Prof. Dr. V. Bangert

Lineare Algebra I

Wintersemester 2006/07

1 Grundlagen

A) Zur Sprache in der Mathematik

Mathematik wird mitgeteilt in der üblichen Sprache, die mit Fachausdrücken angereichert ist und die sehr viel präziser verwendet wird als gewöhnlich. Zur Abkürzung verwenden wir folgende logische Symbole:

(1.1) "Aus der Aussage A folgt die Aussage B", wird abgekürzt als "$A \Rightarrow B$".
Beispiel: Ist x eine rationale Zahl, für die $(x-1)(x^2-2)=0$ gilt, so gilt $x=1$.
Abgekürzt: $x \in \mathbb{Q}$ und $(x-1)(x^2-2)=0 \Rightarrow x=1$.
Andere sprachliche Ausdrucksweisen für "$A \Rightarrow B$":
"Wenn A gilt, so gilt auch B" oder
"A impliziert B" oder
"A ist hinreichend für B" oder
"B ist notwendig für A".

(1.2) "$A \Leftrightarrow B$" bedeutet "$A \Rightarrow B$ und $B \Rightarrow A$".
Beispiel: Die Seiten eines Dreiecks $\triangle$ sind gleich lang $\Leftrightarrow$ Jeder Innenwinkel von $\triangle$ ist $60°$.
Sprachlich wird "$A \Leftrightarrow B$" ausgedrückt durch "A und B sind äquivalent" oder durch "A gilt genau dann , wenn B gilt" oder durch "A gilt dann und nur dann, wenn B gilt" oder durch "A ist notwendig und hinreichend für B".

(1.3) "Es existiert (mindestens) ein ..." wird abgekürzt durch das logische Symbol "$\exists$".
Beispiel: $\exists x \in \mathbb{R} : x^2 = 2$. In Worten: Es existiert eine reelle Zahl x, für die $x^2 = 2$ gilt.

(1.4) "Für alle ..." wird abgekürzt durch das logische Symbol "$\forall$".
Beispiel: $\forall x \in \mathbb{R} : x^2 \geq 0$. In Worten: Für alle reellen Zahlen x gilt $x^2 \geq 0$. (statt "Für alle" auch "Für jede").

(1.5) "A wird durch B definiert" wird abgekürzt als "$A := B$". Beispiel: $e^x := \sum\limits_{n=0}^{\infty} \frac{x^n}{n!}$.
Gemeint ist hier, daß das Symbol e^x durch die rechtsstehende unendliche Reihe definiert wird.

Schließlich: "oder" ist nicht ausschließend (im Gegensatz zu "entweder ... oder").
Beispiel: Die Aussage "Es gilt $1+1=2$ oder es gilt $1-1=0$" ist wahr. Hingegen ist "Entweder es gilt $1+1=2$ oder es gilt $1-1=0$" eine falsche Aussage.

B) Mengen und Abbildungen (naive Mengenlehre)
(Georg Cantor, 1845-1918)

Eine Menge ist eine Zusammenfassung von mathematischen Objekten, die die Elemente der Menge genannt werden. Zwei Mengen sind genau dann gleich, wenn sie dieselben Elemente enthalten.

Bezeichnungen:

"$x \in M$" steht für "x ist Element der Menge M",
"$x \notin M$" steht für "x ist nicht Element der Menge M".

Angabe von Mengen:

(1.6) Aufzählend, z.B. $M = \{1,3,5,7\} (= \{7,5,3,1\} = \{1,1,3,5,7\})$.
Speziell: $\emptyset = \{\quad\}$ bezeichnet die "leere Menge", die dadurch definiert ist, daß sie kein Element enthält.

(1.7) Durch Angabe von Eigenschaften: Sei M eine Menge und $A(x)$ eine Aussage über die Elemente von M. Dann kann man die Menge N betrachten, die aus denjenigen Elementen x von M besteht, für die die Aussage $A(x)$ wahr ist,

$$N = \{x \mid x \in M \text{ und } A(x) \text{ ist wahr}\} = \{x \in M \mid A(x) \text{ ist wahr}\}.$$

Wenn wir davon ausgehen, daß wir die natürlichen, die ganzen und die reellen Zahlen schon kennen, so können wir folgende Symbole für sie einführen:

$\mathbb{N} := \{x \in \mathbb{R} \mid x \text{ ist natürliche Zahl}\} = \{0,1,2,3,\ldots\}$
$\mathbb{Z} := \{x \in \mathbb{R} \mid x \text{ ist ganze Zahl}\} = \{0,1,-1,2,-2,\ldots\}$

Die rationalen Zahlen können dann definiert werden durch

$$\begin{aligned} \mathbb{Q} &:= \{x \in \mathbb{R} \mid \exists p \in \mathbb{Z}, \exists q \in \mathbb{N}, q \neq 0 : x = \tfrac{p}{q}\} \\ (&= \{x \in \mathbb{R} \mid \exists p \in \mathbb{Z}, \exists q \in \mathbb{Z}, q \neq 0 : x = \tfrac{p}{q}\}). \end{aligned}$$

Die Menge der ungeraden natürlichen Zahlen kann z.B. als $\{n \in \mathbb{N} \mid \exists m \in \mathbb{N} : n = 2m+1\}$ beschrieben werden. Dafür ist auch folgende kürzere Schreibweise üblich: $\{2m + 1 \mid m \in \mathbb{N}\}$.

Beispiel: $\{1, 3, 5, 7\} = \{n \mid n$ ist ungerade natürliche Zahl und $n < 9\}$.

(1.8) Def.: Eine Menge A heißt <u>Teilmenge einer Menge</u> B (abgekürzt: $A \subseteq B$), falls jedes Element von A auch Element von B ist (oder äquivalent, falls gilt: $x \in A \Rightarrow x \in B$).

Beispiel: 1) A Menge $\Rightarrow \emptyset \subseteq A$, $A \subseteq A$
2) $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$

Bem.: Für Mengen A und B gilt:

$$A = B \Leftrightarrow A \subseteq B \text{ und } B \subseteq A.$$

(1.9) Def.: Seien A und B Mengen. Dann heißt

(a) $A \cap B := \{x \mid x \in A \text{ und } x \in B\}$ der <u>Durchschnitt</u> (oder die <u>Schnittmenge</u>) <u>von</u> A <u>und</u> B.
(b) $A \cup B := \{x \mid x \in A \text{ oder } x \in B\}$ die <u>Vereinigung von</u> A <u>und</u> B.
(c) $A \setminus B := \{x \mid x \in A \text{ und } x \notin B\}$ das <u>Komplement von</u> B <u>in</u> A.

Es ist oft hilfreich, sich solche Beziehungen zwischen Mengen anhand von Diagrammen zu veranschaulichen.

(1.10) Rechenregeln. Sind A, B, C Mengen, so gilt:

$$\begin{array}{llcll}
\text{(a)} & A \cup B & = & B \cup A,\ A \cap B = B \cap A & \text{“Kommutativgesetze”} \\[1em]
\text{(b)} & (A \cup B) \cup C & = & A \cup (B \cup C) & \text{“Assoziativgesetze”} \\
& (A \cap B) \cap C & = & A \cap (B \cap C) & \\[1em]
\text{(c)} & A \cap (B \cup C) & = & (A \cap B) \cup (A \cap C) & \text{“Distributivgesetze”} \\
& A \cup (B \cap C) & = & (A \cup B) \cap (A \cup C) & \\[1em]
\text{(d)} & A \setminus (B \cup C) & = & (A \setminus B) \cap (A \setminus C) & \text{“de Morgansche} \\
& A \setminus (B \cap C) & = & (A \setminus B) \cup (A \setminus C) & \text{Regeln”}
\end{array}$$

Exemplarisch wird ein Beweis für $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$ gegeben:

Zeige zunächst: $x \in (A \setminus B) \cap (A \setminus C) \Rightarrow x \in A \setminus (B \cup C)$
$x \in (A \setminus B) \cap (A \setminus C) \Rightarrow (x \in A$ und $x \notin B)$ und $(x \in A$ und $x \notin C) \Rightarrow x \in A$ und $(x \notin B$ und $x \notin C) \Rightarrow x \in A$ und $(x \notin B \cup C) \Rightarrow x \in A \setminus (B \cup C)$

Die andere Richtung, nämlich $x \in A \setminus (B \cup C) \Rightarrow x \in (A \setminus B) \cap (A \setminus C)$, kann man dadurch einsehen, daß alle Implikationen “$\Rightarrow$” des vorangehenden Beweises in Wirklichkeit Äquivalenzen “$\Leftrightarrow$” sind. Eine etwas andere Möglichkeit ist, folgendermaßen zu argumentieren.

Aus $B \subseteq B \cup C$ folgt offenbar $A \setminus (B \cup C) \subseteq A \setminus B$, und ebenso $A \setminus (B \cup C) \subseteq A \setminus C$. Die beiden letzten Inklusionen implizieren

$$A \setminus (B \cup C) \subseteq (A \setminus B) \cap (A \setminus C).$$

Es ist wichtig, zu wissen, daß Mengen selbst wieder als Elemente von Mengen auftreten können.

Beispiel: $\{\{0,1\},\{1,2\},\{2,3\}\ldots\} = \{\{n, n+1\} \mid n \in \mathbb{N}\}$.

(1.11) Def.: Ist A eine Menge, so heißt

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

die Potenzmenge von A.

$\mathcal{P}(A)$ ist also die Menge, deren Elemente genau die Teilmengen von A sind.

Beispiel: $A = \{1,2\} \Rightarrow \mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}$.

Ist A eine Menge mit $n(\in \mathbb{N})$ Elementen, so enthält $\mathcal{P}(A)$ 2^n Elemente (Beweis durch vollständige Induktion nach n).

(1.12) Def.: Es seien A und B Mengen. Eine Abbildung $f : A \to B$ von A nach B ist eine Vorschrift, die jedem Element von A ein (und nur ein) Element von B zuordnet.

Bez.: Ist $x \in A$, so bezeichnet $f(x)$ das Element von B, das x durch f zugeordnet wird. Statt $f : A \to B$ schreibt man manchmal auch $A \xrightarrow{f} B$. Eine Abbildung $f : A \to \mathbb{R}$ nennt man oft auch eine Funktion. A heißt der Definitionsbereich von f, B die Zielmenge von f.

Beispiel:

1) $f : \mathbb{R} \to \mathbb{R}, \forall x \in \mathbb{R} : f(x) := x^2 - 1$.
2) Sei $\mathbb{R}_{>0} := \{x \in \mathbb{R} \mid x > 0\}$ und $f : \mathbb{R} \to \mathbb{R}_{>0}$ definiert durch $f(x) = e^x$ für alle $x \in \mathbb{R}$.
3) $f : \mathbb{N} \to \{-1, 1\}, \forall n \in \mathbb{N} : f(n) = (-1)^n = \begin{cases} 1 & \text{falls } n \text{ gerade} \\ -1 & \text{falls } n \text{ ungerade} \end{cases}$
4) Sei A Menge. Dann ist die Identität (oder: die identische Abbildung) $\mathrm{id}_A : A \to A$ von A definiert durch: $\forall x \in A : \mathrm{id}_A(x) = x$.
5) Seien A, B Mengen und $A \subseteq B$. Die Abbildung $i : A \to B$, definiert durch $i(x) = x$ für alle $x \in A$, heißt Inklusionsabbildung von A in B.

(1.13) Def.: Eine Abbildung $f : A \to B$ heißt

(a) injektiv, falls für alle $a \in A$, $b \in A$ gilt: $a \neq b \Rightarrow f(a) \neq f(b)$.

(b) surjektiv, falls für jedes $b \in B$ ein $a \in A$ existiert, so daß $f(a) = b$ gilt.

(c) bijektiv, falls f injektiv und surjektiv ist.

Bem.: Von den obigen Beispielen sind 2), 4) und 5) injektiv und 2), 3) und 4) surjektiv.

Schreibweisen: Sei $f : A \to B$ eine Abbildung.

Ist $C \subseteq A$, so benützen wir die Abkürzung

$$f(C) := \{b \in B \mid \exists a \in C : f(a) = b\}.$$

Die Menge $f(C) \subseteq B$ heißt das Bild von C unter f.

Ist $D \subseteq B$, so schreiben wir

$$f^{-1}(D) := \{a \in A \mid f(a) \in D\}.$$

Die Menge $f^{-1}(D) \subseteq A$ heißt das Urbild von D unter f.

Bem.: $f : A \to B$ surjektiv $\Leftrightarrow$ $f(A) = B$
$f : A \to B$ injektiv $\Leftrightarrow$ Für jedes $b \in B$ enthält $f^{-1}(\{b\})$ höchstens ein Element.

(1.14) Rechenregeln. Sei $f : M \to N$ Abbildung, $A \subseteq M$, $B \subseteq M$, $C \subseteq N$, $D \subseteq N$. Dann gilt:

(a) $f(A \cup B) = f(A) \cup f(B)$
(b) $f(A \cap B) \subseteq f(A) \cap f(B)$
(c) $f(M) \setminus f(A) \subseteq f(M \setminus A)$
(d) $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$
(e) $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$
(f) $f^{-1}(N \setminus C) = M \setminus f^{-1}(C)$

Bew. von (d): $x \in f^{-1}(C \cup D) \Rightarrow f(x) \in C \cup D \Rightarrow f(x) \in C$ oder $f(x) \in D \Rightarrow x \in f^{-1}(C)$ oder $x \in f^{-1}(D) \Rightarrow x \in f^{-1}(C) \cup f^{-1}(D)$.

Daraus folgt: $f^{-1}(C \cup D) \subseteq f^{-1}(C) \cup f^{-1}(D)$.

Offensichtlich gilt $f^{-1}(C) \subseteq f^{-1}(C \cup D)$, $f^{-1}(D) \subseteq f^{-1}(C \cup D)$, also auch $f^{-1}(C) \cup f^{-1}(D) \subseteq f^{-1}(C \cup D)$.

(1.15) Def. (Komposition von Abbildungen). Seien A, B, C, D Mengen mit $B \subseteq C$, und $f : A \to B$, $g : C \to D$ Abbildungen. Dann ist die Abbildung $g \circ f : A \to D$ definiert durch:
Für alle $x \in A$ setzen wir $(g \circ f)(x) := g(f(x))$.

Bem.: $g(f(x))$ läßt sich bilden, da $f(x) \in B$ gilt und B eine Teilmenge des Definitionsbereichs C von g ist.

Bez.: Statt "Komposition" wird auch das Wort "Hintereinanderausführung" benutzt. "$g \circ f$"wird gelesen als "g Kringel f" oder "g nach f".

(1.16) Def.: Ist $f : M \to N$ Abbildung und $A \subseteq M$, so ist die Restriktion (oder Einschränkung) $f|A : A \to N$ von f auf A definiert durch:

$$\text{Für alle } x \in A \text{ ist } (f|A)(x) := f(x).$$

Bem.: Ist $i : A \to M$ die Inklusionsabbildung, so gilt $f|A = f \circ i$.

(1.17) Fakt: Ist $f : M \to N$ eine bijektive Abbildung, so existiert genau eine Abbildung $f^{-1} : N \to M$, so daß $f^{-1} \circ f = \mathrm{id}_M$ gilt. Für diese Abbildung gilt auch $f \circ f^{-1} = \mathrm{id}_N$.

Bez.: f^{-1} heißt Umkehrabbildung von f (oder die zu f inverse Abbildung). Vorsicht: Das Symbol $f^{-1}(C)$ ist für beliebige Abbildungen $f : M \to N$ und $C \subseteq N$ definiert. Die Abbildung $f^{-1} : N \to M$ gibt es aber nur, wenn f bijektiv ist. Wenn die Abbildung $f^{-1} : N \to M$ existiert, so stimmt für jedes $C \subseteq N$ die Bildmenge von C unter f^{-1} (bezeichnet durch $f^{-1}(C)$) mit der Urbildmenge von C unter f (ebenfalls bezeichnet durch $f^{-1}(C)$) überein.

Bew. von (1.17): Die Bijektivität von f besagt gerade, daß es zu jedem $y \in N$ genau ein $x \in M$ gibt, für das $f(x) = y$ gilt. Damit $f^{-1} \circ f = \mathrm{id}_M$ gilt, muß f^{-1} das Element y von N auf dieses $x \in M$ abbilden, und wenn wir $f^{-1} : N \to M$ auf diese Weise definieren, so gilt sowohl $f^{-1} \circ f = \mathrm{id}_M$ als auch $f \circ f^{-1} = \mathrm{id}_N$.

Bsp.:

1) $M = N$, $f = \mathrm{id}_M \Rightarrow f^{-1} = \mathrm{id}_M$ Denn: $\mathrm{id}_M \circ \mathrm{id}_M = \mathrm{id}_M$.

2) $f : \mathbb{R} \to \mathbb{R}_{>0}$, $f(x) = e^x$, ist bijektiv. Die Umkehrabbildung von f ist der natürliche Logarithmus $\ln : \mathbb{R}_{>0} \to \mathbb{R}$, d.h. es gilt für alle $x \in \mathbb{R}$: $\ln(e^x) = x$, und für alle $y \in \mathbb{R}_{>0} : e^{\ln y} = y$.

(1.18) Def.: Sind A und B Mengen, so ist das kartesische Produkt $A \times B$ von A und B die Menge alle geordneten Paare (a, b) mit $a \in A$ und $b \in B$, oder kürzer:

$$A \times B := \{(a, b) \mid a \in A, b \in B\}.$$

Dabei soll der Begriff "geordnetes Paar" so definiert sein, daß $(a, b) = (a', b')$ genau dann gilt, wenn $a = a'$ und $b = b'$ gilt.

Das "kartesische Produkt" ist benannt nach R. Descartes (1596 - 1650), dem Begründer der analytischen Geometrie. Allgemeiner definieren wir zu Mengen $A_1, \ldots, A_n$ deren kartesisches Produkt

$$A_1 \times \ldots \times A_n := \{(a_1, \ldots, a_n) \mid a_1 \in A_1, \ldots, a_n \in A_n\}.$$

Dabei soll wieder gelten

$$(a_1, \ldots, a_n) = (a'_1, \ldots, a'_n) \Leftrightarrow a_1 = a'_1, \ldots, a_n = a'_n.$$

Speziell für $\mathbb{Z}, \mathbb{Q}$ und $\mathbb{R}$ und $n \in \mathbb{N}$, $n \geq 1$, schreiben wir

$$\mathbb{Z}^n := \underbrace{\mathbb{Z} \times \ldots \times \mathbb{Z}}_{n\text{-mal}}$$

und analog $\mathbb{Q}^n$, $\mathbb{R}^n$.

Bsp.: 1) Auf einem kartesische Produkt $A_1 \times \ldots \times A_n$ haben wir die Projektionsabbildung $p_i : A_1 \times \ldots \times A_n \to A_i$ auf die i'te Koordinate,

$$p_i(x_1, \ldots, x_n) := x_i$$

2) Die Addition von reellen Zahlen kann man als Abbildung $+ : \mathbb{R}^2 \to \mathbb{R}$, $+(x, y) := x + y$, betrachten.

Die Elemente von $\mathbb{R}^n$ nennt man n-Tupel reeller Zahlen. Mit diesem Begriff läßt sich mathematisch präzis sagen, was die Lösungsmenge L einer Gleichung

$$a_1x_1 + \ldots + a_nx_n = b$$

mit $a_1, \dots, a_n$, $b \in \mathbb{R}$ ist:

$$L = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid a_1x_1 + \dots + a_nx_n = b\}.$$

Ist eines der $a_1, \dots, a_n$ nicht null, etwa $a_i \neq 0$, so gilt

$$L = ((x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i = \frac{1}{a_i}(b - a_1x_1 \dots - a_{i-1}x_{i-1} - a_{i+1}x_{i+1} \dots - a_nx_n)\}.$$

Gilt $a_i = 0$ für alle $1 \leq i \leq n$, so gilt $L = \emptyset$, falls $b \neq 0$, und $L = \mathbb{R}^n$, falls $b = 0$.

(1.19a) Def.: Seien M, N Mengen. Eine Relation auf (M, N) ist eine Teilmenge von $M \times N$. Ist $M = N$, so spricht man von einer Relation auf M.

Beispiele:

1) Sei $f : M \to N$ Abbildung. Dann ist der Graph von f,

$$\mathrm{graph}(f) := \big\{\big(x, f(x)\big) \mid x \in M\big\}$$

eine Relation auf (M, N).

2) Sei $M = N = \mathbb{R}$ und $R = \big\{(x, y) \in \mathbb{R}^2 \mid x = y^2\big\}$. Es existiert kein $f : \mathbb{R} \to \mathbb{R}$ mit $\mathrm{graph}(f) = R$.

3) Sei $M = N = \mathbb{R}$ und $R = \big\{(x, y) \in \mathbb{R}^2 \mid x \leq y\big\}$. R ist die Relation „kleiner gleich“ .

(1.19) Def.: Sei M Menge und $R \subseteq M \times M$ Relation auf M. R heißt

(a) reflexiv $\Leftrightarrow \forall x \in M : (x, x) \in R$

(b) symmetrisch $\Leftrightarrow (\forall x, y \in M : (x, y) \in R \Rightarrow (y, x) \in R)$

(c) transitiv $\Leftrightarrow (\forall x, y, z \in M : (x, y) \in R$ und $(y, z) \in R \Rightarrow (x, z) \in R)$

(d) Äquivalenzrelation auf $M \Leftrightarrow R$ ist reflexiv, symmetrisch und transitiv.

Bei Äquivalenzrelationen R schreibt man oft $x \sim_R y$ statt $(x, y) \in R$ (oder kurz $x \sim y$, wenn klar ist, welche Äquivalenzrelation R gemeint ist).

Beispiele:

1) Die Relation "kleiner gleich" auf $\mathbb{R}$, d.h.

$$R = \{(x, y) \in \mathbb{R}^2 \mid x \leq y\},$$

ist symmetrisch und transitiv, aber nicht reflexiv. (R ist nicht reflexiv, da etwa $(0, 1) \in R$, aber $(1, 0) \notin R$).

2) Wir sagen, eine ganze Zahl $n \in \mathbb{Z}$ sei durch $p \in \mathbb{Z}$ teilbar, kurz $p|n$, falls ein $k \in \mathbb{Z}$ existiert, so daß $n = kp$ gilt. Für festes $p \in \mathbb{Z}$, $p \geq 2$, definieren wir folgende Relation R auf $\mathbb{Z}$:

$$R := \{(m, n) \in \mathbb{Z}^2 \mid n - m \text{ ist durch } p \text{ teilbar}\}.$$

Wir zeigen, daß R eine Äquivalenzrelation ist:

(a) $\forall m \in \mathbb{Z} : (m, m) \in R$. Denn $m - m = 0$ ist durch p teilbar, $m - m = 0 \cdot p$

(b) Sei $(m, n) \in R$. Dann existiert $k \in \mathbb{Z} : n - m = kp$. Dann gilt: $m - n = (-k) \cdot p$ und $-k \in \mathbb{Z}$. Also $(n, m) \in R$.

(c) Seien $(l, m) \in R$ und $(m, n) \in R$. Dann existieren $j \in \mathbb{Z}$, $k \in \mathbb{Z}$, so daß $m - l = jp$ und $n - m = kp$ gilt. Addition dieser Gleichungen liefert

$$n - l = (k + j)p.$$

Wegen $k + j \in \mathbb{Z}$ folgt daraus $(l, n) \in R$.

3) Eine Zerlegung (oder Klasseneinteilung) einer Menge M, ist eine Teilmenge $\mathcal{M}$ von $\mathcal{P}(M)$ (d.h. $\mathcal{M}$ ist eine Menge, deren Elemente Teilmengen von M sind), so daß folgende Bedingungen erfüllt sind:

(1) $A \in \mathcal{M}, B \in \mathcal{M} \Rightarrow A = B$ oder $A \cap B = \emptyset$

(2) $\bigcup\limits_{A \in \mathcal{M}} A = M$ (, wobei $\bigcup\limits_{A \in \mathcal{M}} A := \{x | \exists A \in \mathcal{M} : x \in A\}$)

Zusätzlich wollen wir fordern, daß $\emptyset \notin \mathcal{M}$.

Bsp.: Die Menge der Schulklassen einer Schule ist eine Zerlegung (oder Klasseneinteilung) der Menge der Schüler dieser Schule.

Wir werden sehen, daß eine Äquivalenzrelation auf M im Grunde das gleiche ist, wie eine Zerlegung von M. Zunächst überlegen wir, wie wir aus einer Zerlegung eine Äquivalenzrelation erhalten: Zu einer Zerlegung $\mathcal{M}$ von M definieren wir die Relation

$$R = \{(x, y) \in M \times M \mid \exists A \in \mathcal{M} : \{x, y\} \subseteq A\}$$

Wir zeigen: R ist eine Äquivalenzrelation.

(a) Sei $x \in M$. Wir wollen zeigen, daß $(x, x) \in R$ gilt. Wegen (2) existiert ein $A \in \mathcal{M}$, so daß $x \in A$ gilt. Also gilt $\{x, x\}(= \{x\}) \subseteq A$, d.h. $(x, x) \in R$.

(b) $\{x, y\} \subseteq A \Leftrightarrow \{y, x\} \subseteq A$

(c) Es gelte $(x, y) \in R$ und $(y, z) \in R \Rightarrow \exists A \in \mathcal{M}$, $B \in \mathcal{M}$ mit $\{x, y\} \subseteq A$, $\{y, z\} \subseteq B \Rightarrow y \in A \cap B$, d.h. $A \cap B \neq \emptyset \overset{(1)}{\Rightarrow} A = B \Rightarrow \{x, z\} \subseteq A \Rightarrow (x, z) \in R$.

Bezeichnungen: Sei $R \subseteq M \times M$ eine Äquivalenzrelation. Wir schreiben $x \sim y$ für $(x, y) \in R$. Ist $x \in M$, so nennen wir

$$[x]_R := \{y \in M \mid x \sim y\}$$

die <u>Äquivalenzklasse</u> von x. Die Menge der Äquivalenzklassen von R wird mit M/R (gesprochen: M modulo R) bezeichnet, d.h.

$$M/R = \{[x]_R \mid x \in M\}.$$

(1.20) Fakt: Sei $R \subseteq M \times M$ eine Äquivalenzrelation. Dann ist M/R eine Zerlegung von M.

Beweis: Offensichtlich ist jedes Element von M/R eine Teilmenge von M (nämlich eine Äquivalenzklasse). Wir zeigen, daß die Menge M/R die Eigenschaften (1) und (2) hat.

Zu (2): Offensichtlich gilt $\bigcup\limits_{[x]_R \in M/R} [x]_R \subseteq M$. Sei umgekehrt $x \in M$. Die Reflexivität (a) von R besagt, daß $x \sim x$ gilt, also $x \in [x]_R$. Daraus folgt $M \subseteq \bigcup\limits_{[x]_R \in M/R} [x]_R$.

Zu (1): Wir zeigen, daß folgende Aussagen $(\alpha), (\beta), (\gamma)$ über Elemente $x \in M, y \in M$ äquivalent sind:

(α) $[x]_R \cap [y]_R \neq \emptyset$

(β) $x \sim y$

(γ) $[x]_R = [y]_R$

1. Schritt: $(\alpha) \Rightarrow (\beta)$: Sei $z \in [x]_R \cap [y]_R$, d.h. es gelten $x \sim z$ und $y \sim z$. Die Symmetrie und Transitivität von $\sim$ implizieren dann $x \sim y$.
2. Schritt: $(\beta) \Rightarrow (\gamma)$: Wegen der Symmetrie der Voraussetzung $x \sim y$ genügt es, $[y]_R \subseteq [x]_R$ zu zeigen. Sei $w \in [y]_R$, d.h. $y \sim w$. Unsere Voraussetzung $x \sim y$ ergibt zusammen mit $y \sim w$ und der Transitivität, daß $x \sim w$ gilt, d.h. $w \in [x]_R$.
3. Schritt: $(\gamma) \Rightarrow (\alpha)$: Wegen $x \in [x]_R$ folgt aus $[x]_R = [y]_R$, daß $[x]_R \cap [y]_R = [x]_R \neq \emptyset$ gilt.

Sind nun A und B Elemente von M/R, etwa $A = [x]_R$ und $B = [y]_R$ für Elemente x, y von M, so gilt wegen $(\alpha) \Leftrightarrow (\gamma)$ entweder $A \cap B = \emptyset$ oder $A = B$. Das beweist (1).

Beispiel: Es sei nun wieder $p \in \mathbb{Z}, p \geq 2$, und wir betrachten die Äquivalenzrelation auf $\mathbb{Z}$, die durch

$$m \sim n \Leftrightarrow n - m \text{ ist durch } p \text{ teilbar}$$

definiert ist, vgl. Bsp. 2 nach (1.19). Dann sind die Äquivalenzklassen von $\sim$ genau die Teilmengen von $\mathbb{Z}$, die aus den ganzen Zahlen bestehen, die bei Division (mit Rest) durch p, den gleichen Rest ergeben. Sie heißen deshalb auch die “Restklassen $\bmod p$”. Es gibt genau p verschiedene solche Restklassen mod$p$, nämlich

$$\begin{array}{lcll} [0]_p & = & \{np \mid n \in \mathbb{Z}\} & \text{Rest 0 bei Division durch } p \\ [1]_p & = & \{np+1 \mid n \in \mathbb{Z}\} & \text{Rest 1 bei Division durch } p \\ & \vdots & & \\ [p-1]_p & = & \{np+(p-1) \mid n \in \mathbb{Z}\} & \text{Rest } p-1 \text{ bei Division durch } p \\ & = & \{np-1 \mid n \in \mathbb{Z}\} & \end{array}$$

Für diese Äquivalenzrelation ist folgende Bezeichnung gebräuchlich:

$$m \sim n \Leftrightarrow m \equiv n \pmod p$$

2 Gruppen und Körper

(2.1) Def.: Eine Gruppe besteht aus einer Menge, hier genannt G, und einer Abbildung von $G \times G$ nach G, hier bezeichnet durch

$$\top : G \times G \to G, \top(a,b) =: a \top b,$$

so daß folgende Eigenschaften $(G_1) - (G_3)$ erfüllt sind:

(G_1) $\forall a,b,c \in G : a \top (b \top c) = (a \top b) \top c$ "Assoziativgesetz"

(G_2) Es existiert ein Element, genannt e, in G, so daß für alle $a \in G$ gilt: $e \top a = a$ "Existenz eines Linksneutralen"

(G_3) Für alle $a \in G$ existiert ein $b \in G$, so daß $b \top a = e$ gilt. "Existenz eines Linksinversen zu jedem $a \in G$"

Eine Gruppe $(G, \top)$ heißt abelsch, falls für alle $a \in G$, $b \in G$ gilt:

$$a \top b = b \top a \qquad \text{"Kommutativgesetz"}$$

Bezeichnungen:

1) $\top$ heißt die innere Verknüpfung der Gruppe $(G, \top)$. Ist die Gruppe $(G, \top)$ abelsch, so schreibt man oft "+" statt "$\top$". Sonst schreibt man meist "$\cdot$" für "$\top$".

2) Eine Gruppe $(G, \top)$ heißt endlich, falls G nur endlich viele Elemente hat, sonst unendlich. Die Anzahl $|G| \in \mathbb{N} \cup \{\infty\}$ der Elemente von G heißt die Ordnung der Gruppe $(G, \top)$.

Bem.: Die abelschen Gruppen sind nach dem norwegischen Mathematiker Niels Henrik Abel (1802-1829) benannt.

Beispiele:

1) $G := \mathbb{Z}$, $\top := + : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, $e := 0 \in \mathbb{Z}$. Dann ist $(\mathbb{Z}, +)$ eine Gruppe, ebenso $(\mathbb{Q}, +)$ und $(\mathbb{R}, +)$. Jedoch ist $(\mathbb{N}, +)$ keine Gruppe: Es müßte $e = 0$ gelten, aber für $n \in \mathbb{N} \setminus \{0\}$ existiert kein Linksinverses (in $\mathbb{N}$).

2) Ebenso sind $(\mathbb{Q}_{>0}, \cdot)$, $(\mathbb{R}_{>0}, \cdot)$, $(\mathbb{Q}\setminus\{0\}, \cdot)$, $(\mathbb{R}\setminus\{0\}, \cdot)$ Gruppen. In diesen Beispielen ist e die Zahl 1. Jedoch ist $(\mathbb{Z} \setminus \{0\}, \cdot)$ keine Gruppe.
Alle Gruppen in 1) und 2) sind abelsch und unendlich.

3) Zu einer Menge $M \neq \emptyset$ sei $S_M := \{f : M \to M \mid f \text{ bijektiv}\}$ und $\top : S_M \times S_M \to S_M$ sei die Hintereinanderausführung "$\circ$", d.h. $\top(f, g) := g \circ f$. Dann ist $(S_M, \circ)$ eine Gruppe mit $e := \mathrm{id}_M$. Das Links- (und Rechts-) inverse von $f \in S_M$ ist die Umkehrabbildung f^{-1} von f. $(S_M, \circ)$ heißt die symmetrische Gruppe (oder Permutationsgruppe) der Menge M. Wir setzen $S_n := S_{\{1,\dots,n\}}$.

Bem.:

a) Ist $|M| \geq 3$, so ist $(S_M, \circ)$ nicht abelsch.

b) Es gilt $|S_n| = n!$

Sowohl in der Mathematik wie in der Physik ist der Begriff "Gruppe" ein wichtiges Mittel, um die Symmetrien eines geometrischen Objekts (oder allgemeinerer Strukturen) zu untersuchen. In der Geometrie betrachtet man etwa die Gruppe der Kongruenzabbildungen der Ebene oder des Raumes (mit der Hintereinanderausführung als innerer Verknüpfung). Ist F eine Teilmenge der Ebene (oder des Raums), so betrachtet man die Menge G_F aller Kongruenzabbildungen g, die F auf sich abbilden (d.h. $g(F) = F$). Sie bilden die Symmetriegruppe $(G_F, \circ)$ von F.

Bsp.:

F allgemeines Dreieck (d.h. mit 3 verschiedenen Seitenlängen)
$\Rightarrow G_F = \{\mathrm{id}\}$

F gleichschenkliges (nicht gleichseitiges) Dreieck $\Rightarrow |G_F| = 2$

F gleichseitiges Dreieck $\Rightarrow |G_F| = 6$

F Kreis oder Kugel $\Rightarrow |G_F| = \infty$.

Wir beweisen nun einige wichtige und nicht ganz offensichtliche Folgerungen aus den Gruppenaxiomen. Es ist dabei wichtig, das Beweisprinzip zu verstehen, während man sich die einzelnen Beweisschritte und ihre Reihenfolge nicht zu merken braucht.

(2.2) Fakt. Sei $(G, \top)$ eine Gruppe. Dann gilt:

(a) Es gibt nur ein Element in G, für das (G_2) gilt (nämlich e). Für alle $a \in G$ gilt $a \top e = a$.

(b) Zu jedem $a \in G$ existiert nur ein $b \in G$, so daß $b \top a = e$ gilt. Für dieses b gilt auch $a \top b = e$.

Bez.: Zu $a \in G$ wird das (!) Element $b \in G$, für das $b \top a = e$ gilt, mit a^{-1} bezeichnet (bzw. mit $-a$ falls $\top := +$).

Bew.: Zeige zunächst:

$$(*) \qquad a, b \in G \text{ und } b \top a = e \Rightarrow a \top b = e$$

Das ist die 2. Behauptung in (2.2)(b).

Bew. von $(*)$: $(G_3) \Rightarrow \exists c \in G : c \top b = e$. Damit gilt:

$$\begin{aligned} a \top b &\overset{(G_2)}{=} e \top (a \top b) = (c \top b) \top (a \top b) \overset{!}{=} c \top (\underbrace{(b \top a)}_{=e} \top b) = c \top (e \top b) \\ &\overset{(G_2)}{=} c \top b = e \end{aligned}$$

Hierbei ergibt sich die Gleichheit $\overset{!}{=}$ wie folgt aus dem Assoziativgesetz (G_1):

$$(c \top b) \top \underbrace{(a \top b)}_{=:d \in G} \overset{(G_1)}{=} c \top (b \top \underbrace{(a \top b)}_{=d}) \overset{(G_1)}{=} c \top ((b \top a) \top b)$$

Bew. von (2.2)(a). Sei $\tilde{e} \in G$ ein Element, so daß für alle $a \in G$ gilt: $\tilde{e} \top a = a$. Diese Gleichung für $a := e$ ergibt $\tilde{e} \top e = e$. Nach $(*)$ folgt daraus $e \top \tilde{e} = e$. Andererseits gilt $e \top \tilde{e} = \tilde{e}$, da e linksneutral ist. Die beiden vorangehenden Gleichungen beweisen $e = \tilde{e}$. Damit ist die erste Behauptung von (2.2)(a) gezeigt, die zweite geht wie folgt:

Sei $a \in G$. Nach (G_3) existiert ein $b \in G$, so daß $b \top a = e$ gilt, und wegen $(*)$ gilt dann auch $a \top b = e$. Damit folgt:

$$a \top e = a \top (b \top a) \overset{(G_1)}{=} (a \top b) \top a = e \top a \overset{(G_2)}{=} a$$

Bew. von (2.2)(b). Seien $a, b, \tilde{b} \in G$, und es gelte $b \top a = e = \tilde{b} \top a$. Wir zeigen $b = \tilde{b}$:

$$\tilde{b} \overset{(a)}{=} \tilde{b} \top e \overset{(*)}{=} \tilde{b} \top (a \top b) \overset{(G_1)}{=} (\tilde{b} \top a) \top b = e \top b \overset{(G_2)}{=} b.$$

Die 2. Behauptung von (2.2)(b) ist gerade $(*)$.

(2.3) Fakt. Sei $(G, \top)$ eine Gruppe und $a \in G$, $b \in G$. Dann gilt:

(a) $(a^{-1})^{-1} = a$

(b) $(a \top b)^{-1} = b^{-1} \top a^{-1}$

Bew.:

(a) Nach Definition ist $(a^{-1})^{-1}$ das eindeutig bestimmte Element, für das $(a^{-1})^{-1} \top a^{-1} = e$ gilt. Wir zeigen, daß auch $a \top a^{-1} = e$ gilt, woraus $(a^{-1})^{-1} = a$ folgt. Nun gilt $a^{-1} \top a = e$ nach Definition von a^{-1}, und wegen (2.2)(b) folgt daraus $a \top a^{-1} = e$.

(b) Nach Definition ist $(a \top b)^{-1}$ das eindeutig bestimmte Element von G, für das $(a \top b)^{-1} \top (a \top b) = e$ gilt. Wir zeigen, daß $(b^{-1} \top a^{-1}) \top (a \top b) = e$ gilt, woraus $(a \top b)^{-1} = b^{-1} \top a^{-1}$ folgt.

$$\begin{aligned}(b^{-1} \top a^{-1}) \top (a \top b) &\overset{(G_1)}{=} b^{-1} \top (a^{-1} \top (a \top b)) \overset{(G_1)}{=} b^{-1} \top ((a^{-1} \top a) \top b) \\ &\overset{(G_2)}{=} b^{-1} \top (e \top b) \overset{(G_2)}{=} b^{-1} \top b = e\end{aligned}$$

Weitere Beispiele von Gruppen:

1) Die einelementige Gruppe, die nur das neutrale Element enthält, ist die einfachste ("triviale") Gruppe, d.h. $G = \{e\}$ mit $e \top e = e$.

2) Die zweielementige Gruppe $G = \{e, a\}$ mit $a \neq e$ und $e \top e = e$, $e \top a = a \top e = a$ und $a \top a = e$.

   Für endliche Gruppen G kann man die Abbildung $\top : G \times G \to G$ als "Gruppentafel" abspeichern.

   Im vorangehenden Fall:

$$\begin{array}{c|cc} \top & e & a \\ \hline e & e & a \\ a & a & e \end{array}$$

3) Sei $p \in \mathbb{N}$, $p \geq 2$, und $\mathbb{Z}_p$ wie am Ende von Kap. 1 definiert. Wir bemerken folgende Tatsache: Sind m, m', n, n' ganze Zahlen, so gilt

   (∗) $$[m]_p = [m']_p \text{ und } [n]_p = [n']_p \Rightarrow [m+n]_p = [m'+n']_p$$

   Bew.: $$\begin{aligned} [m]_p = [m']_p &\quad\Rightarrow\quad \exists j \in \mathbb{Z} : m' = m + jp \\ [n]_p = [n']_p &\quad\Rightarrow\quad \exists k \in \mathbb{Z} : n' = n + kp \end{aligned}$$

   Addition dieser Gleichungen liefert

$$m' + n' = m + n + (j+k)p, \text{ wobei } j + k \in \mathbb{Z},$$

   also $[m+n]_p = [m'+n']_p$, wie behauptet.

Wegen (∗) ist folgende Definition einer inneren Verknüpfung $+ : \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_p$ möglich:

Zu $a, b \in \mathbb{Z}_p$ wählen wir $m \in \mathbb{Z}$, $n \in \mathbb{Z}$ mit $a = [m]_p$, $b = [n]_p$ und definieren

(∗∗) $$a + b := [m+n]_p.$$

Da m durch a und n durch b nicht eindeutig bestimmt ist, muß man eigentlich befürchten, daß $[m+n]_p$ nicht nur von a und b abhängt, sondern auch von der Wahl von m und n (und in diesem Fall wäre (∗∗) zur Definition von $a+b$ nicht geeignet). Wegen (∗) sehen wir jedoch, daß diese Befürchtung grundlos ist, so daß (∗∗) $a+b$ eindeutig definiert.

Fakt: $(\mathbb{Z}_p, +)$ ist abelsche Gruppe.

Bew.: Wir zeigen (G_1). Seien $a, b, c \in G$. Wir wählen $m, n, l \in \mathbb{Z}$ mit $a = [m]_p$, $b = [n]_p$ und $c = [l]_p$. Dann gilt:

$$\begin{aligned} a + (b + c) &= a + [n + l]_p = [m + (n + l)]_p = [(m + n) + l]_p \\ &= [m + n]_p + c = (a + b) + c \end{aligned}$$

Das Assoziativgesetz für $(\mathbb{Z}_p, +)$ folgt also aus dem Assoziativgesetz für $(\mathbb{Z}, +)$. Ähnlich sieht man, daß $[0]_p$ neutrales Element in $(\mathbb{Z}_p, +)$ ist und daß $[-m]_p$ das zu $[m]_p$ inverse Element ist.

Bem.:

1) Für $p = 2$ gilt $|\mathbb{Z}_p| = 2$ und $\mathbb{Z}_2$ hat die gleiche Verknüpfungstafel wie die 2-elementige Gruppe $(G, \top)$ aus Bsp. 1.

$+$	$[0]_2$	$[1]_2$
$[0]_2$	$[0]_2$	$[1]_2$
$[1]_2$	$[1]_2$	$[0]_2$

Wir werden das später dadurch ausdrücken, daß wir $(\mathbb{Z}_2, +)$ und $(G, \top)$ "isomorph" nennen.

2) Wir betrachten die Menge G aller Drehungen um Winkel $\frac{m}{p} \cdot 360°$, $m \in \mathbb{Z}$, um einen festen Punkt der Ebene. Sind $m, m' \in \mathbb{Z}$, so sind die Drehungen um die Winkel $\frac{m}{p} \cdot 360°$ und $\frac{m'}{p} \cdot 360°$ genau dann gleich, wenn $[m]_p = [m']_p$ gilt. Die Menge G dieser Drehungen mit der Hintereinanderausführung $\circ$ als innerer Verknüpfung ist eine Gruppe, die isomorph zu $(\mathbb{Z}_p, +)$ ist. Die Hintereinanderausführung der Drehungen entspricht gerade der Addition der Drehwinkel. Die Gruppe $(G, \circ)$ ist eine geometrische Version der Gruppe $(\mathbb{Z}_p, +)$.

(2.4) Def.: Ein Körper ist eine Menge, genannt K, und zwei innere Verknüpfungen $+ : K \times K \to K$, $(a, b) \mapsto a + b$, und $\cdot : K \times K \to K$, $(a, b) \mapsto a \cdot b =: ab$, so daß gilt

- (K_1) $(K, +)$ ist eine abelsche Gruppe mit neutralem Element $e =: 0$
- (K_2) $(K \setminus \{0\}, \cdot)$ ist eine abelsche Gruppe mit neutralem Element $e =: 1$ (speziell $1 \neq 0$)
- (K_3) $\forall a, b, c \in K : (a + b)c = (ac) + (bc)$ (Distributivgesetz)

Schreibweisen: $a + (-b) =: a - b$, $(ab) + c =: ab + c$, "Punkt vor Strich". Ist $a \in K \setminus \{0\}$, $b \in K$, so $\frac{1}{a} := a^{-1}$, $\frac{b}{a} := ba^{-1} = a^{-1}b$.

(2.3)(a) $\Rightarrow -(-a) = a$
(2.3)(b) $\Rightarrow -(a + b) = (-b) + (-a) = -a - b$

Beispiel: $(\mathbb{R}, +, \cdot), (\mathbb{Q}, +, \cdot)$ sind Körper.
$(\mathbb{Z}, +, \cdot)$ ist kein Körper (z.B. besitzt 2 in $\mathbb{Z}$ kein multiplikatives Inverses).

Bem.:

1) Verzichtet man in (K_2) auf die Bedingung "abelsch" und ergänzt (K_3) durch $c(a+b) = (ca)+(cb)$, so nennt man $(K, +, \cdot)$ einen Schiefkörper.

2) Verzichtet man in (K_2) auf die Existenz von multiplikativen Inversen, so nennt man $(K, +, \cdot)$ einen kommutativen Ring mit 1.

3) Verzichtet man bei den Forderungen an einen Schiefkörper auf die Existenz von multiplikativen Inversen, so nennt man ein solches $(K, +, \cdot)$ einen Ring mit 1.

Beispiel: $(\mathbb{Z}, +, \cdot)$ ist ein kommutativer Ring mit 1.

(2.5) Rechenregeln. Sei $(K, +, \cdot)$ ein Körper (es genügt: ein Ring) und seien $a, b, c \in K$. Dann gilt:

(a) $0 \cdot a = 0$

(b) $a \cdot (-b) = (-a) \cdot b = -(ab)(=: -ab)$

(c) $a(bc) = (ab)c$

Beweis:

(a): Es gilt $0 \cdot a + 0 \cdot a \overset{(K_3)}{=} (0 + 0)a = 0 \cdot a$. Addiert man zur linken und zur rechten Seite dieser Gleichung $-(0 \cdot a)$, so folgt $0 \cdot a = 0$.

(b): Um zu beweisen, daß $a \cdot (-b) = -(ab)$ gilt, müssen wir zeigen, daß $a \cdot (-b)$ das additive Inverse von ab ist:

$$a \cdot (-b) + ab = (-b) \cdot a + b \cdot a \overset{(K_3)}{=} ((-b) + b)a = 0 \cdot a \overset{(a)}{=} 0$$

(c): Nach (K_2) gilt das Assoziativgesetz für $a, b, c \in K \setminus \{0\}$. Gilt $a = 0$ oder $b = 0$ oder $c = 0$, so folgt aus (a): $a(bc) = 0 = (ab)c$.

Der Körper $\mathbb{C}$ der komplexen Zahlen (G. Cardano 1501-1576, C.F. Gauß 1777-1855)

Erweiterungen von Zahlbereichen waren wichtige Revolutionen in der Mathematikgeschichte, die aus dem Bedürfnis entsprangen, Gleichungen, die in dem bekannten Zahlbereich nicht lösbar waren, durch Einführung "neuer Zahlen" doch zu lösen.

So ist etwa im Bereich $\mathbb{N}$ der natürlichen Zahlen eine Gleichung $n+x = m$ nur lösbar, wenn $m \geq n$ ist, während es im Bereich der ganzen Zahlen stets die Lösung $m - n$ gibt. In $\mathbb{Z}$ ist hingegen eine multiplikative Gleichung $n \cdot x = m$ mit $n \neq 0$ nur lösbar, wenn m von n geteilt wird. In den rationalen Zahlen existiert die Lösung $x = \frac{m}{n}$ immer. Für die antike griechische Mathematik war es eine Katastrophe, daß die Gleichung $x^2 = 2$ in $\mathbb{Q}$ keine Lösung besitzt, d.h. daß die Länge $\sqrt{2}$ der Diagonalen in einem Quadrat der Seitenlänge 1 als "Zahl" im damaligen Sinn nicht existierte. In den reellen Zahlen $\mathbb{R}$ hingegen existieren die Lösungen $\pm\sqrt{2}$ von $x^2 = 2$, und der Zwischenwertsatz (siehe Analysis I) zeigt, daß viele andere Gleichungen in $\mathbb{R}$ eine Lösung besitzen, während das in $\mathbb{Q}$ nicht der Fall zu sein braucht. Der Körper $\mathbb{C}$ der komplexen Zahlen erweitert die reellen Zahlen so, daß beliebige "polynomiale Gleichungen" (vom Grad ≥ 1) eine Lösung besitzen. Überraschenderweise genügt es, dafür zu sorgen, daß die Gleichung

$$x^2 + 1 = 0,$$

die in $\mathbb{R}$ natürlich keine Lösung hat, in der Erweiterung eine Lösung, die später die "imaginäre Einheit i" genannt wird, besitzt. Die historische Entwicklung war, daß eine "imaginäre" Rechengröße i eingeführt wurde, für die $i^2 = -1$ gelten sollte, und daß dann mit Ausdrücken der Form $a + ib$ mit $a, b \in \mathbb{R}$ nach den Regeln der Arithmetik (d.h. nach den Körperaxiomen) gerechnet wurde. Für die Multiplikation solcher Ausdrücke ergibt sich demnach

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

Im Folgenden wird ein vom Geheimnis um das Imaginäre befreiter "reeller" Zugang zu den komplexen Zahlen vorgestellt, der auf C.F. Gauß zurückgeht:

Die Gaußsche Zahlenebene

Wir versuchen die Menge $K := \mathbb{R}^2$ durch folgende innere Verknüpfungen $+, \cdot$ zu einem Körper zu machen. Für $(a,b) \in \mathbb{R}^2$, $(c,d) \in \mathbb{R}^2$ definieren wir:

$$\begin{aligned}(a,b)+(c,d) &:= (a+c, b+d)\\ (a,b)\cdot(c,d) &:= (ac-bd, ad+bc).\end{aligned}$$

Dabei wird - nachträglich - $(*)$ als Motivation für die Definition der Multiplikation dienen können. Es folgt leicht aus den entsprechenden Eigenschaften von $(\mathbb{R},+)$, daß $(\mathbb{R}^2,+)$ eine abelsche Gruppe ist, d.h. daß (K_1) gilt. Das neutrale Element von $(\mathbb{R}^2,+)$ ist $(0,0) \in \mathbb{R}^2$, das Inverse von $(a,b) \in \mathbb{R}^2$ ist $(-a,-b) =: -(a,b)$.

Einfache Rechnungen (Übungen!) zeigen, daß die so auf $\mathbb{R}^2$ definierte Multiplikation kommutativ und assoziativ ist. Außerdem gilt für alle $(c,d) \in \mathbb{R}^2$

$$(1,0)\cdot(c,d) = (c,d),$$

d.h. $(1,0)$ ist neutrales Element bezüglich der Multiplikation. Ist $(a,b) \in \mathbb{R}^2$ nicht das neutrale Element von $(\mathbb{R}^2,+)$, d.h. gilt $(a,b) \neq (0,0)$, so ist $\left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2}\right)$ multiplikatives Inverses zu (a,b). Es gilt nämlich

$$\begin{aligned}\left(\tfrac{a}{a^2+b^2}, \tfrac{-b}{a^2+b^2}\right)\cdot(a,b) &= \left(\tfrac{a^2}{a^2+b^2} - \tfrac{-b}{a^2+b^2}\cdot b, \tfrac{ab}{a^2+b^2} - \tfrac{ba}{a^2+b^2}\right)\\ &= (1,0).\end{aligned}$$

Um (K_2) zu zeigen, d.h. um zu zeigen, daß $(\mathbb{R}^2 \setminus \{(0,0)\}, \cdot)$ eine abelsche Gruppe ist, müssen wir uns noch überlegen, daß "$\cdot$" eine innere Verknüpfung auf $\mathbb{R}^2 \setminus \{(0,0)\}$ ist, d.h. daß $(a,b)\cdot(c,d) \neq (0,0)$ ist, wenn $(a,b) \neq (0,0)$ und $(c,d) \neq (0,0)$ gilt. Wäre $(a,b)\cdot(c,d) = (0,0)$, obwohl $(a,b) \neq (0,0)$ und $(c,d) \neq (0,0)$, so folgte

$$\left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2}\right)\cdot((a,b)\cdot(c,d)) = (0,0).$$

Wegen der Assoziativität von "$\cdot$" folgte daraus

$$(1,0)\cdot(c,d) = (0,0),$$

und damit $(c,d) = (0,0)$, im Widerspruch zur Voraussetzung "$(c,d) \neq (0,0)$". Also muß $(a,b)\cdot(c,d) \neq (0,0)$ gelten. Das beendet den Beweis von (K_2). Das

Distributivgesetz (K_3) ergibt sich durch eine leichte Rechnung:

$$\begin{aligned}(a,b)\cdot((a',b')+(a'',b'')) &= (a,b)\cdot(a'+a'',b'+b'')\\ &= (a(a'+a'')-b(b'+b''),a(b'+b'')+b(a'+a''))\\ &= (aa'-bb',ab'+ab')+(aa''-bb'',ab''+a''b)\\ &= (a,b)\cdot(a',b')+(a,b)\cdot(a'',b'').\end{aligned}$$

Insgesamt haben wir gezeigt, daß $(\mathbb{R}^2,+,\cdot)$ ein Körper ist. Dabei ist $(0,0)$ das neutrale Element bezüglich "+" und $(1,0)$ das neutrale Element bezüglich "$\cdot$". Es gilt

$$(\alpha)\qquad (0,1)\cdot(0,1)=(-1,0)=:-(1,0),$$

d.h. $x=(0,1)\in\mathbb{R}^2$ löst die Gleichung

$$x^2=-(1,0),$$

wobei $(1,0)$ das neutrale Element der Multiplikation ist. Nun ist $(\mathbb{R}^2,+,\cdot)$ im Grunde der Körper der komplexen Zahlen, jedoch haben sich historisch andere Bezeichnungen für seine Elemente durchgesetzt, die auch beim Rechnen praktisch sind. Diese Bezeichnungen sind

$$\begin{aligned}&i:=(0,1)\\ &a:=(a,0)\text{ falls }a\in\mathbb{R}.\end{aligned}$$

Damit schreibt sich $(a,b)\in\mathbb{R}^2$ als

$$(a,b)=(a,0)+(0,1)\cdot(b,0)=:a+ib,$$

wobei in der ersten Gleichheit benutzt wird, daß

$$(\beta)\qquad (0,1)\cdot(b,0)=(0,b)$$

gilt. Schließlich folgt aus

$$(\gamma)\qquad (a,0)\cdot(b,0)=(ab,0)\text{ für alle }a,b\in\mathbb{R},$$

daß bei der Identifikation $(a,0)=:a$ die Multiplikation der Paare $(a,0)$ und $(b,0)$ der Multiplikation der reellen Zahlen a und b entspricht. Gleiches gilt wegen $(a,0)+(b,0)=(a+b,0)$ für die Addition. Schließlich sind die Abkürzungen $(0,0)=:0$ und $(1,0)=:1$ damit konsistent, daß dann 0 und 1 das additive bzw. multiplikative Neutralelement des Körpers sind. Schreibt man (a, b) als $a + ib$, so wird aus der Definition des Produkts $(a, b) \cdot (c, d)$ gerade die Gleichung $(*)$. Wir setzen

$$\mathbb{C} := \{a + ib \mid a \in \mathbb{R}, b \in \mathbb{R}\}.$$

Das Vorangehende zeigt, daß $\mathbb{C}$ mit den Verknüpfungen

$$(a + ib) + (c + id) := (a + c) + i(b + d) \tag{2.6}$$

$$(a + ib) \cdot (c + id) := ac - bd + i(ad + bc) \tag{2.7}$$

ein Körper ist, der $\mathbb{R}$ als "Unterkörper" enthält. Das Neutralelement bezüglich "+" ist $0(= 0 + i \cdot 0) \in \mathbb{R} \subseteq \mathbb{C}$ und das Neutralelement bezüglich "$\cdot$" ist $1(= 1 + i \cdot 0) \in \mathbb{R} \subseteq \mathbb{C}$. Die Gleichung (α) schreibt sich jetzt als

$$i^2 = -1. \tag{2.8}$$

Die Identifikation von $\mathbb{C}$ mit $\mathbb{R}^2$, $a + ib \leftrightarrow (a, b)$, ermöglicht eine Veranschaulichung komplexer Zahlen als Punkte der Koordinatenebene $\mathbb{R}^2$.

Es folgt jetzt noch etwas zum Rechnen mit komplexen Zahlen und damit zusammenhängenden Begriffen. Aus der Formel für das multiplikative Inverse von $(a, b) \neq (0, 0)$ ergibt sich: Ist $z = a + ib \in \mathbb{C} \setminus \{0\}$, $a, b \in \mathbb{R}$, so ist das multiplikative Inverse $z^{-1} \left(= \frac{1}{z}\right)$ von z gegeben durch:

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

Ist $z = a+ib$ mit $a, b \in \mathbb{R}$, so nennt man a den Realteil von z, $a =: \mathrm{Re}(z)$, und b den Imaginärteil von z, $b =: \mathrm{Im}(z)$. Damit gilt für alle $z \in \mathbb{C}$ die Identität

$$z = \mathrm{Re}(z) + i\mathrm{Im}(z).$$

Die Konjugation ist die durch

$$z \in \mathbb{C} \to \overline{z} := \mathrm{Re}(z) - i\mathrm{Im}(z) \in \mathbb{C}$$

gegebene Abbildung, und $\overline{z}$ heißt die zu z konjugierte komplexe Zahl. Man berechnet, daß $z \cdot \overline{z} = a^2 + b^2 \in \mathbb{R}_{\geq 0}$ gilt, und definiert den Betrag $|z| \in \mathbb{R}_{\geq 0}$ von $z \in \mathbb{C}$ durch

$$|z| = \sqrt{a^2 + b^2} = \sqrt{z \cdot \overline{z}}.$$

Damit gilt für $z \in \mathbb{C} \setminus \{0\}$:

$$z^{-1} = \frac{a - ib}{a^2 + b^2} = \frac{\overline{z}}{z \cdot \overline{z}} = \frac{\overline{z}}{|z|^2}. \tag{2.9}$$

In der Gaußschen Zahlenebene entspricht der Konjugation gerade die Spiegelung an der x-Achse. Der Betrag $|z| = \sqrt{a^2 + b^2}$ von $z = a + ib \in \mathbb{C}$ entspricht nach dem Satz von Pythagoras gerade dem Abstand des Punkts (a, b) vom 0-Punkt. Die wichtige Polardarstellung $z = re^{i\varphi}$ (wobei $r = |z|$) werden wir erst später in der Vorlesung behandeln.

Mit der sogenannten "Mitternachtsformel" ist leicht einzusehen, daß nicht nur die Gleichung $x^2 + 1 = 0$ in $\mathbb{C}$ eine Lösung besitzt, sondern daß jede quadratische Gleichung

$$ax^2 + bx + c = 0 \text{ mit } a \neq 0$$

in $\mathbb{C}$ eine Lösung besitzt. In Wahrheit gilt sogar folgender erstaunliche Satz, der vermutlich in der Analysis II, sicher in jeder Vorlesung über Funktionentheorie, bewiesen wird.

(2.10) Fundamentalsatz der Algebra. *Ist $n \geq 1$ und sind $a_0, \dots, a_n$ komplexe Zahlen mit $a_n \neq 0$, so existiert ein $z \in \mathbb{C}$, so daß*

$$a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0 = 0$$

gilt.

Es ist vermutlich eine Überraschung, daß es auch Körper K gibt, so daß K eine endliche Menge ist. Einige davon wollen wir jetzt kennenlernen:

Primkörper

Für $p \in \mathbb{N}$, $p \geq 2$, definieren wir eine Multiplikation auf

$$\mathbb{Z}_p = \{[0]_p, \dots, [p-1]_p\} = \{[m]_p \mid m \in \mathbb{Z}\}$$

durch: $a = [m]_p, b = [n]_p \to ab = [mn]_p$. Das ist unabhängig von der Wahl von $m \in a$, $n \in b$, vgl. Blatt 3, Aufgabe 3.

Es ist leicht nachzuprüfen, daß $(\mathbb{Z}_p, +, \cdot)$ ein kommutativer Ring mit $1 (= [1]_p)$ ist.

(2.11) Satz. *Ist $p \in \mathbb{N}$ Primzahl, so ist $(\mathbb{Z}_p, +, \cdot)$ ein Körper.*

Beweis: Zu zeigen ist die Existenz eines multiplikativen Inversen für jedes $a \in \mathbb{Z}_p \setminus \{0\}$.
$a \in \mathbb{Z}_p \setminus \{0\} \Rightarrow a = [m]_p$ für ein $m \in \{1, \ldots, p-1\}$. Wir zeigen: Die Elemente $[1]_p \cdot [m]_p, [2]_p \cdot [m]_p, \ldots, [p-1]_p \cdot [m]_p$ in $\mathbb{Z}_p$ sind alle ungleich $[0]_p$.
Denn $[j]_p[m]_p = [0]_p$ ist äquivalent dazu, daß p die Zahl jm teilt. Da p Primzahl ist, teilt p dann j oder m (betrachte die Primfaktorzerlegungen von j, m und jm), d.h. $[j]_p = [0]_p$ oder $[m]_p = [0]_p$. Also gilt $[j]_p[m]_p \neq [0]_p$, falls $j \in \{1, \ldots, p-1\}$. Damit sind die Elemente $[1]_p \cdot [m]_p, \ldots, [p-1]_p \cdot [m]_p$ alle verschieden, denn aus $[i]_p[m]_p = [j]_p[m]_p$ folgt $[j-i]_p[m]_p = [0]_p$. Daraus folgt, daß

$$\{[1]_p \cdot [m]_p, \ldots, [p-1]_p \cdot [m]_p\} = \{[1]_p, \ldots, [p-1]_p\}$$

gilt. Speziell existiert ein $i \in \{1, \ldots, p-1\}$, so daß $[i]_p[m]_p = [1]_p$ gilt, d.h. $[i]_p$ ist multiplikatives Inverses von $a = [m]_p$.

Insbesondere gibt es den Körper $(\mathbb{Z}_2, +, \cdot)$ mit 2 Elementen, $\mathbb{Z}_2 = \{0, 1\}$.

Die Additionstabelle von $(\mathbb{Z}_2, +)$ findet sich auf S. 17, während für die Multiplikation gilt

$\cdot$	0	1
0	0	0
1	0	1

Speziell gilt also im Körper $(\mathbb{Z}_2, +, \cdot)$, im Gegensatz zum Fall der natürlichen Zahlen, $1 + 1 = 0$. Hier rächt sich, daß man die multiplikativen Neutralelemente in jedem Körper mit "1" bezeichnet, obwohl es sich dabei je nach Körper um ganz verschiedene Elemente handeln kann. Insofern ist der Widerspruch, daß (in $\mathbb{Z}_2$) $1 + 1 = 0$ gilt, während (in $\mathbb{N}$) $1 + 1 = 2 \neq 0$ gilt, nur ein scheinbarer, der auf unvollständige Notation zurückzuführen ist. Er verschwindet, wenn man (wie es richtig wäre, aber nicht üblich ist) die 1 im Körper K etwa als 1_K bezeichnet.

3 Vektorräume

(3.1) Def.: Sei K ein Körper. Ein Vektorraum über K ist eine abelsche Gruppe $(V, +)$ zusammen mit einer Abbildung $K \times V \to V$, $(\alpha, v) \in K \times V \mapsto \alpha v \in V$, so daß für alle $\alpha, \beta \in K$ und alle $v, w \in V$ gilt:

(V_1) $\alpha(\beta v) = (\alpha\beta)v \quad (=: \alpha\beta v)$

(V_2) $(\alpha + \beta)v = (\alpha v) + (\beta v) \quad (=: \alpha v + \beta v)$

(V_3) $\alpha(v + w) = (\alpha v) + (\alpha w) \quad (=: \alpha v + \alpha w)$

(V_4) $1v = v$ (, wobei 1 das Einselement von K bezeichnet.)

Sprechweisen und Bezeichnungen:
"Punkt vor Strich" $\alpha v + \beta w := (\alpha v) + (\beta w)$, außerdem: $v - w := v + (-w)$
Vektorraum über K = K-Vektorraum

Das neutrale Element von $(V, +)$ wird (zunächst) mit $\underline{0}$ bezeichnet, zur Unterscheidung vom neutralen Element 0 von $(K, +)$. Wir nennen $\underline{0}$ den Nullvektor von V.

Elemente von V werden "Vektoren" (des Vektorraums V) genannt, Elemente des Körpers K werden auch "Skalare" genannt.

Beispiele:

1) $K := \mathbb{R}$ und $V := \mathbb{R}^n$, $n \geq 1$, mit den Verknüpfungen

$$\begin{aligned} (x_1, \dots, x_n) + (y_1, \dots, y_n) &:= (x_1 + y_1, \dots, x_n + y_n) \\ r(x_1, \dots, x_n) &:= (rx_1 \dots, rx_n). \end{aligned}$$

   In diesem Fall $\underline{0} = (0, \dots, 0)$

2) Allgemeiner: Sei K beliebiger Körper, $n \in \mathbb{N}$, $n \geq 1$. Dann können wir genau wie in Beispiel 1) die Menge

$$K^n = \{(x_1, \dots, x_n) \mid x_1 \in K, \dots, x_n \in K\}$$

   zu einem Vektorraum über dem Körper K machen.

3) Der "Nullvektorraum". Für jeden Körper K besitzt die einelementige Menge $V = \{\underline{0}\}$ genau eine Struktur als K-Vektorraum.

4) Betrachtet man in 2) den Spezialfall $n = 1$, so sieht man, daß jeder Körper K als K-Vektorraum aufgefaßt werden kann, d.h. wir setzen $V := K$ und nehmen die Addition in K und die Multiplikation $K \times K \to K$ als Verknüpfungen in (3.1).

5) Sei $K := \mathbb{R}$, $M \neq \emptyset$ eine Menge und $V = \{f \mid f : M \to \mathbb{R}\}$.
Für $r \in \mathbb{R}$, $f, g \in V$ definieren wir $f + g \in V$ und $rf \in V$ durch

   (a) $\forall x \in M : (f + g)(x) := f(x) + g(x)$
   (b) $\forall x \in M : (rf)(x) := rf(x)$

   Mit diesen Verknüpfungen ist V ein $\mathbb{R}$-Vektorraum (,wobei $\underline{0} \in V$ die Abbildung ist, die jedes $x \in M$ auf $0 \in \mathbb{R}$ abbildet).

Wir beweisen exemplarisch, daß (V_3) gilt:
Seien $f, g \in V$, $r \in \mathbb{R}$. Zu zeigen ist:

$$r(f + g) = rf + rg \tag{$*$}$$

Auf beiden Seiten von $(*)$ stehen Abbildungen von M nach $\mathbb{R}$. Wir haben also zu zeigen, daß für alle $x \in M$ gilt:

$$(r(f + g))(x) = (rf + rg)(x)$$

Wendet man auf beide Seiten (a) und (b) an (auf der linken Seite zunächst (b) dann (a)), so erhält man für die linke Seite

$$(r(f + g)(x)) \overset{\text{(b)}}{=} r((f + g)(x)) \overset{\text{(a)}}{=} r(f(x) + g(x))$$

und für die rechte Seite

$$(rf + rg)(x) \overset{\text{(a)}}{=} (rf)(x) + (rg)(x) \overset{\text{(b)}}{=} rf(x) + rg(x)$$

Schließlich gilt $r(f(x) + g(x)) = rf(x) + rg(x)$ aufgrund des Distributivgesetzes für die reellen Zahlen.

Exkurs (zum Tag der offenen Tür):

Lineare Gleichungssysteme und das Gaußsche Eliminationsverfahren

Die meisten Probleme der linearen Algebra (und darüber hinaus die meisten numerischen Verfahren für nichtlineare Probleme) werden letztlich auf das Lösen linearer Gleichungssysteme zurückgeführt. Daß es dafür ein Rechenverfahren (einen Algorithmus) gibt, der stets zum Ziel führt und der noch nicht einmal sehr aufwändig ist, ist also eine wichtige Erkenntnis.

Da wir nur die “Grundrechenarten” benötigen werden, können wir lineare Gleichungssysteme in einem beliebigen Körper K betrachten und lösen. Man kann aber zunächst ruhig an $K = \mathbb{R}$ denken. Wir beginnen mit den einfachsten Fällen.

1) Seien $a, b \in K$. Wir suchen nach allen $x \in K$, so daß die Gleichung

$$(I) \qquad ax = b$$

gilt, d.h. wir wollen die Lösungsmenge

$$L_I = \{x \in K \mid ax = b\}$$

bestimmen. Es gibt 3 verschiedene Fälle:

a) Ist $a \neq 0$, so gilt $L_I = \left\{\frac{b}{a}\right\}$

b) Ist $a = 0$ und $b \neq 0$, so gilt $L_I = \emptyset$, vgl. (2.5)(a).

c) Ist $a = 0$ und $b = 0$, so gilt $L_I = K$, vgl. (2.5)(a).

2) Seien $a, b, c \in K$. Wir suchen nach der Lösungsmenge $L_I \subseteq K^2$ der Gleichung

$$(I) \qquad ax + by = c.$$

Es gibt folgende Fälle:

a) Ist $b \neq 0$, so gilt

$$\begin{aligned} L_I &= \left\{(x, y) \in K^2 \mid x \in K \text{ und } y = \tfrac{c}{b} - \tfrac{a}{b}x\right\} \\ &= \left\{(x, \tfrac{c}{b} - \tfrac{a}{b}x) \mid x \in K\right\} \\ &= \left\{(0, \tfrac{c}{b}) + x(1, -\tfrac{a}{b}) \mid x \in K\right\} \end{aligned}$$

b) Ist $b = 0$ und $a \neq 0$, so gilt

$$L_I = \left\{(\frac{c}{a}, y) \mid y \in K\right\} = \left\{(\frac{c}{a}, 0) + y(0, 1) \mid y \in K\right\}$$

c) Ist $a = b = 0$ und $c \neq 0$, so gilt $L_I = \emptyset$.

d) Ist $a = b = c = 0$, so gilt $L_I = K^2$

Bem.: Ist $K = \mathbb{R}$, so ist in den Fällen 2)a) und 2)b) die Lösungsmenge L_I eine Gerade in der Koordinatenebene $\mathbb{R}^2$ (und umgekehrt existiert zu jeder Geraden $G \subseteq \mathbb{R}^2$ eine lineare Gleichung $I : ax + by = c$, so daß $G = L_I$ gilt).

3) Zu $a, b, c, d, e, f \in K$ betrachten wir das Gleichungssystem

$$(I) \qquad \begin{array}{rl} ax + by = e & \quad I_1 \\ cx + dy = f & \quad I_2, \end{array}$$

das aus den beiden Gleichungen I_1 und I_2 besteht, die vom Typ der in 2) behandelten Gleichungen sind. Nach Definition ist $L_I = L_{I_1} \cap L_{I_2}$.

Wir unterscheiden folgende Fälle:

a) Ist $ad - bc \neq 0$, so ist I "eindeutig lösbar", und es gilt

$$L_I = \left\{ \left(\frac{-bf + de}{ad - bc}, \frac{af - ce}{ad - bc} \right) \right\}$$

Das kann man wie folgt einsehen. Ist etwa $a \neq 0$, so "eliminieren" wir mittels I_1, die "Unbekannte" x aus I_2, d.h. wir betrachten

$$I_2' := I_2 - \frac{c}{a} \cdot I_1 : 0 \cdot x + \left(d - \frac{c}{a} b \right) y = f - \frac{c}{a} e$$

Daraus ergibt sich: $y = \frac{af - ce}{ad - bc}$. Da nun y bekannt ist, können wir x aus I_1 berechnen. Mit einigen Umformungen ergibt sich

$$x = \frac{-bf + de}{ad - bc}$$

Die Formel

$$L_I = \left\{ \left(\frac{-bf + de}{ad - bc}, \frac{af - ce}{ad - bc} \right) \right\}$$

gilt stets, wenn $ad - bc \neq 0$ ist (und nicht nur, wenn zusätzlich $a \neq 0$, wie in obiger Rechnung vorausgesetzt). Das kann man z.B. durch Lösen des (einfachen) Gleichungssystems erkennen, für das $a = 0$ und $b \neq 0$ gilt.

