0 Einleitung

Die *Lineare Algebra* ist ein grundlegendes Teilgebiet der Mathematik, das sich mit Vektorräumen und linearen Abbildungen zwischen diesen beschäftigt. Solche Fragestellungen sind typisch in der Physik. Näherungsweise besteht eigentlich immer eine lineare Beziehung zwischen einer (als klein genug angenommenen) Ursache und ihrer Wirkung (Taylorsche Formel aus dem letzten Semester). Physikalische Gesetze liefern die Beziehung zwischen Ursache und Wirkung und sind deshalb in guter Näherung lineare Abbildungen.

Inhalt

1. *Grundbegriffe (z.T. als Wiederholung).* Mengen, Abbildungen, Zahlen, Gruppen, Ringe, Körper
2. *Vektorräume und Lineare Abbildungen.* Lineare Gleichungssysteme, Matrizen, Determinanten
3. *Eigenwerte.* Charakteristisches Polynom, Diagonalisierbarkeit
4. *Euklidische und unitäre Vektorräume.* Skalarprodukte, Spektralsatz
5. *Tensoren.* Dualräume, Tensorprodukte, multilineare Algebra

Literatur

[1] G. Fischer, "Lineare Algebra," Vieweg (2005).

[2] F. Lorentz, "Lineare Algebra 1," Spektrum (2003).

[3] K. Jänich, "Lineare Algebra," Springer (2003).

1 Grundbegriffe

1.1 Mengen

Endliche Mengen kann man durch eine vollständige Liste ihrer Elemente angeben, Schreibweise $X = \{x_1, x_2, \ldots, x_n\}$. Dabei ist die Reihenfolge beliebig, und das gleiche Element darf nicht mehrmals auftreten. Die x_i, $i = 1, \ldots, n$, heißen *Elemente* der Menge X, Schreibweise $x_i \in X$. Wenn x kein Element von X ist, schreiben wir $x \notin X$. Die *leere Menge* $\emptyset$ enthält kein Element. Die Verallgemeinerung auf unendliche Mengen ist Gegenstand der Mengenlehre. Für unsere Zwecke genügt es sich vorzustellen, daß auch unendliche Mengen aus Elementen bestehen und sich die folgenden Konstruktionen übertragen lassen.

Eine Menge Y heißt *Teilmenge* einer Menge X, wenn für jedes Element $y \in Y$ gilt $y \in X$. Wir haben $X = Y \quad \Leftrightarrow \quad X \subset Y$ und $Y \subset X$. Dabei steht $\Leftrightarrow$ für "genau dann, wenn". Wenn zuvor $A \Rightarrow B$ definiert ist als "aus A folgt B", dann bedeutet $A \Leftrightarrow B$ das gleiche wie "sowohl $A \Rightarrow B$ als auch $B \Rightarrow A$".

Typische Charakterisierung von Teilmengen $Y \subset X$:

$$Y := \{x \in X \ : \ x \text{ hat die Eigenschaft } E\} \ .$$

Dabei bedeutet :=, daß der links stehende Ausdruck durch den rechts stehenden definiert wird.

Sind $X_i, \dots, X_n$ Mengen, so kann man die *Vereinigung* definieren als

$$X_i \cup \dots \cup X_n := \{x \ : \ \text{es gibt ein } i \in \{1, 2, \dots, n\} \text{ mit } x \in X_i\}$$

und den *Durchschnitt* als

$$X_i \cap \dots \cap X_n := \{x \ : \ x \in X_i \text{ für alle } i \in \{1, 2, \dots, n\}\} \ .$$

Abkürzende Schreibweise: $\exists$ für "es existiert ein" und $\forall$ für "für alle".

Ist $Y \subset X$, dann ist das *Komplement* definiert als

$$X \setminus Y := \{x \in X \ : \ x \notin Y\} \ .$$

Seien $X_1, \dots, X_n$ Mengen, dann betrachten wir die *geordneten n-Tupel*

$$x = (x_1, x_2, \dots, x_n) \quad \text{mit} \quad x_1 \in X_1 \ , \ x_2 \in X_2 \ , \ \dots x_n \in X_n \ .$$

Nun erklären wir das *direkte Produkt* der Mengen $X_1, \dots, X_n$ als die Menge aller geordneten n-Tupel:

$$X_1 \times X_2 \times \dots \times X_n := \{(x_1, x_2, \dots, x_n) \ : \ x_i \in X_i\} \ .$$

Speziell schreiben wir $X^n := \underbrace{X \times \dots \times X}_{n}$.

Relationen zwischen Elementen einer Menge X werden beschrieben durch eine Teilmenge $R \subset X \times X$. Für $x, y \in X$ schreiben wir $x \sim y \quad \Leftrightarrow \quad (x, y) \in R$. Solche Relationen (mit speziellen Eigenschaften) sind sehr wichtig, deshalb ein Beispiel: Sei X die Menge aller Studierenden in diesem Hörsaal, dann betrachten wir

$$R = \{(x, y) \in X \times X \ : \text{sowohl } x \text{ als auch } y \text{ haben sich für die Übungen im Raum SR0 eingeschrieben, oder nicht}\} \ .$$

Offenbar gilt $(x, x) \in R$, aus $(x, y) \in R$ folgt $(y, x) \in R$ und aus $(x, y) \in R$ sowie $(y, z) \in R$ folgt $(x, z) \in R$. Eine Relation mit solchen Eigenschaften heißt Äquivalenzrelation:

Definition 1.1 Eine Relation $\sim$ auf X heißt *Äquivalenzrelation*, wenn für beliebige Elemente $x, y, z \in X$ gilt

(Ä1) $x \sim x$

(Ä2) $x \sim y \quad \Rightarrow \quad y \sim x$

(Ä3) $x \sim y$ und $y \sim z \quad \Rightarrow \quad x \sim z$

Ist eine Menge X und auf ihr eine Äquivalenzrelation gegeben, so heißt eine Teilmenge $A \subset X$ *Äquivalenzklasse* bezüglich $\sim$, falls

i) $A \neq \emptyset$

ii) $x, y \in A \quad \Rightarrow \quad x \sim y$

iii) $x \in A$, $y \in X$ und $x \sim y \quad \Rightarrow \quad y \in A$

Satz 1.1 *Ist $\sim$ eine Äquivalenzrelation auf einer Menge X, so gehört jedes Element $a \in X$ zu genau einer Äquivalenzklasse. Insbesondere gilt für zwei beliebige Äquivalenzklassen $A, A' \in X$ entweder $A = A'$ oder $A \cap A' = \emptyset$.*

Beweis. Sei $A_a := \{x \in X \ : \ x \sim a\}$. Dann ist A_a eine Äquivalenzklasse, die a enthält (Eigenschaften überprüfen!). Seien A, A' zwei Äquivalenzklassen mit $A \cap A' \neq \emptyset$, dann nehmen wir $a \in A \cap A'$. Per Konstruktion ist für jedes $x \in A$ auch $x \in A'$, also $A \subset A'$ und umgekehrt $A' \subset A$, also $A = A'$. □

Die Menge aller Äquivalenzklassen auf X bezüglich $\sim$ wird mit $X/\sim$ bezeichnet. Sie heißt *Quotientenmenge* von X nach der Äquivalenzrelation $\sim$. Im Beispiel gibt es zwei Äquivalenzklassen:

$A_1 = \{$Studierende, die sich für die Übungen in SR0 eingeschrieben haben$\}$

$A_2 = \{$Studierende, die sich nicht für die Übungen in SR0 eingeschrieben haben$\}$

Somit haben wir $X/\sim \ = \{A_1, A_2\}$. Ein Element $a \in A$ heißt *Repräsentant* der Äquivalenzklasse A. Für Beweise mit Äquivalenzklassen ist es oft zweckmäßig, die Aussage zunächst für einen ausgewählten Repräsentanten zu beweisen und dann zu zeigen, daß der Beweis unabhängig vom Repräsentanten ist. Wir schreiben auch $[a]$ für die eindeutig bestimmte Äquivalenzklasse, die den Repräsentanten a enthält.

1.2 Abbildungen

Abbildungen beschreiben Beziehungen zwischen Mengen. Seien X, Y Mengen, dann versteht man unter einer *Abbildung von X nach Y* eine Vorschrift f, die jedem $x \in X$ eindeutig ein $f(x) \in Y$ zuordnet. Wir schreiben $f : X \to Y$ für die Abbildung zwischen Mengen und $f : x \mapsto f(x)$ für die Zuordnung der Elemente.

Sei $f : X \to Y$ eine Abbildung zwischen Mengen und seien $M \subset X$ und $N \in Y$ Teilmengen. Dann heißt

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

das *Bild* von M in Y unter f und

$$f^{-1}(N) := \{x \in X \ : \ f(x) \in N\} \subset Y$$

das *Urbild* von N in X. Zu beachten ist, daß für eine einelementige Teilmenge $N = \{y\}$ das Urbild $f^{-1}(y) := f^{-1}(\{y\}) \subset X$ aus mehreren Elementen bestehen kann oder auch leer sein kann. Deshalb ist f^{-1} im allgemeinen keine Abbildung von Y nach X.

Definition 1.2 Eine Abbildung $f : X \to Y$ heißt

- *injektiv*, falls für $x, x' \in X$ aus $f(x) = f(x')$ stets $x = x'$ folgt
- *surjektiv*, falls es zu jedem $y \in Y$ ein $x \in X$ gibt mit $y = f(x)$
- *bijektiv*, falls f injektiv und surjektiv ist

Falls f bijektiv ist, dann ist die *Umkehrabbildung* $f^{-1} : Y \to X$ gegeben durch $f^{-1} : y \mapsto x = f^{-1}(y)$ mit $y = f(x)$.

Wichtige surjektive Abbildungen sind die *kanonischen Projektionen* des direkten Produkts von Mengen auf die i-te Komponente

$$\pi_i : X_1 \times \cdots \times X_n \to X_i \ , \quad \pi_i : (x_1, \ldots, x_n) \mapsto x_i \ .$$

Kanonisch bedeutet dabei, daß die Abbildung in der gegebenen Situation ohne Zusatzinformationen eindeutig definiert ist. Ebenso wird im Fall der Äquivalenzrelationen die Abbildung

$$\pi : X \to X/\sim \ , \quad \pi : a \mapsto A_a := \{x \in X \ : \ x \sim a\}$$

als kanonische Projektion bezeichnet.

Bemerkung: Sei X eine endliche Menge von paarweise verschiedenen Elementen und $f : X \to X$ eine Abbildung, so sind die Eigenschaften injektiv, surjektiv und bijektiv äquivalent.

Seien X, Y, Z Mengen und $f : X \to Y$ sowie $g : Y \to Z$ Abbildungen, so ist die *Komposition* dieser Abbildungen gegeben durch

$$g \circ f : X \to Z \ , \quad x \mapsto (g \circ f)(x) := g(f(x)) \ .$$

Satz 1.2 *Die Komposition von Abbildungen ist assoziativ, d.h. für Abbildungen $f : X \to Y$, $g : Y \to Z$ und $h : Z \to W$ gilt*

$$h \circ (g \circ f) = (h \circ g) \circ f : X \to W \ .$$

Beweis. Für $x \in X$ gilt

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

□

1.3 Zahlen

Die einfachste unendliche Menge ist die Menge

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

der *natürlichen Zahlen*. Man kann sie wie folgt einführen:

Definition 1.3 Die natürlichen Zahlen bilden eine Menge $\mathbb{N}$, in der ein Element $0 \in \mathbb{N}$ ausgezeichnet ist und auf der eine Selbstabbildung $S : \mathbb{N} \to \mathbb{N}$, die Nachfolgerabbildung, definiert ist, so daß die folgenden Axiome erfüllt sind:

(S1) $S : \mathbb{N} \to \mathbb{N}$ ist injektiv

(S2) $0 \notin S(\mathbb{N})$

(S3) Wenn eine Teilmenge $M \subset \mathbb{N}$ das Element 0 enthält und durch S in sich abgebildet wird (d.h. $S(M) \subset M$), dann ist $M = \mathbb{N}$.

Das ist die mengentheoretische Beschreibung des Zählens. Die Vorstellung ist, daß S jeder natürlichen Zahl n ihren Nachfolger zuordnet, also $1 := S(0)$, $2 := S(1)$ usw. (S1) präzisiert, daß man beim Zählen nicht mehrmals auf dieselbe Zahl stößt. (S2) besagt, daß 0 der Ausgangspunkt ist, der durch den Zählprozeß nicht erreicht wird (damit ist S nicht surjektiv). (S3) beinhaltet das *Prinzip der vollständigen Induktion*: Wenn eine Eigenschaft E der Zahl 0 zukommt (Induktionsanfang) und für jede Zahl n, welche die Eigenschaft E hat, auch der Nachfolger die Eigenschaft E hat (Induktionsschluß), dann kommt diese Eigenschaft allen natürlichen Zahlen zu.

Mittels der Abbildung S lassen sich rekursiv

- *Addition* durch $m + 0 := m$ und $m + S(n) := S(m + n)$
- *Multiplikation* durch $m \cdot 0$ und $m \cdot S(n) := m \cdot n + m$

definieren, für $m, n \in \mathbb{N}$.

Die Menge der *ganzen Zahlen* $\mathbb{Z}$ kann man sich als Differenzen $m - n$ von natürlichen Zahlen m, n vorstellen. Dabei ist zu beachten, daß die gleiche ganze Zahl auf verschiedene Weise als Differenz darstellbar ist. Die korrekte mengentheoretische Konstruktion geht wie folgt: Wir betrachten auf der Menge $\mathbb{N} \times \mathbb{N}$ die Relation

$$(m, n) \sim (k, l) \quad \Leftrightarrow \quad m + l = n + k\ .$$

Man zeigt, daß $\sim$ eine Äquivalenzrelation ist. Dann ist $\mathbb{Z} = (\mathbb{N} \times \mathbb{N})/\sim$. Eine ganze Zahl ist also eine Äquivalenzklasse, wobei die Äquivalenzrelation gerade der naiven Differenzbildung entspricht. Addition, Multiplikation und Negativbildung lassen sich leicht für die Paare (m, n) und deren Äquivalenzklassen definieren.

Die Menge der *rationalen Zahlen* $\mathbb{Q}$ wird durch eine analoge Konstruktion erhalten, indem man den Bruch $\frac{m}{n}$ korrekt definiert. Wir betrachten auf der Menge $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ die Relation

$$(p, q) \sim (r, s) \quad \Leftrightarrow \quad p \cdot s = q \cdot r\ .$$

Man zeigt, daß $\sim$ eine Äquivalenzrelation ist. Dann ist $\mathbb{Q} = (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}))/\sim$, und die Äquivalenzrelation entspricht gerade der naiven Quotientenbildung.

Die reellen Zahlen $\mathbb{R}$ werden in der Analysis konstruiert. Zunächst betrachtet man Folgen $(r_n) = \{r_0, r_1, \dots\}$ rationaler Zahlen r_i, welche durch natürliche Zahlen $i \in \mathbb{N}$ indiziert sind. Eine solche Folge heißt *Cauchy-Folge* oder *Fundamentalfolge*, falls es zu jedem $\epsilon > 0$ (hier rational angenommen) einen Index $k \in \mathbb{N}$ gibt, so daß $|r_n - r_m| < \epsilon$ für alle $n, m \geq k$ gilt. Wir erklären eine Äquivalenzrelation für Fundamentalfolgen

$$(r_n) \sim (s_n) \quad \Leftrightarrow \quad (r_n - s_n) \text{ konvergiert gegen } 0 \,.$$

Dabei bedeutet "Konvergenz gegen 0", daß es zu jedem rationalen $\epsilon > 0$ einen Index $k \in \mathbb{N}$ gibt, so daß $|r_n - s_n| < \epsilon$ für alle $n \geq k$ gilt. Eine reelle Zahl ist nun eine Äquivalenzklasse von Fundamentalfolgen.

