1 Mengen und Abbildungen

Wir starten mit einigen einführenden Definitionen und Ergebnissen aus der Theorie der Mengen und Abbildungen, die nicht nur Grundlage der Linearen Algebra sondern der gesamten Mathematik sind.
Unsere Darstellung gründet auf den von G. Cantor geprägten (sog. naiven) Mengenbegriff.

> "Eine *Menge* M ist eine Zusammenfassung von unterscheidbaren Objekten"

Ein solches Objekt x heißt *Element* der Menge M (Schreibweise: $x \in M$; ist x nicht Element von M, so schreiben wir $x \notin M$).

Definition 1.1 Es seien A, B Mengen.

1. A heißt *Teilmenge* von B, falls für jedes $x \in A$ auch $x \in B$ gilt. (Schreibweise: $A \subset B$).
2. A und B heißen *gleich* (Schreibweise $A = B$), falls $A \subset B$ und $B \subset A$.
3. Die Menge $B \setminus A := \{x : x \in B \,\text{und}\, x \notin A\}$ heißt *Differenz* von B und A. Ist $A \subset B$, so heißt $A^c := C_B(A) := B \setminus A$ *Komplement* von A (bzgl. B).
4. Die Menge ohne Elemente heißt *leere Menge* (Schreibweise: $\emptyset$).
5. Die Menge $A \cup B := \{x : x \in A \,\text{oder}\, x \in B\}$ heißt *Vereinigung* von A und B.
6. Die Menge $A \cap B := \{x : x \in A \,\text{und}\, x \in B\}$ heißt *Schnitt* von A und B.

Definition 1.2 Es seien A und B Mengen. Dann heißt

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

also die Menge der geordneten Paare von Elementen aus A und B, das *Produkt* oder die *Produktmenge* von A und B.

Beispiel 1.3 Ist $A = \{1, 2\}$ und $B = \{3\}$, so ist

$$A \times B = \{(1, 3), (2, 3)\} .$$

Man beachte, dass $A \times B$ nicht mit $B \times A$ übereinstimmt, da $(a, b) = (\tilde{a}, \tilde{b})$ genau dann gilt, wenn $a = \tilde{a}$ und $b = \tilde{b}$.

Satz 1.4 *Es seien A_1, A_2, A_3 Mengen. Dann gilt*

1. $A_1 \cup A_2 = A_2 \cup A_1$,
   $A_1 \cap A_2 = A_2 \cap A_1$.

2. $A_1 \cup (A_2 \cup A_3) = (A_1 \cup A_2) \cup A_3$,
   $A_1 \cap (A_2 \cap A_3) = (A_1 \cap A_2) \cap A_3$.

   Wir schreiben deshalb auch kurz $A_1 \cup A_2 \cup A_3$.

3. $A_1 \cap (A_2 \cup A_3) = (A_1 \cap A_2) \cup (A_1 \cap A_3)$,
   $A_1 \cup (A_2 \cap A_3) = (A_1 \cup A_2) \cap (A_1 \cup A_3)$.

Beweis.
1. und 2. folgen sofort aus Definition 1.2. Wir beweisen die erste Aussage von 3.
"$\subset$": Dazu sei

$$x \in A_1 \cap (A_2 \cup A_3) \, .$$

Dann ist $x \in A_1$ und $x \in A_2 \cup A_3$.

1. Fall: $x \in A_1$ und $x \in A_2$. Dann ist $x \in A_1 \cap A_2$, also auch $x \in (A_1 \cap A_2) \cup (A_1 \cap A_3)$.
2. Fall: $x \in A_1$ und $x \in A_3$. Dann ist $x \in A_1 \cap A_3$, also auch $x \in (A_1 \cap A_2) \cup (A_1 \cap A_3)$.

Also ist in jedem Fall $x \in (A_1 \cap A_2) \cap (A_1 \cap A_3)$.
Damit gilt $A_1 \cap (A_2 \cup A_3) \subset (A_1 \cap A_2) \cup (A_1 \cap A_3)$.
"$\supset$": Umgekehrt sei $x \in (A_1 \cap A_2) \cup (A_1 \cap A_3)$. Dann ist $x \in A_1 \cap A_2$ oder $x \in A_1 \cap A_3$. In beiden Fällen ist dann $x \in A_1 \cap (A_2 \cup A_3)$. Also folgt $(A_1 \cap A_2) \cup (A_1 \cap A_3) \subset A_1 \cap (A_2 \cup A_3)$.
Die zweite Aussage von 3. als [Ü]. □

Satz 1.5 (Regeln von de Morgan) *Es seien A_1, A_2 Mengen, und es sei B eine Menge mit $A_1 \subset B$ und $A_2 \subset B$. Dann gilt*

1. $C_B(A_1 \cup A_2) = C_B(A_1) \cap C_B(A_2)$.

2. $C_B(A_1 \cap A_2) = C_B(A_1) \cup C_B(A_2)$.

Beweis.
1. "$\subset$": Es sei $x \in C_B(A_1 \cup A_2)$. Dann ist $x \in B$ und $x \notin A_1 \cup A_2$, also $x \in B$ und $x \notin A_1$ sowie $x \notin A_2$. Damit ist $x \in C_B(A_1)$ und $x \in C_B(A_2)$, d. h. $x \in C_B(A_1) \cap C_B(A_2)$.

"$\supset$": Es sei $x \in C_B(A_1) \cap C_B(A_2)$. Dann ist $x \in C_B(A_1)$ und $x \in C_B(A_2)$. Also ist $x \in B$ und $x \notin A_1$ sowie $x \notin A_2$. Dann ist $x \in B$ und $x \notin A_1 \cup A_2$, also $x \in C_B(A_1 \cap A_2)$.

2. [Ü]. □

Definition 1.6 Es seien X und Y Mengen. Eine Teilmenge R von $X \times Y$ heißt *Relation* (zwischen X und Y). Ist speziell $X = Y$, so heißt R *Relation in* X. Eine Relation R zwischen X und Y heißt *Abbildung (von X nach Y)* bzw. *Funktion (von X nach Y)* falls gilt:

a) Für alle $x \in X$ existiert ein $y \in Y$ mit $(x, y) \in R$.
und
b) Sind $(x, y) \in R$ und $(x, \tilde{y}) \in R$ so gilt $y = \tilde{y}$.

Bemerkung und Definition 1.7 Ist R eine Abbildung von X und Y, so ist jedem Wert $x \in X$ genau ein Wert $f(x)$ mit $(x, f(x)) \in R$ zugeordnet. Wir identifizieren R dann auch mit dieser Zuordnungsvorschrift f und schreiben $f : X \to Y$ oder $x \mapsto f(x)$. Weiter heißt X der *Definitionsbereich (von f)* und

$$W(f) := \{f(x) : x \in X\} = \{y \in Y : \exists\, x \in X \text{ mit } y = f(x)\}$$

Wertebereich (von f). Ferner setzen wir für $B \subset Y$

$$f^{-1}(B) := \{x \in X : f(x) \in B\}$$

($f^{-1}(B)$ heißt *Urbild von B unter f*) und für $A \subset X$

$$f(A) := \{f(x) : x \in A\} = \{y \in Y : \exists\, x \in A \text{ mit } y = f(x)\}$$

($f(A)$ heißt *Bild von A unter f*).
Zwei Abbildungen $f_1 : X_1 \to Y$ und $f_2 : X_2 \to Y$ heißen *gleich* falls $X_1 = X_2$ und $f_1(x) = f_2(x)$ für alle $x \in X_1 (= X_2)$ gilt.
Ist $f : X \to Y$ und ist $X_0 \subset X$, so heißt $f_{|X_0} : X_0 \to Y$, definiert durch $f_{|X_0}(x) := f(x)$ für alle $x \in X_0$, *Einschränkung* von f auf X_0.

Satz 1.8 *Es seien X, Y Mengen und $f : X \to Y$. Dann gilt für $A_1, A_2 \subset X$ und $B_1, B_2 \subset Y$*

1. $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$,
2. $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$,
3. $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$,

4. $f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2)$.

Beweis.
1. "$\subset$": Es sei $y \in f(A_1 \cup A_2)$. Dann existiert ein $x \in A_1 \cup A_2$ mit $f(x) = y$. Ist $x \in A_1$, so ist $y = f(x) \in f(A_1) \subset f(A_1) \cup f(A_2)$. Entsprechend ist $y \in f(A_2) \subset f(A_1) \cup f(A_2)$ im Falle $x \in A_2$.
"$\supset$": Nach Definition gilt $f(A_1) \subset f(A_1 \cup A_2)$ und $f(A_2) \subset f(A_1 \cup A_2)$ also $f(A_1) \cup f(A_2) \subset f(A_1 \cup A_2)$.

2. "$\subset$": Es sei $x \in f^{-1}(B_1 \cup B_2)$. Dann ist $f(x) \in B_1 \cup B_2$.
Ist $f(x) \in B_1$, so ist $x \in f^{-1}(B_1)$ also auch $x \in f^{-1}(B_1) \cup f^{-1}(B_2)$. Entsprechend ist $x \in f^{-1}(B_2) \subset f^{-1}(B_1) \cup f^{-1}(B_2)$, falls $f(x) \in B_2$.
"$\supset$": Nach Definition gilt

$$f^{-1}(B_1) \subset f^{-1}(B_1 \cup B_2) \text{ und } f^{-1}(B_2) \subset f^{-1}(B_1 \cup B_2) \ ,$$

also $f^{-1}(B_1) \cup f^{-1}(B_2) \subset f^{-1}(B_1 \cup B_2)$.

3. Es sei $y \in f(A_1 \cap A_2)$. Dann existiert ein $x \in A_1 \cap A_2$ mit $f(x) = y$. Da $x \in A_1$ und $x \in A_2$ ist, folgt $y \in f(A_1) \cap f(A_2)$.

4. [Ü] □

Beispiel 1.9 Es seien

$$\mathbb{N} := \{1, 2, 3, \dots\} = \{\text{natürliche Zahlen}\}$$

und

$$\mathbb{Z} := \{\text{ganze Zahlen}\} = \{0, \pm 1, \pm 2, \dots\} \ .$$

Weiter seien $X = Y = \mathbb{Z}$ und $f : \mathbb{Z} \to \mathbb{Z}$ definiert durch

$$f(x) := \begin{cases} x, & \text{falls} \quad x \geq 0 \\ -x, & \text{falls} \quad x < 0 \end{cases} \ .$$

Dann gilt $W(f) = \mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Weiter ist etwa

$$f^{-1}(\{1, \dots, n\}) = f^{-1}(\{1, \dots, n\} \cup \{-1, -2, -3, \dots\}) = \{-n, \dots, -1, 1, \dots, n\}$$

und $f^{-1}(\{-1, -2, -3, \dots\}) = \emptyset$ sowie $f(\mathbb{N}) = \mathbb{N} = f(\mathbb{Z} - \{0\})$. Ist $\tilde{f} : \mathbb{N}_0 \to \mathbb{N}_0$ definiert durch

$$\tilde{f}(x) := x \qquad (x \in \mathbb{N}_0) \ ,$$

so ist zwar $\tilde{f}(x) = f(x)$ für alle $x \in \mathbb{N}_0$, aber $\tilde{f} \neq f$. Es gilt aber $f_{|\mathbb{N}_0} = \tilde{f}$.

Definition 1.10 Es seien X, Y Mengen. Eine Abbildung $f : X \to Y$ heißt

1. *surjetiv* (oder Abbildung von X *auf* Y), falls $W(f) = Y$ ist,
2. *injektiv* (oder *eineindeutige* Abbildung), falls für alle $y \in W(f)$ die Menge $f^{-1}(\{y\})$ einelementig ist (d. h. sind $x_1, x_2 \in X$ mit $f(x_1) = f(x_2)$, so ist $x_1 = x_2$),
3. *bijektiv*, falls f injektiv und surjektiv ist.

Beispiel 1.11 Es seien f und $\tilde{f}$ wie im B 1.9 Dann ist f weder surjektiv noch injektiv (es gilt $f^{-1}(\{(n\}) = \{n, -n\}$ für alle $n \in \mathbb{N}$); $\tilde{f}$ ist bijektiv.

Definition 1.12 Es seien X, Y, Z Mengen und $f : X \to Y$ sowie $g : Y \to Z$ Abbildungen. Dann heißt $g \circ f : X \to Z$, definiert durch

$$(g \circ f)(x) := g(f(x)) \qquad (x \in X)$$

Verknüpfung von g und f (oder *Hintereinanderausführung von f und g*).

Satz 1.13 *Es seien X, Y, Z, U Mengen und $f : X \to Y, g : Y \to Z$ und $h : Z \to U$ Abbildungen. Dann gilt*

$$h \circ (g \circ f) = (h \circ g) \circ f \ .$$

Beweis.
Es gilt $h \circ (g \circ f) : X \to U$ und $(h \circ g) \circ f : X \to U$. Weiter gilt für $x \in X$

$$\begin{aligned} (h \circ (g \circ f))(x) &= h((g \circ f)(x)) = h(g(f(x))) = (h \circ g)(f(x)) = \\ &= ((h \circ g) \circ f)(x) \ . \end{aligned}$$

Definition 1.14 Es sei $I \neq \emptyset$ eine Menge, und es seien A_α Mengen für alle $\alpha \in I$. (I nennt man dann "Indexmenge".) Dann heißt

$$\bigcup_{\alpha \in I} A_\alpha := \{x : x \in A_\alpha \text{ für ein } \alpha \in I\}$$

Vereinigung der Mengen A_α (über $\alpha \in I$).
Weiter heißt

$$\bigcap_{\alpha \in I} A_\alpha := \{x : x \in A_\alpha \text{ für alle } \alpha \in I\}$$

Durchschnitt der Mengen A_α (über $\alpha \in I$).
Ist speziell I endlich, so kann man ohne Einschränkung $I = \{1, \dots, n\}$ annehmen. Wir schreiben dann auch

$$A_1 \cup \ldots \cup A_n = \bigcup_{j=1}^{n} A_j := \bigcup_{j \in \{1,\ldots,n\}} A_j$$

und

$$A_1 \cap \ldots \cap A_n = \bigcap_{j=1}^{n} A_j := \bigcap_{j \in \{1,\ldots,n\}} A_j \ .$$

Oft betrachtet man auch Vereinigungen und Durchschnitte von Mengensystemen, die nicht indiziert sind: Ist $\mathcal{F}$ ein System von Mengen (d. h. eine Menge von Mengen), so setzt man

$$\bigcup_{A \in \mathcal{F}} A := \bigcup \{A : A \in \mathcal{F}\} := \{x : x \in A \text{ für ein } A \in \mathcal{F}\}$$

und

$$\bigcap_{A \in \mathcal{F}} A := \bigcap \{A : A \in \mathcal{F}\} := \{x : x \in A \text{ für alle } A \in \mathcal{F}\} \ .$$

Beispiel 1.15 Es sei $I = \mathbb{N}$ und

$$A_n := \{k/n : k \in \mathbb{Z}\} \qquad (n \in \mathbb{N}) \ .$$

Dann ist

$$\bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} \{k/n : k \in \mathbb{Z}\} = \mathbb{Q}$$

wobei $\mathbb{Q} := \{\text{rationale Zahlen}\}$ und

$$\bigcap_{n \in \mathbb{N}} A_n = A_1 = \mathbb{Z} \ .$$

Definition 1.16 Es sei I eine Menge und es seien A_α Mengen für alle $\alpha \in I$. Dann heißt

$$\prod_{\alpha \in I} A_\alpha := \{f : I \to \bigcup_{\alpha \in I} A_\alpha : f(\alpha) \in A_\alpha \text{ für alle } \alpha \in I\}$$

Produkt der Mengen A_α. Ist $f \in \prod_{\alpha \in I} A_\alpha$, so schreiben wir auch $f_\alpha := f(\alpha)$ und $(f_\alpha)_{\alpha \in I}$ für f. Gilt $A_\alpha = A$ für alle $\alpha \in I$, so setzt man $A^I := \prod_{\alpha \in I} A$.

Ist I endlich, so kann man ohne Einschränkung $I = \{1, \dots, n\}$ betrachten. Dann schreiben wir

$$(a_1, \dots, a_n) := (a_j)_{j \in \{1,\dots,n\}}$$

und

$$\begin{aligned} A_1 \times \dots \times A_n &:= \prod_{j=1}^{n} A_j := \prod_{j \in \{1,\dots,n\}} A_j \\ &= \{(a_1, \dots, a_n) : a_j \in A_j \text{ für } j = 1, \dots, n\} . \end{aligned}$$

Ein $(a_1, \dots, a_n) \in \prod_{j=1}^{n} A_j$ heißt *n-Tupel*.
Ist $A_j = A$ für $j = 1, \dots, n$, so setzen wir

$$A^n := A^{\{1,\dots,n\}} = \prod_{j=1}^{n} A .$$

Beispiel 1.17 Es sei $A_j = \mathbb{N}$ für $j = 1, \dots, n$. Dann ist

$$\mathbb{N}^n = \prod_{j=1}^{n} \mathbb{N} = \{(a_1, \dots, a_n) : a_j \in \mathbb{N} \text{ für } j = 1, \dots, n\}$$

die Menge aller *n-Tupel* aus natürlichen Zahlen.

Bemerkung und Definition 1.18 Sind X, Y Mengen und ist $f : X \to Y$ injektiv, so ist $f : X \to W(f)$ bijektiv. Wir definieren

$$f^{-1}(y) := x \qquad (y \in W(f))$$

wobei $y = f(x)$. Die Abbildung $f^{-1} : W(f) \to X$ heißt *Umkehrabbildung von* f. Es gilt dann $f^{-1} \circ f : X \to X$ und $(f^{-1} \circ f)(x) = x$ für alle $x \in X$, d. h. $f^{-1} \circ f = \mathrm{id}_X$, wobei $\mathrm{id}_X : X \to X$, definiert durch

$$\mathrm{id}_X(x) := x \qquad (x \in X) ,$$

die sog. *identische Abbildung auf* X bezeichnet. Genauso gilt $f \circ f^{-1} = \mathrm{id}_{W(f)}$. Außerdem ist $f^{-1} : W(f) \to X$ bijektiv.

2 Gruppen und Körper

In diesem Abschnitt beschäftigen wir uns mit Grundbegriffen der Algebra. Dies hat zunächst einmal das Ziel, die "Bühne zu bereiten" für die Einführung des zentralen Begriffs der Linearen Algebra (nämlich des linearen Raumes). Später werden wir nochmal auf die hier dargestellten Dinge zurückkommen.

Definition 2.1 Es sei $G \neq \emptyset$ eine Menge, und es sei $\circ : G \times G \to G$ eine Abbildung. Dann heißt $(G, \circ)$ eine *Gruppe*, falls gilt

(G.1) (Assoziativgesetz):
Für alle $a, b, c \in G$ ist

$$a \circ (b \circ c) = (a \circ b) \circ c \; .$$

(G.2) (Existenz eine (Rechts-) Einselements bzw. neutralen Elements):
Es existiert eine $e \in G$ mit

$$a \circ e = a \qquad (a \in G) \; .$$

(G.3) (Existenz eines (rechts-) inversen Elements):
Für alle $a \in G$ existiert ein $b \in G$ mit

$$a \circ b = e \; .$$

Ferner heißt $(G, \circ)$ *abelsche* (oder *kommutative*) Gruppe, falls zudem gilt

(G.4) (Kommutativgesetz): Für alle $a, b \in G$ gilt

$$a \circ b = b \circ a \; .$$

Satz 2.2 *Es sei $(G, \circ)$ eine Gruppe. Dann gilt:*

1. *Es existiert nur ein $e \in G$ mit $a \circ e = a$ für alle $a \in G$ und dieses erfüllt dann auch $e \circ a = a$ für alle $a \in G$.*
2. *Zu jedem $a \in G$ existiert nur ein $b \in G$ mit $a \circ b = e$ und dieses b erfüllt auch $b \circ a = e$. Wir setzen*

$$a^{-1} := b \; .$$

Beweis.
α) Wir zeigen zunächst die letzte Aussage. Dazu sei $a \in G$ gegeben. Weiter sei b ein nach (G.3) existierendes rechtsinverses Element zu a (bzgl. e).
Dann gilt

$$b \circ (a \circ b) = b \circ e = b \; .$$

Ist c ein rechtsinverses Element zu b, d. h. $b \circ c = e$, so folgt

$$\begin{aligned} e &= b \circ c = (b \circ (a \circ b)) \circ c \overset{\text{G.1}}{=} (b \circ a) \circ (b \circ c) = \\ &= (b \circ a) \circ e \overset{\text{G.2}}{=} b \circ a \; . \end{aligned}$$

Also ist b auch, "linksinverses Element" zu a.

β) Es sei $a \in G$. Dann gilt mit α)

$$e \circ a = (a \circ b) \circ a \overset{\text{G.1}}{=} a \circ (b \circ a) = a \circ e \overset{\text{G.2}}{=} a \, .$$

Also ist e auch ein "Linkseinselement".

γ) Ist $\tilde{e}$ ein weiteres Rechtseinselement, d. h. ist $a \circ \tilde{e} = a$ für alle $a \in G$, so ist insbesondere $e \circ \tilde{e} = e$. Also folgt

$$e = e \circ \tilde{e} \overset{\beta)}{=} \tilde{e} \, .$$

δ) Es sei nun $\tilde{b} \in G$ ein weiteres rechtsinverses Element zu a (neben b aus Aussage 2.). Dann gilt $a \circ \tilde{b} = e$ und damit

$$b \overset{\text{G.2}}{=} b \circ e = b \circ (a \circ \tilde{b}) \overset{\text{G.1}}{=} (b \circ a) \circ \tilde{b} \overset{\alpha)}{=} e \circ \tilde{b} \overset{\beta)}{=} \tilde{b} \, ,$$

d. h. $b = \tilde{b}$ □.

Beispiel 2.3 Es sei $M \neq \emptyset$ eine Menge, und es sei

$$S(M) := \{f : M \to M, f \text{ bijektiv}\}$$

sowie $\circ$ wie in Definition 1.12 (Hintereinanderausführung). Dann ist $(S(M), \circ)$ eine Gruppe

(Denn: Mit $f, g \in S(M)$ ist auch $f \circ g \in S(M)$ ([Ü]); Axiom (G.1) folgt aus Satz 1.13; das neutrale Element e ist offenbar $e = \mathrm{id}_M$ und (G.3) folgt aus B/D 1.18, da $W(f) = M$ für alle $f \in S(M)$).

Die Funktionen $f \in S(M)$ heißen *Permutationen (von M)* und $(S(M), \circ)$ heißt *Permutationsgruppe (von M)*.

Ist speziell $M = \{1, \ldots, n\}$ für ein $n \in \mathbb{N}$, so heißt

$$S_n := S(\{1, \ldots, n\})$$

symmetrische Gruppe von Grad n. Ein $f \in S_n$ schreibt man oft in der Form

$$f = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ f(1) & f(2) & f(3) & \ldots & f(n) \end{pmatrix}.$$

Man kann leicht zeigen ([Ü]): S_1, S_2 sind abelsch, S_n ist für $n \geq 3$ nicht abelsch; für $n = 3$ betrachte man etwa

$$f = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \text{und} \quad g = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} .$$

Definition 2.4 Gegeben seien eine Menge K mit mindestens zwei Elementen und zwei Abbildungen $+ : K \times K \to K$ ($+$ heißt Addition) und $\cdot : K \times K \to K$ ($\cdot$ heißt Multiplikation). Dann heißt $(K, +, \cdot)$ ein *Körper*, falls gilt

(A) $(K, +)$ ist eine abelsche Gruppe (wobei das neutrale Element mit 0 bezeichnet wird).

(M) $(K\backslash\{0\}, \cdot)$ ist eine abelsche Gruppe (wobei das neutrale Element mit 1 bezeichnet wird).

(D) (Distributivgesetze):
Für alle $a, b, c \in K$ gilt

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c) \quad \text{und} \quad (a + b) \cdot c = (a \cdot c) + (b \cdot c) \; .$$

Für $a, b, c, d \in K$ schreiben wir kurz ab statt $a \cdot b$ und $ab + cd$ statt $(ab) + (cd)$ (Punktrechnung vor Strichrechung).

Ist $a \in K$, so schreiben wir $-a$ für das inverse Element bzgl. $+$. Außerdem setzen wir $b - a := b + (-a)$ für $a, b \in K$ und $b/a := b \cdot a^{-1}$ für $b \in K, a \in K \setminus \{0\}$, wobei a^{-1} das inverse Element von a bzgl. $\cdot$ bezeichnet.

Satz 2.5 *Es sei $(K, +, \cdot)$ ein Körper. Dann gilt*

1. *Für alle $a \in K$ gilt $a \cdot 0 = 0 \cdot a = 0$. (Hieraus folgt insbesondere auch, daß $a(bc) = (ab)c$ und $ab = ba$ sowie $a1 = a$ für alle $a, b, c \in K$ gilt.)*
2. *Sind $a, b \in K$ mit $a \cdot b = 0$, so ist $a = 0$ oder $b = 0$.*
3. *Für alle $a, b \in K$ gilt $-(ab) = (-a)b = a(-b)$ (wir schreiben dann: $-ab$).*

Beweis.
1. Es sei $a \in K$. Dann gilt

$$a0 = a(0 + 0) = a0 + a0$$

und damit

$$0 = a0 - (a0) = (a0 + a0) - (a0) = a0 + (a0 - (a0)) = a0 \; .$$

Entsprechend sieht man, dass $0a = 0$ gilt.
2. Es seien $a, b \in K$ mit $a \cdot b = 0$. Ist $b = 0$, so sind wir fertig. Ist $b \neq 0$, so folgt mit 1.

$$a = a1 = a(b/b) = (ab)/b = 0/b = 0 \; ,$$

also $a = 0$.

3. [Ü] □

Beispiel 2.6 1. Wir betrachten $\mathbb{Q} := \{p/q : p \in \mathbb{Z},\ q \in \mathbb{N}\}$ mit der üblichen Addition $+$ und Multiplikation $\cdot$. Dann ist $(\mathbb{Q}, +, \cdot)$ ein Körper.($\rightarrow$ Analysis)

2. Gleiches gilt für $(\mathbb{R}, +, \cdot)$ wobei $\mathbb{R} = \{\text{reelle Zahlen}\}$. ($\rightarrow$ Analysis)

3. Gleiches gilt für $(\mathbb{C}, +, \cdot)$ wobei $\mathbb{C} = \{\text{komplexe Zahlen}\}$. ($\rightarrow$ Analysis)

4. Es sei $K = \{0, 1\}$ mit folgender Addition und Multiplikation

$+$	0	1
0	0	1
1	1	0

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

(d. h. $0 + 0 = 1 + 1 = 0, 1 + 0 = 0 + 1 = 1, 0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0, 1 \cdot 1 = 1$). Dann ist $(K, +, \cdot)$ ein Körper (der sog. Binärkörper) ([Ü]).

3 Die Definition linearer Räume

Einen der zentralen Punkte der Linearen Algebra stellt die Theorie linearer Gleichungssysteme dar. Bevor wir darauf zu sprechen kommen wollen wir eine geeignete "Sprache" hierfür entwickeln. Ausgangspunkt und Mittelpunkt zugleich ist der Begriff des linearen Raumes (oder Vektorraumes).

Definition 3.1 Es seien $K = (K, +, \cdot)$ ein Körper und $V \neq \emptyset$ eine Menge. Ferner seien zwei Abbildungen $+ : V \times V \to V$ (genannt *Addition*) und $\cdot : K \times V \to V$ (genannt *Skalarmultiplikation*) gegeben. Dann heißt $V = (V, +, \cdot)$ ein *linearer Raum (über K)* bzw. *Vektorraum über K* bzw. *K-Vektorraum*, falls gilt:

(AV) $(V, +)$ ist eine abelsche Gruppe. (Das neutrale Element wird dabei wieder mit 0 (oder 0_V) und das inverse Element von $x \in V$ wieder mit $-x$ bezeichnet.)

(M1) Für alle $\lambda, \mu \in K$ und alle $x \in V$ gilt

$$\lambda \cdot (\mu \cdot x) = (\lambda \cdot \mu) \cdot x \; .$$

(M2) Ist 1 das neutrale Element von $(K \setminus \{0\}, \cdot)$, so gilt für alle $x \in V$

$$1 \cdot x = x \; .$$

(D1) Für alle $\lambda \in K$ und alle $x, y \in V$ gilt

$$\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y \; .$$

(D2) Für alle $\lambda, \mu \in K$ und alle $x \in V$ gilt

$$(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x \; .$$

Die Elemente von V heißen hierbei *Vektoren* und die Elemente aus K heißen *Skalare*. Man beachte, dass Addition und Multiplikation sowohl in K als auch in V mit "+" und "$\cdot$" bezeichnet werden, obwohl es sich um verschiedene Operationen handelt! Außerdem schreiben wir wieder kurz λx statt $\lambda \cdot x$.
Wir stellen zunächst einige Rechenregeln zusammen, die sich aus D 3.1 ergeben.

Satz 3.2 *Es sei V ein linearer Raum über K. Dann gilt*

1. *Für alle $x \in V$, $\lambda \in K$ ist $0x = 0$ und $\lambda 0 = 0$.*
2. *Sind $\lambda \in K$ und $x \in V$ so, dass $\lambda \cdot x = 0$, so folgt $\lambda = 0$ oder $x = 0$.*
3. *Für alle $\lambda \in K$ und $x \in V$ ist $-(\lambda x) = \lambda(-x) = (-\lambda)x (=: -\lambda x)$.*

Beweis.
Es seien $x \in V$ und $\lambda \in K$ gegeben.
1. Nach (D2) und (D1) gilt

$$0x + 0x = (0 + 0)x = 0x$$

und

$$\lambda 0 + \lambda 0 = \lambda(0 + 0) = \lambda 0 \, .$$

Also folgt

$$0 = 0x - (0x) = (0x + 0x) - (0x) = 0x + (0x - (0x)) = 0x$$

und

$$0 = \lambda 0 - (\lambda 0) = (\lambda 0 + \lambda 0) - (\lambda 0) = \lambda 0 + (\lambda 0 - (\lambda 0)) = \lambda 0 \, .$$

2. Es gelte $\lambda x = 0$. Ist $\lambda = 0$, so sind wir fertig. Es sei also $\lambda \neq 0$. Dann gilt nach 1. und (M2), (M1)

$$x = 1 \cdot x = \left(\frac{1}{\lambda}\lambda\right) x = \frac{1}{\lambda}(\lambda x) = \frac{1}{\lambda} \cdot 0 = 0 \, .$$

3. Nach 1. und (D2) bzw. (D1) gilt

$$0 = 0x = (\lambda + (-\lambda))x = \lambda x + (-\lambda)x$$

und

$$0 = \lambda 0 = \lambda(x + (-x)) = \lambda x + \lambda(-x) \, .$$

Da das inverse Element bzgl. + eindeutig bestimmt ist, gilt also

$$-(\lambda x) = (-\lambda)x = \lambda(-x) \, .$$

□

Beispiel 3.3 1. Es sei K ein Körper. Dann ist für $n \in \mathbb{N}$

$$K^n := K^{\{1,\dots,n\}} = \{(x_1, \dots, x_n) : x_j \in K \text{ für } j = 1, \dots, n\}$$

mit $+ : K^n \times K^n \to K^n$ definiert durch

$$x + y = (x_1, \dots, x_n) + (y_1, \dots, y_n) := (x_1 + y_1, \dots, x_n + y_n)$$

für $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in K^n$ und $\cdot : K \times K^n \to K^n$, definiert durch

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

für $x = (x_1, \dots, x_n) \in K^n$ und $\lambda \in K$, ein linearer Raum über dem Körper K.
(Denn:

1. Behauptung: (AV) gilt, d. h. $(K^n, +)$ ist eine abelsche Gruppe.
   (G.1) : Es seien $x, y, z \in K^n$. Dann gilt

$$\begin{aligned}(x+y)+z &= [(x_1,\dots,x_n)+(y_1,\dots,y_n)]+(z_1,\dots,z_n) = \\ &= (x_1+y_1,\dots,x_n+y_n)+(z_1,\dots,z_n) = \\ &= ((x_1+y_1)+z_1,\dots,(x_n+y_n)+z_n) \\ &= (x_1+(y_1+z_1),\dots,x_n+(y_n+z_n)) \\ &= (x_1,\dots,x_n)+[(y_1,\dots,y_n)+(z_1,\dots,z_n)] = x+(y+z)\end{aligned}$$

   (G4): Es seien $x = (x_1,\dots,x_n), y = (y_1,\dots,y_n) \in K^n$. Dann gilt

$$x+y = (x_1+y_1,\dots,x_n+y_n) = (y_1+x_1,\dots,y_n+x_n) = y+x\ .$$

   (G2): Mit $0 := 0_V := (0,\dots 0) \in K^n$ gilt für alle $x = (x_1,\dots,x_n) \in X$

$$x+0 = (x_1,\dots,x_n)+(0,\dots,0) = (x_1+0,\dots,x_n+0) = (x_1,\dots,x_n) = x$$

   (G3): Es sei $x = (x_1,\dots,x_n) \in K^n$. Dann gilt für $y := (-x_1,\dots,-x_n) \in K^n$

$$x+y = (x_1,\dots x_n)+(-x_1,\dots,-x_n) = (x_1-x_1,\dots,x_n-x_n) = (0,\dots,0) = 0.$$

2. Behauptung: (M1) gilt. Dazu seien λ, $\mu \in K$ und $x = (x_1,\dots,x_n) \in K^n$. Dann gilt

$$\begin{aligned}\lambda\cdot(\mu\cdot x) &= \lambda\cdot(\mu x_1,\dots,\mu x_n) = (\lambda(\mu x_1),\dots,\lambda(\mu x_n)) = \\ &= ((\lambda\mu)x_1,\cdot,(\lambda\mu)x_n) = (\lambda\mu)(x_1,\dots,x_n)\ .\end{aligned}$$

3. Behauptung: (M2) gilt. Es sei $x = (x_1,\dots,x_n) \in K^n$. Dann gilt

$$1\cdot x = 1\cdot(x_1,\dots,x_n) = (1x_1,\dots,1x_n) = (x_1,\dots,x_n)\ .$$

4. Behauptung: (D1) gilt. Es seien $\lambda \in K$ und $x, y \in K^n$. Dann gilt

$$\begin{aligned}\lambda(x+y) &= \lambda((x_1,\dots,x_n)+(y_1,\dots,y_n)) = \lambda(x_1+y_1,\dots,x_n+y_n) = \\ &= (\lambda(x_1+y_1),\dots,\lambda(x_n+y_n)) = (\lambda x_1+\lambda y_1,\dots,\lambda x_n+\lambda y_n) = \\ &= \lambda(x_1,\dots,x_n)+\lambda(y_1,\dots,y_n) = \lambda x+\lambda y\ .\end{aligned}$$

5. Behauptung: (D2) gilt. Beweis analog zu 4.)

Insbesondere ergibt sich damit zusammen mit B. 2.6:

a) $\mathbb{Q}^n = (\mathbb{Q}^n, +, \cdot)$ ist linearer Raum über $\mathbb{Q}$,

b) $\mathbb{R}^n = (\mathbb{R}^n, +, \cdot)$ ist linearer Raum über $\mathbb{R}$,

c) $\mathbb{C}^n = (\mathbb{C}^n, +, \cdot)$ ist linearer Raum über $\mathbb{C}$.

Da weiter aus D. 2.1 sofort folgt, dass ein linearer Raum über K auch ein linearer Raum über $\tilde{K}$ für jeden Körper $\tilde{K} \subset K$ ist, ist damit auch $\mathbb{R}^n$ linearer Raum über $\mathbb{Q}$ und $\mathbb{C}^n$ linearer Raum über $\mathbb{R}$ und $\mathbb{Q}$.

Veranschaulichung im Falle $\mathbb{R}^2$:
$x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2, \lambda \in \mathbb{R}$.

2. Es sei $X \neq \emptyset$ eine Menge und K ein Körper. Wir definieren für $f, g \in K^X$ und $\lambda \in K$

$$\begin{array}{lcl} (f+g)(x) & := & f(x) + g(x) \\ (\lambda f)(x) & := & \lambda \cdot f(x) \end{array} \quad (x \in X) \; .$$

Dann ist $(K^X, +, \cdot)$ ein linearer Raum über K.
(Beweis wie oben im Spezialfall $X = \{1, \ldots, n\}$.)

4 Unterräume

Wir betrachten jetzt bestimmte Teilmengen von linearen Räumen.

Definition 4.1 Es sei V ein linearer Raum über K, und es sei $U \subset V$ nichtleer. Dann heißt U *(linearer) Unterraum* (oder *(linearer) Teilraum* oder *Untervektorraum)* von V, falls $(U, +_{|U\times U}, \cdot_{|K\times U})$ ein linearer Raum ist. Wir schreiben dann wieder $+$ für $+_{|U\times U}$ und $\cdot$ für $\cdot_{|K\times U}$.

Äußerst nützlich zum Nachweis der Unterraum-Eigenschaft ist.

Satz 4.2 *Es seien V ein linearer Raum über K und $U \subset V$ nichtleer. Dann gilt: U ist Unterraum von V genau dann, wenn die folgenden Bedingungen erfüllt sind:*

a) Für alle $x, y \in U$ ist $x + y \in U$,

b) Für alle $\lambda \in K$, $x \in U$ ist $\lambda x \in U$.

Beweis.
Die Bedingungen a) und b) bedeuten gerade, dass $+_{|U\times U} : U \times U \to U$ und $\cdot_{|K\times U} : K \times U \to U$ gilt, d. h. Addition und Skalarmultiplikation führen nicht aus U heraus.
1. „$\Rightarrow$“ : Ist $U \subset V$ ein Unterraum, so gelten a) und b) nach D.4.1 und D.3.1.
2. „$\Leftarrow$“ : Nach D.3.1 gelten (G 1) und (G 4) für die Addition sowie (M 1), (M 2) und (D 1), (D 2) auch in U. Es bleiben noch (G 2) und (G 3) aus (AV) zu zeigen.
Ist $y \in U$, so gilt $0 = 0 \cdot y \in U$ nach S.3.2.1. und b) , also ist $0 \in U$ neutrales Element bzgl. $+$. Ist $x \in U$, so ist $-x = -(1x) = (-1)x \in U$ nach S.3.2.3. und b). Klar ist dann $x + (-x) = 0$, also existiert zu x ein inverses Element bzgl. $+$. □

Beispiel 4.3 1. Ist V ein beliebiger linearer Raum über K, so sind $U = V$ und $U = \{0\}$ Unterräume von V. Der Raum $\{0\}$ heißt *Nullraum (von V).*
2. Es sei $V = \mathbb{R}^n$ und $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ fest. Dann ist

$$U_a := \{(x_1, \dots, x_n) \in \mathbb{R}^n : \sum_{k=1}^{n} a_k x_k = 0\}$$

ein Unterraum von $\mathbb{R}^n$.
(Denn: Sind $x, y \in U_a$ und ist $\lambda \in \mathbb{R}$, so gilt

$$\sum_{k=1}^{n} a_k(x_k + y_k) = \sum_{k=1}^{n} a_k x_k + \sum_{k=1}^{n} a_k y_k = 0 + 0 = 0$$

und

$$\sum_{k=1}^{n} a_k(\lambda x_j) = \lambda \sum_{k=1}^{n} a_k x_k = \lambda 0 = 0 \ .$$

Nach S.4.2 ist U ein Unterraum von $\mathbb{R}^n$.)
3. Es sei $V = \mathbb{R}^{\mathbb{R}} (= \{f : \mathbb{R} \to \mathbb{R}\})$. Wir betrachten für $n \in \mathbb{N}_0$ und $a_0, \ldots, a_n \in \mathbb{R}$ die Funktion $P : \mathbb{R} \to \mathbb{R}$ (d. h. $P \in V$) mit

$$P(x) = \sum_{\nu=0}^{n} a_\nu x^\nu \qquad (x \in \mathbb{R}) \ .$$

Eine solche Funktion heißt *Polynomfunktion (in* $\mathbb{R}$*)* oder kurz *Polynom (in* $\mathbb{R}$*)*. Weiter setzen wir

$$\Pi := \Pi_{\mathbb{R}} := \{\text{Polynome in } \mathbb{R}\} \ .$$

Dann ist Π ein Unterraum von $\mathbb{R}^{\mathbb{R}}$
(Denn: Sind $P, Q \in \Pi$ und ist $\lambda \in \mathbb{R}$ so existieren $a_0, \ldots, a_n \in \mathbb{R}$ und $b_0, \ldots, b_m \in \mathbb{R}$ mit

$$P(x) = \sum_{\nu=0}^{n} a_\nu x^\nu \quad \text{und} \quad Q(x) = \sum_{\nu=0}^{m} b_\nu x^\nu \qquad (x \in \mathbb{R}) \ .$$

Ohne Einschränkung sei $n \geq m$. Dann gilt für alle $x \in \mathbb{R}$

$$(P+Q)(x) = \sum_{\nu=0}^{n} a_\nu x^\nu + \sum_{\nu=0}^{m} b_\nu x^\nu = \sum_{\nu=0}^{n} (a_\nu + b_\nu) x^\nu \ ,$$

wobei $b_\nu := 0$ für $\nu = m+1, \ldots, n$ im Falle $m < n$ gesetzt ist. Also ist $P + Q$ ein Polynom in $\mathbb{R}$. Weiter gilt

$$(\lambda P)(x) = \lambda \sum_{\nu=0}^{n} a_\nu x^\nu = \sum_{\nu=0}^{n} (\lambda a_\nu) x^\nu \ ,$$

d. h. λP ist ein Polynom in $\mathbb{R}$. Nach S.4.2 ist Π ein Unterraum von $\mathbb{R}^{\mathbb{R}}$.)
Polynome in $\mathbb{C}$ sind genau wie oben definiert indem man überall $\mathbb{R}$ durch $\mathbb{C}$ ersetzt (für die Variable schreibt man dann üblicherweise z statt x). Dann ist

$$\Pi := \Pi_{\mathbb{C}} := \{\text{Polynome in } \mathbb{C}\}$$

ein Unterraum von $\mathbb{C}^{\mathbb{C}}$.

Satz 4.4 *Ist* $\mathcal{F} \neq \emptyset$ *ein System von Unterräumen eines linearen Raumes* V *über* K*, so ist* $U := \bigcap_{W \in \mathcal{F}} W$ *ein Unterraum von* V*.*

Beweis.
Zunächst ist $0 \in W$ für alle $W \in \mathcal{F}$, also $0 \in U$. Es seien $x, y \in U$ und $\lambda \in K$ gegeben. Dann gilt $x, y \in W$ für alle $W \in \mathcal{F}$, also nach S.4.2 auch $x + y \in W$ und $\lambda x \in W$ für alle $W \in \mathcal{F}$. Dies bedeutet wiederum $x + y \in U$ und $\lambda x \in U$. Nach S.4.2 ist U ein Unterraum von V. □

Der Satz besagt, dass jeder Durchschnitt von Unterräumen wieder ein Unterraum ist. Wie man leicht sieht, ist i. a. die Vereinigung von Unterräumen *kein* Unterraum. Wir werden nun einer beliebigen Teilmenge $M \subset V$ einen "kleinsten" Unterraum zuordnen, der M enthält.

Definition 4.5 Es sei M eine Teilmenge eines linearen Raumes V über K. Dann heißt mit $\mathcal{F} := \mathcal{F}_{\mathcal{M}} := \{U : U \text{ Unterraum von } V,\ U \supset M\}$

$$< M > (:= \text{span}\,(M) := \text{lin span}\,(M) := LH(M)) := \bigcap_{U \in \mathcal{F}} U$$

die *lineare Hülle von M*.
2. Sind $U_1, \ldots, U_n$ Unterräume von V, so heißt

$$\sum_{j=1}^{n} U_j :=< \bigcup_{j=1}^{n} U_j >$$

Summe der $U_1, \ldots, U_n$. Gilt zusätzlich $U_j \cap \sum\limits_{\substack{k=1 \\ k \neq j}}^{n} U_k = \{0\}$ für alle $j = 1, \ldots, n$, so heißt

$$\bigoplus_{j=1}^{n} U_j := \sum_{j=1}^{n} U_j$$

direkte Summe der $U_1, \ldots, U_n$.

Bemerkung 4.6 Da $V \supset M$ ein Unterraum von V ist, wird der Durchschnitt in der D.4.5 stets über eine nichtleere Menge gebildet. Weiter gilt nach S.4.4, dass $< M >$ ein Unterraum von V ist. Damit sind auch $\sum\limits_{j=1}^{n} U_j$ und $\bigoplus\limits_{j=1}^{n} U_j$ Unterräume von V. Ausserdem folgt aus der Definition sofort, dass für jeden Unterraum $\tilde{U} \supset M$ gilt $\tilde{U} \supset < M >$ (d. h. $< M >$ ist der "kleinste" Unterraum, der M enthält). Schließlich sei bemerkt, dass $U_1 + U_2 = U_1 \oplus U_2$ genau dann gilt, wenn $U_1 \cap U_2 = \{0\}$ ist.

Unser Ziel ist es nun, eine "explizitere" Darstellung für $< M >$ bzw. $\sum\limits_{j=1}^{n} U_j$ zu finden.

Definition 4.7 Es sei V ein linearer Raum über K. Sind $x_1, \dots, x_n \in V$, so heißt $x \in V$ eine *Linearkombination* der Vektoren $x_1, \dots, x_n$, falls $\lambda_1, \dots, \lambda_n \in K$ existieren mit

$$x = \sum_{j=1}^{n} \lambda_j x_j \ .$$

Satz 4.8 *Es sei V ein linearer Raum über K. Dann gilt*
1. Ist $M \subset V, M \neq \emptyset$, so ist $x \in < M >$ genau dann, wenn ein $n \in \mathbb{N}$ und $x_1, \dots, x_n \in M$ sowie $\lambda_1, \dots, \lambda_n \in K$ existieren mit

$$x = \sum_{j=1}^{n} \lambda_j x_j$$

(d. h. $< M >= \{$Linearkombinationen aus Vektoren aus $M\}$).
2. Sind $U_1, \dots, U_n$ Unterräume von V, so ist $x \in \sum_{j=1}^{n} U_j$ genau dann, wenn $u_j \in U_j$ $(j = 1, \dots, n)$ existieren mit

$$x = \sum_{j=1}^{n} u_j \ .$$

Weiter gilt: die Summe ist direkt, d. h. $\sum_{j=1}^{n} U_j = \bigoplus_{j=1}^{n} U_j$, genau dann, wenn diese Darstellung für jedes $x \in \sum_{j=1}^{n} U_j$ eindeutig ist (d. h. sind $\tilde{u}_j \in U_j$ $(j = 1, \dots, n)$ mit $x = \sum_{j=1}^{n} \tilde{u}_j$, so gilt $u_j = \tilde{u}_j$ für $j = 1, \dots, n$).

Beweis.
1. Es sei

$$U := \{\sum_{j=1}^{n} \lambda_j x_j : \ x_1, \dots, x_n \in M, \ \lambda_1, \dots, \lambda_n \in K, \ n \in \mathbb{N}\} \ .$$

Dann ist zu zeigen: $< M >= U$.
"$\subset$": Wir zeigen $< M > \subset U$. Da $M \subset U$ ist, genügt es nach B.4.6 zu zeigen, dass U ein Unterraum von V ist.
Es seien also $x, y \in U$ und $\lambda \in K$ gegeben. Dann existieren $x_1, \dots, x_n \in M, \lambda_1, \dots, \lambda_n \in K$ mit

$$x = \sum_{j=1}^{n} \lambda_j x_j$$

und $x_{n+1}, \dots, x_m \in M, \lambda_{n+1}, \dots, \lambda_m \in K$ mit

$$y = \sum_{j=n+1}^{m} \lambda_j x_j \ .$$

Also gilt

$$x+y=\sum_{j=1}^{n}\lambda_j x_j+\sum_{j=n+1}^{m}\lambda_j x_j=\sum_{j=1}^{m}\lambda_j x_j\in U$$

und

$$\lambda x=\lambda\sum_{j=1}^{n}\lambda_j x_j=\sum_{j=1}^{n}(\lambda\lambda_j)x_j\in U\ .$$

Nach S.4.2 ist U ein Unterraum von V.
"$\supset$": Es sei $x\in U$. Dann existieren $x_1,\dots,x_n\in M,\lambda_1,\dots,\lambda_n\in K$ mit $x=\sum\limits_{j=1}^{n}\lambda_j x_j$. Da $\{x_1,\dots,x_n\}\subset M\subset<M>$ gilt, und da $<M>$ ein Unterraum von V ist, ist auch $x=\sum\limits_{j=1}^{n}\lambda_j x_j\in<M>$.
2. Es sei

$$U:=\{\sum_{j=1}^{n}u_j : u_j\in U_j \text{ für } j=1,\dots,n\}\ .$$

Dann ist zu zeigen:$\sum_{j=1}^{n}U_j=U$.
"$\subset$": Wir zeigen $\sum\limits_{j=1}^{n}U_j\subset U$. Dazu genügt es nach B.4.6 zu zeigen: U ist ein Unterraum von V (man beachte, dass nach Definition $U_j\subset U$ für $j=1,\dots,n$ gilt).
Es seien also $x,y\in U$ und $\lambda\in K$ gegeben. Dann existieren $u_j,v_j\in U_j$ für $j=1,\dots,n$ mit

$$x=\sum_{j=1}^{n}u_j\quad\text{und}\quad y=\sum_{j=1}^{n}v_j\ .$$

Also gilt

$$x+y=\sum_{j=1}^{n}(u_j+v_j)\in U$$

(da $u_j+v_j\in U_j$ für $j=1,\dots,n$) und

$$\lambda x=\sum_{j=1}^{n}\lambda u_j\in U$$

(da $\lambda u_j\in U_j$ für $j=1,\dots,n$). Nach S.4.2 ist U ein Unterraum von V.
"$\supset$": Wir zeigen $\sum_{j=1}^{n}U_j\supset U$.
Dazu sei $x\in U$. Dann gilt $x=\sum_{j=1}^{n}u_j$ für gewisse $u_j\in U_j\,(j=1,\dots,n)$. Da $\{u_1,\dots,u_n\}\subset\bigcup\limits_{j=1}^{n}U_j\subset<\bigcup\limits_{j=1}^{n}U_j>=\sum\limits_{j=1}^{n}U_j$ gilt, und da $\sum\limits_{j=1}^{n}U_j$ ein Unterraum von V ist, gilt auch $\sum\limits_{j=1}^{n}u_j\in\sum\limits_{j=1}^{n}U_j$.

3. „⇒“:Es sei $x = \sum_{j=1}^{n} u_j = \sum_{j=1}^{n} \tilde{u}_j$, wobei $u_j, \tilde{u}_j \in U_j$ für $j = 1, \dots, n$. Dann gilt

$$0 = \sum_{j=1}^{n} (u_j - \tilde{u}_j) \ .$$

Es sei $j \in \{1, \dots, n\}$. Dann gilt nach 2.

$$U_j \ni \tilde{u}_j - u_j = \sum_{\substack{k=1 \\ k \neq j}}^{n} u_k - \tilde{u}_k \in \sum_{\substack{k=1 \\ k \neq j}}^{n} U_k \ ,$$

also ist $\tilde{u}_j - u_j = 0$ nach D.4.5, d. h. $u_j = \tilde{u}_j$. Da $j \in \{1, \dots, n\}$ beliebig war, folgt die Behauptung.
„⇐“:[Ü]. □

Bemerkung 4.9 Ein besonderer einfacher und wichtiger Spezialfall ist gegeben durch $M = \{x_1, \dots, x_n\} \subset V$ (d. h. M ist endlich). Dann ist

$$< x_1, \dots, x_n > := < \{x_1, \dots, x_n\} > = \{\sum_{j=1}^{n} \lambda_j x_j : \lambda_j \in K \text{ für } j = 1, \dots, n\} \ .$$

Beispiel 4.10 1. Es sei $V = \mathbb{R}^2$ und $x_1 = (1, 0), x_2 = (1, 1), x_3 = (0, 1) \in \mathbb{R}^2$.

Dann gilt

$$\begin{aligned} U_1 = < x_1 > &= \{\lambda x_1 : \lambda \in \mathbb{R}\} = \{(\lambda, 0) : \lambda \in \mathbb{R}\} \\ U_2 = < x_2 > &= \{\lambda x_2 : \lambda \in \mathbb{R}\} = \{(\lambda, \lambda) : \lambda \in \mathbb{R}\} \\ U_3 = < x_3 > &= \{\lambda x_3 : \lambda \in \mathbb{R}\} = \{(0, \lambda) : \lambda \in \mathbb{R}\} \\ < x_1, x_2 > &= \{\lambda_1 x_1 + \lambda_2 x_2 : \lambda_1, \lambda_2 \in \mathbb{R}\} = \mathbb{R}^2 \\ &= < x_1, x_3 > = < x_2, x_3 > \ . \end{aligned}$$

Weiter ist

$$U_1 + U_2 = < U_1 \cup U_2 > = \mathbb{R}^2 = U_1 + U_3 = U_2 + U_3$$

und $U_1 \cap U_2 = \{0\} = U_1 \cap U_3 = U_2 \cap U_3$, also

$$U_1 \oplus U_2 = U_1 \oplus U_3 = U_2 \oplus U_3 = \mathbb{R}^2$$

Aber: $U_1 \cap (U_2 + U_3) = U_1 \cap \mathbb{R}^2 = U_1 \neq \{0\}$, d. h.

$$\mathbb{R}^2 = U_2 + U_2 + U_3 \ ,$$

aber die Summe ist nicht direkt.

2. Es sei $\Pi = \Pi_{\mathbb{C}}$ (oder $\Pi_{\mathbb{R}}$) wie in B.4.3. Dann gilt

$$\Pi = \Pi_g \oplus \Pi_u \ ,$$

wobei

$$\Pi_g := \{\sum_{\nu=0}^{n} a_\nu z^{2\nu} : a_\nu \in \mathbb{C}, \nu = 1, \dots, n \ ; n \in \mathbb{N}\}$$

und

$$\Pi_u := \{\sum_{\nu=0}^{m} b_\nu z^{2\nu+1} : b_\nu \in \mathbb{C}, \nu = 1, \dots, m \ ; m \in \mathbb{N}\} \ .$$

Π_g ist die Menge der "geraden" und Π_u die Menge der "ungeraden" Polynome.
(Denn: Man sieht leicht, dass Π_g und Π_u Unterräume sind, und dass $\Pi_g + \Pi_u = \Pi$ gilt. Es sei $P \in \Pi_u \cap \Pi_g$. Dann gilt mit gewissen $a_\nu \in \mathbb{C}$ für $\nu = 0, \dots, n$ und $b_\nu \in \mathbb{C}$ für $\nu = 0, \dots, m$

$$P(z) = \sum_{\nu=0}^{n} a_\nu z^{2\nu} = \sum_{\nu=0}^{m} b_\nu z^{2\nu+1} \ .$$

Dann gilt für alle $z \in \mathbb{C}$

$$P(z) = \sum_{\nu=0}^{n} a_\nu z^{2\nu} = \sum_{\nu=0}^{m} a_\nu (-1)^{2\nu} z^{2\nu} = P(-z)$$

und

$$P(z) = \sum_{\nu=0}^{m} b_\nu z^{2\nu+1} = -\sum_{\nu=0}^{m} a_\nu (-1)^{2\nu+1} z^{2\nu+1} = -P(-z)$$

also $2P(z) = 0$ und damit $P(z) = 0$. Folglich ist $P = 0_\Pi$.)

Definition 4.11 Es sei V ein linearer Raum über K.
1. Ist $I \neq \emptyset$ so nennen wir ein $(x_\alpha)_{\alpha \in I} \in V^I$ eine *Familie in* V. Die Familie $(x_\alpha)_{\alpha \in I}$ heißt *endlich*, falls I endlich ist.
Ist $J \subset I, J \neq \emptyset$, so heißt $(x_\alpha)_{\alpha \in J}$ eine *Teilfamilie* von $(x_\alpha)_{\alpha \in I}$.
2. Eine Familie $(x_\alpha)_{\alpha \in I}$ in V heißt *Erzeugendensystem von* V, falls

$$< x_\alpha : \alpha \in I >= V$$

(wobei $< x_\alpha : \alpha \in I >:=< \{x_\alpha : \alpha \in I\} >$).
3. Der Raum V heißt *endlich erzeugt* (oder *endlich-dimensional*), falls ein endliches Erzeugendensystem von V existiert, d. h. falls Vektoren $x_1, \dots, x_n \in V$ existieren mit $< x_1, \dots, x_n >= V$. Anderenfalls heißt V *unendlich erzeugt* (oder *unendlich-dimensional*).

Beispiel 4.12 1. Es seien $n \in \mathbb{N}$ und K ein Körper. Dann ist $V = K^n$ endlich-dimensional. Es gilt nämlich etwa

$$K^n =< e_1, \dots, e_n >$$

wobei $e_j = (\delta_{j1,\dots,}\delta_{jn}) \in K^n$ mit

$$\delta_{jk} := \begin{cases} 1, & \text{falls } j = k \\ 0, & \text{sonst} \end{cases} ,$$

d. h. $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$.
(Beachte: Ist $x = (\lambda_1, \dots, \lambda_n) \in K^n$, so gilt $x = \sum_{j=1}^n \lambda_j e_j$.)
2. Der Raum Π aus B.4.3.3 ist nicht endlich-dimensional.
(Denn: Angenommen es existieren Polynome $P_1, \dots, P_n$ mit $P_j \neq 0$ und $< P_1, \dots, P_n >= \Pi$. Dann ist

$$P_j(z) = \sum_{j=1}^{m_j} a_{j\nu} z^\nu$$

für gewisse $a_{j\nu} \in \mathbb{C}$ (oder $\mathbb{R}$) und gwisse $m_1, \dots, m_n \in \mathbb{N}_0$. Wir setzen

$$N := \max\{m_1, \dots, m_n\} \ (\in \mathbb{N}_0) \ .$$

Dann existieren $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ mit

$$z^{N+1} = \sum_{j=1}^{n} \lambda_j P_j(z) =: \sum_{\nu=0}^{N} a_\nu z^\nu \ .$$

Für $z \neq 0$ folgt $1 = \sum\limits_{\nu=0}^{N} a_\nu z^{\nu-N-1}$ und für $|z| > \max\{1, N \cdot \max\limits_{0\leq\nu\leq N} |a_\nu|\}$ ergibt sich

$$1 = |\sum_{0}^{N} a_\nu / z^{N+1-\nu}| \leq \sum_{0}^{N} |a_\nu| \frac{1}{|z|^{N+1-\nu}} \leq \sum_{0}^{N} |a_\nu|/|z| < 1 \ ,$$

also ein Widerspruch.)