Bem.: Im Fall $K = \mathbb{R}$ hat die Bedingung $ad - bc \neq 0$ zur Folge, daß die Lösungsmengen L_{I_1} und L_{I_2} nichtparallele Geraden in $\mathbb{R}^2$ sind. Die eindeutige Lösbarkeit von I entspricht geometrisch also der Tatsache, daß sich nichtparallele Geraden in $\mathbb{R}^2$ in genau einem Punkt schneiden.

b) Gilt $a = b = c = d = e = f = 0$, so folgt $L_I = K^2$.

c) Gilt $a = b = c = d = 0$ und $(e, f) \neq (0, 0)$, so folgt $L_I = \emptyset$.

d) Ist $ad - bc = 0$ und sind nicht alle a, b, c, d gleich 0, etwa $a \neq 0$, so folgt $(c, d) = \frac{c}{a}(a, b)$. Bildet man $I_2' = I_2 - \frac{c}{a} I_1$, so erhält man

$$I_2' \qquad 0 \cdot x + 0 \cdot y = f - \frac{c}{a} e$$

Ist $f - \frac{c}{a}e \neq 0$, so folgt $L_I = \emptyset$, da $L_I \subseteq L_{I_2'} = \emptyset$.
Ist $f - \frac{c}{a}e = 0$, so folgt $L_I = L_{I_1}$, und die Möglichkeiten für L_{I_1} wurden unter 2) aufgelistet.

Die Beispiele 1)-3) sollten klar machen, daß bei größeren linearen Gleichungssystemen (mehr Unbekannte und mehr Gleichungen) die Untersuchung der Lösungsmenge mittels Fallunterscheidungen sehr unübersichtlich würde und daß man deshalb das Problem, die Lösungsmengen zu verstehen, grundsätzlicher angehen muß. Wir werden sehen, daß die Theorie der Vektorräume einen geeigneten mathematischen Rahmen liefert, in dem die Lösungsmengen linearer Gleichungssysteme gut verstanden werden können. Zunächst soll aber erklärt werden, wie solche allgemeinen linearen Gleichungssysteme notiert werden und wie sie gelöst werden können. Ein lineares Gleichungssystem I mit m Gleichungen $I_1, \ldots, I_m$ für n Unbekannte $x_1, \ldots, x_n$ wird wie folgt notiert:

$$(I) \qquad \begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & +\ldots+ & a_{1n}x_n & = & b_1 & & I_1 \\ a_{21}x_1 & + & a_{22}x_2 & +\ldots+ & a_{2n}x_n & = & b_2 & & I_2 \\ \vdots & & \vdots & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & +\ldots+ & a_{mn}x_n & = & b_m & & I_m \end{array}$$

Dabei sind die "Koeffizienten" a_{ij}, für $1 \leq i \leq m$ und $1 \leq j \leq n$, und die "rechten Seiten" b_i, für $1 \leq i \leq m$, gegebene Körperelemente. Es wird eine explizite und möglichst einfache Beschreibung der Lösungsmenge

$$\begin{aligned} L_I &= \{(x_1, \ldots, x_n) \in K^n \mid (x_1, \ldots, x_n) \text{ erfüllt } I_1, I_2, \ldots, I_m\} \\ &= L_{I_1} \cap \ldots \cap L_{I_m} \end{aligned}$$

gesucht.

Das Gaußsche Eliminationsverfahren verwandelt durch (mehrfaches) Anwenden folgender *elementarer Umformungen* das Gleichungssystem I in ein anderes Gleichungssystem $\tilde{I}$, für das $L_I = L_{\tilde{I}}$ gilt und so daß $\tilde{I}$ leicht explizit gelöst werden kann.

Elementare Umformungen von I:

1) *Vertauschen von 2 Gleichungen*: Sei $1 \leq i < j \leq m$. Setze

$$\begin{aligned} I'_i &:= I_j \\ I'_j &:= I_i \\ I'_l &:= I_l \qquad \text{für } l \in \{1, \ldots, m\} \setminus \{i, j\}. \end{aligned}$$

   Offensichtlich gilt $L_{I'} = L_I$.

2) *Multiplikation einer der Gleichungen mit einem $c \in K \setminus \{0\}$*:
   Sei $i \in \{1, \ldots, m\}$ und $c \in K \setminus \{0\}$. Setze

$$\begin{aligned} I'_i &:= cI_i \\ I'_l &:= I_l \qquad \text{für } l \in \{1, \ldots, m\} \setminus \{i\}. \end{aligned}$$

   Dann gilt offensichtlich $L_I \subseteq L_{I'}$ und wegen $c \neq 0$ auch $L_{I'} \subseteq L_I$ (I entsteht aus I' durch Multiplikation der i'ten Gleichung mit c^{-1}).

3) *Addition einer Gleichung zu einer anderen*: Seien i, j in $\{1, \ldots, m\}$, $i \neq j$. Setze

$$\begin{aligned} I'_j &:= I_j + I_i \\ I'_l &:= I_l \qquad \text{für } l \in \{1, \ldots, m\} \setminus \{j\}. \end{aligned}$$

   Offensichtlich gilt $L_I \subseteq L_{I'}$. Umgekehrt erhält man I aus I', wenn man zunächst die i'te Gleichung von I' mit $(-1) \in K$ multipliziert und dann zur j'ten Gleichung von I' addiert. Daraus folgt $L_{I'} \subseteq L_I$.

Das Gaußsche Eliminationsverfahren besteht darin, sukzessive mittels der Umformungen 1)-3) die Unbekannten $x_1, \ldots, x_n$ aus einem Teil der Gleichungen zu eliminieren, und zwar so, daß ein Gleichungssystem in "Stufenform" entsteht. Bevor wir das allgemein formulieren, hier zwei Beispiele (im Fall $K := \mathbb{R}$):

1)

$$
(I) \quad \begin{array}{rrrcrl}
x_1 & + x_2 & + x_3 & = & 2 & \mid \cdot (-2) \quad \mid \cdot (-3) \\
2x_1 & + 4x_2 & + 3x_3 & = & -1 & \\
3x_1 & - x_2 & + 4x_3 & = & 7 &
\end{array}
$$

$$
(I') \quad \begin{array}{rrrcrl}
x_1 & + x_2 & + x_3 & = & 2 & \\
 & 2x_2 & + x_3 & = & -5 & \mid \cdot 2 \\
 & - 4x_2 & + x_3 & = & 1 &
\end{array}
$$

$$
(I'') \quad \begin{array}{rrrcr}
x_1 & + x_2 & + x_3 & = & 2 \\
 & 2x_2 & + x_3 & = & -5 \\
 & & 3x_3 & = & -9
\end{array}
$$

Das Gleichungssystem I is in Stufenform und kann rekursiv, beginnend mit der letzten Gleichung, gelöst werden:

$$
\begin{array}{rcl}
3x_3 = -9 & \Rightarrow & x_3 = -3 \\
x_3 = -3 \text{ und } 2x_2 + x_3 = -5 & \Rightarrow & x_2 = -1 \\
x_2 = -1 \text{ und } x_3 = -3 \text{ und } x_1 + x_2 + x_3 = 2 & \Rightarrow & x_1 = 6
\end{array}
$$

Also: $L_I = L_{I''} = \{(6, -1, -3)\}$.

2) Für beliebiges $a \in \mathbb{R}$ betrachten wir das Gleichungssystem

$$
(I_a) \quad \begin{array}{rrrrcrl}
x_1 & +x_2 & -x_3 & -x_4 & = & -2 & \mid \cdot(-3) \\
3x_1 & +3x_2 & -x_3 & -2x_4 & = & 6 & \\
-x_1 & -x_2 & +3x_3 & +2x_4 & = & a &
\end{array}
$$

$$
(I_a') \quad \begin{array}{rrrrcrl}
x_1 & +x_2 & -x_3 & -x_4 & = & -2 & \\
 & & 2x_3 & +x_4 & = & 12 & \mid \cdot(-1) \\
 & & 2x_3 & +x_4 & = & a-2 &
\end{array}
$$

$$
(I_a'') \quad \begin{array}{rrrrcr}
x_1 & +x_2 & -x_3 & -x_4 & = & -2 \\
 & & 2x_3 & +x_4 & = & 12 \\
 & & & 0 & = & a-14
\end{array}
$$

Das System I_a'' ist in Stufenform, wobei die "Stufenlänge" 2 ist, was daran liegt, daß "zufällig" bei der Elimination von x_1 auch x_2 eliminiert wird und bei der Elimination von x_3 auch x_4 eliminiert wird. Die letzte Gleichung von I_a'' ist als

$$0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4 = a - 14$$

zu interpretieren, so daß wir für $a \neq 14$

$$L_{I_a} = L_{I_a''} = \emptyset$$

erhalten. Für $a = 14$ gilt die letzte Gleichung von I_{14}'' für alle $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$, und wir können, um die 2. Gleichung in I_{14}'' zu lösen, $x_3 \in \mathbb{R}$ beliebig wählen und dazu $x_4 = 12 - 2x_3$ berechnen. Um die 1. Gleichung von I_{14}'' zu lösen, können wir zusätzlich $x_2 \in \mathbb{R}$ beliebig wählen und

$$x_1 = -2 - x_2 + x_3 + x_4 = 10 - x_2 - x_3$$

berechnen. Wir erhalten

$$
\begin{aligned}
L_{I_{14}} &= L_{I_{14}''} = \{(10 - x_2 - x_3, x_2, x_3, 12 - 2x_3) \mid x_2 \in \mathbb{R}, x_3 \in \mathbb{R}\} \\
&= \{(10, 0, 0, 12) + x_2(-1, 1, 0, 0) + x_3(-1, 0, 1, -2) \mid x_2 \in \mathbb{R}, x_3 \in \mathbb{R}\}.
\end{aligned}
$$

Wir hätten z.B. auch $x_1 \in \mathbb{R}$ und $x_4 \in \mathbb{R}$ beliebig wählen und daraus x_2 und x_3 berechnen können. Das ergäbe eine andere Darstellung der Lösungsmenge $L_{I_{14}}$.

Mit diesen Beispielen vor Augen wird nun das Gaußsche Eliminationsverfahren im allgemeinen Fall beschrieben.

Sei I ein lineares Gleichungssystem mit m Gleichungen $I_1, \ldots, I_m$ für n Unbekannte, siehe S. 30.

1. Schritt: Elimination von x_1 aus mindestens $m-1$ Gleichungen:

1. Fall: $a_{11} \neq 0$. Setze

$$\begin{array}{lcll} I_1' & := & \frac{1}{a_{11}} I_1 & \\ I_i' & := & I_i - \frac{a_{i1}}{a_{11}} I_1 & \qquad \text{für } 1 < i \leq m. \end{array}$$

Dann hat das System I' die Form

$$\begin{array}{rrcrcl} x_1 + & a_{12}' x_2 & + \ldots + & a_{1n}' x_n & = & b_1' \\ & a_{22}' x_2 & + \ldots + & a_{2n}' x_n & = & b_2' \\ & \vdots & & \vdots & & \vdots \\ & a_{m2}' x_2 & + \ldots + & a_{mn}' x_n & = & b_m' \end{array}$$

2. Fall: $a_{11} = 0$, aber $a_{i1} \neq 0$ für ein $i \in \{2, \ldots, m\}$. Vertausche I_1 und I_i und wende das Vorgehen aus dem 1. Fall an.

3. Fall: $a_{i1} = 0$ für alle $i \in \{1, \ldots, m\}$. In diesem Fall lassen wir I unverändert.

2. Schritt: Elimination von x_2.

Ist im 1. Schritt der 1. oder der 2. Fall eingetreten, so lassen wir jetzt (und in alle Zukunft) die 1. Gleichung unverändert und behandeln die 2. bis m'te Gleichung wie im 1. Schritt, jetzt aber für x_2 statt x_1. Ist im 1. Schritt der 3. Fall eingetreten, so wenden wir den 1. Schritt auf I an, jetzt aber für x_2 statt x_1. Wir iterieren das Verfahren und erhalten so nach höchstens n Schritten ein Gleichungssystem $\tilde{I}$ im Stufenform, das wir rekursiv, beginnend mit der letzten Gleichung, lösen können und für das $L_{\tilde{I}} = L_I$ gilt.

Damit ist der Exkurs zu Ende und wir machen mit den Vektorräumen weiter.

(3.2) Rechenregeln. Sei V ein K-Vektorraum. Dann gilt für alle $\alpha \in K$ und alle $v \in V$:

(a) $0v = \underline{0}$, $\alpha \underline{0} = \underline{0}$

(b) $\alpha v = \underline{0} \Rightarrow \alpha = 0$ oder $v = \underline{0}$

(c) $(-\alpha)v = \alpha(-v) = -(\alpha v) \quad (=: -\alpha v)$

Bew.:

(a)
$$\begin{aligned} 0v &= 0v + \underline{0} = 0v + (v + (-v)) \overset{(V_4)}{=} (0v + 1v) + (-v) \\ &\overset{(V_2)}{=} (0+1)v + (-v) = 1v + (-v) \overset{(V_4)}{=} v + (-v) = \underline{0} \\ \alpha\underline{0} &= \alpha\underline{0} + (\alpha\underline{0} + (-(\alpha\underline{0}))) \overset{(V_3)}{=} \alpha(\underline{0} + \underline{0}) + (-(\alpha\underline{0})) \\ &= \alpha\underline{0} + (-(\alpha\underline{0})) = \underline{0} \end{aligned}$$

(b) Sei $\alpha v = \underline{0}$ und $\alpha \neq 0$. Wegen (a) folgt daraus $\alpha^{-1}(\alpha v) = \alpha^{-1}\underline{0} = \underline{0}$, und daraus mit (V_1) und (V_4): $\underline{0} = (\alpha^{-1}\alpha)\, v = 1 \cdot v = v$

(c) $(-\alpha)v + \alpha v \overset{(V_2)}{=} ((-\alpha) + \alpha)v = 0v \overset{(a)}{=} \underline{0}$. Also ist $(-\alpha)v$ das additive Inverse von αv, d.h. $(-\alpha)v = -(\alpha v)$.
$\alpha(-v) + \alpha v \overset{(V_3)}{=} \alpha((-v) + v) = \alpha\underline{0} \overset{(a)}{=} \underline{0}$. Wie zuvor folgt daraus:
$\alpha(-v) = -(\alpha v)$.

(3.3) Def.: Sei V K-Vektorraum. Eine Teilmenge U von V heißt Untervektorraum von V, falls gilt:

(a) $U \neq \emptyset$

(b) $v, w \in U \Rightarrow v + w \in U$

(c) $\alpha \in K, v \in U \Rightarrow \alpha v \in U$

Bez.: Wir werden oft statt "Untervektorraum" kurz "Unterraum" schreiben.

Beispiele:

1) $\{\underline{0}\}$ und V sind Unterräume von V. Um einzusehen, daß $\{\underline{0}\}$ Unterraum ist, muß man insbesondere wissen, daß für alle $\alpha \in K$ gilt $\alpha\underline{0} = \underline{0}$, vgl. (3.2)(a).

2) Ist $0 < m < n$, so ist die Menge

$$K^m \times \{\underbrace{0, \ldots, 0)}_{n-m}\} = \{(x_1, \ldots, x_n) \in K^n \mid x_{m+1} = m_{m+2} = \qquad = x_n = 0\}$$

ein Untervektorraum des K-Vektorraums K^n.

3) Sei $M \neq \emptyset$ eine Menge und

$$V = \{f | f : M \to \mathbb{R}\}$$

der auf S. 26, Bsp. 5, definierte $\mathbb{R}$-Vektorraum. Ist A eine Teilmenge von M, so ist

$$U := \{f \in V | \text{ Für alle } a \in A \text{ gilt } f(a) = 0\}$$

ein Unterraum von V.

(3.4) Folgerung. Sei U ein Unterraum des K-Vektorraums V. Dann ist U mit den von V auf U eingeschränkten Verknüpfungen selbst ein K-Vektorraum.

Bew.: Die Bedingung (3.3)(b) besagt gerade, daß $+|U \times U$ eine innere Verknüpfung ist. Daß diese assoziativ und kommutativ ist, folgt direkt aus den entsprechenden Eigenschaften von $+$ auf $V \times V$. Ist $v \in U$, so gilt nach (3.3)(c) auch $(-1)v \in U$, und wegen (V_4) und (3.2)(c) gilt $(-1)v = -v \in U$, d.h. U enthält mit jedem Element auch dessen additives Inverses. Schließlich existiert nach (3.3)(a) ein Element $v \in U$. Dann folgt $-v \in U$ und wegen (3.3)(b) auch $v+(-v) = \underline{0} \in V$. Damit haben wir bewiesen, daß $(U, +|U \times U)$ eine abelsche Gruppe ist. Aus (3.3)(c) folgt, daß die Abbildung $K \times U \to U$, $(\alpha, v) \to \alpha v$, wirklich nur Bilder in U hat. Die Vektorraumaxiome $(V_1)-(V_4)$ für U folgten nun direkt aus den entsprechenden für V.

Wir nennen ein wie auf S. 30 notiertes lineares Gleichungssystem *homogen*, wenn die rechten Seiten $b_1, \ldots, b_m$ der m Gleichungen alle gleich 0 sind, d.h. wenn $b_1 = b_2 = \ldots = b_m = 0$ gilt.

(3.5) Satz. *Die Lösungsmenge $L_I \subseteq K^n$ eines homogenen linearen Gleichungssystems I für n Unbekannte ist ein Untervektorraum von K^n.*

Bew.:

1) Da I homogen ist, gilt $\underline{0} \in L_I$.

2) Sind $x = (x_1, \ldots, x_n)$ und $y = (y_1, \ldots, y_n)$ Elemente von L_I, so wollen wir zeigen, daß auch $x + y \in L_I$ gilt. Wegen $x \in L_I$ und $y \in L_I$ gilt für jedes $i \in \{1, \ldots, m\}$

$$a_{i1}x_1 + \ldots + a_{in}x_n = 0$$

und

$$a_{i1}y_1 + \ldots + a_{in}y_n = 0$$

Addition dieser Gleichungen liefert

$$a_{i1}(x_1 + y_1) + \ldots + a_{in}(x_n + y_n) = 0$$

Da das für jedes $i \in \{1, \ldots, m\}$ gilt, haben wir gezeigt, daß $x + y = (x_1 + y_1, \ldots, x_n + y_n) \in L_I$ gilt.

3) Analog folgt aus $\alpha \in K$ und $x = (x_1, \ldots, x_n) \in L_I$, daß $\alpha x = (\alpha x_1, \ldots, \alpha x_n) \in L_I$ gilt.

(3.6) Fakt. Ist V ein K-Vektorraum und ist $\mathcal{U}$ eine Menge von Unterräumen von V, so ist deren Durchschnitt

$$W := \bigcap_{U \in \mathcal{U}} U = \{v | \forall U \in \mathcal{U} : v \in U\}$$

ein Unterraum von V.

Bew.:

1) Nach (3.5) enthält jedes $U \in \mathcal{U}$ den 0-Vektor $\underline{0}$, also $\underline{0} \in W$.

2) Sind $v, w \in W$, so liegen v und w in jedem $U \in \mathcal{U}$. Da jedes $U \in \mathcal{U}$ ein Unterraum ist, folgt, daß $v + w \in U$ für jedes $U \in \mathcal{U}$ gilt, d.h. es folgt $v + w \in \bigcap_{U \in \mathcal{U}} U = W$.

3) Ebenso folgt aus $\alpha \in K$ und $w \in W$ auch $\alpha w \in W$.

(3.7) Def.: Sei V ein K-Vektorraum, $M \subseteq V$ und $\mathcal{U}_M = \{U | U$ Unterraum von V mit $M \subseteq U\}$. Dann heißt

$$\operatorname{span}(M) := \bigcap_{U \in \mathcal{U}_M} U$$

der <u>von</u> M <u>aufgespannte</u> (oder erzeugte) <u>Untervektorraum</u>. M heißt <u>Erzeugendensystem</u> von V, falls $\operatorname{span}(M) = V$ gilt. V heißt <u>endlich erzeugt</u>, falls es eine endliche Menge $M \subseteq V$ mit $\operatorname{span}(M) = V$ gibt.

Bem.: Wegen (3.6) ist $\operatorname{span}(M)$ wirklich ein Untervektorraum von V. Es folgt direkt aus der Def. (3.7), daß $\operatorname{span}(M)$ der (bzgl. der Inklusion von Mengen) kleinste Unterraum ist, der M enthält.

Bsp.:

1) $M = \emptyset$ oder $M = \{\underline{0}\} \Rightarrow \operatorname{span}(M) = \{\underline{0}\}$
2) $M \subseteq M' \subseteq V \Rightarrow \operatorname{span}(M) \subseteq \operatorname{span}(M')$.
3) M Unterraum von $V \Leftrightarrow \operatorname{span}(M) = M$
   $\operatorname{span}(\operatorname{span}(M)) = \operatorname{span}(M)$

Nach der eher abstrakten Definition von $\operatorname{span}(M)$ gibt der folgende Satz eine konstruktive Beschreibung der Elemente von $\operatorname{span}(M)$.

(3.8) Satz. *Sei V ein K-Vektorraum und $\emptyset \neq M \subseteq V$. Dann gilt*

$\operatorname{span}(M) = \{v | \exists k \in \mathbb{N}_{>0}, v_1, \ldots, v_k \in M, \alpha_1, \ldots, \alpha_k \in K : v = \alpha_1 v_1 + \ldots + \alpha_k v_k\}$.

Bez.: Der Ausdruck $\alpha_1 v_1 + \ldots + \alpha_k v_k$ heißt eine <u>Linearkombination der Vektoren</u> $v_1, \ldots, v_k$. Als Abkürzung werden wir oft das Summenzeichen $\sum$ verwenden,

$$\alpha_1 v_1 + \ldots + \alpha_k v_k =: \sum_{i=1}^{k} \alpha_i v_i.$$

Das Symbol "i" in der vorangehenden Summe ist eine Variable, die die Werte $1, 2, \ldots, k$ annimmt. Sie kann statt mit "i" mit jedem anderen Symbol, dessen Bedeutung noch nicht anderweitig festgelegt ist, bezeichnet werden, d.h. $\sum_{i=1}^{k} \alpha_i v_i = \sum_{r=1}^{k} \alpha_r v_r = \sum_{\gamma=1}^{k} \alpha_\gamma v_\gamma$.

Bew.: Die rechte Seite in (3.8) sei mit U_0 abgekürzt. Wegen $v = 1v$ gilt $M \subseteq U_0$.

1) Als erstes zeigen wir, daß $U_0 \subseteq \operatorname{span}(M)$ gilt. Sei $v \in U_0$, $v = \sum_{i=1}^{k} \alpha_i v_i$ mit $v_i \in M$ und $\alpha_i \in K$ für $1 \leq i \leq k$. Da $\operatorname{span}(M) \supseteq M$ ein Untervektorraum ist, folgt aus $v_i \in M$, $\alpha_i \in K$ mit (3.3)(c), daß $\alpha_i v_i \in \operatorname{span}(M)$ gilt, und dann mit Induktion aus (3.3)(b), daß $v = \sum_{i=1}^{k} \alpha_i v_i \in \operatorname{span}(M)$ gilt.

2) Wir zeigen, daß U_0 Untervektorraum ist. Daraus folgt wegen $M \subseteq U_0$, daß $U_0 \in \mathcal{U}_M$ gilt und damit $\operatorname{span}(M) \subseteq U_0$, vgl. Def. (3.7).
Wegen $M \neq \emptyset$ und $M \subseteq U_0$ gilt $U_0 \neq \emptyset$. Seien v und w in U_0, d.h. es existieren $v_1, \ldots, v_k$, $w_1, \ldots, w_l$ in M und $\alpha_1, \ldots, \alpha_k$, $\beta_1, \ldots, \beta_l$ in K, so daß $v = \sum_{i=1}^{k} \alpha_i v_i$ und $w = \sum_{j=1}^{l} \beta_j w_j$ gilt. Dann folgt

$$v + w = \alpha_1 v_1 + \ldots + \alpha_k v_k + \beta_1 w_1 + \ldots + \beta_l w_l \in U_0$$

Ebenso gilt für alle $\alpha \in K$

$$\alpha v = \alpha \left(\sum_{i=1}^{k} \alpha_i v_i \right) = \sum_{i=1}^{k} (\alpha \alpha_i) v_i \in U_0$$

Aus 1) und 2) folgt nun $\operatorname{span}(M) = U_0$, wie behauptet.

Bsp.:

1) $v \in \mathbb{R}^2 \setminus \{(0,0)\} \Rightarrow \operatorname{span}(\{v\})$ ist die "Gerade" $\{rv \mid r \in \mathbb{R}\}$

2) $v, w \in \mathbb{R}^2 \setminus \{(0,0)\}, w \notin \operatorname{span}(\{v\}) \Rightarrow \operatorname{span}(\{v, w\}) = \mathbb{R}^2$.
(Beweis später)

3) $e_1 := (1, 0, \ldots, 0) \in K^n$, $e_2 := (0, 1, 0, \ldots, 0) \in K^n, \ldots,$
$e_n := (0, \ldots, 0, 1) \in K^n \Rightarrow \operatorname{span}(\{e_1, \ldots, e_n\}) = K^n$

Ist nämlich $x = (x_1, \ldots, x_n) \in K^n$, so gilt $x = \sum\limits_{i=1}^{n} x_i e_i$:

$$\begin{aligned} \sum_{i=1}^{n} x_i e_i &= x_1(1,0,\ldots,0) + x_2(0,1,0,\ldots,0) + \ldots + x_n(0,\ldots,0,1) \\ &= (x_1,0,\ldots,0) + (0,x_2,0,\ldots,0) + \ldots + (0,\ldots,0,x_n) = (x_1,\ldots,x_n) \end{aligned}$$

Wir kommen nun zu einer sehr wichtigen Definition.

(3.9) Def.: Sei V ein K-Vektorraum, $k \in \mathbb{N}_{>0}$ und $v_1, \ldots, v_k$ Vektoren in V. Die Vektoren $v_1, \ldots, v_k$ heißen <u>linear unabhängig</u>, falls gilt:
Sind $\alpha_1 \in K, \ldots, \alpha_k \in K$ und gilt $\alpha_1 v_1 + \ldots + \alpha_k v_k = \underline{0}$, so folgt $\alpha_1 = \ldots = \alpha_k = 0$. Sonst heißen die Vektoren $v_1, \ldots, v_k$ <u>linear abhängig</u>. Eine Teilmenge M von V heißt <u>linear unabhängig</u>, falls gilt: Ist $k \in \mathbb{N}_{>0}$ und sind $v_1, \ldots, v_k$ verschiedene Vektoren in M, so sind die $v_1, \ldots, v_k$ linear unabhängig. Sonst heißt M <u>linear abhängig</u>.

Bem.:

1) $v_1, \ldots, v_k$ linear abhängig $\Leftrightarrow$ es existieren $\alpha_1 \in K, \ldots, \alpha_k \in K$, nicht alle $= 0$, mit $\sum\limits_{i=1}^{k} \alpha_i v_i = \underline{0}$.

2) Mit "die Vektoren $v_1, \ldots, v_k$" ist eine beliebige endliche Folge von Vektoren in V gemeint. Es kann dabei auch vorkommen, daß es ein Paar von Indizes $1 \leq i < j \leq k$ gibt, für das $v_i = v_j$ gilt. Dann sind die Vektoren $v_1, \ldots, v_k$ freilich linear abhängig, da dann $0v_1 + \ldots + 1v_i + \ldots + (-1)v_j + \ldots + 0v_n = 0$ und $1 \neq 0$ (und $(-1) \neq 0$) gilt.

3) Mit "$v_1, \ldots, v_k$ sind verschiedene Vektoren" ist gemeint, daß aus $1 \leq i < j \leq k$ folgt $v_i \neq v_j$. Man sagt dazu auch, die Vektoren $v_1, \ldots, v_k$ seien paarweise verschieden.

4) Eine Menge $M \subseteq V$ ist linear abhängig, wenn es ein $k \in \mathbb{N}_{>0}$ und verschiedene Vektoren $v_1, \ldots, v_k$ in M gibt, die linear abhängig sind.

5) Sei $M \subseteq M' \subseteq V$. Dann gilt:
Ist M linear abhängig, so ist auch M' linear abhängig. Ist M' linear unabhängig, so ist auch M linear unabhängig.

Bsp.:

1) Enthält die Menge $M \subseteq V$ den 0-Vektor $\underline{0}$, so ist M linar abhängig: Wähle $k = 1$, $v_1 := \underline{0}$, $\alpha_1 := 1$. Dann $1 \cdot v_1 = 1 \cdot \underline{0} = \underline{0}$, aber $1 \neq 0$.

2) Ist $v \in V$, so ist die 1-elementige Menge $M = \{v\}$ genau dann linear unabhängig, wenn $v \neq \underline{0}$ gilt.

3) Die Vektoren $e_1, \ldots, e_n$ in K^n sind linear unabhängig. Denn $\alpha_1 e_1 + \ldots + \alpha_n e_n = (\alpha_1, \ldots, \alpha_n)$, also

$$\alpha_1 e_1 + \ldots + \alpha_n e_n = \underline{0} = (0, \ldots, 0) \Leftrightarrow \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0.$$

4) Wir betrachten

$$V = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$$

mit der auf S. 27, Bsp. 5, definierten $\mathbb{R}$-Vektorraumstruktur. Dann sind die Funktionen $\sin : \mathbb{R} \to \mathbb{R}$, $\cos : \mathbb{R} \to \mathbb{R}$ linear unabhängig. Seien nämlich r, s reelle Zahlen, so daß $r \sin + s \cos$ die 0-Funktion ist, d.h. so daß für alle $x \in \mathbb{R}$ gilt

$$r \sin(x) + s \cos(x) = 0.$$

Wir benutzen diese Gleichung für die Fälle $x = 0$ und $x = \pi/2$ und erhalten $r = s = 0$.

5) Sind die Vektoren $(1, 1, 1), (1, 2, 2)$ und $(1, 2, 3)$ in $\mathbb{R}^3$ linear unabhängig? Es seien r, s, t reelle Zahlen, so daß

$$r(1, 1, 1) + s(1, 2, 2) + t(1, 2, 3) = \underline{0}$$

gilt. Die Frage ist, ob daraus $r = s = t = 0$ folgt. Betrachten wir die vorangehende Gleichung komponentenweise, so sehen wir, daß r, s, t

folgendes homogene lineare Gleichungssystem erfüllen

$$(I) \qquad \begin{array}{rrrcl} r & +s & +t & = & 0 \mid \cdot(-1) \\ r & +2s & +2t & = & 0 \\ r & +2s & +3t & = & 0 \end{array}$$

$$(I') \qquad \begin{array}{rrrcl} r & +s & +t & = & 0 \\ & s & +t & = & 0 \mid \cdot(-1) \\ & s & +2t & = & 0 \end{array}$$

$$(I'') \qquad \begin{array}{rrrcl} r & +s & +t & = & 0 \\ & s & +t & = & 0 \\ & & t & = & 0 \end{array}$$

Aus (I'') folgt, daß $r = s = t = 0$ gilt, d.h. die Vektoren $(1,1,1)$, $(1,2,2)$ und $(1,2,3)$ sind linear unabhängig.

(3.10) Fakt. Die Vektoren $v_1, \ldots, v_k$ seien linear unabhängig. Dann gilt:

$$\alpha_1 v_1 + \ldots + \alpha_k v_k = \beta_1 v_1 + \ldots + \beta_k v_k \Rightarrow \alpha_1 = \beta_1, \ldots, \alpha_k = \beta_k$$

Bew.: Die vorausgesetzte Gleichung läßt sich umformulieren in

$$(\alpha_1 - \beta_1)v_1 + \ldots + (\alpha_k - \beta_k)v_k = \underline{0}$$

Da die $v_1, \ldots, v_k$ als linear unabhängig vorausgesetzt sind, folgt $\alpha_1 - \beta_1 = 0, \ldots, \alpha_k - \beta_k = 0$, also $\alpha_1 = \beta_1, \ldots, \alpha_k = \beta_k$.

Nun definieren wir einen zentralen Begriff der linearen Algebra.

(3.11) Definition. Eine Teilmenge B eines K-Vektorraums V heißt eine <u>Basis von V</u>, falls B sowohl linear unabhängig als auch ein Erzeugendensystem von V ist.

Bsp.: $\{e_1, \ldots, e_n\}$ ist eine Basis von K^n.

Bem.: Es sei V ein K-Vektorraum und $B = \{v_1, \ldots, v_n\}$ eine n-elementige Teilmenge von V. Dann ist B genau dann eine Basis von V, wenn es zu jedem $v \in V$ eindeutig bestimmte Elemente $\alpha_1 \in K, \ldots, \alpha_n \in K$ gibt, so daß $v = \sum\limits_{i=1}^{n} \alpha_i v_i$ gilt.

Bew.: Ist B Basis, so folgt die Existenz einer Darstellung $v = \sum\limits_{i=1}^{n} \alpha_i v_i$ aus der Voraussetzung, daß B Erzeugendensystem von V ist, d.h. daß $\mathrm{span}(B) = V$ gilt. Die Eindeutigkeit der $\alpha_1, \ldots, \alpha_n$ folgt aus (3.10). Besitzt umgekehrt jedes $v \in V$ eine Darstellung $v = \sum\limits_{i=1}^{n} \alpha_i v_i$, so ist $B = \{v_1, \ldots, v_n\}$ offenbar ein Erzeugendensystem. Wendet man die Eindeutigkeit dieser Darstellung im Fall $v = \underline{0}$ an, so erhält man die lineare Unabhängigkeit der $v_1, \ldots, v_n$.

Als Vorbereitung auf den nächsten Satz beweisen wir folgenden Hilfssatz.

(3.12) Lemma. Sei V K-Vektorraum, $M \subseteq V$. Dann gilt:

(a) $v \in \mathrm{span}(M) \Rightarrow \mathrm{span}(M \cup \{v\}) = \mathrm{span}(M)$

(b) $v \in \mathrm{span}(M) \setminus M \Rightarrow M \cup \{v\}$ ist linear abhängig.

Bew.: (a) Mit den Beispielen 2) und 3) nach (3.7) folgt

$$\mathrm{span}(M) \subseteq \mathrm{span}(M \cup \{v\}) \subseteq \mathrm{span}(\mathrm{span}(M)) = \mathrm{span}(M),$$

also speziell $\mathrm{span}(M) = \mathrm{span}(M \cup \{v\})$.

(b) Ist $M = \emptyset$, so gilt $\mathrm{span}(M) = \{\underline{0}\}$ und $M \cup \{\underline{0}\} = \{\underline{0}\}$ ist linear abhängig.

Ist $v \in \mathrm{span}(M) \setminus M$ und $M \neq \emptyset$, so existieren nach (3.8) ein $k \in \mathbb{N}_{>0}$, $v_1, \ldots, v_k \in M$ und $\alpha_1, \ldots, \alpha_k \in K$, so daß

$$(*) \qquad v = \sum_{i=1}^{k} \alpha_i v_i$$

gilt. Wir können annehmen, daß für $1 \leq i < j \leq k$ die Vektoren v_i und v_j verschieden sind, indem wir, falls $v_i = v_j$ gilt, die beiden Summanden $\alpha_i v_i$

und $\alpha_j v_j$ zu einem Summanden $(\alpha_i + \alpha_j)v_i$ zusammenfassen. Da $v \notin M$ gilt, ist v von allen v_i, $1 \leq i \leq k$, verschieden. Formen wir $(*)$ um in die Gleichung

$$1v + \sum_{i=1}^{k}(-\alpha_i)v_i = \underline{0},$$

so erhalten wir eine Linearkombination aus verschiedenen Vektoren in $M \cup \{v\}$, die den 0-Vektor darstellt und deren Koeffizienten nicht alle 0 sind, d.h. $M \cup \{v\}$ ist linear abhängig.

Folgender Satz charakterisiert Basen von Vektorräumen als (bezüglich der Inklusion von Mengen) größte linear unabhängige Mengen und als kleinste Erzeugendensysteme.

(3.13) Satz. *Sei V ein K-Vektorraum. Folgende Aussagen über eine Teilmenge $M \subseteq V$ sind äquivalent.*

(a) *M ist Basis.*

(b) M ist linear unabhängig und jede echte Obermenge $M' \subseteq V$ von M ist linear abhängig.

(c) Es gilt span$(M) = V$ und für jede echte Teilmenge M'' von M gilt span$(M'') \neq V$.

Bem.: Hierbei bedeutet "echte Obermenge M' von M", daß $M \subseteq M'$ und $M \neq M'$ gilt. Analog ist "M'' echte Teilmenge von M", wenn $M'' \subseteq M$ und $M'' \neq M$ gilt.

Bew.: Zunächst überzeugt man sich, daß (3.13) im Fall $V = \{\underline{0}\}$ richtig ist. Dann erfüllt nur die leere Menge die Bedingungen (a), (b) oder (c). Wir können also annehmen, daß V nicht nur aus dem 0-Vektor besteht.

1) Wir zeigen (a) $\Rightarrow$ (b). Sei M Basis und $M' \subseteq V$ echte Obermenge von M. Wir wollen zeigen, daß M' linear abhängig ist. Sei $v \in M' \setminus M$. Da span$(M) = V$ gilt, folgt $v \in$ span(M). Dann impliziert (3.12)(b), daß $M \cup \{v\}$ - und damit auch $M' \supseteq M \cup \{v\}$ - linear abhängig ist.

2) Wir zeigen: (b) $\Rightarrow$ (c). Die Teilmenge M von V erfülle die Aussage (b). Wir zeigen zunächst, daß $\operatorname{span}(M) = V$ gilt. Da $M \subseteq \operatorname{span}(M)$ gilt, genügt es, $V \setminus M \subseteq \operatorname{span}(M)$ zu zeigen. Sei also $v \in V \setminus M$. Wegen (b) ist dann $M \cup \{v\}$ linear abhängig, so daß verschiedene $v_1, \ldots, v_k$ in M und $\alpha_1, \ldots, \alpha_k$, $\alpha \in K$ existieren, so daß

$$(*) \qquad \alpha v + \sum_{i=1}^{k} \alpha_i v_i = \underline{0}$$

gilt und so daß nicht $\alpha = \alpha_1 = \ldots = \alpha_k = 0$ gilt. Wäre $\alpha = 0$, so könnten also nicht alle α_i, $1 \leq i \leq k$, gleich 0 sein, und die aus $(*)$ folgende Gleichung

$$\sum_{i=1}^{k} \alpha_i v_i = \underline{0}$$

widerspräche der vorausgesetzen linearen Unabhängigkeit von M. Also gilt $\alpha \neq 0$, und wir erhalten aus $(*)$

$$v = \sum_{i=1}^{k} \left(\frac{-\alpha_i}{\alpha} \right) v_i$$

Nach (3.8) zeigt das, daß $v \in \operatorname{span}(M)$ ist. Das beweist $\operatorname{span}(M) = V$. Ist schließlich M'' eine echte Untermenge von M und $v \in M \setminus M''$, so folgt $v \notin \operatorname{span}(M'')$. Denn nach (3.12)(b) impliziert $v \in \operatorname{span}(M'')$, daß $M'' \cup \{v\} \subseteq M$ linear abhängig ist. Das beweist $\operatorname{span}(M'') \neq V$.

3) Wir zeigen: (c) $\Rightarrow$ (a). Die Teilmenge M von V erfülle die Aussage (c). Es ist zu zeien, daß M linear unabhängig ist. Wäre M linear abhängig, so existierten verschiedene Elemente $v_1, \ldots, v_k$ in M und $\alpha_1, \ldots, \alpha_k \in K$, nicht alle gleich 0, ohne Einschränkung $\alpha_1 \neq 0$, so daß

$$\sum_{i=1}^{k} \alpha_i v_i = \underline{0}$$

gilt. Wegen $\alpha_1 \neq 0$ können wir daraus

$$v_1 = \sum_{i=2}^{k} \left(\frac{-\alpha_i}{\alpha_1} \right) v_i$$

folgern, also $v_1 \in \operatorname{span}(M \setminus \{v_1\})$, vgl. (3.8). Dann folgt aus (3.12)(a), daß $\operatorname{span}(M \setminus \{v_1\}) = \operatorname{span}(M) = V$ gelten müßte, was der vorausgesetzten Minimalität des Erzeugendensystems M widerspricht. Also ist M linear unabhängig.

(3.14) Satz. *Jeder Vektorraum besitzt eine Basis.*

Wir zeigen sogar:

Zusatz: Ist $M_0 \subseteq V$ linear unabhängig, so existiert eine M_0 enthaltende Basis von V.

Zum Angewöhnen beweisen wir (3.14) zunächst für den Fall endlich erzeugter Vektorräume. Sei also $M \subseteq V$ eine endliche Menge mit $\operatorname{span}(M) = V$. Wenn $\operatorname{span}(M \setminus \{v\}) \neq V$ für jedes $v \in M$ gilt, so erfüllt M die Aussage (3.13)(c) und ist deshalb eine Basis. Andernfalls existiert ein $v \in M$, so daß die echte Teilmenge $M \setminus \{v\}$ von M ein Erzeugendensystem von V ist. Wir wenden nun die vorangehenden Argumente auf $M \setminus \{v\}$ statt M an. Da M endlich ist, erhalten wir nach endlich vielen Schritten eine Basis von V.

Auch im Fall allgemeiner Vektorräume ist der Beweis für die Existenz einer Basis nicht schwer. Er beruht allerdings auf dem sogenannten “Lemma von Zorn”, das man als Axiom, als Grundannahme der Mengenlehre ansehen kann. Dieses Lemma ist, wie wir gleich sehen werden, sehr nützlich, aber es macht eine reine Existenzaussage über die “Existenz eines maximalen Elements”. Es weist keinen Weg, ein solches Element auch nur approximativ zu bestimmen.

Nun zur Formulierung des Lemmas von Zorn: Sei X eine Menge und $\emptyset \neq \mathcal{K} \subseteq \mathcal{P}(X)$. $\mathcal{K}$ heißt <u>Kette</u>, falls für alle Elemente M_1, M_2 von $\mathcal{K}$ gilt: $M_1 \subseteq M_2$ oder $M_2 \subseteq M_1$.

(3.15) Lemma von Zorn. Sei X Menge und $\emptyset \neq \mathcal{M} \subseteq \mathcal{P}(X)$. Für jede Kette $\mathcal{K} \subseteq \mathcal{M}$ gelte

$$\bigcup_{M \in \mathcal{K}} M \in \mathcal{M}.$$

Dann existiert ein maximales Element in $\mathcal{M}$, d.h. ein $M \in \mathcal{M}$, so daß für alle $M' \in \mathcal{M}$ gilt: $M \subseteq M' \Rightarrow M = M'$.

Wir beweisen nun Satz (3.14), und zwar in der starken Form des Zusatzes.

Wir setzen

$$\mathcal{M} := \{M | M_0 \subseteq M \subseteq V, M \text{ linear unabhängig}\}.$$

Wir zeigen, daß für $\mathcal{M}$ die Voraussetzungen des Lemmas von Zorn erfüllt sind. Es gilt $\mathcal{M} \subseteq \mathcal{P}(V)$ und $\mathcal{M} \neq \emptyset$, da $M_0 \in \mathcal{M}$. Sei $\mathcal{K} \subseteq \mathcal{M}$ eine Kette. Wir wollen zeigen, daß $\bigcup\limits_{M \in \mathcal{K}} M \in \mathcal{M}$ gilt. Da $\mathcal{K} \neq \emptyset$ und $M_0 \subseteq M \subseteq V$ für jedes $M \in \mathcal{K}$ gilt, folgt $M_0 \subseteq \bigcup\limits_{M \in \mathcal{K}} M \subseteq V$. Es bleibt zu zeigen, daß $\bigcup\limits_{M \in \mathcal{K}} M$ linear unabhängig ist. Seien verschiedene Vektoren $v_1, \ldots, v_k$ in $\bigcup\limits_{M \in \mathcal{K}} M$ gegeben. Dann existieren Mengen $M_1, \ldots, M_k$ in $\mathcal{M}$ mit $v_1 \in M_1$, $v_2 \in M_2, \ldots, v_k \in M_k$. Da $\mathcal{K}$ eine Kette ist, existiert unter den k Mengen $M_1, \ldots, M_k$ eine größte, d.h. es existiert ein $j \in \{1, \ldots, k\}$, so daß für alle $i \in \{1, \ldots, k\}$ gilt:

$$M_i \subseteq M_j.$$

Dann liegen auch die Elemente $v_i \in M_i$ für alle $i \in \{1, \ldots, k\}$ in M_j. Da $M_j \in \mathcal{K} \subseteq \mathcal{M}$ linear unabhängig ist, sind auch die Vektoren $v_1, \ldots, v_k$ linear unabhängig. Damit haben wir gezeigt, daß $\bigcup\limits_{M \in \mathcal{K}} M$ linear unabhängig ist, und damit $\bigcup\limits_{M \in \mathcal{K}} M \in \mathcal{M}$. Das Lemma von Zorn (3.15) ist also auf unsere Menge $\mathcal{M} \subseteq \mathcal{P}(V)$ anwendbar und liefert uns eine größte linear unabhängige Menge $B \subseteq V$, die M_0 enthält. Nun folgt aus Satz (3.13), daß B eine Basis ist.

Bem.: Die Lösungsmenge eines homogenen, linearen Gleichungssystems (mit Koeffizienten und n Unbekannten im Körper K) ist nach Satz (3.5) ein Untervektorraum von K^n. Ein solches Gleichungssystem zu lösen, bedeutet gerade, eine Basis des Lösungsraums zu bestimmen. Das Gaußsche Eliminationsverfahren liefert eine solche Basis.

Bsp.: Für $K := \mathbb{R}$.

$$
(I) \qquad \begin{array}{rrrrcl}
x_1 & + x_2 & - x_3 & - x_4 & = & 0 \\
3x_1 & + 3x_2 & - x_3 & - 2x_4 & = & 0 \\
- x_1 & - x_2 & + 3x_3 & + 2x_4 & = & 0
\end{array}
\quad | \cdot (-3)
$$

$$
(I') \qquad \begin{array}{rrrrcl}
x_1 & + x_2 & - x_3 & - x_4 & = & 0 \\
 & & 2x_3 & + x_4 & = & 0 \quad | \cdot (-1) \\
 & & 2x_3 & + x_4 & = & 0
\end{array}
$$

$$
(I'') \qquad \begin{array}{rrrrcl}
x_1 & + x_2 & - x_3 & - x_4 & = & 0 \\
 & & 2x_3 & + x_4 & = & 0
\end{array}
$$

Um die Lösungsmenge von I'' zu bestimmen, kann man etwa $x_2 \in \mathbb{R}$, $x_3 \in \mathbb{R}$ beliebig wählen und daraus x_1 und x_4 berechnen:

$$
\begin{array}{rcl}
L_I = L_{I''} & = & \{(-x_2 - x_3, x_2, x_3, -2x_3) \mid x_2 \in \mathbb{R}, x_3 \in \mathbb{R}\} \\
 & = & \{x_2(-1,1,0,0) + x_3(-1,0,1,-2) \mid x_2 \in \mathbb{R}, x_3 \in \mathbb{R}\} \\
 & = & \operatorname{span}(\{(-1,1,0,0), (-1,0,1,-2)\})
\end{array}
$$

Dieses Verfahren liefert automatisch eine Basis des Lösungsraums, d.h. die Vektoren, die man erhält, sind nicht nur ein Erzeugendensystem von L_I, sondern auch linear unabhängig. Allerdings gibt es bisweilen Wahlmöglichkeiten, im obigen Beispiel kann man etwa, statt x_2 und x_3 frei zu wählen und x_1 und x_4 aus x_2 und x_3 auszurechnen, auch x_1 und x_4 frei wählen und dann x_2 und x_3 ausrechnen:

$$
\begin{array}{rcl}
x_3 & = & -\frac{1}{2}x_4 \\
x_2 & = & -x_1 + \frac{1}{2}x_4
\end{array}
$$

Das führt zur Darstellung

$$
\begin{array}{rcl}
L_I & = & \{x_1, -x_1 + \frac{1}{2}x_4, -\frac{1}{2}x_4, x_4) \mid x_1 \in \mathbb{R}, x_4 \in \mathbb{R}\} \\
 & = & \operatorname{span}(\{(1,-1,0,0), (0, \frac{1}{2}, -\frac{1}{2}, 1)\})
\end{array}
$$

Diese zwei verschiedenen Darstellungen von L_I resultieren aus der Konstruktion verschiedener Basen von L_I.

Übung: Weisen Sie direkt nach, daß

$$\text{span}(\{(-1,1,0,0),(-1,0,1,-2)\}) = \text{span}(\{(1,-1,0,0),(0,\frac{1}{2},-\frac{1}{2},1)\})$$

gilt.

In diesem Beispiel haben die zwei angegebenen Basen von L_I die gleiche Anzahl von Elementen (nämlich zwei). Man kann sich fragen, ob generell alle Basen eines endlich erzeugten Vektorraums die gleiche Anzahl von Elementen besitzen. Wir wollen uns als nächstes mit dem Beweis beschäftigen, daß das in der Tat so ist.

(3.16) Austauschlemma. Sei B eine Basis von V mit n Elementen $v_1, \ldots, v_n$. Seien $\alpha_1, \ldots, \alpha_n \in K$ und $w = \sum_{i=1}^{n} \alpha_i v_i$. Ist $j \in \{1, \ldots, n\}$ und gilt $\alpha_j \neq 0$, so ist

$$B' := (B \setminus \{v_j\}) \cup \{w\}$$

ebenfalls eine Basis von V.

Bew.: 1) Wir zeigen, daß $\text{span}(B') = V$ gilt. Wegen $\alpha_j \neq 0$ können wir die Gleichung $w = \sum_{i=1}^{n} \alpha_i v_i$ umformen in

$$v_j = \frac{1}{\alpha_j} w + \sum_{i=1, i\neq j}^{n} \left(\frac{-\alpha_i}{\alpha_j} \right) v_i.$$

Daraus folgt $v_j \in \text{span}(B')$ und daraus mit (3.12)(a)

$$\text{span}(B') = \text{span}(B' \cup \{v_j\}).$$

Nun gilt aber $B \subseteq B' \cup \{v_j\}$ und deshalb

$$V = \text{span}(B) \subseteq \text{span}(B' \cup \{v_j\}) (\subseteq V).$$

Aus den beiden vorangehenden Aussagen folgt $\text{span}(B') = V$, d.h. B' ist ein Erzeugendensystem von V.

2) Um die lineare Unabhängigkeit von B' zu zeigen, seien $\beta \in K$, $\beta_1 \in K, \ldots, \beta_{j-1} \in K$, $\beta_{j+1} \in K, \ldots, \beta_n \in K$ gegeben, so daß

$$\beta w + \sum\nolimits_{i=1, i\neq j}^{n} \beta_i v_i = \underline{0}$$

gilt. Wir wollen zeigen, daß daraus $\beta = \beta_1 = \ldots = \beta_{j-1} = \beta_{j+1} = \ldots = \beta_n = 0$ folgt. Mit $w = \sum\limits_{i=1}^{n} \alpha_i v_i$ können wir die vorangehende Gleichung umformen in

$$\beta\alpha_j v_j + \sum\nolimits_{i=1, i\neq j}^{n} (\beta\alpha_i + \beta_i) v_i = \underline{0}.$$

Da die $v_1, \ldots, v_n$ linear unabhängig sind, folgt $\beta\alpha_j = 0$ und $\beta\alpha_i + \beta_i = 0$ für alle $i \in \{1, \ldots, n\} \setminus \{i\}$. Wegen $\alpha_j \neq 0$ folgt daraus $\beta = 0$ und $\beta_i = 0$ für alle $i \in \{1, \ldots, n\} \setminus \{j\}$, wie behauptet.

(3.17) Austauschsatz von Steinitz. *Sei B eine Basis von V mit n Elementen $v_1, \ldots, v_n$, und seien $w_1, \ldots w_m$ linear unabhängige Vektoren in V. Dann gilt:*

(a) $m \leq n$

(b) *Es gibt $n-m$ Vektoren in B, bei geeigneter Numerierung $v_{m+1}, \ldots, v_n$, so daß $\{w_1, \ldots, w_m, v_{m+1}, \ldots, v_n\}$ eine Basis von V ist.*

Bew.: Wir beweisen (3.17) durch vollständige Induktion nach m. Im Fall $m = 0$ sind (a) und (b) offensichtlich richtig. Für den Induktionsschritt seien linear unabhängige $w_1, \ldots, w_m$ gegeben. Wir wenden die Induktionsvorausetzung auf die linear unabhängigen Vektoren $w_1, \ldots, w_{m-1}$ an. Wir erhalten aus (a), daß $m - 1 \leq n$ gilt. Wäre $m - 1 = n$, d.h. $n - m + 1 = 0$, so folgte aus (b), daß $\{w_1, \ldots, w_{m-1}\}$ schon eine Basis von V wäre. Da $w_1, \ldots, w_m$ linear unabhängig sind, widerspricht das der in (3.13) bewiesenen Tatsache, daß jede Basis eine <u>maximale</u> linear unabhängige Menge ist. Es kann also nicht $m-1 = n$ gelten, so daß aus $m-1 \leq n$ sogar $m \leq n$ folgt. Um (b) zu beweisen, benutzen wir das Austauschlemma (3.16). Nach der Induktionsvoraussetzung können wir annehmen, daß $\{w_1, \ldots, w_{m-1}, v_m, \ldots, v_n\}$ eine Basis von V ist. Also existieren $\alpha_1 \in K, \ldots, \alpha_n \in K$, so daß

$$w_m = \sum_{i=1}^{m-1} \alpha_i w_i + \sum_{j=m}^{n} \alpha_j v_j$$

gilt. Da $w_m \notin \operatorname{span}\{w_1, \ldots, w_{m-1}\}$ gilt, vgl. (3.12)(b), existiert ein $j \in \{m, \ldots, n\}$ mit $\alpha_j \neq 0$, o.E. $\alpha_m \neq 0$. Nach (3.16) ist dann

$$(\{w_1, \ldots, w_{m-1}, v_m, \ldots, v_n\} \setminus \{v_m\}) \cup \{w_m\} = \{w_1, \ldots, w_m, v_{m+1}, \ldots, v_n\}$$

eine Basis von V. Das beweist (b).

Der Beweis, der für die Existenz einer Basis in jedem endlich erzeugten Vektorraum gegeben wurde, zeigt, daß jedes endliche Erzeugendensystem eine Basis enthält. Zusammen mit (3.17) folgt, daß je zwei Basen eines endlich erzeugten Vektorraums die gleiche Anzahl von Elementen besitzen.

(3.18) Def.: Ist V endlich erzeugter Vektorraum, so heißt die Anzahl der Elemente einer ($\Rightarrow$ jeder) Basis die Dimension von V, bezeichnet als $\dim V$. Ist V nicht endlich erzeugt, so setzen wir $\dim V := \infty$.

Bsp.: 1) Der triviale Vektorraum $V = \{\underline{0}\}$ besitzt nur die leere Menge als Basis. Es gilt: $V = \{\underline{0}\} \Leftrightarrow \dim V = 0$.
2) Der K-Vektorraum K^n hat die Dimension n, da $\{e_1, \ldots, e_n\}$ eine Basis des K^n ist, vgl. S. 39, Bsp. 3) und S. 41, Bsp. 3).
3) Der Vektorraum der reellen Zahlenfolgen

$$V = \{f | f : \mathbb{N} \to \mathbb{R}\},$$

vgl. S. 27, Bsp. 5, ist unendlich-dimensional, da er die unendliche linear unabhängige Teilmenge

$$M = \{f_n | n \in \mathbb{N}\}$$

besitzt, wobei $f_n : \mathbb{N} \to \mathbb{R}$ durch $f_n(m) = \begin{cases} 1 & \text{falls } m = n \\ 0 & \text{falls } m \neq n \end{cases}$ definiert ist, vgl. Blatt 6, Aufgabe 1.

Bem.: Statt "endlich erzeugter Vektorraum" sagt man meist "endlich-dimensionaler Vektorraum".

(3.19) Folgerung. Sei V Vektorraum, $\dim V =: n < \infty$. Dann gilt:

(a) Ist $M \subseteq V$ linear unabhängig, so gilt $|M| \leq n$, und es gilt genau dann $|M| = n$, wenn M Basis von V ist.

(b) Ist M Erzeugendensystem von V, so gilt $|M| \geq n$, und es gilt genau dann $|M| = n$, wenn M Basis von V ist.

(c) Ist U Untervektorraum von V, so gilt $\dim U \leq \dim V$, und es gilt genau dann $\dim U = \dim V$, wenn $U = V$ gilt.

Bew.:

(a) Aus (3.17) folgt $|M| \leq n$. Gilt $|M| = n$, so ist M also eine maximale linear unabhängige Menge, und (3.13) zeigt, daß M eine Basis von V ist.

(b) Wir haben in (3.14) gezeigt, daß jedes endliche Erzeugendensystem eine Basis enthält. Das beweist $|M| \geq n$. Ein Erzeugendensystem M mit $|M| = n$ ist also minimal, so daß M nach (3.13) eine Basis von V ist.

(c) Eine Basis B von U besteht aus linear unabhängigen Vektoren, so daß aus (a) folgt: $\dim U = |B| \leq n = \dim V$. Gilt $\dim U = n$, so folgt aus (a), daß B auch Basis von V ist. Es gilt deshalb:

$$U = \operatorname{span}(B) = V.$$

Folgendes Beispiel zeigt, daß die vorangehenden Ergebnisse auch bei konkreten Rechnungen nützlich sein können. In Blatt 6, Aufgabe 3, ist zu zeigen, daß $B := \{(1,2),(3,2)\}$ eine Basis des $\mathbb{R}^2$ ist. Daß B ein Erzeugendensystem des $\mathbb{R}^2$ ist, kann man daran sehen, daß sich aus B die Vektoren

$$\begin{array}{rcl} e_1 = (1,0) & = & \frac{1}{2}\left((3,2)-(1,2)\right) \text{ und} \\ e_2 = (0,1) & = & \frac{3}{4}(1,2) - \frac{1}{4}(3,2) \end{array}$$

linear kombinieren lassen, die ein Erzeugendensystem (sogar eine Basis) des $\mathbb{R}^2$ bilden. Die lineare Unabhängigkeit von B folgt dann ohne Weiteres aus (3.19)(b), da $\dim \mathbb{R}^2 = 2$.

(3.20) Def.: Seien U_1, U_2 Unterräume eines Vektorraums V. Dann heißt der Unterraum

$$U_1 + U_2 := \operatorname{span}(U_1 \cup U_2)$$

die <u>Summe von</u> U_1 <u>und</u> U_2. Gilt $U_1 \cap U_2 = \{\underline{0}\}$, so nennt man $\operatorname{span}(U_1 \cup U_2)$ die <u>direkte Summe von</u> U_1 <u>und</u> U_2, bezeichnet durch

$$\operatorname{span}(U_1 \cup U_2) =: U_1 \oplus U_2.$$

Bsp.: 1) Sind U_1, U_2 verschiedene 1-dimensionale Unterräume des $\mathbb{R}^2$ (d.h. Geraden durch $(0,0)$), so gilt $U_1 \cap U_2 = \{(0,0)\}$ und $U_1 \oplus U_2 = \mathbb{R}^2$.
2) Sind U_1, U_2 verschiedene 2-dimensionale Unterräume des $\mathbb{R}^3$ (d.h. Ebenen durch $(0,0,0)$), so gilt $\dim(U_1 \cap U_2) = 1$ und $U_1 + U_2 = \mathbb{R}^3$.
3) Für $0 < m < n$ und $K^m \times \{0\} := \{x \in K^n | x_{m+1} = \ldots = x_n = 0\}$, $\{0\} \times K^{n-m} := \{x \in K^n | x_1 = \ldots = x_m = 0\}$ gilt

$$K^n = (K^m \times \{0\}) \oplus \left(\{0\} \times K^{n-m}\right).$$

(3.21) Fakt. Seien U_1 und U_2 Unterräume eines Vektorraums. Dann gilt:

(a) $U_1 + U_2 = \{u_1 + u_2 | u_1 \in U_1 \text{ und } u_2 \in U_2\}$

(b) Gilt $U_1 \cap U_2 = \{\underline{0}\}$, so existieren für jedes $v \in U_1 \oplus U_2$ genau zwei Vektoren $u_1 \in U_1$ und $u_2 \in U_2$, so daß $v = u_1 + u_2$ gilt.

Bew.:

(a) Wir kürzen $W := \{u_1 + u_2 | u_1 \in U_1 \text{ und } u_2 \in U_2\}$ ab. Da $U_1 + U_2$ ein Unterraum ist, der $U_1 \cup U_2$ enthält, gilt $W \subseteq U_1 + U_2$. Es ist leicht zu zeigen, daß W ein Unterraum ist. Da offensichtlich $U_1 \cup U_2 \subseteq W$ gilt und da $\operatorname{span}(U_1 \cup U_2)$ der kleinste Unterraum ist, der $U_1 \cup U_2$ enthält, folgt $U_1 + U_2 \subseteq W$.

(b) Gilt $v = u_1 + u_2 = u_1' + u_2'$ mit $u_1, u_1' \in U_1$ und $u_2, u_2' \in U_2$, so folgt

$$u_1 - u_1' = u_2 - u_2' \in U_1 \cap U_2 = \{\underline{0}\},$$

also $u_1 = u_1'$ und $u_2 = u_2'$.

(3.22) Lemma.

(a) Ist U Unterraum eines Vektorraums V, so existiert ein Unterraum W von V, so daß $V = U \oplus W$ gilt.

(b) Sind U_1, U_2 Unterräume von V mit $U_1 \cap U_2 = \{\underline{0}\}$, so gilt

$$\dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2.$$

Bem.: Ein Unterraum W wie in (a) heißt ein zu U komplementärer Unterraum. Im allgemeinen gibt es sehr viele zu U komplementäre Unterräume. Ist etwa U ein 1-dimensionaler Unterraum des $\mathbb{R}^2$, so ist jeder 1-dimensionale Unterraum $W \neq U$ komplementär zu U.

Bew.:

(a) Wir wählen eine Basis B' von U. Da B' linear unabhängig ist, folgt aus Satz (3.14) die Existenz einer B' enthaltenden Basis B von V. Wir setzen $W := \text{span}(B \setminus B')$. Dann gilt

$$U + W = \text{span}(\text{span}(B') \cup \text{span}(B \setminus B')) \supseteq \text{span}(B' \cup (B \setminus B')) = V,$$

also $U+W = V$. Um $U \cap W = \{\underline{0}\}$ zu zeigen, beweisen wir, daß jedes $v \in U \cap W$ der 0-Vektor ist. Ist $v \in U \cap W$, so existieren (o.E. verschiedene) $v_1, \ldots, v_k$ in B', $w_1, \ldots, w_l$ in $B \setminus B'$ und $\alpha_1, \ldots, \alpha_k$, $\beta_1, \ldots, \beta_l$ in K, so daß

$$v = \sum_{i=1}^{k} \alpha_i v_i = \sum_{j=1}^{l} \beta_j w_j$$

gilt. Daraus folgt

$$\underline{0} = \sum_{i=1}^{k} \alpha_i v_i + \sum_{j=1}^{l} (-\beta_j) w_j.$$

Da die $v_1, \ldots, v_k$, $w_1, \ldots, w_l$ verschiedene Elemente der Basis B sind, folgt $\alpha_1 = \ldots = \alpha_k = 0 = \beta_1 = \ldots = \beta_l$. Also $v = 0$.

(b) Wir wählen Basen B_1 von U_1, B_2 von U_2. Aus $U_1 \cap U_2 = \{\underline{0}\}$ folgt $B_1 \cap B_2 = \emptyset$. Wir werden zeigen, daß $B_1 \cup B_2$ eine Basis von $U_1 \oplus U_2$ ist, woraus die Behauptung folgt. Aus

$$\text{span}(B_1 \cup B_2) \supseteq U_1 \cup U_2$$

folgt

$$\text{span}(B_1 \cup B_2) = \text{span}(\text{span}(B_1 \cup B_2)) \supseteq \text{span}(U_1 \cup U_2) = U_1 \oplus U_2.$$

Also ist $B_1 \cup B_2$ Erzeugendensystem von $U_1 \oplus U_2$. Um zu zeigen, daß $B_1 \cup B_2$ linear unabhängig ist, sei

$$\underline{0} = \sum_{i=1}^{k} \alpha_i v_i + \sum_{j=1}^{l} \beta_j w_j$$

eine Darstellung des 0-Vektors als Linearkombination verschiedener Vektoren $v_1, \ldots, v_k$ in B_1 und $w_1, \ldots, w_l$ in B_2. Nun folgt aus (3.21)(b), daß $\sum\limits_{i=1}^{k} \alpha_i v_i = \underline{0} = \sum\limits_{j=1}^{l} \beta_j w_j$ gilt. Daraus folgt $\alpha_1 = \ldots = \alpha_k = 0 = \beta_1 = \ldots = \beta_l$, da B_1 und B_2 linear unabhängig sind.

Bem.: (3.22)(b) ist auch im Fall von unendlich-dimensionalen Vektorräumen richtig, wenn man $\infty + \infty := \infty$ setzt und $\infty + n := \infty$ für alle $n \in \mathbb{N}$.

(3.23) Dimensionssatz. *Seien U_1 und U_2 Untervektorräume eines endlich-dimensionalen Vektorraums V. Dann gilt:*

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Bew.: Wir setzen $U_{12} := U_1 \cap U_2$. Nach (3.22)(a) existieren Unterräume W_1, W_2 von V, so dass

$$U_1 = U_{12} \oplus W_1$$

und

$$U_2 = U_{12} \oplus W_2$$

gilt. Wir wollen

$$(*) \qquad U_1 + U_2 = (U_{12} \oplus W_1) \oplus W_2$$

zeigen. Als erstes sehen wir, dass die Summe aus $U_{12} \oplus W_1$ und W_2 direkt ist, da $(U_{12} \oplus W_1) \cap W_2 = U_1 \cap W_2 = U_1 \cap (U_2 \cap W_2) = U_{12} \cap W_2 = \{\underline{0}\}$ gilt. Wegen $U_{12} \oplus W_1 = U_1$ gilt $U_1 \subseteq (U_{12} \oplus W_1) \oplus W_2$ und analog gilt $U_2 = U_{12} \oplus W_2 \subseteq (U_{12} \oplus W_1) \oplus W_2$. Da $(U_{12} \oplus W_1) \oplus W_2$ ein Unterraum ist, folgt $U_1 + U_2 = \operatorname{span}(U_1 \cup U_2) \subseteq (U_{12} \oplus W_1) \oplus W_2$. Die Inklusion $(U_{12} \oplus W_1) \oplus W_2 \subseteq$

$U_1 + U_2$ ist klar, da alle Summanden U_{12}, W_1 und W_2 in $U_1 + U_2$ enthalten sind. Aus (3.22)(b) folgt

$$\dim U_1 = \dim U_{12} + \dim W_1$$
$$\dim U_2 = \dim U_{12} + \dim W_2$$

und aus $(*)$ und (3.22)(b) folgt

$$\dim(U_1 + U_2) = \dim U_{12} + \dim W_1 + \dim W_2.$$

Subtraktion der ersten beiden Gleichungen von der dritten ergibt die Behauptung.

Bez.: Ein Unterraum H eines endlich-dimensionalen Vektorraums V heisst Hyperebene in V, falls $\dim H = \dim V - 1$ gilt.

(3.24) Folgerung. Ist U Unterraum von V und H Hyperebene in V, so gilt entweder $\dim(U \cap H) = \dim U - 1$ oder $U \subseteq H$.

Bew.: Aus (3.23) folgt

$$(*) \qquad \dim(U \cap H) = \dim U + (\dim V - 1) - \dim(U + H)$$

1. Fall: $\dim(U + H) = \dim V$. Dann folgt aus $(*)$:

$$\dim(U \cap H) = \dim U - 1.$$

2. Fall: $\dim(U + H) < \dim V$. Dann folgt aus $\dim H = \dim V - 1$, dass $\dim(H) = \dim(U + H)$ gilt. Wegen $H \subseteq U + H$ und (3.19)(b) erhalten wir $U + H = H$, woraus $U \subseteq H$ folgt.

Wir wollen (3.24) benutzen, um eine Aussage über homogene lineare Gleichungssysteme zu beweisen, die man weniger begrifflich, sondern eher rechnerisch, auch aus dem Gaußschen Eliminationsverfahren erhalten kann.

(3.25) Lemma. Es seien $a_1, \dots, a_n$ Elemente eines Körpers K, die nicht alle gleich 0 sind. Dann ist die Lösungsmenge L_I der Gleichung

$$(I) \qquad a_1 x_1 + \cdots + a_n x_n = 0$$

eine Hyperebene in K^n.

Bew.: Wir können ein $j \in \{1, \dots, n\}$ finden, für das $a_j \neq 0$ gilt. Dann gilt:

$$L_I = \left\{ (x_1, \dots, x_n) \in K^n \mid x_j = \sum_{i=1, i \neq j}^{n} \left(-\frac{a_i}{a_j} \right) x_i \right\}$$

Setzt man für $i \in \{1, \dots, n\} \setminus \{j\}$

$$v_i = e_i - \frac{a_i}{a_j} e_j,$$

so gilt $L_I = \operatorname{span}\{v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n\}$. Es ist leicht zu zeigen, dass die $v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_n$ linear unabhängig sind. Daraus folgt: $\dim L_I = n-1$, wie behauptet.

(3.26) Satz. *Sei K ein Körper und*

$$(I) \qquad \begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n &= 0 \\ \\ a_{k1}x_1 + \dots + a_{kn}x_n &= 0 \end{aligned}$$

ein homogenes lineares Gleichungssystem mit $a_{ij} \in K$ für $1 \leq i \leq k$ und $1 \leq j \leq n$. Dann gilt

$$\dim L_I \geq n - k.$$

Beweis durch Induktion nach k: Sei $k = 1$, d.h. (I) besteht nur aus einer Gleichung $a_{11}x_1 + \dots + a_{1n}x_n = 0$.

1. Fall: $a_{11} = a_{12} = \dots = a_{1n} = 0$. Dann gilt $L_I = K^n$ und damit $\dim L_I = n \geq n - k = n - 1$.

2. Fall: Es existiert ein $j \in \{1, \dots, n\}$, für das $a_{1j} \neq 0$ gilt. Dann besagt (3.25):

$$\dim L_I = n - 1.$$

Für den Induktionsschritt sei (I) ein Gleichungssystem mit $k > 1$ Gleichungen. Wir wenden die Induktionsvoraussetzung auf das Gleichungssystem (I') an, das aus den ersten $k-1$ Gleichungen $I_1, \dots, I_{k-1}$ von (I) gebildet wird. Es gilt

$$L_I = L_{I'} \cap L_{I_k}.$$

Die Induktionsvoraussetzung besagt, dass $\dim L_{I'} \geq n-k+1$ gilt. Sind nicht alle Koeffizienten von I_k gleich null, so ist L_{I_k} nach (3.25) eine Hyperebene. Dann folgt aus (3.24), angewendet auf $U = L_{I'}$, $H = L_{I_k}$:

$$\dim L_I = \dim(L_{I'} \cap L_{I_k}) \geq \dim(L_{I'}) - 1 \geq n - k.$$

Sind alle Koeffizienten von I_k gleich null, so gilt $L_{I_k} = K^n$ und damit $L_I = L_{I'} \cap L_{I_k} = L_{I'}$, und die Induktionsvoraussetzung ergibt

$$\dim L_I = \dim L_{I'} \geq n - k + 1 \geq n - k.$$

(3.27) Def.: Eine Teilmenge A eines Vektorraums V heisst <u>affiner Unterraum von V</u>, falls es ein $v \in V$ und einen k-dimensionalen Untervektorraum U von V gibt, so dass

$$A = \{v + u \mid u \in U\} =: v + U$$

gilt.

Wichtige Bemerkung: Seien $v_0, v_1 \in V$ und U_0, U_1 Untervektorräume von V. Dann gilt:

$$v_0 + U_0 = v_1 + U_1 \Leftrightarrow U_0 = U_1 \text{ und } v_1 - v_0 \in U_0.$$

Ist $A = v + U$ affiner Unterraum, so nennt man den Untervektorraum U den <u>Richtungsraum</u> von A. Er ist nach der vorangehenden Bemerkung eindeutig durch A bestimmt. Wir definieren die Dimension eines affinen Unterraums $A = v + U$ durch $\dim A := \dim U$. Wir nennen einen 1-dimensionalen affinen Unterraum eine <u>affine Gerade</u>, einen 2-dimensionalen affinen Unterraum eine <u>affine Ebene</u> und - im Fall $\dim V = n < \infty$ - einen $(n-1)$-dimensionalen affinen Unterraum eine <u>affine Hyperebene in</u> V.

Einen affinen Unterraum $A = v + U$ sollte man sich als den um den Vektor v parallel verschobenen Untervektorraum U vorstellen. Ein affiner Unterraum A ist genau dann ein Untervektorraum, wenn $\underline{0} \in A$ gilt.

Affine Unterräume treten als Lösungsmengen von inhomogenen linearen Gleichungssystemen auf.

(3.28) Satz. *Sei I ein lineares Gleichungssystem mit k Gleichungen für n Unbekannte mit Koeffizienten in einem Körper K. Dann gilt entweder $L_I = \emptyset$ oder L_I ist ein affiner Unterraum von K^n der Dimension $\geq n-k$. Der Richtungsraum von L_I ist die Lösungsmenge des zu I gehörenden homogenen linearen Gleichungssystems.*

Bem.: Gilt $k < n$ und $L_I \neq \emptyset$, so gibt es also sicher mehr als nur eine Lösung von I.