1.4 Gruppen

Unter einer *Verknüpfung* oder *Komposition* auf einer Menge G versteht man eine Vorschrift, die zwei gegebenen Elementen $a, b \in G$ ein neues Element $a * b \in G$ zuordnet, d.h. eine Abbildung

$$* : G \times G \to G \,, \qquad * : (a, b) \mapsto a * b := *(a, b) \,.$$

Beispiele sind die Addition $* = +$ und Multiplikation $* = \cdot$ in $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$. Ein weiteres wichtiges Beispiel ist die Menge $G = \mathrm{Abb}(X, X) = \{f : X \to X\}$ der Selbstabbildungen von X mit der Komposition $* = \circ$ als Verknüpfung, $f, g \in \mathrm{Abb}(X, X) \;\Rightarrow\; f \circ g \in \mathrm{Abb}(X, X)$.

Definition 1.4 Eine Menge G zusammen mit einer Verknüpfung $*$ heißt *Gruppe*, falls die folgenden Axiome erfüllt sind

- (G1) $(a * b) * c = a * (b * c)$ für alle $a, b, \in G$ (Assoziativgesetz)
- (G2) Es gibt ein neutrales Element $e \in G$ mit $e * a = a$ für alle $a \in G$
- (G3) Zu jedem $a \in G$ gibt es ein $a^{-1} \in G$ (das zu a inverse Element) mit $a^{-1} * a = e$

Die Gruppe heißt *kommutativ* (oder *abelsch*), falls außerdem $a * b = b * a$ für alle $a, b \in G$.

Satz 1.3 *Ist G eine Gruppe, so gilt:*

i) *Das neutrale Element ist eindeutig bestimmt, und außerdem gilt $a * e = a$ für alle $a \in G$.*

ii) *Das inverse Element a^{-1} ist für jedes $a \in G$ eindeutig bestimmt, und außerdem gilt $a * a^{-1} = e$ sowie $(a^{-1})^{-1} = a$ und $(ab)^{-1} = b^{-1} * a^{-1}$.*

iii) *Es gelten die Kürzungsregeln* $a * b' = a * b \;\Rightarrow\; b = b'$ *und* $b' * a = b * a \;\Rightarrow\; b = b'$.

Beweis. Sei $e \in G$ eines der neutralen Elemente und $a \in G$. Zum Inversen $a^{-1} \in G$ gibt es wieder ein Inverses $(a^{-1})^{-1} \in G$, und es gilt

$$\begin{aligned} a * a^{-1} &= e * (a * a^{-1}) = ((a^{-1})^{-1} * a^{-1}) * (a * a^{-1}) = (((a^{-1})^{-1} * a^{-1}) * a) * a^{-1} \\ &= ((a^{-1})^{-1} * (a^{-1} * a)) * a^{-1} = ((a^{-1})^{-1} * e) * a^{-1} = (a^{-1})^{-1} * (e * a^{-1}) \\ &= (a^{-1})^{-1} * a^{-1} = e \end{aligned}$$

und daraus

$$a * e = a * (a^{-1} * a) = (a * a^{-1}) * a = e * a = a \;.$$

Sei $\tilde{e} \in G$ ein weiteres neutrales Element, so gilt aus Sicht von e die Gleichung $e * \tilde{e} = \tilde{e}$ und aus Sicht von $\tilde{e}$ die Gleichung $e * \tilde{e} = e$, also $e = \tilde{e}$. Damit ist i) bewiesen und $a * a^{-1} = e$ aus ii).

Sei $\widetilde{a^{-1}}$ ein weiteres Inverses zu a, so gilt

$$\widetilde{a^{-1}} = \widetilde{a^{-1}} * e = \widetilde{a^{-1}} * (a * a^{-1}) = (\widetilde{a^{-1}} * a) * a^{-1} = e * a^{-1} = a^{-1} \;,$$

also ist das Inverse eindeutig. Damit ist wegen $a * a^{-1} = e$ das Inverse von a^{-1} durch a selbst gegeben. Schließlich gilt

$$(b^{-1} * a^{-1}) * (a * b) = ((b^{-1} * a^{-1}) * a) * b = (b^{-1} * (a^{-1} * a)) * b = (b^{-1} * e) * b = b^{-1} * b = e$$

so daß $b^{-1} * a^{-1}$ das Inverse zu $a * b$ ist.

Die Kürzungsregeln folgen nach Verknüpfung von links/rechts mit a^{-1} unter Verwendung der Assoziativität. □

Beispiele 1.1 (für Gruppen)

- $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ mit der Addition als Verknüpfung, mit $e = 0$ als neutralem Element und dem Negativem $a \mapsto -a$ als Inversen.
- $\mathbb{Q}^* := \mathbb{Q} \setminus \{0\}$ und $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$ mit der Multiplikation als Verknüpfung und $e = 1$ als neutralem Element.
- $S(X) := \{f \in \mathrm{Abb}(X, X) \;:\; f \text{ ist bijektiv}\}$ mit der Komposition $\circ$ als Verknüpfung und der identischen Abbildung $\mathrm{id}_X : x \mapsto x$ als neutralem Element. Diese Gruppe heißt die *symmetrische Gruppe* der Menge X. Falls $X = \{1, 2, \dots, n\}$, so schreibt man $S_n := S(\{1, 2, \dots, n\})$. Die Elemente $\sigma \in S_n$ sind *Permutationen* der Zahlen $1, \dots, n$.
- Drehungen eines starren Körpers um seinen festgehaltenen Schwerpunkt im $\mathbb{R}^3$.

- Drehungen eines unendlichen Kristalls um den Winkel $\frac{2\pi}{n}$ um eine geeignete Achse im Kristall, so daß der Kristall in sich selbst überführt wird.

Definition 1.5 Sei G eine Gruppe mit Verknüpfung $*$. Eine nichtleere Teilmenge $H \subset G$ heißt *Untergruppe*, wenn für alle $a, b \in H$ auch $a * b \in H$ und $a^{-1} \in H$ gilt.

Es folgt automatisch $e \in H$ (durch Wahl von $b = a^{-1}$). Ist G eine Gruppe, so sind die Teilmengen $H = G$ und $H = \{e\}$ Untergruppen. Sie heißen *triviale Untergruppen*.

Beispiele sind $\mathbb{Q}^* \subset \mathbb{R}^*$ bezüglich der Multiplikation oder G als Gruppe der Drehungen eines Körpers und H als Untergruppe der Drehungen um eine feste Achse.

Für das Studium von Untergruppen spielen die *Nebenklassen* eine wichtige Rolle.

Definition 1.6 Sei G eine Gruppe und H eine Untergruppe von G. Dann heißt die Menge $Hg := \{h * g \ : h \in H\}$ die *Rechtsnebenklasse* von G nach H und die Menge $gH := \{h * g \ : h \in H\}$ die *Linksnebenklasse* von G nach H. (Das Element $g \in G$ bleibt festgehalten.)

Satz 1.4 *Ist $g \in H$ und H Untergruppe von G, so ist $Hg = H$.*

Beweis. Nach Definition der Untergruppe ist $Hg \subset H$. Für beliebiges $h \in H$ ist $h * g^{-1} \in H$ und damit $h = (h * g^{-1}) * g \in Hg$. □

Satz 1.5 *Sei G eine Gruppe und H eine Untergruppe von G. Dann gilt für zwei Rechtsnebenklassen Hg und Hk von G nach H entweder $Hg = Hk$ oder $Hg \cap Hk = \emptyset$.*

Beweis. Angenommen, Hg und Hk haben ein gemeinsames Element, d.h es existieren $h_1, h_2 \in H$ mit $h_1 * g = h_2 * k$. Durch Verknüpfung mit h_2^{-1} von links und mit g^{-1} von rechts erhalten wir

$$h_2^{-1} * h_1 * g * g^{-1} = h_2^{-1} * h_2 * k * g^{-1} \ ,$$

d.h. $h_2^{-1} * h_1 = k * g^{-1}$. Die linke Seite dieser Gleichung ist Element von H und somit auch die rechte Seite $k * g^{-1}$. Bildung der Rechtsnebenklasse liefert $H(k*g^{-1}) = H$. Durch Verknüpfung mit g von rechts entsteht $H(k*g^{-1})*g = Hg$, also $Hk = Hg$. □

Für eine beliebige Menge M bezeichnen wir mit $|M|$ die Anzahl ihrer Elemente.

Satz 1.6 *Sei G eine Gruppe mit $|G| < \infty$ Elementen und H eine Untergruppe von G. Dann gilt für ein beliebiges $g \in G$ die Beziehung $|Hg| = |H|$, und insbesondere ist $|G|$ ein Vielfaches von $|H|$.*

Beweis. Sei $k \in Hg$, also $k = h * g$ für ein $h \in H$. Dann ist $h = k * g^{-1}$ eindeutig bestimmt. Wenn $H = \{h_1, \ldots, h_n\}$, dann sind also sämtliche $h_1 * g, \ldots, h_n * g \in Hg$ verschieden. Damit enthält Hg genauso viele Elemente wie H.

Jedes $g \in G$ tritt in mindestens einer Rechtsnebenklasse auf, z.B. in Hg (da $e \in H$). Wenn es in mehreren Rechtsnebenklassen auftritt, dann sind diese gleich. Deshalb zerfällt G in disjunkte nichtleere Rechtsnebenklassen. Ihre Anzahl sei r. Da jede Rechtsnebenklasse $|H|$ Elemente enthält, gilt $|G| = r|H|$. □

Insbesondere besitzt eine aus p Elementen bestehende Gruppe, wenn p eine Primzahl ist, nur die trivialen Untergruppen G und $\{e\}$.

Definition 1.7 Seien G und H Gruppen mit Verknüpfungen $*$ und $\cdot$. Eine Abbildung $\phi : G \to H$ heißt *Homomorphismus (von Gruppen)*, falls

$$\phi(a * b) = \phi(a) \cdot \phi(b) \qquad \text{für alle } a, b \in G\ .$$

Ein Homomorphismus heißt *Isomorphismus*, wenn er bijektiv ist.

Satz 1.7 *Sei $\phi : G \to H$ ein Homomorphismus von Gruppen, dann gilt*

i) *$\phi(e) = \epsilon$, wenn e und ϵ die neutralen Elemente von G und H sind.*

ii) *$\phi(a^{-1}) = \phi(a)^{-1}$ für alle $a \in G$.*

iii) *Ist ϕ ein Isomorphismus, so ist auch die Umkehrabbildung $\phi^{-1} : H \to G$ ein Gruppen-Homomorphismus.*

Beweis.

$$\epsilon \cdot \phi(a) = \phi(a) = \phi(e * a) = \phi(e) \cdot \phi(a)\ ,$$

daraus aus der Kürzungsregel $\epsilon = \phi(e)$. Weiter

$$\epsilon = \phi(e) = \phi(a^{-1} * a) = \phi(a^{-1}) \cdot \phi(a)\ ,$$

also $\phi(a^{-1}) = \phi(a)^{-1}$.

Sei ϕ nun bijektiv und $\alpha = \phi(a)$ und $\beta = \phi(b)$, dann gilt

$$\phi(a * b) = \phi(a) \cdot \phi(b) = \alpha \cdot \beta$$

und daraus

$$\phi^{-1}(\alpha \cdot \beta) = a * b = \phi^{-1}(\alpha) * \phi^{-1}(\beta)\ .$$ □

Beispiele 1.2 (für Gruppenhomomorphismen)

- $G = \mathbb{R}$ mit der Addition als Verknüpfung, $H = \mathbb{R}_+ := \{r \in \mathbb{R}\ ,\ r > 0\}$ mit der Multiplikation als Verknüpfung, und $\phi = \exp : \mathbb{R} \to \mathbb{R}_+$. Es gilt $\exp(a + b) = \exp(a) \cdot \exp(b)$. Dann folgt automatisch $\exp(0) = 1$ und $\exp(-a) = \frac{1}{\exp(a)}$.

  Da exp bijektiv ist, ist auch $\ln := \exp^{-1} : \mathbb{R}_+ \to \mathbb{R}$ ein Homomorphismus. Es gilt $\ln(x \cdot y) = \ln(x) + \ln(y)$. Es folgt automatisch $\ln(1) = 0$ und $\ln(\frac{1}{a}) = -\ln(a)$.

- Wir wählen $G = \mathbb{Z}$ mit der Addition als Verknüpfung. Für jedes $m \in \mathbb{Z}$ ist die Abbildung

  $$\phi_m : \mathbb{Z} \to \mathbb{Z} \,, \qquad \phi_m : a \mapsto m \cdot a$$

  ein Homomorphismus wegen $\phi_m(a)+\phi_m(b) = m \cdot a + m \cdot b = m \cdot (a+b) = \phi_m(a+b)$. Sein Bild

  $$m\mathbb{Z} := \{m \cdot a \ : a \in \mathbb{Z}\}$$

  ist eine Untergruppe von $\mathbb{Z}$, da $(m \cdot a) + (m \cdot b) = m \cdot (a+b) \in m\mathbb{Z}$ und $-(m \cdot a) = m \cdot (-a) \in m\mathbb{Z}$.

Zum Abschluß ein wichtiges Beispiel für Gruppen. Sei $m \in \mathbb{Z}$ mit $m > 0$. Wir definieren eine Äquivalenzrelation für $a, b \in \mathbb{Z}$

$$a \sim b \quad \Leftrightarrow \quad a - b \in m\mathbb{Z} \,.$$

Diese zerlegt $\mathbb{Z}$ in m disjunkte Teilmengen

$$[r]_m := \{r + m \cdot n \ : \ n \in \mathbb{Z}\} \,, \qquad \text{für } r = \{0, 1, \dots, m-1\} \,.$$

Die Äquivalenzklassen $[r]_m$ heißen *Restklassen*, da jedes Element $a \in [r]_m$ bei der Division durch m den Rest r läßt. Wenn m festgelegt ist, schreibt man auch $\bar{r}$ statt $[r]_m$. Für die Quotientenmenge führen wir die folgende Bezeichnung ein:

$$\mathbb{Z}/m\mathbb{Z} := \mathbb{Z}/\sim \ = \{[0]_m, [1]_m, \dots, [m-1]_m\} \,.$$

Am bekanntesten ist die Situation für $m = 10$, wo die Restklasse bei natürlichen Zahlen durch die letzte Ziffer gegeben ist, z.B. $27, 197, -13 \in [7]_{10}$ oder $4, 1024, -26 \in [4]_{10}$. Wichtig ist auch $m = 2$, wo $[0]_2$ alle geraden Zahlen enthält und $[1]_2$ alle ungeraden.

Die einzelnen Elemente $a \in [r]_m$ sind die Repräsentanten der Restklasse. Die Beschränkung auf $r = \{0, 1, \dots, m-1\}$ ist nur eine sinnvolle Bezeichnung; die Definition liefert auch $[r + m \cdot n]_m = [r]_m$ für beliebiges $n \in \mathbb{Z}$. Dann können wir eine Addition für die Restklassen definieren als

$$[a]_m + [b]_m := [a+b]_m \,.$$

Wir haben z.B. $[7]_{10} + [4]_{10} = [1]_{10}$. Zu zeigen wäre, daß diese Definition unabhängig von der Wahl der Repräsentanten ist. Ein Beispiel soll genügen: $[27]_{10} + [4]_{10} = [31]_{10} = [1]_{10}$ oder auch $[197]_{10} + [-26]_{10} = [171]_{10} = [1]_{10}$. Das Beispiel zeigt, daß die Menge $\mathbb{Z}/m\mathbb{Z}$ mit dieser Addition, mit $[r]_m^{-1} := [-r]_m$ als Inversem und $e = [0]_m$ als neutralem Element eine kommutative Gruppe bildet, die als *zyklische Gruppe der Ordnung m* bezeichnet wird.