5 Lineare Unabhängigkeit, Basis und Dimensionen

Wir starten mit einem weiteren zentralen Begriff der Linearen Algebra.

Definition 5.1 Es sei V ein linearer Raum über K.
1. Sind $x_1, \dots, x_n \in V$, so heißt $(x_1, \dots, x_n)$ *linear abhängig*, falls $\lambda_1, \dots, \lambda_n \in K$ existieren mit $\lambda_j \neq 0$ für ein $j \in \{1, \dots, n\}$ und

$$\sum_{j=0}^{n} \lambda_j x_j = 0 \ .$$

Anderenfalls heißt $(x_1, \dots, x_n)$ linear unabhängig.
Man sagt auch "$x_1, \dots, x_n$ sind linear abhängig" bzw. "$x_1, \dots, x_n$ sind linear unabhängig".
2. Eine Familie $(x_\alpha)_{\alpha \in I}$ in V heißt *linear unabhängig*, falls jede endliche Teilfamilie linear unabhängig ist.

Bemerkung 5.2 1. Aus D.5.1 ergibt sich sofort, dass $x_1, \dots, x_n \in V$ linear unabhängig genau dann sind, wenn gilt: Sind $\lambda_1, \dots, \lambda_n \in K$ mit

$$\sum_{j=0}^{n} \lambda_j x_j = 0 \,,$$

so ist $\lambda_1 = \ldots = \lambda_n = 0$.
2. Ein Element $x \in V$ ist linear unabhängig genau dann, wenn $x \neq 0$ ist.
3. Ist $n \geq 2$, so sind $x_1, \dots, x_n \in V$ genau dann linear abhängig, wenn ein $j_0 \in \{1, \dots, n\}$ und $\mu_j \in K \, (j \in \{1, \dots, n\} \setminus \{j_0\})$ existieren mit

$$x_{j_0} = \sum_{\substack{j=1 \\ j \neq j_0}}^{n} \mu_j x_j \,.$$

(Denn: "$\Rightarrow$" : Sind $x_1, \dots, x_n$ linear abhängig, so existieren $\lambda_1, \dots, \lambda_n \in K$ mit $\lambda_{j_0} \neq 0$ für ein $j_0 \in \{1, \dots, n\}$ und $0 = \sum\limits_{j=1}^{n} \lambda_j x_j$. Also folgt

$$x_{j_0} = \sum_{\substack{j=1 \\ j \neq j_0}}^{n} \left(-\frac{\lambda_j}{\lambda_{j_0}} \right) x_j \,.$$

"$\Leftarrow$": Ist umgekehrt $x_{j_0} = \sum\limits_{\substack{j=1 \\ j \neq j_0}}^{n} \mu_j x_j$ für ein $j_0 \in \{1, \dots, n\}$ und $\mu_j \in K \; (j \in \{1, \dots, n\} \setminus \{j_0\})$, so ist mit $\lambda_j := \mu_j \, (j \neq j_0)$ und $\lambda_{j_0} := -1$

$$0 = \sum_{j=1}^{n} \lambda_j x_j \,.)$$

Beispiel 5.3 1. Es sei $V = K^n$. Dann sind $e_1, \dots, e_n$ linear unabhängig in K^n.
(Denn: Sind $\lambda_1, \dots, \lambda_n \in K$ mit

$$(0, \dots, 0) = 0 = \sum_{j=1}^{n} \lambda_j e_j = (\lambda_1, \dots, \lambda_n) \,,$$

so gilt $\lambda_j = 0$ für $j = 1, \dots, n$.)
2. Es sei $V = \Pi$ und für $n \in \mathbb{N}_0$ sei $E_n \in \Pi$ definiert durch

$$E_n(z) = z^n \qquad (z \in \mathbb{C}) \,.$$

Dann ist $(E_0, \dots, E_n)$ linear unabhängig in Π für alle $n \in \mathbb{N}_0$.
(Denn: Es seien $\lambda_0, \dots, \lambda_n \in \mathbb{C}$ mit

$$0 = \sum_{j=0}^{n} \lambda_j z^j = \sum_{j=0}^{n} \lambda_j E_j(z) \qquad (z \in \mathbb{C}) .$$

Angenommen, es existiert ein $j \in \{0, \dots, n\}$ mit $\lambda_j \neq 0$. Wir setzen dann

$$m := \max\{j \in \{0, \dots, n\} : \lambda_j \neq 0\} .$$

Ist $m = 0$, so ist $0 = \lambda_0 \neq 0$, also Widerspruch. Ist $m > 0$, so folgt

$$z^m = \sum_{j=0}^{m-1} \left(-\frac{\lambda_j}{\lambda_m}\right) z^j \qquad (z \in \mathbb{C}) .$$

Dies führt auf den gleichen Widerspruch wie in B.4.12).
Hieraus folgt auch, dass $(E_n)_{n\in\mathbb{N}_0}$ linear unabhängig in Π ist. Außerdem gilt offenbar $\Pi =< E_n : n \in \mathbb{N}_0 >$.

Satz 5.4 *Es sei V ein linearer Raum über K und es seien $x_1, \dots, x_n \in V$. Dann sind aquivalent:*
a) $x_1, \dots, x_n$ sind linear unabhängig.
b) Zu jedem $x \in < x_1, \dots, x_n >$ existieren eindeutig bestimmte $\lambda_1, \dots, \lambda_n \in K$ (d. h. es existiert genau ein $(\lambda_1, \dots, \lambda_n) \in K^n$) mit

$$x = \sum_{j=1}^{n} \lambda_j x_j .$$

Beweis.
1. a) $\Rightarrow$ b): Es sei $x \in < x_1, \dots, x_n >$. Dann existieren $\lambda_1, \dots, \lambda_n \in K$ mit

$$x = \sum_{j=1}^{n} \lambda_j x_j .$$

Es sei $x = \sum_{j=1}^{n} \mu_j x_j$ mit $\mu_1, \dots, \mu_n \in K$. Dann ist

$$0 = \sum_{j=1}^{n} (\lambda_j - \mu_j) x_j .$$

Da $x_1, \dots, x_n$ linear unabhängig sind, folgt $\lambda_j - \mu_j = 0$, d. h. $\lambda_j = \mu_j$ für $j = 1, \dots, n$.
2. b) $\Rightarrow$ a): Angenommen $x_1, \dots, x_n$ sind linear abhängig. Dann existiert ein $(\lambda_1, \dots, \lambda_n) \in K^n \setminus \{0\}$ mit $0 = \sum_{j=1}^{n} \lambda_j x_j$. Also lässt sich 0 darstellen als

$$0 = \sum_{j=1}^{n} \lambda_j x_j = \sum_{j=1}^{n} 0 x_j$$

im Widerspruch zu b). □

Definition 5.5 Es sei V ein linearer Raum über K. Eine Familie $(x_\alpha)_{\alpha\in I}$ in V heißt *(algebraische) Basis* von V, falls $(x_\alpha)_{\alpha\in I}$ ein linear unabhängiges Erzeugendensystem von V ist. Für $V = \{0\}$ bezeichnen wir $\emptyset$ als Basis

Wir werden uns im folgenden i. w. auf die Untersuchung endlich-dimensionaler linearer Räume beschränken. Algebraische Basen spielen in unendlich-dimensionalen Räumen i. a. keine so wichtige Rolle.

Bemerkung 5.6 Aus S.5.4 folgt sofort: Sind $x_1, \dots, x_n \in V$, so sind äquivalent:
a) $(x_1, \dots, x_n)$ ist eine Basis von V.
b) Jedes $x \in V$ besitzt genau eine Darstellung $x = \sum_{j=1}^{n} \lambda_j x_j$ mit $\lambda_1, \dots, \lambda_n \in K$.

Beispiel 5.7 Es sei $V = K^n$, wobei K ein Körper ist (vgl. B.4.12). Dann ist $(e_1, \dots, e_n)$ eine Basis von K^n (B.4.12 und B.5.3). Diese Basis nennen wir im folgenden *Standardbasis* oder *kanonische Basis* von K^n.

Dieses "Standardbeispiel" zeigt, dass in K^n eine Basis aus n Elementen existiert. Wir wollen uns nun allgemeiner folgenden Fragen zuwenden:

1. Existiert in jedem (endlich-dimensionalen) linearen Raum eine Basis?

2. Haben je 2 Basen eines (endlich-dimensionalen) linearen Raumes die gleiche Anzahl von Elementen?

Wir beweisen zunächst den

Satz 5.8 (Basisauswahlsatz) *Es seien $V \neq \{0\}$ ein linearer Raum über K und $(x_1, \dots, x_m)$ ein endliches Erzeugendensystem von V. Dann existiert eine Teilfamilie von $(x_1, \dots, x_m)$, die eine Basis von V ist.*

Beweis.
Wir setzen

$$J := \{j \in \{1, \dots, m\} : x_j \notin < x_1, \dots, x_{j-1} >\}$$

(mit $< x_1, \dots, x_0 > := < \emptyset > = \{0\}$).
1. Wir zeigen $< x_j : j \in J > = V$. Angenommen, $< x_j : j \in J > \neq V$. Dann existiert wegen $< x_1, \dots, x_m > = V$ ein $k \in \{1, \dots, m\} \setminus J$ mit $x_k \notin < x_j : j \in J >$ (sonst würde gelten $\{x_1, \dots, x_n\} \subset < x_j : j \in J >$, also auch $V = < x_1, \dots, x_n > \subset < x_j : j \in J >$).

Wir setzen $k_0 := \min\{k \in \{1,\dots,m\} \setminus J : x_k \notin < x_j : j \in J >\} (\in \{1,\dots,m\} \setminus J)$. Dann ist $< x_1,\dots,x_{k_0-1} > \subset < x_j : j \in J >$, also $x_{k_0} \notin < x_1,\dots,x_{k_0-1} >$. Nach Definition ist also $k_0 \in J$. Widerspruch!
2. Wir zeigen $(x_j)_{j\in J}$ ist linear unabhängig. Angenommen nicht. Dann existieren $\lambda_j \in K$ $(j \in J)$, $\lambda_j \neq 0$ für ein j, mit $0 = \sum\limits_{j\in J} \lambda_j x_j$. Für $j_0 := \max\{j : \lambda_j \neq 0\}$ gilt

$$x_{j_0} = \sum_{\substack{j\in J \\ j<j_0}} \left(-\frac{\lambda_j}{\lambda_{j_0}}\right) x_j \in < x_1,\dots,x_{j_{0-1}} >$$

im Widerspruch zur Definition von J. □

Satz 5.9 *Es sei V ein endlich-dimensionaler linearer Raum über K. Dann existiert eine (endliche) Basis von V.*

Beweis.
Da V (ohne Einschränkung $\neq \{0\}$) endlich-dimensional ist, existiert ein endliches Erzeugendensystem von V. Nach S.5.8 existiert damit auch eine endliche Basis von V. □

Die Aussage von S.5.9 bleibt auch für beliebige lineare Räume richtig. Der Beweis basiert auf einer Anwendung des Auswahlaxioms, worauf wir nicht weiter eingehen wollen.
Wir wenden uns der zweiten der oben angesprochenen Fragen zu. Dazu beweisen wir zunächst folgendes Hilfsresultat.

Satz 5.10 *Es sei V ein linearer Raum über K, und es sei $(x_1,\dots,x_n)$ eine Basis von V. Ist $x = \sum\limits_{j=1}^{n} \lambda_j x_j \in V$ mit $\lambda_k \neq 0$ für ein $k \in \{1,\dots,n\}$, so ist auch $(x_1,\dots,x_{k-1},x,x_{k+1},\dots,x_n)$ eine Basis von V.*

Beweis.
1. Wir zeigen: Für $M := \{x_1,\dots,x_n,x\} \setminus \{x_k\}$ gilt $< M >= V$. Dazu sei $y \in V$ gegeben. Dann existieren $\mu_1,\dots,\mu_n \in K$ mit $y = \sum\limits_{j=1}^{n} \mu_j x_j$. Wegen $\lambda_k \neq 0$ ist $x_k = \frac{1}{\lambda_k} x - \sum\limits_{\substack{j=1 \\ j\neq k}}^{n} \frac{\lambda_j}{\lambda_k} x_j$ und daher

$$\begin{aligned} y &= \sum_{j=1}^{n} \mu_j x_j = \frac{\mu_k}{\lambda_k} x - \sum_{\substack{j=1 \\ j\neq k}}^{n} \frac{\mu_k \lambda_j}{\lambda_k} x_j + \sum_{\substack{j=1 \\ j\neq k}}^{n} \mu_j x_j \\ &= \frac{\mu_k}{\lambda_k} x - \sum_{\substack{j=1 \\ j\neq k}}^{n} (\mu_j - \frac{\mu_k \lambda_j}{\lambda_k}) x_j \ , \end{aligned}$$

also $y \in < M >$. Da $y \in V$ beliebig war, ist $< M >= V$.
2. Es seien $\mu_1, \dots, \mu_n \in K$ mit

$$0 = \mu_k x + \sum_{j \neq k} \mu_j x_j \ .$$

Dann gilt mit $x = \sum_{j=1}^{n} \lambda_j x_j$

$$\mu_k \lambda_k x_k + \sum_{j \neq k} (\mu_k \lambda_j + \mu_j) x_j = 0 \ .$$

Da $x_1, \dots, x_n$ linear unabhängig sind, folgt $\mu_k \lambda_k = 0$ und $\mu_k \lambda_j + \mu_j = 0$ für $j = 1, \dots, n, j \neq k$. Mit $\lambda_k \neq 0$ ergibt sich $\mu_k = 0$, also auch $\mu_j = 0$ für alle $j \in \{1, \dots, n\}$.
□

Hiermit gilt der zentrale

Satz 5.11 (Steinitz'scher Austauschsatz) *Es sei $V \neq \{0\}$ ein linearer Raum über K, und es sei $(x_1, \dots, x_n)$ eine Basis von V. Ferner sei $(y_1, \dots, y_m)$ linear unabhängig in V. Dann ist $m \leq n$, und es existieren $n - m$ Elemente aus $\{x_1, \dots, x_n\}$, die zusammen mit $y_1, \dots, y_m$ eine Basis bilden, d. h. es existieren $j_1, \dots, j_{n-m} \in \{1, \dots, n\}$ so, dass $(y_1, \dots, y_m, x_{j_1}, \dots x_{j_{n-m}})$ eine Basis von V ist, oder es ist $n = m$ und $(y_1, \dots, y_m)$ ist eine Basis von V.*

Beweis.
Wir zeigen folgende Aussage A_m für alle $m \in \mathbb{N}$.
(A_m): Existiert eine linear unabhängige Familie $(y_1, \dots, y_m)$ in V, so ist $m \leq n$. Außerdem existieren im Falle $m < n$ zu jeder solchen Familie $j_1, \dots, j_{n-m} \in \{1, \dots, n\}$ so, dass $(y_1, \dots, y_m, x_{j_1}, \dots, x_{j_{n-m}})$ eine Basis von V bildet, und im Falle $m = n$ ist $(y_1, \dots, y_m)$ eine Basis von V.
1. Induktionsanfang: Für $m = 1$ ist natürlich $m \leq n$. Ist $y \neq 0$, so gilt $y = \sum_{j=1}^{n} \lambda_j x_j$ mit $\lambda_k \neq 0$ für ein $k \in \{1, \dots, n\}$. Dann ist $(y, x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$ nach S.5.10 eine Basis von V.
2. Induktionsschritt von m auf $m+1$: (A_m) gelte für ein $m \in \mathbb{N}$. Zu zeigen ist: (A_{m+1}) gilt.
Existiert keine Familie $(y_1, \dots, y_{m+1})$ in V, die linear unabhängig ist, so ist nichts zu zeigen. Es sei also $(y_1, \dots, y_{m+1})$ in V linear unabhängig. Dann ist auch $(y_1, \dots, y_m)$ in V linear unabhängig. Nach Induktionsvoraussetzung ist $m \leq n$. Angenommen, es ist $m = n$. Dann ist $(y_1, \dots, y_m)$ eine Basis von V und damit ist $y_{m+1} \in < y_1, \dots, y_m >$ im Widerspruch zur linearen Unabhängigkeit von $(y_1, \dots, y_{m+1})$. Also ist $m < n$, d.

h. $m+1 \leq n$ und es existieren $j_1, \ldots, j_{n-m}$ so, dass $(y_1, \ldots, y_m, x_{j_1}, \ldots, x_{j_{n-m}})$ eine Basis von V bilden. Insbesondere ist

$$y_{m+1} = \sum_{\nu=1}^{m} \lambda_\nu y_\nu + \sum_{k=1}^{n-m} \mu_k x_{j_k}$$

für gewisse $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_{n-m} \in K$. Da $(y_1, \ldots, y_{m+1})$ linear unabhängig ist, ist $\mu_k \neq 0$ für ein $k \in \{1, \ldots, n-m\}$. Nach S.5.10 ist dann

$$(y_1, \ldots, y_{m+1}, x_{j_1}, \ldots, x_{j_{k-1}}, x_{j_{k+1}}, \ldots, x_{j_{n-m}})$$

eine Basis von V der gesuchten Form (Man beachte, dass die Basis $n-m-1 = n-(m+1)$ Elemente aus $\{x_1, \ldots, x_n\}$ enthält.) □

Beispiel 5.12 Es sei $V = \mathbb{R}^3$, und es sei $(x_1, x_2, x_3) = (e_1, e_2, e_3)$ die kanonische Basis in V. Ferner seien

$$y_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \qquad y_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Dann sind y_1, y_2 linear unabhängig in V. Nach S.5.11 existiert ein $j \in \{1, 2, 3\}$, so dass (y_1, y_2, x_j) eine Basis von V ist. Es gilt hier: (y_1, y_2, x_1) und (y_1, y_2, x_3) sind Basen von V, aber (y_1, y_2, x_2) ist keine Basis von V.

Als wichtige Konsequenz aus S.5.11 erhalten wir

Bemerkung und Definition 5.13 Es sei V ein linearer Raum über K. Ist V endlichdimensional, so hat jede Basis die gleiche Anzahl n von Elementen.
(Denn: Ist $V = \{0\}$, so ist $\emptyset$ die einzige Basis von V. Es sei also $V \neq \{0\}$. Nach S.5.9 besitzt V eine endliche Basis $(x_1, \ldots, x_n)$. Nun sei $(y_j)_{j \in J}$ eine weitere Basis von V. Dann ist J endlich, denn anderenfalls würden endliche, linear unabhängige Teilfamilien von $(y_j)_{j \in J}$ beliebiger Länge existieren, insbesondere also mit mehr als n Elementen im Widerspruch zu S.5.11. Also hat wieder nach S.5.11 einerseits $(y_j)_{j \in J}$ höchstens n Elemente und andererseits auch mindestens n Elemente, denn die linear unabhängige Familie $(x_1, \ldots, x_n)$ kann nicht mehr Elemente haben als die Basis $(y_j)_{j \in J}$).
Die Zahl

$$\dim(V) := n$$

heißt *Dimension* von V, und V heißt *n-dimensional*.

Beispiel 5.14 1. Es sei $V = K^n$ wie in B.4.12. Dann ist nach B.5.7 $\dim(V) = n$.
2. Es sei $V = \mathbb{C}$. Dann ist $\mathbb{C}$ ein linearer Raum über $\mathbb{C}$ und ein linearer Raum über $\mathbb{R}$. Nach 1. ist $\mathbb{C}$ über $\mathbb{C}$ 1-dimensional. Wie man leicht sieht ist $(1, i)$ eine Basis von $\mathbb{C}$ als linearem Raum über $\mathbb{R}$. Also ist $\mathbb{C}$ über $\mathbb{R}$ 2-dimensional.

Satz 5.15 *Es sei V ein n-dimensionaler linearer Raum über K, und es seien $x_1, \dots, x_n \in V$. Dann sind folgende Aussagen äquivalent:*
a) $(x_1, \dots, x_n)$ ist eine Basis von V.
b) $(x_1, \dots, x_n)$ ist ein Erzeugendensystem von V.
c) $(x_1, \dots, x_n)$ ist linear unabhängig.

Beweis.
1. a) $\Rightarrow$ b) und a) $\Rightarrow$ c) sind klar.
2. b) $\Rightarrow$ a): Nach dem Basisauswahlsatz (S.5.8) existiert ein $J \subset \{1, \dots, n\}$ so, dass $(x_j)_{j \in J}$ eine Basis von V ist. Dann ist nach S.5.13 $|J| = n$, also $J = \{1, \dots, n\}$.
3. c) $\Rightarrow$ a): Es existiert eine Basis aus n Elementen . Nach dem Austauschsatz (S.5.11) folgt aus c), dass $(x_1, \dots, x_n)$ eine Basis von V ist. □

Im Basisauswahlsatz hatten wir gesehen, dass jedes Erzeugendensystem zu einer Basis "ausgedünnt" werden kann. Andererseits gilt auch

Satz 5.16 (Basisergänzungssatz) *Es sei V ein endlich-dimensionaler linearer Raum über K. Ist $(y_1, \dots, y_m)$ linear unabhängig in V, so existiert eine Basis von V, die $(y_1, \dots, y_m)$ als Teilfamilie enthält.*

Beweis.
Es existiert eine Basis $(x_1, \dots, x_n)$ von V. Also ist nach S.5.11 $m \leq n$, und es existiert eine Basis von V, die $(y_1, \dots, y_m)$ enthält. □

Satz 5.17 *Es sei V ein endlich-dimensionaler linearer Raum über K, und es sei $U \subset V$ ein Unterraum. Dann gilt*
1. U ist endlich-dimensional mit $\dim(U) \leq \dim(V)$.
2. Ist $\dim(U) = \dim(V)$, so ist $U = V$.

Beweis.
1. Angenommen, U ist unendlich-dimensional. Dann existiert eine Folge $(x_j)_{j \in \mathbb{N}}$ in U mit $x_1 \neq 0$ und $x_j \notin < x_1, \dots, x_{j-1} >$ für $j \geq 2$.
(Denn: Es ist $U \neq \{0\}$, also existiert ein $x_1 \in U \setminus \{0\}$. Sind $x_1, \dots, x_{j-1}$ wie oben

konstruiert, so existiert, da $< x_1, \dots, x_{j-1} > \neq U$ ist, ein $x_j \in U \setminus < x_1, \dots, x_{j-1} >$. Induktiv ergibt sich $(x_j)_{j\in\mathbb{N}}$.)
Wie im Beweisschritt 2. zu S.5.8 sieht man, dass $(x_1, \dots, x_j)$ für alle $j \in \mathbb{N}$ linear unabhängig in U, also auch in V, ist. Dies widerspricht S.5.11, nachdem jede linear unabhängige Familie in V höchstens $\dim(V)$ Elemente enthält.
Ist $(x_1, \dots, x_m)$ eine Basis von U, so ist $(x_1, \dots, x_m)$ linear unabhängig in U und in V, d. h. es gilt $\dim(U) = m \leq \dim(V)$ nach S.5.11.
2. Ist $\dim(U) = \dim(V) = n$, so existiert eine Basis $(x_1, \dots, x_n)$ von U. Also ist wieder $(x_1, \dots, x_n)$ linear unabhängig in U und in V. Nach S.5.15 ist $(x_1, \dots, x_n)$ eine Basis von V, also $V =< x_1, \dots, x_n >= U$. □

Satz 5.18 (Dimensionsformel für Unterräume) *Es seien U, W Unterräume eines endlich-dimensionalen linearen Raumes V über K. Dann gilt*

$$\boxed{\dim(U + W) + \dim(U \cap W) = \dim(U) + \dim(W)}$$

Beweis.
Es sei $k := \dim(U)$, $\ell := \dim(W)$. Nach S.4.4 ist $U \cap W$ ein Unterraum von V mit $m := \dim(U \cap W) \leq \min(k, \ell)$ nach S.5.17. Es sei $B = (x_1, \dots, x_m)$ (bzw. $B = \emptyset$) eine Basis von $U \cap W$. Diese sei gemäß S.5.16 durch Vektoren $u_{m+1}, \dots, u_k \in U$ zu einer Basis von U (falls $k > m$) und durch $w_{m+1}, \dots, w_\ell \in W$ (falls $\ell > m$) zu einer Basis von W ergänzt.
Es genügt zu zeigen: $M := (x_1, \dots, x_m,\ u_{m+1}, \dots, u_k,\ w_{m+1}, \dots, w_\ell)$ ist linear unabhängig in $U + W$.
(Denn dann ist M auch eine Basis von $U + W$ nach Konstruktion, und damit ist $\dim(U + W) = m + (k - m) + (\ell - m) = k + \ell - m$ wie behauptet.)
Es sei also (mit $\sum_{\emptyset} := 0$)

$$0 = \sum_{j=1}^{m} \alpha_j x_j + \sum_{j=m+1}^{k} \beta_j u_j + \sum_{j=m+1}^{\ell} \gamma_j w_j \qquad (*)$$

für gewisse $\alpha_j, \beta_j, \gamma_j \in K$. Dann ist

$$x := \sum_{j=1}^{m} \alpha_j x_j + \sum_{j=m+1}^{k} \beta_j u_j = -\sum_{j=m+1}^{\ell} \gamma_j w_j \in U \cap W\ .$$

Da $(x_1, \dots, x_m)$ eine Basis von $U \cap W$ ist, gilt auch

$$x = \sum_{j=1}^{m} \tilde{\alpha}_j x_j \left(= \sum^{m} j =_1 \alpha_j x_j + \sum_{j=m+1}^{k} \beta_j u_j \right)$$

mit gewissen $\tilde{\alpha}_j \in K$. Da $(x_1, \dots, x_m, u_{m+1}, \dots, u_k)$ eine Basis von U ist folgt nach B.5.6 $\beta_j = 0\,(j = m+1, \dots, k)$ (und $\tilde{\alpha}_j = \alpha_j\,(j = 1, \dots, m)$). Die lineare Unabhängigkeit von $(x_1, \dots, x_m, w_{m+1}, \dots, w_\ell)$ liefert mit $(*)$ wiederum $\alpha_j = 0$ $(j = 1, \dots, m)$ und $\gamma_j = 0\,(j = m+1, \dots, \ell)$. □

Folgerung 5.19 Unter den Voraussetzungen von S.5.18 ergibt sich sofort: Es ist $U + W = U \oplus W$ genau dann, wenn

$$\dim(U + W) = \dim(U) + \dim(W)\ .$$

(Denn: Ist $U + W$ direkt, so ist $U \cap W = \{0\}$, also $\dim(U \cap W) = 0$. Ist umgekehrt die Formel richtig, so ist $\dim(U \cap W) = 0$, also $U \cap W = \{0\}$ und damit $U + W = U \oplus W$). Weiter ergibt sich induktiv: Sind $U_1, \dots, U_m$ Unterräume von V, so gilt

$$\boxed{U_1 + \cdots + U_m = U_1 \oplus \cdots \oplus U_m}$$

genau dann, wenn

$$\boxed{\dim\left(\sum_{j=1}^{m} U_j\right) = \sum_{j=1}^{m} \dim(U_j)}$$

Beweis. [Ü]

Zum Abschluss beweisen wir noch

Satz 5.20 *Es sei V ein endlich-dimensionaler linearer Raum über K, und es sei $U \subset V$ ein Unterraum. Dann existiert ein Unterraum $W \subset V$ mit*

$$V = U \oplus W\ .$$

Beweis.
Ohne Einschränkung sei $U \neq V$ und $U \neq \{0\}$ (sonst ist $W = \{0\}$ bzw. $W = V$ geeignet). U ist als Unterraum von V endlich-dimensional (S.5.17). Ist $(u_1, \dots, u_m)$ eine Basis von U, so existiert nach S.5.16 eine Basis $(u_1, \dots, u_m, w_1, \dots, w_\ell)$ von V. Dann gilt für $W =< w_1, \dots, w_\ell >$ die Behauptung. (Denn: Nach Konstruktion ist $U + W = V$. Außerdem gilt $\dim(U + W) = m + \ell = \dim(U) + \dim(W)$, also $U + W = U \oplus W$) □

6 Lineare Abbildungen

Bisher haben wir uns beschränkt auf die Untersuchung einzelner linearer Räume. "Lebendig" und die Theorie erst so richtig dadurch, dass Beziehungen zwischen verschiedenen Räumen hergestellt werden. Entscheidend sind dabei lineare Abbildungen.

Definition 6.1 Es seien V, W lineare Räume über einem (gemeinsamen) Körper K. Eine Abbildung $T : V \to W$ heißt *linear* (oder auch *(linearer) Operator*), falls gilt

a) Für alle $v_1, v_2 \in V$ ist

$$T(v_1 + v_2) = T(v_1) + T(v_2)$$

und

b) Für alle $v \in V, \lambda \in K$ ist

$$T(\lambda v) = \lambda T(v) \; .$$

Ist T linear, so schreibt man auch kurz "Tv" statt "$T(v)$".
Weiter setzen wir

$$L(V, W) := \{T : V \to W \text{ linear}\}$$

und $L(V) := L(V, V)$. Ist speziell $W = K$, so heißt T ein *lineares Funktional (auf V)* und $V^* := L(V, K)$ heißt *Dualraum von V*.

Bemerkung 6.2 1. Ist $T \in L(V, W)$, so gilt für $\lambda_1, \ldots, \lambda_n \in K, v_1, \ldots, v_n \in V$

$$T\left(\sum_{j=1}^{n} \lambda_j v_j\right) = \sum_{j=1}^{n} \lambda_j T(v_j)$$

(folgt per Induktion aus D.6.1).
2. Ist $T \in L(V, W)$, so gilt stets $T(0) = 0$
(denn $T(0) = T(0 + 0) = T(0) + T(0)$, also $T(0) = 0$).
3. Sind $v_1, \ldots, v_n$ in V linear abhängig, so sind $Tv_1, \ldots, Tv_n$ in W linear abhängig.
(Denn: Ist

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

für gewisse $\lambda_j \in K$ $(j = 1, \ldots, n)$, wobei $\lambda_j \neq 0$ für ein j, so ist

$$0 = T(0) = T\left(\sum_{j=1}^{n} \lambda_j v_j\right) = \sum_{j=1}^{n} \lambda_j T v_j \; ,$$

also sind $Tv_1, \dots, Tv_n$ linear abhängig.)
Entsprechend gilt damit: Sind $Tv_1, \dots, Tv_n$ linear unabhängig, so sind auch $v_1, \dots, v_n$ linear unabhängig
Man beachte aber: Aus $v_1, \dots, v_n$ linear unabhängig folgt i.a. nicht, dass $Tv_1, \dots, Tv_n$ linear unabhängig sind! (Beispiel: $T : V \to W, T(v) \equiv 0\,(v \in V)$.)

Beispiel 6.3 Es sei K ein Körper, und es sei $V = K^n$. Ferner sei $a = (a_1, \dots, a_n) \in K^n$.
Wir definieren $T : K^n \to K$ durch

$$Tv = T(x_1, \dots, x_n) := \sum_{k=1}^{n} a_k x_k \qquad (v = (x_1, \dots, x_n) \in K^n)\,.$$

Dann ist T linear, d. h. T ist ein lineares Funktional ([Ü]). Allgemeiner gilt: Sind $A_1, \dots, A_m \in K^n$, so ist $T : K^n \to K^m$, definiert durch

$$Tv := (T_1 v, \dots, T_m v) \qquad (v \in K^n)$$

ebenfalls linear. Dabei ist T_j wie oben mit A_j anstelle von $a\,(j = 1, \dots, m)$ definiert, d. h. ist $A_j = (a_1^{(j)}, \dots, a_n^{(j)})$, so gilt

$$T_j(v) = T_j(x_1, \dots, x_n) := \sum_{k=1}^{n} a_k^{(j)} x_k \qquad (v = (x_1, \dots, x_n) \in K^n)\,.$$

(Denn: Sind $v_1, v_2 \in K^n$, so gilt

$$\begin{aligned} T(v_1 + v_2) &= (T_1(v_1 + v_2), \dots, T_m(v_1 + v_2)) = (T_1 v_1 + T_1 v_2, \dots, T_m v_1 + T_m v_2) \\ &= (T_1 v_1, \dots, T_m v_1) + (T_1, v_2, \dots, T_m v_2) = Tv_1 + Tv_2 \end{aligned}$$

Entsprechend sieht man

$$T(\lambda v) = \lambda T v$$

für $\lambda \in K, v \in K^n$).
Ist etwa $K = \mathbb{R}, n = 2, m = 3$ sowie $A_1 = (1, -2), A_2 = (3, 0), A_3 = (-2, 3)$ so ist $T : \mathbb{R}^2 \to \mathbb{R}^3$ mit

$$T(x_1, x_2) = (x_1 - 2x_2, 3x_1, 3x_2 - 2x_1) \qquad ((x_1, x_2) \in \mathbb{R}^2)$$

linear.

Wir versehen jetzt in natürlicher Weise $L(V, W)$ mit einer Vektorraumstruktur.

Satz 6.4 *Es seien V, W lineare Räume über K. Für $T, S \in L(V,W)$ sei $T+S : V \to W$ definiert durch*

$$(T+S)(v) := T(v) + S(v) \qquad (v \in V) .$$

Ferner sei für $\lambda \in K$ die Abbildung $\lambda \cdot T : V \to W$ definiert durch

$$(\lambda \cdot T)(v) := \lambda(Tv) \qquad (v \in V) .$$

Dann gilt $+ : L(V,W) \times L(V,W) \to L(V,W)$ sowie $\cdot : K \times L(V,W) \to L(V,W)$ und $(L(V,W), +, \cdot)$ ist ein linearer Raum über K.
Ist U ein weiterer linearer Raum über K, und ist $T \in L(V,W)$ sowie $S \in L(U,V)$, so ist

$$T \circ S \in L(U,W) .$$

Beweis.
1. $S+T \in L(V,W)$ und $\lambda T \in L(V,W)$: Bedingungen aus D.6.1 prüfen.
$(L(V,W), +, \cdot)$ linearer Raum: Vektorraumaxiome überprüfen.
2. Für $u_1, u_2 \in U$ gilt

$$\begin{aligned} (T \circ S)(u_1 + u_2) &= T(S(u_1+u_2)) = T(Su_1 + Su_2) = \\ &= T(Su_1) + T(Su_2) = (T \circ S)(u_1) + (T \circ S)(u_2) . \end{aligned}$$

Ist $u \in U$ und $\lambda \in K$, so gilt weiterhin

$$\begin{aligned} (T \circ S)(\lambda u) &= T(S(\lambda u)) = T(\lambda(Su)) = \lambda T(S(u)) = \\ &= \lambda (T \circ S)(u) . \end{aligned}$$

Damit ist $S \circ T : \ V \to U$ linear. □

Definition 6.5 Es seien V, W lineare Räume über K, und es sei $T \in L(V,W)$. Ist $T : V \to W$ bijektiv, so heißt T *Isomorphismus*. Existiert ein Isomorphismus $T : V \to W$, so heißen V und W *isomorph*. Ein Isomorphismus $T : V \to V$ heißt auch *Automorphismus (auf V)*. Wir setzen

$$\mathrm{Aut}(V) := \{T : V \to V : T \text{ Automorphismus auf } V\} .$$

Sind zwei Räume isomorph, so sind sie bzgl. der linearen Struktur als "gleich" anzusehen.

Satz 6.6 *Es seien V, W lineare Räume über K, und es sei $T : V \to W$ ein Isomorphismus. Dann ist auch $T^{-1} : W \to V$ ein Isomorphismus. Ist U ein weiterer linearer Raum über K, und ist $S : U \to V$ ein Isomorphismus, so ist auch $T \circ S : U \to V$ ein Isomorphismus. Insbesondere ist also $(Aut(V), \circ)$ eine Gruppe.*

Beweis.
1. Nach B/D.1.18 ist $T^{-1} : W \to V$ bijektiv. Es bleibt zu zeigen: T^{-1} ist linear. Dazu seien $w_1, w_2 \in W$ gegeben. Dann existieren $v_1, v_2 \in V$ mit $T(v_j) = w_j \, (j = 1, 2)$. Also folgt

$$\begin{aligned} T^{-1}(w_1 + w_2) &= T^{-1}(T(v_1) + T(v_2)) = T^{-1}(T(v_1 + v_2)) = \\ &= v_1 + v_2 = T^{-1}(w_1) + T^{-1}(w_2) \, . \end{aligned}$$

Entsprechend sieht man $T^{-1}(\lambda w) = \lambda T^{-1}(w)$ für $w \in W, \lambda \in K$.
2. Mit T und S ist auch $T \circ S$ bijektiv. Außerdem ist $T \circ S$ linear nach S.6.4. □

Satz 6.7 *Es sei V ein n-dimensionaler Vektorraum über K. Dann ist V isomorph zu K^n.*

Beweis.
Es sei $(v_1, \ldots, v_n)$ eine Basis von V. Wir definieren $T : V \to K^n$ durch

$$T(v) = T\left(\sum_{j=1}^{n} \lambda_j v_j\right) := (\lambda_1, \ldots, \lambda_n) \qquad \left(v = \sum_{j=1}^{n} \lambda_j v_j \in V\right)$$

(Beachte: T ist wohldefiniert, da die Darstellung $v = \sum\limits_{j=1}^{n} \lambda_j v_j$ nach B.5.6 eindeutig ist!). Dann ist $T \in L(V, K^n)$
Denn: Sind $v = \sum\limits_{j=1}^{n} \lambda_j v_j$ und $w = \sum\limits_{j=1}^{n} \mu_j v_j \in V$, so gilt

$$\begin{aligned} T(v + w) &= T\left(\left(\sum_{j=1}^{n} \lambda_j v_j\right) + \left(\sum_{j=1}^{n} \mu_j v_j\right)\right) = T\left(\sum_{j=1}^{n} (\lambda_j + \mu_j) v_j\right) = \\ &= (\lambda_1 + \mu_1, \ldots, \lambda_n + \mu_n) = (\lambda_1, \ldots, \lambda_n) + (\mu_1, \ldots, \mu_n) = T(v) + T(w) \, . \end{aligned}$$

Ist $\lambda \in K$, so gilt zudem

$$\begin{aligned} T(\lambda v) &= T\left(\lambda \sum_{j=1}^{n} \lambda_j v_j\right) = T\left(\sum_{j=1}^{n} (\lambda \lambda_j) v_j\right) = (\lambda\lambda_1, \ldots, \lambda\lambda_n) = \\ &= \lambda(\lambda_1, \ldots, \lambda_n) = \lambda T(v) \, . \end{aligned}$$

Außerdem ist T bijektiv.
Denn: Ist $w = (\lambda_1, \ldots, \lambda_n) \in K^n$, so ist $T(v) = w$ für $v = \sum_{j=1}^{n} \lambda_j v_j$, also $v \in T^{-1}(\{w\})$, d. h. T ist surjektiv.
Ist $\tilde{v} \in T^{-1}(\{w\})$, d. h. $T(\tilde{v}) = w = (\lambda_1, \ldots, \lambda_n)$, so gilt $\tilde{v} = \sum_{j=1}^{n} \lambda_j v_j = v$ d. h. T ist injektiv. □

Bemerkung und Definition 6.8 Nach S.6.7 stimmt jeder n-dimensionale Vektorraum über K hinsichtlich seiner linearen Struktur mit dem "Modellraum" K^n überein. Mittels der Abbildung T aus dem Beweis zu S.6.7 lassen sich sämtliche Aussagen über die lineare Struktur von K^n auf V übertragen. Die Abbildung T nennt man *Koordinatenabbildung (bzgl.* $(v_1, \dots v_n)$*)*. Man beachte, dass T wesentlich von der in V gewählten Basis abhängt. Die Umkehrabbildung T^{-1} heißt *kanonischer Basisisomorphismus.*

Beispiel 6.9 Es sei $\Pi_n :=< E_0, \dots, E_n >$ $(= \{\text{Polynome vom Grad } \leq n\}) \subset \Pi$. Dann ist $(E_0, \dots, E_n)$ mit $E_\nu(z) = z^\nu$ eine Basis von Π_n (vgl. B.5.3). Nach S.6.7 ist Π_n isomorph zu $\mathbb{C}^{n+1}$, wobei die Koordinatenabbildung bzgl. $(v_1, \dots, v_{n+1}), v_j = E_{j+1}$ gegeben ist durch

$$T(P) = (\lambda_1, \dots, \lambda_{n+1})$$

für $P(z) = \sum_{\nu=0}^{n} a_\nu z^\nu = \sum_{j=1}^{n+1} \lambda_j v_j$ mit $\lambda_j := a_{j+1}\, (j = 1, \dots, n+1)$.

7 Kern und Bild

Es geht sofort los.

Definition 7.1 Es seien V, W lineare Räume über K, und es sei $T \in L(V, W)$. Dann heißen

$$\text{Kern}(T) := T^{-1}(\{0\})$$

der *Kern von* T und

$$\text{Bild}(T) := W(T) = T(V)$$

das *Bild von* T.

Es gelten folgende einfache Eigenschaften.

Satz 7.2 *Es sei* $T \in L(V, W)$*. Dann gilt*
1. $Kern(T)$ *ist ein Unterraum von* V*.*
2. $Bild(T)$ *ist ein Unterraum von* W*.*
3. T *ist genau dann injektiv, wenn* $Kern(T) = \{0\}$ *ist.*
4. Ist W *endlich-dimensional, so ist* T *genau dann surjektiv, wenn*

$$\dim(Bild(T)) = \dim W$$

gilt.

Beweis.
1. Zunächst ist $\text{Kern}(T) \neq \emptyset$ da $0 \in \text{Kern}(T)$ (B.6.2). Sind $v_1, v_2 \in \text{Kern}(T)$ und $\lambda_1, \lambda_2 \in K$, so gilt

$$T(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 T(v_1) + \lambda_2 T(v_2) = 0 \ ,$$

also auch $\lambda_1 v_1 + \lambda_2 v_2 \in \text{Kern}(T)$.
2. Aus $T(0) = 0$ folgt $0 \in \text{Bild}(T)$. Sind $w_1, w_2 \in \text{Bild}(T), \lambda_1, \lambda_2 \in K$, so existieren $v_1, v_2 \in V$ mit $T(v_j) = w_j \ (j = 1, 2)$. Also gilt

$$\lambda_1 w_1 + \lambda_2 w_2 = T(\lambda_1 v_1 + \lambda_2 v_2) \in \text{Bild}(T) \ .$$

3. "$\Rightarrow$": Ist T injektiv, so ist insbesondere $T^{-1}(\{0\})$ höchstens einpunktig. Aus $T(0) = 0$ folgt $\text{Kern}(T) = T^{-1}(\{0\}) = \{0\}$.
"$\Leftarrow$": Es seien $v_1, v_2 \in V$ mit $T(v_1) = T(v_2)$. Dann ist

$$0 = T(v_1) - T(v_2) = T(v_1 - v_2) \ ,$$

also $v_1 - v_2 = 0$, d. h. $v_1 = v_2$.
4. "$\Leftarrow$" folgt sofort aus S.5.17 und 2.; "$\Rightarrow$" ist klar. □

Beispiel 7.3 (vgl. 6.3) Es sei $V = \mathbb{R}^n, W = \mathbb{R}$, und für ein $(a_1, \ldots, a_n) \in \mathbb{R}^n \setminus \{0\}$ sei

$$T(v) = T(x_1, \ldots, x_n) = \sum_{j=1}^{n} a_j x_j \qquad (v = (x_1, \ldots, x_n) \in \mathbb{R}^n) \ .$$

Dann ist $\text{Kern}(T) = U_a$ aus B.4.3. Ist etwa $n = 2$ und $a = (1, -1)$, so ist

$$U_a = \text{Kern}(T) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = x_2\} \ .$$

Weiter gilt $\text{Bild}\,(T) = \mathbb{R}$, da $a_j \neq 0$ für ein $j \in \{1, \ldots, n\}$, und damit ist $T(e_j) = a_j \neq 0$, d. h. $\text{Bild}(T) \neq \{0\}$. (Dann ist schon $\text{Bild}(T) = \mathbb{R}$).

Eine zentrale Aussage über den Zusammenhang der "Größe" von Kern und Bild in endlich-dimensionalem Fall liefert

Satz 7.4 (Dimensionsformel für lineare Abbildungen) *Es seien V und W lineare Räume über K, wobei V endlich-dimensional sei. Ist $T \in L(V, W)$, so ist Bild(T) endlich-dimensional, und es gilt*

$$\boxed{\dim(\mathit{Kern}(T)) + \dim(\mathit{Bild}(T)) = \dim V}$$

Beweis.
1. Wir zeigen zunächst die Behauptung für den Spezialfall, dass T injektiv ist, also dann

$$\dim(\text{Bild}(T)) = \dim(V) .$$

Es sei $(v_1, \dots, v_n)$ eine Basis von V. Es genügt, zu zeigen $(Tv_1, \dots, Tv_n)$ ist eine Basis von $T(V) = \text{Bild}(T)$.
Zunächst gilt $< Tv_1, \dots, Tv_n >= \text{Bild}(T)$.
(Denn: Da Bild(T) ein Unterraum von W ist, ist

$$< Tv_1, \dots, Tv_n > \subset \text{Bild}(T) .$$

Ist $w \in T(V)$, d. h. $w = T(v)$ für ein $v \in V$, so existieren $\lambda_1, \dots, \lambda_n \in K$ mit $v = \sum_{j=1}^{n} \lambda_j v_j$. Also ist $w = T(v) = \sum_{j=1}^{n} \lambda_j T v_j \in < Tv_1, \dots, Tv_n >$).
Sind $\lambda_1, \dots, \lambda_n \in K$ mit $0 = \sum_{j=1}^{n} \lambda_j T v_j$, so ist $0 = T\left(\sum_{j=1}^{n} \lambda_j v_j\right)$, also $\sum_{j=1}^{n} \lambda_j v_j \in$ Kern(T). Aus Kern$(T) = \{0\}$ folgt $\sum_{j=0}^{n} \lambda_j v_j = 0$, also $\lambda_1 = \dots = \lambda_n = 0$, d. h. $Tv_1, \dots, Tv_n$ sind linear unabhängig.
2. Nun sei $T \in L(V, W)$ beliebig. Nach S.7.2 ist Kern(T) ein Unterraum von V. Also existiert nach S.5.20 ein Unterraum U von V mit

$$\text{Kern}(T) \oplus U = V .$$

Wir zeigen: $T(U) = \text{Bild}(T)$.
$T(U) \subset \text{Bild}(T) (= T(V))$ ist klar.
Es sei $w \in \text{Bild}(T)$, d. h. $w = T(v)$ für ein $v \in V$. Dann existieren $u_1 \in \text{Kern}(T)$, und $u_2 \in U$ mit $v = u_1 + u_2$, also

$$w = T(v) = T(u_1 + u_2) = T(u_1) + T(u_2) = T(u_2) \in T(U) .$$

Damit ist Bild$(T) \subset T(U)$.
Wie man leicht sieht, ist $T_{|U} : U \to V$ injektiv ([Ü]). Also folgt aus 1.

$$\dim(U) = \dim(T(U)) \ (= \dim(\text{Bild}(T))) .$$

Hieraus folgt dann mit F.5.19

$$\dim V = \dim(\text{Kern}(T)) + \dim(U) = \dim(\text{Kern}(T)) + \dim(\text{Bild}(T)) .$$

□

Satz 7.5 *Es seinen V, W endlichdimensionale lineare Räume mit* $\dim(V) = \dim(W)$. *Ist $T \in L(V, W)$ so sind äquivalent:*
a) T ist injektiv,
b) T ist surjektiv,
c) T ist bijektiv.

Beweis.
1. a) $\Rightarrow$ b): Nach S.7.2 ist $\text{Kern}(T) = \{0\}$. Also $\dim(\text{Kern}(T)) = 0$. Damit ist nach S.7.4 $\dim(\text{Bild}(T)) = \dim(V) = \dim(W)$, also ist T surjektiv nach S.7.2.
2. b) $\Rightarrow$ c): Nach Voraussetzung ist $\dim(\text{Bild}(T)) = \dim W = \dim V$, also nach S.7.4 $\dim(\text{Kern}(T)) = 0$, d. h. $\text{Kern}(T) = \{0\}$. Nach S.7.4 ist T injektiv, also bijektiv.
3. c) $\Rightarrow$ a): Nach Definition richtig. □

Bemerkung 7.6 1. Aus S.7.5 ergibt sich insbesonderre folgendes angenehme Kriterium für den Nachweis der Bijektivität von $T \in L(V, W)$: Ist $\dim(V) = \dim(W)$, und ist $\text{Kern}(T) = \{0\}$, so ist $T \in L(V, W)$ ein Isomorphismus.
2. Aus der Dimensionsformel (S.7.4) ergeben sich wichtige Eigenschaften über die (Nicht-) Existenz linearer Abbildungen. Es gilt nämlich:
(i) Ist $\dim(V) > \dim(W)$, so existiert kein injektives $T \in L(V, W)$.
(ii) Ist $\dim(W) > \dim(V)$, so existiert kein surjektives $T \in L(V, W)$.
(Denn: Es sei $T \in L(V, W)$.
(i) Aus S.7.4 folgt

$$\dim(\text{Kern}(T)) = \dim V - \dim(\text{Bild}(T)) \geq \dim V - \dim W > 0 \,,$$

also ist T nicht injektiv.
(ii) Aus S.7.4 folgt

$$\dim(\text{Bild}(T)) = \dim(V) - \dim(\text{Kern}(T)) \leq \dim(V) < \dim(W)$$

also ist T nicht surjektiv.)

8 Matrizen

Der folgende Satz zeigt insbesondere, dass lineare Abbildungen durch Angabe der Werte auf einer Basis festgelegt sind.

Satz 8.1 *Es seien V, W lineare Räume über K, und es seien $(v_1, \ldots, v_n)$ eine Basis von V sowie $u_1, \ldots, u_n \in W$ beliebig. Dann existiert genau ein $T \in L(V, W)$ mit*

$$T(v_j) = u_j \qquad (j = 1, \ldots, n) \,.$$

Beweis.
1. Existenz: Es sei $v \in V$. Dann existieren eindeutig bestimmte $\lambda_1, \ldots, \lambda_n \in K$ mit

$$v = \sum_{j=1}^{n} \lambda_j v_j \,.$$

Wir definieren $T : U \to W$ durch

$$T(v) := \sum_{j=1}^{n} \lambda_j u_j \qquad (v \in V) \ .$$

Dann ist T linear und es gilt $T(v_j) = u_j$ $(j = 1, \dots, n)$.
2. Eindeutigkeit: Ist $\tilde{T} : V \to W$ eine weitere lineare Abbildung mit $\tilde{T}(v_j) = u_j$ für $j = 1, \dots, n$, so gilt für jedes $v \in V, v = \sum_{j=1}^{n} \lambda_j v_j$:

$$\tilde{T}(v) = \tilde{T}\left(\sum_{j=1}^{n} \lambda_j v_j\right) = \sum_{j=1}^{n} \lambda_j \tilde{T}(v_j) = \sum_{j=1}^{n} \lambda_j u_j = T(v) \ ,$$

also $T = \tilde{T}$. □

Bemerkung 8.2 Es sei (unter den Voraussetzungen von S.8.1) W endlich-dimensional mit Basis $(w_1, \dots, w_m)$. Ferner sei T die Abbildung aus S.8.1. Dann existieren eindeutig bestimmte $(a_{11}, \dots, a_{m1}), \dots, (a_{1n}, \dots, a_{mn}) \in K^m$ mit

$$u_k = T(v_k) = \sum_{j=1}^{m} a_{jk} w_j \qquad (k = 1, \dots, n) \qquad (1)$$

Fasst man die Vektoren $a^{(1)} = (a_{11}, \dots, a_{m1}), \dots, a^{(n)} = (a_{1n}, \dots, a_{mn}) \in K^m$ in n Spalten zusammen, so entsteht folgendes Schema:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} =: (a_{jk})_{\substack{j=1,\dots,m \\ k=1,\dots,n}} =: (a_{jk}) \ .$$

Definition 8.3 Sind $n, m \in \mathbb{N}$ und sind $a_{jk} \in K$ für $j = 1, \dots, m, k = 1, \dots, n$, so heißt ein Schema $A = (a_{jk})$ der obigen Form eine $(m \times n)$-*Matrix (mit Einträgen in* K*)*.
Ist $m = n$, so heißt A *quadratische* Matrix.
Weiter setzen wir

$$K^{m \times n} := \{A : A \text{ ist } (m \times n) - \text{Matrix mit Einträgen in } K\} \ .$$

Ist T wie in B.8.2, so heißt die zugehörige Matrix A die *(Koordinaten-) Matrix von* T *bzgl.* $(v_1, \dots, v_n)$ *und* $(w_1, \dots, w_m)$. Dabei sei betont, dass A nicht nur von T, sondern i. a. auch von der speziellen Wahl von $(v_1, \dots, v_n)$ und $(w_1, \dots, w_m)$ abhängt!

Beispiel 8.4 Es sei $V = W = \mathbb{R}^2$ und es sei $T \in L(\mathbb{R}^2, \mathbb{R}^2)$ definiert durch

$$T(x_1, x_2) = (x_1, -x_2) \qquad ((x_1, x_2) \in \mathbb{R}^2) \, .$$

Ist $v_1 = w_1 = (1,0) = e_1, v_2 = w_2 = (0,1) = e_2$, so ist

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

(denn $e_1 = T(e_1) = 1 \cdot e_1 + 0 \cdot e_2, -e_2 = T(e_2) = 0 \cdot e_1 + (-1)e_2$). Ist aber $w_2 = (0,-1)$, so ist

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

(denn: $T(e_2) = -e_2 = 0 \cdot e_1 + 1 \cdot w_2$).

Satz 8.5 *Es seien V, W lineare Räume über K, und es seien $N := (v_1, \ldots, v_n)$ bzw. $M := (w_1, \ldots, w_m)$ Basen von V bzw. W. Dann ist $\varphi = \varphi_{M,N} : L(V,W) \to K^{m \times n}$, definiert durch*

$$\varphi(T) := \varphi_{M,N}(T) := A \qquad (T \in L(V,W)) \, ,$$

wobei A wie in B.8.2, bijektiv.

Beweis.
1. Nach B.8.2 ist $\varphi : L(V,W) \to K^{m \times n}$ wohldefiniert (beachte: A ist durch T und M, N eindeutig festgelegt).
2. φ ist injektiv, denn ist $A = \varphi(T) = \varphi(S)$ für $T, S \in L(V,W)$, so gilt nach B.8.2 (mit $A = (a_{jk})$):

$$T(v_k) = \sum_{j=1}^{m} a_{jk} w_j = S(v_k) \qquad (k = 1, \ldots, n) \, .$$

Aus der Eindeutigkeitsaussage von S.8.1 folgt $T = S$.
3. φ ist surjektiv, denn ist $A = (a_{jk}) \in K^{m \times n}$, so existiert nach S.8.1 ein $T \in L(V,W)$ mit

$$T(v_k) = u_k := \sum_{j=1}^{m} a_{jk} w_j \qquad (k = 1, \ldots, n) \, .$$

Also ist $\varphi(T) = A$. □

Definition 8.6 Es sei K ein Körper
1. Für $A, B \in K^{m\times n}, A = (a_{jk}), B = (b_{jk})$, definieren wir

$$A + B = (a_{jk} + b_{jk}) = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix} .$$

2. Für $A = (a_{jk}) \in K^{m\times n}, \lambda \in K$ definieren wir

$$\lambda A := (\lambda a_{jk}) = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1m} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix} .$$

3. Für $A = (a_{jk}) \in K^{m\times n}, B = (b_{jk}) \in K^{n\times p}$ definieren wir $A \cdot B \in K^{m\times p}$ durch

$$A \cdot B := (c_{jk})_{\substack{j=1,\dots,m \\ k=1,\dots,p}}$$

mit

$$c_{jk} = \sum_{\nu=1}^{n} a_{j\nu} b_{\nu k} \qquad (j = 1, \dots, m, k = 1, \dots, p) .$$

Satz 8.7 *Mit den Bezeichnungen aus S.8.5 und D.8.6 gilt*
1. $K^{m\times n}$ ist ein linearer Raum über K.
2. Die Abbildung $\varphi : L(V, W) \to K^{m\times n}$ ist ein Isomorphismus, d. h. $L(V, W)$ und $K^{m\times n}$ sind isomorph.