Bew.: Sei $L_I \neq \emptyset$. Wähle $v \in L_I$. Wir bezeichnen den Lösungsraum des zu I gehörenden homogenen linearen Gleichungssystems mit U. Wir wollen $L_I = v + U$ zeigen. Ist $w \in L_I$, so erfüllt $w = (w_1, \ldots, w_n)$ für alle $i \in \{1, \ldots, k\}$ die Gleichung

$$a_{i1}w_1 + \cdots + a_{in}w_n = b_i,$$

die ebenfalls von $v = (v_1, \ldots, v_n)$ erfüllt wird:

$$a_{i1}v_1 + \cdots + a_{in}v_n = b_i.$$

Subtrahiert man die zweite Gleichung von der ersten, so sieht man, dass $w-v$ das homogene lineare Gleichungssystem erfüllt, d.h. $w - v \in U$. Daraus folgt $w \in v + U$ und damit $L_I \subseteq v + U$. Ist umgekehrt $u \in U$, so gilt für alle $i \in \{1, \ldots, k\}$

$$a_{i1}u_1 + \cdots + a_{in}u_n = 0.$$

Addiert man diese Gleichung zu der vorangehenden, so folgt $v + u \in L_I$. Das beweist $v + U \subseteq L_I$.

4 Lineare Abbildungen und Matrizen

Wie überall, so sind auch in der Mathematik nicht nur die einzelnen Objekte wichtig, sondern auch die Beziehungen zwischen ihnen, die hier meist durch Abbildungen beschrieben werden, die (in einem zu definierenden Sinn) die Struktur der Objekte erhalten, sogenannte Homomorphismen. Im folgenden seien V und W Vektorräume über demselben Körper K.

(4.1) Definition. Eine Abbildung $L : V \to W$ heißt linear (oder Vektorraum-homomorphismus, falls für alle $v, v_1, v_2 \in V$ und alle $\alpha \in K$ gilt:

$$\begin{array}{llcll} \text{(a)} & L(v_1 + v_2) & = & L(v_1) + L(v_2) & (L \text{ ist "additiv"}) \\ \text{(b)} & L(\alpha v) & = & \alpha L(v) & (L \text{ ist "homogen"}) \end{array}$$

Bez.: Statt "lineare Abbildung" ist auch "linearer Operator" üblich. Im Fall $W = K$ spricht man oft von einem "linearen Funktional" oder einer "Linearform" $L : V \to K$. Die Menge der linearen Abbildungen von V nach W bezeichnen wir mit

$$\mathrm{Hom}(V, W) := \{L \mid L : V \to W \text{ linear}\}.$$

Eine lineare Abbildung $L : V \to V$ heißt Endomorphismus, und wir setzen

$$\mathrm{End}(V) = \{L \mid L : V \to V \text{ linear}\}.$$

Ein Homomorphismus $L : V \to W$ heißt Isomorphismus, falls L bijektiv ist. Zwei K-Vektorräume V und W heißen isomorph, falls es einen Isomorphismus $L : V \to W$ gibt.

Beispiele:

0) $\mathrm{id}_V \in \mathrm{End}(V)$.
   Ist $L : V \to W$ die "0-Abbildung" definiert durch

   $$\forall v \in V : Lv = \underline{0} \in W,$$

   so gilt $L \in \mathrm{Hom}(V, W)$.

   Ist $U \subseteq V$ Unterraum, so ist die Inklusion $i : U \to V$, $\forall u \in U$: $i(u) := u$, eine lineare Abbildung.

1) Für $\alpha \in K$ sei $S_\alpha : V \to V$, $S_\alpha(v) := \alpha v$, die "Streckung" um α. Dann gilt: $S_\alpha \in \mathrm{End}(V)$. Ist $\alpha \neq 0$, so ist S_α Isomorphismus.

2) Wir betrachten $\mathbb{R}$ als $\mathbb{R}$-Vektorraum, vgl. Bsp. 4 nach (3.1). Dann besteht $\mathrm{End}(\mathbb{R})$ gerade aus den Abbildungen $L : \mathbb{R} \to \mathbb{R}$ der Form $x \in \mathbb{R} \mapsto mx \in \mathbb{R}$ für ein $m \in \mathbb{R}$. Analoges gilt für jeden Körper K anstelle von $\mathbb{R}$.

3) Zu reellen Zahlen $a < b$ ist $C^0([a,b],\mathbb{R}) = \{f \mid f : [a,b] \to \mathbb{R} \text{ stetig}\}$ ein $\mathbb{R}$-Vektorraum, vgl. Bsp. 5 nach (3.1). Wir definieren

$$L : C^0([a,b],\mathbb{R}) \to \mathbb{R}, \quad L(f) := \int\limits_a^b f(x)dx$$

   L ist linearer Operator von $C^0([a,b],\mathbb{R})$ nach $\mathbb{R}$ ("Lineares Funktional").

4) $C^1(\mathbb{R},\mathbb{R}) = \{f \mid f : \mathbb{R} \to \mathbb{R} \text{ ist differenierbar und } f' : \mathbb{R} \to \mathbb{R} \text{ ist stetig}\}$ $D : C^1(\mathbb{R},\mathbb{R}) \to C^0(\mathbb{R},\mathbb{R}), D(f) := f'$. D ist linearer Operator.

5) Ist $L \in \mathrm{Hom}(V,W)$ Isomorphismus, so ist die Umkehrabbildung $L^{-1} : W \to V$ ebenfalls linear und damit ein Isomorphismus von W auf V.

Wichtig ist folgender Zusammenhang mit der Analysis: Ganz grob gesprochen ist die Analysis die Kunst, nichtlineare Funktionen (in einer kleinen Umgebung eines festen Punktes x_0) durch lineare Funktionen zu approximieren und aus Eigenschaften der approximierenden linearen Funktion (, mit der man gut explizit rechnen kann,) auf Eigenschaften der ursprünglichen nichtlinearen Funktion zu schließen.

Im Fall von (nichtlinearen) Funktionen $f : \mathbb{R} \to \mathbb{R}$ und $x_0 \in \mathbb{R}$ ist die approximierende lineare Funktion gegeben durch $h \in \mathbb{R} \to f'(x_0)h \in \mathbb{R}$, und es gilt für den durch $f(x_0 + h) = f(x_0) + f'(x_0)h + R(h)$ definierten "Approximationsfehler" $R(h) : \lim\limits_{h\to 0} \frac{R(h)}{h} = 0$. Betrachtet man (z.B. in der Analysis II) Funktionen $f : \mathbb{R}^n \to \mathbb{R}^m$ und $x_0 \in \mathbb{R}^n$, so wird in der analogen Definition der Differenzierbarkeit aus der linearen Funktion $h \in \mathbb{R} \to f'(x_0)h \in \mathbb{R}$ eine lineare Abbildung $\mathrm{Df}(x_0) \in \mathrm{Hom}(\mathbb{R}^n,\mathbb{R}^m)$.
Noch allgemeiner betrachtet man in der Theorie der Differentialgleichungen (und in der Physik und in anderen Wissenschaften) nichtlineare Differentialoperatoren zwischen unendlich-dimensionalen Funktionenräumen, die durch lineare Operatoren zwischen solchen Funktionenräumen approximiert werden. Oft verwendet man in der Physik direkt diese linearisierten Operatoren (z.B. Wellengleichung, Wärmeleitungsgleichung).

Bez.: Ist $L \in \mathrm{Hom}(V,W)$, so heißt

$$\ker L = \{v \in V \mid L(v) = \underline{0} \in W\} = L^{-1}(\{\underline{0}\})$$

der Kern von L und

$$\operatorname{im} L := \{w \in W \mid \exists v \in V : L(v) = w\} = L(V)$$

das Bild von L.

(4.2) Fakt. Sei $L \in \operatorname{Hom}(V, W)$, U_1 Unterraum von V, U_2 Unterraum von W. Dann gilt:

(a) $L(U_1)$ ist Unterraum von W.

(b) $L^{-1}(U_2)$ ist Unterraum von V.

Speziell: $\ker L$ ist Unterraum von V, $\operatorname{im} L$ ist Unterraum von W, und es gilt $L(\underline{0}) = \underline{0}$.

Bew.:

(a) Sind $w_1, w_2 \in L(U_1)$, so existieren $v_1, v_2 \in U_1$, so daß $L(v_1) = w_1$ und $L(v_2) = w_2$ gilt. Da U_1 Unterraum ist, gilt $v_1 + v_2 \in U_1$, also $L(v_1 + v_2) \in L(U_1)$. Die Additivität von L impliziert: $w_1 + w_2 = L(v_1) + L(v_2) = L(v_1 + v_2) \in L(U_1)$.
Analog folgt: Ist $w \in L(U_1), \alpha \in K$, so gilt $\alpha w \in L(U_1)$. Schließlich impliziert $U_1 \neq \emptyset$, daß $L(U_1) \neq \emptyset$ gilt.

(b) Ist $v \in L^{-1}(U_2), \alpha \in K$, so gilt $\alpha L(v) \in U_2$, da U_2 ein Unterraum ist. Die Homogenität von L impliziert: $L(\alpha v) = \alpha L(v) \in U_2$, also $\alpha v \in L^{-1}(U_2)$. Analog folgt: Sind $v_1, v_2 \in L^{-1}(U_2)$, so gilt $v_1 + v_2 \in L^{-1}(U_2)$. Schließlich zeigen wir, daß $L(\underline{0}) = \underline{0}$ gilt. Daraus folgt dann $\underline{0} \in L^{-1}(U_2)$, speziell $L^{-1}(U_2) \neq \emptyset$. $L(\underline{0}) = L(0\underline{0}) = 0L(\underline{0}) = \underline{0}$, vgl. (3.2)(a).

(4.3) Fakt. Sei $L \in \operatorname{Hom}(V, W)$. Folgende Aussagen sind äquivalent.

(a) L ist injektiv.

(b) $\ker L = \{\underline{0}\}$.

(c) Ist $M \subseteq V$ linear unabhängig, so ist $L(M) \subseteq W$ linear unabhängig.

Bew.: (a) $\Rightarrow$ (b): Ist L injektiv, so ist $\underline{0} \in V$ das einzige Element von V, das durch L auf $\underline{0} \in W$ abgebildet wird, also $\ker L = \{\underline{0}\}$.
(b) $\Rightarrow$ (c): Es gelte $\ker L = \{\underline{0}\}$ und $M \subseteq V$ sei linear unabhängig. Seien $w_1, \ldots, w_k$ verschiedene Elemente von $L(M)$, $\alpha_1, \ldots, \alpha_k \in K$, und es gelte $\sum_{i=1}^{k} \alpha_i w_i = \underline{0}$. Wir wollen zeigen, daß $\alpha_1 = \ldots = \alpha_k = 0$ gilt. Zu jedem $w_i \in L(M)$ existiert ein $v_i \in M$ mit $L(v_i) = w_i$. Da die $w_1, \ldots, w_k$ verschieden sind, sind auch die $v_1, \ldots, v_k$ verschieden. Durch Induktion folgt aus der Linearität von L, daß

$$L\left(\sum_{i=1}^{k} \alpha_i v_i\right) = \sum_{i=1}^{k} \alpha_i L(v_i)$$

gilt. Aus $\sum_{i=1}^{k} \alpha_i L(v_i) = \sum_{i=1}^{k} \alpha_i w_i = \underline{0}$ folgt also

$$\sum_{i=1}^{k} \alpha_i v_i \in \ker L = \{\underline{0}\},$$

d.h. $\sum_{i=1}^{k} \alpha_i v_i = \underline{0}$. Da die $v_1, \ldots, v_k$ verschiedene Elemente der linear unabhängigen Menge M sind, folgt $\alpha_1 = \ldots = \alpha_k = 0$.

(c) $\Rightarrow$ (a): Seien $v_1 \neq v_2 \in V$. Dann gilt $v_2 - v_1 \neq \underline{0}$, d.h. die Menge $M = \{v_2 - v_1\} \subseteq V$ ist linear unabhängig. Nach (c) ist dann auch $L(M) = \{L(v_2 - v_1)\} \subseteq W$ linear unabhängig, d.h. $L(v_2 - v_1) \neq \underline{0}$. Nun gilt (vgl. (3.2)(c)):

$$\underline{0} \neq L(v_2 - v_1) = L(v_2 + (-1)v_1) = L(v_2) + L((-1)v_1) = L(v_2) - L(v_1).$$

Also $L(v_1) \neq L(v_2)$, d.h. L ist injektiv.

Bem.:
Die linke Seite eines linearen Gleichungssystems I mit k Gleichungen und n Unbekannten definiert wie folgt eine lineare Abbildung $L \in \mathrm{Hom}(K^n, K^k)$:

$$L(x_1, \ldots, x_n) := (a_{11}x_1 + \ldots + a_{1n}x_n, \ldots, a_{k1}x_1 + \ldots + a_{kn}x_n) \in K^k.$$

Das Problem, für eine gegebene rechte Seite $b = (b_1, \ldots, b_k) \in K^k$ von I die Lösungsmenge L_I zu finden, ist also gerade das Problem, die Urbildmenge $L^{-1}(\{b\})$ zu finden,

$$L_I = L^{-1}(\{b\}).$$

Es gilt also:
I besitzt für die rechte Seite $b = (b_1, \ldots, b_k) \in K^k$ eine Lösung $\Leftrightarrow$ $b \in \operatorname{im} L$.
Speziell gilt: I besitzt für jede rechte Seite $b \in K^k$ eine Lösung $\Leftrightarrow$ L surjektiv.
I besitzt für jede rechte Seite höchstens eine Lösung $\Leftrightarrow$ L injektiv $\Leftrightarrow$ $\ker L = \{\underline{0}\}$.
I besitzt für jede rechte Seite genau eine Lösung $\Leftrightarrow$ L Isomorphismus.

(4.4) Lemma. Sei $L \in \operatorname{Hom}(V, W)$, $M \subseteq V$. Dann gilt:

$$L(\operatorname{span}(M)) = \operatorname{span} L(M).$$

Bew.: (a) Nach (4.2)(a) ist $L(\operatorname{span}(M))$ ein Untervektorraum von W und $L(\operatorname{span}(M))$ enthält $L(M)$. Die Definition (3.7) von "span" zeigt nun, dass $\operatorname{span}(L(M)) \subseteq L(\operatorname{span}(M))$ gilt.
(b) Ist $M = \emptyset$, so ist (4.4) trivialerweise wahr. Ist $M \neq \emptyset$, so lässt sich nach (3.8) jedes $w \in L(\operatorname{span}(M))$ schreiben als

$$w = L\left(\sum_{i=1}^{k} \alpha_i v_i\right)$$

mit $k \in \mathbb{N}_{>0}$,$\{v_1, \ldots, v_k\} \subseteq M$ und $\{\alpha_1, \ldots, \alpha_k\} \subseteq K$. Mit der Linearität von L folgt

$$w = \sum_{i=1}^{k} \alpha_i L(v_i)$$

mit $L(v_i) \in L(M)$, also $w \in \operatorname{span}(L(M))$.
Damit ist $L(\operatorname{span}(M)) \subset \operatorname{span}(L(M))$ bewiesen.

(4.5) Folgerung. Sei $L \in \operatorname{Hom}(V, W)$. Dann sind äquivalent:

(a) L ist Isomorphismus ($\Leftrightarrow$ L bijektiv).

(b) Für jede Basis B von V gilt: $L|B$ ist injektiv und $L(B)$ ist Basis von W.

(c) Es existiert eine Basis B von V, so daß $L|B$ injektiv und $L(B)$ Basis von W ist.

Bew.:

(a) $\Rightarrow$ (b). Sei $B \subseteq V$ Basis $\overset{(4.3)}{\Rightarrow}$ $L(B)$ ist linear unabhängig. $\operatorname{span} L(B) \overset{(4.4)}{=} L(\operatorname{span} B) = L(V) \overset{L \text{ surjektiv}}{=} W \Rightarrow L(B)$ ist Erzeugendensystem von W. L injektiv $\Rightarrow L|B$ injektiv.

(b) $\Rightarrow$ (c): klar, da nach (3.14) eine Basis von V existiert.

(c) $\Rightarrow$ (a). Zeige: L ist injektiv. Sei $v \in \ker L$. Dann existiert $k \in \mathbb{N}_{>0}$ und verschiedene $v_1, \ldots, v_k \in B$ und $\alpha_1, \ldots, \alpha_k \in K$: $v = \sum\limits_{i=1}^{k} \alpha_i v_i$. Daraus folgt

$$(*) \qquad \underline{0} = L(v) = \sum_{i=1}^{k} \alpha_i L(v_i).$$

Da $L|B$ injektiv ist, sind die $L(v_1), \ldots, L(v_k)$ verschiedene Elemente der Basis $L(B)$. Also folgt aus $(*)$: $\alpha_1 = \ldots = \alpha_k = 0$ und daraus $v = 0$. Also $\ker L = \{\underline{0}\}$, d.h. L ist injektiv. $L(V) = L \operatorname{span} B) \overset{(4.4)}{=} \operatorname{span} L(B) = W \Rightarrow L$ ist surjektiv.

(4.6) Fakt. Seien V, W, Z K-Vektorräume, $L_1 \in \operatorname{Hom}(V, W)$, $L_2 \in \operatorname{Hom}(W, Z)$. Dann gilt: $L_2 \circ L_1 \in \operatorname{Hom}(V, Z)$. Sind L_1 und L_2 Isomorphismen, so ist $L_2 \circ L_1 : V \to Z$ Isomorphismus.

Bez.: Zwei K-Vektorräume V und W heißen zueinander isomorph, wenn es einen Isomorphismus $L : V \to W$ gibt (und damit auch den Isomorphismus $L^{-1} : W \to V$, vgl. Bsp. 5 nach (4.1)). Fakt(4.6) zeigt: Ist V isomorph zu W und W isomorph zu Z, so ist auch V isomorph zu Z.

(4.7) Def. Eine <u>geordnete Basis</u> $\mathcal{G}$ eines n-dimensionalen Vektorraums V ist ein n-Tupel von Vektoren $\mathcal{G} = (v_1, \ldots, v_n) \in V^n$, so daß $\{v_1, \ldots, v_n\}$ eine Basis von V ist.

(4.8) Satz. *Sei* $\dim V = n < \infty$, $\mathcal{G} = (v_1, \ldots, v_n)$ *geordnete Basis von* V *und seien* $w_1, \ldots, w_n$ *beliebige Elemente von* W. *Dann existiert genau ein* $L \in \operatorname{Hom}(V, W)$, *so daß* $L(v_1) = w_1, \ldots, L(v_n) = w_n$ *gilt.*

Beweis: Vorbemerkung: Wegen $\operatorname{span}\{v_1, \ldots, v_n\} = V$ existieren zu jedem $v \in V$ Skalare $\alpha_1, \ldots, \alpha_n \in K$, so daß $v = \sum\limits_{i=1}^{n} \alpha_i v_i$ gilt, und da die $v_1, \ldots, v_n$ linear unabhängig sind, sind die $\alpha_1, \ldots, \alpha_n$ eindeutig durch v bestimmt, vgl. (3.10).

(a) Eindeutigkeit von L: Es seien $L,\ L' \in \operatorname{Hom}(V,W)$ und $L(v_i) = L'(v_i)$ für alle $i \in \{1,\dots,n\}$. Sei $v = \sum\limits_{i=1}^{n} \alpha_i v_i \in V$. Dann gilt

$$L(v) = L\left(\sum_{i=1}^{n} \alpha_i v_i\right) = \sum_{i=1}^{n} \alpha_i L(v_i) = \sum_{i=1}^{n} \alpha_i L'(v_i) = L'(\sum \alpha_i v_i) = L'(v).$$

(b) Existenz von L: Für $v = \sum\limits_{i=1}^{n} \alpha_i v_i \in V$ definieren wir

$$L(v) := \sum_{i=1}^{n} \alpha_i w_i \in W.$$

Ist $v = \sum\limits_{i=1}^{n} \alpha_i v_i,\ v' = \sum\limits_{i=1}^{n} \alpha_i' v_i$, so $v + v' = \sum\limits_{i=1}^{n} (\alpha_i + \alpha_i') v_i$ und damit: $L(v+v') = \sum\limits_{i=1}^{n} (\alpha_i + \alpha_i') w_i = \sum\limits_{i=1}^{n} \alpha_i w_i + \sum\limits_{i=1}^{n} \alpha_i' w_i = L(v) + L(v')$, d.h. L ist additiv. Analog folgt, daß L homogen ist, also $L \in \operatorname{Hom}(V,W)$. Offensichtlich gilt $L(v_i) = w_i$ für alle $i \in \{1,\dots,n\}$.

Bem.: Eine (4.8) entsprechende Aussage gilt auch im Fall $\dim V = \infty$: Ist B Basis von V und $\ell : B \to W$ eine Abbildung, so existiert genau ein $L \in \operatorname{Hom}(V,W)$, für das $L|B = \ell$ gilt. Anders ausgedrückt: Jede Abbildung einer Basis von V nach W kann zu genau einer linearen Abbildung von V nach W fortgesetzt werden. Der Beweis verläuft wie der Beweis von (4.8).

(4.9) Satz. *Seien V, W K-Vektorräume,* $\dim V = n < \infty$. *Dann gilt:*

(a) *V ist isomorph zu K^n.*

(b) *V ist isomorph zu W* $\Leftrightarrow \dim V = \dim W$.

Bew.:

(a) Sei $(v_1,\dots,v_n)$ geordnete Basis von V. Nach (4.8) existiert ein $L \in \operatorname{Hom}(V,K^n)$ mit $L(v_1) = e_1,\dots,L(v_n) = e_n$. Nach (4.5) ist L Isomorphismus.

(b) Analog folgt aus $\dim V = \dim W$, daß V und W isomorph sind. Ist umgekehrt $L : V \to W$ Isomorphismus, so folgt aus (4.5), daß $\dim V = \dim W$ gilt.

(4.10) Dimensionssatz für lineare Abbildungen. *Ist* $L \in \mathrm{Hom}(V, W)$ *und* $\dim V < \infty$*, so gilt* $\dim(\ker L) < \infty$*,* $\dim(\operatorname{im} L) < \infty$ *und*

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

Speziell folgt: Ist $\dim V = \dim W$*, so gilt:* L *injektiv* $\Leftrightarrow$ L *surjektiv* $\Leftrightarrow$ L *bijektiv.*

Bew.: Aus (3.22) folgt die Existenz eines zu $\ker L$ komplementären Unterraums U von V, d.h. $\ker L \oplus U = V$ und

$$\dim(\ker L) + \dim U = \dim V.$$

Wir zeigen, daß $L|U : U \to \operatorname{im} L$ ein Isomorphismus ist und damit $\dim U = \dim(\operatorname{im} L)$, woraus mit der vorangehenden Gleichung die Behauptung folgt.

$L|U$ ist injektiv, da $\ker(L|U) = \ker L \cap U = \{\underline{0}\}$, vgl. (4.3). Zu jedem $w \in \operatorname{im} L$ existiert ein $v \in V$ mit $L(v) = w$. Wegen $\ker L \oplus U = V$ existieren $v_1 \in \ker L$ und $u \in U$, so daß $v = v_1 + u$ gilt. Dann folgt $w = L(v) = L(v_1) + L(u) = L(u)$, d.h. $w \in L(U)$. Das zeigt, daß $L|U : U \to \operatorname{im} L$ surjektiv ist.

(4.11) Folgerung. Seien V, W K-Vektorräume, $\dim V = \dim W < \infty$, und $L \in \mathrm{Hom}(V, W)$. Dann ist L genau dann surjektiv, wenn L injektiv ist.

Bew.: L injektiv $\Leftrightarrow \ker L = \{\underline{0}\} \overset{(4.10)}{\Leftrightarrow} \dim(\operatorname{im} L) = \dim V = \dim W$
$\overset{(3.19)(c)}{\Leftrightarrow} \operatorname{im} L = W \Leftrightarrow L$ surjektiv.

Bem.: Es ist wichtig zu wissen, daß (4.11) in unendlich-dimensionalen Vektorräumen nicht wahr ist. Betrachtet man etwa den $\mathbb{R}$-Vektorraum der reellen Zahlenfolgen $V = \{a | a : \mathbb{N} \to \mathbb{R}\}$ und definiert $L \in \mathrm{End}(V)$ durch $L(a) = b$ mit

$$b(n) := \begin{cases} 0 & \text{falls } n \text{ ungerade} \\ a(m) & \text{falls } n = 2m \text{ und } m \in \mathbb{N} \end{cases}$$

so ist L injektiv, aber nicht surjektiv.

Anwendung:

Sei I die linke Seite eines linearen Gleichungssystems mit n Unbekannten und k Gleichungen und Koeffizienten $a_{ij} \in K$. Sei $L : K^n \to K^k$ die zugehörige lineare Abbildung, vgl. Bem. nach (4.3). Dann gilt: $L_{I^{\text{hom}}} = \ker L$ und

$$\{b = (b_1, \ldots, b_k) \in K^k \mid \exists \text{ Lösung von } I \text{ mit der rechten Seite } b\} = \operatorname{im} L.$$

Aus (4.10) folgt: $\dim L_{I^{\text{hom}}} = n - \dim(\operatorname{im} L) \geq n - k$. Das wurde schon in (3.25) gezeigt.

$$\begin{aligned} I \text{ ist für beliebige rechte Seiten } (b_1, \ldots, b_k) \text{ lösbar } &\Leftrightarrow \dim(\operatorname{im} L) = k \\ &\Leftrightarrow \dim(L_{I^{\text{hom}}}) = n - k \end{aligned}$$

d.h. ist $n \geq k$ und ist I für beliebige rechte Seite lösbar, so ist die Lösungsmenge für jede rechte Seite ein $(n-k)$-dimensionaler affiner Unterraum. Ist $k = n$, so gilt: I ist für beliebige rechte Seiten $(b_1, \ldots, b_n)$ lösbar $\Leftrightarrow L_{I^{\text{hom}}} = \{0\} \Leftrightarrow I$ besitzt für jede rechte Seite $(b_1, \ldots, b_n)$ genau eine Lösung.

Sind V und W K-Vektorräume, so besitzt die Menge $\operatorname{Hom}(V, W)$ in natürlicher Weise die Struktur eines K-Vektorraums. Diese Struktur werden wir jetzt definieren.

1) Multiplikation mit Skalaren:
   Ist $\alpha \in K$, $L \in \operatorname{Hom}(V, W)$, so definieren wir $\alpha L : V \to W$ durch:
   Für jedes $v \in V$ sei
   $$(\alpha L)(v) := \alpha L(v).$$
   Es ist leicht zu zeigen, daß damit $\alpha L \in \operatorname{Hom}(V, W)$ gilt. Die Additivität von αL folgt so:
   Seien $v_1, v_2 \in V$. Dann gilt:
   $$\begin{aligned} (\alpha L)(v_1 + v_2) &\overset{\text{Def.}}{=} \alpha \cdot L(v_1 + v_2) \overset{L \text{additiv}}{=} \alpha(L(v_1) + L(v_2)) \\ &\overset{(V_3)}{=} \alpha L(v_1) + \alpha L(v_2) \\ &\overset{\text{Def.}}{=} (\alpha L)(v_1) + (\alpha L)(v_2) \end{aligned}$$

2) Addition:
   Sind $L_1, L_2 \in \operatorname{Hom}(V, W)$, so definieren wir $L_1 + L_2 : V \to W$ durch:
   Für jedes $v \in V$ sei
   $$(L_1 + L_2)(v) := L_1(v) + L_2(v).$$
   Wie oben folgt $L_1 + L_2 \in \operatorname{Hom}(V, W)$.

(4.12) Satz. *Mit dieser Addition und dieser Multiplikation mit Elementen aus K ist* $\mathrm{Hom}(V, W)$ *ein K-Vektorraum.*

Bew.: Es muß gezeigt werden, daß $(\mathrm{Hom}(V, W), +)$ eine abelsche Gruppe ist und daß die Aussagen $(V_1) - (V_4)$ aus (3.1) für $\mathrm{Hom}(V, W)$ gelten. Dabei ist das neutrale Element von $(\mathrm{Hom}(V, W), +)$ die 0-Abbildung, die jedes $v \in V$ auf $\underline{0} \in W$ abbildet. Exemplarisch wird (V_2) für $\mathrm{Hom}(V, W)$ gezeigt: Seien $\alpha, \beta \in K$ und $L \in \mathrm{Hom}(V, W)$. Dann gilt für alle $v \in V$:

$$\begin{aligned} ((\alpha+\beta)L)(v) &\overset{\text{Def.}}{=} (\alpha+\beta)\cdot L(v) \\ &\overset{(V_2)\text{ für } W}{=} \alpha L(v) + \beta L(v) \\ &\overset{\text{Def.}}{=} (\alpha L)(v) + (\beta L)(v) \\ &\overset{\text{Def.}}{=} (\alpha L + \beta L)(v) \end{aligned}$$

Da das für alle $v \in V$ gilt, haben wir $(\alpha + \beta)L = \alpha L + \beta L$ gezeigt.

Wir entwickeln nun die Matrizenrechnung, die ein nützlicher “Kalkül” für das praktische Umgehen und Rechnen mit linearen Abbildungen ist, die aber noch weitere Anwendungen besitzt.

(4.13) Def.: Eine $m \times n$-Matrix A über dem Körper K besteht aus $m \cdot n$ Körperelementen $a_{ij} \in K$, $1 \le i \le m$ und $1 \le j \le n$, die wie folgt als rechteckiges Schema aufgeschrieben werden:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = (a_{ij})_{\substack{1\le i\le m \\ 1\le j\le n}}$$

Bez.: 1) Die Vektoren $(a_{11}, \dots, a_{1n}), \dots, (a_{m1}, \dots, a_{mn})$ in K^n heißen die Zeilenvektoren von A.

2) Die $m \times 1$-Matrizen $\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$ heißen die Spaltenvektoren von A.

3) Ist $m = n$, so heißt A quadratische Matrix.

4) $K^{m\times n} := \{A \mid A \text{ ist } m \times n\text{-Matrix über } K\}$.

Bem.: Die linke Seite eines linearen Gleichungssystems aus m Gleichungen für n Unbekannte ist durch eine $m \times n$-Matrix gegeben.

(4.14) Def.: Seien V, W K-Vektorräume, $\dim V = n$, $\dim W = m$, $\mathcal{G} = (v_1, \ldots, v_n)$ geordnete Basis von V, $\mathcal{G}' = (w_1, \ldots, w_m)$ geordnete Basis von W und $L \in \mathrm{Hom}(V, W)$. Dann heißt die Matrix $A = (a_{ij}) \in K^{m \times n}$, die durch

$$(*) \qquad L(v_j) = \sum_{i=1}^{m} a_{ij} w_i \quad \text{für } 1 \leq j \leq n$$

gegeben ist, die <u>Matrix von L bezüglich $\mathcal{G}$ und $\mathcal{G}'$</u>.

Bez.: Die durch $(*)$ definierte Matrix wird mit $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L)$ bezeichnet. Wir erhalten also eine Abbildung $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'} : \mathrm{Hom}(V, W) \to K^{m \times n}$.

Beachte: 1) Die Stellung der Indizes in $(*)$.
2) Es gilt $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L) \in K^{m \times n}$, wobei $m = \dim W =$ Dimension des Zielraums von L und $n = \dim V =$ Dimension des Urbildraums.

Spezialfälle: 1) Ist $V = W$ und $L = \mathrm{id}_V$, so gilt für jede geordnete Basis $\mathcal{G}$ von $V = W$:

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(\mathrm{id}_V) = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & 1 \end{pmatrix} =: E_n$$

Andere Beschreibung: $E_n = (\delta_{ij}) \in K^{n \times n}$ mit

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

E_n heißt die n-reihige Einheitsmatrix.

2) $W = K$, $L \in \mathrm{Hom}(V, K)$ Linearform. Ist $\mathcal{G} = (v_1, \ldots, v_n)$ geordnete Basis von V, $\mathcal{G}' = (1)$ Basis des K-Vektorraums K, so gilt $\mathrm{Mat}_{\mathcal{G}}^{(1)}(L) = (a_1, \ldots, a_n) \in K^{1 \times n}$, wobei $a_j = L(v_j) \in K$.
3) $V = K$ mit Basis $\mathcal{G} = (1)$, $\mathcal{G}' = (w_1, \ldots, w_m)$ geordnete Basis von W und

$L \in \mathrm{Hom}(K, W)$. Dann gilt

$$\mathrm{Mat}^{\mathcal{G}'}_{(1)}(L) = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix},$$

wobei $L(1) = \sum\limits_{i=1}^{m} a_i w_i$.

(4.15) Fakt. Die Abbildung $\mathrm{Mat}^{\mathcal{G}'}_{\mathcal{G}} : \mathrm{Hom}(V, W) \to K^{m \times n}$ ist bijektiv.

Bew.: 1) Injektivität:
Seien $L_1, L_2 \in \mathrm{Hom}(V, W)$ mit $\mathrm{Mat}^{\mathcal{G}'}_{\mathcal{G}}(L_1) = \mathrm{Mat}^{\mathcal{G}'}_{\mathcal{G}}(L_2)$. Dann gilt nach $(*)$ für alle $j \in \{1, \ldots, n\}$:

$$L_1(v_j) = \sum_{i=1}^{m} a_{ij} w_i = L_2(v_j)$$

Aus (4.8) folgt nun $L_1 = L_2$.

2) Surjektivität: Sei $A = (a_{ij}) \in K^{m \times n}$ gegeben. Für $j \in \{1, \ldots, n\}$ setze

$$\overline{w}_j := \sum_{i=1}^{m} a_{ij} w_i \in W.$$

Nun impliziert (4.8), daß es (genau) ein $L \in \mathrm{Hom}(V, W)$ gibt, so daß für alle $j \in \{1, \ldots, n\}$ gilt:

$$L(v_j) = \overline{w}_j$$

Für dieses L gilt $\mathrm{Mat}^{\mathcal{G}'}_{\mathcal{G}}(L) = A$.

Eine Matrix $A \in K^{m \times n}$ ist im Grunde nur ein etwas anderes aufgeschriebenes Element von $K^{m \cdot n}$. Deshalb läßt sich leicht eine “natürliche ”Vektorraumstruktur auf $K^{m \times n}$ definieren:

1) Multiplikation mit Skalaren:
Ist $\alpha \in K$, $A = (a_{ij} \in K^{m \times n}$, so setzen wir

$$\alpha A := (\alpha a_{ij}) \in K^{m \times n},$$

oder, ausführlicher geschrieben,

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

2) Addition:
Sind $A, B \in K^{m\times n}$, $A = (a_{ij}), B = (b_{ij})$, so definieren wir

$$A + B := (a_{ij} + b_{ij}) \in K^{m\times n},$$

d.h.

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

(4.16) Fakt. Mit diesen Operationen ist $K^{m\times n}$ ein K-Vektorraum der Dimension $m \cdot n$. Die Matrizen $E_{ij} \in K^{m\times n}$, die für $1 \le i \le m$, $1 \le j \le n$ durch

$$E_{ij} = \begin{array}{c} \text{j'te Spalte} \\ \begin{pmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix} \end{array} \text{i'te Zeile}$$

definiert sind, bilden eine Basis von $K^{m\times n}$.

Eine andere Beschreibung von E_{ij} ist

$$E_{ij} = (e_{kl}) \in K^{m\times n} \text{ mit } e_{kl} = \begin{cases} 0 & \text{falls } (k,l) \neq (i,j) \\ 1 & \text{falls } (k,l) = (i,j). \end{cases}$$

Ist $A = (a_{ij}) \in K^{m\times n}$, so gilt $A = \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} a_{ij} E_{ij}$.

(4.17) Folgerung. Seien V, W K-Vektorräume, $\mathcal{G} = (v_1, \dots, v_n)$ geordnete Basis von V, $\mathcal{G}' = (w_1, \dots, w_m)$ geordnete Basis von W. Dann ist $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'} : \mathrm{Hom}(V, W) \to K^{m\times n}$ Isomorphismus von K-Vektorräumen.

Bew.: Nach (4.15) ist $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}$ bijektiv. Es bleibt, die Additivität und Homogenität von $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}$ zu zeigen. Wir führen den Beweis der Homogenität aus: Seien $\alpha \in K$, $L \in \mathrm{Hom}(V,W)$ und $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L) = A = (a_{ij}) \in K^{m\times n}$, d.h. $L(v_j) = \sum\limits_{i=1}^{m} a_{ij}w_i$ für alle $j \in \{1,\dots,n\}$. Dann gilt:
$(\alpha L)(v_j) \overset{\text{Def.}}{=} \alpha L(v_j) = \alpha \sum\limits_{i=1}^{m} a_{ij}w_i = \sum\limits_{i=1}^{m}(\alpha a_{ij})w_i$. Daraus folgt:

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(\alpha L) = (\alpha a_{ij}) \overset{\text{Def.}}{=} \alpha A = \alpha\,\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L).$$

Seien $V, W, \mathcal{G}, \mathcal{G}'$ und $L \in \mathrm{Hom}(V,W)$ wie bisher und Z ein weiterer K-Vektorraum mit geordneter Basis $\mathcal{G}'' = (z_1,\dots,z_k)$ und $J \in \mathrm{Hom}(W,Z)$. Nach (4.6) gilt dann $J \circ L \in \mathrm{Hom}(V,Z)$.

Frage: Wie hängt $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}''}(J\circ L)$ mit $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L)$ und $\mathrm{Mat}_{\mathcal{G}'}^{\mathcal{G}''}(J)$ zusammen?

(4.18) Satz. *Ist* $A = (a_{hl}) = \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}''}(J) \in K^{k\times m}$ *und* $B = (b_{ij}) = \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L) \in K^{m\times n}$, *so gilt* $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}''}(J\circ L) = (c_{hj}) \in K^{k\times n}$ *mit*

$$c_{hj} = \sum_{i=1}^{m} a_{hi}b_{ij}$$

für $1 \le h \le k$, $1 \le j \le n$.

Vorbemerkung über Doppelsummen: Ist $(d_{ij}) \in K^{m\times n}$, so gilt

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{n} d_{ij}\right) = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} d_{ij}\right).$$

Im Fall $m = n = 2$ bedeutet das einfach

$$(d_{11}+d_{12}) + (d_{21}+d_{22}) = (d_{11}+d_{21}) + (d_{12}+d_{22}),$$

und das folgt aus der Kommutativität (und der Assoziativität) der Addition. Im Prinzip kann man den allgemeinen Fall mit vollständiger Induktion beweisen, aber die Aussage sollte auch so offensichtlich sein. Sie gilt statt für Körperelemente $d_{ij} \in K$ auch im Fall, daß die d_{ij} Vektoren eines Vektorraums sind.

Bew. von (4.18): Aus $L(v_j) = \sum\limits_{i=1}^{m} b_{ij} w_i$ und $J(w_i) = \sum\limits_{h=1}^{k} a_{hi} z_h$ folgt

$$\begin{aligned} (J \circ L)(v_j) &= J\left(\sum_{i=1}^{m} b_{ij} w_i\right) = \sum_{i=1}^{m} b_{ij} J(w_i) = \sum_{i=1}^{m} b_{ij} \left(\sum_{h=1}^{k} a_{hi} z_h\right) \\ &= \sum_{i=1}^{m} \left(\sum_{h=1}^{k} b_{ij} a_{hi} z_h\right) = \sum_{h=1}^{k} \left(\sum_{i=1}^{m} b_{ij} a_{hi} z_h\right) = \sum_{h=1}^{k} \left(\sum_{i=1}^{m} a_{hi} b_{ij}\right) z_h. \end{aligned}$$

Dabei wurde in der vorletzten Gleichung die Vorbemerkung auf die Vektoren $d_{ih} := b_{ij} a_{hi} z_h$ angewendet. Die drittletzte Gleichung folgt aus dem Distributivgesetz (V_3), die letzte aus (V_2). Insgesamt zeigt die Gleichungskette, daß $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}''}(J \circ L)$ durch die $k \times n$-Matrix mit den Koeffizienten $c_{hj} := \sum\limits_{i=1}^{m} a_{hi} b_{ij}$ gegeben ist.

(4.19) Def. (Multiplikation von Matrizen): Ist $A = (a_{hl}) \in K^{k \times m}$, $B = (b_{ij}) \in K^{m \times n}$, so heißt die Matrix $C = (c_{hj}) \in K^{k \times n}$ mit

$$c_{hj} = \sum_{i=1}^{m} a_{hi} b_{ij}$$

das Produkt von A und B, bezeichnet durch $C =: AB$.

Bem.: Damit AB definiert ist, muß die Anzahl der Spalten von A gleich der Anzahl von Zeilen von B sein.

Bsp.:

$$\underbrace{\begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}}_{\in \mathbb{R}^{2 \times 3}} \underbrace{\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & 0 \end{pmatrix}}_{\in \mathbb{R}^{3 \times 4}} = \underbrace{\begin{pmatrix} 7 & 4 & 2 & 1 \\ 1 & 3 & 4 & 1 \end{pmatrix}}_{\in \mathbb{R}^{2 \times 4}}$$

$$\underbrace{\begin{pmatrix} 3 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}}_{\in \mathbb{R}^{2 \times 3}} \underbrace{\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}}_{\in \mathbb{R}^{3 \times 1}} = \underbrace{\begin{pmatrix} 6 \\ 3 \end{pmatrix}}_{\in \mathbb{R}^{2 \times 1}}$$

(4.20) Folgerung.

(a) Mit den Bezeichnungen von (4.18) gilt:

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}''}(J \circ L) = \mathrm{Mat}_{\mathcal{G}'}^{\mathcal{G}''}(J)\,\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L).$$

(b) Die Matrizenmultiplikation ist assoziativ: $A(BC) = (AB)C$.

(c) Distributivgesetze:

$$\begin{array}{rcl} A(B_1 + B_2) & = & AB_1 + AB_2 \\ (A_1 + A_2)B & = & A_1B + A_2B. \end{array}$$

(d) $\alpha(AB) = (\alpha A)B = A(\alpha B)$

(e) $A \in K^{m\times n} \Rightarrow E_m A = AE_n = A$

Bem.: Die Matrizenmultiplikation ist nicht kommutativ. Ist etwa $A \in K^{k\times m}$, $B \in K^{m\times n}$ und ist $k \neq n$, so kann man zwar AB bilden, aber BA ist nicht definiert. Ist $k = n \neq m$, so sind zwar AB und BA definiert, liegen aber in verschiedenen Räumen. Ist schließlich $k = n = m > 1$, so findet man leicht Matrizen $A, B \in K^{m\times m}$ mit $AB \neq BA$.

Bew. von (4.20):

(a) Wir haben die Multiplikation von Matrizen gerade so definiert, daß (4.20)(a) gilt, vgl. (4.18) und (4.19).

(b) Mit etwas Übung in solchen Überlegungen kann man direkt sehen, daß die Assoziativität aus (4.15), (4.20)(a) und der Tatsache folgt, daß die Hintereinanderausführung von Abbildungen assoziativ ist. Man kann die Assoziativität aber auch nachweisen, indem man Definition (4.19) benützt und rechnet.

(c) Auch das kann wieder aus entsprechenden Gleichungen für Homomorphismen hergeleitet werden, oder durch direkte Rechnung . Etwa falls $A \in (a_{hi}) \in K^{k\times m}$, $B_1 = (b^1_{ij}) \in K^{m\times n}$, $B_2 = (b^2_{ij}) \in K^{m\times n}$:

$$\sum_{i=1}^{m} a_{hi}(b^1_{ij} + b^2_{ij}) = \sum_{i=1}^{m} a_{hi}b^1_{ij} + \sum_{i=1}^{m} a_{hi}b^2_{ij}.$$

(d) und (e) sind offensichtlich.

(4.21) Folgerung (Abhängigkeit von der Wahl der Basen): Seien $\mathcal{G}$ und $\overline{\mathcal{G}}$ geordnete Basen von V, $\mathcal{G}'$ und $\overline{\mathcal{G}}'$ geordnete Basen von W und $L \in \operatorname{Hom}(V,W)$. Dann gilt:

$$\operatorname{Mat}_{\overline{\mathcal{G}}}^{\overline{\mathcal{G}}'}(L) = \operatorname{Mat}_{\mathcal{G}'}^{\overline{\mathcal{G}}'}(\mathrm{id}_W)\operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L)\operatorname{Mat}_{\overline{\mathcal{G}}}^{\mathcal{G}}(\mathrm{id}_V)$$

Bew.: Mittels (4.20)(a) folgt aus $L = \mathrm{id}_W \circ L \circ \mathrm{id}_V$:

$$\begin{aligned}\operatorname{Mat}_{\overline{\mathcal{G}}}^{\overline{\mathcal{G}}'}(L) = \operatorname{Mat}_{\overline{\mathcal{G}}}^{\overline{\mathcal{G}}'}(\mathrm{id}_W \circ (L \circ \mathrm{id}_V)) &\overset{(4.20)(a)}{=} \operatorname{Mat}_{\mathcal{G}'}^{\overline{\mathcal{G}}'}(\mathrm{id}_W)\operatorname{Mat}_{\overline{\mathcal{G}}}^{\mathcal{G}'}(L \circ \mathrm{id}_V)\\ &\overset{(4.20)(a)}{=} \operatorname{Mat}_{\mathcal{G}'}^{\overline{\mathcal{G}}'}(\mathrm{id}_W)\operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L)\operatorname{Mat}_{\overline{\mathcal{G}}}^{\mathcal{G}}(\mathrm{id}_V).\end{aligned}$$

Wir betrachten nun den Spezialfall $V = K^n$ mit geordneter Basis $\mathcal{G}_n = (e_1, \dots, e_n)$ und $W = K^m$ mit $\mathcal{G}_m = (e_1, \dots, e_m)$.

In diesem Fall führen wir zur Abkürzung ein:

$$\operatorname{Mat} := \operatorname{Mat}_{\mathcal{G}_n}^{\mathcal{G}_m} : \operatorname{Hom}(K^n, K^m) \to K^{m\times n}.$$

Wir identifizieren die Vektoren von K^n mit den einspaltigen Matrizen in $K^{n\times 1}$ durch den Vektorraumisomorphismus

$$(x_1, \dots, x_n) \in K^n \to \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^{n\times 1}.$$

(4.22) Satz. *Mit diesen Identifikationen gilt: Ist* $L \in \operatorname{Hom}(K^n, K^m)$ *und* $\operatorname{Mat}(L) = A = (a_{ij}) \in K^{m\times n}$*, so gilt für alle* $x = (x_1, \dots, x_n) \in K^n$*:*

$$L(x) = A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Erläuterung: Identifiziert man Elemente von K^n bzw. K^m mit Spaltenvektoren, so entspricht die Anwendung eines $L \in \operatorname{Hom}(K^n, K^m)$ auf x gerade der Multiplikation von $\operatorname{Mat}(L)$ mit x.

Bew. von (4.22): Berechnung der linken Seite:

$$\begin{aligned} L(x) &= L\left(\sum_{j=1}^{n} x_j e_j\right) = \sum_{j=1}^{n} x_j L(e_j) = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij} e_i\right) \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j\right) e_i = \left(\sum_{j=1}^{n} a_{1j} x_j, \ldots, \sum_{j=1}^{n} a_{mj} x_j\right). \end{aligned}$$

Berechnung der rechten Seite:

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum\limits_{j=1}^{n} a_{1j} x_j \\ \vdots \\ \sum\limits_{j=1}^{n} a_{mj} x_j \end{pmatrix}.$$

Bem.: Speziell folgt aus (4.22), daß $L(e_j)$, aufgefaßt als Spaltenvektor, gerade die j'te Spalte $\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$ der Matrix $A = \mathrm{Mat}(L)$ ist.

(4.23) Def.: Sei $A \in K^{m \times n}$. Der Rang von A, bezeichnet als $\mathrm{rg}(A)$, ist die maximale Anzahl linear unabhängiger Spaltenvektoren von A,

$$\mathrm{rg}(A) := \dim\left(\mathrm{span}\left\{\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \middle| j \in \{1, \ldots, n\}\right\}\right).$$

(4.23) enthält implizit folgende Aussage, die vielleicht nicht völlig offensichtlich ist: Sind $v_1, \ldots, v_k$ Elemente eines Vektorraums, so gilt

$$\dim(\mathrm{span}\{v_1, \ldots, v_k\}) = \max\{|M| \mid M \subseteq \{v_1, \ldots, v_k\}, M \text{ linear unabhängig}\}.$$

Man kann das wie folgt begründen. Der Beweis von (3.14) (im endlich erzeugten Fall) zeigt, daß $\{v_1, \ldots, v_k\}$ eine Basis B von $\mathrm{span}\{v_1, \ldots, v_k\}$ enthält. Es gibt also eine linear unabhängige Teilmenge B von $\{v_1, \ldots, v_k\}$ mit $|B| = \dim(\mathrm{span}\{v_1, \ldots, v_k\})$. Andererseits zeigt (3.19)(a), daß keine linear unabhängige Teilmenge von $\{v_1, \ldots, v_k\}$ mehr als $\dim(\mathrm{span}\{v_1, \ldots, v_k\})$ Elemente enthält.

Bem.: 1) Da unter n Vektoren in einem m-dimensionalen Vektorraum höchstens m linear unabhängig sind, vgl. (3.19)(a), gilt $\operatorname{rg}(A) \leq \min\{m, n\}$.
2) Wir werden in Kap. 6 zeigen, daß $\operatorname{rg}(A)$ auch die Maximalzahl linear unabhängiger Zeilenvektoren von A ist.

(4.24) Fakt. Seien V, W K-Vektorräume mit geordneten Basen $\mathcal{G} = (v_1, \ldots, v_n)$ und $\mathcal{G}' = (w_1, \ldots, w_m)$, $L \in \operatorname{Hom}(V, W)$ und $A = \operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L)$. Dann gilt

$$\operatorname{rg}(A) = \dim(\operatorname{im} L)$$

Bew. 1) Wir betrachten zunächst den Spezialfall, daß $V = K^n$, $\mathcal{G} = \mathcal{G}_n$ und $W = K^m$, $\mathcal{G}' = \mathcal{G}_m$ gilt. Mit der Bemerkung vor (4.23) folgt

$$\operatorname{im} L = \operatorname{span}\{L(e_j) \mid 1 \leq j \leq n\} = \operatorname{span}\left\{ \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \middle| 1 \leq j \leq n \right\}.$$

Das ergibt die Behauptung.

2) Im allgemeinen Fall betrachten wir die Isomorphismen $J_1 : K^n \to V$, $J_2 : W \to K^m$, die durch

$$\begin{aligned} J_1(e_j) &= v_j \quad \text{für } 1 \leq j \leq n \\ J_2(w_i) &= e_i \quad \text{für } 1 \leq i \leq m \end{aligned}$$

definiert sind und setzen $\tilde{L} := J_2 \circ L \circ J_1 \in \operatorname{Hom}(K^n, K^m)$. Dann gilt nach (4.20)(a)

$$\operatorname{Mat}(\tilde{L}) = M_{\mathcal{G}'}^{\mathcal{G}_m}(J_2) \operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L) \operatorname{Mat}_{\mathcal{G}_n}^{\mathcal{G}}(J_1)$$

und die Definitionen von J_1 und J_2 zeigen direkt, daß

$$\operatorname{Mat}_{\mathcal{G}_n}^{\mathcal{G}}(J_1) = E_n \text{ und } \operatorname{Mat}_{\mathcal{G}'}^{\mathcal{G}_m}(J_2) = E_m$$

gilt. Daraus folgt

$$\operatorname{Mat}(\tilde{L}) = \operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L) = A$$

Da J_1, J_2 Isomorphismen sind, gilt $\operatorname{im} \tilde{L} = J_2(\operatorname{im} L)$ und

$$\dim(\operatorname{im} L) = \dim(\operatorname{im} \tilde{L}).$$

Mittels 1) folgt nun

$$\operatorname{rg}(A) = \dim(\operatorname{im} \tilde{L}) = \dim(\operatorname{im} L)$$

Bem.: In der Vorlesung war der Teil 2) des Beweises etwas anders geführt worden. Insbesondere entsprechen die hier verwendeten J_1, J_2 den Umkehrabbildungen der in der Vorlesung verwendeten.

Folgerung (siehe Übungsblatt 11, Aufgabe 2)

$$\operatorname{rg}(AB) \leq \min\{\operatorname{rg}(A), \operatorname{rg}(B)\}.$$

Das lineare Gleichungssystem

$$\text{(I)} \qquad \begin{array}{cccc} a_{11}x_1 & + \ldots + & a_{1n}x_n & = & b_1 \\ \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + \ldots + & a_{mn}x_n & = & b_m \end{array}$$

läßt sich mit der Matrix $A = (a_{ij}) \in K^{m\times n}$ und $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$ kürzer schreiben als

$$\text{(I)} \qquad Ax = b.$$

Aus dem Dimensionssatz für lineare Abbildungen (4.10) und aus (3.28) folgt: Ist der Lösungsraum L_I von I nicht leer, so ist L_I ein affiner Unterraum von K^n der Dimension $n - \operatorname{rg}(A)$.

(4.25) Satz (Rangkriterium). *Das lineare Gleichungssystem (I) besitzt genau dann eine Lösung, wenn*

$$\operatorname{rg}(A) = \operatorname{rg}(A|b)$$

gilt, wobei $(A|b) \in K^{m\times(n+1)}$ die Matrix

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} & b_m \end{pmatrix}$$

bezeichnet.

Bew.: Da Mat : $\mathrm{Hom}(K^n, K^m) \to K^{m\times n}$ nach (4.15) surjektiv ist, existiert $L \in \mathrm{Hom}(K^n, K^m)$ mit $\mathrm{Mat}(L) = A$. Dann gilt:

$$\begin{aligned}
\text{(I) besitzt eine Lösung} \quad &\Leftrightarrow \quad b \in \operatorname{im} L \\
&\Leftrightarrow \quad b \in \operatorname{span}\left\{ \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \middle| j \in \{1, \ldots, n\} \right\} \\
&\Leftrightarrow \quad \mathrm{rg}(A) = \mathrm{rg}(A|b)
\end{aligned}$$

Es wird noch eine Bemerkung zur Anwendung des Gaußschen Eliminationsverfahrens in Matrixschreibweise gegeben. Seien dazu $A \in K^{m\times n}$, $b \in K^{m\times 1}$ und I das lineare Gleichungssystem

$$\text{(I)} \qquad Ax = b$$

Wir wenden nun das Gaußsche Eliminationsverfahren auf $(A|b)$ an, verwandeln A in eine Matrix $\overline{A} = (\overline{a}_{ij}) \in K^{m\times n}$ in Stufenform (speziell $\overline{a}_{ij} = 0$, falls $i > j$) und erhalten $(\overline{A}|\overline{b}) \in K^{m\times(n+1)}$, so daß L_I mit der Lösungsmenge $L_{\overline{I}}$ von

$$\overline{A}x = \overline{b}$$

übereinstimmt. $L_{\overline{I}}$ kann leicht bestimmt werden. Außerdem gilt

$$\mathrm{rg}(A) = \mathrm{rg}(\overline{A})$$

und $\mathrm{rg}(\overline{A})$ ist leichter zu bestimmen. Diese letzte Gleichung ergibt sich wie folgt. Wegen (4.10) gilt

$$\mathrm{rg}(A) + \dim\{x \in \mathbb{R}^n \mid Ax = 0\} = n.$$

Es ist nun die wichtigste Eigenschaft des Gaußschen Eliminationsverfahrens, daß $Ax = 0$ genau dann gilt, wenn $\overline{A}x = 0$. Man erhält also $\mathrm{rg}(A) = n - \dim\{x \in \mathbb{R}^n \mid Ax = 0\} = n - \dim\{x \in \mathbb{R}^n \mid \overline{A}x = 0\} = \mathrm{rg}(\overline{A})$.

Explizites Beispiel: Wir betrachten über $K = \mathbb{R}$:

$$(I) \qquad \begin{array}{ccccccc}
x_1 & +x_2 & -x_3 & -x_4 & = & -2 \\
3x_1 & +3x_2 & -x_3 & -2x_4 & = & 6 \\
-x_1 & -x_2 & +3x_3 & +2x_4 & = & 14
\end{array}$$

Dann ist

$$
\left(A\middle|b\right) = \begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 3 & 3 & -1 & -2 & 6 \\ -1 & -1 & 3 & 2 & 14 \end{pmatrix} \quad \begin{array}{l} \mid \cdot (-3) \\ + \\ + \end{array}
$$

$$
\begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 0 & 0 & 2 & 1 & 12 \\ 0 & 0 & 2 & 1 & 12 \end{pmatrix} \quad \begin{array}{l} \\ \mid \cdot (-1) \quad \mid \cdot (-\tfrac{1}{2}) \\ + \end{array}
$$

$$
\begin{pmatrix} 1 & 1 & -1 & -1 & -2 \\ 0 & 0 & 1 & \tfrac{1}{2} & 6 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad = \left(\overline{A} \mid \overline{b}\right).
$$

Explizite Lösung:

$$
\begin{aligned} x_3 + \frac{1}{2}x_4 &= 6 \\ x_1 + x_2 - x_3 - x_4 &= 2\,. \end{aligned}
$$

Wählt man x_4 und x_2 als beliebige reelle Zahlen, so erhält man

$$
L_I = (4,0,6,0) + \operatorname{span}\left\{(\frac{1}{2}, 0, -\frac{1}{2}, 1), (-1,1,0,0)\right\}.
$$

Es gilt $\operatorname{rg}(A) = \operatorname{rg}(\overline{A}) = 2$ und $\dim L_I = 2 = 4 - \operatorname{rg}(A)$.

5 Die allgemeine lineare Gruppe $GL_n(K)$

Wir betrachten nun einen Spezialfall des letzten Kapitels, nämlich daß $V = W$ und damit $\operatorname{Hom}(V,W) = \operatorname{End}(V)$ gilt. Mit $\mathcal{G} = (v_1, \ldots, v_n)$ sei eine geordnete Basis von V bezeichnet.

Es gilt:
$(\operatorname{End}(V), +, \circ)$ ist ein Ring mit $1 = \mathrm{id}_V$, der Endomorphismenring von V.
$(K^{n\times n}, +, \cdot)$ ist ein Ring mit $1 = E_n$, der Ring der $n \times n$-Matrizen über K.

(5.1) Folgerung. $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}} : \mathrm{End}(V) \to K^{n\times n}$ ist Isomorphismus von Ringen.

Bew.: Die Additivität und Bijektivität von $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}} : \mathrm{End}(V) \to K^{n\times n}$ folgt aus (4.17), die Multiplikativität aus (4.20).

(5.2) Lemma. Sei $(R, +, \cdot)$ Ring mit 1 und

$$R^* = \{r \in R \mid \exists s \in R : sr = rs = 1\}.$$

Dann ist $(R^*, \cdot)$ eine Gruppe (mit neutralem Element 1).

Bez.: Ein $r \in R^*$ heißt Einheit von R. R^* heißt die Gruppe der Einheiten von R.

Bem.: Zu jedem $r \in R^*$ gibt es nur ein $s \in R$ mit $sr = rs = 1$. Dieses s wird als r^{-1} bezeichnet.

Bsp.: 1) Ist K ein Körper, so ist $K^* = K \setminus \{0\}$.
2) Die Einheiten des Rings $(\mathbb{Z}, +, \cdot)$ sind 1 und -1.

Bew. von (5.2). 1) Seien $r_1, r_2 \in R^*$. Wir zeigen: $r_1 r_2 \in R^*$. Seien $s_1, s_2 \in R$ mit $s_1 r_1 = r_1 s_1 = 1$ und $s_2 r_2 = r_2 s_2 = 1$. Dann gilt

$$(s_2 s_1)(r_1 r_2) = s_2(s_1 r_1)r_2 = s_2 1 r_2 = s_2 r_2 = 1$$

und

$$(r_1 r_2)(s_2 s_1) = r_1(r_2 s_2)s_1 = r_1 s_1 = 1.$$

Also: $r_1 r_2 \in R^*$.

2) Wegen $1 \cdot 1 = 1$ gilt $1 \in R^*$ und $1^{-1} = 1$. Also enthält R^* ein neutrales Element bzgl. $\cdot$.

3) Gilt $r \in R^*$ und $sr = rs = 1$, so folgt $s \in R^*$, d.h. r besitzt ein Links- (und Rechts-) inverses in R^*.

Bsp.: $\mathrm{End}(V)^* = \{L \in \mathrm{End}(V) \mid L \text{ Isomorphismus }\}$. Ein $L \in \mathrm{End}(V)^*$ heißt Automorphismus von V.
$(K^{n\times n})^* = \{A \in K^{n\times n} \mid \exists B \in K^{n\times n} : BA = AB = E_n\}$.

Zu $A \in (K^{n\times n})^*$ wird das Element $B \in K^{n\times n}$ mit $BA = AB = E_n$ als A^{-1} bezeichnet, d.h. $A^{-1} \in K^{n\times n}$ ist eindeutig bestimmt durch $A^{-1}A = AA^{-1} = E_n$. Es gilt $A^{-1} \in (K^{n\times n})^*$.

(5.3) Def.: $\mathrm{Aut}(V) := \{L \in \mathrm{End}(V) \mid L \text{ Isomorphismus}\}$ heißt die Gruppe der Automorphismen von V.
$\mathrm{GL}_n(K) := (K^{n\times n})^*$ heißt die allgemeine lineare Gruppe der Stufe n.
$A \in \mathrm{GL}_n(K)$ heißt reguläre oder invertierbare oder nicht ausgeartete $(n \times n)$-Matrix.

Bez.: Ein $A \in K^{n\times n} \setminus \mathrm{GL}_n(K)$ heißt singulär oder ausgeartet.

(5.4) Folgerung (aus (5.1)). Ist $\mathcal{G}$ geordnete Basis von V, so ist

$$\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}} \mid \mathrm{Aut}(V) : \mathrm{Aut}(V) \to \mathrm{GL}_n(K)$$

ein Gruppenisomorphismus.

Wegen $\dim V = n < \infty$ und (4.10) sind die Automorphismen von V genau die injektiven $L \in \mathrm{End}(V)$ oder genau die surjektiven $L \in \mathrm{End}(V)$. Letzteres kann auch durch $\dim(\mathrm{im}\, L) = n$ gekennzeichnet werden. Analog gilt:

(5.5) Fakt. Ein $A \in K^{n\times n}$ ist genau dann regulär, wenn $\mathrm{rg}(A) = n$ gilt. Besitzt $A \in K^{n\times n}$ ein Linksinverses oder besitzt $A \in K^{n\times n}$ ein Rechtsinverses, so gilt $A \in \mathrm{GL}_n(K)$ und das Links- bzw. Rechtsinverse ist das Inverse A^{-1} von A.

Bew.: 1) Wir betrachten das $L \in \mathrm{End}(K^n)$, für das $\mathrm{Mat}(L) = A$ gilt. Dann folgt:

$$\begin{aligned} A \in \mathrm{GL}_n(K) &\overset{(5.4)}{\Leftrightarrow} L \in \mathrm{Aut}(K^n) \Leftrightarrow L \text{ Isomorphismus } \overset{(4.10)}{\Leftrightarrow} L \text{ surjektiv} \\ &\Leftrightarrow \dim(\mathrm{im}\, L) = n \overset{(4.24)}{\Leftrightarrow} \mathrm{rg}(A) = n. \end{aligned}$$

2) Ist etwa $B \in K^{n\times n}$ und gilt $BA = E_n$, so wähle $J \in \mathrm{Hom}(K^n, K^n)$ mit $\mathrm{Mat}(J) = B$. Dann gilt $\mathrm{Mat}(J \circ L) = \mathrm{Mat}(J)\,\mathrm{Mat}(L) = BA = E_n$, also $J \circ L = \mathrm{id}_V$. Daraus folgt, daß L injektiv ist, also $L \in \mathrm{Aut}(V)$. Mit (5.4) folgt $A = \mathrm{Mat}(L) \in \mathrm{GL}_n(K)$. Der Beweis für den Fall $AB = E_n$ ist ähnlich.

Bem.: Hier ist eine Charakterisierung der regulären $A \in K^{n\times n}$ durch Eigenschaften des zugehörigen Gleichungssystems $Ax = b$: Ist $A \in K^{n\times n}$, so gilt $A \in \mathrm{GL}_n(K)$ genau dann, wenn das lineare Gleichungssystem $Ax = b$ für jede rechte Seite $b \in \mathbb{R}^n$ genau eine Lösung x, nämlich $x = A^{-1}b$, hat.

Bem. (ohne Beweis): In einem mathematisch präzisierbaren Sinn sind die meisten Matrizen in $\mathbb{R}^{n\times n}$ (oder in $\mathbb{C}^{n\times n}$) regulär. Aber es gibt natürlich auch viele (insbesondere ∞ viele) singuläre (für $n > 1$), da die singulären Matrizen genau die mit n linear abhängigen Spaltenvektoren sind.

(5.6) Folgerung (aus (4.21)). Seien $\mathcal{G}, \overline{\mathcal{G}}$ geordnete Basen von $V, L \in \mathrm{End}(V)$ und $P := \mathrm{Mat}^{\mathcal{G}}_{\overline{\mathcal{G}}}(\mathrm{id}_V)$. Dann gilt $P \in \mathrm{GL}_n(K)$, $P^{-1} = \mathrm{Mat}^{\overline{\mathcal{G}}}_{\mathcal{G}}(\mathrm{id}_V)$ und

$$\mathrm{Mat}^{\overline{\mathcal{G}}}_{\overline{\mathcal{G}}}(L) = P^{-1}\,\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L)P.$$

Bew.: Nach (4.20)(a) gilt $\mathrm{Mat}^{\overline{\mathcal{G}}}_{\mathcal{G}}(\mathrm{id}_V)\,\mathrm{Mat}^{\mathcal{G}}_{\overline{\mathcal{G}}}(\mathrm{id}_V) = \mathrm{Mat}^{\overline{\mathcal{G}}}_{\overline{\mathcal{G}}}(\mathrm{id}_V) = E_n$.
Aus (5.5) folgt nun $P \in \mathrm{GL}_n(K)$ und $\mathrm{Mat}^{\overline{\mathcal{G}}}_{\mathcal{G}}(\mathrm{id}_V) = P^{-1}$. Damit folgt die letzte Behauptung von (5.6) direkt aus (4.21).

(5.7) Def. Eine Matrix $\tilde{A} \in K^{n\times n}$ heißt <u>ähnlich zur</u> Matrix $A \in K^{n\times n}$, wenn es ein $P \in \mathrm{GL}_n(K)$ gibt, so daß

$$\tilde{A} = P^{-1}AP$$

gilt.

Bem.: 1) "ähnlich" ist eine Äquivalenzrelation auf $K^{n\times n}$.

2) (5.6) besagt, daß für ein festes $L \in \mathrm{End}(V)$ und für beliebige geordnete Basen $\mathcal{G}$ von V die Matrizen $\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L)$ alle zueinander ähnlich sind.

<u>Problem</u>: Man finde zu gegebenem $A \in K^{n\times n}$ eine möglichst einfache zu A ähnliche Matrix $\tilde{A}$. Der "Idealfall", der im allgemeinen nicht erreicht werden kann, ist eine Diagonalmatrix $\tilde{A}$, d.h. eine Matrix $\tilde{A} = (\tilde{a}_{ij})$ mit $\tilde{a}_{ij} = 0$ für $i \neq j$, d.h.

$$\tilde{A} = \begin{pmatrix} \tilde{a}_{11} & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \tilde{a}_{nn} \end{pmatrix}.$$

Dieses Problem ist sehr viel schwieriger als die bisher betrachteten, und wir werden es im nächsten Semester lösen.

Hin und wieder will man das Inverse einer Matrix $A \in \mathrm{GL}_n(K)$ auch explizit ausrechnen. Wir lernen nun einen Algorithmus kennen, der für gegebenes $A \in K^{n\times n}$ entscheidet, ob A in $\mathrm{GL}_n(K)$ liegt und mit dem man A^{-1} recht effizient ausrechnen kann, wenn $A \in \mathrm{GL}_n(K)$ gilt. Dieser Gaußsche Algorithmus ist ganz eng mit dem Gaußschen Eliminationsverfahren verwandt.

(5.8) Berechnung der inversen Matrix.

Zu gegebenem $A \in K^{n\times n}$ betrachten wir die Matrix

$$(A \mid E_n) = \left(\begin{array}{c|ccc} & 1 & & 0 \\ A & & \ddots & \\ & 0 & & 1 \end{array}\right) \in K^{n\times(2n)}.$$

Wir wenden nun das Gaußsche Verfahren auf die Matrix A an und vollziehen die entsprechenden Zeilenoperationen an der rechtsstehenden Matrix (die im 1-Schritt E_n ist) nach. Man kann sich vorstellen, daß diese Matrix aus n Spaltenvektoren (im 1. Schritt $e_1, \dots, e_n$) besteht, und das Verfahren ist ganz analog zu dem am Ende von Kapitel 4 beschriebenen Verfahren für die Matrix $(A \mid b) \in K^{n\times(n+1)}$. Nach höchstens n Schritten erhält man eine Matrix $(\overline{A} \mid \overline{B})$ mit $\overline{a}_{ij} = 0$ für $i > j$, wobei $\mathrm{rg}(\overline{A}) = \mathrm{rg}(A)$ gilt, vgl. S. 68. In Lemma (5.9) werden wir sehen, daß $\mathrm{rg}(\overline{A}) = n$ gilt, falls alle $\overline{a}_{ii}$ ungleich null sind. Es folgt also

$$A \in \mathrm{GL}_n(K) \Leftrightarrow \overline{a}_{ii} \neq 0 \text{ für alle } i \in \{1, \dots, n\}.$$

Wenn $\overline{a}_{ii} = 0$ für ein $i \in \{1, \dots, n\}$ gilt, so bricht das Verfahren ab: Wir haben erkannt, daß $A \notin \mathrm{GL}_n(K)$ gilt und daß keine zu A inverse Matrix existiert. Andernfalls können wir annehmen, daß $\overline{a}_{ii} = 1$ für alle $i \in \{1, \dots, n\}$ gilt. Durch geeignete Addition der letzten, der vorletzten, der 2. Zeile zu den darüberliegenden verwandelt man $(\overline{A} \mid \overline{B})$ in eine Matrix $(E_n \mid \overline{\overline{B}}) \in K^{n\times(2n)}$. Wir zeigen nun, daß $\overline{\overline{B}}$ die gesuchte Matrix A^{-1} ist.

Betrachtet man während des ganzen Verfahrens nur die j'te Spalte der rechtsstehenden Matrix, so sehen wir (vgl. das am Ende von Kap. 4 beschriebene Verfahren):

<u>Die Lösungsmenge der Gleichung</u> $Ax = e_j$ <u>stimmt mit der Lösungsmenge der Gleichung</u> $E_n x = j$'te Spalte von $\overline{\overline{B}} = \overline{\overline{B}} e_j$ <u>überein.</u>

Für die Lösung $\overline{\overline{B}}e_j$ der Gleichung $E_n x (= x) = \overline{\overline{B}}e_j$ gilt also auch $A(\overline{\overline{B}}e_j) = e_j$. Daraus folgt:

$$A\overline{\overline{B}}e_j = e_j \text{ für alle } j \in \{1, \dots, n\}.$$

Da $A\overline{\overline{B}}e_j$ die j'te Spalte von $A\overline{\overline{B}} \in K^{n\times n}$ ist, folgt aus der vorangehenden Gleichung

$$A\overline{\overline{B}} = E_n.$$

Nach (5.5) impliziert das $\overline{\overline{B}} = A^{-1}$.

Wir geben nun ein explizites Beispiel für die Durchführung dieses Verfahrens.

Berechnung von A^{-1} für $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 1 \end{pmatrix} \in \mathbb{Q}^{3\times 3}$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 1 & 2 & 1 & 0 & 0 & 1 \end{array}\right) \qquad \begin{array}{cc} - & - \\ & \\ & \end{array}$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right) \qquad \begin{array}{cc} & \\ + & \cdot(-1) \\ & \end{array}$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -2 & 1 & 1 \end{array}\right) \qquad \begin{array}{cc} & \\ & \\ + & - \end{array}$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 3 & -1 & -1 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -2 & 1 & 1 \end{array}\right) \qquad \begin{array}{c} \\ - \\ \end{array}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 4 & -1 & -2 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -2 & 1 & 1 \end{array}\right) \Rightarrow A^{-1} = \begin{pmatrix} 4 & -1 & -2 \\ -1 & 0 & 1 \\ -2 & 1 & 1 \end{pmatrix} \in \mathrm{GL}_3(\mathbb{Q}) \subseteq \mathbb{Q}^{3\times 3}.$$

Speziell folgt: $A \in \mathrm{GL}_3(\mathbb{Q})$.

Zur Begründung des Verfahrens (5.8) ist noch folgendes Lemma nachzutragen.