1.5 Ringe

Bei Gruppen hat man eine einzige Verknüpfung, z.B. Addition oder Multiplikation. Im weiteren brauchen wir Mengen, deren Elemente mittels Addition *und* Multiplikation verknüpft werden können (z.B. die Zahlen). Solche Mengen (mit brauchbaren Eigenschaften) heißen Ringe.

Definition 1.8 Eine Menge R zusammen mit zwei Verknüpfungen

$$\begin{aligned} + :& R \times R \to R\,, \qquad && + : (a,b) \mapsto a+b \qquad && (Addition) \\ \cdot :& R \times R \to R\,, \qquad && \cdot : (a,b) \mapsto a \cdot b \qquad && (Multiplikation) \end{aligned}$$

heißt *Ring*, wenn folgendes gilt:

(R1) R zusammen mit der Addition ist eine kommutative Gruppe.

(R2) Die Multiplikation ist assoziativ.

(R3) Es gelten die Distributivgesetze

$$a \cdot (b+c) = a \cdot b + a \cdot c \qquad \text{und} \qquad (a+b) \cdot c = a \cdot c + b \cdot c$$

für alle $a, b, c \in R$.

Ein Ring heißt kommutativ, falls $a \cdot b = b \cdot a$ für alle $a, b \in R$. Ein Element $1 \in R$ heißt Einselement, wenn $1 \cdot a = a \cdot 1 = a$ für alle $a \in R$.

In (R3) ist die Konvention benutzt, daß die Multiplikation stärker bindet als die Addition.

Ist R ein Ring und $0 \in R$ das Nullelement, d.h. das neutrale Element der kommutativen Gruppe $(R, +)$, dann gilt $0 \cdot a = (0+0) \cdot a = 0 \cdot a + 0 \cdot a$, also $0 \cdot a = 0$ und analog $a \cdot 0 = 0$.

Beispiele 1.3 (für Ringe)

- die Zahlbereiche $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$,
- die Menge der reellwertigen stetigen Funktionen über einem Intervall,

  $$R = \{f : I \to \mathbb{R}\}\,, \qquad I \subset \mathbb{R}$$

  mit

  $$(f_1 + f_2)(x) := f_1(x) + f_2(x) \qquad \text{und} \qquad (f_1 \cdot f_2)(x) := f_1(x) \cdot f_2(x)\,.$$

- Die Gruppe $(\mathbb{Z}/m\mathbb{Z}, +)$ ist ebenfalls ein Ring mit der Multiplikation $[a]_m \cdot [b]_m := [a \cdot b]_m$.

Definition 1.9 Ein Ring R heißt *nullteilerfrei*, wenn aus $a \cdot b = 0$ stets folgt $a = 0$ oder $b = 0$.

Satz 1.8 *Der Restklassenring $\mathbb{Z}/m\mathbb{Z}$ ist genau dann nullteilerfrei, wenn m eine Primzahl ist.*

Beweis. Ist m keine Primzahl, so gilt $m = k \cdot l$ mit $k, l \neq m, 0$. Dann ist $[k]_m \neq [0]_m$ und $[l]_m \neq [0]_m$, aber $[k \cdot l]_m = [m]_m = [0]_m$. Ist andererseits m eine Primzahl und $[k \cdot l]_m = [0]_m$, so gilt $k \cdot l = m \cdot r$ für $r \in \mathbb{Z}$, so daß k oder l durch m teilbar sind, also $[k]_m = [0]_m$ oder $[l]_m = [0]_m$. □

Die für Gruppen kennengelernten Konzepte von Untergruppen und Homomorphismen von Gruppen übertragen sich auf Ringe:

Definition 1.10 Ist R ein Ring und $R' \subset R$ eine Teilmenge, so heißt R' *Unterring*, wenn R' bezüglich der Addition eine kommutative Untergruppe ist und bezüglich der Multiplikation gilt $a' \cdot b' \in R'$ für alle $a', b' \in R$.

Sind R und S Ringe mit Verknüpfungen $+, \cdot$ bzw. $\oplus, \odot$, so heißt die Abbildung $\phi : R \to S$ ein *Homomorphismus von Ringen*, falls für alle $a, b \in R$ gilt

$$\phi(a+b) = \phi(a) \oplus \phi(b) \quad \text{und} \qquad \phi(a \cdot b) = \phi(a) \odot \phi(b) \ .$$

Beispiele für Unterringe sind $\mathbb{Z} \subset \mathbb{Q} \subset R$ und $m\mathbb{Z} \subset \mathbb{Z}$. Die Abbildung $\phi : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$, $\phi : a \mapsto [a]_m$ ist ein Beispiel für einen Ringhomomorphismus.

1.6 Körper

Wenn in einem Ring auch die Multiplikation zu einer Gruppe wird (abgesehen vom Nullelement), sprechen wir von einem Körper.

Definition 1.11 Eine Menge K zusammen mit zwei Verknüpfungen

$$\begin{aligned} + :& K \times K \to K \ , \qquad + : (a,b) \mapsto a+b \qquad && (\textit{Addition}) \\ \cdot :& K \times K \to K \ , \qquad \cdot : (a,b) \mapsto a \cdot b \qquad && (\textit{Multiplikation}) \end{aligned}$$

heißt *Körper*, wenn folgendes gilt:

(K1) K zusammen mit der Addition ist eine kommutative Gruppe. Ihr neutrales Element wird mit 0 bezeichnet und das zu $a \in K$ inverse Element mit $-a$.

(K2) Sei $K^* := K \setminus \{0\}$, dann ist für $a, b \in K^*$ auch $a \cdot b \in K^*$, und K^* mit dieser Multiplikation ist eine kommutative Gruppe. Ihr neutrales Element wird mit 1 bezeichnet und das zu $a \in K^*$ inverse Element mit $a^{-1} = 1/a$. Man schreibt auch $a/b = a^{-1} \cdot b = b \cdot a^{-1}$.

(K3) Es gelten die Distributivgesetze

$$a \cdot (b+c) = a \cdot b + a \cdot c \qquad \text{und} \qquad (a+b) \cdot c = a \cdot c + b \cdot c$$

für alle $a, b, c \in K$.

Körper spielen eine zentrale Rolle in linearen Abbildungen. Beispiele für Körper sind $\mathbb{Q}$ und $\mathbb{R}$. In einem Körper gilt

i) $1 \neq 0$ (da $1 \in K^*$ und $0 \notin K^*$)

ii) $a \cdot 0 = 0 \cdot a = 0$ (wie zuvor)

iii) $a \cdot b = 0 \quad \Rightarrow \quad a = 0$ oder $b = 0$ (aus (K2))

iv) $a \cdot (-b) = -(a \cdot b)$ und $(-a) \cdot (-b) = a \cdot b$

v) $a \cdot b = a \cdot c$ und $a \neq 0 \quad \Rightarrow \quad b = c$

Satz 1.9 *Ist $p \in \mathbb{N}$ eine Primzahl, so ist $\mathbb{Z}_p := \mathbb{Z}/p\mathbb{Z}$ ein Körper.*

Beweis. Zu zeigen ist, daß $K^* := (\mathbb{Z}/p\mathbb{Z}) \setminus \{[0]_p\}$ eine kommutative Gruppe ist. Kommutativität und Assoziativität folgt aus der Definition. Das neutrale Element ist $[1]_p$. Bleibt die Existenz des inversen Elements. Das kann auf verschiedene Weise bewiesen werden. Wir geben das Inverse explizit an durch den

Kleinen Satz von Fermat *Ist p eine Primzahl und $a \in \mathbb{Z}$, so ist $a^p - a$ durch p teilbar.*

Ist a selbst nicht durch p teilbar, so muß $a^{p-1} - 1$ durch p teilbar sein. Das bedeutet $([a]_p)^{-1} = ([a]_p)^{p-2} = \underbrace{[a]_p \cdots [a]_p}_{(p-2) \text{ mal}}$.

Beweis des kleinen Satzes von Fermat. Wir können $a \in \mathbb{N}$ annehmen, denn für ungerades p gilt $(-a)^p - (-a) = -(a^p - a)$, während für $p = 2$ gilt $(-a)^2 - (-a) = a^2 - a + 2a$.

Der kürzeste Beweis verwendet die vollständige Induktion und die binomische Formel. Die Aussage ist richtig für $a = 0$, da $0^p - 0 = 0$ durch p teilbar ist. Angenommen, $a^p - a$ sei durch p teilbar. Dann betrachten wir

$$\begin{aligned}
&(a+1)^p - (a+1) \\
&= \Big(a^p + \underbrace{\binom{p}{1} a^{p-1} + \binom{p}{2} a^{p-2} + \cdots + \binom{p}{p-1} a}_{\text{durch } p \text{ teilbar}} + 1\Big) - (a+1)\,.
\end{aligned}$$

Die Binomialkoeffizienten $\binom{p}{k} = \frac{p(p-1)\cdots(p-k+1)}{1 \cdot 2 \cdots k}$ sind ganze Zahlen und somit für $0 < k < p$ durch p teilbar. Damit folgt die Teilbarkeit von $(a+1)^p - (a+1)$ durch p aus der Teilbarkeit von $a^p - a$ durch p. $\square$

Ist z.B. $p = 5$, so haben wir $[2]_5^{-1} = [2^3]_5 = [3]_5$ und $[4]_5^{-1} = [4^3]_5 = [4]_5$. Da $(\mathbb{Z}/5\mathbb{Z}) \setminus \{[0]_5\}$ bezüglich der Multiplikation eine aus 4 Elementen bestehende Gruppe ist, kann es Untergruppen aus zwei Elementen geben. Eine solche ist $H = \{[1]_5, [4]_5\}$.

Ein besonders wichtiger Körper ist der Körper $\mathbb{C}$ der komplexen Zahlen. Wir definieren für die Menge $\mathbb{R} \times \mathbb{R}$ eine Addition durch

$$(a, b) + (c, d) := (a + c, b + d)$$

und eine Multiplikation durch

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

Offensichtlich ist $(0, 0)$ das neutrale Element der Addition, das Negative zu (a, b) ist $(-a, -a)$, und $(1, 0)$ ist das Einselements bezüglich der Multiplikation. Das multiplikative Inverse zu $(a, b) \neq (0, 0)$ ist

$$(a, b)^{-1} := \Big(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\Big) .$$

Durch direktes (z.T. etwas längliches) Nachrechnen bestätigt man, daß $\mathbb{R} \times \mathbb{R}$ mit diesen Verknüpfungen ein Körper ist. Er heißt *Körper der komplexen Zahlen* und wird mit $\mathbb{C}$ bezeichnet.

Die durch $\iota : a \mapsto (a, 0)$ definierte Abbildung $\iota : \mathbb{R} \to \mathbb{C}$ ist injektiv. Wegen $(a, 0) + (b, 0) = (a + b, 0)$ und $(a, 0) \cdot (b, 0) = (ab, 0)$ kann man deshalb $\mathbb{R}$ mit seinem Bild $\iota(\mathbb{R}) = \mathbb{R} \times \{0\} \subset \mathbb{C}$ identifizieren und somit die reellen Zahlen als Unterkörper der komplexen Zahlen auffassen. Es zeigt sich, daß viele Beziehungen für reelle Zahlen sehr viel einfacher beweisen und verstehen kann, wenn man sie als Einschränkung von Beziehungen für komplexe Zahlen auf $\mathbb{R} \times \{0\}$ auffaßt.

Wegen $(a, b) = (a, 0) + (0, 1) \cdot (b, 0)$ erlaubt es die Identifikation von $\mathbb{R}$ mit $\mathbb{R} \times \{0\} \subset \mathbb{C}$, eine komplexe Zahl zu schreiben als $z = (a, b) = a + \mathrm{i}b$. Dabei heißt $\mathrm{i} := (0, 1) \in \mathbb{C}$ die *imaginäre Einheit*. Für sie gilt $\mathrm{i}^2 = (0, 1) \cdot (0, 1) = (-1, 0) = -1$. Für $z = a + \mathrm{i}b \in \mathbb{C}$ heißt

- $\mathrm{Re}(z) := a \in \mathbb{R}$ der *Realteil* von z,
- $\mathrm{Im}(z) := b \in \mathbb{R}$ der *Imaginärteil* von z
- $\bar{z} := a - \mathrm{i}b$ die zu z *konjugiert komplexe Zahl*.

Die Abbildung $z \mapsto \bar{z}$ heißt *komplexe Konjugation*. Es gilt

$$\begin{aligned} \overline{z + w} &= \bar{z} + \bar{w} , \\ \overline{z \cdot w} &= \bar{z} \cdot \bar{w} , \\ z \in \mathbb{R} \quad &\Leftrightarrow \quad z = \bar{z} . \end{aligned}$$

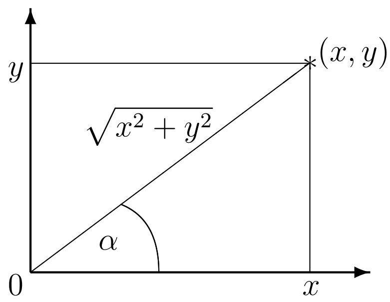
Für $z = a + \mathrm{i}b$ ist $z \cdot \bar{z} = (a + \mathrm{i}b) \cdot (a - \mathrm{i}b) = a^2 + b^2$ stets eine nichtnegative reelle Zahl. Es läßt sich deshalb die Wurzel ziehen, und wir definieren

$$|z| := \sqrt{z \cdot \bar{z}} = \sqrt{a^2 + b^2} \in \mathbb{R}_+$$

als den *Betrag* von z. Für beliebige $z, w \in \mathbb{C}$ gilt

$$\begin{aligned} |z+w| &\leq |z|+|w| \qquad \text{(Dreiecksungleichung)} \\ |z \cdot w| &= |z| \cdot |w| \\ |\bar{z}| &= |z| \end{aligned}$$

Es ist nun sehr intuitiv, eine komplexe Zahl $z = x + \mathrm{i}y$ als Punkt (x, y) in einem kartesischen Koordinatensystem in der Ebene (Gaußsche Zahlenebene) darzustellen. Nach Pythagoras ist dann $|z|$ gerade der Abstand dieses Punktes vom Ursprung $0 = (0, 0)$.



Real- und Imaginärteil ergeben sich zu $x = |z| \cos\alpha$ und $y = |z| \sin\alpha$, so daß wir folgende Darstellung einer komplexen Zahl erhalten:

$$z = |z|(\cos\alpha + \mathrm{i}\sin\alpha) \;, \qquad 0 \leq \alpha < 2\pi \;.$$

Es gilt Eulers Formel

$$\cos\alpha + \mathrm{i}\sin\alpha = \mathrm{e}^{\mathrm{i}\phi} \;,$$

wobei die Exponentialfunktion durch die Reihe

$$\mathrm{e}^z := \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

definiert wird. In der Analysis wird bewiesen, daß diese Reihe konvergiert. Es gilt $\mathrm{e}^z \cdot \mathrm{e}^w = \mathrm{e}^{z+w}$ und $\overline{\mathrm{e}^z} = \mathrm{e}^{\bar{z}}$. Daraus folgt für $\alpha \in \mathbb{R}$

$$\mathrm{e}^{\mathrm{i}\alpha}\mathrm{e}^{-\mathrm{i}\alpha} = \mathrm{e}^0 = 1 \qquad \text{und} \qquad 1 = \left|\mathrm{e}^{\mathrm{i}\alpha}\mathrm{e}^{-\mathrm{i}\alpha}\right| = \left|\mathrm{e}^{\mathrm{i}\alpha}\right| \cdot \left|\mathrm{e}^{-\mathrm{i}\alpha}\right| = \left|\mathrm{e}^{\mathrm{i}\alpha}\right|^2$$

also $\left|\mathrm{e}^{\mathrm{i}\alpha}\right| = 1$ für alle $\alpha \in \mathbb{R}$. Das bedeutet, daß es für jede komplexe Zahl z, welche in der Gaußschen Zahlenebene auf dem Einheitskreis um den Ursprung 0 mit Radius 1 liegt, ein $\alpha \in \mathbb{R}$ gibt mit $z = \mathrm{e}^{\mathrm{i}\alpha}$. Tatsächlich sind $\cos\alpha$ und $\sin\alpha$ gerade als Real- und Imaginärteil von $\mathrm{e}^{\mathrm{i}\alpha}$ definiert. Der Winkel $\alpha = \arg(z)$ heißt das *Argument* der komplexen Zahl $z = |z|\mathrm{e}^{\mathrm{i}\alpha}$. Es gilt $\arg(z) = \arctan\frac{\mathrm{Im}(z)}{\mathrm{Re}(z)}$.

