begin{pmatrix} \vdots \\ b_i \\ \vdots \end{pmatrix}.$$

(D2) *Ist* $a_i = a_j$ *für zwei Indizes* $i \neq j$*, so gilt*

$$\det A = 0.$$

(D3) *Für die Einheitsmatrix* E *gilt*

$$\det E = 1.$$

In dem folgenden Satz werden die wichtigsten Eigenschaften der Determinante zusammengefasst:

Satz 9.2. *Es sei* $\det$ *eine Determinante auf* $M_{n,n}(\mathbb{K})$*. Dann gilt:*

[31] Weierstrass, Karl Theodor Wilhelm, 31. 10. 1815 (Ostenfelde) – 19. 2. 1897 (Berlin), sein Hauptwerk befasst sich mit der logisch korrekten Fundierung der Analysis und dem Aufbau der Funktionentheorie auf Basis der Potenzreihenentwicklungen, (en.wikipedia.org/wiki/Karl_Weierstrass).

[32] Die Punkte „$\vdots$" in der folgenden Formel bedeuten, dass in den vorkommenden Matrizen die von der i-ten Zeile verschiedenen Zeilen jeweils gleich sind.

(i) $\det(\lambda A) = \lambda^n \det A$ *für alle* $\lambda \in \mathbb{K}$.

(ii) *Ist* $B = E_n(\lambda, i)A$ *mit* $\lambda \in \mathbb{K}$ *und* $i \in \{1, \dots, n\}$, *so gilt*

$$\det B = \lambda \det A.$$

(iii) *Ist* $a_i = 0$ *für ein* $i \in \{1, \dots, n\}$, *so gilt* $\det A = 0$.

(iv) *Ist* $B = E_n(i, j)A$ *mit* $i, j \in \{1, \dots, n\}$ *und* $i \neq j$, *so gilt*

$$\det B = -\det A.$$

(v) *Ist* $B = E_n(\lambda, i, j)A$ *mit* $\lambda \in \mathbb{K}$, $i, j \in \{1, \dots, n\}$ *und* $i \neq j$, *so gilt*

$$\det B = \det A.$$

(vi) *Es sei*

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

Dann gilt $\det A = \lambda_1 \cdot \ldots \cdot \lambda_n$.

(vii) *Es sei* $n \geq 2$, $n = n_1 + n_2$ *mit* $n_1 \geq 1$, $n_2 \geq 1$ *und*

$$A = \begin{pmatrix} A_1 & C \\ 0 & A_2 \end{pmatrix} \quad \textit{mit } A_1 \in M_{n_1,n_1}(\mathbb{K}),\ A_2 \in M_{n_2,n_2}(\mathbb{K}).$$

Dann gilt

$$\det A = \det A_1 \det A_2.$$

(viii) $\det A = 0 \iff \operatorname{rang} A < n$.

(ix) *Für* $A, B \in M_{n,n}(\mathbb{K})$ *gilt:* $\det AB = \det A \det B$.

Ist A *regulär, so gilt:* $\det A^{-1} = \dfrac{1}{\det A}$.

Beweis. 1. Wegen (D1) gilt

$$\det(\alpha A) = \det \begin{pmatrix} \alpha a_1 \\ \vdots \\ \alpha a_n \end{pmatrix} = \alpha \det \begin{pmatrix} a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_n \end{pmatrix} = \cdots = \alpha^n \det \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \alpha^n \det A.$$

2. Die Matrix B entsteht aus A durch Multiplikation der i-ten Zeile mit λ. Das Ergebnis folgt sofort aus (D1).

3. Dies folgt sofort aus (ii) mit $\lambda = 0$.

4. B entsteht aus A durch Vertauschung der i-ten mit der j-ten Zeile. In den folgenden Matrizen sind jeweils die k-ten Zeilen mit $k \neq i$ und $k \neq j$ gleich. Es gilt unter Benutzung

von D(2):

$$\det\begin{pmatrix}\vdots\\a_j\\\vdots\\a_i\\\vdots\end{pmatrix}+\det\begin{pmatrix}\vdots\\a_i\\\vdots\\a_j\\\vdots\end{pmatrix}=\det\begin{pmatrix}\vdots\\a_i\\\vdots\\a_i\\\vdots\end{pmatrix}+\det\begin{pmatrix}\vdots\\a_i\\\vdots\\a_j\\\vdots\end{pmatrix}+\det\begin{pmatrix}\vdots\\a_j\\\vdots\\a_i\\\vdots\end{pmatrix}+\det\begin{pmatrix}\vdots\\a_j\\\vdots\\a_j\\\vdots\end{pmatrix}$$

$$=\det\begin{pmatrix}\vdots\\a_i\\\vdots\\a_i+a_j\\\vdots\end{pmatrix}+\det\begin{pmatrix}\vdots\\a_j\\\vdots\\a_i+a_j\\\vdots\end{pmatrix}=\det\begin{pmatrix}\vdots\\a_i+a_j\\\vdots\\a_i+a_j\\\vdots\end{pmatrix}=0.$$

Hieraus folgt

$$\det B=\det\begin{pmatrix}\vdots\\a_j\\\vdots\\a_i\\\vdots\end{pmatrix}=-\det\begin{pmatrix}\vdots\\a_i\\\vdots\\a_j\\\vdots\end{pmatrix}=-\det A.$$

5. B entsteht aus A durch Addition des λ-fachen der j-ten Zeile zur i-ten Zeile. Wegen (D1) gilt

$$\det B=\det A+\lambda\det\tilde{A},$$

wobei die i-te Zeile von $\tilde{A}$ durch a_j gegeben ist, während alle übrigen Zeilen von $\tilde{A}$ mit jenen von A übereinstimmen. In $\tilde{A}$ sind somit die i-te und die j-te Zeile gleich. Aus (D2) folgt daher $\det\tilde{A}=0$ und somit $\det B=\det A$.

6. Ist eines der Elemente λ_i null, so existiert ein Index $i_0\in\{1,\dots,n\}$ mit $\lambda_{i_0}=0$ und $\lambda_i\neq 0$ für $i=i_0+1,\dots,n$. Im Falle $i_0=n$ ist die n-te Zeile der Matrix eine Nullzeile. Wegen 3. ist daher die Determinante der Matrix null. Ist $i_0>n$, so können durch Addition entsprechender Vielfacher der n-ten Zeile zu den Zeilen $1,\dots,n-1$ die Elemente $\alpha_{i,n}$, $i=1,\dots,n-1$ annulliert werden. Im nächsten Schritt annulliert man die Elemente $\alpha_{i,n-1}$, $i=1,\dots,n-2$ durch Addition entsprechender Vielfacher der $i-1$-sten Zeile zu den Zeilen $a_1,\dots,a_{n-2}$. Auf diese Weise fortfahrend annulliert man die Elemente $\alpha_{i,j}$ mit $i=i_0+1,\dots,n$ und $j=1,\dots,i+1$ durch Zeilenoperationen, die gemäß (v) die Determinante unverändert lassen. Die i_0-te Zeile der erhaltenen Matrix ist eine Nullzeile. Daher gilt $\det A=0=\lambda_1\cdot\ldots\cdot\lambda_n$.

Gilt $\lambda_i\neq 0$, $i=1,\dots,n$, so kann durch Addition geeigneter Vielfacher der Zeilen von A (beginnend mit der n-ten Zeile) die Matrix A zur Diagonalmatrix $B=\operatorname{diag}(\lambda_1,\dots,\lambda_n)$ umgeformt werden. Wegen (v) gilt $\det A=\det B$. Wiederholte Anwendung von (ii) ergibt $\det B=\lambda_1\cdot\ldots\cdot\lambda_n$.

7. Durch elementare Zeilenumformungen des Typs (I) und (II) in den Zeilen $1,\dots,n_1$ wird die Matrix A zu

$$\tilde{A}=\begin{pmatrix}\tilde{A}_1&\tilde{C}\\0&A_2\end{pmatrix}\quad\text{mit}\quad\tilde{A}_1=\begin{pmatrix}\lambda_1&*&\cdots&*\\0&\ddots&\ddots&\vdots\\\vdots&\ddots&\ddots&*\\0&\cdots&0&\lambda_{n_1}\end{pmatrix}$$

umgeformt. Ist hierbei k die Anzahl der elementaren Zeilenumformungen vom Typ (I), so gilt

$$\det A_1=(-1)^k\det\tilde{A}_1=(-1)^k\lambda_1\cdot\ldots\cdot\lambda_{n_1}.$$

Analog erhält man aus $\tilde{A}$ durch elementare Zeilenumformungen der Typen (I) und (II) (darunter ℓ Umformungen des Typs (I)) in den Zeilen $n_1+1,\dots,n$ die Matrix

$$\tilde{\tilde{A}} = \begin{pmatrix} \tilde{A}_1 & \tilde{C} \\ 0 & \tilde{A}_2 \end{pmatrix} \quad \text{mit} \quad \tilde{A}_2 = \begin{pmatrix} \lambda_{n_1+1} & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Für die Matrix A_2 gilt

$$\det A_2 = (-1)^\ell \det \tilde{A}_2 = (-1)^\ell \lambda_{n_1+1} \cdot \ldots \cdot \lambda_n.$$

Insgesamt erhalten wir

$$\det A_1 \det A_2 = (-1)^{k+\ell} \lambda_1 \cdot \ldots \cdot \lambda_n.$$

Andererseits wird die Matrix A durch die verwendeten elementaren Umformungen des Typs (I) und (II) (darunter $k+\ell$ Umformungen des Typs (I)) auf die Matrix $\tilde{\tilde{A}}$ transformiert. Aus (iv) und (vi) folgt

$$\det A = (-1)^{k+\ell} \det \tilde{\tilde{A}} = (-1)^{k+\ell} \lambda_1 \cdot \ldots \cdot \lambda_n = \det A_1 \det A_2.$$

8. Durch elementare Zeilenumformungen des Typs (I) und (II) kann die Matrix A zu

$$\tilde{A} = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

umgeformt werden. Wegen (iv) und (v) gilt

$$\det A = \pm \det \tilde{A} = \lambda_1 \cdot \ldots \cdot \lambda_n.$$

Es ist $\operatorname{rang} A = n$ genau dann, wenn $\lambda_i \neq 0$, $i = 1,\dots,n$, d.h. es gilt $\operatorname{rang} A = n$ genau dann, wenn $\det A \neq 0$.

9. *Fall 1:* $\det A = 0$.
In diesem Fall gilt (wegen (viii)) $\operatorname{rang} A < n$. Die Sylvestersche Ungleichung (7.4) impliziert

$$\operatorname{rang} AB \le \min\bigl(\operatorname{rang} A, \operatorname{rang} B\bigr) < n.$$

Somit gilt wegen (viii) auch $\det AB = 0$.
Fall 2: $\det A \neq 0$.
Da $\operatorname{rang} A = n$ ist, folgt aus Korollar 8.9, dass A Produkt endlich vieler Elementarmatrizen ist:

$$A = E_1 \cdot \ldots \cdot E_m.$$

Es sei $\tilde{E}$ eine $n \times n$-Elementarmatrix und $C \in M_{n,n}(\mathbb{K})$.

a) $\tilde{E} = E_n(i,j)$. In diesem Fall gilt $\det \tilde{E} = -1$ ($\tilde{E}$ entsteht aus der Einheitsmatrix E durch Vertauschung der i-ten und der j-ten Zeile) und $\tilde{E}C$ entsteht aus C durch Vertauschung der i-ten mit der j-ten Zeile (Satz 8.8, b)). Es gilt somit

$$\det \tilde{E}C = -\det C = \det \tilde{E} \det C.$$

b) $\tilde{E} = E_n(\lambda, i, j)$. In diesem Fall gilt $\det \tilde{E} = 1$ und $\tilde{E}C$ entsteht aus C durch Addition des λ-fachen der j-ten Zeile zur i-ten Zeile. Es gilt somit

$$\det \tilde{E}C = \det C = \det \tilde{E} \det C.$$

c) $\tilde{E} = E_n(\lambda, i)$. Es gilt $\det \tilde{E} = \lambda$. $\tilde{E}C$ entsteht aus C durch Multiplikation der i-ten Zeile mit λ. Somit gilt

$$\det \tilde{E}C = \lambda \det C = \det \tilde{E} \det C.$$

Damit ist gezeigt,dass die zu beweisende Aussage gilt, falls der erste Faktor eine Elementarmatrix ist. Damit erhalten wir:

$$\begin{aligned}
\det AB &= \det(E_1 \cdot \ldots \cdot E_n B) = \det E_1 \det(E_2 \cdot \ldots \cdot E_m B) \\
&= \det E_1 \det E_2 \det(E_3 \cdot \ldots \cdot E_m B) = \cdots = \det E_1 \cdot \ldots \cdot \det E_m \det B \\
&= \det E_1 \cdot \ldots \cdot \det E_{m-2} \det(E_{m-1} E_m) \det B \\
&= \det E_1 \cdot \ldots \cdot \det E_{m-3} \det(E_{m-2} E_{m-1} E_m) \det B = \cdots = \det(E_1 \cdot \ldots \cdot E_m) \det B \\
&= \det A \det B.
\end{aligned}$$

Ist die Matrix A regulär, so folgt aus $AA^{-1} = E$ sofort

$$\det A \det A^{-1} = \det AA^{-1} = \det E = 1.$$

□

Ist die Charakteristik des Körpers $\mathbb{K}$ *ungleich* 2, *so folgt aus der Aussage* (iv) *von Satz* 9.2 *die Eigenschaft* (D2).

Beweis. Es seien die i-te und die j-te Zeile, $i \neq j$, der Matrix $A \in M_{n,n}(\mathbb{K})$ gleich. Vertauscht man diese zwei Zeilen, so ändert sich die Matrix nicht. Wegen (iv) gilt jedoch

$$\det A = -\det A,$$

d.h. $0 = \det A + \det A = (1+1) \det A$. Da die Charakteristik des Körpers ungleich 2 ist, gilt $1 + 1 \neq 0$. Es muss daher $\det A = 0$ sein. □

Alle Aussagen über Determinanten und Zeilen einer Matrix gelten in analoger Form auch für Determinanten und die Spalten einer Matrix. Dies folgt unmittelbar aus den folgenden Aussagen.

Definition 9.3. *Es sei* $A = (\alpha_{ij}) \in M_{r,n}(\mathbb{K})$ *gegeben. Die Matrix* $B = (\beta_{ij})$ *mit* $\beta_{ij} = \alpha_{ji}$, $i = 1, \ldots, n$, $j = 1, \ldots, r$, *heißt die zu* A ***transponierte Matrix***, $B = A^{\mathsf{T}}$.

A^{T} entsteht aus A durch Spiegelung an der Hauptdiagonalen. Die i-te Zeile bzw. Spalte von A^{T} ist die i-te Spalte bzw. Zeile von A.

Satz 9.4. *Es sei* $A \in M_{n,n}(\mathbb{K})$ *gegeben. Dann gilt:*

$$\det A = \det A^{\mathsf{T}}.$$

Beweis. a) Es sei $\operatorname{rang} A < n$. Dies ist gleichbedeutend mit $\operatorname{rang} A^{\mathsf{T}} < n$. Wegen Satz 9.2, (viii), ist daher $\det A = 0$ äquivalent zu $\det A^{\mathsf{T}} = 0$.

b) Es sei $\operatorname{rang} A = n$. Dann ist auch $\operatorname{rang} A^{\mathsf{T}} = n$. Gemäß Korollar 8.9 ist A Produkt endlich vieler Elementarmatrizen,

$$A = E_1 \cdot \ldots \cdot E_m.$$

Für A^T gilt $A^\mathsf{T} = {E_m}^\mathsf{T} \cdot \ldots \cdot {E_1}^\mathsf{T}$ und wegen Satz 9.2, (ix),

$$\det A^\mathsf{T} = \det {E_1}^\mathsf{T} \cdot \ldots \cdot \det {E_m}^\mathsf{T}. \tag{9.1}$$

Aus der speziellen Gestalt der Elementarmatrizen (siehe Seite 131) folgt sofort

$$E_n(i,j)^\mathsf{T} = E_n(i,j), \quad E_n(\lambda,i,j)^\mathsf{T} = E_n(\lambda,j,i), \quad E_n(\lambda,i)^\mathsf{T} = E_n(\lambda,i)$$

und $\det E_n(i,j)^\mathsf{T} = \det E_n(i,j) = -1$, $\det E_n(\lambda,i,j)^\mathsf{T} = \det E_n(\lambda,j,i) = 1$, $\det E_n(\lambda,i)^\mathsf{T} = \det E_n(\lambda,i) = \lambda$. Somit gilt $\det \tilde{E}^\mathsf{T} = \det \tilde{E}$ für jede Elementarmatrix $\tilde{E}$. Aus (9.1) folgt daher

$$\det A^\mathsf{T} = \det E_1 \cdot \ldots \cdot \det E_m = \det A.$$

□

Das folgende Resultat zeigt, dass es höchstens eine Determinante auf $M_{n,n}(\mathbb{K})$ mit den Eigenschaften (D1) – (D3) geben kann.

Satz 9.5. *Es seien* det *und* Det *Abbildungen* $M_{n,n}(\mathbb{K}) \to \mathbb{K}$, *die beide die Eigenschaften* (D1) – (D3) *besitzen. Dann gilt*

$$\det A = \operatorname{Det} A \quad \textit{für alle } A \in M_{n,n}(\mathbb{K}).$$

Beweis. Satz 9.2 gilt sowohl für det als auch für Det. Daher stimmen die zwei Determinanten für Elementarmatrizen überein. Hieraus folgt wiederum (Korollar 8.9), dass $\det A = \operatorname{Det} A$ für reguläre Matrizen A gilt. Ist A nicht regulär, so muß wegen Satz 9.2, (viii), $\det A = \operatorname{Det} A = 0$ gelten. Damit ist $\det \equiv \operatorname{Det}$ gezeigt. □

Um zu zeigen, dass tatsächlich eine Determinante existiert, müssen wir einige Vorbereitungen treffen.

Es sei die Matrix $A = (a_{i,j})_{i,j=1,\ldots,n}$ gegeben. Wir definieren die Zeilenvektoren

$$\begin{aligned} \alpha_i &= \big(a_{i,1} \cdots a_{i,n}\big), \\ e_i &= \big(0 \cdots 0\ 1\ 0 \cdots 0\big) \quad (\text{“1” an der } i\text{-ten Stelle}). \end{aligned}$$

Wegen (D1) gilt

$$\begin{aligned} \det A &= \sum_{i_1=1}^n a_{1,i_1} \det \begin{pmatrix} e_{i_1} \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} = \sum_{i_1=1}^n \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^n a_{1,i_1} a_{2,i_2} \det \begin{pmatrix} e_{i_1} \\ e_{i_2} \\ \alpha_3 \\ \vdots \\ \alpha_n \end{pmatrix} \\ &= \cdots = \sum_{i_1=1}^n \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^n \cdots \sum_{\substack{i_n=1 \\ i_n \notin \{i_1,\ldots,i_{n-1}\}}} a_{1,i_1} a_{2,i_2} \cdots a_{n,i_n} \det \begin{pmatrix} e_{i_1} \\ e_{i_2} \\ \vdots \\ e_{i_n} \end{pmatrix}. \end{aligned}$$

Für die in der Summe aufscheinenden n-Tupel $(i_1, \ldots, i_n)$ sind die Zahlen $i_1, \ldots, i_n$ paarweise verschieden und es kommen sämtliche n-Tupel dieser Form vor. Für ein bestimmtes n-Tupel $(i_1, \ldots, i_n)$ wird durch

$$\pi(k) = i_k, \quad k = 1, \ldots, n,$$

eine bijektive Abbildung $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$ definiert. Es bezeichne $\mathcal{S}_n$ die Menge aller bijektiven Abbildungen $\{1, \dots, n\} \to \{1, \dots, n\}$. Dann können wir die Summe von oben wie folgt schreiben:

$$\det A = \sum_{\pi \in \mathcal{S}_n} a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)} \det \begin{pmatrix} e_{\pi(1)} \\ e_{\pi(2)} \\ \vdots \\ e_{\pi(n)} \end{pmatrix}. \tag{9.2}$$

9.2 Permutationen

Definition 9.6. *Es sei $M = \{1, \dots, n\}$. Eine bijektive Abbildung $\pi : M \to M$ heißt eine* ***Permutation****. Wir schreiben*

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

Der Name Permutation wird durch folgenden Sachverhalt erklärt: Das n-Tupel der Bildelemente $(\pi(1), \dots, \pi(n))$ entsteht aus $(1, \dots, n)$ durch Umordnen der Elemente $1, \dots, n$.

Satz 9.7. *Die Menge $\mathcal{S}_n$ aller Permutationen von $M = \{1, \dots, n\}$ ist eine Gruppe (mit der Komposition von Abbildungen als algebraische Operation). $\mathcal{S}_n$ heißt die* ***symmetrische Gruppe*** *und enthält $n!$ Elemente.*

Beweis. Da Permutationen spezielle Abbildungen sind, folgt die Assoziativität der algebraischen Verknüpfung von Satz 1.22, a). Einselement ist die identische Abbildung, $\epsilon = id_M$. Das inverse Element zu π ist die Umkehrabbildung π^{-1}. Da einer Permutation eindeutig eine bestimmte Anordnung der n Elemente $1, \dots, n$ entspricht, ist die Anzahl der verschiedenen Permutationen $n!$. □

Definition 9.8. *$\pi \in \mathcal{S}_n$ heißt ein* ***Zyklus*** *der Ordnung k, $k \leq n$, wenn es k Elemente $i_1, \dots, i_k$ aus M gibt mit*

$$\pi(i_m) = i_{m+1}, \quad m = 1, \dots, k-1, \; \pi(i_k) = i_1 \text{ und}$$
$$\pi(i) = i \text{ für } i \notin \{i_1, \dots, i_k\}.$$

Wir schreiben $\pi = (i_1, \dots, i_k)$. Ein Zyklus der Ordnung 2 heißt eine ***Transposition****.*

Die identische Abbildung können wir als Zyklus der Ordnung 1 auffassen. Ist $\pi \in \mathcal{S}_n$ ein Zyklus der Ordnung k, so gilt $\pi^k = \epsilon$.[33] (Beweis: Es sei $\pi = (i_1, \dots, i_k)$. Für $i \in \{i_1, \dots, i_k\}$ gilt $\pi^k(i) = i$.)

Satz 9.9. a) *Jedes $\pi \in \mathcal{S}_n$ läßt sich als Produkt elementefremder Zyklen schreiben. Die Darstellung ist bis auf die Reihenfolge der Zyklen eindeutig.*
b) *Jedes $\pi \in \mathcal{S}_n$, $n \geq 2$, läßt sich als Produkt von Transpositionen schreiben.*

Beweis. a) Es ist klar, dass elementefremde Zyklen vertauschbar sind. Es sei nun $\pi \in \mathcal{S}_n$ gegeben. Ist $\pi = \epsilon$, so ist nichts zu beweisen. Es sei daher $\pi \neq \epsilon$. i_1 sei das kleinste Element in $\{1, \dots, n\}$ mit $\pi(i_1) \neq i_1$. In der Folge

$$i_1, \pi(i_1), \pi^2(i_1), \pi^3(i_1), \dots$$

[33] π^k ist durch $\pi^0 = \epsilon$, $\pi^k = \pi(\pi^{k-1})$, $k = 1, 2, \dots$, definiert.

können nicht alle Zahlen verschieden sein. Es existieren daher $0 \leq j < m$ mit $\pi^j(i_1) = \pi^m(i_1) = \pi^j\big(\pi^{m-j}(i_1)\big)$. Da π^j injektiv ist, gilt $i_1 = \pi^{m-j}(i_1)$. Wir können daher

$$k_1 := \min\{m \in \mathbb{N} \mid i_1 = \pi^m(i_1)\}$$

definieren. Wegen $\pi(i_1) \neq i_1$ gilt $k_1 \geq 2$. Die Zahlen $i_1, \dots, \pi^{k_1-1}(i_1)$ sind paarweise verschieden. Wäre $\pi^j(i_1) = \pi^m(i_1)$ für j, m mit $0 \leq j < m \leq k_1 - 1$, so würde wie oben $i_1 = \pi^{m-j}(i_1)$ folgen in Widerspruch zur Definition von k_1 (man beachte $1 \leq m-j \leq k_1 - 1$). Die Permutation

$$\pi_1 = \big(i_1, \pi(i_1), \dots, \pi^{k_1-1}\big)$$

ist somit ein Zyklus der Länge k_1.

Für die Menge $M_1 := \{1, \dots, n\} \setminus \{i_1, \dots, \pi^{k_1-1}(i_1)\}$ sind zwei Fälle zu unterscheiden:

Fall 1: $\pi(i) = i$ für alle $i \in M_1$.
In diesem Fall ist $\pi = \pi_1$.

Fall 2: Es gibt mindestens ein $i \in M_1$ mit $\pi(i) \neq i$.
Wir setzen $i_2 := \min\{i \in M_1 \mid \pi(i) \neq i\}$. Analog wie oben können wir

$$k_2 := \min\{m \in \mathbb{N} \mid i_2 = \pi^m(i_2)\}$$

definieren. Die Zahlen $i_2, \dots, \pi^{k_2-1}(i_2)$ sind paarweise verschieden und

$$\pi_2 := \big(i_2, \dots, \pi^{k_2-1}(i_2)\big)$$

ist ein Zyklus der Länge k_2. Es gilt darüberhinaus

$$\{i_1, \dots, \pi^{k_1-1}(i_1)\} \cap \{i_2, \dots, \pi^{k_2-1}(i_2)\} = \emptyset.$$

Wäre der Durchschnitt nicht-leer, so hätte man $\pi^m(i_1) = \pi^p(i_2)$ mit Zahlen $0 \leq m \leq k_1 - 1$, $0 \leq p \leq k_2 - 1$. Daraus würde $i_1 = \pi^{p-m}(i_2)$ oder $i_2 = \pi^{m-p}(i_1)$ folgen (je nachdem $p - m \geq 0$ oder $m - p \geq 0$). Aus $i_1 = \pi^{p-m}(i_2)$ folgt aber auch $\pi^{k_2-p+m}(i_1) = \pi^{k_2-p+m}\pi^{p-m}(i_2) = \pi^{k_2}(i_2) = i_2$, d.h. es würde in beiden Fällen $i_2 \in \{i_1, \dots, \pi^{k_1-1}(i_1)\}$ gelten in Widerspruch zur Wahl von i_2.

Für den nächsten Schritt definieren wir

$$M_2 := \{1, \dots, n\} \setminus \{i_1, \dots, \pi^{k_1-1}(i_1), i_2, \dots, \pi^{k_2-1}(i_2)\}.$$

Man hat wieder analog wie oben zwei Fälle zu unterscheiden. In endlich vielen Schritten erhalten wir eine Darstellung von π der verlangten Art.

b) Es genügt Zyklen als Produkt von Transpositionen zu schreiben. Es sei $\pi = (i_1, \dots, i_k)$, $k > 1$. Dann gilt:

$$\pi = (i_1, i_2)(i_2, i_3) \cdot \ldots \cdot (i_{k-1}, i_k).$$

Für $\pi = \epsilon$ gilt beispielsweise $\epsilon = (1\,2)(1\,2)$. □

Beispiel 9.10. Es sei $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 4 & 5 & 6 & 1 & 7 & 2 & 8 \end{pmatrix} \in \mathcal{S}_8$ gegeben. Die Darstellung von π als Kompositum elementefremder Zyklen ist durch $\pi = (1\,3\,5)(2\,4\,6\,7)(8) = (1\,3\,5)(2\,4\,6\,7)$ gegeben. Man beachte $(8) = \epsilon$. ◊

Satz 9.11. *Gilt für Transpositionen τ_i, $i = 1, \dots, k$, und $\tilde{\tau}_j$, $j = 1, \dots, m$, in $\mathcal{S}_n$*

$$\tau_1 \cdot \ldots \cdot \tau_k = \tilde{\tau}_1 \cdot \ldots \cdot \tilde{\tau}_m,$$

so sind k und m zugleich gerade oder ungerade, d.h. $(-1)^k = (-1)^m$.

Beweis. Wir betrachten die Funktion

$$P(x_1, \ldots, x_n) = \prod_{\substack{i,j=1 \\ i<j}}^{n} (x_i - x_j).$$

Es sei nun eine Transposition $\tau = (i, k)$, $i < k$, gegeben. Dann ist leicht einzusehen, dass $P(x_{\tau(1)}, \ldots, x_{\tau(n)}) = P(x_1, \ldots, x_{i-1}, x_k, x_{i+1}, \ldots, x_{k-1}, x_i, x_{k+1}, \ldots, x_n) = \pm P(x_1, \ldots, x_n)$ gilt. Um das genaue Vorzeichen zu bestimmen, müssen wir nur jene Faktoren betrachten, in welchen x_i oder x_k vorkommen. Es sind dies die folgenden:

$P(x_1, \ldots, x_n)$	$P(x_{\tau(1)}, \ldots, x_{\tau(n)})$
a) $x_1 - x_i, \ldots, x_{i-1} - x_i$	a') $x_1 - x_k, \ldots, x_{i-1} - x_k$
b) $x_1 - x_k, \ldots, x_{i-1} - x_k$	b') $x_1 - x_i, \ldots, x_{i-1} - x_i$
c) $x_i - x_{i+1}, \ldots, x_i - x_{k-1}$	c') $x_k - x_{i+1}, \ldots, x_k - x_{k-1}$
d) $x_i - x_k$	d') $x_k - x_i$
e) $x_i - x_{k+1}, \ldots, x_i - x_n$	e') $x_k - x_{k+1}, \ldots, x_k - x_n$
f) $x_{i+1} - x_k, \ldots, x_{k-1} - x_k$	f') $x_{i+1} - x_i, \ldots, x_{k-1} - x_i$
g) $x_k - x_{k+1}, \ldots, x_k - x_n$	g') $x_i - x_{k+1}, \ldots, x_i - x_n$

Die Faktoren der Gruppen a), b), e) und g) für $P(x_1, \ldots, x_n)$ kommen auch in $P(x_{\tau(1)}, \ldots$ $\ldots, x_{\tau(n)})$ vor und zwar in den Gruppen b'), a'), g') und e'). Die Faktoren der Gruppen c), d) und f) für $P(x_1, \ldots, x_n)$ erscheinen in $P(x_{\tau(1)}, \ldots, x_{\tau(n)})$ mit negativem Vorzeichen in den Gruppen f'), d') und c'). Die Anzahl der Vorzeichenwechsel ist daher gleich der Anzahl der Faktoren in den Gruppen c), d) und f):

$$k - 1 - i + 1 + k - 1 - i = 2k - 2i - 1.$$

Im Fall $k = i + 1$ fehlen in der Tabelle die Gruppen c), f) und c'), f'). Für den Vorzeichenwechsel sind nur die Gruppen d) und d') maßgebend. Die Anzahl der Faktoren mit unterschiedlichem Vorzeichen ist in diesem Fall 1.

Somit gilt in jedem Fall: $P(x_{\tau(1)}, \ldots, x_{\tau(n)}) = -P(x_1, \ldots, x_n)$. Ist nun $\pi = \tau_1 \cdot \ldots \cdot \tau_k = \tilde{\tau}_1 \cdot \ldots \cdot \tilde{\tau}_m$, so folgt durch wiederholte Anwendung des eben bewiesenen Resultates für Transpositionen

$$P(x_{\pi(1)}, \ldots, x_{\pi(n)}) = (-1)^k P(x_1, \ldots, x_n) = (-1)^m P(x_1, \ldots, x_n)$$

und daher auch $(-1)^k = (-1)^m$. $\square$

Definition 9.12. *Es sei $\pi \in \mathcal{S}_n$ gegeben. Das **Signum** von π ist definiert als*

$$\operatorname{sign} \pi := (-1)^k,$$

wobei π als Produkt von k Transpositionen dargestellt werden kann. Ist $\operatorname{sign} \pi = 1$, *so heißt π eine **gerade Permutation**, sonst eine **ungerade Permutation**.*

Satz 9.11 zeigt, dass $\operatorname{sign} \pi$ eindeutig definiert ist. Zyklen gerader Ordnung sind ungerade Permutationen. Dies folgt aus der in Beweis zu Satz 9.9, b), angegebenen Darstellung von Zyklen als Produkt von Transpositionen.

*Die Menge $\mathcal{A}_n$ der geraden Permutationen aus $\mathcal{S}_n$ ist bezüglich der auf $\mathcal{S}_n$ erklärten Verknüpfung (das ist die Bildung des Kompositums zweier Permutationen) eine Gruppe, die **alternierende Gruppe**.*

Beweis. Es ist klar, dass das Kompositum zweier gerader Permutationen wieder eine gerade Permutation ist. Ebenso ist für $\pi \in \mathcal{A}_n$ stets auch $\pi^{-1} \in \mathcal{A}_n$. Aus diesen zwei Aussagen folgt sofort, dass auch $\varepsilon \in \mathcal{A}_n$ gilt. Ebenso klar ist, dass das Assoziativgesetz für die Bildung des Kompositums von geraden Permutationen gilt. □

Da jede Permutation entweder gerade oder ungerade ist gilt:

$$\mathcal{S}_n = \mathcal{A}_n \cup (\mathcal{S}_n \setminus \mathcal{A}_n) \quad \text{und} \quad \mathcal{A}_n \cap (\mathcal{S}_n \setminus \mathcal{A}_n) = \emptyset. \tag{9.3}$$

Wir werden später die folgenden Aussagen benötigen:

Satz 9.13. a) $\pi \to \pi^{-1}$ *definiert eine bijektive Abbildung* $\mathcal{S}_n \to \mathcal{S}_n$. *Ferner gilt* $\operatorname{sign} \pi^{-1} = \operatorname{sign} \pi$ *für alle* $\pi \in \mathcal{S}_n$.

b) *Es sei* $\sigma \in \mathcal{S}_n$ *gegeben.* $\pi \to \sigma\pi$ *bzw.* $\pi \to \pi\sigma$ *definiert eine bijektive Abbildung* $\mathcal{S}_n \to \mathcal{S}_n$. *Ferner ist* $\operatorname{sign} \pi\sigma = \operatorname{sign} \pi \cdot \operatorname{sign} \sigma$.

c) *Es sei* $\tau \in \mathcal{S}_n$ *eine Transposition und* $\Phi : \mathcal{S}_n \to \mathcal{S}_n$ *die durch* $\pi \to \pi\tau$, $\pi \in \mathcal{S}_n$ *definierte bijektive Abbildung. Dann ist* $\Phi\,\big|_{\mathcal{A}_n}$ *eine bijektive Abbildung* $\mathcal{A}_n \to \mathcal{S}_n \setminus \mathcal{A}_n$.

Beweis. a) Für $\pi \in \mathcal{S}_n$ gilt $\pi = (\pi^{-1})^{-1}$, d.h. die Abbildung ist surjektiv und damit auch bijektiv.

b) Aus $\pi_1\sigma = \pi_2\sigma$ folgt unmittelbar $\pi_1 = \pi_2$, d.h. die Abbildung ist injektiv und somit auch bijektiv.

Die Aussagen über das Signum von π^{-1} bzw. $\pi\sigma$ sind trivial.

c) Für $\pi \in \mathcal{A}_n$ ist $\operatorname{sign} \Phi(\pi) = \operatorname{sign} \pi \cdot \operatorname{sign} \tau = -1$, d.h. es gilt $\Phi(\mathcal{A}_n) \subset \mathcal{S}_n \setminus \mathcal{A}_n$. Für $\pi \in \mathcal{S}_n \setminus \mathcal{A}_n$ ist $\pi^* = \pi\tau \in \mathcal{A}_n$ und es gilt $\Phi(\pi^*) = \pi^*\tau = \pi(\tau\tau) = \pi$. Damit ist $\Phi(A_n) = \mathcal{S}_n \setminus \mathcal{A}_n$ gezeigt, d.h. Φ ist surjektiv. Injektivität von Φ ist klar. □

Mit Hilfe von Punkt c) des Satzes sieht man sofort, dass $\mathcal{A}_n$ $n!/2$ Elemente enthält.

9.3 Existenz der Determinante

In diesem Abschnitt zeigen wir, dass es tatsächlich eine Determinante auf $M_{n,n}(\mathbb{K})$ gibt. Ausgangspunkt ist die Darstellung (9.2), die für jede Determinante gelten muss. Die in (9.2) aufscheinenden Matrizen

$$E_\pi = \begin{pmatrix} e_{\pi(1)} \\ e_{\pi(2)} \\ \vdots \\ e_{\pi(n)} \end{pmatrix}$$

entstehen aus der Einheitsmatrix E, indem man die Spalten der Einheitsmatrix der Permutation π unterwirft, d.h. die erste Spalte von E wird zur $\pi(1)$-sten Spalte von E_π, die zweite Spalte von E zur $\pi(2)$-ten Spalte von E_π etc. Ist $\pi = \tau_1 \cdots \tau_k$ mit Transpositionen $\tau_1 = (i_1, j_1)$,…,$\tau_k = (i_k, j_k)$, so entsteht E_π aus E, indem man zuerst in E die i_k-te mit der j_k-ten Spalte vertauscht, dann in der erhaltenen Matrix die i_{k-1}-te Spalte mit der j_{k-1}-ten Spalte etc. E_π entsteht somit aus E durch k-maliges Vertauschen von Spalten. Daher gilt

$$\det E_\pi = (-1)^k \det E = (-1)^k = \operatorname{sign} \pi.$$

Damit ist gezeigt, dass jede Determinante auf $M_{n,n}(\mathbb{K})$ durch die folgende Formel gegeben sein muss (dies beweist ebenfalls die Eindeutigkeit der Determinante):

$$\det A = \sum_{\pi \in \mathcal{S}_n} (\operatorname{sign} \pi) a_{1,\pi(1)} \cdots a_{n,\pi(n)}, \quad A = \big(a_{i,j}\big)_{i,j=1,\dots,n}.$$

Es bleibt noch zu zeigen, dass durch diese Formel tatsächlich eine Determinante definiert wird.

Satz 9.14 (Leibnitz[34]-Formel). *Es sei* $A = (\alpha_{ij}) \in M_{n,n}(\mathbb{K})$ *vorgegeben. Durch*

$$\det A := \sum_{\pi \in \mathcal{S}_n} (\operatorname{sign} \pi) \alpha_{1,\pi(1)} \cdot \alpha_{2,\pi(2)} \cdot \ldots \cdot \alpha_{n,\pi(n)}.$$

wird eine Determinante auf $M_{n,n}(\mathbb{K})$ *definiert.*

Beweis. Wir müssen die Eigenschaften (D1) – (D3) nachweisen.
a) Es sei $a_i = \lambda b_i + \mu c_i$, wobei $b_i = (\beta_{i1}, \ldots, \beta_{in})$ und $c_i = (\gamma_{i1}, \ldots, \gamma_{in})$. Dann gilt

$$\begin{aligned}
\det A &= \sum_{\pi \in \mathcal{S}_n} (\operatorname{sign} \pi) \alpha_{1,\pi(1)} \cdot \ldots \cdot \big(\lambda \beta_{i,\pi(i)} + \mu \gamma_{i,\pi(i)}\big) \cdot \ldots \cdot \alpha_{n,\pi(n)} \\
&= \lambda \sum_{\pi \in \mathcal{S}_n} (\operatorname{sign} \pi) \alpha_{1,\pi(1)} \cdot \ldots \cdot \beta_{i,\pi(i)} \cdot \ldots \cdot \alpha_{n,\pi(n)} \\
&\qquad + \mu \sum_{\pi \in \mathcal{S}_n} (\operatorname{sign} \pi) \alpha_{1,\pi(1)} \cdot \ldots \cdot \gamma_{i,\pi(i)} \cdot \ldots \cdot \alpha_{n,\pi(n)} \\
&= \lambda \det B + \mu \det C,
\end{aligned}$$

wobei B die Zeilen $a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n$ hat und C die Zeilen $a_1, \ldots, a_{i-1}, c_i, a_{i+1}, \ldots, a_n$. Damit ist (D1) bewiesen.
b) Für zwei Indizes $i, j \in \{1, \ldots, n\}$ mit $i \neq j$ sei $a_i = a_j$, d.h. es ist

$$\alpha_{i,k} = \alpha_{j,k}, \quad k = 1, \ldots, n. \tag{9.4}$$

Wir setzen $\tau = (i\ j)$. Dann definiert $\Phi(\pi) = \pi\tau$, $\pi \in \mathcal{S}_n$, eine bijektive Abbildung $\Phi : \mathcal{S}_n \to \mathcal{S}_n$ (Satz 9.13, b)). Die Einschränkung $\Phi\,|_{\mathcal{A}_n}$ ist eine bijektive Abbildung $\mathcal{A}_n \to \mathcal{S}_n \setminus \mathcal{A}_n$. Daher gilt

$$\begin{aligned}
\det A = &\sum_{\pi \in \mathcal{A}_n} \alpha_{1,\pi(1)} \cdot \ldots \cdot \alpha_{i,\pi(i)} \cdot \ldots \cdot \alpha_{j,\pi(j)} \cdot \ldots \cdot \alpha_{n,\pi(n)} \\
&- \sum_{\pi \in \mathcal{A}_n} \alpha_{1,\pi(\tau(1))} \cdot \ldots \cdot \alpha_{i,\pi(\tau(i))} \cdot \ldots \cdot \alpha_{j,\pi(\tau(j))} \cdot \ldots \cdot \alpha_{n,\pi(\tau(n))} \\
= &\sum_{\pi \in \mathcal{A}_n} \alpha_{1,\pi(1)} \cdot \ldots \cdot \alpha_{i,\pi(i)} \cdot \ldots \cdot \alpha_{j,\pi(j)} \cdot \ldots \cdot \alpha_{n,\pi(n)} \\
&- \sum_{\pi \in \mathcal{A}_n} \alpha_{1,\pi(1)} \cdot \ldots \cdot \alpha_{i,\pi(j)} \cdot \ldots \cdot \alpha_{j,\pi(i)} \cdot \ldots \cdot \alpha_{n,\pi(n)}.
\end{aligned}$$

Beachtet man (9.4), so folgt $\det A = 0$. Damit ist (D2) bewiesen.
c) Es gilt

$$\det E = \sum_{\pi \in \mathcal{S}_n} (\operatorname{sign} \pi) \delta_{1,\pi(1)} \cdot \ldots \cdot \delta_{n,\pi(n)},$$

wobei

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

[34]Gottfried Wilhelm Leibnitz, 1. 7. 1646 (Leipzig) – 14. 11. 1716 (Hannover), Philosoph, Mathematiker, Physiker Historiker, Diplomat, Doktor des weltlichen und des Kirchenrechtes, der Universalgelehrte seiner Zeit, Begründer der Infinitesimalrechnung (unabhängig von Newton), die Symbole dy/dx und $\int f\,dx$ gehen auf Leibitz zurück, (de.freepedia.org/Gottfried_Wilhelm_Leibnitz.html).

Der einzige Summand $\neq 0$ ist jener mit $i = \pi(i)$ für $i = 1, \ldots, n$, d.h. jener mit $\pi = \varepsilon$. Daraus folgt $\det E = 1$, d.h. (D3) ist erfüllt. $\square$

Die Produkte $\alpha_{1,\pi(1)} \cdot \ldots \cdot \alpha_{n,\pi(n)}$, $\pi \in \mathcal{S}_n$, sind genau jene Produkte aus n Elementen der Matrix A, in denen genau ein Element aus jeder Zeile und Spalte von A als Faktor vorkommt.

Beispiel 9.15. n = 2.

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}.$$

$\mathcal{S}_2$ besteht aus zwei Permutationen: $\epsilon, (1,2)$ mit $\operatorname{sign} \epsilon = 1$ und $\operatorname{sign}(1,2) = -1$. Daher ist

$$\det A = \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}.$$

$\diamondsuit$

Beispiel 9.16. $n = 3$.

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix}.$$

Die geraden Permutationen von $\mathcal{S}_3$ sind

$$\epsilon,\ (1\ 2\ 3),\ (1\,3\,2),$$

die ungeraden

$$(1\,2),\ (1\,3),\ (2\,3).$$

Daher ist

$$\begin{aligned} \det A = {} & \alpha_{11}\alpha_{22}\alpha_{33} + \alpha_{12}\alpha_{23}\alpha_{31} + \alpha_{13}\alpha_{21}\alpha_{32} \\ & - \alpha_{12}\alpha_{21}\alpha_{33} - \alpha_{13}\alpha_{22}\alpha_{31} - \alpha_{11}\alpha_{23}\alpha_{32}. \end{aligned}$$

Diese Formel ist auch unter dem Namen „***Regel von Sarrus***“[35] bekannt. $\diamondsuit$

Da $\mathcal{S}_n$ aus $n!$ Elementen besteht, wird die Berechnung von $\det A$ an Hand der Definition für große n sehr aufwendig. Effizientere Methoden zur Berechnung von Determinanten basieren auf den Eigenschaften, die in Satz 9.2 aufgelistet sind. Die Aussagen (iv), (v) und (vi) aus Satz 9.2 ergeben den folgenden Algorithmus zur Berechnung von $\det A$:

Dreiecksalgorithmus. Durch Vertauschung von Zeilen der Matrix A bzw. Addition von Vielfachen einer Zeile zu anderen Zeilen transformiere man A auf eine obere Dreiecksmatrix A' mit den Elementen $\beta_1, \ldots, \beta_n$ in der Hauptdiagonalen. Dann ist

$$\det A = (-1)^k \beta_1 \cdot \ldots \cdot \beta_n,$$

wobei k die Anzahl der Zeilenvertauschungen ist, die vorgenommen wurden.

„Unterdeterminanten“ einer Matrix spielen in einer Reihe von theoretischen Überlegungen eine wichtige Rolle:

Definition 9.17. *Es sei $A = (\alpha_{ij}) \in M_{r,n}(\mathbb{K})$ und $1 \leq s \leq \min(r,n)$ gegeben.*

$$A^{i_1,\ldots,i_s}_{j_1,\ldots,j_s} := \det \begin{pmatrix} \alpha_{i_1,j_1} & \ldots & \alpha_{i_1,j_s} \\ \vdots & & \vdots \\ \alpha_{i_s,j_1} & \ldots & \alpha_{i_s,j_s} \end{pmatrix}, \quad \begin{array}{l} 1 \leq i_1 < \cdots < i_s \leq r, \\ 1 \leq j_1 < \cdots < j_s \leq n, \end{array}$$

[35] Sarrus, Pierre Frédéric, 10. 3. 1798 (Saint-Affriques) – 20. 11. 1861 (ebenda), wichtige Beiträge zur Variationsrechnung (de.wikipedia.org/wiki/Pierre_Frederic_Sarrus).

heißt ein ***Minor*** *s-ter Ordnung von A. Ein Minor s-ter Ordnung mit $i_1 = j_1, \dots, i_s = j_s$ heißt* ***Hauptminor*** *von A.*

Satz 9.18. *Es sei $A \in M_{r,n}(\mathbb{K})$ gegeben. Es ist genau dann $\operatorname{rang} A = k$, $1 \leq k \leq \min(r,n)$, wenn gilt:*

(i) *Es gibt einen von von Null verschiedenen Minor k-ter Ordnung von A.*

(ii) *Sämtliche Minoren $(k+1)$-ter Ordnung von A sind Null.*

Für $k = \min(r,n)$ entfällt (ii), für $k = 0$ entfällt (i).

Beweis. $\operatorname{rang} A = 0$ ist trivial. $\operatorname{rang} A = k$ gleichbedeutend damit, dass k Zeilen von A linear unabhängig sind. Es seien dies die Zeilen $i_1, \dots, i_k$. Die Matrix

$$A' = \begin{pmatrix} \alpha_{i_1,1} & \cdots & \alpha_{i_1,n} \\ \vdots & & \vdots \\ \alpha_{i_k,1} & \cdots & \alpha_{i_k,n} \end{pmatrix}$$

hat den Rang k. Es existieren daher k linear unabhängige Spalten, etwa $j_1, \dots, j_k$. Dann ist aber

$$\operatorname{rang} A'' = \operatorname{rang} \begin{pmatrix} \alpha_{i_1,j_1} & \cdots & \alpha_{i_1,j_k} \\ \vdots & & \vdots \\ \alpha_{i_k,j_1} & \cdots & \alpha_{i_k,j_k} \end{pmatrix} = k$$

und wegen Satz 9.2, (viii), auch $A^{i_1,\dots,i_k}_{j_1,\dots,j_k} \neq 0$. Ist umgekehrt der Minor $A^{i_1,\dots,i_k}_{j_1,\dots,j_k} \neq 0$, so hat die Matrix A'' und damit auch A' den Rang k. Daraus folgt $\operatorname{rang} A \geq k$.
Wir haben damit die folgenden Implikationen bewiesen:

a) Aus $\operatorname{rang} A = k$ folgt, dass ein von Null verschiedener Minor k-ter Ordnung existiert.

b) Existiert ein von Null verschiedener Minor k-ter Ordnung, so ist $\operatorname{rang} A \geq k$.