(5.9) Lemma. Sei $A = (a_{ij}) \in K^{n \times n}$ eine "obere Dreiecksmatrix", d.h. es gilt

$$(*) \qquad a_{ij} = 0 \text{ für alle } i > j.$$

Dann gilt:

$$\mathrm{rg}(A) = n \Leftrightarrow a_{ii} \neq 0 \text{ für alle } i \in \{1, \ldots, n\}.$$

Bew.: Wir betrachten das homogene lineare Gleichungssystem

$$(I) \qquad Ax = 0$$

und zeigen, daß (I) genau dann eine Lösung $x \neq 0$ besitzt, wenn mindestens eines der a_{ii} gleich 0 ist. Wegen

$$\mathrm{rg}(A) = n - \dim\{x \mid Ax = 0\}$$

folgt daraus die Behauptung.
Gilt $a_{ii} \neq 0$ für alle $i \in \{1, \ldots, n\}$, so zeigt das sukzessive Lösungsverfahren für (I) sofort, daß $x = 0$ die einzige Lösung von (I) ist. Andernfalls sei $j \in \{1, \ldots, n\}$ der größte Index, für den $a_{jj} = 0$ gilt. Wegen $(*)$ reduziert sich dann die j'te Gleichung (I_j) auf

$$(I_j) \qquad 0 \cdot x_1 + \ldots + 0 \cdot x_j + a_{jj+1}x_{j+1} + \ldots + a_{jn}x_n = 0.$$

Andererseits folgt wie oben aus $a_{j+1j+1} \neq 0, \ldots, a_{nn} \neq 0$, daß $x_n = 0, \ldots, x_{j+1} = 0$ für jedes x gilt, das die Gleichungen $(I_n), \ldots, (I_{j+1})$ erfüllt. Die Gleichung (I_j) folgt demnach aus den Gleichungen $(I_{j+1}), \ldots, (I_n)$. Man erhält also für jede Wahl von $x_j \in K$ mindestens eine Lösung $x \in L_I$, und insbesondere eine Lösung $x \neq 0$.

6 Dualität

Sei K ein Körper und V ein K-Vektorraum.

(6.1) Def.: Der K-Vektorraum.

$$V^* := \mathrm{Hom}(V, K) = \{l : V \to K \mid l \text{ Linearform}\}$$

heißt der <u>Dualraum von</u> V.

Bem.: Die Struktur von $\operatorname{Hom}(V, K)$ als K-Vektorraum ist in (4.12) erklärt.

Im Spezialfall $V = K^n$ besteht $(K^n)^*$ aus allen Abbildungen l der Form

$$l(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i x_i$$

mit $a_i = l(e_i) \in K$. $(K^n)^*$ kann also als Menge der linken Seiten von einzeiligen linearen Gleichungen interpretiert werden.
Ist $\dim V = n < \infty$ und ist $\mathcal{G} = (v_1, \ldots, v_n)$ eine geordnete Basis von V, $\mathcal{G}' = (1)$ Basis von K, so gilt für jedes $l \in V^*$:

$$\operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}'}(l) = (a_1, \ldots, a_n) \in K^{1\times n} \text{ mit } a_i = l(v_i).$$

Speziell gilt dann $\dim V^* = n = \dim V$, vgl. auch (4.16) und (4.17).

(6.2) Satz. *Sei* $\dim V = n < \infty$ *und* $\mathcal{G} = (v_1, \ldots, v_n)$ *geordnete Basis von* V*. Für* $i \in \{1, \ldots, n\}$ *sei* $l_i \in V^*$ *definiert durch*

$$l_i(v_j) = \begin{cases} 0 & \text{falls } i \neq j \\ 1 & \text{falls } i = j \end{cases}$$

Dann ist $\mathcal{G}^* := (l_1, \ldots, l_n)$ *eine geordnete Basis von* V^**, genannt die zu* $\mathcal{G}$ *duale Basis. Für alle* $l \in V^*$ *gilt*

$$l = \sum_{i=1}^{n} l(v_i) l_i.$$

Bem.: Nach Satz (4.8) ist $l_i \in V^*$ durch obige Gleichung eindeutig definiert. Kürzer kann man dafür $l_i(v_j) = \delta_{ij}$ schreiben.

Bew.: 1) Lineare Unabhängigkeit von $\mathcal{G}^*$: Seien $\alpha_1 \in K, \ldots, \alpha_n \in K$ und

$$\sum_{i=1}^{n} \alpha_i l_i = 0 \in V^*.$$

Dann gilt für alle $j \in \{1, \ldots, n\}$:

$$0 = \left(\sum_{i=1}^{n} \alpha_i l_i\right)(v_j) = \sum_{i=1}^{n} \alpha_i l_i(v_j) = \sum_{i=1}^{n} \alpha_i \delta_{ij} = \alpha_j.$$

Das beweist, daß $\mathcal{G}^*$ linear unabhängig ist.
2) Wir zeigen, daß für jedes $l \in V^*$

$$l = \sum_{i=1}^{n} l(v_i) l_i$$

gilt. Speziell folgt hieraus, daß $\mathcal{G}^*$ ein Erzeugendensystem für V^* ist.

Wir zeigen, daß die vorangehende Gleichung angewandt auf irgendein Basiselement v_j richtig ist. Daraus folgt die Gleichheit der Linearformen, vgl. Satz (4.8). Es gilt:

$$\left(\sum_{i=1}^{n} l(v_i) l_i \right) (v_j) = \sum_{i=1}^{n} l(v_i) l_i(v_j) = \sum_{i=1}^{n} l(v_i) \delta_{ij} = l(v_j).$$

(6.3) Def.: Sind V, W K-Vektorräume und ist $L \in \mathrm{Hom}(V, W)$, so heißt die Abbildung

$$L : W^* \to V^*,$$

die für $l \in W^*$ durch

$$L^*(l) := l \circ L$$

definiert ist, die zu L duale Abbildung.

Bem.: Explizit bedeutet die vorangehende Gleichung: Für alle $v \in V$ gilt $(L^*(l))(v) = l(L(v))$.

(6.4) Fakt. (a) $(\mathrm{id}_V)^* = \mathrm{id}_{V^*}$.
(b) Sind V, W, Z K-Vektorräume und $L \in \mathrm{Hom}(V, W)$, $J \in \mathrm{Hom}(W, Z)$, so gilt

$$(J \circ L)^* = L^* \circ J^*$$

(c) Es gilt $L^* \in \mathrm{Hom}(W^*, V^*)$, und die Abbildung

$$L \in \mathrm{Hom}(V, W) \to L^* \in \mathrm{Hom}(W^*, V^*)$$

ist ein injektiver Homomorphismus.

Bem.: 1) Explizit bedeutet die letzte Behauptung, daß für alle L_1, $L_2 \in \mathrm{Hom}(V, W)$, α_1, $\alpha_2 \in K$ gilt:

$$\begin{array}{rcl} (\alpha_1 L_1 + \alpha_2 L_2)^* & = & \alpha_1 L_1^* + \alpha_2 L_2^* \\ L_1 \neq L_2 & \Rightarrow & L_1^* \neq L_2^* \end{array}$$

2) In der Sprache der Kategorien besagen die Eigenschaften (6.4)(a) und (b) gerade, daß * ein kontravarianter Funktor ist.

Bew.:

(b) Für alle $l \in Z^*$ gilt:

$$\begin{array}{rcl} (J \circ L)^*(l) & \overset{(6.3)}{=} & l \circ (J \circ L) = (l \circ J) \circ L \\ & \overset{(6.3)}{=} & (J^*(l)) \circ L \\ & \overset{(6.3)}{=} & L^*(J^*(l)) = (L^* \circ J^*)(l) \end{array}$$

(c) Für alle $l \in W^*$ gilt:

$$\begin{array}{rcl} (\alpha_1 L_1 + \alpha_2 L_2)^*(l) & \overset{(6.3)}{=} & l \circ (\alpha_1 L_1 + \alpha_2 L_2) \\ & \overset{l \text{ linear}}{=} & \alpha_1 (l \circ L_1) + \alpha_2 (l \circ L_2) \\ & \overset{(6.3)}{=} & \alpha_1 (L_1^*(l)) + \alpha_2 (L_2^*(l)) \\ & = & (\alpha_1 L_1^* + \alpha_2 L_2^*)(l) \end{array}$$

Dabei beruht die letzte Gleichung auf der Definition der K-Vektorraumstruktur von $\mathrm{Hom}(W^*, V^*)$, vgl. (4.12). Da $L \to L^*$ eine lineare Abbildung ist, folgt die Injektivität von $L \to L^*$ aus

$$L \neq 0 \Rightarrow L^* \neq 0.$$

Um das zu beweisen, wählen wir ein $v \in V$ mit $L(v) \neq 0$. Wir ergänzen $w := L(v)$ zu einer Basis $\mathcal{B}$ von W und definieren $l \in W^*$ durch

$$l(b) = \begin{cases} 0 & \text{falls } b \in \mathcal{B} \setminus \{w\} \\ 1 & \text{falls } b = w, \end{cases}$$

vgl. (4.8). Dann gilt

$$(L^*(l))(v) = l(L(v)) = l(w) = 1,$$

also $L^*(l) \neq 0$, also $L^* \neq 0$, wie behauptet.

Bem.: Sind V und W endlichdimensional, so gilt $\dim(\mathrm{Hom}(V,W)) = \dim V \cdot \dim W = \dim(\mathrm{Hom}(W^*,V^*))$. In diesem Fall folgt aus (6.4)(c), daß $L \in \mathrm{Hom}(V,W) \to L^* \in \mathrm{Hom}(W^*,V^*)$ ein Isomorphismus ist, vgl. (4.11). Es ist nicht schwierig einzusehen, daß dafür schon $\dim W < \infty$ genügt. Ist jedoch W unendlichdimensional, so ist $L \in \mathrm{Hom}(V,W) \to L^* \in \mathrm{Hom}(W^*,V^*)$ nicht surjektiv.

(6.5) Def.: Ist $A = (a_{ij}) \in K^{m\times n}$, so heißt die Matrix $B = (b_{ji}) \in K^{n\times m}$ mit $b_{ji} := a_{ij}$ für $1 \leq j \leq n$, $1 \leq i \leq m$ die transponierte Matrix zu A, bezeichnet durch $B =: A^T$.

Die j'te Zeile von A^T ist also gerade die j'te Spalte von A.

(6.6) Satz. *Seien V und W endlichdimensionale K-Vektorräume mit geordneten Basen $\mathcal{G}_V$ und $\mathcal{G}_W$ und sei $L \in \mathrm{Hom}(V,W)$. Dann ist die Matrix von $L^* : W^* \to V^*$ bezüglich der Dualbasen $\mathcal{G}_W^*$ und $\mathcal{G}_V^*$ gerade die transponierte Matrix zur Matrix von L bezüglich $\mathcal{G}_V$ und $\mathcal{G}_W$, d.h.*

$$\mathrm{Mat}_{\mathcal{G}_W^*}^{\mathcal{G}_V^*}(L^*) = \left(\mathrm{Mat}_{\mathcal{G}_V}^{\mathcal{G}_W}(L)\right)^T.$$

Bew.: Seien $\mathcal{G}_V = (v_1,\ldots,v_n)$, $\mathcal{G}_V^* = (l_1,\ldots,l_n)$, $\mathcal{G}_W = (w_1,\ldots,w_m)$, $\mathcal{G}_W^* = (f_1,\ldots,f_m)$. Die Matrizen $(a_{ij}) = \mathrm{Mat}_{\mathcal{G}_V}^{\mathcal{G}_W}(L) \in K^{m\times n}$ und $(b_{ij}) = \mathrm{Mat}_{\mathcal{G}_V^*}^{\mathcal{G}_W^*}(L^*) \in K^{n\times m}$ sind durch

$$(*) \qquad L(v_j) = \sum_{i=1}^m a_{ij} w_i \qquad \text{für } 1 \leq j \leq n$$

und

$$(**) \qquad L^*(f_i) = \sum_{j=1}^n b_{ji} l_j \qquad \text{für } 1 \leq i \leq m$$

definiert. Aus $(**)$ folgt

$$L^*(f_i)(v_j) = \sum_{k=1}^n b_{ki} l_k(v_j) = \sum_{k=1}^n b_{ki}\delta_{kj} = b_{ji}.$$

Nach Definition von L^* gilt

$$\begin{aligned} L^*(f_i)(v_j) &= f_i(L(v_j)) \overset{(*)}{=} f_i\left(\sum_{k=1}^{m} a_{kj} w_k\right) \\ &= \sum_{k=1}^{m} a_{kj}\delta_{ik} = a_{ij}. \end{aligned}$$

Die vorangehenden Gleichungen zeigen $b_{ji} = a_{ij}$, und das ist gerade die Behauptung.

Übersetzt man Satz (6.6) in die Matrizenrechnung, so erhält man:

(6.7) Folgerung. Ist $A \in K^{m\times n}$, $B \in K^{n\times p}$, so gilt

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

Bew.: Wir wählen K-Vektorräume V, W, Z mit $\dim V = p$, $\dim W = n$ und $\dim V = m$ und geordneten Basen $\mathcal{G}_V$, $\mathcal{G}_W$ und $\mathcal{G}_Z$. Nach (4.15) existieren $L \in \operatorname{Hom}(V, W)$, $J \in \operatorname{Hom}(W, Z)$ mit $\operatorname{Mat}^{\mathcal{G}_W}_{\mathcal{G}_V}(L) = B$ und $\operatorname{Mat}^{\mathcal{G}_Z}_{\mathcal{G}_W}(J) = A$. Dann gilt:

$$\begin{aligned} (AB)^T &= \left(\operatorname{Mat}^{\mathcal{G}_Z}_{\mathcal{G}_W}(J)\operatorname{Mat}^{\mathcal{G}_W}_{\mathcal{G}_V}(L)\right)^T \overset{(4.20)(a)}{=} \left(\operatorname{Mat}^{\mathcal{G}_Z}_{\mathcal{G}_V}(J\circ L)\right)^T \\ &\overset{(6.6)}{=} \operatorname{Mat}^{\mathcal{G}_V^*}_{\mathcal{G}_Z^*}((J\circ L)^*) \overset{(6.4)(b)}{=} \operatorname{Mat}^{\mathcal{G}_V^*}_{\mathcal{G}_Z^*}(L^*\circ J^*) \\ &\overset{(4.20)(a)}{=} \operatorname{Mat}^{\mathcal{G}_V^*}_{\mathcal{G}_W^*}(L^*)\operatorname{Mat}^{\mathcal{G}_W^*}_{\mathcal{G}_Z^*}(J^*) \overset{(6.6)}{=} B^T A^T. \end{aligned}$$

Hier ist ein Beispiel, das zeigt, wie man die "Funktoreigenschaften" (6.4)(a) und (b) benutzen kann.

(6.8) Lemma. Ist $L : V \to W$ Isomorphismus, so ist $L^* : W^* \to V^*$ Isomorphismus, und es gilt $(L^*)^{-1} = (L^{-1})^*$.

Bew.: Sei L Isomorphismus und $L^{-1} : W \to V$ der inverse Isomorphismus, d.h. es gilt $L^{-1}\circ L = \mathrm{id}_V$, $L\circ L^{-1} = \mathrm{id}_W$. Dann folgt aus (6.4)(a) und (b):

$$L^*\circ (L^{-1})^* = (L^{-1}\circ L)^* = (\mathrm{id}_V)^* = \mathrm{id}_{V^*}$$

und

$$(L^{-1})^*\circ L^* = (L\circ L^{-1})^* = (\mathrm{id}_W)^* = \mathrm{id}_{W^*}.$$

Diese Gleichungen zeigen, daß L^* bijektiv ist und daß $(L^{-1})^*$ die Umkehrabbildung von L^* ist.

(6.9) Satz. *Seien V, W K-Vektorräume und $L \in \mathrm{Hom}(V,W)$. Dann gelten:*

(a) L surjektiv $\Rightarrow$ L^* injektiv

(b) L injektiv $\Rightarrow$ L^* surjektiv

Bew.: (a) Wir zeigen, daß $\ker(L^*) = 0$ gilt, wenn L surjektiv ist. Sei $l \in W^*$ und $L^*(l) = l \circ L = 0$. Da $L(V) = W$ gilt, folgt $l = 0$.
(b) Wir schreiben L als $L = i \circ \tilde{L}$, wobei $i : L(V) \to W$ die Inklusionsabbildung ist und $\tilde{L} : V \to L(V)$ sich nur dadurch von L unterscheidet, daß wir den Zielbereich W durch den Unterraum $L(V)$ von W ersetzen.

Die Injektivität von L impliziert nun, daß $\tilde{L}$ ein Isomorphismus ist, und mit (6.8) folgt, daß $\tilde{L}^* : L(V)^* \to V^*$ ein Isomorphismus ist. Wegen $L^* = \tilde{L}^* \circ i^*$ genügt es also zu zeigen, daß $i^* : W^* \to L(V)^*$ surjektiv ist. Nach (3.22)(a) existiert ein zu $L(V)$ komplementärer Untervektorraum U von W, d.h. es gilt $L(V) \oplus U = W$. Ist nun $l \in L(V)^*$ gegeben, so setzen wir l auf W fort, indem wir für $w = w_1 + u \in W$ mit $w_1 \in L(V), u \in U$ definieren:

$$\tilde{l}(w) := l(w_1).$$

Dann gilt $\tilde{l} \in W^*$ und $i^*(\tilde{l}) = \tilde{l} \circ i = l$. Das zeigt die Surjektivität von i^* und beendet den Beweis von (b).

(6.10) Folgerung. Ist $L \in \mathrm{Hom}(V,W)$ und ist W endlich dimensional, so gilt

$$\dim(\mathrm{im}\, L) = \dim(\mathrm{im}\, L^*).$$

Bew.: Wir schreiben wieder L als $L = i \circ \tilde{L}$, wobei $\tilde{L} : V \to L(V)$ durch $\tilde{L}(v) = L(v)$ für alle $v \in V$ definiert ist und $i : L(V) \to W$ die Inklusionsabbildung bezeichnet. Dann ist $\tilde{L}$ surjektiv und i ist injektiv, und es gilt

$$L^* = \tilde{L}^* \circ i^*.$$

Nach (6.9) ist $i^* : W^* \to L(V)^*$ surjektiv und $\tilde{L}^* : L(V)^* \to V^*$ injektiv. Daraus folgt:

$$\dim(\mathrm{im}\, L^*) = \dim(\mathrm{im}\, \tilde{L}^*) = \dim(L(V)^*).$$

Da $L(V) \subseteq W$ endlichdimensional ist, gilt $\dim(L(V)^*) = \dim L(V)$, vgl. (6.2). Daraus folgt

$$\dim(\mathrm{im}\, L^*) = \dim L(V) = \dim(\mathrm{im}\, L).$$

(6.11) Folgerung. Für jede Matrix A gilt $\mathrm{rg}(A) = \mathrm{rg}(A^T)$.

Bem.: Da die Spalten von A^T gerade die Zeilen von A sind, besagt (6.11), daß die Maximalzahl linear unabhängiger Spaltenvektoren von A gleich der Maximalzahl linear unabhängiger Zeilenvektoren von A ist. Diese Aussage drückt man oft kurz durch "Spaltenrang = Zeilenrang" aus.

Bew.: Wir wählen K-Vektorräume V, W mit geordneten Basen $\mathcal{G}_V$, $\mathcal{G}_W$ und $L \in \mathrm{Hom}(V, W)$ mit

$$\mathrm{Mat}_{\mathcal{G}_V}^{\mathcal{G}_W}(L) = A,$$

vgl. (4.15). Ist $A \in K^{m \times n}$, so gilt $\dim V = n$, $\dim W = m$ $(< \infty)$. Aus (6.6) folgt

$$\mathrm{Mat}_{\mathcal{G}_W^*}^{\mathcal{G}_V^*}(L^*) = A^T$$

und (4.24) impliziert $\mathrm{rg}(A) = \dim(\mathrm{im}\, L)$ und $\mathrm{rg}(A^T) = \dim(\mathrm{im}\, L^*)$. Aus (6.10) folgt nun

$$\mathrm{rg}(A) = \dim(\mathrm{im}\, L) \overset{(6.10)}{=} \dim(\mathrm{im}\, L^*) = \mathrm{rg}(A^T).$$

7 Die Determinante

Die Determinante wird sich als Abbildung herausstellen, die jeder quadratischen Matrix A (über einem Körper K) ein Element $\det A \in K$ zuordnet, und zwar so, daß

$$A \in \mathrm{GL}_n(K) \Leftrightarrow \det A \neq 0$$

gilt. Etwas allgemeiner werden wir jedem Endomorphismus L eines endlichdimensionalen K-Vektorraums seine Determinante $\det L \in K$ zuordnen, wobei $\det L \neq 0$ genau die Automorphismen L charakterisiert. Im Fall $K = \mathbb{R}$ hängt die Determinante eng mit dem Volumenbegriff zusammen, und dieser Zusammenhang wird hier benutzt, um die Determinante und ihre Eigenschaften zu motivieren:

Es bezeichne V einen n-dimensionalen $\mathbb{R}$-Vektorraum. Sind $v_1, \ldots, v_n \in V$, so heißt

$$P(v_1, \ldots, v_n) := \left\{ \sum_{i=1}^{n} t_i v_i \,\middle|\, \forall i \in \{1, \ldots, n\} : 0 \leq t_i \leq 1 \right\}$$

das von $v_1, \dots, v_n$ aufgespannte Parallelotop. Sind die $v_1, \dots, v_n$ linear abhängig, so liegt $P(v_1, \dots, v_n)$ in dem echten Untervektorraum $\text{span}\{v_1, \dots, v_n\}$ von V, und dann nennen wir $P(v_1, \dots, v_n)$ ein ausgeartetes Parallelotop.

Bsp.: Ist $n = 2$, so ist ein nicht ausgeartetes Parallelotop gerade ein Parallelogramm, dessen eine Ecke der 0-Punkt ist. Ist $n = 3$, so nennt man ein Parallelotop auch einen Spat.

Es ist wichtig zu erkennen, daß wir in dieser Situatio nicht einfach von "dem (n-dimensionalen) Volumen eines Parallelotops" sprechen können, sondern daß wir hierfür ein neues mathematisches Objekt benötigen, mit dessen Hilfe dieses Volumen definiert werden kann. Ähnlich verhält es sich mit der "Länge von Vektoren", die ebenfalls nicht in den Vektorraumaxiomen auftritt, und deshalb (im Fall $K = \mathbb{R}$) erst (mit Hilfe einer "Norm" oder eines "Skalarprodukts", siehe LA II) definiert werden muß.

Zur Definition eines Volumens von Parallelotopen werden wir folgendermaßen vorgehen: Wir werden eine Reihe von Eigenschaften aufstellen, die dieser Volumenbegriff haben sollte, und dann zeigen, daß durch diese Eigenschaften das Volumen von Parallelotopen schon eindeutig (bis auf einen Normierungsfaktor) bestimmt ist. Weiter erhalten wir so auch Rechenmethoden zur Bestimmung dieses Volumens (und insbesondere die "Determinante"). Die geforderten Eigenschaften sind dabei unserer Anschauung und unserer Erfahrung mit einfacheren Situationen (etwa $n = 2$ oder $n = 3$) entnommen.

Bem.: Dieses Vorgehen unterscheidet sich grundlegend vom Ansatz der Schulmathematik. Dort nimmt man an, daß ein Flächen- oder Rauminhaltsbegriff mit (meist nicht explizit genannten) "offensichtlichen" Eigenschaften existiert und beschäftigt sich damit, diese Flächen - oder Rauminhalte für einfache Klassen von Figuren formelmäßig zu berechnen. Die (nicht immer mathematisch vollständige) Herleitung dieser Formeln beruht auf den erwähnten "offensichtlichen" Eigenschaften. Der Vorteil unseres Vorgehens besteht einerseits darin, daß es mathematisch exakt ist und insbesondere echte Beweise möglich macht, und andererseits in der Möglichkeit, auch Situationen zu behandeln, die der unmittelbaren Anschauung nicht zugänglich sind (wie etwa den Fall $\dim V = n > 3$).

Zunächst müssen wir uns überlegen, was für ein mathematisches Objekt (Teilmenge von ?, Abbildung von ? nach ?) wir benutzen können, um ein Volumen für Parallelotope zu definieren. Da ein Parallelotop durch n Vektoren $v_1, \ldots, v_n \in V$ gegeben ist, wird ein Volumenbegriff zu einer Funktion führen, die den Vektoren $v_1, \ldots, v_n$ eine reelle Zahl (nämlich das Volumen von $P(v_1, \ldots, v_n)$) zuordnet, d.h. zu einer Funktion

$$D : \underbrace{V \times \ldots \times V}_{n \text{ Faktoren}} = V^n \to \mathbb{R}. \tag{1}$$

Hier könnte (und sollte) man einwenden, daß das Volumen ja nichtnegativ sein sollte, so daß der Zielbereich von D nicht $\mathbb{R}$ sondern $\mathbb{R}_{\geq 0}$ sein sollte. Es stellt sich allerdings später heraus, daß das natürliche mathematische Objekt vorzeichenbehaftete Volumina definiert, wobei das Vorzeichen mit der “Orientierung” des Vektorsystems $(v_1, \ldots, v_n)$ (im Fall $n = 3$: Links- oder Rechtssystem) zusammenhängt. Wir stellen nun folgende Forderungen an die Funktion D:

(2) Für alle $(v_1, \ldots, v_n) \in V^n$, alle $i \in \{1, \ldots, n\}$ und $r \in \mathbb{R}$ soll gelten:

$$D(v_1, \ldots, v_{i-1}, rv_i, v_{i+1}, \ldots, v_n) = rD(v_1, \ldots, v_i, \ldots, v_n).$$

(3) Für alle $1 \leq i < j \leq n$ und alle $(v_1, \ldots, v_n) \in V^n$ soll gelten:

$$\begin{aligned} D(v_1, \ldots, \overset{i\text{'te Stelle}}{v_i}, \ldots, \overset{j}{v_j}, \ldots, v_n) &= D(v_1, \ldots, v_i \overset{i}{+} v_j, \ldots, \overset{j}{v_j}, \ldots, v_n) \\ &= D(v_1, \ldots, \overset{i}{v_i}, \ldots, v_i \overset{j}{+} v_j, \ldots, v_n). \end{aligned}$$

Dabei hat (2) die anschauliche Bedeutung, daß sich das Volumen von $P(v_1,$ $\ldots$, $v_n)$ um den Faktor r ändern sollte, wenn man eine Kante v_i um den Faktor r verlängert bzw. verkürzt, d.h. wenn man v_i durch rv_i ersetzt. Die anschauliche Bedeutung der Forderung (3) ist die “Scherungsinvarianz” des Volumens, die man sich etwa an Parallelogrammen (Fall $n = 2$) illustrieren kann.

Nach diesen langen motivierenden Vorüberlegungen kommen wir jetzt zu den mathematischen Definitionen.

(7.1) Def.: Sei V ein K-Vektorraum und $k \in \mathbb{N}_{>0}$. Eine Abbildung $f : V^k \to K$ heißt

(a) <u>k-linear</u> (k-Linearform auf V), falls f in jedem Argument $v_i \in V$ linear ist, d.h. ist $i \in \{1, \ldots, k\}$ und sind $v_1, \ldots, v_{i-1}$, v, w, $v_{i+1}, \ldots, v_k \in V$ und α, $\beta \in K$, so gilt

$$\begin{aligned} f(v_1, \ldots, v_{i-1}, \alpha v + \beta w, v_{i+1}, \ldots, v_k) &= \alpha f(v_1, \ldots, \overset{i}{v}, \ldots, v_k) \\ &+ \beta f(v_1, \ldots, \overset{i}{w}, \ldots, v_k). \end{aligned}$$

(b) <u>alternierend</u> (oder schiefsymmetrisch), falls für alle $1 \leq i < j \leq k$ und alle $(v_1, \ldots, v_k) \in V^k$ gilt:

$$v_i = v_j \Rightarrow f(v_1, \ldots, v_k) = 0.$$

Bsp.: 1-Linearformen sind gerade die im letzten Kapitel auftretenden Linearformen, d.h. die Elemente von V^*. 2-Linearformen werden Bilinearformen genannt und sie werden in der LA II eine wichtige Rolle spielen. Explizite Beispiele von Bilinearformen auf $\mathbb{R}^2$ sind etwa das “Standardskalarprodukt” $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$

$$g((x_1, x_2), (y_1, y_2)) := x_1 y_1 + x_2 y_2$$

oder die “Standarddeterminantenform” $D_0 : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$

$$D_0((x_1, x_2), (y_1, y_2)) := x_1 y_2 - x_2 y_1.$$

Dabei ist D_0 alternierend, während g nicht alternierend (sondern “symmetrisch”) ist.

(7.2) Lemma. Sei V K-Vektorraum, $\dim V = n$, und $f : V^n \to K$. Dann gilt: f erfüllt genau dann die Eigenschaften (2) und (3) (für K statt $\mathbb{R}$), wenn f eine alternierende n-Linearform ist.

Ein Beweis von (7.2) wurde nicht vorgeführt, kann aber in R. Walter, Einführung in die Lineare Algebra, S. 146, Satz D, nachgelesen werden. Blatt 13, Aufgabe 4, behandelt den Fall $\dim V = 2$.

Die Fragen nach Existenz, Eindeutigkeit, expliziter Berechenbarkeit eines Volumenbegriffs für Parallelotope mit den Eigenschaften (1) - (3) führen also zu folgenden Fragen:

Gibt es alternierende n-Formen $f \neq 0$ auf V und, wenn ja, wie viele? Können wir eine explizite Formel für $f(v_1, \ldots, v_n)$ finden?

Ganz analog zur Vektorraumstruktur von $\mathrm{Hom}(V, W)$ und V^* kann man die Menge der k-Linearformen auf V zu einem K-Vektorraum machen, der die alternierenden k-Linearformen als Unterraum enthält. Als Antwort auf die 1. Frage werden wir in (7.12) zeigen, daß die alternierenden n-Formen auf einem n-dimensionalen Vektorraum einen 1-dimensionalen Vektorraum bilden. Der Weg zu dieser Erkenntnis ist mit einigen Vorüberlegungen gepflastert.

(7.3) Lemma. Für jede alternierende k-Linearform $f : V^k \to K$ gilt:

(a) Ist $1 \leq i < j \leq n$ und vertauscht man das i'te und das j'te Argument von f, so ändert f sein Vorzeichen, d.h.

$$\begin{aligned} & f(v_1, \ldots, v_{i-1}, \overset{i}{v}, \ldots, v_{j-1}, \overset{j}{w}, \ldots, v_k) \\ = & -f(v_1, \ldots, v_{i-1}, \overset{i}{w}, \ldots, v_{j-1}, \overset{j}{v}, \ldots, v_k). \end{aligned}$$

(b) Sind $v_1, \ldots, v_k$ linear abhängig, so gilt $f(v_1, \ldots, v_k) = 0$.

(c) Ist $1 \leq i \neq j \leq k$, $v_1, \ldots, v_n \in V$ und $\lambda \in K$, so gilt

$$f(v_1, \ldots, v_{i-1}, v_i + \lambda v_j, v_{i+1}, \ldots, v_n) = f(v_1, \ldots, v_i, \ldots, v_n).$$

Bew.: (a)

$$\begin{aligned} 0 \overset{f \text{alt.}}{=} & \ f(v_1, \ldots, v_{i-1}, v \overset{i}{+} w, \ldots, v_{j-1}, v + w, \ldots, v_n) \overset{f \text{ linear im i'ten Argument}}{=} \\ & \ f(v_1, \ldots, v_{i-1}, \overset{i}{v}, \ldots, v_{j-1}, v \overset{j}{+} w, \ldots, v_n) \\ + & \ f(v_1, \ldots, v_{i-1}, \overset{i}{w}, \ldots, v_{j-1}, v \overset{j}{+} w, \ldots, v_n) \\ = & \ f(v_1, \ldots, v, \ldots, v, \ldots, v_n) + f(v_1, \ldots, v, \ldots, w, \ldots, v_n) \\ + & \ f(v_1, \ldots, w, \ldots, v, \ldots, v_n) + f(v_1, \ldots, w, \ldots, w, \ldots, v_n) \\ \overset{f \text{alt.}}{=} & \ f(v_1, \ldots, v, \ldots, w, \ldots, v_n) + f(v_1, \ldots, w, \ldots, v, \ldots, v_n). \end{aligned}$$

(b) Sind $v_1, \ldots, v_k$ linear abhängig, so existiert ein $i \in \{1, \ldots, k\}$ und Skalare α_j für $j \in \{1, \ldots, n\} \setminus \{i\}$, so daß gilt:

$$v_i = \sum\nolimits_{\substack{j=1 \\ j \neq i}}^{n} \alpha_j v_j.$$

Daraus folgt:

$$\begin{aligned} f(v_1, \ldots, v_i, \ldots, v_k) &= f(v_1, \ldots, v_{i-1}, \textstyle\sum_{\substack{j=1 \\ j \neq i}}^n \alpha_j v_j, v_{i+1}, \ldots, v_n) \\ &= \textstyle\sum_{\substack{j=1 \\ j \neq i}}^n \alpha_j f(v_1, \ldots, \overset{i}{v_j}, \ldots, v_n) \\ &\overset{f \text{alt.}}{=} 0 \end{aligned}$$

(c) Man benutzt die Linearität von f im j'ten Argument und wendet die Alternierungseigenschaft von f an.

Wir wollen nun (7.3) verallgemeinern und herausfinden, wie sich $f(v_1, \ldots, v_k)$ verändert, wenn man die Argumente $v_1, \ldots, v_k$ einer beliebigen Permutation unterwirft (statt - wie in (7.3)(a) - nur zwei der Argumente zu vertauschen). Dazu dienen folgende Vorüberlegungen über Permutationen.

Wir erinnern an Bsp. 3) nach (2.1), die symmetrische (oder Permutations-) Gruppe S_n von $\{1, \ldots, n\}$:

$$S_n = \{\sigma \mid \sigma : \{1, \ldots, n\} \to \{1, \ldots n\} \text{ bijektiv}\}.$$

Die Gruppenstruktur ist die Hintereinanderausführung $\circ$.

Bez.: Eine Permutation $\sigma \in S_n$ heißt Transposition, falls es zwei Zahlen $i \neq j$ in $\{1, \ldots, n\}$ gibt, so daß gilt:

$$\sigma(i) = j, \sigma(j) = i \text{ und } \sigma(k) = k \text{ für alle } k \in \{1, \ldots, n\} \setminus \{i, j\}.$$

Bem.: Ist f alternierende k-Linearform auf V und $\sigma \in S_k$ eine Transposition, so läßt sich (7.3)(a) umformulieren als:

$$f(v_1, \ldots, v_k) = -f(v_{\sigma(1)}, \ldots, v_{\sigma(k)}).$$

In Analogie zum Summenzeichen $\sum$ führen wir für das Produkt von endlich vielen Körperelementen $a_1, \ldots, a_n$ die Abkürzung

$$\prod_{i=1}^n a_i := a_1 \cdot \ldots \cdot a_n$$

ein. Wegen der Kommutativität von $+$ und $\cdot$ gilt für jedes $\sigma \in S_n$:

$$\sum_{i=1}^n a_{\sigma(i)} = \sum_{i=1}^n a_i \text{ und } \prod_{i=1}^n a_{\sigma(i)} = \prod_{i=1}^n a_i.$$

(7.4) Def.: Ist $\sigma \in S_n$, so heißt

$$\operatorname{sgn}(\sigma) := \prod_{1\leq i<j\leq n} \frac{\sigma(j)-\sigma(i)}{j-i}$$

das Signum von σ.

Bsp.: 1) $\sigma = \mathrm{id}_{\{1,\ldots,n\}} \Rightarrow \operatorname{sgn}(\sigma) = 1$
2) Ist $\sigma \in S_2$ definiert durch $\sigma(1) := 2$, $\sigma(2) := 1$, so gilt

$$\operatorname{sgn}(\sigma) = \frac{1-2}{2-1} = -1.$$

Bem.: 1) Da in $\prod\limits_{1\leq i<j\leq n} \frac{\sigma(j)-\sigma(i)}{j-i}$ die Faktoren im Zähler bis auf Anordnung und Vorzeichen mit den Faktoren im Nenner übereinstimmen, gilt $\operatorname{sgn}(\sigma) \in \{1,-1\}$.
2) Die Fehlstandszahl $f(\sigma)$ einer Permutation $\sigma \in S_n$ ist die Anzahl der Paare (i,j) mit $1 \leq i < j \leq n$, für die $\sigma(i) > \sigma(j)$ gilt. Die vorangehende Überlegung zeigt, daß wir für jedes dieser Paare (nach dem Kürzen) eine -1 im Produkt $\prod\limits_{1\leq i<j\leq n} \frac{\sigma(j)-\sigma(i)}{j-i}$ erhalten, während alle anderen Paare zu einer $+1$ führen. Es gilt deshalb

$$\operatorname{sgn}(\sigma) = (-1)^{f(\sigma)}.$$

Bez.: Eine Permutation $\sigma \in S_n$ heißt <u>gerade</u>, falls $\operatorname{sgn}(\sigma) = 1$ gilt, sonst <u>ungerade</u>.

(7.5) Fakt. Ist $\sigma \in S_n$ eine Transposition, so gilt $\operatorname{sgn}(\sigma) = -1$.

Bew.: Vertauscht σ die Zahlen i und j mit $1 \leq i < j \leq n$, so hat σ genau $2(j-i)-1$ Fehlstände, also:

$$\operatorname{sgn}(\sigma) = (-1)^{2(j-i)-1} = -1.$$

(7.6) Satz. *Für alle* σ, $\tau \in S_n$ *gilt* $\operatorname{sgn}(\sigma \circ \tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$, *d.h.* σ *ist Homomorphismus von* $(S_n, \circ)$ *in die Gruppe(!)* $(\{\pm 1\}, \cdot)$.

Bew.: Vorbemerkung: Wegen $\frac{\sigma(j)-\sigma(i)}{j-i} = \frac{\sigma(i)-\sigma(j)}{i-j}$ kann man in der Definition von $\operatorname{sgn}(\sigma)$, statt über alle Paare (i,j) mit $1 \le i < j \le n$ zu multiplizieren, über irgendeine Menge von Paaren (i,j) multiplizieren, die die Eigenschaft hat, daß die Menge der zugehörigen Paarmengen $\{i,j\}$ die Menge aller 2-elementigen Teilmengen von $\{1,\dots,n\}$ ist.

$$\begin{aligned}
\operatorname{sgn}(\sigma \circ \tau) &= \prod_{1\le i<j\le n} \tfrac{\sigma(\tau(j))-\sigma(\tau(i))}{j-i} \\
&= \underbrace{\prod_{1\le i<j\le n} \frac{\sigma(\tau(j))-\sigma(\tau(i))}{\tau(j)-\tau(i)}}_{\substack{=\operatorname{sgn}(\sigma)\\ \text{(vgl. Vorbemerkung)}}} \quad \underbrace{\prod_{1\le i<j\le n} \frac{\tau(j)-\tau(i)}{j-i}}_{=:\operatorname{sgn}(\tau)}
\end{aligned}$$

(7.7) Folgerung. (a) Für alle $\sigma \in S_n$ gilt $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$.
(b) Sei $\tau \in S_n$ ungerade. Dann bildet die Abbildung $\sigma \in S_n \to \tau \circ \sigma \in S_n$ die Menge der geraden Permutationen bijektiv auf die Menge der ungeraden Permutationen ab.

Bew.: (a) $1 = \operatorname{sgn}(\mathrm{id}_{\{1,\dots,n\}}) = \operatorname{sgn}(\sigma \circ \sigma^{-1}) \overset{(7.6)}{=} \operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma^{-1})$.

Wegen $\operatorname{sgn}(\sigma) \in \{\pm 1\}$ folgt aus $\operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma^{-1}) = 1$, daß $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$ gilt.

b) Zunächst ist zu zeigen, daß $\tau \circ \sigma$ ungerade ist, wenn σ gerade ist:
$\operatorname{sgn}(\tau \circ \sigma) \overset{(7.6)}{=} \operatorname{sgn}(\tau) \cdot \operatorname{sgn}(\sigma) = (-1) \cdot 1 = -1$.

Die Abbildung $\sigma \to \tau \circ \sigma$ ist injektiv, da aus $\tau \circ \sigma_1 = \tau \circ \sigma_2$ folgt:

$$\tau^{-1} \circ (\tau \circ \sigma_1) = \sigma_1 = \tau^{-1} \circ (\tau \circ \sigma_2) = \sigma_2.$$

Zum Nachweis der Surjektivität sei $\sigma \in S_n$ eine beliebige ungerade Permutation. Wegen (a) ist τ^{-1} ungerade und mit (7.6) folgt, daß $\tilde{\sigma} := \tau^{-1} \circ \sigma$ gerade ist. Die gerade Permutation $\tilde{\sigma}$ wird auf $\tau \circ \tilde{\sigma} = \tau \circ (\tau^{-1} \circ \sigma) = \sigma$ abgebildet.

(7.8) Fakt. Ist $n \ge 2$ und $\sigma \in S_n$, so existiert ein $j \in \mathbb{N}$ und Transpositionen $\tau_1, \dots, \tau_j$, so daß $\sigma = \tau_j \circ \ldots \circ \tau_1$ gilt. Es gilt dann $\operatorname{sgn}(\sigma) = (-1)^j$.

Bew.: Die erste Aussage kann mit vollständiger Induktion nach n bewiesen werden (Blatt 13, Aufgabe 1). Dann folgt $\operatorname{sgn}(\sigma) = (-1)^j$ per Induktion aus (7.6).

(7.9) Lemma. Sei V K-Vektorraum und $f : V^k \to K$ k-linear und alternierend. Dann gilt für alle $(v_1, \ldots, v_k) \in V^k$ und alle $\sigma \in S_k$:

$$f(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = \operatorname{sgn}(\sigma) f(v_1, \ldots, v_k).$$

Bew.: Nach (7.8) können wir σ als $\sigma = \tau_j \circ \ldots \circ \tau_1$ mit Transpositionen $\tau_1, \ldots, \tau_j \in S_k$ darstellen, d.h. wir erhalten $(v_{\sigma(1)}, \ldots, v_{\sigma(k)})$ aus $(v_1, \ldots, v_k)$, indem wir sukzessive j mal zwei Elemente der Argumentliste miteinander vertauschen. Nach (7.3)(a) führt jede dieser Vertauschungen zu einem Vorzeichenwechsel des Wertes von f. Nach j Schritten erhalten wir:

$$f(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = (-1)^j f(v_1, \ldots, v_k).$$

Andererseits gilt $(-1)^j = \operatorname{sgn}(\sigma)$, vgl. (7.8).

(7.10) Satz. *Sei V K-Vektorraum,* $\dim V = n$, $(v_1, \ldots, v_n)$ *geordnete Basis von V und $\alpha \in K$. Dann gibt es genau eine alternierende n-Linearform $f : V^n \to K$ mit $f(v_1, \ldots, v_n) = \alpha$ und für dieses f gilt: Ist $(w_1, \ldots, w_n) \in V^n$, $w_j = \sum_{i=1}^{n} a_{ij} v_i$ für $1 \le j \le n$, so*

$$(*) \qquad f(w_1, \ldots, w_n) = \alpha \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \cdot a_{\sigma(2)2} \cdot \ldots \cdot a_{\sigma(n)n}.$$

Speziell gilt: Ist $f(v_1, \ldots, v_n) = 0$ für eine Basis $(v_1, \ldots, v_n)$, so ist $f \equiv 0$.

Bew.: (a) Eindeutigkeit: Sei f alternierende n-Form, $f(v_1, \ldots, v_n) = \alpha$. Dann gilt

$$\begin{aligned}
f(w_1, \ldots, w_n) &= f\left(\sum_{i_1=1}^{n} a_{i_1 1} v_{i_1}, \sum_{i_2=1}^{n} a_{i_2 2} v_{i_2}, \ldots, \sum_{i_n=1}^{n} a_{i_n n} v_{i_n} \right) \\
&\overset{f\ n\text{-linear}}{=} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \ldots \sum_{i_n=1}^{n} a_{i_1 1} \cdot a_{i_2 2} \cdot \ldots \cdot a_{i_n n} \quad \underbrace{f(v_{i_1}, v_{i_2}, \ldots, v_{i_n})}_{=0,\text{ falls } k \to i_k \text{ nicht injektiv}} \\
&\overset{f\ \text{alternierend}}{=} \sum_{\sigma \in S_n} a_{\sigma(1)1} \cdot a_{\sigma(2)2} \cdot \ldots \cdot a_{\sigma(n)n} f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) \\
&\overset{(7.9)}{=} \alpha \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \cdot \ldots \cdot a_{\sigma(n)n}
\end{aligned}$$

(b) Wir zeigen: $\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdot \ldots \cdot a_{n\sigma(n)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \cdot \ldots \cdot a_{\sigma(n)n}$.

Denn:

$$\sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdot \ldots \cdot a_{n\sigma(n)}$$

$$\overset{\sigma\in S_n\to\sigma^{-1}\in S_n \text{ bijektiv}}{=} \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma^{-1}) \underbrace{a_{1\sigma^{-1}(1)} \cdot \ldots \cdot a_{n\sigma^{-1}(n)}}_{=\prod_{i=1}^n b_i \text{ für } b_i := a_{i\sigma^{-1}(i)}}$$

$$\overset{(7.7)(a)}{=} \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \underbrace{a_{\sigma(1)1} \cdot \ldots \cdot a_{\sigma(n)n}}_{=\prod_{i=1}^n b_{\sigma(i)}}$$

(c) Existenz: Definiere f durch $(*)$. Man sieht leicht, daß f n-linear ist. Außerdem gilt wegen $v_j = \sum\limits_{i=1}^n \delta_{ij} v_i$ (mit $\delta_{ij} = 1$ für $i = j$, $\delta_{ij} = 0$ für $i \neq j$):

$$f(v_1, \ldots, v_n) = \alpha \sum \operatorname{sgn}(\sigma) \delta_{\sigma(1)1} \cdot \ldots \cdot \delta_{\sigma(n)n} = \alpha \operatorname{sgn}(\mathrm{id}) = \alpha.$$

Bleibt zu zeigen, daß $f(w_1, \ldots, w_n) = 0$ gilt, falls für ein Paar (i, j) mit $1 \leq i < j \leq n$ gilt: $w_i = w_j$. Ist $w_k = \sum\limits_{l=1}^n a_{lk} v_l$, $1 \leq k \leq n$, so folgt aus $w_i = w_j$:

$$(**) \qquad a_{li} = a_{lj} \text{ für } 1 \leq l \leq n.$$

Sei τ die Transposition, die i und j vertauscht.

$$f(w_1, \ldots, w_n) \overset{(b)}{=} \alpha \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdot \ldots \cdot a_{n\sigma(n)} \overset{(7.7)(b)}{=}$$
$$= \alpha \sum_{\sigma\in S_n \text{gerade}} a_{1\sigma(1)} \cdot \ldots \cdot a_{n\sigma(n)} - \alpha \sum_{\sigma\in S_n \text{gerade}} a_{1\tau\circ\sigma(1)} \cdot \ldots \cdot a_{n\tau\circ\sigma(n)} = 0,$$

da aus $(**)$ folgt: $a_{l\sigma(l)} = a_{l\tau\circ\sigma(l)}$ für alle $l \in \{1, \ldots, n\}$ und alle $\sigma \in S_n$.

(7.11) Def.: Sei V K-Vektorraum, $\dim V = n$. Eine alternierende n-Linearform $D \neq 0$ auf V heißt Determinantenform auf V.

Erinnerung: $D \neq 0$ bedeutet: Es existiert $(v_1, \ldots, v_n) \in V^n$: $D(v_1, \ldots, v_n) \neq 0$.

Bsp.: $D_0 : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, $D_0((x_1, x_2), (y_1, y_2)) = x_1 y_2 - x_2 y_1$. Dann gilt:

$$D_0(e_1, e_2) = 1$$

(7.12) Folgerung. Sei D Determinantenform auf dem n-dimensionalen K-Vektorraum V und $(v_1, \ldots, v_n) \in V^n$. Dann gilt

(a) $(v_1, \ldots, v_n)$ Basis von $V \Leftrightarrow D(v_1, \ldots, v_n) \neq 0$.

(b) Ist $f : V^n \to K$ alternierende n-Linearform, so existiert $\alpha \in K$:

$$f = \alpha D.$$

Bem.: 1) (7.12)(b) besagt, daß der K-Vektorraum der alternierenden n-Linearformen auf V (mit $\dim V = n$) 1-dimensional ist. Die Determinantenformen sind die Elemente $\neq 0$ dieses Vektorraumes.

2) $f = \alpha D$, $(v_1, \ldots, v_n)$ Basis von $V \Rightarrow \alpha = \frac{f(v_1,\ldots,v_n)}{D(v_1,\ldots,v_n)}$

Bew.: (a) Sei $(v_1, \ldots, v_n)$ Basis von V. Dann folgt aus (7.10) und $D \neq 0$, daß $D(v_1, \ldots, v_n) \neq 0$ gilt. Ist $D(v_1, \ldots, v_n) \neq 0$, so folgt aus (7.3)(b), daß die $v_1, \ldots, v_n$ linear unabhängig sind, also - wegen $\dim V = n$ - eine Basis von V bilden.

(b) Sei $(v_1, \ldots, v_n)$ eine Basis von V und $\alpha := \frac{f(v_1,\ldots,v_n)}{D(v_1,\ldots,v_n)}$. Dann sind f und αD alternierende n-Linearformen, die auf der Basis $(v_1, \ldots, v_n)$ den gleichen Wert annehmen. Also folgt aus (7.10) (Eindeutigkeit): $f = \alpha D$.

Einschub: Die Determinante quadratischer Matrizen.

In K^n haben wir die Standardbasis $(e_1, \ldots, e_n)$, die - nach (7.10) - eine "Standarddeterminantenform" $D_0 : K^n \times \ldots \times K^n \to K$ durch

$$D_0(e_1, \ldots, e_n) = 1$$

eindeutig bestimmt. Ist $A = (a_{ij})_{1 \leq i,j \leq n} \in K^{n \times n}$, so seien $A_j := \sum\limits_{i=1}^{n} a_{ij} e_i$ die "Spaltenvektoren" von A.

(7.13) Def. Die Determinante $\det : K^{n \times n} \to K$ ist definiert durch

$$\det A := D_0(A_1, \ldots, A_n)$$

Bezeichnungsweise: $\det A = \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}$

(7.14) Folgerung (Gottfried Wilhelm Leibniz 1646-1716). Ist $A = (a_{ij})_{1\leq i,j\leq n} \in K^{n\times n}$, so gilt

$$\begin{aligned}\det A &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1}\cdot\ldots\cdot a_{\sigma(n)n}\\ &= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)}\cdot\ldots\cdot a_{n\sigma(n)}.\end{aligned}$$

Bew.:
Die erste Gleichung folgt aus (7.10)(∗), wenn man $\det(A) = D_0(A_1,\ldots,A_n)$, $A_j = \sum\limits_{i=1}^{n} a_{ij}e_i$ und $D_0(e_1,\ldots,e_n) = 1$ benutzt. Die zweite Gleichung ist gerade Teil (b) des Beweises von (7.10).

Speziell: $n = 1: \quad A = (a_{11}) \to \quad \det A = a_{11}$
$n = 2: \quad A = \binom{a_{11}a_{12}}{a_{21}a_{22}} \to \quad \det A = a_{11}a_{22} - a_{12}a_{21}$

$n = 3$: Die Regel von Sarrus (Pierre Frédéric Sarrus 1798-1861, franz. Mathematiker) ist folgende Merkregel zur Berechnung der Determinante von (3×3)-Matrizen $A = (a_{ij}) \in K^{3\times 3}$.

Spezialisiert man (7.14) auf den Fall $n = 3$, so erhält man:

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

Diese Formel kann man sich anhand folgenden Schemas merken:

$$\begin{array}{ccccccccc} a_{11} & & a_{12} & & a_{13} & & a_{11} & & a_{12}\\ & \diagdown & & \times & & \times & & \diagup & \\ a_{21} & & a_{22} & & a_{23} & & a_{21} & & a_{22}\\ & \diagup & & \times & & \times & & \diagdown & \\ a_{31} & & a_{32} & & a_{33} & & a_{31} & & a_{32}\end{array}$$

(7.15) Satz *(einige Eigenschaften der Determinante).*

(a) $\det E_n = 1$

(b) $\det A \neq 0 \Leftrightarrow A \in \mathrm{GL}_n(K)$

(c) $\det A = \det(A^T)$

(d) $\det A$ *ist n-linear und alternierend in den Spaltenvektoren von A.*

(e) $\det A$ *ist n-linear und alternierend in den Zeilenvektoren von A.*

Bew.: (a) $\det E_n = D_0(e_1, \ldots, e_n) = 1$

(b) $\det A \neq 0 \Leftrightarrow D_0(A_1, \ldots, A_n) \neq 0 \overset{(7.12)(a)}{\Leftrightarrow} A_1, \ldots, A_n$ linear unabhängig

$$\overset{(4.23)}{\Leftrightarrow} \operatorname{rg}(A) = n \overset{(5.5)}{\Leftrightarrow} A \in \mathrm{GL}_n(K)$$

(c) Die 2. Gleichung in (7.14) impliziert $\det(A) = \det(A^T)$.

(d) Dies ist eine Umformulierung von $\det(A) = D_0(A_1, \ldots, A_n)$ und der Tatsache, daß D_0 n-linear und alternierend ist.

(e) Da die Zeilenvektoren von A gerade die Spaltenvektoren von A^T sind, folgt (e) aus (c) und (d).

Es folgen einige Konsequenzen von Satz (7.15):

Aus (7.15)(b) und (c) folgt:

$$\begin{aligned} \det A = 0 \quad &\Leftrightarrow \quad \text{Die Spaltenvektoren von } A \text{ sind linear abhängig.} \\ &\Leftrightarrow \quad \text{Die Zeilenvektoren von } A \text{ sind linear abhängig.} \end{aligned}$$

Aus (7.15)(d) folgt:

$$\begin{vmatrix} a_{11} & \ldots & \alpha a_{1j} & \ldots & a_{n1} \\ \vdots & & & & \vdots \\ a_{m1} & \ldots & \alpha a_{nj} & \ldots & a_{nn} \end{vmatrix} = \alpha \begin{vmatrix} a_{11} & \ldots & a_{n1} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}$$

Induktiv folgt daraus:

$$\det(\alpha A) = \alpha^n \det A.$$

Vertauscht man 2 Spalten von A, so ändert sich das Vorzeichen der Determinante. Addiert man zu einer Spalte eine Linearkombination der anderen Spalten, so ändert sich $\det A$ nicht, z.B.

$$\begin{vmatrix} a_{11} + \alpha a_{12} & a_{12} & \ldots & a_{1n} \\ a_{21} + \alpha a_{22} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} + \alpha a_{n2} & a_{n2} & \ldots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix}$$

Aus (7.15)(e) folgen die analogen Aussagen für die Zeilen.

(7.16) Berechnung von Determinanten (Gaußscher Algorithmus)

Es wird ein Verfahren vorgestellt, das mit relativ geringem Rechenaufwand die Berechnung von Determinanten ermöglicht. Es ist eng mit dem Gaußschen Eliminationsverfahren (siehe 3. Kapitel) verwandt.

Zunächst wird gezeigt, daß die Berechnung der Determinante für obere (oder untere) Dreiecksmatrizen sehr leicht ist: Die Determinante ist das Produkt der Diagonalelemente. Wir erinnern daran, daß eine Matrix $A = (a_{ij}) \in K^{n\times n}$ eine obere (bzw. untere) Dreiecksmatrix genannt wird, falls $a_{ij} = 0$ für alle $1 \leq j < i \leq n$ (bzw. für alle $1 \leq i < j \leq n$) gilt.

(7.17) Lemma. Ist $A = (a_{ij}) \in K^{n\times n}$ eine obere oder eine untere Dreiecksmatrix, so gilt

$$\det A = a_{11} \cdot \ldots \cdot a_{nn}.$$

Bew.: Gilt $a_{ij} = 0$ für alle $1 \leq j < i \leq n$, so gilt für jede Permutation $\sigma \in S_n$:

$$\begin{aligned} a_{\sigma(1)1} \cdot \ldots \cdot a_{\sigma(n)n} \neq 0 \quad &\Rightarrow \quad \sigma(i) \leq i \text{ für alle } i \in \{1, \ldots, n\} \\ &\Rightarrow \quad \sigma = \mathrm{id}_{\{1,\ldots,n\}} \end{aligned}$$

In der Leibnizformel (7.14) ist also höchstens der zu $\sigma = \mathrm{id}_{\{1,\ldots,n\}}$ gehörende Summand ungleich 0. Wegen $\mathrm{sgn}(\mathrm{id}_{\{1,\ldots,n\}}) = 1$ folgt daraus die Behauptung.

Das Verfahren zur Berechnung der Determinante einer beliebigen Matrix $A \in K^{n\times n}$ besteht darin, durch Addition von Vielfachen von Zeilen zu anderen Zeilen und - falls nötig - durch Vertauschen von Zeilen A in eine obere Dreiecksmatrix $\overline{A} = (\overline{a}_{ij}) \in K^{n\times n}$ zu verwandeln. Dann gilt mit (7.17), wenn k die Anzahl der ausgeführten Zeilenvertauschungen bezeichnet:

$$\det A = (-1)^k \det \overline{A} = (-1)^k \overline{a}_{11} \cdot \ldots \cdot \overline{a}_{nn}$$

Bsp.:

$$\begin{aligned} \begin{vmatrix} 0 & 2 & 3 \\ 2 & 8 & 6 \\ 7 & 8 & 10 \end{vmatrix} &= - \begin{vmatrix} 2 & 8 & 6 \\ 0 & 2 & 3 \\ 7 & 8 & 10 \end{vmatrix} \begin{matrix} \mid \cdot (-\frac{7}{2}) \\ \\ \longleftarrow + \end{matrix} \\ &= - \begin{vmatrix} 2 & 8 & 6 \\ 0 & 2 & 3 \\ 0 & -20 & -11 \end{vmatrix} \begin{matrix} \\ \mid \cdot 10 \\ \longleftarrow + \end{matrix} = - \begin{vmatrix} 2 & 8 & 6 \\ 0 & 2 & 3 \\ 0 & 0 & 19 \end{vmatrix} = -76 \end{aligned}$$

Als nächstes wird der Laplacesche Entwicklungssatz behandelt, der bei der Berechnung von Determinanten von Matrizen besonderer Bauart (vgl. z. B. Blatt 14, Aufgabe 3) und bei manchen theoretischen Überlegungen hilfreich sein kann.

(7.18) Lemma. Sei $\tilde{A} \in K^{(n-1)\times(n-1)}$ und

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{A} & \\ 0 & & & \end{pmatrix} \in K^{n\times n}\,.$$

Dann gilt

$$\det A = \det \tilde{A}.$$

Bew.: Betrachten wir $\det A$ als Funktion der Spaltenvektoren $\tilde{A}_1, \ldots, \tilde{A}_{n-1} \in K^{n-1}$ von $\tilde{A}$, so ist $\det A$ alternierende $(n-1)$-Linearform auf K^{n-1}. Für $\tilde{A} = E_{n-1} = (e_1 \ldots e_{n-1})$ gilt $\det A = 1$. Damit folgt aus der Eindeutigkeitsaussage in (7.10), daß $\det A = \det \tilde{A}$ gilt.

Bem.: Mit einem analogen Beweis kann man für alle $A \in K^{n\times n}$, $B \in K^{m\times m}$ und $C \in K^{n\times m}$ zeigen:

$$\det \begin{pmatrix} A & C \\ 0 & B \end{pmatrix} = \det A \cdot \det B,$$

vgl.Blatt 14, Aufgabe 1.

Für die Formulierung des Laplaceschen Entwicklungsstatzes ist folgende Bezeichnung wichtig. Zu $A \in K^{n\times n}$ und $1 \le i, j \le n$ bezeichne $A_{ij} \in K^{(n-1)\times(n-1)}$ die Matrix, die aus A durch Streichen der i'ten Zeile und der j'ten Spalte entsteht.

Bsp.:

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 8 & 7 & 6 & 5 \\ 4 & 3 & 2 & 1 \end{pmatrix} \Rightarrow A_{23} = \begin{pmatrix} 1 & 2 & 4 \\ 8 & 7 & 5 \\ 4 & 3 & 1 \end{pmatrix}.$$

(7.19) Entwicklungssatz. *Für alle $A \in K^{n\times n}$, $j \in \{1,\dots,n\}$ gilt:*

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

Für alle $i \in \{1,\dots,n\}$ gilt:

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

Bsp.:

$$\begin{vmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 0 & 2 \\ 3 & 1 & 1 & 1 \\ 1 & 0 & 0 & 2 \end{vmatrix} \overset{\text{Entwicklung nach der 3. Spalte}}{=} (-1)^{1+3} \begin{vmatrix} 0 & 1 & 2 \\ 3 & 1 & 1 \\ 1 & 0 & 2 \end{vmatrix} + (-1)^{3+3} \begin{vmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 2 \end{vmatrix} = -2$$

$$\overset{\text{oder Entwicklung nach der 4. Zeile}}{=} (-1)^{1+3} \begin{vmatrix} 2 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 1 \end{vmatrix} + (-1)^{4+4} \cdot 2 \cdot \begin{vmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 3 & 1 & 1 \end{vmatrix} = -2.$$

Bew.: $A_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix}, \dots, A_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{pmatrix}$ seien die Spaltenvektoren von A.

$$\begin{aligned} \det A &= D_0(A_1,\dots,A_n) = D_0(A_1,\dots,A_j = \textstyle\sum a_{ij} e_i, \dots, A_n) \\ &= \sum_{i=1}^{n} a_{ij} D_0(A_1,\dots,A_{j-1}, e_i, A_{j+1}\dots,A_n) \end{aligned}$$

Es bleibt zu zeigen: $D_0(A_1,\dots,A_{j-1}, e_i, A_{j+1}\dots,A_n) = (-1)^{i+j} \det A_{ij}$

$$D_0(A_1,\dots,A_{j-1},e_i,A_{j+1}\dots,A_n)=(-1)^{j-1}D_0(e_i,A_1,\dots,A_{j-1},A_{j+1},\dots,A_n)$$

$$= \quad (-1)^{j-1} \left|\begin{array}{ccc|cc} 0 & a_{11} & \dots & \dots & a_{1n} \\ \vdots & & & & \vdots \\ 1 & a_{i1} & \dots & \dots & a_{in} \\ \vdots & & & & \vdots \\ 0 & a_{n1} & \dots & \dots & a_{nn} \end{array}\right| = (-1)^{i+j-2} \left|\begin{array}{ccc|cc} 1 & a_{i1} & \dots & \dots & a_{in} \\ 0 & a_{11} & \dots & \dots & a_{1n} \\ \vdots & \vdots & & & \vdots \\ 0 & a_{i-1,1} & \dots & \dots & a_{i-1,n} \\ 0 & a_{i+1,1} & \dots & \dots & a_{i+1,n} \\ \vdots & \vdots & & & \vdots \\ 0 & a_{n1} & \dots & \dots & a_{nn} \end{array}\right|$$

$$= \quad (-1)^{i+j} \left|\begin{array}{ccc|cc} 1 & 0 & \dots & \dots & 0 \\ 0 & a_{11} & \dots & \dots & a_{1n} \\ \vdots & \vdots & & & \vdots \\ 0 & a_{i-1,1} & \dots & \dots & a_{i-1,n} \\ 0 & a_{i+1,1} & \dots & \dots & a_{i+1,n} \\ \vdots & \vdots & & & \vdots \\ 0 & a_{n1} & \dots & \dots & a_{nn} \end{array}\right|$$

$$\overset{(7.18)}{=} \quad (-1)^{i+j} \quad \det A_{ij}$$

(hierbei soll der senkrechte Strich in den Determinanten andeuten, daß die $(j+1)$'te Spalte gestrichen ist!)

Damit ist der Einschub über die Determinante quadratischer Matrizen und ihre Berechnung abgeschlossen. In der nächsten Definition wird der zur Motivation der Determinantenformen benutzte Zusammenhang mit dem n-dimensionalen Volumenbegriff präzisiert.

(7.20) Def.: Es sei V ein n-dimensionaler $\mathbb{R}$-Vektorraum mit Determinantenform $D : V^n \to \mathbb{R}$. Ist $(v_1,\dots,v_n)\in V^n$ und $P = P(v_1,\dots,v_n)$ das von $v_1,\dots,v_n$ erzeugte Polytop, so heißt

$$\mathrm{vol}_D(P) := |D(v_1,\dots,v_n)|$$

das <u>Volumen</u> <u>von</u> P <u>bezüglich</u> D.

Ist $V=\mathbb{R}^n$, so wählt man $D := D_0$ und nennt

$$\mathrm{vol}_n(P) := |D_0(v_1,\dots,v_n)|$$

das $\underline{n\text{-dimensionale}}$ Volumen von $P \subseteq \mathbb{R}^n$.

Bem.: 0) Für den Einheitswürfel $P = P(e_1, \ldots, e_n) \subseteq \mathbb{R}^n$ gilt:

$$\mathrm{vol}_n(P) = 1.$$

1) In Aufgabe 3 c) von Blatt 13 sollte gezeigt werden, daß für jede Basis $(v_1, \ldots, v_n)$ die Menge $P = P(v_1, \ldots, v_n) \subseteq V$ die Menge $\{v_1, \ldots, v_n\}$ eindeutig bestimmt. Mit (7.9) folgt daraus, daß $\mathrm{vol}_D(P)$ nur von P und nicht von der gewählten Darstellung $P = P(v_1, \ldots, v_n)$ abhängt.

2) Ist $P = P(v_1, \ldots, v_n)$ Parallelotop in $\mathbb{R}^n$ und $v_j = (v_{1j}, \ldots, v_{nj})$, so gilt

$$\mathrm{vol}_n(P) = \left| \det \begin{pmatrix} v_{11} & \ldots & v_{1n} \\ \vdots & & \vdots \\ v_{n1} & \ldots & v_{nn} \end{pmatrix} \right|$$

3) Ist $\tilde{D}$ eine weitere Determinantenform auf V, so existiert nach (7.12) eine $\alpha \in \mathbb{R}_{>0}$, so daß für alle Parallelotope $P \subseteq V$ gilt:

$$\mathrm{vol}_{\tilde{D}}(P) = \alpha \, \mathrm{vol}_D(P).$$

4) Ist $r \geq 0$ und $rP := \{rv | v \in P\} = P(rv_1, \ldots, rv_n)$, so gilt:

$$\mathrm{vol}_D(rP) = r^n \, \mathrm{vol}_D(P).$$

Als nächstes definieren wir die Determinante von Endomorphismen endlichdimensionaler Vektorräume über einem beliebigen Körper K.

(7.21) Def.: Sei V n-dimensionaler K-Vektorraum und $L \in \mathrm{End}(V)$. Dann ist für jede geordnete Basis $(v_1, \ldots, v_n)$ von V und jede Determinantenform D auf V der Quotient $D(L(v_1), \ldots, L(v_n))/D(v_1, \ldots, v_n)$ das selbe Element von K und

$$\det L := \frac{D(L(v_1), \ldots, L(v_n))}{D(v_1, \ldots, v_n)}$$

heißt die Determinante von L.

Es muß begründet werden, warum dieses $\det L$ nicht von der Wahl der geordneten Basis und der Wahl der Determinantenform abhängt. Zunächst folgt aus (7.12)(a), daß $D(v_1, \ldots, v_n) \neq 0$ ist, so daß $\det L$ überhaupt definiert ist. Aus (7.12)(b) folgt die Unabhängigkeit von der Wahl der Determinantenform. Um die Unabhängigkeit von der Wahl der Basis zu beweisen, betrachtet man die Abbildung

$$f : V^n \to K, f(w_1, \ldots, w_n) = D(L(w_1), \ldots, L(w_n)).$$

Dann ist f alternierende n-Linearform und (7.12)(b) zeigt die Existenz eines $\alpha \in K$, für das $f = \alpha D$ gilt. Also gilt für jede Basis $(v_1, \ldots, v_n)$ von V:

$$\frac{D(L(v_1), \ldots, L(v_n))}{D(v_1, \ldots, v_n)} = \alpha (= \det L).$$

(7.22) Satz (Eigenschaften von $\det L$). *Sei V K-Vektorraum,* $\dim V = n < \infty$, *und seien L, L_1, $L_2 \in \operatorname{End}(V)$, $\alpha \in K$. Dann gilt:*

(a) $\det(\mathrm{id}_V) = 1$.

(b) $L \in \operatorname{Aut}(V) \Leftrightarrow \det L \neq 0$.

(c) $\det(L_2 \circ L_1) = \det L_2 \cdot \det L_1$.

(d) $L \in \operatorname{Aut}(V) \Rightarrow \det(L^{-1}) = (\det L)^{-1}$.

(e) $\det(\alpha L) = \alpha^n \det L$.

Bew.: (b)

$$\begin{array}{lcl} L \in \operatorname{Aut}(V) & \Leftrightarrow & \text{Für jede Basis } (v_1, \ldots, v_n) \text{ von } V \text{ ist } (L(v_1), \ldots, L(v_n)) \\ & & \text{Basis von } V. \\ & \overset{(7.21)}{\Leftrightarrow} & \det L \neq 0. \end{array}$$

(c) 1. Fall:

$$\begin{array}{lcl} \det L_1 = 0 & \Rightarrow & (L_1(v_1), \ldots, L_1(v_n)) \text{ linear abhängig} \\ & \overset{L_2 \text{ linear}}{\Rightarrow} & (L_2(L_1(v_1)), \ldots, L_2(L_1(v_1))) \text{ linear abhängig} \\ & \Rightarrow & \det(L_2 \circ L_1) = 0 \end{array}$$

Es gilt in diesem Fall also:

$$\det L_2 \cdot \det L_1 = 0 = \det(L_2 \circ L_1).$$

2. Fall: $\det L_1 \neq 0$. Für jede Basis $(v_1, \ldots, v_n)$ von V ist dann auch

$$(L_1(v_1), \ldots, L_1(v_n))$$

eine Basis von V. Dann gilt

$$\begin{aligned}
\det(L_2 \circ L_1) &= \tfrac{D(L_2(L_1(v_1)),\ldots,L_2(L_1(v_n)))}{D(v_1,\ldots,v_n)} \\
&= \tfrac{D(L_2(L_1(v_1)),\ldots,L_2(L_1(v_n)))}{D(L_1(v_1),\ldots,L_1(v_n)))} \cdot \tfrac{D(L_1(v_1),\ldots,L_1(v_n)))}{D(v_1,\ldots,v_n)} \\
&= \det L_2 \cdot \det L_1.
\end{aligned}$$

Behauptung (d) folgt aus (a) und (c), und (e) folgt aus der n-Linearität von D.

Im Fall endlichdimensionaler reeller Vektorräume kann der Betrag $|\det L|$ der Determinante eines Endomorphismus L als Maßzahl für die *Volumenverzerrung* durch L interpretiert werden. Das spielt in der mehrdimensionalen Integration eine wichtige Rolle und wird wie folgt präzisiert:

(7.23) Fakt. Ist V endlichdimensionaler $\mathbb{R}$-Vektorraum und $L \in \operatorname{End}(V)$, so gilt für jedes Parallelotop $P \subseteq V$ und jede Determinantenform D auf V:

$$\operatorname{vol}_D(L(P)) = |\det L| \operatorname{vol}_D(P).$$

Bew.: Aus der Homomorphieeigenschaft von L folgt für alle $(v_1, \ldots, v_n) \in V^n$:

$$L(P(v_1, \ldots, v_n)) = P(L(v_1), \ldots, L(v_n)).$$

Also gilt für $P = P(v_1, \ldots, v_n)$:

$$\begin{aligned}
|\det L| \operatorname{vol}_D(P) &\overset{(7.20)}{=} |\det L||D(v_1, \ldots, v_n)| \overset{(7.21)}{=} |D(L(v_1), \ldots, L(v_n))| \\
&\overset{(7.20)}{=} \operatorname{vol}_D(P(L(v_1), \ldots, L(v_n))) = \operatorname{vol}_D(L(P))
\end{aligned}$$

(7.24) Fakt (Zusammenhang mit der Determinante von Matrizen). Sei V K-Vektorraum, $\mathcal{G} = (v_1, \ldots, v_n)$ geordnete Basis von V und $L \in \operatorname{End}(V)$. Dann gilt

$$\det L = \det(\operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)).$$

Speziell: Ist $L \in \mathrm{End}(K^n)$ und $A = \mathrm{Mat}(L)$, d. h. ist L durch

$$L(x) = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

gegeben, so gilt $\det L = \det A$.