2 Vektorräume und lineare Abbildungen

2.1 Vektorräume

Vektorräume sind der zentrale Gegenstand der linearen Algebra.

Definition 2.1 Sei K ein Körper. Eine Menge V zusammen mit einer inneren Verknüpfung $+ : V \times V \to V$ (der Addition) und einer äußeren Verknüpfung $\cdot : K \times V \to V$ (Multiplikation mit Skalaren) heißt *Vektorraum über K*, wenn

(V1) V zusammen mit der Addition ist eine kommutative Gruppe. (Wie üblich wird das neutrale Element wird mit $0 \in V$ bezeichnet und das zu $v \in V$ inverse Element mit $-v$.)

(V2) Für die Multiplikation mit Skalaren gilt

$$
\begin{aligned}
(\lambda + \mu) \cdot v &= \lambda \cdot v + \mu \cdot v \ , \qquad & \lambda \cdot (v + w) &= \lambda \cdot v + \lambda \cdot w \ , \\
\lambda \cdot (\mu \cdot v) &= (\lambda\mu) \cdot v \ , & 1 \cdot v &= v \ ,
\end{aligned}
$$

für alle $\lambda, \mu \in K$ und $v, w \in V$.

Ein Element eines Vektorraumes V heißt *Vektor*.

Die wichtigsten Fälle sind Vektorräume über $K = \mathbb{R}$ und $K = \mathbb{C}$. Wir sprechen dann von reellen bzw. komplexen Vektorräumen.

Satz 2.1 *In einem Vektorraum über K gilt*

i) $0 \cdot v = 0 \in V \quad \forall v \in V$

ii) $\lambda \cdot 0 = 0 \in V \quad \forall \lambda \in K$

iii) $(-1) \cdot v = -v \quad \forall v \in V$

iv) $\lambda \cdot v = 0 \in V \quad \Leftrightarrow \quad \lambda = 0$ *oder* $v = 0$

Beweis. Beweis von i),ii),iii) durch gleiche Methoden wie bei Ringen.
iv): Ist $\lambda \cdot v = 0$, aber $\lambda \neq 0$, so gilt $v = 1 \cdot v = (\lambda^{-1}\lambda) \cdot v = \lambda^{-1} \cdot (\lambda \cdot v) = 0$. □

Beispiele 2.1 (für Vektorräume)

- $K^n = \{x = (x_1, \ldots, x_n) \ : \ x_i \in K\}$ ist ein Vektorraum mit Addition

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

und skalarer Multiplikation

$$\lambda \cdot (x_1, \ldots, x_n) := (\lambda x_1, \ldots, \lambda x_n) \ .$$

- Die Menge $M(m \times n, K) := \left\{ \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} : a_{ij} \in K \right\}$ der *Matrizen* mit m Zeilen und n Spalten und Einträgen aus K ist ein Vektorraum mit komponentenweiser Addition und skalarer Multiplikation: Schreiben wir $A = (a_{ij}), B = (b_{ij}) \in M(m \times n, K)$, so ist

$$A + B := (a_{ij} + b_{ij}) , \qquad \lambda \cdot A := (\lambda a_{ij}) .$$

  Es ist $M(m \times n, K) = K^{m \cdot n}$, jedoch ist die Schreibweise in rechteckiger Anordnung im weiteren sehr nützlich.

- $\mathbb{C}$ kann als reeller Vektorraum aufgefaßt werden durch $\cdot : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$, $(\lambda, x + \mathrm{i}y) \mapsto \lambda x + \mathrm{i}\lambda y$.

Definition 2.2 Sei V ein Vektorraum. Eine Teilmenge $W \subset V$ heißt *Untervektorraum*, wenn

(UV1) $W \neq \emptyset$

(UV2) $v, w \in W \quad \Rightarrow \quad v + w \in W$ (d.h. W ist abgeschlossen bezüglich der Addition)

(UV3) $v \in W\ ,\ \lambda \in K \quad \Rightarrow \quad \lambda \cdot v \in W$ (d.h. W ist abgeschlossen bezüglich der Multiplikation mit Skalaren)

Ein Untervektorraum ist automatisch ein Vektorraum. Beispiele sind

- der Nullvektor $W = \{0\}$
- Vielfache $W = \{\lambda \cdot v \ : \ \lambda \in K\}$ eines ausgewählten Vektors $v \in V$, speziell jede Gerade in der Ebene durch den Nullpunkt, $W = \{(x, y) \in \mathbb{R}^2 \ : \ ax + by = 0\}$ für $a, b \in \mathbb{R}$.

Satz 2.2 *Seien W_i mit $i \in I$ (Indexmenge) jeweils Untervektorräume von V, so ist der Durchschnitt $W := \bigcap_{i \in I} W_i \subset V$ wieder ein Untervektorraum von V.*

Beweis. $0 \in W_i\ \forall i \in I$, also $0 \in W$ (damit ist W nicht leer). Seien $v, w \in W$, so sind $v, w \in W_i\ \forall i \in I$. Dann ist auch $v + w \in W_i\ \forall i \in I$ und somit $v + w \in W$. Analog ist $\lambda \cdot v \in W$. $\square$

Dagegen ist die Vereinigung von Untervektorräumen im allgemeinen nicht wieder ein Untervektorraum. Man kann eine solche Vereinigung aber zu einen Vektorraum abschließen.

Seien $v_1, \dots, v_r \in V$ (nicht notwendig verschiedene) Vektoren aus V. Ein Vektor $v \in V$ heißt *Linearkombination* von $v_1, \dots, v_r$, wenn es $\lambda_1, \dots, \lambda_r \in K$ gibt, so daß

$$v = \lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r = \sum_{i=1}^{r} \lambda_i \cdot v_i \ .$$

Wir bezeichnen mit

$$\operatorname{span}_K(v_1, \dots, v_r) = \left\{ \sum_{i=1}^{r} \lambda_i \cdot v_i \ : \ \lambda_i \in K \right\}$$

den Raum aller Linearkombinationen von $v_1, \dots, v_r$. Offenbar ist $\operatorname{span}_K(v_1, \dots, v_r)$ ein Untervektorraum von V, er heißt der durch $v_1, \dots, v_r$ *aufgespannte* (oder *erzeugte*) *Untervektorraum.*

Die Definition läßt sich verallgemeinern auf Untervektorräume, die von einer Familie $(v_i)_{i \in I}$ aus möglicherweise unendlich vielen Vektoren aufgespannt wird. Dann ist $\operatorname{span}_K(v_i)_{i \in I}$ definiert als die Menge aller *endlichen* Linearkombinationen, d.h. zu jedem $v \in \operatorname{span}_K(v_i)_{i \in I}$ gibt es Indizes $i_1, \dots, i_r \in I$ und Skalare $\lambda_1, \dots, \lambda_r \in K$, so daß $v = \sum_{j=1}^{r} \lambda_j v_{i_j}$.

Beispiele 2.2 (für Linearkombinationen)

- Für $V = \mathbb{R}^3$ und $v_1, v_2 \in V$ ist $\operatorname{span}_{\mathbb{R}}(v_1)$ die Gerade durch 0 und v_1, falls $v_1 \neq 0$, und $\operatorname{span}_{\mathbb{R}}(v_1, v_2)$ die Ebene durch $0, v_1, v_2$, falls $v_1 \neq 0$ und $v_2 \notin \operatorname{span}_{\mathbb{R}}(v_1)$.
- Im Vektorraum K^n setzen wir $e_i = (0, \dots, 0, 1, 0, \dots, 0)$, wobei die 1 an der i-ten Stelle steht. Dann ist $\operatorname{span}_K(e_i)_{i=1,\dots,n} = K^n$.

Im weiteren wird es wichtig sein, ob sich ein Vektor $v \in V$ in eindeutiger Weise als Linearkombination von vorgegebenen Vektoren $v_1, \dots, v_r$ darstellen läßt oder nicht.

Definition 2.3 Sei V ein Vektorraum über K. Eine Familie von endlich vielen Vektoren $v_1, \dots, v_r$ heißt *linear unabhängig* (bzw. die Vektoren $v_1, \dots, v_r$ heißen linear unabhängig), wenn aus $\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r = 0$ für $\lambda_1, \dots, \lambda_r \in K$ folgt $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$.

Eine beliebige Familie $(v_i)_{i \in I}$ heißt linear unabhängig, wenn jede endliche Teilfamilie linear unabhängig ist. Entsprechend heißt eine Familie $(v_i)_{i \in I}$, die nicht linear unabhängig ist, *linear abhängig.* In diesem Fall gibt es also $v_{i_1}, \dots, v_{i_r} \neq 0$ mit $i_1, \dots, i_r \in I$ und $\lambda_1, \dots, \lambda_r \neq 0$ mit $\sum_{j=1}^{r} \lambda_j \cdot v_{i_j} = 0$.

Satz 2.3 *Für eine Familie $(v_i)_{i \in I}$ sind folgende Bedingungen äquivalent:*

i) $(v_i)_{i \in I}$ *ist linear unabhängig*

ii) *Jeder Vektor* $v \in span_K(v_i)_{i\in I}$ *läßt sich in eindeutiger Weise als Linearkombination von Vektoren aus* $(v_i)_{i\in I}$ *darstellen.*

Beweis. i)⇒ii). Sei $v \in \mathrm{span}_K(v_i)_{i\in I}$ auf zwei verschiedene Arten als Linearkombination darstellbar,

$$v = \sum_{i\in I} \lambda_i \cdot v_i = \sum_{i\in I} \mu_i \cdot v_i$$

wobei nur endlich viele Skalare λ_i, μ_i ungleich Null sind. Es gibt also eine endliche Teilmenge $J \in I$ (Vereinigung der Indizes i, für die $\lambda_i \neq 0$ oder $\mu_i \neq 0$), so daß

$$\sum_{i\in J} (\lambda_i - \mu_i) \cdot v_i = 0 \ .$$

Da nach Voraussetzung i) jede endliche Teilfamilie von $(v_i)_{i\in I}$ linear unabhängig ist, ist $\lambda_i = \mu_i$ für alle $i \in J$ und weiter für alle $i \in I$ (auf dem Komplement $I \setminus J$ ist $\lambda_i = \mu_i = 0$).

ii)⇒i). Der Nullvektor läßt sich als Linearkombination von $(v_i)_{i\in I}$ darstellen, in der sämtliche Skalare Null sind. Ist die Linearkombination eindeutig, so ist $(v_i)_{i\in I}$ linear unabhängig. □

2.2 Erzeugendensystem, Basis und Dimension

Definition 2.4 Eine Familie $\mathcal{B} = (v_i)_{i\in I}$ von Vektoren $v_i \in V$ heißt *Erzeugendensystem* von V, wenn $V = \mathrm{span}_K(v_i)_{i\in I}$, wenn also jeder Vektor $v \in V$ eine endliche Linearkombination von $(v_i)_{i\in I}$ ist.

Ein Erzeugendensystem $\mathcal{B} = (v_i)_{i\in I}$ heißt *Basis* von V, wenn $\mathcal{B}$ linear unabhängig ist.

Der Vektorraum V heißt *endlich erzeugt*, falls es ein endliches Erzeugendensystem $\mathcal{B} = (v_1, \ldots, v_r)$ von V gibt.

Beispiele 2.3 (für Erzeugendensysteme und Basen)

- Im Vektorraum K^n ist $\mathcal{B} = (e_1, \ldots, e_n)$, mit $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, ein Erzeugendensystem, denn jeder Vektor $x = (x_1, \ldots, x_n)$ läßt sich schreiben als $x = \sum_{i=1}^n x_i e_i$ mit $x_i \in K$. Das Erzeugendensystem ist linear unabhängig und deshalb eine Basis, die *Standardbasis* des K^n.
- Im Vektorraum $M(m \times n, K)$ der $m \times n$-Matrizen ist $\mathcal{B} = (E_{ij})_{\substack{i=1,\ldots m\\ j=1,\ldots,n}}$ mit

$$E_{ij} = \begin{pmatrix} & & & 0 & & & \\ & O & & \vdots & & O & \\ & & & 0 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & 0 & & & \\ & O & & \vdots & & O & \\ & & & 0 & & & \end{pmatrix} ,$$

wobei die 1 im Schnittpunkt der i-ten Zeile mit der j-ten Spalte steht, ein Erzeugendensystem. Jede Matrix $A = (a_{ij})$ läßt sich darstellen als $A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} E_{ij}$. Das Erzeugendensystem ist linear unabhängig und deshalb eine Basis, die *Standardmatrixbasis*.

- In $\mathbb{C}$, aufgefaßt als reeller Vektorraum, ist $\mathcal{B} = (1, \mathrm{i})$ eine Basis, denn jede komplexe Zahl kann als $z = x \cdot 1 + y \cdot \mathrm{i}$ mit $x, y \in \mathbb{R}$ geschrieben werden.

  Fassen wir $\mathbb{C} = \mathbb{C}^1$ dagegen als komplexen Vektorraum auf, dann ist $\mathcal{B} = (1, \mathrm{i})$ zwar ein Erzeugendensystem, aber keine Basis mehr, denn 1 und i sind nicht mehr linear unabhängig über $\mathbb{C}$: $1 \cdot 1 + \mathrm{i} \cdot \mathrm{i} = 0$.

Ist $\mathcal{B} = (v_1, \ldots, v_n)$ eine endliches Erzeugendensystem bzw. eine endliche Basis, dann ist jede Permutation (Umordnung) der Vektoren v_i wieder ein endliches Erzeugendensystem bzw. eine endliche Basis. Z.B. ist $\mathcal{B} = (v_1, v_2, v_3)$ mit $v_1 = (1, 0, 0)$, $v_2 = (0, 0, 1)$, $v_3 = (0, 1, 0)$ eine Basis von K^3, aber es ist übersichtlicher, statt $\mathcal{B}$ mit der durch Umordnung erhaltenen Standardbasis (v_1, v_3, v_2) zu arbeiten.