Der Beweis des Satzes ist nun einfach. Es sei $\operatorname{rang} A = k$. Dann gilt (i) wegen der schon bewiesenen Implikation a) . Würde (ii) nicht gelten, so gäbe es einen Minor $(k+1)$-ter Ordnung $\neq 0$ und wegen b) wäre $\operatorname{rang} A \geq k+1$. Damit ist auch (ii) bewiesen. Umgekehrt folgt $\operatorname{rang} A \geq k$ aus (i). Wäre $\operatorname{rang} A \geq k+1$, so gäbe es $k+1$ linear unabhängige Zeilen von A und die daraus gebildete Matrix A' hätte Rang $k+1$. Wegen a) hätte man einen von Null verschiedenen Minor $(k+1)$-ter Ordnung von A', der auch ein Minor von A wäre. Dies kann wegen (ii) nicht sein. Daher ist $\operatorname{rang} A \leq k$. □

Um Determinanten für Endomorphismen endlich-dimensionaler Vektorräume definieren zu können, benötigen wir das folgende Resultat:

Sind die Matrizen $A, B \in M_{n,n}(\mathbb{K})$ ähnlich, so gilt

$$\det A = \det B.$$

Beweis. Es ist $B = S^{-1}AS$ mit einer regulären $n \times n$-Matrix S. Aus Satz 9.2, (ix), folgt $\det B = \det S^{-1} \det A \det S = (\det S)^{-1} \det A \det S = \det A$. □

In Hinblick auf die eben bewiesene Aussage können wir definieren:

Definition 9.19. *Es sei $\phi \in L(X,X)$ und $\dim X = n$. A sei die Matrix von ϕ bzgl. einer Basis von X. Wir definieren*

$$\det \phi := \det A.$$

Da ähnliche Matrizen dieselbe Determinante haben, ist diese Definition sinnvoll. Es ist klar, dass sich viele der oben angegebenen Eigenschaften der Determinante auf Determinanten von Endomorphismen übertragen lassen. Beispielsweise ist ϕ invertierbar genau dann, wenn $\det\phi \neq 0$ ist. Wir überlassen es dem Leser, derartige Aussagen zu formulieren.

9.4 Der Entwicklungssatz von Laplace

Es sei $A = (\alpha_{ij}) \in M_{n,n}(\mathbb{K})$ gegeben. Ein Element α_{ij} kommt als Faktor in dem Produkt $\alpha_{1,\pi(1)} \cdot \ldots \cdot \alpha_{n,\pi(n)}$ genau dann vor, wenn $\pi(i) = j$ ist. Die Summe aller Summanden in $\det A$, welche α_{ij} als Faktor enthalten, ist daher durch

$$\alpha_{ij} \sum_{\substack{\pi\in\mathcal{S}_n\\ \pi(i)=j}} (\operatorname{sign}\pi)\alpha_{1,\pi(1)} \cdot \ldots \cdot \alpha_{i-1,\pi(i-1)}\alpha_{i+1,\pi(i+1)} \cdot \ldots \cdot \alpha_{n,\pi(n)} \tag{9.5}$$

gegeben. Die Matrix $\tilde{A}_{ij} = (\tilde{\alpha}_{k,\ell})_{k,\ell=1,\ldots,n}$ sei durch

$$\tilde{\alpha}_{k,\ell} = \begin{cases} \tilde{\alpha}_{k,\ell} & \text{für } k \neq i, \\ 0 & \text{für } k = i \text{ und } \ell \neq j, \\ 1 & \text{für } k = i \text{ und } \ell = j, \end{cases}$$

definiert, d.h.

$$\tilde{A}_{ij} = \begin{array}{cl} \begin{pmatrix} \alpha_{1,1} & \ldots\ldots\ldots\ldots\ldots\ldots & \alpha_{1,n} \\ \vdots & & \vdots \\ \alpha_{i-1,1} & \ldots\ldots\ldots\ldots\ldots\ldots & \alpha_{i-1,n} \\ 0 & \cdots \; 0 \;\; 1 \;\; 0 \; \cdots & 0 \\ \alpha_{i+1,1} & \ldots\ldots\ldots\ldots\ldots\ldots & \alpha_{i+1,n} \\ \vdots & & \vdots \\ \alpha_{n,1} & \ldots\ldots\ldots\ldots\ldots\ldots & \alpha_{n,n} \end{pmatrix}. & -\, i\text{-te Zeile} \\ \begin{array}{c} | \\ j\text{-te Spalte} \end{array} & \end{array}$$

Man sieht nun sofort, dass

$$\det \tilde{A}_{ij} = \sum_{\substack{\pi\in\mathcal{S}_n\\ \pi(i)=j}} (\operatorname{sign}\pi)\alpha_{1,\pi(1)} \cdot \ldots \cdot \alpha_{i-1,\pi(i-1)}\alpha_{i+1,\pi(i+1)} \cdot \ldots \cdot \alpha_{n,\pi(n)} \tag{9.6}$$

gilt. Denn für eine Permutation π mit $\pi(i) \neq j$ ist $\tilde{\alpha}_{i,\pi(i)} = 0$, während $\tilde{\alpha}_{i,\pi(i)} = 1$ für π mit $\pi(i) = j$ gilt.

Durch Addition von entsprechenden Vielfachen der i-ten Zeile in $\tilde{A}_{ij}$ zu den anderen Zeilen erhält man (Satz 9.2, (v))

$$\det \tilde{A}_{ij} = \det \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1,j-1} & 0 & \alpha_{1,j+1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{i-1,1} & \cdots & \alpha_{i-1,j-1} & 0 & \alpha_{i-1,j+1} & \cdots & \alpha_{i-1,n} \\ 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ \alpha_{i+1,1} & \cdots & \alpha_{i+1,j-1} & 0 & \alpha_{i+1,j+1} & \cdots & \alpha_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{n,1} & \cdots & \alpha_{n,j-1} & 0 & \alpha_{n,j+1} & \cdots & \alpha_{nn} \end{pmatrix}.$$

Durch Vertauschung von Zeilen und Spalten erreicht man schließlich (siehe Definition 9.1, (D1), Satz 9.4 und Satz 9.2, (vii))

$$\det \tilde{A}_{ij} = (-1)^{i+j} \det \begin{pmatrix} 1 & 0 \\ 0 & A_{ij} \end{pmatrix} = (-1)^{i+j} \det A_{ij}, \tag{9.7}$$

wobei die $(n-1) \times (n-1)$-Matrix A_{ij} aus A durch Streichen der i-ten Zeile und der j-ten Spalte entsteht. Es sind nun folgende Bezeichnungen üblich:

Definition 9.20. *$A = (\alpha_{ij}) \in M_{n,n}(\mathbb{K})$ sei gegeben. Die Matrizen $A_{ij} \in M_{n-1,n-1}(\mathbb{K})$ seien wie oben definiert, $i, j = 1, \ldots, n$. Die Zahl $\operatorname{adj} \alpha_{ij} := (-1)^{i+j} \det A_{ij}$ heißt das zu α_{ij}* ***adjunkte Element*** *(oder das* ***algebraische Komplement*** *von α_{ij}), $i, j = 1, \ldots, n$. Die Matrix*

$$\operatorname{adj} A := (\operatorname{adj} \alpha_{ij})^{\mathsf{T}}_{i,j=1,\ldots,n}$$

heißt die zu A ***adjunkte Matrix*** *oder die zu A* ***Adjunkte****.*

Mit diesen Bezeichnungen ist die Summe aller Summanden in $\det A$ welche α_{ij} als Faktor enthalten durch (siehe (9.5) – (9.7))

$$\alpha_{ij} \operatorname{adj} \alpha_{ij}$$

gegeben. Da

$$\mathcal{S}_n = \bigcup_{j=1}^{n} \{\pi \in \mathcal{S}_n \mid \pi(i) = j\}$$

gilt und $\{\pi \in \mathcal{S}_n \mid \pi(i) = j_1\} \cap \{\pi \in \mathcal{S}_n \mid \pi(i) = j_2\} = \emptyset$ für $j_1 \neq j_2$ ist, erhalten wir insgesamt

$$\det A = \sum_{j=1}^{n} \alpha_{ij} \operatorname{adj} \alpha_{ij}.$$

Damit haben wir bewiesen:

Satz 9.21 (Entwicklungssatz von Laplace[36]). *Die Matrix $A = (\alpha_{ij}) \in M_{n,n}(\mathbb{K})$ sei gegeben. Für jedes $i = 1, \ldots, n$ gilt (Entwicklung nach der i-ten Zeile):*

$$\det A = \sum_{j=1}^{n} \alpha_{ij} \operatorname{adj} \alpha_{ij}.$$

Für jedes $j = 1, \ldots, n$ gilt (Entwicklung nach der j-ten Spalte):

$$\det A = \sum_{i=1}^{n} \alpha_{ij} \operatorname{adj} \alpha_{ij}.$$

Die Entwicklung nach der j-ten Spalte ergibt sich aus $\det A = \det A^{\mathsf{T}}$, wenn $\det A^{\mathsf{T}}$ nach der j-ten Zeile entwickelt wird.

Satz 9.22. *Es sei $A \in M_{n,n}(\mathbb{K})$ gegeben. Dann gilt:*

$$A(\operatorname{adj} A) = (\operatorname{adj} A)A = \det A \cdot E.$$

[36] Laplace, Pierre-Simon, 23. 3. 1749 (Beaumont-en-Auge, Normandie) – 5. 3. 1827 (Paris), Beiträge zu Differenzen- und Differentialgleichungen, zur Wahrscheinlichkeitstheorie, zur mathematischen Astronomie (www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Laplace.html).

Beweis. Wir beweisen zuerst die Aussage für $A(\operatorname{adj} A) = (\gamma_{ij})$. Nach Satz 9.21 gilt $\gamma_{ii} = \sum\limits_{k=1}^{n} \alpha_{ik} \operatorname{adj} \alpha_{ik} = \det A$, $i = 1, \ldots, n$. Für $i \neq j$ erhalten wir $\gamma_{ij} = \sum\limits_{k=1}^{n} \alpha_{ik} \operatorname{adj} \alpha_{jk}$. Nach Satz 9.21 ist $\gamma_{ij} = \det \tilde{A}$ mit

$$\tilde{A} = \begin{array}{l} \begin{pmatrix} \alpha_{11} & \cdots\cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{i1} & \cdots\cdots & \alpha_{in} \\ \vdots & & \vdots \\ \alpha_{i1} & \cdots\cdots & \alpha_{in} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots\cdots & \alpha_{nn} \end{pmatrix} \end{array} \begin{array}{l} \\ \\ - i\text{-te Zeile} \\ \\ - j\text{-te Zeile} \\ \\ \\ \end{array}$$

(Entwicklung von $\det \tilde{A}$ nach der j-ten Zeile). In $\tilde{A}$ sind aber zwei Zeilen gleich, d.h. es ist $\gamma_{ij} = 0$ für $i \neq j$. Damit ist $A(\operatorname{adj} A) = \det A \cdot E$ gezeigt. Die Aussage für $(\operatorname{adj} A)A$ folgt analog, indem man den Laplace'schen Entwicklungssatz für die Spalten verwendet. □

Korollar 9.23. *Es sei $A \in M_{n,n}(\mathbb{K})$ regulär. Dann gilt*

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

Beweis. Nach Satz 9.22 gilt für beliebige Matrizen

$$A \operatorname{adj} A = \det A \cdot E.$$

Durch Multiplikation beider Seiten mit $\frac{1}{\det A} A^{-1}$ (weil A regulär ist, gilt $\det A \neq 0$ und A^{-1} existiert) erhalten wir die zu beweisende Formel. □

Ist $A \in M_{n,n}(\mathbb{K})$ regulär, so ist die Lösung des Gleichungssystems

$$A\xi = \beta,\ \beta \in \mathbb{K}^n,$$

durch $\xi = A^{-1}\beta$ gegeben. Auf Grund des Korollars 9.23 erhalten wir

$$\xi = \frac{1}{\det A} (\operatorname{adj} A)\beta.$$

Die j-te Koordinate ξ_j ist daher durch

$$\xi_j = \frac{1}{\det A} \sum_{k=1}^{n} \beta_k \operatorname{adj} \alpha_{kj}$$

gegeben. Wir definieren die Matrix

$$A_j := \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1,j-1} & \beta_1 & \alpha_{1,j+1} & \cdots & \alpha_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \alpha_{n1} & \ldots & \alpha_{n,j-1} & \beta_n & \alpha_{n,j+1} & \ldots & \alpha_{nn} \end{pmatrix},$$

d.h. A_j entsteht aus A, indem man die j-te Spalte durch den Vektor β ersetzt. Mit Hilfe von Satz 9.21 (Entwicklung von $\det A_j$ nach der j-ten Spalte) folgt

$$\det A_j = \sum_{k=1}^{n} \beta_k \operatorname{adj} \alpha_{kj}.$$

Wir haben somit insgesamt bewiesen:

Satz 9.24 (Cramersche Regel[37]). *Es sei $A \in M_{n,n}(\mathbb{K})$ regulär. Für $\beta \in \mathbb{K}^n$ und $j = 1, \ldots, n$ sei A_j die Matrix, welche aus A entsteht, indem die j-te Spalte durch β ersetzt wird. Dann ist die eindeutig bestimmte Lösung $\xi = (\xi_1, \ldots, \xi_n)^\mathsf{T}$ von*

$$A\xi = \beta$$

durch

$$\xi_i = \frac{\det A_i}{\det A}, \quad i = 1, \ldots, n,$$

gegeben.

9.5 Determinantenformen

Definition 9.25. *Gegeben sei ein Vektorraum X über $\mathbb{K}$ und eine Abbildung $\Phi : \underbrace{X \times \cdots \times X}_{k\text{-mal}} \to \mathbb{K}$. Φ heißt genau dann eine **k-fache Linearform** auf X, wenn gilt:*
Für alle $i = 1, \ldots, k$ ist

$$\begin{aligned}\Phi(a_1, \ldots, a_{i-1}, \alpha a_i + \beta \tilde{a}_i, a_{i+1}, \ldots, a_k)& \\ = \alpha\Phi(a_1, \ldots, a_i, \ldots, a_k) + \beta\Phi(a_1, \ldots, \tilde{a}_i, \ldots, a_k),&\end{aligned}$$

*für $a_1, \ldots, a_i, \tilde{a}_i, \ldots, a_k \in X$ und $\alpha, \beta \in \mathbb{K}$. Eine k-fache Linearform auf X heißt **schiefsymmetrisch**, wenn für $\pi \in \mathcal{S}_n$ und $a_1, \ldots, a_k \in X$ gilt:*

$$\Phi(a_{\pi(1)}, \ldots, a_{\pi(k)}) = \operatorname{sign} \pi \Phi(a_1, \ldots, a_k).$$

*Ist $\dim X = n < \infty$, so heißt eine n-fache schiefsymmetrische Linearform $\Phi \not\equiv 0$ auch **Determinantenform** auf X.*

Beispiel 9.26. $\mathcal{B} = (b_1, \ldots, b_n)$ sei eine geordnete Basis von X. Durch

$$\Phi(a_1, \ldots, a_n) := \det(\alpha_1, \ldots, \alpha_n),$$

wobei α_j der Koordinatenvektor von a_j bzgl. $\mathcal{B}$ ist, $i = 1, \ldots, n$, wird eine Determinantenform auf X definiert. $\Diamond$

Über alle möglichen Determinantenformen gibt der folgende Satz Auskunft:

Satz 9.27. *X sei ein n-dimensionaler Vektorraum über $\mathbb{K}$ und Φ_1, Φ_2 seien n-fache, schiefsymmetrische Linearformen auf X mit $\phi_1 \not\equiv 0$ (d.h. insbesondere, dass Φ_1 eine Determinantenform auf X ist). Dann existiert ein $\lambda \in \mathbb{K}$ mit*

$$\Phi_2(a_1, \ldots, a_n) = \lambda\Phi_1(a_1, \ldots, a_n)$$

für alle $a_1, \ldots, a_n \in X$.

Beweis. Wir wählen eine geordnete Basis $\mathcal{B} = (b_1, \ldots, b_n)$ von X und definieren

$$\Phi_0(a_1, \ldots, a_n) = \det(\alpha_1, \ldots, \alpha_n),$$

[37] Cramer, Gabriel, 31. 7. 1704 (Genf) – 4. 1. 1752 (Bagnols-sur-Cèze), weitgestreute Interessen, Hauptwerk "Introduction à l'analyse des lignes courbes algébriques" (1750) (www-history.mcs.st-andrews.ac.uk/history/Mathematicians/Cramer.html).

α_i der Koordinatenvektor von a_i bzgl. $\mathcal{B}$, $i = 1, \dots, n$. Φ_0 ist eine Determinantenform auf X. Φ sei eine beliebige n-fache schiefsymmetrische Linearform auf X. Für $a_1, \dots, a_n \in X$ ist

$$a_j = \alpha_{1j} b_1 + \cdots + \alpha_{nj} b_n, \quad j = 1, \dots, n.$$

Wegen der n-fachen Linearität von Φ erhalten wir

$$\begin{aligned}\Phi(a_1, \dots, a_n) &= \sum_{i_1=1}^{n} \cdots \sum_{i_n=1}^{n} \alpha_{1,i_1} \cdot \ldots \cdot \alpha_{n,i_n} \Phi(b_{i_1}, \dots, b_{i_n}) \\ &= \sum_{\pi \in \mathcal{S}_n} \alpha_{1,\pi(1)} \cdot \ldots \cdot \alpha_{n,\pi(n)} \Phi(b_{\pi(1)}, \dots, b_{\pi(n)}).\end{aligned}$$

Wegen der Schiefsymmetrie von Φ erhalten wir $\Phi(b_{\pi(1)}, \dots, b_{\pi(n)}) = \operatorname{sign} \pi \Phi(b_1, \dots, b_n)$ und daher

$$\begin{aligned}\Phi(a_1, \dots, a_n) &= \Phi(b_1, \dots, b_n) \cdot \det(\alpha_1, \dots, \alpha_n) \\ &= \Phi(b_1, \dots, b_n) \Phi_0(a_1, \dots, a_n)\end{aligned}$$

für beliebige $a_1, \dots, a_n \in X$. Damit ist $\Phi = \lambda \Phi_0$ mit $\lambda = \Phi(b_1, \dots, b_n)$ gezeigt. Somit gilt $\Phi_1 = \lambda_1 \Phi_0$ und $\Phi_2 = \lambda_2 \Phi_0$ und daher $\Phi_2 = \frac{\lambda_2}{\lambda_1} \Phi_1$. Wegen $\Phi_1 \not\equiv 0$ ist $\lambda_1 \neq 0$. $\square$

9.6 Vektorräume mit Orientierung

Für diesen Abschnitt sei X ein reeller Vektorraum mit $\dim X = n < \infty$. Auf Grund von Satz 9.27 ist die folgende Definition sinnvoll:

Definition 9.28. Φ_1, Φ_2 *seien Determinantenformen auf* X. Φ_1 *und* Φ_2 *heißen genau dann äquivalent, wenn ein* $\lambda > 0$ *existiert mit*

$$\Phi_1 \equiv \lambda \Phi_2.$$

Es ist klar, dass für Determinantenformen auf X stets $\Phi_1 = \lambda \Phi_2$ mit $\lambda \neq 0$, $\lambda \in \mathbb{R}$ gilt. Durch die eben definierte Äquivalenzrelation zerfällt die Menge der Determinantenformen auf X in zwei Klassen, welche wir mit $\mathfrak{P}$ und $\mathfrak{N}$ bezeichnen.

Definition 9.29. a) $(\mathfrak{P}, \mathfrak{N})$ *heißt eine* ***Orientierung*** *auf* X *und* $(X, \mathfrak{P}, \mathfrak{N})$ *heißt ein* ***orientierter Vektorraum****.*
b) *Eine geordnete Basis* $\mathcal{B} = (a_1, \dots, a_n)$ *von* X *repräsentiert genau dann die* ***positive Orientierung*** *von* X*, wenn* $\Phi(a_1, \dots, a_n) > 0$ *für eine (und damit alle) Determinantenformen* Φ *aus* $\mathfrak{P}$ *gilt.*

Es ist klar, dass auf X zwei Orientierungen möglich sind: $(\mathfrak{P}, \mathfrak{N})$ und $(\mathfrak{N}, \mathfrak{P})$. Zu gegebener Basis $\mathcal{B} = (b_1, \dots, b_n)$ kann die Orientierung auf X immer so gewählt werden, dass $\mathcal{B}$ die positive Orientierung repräsentiert. Dazu definiert man Determinantenformen Φ_0 durch

$$\Phi_0(a_1, \dots, a_n) := \det(\alpha_1, \dots, \alpha_n)$$

(α_i der Koordinatenvektor von a_i bzgl. $\mathcal{B}$, $i = 1, \dots, n$) und wählt auf X die Orientierung $(\mathfrak{P}, \mathfrak{N})$, wobei $\mathfrak{P}$ die Klasse mit $\Phi_0 \in \mathfrak{P}$ ist. Es ist dann $\Phi_0(b_1, \dots, b_n) = \det E = 1 > 0$.

Beispiel 9.30. $n = 1$.

Es sei $b \neq 0$ und $\mathcal{B} = \{b\}$. Die Determinantenform Φ_0 ist durch

$$\Phi_0(a) = \alpha$$

gegeben, wobei α durch $a = \alpha b$ bestimmt ist. Jeder Vektor a mit $a = \alpha b$, $\alpha > 0$, ergibt eine Basis, welche die positive Orientierung repräsentiert. $\Diamond$

Beispiel 9.31. $n = 2$.
X sei ein zweidimensionaler reeller Vektorraum. Es sei (b_1, b_2) eine Basis von X und Φ_0 die zugehörige Determinantenform (d.h. $\Phi_0(a_1, a_2) = \det(\xi_1, \xi_2)$, wobei ξ_i der Koordinatenvektor von a_i bzgl. (b_1, b_2) ist, $i = 1, 2$). Mit der Basis (b_1, b_2) assoziieren wir durch

$$\langle a, b\rangle_1 := \alpha_1\beta_1 + \alpha_2\beta_2$$

ein inneres Produkt auf X, wobei $a = \alpha_1 b_1 + \alpha_2 b_2$, $b = \beta_1 b_1 + \beta_2 b_2$ ist (siehe Abschnitt 2.2). Ferner definieren wir $\|a\|_1 := \langle a, a\rangle_1^{1/2}$ für $a \in X$. Für Vektoren $a, b \in X$ folgt aus der Definition von Φ_0:

$$\Phi_0(a, b) = \alpha_1\beta_2 - \alpha_2\beta_1.$$

Man zeigt nun leicht

$$\Phi_0(a, b)^2 = \|a\|_1^2\|b\|_1^2 - \langle a, b\rangle_1^2, \quad a, b \in X. \tag{9.8}$$

Daraus folgt

$$\Big(\frac{\langle a, b\rangle_1}{\|a\|_1\|b\|_1}\Big)^2 + \Big(\frac{\Phi_0(a, b)}{\|a\|_1\|b\|_1}\Big)^2 = 1.$$

Es gibt daher genau ein $\theta \in (-\pi, \pi]$ derart, dass

$$\cos\theta = \frac{\langle a, b\rangle_1}{\|a\|_1\|b\|_1}, \quad \sin\theta = \frac{\Phi_0(a, b)}{\|a\|_1\|b\|_1}. \tag{9.9}$$

Der durch (9.9) eindeutig festgelegte Winkel θ heißt der ***orientierte Winkel*** von a nach b. Der orientierte Winkel vom Basiselement b_1 nach dem Basiselement b_2 ist $\pi/2$. Dies folgt aus $\langle b_1, b_2\rangle_1 = 0$ und $\Phi_0(b_1, b_2) = 1$.

Es sei nun a_1, a_2 eine beliebige geordnete Basis von X. Der orientierte Winkel von a_1 nach a_2 sei θ_0. Mit Hilfe von (9.9) folgt sofort:

> *Die Basis (a_1, a_2) repräsentiert genau dann die durch (b_1, b_2) gegebene positive Orientierung auf X, wenn*
>
> $$0 < \theta_0 < \pi$$
>
> *gilt.*

$\Diamond$

Beispiel 9.32. $n = 3$.
Es sei (e_1, e_2, e_3) eine Orthonormalbasis für die Vektoren im $\mathbb{R}^3$. In Abschnitt 2.4 haben wir gesehen, dass das Raumprodukt dreier Vektoren eine Determinantenform definiert (Satz 2.10, Aussagen a), c) und f)). Aus Satz 2.10, b), folgt sofort, dass zwei Basen genau dann dieselbe Orientierung repräsentieren, wenn sie beide zugleich ein Rechtssystem bzw. ein Linkssystem bilden. $\Diamond$

Es sei nun $(\mathcal{P}, X)$ ein n-dimensionaler affiner Raum (siehe Abschnitt 3.1). Ist auf X eine Orientierung gegeben, so nennen wir auch den affinen Raum orientiert. Es sei $\mathcal{B} = (b_1, \ldots, b_n)$ eine Basis von X, welche die positive Orientierung auf X repräsentiert. Wie wir eingangs gesehen haben, existiert eine Determinantenform Φ_0 auf X mit $\Phi_0(b_1, \ldots, b_n) = 1$.

Durch $n+1$ Punkte $P_0, P_1, \ldots, P_n \in \mathcal{P}$ wird ein Parallelflach $F(P_0, P_1, \ldots, P_n)$ bestimmt:

$$F(P_0, P_1, \ldots, P_n) = \big\{P \in \mathcal{P} \mid \overrightarrow{P_0\,P} = \sum_{i=1}^{n} \alpha_i \overrightarrow{P_0\,P_i},\ \alpha_i \geq 0,\ \sum_{i=1}^{n} \alpha_i < 1\big\}.$$

Das ***orientierte Volumen*** von $F(P_0, P_1, \ldots, P_n)$ wird durch

$$V_F := \Phi_0(\overrightarrow{P_0\,P_1}, \ldots, \overrightarrow{P_0\,P_n})$$

definiert. Unmittelbar aus der Definition folgt, dass das Volumen invariant gegenüber Translationen ist, d.h. es gilt

$$V_F = V_{F_1}$$

wobei $F = F(P_0, \ldots, P_n)$ und $F_1 = F(Q_0, \ldots, Q_n)$ mit $\overrightarrow{P_i\,Q_i} = a$, $i = 0, \ldots, n$, für einen festen Vektor $a \in X$. Denn man sieht sofort, dass $\overrightarrow{Q_0\,Q_i} = \overrightarrow{P_0\,Q_0} + \overrightarrow{P_0\,P_i} + \overrightarrow{P_i\,Q_i} = -a + \overrightarrow{P_0\,P_i} + a = \overrightarrow{P_0\,P_i}$, $i = 1, \ldots, n$. $V_F = 0$ gilt genau dann, wenn die Vektoren $\overrightarrow{P_0\,P_1}, \ldots, \overrightarrow{P_0\,P_n}$ linear abhängig sind. Repräsentieren die Vektoren $\overrightarrow{P_0\,P_1}, \ldots, \overrightarrow{P_0\,P_n}$ die positive Orientierung in X, so ist $V_F > 0$.

Beispiel 9.33. $n = 1$.
Es sei $\mathcal{B} = \{b\}$. Für zwei Punkte $P_0, P_1 \in X$ ist $F(P_0, P_1)$ die Strecke zwischen P_0 und P_1. V_F ist dann die orientierte Länge dieser Strecke, $V_F = \Phi_0(\overrightarrow{P_0\,P_1}) = \alpha\Phi_0(b) = \alpha$, wobei $\overrightarrow{P_0\,P_1} = \alpha b$ ist. $\Diamond$

Beispiel 9.34. $n = 2$.
Es sei $\mathcal{B} = (b_1, b_2)$ eine Orthonormalbasis von X. Für drei Punkte P_0, P_1, P_2 setzen wir $\overrightarrow{P_0\,P_1} = a_1$, $\overrightarrow{P_0\,P_2} = a_2$. Dann ist $F(P_0, P_1, P_2)$ das durch P_0, P_1, P_2 bestimmte Parallelogramm und V_F dessen Volumen. Es sei $(\alpha_1, \alpha_2)^\mathsf{T}$ bzw. $(\beta_1, \beta_2)^\mathsf{T}$ der Koordinatenvektor von a_1 bzw. a_2 bezüglich $\mathcal{B}$. Dann gilt

$$V_F = \Phi_0(a_1, a_2) = \det\begin{pmatrix}\alpha_1 & \beta_1\\ \alpha_2 & \beta_2\end{pmatrix} = \alpha_1\beta_2 - \alpha_2\beta_1.$$

Ist X der Raum der Vektoren im $\mathbb{R}^2$, so stimmt $|V_F|$ mit dem üblicherweise definierten Flächeninhalt überein. Denn aus (9.8) folgt sofort $\Phi_0(a_1, a_2)^2 = \|a_1\|^2\|a_2\|^2\big(1 - \cos^2\theta\big) = \|a_1\|^2\|a_2\|^2\sin^2\theta$, d.h. $|V_F| = \|a_1\|\,\|a_2\|\,|\sin\theta|$. Man beachte, dass $\|a_2\|\,|\sin\theta|$ die Höhe des Parallelogramms ist. $\Diamond$

Beispiel 9.35. $n = 3$.
Vergleiche Satz 2.10, d). $\Diamond$

Kapitel 10

Eigenwerte und Eigenvektoren

Definition 10.1. *X sei ein Vektorraum über $\mathbb{K}$. Der Endomorphismus $\varphi \in L(X,X)$ und $\lambda \in \mathbb{K}$ seien gegeben. λ heißt genau dann* ***Eigenwert*** *von φ, wenn es ein $a \in X$, $a \neq o$, gibt mit*

$$\varphi(a) = \lambda a.$$

Ist λ Eigenwert von φ, so heißt jeder Vektor $a \neq o$ mit $\varphi(a) = \lambda a$ ein ***Eigenvektor*** *von φ zu λ. Die Menge $\sigma(\varphi) := \{\lambda \in \mathbb{K} \mid \lambda \text{ ist ein Eigenwert von } \varphi\}$ heißt das* ***Spektrum*** *von φ.*

Aus der Definition folgt unmittelbar, dass $a \neq o$ genau dann Eigenvektor zu λ ist, wenn

$$a \in \ker(\varphi - \lambda\varepsilon)$$

gilt.

Für den Rest dieses Abschnittes sei $\dim X = n < \infty$. Ist A die Matrix von φ bzgl. einer geordneten Basis von X und a ein Eigenvektor von φ zum Eigenwert λ, so gilt für den Koordinatenvektor ξ von a siehe Satz 5.18)

$$A\xi = \lambda\xi.$$

Dies ist gleichbedeutend damit, dass ξ eine nichttriviale Lösung des homogenen Gleichungssystems

$$(A - \lambda E)\xi = 0$$

ist.

Ist A eine $n \times n$-Matrix über $\mathbb{K}$, so nennen wir in analoger Weise $\lambda \in \mathbb{K}$ einen ***Eigenwert*** von A, wenn es einen Vektor $\xi \neq 0$ aus $\mathbb{K}^n$ gibt mit $A\xi = \lambda\xi$. Jeder Vektor ξ mit dieser Eigenschaft heißt dann ***Eigenvektor*** von A zu λ.

Satz 10.2. *Es seien $\varphi \in L(X,X)$ und $\lambda \in \mathbb{K}$ gegeben. $\mathcal{B}$ sei eine geordnete Basis von X und es sei $A = M(\mathcal{B})$.*

a) *Die folgenden Aussagen sind äquivalent:*

(i) *λ ist Eigenwert von φ.*

(ii) $\operatorname{rang}(\varphi - \lambda\varepsilon) = \operatorname{rang}(A - \lambda E) < n$.

(iii) $\det(A - \lambda E) = 0$.

b) *Es sei $x \in X$ mit dem Koordinatenvektor ξ bzgl. $\mathcal{B}$ gegeben. Dann sind die folgenden Aussagen äquivalent:*

(i) *$x \in X$ ist Eigenvektor von φ zu λ.*

(ii) $x \in \ker(\varphi - \lambda\varepsilon)$, $x \neq o$.

(iii) ξ *ist nicht-triviale Lösung von* $(A - \lambda E)\xi = 0$.

Beweis. a) Nach Definition 10.1 ist λ genau dann Eigenwert, wenn es einen Vektor $x \neq o$ in $\ker(\varphi - \lambda\varepsilon)$ gibt. Dies ist gleichbedeutend mit $\operatorname{rang}(\varphi - \lambda\varepsilon) < n$ (vgl. Satz 5.12). Dies ist wiederum äquivalent mit $\operatorname{rang}(A - \lambda E) < n$ bzw. $\det(A - \lambda E) = 0$ (vgl. Satz 5.28 und Satz 9.2, (viii)).

Die Aussage b) folgt unmittelbar aus der Definition des Eigenvektors. □

Mit Hilfe von Satz 9.14 folgt unmittelbar, dass $\det(\lambda E - A)$ ein normiertes Polynom vom Grade n in λ ist. Denn es ist $\det(\lambda E - A) = (\lambda - \alpha_{11}) \cdot \dots \cdot (\lambda - \alpha_{nn}) +$ Terme, in denen weniger als n Faktoren der Form $\lambda - \alpha_{ii}$ vorkommen.

Definition 10.3. *A sei eine $n \times n$-Matrix über $\mathbb{K}$. Das Polynom $\chi_A(\lambda) = \det(\lambda E - A)$ heißt das **charakteristische Polynom** von A.*

Ist $\varphi \in L(X, X)$ und A die Matrix von φ bzgl. einer geordneten Basis von X, so gilt $\det(\lambda\varepsilon - \varphi) = \det(\lambda E - A)$. Das Polynom $\chi_\varphi(\lambda) = \det(\lambda\varepsilon - \varphi) = \chi_A(\lambda)$ heißt das charakteristische Polynom von φ.

Erste einfache Aussagen über das charakteristische Polynom fassen wir in dem folgenden Satz zusammen:

Satz 10.4. a) *Sind die $n \times n$-Matrizen A und $\tilde{A}$ ähnlich, so gilt*

$$\chi_A = \chi_{\tilde{A}}.$$

b) *Für jede $n \times n$-Matrix A gilt*

$$\chi_{A^T} = \chi_A.$$

c) *Es sei $\lambda^n + q_{n-1}\lambda^{n-1} + \dots + q_0$ das charakteristische Polynom der $n \times n$-Matrix A. Dann gilt:*

$$q_{n-i} = (-1)^i \sum_{1 \le j_1 < \dots < j_i \le n} A^{j_1, \dots, j_i}_{j_1, \dots, j_i},$$

d.h. $(-1)^i q_{n-i}$ ist die Summe aller Hauptminoren der Ordnung i von A. Insbesondere gilt

$$q_0 = (-1)^n \det A \quad \text{und} \quad q_{n-1} = -\operatorname{sp} A,$$

*wobei $\operatorname{sp} A := \sum_{i=1}^n \alpha_{ii}$ die **Spur** von A bezeichnet.*

Beweis. a) Aus $\tilde{A} = S^{-1}AS$ folgt auch $\lambda E - \tilde{A} = S^{-1}(\lambda E - A)S$ und (wegen Satz 9.2, (ix)) $\det(\lambda E - \tilde{A}) = \det(\lambda E - A)$.

b) Dies ist klar wegen Satz 9.4, b).

c) Es ist

$$\det(\lambda E - A) = \sum_{\pi \in \mathcal{S}_n} (\operatorname{sign} \pi) \tilde{\alpha}_{1,\pi(1)} \cdot \dots \cdot \tilde{\alpha}_{n,\pi(n)},$$

wobei $\tilde{\alpha}_{ij}$ die Elemente von $\lambda E - A$ bezeichnet,

$$\tilde{\alpha}_{ij} = \begin{cases} \lambda - \alpha_{ii} & \text{für } j = i, \\ -\alpha_{ij} & \text{für } i \neq j. \end{cases}$$

Wir erhalten Terme der Form $\gamma\lambda^{n-k}$, $k = 1, \dots, n$, genau dann, wenn wir $1 \le i_1 < i_2 < \dots < i_{n-k} \le n$ wählen und ein Permutation $\pi \in \mathcal{S}_n$ mit

$$\pi(i_1) = i_1, \dots, \pi(i_{n-k}) = i_{n-k}. \tag{10.1}$$

Ferner ist beim Ausmultiplizieren des Produktes $\tilde{\alpha}_{1\pi(1)} \cdots \cdot \tilde{\alpha}_{n,\pi(n)}$ in den Faktoren $\tilde{\alpha}_{i_\kappa,\pi(i_\kappa)} = \tilde{\alpha}_{i_\kappa,i_\kappa}$, $\kappa = 1, \ldots, n-k$, der Summand λ und in den möglicherweise noch vorhandenen Faktoren der Form $\tilde{\alpha}_{jj}$, $j \notin \{i_1, \ldots, i_{n-k}\}$, der Summand $-\alpha_{jj}$ zu berücksichtigen. Man erhält

$$(-1)^k \lambda^{n-k} (\operatorname{sign} \pi) \prod_{\substack{j=1 \\ j \notin \{i_1,\ldots,i_{n-k}\}}}^{n} \alpha_{j,\pi(j)}.$$

Für eine Permutation $\pi \in \mathcal{S}_n$ mit (10.1) ist $\pi' = \pi \mid_{\{1,\ldots,n\} \setminus \{i_1,\ldots,i_{n-k}\}}$ eine Permutation von $\{1, \ldots, n\} \setminus \{i_1, \ldots, i_{n-k}\}$ mit $\operatorname{sign} \pi' = \operatorname{sign} \pi$. Daher erhalten wir für eine feste Wahl von $1 \le i_1 < \cdots < i_{n-k} \le n$ insgesamt den Beitrag

$$(-1)^k \lambda^{n-k} \sum_{\pi'} (\operatorname{sign} \pi') \prod_{\kappa=1}^{k} \alpha_{j_\kappa,\pi'(j_\kappa)} = (-1)^k \lambda^{n-k} A^{j_1,\ldots,j_k}_{j_1,\ldots,j_k},$$

wobei die Summe über alle Permutationen von $\{j_1, \ldots, j_k\} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_{n-k}\}$ zu erstrecken ist. Berücksichtigt man noch sämtliche Möglichkeiten für die Wahl von $i_1, \ldots, i_{n-k}$, so erhält man

$$q_{n-k} = (-1)^k \sum_{1 \le j_1 < \cdots < j_k \le n} A^{j_1,\ldots,j_k}_{j_1,\ldots,j_k}.$$

Für $k = 1$ erhält man

$$q_{n-1} = -\sum_{j=1}^{n} A^j_j = -\sum_{j=1}^{n} \alpha_{jj} = -\operatorname{sp} A,$$

während $k = n$ auf

$$q_0 = (-1)^n A^{1,\ldots,n}_{1,\ldots,n} = (-1)^n \det A$$

führt. $\square$

Satz 10.5. *A sei eine $n \times n$-Matrix.*

a) *A ist genau dann regulär, wenn 0 kein Eigenwert von A ist.*

b) *Es sei A regulär. Dann ist λ genau dann ein Eigenwert von A, wenn λ^{-1} Eigenwert von A^{-1} ist.*

c) *Es sei $B = \alpha_m A^m + \cdots + \alpha_1 A + \alpha I$, $\alpha_i \in \mathbb{K}$. Ist λ Eigenwert von A und a ein Eigenvektor von A zu λ, so ist $\alpha_m \lambda^m + \cdots + \alpha_1 \lambda + \alpha_0$ Eigenwert von B und a ein zugehöriger Eigenvektor von B.*

Beweis. a) A regulär ist äquivalent zu $\det A \neq 0$. Wegen $q_0 = (-1)^n \det A$ (Satz 10.4, c)) ist dies wiederum äquivalent mit der Tatsache, dass 0 keine Nullstelle von χ_A ist.

b) Für $a \neq o$ ist $Aa = \lambda a$ wegen der Regularität von A gleichwertig mit $a = \lambda A^{-1} a$. Wegen $\lambda \neq 0$ ist die letzte Beziehung äquivalent zu $A^{-1} a = (1/\lambda) a$.

c) Aus $Aa = \lambda a$, $a \neq 0$, folgt sofort $Ba = \alpha_m A^m a + \cdots + \alpha_0 a = \alpha_m \lambda^m a + \cdots + \alpha_0 a = (\alpha_m \lambda^m + \cdots + \alpha_0) a$. $\square$

Wir überlassen es dem Leser die zu Satz 10.5 analogen Aussagen für Endomorphismen n-dimensionaler Vektorräume zu formulieren.

Da für $A \in M_{n,n}(\mathbb{K})$ das charakteristische Polynom χ_A den Grad n hat, kann A höchstens n Eigenwerte haben. χ_A muß aber nicht eine Nullstelle in $\mathbb{K}$ besitzen. Beispielsweise ist für $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in M_{2,2}(\mathbb{R})$

$$\chi_A(\lambda) = \lambda^2 + 1,$$

d.h. A besitzt als reelle Matrix keine Eigenwerte. Als Matrix in $M_{2,2}(\mathbb{C})$ besitzt A dagegen die Eigenwerte $\lambda_1 = i$, $\lambda_2 = -i$.

Es sei λ_0 Eigenwert von $A \in M_{n,n}(\mathbb{K})$. Dann ist $\chi_A(\lambda) = (\lambda - \lambda_0)p(\lambda)$, p ein Polynom von Grad $n-1$. Ist $p(\lambda_0) = 0$, so gilt $\chi_A(\lambda) = (\lambda - \lambda_0)^2 p_1(\lambda)$, grad $p_1 = n-2$. Auf diese Weise fortfahrend erhält man eine natürliche Zahl $m \leq n$ mit

$$\chi_A(\lambda) = (\lambda - \lambda_0)^m q(\lambda), \quad q(\lambda_0) \neq 0,$$

wobei $q(\lambda)$ ein Polynom vom Grad $n - m$ ist. λ_0 ist m-fache Nullstelle von $\chi_A(\lambda)$. m heißt die ***algebraische Vielfachheit*** des Eigenwertes λ_0.

Ist $\mathbb{K} = \mathbb{C}$, so gilt (wegen des Fundamentalsatzes der Algebra) für das charakteristische Polynom von $A \in M_{n,n}(\mathbb{C})$

$$\chi_A(\lambda) = (\lambda - \lambda_1)^{m_1} \cdot \dots \cdot (\lambda - \lambda_s)^{m_s},$$

wobei $\lambda_i \neq \lambda_j$ für $i \neq j$ und $m_i \geq 1$ gilt. Da grad $\chi_A = n$ ist, muß

$$m_1 + \dots + m_s = n$$

sein. Zählt man wie üblich jeden Eigenwert entsprechend seine algebraischen Vielfachheit, so kann man festhalten:

Jede Matrix $A \in M_{n,n}(\mathbb{C})$ besitzt n Eigenwerte.

Satz 10.6. *Es sei $\varphi \in L(X,X)$, $\dim X = n$, gegeben. $\lambda_1, \dots, \lambda_k \in \mathbb{K}$ seien paarweise verschiedene Eigenwerte von φ und a_i sei ein Eigenvektor von φ zu λ_i, $i = 1, ..., k$. Dann sind die Vektoren $a_1, \dots, a_k$ linear unabhängig.*

Beweis. Zunächst gilt

$$(\varphi - \lambda_i \varepsilon)(a_j) = \begin{cases} 0 \text{ für } i = j, \\ \varphi(a_j) - \lambda_i a_j = (\lambda_j - \lambda_i)a_j \text{ für } i \neq j. \end{cases}$$

Aus $\alpha_1 a_1 + \dots + \alpha_k a_k = 0$ folgt

$$\begin{aligned} 0 &= (\varphi - \lambda_2 \varepsilon) \cdot \dots \cdot (\varphi - \lambda_k \varepsilon)(\alpha_1 a_1 + \dots + \alpha_k a_k) \\ &= \alpha_1 (\lambda_1 - \lambda_2) \cdot \dots \cdot (\lambda_1 - \lambda_k) a_1 \end{aligned}$$

und wegen der Voraussetzung über die λ_j

$$\alpha_1 = 0.$$

Aus $\alpha_2 a_2 + \dots + \alpha_k a_k = 0$ erhält man analog $\alpha_2 = 0$. Nach endlich vielen Schritten hat man schließlich $\alpha_1 = \dots = \alpha_k = 0$. □

Aus Satz 10.6 erhält man sofort: Besitzt $\varphi \in L(X,X)$ genau n paarweise verschiedene Eigenwerte $\lambda_1, \dots, \lambda_n$ mit zugehörige Eigenvektoren $a_1, \dots, a_n$, so ist $\mathcal{B} = (a_1, \dots, a_n)$ eine geordnete Basis von X. Wegen $\varphi(a_i) = \lambda_i a_i$, $i = 1, \dots, n$, ist die Matrix A von φ bzgl. $\mathcal{B}$ von der Form $A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Dies bedeutet, dass jede Matrix von φ bzgl. irgendeiner Basis von X zu A ähnlich ist, d.h. diagonalisierbar ist. Es gilt allgemein der folgende Satz:

Satz 10.7. *$A \in M_{n,n}(\mathbb{K})$ sei gegeben. A ist genau dann diagonalisierbar (d.h. ähnlich zu einer Diagonalmatrix), wenn es n linear unabhängige Eigenvektoren von A gibt.*

Beweis. 1. $\xi_1, \dots, \xi_n$ seien linear unabhängige Eigenvektoren von A, $A\xi_i = \lambda_i \xi_i$, $i = 1, \dots, n$. Diese Vektoren bilden eine Basis von $\mathbb{K}^n$ (vgl. Bemerkung 4.56). A ist die Matrix der Abbildung φ_A bzgl. der kanonischen Basis von $\mathbb{K}^n$ (siehe (6.3)). Bezüglich der Basis $(\xi_1, \dots, \xi_n)$

ist die Matrix von φ_A durch

$$J = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

gegeben. Da A bzw. J jeweils die Matrix von φ_A bzgl. einer Basis des $\mathbb{K}^n$ ist, sind A und J ähnlich.
2. Es seien J und A ähnlich, d.h. es gilt

$$A = S^{-1}JS \quad \text{mit } S \text{ regulär.}$$

Bezeichnet e_i, $i = 1, \ldots, n$, den i-ten Vektor der kanonischen Basis von $\mathbb{K}^n$, so gilt $Je_i = \lambda_i e_i$, $i = 1, \ldots, n$. Daraus folgt $S^{-1}JSS^{-1}e_i = \lambda_i S^{-1}e_i$, d.h. $AS^{-1}e_i = \lambda_i S^{-1}e_i$. Die Vektoren $S^{-1}e_i$, $i = 1, \ldots, n$, sind somit Eigenvektoren von A zu den Eigenwerten λ_i. $S^{-1}e_i$ ist die i-te Spalte von S^{-1}. Da S^{-1} regulär ist, sind die Vektoren $S^{-1}e_i$, $i = 1, \ldots, n$, linear unabhängig.
□

Korollar 10.8. *$A \in M_{n,n}(\mathbb{K})$ sei gegeben. $\lambda_1, \ldots, \lambda_s$ seien die paarweise verschiedenen Eigenwerte von A, m_i die Vielfachheit von λ_i, $i = 1, \ldots, s$. Dann gilt: A ist genau dann diagonalisierbar, wenn*

(i) $m_1 + \cdots + m_s = n$,

(ii) $\operatorname{rang}(A - \lambda_i E) = n - m_i$, $i = 1, \ldots, s$.