Beweis.
1. Entweder direkt nachrechnen oder durch Anwendung von B.3.3.2 (Beachte im zweiten Fall:

$$K^{m\times n} = K^{\{1,\dots,m\}\times\{1,\dots,n\}} ,$$

denn $A = (a_{jk}) \in K^{m\times n}$ ist eine Schreibweise für die Abbildung $f : \{1, \dots, m\} \times \{1, \dots, n\} \to K$,

$$f(j, k) = a_{jk} \quad (j = 1, \dots, m, k = 1, \dots, n).)$$

2. Nach S.8.5 genügt es, zu zeigen: φ ist linear. Dazu seien $T, S \in L(V, W)$ und $\lambda \in K$. Dann gilt mit

$$\varphi(T) = A = (a_{jk}) , \quad \varphi(S) = B = (b_{jk}) ,$$

und $N = (v_1, \dots, v_n), M = (w_1, \dots, w_m)$

$$\begin{aligned}(T+S)(v_k) &= T(v_k) + S(v_k) = \sum_{j=1}^{m} a_{jk} w_j + \sum_{j=1}^{m} b_{jk} w_j \\ &= \sum_{j=1}^{n} (a_{jk} + b_{jk}) w_j \qquad (k = 1, \dots, n)\end{aligned}$$

also ist $A+B = (a_{jk}+b_{jk})$ die Matrix von $T+S$ bzgl. M, N, d. h. $\varphi(T+S) = A+B = \varphi(T) + \varphi(S)$. Entsprechend sieht man, dass λA die Matrix von λT bzgl. M, N, also $\varphi(\lambda T) = \lambda A = \lambda \varphi(T)$ ist. □

Der Satz besagt also, dass man $L(V, W)$ von der linearen Struktur her mit $K^{m \times n}$, identifizieren kann, d. h. um lineare Abbildungen auf endlich-dimensionalen Räumen zu studieren, reicht es, Matrizen zu untersuchen. Dabei zeigt sich, dass sich nicht nur die lineare Struktur überträgt. Es gilt nämlich

Satz 8.8 *Es seien V, W, U endlich-dimensionale lineare Räume über K mit Basen $N = (v_1, \dots, v_n), M = (w_1, \dots, w_m)$ sowie $P = (u_1, \dots, u_p)$. Sind $T \in L(V, W)$ bzw. $S \in T(U, V)$ mit Matrix A bzgl. M, N bzw. Matrix B bzgl. N, P, so ist $A \cdot B$ die Matrix von $T \circ S$ bzgl. M, P, d. h.*

$$\varphi_{M,P}(T \circ S) = A \cdot B = \varphi_{M,N}(T) \cdot \varphi_{N,P}(S) \, .$$

Beweis.
Es seien

$$A =: (a_{jk}) \in K^{m \times n} \text{ und } \quad B =: (b_{jk}) \in K^{n \times p} \, .$$

Dann gilt

$$A \cdot B = (\sum_{\nu=1}^{n} a_{j\nu} b_{\nu k})_{\substack{j=1,\dots,m \\ k=1,\dots,p}} \, .$$

Außerdem ist

$$Tv_\nu = \sum_{j=1}^{m} a_{j\nu} w_j \qquad (\nu = 1, \dots, n)$$

und

$$Su_k = \sum_{\nu=1}^{n} b_{\nu k} v_\nu \qquad (k = 1, \dots, p) \, ,$$

also

$$\begin{aligned}(T \circ S)(u_k) &= T(Su_k) = \sum_{\nu=1}^{n} b_{\nu k} T v_\nu = \sum_{\nu=1}^{n} b_{\nu k} \sum_{j=1}^{m} a_{j\nu} w_j = \\ &= \sum_{j=1}^{m} w_j (\sum_{\nu=1}^{n} a_{j\nu} b_{\nu k}) \,,\end{aligned}$$

d. h. AB ist die Matrix von $T \circ S$ bzgl. M und P. □

Vereinbarung: Wir schreiben ab jetzt Vektoren in K^p, wobei $p \in \mathbb{N}$, als Spaltenvektoren, d. h.

$$K^p = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} : x_j \in K \text{ für } j = 1, \ldots, p \right\} .$$

Dann entspricht K^p der Menge der $(p \times 1)$-Matrizen $K^{p\times 1}$. Ist $C = (c_{jk}) \in K^{m\times n}$ und ist $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n = K^{n\times 1}$, so setzen wir

$$C \cdot x := C \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum\limits_{k=1}^{n} c_{1k} x_k \\ \vdots \\ \sum\limits_{k=1}^{n} c_{mk} x_k \end{pmatrix} \in K^m = K^{m\times 1} .$$

Bemerkung und Definition 8.9 Es sei K ein Körper und $V = K^n, W = K^m$. Weiter seien $N = (e_1, \ldots, e_n)$ bzw. $M = (e_1, \ldots, e_m)$ die kanonischen Basen (vgl. B.4.12 und B.5.7). Ist $T \in L(V, W)$ und $A = (a_{jk})$ die Matrix von T bzgl. der kanonischen Basen, d. h. $A = \varphi_{M,N}(T)$, so gilt

$$T(e_k) = \sum_{j=1}^{m} a_{jk} e_j = \sum_{j=1}^{m} a_{jk} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} ,$$

also die k-te Spalte von A.

Damit erhält man allgemein für $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$

$$T(x) = T\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = T\left(\sum_{k=1}^{n} x_k e_k\right) = \sum_{k=1}^{n} x_k \cdot T(e_k) =$$

$$= \sum_{k=1}^{n} x_k \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k \end{pmatrix} = A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

also

$$\boxed{T(x) = A \cdot x} \quad .$$

Ist umgekehrt eine Matrix $A \in K^{m\times n}$ gegeben, so ist durch

$$T(x) := Ax \qquad (x \in K^n)$$

ein $T = T_A \in L(K^n, K^m)$ definiert mit zugehöriger Matrix A bzgl. der kanonischen Basen, d. h. $A = \varphi_{M,N}(T_A)$.
So lassen sich Eigenschaften von T auf A und von A auf T übertragen. Man setzt etwa

$$\text{Bild}(A) := \text{Bild}(T) \quad \text{und} \quad \text{Kern}(A) := \text{Kern}(T) \; .$$

$\text{Bild}(A)$ heißt *Bild* von A und $\text{Kern}(A)$ heißt *Kern* von A.

Ist $S \in L(K^p, K^n)$, und ist B die Matrix von S bzgl. der kanonischen Basen, so gilt also nach S.8.8 auch

$$(T \circ S)(x) = (A \cdot B)x \qquad (x \in K^p) \; .$$

Es ist daher günstig, Rechenreglen für die Matrixmultiplikation zur Verfügung zu haben.

Satz 8.10 *Es sei K ein Körper.*
1. Für $A \in K^{m\times n}, B \in K^{n\times p}$ und $\lambda, \mu \in K$ gilt

$$(\lambda A)(\mu B) = (\lambda \cdot \mu)(AB) \; .$$

2. Für $A \in K^{m\times n}, B \in K^{n\times p}, C \in K^{p\times r}$ *gilt*

$$(AB)C = A(BC) .$$

3. Für $A \in K^{m\times n}, B, C \in K^{n\times p}$ *gilt*

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

4. Für $A, B \in K^{m\times n}, C \in K^{n\times p}$ *gilt*

$$(A + B)C = AC + BC .$$

Beweis. [Ü]

Bemerkung 8.11 Man beachte, dass für $A \in K^{m\times n}, B \in K^{n\times p}$ zwar AB, i. a. aber *nicht* BA definiert ist. Außerdem gilt auch in dem Fall, dass BA definiert ist (nämlich $p = m$), i. a. *nicht* $AB = BA$!
Betrachten wir etwa $A, B \in \mathbb{R}^{2\times 2}$ mit

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} , \; B = \begin{pmatrix} -3 & 2 \\ 1 & -2 \end{pmatrix} ,$$

so gilt

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

und

$$BA = \begin{pmatrix} -3 & 2 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -3 & -4 \end{pmatrix} \neq AB .$$

Wir kommen jetzt zu einer wichtigen Kenngröße allgemeiner Matrizen.

Definition 8.12 Es sei $A = (a_{jk}) \in K^{m\times n}$. Dann heißt

$$Rg(A) := \dim(\text{Bild}(A))$$

der *Rang* von A. Weiter seien $a^{(1)}, \ldots , a^{(n)}$ die Spalten von A, d. h.

$$a^{(k)} = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} \qquad (k = 1, \ldots , n)$$

sowie $A_1, \ldots, A_m$ die Zeilen von A, d. h.

$$A_j = (a_{j1}, \ldots, a_{jn}) \qquad (j = 1, \ldots, m) .$$

Dann heißen

$$Srg(A) := \dim < a^{(1)}, \ldots, a^{(n)} >$$

der *Spaltenrang* von A und

$$Zrg(A) := \dim < A_1, \ldots, A_m >$$

der *Zeilenrang* von A.

Bemerkung 8.13 Nach B.8.9 gilt $Rg(A) = Srg(A)$, denn

$$\text{Bild}(A) = \text{Bild}(T) =< T(e_1), \ldots, T(e_n) >$$

und $T(e_k) = a^{(k)} (k = 1, \ldots, n)$. Also ist

$$Rg(A) = \dim(\text{Bild}(A)) = \dim < a^{(1)}, \ldots, a^{(n)} >= Srg(A) .$$

Unser Ziel ist es nun, zu zeigen, dass auch $Rg(A) = Zrg(A)$ gilt. Dazu brauchen wir den Begriff der Transponierten einer Matrix, der von allgemeiner Bedeutung ist.

Definition 8.14 Es sei $A = (a_{jk}) \in K^{m \times n}$. Dann heißt die Matrix $A^T = (b_{jk}) \in K^{n \times m}$, definiert durch

$$b_{jk} := a_{kj} \qquad (j = 1, \ldots, n, k = 1, \ldots, m)$$

die *Transponierte* von A.

Für $x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \in K^{m \times 1}$ ist insbesondere damit

$$x^T = (x_1, \ldots, x_m) \in K^{1 \times m} ,$$

d. h. x^T ist "x als Zeilenvektor".

Wir stellen einige elementare Eigenschaften zusammen.

Satz 8.15 *1. Für $A, B \in K^{m \times n}$ gilt $(A + B)^T = A^T + B^T$.*
2. Für $A \in K^{m \times n}$ gilt $(A^T)^T = A$.
3. Für $A \in K^{m \times n}, B \in K^{n \times p}$ gilt $(AB)^T = B^T A^T$.

Beweis.
1. und 2. sind klar.
3. Ist $A = (a_{jk}), B = (b_{jk})$, so gilt

$$(AB)^T = \left(\sum_{\nu=1}^{n} a_{j\nu} b_{\nu k}\right)^T = \left(\sum_{\nu=1}^{n} a_{k\nu} b_{\nu j}\right) = \left(\sum_{\nu=1}^{n} b_{\nu j} a_{k\nu}\right) = B^T A^T \ .$$

□

Es glt nun folgender zentrale Satz.

Satz 8.16 *Ist $A \in K^{m\times n}$, so gilt*

$$Rg(A) = Rg(A^T) \ .$$

Beweis.
Es sei $A = (a_{jk})$, und es seien

$$a^{(k)} = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} \qquad (k = 1, \dots, n)$$

die Spalten von A sowie

$$A_j = (a_{j1}, \dots, a_{jn}) \qquad (j = 1, \dots, m)$$

die Zeilen von A. Dann sind $A_j^T (j = 1, \dots, m)$ die Spalten von A^T. Es sei

$$U :=< A_1^T, \dots, A_m^T > \subset K^n \ .$$

Dann ist $s := \dim(U) = Rg(A^T)$ nach B.8.13. Es sei $(u_1, \dots, u_s)$ eine Basis von U. Dann existieren $c_{j\nu} \in K$ mit

$$A_j = \sum_{\nu=1}^{s} c_{j\nu} u_\nu^T \qquad (j = 1, \dots, m) \ ,$$

d. h. mit $u_\nu^T = (u_{\nu 1}, \dots, u_{\nu n})$ $(\nu = 1, \dots, s)$ gilt

$$(a_{j1}, \dots, a_{jn}) = \sum_{\nu=1}^{s} c_{j\nu} (u_{\nu 1}, \dots, u_{\nu n}) \qquad (j = 1, \dots, m) \ .$$

Vergleich der einzelnen Komponenten ergibt

$$a_{jk} = \sum_{\nu=1}^{s} c_{j\nu} u_{\nu k} \qquad (j = 1, \dots, m; k = 1, \dots, n)$$

d. h.

$$\begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} = \sum_{\nu=1}^{s} u_{\nu k} \begin{pmatrix} c_{1\nu} \\ \vdots \\ c_{m\nu} \end{pmatrix} \qquad (k = 1, \ldots, n) \; .$$

Mit

$$c_\nu := \begin{pmatrix} c_{1\nu} \\ \vdots \\ c_{m\nu} \end{pmatrix} \qquad (\nu = 1, \ldots, s)$$

gilt also

$$a^{(k)} = \sum_{\nu=1}^{s} u_{\nu k} c_\nu \qquad (k = 1, \ldots, n) \; .$$

Damit ist

$$\text{Bild}(A) =< a^{(1)}, \ldots, a^{(n)} > \subset < c_1, \ldots, c_s > \; ,$$

also (nach dem Basisauswahlsatz)

$$Rg(A) = \dim(\text{Bild}(A)) \leq s = Rg(A^T) \; .$$

Aus $A = (A^T)^T$ ergibt sich durch die gleiche Argumentation

$$Rg(A^T) \leq Rg((A^T)^T) = Rg(A) \; ,$$

also insgesamt die Behauptung. □

Hieraus ergibt sich sofort

Folgerung 8.17 Ist $A \in K^{m \times n}$, so gilt

$$Rg(A) = Srg(A) = Zrg(A) \; .$$

(Denn: Es gilt offenbar $Zrg(A) = Srg(A^T)$. Nach B.8.13 und S.8.16 ist

$$Zrg(A) = Srg(A^T) = Rg(A^T) = Rg(A) \; ,$$

also mit B.8.13:

$$Rg(A) = Srg(A) = Zrg(A) \; .)$$

Eine wichtige Aussage über den Rang des Produktes liefert

Satz 8.18 *Sind $A \in K^{m\times n}$ und $B \in K^{n\times p}$, so gilt*

$$Rg(AB) \leq \min(Rg(A), Rg(B)) .$$

Beweis.
Wir setzen

$$T(x) := Ax \quad (x \in K^n) , \qquad S(x) := Bx \quad (x \in K^p) .$$

Dann gilt nach S.8.8 bzw. B.D.8.9

$$(T \circ S)(x) = ABx \qquad (x \in K^p) .$$

Aus $\text{Bild}(T \circ S) \subset \text{Bild}(T)$ folgt

$$Rg(AB) = \dim(\text{Bild}(T \circ S)) \leq \dim(\text{Bild}(T)) = Rg(A) .$$

Weiter ist mit S.8.16 durch die gleiche Argumentation

$$Rg(AB) = Rg((AB)^T) = Rg(B^T A^T) \leq Rg(B^T) = Rg(B) .$$

also insgesamt

$$Rg(AB) \leq \min(Rg(A), Rg(B)) .$$

□

Bemerkung und Definition 8.19 Für $p \in \mathbb{N}$ heißt

$$E := E_p := (\delta_{jk})_{j,k=1,\dots,p} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

(p-reihige) Einheitsmatrix. Es gilt damit für $A \in K^{m\times n}$

$$AE_n = A \qquad \text{und} \qquad E_m A = A .$$

Definition 8.20 Es sei $A \in K^{n\times n}$. Dann heißt A *invertierbar* falls eine Matrix $B \in K^{n\times n}$ mit $AB = E_n$ existiert.

Satz 8.21 *Es sei $A \in K^{n\times n}$. Ist A invertierbar, so existiert genau ein $B \in K^{n\times n}$ mit $AB = E_n$. Außerdem gilt $CA = E_n$ genau dann, wenn $B = C$ ist.*

Beweis.
1. Es sei $B \in K^{n\times n}$ so, dass $AB = E$. Wir betrachten $T, S \in L(K^n)$ mit

$$T(x) := A \cdot x \ , \ S(x) := B \cdot x \qquad (x \in K^n) \ .$$

(Es gilt nach B.D.8.9: $\varphi(T) = A, \varphi(S) = B$, wobei $\varphi = \varphi_{M,M}$ und M die kanonische Basis in K^n sind.)
Dann gilt nach S.8.8: AB ist Matrix von $T \circ S$ (bzgl. kanonischer Basen), also nach B.8.9

$$(T \circ S)(x) = ABx = Ex = x \qquad (x \in K^n) \ .$$

d. h. $T \circ S = \mathrm{id}_{K^n}$. Damit ist T surjektiv und nach S.7.5 auch bijektiv. Also ist $S = T^{-1}$ und damit ist B Matrix zu T^{-1}, d. h. $\varphi(T^{-1}) = B$. Insbesondere folgt für jedes $\tilde{B}$ mit $A\tilde{B} = E$ genauso $\tilde{B} = \varphi(T^{-1})$, also $B = \tilde{B}$.
2. Aus $T \circ T^{-1} = \mathrm{id}_{K^n} = T^{-1} \circ T$ folgt

$$BA = \varphi(T^{-1})\varphi(T) = \varphi(T^{-1} \circ T) = \varphi(\mathrm{id}_{K^n}) = E \ .$$

3. Ist $C \in K^{n\times n}$ mit $CA = E$, so ist nach 1. und 2., angewandt auf (C, A) statt (A, B), auch $AC = E$, also ist $C = B$ nach 1. □

Definition 8.22 Ist $A \in K^{n\times n}$ invertierbar, so heißt die (nach S.8.21 eindeutig bestimmte) Matrix B mit $AB = E$ *Inverse* zu A. Wir schreiben $B =: A^{-1}$.

Beispiel 8.23 Es sei $A \in \mathbb{R}^{2\times 2}$ mit $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Dann gilt

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E, \quad \text{also} \quad A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \ .$$

Satz 8.24 *Es seien $A, B \in K^{n\times n}$ invertierbar. Dann gilt*
1. A^{-1} ist invertierbar mit $(A^{-1})^{-1} = A$.
2. AB ist invertierbar mit $(AB)^{-1} = B^{-1}A^{-1}$.
3. Ist $\lambda \in K \setminus \{0\}$, so ist λA invertierbar mit $(\lambda A)^{-1} = \lambda^{-1}A^{-1}$.

Beweis.
1. Nach S.8.21 ist $AA^{-1} = E = A^{-1}A$, also ist nach Definition A^{-1} invertierbar mit $(A^{-1})^{-1} = A$.
2. Es gilt $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AEA^{-1} = AA^{-1} = E$, also ist nach Definition AB invertierbar mit $(AB)^{-1} = B^{-1}A^{-1}$.
3. Es gilt $(\lambda A)(\frac{1}{\lambda} \cdot A^{-1}) = (\lambda\frac{1}{\lambda})(AA^{-1}) = E$, also ist λA invertierbar mit $(\lambda A)^{-1} = \lambda^{-1}A^{-1}$. □

Bemerkung 8.25 Aus S.8.10 und S.8.24 folgt, dass die Menge

$$GL_n = GL_n(K) := \{A \in K^{n\times n} : A \text{ invertierbar}\}$$

mit der Matrixmultiplikation als Verknüpfung eine Gruppe bildet (mit E als neutralem Element). GL_n ist für $n \geq 2$ i. a. nicht kommutativ.

Der folgende Satz ist i. w. eine Neuformulierung von S.7.5.

Satz 8.26 *Es sei $A \in K^{n\times n}$. Dann sind äquivalent:*
a) A ist invertierbar,
b) $Rg(A) = n$,
c) $Kern(A) = \{0\}$.

Beweis.
Es sei $T \in L(K^n)$ mit $T(x) = Ax$ $(x \in K^n)$.
1. a) $\Rightarrow$ b): Ist A invertierbar, so gilt (vgl. Beweis zu S.8.21): T ist bijektiv. Also ist T insbesondere surjektiv und damit

$$Rg(A) = \dim(\text{Bild}(A)) = \dim \text{Bild}(T) = n\ .$$

2. b) $\Rightarrow$ c): Ist $Rg(A) = \dim(\text{Bild}(T)) = n$, so gilt nach der Dimensionsformel für lineare Abbildungen

$$\dim(\text{Kern}(T)) = n - \dim(\text{Bild}(T)) = 0\ ,$$

also $\text{Kern}(T) = \text{Kern}(A) = \{0\}$.
3. c $\Rightarrow$ a): Ist $\text{Kern}(T) = \text{Kern}(A) = \{0\}$, so ist T injektiv also auch bijektiv nach S.7.5. Ist $S = T^{-1}$, so gilt für die Matrix B von S (bzgl. der kanonischen Basen)

$$AB = \varphi(T)\varphi(T^{-1}) = \varphi(T \circ T^{-1}) = \varphi(Id_{K^n}) = E\ ,$$

also ist A invertierbar. □

Satz 8.27 *Es sei $A \in K^{n\times n}$ invertierbar. Dann gilt für $B \in K^{n\times p}$ und $C \in K^{m\times n}$*

$$Rg(AB) = Rg(B) \quad \textit{und} \quad Rg(CA) = Rg(C)\ .$$

Beweis.
Aus S.8.18 folgt

$$Rg(B) = Rg(A^{-1}(AB)) \leq Rg(AB) \leq Rg(B)$$

und

$$Rg(C) = Rg((CA)A^{-1}) \leq Rg(CA) \leq Rg(C)\ .$$

□

9 Lineare Gleichungssysteme

Wir kommen nun zu einem der Kernthemen der linearen Algebra, den linearen Gleichungssystemen. Wir werden sehen, dass für eine systematische Behandlung solcher Systeme die in den ersten Abschnitten dargestellten Begriffe und Ergebnisse sehr nützlich sind.

Definition 9.1 Es sei K ein Körper (wobei wir jetzt meist $K = \mathbb{R}$ oder $K = \mathbb{C}$ betrachten), und es seien $a_{jk} \in K, b_j \in K$ $(j = 1, \dots, m;\ k = 1, \dots n)$. Dann heißt das System von Gleichungen

$$\begin{array}{ccccccc} a_{11}x_1 & +\cdots+ & a_{1n}x_n & = & \sum\limits_{k=1}^{n} a_{1k}x_k & = b_1 & \\ a_{21}x_1 & +\cdots+ & a_{2n}x_n & = & \sum\limits_{k=1}^{n} a_{2k}x_k & = b_2 & \\ \vdots & & \vdots & & \vdots & \vdots & \\ a_{m1}x_1 & +\cdots+ & a_{mn}x_n & = & \sum\limits_{n=1}^{n} a_{mk}x_k & = b_m & \end{array} \qquad (LGS)$$

ein *lineares Gleichungssystem (LGS) (in den Unbekannten $x_1, \dots, x_n$)*

Ein Vektor $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$, für den (LGS) erfüllt ist, heißt *Lösung* des LGS. Das LGS heißt *homogen*, falls $b_j = 0$ für $j = 1, \dots, m$ gilt; ansonsten heißt es *inhomogen.*

Die Verbindung zu dem bisherigen Inhalt der Vorlesung ist offensichtlich.
Ist $A = (a_{jk}), b = (b_1, \dots, b_m)^T$ und $x = (x_1, \dots, x_n)^T$, so lässt sich (LGS) auch schreiben als

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

oder kurz als

$$Ax = b\ .$$

Die Matrix A heißt *(Koeffizienten-) Matrix* des LGS. Die $(m \times (n+1))$ Matrix $(A|b)$ mit

$$(A|b) = \begin{pmatrix} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{mn} & \cdots & a_{mn} & b_m \end{pmatrix}$$

heißt *erweiterte Koeffizientenmatrix* des LGS.

Ein erstes Lösbarkeitskriterium liefert

Satz 9.2 *Es seien $A = (a_{jk}) \in K^{m\times n}$ und $b \in K^m$. Dann sind aquivalent:*

a) $Ax = b$ ist lösbar (d. h. es existiert ein $x \in K^n$ mit $Ax = b$) ,

b) $b \in < a^{(1)}, \ldots, a^{(n)} >$, wobei $a^{(k)}$ die Spalten von A sind,

c) $Rg(A) = Rg(A|b)$.

Beweis.
1. die Äquivalenz von a) und b) ergibt sofort daraus, dass man $Ax = b$ auch schreiben kann als

$$\sum_{k=1}^{n} x_k a^{(k)} = b \, .$$

2. b) $\Rightarrow$ c): Ist $b \in < a^{(1)}, \ldots, a^{(n)} >$, so ist

$$< a^{(1)}, \ldots, a^{(n)} > = < a^{(1)}, \ldots, a^{(n)}, b > \, ,$$

also $Rg(A) = \dim(< a^{(1)}, \ldots, a^{(n)} >) = \dim(< a^{(1)}, \ldots, a^{(n)}, b >) = Rg(A|b)$.

3. c) $\Rightarrow$ b): Da $< a^{(1)}, \ldots, a^{(n)} > \subset < a^{(1)}, \ldots, a^{(n)}, b >$ gilt, folgt aus

$$\dim(< a^{(1)}, \ldots, a^{(n)} >) = \dim(< a^{(1)}, \ldots, a^{(n)}, b >)$$

bereits $< a^{(1)}, \ldots, a^{(n)} > = < a^{(1)}, \ldots, a^{(n)}, b >$ nach S.5.17, also ist

$$b \in < a^{(1)}, \ldots, a^{(n)} > \, .$$

□

Beispiel 9.3 Wir betrachten das LGS

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

also

$$Ax = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 6 \\ 6 \end{pmatrix} = b \, .$$

Dann gilt

$$b = \begin{pmatrix} 6 \\ 6 \\ 6 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$$

also $b \in < \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} >$, und eine Lösung ist gegeben durch

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} .$$

Wir wollen zunächst einige allgemeine Aussagen über die Struktur der Lösungsmenge eines LGS machen.

Satz 9.4 *Es seien $A \in K^{m\times n}$ und $b \in K^m$. Wir setzen*

$$Lös(A,b) := \{x \in K^n : Ax = b\}$$

($Lös(A,b)$ heißt Lösungsmenge des LGS). Dann gilt
1. $Lös(A,0)$ ist ein Unterraum von K^n mit $\dim(Lös(A,0)) = n - RgA$.
2. Ist $Lös(A,b) \neq \emptyset$, und ist $y \in Lös(A,b)$, so ist

$$Lös(A,b) = y + Lös(A,0) := \{y + x : x \in Lös(A,0)\}\ (= \{y + x : Ax = 0\})$$

Beweis.
1. Es gilt

$$\text{Lös}(A,0) = \text{Kern}(A) = \text{Kern}(T) ,$$

wobei $T(x) = Ax$ ($x \in K^n$). Also ist nach S.7.2 Lös$(A,0)$ ein Unterraum von K^n. (Insbesondere ist stets $0 \in$ Lös$(A,0)$!).
Aus der Dimensionsformel für lineare Abbildungen (S.7.4) folgt weiter

$$\dim(\text{Lös}(A,0)) = \dim(\text{Kern}(A)) = n - \dim(\text{Bild}(A)) = n - Rg(A) .$$

2. "$\subset$": Es sei $z \in$ Lös(A,b). Dann gilt für $x := z - y$:

$$Ax = A(z-y) = Az - Ay = b - b = 0 ,$$

also $x \in \text{Lös}(A,0)$ und $z = y + x$, d. h. $z \in y + \text{Lös}(A,0)$.
3. "$\supset$" : Ist $z \in y + \text{Lös}(A,0)$, d. h. es existiert ein $x \in \text{Lös}(A,0)$ mit $z = y + x$, so gilt

$$Az = A(y + x) = Ay + Ax = Ay = b \ ,$$

also $z \in \text{Lös}(A,b)$. □

Die Aussage 2. besagt anschaulich folgendes: Ist y eine spezielle Lösung des inhomogenen Systems $Ax = b$, so ergeben sich alle Lösungen z durch $z = y + x$, wobei x die Lösungen des homogenen Systems "durchläuft", d. h. ist $r := Rg(A) < n$, und ist $(v_1, \dots, v_{n-r})$ eine Basis von $\text{Lös}(A,0)$, so ist

$$\text{Lös}(A,b) = \{y + \sum_{\nu=1}^{n-r} \lambda_\nu v_\nu : \lambda_\nu \in K \text{ für } \nu = 1, \dots, n-r\} \ .$$

Ist $r = n$, so ist $\text{Lös}(A,0) = \{0\}$ und $\text{Lös}(A,b) = \{y\}$.

Satz 9.5 *Es sei $A \in K^{m\times n}$. Dann gilt*
1. $Ax = b$ ist für alle $b \in K^m$ genau dann lösbar, wenn $Rg(A) = m$ ist.
2. $Ax = b$ hat für alle $b \in K^m$ höchstens eine Lösung genau dann, wenn $Rg(A) = n$ ist.
3. $Ax = b$ ist für alle $b \in K^m$ eindeutig lösbar genau dann, wenn $Rg(A) = n = m$ (insbesondere also $n = m$) gilt. In diesem Fall ist

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

die eindeutig bestimmte Lösung.

Beweis.
1. $Ax = b$ ist nach S.9.2 genau dann für alle $b \in K^m$ lösbar, wenn

$$K^m =< a^{(1)}, \dots, a^{(n)} >= \text{Bild}(A) \ .$$

Da $\text{Bild}(A)$ stets ein Unterraum von K^m ist, ist dies (S.5.17) wiederum äquivalent zu

$$Rg(A) = \dim(\text{Bild}(A)) = \dim(K^m) = m \ .$$

2. Nach S.9.4.1. ist $\text{Lös}(A,0) = \{0\}$ genau dann, wenn $Rg(A) = n$ ist. Nach S.9.4.2. ist dies genau dann der Fall, wenn $\text{Lös}(A,b)$ höchstens einpunktig ist.
3. Nach 1. und 2. ist $Ax = b$ eindeutig lösbar für alle $b \in K^m$ genau dann, wenn $Rg(A) = n = m$ gilt.
Gilt nun $Rg(A) = n = m$, so ist $A \in K^{n\times n}$ invertierbar nach S.8.26. Ist $b \in K^m$, so gilt für $x = A^{-1}b$

$$Ax = AA^{-1}b = b \ ,$$

d. h. $\text{Lös}(A,b) = \{A^{-1}b\}$. □

Definition 9.6 Es seien $A, \tilde{A} \in K^{m\times n}$ und $b, \tilde{b} \in K^m$. Dann heißen die LGS'e $Ax = b$ und $\tilde{A}x = \tilde{b}$ *äquivalent*, falls

$$\text{Lös}(A, b) = \text{Lös}(\tilde{A}, \tilde{b})$$

gilt.

Satz 9.7 *Ist $C \in K^{m\times m}$ invertierbar, und sind $A \in K^{m\times n}$ sowie $b \in K^m$, so ist $Ax = b$ äquivalent zu $CAx = Cb$.*

Beweis.
Ist $x \in \text{Lös}(A, b)$, so gilt $CAx = Cb$. Ist umgekehrt $x \in \text{Lös}(CA, Cb)$, so ist

$$Ax = (C^{-1}C)Ax = C^{-1}(CAx) = C^{-1}(Cb) = b$$

also $x \in \text{Lös}(A, b)$. □

Wir haben uns bisher beschränkt auf Strukturaussagen über die Lösungsmenge von Gleichungssystemen. Es drängt sich nun die interesannte Frage auf, wie man an Lösungen herankommt.

10 Der Gauß'sche Algorithmus

Definition 10.1 Es sei $A \in K^{m\times n}$.
1. Unter den *elementaren Zeilenoperationen* (oder *elementaren Zeilenumformungen*) verstehen wir die Operationen

$Z_j \leftrightarrow Z_k$:	Vertauschen der Zeilen j und k
$Z_j + \lambda Z_k$:	Addition des λ-fachen der k-ten Zeile zur j-ten Zeile ($\lambda \in K$; $j \neq k$)
λZ_j :	Multiplikation der j-ten Zeile mit $\lambda \in K \setminus \{0\}$.

2. Die *elementaren Spaltenoperationen* $S_j \leftrightarrow S_k, S_j + \lambda S_k, \lambda S_j$ sind analog mit "Spalte" statt "Zeile" definiert.

Beispiel 10.2 Es sei

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

Dann läßt sich A etwa folgendermaßen mittels elementarer Zeilenoperationen transformieren:

$$\begin{array}{l} \\ \overset{Z_1 \leftrightarrow Z_2}{\longrightarrow} \\ \overset{Z_3 - Z_1}{\longrightarrow} \\ \overset{Z_3 - 2Z_2}{\longrightarrow} \\ \overset{(-\frac{1}{4})Z_3}{\longrightarrow} \end{array} \begin{pmatrix} 1 & -1 & 3 & -1 & -2 \\ 0 & -1 & 1 & -1 & 0 \\ 1 & 1 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 & 3 & -1 & -2 \\ 0 & 1 & 1 & -1 & 0 \\ 0 & 2 & -2 & 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 & 3 & -1 & -2 \\ 0 & 1 & 1 & -1 & 0 \\ 0 & 0 & -4 & 4 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 & 3 & -1 & -2 \\ 0 & 1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 & -1 \end{pmatrix}$$

Die elementaren Zeilen- (Spalten-) Operationen lassen sich jeweils durch eine Matrixmultiplikation beschreiben:

Definition 10.3 Es sei $m \in \mathbb{N}$. Wir schetzen für $j, k = 1, \ldots, m$

$$E^{(j,k)} := (e_{\nu\mu})_{\nu,\mu=1,\ldots,m}$$

mit $e_{\nu\mu} := 1$ falls $(\nu, \mu) = (j, k)$ und $e_{\nu\mu} := 0$ sonst, d. h.

$$E^{(j,k)} = \begin{pmatrix} 0 & \ldots & & & & & & 0 \\ \vdots & & & & & & & \\ \vdots & & & 0 & & & & \vdots \\ 0 & \ldots & 0 & 1 & 0 & \ldots & \ldots & 0 \\ \vdots & & & 0 & & & & \vdots \\ \vdots & & & & & & & \\ 0 & \ldots & & & & & & 0 \end{pmatrix} \leftarrow j - \text{te Zeile}$$

$$\uparrow$$

k-te Spalte

Dann heißen für $j,k = 1,\dots,m$ und $\lambda \in K$ die Matrizen

$$P^{(j,k)} := E - E^{(j,j)} - E^{(k,k)} + E^{(j,k)} + E^{(k,j)}$$

$$= \begin{pmatrix}
1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\
0 & \ddots & & & & & & & & & \vdots \\
\vdots & & 1 & & & & & & & & \vdots \\
\vdots & & & 0 & \dots & & 0 & 1 & & & \vdots \\
\vdots & & & \vdots & 1 & & & \vdots & & & \vdots \\
\vdots & & & \vdots & & \ddots & & \vdots & & & \vdots \\
\vdots & & & 0 & & & 1 & \vdots & & & \vdots \\
\vdots & & & 1 & 0 & \dots & 0 & 0 & & & \vdots \\
\vdots & & & & & & & & 1 & & \vdots \\
 & & & & & & & & & \ddots & 0 \\
0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 & 1
\end{pmatrix}
\begin{matrix} \\ \\ \\ \\ \leftarrow j \\ \\ \\ \\ \leftarrow k \\ \\ \\ \end{matrix}$$

$$F^{(j,k)}(\lambda) := E + \lambda E^{(j,k)} = \begin{pmatrix}
1 & 0 & \dots & \dots & & 0 \\
0 & 1 & & \lambda & & \vdots \\
\vdots & & \ddots & & & \vdots \\
\vdots & & & & & \vdots \\
\vdots & & & & 1 & 0 \\
0 & \dots & \dots & \dots & 0 & 1
\end{pmatrix} \qquad (j \neq k)$$

sowie

$$F^{(j)}(\lambda) := E + (\lambda - 1)E^{(j,j)} = \begin{pmatrix}
1 & 0 & \dots & \dots & \dots & \dots & 0 \\
\vdots & \ddots & & & & & \vdots \\
\vdots & & 1 & & & & \vdots \\
\vdots & & & \lambda & & & \vdots \\
\vdots & & & & 1 & & \vdots \\
\vdots & & & & & \ddots & 0 \\
0 & \dots & \dots & \dots & \dots & 0 & 1
\end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow j \\ \\ \\ \\ \end{matrix} \quad (\lambda \neq 0)$$

$$\uparrow$$
$$j$$

$(m \times m)$-Elementarmatrizen.

Satz 10.4 *Es sei $A \in K^{m\times n}$. Dann gilt:*
Die elementaren Zeilenumformungen lassen sich wie folgt beschreiben

$$\begin{array}{lll} A & \overset{Z_j\leftrightarrow Z_k}{\longrightarrow} & P^{(j,k)}\cdot A \\ A & \overset{Z_j+\lambda Z_k}{\longrightarrow} & F^{(j,k)}(\lambda)\cdot A \\ A & \overset{\lambda Z_j}{\longrightarrow} & F^{(j)}(\lambda)\cdot A \end{array}$$

Entsprechend gilt für die elementaren Spaltenumformungen

$$\begin{array}{lll} A & \overset{S_j\leftrightarrow S_k}{\longrightarrow} & A\cdot P^{(k,j)} \\ A & \overset{S_j+\lambda S_k}{\longrightarrow} & A\cdot F^{(k,j)}(\lambda) \\ A & \overset{\lambda S_j}{\longrightarrow} & A\cdot F^{(j)}(\lambda) \end{array}$$

Beweis.
Wir beschränken uns auf Zeilenoperationen. Es gilt für $j,k = 1,\dots,m$

$$E^{(j,k)}\cdot A = \begin{pmatrix} 0 & \dots & \dots & \dots & 0 & \dots & 0 \\ \vdots & & & & \vdots & & \vdots \\ 0 & \dots & \dots & 0 & 1 & \dots & 0 \\ \vdots & & & & 0 & & \vdots \\ \vdots & & & & \vdots & & \vdots \\ \vdots & & & & \vdots & & \vdots \\ 0 & \dots & \dots & \dots & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} a_{11} & \dots & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{j1} & \dots & & a_{jn} \\ \vdots & & & \vdots \\ a_{k1} & \dots & & a_{kn} \\ \vdots & & & \vdots \\ a_{m1} & \dots & & a_{mn} \end{pmatrix} =$$

$$= \begin{pmatrix} 0 & \dots & \dots & 0 \\ \vdots & & & \vdots \\ a_{k1} & \dots & \dots & a_{kn} \\ 0 & \dots & \dots & 0 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ A_k \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} \leftarrow j-\text{te Zeile} \quad ,$$

d. h. $E^{(j,k)}\cdot A$ "schneidet aus A die k-te Zeile A_k aus und setzt diese in die j-te Zeile".
Hieraus ergibt sich die Behauptung durch Nachrechnen. □

Wichtig für das folgende ist weiter

Satz 10.5 *Die Elementarmatrizen sind invertierbar, und es gilt*

$$(P^{(j,k)})^{-1} = P^{(j,k)}\ , \quad (F^{(j,k)}(\lambda))^{-1} = F^{(j,k)}(-\lambda) \qquad (j\neq k)$$

sowie

$$(F^{(j)}(\lambda))^{-1} = F^{(j)}(1/\lambda) \qquad (\lambda \neq 0)\ .$$

Beweis.. Es gilt $P^{(j,k)} \cdot P^{(j,k)} = E, F^{(j,k)}(\lambda)F^{(j,k)}(-\lambda) = E$ und $F^{(j)}(\lambda)F^{(j)}(1/\lambda) = E$. □

Folgerung 10.6 Es seien $A, \tilde{A} \in K^{m\times n}$ und $b, \tilde{b} \in K^m$. Ist die Matrix $(\tilde{A}|\tilde{b}) \in K^{m\times(n+1)}$ durch elementare Zeilenoperationen aus $(A|b)$ entstanden, so sind die Gleichungssysteme

$$Ax = b \qquad \text{und} \qquad \tilde{A}x = \tilde{b}$$

äquivalent.
(Denn: Dies ergibt sich durch Kombination von S.9.7, S.10.4 und S. 10.5 und S.8.24.2. Man beachte dabei, dass $C(A|b) = (CA|Cb)$ gilt.)

Wir betrachten zunächst Gleichungssysteme

$$Ax = b$$

mit invertierbarer Matrix $A \in K^{n\times n}$ und $b \in K^n$. Nach S.9.5 ist die (eindeutig bestimmte) Lösung gegeben durch $x = A^{-1}b$. Allerdings erweist sich die Berechnung von A^{-1} als i. a. sehr aufwendig. Das im folgenden dargestellte Verfahren ist weniger rechenintensiv und hat zudem den Vorteil, dass es leicht modifiziert für allgemeine Matrizen $A \in K^{m\times n}$ anwendbar ist.

Satz 10.7 *Es seien $A \in K^{n\times n}$ und $b \in K^n$ mit $Rg(A) = n$ (d. h. A ist invertierbar). Dann lässt sich $(A|b)$ durch elementare Zeilenoperationen in eine Matrix $(R|c)$ überführen, wobei $R \in K^{n\times n}$ eine obere Dreiecksmatirx ist, d. h. R hat die Form*

$$R = \begin{pmatrix} r_{11} & \dots & \dots & r_{1n} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & & \vdots \\ 0 & \dots & 0 & r_{nn} \end{pmatrix},$$

(also $r_{jk} = 0$ für $j > k$) und $r_{jj} \neq 0$ für $j = 1, \dots, n$. Insbesondere sind die Systeme $Ax = b$ und $Rx = c$ äquivalent.

Beweis.
Zunächst können wir durch eventuelles Vertauschen zweier Zeilen erreichen, dass $(A|b)$

übergeht in $(\tilde{A}|\tilde{b})$ mit $\tilde{a}_{11} \neq 0$. (Beachte: wäre $a_{j1} = 0$ für $j = 1, \ldots, n$, so wäre $\mathrm{Rg}\,(A) < n$.) Nun subtrahieren wir das $\frac{\tilde{a}_{j1}}{\tilde{a}_{11}}$-fache der ersten Zeile von $(\tilde{A}|\tilde{b})$ von der j-ten Zeile von $(\tilde{A}|\tilde{b})$ $(j = 2, \ldots, n)$. Dann entsteht eine Matrix $(A_1|b_1)$ der Form

$$(A_1|b_1) = \begin{pmatrix} a_{11}^{(1)} & & \ldots & a_{1n}^{(1)} & b_1^{(1)} \\ 0 & a_{22}^{(1)} & \ldots & a_{2n}^{(1)} & b_2^{(1)} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & a_{n2}^{(1)} & \ldots & a_{nn}^{(1)} & b_n^{(1)} \end{pmatrix},$$

d. h. der erste Eintrag $a_{j1}^{(1)}$ der j-ten Zeile ist 0 für $j = 2, \ldots, n$ (und $a_{11}^{(1)} \neq 0$). Nun verfahren wir entsprechend mit der $((n-1) \times n)$-Matrix

$$(B_1|d_1) := \begin{pmatrix} a_{22}^{(1)} & \ldots & a_{2n}^{(1)} & b_2^{(1)} \\ \vdots & & \vdots & \vdots \\ a_{n2}^{(1)} & \ldots & a_{nn}^{(1)} & b_n^{(1)} \end{pmatrix}$$

anstelle von $(A|b)$. Dabei ist wichtig, dass nach S.8.27 und S.10.4/5 $\mathrm{Rg}\,(A_1) = \mathrm{Rg}\,(A) = n$ gilt. Damit ist $\mathrm{Rg}\,(B_1) = n - 1$ (wäre $\mathrm{Rg}\,(B_1) < n - 1$, so wäre eine Zeile von B_1 Linearkombination der restlichen, was dann auch für A_1 gelten würde). Es entsteht nach diesem Umformen eine Matrix $(A_2|b_2)$ der Form

$$(A_2|b_2) = \begin{pmatrix} a_{11}^{(1)} & \ldots & & & a_{1n}^{(1)} & b_1^{(1)} \\ 0 & a_{22}^{(2)} & \ldots & & a_{2n}^{(2)} & b_2^{(2)} \\ \vdots & 0 & a_{33}^{(2)} & \ldots & a_{3n}^{(2)} & b_3^{(2)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & a_{n3}^{(2)} & \ldots & a_{nn}^{(2)} & b_n^{(2)} \end{pmatrix}$$

mit $a_{22}^{(2)} \neq 0$ und $a_{j1}^{(2)} = 0$ für $j = 2, \ldots, n$ sowie $a_{j2}^{(2)} = 0$ für $j = 3, \ldots, n$. So fortfahrend erhalten wir eine Kette von Matrizen $(A_m|b_m)$ der Form

$$(A_m|b_m) = \begin{pmatrix} \otimes & \ldots & * & * & \ldots & * & * \\ 0 & \ddots & & & & & \vdots \\ \vdots & \ddots & \otimes & * & \ldots & * & * \\ \vdots & & 0 & * & \ldots & * & * \\ \vdots & & \vdots & \vdots & & & \vdots \\ 0 & & 0 & * & \ldots & * & * \end{pmatrix} \leftarrow m \qquad (m = 1, \ldots, n-1)$$

wobei die Einträge $\otimes$ nicht verschwinden. Dabei ist $(A_m|b_m)$ aus $(A_0|b_0) := (A|b)$ durch elementare Zeilenumformungen (der Form $Z_j \leftrightarrow Z_k$ und $Z_j + \lambda Z_k$) entstanden.

Für $m = n - 1$ ist

$$(A_{n-1}|b_{n-1}) = \begin{pmatrix} \otimes & & \dots & \dots & * & * \\ 0 & \ddots & & & \vdots & \vdots \\ \vdots & \ddots & & & \vdots & \vdots \\ \vdots & & & & * & \vdots \\ 0 & \dots & \dots & 0 & \otimes & * \end{pmatrix} =: (R|c)$$

mit $R = (r_{jk})$, $c = (c_1, \dots, c_n)^T$ und $r_{j_k} = 0$ für $j > k$ und $k = 1, \dots, n-1$ sowie $r_{jj} \neq 0$ für $j = 1, \dots, n$. Nach F.10.6 ist $Rx = c$ äquivalent zu $Ax = b$. □

Bemerkung 10.8 Das im Beweis dargestellte Verfahren zur Berechnung von $(R|c)$ aus $(A|b)$ heißt *Gauß- Verfahren* oder *Gauß- Algorithmus*. Hat man das LGS $Ax = b$ durch das Gauß-Verfahren in das äquivalente System $Rx = c$ überführt, so läss t sich die Lösung

$$x = A^{-1}b = R^{-1}c$$

leicht berechnen: Die Lösung $x = (x_1, \dots, x_n)^T \in K^n$ erfüllt

$$Rx = \begin{pmatrix} r_{11} & \dots & & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & r_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = c$$

genau dann, wenn

$$x_n = c_n / r_{nn}$$

und

$$x_j = \frac{1}{r_{jj}}(c_j - \sum_{k=j+1}^{n} r_{jk} x_k) \qquad (j = n-1, n-2, \dots, 1)\ ,$$

d. h. x lässt sich sukzessive "von hinten nach vorne" berechnen, da x_j nur von den dann bekannten Größen r_{jk}, c_j und $x_{j+1}, \dots, x_n$ abhängt!

Jetzt endlich ein

Beispiel 10.9 Wir betrachten das LGS

$$\begin{array}{rcrcrcr} 3x_1 & + & x_2 & - & 3x_3 & = & 8 \\ -6x_1 & & & + & 5x_3 & = & -11\ , \\ 3x_1 & + & 5x_2 & - & 7x_3 & = & 20 \end{array}$$

also

$$(A|b) = (A_0|b_0) = \begin{pmatrix} 3 & 1 & -3 & 8 \\ -6 & 0 & 5 & -11 \\ 3 & 5 & -7 & 20 \end{pmatrix} .$$

1. Schritt $(Z_2 + 2Z_1;\ Z_3 - Z_1)$:

$$(A_1|b_1) = \begin{pmatrix} 3 & 1 & -3 & 8 \\ 0 & 2 & -1 & 5 \\ 0 & 4 & -4 & 12 \end{pmatrix} ,$$

2. Schritt $(Z_3 - 2Z_2)$:

$$(R|c) = (A_2|b_2) = \begin{pmatrix} 3 & 1 & -3 & 8 \\ 0 & 2 & -1 & 5 \\ 0 & 0 & -2 & 2 \end{pmatrix} .$$

Damit ist

$$\begin{aligned} x_3 &= \frac{c_3}{r_{33}} = -1 \\ x_2 &= \frac{1}{r_{22}} \left(c_2 - \sum_{k=3}^{3} r_{2k} x_k \right) = \frac{1}{r_{22}} (5 - (-1)(-1)) = \frac{4}{2} = 2 \\ x_1 &= \frac{1}{r_{11}} \left(c_1 - \sum_{k=2}^{3} r_{1k} x_k \right) = \frac{1}{r_{11}} (8 - (1 \cdot 2 - 3 \cdot (-1))) = \frac{3}{3} = 1 , \end{aligned}$$

also

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} .$$

Bemerkung 10.10 1. In S.10.7 hatten wir vorausgesetzt, dass A vollen Rang hat, also invertierbar ist. Was passiert, wenn die (quadratische) Matrix A nicht invertierbar ist? Dann wird das Gauß'sche Verfahren in einem der Iterationsschritte zusammenbrechen, und zwar an dem Punkt, an dem eine Zeile mit einem nichtverschwindenen Eintrag "eingewechselt" werden muß oder spätestens damit, dass $r_{nn} = 0$ ist. Also: wenn das Verfahren bis zum letzten Schritt funktioniert, muß A notwendig vollen Rang gehabt haben.
2. Eine Erweiterung der Gauß-Algorithmus führt auch auf ein Verfahren zur Berechnug von A^{-1}:

Wir betrachten die Matrix $(A|E_n) \in K^{n\times(2n)}$. Führt man die Schritte des Gauß-Algorithmus wie im Beweis zu S. 10.7 beschrieben mit $(A|E_n)$ statt $(A|b)$ aus (also mit "rechter Seite" E_n), so wird $(A|E_n)$ übergeführt in eine Matrix $(R|C) \in K^{n\times(2n)}$ mit oberer Dreiecksmatrix $R = (r_{j_k})$ und $r_{jj} \neq 0$ für $j = 1, \ldots, n$. Durch weitere elementare Zeilenoperationen der Form $Z_j + \lambda Z_k$ und λZ_j kann $(R|C)$ übergeführt werden in $(E_n|B)$. Dabei gilt dann $B = A^{-1}$. (Denn: Ist $b^{(k)}$ die k-te Spalte von B, und ist $e^{(k)}$ der k-te Einheitsvektor in K^n so gilt nach Konstruktion $Ax = e^{(k)}$ genau dann, wenn $x = E_n x = b^{(k)}$, d. h. $Ab^{(k)} = e^{(k)}$ für $k = 1, \ldots, n$ bzw. $AB = E_n$ ist). Dies zeigt auch, das jede invertierbare Matrix A ein Produkt aus Elementarmatrizen ist. (Denn es ist $E_n = Z \cdot A$, wobei Z ein Produkt aus Elementarmatrizen ist. Also ist nach S.10.5 auch $A = Z^{-1}$ Produkt aus Elementarmatrizen.)

Bemerkung 10.11 Wir betrachten die Iterationsschritte des Gauß'schen Algorithmus noch einmal etwas näher. Dabei untersuchen wir anstelle der erweiterten Matrix $(A|b)$ nur die $(n\times n)$-Matrix A. Zunächst sei A so, dass im Gauß'schen Verfahren keine Zeilenvertauschungen vorgenommen werden müssen. Dann lässt sich der erste Schritt folgendermaßen formalisieren

$$A_1 = L_1 A (= L_1 A_0)$$

wobei A_1 wie im Beweis zu S.10.7 und

$$\begin{aligned} L_1 &= \begin{pmatrix} 1 & 0 & \ldots & \ldots & 0 \\ -\ell_{21} & 1 & 0 & & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 & 0 \\ -\ell_{n1} & 0 & & 0 & 1 \end{pmatrix} \\ &= F^{(n,1)}(-\ell_{n1}) \cdots F^{(2,1)}(-\ell_{21}) = E + \sum_{j=2}^{n} (-\ell_{j1}) \cdot E^{(j,1)} \end{aligned}$$

mit

$$\ell_{j1} = \frac{a_{j1}}{a_{11}} = \frac{a_{j1}^{(0)}}{a_{11}^{(0)}} \qquad (j = 2, \ldots, n) \, .$$

Allgemeiner gilt für $m = 1, \ldots, n-1$

$$A_m = L_m A_{m-1}$$

wobei A_{m-1}, A_m wie im Beweis zu S.10.7 und

$$L_m = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & & & & & & \vdots \\ \vdots & \ddots & \ddots & & & & & \\ \vdots & & 0 & 1 & & & & \\ \vdots & & \vdots & -\ell_{m+1,m} & 1 & & & \vdots \\ \vdots & & \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & & 0 \\ 0 & \dots & 0 & -\ell_{n,m} & 0 & \dots & & 1 \end{pmatrix}$$

$$= F^{(n,m)}(-\ell_{n,m}) \cdots F^{(m+1,m)}(-\ell_{m+1,n}) = E + \sum_{j=m+1}^{n} (-\ell_{jm}) E^{(j,m)}$$

mit

$$\ell_{jm} = \frac{a_{jm}^{(m-1)}}{a_{mm}^{(m-1)}} \qquad (j = m+1, \cdots, n) \, .$$

(Matrizen dieser Form heißen *Frobenius-Matrizen*. Wir setzen zur Abkürzung

$$\mathcal{L}_{n,m} := \{L \in K^{n \times n} : L = E + \sum_{j=m+1}^{n} \lambda_j E^{(j,m)}, \ \lambda_j \in K \text{ für } j = m+1, \dots, n\} \, .)$$

Also ergibt sich insgesamt

$$L_{n-1} L_{n-2} \cdots L_2 L_1 A = A_{n-1} = R \, .$$

Weiter sieht man mit S.10.5: Die Matrix L_m ist invertierbar mit

$$L_m^{-1} = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & & & & & & \vdots \\ \vdots & \ddots & \ddots & & & & & \\ \vdots & & 0 & 1 & & & & \\ \vdots & & \vdots & \ell_{m+1,m} & 1 & & & \vdots \\ \vdots & & \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & & 0 \\ 0 & \dots & 0 & \ell_{n,m} & 0 & \dots & & 1 \end{pmatrix}$$

$$= F^{(m+1,m)}(\ell_{m+1,m}) \cdots F^{(n,m)}(\ell_{n,m}) = E + \sum_{j=m+1}^{n} \ell_{jm} E^{(j,m)}$$

Damit folgt

$$A = L_1^{-1} \cdot L_2^{-1} \cdots L_{n-1}^{-1} \cdot R \,,$$

d. h. mit der unteren Dreiecksmatrix

$$L := L_1^{-1} \cdots L_{n-1}^{-1} = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ \ell_{21} & 1 & \ddots & & \vdots \\ \vdots & \ell_{32} & \ddots & \ddots & \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \ell_{n1} & \ell_{n2} & \cdots & \ell_{n,n-1} & 1 \end{pmatrix}$$

ist

$$\boxed{A = LR} \,.$$

Eine solche Zerlegung von A hat den Vorteil, dass die Lösung eines LGS $Ax = b$ für $b \in \mathbb{R}^n$ durch Lösen der beiden Gleichungssysteme

$$Ly = b$$

und

$$Rx = y$$

berechnet werden kann. Hierbei haben beide Koeffizientenmatrizen Dreiecksgestalt, d. h. die Gleichungssysteme können rekursiv (vgl. B. 10.8) gelöst werden. Dies ist insbesondere dann von Vorteil, wenn $Ax = b$ für verschiedene rechte Seiten b gelöst werden muss.
Man kann zeigen ([Ü]), dass eine solche Zerlegung für invertierbare Matrizen stets eindeutig ist, d. h. ist

$$\mathcal{L}_n := \{L = (\ell_{jk}) \in K^{n \times n} : \ell_{jk} = 0 \text{ für } j < k,\ \ell_{jj} = 1 \text{ für } j = 1, \dots, n\} \,,$$

so existieren höchstens ein $L \in \mathcal{L}_n$ sowie eine obere Dreiecksmatrix $R = (r_{jk})$ mit $r_{jj} \neq 0$ $(j = 1, \dots, n)$ und $A = LR$. Man nennt diese Zerlegung dann LR-Zerlegung von A.

Ist A so, dass im Gauß'schen Algorithmus Zeilen vertauscht werden müssen, so sieht man wie oben: Es existieren Frobenius-Matrizen $L_1, \dots, L_{n-1}$ und Vertauschungsmatrizen

$$P^{(1,k_1)}, \dots, P^{(n-1,k_{n-1})}$$

mit $k_m \geq m$ (wobei $P^{(m,k_m)} = E$, falls im m-ten Schritt keine Vertauschung vorgenommen wurde) und so, dass

$$L_{n-1}P^{(n-1,k_{n-1})}L_{n-2}P^{(n-2,k_{n-2})} \cdot \ldots \cdot L_1 P^{(1,k_1)} \cdot A = R\,.$$

Man kann zeigen ([Ü]), dass für jede Frobenius Matrix $F \in \mathcal{L}_{n,m}$ und $j,k > m$ eine Frobenius-Matrix $F' \in \mathcal{L}_{n,m}$ existiert mit

$$P^{(j,k)}F = F'P^{(j,k)}\,.$$

Also folgt, dass mit gewissen Frobenius-Matrizen $\tilde{L}_m \in \mathcal{L}_{n,m}$ $(m = 1, \ldots, n-1)$ gilt

$$\tilde{L}_{n-1} \cdot \ldots \cdot \tilde{L}_1 P^{(n-1,k_{n-1})} \cdot \ldots \cdot P^{(1,k_1)}A = R\,.$$

Mit $P := P^{(n-1,k_{n-1})} \cdots P^{(1,k_1)}$ und $L := (\tilde{L}_{n-1} \cdots \tilde{L}_1)^{-1} = \tilde{L}_1^{-1} \cdots \tilde{L}_{n-1}^{-1}$ gilt dann

$$\boxed{PA = LR}\,.$$

Hierbei ist P (eine sog. Permutationsmatrix) invertierbar, und $L \in \mathcal{L}_n$ wieder eine untere Dreiecksmatrix. Ähnlich wie oben kann die Lösung von $Ax = b$ für beliebiges $b \in K^n$ durch Lösen der Systeme

$$Ly = Pb$$

und

$$Rx = y$$

berechnet werden.

Der folgende Satz liefert eine Charakterisierung der Matrizen, für die die LR-Zerlegung existiert.

Satz 10.12 *Es sei $A = (a_{jk}) \in GL_n(K)$. Genau dann existiert die LR-Zerlegung von A, wenn die Matrizen*

$$A^{(r)} := (a_{jk})_{j,k=1,\ldots,r} \in K^{r\times r}$$

für $r = 1, \ldots, n$ invertierbar sind.