Für die Definition einer Basis gibt es mehrere äquivalente Möglichkeiten. Wir beschränken uns zunächst auf endlich erzeugte Vektorräume:

Satz 2.4 *Für eine Familie $\mathcal{B} = (v_1, \ldots, v_n)$ von Vektoren $v_i \in V$, $V \neq \{0\}$, sind folgende Bedingungen äquivalent:*

i) *$\mathcal{B}$ ist ein linear unabhängiges Erzeugendensystem von V (also eine Basis).*

ii) *$\mathcal{B}$ ist ein Erzeugendensystem von V, aber $(v_1, \ldots, v_{r-1}, v_r, \ldots, v_n)$ ist für jedes $1 \leq r \leq n$ kein Erzeugendensystem mehr.*

iii) *Zu jedem $v \in V$ gibt es eindeutig bestimmte $\lambda_1, \ldots, \lambda_n \in K$ mit $v = \sum_{i=1}^{n} \lambda_i \cdot v_i$ (Eindeutigkeit der Zerlegung von v nach der Basis).*

iv) *$\mathcal{B}$ ist linear unabhängig, während $(v_1, \ldots, v_n, v)$ linear abhängig ist für alle $v \in V$.*

Beweis. i)⇒ii). Wäre $(v_1, \ldots, v_{r-1}, v_r, \ldots, v_n)$ ein Erzeugendensystem, dann gibt es $\lambda_1, \ldots, \lambda_{r-1}, \lambda_{r+1}, \ldots, \lambda_n \in K$ mit $v_r = \sum_{i=1}^{r-1} \lambda_i \cdot v_i + \sum_{j=r+1}^{n} \lambda_j \cdot v_j$. Damit ist $0 = \sum_{i=1}^{r-1} \lambda_i \cdot v_i + (-1) \cdot v_r + \sum_{j=r+1}^{n} \lambda_j \cdot$, so daß $\mathcal{B}$ nicht linear unabhängig wäre.

ii)⇒iii). Angenommen, es gibt ein $v \in V$ und für dieses zwei verschiedene Darstellungen

$$v = \sum_{i=1}^{n} \lambda_i \cdot v_i \qquad \text{und} \qquad v = \sum_{i=1}^{n} \mu_i \cdot v_i$$

mit $\lambda_r \neq \mu_r$ für mindestens ein r. (Es gibt mindestens eine Darstellung, da $\mathcal{B}$ Erzeugendensystem.) Subtraktion beider Gleichungen und Division durch $(\lambda_r -$

μ_r) ergibt

$$v_r = \sum_{i=1}^{r-1} \frac{\lambda_i - \mu_i}{\mu_r - \lambda_r} \cdot v_i + \sum_{j=r+1}^{n} \frac{\lambda_j - \mu_j}{\mu_r - \lambda_r} \cdot v_j$$

Das würde bedeuten, daß $(v_1, \ldots, v_{r-1}, v_r, \ldots, v_n)$ ein Erzeugendensystem wäre, denn in einer Linearkombination ließe sich v_r durch die v_i, v_j mit $i, j \neq r$ ausdrücken, im Widerspruch zu ii).

iii)$\Rightarrow$iv). Nach Satz 2.3 ist $\mathcal{B}$ linear unabhängig. Andererseits gibt es für jedes $v \in V$ Skalare $\lambda_1, \ldots, \lambda_n \in K$ mit $0 = \sum_{i=1}^n \lambda_i \cdot v_i + (-1) \cdot v$, so daß $(v_1, \ldots, v_r, v)$ nicht linear unabhängig ist.

iv)$\Rightarrow$i). Da $(v_1, \ldots, v_n, v)$ nicht linear unabhängig ist, gibt es $\lambda_1, \ldots, \lambda_n, \lambda \in K$ mit $0 = \sum_{i=1}^n \lambda_i \cdot v_i + \lambda \cdot v$. Da $\mathcal{B}$ linear unabhängig, ist $\lambda \neq 0$, so daß

$$v = \sum_{i=1}^{n} \frac{-\lambda_i}{\lambda} \cdot v_i \ .$$

Damit ist $\mathcal{B}$ ein Erzeugendensystem. $\square$

Die wichtigste Eigenschaft einer Basis ist wohl die Eindeutigkeit der Zerlegung eines gegebenen Vektors nach einer Basis des Vektorraums. Ist eine Basis fixiert, dann kann man an Stelle des Vektors $v \in V$ mit einer Folge von Zahlen $(\lambda_1, \ldots, \lambda_n)$ mit $\lambda_i \in K$ arbeiten. Diese Zahlen λ_i heißen die *Koordinaten* von $v = \sum_{i=1}^n \lambda_i \cdot v_i$ bezüglich der Basis $(v_1, \ldots, v_n)$.

Allgemeiner definiert jede Familie $(v_1, \ldots, v_n)$ von Vektoren aus V eine Abbildung $K^n \to V$ durch $(\lambda_1, \ldots, \lambda_n) \mapsto \sum_{i=1}^n \lambda_i \cdot v_i$. Diese Abbildung ist genau dann

- injektiv, wenn die v_i linear unabhängig sind,
- surjektiv, wenn die v_i ein Erzeugendensystem bilden,
- bijektiv, wenn die v_i eine Basis bilden.

Man kann also viele Probleme, die Vektorräume beinhalten, auf Probleme im Raum K^n der Koordinaten zurückführen. Damit das aber gelingt, ist es einerseits notwendig, erst einmal eine Basis zu identifizieren. Die äquivalenten Formulierungen aus Satz 2.4 sind dabei nützlich. Und dann ist zu klären, wie die Ergebnisse von der Wahl der Basis abhängen. Wir müssen uns dazu weitere Eigenschaften der Basen erarbeiten.

Satz 2.5 (Basisauswahlsatz) *Aus jedem endlichen Erzeugendensystem läßt sich eine Basis auswählen. Insbesondere hat jeder endlich erzeugte Vektorraum eine Basis.*

Beweis. Man nehme aus einem endlichen Erzeugendensystem einzelne Vektoren weg, so daß die reduzierte Familie immer noch ein Erzeugendensystem des Vektorraumes bleibt. Ist das nicht mehr möglich, so liegt eine Basis vor. $\square$

Satz 2.6 (Austauschlemma von Steinitz) *Ist $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis eines Vektorraumes V über K und w ein weiterer vom Nullvektor verschiedener Vektor aus V, so kann man einen der Basisvektoren gegen w "austauschen", d.h. es existiert ein Index $1 \leq r \leq n$, so daß $(v_1, \ldots, v_{r-1}, w, v_{r+1}, \ldots, v_n)$ ebenfalls eine Basis von V ist.*

Beweis. Da $\mathcal{B}$ eine Basis, gibt es $\lambda_1, \ldots, \lambda_n \in K$ mit $w = \sum_{i=1}^n \lambda_i \cdot v_i$. Da $w \neq 0$, gibt es mindestens einen Skalar $\lambda_r \neq 0$. Damit gilt

$$v_r = \frac{1}{\lambda_r} \cdot w + \sum_{i=1}^{r-1} \frac{-\lambda_i}{\lambda_r} \cdot v_i + \sum_{j=r+1}^{n} \frac{-\lambda_j}{\lambda_r} \cdot v_j \ .$$

Sei $v \in V$ ein beliebiger Vektor, dann gibt es $\mu_1, \ldots, \mu_n \in K$ mit $v = \sum_{i=1}^n \mu_i \cdot v_i$. Einsetzen von v_r liefert

$$v = \sum_{i=1}^{r-1} \left(\mu_i - \frac{\lambda_i}{\lambda_r}\right) \cdot v_i + \frac{\mu_r}{\lambda_r} \cdot w + \sum_{j=r+1}^{n} \left(\mu_j - \frac{\lambda_j}{\lambda_r}\right) \cdot v_j \ .$$

Damit ist $(v_1, \ldots, v_{r-1}, w, v_{r+1}, \ldots, v_n)$ ein Erzeugendensystem.

Zur Untersuchung der linearen Unabhängigkeit sei

$$0 = \sum_{i=1}^{r-1} \mu_i \cdot v_i + \mu \cdot w + \sum_{j=r+1}^{n} \mu_j \cdot v_j \ .$$

Einsetzen von w liefert

$$0 = \sum_{i=1}^{r-1} (\mu_i + \mu\lambda_i) \cdot v_i + \mu\lambda_r \cdot v_r + \sum_{j=r+1}^{n} (\mu_j + \mu\lambda_j) \cdot v_j \ .$$

Da $\mathcal{B}$ linear unabhängig, folgt $\mu_i + \mu\lambda_i = 0$ für $i \neq r$ und $\mu\lambda_r = 0$, also $\mu = 0$ und dann $\mu_i = 0$. □

Wir zeigen nun, daß alle Basen eines endlich erzeugten Vektorraumes aus der gleichen Anzahl an Vektoren bestehen.

Satz 2.7 (Austauschsatz) *In einem Vektorraum V über K sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis und $(w_1, \ldots, w_r)$ eine linear unabhängige Familie von Vektoren aus V. Dann ist $r \leq n$, und es gibt paarweise verschiedene Indizes $i_1, \ldots, i_r \in \{1, 2, \ldots, n\}$, so daß man nach Austausch von v_{i_j} durch w_j für alle $1 \leq j \leq r$ wieder eine Basis von V erhält. Nach Permutation der Indizes (Umnumerierung) erreicht man, daß $B^* = (w_1, \ldots, w_r, v_{r+1}, \ldots, v_n)$ eine Basis von V ist.*

Beweis. Wir tauschen schrittweise die Basisvektoren aus. Im ersten Schritt finden wir ein $1 \leq s \leq n$, so daß $(v_1, \ldots, v_{s-1}, w_1, v_{s+1}, \ldots, v_n)$ eine Basis ist. Nun nennen wir $v_1 \mapsto v_s$ und ordnen durch Vertauschen von $v_s \leftrightarrow w_1$ die Basis um, so daß $(w_1, v_2, \ldots, v_n)$ eine Basis ist.

Dann gibt es $\mu_1, \lambda_2, \ldots, \lambda_n \in K$ mit $w_2 = \mu_1 \cdot w_1 + \sum_{i=2}^n \lambda_i \cdot v_i$. Mindestens eines der λ_i ist von 0 verschieden, denn sonst wäre $w_2 = \mu_1 \cdot w_1$, und die Familie $(w_1, \ldots, w_r)$ wäre nicht linear unabhängig. Es läßt sich deshalb ein v_s mit $2 \leq s \leq n$ durch w_2 austauschen, so daß nach Umnumerierung $(w_1, w_2, v_3 \ldots, v_n)$ eine Basis von V ist.

Ist $r \leq n$, so führt die wiederholte Anwendung dieses Verfahrens auf eine Basis $(w_1, \ldots, w_r, v_{r+1}, \ldots, v_n)$. Wäre $r > n$, dann erhalten wir im n-ten Schritt eine Basis $(w_1, \ldots, w_n)$ von V. Dann ist w_{n+1} nach dieser Basis zerlegbar, so daß $(w_1, \ldots, w_r)$ für $r > n$ nicht mehr linear unabhängig wäre. □

Satz 2.8 *Sei V ein endlich erzeugter Vektorraum. Dann besteht jede Basis von V aus der gleichen Anzahl von Vektoren.*

Beweis. Seien zwei Basen $\mathcal{B}_1 = (v_1, \ldots, v_n)$ und $\mathcal{B}_2 = (w_1, \ldots, w_r)$ von V gegeben. Ist $n < r$, dann wäre nach Satz 2.7 $\mathcal{B}_2$ nicht linear unabhängig, also keine Basis. Ist $n > r$, dann wäre nach Satz 2.7 $\mathcal{B}_1$ nicht linear unabhängig, also keine Basis. Folglich gilt $n = r$. □

Definition 2.5 In einem Vektorraum V über K heißt die durch

$$\dim_K V := \begin{cases} 0 & \text{falls } V = \{0\}, \\ n & \text{falls } V \text{ endlich erzeugt ist und eine aus } n \text{ Vektoren} \\ & \text{bestehende Basis besitzt,} \\ \infty & \text{falls } V \text{ nicht endlich erzeugt ist} \end{cases}$$

definierte Zahl die *Dimension* von V.

Offenbar ist $\dim_K(K^n) = n$ und $\dim_{\mathbb{R}}(\mathbb{C}) = 2$, aber $\dim_{\mathbb{C}}(\mathbb{C}) = 1$. Wir schreiben auch $\dim(V)$ statt $\dim_K(V)$, wenn der Körper klar ist. Die Dimension ist eine entscheidende Charakterisierung eines Vektorraumes.

Satz 2.9 *Ist $W \subset V$ Untervektorraum eines endlich erzeugten Vektorraumes V, so ist auch W endlich erzeugt, und es gilt*

i) $\dim(W) \leq \dim(V)$,

ii) *aus* $\dim(W) = \dim(V)$ *folgt* $W = V$.

Beweis. Sei $n := \dim_K(V)$. Gäbe es in W mehr als n linear unabhängige Vektoren $w_1, \ldots, w_r \in W$, so wären diese auch als Elemente von V linear unabhängig, im Widerspruch zur Eindeutigkeit der Dimension. Also ist $r \leq n$.

W ist durch höchstens n linear unabhängige Vektoren $(w_1, \ldots, w_n)$ erzeugt, denn wäre $(w_1, \ldots, w_n)$ kein Erzeugendensystem, dann gibt es ein $w \in W$ mit $w \notin \operatorname{span}_K(w_1, \ldots, w_n)$. Dann wäre $(w_1, \ldots, w_n, w)$ linear unabhängig, im Widerspruch zur Dimension von V.

Sei nun $\dim(W) = \dim(V) = n$ und $(w_1, \dots, w_n)$ eine Basis von W. Wäre $V \neq W$, so gibt es ein $v \in V \setminus W$, das keine Linearkombination von $(w_1, \dots, w_n)$ ist. Damit wäre $(w_1, \dots, w_n, v)$ linear unabhängig, was durch die Dimension ausgeschlossen ist. □

Dieser Satz ist äußerst wichtig. Wenn wir schon wissen, daß ein Vektorraum V die Dimension n hat und n linear unabhängige Vektoren $v_1, \dots, v_n \in V$ gegeben sind, dann ist $W = \mathrm{span}(v_1, \dots, v_n)$ ein Untervektorraum, in dem $(v_1, \dots, v_n)$ eine Basis ist. Aus der Gleichheit der Dimensionen folgt $W = V$, also ist $(v_1, \dots, v_n)$ ein Erzeugendensystem von V. Es ist wesentlich einfacher, die lineare Unabhängigkeit zu überprüfen als die Eigenschaft des Erzeugendensystems. Daraus wird die Bedeutung der Dimension ersichtlich.

Bei der Untersuchung der linearen Unabhängigkeit ist das Austauschlemma sehr hilfreich.

Beispiele 2.4 Sei $V = \mathbb{R}^4$ und $v_1 = (2,5,3,7)$, $v_2 = (1,1,-3,6)$, $v_3 = (1,4,0,1)$, $v_4 = (2,-1,5,1)$. Ist die Familie $\mathcal{B}_1 = (v_1, v_2, v_3, v_4)$ linear unabhängig? Wenn ja, dann ist $\mathcal{B}_1$ auch eine Basis (insbesondere ein Erzeugendensystem), da $\mathbb{R}^4$ die Dimension 4 hat.

Wir betrachten den Vektor $w_2 = v_1 - 2v_2 = (0,3,9,-5)$. Wenn $\mathcal{B}_1$ eine Basis ist, so können wir nach dem Austauschlemma v_1 oder v_2 durch w austauschen, ohne die Basiseigenschaft zu verlieren. Also ist $\mathcal{B}_2 = (v_1, w_2, v_3, v_4)$ eine Basis, wenn $\mathcal{B}_1$ eine ist. Sei umgekehrt $\mathcal{B}_2$ eine Basis, dann ist nach Austausch von w_2 durch $v_2 = \frac{1}{2} \cdot v_1 - \frac{1}{2} \cdot w_2$ auch $\mathcal{B}_1$ eine Basis.

Im nächsten Schritt betrachten wir $w_3 = v_1 - 2v_3 = (0,-3,3,5)$. Dann ist $\mathcal{B}_3 = (v_1, w_2, w_3, v_4)$ genau dann eine Basis, wenn $\mathcal{B}_1$ eine ist.