Im Falle der Diagonalisierbarkeit ist A ähnlich zu

$$J = \operatorname{diag}(\underbrace{\lambda_1, \ldots, \lambda_1}_{m_1}, \underbrace{\lambda_2, \ldots, \lambda_2}_{m_2}, \ldots \quad \ldots, \underbrace{\lambda_s, \ldots, \lambda_s}_{m_s}).$$

Beweis. 1. A sei diagonalisierbar, d.h. $J = S^{-1}AS$, S regulär, wobei ohne Beeinträchtigung der Allgemeinheit angenommen werden kann, dass J die im Satz angegebene Form hat. Die Matrix J hat offensichtlich die paarweise verschiedenen Eigenwerte $\lambda_1, \ldots, \lambda_s$, λ_i von der Vielfachheit m_i. Es ist klar, dass $m_1 + \cdots + m_s = n$ gilt. A hat daher dieselben Eigenwerte und es gilt (i) (vgl. Satz 10.4, a)). Ferner ist

$$J - \lambda_i E = \operatorname{diag}(\underbrace{\lambda_1 - \lambda_i, \ldots, \lambda_1 - \lambda_i}_{m_1}, \ldots, \underbrace{0, \ldots, 0}_{m_i}, \ldots, \underbrace{\lambda_s - \lambda_i, \ldots, \lambda_s - \lambda_i}_{m_s}),$$

woraus $\operatorname{rang}(J - \lambda_i E) = \operatorname{rang}(A - \lambda_i E) = n - m_i$ folgt.
2. Es seien die Bedingungen (i) und (ii) erfüllt. Aus (ii) folgt, dass es m_i linear unabhängige Eigenvektoren $a_1^i, \ldots, a_{m_i}^i$ von A zum Eigenwert λ_i gibt (vgl. Satz 10.2, b)). Wegen (i) erhält man insgesamt n Eigenvektoren $a_1^i, \ldots, a_{m_i}^i$, $i = 1, \ldots, s$. Angenommen, es gilt $\sum\limits_{i=1}^{s} \sum\limits_{j=1}^{m_i} \alpha_j^i a_j^i = o$. Wir setzen $b_i = \sum\limits_{j=1}^{m_i} \alpha_j^i a_j^i$, $i = 1, \ldots, s$. Ist $b_i \neq o$, so ist b_i ein Eigenvektor von A zum Eigenwert λ_i. Aus $\sum\limits_{i=1}^{s} b_i = o$ folgt aber $b_i = o$, $i = 1, \ldots, s$ (wegen Satz 10.6). Aus $b_i = o$ folgt $\alpha_1^i = \cdots = \alpha_{m_i}^i = 0$, da die Vektoren $a_1^i, \ldots, a_{m_i}^i$ linear unabhängig sind. Damit ist gezeigt, dass die Vektoren a_j^i, $j = 1, \ldots, m_i$, $i = 1, \ldots, s$, eine Basis des $\mathbb{K}^n$ bilden. Nach Satz 10.7 ist A daher diagonalisierbar. □

Im Falle $\mathbb{K} = \mathbb{C}$ ist die Bedingung (i) von Korollar 10.8 stets erfüllt. Es sei λ_0 eine m-fache Nullstelle des charakteristischen Polynoms $\chi_A(\lambda)$, $A \in M_{n,n}(\mathbb{K})$. $k = \dim\ker(\lambda_0 E - A) = n - \operatorname{rang}(\lambda_0 E - A)$ heißt die ***geometrische Vielfachheit*** des Eigenwertes λ_0 von A. Die Bedingung (ii) des obigen Korollars bedeutet dann: Für jeden Eigenwert von A stimmen die algebraische und die geometrische Vielfachheit überein. Im allgemeinen gilt:

Ist λ_0 m-fache Nullstelle des charakteristischen Polynoms von $A \in M_{n,n}(\mathbb{K})$, so gilt

$$k = n - \operatorname{rang}(\lambda_0 E - A) \leq m.$$

Beweis. $a_1, \dots, a_k$ seien linear unabhängige Eigenvektoren von A zu λ_0. Wir ergänzen diese Vektoren zu einer Basis von $\mathbb{K}^n$: $\mathcal{B} = (a_1, \dots, a_k, a_{k+1}, \dots, a_n)$. Die Matrix des Endomorphismus φ_A (vgl. (6.3)) bezgl. $\mathcal{B}$ hat die Gestalt

$$\tilde{A} = \left(\begin{array}{ccc|c} \lambda_0 & & 0 & \\ & \ddots & & A_{12} \\ 0 & & \lambda_0 & \\ \hline & 0 & & A_{22} \end{array}\right).$$

Wegen Satz 9.2, (vii), gilt $\det(\lambda E - \tilde{A}) = (\lambda - \lambda_0)^k \chi_{A_{22}}(\lambda)$. Wegen $\det(\lambda E - A) = \det(\lambda E - \tilde{A})$, ist die algebraische Vielfachheit von λ_0 mindestens k. $\square$

Kapitel 11

Das Minimalpolynom

Es sei X ein Vektorraum über $\mathbb{K}$. Für diesen Abschnitt sei stets $\dim X = n$. Ist $p(\lambda) = \alpha_m \lambda^m + \cdots + \alpha_1 \lambda + \alpha_0$, $\alpha_i \in \mathbb{K}$, ein Polynom und $\varphi \in L(X, X)$, so kann man λ durch φ ersetzen und erhält den Endomorphismus $p(\varphi) = \alpha_m \varphi^m + \cdots + \alpha_1 \varphi + \alpha_0 \epsilon$. Ist A die Matrix von φ bzgl. irgendeiner Basis von X, so ist $p(A) = \alpha_m A^m + \cdots + \alpha_1 A + \alpha_0 E$ die Matrix von $p(\varphi)$ bzgl. derselben Basis (vgl. Abschnitt 7.2).

Satz 11.1 (Cayley[38]-Hamilton[39]). *Es sei $A \in M_{n,n}(\mathbb{K})$ gegeben. Dann gilt*

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

Beweis. Es sei $\lambda \in \mathbb{K}$ gewählt. Dann gilt nach Satz 9.22

$$(\lambda E - A)\operatorname{adj}(\lambda E - A) = \det(\lambda E - A) \cdot E = \chi_A(\lambda) \cdot E. \tag{11.1}$$

Die Elemente von $\operatorname{adj}(\lambda E - A)$ sind Polynome in λ vom Grad $\leq n-1$. Daher gilt mit konstanten $n \times n$-Matrizen $C_0, \ldots, C_{n-1}$

$$\operatorname{adj}(\lambda E - A) = C_0 + \lambda C_1 + \cdots + \lambda^{n-1} C_{n-1}.$$

Aus (11.1) folgt (mit $\chi_A(\lambda) = q_0 + q_1\lambda + \cdots + q_n\lambda^n, q_n = 1$)

$$\begin{aligned} -AC_0 + (C_0 - AC_1)\lambda + \cdots + (C_{n-2} - AC_{n-1})\lambda^{n-1} + C_{n-1}\lambda^n& \\ = (q_0 + \lambda q_1 + \cdots + q_n\lambda^n)E.& \end{aligned}$$

Durch Koeffizientenvergleich erhält man

$$\begin{aligned} AC_0 + q_0 E &= 0, \\ AC_1 - C_0 + q_1 E &= 0, \\ &\vdots \\ AC_{n-1} - C_{n-2} + q_{n-1}E &= 0, \\ -C_{n-1} + q_n E &= 0. \end{aligned}$$

[38] Cayley, Arthur, 16. 8. 1821 (Richmond) – 26. 1. 1895 (Cambridge), zunchst Rechtsanwalt, 1863 Professor in Cambridge, Beitrge zur Geometrie, Analysis, Angewandten Mathematik und Analytischen Mechanik (www.mathe.tu-freiberg.de/ hebisch/cafe/cayley.html bzw. en.wikipedia.org/wiki/Arthur_Cayley).

[39] Hamilton, William Rowan, 4. 8. 1805 (Dublin) – 2. 9. 1865 (Dublin), war Mathematiker, Physiker und Astronom, wichtige Beitrge zur Optik, Dynamik und Algebra (en.wikipedia.org/wiki/William_Rowan_Hamilton).

Durch Multiplikation der i-ten Gleichung mit A^{i-1} von rechts und nachfolgender Addition aller Gleichungen erhält man

$$0 = q_0 E + q_1 A + \cdots + q_n A^n = \chi_A(A).$$

□

Wir überlassen es dem Leser, den Satz von Cayley-Hamilton für Endomorphismen zu formulieren. Um für eine Matrix $A \in M_{n,n}(\mathbb{K})$ bzw. einen Endomorphismus $\varphi \in L(X,X)$ die Menge aller Polynome $p(\lambda)$ mit $p(A) = 0$ bzw. $p(\varphi) = 0$ zu bestimmen, holen wir etwas weiter aus.

Es sei $\varphi \in L(X,X)$ gegeben. Ferner sei $x \in X$, $x \neq o$, gewählt. Wir bilden die Vektoren

$$x, \varphi(x), \varphi^2(x), \ldots.$$

Wegen $\dim X = n$ und $x \neq o$ existiert ein p mit $1 \leq p \leq n$ derart, daß die Vektoren $x, \varphi(x), \ldots, \varphi^{p-1}(x)$ linear unabhängig und die Vektoren $x, \ldots, \varphi^p(x)$ linear abhängig sind. Daraus folgt mit eindeutig bestimmten Konstanten $\gamma_i \in \mathbb{K}$, $i = 0, \ldots, p-1$,

$$\begin{aligned} o &= \varphi^p(x) + \gamma_{p-1}\varphi^{p-1}(x) + \cdots + \gamma_1\varphi(x) + \gamma_0 x \\ &= (\varphi^p + \gamma_{p-1}\varphi^{p-1} + \cdots + \gamma_1\varphi + \gamma_0\varepsilon)x \\ &= f(\varphi)(x), \end{aligned}$$

wobei $f(\lambda) := \lambda^p + \gamma_{p-1}\lambda^{p-1} + \cdots + \gamma_0$. Für $x = o$ gilt $\varepsilon(x) = o$, d.h. wir können $f(\lambda) = 1$ setzen. Für das Polynom f gilt:

Satz 11.2. a) *f ist das eindeutig bestimmte normierte Polynom minimalen Grades mit $f(\varphi)(x) = o$. $x = o$ ist gleichbedeutend mit $f = 1$.*

b) *Ist g ein weiteres Polynom mit $g(\varphi)(x) = o$, so gilt mit einem Polynom h*

$$g = f \cdot h,$$

d.h. f teilt g.

Beweis. Es sei q ein Polynom mit $0 \leq r = \operatorname{grad} q < p = \operatorname{grad} f$ und $q(\varphi)x = o$. Wir können q als normiert annehmen. $q(\varphi)x = o$ bedeutet, daß die Vektoren $x, \varphi(x), \ldots, \varphi^r(x)$ linear abhängig sind, in Widerspruch zur Wahl von p. Damit ist $\operatorname{grad} q \geq p$ für jedes Polynom $q \neq 0$ mit $q(\varphi)(x) = o$ gezeigt.

Es sei nun g ein Polynom mit $g(\varphi)(x) = o$. Wegen $\operatorname{grad} g \geq \operatorname{grad} f$ existieren Polynome h und h_1 mit

$$g = hf + h_1$$

und $\operatorname{grad} h_1 < \operatorname{grad} f$ (siehe Satz A.7). Aus der letzten Beziehung folgt

$$o = g(\varphi)(x) = h(\varphi)f(\varphi)(x) + h_1(\varphi)(x) = h_1(\varphi)(x).$$

Aus $\operatorname{grad} h_1 < \operatorname{grad} f$ und der zu Beginn dieses Beweises bewiesenen Aussage folgt

$$h_1 = 0,$$

d.h. $g = fh$.

Ist $\tilde{f}$ ein normiertes Polynom mit $\tilde{f}(\varphi)(x) = o$ und $\operatorname{grad} \tilde{f} = \operatorname{grad} f$, so folgt $\tilde{f} = fh$ mit $\operatorname{grad} h = 0$, d.h. h ist eine Konstante. Da f und $\tilde{f}$ normiert sind, muß $h = 1$ sein. Somit gilt $\tilde{f} = f$, wodurch die Eindeutigkeit von f bewiesen ist. □

Definition 11.3. *Das in Satz* 11.2, *a), charakterisierte Polynom f heißt der* ***φ-Annullator*** *von x.*

Ist $A \in M_{n,n}(\mathbb{K})$ die Matrix von φ bzgl. einer geordneten Basis von X und ξ der Koordinatenvektor von $x \in X$, $x \neq o$, so erhält man den φ-Annullator von x wie folgt:

Man bestimmt die natürliche Zahl p so, daß die Vektoren

$$\xi, A\xi, \ldots, A^{p-1}\xi$$

linear unabhängig und die Vektoren

$$\xi, A\xi, \ldots, A^{p}\xi$$

linear abhängig sind. Dann muß

$$A^p\xi = -\gamma_{p-1}A^{p-1}\xi - \cdots - \gamma_0\xi$$

sein.

Der φ-Annullator von x ist dann (wegen Satz 5.18 und Abschnitt 7.2 ist die obige Beziehung gleichbedeutend mit $\varphi^p(x) = -\gamma_{p-1}\varphi^p(x) - \cdots - \gamma_0 x$)

$$f(\lambda) = \lambda^p + \gamma_{p-1}\lambda^{p-1} + \cdots + \gamma_0.$$

Wir nennen f auch den A-Annullator von ξ.

Es sei $a_1, \ldots, a_n$ eine geordnete Basis von X und f_i seien die φ-Annullatoren von a_i, $i = 1, \ldots, n$. Das Polynom g sei das normierte kleinste gemeinsame Vielfache der Polynome $f_1, \ldots, f_n$, d.h. es gilt $g = f_i h_i$, $i = 1, \ldots, n$, mit Polynomen h_i und aus $g_1 = f_i\tilde{h}_i$ mit Polynomen $\tilde{h}_i$, $i = 1, \ldots, n$, folgt $g_1 = gh$ (vgl. Definition A.5, c)).

Satz 11.4. a) *Das Polynom g ist das eindeutig bestimmte normierte Polynom minimalen Grades mit*

$$g(\varphi) = 0,$$

d.h. insbesonders, daß g unabhängig von der Wahl der Basis $(a_1, \ldots, a_n)$ ist.

b) *Ist $\tilde{g}$ ein Polynom mit $\tilde{g}(\varphi) = 0$, so ist*

$$\tilde{g} = gh$$

mit einem Polynom h.

Beweis. Aus $f_i(\varphi)(a_i) = o$ folgt $g(\varphi)(a_i) = h_i(\varphi)f_i(\varphi)(a_i) = h_i(\varphi)(o) = o$, $i = 1, \ldots, n$, d.h. $g(\varphi)$ ist die Nullabbildung. Es sei nun $\tilde{g} \neq 0$ ein Polynom mit $\tilde{g}(\varphi) = 0$. Daraus folgt $\tilde{g}(\varphi)(a_i) = o$, $i = 1, \ldots, n$. Nach Satz 11.2 folgt $\tilde{g} = f_i\tilde{h}_i$, $i = 1, \ldots, n$, d.h. $\tilde{g}$ ist ein gemeinsames Vielfaches der φ-Annullatoren f_i von a_i. Daraus folgt aber $\tilde{g} = gh$ mit einem Polynom h, da g ein kleinstes gemeinsames Vielfaches der f_i ist. Damit ist b) bewiesen.

Ferner folgt $\operatorname{grad}\tilde{g} > \operatorname{grad} g$, falls $\tilde{g} \neq 0$ und $\tilde{g}(\varphi) = 0$ ist. Damit ist die Minimalität des Grades von g gezeigt.

Ist $\tilde{g}$ normiert mit $\operatorname{grad}\tilde{g} = \operatorname{grad} g$ und $\tilde{g}(\varphi) = 0$, so folgt nach b) $\tilde{g} = gh$, h eine Konstante. Da $\tilde{g}$ und g normiert sind, muß $h = 1$ gelten. Damit ist auch die Eindeutigkeit von g bewiesen. □

Definition 11.5. *Das durch Satz* 11.4, *a) charakterisierte Polynom g heißt das* ***Minimalpolynom*** *von φ, $g = \mu_\varphi$.*

Ist $A \in M_{n,n}(\mathbb{K})$ die Matrix von φ bzgl. irgendeiner Basis von X, so ist μ_φ das eindeutig bestimmte normierte Polynom minimalen Grades mit $\mu_\varphi(A) = 0$. Dies ist klar, da für jedes Polynom p die Matrix $p(A)$ die Matrix von $p(\varphi)$ bzgl. derselben Basis ist.

Wir nennen μ_φ daher auch das Minimalpolynom von A und schreiben μ_A an Stelle von μ_φ.

Beispiel 11.6. Es sei

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 2 & 0 & 0 \\ -1 & 1 & 2 & 1 \\ -1 & 1 & 0 & 3 \end{pmatrix} \in M_{n,n}(\mathbb{R})$$

gegeben. Wir berechnen zunächst die A-Annullatoren für $e_1, \ldots, e_4$:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad Ae_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \end{pmatrix}, \quad A^2 e_1 = \begin{pmatrix} 0 \\ 0 \\ -4 \\ -4 \end{pmatrix} = 4Ae_1 - 4e_1.$$

Wir sehen, daß e_1, Ae_1 linear unabhängig sind. Andererseits gilt $A^2e_1 - 4Ae_1 + 4e_1 = (A^2 - 4A + 4E)e_1 = 0$. Der A-Annullator f_1 von e_1 ist daher

$$f_1(\lambda) = \lambda^2 - 4\lambda + 4 = (\lambda - 2)^2.$$

Für e_2 erhalten wir:

$$e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad Ae_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}, \quad A^2 e_2 = \begin{pmatrix} 4 \\ 4 \\ 4 \\ 4 \end{pmatrix} = 4Ae_2 - 4e_2,$$

d.h. $(A^2 - 4A + 4E)e_2 = 0$ und

$$f_2(\lambda) = (\lambda - 2)^2.$$

Für e_3 gilt:

$$e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad Ae_3 = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \end{pmatrix} = 2e_3,$$

d.h. $(A - 2E)e_3 = 0$ und

$$f_3(\lambda) = \lambda - 2.$$

Schließlich erhalten wir für e_4:

$$e_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \quad Ae_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 3 \end{pmatrix}, \quad A^2 e_4 = \begin{pmatrix} 4 \\ 0 \\ 4 \\ 8 \end{pmatrix} = 4Ae_4 - 4e_4,$$

d.h. $(A^2 - 4A + 4E)e_4 = 0$ und

$$f_4(\lambda) = (\lambda - 2)^2.$$

Das normierte kleinste gemeinsame Vielfache der Polynome $f_1, \ldots, f_4$ ist das Minimalpolynom von A,

$$\mu_A(\lambda) = (\lambda - 2)^2.$$

$\Diamond$

Über den Zusammenhang von Minimalpolynom und charakteristischem Polynom gilt:

Satz 11.7. *Es sei $A \in M_{n,n}(\mathbb{K})$. Dann gilt:*

a) *Das Minimalpolynom μ_A ist Teiler des charakteristischen Polynoms χ_A.*

b) *Jede Nullstelle λ_0 von χ_A ist auch Nullstelle von μ_A.*

Beweis. Ist klar wegen Satz 11.1 und Satz 11.4, b). Ist λ_0 Nullstelle von χ_A, so existiert ein $\xi \neq 0$ mit $A\xi = \lambda_0 \xi$ bzw. $(A - \lambda_0 E)\xi = 0$, d.h. $\lambda - \lambda_0$ ist der A-Annullator von ξ. Da trivialerweise $\mu_A(A)\xi = 0$ gilt, folgt wegen Satz 11.2, b), mit einem Polynom h

$$\mu_A(\lambda) = (\lambda - \lambda_0)h(\lambda).$$

□

Korollar 11.8. *Es sei $\mathbb{K} = \mathbb{C}$ und $A \in M_{n,n}(\mathbb{C})$. Dann hat das charakteristische Polynom χ_A die Gestalt*

$$\chi_A(\lambda) = (\lambda - \lambda_1)^{m_1} \cdot \dots \cdot (\lambda - \lambda_s)^{m_s},$$

wobei $\lambda_1, \dots, \lambda_s$ die paarweise verschiedenen Eigenwerte von A sind und $m_i \geq 1$, $m_1 + \dots + m_s = n$ gilt. Das Minimalpolynom μ_A hat dann die Gestalt:

$$\mu_A(\lambda) = (\lambda - \lambda_1)^{k_1} \cdot \dots \cdot (\lambda - \lambda_s)^{k_s}$$

mit $1 \leq k_i \leq m_i$, $i = 1, \dots, s$.

Beweis. Da μ_A Teiler von χ_A ist, folgt zunächst, daß μ_A die angegebene Form mit $0 \leq k_i \leq m_i$ hat. Wegen Satz 11.7, b) kann jedoch nicht $k_i = 0$ sein. □

Kapitel 12

Die Jordansche Normalform

Für diesen Abschnitt setzen wir generell voraus, daß $\mathbb{K} = \mathbb{C}$ gilt. Dies garantiert, daß jedes Polynom als Produkt von Potenzen linearer Faktoren geschrieben werden kann. Ferner sei X stets ein linearer Raum über $\mathbb{C}$ mit $\dim X = n$.

Ist $x \in X$ ein Eigenvektor von $\varphi \in L(X, X)$ so gilt $\varphi([x]) \subset [x] = \{\alpha x \mid \alpha \in \mathbb{C}\}$. Die lineare Hülle ist ein invarianter Unterraum im Sinne der folgenden Definition:

Definition 12.1. *Es sei $\varphi \in L(X, X)$ und U ein Unterraum von X. U heißt genau dann **φ-invariant**, wenn $\varphi(U) \subset U$ gilt.*

Ist U ein φ-invarianter Unterraum, so ist die Einschränkung von φ auf U ein Endomorphismus von U, $\varphi|_U \in L(U, U)$.

Die Bedeutung φ-invarianter Unterräume ersieht man aus folgendem Satz:

Satz 12.2. *Es sei $\varphi \in L(X, X)$ und U ein φ-invarianter Unterraum von X. $\mathcal{B} = (a_1, \ldots, a_k, a_{k+1}, \ldots, a_n)$ sei eine geordnete Basis von X derart, daß $(a_1, \ldots, a_k)$ eine Basis von U ist (d.h. eine Basis von U wird zu einer Basis von X ergänzt). Die Matrix A von φ bzgl. $\mathcal{B}$ hat die Gestalt*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

wobei A_{11} eine $k \times k$-Matrix, A_{12} eine $k \times (n-k)$-Matrix und A_{22} eine $(n-k) \times (n-k)$-Matrix ist. Die Matrix A_{11} ist die Matrix von $\varphi|_U$ bzgl. der Basis $(a_1, \ldots, a_k)$ von U.

Beweis. Da wegen der φ-Invarianz von U die Vektoren $\varphi(a_i) \in U$ sind, $i = 1, \ldots, k$, folgt

$$\varphi(a_i) = \alpha_{1i} a_1 + \cdots + \alpha_{ki} a_k + 0 \cdot a_{k+1} + \cdots + 0 \cdot a_k,$$

$i = 1, \ldots, k$. Da $(\alpha_{1i}, \ldots, \alpha_{ki}, 0, \ldots, 0)^{\mathsf{T}}$ der i-te Spaltenvektor von A ist, folgt die Behauptung über die Gestalt von A. Ferner ist $(\alpha_{1i}, \ldots, \alpha_{ki})^{\mathsf{T}}$, $i = 1, \ldots, k$, der i-te Spaltenvektor der Matrix von $\varphi|_U$ bzgl. der Basis $(a_1, \ldots, a_k)$, d.h. A_{11} ist die Matrix von $\varphi|_U$ bzgl. $(a_1, \ldots, a_k)$.
□

Korollar 12.3. *Es sei $\varphi \in L(X, X)$ und U_1, U_2 seien φ-invariante Unterräume von X mit $X = U_1 \oplus U_2$. $\mathcal{B}_1 = (a_1, \ldots, a_k)$ bzw. $\mathcal{B}_2 = (a_{k+1}, \ldots, a_n)$ seien geordnete Basen von U_1 bzw. U_2. Die Matrix A von φ bzgl. der Basis $\mathcal{B} = (a_1, \ldots, a_k, a_{k+1}, \ldots, a_n)$ von X hat die Gestalt*

$$A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix},$$

wobei die $k \times k$-Matrix A_{11}, die Matrix von $\varphi\,|_{U_1}$ bzgl. $\mathcal{B}_1$ und die $(n-k) \times (n-k)$-Matrix A_{22} die Matrix von $\varphi\,|_{U_2}$ bzgl. $\mathcal{B}_2$ ist.

Beweis. Nach Satz 12.2 hat A die dort angegebene Gestalt. Wegen der φ-Invarianz von U_2 gilt

$$\varphi(a_i) = 0a_1 + \cdots + 0a_k + \alpha_{k+1,i}a_{k+1} + \cdots + \alpha_{n,i}a_n,$$

$i = k+1, \ldots, n$. Daraus folgt $A_{12} = 0$ und die Tatsache, daß A_{22} die Matrix von $\varphi\,|_{U_2}$ bzgl. $\mathcal{B}_2$ ist. $\square$

Lemma 12.4. *Es sei $\varphi \in L(X,X)$ und $g \in \mathbb{C}[t]$ normiert. Gilt $g = g_1g_2$ mit normierten teilerfremden Polynomen g_1, g_2, so folgt*

$$\ker g(\varphi) = \ker g_1(\varphi) \oplus \ker g_2(\varphi).$$

Beweis. Aus $g_1(\varphi)(x) = o$ bzw. $g_2(\varphi)(x) = o$ folgt $g(\varphi)(x) = g_2(\varphi)g_1(\varphi)(x) = g_1(\varphi)g_2(\varphi)(x) = o$, d.h.

$$\ker g_i(\varphi) \subset \ker g(\varphi), \quad i = 1, 2.$$

Es sei nun $x \in \ker g_1(\varphi) \cap \ker g_2(\varphi)$, d.h. $g_1(\varphi)(x) = o = g_2(\varphi)(x)$. Ist f der φ-Annullator von x so folgt nach Satz 11.2, b),

$$f \mid g_1 \text{ und } f \mid g_2.$$

Da g_1, g_2 teilerfremd sind muß $f = 1$ gelten, woraus $x = o$ folgt (Satz 11.2, a)). Damit ist

$$\ker g_1(\varphi) \cap \ker g_2(\varphi) = \{o\}$$

gezeigt.

Es sei nun $x \in \ker g(\varphi)$. Da g_1, g_2 teilerfremd sind, gilt mit Polynomen h, k (Satz A.8, b))

$$1 = hg_1 + kg_2.$$

Ersetzt man in dieser Beziehung die Unbestimmte durch φ, so erhält man

$$\varepsilon = h(\varphi)g_1(\varphi) + k(\varphi)g_2(\varphi),$$

woraus

$$x = h(\varphi)g_1(\varphi)(x) + k(\varphi)g_2(\varphi)(x)$$

folgt. Wir setzen $x_1 = k(\varphi)g_2(\varphi)(x)$ und $x_2 = h(\varphi)g_1(\varphi)(x)$. Dann gilt

$$\begin{aligned} g_1(\varphi)(x_1) &= g_1(\varphi)k(\varphi)g_2(\varphi)(x) = k(\varphi)g(\varphi)(x) = o, \\ g_2(\varphi)(x_2) &= g_2(\varphi)h(\varphi)g_1(\varphi)(x) = h(\varphi)g(\varphi)(x) = o, \end{aligned}$$

d.h. $x_1 \in \ker g_1(\varphi)$, $x_2 \in \ker g_2(\varphi)$. Damit ist $x = x_1 + x_2$ mit $x_i \in \ker g_i(\varphi)$ gezeigt. $\square$

Satz 12.5 (1. Zerlegungssatz). *Es sei $\varphi \in L(X,X)$ mit dem Minimalpolynom $\mu_\varphi(\lambda) = (\lambda - \lambda_1)^{k_1} \ldots (\lambda - \lambda_s)^{k_s}$ gegeben. Hierbei sind $\lambda_1, \ldots, \lambda_s$ die paarweise verschiedenen Eigenwerte von φ. Dann gilt:*

a) $X_i = \ker(\varphi - \lambda_i\varepsilon)^{k_i} \neq \{o\}$, $i = 1, \ldots, s$.

b) X_i *ist φ-invariant,* $i = 1, \ldots, s$.

c) $X = X_1 \oplus \cdots \oplus X_s$.

d) $(\lambda - \lambda_i)^{k_i}$ *ist das Minimalpolynom von* $\varphi_i = \varphi\,|_{X_i}$, $i = 1, \ldots, s$.

e) $\dim X_i = m_i$, *m_i die algebraische Vielfachheit von λ_i, $i = 1, \ldots, s$.*
f) *Für $i = 1, \ldots, s$ gilt:*

$$\{o\} \underset{\neq}{\subseteq} \ker(\varphi - \lambda_i \varepsilon) \underset{\neq}{\subseteq} \cdots \underset{\neq}{\subseteq} \ker(\varphi - \lambda_i \varepsilon)^{k_i} = \ker(\varphi - \lambda_i \varepsilon)^{k_i + \nu},$$

und

$$n > \operatorname{rang}(\varphi - \lambda_i \varepsilon) > \cdots > \operatorname{rang}(\varphi - \lambda_i \varepsilon)^{k_i} = \operatorname{rang}(\varphi - \lambda_i \varepsilon)^{k_i + \nu},$$

$\nu = 1, 2, \ldots$.

Beweis. a) Für jeden Eigenvektor x zum Eigenwert λ_i gilt $(\varphi - \lambda_i \varepsilon)x = o$ und daher auch $\ker(\varphi - \lambda_i \varepsilon)^{k_i} \neq \{o\}$.
b) Es sei $x \in X_i$. Dann gilt $(\varphi - \lambda_i \varepsilon)^{k_i} \varphi(x) = \varphi(\varphi - \lambda_i \varepsilon)^{k_i}(x) = \varphi(o) = o$, d.h. $\varphi(x) \in X_i$.
c) Da $(\lambda - \lambda_1)^{k_1}$ und $g_1(\lambda) = (\lambda - \lambda_2)^{k_2} \cdot \ldots \cdot (\lambda - \lambda_s)^{k_s}$ teilerfremd sind, folgt wegen Lemma 12.4

$$\begin{aligned} X = \ker \mu_\varphi(\varphi) &= \ker(\varphi - \lambda_1 \varepsilon)^{k_1} \oplus \ker g_1(\varphi) \\ &= X_1 \oplus \ker g_1(\varphi). \end{aligned}$$

Analog erhält man

$$\ker g_1(\varphi) = \ker(\varphi - \lambda_2 \varepsilon)^{k_2} \oplus \ker g_2(\varphi)$$

mit $g_2(\lambda) = (\lambda - \lambda_3)^{k_3} \cdot \ldots \cdot (\lambda - \lambda_s)^{k_s}$, d.h.

$$X = X_1 \oplus X_2 \oplus \ker g_2(\varphi).$$

Nach endlich vielen Schritten erhält man $X = X_1 \oplus \cdots \oplus X_s$.
d) Da X_i φ-invariant ist, gilt $\varphi_i = \varphi\,|_{X_i} \in L(X_i, X_i)$. Es sei μ_i das Minimalpolynom von φ_i. Aus $(\varphi - \lambda_i \varepsilon)^{k_i}(X_i) = \{o\}$ und $(\varphi_i - \lambda_i \varepsilon)^{k_i}(X_i) = (\varphi - \lambda_i \varepsilon)^{k_i}(X_i)$ folgt, daß $(\varphi_i - \lambda_i \varepsilon)^{k_i} = 0$ ($=$ Nullabbildung auf X_i) ist. Nach Satz 11.4, b), gilt

$$\mu_i \mid (\lambda - \lambda_i)^{k_i},$$

d.h. $\mu_i = (\lambda - \lambda_i)^{\tilde{k}_i}$ mit $\tilde{k}_i \leq k_i$. Angenommen, es sei $\tilde{k}_i < k_i$. Wir setzen

$$\tilde{\mu}(\lambda) = (\lambda - \lambda_1)^{k_1} \cdot \ldots \cdot (\lambda - \lambda_i)^{\tilde{k}_i} \cdot \ldots \cdot (\lambda - \lambda_s)^{k_s}.$$

Für $x \in X$ gibt es wegen Teil c) des Beweises eindeutig bestimmte Elemente $x_j \in X_j$, $j = 1, \ldots, s$, mit

$$x = x_1 + \cdots + x_s.$$

Daraus folgt

$$\tilde{\mu}(\varphi)(x) = \sum_{j=1}^{s} \tilde{\mu}(\varphi)(x_j).$$

Aus $(\varphi - \lambda_j \varepsilon)^{k_j}(x_j) = o$, $j = 1, \ldots, s$, $j \neq i$, und $(\varphi - \lambda_i \varepsilon)^{\tilde{k}_i}(x_i) = (\varphi_i - \lambda_i \varepsilon)^{\tilde{k}_i}(x_i) = o$ folgt $\tilde{\mu}(\varphi)(x) = 0$. Da x beliebig aus X war, folgt $\tilde{\mu}(\varphi) = 0$. Wegen $\tilde{k}_i < k_i$ folgt $\operatorname{grad} \tilde{\mu} < \operatorname{grad} \mu_\varphi$ in Widerspruch zur Tatsache, daß μ_φ das Minimalpolynom von φ ist.
e) Die Aussagen a) – c) des Satzes zeigen, daß X die direkte Summe der nicht-trivialen φ-invarianten Unterräume X_i ist. Wählt man für jedes X_i eine geordnete Basis $\mathcal{B}_i$ dann ist $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_s)$ eine geordnete Basis von X. Nach dem Korollar 12.3 hat die Matrix A von φ bzgl. $\mathcal{B}$ die Gestalt:

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_s \end{pmatrix},$$

wobei A_i eine $(\dim X_i) \times (\dim X_i)$-Matrix ist, $i = 1, \ldots, s$.

Da $(\lambda - \lambda_i)^{k_i}$ das Minimalpolynom von φ_i ist, muß das charakteristische Polynom von φ_i und damit auch von A_i die Gestalt $(\lambda - \lambda_i)^{\tilde{k}_i}$ mit $\tilde{k}_i \geq k_i$ haben (Satz 11.7). Aus der speziellen Gestalt von A folgt (Satz 9.2, (vii))

$$\chi_\varphi(\lambda) = \chi_A(\lambda) = \chi_{A_1}(\lambda) \cdot \cdots \cdot \chi_{A_s}(\lambda) = (\lambda - \lambda_1)^{\tilde{k}_1} \cdot \cdots \cdot (\lambda - \lambda_s)^{\tilde{k}_s}.$$

Andererseits ist

$$\chi_A = (\lambda - \lambda_1)^{m_1} \cdot \cdots \cdot (\lambda - \lambda_s)^{m_s},$$

woraus $\tilde{k}_i = m_i$, $i = 1, \ldots, s$, folgt, d.h. $\chi_{A_i}(\lambda) = (\lambda - \lambda_i)^{m_i}$. A_i ist daher eine $m_i \times m_i$-Matrix, woraus $\dim X_i = m_i$ folgt.

f) $\{o\} \subsetneqq \ker(\varphi - \lambda_i\varepsilon)$ ist bereits unter Teil a) bewiesen. Wir setzen $\psi = \varphi - \lambda_i\varepsilon$. Aus $\psi^k(x) = o$ folgt stets $\psi^{k+1}(x) = \psi(\psi^k(x)) = o$, d.h. es gilt immer $\ker\psi^k \subset \ker\psi^{k+1}$.

Es sei nun $\ker\psi^m = \ker\psi^{m+1}$. Für $x \in \ker\psi^{m+2}$ gilt $o = \psi^{m+2}(x) = \psi^{m+1}(\psi(x))$, d.h. $\psi(x) \in \ker\psi^{m+1}$. Wegen $\ker\psi^{m+1} = \ker\psi^m$ folgt $\psi(x) \in \ker\psi^m$, d.h. $o = \psi^m(\psi(x)) = \psi^{m+1}(x)$. Damit ist $x \in \ker\psi^{m+1}$ gezeigt. Es ist somit $\ker\psi^{m+2} \subset \ker\psi^{m+1}$. Da die umgekehrte Inklusion stets gilt, folgt $\ker\psi^{m+2} = \ker\psi^{m+1}$. In analoger Weise fortfahrend sieht man:

Aus $\ker\psi^m = \ker\psi^{m+1}$ *folgt* $\ker\psi^{m+\nu} = \ker\psi^m$
für $\nu = 1, 2, \ldots$.

Es sei nun $\ker\psi^m = \ker\psi^{m+1}$ für ein $m < k_i$. Daraus folgt $\ker\psi^m = \ker\psi^{k_i} = X_i$, d.h. $\{o\} = (\varphi - \lambda_i\varepsilon)^m(X_i) = (\varphi_i - \lambda_i\varepsilon)^m(X_i)$. Es ist somit $(\varphi_i - \lambda_i\varepsilon)^m = 0$. Nach Satz 11.4, b), müßte das Minimalpolynom $(\lambda - \lambda_i)^{k_i}$ von φ_i (siehe Teil d) des Beweises) Teiler von $(\lambda - \lambda_i)^m$ sein. Dieser Widerspruch zeigt, daß $m \geq k_i$ sein muß.

Es sei $\ker\psi^{k_i} \subsetneqq \ker\psi^{k_i+1}$. Wir wählen $y \in \ker\psi^{k_i+1} \setminus \ker\psi^{k_i}$, d.h.

$$(\varphi - \lambda_i\varepsilon)^{k_i+1}(y) = o \quad \text{und} \quad (\varphi - \lambda_i\varepsilon)^{k_i}(y) \neq o.$$

Die erste Gleichung zeigt, daß der φ-Annullator von y ein Teiler von $(\lambda - \lambda_i)^{k_i+1}$ sein muß. Wegen der zweiten Gleichung muß $(\lambda - \lambda_i)^{k_i+1}$ der φ-Annullator von y sein. Da trivialerweise $\mu_\varphi(\varphi)(y) = o$ ist, müßte nach Satz 11.2, b),

$$(\lambda - \lambda_i)^{k_i+1} \mid \mu_\varphi$$

gelten. Dieser Widerspruch beweist $\ker\psi^{k_i} = \ker\psi^{k_i+1}$. Damit ist die Aussage über die Kerne von $(\varphi - \lambda_i\varepsilon)^j$ vollständig bewiesen.

Aus $\ker\psi^j \subsetneqq \ker\psi^{j+1}$ folgt wegen Satz 5.12

$$\operatorname{rang}\psi^j = n - \dim(\ker\psi^j) > n - \dim(\ker\psi^{j+1}) = \operatorname{rang}\psi^{j+1}.$$

Damit ist auch die Aussage über $\operatorname{rang}\psi^j$ bewiesen. □

Die Aussage f) aus Satz 12.5 charakterisiert den Exponenten k_i des Linearfaktors $\lambda - \lambda_i$ im Minimalpolynom:

k_i *ist die kleinste natürliche Zahl mit der Eigenschaft*
$\operatorname{rang}(\varphi - \lambda_i\varepsilon)^{k_i} = \operatorname{rang}(\varphi - \lambda_i\varepsilon)^{k_i+1}$.

Ob eine Matrix diagonalisierbar ist, kann auch an Hand des Minimalpolynomes entschieden werden:

Satz 12.6. *Die Matrix $A \in M_{n,n}(\mathbb{C})$ ist genau dann diagonalisierbar, wenn das Minimalpolynom μ_A die Gestalt*

$$\mu_A(\lambda) = (\lambda - \lambda_1) \cdot \cdots \cdot (\lambda - \lambda_s)$$

hat (d.h. $k_i = 1,\ i = 1, \ldots, s$*).*

Beweis. Das Minimalpolynom habe die angegebene Gestalt. Die Unterräume X_i aus Satz 12.5 sind in diesem Fall durch $X_i = \ker(\varphi - \lambda_i \varepsilon)$ gegeben. Jeder Vektor $x \neq o$ aus X_i ist somit Eigenvektor von φ zum Eigenwert λ_i. Wegen $\dim X_i = m_i$ ist somit die geometrische Vielfachheit von λ_i gleich der algebraischen Vielfachheit, $i = 1, \ldots, s$. Nach dem Korollar 10.8 (siehe auch die Bemerkungen in Anschluß an den Beweis dieses Korollars) ist A diagonalisierbar.

Es sei nun A diagonalisierbar. Nach Satz 10.7 existiert eine Basis des $\mathbb{C}^n$ bestehend aus Eigenvektoren $\xi_1, \ldots, \xi_n$ von A. Es sei ξ_i Eigenvektor zum Eigenwert λ_0. Der A-Annullator von ξ_i ist wegen $A\xi_i - \lambda_0 \xi_i = 0$ durch $\lambda - \lambda_0$ gegeben. Da das Minimalpolynom von A das normierte kleinste gemeinsame Vielfache der A-Annullatoren der ξ_i ist, folgt die angegebene Gestalt von μ_A. □

Um zur Jordan-Normalform zu kommen, müssen wir noch für die Unterräume X_i Basen $\mathcal{B}_i$ so wählen, daß die Matrizen A_i von $\varphi_i = \varphi\,|_{X_i}$ bzgl. $\mathcal{B}_i$ eine möglichst einfache Gestalt haben. Nach Satz 12.5, d), ist $(\lambda - \lambda_i)^{k_i}$ das Minimalpolynom von φ_i bzw. A_i. Es genügt daher, die folgende Situation zu betrachten:

$$\text{(V)} \qquad \begin{gathered} \text{Für } \varphi \in L(X,X), \dim X = n, \text{ sei} \\ \mu_\varphi(\lambda) = (\lambda - \lambda_0)^k \\ \text{das Minimalpolynom.} \end{gathered}$$

Definition 12.7. *Es liege die durch* (V) *beschriebene Situation vor. Ferner sei* $a \in X$ *gegeben. Die Folge der Vektoren*

$$a, (\varphi - \lambda_0 \varepsilon)(a), \ldots, (\varphi - \lambda_0 \varepsilon)^{\ell - 1}(a)$$

heißt genau dann die von a *erzeugte* ***Jordankette***[40]*, wenn* $(\varphi - \lambda_0 \varepsilon)^{\ell-1}(a) \neq o$ *und* $(\varphi - \lambda_0 \varepsilon)^{\ell}(a) = o$ *gilt. Die Zahl* ℓ *heißt die* ***Länge*** *der Jordankette.*

Da $(\varphi - \lambda_0 \varepsilon)^k(a) = o$ für alle $a \in X$ gilt ($(\lambda - \lambda_0)^k$ ist das Minimalpolynom von φ!), folgt stets $\ell \leq k$. Die Wichtigkeit von Jordanketten zeigt der folgende Satz:

Satz 12.8. *Es sei* (V) *erfüllt und* $a \in X$ *erzeuge eine Jordankette der Länge* $\ell \leq k$*. Dann gilt:*

a) $(\lambda - \lambda_0)^\ell$ *ist der* φ*-Annullator von* a *und die Vektoren*

$$a, (\varphi - \lambda_0 \varepsilon)(a), \ldots, (\varphi - \lambda_0 \varepsilon)^{\ell - 1}(a)$$

sind linear unabhängig.

b) *Der Unterraum* $U = [a, (\varphi - \lambda_0 \varepsilon)(a), \ldots, (\varphi - \lambda_0 \varepsilon)^{\ell-1}(a)]$ *ist* φ*-invariant.* $(\lambda - \lambda_0)^\ell$ *ist das Minimalpolynom von* $\varphi\,|_U$.

c) *Die Matrix* A *von* $\varphi\,|_U$ *bzgl. der geordneten Basis* $\mathcal{B} = (a, (\varphi - \lambda_0 \varepsilon)(a), \ldots$ $\ldots, (\varphi - \lambda_0 \varepsilon)^{\ell-1}(a))$ *hat die Gestalt:*

$$A = \begin{pmatrix} \lambda_0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda_0 \end{pmatrix} \in M_{\ell,\ell}(\mathbb{C}).$$

[40] Jordan, Marie Ennemond Camille, 5. 1. 1838 (Lyons) – 22. 1. 1922 (Paris), grundlegende Arbeiten in Gruppentheorie und Verfasser eines einflussreichen Lehrbuches "Cours d'analyse" (en.wikipedia.org/wiki/Camille_Jordan).

d) *Aus $U = U_1 \oplus U_2$ mit φ-invarianten Unterräumen U_1, U_2 folgt $U_1 = \{o\}$ oder $U_2 = \{o\}$, d.h. U kann nicht mehr als direkte Summe von φ-invarianten nicht-trivialen Unterräumen dargestellt werden.*

Beweis. a) Es sei p der φ-Annullator von a. Nach Satz 11.2, b), gilt $p \mid (\lambda - \lambda_0)^\ell$, d.h. $p = (\lambda - \lambda_0)^{\tilde{\ell}}$ mit $\tilde{\ell} \leq \ell$. Wegen $(\varphi - \lambda_0 \varepsilon)^{\ell-1}(a) \neq o$ gilt $\tilde{\ell} = \ell$.

Es sei nun

$$o = \alpha_0 a + \alpha_1 (\varphi - \lambda_0 \varepsilon)(a) + \cdots + \alpha_{\ell-1}(\varphi - \lambda_0 \varepsilon)^{\ell-1}(a) = q(\varphi)(a)$$

mit $q(\lambda) = \alpha_0 + \alpha_1(\lambda - \lambda_0) + \cdots + \alpha_{\ell-1}(\lambda - \lambda_0)^{\ell-1}$. Wegen $\operatorname{grad} q < \ell$ und der Tatsache, daß $(\lambda - \lambda_0)^\ell$ der φ-Annullator von a ist, folgt $q \equiv 0$, d.h. $\alpha_0 = \cdots = \alpha_{\ell-1} = 0$. Die Vektoren der von a erzeugten Jordankette sind somit linear unabhängig.

b) Aus

$$\begin{aligned}\varphi\big((\varphi - \lambda_0 \varepsilon)^j(a)\big) &= (\varphi - \lambda_0 \varepsilon)(\varphi - \lambda_0 \varepsilon)^j(a) + \lambda_0(\varphi - \lambda_0 \varepsilon)^j(a)\\ &= (\varphi - \lambda_0 \varepsilon)^{j+1}(a) + \lambda_0(\varphi - \lambda_0 \varepsilon)^j(a), \quad j = 0, \ldots, \ell-2,\end{aligned}$$

und

$$\begin{aligned}\varphi\big((\varphi - \lambda_0 \varepsilon)^{\ell-1}(a)\big) &= (\varphi - \lambda_0 \varepsilon)^\ell(a) + \lambda_0(\varphi - \lambda_0 \varepsilon)^{\ell-1}(a)\\ &= \lambda_0(\varphi - \lambda_0 \varepsilon)^{\ell-1}(a)\end{aligned}$$

folgt die φ-Invarianz von U und die unter c) angegebene Gestalt von A. Man beachte, daß für $j = 0, \ldots, \ell - 2$ der Koordinatenvektor von $\varphi\big((\varphi - \lambda_0 \varepsilon)^j(a)\big)$ bzgl. der Basis $\mathcal{B}$ durch $(0, \ldots, 0, \lambda_0, 1, 0, \ldots, 0)^\mathsf{T}$ gegeben ist, wobei λ_0 an der j-ten Stelle steht, und der Koordinatenvektor von $\varphi\big((\varphi - \lambda_0 \varepsilon)^{\ell-1}(a)\big)$ durch $(0, \ldots, 0, \lambda_0)^\mathsf{T}$.

c) Aus $(\varphi - \lambda_0 \varepsilon)^\ell(a) = o$ folgt sofort $(\varphi - \lambda_0 \varepsilon)^\ell(U) = \{o\}$, d.h. $\mu_{\varphi|_U}$ ist ein Teiler von $(\lambda - \lambda_0)^\ell$. Wegen $(\varphi - \lambda_0 \varepsilon)^{\ell-1}(a) \neq o$ folgt $\mu_{\varphi|_U} = (\lambda - \lambda_0)^\ell$.

d) Angenommen, es sei $U_1 \neq \{o\}$ und $U_2 \neq \{o\}$, d.h. $\dim U_i = \ell_i$ mit $1 \leq \ell_i < \ell$, $i = 1, 2$. Wegen $U = U_1 \oplus U_2$ gibt es eindeutig bestimmte Vektoren $a_i \in U_i$ mit $a = a_1 + a_2$.