Beweis.
"⇒": Ist $A = LR$ die LR-Zerlegung, so schreiben wir diese in Blockform (die Bedeutung der Blöcke ergibt sich aus $A^{(r)} \in K^{r\times r}$)

$$\begin{pmatrix} A^{(r)} & B^{(r)} \\ C^{(r)} & D^{(r)} \end{pmatrix} = \begin{pmatrix} L^{(r)} & 0 \\ L_{21}^{(r)} & L_{22}^{(r)} \end{pmatrix} \begin{pmatrix} R^{(r)} & R_{12}^{(r)} \\ 0 & R_{22}^{(r)} \end{pmatrix} =$$
$$= \begin{pmatrix} L^{(r)}R^{(r)} & L^{(r)}R_{12}^{(r)} \\ L_{21}^{(r)}R^{(r)} & L_{21}^{(r)}R_{12}^{(r)} + L_{22}^{(r)}R_{22}^{(r)} \end{pmatrix}.$$

Es folgt $A^{(r)} = L^{(r)}R^{(r)}$. Da $L^{(r)}$ und $R^{(r)}$ invertierbar sind, ist auch $A^{(r)}$ invertierbar.
"⇐": (Induktion nach n)
Der Fall $n = 1$ ist klar. Es sei $A \in K^{(n+1)\times(n+1)}$ so, dass $A^{(r)}$ invertierbar ist für $r = 1, \dots, n+1$. Dann gilt nach Induktionsvoraussetzung

$$A = \begin{pmatrix} A^{(n)} & b \\ a^T & \alpha \end{pmatrix} = \begin{pmatrix} L^{(n)}R^{(n)} & b \\ a^T & \alpha \end{pmatrix}$$

mit gewissen $a, b \in K^n, \alpha \in K$ und einer oberen Dreiecksmatrix $R^{(n)}$ mit $r_{jj}^{(n)} \neq 0$ sowie einer Matrix $L^{(n)} \in \mathcal{L}_n$. Also folgt

$$\begin{pmatrix} (L^{(n)})^{-1} & 0 \\ 0^T & 1 \end{pmatrix} A = \begin{pmatrix} R^{(n)} & b' \\ a^T & \alpha \end{pmatrix}$$

mit $b' = (L^{(n)})^{-1}b$ und $(L^{(n)})^{-1} \in \mathcal{L}_n$ (man beachte: $\mathcal{L}_n$ ist eine Gruppe ([Ü])). Durch elementare Zeilenoperationen der Form $Z_j + \lambda Z_k$ lässt sich die Matrix auf der rechten Seite in eine obere Dreiecksmatrix R mit $r_{jj} \neq 0$ überführen. Also existiert nach B.10.11 eine Matrix $\tilde{L} \in \mathcal{L}_{n+1}$ mit

$$\tilde{L} \begin{pmatrix} (L^{(n)})^{-1} & 0 \\ 0^T & 1 \end{pmatrix} A = R\,.$$

Da $\mathcal{L}_{n+1}$ eine Gruppe ist, gilt für die untere Dreiecksmatrix

$$L := \begin{pmatrix} L^{(n)} & 0 \\ 0^T & 1 \end{pmatrix} \tilde{L}^{-1} \in \mathcal{L}_{n+1}$$

die Zerlegung $A = LR$. □

Bisher haben wir uns lediglich mit einer Lösungsmethode für lineare Gleichungssysteme mit invertierbarer Koeffizientenmatrix A beschäftigt. Wir werden jetzt auf Gleichungssysteme mit beliebiger Matrix A eingehen.

Definition 10.13 Es sei $B = (b_{jk}) \in K^{m\times n}$, und es sei $r \in \{1, \dots, \min(n, m)\}$. Man sagt, die Matrix B hat *Zeilenstufenform (mit r Stufen)*, wenn es Spaltenindizes $k_1, \dots, k_r \in \{1, \dots, n\}$ mit folgenden Eigenschaften gibt:

a) $k_j \leq k_{j+1}$ für $j = 1, \ldots, r-1$

b) $b_{j,k_j} \neq 0$ für $j = 1, \ldots, r$

c) $b_{jk} = 0$ für $j = r+1, \ldots, m;\ k = 1, \ldots, n$ und für $j = 1, \ldots, r;\ k < k_j$

d. h. B hat die Form

$$B = \begin{pmatrix} 0 & \ldots & 0 & b_{1,k_1} & \ldots & & & & & \\ 0 & \ldots & \ldots & \ldots & 0 & b_2, k_2 & & & & \\ 0 & \ldots & \ldots & \ldots & \ldots & 0 & & & & \\ \vdots & & & & & & \ddots & & & \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & b_{r,k_r} & \ldots & \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ \vdots & & & & & & & & & \vdots \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \end{pmatrix} .$$

Satz 10.14 *Ist $B = (b_{jk}) \in K^{m \times n}$ eine Matrix in Zeilenstufenform mit r Stufen, so ist $Rg(B) = r$.*

Beweis..
Sind $B_1 \ldots, B_r$ die ersten r Zeilen von B, so gilt

$$Rg(B) = Zrg(B) = \dim < B_1 \ldots, B_r > \leq r \ .$$

Wir zeigen $B_1 \ldots, B_r$ sind linear unabhängig (dann ist $Rg(B) = Zrg(B) = r$). Dazu seien $\lambda_1, \ldots, \lambda_r \in K$ so, dass

$$\sum_{\nu=1}^{r} \lambda_\nu B_\nu = 0^T \ ,$$

d. h.

$$\sum_{\nu=1}^{r} \lambda_\nu b_{\nu k} = 0 \qquad \text{für } k = 1, \ldots, n \ .$$

Insbesondere gilt

$$\sum_{\nu=1}^{r} \lambda_\nu b_{\nu,k_j} = 0 \qquad \text{für } j = 1, \ldots, r \ .$$

Für $j = 1$ ergibt sich (da $b_{\nu,k_1} = 0$ für $\nu > 1$)

$$0 = \sum_{\nu=1}^{r} \lambda_\nu b_{\nu,k_1} = \lambda_1 b_{\nu,k_1} \ ,$$

also $\lambda_1 = 0$. Hieraus folgt für $j = 2$

$$0 = \sum_{\nu=2}^{r} \lambda_\nu b_{\nu,k_2} = \lambda_2 b_{\nu,k_2} \ ,$$

also $\lambda_2 = 0$. So fortfahrend ergibt sich $\lambda_1 = \lambda_2 = \ldots = \lambda_r = 0$, also sind $B_1, \ldots, B_r$ linear unabhängig. □

Satz 10.15 *Jede Matrix $A \in K^{m\times n}, A \neq 0$, lässt sich durch elementare Zeilenumformungen in eine Matrix $B \in K^{m\times n}$ transformieren, die Zeilenstufenform (mit $Rg(A)$ Stufen) hat.*

Beweis.
Da $A \neq 0$ ist, existiert ein kleinster Spaltenindex k_1 mit $a^{(k_1)} \neq 0$. Ist etwa $a_{jk_1} \neq 0$, so vertauschen wir die erste mit der j-ten Zeile (diesen Eintrag nennen wir b_{1k_1}). Durch Umformungen der Form $Z_j + \lambda Z_1$ können wir dann A transformieren in die Form

$$\begin{pmatrix} 0 & \cdots & 0 & b_{1,k_1} & \cdots & & \\ \vdots & & \vdots & 0 & & & \\ \vdots & & \vdots & \vdots & & & \\ \vdots & & \vdots & \vdots & & & A_1 \\ \vdots & & \vdots & \vdots & & & \\ 0 & \ldots & 0 & 0 & & & \end{pmatrix}$$

Ist $A_1 \in K^{(m-1)\times(n-k_1)}$ die Nullmatrix, so sind wir fertig. Ist $A_1 \neq 0$, so wird die gleiche Prozedur auf A_1 angewandt. Nach endlich vielen Schritten landet man bei einer Matrix B in Zeilenstufenform. □

Folgerung 10.16 Ist $A \in K^{m\times n}, A \neq 0$ und ist $b \in K^m$, so lässt sich $(A|b)$ durch elementare Zeilenumformungen in eine Matrix $(A'|b')$ transformieren, die Zeilenstufenform hat. Die LGS'e $Ax = b$ und $A'x = b'$ sind nach F. 10.6 äquivalent.
Ist

$$(A'|b') = \begin{pmatrix} 0 & \cdots & 0 & a'_{1,k_1} & & & & & c_1 \\ \vdots & & \vdots & 0 & a'_{2,k_2} & & & & c_2 \\ \vdots & & \vdots & & 0 & & & & \vdots \\ \vdots & & \vdots & & & 0 & a'_{r,k_r} & \ldots & c_r \\ \vdots & & \vdots & & & & & 0 & c_{r+1} \\ \vdots & & \vdots & & & & & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & \cdots & \cdots & \cdots & 0 & c_m \end{pmatrix}$$

mit $a'_{j,k_j} \neq 0$ für $j = 1, \ldots, r$, so ist $A'x = b'$ (bzw. $Ax = b$) genau dann lösbar, wenn $c_{r+1} = \ldots = c_m = 0$ gilt. (In diesem Fall ist $Rg(A') = Rg(A'|b')$ anderenfalls ist $Rg(A') = r < Rg(A'|b')$)

Bemerkung 10.17 Wie kann man im Falle $c_{r+1} = \cdots = c_m = 0$ das System lösen? Durch geeignetes Vertauschen von *Spalten* lässt sich (die um die letzten $m-r$, nur aus Nullen bestehenden Zeilen "verkleinerte" Matrix) in die Form

$$(A''|c) = \begin{pmatrix} a''_{11} & \cdots & \cdots & \cdots & a''_{1r} & \cdots & a''_{1n} & c_1 \\ 0 & a''_{22} & \cdots & & a''_{2r} & \cdots & a''_{2n} & c_2 \\ \vdots & 0 & & & \vdots & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a''_{rr} & \cdots & a''_{rn} & c_r \end{pmatrix}$$

mit $a''_{jj} \neq 0$ für $j = 1, \ldots, r$ bringen. Wir berechnen die Lösungen y von $A''y = c$. Die Lösungen x von $A'x = b'$ (bzw. $Ax = b$) ergeben sich dann durch Rückvertauschen der Variablen $y_1, \ldots, y_n$.
Nach S.9.4 haben wir eine spezielle Lösung $y^{(0)}$ der inhomogenen Gleichung und eine Basis von Lös $(A'', 0)$ zu berechnen.

1. Berechnung einer speziellen Lösung $y^{(0)}$ von $A''y = c$
Mit $y^{(0)} = (y_1, \ldots, y_n)^T$ setzen wir $y_{r+1} = \cdots = y_n = 0$. Dann kann man die $y_1, \ldots, y_r$ "von hinten nach vorne" wie in B.10.8 berechnen (mit

$$R = \begin{pmatrix} a''_{11} & \cdots & & a''_{1r} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 0 & a''_{rr} \end{pmatrix}$$

und rechter Seite $c = (c_1, \ldots, c_r)^T$).
2. Berechnung einer Basis von Lös $(A'', 0)$:
O. E. können wir $n > r$ annehmen (sonst ist Lös$(A'', 0) = \{0\}$). Wir setzen $A'' = (R|S)$ mit

$$R = \begin{pmatrix} a''_{11} & \cdots & a''_{1r} \\ & \ddots & \vdots \\ 0 & & a''_{rr} \end{pmatrix} \quad \text{und} \quad S = \begin{pmatrix} a''_{1,r+1} & \cdots & a''_{1n} \\ \vdots & & \vdots \\ a''_{r,r+1} & \cdots & a''_{rn} \end{pmatrix}.$$

Schreiben wir $y = (y_1, \ldots, y_n)^T \in K^n$ als $(u, v)^T$ mit $u = (y_1, \ldots, y_r)^T$ und $v = (y_{r+1}, \ldots, y_n)^T$, so ist

$$\text{Lös } (A'', 0) = \{y : A''y = 0\} = \{(u, v) \in K^r \times K^{n-r} : Ru = -Sv\}$$

(denn: $A''y - A''\binom{u}{v} = Ru + Sv$). Für $v = e^{(k)}$ ($= k$-ter Einheitsvektor in K^{n-r}) berechnen wir die Lösung $u^{(k)}$ von

$$Ru = -Sv = -Se^{(k)} = -\begin{pmatrix} a''_{1,k+r} \\ \vdots \\ a''_{r,k+r} \end{pmatrix} \qquad (k = 1, \dots, n-r)$$

nach dem Verfahren aus B.10.8. Dann ist $y^{(1)}, \dots, y^{(n-r)}$ mit

$$y^{(k)} := \binom{u^{(k)}}{e^{(k)}} \qquad (k = 1, \dots, n-r)$$

eine Basis von $\mathrm{Lös}\,(A'', 0)$ (denn $y^{(1)}, \dots, y^{(n-1)}$ sind linear unabhängig nach Konstruktion und $\dim(A'', 0) = n - r$).

Beispiel 10.18 Wir betrachten das LGS

$$Ax = \begin{pmatrix} 1 & -2 & 2 & -1 \\ 2 & -4 & -5 & 1 \\ -1 & 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix} = b\,.$$

Dann ist

$$\begin{aligned} (A|b) &= \begin{pmatrix} 1 & -2 & 2 & -1 & 4 \\ 2 & -4 & -5 & 1 & -1 \\ -1 & 2 & 1 & 0 & -1 \end{pmatrix} \\ \overset{\substack{Z_2-2Z_1 \\ Z_3+Z_1}}{\longrightarrow} & \begin{pmatrix} 1 & -2 & 2 & -1 & 4 \\ 0 & 0 & -9 & 3 & -9 \\ 0 & 0 & 3 & -1 & 3 \end{pmatrix} \\ \overset{Z_3+\frac{1}{3}Z_2}{\longrightarrow} & \begin{pmatrix} 1 & -2 & 2 & -1 & 4 \\ 0 & 0 & -9 & 3 & -9 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = (A'|b')\,. \end{aligned}$$

Vertauschen der 4. und 2. Spalte (und Weglassen der letzten Zeile) ergibt

$$(A''|c) = \begin{pmatrix} 1 & -1 & 2 & -2 & 4 \\ 0 & 3 & -9 & 0 & -9 \end{pmatrix} = (R|S|c)$$

also die gewünschte Form.
Eine spezielle Lösung $y^{(0)}$ der inhomogenen Gleichung ergibt sich mit $y_3 = y_4 = 0$ aus

$$R\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 4 \\ -9 \end{pmatrix}\,,$$

also

$$y_2 = -3, y_1 = 1 \ ,$$

und damit

$$y^{(0)} = \begin{pmatrix} 1 \\ -3 \\ 0 \\ 0 \end{pmatrix} .$$

Eine Basis $(y^{(1)}, y^{(2)})$ von Lös $(A'', 0)$ erhält man aus

$$Ru = \begin{pmatrix} 1 & -1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -2 \\ 9 \end{pmatrix} = -Se^{(1)}$$

und

$$Ru = \begin{pmatrix} 1 & -1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} = -Se^{(2)} \ .$$

Man berechnet

$$y^{(1)} = \begin{pmatrix} 1 \\ 3 \\ 1 \\ 0 \end{pmatrix} \qquad \text{und} \qquad y^{(2)} = \begin{pmatrix} 2 \\ 0 \\ 0 \\ 1 \end{pmatrix} \ .$$

Also ist

$$\text{Lös}\,(A'', c) = y^{(0)} + \text{Lös}\,(A'', 0) = \left\{ \begin{pmatrix} 1 \\ -3 \\ 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 3 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 2 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \lambda, \mu \in \mathbb{R} \right\} .$$

Durch Rückvertauschen der Variablen erhält man

$$\text{Lös}\,(A, b) = \text{Lös}\,(A', b') = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ -3 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 0 \\ 1 \\ 3 \end{pmatrix} + \mu \begin{pmatrix} 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \lambda, \mu \in \mathbb{R} \right\} .$$

11 Determinanten

Bevor wir zu einem weiteren zentralen Begriff der Linearen Algebra, nämlich der Determinante einer Matrix, kommen, befassen wir uns vorbereitend noch einmal etwas genauer mit der Symmetrischen Gruppe S_n, d. h. der Menge aller bijektiven Abbildungen auf $\{1, \dots, n\}$ mit der Hintereinanderausführung als Verknüpfung (siehe B. 2.3).

Definition 11.1 Es sei $n \geq 2$. Ein $\sigma \in S_n$ heißt *Transposition (der Elemente j und k aus $\{1, \dots, n\}$)*, falls $j \neq k$ und $\sigma(j) = k$ sowie $\sigma(m) = m$ für alle $m \in \{1, \dots, n\} \setminus \{j, k\}$. Wir schreiben dann $\sigma =: [j, k]$. Es gilt dabei $[j, k] = [k, j]$ und $[j, k]^{-1} = [j, k]$.

Satz 11.2 *Es sein $n \in \mathbb{N}$. Dann gilt*

1. $|S_n| = n!$.
2. *Für $n \geq 2$ ist jedes $\sigma \in S_n$ darstellbar als Produkt (bzgl. $\circ$) aus höchstens n Transpositionen.*

Beweis.
1. Induktionsanfang $n = 1$: Es ist $|S_1| = |\{id_{\{1\}}\}| = 1$.
2. Induktionsschritt $n \to n + 1$: Es gilt

$$S_{n+1} = \bigcup_{j=1}^{n+1} A_j \ ,$$

wobei $A_j := \{\sigma \in S_{n+1} : \sigma(j) = n+1\}$ und die Zerlegung disjunktiv ist. Jedem $\sigma \in A_j$ entspricht in eineindeutiger Weise ein $\tau \in S_n$ (nämlich τ mit $\tau(k) := \sigma(k)$ für $k < j$ und $\tau(k) := \sigma(k + 1)$ für $k \geq j$). Nach Induktionsvoraussetzung ist also

$$|A_j| = |S_n| = n! \qquad (j = 1, \dots, n + 1)$$

und damit

$$|S_{n+1}| = \sum_{j=1}^{n+1} |A_j| = (n + 1)! \ .$$

2. Ist $\sigma = id$, so gilt $\sigma = [1, 2] \circ [1, 2]$. Ist $\sigma \neq id$, so existiert ein kleinstes $j_1 \in \{1, \dots, n\}$ mit $\sigma(j_1) = k_1 \neq j_1$. Für $\sigma_1 := [j_1, k_1] \circ \sigma$ folgt dann $\sigma_1(j) = j$ für $j \leq j_1$. Im Falle $\sigma_1 = id$ gilt $\sigma = [j_1, k_1]$ und wir sind fertig. Ist $\sigma_1 \neq id$, so existiert ein kleinstes $j_2 \in \{1, \dots, n\}$ mit $k_2 = \sigma_1(j_2) \neq j_2$ (dabei ist $j_2 > j_1$). Wir betrachten $\sigma_2 := [j_2, k_2] \circ \sigma_1$.

Dann ist $\sigma_2(j) = j$ für $j \leq j_2$. So fortfahrend gelangen wir nach $m \leq n-1$ Schritten zu einer Darstellung

$$\sigma_m = [j_m, k_m] \circ \ldots \circ [j_1, k_2] \circ \sigma = id \; .$$

Da $[j,k]^{-1} = [j,k]$ gilt, ist

$$\sigma = [j_1, k_1] \circ \ldots \circ [j_m, k_m] \; .$$

□

Beispiel 11.3 Es sei

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 1 & 5 & 2 \end{pmatrix} \; .$$

Dann gilt

$$\begin{aligned} \sigma_1 = [1,4] \circ \sigma &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 5 & 2 \end{pmatrix} \\ \sigma_2 = [2,3] \circ \sigma_1 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 3 \end{pmatrix} \\ \sigma_3 = [3,4] \circ \sigma_2 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 5 & 4 \end{pmatrix} \\ \sigma_4 = [4,5] \circ \sigma_3 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = id \end{aligned}$$

d. h.

$$id = \sigma_4 = [4,5] \circ [3,4] \circ [2,3] \circ [1,4] \circ \sigma$$

also

$$\sigma = [1,4] \circ [2,3] \circ [3,4] \circ [4,5] \; .$$

Definition 11.4 Es sei $\sigma \in S_n$.
1. Jedes Paar $(j,k) \in \{1, \ldots, n\}^2$ mit $j < k$ und $\sigma(j) > \sigma(k)$ heißt *Inversion* (oder *Verstellung*) von σ.
2. Ist $I(\sigma)$ die Anzahl der Inversionen von σ, so heißt σ *gerade* (bzw. *ungerade*) falls $I(\sigma)$ gerade (bzw. ungerade), ist. Die Zahl

$$\operatorname{sign}(\sigma) := (-1)^{I(\sigma)}$$

heißt *Signum* (oder *Vorzeichen*) von σ.

Bemerkung 11.5 1. Für jedes $\sigma \in S_n$ gilt

$$\operatorname{sign}(\sigma) = \prod_{1 \le j < k \le n} \frac{\sigma(k) - \sigma(j)}{k - j}$$

(Denn: Es ist

$$\begin{aligned}
\prod_{j<k} (\sigma(k) - \sigma(j)) &= \prod_{\substack{j<k \\ \sigma(j)<\sigma(k)}} (\sigma(k) - \sigma(j)) \cdot \prod_{\substack{j<k \\ \sigma(j)>\sigma(k)}} (\sigma(k) - \sigma(j)) \\
&= \prod_{\substack{j<k \\ \sigma(j)<\sigma(k)}} (\sigma(k) - \sigma(j)) \cdot \prod_{\substack{j<k \\ \sigma(j)>\sigma(k)}} |\sigma(k) - \sigma(j)| \cdot (-1)^{I(\sigma)} \\
&= (-1)^{I(\sigma)} \prod_{j<k} |\sigma(k) - \sigma(j)| = (-1)^{I(\sigma)} \prod_{j<k} (k - j) \ ,
\end{aligned}$$

wobei bei der letzten Gleichheit zu beachten ist, dass aufgrund der Bijektivität von σ beide Produkte die gleichen Faktoren enthalten).

2. Ist $\sigma = [j, k]$ eine Transposition, so ist σ ungerade, d. h.

$$\operatorname{sign}(\sigma) = -1$$

(Denn: o. E. sei $j < k$. Dann hat σ genau die Inversionen

$$(j, j+1), \ldots, (j, k) \quad \text{und} \quad (j+1, k), \ldots, (k-1, k) \ ,$$

d. h. $I(\sigma) = 2(k - j) - 1$ ist ungerade).

Von zentraler Bedeutung ist nun

Satz 11.6 *1. Es seien $\sigma, \tau \in S_n$. Dann gilt*

$$\operatorname{sign}(\sigma \circ \tau) = \operatorname{sign}(\sigma) \cdot \operatorname{sign}(\tau) \ .$$

2. Ist $\sigma \in S_n$ Produkt aus m Transposition, so ist $\operatorname{sign}(\sigma) = (-1)^m$.

Beweis.
1. Mit $\beta_{kj} := \dfrac{\sigma(\tau(k)) - \sigma(\tau(j))}{\tau(k) - \tau(j)}$ für $j \neq k$ gilt nach B.11.5.1

$$\begin{aligned}
\operatorname{sign}(\sigma \circ \tau) &= \prod_{j<k} \frac{\sigma(\tau(k)) - \sigma(\tau(j))}{k - j} = \\
&= \prod_{j<k} \frac{\sigma(\tau(k)) - \sigma(\tau(j))}{\tau(k) - \tau(j)} \prod_{j<k} \frac{\tau(k) - \tau(j)}{k - j} \\
&= \operatorname{sign}(\tau) \prod_{\substack{j<k \\ \tau(j)<\tau(k)}} \beta_{kj} \cdot \prod_{\substack{j<k \\ \tau(j)>\tau(k)}} \beta_{kj} \\
&\overset{\beta_{jk}=\beta_{kj}}{=} \operatorname{sign}(\tau) \prod_{\substack{j<k \\ \tau(j)<\tau(k)}} \beta_{kj} \prod_{\substack{j>k \\ \tau(j)<\tau(k)}} \beta_{kj} \\
&= \operatorname{sign}(\tau) \cdot \prod_{\tau(j)<\tau(k)} \frac{\sigma(\tau(k)) - \sigma(\tau(j))}{\tau(k) - \tau(j)} \\
&= \operatorname{sign}(\tau) \cdot \prod_{j<k} \frac{\sigma(k) - \sigma(j)}{k - j} = \operatorname{sign}(\tau) \cdot \operatorname{sign}(\sigma) \, .
\end{aligned}$$

wobei die vorletzte Gleichheit aus der Bijektivität von τ folgt.
2. Folgt auch 1. und B.11.5.2. □

Definition 11.7 Es sei K ein Körper. Eine Abbildung $d : K^{n\times n} \to K$ heißt Determinatnenabbildung, falls folgende Bedingungen gelten:

a) d ist linear in jeder Zeile, d. h. sind $j \in \{1, \ldots, n\}$, $A = \begin{pmatrix} A_1 \\ \vdots \\ A_1 \end{pmatrix} \in K^{n\times n}$ und $b \in K^n, \lambda, \mu \in K$, so gilt

$$d \begin{pmatrix} A_1 \\ \vdots \\ \lambda A_j + \mu b^T \\ \vdots \\ A_n \end{pmatrix} = \lambda \cdot d(A) + \mu \cdot d \begin{pmatrix} A_1 \\ \vdots \\ b^T \\ \vdots \\ A_n \end{pmatrix} ,$$

b) $d(A) = 0$, falls zwei Zeilen von A übereinstimmen,

d) $d(E) = 1$.

Bemerkung 11.8 Ist d eine Determinantenabbildung, so gilt für

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix} \in K^{n \times n}$$

und $\lambda \in K$ sowie $j \neq k$

$$d \begin{pmatrix} A_1 \\ \vdots \\ A_j + \lambda A_k \\ \vdots \\ A_n \end{pmatrix} = d\left(F^{(j,k)}(\lambda) \cdot A\right) = d(A)$$

(d. h. Addition eines Vielfachen einer Zeile zu einer anderen Zeile ändert d nicht) und

$$d(P^{(j,k)} \cdot A) = -d(A)$$

(d. h. Vertauschen zweier Zeilen bewirkt Vertauschung des Vorzeichens).

(Denn: o. E. sei $j < k$. Dann gilt

$$d \begin{pmatrix} A_1 \\ \vdots \\ A_j + \lambda A_k \\ \vdots \\ A_k \\ \vdots \\ A_n \end{pmatrix} = d(A) + \lambda d \begin{pmatrix} A_1 \\ \vdots \\ A_k \\ \vdots \\ A_k \\ \vdots \\ A_n \end{pmatrix} = d(A)$$

und

$$0 = d\begin{pmatrix} A_1 \\ \vdots \\ A_j + A_k \\ \vdots \\ A_k + A_j \\ \vdots \\ A_n \end{pmatrix} = d\begin{pmatrix} A_1 \\ \vdots \\ A_j \\ \vdots \\ A_k + A_j \\ \vdots \\ A_n \end{pmatrix} + d\begin{pmatrix} A_1 \\ \vdots \\ A_k \\ \vdots \\ A_k + A_j \\ \vdots \\ A_n \end{pmatrix}$$

$$= d(A) + d\begin{pmatrix} A_1 \\ \vdots \\ A_k \\ \vdots \\ A_j \\ \vdots \\ A_n \end{pmatrix} = d(A) + d(P^{(j,k)}A)\ .)$$

Der folgende Satz ist grundlegend für die Theorie der Determinanten.

Satz 11.9 *Es seien K ein Körper und $n \in \mathbb{N}$. Dann existiert genau eine Determinatenabbildung* $\det : K^{n\times n} \to K$*, und es gilt für $A = (a_{jk}) \in K^{n\times n}$:*

$$\det A = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \qquad (2)$$

Beweis.

1. Existenz: Es sei $d : K^{n\times n} \to K$ definiert durch

$$d(A) := \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) \cdot a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \qquad (A = (a_{jk}) \in K^{n\times n})\ .$$

Wir zeigen, dass d die Bedingungen aud D.11.7 erfüllt:
Es seinen $A, b = (b_1, \ldots, b_n)^T, \lambda, \mu$ wie D.11.7 a). Dann gilt

$$\sum_{\sigma\in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} \cdots a_{j-1,\sigma(j-1)} (\lambda a_{j,\sigma(j)} + \mu b_{\sigma(j)}) a_{j+1,\sigma(j+1)} \cdots a_{n,\sigma(n)}$$

$$= \lambda \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) \cdot a_{1,\sigma(1)} \cdots a_{j,\sigma(j)} \cdots a_{n,\sigma(n)} +$$

$$+ \mu \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) \cdot a_{1,\sigma(1)} \cdots b_{\sigma(j)} \cdots a_{n,\sigma(n)}\ .$$

Dies ist nichts anderes als die Bedingung D.11.7 a) für d.

Es sei nun $A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix} \in K^{n\times n}$ mit $A_j = A_k$ für ein Paar (j,k) mit $j < k$. Es gilt mit $\tau := [j,k]$ nach S.11.6 und B.11.5.2 $\operatorname{sign}(\sigma\circ\tau) = -\operatorname{sign}(\sigma)$ für alle $\sigma \in S_n$ also

$$\begin{aligned} d(A) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)}\cdots a_{n,\sigma(n)} = \\ &\quad \sum_{\substack{\sigma\in S_n \\ \sigma\text{gerade}}} \operatorname{sign}(\sigma) a_{1,\sigma(1)}\cdots a_{j,\sigma(j)}\cdots a_{k,\sigma(k)}\cdots a_{n,\sigma(n)} + \\ &\quad \sum_{\substack{\sigma\in S_n \\ \sigma\text{gerade}}} -\operatorname{sign}(\sigma) a_{1,\sigma(\tau(1))}\cdots a_{j,\sigma(\tau(j))}\cdots a_{k,\sigma(\tau(k))}\cdots a_{n,\sigma(\tau(n))} \ . \end{aligned}$$

(Man beachte: Durch $\sigma \mapsto \sigma\circ\tau$ ist eine bijektive Abbildung zwischen $\{\sigma \in S_n : \sigma \text{ gerade}\}$ und $\{\sigma \in S_n : \sigma \text{ ungerade}\}$ gegeben.) Weiter gilt

$$a_{j,\sigma(\tau(j))} = a_{j,\sigma(k)} = a_{k,\sigma(k)} \text{ und } a_{k,\sigma(\tau(k))} = a_{k,\sigma(j)} = a_{j,\sigma(j)}$$

sowie $a_{m,\sigma(\tau(m))} = a_{m,\sigma(m)}$ für $m \neq j,k$. Also ist der erste Summand das Negative des zweiten, d. h. $d(A) = 0$.

Ist $A = E = (\delta_{jk})$ so ist von der Summe

$$d(E) = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\delta_{1,\sigma(1)}\cdots\delta_{n,\sigma(n)}$$

nur der Summand für $\sigma = id$ nicht 0 und dieser hat den Wert 1, also $d(E) = 1$.

2. Eindeutigkeit: Es seien d_1 und d_2 Determinantenabbildungen, und es sei $A \in K^{n\times n}$, wobei o. E. $A \neq 0$. Wir zeigen: $d_1(A) = d_2(A)$. Dazu sei B eine gemäß S.10.15 aus A durch elementare Zeilenumformungen (der Form $Z_j \leftrightarrow Z_k$ und $Z_j + \lambda Z_k$) entstandene Matrix in Zeilenstufenform. Nach B. 11.8 gilt

$$d_i(B) = (-1)^m d_i(A) \qquad (i = 1,2)$$

wobei m die Anzahl der Zielenvertauschungen beim Übergang von A nach B bezeichnet. Also genügt es, zu zeigen:

$$d_1(B) = d_2(B) \ .$$

Es sei $B = \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix} = (b_{jk})$.

1. Fall: $Rg(B) < n$. Dann ist $B_n = 0$. Durch Addition von B_1 zu $B_n = 0$ entsteht eine Matrix $B^{(1)}$ mit $\det_i(B^{(1)}) = \det_i(B)$ $(i = 1,2)$ und mit zwei gleichen Zeilen. Also ist

$$d_1(B) = d_2(B) = 0 \ .$$

2. Fall: $Rg(B) = n$. Dann ist $b_{jj} \neq 0$ für $j = 1, \ldots, n$. Für

$$B^{(1)} := \begin{pmatrix} B_1/b_{11} \\ \vdots \\ B_n/b_{nn} \end{pmatrix} =: (b_{jk}^{(1)})$$

gilt dann $b_{jj}^{(1)} = 1$ $(1 \leq j \leq n)$ und nach D.11.7 a)

$$d_i(B^{(1)}) = d_i(B)/\prod_{j=1}^{n} b_{jj} \qquad (i = 1, 2) \ .$$

Durch weitere elementare Zeilenumformungen der Form $Z_j + \lambda Z_k$ (Addition des $(-b_{jn}^{(1)})$-fachen der letzten Zeile zur j-ten Zeile für $j = 1, \ldots, n-1$ u. s. w.; vgl. B.10.10) lässt sich $B^{(1)}$ in die Einheitsmatrix E transformieren. Nach B.11.8 gilt dann $d_i(B^{(1)}) = d_i(E) = 1$. Also gilt insgesamt

$$d_i(B) = \prod_{j=1}^{n} b_{jj} \cdot d_i(B^{(1)}) = \prod_{j=1}^{n} b_{jj} \qquad (i = 1, 2) \ ,$$

also insbesondere $d_1(B) = d_2(B)$. □

Beispiel 11.10 Für $n = 2$ gilt $S_2 = \{id, \sigma\}$, wobei $\sigma = [2, 1]$, also

$$\begin{aligned} \det A &= \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \\ &= \operatorname{sign}(id) a_{1,id(1)} a_{2,id(1)} + \operatorname{sign}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} \\ &= a_{11}a_{22} - a_{12}a_{21} \ . \end{aligned}$$

Weiter überlegt man sich leicht, dass für $n = 3$ gilt

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

Bemerkung 11.11 1. Aus der Formel (2) ergibt sich insbesondere

$$\det(A) = \det(A^T)$$

für alle $A \in K^{n \times n}$.

(Denn: Ist $A^T = (b_{jk}) = (a_{kj})$, so gilt, da $\operatorname{sign}(\sigma) = \operatorname{sign}(\sigma^{-1})$ nach S.11.6,

$$\begin{aligned}\det A &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \\ &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma^{-1}) a_{\sigma^{-1}(\sigma(1)),\sigma(1)} \cdots a_{\sigma^{-1}(\sigma(n)),\sigma(n)} \\ &= \sum_{\tau\in S_n} \operatorname{sign}(\tau) a_{\tau(1),1} \cdots a_{\tau(n),n} \\ &= \sum_{\tau\in S_n} \operatorname{sign}(\tau) b_{1,\tau(1)} \cdots b_{n,\tau(n)} = \det A^T \ .)\end{aligned}$$

2. Aus dem Beweis zu S. 11.9 ergibt sich auch folgende Charakterisierung der Invertierbarkeit einer Matrix: Ist $A \in K^{n\times n}$, so gilt

$$\boxed{A \text{ invertierbar} \Leftrightarrow \det(A) \neq 0}$$

(Denn: Ist B wie im Beweisteil 2. zu S.11.9 so gilt $Rg(A) = Rg(B)$ und nach dem Beweisteil 2. damit

$$\det A = (-1)^m \det B = \begin{cases} (-1)^m \prod_{j=1}^{n} b_{jj} \neq 0, & \text{falls} \quad Rg(A) = n \\ 0 & \text{falls} \quad Rg(A) < n \end{cases} .)$$

Eine zentrale Eigenschaft der Determinantenabbildung liefert

Satz 11.12 *Es seien $A, B \in K^{n\times n}$. Dann gilt*

$$\det(AB) = \det(A)\det(B) \ .$$

Beweis.
1. Fall: $Rg(B) < n$. Dann ist $Rg(AB) < n$ nach S. 8.18, also

$$\det(A)\det(B) = 0 = \det(AB)$$

nach B. 11.11.2.
2. Fall: $Rg(B) = n$. Dann ist $\det(B) \neq 0$ nach B. 11.11.2 .
Wir betrachten $d : K^{n\times n} \to K$ mit

$$d(A) := \frac{\det(AB)}{\det B} \qquad (A \in K^{n\times n}) \ .$$

Wir zeigen: d ist eine Determinantenabbildung, also $d = \det$ nach S. 11.9. Dies ist die Behauptung.

Es sei $A = (a_{jk}) = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}$ und $b \in K^n, \lambda, \mu \in K$. Dann gilt

$$
\begin{aligned}
d\left(\begin{pmatrix} A_1 \\ \vdots \\ \lambda A_j + \mu b^T \\ \vdots \\ A_n \end{pmatrix}\right) &= \frac{1}{\det B} \det\left(\begin{pmatrix} A_1 \\ \vdots \\ \lambda A_j + \mu b^T \\ \vdots \\ A_n \end{pmatrix} B\right) = \\
&= \frac{1}{\det B} \det\left(\begin{pmatrix} A_1 B \\ \vdots \\ \lambda A_j B + \mu b^T B \\ \vdots \\ A_n B \end{pmatrix}\right) = \\
&= \frac{1}{\det B} \cdot \left[\lambda \det\left(\begin{pmatrix} A_1 B \\ \vdots \\ A_j B \\ \vdots \\ A_n B \end{pmatrix}\right) + \mu \det\left(\begin{pmatrix} A_1 B \\ \vdots \\ b^T B \\ \vdots \\ A_n B \end{pmatrix}\right)\right] \\
&= \lambda d(A) + \mu d\left(\begin{pmatrix} A_1 \\ \vdots \\ b^T \\ \vdots \\ A_n \end{pmatrix}\right) ,
\end{aligned}
$$

also ist a) aus D. 11.7 erfüllt. Gilt $A_j = A_k$ für $j < k$, so ist $A_j B = A_k B$, d. h. zwei Zeilen von AB stimmen überein. Also ist

$$d(A) = \frac{1}{\det B} \cdot \det(AB) = 0 .$$

Ist schließlich $A = E$, so ist $AB = B$, also $d(E) = 1$. □

Wir ergänzen diesen Abschnitt durch eine zumindest theoretisch nützliche Formel für die Berechnung von Determinanten.

Definition 11.13 Es seien $A = (a_{jk}) \in K^{n\times n}$, wobei $n > 1$, und $j, k \in \{1, \dots, n\}$. Die Matrix $S_{jk}(A) \in K^{(n-1)\times(n-1)}$, die aus A durch Streichen der j-ten Zeile und der k-ten Spalte entsteht heißt *Streichungsmatrix (bzgl.* (j,k)*)* von A.
(Ist etwa

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

so ist

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

Satz 11.14 (Laplace'scher Entwicklungssatz) *Es sei* $A = (a_{jk}) \in K^{n\times n}$. *Dann gilt für* $k = 1, \dots, n$

$$\det A = \sum_{j=1}^{n} a_{jk}(-1)^{j+k} \det(S_{jk}(A)) \tag{3}$$

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

$$\det A = \sum_{k=1}^{n} a_{jk}(-1)^{j+k} \det(S_{jk}(A)) \tag{4}$$

("Entwicklung nach der j*-ten Zeile").*

Beweis.
Wir zeigen (4). Die Formel (3) ergibt sich damit aus $\det(A) = \det(A^T)$.
Es sei $A_{jk} \in K^{n\times n}$ die Matrix, die aus A durch Ersetzen der j-ten Zeile A_j durch e_k^T und Ersetzen der k-ten Spalte $a^{(k)}$ durch e_j entsteht, d. h.

$$A_{jk} = \begin{pmatrix} a_{11} & \dots & 0 & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ a_{n1} & & 0 & & a_{mn} \end{pmatrix} \leftarrow j$$

$$\uparrow$$

$$k$$

Behauptung: Es gilt

$$\sum_{k=1}^{n} a_{jk} \det A_{jk} = \det A\,.$$

(Denn: Ist $A'_{jk} \in K^{n\times n}$ die Matrix, in der die j-te Zeile A_j durch e_k^T ersetzt ist, so läßt sich A_{jk} aus A'_{jk} durch Zeilenoperationen vom Typ $(Z_\nu + \lambda Z_\mu)$ gewinnen (man addiere zur i-ten Zeile von A'_{jk} das $(-a_{ik})$-fache der j-ten Zeile von A'_{jk} $(i \neq j)$). Also gilt nach B.11.8 $\det A_{jk} = \det A'_{jk}$. Schreibt man $A_j = \sum\limits_{k=1}^{n} a_{jk} e_k^T$, so ergibt sich aus D.11.7 a)

$$\det A = \sum_{k=1}^{n} a_{jk} \det(A'_{jk}) = \sum_{k=1}^{n} a_{jk} \det(A_{jk}) \text{ .)}$$

Durch $(k-1)$ Vertauschungen (falls $k > 1$) benachbarter Spalten erhalten wir aus A_{jk} eine Matrix $B_{jk} \in K^{n\times n}$, in der die $(m+1)$-te Spalte die m-te Spalte von A_{jk} ist $(m < k)$, und deren erste Spalte die k-te Spalte e_j von A_{jk} ist. Also ist nach B. 11.8 (mit "Spalten" statt "Zeilen"; man beachte dabei: aus B.11.11.1 folgt, dass die Eigenschaften aus D.11.7 und B.11.8 auch mit "Spalte" statt "Zeile" gelten)

$$\det A_{jk} = (-1)^{k-1} \det B_{jk} \text{ .}$$

Entsprechend erhält man durch $j-1$ (falls $j > 1$) Vertauschungen von Zeilen aus B_{jk} eine Matrix C_{jk} mit $\det(B_{jk}) = (-1)^{j-1} \det(C_{jk})$, wobei C_{jk} von der Form

$$C_{jk} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & S_{jk}(A) & \\ 0 & & & \end{pmatrix} =: (d_{\mu\nu})_{\mu,\nu=1,\dots,n}$$

ist. Hierfür gilt

$$\begin{aligned} \det(C_{jk}) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma) d_{1,\sigma(1)} \cdot \prod_{\mu=2}^{n} d_{\mu,\sigma(\mu)} \\ &= \sum_{\substack{\sigma\in S_n \\ \sigma(1)=1}} \operatorname{sign}(\sigma) \cdot d_{11} \cdot \prod_{\mu=2}^{n} d_{\mu,\sigma(\mu)} \\ &= \sum_{\tau\in S_{n-1}} \operatorname{sign}(\tau) \cdot \prod_{\mu=1}^{n-1} d_{\mu+1,\tau(\mu)+1} = \det(S_{jk}(A)) \text{ .} \end{aligned}$$

(Man beachte: Jedem $\sigma \in S_n$ mit $\sigma(1) = 1$ entspricht genau ein $\tau \in S_{n-1}$, nämlich τ mit $\tau(\mu) := \sigma(\mu+1) - 1$, und es gilt dabei $\operatorname{sign}(\sigma) = \operatorname{sign}(\tau)$.)
Also gilt insgesamt

$$\begin{aligned} \det A &= \sum_{k=1}^{n} a_{jk} \det A_{jk} = \sum_{k=1}^{n} a_{jk} (-1)^{(j-1)+(k-1)} \det(C_{jk}) \\ &= \sum_{k=1}^{n} a_{jk} (-1)^{j+k} \det(S_{jk}(A)) \text{ .} \end{aligned}$$

□

Beispiel 11.15 Es sei

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

Dann gilt (Entwicklung nach der 2. Zeile):

$$\begin{aligned} \det A &= \sum_{k=1}^{4} a_{2k}(-1)^{2+k} \det(S_{2k}(A)) = \\ &= 4 \cdot (-1) \det \begin{pmatrix} 2 & 3 & 2 \\ 1 & 2 & 0 \\ 1 & 3 & 1 \end{pmatrix} + 1 \cdot \det \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 2 & 1 & 3 \end{pmatrix} = \\ &= -4(4+6-4-3) + 1(3+8+9-6-2-18) = -18 . \end{aligned}$$

Bemerkung 11.16 Determinanten können auch dazu verwandt werden, die Lösung $x = A^{-1}b$ eines LGS $Ax = b$ mit invertierbarer Matrix $A \in K^{n\times n}$ darzustellen: Sind $a^{(k)}$ die Spalten von A, so gilt

$$x_k = \frac{1}{\det(A)} \cdot \det(a^{(1)}, \dots, a^{(k-1)}, b, a^{(k+1)}, \dots, a^{(n)}) \qquad (k = 1, \dots, n) .$$

(Denn: Für $x = (x_1, \dots, x_n)^T$ gilt

$$\sum_{\nu=1}^{n} x_\nu a^{(\nu)} = b .$$

Also folgt mit D.11.7 a) (und B.11.11.1)

$$\begin{aligned} & \det(a^{(1)}, \dots, a^{(k-1)}, b, a^{(k+1)}, \dots, a^{(n)}) = \\ = \ & \det(a^{(1)}, \dots, a^{(k-1)}, \sum_{\nu=1}^{n} x_\nu a^{(\nu)}, a^{(k+1)}, \dots, a^{(n)}) = \\ = \ & \sum_{\nu=1}^{n} x_\nu \det(a^{(1)}, \dots, a^{(k-1)}, a^{(\nu)}, a^{(k+1)}, \dots, a^{(n)}) = x_k \det(A) .) \end{aligned}$$

Darüber hinaus lässt sich auch die Inverse A^{-1} durch Determinanten darstellen: Ist $A^{-1} =: (a_{jk}^{(-1)})$, so gilt

$$a_{jk}^{(-1)} = \frac{1}{\det A}(-1)^{j+k} \det(S_{kj}(A)) \qquad (j, k = 1 \dots, n)$$

(siehe [Ü]).

12 Eigenwerte

Ein wesentliches Anliegen der Linearen Algebra besteht darin, Aussagen über das Abbildungsverhalten linearer Selbstabbildungen T (d. h. $T \in L(V)$) zu machen. Besonders einfach ist die Wirkung von T auf ein $v \in V$, wenn Tv lediglich eine "Streckung" von v bewirkt, d. h. $Tv = \lambda v$ für ein $\lambda \in K$ gilt. Dies führt auf

Definition 12.1 Es sei V ein linearer Raum über K, und es sei $T \in L(V)(= L(V,V))$. Ein $\lambda \in K$ heißt *Eigenwert von* T, falls ein $v \in V, v \neq 0$, existiert mit

$$Tv = \lambda v \qquad (\text{bzw. } (T - \lambda I)(v) = 0) \ ,$$

wobei $I := id_V$. Ist λ Eigenwert von T, so heißt $\mathrm{Kern}(T - \lambda I)$ *Eigenraum (von T) zu* λ und jedes $v \in \mathrm{Kern}(T - \lambda I)$ heißt *Eigenvektor (von T) zu* λ.
Indem wir $A \in K^{n\times n}$ wieder mit $T \in L(K^n)$ mit $Tx = Ax$ $(x \in K^n)$ identifizieren (siehe B./D.8.9) können wir damit auch von "Eigenwert von A" bzw. "Eigenraum" oder "Eigenvektor von A" reden.
Offenbar ist λ genau dann Eigenwert von A, wenn das LGS $(A-\lambda E)x = 0$ eine Lösung $x \neq 0$ hat, also genau dann, wenn $\det(A - \lambda E) = 0$ ist (B. 11.11).

Beispiel 12.2 1. Es sei

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

Dann sind ± 1 Eigenwerte von A (denn

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

2. Es sei

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

Dann gilt

$$\det(A - \lambda E) = \det \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix} = \lambda^2 + 1 \ .$$

Also: Ist $K = \mathbb{R}$, so ist $\det(A - \lambda E) \neq 0$ für alle $\lambda \in \mathbb{R}$, d. h. A hat keinen Eigenwert. Ist andererseits $K = \mathbb{C}$, so gilt $\det(A - \lambda E) = 0$ für $\lambda = \pm i$, d. h. $\lambda_{1,2} = \pm i$ sind Eigenwerte von A.

Insbesondere zeigt das Beispiel, dass nicht jede Matrix $A \in \mathbb{R}^{2\times 2}$ einen Eigenwert besitzt!

Wir betrachten jetzt oft $K = \mathbb{R}$ oder $K = \mathbb{C}$. In diesem Fall schreiben wir $K =: \mathbb{K}$ wobei dann $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.

Satz 12.3 *Es sei $A \in \mathbb{K}^{n\times n}$. Dann ist $P : \mathbb{K} \to \mathbb{K}$, definiert durch*

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

ein Polynom vom Grad n (mit höchstem Koeffizient $(-1)^n$).

Beweis.
Es gilt $A - \lambda E = (\tilde{a}_{jk})$ mit

$$\tilde{a}_{jk} = \begin{cases} a_{jj} - \lambda & , \text{ falls } j = k \\ a_{jk} & , \text{ falls } j \neq k \end{cases} .$$

Also folgt aus (2) mit gewissen $\alpha_\nu \in \mathbb{K}$

$$\begin{aligned} \det(A - \lambda E) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\tilde{a}_{1,\sigma(1)} \cdots \tilde{a}_{n,\sigma(n)} = \\ &= (a_{11} - \lambda) \cdots (a_{nn} - \lambda) + \sum_{\substack{\sigma\in S_n \\ \sigma\neq id}} \operatorname{sign}(\sigma)\tilde{a}_{1,\sigma(1)} \cdots \tilde{a}_{n,\sigma(n)} \\ &= (-1)^n\lambda^n + \sum_{\nu=0}^{n-1} \alpha_\nu\lambda^\nu + \sum_{\substack{\sigma\in S_n \\ \sigma\neq id}} \tilde{a}_{1,\sigma(1)} \cdots \tilde{a}_{n,\sigma(n)} . \end{aligned}$$

Da in jedem Summanden von $\sum_{\sigma\in S_n\setminus\{id\}} \tilde{a}_{1,\sigma(1)} \cdots \tilde{a}_{n,\sigma(n)}$ Faktoren der Art $(a_{jj} - \lambda)$ höchstens $(n-1)$-mal auftauchen, ist insgesamt

$$P(\lambda) = \det(A - \lambda E) = (-1)^n\lambda^n + \sum_{\nu=0}^{n-1} \beta_\nu\lambda^\nu$$

mit gewissen $\beta_\nu \in \mathbb{K}$. □

Bemerkung und Definition 12.4 Es sei $A \in \mathbb{K}^{n\times n}$. Dann sind die Eigenwerte von A genau die Nullstellen des Polynoms P_A. Das Polynom $P_A \in \Pi_n$ heißt *charaktristisches Polynom von A*

Definition 12.5 Es seien $A, B \in K^n$. Dann heißen A und B *ähnlich*, falls eine invertierbare Matrix $C \in K^{n\times n}$ existiert mit

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

Damit gilt:

Satz 12.6 *Sind A, $B \in \mathbb{K}^{n\times n}$ ähnlich, so ist*

$$P_A = P_B \ .$$

Beweis.
Es sei $C \in GL_n(\mathbb{K})$ mit $B = C^{-1}AC$. Dann gilt für $\lambda \in \mathbb{K}$

$$B - \lambda E = C^{-1}AC - C^{-1}(\lambda E)C = C^{-1}(A - \lambda E)C \ ,$$

also mit S. 11.12 (man beachte: $\det(C^{-1})\det(C) = \det(C^{-1}C) = \det(E) = 1$)

$$\begin{aligned} P_B(\lambda) &= \det(B - \lambda E) = \det(C^{-1})\det(A - \lambda E)\det(C) = \\ &= \det(A - \lambda E) = P_A(\lambda) \qquad (\lambda \in \mathbb{K}) \ . \end{aligned}$$

□

Satz 12.7 (Basistransformationssatz) *Es sei V ein endlich-dimensionaler linearer Raum über K, und es sei $T \in L(V)$. Ferner seien M und N Basen von V, und es seinen A bzw. B die Matrizen von T bzgl. M, M bzw. N, N, d. h.*

$$A = \varphi_M(T) \ , \qquad B = \varphi_N(T) \ ,$$

wobei $\varphi_M := \varphi_{M,M}$ wie in S.8.5. Ist $C = \varphi_{M,N}(I)$, so gilt

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

Insbesondere sind A und B ähnlich.

Beweis.
Zunächst gilt nach S.8.8

$$E = \varphi_{M,M}(I) = \varphi_{M,N}(I) \cdot \varphi_{N,M}(I) = C \cdot \varphi_{N,M}(I)$$

d. h. $C^{-1} = \varphi_{N,M}(I)$. Also gilt wieder nach S.8.8

$$B = \varphi_N(T) = \varphi_{N,N}(I \circ T \circ I) = \varphi_{N,M}(I) \cdot \varphi_{M,M}(T) \cdot \varphi_{M,N}(I) = C^{-1}AC \ .$$

□

Definition 12.8 Es seien V ein n-dimensionaler linearer Raum über $\mathbb{K}$ und $T \in L(V)$. Ist M eine Basis von V, so heißt das Polynom vom Grad n

$$P := P_T := P_{\varphi_M(T)}$$

charakteristisches Polynom von T. (Wichtig: Nach S.12.6 und S.12.7 ist P_T unabhängig von der Wahl von M!).

Satz 12.9 *Mit den Bezeichnungen aus D. 12.8 sind für $\lambda \in \mathbb{K}$ äquivalent*

a) λ ist Eigenwert von T,

b) λ ist Eigenwert von $\varphi_M(T)$,

c) $P_T(\lambda) = 0$.

Beweis.
1. Ist $\psi_M : V \to \mathbb{K}^n$ die Koordinatenabbildung bzgl. M (vgl. B./D.6.8) und $A := \varphi_M(T)$, so gilt

$$Ax = (\psi_M \circ T \circ \psi_M^{-1})x \qquad (x \in K^n) \; .$$

(Denn: Für $k = 1, \ldots, n$ gilt $\psi_M(v_k) = e_k$ bzw. $\psi_M^{-1}(e_k) = v_k$. Nach der Definition der Koordinatenmatrix $A = (a_{jk})$ gilt damit für $k = 1, \ldots, n$

$$\begin{aligned}(\psi_M \circ T \circ \psi_M^{-1})e_k &= \psi_M(Tv_k) = \psi_M(\sum_{j=1}^{n} a_{jk}v_j) = \\ &= \sum_{j=1}^{n} a_{jk}\psi_M(v_j) = \sum_{j=1}^{n} a_{jk}e_j = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{pmatrix} = Ae_k \; .\end{aligned}$$

Da $x \mapsto Ax$ und $\psi_M \circ T \circ \psi_M^{-1}$ linear auf K^n sind, folgt die Behauptung mit S.8.1.)
2. a) $\Rightarrow$ b): Ist $Tv = \lambda v$ für ein $v \in K \setminus \{0\}$, so gilt für $x = \psi_M(v) \in \mathbb{K}^n \setminus \{0\}$ nach 1.

$$Ax = A\psi_M(v) = \psi_M(Tv) = \psi_M(\lambda v) = \lambda\psi_M(v) = \lambda x \; .$$

b) $\Rightarrow$ a): Ist $Ax = \lambda x$ für ein $x \neq 0$, so gilt für $v = \psi_M^{-1}(x) \in V \setminus \{0\}$ nach 1.

$$Tv = T\psi_M^{-1}(x) = \psi_M^{-1}Ax = \psi_M^{-1}(\lambda x) = \lambda\psi_M^{-1}(x) = \lambda v \; ,$$

also ist λ Eigenwert von T.
3. Die Äquivalenz von b) und c) ergibt sich sofort aus D./B.12.4 und D.12.8. □

In B. 12.2 hatten wir gesehen, dass für gerades n lineare Abbildungen auf $\mathbb{R}^n$ i. a. keine Eigenwerte besitzen (man setze für $n = 2$ etwa $Tx = Ax$ mit $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$). Unter Verwendung des Fundamentalsatzes der Algebra bzw. des Zwischenwertsatzes aus der Analysis können wir jedoch folgendes wichtige Ergebnis beweisen.

Satz 12.10 *Es sei V linearer Raum über $\mathbb{K}$ mit $\dim(V) = n \in \mathbb{N}$, und es sei $T \in L(V)$*
1. Ist $\mathbb{K} = \mathbb{C}$, so hat T mindestens einen Eigenwert.
2. Ist $\mathbb{K} = \mathbb{R}$, und ist n ungerade, so hat T mindestens einen Eigenwert.