Bew.: Die Matrix $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = A = (a_{ij}) \in K^{n \times n}$ ist durch

$$L(v_j) = \sum_{i=1}^{n} a_{ij} v_i \qquad \text{für } 1 \leq j \leq n$$

definiert. Es gilt:

$$\begin{aligned} \det L &= \tfrac{D(L(v_1),\ldots,L(v_n))}{D(v_1,\ldots,v_n)} \\ &= \frac{D\left(\sum\limits_{i_1=1}^{n} a_{i_1 1} v_{i_1}, \ldots, \sum\limits_{i_n=1}^{n} a_{i_n n} v_{i_n}\right)}{D(v_1,\ldots,v_n)} \overset{\text{Bew. von (7.10)}}{=} \det A. \end{aligned}$$

Bem.: Sind $\mathcal{G}$, $\mathcal{G}'$ verschiedene geordnete Basen von V und ist $L \in \mathrm{End}(V)$, so wird im allgemeinen

$$\det L \neq \det \left(\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(L) \right)$$

gelten, vgl. (4.20)(a) und (4.21).

(7.25) Folgerung (weitere Eigenschaften von det).

(a) Für alle $A, B \in K^{n \times n}$ gilt: $\det(AB) = \det A \cdot \det B$.

(b) Für alle $A \in \mathrm{GL}_n(K)$ gilt: $\det(A^{-1}) = (\det A)^{-1}$.

Bew.: Das folgt aus (4.20), (7.22) und (7.24).

Bem.: Aus (7.25) folgt, daß $\det : (\mathrm{GL}_n(K), \cdot) \to (K \setminus \{0\}, \cdot)$ ein Gruppenhomomorphismus ist. Sein Kern ist die wichtige Untergruppe

$$\mathrm{SL}_n(K) = \{A \in \mathrm{GL}_n(K) | \det A = 1\}$$

von $\mathrm{GL}_n(K)$, genannt die <u>spezielle lineare Gruppe</u> (über K).

Wozu ist die Determinante nütze?

Als wichtige und nützliche Eigenschaft der Determinante haben wir kennengelernt, daß durch ihr Nichtverschwinden (d.h. $\det \neq 0$) die Automorphismen unter den Endomorphismen und die invertierbaren unter den quadratischen Matrizen charakterisiert werden. Im Fall reeller Vektorräume ist die Determinante bei der Definition von Volumina von Parallelotopen und ihr Betrag als "Volumenverzerrung" von Endomorphismen wichtig. Das Vorzeichen der Determinante (im Fall $K = \mathbb{R}$) ist eng verknüpft mit dem Problem der

Orientierung von endlichdimensionalen $\mathbb{R}$-Vektorräumen.

Ist V ein n-dimensionaler $\mathbb{R}$-Vektorraum, so nennt man zwei geordnete Basen $\mathcal{G} = (v_1, \ldots, v_n)$ und $\mathcal{G}' = (v'_1, \ldots, v'_n)$ <u>gleich orientiert</u>, falls der Endomorphismus $L \in \mathrm{End}(V)$, der durch $L(v_i) = v'_i$ (für $1 \leq i \leq n$) definiert ist, positive Determinante hat. Aus (7.22) folgt leicht, daß "gleich orientiert" eine Äquivalenzrelation auf der Menge der geordneten Basen ist, die genau zwei Äquivalenzklassen hat. Die Auswahl einer dieser beiden Äquivalenzklassen von gleich orientierten Basen nennt man eine <u>Orientierung von</u> V, und die geordneten Basen in dieser Äquivalenzklasse nennt man <u>positiv orientiert</u> (bezüglich der gewählten Orientierung). Die <u>kanonische Orientierung des</u> $\mathbb{R}^n$ ist durch die Auswahl der Äquivalenzklasse definiert, die die Basis $(e_1, \ldots, e_n)$ enthält, und eine geordnete Basis des $\mathbb{R}^n$ heißt positiv orientiert, falls sie gleich orientiert ist wie $(e_1, \ldots, e_n)$, d.h. eine Basis $(v_1, \ldots, v_n)$ des $\mathbb{R}^n$ ist positiv orientiert, wenn die Matrix mit dem Spaltenvektoren $(v_1, \ldots, v_n)$ positive Determinante hat.

Bsp.: Ist $(v_1, \ldots, v_n)$ Basis von V und $\sigma \in S_n$, so sind $(v_1, \ldots, v_n)$ und $(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$ genau dann gleich orientiert, wenn $\mathrm{sgn}(\sigma) = 1$ gilt.

Bew.: Ist $L \in \mathrm{End}(V)$ durch $L(v_i) := v_{\sigma(i)}$ definiert und ist D Determinantenform auf V, so gilt

$$\det L = \frac{D(v_{\sigma(i)}, \ldots, v_{\sigma(n)})}{D(v_1, \ldots, v_n)} \overset{(7.9)}{=} \mathrm{sgn}(\sigma).$$

Nach der Leibnizformel (7.14) ist $\det A$ eine Summe von Produkten, die aus n Komponenten der Matrix $A = (a_{ij}) \in K^{n \times n}$ gebildet werden, d.h. $\det A$

hängt polynomial von den Komponenten von A ab. Insbesondere im Fall $K = \mathbb{R}$ kann man dies benutzen, um zu zeigen:

"Die meisten" $A \in \mathbb{R}^{n\times n}$ sind regulär.

(Zur Erinnerung: A regulär $\Leftrightarrow \det A \neq 0$, vgl. (7.15) und (5.3)).

Dabei muß natürlich erst mathematisch definiert werden, was "die meisten" bedeuten soll. Das ist sogar auf verschiedene Weisen möglich und wird hier nicht durchgeführt. In jedem Fall ist der entscheidende Punkt, daß die Nullstellenmenge $\{A \in \mathbb{R}^{n\times n} | \det A = 0\}$ der polynomialen Funktion det eine "kleine" Teilmenge von $\mathbb{R}^{n\times n}$ ist. Die vorangehende Aussage ist zu folgenden Aussagen äquivalent:

"Die meisten" linearen Gleichungssysteme mit n Gleichungen für n Unbekannte (über dem Körper $\mathbb{R}$) sind eindeutig lösbar.

oder:

Im $\mathbb{R}^n$ schneiden sich ein k-dimensionaler und ein $(n-k)$-dimensionaler affiner Unterraum "typischerweise"in einem Punkt.

Der Zusammenhang zwischen den vorangehenden Aussagen besteht darin, daß ein k-dimensionaler affiner Unterraum durch $n-k$ lineare Gleichungen definiert werden kann, ein $(n-k)$-dimensionaler durch k lineare Gleichungen. Die Schnittmenge zweier solcher Unterräume ist also gerade die Lösungsmenge eines linearen $n \times n$-Gleichungssystems.

Auf der Determinante beruht auch folgende Lösungsformel für lineare $n\times n$-Gleichungssysteme:

(7.26) Cramersche Regel (Gabriel Cramer 1704-1752, schweiz. Mathematiker).

Sei $A = (a_{ij}) \in \mathrm{GL}_n(K)$, $b \in K^n$. Die Lösung $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$ des

Gleichungssystems

$$\text{(I)} \qquad Ax = b$$

ist durch

$$x_k = \frac{1}{\det A} \begin{vmatrix} a_{11} & \dots & a_{1k-1} & b_1 & a_{1k+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{nk-1} & b_n & a_{nk+1} & \dots & a_{nn} \end{vmatrix} \qquad (\text{ für } k = 1, \dots, n)$$

gegeben.

Bew.: Zunächst sei bemerkt, daß aus $A \in \mathrm{GL}_n(K)$ folgt, daß die Abbildung $x \in K^n \to Ax \in K^n$ bijektiv ist, vgl. (7.22) und (7.24). Es existiert also für jedes $b \in K^n$ genau ein $x \in K^n$, das (I) löst. Bezeichnen $A_1, \dots, A_n$ die Spaltenvektoren von A, so ist $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$ genau dann Lösung von (I), wenn $x_1 A_1 + \dots + x_n A_n = b$ gilt. Daraus folgt für die Lösung x von (I):

$$\begin{aligned} D_0(A_1, \dots, A_{k-1}, b, A_{k+1}, \dots, A_n) &= D_0(A_1, \dots, A_{k-1}, \sum_{i=1}^n x_i A_i, A_{k+1}, \dots, A_n) \\ &= \sum_{i=1}^n x_i \underbrace{D_0(A_1, \dots, A_{k-1}, A_i, A_{k+1}, \dots, A_n)}_{=\delta_{ik} D_0(A_1, \dots, A_n)} = x_k \det A \end{aligned}$$

Daraus folgt $x_k = (\det A)^{-1} D_0(A_1, \dots, A_{k-1}, b, A_{k+1}, \dots, A_n)$ und das ist genau die Behauptung.

Weiter spielt die Determinante eine wichtige Rolle bei der Berechnung von Eigenwerten von Endomorphismen. Diese werden uns in der Linearen Algebra II oft wiederbegegnen.

Eigenwerte und Eigenvektoren.

(7.27) Def.: Sei V K-Vektorraum, $L \in \mathrm{End}(V)$. $\lambda \in K$ heißt Eigenwert (EW) von L, falls ein $v \in V \setminus \{0\}$ existiert mit $L(v) = \lambda v$. $v \in V$ heißt Eigenvektor

(EV) <u>von</u> L, falls $v \neq 0$ gilt und ein $\lambda \in K$ existiert mit $L(v) = \lambda v$. In diesem Fall heißt v ein Eigenvektor zum Eigenwert λ.

Bem.: v EV zum EW λ, $\alpha \in K \setminus \{0\} \Rightarrow \alpha v$ EV zum EW λ.

Bsp.:

1) Ist $L = \lambda\, \mathrm{id}_V$, so ist λ der einzige EW von L und jedes $v \in V \setminus \{0\}$ ist EV zum EW λ.

2) Ist $L : K^n \to K^n$ definiert durch $L(e_i) = \lambda_i e_i$ für alle $i \in \{1, \ldots, n\}$, d.h.

$$\mathrm{Mat}(L) = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix},$$

so sind $\lambda_1, \ldots, \lambda_n$ Eigenwerte von L und $e_1, \ldots, e_n$ Eigenvektoren von L.

3) $L : \mathbb{R}^2 \to \mathbb{R}^2$, $L(x_1, x_2) = (x_1 + x_2, x_2)$ "Scherung", mit

$$\mathrm{Mat}(L) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Dann ist $\lambda = 1$ der einzige EW von L und $x \in \mathbb{R}^2$ ist genau dann EV von L, wenn $x = (x_1, 0)$ für ein $x_1 \neq 0$ gilt. Eine Begründung dafür wird nach Satz (7.28) gegeben werden.

4) $L : \mathbb{R}^2 \to \mathbb{R}^2$, $L(x_1, x_2) = (ax_1 - bx_2, bx_1 + ax_2)$ mit

$$\mathrm{Mat}(L) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Ist $b \neq 0$, so hat L keinen (reellen) EW. (Begründung nach Satz (7.28)). Geometrisch ist L eine "Drehstreckung". Identifiziert man $\mathbb{R}^2$ mit $\mathbb{C}$ durch $(x_1, x_2) \leftrightarrow x_1 + ix_2$, vgl. Kap. 2, so kann man L schreiben als: $L(x_1 + ix_2) = (a + ib)(x_1 + ix_2)$. Die Abbildung, die $z \in \mathbb{C}$ die Zahl $(a + ib)z \in \mathbb{C}$ zuordnet, ist also eine Drehstreckung.

5) "Gekoppelte Schwingungen": Gegeben $A \in \operatorname{End}(\mathbb{R}^n)$, gesucht Lösungen $x : \mathbb{R} \to \mathbb{R}^n$ der "Differentialgleichung":

$$(*) \qquad x''(t) = A(x(t)) \text{ für alle } t \in \mathbb{R}.$$

Ist $v \in \mathbb{R}^n$ EV von A zum EW $\lambda \in \mathbb{R}$ und ist $f : \mathbb{R} \to \mathbb{R}$ Lösung von $f''(t)$-$\lambda f(t) = 0$ (diese sind explizit bekannt!), so ist $x(t) := f(t)v$ Lösung von $(*)$:

$$x''(t) = f''(t)v = \lambda f(t)v = f(t)A(v) = A(f(t)v) = A(x(t)).$$

(7.28) Satz. *Sei* $1 \leq \dim V < \infty$, $L \in \operatorname{End}(V)$. *Dann gilt:*

$$\lambda \text{ EW } \textit{von } L \Leftrightarrow \det(L - \lambda \operatorname{id}_V) = 0.$$

Bew.:

$$\begin{aligned} \det(L - \lambda \operatorname{id}_V) = 0 &\overset{(7.22)(b)}{\Leftrightarrow} (L - \lambda \operatorname{id}_V) \notin \operatorname{Aut}(V) \overset{(4.11)}{\Leftrightarrow} \\ \operatorname{kern}(L - \lambda \operatorname{id}_V) \neq \{0\} &\Leftrightarrow \exists v \in V \setminus \{0\} : (L - \lambda \operatorname{id}_V)(v) = 0 \\ &\Leftrightarrow \exists v \in V \setminus \{0\} : L(v) = \lambda v. \end{aligned}$$

Bem.: Ist $\operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = (a_{ij}) \in K^{n \times n}$, so gilt $\operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}}(L - \lambda \operatorname{id}_V) = (a_{ij} - \lambda \delta_{ij})_{1 \leq i,j \leq n}$. Also nach (7.24) und (7.14):

$$\begin{aligned} \det(L - \lambda \operatorname{id}_V) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma)(a_{1\sigma(1)} - \lambda \delta_{1\sigma(1)}) \cdot \ldots \cdot (a_{n\sigma(n)} - \lambda \delta_{n\sigma(n)}) \\ &= (-1)^n \lambda^n + \left(\sum_{i=1}^{n} a_{ii} \right) (-1)^{n-1} \lambda^{n-1} + ?\lambda^{n-2} + \ldots + ?\lambda + \det L, \end{aligned}$$

mit von den a_{ij} abhängenden Koeffizienten ?, die uns im Moment nicht weiter interessieren. Einen Ausdruck dieser Art nennt man "ein Polynom in (der Variablen) λ".

Bez.: $P_L(\lambda) := \det(L - \lambda \operatorname{id}_V)$ heißt das charakteristische Polynom von L.

Satz (7.28) besagt, daß die Eigenwerte eines Endomorphismus L genau die Nullstellen seines charakteristischen Polynoms P_L sind.

Wenn man weiß, daß $\lambda \in K$ Eigenwert von L ist, so erhält man die zugehörigen Eigenvektoren als Lösungen $v \neq 0$ des homogenen linearen $(n \times n)$-Gleichungssystems $(L - \lambda \operatorname{id}_V)(v) = 0$.

Bsp. 3) $L : \mathbb{R}^2 \to \mathbb{R}^2$, $L(x_1, x_2) = (x_1 + x_2, x_2)$ mit

$$\operatorname{Mat}(L) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Dann ist $\mathrm{Mat}(L - \lambda \mathrm{id}_{\mathbb{R}^2}) = \begin{pmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{pmatrix}$ und damit $P_L(\lambda) = (1-\lambda)^2$. Also ist $\lambda = 1$ einziger Eigenwert von L.

Aus $(L - \mathrm{id}_V)(x_1, x_2) = (x_2, 0)$ folgt, daß die Vektoren $x = (x_1, 0)$ mit $x_1 \neq 0$ genau die Eigenvektoren von L zum Eigenwert 1 sind.

Bsp. 4) Wir betrachten das $L \in \mathrm{End}(\mathbb{R}^2)$ mit

$$\mathrm{Mat}(L) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Dann gilt $P_L(\lambda) = \begin{vmatrix} a-\lambda & -b \\ b & a-\lambda \end{vmatrix} = (a-\lambda)^2 + b^2$.

Falls $b \neq 0$ ist, so hat P_L keine (reelle) Nullstelle, d.h. L besitzt keinen (reellen) Eigenwert.

Die Antwort auf die Frage, ob jeder Endomorphismus eines endlichdimensionalen K-Vektorraums einen Eigenwert besitzt, hängt vom Körper K ab. Das vorangehende Beispiel 4) zeigt, daß nicht jeder Endomorphismus des $\mathbb{R}^2$ einen Eigenwert besitzt, während der Fundamentalsatz der Algebra (2.10) impliziert, daß jeder Endomorphismus eines $\mathbb{C}$-Vektorraums V mit $1 \leq \dim V < \infty$ einen Eigenwert (und damit auch Eigenvektoren) besitzt.

Zum Abschluß beweisen wir

(7.29) Satz. *Sei V endlichdimensionaler K-Vektorraum und $L \in \mathrm{End}(V)$. Besitzt L n verschiedene Eigenwerte $\lambda_1, \ldots, \lambda_n \in K$, so ist L diagonalisierbar, d.h. es existiert eine geordnete Basis $\mathcal{G}$ von V, so daß*

$$\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L) = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & 0 & \\ & 0 & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

gilt.

Bem.: 1) Zum Begriff "diagonalisierbar" siehe (5.7) und die auf (5.7) folgenden Bemerkungen.

2) Daß L n verschiedene Eigenwerte besitzt, ist nach (7.28) dazu äquivalent, daß P_L n verschiedene Nullstellen besitzt.

Bew.: Es seien $v_1, \ldots, v_n$ Eigenvektoren zu den Eigenwerten $\lambda_1, \ldots, \lambda_n$. Wir zeigen zunächst, daß $v_1, \ldots, v_n$ linear unabhängig sind. Da $v_1 \neq 0$ ist, ist die Menge

$$\{k | k \in \{1, \ldots, n\} \text{ und } v_1, \ldots, v_k \text{ sind linear unabhängig}\}$$

nicht leer und besitzt deshalb ein maximales Element k_0.

Gilt $k_0 = n$, so ist unsere Behauptung bewiesen. Ist $k_0 < n$, so sind wegen der Maximalität von k_0 die Vektoren $v_1, \ldots, v_{k_0+1}$ linear abhängig. Es existieren also $\alpha_1, \ldots, \alpha_{k_0+1} \in K$, die nicht alle gleich null sind, so daß

$$\sum_{i=1}^{k_0+1} \alpha_i v_i = 0$$

gilt. Wegen $v_{k_0+1} \neq 0$ existiert sogar ein $i_0 \in \{1, \ldots, k_0\}$ mit $\alpha_{i_0} \neq 0$. Auf die vorangehende Gleichung wenden wir L an und erhalten

$$0 = L\left(\sum_{i=1}^{k_0+1} \alpha_i v_i\right) = \sum_{i=1}^{k_0+1} \alpha_i L(v_i) = \sum_{i=1}^{k_0+1} \alpha_i \lambda_i v_i.$$

Von dieser Gleichung subtrahieren wir die Gleichung

$$0 = \lambda_{k_0+1}\left(\sum_{i=1}^{k_0+1} \alpha_i v_i\right) = \sum_{i=1}^{k_0+1} \alpha_i \lambda_{k_0+1} v_i.$$

Das ergibt:

$$0 = \sum_{i=1}^{k_0} \alpha_i (\lambda_i - \lambda_{k_0+1}) v_i.$$

Wegen $\lambda_{i_0} - \lambda_{k_0+1} \neq 0$ und $\alpha_{i_0} \neq 0$ widerspricht das der linearen Unabhängigkeit der Vektoren $v_1, \ldots, v_{k_0}$. Dieser Widerspruch zeigt, daß der Fall $k_0 < n$ nicht eintreten kann. Damit ist die lineare Unabhängigkeit der $v_1, \ldots, v_n$ gezeigt und wir können die geordnete Basis $\mathcal{G} = (v_1, \ldots, v_n)$ von V betrachten. Wegen $L(v_i) = \lambda_i v_i$ für $i \in \{1, \ldots, n\}$ erhalten wir

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & 0 & \\ & 0 & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

wie behauptet.

8 Euklidische Vektorräume

In diesem Kapitel behandeln wir eine zusätzliche Struktur auf reellen Vektorräumen, nämlich ein Skalarprodukt. Mit dessen Hilfe ist es möglich, die Begriffe der euklidischen Geometrie wie etwa Länge von Vektoren, Abstand von Vektoren, Winkel zwischen Vektoren, Orthogonalprojektion, Drehung etc. zu definieren und mit ihnen zu arbeiten. Als wichtiges Beispiel eines Skalarprodukts auf einem ∞-dimensionalen $\mathbb{R}$-Vektorraum werden wir das "L^2-Skalarprodukt" auf dem $\mathbb{R}$-Vektorraum der stetigen Funktion $f : [0, 2\pi] \to \mathbb{R}$ betrachten. Als Anwendung der Überlegungen zu den euklidischen Vektorräumen werden wir hier die algebraischen Aspekte der Theorie der Fourierreihen kennenlernen.

Zunächst sei an Definition (7.1) erinnert. Eine Bilinearform (=2-Linearform) auf einen K-Vektorraum V ist eine Abbildung

$$b : V \times V \to K,$$

die in jedem der beiden Argumente linear ist, vgl. (7.1).

(8.1) Def.: Eine Bilinearform $b : V \times V \to K$ auf einem K-Vektorraum V heißt <u>symmetrisch</u>, falls für alle $v, w \in V$ gilt:

$$b(v, w) = b(w, v)$$

In der nächsten Definition ist die Beschränkung auf den Fall $K = \mathbb{R}$ wesentlich, da wir die Bedingung "$b(v, v) > 0$" fordern.

(8.2) Def.: Sei V ein $\mathbb{R}$-Vektorraum. Ein <u>Skalarprodukt</u> b <u>auf</u> V ist eine symmetrische Bilinearform $b : V \times V \to \mathbb{R}$, so daß für alle $v \in V \setminus \{0\}$ gilt:

$$b(v, v) > 0.$$

Bez.: Skalarprodukte werden wir in der Regel mit dem Symbol

$$\langle v, w \rangle \quad (\text{statt } b(v, w))$$

schreiben. Ein <u>euklidischer Vektorraum</u> ist ein Paar $(V, \langle , \rangle)$ bestehend aus einem $\mathbb{R}$-Vektorraum V und einem Skalarprodukt $\langle , \rangle$ auf V.

Bem.: Für jede Bilinearform b gilt $b(0, w) = b(w, 0) = 0$ für alle $w \in V$, denn $b(0, w) = b(0 \cdot 0, w) = 0b(0, w) = 0$, vgl. (2.5)(a) und (3.2)(a). (Hier und in Zukunft wird – wie allgemein üblich – der Nullvektor eines Vektorraums einfach mit 0 (statt mit $\underline{0}$) bezeichnet. Es wird aus dem Zusammenhang klar, ob $0 \in K$ oder $0 \in V$ gemeint ist!).

Bsp. 1) Sind $a_1, \dots, a_n$ reelle Zahlen, so definiert

$$b(x, y) := \sum_{i=1}^{n} a_i x_i y_i$$

eine symmetrische Bilinearform auf $\mathbb{R}^n$. Dieses b ist genau dann ein Skalarprodukt, wenn $a_i > 0$ für alle $i \in \{1, \dots, n\}$ gilt. Das <u>Standardskalarprodukt auf $\mathbb{R}^n$</u> ist durch

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i$$

gegeben.

2) Sei V der $\mathbb{R}$-Vektorraum

$$C^0([0, 2\pi], \mathbb{R}) = \{f | f : [0, 2\pi] \to \mathbb{R} \text{ stetig}\}.$$

Auf diesem V ist

$$\langle f, g \rangle_{L^2} := \int_0^{2\pi} f(t) g(t) dt$$

ein Skalarprodukt, genannt das L^2-Skalarprodukt.

(8.3) Def.: Sei $(V, \langle , \rangle)$ euklidischer Vektorraum. Dann heißt $\| v \| := \sqrt{\langle v, v \rangle} \geq 0$ die <u>Norm</u> (oder <u>Länge</u>) von $v \in V$ (bzgl. $\langle , \rangle$) und $\| v - w \|$ der <u>Abstand</u> zwischen v und w.

(8.4) Satz. *Sei $(V, \langle , \rangle)$ euklidischer Vektorraum. Dann gilt für alle $v, w \in V, \lambda \in \mathbb{R}$:*

(a) $\| v \| \geq 0$ *und:* $\| v \| = 0 \Leftrightarrow v = 0$

(b) $\| \lambda v \| = |\lambda| \, \| v \|$

(c) $|\langle v, w\rangle| \leq \| v \| \| w \|$ *(Cauchy-Schwarzsche Ungleichung) mit "="* $\Leftrightarrow$ *v und w sind linear abhängig.*

(d) $\| v + w \| \leq \| v \| + \| w \|$ *(Dreiecksungleichung)*

Bew.:

(a) ist klar

(b) $\| \lambda v \| = \sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{\lambda^2 \langle v, v\rangle} = |\lambda| \, \| v \|$

(c) Es genügt, den Fall $w \neq 0$ zu betrachten. Dann gilt für alle $t \in \mathbb{R}$:

$$\begin{aligned} 0 \leq \langle v + tw, v + tw\rangle &= \| v \|^2 + 2t\langle v, w\rangle + t^2 \| w \|^2 \\ &= \left(t \| w \| + \tfrac{\langle v,w\rangle}{\| w \|}\right)^2 + \| v \|^2 - \tfrac{\langle v,w\rangle^2}{\| w \|^2} \end{aligned}$$

Für $t_0 = -\frac{\langle v,w\rangle}{\| w \|^2}$ ist der erste Summand gleich 0. Daraus folgt (c). Gleichheit in (c) tritt genau dann ein, wenn $\langle v + t_0 w, v + t_0 w\rangle = 0$ gilt. Daraus folgt, daß $v + t_0 w = 0$ gilt, d.h. v und w sind linear abhängig. Umgekehrt gilt für linear abhängige v und w offensichtlich Gleichheit in (c).

(d)

$$\begin{aligned} \| v + w \|^2 &= \langle v + w, v + w\rangle = \| v \|^2 + 2\langle v, w\rangle + \| w \|^2 \\ &\overset{(c)}{\leq} \| v \|^2 + 2 \| v \| \| w \| + \| w \|^2 = (\| v \| + \| w \|)^2. \end{aligned}$$

(8.5) Def. Sei $(V, \langle , \rangle)$ euklidischer Vektorraum und $v, w \in V \setminus \{0\}$. Dann ist der Winkel $\varphi = \varphi(v, w) \in [0, \pi]$ zwischen v und w definiert durch:

$$\cos \varphi = \frac{\langle v, w\rangle}{\| v \| \| w \|} \quad \text{bzw.} \quad \varphi = \arccos \frac{\langle v, w\rangle}{\| v \| \| w \|} \in [0, \pi]$$

Bem. 1) (8.4)(c) $\Rightarrow -1 \leq \frac{\langle v,w\rangle}{\| v \| \| w \|} \leq 1$. $\cos : [0, \pi] \to [-1, 1]$ ist bijektiv mit Umkehrabbildung arccos.

2) $\varphi(v, w) = \varphi(w, v)$

3) v und w heißen orthogonal (senkrecht) zueinander $\Leftrightarrow \langle v, w\rangle = 0$. Sind $v, w \in V \setminus \{0\}$, so sind v und w genau dann orthogonal, wenn $\varphi(v, w) = \frac{\pi}{2}$, vgl. Def. (8.5) und $\cos\frac{\pi}{2} = 0$. Sind v und w orthogonal zueinander, so gilt der Satz von Pythagoras:

$$\| v + w \|^2 = \langle v + w, v + w\rangle = \| v \|^2 + 2\langle v, w\rangle + \| w \|^2 = \| v \|^2 + \| w \|^2$$

Die Definition des Winkels ist so eingerichtet, daß der "Kosinussatz" gilt:

(8.6) Folgerung (Kosinussatz). Für alle $v, w \in V \setminus \{0\}$ gilt:

$$\| v - w \|^2 = \| v \|^2 + \| w \|^2 - 2 \| v \| \| w \| \cos\varphi(v, w)$$

Bew.:

$$\| v - w \|^2 = \| v \|^2 - 2\langle v, w\rangle + \| w \|^2 = \| v \|^2 + \| w \|^2 - 2\cos\varphi \| v \| \| w \| .$$

(8.7) Def.: Eine Teilmenge $S \subseteq V$ heißt Orthonormalsystem (ONS), falls gilt

(i) Für alle $v \in S$ gilt $\| v \| = 1$ ("*alle $v \in S$ normiert*") und

(ii) $v, w \in S$, $v \neq w \Rightarrow \langle v, w\rangle = 0$ ("*je zwei orthogonal*")

Eine Orthonormalbasis (ONB) ist ein Orthonormalsystem, das V erzeugt.

Bem.: S ONS $\Rightarrow$ S linear unabhängig: $v_1, \ldots, v_n \in S$ verschieden, $r_1, \ldots, r_n \in \mathbb{R}$ und $\sum_{i=1}^{n} r_i v_i = 0 \Rightarrow 0 = \langle \sum_{i=1}^{n} r_i v_i, \sum_{j=1}^{n} r_j v_j\rangle = \sum_{i,j=1}^{n} r_i r_j \langle v_i, v_j\rangle = \sum_{i=1}^{n} r_i^2 \Rightarrow$ alle $r_i = 0$.

Bsp.:

1) $V = \mathbb{R}^n$, $\langle x, y\rangle := \sum_{i=1}^{n} x_i y_i \Rightarrow \{e_1, \ldots, e_n\}$ ist ONB von $(\mathbb{R}^n, \langle,\rangle)$.

2) $V = C^0([0, 2\pi], \mathbb{R})$, $\langle,\rangle = \langle,\rangle_{L^2}$.
Sei $f_0 \in V$ definiert durch $f_0(x) := \frac{1}{\sqrt{2\pi}}$ und für $k \geq 1$:

$$f_k(x) := \frac{1}{\sqrt{\pi}} \cos(kx), g_k(x) := \frac{1}{\sqrt{\pi}} \sin(kx).$$

Dann ist $S = \{f_k \mid k \in \mathbb{N}\} \cup \{g_k \mid k \in \mathbb{N} \setminus \{0\}\}$ ein ONS (vgl. Anwesenheitsaufgabe).

(8.8) Satz:

(a) (Besselsche Ungleichung). Ist S ONS*, $v \in V$ und $E \subseteq S$ endlich, so gilt*

$$\sum_{e \in E} \langle v, e \rangle^2 \leq \| v \|^2 .$$

(b) Ist $B = (v_1, \ldots, v_n)$ ONB *von V, so gilt*

$$v = \sum_{i=1}^{n} \langle v, v_i \rangle v_i \ \textit{und} \ \| v \|^2 = \sum_{i=1}^{n} \langle v, v_i \rangle^2.$$

Bew.: (a) $0 \leq \| v - \sum\limits_{e \in E} \langle v, e \rangle e \|^2 = \| v \|^2 - 2 \sum\limits_{e \in E} \langle v, e \rangle^2 + \sum\limits_{e \in E} \langle v, e \rangle^2$

(b) $(v_1, \ldots, v_n)$ Basis $\Rightarrow \exists r_1, \ldots, r_n \in \mathbb{R} : v = \sum\limits_{j=1}^{n} r_j v_j \Rightarrow \langle v, v_i \rangle = \sum\limits_{j=1}^{n} r_j \langle v_j, v_i \rangle = r_i$

Bem.: Ein ONS $S \subseteq V$ heißt vollständiges ONS, wenn für alle $v \in V$ gilt

$$\sum_{e \in S} \langle v, e \rangle^2 = \| v \|^2 .$$

Dabei wird $\sum\limits_{e \in S} \langle v, e \rangle^2$ definiert durch

$$\sum_{e \in S} \langle v, e \rangle^2 = \sup \left\{ \sum_{e \in E} \langle v, e \rangle^2 \mid E \text{ endliche Teilmenge von } S \right\}.$$

Es ist eine wichtige Tatsache, daß das ONS

$$S = \{ f_k \mid k \in \mathbb{N} \} \cup \{ g_k \mid k \in \mathbb{N}_{>0} \}$$

ein vollständiges ONS in $C^0([0, 2\pi], \mathbb{R})$ ist. Wir werden den Beweis dafür, der in die Analysis gehört, hier nicht durchführen.

Bsp. 1) Im $\mathbb{R}^n$ mit dem Standardskalarprodukt $\langle , \rangle$ gilt für alle $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$:

$$x = \sum_{i=1}^{n} \langle x, e_i \rangle e_i = \sum_{i=1}^{n} x_i e_i \text{ und } \|x\|^2 = \sum_{i=1}^{n} \langle x, e_i \rangle^2 = \sum_{i=1}^{n} x_i^2.$$

2) Zu $f \in C^0([0,2\pi],\mathbb{R})$ heißen

$$\begin{array}{ll} a_k := a_k(f) := \frac{1}{\pi}\int_0^{2\pi} f(t)\cos(kt)dt & \text{für } k \in \mathbb{N} \text{ und} \\ b_k := b_k(f) := \frac{1}{\pi}\int_0^{2\pi} f(t)\sin(kt)dt & \text{für } k \in \mathbb{N}_{>0} \end{array}$$

die <u>Fourierkoeffizienten</u> von f.

Es gilt dann für das ONS $S = \{f_k | k \in \mathbb{N}\} \cup \{g_k | k \in \mathbb{N}_{>0}\}$:

$$a_k = a_k(f) = \begin{cases} \sqrt{\frac{2}{\pi}}\langle f, f_k\rangle_{L^2} & \text{für } k = 0 \\ \sqrt{\frac{1}{\pi}}\langle f, f_k\rangle_{L^2} & \text{für } k \in \mathbb{N}_{>0} \end{cases}$$

und

$$b_k = b_k(f) = \frac{1}{\sqrt{\pi}}\langle f, g_k\rangle \quad \text{für } k \in \mathbb{N}_{>0}.$$

Die Besselsche Ungleichung (8.8)(a) besagt nun

$$(*) \qquad \frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2) \le \frac{1}{\pi}\int_0^{2\pi} f^2(t)dt$$

und die (hier nicht bewiesene) Vollständigkeit von S zeigt, daß in der Ungleichung $(*)$ in Wirklichkeit stets die Gleichheit gilt.

(8.9) Satz (Gram-Schmidtsches Orthonormalisierungsverfahren). *Sei* $(V, \langle,\rangle)$ *euklidischer Vektorraum.*

a) *Sind* $v_1, \ldots, v_k$ *linear unabhängige Vektoren in* V*, so existiert ein ONS* $w_1, \ldots, w_k$ *in* V*, so daß für alle* $i \in \{1, \ldots, k\}$ *gilt:*

$$(*) \qquad \mathrm{span}\{w_1, \ldots, w_i\} = \mathrm{span}\{v_1, \ldots, v_i\}$$

b) *Ist* $\dim V < \infty$*, so existiert eine Orthonormalbasis (ONB) von* V*.*

Bew.: (a) Durch Induktion nach k.
Induktionsanfang: Da ein einzelner Vektor v_1 genau dann linear unabhängig ist, wenn $v_1 \neq 0$ gilt, können wir $w_1 := \frac{1}{\|v_1\|}v_1$ definieren.
Induktionsschritt: Die Induktionsvoraussetzung angewendet auf $v_1, \ldots, v_{k-1}$ liefert ein ONS $w_1, \ldots, w_{k-1}$, so daß $(*)$ für $1 \le i \le k-1$ gilt. Wir definieren

$$\tilde{w}_k := v_k - \sum_{i=1}^{k-1}\langle v_k, w_i\rangle w_i.$$

Wir benutzen $(*)$ für $i = k-1$, um einzusehen, daß $\tilde{w}_k \in \operatorname{span}\{v_1, \dots, v_k\}$, $v_k \in \operatorname{span}\{w_1, \dots, w_{k-1}, \tilde{w}_k\}$ und $\tilde{w}_k \neq 0$ gilt. Speziell folgt

$$\operatorname{span}\{v_1, \dots, v_k\} = \operatorname{span}\{w_1, \dots, w_{k-1}, \tilde{w}_k\}.$$

Für $j \in \{1, \dots, k-1\}$ gilt

$$\begin{aligned} \langle \tilde{w}_k, w_j \rangle &= \langle v_k - \sum_{i=1}^{k-1} \langle v_k, w_i \rangle w_i, w_j \rangle \\ &= \langle v_k, w_j \rangle - \sum_{i=1}^{k-1} \langle v_k, w_i \rangle \langle w_i, w_j \rangle = 0. \end{aligned}$$

Wir setzen $w_k := \frac{1}{\|\tilde{w}_k\|} \tilde{w}_k$ und erhalten ein ONS $w_1, \dots, w_k$, das $(*)$ für $1 \leq i \leq k$ erfüllt.

b) Ist $v_1, \dots, v_n$ Basis von V, so ist das nach (a) konstruierte ONS $w_1, \dots, w_n$ eine ONB von V.

(8.10) Def.: Sei M Teilmenge eines euklidischen Vektorraums $(V, \langle , \rangle)$. Dann heißt

$$M^\perp = \{v \in V \mid \forall w \in M : \langle v, w \rangle = 0\}$$

der <u>Orthogonalraum von</u> M.

(8.11) Fakten. (a) $M^\perp$ ist Untervektorraum von V.

(b) $M_1 \subseteq M_2 \subseteq V \Rightarrow M_2^\perp \subseteq M_1^\perp$

(c) $M^\perp = (\operatorname{span}(M))^\perp$

(d) $M \subseteq (M^\perp)^\perp$

(e) $M \cap M^\perp \subseteq \{0\}$

Bew.: (a) Zu $w \in M$ definiere $l_w \in V^*$ durch $l_w(v) = \langle v, w \rangle$. Dann gilt $M^\perp = \bigcap\limits_{w \in M} \ker(l_w)$. Da $\ker(l_w)$ ein Untervektorraum ist und jeder Durchschnitt von Untervektorräumen ein Untervektorraum ist, siehe (3.6), folgt (a).
(b) klar.
(c) Wegen $M \subseteq \operatorname{span}(M)$ und (b) folgt $\operatorname{span}(M)^\perp \subseteq M^\perp$. Ist $v \in M^\perp$ und

$w \in \mathrm{span}(M)$, so existieren $k \in \mathbb{N}$ und $w_1, \dots, w_k \in M$, $r_1, \dots, r_k \in \mathbb{R}$, so daß $w = \sum\limits_{i=1}^{k} r_i w_i$ gilt. Dann folgt

$$\langle v, w \rangle = \sum_{i=1}^{k} r_i \langle v, w_i \rangle = 0.$$

Das beweist $v \in \mathrm{span}(M)^\perp$.
(d) Ist $w \in M$, so gilt für alle $v \in M^\perp$: $\langle w, v \rangle = \langle v, w \rangle = 0$. Also: $w \in (M^\perp)^\perp$.
(e) Ist $v \in M \cap M^\perp$, so gilt speziell $\langle v, v \rangle = 0$. Also nach (8.2): $v = 0$.

Bsp.: In $\mathbb{R}^n$ mit dem Standardskalarprodukt gilt für jedes $k \in \{1, \dots, n-1\}$:

$$\{e_1, \dots, e_k\}^\perp = (\mathbb{R}^k \times \{0\})^\perp = \{0\} \times \mathbb{R}^{n-k}.$$

(8.12) Satz. *Sei U Untervektorraum des euklidischen Vektorraums $(V, \langle , \rangle)$. Ist $\dim U < \infty$, so gilt*

$$V = U \oplus U^\perp.$$

Bem.: 1) Aus $V = U \oplus U^\perp$ folgt $(U^\perp)^\perp = U$.
2) Blatt 2, Aufgabe 3, zeigt, daß (8.12) ohne die Voraussetzung "$\dim U < \infty$" nicht allgemein gilt.

Bew.: Nach (8.11)(e) gilt $U \cap U^\perp = \{0\}$, d.h. wir haben eine direkte Summe $U \oplus U^\perp \subseteq V$. Es bleibt zu zeigen, daß $U \oplus U^\perp = V$ gilt. Nach (8.9)(b) existiert eine ONB $v_1, \dots, v_k$ von U. Offensichtlich gilt für jedes $v \in V$

$$(*) \qquad v = \sum_{i=1}^{k} \langle v, v_i \rangle v_i + (v - \sum_{i=1}^{k} \langle v, v_i \rangle v_i)$$

mit $\sum\limits_{i=1}^{k} \langle v, v_i \rangle v_i \in U$. Ist $j \in \{1, \dots, k\}$, so gilt

$$\langle v - \sum_{i=1}^{k} \langle v, v_i \rangle v_i, v_j \rangle = \langle v, v_j \rangle - \sum_{i=1}^{k} \langle v, v_i \rangle \langle v_i, v_j \rangle = 0,$$

d.h. $v - \sum\limits_{i=1}^{k} \langle v, v_i \rangle v_i \in \{v_1, \dots, v_k\}^\perp = U^\perp$. Also zeigt $(*)$, daß $v \in U \oplus U^\perp$ gilt, d.h. $U \oplus U^\perp = V$.

(8.13) Def.: Sei U Untervektorraum eines euklidischen Vektorraums $(V, \langle , \rangle)$, so daß $V = U \oplus U^\perp$ gilt, d.h. für alle $v \in V$ existiert genau ein Paar $(u_0, u_1) \in U \times U^\perp$, so daß $v = u_0 + u_1$ gilt. Dann heißt die Abbildung

$$P_U : V \to U, P_U(v) = u_0$$

die <u>Orthogonalprojektion auf</u> U.

(8.14) Fakten. Sei $U \subseteq V$ Untervektorraum mit $U \oplus U^\perp = V$. Dann gilt:

(a) $P_U \in \mathrm{Hom}(V, U)$

(b) $P_U|U = \mathrm{id}_U, P_U|U^\perp = 0$

(c) Ist $\dim U = k < \infty$ und $\{v_1, \ldots, v_k\}$ ONB von U, so gilt für alle $v \in V$:

$$P_U(v) = \sum_{i=1}^{k} \langle v, v_i \rangle v_i$$

(d) Ist $\dim U^\perp = j < \infty$ und $\{w_1, \ldots, w_j\}$ ONB von $U^\perp$, so gilt

$$P_U(v) = v - \sum_{i=1}^{j} \langle v, w_i \rangle w_i$$

Bew.: Die Aussagen (a) und (b) sind leicht einzusehen, Aussage (c) ist im Beweis von (8.12) enthalten und Aussage (d) läßt sich analog beweisen.

Bsp.: Im $\mathbb{R}^n$ mit dem Standardskalarprodukt ist die Orthogonalprojektion auf $U = \mathbb{R}^k \times \{0\} \subseteq \mathbb{R}^n$ durch

$$P_{\mathbb{R}^k \times \{0\}}(x_1, \ldots, x_n) = (x_1, \ldots, x_k, 0, \ldots, 0)$$

gegeben.

Eine wichtige Eigenschaft der Orthogonalprojektion P_U auf U ist, daß $P_U(v)$ der Punkt auf U ist, der v am nächsten liegt. Das kommt in folgendem Satz zum Ausdruck.

(8.15) Satz. *Sei U Untervektorraum eines euklidischen Vektorraums $(V, \langle , \rangle)$, es gelte $V = U \oplus U^\perp$ und es sei $v \in V$. Dann gilt für alle $u \in U$*

$$\| v - u \| \geq \| v - P_U(v) \|,$$

mit Gleichheit nur für $u = P_U(v)$.

Bew.:

$$\begin{aligned}\| v-u \|^2 &= \| \underbrace{(v-P_U(v))}_{\in U^\perp} + \underbrace{(P_U(v)-u)}_{\in U} \|^2 \\ &= \| v-P_U(v) \|^2 + \underbrace{\| (P_U(v)-u) \|^2}_{\geq 0} \\ &\geq \| v-P_U(v) \|^2\end{aligned}$$

mit Gleichheit nur für $u = P_U(v)$.

Bez.: Sind M_0, M_1 nichtleere Teilmengen von V, so nennen wir

$$d(M_0, M_1) := \inf\{\| v_0 - v_1 \| | v_0 \in M_0, v_1 \in M_1\}$$

den Abstand von M_0 und M_1. Offenbar gilt $d(M_0, M_1) = 0$, falls $M_0 \cap M_1 \neq \emptyset$. Mit dieser Bezeichnung läßt sich die Aussage in Satz (8.15) wie folgt formulieren:

$d(v, U) = \| v - P_U(v) \|$ und $P_U(v)$ ist das einzige Element von U, für das diese Gleichung gilt.

Wir haben folgendes Rechenverfahren zur Bestimmung von $P_U(v)$ und $d(v, U)$, falls $\dim U < \infty$ oder $\dim U^\perp < \infty$ (und $U \oplus U^\perp = V$):

Je nachdem, was leichter ist, bestimmt man entweder eine Basis von U oder eine von $U^\perp$. Diese macht man mit dem Gram-Schmidtschen Verfahren (8.9) zu einer Orthonormalbasis, und erhält entweder eine ONB$\{v_1, \ldots, v_k\}$ von U oder eine ONB$\{w_1, \ldots, w_j\}$ von $U^\perp$.

Im 1. Fall berechnen wir

$$P_U(v) = \sum_{i=1}^{k} \langle v, v_i \rangle v_i \text{ und } d(v, U) = \| v - \sum_{i=1}^{k} \langle v, v_i \rangle v_i \| .$$

Im 2. Fall berechnen wir

$$P_U(v) = v - \sum_{i=1}^{j} \langle v, w_i \rangle w_i \text{ und } d(v, U) = \left(\sum_{i=1}^{j} \langle v, w_i \rangle^2 \right)^{\frac{1}{2}} .$$

Diese Aussagen und Rechnungen sollten wenigstens im Fall des $\mathbb{R}^2$ und $\mathbb{R}^3$ mit den Standardskalarprodukten sehr anschaulich und leicht verständlich sein. Wir übertragen diese Erkenntnisse auf das Problem der "Approximation" von stetigen Funktionen durch Summen von trigonometrischen Funktionen, d.h. auf Fourierreihen. Dadurch wird dieses Problem aus der Analysis unserer geometrischen Anschauung, die durch den uns umgebenden euklidischen Raum geprägt ist, zugänglich.

Wir betrachten wieder $V = C^0([0, 2\pi], \mathbb{R})$ mit dem L^2-Skalarprodukt $\langle , \rangle_{L^2}$ und dem (vollständigen) ONS bestehend aus den Funktionen

$$f_0(t) = \frac{1}{\sqrt{2\pi}}, f_k(t) = \frac{1}{\sqrt{\pi}} \cos(kt) \text{ und } g_k(t) = \frac{1}{\sqrt{\pi}} \sin(kt) \text{ für } k \in \mathbb{N}_{>0}.$$

Die Fourierkoeffizienten $a_k = a_k(f)$ und $b_k = b_k(f)$ sind bis auf Normierungsfaktoren die Skalarprodukte $\langle f, f_k \rangle_{L^2}$ und $\langle f, g_k \rangle_{L^2}$, vgl. Bsp. 2 vor (8.9). Wir betrachten

$$U_n := \text{span}\{f_0, \ldots, f_n, g_1, \ldots, g_n\} \subseteq V$$

und berechnen für $f \in V$

$$\begin{aligned} P_{U_n}(f) &= \langle f, f_0 \rangle f_0 + \sum_{k=1}^{n} \langle f, f_k \rangle f_k + \langle f, g_k \rangle g_k \\ &= \tfrac{a_0}{2} + \sum_{k=1}^{n} a_k \cos(kt) + b_k \sin(kt) \end{aligned}$$

Satz (8.15) besagt nun, daß $P_{U_n}(f)$ das Element von U_n ist, das f bezüglich der L^2-Norm am besten approximiert. Verwenden wir die Vollständigkeit unseres ONS, d.h. die Gleichheit in der Besselschen Ungleichung (8.8)(a), so erhalten wir

$$\| f - P_{U_n}(f) \|_{L^2}^2 = \pi \sum_{k=n+1}^{\infty} (a_k^2 + b_k^2).$$

Da die Reihe $\sum\limits_{k=1}^{\infty} (a_k^2 + b_k^2)$ aufgrund der Besselschen Ungleichung konvergiert, folgt

$$\lim_{n \to \infty} \| f - P_{U_n}(f) \|_{L^2} = 0,$$

d.h. die endlichen Abschnitte $P_{U_n}(f)$ der "Fourierreihe" konvergieren bezüglich der L^2-Norm gegen f.

Wir verallgemeinern nun (8.15) auf die Berechnung des Abstands zweier beliebiger affiner Unterräume eines endlich-dimensionalen euklidischen Vektorraums $(V, \langle , \rangle)$.

(8.16) Satz. *Seien* $A_0 = v_0 + U_0$ *und* $A_1 = v_1 + U_1$ *affine Unterräume von* V. *Dann gilt*

$$d(A_0, A_1) = \| P_{(U_0+U_1)^\perp}(v_1 - v_0) \|$$

Ein Punktepaar $(z_0, z_1) \in A_0 \times A_1$ *mit* $\| z_1 - z_0 \| = d(A_0, A_1)$ *erhält man wie folgt. Wir setzen* $w := P_{(U_0+U_1)^\perp}(v_1 - v_0)$. *Dann gilt* $(v_1 - v_0) - w \in U_0 + U_1$, *so daß* $u_0 \in U_0$ *und* $u_1 \in U_1$ *existieren, so daß*

$$v_1 - v_0 = u_0 + u_1 + w$$

gilt. Wir setzen dann

$$z_0 := v_0 + u_0, z_1 := v_1 - u_1.$$

Jedes weitere Punktepaar $(z_0', z_1') \in A_0 \times A_1$ *mit* $\| z_1' - z_0' \| = d(A_0, A_1)$ *ist dann von der Form*

$$z_0' = z_0 + u, z_1' = z_1 + u$$

für ein $u \in U_0 \cap U_1$.

Bem.: 1) Ist $(z_0, z_1) \in A_0 \times A_1$, so gilt für alle $u \in U_0 \cap U_1$:

$$z_0' := z_0 + u \in A_0, z_1' := z_1 + u \in A_1, \text{ und } \| z_1' - z_0' \| = \| z_1 - z_0 \|$$

2) Gilt $U_0 \cap U_1 = \{0\}$, so existiert genau ein Paar $(z_0, z_1) \in A_0 \times A_1$ mit $\| z_1 - z_0 \| = d(A_0, A_1)$.

Bew.: (a) Es sei $z_0 := v_0 + u_0$, $z_1 := v_1 - u_1$. Dann gilt

$z_0 \in A_0$ und $z_1 \in A_1$ und $z_1 - z_0 = v_1 - v_0 - u_0 - u_1 = w = P_{(U_0+U_1)^\perp}(v_1 - v_0)$,

nach Wahl der Vektoren u_0, u_1 und w.

(b) Seien $z_0' \in A_0$, $z_1' \in A_1$ beliebige Punkte auf A_0 und A_1. Dann existieren $u_0' \in U_0$ und $u_1' \in U_1$, so daß

$$z_0' = z_0 + u_0' \text{ und } z_1' = z_1 + u_1'$$

gilt. Damit erhalten wir

$$\begin{aligned} \| z_1' - z_0' \|^2 &= \| \underbrace{(z_1 - z_0)}_{\in (U_0+U_1)^\perp} + \underbrace{(u_1' - u_0')}_{\in U_0+U_1} \|^2 = \| z_1 - z_0 \|^2 + \underbrace{\| u_1' - u_0' \|^2}_{\geq 0} \\ &\geq \| z_1 - z_0 \|^2 . \end{aligned}$$

Außerdem gilt $\| z_1' - z_0' \| = d(A_0, A_1)$ genau dann, wenn $u_1' - u_0' = 0$ gilt, d.h. wenn $u := u_1' = u_0' \in U_0 \cap U_1$ und $z_1' = z_1 + u$, $z_2' = z_2 + u$.

Ein Rechenverfahren zur Bestimmung von $d(A_0, A_1)$ erhalten wir wie folgt. Bestimme eine Basis von $(U_0+U_1)^\perp$ und orthonormalisiere sie. Ist $\{w_1, \ldots, w_j\}$ die so erhaltene ONB von $(U_0 + U_1)^\perp$, so gilt

$$d(A_0, A_1) = \left(\sum_{i=1}^{j} \langle v_1 - v_0, w_i \rangle^2 \right)^{\frac{1}{2}}$$

Als nächstes definieren wir die "strukturverträglichen Abbildungen" zwischen euklidischen Vektorräumen.

(8.17) Def.: Seien $(V, \langle , \rangle_V)$ und $(W, \langle , \rangle_W)$ euklidische Vektorräume. Eine Abbildung $L : V \to W$ heißt orthogonal (bzgl. $\langle , \rangle_V$ und $\langle , \rangle_W$), falls $L \in \mathrm{Hom}(V, W)$ gilt und falls für alle $v_1, v_2 \in V$ gilt:

$$\langle v_1, v_2 \rangle_V = \langle L(v_1), L(v_2) \rangle_W$$

Bem. 1) Orthogonale Abbildungen L erhalten also die Norm von Vektoren (d.h. $\| v \| = \| L(v) \| = 0$) und die Winkel zwischen Vektoren (d.h. $\varphi(v, w) = \varphi(L(v), L(w))$, vgl. Def. (8.5)).
2) Orthogonale Abbildungen L sind stets injektiv, denn $v \in \ker L$ impliziert

$$\| v \| = \| L(v) \| = 0,$$

also $v = 0$. Speziell sind orthogonale Abbildungen eines endlich-dimensionalen euklidischen Vektorraums in sich stets bijektiv, vgl. (4.11).

Bsp.: Sei $U \subseteq V$ Untervektorraum, für den $U \oplus U^\perp = V$ gilt, und P_U die Orthogonalprojektion auf U. Wir zerlegen jedes $v \in V$ als $v = u_0 + u_1$ mit $u_0 \in U$, $u_1 \in U^\perp$ (d.h. $u_0 = P_U(v)$, $u_1 = v - P_U(v)$) und definieren

$$S_U : V \to V,\ S_U(v) = S_U(u_0 + u_1) := u_0 - u_1.$$

Dann ist S_U orthogonale Abbildung. S_U heißt die Spiegelung an U. Wegen $u_0 = P_U(v), u_1 = v - P_U(v)$ gilt

$$S_U(v) = 2P_U(v) - v.$$

Bez.: Ist $(V, \langle , \rangle)$ euklidischer Vektorraum, so sei mit

$$\mathrm{O}(V) = \{L \in \mathrm{Aut}(V) \mid L \text{ orthogonal}\}$$

die orthogonale Gruppe von $(V,\langle,\rangle)$ bezeichnet. (Statt $\mathrm{O}(V)$ sollte man besser $\mathrm{O}(V,\langle,\rangle)$ schreiben, aber das ist nicht üblich, da zu umständlich). $\mathrm{O}(V)$ ist eine Untergruppe von $(\mathrm{Aut}(V),\circ)$. Ist $\dim V<\infty$, so ist ein orthogonales $L\in\mathrm{End}(V)$ automatisch bijektiv.

(8.18) Fakten. Seien $(V,\langle,\rangle)$ und $W,\langle,\rangle_W)$ euklidische Vektorräume, $\dim V = n<\infty$ und $L\in\mathrm{Hom}(V,W)$. Dann gilt:

(a) Ist L orthogonal und $(v_1,\dots,v_n)$ ONB von V, so ist $(L(v_1),\dots,L(v_n))$ ONS in W.

(b) Ist $(v_1,\dots,v_n)$ ONB von V und ist $(L(v_1),\dots,L(v_n))$ ONS in W, so ist L orthogonal.

Bew.: (a) $\langle L(v_i),L(v_j)\rangle_W=\langle v_i,v_j\rangle_V=\delta_{ij}$.
(b) Ist $v=\sum\limits_{i=1}^n\alpha_i v_i, w=\sum\limits_{j=1}^n\beta_j v_j$, so folgt

$$\langle L(v),L(w)\rangle_W=\sum_{i,j=1}^n\alpha_i\beta_j\underbrace{\langle L(v_i),L(v_j)\rangle}_{=\delta_{ij}}=\sum_{i=1}^n\alpha_i\beta_i=\langle v,w\rangle_V.$$

(8.19) Folgerung. Seien $(V,\langle,\rangle_V)$ und $(W,\langle,\rangle_W)$ euklidische Vektorräume mit $\dim V=\dim W<\infty$. Dann existiert ein orthogonaler Isomorphismus $L: V\to W$. Speziell ist jeder n-dimensionale euklidische Vektorraum orthogonal isomorph zum $\mathbb{R}^n$ mit dem Standardskalarprodukt.

Bew.: Sei $(v_1,\dots,v_n)$ ONB von V, $(w_1,\dots,w_n)$ ONB von W. Definiere $L\in\mathrm{Hom}(V,W)$ durch

$$L(v_i)=w_i \text{ für } i\in\{1,\dots,n\}.$$

Dann ist L Isomorphismus (vgl. (4.5)) und (8.18)(b) impliziert, daß L orthogonal ist.

(8.20) Satz. *Sei* $(V,\langle,\rangle)$ *euklidischer Vektorraum,* $\mathcal{G}=(v_1,\dots,v_n)$ ONB *von* V *und* $L\in\mathrm{End}(V)$. *Dann gilt:*

$$L\in\mathrm{O}(V)\Leftrightarrow\ \textit{Für } A:=\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L)\ \textit{gilt}: A^TA=E_n.$$

Speziell gilt $\det L\in\{\pm1\}$.

Bem.: Offenbar ist $A^TA = E_n$ äquivalent zu $A^T = A^{-1}$ und zu $AA^T = E_n$.

Bew.: Sei $L \in \mathrm{O}(V)$ und $(a_{ij}) := A = \mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L)$. Dann gilt für alle $1 \leq i, j \leq n$:

$$\begin{aligned} \delta_{ij} = \langle v_i, v_j \rangle &= \langle L(v_i), L(v_j) \rangle = \langle \sum_{k=1}^{n} a_{ki} v_k, \sum_{l=1}^{n} a_{lj} v_l \rangle \\ &= \sum_{k,l=1}^{n} a_{ki} a_{lj} \underbrace{\langle v_k, v_l \rangle}_{=\delta_{kl}} = \sum_{k=1}^{n} a_{ki} a_{kj} = (A^T A)_{ij}, \end{aligned}$$

also $A^TA = E_n$. Umgekehrt zeigt diese Gleichungskette: Aus $A^TA = E_n$ folgt $\langle L(v_i), L(v_j) \rangle = \delta_{ij}$ für alle $1 \leq i, j \leq n$. Daraus folgt nach (8.18)(b), daß L orthogonal ist. Schließlich folgt aus $A^TA = E_n$, daß $1 = \det E_n = \det(A^TA) \overset{(7.25)}{=} \det A^T \cdot \det A \overset{(7.15)}{=} (\det A)^2$ gilt, also $\det L = \det A \in \{\pm 1\}$.

Wir führen folgende Bezeichnungen ein:

$$\begin{array}{rcll} \mathrm{SO}(V) &=& \{L \in \mathrm{O}(V) \mid \det L = 1\} & \\ \mathrm{O}(n) &=& \{A \in \mathbb{R}^{n \times n} \mid A^TA = E_n\} & \text{“orthogonale Gruppe”} \\ \mathrm{SO}(n) &=& \{A \in \mathrm{O}(n) \mid \det A = 1\} & \text{“spezielle orthogonale Gruppe”} \end{array}$$

Die Matrizen in $\mathrm{O}(n)$ heißen “orthogonale Matrizen”.

Ist $\mathcal{G}$ ONB von V, so ist

$$\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}} \mid \mathrm{O}(V) : \mathrm{O}(V) \to \mathrm{O}(n)$$

ein Isomorphismus der Untergruppe $\mathrm{O}(V)$ von $\mathrm{Aut}(V)$ auf die Untergruppe $\mathrm{O}(n)$ von $\mathrm{GL}(n, \mathbb{R})$, der die Untergruppe $\mathrm{SO}(V)$ von $\mathrm{O}(V)$ auf die Untergruppe $\mathrm{SO}(n)$ von $\mathrm{O}(n)$ abbildet.

Bsp.: Ist $U \subseteq V$ ein k-dimensionaler Untervektorraum, so existiert eine ONB $\mathcal{G} = (v_1, \ldots, v_n)$ von V, so daß $(v_1, \ldots, v_k)$ ONB von U ist. Dann gilt

$$\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(S_U) = \begin{pmatrix} \overbrace{\begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix}}^{k} & 0 \\ 0 & \underbrace{\begin{matrix} -1 & & \\ & \ddots & \\ & & -1 \end{matrix}}_{n-k} \end{pmatrix}$$

und $\det(S_U) = (-1)^{n-k}$. Die Spiegelung S_U an U ist also genau dann Element von $\mathrm{SO}(V)$, wenn $n-k$, genannt die Kodimension von U in V, gerade ist.

Bem.: Eine Matrix $A \in \mathbb{R}^{n\times n}$ liegt genau dann in $\mathrm{O}(n)$, wenn ihre Spaltenvektoren (oder ihre Zeilenvektoren) eine ONB des $\mathbb{R}^n$ mit dem kanonischen Skalarprodukt bilden. A liegt genau dann in $\mathrm{SO}(n)$, wenn die Spaltenvektoren $(A_1, \ldots, A_n)$ eine positiv orientierte ONB des $\mathbb{R}^n$ bilden.

Bsp.: Die Matrix $\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$ ist für alle $\varphi \in \mathbb{R}$ in $\mathrm{SO}(n)$.

9 Die Gramsche Determinante und das Kreuzprodukt

(J.P. Gram, dänischer Mathematiker 1850-1916).

Zunächst werden wir uns überlegen, daß es in euklidischen Vektorräumen einen natürlichen Volumenbegriff für Parallelotope gibt.

(9.1) Satz. *Sei V euklidischer Vektorraum, $0 < \dim V = n < \infty$. Dann existieren genau zwei Determinantenformen $\pm D$ auf V, so daß gilt:*

$$\textit{Ist } v_1, \ldots, v_n \text{ ONB } \textit{ von } V, \textit{so gilt } |\pm D(v_1, \ldots, v_n)| = 1.$$

Bez.: Ein solches D heißt normierte Determinantenform des euklidischen Vektorraums V.

Bew.: Existenz: Sei $(\overline{v}_1, \ldots, \overline{v}_n)$ eine (feste) ONB von V. Nach (7.10) existiert genau eine Determinantenform D von V mit

$$D(\overline{v}_1, \ldots, \overline{v}_n) = 1.$$

Wir müssen zeigen, daß dann für jede ONB $(v_1, \ldots, v_n)$ von V

$$|D(v_1, \ldots, v_n)| = 1$$

gilt. Sei $L \in \operatorname{End} V$ durch $L(\overline{v}_i) = v_i$ für $1 \le i \le n$ definiert. Nach Definition (7.21) gilt:

$$D(v_1, \ldots, v_n) = \det L \cdot D(\overline{v}_1, \ldots, \overline{v}_n) = \det L.$$

Nach (8.20) gilt: $\det L \in \{\pm 1\}$, also

$$|D(v_1, \ldots, v_n| = |\det L| = 1.$$

Daß $\pm D$ die einzigen solchen Determinantenformen sind, folgt direkt aus (7.10).

Aus Satz (9.1) folgt, daß in endlich-dimensionalen euklidischen Vektorräumen $(V, \langle,\rangle)$ ein natürlicher Volumenbegriff für Parallelotope existiert. Ist D eine normierte Determinantenform und sind $w_1, \ldots, w_n \in V$, so definieren wir

$$\mathrm{vol}_n^{\langle,\rangle}(P(w_1, \ldots, w_n)) := |D(w_1, \ldots, w_n)|.$$

Ist $(v_1, \ldots, v_n)$ eine ONB von V, so nennt man $P(v_1, \ldots, v_n)$ einen Würfel der Kantenlänge 1, und es folgt:

$$\mathrm{vol}_n^{\langle,\rangle}(P(v_1, \ldots, v_n)) = 1.$$

Die Gramsche Determinante berechnet $\mathrm{vol}_n^{\langle,\rangle}(P(w_1, \ldots, w_n))$ mittels der Skalarprodukte $\langle w_i, w_k\rangle$ für $1 \leq i, k \leq n$:

(9.2) Satz (Gramsche Determinante). *Sei $(V, \langle,\rangle)$ n-dimensionaler euklidischer Vektorraum und $w_1, \ldots, w_n \in V$. Dann gilt*

$$\mathrm{vol}_n^{\langle,\rangle}(P(w_1, \ldots, w_n)) = \sqrt{\det((\langle w_i, w_k\rangle)_{1\leq i,k\leq n})}.$$

Bsp.: Wir betrachten $\mathbb{R}^2$ mit dem Standardskalarprodukt. Dann ist D_0 eine normierte Determinantenform, da $D_0(e_1, e_2) = 1$ ist. In diesem Fall sagt (9.2):

$$\begin{aligned}\mathrm{vol}_2^{\langle,\rangle}(P(w_1, w_2)) &= \sqrt{\det\begin{pmatrix} \langle w_1, w_1\rangle & \langle w_1, w_2\rangle \\ \langle w_2, w_1\rangle & \langle w_2, w_2\rangle \end{pmatrix}} = \sqrt{\|w_1\|^2\|w_2\|^2 - \langle w_1, w_2\rangle^2} \\ &= \|w_1\|\|w_2\|\sqrt{1 - \cos^2\varphi} = \|w_1\|\|w_2\| \sin\varphi,\end{aligned}$$

wobei φ den Winkel zwischen w_1 und w_2 bezeichnet. Das ist die Formel für die Fläche des Parallelogramms $P(w_1, w_2)$ mit den Seitenlängen $\|w_1\|$ und $\|w_2\|$ und dem eingeschlossenen Winkel φ.

Beweis von (9.2): Sei $\mathcal{G} = (v_1, \ldots, v_n)$ eine ONB von V und sei $L \in \mathrm{End}(V)$ durch $L(v_j) = w_j$ für $1 \leq j \leq n$ definiert. Dann ist $A = (a_{ij})_{1\leq i,j\leq n} = \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)$ durch

$$w_j = \sum_{i=1}^{n} a_{ij} v_i$$

definiert, und es gilt für jede normierte Determinantenform D:

$$(*)\mathrm{vol}_n^{\langle,\rangle}(P(w_1,\dots,w_n)) = |D(w_1,\dots,w_n)| = |\det A||D(v_1,\dots,v_n)| = |\det A|.$$

Andererseits:

$$\begin{aligned}\langle w_i, w_j\rangle &= \left\langle \sum_{k=1}^n a_{ki}v_k, \sum_{l=1}^n a_{lj}v_l \right\rangle = \sum_{k,l=1}^n a_{ki}a_{lj} \underbrace{\langle v_k, v_l\rangle}_{=\delta_{kl}} = \sum_{k=1}^n a_{ki}a_{kj} \\ &= (A^T \cdot A)_{ij}\end{aligned}$$

Also: $(\langle w_i, w_j\rangle)_{1\le i,j\le n} = A^T A$.

Daraus folgt: $(**)$ $\det(\langle w_i, w_j\rangle)_{1\le i,j\le n} = \det(A^T A) = (\det A)^2$.

Aus $(*)$ und $(**)$ folgt die Behauptung.

Aufgrund der Tatsache, daß sich $\mathrm{vol}_n^{\langle,\rangle}(P(w_1,\dots,w_n))$ allein aus den Skalarprodukten $\langle w_i, w_j\rangle$, für $1 \le i,j \le n$, berechnen läßt, kann man vermuten, daß in euklidischen Vektorräumen auch ein natürlicher k-dimensionaler Volumenbegriff für k-dimensionale Parallelotope mit $k < \dim V = n$ existiert. Das von k Vektoren $w_1,\dots,w_k$ aufgespannte Parallelotop $P(w_1,\dots,w_k) \subseteq V$ ist durch

$$P(w_1,\dots,w_k) = \left\{ \sum_{i=1}^k s_i w_i \mid \forall i \in \{1,\dots,k\} : 0 \le s_i \le 1 \right\}$$

definiert. In Analogie zu (9.2) definieren wir

(9.3) Def.: Sind $w_1,\dots,w_k \in V$, so ist das k-dimensionale Volumen von $P(w_1,\dots,w_k)$ (bzgl. $\langle,\rangle$) definiert als

$$\mathrm{vol}_k^{\langle,\rangle}(P(w_1,\dots,w_k)) = \sqrt{\det(\langle w_i, w_j\rangle)_{1\le i,j\le k}}$$

Wegen (9.2) hat diese Definition folgende Eigenschaft: Ist U ein k-dimensionaler Untervektorraum von V, der die Vektoren $w_1,\dots,w_k$ enthält, so induziert $\langle,\rangle$ ein Skalarprodukt

$$\langle,\rangle_U := \langle,\rangle \mid U \times U$$

auf U. Ist D^U normierte Determinantenform auf $(U, \langle,\rangle_U)$, so gilt

$$\mathrm{vol}_k^{\langle,\rangle}(P(w_1, \dots, w_k)) = |D^U(w_1, \dots, w_k)|.$$

Eine wichtige Eigenschaft dieses Volumenbegriffs ist, daß er unter orthogonalen Abbildungen invariant ist. Das bedeutet Folgendes:

Sind $(V, \langle,\rangle_V)$ und $(W, \langle,\rangle_W)$ euklidische Vektorräume, so gilt für jede orthogonale Abbildung $L : V \to W$ und jedes Parallelotop $P = P(v_1, \dots, v_k)$ in V

$$(*) \qquad \mathrm{vol}_k^{\langle,\rangle_W}(L(P)) = \mathrm{vol}_k^{\langle,\rangle_V}(P).$$

Gleichung $(*)$ folgt aus der Beziehung $L(P(v_1, \dots, v_k)) = P(L(v_1), \dots, L(v_k))$:

$$\begin{aligned}\mathrm{vol}_k^{\langle,\rangle_W}(L(P)) &= \sqrt{\det(\langle L(v_i), L(v_j)\rangle_W)_{1\le i,j\le k}} \\ &= \sqrt{\det(\langle v_i, v_j\rangle_V)_{1\le i,j\le k}} = \mathrm{vol}_k^{\langle,\rangle_V}(P).\end{aligned}$$

Definition (9.3) ist die (lineare) Grundlage für die Definition des "k-dimensionalen Flächeninhalts" von "k-dimensionalen Flächen" im euklidischen $\mathbb{R}^n$ und für die Definition von Integralen über solche k-dimensionalen Flächen. Diese Objekte werden in der Analysis behandelt.

Das Kreuzprodukt

In einem n-dimensionalen, orientierten, euklidischen Vektorraum V definieren wir eine Abbildung "Kreuzprodukt", die $n-1$ Vektoren $w_1, \dots, w_{n-1}$ in V einen Vektor in V, bezeichnet als $w_1 \times \dots \times w_{n-1}$, zuordnet und die angenehme algebraische und interessante geometrische Eigenschaften hat.

In einem solchen Vektorrau V gibt es genau eine normierte Determinantenform D, die auf positiv orientierten Basen positive Werte annimmt, d.h.

$$(v_1, \dots, v_n) \text{ positive orientierte ONB von } V \Rightarrow D(v_1, \dots, v_n) = 1.$$

Diese Determinantenform D nennen wir die kanonische Determinantenform von V. Betrachten wir $\mathbb{R}^n$ mit dem Standardskalarprodukt und der üblichen Orientierung, so ist $(e_1, \dots, e_n)$ eine positiv orientierte ONB von $\mathbb{R}^n$. Da für die Standarddeterminantenform D_0 des $\mathbb{R}^n$

$$D_0(e_1, \dots, e_n) = 1$$

gilt, ist D_0 die kanonische Determinantenform dieses "Standard-$\mathbb{R}^n$".

(9.4) Lemma (von Riesz). Sei $(V, \langle , \rangle)$ endlichdimensionaler euklidischer Vektorraum. Dann existiert zu jeder Linearform $\varphi \in V^*$ genau ein Vektor $v \in V$, so daß für alle $w \in V$ gilt:

$$\varphi(w) = \langle v, w \rangle.$$

Bew.: Sei $v_1, \ldots, v_n$ eine ONB von V und $v := \sum\limits_{i=1}^{n} \varphi(v_i) v_i \in V$. Dann gilt für alle $w := \sum\limits_{i=1}^{n} \alpha_i v_i \in V$:

$$\varphi(w) = \sum \alpha_i \varphi(v_i) = \left\langle \sum_{i=1}^{n} \alpha_i v_i, \sum_{j=1}^{n} \varphi(v_j) v_j \right\rangle = \langle w, v \rangle = \langle v, w \rangle.$$

Die Eindeutigkeit von v folgt aus der auch sonst wichtigen Tatsache, daß jeder Vektor $v \in V$ durch seine Skalarprodukte $\langle v, w \rangle$ mit allen Vektoren $w \in V$ (es genügt mit allen Vektoren einer Basis von V) eindeutig bestimmt ist:

Gilt für $v \in V$ und $\tilde{v} \in V$, daß

$$\langle v, w \rangle = \langle \tilde{v}, w \rangle$$

für alle $w \in V$ gilt, so folgt

$$\langle v - \tilde{v}, w \rangle = 0$$

für alle $w \in V$. Für $w := v - \tilde{v}$ erhalten wir

$$0 = \langle v - \tilde{v}, v - \tilde{v} \rangle = \| v - \tilde{v} \|^2,$$

also $v - \tilde{v} = 0$ aufgrund der positiven Definitheit des Skalarprodukts.

(9.5) Def.: Sei $(V, \langle , \rangle)$ orientierter euklidischer Vektorraum mit $3 \leq \dim V = n < \infty$ und kanonischer Determinantenform D. Zu je $n-1$ Vektoren $w_1, \ldots, w_{n-1}$ von V existiert genau ein Vektor $w \in V$, so daß für alle $v \in V$ gilt

$$D(w_1, \ldots, w_{n-1}, v) = \langle w, v \rangle.$$

Dieses $w \in V$ heißt das <u>Kreuzprodukt</u> von $w_1, \ldots, w_{n-1}$ und wird mit

$$w =: w_1 \times \ldots \times w_{n-1}$$

bezeichnet.