Wegen der φ-Invarianz von U_i gilt $\varphi^\nu(a_i) \in U_i$, $\nu = 0, 1, 2, \ldots$. Daraus folgt, daß für den φ-Annullator p_i von a_i gilt:

$$\operatorname{grad} p_i \leq \ell_i.$$

p_i teilt das Minimalpolynom $(\lambda - \lambda_0)^\ell$ von $\varphi\,|_U$. Daraus folgt

$$p_i(\lambda) = (\lambda - \lambda_0)^{\tilde{\ell}_i} \text{ mit } \tilde{\ell}_i \leq \ell_i.$$

Es sei $\rho = \max(\tilde{\ell}_1, \tilde{\ell}_2)$. Dann gilt $\rho < \ell$ und

$$(\varphi - \lambda_0 \varepsilon)^\rho(a) = (\varphi - \lambda_0 \varepsilon)^\rho(a_1) + (\varphi - \lambda_0 \varepsilon)^\rho(a_2) = o$$

in Widerspruch zur Tatsache, daß $(\lambda - \lambda_0)^\ell$ der φ-Annullator von a ist. $\square$

Wählt man $\tilde{\mathcal{B}} = \big((\varphi - \lambda_0 \varepsilon)^{\ell-1}(a), \ldots, (\varphi - \lambda_0 \varepsilon)(a), a\big)$ als Basis von U, so ist die Matrix von $\varphi\,|_U$ bzgl. $\tilde{\mathcal{B}}$ durch

$$A = \begin{pmatrix} \lambda_0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_0 \end{pmatrix} \in M_{\ell,\ell}(\mathbb{C})$$

gegeben. Dies folgt unmittelbar aus dem Beweis zu Aussage b) von Satz 12.8.

Definition 12.9. *U sei ein Unterraum von X und $a_1, \dots, a_k \in X$ seien gegeben. Die Vektoren $a_1, \dots, a_k$ heißen genau dann **linear unabhängig relativ zu** U, wenn aus $\alpha_1 a_1 + \cdots + \alpha_k a_k \in U$ mit $\alpha_i \in \mathbb{C}$ stets $\alpha_1 = \cdots = \alpha_k = 0$ folgt.*

Sind $a_1, \dots, a_k$ linear unabhängig relativ zu U, so sind $a_1, \dots, a_k$ linear unabhängig (im Sinne von Definition 4.36), denn aus $\alpha_1 a_1 + \cdots + \alpha_k a_k = o$ folgt $\alpha_1 a_1 + \cdots + \alpha_k a_k \in U$ und daher $\alpha_1 = \cdots = \alpha_k = 0$. Ein Vektor $a \in X$ ist genau dann linear unabhängig relativ zu U, wenn $a \notin U$ ist.

Satz 12.10. *Es sei U ein Unterraum von X. Die Maximalzahl r von relativ zu U linear unabhängigen Vektoren aus X ist durch*

$$r = \dim X - \dim U$$

gegeben.

Beweis. Wir setzen $\dim X = n$ und $\dim U = m \leq n$. Der Fall $m = n$ ist trivial. Es sei daher $m < n$. Wir wählen die Basis $a_1, \dots, a_m, a_{m+1}, \dots, a_n$ von X derart, daß $a_1, \dots, a_m$ eine Basis von U bilden.

Aus $\alpha_{m+1} a_{m+n} + \cdots + \alpha_n a_n \in U$ folgt

$$\alpha_{m+1} a_{m+1} + \cdots + \alpha_n a_n = \alpha_1 a_1 + \cdots + \alpha_m a_m$$

mit geeigneten $\alpha_1, \dots, \alpha_m$. Aus der letzten Beziehung folgt aber $\alpha_1 = \cdots = \alpha_m = \alpha_{m+1} = \cdots = \alpha_n = 0$, d.h. die Vektoren $a_{m+1}, \dots, a_n$ sind linear unabhängig relativ zu U. Dies beweist

$$r \geq n - m.$$

Es seien nun $b_1, \dots, b_r$ linear unabhängig relativ zu U. Aus

$$\alpha_1 a_1 + \cdots + \alpha_m a_m + \beta_1 b_1 + \cdots + \beta_r b_r = o$$

folgt $\beta_1 b_1 + \cdots + \beta_r b_r = -\alpha_1 a_1 - \cdots - \alpha_m a_m \in U$. Daraus folgt $\beta_1 = \cdots = \beta_r = 0$. Somit gilt $\alpha_1 a_1 + \cdots + \alpha_m a_m = o$, woraus auch $\alpha_1 = \cdots = \alpha_m = 0$ folgt. Die Vektoren $a_1, \dots, a_m, b_1, \dots, b_r$ sind somit linear unabhängig, d.h. $r + m \leq n$ bzw.

$$r \leq n - m.$$

Damit ist $r = n - m$ bewiesen. □

Lemma 12.11. *Es gelte* (V) *und $a_1, \dots, a_k$ seien Vektoren aus* $\ker(\varphi - \lambda_0 \varepsilon)^\ell$, *$\ell \leq k$, welche linear unabhängig relativ zu* $\ker(\varphi - \lambda_0 \varepsilon)^{\ell-1}$ *sind. Dann sind die Vektoren*

$$(\varphi - \lambda_0 \varepsilon)^j (a_1), \dots, (\varphi - \lambda_0 \varepsilon)^j (a_k)$$

in $\ker(\varphi - \lambda_0 \varepsilon)^{\ell-j}$ *linear unabhängig relativ zu* $\ker(\varphi - \lambda_0 \varepsilon)^{\ell-j-1}$, *$j = 1, \dots, \ell - 1$.*

Beweis. Wegen $(\varphi - \lambda_0 \varepsilon)^{\ell-j}((\varphi - \lambda_0 \varepsilon)^j (a_i)) = (\varphi - \lambda_0 \varepsilon)^\ell (a_i) = o$ ist klar, daß die Vektoren $(\varphi - \lambda_0 \varepsilon)^j (a_i)$ Elemente aus $\ker(\varphi - \lambda_0 \varepsilon)^{\ell-j}$ sind, $i = 1, \dots, k$.

Es sei nun

$$\alpha_1 (\varphi - \lambda_0 \varepsilon)^j (a_1) + \cdots + \alpha_k (\varphi - \lambda_0 \varepsilon)^j (a_k) \in \ker(\varphi - \lambda_0 \varepsilon)^{\ell-j-1}.$$

Daraus folgt

$$\begin{aligned} o &= (\varphi - \lambda_0\varepsilon)^{\ell-j-1}\big(\alpha_1(\varphi - \lambda_0\varepsilon)^j(a_1) + \cdots + \alpha_k(\varphi - \lambda_0\varepsilon)^j(a_k)\big) \\ &= (\varphi - \lambda_0\varepsilon)^{\ell-1}(\alpha_1 a_1 + \cdots + \alpha_k a_k), \end{aligned}$$

d.h. $\alpha_1 a_1 + \cdots + \alpha_k a_k \in \ker(\varphi - \lambda_0\varepsilon)^{\ell-1}$. Somit ist $\alpha_1 = \cdots = \alpha_k = 0$. $\square$

Satz 12.12. *Es gelte* (V). *Für die Zahlen*

$$\rho_i := \operatorname{def}(\varphi - \lambda_0\varepsilon)^i - \operatorname{def}(\varphi - \lambda_0\varepsilon)^{i-1}, \quad i = 1, 2, \dots ,$$

gilt:

a) $\rho_i = 0$ *für* $i = k+1, k+2, \dots$ *und* $\rho_i > 0$ *für* $i = 1, \dots, k$.

b) $\rho_{i+1} \leq \rho_i$, $i = 1, 2, \dots$.

c) *Es gibt eindeutig bestimmte natürliche Zahlen* q *und* $k_1, \dots, k_q$ *derart, daß*

$$k = k_1 > k_2 > \cdots > k_q \geq 1$$

und

$$0 = \rho_{k_1+1} < \rho_{k_1} = \cdots = \rho_{k_2-1} < \rho_{k_2} = \cdots = \rho_{k_q-1} < \rho_{k_q} = \cdots = \rho_1$$

gelten.

d) *Für die Zahlen* $r_i := \rho_{k_i} - \rho_{k_i-1}$, $i = 2, \dots, q$, *und* $r_1 := \rho_{k_1}$ *gilt*

$$r_1 + \cdots + r_j = \rho_{k_j}, \quad j = 1, \dots, q,$$

und

$$r_1 k_1 + \cdots + r_q k_q = n (= \dim X). \tag{12.1}$$

Beweis. Die Aussage a) folgt aus Satz 12.5, f). Nach Satz 12.10 gibt es in $\ker(\varphi - \lambda_0\varepsilon)^{i+1}$ genau ρ_{i+1} relativ zu $\ker(\varphi - \lambda_0\varepsilon)^i$ linear unabhängige Vektoren $a_1, \dots, a_{\rho_{i+1}}$. Wegen Hilfssatz 12.11 sind dann die Vektoren $(\varphi - \lambda_0\varepsilon)(a_j)$, $j = 1, \dots, \rho_{i+1}$, relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{i-1}$ linear unabhängige Vektoren aus $\ker(\varphi - \lambda_0\varepsilon)^i$. Somit ist

$$\rho_{i+1} \leq \rho_i, \quad i = 1, 2, \dots .$$

Die Aussage c) ist dann unmittelbar klar.

Für die Zahlen $r_1, \dots, r_q$ ist nur die Gleichung (12.1) nicht trivial. Wegen der Definition der r_j gilt

$$\begin{aligned} r_1 k_1 + \cdots + r_q k_q &= \rho_{k_1} k_1 + (\rho_{k_2} - \rho_{k_1}) k_2 + \cdots + (\rho_{k_q} - \rho_{k_{q-1}}) k_q \\ &= \rho_{k_1}(k_1 - k_2) + \rho_{k_2}(k_2 - k_3) + \cdots \\ &\quad \cdots + \rho_{k_{q-1}}(k_{q-1} - k_q) + \rho_{k_q} k_q. \end{aligned}$$

Da $\rho_{k_j} = \cdots = \rho_{k_{j+1}-1}$, $j = 1, \dots, q-1$, bzw. $\rho_{k_q} = \cdots = \rho_1$ gilt, erhalten wir auf Grund der Definition der ρ_i die Gleichung

$$\begin{aligned} \rho_{k_j}(k_j - k_{j+1}) &= \rho_{k_j} + \rho_{k_j+1} + \cdots + \rho_{k_{j+1}-1} \\ &= \operatorname{def}(\varphi - \lambda_0\varepsilon)^{k_j} - \operatorname{def}(\varphi - \lambda_0\varepsilon)^{k_{j+1}}, \ j = 1, \dots, q-1, \end{aligned}$$

bzw.

$$\rho_{k_q} k_q = \rho_{k_q} + \cdots + \rho_1 = \operatorname{def}(\varphi - \lambda_0\varepsilon)^{k_q}.$$

Somit gilt insgesamt

$$r_1k_1 + \cdots + r_qk_q = \operatorname{def}(\varphi - \lambda_0\varepsilon)^{k_1} = \operatorname{def}(\varphi - \lambda_0\varepsilon)^k = n$$

(beachte, daß $(\varphi - \lambda_0\varepsilon)^k$ die Nullabbildung ist). $\square$

Man beachte, daß $r_1 + \cdots + r_j$ die Maximalzahl von relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{\nu-1}$ linear unabhängigen Vektoren in $\ker(\varphi - \lambda_0\varepsilon)^\nu$ für $\nu = k_{j+1} - 1, \ldots, k_j$ im Falle $j > q$ bzw. für $\nu = 1, \ldots, k_q$ im Falle $j = q$ ist (beachte $\rho_{k_j} = \cdots = \rho_{k_{j+1}-1}$).

Wir können nun den folgenden Algorithmus durchführen, um eine Basis von X zu konstruieren:

Schritt 1: Die Maximalzahl von relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{k_1-1}$ linear unabhängigen Vektoren in $\ker(\varphi - \lambda_0\varepsilon)^{k_1}$ ist r_1. Wir wählen daher r_1 Vektoren

$$a_1^1, \ldots, a_{r_1}^1 \in \ker(\varphi - \lambda_0\varepsilon)^{k_1},$$

die relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{k_1-1}$ linear unabhängig sind.[41] Wir bilden die Jordanketten

$$a_j^1, (\varphi - \lambda_0\varepsilon)(a_j^1), \ldots, (\varphi - \lambda_0\varepsilon)^{k_1-1}(a_j^1), \quad j = 1, \ldots, r_1. \tag{12.2}$$

Beachte, daß wegen der Wahl der Vektoren a_j^1 die damit gebildeten Jordanketten tatsächlich die Länge k_1 haben. Als Ergebnis von Schritt 1 erhalten wir die k_1r_1 Vektoren (12.2).

Schritt 2: Die Maximalzahl von relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{k_2-1}$ linear unabhängigen Vektoren in $\ker(\varphi - \lambda_0\varepsilon)^{k_2}$ ist $r_1 + r_2$. Durch $(\varphi - \lambda_0\varepsilon)^{k_1-k_2}(a_j^1)$, $j = 1, \ldots, r_1$, erhält man bereits r_1 Vektoren in $\ker(\varphi - \lambda_0\varepsilon)^{k_2}$, die relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{k_2-1}$ linear unabhängig sind (Hilfssatz 12.11). Wir wählen daher Vektoren $a_1^2, \ldots, a_{r_2}^2 \in \ker(\varphi - \lambda_0\varepsilon)^{k_2}$ derart, daß die Vektoren

$$a_1^2, \ldots, a_{r_2}^2 \quad \text{und} \quad (\varphi - \lambda_0\varepsilon)^{k_1-k_2}(a_j^1),\ j = 1, \ldots, r_1,$$

relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{k_2-1}$ linear unabhängig sind. Wir bilden die Jordanketten

$$a_j^2, (\varphi - \lambda_0\varepsilon)(a_j^2), \ldots, (\varphi - \lambda_0\varepsilon)^{k_2-1}(a_j^2), \quad j = 1, \ldots, r_2, \tag{12.3}$$

und erhalten damit im zweiten Schritt die k_2r_2 neuen Vektoren (12.3).

$$\vdots$$

$$\vdots$$

Schritt q: Für die in den Schritten 1 bis $q-1$ gewählten Vektoren $a_1^j, \ldots, a_{r_j}^j$, $j = 1, \ldots, q-1$, sind nach Hilfssatz 12.11 die Vektoren

$$\begin{aligned} &(\varphi - \lambda_0\varepsilon)^{k_1-k_q}(a_j^1), \quad j = 1, \ldots, r_1, \\ &\quad \vdots \\ &(\varphi - \lambda_0\varepsilon)^{k_{q-1}-k_q}(a_j^{q-1}), \quad j = 1, \ldots, r_{q-1}, \end{aligned} \tag{12.4}$$

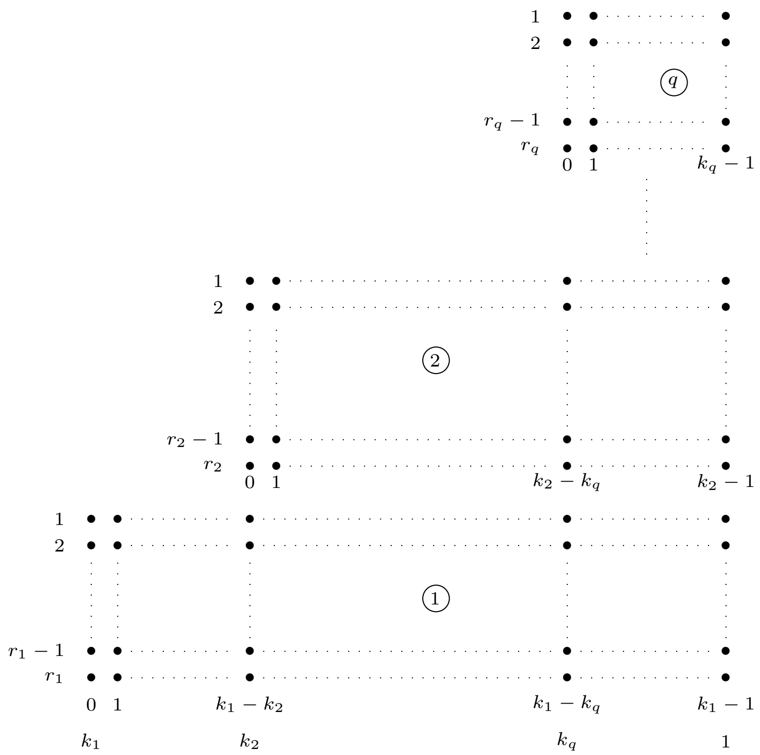
[41] Weiter unten wird gezeigt, wie man solche Vektoren berechnen kann.

relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{k_q-1}$ linear unabhängig. Wir wählen Vektoren $a_1^q, \ldots$ $\ldots, a_{r_q}^q \in \ker(\varphi - \lambda_0\varepsilon)^{k_q}$ derart, daß diese Vektoren und die Vektoren (12.4) relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{k_q-1}$ linear unabhängig sind. Wir bilden die Jordanketten

$$a_j^q, (\varphi - \lambda_0\varepsilon)(a_j^q), \ldots, (\varphi - \lambda_0\varepsilon)^{k_q-1}(a_j^q), \quad j = 1, \ldots, r_q, \tag{12.5}$$

und erhalten somit im q-ten Schritt die $k_q r_q$ neuen Vektoren (12.5).

Zur Veranschaulichung möge die untenstehende Zeichnung dienen, in der die in den Schritten 1 bis q konstruierten Vektoren durch Punkte dargestellt sind. In dem Teil des Schemas, der die im Schritt i, $i = 1, \ldots, q$, konstruierte Gruppe von Vektoren symbolisiert, steht der Punkt in der Zeile j und in der Spalte κ für den Vektor $(\varphi - \lambda_0\varepsilon)^\kappa(a_j^i)$, $\kappa = 0, \ldots, k_i - 1$, $j = 1, \ldots, r_i$. Die Vektoren des Schemas, die in der von rechts gezählten ℓ-ten Spalte des Schemas stehen, sind sämtliche aus $\ker(\varphi - \lambda_0\varepsilon)^\ell \setminus \ker(\varphi - \lambda_0\varepsilon)^{\ell-1}$, $\ell = 1, \ldots, k_1$, und linear unabhängig relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{\ell-1}$.



In den Schritten 1 bis q erhalten wir insgesamt $n = k_1 r_1 + \cdots + k_q r_q$ Vektoren. Es gilt nun:

Satz 12.13. *Die in den Schritten* 1 *bis* q *konstruierten Vektoren sind linear unabhängig und bilden somit eine Basis von* X.

Beweis. Es sei eine Linearkombination der konstruierten Vektoren gleich dem Nullvektor. Da nach Konstruktion sämtliche Vektoren, ausgenommen die Vektoren $a_1^1, \ldots, a_{r_1}^1$, in $\ker(\varphi - \lambda_0\varepsilon)^{k_1-1}$ liegen, folgt

$$o = \alpha_1 a_1^1 + \cdots + \alpha_{r_1} a_{r_1}^1 + a$$

mit $a \in \ker(\varphi - \lambda_0\varepsilon)^{k_1-1}$, d.h.

$$\alpha_1 a_1^1 + \cdots + \alpha_{r_1} a_{r_1}^1 \in \ker(\varphi - \lambda_0\varepsilon)^{k_1-1}.$$

Da die Vektoren $a_1^1, \ldots, a_{r_1}^1$ relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{k_1-1}$ linear unabhängig sind, folgt daraus $\alpha_1 = \cdots = \alpha_{r_1} = 0$.

Es sei bereits gezeigt, daß die Koeffizienten für alle Vektoren, die in der obenstehenden Zeichnung durch Punkte links von der $(\ell-1)$-ten Spalte (von rechts gezählt) symbolisiert werden, Null sind. Da dann die Linearkombination der übrigen Vektoren gleich dem Nullvektor sein muß, ist die Linearkombination der Vektoren in der ℓ-ten Spalte in $\ker(\varphi - \lambda_0\varepsilon)^{\ell-1}$ enthalten. Laut Konstruktion der Vektoren sind die Vektoren der ℓ-ten Spalte linear unabhängig relativ zu $\ker(\varphi - \lambda_0\varepsilon)^{\ell-1}$. Daher sind die Koeffizienten für die Vektoren der ℓ-ten Spalte sämtliche Null.

Nach endlich vielen Schritten sehen wir, daß sämtliche Koeffizienten der Linearkombination Null sein müssen. Die in den Schritten 1 bis q konstruierten n Vektoren sind daher linear unabhängig und bilden wegen $\dim X = n$ eine Basis von X. □

Satz 12.14. *Es sei $\varphi \in L(X,X)$ gegeben und es gelte die Voraussetzung* (V). *$\mathcal{B}$ bezeichne die in den Schritten* 1 *bis q konstruierte geordnete Basis von X. Die Matrix J von φ bzgl. $\mathcal{B}$ hat die Gestalt*

$$J = \operatorname{diag}(J_1^1, \ldots, J_{r_1}^1, J_1^2, \ldots, J_{r_2}^2, \ldots\ldots, J_1^q, \ldots, J_{r_q}^q),$$

wobei

$$J_i^j = \begin{pmatrix} \lambda_0 & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda_0 \end{pmatrix} \in M_{k_j,k_j}(\mathbb{C}),$$

$i = 1, \ldots, r_j$, $j = 1, \ldots, q$. Die Zahlen $q, r_1, \ldots, r_q$ und $k_1, \ldots, k_q$ sind eindeutig durch φ bestimmt.

Beweis. Die Aussage des Satzes über die Gestalt von J folgt unmittelbar aus Satz 12.8, da J_i^j die Matrix von $\varphi|_U$ mit $U = [a_i^j, (\varphi - \lambda_0\varepsilon)(a_i^j), \ldots \ldots, (\varphi - \lambda_0\varepsilon)^{k_j-1}(a_i^j)]$ ist. Die Eindeutigkeit der Zahlen q, r_j, k_j ist wegen Satz 12.12 klar. □

Ist A die Matrix von φ bzgl. irgendeiner geordneten Basis von X, so heißt J die ***Jordansche Normalform*** von A. Die Matrizen J_i^j heißen ***Jordanblöcke*** zum Eigenwert λ_0. Ist man nur an der Jordanschen Normalform interessiert, genügt es die Zahlen $q, r_1, \ldots, r_q$ und $k_1, \ldots, k_q$ zu berechnen. Dies kann an Hand der Formeln aus Satz 12.12 geschehen. Will man auch eine Transformationsmatrix S mit $J = S^{-1}AS$ bestimmen, muß man auch die Schritte 1 bis q durchführen (vgl. Satz 8.4 und Definition 8.1), um die Vektoren der neuen Basis von X zu konstruieren.

Um im j-ten Schritt die Vektoren $a_1^j, \ldots, a_{r_j}^j$ zu bestimmen, müssen wir folgende Aufgabe lösen:

> *Es sei r die Maximalzahl von Vektoren aus* $\ker(\varphi - \lambda_0\varepsilon)^\ell$*, die relativ zu* $\ker(\varphi - \lambda_0\varepsilon)^{\ell-1}$ *linear unabhängig sind. $b_1, \ldots, b_m$, $m < r$, seien Vektoren aus* $\ker(\varphi - \lambda_0\varepsilon)^\ell$*, die relativ zu* $\ker(\varphi-\lambda_0\varepsilon)^{\ell-1}$ *linear unabhängig sind. Gesucht sind Vektoren $b_{m+1}, \ldots, b_r$ derart, daß $b_1, \ldots, b_m$ relativ zu* $\ker(\varphi - \lambda_0\varepsilon)^{\ell-1}$ *linear unabhängig sind.*

Die Vektoren $b_{m+1}, \ldots, b_r$ können wie folgt bestimmt werden:

1. Man bestimmt eine Basis $(c_1, \ldots, c_k)$ von $\ker(\varphi-\lambda_0\varepsilon)^{\ell-1}$. Die Vektoren $c_1, \ldots, c_k, b_1, \ldots, b_m$ sind linear unabhängige Vektoren aus $\ker(\varphi - \lambda_0\varepsilon)^\ell$.

Beweis. Aus $\gamma_1 c_1 + \cdots + \gamma_k c_k + \beta_1 b_1 + \cdots + \beta_m b_m = 0$ folgt $\beta_1 b_1 + \cdots + \beta_m b_m \in \ker(\varphi - \lambda_0 \varepsilon)^{\ell-1}$ und somit $\beta_1 = \cdots = \beta_m = 0$. Aus $\gamma_1 c_1 + \cdots + \gamma_k c_k = 0$ folgt aber $\gamma_1 = \cdots = \gamma_k = 0$. □

2. Man ergänzt die Vektoren $c_1, \ldots, c_k$, $b_1, \ldots, b_m$ durch Vektoren $b_{m+1}, \ldots$ $\ldots, b_r$ zu einer Basis von $\ker(\varphi - \lambda_0 \varepsilon)^\ell$. Dazu wählt man eine Basis $(d_1, \ldots, d_{k+r})$ von $\ker(\varphi - \lambda_0 \varepsilon)^\ell$ und ersetzt $m + k$ dieser Vektoren durch $c_1, \ldots, c_k, b_1, \ldots, b_m$ (Austauschsatz von Steinitz). Die übrig gebliebenen Vektoren d_j sind die gesuchten.

Beweis. Wir müssen noch zeigen, daß die Vektoren $b_1, \ldots, b_r$ tatsächlich relativ zu $\ker(\varphi - \lambda_0 \varepsilon)^{\ell-1}$ linear unabhängig sind.

Es sei $\beta_1 b_1 + \cdots + \beta_r b_r \in \ker(\varphi - \lambda_0 \varepsilon)^{\ell-1}$. Dann gilt mit Konstanten $\gamma_1, \ldots, \gamma_k \in \mathbb{C}$

$$\beta_1 b_1 + \cdots + \beta_r b_r + \gamma_1 c_1 + \cdots + \gamma_k c_k = o.$$

Da die Vektoren $c_1, \ldots, c_k$, $b_1, \ldots, b_r$ eine Basis von $\ker(\varphi - \lambda_0 \varepsilon)^\ell$ bilden, folgt $\beta_1 = \cdots = \beta_r = 0$ (und $\gamma_1 = \cdots = \gamma_k = 0$). □

Beispiel 12.15. Gegeben sei

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

Das Minimalpolynom von A ist (siehe Beispiel 11.6)

$$\mu_A(\lambda) = (\lambda - 2)^2.$$

Es ist $k_1 = k = 2$ und $\operatorname{rang}(A - 2E)^2 = \operatorname{rang} 0 = 0$ und

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

Somit gilt $\rho_i = 0$, $i \geq 3$, und

$$\rho_2 = 4 - 3 = 1, \; \rho_1 = 3 - 0 = 3.$$

Daraus folgt $q = 2$, $k_1 = 2$, $k_2 = 1$ und

$$r_1 = 1 - 0 = 1, \; r_2 = 3 - 1 = 2.$$

Die Jordansche Normalform von A ist daher (siehe Satz 12.14) durch

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

gegeben. Um eine Matrix S mit $J = S^{-1} A S$ zu bestimmen, müssen wir eine Basis von $\mathbb{C}^4$ gemäß den Schritten 1 bis 2 berechnen:

Schritt 1. Es ist ein Vektor $\xi \in \ker(A - 2E)^2 = \mathbb{C}^4$ mit $\xi \notin \ker(A - 2E)$ zu bestimmen. (Man beachte: Ein Vektor ξ ist linear unabhängig relativ zu $\ker(A - 2E)$ genau dann, wenn $\xi \notin \ker(A - 2E)$ gilt.)

Eine Basis von $\ker(A-2E)$ (bzw. eine Lösungsbasis von $(A-2E)\eta = 0$) ist durch

$$\eta_1 = \begin{pmatrix}1\\1\\0\\0\end{pmatrix}, \quad \eta_2 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \quad \eta_3 = \begin{pmatrix}0\\-1\\0\\1\end{pmatrix}$$

gegeben. Wir müssen die Vektoren η_1, η_2, η_3 durch einen Vektor a^1 zu einer Basis von $\ker(A-2E)^2 = \mathbb{C}^4$ ergänzen. In der kanonischen Basis von $\mathbb{C}^4$ können wir beispielsweise e_1, e_2, e_3 durch die Vektoren η_1, η_3, η_2 ersetzen (vgl. Hilfssatz 4.47). Nach dem oben angegebenen Algorithmus können wir $a^1 = e_4$ wählen. Die von a^1 erzeugte Jordan-Kette ist

$$a^1 = e_4 = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}, \quad (A-2E)a^1 = \begin{pmatrix}1\\0\\1\\1\end{pmatrix}.$$

Im Schritt 2 müssen wir $r_2 = 2$ Vektoren a_1^2, a_2^2 in $\ker(A-2E)$ derart bestimmen, daß $(A-2E)a^1, a_1^2, a_2^2$ linear unabhängig relativ zu $\ker(A-2E)^0 = \{0\}$ sind, d.h. $(A-2E)a^1, a_1^2, a_2^2$ müssen einfach linear unabhängig sein.

Die Vektoren η_1, η_2, η_3 bilden eine Basis von $\ker(A-2E)$. Der Vektor $(A-2E)a^1$ besitzt die Darstellung

$$(A-2E)a^1 = \eta_1 + \eta_2 + \eta_3.$$

Wir können daher jeden der Vektoren η_i durch $(A-2E)a^1$ ersetzen. Wir wählen etwa η_1. Dann ist $a_1^2 = \eta_2$, $a_2^2 = \eta_3$. Die gesuchte Basis von $\mathbb{C}^4$ ist daher

$$a^1 = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}, \quad (A-2E)a^1 = \begin{pmatrix}1\\0\\1\\1\end{pmatrix}, \quad a_1^2 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \quad a_2^2 = \begin{pmatrix}0\\-1\\0\\1\end{pmatrix}.$$

Eine Matrix S mit $J = S^{-1}AS$ ist durch

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

gegeben. Als Probe verifiziere man $JS = AS$.

Bei diesem Beispiel hätte man in Schritt 1 auch die Vektoren e_2, e_3, e_4 durch η_1, η_2, η_3 ersetzen können, was zu $a^1 = e_1$ geführt hätte. Dann ist

$$a^1 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \quad (A-2E)a^1 = \begin{pmatrix}-1\\0\\-1\\-1\end{pmatrix}.$$

Im zweiten Schritt hätten wir $(A-2E)a^1 = -\eta_1 - \eta_2 - \eta_3$ und könnten etwa $a_1^2 = \eta_1$, $a_2^2 = \eta_2$ wählen. Dies würde die Matrix

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

liefern, für die ebenfalls $J = S^{-1}AS$ gilt. $\Diamond$

Beispiel 12.16. Es sei

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

Durch Entwicklung von $\det(A-\lambda E)$ erhält man sofort $\chi_A(\lambda) = (\lambda-2)^8$, d.h. $\mu_A(\lambda) = (\lambda-2)^k$. Eine einfache Rechnung zeigt

$$\operatorname{rang}(A-2E) = 5,\ \operatorname{rang}(A-2E)^2 = 2,\ \operatorname{rang}(A-2E)^3 = 1 \text{ und } \operatorname{rang}(A-2E)^4 = 0.$$

Daraus folgt $\mu_A(\lambda) = (\lambda-2)^4$, d.h. $k_1 = k = 4$ und

$$\rho_4 = 8-7 = 1,\ \rho_3 = 7-6 = 1,\ \rho_2 = 6-3 = 3,\ \rho_1 = 3-0 = 3.$$

Daher ist $q = 2$ und

$$k_1 = 4,\ k_2 = 2,\ r_1 = 1,\ r_2 = 2.$$

Die Jordansche Normalform von A ist daher durch

$$J = \left(\begin{array}{cccc:cc:cc} 2 & & & 0 & & & & \\ 1 & 2 & & & & 0 & & \\ & 1 & 2 & & & & & \\ 0 & & 1 & 2 & & & & \\ \hdashline & & & & 2 & 0 & & \\ & & & & 1 & 2 & & \\ \hdashline & & 0 & & & & 2 & 0 \\ & & & & & & 1 & 2 \end{array}\right).$$

gegeben. Um die Transformationsmatrix S zu bestimmen, berechnen wir zunächst eine Basis von $\ker(A-2E)^3$. Eine einfache Rechnung zeigt, daß $e_2, e_3, \ldots, e_8$ eine Basis von $\ker(A-2E)^3$ bilden. Diese Basis kann durch e_1 zu einer Basis von $\ker(A-2E)^4 = \mathbb{C}^8$ ergänzt werden. Dies liefert $a^1 = e_1$. Die zugehörige Jordan-Kette ist

$$e_1, (A-2E)e_1 = e_2 + e_6 - e_8, (A-2E)^2 e_1 = e_3, (A-2E)^3 e_1 = e_4.$$

In Schritt 2 sind zwei Vektoren a_1^2, a_2^2 aus $\ker(A-2E)^2$ derart zu bestimmen, daß a_1^2, a_2^2 und $(A-2E)^2 e_1 = e_3$ relativ zu $\ker(A-2E)$ linear unabhängig sind. Eine Basis von $\ker(A-2E)$ ist durch e_2, e_4, e_8 gegeben.

Eine Basis von $\ker(A-2E)^2$ ist etwa durch die Vektoren

$$e_2, e_3, e_4, e_5, e_6 - e_7, e_8.$$

gegeben. Die Vektoren e_2, e_4, e_8 und $(A-2E)^2 e_1 = e_3$ können daher durch e_5 und $e_6 - e_7$ zu einer Basis von $\ker(A-2E)^2$ ergänzt werden. Dies ergibt

$$a_1^2 = e_5,\ a_2^2 = e_6 - e_7$$

und die zugehörige Jordan-Ketten

$$e_5,\ (A-2E)e_5 = e_2,$$
$$e_6 - e_7,\ (A-2E)(e_6-e_7) = -e_8.$$

Die Matrix S ist somit durch

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

gegeben. $\Diamond$

Beispiel 12.17. Es sei

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

Eine einfache Rechnung zeigt

$$\chi_A(\lambda) = (\lambda-1)^3(\lambda-2)^3.$$

Wir betrachten zunächst A auf $\ker(A-E)^3$. Man erhält

$$\operatorname{rang}(A-E) = 5,\ \operatorname{rang}(A-E)^2 = 4,\ \operatorname{rang}(A-E)^3 = 3.$$

Nach Satz 12.5, Punkt 5, besitzt das Minimalpolynom von A den Faktor $(\lambda-1)^3$. Aus den Rangzahlen folgt

$$\rho_3 = 3-2 = 1,\ \rho_2 = 2-1 = 1,\ \rho_3 = 1-0 = 1.$$

Es gilt somit $q = 1$ und

$$k_1 = 3,\ r_1 = 1.$$

Die Jordansche Normalform von $A\,|_{\ker(A-E)^3}$ ist daher durch

$$J_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

gegeben. Um die zugehörige Basis von $\ker(A-E)^3$ zu berechnen führen wir Schritt 1 durch.

Die Vektoren e_1, e_2, e_6 bilden eine Basis von $\ker(A-E)^3$. Wegen $(A-E)^2e_1 \neq o$, d.h. $e_1 \notin \ker(A-E)^2$, ist e_1 linear unabhängig relativ zu $\ker(A-E)^2$. Die zugehörige Jordan-Kette ist

$$e_1, (A-E)e_1 = -e_1 + 2e_2 + e_6,\ (A-e)^2e_1 = -e_2.$$

Diese Vektoren bilden die gesuchte Basis von $\ker(A-E)^3$.

Für den Eigenwert $\lambda = 2$ gilt:

$$\operatorname{rang}(A-2E) = 4,\ \operatorname{rang}(A-2E)^2 = 3,\ \operatorname{rang}(A-2E)^3 = 3.$$

Das Minimalpolynom von A enthält somit den Faktor $(\lambda-2)^2$, d.h.

$$\mu_A(\lambda) = (\lambda-1)^3(\lambda-2)^2.$$

Ferner gilt:

$$\rho_2 = 3-2 = 1,\ \rho_1 = 2-0 = 2$$

und $q = 2$

$$k_1 = 2,\ k_2 = 1,\ r_1 = 1,\ r_2 = 1.$$

Die Jordansche Normalform von $A\,|_{\ker(A-2E)^2}$ ist daher durch

$$J_2 = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

gegeben. Zur Bestimmung einer entsprechenden Basis von $\ker(A-2E)^2$ müssen wir die Schritte 1 und 2 durchführen:

Schritt 1. Eine Basis von $\ker(A-2E)$ ist (e_2-e_3, e_5), während eine Basis von $\ker(A-2E)^2$ durch (e_2-e_3, e_4, e_5) gegeben ist. Es ist daher $a^1 = e_4$. Die zugehörige Jordan-Kette ist

$$e_4, (A-2E)e_4 = -e_2+e_3.$$

Schritt 2. $(A-2E)e_4 = -e_2+e_3$ ist zu einer Basis von $\ker(A-2E)$ zu ergänzen. Dies liefert etwa $a^2 = e_5$.

Die gesuchte Basis von $\ker(A-2E)^2$ ist daher

$$e_4, (A-2E)e_4 = -e_2+e_3, e_5.$$

Wir erhalten somit insgesamt (vgl. die Bemerkungen im Anschluß an den Beweis von Satz 12.5)

$$J = \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix}$$

mit der Transformationsmatrix

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

$\Diamond$

Eine reelle Matrix A, aufgefaßt als Matrix in $M_{n,n}(\mathbb{C})$, wird im allgemeinen auch nicht-reelle Eigenwerte besitzen. Die Jordansche Normalform J von A und die zugehörigen Transformationsmatrizen S mit $S^{-1}AS = J$ werden daher im allgemeinen nicht-reell sein. Es erhebt sich daher die Frage, auf welche Form man eine reelle Matrix mittels Basistransformationen im $\mathbb{R}^n$ bringen kann. Eine von mehreren möglichen reellen Normalformen kann man mittels der Jordanschen Normalform gewinnen. Ausgangspunkt ist der folgende Hilfssatz:

Lemma 12.18. *Es sei $A \in M_{n,n}(\mathbb{R})$ gegeben. Für beliebige $\lambda \in \mathbb{C}\setminus\mathbb{R}$, $\ell \in \mathbb{N}$ und $a \in \mathbb{C}^n$ gilt $a \in \ker(A-\lambda E)^\ell$ genau dann, wenn $\bar{a} \in \ker(A-\bar{\lambda}E)^\ell$ ist.*

Beweis. Es gilt $(A-\lambda E)^\ell a = 0$ genau dann, wenn $0 = \bar{0} = \overline{(A-\lambda E)^\ell a} = (A-\bar{\lambda}E)^\ell \bar{a}$ ist. $\square$

Aus Hilfssatz 12.18 folgt insbesondere, daß zu einem nicht-reellen Eigenwert λ_0 von $A \in M_{n,n}(\mathbb{R})$ stets auch $\bar{\lambda}_0$ Eigenwert von A ist. Ferner stimmen die algebraische Vielfachheit von λ_0 und $\bar{\lambda}_0$ sowie die Exponenten der Wurzelfaktoren von λ_0 und $\bar{\lambda}_0$ im Minimalpolynom von A überein (vgl. Satz 12.5, a), e) und f)).

Es sei nun λ_0 ein nicht-reeller Eigenwert von A und m bzw. k sei die algebraische Vielfachheit von λ_0 und $\bar{\lambda}_0$ bzw. der Exponent der Wurzelfaktoren $\lambda - \lambda_0$ und $\lambda - \bar{\lambda}_0$ im Minimalpolynom von A. Dann ist $(a_1, \ldots, a_m)$ genau dann eine geordnete Basis von $\ker(A - \lambda_0 E)^k$, wenn $(\bar{a}_1, \ldots, \bar{a}_m)$ eine geordnete Basis von $\ker(A - \bar{\lambda}_0 E)^k$ ist. Dies folgt unmittelbar aus Hilfssatz 12.18 und der Tatsache, daß Vektoren $a_1, \ldots, a_m \in \mathbb{C}^n$ genau dann linear unabhängig sind, wenn dies für die Vektoren $\bar{a}_1, \ldots, \bar{a}_m$ gilt. Wegen $\ker(A - \lambda_0 E)^k \cap \ker(A - \bar{\lambda}_0 E)^k = \{0\}$ sind die Vektoren $a_1, \ldots, a_m, \bar{a}_1, \ldots, \bar{a}_m$ linear unabhängig (und bilden daher eine Basis von $\ker(A - \lambda_0 E)^k \oplus \ker(A - \bar{\lambda}_0 E)^k$).

Lemma 12.19. *Die Vektoren $a_1, \ldots, a_m, \bar{a}_1, \ldots, \bar{a}_m \in \mathbb{C}^n$, $2m \leq n$, sind in $\mathbb{C}^n$ genau dann linear unabhängig, wenn die Vektoren*

$$\operatorname{Re} a_1, \operatorname{Im} a_1, \ldots, \operatorname{Re} a_m, \operatorname{Im} a_m \in \mathbb{R}^n$$

in $\mathbb{R}^n$ linear unabhängig sind.

Beweis. Wir setzen $b_j = \operatorname{Re} a_j$ und $c_j = \operatorname{Im} a_j$, $j = 1, \ldots, m$. Wegen $b_j = (a_j + \bar{a}_j)/2$ und $c_j = (a_j - \bar{a}_j)/2i$, $j = 1, \ldots, m$, entsteht $(b_1, c_1, \ldots, b_m, c_m)$ aus $(a_1, \ldots, a_m, \bar{a}_1, \ldots, \bar{a}_m)$ durch elementare Umformungen. Daher sind beide $2m$-Tupel von Vektoren im $\mathbb{C}^n$ zugleich linear unabhängig oder linear abhängig. Dies beweist die Aussage, wenn man noch beachtet, daß Vektoren aus $\mathbb{R}^n$ genau dann in $\mathbb{C}^n$ linear unabängig sind, wenn sie in $\mathbb{R}^n$ linear unabhängig sind. $\square$

Satz 12.20. *Es sei $A \in M_{n,n}(\mathbb{R})$ gegeben. $\lambda_1 \ldots, \lambda_s$ seien die paarweise verschiedenen reellen Nullstellen von χ_A und $\mu_1, \ldots, \mu_k, \bar{\mu}_1, \ldots, \bar{\mu}_k$ die paarweise verschiedenen nicht-reellen Nullstellen von χ_A. Dann existiert eine reguläre Matrix $S \in M_{n,n}(\mathbb{R})$ mit*

$$S^{-1}AS = \operatorname{diag}(J_1, \ldots, J_r, K_1, \ldots, K_t),$$

wobei $J_1, \ldots, J_r$ die zu den Eigenwerten $\lambda_1, \ldots, \lambda_s$ gehörigen Jordan-Blöcke in der Jordanschen Normalform von A sind und $K_1, \ldots, K_t$ aus den zu $\mu_1, \ldots, \mu_k$ gehörigen Jordan-Blöcken in der Jordanschen Normalform von A wie folgt entstehen: Ist

$$J(\mu) = \begin{pmatrix} \mu & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \mu \end{pmatrix} \in M_{\ell,\ell}(\mathbb{C}),$$

so ist

$$K(\mu) = \begin{pmatrix} R(\mu) & & & 0 \\ E_2 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & E_2 & R(\mu) \end{pmatrix} \in M_{2\ell,2\ell}(\mathbb{R})$$

mit

$$E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad R(\mu) = \begin{pmatrix} \operatorname{Re} \mu & -\operatorname{Im} \mu \\ \operatorname{Im} \mu & \operatorname{Re} \mu \end{pmatrix}.$$

Beweis. a) Ist $a, (A-\lambda E)a, \ldots, (A-\lambda E)^{\ell-1}a$ mit $a \in \mathbb{C}^n$ eine Jordan-Kette zum Eigenwert $\lambda \in \mathbb{R}$, so erzeugt $\operatorname{Re} a$ oder $\operatorname{Im} a$ ebenfalls eine Jordan-Kette der Länge ℓ zu λ. Daraus folgt, daß die weiter oben in den Schritten 1 bis q konstruierten Basisvektoren von $\ker(A-\lambda_j E)^{k_j}$, $j = 1, \ldots, s$, in $\mathbb{R}^n$ gewählt werden können (k_j ist der Exponent von $\lambda-\lambda_j$ im Minimalpolynom von A).

b) Ist λ ein nicht-reeller Eigenwert von A und $a, (A-\lambda E)a, \ldots, (A-\lambda E)^{\ell-1}a$ eine Jordan-Kette von A zu λ mit $a \in \mathbb{C}^n$, so erzeugt $\bar{a}$ eine Jordan-Kette der Länge ℓ von A zum Eigenwert $\bar{\lambda}$. Aus

$$\begin{aligned} A(A-\lambda E)^j a &= \lambda(A-\lambda E)^j a + (A-\lambda E)^{j+1}a, \quad j = 0, \ldots, \ell-2, \\ A(A-\lambda E)^{\ell-1}a &= \lambda(A-\lambda E)^{\ell-1}a \end{aligned}$$

folgt

$$\begin{aligned} A \operatorname{Re}(A-\lambda E)^j a &= \operatorname{Re}\lambda \operatorname{Re}(A-\lambda E)^j a - \operatorname{Im}\lambda \operatorname{Im}(A-\lambda E)^j a \\ &\qquad + \operatorname{Re}(A-\lambda E)^{j+1}a,\ j = 0, \ldots, \ell-2, \\ A \operatorname{Im}(A-\lambda E)^j a &= \operatorname{Im}\lambda \operatorname{Re}(A-\lambda E)^j a + \operatorname{Re}\lambda \operatorname{Im}(A-\lambda E)^j a \\ &\qquad + \operatorname{Im}(A-\lambda E)^{j+1}a,\ j = 0, \ldots, \ell-2, \\ A \operatorname{Re}(A-\lambda E)^{\ell-1}a &= \operatorname{Re}\lambda \operatorname{Re}(A-\lambda E)^{\ell-1}a - \operatorname{Im}\lambda \operatorname{Im}(A-\lambda E)^{\ell-1}a, \\ A \operatorname{Im}(A-\lambda E)^{\ell-1}a &= \operatorname{Im}\lambda \operatorname{Re}(A-\lambda E)^{\ell-1}a + \operatorname{Re}\lambda \operatorname{Im}(A-\lambda E)^{\ell-1}a. \end{aligned}$$

Bezüglich der Basis

$$\big(a, \ldots, (A-\lambda E)^{\ell-1}a, \bar{a}, \ldots, (A-\bar{\lambda}E)^{\ell-1}\bar{a}\big)$$

von $U := [a, \ldots, (A-\bar{\lambda}E)^{\ell-1}\bar{a}]$ ist daher die Matrix von $\varphi_A\,|_U$ durch die Jordansche Normalform $\operatorname{diag}(J(\lambda), J(\bar{\lambda})) \in M_{2\ell,2\ell}(\mathbb{C})$ gegeben. Bezüglich der Basis

$$\begin{aligned} \big(\operatorname{Re} a, -\operatorname{Im} a, \operatorname{Re}(A-\lambda E)a, -\operatorname{Im}(A-\lambda E)a, \ldots \\ \ldots, \operatorname{Re}(A-\lambda E)^{\ell-1}a, -\operatorname{Im}(A-\lambda E)^{\ell-1}a\big) \end{aligned}$$

von U ist die Matrix von $\varphi_A\,|_U$ durch $K(\mu) \in M_{2\ell,2\ell}(\mathbb{R})$ gegeben. $\square$

Kapitel 13

Vektorräume mit innerem Produkt

Für Vektoren im $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ hatten wir seinerzeit (vgl. Abschnitt 2.2) durch geometrische Überlegungen ein inneres Produkt eingeführt. In diesem Kapitel wollen wir Vektorräume mit innerem Produkt auf einer etwas abstrakteren Ebene betrachten. Es sei stets (ausgenommen in Abschnitt 13.3)

$$\mathbb{K} = \mathbb{R} \text{ oder } \mathbb{K} = \mathbb{C}.$$

13.1 Inneres Produkt

Definition 13.1. *X sei ein Vektorraum über $\mathbb{K}$. Eine Abbildung $\Phi : X \times X \to \mathbb{K}$ heißt genau dann ein **inneres Produkt** auf X, wenn gilt:*

(i) *$\Phi(x, \alpha y_1 + \beta y_2) = \alpha\Phi(x, y_1) + \beta\Phi(x, y_2)$ für alle $x, y_1, y_2 \in X$ und $\alpha, \beta \in \mathbb{K}$ (Linearität im zweiten Argument).*

(ii) *$\Phi(y, x) = \overline{\Phi(x, y)}$ für alle $x, y \in X$.*

(iii) *$\Phi(x, x) \geq 0$ für alle $x \in X$ und $\Phi(x, x) = 0$ genau dann, wenn $x = o$ (positive Definitheit).*

Aus der Eigenschaft (ii) folgt $\Phi(x, x) = \overline{\Phi(x, x)}$, d.h. $\Phi(x, x)$ ist stets reell. Daher ist es unter (iii) sinnvoll, $\Phi(x, x) \geq 0$ zu verlangen. Ist $\mathbb{K} = \mathbb{R}$, so bedeuten (i) und (ii), dass Φ eine symmetrische ***Bilinearform*** auf X ist (statt 2-fache Linearform sagt man Bilinearform, vgl. Definition 9.25). Eine Abbildung Φ mit den Eigenschaften (i) und (ii) wird auch ***Sesquilinearform*** genannt. Aus (i) und (iii) folgt

$$\Phi(\alpha x_1 + \beta x_2, y) = \bar{\alpha}\Phi(x_1, y) + \bar{\beta}\Phi(x_2, y).$$

Es ist üblich $\langle x, y\rangle$ an Stelle von $\Phi(x, y)$ zu schreiben. Die folgende Schlußweise wird häufig verwendet werden:

Lemma 13.2. *X sei ein Vektorraum über $\mathbb{K}$ mit innerem Produkt $\langle\cdot,\cdot\rangle$ und $y \in X$ sei gegeben. Es ist genau dann $\langle y, x\rangle = 0$ für alle $x \in X$, wenn $y = o$ gilt.*

Beweis. Es sei $\langle y, x\rangle = 0$ für alle $x \in X$ und $y \neq o$. Für $x = y$ folgt $\langle y, y\rangle = 0$ in Widerspruch zur positiven Definitheit von $\langle\cdot,\cdot\rangle$. Somit gilt $y = o$. Ist umgekehrt $y = o$, so folgt aus $\langle o, x\rangle = \langle o + o, x\rangle = \langle o, x\rangle + \langle o, x\rangle$ auch $\langle o, x\rangle = 0$ für alle $x \in X$. ☐

Wie im Falle der Vektoren im $\mathbb{R}^2$ bzw. $\mathbb{R}^3$ definieren wir:

Definition 13.3. *X sei ein Vektorraum über $\mathbb{K}$ mit innerem Produkt $\langle\cdot,\cdot\rangle$.*
a) Zwei Vektoren $x, y \in X$ heißen genau dann ***orthogonal****, $x \perp y$, wenn*

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

gilt.
b) Ist M eine Teilmenge von X, so heißt

$$M^{\perp} = \{x \in X \mid x \perp y \text{ für alle } y \in M\}$$

das ***orthogonale Komplement*** *von M.*

Wie seinerzeit (vgl. Abschnitt 2.2) können wir mit Hilfe eines inneren Produktes die Norm eines Vektors definieren:

$$\|x\| := \langle x, x\rangle^{1/2} \quad \text{für alle } x \in X. \tag{13.1}$$

Satz 13.4. *X sei ein Vektorraum über $\mathbb{K}$ mit innerem Produkt $\langle\cdot,\cdot\rangle$. Dann gilt:*
a) *Für alle $x, y \in X$ ist*

$$|\langle x, y\rangle| \leq \|x\|\,\|y\| \quad (\textit{Schwarzsche Ungleichung}).$$

Das Gleichheitszeichen gilt dann und nur dann, wenn x und y linear abhängig sind.
b) *Die Abbildung $\|\cdot\| : X \to \mathbb{R}$ ist eine* ***Norm*** *auf X, d.h. es gilt:*

(N1) $\|x\| \geq 0$ *für alle $x \in X$ und $\|x\| = 0$ genau dann, wenn $x = o$.*

(N2) $\|\alpha x\| = |\alpha|\,\|x\|$ *für alle $x \in X$ und $\alpha \in \mathbb{K}$.*

(N3) $\|x + y\| \leq \|x\| + \|y\|$ *für alle $x, y \in X$.*

Beweis. a) Der Beweis ist analog dem seinerzeit in Abschnitt 2.2 gegebenen. Für $y = o$ ist nichts zu beweisen. Es sei also $y \neq o$. Wir bestimmen $\alpha \neq 0$ derart, dass $\alpha y \perp x - \alpha y$ ist. Dies ist dann der Fall, wenn

$$0 = \langle \alpha y, x - \alpha y\rangle = \bar{\alpha}\langle y, x - \alpha y\rangle = \bar{\alpha}\langle y, x\rangle - \bar{\alpha}\alpha\langle y, y\rangle.$$

Wegen $\bar{\alpha} \neq 0$ und $\langle y, y\rangle \neq 0$ folgt daraus

$$\alpha = \frac{\overline{\langle x, y\rangle}}{\|y\|^2}.$$

Es ist $x = \alpha y + x - \alpha y$. Wegen $\alpha y \perp x - \alpha y$ erhalten wir

$$\begin{aligned}\|x\|^2 &= \big\langle \alpha y + x - \alpha y, \alpha y + x - \alpha y\big\rangle = |\alpha|^2\|y\|^2 + \|x - \alpha y\|^2 \\ &\geq |\alpha|^2\|y\|^2 = \frac{|\langle x, y\rangle|^2}{\|y\|^4}\|y\|^2 = \frac{|\langle x, y\rangle|^2}{\|y\|^2},\end{aligned}$$

woraus die Schwarzsche Ungleichung unmittelbar folgt. Das Gleichheitszeichen gilt genau dann, wenn in der Ungleichung oben Gleichheit gilt, d.h. $x = \alpha y$.
b) Die Eigenschaft (i) folgt unmittelbar aus der Definitheit des inneren Produktes. (ii) ist klar wegen $\|\alpha x\|^2 = \langle \alpha x, \alpha x\rangle = \bar{\alpha}\alpha\langle x, x\rangle = |\alpha|^2\|x\|^2$.