Beweis.
1. Nach dem Fundamentalsatz der Algebra gilt: Ist $P \in \Pi_{\mathbb{C}}$ mit $P(z) = \sum_{\nu=0}^{n} a_\nu z^\nu$ und $a_n \neq 0$ (d. h. $\deg(P) = n$), so existieren $z_1, \dots, z_n \in \mathbb{C}$ mit

$$P(z) = a_n \cdot \prod_{k=1}^{n} (z - z_k) \qquad (z \in \mathbb{C}) ,$$

d. h. P "zerfällt in Linearfaktoren". Insbesondere hat P also Nullstellen (in $\mathbb{C}$). Wendet man dies auf $P = P_T$ an, so ergibt sich 1. aus S.12.9 (man beachte, dass $a_n = (-1)^n \neq 0$ nach S. 12.3).
2. Ist $P = P_T$ so ist nach S. 12.3

$$P(x) = \sum_{\nu=0}^{n} a_\nu x^\nu \qquad (x \in \mathbb{R})$$

mit $a_n = -1$ (da n ungerade). Also gilt für $x \neq 0$

$$\frac{P(x)}{x^n} = -1 + \sum_{\nu=0}^{n-1} a_\nu x^{\nu-n} \to -1 \qquad (x \to \pm\infty) ,$$

d. h. es existiert ein $R > 0$ mit

$$P(x) < -\frac{x^n}{2} < 0 \quad \text{für } x > R \qquad \text{und} \qquad P(x) > -\frac{x^n}{2} > 0 \quad \text{für } x < -R .$$

Nach dem Zwischenwertsatz für stetige Funktionen (siehe Analysis) existiert ein $\lambda \in \mathbb{R}$ mit $P(\lambda) = 0$. Also ist λ nach S. 12.9 Eigenwert von T. □

Zum Abschluss dieses Abschnitts wollen wir eine Charakterisierung "diagonalisierbarer" linearer Abbildungen herleiten. Dazu beweisen wir zunächst:

Satz 12.11 *Es sei V ein linearer Raum über K, und es es $T \in L(V)$. Ferner seien $\lambda_1, \dots, \lambda_m$ Eigenwerte von T mit $\lambda_j \neq \lambda_k$ für $k \neq j$.*
1. Sind $v_1, \dots, v_m \in V \setminus \{0\}$ zugehörige Eigenvektoren (d. h. $Tv_j = \lambda_j v_j$ $(j = 1, \dots, m)$), so sind $v_1, \dots, v_m$ linear unabhängig.
2. Die Summe der zugehörigen Eigenräume ist direkt, d. h.

$$\sum_{j=1}^{m} \mathrm{Kern}(T - \lambda_j I) = \bigoplus_{j=1}^{m} \mathrm{Kern}(T - \lambda_j I) \qquad (j = 1, \dots, m) .$$

Beweis.
1. Angenommen, $v_1, \dots, v_m$ sind linear abhängig. Nach dem Beweisschritt 2. zu S.5.8 existiert

$$k := \min\{j \in \{1, \dots, m\} : v_j \in < v_1, \dots, v_{j-1} >\}$$

(mit $< \emptyset >= \{0\}$), und es gilt $k \geq 2$ und $v_1, \ldots, v_{k-1}$ sind linear unabhängig. Sind $\beta_1, \ldots, \beta_{k-1} \in K$ so, dass

$$v_k = \sum_{j=1}^{k-1} \beta_j v_j \ ,$$

so folgt

$$Tv_k = T\left(\sum_{j=1}^{k-1} \beta_j v_j\right) = \sum_{j=1}^{k-1} \beta_j \lambda_j \nu_j \ .$$

Also ergibt sich

$$0 = \lambda_k v_k - Tv_k = \sum_{j=1}^{k-1} \beta_j \lambda_k v_j - \sum_{j=1}^{k-1} \beta_j \lambda_j v_j = \sum_{j=1}^{k-1} \beta_j (\lambda_k - \lambda_j) v_j \ .$$

Da $v_1, \ldots, v_{k-1}$ linear unabhängig sind, folgt $\beta_j(\lambda_k - \lambda_j) = 0$ für $j = 1, \ldots, k-1$, also nach Voraussetzung $\beta_j = 0$ für $j = 1, \ldots, k-1$ und damit $v_k = 0$, im Widerspruch zur Voraussetzung.
2. Wir zeigen: Die Summe ist direkt. Dazu sei $U_j := \text{Kern}(T - \lambda_j \cdot I)$ $(j = 1, \ldots, m)$ und

$$u \in U_j \cap \sum_{k \neq j} U_k \ .$$

Dann gilt $T(u) = \lambda_j u$ und $u = \sum\limits_{k \neq j} u_k$ mit $u_k \in U_k$ $(k = 1, \ldots, m;\ k \neq j)$.
Angenommen, es ist $u \neq 0$. Dann ist $I = \{k \neq j : u_k \neq 0\} \neq \emptyset$ und $u_j = \sum\limits_{k \in I} u_k$. Da $T(u_k) = \lambda_k u_k$ für $k \in I$ gilt, widerspricht dies 1. Also ist $u = 0$ und damit ist Summe direkt. □

Wir wenden uns dem Problem der "Diagonalisierbarkeit" linearer Abbildungen zu. Worum geht es dabei? Angenommen, $T \in L(V)$ ist so, dass eine Basis $(v_1, \ldots, v_n)$ aus Eigenvektoren existiert. Dann ist das Abbildungsverhalten von T sehr übersichtlich: Es gilt nämlich mit $Tv_k = \lambda_k v_k$ für alle $v \in V$

$$T(v) = T\left(\sum_{k=1}^{n} \beta_k v_k\right) = \sum_{k=1}^{n} \beta_k T v_k = \sum_{k=1}^{n} \beta_k \lambda_k v_k \qquad \left(v = \sum_{k=1}^{n} \beta_k v_k\right) \ .$$

Ist $A = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix} \in K^{n \times n}$, so hat $T \in L(K^n)$ mit $Tx = Ax$ offenbar obige Eigenschaft für $(v_1, \ldots, v_n) = (e_1, \ldots, e_n)$
(denn es gilt $T(e_k) = Ae_k = \lambda_k e_k$ für $k = 1, \ldots, n$).

Definition 12.12 1. Eine Matrix $D = (d_{jk}) \in K^{n\times n}$ heißt *Diagonalmatrix*, falls $\lambda_1, \dots, \lambda_n \in K$ existieren mit

$$d_{jk} = \left\{ \begin{array}{ccc} \lambda_k & , & \text{falls} \quad j = k \\ 0 & , & \text{falls} \quad j \neq k \end{array} \right\} = \lambda_k \delta_{jk} \ ,$$

d. h.

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

Wir schreiben dann $D =: \operatorname{diag}(\lambda_1, \dots, \lambda_n)$.
2. Ist V ein linearer Raum über K, so heißt $T \in L(V)$ *diagonalisierbar* falls eine Basis von V aus Eigenvektoren existiert.
Ist $A \in K^{n\times n}$, so heißt A *diagonalisierbar*, falls $x \mapsto Ax \in L(K^n)$ diagonalisierbar ist.

Es gilt dann:

Satz 12.13 *Es sei V ein n-dimensionaler linearer Raum über K, und es sei $T \in L(V)$. Dann sind äquivalent:*
a) T ist diagonalisierbar.
b) Es existiert eine Basis M von V so, dass $\varphi_M(T)$ eine Diagonalmatrix ist.
c) Es ist $V = \bigoplus_{j=1}^{m} Kern(T - \lambda_j \cdot I)$, wobei $\lambda_1, \dots, \lambda_m$ die paarweise verschiedenen Eigenwerte von T sind.
d) Es ist $\dim(V) = \sum_{j=1}^{m} \dim(Kern(T - \lambda_j I))$, wobei $\lambda_1, \dots, \lambda_m$ wie in c).

Beweis.
a) $\Rightarrow$ b): Ist $M = (v_1, \dots, v_n)$ eine Basis aus Eigenvektoren von T, so existieren $\mu_1, \dots, \mu_n \in K$ mit $Tv_k = \mu_k v_k$ $(k = 1, \dots, n)$. Also gilt nach Definiton von $\varphi_M(T) = A = (a_{jk})$:

$$a_{jk} = \left\{ \begin{array}{ccc} \mu_k & , & \text{falls} \quad j = k \\ 0 & , & \text{sonst} \end{array} \right\} (= \mu_k \delta_{jk}) \ ,$$

d. h. $\varphi_M(T) = \operatorname{diag}(\mu_1, \dots, \mu_n)$.
b) $\Rightarrow$ a): Existiert eine Basis $M = (v_1, \dots, v_n)$ von V mit $A = \varphi_M(T) = \operatorname{diag}(\mu_1, \dots, \mu_n)$, so gilt für $k = 1, \dots, n$

$$T(v_k) = \sum_{j=n}^{n} a_{jk} v_j = \mu_k v_k$$

(wobei $A = (a_{jk})$). Also sind $v_1, \dots, v_n$ Eigenvektoren.
a) $\Rightarrow$ c): Nach Voraussetzung existiert eine Basis $M = (v_1, \dots, v_n)$ aus Eigenvektoren mit zugehörigen Eigenwerten $\mu_1, \dots, \mu_n$. Also gilt für jedes $v \in V$, $v = \sum_{k=1}^{n} \beta_k v_k$,

$$v = \sum_{k=1}^{n} \beta_k v_k = \sum_{j=1}^{m} \sum_{k:\mu_k=\lambda_j} \beta_k v_k \in \sum_{j=1}^{m} \mathrm{Kern}(T - \lambda_j \cdot I)$$

(Man beachte dabei:

$$T(\sum_{k:\mu_k=\lambda_j} \beta_k v_k) = \sum_{k:\mu_k=\lambda_j} \beta_k T v_k = \lambda_j \sum_{k:\mu_k=\lambda_j} \beta_k v_k \ ,$$

d. h. $\sum_{k:\mu_k=\lambda_j} \beta_k v_k \in \mathrm{Kern}(T - \lambda_j \cdot I)$).
Also ist $V = \sum_{j=1}^{m} \mathrm{Kern}(T - \lambda_j \cdot I)$.
Nach S.12.11.2 ist die Summe direkt.
c) $\Rightarrow$ d): Ergibt sich sofort aus F. 5.19.
d) $\Rightarrow$ a): Es gelte

$$\dim V = \sum_{j=1}^{m} \dim(\mathrm{Kern}(T - \lambda_j \cdot I)) \ .$$

Wir wählen Basen $(v_1^{(j)}, \dots, v_{d_j}^{(j)})$ von $U_j := \mathrm{Kern}(T - \lambda_j \cdot I)$. Dann ist $n = \sum_{j=1}^{m} d_j$, d. h.

$$M := (v_1^{(1)}, \dots, v_{d_1}^{(1)}, \dots, v_1^{(m)}, \dots, v_{d_m}^{(m)})$$

besteht aus n Elementen. Wir zeigen: M ist linear unabhängig (dann ist M auch Basis von V nach S. 5.15, also eine Basis aus Eigenvektoren).
Es seien also $\alpha_k^{(j)} \in K$ für $k = 1, \dots, d_j$; $j = 1, \dots, m$ mit

$$0 = \sum_{j=1}^{m} \sum_{k=1}^{d_j} \alpha_k^{(j)} v_k^{(j)} = \sum_{j=1}^{m} u_j$$

wobei $u_j = \sum_{k=1}^{d_j} \alpha_k^{(j)} v_k^{(j)} \in U_j$. Aus S.12.11.2 folgt, dass $u_j = 0$ für $j = 1, \dots, m$ gelten muss, also

$$\sum_{k=1}^{d_j} \alpha_k^{(j)} v_k^{(j)} = 0 \qquad (j = 1, \dots, m) \ .$$

Da $(v_1^{(j)}, \dots, v_{d_j}^{(j)})$ für $j = 1, \dots, m$ linear unabhängig sind, folgt $\alpha_k^{(j)} = 0$, für $k = 1, \dots, d_j$ und $j = 1, \dots, m$, d. h. M ist linear unabhängig. $\square$

Bemerkung 12.14 1. Insbesondere folgt aus S. 12.13 (oder auch schon aus S. 12.11), dass ein $T \in L(V)$ sicher dann diagonalisierbar ist, wenn $n = \dim(V)$ paarweise verschiedene Eigenwerte existieren. Dies ist jedoch nicht notwendig für die Diagonalisierbarkeit (etwa $T = id_{K^n}$ hat nur den Eigenwert $\lambda = 1$, ist aber diagonalisierbar, denn die kanonische Basis ist eine Basis aus Eigenvektoren).
2. Notwendig für die Diagonaliserbarkeit ist natürlich, dass zumindest ein Eigenwert existiert. Dies ist jedoch nicht hinreichend.
(Man betrachte etwa $T \in L(\mathbb{R}^2)$ mit

$$Tx = Ax := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \left(x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \right) .$$

Dann gilt

$$P_T(\lambda) = P_A(\lambda) = \det(A - \lambda E) = \det \begin{pmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 ,$$

d. h. $\lambda = 1$ ist einziger Eigenwert von A. Es gilt

$$0 = (T - I)x = (A - E)x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

genau dann, wenn $x_2 = 0$ ist, d. h.

$$\mathrm{Kern}(T - I) = \mathrm{Kern}(A - E) = \mathbb{R} \times \{0\} ,$$

also $\dim(\mathrm{Kern}(T - I)) = 1 < 2 = \dim(V)$. Nach S. 12.13 ist also T nicht diagonalisierbar.)
3. Eine Matrix $A \in K^{n \times n}$ ist diagonalisierbar genau dann, wenn ein $C \in GL_n(K)$ und $\lambda_1, \ldots, \lambda_n \in K$ existieren mit

$$C^{-1}AC = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) .$$

(Denn: Ist A diagonalisierbar, so existiert nach S.12.13 eine Basis M von K^n so, dass die Matrix von $x \mapsto Ax$ bzgl. M eine Diagonalmatrix ist. Nach S. 12.7 existiert dann ein C wie behauptet (man beachte dabei: A ist die Matrix von $x \mapsto Ax$ bzgl. der kanonischen Basis).
Existiert umgekehrt ein solches C, so ist $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ die Matrix von $x \mapsto Ax$ bzgl. $(Ce_1, \ldots, Ce_n)$, denn es gilt

$$ACe_k = C\mathrm{diag}(\lambda_1, \ldots, \lambda_n)e_k = C\lambda_k e_k = \lambda_k Ce_k$$

für $k = 1, \ldots, n$.)

Definition 12.15 Es sei V ein endlich-dimensionaler linearer Raum über K, und es sei $T \in L(V)$. Ist $\lambda \in K$ Eigenwert von T, so heißt $\dim(\mathrm{Kern}(T - \lambda \cdot I))$ *geometrische Vielfachheit* von λ.
(Nach S. 12.13 ist T genau dann diagonalisierbar, wenn die Summe der geometrischen Vierfachheiten aller Eigenwerte $= \dim(V)$ ist.)

13 Skalarprodukte und Normen

Ist $x = \binom{x_1}{x_2} \in \mathbb{R}^2$, so ist nach dem Satz von Pythagoras bekanntlich der Abstand von 0 zu x (bzw. die "Länge" des Vektors x) gegeben durch

$$||x|| := \sqrt{x_1^2 + x_2^2} = \sqrt{x^T x} \, .$$

Entsprechend gilt für $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$

$$||x|| = \sqrt{x_1^2 + x_2^2 + x_3^2} = \sqrt{x^T x} \, .$$

Weiterhin kann man sich überlegen, dass der "Winkel α zwischen x und y" in $\mathbb{R}^2 \setminus \{0\}$ gegeben ist durch

$$\alpha = \arccos\left(\frac{x^T y}{||x||\,||y||}\right) \, ,$$

wobei man $< x, y >:= x^T y$ als Skalarprodukt von x und y bezeichnet. Um Analysis oder Geometrie in allgmeineren linearen Räumen treiben zu können, brauchen wir Abstände, Normen oder Skalarprodukte in diesen Räumen.

Definition 13.1 Es sei V ein linearer Raum über $\mathbb{K}$. Eine Abbildung $< \cdot, \cdot >: V \times V \to \mathbb{K}$ heißt *Skalarprodukt* (auf V) (oder *inneres Produkt* (auf V)), falls folgende Bedingungen gelten:
(S.1) $< x, x > \geq 0$ für alle $x \in V$ und $< x, x >= 0$ gilt genau dann wenn $x = 0$ ist,
(S.2) $< x, y >= \overline{< y, x >}$ für alle $x, y \in V$,
(S.3) $x \mapsto < x, y >$ ist linear für alle $y \in V$ (d. h. für alle $y \in V, x_1, x_2 \in V, \lambda_1, \lambda_2 \in \mathbb{K}$ gilt

$$< \lambda_1 x_1 + \lambda_2 x_2, y >= \lambda_1 < x_1, y > + \lambda_2 < x_2, y >).$$

Einen linearen Raum mit Skalarprodukt $(V, < \cdot, \cdot >)$ nennt man *unitären Raum*. Im Falle $\mathbb{K} = \mathbb{R}$ spricht man auch von einem *euklidischen Raum*.

Bemerkung 13.2 1. Für $z \in \mathbb{C}$ bedeutet "$z \geq 0$", daß z reell und nichtnegativ ist, d. h. auch im Falle $\mathbb{K} = \mathbb{C}$ ist $< x, x >$ stets reell (und ≥ 0).
2. "$\overline{\cdot}$" in (S.2) bedeutet komplexe Konjugation, d. h. ist $z = x + iy$ mit $x, y \in \mathbb{R}$, so ist $\overline{z} = x - iy$. Ist $\mathbb{K} = \mathbb{R}$, so ist als (S.2) als $< x, y >=< y, x >$ $(x, y \in V)$ zu lesen.
3. Aus (S.2) und (S.3) folgt, daß $y \mapsto < x, y >$ "konjugiert-linear" ist, d. h. für $\lambda_1, \lambda_2 \in \mathbb{K}$ und $y_1, y_2 \in V$ gilt

$$< x, \lambda_1 y_1 + \lambda_2 y_2 >= \overline{\lambda_1} < x, y_1 > + \overline{\lambda_2} < x, y_2 >$$

(und dies für alle $x \in V$).

(denn:

$$\begin{aligned} < x, \lambda_1 y_1 + \lambda_2 y_2 > &= \overline{< \lambda_1 y_1 + \lambda_2 y_2, x >} = \\ &= \overline{\lambda_1 < y_1, x > + \lambda_2 < y_2, x >} = \overline{\lambda_1} \overline{< y_1, x >} + \overline{\lambda_2} \overline{< y_2, x >} \\ &= \overline{\lambda_2} < x, y_1 > + \overline{\lambda_2} < x, y_2 >) \end{aligned}$$

Beispiel 13.3 1. Es sei $V = \mathbb{K}^n$ und $< \cdot, \cdot >: \mathbb{K}^n \times \mathbb{K}^n \to \mathbb{K}$ definiert durch

$$< x, y > := x^T \overline{y} = \sum_{j=1}^{n} x_j \overline{y_j} \qquad (x = (x_1, \dots, x_n)^T, y = (y_1, \dots, y_n)^T \in \mathbb{K}^n)$$

(wobei $\overline{y} := (\overline{y_1}, \dots, \overline{y_n})$). Dann ist $< \cdot, \cdot >$ ein Skalarprodukt auf $\mathbb{K}^n$, das auch als *kanonisches Skalarprodukt* bezeichnet wird ([Ü]).
2. Es sei $[a, b] \subset \mathbb{R}$ ein kompaktes Intervall. Wir setzen

$$C[a, b] := C([a, b], \mathbb{K}) := \{f : [a, b] \to \mathbb{K} : f \text{ stetig auf } [a, b]\} .$$

Dann ist $C[a, b]$ ein Unterraum von $\mathbb{K}^{[a,b]} = \{f : [a, b] \to \mathbb{K}\}$, also ein linearer Raum über $\mathbb{K}$.
(In der Analysis zeigt man: Sind $f, g : [a, b] \to \mathbb{K}$ stetig und $\lambda \in \mathbb{K}$, so sind auch $f + g$ und λf stetig auf $[a, b]$).
Wir definieren $< \cdot, \cdot >: C[a, b] \times C[a, b] \to \mathbb{K}$ durch

$$< f, g > := \int_a^b f(t) \overline{g(t)} dt \qquad (f, g \in C[a, b]) .$$

Dann ist $< \cdot, \cdot >$ ein Skalarprodukt auf $C[a, b]$.
(Denn: (S. 2) und (S. 3) ergeben sich aus bekannten Rechenregeln für Integrale ($\to$ Analysis). Da $f(t)\overline{f}(t) = |f(t)|^2 \geq 0$ für alle $t \in [a, b]$ ist, ist $< f, f > \geq 0$ ($\to$ Analysis). Wichtig, und nicht so klar, ist, dass aus $f \neq 0$ schon $< f, f > > 0$ folgt.)

Definition 13.4 Es sei V ein linearer Raum über $\mathbb{K}$. Eine Abbildung $|| \cdot || : V \to \mathbb{R}$ heißt *Norm* (auf V), falls die folgenden Bedingungen gelten:

(N.1) $||x|| \geq 0$ für alle $x \in V$, und es gilt $||x|| = 0$ genau dann, wenn $x = 0$ ist,

(N.2) $||\lambda x|| = |\lambda| \, ||x||$ für alle $x \in V, \lambda \in \mathbb{K}$,

(N.3) (Dreiecksungleichung)
$||x + y|| \leq ||x|| + ||y||$ für alle $x, y \in V$.

$(V, || \cdot ||)$ heißt dann ein *normierter Raum*.

Der folgende Satz zeigt, dass jedes Skalarprodukt eine Norm induziert:

Satz 13.5 *Es sei $(V, <\cdot,\cdot>)$ ein unitärer Raum über $\mathbb{K}$. Wir setzen für $x \in V$*

$$||x|| := \sqrt{<x,x>}\ .$$

Dann gilt:
1. Für alle $x, y \in V$ ist

$$||x+y||^2 = ||x||^2 + ||y||^2 + <x,y> + <y,x>\ .$$

2. (Cauchy-Schwarz'sche Ungleichung)
Für alle $x, y \in V$ gilt

$$|<x,y>| \leq ||x||\ ||y||\ .$$

3. $||\cdot||$ ist eine Norm auf V, die sog. induzierte Norm auf V.

Beweis.
1. Für $x, y \in V$ gilt

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

2. Ist $y = 0$, so ist $<x,y> = <x,0> = 0$, also die Behauptung trivial. Es sei also $y \neq 0$. Für alle $\lambda \in \mathbb{K}$ gilt nach 1., (S.1), (S. 3) sowie B. 13.2.3

$$\begin{aligned} 0 &\leq <x-\lambda y, x-\lambda y> = ||x||^2 + ||-\lambda y||^2 + <x,-\lambda y> + <-\lambda y, x> \\ &= ||x||^2 + \lambda\overline{\lambda} <y,y> - \overline{\lambda} <x,y> - \lambda <y,x>\ . \end{aligned}$$

Für $\lambda := <x,y> / ||y||^2 = <x,y> / <y,y>$ ergibt sich

$$0 \leq ||x||^2 + <x,y>\overline{<x,y>}/||y||^2 - \overline{<x,y>}<x,y>/||y||^2 - <x,y><y,x>/||y||^2\ .$$

also (da $<y,y> > 0$)

$$|<x,y>|^2 = <x,y>\overline{<x,y>} = <x,y><y,x> \leq ||x||^2 \cdot ||y||^2 = (||x||\ ||y||)^2\ .$$

3. (N.1) ergibt sich sofort aus (S.1).
Sind $x \in V, \lambda \in \mathbb{K}$, so gilt

$$||\lambda x|| = <\lambda x, \lambda x>^{1/2} = (\lambda\overline{\lambda} <x,x>)^{1/2} = |\lambda|\,||x||\ ,$$

also (N.2). Schließlich gilt für $x, y \in V$ nach 1. und 2.

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

also (N.3). □

Beispiel 13.6 1. Es sei $V = \mathbb{K}^n$ und $< \cdot, \cdot >$ das kanonische Skalarprodukt. Dann ist nach S. 13.5

$$||x||_2 := ||x|| := \sqrt{x^T \overline{x}} = \sqrt{\sum_{j=1}^{n} |x_j|^2} \qquad (x = (x_1, \dots, x_n)^T \in \mathbb{K}^n)$$

eine Norm auf $\mathbb{K}^n$ (im Falle $\mathbb{K} = \mathbb{R}$ heißt $|| \cdot ||_2$ euklidische Norm). Die Cauchy-Schwarz'sche Ungleichung lautet hier

$$|\sum_{j=1}^{n} x_j \overline{y}_j| \leq \sqrt{\sum_{j=1}^{n} |x_j|^2} \sqrt{\sum_{j=1}^{n} |y_j|^2} \qquad (x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{K}) \; .$$

Weitere wichtige Normen auf $\mathbb{K}^n$ sind

$$||x||_\infty := \max_{j=1,\dots,n} |x_j| \qquad (x = (x_1, \dots, x_n)^T \in \mathbb{K}^n)$$

und

$$||x||_1 := \sum_{j=1}^{n} |x_j| \qquad (x = (x_1, \dots, x_n)^T \in \mathbb{K}^n)$$

(das $||\cdot||_\infty, ||\cdot||_1$ Normen auf $\mathbb{K}^n$ sind, ergibt sich leicht aus Eigenschaften des Betrages $|\cdot|$ auf $\mathbb{K}$).
2. Es sei $V = C[a, b]$ mit $< \cdot, \cdot >$ aus B. 13.3.2. Dann ist nach S. 13.5

$$||f||_2 := ||f|| := \sqrt{\int_a^b |f(t)|^2 dt} \qquad (f \in C[a, b])$$

eine Norm auf $C[a, b]$. Weitere wichtige Normen auf $C[a, b]$ sind

$$||f||_\infty := \sup_{t \in [a,b]} |f(t)| \qquad (f \in C[a, b])$$

und

$$||f||_1 := \int_a^b |f(t)| dt \qquad (f \in C[a, b])$$

(das $||f||_\infty, ||f||_1$ Normen auf $C[a, b]$ sind folgt aus Standardergebnissen der Analysis)

Bemerkung 13.7 Aus S. 13.5.1. ergeben sich leicht folgende Indentitäten, die in jedem unitären Raum für die induzierte Norm gelten:
1. (*Parallelogrammidentität*)

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

2. (*Polarisierungsidentität*)

$$< x, y >= \begin{cases} \frac{1}{4}(||x+y||^2 - ||x-y||^2), & \text{falls} \quad \mathbb{K} = \mathbb{R} \\ \\ \frac{1}{4}(||x+y||^2 - ||x-y||^2 + i||x+iy||^2 - i||x-iy||^2), & \text{falls} \quad \mathbb{K} = \mathbb{C} \end{cases} .$$

(Denn: Nach S. 13.5.1 gilt (beachte: $||y|| = ||-y||$)

$$\begin{aligned} ||x+y||^2 + ||x-y||^2 &= 2||x||^2 + 2||y||^2 + < x, y > + < y, x > + < x, -y > + < -y, x > \\ &= 2||x||^2 + 2||y||^2 \end{aligned}$$

und im Falle $\mathbb{K} = \mathbb{R}$

$$\begin{aligned} ||x+y||^2 - ||x-y||^2 &= < x, y > + < y, x > - < x, -y > - < -y, x >= \\ &= 2 < x, y > +2 < y, x >= 4 < x, y > \ . \end{aligned}$$

Im Falle $\mathbb{K} = \mathbb{C}$ gilt wieder mit S. 13.5.1

$$\begin{aligned} & ||x+y||^2 - ||x-y||^2 + i||x+iy||^2 - i||x-iy||^2 = \\ = \ & 2 < x, y > +2 < y, x > +2i < x, iy > +2i < iy, x > \\ = \ & 2 < x, y > +2 < y, x > +2 < x, y > -2 < y, x >= 4 < x, y >) \end{aligned}$$

Hieraus folgt, dass z. B. auf $\mathbb{K}^n$ die Normen $||\cdot||_\infty$ und $||\cdot||_1$ (für $n \geq 2$) nicht von einem Skalarprodukt herrühren.
(Denn:

$$2(||e_1||_\infty^2 + ||e_2||_\infty^2) = 4 \neq 2 = ||e_1 + e_2||_\infty^2 + ||e_1 - e_2||_\infty^2$$

und

$$2(||e_1||_1^2 + ||e_2||_1^2) = 4 \neq 8 = ||e_1 + e_2||_1^2 + ||e_1 - e_2||_1^2) \ .$$

14 Orthogonalität

Zu Beginn des Abschnittes 13 hatten wir bemerkt, dass der "Winkel" zwischen zwei Vektoren $x, y \in \mathbb{R}^2$ gegeben ist durch

$$\alpha = \arccos\left(\frac{x^T y}{||x||\ ||y||}\right) \qquad (\in [0, \pi])$$

(man beachte: $\frac{x^T y}{||x||\ ||y||} \in [-1, 1]$ nach der Cauchy-Schwarz'schen Ungleichung). Insbesondere gilt $\alpha = \pi/2\,(\hat{=}\,90$ Grad) für $x^T y = 0$, d. h. ist $x^T y = 0$, so stehen x und y senkrecht aufeinander.

Definition 14.1 Es sei $V = (V, < \cdot, \cdot >)$ ein unitärer Raum. Ferner seien $x, y \in V$ und $M, N \subset V$.
1. x, y heißen *orthogonal* (oder *senkrecht zueinander*), falls $< x, y >= 0$ ist. Wir schreiben dann $x \perp y$.
2. x heißt *orthogonal zu* M (oder *senkrecht stehend auf* M), falls $x \perp y$ für alle $y \in M$. Wir schreiben dann $x \perp M$.
3. M und N heißen *orthogonal* (oder *senkrecht zueinander*), falls $x \perp y$ für alle $x \in M, y \in N$. Wir schreiben dann $M \perp N$.
4. Die Menge

$$M^{\perp} := \{x \in V : x \perp y \quad \text{für alle } y \in M\}$$

heißt *orthogonales Komplement* von M.

Beispiel 14.2 1. Ist $V = \mathbb{K}^n$ mit kanonischen Skalarprodukt, so sind die kanonischen Basisvektoren $e_1, \ldots, e_n$ paarweise orthogonal (denn $< e_j, e_k >^2 = e_j^T e_k = \delta_{jk}$ $(j, k = 1, \ldots, n)$).
2. Es sei $V = (C[-\pi, \pi], \mathbb{C})$. Dann sind die Funktionen $e^{ik\,\cdot}$ $(k \in \mathbb{Z})$, d. h.

$$t \mapsto e^{ikt} \qquad (t \in [-\pi, \pi],\ k \in \mathbb{Z})$$

paarweise orthogonal.
(Denn: Die Funktion $e^{ik\,\cdot}$ ist 2π-periodisch. Also gilt für $k \neq j$ ($\rightarrow$ Analysis)

$$\begin{aligned} < e^{ij\,\cdot}, e^{ik\,\cdot} > &= \int_{-\pi}^{\pi} e^{ijt}\overline{e^{ikt}}dt = \int_{-\pi}^{\pi} e^{ijt}e^{-ikt}dt \\ &= \int_{-\pi}^{\pi} e^{i(j-k)t}dt = \frac{1}{i(j-k)} e^{i(j-k)t}|_{t=-\pi}^{\pi} = 0\ .) \end{aligned}$$

Enstprechend sind in $V = (C[-\pi, \pi], \mathbb{R})$ die Funktionen

$$\begin{matrix} \cos(k\,\cdot) \\ \sin(k\,\cdot) \end{matrix} \qquad (t \in [-\pi, \pi], k \in \mathbb{N}_0)$$

paarweise orthogonal (d. h. $\cos(k\,\cdot) \perp \cos(j\,\cdot), \sin(k\,\cdot) \perp \sin(j\,\cdot)$ für $k \neq j$ und $\sin(k\,\cdot) \perp \cos(k\,\cdot)$ für alle $k \in \mathbb{N}_0$).
3. Es sei $V = \mathbb{R}^n$ mit kanonischem Skalarprodukt. Für ein $a = (a_1, \ldots, a_n)^T \in \mathbb{R}^n \setminus \{0\}$ sei

$$M := M_a := \{\lambda a : \lambda \in \mathbb{R}\} =< a > \ .$$

Dann ist

$$\begin{aligned} M^\perp &= \{x \in \mathbb{R}^n : \lambda a \perp x \quad \text{für alle } \lambda \in \mathbb{R}^n\} = \\ &= \{x = (x_1, \dots, x_n)^T : \lambda \sum_{j=1}^n a_j x_j = 0 \text{ für alle } \lambda \in \mathbb{R}\} \\ &= \{x = (x_1, \dots, x_n)^T : \sum_{j=1}^n a_j x_j = 0\} = \{a\}^\perp =: U_a \end{aligned}$$

mit U_a aus B. 4.3.2. Insbesondere ist mit obigen Bezeichnungen

$$a \perp U_a \, .$$

Der folgende Satz ist eine Verallgemeinerung des klassischen "Pythagoras".

Satz 14.3 (Pythagoras) *Es sei V ein unitärer Raum, und es seien $x, y \in V$. Sind x, y orthogonal, so gilt*

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

(wobei $|| \cdot || =< \cdot, \cdot >^{1/2}$ die induzierte Norm).

Beweis.
Der Beweis ergibt sich sofort aus S. 13.5.1 und $< x, y >=< y, x >= 0$. □

Bemerkung 14.4 Aus S. 13.5.1 folgt auch, dass

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

genau dann gilt, wenn $< x, y > + < y, x >=< x, y > + \overline{< x, y >} = 2 \operatorname{Re} < x, y >= 0$ ist. Also ist im Falle $\mathbb{K} = \mathbb{R}$ die Bedingung $x \perp y$ auch notwendig.

Wir stellen einige Eigenschaften des orthogonalen Komplements zusammen.

Satz 14.5 *Es seien $M, N \subset V$, wobei V ein unitärer Raum ist. Dann gilt*
1. $M^\perp$ ist ein Unterraum von V.
2. $M \subset M^{\perp\perp} (:= (M^\perp)^\perp)$.
3. Aus $M \subset N$ folgt $N^\perp \subset M^\perp$.
4. $M^\perp =< M >^\perp$.
5. $M \cap M^\perp \subset \{0\}$.
6. Ist M ein Unterraum, so ist die Summe $M + M^\perp$ direkt, d. h. $M + M^\perp = M \oplus M^\perp$.

Beweis.
1. Es seien $x_1, x_2 \in M^\perp$ und $\lambda_1, \lambda_2 \in \mathbb{K}$. Dann gilt für alle $y \in M$

$$< \lambda_1 x_1 + \lambda_2 x_2, y >= \lambda_1 < x_1, y > + \lambda_2 < x_2, y >= 0 \ ,$$

d. h. $\lambda_1 x_1 + \lambda_2 x_2 \in M^\perp$. Also ist $M^\perp$ ein Unterraum nach S. 4.2.
2. Ist $y \in M$, so gilt für alle $x \in M^\perp$ nach Definition $x \perp y$. Also ist wieder nach Definition $y \in (M^\perp)^\perp$.
3. Es sei $M \subset N$. Ist $x \in N^\perp$, so gilt $x \perp y$ für alle $y \in N$, also auch $x \perp y$ für alle $y \in M$, und damit ist $x \in M^\perp$.
4. Aus 3. folgt $< M >^\perp \subset M^\perp$. Ist umgekehrt $x \in M^\perp$, so gilt $x \perp y$ für alle $y \in M$. Ist $z \in < M >$, so existieren $\lambda_1, \ldots, \lambda_n \in \mathbb{K}, y_1, \ldots, y_n \in M$ mit $z = \sum_{j=1}^{n} \lambda_j y_j$. Also gilt

$$< z, x >= \sum_{j=1}^{n} \lambda_j < y_j, x >= 0 \ ,$$

d. h. $x \perp z$. Da $z \in < M >$ beliebig war, gilt $x \in < M >^\perp$.
5. Ist $x \in M \cap M^\perp$, so gilt $x \perp y$ für alle $y \in M$, also insbesondere für $y = x$. Folglich ist $< x, x >= 0$ und damit $x = 0$ nach (S.1).
6. Folgt sofort aus 5. □

Definition 14.6 Es sei $V = (V, < \cdot, \cdot >)$ ein unitärer Raum, und es sei $I \neq \emptyset$ sowie $(x_\alpha)_{\alpha \in I} (\in V^I)$ eine Familie von Vektoren aus V.
1. $(x_\alpha)_{\alpha \in I}$ heißt *Orthogonalsystem* (OGS), falls $x_\alpha \perp x_\beta$ für alle $\alpha, \beta \in I, \alpha \neq \beta$.
2. $(x_\alpha)_{\alpha \in I}$ heißt *Orthnormalsystem* (ONS), falls $(x_\alpha)_{\alpha \in I}$ ein Orthogonalsystem ist mit $||x_\alpha|| = 1$ für alle $\alpha \in I$.
3. $(x_\alpha)_{\alpha \in I}$ heißt *Orthonormalbasis* (ONB), falls $(x_\alpha)_{\alpha \in I}$ ein Orthonormalsystem und eine Basis ist.

Bemerkung 14.7 1. Ist $(x_\alpha)_{\alpha \in I}$ ein OGS mit $x_\alpha \neq 0$ für alle $\alpha \in I$, so ist $(x_\alpha / ||x||_\alpha)_{\alpha \in I}$ ein ONS (wobei $x/\lambda := \lambda^{-1} \cdot x$ für $x \in V, \lambda \in \mathbb{K}$).
2. $(x_\alpha)_{\alpha \in I}$ ist genau dann ein ONS, wenn $< x_\alpha, x_\beta >= \delta_{\alpha\beta} := \begin{cases} 1, & \alpha = \beta \\ 0, & \alpha \neq \beta \end{cases}$ gilt.
3. Ist $(x_\alpha)_{\alpha \in I}$ ein OGS, so gilt für alle endlichen $J \subset I$:

$$|| \sum_{j \in J} x_j ||^2 = \sum_{j \in J} ||x_j||^2 \tag{5}$$

(allgemeiner Satz von Pythagoras).

Beispiel 14.8 (vgl. B. 14.2)
1. Ist $V = \mathbb{K}^n$, so ist $(e_1, \dots, e_n)$ eine ONB in V.
2. Ist $V = (C[-\pi, \pi], \mathbb{C})$ so ist $(e^{ik\cdot})_{k\in\mathbb{Z}}$ ein OGS in V. In $V = (C[-\pi, \pi], \mathbb{R})$ ist $(\cos(\cdot), \sin(\cdot), \cos(2\cdot), \sin(2\cdot), \dots)$ ein OGS.

Satz 14.9 *Es sei V ein unitärer Raum, und es sei $(x_\alpha)_{\alpha\in I}$ ein OGS in V. Ist $x_\alpha \neq 0$ für alle $\alpha \in I$, so ist $(x_\alpha)_{\alpha\in I}$ linear unabhängig.*

Beweis.
Es sei $J \subset I$ endlich und es seien $\lambda_j \in \mathbb{K}$ $(j \in J)$ mit

$$0 = \sum_{j\in J} \lambda_j x_j \ .$$

Nach dem Satz von Pythagoras (5) ist

$$0 = ||\sum_{j\in J} \lambda_j x_j||^2 = \sum_{j\in J} |\lambda_j|^2 ||x_j||^2 \ .$$

Aus $||x_j||^2 > 0$ $(j \in J)$ folgt $|\lambda_j|^2 = 0$, d. h. $\lambda_j = 0$ für alle $j \in J$. □

Ist $(x_1, \dots, x_n)$ eine Basis eines (beliebigen) linearen Raumes V, so hat bekanntlich jedes $x \in V$ eine eindeutige Darstellung $x = \sum\limits_{j=1}^{n} \lambda_j x_j$. Ist V unitär und $(x_1, \dots, x_n)$ eine ONB, so lassen sich die Koeffizienten λ_j dabei "explizit" angeben.

Satz 14.10 *Es sei V ein unitärer Raum, und es sei $(x_1, \dots, x_n)$ eine ONB von V. Dann gilt für jedes $x \in V$*

$$x = \sum_{j=1}^{n} < x, x_j > x_j \tag{6}$$

und die sog. Parseval'sche Gleichung:

$$||x||^2 = \sum_{j=1}^{n} | < x, x_j > |^2 \ . \tag{7}$$

Beweis.
Da $(x_1, \dots, x_n)$ eine Basis von V ist, existieren $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ mit

$$x = \sum_{j=1}^{n} \lambda_j x_j \ .$$

Also folgt für $k = 1, \dots, n$

$$\begin{aligned} < x, x_k > &= < \sum_{j=1}^{n} \lambda_j x_j, x_k >= \sum_{j=1}^{n} \lambda_j < x_j, x_k >= \\ &= \sum_{j=1}^{n} \lambda_j \delta_{jk} = \lambda_k \ . \end{aligned}$$

Weiter gilt nach dem Satz Pythagoras (5)

$$||x||^2 = || \sum_{j=1}^{n} < x, x_j > x_j||^2 = \sum_{j=1}^{n} | < x, x_j > |^2 ||x_j||^2 = \sum_{j=1}^{n} | < x, x_j > |^2 \ .$$

□

In der Darstellungsformel (6) liegt einer der wesentlichen Vorteile einer ONB gegenüber einer beliebigen Basis. Der folgende Satz gibt ein Verfahren an, nach dem man aus einer Basis eines unitären Raumes eine ONB konstruieren kann.

Satz 14.11 (Schmidt'sches Orthogonalisierungsverfahren) *Es sei V ein (nicht notwendig endlich-dimensionaler) unitärer Raum. Sind $x_1, \dots, x_n$ linear unabhängig in V, so ist durch*

$$y_1 := x_1, \qquad y_k := x_k - \sum_{\nu=1}^{k-1} \frac{< x_k, y_\nu >}{||y_\nu||^2} y_\nu \tag{8}$$

für $k = 2, \dots, n$ ein OGS $(y_1, \dots, y_n)$ in V definiert mit

$$span(y_1, \dots, y_k) = span(x_1, \dots, x_k) \quad für \quad k = 1, \dots, n \ .$$

Beweis.
Wir zeigen per Induktion nach $m(\leq n)$: Durch (8) ist ein OGS $(y_1, \dots, y_m)$ definiert mit $\text{span}(y_1, \dots, y_m) = \text{span}(x_1, \dots, x_m)$
$\underline{m = 1}$: Für $y_1 := x_1$ gilt: (y_1) ist ein OGS mit $\text{span}(x_1) = \text{span}(y_1)$
$\underline{m \to m+1}$: Nach Induktionsvoraussetzung ist durch (8) ein OGS $(y_1, \dots, y_m)$ definiert mit $\text{span}(y_1, \dots, y_m) = \text{span}(x_1, \dots, x_m)$. Insbesondere ist $y_k \neq 0$ für $k = 1, \dots, m$ (da $\dim \text{span}(y_1, \dots, y_m) = \dim \text{span}(x_1, \dots, x_m) = m$). Damit können wir

$$y_{m+1} := x_{m+1} - \sum_{\nu=1}^{m} \frac{< x_{m+1}, y_\nu >}{||y_\nu||^2} y_\nu$$

setzen. Es gilt dann für $k = 1, \dots, m$

$$\begin{aligned} < y_{m+1}, y_k > &= < x_{m+1}, y_k > - \sum_{\nu=1}^{m} \frac{< x_{m+1}, y_\nu >}{||y_\nu||^2} < y_\nu, y_k > \\ &= < x_{m+1}, y_k > - < x_{m+1}, y_k >= 0 \ , \end{aligned}$$

also ist, da $(y_1, \dots, y_m)$ ein OGS ist, auch $(y_1, \dots, y_{m+1})$ ein OGS. Aus der Definition von y_{m+1} folgt

$$x_{m+1} \in \text{span}(y_1, \dots, y_{m+1}) .$$

Da $\text{span}(y_1, \dots, y_m) = \text{span}(x_1, \dots, x_m)$ gilt ist deshalb

$$\text{span}(x_1, \dots, x_{m+1}) \subset \text{span}(y_1, \dots, y_{m+1}) .$$

Andererseits ist nach Definition $y_{m+1} \in \text{span}(x_{m+1}, y_1, \dots, y_m) = \text{span}(x_1, \dots, x_{m+1})$, also auch

$$\text{span}(y_1, \dots, y_{m+1}) \subset \text{span}(x_1, \dots, x_{m+1}) .$$

□

Folgerung 14.12 Unter den Voraussetzungen von S. 14.11 gilt: Ist $u_j := \frac{y_j}{||y_j||}$ für $j = 1, \dots, n$, so ist $(u_1, \dots, u_n)$ ein ONS in V mit $\text{span}(x_1, \dots, x_k) = \text{span}(u_1, \dots, u_k)$ für $k = 1, \dots, n$. Weiterhin gilt: Ist V n-dimensional, so ist $(u_1, \dots, u_n)$ eine ONB von V. Insbesondere hat also jeder endlich-dimensionale unitäre Raum $V \neq \{0\}$ eine ONB.

Denn: Nach S. 14.11 ist $(y_1, \dots, y_n)$ ein OGS mit $y_k \neq 0$ für $k = 1, \dots, n$, also folgt die erste Behauptung aus B. 14.7.1. Nach S.14.11 gilt dabei $\text{span}(x_1, \dots, x_k) = \text{span}(u_1, \dots, u_k)$ für $k = 1, \dots, n$. Ist V endlich-dimensional mit $\dim(V) = n$, so ist $(x_1, \dots, x_n)$ eine Basis von V, also ist auch $(u_1, \dots, u_n)$ eine Basis von V (beachte: $\text{span}(u_1, \dots u_n) = \text{span}(x_1, \dots, x_n)$).

Beispiel 14.13 Es sei $V = \mathbb{R}^2$, und es seien $x_1 = \binom{4}{2}, x_2 = \binom{1}{3}$. Dann ist nach S. 14.11 durch

$$y_1 := x_1 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

und

$$\begin{aligned} y_2 &:= x_2 - \frac{< x_2, y_1 >}{||y_1||^2} y_1 \\ &= \begin{pmatrix} 1 \\ 3 \end{pmatrix} - \frac{(1,3) \cdot \binom{4}{2}}{||\binom{4}{2}||^2} \begin{pmatrix} 4 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 3 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \end{aligned}$$

ein OGS in $\mathbb{R}^2$ und durch

$$\left(\frac{y_1}{||y_1||} , \frac{y_2}{||y||_2} \right) = \left(\frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix} , \frac{1}{\sqrt{5}} \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right)$$

eine ONB vom $\mathbb{R}^2$ gegeben.
Mit $U :=< x_1 >=< y_1 >$ gilt hier weiterhin

$$U^\perp = \{y_1\}^\perp =< y_2 >$$

und $\mathbb{R}^2 =< y_1 > + < y_2 >= U \oplus U^\perp$.
Allgemeiner gilt:

Satz 14.14 *Es sei V ein unitärer Raum, und es sei $U \subset V$ ein endlich-dimensionaler Unterraum. Dann gilt*
1. $V = U \oplus U^\perp$,
2. $U^{\perp\perp} = U$.

Beweis.
1. Nach S. 14.5.6 genügt es, zu zeigen: $V \subset U + U^\perp$. O. E. sei $U \neq \{0\}$, also $n := \dim(U) > 0$. Nach F. 14.12 existiert eine ONB $(u_1, \ldots, u_n)$ von U. Es sei $x \in V$ gegeben. Wir setzen

$$u := \sum_{j=1}^{n} < x, u_j > u_j \in U \ .$$

Es genügt, zu zeigen: $x - u \in U^\perp$ (dann ist $x = u + (x - u) \in U + U^\perp$).
Es gilt für $k = 1, \ldots, n$

$$\begin{aligned} < x - u, u_k > &= < x, u_k > - < u, u_k >= \\ &= < x, u_k > - \sum_{j=1}^{n} < x, u_j >< u_j, u_k > \\ &= < x, u_k > - < x, u_k >= 0 \end{aligned}$$

also $x - u \in \{u_1, \ldots, u_n\}^\perp = \text{span}(u_1, \ldots, u_n)^\perp = U^\perp$.
2. Nach S. 14.5.2 ist $U \subset U^{\perp\perp}$.
Es sei $x \in U^{\perp\perp}$. Nach 1. existieren $u \in U, v \in U^\perp$ mit

$$x = u + v \ .$$

Dann gilt

$$v = x - u \in U^{\perp\perp} + U = U^{\perp\perp} \ .$$

Nach S.14.5.5 ist $v = 0$ und damit $x = u \in U$. Also ist auch $U^{\perp\perp} \subset U$. □

Bemerkung 14.15 Die Aussage des S. 14.14 wird i. a. falsch, wenn man auf die Voraussetzung "$\dim(U) < \infty$" verzichtet. Mit Hilfe des Weierstraß'schen Approximationssatzes (→ Analysis) kann man zeigen, daß etwa für $V = C[a, b]$ (mit dem Skalarprodukt aus B. 13.3.2) und den Unterraum

$$U := \{P_{|[a,b]} : P \text{ Polynom}\}$$

gilt: $U^\perp = \{0\}$, also $U \oplus U^\perp = U \neq V$.

15 Projektionen

Im vorigen Abschnitt haben wir gesehen, daß für einen unitären Raum V und einen endlich- dimensionalen Unterraum U stets $V = U \oplus U^{\perp}$ gilt. In S. 5.20 hatten wir gezeigt, daß für einen endlich-dimensionalen linearen Raum und einen Unterraum U stets ein Unterraum W existiert mit $V = U \oplus W$. Wir betrachten zunächst wieder allgemeine lineare Räume.

Definition 15.1 Es sei V ein linearer Raum über K, und es seien U, W Unterräume von V so, dass $V = U \oplus W$ gilt. Dann heißt $P := P_{U,W} : V \to V$, definiert durch

$$P(v) := P_{U,W}(v) := u \qquad (v \in V) \ ,$$

wobei $v = u + w$ mit $u \in U, w \in W$, *Projektion auf U (längs W).*

Bemerkung 15.2 Ist $P = P_{U,W}$ wie in D. 15.1, so gilt, wie man leicht sieht:
1. $P \in L(V)$
2. $\text{Bild}(P) = U$,
3. $\text{Kern}(P) = W$,
4. $P^2 (:= P \circ P) = P$.

Umgekehrt kann man zeigen ([Ü]):
Ist $T \in L(V)$ mit $T = T^2$, so gilt $T = P_{\text{Bild}(T),\text{Kern}(T)}$, d. h. T ist Projektion auf $\text{Bild}(T)$ längs $\text{Kern}(T)$.

Beispiel 15.3 Es sei $V = \mathbb{R}^2$. Dann ist

$$V = < \begin{pmatrix} 1 \\ 0 \end{pmatrix} > \oplus < \begin{pmatrix} 1 \\ 1 \end{pmatrix} > = U \oplus W \ .$$

Hier ist $P = P_{U,W}$ gegeben durch

$$P \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ .$$

Definition 15.4 Es sei V ein unitärer Raum, und es sei $U \subset V$ ein Unterraum mit $V = U \oplus U^{\perp}$. Dann heißt $P_U := P_{U,U^{\perp}}$ *orthogonale Projektion (von V) auf U*:

Der folgende Satz verdeutlicht die Relevanz orthogonaler Projektionen für die Approximation in unitären Räumen.

Satz 15.5 (Projktionssatz) *Es sei V ein unitärer Raum, und es sei $U \subset V$ ein Unterraum mit $V = U \oplus U^{\perp}$. Dann gilt für alle $x \in V$*

$$||x - P_U x|| = \text{dist}(x, U) := \inf_{y \in U} ||x - y|| \,,$$

und für alle $y \in U, y \neq P_U x$, ist $||x - y|| > \text{dist}(x, U)$ (d. h. $P_U x$ ist die eindeutig bestimmte Lösung des Problems "Minimiere den Abstand $||x - y||$ über alle $y \in U$").

Beweis.
Es sei $x \in V$. Dann gilt

$$x = u + w = P_U(x) + w$$

mit $u \in U, v \in U^{\perp}$, also $x - P_U x = w \in U^{\perp}$. Für alle $y \in U$ gilt damit nach dem Satz von Pythagoras (beachte $P_U x - y \in U$)

$$||x - y||^2 = ||x - P_U x + P_U x - y||^2 = ||x - P_U x||^2 + ||P_U x - y||^2 \geq ||x - P_U x||^2 \,,$$

also $||x - P_U x|| = \inf_{y \in U} ||x - y||$ und für $y \neq P_U x$ gilt $||P_U x - y||^2 > 0$, also $||x - y|| > ||x - P_U x||$. □

In Verallgemeinerung von S. 14.10 gilt

Satz 15.6 *Es sei V ein unitärer Raum, und es sei $U \subset V$ ein (endlich-dimensionaler) Unterraum mit ONB $(u_1, \ldots, u_m)$. Dann gilt für alle $x \in V$*

$$P_U x = \sum_{j=1}^{m} < x, u_j > u_j$$

und es gilt die sog. Bessel'sche Ungleichung

$$\sum_{j=1}^{m} | < x, u_j > |^2 \leq ||x||^2 \,.$$

Beweis.
Es sei $x \in V$. Aus dem Beweis zu S. 14.14.1 ergibt sich, dass für

$$u := \sum_{j=1}^{m} < x, u_j > u_j \in U$$

gilt

$$x - u \in U^{\perp}.$$

Also ist $x = u + (x - u)$ mit $u \in U$ und $x - u \in U^\perp$. Aus der Eindeutigkeit dieser Darstellung folgt $u = P_U x$.
Weiter gilt

$$< x, u_j >=< u, u_j > + < x - u, u_j >=< u, u_j > \qquad (j = 1, \dots, m)$$

und damit nach der Parseval'schen Gleichung (7) und dem Satz von Pythagoras (5)

$$\sum_{j=1}^{m} | < x, u_j > |^2 = \sum_{j=1}^{m} | < u, u_j > |^2 = ||u||^2 \leq ||u||^2 + ||x - u||^2 = ||x||^2 \ .$$

□

Hiermit erhält man eine kompakte Darstellung im Schmidt'schen Orthogonalisierungsverfahren:

Folgerung 15.7 Mit den Voraussetzungen und Bezeichnungen von S. 14.11 gilt

$$y_k = x_k - P_{U_{k-1}}(x_k) \qquad (k = 2, \dots, n) \ ,$$

wobei $U_{k-1} := \text{span}(x_1, \dots, x_{k-1}) = \text{span}(y_1, \dots, y_{k-1})$.

Denn: Nach S. 14.11 ist mit $u_\nu := y_\nu / ||y_\nu||$

$$y_k = x_k - \sum_{\nu=1}^{k-1} < x_k, y_\nu / ||y_\nu|| > \cdot (y_\nu / ||y_\nu||) = x_k - \sum_{\nu=1}^{k-1} < x_k, u_\nu > u_\nu \ .$$

Da $(u_1, \dots, u_{k-1})$ eine ONB vom U_{k-1} ist, folgt die Darstellung aus S. 15.6.

Beispiel 15.8 Es sei $V = (C[-\pi, \pi], \mathbb{C})$ mit dem Skalarprodukt aus B. 13.3.2.
Für $n \in \mathbb{N}$ sei

$$U_n := \text{span}(e^{ik\,\cdot}, k = -n, \dots, 0, \dots, n) \ .$$

Dann ist $\left(\frac{1}{\sqrt{2\pi}} e^{ik\,\cdot}\right)_{k=-n,\dots,n}$ eine ONB von U_n (vgl. B. 14.2). Für jedes $f \in (C[-\pi, \pi]), \mathbb{C})$ ist

$$\begin{aligned} S_n(t) := P_{U_n}(f)(t) &= \sum_{k=-n}^{n} < f, \frac{e^{ik\,\cdot}}{\sqrt{2\pi}} > \frac{e^{ikt}}{\sqrt{2\pi}} \\ &= \sum_{k=-n}^{n} a_k e^{ikt} \qquad (t \in [-\pi, \pi]) \end{aligned}$$

mit

$$a_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) e^{\overline{iks}} ds = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) e^{-iks} ds$$

(S_n heißt n-te Teilsumme der Fourier-Reihe $\sum_{k=-\infty}^{\infty} a_k e^{ik\cdot}$ von f). Es gilt nach S. 15.5 und 15.6

$$||S_n - f||_2 \leq ||T - f||_2 \qquad (T \in U_n)$$

und

$$\begin{aligned} \int_{-\pi}^{\pi} |f(t)|^2 dt &= ||f||^2 \geq ||S_n||^2 = \\ &= \sum_{k=-n}^{n} | < f, \frac{e^{ik\cdot}}{\sqrt{2\pi}} > |^2 = 2\pi \sum_{k=-n}^{n} |a_k|^2 \ . \end{aligned}$$

Bemerkung 15.9 Wir betrachten nochmals unter den Voraussetzungen von S. 15.5 das Optimierungsproblem.

$$\text{minimiere} \qquad ||x - y|| \qquad \text{über} \quad y \in U \ .$$

Nach S. 15.5 ist $x^* := P_U x$ die eindeutig bestimmte Lösung. In S. 15.6 haben wir gesehen, daß x^* sehr einfach zu bestimmen ist, wenn eine ONB von U gegeben ist. Was lässt sich über die Berechnung von x^* sagen, wenn lediglich irgendeine Basis $(u_1, \ldots, u_m)$ von U gegeben ist?
Es gilt (siehe Beweis zu S. 15.5 und S.15.6): x^* ist charakterisiert durch

$$x - x^* = x - P_U x \perp U \ ,$$

also

$$x - x^* \perp u_j \qquad (j = 1, \ldots, m)$$

(sog. *Normalengleichungen*) bzw.

$$< x, u_j > = < x^*, u_j > \qquad (j = 1, \ldots, m) \ .$$

Wir suchen $\alpha_1, \ldots, \alpha_m \in \mathbb{K}$ so, daß $x^* = \sum_{k=1}^{m} \alpha_k u_k$, d. h.

$$< x, u_j > = \sum_{k=1}^{m} \alpha_k < u_k, u_j > \qquad (j = 1, \ldots, m) \ .$$

Mit der Matrix

$$B = (b_{jk}) = (< u_k, u_j >) \in \mathbb{K}^{m \times m}$$

(B^T heißt *Gram'sche Matrix* (von $(u_1, \ldots, u_m)$) ist also $(\alpha_1, \ldots, \alpha_m)$ Lösung des LGS

$$B \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} < x, u_1 > \\ \vdots \\ < x, u_m > \end{pmatrix} \ .$$

Aus obigen Überlegungen folgt, dass das System genau eine Lösung hat, also ist stets B invertierbar.