Der dritte Schritt liefert $w_4 = v_1 - v_4 = (0,6,-2,6)$. Dann ist $\mathcal{B}_4 = (v_1, w_2, w_3, w_4)$ genau dann eine Basis, wenn $\mathcal{B}_1$ eine ist.

Die Frage, ob $\mathcal{B}_1$ eine Basis ist, ist nun darauf zurückgeführt, ob (w_2, w_3, w_4) linear unabhängig sind. Denn jede Linearkombination von (w_2, w_3, w_4) hat an der ersten Stelle eine 0, d.h. $\mathrm{span}_{\mathbb{R}}(w_2, w_3, w_4) \subset \{0\} \times \mathbb{R}^3$. Also folgt aus $\lambda_1 v_1 + \sum_{\imath=2}^{4} \lambda_i \cdot w_i = 0$, daß $\lambda_1 = 0$ und $\sum_{\imath=2}^{4} \lambda_i \cdot w_i = 0$ sein muß. Wir haben also das Problem, ob $\mathcal{B}_1$ eine Basis von $\mathbb{R}^4$ ist, auf das niederdimensionale Problem zurückgeführt, ob (w_2, w_3, w_4) eine Basis von $\{0\} \times \mathbb{R}^3$ ist.

Durch wiederholte Anwendung dieses Verfahrens werden w_3, w_4 durch geeignete Linearkombination[1] u_3, u_4 von w_2, w_3, w_4 ausgetauscht, so daß die ersten beiden Stellen von u_3, u_4 Null sind, also $\mathrm{span}_{\mathbb{R}}(u_3, u_4) \subset \{0\} \times \{0\} \times \mathbb{R}^2$. Damit ist (w_2, w_3, w_4) eine Basis von $\{0\} \times \mathbb{R}^3$, wenn (u_3, u_4) eine Basis von $\{0\}^2 \times \mathbb{R}^2$ ist.

Im letzten Schritt tauschen wir u_4 durch eine Linearkombination[2] x_4 von u_3, u_4 aus mit $\mathrm{span}_{\mathbb{R}}(x_4) \subset \{0\}^3 \times \mathbb{R}^1$. Nun ist (x_4) genau dann eine Basis von $\{0\}^3 \times \mathbb{R}^1$,

[1] z.B. $u_3 = w_2 + w_3 = (0,0,12,0)$ und $u_4 = 2w_2 - w_4 = (0,0,20,-16)$

[2] z.B. $x_4 = 5u_3 - 3u_4 = (0,0,0,48)$

wenn $x_4 \neq 0$ ist, was einfach zu entscheiden ist. Aus der Kette von Beziehungen folgt, daß $\mathcal{B}_1$ genau dann eine Basis von $\mathbb{R}^4$ ist, wenn $x_4 \neq 0$ ist (das ist hier der Fall).

Dieses Verfahren, welches schrittweise das Problem auf einfachere zurückführt, ist eine Variante des Gaußschen Algorithmus, den wir später behandeln werden.

Unter Umständen zwingt uns das Verfahren, die Reihenfolge der Vektoren in der Familie zu ändern. Sei z.B. $v_1 = (0, 1, 2)$, $v_2 = (1, 1, 1)$, $v_3 = (1, 2, 2)$. Dann können wir v_2 nicht durch eine Linearkombination w_2 aus v_1, v_2 ersetzen, deren erste Stelle 0 ist. Stattdessen ist $v_1 \mapsto v_2$ zu tauschen und dann v_3 durch eine Linearkombination w_3 aus v_2, v_3 auszutauschen.

Ein weiteres nützliches Hilfsmittel ist:

Satz 2.10 (Basisergänzungssatz) *In einem endlich erzeugten Vektorraum V sei eine Familie $(w_1, \dots, w_r)$ von linear unabhängigen Vektoren gegeben. Dann läßt sich diese Familie zu einer Basis $(w_1, \dots, w_r, v_{r+1}, \dots, v_n)$ von V ergänzen.*

Beweis. Man nehme irgendeine Basis von V und wende den Austauschsatz an. □

Bemerkungen zu nicht endlich erzeugten Vektorräumen

In nicht endlich erzeugten (also unendlich-dimensionalen) Vektorräumen gibt es einige Besonderheiten.

Satz 2.11 *Ist V nicht endlich erzeugt, so gibt es eine unendliche linear unabhängige Familie.*

Beweis. Ist $(v_1, \dots, v_n)$ linear unabhängig, dann gibt es einen Vektor $v \in V$, so daß auch $(v_1, \dots, v_n, v)$ linear unabhängig ist. Sonst wäre V endlich erzeugt, im Gegensatz zur Voraussetzung. □

Unter Benutzung des *Auswahlaxioms* gilt:

Satz 2.12 *Jeder Vektorraum besitzt eine Basis.*

Jedoch läßt sich in nicht endlich erzeugten Vektorräumen eine Basis im allgemeinen nicht angeben. Im Vektorraum der reellen Zahlenfolgen $V = \{(a_1, a_2, \dots) \,:\, a_i \in \mathbb{R}\}$ sind die Vektoren

$$\{(1, 0, 0, 0, \dots), (0, 1, 0, 0, \dots), (0, 0, 1, 0, \dots), \dots\}$$

zwar linear unabhängig, aber sie bilden keine Basis (auch kein Erzeugendensystem), denn z.B. wird die Folge $(1, 1, 1, \dots)$ nicht davon *als endliche Linearkombination* erzeugt.

In unendlich-dimensionalen unitären Vektorräumen ist es möglich (und üblich), sogenannte Orthonormalbasen zu betrachten, so daß jeder Vektor sich als eindeutige unendliche Linearkombination darstellen läßt (z.B. bilden obige Vektoren eine Orthonormalbasis im Vektorraum der reellen Zahlenfolgen). Unter Verwendung des Auswahlaxioms existiert zwar eine Basis, so daß jeder Vektor eine endliche Linearkombination ist, diese Basis ist aber viel größer (überabzählbar) als die sehr natürliche Orthonormalbasis.

Zwar kann Satz 2.4 über die äquivalenten Eigenschaften einer Basis auch für unendlich-dimensionale Vektorräume bewiesen werden (ohne Auswahlaxiom), da eine solche Basis im allgemeinen aber nicht angegeben werden kann, haben echte Basen in unendlich-dimensionalen Vektorräumen kaum eine praktische Bedeutung. Die Verwendung von Orthonormalbasen ist dann sinnvoller, allerdings erfordert das dann Hilfsmittel aus der Analysis. Diese Strukturen werden später behandelt.

2.3 Direkte Summe von Vektorräumen

Sei $(v_1, \ldots, v_n)$ eine Basis von V, dann erzeugt jeder Basisvektor v_i einen eindimensionalen Untervektorraum $W_i = Kv_i := \mathrm{span}_K(v_i)$. Da jeder Vektor $v \in V$ als Summe $v = \sum_{i=1}^n w_i$ mit $w_i \in Kv_i$ dargestellt werden kann, ist es sinnvoll, V selbst als Summe der Untervektorräume aufzufassen. Diese Summe läßt sich dann verallgemeinern auf höherdimensionale Untervektorräume:

Definition 2.6 Sei W ein Vektorraum und $W_1, \ldots, W_k$ Untervektorräume von W. Ein Vektorraum V heißt *direkte Summe* der Untervektorräume $W_1, \ldots, W_k$, bezeichnet mit

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k \;,$$

wenn gilt:

(DS1) $V = W_1 + \cdots + W_k := \{v \in W \;:\; \text{es gibt } w_i \in W_i \text{ mit } v = w_1 + \cdots + w_k\}$

(DS2) Von Null verschiedene Vektoren $w_1 \in W_1, \ldots, w_k \in W_k$ sind linear unabhängig in W.

Beispiele 2.5 Sei $V = \mathbb{R}^3$ und $v_1, v_2, v_3 \in V$ linear unabhängig, dann ist z.B. $V = \mathbb{R}v_1 \oplus \mathbb{R}v_2 \oplus \mathbb{R}v_3$ oder $V = \mathrm{span}_{\mathbb{R}}(v_1, v_3) \oplus \mathbb{R}v_2$.

Satz 2.13 *Sind $W_1, \ldots, W_k$ endlich-dimensionale Untervektorräume, so gilt $\dim(W_1 \oplus \cdots \oplus W_k) = \dim(W_1) + \cdots + \dim(W_k)$.*

Beweis. Es existieren Basen $(w_1^{(i)}, \ldots, w_{r_i}^{(i)})$ der r_i-dimensionalen Untervektorräume W_i. Wir zeigen, daß

$$\mathcal{B} := (w_1^{(1)}, \ldots, w_{r_1}^{(1)}, \ldots, w_1^{(k)}, \ldots, w_{r_k}^{(k)})$$

eine Basis von $V := W_1 \oplus \cdots \oplus W_k$ ist. Klar ist, daß $\mathcal{B}$ ein Erzeugendensystem von V ist. Es verbleibt der Beweis der linearen Unabhängigkeit. Sei also

$$0 = \sum_{i_1=1}^{r_1} \lambda_{i_1} \cdot w_{i_1}^{(1)} + \cdots + \sum_{i_k=1}^{r_k} \lambda_{i_k} \cdot w_{i_k}^{(k)} .$$

Wir setzen $v_j := \sum_{i_j=1}^{r_j} \lambda_{i_j} \cdot w_{i_j}^{(j)} \in W_j$. Dann ist $v_1 + \cdots + v_k = 0$. Nach (DS2) sind von Null verschiedene $v_j \in W_j$ linear unabhängig, also folgen aus $v_1 + \cdots + v_k = 0$ die Gleichungen $v_j = 0$ für alle j und dann, da $(w_{i_j}^{(j)})$ Basis von W_j ist, $\lambda_{i_j} = 0$ für alle i, j. □

Die Bedingung (DS2) ist entscheidend dafür, daß die Dimension der direkten Summe von Vektorräumen gleich der Summe der Dimensionen der Untervektorräume ist. Arbeitet man einfach nur mit der Summe (DS1) von Untervektorräumen, so gilt:

Satz 2.14 *Für endlich-dimensionale Untervektorräume $W_1, W_2 \subset V$ gilt $\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$.*

Beweis. Wir erinnern daran, daß $W_1 \cap W_2$ ein endlich-dimensionaler Untervektorraum ist. Man nehme eine Basis $(v_1, \ldots, v_m)$ und ergänze sie zu Basen $(v_1, \ldots, v_m, w_1, \ldots, w_r)$ von W_1 und $(v_1, \ldots, v_m, w'_1, \ldots, w'_s)$ von W_2. Dann ist $\dim(W_1 + W_2) = m + r + s$ sowie $\dim(W_1) = m + r$ und $\dim(W_2) = m + s$. □

Satz 2.15 *Sei V eine endlich-dimensionaler Vektorraum und $W \subset V$ ein Untervektorraum. Dann gibt es einen (nicht eindeutig bestimmten) Untervektorraum $W' \subset V$ mit $V = W \oplus W'$. (W' heißt* direkter Summand *von V zu W.)*

Beweis. Ist $\dim(W) = \dim(V)$, so ist $V = W \oplus \{0\}$. Sei nun $\dim(W) < \dim(V)$, $(v_1, \ldots, v_n)$ eine Basis von V und $(w_1, \ldots, w_r)$ eine Basis von W. Nach dem Basisaustauschsatz gibt es nach möglicher Umnumerierung der Indizes eine Basis $(w_1, \ldots, w_r, v_{r+1}, \ldots, v_n)$ von V. Also ist $W' := \mathrm{span}(v_{r+1}, \ldots, v_n)$ ein direkter Summand. □

2.4 Lineare Abbildungen

Definition 2.7 Seien V, W Vektorräume über K. Eine Abbildung $F : V \to W$ heißt *linear* (genauer: K-linear) oder *Homomorphismus* von Vektorräumen (über K), wenn

(L) $F(\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2) = \lambda_1 \cdot F(v_1) + \lambda_2 \cdot F(v_2)$ für alle $v_1, v_2 \in V$ und $\lambda_1, \lambda_2 \in K$.

Eine lineare Abbildung $F : V \to W$ heißt

- *Isomorphismus*, wenn F bijektiv ist,
- *Endomorphismus*, wenn $W = V$ ist,

- *Automorphismus*, wenn F bijektiv und $W = V$ ist.

Beispiele 2.6

i) *Skalentransformationen (Dilatationen).* Sei V ein reeller Vektorraum und $c > 0$, dann definiert $F_c : v \mapsto c \cdot v$ einen Automorphismus von V (eine lineare Abbildung $F_c : V \to V$, die bijektiv ist mit $F_c^{-1} = F_{\frac{1}{c}} : v \mapsto \frac{1}{c} \cdot v$).

ii) *Drehungen in der Ebene.* Sei $V = \mathbb{R}^2$ und F_α die Drehung um den Nullpunkt mit dem Winkel α,

$$F_\alpha : (x, y) \mapsto F_\alpha(x, y) := (x \cos\alpha - y \sin\alpha, x \sin\alpha + y \cos\alpha) \ .$$

F ist ebenfalls ein Automorphismus mit $F_\alpha^{-1} = F_{-\alpha}$.

iii) *Matrizen.* Sei $V = W = \mathbb{R}^2$ und

$$F_a : (x_1, x_2) \mapsto (a_{11}x_1 + a_{12}x_2, a_{21}x_1 + a_{22}x_2) \ .$$

Dann ist $F_a : \mathbb{R}^2 \to \mathbb{R}^2$ eine lineare Abbildung, welche durch die Matrix $a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in M(2 \times 2, \mathbb{R})$ parametrisiert wird. Die beiden Beispiele i) und ii) sind Spezialfälle mit $a = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$ (für $V = \mathbb{R}^2$) und $a = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$. Für allgemeines a ist F ein Endomorphismus. Es wird wichtig sein zu untersuchen, wann F bijektiv ist, also ein Automorphismus ist.

iv) *Projektionen der Ebene auf eine Gerade durch den Nullpunkt.* Sei $V = \mathbb{R}^2$ und (v_1, v_2) eine Basis, z.B. $v_1 = (1, 0)$ und $v_2 = (0, 1)$. Dann definiert die eindeutige Zerlegung von v nach der Basis $v = \lambda_1 \cdot v_1 + \lambda_2 \cdot v_2$ zwei Projektionen

$$\begin{aligned} F_1 &: \lambda_1 \cdot v_1 + \lambda_2 \cdot v_2 \mapsto \lambda_1 \cdot v_1 \ , \\ F_2 &: \lambda_1 \cdot v_1 + \lambda_2 \cdot v_2 \mapsto \lambda_2 \cdot v_2 \ . \end{aligned}$$

Diese Projektionen F_1, F_2 sind lineare Abbildungen.

v) In Analogie zu Beispiel iii) definieren wir $F_a : \mathbb{R}^2 \to \mathbb{R}^1$ durch $F_a : (x_1, x_2) \mapsto a_{11}x_1 + a_{12}x_2$. Dann ist F_a eine lineare Abbildung, die durch die Matrix $a = \begin{pmatrix} a_{11} & a_{12} \end{pmatrix} \in M(1 \times 2, \mathbb{R})$ parametrisiert wird. Beispiel iv) für $v_1 = (1, 0)$ und $v_2 = (0, 1)$ ist ein Spezialfall mit $F_1 \Leftrightarrow a = (1 \ 0)$ und $F_2 \Leftrightarrow a = (0 \ 1)$.

vi) *Projektionen auf einen Untervektorraum.* In Verallgemeinerung von Beispiel iv) sei V ein n-dimensionaler Vektorraum, $W \subset V$ ein Untervektorraum und $V = W \oplus W'$. Dann ist die Projektion $F_W : V \to W$ eine lineare Abbildung, die wie folgt erhalten wird. Man nehme eine Basis $(w_1, \ldots, w_r)$ von W und eine Basis $(w'_{r+1}, \ldots, w'_n)$ von W' (also $n-r$ Vektoren von V, so daß $(w_1, \ldots, w_r, w'_{r+1}, \ldots, w'_n)$ eine Basis von V ist) und zerlege $v \in V$ nach dieser Basis, $v = \sum_{i=1}^r \lambda_i \cdot v_i + \sum_{j=r+1}^n \mu_j \cdot w'_j$. Dann ist $F_W : v \mapsto \sum_{i=1}^r \lambda_i \cdot v_i$. Diese Abbildung F_W ist unabhängig vom direkten Summanden W' (welcher nicht eindeutig war) und unabhängig von der Basis von W.

vii) *Differentiation.* Für ein offenes Intervall $I = (a, b) \subset \mathbb{R}$ bezeichnen $V = C^1(I)$ und $W = C(I)$ die (unendlich-dimensionalen) Vektorräume der einmal stetig differenzierbaren bzw. der stetigen reellwertigen Funktionen auf I, mit $(\lambda_1 \cdot f_1 + \lambda_2 \cdot f_2)(x) := \lambda_1 \cdot f_1(x) + \lambda_2 \cdot f_2(x)$ für $x \in (a, b)$. Dann ist das Differential $D : V \to W$ mit $D : f \mapsto f'$ eine lineare Abbildung. Ist $V = C^\infty(I)$ der Vektorraum der beliebig oft differenzierbaren Funktionen über I, dann ist das Differential $D : V \to V$ ein Endomorphismus.

viii) *Integration.* Für ein abgeschlossenes Intervall $I = [a, b] \subset \mathbb{R}$ bezeichne $V = C(I)$ den (unendlich-dimensionalen) Vektorraum der stetigen reellwertigen Funktionen über I. Dann ist das Integral $\int : V \to \mathbb{R}$ mit $\int : f \mapsto \int_a^b f(x)\, dx$ eine lineare Abbildung.