Die Dreiecksungleichung folgt mit Hilfe der Schwarzschen Ungleichung:

$$\begin{aligned}\|x + y\|^2 &= \langle x + y, x + y\rangle = \|x\|^2 + \langle y, x\rangle + \langle x, y\rangle + \|y\|^2 \\ &= \|x\|^2 + 2\mathrm{Re}\,\langle x, y\rangle + \|y\|^2 \leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.\end{aligned}$$

□

Satz 13.5. *X sei ein Vektorraum über $\mathbb{K}$ mit innerem Produkt $\langle \cdot, \cdot \rangle$. Für die durch* (13.1) *definierte Norm auf X gilt die* ***Parallelogrammregel****,*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad \textit{für alle } x, y \in X.$$

Beweis. Man addiere die Identitäten

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

□

Beispiel 13.6. Durch

$$\|x\|_1 = |\xi_1| + |\xi_2| \quad \text{für alle } x = (\xi_1, \xi_2)^\mathsf{T} \in \mathbb{R}^2$$

wird eine Norm auf dem $\mathbb{R}^2$ definiert (d.h. für $\| \cdot \|_1$ gelten die Eigenschaften (N1) - (N3) aus Satz 13.4, b)). Für die Vektoren $x = (1, 0)^\mathsf{T}$, $y = (0, 1)^\mathsf{T}$ ist $\|x + y\|_1^2 + \|x - y\|_1^2 = 8$ und $2\|x\|_1^2 + 2\|y\|_1^2 = 4$. Die Parallelogrammregel ist daher für $\| \cdot \|_1$ nicht erfüllt. Es gibt somit kein inneres Produkt Φ auf $\mathbb{R}^2$ mit $\|x\|_1^2 = \Phi(x, x)$ für alle $x \in \mathbb{R}^2$. $\Diamond$

Im folgenden sei X stets ein Vektorraum über $\mathbb{K}$ mit innerem Produkt $\langle \cdot, \cdot \rangle$.

Definition 13.7. *M sei Teilmenge von X.*
a) *M heißt genau dann* ***orthonormiert****, wenn gilt:*

(i) $\|x\| = 1$ *für alle* $x \in M$.

(ii) $x \perp y$ *für alle* $x, y \in M$, $x \neq y$.

b) M heißt genau dann eine ***Orthonormalbasis*** *von X, wenn M orthonormiert und eine Basis von X ist.*

Satz 13.8. *M sei eine orthonormierte Teilmenge von X. Dann ist M linear unabhängig.*

Beweis. Es seien paarweise verschiedene Elemente $a_1, \ldots, a_k \in M$ gewählt und es sei $\alpha_1 a_1 + \cdots + \alpha_k a_k = o$. Daraus folgt wegen $\langle a_j, a_i \rangle = 0$ für $i \neq j$ und $\langle a_j, a_j \rangle = \|a_j\|^2 = 1$

$$0 = \left\langle a_j, \alpha_1 a_1 + \cdots + \alpha_k a_k \right\rangle = \alpha_j, \quad j = 1, \ldots, k.$$

Es gilt somit $\alpha_1 = \cdots = \alpha_k = 0$, d.h. $a_1, \ldots, a_k$ sind linear unabhängig und daher auch M (siehe Definition 4.36, c)). □

Satz 13.9. *Die Vektoren $a_1, \ldots, a_k \in X$ seien linear unabhängig. Dann gibt es Vektoren $b_1, \ldots, b_k \in X$ derart, dass $\{b_1, \ldots, b_k\}$ orthonormiert ist und*

$$[b_1, \ldots, b_i] = [a_1, \ldots, a_i], \quad i = 1, \ldots, k,$$

gilt.

Beweis. In dem Beweis werden die Vektoren $b_1, \ldots, b_k$ aus den Vektoren $a_1, \ldots, a_k$ konstruiert. Man nennt dieses Konstruktionsverfahren das ***Gram***[42]***-Schmidtsche***[43] ***Orthonormierungsverfahren***[44].

Wir setzen

$$b_1 := \frac{a_1}{\|a_1\|}.$$

Dann ist $\|b_1\| = 1$ und $[b_1] = [a_1]$.

Im nächsten Schritt bestimmen wir λ derart, dass $\tilde{b}_2 = a_2 + \lambda b_1$ orthogonal zu b_1 ist. Dies führt auf die Gleichung

$$0 = \langle b_1, \tilde{b}_2 \rangle = \langle b_1, a_2 + \lambda b_1 \rangle = \lambda \|b_1\|^2 + \langle b_1, a_2 \rangle,$$

d.h. $\lambda = -\langle b_1, a_2 \rangle$ (beachte $\|b_1\| = 1$).

Es ist $\tilde{b}_2 \neq 0$, da b_1, a_2 linear unabhängig sind. Daher können wir

$$b_2 := \frac{\tilde{b}_2}{\|\tilde{b}_2\|}$$

setzen. Klarerweise ist auch $b_2 \perp b_1$. Nach Konstruktion gilt $b_1, b_2 \in [a_1, a_2]$. b_1, b_2 sind linear unabhängig (da orthonormiert). Daher kann die Basis a_1, a_2 von $[a_1, a_2]$ durch b_1, b_2 ersetzt werden (Austauschsatz von Steinitz, Satz 4.48), woraus

$$[b_1, b_2] = [a_1, a_2]$$

folgt.

Als nächstes bestimmen wir $\mu, \nu \in \mathbb{K}$ derart, dass

$$\tilde{b}_3 = a_3 + \mu b_1 + \nu b_2$$

orthogonal zu b_1, b_2 ist. Die Zahlen μ und ν bestimmt man aus

$$0 = \langle b_1, \tilde{b}_3 \rangle = \langle b_1, a_3 \rangle + \mu \quad \text{und} \quad 0 = \langle b_2, \tilde{b}_3 \rangle = \langle b_2, a_3 \rangle + \nu$$

(beachte $\langle b_1, b_2 \rangle = 0$). Der Vektor $\tilde{b}_3$ ist eine nicht-triviale Linearkombination von a_1, a_2, a_3 (b_1, b_2 sind Linearkombinationen von a_1 bzw. a_1, a_2). Wegen der linearen Unabhängigkeit von a_1, a_2, a_3 ist daher $\tilde{b}_3 \neq o$ und wir können

$$b_3 := \frac{\tilde{b}_3}{\|\tilde{b}_3\|}$$

setzen. Klarerweise sind b_1, b_2, b_3 orthonormiert. Ferner gilt $b_1, b_2, b_3 \in [a_1, a_2, a_3]$, woraus wieder mit Hilfe des Austauschsatzes von Steinitz

$$[b_1, b_2, b_3] = [a_1, a_2, a_3]$$

folgt.

Auf diese Weise fortfahrend erhält man schließlich die Vektoren $b_1, \ldots, b_k$. □

[42] Gram, Jorgen Pedersen, 27. 6. 1850 (Nustrup, Dänemark) – 29. 4. 1916 (Kopenhagen), Beiträge zur Reinen Mathematik (Wahrscheinlichkeitstheorie, Zahlentheorie) und zu praktischen Anwendungen (z.B. in der Forstwirtschaft), hat nie an einer Universität unterrichtet, arbeitete für die Hafnia Versicherungsgesellschaft (www-groups.dcs.st-and.ac.uk/ history/Mathematicians/Gram.html).

[43] Schmidt, Erhard, 13. 1. 1876 (Dorpat, Estland) – 6. 12. 1959 (Berlin), einer der Mitbegründer der abstrakten Funktionalanalysis (www-groups.dcs.st-and.ac.uk/ history/Mathematicians/Schmidt.html).

[44] Dieses Verfahren geht letztendlich auf Laplace zurück.

Korollar 13.10. *Ein endlich-dimensionaler Vektorraum mit innerem Produkt besitzt immer eine Orthonormalbasis.*

Beweis. Es sei $\{a_1, \ldots, a_k\}$ eine beliebige Basis von X. Dann gibt es Vektoren $b_1, \ldots, b_k$ derart, dass $\{b_1, \ldots, b_k\}$ orthonormiert ist und $[b_1, \ldots, b_k] = [a_1, \ldots, a_k] = X$ gilt. □

Orthonormalbasen gestatten eine besonders einfache koordinatenweise Darstellung des inneren Produktes.

Satz 13.11. *Es sei* $\dim X = n$ *und* $\mathcal{B}$ *sei eine geordnete Orthonormalbasis von* X.[45] *Die Koordinatenvektoren von* $x, y \in X$ *bzgl.* $\mathcal{B}$ *seien* ξ *bzw.* η. *Dann gilt:*

$$\langle x, y\rangle = \sum_{i=1}^{n} \bar{\xi}_i \eta_i = \xi^* \eta.$$

Hier bezeichnet ξ^* *den Vektor* $\bar{\xi}^{\mathsf{T}} = (\bar{\xi}_1, \ldots, \bar{\xi}_n)$ *und das Produkt* $\xi^*\eta$ *ist als Matrixprodukt auszuführen. Insbesondere gilt*

$$\|x\| = \Big(\sum_{i=1}^{n} |\xi_i|^2\Big)^{1/2} = (\xi^*\xi)^{1/2}, \quad x \in X.$$

Beweis. Es sei $\mathcal{B} = (b_1, \ldots, b_n)$. Aus $x = \xi_1 b_1 + \cdots + \xi_n b_n$ und $y = \eta_1 b_1 + \cdots + \eta_n b_n$ folgt unter Verwendung von $\langle b_i, b_j\rangle = 0$ für $i \neq j$ und $= 1$ für $i = j$ sofort

$$\langle x, y\rangle = \bar{\xi}_1 \eta_1 + \cdots + \bar{\xi}_n \eta_n.$$

Die Darstellung von $\|x\|$ ist dann trivial. □

Ist X ein reeller oder komplexer Vektorraum mit $\dim X = n$ und $\mathcal{B}$ eine beliebige geordnete Basis von X, so wird durch

$$\langle x, y\rangle_1 := \xi^* \eta,$$

wobei ξ bzw. η den Koordinatenvektor von x bzw. y bzgl. $\mathcal{B}$ bezeichnet, ein inneres Produkt auf X definiert. $\mathcal{B}$ ist dann bzgl. dieses inneren Produktes eine geordnete Orthonormalbasis (Beweis als Übung).

13.2 Orthogonale Projektion

Bei vielen Fragestellungen ist die folgende Aufgabe zu lösen: Gegeben seien ein Unterraum U von X und ein Vektor $x \in X$. Gesucht ist ein Vektor $x_U \in U$ derart, dass

$$\|x - x_U\| = \min_{y \in U} \|x - y\| \tag{13.2}$$

gilt. Eine überaus nützliche Charakterisierung der Lösung dieses Problems gibt der folgende Satz:

Satz 13.12. *Es seien ein Unterraum* U *von* X *und ein Vektor* $x \in X$ *gegeben. Für einen Vektor* $x_U \in U$ *gilt genau dann* (13.2), *wenn*

$$\langle x - x_U, y\rangle = 0 \quad \textit{für alle } y \in U \tag{13.3}$$

[45] $\mathcal{B} = (b_1, \ldots, b_n)$ heißt eine geordnete Orthonormalbasis von X, wenn $\{b_1, \ldots, b_n\}$ eine Orthonormalbasis von X im Sinne von Definition 13.7, b), ist.

gilt. x_U *heißt die* ***orthogonale Projektion*** *von* x *auf* U*.*

Beweis. Es gelte zunächst (13.3). Für beliebiges $h \in U$ gilt

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

woraus sofort (13.2) folgt.

Es sei nun (13.3) nicht erfüllt, d.h. es existiert ein $y_0 \in U$ mit $\langle x-x_U, y_0\rangle \neq 0$. Für $\lambda \in \mathbb{K}$ gilt

$$\begin{aligned}\|x-(x_U+\lambda y_0)\|^2 &= \|x-x_U\|^2 - 2\,\mathrm{Re}\langle x-x_U, \lambda y_0\rangle + |\lambda|^2\|y_0\|^2\\ &= \|x-x_U\|^2 - 2\,\mathrm{Re}\,\lambda\,\mathrm{Re}\langle x-x_U, y_0\rangle\\ &\quad + 2\,\mathrm{Im}\,\lambda\,\mathrm{Im}\langle x-x_U, y_0\rangle + |\lambda_0|^2\|y_0\|^2.\end{aligned}$$

Durch geeignete Wahl von $\mathrm{Re}\,\lambda$ bzw. $\mathrm{Im}\,\lambda$ kann man stets erreichen, dass

$$\|x-(x_U+\lambda y_0)\| < \|x-x_U\|$$

gilt. □

Die in Satz 13.12 gegebene Charakterisierung von x_U erlaubt es nun, im Falle eines endlich-dimensionalen Unterraumes U die Existenz und Eindeutigkeit von x_U zu beweisen.

Satz 13.13. *Es seien die Voraussetzungen von Satz* 13.12 *erfüllt. Ferner sei* $\dim U = k$ *und* $x_1, \ldots, x_k$ *sei eine Basis von* U*. Dann existiert genau ein Vektor* $x_U = \alpha_1 x_1 + \cdots + \alpha_k x_k$ *mit* (13.2)*. Der Vektor* $\alpha = (\alpha_1, \ldots, \alpha_k)^\mathsf{T}$ *ist Lösung des Gleichungssystems*

$$G(x_1, \ldots, x_k)\alpha = \big(\langle x_1, x\rangle, \ldots, \langle x_k, x\rangle\big)^\mathsf{T}.$$

Die Matrix

$$G(x_1, \ldots, x_k) = \begin{pmatrix}\langle x_1, x_1\rangle & \cdots & \langle x_1, x_k\rangle\\ \vdots & & \vdots\\ \langle x_k, x_1\rangle & \cdots & \langle x_k, x_k\rangle\end{pmatrix}$$

heißt die ***Gramsche Matrix*** *der Vektoren* $x_1, \ldots, x_k$*.*

Beweis. Die Bedingung (13.3) ist äquivalent zu $\langle x-x_U, x_i\rangle = 0$ bzw. zu

$$\langle x_i, x-x_U\rangle = 0, \quad i = 1, \ldots, k.$$

Unter Berücksichtigung von $x_U = \alpha_1 x_1 + \cdots + \alpha_k x_k$ ist dies wiederum gleichwertig mit

$$\langle x_i, x\rangle = \sum_{j=1}^{k} \alpha_j \langle x_i, x_j\rangle, \quad i = 1, \ldots, k,$$

bzw. mit

$$G(x_1, \ldots, x_k)\alpha = \big(\langle x_1, x\rangle, \ldots, \langle x_k, x\rangle\big)^\mathsf{T}.$$

Die Existenz und Eindeutigkeit von x_U folgt nun unmittelbar aus dem folgenden Satz:

Satz 13.14. *Es seien* $x_1, \ldots, x_k$ *Vektoren aus* X*. Dann gilt:*

a) *Es ist dann und nur dann* $\det G(x_1, \ldots, x_k) = 0$*, wenn die Vektoren* $x_1, \ldots, x_k$ *linear abhängig sind.*

b) *Es ist stets* $\det G(x_1,\dots,x_k) \geq 0$.

Beweis. a) Es seien $x_1,\dots,x_k$ linear abhängig, d.h. es gilt $\alpha_1 x_1 + \cdots + \alpha_k x_k = o$ mit $(\alpha_1,\dots,\alpha_k) \neq (0,\dots,0)$. Daraus folgt

$$\begin{aligned}&\alpha_1\langle x_i,x_1\rangle + \cdots + \alpha_k\langle x_i,x_k\rangle\\ &\qquad = \big\langle x_i, \alpha_1 x_1 + \cdots + \alpha_k x_k\big\rangle = 0, \quad i=1,\dots,k,\end{aligned}$$

d.h. die Spalten der Gramschen Matrix sind linear abhängig. Es gilt somit $\det G(x_1,\dots,x_k) = 0$.
b) Die Aussage b) folgt unmittelbar aus $\det G(x_1) = \|x\|^2$ und dem folgenden Hilfssatz:

Lemma 13.15. *Es seien linear unabhängige Vektoren* $x_1,\dots,x_k \in X$ *und* $x \in X$ *gegeben.* x_U *sei die Projektion von* x *auf* $U := [x_1,\dots,x_k]$. *Dann gilt:*

$$\|x - x_U\|^2 = \frac{\det G(x_1,\dots,x_k,x)}{\det G(x_1,\dots,x_k)}.$$

Beweis. Setzen wir $x_U = \alpha_1 x_1 + \cdots + \alpha_k x_k$, so ist wegen Satz 13.13 und Satz 9.24 (Cramersche Regel)

$$\alpha_i = \frac{\det G(x_1,\dots,x_{i-1},x_{i+1},\dots,x_k)}{\det G(x_1,\dots,x_k)}, \quad i=1,\dots,k.$$

Wir definieren die Matrizen H_i, $i=1,\dots,k$, und H durch

$$H_i = \left(\begin{array}{ccc|c} \langle x_1,x_1\rangle & \dots & \langle x_1,x_k\rangle & \langle x_1,x\rangle \\ \vdots & & \vdots & \vdots \\ \langle x_k,x_1\rangle & \dots & \langle x_k,x_k\rangle & \langle x_k,x\rangle \\ \hline & {e_i}^{\mathsf{T}} & & 0 \end{array}\right),$$

$$H = \left(\begin{array}{ccc|c} \langle x_1,x_1\rangle & \dots & \langle x_1,x_k\rangle & \langle x_1,x\rangle \\ \vdots & & \vdots & \vdots \\ \langle x_k,x_1\rangle & \dots & \langle x_k,x_k\rangle & \langle x_k,x\rangle \\ \hline 0 & \cdots & 0 & 1 \end{array}\right),$$

wobei ${e_i}^{\mathsf{T}} = (0,\dots,0,1,0,\dots,0)$ mit der Eins an der i-ten Stelle gesetzt wurde. Durch Entwicklung nach der letzten Zeile (Satz 9.21) sieht man leicht, dass

$$\begin{aligned}\det H_i &= -\det G(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_k), \quad i=1,\dots,k,\\ \det H &= \det G(x_1,\dots,x_k)\end{aligned}$$

gilt. Es ist daher

$$x - x_U = \frac{1}{\det G(x_1,\dots,x_k)}\big(\det H\, x + \det H_1\, x_1 + \cdots + \det H_k\, x_k\big)$$

und (beachte $x - x_U \perp x_U$)

$$\begin{aligned}\|x - x_U\|^2 &= \langle x - x_U, x - x_U\rangle\\ &= \frac{1}{\det G(x_1,\dots,x_k)}\Big(\det H\,\langle x,x\rangle + \sum_{i=1}^{k} \det H_i\,\langle x,x_i\rangle\Big).\end{aligned}$$

Auf Grund der speziellen Gestalt der Matrizen H, $H_1, \dots, H_k$ und der Linearität der Determinante in den Zeilen (vgl. Definition 9.1) folgt

$$\|x - x_U\|^2 = \frac{\det G(x_1, \dots, x_k, x)}{\det G(x_1, \dots, x_k)}.$$

□

Mit Hilfe von Hilfssatz 13.15 erhält man:

Satz 13.16 (Hadamardsche[46]Ungleichung). a) *Es seien die Vektoren $x_1, \dots, x_k \in X$ gegeben. Dann ist*

$$\det G(x_1, \dots, x_k) \leq \prod_{i=1}^{k} \|x_i\|^2,$$

wobei das Gleichheitszeichen genau dann gilt, wenn die Vektoren $x_1, \dots, x_k$ orthogonal sind.
b) *Es sei die Matrix $A = (\alpha_{ij})_{i,j=1,\dots,n} \in M_{n,n}(\mathbb{K})$ gegeben. Dann ist*

$$|\det A|^2 \leq \Big(\sum_{i=1}^{n} |\alpha_{i1}|^2\Big) \cdot \dots \cdot \Big(\sum_{i=1}^{n} |\alpha_{in}|^2\Big)$$

Beweis. a) Die Aussage folgt unmittelbar aus

$$\det G(x_1, \dots, x_k, x) \leq \det G(x_1, \dots, x_k)\|x\|^2$$

für beliebige Vektoren $x_1, \dots, x_k, x \in X$. Diese Ungleichung ist trivial, falls $x_1, \dots, x_k, x$ linear abhängig sind (vgl. Satz 13.14, a)). Es seien daher $x_1, \dots, x_k, x$ linear unabhängig und x_U sei die orthogonale Projektion von x auf $U := [x_1, \dots, x_k]$. Dann ist (wegen $x_U \perp x - x_U$ und Hilfssatz 13.15)

$$\begin{aligned}\|x\|^2 &= \langle x_U + x - x_U, x_U + x - x_U\rangle = \langle x_U, x_U\rangle + \langle x - x_U, x - x_U\rangle \\ &\geq \langle x - x_U, x - x_U\rangle = h^2 = \frac{\det G(x_1, \dots, x_k, x)}{\det G(x_1, \dots, x_k)}.\end{aligned}$$

Wegen $x_U = o$ im Falle $x \perp U$ gilt das Gleichheitszeichen genau im Falle $x \perp U$.
b) Es sei $(e_1, \dots, e_n)$ eine Orthonormalbasis von X. Wir setzen $a_j = \sum_{i=1}^n \alpha_{ij} e_i$, $j = 1, \dots, n$. Dann folgt aus a) die Ungleichung

$$\det G(a_1, \dots, a_n) \leq \prod_{j=1}^{n} \|a_j\|^2 = \prod_{j=1}^{n} \Big(\sum_{i=1}^{n} |\alpha_{ij}|^2\Big).$$

Da $(e_1, \dots, e_n)$ eine Orthonormalbasis von X ist, gilt $\langle a_j, a_k\rangle = \sum_{i=1}^n \bar{\alpha}_{ij}\alpha_{ik}$. Es ist daher

$$G(a_1, \dots, a_n) = \Big(\sum_{i=1}^{n} \bar{\alpha}_{ij}\alpha_{ik}\Big)_{j,k=1,\dots,n} = \bar{A}^{\mathsf{T}} A$$

und folglich

$$\det G(a_1, \dots, a_n) = \det \bar{A} \cdot \det A = \overline{\det A} \cdot \det A = |\det A|^2.$$

□

[46]Hadamard, Jacques Solomon, 8. 12. 1865 (Versailles) – 17. 10. 1963 (Paris), Beweis des Primzahlsatzes, Konzept des "korrekt gestellten Problemes" von Rand- und Anfangswertproblemen für partielle Differentialgleichungen (en.wikipedia.ord/wiki/Jacques_Hadamard und www-groups.dcs.st-and.ac.uk/ history/Mathematicians/Hadamard.html).

13.3 Dualräume und duale Abbildungen

In diesem Abschnitt seien X, $Y, \ldots$ beliebige endlich-dimensionale Vektorräume über einem Körper $\mathbb{K}$.

Definition 13.17. a) *Der Vektorraum $X' := L(X, \mathbb{K})$ heißt der zu X **duale Raum**.*
b) *Es sei $\dim X = n$. Ist $\mathcal{B} = (e_1, \ldots, e_n)$ eine geordnete Basis von X, so heißt $\mathcal{B}' = (e_1', \ldots, e_n')$, $e_i' \in X'$, genau dann die zu $\mathcal{B}$ **duale Basis** von X', wenn gilt:*

$$e_i'(e_j) = \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{für } i \neq j, \end{cases} \quad i, j = 1, \ldots, n.$$

Wegen Satz 5.16, b), ist $\dim X' = n \cdot 1 = n$. Die zur Basis $\mathcal{B}$ von X und zur Basis 1 von $\mathbb{K}$ gehörende kanonische Basis von $L(X, \mathbb{K})$ (vgl. Abschnitt 5.3) ist gerade die zu $\mathcal{B}$ duale Basis.

Definition 13.18. *Es sei $\varphi \in L(X, Y)$ gegeben. Die durch*

$$\big(\varphi'(g)\big)(x) := g\big(\varphi(x)\big), \quad g \in Y', \ x \in X,$$

*definierte Abbildung $\varphi' \in L(Y', X')$ heißt die zu φ **duale Abbildung**.*

Man überzeugt sich leicht davon, dass $\varphi'(g)$ für $g \in Y'$ ein Element in X' und φ' eine lineare Abbildung $Y' \to X'$ ist. Bezüglich der Matrixdarstellung der zu φ dualen Abbildung gilt:

Satz 13.19. *Es sei $\varphi \in L(X, Y)$ gegeben und $\mathcal{B}_X$ bzw. $\mathcal{B}_Y$ seien geordnete Basen von X bzw. Y und $\mathcal{B}_X'$ bzw. $\mathcal{B}_Y'$ die zugehörigen dualen Basen von X' bzw. Y'. Ist $A = M(\varphi; \mathcal{B}_X, \mathcal{B}_Y)$, so ist $A^\mathsf{T} = M(\varphi'; \mathcal{B}_Y', \mathcal{B}_X')$.*

Beweis. Es sei $\mathcal{B}_X = (e_1, \ldots, e_n)$, $\mathcal{B}_Y = (f_1, \ldots, f_r)$ und $\mathcal{B}_X' = (e_1', \ldots, e_n')$, $\mathcal{B}_Y' = (f_1', \ldots, f_r')$. Ist $A = (\alpha_{ij})_{\substack{i=1,\ldots,r \\ j=1,\ldots,n}} = M(varphi; \mathcal{B}_X, \mathcal{B}_Y)$, so gilt

$$\varphi(e_j) = \alpha_{1j} f_1 \cdots + \alpha_{rj} f_r, \quad j = 1, \ldots, n.$$

Für die duale Abbildung φ' gilt auf Grund der Definition von φ' und der Definition der dualen Basis

$$\begin{aligned} \big(\varphi'(f_i')\big)(e_j) = f_i'\big(\varphi(e_j)\big) &= f_i'\big(\alpha_{1j} f_1 + \cdots + \alpha_{rj} f_r\big) \\ &= \alpha_{ij}, \quad i = 1, \ldots, r, \ j = 1, \ldots, n, \end{aligned}$$

woraus

$$\varphi'(f_i') = \alpha_{i1} e_1' + \cdots + \alpha_{in} e_n', \quad i = 1, \ldots, r, \tag{13.4}$$

folgt. Dies bedeutet, dass die i-te Spalte der Matrix $M(\varphi'; \mathcal{B}_Y', \mathcal{B}_X')$ durch $(\alpha_{i1}, \ldots, \alpha_{in})^\mathsf{T}$ gegeben ist. Hieraus folgt die Behauptung. □

Satz 13.20. a) *Es seien $\varphi, \psi \in L(X, Y)$ gegeben. Dann ist*

$$(\varphi + \psi)' = \varphi' + \psi'.$$

b) *Es seien $\varphi \in L(Y, Z)$, $\psi \in L(X, Y)$ gegeben. Dann ist*

$$(\varphi\psi)' = \psi'\varphi'.$$

Beweis. a) Für $g \in Y'$ und $x \in X$ gilt

$$\begin{aligned}\big((\varphi+\psi)'(g)\big)(x) &= g\big((\varphi+\psi)(x)\big) = g\big(\varphi(x)+\psi(x)\big) = g\big(\varphi(x)\big)+g\big(\psi(x)\big)\\ &= \big(\varphi'(g)\big)(x)+\big(\psi'(g)\big)(x) = \big(\varphi'(g)+\psi'(g)\big)(x).\end{aligned}$$

Daraus folgt $(\varphi+\psi)'(g) = \varphi'(g)+\psi'(g)$ für alle $g \in Y'$ und somit $(\varphi+\psi)' = \varphi'+\psi'$.
b) Für $h \in Z'$ und $x \in X$ gilt

$$\begin{aligned}\big((\varphi\psi)'(h)\big)(x) &= h\big((\varphi\psi)(x)\big) = h\big(\varphi(\psi(x))\big)\\ &= \big(\varphi'(h)\big)(\psi(x)) = \big(\psi'(\varphi'(h))\big)(x),\end{aligned}$$

d.h. $(\varphi\psi)'(h) = \psi'(\varphi'(h))$ für alle $h \in Z'$, woraus schließlich $(\varphi\psi)' = \psi'\varphi'$ folgt. □

Da man eine Matrix immer als Matrix einer linearen Abbildung zwischen arithmetischen Vektorräumen auffassen kann (vgl. Abschnitt 7.2), folgt aus Satz 13.20 sofort:

Korollar 13.21. a) *Für Matrizen $A, B \in M_{rn}(\mathbb{K})$ gilt*

$$(A+B)^{\mathsf{T}} = A^{\mathsf{T}} + B^{\mathsf{T}}.$$

b) *Für Matrizen $A \in M_{sr}(\mathbb{K})$, $B \in M_{rn}(\mathbb{K})$ gilt*

$$(AB)^{\mathsf{T}} = B^{\mathsf{T}} A^{\mathsf{T}}.$$

Natürlich kann man dieses Korollar auch direkt, ohne Bezugnahme auf duale Abbildungen beweisen.

13.4 Adjungierte Abbildungen

Es sei nun wieder $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Ist X ein Vektorraum über $\mathbb{K}$ mit innerem Produkt $\langle\cdot,\cdot\rangle$, so wird für jeden Vektor $h \in X$ durch $x \to \langle h, x\rangle$ eine lineare Abbildung $X \to \mathbb{K}$ definiert. Der folgende Satz zeigt, dass man im Falle $\dim X < \infty$ auf diese Weise alle Elemente in $X' = L(X,\mathbb{K})$ erhält.

In diesem Abschnitt sind $X, Y, \ldots$ stets endlich-dimensionale Vektorräume über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$ mit innerem Produkt.

Satz 13.22 (Riesz[47]). *Für jedes $f \in X'$ existiert genau ein Vektor $h_f \in X$ mit*

$$f(x) = \langle h_f, x\rangle, \quad x \in X.$$

Die durch $f \to h_f$ definierte Abbildung $\iota_X : X' \to X$ ist bijektiv und ***konjugiert linear*** *(d.h. es ist $\iota_X(\alpha f + \beta g) = \bar{\alpha}\iota_X(f) + \bar{\beta}\iota_X(g)$ für alle $f, g \in X'$ und $\alpha, \beta \in \mathbb{K}$).*[48]

Beweis. Wir wählen eine Orthonormalbasis $(e_1, \ldots, e_n)$ von X. Dann gilt $f(x) = \xi_1 f(e_1) + \cdots + \xi_n f(e_n)$ für $x = \xi_1 e_1 + \cdots + \xi_n e_n \in X$. Definieren wir

$$h_f = \overline{f(e_1)}e_1 + \cdots + \overline{f(e_n)}e_n,$$

so gilt für alle $x = \xi_1 e_1 + \cdots + \xi_n e_n \in X$

$$\langle h_f, x\rangle = f(e_1)\xi_1 + \cdots + f(e_n)\xi_n = f(x).$$

[47] Riesz, Marcel, 16. 11. 1886 (Györ, Ungarn) – 4. 9. 1969 (Lund, Schweden), Arbeitsgebiete: Funktionalanalysis, Partielle Differentialgleichungen, Zahlentheorie, seit 1908 in Schweden (www-groups.dcs.st-and.ac.uk/ history/Mathematicians/Riesz_Marcel.html).

[48] Im Falle $\mathbb{K} = \mathbb{R}$ ist natürlich „konjugiert linear“ dasselbe wie „linear“.

Der Vektor h_f ist eindeutig durch f bestimmt. Denn wäre $\tilde{h} \in X$ ein weiterer Vektor mit $\langle \tilde{h}, x\rangle = f(x)$, $x \in X$, so würde $\langle h_f - \tilde{h}, x\rangle = 0$ für alle $x \in X$ gelten, woraus wegen Hilfssatz 13.2 $h_f = \tilde{h}$ folgt.

Gilt $h_f = h_g$ für $f, g \in X'$, so folgt $f(x) = \langle h_f, x\rangle = \langle h_g, x\rangle = g(x)$ für alle $x \in X$, d.h. $f = g$. Damit ist die Injektivität von ι_X gezeigt. Für $h \in X$ definieren wir $f \in X'$ durch $f(x) = \langle h, x\rangle$, $x \in X$. Dann ist offensichtlich $\iota_X(f) = h$, d.h. ι_X ist auch surjektiv. Für $f, g \in X'$ und $\alpha, \beta \in \mathbb{K}$ gilt

$$\begin{aligned}\big\langle \iota_X(\alpha f + \beta g), x\big\rangle &= (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x)\\ &= \alpha\langle \iota_X(f), x\rangle + \beta\langle \iota_X(g), x\rangle = \big\langle \bar{\alpha}\iota_X(f) + \bar{\beta}\iota_X(g), x\big\rangle\end{aligned}$$

für alle $x \in X$. Daraus folgt (siehe Hilfssatz 13.2)

$$\iota_X(\alpha f + \beta g) = \bar{\alpha}\iota_X(f) + \bar{\beta}\iota_X(g).$$

□

In Hinblick auf den eben bewiesenen Satz können wir lineare Abbildungen $X \to \mathbb{K}$ durch Elemente aus X repräsentieren. Einer Abbildung $\varphi \in L(X, Y)$ können wir dementsprechend eine Abbildung $Y \to X$ zuordnen:

Definition 13.23. *Es sei $\varphi \in L(X, Y)$ gegeben. Die Abbildung*

$$\varphi^* := \iota_X \varphi' \iota_Y^{-1} \in L(Y, X)$$

heißt die zu φ ***adjungierte Abbildung****.*

Man überzeugt sich leicht davon, dass φ^* tatsächlich eine lineare Abbildung $Y \to X$ ist. Die folgende Charakterisierung von φ^* wird vielfach für die Definition der adjungierten Abbildung verwendet:

Satz 13.24. *Es sei $\varphi \in L(X, Y)$ gegeben. Eine Abbildung $\tilde{\varphi} \in L(Y, X)$ ist genau dann die zu φ adjungierte Abbildung, wenn*

$$\langle \tilde{\varphi}(y), x\rangle = \langle y, \varphi(x)\rangle \tag{13.5}$$

für alle $y \in Y$ und $x \in X$ gilt.

Beweis. a) Es sei $\tilde{\varphi} = \varphi^*$, φ^* die zu φ adjungierte Abbildung. Dann folgt mit Hilfe der Definition der adjungierten Abbildung und der Abbildungen ι_X bzw. ι_Y

$$\begin{aligned}\langle \varphi^*(y), x\rangle &= \langle (\iota_X \varphi' \iota_Y^{-1})(y), x\rangle = \big(\varphi'(\iota_Y^{-1}(y))\big)(x)\\ &= \big(\iota_Y^{-1}(y)\big)\big(\varphi(x)\big) = \langle y, \varphi(x)\rangle\end{aligned}$$

für beliebige $y \in Y$ und $x \in X$.
b) Für eine Abbildung $\tilde{\varphi} \in L(Y, X)$ gelte nun

$$\langle \tilde{\varphi}(y), x\rangle = \langle y, \varphi(x)\rangle, \quad y \in Y,\ x \in X. \tag{13.6}$$

Aus $\langle \tilde{\varphi}(y), x\rangle = \big(\iota_X^{-1}(\tilde{\varphi}(y))\big)(x)$ und $\langle y, \varphi(x)\rangle = \big(\iota_Y^{-1}(y)\big)\big(\varphi(x)\big) = \big(\varphi'(\iota_Y^{-1}(y))\big)(x)$ für alle $x \in X$ folgt mittels (13.6)

$$\iota_X^{-1}(\tilde{\varphi}(y)) = \varphi'(\iota_Y^{-1}(y)),$$

d.h. $\tilde{\varphi}(y) = \big(\iota_X \varphi' \iota_Y^{-1}\big)(y) = \varphi^*(y)$ für alle $y \in Y$. □

Eine unmittelbare Folgerung aus (13.5) ist die folgende Aussage:

Korollar 13.25. *Es sei $\varphi \in L(X,Y)$ gegeben. Dann gilt*

$$(\varphi^*)^* = \varphi.$$

In Analogie zu Satz 13.20 gilt:

Satz 13.26. a) *Es seien $\varphi, \psi \in L(X,Y)$ gegeben. Dann ist*

$$(\varphi + \psi)^* = \varphi^* + \psi^*.$$

b) *Es seien $\varphi \in L(Y,Z)$, $\psi \in L(X,Y)$ gegeben. Dann ist*

$$(\varphi\psi)^* = \psi^*\varphi^*.$$

Beweis. Wir beweisen nur die Aussage b). Für beliebige x$\in X$ und $z \in Z$ gilt auf Grund von Satz 13.24

$$\begin{aligned}\langle(\psi^*\varphi^*)(z), x\rangle &= \langle\psi^*(\varphi^*(z)), x\rangle = \langle\varphi^*(z), \psi(x)\rangle \\ &= \langle z, \varphi(\psi(x))\rangle = \langle z, (\varphi\psi)(x)\rangle,\end{aligned}$$

woraus (wieder wegen Satz 13.24) $\psi^*\varphi^* = (\varphi\psi)^*$ folgt. □

X sei ein Vektorraum mit innerem Produkt und $\dim X = n$. Ist $(e_1, \ldots, e_n)$ eine Basis von X und $(e_1', \ldots, e_n')$ die dazugehörige duale Basis von X', so ist $(\iota_X(e_1'), \ldots, \iota_X(e_n'))$ wieder eine Basis von X (wir überlassen den einfachen Beweis der linearen Unabhängigkeit der Vektoren $\iota_X(e_i')$ dem Leser). Wegen der Beziehung

$$\langle\iota_X(e_i'), e_j\rangle = e_i'(e_j) = \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{für } i \neq j, \end{cases} \qquad i, j = 1, \ldots, n,$$

gilt $\iota_X(e_i') = e_i$, $i = 1, \ldots, n$, genau dann, wenn $(e_1, \ldots, e_n)$ eine Orthonormalbasis von X ist. Für die Matrixdarstellung von adjungierten Abbildungen werden wir daher meist Orthonormalbasen zugrundelegen.

Satz 13.27. *Es sei $\varphi \in L(X,Y)$ gegeben. Ferner seien $\mathcal{B}_X$ bzw. $\mathcal{B}_Y$ Orthonormalbasen von X bzw. Y. Ist $A = M(\varphi; \mathcal{B}_X, \mathcal{B}_Y)$, so ist*

$$A^* := \bar{A}^{\mathsf{T}}$$

die Matrix von φ^* bzgl. $\mathcal{B}_Y$ und $\mathcal{B}_X$.

Beweis. Es sei $\mathcal{B}_X = (e_1, \ldots, e_n)$, $\mathcal{B}_Y = (f_1, \ldots, f_r)$, $A = (\alpha_{ij})_{\substack{i=1,\ldots,r \\ j=1,\ldots,n}}$ und $A^* = (\beta_{k\ell})_{\substack{k=1,\ldots,n \\ \ell=1,\ldots,r}}$. Dann gilt (siehe Satz 5.17) $\varphi^*(f_j) = \beta_{1j}e_1 + \cdots + \beta_{nj}e_n$, $j = 1, \ldots, r$, $\varphi(e_i) = \alpha_{1i}f_1 + \cdots + \alpha_{ri}f_r$, $i = 1, \ldots, n$, und wegen Gleichung (13.5)

$$\bar{\beta}_{ij} = \langle\varphi^*(f_j), e_i\rangle = \langle f_j, \varphi(e_i)\rangle = \alpha_{ji},$$

d.h. $\beta_{ij} = \bar{\alpha}_{ji}$, $i = 1, \ldots, n$, $j = 1, \ldots, r$. □

Im Falle $\mathbb{K} = \mathbb{R}$ ist natürlich $A^* = A^{\mathsf{T}}$. Eine direkte Folge von Satz 13.27 sind die folgenden Aussagen:

Korollar 13.28. a) *Es sei $\varphi \in L(X,Y)$ gegeben. Dann ist*

$$\operatorname{rang}\varphi^* = \operatorname{rang}\varphi.$$

b) *Es sei $\varphi \in L(X,X)$ gegeben. Dann ist $\lambda \in \mathbb{K}$ genau dann ein Eigenwert von φ, wenn $\bar{\lambda}$ ein Eigenwert von φ^* ist.*

Beweis. a) Ist A die Matrix von φ bzgl. Orthonormalbasen in X und Y, so ist A^* die Matrix von φ^* bzgl. dieser Basen. Dann gilt $\operatorname{rang}\varphi^* = \operatorname{rang}A^* = \operatorname{rang}\bar{A}^{\mathsf{T}} = \operatorname{rang}\bar{A} = \operatorname{rang}A = \operatorname{rang}\varphi$.
b) Wir wählen eine Orthonormalbasis $\mathcal{B}$ von X. Die Matrix von φ und φ^* bzgl. $\mathcal{B}$ ist A bzw. A^*. Das Resultat folgt wegen Satz 10.2 aus $\det(A^* - \bar{\lambda}E) = \det\overline{(A-\lambda E)}^{\mathsf{T}} = \det\overline{(A-\lambda E)} = \overline{\det(A-\lambda E)} = 0$. □

Betreffend die Eigenvektoren von φ und φ^* gilt:

Satz 13.29. *Es sei $\varphi \in L(X,X)$ gegeben. Ferner sei a ein Eigenvektor von φ zum Eigenwert λ und b ein Eigenvektor von φ^* zum Eigenwert μ. Dann gilt*

$$a \perp b,$$

falls $\mu \neq \bar{\lambda}$ ist.[49]

Beweis. Aus $\varphi(a) = \lambda a$ und $\varphi^*(b) = \mu b$ folgt

$$\bar{\lambda}\langle a,b\rangle = \langle \varphi(a), b\rangle = \langle a, \varphi^*(b)\rangle = \mu\langle a,b\rangle,$$

d.h. $(\bar{\lambda} - \mu)\langle a,b\rangle = 0$. Wegen $\bar{\lambda} \neq \mu$ folgt $\langle a,b\rangle = 0$. □

Es besteht ein wichtiger Zusammenhang zwischen dem Kern einer linearen Abbildung und dem Bildbereich der dazu adjungierten Abbildung:

Satz 13.30. *Es sei $\varphi \in L(X,Y)$ gegeben. Dann gilt*

$$\ker\varphi = \varphi^*(Y)^\perp, \quad \ker\varphi^* = \varphi(X)^\perp$$

und

$$X = \ker\varphi \oplus \varphi^*(Y), \quad Y = \ker\varphi^* \oplus \varphi(X).$$

Beweis. a) Es sei $a \in \ker\varphi$. Dann gilt für alle $y \in Y$

$$\langle a, \varphi^*(y)\rangle = \langle \varphi(a), y\rangle = \langle o, y\rangle = 0,$$

d.h. $\ker\varphi \subset \varphi^*(Y)^\perp$. Ist umgekehrt $a \in \varphi^*(Y)^\perp$, so gilt

$$0 = \langle a, \varphi^*(y)\rangle = \langle \varphi(a), y\rangle \quad \text{für alle } y \in Y,$$

d.h. $\varphi(a) = o$ (vgl. Hilfssatz 13.2). Es gilt somit auch $\varphi^*(Y)^\perp \subset \ker\varphi$.
b) Für $x \in X$ sei x_1 die orthogonale Projektion auf $\varphi^*(Y)$. Wir setzen $x_2 = x - x_1$. Dann ist $x_2 \perp \varphi^*(Y)$ (vgl. Satz 13.12 und Satz 13.13). Dies bedeutet $x_2 \in \ker\varphi$. Somit gilt $X = \ker\varphi + \varphi^*(Y)$. Offensichtlich ist die Summe direkt.