Beispiel 15.10 Es sei $V = \mathbb{R}^4$ (mit kanonischem Skalarprodukt), und es seien $x = (3, -3, 0, -3)^T$ sowie $U = \text{span}(u_1, u_2)$, wobei $u_1 := (1, 0, -1, -1)^T$, $u_2 := (0, 2, 1, 2)^T$. Dann gilt

$$B = \begin{pmatrix} < u_1, u_1 > & < u_2, u_1 > \\ < u_1, u_2 > & < u_2, u_2 > \end{pmatrix} = \begin{pmatrix} 3 & -3 \\ -3 & 9 \end{pmatrix} .$$

Also ist das LGS

$$\begin{pmatrix} 3 & -3 \\ -3 & 9 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} < x, u_1 > \\ < x, u_2 > \end{pmatrix} = \begin{pmatrix} 6 \\ -12 \end{pmatrix}$$

zu lösen. Es gilt

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} ,$$

also ist

$$P_U x = x^* = 1 \cdot u_1 + (-1) u_2 = \begin{pmatrix} 1 \\ -2 \\ -2 \\ -3 \end{pmatrix} .$$

Bemerkung 15.11 Allgemein gilt für $V = \mathbb{K}^n$ und linear unabhängige $u_1, \ldots, u_m \in \mathbb{K}^n$ mit $U := \text{span}(u_1, \ldots, u_m)$ und $A := (u_1, \ldots, u_m)$ (d. h. die u_j sind die Spalten von A)

$$\overline{A}^T A = (\overline{u_j}^T u_k)_{j,k} = (< u_k, u_j >)_{j,k} ,$$

wobei $\overline{B} := (\overline{b_{jk}})$ falls $B = (b_{jk})$, und

$$\overline{A}^T x = \begin{pmatrix} \overline{u_1}^T x \\ \vdots \\ \overline{u_n}^T x \end{pmatrix} = \begin{pmatrix} < x, u_1 > \\ \vdots \\ < x, u_m > \end{pmatrix}$$

d. h. die Normalengleichungen haben hier die Form

$$\boxed{\overline{A}^T A \, \alpha = \overline{A}^T x}$$

($x^* = \sum_{k=1}^{m} \alpha_k u_k$ minimiert $||x - y||^2 = \sum_{j=1}^{n} |x_j - y_j|^2$ über alle $y \in U$; sog. "kleinste-Quadrate Approximation" an x).

16 Selbstadjungierte und normale Operatoren

Wir betrachten im folgenden spezielle Klassen linearer Abbildungen auf unitären Räumen. Soweit möglich wollen wir wieder die Fälle $\mathbb{K} = \mathbb{R}$ und $\mathbb{K} = \mathbb{C}$ einheitlich behandeln. Vorbereitend zeigen wir

Satz 16.1 *Es sei $(V, <\cdot,\cdot>)$ ein endlich-dimensionaler unitärer Raum. Ist $\varphi \in V^* = L(V, \mathbb{K})$, so existiert genau ein $y \in V$ mit*

$$\varphi = <\cdot, y>$$

(d. h. $\varphi(x) = <x, y>$ für alle $x \in V$). Also ist $V^* = \{<\cdot, y>: y \in V\}$.

Beweis.
1. Existenz: Es sei $(x_1, \ldots, x_n)$ eine ONB von V (existiert nach F. 14.12). Dann gilt für $x \in V$ nach S. 14.10

$$x = \sum_{j=1}^{n} <x, x_j> x_j \ ,$$

also

$$\begin{aligned} \varphi(x) &= \varphi(\sum_{j=1}^{n} <x, x_j> x_j) = \sum_{j=1}^{n} <x, x_j> \varphi(x_j) = \\ &= <x, \sum_{j=1}^{n} \overline{\varphi(x_j)} x_j> \ , \end{aligned}$$

d. h. mit $y := \sum\limits_{j=1}^{n} \overline{\varphi(x_j)} x_j$ gilt $\varphi = <\cdot, y>$.
2. Eindeutigkeit: Es seien $y_1, y_2 \in V$ mit $<\cdot, y_1> = \varphi = <\cdot, y_2>$. Dann gilt für alle $x \in V$

$$0 = <x, y_1> - <x, y_2> = <x, y_1 - y_2> \ ,$$

also insbesondere

$$0 = <y_1 - y_2, y_1 - y_2> = ||y_1 - y_2||^2 \ .$$

Damit ist $y_1 = y_2$.
3. Nach (S.3) ist $<\cdot, y> \in V^*$ für alle $y \in V$, also gilt $V^* = \{<\cdot, y>: y \in V\}$ nach 1. □

Damit ist folgende Definition möglich.

Definition 16.2 Es seien $(V, <\cdot,\cdot>=<\cdot,\cdot>_V)$ und $(W, <\cdot,\cdot>=<\cdot,>_W)$ endlich-dimensionale unitäre Räume, und es sei $T \in L(V,W)$. Für $w \in W$ ist

$$<\cdot, w> \circ T \in V^* .$$

Also existiert nach S. 16.1 genau ein $u \in V$ mit $<\cdot, w> \circ T =<\cdot, u>$. Wir setzen

$$T^* w := u .$$

Damit gilt für alle $v \in V, w \in W$

$$\boxed{<Tv, w>=<v, T^* w> \ .}$$

Die Abbildung $T^* : W \to V$ heißt *Adjungierte* von T.

Satz 16.3 *Mit den Bezeichnungen aus D. 16.2 gilt*
1. $T^* \in L(W,V)$.
2. $T^{**} := (T^*)^* = T$.
3. Für $\mu \in \mathbb{K}$ *ist* $(\mu T)^* = \overline{\mu} T^*$.
4. Für $S \in L(V,W)$ *ist* $(S+T)^* = S^* + T^*$.
5. Ist $(U, <\cdot,\cdot>=<\cdot,\cdot>_U)$ *ein weiterer endlich-dimensionaler unitärer Raum und ist* $S \in L(U,V)$*, so ist*

$$(T \circ S)^* = S^* \circ T^* .$$

Beweis.
1. Es seien $w_1, w_2 \in W$ und $\lambda_1, \lambda_2 \in \mathbb{K}$. Dann gilt für alle $v \in V$

$$\begin{aligned} <Tv, \lambda_1 w_1 + \lambda_2 w_2> &= \overline{\lambda_1} <Tv, w_1> + \overline{\lambda_2} <Tv, w_2> \\ &= \overline{\lambda_1} <v, T^* w_1> + \overline{\lambda_2} <v, T^* w_2> \\ &= <v, \lambda_1 T^* w_1 + \lambda_2 T^* w_2> \ . \end{aligned}$$

Also erfüllt $u := \lambda_1 T^* w_1 + \lambda_2 T^* w_2$ die Bedingung aus D. 16.2 für $w = \lambda_1 w_1 + \lambda_2 w_2$ und damit ist $T^*(\lambda_1 w_1 + \lambda_2 w_2) = \lambda_1 T^* w_1 + \lambda_2 T^* w_2$.
2. Nach D. 16.2 (angewandt auf T und T^*) gilt für alle $v \in V, w \in W$

$$\begin{aligned} <Tv, w> &= <v, T^* w> = \overline{<T^* w, v>} = \\ &= \overline{<w, T^{**} v>} = <T^{**} v, w> , \end{aligned}$$

also insbesondere für $v \in V$

$$<Tv - T^{**} v, w> = 0$$

für $w = Tv - T^{**} v$. Damit ist $Tv - T^{**} v = 0$, d. h. $Tv = T^{**} v$.
(Beweis von 3., 4., 5. als [Ü]) □

Beispiel 16.4 Es sei $T : \mathbb{R}^3 \to \mathbb{R}^2$ mit

$$T(x_1, x_2, x_3) := (3x_2 + x_3, 2x_1) = \begin{pmatrix} 0 & 3 & 1 \\ 2 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} =: A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} .$$

Dann gilt für $(y_1, y_2) \in \mathbb{R}^2$, $(x_1, x_2, x_3) \in \mathbb{R}^3$

$$\begin{aligned} < (x_1, x_2, x_3), T^*(y_1, y_2) > &= < T(x_1, x_2, x_3), (y_1, y_2) > \\ &= 3x_2y_1 + x_3y_1 + 2x_1y_2 \\ &= < (x_1, x_2, x_3) \, , \, (2y_2, 3y_1, y_1) > \end{aligned}$$

d. h. nach Definition

$$T^*(y_1, y_2) = (2y_2, 3y_1, y_1) = \begin{pmatrix} 0 & 2 \\ 3 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = A^T \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} .$$

Allgemeiner gilt

Bemerkung und Definition 16.5 Es seien V, W, T wie in D. 16.2. Weiter seien $N = (v_1, \dots, v_n)$ bzw. $M = (w_1, \dots, w_m)$ ONB von V bzw. W. Ist $A = (a_{jk}) = \varphi_{M,N}(T)$ die Matrix von T bzgl. M, N, und setzt man

$$A^* := (\overline{a_{kj}})_{k=1,\dots,n, j=1,\dots,m}$$

($A^* = \overline{A}^T$ ist die *konjungierte Transponierte* von A), so ist $A^* = \varphi_{N,M}(T^*)$ d. h. A^* ist die Matrix von T^* bzgl. N, M.
Ist $\mathbb{K} = \mathbb{R}$, so ist dabei $A^* = A^T$.
(Denn: Es sei $B = (b_{kj})_{k=1,\dots,n, j=1,\dots,m} = \varphi_{N,M}(T^*)$. Dann gilt für $j = 1, \dots, m$; $k = 1, \dots, n$

$$\begin{aligned} a_{jk} &= \sum_{\nu=1}^{m} a_{\nu k} < w_\nu, w_j > = < \sum_{\nu=1}^{m} a_{\nu k} w_\nu, w_j > = < Tv_k, w_j > \\ &= < v_k, T^* w_j > = < v_k, \sum_{\nu=1}^{n} b_{\nu j} v_\nu > = \sum_{\nu=1}^{n} \overline{b_{\nu j}} < v_k, v_\nu > = \overline{b_{kj}} \end{aligned}$$

bzw. $b_{kj} = \overline{a_{jk}}$. Also ist $B = \overline{A}^T = A^*$
Ist insbesondere $V = W$ und $N = M$ eine ONB sowie $T \in L(V)$, so gilt $T = T^*$ genau dann, wenn $A = A^*$ ist.

Definition 16.6 1. Es sei $(V, <\cdot,\cdot>)$ ein endlich-dimensionaler unitärer Raum. Eine Abbildung $T \in L(V)$ heißt *selbstadjungiert* (oder *hermitesch*) falls $T = T^*$ gilt (d. h. $< Tv_1, v_2 >=< v_1, Tv_2 >$ für alle $v_1, v_2 \in V$).
Weiter heißt T *unitär diagonalisierbar*, falls eine ONB von V aus Eigenvektoren existiert.
2. Eine Matrix $A \in \mathbb{K}^{n\times n}$ heißt *hermitesch*, falls $A = A^*$ gilt (d. h. falls $x \mapsto Ax \in L(\mathbb{K}^n)$, wobei $\mathbb{K}^n$ mit dem kanonischen Skalarprodukt versehen ist, hermitesch ist; vgl. B./D. 16.5). Im Falle $\mathbb{K} = \mathbb{R}$ heißt A mit $A = A^* = A^T$ auch *symmetrisch*.

Satz 16.7 *Es sei $(V, <\cdot,\cdot>)$ ein endlich-dimensionaler unitärer Raum. Ist $T \in L(V)$ selbstadjungiert, so gilt*
1. T hat einen Eigenwert.
2. Alle Eigenwerte von T sind reell.
3. Sind λ_1, λ_2 Eigenwerte mit $\lambda_1 \neq \lambda_2$ und sind v_1, v_2 zugehörige Eigenvektoren, so sind v_1, v_2 orthogonal.

Beweis.
1. Es sei $M = (v_1, \dots, v_n)$ eine ONB von V (existiert nach F. 14.12). Ist $A = \varphi_M(T)$, so ist $A = A^*$ nach B./D. 16.5, d. h. A ist hermitesch.
2. Es sei $\lambda \in \mathbb{K}$ ein Eigenwert von T (falls existent). Ist $v \neq 0$ Eigenvektor zu λ, so gilt

$$\lambda < v, v >=< \lambda v, v >=< Tv, v >=< v, Tv >=< v, \lambda v >= \overline{\lambda} < v, v >$$

also $\lambda = \overline{\lambda}$ (da $< v, v >\neq 0$), d. h. $\lambda \in \mathbb{R}$.
3. Ist $\mathbb{K} = \mathbb{C}$, so hat T einen Eigenwert nach S. 12.10.
Es sei also $\mathbb{K} = \mathbb{R}$ und es sei $A = \varphi_M(T) \in \mathbb{R}^{n\times n} \subset \mathbb{C}^{n\times n}$. Dann hat A (als Matrix in $\mathbb{C}^{n\times n}$) einen Eigenwert λ (wende S. 12.10 auf $z \mapsto Az \in L(\mathbb{C}^n)$ an). Nach 2. ist $\lambda \in \mathbb{R}$, da A hermitesch ist. Es existiert ein $z \in \mathbb{C}^n \setminus \{0\}$ mit $Az = \lambda z$. Ist $z = x + iy$ mit $x, y \in \mathbb{R}^n$, so folgt

$$Ax + iAy = A(x + iy) = Az = \lambda z = \lambda x + i\lambda y \ ,$$

also (Vergleich von Real- und Imaginärteil, wobei man beachte, dass A und λ reell sind)

$$Ax = \lambda x \qquad \text{und} \qquad Ay = \lambda y \ .$$

Aus $z \neq 0$ folgt $x \neq 0$ oder $y \neq 0$. Also hat auch $x \mapsto Ax \in L(\mathbb{R}^n)$ den Eigenwert λ. Damit hat auch T den Eigenwert λ (S. 12.9) und es gilt 1.
4. Es seien λ_1, λ_2 Eigenwerte mit zugehörigen Eigenvektoren v_1, v_2. Dann gilt (beachte λ_2 reell)

$$\begin{aligned} \lambda_1 < v_1, v_2 > &= < \lambda_1 v_1, v_2 >=< Tv_1, v_2 >=< v_1, Tv_2 >= \\ &= < v_1, \lambda_2 v_2 >= \lambda_2 < v_1, v_2 > \ . \end{aligned}$$

Aus $\lambda_1 \neq \lambda_2$ folgt $< v_1, v_2 >= 0$, also v_1, v_2 orthogonal. □

Damit können wir folgenden zentralen Satz der Linearen Algebra beweisen.

Satz 16.8 (Spektralsatz für selbstadjungierte Operatoren; Hauptachsentransformation) *Es sei $(V, < \cdot, \cdot >)$ ein endlich-dimensionaler unitärer Raum, und es sei $T \in L(V)$ selbstadjungiert. Dann existiert eine ONB von V aus Eigenvektoren von T, d. h. T ist unitär diagonalisierbar.*

Beweis.
Wir beweisen die Behauptung durch Induktion nach $n = \dim(V)$. Für $n = 1$ ist die Behauptung offenbar erfüllt.
Es sei V ein $(n+1)$-dimensionaler unitärer Raum, und es sei $T \in L(V)$ selbstadjungiert. Dann hat T nach S. 16.7 einen (reellen) Eigenwert λ. Es sei u ein Eigenvektor zu λ mit $||u|| = 1$ und $U := \text{span}(u)$. Wir betrachten $S := T_{|U^\perp}$ (d. h. S ist die Einschränkung von T auf das orthogonale Komplement von U). Dann ist $S \in L(U^\perp, V)$. Wir zeigen: $S(U^\perp) \subset U^\perp$ (also $S \in L(U^\perp)$).
(Denn: Es sei $v \in U^\perp$. Dann gilt

$$< u, Tv >=< Tu, v >=< \lambda u, v >= \lambda < u, v >= 0 \ ,$$

also $Sv = Tv \in U^\perp$).
Weiter gilt für $v_1, v_2 \in U^\perp$

$$< Sv_1, v_2 >=< Tv_1, v_2 >=< v_1, Tv_2 >=< v_1, Sv_2 > \ .$$

also ist S selbstadjungiert. Da $\dim(U^\perp) = n$ ist, existiert nach Induktionsvoraussetzung eine ONB $(u_1, \ldots, u_n)$ von $U^\perp$ aus Eigenvektoren von S. Dann sind $u_1, \ldots, u_n$ natürlich auch Eigenvektoren von T. Also ist $(u_1, \ldots, u_n, u)$ eine ONB von V aus Eigenvektoren von T. □

Bemerkung 16.9 1. Unter den Voraussetzungen von S. 16.8 seien $\lambda_1, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte von T, und es sei $(v_1, \ldots, v_n)$ eine ONB aus Eigenvektoren. Wir setzen

$$\{v_1, \ldots, v_n\} = \bigcup_{j=1}^{m} \{u_1^{(j)}, \ldots, u_{k_j}^{(j)}\} \ ,$$

wobei $u_1^{(j)}, \ldots, u_{k_j}^{(j)}$ die Eigenvektoren zu λ_j sind $(j = 1, \ldots, m)$. Dann ist $(u_1^{(j)}, \ldots, u_{k_j}^{(j)})$

eine ONB von $U_j := \mathrm{Kern}(T - \lambda_j I)$. Also gilt für alle $v \in V$ nach S. 14.10 und S. 15.6

$$\begin{aligned} T(v) &= T\left(\sum_{\ell=1}^{n} < v, v_\ell > v_\ell\right) = \\ &= T\left(\sum_{j=1}^{m}\sum_{\mu=1}^{k_j} < v, u_\mu^{(j)} > u_\mu^{(j)}\right) = \\ &= \sum_{j=1}^{m}\sum_{\mu=1}^{k_j} < v, u_\mu^{(j)} > Tu_\mu^{(j)} = \sum_{j=1}^{m}\lambda_j \sum_{\mu=1}^{k_j} < v, u_\mu^{(j)} > u_\mu^{(j)} \\ &= \sum_{j=1}^{m}\lambda_j P_{U_j}(v)\ , \end{aligned}$$

und damit

$$T = \sum_{j=1}^{m}\lambda_j P_{U_j}\ . \tag{9}$$

Hierbei gilt $\sum\limits_{j=1}^{m} P_{U_j} = I$ und $P_{U_j}P_{U_k} = \delta_{jk}P_{U_k}$ $(j,k = 1,\dots,m)$([Ü]). Die Darstellung (9) heißt *Spektralzerlegung* von T.

2. Ist $(V, < \cdot,\cdot >)$ ein endlich-dimensionaler unitärer Raum, so sind für beliebiges $T \in L(V)$ die linearen Abbildungen $T\circ T^*$ und $T^*\circ T$ selbstadjungiert (denn $(T\circ T^*)^* = T^{**}\circ T^* = T\circ T^*$ und entsprechend für $T^*\circ T$).

3. Im Falle $\mathbb{K} = \mathbb{R}$ ist die Selbstadjungiertheit von T auch notwendig für die unitäre Diagonalisierbarkeit.

(Denn: Es sei T so, dass eine ONB $M = (v_1,\dots,v_n)$ aus Eigenvektoren von T (zu den Eigenwerten $\lambda_1,\dots,\lambda_n \in \mathbb{R}$) existiert. Dann ist $A = \varphi_M(T) = \mathrm{diag}(\lambda_1,\dots,\lambda_n)$. Also ist nach D./B. 16.5 auch (beachte $\lambda_j \in \mathbb{R}$)

$$\varphi_M(T^*) = A^* = (\mathrm{diag}(\lambda_1,\dots,\lambda_n))^* = \mathrm{diag}(\lambda_1,\dots,\lambda_n) = A = \varphi_M(T)$$

d. h. es gilt $T = T^*$.)

Im Gegensatz dazu gibt es im Falle $\mathbb{K} = \mathbb{C}$ weitere unitär diagonalisierbare lineare Abbildungen, wie wir im folgenden sehen werden.

Definition 16.10 Es sei $(V, < \cdot,\cdot >)$ ein endlich-dimensionaler unitärer Raum über $\mathbb{K}$. Ein $T \in L(V)$ heißt *normal*, falls gilt

$$T\circ T^* = T^*\circ T\ .$$

Entsprechend heißt eine Matrix $A \in \mathbb{K}^{n\times n}$ *normal*, falls

$$AA^* = A^*A$$

gilt, d. h., falls $x \mapsto Ax \in L(\mathbb{K}^n)$ normal ist, wobei $\mathbb{K}^n$ mit dem kanonischen Skalarprodukt versehen ist.

Bemerkung 16.11 Offenbar ist jede selbstadjungierte lineare Abbildung (bzw. jede hermitesche Matrix) normal. Die Umkehrung gilt aber nicht. So ist etwa

$$A = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$$

nicht hermitesch, aber normal, da,

$$\begin{aligned} AA^* = AA^T &= \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix} = A^T A = A^* A \, . \end{aligned}$$

Der folgende Satz gibt eine Charakterisierung normaler Operatoren .

Satz 16.12 *Es sei $(V, < \cdot, \cdot >)$ ein endlich-dimensionaler unitärer Raum. Ein $T \in L(V)$ ist genau dann normal, wenn*

$$||Tv|| = ||T^* v||$$

für alle $v \in V$ gilt.

Beim Beweis verwenden wir

Satz 16.13 *Es sei $(V, < \cdot, \cdot >)$ ein endlich-dimensionaler unitärer Raum, und es sei $T \in L(V)$ mit*

$$< Tv, v >= 0 \qquad (v \in V) \, .$$

Ist $\mathbb{K} = \mathbb{C}$, oder ist $\mathbb{K} = \mathbb{R}$ und T selbstadjungiert, so ist $T = 0$.

Beweis.
1. Es sei $\mathbb{K} = \mathbb{C}$. Dann gilt für $u, w \in V$

$$\begin{aligned} < Tu, w > \; = \; & \frac{1}{4}[< T(u+w) \, , \, u+w > - < T(u-w) \, , \, u-w > \\ & + i < T(u+iw) \, , \, u+iw > - i < T(u-iw) \, , \, u-iw >] \, . \end{aligned}$$

Also ist nach Voraussetzung $< Tu, w >= 0$, d. h. insbesondere $< Tu, Tu >= 0$ und damit $Tu = 0$.

2. Ist $\mathbb{K} = \mathbb{R}$ und T selbstdajungiert, so gilt für $u, w \in V$

$$< Tw, u >=< w, Tu >=< Tu, w >$$

und damit ist

$$< Tu, w >= \frac{1}{4} [< T(u+w), u+w > - < T(u-w), u-w >] .$$

Nach Voraussetzung ist wieder $< Tu, w >= 0$ also insbesondere $< Tu, Tu >= 0$ und damit $Tu = 0$. □

Beweis zu Satz 16.12.
Es gilt mit S. 16.13 (man beachte, dass $T^*T - TT^*$ selbstadjungiert ist):

$$\begin{aligned} T \text{ normal} &\Leftrightarrow T^*T - TT^* = 0 \\ &\Leftrightarrow < (T^*T - TT^*)v, v >= 0 \quad (v \in V) \\ &\Leftrightarrow < T^*Tv, v >=< TT^*v, v > \quad (v \in V) \\ &\Leftrightarrow < Tv, Tv >=< T^*v, T^*v > \quad (v \in V) \\ &\Leftrightarrow ||Tv||^2 = ||T^*v||^2 \quad (v \in V) . \end{aligned}$$

□

Hieraus ergibt sich

Satz 16.14 *Es sei $(V, < \cdot, \cdot >)$ ein endlich-dimensionaler unitärer Raum, und es sei $T \in L(V)$ normal. Ist $v \in V$ ein Eigenvektor zum Eigenwert λ von T, so ist v auch ein Eigenvektor zum Eigenwert $\overline{\lambda}$ von T^*.*

Beweis.
Da T normal ist, ist auch $T - \lambda I$ normal (denn

$$\begin{aligned} &(T - \lambda I)(T - \lambda I)^* = (T - \lambda I)(T^* - \overline{\lambda} I) = \\ &= TT^* - \overline{\lambda} T - \lambda T^* + |\lambda|^2 I = T^*T - \lambda T^* - \overline{\lambda} T + |\lambda|^2 I \\ &= (T^* - \overline{\lambda} I)(T - \lambda I) = (T - \lambda I)^*(T - \lambda I)) . \end{aligned}$$

Nach S. 16.12 gilt

$$0 = ||(T - \lambda I)v|| = ||(T - \lambda I)^* v|| = ||(T^* - \overline{\lambda} I)v|| ,$$

d. h. v ist Eigenvektor zum Eigenwert $\overline{\lambda}$. □

Damit kommen wir zur Charakterisierung unitär diagonalisierbarer Operatoren im komplexen Fall.

Satz 16.15 (Spektralsatz für normale komplexe Operatoren) *Es sei V ein endlichdimensionaler unitärer Raum über $\mathbb{C}$ und es sei $T \in L(V)$. Genau dann existiert eine ONB aus Eigenvektoren von T, wenn T normal ist.*

Beweis.
"$\Rightarrow$": Es sei M eine ONB aus Eigenvektoren von T. Dann ist $D = \varphi_M(T)$ eine Diagonalmatrix. Nach B./D. 16.5 ist $\overline{D} = D^* = \varphi_M(T^*)$ und damit gilt

$$\varphi_M(T \circ T^*) = \varphi_M(T)\varphi_M(T^*) = D\overline{D} = \overline{D}D = \varphi_M(T^* \circ T) \; .$$

Da φ_M injektiv ist, folgt $T \circ T^* = T^* \circ T$, also ist T normal.
"$\Leftarrow$": Der Beweis verläuft vollkommen analog zum Beweis von S. 16.8, wenn wir folgendes zeigen können:
Es sei λ ein Eigenwert von T (existiert, da $\mathbb{K} = \mathbb{C}$) mit zugehörigem Eigenvektor u (mit $||u|| = 1$), und es sei $U := \text{span}(u)$. Dann ist $S := T_{|U^\perp} \in L(U^\perp)$ und S ist normal.
Denn: Es sei $v \in U^\perp$. Dann gilt mit S. 16.14

$$< Sv, u >=< Tv, u >=< v, T^*u >=< v, \overline{\lambda}u >= \lambda < v, u >= 0 \; ,$$

also ist $Sv \in U^\perp$, d. h. $S(U^\perp) \subset U^\perp$ und damit $S \in L(U^\perp)$.
Wieder sei $v \in U^\perp$. Dann gilt entsprechend

$$< u, T^*v >=< Tu, v >= \lambda < u, v >= 0 \; ,$$

also $T^*v \in U^\perp$, d. h. $T^*(U^\perp) \subset U^\perp$.
Nun seien $v_1, v_2 \in U^\perp$. Dann ist

$$< Sv_1, v_2 >=< Tv_1, v_2 >=< v_1, T^*v_2 > \; .$$

Da $T^*v_2 \in U^\perp$ gilt, ist $T^*v_2 = S^*v_2$ nach Definition von S^*, d. h. $S^* = T^*_{|U^\perp}$. Also folgt

$$S \circ S^* = (T_{|U^\perp}) \circ (T^*_{|U^\perp}) = (T^*_{|U^\perp}) \circ (T_{|U^\perp}) = S^* \circ S \; ,$$

d. h. S ist normal. □

Beispiel 16.16 Es sei

$$A = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$$

(vgl. B. 16.11). Dann ist

$$P(\lambda) = \det(A - \lambda E) = \begin{pmatrix} 2-\lambda & -3 \\ 3 & 2-\lambda \end{pmatrix} = (2-\lambda)^2 + 9 = 0$$

für $\lambda_{1,2} = 2 \pm 3i$. Man berechnet, dass

$$(i,1)^T \qquad \text{Eigenvektor zu} \qquad \lambda_1 = 2 + 3i$$

und

$$(-i,1)^T \qquad \text{Eigenvektor zu} \qquad \lambda_2 = 2 - 3i$$

ist. Es gilt

$$(i,1)\begin{pmatrix} -i \\ 1 \end{pmatrix} = i(\overline{-i}) + 1 = 0 \,,$$

also ist $\left(\frac{1}{\sqrt{2}}(i,1)^T, \frac{1}{\sqrt{2}}(-i,1)^T\right)$ eine ONB aus Eigenvektoren in $\mathbb{C}^2$

Bemerkung 16.17 Wie in B. 16.9 sieht man: Ist $\mathbb{K} = \mathbb{C}$, und ist $T \in L(V)$ normal, so gilt nach S. 16.15

$$T = \sum_{j=1}^{m} \lambda_j P_{U_j} \,,$$

wobei $\lambda_1, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte von T und

$$U_j := \text{Kern}(T - \lambda_j \cdot I) \qquad (j = 1, \ldots, m) \,,$$

die zugehörigen Eigenräume sind. Wieder gilt wie in B. 16.9 dabei: $\sum\limits_{j=1}^{m} P_{U_j} = I$ und $P_{U_j} P_{U_k} = \delta_{jk} P_{U_j}$. Man beachte, dass hierbei (im Gegensatz zur Situation in B. 16.9) die λ_j i. a. nicht-reell sind.

17 Unitäre Operatoren und QR-Zerlegung

Wir betrachten jetzt eine weitere wichtige Klasse linearer Abbildungen auf unitären Räumen.

Definition 17.1 Es sei $(V, < \cdot, \cdot >)$ ein unitärer Raum. Ein $T \in L(V)$ heißt *unitär* (oder *isometrisch* oder *Isometrie*), falls gilt

$$||Tv|| = ||v|| \qquad \text{für alle} \quad v \in V \,.$$

Im Falle $\mathbb{K} = \mathbb{R}$ heißt T dann auch *orthogonal*.
Eine Matrix $A \in \mathbb{K}^{n \times n}$ heißt *unitär* (bzw. *orthogonal* für $\mathbb{K} = \mathbb{R}$), falls $x \mapsto Ax \in L(\mathbb{K}^n)$ (mit kanonischem Skalarprodukt) unitär ist.

Beispiel 17.2 1. Es sei $(v_1, \dots, v_n)$ eine ONB von V, und es seien $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ mit $|\lambda_j| = 1$ für $j = 1, \dots, n$. Dann ist $T \in L(V)$ mit $T(v_k) = \lambda_k v_k$ $(k = 1, \dots, n)$, d. h.

$$T(v) = T\left(\sum_{j=1}^{n} < v, v_j > v_j\right) = \sum_{j=1}^{n} < v, v_j > \lambda_j v_j \qquad (v \in V) ,$$

unitär
(Denn: Für $v \in V$ gilt nach (7)

$$||Tv||^2 = \sum_{j=1}^{n} | < v, v_j > \lambda_j|^2 = \sum_{j=1}^{n} | < v, v_j > |^2 = ||v||^2 .)$$

Insbesondere ist $A = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{K}^{n\times n}$ mit $|\lambda_j| = 1$ für $j = 1, \dots, n$ unitär.
2. (Spiegelungen) Es sei $a \in V$ mit $||a|| = 1$. Wir betrachten $S = S_a : V \to V$ mit

$$S_a(v) := v - 2 < v, a > a \qquad (v \in V) .$$

Dann gilt
1. $S_a(a) = -a$, $S_a(b) = b$ für alle $b \in < a >^{\perp}$,
2. $S_a^2 = S_a \circ S_a = I$,
3. $||S_a v|| = ||v||$ für alle $v \in V$ (d. h. S_a ist unitär).
(Denn: Es gilt $V =< a > \oplus < a >^{\perp}$ nach S. 14.14, d. h. jedes $v \in V$ hat genau eine Darstellung

$$v = \lambda a + b \qquad \text{mit } \lambda \in \mathbb{K},\ b \perp a .$$

Daraus folgt

$$\begin{aligned} S_a(v) &= \lambda a + b - 2 < \lambda a + b, a > a = \\ &= \lambda a + b - 2\lambda \underbrace{< a, a >}_{=1} a - 2\lambda \underbrace{< b, a >}_{=0} a = -\lambda a + b \end{aligned}$$

(d. h. S_a ist "Spiegelung an $< a >^{\perp}$"). Hieraus ergeben sich sofort 1. und 2. Außerdem gilt mit dem Satz von Pythagoras:

$$||S_a(v)||^2 = || - \lambda a + b||^2 = | - \lambda|^2 ||a||^2 + ||b||^2 = |\lambda|^2 ||a||^2 + ||b||^2 = ||v||^2 .$$

Also ist S_a unitär.)
Im Falle $V = \mathbb{K}^n$ ist (da $a\overline{a}^T \cdot x = a(\overline{a}^T x) = \overline{a}^T x \cdot a = x^T \overline{a} \cdot a$)

$$Q_a := E - 2a\overline{a}^T = E - 2aa^*$$

die Matrix von S_a bzgl. der kanonischen Basis; Q_a ist also unitär. Außerdem ist Q_a auch hermitesch, denn

$$\begin{aligned} Q_a^* &= (E - 2aa^*)^* = E^* - 2(aa^*)^* = \\ &= E - 2a^{**}a^* = E - 2aa^* = Q_a . \end{aligned}$$

Matrizen der Form Q_a heißen *Householder-Matrizen.*

Wir stellen im folgenden Satz verschiedene Charakterisierungen unitärer Abbildungen zusammen.

Satz 17.3 *Es sei $(V, < \cdot, \cdot >)$ ein endlich-dimensionaler unitärer Raum, und es sei $T \in L(V)$. Dann sind folgende Aussagen äquivalent.*
a) T ist unitär,
b) $< Tv_1, Tv_2 >=< v_1, v_2 >$ für alle $v_1, v_2 \in V$,
c) $T^* \circ T = I$,
d) Ist $(v_1, \dots, v_n)$ eine ONB von V, so ist $(Tv_1, \dots, Tv_n)$ eine ONB von V,
e) Es existiert eine ONB $(v_1, \dots, v_n)$ von V so, dass $(Tv_1, \dots, Tv_n)$ eine ONB von V ist,
f) T ist bijektiv mit $T^{-1} = T^*$,
g) T^* ist unitär.

Beweis.
1. a) $\Rightarrow$ b): Ist $\mathbb{K} = \mathbb{R}$, so folgt aus der Polarisierungsidentität (B. 13.7)

$$\begin{aligned} < Tv_1, Tv_2 > &= \frac{1}{4}(||Tv_1 + Tv_2||^2 - ||Tv_1 - Tv_2||^2) = \\ &= \frac{1}{4}(||T(v_1 + v_2)||^2 - ||T(v_1 - v_2)||^2) = \\ &= \frac{1}{4}(||v_1 + v_2||^2 - ||v_1 - v_2||^2) =< v_1, v_2 > \ . \end{aligned}$$

Eine entsprechende Rechnung liefert die Behauptung im Fall $\mathbb{K} = \mathbb{C}$.
2. b) $\Rightarrow$ c): Für alle $v_1, v_2 \in V$ gilt (nach Definition von T^*)

$$\begin{aligned} < ((T^* \circ T) - I)v_1, v_2 > &= < (T^* \circ T)v_1, v_2 > - < v_1, v_2 >= \\ &= < Tv_1, Tv_2 > - < v_1, v_2 >= 0 \ , \end{aligned}$$

also insbesondere für $v_2 = (T^* \circ T - I)v_1$. Hieraus folgt

$$||(T^* \circ T - I)v_1||^2 = 0 \ ,$$

d. h. $(T^* \circ T - I)v_1 = v_1$ bzw. $T^* \circ T = I$.
3. c) $\Rightarrow$ d): Es sei $(v_1, \dots, v_n)$ eine ONB von V. Dann gilt für $j, k \in \{1, \dots, n\}$

$$< Tv_j, Tv_k >=< v_j, (T^* \circ T)v_k >=< v_j, v_k >= \delta_{jk} \ .$$

4. d) $\Rightarrow$ e) ist klar (F. 14.12).

5. e) $\Rightarrow$ a): Es sei $(v_1, \dots, v_n)$ eine ONB von V so, dass $(Tv_1, \dots, Tv_n)$ ebenfalls eine ONB ist. Für $v \in V$ gilt dann mit Pythagoras und (7)

$$\begin{aligned} ||Tv||^2 &= ||T(\sum_{j=1}^{n} < v, v_j > v_j)||^2 = ||\sum_{j=1}^{n} < v, v_j > Tv_j||^2 \\ &= \sum_{j=1}^{n} | < v, v_j > |^2 = ||v||^2 . \end{aligned}$$

Damit sind a) bis e) äquivalent.
6. c) $\Leftrightarrow$ f): Ist $T^* \circ T = I$, so ist T injektiv, also auch bijektiv nach S. 7.5. Außerdem ist $T^{-1} = T^*$. Umgekehrt folgt aus f) natürlich auch c).
7. f) $\Leftrightarrow$ g): Es gilt $T^{-1} = T^*$ genau dann, wenn $(T^*)^{-1} = T (= T^{**})$ gilt. Also folgt aus der (bewiesenen) Äquivalenz von a) und f) auch die Äquivalenz von f) und g). □

Folgerung 17.4 1. Nach S. 17.3 ist jede unitäre lineare Abbildung auf (einem endlichdimensionalen unitären Raum) V auch normal (denn: $T^* \circ T = I = (T^*)^* \circ T^* = T \circ T^*$). Also folgt aus S. 16.15, dass im Falle $\mathbb{K} = \mathbb{C}$ eine ONB aus Eigenvektoren existiert (d. h. T ist unitär diagonalisierbar). Außerdem folgt aus der Definition sofort, dass alle Eigenwerte den Betrag 1 haben.
2. Es sei $C \in \mathbb{K}^{n \times n}$, und es seinen $c^{(k)} = Ce_k$ für $k = 1, \dots, n$ die Spalten von C. Dann ist C genau dann unitär, wenn $(c^{(1)}, \dots, c^{(n)})$ eine ONB von $\mathbb{K}^n$ ist. Weiter ist dies genau dann der Fall, wenn C invertierbar ist mit

$$C^{-1} = C^* .$$

(Denn: Die Behauptungen ergeben sich durch Anwendung von S.17.3 unter Beachtung, dass $(e_1, \dots, e_n)$ eine ONB von $\mathbb{K}^n$ mit dem kanonischen Skalarprodukt ist.)

Im Fall $\mathbb{K} = \mathbb{R}$ ist nicht jede orthogonale lineare Abbildung diagonalisierbar, wie das folgende Beispiel zeigt.

Beispiel 17.5 Für $\alpha \in (0, \pi)$ sei

$$A = A_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

(A beschreibt die Drehung um den Winkel α um 0 in $\mathbb{R}^2$). Dann gilt

$$Ae_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} , \; Ae_2 = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix} .$$

Also ist

$$< Ae_1, Ae_2 > = -\cos\alpha \sin\alpha + \cos\alpha \sin\alpha = 0$$

und

$$||Ae_1||^2 = \cos^2\alpha + \sin^2\alpha = 1 = ||Ae_2||^2 \ ,$$

d. h. (Ae_1, Ae_2) ist eine ONB von $\mathbb{R}^2$. Nach F. 17.4.2 ist A orthogonal. Es gilt

$$\begin{aligned} P(\lambda) = \det(A - \lambda E) &= (\cos\alpha - \lambda)^2 + \sin^2\alpha = 1 - 2\lambda\cos\alpha + \lambda^2 \\ &\geq 1 + \lambda^2 - 2|\lambda\cos\alpha| > 1 + \lambda^2 - 2|\lambda| = (1 - |\lambda|)^2 \geq 0 \ , \end{aligned}$$

d. h. $P(\lambda) > 0$ für alle $\lambda \in \mathbb{R}$. Also hat A keine (reellen) Eigenwerte und ist damit insbesondere nicht diagonalisierbar.

Folgerung 17.6 Es sei $A \in \mathbb{K}^{n\times n}$ symmetrisch (im Falle $\mathbb{K} = \mathbb{R}$) oder normal (im Falle $\mathbb{K} = \mathbb{C}$). Dann existiert nach S. 16.8 bzw. S. 16.15 (Spektralsätze) eine ONB $(v_1, \ldots, v_n)$ von $\mathbb{K}^n$ aus Eigenvektoren (zu den Eigenwerten $\lambda_1, \ldots, \lambda_n$) von A. Ist $C \in \mathbb{K}^{n\times n}$ die Matrix mit den Spalten $v_1, \ldots, v_n$, d. h. $C = (Ce_1, \ldots, Ce_n) = (v_1, \ldots, v_n)$, so ist

$$A = C \cdot \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \cdot C^{-1}$$

(vgl. B. 12.14.3). Nach F. 17.4.2 ist C unitär (bzw. orthogonal im Falle $\mathbb{K} = \mathbb{R}$), d. h.

$$C^{-1} = C^* \qquad (= C^T \quad \text{für } \mathbb{K} = \mathbb{R}) \ .$$

Damit gilt im Falle $\mathbb{K} = \mathbb{R}$

$$\boxed{A = C \cdot \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \cdot C^T}$$

(dies wird auch als *Hauptachsentransformation einer symmetrischen Matrix* bezeichnet) und im Falle $\mathbb{K} = \mathbb{C}$

$$\boxed{A = C \cdot \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \cdot C^*}$$

(mit reellen $\lambda_n, \ldots, \lambda_n$ im Falle einer hermiteschen Matrix A).

Im Abschnitt 10 haben wir gesehen, wie man die LR-Zerlegung bestimmter invertierbarer Matrizen berechnen kann. Wir werden jetzt eine andere Zerlegung kennenlernen, die auf unitären Transformationen beruht. Wesentlicher Vorteil dieser Zerlegung ist die größere numerische Stabilität.

Wir werden zunächst noch einmal näher auf Householder-Matrizen eingehen.

Satz 17.7 *Es sei* $x = (x_1, \dots, x_n)^T \in \mathbb{K}^n$. *Dann existiert ein* $a \in \mathbb{K}$ *mit* $Q_a x \in < e_1 >$, *d. h.* $Q_a x = \lambda e_1$ *für ein* $\lambda \in \mathbb{K}$, *wobei* $e_1 = (1, 0, \dots, 0)^T$.

Beweis.
Zunächst gilt: Ist $y = Q_a x$, so ist (beachte: $Q_a = Q_a^*$)

$$< y, x >=< Q_a x, x >=< x, Q_a^* x >=< x, Q_a x >= \overline{< Q_a x, x >} ,$$

also $< y, x > \in \mathbb{R}$.
Wir versuchen, a durch die Forderung $Q_a x = \lambda e_1$ zu bestimmen. Aus $||Q_a x|| = ||x||$ folgt zunächst $|\lambda| = ||x||$. Da $\lambda \overline{x}_1 =< \lambda e_1, x >$ reell sein muss, ist

$$\lambda = \pm e^{i\alpha} ||x||$$

wobei $\alpha := \arg(x_1)$ (d. h. $x_1 = e^{i\alpha}|x_1|$) im Falle $x_1 \neq 0$. (Weiter setzen wir $\alpha := 0$ falls $x_1 = 0$.) Aus

$$x - 2 < x, a > a = \lambda e_1$$

und $||a|| = 1$ folgt

$$a = \frac{x - \lambda e_1}{||x - \lambda e_1||}$$

falls $x \notin < e_1 >$; im Falle $x \in < e_1 >$ wählen wir $a = e_1$.
Weiter folgt mit $x_1 = e^{i\alpha}|x_1|$

$$\begin{aligned} ||x - \lambda e_1||^2 &= || x \mp ||x|| e^{i\alpha} e_1 ||^2 = |x_1 \mp ||x|| e^{i\alpha}|^2 + \sum_{j=2}^{n} |x_j|^2 \\ &= (|x_1| \mp ||x||)^2 + \sum_{j=2}^{n} |x_j|^2 . \end{aligned}$$

Wir wählen das Vorzeichen + (dann ist jedenfalls $||x|| + |x_1| > 0$, wobei o. E. $x \neq 0$ vorausgesetzt wird), d. h. $\lambda = -||x|| e^{i\alpha}$. Dann gilt

$$(|x_1| + ||x||)^2 = |x_1|^2 + 2||x|| \, |x_1| + ||x||^2 ,$$

also

$$||x - \lambda e_1||^2 = 2||x||^2 + 2||x|| \, |x_1| .$$

Folglich ist

$$a = \frac{1}{(2||x||^2 + 2||x|| \, |x_1|)^{1/2}} (x + ||x|| e^{i\alpha} e_1) .$$

(Dies gilt auch im Falle $x \in < e_1 > \setminus \{0\}$; für $x = 0$ wähle man etwa $a = e_1$).
Umgekehrt sieht man, dass mit diesem a die Behauptung gilt ([Ü]). □

Hiermit lässt sich zeigen

Satz 17.8 *Es sei $A \in \mathbb{K}^{n\times n}$. Dann existieren eine unitäre Matrix $Q \in \mathbb{K}^{n\times n}$ und eine obere Dreiecksmatrix $R \in \mathbb{K}^{n\times n}$ mit*

$$A = Q \cdot R$$

(eine solche Zerlegung von A heißt QR-Zerlegung).

Beweis.
Wir konstruieren $(n-1)$ unitäre und hermitesche Matrizen $Q_1, \dots, Q_{n-1}$ (die i. w. aus Householder-Matrizen zusammengesetzt sind) so, dass

$$Q_{n-1} \cdots Q_1 A = R$$

mit einer oberen Dreiecksmatrix R. Dann gilt mit der unitären Matrix

$$Q := (Q_{n-1} \cdots Q_1)^{-1} = Q_1^{-1} \cdots Q_{n-1}^{-1} = Q_1^* \cdots Q_{n-1}^* = Q_1 \cdots Q_{n-1}$$

die Behauptung. Wir setzen $A = A_0 := (a_0^{(1)}, \dots, a_0^{(n)})$, d. h. $a^{(k)}$ sind die Spalten von A. Es sei Q_1 eine Householder-Matrix so, dass für ein $r_1 \in \mathbb{K}$

$$Q_1 \cdot a_0^{(1)} = r_1 e_1$$

(existiert nach S. 17.7). Dann gilt für $A_1 := Q_1 \cdot A$

$$A_1 = \begin{bmatrix} r_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & \tilde{A}_1 & \\ 0 & & & \end{bmatrix}$$

mit $\tilde{A}_1 \in \mathbb{K}^{(n-1)\times(n-1)}$. Nun beachten wir $\tilde{A}_1 =: (a_1^{(1)}, \dots, a_1^{(n-1)})$ und wählen $\tilde{Q}_2 \in \mathbb{K}^{(n-1)\times(n-1)}$ gemäß S. 17.7 so, dass

$$\tilde{Q}_2 a_1^{(1)} = r_2 e_1$$

für ein $r_2 \in \mathbb{K}$. Für die unitäre und hermitesche Matrix

$$Q_2 := \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{Q}_2 & \\ 0 & & & \end{bmatrix} \in \mathbb{K}^{n\times n}$$

gilt dann

$$A_2 := Q_2 \cdot A_1 = Q_2 Q_1 \cdot A = \begin{bmatrix} r_1 & * & \cdots & \cdots & * \\ 0 & r_2 & * & \cdots & * \\ \vdots & 0 & & & \\ \vdots & \vdots & & \tilde{A}_2 & \\ 0 & 0 & & & \end{bmatrix}$$

mit $\tilde{A}_2 \in \mathbb{K}^{(n-2)\times(n-2)}$. So fortfahrend erhalten wir im j-ten Schritt eine unitäre und hermitesche Matrix Q_j der Form

$$Q_j = \begin{bmatrix} E_{j-1} & 0 \\ 0 & \tilde{Q}_j \end{bmatrix},$$

wobei $\tilde{Q}_j \in \mathbb{K}^{(n-j+1)\times(n-j+1)}$ eine Householder-Matrix ist, und die

$$A_j = Q_j A_{j-1} = \begin{bmatrix} r_1 & * & \dots & & & * \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & & r_j & * & \dots & * \\ \vdots & & 0 & & & \\ \vdots & & \vdots & & \tilde{A}_j & \\ 0 & & 0 & & & \end{bmatrix}$$

erfüllt. Nach $(n-1)$ Schritten ist schließlich

$$R := A_{n-1} = Q_{n-1} A_{n-2} = \begin{bmatrix} r_1 & * & \cdots & * \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & * \\ 0 & \dots & 0 & r_n \end{bmatrix} = Q_{n-1} \cdots Q_1 \cdot A$$

wie behauptet. □

Bemerkung 17.9 1. Das im Beweis zu S. 17.8 dargestellte Verfahren zur Berechnung einer QR-Zerlegung heißt Householder-Verfahren. Es gibt ein alternatives Verfahren, das auf dem Schmidtschen Orthogonalisierungsverfahren beruht.
2. Man kann zeigen, dass zu jeder invertierbaren Matrix A eine QR-Zerlegung existiert mit $r_{jj} > 0$ für alle j, wobei $R = (r_{jk})$. Mit dieser Normierung ist die Zerlegung dann eindeutig ([Ü]). Man spricht dann auch von der QR-Zerlegung von A.
3. Hat man für eine invertierbare Matrix $A \in \mathbb{K}^{n\times n}$ eine QR-Zerlegung bzw. nur die Matrizen $Q_1, \dots, Q_{n-1}$ und R bestimmt, so ist die Lösung $A^{-1}b$ von $Ax = b$ für $b \in \mathbb{K}^n$ leicht zu berechnen durch

$$Rx = y \qquad \text{und} \qquad Qy = b\,.$$

Dabei ist y sofort gegeben durch $y = Q^{-1}b = Q_{n-1} \cdots Q_1 \cdot b$ und $Rx = y$ kann wie in B. 10.8 gelöst werden. Der wesentliche Vorteil gegenüber der LR-Zerlegung (die nur etwa die Hälfte des Rechenaufwandes erfordert) liegt in der größeren numerischen Stabilität (→ Numerik).

Beispiel 17.10 Es sei

$$A = \begin{pmatrix} 3 & 1 & 2 \\ 0 & 3 & -1 \\ 4 & 8 & 6 \end{pmatrix} .$$

Dann gilt (siehe Beweis zu S. 17.7 und S. 17.8):

$$Q_1 = E - 2a_1 a_1^T \qquad \text{und} \qquad a_1 = \frac{1}{\sqrt{80}} \begin{pmatrix} 8 \\ 0 \\ 4 \end{pmatrix} ,$$

d. h.

$$Q_1 = \begin{pmatrix} -3/5 & 0 & -4/5 \\ 0 & 1 & 0 \\ -4/5 & 0 & 3/5 \end{pmatrix} .$$

Also ist

$$A_1 = Q_1 A = \begin{pmatrix} -5 & -7 & -6 \\ 0 & 3 & -1 \\ 0 & 4 & 2 \end{pmatrix} .$$

Hieraus folgt mit $\tilde{A}_1 = \begin{pmatrix} 3 & -1 \\ 4 & 2 \end{pmatrix}$

$$\tilde{Q}_2 = E - 2a_2 a_2^T \qquad \text{wobei} \qquad a_2 = \frac{1}{\sqrt{80}} \begin{pmatrix} 8 \\ 4 \end{pmatrix} ,$$

d. h.

$$\tilde{Q}_2 = \begin{pmatrix} -3/5 & -4/5 \\ -4/5 & 3/5 \end{pmatrix}$$

bzw.

$$Q_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -3/5 & -4/5 \\ 0 & -4/5 & 3/5 \end{pmatrix} .$$

Also ist

$$R = A_2 = Q_2 \cdot A_1 = \begin{pmatrix} -5 & -7 & -6 \\ 0 & -5 & -1 \\ 0 & 0 & 2 \end{pmatrix} = Q_2 Q_1 \cdot A .$$

Schließlich gilt

$$\begin{aligned} Q &= (Q_2Q_1)^{-1} = Q_1^{-1}Q_2^{-1} = Q_1^*Q_2^* = Q_1Q_2 \\ &= \frac{1}{25}\begin{pmatrix} -15 & 16 & -12 \\ 0 & -15 & -20 \\ -20 & -12 & 9 \end{pmatrix} . \end{aligned}$$

18 Definite Operatoren

Es sei V ein endlich-dimensionaler unitärer Raum über $\mathbb{K}$, und es sei $T \in L(V)$ selbstadjungiert. Dann gilt für alle $v \in V$

$$< Tv, v >=< v, T^*v >=< v, Tv >= \overline{< Tv, v >} ,$$

d. h. $< Tv, v > \in \mathbb{R}$.

Definition 18.1 Es sei V ein endlich-dimensionaler unitärer Raum, und es sei $T \in L(V)$ selbstadjungiert. Dann heißt T
1. *positiv semidefinit*, falls $< Tv, v > \geq 0$ für alle $v \in V$.
2. *positiv definit*, falls $< Tv, v > > 0$ für alle $v \in V \setminus \{0\}$.
3. *negativ (semi-)definit*, falls $-T$ positiv (semi-)definit ist,
4. *indefinit*, falls T weder positiv noch negativ semidefinit ist.
Wie üblich heißt eine Matrix $A \in \mathbb{K}^{n\times n}$ *positiv (negativ) (semi-)definit* bzw. *indefinit*, falls das Entsprechende für $x \mapsto Ax \in L(\mathbb{K}^n)$ (mit kanonischem Skalarprodukt) gilt. Ferner heißt $q = q_A : \mathbb{K}^n \to \mathbb{K}$ mit

$$q(x) = q_A(x) = x^*Ax =< Ax, x > \qquad (x \in \mathbb{K}^n)$$

quadratische Form (von A).

Bemerkung 18.2 Ist $A \in \mathbb{K}^{n\times n}$ beliebig, so ist $A^*A \in \mathbb{K}^{n\times n}$ positiv semidefinit (denn $< A^*Ax, x >=< Ax, Ax > \geq 0$). Weiter ist A^*A positiv definit, wenn $< Ax, Ax >> 0$ für alle $x \in \mathbb{K}^n \setminus \{0\}$, also genau dann, wenn A invertierbar ist (warum?).

Eine wesentliche Verallgemeinerung dieses Sachverhaltes liefert

Satz 18.3 *Es sei $(V, < \cdot, \cdot >)$ ein endlich-dimensionaler unitärer Raum, und es sei $T \in L(V)$. Dann sind äquivalent:*
a) T ist positiv semidefinit,
b) T ist selbstadjungiert und alle Eigenwerte sind ≥ 0,
c) es existiert ein positiv semidefinites $S \in L(V)$ mit $S^2 = T$,
d) es existiert ein selbstadjungiertes $S \in L(V)$ mit $S^2 = T$,
e) es existiert ein $S \in L(V)$ mit $S^* \circ S = T$.

Beweis.
a) $\Rightarrow$ b): Nach Definition ist T selbstadjungiert. Weiter gilt: Ist λ ein Eigenwert von T und $v \neq 0$ ein zugehöriger Eigenvektor, so gilt

$$0 \leq < Tv, v > = < \lambda v, v > = \lambda < v, v > ,$$

also $\lambda \geq 0$ (da $(v, v) > 0$).
b) $\Rightarrow$ c): Nach dem Spektralsatz (S. 16.8) existiert eine ONB $(v_1, \ldots, v_n)$ aus Eigenvektoren zu den Eigenwerten $\lambda_1, \ldots, \lambda_n \in [0, \infty)$. Wir definieren $S \in L(V)$ durch

$$S(v_k) := \sqrt{\lambda_k} v_k \qquad (k = 1, \ldots, n)$$

d. h. $S(v) = S(\sum\limits_{j=1}^{n} < v, v_j > v_j) = \sum\limits_{j=1}^{n} \sqrt{\lambda_j} < v, v_j > v_j$. Dann gilt

$$< Sv, v > = \sum_{j=1}^{n} \sqrt{\lambda_j} < v, v_j > < v_j, v > = \sum_{j=1}^{n} \sqrt{\lambda_j} \, | < v, v_j > |^2 \geq 0$$

und

$$\begin{aligned} < v, Sv > &= < v, \sum_{j=1}^{n} \sqrt{\lambda_j} < v, v_j > v_j > = \sum_{j=1}^{n} \sqrt{\lambda_j} \, \overline{< v, v_j >} < v, v_j > = \\ &= \sum_{j=1}^{n} \sqrt{\lambda_j} \, | < v, v_j > |^2 \; (= < Sv, v >) \, . \end{aligned}$$

Also ist S positiv semidefinit. Aus $S^2(v_k) = \sqrt{\lambda_k}\sqrt{\lambda_k} v_k = Tv_k$ für $k = 1, \ldots, n$ folgt $S = T^2$.
c) $\Rightarrow$ d) und d) $\Rightarrow$ e) sind klar.
e) $\Rightarrow$ a): Es sei $S \in L(V)$ mit $S^* \circ S = T$. Dann ist $T = T^*$ und es gilt für alle $v \in V$

$$< Tv, v > = < (S^* \circ S)v, v > = < Sv, Sv > \geq 0 \, .$$

Also ist T positiv semidefinit. □

Bemerkung 18.4 Es sei V ein endlich-dimensionaler unitärer Raum, und es sei $T \in L(V)$.
1. Wie im Beweis zu S. 18.3 sieht man, dass ein T genau dann positiv definit ist, wenn T selbstadjungiert ist mit durchweg positiven Eigenwerten. (Denn: Ist T positiv definit, so gilt $\lambda > 0$ im Beweisschritt a) $\Rightarrow$ b). Ist umgekehrt T selbstadjungiert mit positiven Eigenwerten $\lambda_1, \ldots, \lambda_n$, so ist

$$< Tv, v > = \sum_{j=1}^{n} \lambda_j < v, v_j > < v_j, v > = \sum_{j=1}^{n} \lambda_j \, | < v, v_j > |^2 > 0$$

für $v \neq 0$, da $< v, v_j > \neq 0$ für mindestens ein j.)

Entsprechende Aussagen gelten für "negativ definit". Außerdem ist ein selbstadjungiertes $T \in L(V)$ genau dann indefinit, wenn sowohl positive als auch negative Eigenwerte existieren.
2. Im Allgemeinen existieren viele selbstadjungierte $S \in L(V)$ mit $S^2 = T$:
Ist etwa $A = E$ in $\mathbb{K}^{2\times 2}$, so gilt für alle $\vartheta \in \mathbb{R}$

$$E = Q_\vartheta^2$$

mit

$$Q_\vartheta := \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix} .$$

(Q_ϑ ist Spiegelung an der Gerade G_ϑ mit Winkel $\vartheta/2$ zur x_1-Achse). Man kann jedoch zeigen, dass nur ein positiv semidefinites $S \in L(V)$, nämlich das aus dem Beweis zu S. 18.3, mit $S^2 = T$ existiert (auf den Beweis wollen wir verzichten). Man schreibt dann auch

$$S := \sqrt{T}$$

für dieses S.

Insbesondere zeigt S. 18.3, dass jede positiv definite Matrix $A \in \mathbb{K}^{n\times n}$ eine Darstellung $A = B^*B$ mit (in diesem Fall invertierbarer) Matrix $B \in \mathbb{K}^{n\times n}$ hat. Der folgende Satz zeigt, dass B sogar als Dreiecksmatrix gewählt werden kann.

Satz 18.5 *E sei $A \in \mathbb{K}^{n\times n}$. Genau dann ist A positiv definit, wenn eine obere Dreiecksmatrix $R = (r_{jk}) \in \mathbb{K}^{n\times n}$ existiert mit*

$$A = R^*R$$

und $r_{jj} \neq 0$ für $j = 1, \dots, n$.