Diese sehr unterschiedlichen Beispiele zeigen, daß lineare Abbildungen häufig auftreten. Kenntnisse der allgemeinen Eigenschaften von linearen Abbildungen und von Methoden zu ihrer Untersuchung sind deshalb sehr wichtig.

Satz 2.16 *Sei $F : V \to W$ eine lineare Abbildung. Dann gilt*

i) $F(0) = 0$, $F(v - w) = F(v) - F(w)$, $F(\lambda_1 \cdot v_1 + \cdots + \lambda_n \cdot v_n) = \lambda_1 \cdot F(v_1) + \cdots + \lambda_n \cdot F(v_n)$

ii) *Ist die Familie $(v_i)_{i \in I}$ linear abhängig in V, so ist die Familie $(F(v_i))_{i \in I}$ linear abhängig in W*

iii) *Sind $V' \subset V$ und $W' \subset W$ Untervektorräume, so ist $F(V') \subset W$ und $F^{-1}(W') \subset V$ Untervektorräume.*

iv) $\dim(F(V)) \leq \dim(V)$

v) *Ist F ein Isomorphismus, so ist auch das Urbild $F^{-1} : W \to V$ eine lineare Abbildung (und damit ein Isomorphismus).*

Beweis. i) ist klar.

ii) Sei $0 = \sum_{i \in I} \lambda_i \cdot v_i$, wobei mindestens einer und insgesamt endlich viele Skalare λ_i ungleich 0 sind. Dann ist auch $0 = F(0) = \sum_{i \in I} \lambda_i \cdot F(v_i)$.

iii) $0 \in V'$ und $0 = F(0) \in F(V')$, also $F(V') \neq \emptyset$. Sind $w_1, w_2 \in F(V')$, dann gibt es v_1, v_2 mit $w_1 = F(v_1)$ und $w_2 = F(v_2)$. Also $\lambda_1 \cdot w_1 + \lambda_2 \cdot w_2 = \lambda_1 \cdot F(v_1) + \lambda_2 \cdot F(v_2) = F(\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2) \in F(V')$, und $F(V')$ ist Untervektorraum.

Sind $v_1, v_2 \in F^{-1}(W')$, so ist $F(v_1), F(v_2) \in W'$ und auch $\lambda_1 \cdot F(v_1) + \lambda_2 \cdot F(v_2) \in W'$, da W' Untervektorraum. Also ist $F(\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2) = \lambda_1 \cdot F(v_1) + \lambda_2 \cdot F(v_2) \in W'$ und damit $\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2 \in F^{-1}(W')$, und $F^{-1}(W')$ ist Untervektorraum.

iv) Sind $F(v_1), \ldots, F(v_n)$ linear unabhängig, so sind auch $v_1, \ldots, v_n$ linear unabhängig, denn das Gegenteil führt wegen ii) zu einem Widerspruch.

v) Seien $w_1, w_2 \in W$ und $v_1, v_2 \in V$ mit $F(v_1) = w_1$ und $F(v_2) = w_2$. Dann ist $F(\lambda_1 \cdot v_1 + \lambda_2 \cdot v_2) = \lambda_1 \cdot w_1 + \lambda_2 \cdot w_2$. Anwenden von F^{-1} liefert $F^{-1}(\lambda_1 \cdot w_1 + \lambda_2 \cdot w_2) = \lambda_1 \cdot F^{-1}(w_1) + \lambda_2 \cdot F^{-1}(w_2)$. □

Seien V, W Vektorräume über K, dann bezeichnen wir mit

$$\mathrm{Hom}_K(V, W) := \{F : V \to W \ : \ F \text{ ist } K\text{-linear}\}$$

die Menge aller K-linearen Abbildungen (aller Homomorphismen) von V nach W. Die Menge $\mathrm{Hom}_K(V, W)$ ist ein Vektorraum mit Verknüpfungen

$$(\lambda_1 \cdot F_1 + \lambda_2 \cdot F_2)(v) := \lambda_1 \cdot F_1(v) + \lambda_2 \cdot F_2(v) \ .$$

Insbesondere ist der Nullvektor die lineare Abbildung $0 : v \mapsto 0$ für alle $v \in V$, und die negative Abbildung ist $-F : v \mapsto -F(v)$. Ist der Körper klar, dann schreiben wir $\mathrm{Hom}(V, W)$ statt $\mathrm{Hom}_K(V, W)$

Ist $V = W$, so schreiben wir $\mathrm{End}(V) := \mathrm{Hom}(V, V)$, und $\mathrm{End}(V)$ ist wieder ein Vektorraum. Entsprechend bezeichnet

$$\mathrm{Aut}_K(V) := \{F : V \to V \ : \ F \text{ ist } K\text{-linear und bijektiv}\}$$

die Menge der bijektiven linearen Abbildungen von V nach V. Jedoch ist $\mathrm{Aut}(V)$ kein Vektorraum, denn für $F \in \mathrm{Aut}(V)$ ist $0 \cdot F = 0 \notin \mathrm{Aut}(V)$.

Lineare Abbildungen lassen sich ganz analog wie allgemeine Abbildungen hintereinander ausführen.

Satz 2.17 *Sind U, V, W Vektorräume und $G : U \to V$ sowie $F : V \to W$ lineare Abbildung, dann ist auch die Komposition $F \circ G : U \to W$ eine lineare Abbildung.*

Beweis.

$$\begin{aligned}
(F \circ G)(\lambda_1 \cdot u_1 + \lambda_2 \cdot u_2) &:= F(G(\lambda_1 \cdot u_1 + \lambda_2 \cdot u_2)) \\
&= F(\lambda_1 \cdot G(u_1) + \lambda_2 \cdot G(u_2)) \\
&= \lambda_1 \cdot F(G(u_1)) + \lambda_2 \cdot F(G(u_2)) \\
&= \lambda_1 \cdot (F \circ G)(u_1) + \lambda_2 \cdot (F \circ G)(u_2) \ .
\end{aligned}$$

□

Ohne Beweis erwähnen wir:

Satz 2.18 *Ist V ein Vektorraum, so ist* $\mathrm{End}(V)$ *ein Ring mit der Komposition als Multiplikation, und* $\mathrm{Aut}(V)$ *ist eine Gruppe mit der Komposition als Verknüpfung.*

Definition 2.8 Ist $F : V \to W$ eine lineare Abbildung, so heißt

- $\mathrm{im}(F) := F(V) \subset W$ das *Bild* von F,
- $\ker(F) := F^{-1}(0) \subset V$ der *Kern* von F.

Satz 2.19 *Sei $F : V \to W$ linear. Dann gilt:*

i) $\mathrm{im}(F) \subset W$ *und* $\ker(F) \subset V$ *sind Untervektorräume.*

ii) F *surjektiv* $\Leftrightarrow$ $\mathrm{im}(F) = W$

iii) F *injektiv* $\Leftrightarrow$ $\ker(F) = \{0\}$

iv) *Ist F injektiv und sind $v_1, \ldots, v_n$ linear unabhängig, so sind auch $F(v_1), \ldots, F(v_n)$ linear unabhängig.*

Beweis. i) ist in Satz 2.16.iii) bewiesen und ii) ist die Definition der Surjektivität.

iii) $0 \in \ker(F)$. Wenn F bijektiv, so folgt aus $F(v) = 0 = F(0)$, daß $v = 0$. Gibt es umgekehrt $v_1 \neq v_2 \in V$ mit $F(v_1) = F(v_2)$, so wäre $F(v_1 - v_2) = 0$, also $0 \neq v_1 - v_2 \in \ker(F)$.

iv) Sei $\lambda_1 \cdot F(v_1) + \cdots + \lambda_n \cdot F(v_n) = 0$, dann gilt $0 = F(\lambda_1 \cdot v_1 + \cdots + \lambda_n \cdot v_n)$ nach Linearität und $\lambda_1 \cdot v_1 + \cdots + \lambda_n \cdot v_n = 0$ nach iii). Also ist $\lambda_i = 0$. □

Definition 2.9 Ist $F : V \to W$ linear, so heißt die Zahl $\mathrm{rang}(F) := \dim(\mathrm{im}(F))$ der *Rang* der Abbildung F.

Der Rang ist eine wichtige Charakterisierung einer linearen Abbildung.

Satz 2.20 *Ist V ein endlich-dimensionaler Vektorraum und $F : V \to W$ eine lineare Abbildung, so gilt*

$$\dim(V) = \dim(\mathrm{im}(F)) + \dim(\ker(F)) \; .$$

Beweis. Wir wissen $\dim(\mathrm{im}(F)) \leq \dim(V)$ nach Satz 2.16.iv) und $\dim(\ker(F)) \leq \dim(V)$ nach Satz 2.19. Seien also $(v_1, \ldots, v_k)$ eine Basis von $\ker(F)$ und $(w_1, \ldots, w_r)$ eine Basis von $\mathrm{im}(F)$. Wir wählen $u_1, \ldots, u_r \in V$ mit $F(u_i) = w_i$ für alle $1 \leq i \leq r$. Für $v \in V$ gibt es eine eindeutige Darstellung $F(v) = \lambda_1 \cdot w_1 + \cdots + \lambda_r \cdot w_r$. Für die so bestimmten λ_i konstruieren wir $v' := \lambda_1 \cdot u_1 + \cdots + \lambda_r \cdot u_r$. Dann gilt $F(v - v') = 0$, also $v - v' \in \ker(F)$. Es existieren also eindeutig bestimmte $\mu_1, \ldots, \mu_k$ mit $v - v' = \mu_1 \cdot v_1 + \cdots + \mu_k \cdot v_k$. Also gilt

$$v = \lambda_1 \cdot u_1 + \cdots + \lambda_r \cdot u_r + \mu_1 \cdot v_1 + \cdots + \mu_k \cdot v_k \; .$$

Damit ist V durch die Familie $\mathcal{B} = (u_1, \ldots, u_r, v_1, \ldots, v_k,)$ erzeugt. Zum Beweis der linearen Unabhängigkeit sei $\lambda_1 \cdot u_1 + \cdots + \lambda_r \cdot u_r + \mu_1 \cdot v_1 + \cdots + \mu_k \cdot v_k = 0$. Anwenden von F ergibt $F(\lambda_1 \cdot u_1 + \cdots + \lambda_r \cdot u_r) = \lambda_1 \cdot w_1 + \cdots + \lambda_r \cdot w_r = 0$, und damit $\lambda_i = 0$. Das liefert $\mu_1 \cdot v_1 + \cdots + \mu_k \cdot v_k = 0$ und damit $\mu_j = 0$. Damit ist $\mathcal{B}$ eine Basis von V aus $\dim(\mathrm{im}(F)) + \dim(\ker(F))$ Vektoren. □

Satz 2.21 *Zwischen endlich-dimensionalen Vektorräumen V, W gibt es genau dann einen Isomorphismus, wenn* $\dim(V) = \dim(W)$.

Beweis. F ist Isomorphismus, wenn linear, injektiv und surjektiv. F injektiv heißt $\dim(\ker(F)) = 0$ und F surjektiv heißt $\mathrm{im}(F) = W$. Aus dem vorigen Satz folgt dann die Behauptung. □

Satz 2.22 *Ist* $\dim(V) = \dim(W) < \infty$ *und ist F linear, dann sind für F Injektivität, Surjektivität und Bijektivität äquivalent.*

Beweis. injektiv$\Rightarrow$ surjektiv: F injektiv, dann $\ker(F) = \{0\}$ und $\dim(V) = \dim(\mathrm{im}(F))$. Also $\dim(\mathrm{im}(F)) = \dim(W)$ und dann $\mathrm{im}(F) = W$.

surjektiv$\Rightarrow$injektiv: F surjektiv, dann $\dim(\mathrm{im}(F)) = \dim(W) = \dim(V)$. Also $\dim(\ker(F)) = 0$ und $\ker(F) = \{0\}$, somit F injektiv. □

Satz 2.23 *Sei $F : V \to W$ linear und $\mathcal{B} = (u_1, \ldots, u_r, v_1, \ldots, v_k)$ eine Basis von V mit* $\ker(F) = \mathrm{span}(v_1, \ldots, v_k)$. *Mit* $U := \mathrm{span}(u_1, \ldots, u_r)$ *gilt:*

i) $V = U \oplus \ker(F)$

ii) *Die Einschränkung* $F\big|_U : U \to \mathrm{im}(F)$ *ist ein Isomorphismus.*

iii) *Sei $P_U : V = U \oplus \ker(F) \to U$ die Projektion auf den ersten Summanden, definiert für $v = u + v' \in V$ mit $u \in U$ und $v' \in \ker(F)$ durch $P_U(v) := u$, so gilt $F = F\big|_U \circ P_U$.*

Beweis. i) ist klar, denn U und $\ker(F)$ sind linear unabhängig.

ii) $\ker(F\big|_U) = \ker(F) \cap U = \{0\}$, also ist $F\big|_U : U \to \mathrm{im}(F)$ injektiv, außerdem surjektiv und damit bijektiv.

iii) Für $v = u + v'$ ist $F(v) = F(u) + F(v') = F(u) = F\big|_U(u) = F\big|_U(P(v))$. □

Zu beachten ist, daß U in der direkten Summe $V = U \oplus \ker(F)$ nicht eindeutig definiert ist. Die $F\big|_U$ sind also, je nach Wahl von U, verschiedene Isomorphismen.

2.5 Matrizen

Die Forderung der Linearität ist eine sehr einschränkende Bedingung. Es geht nun darum zu untersuchen, durch welche Eigenschaften eine lineare Abbildung eindeutig fixiert wird.