Der Beweis für die übrigen Aussagen folgt mittels $\varphi = (\varphi^*)^*$ sofort. □

[49] Im Falle $\mathbb{K} = \mathbb{R}$ bezieht sich dieser Satz natürlich auf reelle Eigenwerte.

Kapitel 14

Spezielle lineare Abbildungen

In diesem Kapitel bezeichnet $X, Y, \ldots$ stets endlich-dimensionale Vektorräume über $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$ mit innerem Produkt.

14.1 Normale Abbildungen und Matrizen

Definition 14.1. a) *Eine Abbildung $\varphi \in L(X, X)$ heißt genau dann **normal**, wenn*

$$\varphi\varphi^* = \varphi^*\varphi$$

gilt.
b) *Eine Matrix $A \in M_{n,n}(\mathbb{K})$ heißt genau dann **normal**, wenn*

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

gilt.

Es ist offensichtlich, dass die Matrix einer normalen Abbildung bzgl. einer Orthonormalbasis eine normale Matrix ist.

Satz 14.2. *Eine Abbildung $\varphi \in L(X, X)$ ist genau dann normal, wenn*

$$\langle \varphi(x), \varphi(y) \rangle = \langle \varphi^*(x), \varphi^*(y) \rangle \tag{14.1}$$

für alle $x, y \in X$ gilt. Insbesondere gilt für normale Abbildungen stets

$$\|\varphi(x)\| = \|\varphi^*(x)\|, \quad x \in X. \tag{14.2}$$

Beweis. Ist φ normal, so gilt für beliebige $x, y \in X$

$$\langle \varphi(x), \varphi(y) \rangle = \langle x, (\varphi^*\varphi)(y) \rangle = \langle x, (\varphi\varphi^*)(y) \rangle = \langle \varphi^*(x), \varphi^*(y) \rangle.$$

Umgekehrt folgt aus (14.1)

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

für alle $x, y \in X$. Aus dieser Gleichung folgt wegen Hilfssatz 13.2

$$(\varphi^*\varphi)(y) = (\varphi\varphi^*)(y) \quad \text{für alle } y \in X.$$

□

Eine direkte Folgerung aus (14.2) ist:

Korollar 14.3. *Es sei $\varphi \in L(X,X)$ normal. Dann ist*

$$\ker \varphi = \ker \varphi^*$$

und

$$X = \ker \varphi \oplus \varphi(X).$$

Beweis. Wegen (14.2) gilt $x \in \ker \varphi$ genau dann, wenn $x \in \ker \varphi^*$ ist. Die Darstellung von X als direkte Summe folgt aus Satz 13.30. □

Satz 14.4. *Es sei $\varphi \in L(X,X)$ normal. Dann gilt:*

a) *Ein Vektor $a \in X$ ist genau dann Eigenvektor von φ zum Eigenwert λ, wenn a Eigenvektor von φ^* zum Eigenwert $\bar{\lambda}$ ist.*

b) *Ist a bzw. b Eigenvektor von φ zum Eigenwert λ bzw. μ mit $\lambda \neq \mu$, so gilt*

$$a \perp b$$

c) *Es ist*

$$\varphi^k(X) = \varphi(X), \quad k = 1, 2, \dots,$$

und folglich

$$\operatorname{rang} \varphi^k = \operatorname{rang} \varphi, \quad k = 1, 2, \dots .$$

Beweis. Man sieht leicht, dass für beliebige $\lambda \in \mathbb{K}$ mit φ auch die Abbildungen $(\varphi - \lambda\epsilon)^k$, $k = 1, 2, \dots$, normal sind. Nach Korollar 14.3 gilt somit

$$\ker(\varphi - \lambda\epsilon) = \ker(\varphi^* - \bar{\lambda}\epsilon),$$

d.h. $(\varphi - \lambda\epsilon)(a) = o$ ist gleichwertig mit $(\varphi^* - \bar{\lambda}\epsilon)(a) = o$. Dies beweist die Aussage a).

Wegen a) ist b Eigenvektor von φ^* zum Eigenwert $\bar{\mu}$. Nach Satz 13.29 gilt $a \perp b$, falls $\bar{\mu} \neq \bar{\lambda}$, d.h. $\mu \neq \lambda$ gilt.

Um c) zu beweisen genügt es, $\varphi^2(X) = \varphi(X)$ zu zeigen. Es ist stets $\varphi^2(X) \subset \varphi(X)$. Es sei nun $x \in \varphi(X)$ gegeben. Für ein $y \in X$ gilt dann $x = \varphi(y)$. Wegen Korollar 14.3 ist $y = y_1 + y_2$ mit $y_1 \in \ker \varphi$ und $y_2 \in \varphi(X)$. Daraus folgt

$$x = \varphi(y_1 + y_2) = \varphi(y_2) \in \varphi^2(X).$$

□

Die Aussage c) des eben bewiesenen Satzes hat weitreichende Konsequenzen für die Struktur normaler Matrizen:

Satz 14.5. *Es sei $A \in M_{n,n}(\mathbb{C})$ gegeben. Dann ist A genau dann normal, wenn es eine reguläre Matrix $U \in M_{n,n}(\mathbb{C})$ gibt, sodass*

$$U^* = U^{-1},$$
$$U^* A U = \operatorname{diag}(\lambda_1, \dots, \lambda_n),$$

gilt. $\lambda_1, \dots, \lambda_n$ sind natürlich die Eigenwerte von A.

Beweis. Wegen $\mathbb{K} = \mathbb{C}$ besitzt die Matrix A genau n Eigenwerte, wobei jeder Eigenwert entsprechend seiner algebraischen Vielfachheit gezählt wird. Es sei λ_i ein Eigenwert von A und k_i sei der Exponent von $\lambda - \lambda_i$ im Minimalpolynom von A. Die Matrix A ist die Matrix des Endomorphismus φ_A bzgl. der kanonischen Basis von $\mathbb{C}^n$ (vgl. Abschnitt 7.2). Da die kanonische Basis von $\mathbb{C}^n$ orthonormiert ist, muß φ_A normal sein. Auf Grund von Satz 12.5, f), und Satz 14.4, c), (angewandt auf $\varphi_A - \lambda_i \epsilon$) muß $k_i = 1$ sein. Aus Satz 12.6 folgt, dass A diagonalisierbar ist.

Auf Grund von Satz 10.7 gibt es eine Basis von $\mathbb{C}^n$ bestehend aus Eigenvektoren von A. Ist λ ein Eigenwert von A mit algebraischer Vielfachheit k, so existieren k linear unabhängige Eigenvektoren von A zu λ (Korollar 10.8). Diese Eigenvektoren können wir orthonormiert annehmen (Satz 13.9). Nach Satz 14.4, b), sind Eigenvektoren zu verschiedenen Eigenwerten von A orthogonal. Insgesamt können wir daher annehmen, dass eine Orthonormalbasis $a_1, \dots, a_n$ von $\mathbb{C}^n$ bestehend aus Eigenvektoren von A existiert. Wir setzen $U := (a_1, \dots, a_n)$. Wegen $a_i^* a_j = 0$ für $i \neq j$ und $a_i^* a_i = 1$, $i, j = 1, \dots, n$, gilt

$$U^* A U = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \quad \text{und} \quad U^* U = E,$$

d.h. insbesondere auch $U^* = U^{-1}$.

Ist umgekehrt $U^* A U = D$ mit einer regulären Matrix $U \in M_{n,n}(\mathbb{C})$ und einer Diagonalmatrix D, wobei $U^* = U^{-1}$ gilt, so ist $A = U D U^*$ und $A^* = U \bar{D} U^*$. Daraus folgt

$$A A^* = U D \bar{D} U^* = U \bar{D} D U^* = A^* A.$$

□

Eine unmittelbare Folgerung aus dem oben bewiesenen Satz und aus Satz 14.4, b), ist:

Korollar 14.6. *Es sei $\varphi \in L(X, X)$ normal. Dann existiert ein Polynom $p \in \mathbb{C}[\lambda]$ mit*

$$\varphi^* = p(\varphi) \quad \textit{und} \quad \varphi = \bar{p}(\varphi^*),$$

wobei $\bar{p} = p_1 - i p_2$ für $p = p_1 + i p_2$, $p_1, p_2 \in \mathbb{R}[\lambda]$ ist. Die analoge Aussage gilt für normale Matrizen in $M_{n,n}(\mathbb{K})$.

Ist die Matrix von φ bzgl. einer Orthonormalbasis von X reell, so kann auch $p \in \mathbb{R}[\lambda]$ gewählt werden. Insbesondere kann $p \in \mathbb{R}[\lambda]$ gewählt werden, falls φ eine normale Abbildung eines reellen Vektorraumes ist.

Beweis. Es sei zunächst $\mathbb{K} = \mathbb{C}$. Auf Grund von Satz 14.5 existiert eine Orthonormalbasis von X, sodass die Matrizen von φ bzw. φ^* bzgl. dieser Basis Diagonalmatrizen sind, $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ bzw. $\bar{D} = \operatorname{diag}(\bar{\lambda}_1, \dots, \bar{\lambda}_n)$. Wir wählen ein Polynom $p \in \mathbb{C}[\lambda]$ mit

$$p(\lambda_i) = \bar{\lambda}_i, \quad i = 1, \dots, n.$$ [50]

Dann ist auch $\lambda_i = \bar{p}(\bar{\lambda}_i)$, $i = 1, \dots, n$, und

$$p(D) = \operatorname{diag}\big(p(\lambda_1), \dots, p(\lambda_n)\big) = \bar{D}, \quad \bar{p}(\bar{D}) = D.$$

Daraus folgt sofort die Behauptung, weil $p(D)$ bzw. $\bar{p}(\bar{D})$ die Matrix von $p(\varphi)$ bzw. $\bar{p}(\varphi^*)$ bzgl. der gewählten Orthonormalbasis ist (vgl. die Ausführungen am Ende von Anhang A).

Es sei nun $A \in M_{n,n}(\mathbb{R})$ die Matrix von φ bzgl. einer Orthonormalbasis von X. Dann ist $A^* = A^{\mathsf{T}} \in M_{n,n}(\mathbb{R})$ die Matrix von φ^* bzgl. derselben Basis. Nach der bereits bewiesenen Aussage existiert ein Polynom $p = p_1 + i p_2$, $p_1, p_2 \in \mathbb{R}[\lambda]$, mit

$$A^* = p(A) = p_1(A) + i p_2(A).$$

[50] Polynome mit dieser Eigenschaft existieren immer.

Wegen $A^* \in M_{n,n}(\mathbb{R})$ muß $p_2(A) = 0$ sein, d.h. $A^* = p_1(A)$ mit $p_1 \in \mathbb{R}[\lambda]$ und folglich auch $A = p_1(A^*)$.

Ist X reell und $\varphi \in L(X,X)$ normal, so ist die Matrix A von φ bzgl. irgendeiner Orthonormalbasis von X ebenfalls reell und die Behauptung folgt aus der eben bewiesenen Aussage über A. □

Korollar 14.7. *Es sei $\varphi \in L(X,X)$ normal. Ist der Unterraum U von X φ-invariant, so ist auch $U^\perp$ φ-invariant.*

Beweis. Es sei $z \in U^\perp$ gewählt. Dann gilt wegen der φ-Invarianz von U

$$\langle \varphi^*(z), u\rangle = \langle z, \varphi(u)\rangle = 0$$

für alle $u \in U$, d.h. $\varphi^*(z) \in U^\perp$. Damit ist gezeigt, dass $U^\perp$ φ^*-invariant ist.[51] Auf Grund von Korollar 14.6 ist $\varphi = \bar{p}(\varphi^*)$ mit $\bar{p} \in \mathbb{K}[\lambda]$. Daraus folgt, dass $U^\perp$ ebenfalls φ-invariant ist. Man beachte, dass im Falle $\mathbb{K} = \mathbb{R}$ auch $\bar{p} \in \mathbb{R}[\lambda]$ ist. □

14.2 Unitäre Abbildungen und Matrizen

Die in Satz 14.5 aufscheinende Transformationsmatrix U gehört einer Klasse von Matrizen an, die eine wichtige Rolle bei Basistransformationen spielt:

Definition 14.8. a) *Es sei $\varphi \in L(X,X)$ gegeben. φ heißt genau dann **unitär** oder **isometrisch**, wenn φ invertierbar ist und*

$$\varphi^* = \varphi^{-1}$$

gilt.

b) *Eine Matrix $U \in M_{n,n}(\mathbb{K})$ heißt genau dann **unitär**, wenn U regulär ist und*

$$U^* = U^{-1}$$

*gilt. Im Falle $\mathbb{K} = \mathbb{R}$ heißt U **orthogonal**.*

Offensichtlich ist eine unitäre Abbildung $\varphi \in L(X,X)$ wegen $\varphi^*\varphi = \varphi^{-1}\varphi = \varphi\varphi^{-1} = \varphi\varphi^*$ immer normal. Die Matrix einer unitären Abbildung bzgl. einer Orthonormalbasis von X ist eine unitäre bzw. orthogonale Matrix (vgl. Satz 13.27 und Abschnitt 7.2). Das Kompositum $\varphi\psi$ unitärer Abbildungen $\varphi, \psi \in L(X,X)$ ist wegen $(\varphi\psi)^* = \psi^*\varphi^* = \psi^{-1}\varphi^{-1} = (\varphi\psi)^{-1}$ wieder unitär. Zu einer unitären Abbildung $\varphi \in L(X,X)$ ist φ^{-1} ebenfalls unitär (wegen $(\varphi^{-1})^* = (\varphi^*)^* = \varphi = (\varphi^{-1})^{-1}$). Die unitären Abbildungen aus $L(X,X)$ bilden somit eine Gruppe (bzgl. der Komposition von Abbildungen), die sogenannte ***unitäre Gruppe***. Für unitäre bzw. orthogonale Matrizen gelten die analogen Aussagen.

Die unitären Abbildungen können durch eine Reihe von weiteren Eigenschaften charakterisiert werden:

Satz 14.9. *Es sei $\varphi \in L(X,X)$ gegeben. Dann sind die folgenden Aussagen äquivalent:*

(i) *φ ist unitär.*

(ii) *Es ist*

$$\|\varphi(x)\| = \|x\|$$

für alle $x \in X$.

[51] Man beachte, dass dies für beliebige Abbildungen $\varphi \in L(X,X)$ gilt.

(iii) *Es gilt*
$$\langle \varphi(x), \varphi(y) \rangle = \langle x, y \rangle$$
für alle $x, y \in X$.

(iv) *Ist* $(e_1, \ldots, e_n)$ *eine Orthonormalbasis von* X, *so ist auch* $(\varphi(e_1), \ldots$ $\ldots, \varphi(e_n))$ *eine Orthonormalbasis von* X.

Beweis. Wir führen den Beweis, indem wir die Implikationen (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (i) beweisen.
1. (i) $\Longrightarrow$ (ii).
Es sei φ unitär und $x \in X$ gewählt. Dann ist
$$\|\varphi(x)\|^2 = \langle \varphi(x), \varphi(x) \rangle = \langle x, (\varphi^*\varphi)(x) \rangle = \langle x, x \rangle = \|x\|^2.$$
2. (ii) $\Longrightarrow$ (iii).
Für $x, y \in X$ gilt wegen (ii) die Beziehung
$$\langle \varphi(x+y), \varphi(x+y) \rangle = \langle x+y, x+y \rangle = \langle x, x \rangle + 2\,\mathrm{Re}\langle x, y \rangle + \langle y, y \rangle.$$
Andererseits gilt (ebenfalls wegen (ii)):
$$\begin{aligned}\langle \varphi(x+y), \varphi(x+y) \rangle &= \langle \varphi(x) + \varphi(y), \varphi(x) + \varphi(y) \rangle \\ &= \langle \varphi(x), \varphi(x) \rangle + 2\,\mathrm{Re}\langle \varphi(x), \varphi(y) \rangle + \langle \varphi(y), \varphi(y) \rangle \\ &= \langle x, x \rangle + 2\,\mathrm{Re}\langle \varphi(x), \varphi(y) \rangle + \langle y, y \rangle.\end{aligned}$$
Wir erhalten somit
$$\mathrm{Re}\langle \varphi(x), \varphi(y) \rangle = \mathrm{Re}\langle x, y \rangle \quad \text{für alle } x, y \in X. \tag{14.3}$$
Im Falle $\mathbb{K} = \mathbb{R}$ ist damit (iii) bewiesen. Im Falle $\mathbb{K} = \mathbb{C}$ können wir in (14.3) x durch ix ersetzen und erhalten
$$\mathrm{Im}\langle \varphi(x), \varphi(y) \rangle = \mathrm{Im}\langle x, y \rangle \quad \text{für alle } x, y \in X.$$
Dies und (14.3) beweisen (iii).
3. (iii) $\Longrightarrow$ (iv).
Ist $(e_1, \ldots, e_n)$ eine Orthonormalbasis von X, so ist wegen (iii) auch $(\varphi(e_1), \ldots, \varphi(e_n))$ eine Orthonormalbasis.
4. (iv) $\Longrightarrow$ (i).
Es sei $(e_1, \ldots, e_n)$ irgendeine Orthonormalbasis von X. Wegen (iv) gilt $\operatorname{rang} \varphi = n$, d.h. φ ist invertierbar. Da $(\varphi(e_1), \ldots, \varphi(e_n))$ ebenfalls eine Orthonormalbasis von X ist, gilt
$$\langle e_i, e_j \rangle = \langle \varphi(e_i), \varphi(e_j) \rangle = \langle e_i, (\varphi^*\varphi)(e_j) \rangle, \quad i, j = 1, \ldots, n.$$
Für festes j bedeutet dies nach Hilfssatz 13.2
$$(\varphi^*\varphi)(e_j) = e_j, \quad j = 1, \ldots, n,$$
woraus (siehe Satz 5.8) $\varphi^*\varphi = \epsilon$ folgt. $\square$

Da in den Spalten einer Matrix $U \in M_{n,n}(\mathbb{K})$ die Bilder der Vektoren der kanonischen Basis von $\mathbb{K}^n$ stehen, folgt aus Satz 14.9 sofort:

Eine Matrix $U \in M_{n,n}(\mathbb{K})$ *ist genau dann unitär bzw. orthogonal, wenn die Spalten von* U *eine Orthonormalbasis von* $\mathbb{K}^n$ *bilden.*

Aus Satz 14.9 und aus der Definition unitärer Abbildungen bzw. Matrizen folgt sofort:

Satz 14.10. *Es sei $\varphi \in L(X,X)$ unitär. Dann gilt:*
a) *Ist $\lambda \in \mathbb{K}$ ein Eigenwert von φ, so gilt $|\lambda| = 1$.*
b) *Es ist $|\det \varphi| = 1$.*
Analoge Aussagen gelten für orthogonale Matrizen.

Beweis. a) Ist λ Eigenwert von φ und x ein zugehöriger Eigenvektor, so gilt (siehe Satz 14.9, (ii))

$$\|x\| = \|\varphi(x)\| = \|\lambda x\| = |\lambda|\,\|x\|.$$

Wegen $\|x\| \neq 0$ folgt daraus $|\lambda| = 1$.
b) Es sei U die Matrix von φ bzgl. einer Orthonormalbasis von X. Dann gilt $\det \varphi = \det U$. Wegen Satz 9.2, (ix), folgt

$$\det U^* \cdot \det U = \det E = 1.$$

Ferner gilt $\det U^* = \det \bar{U}^{\mathsf{T}} = \det \bar{U} = \overline{\det U}$. Damit ist

$$1 = \det U^* \cdot \det U = \overline{\det U} \cdot \det U = |\det U|^2.$$

□

Man beachte, dass in Aussage a) von Satz 14.10 nicht die Existenz von Eigenwerten behauptet wird. Im Falle $\mathbb{K} = \mathbb{R}$ kann eine orthogonale Matrix überhaupt keine Eigenwerte (in $\mathbb{R}$) besitzen, wie das Beispiel $U = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ zeigt. Der folgende Satz charakterisiert die Struktur isometrischer Abbildungen reeller Vektorräume:

Satz 14.11. *Es sei $U \in M_{n,n}(\mathbb{R})$ orthogonal, $\sigma_j = \pm 1$, $j = 1, \dots, p$, seien die reellen Eigenwerte von U und $\lambda_j = e^{i\theta_j}$, $\bar{\lambda}_j = e^{-i\theta_j}$, $0 \leq \theta_j < 2\pi$, $j = 1, \dots, k$, seien die nicht-reellen Eigenwerte von U (U aufgefaßt als Matrix in $M_{n,n}(\mathbb{C})$), $n = p + 2k$. Dann gibt es eine orthogonale Matrix $S \in M_{n,n}(\mathbb{R})$ mit*

$$S^{\mathsf{T}} U S = \begin{pmatrix} \sigma_1 & & & & & & & \\ & \ddots & & & & O & & \\ & & \sigma_p & & & & & \\ & & & \cos\theta_1 & -\sin\theta_1 & & & \\ & & & \sin\theta_1 & \cos\theta_1 & & & \\ & & & & & \ddots & & \\ & & O & & & & \cos\theta_k & -\sin\theta_k \\ & & & & & & \sin\theta_k & \cos\theta_k \end{pmatrix}.$$

Dies bedeutet insbesondere, dass zu jeder isometrischen Abbildung φ eines reellen, n-dimensionalen Vektorraumes eine geordnete Orthonormalbasis $\mathcal{B}$ existiert, sodass die Matrix von $M(\varphi, \mathcal{B})$ die oben angegebene Gestalt besitzt.

Beweis. Die Matrix U ist normal. Daher ist U nach Satz 14.5 unitär-ähnlich zur Diagonalmatrix $D = \operatorname{diag}(\sigma_1, \dots, \sigma_p, \lambda_1, \dots, \lambda_k, \bar{\lambda}_1, \dots, \bar{\lambda}_k)$,

$$\tilde{S}^* U S = D,$$

wobei die Spalten von $\tilde{S}$ die Eigenvektoren von U zu den Eigenwerten $\sigma_1, \dots, \sigma_p, \lambda_1, \dots, \lambda_k, \bar{\lambda}_1, \dots, \bar{\lambda}_k$ in $\mathbb{C}^n$ sind. Die im Satz angegebene Form der Matrix $S^{\mathsf{T}} U S$ ist gerade die reelle Normalform von U (vgl. Satz 12.20). Ohne Einschränkung können wir annehmen, dass

$$\tilde{S} = (\epsilon_1, \dots, \epsilon_p, \alpha_1, \dots, \alpha_k, \bar{\alpha}_1, \dots, \bar{\alpha}_k)$$

gilt mit $\epsilon_j \in \mathbb{R}^n$, $j = 1, \ldots, p$, und $\alpha_j \in \mathbb{C}^n$, $j = 1, \ldots, k$. In leichter Modifikation (nämlich durch Einfügen der Faktoren „$\sqrt{2}$") der im Beweis von Satz 12.20 konstruierten Transformationsmatrix setzen wir

$$S = \big(\epsilon_1, \ldots, \epsilon_p, \sqrt{2}\beta_1, \sqrt{2}\gamma_1, \ldots, \sqrt{2}\beta_k, \sqrt{2}\gamma_k\big) \tag{14.4}$$

mit $\beta_j = \operatorname{Re}\alpha_j$, $\gamma_j = -\operatorname{Im}\alpha_j$, $j = 1, \ldots, k$. Da die Spalten von $\tilde{S}$ in $\mathbb{C}^n$ orthonormiert sind, ist klar, dass dies auch für $\epsilon_1, \ldots, \epsilon_p$ in $\mathbb{R}^n$ gilt. Ferner folgt aus $0 = \alpha_j^*\epsilon_\ell = ({\beta_j}^\mathsf{T} + i{\gamma_j}^\mathsf{T})\epsilon_\ell = {\beta_j}^\mathsf{T}\epsilon_\ell + i{\gamma_j}^\mathsf{T}\epsilon_\ell$ sofort ${\beta_j}^\mathsf{T}\epsilon_\ell = 0$ und ${\gamma_j}^\mathsf{T}\epsilon_\ell = 0$, $j = 1, \ldots, k$, $\ell = 1, \ldots, p$.

Für $j \neq \ell$ folgen aus $\alpha_j^*\alpha_\ell = 0 = \alpha_j^*\bar{\alpha}_\ell$ die Gleichungen

$$\begin{aligned} {\beta_j}^\mathsf{T}\beta_\ell + {\gamma_j}^\mathsf{T}\gamma_\ell = 0, {\beta_j}^\mathsf{T}\beta_\ell - {\gamma_j}^\mathsf{T}\gamma_\ell = 0,\\ {\gamma_j}^\mathsf{T}\beta_\ell - {\beta_j}^\mathsf{T}\gamma_\ell = 0, {\beta_j}^\mathsf{T}\gamma_\ell + {\gamma_j}^\mathsf{T}\beta_\ell = 0. \end{aligned}$$

Hieraus erhält man sofort ${\beta_j}^\mathsf{T}\beta_\ell = {\gamma_j}^\mathsf{T}\gamma_\ell = {\beta_j}^\mathsf{T}\gamma_\ell = {\gamma_j}^\mathsf{T}\beta_\ell = 0$, $j, \ell = 1, \ldots, k$, $j \neq \ell$.

Schließlich folgen aus $\alpha_j^*\bar{\alpha}_j = 0$ und $\alpha_j^*\alpha_j = 1$ die Gleichungen

$${\beta_j}^\mathsf{T}\beta_j - {\gamma_j}^\mathsf{T}\gamma_j + 2i{\beta_j}^\mathsf{T}\gamma_j = 0 \quad \text{und} \quad {\beta_j}^\mathsf{T}\beta_j + {\gamma_j}^\mathsf{T}\gamma_j = 1.$$

Daraus erhält man ${\beta_j}^\mathsf{T}\gamma_j = 0$ und ${\beta_j}^\mathsf{T}\beta_j = {\gamma_j}^\mathsf{T}\gamma_j = 1/2$, $j = 1, \ldots, k$. Es gilt daher $\|\sqrt{2}\beta_j\| = \|\sqrt{2}\gamma_j\| = 1$.

Durch direktes Nachrechnen bestätigt man übrigens leicht, dass $US = SD$ gilt. □

Aus Satz 14.11 folgt nun unmittelbar, dass für eine isometrische Transformation $\varphi \in L(X, X)$, der reelle Vektorraum X die direkte Summe φ-invarianter, paarweise orthogonaler Unterräume der Dimensionen 1 oder 2 ist. Die ein-dimensionalen Unterräume werden von den reellen Eigenvektoren von φ zu den Eigenwerten 1 bzw. -1 aufgespannt, während die zwei-dimensionalen Unterräume jeweils vom Realteil und Imaginärteil der komplexen Eigenvektoren von φ zu den nicht-reellen Eigenwerten aufgespannt werden. Auf den ein-dimensionalen Unterräumen ist die Einschränkung von φ entweder die Identität (falls $\sigma_j = 1$) oder die Spiegelung am Ursprung (falls $\sigma_j = -1$). Wir wollen noch klären, wie φ eingeschränkt auf die zwei-dimensionalen Unterräume wirkt.

Es sei $\dim X = 2$. $\varphi \in L(X, X)$ sei eine isometrische Abbildung mit einem Paar $\lambda = e^{i\theta}, \bar{\lambda}$, $0 \leq \theta < 2\pi$, nicht-reeller Eigenwerte. Bezüglich einer geordneten Orthonormalbasis $\mathcal{B} = (e_1, e_2)$ von X sei

$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

die Matrix von φ bzgl. $\mathcal{B}$. Zur Basis $\mathcal{B}$ definieren wir auf X die Determinantenform Φ_0 durch

$$\Phi_0(x, y) = \det\begin{pmatrix} \xi_1 & \eta_1 \\ \xi_2 & \eta_2 \end{pmatrix},$$

wobei $(\xi_1, \xi_2)^\mathsf{T}$ bzw. $(\eta_1, \eta_2)^\mathsf{T}$ der Koordinatenvektor von x bzw. y bzgl. $\mathcal{B}$ ist. Dann repräsentiert $\mathcal{B}$ die durch Φ_0 auf X definierte positive Orientierung (vgl. Abschnitt 9.6).

Es sei nun $x = \xi_1 e_1 + \xi_2 e_2$ gegeben. Dann ist

$$\varphi(x) = \xi_1\varphi(e_1) + \xi_2\varphi(e_2) = \big(\xi_1\cos\theta - \xi_2\sin\theta\big)e_1 + \big(\xi_1\sin\theta + \xi_2\cos\theta\big)e_2.$$

Somit gilt

$$\Phi_0(x, \varphi(x)) = \det\begin{pmatrix} \xi_1 & \xi_1\cos\theta - \xi_2\sin\theta \\ \xi_2 & \xi_1\sin\theta + \xi_2\cos\theta \end{pmatrix} = (\xi_1^2 + \xi_2^2)\sin\theta$$

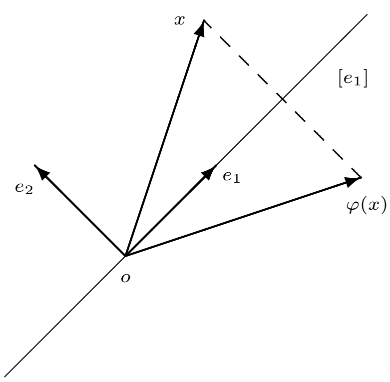
und
$$\langle x, \varphi(x) \rangle = (\xi_1^2 + \xi_2^2) \cos\theta.$$
Ferner ist $\|\varphi(x)\| = \|x\| = (\xi_1^2 + \xi_2^2)^{1/2}$. Es ist daher
$$\frac{\langle x, \varphi(x)\rangle}{\|x\|\,\|\varphi(x)\|} = \cos\theta, \quad \frac{\Phi_0(x, \varphi(x))}{\|x\|\,\|\varphi(x)\|} = \sin\theta.$$
Der orientierte Winkel von x nach $\varphi(x)$ ist somit für alle $x \in X$ durch θ gegeben (vgl. (9.9). Die Abbildung φ kann daher als Drehung um den Winkel θ (im Sinne der durch Φ_0 gegebenen Orientierung von X) interpretiert werden.

Wir geben eine vollständige Diskussion isometrischer Abbildungen in den Fällen $\dim X = 2$ und $\dim X = 3$:

1. $\dim X = 2$.

a) Es sei $\det\varphi = 1$. In diesem Fall besitzt φ ein Paar $\lambda_1, \lambda_2 = \bar\lambda_1$ von Eigenwerten mit $|\lambda_1| = 1$. Es sei (e_1, e_2) eine beliebige Orthonormalbasis von X, welche die positive Orientierung von X repräsentiert. Dann ist
$$\varphi(e_1) = \alpha e_1 + \beta e_2, \quad \alpha^2 + \beta^2 = 1.$$
Aus $\det\varphi = 1$, $\|\varphi(e_2)\|^2 = 1$ und $\langle \varphi(e_1), \varphi(e_2)\rangle = 0$ folgt nach einfacher Rechnung
$$\varphi(e_2) = -\beta e_1 + \alpha e_2.$$
Wir setzen $\alpha = \cos\theta$, $\beta = \sin\theta$ mit $0 \le \theta < 2\pi$. Dann ist
$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$
die Matrix von φ bzgl. (e_1, e_2). Wie wir oben gesehen haben, ist φ eine Drehung um den Winkel θ. Im Falle $\theta = 0$ bzw. $\theta = \pi$ ist $\varphi = \epsilon$ bzw. $\varphi = -\epsilon$.

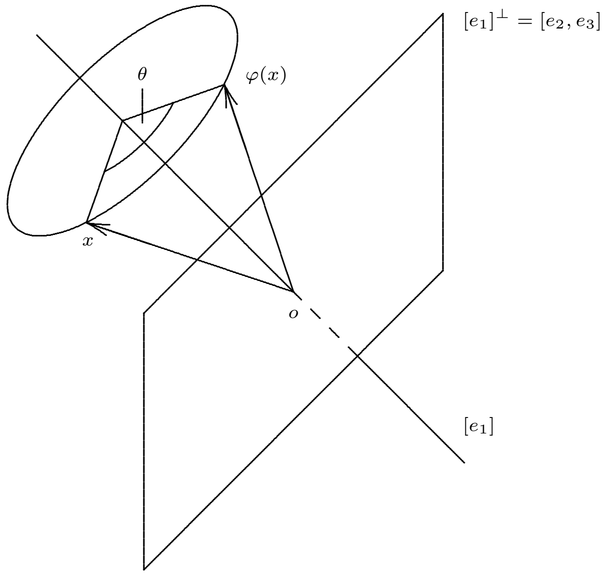
b) Es sei $\det\varphi = -1$. Die Eigenwerte von φ sind durch $\lambda_1 = 1$ und $\lambda_2 = -1$ gegeben. Die zugehörigen Eigenvektoren e_1, e_2 bilden eine Orthonormalbasis von X. An Hand von
$$\varphi(\xi_1 e_1 + \xi_2 e_2) = \xi_1 e_1 - \xi_2 e_2$$
sieht man, dass φ eine Spiegelung an dem eindimensionalen Unterraum $[e_1]$ ist.



2. $\dim X = 3$.

a) Es sei $\det\varphi = 1$. Dann muß $\lambda_1 = 1$ ein Eigenwert von φ sein. Für die übrigen Eigenwerte muß $\lambda_3 = \bar\lambda_2$ gelten. Wegen Satz 14.11 existiert eine Orthonormalbasis $\mathcal{B} = (e_1, e_2, e_3)$ von X, wobei e_1 Eigenvektor von φ zum Eigenwert $\lambda_1 = 1$ ist, sodass
$$U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \quad 0 \le \theta < 2\pi,$$

die Matrix von φ bzgl. $\mathcal{B}$ ist. Die Abbildung φ ist daher eine Drehung um den Winkel θ mit der Drehachse $[e_1]$.



b) Es sei $\det \varphi = -1$. Ein Eigenwert ist $\lambda_1 = -1$, während für die übrigen $\lambda_3 = \bar{\lambda}_2$ mit $|\lambda_2| = 1$ gelten muß. Wie im Fall a) sieht man, dass eine Orthonormalbasis (e_1, e_2, e_3) von X existiert, sodass

$$U = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \quad 0 \le \theta < 2\pi,$$

die Matrix von φ bzgl. dieser Basis ist. Wegen

$$U = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$$

ist φ das Kompositum einer Drehung um den Winkel θ mit der Drehachse $[e_1]$ und einer Spiegelung an dem zweidimensionalen Unterraum $[e_1]^\perp = [e_2, e_3]$.

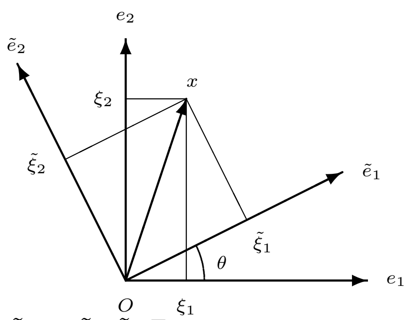
Eine orthogonale Matrix $U \in M_{n,n}(\mathbb{R})$ kann auch (wie jede andere reguläre Matrix) als Matrix einer Basistransformation aufgefaßt werden (siehe Definition 8.1 bzw. Satz 8.2, e)). Wir wollen dies für den Fall $\dim X = 2$ genauer diskutieren. In diesem Fall hat U die Gestalt

$$U = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad 0 \le \theta < 2\pi.$$

Ist (e_1, e_2) eine beliebige Orthonormalbasis von X, so ist die neue Basis durch

$$\begin{aligned} \tilde{e}_1 &= \cos\theta\, e_1 + \sin\theta\, e_2 \\ \tilde{e}_2 &= -\sin\theta\, e_1 + \cos\theta\, e_2 \end{aligned}$$

gegeben, entsteht somit aus (e_1, e_2) durch eine Drehung um den Winkel θ im Sinne der positiven Orientierung von X.



Bezeichnet $\xi = (\xi_1, \xi_2)^\mathsf{T}$ bzw. $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2)^\mathsf{T}$ den Koordinatenvektor von $x \in X$ bzgl. (e_1, e_2) bzw. $(\tilde{e}_1, \tilde{e}_2)$, so gilt $\xi = U\tilde{\xi}$, d.h. $\tilde{\xi} = U^\mathsf{T}\xi$ (vgl. Satz 8.2, b)) oder ausführlicher geschrieben

$$\begin{aligned} \tilde{\xi}_1 &= \cos\theta\xi_1 - \sin\theta\xi_2, \\ \tilde{\xi}_2 &= \sin\theta\xi_1 + \cos\theta\xi_2. \end{aligned}$$

Wie wir gesehen haben, ist eine Matrix $A \in M_{n,n}(\mathbb{C})$ genau dann normal, wenn sie zu einer Diagonalmatrix $D \in M_{n,n}(\mathbb{C})$ unitär-ähnlich ist (siehe Satz 14.5). Es erhebt sich nun die Frage, auf welche Form eine beliebige Matrix aus $M_{n,n}(\mathbb{C})$ durch Ähnlichkeitstransformationen mit unitären Matrizen gebracht werden kann.

Satz 14.12 (Schur[52]). *Es sei $A \in M_{n,n}(\mathbb{C})$ gegeben. Dann existiert eine unitäre Matrix $U \in M_{n,n}(\mathbb{C})$ derart, dass U^*AU eine obere Dreiecksmatrix ist,*

$$U^*AU = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ & \ddots & \ddots & \vdots \\ & 0 & \ddots & * \\ & & & \lambda_n \end{pmatrix},$$

wobei $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A sind. Ist A reell und sind die Eigenwerte $\lambda_1, \ldots, \lambda_n$ reell, so kann $U \in M_{n,n}(\mathbb{R})$ gewählt werden (d.h. U ist orthogonal).

Beweis. Der Beweis erfolgt mittels vollständiger Induktion nach n.

Für $n = 1$ ist die Aussage trivial. Die Aussage sei für $n-1$ richtig. Zu einem Eigenwert λ_0 von A wählen wir einen Eigenvektor y mit $\|y\| = 1$ und ergänzen diesen zu einer Orthonormalbasis $(y, x_2, \ldots, x_n)$ von $\mathbb{C}^n$. Die Matrix $Q := (y, x_2, \ldots, x_n)$ ist unitär. Ferner gilt

$$\begin{aligned} Q^*AQ &= \begin{pmatrix} y^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix} A\,(y, x_2, \ldots, x_n) = \begin{pmatrix} y^* \\ x_2^* \\ \vdots \\ x_n^* \end{pmatrix} (\lambda_0 y, Ax_2, \ldots, Ax_n) \\ &= \left(\begin{array}{c|c} \lambda_0 & \beta \\ \hline 0_{n-1} & A_{n-1} \end{array}\right), \end{aligned}$$

wobei 0_{n-1} den Nullvektor in $\mathbb{C}^n$ bezeichnet und $\beta = (x_2^* A x_2, \ldots, x_n^* A x_n)$,

$$A_{n-1} = \begin{pmatrix} x_2^* \\ \vdots \\ x_n^* \end{pmatrix} A\,(x_2, \ldots, x_n) \in M_{n-1,n-1}(\mathbb{C})$$

[52] Schur, Issai, 10. 1. 1875 (Mogilyov, Weissrussland) – 10. 1. 1941 (Tel Aviv), Darstellung von Gruppen, Kombinatorik, Schurzerlegung einer Matrix (en.wikipedia.org/wiki/Schur).

gesetzt wurde. Laut Induktionsvoraussetzung existiert eine unitäre Matrix $V \in M_{n-1,n-1}(\mathbb{C})$ derart, dass $V^* A_{n-1} V$ eine obere Dreiecksmatrix ist. Wir setzen

$$U := Q \begin{pmatrix} 1 & 0_{n-1}^* \\ 0_{n-1} & V \end{pmatrix}.$$

Dann ist U ebenfalls unitär und es gilt

$$\begin{aligned} U^* A U &= \begin{pmatrix} 1 & 0_{n-1}^* \\ 0_{n-1} & V^* \end{pmatrix} Q^* A Q \begin{pmatrix} 1 & 0_{n-1}^* \\ 0_{n-1} & V \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0_{n-1}^* \\ 0_{n-1} & V^* \end{pmatrix} \begin{pmatrix} \lambda_0 & \beta V \\ 0_{n-1} & A_{n-1} V \end{pmatrix} = \begin{pmatrix} \lambda_0 & \beta V \\ 0_{n-1} & V^* A_{n-1} V \end{pmatrix}. \end{aligned}$$

Die Behauptung über reelle Matrizen folgt unmittelbar aus dem Beweis. □

Beispiel 14.13. Es sei

$$A = \begin{pmatrix} 9 & -1 & -2 \\ 2 & 6 & -2 \\ 0 & 1 & 5 \end{pmatrix}$$

gegeben. Eine einfache Rechnung zeigt, dass $\lambda_1 = 6$, $\lambda_2 = \lambda_3 = 7$ die Eigenwerte von A sind. Für $\lambda_1 = 6$ erhält man

$$(A - \lambda_1 E) = \begin{pmatrix} 3 & -1 & -2 \\ 2 & 0 & -2 \\ 0 & 1 & -1 \end{pmatrix}.$$

Somit ist $y = \frac{1}{\sqrt{3}}(1,1,1)^\mathsf{T}$ ein normierter Eigenvektor von A zum Eigenwert λ_1. Dieser Eigenvektor y kann beispielsweise durch

$$x_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad x_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

zu einer Orthonormalbasis des $\mathbb{R}^3$ ergänzt werden, d.h. es ist

$$Q = \frac{1}{6} \begin{pmatrix} 2\sqrt{3} & 3\sqrt{2} & \sqrt{6} \\ 2\sqrt{3} & -3\sqrt{2} & \sqrt{6} \\ 2\sqrt{3} & 0 & -2\sqrt{6} \end{pmatrix}.$$

Ferner gilt

$$Q^\mathsf{T} A Q = \begin{pmatrix} 6 & \frac{5}{6}\sqrt{6} & \frac{5}{6}\sqrt{18} \\ 0 & 7 & 0 \\ 0 & \frac{4}{3}\sqrt{3} & 7 \end{pmatrix}.$$

Somit müssen wir $A_2 = \begin{pmatrix} 7 & 0 \\ \frac{4}{3}\sqrt{3} & 7 \end{pmatrix}$ auf obere Dreieckform bringen. Dies wird durch $V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ gewährleistet: $V^T A_2 V = \begin{pmatrix} 7 & \frac{4}{3}\sqrt{3} \\ 0 & 7 \end{pmatrix}$. Die gesuchte Matrix U ist somit durch

$$\begin{aligned} U = Q \begin{pmatrix} 1 & 0 \\ 0 & V \end{pmatrix} &= \frac{1}{6} \begin{pmatrix} 2\sqrt{3} & 3\sqrt{2} & \sqrt{6} \\ 2\sqrt{3} & -3\sqrt{2} & \sqrt{6} \\ 2\sqrt{3} & 0 & -2\sqrt{6} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ &= \frac{1}{6} \begin{pmatrix} 2\sqrt{3} & \sqrt{6} & 3\sqrt{2} \\ 2\sqrt{3} & \sqrt{6} & -3\sqrt{2} \\ 2\sqrt{3} & -2\sqrt{6} & 0 \end{pmatrix} \end{aligned}$$

gegeben. Tatsächlich prüft man leicht nach, dass

$$U^\mathsf{T} A U = \begin{pmatrix} 6 & \frac{5}{2}\sqrt{2} & \frac{5}{6}\sqrt{6} \\ 0 & 7 & \frac{4}{3}\sqrt{3} \\ 0 & 0 & 7 \end{pmatrix}$$

gilt. $\Diamond$

14.3 Selbstadjungierte Abbildungen

Auch in diesem Abschnitt bezeichnet $X, Y, \ldots$ stets einen endlich-dimensionalen Vektorraum über $\mathbb{K} = \mathbb{C}$ oder $\mathbb{K} = \mathbb{R}$ mit innerem Produkt.

Eine besonders wichtige Klasse normaler Abbildungen wird in diesem Abschnitt behandelt:

Definition 14.14. a) *Eine lineare Abbildung $\varphi \in L(X, X)$ heißt genau dann **selbstadjungiert**, wenn*

$$\varphi^* = \varphi$$

gilt.

b) *Eine Matrix $A \in M_{n,n}(\mathbb{C})$ heißt genau dann **hermitesch**, wenn*

$$A^* = A$$

*gilt. Ist $A \in M_{n,n}(\mathbb{R})$ (und daher $A^\mathsf{T} = A$), so heißt A **symmetrisch**.*

Aus Satz 13.27 folgt sofort, dass die Matrix einer selbstadjungierten Abbildung bzgl. einer Orthonormalbasis von X hermitesch bzw. symmetrisch ist. Offensichtlich sind selbstadjungierte Abbildungen bzw. hermitesche oder symmetrische Matrizen stets normal.

Satz 14.15. a) *Eine normale Abbildung $\varphi \in L(X, X)$ ist genau dann selbstadjungiert, wenn sie n reelle Eigenwerte besitzt ($n = \dim X$).*

b) *Eine normale Matrix $A \in M_{n,n}(\mathbb{C})$ ist genau dann hermitesch, wenn sämtliche Eigenwerte von A reell sind. Insbesondere ist eine hermitesche Matrix stets zu einer reellen Diagonalmatrix unitär-ähnlich.*

c) *Ist $A \in M_{n,n}(\mathbb{R})$ symmetrisch, so existiert eine orthogonale Matrix $S \in M_{n,n}(\mathbb{R})$ mit*

$$S^\mathsf{T} A S = \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

wobei $\lambda_1, \ldots, \lambda_n$ die (in diesem Falle reellen) Eigenwerte von A sind.

Beweis. a) Es sei zunächst $\mathbb{K} = \mathbb{C}$ und $\varphi \in L(X, X)$ sei normal. Nach Satz 14.5 existiert eine Orthonormalbasis $\mathcal{B}$ von X, sodass die Matrizen von φ bzw. φ^* bzgl. $\mathcal{B}$ durch

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \quad \text{bzw.} \quad D^* = \bar{D} = \operatorname{diag}(\bar{\lambda}_1, \ldots, \bar{\lambda}_n)$$

gegeben. Aus $\varphi^* = \varphi$ folgt $\lambda_i = \bar{\lambda}_i$, $i = 1, \ldots, n$, d.h. alle Eigenwerte von φ sind reell. Sind umgekehrt die Eigenwerte von φ reell, so gilt $D = \bar{D}$ und somit $\varphi^* = \varphi$. Damit ist auch die Aussage b) bewiesen.

Ist $A \in M_{n,n}(\mathbb{R})$ symmetrisch, so ist A als Matrix in $M_{n,n}(\mathbb{C})$ hermitesch. Nach b) besitzt A n reelle Eigenwerte und ist unitär-ähnlich zur reellen Diagonalmatrix $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ mit den Eigenwerten $\lambda_i \in \mathbb{R}$, $i = 1, \ldots, n$. Ist λ_0 ein Eigenwert von A mit der algebraischen Vielfachheit m_0, so folgt wegen Korollar 10.8

$$\operatorname{rang}(A - \lambda_0 E) = n - m_0.$$

Dies bedeutet, dass zu λ_0 genau m_0 linear unabhängige Eigenvektoren von A in $\mathbb{R}^n$ existieren, die wir ohne Einschränkung als orthonormiert annehmen können. Eigenvektoren von A zu verschiedenen Eigenvektoren sind ohnehin orthogonal zueinander (siehe Satz 14.4, b)). Somit existiert eine Orthonormalbasis $\mathcal{B} = (e_1, \ldots, e_n)$ des $\mathbb{R}^n$ bestehend aus Eigenvektoren von A. S sei die Matrix, in deren Spalten die Vektoren $e_i \in \mathbb{R}^n$ stehen. Dann gilt $S^{\mathsf{T}}S = E$ und $S^{\mathsf{T}}AS = D$. □

Beispiel 14.16. Die Matrix

$$A = \begin{pmatrix} 10 & -3+3i & 3-3i \\ -3-3i & 7 & -3 \\ 3+3i & -3 & 7 \end{pmatrix}$$

ist hermitesch mit dem charakteristischen Polynom $(\lambda-4)^2(\lambda-16)$. Eine einfache Rechnung ergibt die orthonormalen Eigenvektoren

$$u_1 = \frac{1}{\sqrt{6}}\big(-(1-i), 0, 2\big)^{\mathsf{T}}, \quad u_2 = \frac{1}{\sqrt{12}}\big(1-i, 3, 1\big)^{\mathsf{T}}$$

zum Eigenwert 4 und den Eigenvektor

$$u_3 = \frac{1}{2}\big(1-i, -1, 1\big)^{\mathsf{T}}$$

zum Eigenwert 16. Die Matrix

$$U = \begin{pmatrix} -\frac{1-i}{\sqrt{6}} & \frac{1-i}{\sqrt{12}} & \frac{1-i}{2} \\ 0 & \frac{3}{\sqrt{12}} & -\frac{1}{2} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{12}} & \frac{1}{2} \end{pmatrix}$$

ist unitär und es gilt

$$U^*AU = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 16 \end{pmatrix}.$$

◊

Wir wollen einen zweiten, von dem oben gegebenen Beweis für Satz 14.15 verschiedenen Beweis geben, der eine wichtige Extremaleigenschaft der Eigenwerte selbstadjungierter Abbildungen bzw. hermitescher oder symmetrischer Matrizen wiederspiegelt.