Beweis.
1. "⇒" Es sei A positiv definit. Dann existieren nach S. 18.3 eine (positiv semidefinite) Matrix B mit $B^*B = A$. Da A positiv definit ist, ist B invertierbar (0 kann kein Eigenwert von B sein). Nach S. 17.8 existieren eine unitäre Matrix Q und eine obere Dreiecksmatrix $R = (r_{jk})$ mit $B = QR$. Dabei ist $\det(R) \neq 0$ und deshalb $r_{jj} \neq 0$ ($j = 1, \dots, n$). Es gilt

$$A = B^*B = (QR)^*(QR) = R^*Q^*QR = R^*R .$$

2. "⇐" Ist $A = R^*R$, so ist A positiv semidefinit nach B. 18.2. Da $r_{jj} \neq 0$ für $j = 1, \dots, n$ d. h. $\det(R) \neq 0$ ist, ist R invertierbar. Also ist A positiv definit nach B. 18.2. □

Bemerkung 18.6 In der Situation von S. 18.5 kann R so gewählt werden, dass $r_{jj} > 0$ für alle $j \in \{1, \dots, n\}$. Mit dieser Normierung ist R dann eindeutig bestimmt und die Zerlegung $A = R^*R$ heißt *Cholesky-Zerlegung* von A.
(Denn:
1. Existenz: Nach S. 18.5 existiert eine (invertierbare) obere Dreiecksmatrix $R = (r_{jk})$ mit $r_{jj} \neq 0$ und $A = R^*R$. Es sei

$$D = \operatorname{diag}(e^{-i\alpha_1}, \dots, e^{-i\alpha_n}) ,$$

wobei $\alpha_j = \arg(r_{jj})$ (d. h. $r_{jj} = |r_{jj}|e^{i\alpha_j}$). Dann gilt

$$\overline{D}D = \operatorname{diag}(e^{i\alpha_1}e^{-i\alpha_1}, \dots, e^{i\alpha_n}e^{-i\alpha_n}) = E$$

und

$$\tilde{R} = DR = \begin{bmatrix} |r_{11}| & * & \dots & * \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & |r_{nn}| \end{bmatrix} ,$$

d. h. $\tilde{R}$ ist eine obere Dreiecksmatrix mit positivem Diagonalelementen. Außerdem gilt

$$\tilde{R}^*\tilde{R} = (DR)^*DR = R^*\overline{D}DR = R^*R = A .$$

2. Eindeutigkeit: Es seien R und S obere Dreiecksmatrizen mit $r_{jj} > 0$ und $s_{jj} > 0$ für $j = 1, \dots, n$ sowie $R^*R = A = S^*S$. Dann sind R^{-1} und S^{-1} (und damit auch RS^{-1} sowie SR^{-1}) obere Dreiecksmatrizen. Aus $R^*R = S^*S$ folgt

$$SR^{-1} = (S^*)^{-1}R^* = (S^{-1})^*R^* = (RS^{-1})^* ,$$

d. h. $SR^{-1} = D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Also ist $S = DR$ und damit $R^*R = (DR)^*DR = R^*\overline{D}DR$.
Hieraus folgt $E = \overline{D}D = \operatorname{diag}(|\lambda_1|^2, \dots, |\lambda_n|^2)$, d. h. $|\lambda_j| = 1$ für $j = 1, \dots, n$. Außerdem folgt aus $S = DR$, dass $s_{jj} = \lambda_j r_{jj}$ für $j = 1, \dots, n$ gilt, also $\lambda_j = s_{jj}/r_{jj} > 0$. Damit ist $\lambda_j = 1$ für $j = 1, \dots, n$, d. h. $D = E$ und damit $S = R$.)

Bemerkung und Definition 18.7 Es seien V bzw. W endlich-dimensionale unitäre Räume (über dem gemeinsamen Körper $\mathbb{K}$), und es sei $T \in L(V, W)$ beliebig. Dann ist $T^* \circ T \in L(V)$ positiv semidefinit. Sind $\mu_1, \dots, \mu_n$ die Eigenwerte von $T^* \circ T$, so gilt $\mu_j \geq 0$ für $j = 1, \dots, n$. Die Zahlen

$$\lambda_j := \sqrt{\mu_j} \qquad (j = 1, \dots, n)$$

(oder auch nur die positiven dieser Zahlen) heißen dann *Singulärwerte von* T.
Wie üblich sind die *Singulärwerte* von $A \in \mathbb{K}^{m\times n}$ definiert als die Singulärwerte von $x \mapsto Ax \in L(\mathbb{K}^n, \mathbb{K}^m)$ (mit kanonischem Skalarprodukt).
Damit gilt

Satz 18.8 (Singulärwertzerlegung) *Es seien V, W endlich-dimensionale unitäre Räume über $\mathbb{K}$, und es sei $T \in L(V,W)$. Ferner seinen $\lambda_1, \dots, \lambda_r$ die positiven Singulärwerte von T. Dann existieren eine ONB $N = (v_1, \dots, v_n)$ von V und eine ONB $M = (w_1, \dots, w_m)$ von W so, dass*

$$\varphi_{M,N}(T) = D = (d_{jk}) \ \in \mathbb{K}^{m \times n}$$

mit

$$d_{jk} := \begin{cases} \lambda_k & , \quad \textit{falls } j = k,\ k = 1, \dots, r \\ 0 & , \quad \textit{sonst} \end{cases} .$$

Beweis.
Da T^*T positiv semidefinit ist, existiert nach S. 16.8 eine ONB $N = (v_1, \dots, v_n)$ von V aus Eigenvektoren von T^*T zu den Eigenwerten $\mu_1, \dots, \mu_n \geq 0$ von T^*T. O. E. sei $T \neq 0$. Dann ist auch $T^*T \neq 0$. (Denn: Ist $T^*T = 0$, so ist $< Tv_k, Tv_k >=< v_k, T^*Tv_k >= 0$, d. h. $Tv_k = 0$ für $k = 1, \dots, n$ und damit $T = 0$.)
Weiter sei o. E. $r \in \{1, \dots, n\}$ so, dass $\mu_1, \dots, \mu_r > 0$ und $\mu_{r+1}, \dots, \mu_n = 0$. Dann gilt für $j, k = 1, \dots, n$

$$< Tv_j, Tv_k >=< v_j, T^*Tv_k >= \mu_k < v_j, v_k >= \mu_k \delta_{jk} ,$$

d. h. $w_1 = \frac{1}{\sqrt{\mu_1}} Tv_1, \dots, w_r = \frac{1}{\sqrt{\mu_r}} Tv_r$ ist ein ONS in W (und $r \leq m$). Nach dem dem Schmidt'schen Orthogonalisierungsverfahren und dem Basisergänzungssatz läßt sich $w_1, \dots, w_r$ zu einer ONB $M = (w_1, \dots, w_m)$ von W fortsetzen. Dann gilt für $D = (d_{jk}) = \varphi_{M,N}(T)$ nach S. 14.10

$$\sum_{j=1}^{m} d_{jk} w_j = Tv_k = \sum_{j=1}^{m} < Tv_k, w_j > w_j ,$$

d. h. $d_{jk} =< Tv_k, w_j >$. Für $j \leq r$ (und $k \leq n$) ist

$$d_{jk} =< Tv_k, w_j >=< Tv_k, T\left(\frac{v_j}{\sqrt{\mu_j}}\right) >= \frac{1}{\sqrt{\mu_j}} < Tv_k, Tv_j >= \sqrt{\mu_j}\delta_{jk} = \sqrt{\mu_k}\delta_{jk} .$$

Weiter gilt für $j > r, k \leq r$

$$d_{jk} =< Tv_k, w_j >= \sqrt{\mu_k} < \frac{1}{\sqrt{\mu_k}} Tv_k, w_j >= \sqrt{\mu_k} < w_k, w_j >= 0$$

und für $j > r, k > r$ ist $T^*Tv_k = 0$, also

$$0 =< v_k, T^*Tv_k >=< Tv_k, Tv_k > ,$$

d. h. $Tv_k = 0$ und damit $d_{jk} =< Tv_k, w_j >= 0$. Insgesamt ist $\varphi_{M,N}(T) = (d_{jk})$ mit

$$d_{jk} := \begin{cases} \lambda_k & , \quad \text{falls } j = k,\ k = 1, \dots, r \\ 0 & , \quad \text{sonst} \end{cases} .$$

□

Bemerkung 18.9 Insbesondere ergibt sich aud S. 18.8, dass für jede Matrix $A \in \mathbb{K}^{m\times n}$ unitäre Matrizen $B \in \mathbb{K}^{m\times m}$ und $C \in \mathbb{K}^{n\times n}$ existieren mit

$$A = BDC^* \ ,$$

wobei $D = (d_{jk})$ wie in S.18.8 und $\lambda_1, \dots, \lambda_r$ die positiven Singulärwerte von A sind. Eine solche Zerlegung nennt man auch *Singulärwertzerlegung* von A.
(Denn: Für $C = (v_1, \dots, v_n) = (Ce_1, \dots, Ce_n)$ und $B = (w_1, \dots, w_m) = (Be_1, \dots, Be_n)$ gilt nach S. 18.8 für $k = 1, \dots, n$

$$ACe_k = Av_k = \left\{ \begin{array}{lll} \lambda_k w_k & , \text{ falls } & k \le r \\ 0 & , \text{ falls } & k > r \end{array} \right\} = \left\{ \begin{array}{lll} B\lambda_k e_k & , \text{ falls } & k \le r \\ B0 & , \text{ falls } & k > r \end{array} \right\} = BDe_k \ .$$

Also ist $AC = BD$ und damit, da C unitär ist, $A = BDC^{-1} = BDC^*$.)

19 Triangulierbarkeit und der Satz von Cayley-Hamilton

Wir betrachten im folgenden wieder allgemeine lineare Räume V über K allerdings meist über $K = \mathbb{K}$. Zunächst wollen wir untersuchen, welche $T \in L(V)$ triangulierbar sind, d. h. für eine geeignete Basis M ist die Matrix von V bzgl. M eine Dreiecksmatrix.

Definition 19.1 Es sei V ein linearer Raum über K, und es sei $T \in L(V)$. Ein Unterraum U von V heißt *invariant (unter T)*, falls $Tu \in U$ für alle $u \in U$ (d. h. $T(U) \subset U$).

Damit gilt

Satz 19.2 *Es sei V ein endlich-dimensionaler linearer Raum über K und $M = (v_1, \dots, v_n)$ eine Basis von V. Ferner sei $T \in L(V)$. Dann sind äquivalent*
a) $\varphi_M(T)$ ist eine obere Dreiecksmatrix,
b) $Tv_k \in < v_1, \dots, v_k >$ für jedes $k \in \{1, \dots, n\}$,
c) $< v_1, \dots, v_k >$ ist invariant unter T für jedes $k \in \{1, \dots, n\}$.

Beweis.
1. Die Äquivalenz von a) und b) ergibt sich direkt aus der Definition von $\varphi_M(T)$. ([Ü])
2. c) $\Rightarrow$ b) ist klar. Bleibt also noch b) $\Rightarrow$ c) zu zeigen. Es sei $k \in \{1, \dots, n\}$. Nach b) ist

$$Tv_1 \in < v_1 > \subset < v_1, \dots, v_k >, \dots, Tv_k \in < v_1, \dots, v_k > \ .$$

also auch

$$< Tv_1, \dots, Tv_k > \subset < v_1, \dots, v_k > \qquad (k = 1, \dots, n)$$

und damit

$$T(< v_1, \dots, v_k >) \subset < v_1, \dots, v_k > \qquad (k = 1, \dots, n)\ .$$

□

Definition 19.3 Es sei V ein endlich-dimensionaler linearer Raum über K, und es sei $T \in L(V)$. T heißt *triangulierbar*, falls eine Basis M von V so existiert, dass $\varphi_M(T)$ eine obere Dreiecksmatrix ist. Eine Matrix $A \in K^{n \times n}$ heißt *triangulierbar*, falls $x \mapsto Ax \in L(K^n)$ triangulierbar ist.

Satz 19.4 *Es sei V ein n-dimensionaler linearer Raum über $\mathbb{K}$, und es sei $T \in L(V)$. Dann ist T genau dann triangulierbar, wenn P_T in Linearfaktoren zerfällt (d. h. $P_T(\lambda) = \prod_{j=1}^{n} (\lambda_j - \lambda)$ für gewisse $\lambda_1, \dots, \lambda_n \in \mathbb{K}$). Insbesondere ist im Falle $\mathbb{K} = \mathbb{C}$ jedes $T \in L(V)$ triangulierbar.*

Beweis.
1. Ist T triangulierbar, d. h. existiert eine Basis M von V so, dass $\varphi_M(T) = (r_{jk})$ eine obere Dreiecksmatrix ist, so ist auch $\varphi_M(T) - \lambda E$ obere Dreiecksmatrix, und es gilt

$$P_T(\lambda) = P_{\varphi_M(T)}(\lambda) = \prod_{j=1}^{n} (r_{jj} - \lambda) =: \prod_{j=1}^{n} (\lambda_j - \lambda)$$

mit $r_{jj} =: \lambda_j$.
2. Wir zeigen die Rückrichtung per Induktion nach $n = \dim(V)$. Für $n = 1$ ist nichts zu zeigen.
Es sei $\dim(V) = n$ und $T \in L(V)$ mit

$$P_T(\lambda) = \prod_{j=1}^{n} (\lambda_j - \lambda) \qquad (\lambda \in \mathbb{K}) \, .$$

Dann sind $\lambda_1, \dots, \lambda_n$ Eigenwerte von T (S. 12.9). Es sei $v_1 \neq 0$ ein Eigenvektor zu λ_1, und es sei $U :=< v_1 >$. Wir wählen eine Unterraum W von V so, dass

$$U \oplus W = V$$

(S. 5.20) und setzen $S := P_{W,U} \circ T_{|W} : W \to W$. Dann ist $S \in L(W)$ (vgl. B. 15.2).
Wir zeigen: $P_S(\lambda) = \prod_{j=2}^{n} (\lambda_j - \lambda)$.
(Denn: Es sei $N = (v_2, \dots, v_n)$ eine Basis von W. Dann ist $(v_1, \dots, v_n) =: M$ eine Basis von V und es gilt (da $Tv_1 = \lambda_1 v_1$)

$$A = (a_{jk}) := \varphi_M(T) = \begin{bmatrix} \lambda_1 & * & \dots & * \\ 0 & & & \\ \vdots & & B & \\ \vdots & & & \\ 0 & & & \end{bmatrix}$$

mit $B \in \mathbb{K}^{(n-1)\times(n-1)}$. Für $k = 2, \dots, n$ gilt

$$\begin{aligned} Sv_k &= P_{W,U}(Tv_k) = P_{W,U} \left(\sum_{j=1}^{n} a_{jk} v_j \right) = \\ &= \sum_{j=1}^{n} a_{jk} P_{W,U} v_j = \sum_{j=2}^{n} a_{jk} v_j \, , \end{aligned}$$

d. h. es ist $B = \varphi_N(S)$. Damit gilt (Entwicklung nach 1. Spalte)

$$P_T(\lambda) = \det(A - \lambda E) = (\lambda_1 - \lambda)\det(B - \lambda E) = (\lambda_1 - \lambda)P_S(\lambda)$$

und deshalb

$$P_S(\lambda) = \prod_{j=2}^{n}(\lambda_j - \lambda) \qquad (\lambda \in \mathbb{K})) \; .$$

Nach Induktionsvoraussetzung (beachte $\dim(W) = n - 1$) können wir deshalb N so wählen, dass $B = \varphi_N(S)$ eine obere Dreiecksmatrix ist. Dann ist auch $A = \varphi_M(T)$ eine obere Dreiecksmatrix.

3. Ist $\mathbb{K} = \mathbb{C}$, so zerfällt P_T nach dem Fundamentalsatz der Algebra stets in Linearfaktoren, also ist jedes $T \in L(V)$ triangulierbar. □

Definition 19.5 Es seien V ein linearer Raum über K, und es seien $T, T_1, \ldots, T_k \in L(V)$.

1. Wir setzen $\prod_{\nu=1}^{k} T_\nu := T_1 \circ \ldots \circ T_k$ sowie $T^k := \prod_{\nu=1}^{k} T$ für $k \in \mathbb{N}$ und $T^0 := I$. Entsprechend setzen wir für $A, A_1, \ldots, A_k \in K^{n\times n}$ auch $\prod_{\nu=1}^{k} A_\nu = A_1 \cdot \ldots \cdot A_k$; $A^k := \prod_{\nu=1}^{k} A$ für $k \in \mathbb{N}$ sowie $A^0 := E = E_n$.

2. Ist $K = \mathbb{K}$ und $P \in \Pi_{\mathbb{K}}$, $P(z) = \sum_{\nu=0}^{n} a_\nu z^\nu$, so setzen wir

$$P(T) := \sum_{\nu=0}^{n} a_\nu T^\nu \in L(V) \; .$$

Entsprechend setzen wir für $A \in \mathbb{K}^{n\times n}$

$$P(A) := \sum_{\nu=0}^{n} a_\nu A^\nu \in \mathbb{K}^{n\times n} \; .$$

Bemerkung 19.6 Sind $P, Q \in \Pi_{\mathbb{K}}$, so gilt für alle $T \in L(V)$

$$(PQ)(T) = P(T) \circ Q(T)$$

und damit auch

$$P(T) \circ Q(T) = Q(T) \circ P(T)$$

(Denn: Ist $P(z) = \sum_{\mu=0}^{n} a_\mu z^\nu$, $Q(z) = \sum_{\nu=0}^{m} b_\nu z^\nu$, so ist mit $a_\mu = b_\nu := 0$ für $\mu > n$ bzw. $\nu > m$

$$\begin{aligned}(PQ)(T) &= \sum_{k=0}^{n+m} T^k \sum_{\mu=0}^{k} a_\mu b_{k-\mu} = \sum_{k=0}^{n+m}\sum_{\mu=0}^{k} a_\mu T^\mu \circ b_{k-\mu}T^{k-\mu} \\ &= \left(\sum_{\mu=0}^{n} a_\mu T^\mu\right) \circ \left(\sum_{\nu=0}^{m} b_\nu T^\nu\right) = P(T) \circ Q(T)) \; .\end{aligned}$$

Damit gilt folgender bemerkenswerte Satz.

Satz 19.7 (Cayley-Hamilton) *Es sei V ein endlich-dimensionaler linearer Raum über $\mathbb{K}$, und es sei $T \in L(V)$ triangulierbar. Ist P_T das charakteristische Polynom von T, so gilt*

$$P_T(T) = 0 \, .$$

Beweis.
Es sei $M = (v_1, \ldots, v_n)$ eine Basis von V so, dass $\varphi_M(T) = R = (r_{jk})$ eine obere Dreiecksmatrix ist. Es reicht, zu zeigen

$$P_T(T)v_k = 0 \qquad (k = 1, \ldots, n) \, .$$

Es gilt

$$P_T(z) = P_{\varphi_M(T)}(z) = \prod_{j=1}^{n}(r_{jj} - z) = \prod_{j=1}^{n}(\lambda_j - z)$$

mit $\lambda_j = r_{jj}$. Also ist nach B. 19.6 für $k = 1, \ldots, n$ (mit $\prod_{\emptyset} := I$)

$$\begin{aligned} P_T(T) &= \prod_{j=1}^{n}(\lambda_j T^0 - T^1) = (-1)^n \prod_{j=1}^{n}(T - \lambda_j I) \\ &= (-1)^n \prod_{j=k+1}^{n}(T - \lambda_j I) \prod_{j=1}^{k}(T - \lambda_j I) \, . \end{aligned}$$

Also reicht es, für $k = 1, \ldots, n$

$$\left(\prod_{j=1}^{k}(T - \lambda_j I) \right) v_k = 0$$

zu zeigen. Für $k = 1$ gilt $Tv_1 = \lambda_1 v_1$, d. h. $(T - \lambda_1 I)v_1 = 0$.
Es gelte die Behauptung für $m = 1, \ldots, k-1$, d. h.

$$\prod_{j=1}^{m}(T - \lambda_j I)v_m = 0 \qquad (m = 1, \ldots, k-1) \, .$$

Dann ist auch $\prod_{j=1}^{k-1}(T-\lambda_j I)v_m = \prod_{j=m+1}^{k-1}(T-\lambda_j I) \prod_{j=1}^{m}(T-\lambda_j I)v_m = 0$ für $m = 1, \ldots k-1$.
Aus $\varphi_M(T) = R = (r_{\mu k})$ mit $r_{\mu k} = 0$ für $\mu > k$ folgt

$$Tv_k = \sum_{\mu=1}^{k} r_{\mu k} v_\mu = \sum_{\mu=1}^{k-1} r_{\mu k} v_\mu + \lambda_k v_k$$

d. h.

$$(T - \lambda_k I)v_k = \sum_{\mu=1}^{k-1} r_{\mu k} v_\mu \ .$$

Also gilt auch

$$\prod_{j=1}^{k}(T - \lambda_j I)v_k = \prod_{j=1}^{k-1}(T - \lambda_j I)(T - \lambda_k I)v_k = \sum_{\mu=1}^{k-1} r_{\mu k} \prod_{j=1}^{k-1}(T - \lambda_j I)v_\mu = 0$$

d. h. die Behauptung gilt für k. Damit ist $P_T(T)v_k = 0$ für $k = 1, \dots, n$. □

Bemerkung 19.8 Im Falle $\mathbb{K} = \mathbb{C}$ ist nach S. 19.4 die Voraussetzung "T triangulierbar" stets erfüllt, d. h. $P_T(T) = 0$ gilt für alle $T \in L(V)$. Damit ergibt sich für alle $A \in \mathbb{C}^{n\times n}$

$$P_A(A) = 0 \ .$$

Ist $A \in \mathbb{R}^{n\times n}$, so lässt sich A natürlich auch als Matrix in $\mathbb{C}^{n\times n}$ auffassen. Also gilt auch hier

$$P_A(A) = 0 \ .$$

Hieraus folgt, dass die Aussage von S. 19.7, also $P_T(T) = 0$, auch im Falle $\mathbb{K} = \mathbb{R}$, ohne die Voraussetzung "T triangulierbar" richtig bleibt ([Ü]).

Beispiel 19.9 Es sei

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

Dann gilt

$$P_A(\lambda) = \det(A - \lambda E) = \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix} = \lambda^2 + 1 \ .$$

Also ist nach dem Satz von Cayley-Hamilton

$$P_A(A) = A^2 + E = 0 \ .$$

Tatsächlich rechnet man nach

$$A^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \ ,$$

d. h. $A^2 + E = 0$.

20 Jordan'sche Normalform

Definition 20.1 Es sei V ein linearer Raum über K, und es sei $T \in L(V)$. Ist λ ein Eigenwert von T, so heißt $v \in V$ *verallgemeinerter Eigenvektor* (oder auch *Hauptvektor*) (von T), falls

$$(T - \lambda I)^j v = 0$$

für ein $j \in \mathbb{N}$, d. h. $v \in \text{Kern}(T - \lambda I)^j$ für ein $j \in \mathbb{N}$. (Wie üblich sind *verallgemeinerte Eigenvektoren* einer Matrix $A \in K^{n \times n}$ über $x \mapsto Ax$ definiert.)

Beispiel 20.2 Es sei $T \in L(\mathbb{K}^3)$ mit

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_2 \\ 0 \\ x_3 \end{pmatrix} .$$

Dann sind $0, 1$ die Eigenwerte von T (denn: $\det(A - \lambda E) = (-\lambda)^2(1 - \lambda)$). Weiter ist e_1 Eigenvektor zu $\lambda = 0$, und e_3 ist Eigenvektor zu $\lambda = 1$. Außerdem sieht man: $\dim(\text{Kern}(T - \lambda I)) = 1$ für $\lambda = 0, 1$.
Es gilt

$$T^2 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ 0 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix} ,$$

also ist e_2 verallgemeinerter Eigenvektor zum Eigenwert $\lambda = 0$. Insbesondere sieht man, dass eine Basis - nämlich (e_1, e_2, e_3) - aus verallgemeinerten Eigenvektoren existiert.

Ist $T \in L(V)$ so gilt

$$\{0\} = \text{Kern}(T^0) \subset \text{Kern}(T^1) \subset \text{Kern}(T^2) \subset \dots .$$

Weiter kann man leicht zeigen:

Satz 20.3 *Es sei T ein linearer Raum über K, und es sei $T \in L(V)$.*
1. Ist $Kern(T^m) = Kern(T^{m+1})$ für ein $m \in \mathbb{N}_0$, so gilt $\{0\} = Kern(T^0) \subset \dots \subset Kern(T^m) = Kern(T^{m+1}) = Kern(T^{m+2}) = \dots$.
2. Ist $\dim(V) = n$, so gilt

$$Kern(T^n) = Kern(T^{n+1}) \; (= Kern(T^{n+2}) = \dots) .$$

Beweis.
1. Es sei $k \in \mathbb{N}$. Wir zeigen $\mathrm{Kern}(T^{m+k}) = \mathrm{Kern}(T^{m+k+1})$. Klar ist, dass "$\subset$" gilt. Umgekehrt sei $v \in \mathrm{Kern}(T^{m+k+1})$. Dann ist

$$0 = T^{m+k+1}v = T^{m+1}(T^k v) \,,$$

also ist

$$T^k v \in \mathrm{Kern}(T^{m+1}) = \mathrm{Kern}(T^m)$$

und damit

$$0 = T^m(T^k v) = T^{m+k}v \,.$$

Folglich ist $v \in \mathrm{Kern}(T^{m+k})$, d. h. "$\supset$" gilt auch.
2. Nach 1. reicht es, $\mathrm{Kern}(T^n) = \mathrm{Kern}(T^{n+1})$ zu zeigen. Angenommen, nicht. Dann ist nach 1.

$$\{0\} = \mathrm{Kern}(T^0) \subsetneqq \mathrm{Kern}(T^1) \subsetneqq \mathrm{Kern}(T^2) \subsetneqq \mathrm{Kern}(T^{n+1}) \,.$$

Hieraus folgt

$$\dim(\mathrm{Kern}(T^{n+1})) > \dim(\mathrm{Kern}(T^n)) > \dots \,,$$

also $\dim(\mathrm{Kern}(T^{n+1})) \geq n+1$. Da $\mathrm{Kern}(T^{n+1})$ ein Unterraum von V ist, ergibt sich ein Widerspruch. □

Damit erhält man sofort

Bemerkung und Definition 20.4 Es sei V ein n-dimensionaler linearer Raum , und es sei $T \in L(V)$. Ist λ ein Eigenwert von T, so ist V verallgemeinerter Eigenvektor genau dann, wenn $v \in \mathrm{Kern}(T-\lambda I)^n$ ist. Der Unterraum $\mathrm{Kern}(T-\lambda I)^n$ heißt *Hauptraum* zu λ (von T). Entsprechend heißt für eine Matrix $A \in K^{n\times n}$ mit Eigenwert λ der Unterraum $\mathrm{Kern}(A-\lambda E)^n$ *Hauptraum* zu λ (von A).
(Denn: Ist $v \in \mathrm{Kern}(T-\lambda I)^n$, so ist V verallgemeinerter Eigenvektor. Ist umgekehrt v ein verallgemeinerter Eigenvektor von T, so ist

$$v \in \mathrm{Kern}(T-\lambda I)^j$$

für ein $j \in \mathbb{N}$. Aus S. 20.3 ergibt sich dann auch $v \in \mathrm{Kern}(T-\lambda I)^n$).

Bemerkung 20.5 In Analogie zu S. 20.3 gilt für Bilder: Ist $T \in L(V)$, so ist

$$V = \mathrm{Bild}(T^0) \supset \mathrm{Bild}(T^1) \supset \mathrm{Bild}(T^2) \supset \cdots \,.$$

Außerdem gilt im Falle $\dim(V) = n$:

$$\text{Bild}(T^n) = \text{Bild}(T^{n+1}) = \text{Bild}(T^{n+2}) = \cdots$$

(Denn: Nach der Dimensionsformel (S. 7.4) und S. 20.3 ist

$$\dim(\text{Bild}(T^n)) = \dim(\text{Bild}(T^{n+1})) = \cdots ,$$

und damit ist nach der vorherigen Inklusionskette

$$\text{Bild}(T^n) = \text{Bild}(T^{n+1}) = \cdots .)$$

Wir kommen nun zu einem zentralen Satz über Haupträume

Satz 20.6 *Es sei V ein n-dimensionaler linearer Raum über $\mathbb{K}$, und es sei $T \in L(V)$ triangulierbar. Ferner seien $\lambda_1, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte von T. Ist mit $k_1, \ldots, k_m \in \{1, \ldots, n\}$*

$$P_T(\lambda) = \prod_{j=1}^{m} (\lambda_j - \lambda)^{k_j} ,$$

(die Zahl k_j heißt algebraische Vielfachheit von λ_j), so gilt

$$k_j = \dim(\mathit{Kern}(T - \lambda_j I)^n) .$$

Beweis.
Ist $\varphi_M(T) = R = (r_{jk})$ mit $r_{jk} = 0$ für $j > k$, so ist

$$P_T(\lambda) = P_{\varphi_M(T)}(\lambda) = \det(R - \lambda E) = \prod_{\ell=1}^{n} (r_{\ell\ell} - \lambda) = \prod_{j=1}^{m} (\lambda_j - \lambda)^{k_j}$$

für gewisse $k_j \in \{1, \ldots, n\}$ wobei $\lambda_1, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte von T sind. Es ist also zu zeigen:
Für $j = 1, \ldots, m$ steht die Zahl λ_j genau $\dim(\text{Kern}(T - \lambda_j I)^n)$-mal auf der Diagonalen von R.
Wir beweisen die Behauptung durch Induktion nach $n = \dim(V)$. Für $n = 1$ ist die Behauptung klar.
Für ein $n \in \mathbb{N}, n \geq 2$ sei $(v_1, \ldots, v_n)$ eine Basis von V so, dass $\varphi_M(T)$ obere Dreiecksmatrix ist, d. h.

$$R = \varphi_M(T) = \begin{bmatrix} \mu_1 & & & \\ & \ddots & * & \\ 0 & & \mu_{n-1} & \\ & & & \mu_n \end{bmatrix} = (r_{jk})_{j,k=1,\ldots,n} .$$

Es sei $j \in \{1, \dots, m\}$. O. E. sei $\lambda_j = 0$ (sonst betrachte $(T - \lambda_j I)$ statt T im folgenden). Ferner sei $U := < v_1, \dots, v_{n-1} >$.
Dann ist U invariant unter T (S. 19.2), also $S = T_{|U} \in L(U)$. Weiter ist

$$\tilde{R} = \begin{bmatrix} \mu_1 & & \\ & \ddots & * \\ 0 & & \mu_{n-1} \end{bmatrix} = (r_{jk})_{j,k=1,\dots,n-1} \ .$$

die Matrix von S bzgl. $(v_1, \dots, v_{n-1})$. Nach Induktionsvoraussetzung erscheint 0 auf der Diagonale von $\tilde{R}$ genau $\dim(\mathrm{Kern}(S^{n-1})) \in \{0, \dots, n-1\}$ mal. Da $\dim(U) = n-1$ ist, folgt aus S. 20.3 $\mathrm{Kern}(S^{n-1}) = \mathrm{Kern}(S^n)$, also erscheint 0 auf der Diagonale von $\tilde{R}$ auch $\dim(\mathrm{Kern}(S^n))$ mal.

1. Fall: $\mu_n \neq 0$. Wir zeigen: Dann ist $\mathrm{Kern}(T^n) \subset U$
(Denn: Es gilt

$$\varphi_M(T^n) = R^n = \begin{bmatrix} \mu_1^n & & & \\ & \ddots & & \\ & & \mu_{n-1}^n & \\ & & & \mu_n^n \end{bmatrix} =: \left(r_{jk}^{(n)}\right)$$

also ist

$$T^n v_n = \sum_{j=1}^{n-1} r_{jn}^{(n)} v_j + \mu_n^n v_n =: u + \mu_n^n v_n$$

für ein $u \in U$. Es sei $v \in \mathrm{Kern}(T^n)$. Dann ist (da $U \oplus < v_n >= V$)

$$v = \tilde{u} + a v_n$$

für ein $\tilde{u} \in U, a \in K$. Also gilt

$$0 = T^n v = T^n \tilde{u} + a T^n v_n = \underbrace{T^n \tilde{u} + a u}_{\in U} + a \mu_n^n v_n \ .$$

Hieraus folgt $a \mu_n^n v_n = 0$ (da $V = U \oplus < v_n >$), also $a = 0$, d. h. $v = \tilde{u} \in U$. Folglich ist $\mathrm{Kern}(T^n) \subset U$).
Aus $\mathrm{Kern}(T^n) \subset U$ folgt $\mathrm{Kern}(S^n) = \mathrm{Kern}((T_{|U})^n) = \mathrm{Kern}(T^n)$. Also erscheint 0 auf der Diagonalen von R genau $\dim(\mathrm{Kern}(S^n)) = \dim(\mathrm{Kern}(T^n))$ mal, d. h. $k_j = \dim(\mathrm{Kern}(T^n))$, wie behauptet.
2. Fall: $\mu_n = 0$. Wir zeigen:

$$\dim(\mathrm{Kern}(T^n)) = \dim(\mathrm{Kern}(S^n)) + 1 \ ,$$

was impliziert, dass 0 auf der Diagonalen von R genau $\dim(\mathrm{Kern}(T^n))$ mal die 0 steht, wie behauptet.
Nach der Dimensionsformel für Unterräume (S. 5.18) gilt

$$\begin{aligned}\dim(\mathrm{Kern}(T^n)) &= \dim(U \cap \mathrm{Kern}(T^n)) + \dim(U + \mathrm{Kern}(T^n)) - \dim(U) \\ &= \dim(\mathrm{Kern}(S^n)) + \dim(U + \mathrm{Kern}(T^n)) - (n-1) \ .\end{aligned}$$

Wir zeigen: Es existiert ein $w \in \mathrm{Kern}(T^n) \setminus U$.
Dazu machen wir den Ansatz

$$w = u - v_n$$

mit einem $u \in U$. Dann ist jedenfalls $w \notin U$. Wie könnte u aussehen?
Es gilt

$$T^n(u - v_n) = T^n u - T^n v_n \ ,$$

also existiert ein u (und damit ein w) wie gewünscht, falls $T^n u = T^n v_n$, d. h. , falls $T^n v_n \in \mathrm{Bild}((T^n)_{|U}) = \mathrm{Bild}(S^n)$. Da $R = \varphi_M(T)$ ist, gilt

$$Tv_n = \sum_{j=1}^{n-1} r_{jn} v_j + \mu_n v_n = \sum_{j=1}^{n-1} r_{jn} v_j \in U \ ,$$

also

$$T^n v_n = T^{n-1}(Tv_n) \in \mathrm{Bild}((T^{n-1})_{|U}) = \mathrm{Bild}(S^{n-1}) = \mathrm{Bild}(S^n)$$

nach B. 20.5. Also existiert ein $w \in \mathrm{Kern}(T^n) \setminus U$. Hieraus ergibt sich

$$n = \dim(V) \geq \dim(U + \mathrm{Kern}(T^n)) > \dim(U) = n - 1 \ ,$$

also $\dim(U + \mathrm{Kern}(T^n)) = n$. Nach obiger Dimensionsformel ist

$$\dim(\mathrm{Kern}(T^n)) = \dim(\mathrm{Kern}(S^n)) + 1 \ .$$

□

Damit können wir folgenden Struktursatz für lineare Abbildungen beweisen (vgl. S. 12.13)

Satz 20.7 *Es seinen V ein n-dimensionaler linearer Raum über $\mathbb{K}$ und $T \in L(V)$. Sind $\lambda_1, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte von T mit den zugehörigen Haupträumen $U_1, \ldots, U_m$, so gilt :*

1. U_j ist invariant unter T $(j = 1, \ldots, m)$,
2. Ist T triangulierbar, so gilt

$$V = \bigoplus_{j=1}^{m} U_j \left(= \bigoplus_{j=1}^{m} Kern(T - \lambda_j I)^n \right) \ .$$

Beweis.
1. Nach B. 19.6 ist

$$(T-\lambda_j I)^n T = T(T-\lambda_j I)^n \ ,$$

also gilt für $u \in U_j$

$$(T-\lambda_j I)^n Tu = T(T-\lambda_j I)^n u = T0 = 0 \ ,$$

d. h. $Tu \in U_j$.
2. Nach S. 20.6 gilt

$$\sum_{j=1}^{m} \dim(U_j) = \sum_{j=1}^{m} k_j = n \ .$$

Es sei $U := \sum\limits_{j=1}^{m} U_j$. Wir zeigen $U = V$. Dann folgt 2. aus F. 5.19.

Mit U_j $(j = 1, \dots, m)$ ist auch $U = \sum\limits_{j=1}^{m} U_j$ invariant unter T. Also ist $S := T_{|U} \in L(U)$. Außerdem ist S triangulierbar.
(Denn: Ist $U = V$, so ist $S = T$ triangulierbar. Ist $U \neq V$ und ist $N = (v_1, \dots, v_r)$ eine Basis von U, so wählen wir Vektoren $v_{r+1}, \dots, v_n$ so, dass $M = (v_1, \dots, v_n)$ eine Basis von V ist.
Dann gilt für $A = (a_{jk}) = \varphi_M(T)$

$$Sv_k = Tv_k = \sum_{j=1}^{n} a_{jk} v_j = \sum_{j=1}^{r} a_{jk} v_j \qquad (k = 1, \dots, r) \ ,$$

also $a_{jk} = 0$ für alle (k, j) mit $k \leq r$, $j > r$, d. h.

$$A = \begin{bmatrix} B & C \\ \\ O & D \end{bmatrix}$$

mit $B \in \mathbb{K}^{r\times r}, C \in \mathbb{K}^{r\times(n-r)}, D \in K^{(n-r)\times(n-r)}$. Außerdem gilt $\varphi_N(S) = B$. Hieraus folgt ([Ü])

$$\begin{aligned} P_T(\lambda) &= P_{\varphi_M(T)}(\lambda) = \det(A - \lambda E_n) = \det(B - \lambda E_r)\det(D - \lambda E_{n-r}) \\ &= P_S(\lambda) \cdot \det(D - \lambda E_{n-r}) \ . \end{aligned}$$

Da T triangulierbar ist, hat P_T genau n Nullstellen incl. Vielfachheiten. Also haben $P_S \in \Pi_r$ r Nullstellen und $\det(D - \lambda E_{n-r}) \in \Pi_{n-r}$ $n - r$ Nullstellen. Damit zerfällt auch P_S in Linearfaktoren). Nach Definition hat S dieselben Eigenwerte wie T. Also gilt

$$P_S(\lambda) = \prod_{j=1}^{m} (\lambda_j - \lambda)^{\tilde{k_j}} \ .$$

Außerdem stimmen auch die Haupträume von S zu den Eigenwerten λ_j mit U_j überein, d. h.

$$\tilde{k_j} = \dim(U_j) \ .$$

Anwendung von S. 20.6 auf S ergibt

$$\dim(U) = \sum_{j=1}^{m} \dim(U_j) = n \ (= \dim(V))$$

also $U = V$. □

Als Konsequenz erhalten wir

Folgerung 20.8 Sind V ein endlich-dimensionaler linearer Raum über $\mathbb{K}$ und $T \in L(V)$ triangulierbar, so existiert eine Basis von V aus verallgemeinerten Eigenvektoren von T.
Denn: Man wähle Basen von U_j $(j = 1, \dots, m)$ und setze diese zusammen.

Definition 20.9 Es sei V in linearer Raum über K, und es sei $T \in L(V)$. Dann heißt T *nilpotent*, falls $T^m = 0$ für ein $m \in \mathbb{N}$ gilt. Entsprechend heißt ein $A \in K^{n\times n}$ *nilpotent*, falls $A^m = 0$ für ein $m \in \mathbb{N}$.

Beispiel 20.10 1. Für

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

git $A^2 = 0$, also ist A nilpotent.
2. Es sei $T : \Pi_n \to \Pi_n$ definiert durch

$$T(P) := P' \qquad (P \in \Pi_n) \ .$$

Dann ist $T \in L(\Pi_n)$ ([Ü]) und es gilt

$$T^{n+1}(P) = P^{(n+1)} = 0 \qquad (P \in \Pi_n) \ ,$$

also $T^{n+1} = 0$ d. h. T ist nilpotent.

Bemerkung 20.11 1. Ist V ein n-dimensionaler linerarer Raum, und ist $T \in L(V)$ nilpotent, so ist $T^n = 0$
(Denn: Ist T nilpotent, so ist jedes $v \in V$ verallgemeinerter Eigenvektor zum Eigenwert 0. Also ist $\mathrm{Kern}(T^n) = V$ nach S. 20.4 .)
2. Ist $T \in L(V)$, wobei V ein n-dmensionaler linearer Raum, und ist λ ein Eigenwert T, so ist $(T - \lambda I)_{|\mathrm{Kern}(T-\lambda I)^n}$ nilpotent.
(Denn: $U := \mathrm{Kern}(T-\lambda I)^n$ ist invariant unter T. Also gilt $((T-\lambda I)_{|U})^n = (T-\lambda I)^n_{|U} = 0$.)

Im folgenden Satz liegt der wesentliche Schritt zum Beweis der Jordan-Normalform.

Satz 20.12 *Es sei V ein n-dimensionaler linearer Raum über K. Ferner sei $N \in L(V)$ nilpotent und für $v \in V$, $v \neq 0$ sei*

$$m(v) := m_N(v) := \max\{m \in \mathbb{N}_0 : N^m v \neq 0\} .$$

Dann existieren Vektoren $v_1, \dots, v_k \in V$ mit folgenden Eigenschaften:
a) $(v_1, Nv_1, \dots, N^{m(v_1)}v_1, \dots, v_k, Nv_k, \dots, N^{m(v_k)}v_k)$ ist eine Basis von V,
b) $(N^{m(v_1)}v_1, \dots, N^{m(v_k)}v_k)$ ist eine Basis von Kern(N).

Beweis.
Wir zeigen die Behauptung per Induktion nach $n = \dim V$. Für $n = 1$ ist die Behauptung klar. Es sei $n \in \mathbb{N}$, und es sei $N \in L(V)$ nilpotent, wobei $\dim(V) = n$. Dann ist N nicht surjektiv (sonst wäre auch N^m surjektiv für alle m), also ist $\dim(\text{Bild}(N)) < n = \dim(V)$. Wir betrachten $\tilde{N} := N_{|\text{Bild}(N)}$. Dann ist $\tilde{N} \in L(\text{Bild}(N))$ und nilpotent. Nach Induktionsvoraussetzung existieren $u_1, \dots, u_j \in \text{Bild}(N)$ so, dass
(i) $(u_1, Nu_1, \dots, N^{m(u_1)}u_1, \dots, u_j, Nu_j, \dots, N^{m(u_j)}u_j)$ eine Basis von Bild(N)
und
(ii) $(N^{m(u_1)}u_1, \dots, N^{m(u_j)}u_j)$ eine Basis von $\text{Kern}(\tilde{N}) = \text{Kern}(N) \cap \text{Bild}(N)$
ist (man beachte dabei $m(u) = m_N(u) = m_{\tilde{N}}(u)$ für alle $u \in \text{Bild}(N)$).
Da $\{u_1, \dots, u_j\} \subset \text{Bild}(N)$ ist, existieren $v_1, \dots, v_j \in V$ mit $Nv_r = u_r$ für $r = 1, \dots, j$. Dann gilt $m(v_r) = m(u_r) + 1$ für $r = 1, \dots, j$ (beachte $u_r \neq 0$). Wir wählen einen Unterraum W von Kern(N) mit

$$\text{Kern}(N) = (\text{Kern}(N) \cap \text{Bild}(N)) \oplus W$$

(existiert nach S. 5.20) und eine Basis $(v_{j+1}, \dots, v_k)$ von W (dabei ist $k = \dim(\text{Kern}(N))$.
Aus $v_{j+1}, \dots, v_k \in \text{Kern}(N)$ folgt $m(v_r) = 0$ für $r = j+1, \dots, k$.
Wir zeigen, dass für die so konstruierten $v_1, \dots, v_k$ die Bedingungen a) und b) gelten.
zu a): Wir zeigen zunächst: Das System in a) ist linear unabhängig.
Dazu seien $a_{r,s} \in K$ mit

$$0 = \sum_{r=1}^{k} \sum_{s=0}^{m(v_r)} a_{r,s} N^s(v_r) .$$

Anwenden von N ergibt

$$\begin{aligned} 0 = N(0) &= \sum_{r=1}^{k} \sum_{j=0}^{m(v_r)} a_{r,s} N^s(Nv_r) = \sum_{r=1}^{j} \sum_{s=0}^{m(u_r)+1} a_{r,s} N^s(u_r) \\ &= \sum_{r=1}^{j} \sum_{s=0}^{m(u_r)} a_{r,s} N^s(u_r) . \end{aligned}$$

Also folgt aus (i)

$$a_{r,s} = 0 \quad \text{für} \quad r = 1, \dots, j \quad \text{und} \quad s = 0, \dots, m(v_r) - 1 \; .$$

Hieraus ergibt sich (beachte: $m(v_r) \neq 0$)

$$0 = \underbrace{\sum_{r=1}^{j} a_{r,m(v_r)} N^{m(v_r)}(v_r)}_{\in \mathrm{Kern}(N) \cap \mathrm{Bild}(N)} + \underbrace{\sum_{r=j+1}^{k} a_{r,0} v_r}_{\in W} \; .$$

Da die Summe von $\mathrm{Kern}(N) \cap \mathrm{Bild}(N)$ und W direkt ist, erhalten wir

$$0 = \sum_{r=1}^{j} a_{r,m(v_r)} N^{m(v_r)}(v_r) = \sum_{r=1}^{j} a_{r,m(v_r)} N^{m(u_r)}(u_r)$$

und

$$0 = \sum_{r=j+1}^{k} a_{r,0} v_r \; .$$

Aus (ii) ergibt sich

$$a_{r,m(v_r)} = 0 \quad \text{für} \quad r = 1, \dots, j$$

und da $(v_{j+1}, \dots, v_k)$ eine Basis von W ist, folgt auch

$$a_{r,0} = 0 \quad \text{für} \quad r = j+1, \dots, k \; .$$

Damit sind alle $a_{r,s} = 0$, d. h. das System in a) ist linear unabhängig.
Weiter impliziert (i)

$$\dim(\mathrm{Bild}(N)) = \sum_{r=1}^{j} (m(u_r) + 1) = \sum_{r=1}^{j} m(v_r) \; .$$

Das System in a) besteht also aus

$$\sum_{r=1}^{k} (m(v_r) + 1) = k + \sum_{r=1}^{j} m(v_r) = \dim(\mathrm{Kern}(N)) + \dim(\mathrm{Bild}(N))$$

d. h. nach S. 7.4 aus $\dim(V) = n$ Vektoren. Damit ist das System in a) eine Basis von V.
Zu b): Es gilt

$$(N^{m(v_1)} v_1, \dots, N^{m(v_k)} v_k) = (N^{m(u_1)} u_1, \dots, N^{m(u_j)} u_j, v_{j+1}, \dots, v_k) \; .$$

Also ist nach (ii) und der Definition von $(v_{j+1}, \dots, v_k)$ dieses System eine Basis von $\mathrm{Kern}(N)$. □

Beispiel 20.13 1. Es sei $V = \Pi_n$ und $N(P) := P'$ $(P \in \Pi_n)$ (vgl. B. 20.10). Für $v_1 \in \Pi_n$ mit $v_1(z) = x^n$, gilt dann $m(v_1) = n$ und

$$(v_1, Nv_1, \ldots, N^{m(v_1)}v_1) = (x^n, nx^{n-1}, \ldots, n!x^0)$$

ist eine Basis von $V = \Pi_n$. Außerdem ist $(N^{m(v_1)}v_1) = n!x^0)$ eine Basis von $\text{Kern}(N) = \Pi_0$.
2. Es sei $V = \mathbb{K}^4$ und $N(v) := Av$ mit

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

Dann ist A (d. h. N) nilpotent, denn es gilt $A^2 = 0$.
Für

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = e_2 , \quad v_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = e_4$$

gilt

$$\begin{aligned} Nv_1 = Ae_2 = e_1 \quad &, \quad Nv_2 = Ae_4 = e_3 \\ N^2v_1 = Ae_1 = 0 \quad &, \quad N^2v_2 = Ae_3 = 0, \end{aligned}$$

d. h. $m(v_1) = m(v_2) = 1$. Hier ist

$$(v_1, Nv_1, v_2, Nv_2) = (e_2, e_1, e_4, e_3) \quad \text{Basis von} \quad V = \mathbb{K}^4$$

und

$$(Nv_1, Nv_2) = (e_1, e_3) \quad \text{Basis von Kern}(N) .$$

Satz 20.14 *Es sei V ein linearer Raum über K, und es sei $T \in L(V)$. Weiter seinen $U_1, \ldots, U_m$ invariante Unterräume und es gelte $V = \bigoplus_{j=1}^{m} U_j$. Ist $M = (v_1, \ldots, v_n) = (v_{d_0}, \ldots, v_{d_1}, \ldots, v_{d_{m-1}+1}, \ldots, v_{d_m})$, wobei $(v_{d_{j-1}+1}, \ldots, v_{d_j})$ eine Basis von U_j ist (mit $d_0 := 0$ und $d_m := n$), so gilt*

$$A = \varphi_M(T) = \begin{bmatrix} B_1 & & & O \\ & B_2 & & \\ & & \ddots & \\ O & & & B_m \end{bmatrix}$$

wobei $B_j \in K^{(d_j-d_{j-1})\times(d_j-d_{j-1})}$ *die Matrix von* $T_{|U_j}$ *bzgl.* $(v_{d_{j-1}+1},\dots,v_{d_j})$ *ist (d. h. A hat Blockdiagonalgestalt).*

Beweis.
Ist $A=(a_{jk})$ so gilt für $j=1,\dots,m$ und $k=d_{j-1}+1,\dots,d_j$

$$Tv_k=\sum_{\mu=0}^{n}a_{\mu k}v_\mu=\sum_{\ell=1}^{m}\sum_{\mu=d_{\ell-1}+1}^{d_\ell}a_{\mu k}v_\mu=\sum_{\mu=d_{j-1}+1}^{d_j}a_{\mu k}v_\mu\ ,$$

d. h. $a_{\mu k}=0$ für $\mu\notin\{d_{j-1}+1,\dots,d_j\}$. Damit hat A eine Blockdiagonalgestalt wie behauptet. □

Folgerung 20.15 Es seien $V,N,v_1,\dots,v_k$ wie in S. 20.12. Ist M die Basis von V aus S. 20.12 in "absteigender Reihenfolge" d. h.

$$M=(N^{m(v_1)}v_1,\dots,Nv_1,v_1,\dots,N^{m(v_k)}v_k,\dots,v_k)\ ,$$

so gilt für $A=\varphi_M(N)$:

$$A=\begin{bmatrix}B_1 & & & O\\ & B_2 & & \\ & & \ddots & \\ O & & & B_k\end{bmatrix}$$

wobei für $m(v_r)>0$

$$B_r=\begin{bmatrix}0&1&0&\dots&0\\ \vdots&\ddots&\ddots&\ddots&\vdots\\ \vdots&&\ddots&\ddots&0\\ \vdots&&&\ddots&1\\ 0&\dots&\dots&\dots&0\end{bmatrix}\in K^{m(v_r)+1,m(v_r)+1}$$

bzw. $B=(0)$ für $m(v_r)=0$.
Denn: Für $r\in\{1,\dots,k\}$ wird der erste Vektor des Tupels $(N^{m(v_r)}v_r,\dots,Nv_r,v_r)$ durch N auf 0 abgebildet und die folgenden jeweils auf ihren Vorgänger. Insbesondere sind die Unterräume $W_r=<N^{m(v_r)}v_r,\dots,v_r>$ invariant unter N und es gilt $V=\bigoplus_{r=1}^{k}W_r$. Nach S.20.14 hat A Blockdiagonalgestalt wie oben angegeben mit gewissen

$B_1, \ldots, B_k$. Außerdem liefern die Matrizen

$$B_r = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \ldots & \ldots & \ldots & 0 \end{bmatrix}.$$

innerhalb von W_r gerade das entsprechende Abbildungsverhalten.

Definition 20.16 1. Es sei $\lambda \in \mathbb{K}$. Eine Matrix A der Form

$$A = \begin{bmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \ldots & 0 & \ddots & \lambda \end{bmatrix} \in \mathbb{K}^{n \times n}$$

für $n \geq 2$ bzw. $A = (\lambda)$ für $n = 1$ heißt *Jordan-Matrix*.
2. Es sei V ein endlich-dimensionaler linearer Raum über $\mathbb{K}$, und es sei $T \in L(V)$. Eine Basis M von V heißt *Jordan-Basis (für T)*, falls $\varphi_M(T)$ folgende Blockdiagonalgestalt hat:

$$\varphi_M(T) = \begin{bmatrix} A_1 & & & \\ & A_2 & 0 & \\ 0 & & \ddots & \\ & & & A_p \end{bmatrix}$$

wobei A_j für $j = 1, \ldots, p$ eine Jordan-Matrix ist.

Damit gilt schließlich

Satz 20.17 (Jordan'sche Normalform) *Es sei V ein endlich-dimensionaler linearer Raum über $\mathbb{K}$, und es sei $T \in L(V)$ triangulierbar. Dann existiert eine Jordan-Basis für T aus Hauptvektoren von T.*

Beweis.
Es sei $T \in L(V)$ triangulierbar, und es seien $\lambda_1, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte von T. Dann gilt nach S. 20.7

$$V = \bigoplus_{j=1}^{m} U_j$$

mit $U_j := \text{Kern}(T - \lambda_j I)^n$ und die U_j sind invariant. Außerdem ist $(T - \lambda_j I)_{|U_j}$ nilpotent für $j = 1, \ldots, m$ nach B. 20.11.2. Also existiert nach F. 20.15 für $j = 1, \ldots, m$ eine Jordan-Basis für $(T - \lambda_j I)_{|U_j}$ und diese ist dann auch Jordan-Basis für $T_{|U_j} = (T - \lambda_j I)_{|U_j} + \lambda I_{U_j}$ (bestehend aus Hauptvektoren zum Eigenwert λ_j). Setzt man diese Basen zusammen, so erhält man nach S.20.14 eine Jordan-Basis für T aus Hauptvektoren von T. □

21 Elementare Zahlentheorie

Wir starten mit einigen Ergebnissen aus der elementaren Zahltentheorie.

Definition 21.1 Es seien $a, b \in \mathbb{Z} = \{\text{ganze Zahlen}\}$. Man sagt, a sei *Teiler von* b (bzw. a *teilt* b) (kurz: $a|b$), falls $b/a \in \mathbb{Z}$ gilt, d. h. falls ein $q \in \mathbb{Z}$ existiert mit $b = qa$.

Es gilt damit

Satz 21.2 .
1. Für alle $a \in \mathbb{Z}$ ist $\pm 1|a$ und $\pm a|a$.
2. Sind $a, b \in \mathbb{Z}$ mit $a|b$, $b|c$, so ist $a|c$.
3. Sind $a, b_j \in \mathbb{Z}$, $j = 1, \dots, n$, mit $a|b_j$ für $j = 1, \dots, n$, so ist

$$a| \sum_{j=1}^{n} c_j b_j \qquad \text{für alle} \quad c_1, \dots, c_n \in \mathbb{Z} \, .$$

Beweis. [Ü]

Definition 21.3 Es seien $a, b \in \mathbb{Z}$ mit $a \neq 0$ oder $b \neq 0$. Dann heißt

$$d := \max\{k \in \mathbb{N} : k|a \text{ und } k|b\}$$

größter gemeinsamer Teiler von a und b (kurz: $ggT(a, b)$).
(Für $a = b = 0$ setzen wir $ggT(0, 0) := 0$.) Im Falle $ggT(a, b) = 1$ heißen a und b *teilerfremd*.

Satz 21.4 (Division mit Rest) *Es seien $a, b \in \mathbb{Z}, b \neq 0$. Dann existieren $q, r \in \mathbb{Z}$ mit*

$$a = qb + r \qquad \text{und} \quad r \in \{0, \dots, |b| - 1\} \, .$$

Beweis.
Da $b \neq 0$, ist, gilt

$$L := \mathbb{N}_0 \cap \{a - xb : x \in \mathbb{Z}\} \neq \emptyset \, .$$

Für $r := \min L$ und q so, dass $a - qb = r$ gilt dann die Behauptung. □

Zur Abkürzung schreiben wir im folgenden für Mengen $A, B \subset \mathbb{C}$ und $\lambda, \mu \in \mathbb{C}$

$$\begin{aligned} \lambda A &:= \{\lambda a : a \in A\} \, , \; \mu \pm \lambda A = \{\mu \pm \lambda a : a \in A\} \\ \mu B \pm \lambda A &= \{\mu b \pm \lambda a : a \in A, b \in B\} \, . \end{aligned}$$

Es gilt damit

Satz 21.5 *Es seien $a, b \in \mathbb{Z}$, und es sei $d := ggT(a, b)$. Dann ist*

$$\{ax + by : x, y \in \mathbb{Z}\} = a\mathbb{Z} + b\mathbb{Z} = d\mathbb{Z} .$$

Insbesondere sind a, b genau dann teilerfremd, wenn gilt

$$a\mathbb{Z} + b\mathbb{Z} = \mathbb{Z} .$$

Beweis.
Ist $a = 0$ oder $b = 0$, so sieht man leicht, dass die Behauptung gilt. Es sei also $a \neq 0$ und $b \neq 0$, und es sei $m := \min(\mathbb{N} \cap L)$ wobei

$$L := a\mathbb{Z} + b\mathbb{Z} = ax + by : x, y \in \mathbb{Z}\} .$$

Aus $d|a$ und $d|b$ folgt $d|ax + by$ für alle $x, y \in \mathbb{Z}$ (S. 21.2.3), also gilt

$$L \subset d\mathbb{Z}$$

und insbesondere $d|m$.
Andererseits gilt $m|a$.
(Denn: Es gilt $m\mathbb{Z} \subset L$ und $a \in L$, also auch $a - m\mathbb{Z} \subset L$. Division von a durch m mit Rest (nach S. 21.4) kann keinen Rest $r \neq 0$ ergeben (wäre $a = qm + r$ mit $r \in \{1, \ldots, m-1\}$ so wäre $r = a - qm \in a - m\mathbb{Z} \subset L$ im Widerspruch zur Minimalität von m). Also ist m Teiler von a.)
Entsprechend sieht man $m|b$. Also ist $m \leq d$ nach Definition von d. Mit $d|m$ folgt $m = d$, also

$$d\mathbb{Z} = m\mathbb{Z} \subset L .$$

Insgesamt ergibt sich $L = d\mathbb{Z}$. □

Folgerung 21.6 Es seien $a, b, c \in \mathbb{Z}$ mit $ggT(a, b) = 1$. Dann gilt
1. Es existieren $x, y \in \mathbb{Z}$ mit $xa + yb = 1$.
2. Ist $a|bc$, so ist $a|c$.
3. Ist $a|c$ und $b|c$, so ist $ab|c$.