Bem.: Es ist ausschließlich das Kreuzprodukt von $n-1$ Vektoren definiert, also etwa $v \times w$ für v und w in $\mathbb{R}^3$, <u>nicht aber</u> $v \times w$ für v und w in $\mathbb{R}^4$!

Begründung für die Existenz von w in (9.5): Die Abbildung

$$v \in V \to D(w_1, \ldots, w_{n-1}, v) \in \mathbb{R}$$

ist eine Linearform. Somit folgt die Existenz von w aus Lemma (9.4).

(9.6) Eigenschaften des Kreuzprodukts.

(a) Die Abbildung $(w_1, \ldots, w_{n-1}) \in V^{n-1} \to w_1 \times \ldots \times w_{n-1} \in V$ ist multilinear und alternierend. Speziell gilt: $w_1, \ldots, w_{n-1}$ linear abhängig $\Rightarrow w_1 \times \ldots \times w_{n-1} = 0$.

(b) Ist $L \in \mathrm{O}(V)$, so gilt

$$L(w_1 \times \ldots \times w_{n-1}) = \det L \cdot (L(w_1) \times \ldots \times L(w_{n-1})).$$

(c) "Geometrische Definition des Kreuzprodukts":
$w_1 \times \ldots \times w_{n-1}$ ist der eindeutig bestimmte Vektor in V, für den gilt:

(i) $w_1 \times \ldots \times w_{n-1} \in (\mathrm{span}\{w_1, \ldots, w_{n-1}\})^\perp$.

(ii) $\| w_1 \times \ldots \times w_{n-1} \| = \mathrm{vol}_{n-1}^{\langle,\rangle}(P(w_1, \ldots, w_{n-1}))$.

(iii) Sind $w_1, \ldots, w_{n-1}$ linear unabhängig, so ist $(w_1, \ldots, w_{n-1}, w_1 \times \ldots \times w_{n-1})$ positiv orientierte Basis von V.

Bew.: a) Exemplarisch wird die Alternierungseigenschaft gezeigt:
Es sei $1 \leq i < j \leq n-1$ und $w_i = w_j$. Dann gilt für alle $v \in V$

$$\langle w_1 \times \ldots \times w_{n-1}, v \rangle = D(w_1, \ldots, w_{n-1}, v) = 0,$$

da D alternierend ist. Daraus folgt $w_1 \times \ldots \times w_{n-1} = 0$, vgl. (9.4).

b) Da L bijektiv ist, können wir für alle $v \in V$ schreiben

$$\begin{aligned}
&\langle L(w_1 \times \ldots \times w_{n-1}), v \rangle = \langle L(w_1 \times \ldots \times w_{n-1}), L(L^{-1}(v)) \rangle \\
\overset{L \in \mathrm{O}(V)}{=}\;& \langle w_1 \times \ldots \times w_{n-1}, L^{-1}(v) \rangle = D(w_1, \ldots, w_{n-1}, L^{-1}(v)) \\
\overset{\text{Def. von } \det(L)}{=}\;& \det(L)^{-1} D(L(w_1), \ldots, L(w_{n-1}), v) \\
\overset{\det(L) \in \{\pm 1\}}{=}\;& \langle \det(L) \cdot (L(w_1) \times \ldots \times L(w_{n-1})), v \rangle
\end{aligned}$$

Daraus folgt die Behauptung.

(c) Exemplarisch wird (ii) bewiesen. Sind $w_1, \ldots, w_{n-1}$ linear abhängig, so sind beide Seiten null, vgl. (a) und die Bemerkungen nach (9.3). Sind $w_1, \ldots, w_{n-1}$ linear unabhängig, so existiert ein $v \in V$, so daß $(w_1, \ldots, w_{n-1}, v)$ eine Basis von V ist, also

$$0 \neq D(w_1, \ldots, w_{n-1}, v) = \langle w_1 \times \ldots \times w_{n-1}, v \rangle$$

und damit folgt

$$w_1 \times \ldots \times w_{n-1} \neq 0.$$

Wir setzen

$$w := \frac{w_1 \times \ldots \times w_{n-1}}{\| w_1 \times \ldots \times w_{n-1} \|}.$$

Es sei $U := \operatorname{span}\{w_1, \ldots, w_{n-1}\}$ und $D_U : U^{n-1} \to \mathbb{R}$ die Abbildung

$$D_U(u_1, \ldots, u_{n-1}) := D(u_1, \ldots, u_{n-1}, w).$$

Ist $v_1, \ldots, v_{n-1}$ ONB von U, so folgt aus (a), daß $v_1, \ldots, v_{n-1}, w$ eine ONB von V ist, also

$$|D_U(v_1, \ldots, v_{n-1})| = |D(v_1, \ldots, v_{n-1}, w)| = 1.$$

D_U ist also eine normierte Determinantenform auf U und die Bemerkungen nach (9.3) zeigen, daß

$$\operatorname{vol}_{n-1}^{\langle,\rangle}(P(w_1, \ldots, w_{n-1})) = |D_U(w_1, \ldots, w_{n-1})| = |D(w_1, \ldots, w_{n-1}, w)|$$

gilt. Andererseits gilt nach Def. (9.5)

$$\begin{aligned} D(w_1, \ldots, w_{n-1}, w) &= \langle w_1 \times \ldots \times w_{n-1}, w \rangle \\ &= \left\langle w_1 \times \ldots \times w_{n-1}, \tfrac{w_1 \times \ldots \times w_{n-1}}{\| w_1 \times \ldots \times w_{n-1} \|} \right\rangle = \| w_1 \times \ldots \times w_{n-1} \| . \end{aligned}$$

Zusammen mit der vorangehenden Gleichung beweist das Aussage (ii).

Schließlich leiten wir eine explizite Formel her, die es erlaubt, $w_1 \times \ldots \times w_{n-1}$ aus den Komponenten von $w_1, \ldots, w_{n-1}$ bezüglich einer positiv orientierten ONB zu berechnen.

(9.7) Fakt. Ist $(v_1, \ldots, v_n)$ positiv orientierte ONB von V und $w_j =: \sum\limits_{i=1}^{n} a_{ij} v_i$ für $1 \leq j \leq n-1$, so gilt

$$w_1 \times \ldots \times w_{n-1} = \sum_{i=1}^{n} \big((-1)^{n+i} \det A_i)\big)\, v_i,$$

wobei $A^i \in \mathbb{R}^{(n-1)\times(n-1)}$ aus $A := (a_{ij})_{\substack{1\leq i\leq n\\ 1\leq j\leq n-1}} \in \mathbb{R}^{n\times(n-1)}$ durch Streichen der i'ten Zeile entsteht.

Bem.: Es gilt $a_{ij} = \langle v_i, w_j\rangle$ und die Spalten von A bestehen gerade aus den Komponenten der w_j bezüglich der gegebenen, positiv orientierten ONB $(v_1, \ldots, v_n)$.

Bsp.: Im Fall von $V = \mathbb{R}^3$ mit der üblichen Orientierung und dem Standardskalarprodukt betrachten wir die positiv orientierte ONB (e_1, e_2, e_3) und

$$\begin{aligned} w_1 &=: \ a = (a_1, a_2, a_3) = \sum_{i=1}^{3} a_i\, e_i \in \mathbb{R}^3 \text{ und} \\ w_2 &=: \ b = (b_1, b_2, b_3) = \sum_{i=1}^{3} b_i\, e_i \in \mathbb{R}^3. \end{aligned}$$

Dann ist $A \in \mathbb{R}^{3\times 2}$ die Matrix

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix}$$

und es gilt:

$$a \times b = \sum_{i=1}^{3} \big((-1)^{i+3} \det A_i\big)\, e_i = (a_2b_3 - a_3b_2, -a_1b_3 + a_3b_1, a_1b_2 - a_2b_1).$$

Vorbemerkung. Sind $z_1, \ldots, z_n$ beliebige Vektoren in V und $z_j =: \sum\limits_{i=1}^{n} z_{ij} v_i$, d.h. $z_{ij} = \langle z_j, v_i\rangle$, so gilt

$$D(z_1, \ldots, z_n) = \det((z_{ij})_{1\leq i,j\leq n}).$$

Es ist nämlich leicht zu sehen, daß die rechte Seite als Funktion von $z_1, \ldots, z_n$ eine alternierende n-Form ist, die für $z_1 := v_1, \ldots, z_n := v_n$ den Wert 1 annimmt, gerade wie D.

Bew. von (9.7): Für alle $x = \sum\limits_{i=1}^{n} x_i v_i \in V$ gilt aufgrund der Vorbemerkung:

$$\begin{aligned} D(w_1, \ldots, w_{n-1}, x) &= \begin{vmatrix} a_{11} & \ldots & a_{1n-1} & x_1 \\ \vdots & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ a_{n1} & \ldots & a_{nn-1} & x_n \end{vmatrix} \overset{\substack{\text{Entwicklung nach} \\ \text{der letzten Spalte}}}{=} \sum_{i=1}^{n} (-1)^{n+i} \det(A_i) x_i \\ &= \left\langle \sum_{i=1}^{n} \left((-1)^{n+i} \det A_i \right) v_i, x \right\rangle . \end{aligned}$$

Also: $w_1 \times \ldots \times w_{n-1} = \sum\limits_{i=1}^{n} \left((-1)^{n+i} \det A_i \right) v_i$.

Bsp.: Es sei $V = \mathbb{R}^{n+1}$ mit üblicher Orientierung und Standardskalarprodukt. Es sei $e_1, \ldots, e_n$ die Standardbasis des $\mathbb{R}^n$ und $s_1, \ldots, s_n$ reelle Zahlen. Wir betrachten die Vektoren $w_1 := (e_1, s_1), \ldots, w_n := (e_n, s_n)$ im $\mathbb{R}^{n+1}$. Dann gilt:

$$\mathrm{vol}_n^{\langle,\rangle}(P(w_1, \ldots, w_n)) = \sqrt{1 + \sum_{i=1}^{n} (s_i)^2}$$

Bew.: Nach (9.6)(c)(ii) und (9.7) gilt:

$$\mathrm{vol}_n(P(w_1, \ldots, w_n)) \overset{(c)}{=} \| w_1 \times \ldots \times w_n \| \overset{(9.7)}{=} \sqrt{\sum_{i=1}^{n+1} (\det A_i)^2},$$

wobei A_i aus der $((n+1) \times n)$-Matrix

$$A = \begin{pmatrix} 1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & \ldots & \ldots & 0 \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ 0 & & \ldots & \ldots & 1 \\ s_1 & s_2 & \ldots & \ldots & s_n \end{pmatrix}$$

durch Streichen der i'ten Zeile entsteht. Offensichtlich gilt $\det A_{n+1} = 1$ und $|\det A_i| = |s_i|$ für $1 \leq i \leq n$.

10 Adjungierte Abbildung und Diagonalisierung symmetrischer Matrizen

Folgendes Lemma bereitet die Definition der adjungierten Abbildung vor.

(10.1) Lemma. Seien $(V, \langle , \rangle_V)$ und $(W, \langle , \rangle_W)$ euklidische Vektorräume und $L \in \mathrm{Hom}(V, W)$. Dann existiert höchstens eine Abbildung $L^* : W \to V$, so daß für alle $v \in V$ und alle $w \in W$ gilt

$$(*) \qquad \langle L(v), w \rangle_W = \langle v, L^*(w) \rangle_V .$$

Falls L^* existiert, so gilt $L^* \in \mathrm{Hom}(W, V)$. Ist V endlich-dimensional, so existiert L^*.

Bew.: Sind L^* und $\tilde{L}^* : W \to V$ Abbildungen, so daß $(*)$ für alle $v \in V$ und $w \in W$ gilt, so gilt für jedes $w \in W$

$$\langle v, L^*(w) \rangle_V = \langle L(v), w \rangle_W = \langle v, \tilde{L}^*(w) \rangle_V$$

für alle $v \in V$. Mit der Eindeutigkeitsausage in (9.4) folgt $L^*(w) = \tilde{L}^*(w)$, d.h. $L^* = \tilde{L}^*$.

Wir nehmen an, daß L^* existiert, und zeigen, daß $L^* : W \to V$ linear ist. Es gilt für $w_1, w_2 \in W$ und alle $v \in V$

$$\begin{aligned} \langle v, L^*(w_1 + w_2) \rangle_V &= \langle L(v), w_1 + w_2 \rangle_W = \langle L(v), w_1 \rangle_W + \langle L(v), w_2 \rangle_W \\ &= \langle v, L^*(w_1) \rangle_V + \langle v, L^*(w_2) \rangle_V = \langle v, L^*(w_1) + L^*(w_2) \rangle_V . \end{aligned}$$

Daraus folgt $L^*(w_1 + w_2) = L^*(w_1) + L^*(w_2)$. Es sei V endlichdimensional und $w \in W$ gegeben. Wir betrachten die Linearform

$$v \in V \to \langle L(v), w \rangle_W \in \mathbb{R}$$

in V^*. Nach (9.4) existiert (genau) ein $\tilde{v} \in V$, so daß für alle $v \in V$ gilt

$$\langle v, \tilde{v} \rangle_V = \langle L(v), w \rangle .$$

Wir setzen $L^*(w) := \tilde{v}$, und dann ist $(*)$ für alle $v \in V$ erfüllt.

Bez.: Wir nennen L^* (, falls L^* existiert,) die <u>zu L adjungierte Abbildung</u>.

(10.2) Fakt. Seien V und W endlichdimensionale euklidische Vektorräume mit Orthonormalbasen $\mathcal{G}_V = (v_1, \dots, v_n)$, $\mathcal{G}_W = (w_1, \dots, w_m)$ und $L \in \mathrm{Hom}(V, W)$. Dann gilt

$$\mathrm{Mat}_{\mathcal{G}_W}^{\mathcal{G}_V}(L^*) = (\mathrm{Mat}_{\mathcal{G}_V}^{\mathcal{G}_W}(L))^T.$$

Bew.: Sei $\mathrm{Mat}_{\mathcal{G}_V}^{\mathcal{G}_W}(L) = (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}}$ und $\mathrm{Mat}_{\mathcal{G}_W}^{\mathcal{G}_V}(L^*) = (b_{lk})_{\substack{1\le l\le n\\ 1\le k\le m}}$, d.h. $L(v_j) = \sum\limits_{i=1}^{m} a_{ij} w_i$ für $1 \le j \le n$ und $L^*(w_k) = \sum\limits_{l=1}^{n} b_{lk} v_l$ für $1 \le k \le m$. Da $\mathcal{G}_V$ und $\mathcal{G}_W$ Orthonormalbasen sind, folgt

$$\langle L(v_j), w_k\rangle_W = a_{kj} = \langle v_j, L^*(w_k)\rangle_V = b_{jk}.$$

Das ist gerade die Behauptung.

(10.3) Def.: Sei $(V, \langle , \rangle)$ euklidischer Vektorraum und $L \in \mathrm{End}(V)$. L heißt <u>selbstadjungiert</u> (oder <u>symmetrisch</u>), falls $L = L^*$ gilt, d.h. falls für alle $v, w \in V$ gilt:

$$\langle L(v), w\rangle = \langle v, L(w)\rangle.$$

(10.4) Folgerung. Sei $\dim V < \infty$, $L \in \mathrm{End}(V)$ und $\mathcal{G}$ ONB von V. Dann gilt:

$$L \text{ selbstadjungiert } \Leftrightarrow \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)^T.$$

Bew.: Ist L selbstadjungiert, so folgt $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)^T$ aus (10.2). Gilt $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)^T$, so folgt aus (10.2), daß $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L^*) = \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)$ gilt. Daraus folgt $L = L^*$, vgl. (4.15).

Im endlichdimensionalen Fall sind die selbstadjungierten Endomorphismen also gerade die, die bezüglich einer ONB durch symmetrische Matrizen beschrieben werden. Hier ist ein typisches Beispiel, wie selbstadjungierte Endomorphismen unendlichdimensionaler Vektorräume in der Analysis auftreten:

Es sei $V := \{f \mid f : \mathbb{R} \to \mathbb{R}$ beliebig oft differenzierbar und periodisch mit Periode $1\}$ mit dem L^2-Skalarprodukt

$$\langle f, g\rangle_{L^2} = \int_0^1 f(x)g(x)dx.$$

Dann ist der Endomorhismus (!)

$$L : V \to V, L(f) := f''$$

selbstadjungiert, d.h. für alle $f, g \in V$ gilt

$$\int_0^1 f''(x)g(x)dx = \int_0^1 f(x)g''(x)dx.$$

Das läßt sich durch zweimalige partielle Integration zeigen.

Die folgende Definition bezieht sich auf Endomorphismen eines allgemeinen K-Vektorraums. Insbesondere spielen dabei Skalarprodukte keine Rolle.

(10.5) Def.: Sei V Vektorraum über einem Körper K und $L \in \mathrm{End}(V)$. Ein Untervektorraum U von V heißt $\underline{L\text{-invariant}}$, falls $L(U) \subseteq U$ gilt.

Bem.: Ein $v \in V$ ist genau dann Eigenvektor (EV) von L (vgl. Def. (7.27)), wenn $\mathrm{span}\{v\}$ ein 1-dimensionaler L-invarianter Untervektorraum von V ist.

Eine sehr wichtige Eigenschaft von selbstadjungierten Endomorphismen ist die folgende:

(10.6) Fakt. Ist $(V, \langle , \rangle)$ euklidischer Vektorraum, ist $L \in \mathrm{End}(V)$ selbstadjungiert und U L-invarianter Untervektorraum von V, so ist auch $U^\perp$ L-invariant.

Bew.: $v \in U^\perp \Rightarrow \forall u \in U : \langle v, u \rangle = 0 \overset{L(U)\subseteq U}{\Rightarrow} \forall u \in U : \langle v, L(u) \rangle = 0 \overset{L\text{selbstadj.}}{\Rightarrow}$ $\forall u \in U : \langle L(v), u \rangle = 0 \Rightarrow L(v) \in U^\perp$.

Bem.: Eine analoge Aussage gilt für orthogonale Selbstabbildungen eines endlichdimensionalen euklidischen Vektorraums.

Folgendes Objekt ist nützlich beim Beweis der Existenz reeller Eigenwerte von selbstadjungierten Endomorphismen eines endlichdimensionalen euklidischen Vektorraums. Es ist benannt nach John William Strutt, 3rd Baron Rayleigh (1842-1919), einem englischen Physiknobelpreisträger und Rektor der University of Cambridge.

(10.7) Def.: Sei $(V, \langle , \rangle)$ euklidischer Vektorraum und $L \in \mathrm{End}(V)$ selbstadjungiert.

Die Abbildung

$$R_L : V \setminus \{0\} \to \mathbb{R}, R_L(v) := \frac{\langle L(v), v\rangle}{\langle v, v\rangle},$$

heißt der Rayleighquotient zu L.

Bem.: Es gilt offensichtlich für alle $v \in V \setminus \{0\}, \lambda \in \mathbb{R} \setminus \{0\}$: $R_L(\lambda v) = R_L(v)$.

Der folgende Satz zeigt, warum der Rayleighquotient von Bedeutung ist.

(10.8) Satz. *Sei* $(V, \langle , \rangle)$ *n-dimensionaler euklidischer Vektorraum,* $1 \leq n < \infty$*. Dann gelten:*

(a) *Es existieren* $v_-, v_+ \in V \setminus \{0\}$*, so daß für alle* $v \in V \setminus \{0\}$ *gilt:*

$$R_L(v_-) \leq R_L(v) \leq R_L(v_+).$$

(b) *Jeder Eigenwert* λ *von* L *liegt im Intervall* $[R_L(v_-), R_L(v_+)]$.

(c) v_- *und* v_+ *sind Eigenvektoren von* L *zu den Eigenwerten* $R_L(v_-)$ *und* $R_L(v_+)$.

Bem.: Wichtig ist für uns vor allem Teil (c), aus dem die Existenz eines (reellen) Eigenwerts von L folgt.

Satz (10.8)(c) ist allerdings ein reines Existenzresultat, das zur expliziten Berechnung eines Eigenwerts nicht tauglich ist. Diese geschieht in der Regel mittels des charakteristischen Polynoms, siehe Satz (7.28). Zur approximativen Berechnung von Eigenwerten ist (10.8) jedoch nützlich.

Bew.: (a) Es ist leicht zu sehen, daß

$$S := \{v \in V \mid \| v \| = 1\}$$

eine beschränkte und (im Sinne der Topologie) abgeschlossene Teilmenge von V ist. Da V endlich dimensional ist, folgt aus dem Satz von Bolzano-Weierstraß (siehe Analysis), daß S eine kompakte Teilmenge von V ist (nach (8.19) kann man sich beim Beweis auf den Fall beschränken, daß $(V, \langle , \rangle)$ der $\mathbb{R}^n$ mit dem Standardskalarprodukt ist). Nun ist $R_L | S : S \to \mathbb{R}$ eine stetige Funktion auf einer kompakten Teilmenge $S \subseteq V$. In der Analysis wird gezeigt, daß es dann $v_-, v_+ \in S$ gibt, so daß für alle $v \in S$ gilt

$$R_L(v_-) \leq R_L(v) \leq R_L(v_+).$$

Ist $v \in V \setminus \{0\}$, so gilt $\frac{v}{\|v\|} \in S$ und $R_L(v) = R_L\left(\frac{v}{\|v\|}\right)$, so daß die vorangehenden Ungleichungen in der Tat für alle $v \in V \setminus \{0\}$ gelten.

(b) Ist v Eigenvektor zum Eigenwert λ von L, so gilt offenbar $R_L(v) = \lambda$, also $R_L(v_-) \leq \lambda = R_L(v) \leq R_L(v_+)$, nach (a).

(c) Es sei v_- wie in (a) und o.E. $\| v_- \| = 1$. Für einen beliebigen Vektor $v \in V$ betrachten wir die in einer Umgebung von $t = 0$ definierte Funktion

$$f(t) := R_L(v_- + tv).$$

Es gilt $f(t) = \frac{g(t)}{h(t)}$, wobei $g(t)$ und $h(t)$ die (höchstens) quadratischen Polynome

$$\begin{aligned} g(t) &= \langle L(v_-) + tL(v), v_- + tv\rangle = \langle L(v_-), v_-\rangle \\ &\quad +t(\langle L(v_-), v\rangle + \langle L(v), v_-\rangle) + t^2\langle L(v), v\rangle \\ &\overset{L\text{selbstadj.}}{=} \langle L(v_-), v_-\rangle + 2t\langle L(v_-), v\rangle + t^2\langle L(v), v\rangle \\ &\overset{\|v_-\|=1}{=} R_L(v_-) + 2t\langle L(v_-), v\rangle + t^2\langle L(v), v\rangle \end{aligned}$$

und

$$\begin{aligned} h(t) &= \langle v_-, v_-\rangle + 2t\langle v_-, v\rangle + t^2\langle v, v\rangle \\ &= 1 + 2t\langle v_-, v\rangle + t^2\langle v, v\rangle \end{aligned}$$

sind. Insbesondere sind f, g und h differenzierbar an der Stelle $t = 0$. Da f nach Voraussetzung an der Stelle $t = 0$ sein Minimum annimmt, folgt

$$\begin{aligned} 0 = f'(0) = \tfrac{h(0)g'(0) - g(0)h'(0)}{h(0)^2} &= 2\langle L(v_-), v\rangle - R_L(v_-) \cdot 2\langle v_-, v\rangle \\ &= 2\langle L(v_-) - R_L(v_-)v_-, v\rangle. \end{aligned}$$

Da diese Gleichung für alle $v \in V$ gilt, folgt

$$L(v_-) - R_L(v_-)v_- = 0,$$

d.h. v_- ist Eigenvektor von L zum Eigenwert $R_L(v_-)$. Der Beweis für v_+ verläuft analog.

Aus (10.6) und (10.8) ergibt sich leicht das folgende Hauptergebnis dieses Kapitels:

(10.9) Satz. *Sei* $(V, \langle , \rangle)$ *euklidischer Vektorraum,* $1 \leq \dim V = n < \infty$, *und* $L \in \mathrm{End}(V)$ *selbstadjungiert. Dann existiert eine* ONB $\mathcal{G} = (v_1, \ldots, v_n)$ *aus Eigenvektoren von* L. *Es gilt also*

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

wobei λ_i *der Eigenwert von* L *zum Eigenvektor* v_i *ist.*

Der Beweis erfolgt durch vollständige Induktion nach $n = \dim V$. Der Fall $n = 1$ ist offensichtlich. Zum Induktionsschritt bemerken wir, daß nach (10.8) ein Eigenvektor v von L existiert und daß nach (10.6)

$$U := (\mathrm{span}\{v\})^{\perp}$$

ein L-invarianter Untervektorraum von V ist. Wir überlegen nun, daß auf $L|U \in \mathrm{End}(U)$ die Induktionsvoraussetzung anwendbar ist: $L|U$ ist selbstadjungiert bezüglich des Skalarprodukts $\langle , \rangle_U := \langle , \rangle | U \times U$ auf U, und nach (8.12) gilt $\dim U = n - 1$. Also existiert eine ONB$(v_1, \ldots, v_{n-1})$ (bzgl. $\langle , \rangle_U$) von U, die aus Eigenvektoren von $L|U$ ($\Rightarrow$ von L) besteht. Wegen $v \in U^{\perp}$ ist dann $(v_1, \ldots, v_{n-1}, \frac{v}{\|v\|})$ eine ONB, die aus Eigenvektoren von L besteht.

Übersetzen wir die Aussage von (10.9) in die Sprache der Matrizen, so erhalten wir:

(10.10) Folgerung. Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch, d.h. es gelte $A = A^T$. Dann existiert eine orthogonale Matrix $S \in \mathrm{O}(n)$ (d.h. es gilt $S^T = S^{-1}$), so daß $S^T A S$ eine Diagonalmatrix ist.

Bem.: Damit ist für symmetrische reelle Matrizen das "Normalformenproblem" mit optimalem Ergebnis gelöst, vgl. Bem. 2) nach (5.7).

Bew.: Wir definieren $L \in \mathrm{End}(\mathbb{R}^n)$ durch die Forderung $\mathrm{Mat}(L) = A$. Nach (10.4) ist L selbstadjungiert bezüglich des Standardskalarprodukts auf $\mathbb{R}^n$. Nach (10.9) existiert eine ONB $\mathcal{G} = (v_1, \ldots, v_n)$ des $\mathbb{R}^n$ aus Eigenvektoren von L. Aus (8.18) folgt, daß

$$S := \mathrm{Mat}_{\mathcal{G}}^{(e_1, \ldots, e_n)}(\mathrm{id}_{\mathbb{R}^n}) \in \mathrm{O}(n)$$

gilt. Nun folgt aus (5.6), daß $S^{-1}AS$ gerade die Matrix von L bezüglich der Basis $\mathcal{G}$ ist, d.h.

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = S^T AS,$$

und $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)$ ist eine Diagonalmatrix, da die Basis $\mathcal{G}$ aus Eigenvektoren von L besteht.

11 Die Normalform orthogonaler Abbildungen

Wir benützen die Begriffe und Bezeichnungen, die am Ende des 8. Kapitels (ab (8.17)) eingeführt wurden.

Wir untersuchen zunächst die Gruppe $\mathrm{O}(\mathbb{R}^2)$ der orthogonalen Selbstabbildungen des $\mathbb{R}^2$ mit dem Standardskalarprodukt.

Beispiele von Elementen in $\mathrm{O}(\mathbb{R}^2)$:

1) "Drehungen": Sei $\varphi \in \mathbb{R}$ und $D_\varphi \in \mathrm{End}(\mathbb{R}^2)$ mit

$$\mathrm{Mat}(D_\varphi) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

Dann gilt

$$\begin{aligned} \mathrm{Mat}(D_\varphi)^T \cdot \mathrm{Mat}(D_\varphi) &= \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \cdot \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\varphi + \sin^2\varphi & 0 \\ 0 & \cos^2\varphi + \sin^2\varphi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \end{aligned}$$

d.h. $\mathrm{Mat}(D_\varphi) \in O(2)$ und damit $D_\varphi \in \mathrm{O}(\mathbb{R}^2)$. Wegen

$$\begin{vmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{vmatrix} = 1 \text{ gilt sogar } D_\varphi \in \mathrm{SO}(\mathbb{R}^2), \mathrm{Mat}(D_\varphi) \in \mathrm{SO}(2).$$

Offenbar gilt $D_{\varphi+2\pi} = D_\varphi$ und $D_\varphi \neq D_\psi$, falls $0 \leq \varphi < \psi < 2\pi$.

Die Abbildung $D_\varphi \in \mathrm{O}(\mathbb{R}^2)$ heißt die <u>Drehung des</u> $\mathbb{R}^2$ <u>um den Winkel</u> φ.

2) "Spiegelungen an Geraden durch $0 \in \mathbb{R}^2$": Sei $U \subseteq \mathbb{R}^2$ 1-dimensionaler Untervektorraum. Wähle eine ONB $\mathcal{G} = (v_1, v_2)$ des $\mathbb{R}^2$ mit $\operatorname{span}\{v_1\} = U$. Definiere "die Spiegelung $S_U \in \operatorname{End}(\mathbb{R}^2)$ an U" durch

$$S_U(v_1) = v_1, S_U(v_2) = -v_2.$$

Dann gilt $S_U \in \mathrm{O}(\mathbb{R}^2) \setminus \mathrm{SO}(\mathbb{R}^2)$, da $\operatorname{Mat}_{\mathcal{G}}^{\mathcal{G}}(S_U) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \in \mathrm{O}(2) \setminus \mathrm{SO}(2)$.

S_U ist durch U eindeutig bestimmt, vgl. Blatt 6, Anwesenheitsaufgabe 1.

(11.1) Fakt. Es gilt:

(a) $\mathrm{SO}(\mathbb{R}^2) = \{D_\varphi | \varphi \in [0, 2\pi)\}$.

(b) $\mathrm{O}(\mathbb{R}^2) \setminus \mathrm{SO}(\mathbb{R}^2) = \{S_U | U$ 1-dimensionaler Unterraum des $\mathbb{R}^2\}$.

(c) $\forall \varphi, \psi \in \mathbb{R} : D_{\varphi+\psi} = D_\varphi \circ D_\psi (= D_\psi \circ D_\varphi), D_0 = \mathrm{id}_{\mathbb{R}^2}$ (d.h. $\varphi \in (\mathbb{R}, +) \to D_\varphi \in (\mathrm{SO}(\mathbb{R}^2), \circ)$ ist surjektiver Gruppenhomomorphismus).

$$D_\varphi = \mathrm{id}_{\mathbb{R}^2} \Leftrightarrow \varphi \in \{2\pi k | k \in \mathbb{Z}\}$$

(d) $S_{U_2} \circ S_{U_1} = D_{2\varphi}$, wobei φ der Winkel ist, um den man U_1 drehen muß, um U_2 zu erhalten, d.h. $D_\varphi(U_1) = U_2$. (φ ist bis auf additive Vielfache $\varphi + k\pi, k \in \mathbb{Z}$, von π bestimmt.)

(e) $D_{2\varphi} \circ S_U = S_{D_\varphi(U)}, S_U \circ D_{2\varphi} = S_{D_{-\varphi}(U)}$.

Insbesondere folgt aus (c), (d) bzw. (e), daß $\mathrm{SO}(\mathbb{R}^2)$ abelsch ist, während $\mathrm{O}(\mathbb{R}^2)$ nicht abelsch ist.

Bew.:

(a) Sei $L \in \mathrm{SO}(\mathbb{R}^2)$, $L(e_1) =: (x, y) \in \mathbb{R}^2$. Dann gilt $x^2 + y^2 = 1$ und mit den Kenntnissen aus der Analysis I kann man einsehen, daß es genau ein $\varphi \in [0, 2\pi)$ gibt mit $(x, y) = (\cos\varphi, \sin\varphi)$, d.h. $L(e_1) = D_\varphi(e_1)$. Dann sind $L(e_2)$ und $D_\varphi(e_2)$ beides Einheitsvektoren, die zu $L(e_1) = D_\varphi(e_1)$ orthogonal sind und zusammen mit $L(e_1) = D_\varphi(e_1)$ eine positiv orientierte Basis bilden. Da es nur einen solchen Vektor gibt, gilt auch $L(e_2) = D_\varphi(e_2)$, also $L = D_\varphi$ (vgl. Blatt 6, Anwesenheitsaufgabe 1).

(b) Sei $L \in \mathrm{O}(2) \setminus \mathrm{SO}(2)$. Um zu zeigen, daß L eine Spiegelung ist, suchen wir einen 1-dimensionalen Unterraum U, der punktweise von L festgelassen wird, d.h. einen Vektor $v \in \mathbb{R}^2 \setminus \{0\}$ mit $L(v) = v$ ($\Leftrightarrow v$ EV zum EW 1 von L). Ist $\mathrm{Mat}(L) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, so gilt wegen $\det L = -1$: $\det(L - \lambda \mathrm{id}_{\mathbb{R}^2}) = \lambda^2 - (a+d)\lambda - 1 =: p(\lambda)$. Wegen $p(0) = -1$ und $\lim\limits_{\lambda \to \pm\infty} p(\lambda) = \infty$, hat $p(\lambda)$ zwei Nullstellen (d.h. L zwei Eigenwerte, vgl. Satz (7.28)) $\lambda_2 < 0 < \lambda_1$. Da L orthogonal ist, hat jeder Eigenwert von L den Betrag 1, also $\lambda_2 = -1, \lambda_1 = 1$. Sei v_1 EV von L zum EW $\lambda_1 = 1$ und (v_1, v_2) ONB von $\mathbb{R}^2$. Aus $L \in \mathrm{O}(\mathbb{R}^2)$, $L \neq \mathrm{id}_{\mathbb{R}^2}$ und $L(v_1) = v_1$, folgt $L(v_2) = -v_2$, d.h. $L = S_U$ für $U := \mathrm{span}\{v_1\}$.

(c) folgt aus den "Additionstheoremen" für sin und cos.

(d) Es gelte $D_\varphi(U_1) = U_2$. Da $\det(S_{U_2} \circ S_{U_1}) = \det(S_{U_2}) \det(S_{U_1}) = (-1)^2 = 1$ ist, ist $S_{U_2} \circ S_{U_1} \in \mathrm{SO}(\mathbb{R}^2)$, d.h. $S_{U_2} \circ S_{U_1}$ ist eine Drehung.

Wie in der Vorlesung mit einem einfachen elementargeometrischen Argument (und ähnlich für (e)) gezeigt wurde, ist $S_{U_2} \circ S_{U_1}$ in der Tat eine Drehung um den Winkel 2φ. Dieses Argument muß aber in der hier aufgebauten "Analytischen Geometrie" durch eine Rechnung bewiesen werden. Es ist nun so, daß solche Rechnungen statt mit (2×2)-Matrizen sehr viel weniger aufwendig mit komplexen Zahlen ausgeführt werden können. Deshalb zunächst der

Exkurs: Beschreibung von Drehungen und Spiegelungen des $\mathbb{R}^2$ mit Hilfe der komplexen Zahlen.

Identifizieren wir wie üblich $(x, y) \in \mathbb{R}^2$ mit $x + iy \in \mathbb{C}$ und definieren (!) wir für $\varphi \in \mathbb{R}$

$$e^{i\varphi} := \cos\varphi + i \sin\varphi$$

(für diese Definition gibt es einen mathematischen Hintergrund, der uns jetzt nicht zu interessieren braucht), so berechnet man:

$$D_\varphi(z) = e^{i\varphi} \cdot z.$$

Die Additionstheoreme für sin und cos sind äquivalent zur Gleichung $e^{i(\varphi+\psi)} = e^{i\varphi} \cdot e^{i\psi}$ für alle $\varphi, \psi \in \mathbb{R}$. Außerdem gilt $\overline{e^{i\varphi}} = e^{-i\varphi}$ und $|e^{i\varphi}| = \sqrt{\cos^2 + \sin^2 \varphi} = 1$. Ist $U = \mathrm{span}_{\mathbb{R}}\{e^{i\psi}\} := \{s e^{i\psi} \mid s \in \mathbb{R}\}$, so zeigen wir, daß S_U durch

$$S_U(z) = e^{2i\psi}\overline{z}$$

gegeben ist: Die Abbildung $z \in \mathbb{C} \simeq \mathbb{R}^2 \to \overline{z} \in \mathbb{C} \simeq \mathbb{R}^2$ ist gerade die Spiegelung an der x-Achse, so daß die Abbildung

$$z \to e^{2i\psi}\overline{z} = D_{2\psi}(\overline{z})$$

in $\mathrm{O}(\mathbb{R}^2) \setminus \mathrm{SO}(\mathbb{R}^2)$ liegt. Wendet man sie auf $z = e^{i\psi}$ an, so erhält man $e^{i\psi} \to e^{2i\psi} \cdot e^{-i\psi} = e^{i\psi}$. Da auch $S_U(e^{i\psi}) = e^{i\psi}$ gilt, folgt

$$S_U(z) = e^{2i\psi}\overline{z}$$

für alle $z \in \mathbb{C}$, denn beide Abbildungen sind Spiegelungen, die die Gerade $\mathrm{span}_{\mathbb{R}}\{e^{2i\psi}\}$ punktweise fest lassen.

Wir kommen nun zu einem "analytischen" Beweis von (11.1)(d) und (e):

(d): Es sei $U_1 = \mathrm{span}_{\mathbb{R}}\{e^{i\psi_1}\} = D_{\psi_1}(\mathbb{R} \times \{0\})$ und $U_2 = \mathrm{span}_{\mathbb{R}}\{e^{i\psi_2}\} = D_{\psi_2}(\mathbb{R} \times \{0\})$.

Dann gilt $D_{\psi_2-\psi_1}(U_1) = D_{\psi_2-\psi_1}(D_{\psi_1}(\mathbb{R}\times\{0\}) \overset{(c)}{=} D_{\psi_2}(\mathbb{R}\times\{0\}) = U_2$, d.h. für $\varphi := \psi_2 - \psi_1$ gilt $D_\varphi(U_1) = U_2$. Andererseits berechnen wir für alle $z \in \mathbb{C}$:

$$S_{U_2} \circ S_{U_1}(z) = S_{U_2}(e^{2i\psi_1}\overline{z}) = e^{2i\psi_2}\overline{(e^{2i\psi_1}\overline{z})} = e^{i2(\psi_2-\psi_1)}z = D_{2\varphi}(z)$$

(e): Ist $U = \mathrm{span}_{\mathbb{R}}\{e^{i\psi}\}$, so gilt

$$D_{2\varphi} \circ S_U(z) = e^{2i\varphi}e^{2i\psi}\overline{z} = e^{2i(\varphi+\psi)}\overline{z} = S_{D_\varphi(U)}(z),$$

da $D_\varphi(U) = \mathrm{span}_{\mathbb{R}}\{D_\varphi(e^{i\psi})\} = \mathrm{span}_{\mathbb{R}}\{e^{i(\varphi+\psi)}\}$.

Ebenso erhalten wir:

$$S_U \circ D_{2\varphi}(z) = e^{2i\psi}(e^{-2i\varphi}\overline{z}) = e^{2i(\psi-\varphi)}\overline{z} = S_{D_{-\varphi}(U)}(z).$$

Normalform von orthogonalen Abbildungen

Wir erinnern uns an Definition (10.5):
Sei $L \in \mathrm{End}(V)$. Ein Untervektorraum U von V heißt <u>L-invariant</u>, falls $L(U) \subseteq U$ gilt.

Bem.: $\{0\}$ und V sind L-invariant für jedes $L \in \mathrm{End}(V)$.

Ein großes Ziel bei der Untersuchung eines $L \in \mathrm{End}(V)$ ist es, L-invariante Unterräume $U_1 \neq \{0\}, U_2 \neq \{0\}$ zu finden, so daß V die direkte Summe von

U_1 und U_2 ist, d.h. $V = U_1 \oplus U_2$. Dann reduziert sich die Untersuchung von L auf die Untersuchung von $L|U_1 \in \mathrm{End}(U_1)$ und $L|U_2 \in \mathrm{End}(U_2)$. (Leider ist das nicht immer möglich, z.B. nicht für die "Scherung" $L \in \mathrm{End}(\mathbb{R}^2)$ mit $\mathrm{Mat}(L) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.)

Man wird natürlich versuchen, solche L-invarianten Unterräume U_1 und U_2 in noch kleinere L-invariante Unterräume zu zerlegen, und dazu benötigen wir den Begriff der direkten Summe von endlich vielen Unterräumen (vgl. (3.20)-(3.22) und Blatt 6, Aufgabe 2): Sind $U_1, \ldots, U_k$ Unterräume eines Vektorraums V, so heißt V die direkte Summe von $U_1, \ldots, U_k$ (geschrieben $V = U_1 \oplus \ldots \oplus U_k = \bigoplus\limits_{i=1}^{k} U_i$), falls gilt:

(i) $V = \mathrm{span}(\bigcup\limits_{i=1}^{k} U_i)$ und

(ii) Für alle $i \in \{1, \ldots, k\}$ gilt: $U_i \cap \mathrm{span}(\bigcup\limits_{\substack{j=1 \\ j \neq i}}^{k} U_j) = \{0\}$.

Daraus folgt: Ist B_i für $i = 1, \ldots, k$ eine Basis von U_i, so gilt $B_i \cap B_j = \emptyset$ für $i \neq j$, und $B := \bigcup\limits_{i=1}^{k} B_i$ ist Basis von V. Ist $\dim V < \infty$, so folgt:

$$\dim V = \sum_{i=1}^{k} \dim U_i.$$

Im Fall eines euklidischen Vektorraums $(V, \langle , \rangle)$ ist folgender Spezialfall der direkten Summe wichtig:

(11.2) Def.: Seien $U_1, \ldots, U_k$ Unterräume von V. Dann heißt V die orthogonale Summe von $U_1, \ldots, U_k$, falls gilt:

(i) $V = \mathrm{span}(\bigcup\limits_{i=1}^{k} U_i)$ und

(ii) Für alle $1 \leq i \neq j \leq k$ gilt $U_i \subseteq U_j^{\perp}$.

Zur Bedingung (ii) sagt man, die U_i, $1 \leq i \leq k$, seien paarweise orthogonal.

Aus (11.2)(i) und (ii) folgt, daß V die direkte Summe von $U_1, \ldots, U_k$ ist.

(11.3) Satz. *Sei* $(V, \langle , \rangle)$ *n-dimensionaler euklidischer Vektorraum und* $L \in \mathrm{O}(V)$. *Dann existieren* $k \in \mathbb{N}$, *2-dimensionale L-invariante Unterräume* $U_1, \ldots, U_k$ *von* V *und L-invariante Unterräume* U_- *und* U_+ *von* V, *so daß* V *die orthogonale Summe von* $U_1, \ldots, U_k$, U_-,U_+ *ist und so daß gilt*

(i) $L|U_i \in \mathrm{SO}(U_i) \setminus \{\pm \mathrm{id}_{U_i}\}$ *für* $1 \leq i \leq k$

(ii) $L|U_- = -(\mathrm{id}_{U_-})$

(iii) $L|U_+ = \mathrm{id}_{U_+}$

Bem.: 1) Es kann $k = 0$ oder $U_- = \{0\}$ oder $U_+ = \{0\}$ gelten.
2) Jedes $v \in V$ läßt sich dann eindeutig darstellen als

$$v = u_1 + \ldots + u_k + u_- + u_+$$

mit $u_i \in U_i$ für $1 \leq i \leq k$, $u_- \in U_-$ und $u_+ \in U_+$. Es gilt dann:

$$L(v) = L(u_1) + \ldots + L(u_k) - u_- + u_+.$$

L setzt sich also aus k "Drehungen" in den 2-dimensionalen Unterräumen U_i, $1 \leq i \leq k$, aus der "Punktspiegelung" $-(\mathrm{id}_{U_-})$ im Unterraum U_- und der Identität auf U_+ zusammen.

3) Es gilt $2k + \dim U_- + \dim U_+ = n$. Ist n ungerade, so folgt $\dim U_- + \dim U_+ \neq 0$.

4) $\det(L) = (-1)^{\dim U_-}$, vgl. LA I, Blatt 14, Aufgabe 1.

Umformuliert für Matrizen besagt (11.3):

(11.3)' Satz. *Sei* $L \in \mathrm{O}(V)$. *Dann existieren eine* ONB $\mathcal{G} = (v_1, \ldots, v_n)$ *von*

V, $k \in \mathbb{N}$ *und* $\varphi_1, \dots, \varphi_k \in (0, \pi)$, *so daß gilt:*

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \begin{pmatrix} \boxed{\begin{matrix}\cos\varphi_1 & -\sin\varphi_1\\ \sin\varphi_1 & \cos\varphi_1\end{matrix}} & & & & & & & & & \\ & \boxed{\begin{matrix}\cos\varphi_2 & -\sin\varphi_2\\ \sin\varphi_2 & \cos\varphi_2\end{matrix}} & & & & & 0 & & & \\ & & \ddots & & & & & & & \\ 0 & & & \boxed{\begin{matrix}\cos\varphi_k & -\sin\varphi_k\\ \sin\varphi_k & \cos\varphi_k\end{matrix}} & & & & & & \\ & & & & -1 & & & & & \\ & & & & & \ddots & & & & \\ & & & & & & -1 & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{pmatrix}$$

Äquivalent dazu ist: *Zu jedem* $A \in O(n)$ *existiert ein* $B \in O(n)$, *so daß* $B^T AB$ *die obige Form hat.*

Der wichtigste Schritt im Beweis von (11.3) ist

(11.4) Lemma. Sei $\dim V > 0$ und $L \in \mathrm{O}(V)$. Dann existiert ein L-invarianter Unterraum U von V mit $0 < \dim U \leq 2$.

Bew.: Wir betrachten die Hilfsabbildung

$$H := L + L^* \in \mathrm{End}(V).$$

Dieses H ist selbstadjungiert, da $H^* = L^* + L^{**} = L^* + L = H$ gilt. Nach Satz (10.8)(c) existiert ein EW $\lambda \in \mathbb{R}$ von H mit zugehörigem Eigenvektor $v \neq 0$. Wegen $L \in \mathrm{O}(V)$ gilt $L^* = L^{-1}$, und damit

$$H(v) = L(v) + L^{-1}(v) = \lambda v.$$

Wir wenden L auf diese Gleichung an und erhalten

$$L^2(v) + v = \lambda L(v)$$

oder

$$L^2(v) = \lambda L(v) - v.$$

Daraus folgt, daß $U := \mathrm{span}\{v, L(v)\}$ L-invariant ist. Wegen $v \neq 0$ gilt $0 < \dim U \leq 2$.

Bem.: $v \in V$ ist EV von H zum EW $+2$ (bzw. -2) $\Leftrightarrow$ $v \in V$ ist EV von L zum EW $+1$ (bzw. -1), vgl. Blatt 7, Aufgabe 1.

Beweis von (11.3): Durch vollständige Induktion nach $n = \dim V$.

Induktionsanfang: Ist $\dim V = 1$, so gilt $\mathrm{O}(V) = \{\pm \mathrm{id}_V\}$. Die Behauptung ist richtig mit $k = 0$ und entweder $U_+ = V$ oder $U_- = V$.

Induktionsschritt: Sei $L \in \mathrm{O}(V)$ und $\dim V = n > 1$. Wegen (8.19) genügt es, den Fall zu betrachten, daß $(V, \langle, \rangle)$ der $\mathbb{R}^n$ mit dem Standardskalarprodukt ist. Aus Lemma (11.4) folgt, daß ein L-invarianter Unterraum $U \subseteq \mathbb{R}^n$ existiert mit $0 < \dim U \leq 2$.

(a) Gilt $n = 2$ und $U = \mathbb{R}^n$, so folgt die Behauptung aus (11.1):
Entweder L ist eine Spiegelung ($\Rightarrow \mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ für eine geeignete ONB des $\mathbb{R}^2$), oder $L = D_\varphi$ für ein $\varphi \in [0, 2\pi)$.
Ist $\varphi = 0$, so $L = \mathrm{id}_{\mathbb{R}^2}$ ($\Rightarrow k = 0$, $U_- = \{0\}$, $U_+ = \mathbb{R}^2$).
Ist $\varphi = \pi$, so $L = -\mathrm{id}_{\mathbb{R}^2}$ ($\Rightarrow k = 0$, $U_- = \mathbb{R}^2$, $U_+ = \{0\}$). Sonst gilt $D_\varphi \in \mathrm{SO}(\mathbb{R}^2) \setminus \{\pm \mathrm{id}_{\mathbb{R}^2}\}$.

(b) Gilt $\dim U < n$, so folgt aus (8.12)

$$\mathbb{R}^n = U \oplus U^\perp$$

Da U L-invariant ist und L orthogonal ist, ist auch $U^\perp$ L-invariant (Übung!). Wegen $\dim U^\perp = n - \dim U < n$ ist auf $L|U^\perp$ die Induktionsvoraussetzung anwendbar. Daraus folgt zusammen mit unseren Kenntnissen über $O(\mathbb{R})$ und $O(\mathbb{R}^2)$ die Behauptung.

Die einzige zusätzliche Information, die man zum Beweis von (11.3)' benötigt, ist folgende: Ist $\dim V = 2$ und $L \in \mathrm{SO}(V) \setminus \{\pm \mathrm{id}_V\}$, so existiert eine ONB $\mathcal{G}$ von V und $\varphi \in (0, \pi)$, so daß $\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$ gilt. Das sieht man so ein. Ist $\tilde{\mathcal{G}} = (v_1, v_2)$ irgendeine ONB von V, so existiert $\tilde{\varphi} \in (0, \pi) \cup (\pi, 2\pi)$ mit

$$\mathrm{Mat}^{\tilde{\mathcal{G}}}_{\tilde{\mathcal{G}}}(L) = \begin{pmatrix} \cos\tilde{\varphi} & -\sin\tilde{\varphi} \\ \sin\tilde{\varphi} & \cos\tilde{\varphi}. \end{pmatrix}$$

Ist $\tilde{\varphi} \in (\pi, 2\pi)$, so betrachten wir $\mathcal{G} = (v_1, -v_2)$ und erhalten

$$\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L) = \begin{pmatrix} \cos\tilde{\varphi} & \sin\tilde{\varphi} \\ -\sin\tilde{\varphi} & \cos\tilde{\varphi}. \end{pmatrix}$$

Wegen $\cos(-\tilde{\varphi}) = \cos\tilde{\varphi}$, $\sin(-\tilde{\varphi}) = -\sin\tilde{\varphi}$ folgt mit $\varphi := 2\pi - \tilde{\varphi} \in (0, \pi)$:

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi, \end{pmatrix}$$

wie behauptet.

Die anschauliche Begründung für die letzte Überlegung ist wie folgt: Ist $\dim V = 2$ und $L \in \mathrm{SO}(V)$, so ist der "Drehwinkel" $\tilde{\varphi}$ von L erst nach Wahl einer Orientierung ($\Rightarrow$ eines "positiven Drehsinns") für V definiert. Ändert man die Orientierung, so geht $\tilde{\varphi}$ in $-\tilde{\varphi}$ über. Ist $\tilde{\varphi} \in (\pi, 2\pi)$, so ist $-\tilde{\varphi} \in (-2\pi, -\pi)$ und statt $-\tilde{\varphi}$ kann man natürlich auch $\varphi := 2\pi - \tilde{\varphi} \in (0, \pi)$ nehmen.

Oft ist es wichtig zu wissen, ob die Unterräume in (11.3) und ob die Normalform in (11.3)' und die zugehörige ONB eindeutig durch L bestimmt sind.

Hierzu kann man folgendes sagen:
Es gilt $U_+ = \ker(L - \mathrm{id}_V)$, $U_- = \ker(L + \mathrm{id}_V)$, so daß U_-, U_+ und damit auch $k = \frac{1}{2}(n - (\dim U_- + \dim U_+))$ eindeutig durch L bestimmt sind. Die $\varphi_1, \ldots, \varphi_k \in (0, \pi)$ sind (bis auf ihre Reihenfolge) eindeutig durch L bestimmt (und damit auch die Normalform in (11.3)'). Aus (11.3)' folgt nämlich sofort:

$$\det(L - \lambda \mathrm{id}) = \prod_{j=1}^{k} \underbrace{(\lambda^2 - 2(\cos\varphi_j)\lambda + 1)}_{(\lambda - e^{i\varphi_j})(\lambda - e^{-i\varphi_j})} \cdot (1+\lambda)^{\dim U_-} \cdot (1-\lambda)^{\dim U_+},$$

so daß $e^{i\varphi_j}$ und $e^{-i\varphi_j}$ für $j = 1, \ldots, k$ gerade die Nullstellen in $\mathbb{C} \setminus \mathbb{R}$ des charakteristischen Polynoms von L sind. Also sind $\cos\varphi_j = \frac{1}{2}(e^{i\varphi_j} + e^{-i\varphi_j})$ und damit auch $\varphi_j \in (0, \pi)$ eindeutig durch L bestimmt.

Eine ONB, in der die Matrix von L die Normalform (11.3)' annimmt, ist nicht eindeutig durch L bestimmt. Z.B. ist die Matrix einer Drehung, $D_\varphi \in \mathrm{SO}(\mathbb{R}^2)$, $\varphi \in (0, \pi)$, bezüglich jeder positiv orientierten ONB des $\mathbb{R}^2$ in der Normalform $\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$.

Wir geben die möglichen Normalformen einer Matrix $A \in \mathrm{O}(3)$ an und die geometrischen Bezeichnungen für die orthogonalen Abbildungen des $\mathbb{R}^3$, die bezüglich einer geeigneten ONB diese Normalform haben.

a) $\begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix}$ eigentliche Drehung

b) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_3$ Identität

c) $\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ Geradenspiegelung

a') $\begin{pmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & -1 \end{pmatrix}$ Drehspiegelung

b') $\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = -E_3$ Punktspiegelung

c') $\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ (Ebenen-)Spiegelung

Bemerkungen zur Berechnung der Normalform von orthogonalen Matrizen A:

Die Unterräume $U_\pm$ erhält man als Lösungsräume der homogenen linearen Gleichungssysteme

$$\big(A \mp E_n\big)x^T = 0\,.$$

Die Dimensionen von $U_\pm$ geben die Anzahl der Einträge ± 1 in der Diagonalen der Normalform an. Man kann die Normalform angeben, wenn man alle (auch die komplexen) Nullstellen des charakteristischen Polynoms

$$p(A) = \det(A - \lambda E)$$

von A mit ihrem Vielfachheiten kennt. Diese Nullstellen liegen auf dem Einheitskreis

$$S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$$

und jedes nichtreelle Nullstellenpaar $z = e^{i\varphi}, \overline{z} = e^{-i\varphi}$ mit $\varphi \in (0,\pi)$ führt zu einem "Drehkästchen"

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

in der Normalform.

Im allgemeinen wird es nicht möglich sein, diese Nullstellen explizit zu bestimmen. Im Fall von $A \in \mathrm{O}(3)$ erhält man jedoch:

$$\cos\varphi \quad = \quad \tfrac{1}{2}(\mathrm{spur}(A) - \det A)\,,$$

was die Berechnung enorm vereinfacht.

Alternativ kann man, wie im Beweis von Satz (11.3), zu $A \in \mathrm{O}(n)$ die symmetrische Matrix $B = A + A^T = A + A^{-1}$ betrachten. Nach (11.3) existiert eine Matrix $S \in \mathrm{O}(n)$, so daß $S^T\,\mathrm{AS}$ die in (11.3)' angegebene Normalform ist. Dann gilt

$$S^T\,\mathrm{BS} = S^T\,\mathrm{AS} + S^T A^T S = S^T\,\mathrm{AS} + (S^T\,\mathrm{AS})^T,$$

d.h. $S^T\,\mathrm{BS}$ ist die Diagonalmatrix

$$S^T BS = \begin{pmatrix} 2\cos\varphi_1 & & & & & & & & & & & \\ & 2\cos\varphi_1 & & & & & & & & & 0 & \\ & & 2\cos\varphi_2 & & & & & & & & & \\ & & & 2\cos\varphi_2 & & & & & & & & \\ & & & & \ddots & & & & & & & \\ & & & & & -2 & & & & & & \\ & 0 & & & & & \ddots & & & & & \\ & & & & & & & -2 & & & & \\ & & & & & & & & 2 & & & \\ & & & & & & & & & \ddots & & \\ & & & & & & & & & & & 2 \end{pmatrix}$$

Daraus folgt, daß sich die φ_i in der Normalform von A aus den EWen $\lambda_i \in (-2,2)$ von B berechnen lassen, und zwar als

$$\varphi_i = \arccos(\frac{\lambda_i}{2}) \in (0,\pi).$$

Es ist oft einfacher, die EWe λ_i der symmetrischen Matrix $B = A + A^T$ zu bestimmen, als direkt die von A. Dazu folgendes Rechenbeispiel im $\mathbb{R}^4$.

Bsp.: Für die Matrix

$$A = \frac{1}{9\sqrt{2}} \begin{pmatrix} 9 & 4 & 8 & 1 \\ -4 & 9 & -1 & 8 \\ -8 & 1 & 9 & -4 \\ -1 & -8 & 4 & 9 \end{pmatrix}$$

gilt $AA^T = E_4$, d.h. $A \in \mathrm{O}(4)$. Nun ist $A + A^T$ die Diagonalmatrix $\sqrt{2}E_4$, d.h. alle EWe von $A + A^T$ sind $\sqrt{2}$. Deshalb hat A nicht die EWe $+1$ oder -1, und die Normalform von A besteht aus 2 Drehkästchen mit dem Drehwinkel

$$\varphi_1 = \varphi_2 = \arccos\left(\frac{\sqrt{2}}{2}\right) \in (0, \pi),$$

d.h. $\varphi_1 = \varphi_2 = \frac{\pi}{4}$ und $\sin\varphi_1 = \sin\varphi_2 = \frac{\sqrt{2}}{2}$. Es existiert also ein $S \in \mathrm{O}(4)$, so daß

$$S^T\,\mathrm{AS} = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

gilt.

Nach so vielen Worten möchte man hoffen, nun alles über orthogonale Abbildungen zu wissen. Weit gefehlt! Über die Gruppenstruktur von $\mathrm{SO}(n)$ für $n \geq 3$ wurde etwa noch nichts gesagt. Man möchte auch gern die Elemente von $\mathrm{SO}(3)$ durch 3 (warum gerade 3?) reelle "Parameter" beschreiben (so wie wir die Elemente von $\mathrm{SO}(2)$ durch den Drehwinkel $\varphi \mod 2\pi$ beschrieben haben). Aber geht das und wie am besten? Da gibt es Fragen und Antworten, genug für ein ganzes Mathematikstudium...

12 Symmetrische Bilinearformen und quadratische Formen

A) Der Fall eines beliebigen Koeffizientenkörpers der Charakteristik $\neq 2$.

Es sei K ein Körper, in dem $2 := 1 + 1 \neq 0$ gilt. (Körper mit $2 \neq 0$ heißen von der Charakteristik $\mathrm{char}(K) \neq 2$). Wir erinnern uns daran, was eine symmetrische Bilinearform b auf einem K-Vektorraum V ist (vgl. (7.1) und (8.1)): b ist eine Abbildung

$$b : V \times V \to K,$$

so daß für alle $v, w, z \in V$ und alle $\alpha, \beta \in K$ gilt:

(a) $b(\alpha v + \beta w, z) = \alpha b(v, z) + \beta b(w, z)$

(b) $b(v, w) = b(w, v)$

Aus (a) und (b) folgt, daß b auch im 2. Argument linear ist.

Bsp.: 1) Jedes Skalarprodukt auf einem $\mathbb{R}$-Vektorraum ist eine symmetrische Bilinearform.
2) Auf dem $\mathbb{R}^4$ ist das "Lorentzprodukt", definiert durch

$$b(x, y) = \sum_{i=1}^{3} x_i y_i - c^2 x_4 y_4,$$

eine symmetrische Bilinearform, die weder positiv noch negativ definit ist. Dieses b ist das Grundobjekt der speziellen Relativitätstheorie (c ist die Lichtgeschwindigkeit).
3) Ist $f \in C^2(\mathbb{R}^n, \mathbb{R})$, so definiert man in der Analysis die 2. Ableitung $D^2 f(x_0)$ von f an der Stelle x_0 und $D^2 f(x_0)$ ist eine symmetrische Bilinearform auf dem $\mathbb{R}^n$.

(12.1) Def.: Ist b symmetrische Bilinearform auf dem Vektorraum V, so heißt

$$Q_b : V \to K, Q_b(v) := b(v, v)$$

die zu b gehörige quadratische Form.

Bsp.: 1) Ist b ein Skalarprodukt, so ist Q_b gerade das Quadrat der zugehörigen Norm.
2) Ist $A = (a_{ij}) \in K^{n \times n}$ symmetrisch und $x, y \in K^n \simeq K^{n \times 1}$, so definiert

$$b(x, y) = \sum_{i,j=1}^{n} a_{ij} x_i y_j = x^T A y$$

eine symmetrische Bilinearform. Die zugehörige quadratische Form ist

$$Q_b(x) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = x^T A x$$

3) Ist $f \in C^2(\mathbb{R}^n, \mathbb{R})$ und $x_0 \in \mathbb{R}^n$, so ist "die Taylorentwicklung von f um x_0 bis zur Ordnung 2" folgende Aussage: Für alle $h \in \mathbb{R}^n$ gilt

$$f(x_0 + h) = f(x_0) + Df(x_0)(h) + \frac{1}{2} D^2 f(x_0)(h,h) + R(x_0, h) \parallel h \parallel^2$$

für eine Funktion $R(x_0, h)$ mit $\lim\limits_{h \to 0} R(x_0, h) = 0$. Hier tritt also gerade die zur symmetrischen Bilinearform $D^2 f(x_0)$ gehörige quadratische Form $h \to D^2 f(x_0)(h,h)$ auf.

(12.2) Fakt. Jede symmetrische Bilinearform b ist durch ihre quadratische Form Q_b eindeutig bestimmt (d.h. sind b, b' symmetrische Bilinearformen auf V und gilt $Q_b = Q_{b'}$, so gilt $b = b'$).

Bew.: Für alle $v, w \in V$ gilt

$$b(v, w) = \frac{1}{2}(Q_b(v + w) - Q_b(v) - Q_b(w)).$$

(Hier wird verwendet, daß wegen $2 \neq 0$ das Inverse $2^{-1} = \frac{1}{2}$ von 2 in K existiert!)

Obige Gleichung ergibt sich aus folgender Rechnung:

$$\begin{aligned} Q_b(v + w) - Q_b(v) - Q_b(w) &= b(v + w, v + w) - b(v, v) - b(w, w) \\ &= b(v, w) + b(w, v) = 2b(v, w) \end{aligned}$$

(12.3) Def.: Es sei $1 \leq \dim V = n < \infty$ und $\mathcal{G} = (v_1, \ldots, v_n)$ geordnete Basis von V. Ist $b : V \times V \to K$ symmetrische Bilinearform, so heißt die symmetrische Matrix

$$B := (b(v_i, v_j))_{1 \leq i,j \leq n} \in K^{n \times n}$$

die Matrix von b bezüglich $\mathcal{G}$.

Bsp.: 1) Die Matrix bezüglich der Standardbasis der symmetrischen Bilinearform b auf K^n, die im obigen Beispiel 2) mittels der symmetrischen Matrix

A definiert wurde, ist gerade dieses A.
2) Ist $f \in C^2(\mathbb{R}^n, \mathbb{R})$ und $x_0 \in \mathbb{R}^n$, so ist die Matrix von $D^2 f(x_0)$ bezüglich der Standardbasis $(e_1, \ldots, e_n)$ des $\mathbb{R}^n$ gerade die Matrix der zweiten partiellen Ableitungen $(\partial_i \partial_j f(x_0))_{1 \leq i,j \leq n}$ von f an der Stelle x_0.

(12.4) Fakt. Sei $\mathcal{G} = (v_1, \ldots, v_n)$ geordnete Basis von V. Dann ist die Abbildung, die jeder symmetrischen Bilinearform auf V ihre Matrix bezüglich $\mathcal{G}$ zuordnet, eine Bijektion auf die Menge der symmetrischen $(n \times n)$-Matrizen mit Koeffizienten in K.

Bem.: Die Menge der symmetrischen Bilinearformen auf V besitzt eine offensichtliche Struktur als K-Vektorraum, und bezüglich dieser Struktur ist die Abbildung in (12.4) ein Isomorphismus auf den Untervektorraum der symmetrischen Matrizen in $K^{n \times n}$.

Bew.: (i) Surjektivität: Ist $B = (b_{ij}) \in K^{n \times n}$ symmetrisch und sind $v = \sum\limits_{i=1}^{n} x_i v_i \in V$, $w = \sum\limits_{i=1}^{n} y_i v_i \in V$, so definieren wir

$$b(v, w) := \sum_{i,j=1}^{n} b_{ij} x_i y_j.$$

Dann ist b symmetrische Bilinearform, deren Matrix bezüglich $\mathcal{G} = (v_1, \ldots, v_n)$ gerade die gegebene Matrix $B = (b_{ij})$ ist.

(ii) Injektivität: Sind b und b' symmetrische Bilinearformen auf V und gilt $b(v_i, v_j) = b'(v_i, v_j)$ für alle $1 \leq i, j \leq n$, so folgt für alle $v = \sum\limits_{i=1}^{n} x_i v_i$, $w = \sum\limits_{j=1}^{n} y_j v_j$ in V:

$$b(v, w) = \sum_{i,j=1}^{n} x_i y_j b(v_i, v_j) = \sum_{i,j=1}^{n} x_i y_j b'(v_i, v_j) = b'(v, w).$$

(12.5) Def.: Sei b symmetrische Bilinearform auf V. Dann heißt der Untervektorraum

$$N_b := \{v \in V \mid \forall w \in V : b(v, w) = 0\}$$

von V der <u>Ausartungsraum von</u> b, und b heißt <u>nichtausgeartet</u>, wenn $N_b = \{0\}$ gilt.

Bem.: b nichtausgeartet $\Leftrightarrow \forall v \in V \exists w \in V : b(v, w) \neq 0$.

Exkurs über "Koordinaten und Koordinatentransformationen"

Eine geordnete Basis $\mathcal{G} = (v_1, \ldots, v_n)$ eines K-Vektorraums V gibt die Möglichkeit jeden Vektor $v = \sum\limits_{i=1}^{n} x_i v_i \in V$ "durch seine Koordinaten $(x_1, \ldots, x_n) \in K^n$ bezüglich $\mathcal{G}$ zu beschreiben". Präziser ausgedrückt heißt das, daß die Abbildung

$$I_{\mathcal{G}} : V \to K^n, I_{\mathcal{G}}\left(\sum_{i=1}^{n} x_i v_i\right) = (x_1, \ldots, x_n)$$

ein Isomorphismus ist. $I_{\mathcal{G}}$ ist gerade der Isomorphismus, für den $I_{\mathcal{G}}(v_i) = e_i$ für $1 \leq i \leq n$ gilt. Umgekehrt existiert für jeden Isomorphismus $I : V \to K^n$ eine geordnete Basis $\mathcal{G} = (v_1, \ldots, v_n)$ von V, so daß $I = I_{\mathcal{G}}$ gilt, nämlich $v_i := I^{-1}(e_i)$ für $1 \leq i \leq n$.

Man kann also geordnete Basen $\mathcal{G}$ mit Isomorphismen von V auf K^n identifizieren und diese Identifikation besteht gerade darin, daß jedes $\mathcal{G} = (v_1, \ldots, v_n)$ als "Koordinatensystem" angesehen wird, mit dessen Hilfe man einen Vektor $v = \sum\limits_{i=1}^{n} x_i v_i$ durch seine Koordinaten $(x_1, \ldots, x_n) \in K^n$ bezüglich $\mathcal{G}$ beschreibt. Man fragt sich dann, wie die Koordinaten von $v \in V$ bezüglich verschiedener Basis $\mathcal{G} = (v_1, \ldots, v_n)$ und $\mathcal{G}' = (v'_1, \ldots, v'_n)$ miteinander zusammenhängen. Die Abbildung

$$L = L_{\mathcal{G},\mathcal{G}'} : K^n \to K^n,$$

die die Koordinaten von $v \in V$ bezüglich $\mathcal{G}$ in die Koordinaten von $v \in V$ bezüglich $\mathcal{G}'$ verwandelt, nennt man die Koordinatentransformation von den Koordinaten bezüglich $\mathcal{G}$ in die Koordinaten bezüglich $\mathcal{G}'$. L ist also dadurch definiert, daß

$$L \circ I_{\mathcal{G}} = I_{\mathcal{G}'} \text{ oder } L = I_{\mathcal{G}'} \circ I_{\mathcal{G}}^{-1}$$

gilt. Insbesondere gilt $L \in \mathrm{Aut}(K^n)$. Ist $T = (t_{ij}) \in K^{n \times n}$ durch

$$(*) \qquad v_j = \sum_{i=1}^{n} t_{ij} v'_i \quad \text{für } 1 \leq j \leq n$$

definiert, d.h. $T = \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}'}(\mathrm{id}_V)$, so gilt

$$\mathrm{Mat}(L) = T,$$

denn $L(e_j) = I_{\mathcal{G}'} \circ I_{\mathcal{G}}^{-1}(e_j) = I_{\mathcal{G}'}(v_j) = I_{\mathcal{G}'}\left(\sum\limits_{i=1}^{n} t_{ij} v_i'\right) = \sum\limits_{i=1}^{n} t_{ij} e_i.$

Explizit bedeutet das: Sind $(x_1, \ldots, x_n)$ die Koordinaten von v bezüglich $\mathcal{G}$ und $(x_1', \ldots, x_n')$ die Koordinaten von v bezüglich $\mathcal{G}'$, so gilt

$$(**) \qquad x_i' = \sum_{i=1}^{n} t_{ij} x_j \quad \text{für } 1 \leq j \leq n,$$

wobei die Matrix (t_{ij}) durch $(*)$ definiert ist. Man beachte dabei, daß in $(*)$ die v_j durch die v_j' ausgedrückt werden. $T = (t_{ij})$ ist gerade die Inverse der Matrix $S = (s_{ij})$, die durch

$$v_j' = \sum_{i=1}^{n} s_{ij} v_i \text{ für } 1 \leq j \leq n$$

definiert ist.

Ist umgekehrt eine geordnete Basis $\mathcal{G}$ von V und ein $L \in \mathrm{Aut}(K^n)$ gegeben, so existiert genau eine geordnete Basis $\mathcal{G}'$ von V, so daß $L \circ I_{\mathcal{G}} = I_{\mathcal{G}'}$ gilt, d.h. so daß L die Koordinatentransformation zu $\mathcal{G}$ und $\mathcal{G}'$ ist.

Zurück zu den Bilinearformen: Mit Hilfe des Isomorphismus $I_{\mathcal{G}} : V \to K^n$ läßt sich leicht folgende Aussage beweisen.

(12.6) Fakt. Sei b symmetrische Bilinearform auf V, $\mathcal{G} = (v_1, \ldots, v_n)$ Basis von V und B die Matrix von b bezüglich $\mathcal{G}$. Dann gilt:

(a) b ist genau dann nicht ausgeartet, wenn B nicht ausgeartet ist (d.h. wenn $\det B \neq 0$ gilt).

(b) Es sei $I_{\mathcal{G}} : V \to K^n \equiv K^{n\times 1}$ der Isomorphismus, der durch

$$I_{\mathcal{G}}(v_i) = e_i \text{ für alle } i \in \{1, \ldots, n\}$$

definiert ist. Dann gilt

$$I_{\mathcal{G}}(N_b) = \{x \in K^{n\times 1} | Bx = 0\}.$$

Bem.: Faßt man B durch $x \to Bx$ als Endomorphismus von $K^{n\times 1}$ auf, so ist $I_{\mathcal{G}}(N_b)$ also gerade der Kern von B.

Bew.: Es genügt, (b) zu beweisen. Aus der Definition von B folgt leicht, daß

$$b(v, w) = I_{\mathcal{G}}(v)^T B I_{\mathcal{G}}(w)$$

gilt. Ist nun $x = I_{\mathcal{G}}(v) \in N_b$, so folgt

$$0 = b(v, w) = (x^T B)(I_{\mathcal{G}}(w)) = (Bx)^T (I_{\mathcal{G}}(w))$$

für alle $w \in V$. Da $I_{\mathcal{G}}$ surjektiv ist, folgt

$$(Bx)^T y = 0$$

für alle $y \in K^{n\times 1}$, und daraus $Bx = 0$. Ist umgekehrt $x \in K^{n\times 1}$ und $Bx = 0$, so zeigt die gleiche Rechnung, daß für alle $w \in V$

$$b((I_{\mathcal{G}})^{-1}(x)), w) = (Bx)^T (I_{\mathcal{G}}(w)) = 0$$

gilt, d.h. $I_{\mathcal{G}}^{-1}(x) \in N_b$.

Als nächstes untersuchen wir, wie die Matrizen einer symmetrischen Bilinearform bezüglich verschiedener Basen miteinander zusammenhängen.