Es sei $\varphi \in L(X, X)$ gegeben und $\mathcal{B}$ sei eine Orthonormalbasis von X. Mit A bezeichnen wir die Matrix von φ bzgl. $\mathcal{B}$. Für $\xi \in \mathbb{K}^n$, $\xi \neq 0$, definieren wir

$$F(\xi) = \frac{\xi^* A \xi}{\xi^* \xi}.$$

Aus $\overline{\langle x, \varphi(x)\rangle} = \langle \varphi(x), x\rangle = \langle x, \varphi(x)\rangle$ folgt, dass $\langle x, \varphi(x)\rangle$ stets reell ist. Damit ist auch $\xi^* A\xi = \langle x, \varphi(x)\rangle$ reell für alle $\xi \in \mathbb{K}^n$. Dasselbe gilt für $\xi^*\xi$. Somit ist $F(\xi) \in \mathbb{R}$ für $\xi \neq 0$.

Aus $\xi^* A \xi = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} \bar{\xi}_i \alpha_{ij} \xi_i$ und der Tatsache, dass $\xi_i \to \bar{\xi}_i$ stetig ist, folgt, dass $\xi^* A\xi$ eine stetige Funktion von $\xi_1, \ldots, \xi_n$ ist. $\xi^*\xi = |\xi_1|^2 + \cdots + |\xi_n|^2$ ist ebenfalls stetig in $\xi_1, \ldots, \xi_n$. Daher ist $F(\xi)$ als Funktion von $\xi_1, \ldots, \xi_n$ überall dort stetig, wo $\xi^*\xi \neq 0$ ist, d.h. für $\xi \neq 0$.

Die Menge

$$K = \{\xi \in \mathbb{K}^n \mid \|\xi\| = 1\}$$

ist eine beschränkte und abgeschlossene Teilmenge von $\mathbb{K}^n \setminus \{0\}$, d.h. K ist kompakt. Nach einem Satz der Analysis existiert ein $f_1 \in K$ mit

$$F(f_1) = \min_{\xi \in K} F(\xi).$$

Für beliebiges $\xi \in \mathbb{K}^n$, $\xi \neq 0$, ist $\frac{1}{\|\xi\|}\xi \in K$ und $F(\xi) = F\Big(\frac{1}{\|\xi\|}\xi\Big)$. Daraus folgt

$$F(\xi) \geq F(f_1) \ \text{ für alle } \ \xi \in \mathbb{K}^n \setminus \{0\}.$$

Es sei nun η beliebig aus $\mathbb{K}^n$. Wir definieren für $t \in \mathbb{R}$ mit $f_1 + t\eta \neq 0$

$$G(t) = F(f_1 + t\eta).$$

Explizit ist $G(t)$ durch

$$\begin{aligned} G(t) &= \frac{(f_1+t\eta)^* A (f_1+t\eta)}{(f_1+t\eta)^*(f_1+t\eta)} \\ &= \frac{f_1^* A f_1 + t(\eta^* A f_1 + f_1^* A\eta) + t^2\eta^*\eta}{f_1^* f_1 + t(\eta^* f_1 + f_1^*\eta) + t^2\eta^*\eta} \end{aligned}$$

gegeben. Daraus folgt, dass $G(t)$ in einer Umgebung von $t = 0$ differenzierbar ist. Für $t = 0$ hat $G(t)$ das Minimum $G(0) = F(f_1)$ (beachte, dass $F(f_1) \leq F(f_1 + t\eta)$ für alle $t \neq 0$ ist). Daher muß $G'(0) = 0$ sein, d.h.

$$\begin{aligned} 0 = G'(0) &= \frac{(\eta^* A f_1 + f_1^* A\eta) f_1^* f_1 - (\eta^* f_1 + f_1^*\eta) f_1^* A f_1}{f_1^* f_1} \\ &= 2\eta^* A f_1 - 2(\eta^* f_1)(f_1^* A f_1). \end{aligned}$$

Man beachte, dass $f_1^*\eta = \eta^* f_1$, $f_1^* A\eta = \eta^* A^* f_1 = \eta^* A f_1$ und $f_1^* f_1 = 1$ gilt. Aus obiger Gleichung folgt

$$\eta^*\big(Af_1 - (f_1^* A f_1) f_1\big) = 0.$$

Diese Beziehung muß für alle η gelten. Für $\eta = Af_1 - (f_1^* A f_1) f_1$ erhalten wir

$$\|Af_1 - (f_1^* A f_1) f_1\|^2 = 0,$$

d.h.

$$Af_1 = (f_1^* A f_1) f_1.$$

Dies zeigt, dass f_1 ein Eigenvektor von A zum Eigenwert $\lambda_1 = f_1^* A f_1$ ist. λ_1 ist auch ein Eigenwert von φ. Der Vektor a_1 mit dem Koordinatenvektor f_1 bzgl. $\mathcal{B}$ ist der zu λ_1 gehörige Eigenvektor von φ. Es ist $\|a_1\| = 1$. λ_1 ist durch

$$\lambda_1 = \min_{x \neq o} \frac{\langle x, \varphi(x)\rangle}{\|x\|^2}$$

charakterisiert.

Es sei nun $U = [a_1]$. U ist φ-invariant, da $\varphi(a_1) = \lambda_1 a_1$ gilt. Nach Korollar 14.7 ist auch $U^\perp$ φ-invariant. Die Abbildung $\varphi_1 = \varphi \mid U^\perp$ ist somit ein Endomorphismus $U^\perp \to U^\perp$. Für beliebige $x, y \in U^\perp$ gilt

$$\langle \varphi_1(x), y\rangle = \langle \varphi(x), y\rangle = \langle x, \varphi(y)\rangle = \langle x, \varphi_1(y)\rangle,$$

d.h. φ_1 ist selbstadjungiert. Wie oben beweist man, dass ein Eigenwert λ_2 mit zugehörigem Eigenvektor $a_2 \in U^\perp$, $\|a_2\| = 1$, von φ_1 existiert. Aus $\varphi(a_2) = \varphi_1(a_2) = \lambda_2 a_2$ folgt, dass λ_2 Eigenwert von φ mit dem Eigenvektor a_2 ist. Wegen $a_2 \in U^\perp$ gilt $a_1 \perp a_2$. λ_2 ist durch

$$\lambda_2 = \min_{\substack{x \neq o \\ x \perp a_1}} \frac{\langle x, \varphi(x) \rangle}{\|x\|^2}$$

charakterisiert. Der Unterraum $U = [a_1, a_2]$ ist φ-invariant. Man kann daher die obige Konstruktion auf $U^\perp$ durchführen. Nach n Schritten erhält man die Eigenwerte $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ mit zugehörigen Eigenvektoren $a_1, \ldots, a_n$, welche eine Orthonormalbasis von X bilden.

Damit haben wir das folgende Ergebnis bewiesen:

Satz 14.17. *Es sei $\varphi \in L(X, X)$ selbstadjungiert. Dann besitzt φ die reellen Eigenwerte $\lambda_1 \leq \cdots \leq \lambda_n$, die folgendermaßen charakterisiert werden können:*

$$\lambda_1 = \min_{x \neq o} \frac{\langle x, \varphi(x) \rangle}{\|x\|^2},$$
$$\lambda_j = \min_{\substack{x \neq o \\ x \in [a_1, \ldots, a_{j-1}]^\perp}} \frac{\langle x, \varphi(x) \rangle}{\|x\|^2}, \quad j = 2, \ldots, n,$$

wobei a_1 eine Minimumstelle von $\langle x, \varphi(x) \rangle / \|x\|^2$ auf $\{x \in X \mid \|x\| = 1\}$ ist und a_k jeweils eine Minimumstelle auf $\{x \in X \mid \|x\| = 1\} \cap [a_1, \ldots, a_{k-1}]^\perp$ ist, $k = 2, \ldots, n$. Die Vektoren $a_1, \ldots, a_k$ sind Eigenvektoren von φ zu den Eigenwerten $\lambda_1, \ldots, \lambda_n$ und bilden eine Orthonormalbasis von X.

In dem obigen Beweis hätten wir auch $\max_{\xi \in V} F(\xi)$ betrachten können. Dies hätte im ersten Schritt

$$\lambda_n = \max_{x \neq o} \frac{\langle x, \varphi(x) \rangle}{\|x\|^2}$$

geliefert.

Der Quotient $\dfrac{\langle x, \varphi(x) \rangle}{\|x\|^2}$ bzw. $\dfrac{\xi^* A \xi}{\xi^* \xi}$ heißt der ***Rayleigh***[53]***-Quotient*** von φ bzw. A.

Eine spezielle Klasse von selbstadjungierten Abbildungen erhält man in Zusammenhang mit der Minimierungsaufgabe (13.2).

Definition 14.18. *Es sei $\pi \in L(X, X)$ gegeben. π heißt genau dann eine **orthogonale Projektion**, wenn*

$$\pi^* = \pi \quad \textit{und} \quad \pi^2 = \pi \tag{14.5}$$

gilt.

Der Zusammenhang dieser Begriffsbildung mit Abschnitt 13.2 wird durch den folgenden Satz hergestellt:

Satz 14.19. a) *Es sei $\pi \in L(X, X)$ eine orthogonale Projektion und $U := \pi(X)$. Dann ist $x_U := \pi(x)$ für alle $x \in X$ die Lösung der Minimierungsaufgabe* (13.2)*. Ferner gilt* $\ker \pi = U^\perp$.

b) *Es sei der Unterraum U von X gegeben. Die Abbildung $\pi \in L(X, X)$ ist durch $\pi(x) := x_U$, $x \in X$, definiert, wobei x_U die Lösung von Aufgabe* (13.2) *ist. Dann ist π eine orthogonale Projektion mit $U = \pi(X)$.*

[53] John William Strutt, 3rd Baron Rayleigh, 12. 11. 1842 (Langford Grove, England) – 30. 6. 1919 (Terlins Place bei Witham, England), seit 1873 Baron Rayleigh, Nobelpreis für Physik 1904, mit William Ramsay Entdecker des Edelgases Argon (en.wikipedia.org/wiki/John_Strutt%2C_3rd_Baron_Rayleigh).

Beweis. a) Wir wählen $x \in X$. Für beliebige $y \in U$ gilt wegen (14.5)

$$\begin{aligned}\langle x - \pi(x), y\rangle &= \langle x - \pi(x), \pi(y)\rangle = \langle \pi(x) - \pi^2(x), y\rangle \\ &= \langle \pi(x) - \pi(x), y\rangle = 0.\end{aligned}$$

Man beachte, dass $y = \pi(z)$ mit einem $z \in X$ und daher $\pi(y) = \pi^2(z) = \pi(z) = y$ gilt. Nach Satz 13.12 ist $\pi(x)$ die Lösung von (13.2). Für $x \in \ker \pi$ und $y \in \pi(X)$ gilt $y = \pi(z)$ mit einem $z \in X$ und $\langle x, y\rangle = \langle x, \pi(z)\rangle = \langle \pi(x), z\rangle = \langle o, z\rangle = 0$, d.h. $\ker \pi \perp U := \pi(X)$. Da auch $X = \ker \pi \oplus \pi(X)$ gilt (vgl. Korollar 14.3), muß $\ker \pi = U^\perp$ sein.

b) Wir setzen $U := \pi(X)$. Ein Vektor $x \in X$ ist genau dann in $\ker \pi$, wenn $x_U = o$ gilt. Wegen Satz 13.13 bedeutet dies $\langle x, y\rangle = 0$ für alle $y \in U$. Damit ist $\ker \pi = U^\perp$ gezeigt. Dass $X = \ker \pi \oplus \pi(X)$ gilt, ist wegen Korollar 14.3 klar.

Für $x \in X$ ist $x = \pi(x) + (x - \pi(x))$. Wegen der Definition von $\pi(x)$ gilt $\pi(x) \in U$ und (wegen Satz 13.13) $x - \pi(x) \in U^\perp$. Daher erhalten wir für beliebige $x, y \in X$

$$\begin{aligned}\langle x, \pi(y)\rangle &= \langle \pi(x) + (x - \pi(x)), \pi(y)\rangle = \langle \pi(x), \pi(y)\rangle + \langle x - \pi(x), \pi(y)\rangle \\ &= \langle \pi(x), \pi(y)\rangle = \langle \pi(x), \pi(y) - y + y\rangle \\ &= \langle \pi(x), y\rangle + \langle \pi(x), \pi(y) - y\rangle = \langle \pi(x), y\rangle.\end{aligned}$$

Daraus folgt $\pi^* = \pi$.

Für $x \in X$ ist $\pi(x) = x_U$ die Lösung von (13.2) und $\pi(x_U) = \pi^2(x)$ die Lösung von (13.2) mit x_U an Stelle von x. Daher gilt (wieder wegen Satz 13.13)

$$0 = \langle x_U - \pi(x_U), y\rangle \quad \text{für alle } y \in U.$$

Daraus folgt (für $y = x_U - \pi(x_U) \in U$) $x_U = \pi(x_U)$ für alle $x \in X$, d.h. $\pi^2 = \pi$. □

14.4 Positiv definite Matrizen

Eine besonders wichtige Rolle spielen selbstadjungierte Abbildungen, deren Eigenwerte sämtliche positiv bzw. nicht-negativ sind. Im folgenden beschränken wir uns auf Matrizen und überlassen es dem Leser, die Resultate für lineare Abbildungen zu formulieren.

Definition 14.20. *Es sei $A \in M_{n,n}(\mathbb{C})$ hermitesch.*

a) *A heißt genau dann **positiv definit**, $A > 0$ (bzw. **positiv semidefinit**, $A \geq 0$), wenn sämtliche Eigenwerte von A positiv (bzw. nicht-negativ) sind.*

b) *A heißt genau dann **negativ definit** (bzw. **negativ semidefinit**), wenn $-A$ positiv definit (bzw. positiv semidefinit) ist.*

c) *A heißt genau dann **indefinit**, wenn A sowohl positive als auch negative Eigenwerte besitzt.*

Wie für nicht-negative Zahlen existiert auch für positiv semidefinite Matrizen eine „Quadratwurzel“:

Satz 14.21. *Die Matrix $A \in M_{n,n}(\mathbb{C})$ sei positiv definit (bzw. positiv semidefinit). Dann existiert eine positiv definite (bzw. positiv semidefinite) Matrix H mit*

$$H^2 = A \quad \textit{und} \quad \operatorname{rang} A = \operatorname{rang} H.$$

*H heißt die **Quadratwurzel** von A, $H = A^{1/2}$.*

Beweis. $A \geq 0$ bedeutet $0 \leq \lambda_1 \leq \cdots \leq \lambda_n$ für die Eigenwerte λ_i von A. Wir setzen

$$H_0 := \operatorname{diag}\big(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2}\big).$$

Da A hermitesch ist, existiert eine unitäre Matrix U mit $A = UDU^*$, wobei

$$D = \operatorname{diag}\bigl(\lambda_1, \ldots, \lambda_n\bigr)$$

gilt. Daher ist $D = H_0^2$.

Setzen wir $H := UH_0U^*$, so erhalten wir

$$H^2 = UH_0U^*UH_0U^* = UH_0^2U^* = UDU^* = A.$$

Offensichtlich ist H hermitesch. Klarerweise sind alle Eigenwerte von H positiv (bzw. nichtnegativ) falls dies für die Eigenwerte von A gilt. □

Eine Charakterisierung positiv definiter bzw. positiv semidefiniter Matrizen erfolgt über die mit einer hermiteschen Matrix assoziierten quadratischen Form:

Satz 14.22. a) *Eine hermitesche Matrix A ist genau dann positiv definit (bzw. positiv semidefinit), wenn $\xi^* A\xi > 0$ für alle $\xi \neq 0$ (bzw. $\xi^* A\xi \geq 0$ für alle ξ) gilt.*

b) *Eine hermitesche Matrix A ist genau dann indefinit, wenn es Vektoren $\xi_1, \xi_2 \in \mathbb{C}^n$ gibt mit $\xi_1^* A\xi_1 < 0 < \xi_2^* A\xi_2$.*

Beweis. a) Es sei $\xi^* A\xi \geq 0$ für alle $\xi \neq 0$. $\lambda_1, \ldots, \lambda_n$ seien die Eigenwerte von A mit dazugehörigen orthonormalen Eigenvektoren $u_1, \ldots, u_n$. Dann gilt

$$0 \leq u_i^* Au_i = u_i^*(\lambda_i u_i) = \lambda_i u_i^* u_i = \lambda_i,$$

für alle $i = 1, \ldots, n$, d.h. A ist positiv semidefinit. Ist $\xi^* A\xi > 0$ für alle $\xi \neq 0$, so folgt $\lambda_i > 0$, $i = 1, \ldots, n$, d.h. A ist positiv definit.

Ist umgekehrt A positiv semidefinit, so existiert $H = A^{1/2}$ und ist ebenfalls positiv semidefinit. Daher gilt

$$\xi^* A\xi = \xi^*(HH)\xi = \xi^* H^* H\xi = (H\xi)^*(H\xi) \geq 0 \quad \text{für alle } \xi \in \mathbb{C}^n.$$

Ist A positiv definit, so ist auch H positiv definit, d.h. insbesondere $\operatorname{rang} H = n$. Somit ist $H\xi = 0$ nur möglich für $\xi = 0$. □

14.5 Die Singulärwertzerlegung einer Matrix

Für jede Matrix $A \in M_{m,n}(\mathbb{C})$ sind die Matrizen A^*A und AA^* offensichtlich positiv semidefinit. Wir zeigen zunächst, dass ihre positiven Eigenwerte übereinstimmen:

Satz 14.23. *Es sei $A \in M_{m,n}(\mathbb{C})$ gegeben. Dann stimmen die positiven Eigenwerte von A^*A und AA^* (bzw. von $(A^*A)^{1/2}$ und $(AA^*)^{1/2}$) und deren algebraische Vielfachheit überein.*

Beweis. Es sei $\lambda > 0$ ein Eigenwert von A^*A mit der algebraischen Vielfachheit k und $u_1, \ldots, u_k$ seien orthonormierte Eigenvektoren von A^*A zum Eigenwert λ (beachte, dass A^*A diagonalisierbar und daher k auch die Dimension von $\ker(A^*A - \lambda E)$ ist). Dann gilt

$$(AA^*)Au_i = A(A^*Au_i) = \lambda Au_i, \quad i = 1, \ldots, k. \tag{14.6}$$

Ferner ist

$$u_j^* A^* Au_i = (Au_j)^*(Au_i), \quad i, j = 1, \ldots, k,$$

und andererseits

$$u_j^* A^* Au_i = u_j^*(\lambda u_i) = \lambda u_j^* u_i = \begin{cases} \lambda & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases}$$

Daraus folgt

$$(Au_j)^*(Au_i) = \begin{cases} \lambda & \text{für } i = j, \\ 0 & \text{für } i \neq j, \end{cases}$$

d.h. es ist $Au_i \neq 0$, $i = 1, \ldots, k$ und die Vektoren $Au_1, \ldots, Au_k$ sind orthogonal. Wegen (14.6) ist daher λ auch ein Eigenwert von AA^* mit den orthogonalen Eigenvektoren $Au_1, \ldots, Au_k$. Die algebraische Vielfachheit von λ als Eigenwert von AA^* ist daher mindestens k. Durch Vertauschung der Rollen von A und A^* folgt die Behauptung. □

Definition 14.24. *Es sei die Matrix $A \in M_{m,n}(\mathbb{C})$ gegeben. Die Eigenwerte $\sigma_1, \ldots, \sigma_n$ der Matrix $(A^*A)^{1/2} \in M_{n,n}(\mathbb{C})$ heißen die **singulären Werte** von A.*

Lemma 14.25. *Für jede Matrix $A \in M_{m,n}(\mathbb{C})$ gilt*

$$\operatorname{rang} A = \operatorname{rang} A^* = \operatorname{rang} A^*A = \operatorname{rang} AA^*.$$

Beweis. Es ist nur $\operatorname{rang} A = \operatorname{rang} A^*A$ zu beweisen. Mit Hilfe von Satz 5.12 folgt

$$\operatorname{rang} A = n - \operatorname{def} A = \operatorname{rang} A^*A + \operatorname{def} A^*A - \operatorname{def} A.$$

Offensichtlich ist $\ker A \subset \ker A^*A$. Ist umgekehrt $x \in \ker A^*A$, d.h. $A^*Ax = 0$, so folgt $0 = x^*A^*Ax = (Ax)^*Ax = \|Ax\|^2$ und daher $Ax = 0$. Damit ist auch $\ker A^*A \subset \ker A$ gezeigt. Es gilt somit $\operatorname{def} A^*A = \operatorname{def} A$. □

Aus Lemma 14.25 und Satz 14.23 folgt:

Satz 14.26. *Es sei $A \in M_{m,n}(\mathbb{C})$ mit $\operatorname{rang} A = r \leq \min(m, n)$ gegeben. Dann besitzen A und A^* genau dieselben r positiven singulären Werte $\sigma_1, \ldots, \sigma_r$, während die übrigen singulären Werte jeweils Null sind.*

Satz 14.27. *Es seien $A \in M_{n,n}(\mathbb{C})$ und die unitäre Matrix $U \in M_{n,n}(\mathbb{C})$ gegeben. Dann besitzen die Matrizen*

$$A, \ AU, \ UA \text{ und } U^*AU$$

dieselben singulären Werte $\sigma_1, \ldots, \sigma_n$. Insbesondere besitzen unitär-ähnliche Matrizen dieselben singulären Werte.

Beweis. Aus

$$(AU)(AU)^* = AUU^*A^* = AA^*$$

folgt sofort, dass A und AU dieselben singulären Werte haben. Wegen Satz 14.26 gilt dies auch für $A, A^*, A^*U^* = (UA)^*$ und UA. □

Im Falle normaler Matrizen besteht ein direkter Zusammenhang zwischen Eigenwerten und singulären Werten:

Satz 14.28. *Es sei $A \in M_{n,n}(\mathbb{C})$ normal mit den Eigenwerten $\lambda_1, \ldots, \lambda_n$. Dann sind die singulären Werte von A durch*

$$\sigma_i = |\lambda_i|, \quad i = 1, \ldots, n,$$

gegeben.

Beweis. Nach Satz 14.5 existiert eine unitäre Matrix $U \in M_{n,n}(\mathbb{C})$ mit $U^*AU = D := \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Die Matrizen A und D haben dieselben singulären Werte (Satz 14.27). Für die Matrix D ist das Ergebnis wegen $(D^*D)^{1/2} = \operatorname{diag}(|\lambda_1|, \dots, |\lambda_n|)$ trivial. □

Beispiel 14.29. A sei die Matrix aus Beispiel 12.15 mit den Eigenwerten

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 2.$$

Eine einfache Rechnung ergibt

$$A^*A = \begin{pmatrix} 3 & -1 & -2 & -3 \\ -1 & 7 & 2 & 5 \\ -2 & 2 & 4 & 2 \\ -3 & 5 & 2 & 11 \end{pmatrix}$$

Die Eigenwerte von A^*A sind

$$\mu_1 = 16,\ \mu_2 = \mu_3 = 4,\ \mu_4 = 1.$$

Die singulären Werte von A sind somit durch

$$\sigma_1 = 4,\ \sigma_2 = \sigma_3 = 2,\ \sigma_4 = 1$$

gegeben. Dieses Beispiel zeigt, dass die Aussage von Satz 14.28 für beliebige Matrizen in $M_{n,n}(\mathbb{C})$ im allgemeinen nicht gilt. ◊

Für beliebige $n \times n$-Matrizen gilt jedoch:

Satz 14.30. *Es sei $A \in M_{n,n}(\mathbb{C})$ gegeben und $\sigma_1, \dots, \sigma_n$ seien die singulären Werte von A. Dann gilt*

$$|\det A| = \sigma_1 \cdot \dots \cdot \sigma_n.$$

Beweis. Die Aussage folgt sofort aus

$$\sigma_1^2 \cdot \dots \cdot \sigma_n^2 = \det A^*A = \det A^* \det A = |\det A|^2.$$

□

Die singulären Werte einer Matrix spielen eine fundamentale Rolle bei einer speziellen Faktorisierung von Matrizen, der sogenannten Singulärwertzerlegung:

Satz 14.31 (Singulärwertzerlegung). *Es sei eine Matrix $A \in M_{m,n}(\mathbb{C})$ mit $\operatorname{rang} A = r$ $(\leq \min(m,n))$ gegeben. Ferner seien $\lambda_1 \geq \dots \geq \lambda_r > 0$ und $\lambda_{r+1} = \dots = \lambda_n = 0$ die Eigenwerte von A^*A (beachte $\operatorname{rang} A^*A = \operatorname{rang} A = r$). Dann gilt:*

a) *Es sei*

$$U^*AV = \Sigma = \big(\sigma_{ij}\big)_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \in M_{m,n}(\mathbb{R}) \tag{14.7}$$

mit unitären Matrizen $U \in M_{m,m}(\mathbb{C})$, $V \in M_{n,n}(\mathbb{C})$, wobei

$$\begin{aligned} &\sigma_{ij} = 0 \text{ für } i \neq j,\ i = 1, \dots, m,\ j = 1, \dots, n, \\ &\sigma_{ii} \geq 0, \quad i = 1, \dots, k := \min(m,n) \end{aligned} \tag{14.8}$$

gilt. Dann ist

$$\sigma_{ii} = 0, \quad i = r+1, \dots, k, \quad \text{und} \quad \sigma_{ii} = \lambda_i^{1/2}, \quad i = 1, \dots, r. \tag{14.9}$$

b) *Es sei* $(v_1, \dots, v_n)$ *eine Orthonormalbasis des* $\mathbb{C}^n$ *bestehend aus Eigenvektoren von* A^*A, *wobei* v_i *Eigenvektor zum Eigenwert* λ_i *ist,* $i = 1, \dots, n$. *Ferner sei*

$$u_i = \frac{1}{\lambda_i^{1/2}} A v_i, \quad i = 1, \dots, r.$$

Dann sind die Vektoren $u_1, \dots, u_r$ *orthonormiert. Die Vektoren* $u_{r+1}, \dots, u_m$ *seien so gewählt, dass* $u_1, \dots, u_m$ *eine Orthonormalbasis des* $\mathbb{C}^m$ *bilden. Mit den unitären Matrizen* $U = (u_1, \dots, u_m)$, $V = (v_1, \dots, v_n)$ *gilt*

$$U^* A V = \Sigma,$$

wobei für die Matrix $\Sigma \in M_{m,n}(\mathbb{R})$ *(*14.8*) und (*14.9*) gelten. Ferner ist*

$$\operatorname{bild} A = [u_1, \dots, u_r], \quad \ker A = [v_{r+1}, \dots, v_n].$$

Beweis. a) Aus $A = U \Sigma V^*$ folgt

$$A^* A = V \Sigma^{\mathsf{T}} U^* U \Sigma V^* = V \Sigma^{\mathsf{T}} \Sigma V^*.$$

Offensichtlich ist

$$\Sigma^{\mathsf{T}} \Sigma = \operatorname{diag}\bigl(\sigma_{11}^2, \dots, \sigma_{kk}^2, 0, \dots, 0\bigr).$$

Da die Eigenwerte von A^*A und $\Sigma^{\mathsf{T}}\Sigma$ übereinstimmen, gilt (14.9).
b) Wir beweisen zunächst, dass die Vektoren $u_1, \dots, u_r$ orthonormiert sind. Es gilt

$$\begin{aligned} u_i^* u_j &= \frac{1}{(\lambda_i \lambda_j)^{1/2}} (A v_i)^* A v_j = \frac{1}{(\lambda_i \lambda_j)^{1/2}} v_i^* A^* A v_j \\ &= \frac{1}{(\lambda_i \lambda_j)^{1/2}} \lambda_j v_i^* v_j = \Bigl(\frac{\lambda_j}{\lambda_i}\Bigr)^{1/2} v_i^* v_j = \begin{cases} 0 & \text{für } i \neq j, \\ 1 & \text{für } i = j. \end{cases} \end{aligned}$$

Wegen der Definition der Vektoren $u_1, \dots, u_r$ ist

$$A v_j = \lambda_j^{1/2} u_j, \quad j = 1, \dots, r. \tag{14.10}$$

Da $\ker A = \ker A^*A$ gilt (vgl. Hilfssatz 14.25), folgt aus $A^*Av_j = 0$, $j = r+1, \dots, n$, auch

$$A v_j = 0, \quad j = r+1, \dots, n. \tag{14.11}$$

Aus (14.10) und (14.11) folgt auch

$$\operatorname{bild} A = [u_1, \dots, u_r], \quad \ker A = [v_{r+1}, \dots, v_n]. \tag{14.12}$$

Schließlich ist

$$\begin{aligned} U^* A V &= U^* \bigl(\lambda_1^{1/2} u_1, \dots, \lambda_r^{1/2} u_r, 0, \dots, 0\bigr) \\ &= \begin{pmatrix} \lambda_1^{1/2} & & & \\ & \ddots & & 0 \\ & & \lambda_r^{1/2} & \\ & 0 & & \end{pmatrix} \in M_{m,n}(\mathbb{R}). \end{aligned}$$

□

Jede Darstellung von $A \in M_{m,n}(\mathbb{C})$ der Form

$$A = U \Sigma V^*$$

mit unitären Matrizen $U \in M_{m,m}(\mathbb{C})$ und $V \in M_{n,n}(\mathbb{C})$ und einer reellen Matrix $\Sigma \in M_{m,n}(\mathbb{R})$, welche (14.8) erfüllt, heißt eine ***Singulärwertzerlegung*** von A. Nach dem eben bewiesenen Satz existiert stets eine Singulärwertzerlegung von A. Ferner ist die Matrix Σ eindeutig bestimmt, während dies für die Matrizen U und V im allgemeinen nicht gilt.

Ist $A = U\Sigma V^*$ eine Singulärwertzerlegung von A, so ist offensichtlich $A^* = V\Sigma^{\mathsf{T}}U^*$ eine Singulärwertzerlegung von A^*.

Beispiel 14.32. Es sei

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

Dann ist

$$A^T A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad AA^T = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Die Eigenwerte von $A^T A$ sind durch $\lambda_1 = 2$, $\lambda_2 = 1$ gegeben. Zugehörige Eigenvektoren sind $v_1 = (1,0)^{\mathsf{T}}$ und $v_2 = (0,1)^{\mathsf{T}}$. Die Eigenwerte von AA^T müssen $\lambda_1 = 2$, $\lambda_2 = 1$ und $\lambda_3 = 0$ sein (Satz 14.23). Zugehörige Eigenvektoren sind

$$u_1 = \frac{1}{\sqrt{2}} A v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad u_2 = A v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Die Singulärwertzerlegung von A ist somit durch

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

gegeben. $\Diamond$

14.6 Die Pseudoinverse einer Matrix

Die Singulärwertzerlegung einer Matrix eröffnet einen systematischen Weg zur Lösung der folgenden Minimierungsaufgabe:

$(\mathcal{M}_{A,\beta})$ Gegeben seien eine Matrix $A \in M_{m,n}(\mathbb{C})$ und ein Vektor $\beta \in \mathbb{C}^m$. Gesucht sind Vektoren $\xi \in \mathbb{C}^n$ mit

$$\|A\xi - \beta\| = \inf_{\eta \in \mathbb{C}^n} \|A\eta - \beta\|.$$

Ist das lineare Gleichungssystem $(\mathcal{S}_{A,\beta})$ lösbar, so stimmen offensichtlich die Lösungsmengen von $(\mathcal{S}_{A,\beta})$ und $(\mathcal{M}_{A,\beta})$ überein.

Es sei nun $A = U\Sigma V^*$ eine Singulärwertzerlegung von A und $\sigma_1, \dots, \sigma_r$ seien die positiven singulären Werte von A, $r = \operatorname{rang} A$. Ist ξ eine Lösung von $(\mathcal{M}_{A,\beta})$, so gilt (beachte, dass U und V unitär sind)

$$\begin{aligned} \|A\xi - \beta\| &= \|U^*(A\xi - \beta)\| = \|\Sigma V^*\xi - U^*\beta\| \\ &\le \inf_{\eta \in \mathbb{C}^n} \|U^*(A\eta - \beta)\| = \inf_{\eta \in \mathbb{C}^n} \|\Sigma V^*\eta - U^*\beta\| \\ &= \inf_{\tilde{\eta} \in \mathbb{C}^n} \|\Sigma\tilde{\eta} - U^*\beta\|. \end{aligned}$$

Dies bedeutet, dass $\zeta := V^*\xi$ eine Lösung der Minimierungsaufgabe $(\mathcal{M}_{\Sigma,U^*\beta})$ ist. Umgekehrt gilt für jede Lösung ζ von $(\mathcal{M}_{\Sigma,U^*\beta})$

$$\begin{aligned}\|\Sigma\zeta - U^*\beta\| &= \|U^*(U\Sigma V^*V\zeta - \beta\| = \|AV\zeta - \beta\| \\ &\le \inf_{\tilde\eta\in\mathbb{C}^n} \|\Sigma\tilde\eta - U^*\beta\| = \inf_{\tilde\eta\in\mathbb{C}^n} \|AV\tilde\eta - \beta\| \\ &= \inf_{\eta\in\mathbb{C}^n} \|A\eta - \beta\|,\end{aligned}$$

d.h. $\xi := V\zeta$ ist eine Lösung von $(\mathcal{M}_{A,\beta})$.

Die Lösungen von $(\mathcal{M}_{\Sigma,U^*\beta})$ können einfach berechnet werden. Wir setzen $U^*\beta = \delta = (\delta_1, \ldots, \delta_m)^\mathsf{T}$ und $\tilde\eta = (\tilde\eta_1, \ldots, \tilde\eta_n)^\mathsf{T}$. Dann ist

$$\begin{aligned}\|\Sigma\tilde\eta - \delta\|^2 = |\sigma_1\tilde\eta_1 - \delta_1|^2 + \cdots + |\sigma_r\tilde\eta_r|^2 \\ + |\delta_{r+1}|^2 + \cdots + |\delta_m|^2.\end{aligned}$$

Sämtliche Lösungen $\zeta = (\zeta_1, \ldots, \zeta_n)^\mathsf{T}$ von $(\mathcal{M}_{\Sigma,U^*\beta})$ sind daher durch

$$\zeta_i = \frac{\delta_i}{\sigma_i}, \quad i = 1, \ldots, r, \text{ und } \zeta_i \text{ beliebig aus } \mathbb{C},\ i = r+1, \ldots, n,$$

gegeben. Beachten wir noch, dass durch $\xi = V\zeta$ die Lösungen von $(\mathcal{M}_{A,\beta})$ gegeben sind, so erhalten wir:

Satz 14.33. *Es sei $A = U\Sigma V^*$ eine Singulärwertzerlegung von A und $\sigma_1, \ldots, \sigma_r$ seien die positiven singulären Werte von A, $r = \operatorname{rang} A$. Dann sind sämtliche Lösungen ξ von $(\mathcal{M}_{A,\beta})$ durch*

$$\xi = \sum_{i=1}^{r} \frac{\delta_i}{\sigma_i} v_i + \sum_{i=r+1}^{n} \zeta_i v_i, \quad \zeta_i \in \mathbb{C},\ i = r+1, \ldots, n, \tag{14.13}$$

gegeben, wobei $v_1, \ldots, v_n$ die Spalten von V sind und $U^*\beta = (\delta_1, \ldots, \delta_m)^\mathsf{T}$ ist.

Für $r < n$ besitzt $(\mathcal{M}_{A,\beta})$ eine $(n-r)$-parametrige Schar von Lösungen. In diesem Fall ist man sehr oft an Lösungen mit minimaler Norm interessiert.

Definition 14.34. *Es sei das Problem $(\mathcal{M}_{A,\beta})$ vorgelegt. Eine Lösung $\xi^\#$ von $(\mathcal{M}_{A,\beta})$ heißt genau dann eine **Pseudonormallösung** von $(\mathcal{M}_{A,\beta})$, wenn*

$$\|\xi^\#\| \le \|\xi\|$$

für alle Lösungen ξ von $(\mathcal{M}_{A,\beta})$ gilt.

Satz 14.35. *Es sei das Problem $(\mathcal{M}_{A,\beta})$ vorgelegt und es seien die Voraussetzungen aus Satz 14.33 erfüllt. Dann ist*

$$\xi^\# = \sum_{i=1}^{r} \frac{\delta_i}{\sigma_i} v_i = V\Sigma^\# U^*\beta$$

die eindeutig bestimmte Pseudonormallösung von $(\mathcal{M}_{A,\beta})$, wobei die Matrix $\Sigma^\# = (\sigma_{ij}^\#) \in M_{n,m}(\mathbb{R})$ durch

$$\sigma_{ij}^\# = \begin{cases} \frac{1}{\sigma}i & \text{für } i = j \text{ und } i = 1, \ldots, r, \\ 0 & \text{sonst,} \end{cases}$$

gegeben ist.

Beweis. Jede Lösung ξ von $(\mathcal{M}_{A,\beta})$ ist durch (14.13) gegeben. Daraus folgt wegen der Orthonormiertheit der Vektoren $v_1, \ldots, v_n$

$$\|\xi\|^2 = \|\xi^\#\| + \sum_{i=r+1}^{n} |\zeta_i|^2 \geq \|\xi^\#\|^2,$$

d.h. $\xi^\#$ ist eine Pseudonormallösung. $\xi^\#$ ist eindeutig bestimmt, da für jede von $\xi^\#$ verschiedene Lösung von $(\mathcal{M}_{A,\beta})$ mindestens eine der Zahlen $\zeta_{r+1}, \ldots, \zeta_n$ von Null verschieden sein müßte.

Die Darstellung $\xi^\# = V\Sigma^\# U^*\beta$ folgt durch einfaches Verifizieren.

Da bei fix gegebener Matrix $A \in M_{m,n}(\mathbb{C})$ die Lösung $\xi^\#$ der Aufgabe $(\mathcal{M}_{A,\beta})$ durch β eindeutig bestimmt ist, muß die Matrix $V\Sigma^\# U^*$ unabhängig von der speziellen Singulärwertzerlegung von A sein. Man beachte, dass wohl Σ und infolgedessen auch $\Sigma^\#$ jedoch im allgemeinen nicht U und V eindeutig durch A bestimmt sind. □

Definition 14.36. *Die durch $A \in M_{m,n}(\mathbb{C})$ eindeutig bestimmte Matrix*

$$A^\# := V\Sigma^\# U^* \in M_{n,m}(\mathbb{C})$$

*heißt die **Pseudoinverse** (oder auch **Moore**[54]**-Penrose**[55]**-Inverse**) von A.*

Ist $A \in M_{n,n}(\mathbb{C})$ regulär, so sind die singulären Werte $\sigma_1, \ldots, \sigma_n$ von A sämtliche positiv. Daher ist in jeder Singulärwertzerlegung $A = U\Sigma V^*$ von A die Matrix Σ regulär. Dies hat $\Sigma^\# = \Sigma^{-1}$ und

$$A^\# = V\Sigma^{-1}U^* = \big(U\Sigma V^*\big)^{-1} = A^{-1}$$

zur Folge.

Die Eindeutigkeit der Pseudoinversen einer Matrix kann auch ohne Bezugnahme auf das Minimierungsproblem $(\mathcal{M}_{A,\beta})$ bewiesen werden:

Satz 14.37. *Es sei $A \in M_{m,n}(\mathbb{C})$ gegeben.*

a) *Die Pseudoinverse $A^\#$ ist die einzige Matrix aus $M_{n,m}(\mathbb{C})$, welche die folgenden Gleichungen erfüllt:*

$$\begin{aligned} &\text{(i)} && AA^\# = (AA^\#)^*,\\ &\text{(ii)} && A^\# A = (A^\# A)^*,\\ &\text{(iii)} && AA^\# A = A,\\ &\text{(iv)} && A^\# AA^\# = A^\#. \end{aligned}$$

b) *Es gilt*

$$(A^*)^\# = (A^\#)^* \quad \textit{und} \quad (A^\#)^\# = A.$$

c) *Es gilt*

$$A^\# = (A^*A)^\# A^* = A^*(AA^*)^\#.$$

Ist insbesondere eine der Matrizen $A^*A \in M_{n,n}(\mathbb{C})$ oder $AA^* \in M_{m,m}(\mathbb{C})$ regulär, so ist $A^\# = (A^*A)^{-1}A^*$ bzw. $A^\# = A^*(AA^*)^{-1}$.

[54] Moore, Eliakim Hastings, 26. 1. 1862 (Marietta, Ohio) – 30. 12. 1932 (Chicago), Beiträge zur abstrakten Algebra, zur Geometrie und den Grundlagen der Analysis (en.wikipedia.org/wiki/E._H._Moore).

[55] Penrose Roger, 8. 8. 1931, Mathematische Physik (insbesondere Allgemeine Relativitätstheorie und Kosmologie) entdeckte 1955 als Student unabhängig von E. H. Moore die sog. Moore-Penrose Inverse (en.wikipedia.org/wiki/Roger_Penrose).

Beweis. Es sei $\operatorname{rang} A = r$ und $A = U\Sigma V^*$ eine Singulärwertzerlegung von A mit den positiven singulären Werten $\sigma_1, \ldots, \sigma_r$.
a) Aus $A^\# = V\Sigma^\# U^*$ folgt

$$AA^\# = U\Sigma V^* V\Sigma^\# U^* = U\Sigma\Sigma^\# U^*.$$

Offensichtlich ist $\Sigma\Sigma^\# = \operatorname{diag}(1, \ldots, 1, 0, \ldots, 0) \in M_{m,m}(\mathbb{R})$ mit r Einsen in der Hauptdiagonale und daher $(\Sigma\Sigma^\#)^* = \Sigma\Sigma^\#$. Es gilt somit

$$AA^\# = U(\Sigma\Sigma^\#)^* U^* = (U\Sigma\Sigma^\# U^*)^* = (AA^\#)^*.$$

Der Beweis für (ii) verläuft analog.

Wegen $\Sigma\Sigma^\#\Sigma = \Sigma$ und $\Sigma^\#\Sigma\Sigma^\# = \Sigma^\#$ folgen in analoger Weise die Beziehungen (iii) und (iv).

Es sei nun $B \in M_{n,m}(\mathbb{C})$ eine Matrix, für die ebenfalls die Beziehungen (i) – (iv) gelten (mit B an Stelle von $A^\#$). Unter Verwendung von (i) – (iv) für $A^\#$ und B folgt:

$$\begin{aligned}
B &\overset{\text{(iv)}}{=} B\underline{AB} \overset{\text{(i)}}{=} B(AB)^* = BB^*\underline{A^*} \overset{\text{(iii)}}{=} BB^*A^*(A^\#)^*A^* \\
&= BB^*A^*\underline{(AA^\#)^*} \overset{\text{(i)}}{=} BB^*A^*AA^\# = B\underline{(AB)^*}AA^\# \overset{\text{(i)}}{=} \underline{BAB}AA^\# \\
&\overset{\text{(iv)}}{=} BA\underline{A^\#} \overset{\text{(iv)}}{=} BA\underline{A^\#A}A^\# \overset{\text{(ii)}}{=} BA(A^\#A)^*A^\# = \underline{BA}A^*(A^\#)^*A^\# \\
&\overset{\text{(ii)}}{=} (BA)^*A^*(A^\#)^*A^\# = \underline{A^*B^*A^*}(A^\#)^*A^\# \overset{\text{(iii)}}{=} A^*(A^\#)^*A^\# \\
&= (A^\#A)^*A^\# \overset{\text{(ii)}}{=} A^\#AA^\# \overset{\text{(iv)}}{=} A^\#.
\end{aligned}$$

b) Aus $AA^\# = \big((A^\#)^*A^*\big)^*$ und $(AA^\#)^* = (A^\#)^*A^*$ folgt wegen (i) die Beziehung $(A^\#)^*A^* = \big((A^\#)^*A^*\big)^*$, das ist (ii) für A^* und $(A^\#)^*$. Analog sieht man, dass auch (i) für A^* und $(A^\#)^*$ gilt. Durch Bildung der konjugiert-transponierten Matrizen in (iii) und (iv) sieht man, dass diese Beziehungen auch für A^* und $(A^\#)^*$ gelten. Daraus folgt $(A^\#)^* = (A^*)^\#$.

Da (i) – (iv) invariant gegenüber einer Vertauschung von A und $A^\#$ sind, folgt $(A^\#)^\# = A$.
c) Es sei $A = U\Sigma V^*$ eine Singulärwertzerlegung von A. Daraus erhält man die folgende Singulärwertzerlegung von A^*A:

$$A^*A = V\Sigma^{\mathsf{T}}U^*U\Sigma V^* = V\Sigma^{\mathsf{T}}\Sigma V^*.$$

Somit ist $(A^*A)^\# = V(\Sigma^{\mathsf{T}}\Sigma)^\# V^*$ und

$$(A^*A)^\# A^* = V(\Sigma^{\mathsf{T}}\Sigma)^\# V^*V\Sigma^{\mathsf{T}}U^* = V(\Sigma^{\mathsf{T}}\Sigma)^\#\Sigma^{\mathsf{T}}U^*.$$

Eine einfache Rechnung zeigt $(\Sigma^{\mathsf{T}}\Sigma)^\#\Sigma^{\mathsf{T}} = \Sigma^\#$, woraus

$$(A^*A)^\# A^* = A^\#$$

folgt. Der Beweis für $A^\# = A^*(AA^*)^\#$ ist analog. □

Durch die Aussage c) des eben bewiesenen Satzes wird die Berechnung von Pseudoinversen für beliebige Matrizen $A \in M_{m,n}(\mathbb{C})$ auf die Berechnung von Pseudoinversen für positiv semidefinite Matrizen der Dimension m bzw. n zurückgeführt.

Abschließend wollen wir noch eine geometrische Interpretation der Pseudoinversen geben. Aus (14.10) und (14.12) folgt sofort, dass

$$A\,|_{(\ker A)^\perp} \colon (\ker A)^\perp \to \operatorname{bild} A$$

bijektiv ist. Bezeichnet e_i, $i = 1, \ldots, m$, die Vektoren der kanonischen Basis des $\mathbb{C}^m$ und $\tilde{e}_i$, $i = 1, \ldots, n$, die Vektoren der kanonischen Basis des $\mathbb{C}^n$, so gilt

$$\begin{aligned} A^\# u_j &= V\Sigma^\# U^* u_j = V\Sigma^\# e_j = \frac{1}{\sigma_j} V\tilde{e}_j \\ &= \frac{1}{\sigma_j} v_j, \quad j = 1, \ldots, r. \end{aligned}$$

Dies beweist, dass

$$A^\# \mid_{\operatorname{bild} A} : \operatorname{bild} A \to (\ker A)^\perp$$

ebenfalls bijektiv ist. Wegen $AA^\# u_j = A(1/\sigma_j)v_j = u_j$, $j = 1, \ldots, r$, (siehe (14.10)) ist

$$A^\# \mid_{\operatorname{bild} A} = \left(A \mid_{(\ker A)^\perp}\right)^{-1}.$$

Offensichtlich ist

$$A^\# u_j = V\Sigma^\# e_j = V0 = 0, \quad j = r+1, \ldots, m,$$

d.h. $A^\# \mid_{(\operatorname{bild} A)^\perp}$ ist die Nullabbildung. Das folgende Diagramm illustriert die Situation:

$$\begin{tikzcd}
\mathbb{C}^n & & \mathbb{C}^m \\
\| & & \| \\
(\ker A)^\perp \arrow[rr, shift left, "A"] & & \operatorname{bild} A \arrow[ll, shift left, "A^\#"] \\
\oplus & & \oplus \\
\ker A \arrow[rrd, "A"] & & (\operatorname{bild} A)^\perp \arrow[lld, "A^\#"'] \\
\{0\} & & \{0\}
\end{tikzcd}$$

Wir wollen noch anmerken, dass $AA^\#$ eine Projektion des $\mathbb{C}^m$ mit $\operatorname{bild} AA^\# = \operatorname{bild} A$ und $A^\# A$ eine Projektion des $\mathbb{C}^n$ mit $\operatorname{bild} A^\# A = (\ker A)^\perp$ ist. Denn aus Satz 14.37, a), folgt

$$AA^\# = (AA^\#)^* \quad \text{und} \quad (AA^\#)^2 = (AA^\# A)A^\# = AA^\#.$$

Analoges gilt für $A^\# A$.