Denn:
1. Die erste Aussage ergibt sich direkt aus der letzten Aussage von S. 21.5.
2. Nach 1. existieren $x, y \in \mathbb{Z}$ mit $xa + yb = 1$. Also ist $c = xac + ybc$. Da $a|bc$ und $a|a$ ist $a|c$ nach S. 21.2.3.
3. Wie in 2. ist $c = xac+ybc$. Da $c = ua = bv$ für gewisse $u, v \in \mathbb{Z}$, ist $c = xabv+ybau = ab(xv + yu)$. Also ist $ab|c$.

Definition 21.7 Es sei $p \in \mathbb{N}, p \geq 2$. Dann heißt p *Primzahl* falls p nur die Teiler ± 1 und $\pm p$ besitzt.

Bemerkung 21.8 Sind $a, b \in \mathbb{Z}$ und ist p eine Primzahl, so folgt aus $p|ab$ schon $p|a$ oder $p|b$. (Denn: Ist p kein Teiler von a, so ist $ggT(a,p) = 1$. Also ist $p|b$ nach F. 21.6.2.)

Satz 21.9 (Primfaktorzerlegung; Fundamentalsatz der Arithmetik) *Es sei $n \in \mathbb{N}, n \geq 2$. Dann existieren Primzahlen $p_1, \ldots, p_k$ mit $n = \prod_{j=1}^{k} p_j$ und diese Zerlegung ist bis auf die Reihenfolge der Faktoren eindeutig.*

Beweis.
1. Existenz: (Induktion nach n)
$\underline{n = 2}$: Für $n = 2$ ist die Behauptung klar.
$\underline{n - 1 \to n \ (n \geq 3)}$: Ist n eine Primzahl, so ist nichts zu zeigen. Es sei also n keine Primzahl. Dann existiert ein Primteiler p von n, d. h. es existiert eine Primzahl p mit $p|n$. (Denn: Man wähle $p := \min\{k > 1 : k|n\}$. Dann ist p eine Primzahl, denn sonst hätte p einen Teiler p_0 mit $1 < p_0 < p$, und damit wäre p_0 auch Teiler von n im Widerspruch zur Minimalität von p.)
Also ist $n = pq$ für ein $q \in \mathbb{N}$ mit $1 < q < n$. Nach Induktionsvoraussetzung ist q Produkt aus Primfaktoren, und damit gilt dasselbe für n.
2. Eindeutigkeit: Wir zeigen per Induktion nach k:
Ist n Produkt aus k Primzahlen, so besteht jede Darstellung von n als Produkt aus Primzahlen aus k Faktoren, und diese Faktoren sind bis auf die Reihenfolge dieselben.
$\underline{k = 1}$: Ist $n = p$ eine Primzahl, so gilt die Behauptung nach Definition.
$\underline{k - 1 \to k \ (k \geq 2)}$: Es sei

$$n = \prod_{j=1}^{k} p_j = \prod_{j=1}^{m} q_j$$

mit Primzahlen p_j und q_j. Aus $p_k | \prod_{j=1}^{m} q_j$ folgt aus B. 21.8 (induktiv), dass $p_k | q_{j_0}$ für ein $j_0 \in \{1, \ldots, m\}$, also $p_k = q_{j_0}$ (da q_{j_0} Primzahl). Nach Umnummerierung können wir o. E. $q_{j_0} = q_m$ annehmen. Dann ist

$$n = \prod_{j=1}^{k-1} p_j = \prod_{j=1}^{m-1} q_j \ .$$

Nach Induktionsvoraussetzung ist $k-1 = m-1$ und die Faktoren $p_1, \ldots, p_{k-1}$ stimmen bis auf die Reihenfolge mit den $q_1, \ldots, q_{k-1}$ überein. Nach obiger Konstruktion gilt die Behauptung damit auch für k. □

Definition 21.10 1. Es seien $a, b \in \mathbb{Z}, m \in \mathbb{N}_0$. Dann heißen a und b *kongruent modulo* m, falls

$$m|(a-b) \ .$$

Wir schreiben dann

$$a \equiv b \bmod m \ .$$

Bemerkung 21.11 1. Man sieht leicht, dass durch

$$a \sim b :\Leftrightarrow a \equiv b \bmod m$$

eine Äquivalenzrelation auf $\mathbb{Z}$ gegeben ist. Die Äquivalenzklassen werden *Restklassen (modulo m)* genannt. Wir schreiben kurz für die Restklasse mit dem Repräsentanten $a \in \mathbb{Z}$ (d. h. für die Menge $\{a + km : k \in \mathbb{Z}\}$)

$$[a] := [a]_m := a \bmod m := a + m\,\mathbb{Z}$$

und außerdem setzen wir

$$\mathbb{Z}_m := \mathbb{Z}/m\mathbb{Z} := \{[a] : a \in \mathbb{Z}\} \ ,$$

d. h. $\mathbb{Z}_m$ ist die Restklassenmenge.
Es gilt dabei

$$\mathbb{Z}_m = \{[0], [1], \ldots [m-1]\}$$

für $m \neq 0$ (und $\mathbb{Z}_0 = \mathbb{Z}$).
2. Auf $\mathbb{Z}_m$ kann man repräsentantenweise eine Addition und eine Multiplikation einführen:

$$[a] + [b] := [a + b]$$

und

$$[a] \cdot [b] = [a \cdot b]$$

für $a, b \in \mathbb{Z}$. Dabei ist wichtig, dass die Operationen unabhängig von der Auswahl der Repräsentanten sind.
(Es gilt etwa für die Multiplikation: Ist $[a] = [\tilde{a}], [b] = [\tilde{b}]$, so ist $m|(a-\tilde{a})$ und $m|(b-\tilde{b})$. Also ist nach S. 21.2.3 auch $m|(a-\tilde{a})b+\tilde{a}(b-\tilde{b})$, d. h. $m|(ab-\tilde{a}\tilde{b})$. Also ist $[ab] = [\tilde{a}\tilde{b}]$.)

3. Es sei $m \in \mathbb{N}$ und $\zeta := e^{2\pi i/m}$ eine m-te Einheitswurzel ($\zeta^m = 1$). Dann ist durch $[a] = [a]_m \mapsto \zeta^a$ eine bijektive Abbildung von $\mathbb{Z}_m$ nach $\{\zeta^a : a \in \mathbb{Z}\} = \{\zeta^a : a = 0, 1, \ldots, m-1\}$ gegeben. Dabei entsprechen die Restklassenoperationen aus 2. gewissen

Operationen in $\mathbb{C}$, nämlich $[a]+[b]$ entspricht ζ^{a+b} und $[a][b]$ entspricht $\zeta^{a\cdot b}$.
Ist etwa $m=4$, so gilt

$$[2+3] = 5 \bmod 4 = [1]$$

sowie

$$[2\cdot 3] = 6 \bmod 4 = [2]$$

und entsprechend

$$\zeta^{2+3} (= \zeta^2\zeta^3) = \zeta^5 = \zeta^1$$

sowie

$$\zeta^{2\cdot 3} (= (\zeta^2)^3) = (-1)^3 = -1 = \zeta^2 \ .$$

Aus den Rechenregeln im Körper $\mathbb{C}$ oder durch direktes Nachrechnen erhält man leicht

Satz 21.12 *Es sei $m \in \mathbb{N}$. Dann gilt*
1. $(\mathbb{Z}_m, +)$ ist eine abelsche Gruppe (mit Nullelement $[0]$ und $-[a] = [-a]$).
2. Für alle $[a], [b], [c] \in \mathbb{Z}_m$ ist

$$([a][b])[c] = [a]([b][c])$$

und

$$([a]+[b])[c] = [a][c] + [b][c]$$

sowie

$$[a][b] = [b][a] \ .$$

3. Für alle $[a] \in \mathbb{Z}_m$ gilt $[a][1] (= [1][a]) = [a]$
(d. h. $(\mathbb{Z}_m, +, \cdot)$ ist ein "kommutativer Ring mit Einselement").

Bemerkung 21.13 Im allgemeinen ist $(\mathbb{Z}_m, +, \cdot)$ kein Körper, d. h. i. a. existiert nicht für alle $[a] \in \mathbb{Z}_m \setminus \{[0]\}$ ein $[x]$ mit $[a][x] = [1]$. Zum Beispiel hat die Gleichung

$$[3]_6[x]_6 = [1]_6$$

keine Lösung $[x]_6$, denn $[3]_6[x]_6 \in \{[3]_6, [0]_6\}$ für alle $x \in \mathbb{Z}$. Außerdem können Nullteiler existieren, d. h. nichtverschwindende Lösungen der Gleichung $[a][b] = [0]$. So gilt etwa

$$[3]_6[2]_6 = [0]_6 \ .$$

Wir wollen jetzt zeigen, dass in gewissen Teilsystemen von $\mathbb{Z}_m$ solche Probleme nicht auftreten.

Bemerkung und Definition 21.14 Sind $a, b \in \mathbb{Z}, m \in \mathbb{N}$ so folgt aus $a \equiv b \bmod m$ und $ggT(a, m) = 1$ auch $ggT(b, m) = 1$.
(Denn: Nach S. 21.5 hat die Gleichung $ax + my = 1$ eine Lösung $(x, y) \in \mathbb{Z}^2$. Für $b = a + km$ gilt dann

$$bx + m(y - kx) = ax + kmx + my - kmx = 1 \ ,$$

d. h. $\{bu + mv : u, v \in \mathbb{Z}\} = \mathbb{Z}$. Also ist wieder nach S. 21.5 $ggT(b, m) = 1$.)
Man bezeichnet im Falle $ggT(a, m) = 1$ die Restklasse $[a]_m$ (repräsentantenunabhängig) als *prime Restklasse (mod m)*. Außerdem setzen wir

$$\mathbb{Z}_m^* := (\mathbb{Z}/m\mathbb{Z})^* = \{[a]_m : [a]_m \text{ prime Restklasse}\} \ .$$

Es gilt damit:

Satz 21.15 *Es sei $m \in \mathbb{N}$. Dann ist $(\mathbb{Z}_m^*, \cdot)$ (mit $\cdot$ wie in B. 21.11.2) eine abelsche Gruppe.*

Beweis.
Offensichtlich ist $[1] \in \mathbb{Z}_m^*$. Also genügt es nach S. 21.12, zu zeigen: mit $[a], [b] \in \mathbb{Z}_m^*$ ist auch $[a][b] \in \mathbb{Z}_m^*$ und die Gleichung $[a][x] = [1]$ hat für jedes $[a] \in \mathbb{Z}_m^*$ eine Lösung in $\mathbb{Z}_m^*$ (vgl. D. 2.1).
Es seien $a, b \in \mathbb{Z}, a, b$ teilerfremd zu m. Dann existieren nach S. 21.5 $(x, y) \in \mathbb{Z}^2$ und $(u, v) \in \mathbb{Z}^2$ mit

$$ax + my = 1, \qquad bu + mv = 1 \ .$$

Die erste Gleichung impliziert $ax = 1 \bmod m$, d. h. $[a][x] = [ax] = [1]$, und außerdem gilt $1 \in x\mathbb{Z} + m\mathbb{Z}$. Also ist $ggT(x, m) = 1$ nach S. 21.5. Damit ist $[x] \in \mathbb{Z}_m^*$. Weiter gilt

$$1 = (ax + my)(bu + mv) = (ab)(xu) + m(axv + ybu + myv) \ ,$$

d. h. $1 \in ab\mathbb{Z} + m\mathbb{Z}$ und damit ist $ggT(ab, m) = 1$ nach S. 21.5, also $[a][b] = [ab] \in \mathbb{Z}_m^*$. □

Beispiel 21.16 Für $m = 6$ gilt $ggT(1, 6) = ggT(5, 6) = 1$ und $ggT(0, 6) = 6$, $ggT(2, 6) = 2$, $ggT(3, 6) = 3$, $ggT(4, 6) = 2$, also ist

$$\mathbb{Z}_6^* = \{[1], [5]\} \ .$$

Nach S. 21.15 ist $(\mathbb{Z}_6^*, \cdot)$ eine Gruppe.

Ganz anders sind die Verhältnisse im Falle von Primzahlen m.

Folgerung 21.17 Es sei $p \in \mathbb{N}$ eine Primzahl. Dann gilt $ggT(a,p) = 1$ für alle $a \in \{1,\dots,p-1\}$, d. h.

$$\mathbb{Z}_p^* = \{[1],[2],\dots,[p-1]\} = \mathbb{Z}_p \setminus \{[0]\} .$$

Also ist nach S. 21.15 $\mathbb{Z}_p \setminus \{[0]\}$ eine abelsche Gruppe und folglich ist nach S. 21.12 $(\mathbb{Z}_p, +, \cdot)$ ein Körper (mit p Elementen) .

Eine nette Anwendung von Kongruenzen sind bekannte Teilbarkeitskriterien:

Satz 21.18 *Es sei $N \in \mathbb{N}$,*

$$N = \sum_{\nu=0}^{n} a_\nu 10^\nu \qquad \textit{mit } a_\nu \in \{0,\dots,9\} \textit{ für } \nu = 0,\dots,n .$$

Dann gilt

1. $3|N$ *genau dann wenn* $3| \sum_{\nu=0}^{n} a_\nu$

2. $9|N$ *genau dann wenn* $9| \sum_{\nu=0}^{n} a_\nu$

Beweis.
1. Aus $10 \equiv 1 \bmod 3$ folgt $10^\nu \equiv 1 \bmod 3$ (denn $[10^\nu]_3 = ([10]_3)^\nu = [1]$ für alle $\nu \in \mathbb{N}_0$). Also gilt

$$\begin{aligned} [N]_3 &= \left[\sum_{\nu=0}^{n} a_\nu 10^\nu\right]_3 = \sum_{\nu=0}^{n} [a_\nu]_3 [10^\nu]_3 = \sum_{\nu=0}^{n} [a_\nu]_3 \\ &= \left[\sum_{\nu=0}^{n} a_\nu\right]_3 , \end{aligned}$$

d. h. $N \equiv \sum_{\nu=0}^{n} a_\nu \bmod 3$. Insbesondere gilt also $3|N$ genau dann wenn $3| \sum_{\nu=0}^{n} a_\nu$.
2. Analog. □

Von grundlegender Bedeutung ist folgendes Ergebnis über simultane Kongruenzen

Satz 21.19 (Chinesischer Restsatz) *Es seien $m_1,\dots,m_n \in \mathbb{N}$ paarweise teilerfremd, und es seien $a_1,\dots,a_n \in \mathbb{Z}$. Dann existiert ein $x \in \mathbb{Z}$ so, dass*

$$x \equiv a_j \mod m_j \qquad \textit{für } j = 1,\dots,n .$$

Dabei ist $x \bmod m_1 \cdot \ldots \cdot m_n$ eindeutig bestimmt und mit x ist jedes $y \in [x]_{m_1 \cdot \ldots \cdot m_n}$ eine Lösung.

Beweis.
1. Eindeutigkeit: Sind $x, y \in \mathbb{Z}$ Lösungen des Systems von Kongruenzen, so gilt

$$m_j | (x - y) \qquad (j = 1, \dots, m) \; .$$

Da die m_j paarweise teilerfremd sind, ist dies äquivalent zu

$$\prod_{j=1}^{n} m_j | (x - y) \; .$$

(Denn: Induktiv ergibt sich aus F. 21.6.3: Ist $c \in \mathbb{Z}$, so folgt aus

$$m_j | c \qquad (j = 1, \dots, n)$$

auch

$$\prod_{j=1}^{n} m_j | c \; .$$

Man beachte dabei: Für $u, v \in \mathbb{Z}, \; k \in \mathbb{N}$ folgt aus $ggT(u, k) = ggT(v, k) = 1$ schon $ggT(uv, k) = 1$ nach dem Beweis zu S. 21.15).
Also ist $x \equiv y \bmod \prod_{j=1}^{n} m_j$.
2. Existenz: Wir betrachten die Abbildung

$$b : \mathbb{Z}_{\prod_{j=1}^{n} m_j} \to \prod_{j=1}^{n} \mathbb{Z}_{mj} \; ,$$

definiert durch

$$b\left([x]_{\prod_{j=1}^{n} m_j} \right) := ([x]_{m_1}, \dots, [x]_{m_n}) \; .$$

Sind $[x]_{\prod m_j}, [y]_{\prod m_j} \in \mathbb{Z}_{\prod m_j}$ mit $b([x]_{\prod m_j}) = b([y]_{\prod m_j})$, so gilt für $j = 1, \dots, n$

$$x \equiv y \qquad \bmod m_j \; .$$

Also ist $x \equiv y \bmod \prod m_j$ nach1., d. h. $[x]_{\prod m_j} = [y]_{\prod m_j}$. Damit ist b injektiv. Da

$$|\mathbb{Z}_{\prod m_j}| = \prod_{j=1}^{n} m_j = |\prod_{j=1}^{n} \mathbb{Z}_{m_j}|$$

gilt, ist b auch bijektiv, d. h. für beliebige $a_1, \dots, a_n \in \mathbb{Z}$ existiert ein $x \in \mathbb{Z}$ mit

$$[x]_{m_j} = [a_j]_{m_j} \qquad \text{bzw.} \qquad x \equiv a_j \bmod m_j \; .$$

□

Beispiel 21.20 Wir betrachten die Kongruenzen

$$\begin{aligned} x &\equiv 2 \mod 3 \\ x &\equiv 3 \mod 5 \\ x &\equiv 2 \mod 7\,. \end{aligned}$$

Dann ist $x = 23$ eine Lösung. Nach S. 21.19 ist $x \bmod 105 = 3 \cdot 5 \cdot 7$ eindeutig bestimmt. Außerdem ist jedes $y \in [x]_{105}$, d. h. jedes y mit $y = 23 + k \cdot 105$ für ein $k \in \mathbb{Z}$, ebenfalls Lösung.

Definition 21.21 Für $n \in \mathbb{N}, n \geq 2$, sei

$$\begin{aligned} \varphi(n) := |\mathbb{Z}_n^*| &= \text{Anzahl der primen Restklassen mod } n \\ &= \text{Anzahl der } a \in \mathbb{N} \text{ mit } a \leq n \text{ und } ggT(a,n) = 1 \end{aligned}$$

und für $n = 1$ setzen wir $\varphi(1) := 1$. Die Funktion $\varphi : \mathbb{N} \to \mathbb{N}$ heißt *Eulersche Funktion* (oder *Eulersche Phi-Funktion*).

Es gilt

Satz 21.22 *1. Die Eulersche Funktion ist multiplikativ, d. h. für alle teilerfremden $n, m \in \mathbb{N}$ gilt*

$$\varphi(nm) = \varphi(n)\varphi(m)\,.$$

2. Ist p eine Primzahl, so ist $\varphi(p^\nu) = p^\nu - p^{\nu-1}$ für alle $v \in \mathbb{N}$. Insbesondere ist $\varphi(p) = p - 1$.
3. Für beliebige $n \in \mathbb{N}$ ist

$$\varphi(n) = n \prod_{\substack{p \text{ prim} \\ p|n}} \left(1 - \frac{1}{p}\right)\,.$$

Beweis.
1. Es sei b wie im Beweis zu S. 21.19 (mit $m_1 = m, m_2 = n$). Dann ist $b^* := b_{|\mathbb{Z}_{mn}^*}$ eine bijektive Abbildung von $\mathbb{Z}_{mn}^*$ nach $\mathbb{Z}_m^* \times \mathbb{Z}_n^*$.
(Denn: Ist $x \in \mathbb{Z}$ mit $[x]_{mn} \in \mathbb{Z}_{mn}^*$, so gilt $ggT(mn, x) = 1$, also auch $ggT(m, x) = ggT(n, x) = 1$. Also ist $b([x]_{mn}) = ([x]_m, [x]_n) \in \mathbb{Z}_m^* \times \mathbb{Z}_n^*$.
Ist andererseits $([x]_m, [x]_n) \in \mathbb{Z}_m^* \times \mathbb{Z}_n^*$, so ist $ggT(x, m) = ggT(x, n) = 1$, also auch $ggT(x, mn) = 1$ (vgl. Beweis zu S. 21.15). Also ist $[x]_{mn} \in \mathbb{Z}_{mn}^*$. Da $b : \mathbb{Z}_{mn} \to \mathbb{Z}_m \times \mathbb{Z}_n$ bijektiv ist, ist damit auch $b^* : \mathbb{Z}_{mn}^* \to \mathbb{Z}_m^* \times \mathbb{Z}_n^*$ bijektiv).
Aus $\varphi(mn) = |\mathbb{Z}_{mn}^*| = |\mathbb{Z}_m^* \times \mathbb{Z}_n^*| = |\mathbb{Z}_m^*||\mathbb{Z}_n^*| = \varphi(n)\varphi(m)$ folgt 1.

2. Ist p prim und $\nu \in \mathbb{N}$, so sind die $a \in \{1, \dots, p^\nu\}$, die nicht teilerfremd zu p^ν sind, genau die $p^{\nu-1}$ Zahlen

$$p, 2p, 3p, \dots, p^{\nu-1}p \, .$$

Also ist $\varphi(p^\nu) = p^\nu - p^{\nu-1}$.
3. Es sei $n \in \mathbb{N}$ beliebig. Dann existieren (eindeutig bestimmte) paarweise verschiedene Primzahlen $p_1, \dots, p_k$ und $\nu_1, \dots, \nu_k \in \mathbb{N}$ so, dass

$$n = p_1^{\nu_1} \cdots p_k^{\nu_k} \, .$$

Da $p_j^{\nu_j}$ paarweise teilerfremd sind, gilt nach 1. und 2.

$$\begin{aligned} \varphi(n) &= \prod_{j=1}^{k} \varphi(p_j^{\nu_j}) = \prod_{j=1}^{k} p_j^{\nu_j} - p_j^{\nu_j - 1} = \\ &= \left(\prod_{j=1}^{k} p_j^{\nu_j} \right) \left(\prod_{j=1}^{k} \left(1 - \frac{1}{p_j} \right) \right) = n \prod_{j=1}^{k} \left(1 - \frac{1}{p_j} \right) = n \prod_{\substack{p \text{ prim} \\ p \mid n}} \left(1 - \frac{1}{p} \right) \, . \end{aligned}$$

□

22 Untergruppen und Homomorphismen

In Abschnitt 2 haben wir Gruppen definiert, haben Beispiele kennengelernt und einige einfache Eigenschaften hergeleitet. Wir werden jetzt einen genaueren Blick auf Gruppen werfen, wobei das Schwergewicht auf endlichen Gruppen liegen wird.
Ähnlich wie Unterräume bei linearen Räumen spielen "Untergruppen" bei Gruppen eine wichtige Rolle.

Definition 22.1 Es sei $G = (G, \circ)$ eine Gruppe. Eine Menge $U \subset G$ heißt *Untergruppe von* G wenn $(U, \circ_{|U \times U})$ eine Gruppe ist (d. h. wenn mit $a, b \in U$ auch $a \circ b \in U$ gilt, und wenn $e \in U$ und mit $a \in U$ auch $a^{-1} \in U$ ist). Statt $a \circ b$ schreiben wir auch kurz ab.

Bemerkung 22.2 Eine Menge $U \subset G, U \neq \emptyset$ ist schon dann eine Untergruppe, wenn mit $a, b \in U$ auch $a \circ b^{-1} \in U$ gilt.
(Denn: Ist $a \in U$, so ist $e = a \circ a^{-1} \in U$. Also ist auch $a^{-1} = e \circ a^{-1} \in U$. Sind $a, b \in U$ so gilt also auch $a \circ b = a \circ (b^{-1})^{-1} \in U$.)

Beispiel 22.3 1. Ist G eine beliebige Gruppe, so sind $U = G$ und $U = \{e\}$ Untergruppen (die sog. trivialen Untergruppen).

2. Ist $(G, \circ) = (\mathbb{C}, +)$, so haben wir folgende Kette ineinander geschachtelter Untergruppen (wobei $m \in \mathbb{N}$)

$$\{0\} \subset m\mathbb{Z} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C} \, .$$

3. Ist $(G, \circ) = (\mathbb{C} \setminus \{0\}, \cdot)$, so haben wir folgende Kette ineinander geschachtelter Untergruppen:

$$\{1\} \subset \{\pm 1\} \subset \mathbb{Q} \setminus \{0\} \subset \mathbb{R} \setminus \{0\} \subset \mathbb{C} \setminus \{0\} \, .$$

4. Es sei K ein Körper und $(G, \circ) = (GL_n(K), \cdot)$ die Gruppe der invertierbaren $(n \times n)$ Matrizen über K mit der Matrix-Multiplikation (vgl. B. 8.25). Dann ist

$$SL_n(K) := \{A \in GL_n(K) : \det A = 1\}$$

eine Untergruppe von $GL_n(K)$. (Sind $A, B \in SL_n(K)$, so gilt $\det(AB^{-1}) = \det A / \det B = 1$, also ist $AB^{-1} \in SL_n(K)$).
Weiter ist für $K = \mathbb{K}$

$$U_n := \{A \in GL_n(\mathbb{K}) : A \text{ unitär}\}$$

eine Untergruppe von $GL_n(\mathbb{K})$ ([Ü]), die sog. unitäre Gruppe. Für $\mathbb{K} = \mathbb{R}$ heißt $O_n := U_n$ auch orthogonale Gruppe. Die Menge

$$SO_n := O_n^+ := \{A \in O_n : \det A = 1\}$$

ist eine Untergruppe von O_n (die sog. spezielle orthogonale Gruppe).

Definition 22.4 Es sei $(G, \circ)$ eine Gruppe. Dann heißt

$$\operatorname{ord}(G) := |G| \; (\in \mathbb{N} \cup \{\infty\})$$

die *Ordnung von G*.

Neben den Gruppen $(\mathbb{Z}_m, +)$ und $(\mathbb{Z}_m^*, \cdot)$ spielen in der Theorie endlicher Gruppen die symmetrischen Gruppen S_n eine wichtige Rolle.

Beispiel 22.5 (vgl. B. 2.3 und Abschnitt 11)
Es sei $\sigma \in S_n$ (d. h. $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ bijektiv). Neben der Schreibweise

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

kann σ auch in der sog. Zykelschreibweise dargestellt werden: Jedes σ ist Produkt aus Zykeln, d. h.

$$\sigma = \left[a_1^{(1)}, \ldots, a_{m_1}^{(1)}\right] \circ \cdots \circ \left[a_1^{(k)}, \ldots, a_{m_k}^{(k)}\right] \, ,$$

wobei ein Zykel $[a_1, \ldots, a_m]$ eine Abkürzung für die Abbildung $\tau \in S_n$, mit $\tau(a_j) = a_{j+1}$ für $j = 1, \ldots, m-1$ und $\tau(a_m) = a_1$ sowie $\tau(a) = a$ für $a \notin \{a_1, \ldots, a_m\}$ ist (d. h. $a_1 \to a_2 \to a_3 \to \ldots \to a_m \to a_1$). Eine solche Darstellung ist i. a. keineswegs eindeutig. Ist etwa

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 2 & 4 & 5 & 1 & 7 & 6 \end{pmatrix} .$$

So ist

$$\sigma = [1, 3, 4, 5] \circ [6, 7] .$$

Dabei sind die Zykeln der Länge 2 Transpositionen (D. 11.1). Zykeln der Länge 1 (wie oben im Beispiel 2) werden nicht aufgeführt.
Es gilt damit etwa: S_3 hat die nichttrivialen Untergruppen

$$U_1 = \{id, [1, 2, 3], [1, 3, 2]\}$$

der Ordnung 3 und

$$U_2 = \{id, [1, 2]\},\ U_3 = \{id, [1, 3]\},\ U_4 = \{id, [2, 3]\}$$

der Ordnung 2. Dabei heißt U_1 Drehgruppe und U_2, U_3, U_4 heißen Spiegelungsgruppen. (Nummeriert man die Ecken eines gleichseitigen Dreiecks mit 1, 2, 3, so entsprechen den Symmetriebewegungen des Dreiecks die obigen Permutationen der Eckpunkte:

entspricht $[1, 2, 3]$

entspricht $[1, 3, 2]$

entspricht $[2, 3]$

entspricht $[1, 2]$

entspricht $[1, 3]$)

Definition 22.6 Es seien $(G, \circ)$ und $(H, \bullet)$ Gruppen. Eine Abbildung $h : G \to H$ heißt *(Gruppen-) Homomorphismus*, falls

$$h(a \circ b) = h(a) \bullet h(b)$$

für alle $a, b \in G$ gilt. Ein Homomorphismus h heißt *Injektion* (oder *Einbettung*), falls h injektiv ist. Ein bijektiver Homomorphismus heißt *Isomorphismus*. Existiert ein Isomorphismus $h : G \to H$, so heißen G und H *isomorph*.

Beispiel 22.7 1. Es sei $h : (\mathbb{Z}, +) \to (\mathbb{Z}_m, +)$ definiert durch

$$h(a) := [a] \qquad (a \in \mathbb{Z}) .$$

Dann ist h ein Homomorphismus (denn

$$h(a + b) = [a + b] = [a] + [b] = h(a) + h(b))$$

2. Ist $h : (\mathbb{R}, +) \to (\mathbb{C}, +)$ definiert durch $h(x) = x+i0$ $(x \in \mathbb{R})$, so ist h eine Einbettung (die natürliche Einbettung von $\mathbb{R}$ in $\mathbb{C}$).

Wir stellen einige elementare Eigenschaften von Homomorphismen zusammen:

Satz 22.8 *Es seinen $(G, \circ)$ und $(H, \bullet)$ Gruppen, und es sei $h : G \to H$ ein Homomorphismus. Dann gilt*
1. $h(e) = e$.
2. $h(a^{-1}) = (h(a))^{-1}$ für alle $a \in G$.
3. Ist $U \subset G$ eine Untergruppe von G, so ist $h(U) \subset H$ eine Untergruppe von H.
4. Ist $V \subset H$ eine Untergruppe von H, so ist $h^{-1}(V) \subset G$ eine Untergruppe von G.
(Insbesondere ist damit

$$Kern(h) := h^{-1}(\{e\})$$

eine Untergruppe von G).
5. h ist genau dann injektiv, wenn $Kern(h) = \{e\}$ ist.

Beweis.
1. Es gilt

$$h(e) = h(e \circ e) = h(e) \bullet h(e) ,$$

also ist $e = h(e)h(e)^{-1} = h(e)h(e)h(e)^{-1} = h(e)$.
2. Nach 1. gilt für $a \in G$

$$h(a)h(a^{-1}) = h(aa^{-1}) = h(e) = e ,$$

also $(h(a))^{-1} = h(a^{-1})$ nach S.2.2.2.
3. und 4.: [Ü]
5. Ist h injektiv, so ist $h^{-1}(\{e\}) = \text{Kern}(h)$ höchstens einelementig. Da Kern(h) ein Unterraum von G ist, ist Kern$(h) = \{e\}$.
Es sei umgekehrt Kern$(h) = \{e\}$. Dann gilt für alle $a, b \in G$ mit $h(a) = h(b)$.

$$h(ab^{-1}) = h(a)h(b^{-1}) = h(a)(h(b))^{-1} = e ,$$

also $ab^{-1} = e$ und damit $a = ab^{-1}b = eb = b$. □

Satz 22.9 *Es seien G, H Gruppen, und es sei $h : G \to H$ ein Isomorphismus. Dann ist auch $h^{-1} : H \to G$ ein Homomorphismus (und damit ein Isomorphismus).*

Beweis.
Es seien $u, v \in H$. Dann existieren $a, b \in G$ mit $u = h(a), v = h(b)$. Also gilt

$$h^{-1}(uv) = h^{-1}(h(a)h(b)) = h^{-1}(h(ab)) = ab = h^{-1}(u)h^{-1}(v) \ .$$

□

Ist $h : G \to H$ eine Injektion, so ist $h : G \to h(G) \subset H$ ein Isomorphismus zwischen G und der Untergruppe $h(G)$ von H. Dies kann man ausnutzen um die Rolle der symmetrischen Gruppen zu verstehen:

Satz 22.10 (Cayley) *Jede endliche Gruppe der Ordnung n ist isomorph zu einer Untergruppe der symmetrischen Gruppe S_n.*

Beweis.
Es sei G eine endliche Gruppe. Wir müssen eine Einbettung $h : G \to S_n$ konstruieren. Es sei $G = \{a_1, \dots, a_n\}$, und es sei $a \in G$. Die Multiplikation (von links) mit a bewirkt eine Permutation von G, d. h. es existiert genau ein $\sigma = \sigma_a \in S_n$ mit

$$aa_j = a_{\sigma(j)} \qquad (j = 1, \dots, n)$$

(wichtig: Ist $aa_j = aa_k$, so gilt $a^{-1}aa_j = a^{-1}aa_k$, also $a_j = a_k$). Es gilt dabei für $a, b \in G$

$$\sigma_a\sigma_b = \sigma_{ab}$$

(Denn: Für $j = 1, \dots, n$ gilt

$$(ab)a_j = a(ba_j) = aa_{\sigma_b(j)} = a_{\sigma_a(\sigma_b(j))} = a_{(\sigma_a\sigma_b)(j)} \ ,$$

d. h. $\sigma_{ab} = \sigma_a\sigma_b$).
Also ist $h : G \to S_n$ mit $h(a) := \sigma_a$ ein Homomorphismus. Außerdem gilt: Ist $id = h(a) = \sigma_a$, so ist $aa_j = a_{id(j)} = a_j$ für $j = 1, \dots, n$ und damit $a = e$. Also ist $\mathrm{Kern}(h) = \{e\}$, d. h. h ist injektiv und damit eine Einbettung. □

Definition 22.11 1. Es sei $(G, \circ)$ eine Gruppe. Für $x \in G$ setzen wir

$$x^n := \underbrace{x \circ \cdots \circ x}_{n\text{mal}} \ , \quad x^{-n} := (x^{-1})^n \ (= (x^n)^{-1})$$

für $n \in \mathbb{N}$ (und $x^0 := e$). (Ist $(G, \circ)$ eine "additive Gruppe", d. h. "$\circ$" = "+", so schreibt man üblicherweise nx statt x^n).

Für $x \in G$ setezn wir zudem

$$< x >:= \{x^n : n \in \mathbb{Z}\} .$$

(Dabei ist $< x >$ eine Untergruppe von G, wie man leicht sieht.)
2. Die Gruppe $(G, \circ)$ heißt *zyklisch*, falls $G =< x >$ für ein $x \in G$ gilt.
Das Element x heißt dann *erzeugendes Element* von G.

Beispiel 22.12 1. Es sei $(G, \circ) = (\mathbb{Z}, +)$. Dann gilt

$$\mathbb{Z} =< 1 >=< -1 >$$

und ± 1 sind die einzigen erzeugenden Elemente.
2. Es sei $(G, \circ) = (\mathbb{Z}_m, +)$. Dann gilt

$$\mathbb{Z}_m =< [1] > \ .$$

Außerdem gilt für jede prime Restklasse $[a]$ auch $\mathbb{Z}_m =< [a] >$ ([Ü]).

Das Repertoire an zyklischen Gruppen ist damit (bis auf Isomorphie), schon erschöpft:

Satz 22.13 *Jede zyklische Gruppe $G =< x >$ ist entweder isomorph zu $(\mathbb{Z}, +)$ oder zu $(\mathbb{Z}_m, +)$ für ein $m \in \mathbb{N}$.*

Beweis.
1. Fall: Es ist $x^n \neq x^k$ für alle $n, k \in \mathbb{Z}, n \neq k$. Dann ist durch $h : \mathbb{Z} \to < x >$ mit $h(n) := x^n$ ein Isomorphismus von $\mathbb{Z}$ nach $G =< x >$ definiert.
(Denn: Für $n, k \in \mathbb{Z}$ gilt

$$h(n + k) = x^{n+k} = x^n \circ x^k = h(n) \circ h(k) ,$$

also ist h ein Homomorphismus. Ist $y \in G =< x >$, so existiert ein $n \in \mathbb{N}$ mit $x^n = y$, d. h. $y = h(n)$. Sind schließlich $n, k \in \mathbb{Z}$ mit $h(n) = h(k)$, d. h. $x^n = x^k$, so ist nach Voraussetzung $n = k$, also ist h bijektiv.)
2. Fall: Es existieren $n, k \in \mathbb{Z}$, $n \neq k$, mit $x^n = x^k$. Dann ist h aus "1. Fall" ein surjektiver Homomorphismus, aber nicht mehr injektiv. Also existiert nach S. 22.8.5 ein $n \in \text{Kern}(h) \setminus \{0\}$. Da $\text{Kern}(h)$ eine Gruppe ist, ist mit n auch $-n \in \text{Kern}(h)$. Also existiert ein minimales $m \in \mathbb{N}$ mit $m \in \text{Kern}(h)$. Dann sind x^j, $j = 0, \ldots, m-1$ paarweise verschieden (ansonsten gäbe es $i, j \in \{0, \ldots, m-1\}, i < j$, mit $x^i = x^j$, also $x^{j-i} = e$ und damit $j - i \in \text{Kern}(h)$ mit $0 < j - i < m$).
Aus $x^m = e$ folgt $x^{n+km} = x^n \circ x^{km} = x^n \circ (x^m)^k = x^n$ für alle $n, k \in \mathbb{Z}$, d. h. $h(n) = h(n + km)$ für alle n, k. Also ist durch

$$\tilde{h}([n]_m) := h(n) \qquad (n \in \mathbb{Z})$$

eine Abbildung $\tilde{h} : (\mathbb{Z}_m, +) \to < x >= G$ (wohl-)definiert. Aus obigen Überlegungen folgt, dass $\tilde{h}$ ein Isomorphismus ist. □

Bemerkung und Definition 22.14 Es sei $(G, \circ)$ eine Gruppe und es sei $x \in G$. Dann heißt

$$\text{ord}(x) := \text{ ord } (< x >)$$

Ordnung von x.
Es ist $x^n = e$ für ein $n \in \mathbb{Z}$ genau dann, wenn $n = 0$ oder wenn $\text{ord}(x) < \infty$ und n Vielfaches von $\text{ord}\,(x)$ ist, d. h. $n = k\,\text{ord}\,(x)$ für ein $k \in \mathbb{Z}$. (Dies ergibt sich aus der Isomorphie zwischen $< x >$ und $\mathbb{Z}_m$ mit $m = \text{ord}(x)$ im Falle $\text{ord}(x) < \infty$ und der Isomorphie zwischen $< x >$ und $\mathbb{Z}$ im Falle $\text{ord}(x) = \infty$.)

Bemerkung und Definition 22.15 Es sei $(G, \circ)$, eine Gruppe, und es sei $H \subset G$ eine Untergruppe. Für $a, b \in G$ definieren wir

$$a \sim b : \quad \Leftrightarrow \quad ab^{-1} \in H \quad (\Leftrightarrow\ a \in Hb := \{hb : h \in H\}) \ .$$

Man sieht leicht, dass $\sim$ eine Äquivalenzrelation auf G ist. Die Äquivalenzklassen sind die Teilmengen Hb, die *Rechtsrestklassen* genannt werden. (Durch Betrachtung von $b^{-1}a$ anstelle von ab^{-1} erhält man *Linksrestklassen* bH; für abelsche Gruppen gilt natürlich $Hb = bH$).

Beispiel 22.16 1. Es sei $(G, \circ) = (\mathbb{Z}, +)$. Für $m \in \mathbb{N}$ ist $m\mathbb{Z}$ eine Untergruppe von $\mathbb{Z}$. Hier sind die Rechtsrestklassen (und Linksrestklassen) gerade die üblichen Restklassen $[a]_m$
(Denn: Sind $a, b \in \mathbb{Z}$, so gilt $a - b \in m\mathbb{Z}$ genau dann, wenn $a = b + km$ für ein $k \in \mathbb{Z}$, also $a \in m\mathbb{Z} + b$, d. h. $[a]_m = [b]_m$)
2. Es sei $(G, \circ) = (S_3, \circ)$. Ist $H = \{id, [1,2,3], [1,3,2]\}$ (vgl. B. 22.5) so existieren die zwei Rechts- (und Links-)restklassen

$$R_1 = H \cdot id = H$$

und

$$R_2 = H \cdot [1,2] = \{[1,2], [1,3], [2,3]\} \ .$$

Ist $H = \{id, [1,2]\}$, so existieren die drei Rechtsrestklassen

$$\begin{aligned} &R_1 = H \cdot id \ , \quad R_2 = H \cdot [1,3] = \{[1,3],\ [1,3,2]\} \ , \\ &R_3 = H \cdot [2,3] = \{[2,3],\ [1,2,3]\} \end{aligned}$$

und die drei Linksrestklassen

$$\begin{aligned} &R_1 = id \cdot H \ , \quad R_2 = [1,3] \cdot H = \{[1,3], [1,2,3]\} \ , \\ &R_3 = [2,3] \cdot H = \{[2,3], [1,3,2]\} \ . \end{aligned}$$

(Hier stimmen die Rechts- und die Linksklassen zu $[1,3]$ und $[2,3]$ nicht überein.)
3. Es sei $(G,\circ) = (\mathbb{R}^2,+)$ und es sei $H = \mathbb{R}x = \{\lambda x : \lambda \in \mathbb{R}\}$ für ein $x \in \mathbb{R}^2 \setminus \{0\}$. Dann sind die Restklassen Hb gegeben durch

$$Hb = \mathbb{R}x + b \qquad (b \in \mathbb{R}^2) \ .$$

Bemerkung und Definition 22.17 Es seien $(G,\circ)$ eine Gruppe und $H \subset G$ eine Untergruppe. Dann heißt

$$G : H := |\{Hb : b \in G\}| \ (\in \mathbb{N} \cup \{\infty\}) \ ,$$

also die Anzahl der Rechtsrestklassen bzgl. H, der *Index von H in G*. Dabei sieht man leicht, dass $G : H$ auch mit der Anzahl der Linksrestklassen bzgl. H übereinstimmt ([Ü]).

Es gilt damit folgendes elementare Ergebnis.

Satz 22.18 (Lagrange) *Ist G eine endliche Gruppe und H eine Untergruppe, so ist*

$$ord\,G := (G : H) \cdot ord\,(H) \ .$$

Insbesondere teilt also die Ordnung von H die Ordnung von G.

Beweis.
Der Beweis ergibt sich sofort daraus, dass die Äquivalenzklassen Hb eine Zerlegung von G in $G : H$ Teilmengen bewirken, und dass $|Hb| = |H| = \operatorname{ord} H$ für alle $b \in G$ gilt (die Abbildung $\varphi : H \to Hb$ mit $\varphi(h) = hb$ ist bijektiv). □

Also Konsequenzen erhalten wir

Satz 22.19 (Euler) *Es sei G eine endliche Gruppe, und es sei $x \in G$. Dann ist $ord(x)$ ein Teiler von $ord(G)$. Insbesondere gilt stets*

$$x^{ord\,(G)} = e \ .$$

Beweis.
Die Behauptung ergibt sich sofort aus dem Satz von Lagrange mit $H =< x >$ und aus B. /D. 22.14. □

Wendet man dies speziell auf die primen Restklassengruppen $\mathbb{Z}_m^*$ an, so ergeben sich wichtige Zahlentheoretische Konsequenzen.

Folgerung 22.20 1. Sind $a \in \mathbb{Z}, m \in \mathbb{N}$ mit $ggT(a,m) = 1$, so ist

$$a^{\varphi(m)} \equiv 1 \mod m \,.$$

2. (Kleiner Satz von Fermat) Sind $a \in \mathbb{Z}$ und p eine Primzahl, so gilt

$$a^p \equiv a \mod p$$

und in dem Fall, dass p kein Teiler von a ist (d. h. $ggT(a,p) = 1$)

$$a^{p-1} \equiv 1 \mod p \,.$$

Denn: Nach Definition gilt $|\mathbb{Z}_m^*| = \varphi(m)$. Also folgt 1. sofort aus S. 22.19. Ist $m = p$, wobei p prim, so ist $\varphi(p) = p-1$ nach S. 21.12 und damit folgt die zweite Aussage von 2. aus 1. Durch Multiplikation von $a^{p-1} \equiv 1 \bmod p$ mit a ergibt sich die erste Aussage im Falle $ggT(a,p) = 1$. Ist $p|a$, d. h. $a = kp$ für ein $k \in \mathbb{Z}$, so gilt

$$a \equiv 0 \mod p$$

also auch

$$a^p \equiv 0 \equiv a \mod p \,.$$

Anwendungsbeispiel 22.21 (RSA-Kryptographie-Verfahren) Eine (noch recht junge) Anwendung der obigen Folgerung 22.20 ergibt sich für Verschlüsselungsverfahren. Das von Rivest, Shamir und Adleman im Jahre 1978 vorgestellte Verfahren beruht auf folgender Beobachtung:
Es seien p, q Primzahlen (möglichest groß ; etwa je 100 Stellen) $p \neq q$, und es sei $N = pq$. Dann gilt $\varphi(N) = (p-1)(q-1)$ nach S. 21.22. Weiter sei $s \in \mathbb{N}$ teilerfremd zu $\varphi(N)$. Dann existieren nach S. 21.5 $t, k \in \mathbb{Z}$ mit

$$st = k\varphi(N) + 1$$

(d. h. $st \equiv 1 \bmod \varphi(N)$). O. E. können wir $t, k \in \mathbb{N}$ annehmen.

Behauptung: Dann gilt für alle $a \in \mathbb{N}$

$$(a^s)^t = a^{st} \equiv a \mod N \tag{10}$$

Denn: Ist $ggT(a,N) = 1$ so gilt nach F. 22.20.1

$$a^{st} = a^{k\varphi(N)+1} = (a^{\varphi(N)})^k \, a \equiv a \bmod N \,.$$

Da $\varphi(N) = (p-1)(q-1)$ gilt, folgt aus $st \equiv 1 \bmod \varphi(N)$ auch $st \equiv 1 \bmod (p-1)$ und $st \equiv 1 \bmod (q-1)$, d. h. es existieren $\ell, m \in \mathbb{N}$ mit $st = \ell(p-1) + 1 = m(q-1) + 1$

(nämlich etwa $\ell = k(p-1)$ und $m = k(q-1)$). Also folgt aus F. 22.20.2 im Falle $a \not\equiv 0 \bmod p$ (beachte $(\mathbb{Z}_p \setminus \{[0]\}, \cdot)$ ist eine Gruppe)

$$[a^{st}]_p = [a^{\ell(p-1)}a]_p = [a]_p^{\ell(p-1)}[a]_p = [a^p]_p^{\ell} \cdot [a]_p^{-\ell}[a]_p = [a]_p^{\ell}[a]_p^{-\ell}[a]_p = [a]_p$$

(und im Falle $a \equiv 0 \bmod p$ natürlich auch $a^{st} \equiv a \equiv 0 \bmod p$).
Entsprechend gilt

$$a^{st} \equiv a \bmod q \ .$$

Da p, q teilerfremd sind, gilt mit F. 21.6.3

$$a^{st} \equiv a \bmod N = pq \ .$$

Wie kann man die Beziehung (10) nutzen?
Will man eine Nachricht, die ohne Einschränkung ein Vektor $(a_1, \ldots, a_m)$ natürlicher Zahlen sein soll, übermitteln, ohne dass Unbefugte die Nachricht verstehen können, so kann man, nachdem man p, q, s, t wie oben festgelegt hat, die verschlüsselte Information $(a_1^s, \ldots, a_m^s)$ wie auch s und N öffentlich übertragen. Die Nachricht $(a_1, \ldots, a_m)$ ist dann aus der Gleichung (10) sehr leicht rekonstruierbar, wenn man zusätzlich t bzw. $\varphi(N)$ bzw. p und q kennt (t ist aus $\varphi(N)$ leicht berechenbar). Die Prmfaktorzerlegung $N = pq$ ist für ein gegebenes großes N jedoch eine sehr schwierige Aufgabe, die für genügend große p, q (mit den derzeit bekannten Methoden) praktisch unmöglich ist.

23 Normalteiler

Ein (bis heute unerreichtes) Fernziel ist es natürlich, alle (endlichen) Gruppen zu kennen bzw. zu klassifizieren, wobei isomorphe Gruppen nicht als verschieden angesehen werden.
Für Primzahlordnungen liegen die Verhältnisse besonders einfach. Es gilt nämlich

Satz 23.1 *Ist p eine Primzahl, so ist jede Gruppe G der Ordnung p isomorph zu $(\mathbb{Z}_p, +)$.*

Beweis. Es sei $x \in G \setminus \{e\}$. Dann hat x nach dem Satz von Euler (S. 22.19) die Ordnung p, d. h. $< x >= G$. Nach S. 22.13 ist G damit isomorph zu $(\mathbb{Z}_p, +)$. □

Im allgemeinen Fall spielen gewisse Untergruppen eine zentrale Rolle:

Definition 23.2 Es sei G eine Gruppe und $H \subset G$ eine Untergruppe. Dann heißt H *Normalteiler (von G)*, falls $Hg = gH$ für alle $g \in G$ gilt (d. h. falls alle Rechts- und Linksrestklassen bzgl. H übereinstimmen). Mann schreibt dann $H \lhd G$.

Bemerkung 23.3 Es sei G eine Gruppe, und es sei H eine Untergruppe von G. Dann heiss t

$$H^g := gHg^{-1} := \{ghg^{-1} : h \in H\}$$

zu h konjugierte Untergruppe. Man rechnet leicht nach, dass H^g stets eine Untergruppe von G ist ([Ü]). Es gilt damit: H ist genau dann Normalleiter von G, wenn $H^g = H$ für alle $g \in G$ gilt.
(Denn: "⇒": Es gelte $gH = Hg$ für alle $g \in G$. Wir zeigen:
Dann ist $H^g \subset H$ für alle $g \in G$ (und damit gilt dann auch $H^{g^{-1}} \subset H$, also $H = (H^{g^{-1}})^g \subset H^g$, d. h. $H = H^g$ für alle $g \in G$).
Es sei also $a \in H^g$ d. h. $a = ghg^{-1}$ für ein $h \in H$. Da $gh \in gH = Hg$ ist, existiert ein $\tilde{h} \in H$ mit $gh = \tilde{h}g$. Also ist $a = \tilde{h}gg^{-1} = \tilde{h} \in H$.
"⇐": Es gelte $H^g = H$ für alle $g \in G$, und es sei $a \in Hg$, d. h. $a = hg$ für ein $h \in H$. Dann existiert ein $\tilde{h} \in H$ mit $h = g\tilde{h}g^{-1}$. Folglich ist $a = hg = g\tilde{h}g^{-1}g = g\tilde{h} \in gH$. Also ist $Hg \subset gH$. Entsprechend sieht man, dass $gH \subset Hg$ gilt (beachte dabei: $H^{g^{-1}} = H$).)

Beispiel 23.4 1. Ist G abelsch, so ist jede Untergruppe $H \subset G$ Normalteiler.
2. Es sei $G = GL_n(K)$ und $H = SL_n(K)$ (vgl. B. 22.3.4).
Ist $A \in GL_n(\mathbb{K})$, so gilt für alle $B \in SL_n(\mathbb{K})$:

$$\det(ABA^{-1}) = \det(A)\det(B)\det(A^{-1}) = \det(B) = 1\ ,$$

d. h. $ABA^{-1} \in SL_n(\mathbb{K})$. Damit gilt $H^A \subset H$.
Da A beliebig war, ist auch $H = (H^{A^{-1}})^A \subset H^A$, d. h. $H^A = H$ und damit ist H nach B. 23.3 Normalteiler in G.
3. Es sei $G = SL_2(\mathbb{K})$ und

$$H = \left\{ B = B_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} : t \in \mathbb{K} \right\} .$$

Dann ist H eine Untergruppe von $SL_2(\mathbb{K})$.
(Denn: Es gilt
$(B_t B_s = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & s+t \\ 0 & 1 \end{pmatrix} = B_{s+t}$, also $(B_t)^{-1} = B_{-t}$.)
Für $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ gilt

$$AB_tA^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -t & 1 \end{pmatrix}$$

d. h. für $t \neq 0$ ist $AB_tA^{-1} \notin H$, also $H^A \not\subset H$. Damit ist H kein Normalteiler.

Satz 23.5 *Es seien U und V Gruppen und es sei $f : U \to V$ ein Homomorphismus. Ist $N \subset V$ ein Normalteiler von V, so ist $f^{-1}(N)$ ein Normalteiler von U. Insbesondere ist Kern(f) Normalteiler von U.*

Beweis. Es genügt, zu zeigen: Für alle $g \in G$ ist $(f^{-1}(N))^g \subset f^{-1}(N)$. Ist $g \in G$ und $u \in (f^{-1}(N))^g$ so existiert ein $\tilde{u} \in f^{-1}(N)$ mit $u = g\tilde{u}g^{-1}$. Also ist

$$f(u) = f(g\tilde{u}g^{-1}) = f(g)f(\tilde{u})(f(g))^{-1} \in f(g)N(f(g))^{-1} = N ,$$

d. h. $u \in f^{-1}(N)$.
Da $\{e\}$ ein Normalteiler von V ist, ist insbesondere

$$\text{Kern}(f) = f^{-1}(\{e\})$$

Normalteiler von U. □

Die Normalteiler sind unter allen Untergruppen deswegen ausgezeichnet, weil die Restklassen von Normalteilern (wie die Restklassen von $m\mathbb{Z}$ in $\mathbb{Z}$) durch Definition einer Restklassenverknüpfung zu einer Gruppe gemacht werden können.

Satz 23.6 *Es seien G eine Gruppe und H ein Normalteiler in G. Dann ist die Menge*

$$G/H := \{Hg : g \in G\} \qquad (= \{gH : g \in G\})$$

eine Gruppe bzgl. der repräsentantenweise definierten Multiplikation

$$Hg \cdot Hh := H(g \cdot h)$$

(die sog. Faktorgruppe G/H). Die kanonische Projektion

$$\pi : G \to G/H\ , \qquad \pi(g) := Hg$$

ist ein surjektiver Gruppenhomomorphismus mit $Kern(\pi) = H$, und es gilt

$$ord(G/H) = G : H\ .$$

Beweis. Wesentlich ist, dass die Multiplikation wohldefiniert ist, d. h. unabhängig von den Repräsentanten g, h ist. Es seien also g, h, a, b mit $Hg = Ha$ und $Hh = Hb$ gegeben. Dann gilt auch $gH = aH$, und folglich ist $g = ge \in aH$ und $h \in Hb$. Also ist $a^{-1}g \in H$ und $hb^{-1} \in H$.
Dann ist auch

$$a^{-1}ghb^{-1} \in H$$

und also

$$ghb^{-1}a^{-1} = aa^{-1}ghb^{-1}a^{-1} \in aHa^{-1} = H^a = H\ .$$

Dies bedeutet wiederum $gh \in Hab$ bzw. $Hgh = Hab$. Klar ist, dass $He = H$ das neutrale Element von G/H und Ha^{-1} das zu Ha inverse Element ist.
Aus diesen Überlegungen ergibt sich auch, dass π ein surjektiver Homomorphismus ist. Ist $\pi(g) = He$, so ist $Hg = He = H$ und damit $g \in H$. Umgekehrt folgt aus $g \in H$ auch wieder $Hg = H = He$, also $\pi(g) = He$. Also ist $\mathrm{Kern}(\pi) = H$.
Nach der Definition von $G : H$ (B./D. 22.17) ist $\mathrm{ord}(G/H) = G : H$. □

Der folgende Satz liefert eine gewisse Umkehrung von S. 23.6.

Satz 23.7 *Es seien G und B Gruppen, und es sei $f : G \to B$ ein surjektiver Gruppenhomomorphismus. Dann ist B isomorph zur Faktorgruppe $G/Kern(f)$.*

Beweis. Wir zeigen: Durch $i(\mathrm{Kern}(f) \cdot g) := f(g)$ ist eine Abbildung $i : G/\mathrm{Kern}(f) \to B$ wohldefiniert.
(Denn: Ist $\mathrm{Kern}(f) \cdot g = \mathrm{Kern}(f) \cdot h$ für $g, h \in G$, so gilt

$$gh^{-1} \in \mathrm{Kern}(f)\ .$$

Also ist $e = f(gh^{-1}) = f(g)(f(h))^{-1}$, d. h. $f(g) = f(h)$. Damit ist $i(\text{Kern}(f) \cdot g) = i(\text{Kern}(f) \cdot h))$.
Man sieht sofort, dass i ein Homomorphismus ist, und nach Definition ist i surjektiv. Außerdem gilt $\text{Kern}(i) = \{\text{Kern}(f)\}$, also besteht $\text{Kern}(i)$ genau aus dem neutralen Element in $G/\text{Kern}(f)$. Damit ist i auch injektiv nach S. 22.8.5. □