(12.7) Fakt. Seien $\mathcal{G} = (v_1, \dots, v_n)$ und $\mathcal{G}' = (v_1', \dots, v_n')$ geordnete Basen von V und $S = \mathrm{Mat}_{\mathcal{G}'}^{\mathcal{G}}(\mathrm{id}_V)$, d.h. $S = (s_{ki}) \in \mathrm{GL}(n, K)$ mit $v_i' = \sum\limits_{k=1}^{n} s_{ki} v_k$ für $1 \le i \le n$. Ist b symmetrische Bilinearform auf V, so gilt für die Matrix B von b bezüglich $\mathcal{G}$ und die Matrix B' von b bezüglich $\mathcal{G}'$:

$$B' = S^T B S.$$

Bew.: Es gilt $B' = (b(v_i', v_j'))_{1\le i,j\le n}$, $B = (b(v_i, v_j))_{1\le i,j\le n}$ und

$$b(v_i', v_j') = b\left(\sum_{k=1}^{n} s_{ki} v_k, \sum_{l=1}^{n} s_{lj} v_l\right) = \sum_{l,j=1}^{n} s_{ki} b(v_k, v_l) s_{lj} = (S^T B S)_{ij}.$$

Bem.: 1) Man beachte den Unterschied zum Fall der Matrizen von Endomorphismen. Nach (5.6) gilt

$$\mathrm{Mat}_{\mathcal{G}'}^{\mathcal{G}'}(L) = S^{-1} \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) S.$$

2) Insbesondere gilt i.a.

$$\det B' = (\det S)^2 \det B \neq \det B.$$

Das weist daraufhin, daß für symmetrische Bilinearformen b allein keine "Determinante von b" definiert ist.

Es ist überraschend einfach, symmetrische Bilinearformen zu "diagonalisieren":

(12.8) Satz. *Sei V n-dimensionaler K-Vektorraum,* $\mathrm{char}(K) \neq 2$.

(a) *Ist b symmetrische Bilinearform auf V, so existiert eine Basis $\mathcal{G} = (v_1, \ldots, v_n)$ von V, für die*

$$b(v_i, v_j) = 0 \qquad \textit{für } 1 \leq i \neq j \leq n$$

gilt. Die zu b gehörige quadratische Form Q_b ist dann durch

$$Q_b\Big(\sum_{i=1}^{n} x_i v_i\Big) = \sum_{i=1}^{n} d_i x_i^2$$

gegeben mit $d_i := b(v_i, v_i)$.

(b) *Ist $B \in K^{n \times n}$ symmetrisch, so existiert $S \in \mathrm{GL}(n, K)$, so daß $S^T B S$ eine Diagonalmatrix ist.*

Bew.: Durch Induktion nach $n = \dim V$. Im Fall $n = 1$ ist nichts zu beweisen.

Induktionsschritt: 1. Fall: Ist $b(v, v) = 0$ für alle $v \in V$, so folgt aus (12.2), dass $b = 0$ gilt. Im Fall $b = 0$ ist die Behauptung trivial, da die Matrix von b bezüglich jeder Basis die 0-Matrix ist.

2. Fall: Wir können ein $v \in V$ wählen, für das $b(v, v) \neq 0$ gilt. Dann ist die Abbildung

$$l : V \to K, \quad l(w) := b(v, w)$$

eine Linearform auf V mit $l(v) \neq 0$. Nach dem Dimensionssatz für lineare Abbildungen (4.10) ist

$$(*) \qquad U := \ker(l) = \big\{ w \in V \mid b(v, w) = 0 \big\}$$

ein $(n-1)$-dimensionaler Unterraum von V. Jetzt wenden wir die Induktionsvoraussetzung auf $\tilde{b} := b\big|U \times U$ an und erhalten eine Basis $v_1, \ldots, v_{n-1}$ von U, so dass $b(v_i, v_j) = 0$ für alle $1 \leq i \neq j \leq n-1$ gilt. Wegen $v \notin U$

ist dann $v_1, \ldots, v_{n-1}, v_n := v$ eine Basis von V, und wegen $(*)$ gilt auch $b(v_i, v_n) = b(v_n, v_i) = 0$ für alle $1 \leq i < n$, d.h. für die Basis $\mathcal{G} = (v_1, \ldots, v_n)$ ist unsere Behauptung erfüllt.

Bez.: Wir werden Basen $\mathcal{G} = (v_1, \ldots, v_n)$ mit $b(v_i, v_j) = 0$ für alle $1 \leq i \neq j \leq n$ b-Orthogonalbasen nennen.

Bem.: 1) Es gibt i.a viele b-orthogonale Basen. Ist etwa $K = \mathbb{R}$ und b ein Skalarprodukt, so ist jede ONB für b auch b-orthogonal.
2) Ersetzt man eine b-Orthogonalbasis $\mathcal{G} = (v_1, \ldots, v_n)$ durch $\mathcal{G}' = (\alpha_1 v_1, \ldots, \alpha_n v_n)$ für beliebige $\alpha_i \in K \setminus \{0\}$, so ist $\mathcal{G}'$ ebenfalls b-orthogonal, und es gilt

$$d_i' := b(v_i', v_i') = \alpha_i^2 b(v_i, v_i) =: \alpha_i^2 d_i.$$

Wir sehen daran, daß die Diagonalelemente d_i keineswegs durch b eindeutig bestimmt sind. Ist $K = \mathbb{C}$, so kann man für alle $i \in \{1, \ldots, n\}$, für die $d_i \neq 0$ gilt, erreichen daß $d_i' = 1$ ist, indem man α_i als Lösung der Gleichung $\alpha_i^2 = \frac{1}{d_i}$ wählt, d.h. $\alpha_i = \pm\sqrt{\frac{1}{d_i}} \in \mathbb{C}$. Man erhält im Fall $K = \mathbb{C}$ eine eindeutig bestimmte <u>Normalform</u>

$$(b(v_i, v_j))_{1 \leq i,j \leq n} = \begin{pmatrix} \overbrace{\begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix}}^{r_b} & 0 \\ 0 & \underbrace{\begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix}}_{\dim N_b} \end{pmatrix}.$$

Im Fall $K = \mathbb{C}$ besitzen also alle symmetrischen Bilinearformen, die den gleichen Rang r_b haben, die gleiche Normalform. Man sagt dazu, der Rang r_b sei die einzige Invariante einer symmetrischen Bilinearform auf einem n-dimensionalen $\mathbb{C}$-Vektorraum.

Der Beweis von Satz (12.8) enthält ein Verfahren für die Bestimmung einer b-Orthogonalbasis $\mathcal{G} = (v_1, \ldots, v_n)$. Es genügt dabei, homogene lineare Gleichungen zu lösen (Bestimmung von U im Beweis von (12.8)) und für b (und Restriktionen von b auf Unterräume) Elemente v zu suchen, für die $b(v, v) \neq 0$ ist (und das ist ganz leicht, wenn b explizit gegeben ist). Eine Alternative ist folgendes

(12.9) *Verfahren der quadratischen Ergänzung*: Es sei b symmetrische Bilinearform auf V, die bezüglich einer gegebenen Basis $\mathcal{G} = (v_1, \ldots, v_n)$ durch die symmetrische Matrix $B = (b_{ij}) \in K^{n \times n}$ gegeben ist. Wir betrachten ihre quadratische Form

$$Q_b\left(\sum_{i=1}^n x_i v_i\right) = \sum_{i,j=1}^n b_{ij} x_i x_j.$$

Unser Ziel ist, eine Koordinatentransformation

$$x'_i = \sum_{j=1}^n t_{ij} x_j \text{ für } 1 \leq i \leq n$$

mit $T := (t_{ij})_{1 \leq i,j \leq n} \in \mathrm{GL}(n, K)$ zu finden, so daß

$$Q_b\left(\sum_{i=1}^n x_i v_i\right) = \sum_{i=1}^n d_i (x'_i)^2$$

gilt. Ist dann $S = (s_{ij}) = T^{-1}$, so gilt für die durch

$$(*) \qquad v'_j := \sum_{i=1}^n s_{ij} v_i$$

definierte Basis $\mathcal{G}' = (v'_1, \ldots, v'_n)$

$$b(v'_i, v'_j) = \delta_{ij} d_i,$$

vgl. (12.2) und den Exkurs über Koordinaten und Koordinatentransformationen. Das Verfahren hat die Eigenschaft, daß stets $t_{ij} = 0$ für $i > j$ gilt, so daß die Berechnung der Inversen T^{-1} sehr einfach ist.

Wir schildern das Verfahren zunächst an einem Beispiel einer explizit gegebenen symmetrischen Bilinearform b auf einem 3-dimensionalen $\mathbb{R}$-Vektorraum V. Bezüglich einer gegebenen Basis $\mathcal{G} = (v_1, v_2, v_3)$ von V habe b die Matrix

$$(b(v_i, v_j))_{1 \leq i,j \leq 3} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Dann gilt

$$\begin{aligned}
Q_b\left(\sum_{i=1}^{3} x_i v_i\right) &= x_1^2 + 4x_1x_2 + 2x_1x_3 + x_2^2 + 2x_2x_3 + x_3^2 \\
&= x_1^2 + 2x_1(2x_2 + x_3) + x_2^2 + 2x_2x_3 + x_3^2 \\
&\overset{\text{quadrat. Ergänzung}}{=} (x_1 + (2x_2 + x_3))^2 - (2x_2 + x_3)^2 + x_2^2 \\
&\quad +2x_2x_3 + x_3^2.
\end{aligned}$$

Wir setzen $x_1' := x_1 + 2x_2 + x_3$ und rechnen weiter

$$\begin{aligned}
&= (x_1')^2 - 3x_2^2 - 2x_2x_3 \\
&\overset{\text{quadrat. Ergänzung}}{=} (x_1')^2 - 3(x_2 + \tfrac{1}{3}x_3)^2 + \tfrac{1}{3}x_3^2.
\end{aligned}$$

Wir setzen $x_2' := x_2 + \frac{1}{3}x_3$, $x_3' := x_3$ und erhalten in diesen Koordinaten folgende Diagonalform für Q_b:

$$Q_b\left(\sum_{i=1}^{3} x_i v_i\right) = (x_1')^2 - 3(x_2')^2 + \frac{1}{3}(x_3')^2.$$

Für die meisten Zwecke genügt es, diese Diagonalform von Q_b zu kennen. Wir können aber auch noch, wie oben beschrieben, die zu dieser Koordinatentransformation $x_i' = \sum\limits_{j=1}^{3} t_{ij}x_j$ gehörende neue Basis von V bestimmen: Nach Definition von x_1', x_2', x_3' gilt

$$T = (t_{ij}) = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix}.$$

Durch "Umstellen" der Gleichungen für x_1', x_2', x_3' berechnen wir x_1, x_2, x_3 aus x_1', x_2', x_3':

$$\begin{aligned}
x_3 &= x_3' \\
x_2 &= x_2' - \tfrac{1}{3}x_3 = x_2' - \tfrac{1}{3}x_3' \\
x_1 &= x_1' - 2x_2 - x_3 = x_1' - 2x_2' + \tfrac{2}{3}x_3' - x_3' = x_1' - 2x_2' - \tfrac{1}{3}x_3'
\end{aligned}$$

Also

$$S = T^{-1} = \begin{pmatrix} 1 & -2 & -\frac{1}{3} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix}$$

und damit nach $(*)$

$$\begin{array}{rcl} v_1' &=& v_1 \\ v_2' &=& -2v_1 + v_2 \\ v_3' &=& -\frac{1}{3}v_1 - \frac{1}{3}v_2 + v_3. \end{array}$$

Bezüglich dieser Basis gilt dann

$$Q_b\left(\sum_{i=1}^{3} x_i' v_i'\right) = (x_1')^2 - 3(x_2')^2 + \frac{1}{3}(x_3')^2$$

und die Matrix von b bezüglich (v_1', v_2', v_3') ist

$$(b(v_i', v_j'))_{1\le i,j\le 3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \left(= S^T \left(\begin{smallmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{smallmatrix}\right) S\right).$$

Es sollte klar sein, wie dieses an einem Beispiel vorgestellte Verfahren auch im allgemeinen funktioniert - jedenfalls mit folgender

Zusatzüberlegung. Der 1. Schritt ist nicht durchführbar, wenn $b(v_1, v_1) = 0$ gilt, d.h. wenn der Term $b(v_1, v_1)x_1^2$ in der Darstellung von $Q_b(\sum x_i v_i)$ verschwindet. Dann hilft folgende Fallunterscheidung weiter:

a) Falls ein i existiert mit $b(v_i, v_i) \neq 0$, so vertausche v_i mit v_1, d.h. $x_i' := x_1$, $x_1' := x_i$, $x_j' := x_j$ für $j \notin \{1, i\}$.

b) Falls $b(v_i, v_i) = 0$ für alle $i \in \{1, \ldots, n\}$ gilt, aber $b \neq 0$, so existieren $1 \le i \neq j \le n$ mit $b(v_i, v_j) \neq 0$. Setze dann $v_i' = v_i + v_j$, $v_l' = v_l$ für $l \neq i$ (d.h. $x_j' = x_j - x_i$, $x_l' = x_l$ für $l \neq j$). Dann gilt

$$b(v_i', v_i') = b(v_i + v_j, v_i + v_j) = 2b(v_i, v_j) \neq 0,$$

da $2 \neq 0$ und $b(v_i, v_j) \neq 0$.
Analog kann man verfahren, wenn in einem der weiteren Schritte der nächste quadratische Term fehlt.

Bsp.: $Q_b\left(\sum\limits_{i=1}^{3} x_i v_i\right) = x_1 x_2 - x_2 x_3.$

Setze $x_1' := x_1$, $x_2' := x_2 - x_1$, $x_3' = x_3$. Dann gilt

$$Q_b\left(\sum_{i=1}^{3} x_i v_i\right) = x_1'(x_2' + x_1') - (x_2' + x_1')x_3' = (x_1')^2 + x_1'x_2' - x_1'x_3' - x_2'x_3'.$$

B) Symmetrische Bilinearformen auf reellen Vektorräumen

Ab jetzt betrachten wir Vektorräume über dem Körper $K = \mathbb{R}$. Im Gegensatz zum Fall $K = \mathbb{C}$ gibt es im Fall $K = \mathbb{R}$ verschiedene Typen von symmetrischen Bilinearformen. Der Grund dafür sind die Anordnungseigenschaften von $\mathbb{R}$. Die folgenden Überlegungen sind in der Analysis von Bedeutung bei der Untersuchung der kritischen Punkte einer Funktion $f \in C^2(\mathbb{R}^n, \mathbb{R})$ mittels der 2. Ableitung (=Hesseform) von f in diesen Punkten.

(12.9) Def.: Sei b symmetrische Bilinearform auf einem $\mathbb{R}$-Vektorraum V.

(a) b heißt positiv definit, falls für alle $v \in V \setminus \{0\}$ gilt:

$$b(v, v) > 0.$$

(b) b heißt positiv semidefinit, falls für alle $v \in V$ gilt:

$$b(v, v) \geq 0.$$

(c) b heißt negativ definit, bzw. negativ semidefinit falls $-b$ positiv definit bzw. positiv semidefinit ist.

(d) b heißt indefinit, wenn b weder positiv semidefinit noch negativ semidefinit ist.

Bem.: 1) Die Bedingung (12.9)(a) kennen wir natürlich schon aus Def. (8.2), und sie besagt gerade, daß b ein Skalarprodukt ist.
2) Daß b indefinit ist, kann explizit auf folgende Weise ausgedrückt werden: Es existieren $v, w \in V$, so daß $b(v, v) < 0$ und $b(w, w) > 0$ gilt.
3) Sei $f \in C^2(\mathbb{R}^n, \mathbb{R})$ und $x_0 \in \mathbb{R}^n$ ein kritischer Punkt von f, d.h. $\mathrm{Df}(x_0) = 0$. Dann gilt:

$$\begin{array}{lcl} D^2 f(x_0) \text{ positiv definit} & \Rightarrow & x_0 \text{ ist lokales Minimum von } f \\ x_0 \text{ lokales Minimum von } f & \Rightarrow & D^2 f(x_0) \text{ positiv semidefinit} \\ D^2 f(x_0) \text{ negativ definit} & \Rightarrow & x_0 \text{ ist lokales Maximum von } f \\ x_0 \text{ lokales Maximum von } f & \Rightarrow & D^2 f(x_0) \text{ negativ semidefinit} \end{array}$$

Ist $D^2 f(x_0)$ indefinit, so ist nach den vorangehenden Aussagen x_0 kein lokales Extremum von f. Im von der Schule bekannten Fall $n = 1$ ist $D^2 f(x_0)$ einfach

durch $f''(x_0)$ gegeben. Für $n = 1$ ist ein kritischer Punkt x_0 höchstens im "Ausnahmefall", daß $f''(x_0) = 0$ gilt, kein lokales Extremum von f. Für $n > 1$ ist der Fall, daß $D^2f(x_0)$ indefinit ist, keineswegs ein Ausnahmefall.
4) Ist $\mathcal{G} = (v_1, \ldots, v_n)$ b-Orthogonalbasis und $d_i := b(v_i, v_i)$, so gilt:

$$\begin{array}{lcl} b \text{ positiv definit} & \Leftrightarrow & d_i > 0 \text{ für alle } i \in \{1, \ldots, n\} \\ b \text{ positiv semidefinit} & \Leftrightarrow & d_i \geq 0 \text{ für alle } i \in \{1, \ldots, n\} \\ b \text{ indefinit} & \Leftrightarrow & \text{Es existieren } i \text{ und } j \text{ in } \{1, \ldots, n\} \text{ mit} \\ & & d_i > 0, d_j < 0 \end{array}$$

Aus Satz (12.8) erhalten wir leicht

(12.10) Satz. *Sei V n-dimensionaler $\mathbb{R}$-Vektorraum und b eine symmetrische Bilinearform auf V. Dann existiert eine geordnete Basis $\mathcal{G} = (v_1, \ldots, v_n)$ von V, so daß die Matrix B von b bezüglich $\mathcal{G}$ folgende Normalform hat:*

$$B = \begin{pmatrix} \overbrace{\begin{matrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & \ddots & \\ & & & & & -1 \end{matrix}}^{r_b} & 0 \\ 0 & \underbrace{\begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix}}_{\dim N_b} \end{pmatrix} .$$

Bew.: Nach (12.8) existiert eine b-orthogonale Basis $\mathcal{G}' = (v'_1, \ldots, v'_n)$ von V. Wir können annehmen, daß $d'_i = b(v'_i, v'_i) > 0$ für $1 \leq i \leq p$ gilt und daß $d'_i < 0$ für $p < i \leq r_b$ gilt. Wir betrachten die neue Basis $\mathcal{G} = (v_1, \ldots, v_n)$, die durch

$$\begin{array}{lll} v_i := |d'_i|^{-\frac{1}{2}} v'_i & \text{für} & i \in \{1, \ldots, r_b\} \\ v_i := v'_i & \text{für} & i \in \{r_b + 1, \ldots, n\} \end{array}$$

definiert ist. Die Matrix von b bezüglich $\mathcal{G}$ hat dann die oben angegebene Normalform.

Es ist nicht ohne weiteres klar, daß die Anzahl der positiven bzw. negativen Diagonalelemente in der in (12.10) angegebenen "Normalform" eindeutig durch b bestimmt ist; es könnte ja sein, daß für verschiedene b-Orthogonalbasen diese Zahlen verschieden sind. Wir werden als nächstes zeigen, daß das nicht so ist, d.h. daß die Anzahl der positiven (und der negativen) Diagonalelemente eindeutig durch b festgelegt ist. Offensichtlich ist die Anzahl der negativen Diagonalelemente gerade der Rang r_b vermindert um die Anzahl der positiven Diagonalelemente.

(12.11) Def.: Sei b symmetrische Bilinearform auf einem $\mathbb{R}$-Vektorraum V. Dann heißt

$$\operatorname{ind}(b) := \sup\{\dim U \mid U \subseteq V \text{ Untervektorraum}, b|U \times U \text{ negativ definit}\}$$

der Index von b.

Bem.: Ist $\dim V = \infty$, so ist $\operatorname{ind}(b) = \infty$ möglich. Ist $\operatorname{ind}(b) < \infty$, so ist das Supremum in (12.11) offenbar ein Maximum, d.h. es existiert ein Untervektorraum U, auf dem b negativ definiert ist und für den $\dim U = \operatorname{ind}(b)$ gilt. Wir werden zeigen, daß das auch im Fall $\operatorname{ind}(b) = \infty$ richtig ist. Wir können also in (12.11) statt "sup" auch "max" schreiben.

Bsp.: Wir betrachten auf $\mathbb{R}^3$ die quadratische Form

$$Q_b(x_1, x_2, x_3) = x_1^2 + x_2^2 - x_3^2.$$

Die Nullstellenmenge

$$Z_b = \{x \in \mathbb{R}^3 \mid Q_b(x) = 0\}$$

ist ein Doppelkegel: Für jedes $c \in \mathbb{R}$ ist

$$Z_b \cap (\mathbb{R}^2 \times \{c\}) = \{(x_1, x_2, c) \mid x_1^2 + x_2^2 = c^2\}$$

gerade eine Kreislinie vom Radius $|c|$. Im "Inneren" des Doppelkegels gilt $Q_b < 0$, im "Äußeren" $Q_b > 0$. Auf jeder Geraden durch $0 \in \mathbb{R}^3$, die das Innere von Q_b trifft, ist b negativ definit. Es gilt also $\operatorname{ind}(b) \geq 1$. Andererseits schneidet jede Ebene (durch 0) im $\mathbb{R}^3$ das Äußere von Q_b, d.h. b ist auf keinem 2-dimensionalen Untervektorraum des $\mathbb{R}^3$ negativ (semi-) definit. Es folgt $\operatorname{ind}(b) = 1$, wie zu erwarten.

Der folgende Satz zeigt, daß für verschiedene b-Orthogonalsysteme die Anzahl der negativen Diagonalelemente in der Matrix von b gleich ist und mit $\operatorname{ind}(b)$ übereinstimmt.

(12.12) Trägheitssatz von Sylvester (James Joseph Sylvester 1814-1897, engl. Mathematiker). *Sei b symmetrische Bilinearform auf einem n-dimensionalen $\mathbb{R}$-Vektorraum V und sei $\mathcal{G} = (v_1, \dots, v_n)$ eine b-Orthogonalbasis von V. Dann gilt*

$$\#\{i | i \in \{1, \dots, n\} \text{ und } b(v_i, v_i) < 0\} = \operatorname{ind}(b).$$

Bew.: Auf $U_0 := \operatorname{span}\{v_i | i \in \{1, \dots, n\}, b(v_i, v_i) < 0\}$ ist b negativ definit und auf $U_1 := \operatorname{span}\{v_i | i \in \{1, \dots, n\}, b(v_i, v_i) \geq 0\}$ ist b positiv semidefinit, vgl. Blatt 8, Aufgabe 1. Nach Definition von $\operatorname{ind}(b)$ folgt

$$(*) \qquad \dim U_0 \leq \operatorname{ind}(b).$$

Ist andererseits $U \subseteq V$ ein Untervektorraum, auf dem b negativ definit ist und für den $\dim U = \operatorname{ind}(b)$ gilt, so folgt $U \cap U_1 = \{0\}$, also

$$\dim U + \dim U_1 = \dim(U + U_1) \leq \dim V = n,$$

vgl. den Dimensionssatz (3.23). Daraus folgt

$$\operatorname{ind}(b) = \dim U \leq n - \dim U_1 = \dim U_0.$$

Zusammen mit $(*)$ folgt $\operatorname{ind}(b) = \dim U_0$, und das ist gerade die Behauptung.

(12.13) Folgerung. Die in Satz (12.10) als "Normalform von b" bezeichnete Matrix ist unabhängig von der Wahl der (geeignet normierten) b-Orthogonalbasis und damit eindeutig durch b bestimmt.

Um den Rang r_b und den Index $\operatorname{ind}(b)$ (und damit die Normalform (12.10)) einer symmetrischen Bilinearform konkret zu berechnen, kann man das Verfahren der quadratischen Ergänzung (12.9) benutzen. Ist danach

$$Q_b\left(\sum_{i=1}^n x_i v_i\right) = \sum_{i=1}^n d_i (x_i')^2,$$

so ist

$$r_b = \#\{i | d_i \neq 0\}$$

und nach (12.12)

$$\operatorname{ind}(b) = \#\{i | d_i < 0\}.$$

Es folgt eine Version des Satzes von Sylvester, die ausschließlich in der Sprache der Matrizen formuliert ist:

(12.12)' Satz. *Sei* $B \in \mathbb{R}^{n \times n}$ *symmetrisch und* $S, S \in \mathrm{GL}(n, \mathbb{R})$ *so, daß*

$$S^T B S =: D \text{ und } \tilde{S}^T B \tilde{S} =: \tilde{D}$$

Diagonalmatrizen sind. Dann haben D *und* $\tilde{D}$ *die gleiche Anzahl negativer (und die gleiche Anzahl positiver) Diagonalelemente.*

Bew.: Sei $B = (b_{ij})_{1 \le i,j \le n}$ und b die symmetrische Bilinearform auf $\mathbb{R}^n$ mit $b(e_i, e_j) = b_{ij}$ für $1 \le i, j \le n$. Nach (12.7) sind

$$v_i := \sum_{j=1}^{n} s_{ji} e_j, \quad , 1 \le i \le n$$

und

$$\tilde{v}_i := \sum_{j=1}^{n} s_{ji} e_j, \quad , 1 \le i \le n$$

b-Orthogonalbasen des $\mathbb{R}^n$ bezüglich derer b die Matrix D bzw. $\tilde{D}$ hat. Nun folgt die Behauptung aus (12.12) (zur Normierung vgl. den Beweis zu (12.10)).

Als nächstes kommt eine "invariante" Umformulierung von Satz (12.12), d.h. eine Umformulierung in der keine Basen auftreten.

Bez.: Ein Untervektorraum $U \subseteq V$ heißt b-negativ, falls $b|U \times U$ negativ definit ist. U heißt ein maximaler b-negativer Untervektorraum, falls jeder U enthaltende b-negative Untervektorraum gleich U ist.

(12.12)" Satz. *Sei* b *symmetrische Bilinearform auf einem endlich-dimensionalen* $\mathbb{R}$*-Vektorraum* V*. Dann gilt für jeden maximalen* b*-negativen Untervektorraum* U*:*

$$\dim U = \mathrm{ind}(b).$$

Bew.: Der Beweis von (12.8) zeigt, daß man eine b-Orthogonalbasis $\mathcal{G} = (v_1, \ldots, v_n)$ von V so konstruieren kann, daß $U = \mathrm{span}\{v_1, \ldots, v_{\dim U}\}$ gilt. Da U maximal b-negativ ist, gilt dann $b(v_i, v_i) \ge 0$ für $i > \dim U$. Daraus folgt

$$\dim U = \#\{i | i \in \{1, \ldots, n\}, b(v_i, v_i) < 0\} \overset{(12.12)}{=} \mathrm{ind}(b).$$

Wir erwähnen noch eine Anwendung von (12.12) auf den Fall $\mathrm{ind}(b) = \infty$.

Folgerung. Sei b symmetrische Bilinearform auf einem $\mathbb{R}$-Vektorraum V, und es gelte $\mathrm{ind}(b) = \infty$.

(a) Ist U maximaler b-negativer Untervektorraum von V, so gilt $\dim U = \infty$.

(b) Es existiert ein maximaler b-negativer Untervektorraum von V.

Bem.: Die Folgerung zeigt, daß man in Definition (12.11) "sup" durch "max" ersetzen kann.

Bew.: (a) Annahme: $\dim U < \infty$. Wegen $\mathrm{ind}(b) = \infty$ können wir einen b-negativen Untervektorraum $\tilde{U}$ von V finden, für den

$$(*) \qquad \dim U < \dim \tilde{U} < \infty$$

gilt. Dann ist $\tilde{V} := U + \tilde{U} \subseteq V$ endlichdimensional und $\tilde{b} := b|\tilde{V} \times \tilde{V}$ ist symmetrische Bilinearform auf $\tilde{V}$, für die U ein maximaler $\tilde{b}$-negativer Untervektorraum von $\tilde{V}$ ist. Wegen (12.12)" gilt

$$\dim U = \mathrm{ind}(\tilde{b}).$$

Da $\tilde{b}|\tilde{U} \times \tilde{U}$ negativ definit ist, folgt

$$\dim \tilde{U} \le \mathrm{ind}(\tilde{b}),$$

im Widerspruch zu $(*)$.

(b) Das Lemma von Zorn (3.15) angewendet auf die Menge

$$\mathcal{M} = \{U | U \; b\text{-negativer Untervektorraum von } V\} \subseteq \mathcal{P}(V)$$

liefert die Existenz eines maximalen b-negativen Untervektorraums von V.

(12.14) Def.: Eine symmetrische Matrix $B \in \mathbb{R}^{n \times n}$ heißt positiv definit, falls für alle $x \in \mathbb{R}^{n \times 1} \setminus \{0\}$ gilt:

$$x^T B x > 0.$$

Bem.: Sei b symmetrische Bilinearform auf einem n-dimensionalen $\mathbb{R}$-Vektorraum V, $\mathcal{G} = (v_1, \ldots, v_n)$ Basis von V und B die Matrix von b bezüglich $\mathcal{G}$. Dann gilt

$$b \text{ positiv definit} \Leftrightarrow B \text{ positiv definit}$$

Denn: $Q_b\left(\sum\limits_{i=1}^{n} x_i v_i\right) = x^T B x$ für $x^T := (x_1, \ldots, x_n)$.

Zum Abschluß dieses Abschnitts wird gezeigt, wie man mittels der "Hauptunterdeterminanten" von B feststellen kann, ob B positiv definit ist.

Bez.: Zu $B = (b_{ij})_{1\le i,j\le n}$ und $k \in \{1, \dots, n\}$ sei

$$B_k := (b_{ij})_{1\le i,j\le k}.$$

Dann heißt

$$\det B_k = \begin{vmatrix} b_{11} & \dots & b_{1k} \\ \vdots & & \vdots \\ b_{k1} & \dots & b_{kk} \end{vmatrix}$$

die k'te Hauptunterdeterminante von B. Es gilt also $B_1 = (b_{11})$, $\det B_1 = b_{11}$, und $B_n = B$, $\det B_n = \det B$.

(12.15) Satz. *Eine symmetrische Matrix $B \in \mathbb{R}^{n\times n}$ ist genau dann positiv definit, wenn für jedes $k \in \{1, \dots, n\}$ die Hauptunterdeterminante* $\det B_k$ *positiv ist.*

Bem.: Zum konkreten Rechnen taugt Satz (12.15) nur in kleinen Dimensionen, z.B. für $n = 2$, oder für speziell gebaute B. Im allgemeinen wird das Verfahren der quadratischen Ergänzung sehr viel günstiger sein, um in einem konkreten Fall zu entscheiden, ob B positiv definit ist oder nicht. Entscheidend für den Beweis ist folgende

Vorbemerkung: Ist b symmetrische Bilinearform auf einem n-dimensionalen $\mathbb{R}$-Vektorraum V, sind $\mathcal{G}$ und $\mathcal{G}'$ geordnete Basen von V und B und B' die Matrizen von b bezüglich $\mathcal{G}$ und $\mathcal{G}'$, so existiert nach (12.7) eine Matrix $S \in \mathrm{GL}(n, \mathbb{R})$, so daß

$$B' = S^T B S$$

gilt. Daraus folgt $\det B' = (\det S)^2 \det B$ mit $(\det S)^2 > 0$, und damit

$$(*) \qquad \det B > 0 \Leftrightarrow \det B' > 0.$$

Ist b positiv definit, so kann man für $\mathcal{G}'$ eine ONB bezüglich b wählen. Dann gilt $B' = E_n$ und $\det B' = 1 > 0$. Ist also b positiv definit und ist B die Matrix von b bezüglich einer (beliebigen) Basis $\mathcal{G}$ von V, so gilt

$$(**) \qquad \det B > 0.$$

Beweis von (12.15): Wir definieren durch

$$b(x, y) := x^T B y$$

eine symmetrische Bilinearform auf $\mathbb{R}^n \simeq \mathbb{R}^{n\times 1}$, deren Matrix bezüglich $\mathcal{G} = (e_1, \ldots, e_n)$ gerade B ist.

(a) B sei positiv definit. Dann ist b positiv definit und damit auch $b_k := b|(\mathbb{R}^k \times \{0\}) \times (\mathbb{R}^k \times \{0\})$ für jedes $k \in \{1, \ldots, n\}$. Nun ist B_k gerade die Matrix des Skalarprodukts b_k bezüglich der Basis $(e_1, \ldots, e_k)$ von $\mathbb{R}^k \times \{0\}$, also $\det(B_k) > 0$ nach $(**)$.

(b) Für jedes $k \in \{1, \ldots, n\}$ gelte $\det B_k > 0$. Wir zeigen durch Induktion nach k, daß dann $b_k := b|(\mathbb{R}^k \times \{0\}) \times (\mathbb{R}^k \times \{0\})$ für alle $k \in \{1, \ldots, n\}$ positiv definit ist. Also ist $b = b_n$ positiv definit, und damit auch die Matrix B von b bezüglich $(e_1, \ldots, e_n)$.

Induktionsanfang: Für $k = 1$ gilt $b(e_1, e_1) = b_{11} = \det(B_1) > 0$. Also ist b_1 positiv definit.

Induktionsschritt: Nach Induktionsvoraussetzung ist b_{k-1} positiv definit. Wir können also ein ONB $(v_1, \ldots, v_{k-1})$ von $\mathbb{R}^{k-1} \times \{0\}$ für b_{k-1} wählen, d.h. es gilt

$$(+) \qquad b(v_i, v_j) = \delta_{ij}$$

für $1 \le i, j \le k-1$. Für ein beliebiges $x \in \mathbb{R}^k \times \{0\}$ betrachten wir das lineare Gleichungssystem

$$(++) \qquad \begin{array}{rcl} b(x, v_1) & = & 0 \\ & \vdots & \\ b(x, v_{k-1}) & = & 0. \end{array}$$

Das ist ein System bestehend aus $k-1$ linearen Gleichungen für die k Unbekannten $(x_1, \ldots, x_k)$. Der Lösungsraum hat dann mindestens die Dimension 1, siehe z.B. (3.28). Es gibt also ein $x \neq 0$ in $\mathbb{R}^k \times \{0\}$ mit $b(x, v_i) = 0$ für $1 \le i \le k-1$. Daraus folgt $b(x, v) = 0$ für alle $v \in \mathbb{R}^{k-1} \times \{0\}$. Wäre $x \in \mathbb{R}^{k-1} \times \{0\}$, so folgte $b(x, x) = 0$, und daraus $x = 0$, da b auf $\mathbb{R}^{k-1} \times \{0\}$ positiv definit ist. Es gilt also $x \notin \mathbb{R}^{k-1} \times \{0\} = \operatorname{span}\{v_1, \ldots, v_{k-1}\}$, so daß wir $v_1, \ldots, v_{k-1}$ durch $v_k := x$ zu einer Basis von $\mathbb{R}^k \times \{0\}$ ergänzen können. Nun ist B_k die Matrix von b_k bezüglich $(e_1, \ldots, e_k)$, während die Matrix von

b_k bezüglich $(v_1, \ldots, v_k := x)$ wegen (+) und (++) die Diagonalmatrix

$$B'_k = \begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ 0 & & & b(v_k, v_k) \end{pmatrix}$$

ist. Wegen $(*)$ folgt aus unserer Voraussetzung $\det(B_k) > 0$, daß auch $\det(B'_k) = b(v_k, v_k) > 0$ gilt. Die Matrix B'_k von b_k bezüglich $(v_1, \ldots, v_k)$ ist demnach positiv definit, und damit ist b_k selbst positiv definit. Das beendet den Induktionsschritt und den Beweis von (12.15).

C) Symmetrische Bilinearformen auf endlichdimensionalen euklidischen Vektorräumen.

In diesem Abschnitt werde wir zeigen, daß symmetrische Bilinearformen auf endlichdimensionalen euklidischen Vekorräumen sogar durch eine Orthonormalbasis diagonalisiert werden können. Das ist eine erhebliche Verschärfung von Satz (12.8), der nur die Existenz irgendeiner Basis verspricht, die die symmetrische Bilinearform diagonalisiert. Andererseits sind solche Orthonormalbasen i.a. auch sehr viel schwieriger zu berechnen. Wir stellen zunächst einen Zusammenhang zwischen symmetrischen Bilinearformen und selbstadjungierten Endomorphismen her.

(12.6) Fakt. Es sei $(V, \langle , \rangle)$ ein endlichdimensionaler euklidischer Vektorraum. Dann existiert zu jeder symmetrischen Bilinearform b auf V genau ein selbstadjungierter Endomorphismus $L = L_b$ von V, so daß für alle $v, w \in V$ gilt:

$$(*) \qquad b(v, w) = \langle L(v), w \rangle$$

Bew.: Wir ordnen jedem $v \in V$ die Linearform $l_v \in V^*$,

$$l_v(w) = b(v, w) \qquad \text{für alle } w \in V,$$

zu. Nach Lemma (9.4) existiert genau ein $v' \in V$, so daß für alle $w \in V$ gilt:

$$l_v(w) = \langle v', w \rangle.$$

Damit $(*)$ gilt, muß also $L(v) = v'$ gesetzt werden, insbesondere ist $L(v)$ durch $(*)$ eindeutig bestimmt. Wir erhalten also eine Abbildung $L : V \to V$, so daß $(*)$ gilt, und müssen nur noch zeigen, daß $L \in \mathrm{End}(V)$ gilt. Die Linearität von L folgt so: Für alle $v_1, v_2, w \in V$ gilt

$$\begin{array}{rcl} b(v_1, w) & = & \langle L(v_1), w\rangle \\ b(v_2, w) & = & \langle L(v_2), w\rangle \end{array}$$

Addition dieser Gleichungen liefert

$$b(v_1 + v_2, w) = \langle L(v_1) + L(v_2), w\rangle.$$

Andererseits gilt nach Definition von L

$$b(v_1 + v_2, w) = \langle L(v_1 + v_2), w\rangle.$$

Da die beiden letzten Gleichungen für alle $w \in V$ gelten, folgt aus der Eindeutigkeitsaussage von Lemma (9.4):

$$L(v_1 + v_2) = L(v_1) + L(v_2).$$

Analog zeigt man $L(\alpha v) = \alpha L(v)$ für alle $\alpha \in \mathbb{R}$, $v \in V$. Etwas eleganter ist folgende alternative Überlegung. Man zeigt, daß die Abbildungen

$$l : V \to V^*, l(v) = l_v$$

und

$$i : V \to V^*, i(v)(w) := \langle v, w\rangle,$$

Homomorphismen von V nach V^* sind. Lemma (9.4) zeigt, daß i sogar ein Isomorphismus ist. Dann gilt

$$L = i^{-1} \circ l,$$

da aus folgender Gleichungskette $i \circ L = l$ folgt. Für alle $v, w \in V$ gilt:

$$((i \circ L)(v))(w) \stackrel{\text{Def. von } i}{=} \langle L(v), w\rangle \stackrel{\text{Def. von } L}{=} b(v, w) \stackrel{\text{Def. von } l}{=} (l(v))(w).$$

Bem.: 1) Ist $\mathcal{G} = (v_1, \ldots, v_n)$ ONB, so gilt

$$(b(v_i, v_j))_{1 \le i,j \le n} = \mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L)$$

2) Die Abbildung, die jeder symmetrischen Bilinearform b auf V den zugehörigen selbstadjungierten Endomorphismus $L \in \mathrm{End}(V)$ zuordnet, ist ein Vektorraumisomorphismus. Beide Vektorräume haben die Dimension $\frac{1}{2}n(n+1)$,

falls $n = \dim V$. Das kann man daran sehen, daß beide isomorph zum Vektorraum der symmetrischen $(n \times n)$-Matrizen sind, der offensichtlich die Dimension $\frac{1}{2}n(n+1)$ hat.

Es folgt das Hauptergebnis dieses Abschnitts, das eine Umformulierung von Satz (10.9) mittels (12.16) ist.

(12.17) Satz. *Sei b symmetrische Bilinearform auf einem euklidischen Vektorraum $(V, \langle , \rangle)$ mit* $\dim V =: n < \infty$. *Dann existiert eine* ONB $\mathcal{G} = (v_1, \ldots, v_n)$ *von $(V, \langle , \rangle)$, die gleichzeitig b-orthogonal ist.*

Bew.: Nach Satz (10.9) existiert eine ONB $\mathcal{G} = (v_1, \ldots, v_n)$ von $(V, \langle , \rangle)$ aus Eigenvektoren von $L = L_b$, d.h. es existieren $\lambda_i \in \mathbb{R}$ für $1 \leq i \leq n$, so daß

$$L(v_i) = \lambda_i v_i$$

gilt. Dann gilt für $1 \leq i \neq j \leq n$:

$$b(v_i, v_j) \overset{\text{Def. von } L}{=} \langle L(v_i), v_j \rangle = \lambda_i \langle v_i, v_j \rangle \overset{\mathcal{G} \text{ ONB}}{=} 0.$$

Bem.: 1) Satz (12.17) ist genau in einem Punkt stärker als Satz (12.8): Statt einer beliebigen b-Orthogonalbasis finden wir sogar eine ONB (bzgl. $\langle , \rangle$), die b-orthogonal ist. Während eine beliebige b-Orthogonalbasis recht leicht explizit berechnet werden kann, etwa durch quadratische Ergänzung, ist die Berechnung einer b-orthogonalen ONB oft nicht explizit möglich, da die Eigenwerte von L benötigt werden.

2) Für jede b-orthogonale ONB $\mathcal{G} = (v_1, \ldots, v_n)$ sind die Diagonalelemente $b(v_i, v_i)$ gerade die Eigenwerte von L, und somit unabhängig von der Wahl von $\mathcal{G}$.

3) Sind alle Eigenwerte von L verschieden, so ist eine b-orthogonale ONB $\mathcal{G} = (v_1, \ldots, v_n)$ nahezu eindeutig: Jedes andere solche $\mathcal{G}'$ ist von der Form $\mathcal{G}' = (\epsilon_1 v_{\sigma(1)}, \ldots, \epsilon_n v_{\sigma(n)})$ für ein $\sigma \in S_n$ und Zahlen $\epsilon_i \in \{\pm 1\}$.

Wir erwähnen noch eine Umformulierung von (12.17) für die zu $\langle , \rangle$ und b gehörigen quadratischen Formen.

(12.17)' Satz (Simultane Diagonalisierung quadratischer Formen). *Sei V n-dimensionaler $\mathbb{R}$-Vektorraum, b und g symmetrische Bilinearformen auf V und g positiv definit. Dann existiert eine Basis $\mathcal{G} = (v_1, \ldots, v_n)$ von V und*

Zahlen $d_i \in \mathbb{R}$ *für* $1 \le i \le n$, *so daß für alle* $v = \sum\limits_{i=1}^{n} x_i v_i \in V$ *gilt:*

$$Q_g(\sum_{i=1}^{n} x_i v_i) = \sum_{i=1}^{n} x_i^2$$

und

$$Q_b(\sum_{i=1}^{n} x_i v_i) = \sum_{i=1}^{n} d_i x_i^2.$$

Bew.: Verwendet man in (12.17) als Skalarprodukt $\langle , \rangle := g$, so folgt (12.17)' unmittelbar aus (12.17).

Die im wesentlichen äquivalenten Aussagen (10.9), (10.10), (12.17) und (12.17)' spielen auch in der Geometrie eine Rolle und sind dort unter dem Namen "Satz über die Hauptachsentransformation" bekannt. Das soll noch etwas erläutert werden: Ist b eine symmetrische Bilinearform auf dem $\mathbb{R}^n$ mit dem Standardskalarprodukt, so nennt man die Menge

$$E_b = \{x \in \mathbb{R}^n \mid b(x,x) = 1\}$$

eine Quadrik in $\mathbb{R}^n$, und speziell, falls b positiv definit ist, ein Ellipsoid (mit Mittelpunkt $0 \in \mathbb{R}^n$). Explizit ausgedrückt ist dabei $b(x,x) = 1$ eine quadratische Gleichung vom Typ $\sum\limits_{i,j=1}^{n} b_{ij} x_i x_j = 1$.

Satz (12.17) liefert nun eine ONB $\mathcal{G} = (v_1, \dots, v_n)$, d.h. ein neues "Kartesisches Koordinatensystem" für $\mathbb{R}^n$, so daß

$$b(\sum_{i=1}^{n} x_i' v_i, \sum_{i=1}^{n} x_i' v_i) = \sum_{i=1}^{n} d_i (x_i')^2$$

gilt. Ist b positiv definit, so nennt man die neuen Koordinatenachsen $\mathbb{R}v_i$, $1 \le i \le n$, Hauptachsen des Ellipsoids E_b. Die Hauptachsen durchstoßen das Ellipsoid in den Punkten

$$E_b \cap \mathbb{R}v_i = \{\pm \frac{1}{\sqrt{d_i}} v_i\},$$

und die Strecken von $0 \in \mathbb{R}^n$ nach $\pm\frac{1}{\sqrt{d_i}} v_i$ nennt man die Hauptachsenabschnitte von E_b.

13 Polynomringe

Das Umgehen mit Polynomen, d.h. mit Ausdrücken der Form

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

ist aus der Schule vertraut, falls die Koeffizienten $a_0, \ldots, a_n$ ganze oder rationale oder reelle Zahlen sind. In diesen Fällen kann man das Polynom mit einer Abbildung

$$\tilde{p} : \mathbb{R} \to \mathbb{R},\ \tilde{p}(x) := \sum_{i=0}^{n} a_ix^i$$

identifizieren. Es gilt dann $\tilde{p}(0) = a_0, (\tilde{p})'(0) = a_1, (\tilde{p})''(0) = 2a_2$ und die k'te Ableitung von $\tilde{p}$ ausgewertet an der Stelle $x = 0$ ist gerade $k!a_k$, d.h. die Funktion $\tilde{p}$ bestimmt die $a_0, \ldots, a_n$ eindeutig. Wenn man nun Polynome mit Koeffizienten a_i aus einem beliebigen Körper K betrachten will, entsteht im Fall von endlichen Körpern folgende Schwierigkeit: es „existieren" unendlich viele verschiedene Polynome, z.B. die „Monome" $1, x, x^2, \ldots, x^n, \ldots,$ aber nur endlich viele verschiedene Abbildung $\tilde{p} : K \to K$. Ein Polynom $\sum\limits_{i=0}^{n} a_ix^i$ mit $a_i \in K$ kann also nicht durch die zugehörige Abbildung $\tilde{p} : K \to K$, $\tilde{p}(x) := \sum\limits_{i=0}^{n} a_ix^i$, eindeutig bestimmt sein. Diese Tatsache ruft die Frage hervor, was denn dann ein Polynom mit Koeffizienten $a_i \in K$ wirklich „ist" (oder besser „sein soll"), eine Frage, die wir eigentlich schon bei der Einführung des charakteristischen Polynoms (nach (7.28)) hätten stellen sollen. Das Ziel dieses Kapitels ist eine befriedigende Antwort auf diese Frage. Dann werden wir noch die Polynomdivision (mit Rest) kennenlernen und uns mit der eindeutigen Zerlegung von Polynomen in Primfaktoren beschäftigen.

Zu einem beliebigen Körper K betrachten wir die Menge

$$K^{\mathbb{N}} := \{f \mid f : \mathbb{N} \to K\},$$

d.h. ein Element $f \in K^{\mathbb{N}}$ ist eine Folge $f = (f_0, f_1, \ldots, f_n, \ldots)$ mit $f_n \in K$ für alle $n \in \mathbb{N}$.

(13.1) Def.: Zu $f, g \in K^{\mathbb{N}}$, $a \in K$ definieren wir

(i) $f + g \in K^{\mathbb{N}}$ durch $f + g := (f_0 + g_0, f_1 + g_1, \ldots, f_n + g_n, \ldots)$

(ii) $af \in K^{\mathbb{N}}$ durch $af := (af_0, af_1, \ldots, af_n, \ldots)$

(iii) $f \cdot g \in K^{\mathbb{N}}$ durch $f \cdot g := ((f \cdot g)_0, (f \cdot g)_1, \ldots, (f \cdot g)_n, \ldots)$,
wobei $(f \cdot g)_0 := f_0 \cdot g_0, (f \cdot g)_1 := f_0 \cdot g_1 + f_1 \cdot g_0, (f \cdot g)_2 := f_0 \cdot g_2 + f_1 \cdot g_1 + f_2 \cdot g_0$
und $(f \cdot g)_n := \sum\limits_{j=0}^{n} f_j \cdot g_{n-j} = \sum\limits_{\substack{(j,k) \in \mathbb{N} \times \mathbb{N} \\ j+k=n}} f_j \cdot g_k$.

Bem.: 1) Die Ausdrücke in (iii) sind den Ausdrücken nachgebildet, die beim „schulmäßigen“ Multiplizieren von Polynomen auftreten:

$$(a_0+a_1x+\ldots+a_nx^n)\cdot(b_0+b_1x+\ldots+b_mx^m) = a_0b_0+(a_0b_1+a_1b_0)x+(a_0b_2+a_1b_1+a_2b_0)x^2+\ldots$$

2) Für $a \in K$, $f \in K^{\mathbb{N}}$ gilt $af = (a, 0, \ldots, 0, \ldots) \cdot f$.

Ein Ring $(R, +, \cdot)$ heißt <u>nullteilerfrei</u>, falls für alle $a, b \in R$ gilt:

$$ab = 0 \Rightarrow a = 0 \text{ oder } b = 0.$$

(13.2) Fakt.

(a) $(K^{\mathbb{N}}, +)$ ist ein ∞-dimensionaler K-Vektorraum mit dem Nullelement $0 = (0, \ldots, 0, \ldots) \in K^{\mathbb{N}}$.

(b) $(K^{\mathbb{N}}, +, \cdot)$ ist ein nullteilerfreier kommutativer Ring mit $1 = (1, 0, \ldots, 0, \ldots) \in K^{\mathbb{N}}$.

Bem.: Die Definition des „kommutativen Rings mit 1“ unterscheidet sich von der des Körpers (vgl. (2.4)) nur dadurch, daß auf die Forderung der Existenz von multiplikativen Inversen verzichtet wird. (Man fordert üblicherweise auch <u>nicht</u>, daß $1 \neq 0$ gilt, aber das ist für $K^{\mathbb{N}}$ natürlich erfüllt).

Bez.: $(K^{\mathbb{N}}, +, \cdot)$ heißt der Ring der formalen Potenzreihen über K, meist bezeichnet durch $K[[x]]$.

Der Nachweis der in (13.2) behaupteten Eigenschaften ist leicht. Am längsten dauert der Nachweis der Assoziativität der Multiplikation, den man wie folgt durchführen kann:

Seien $f, g, h \in K^{\mathbb{N}}$ und $n \in \mathbb{N}$. Dann gilt:

$$\begin{aligned}((f\cdot g)\cdot h)_n &= \sum_{j+k=n}(f\cdot g)_j h_k = \sum_{j+k=n}\left(\sum_{l+m=j} f_l\cdot g_m\right)\cdot h_k = \sum_{l+m+k=n} f_l\cdot g_m\cdot h_k \\ (f\cdot(g\cdot h))_n &= \sum_{l+j=n} f_l\cdot(g\cdot h)_j = \sum_{l+j=n} f_l\cdot\left(\sum_{m+k=j} g_m\cdot h_k\right) = \sum_{l+m+k=n} f_l\cdot g_m\cdot h_k\end{aligned}$$

(13.3) Fakt. Die Teilmenge

$$K[x] = \{f \in K^{\mathbb{N}} \mid f_n \neq 0 \text{ nur für endlich viele } n \in \mathbb{N}\}$$

ist abgeschlossen bezüglich $+$ und $\cdot$, enthält $1 = (1, 0, \ldots, 0, \ldots)$, und ist ein Unterring (mit 1) von $(K^{\mathbb{N}}, +, \cdot)$.

Die Abgeschlossenheit bzgl. $\cdot$ folgt aus dem Beweis zu (13.5).

(13.4) Def.:

(a) $(K[x], +, \cdot)$ heißt der Polynomring von K.

(b) Ist $f \in K[x] \setminus \{0\}$, so heißt

$$\operatorname{grad} f := \max\{i \in \mathbb{N} \mid f_i \neq 0\} \in \mathbb{N}$$

der Grad von f. Ist $f = 0 \in K[x]$, so setzen wir $\operatorname{grad} f = -\infty$.

(13.5) Fakt. Für alle $f, g \in K[x]$ gilt:

$$\operatorname{grad}(f\cdot g) = \operatorname{grad} f + \operatorname{grad} g.$$

Bew.: Ist $f = 0$ oder $g = 0$, so $f \cdot g = 0$, also

$$\operatorname{grad}(f\cdot g) = -\infty = \operatorname{grad} f + \operatorname{grad} g.$$

Ist $\operatorname{grad} f = m \in \mathbb{N}$, $\operatorname{grad} g = n \in \mathbb{N}$, so gilt $f_m \neq 0$, $g_n \neq 0$ und $f_i = 0$ für $i > m$, $g_j = 0$ für $j > n$. Daraus folgt

$$(f\cdot g)_{m+n} = \sum_{\substack{(i,j)\in\mathbb{N}\times\mathbb{N} \\ i+j=m+n}} f_i g_j = f_m \cdot g_n \neq 0$$

und $(f \cdot g)_k = 0$ für $k > m + n$. Also

$$\operatorname{grad}(f \cdot g) = m + n = \operatorname{grad} f + \operatorname{grad} g.$$

Um zur üblichen Darstellung von Polynomen zu kommen, führen wir folgende Bezeichnung ein:

$$x := (0, 1, 0, \ldots, 0, \ldots) \in K[x].$$

Außerdem definieren wir wie üblich $x^0 := (1, 0, \ldots, 0, \ldots)$ und rekursiv $x^{n+1} := x \cdot x^n$ für $n \in \mathbb{N}$.

In diesem Zugang zu den Polynomen ist also x keine „Variable“ oder „Unbekannte“, sondern ein festes Element des Rings $K[x]$.

(13.6) Fakt. Die Menge $\{x^n \mid n \in \mathbb{N}\}$ ist eine Basis des K-Vektorraums $K[x]$. Es gilt $x^n = (\underbrace{0, \ldots, 0}_{n\text{-mal}}, 1, 0, \ldots) \in K[x]$.

Bew.: Die erste Behauptung folgt aus der zweiten. Die zweite beweisen wir durch Induktion. Die Gleichung $x^0 = (1, 0, \ldots, 0, \ldots)$, die nach Definition von x^0 gilt, ist der Induktionsanfang. Mit der Bezeichnung $\delta_{ij} := 0$ für $i \neq j$ und $\delta_{ii} := 1$ können wir $x = (\delta_{01}, \delta_{11}, \delta_{21}, \ldots)$ und - nach Induktionsvoraussetzung - $x^n = (\delta_{0n}, \delta_{1n}, \ldots, \delta_{nn}, \ldots)$ schreiben. Also gilt für jedes $i \in \mathbb{N}$:

$$(x^{n+1})_i = (x \cdot x^n)_i = \sum_{j+k=i} \delta_{j1}\delta_{kn} = \begin{cases} 1 \text{ falls } i = n+1 \\ 0 \text{ falls } i \neq n+1. \end{cases}$$

(13.7) Folgerung. Jedes Polynom $f \in K[x] \setminus \{0\}$ von Grad $\operatorname{grad} f = m$ besitzt genau eine Darstellung

$$f = f_0 + f_1 x + f_2 x^2 + \ldots + f_m x^m \text{ mit } f_m \neq 0.$$

Bem.: Bezüglich dieser Darstellung gehen die in (13.1) definierten Operationen in die für Polynome „üblichen“ über.

(13.8) Satz (Division mit Rest). *Seien* $f \in K[x]$, $g \in K[x]$, *und es gelte* $\operatorname{grad} f \geq \operatorname{grad} g \geq 0$. *Dann existieren* $h, r \in K[x]$, *so daß*

$$f = gh + r$$

und

$$\operatorname{grad} r < \operatorname{grad} g \qquad (\textit{möglicherweise } r = 0)$$

gelten.

Bew.: Durch Induktion nach $n := \operatorname{grad} f$.

Induktionsanfang: Gilt $\operatorname{grad} f = 0$, so auch $\operatorname{grad} g = 0$, also $f = f_0 \neq 0$, $g = g_0 \neq 0$, und wir können als $h := \frac{f_0}{g_0}$ und $r = 0$ nehmen. Für den Induktionsschritt betrachten wir $f = \sum\limits_{i=0}^{n} f_i x^i$, $g = \sum\limits_{j=0}^{m} g_j x^j$ mit $0 \leq m \leq n$, $f_n \neq 0, g_m \neq 0$. Wir definieren

$$r_1 := f - \frac{f_n}{g_m} x^{n-m} \cdot g.$$

Dann gilt:

$$(*) \qquad f = g \cdot \left(\frac{f_n}{g_m} x^{n-m} \right) + r_1 \text{ und } \operatorname{grad} r_1 < n.$$

1. Fall: $\operatorname{grad} r_1 < \operatorname{grad} g$. Dann erhalten wir die Behauptung, indem wir $h := \frac{f_n}{g_m} x^{n-m}$ und $r := r_1$ setzen.

2. Fall: $\operatorname{grad} r_1 \geq \operatorname{grad} g$. Wegen $n > \operatorname{grad} r_1 \geq \operatorname{grad} g \geq 0$ ist auf r_1 die Induktionsvoraussetzung anwendbar und wir erhalten $h_1, r \in K[x]$, so daß $r_1 = gh_1 + r$ und $\operatorname{grad} r < \operatorname{grad} g$ gelten.

Dann folgt mit $(*)$:

$$f = g \cdot \left(\frac{f_n}{g_m} x^{n-m} + h_1 \right) + r.$$

Setzen wir $h := \frac{f_n}{g_m} x^{n-m} + h_1$, so erhalten wir wegen $\operatorname{grad} r < \operatorname{grad} g$ die Behauptung.

Bem.: Der Beweis von (13.8) besteht in dem üblichen Rechenverfahren zur Polynomdivision, das in die Form eines Beweises gebracht wurde. Umgekehrt liefert der Beweis auch dieses Rechenverfahren, das man beherrschen muß.

Einem Polynom $p = \sum\limits_{i=0}^{n} a_i x^i \in K[x]$ ordnen wir die Abbildung

$$\tilde{p} : K \to K,\ \tilde{p}(b) := \sum_{i=0}^{n} a_i b^i \text{ zu.}$$

Wie zu Beginn des Kapitels begründet, ist die Abbildung

$$p \in K[x] \to \tilde{p} \in \mathrm{Abb}(K, K)$$

nicht injektiv, wenn K ein endlicher Körper ist. Hierbei bezeichnet $\mathrm{Abb}(K, K)$ die Menge der Abbildungen von K in sich, die mit punktweiser Addition und Multiplikation einen Ring bildet. Bezüglich dieser Struktur ist die Abbildung $p \in K[x] \to \tilde{p} \in \mathrm{Abb}(K, K)$ ein Ringhomomorphismus, d.h. es gilt für alle $p, q \in K[x]$

$$\begin{array}{rcl} (p+q)^\sim & = & \tilde{p} + \tilde{q} \\ (p \cdot q)^\sim & = & \tilde{p} \cdot \tilde{q} \end{array}$$

und außerdem für alle $a \in K$

$$(ap)^\sim = a\tilde{p}.$$

In Zukunft werden wir, wie das allgemein üblich ist, statt $\tilde{p}(b)$ einfach $p(b)$ schreiben.

Im nächsten Kapitel wird es eine wichtige Rolle spielen, daß man in Polynome $p \in K[x]$ nicht nur Körperelemente $b \in K$, sondern auch Matrizen $A \in K^{m \times m}$ für beliebiges $m > 0$ und Endomorphismen L eines K-Vektorraums V "einsetzen" kann. Wir definieren für $p = \sum\limits_{i=0}^{n} a_i x^i \in K[x]$ und $A \in K^{m \times m}$

$$p(A) := \sum_{i=0}^{n} a_i A^i,$$

und für $L \in \mathrm{End}(V)$

$$p(L) := \sum_{i=0}^{n} a_i L^i,$$

wobei $L^0 := \mathrm{id}_V$ und $L^{i+1} := L \circ L^i$ für alle $i \in \mathbb{N}$. Auch hier gilt wieder für alle $a \in K$, p, $q \in K[x]$ und $A \in K^{m \times m}$:

$$\begin{array}{rcl} (p+q)(A) & = & p(A) + q(A) \\ (p \cdot q)(A) & = & p(A)q(A) \\ (ap)(A) & = & ap(A), \end{array}$$

und Entsprechendes für $p(L)$ mit $L \in \mathrm{End}(V)$. Exemplarisch beweisen wir die Gleichung

$$(p \cdot q)(A) = p(A)q(A).$$

Ist $p = \sum\limits_{i=0}^{n} a_i x^i$, $q = \sum\limits_{j=0}^{k} b_j x^j$, so gilt $p \cdot q = \sum\limits_{l=0}^{n+k} \left(\sum\limits_{i+j=l} a_i b_j \right) x^l$, und damit

$$(p \cdot q)(A) = \sum_{l=0}^{n+k} \left(\sum_{i+j=l} a_i b_j \right) A^l.$$

Andererseits berechnen wir in $K^{m \times m}$:

$$p(A)q(A) = \left(\sum_{i=0}^{n} a_i A^i \right) \left(\sum_{j=0}^{k} b_j A^j \right) = \sum_{i=0}^{n} \sum_{j=0}^{k} a_i b_j A^{i+j} = \sum_{l=0}^{n+k} \left(\sum_{i+j=l} a_i b_j \right) A^l.$$

(13.9) Def.: Ein $a \in K$ heißt <u>Nullstelle von</u> $p \in K[x]$, wenn $p(a) = 0$ gilt.

Es ist Teil des üblichen Schulstoffs, daß jede natürliche Zahl eine eindeutige Zerlegung als Produkt von Primzahlen besitzt. Unser nächstes Ziel ist, eine analoge Aussage im Ring $K[x]$ zu bewiesen. Wir werden uns auf den Fall des Rings $K[x]$ beschränken, obwohl es in der Zahlentheorie und Algebra allgemeinere (und natürlichere) Formulierungen dieses Satzes gibt.

(13.10) Def.: Es sei R ein nullteilerfreier, kommutativer Ring mit $1 (\neq 0)$.

(i) Seien $a, b \in R \setminus \{0\}$. Wir sagen "$a$ <u>teilt</u> b" (kurz: $a|b$), falls ein $c \in R$ existiert, so daß $ac = b$ gilt.

(ii) Ein $a \in R$ heißt <u>Einheit</u>, falls a ein multiplikatives Inverses besitzt, d.h. falls ein $b \in R$ existiert, so daß $ab = 1$ gilt.

(iii) Ein $a \in R \setminus \{0\}$ heißt <u>prim</u>, falls a keine Einheit ist und falls für jeden Teiler b von a gilt: b ist Einheit oder es existiert eine Einheit $c \in R$, so daß $b = ac$ gilt.

In (5.2) hatten wir schon eingesehen, daß die Menge

$$R^* = \{a \in R \mid a \text{ Einheit } \}$$

mit der Verknüpfung $\cdot$ eine Gruppe ist. Wir setzen hier voraus, daß R kommutativ ist, und daraus folgt, daß $(R^*, \cdot)$ abelsch ist. Es ist leicht zu sehen, daß jede Einheit $c \in R$ jedes $a \in R \setminus \{0\}$ teilt und daß für jedes $a \in R \setminus \{0\}$ und jede Einheit $c \in R$ das Produkt ac ein Teiler von a ist. Die Bedingung "a ist prim"besagt also gerade, daß a nur diese offensichtlichen Teiler besitzt. Insbesondere sehen wir, daß für natürliche Zahlen die Definition (13.10)(iii) (für den Fall $R = \mathbb{Z}$) mit der üblichen Definition von "Primzahl" übereinstimmt. In Algebra und Zahlentheorie heißt die Bedingung (13.10)(iii) allerdings üblicherweise "irreduzibel".

Bem. 1) Sei $a \in R \setminus \{0\}$ keine Einheit. Dann ist a genau dann nicht prim, wenn es $b,\, c \in R \setminus R^*$ gibt, so daß $a = bc$ gilt.
2) Im Fall $R = \mathbb{Z}$ gilt $R^* = \{+1, -1\} \subseteq \mathbb{Z}$.
3) Die Einheiten im Ring $K[x]$ sind genau die Polynome vom Grad 0, d.h. $K[x]^* = K \setminus \{0\}$.
4) Jedes $p \in K[x]$ mit $\operatorname{grad} p = 1$ ist prim.

Hier die kurze Begründung für Bemerkung 3):

Ist $p \in K[x]$ Einheit, so existiert $q \in K[x]$ mit $p \cdot q = 1$. Mit (13.5) folgt daraus
$$0 = \operatorname{grad}(p \cdot q) = \operatorname{grad} p + \operatorname{grad} q,$$
also $\operatorname{grad} p = \operatorname{grad} q = 0$. Gilt umgekehrt $\operatorname{grad} p = 0$, so ist $p = a_0 \in K\setminus\{0\} \subseteq K[x]$ und besitzt das multiplikative Inverse $q = a_0^{-1}$.

(13.11) Lemma. Sei $p \in K[x] \setminus \{0\}$. Ein $a \in K$ ist genau dann Nullstelle von p, wenn $(x - a)$ Teiler von p ist, d.h. wenn ein $h \in K[x]$ existiert, so daß
$$p = (x - a)h$$
gilt.

Bew.: (i) Sei $p(a) = 0$. Wegen $p \neq 0$ gilt dann $\operatorname{grad} p \geq 1$. Division mit Rest von p durch das Polynom $x - a$ liefert $h \in K[x]$ und $r \in K[x]$ mit $\operatorname{grad} r < 1$, so daß
$$p = (x - a)h + r$$
gilt. Daraus folgt $0 = p(a) = r(a)$ und daraus, wegen $\operatorname{grad} r = 1$, daß $r = 0$ gilt.
(ii) Gilt $p = (x - a)h$, so folgt $p(a) = (a - a)h(a) = 0 \cdot h(a) = 0$. Genau genommen verwenden wir hier die Gleichung $(p \cdot q)^\sim = \tilde{p} \cdot \tilde{q}$:
$$\tilde{p}(a) = (x - a)^\sim(a)\tilde{h}(a) = (a - a)\tilde{h}(a) = 0.$$

Bsp.: Das Polynom $p = x^2 + 1$, aufgefaßt als Element von $\mathbb{R}[x]$, ist prim, während $p = x^2 + 1$ als Element von $\mathbb{C}[x]$ nicht prim ist: $p = (x + i)(x - i)$.

(13.12) Fakt. Sei $R = \mathbb{Z}$ oder $R = K[x]$ für einen Körper K. Dann besitzt jedes $a \in R\backslash(R^* \cup \{0\})$ eine Zerlegung in Primfaktoren, d.h. es existieren $m \in \mathbb{N}_{>0}$ und prime Elemente $p_1, \ldots, p_m$ in R (die nicht notwendig verschieden sind), so daß

$$a = p_1 \cdot \ldots \cdot p_m = \prod_{i=1}^{m} p_i$$

gilt.

Bew.: Im Fall $R = K[x]$: Durch Induktion nach $n := \operatorname{grad} p \geq 1$.
Induktionsanfang: Ist $\operatorname{grad} a = 1$, so ist a prim, siehe Bem. 4).
Induktionsschritt: Ist $\operatorname{grad} a = n > 1$ und ist a nicht prim, so existieren nach Bem. 1) und 3) Polynome a_1, a_2 mit $a = a_1 \cdot a_2$ und $\operatorname{grad} a_1 > 0$, $\operatorname{grad} a_2 > 0$. Daraus folgt $\operatorname{grad} a_1 < n$, $\operatorname{grad} a_2 < n$, vgl. (13.5). Nach Induktionsvoraussetzung erhalten wir Primfaktorzerlegungen von a_1 und a_2. Deren Produkt ist eine Primfaktorzerlegung von a. Ein analoger Beweis kann für den Fall $R = \mathbb{Z}$ geführt werden. Man ersetzt dabei den Grad eines Polynoms durch den Betrag einer ganzen Zahl.

Die Eindeutigkeit von Primfaktorzerlegungen ist schwieriger zu beweisen als deren Existenz. Wir wenden uns jetzt diesem Eindeutigkeitsproblem zu.

(13.13) Def.: Zwei Elemente $a, b \in R \setminus \{0\}$ heißen teilerfremd, falls jedes $d \in R \setminus \{0\}$, das sowohl a als auch b teilt, eine Einheit ist.

(13.14) Satz. Sei $R = \mathbb{Z}$ oder $R = K[x]$. Sind a, $b \in R \setminus \{0\}$ teilerfremd, so existieren r, $s \in R$, so daß

$$ra + sb = 1$$

gilt.

Bew.: Im Fall $R = K[x]$: Wir zeigen, daß die Menge

$$I := \{ga + hb | g, h \in R\}$$

eine Einheit, und damit auch das Element $1 \in R$, enthält. I hat offensichtlich folgende Eigenschaften:

$$\begin{array}{rcl} p, q \in I & \Rightarrow & p + q \in I \\ d \in R, p \in I & \Rightarrow & dp \in I \end{array}$$

(Allgemein heißt eine Teilmenge $I \neq \emptyset$ von R mit diesen beiden Eigenschaften ein Ideal in R). Wir können ein Element $k \in I \setminus \{0\}$ minimalen Grades in

$I \setminus \{0\}$ wählen, d.h. für alle $\tilde{k} \in I \setminus \{0\}$ gilt $\operatorname{grad} k \leq \operatorname{grad} \tilde{k}$. Wir zeigen, daß k Teiler von a und b ist. Da $a \in I \setminus \{0\}$ ist, gilt $\operatorname{grad} k \leq \operatorname{grad} a$, so daß nach (13.8) Polynome h, $r \in K[x]$ mit $\operatorname{grad} r < \operatorname{grad} k$ existieren mit

$$a = kh + r.$$

Wegen $a \in I$ und $k \in I$ folgt $r = a - kh \in I$, und $\operatorname{grad} r < \operatorname{grad} k$ impliziert nun, daß $r = 0$ gilt, d.h. $k|a$. Ebenso folgt $k|b$. Da a und b teilerfremd sind, ist $k \in I$ eine Einheit. Der Beweis im Fall $K = \mathbb{Z}$ ist analog, wobei wieder statt des Grades von Polynomen der Betrag von ganzen Zahlen verwendet wird. Die zu (13.8) analoge "Teilbarkeit mit Rest" von ganzen Zahlen ist Stoff der Grundschule.

Folgende Aussage ist der Kern der Eindeutigkeit der Primfaktorzerlegung.

(13.15) Folgerung. Sei $R = \mathbb{Z}$ oder $R = K[x]$. Ist $p \in R$ prim und teilt p ein Produkt $ab \in R \setminus \{0\}$, so teilt p mindestens einen der Faktoren a, b.

Bem.: Für den Fall $R = \mathbb{Z}$ wurde diese Aussage schon im Beweis von Satz (2.11) verwendet.

Bew.: Wir setzen voraus, daß p Teiler von ab und nicht Teiler von a ist, und zeigen, daß p dann b teilt. Da p prim ist und a nicht teilt, sind p und a teilerfremd. Nch (13.14) gibt es r, $s \in R$, so daß

$$ra + sp = 1$$

gilt. Wir multiplizieren diese Gleichung mit b und erhalten

$$rab + spb = b.$$

Da p Teiler von ab ist, existiert ein $c \in R$ mit $pc = ab$. Wir setzen dies in die vorausgehende Gleichung ein und klammern p aus:

$$p(rc + sb) = b.$$

Das zeigt $p|b$.

(13.16) Satz. *Sei $R = \mathbb{Z}$ oder $R = K[x]$ für einen Körper K. Die nach (13.12) existierende Primfaktorzerlegung ist in folgendem Sinn eindeutig: Ist $a \in R \setminus (R^* \cup \{0\})$ und sind*

$$a = \prod_{i=1}^{n} p_i = \prod_{j=1}^{n'} p'_j$$

Primfaktorzerlegungen von a, so gilt $n = n'$ und es existieren $\sigma \in S_n$ und Einheiten $c_i \in R$, so daß für alle $i \in \{1, \ldots, n\}$ gilt

$$p'_{\sigma(i)} = c_i p_i.$$

Unwichtige Bem.: Es gilt dann $\prod_{i=1}^{n} c_i = 1$.

Bew.: Durch Induktion nach der Minimalzahl $n(a)$ von Primfaktoren, die zur Darstellung von a benötigt werden.

Induktionsanfang: Ist $n(a) = 1$, so ist a prim. Nach Definition von "prim" ist dann $a = a$ die einzige Möglichkeit für eine Primfaktorzerlegung von a.