Kapitel 15

Kurven zweiter Ordnung

$(\mathcal{P}, X)$ sei ein euklidischer Raum der Dimension 2 (vgl. Abschnitt 3.1). Das Koordinatensystem (siehe Definition 3.2) $(O; e_1, e_2)$ sei so gewählt, dass $\mathcal{B} = (e_1, e_2)$ eine Orthonormalbasis von X ist. Die Punktmenge $\mathcal{K}$ sei durch

$$\mathcal{K} = \{P \in \mathcal{P} \mid \alpha\xi_1^2 + 2\beta\xi_1\xi_2 + \gamma\xi_2^2 + \delta_1\xi_1 + \delta_2\xi_2 + \eta = 0\}$$

gegeben, wobei $\alpha, \beta, \gamma, \delta_1, \delta_2, \eta$ gegebene reelle Zahlen sind und ξ_1, ξ_2 die kartesischen Koordinaten des Punktes P bzgl. des Koordinatensystems $(O; e_1, e_2)$. Wir nennen $\mathcal{K}$ eine ***Kurve zweiter Ordnung***.

Wir bezeichnen den Ortsvektor von P mit x, $x = \overrightarrow{O\,P}$, und definieren die Matrix A bzw. den Vektor $d \in \mathbb{R}^2$ durch

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}, \qquad d = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}.$$

Damit erhalten wir

$$\mathcal{K} = \{P \in \mathcal{P} \mid \text{Für } x = \overrightarrow{O\,P} \text{ gilt } x^\mathsf{T} A x + d^\mathsf{T} x + \eta = 0\}.$$

Wir nennen

$$x^\mathsf{T} A x + d^\mathsf{T} x + \eta = 0 \tag{15.1}$$

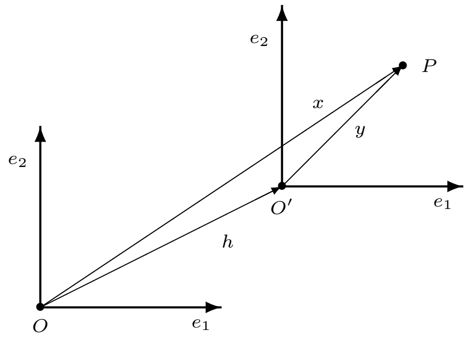
die Gleichung der Kurve zweiter Ordnung.

Unser Ziel ist es, die Gleichung (15.1) durch Einführung eines neuen Koordinatensystems weitestgehend zu vereinfachen. Dabei wird wesentlich sein, dass die reelle Matrix A symmetrisch ist. Wir können also die Resultate des vorhergehenden Abschnittes über hermitesche Matrizen verwenden.

Fall 1: $\det A \neq 0$.
Wir versuchen, durch Wahl eines neuen Koordinatenursprunges O' den Term $d^\mathsf{T} x$ in Gleichung (15.1) zum Verschwinden zu bringen.

Es seien $(O; e_1, e_2)$ und $(O'; e_1, e_2)$ zwei Koordinatensysteme. h sei der Ortsvektor von O' bzgl. des ersten Koordinatensystems, $h = \overrightarrow{O\,O'}$, und x bzw. y sei der Ortsvektor eines beliebigen Punktes P bzgl. des ersten bzw. zweiten Koordinatensystems. Dann gilt

$$x = y + h.$$



Aus Gleichung (15.1) für die Ortsvektoren der Punkte von $\mathcal{K}$ bzgl. $(O; e_1, e_2)$ folgt

$$\begin{aligned} 0 &= x^\mathsf{T} Ax + d^\mathsf{T} x + \eta = (y+h)^\mathsf{T} A(y+h) + d^\mathsf{T}(y+h) + \eta \\ &= y^\mathsf{T} Ay + (2h^\mathsf{T} A + d^\mathsf{T})y + h^\mathsf{T} Ah + d^\mathsf{T} h + \eta. \end{aligned}$$

In dieser Gleichung für die Ortsvektoren der Punkte von $\mathcal{K}$ bzgl. $(O'; e_1, e_2)$ können wir

$$2h^\mathsf{T} A + d^\mathsf{T} = 0 \quad \text{bzw.} \quad 2Ah + d = 0$$

fordern. Diese Gleichung ist wegen $\det A \neq 0$ stets lösbar:

$$h = -\frac{1}{2} A^{-1} d.$$

Mit diesem h erhalten wir die Gleichung

$$y^\mathsf{T} Ay + \tilde{\eta} = 0, \tag{15.2}$$

wobei $\tilde{\eta} = h^\mathsf{T} Ah + d^\mathsf{T} h + \eta = \eta - \frac{1}{4} d^\mathsf{T} A^{-1} d$.

Wegen $\det A \neq 0$ sind die Eigenwerte λ_1, λ_2 von A von Null verschieden. Nach Satz 14.15, c), existiert eine orthogonale Matrix S derart, dass $S^\mathsf{T} AS = \operatorname{diag}(\lambda_1, \lambda_2)$ gilt. Man erhält S, indem man orthonormierte Eigenvektoren a_1, a_2 von A zu λ_1 und λ_2 bestimmt und $S = (a_1, a_2)$ setzt. Wir können a_1, a_2 so wählen, dass $\det S = 1$ gilt (ist $\det S = -1$, so ersetze man etwa a_2 durch $-a_2$).

Durch S ist eine Basis (e_1', e_2') von X derart bestimmt, dass S dem Paar (e_1, e_2) und (e_1', e_2') von Basen zugeordnet ist (im Sinne von Definition 8.1). Der Koordinatenvektor von e_i' bzgl. (e_1, e_2) ist a_i. Ist z der Koordinatenvektor von $\overrightarrow{O'P}$ bzgl. (e_1', e_2') und y der Koordinatenvektor von $\overrightarrow{O'P}$ bzgl. (e_1, e_2), so gilt (siehe Satz 8.2, b))

$$y = Sz.$$

Setzt man in (15.2) ein, so erhalten wir

$$z^\mathsf{T} S^\mathsf{T} ASz + \tilde{\eta} = 0,$$

d.h. wegen $S^\mathsf{T} AS = \operatorname{diag}(\lambda_1, \lambda_2)$

$$\lambda_1 z_1^2 + \lambda_2 z_2^2 + \tilde{\eta} = 0. \tag{15.3}$$

Gleichung (15.3) ist die Gleichung für die Ortsvektoren der Punkte von $\mathcal{K}$ bzgl. $(O'; e_1', e_2')$. Bevor wir eine genauere Diskussion von (15.3) durchführen, wollen wir an die geometrische Bedeutung des Basiswechsels $(e_1, e_2) \to (e_1', e_2')$ erinnern (siehe Abschnitt 14.2):

Die Matrix S hat die Form

$$S = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad 0 \leq \theta < 2\pi,$$

und beschreibt eine Drehung des Koordinatensystems um den Winkel θ gegen den Uhrzeigersinn.

Diskussion von Gleichung (15.3)***:***
Wir setzen $\rho = -\tilde{\eta}$ und erhalten

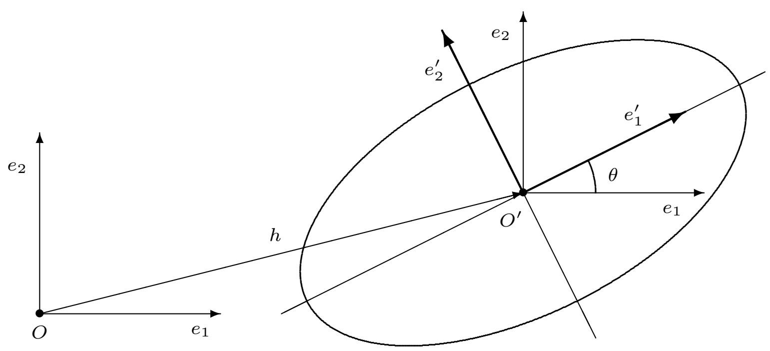
$$\lambda_1 z_1^2 + \lambda_2 z_2^2 = \rho. \tag{15.4}$$

Für das folgende beachte man $\lambda_1 \neq 0$ und $\lambda_2 \neq 0$.

Fall 1.1: $\operatorname{sign}\lambda_1 = \operatorname{sign}\lambda_2$.
a) $\operatorname{sign}\rho \neq \operatorname{sign}\lambda_1$.
Gleichung (15.4) ist für reelle z_1, z_2 nicht möglich. $\mathcal{K}$ hat keine Punkte.
b) $\rho = 0$.
Die einzige Lösung von (15.4) ist $z_1 = z_2 = 0$, d.h. $\mathcal{K}$ besteht nur aus einem Punkt, nämlich O'.
c) $\operatorname{sign}\rho = \operatorname{sign}\lambda_1$.
Wir können annehmen, dass λ_1, λ_2 und ρ positiv sind. Gleichung (15.4) kann man zu

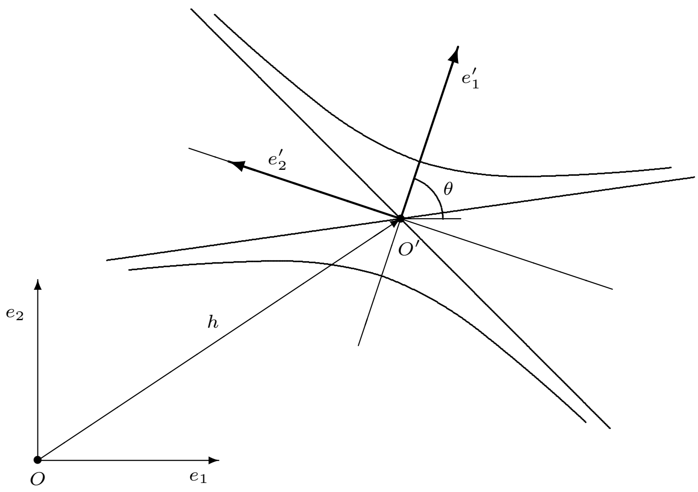
$$\frac{\lambda_1}{\rho} z_1^2 + \frac{\lambda_2}{\rho} z_2^2 = 1.$$

umformen. Daraus erkennt man, dass $\mathcal{K}$ eine Ellipse mit den Halbachsen $(\rho/\lambda_1)^{1/2}$ bzw. $(\rho/\lambda_2)^{1/2}$ ist. Mittelpunkt der Ellipse ist O'. Die Richtungen der Halbachsen werden durch die Basisvektoren e_1', e_2' gegeben, deren Koordinatenvektoren bzgl. (e_1, e_2) durch a_1, a_2 gegeben sind.



Fall 1.2: $\operatorname{sign}\lambda_1 \neq \operatorname{sign}\lambda_2$.
Ohne Einschränkung können wir $\lambda_1 > 0 > \lambda_2$ annehmen.
a) $\rho \neq 0$.
An Hand von (15.4) sieht man sofort, dass eine Hyperbel vorliegt. Mittelpunkt der Hyperbel

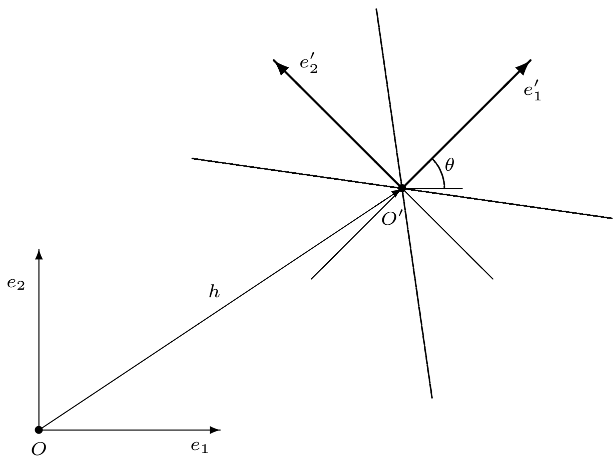
ist O'.



b) $\rho = 0$.
Gleichung (15.4) kann zu

$$\bigl(\sqrt{|\lambda_1|}z_1 - \sqrt{|\lambda_2|}z_2\bigr)\bigl(\sqrt{|\lambda_1|}z_1 + \sqrt{|\lambda_2|}z_2\bigr) = 0$$

umgeformt werden. Es handelt sich um zwei sich in O' schneidende Gerade.



Fall 2: $\det A = 0$.
Hier können wir keine Verschiebung vornehmen. Ein trivialer Fall ist $A = 0$, d.h. (15.1) hat die Form $d^{\mathsf{T}}x + \eta = 0$. Es liegt eine Gerade vor.

Für das weitere sei $A \neq 0$. Ohne Einschränkung können wir für die Eigenwerte von A

$$\lambda_1 \neq 0 \quad \text{und} \quad \lambda_2 = 0$$

annehmen.

Wir bestimmen wie im Fall 1 die Matrix S so, dass $S^{\mathsf{T}}S = E$, $\det S = 1$ und $S^{\mathsf{T}}AS = \operatorname{diag}(\lambda_1, 0)$ gilt. Als neues Koordinatensystem wählen wir $(O; e_1', e_2')$. Für die Ortsvektoren y der Punkte von $\mathcal{K}$ bzgl. des neuen Koordinatensystems erhalten wir aus (15.1) ($x = Sy$)

$$\lambda_1 y_1^2 + \alpha_1 y_1 + \alpha_2 y_2 + \eta = 0,$$

wobei $d^\mathsf{T} S = (\alpha_1, \alpha_2)$ gesetzt wurde. Diese Gleichung können wir zu

$$\begin{aligned}\lambda_1 \left(y_1^2 \frac{\alpha_1}{\lambda_1} y_1 + \left(\frac{\alpha_1}{2\lambda_1}\right)^2 \right) - \lambda_1 \frac{\alpha_1^2}{4\lambda_1^2} + \alpha_2 y_2 + \eta \\ = \lambda_1 \Big(y_1 + \frac{\alpha_1}{2\lambda_1} \Big)^2 + \alpha_2 y_2 + \eta - \frac{\alpha_1^2}{4\lambda_1} = 0\end{aligned}$$

umformen. Wir unterscheiden die folgenden Fälle:

Fall 2.1: $\alpha_2 = 0$. Wir erhalten eine Gleichung der Form

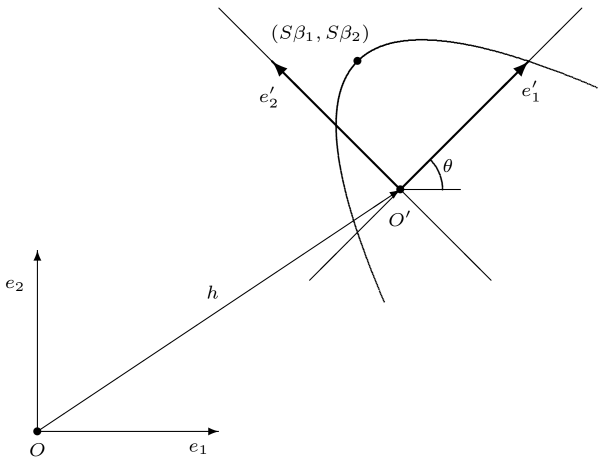
$$(y_1 - \beta_1)^2 = \rho.$$

Für $\rho \geq 0$ besteht $\mathcal{K}$ aus zwei parallen Geraden, die auch zusammenfallen können ($\rho = 0$). Im Falle $\rho < 0$ hat $\mathcal{K}$ keine Punkte.

Fall 2.2: $\alpha_2 \neq 0$.
Wir erhalten eine Gleichung der Form

$$(y_1 - \beta_1)^2 = \kappa(y_2 - \beta_2), \quad \kappa \neq 0.$$

$\mathcal{K}$ ist eine Parabel. Die Koordinaten des Scheitelpunktes bzgl. $(O; e_1', e_2')$ sind β_1, β_2.



Beispiel 15.1. Es sei die Gleichung

$$\xi_1^2 - 2\xi_1\xi_2 + \xi_2^2 - \xi_1 + \xi_2 - 2 = 0$$

gegeben. Es ist

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad d = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \eta = -2.$$

Wegen $\det A = 0$ liegt der Fall 2 vor. Es ist $\lambda_1 = 2$, $\lambda_2 = 0$. Die Matrix S ist durch

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

gegeben. Aus $\cos\theta = \frac{1}{\sqrt{2}}$ und $\sin\theta = -\frac{1}{\sqrt{2}}$ folgt $\theta = \frac{7\pi}{4}$.

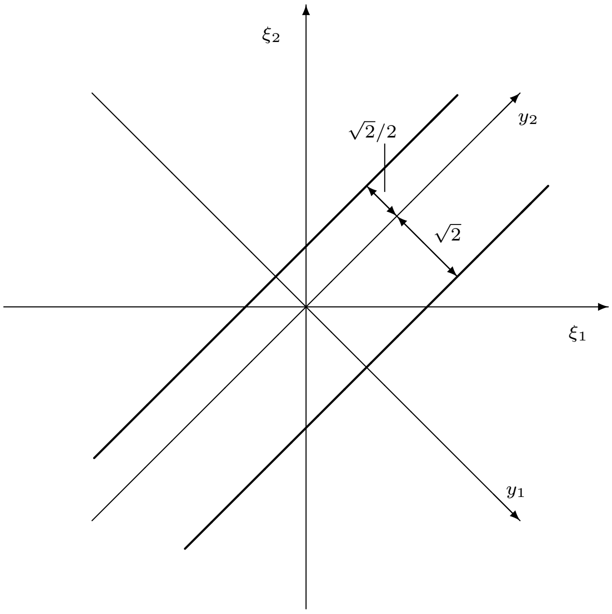
Durch die Transformation $x = Sy$ erhält man die Gleichung

$$2y_1^2 - \frac{2}{\sqrt{2}} y_1 - 2 = 0 \quad \text{bzw.} \quad \Big(y_1 - \frac{1}{2\sqrt{2}} \Big)^2 = 1 + \frac{1}{8} = \frac{9}{8}.$$

Es liegt der Fall 2.1 vor. Wir erhalten zwei parallele Gerade, deren Gleichungen im gedrehten Koordinatensystem

$$y_1 = \frac{2}{\sqrt{2}} = \sqrt{2} \quad \text{und} \quad y_1 = -\frac{1}{\sqrt{2}} = -\frac{\sqrt{2}}{2}$$

sind. $\Diamond$



Beispiel 15.2. Es sei die Gleichung

$$\xi_1^2 + 2\xi_1\xi_2 + \xi_2^2 - 2\xi_1 + \xi_2 - 4 = 0$$

gegeben. Hier ist

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad d = \begin{pmatrix} -2 \\ 1 \end{pmatrix}, \quad \eta = -4.$$

Ferner ist $\det A = 0$ und $\lambda_1 = 2$, $\lambda_2 = 0$. Die mit Hilfe der orthonormierten Eigenvektoren von A gebildete Matrix S ist

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

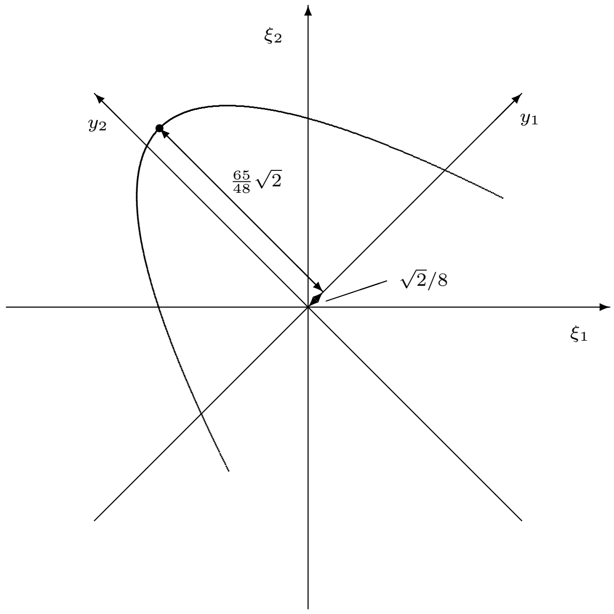
Daraus folgt $\cos\theta = \frac{1}{\sqrt{2}}$, $\sin\theta = \frac{1}{\sqrt{2}}$, d.h. $\theta = \frac{\pi}{4}$. Mit Hilfe der Koordinatentransformation $x = Sy$ erhalten wir die Gleichung

$$2y_1^2 - \frac{1}{\sqrt{2}}y_1 + \frac{3}{\sqrt{2}}y_2 - 4 = 0$$

bzw. nach einigen Umformungen

$$\Big(y_1 - \frac{\sqrt{2}}{8}\Big)^2 = -\frac{3\sqrt{2}}{4}\Big(y_2 - \frac{65\sqrt{2}}{48}\Big).$$

Es liegt eine Parabel vor. $\Diamond$



Beispiel 15.3. Es sei die Gleichung

$$11\xi_1^2 + 6\xi_1\xi_2 + 19\xi_2^2 - 80 = 0$$

gegeben. Es ist

$$A = \begin{pmatrix} 11 & 3 \\ 3 & 19 \end{pmatrix}, \quad d = 0, \quad \eta = -80.$$

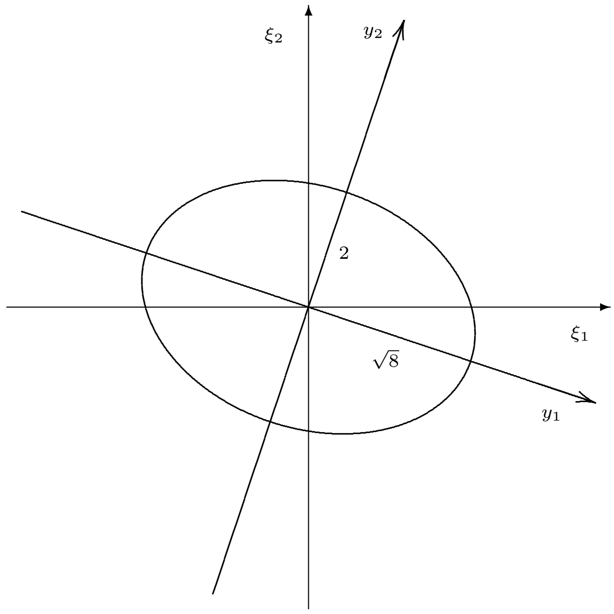
Die Eigenwerte von A sind $\lambda_1 = 10$, $\lambda_2 = 20$ mit den zugehörigen Eigenvektoren $a_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ -1 \end{pmatrix}$, $a_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 3 \end{pmatrix}$. Daher ist

$$S = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix}.$$

Durch die Transformation $x = Sy$ erhalten wir

$$10y_1^2 + 20y_2^2 = 80.$$

Es liegt eine Ellipse mit den Halbachsen $\sqrt{8}$ und 2 vor. Aus $\cos\theta = \frac{3}{\sqrt{10}}$ und $\sin\theta = -\frac{1}{\sqrt{10}}$ kann man den Drehwinkel berechnen, $\theta = -18.435°$ bzw. $\theta = 341.565°$. $\diamondsuit$



ξ_2
y_2
2
ξ_1
$\sqrt{8}$
y_1

Anhang A

Einiges über Polynome

In diesem Anhang stellen wir die wesentlichsten in den vorangehenden Kapiteln benötigten Resultate über Polynome zusammen.

Definition A.1. *Es sei R eine Menge mit zwei inneren Verknüpfungen, die wir mit „$+$“ und „$\cdot$“ bezeichnen.* $(R, +, \cdot)$ *heißt genau dann ein* ***kommutativer Ring****, wenn gilt:*

(i) $(R, +)$ *ist eine abelsche Gruppe.*

(ii) *Die innere Verknüpfung „$\cdot$“ ist assoziativ und kommutativ, d.h.* $(ab)c = a(bc)$ *und* $ab = ba$ *für alle* $a, b, c \in R$.

(iii) $(a + b)c = ac + bc$ *für alle* $a, b, c \in R$.

Ein Beispiel eines kommutativen Ringes ist etwa $(\mathbb{Z}, +, \cdot)$. Natürlich ist auch jeder Körper ein kommutativer Ring.

Definition A.2. *Es sei $\mathbb{K}$ ein Körper und K^* ein kommutativer Ring mit $K^* \supset \mathbb{K}$.*
a) *Ein Element $t \in K^*$ heißt* ***Unbestimmte über*** $\mathbb{K}$*, wenn gilt:*

(i) $1 \cdot t = t$.

(ii) *Für beliebiges $k \in \mathbb{N}$ und beliebige $\alpha_0, \ldots, \alpha_k \in \mathbb{K}$ folgt aus $\alpha_0 + \alpha_1 t + \cdots + \alpha_k t^k = 0$ stets $\alpha_0 = \alpha_1 = \cdots = \alpha_k = 0$.*

b) *Ein Element $f \in K^*$ von der Form*

$$f = \alpha_0 + \alpha_1 t + \cdots + \alpha_k t^k \quad \text{mit } \alpha_\kappa \in \mathbb{K},\ \kappa = 0, \ldots, k,$$

heißt ein Polynom *in t. Die α_j heißen die* ***Koeffizienten*** *von f. Die Menge aller Polynome wird mit $\mathbb{K}[t]$ bezeichnet und bildet einen kommutativen Ring, den* ***Polynomring*** *in einer Unbestimmten über $\mathbb{K}$.*

Es ist klar, dass $\mathbb{K} \subset \mathbb{K}[t]$ gilt. Die Forderung (ii) an t hat zur Folge, dass zwei Polynome $f = \alpha_0 + \alpha_1 t + \ldots$ und $g = \beta_0 + \beta_1 t + \ldots$ dann und nur dann gleich sind, wenn $\alpha_i = \beta_i$, $i = 0, 1, \ldots$ gilt.

Es sei $f = \alpha_0 + \alpha_1 t + \cdots + \alpha_k t^k \in \mathbb{K}[t]$ vorgelegt. Ist $\alpha_k \neq 0$, so heißt k der ***Grad von*** f, $k = \operatorname{grad} f$. Für das Nullpolynom $f \equiv 0$ definieren wir $\operatorname{grad} f = -\infty$. Gilt $\operatorname{grad} f = k \geq 0$ und ist $\alpha_k = 1$, so heißt f ***normiert***.[56]

Für den Grad eines Polynoms gelten folgende einfache Regeln:

[56] im Englischen ist dafür der Ausdruck „monic polynomial“ üblich.

Für Polynome f, g aus $\mathbb{K}[t]$ gilt:

$$\operatorname{grad}(f+g) \leq \max(\operatorname{grad} f, \operatorname{grad} g), \tag{A.1}$$
$$\operatorname{grad} fg = \operatorname{grad} f + \operatorname{grad} g. \tag{A.2}$$

In diesen Formeln ist $-\infty < n$, $-\infty + n = n + (-\infty) = -\infty$ für $n = 0, 1, 2, \ldots$ zu setzen.

Beweis. Es ist $f + g = \alpha_0 + \beta_0 + (\alpha_1 + \beta_1)t + \ldots$. Wegen $\alpha_i = 0$ und $\beta_i = 0$ für $i > k = \max(\operatorname{grad} f, \operatorname{grad} g)$ ist $\alpha_i + \beta_i = 0$ für $i > k$, d.h. $\operatorname{grad}(f+g) \leq k$.

Es sei $\operatorname{grad} f = m$ und $\operatorname{grad} g = n$. Dann ist $fg = \alpha_0\beta_0 + \cdots + \alpha_m\beta_m t^{m+n}$. Wegen $\alpha_m \neq 0$, $\beta_n \neq 0$ ist $\operatorname{grad} fg = m+n$. Formel (A.2) ist trivial, wenn f oder g das Nullpolynom ist. □

Eine unmittelbare Folgerung von (A.2) ist:

$$\textit{Es seien } f, g \in \mathbb{K}[t]. \textit{ Aus } fg = 0 \textit{ folgt } f = 0 \textit{ oder } g = 0. \tag{A.3}$$

Beweis. Es sei $fg = 0$, d.h. $\operatorname{grad} fg = -\infty = \operatorname{grad} f + \operatorname{grad} g$ wegen (A.2). Dann muß aber $\operatorname{grad} f = -\infty$ oder $\operatorname{grad} g = -\infty$ sein. □

Auf Grund der Aussage (A.3) gilt die ***Kürzungsregel*** für Polynome:

Aus $fg = fh$, $f \neq 0$, folgt $g = h$.

Es sei $f \in \mathbb{K}[t]$, $f = \alpha_0 + \alpha_1 t + \cdots + \alpha_k t^k$. Ersetzt man die Unbestimmte t durch $\lambda \in \mathbb{K}$, so erhalten wir das Element $f(\lambda) = \alpha_0 + \alpha_1\lambda + \cdots + \alpha_k\lambda^k \in \mathbb{K}$. Es seien $f, g, h \in \mathbb{K}[t]$. Dann gelten die folgenden Ersetzungsregeln:

Aus $f + g = h$ bzw. $fg = h$ folgt

$$f(\lambda) + g(\lambda) = h(\lambda) \quad \textit{bzw.} \quad f(\lambda)g(\lambda) = h(\lambda)$$

für alle $\lambda \in \mathbb{K}$. Insbesondere folgt für $f, g \in \mathbb{K}[t]$ aus $f = g$ stets

$$f(\lambda) = g(\lambda) \quad \textit{für alle } \lambda \in \mathbb{K}.$$

Letztere Aussage ist im allgemeinen nicht umkehrbar. Es sei $\mathbb{K}$ der Körper mit den zwei Elementen $0, 1$. Für die Polynome $f = t^2 + 1$ und $g = t + 1$ gilt $f(0) = g(0) = 1$, $f(1) = g(1) = 0$. Es ist jedoch $f \neq g$.

Lemma A.3. *Es seien $f \in \mathbb{K}[t]$ und $\lambda_0 \in \mathbb{K}$ gegeben. λ_0 heißt genau dann **Nullstelle** von f, wenn $f(\lambda_0) = 0$ gilt. Ist $\operatorname{grad} f = n > -\infty$ und λ_0 Nullstelle von f, so existiert genau ein Polynom $g \in \mathbb{K}[t]$ und eine natürliche Zahl k, $1 \leq k \leq n$, mit*

$$f = (t - \lambda_0)^k g \quad \textit{und} \quad g(\lambda_0) \neq 0.$$

*k heißt die **Vielfachheit der Nullstelle** λ_0.*

Beweis. Es sei $f = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n$. Durch Nachrechnen bestätigt man leicht die folgende Identität:

$$\begin{aligned} f = (t - \lambda_0)\Big(&\alpha_n t^{n-1} + (\alpha_{n-1} + \lambda_0\alpha_n)t^{n-2} \\ &+ (\alpha_{n-2} + \lambda_0\alpha_{n-1} + \lambda_0^2\alpha_n)t^{n-3} + \cdots + (\alpha_1 + \lambda_0\alpha_2 + \cdots + \lambda_0^{n-1}\alpha_n)\Big) \\ &+ \alpha_0 + \lambda_0\alpha_1 + \cdots + \alpha_n\lambda_0^n =: (t - \lambda_0)g + f(\lambda_0). \end{aligned}$$

Für das Polynom g gilt $\operatorname{grad} g = n - 1$. Wegen $f(\lambda_0) = 0$ erhalten wir

$$f = (t - \lambda_0)g.$$

Ist $g(\lambda_0) \neq 0$, so ist $k = 1$. Im Falle $g(\lambda_0) = 0$ erhalten wir analog $g = (t - \lambda_0)g_1$, d.h. $f = (t - \lambda_0)^2 g_1$. Im Falle $g_1(\lambda_0) \neq 0$ ist $k = 2$. Nach endlich vielen Schritten erhalten wir die angegebene Darstellung. □

Satz A.4. *Es sei* $f \in \mathbb{K}[t]$ *mit* $\operatorname{grad} f = n$. $\lambda_1, \ldots, \lambda_s \in \mathbb{K}$ *seien die paarweise verschiedenen Nullstellen von* f, k_i *sei die Vielfachheit von* λ_i. *Dann gilt:*

$$s \leq k_1 + \cdots + k_s \leq n.$$

Beweis. Wegen $k_i \geq 1$ ist die linke Ungleichung trivial. Da λ_1 k_1-fache Nullstelle ist, gilt:

$$f = (t - \lambda_1)^{k_1} g_1, \; g_1(\lambda_1) \neq 0.$$

Aus $0 = f(\lambda_2) = (\lambda_2 - \lambda_1)^{k_1} g_1(\lambda_2)$ und $\lambda_1 \neq \lambda_2$ folgt $g_1(\lambda_2) = 0$. Daraus folgt

$$f = (t - \lambda_1)^{k_1}(t - \lambda_2)^{k_2} g_2(\lambda), \; g_2(\lambda_1) \neq 0, \; g_2(\lambda_2) \neq 0.$$

Nach endlich vielen Schritten erhalten wir die Darstellung

$$f = (t - \lambda_1)^{k_1} \cdot \cdots \cdot (t - \lambda_s)^{k_s} g_s,$$

wobei $g_s \neq 0$ ist. Nach Formel (A.2) folgt

$$\operatorname{grad} f = k_1 + \cdots + k_s + \operatorname{grad} g \geq k_1 + \cdots + k_s.$$

□

Ist der Körper $\mathbb{K}$ algebraisch abgeschlossen (d.h. jedes Polynom aus $\mathbb{K}[t]$ vom Grad ≥ 1 besitzt mindestens eine Nullstelle in $\mathbb{K}$), so kann die Aussage von Satz A.4 verschärft werden zu:

$$k_1 + \cdots + k_s = n,$$

d.h. $f = (t - \lambda_1)^{k_1} \cdot \cdots \cdot (t - \lambda_s)^{k_s} \alpha, \quad \alpha \neq 0$.

Das Beweisverfahren zu Satz A.4, bricht nämlich erst ab, wenn $\operatorname{grad} g_s = 0$ ist. Das für uns wichtigste Beispiel eines abgebraisch abgeschlossenen Körpers ist der Körper $\mathbb{C}$ der komplexen Zahlen.

Definition A.5. *Es seien* f, g, d, v *aus* $\mathbb{K}[t]$ *gegeben.*
a) f *heißt genau dann ein* ***Teiler*** *von* g, $f \mid g$, *wenn*

$$g = fh \quad \text{mit einem Polynom} \quad h \in \mathbb{K}[t]$$

gilt. Ist $0 < \operatorname{grad} h < \operatorname{grad} g$, *so heißt* f *ein* ***nicht-trivialer Teiler*** *von* g.
b) d *heißt genau dann ein* ***größter gemeinsamer Teiler*** (ggT) *von* f *und* g, *wenn gilt:*

(i) $d \mid f$ *und* $d \mid g$.

(ii) *Gilt* $d^* \mid f$ *und* $d^* \mid g$ *für ein* $d^* \in \mathbb{K}[t]$, *so folgt* $d^* \mid d$.

c) v *heißt genau dann ein* ***kleinstes gemeinsames Vielfaches*** (kgV) *von* f *und* g, *wenn gilt:*

(i) *$f \mid v$ und $g \mid v$.*

(ii) *Gilt $f \mid v^*$ und $g \mid v^*$ für ein $v^* \in \mathbb{K}[t]$, so folgt $v \mid v^*$.*

Jedes $f \in \mathbb{K}[t]$ ist Teiler des Nullpolynoms. Das Nullpolynom ist nur Teiler von sich selbst. Ist daher $f \neq 0$ oder $g \neq 0$, so ist jeder ggT von f und g vom Nullpolynom verschieden.

Sind d und $\tilde{d}$ zwei ggT *von f und g mit $fg \neq 0$, so gilt: $\tilde{d} = \alpha d$, $\alpha \in \mathbb{K}$, $\alpha \neq 0$.*

Beweis. Aus Definition A.5, b), (ii) folgt $d \mid \tilde{d}$ und $\tilde{d} \mid d$. Da nicht $f = 0$ und $g = 0$ sein kann, gilt $d \neq 0$ und $\tilde{d} \neq 0$. Die Teilbarkeitsbeziehungen haben aber $\operatorname{grad} d \leq \operatorname{grad} \tilde{d} \leq \operatorname{grad} d$ zur Folge, d.h. $\operatorname{grad} d = \operatorname{grad} \tilde{d}$. Daraus und aus $d \mid \tilde{d}$ folgt

$$\tilde{d} = \alpha d \quad \text{mit} \quad \alpha \in \mathbb{K},\ \alpha \neq 0.$$

Ist umgekehrt d ein ggT von f, g, so ist auch $\tilde{d} = \alpha d$, $\alpha \in \mathbb{K}$, $\alpha \neq 0$, ein ggT von f, g. □

Analog gilt für zwei kleinste gemeinsame Vielfache $v, \tilde{v}$ von f, g die Beziehung $v = \alpha \tilde{v}$ mit $\alpha \in \mathbb{K}$, $\alpha \neq 0$.

Wir überlassen es dem Leser die Definitionen A.5, b) und c) für mehr als zwei Polynome zu formulieren.

Definition A.6. *Polynome $f, g \in \mathbb{K}[t]$ heißen genau dann **teilerfremd**, wenn* 1 *ein* ggT *von f, g ist, d.h. es existieren keine gemeinsamen Teiler von f und g mit positivem Grad.*

Die folgende Aussage ist für das Rechnen mit Polynomen von fundamentaler Bedeutung:

Satz A.7. *Es seien $f, g \in \mathbb{K}[t]$ mit $\operatorname{grad} f \geq \operatorname{grad} g \geq 0$ gegeben. Dann gibt es Polynome $h, r \in \mathbb{K}[t]$ mit*

$$f = gh + r$$

und

$$\operatorname{grad} r < \operatorname{grad} g.$$

Beweis. Es sei $\operatorname{grad} g = n$, $\operatorname{grad} f = n + k$ und

$$\begin{aligned} f &= \alpha_{n+k} t^{n+k} + \cdots + \alpha_n t^n + \cdots + \alpha_1 t + \alpha_0, \\ g &= \beta_n t^n + \cdots + \beta_0. \end{aligned}$$

Durch Division mit Rest erhält man:

$$\begin{aligned} f &= \frac{\alpha_{n+k}}{\beta_n} t^k g + \Big(\alpha_{n+k-1} - \alpha_{n+k} \frac{\beta_{n-1}}{\beta_n}\Big) t^{n+k-1} + \ldots \\ &\quad \cdots + \Big(\alpha_k - \alpha_{n+k} \frac{\beta_0}{\beta_n}\Big) t^k + \alpha_{k-1} t^{k-1} + \cdots + \alpha_0 \\ &= \tilde{\alpha}_{n+k} t^k g + g_1, \end{aligned}$$

mit $\operatorname{grad} g_1 \leq n + k - 1$ und $\tilde{\alpha}_{n+k} := \alpha_{n+k}/\beta_n$. Ist $\operatorname{grad} g_1 \geq \operatorname{grad} g$, so erhalten wir ganz analog

$$g_1 = \tilde{\alpha}_{n+k-1} t^{k-1} g + g_2 \quad \text{mit} \quad \operatorname{grad} g_2 \leq n + k - 2,$$

d.h. $f = (\tilde{\alpha}_{n+k} t^k + \tilde{\alpha}_{n+k-1} t^{k-1}) g + g_2$. Das Verfahren bricht nach endlich vielen Schritten ab, und zwar dann, wenn der Grad des Restpolynoms $\leq n - 1 = \operatorname{grad} g - 1$ ist. □

Es seien $f, g \in \mathbb{K}[t]$ mit $\operatorname{grad} f \geq \operatorname{grad} g \geq 0$ gegeben. Dann existieren eine natürliche Zahl n und Polynome $q_1, \ldots, q_n$, $r_1, \ldots, r_{n-1}$ mit

$$\operatorname{grad} g > \operatorname{grad} r_1 > \cdots > \operatorname{grad} r_{n-1} \geq 0$$

und

$$\begin{aligned}
f &= q_1 g + r_1, && (1)\\
g &= q_2 r_1 + r_2, && (2)\\
r_1 &= q_3 r_2 + r_3, && (3)\\
&\vdots && \vdots\\
r_{n-3} &= q_{n-1} r_{n-2} + r_{n-1}, && (n-1),\\
r_{n-2} &= q_n r_{n-1}. && (n)
\end{aligned}$$

Das Bestehen der Gleichungen (1) – $(n-1)$ und die Aussage über den Grad der Polynome $r_1, \ldots, r_n$ sind wegen Satz A.7 klar. Da es nur endlich viele natürliche Zahlen $\leq \operatorname{grad} g$ gibt, muß schließlich das Restpolynom Null werden. Das durch die Gleichungen (1) – (n) beschriebene Verfahren der fortgesetzten Division mit Rest heißt der *Euklidische Algorithmus*. Die Bedeutung des Euklidischen Algorithmus zeigt der folgende Satz:

Satz A.8. a) r_{n-1} *ist ein* ggT *von* f *und* g. *Daraus folgt insbesondere, dass zu zwei Polynomen stets ein* ggT *existiert.*
b) *Es gibt Polynome* h *und* k *mit*

$$r_{n-1} = hf + kg.$$

Insbesondere läßt sich dann jeder ggT d *von* f *und* g *in der Form* $d = \tilde{h} f + \tilde{k} g$ *mit* $\tilde{h}, \tilde{k} \in \mathbb{K}[t]$ *schreiben.*

Beweis. a) Aus Gleichung (n) folgt $r_{n-1} \mid r_{n-2}$. Damit folgt $r_{n-1} \mid r_{n-3}$ aus Gleichung $(n-1)$. Auf diese Weise fortfahrend erhält man $r_{n-1} \mid g$ und $r_{n-1} \mid f$.

Es sei nun d ein Polynom mit $d \mid f$ und $d \mid g$. Dann folgt $d \mid r_1$ aus Gleichung (1). Mit Hilfe von (2) folgt daraus zusammen mit $d \mid g$ auch $d \mid r_2$. Man erhält schließlich $d \mid r_{n-1}$.
b) Unter sukzessiver Benutzung der Gleichungen $(n-1)$ bis (1) erhalten wir

$$\begin{aligned}
r_{n-1} &= r_{n-3} - r_{n-2} q_{n-1} = r_{n-3} - (r_{n-4} - r_{n-3} q_{n-2}) q_{n-1}\\
&= (1 + q_{n-2} q_{n-1}) r_{n-3} - q_{n-1} r_{n-4} = \ldots\\
&= hf + kg
\end{aligned}$$

mit Polynomen h, k. $\square$

Beispiel A.9. Es sei $\mathbb{K} = \mathbb{R}$ und

$$f = t^5 + t^4 + t^3 + t^2 + t + 1, \quad g = t^4 + t^3 + 2t^2 + t + 1.$$

Die Durchführung des Enklidischen Algorithmus liefert:

$$\begin{aligned}
t^5 + t^4 + t^3 + t^2 + t + 1 &= t(t^4 + t^3 + 2t^2 + t + 1) + (-t^3 + 1),\\
t^4 + t^3 + 2t^2 + t + 1 &= (-t)(-t^3 + 1) + t^3 + 2t^2 + 2t + 1,\\
-t^3 + 1 &= (-1)(t^3 + 2t^2 + 2t + 1) + 2t^2 + 2t + 2,\\
t^3 + 2t^2 + 2t + 1 &= \frac{1}{2}(2t^2 + 2t + 2) + t^2 + t + 1\\
2t^2 + 2t + 2 &= 2(t^2 + t + 1).
\end{aligned}$$

Es ist somit $d = t^2+t+1$ ein ggT von f und g. Nach dem Schema des Beweises zu Satz A.8, b), erhält man:

$$d = t^2 + t + 1 = (-\frac{t^2}{2} + \frac{t}{2})f + (\frac{t^3}{2} - \frac{t^2}{2} - \frac{t}{2} + 1)g.$$

$\Diamond$

Eine nützliche Folgerung aus Satz A.8 ist:

Satz A.10. *Es seien $f, g, h \in \mathbb{K}[t]$ gegeben. Sind f und g teilerfremd, so folgt aus $f \mid gh$*

$$f \mid h.$$

Beweis. Auf Grund der Voraussetzung ist 1 ein ggT von f und g. Nach Satz A.8, b), gilt

$$1 = k_1 f + k_2 g$$

mit Polynomen k_1, k_2. Daraus folgt

$$h = k_1 f h + k_2 g h.$$

Wegen $f \mid gh$ folgt daraus sofort $f \mid h$. $\square$

Die Berechnung eines kgV zweier Polynome f, g kann nach folgendem Schema erfolgen:

Es sei d ein ggT *von f und g. Dann ist $f = f_1 d$, $g = g_1 d$. Ein* kgV *von f und g ist dann durch*

$$v = f_1 g_1 d$$

gegeben.

Beweis. Es ist klar, dass $f \mid v$ und $g \mid v$ gilt. Es sei nun w ein Polynom mit $f \mid w$ und $g \mid w$, d.h.

$$w = f w_1 = g w_2.$$

Daraus folgt $f_1 d w_1 = g_1 d w_2$ und wegen $d \neq 0$

$$f_1 w_1 = g_1 w_2,$$

d.h. $g_1 \mid f_1 w_1$. Sind f_1 und g_1 teilerfremd, folgt daraus (Satz A.10) $g_1 \mid w_1$, d.h. $w_1 = g_1 w_3$. Somit gilt $w = f_1 d g_1 w_3 = v w_3$, d.h. $v \mid w$. Wir müssen noch zeigen, dass f_1, g_1 teilerfremd sind. Wäre h ein Teiler von f_1 und g_1 mit $\operatorname{grad} h > 0$, so wäre $\tilde{d} = hd$ ein Teiler von f und g mit $\operatorname{grad} \tilde{d} > \operatorname{grad} d$ in Widerspruch zur Tatsache, dass d ein ggT von f und g ist. $\square$

Es seien $f = \alpha_0 + \alpha_1 t + \cdots + \alpha_k t^k \in \mathbb{K}[t]$ und $\varphi \in L(X, X)$ gegeben, wobei X ein Vektorraum über $\mathbb{K}$ ist. Dann wird durch

$$f(\varphi) = \alpha_0 \epsilon + \alpha_1 \varphi + \cdots + \alpha_k \varphi^k$$

ein Endomorphismus von X definiert. Ist $A \in M_{n,n}(\mathbb{K})$, so können wir analog die Matrix

$$f(A) = \alpha_0 E + \alpha_1 A + \cdots + \alpha_k A^k \in M_{n,n}(\mathbb{K})$$

definieren. Ist $\dim X = n$ und A die Matrix von $\varphi \in L(X, X)$ bzgl. einer Basis von X, so ist $f(A)$ die Matrix von $f(\varphi)$ bzgl. derselben Basis. Dies folgt aus der Definition der Summe und des Produktes von Matrizen (siehe Abschnitt 7.2).

Es seien $f, g, h \in \mathbb{K}[t]$ und $\varphi \in L(X, X)$, $A \in M_{n,n}(\mathbb{K})$ gegeben. Dann folgt aus

$$f + g = h \quad \text{bzw.} \quad fg = h$$

auch $f(\varphi) + g(\varphi) = h(\varphi)$, $f(A) + g(A) = h(A)$ bzw.

$$f(\varphi)g(\varphi) = h(\varphi), \quad f(A)g(A) = h(A).$$

Insbesondere gilt wegen $fg = gf$ stets $f(\varphi)g(\varphi) = g(\varphi)f(\varphi)$ und $f(A)g(A) = g(A)f(A)$.

Index