Induktionsschritt: Es sei $n(a) =: n > 1$. Es genügt, die Behauptung in dem Fall zu beweisen, daß die erste Darstellung von a die minimale Anzahl n von Faktoren hat:

$$a = \prod_{i=1}^{n} p_i = \prod_{j=1}^{n'} p'_j \text{ mit } n = n(a) \leq n'.$$

Dann gilt $p_n | a$, d.h. $p_n \Big| \prod_{j=1}^{n'} p'_j$. Induktiv folgt aus (3.15), daß p_n einen der Faktoren $p'_1, \ldots, p'_{n'}$ teilt, o.E. $p_n | p'_{n'}$. Da $p'_{n'}$ prim ist, existiert eine Einheit $c_n \in R$, so daß $p'_{n'} = c_n p_n$ gilt. Dann gilt

$$a = p_n \prod_{i=1}^{n-1} p_i = p_n c_n \prod_{j=1}^{n'-1} p'_j$$

und damit

$$p_n \left(\prod_{i=1}^{n-1} p_i - c_n \prod_{j=1}^{n'-1} p'_j \right) = 0.$$

Da R nullteilerfrei ist, folgt

$$\tilde{a} := \prod_{i=1}^{n-1} p_i = c_n \prod_{j=1}^{n'-1} p'_j,$$

insbesondere $n(\tilde{a}) \leq n - 1 < n(a)$. Wir können nun die Induktionsvoraussetzung auf $\tilde{a}$ anwenden und erhalten $n - 1 = n' - 1$, also $n = n'$ wie behauptet. Außerdem stimmen nach Induktionsvoraussetzung die $p_i, 1 \leq i \leq n - 1$, bis auf Anordnung und Multiplikation mit Einheiten mit den $p'_j, 1 \leq j \leq n - 1$, überein. Zusammen mit $p'_n = c_n p_n$ ergibt das die Behauptung.

Bez.: Ein Polynom $p \in K[x]$ zerfällt in Linearfaktoren, falls jeder Primfaktor von p Grad 1 hat, d.h. falls $a \in K \setminus \{0\}$ und $a_1, \ldots, a_n \in K$ mit $n = \operatorname{grad} p$ existieren, so daß

$$p = a \prod_{i=1}^{n} (x - a_i).$$

Oft werden wir statt "zerfällt in Linearfaktoren " einfach "zerfällt" sagen. Manchmal werde wir den Körper K nochmals erwähnen, etwa: $p(x) = x^2 + 1$ zerfällt über $\mathbb{C}$, aber nicht über $\mathbb{R}$.

Bem.: Zusammen mit (13.11) zeigt der Fundamentalsatz der Algebra (2.10), daß ein $p \in \mathbb{C}[x]$ genau dann prim ist, wenn $\operatorname{grad} p = 1$ ist. Also zerfällt jedes Polynom über $\mathbb{C}$.

Die Eindeutigkeit der Primfaktorzerlegung in $K[x]$ hat folgende Konsequenz, die im nächsten Kapitel verwendet werden wird.

(13.17) Folgerung. Sind $f, g, h \in K[x] \setminus \{0\}$, gilt $f = gh$ und zerfällt f, so zerfallen auch g und h.

Bew.: Das Produkt von Primfaktorenzerlegungen von g und h ist eine Primfaktorzerlegung von f. Andererseits läßt sich f nach Voraussetzung als Produkt von Primfaktoren von Grad 1 schreiben. Da die Primfaktorzerlegung im Sinn von (13.16) eindeutig ist, besitzen g und h ebenfalls nur Primfaktoren von Grad 1, d.h. g und h zerfallen.

Zum Abschluß dieses Kapitels greifen wir die Frage auf, was denn nun das charakteristische Polynom einer Matrix $A \in K^{n \times n}$ sein soll. Im Fall eines endlichen Körpers können wir ja "das Polynom $\det(A - xE_n)$" nicht einfach als Abbildung ansehen, die $x \in K$ auf $\det(A - xE_n) \in K$ abbildet.

Die einfachste Antwort auf diese Frage ist, daß wir auch für Matrizen $B = (b_{ij})_{1 \leq i,j \leq n}$, deren Koeffizienten b_{ij} in einem kommutativen Ring R mit 1 liegen, mit Hilfe der Leibnizformel (7.14) die Determinante von B durch

$$\det B = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} \cdot \ldots \cdot b_{n\sigma(n)} \in R$$

definieren können. Ist nun K ein Körper und $A \in K^{n \times n}$, so können wir $A - xE_n$ als Matrix mit Koeffizienten $a_{ij} - x\delta_{ij}$ im Ring $R = K[x]$ ansehen, und damit ist $\det(A - xE_n) \in K[x]$ definiert. Es ist leicht zu sehen, daß das Polynom $p_A(x) := \det(A - xE_n) \in K[x]$ den Grad n hat. Nun würden wir gern viele Rechenregeln, insbesondere den Determinantenproduktsatz (7.25), auch für quadratische Matrizen über einem Ring R verwenden, aber bewiesen haben wir sie nur für den Fall, daß R ein Körper ist. Hier ist Blatt 9, Aufgabe 3 (korrigiert durch die Forderung, daß statt $R \times R$ nur $R \times (R \setminus \{0\})$ betrachtet wird) hilfreich. In dieser Aufgabe wird gezeigt, daß ein nullteilerfreier, kommutativer Ring R mit $1 (\neq 0)$ zu einem Körper K erweitert werden kann, in ähnlicher Weise wie man $\mathbb{Z}$ zu $\mathbb{Q}$ erweitert. Im Fall $R = K[x]$ benutzt man für diesen "Quotientenkörper" von $K[x]$ das Symbol

$$K(x) = \left\{ \frac{p}{q} \mid p \in K[x], q \in K[x] \setminus \{0\} \right\}.$$

Unsere in Kapitel 7 entwickelte Determinantenrechnung ist für Matrizen mit Koeffizienten in beliebigen Körpern entwickelt, und also auch für den Fall des Körpers $K(x)$ gültig. Das zeigt, daß wir mit Determinaten der Form $\det(A - xE_n) \in K[x]$, für $A \in K^{n\times n}$, umgehen können, wie wir das gewohnt sind. Insbesondere gilt für alle $B \in \mathrm{GL}_n(K)$:

$$(*) \quad \begin{aligned} \det(B^{-1}AB - xE_n) &= \det(B^{-1}(A - xE_n)B) \\ &= \det B^{-1} \cdot \det(A - xE_n) \cdot \det B = \det(A - xE_n). \end{aligned}$$

Das zeigt, daß wir für $L \in \mathrm{End}(V)$, V n-dimensionaler K-Vektorraum, das charakteristische Polynom $p_L \in K[x]$ von L wie folgt definieren können: Wir wählen eine Basis $\mathcal{G}$ und setzen

$$p_L = \det(\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L) - xE_n).$$

Nach (5.6) und $(*)$ ist diese Definition unabhängig von der gewählten Basis.

14 Die Jordansche Normalform

Etwas verkürzt ausgedrückt geht es in diesem Kapitel darum, zu einem $L \in \mathrm{End}(V)$, $\dim V < \infty$, eine Basis $\mathcal{G}$ zu finden, für die $\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L)$ eine möglichst einfache Form hat. Der Name "Jordansche Normalform" geht auf den franz. Mathematiker Camille Jordan (1838-1922) zurück.

(14.1) Def.: (a) $A \in K^{n\times n}$ heisst <u>diagonalisierbar</u>, falls ein $B \in GL(n, K)$ existiert, so dass $B^{-1}AB$ eine Diagonalmatrix ist.

(b) $L \in \mathrm{End}(V)$ heisst diagonalisierbar, falls eine Basis von V existiert, die aus Eigenvektoren von L besteht.

Bem.: 1) Ist $\mathcal{G} = (v_1, \ldots, v_n)$ Basis aus Eigenvektoren von L, so gilt

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

wobei λ_i der Eigenwert von L zum Eigenvektor v_i ist.
2) Ist $\dim V < \infty$, $L \in \mathrm{End}(V)$ und $\mathcal{G}$ eine beliebige Basis von V, so gilt: L ist genau dann diagonalisierbar, wenn $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)$ diagonalisierbar ist. Das folgt aus (5.6).

Bsp.: 1) Ist $A \in \mathbb{R}^{n\times n}$ symmetrisch, d.h. $A = A^T$, so ist A diagonalisierbar, vgl. Folgerung (10.10).
2) Ist V endlichdimensionaler euklidischer Vektorraum und ist $L \in \mathrm{End}(V)$ selbstadjungiert, so ist L diagonalisierbar, vgl. Satz (10.9). Es gibt dann sogar eine ONB von V, so dass $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L)$ Diagonalmatrix ist.
3) Satz (7.29) besagt: Ist $\dim V = n$, $L \in \mathrm{End}(V)$ und besitzt L n verschiedene Eigenwerte, so ist L diagonalisierbar.

(14.2) Def.: Sei $L \in \mathrm{End}(V)$ und $\lambda \in K$ Eigenwert von L.

(a) $E(\lambda) := \ker\big(L - \lambda \mathrm{id}_V\big) = \big\{v \in V \mid L(v) = \lambda v\big\}$ heisst der Eigenraum von L zum Eigenwert λ.

(b) $E'(\lambda) := \bigcup_{k\in\mathbb{N}_{>0}} \ker\big((L - \lambda \mathrm{id}_V)^k\big) = \big\{v \in V \mid \exists k \in \mathbb{N}_{>0} : (L - \lambda \mathrm{id}_V)^k(v) = 0\big\}$ heisst der Hauptraum von L zum Eigenwert λ.

Bem.: 1) $E(\lambda)$ ist Untervektorraum von V und, da λ EW von L ist, gilt $\dim E(\lambda) \geq 1$. $E(\lambda) \setminus \{0\}$ ist gerade die Menge der Eigenvektoren von L zum EW λ.
2) $E(\lambda) \subseteq E'(\lambda)$, und es kann $E(\lambda) \neq E'(\lambda)$ gelten: Ist $L \in \mathrm{End}(\mathbb{R}^2)$ die durch $L(e_1) = e_1$, $L(e_2) = e_1 + e_2$ definierte "Scherung", so gilt $E(1) = \mathrm{span}\{e_1\}$, $E'(1) = \mathbb{R}^2$.
3) $E'(\lambda)$ ist Untervektorraum von V. Es gilt nämlich für alle $k \in \mathbb{N}_{>0}$:

$$\ker\big(L - \lambda \mathrm{id}_V\big)^k \subseteq \ker\big(L - \lambda \mathrm{id}_V\big)^{k+1},$$

und es ist leicht einzusehen, dass die Vereinigung einer aufsteigenden Kette von Untervektorräumen selbst eine Untervektorraum ist. Wir werden später zeigen, dass in Wahrheit für alle $k \geq n$ gilt:

$$\ker\big(L - \lambda \mathrm{id}_V\big)^k = \ker\big(L - \lambda \mathrm{id}_V\big)^n .$$

4) $E(\lambda)$ und $E'(\lambda)$ sind L-invariant. Die L-Invarianz von $E'(\lambda)$ sieht man so: Ist $v \in E'(\lambda)$, so existiert $k \in \mathbb{N}_{>0}$ mit $\big(L - \lambda \mathrm{id}_V\big)^k(v) = 0$. Mit $L \circ \big(L - \lambda \mathrm{id}_V\big) = \big(L - \lambda \mathrm{id}_V\big) \circ L$ folgt dann:

$$\big(L - \lambda \mathrm{id}_V\big)^k\big(L(v)\big) = \big((L - \lambda \mathrm{id}_V)^k \circ L\big)(v) = \big(L \circ (L - \lambda \mathrm{id}_V)^k\big)(v) = 0 .$$

Also $L(v) \in \ker(L - \lambda \mathrm{id}_V)^k \subseteq E'(\lambda)$.

Etwas allgemeiner als Satz (7.29) ist:

(14.3) Satz (Diagonalisierbarkeitskriterium). *Sei V endlichdimensionaler K-Vektorraum und $L \in \mathrm{End}(V)$. Dann sind äquivalent:*

(a) $\dim V = \sum\limits_{\lambda \,\mathrm{EW\ von}\, L} \dim E(\lambda)$

(b) $V = \bigoplus\limits_{\lambda \,\mathrm{EW\ von}\, L} E(\lambda)$

(c) *L ist diagonalisierbar.*

Bem.: Zur Definition der direkten Summe von Unterräumen siehe S. 156.

Bew.: Der einzige nichttriviale Beweisschritt ist der Beweis der Implikation $(a) \Rightarrow (b)$.

Wir verwenden dabei folgende Aussage aus dem Beweis von (7.29):

$(*)$ Sind $v_1, \ldots, v_s$ Eigenvektoren von V zu verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_s$, so sind die $v_1, \ldots, v_s$ linear unabhängig.

Sind nun $\lambda_1, \ldots, \lambda_m$ die verschiedenen Eigenwerte von L und $B_1, \ldots, B_m$ Basen von $E(\lambda_1), \ldots, E(\lambda_m)$, so gilt wegen $E(\lambda_i) \cap E(\lambda_j) = \{0\}$ für $1 \leq i \neq j \leq m$ auch $B_i \cap B_j = \emptyset$ für $1 \leq i \neq j \leq m$. Aus (a) folgt nun für $B := B_1 \cup \ldots \cup B_m$:

$$\# B = \dim V.$$

Wenn wir zeigen können, daß B linear unabhängig (und damit eine Basis) ist, so ist (b) bewiesen. Wäre B nicht linear unabhängig, so existierte eine nichttriviale Linearkombination aus Elementen von B, die $0 \in V$ darstellt. Sammelt man für jedes $i \in \{1, \ldots, m\}$ die Terme, die Elemente von B_i enthalten, so erhalten wir Vektoren $v_1 \in E(\lambda_1), \ldots, v_m \in E(\lambda_m)$, die nicht alle gleich 0 sind und für die

$$v_1 + \ldots + v_m = 0$$

gilt. Betrachten wir nur diejenigen v_i, die nicht gleich 0 sind, so erhalten wir einen Widerspruch zu $(*)$.

Bem.: 1) Will man (14.3) auf einen konkreten Endomorphismus L anwenden, so muß man seine Eigenwerte, d.h. die Nullstellen seines charakteristischen Polynoms P_L bestimmen. Das wird meist nicht explizit, sondern nur approximativ (mit Hilfe der Numerik) möglich sein. Kennt man die Eigenwerte, so ist die Bestimmung der Eigenräume leicht. Sie sind die Lösungsmengen von homogenen linearen Gleichungssystemen

$$E(\lambda) = \{v \in V | L(v) - \lambda v = 0\}.$$

2) Es ist wichtig zu wissen, daß im Fall des Körpers $K = \mathbb{C}$ "die meisten" $A \in \mathbb{C}^{n\times n}$ diagonalisierbar sind. Das folgt nach (7.29) daraus, daß "die meisten" Polynome vom Grad n über $\mathbb{C}$ auch n verschiedene Nullstellen besitzen. Was genau dabei "die meisten" bedeutet, wird hier nicht erklärt und gehört eher in die Analysis III (Stichwort: Menge von Maß 0) oder Funktionalanalysis (Stichwort: Satz von Baire).

Wir beginnen mit einigen Tatsachen zu den Haupträumen.

(14.4) Fakt. Sei λ Eigenwert von $L \in \mathrm{End}(V)$. Dann ist λ der einzige Eigenwert von $L|E'(\lambda)$.

Bew.: (a) Wegen $\{0\} \neq E(\lambda) \subseteq E'(\lambda)$ ist λ Eigenwert von $L|E'(\lambda)$.
(b) Ist μ ein Eigenwert von $L|E'(\lambda)$ mit zugehörigem Eigenvektor $v \in E'(\lambda)$, so gilt für alle $k \in \mathbb{N}_{>0}$:

$$(L - \lambda\,\mathrm{id}_V)^k(v) = (\mu - \lambda)^k v.$$

Wegen $v \in E'(\lambda)$ existiert ein $k \in \mathbb{N}_{>0}$, so daß

$$(L - \lambda\,\mathrm{id}_V)^k(v) = 0.$$

gilt. Die vorangehenden Gleichungen zeigen wegen $v \neq 0$, daß $\mu = \lambda$ gilt.

Für den Rest dieses Kapitels sei V stets ein endlichdimensionaler Vektorraum über einem Körper K und $L : V \to V$ sei ein Endomorphismus von V.

(14.5) Lemma. Sei λ Eigenwert von $L \in \mathrm{End}(V)$. Dann existiert die kleinste Zahl $f(\lambda) \in \mathbb{N}$, genannt der Index des Eigenwerts λ von L, so daß

$$E'(\lambda) = \ker(L - \lambda \,\mathrm{id}_V)^{f(\lambda)}$$

gilt. Weiter gilt

$$V = E'(\lambda) \oplus \mathrm{im}(L - \lambda \,\mathrm{id}_V)^{f(\lambda)}.$$

Bem.: Wegen $L \circ (L - \lambda \,\mathrm{id}_V)^{f(\lambda)} = (L - \lambda \,\mathrm{id}_V)^{f(\lambda)} \circ L$ ist der Unterraum $\mathrm{im}(L - \lambda \,\mathrm{id}_V)^{f(\lambda)}$ L-invariant.

Bew.: Sei $v_1, \dots, v_l$ eine Basis von $E'(\lambda)$. Nach Definition von $E'(\lambda)$ existiert für jedes $i \in \{1, \dots, l\}$ das kleinste $k_i \in \mathbb{N}$ mit $(L - \lambda \,\mathrm{id}_V)^{k_i}(v_i) = 0$.

Sei $f(\lambda) := \max\limits_{1 \le i \le l} k_i$. Dann gilt $(L - \lambda \,\mathrm{id}_V)^{f(\lambda)}(v_i) = 0$ für alle Elemente v_i unserer Basis von $E'(\lambda)$, also

$$E'(\lambda) \subseteq \ker(L - \lambda \,\mathrm{id}_V)^{f(\lambda)}.$$

Nach Definition von $E'(\lambda)$ gilt auch die umgekehrte Inklusion, also $E'(\lambda) = \ker(L - \lambda \,\mathrm{id}_V)^{f(\lambda)}$. Offenbar ist $f(\lambda)$ die kleinste Zahl mit dieser Eigenschaft. Zum Beweis der zweiten Behauptung nehmen wir an, daß

$$v \in E'(\lambda) \cap \mathrm{im}(L - \lambda \,\mathrm{id}_V)^{f(\lambda)}$$

gilt, und zeigen, daß dann $v = 0$ ist. Es existiert also ein $w \in V$ mit

$$v = (L - \lambda \,\mathrm{id}_V)^{f(\lambda)}(w).$$

Da $v \in E'(\lambda)$ ist, gilt $(L - \lambda \,\mathrm{id}_V)^{f(\lambda)}(v) = 0$, also

$$(L - \lambda \,\mathrm{id}_V)^{2f(\lambda)}(w) = 0.$$

Das zeigt, daß auch $w \in E'(\lambda)$ gilt, und aus der ersten Aussage von (14.5) schließen wir

$$0 = (L - \lambda \,\mathrm{id}_V)^{f(\lambda)}(w) = v.$$

Das zeigt, daß die Summe aus $E'(\lambda)$ und $\mathrm{im}(L-\lambda\,\mathrm{id}_V)^{f(\lambda)}$ direkt ist. Wegen $E'(\lambda)=\ker(L-\lambda\,\mathrm{id}_V)^{f(\lambda)}$ implizieren die Dimensionssätze (3.23) und (4.10)

$$\dim(E'(\lambda)\oplus \mathrm{im}(L-\lambda\,\mathrm{id}_V)^{f(\lambda)})=\dim V.$$

Daraus folgt die Behauptung, vgl. (3.19)(c).

(14.6) Lemma. Seien $\lambda_1,\ldots,\lambda_s$ verschiedene Eigenwerte von $L\in\mathrm{End}(V)$. Dann ist die Summe der Hauptträume $E'(\lambda_i)$ für $1\le i\le s$ direkt.

Bew.: Durch vollständige Induktion nach s. Für $s=1$ ist nichts zu zeigen.

Für den Induktionsschritt müssen wir zeigen, daß für alle $i\in\{1,\ldots,s\}$ gilt

$$E'(\lambda_i)\cap \mathrm{span}(\bigcup_{\substack{j=1\\ j\neq i}}^{s} E'(\lambda_j))=\{0\},$$

vgl. S. 156. Da diese Aussage von der Numerierung der λ_i unabhängig ist, genügt es, sie für den Fall $i=s$ zu beweisen. Wir müssen also zeigen, daß jedes $v\in E'(\lambda_s)$, das sich in der Form

$$v=\sum_{i=1}^{s-1} v_i$$

mit $v_i\in E'(\lambda_i)$ schreiben läßt, notwendig der 0-Vektor ist. Wegen $v\in E'(\lambda_s)$ gilt nach (14.5)

$$(*)\qquad 0=(L-\lambda_s\,\mathrm{id}_V)^{f(\lambda_s)}(v)=\sum_{i=1}^{s-1}(L-\lambda_s\,\mathrm{id}_V)^{f(\lambda_s)}(v_i).$$

Die Hauptträume $E'(\lambda_i)$ sind L-invariant und daher auch $(L-\lambda_s\,\mathrm{id}_V)^{f(\lambda_s)}$-invariant. Also folgt aus $v_i\in E'(\lambda_i)$ auch $(L-\lambda_s\,\mathrm{id}_V)^{f(\lambda_s)}(v_i)\in E'(\lambda_i)$ für $1\le i\le s-1$. Die Induktionsvoraussetzung und $(*)$ implizieren nun, daß für alle $i\in\{1,\ldots,s-1\}$ gilt.

$$(L-\lambda_s\,\mathrm{id}_V)^{f(\lambda_s)}(v_i)=0,$$

also $v_i\in E'(\lambda_i)\cap E'(\lambda_s)$. Wir zeigen nun, daß der Durchschnitt von Haupträumen zu verschiedenen Eigenwerten stets $\{0\}$ ist. Daraus folgt dann $v_i=0$

für $1 \leq i \leq s-1$, und damit $v = \sum_{i=1}^{s-1} v_i = 0$, wie behauptet. Seien also $\lambda \neq \mu$ Eigenwerte von L und $w \in E'(\lambda) \cap E'(\mu)$. Sei $k \in \mathbb{N}$ die kleinste Zahl, für die $(L - \mu\,\mathrm{id}_V)^k(w) = 0$ gilt. Wäre $w \neq 0$, so würde $k \geq 1$ gelten und $(L - \mu\,\mathrm{id}_V)^{k-1}(w)$ wäre ein Eigenvektor von $L|E'(\lambda)$ zum Eigenwert μ. Das stünde im Widerspruch zu (14.4), wonach λ der einzige Eigenwert von $L|E'(\lambda)$ ist. Also gilt $w = 0$, was zu beweisen war.

Die Grundidee bei der Untersuchung eines $L \in \mathrm{End}(V)$ ist es, den Vektorraum V in eine direkte Summe von möglichst kleinen L-invarianten Untervektorräumen zu zerlegen. Diagonalisierbarkeit von L bedeutet ja gerade, daß V als direkte Summe von 1-dimensionalen L-invarianten Unterräumen dargestellt werden kann. Der erste wichtige Schritt in dieser Richtung ist der *1. Zerlegungssatz:*

(14.7) Satz. *Sei V endlichdimensionaler K-Vektorraum, $L \in \mathrm{End}(V)$ und das charakteristische Polynom P_L von L zerfalle. Sind $\lambda_1, \ldots, \lambda_s$ die verschiedenen Eigenwerte von L, so gilt*

$$V = E'(\lambda_1) \oplus \ldots \oplus E'(\lambda_s).$$

Bew.: Nach Lemma (14.6) ist die Summe direkt, so daß nur zu zeigen bleibt, daß sie ganz V ist, d.h. daß $V = E'(\lambda_1) + \ldots + E'(\lambda_s)$ gilt. Wir zeigen das durch Induktion nach $n := \dim V$. Im Fall $n = 1$ ($\Rightarrow s = 1$) ist nichts zu beweisen. Im Induktionsschritt benutzen wir die zweite Aussage von Lemma (14.5): Für den L-invarianten Unterraum

$$U := \mathrm{im}(L - \lambda_1\,\mathrm{id}_V)^{f(\lambda_1)}$$

gilt $V = E'(\lambda_1) \oplus U$. Wir wollen die Induktionsvoraussetzung auf $\overline{L} := L|U \in \mathrm{End}(U)$ anwenden. Wegen $\dim E'(\lambda_1) \geq 1$ gilt ja $\dim U = \dim V - \dim E'(\lambda_1) \leq n-1$. Wir müssen uns aber auch überzeugen, daß das charakteristische Polynom $P_{\overline{L}}$ von $\overline{L}$ ebenfalls zerfällt. Aus $V = E'(\lambda_1) \oplus U$ folgt

$$(*) \qquad P_L = P_{L|E'(\lambda_1)} \cdot P_{\overline{L}},$$

vgl. etwa LA I, Blatt 10, Aufgabe 1, oder direkt die Leibnizformel (7.14). Folgerung (13.17) zeigt also, daß $P_{L|E'(\lambda_1)}$ und $P_{\overline{L}}$ zerfallen. Als nächstes überlegen wir uns, daß $P_{\overline{L}}$ genau die Nullstellen $\lambda_2, \ldots, \lambda_s$ besitzt. Da $L|E'(\lambda_1)$

nach (14.4) nur den EW λ_1 besitzt, hat das zerfallende Polynom $P_{L|E'(\lambda_1)}$ nur die Nullstelle λ_1, d.h.

$$P_{L|E'(\lambda_1)} = (\lambda_1 - x)^{\dim E'(\lambda_1)}.$$

Für $2 \leq i \leq s$ gilt also $P_{L|E'(\lambda_1)}(\lambda_i) \neq 0$ und $P_L(\lambda_i) = 0$, so daß aus $(*)$ folgt

$$P_{\overline{L}}(\lambda_i) = 0$$

für $2 \leq i \leq s$. $P_{\overline{L}}$ besitzt keine weiteren Nullstellen, da diese auch Nullstellen von P_L sein müssten, so daß höchstens noch λ_1 möglich wäre. Es gilt jedoch $P_{\overline{L}}(\lambda_1) \neq 0$, da $\overline{L} = L|U$ wegen $U \cap E(\lambda_1) = \{0\}$ nicht den Eigenwert λ_1 hat. Die Induktionsvoraussetzung ergibt also

$$U = E'_{\overline{L}}(\lambda_2) + \ldots + E'_{\overline{L}}(\lambda_s),$$

wobei $E'_{\overline{L}}(\lambda_i) \subseteq U$ den Hauptraum von $\overline{L}$ zum Eigenwert λ_i von $\overline{L}$ bezeichnet. Wegen $V = E'(\lambda_1) \oplus U$ und $E'_{\overline{L}}(\lambda_i) \subseteq E'(\lambda_i)$ folgt

$$V = E'(\lambda_1) + \ldots + E'(\lambda_s),$$

und genau das war noch zu beweisen.

Satz (14.7) reduziert die Untersuchung von Endomorphismen mit zerfallendem charakteristischen Polynom auf den Fall von Endomorphismen, die nur einen einzigen Eigenwert, etwa $\lambda \in K$, besitzen und deren charakteristisches Polynom von der Form $(\lambda - x)^n$ ist. Etwas expliziter werden wir jetzt für einen Eigenwert λ von $L \in \mathrm{End}(V)$ versuchen, $L|E'(\lambda)$ zu verstehen. Für $T_\lambda := (L - \lambda\,\mathrm{id}_V)|E'(\lambda) \in \mathrm{End}(E'(\lambda))$ gilt dann nach (14.5)

$$(T_\lambda)^{f(\lambda)} = 0 \in \mathrm{End}(E'(\lambda))$$

(14.8) Def.: Ein Endomorphismus $T \in \mathrm{End}(V)$ heißt <u>nilpotent</u>, falls ein $k \in \mathbb{N}_{>0}$ existiert, so daß $T^k = 0$ gilt. Das kleinste solche $k \in \mathbb{N}_{>0}$ heißt der <u>Nilpotenzgrad</u> $g(T)$ <u>von</u> T.

Bsp.: 1) $T_\lambda := (L - \lambda\,\mathrm{id}_V)|E'(\lambda)$ ist nilpotent vom Nilpotenzgrad $f(\lambda)$.
2) Besitzt ein Endomorphismus $T \in \mathrm{End}(V)$ bezüglich einer Basis von V eine Matrix $A \in K^{n\times n}$ der Form

$$A = \begin{pmatrix} 0 & \ldots & \ldots & 0 \\ * & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ * & \ldots & * & 0 \end{pmatrix},$$

so ist L nilpotent, da $A^n = 0$ gilt. (Übung!).

(14.9) Lemma. Sei $0 \neq T \in \mathrm{End}(V)$ nilpotent und $g := g(T)$. Ist $v \in V \setminus \ker(T^{g-1})$, so ist $\mathrm{span}\{v, T(v), \ldots, T^{g-1}(v)\}$ ein g-dimensionaler T-invarianter Untervektorraum von V. Speziell folgt $g(T) \leq \dim V$.

Bew.: (a) T-Invarianz von $W := \mathrm{span}\{v, T(v), \ldots, T^{g-1}(v)\}$: Sei $w \in W$, d.h. $w = \sum\limits_{i=0}^{g-1} a_i T^i(v)$ mit $a_i \in K$ für $0 \leq i \leq g-1$. Dann gilt

$$T(w) = \sum_{i=0}^{g-1} a_i T^{i+1}(v) = \sum_{i=1}^{g-1} a_{i-1} T^i(v) \in W,$$

da $T^g(v) = 0$ ist.

(b) Lineare Unabhängigkeit von $v, T(v), \ldots, T^{g-1}(v)$: Sei $\sum\limits_{i=0}^{g-1} a_i T^i(v) = 0$. Wir zeigen durch Induktion nach i, daß $a_i = 0$ gilt. Induktionsanfang: Wir wenden T^{g-1} auf die Gleichung $\sum\limits_{i=0}^{g-1} a_i T^i(v) = 0$ an und erhalten

$$\sum_{i=0}^{g-1} a_i T^{g+i-1}(v) = 0.$$

Da $T^k(v) = 0$ für $k \geq g$ gilt, gilt $T^{g+i-1}(v) = 0$ für $i \geq 1$, d.h.

$$0 = \sum_{i=0}^{g-1} a_i T^{g+i-1}(v) = a_0 T^{g-1}(v).$$

Nun folgt aus der Voraussetzung $T^{g-1}(v) \neq 0$, daß $a_0 = 0$ gilt.

Induktionsschritt: Wir können annehmen, daß $a_0 = \ldots = a_{i-1} = 0$ gilt, also $\sum\limits_{j=i}^{g-1} a_j T^j(v) = 0$. Auf diese Gleichung wenden wir T^{g-i-1} an und erhalten mit der gleichen Begründung wie im Induktionsanfang

$$a_i = 0.$$

(14.10) Def.: Sei $L \in \mathrm{End}(V)$. Ein Untervektorraum U von V heißt L-irreduzibel, falls U L-invariant ist und U nicht als direkte Summe von zwei L-invarianten Untervektorräumen positiver Dimension dargestellt werden kann, d.h. gilt $U = U_1 \oplus U_2$ mit L-invarianten Untervektorräumen U_1 und U_2, so gilt $U_1 = \{0\}$ oder $U_2 = \{0\}$. Ein L-invarianter Unterraum U heißt L-reduzibel, falls U nicht L-irreduzibel ist.

(14.11) Lemma. Sei $T \in \mathrm{End}(V)$ nilpotent und $g(T) = n = \dim V$. Dann ist V T-irreduzibel und für jedes $v \in V \setminus \ker(T^{n-1})$ ist $\mathcal{G} = (v, T(v), \dots, T^{n-1}(v))$ eine Basis von V. Es gilt

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(T) = \begin{pmatrix} 0 & \dots & \dots & \dots & 0 \\ 1 & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

Bew.: Seien U_1, U_2 T-invariante Untervektorräume von V, so daß $V = U_1 \oplus U_2$ gilt. Dann sind $T|U_1$ und $T|U_2$ nilpotent und aus (14.9) folgt $g(T|U_1) \leq \dim U_1$, $g(T|U_2) \leq \dim U_2$. Daraus folgt

$$n = g(T) = \max\{g(T|U_1), g(T|U_2)\} \leq \max\{\dim U_1, \dim U_2\}.$$

Also hat einer der beiden Unterräume die Dimension n, o.E. $\dim U_1 = n$. Daraus folgt $U_1 = V$ und $U_2 = \{0\}$. Das beweist, daß V T-irreduzibel ist. Ist $v \in V \setminus \ker(T^{n-1})$, so folgt aus (14.9), daß $\mathcal{G} = (v, T(v), \dots, T^{n-1}(v))$ eine Basis von V ist. Daraus folgt direkt die behauptete Form von $\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(T)$.

Bem.: Sei $L \in \mathrm{End}(V)$, λ Eigenwert von L und $U \subseteq E'(\lambda)$ ein L-invarianter Untervektorraum, so daß für $T := (L - \lambda\,\mathrm{id}_V)|U$ gilt:

$$g(T) = \dim U.$$

Nach (14.11) existiert dann eine Basis $\mathcal{G}$ von U, so daß

$$\mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L|U) = \begin{pmatrix} \lambda & 0 & \dots & \dots & 0 \\ 1 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & \lambda \end{pmatrix}$$

gilt. Eine solche Matrix nennt man einen "Jordanblock".

Wir werden als nächstes zeigen, daß sich $E'(\lambda)$ als direkte Summe von Unterräumen dieses Typs darstellen läßt. Dazu zunächst folgender einfacher Hilfssatz:

(14.12) Lemma. Sei $V \neq \{0\}$ und $L \in \mathrm{End}(V)$. Dann existiert ein $k \in \mathbb{N}_{>0}$ und L-irreduzible Untervektorräume $U_1 \neq \{0\}, \ldots, U_k \neq \{0\}$, so daß

$$V = U_1 \oplus \ldots \oplus U_k$$

gilt.

Bew.: Durch Induktion nach $n := \dim V$. Für $n = 1$ ist nichts zu beweisen. Im Induktiosschritt gibt es zwei Fälle:
(a) V ist L-irreduzibel. Dann ist die Behauptung für $k := 1$ und $U_1 := V$ erfüllt.
(b) Es existieren L-invariante Unterräume U_1, U_2 von V, so daß $V = U_1 \oplus U_2$ und $U_1 \neq \{0\}$, $U_2 \neq \{0\}$ gilt. Dann kann man die Induktionsvoraussetzung auf U_1 und U_2 anwenden und erhält daraus die gesuchte Zerlegung von V.

Der zweite wesentliche Schritt zur Konstruktion der Jordanschen Normalform ist folgende Umkehrung von (14.11).

(14.13) Lemma (2. Zerlegungssatz). Sei $T \in \mathrm{End}(V)$ nilpotent und $g = g(T) < n = \dim V$. Dann ist V T-reduzibel.

Bew.: Wegen $T^{g-1} \neq 0$ können wir ein $v \in V \setminus \ker(T^{g-1})$ wählen und dazu ein $l \in V^*$ mit $l(T^{g-1}(v)) \neq 0$. Wir definieren

$$\begin{aligned} &U_1 := \mathrm{span}\{v, \ldots, T^{g-1}(v)\} \quad \text{ und} \\ &U_2 := \ker(l) \cap \ker(l \circ T) \cap \ldots \cap \ker(l \circ T^{g-1}). \end{aligned}$$

Dann ist U_1 T-invariant, und nach (14.9) gilt $0 < \dim U_1 = g < n$. Wir haben (14.13) bewiesen, wenn wir zeigen können, daß U_2 T-invariant ist und daß $V = U_1 \oplus U_2$ gilt.

T-Invarianz von U_2: Ist $w \in U_2$, so gilt $l(w) = 0, \ldots, l(T^{g-1}(w)) = 0$, also speziell $l(T(w)) = 0, \ldots, l(T^{g-2}(T(w)) = 0$ und außerdem wegen $T^g(w) = 0$: $l(T^g(w)) = l(T^{g-1}(T(w)) = 0$. Zusammen zeigen diese Gleichungen, daß $T(w) \in U_2$ gilt. Also ist U_2 T-invariant.

Zeige $U_1 \cap U_2 = \{0\}$: Sei $w \in U_1 \cap U_2$. Da $w \in U_1$ ist, existieren $a_0, \ldots, a_{g-1} \in K$ mit $w = \sum\limits_{i=0}^{g-1} a_i T^i(v)$. Da $w \in U_2$ ist, gilt $l(T^{g-1}(w))) \ldots = l(w) = 0$.

Wegen $T^{g-1}(w) = a_0 T^{g-1}(v)$ gilt $0 = l(T^{g-1}(w)) = a_0 l(T^{g-1}(v))$, und wegen $l(T^{g-1}(v)) \neq 0$ folgt daraus $a_0 = 0$. Wie im Beweis von (14.9) folgt induktiv $a_1 = 0, \ldots, a_{g-1} = 0$, also $w = 0$. Das beweist $U_1 \cap U_2 = \{0\}$. Es liegt also eine direkte Summe $U_1 \oplus U_2$ vor, und wir haben

$$\dim U_1 \oplus U_2 = \dim U_1 + \dim U_2.$$

Es gilt $\dim U_1 = g$ und aus (3.24) folgt $\dim U_2 \geq n - g$, also

$$\dim U_1 \oplus U_2 \geq g + (n-g) = n = \dim V.$$

Da $U_1 \oplus U_2$ Untervektorraum von V ist, folgt daraus $V = U_1 \oplus U_2$, wie behauptet.

Zusammen besagen (14.11) und (14.13) gerade, daß für ein nilpotentes $0 \neq T \in \mathrm{End}(V)$ gilt:

$$V \ T\text{-irreduzibel} \Leftrightarrow g(T) = \dim V.$$

Auf jedem T-irreduziblen Unterraum U von V ist $T|U$ also von der einfachen, in (14.11) beschriebenen Bauart.

Wir beweisen nun zunächst den Satz über die Jordansche Normalform in der Formulierung für Endomorphismen und werden das später in eine Aussage über Matrizen übersetzen.

(14.14) Satz (Jordan-Zerlegung). *Sei $L \in \mathrm{End}(V)$ und das charakteristische Polynom von L zerfalle. Sind $\lambda_1, \ldots, \lambda_s$ die verschiedenen EWe von L, so gilt*

$$L = E'(\lambda_1) \oplus \ldots \oplus E'(\lambda_s).$$

Jeder der Haupträume $E'(\lambda_t), 1 \leq t \leq s$, zerfällt in L-irreduzible Unterräume

$$E'(\lambda_t) = U_t^1 \oplus \ldots \oplus U_t^{k_t}$$

und für jedes $i \in \{1, \ldots, k_t\}$ ist $(L - \lambda_t \,\mathrm{id}_V)|U_t^i =: S_t^i \in \mathrm{End}(U_t^i)$ nilpotent vom Nilpotenzgrad $g(S_t^i) = \dim U_t^i$.

Bew.: Die erste Aussage ist gerade der 1. Zerlegungssatz (14.7). Wir wenden dann (14.12) auf $L|E'(\lambda_t)$ an und erhalten eine Zerlegung von $E'(\lambda_t)$ in L-irreduzible Unterräume, genannt $U_t^1, \ldots, U_t^{k_t}$. Nach (14.5) wissen wir, daß $(L - \lambda_t \,\mathrm{id}_V)|U_t^i =: S_t^i$ für jedes $t \in \{1, \ldots, s\}$ und jedes $i \in \{1, \ldots, k_t\}$

nilpotent ist. Da U_t^i L-irreduzibel ist, ist U_t^i auch S_t^i-irreduzibel. Aus dem 2. Zerlegungssatz (14.13) folgt nun, daß der Nilpotenzgrad $g(S_t^i)$ von S_t^i gleich der Dimension von U_t^i ist, d.h. daß S_t^i von dem einfachen, durch (14.11) beschriebenen Typ ist.

Nach (14.11) existiert zu jedem der L-irreduziblen Unterräume U_t^i, $1 \le t \le s$, $1 \le i \le k_t$ eine Basis $\mathcal{G}_t^i$, so daß

$$\mathrm{Mat}_{\mathcal{G}_t^i}^{\mathcal{G}_t^i}(S_t^i) = \begin{pmatrix} 0 & \dots & \dots & \dots & 0 \\ 1 & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$$

gilt. Wegen $L|U_t^i = S_t^i + \lambda_t \mathrm{id}_{U_t^i}$ folgt

$$(*) \qquad \mathrm{Mat}_{\mathcal{G}_t^i}^{\mathcal{G}_t^i}(L|U_t^i) = \begin{pmatrix} \lambda_t & \dots & \dots & \dots & 0 \\ 1 & \ddots & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & \lambda_t \end{pmatrix} =: J_t^i.$$

Damit erhalten wir

(14.15) Satz (Jordansche Normalform). *Sei $L \in \mathrm{End}(V)$ und das charakteristische Polynom P_L von L zerfalle. Die verschiedenen Nullstellen von P_L seien $\lambda_1, \dots, \lambda_s$. Dann existiert eine Basis $\mathcal{G}$ von V, so daß*

$$(**) \qquad \mathrm{Mat}_{\mathcal{G}}^{\mathcal{G}}(L) = \begin{pmatrix} J_1^1 & & & & 0 \\ & \ddots & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & & J_s^{k_s} \end{pmatrix}$$

gilt, wobei die in der Diagonale stehenden „Jordanblöcke“ J_t^i $(\dim U_t^i) \times (\dim U_t^i)$-Matrizen vom Typ $(*)$ sind.

Interpretiert für Matrizen $A \in K^{n \times n}$ besagt (14.15):

(14.16) Folgerung. Sei $A \in K^{n\times n}$ und das charakteristische Polynom von A zerfalle. Dann existiert ein $B \in \mathrm{GL}_n(K)$, so daß $B^{-1}AB$ die Form $(**)$ hat.

Bem.:

1) Ist $K = \mathbb{C}$, so ist die Voraussetzung „das charakteristische Polynom zerfalle“ nach dem Fundamentalsatz der Algebra stets erfüllt. Ist K ein Unterkörper von $\mathbb{C}$, z.B. $K = \mathbb{Q}$ oder $K = \mathbb{R}$, und zerfällt das charakteristische Polynom in $K[x]$ nicht, so können wir uns durch „Komplexifizierung“ der Situation helfen, d.h. ist $L \in \mathrm{End}(\mathbb{R}^n)$, so betrachten wir die $\mathbb{C}$-lineare Fortsetzung $\tilde{L} \in \mathrm{End}(\mathbb{C}^n)$ von L, vgl. Blatt 6, Aufgabe 1. In der Algebra konstruiert man zu jedem Körper K einen K enthaltenden Körper $\tilde{K}$, in dem jedes Polynom zerfällt. Also kann man die Voraussetzung „das charakteristische Polynom zerfalle“ auch für beliebige Körper durch ein der Komplexifizierung ähnliches Verfahren erzwingen.

2) Wie bereits gesagt ist das Hauptproblem bei der Berechnung der Jordanschen Normalform – die Berechnung der EWe $\lambda_1, \dots, \lambda_s$– i.a. nicht explizit lösbar. Kennt man $\lambda_1, \dots, \lambda_s$, so ist die Jordansche Normalform und eine zugehörige Basis durch Lösen von linearen Gleichungssystemen berechenbar. Dazu noch einige Hinweise:
Zu berechnen sind für jedes $t \in \{1, \dots, s\}$ die Zahl k_t der L-irreduziblen Unterräume $U_t^1, \dots, U_t^{k_t}$, in die $E'(\lambda_t)$ zerfällt, und die Dimensionen der U_t^i. Offenbar gilt

$$\sum_{i=1}^{k_t} \dim U_t^i = \dim E'(\lambda_t) \text{ und } \sum_{t=1}^{s} \dim E'(\lambda_t) = \dim V.$$

Berechnet man P_L mit der Jordanschen Normalform, so sieht man, daß $\dim E'(\lambda_t)$ gerade die Vielfachheit der Nullstelle λ_t ist. $E'(\lambda_t)$ kann nach (14.2), (14.5) durch Lösen homogener linearer Gleichungen bestimmt werden. Die Zerlegung von $E'(\lambda_t)$ in L-irreduzible Unterräume kann mit dem im Beweis von (14.13) (2. Zerlegungssatz) verwendeten Verfahren explizit durchgeführt werden.

Bsp.: *Berechnung der Jordanschen Normalform von*

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

Man berechnet $P_A(\lambda) = \det(A - \lambda E_3) = -(\lambda - 2)^3$. Da P_A zerfällt und $\lambda = 2$ der einzige Eigenwert von A ist, ist

$$A - 2E_3 =: T$$

nilpotent, vgl. (14.7) und (14.5). Der Nilpotenzgrad $g = g(T)$ von T, d.h. die kleinste Zahl $g \in \mathbb{N}_{>0}$, für die $T^g = 0$ gilt, ist 3, da

$$T^2 = \begin{pmatrix} 0 & 2 & 2 \\ 0 & -2 & -2 \\ 0 & 2 & 0 \end{pmatrix} \neq 0$$

gilt, vgl. (14.9). Nach (14.11) ist $\mathbb{R}^3$ T-irreduzibel und (14.15) impliziert, daß die Jordansche Normalform $\tilde{A} \in \mathbb{R}^{3\times 3}$ aus einem einzigen Jordanblock zum Eigenwert $\lambda = 2$ besteht, d.h.

$$\tilde{A} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

Berechnung einer Matrix $B \in \mathrm{GL}(3, \mathbb{R})$*, so daß*

$$B^{-1}AB = \tilde{A}$$

gilt: Nach (14.9) gilt für jedes $v \in \mathbb{R}^3 \setminus \ker(T^2)$, daß

$$\mathcal{G} = (v, T(v), T^2(v))$$

eine Basis des $\mathbb{R}^3$ ist, so daß $\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(A) = \tilde{A}$ gilt. Wir identifizieren hierbei A und T mit den Endomorphismen $v \in \mathbb{R}^{3\times 1} \to Av \in \mathbb{R}^{3\times 1}$ und $v \in \mathbb{R}^{3\times 1} \to Tv \in \mathbb{R}^{3\times 1}$. Wir können etwa $v = \left(\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right)$ wählen und erhalten $Tv = \left(\begin{smallmatrix} 3 \\ -1 \\ 1 \end{smallmatrix}\right)$, $T^2v = \left(\begin{smallmatrix} 2 \\ -2 \\ 2 \end{smallmatrix}\right)$. Nach (5.6) gilt dann für

$$B = \begin{pmatrix} 0 & 3 & 2 \\ 0 & -1 & -2 \\ 1 & 1 & 2 \end{pmatrix} \left(= \mathrm{Mat}^{(e_1,e_2,e_2)}_{\mathcal{G}}(\mathrm{id}_{\mathbb{R}^3})\right):$$

$$B^{-1}AB = \mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(A) = \tilde{A}.$$

Aus der Theorie wissen wir, daß das so ist, und man kann es durch Nachrechnen bestätigen.

Im vorangehenden Beispiel wurde von der Berechnung “der” Jordanschen Normalform gesprochen, obwohl wir uns über die Eindeutigkeit der Jordanschen Normalform noch keine Gedanken gemacht haben. Das soll jetzt nachgeholt werden. Daß es zu einem $L \in \mathrm{End}(V)$ recht verschiedene Basen $\mathcal{G}$ von V geben kann, so daß $\mathrm{Mat}^{\mathcal{G}}_{\mathcal{G}}(L)$ Jordansche Normalform hat, sehen wir am vorangehenden Beispiel: Der erste Vektor v der Basis kann dort ganz beliebig in $\mathbb{R}^3 \setminus \ker(T^2)$ gewählt werden. Wir werden jedoch zeigen, daß die Jordansche Normalform $(**)$ selbst bis auf die Reihenfolge der Jordanblöcke eindeutig bestimmt ist. Der entscheidende Schritt zu dieser Eindeutigkeit ist:

(14.17) Satz. *Sei $T \in \mathrm{End}(V)$ nilpotent und*

$$V = Z_1 \oplus \ldots \oplus Z_s$$

eine Zerlegung von V als direkte Summe von T-irreduziblen Unterräumen $Z_\sigma \neq \{0\}$. Für $l \in \mathbb{N}_{>0}$ sei N_l die Anzahl der $\sigma \in \{1, \ldots, s\}$, für die $\dim Z_\sigma = l$ *gilt, und für $l \in \mathbb{N}$ sei*

$$R_l := \mathrm{rg}(T^l).$$

Dann gilt für alle $l \in \mathbb{N}_{>0}$

$$N_l = R_{l+1} - 2R_l + R_{l-1}.$$

Bem.: Die Bedeutung von (14.17) liegt darin, daß die Zahlen N_l, die a priori von der gewählten Zerlegung abhängen könnten, in Wahrheit durch T eindeutig bestimmt sind, da sie aus den Zahlen $R_l = \mathrm{rg}(T^l)$ berechnet werden können.

Bew.: Wir bemerken zunächst:

(1) Ist $l > g(T)$, so ist $N_l = 0$.

Das folgt mittels (14.13) aus der Irreduzibilität der Z_σ:

$$\dim Z_\sigma \overset{(14.13)}{=} g(T|Z_\sigma) \leqq g(T).$$

Wir drücken nun die Zahlen R_l durch die Zahlen N_l aus. Dazu definieren wir für $1 \leq l \leq g(T)$:

$$W_l := \bigoplus_{\dim Z_\sigma = l} Z_\sigma. \tag{2}$$

Dann gilt wegen (1):

$$V = W_1 \oplus \ldots \oplus W_{g(T)}. \tag{3}$$

Da alle auftretenden Unterräume T-invariant sind, gilt für alle $j \in \{1, \ldots, g(T)\}$

$$T^j(W_l) = \bigoplus_{\dim Z_\sigma = l} T^j(Z_\sigma) \tag{4}$$

und

$$T^j(V) = T^j(W_1) \oplus \ldots \oplus T^j(W_{g(T)}). \tag{5}$$

Wir bestimmen nun $\dim T^j(Z_\sigma)$ aus $l := \dim Z_\sigma$. Nach (14.9) und (14.13) existiert ein $v \in Z_\sigma$, so daß

$$v, T(v), \ldots, T^{l-1}(v)$$

eine Basis von Z_σ ist, wobei $T^l(v) = 0$ gilt. Daraus folgt, daß für $j < l$ $T^j(v), \ldots, T^{l-1}(v)$ eine Basis von $T^j(Z_\sigma)$ ist, also

$$\dim T^j(Z_\sigma) = \begin{cases} l - j & \text{für } j < l \\ 0 & \text{für } j \geq l \end{cases} \tag{6}$$

Nach Definition von N_l und nach (2), (4) und (6) gilt für $j < l$

$$\dim(T^j(W_l)) = (l - j)N_l.$$

Aus (5) folgt nun für alle $j \in \mathbb{N}$:

$$\begin{aligned} R_j &= \dim T^j(V) = \sum_{l=1}^{g(T)} T^j(W_l) \overset{(6)}{=} \sum_{l=j+1}^{g(T)} T^j(W_l) \overset{(6)}{=} \sum_{l=j+1}^{g(T)} (l-j)N_l \\ &= \sum_{i=1}^{g(T)-j} iN_{j+i} \overset{(1)}{=} \sum_{i \geq 1} iN_{j+i}. \end{aligned}$$

Daraus folgt für alle $j \in \mathbb{N}_{>0}$

$$\begin{aligned} R_{j+1} - 2R_j + R_{j-1} &= \sum_{i \geq 1} iN_{j+1+i} - 2\sum_{i\geq 1} iN_{j+i} + \sum_{i\geq 1} iN_{j-1+i} \\ &= \sum_{i\geq 2}(i-1)N_{j+i} - 2\sum_{i\geq 1} iN_{j+i} + \sum_{i \geq 0}(i+1)N_{j+i} \\ &= \sum_{i\geq 2}((i-1) - 2i + (i+1))N_{j+i} - 2N_{j+1} + N_j + 2N_{j+1} \\ &= N_j, \end{aligned}$$

wie behauptet.

Wir betrachten nun den allgemeinen Fall: Es sei $L \in \operatorname{End}(V)$ und das charakteristische Polynom P_L von L zerfalle. Es sei $\mathcal{G}$ eine Basis von V, so daß $\operatorname{Mat}^{\mathcal{G}}_{\mathcal{G}}(L) =: \tilde{A}$ folgende Form hat

$$\tilde{A} = \begin{pmatrix} J_1^1 & & 0 \\ & \ddots & \\ 0 & & J_s^{k_s} \end{pmatrix},$$

wobei für $1 \leq t \leq s$ und $1 \leq i \leq k_t$ die Matrix J_t^i ein Jordanblock der Form

$$I_t^i = \begin{pmatrix} \lambda_t & 0 & \dots & \dots & 0 \\ 1 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & \lambda_t \end{pmatrix}$$

ist und die λ_t für $1 \leq t \leq s$ verschiedene Körperelemente sind. Da $P_L = \det(\tilde{A} - xE_n)$ gilt und $\det(\tilde{A} - xE_n)$ genau die Nullstellen $\lambda_1, \dots, \lambda_s$ hat, sind $\lambda_1, \dots, \lambda_s$ genau die Eigenwerte von L.

Sei für $t \in \{1, \dots, s\}$ mit $\mathcal{G}_t \subseteq \mathcal{G}$ die Teilmenge von $\mathcal{G}$ bezeichnet, die zu den Jordanblöcken $J_t^1, \dots, J_t^{k_t}$ gehört. Durch Betrachtung der Potenzen $(\tilde{A} - \lambda_t E_n)^j$ sehen wir, daß

$$\operatorname{span}(\mathcal{G}_t) = E'(\lambda_t)$$

gilt. Speziell ist die Sume der Zeilenzahlen $\nu_t^1, \dots, \nu_t^{k_t}$ der Jordanblöcke $J_t^1, \dots, J_t^{k_t}$ gerade die Dimension von $E'(\lambda_t)$ und

$$T_t := (L - \lambda_t \operatorname{id}_V)|E'(\lambda_t)$$

ist nilpotent vom Grad $g(T_t) = \max\{\nu_t^1, \ldots, \nu_t^{k_t}\}$. Aus den Zahlen $R_{l,t} := \operatorname{rg}(T_t^l)$ lassen sich nach (14.17) die Anzahlen

$$N_{l,t} := \#\{i | 1 \leq i \leq k_t, \nu_t^i = l\}$$

der Jordanblöcke I_t^i zum Eigenwert λ_t berechnen, die Zeilenzahl $\nu_t^i = l$ haben:

$$(+) \qquad N_{l,t} = R_{l+1,t} - 2R_{l,t} + R_{l-1,t}.$$

Wir erhalten also:

(14.18) Satz (Eindeutigkeit der Jordanschen Normalform). *Es sei $L \in \operatorname{End}(V)$ und das charakteristische Polynom P_L von L zerfalle. Dann ist die Jordansche Normalform $\tilde{A} \in K^{n\times n}$ von L (bis auf die Reihenfolge der Jordanblöcke) eindeutig durch L bestimmt. Genauer: Sind $\lambda_1, \ldots, \lambda_s$ die verschiedenen Eigenwerte von L, so ist die Anzahl $N_{l,t}$ der Jordanblöcke in $\tilde{A}$ der Zeilenzahl $l \in \mathbb{N}_{>0}$ zum Eigenwert λ_t, $1 \leq t \leq s$, gegeben durch (+).*

Abgesehen von der Einsicht in die möglichen Typen von linearen Abbildungen, die die Sätze (14.14) und (14.15) vermitteln, ist die Jordansche Normalform auch von praktischem Nutzen. Es kommt z.B. oft vor, daß hohe Potenzen A^m einer Matrix $A \in K^{n\times n}$ berechnet werden sollen Für allgemeine Matrizen A ist das mit sehr großem Rechenaufwand verbunden. Kann man jedoch ein $B \in \mathrm{GL}_n(K)$ finden, so daß $A = B^{-1}JB$ gilt und J Jordansche Normalform hat, so gilt $A^m = B^{-1}J^mB$, wobei J^m sehr leicht zu berechnen ist. Eine damit zusammenhängende theoretische Erkenntnis ist der

(14.19) Satz (von Cayley-Hamilton). *Setzt man $L \in \operatorname{End}(V)$ in sein charakteristisches Polynom P_L ein, so erhält man $P_L(L) = 0 \in \operatorname{End}(V)$.*

Vorbemerkung: Wenn L diagonalisierbar ist, so ist (14.19) leicht einzusehen, Übung!

Bew.: Wir führen den Beweis nur für den Fall durch, daß P_L zerfällt. Wie in Bem. 1 nach (14.16) gesagt wurde, ist das keine wirkliche Einschränkung. Wir verwenden (14.14). Offenbar genügt es zu zeigen, daß für jeden EW λ_t von L gilt:

$$P_L(L) \mid E'(\lambda_t) = 0.$$

Wegen (14.9) gilt für $n_t := \dim E'(\lambda_t)$:

$$(*) \qquad (L - \lambda_t \operatorname{id}_V \mid E'(\lambda_t))^{n_t} = 0.$$

Andererseits enthält P_L den Faktor $(x-\lambda_t)^{n_t}$. Deshalb existiert ein $Q \in \mathrm{End}(V)$ mit

$$(**) \qquad P_L(L) = Q \circ (L - \lambda_t \,\mathrm{id}_V)^{n_t}.$$

Aus $(*)$ und $(**)$ folgt $P_L(L) \mid E'(\lambda_t) = 0$, wie behauptet.

Eine wichtige Anwendung findet die Jordansche Normalform bei der expliziten Lösung von linearen Differentialgleichungssystemen mit konstanten Koeffizienten.

Problemstellung. Gegeben ist eine quadratische Matrix $A \in \mathbb{R}^{n\times n}$. Gesucht sind alle Vektorfunktionen

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \mathbb{R} \to \mathbb{R}^n \simeq \mathbb{R}^{n\times 1}$$

so daß für alle $t \in \mathbb{R}$ gilt:

$$(*) \qquad x'(t) = Ax(t).$$

Dabei ist $x'(t) := \begin{pmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{pmatrix}$. Ist $A = (a_{ij})_{1\le i,j\le n}$, so schreibt sich $(*)$ als

$$x_i'(t) = \sum_{j=1}^{n} a_{ij}x_j(t) \text{ für } 1 \le i \le n \text{ (und alle } t \in \mathbb{R}).$$

Aus der Theorie der gewöhnlichen Differentialgleichungen (vgl. Analysis II) folgt: *für alle* $x_0 \in \mathbb{R}^n$ *existiert genau eine Lösung* $x : \mathbb{R} \to \mathbb{R}^n$ *von* $(*)$, *für die* $x(0) = x_0$ *gilt.* Man nennt dieses x die Lösung von $(*)$ zum Anfangswert x_0. Wir werden sehen, wie man diese Lösung explizit angeben kann, und damit auch diese Existenz - und Eindeutigkeitsaussage beweisen.

Bsp.: “Entkoppelte Systeme”.
Ist A eine Diagonalmatrix mit den Diagonalelementen $\lambda_1, \dots, \lambda_n$, so ist $x : \mathbb{R} \to \mathbb{R}^n$ genau dann Lösung von $(*)$, wenn für alle $t \in \mathbb{R}$ und alle $i \in \{1, \dots, n\}$ gilt:

$$x_i'(t) = \lambda_i x_i(t).$$

Jede dieser n Gleichungen kann mit Schulkenntnissen gelöst werden:

$$x_i(t) = x_i(0)e^{\lambda_i t}$$

und

$$x(t) = \begin{pmatrix} x_1(0)e^{\lambda_1 t} \\ \vdots \\ x_n(0)e^{\lambda_n t} \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} \begin{pmatrix} x_1(0) \\ \vdots \\ x_n(0). \end{pmatrix}$$

Bem.: Man kann für beliebige $A \in \mathbb{R}^{n\times n}$ zeigen, daß die Reihe $\sum\limits_{k=0}^{infty} \frac{A^k}{k!}$ in $\mathbb{R}^{n\times n}$ konvergiert, und man bezeichnet ihren Limes mit $\exp(A)$. Ist A diagonal mit den Diagonalelementen $\lambda_1, \ldots, \lambda_n$, so gilt $\exp(tA) = \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix}$, d.h. unsere Lösung kann als $x(t) = (\exp tA)x_0$ dargestellt werden.

Ist etwa A diagonalisierbar und P so gewählt, daß $P^{-1}AP = B$ eine Diagonalmatrix ist, so kann man die Lösungen y des entkoppelten Systems $y' = By$ explizit angeben (s.o.), und erhält mit $x := P^{-1}y$ alle Lösungen des ursprünglichen “gekoppelten” Systems $x' = Ax$. Nach Folgerung (10.10) ist es (im Prinzip) immer möglich, das System $x' = Ax$ auf diese Weise zu entkoppeln, falls A symmetrisch ist. Ist A nicht (reell) diagonalisierbar, so können wir versuchen, die komplexe Jordansche Normalform von A zu expliziten Löpsung von $(*)$ zu verwenden. Wir müssen dann eine komplexe Versin unserer bisherigen Lösungstheorie entwickeln: Es sei $A \in \mathbb{C}^{n\times n}$ von

$$(**) \qquad z' = Az.$$

Dabei ist $z'(t) = \begin{pmatrix} z_1'(t) \\ \vdots \\ z_n'(t) \end{pmatrix}$, wobei für $z_j(t) = x_j(t) + iy_j(t)$ definiert wird:

$$z_j'(t) = x_j'(t) + iy_j'(t).$$

Wichtige Bemerkung: Sei $A \in \mathbb{R}^{n\times n} \subseteq \mathbb{C}^{n\times n}$, und $z : \mathbb{R} \to \mathbb{C}^n$, $z = x + iy$ mit $x = Re(z) : \mathbb{R} \to \mathbb{R}^n$, $y = \mathrm{Im}(z) : \mathbb{R} \to \mathbb{R}^n$. Dann gilt:

$$z' = Az \Leftrightarrow x' = Ax \text{ und } y' = Ay.$$

Bew.: $z' = Az \Rightarrow X' = Re(z') = Re(Az) \overset{A \in \mathbb{R}^{n\times n}}{=} ARe(z) = Ax$, und analog $y' = Ay$

$x' = Ax$ und $y' = Ay \Rightarrow z' = x' + iy' = Ax + iAy = A(x + iy) = Az$.

Als nächstes wollen wir $(**)$ lösen, falls A eine komplexe Diagonalmatrix ist. Dazu müssen wir die entkoppelten Gleichungen

$$z_j' = \lambda_j z_j$$

mit $\lambda_j \in \mathbb{C}$ und $z_j : \mathbb{R} \to \mathbb{C}$ lösen. Dazu verwenden wir die Exponentialfunktion mit komplexen Argumenten, die entweder durch

$$e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

oder für $a = \mathrm{Re}(z), b = \mathrm{Im}(z)$ durch

$$e^{a+ib} = e^a(\cos b + i \sin b)$$

definiert werden kann. In jedem Fall kann man zeigen, daß die Lösungen von $z_j' = \lambda_j z_j$ durch

$$z_j(t) = z_j(0)e^{\lambda_j t}$$

gegeben sind. Für

$$A = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \in \mathbb{C}^{n\times n}$$

sind die Lösungen $z : \mathbb{R} \to \mathbb{C}^n$ also wieder durch

$$z(t) = \left(e^{\lambda_1 t} z_1(0), \dots, e^{\lambda_n t} z_n(0)\right) = (\exp(tA))z(0)$$

gegeben.

als nächstes betrachten wir $A \in \mathbb{R}^{n\times n}$, $p \in \mathrm{GL}(n, \mathbb{C})$ und $B := P^{-1}AP \in \mathbb{C}^{n\times n}$ und überlegen un s, wie wir aus den Lösungen $z : \mathbb{R} \to \mathbb{C}^n$ von $z' = Bz$ die Lösungen $x : \mathbb{R} \to \mathbb{R}^n$ von $x' = Ax$ gewinnen können. Es gilt: Ist $x_0 \in \mathbb{R}^n$ und $z : \mathbb{R} \to \mathbb{C}^n$ die Lösung von $z' = Bz$ mit $z(0) = P^{-1}x_0 \in \mathbb{C}^n$, so ist

$$x := Pz$$

die Lösung von $x' = Ax$ mit $x(0) = x_0$. Insbesondere gilt automatisch $x(t) \in \mathbb{R}^n$ für alle $t \in \mathbb{R}$ (, obwohl $P \in \mathrm{GL}(n, \mathbb{C})$ und $z(t) \in \mathbb{C}^n$).

Bew.: $x' = (Pz)' = Pz' = P(P^{-1}AP)z = A(Pz) = Ax$ und $x(0) = Pz(0) = PP^{-1}x_0 = x_0$. Es gilt $x = Pz : \mathbb{R} \to \mathbb{R}^n \subseteq \mathbb{C}^n$, da der Imaginärteil $y(t) := \Im(Pz(t))$ ebenfalls die Gleichung

$$y' = Ay$$

löst, und zar mit Anfangswert $y(0) = \Im Pz(0) = \Im x_0 = 0$. Aus der Eindeutigkeit der Lösung zu gegebenem Anfangswert folgt nun, daß $y = 0$ gilt, d.h. tatsächlich ist $Pz : \mathbb{R} \to \mathbb{R}^n$.

Bsp.: Ist $A \in \mathbb{R}^{n\times n}$ komplex diagonalisierbar, d.h. existiert $P \in \mathrm{GL}(n, \mathbb{C})$, so daß

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \in \mathbb{C}^{n\times n}$$

gilt, so ist die Lösung $x : \mathbb{R} \to \mathbb{R}^n$ von $x' = Ax$ mit Anfangswert $x(0) \in \mathbb{R}^n$ gegeben durch

$$x(t) = Pz(t) = P \begin{pmatrix} e^{\lambda_1 t} z_1(0) \\ \vdots \\ e^{\lambda_n t} z_n(0) \end{pmatrix},$$

wobei $z(0) = \begin{pmatrix} z_1(0) \\ \vdots \\ z_n(0) \end{pmatrix} := P^{-1}x(0) \in \mathbb{C}^n$.

Als nächstes betrachten wir den Fall, daß $A \in \mathbb{C}^{n\times n}$ als Jordansche Normalform einen einzigen Jordanblock hat, d.h. daß ein $P \in \mathrm{GL}(n, \mathbb{C})$ existiert, so daß

$$P^{-1}AP = J \text{ mit } J = \begin{pmatrix} \lambda & 0 & \dots & \dots & 0 \\ 1 & \ddots & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & \lambda \end{pmatrix} \text{ für ein } \lambda \in \mathbb{C}$$

gilt. Ist $p = \sum_{k=0}^{m} a_k x^k \in \mathbb{C}[c]$ ein komplexes Polynom, so nennen wir $p' :=$

$\sum\limits_{k=1}^{m} ka_k x^{k-1} \in \mathbb{C}[x]$ die Ableitung von p, und wir setzen induktiv für $k \in \mathbb{N}$:

$$p^{(0)} := p \text{ und } p^{(k)} := \left(p^{(k-1)}\right)'.$$

(14.21) Lemma. Ist $p \in \mathbb{C}[x]$ ein Polynom vom Grad $\leq n-1$, so ist $z : \mathbb{R} \to \mathbb{C}^n$,

$$(+) \qquad z(t) := e^{\lambda t}\left(p^{n-1}(t), p^{n-2}(t), \ldots, p'(t), p(t)\right),$$

eine Lösung von $z' = Jz$. Umgekehrt existiert zu jeder Lösung z von $z' = Jz$ ein Polynom $p \in \mathbb{C}[x]$ vom Grad $\leq n-1$, so daß (t) gilt.

Bem.: Will man $z' = Jz$ mit dem Anfangswert $z(0) = c \in \mathbb{C}^n$ lösen, so muß man als Polynom

$$p = \sum_{k=0}^{n-1} \frac{c_{n-k}}{k!} x^k \in \mathbb{C}[x]$$

wählen, da dann

$$P^{(k)}(0) = c_{n-k}$$

gilt, d.h.

$$z(0) = \left(p^{(n-1)}(0), \ldots, p(0)\right) = (c_1, c_2, \ldots, c_n) = c.$$

Bew.: Differentiation von (+) ergibt

$$z'(t) = \lambda z(t) + e^{\lambda t}(0, p^{(n-1)}(t), \ldots, p'(t)) = Jz(t).$$

Ist umgekehrt z Lösung von $z' = Jz$, so setzen wir $q(t) := e^{-\lambda t} z(t)$. Wir schreiben $J = T + \lambda E_n$ mit der nilpotenten Matrix T vom Nilpotenzgrad $n-1$. Dann gilt

$$\begin{aligned} q * (t) &= -\lambda e^{-\lambda t} z(t) + e^{-\lambda t} z'(t) = -\lambda q(t) + e^{-\lambda t}(T + \lambda E_n) z(t) \\ &= e^{-\lambda t} T z(t) = T q(t). \end{aligned}$$

Induktiv folgt $q^{(n)}(t) = T^n q(t) = 0$. Also besteht $q = (q_1, \ldots, q_n)$ aus Polynomen von Grad $\leq n-1$ und aus $q' = Tq$ folgt $q_1' = 0$, $q_2' = q_1, \ldots, q_n' = q_{n-1}$, d.h. $(q_1, \ldots, q_n) = \left(q_n^{(n-1)}, q_n^{(n-2)}, \ldots, q_n', q_n\right)$.

Bem.: Wenn wir benutzen, daß es zu gegebenem Anfangswert $c \in \mathbb{C}^n$ genau eine Lösung z von $z' = Jz$ mit $z(0) = c$ gibt, so ist der zweite Teil des vorangehenden Beweises überflüssig. Denn die vorangehende Bemerkung zeigt, daß es zu jedem $c \in \mathbb{C}^n$ eine Lösung z vom Typ (+) gibt mit $z(0) = C$, also ist jede Lösung vom Typ (+).

Im Fall eines $A \in \mathbb{C}^{n\times n}$ mit "allgemeiner" Jordanscher Normalform

$$\tilde{A} = \begin{pmatrix} J_1^1 & & 0 \\ & \ddots & \\ 0 & & J_s^{k_s} \end{pmatrix}$$

setzt sich jede Lösung z von $z' = \tilde{A}z$ aus Lösungen vom Typ (+) für die einzelnen Jordanblöcke J_t^i, $1 \leq t \leq s$, $1 \leq i \leq k_t$, zusammen.

Auch in diesem allgemeinen Fall kann man zeigen, daß die Lösungen z von $z' = \tilde{A}z$ durch

$$z(t) = \exp(t\tilde{A})z(0)$$

gegeben sind, da

$$\exp(t\tilde{A}) := \sum_{k=0}^{\infty} \frac{((t\tilde{A})^k)}{k!}$$

leicht zu berechnen ist. Ist $P \in \mathrm{GL}(n,\mathbb{C})$ und $P^{-1}AP = \tilde{A}$, so gilt $P^{-1}\exp(tA)P = \exp(t\tilde{A})$, so daß die Lösungen $w : \mathbb{R} \to \mathbb{C}^n$ von $w' = Aw$ durch

$$w(t) = (P^{-1}\exp(t\tilde{A}))(Pw(0)) = (\exp(tA))w(0)$$

gegeben sind.

Bsp.: Wir betrachten das lineare Differentialgleichungssystem

$$(*) \qquad x' = Ax$$

für Vektorfunktionen $x : \mathbb{R} \to \mathbb{R}^3$ mit der Matrix $A = \left(\begin{smallmatrix} 3 & 2 & -3 \\ 4 & 10 & -12 \\ 3 & 6 & -7 \end{smallmatrix}\right)$. Explizit bedeutet $(*)$ dann:

$$\begin{aligned} x_1'(t) &= 3x_1(t) + 2x_2(t) - 3x_3(t) \\ x_2'(t) &= 4x_1(t) + 10x_2(t) - 12x_3(t) \\ x_3'(t) &= 3x_1(t) + 6x_2(t) - 7x_3(t) \end{aligned}$$

Wir können die Jordansche Normalform $\tilde{A}$ von A berechnen und eine Matrix $B \in \mathrm{GL}(n,\mathbb{C})$, so daß $B^{-1}AB = \tilde{A}$ gilt (Übung!). Es ergibt sich

$$\tilde{A} = \left(\begin{smallmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{smallmatrix}\right)$$

und (nicht eindeutig!)

$$B^{-1} = \begin{pmatrix} 1 & 2 & -3 \\ 0 & \frac{1}{4} & 0 \\ 0 & -\frac{3}{4} & 1 \end{pmatrix}.$$

Zu dem Jordanblock $\left(\begin{smallmatrix} 2 & 0 \\ 1 & 2 \end{smallmatrix}\right)$ gehören die Lösungen

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = e^{2t} \left(\begin{smallmatrix} c_1 \\ c_2 + c_1 t \end{smallmatrix}\right)$$

und zum Eigenvektor e_3 von $J : y_3(t) = c_3 e^{2t}$.

Die Lösungen von $y' = J <$ sind dann $q(t) = e^{2t} \begin{pmatrix} c_1 \\ c_2 + c_1 t \\ c_3 \end{pmatrix}$ mit beliebigen $c_1, c_2, c_3 \in \mathbb{R}$.

Die Lösungen x von $(*)$ sind $x = B^{-1} y$ mit obigem B^{-1}.