Satz 2.24 *Gegeben seien endlich-dimensionale Vektorräume V und W und Vektoren $v_1, \ldots, v_r \in V$ und $w_1, \ldots, w_r \in W$. Dann gilt:*

i) *Sind die $v_1, \ldots, v_r$ linear unabhängig, so gibt es mindestens eine lineare Abbildung $F : V \to W$ mit $F(v_i) = w_i$ für alle $1 \leq i \leq r$.*

ii) *Ist $(v_1, \ldots, v_r)$ eine Basis, so gibt es genau eine lineare Abbildung $F : V \to W$ mit $F(v_i) = w_i$ für alle $1 \leq i \leq r$. Für diese Abbildung F gilt*

- $\mathrm{im}(F) = \mathrm{span}(w_1, \dots, w_r)$
- *F injektiv* $\Leftrightarrow$ *$(w_1, \dots, w_r)$ ist linear unabhängig.*

Beweis. ii) Jedes $v \in V$ hat eine eindeutige Darstellung $v = \lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r$. Wegen $F(v_i) = w_i$ folgt aus der angenommenen Linearität $F(v) = \lambda_1 \cdot w_1 + \cdots + \lambda_r \cdot w_r$. Damit gibt es höchstens eine lineare Abbildung mit diesen Eigenschaften. Zu zeigen ist aber noch, daß F wirklich linear ist. Sei dazu $v' = \mu_1 \cdot v_1 + \cdots + \mu_r \cdot v_r$ ein weiterer Vektor, dann ist

$$\begin{aligned} \lambda \cdot v + \mu \cdot v' &= \sum_{i=1}^{r} (\lambda\lambda_i + \mu\mu_i) \cdot v_i \qquad \text{und} \\ F(\lambda \cdot v + \mu \cdot v') &= \sum_{i=1}^{r} (\lambda\lambda_i + \mu\mu_i) \cdot w_i = \lambda\Big(\sum_{i=1}^{r} \lambda_i \cdot w_i\Big) + \mu\Big(\sum_{i=1}^{r} \mu_i \cdot w_i\Big) \\ &= \lambda \cdot F(v) + \mu \cdot F(v') \ . \end{aligned}$$

Klar ist, daß $\mathrm{im}(F) \subset \mathrm{span}(w_1, \dots, w_r)$. Andererseits ist $w = \lambda_1 \cdot w_1 + \cdots + \lambda_r w_r = F(\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r)$ und damit $\mathrm{span}(w_1, \dots, w_r) \subset \mathrm{im}(F)$.

Sei $(w_1, \dots, w_r)$ linear abhängig, so gibt es mindestens ein $\lambda_i \neq 0$, so daß $0 = \lambda_1 w_1 + \dots \lambda_r w_r = F(\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r)$. Da $(v_1, \dots, v_r)$ eine Basis ist, ist $\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r \neq 0$, und damit ist F nicht injektiv. Das bedeutet F injektiv $\Rightarrow$ $(w_1, \dots, w_r)$ linear unabhängig. Ist umgekehrt $(w_1, \dots, w_r)$ linear unabhängig, dann betrachten wir $F(v) = 0$ mit $v = \lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r$, also $F(\lambda_1 \cdot v_1 + \cdots + \lambda_r \cdot v_r) = \lambda_1 \cdot w_1 + \cdots + \lambda_r \cdot w_r$, was zu $\lambda_i = 0$ und damit $v = 0$ führt. Also ist $\ker(F) = \{0\}$ und F ist injektiv.

i) Ist $(v_1, \dots, v_r)$ linear unabhängig, aber keine Basis, dann können wir eine Basis $(v_1, \dots, v_r, v_{r+1}, \dots, v_n)$ von V finden. Für diese geben wir beliebige Vektoren $w_{r+1}, \dots, w_n$ vor und finden für jede Wahl von $w_{r+1}, \dots, w_n$ eine lineare Abbildung F mit $F(v_i) = w_i$. □

Satz 2.25 *Ist $\mathcal{B} = (w_1, \dots, w_n)$ eine Basis eines Vektorraumes W, so gibt es einen genau einen Isomorphismus $\Phi_{\mathcal{B}} : K^n \to W$ mit $\Phi_{\mathcal{B}}(e_i) = w_i$ für alle $1 \leq i \leq n$, wobei $(e_i)_{i=1,\dots,n}$ die Standardbasis ist.*

Beweis. Folgt aus Satz 2.24.ii) □

Ist also die Basis $\mathcal{B}$ eines n-dimensionalen Vektorraums V gegeben, dann kann man V mit K^n identifizieren über den Isomorphismus $\Phi_{\mathcal{B}}$. Mir werden nun sehen, wie das auch die lineare Abbildung festlegt.

Satz 2.26 *Zu jeder linearen Abbildung $F : K^n \to K^m$ gibt es genau eine Matrix $A = (a_{ij}) \in M(m \times n, K)$ mit $F(e_j) = \sum_{i=1}^{m} a_{ij} \cdot e_i$ für alle $1 \leq j \leq n$.*

Beweis. Wir setzen $F(e_j) = w_j \in K^m$ für $j = 1, \ldots, n$ und zerlegen w_j nach der Basis $(e_i)_{i=1,\ldots,m}$: $w_j = \sum_{i=1}^m a_{ij} \cdot e_i$. $\square$

Ein Vektor $v \in K^n$ hat bezüglich der Standardbasis die Darstellung $v = \sum_{j=1}^n x_j \cdot e_j$. Dabei sind die Skalare x_j die Koordinaten von v. Dann gilt für die lineare Abbildung aus Satz 2.26

$$
\begin{aligned}
F(v) = F\Big(\sum_{j=1}^n x_j \cdot e_j\Big) &= \sum_{j=1}^n x_j \cdot F(e_j) && \text{(Linearität von } F\text{)} \\
&= \sum_{j=1}^n x_j \cdot \Big(\sum_{i=1}^m a_{ij} \cdot e_i\Big) && \text{(Satz 2.26)} \\
&= \sum_{i=1}^m \Big(\sum_{j=1}^n a_{ij} x_j\Big) \cdot e_i && \text{(Umordnen der Summen)}
\end{aligned}
$$

Nun können wir $F(v) = w \in K^m$ wieder nach der Standardbasis des K^m zerlegen und erhalten so für die Koordinaten y_i von $w = \sum_{i=1}^m y_i \cdot e_i$

$$
F(e_j) = \sum_{i=1}^m a_{ij} \cdot e_i \ , \ v = \sum_{j=1}^n x_j \cdot e_j \ , \ F(v) = w = \sum_{i=1}^m y_i \cdot e_i \quad \Rightarrow \quad y_i = \sum_{j=1}^n a_{ij} x_j \ . \tag{*}
$$

Es ist nun üblich (und zweckmäßig), die Koordinaten der Vektoren $v \in K^n$ und $w \in K^m$ als sogenannte *Spaltenvektoren* darzustellen. Wir schreiben

$$
v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} , \qquad w = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} .
$$

Die Gleichung (*) schreibt sich nun als

$$
\begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \ldots & a_{1j} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & \ldots & a_{ij} & \ldots & a_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mj} & \ldots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{pmatrix} .
$$

Die Rechenregel zum Erhalt der i-ten Koordinate w_i ist also, die i-te Zeile $(a_{i1} \ldots a_{ij} \ldots a_{in})$ der Matrix $A = (a_{ij})$ komponentenweise mit dem Spaltenvektor v zu multiplizieren, d.h. man multipliziert den 1. Eintrag der Zeile (also a_{i1}) mit dem 1. Eintrag der Spalte (also x_1), dann den 2. Eintrag der Zeile mit dem 2. Eintrag der Spalte usw. bis zur Multiplikation des n. Eintrags der Zeile mit dem n. Eintrag der Spalte. Schließlich addiert man diese n Zahlen. Das so konstruierte

Produkt $\cdot : M(m \times n, K) \times K^n \to K^m$ heißt das Matrixprodukt einer $m \times n$-Matrix mit einem n-Spaltenvektor. Entsprechend schreiben wir $w = A \cdot v$ für das Matrixprodukt der Matrix $A = (a_{ij})$ mit dem Spaltenvektor v.

Wir sehen also, daß wir auf diese Weise Matrizen $A \in M(m \times n, K)$ mit linearen Abbildungen $F_A : K^n \to K^m$ identifizieren können. Diese Identifikation läßt sich verallgemeinern auf beliebige endlich-dimensionale Vektorräume, sobald eine Basis festgelegt ist.

Satz 2.27 *Gegeben seien K-Vektorräume V mit Basis $\mathcal{A} = (v_1, \ldots, v_n)$ und W mit Basis $\mathcal{B} = (w_1, \ldots, w_m)$.*

i) *Dann gibt es zu jeder linearen Abbildung $F : V \to W$ genau eine Matrix $A = (a_{ij}) \in M(m \times n, K)$, so daß $F(v_j) = \sum_{i=1}^{m} a_{ij} \cdot w_i$.*

ii) *Die so erhaltene Abbildung*

$$M_{\mathcal{B}}^{\mathcal{A}} : \mathrm{Hom}(V, W) \to M(m \times n, K) \, , \qquad M_{\mathcal{B}}^{\mathcal{A}} : F \mapsto A = M_{\mathcal{B}}^{\mathcal{A}}(F)$$

ist ein Isomorphismus von Vektorräumen über K, und insbesondere gilt

$$M_{\mathcal{B}}^{\mathcal{A}}(\lambda \cdot F + \mu \cdot G) = \lambda \cdot M_{\mathcal{B}}^{\mathcal{A}}(F) + \mu \cdot M_{\mathcal{B}}^{\mathcal{A}}(G) \, .$$

Beweis. i) Durch Komposition der Isomorphismen $\Phi_{\mathcal{A}}$ und $\Phi_{\mathcal{B}}$ aus Satz 2.25 mit der eindeutigen Matrix aus Satz 2.26:

$$F = \Phi_{\mathcal{B}} \circ A \circ (\Phi_{\mathcal{A}})^{-1} : V \to W \, , \qquad F(v) := \Phi_{\mathcal{B}}(A \cdot ((\Phi_{\mathcal{A}})^{-1}(v))) \, .$$

Tatsächlich ist zunächst $(\Phi_{\mathcal{A}})^{-1}(v_j) = e_j \in K^n$. Matrixmultiplikation liefert $A \cdot ((\Phi_{\mathcal{A}})^{-1}(v_j)) = \sum_{i=1}^{m} a_{ij} \cdot e_i \in K^m$. Anwenden von $\Phi_{\mathcal{B}}$ liefert $\Phi_{\mathcal{B}}(A \cdot ((\Phi_{\mathcal{A}})^{-1}(v_j))) = \sum_{i=1}^{m} a_{ij} \cdot w_i \in W$.

ii) Die Konstruktion liefert

$$A = M_{\mathcal{B}}^{\mathcal{A}}(F) = (\Phi_{\mathcal{B}})^{-1} \circ F \circ \Phi_{\mathcal{A}} : K^n \to K^m \, .$$

Als Komposition linearer Abbildungen ist $M_{\mathcal{B}}^{\mathcal{A}}$ linear. Da $\mathcal{A}$ eine Basis ist, gibt es nach Satz 2.24.ii) genaus eine lineare Abbildung F mit $F(v_j) = \widetilde{w_j} := \sum_{i=1}^{m} a_{ij} \cdot w_i$. Also ist $M_{\mathcal{B}}^{\mathcal{A}}$ bijektiv. □

Sind also die Basen der Vektorräume V und W festgelegt, dann kann man die Vektorräume mit den Standardräumen K^n, K^m der Koordinaten identifizieren und lineare Abbildungen $F : V \to W$ mit Matizen $A \in M(m \times n, K)$. Eine oftmals sinnvolle Wahl besteht aus Basen, in denen für die Koordinatenmatrix die Einheitsmatrix entsteht:

Definition 2.10 Die Matrix

$$E_r = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \in M(r \times r, K) \; ,$$

in der sämtliche Einträge auf der *Diagonalen* gleich 1 und alle anderen Einträge gleich 0 sind, heißt $(r \times r)$-*Einheitsmatrix*. Ihre Komponenten sind also $E_r = (\delta_{ij})$, wobei

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

das *Kronecker-Symbol* ist.

Satz 2.28 *Sei $F : V \to W$ linear und $\dim(V) = n$, $\dim(W) = m$, $\dim(\mathrm{im}(F)) = \mathrm{rang}(F) = r$. Dann gibt es Basen $\mathcal{A}$ von V und $\mathcal{B}$ von W mit*

$$M_{\mathcal{B}}^{\mathcal{A}}(F) = \begin{pmatrix} E_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix} \; .$$

Dabei ist $0_{p\times q} \in M(p \times q, K)$ die Matrix, in der sämtliche Einträge gleich 0 sind.

Beweis. Wir wählen eine Basis $(w_1, \dots, w_r)$ von $\mathrm{im}(F)$ und ergänzen sie zu einer Basis $\mathcal{B} = (w_1, \dots, w_r, w_{r+1}, \dots, w_m)$ von W. Dann wählen wir wie im Beweis von Satz 2.20 Vektoren $u_1, \dots u_r \in V$ mit $F(u_i) = w_i$ für alle $1 \leq i \leq r$ und ergänzen sie mit einer Basis $(v_1, \dots, v_k)$ von $\ker(F)$ zu einer Basis $\mathcal{A} = (u_1, \dots, u_r, v_1, \dots, v_k)$ von V. Dann hat $M_{\mathcal{B}}^{\mathcal{A}}(F)$ die angegebene Form. □

2.6 Multiplikation von Matrizen

Wir wissen, daß die Komposition von linearen Abbildungen wieder eine lineare Abbildung ist, und daß nach Wahl der Basen jeder linearen Abbildung eine eindeutig bestimmte Matrix entspricht. Am einfachsten ist die Situation für die Vektorräume K^n mit der Standardbasis. Wir bezeichnen in diesem Fall mit $M_m^n := M_{\mathcal{B}}^{\mathcal{A}} : \mathrm{Hom}(K^n, K^m) \to M(m \times n, K)$ den Isomorphismus aus Satz 2.27.

Seien nun $G : K^r \to K^n$ und $F : K^n \to K^m$ lineare Abbildungen, die bezüglich der Standardbasen durch Matrizen $B \in M(n \times r, K)$ und $A \in M(m \times n, K)$ beschrieben werden. Dann ist $F \circ G : K^r \to K^m$ wieder eine lineare Abbildung, welche durch eine Matrix $C \in M(m \times r, K)$ beschrieben wird. Es ist naheliegend, die Matrix C als Produkt $C = A \cdot B$ der Matrizen A und B zu betrachten. Das ist eine sinnvolle Bezeichnung, da dann

$$F \circ G = (M_m^n)^{-1}(A) \circ (M_n^r)^{-1}(B) = (M_m^r)^{-1}(A \cdot B) \; .$$

Man sagt, die Komposition der Homomorphismen F und G induziert über die mit M_q^p bezeichneten Vektorraum-Isomorphismen das Produkt von Matrizen. Wichtig ist dabei, daß nur "aneinander passende" Abbildungen komponiert werden können, so daß nur Produkte

$$\cdot : M(m \times n, K) \times M(n \times r, K) \to M(m \times r, K)$$

erklärt sind. Die Zahl der Spalten der linken Matrix muß also gleich der Zahl der Zeile der rechten Matrix sein.

Da die Komposition von (sogar allgemeinen, insbesondere linearen) Abbildungen assoziativ ist (Satz 1.2), ist auch das resultierende Matrixprodukt assoziativ, d.h. für alle $A \in M(m \times n, K)$, $B \in M(n \times r, K)$ und $C \in M(r \times s, K)$ gilt $(A \cdot B) \cdot C = A \cdot (B \cdot C)$.

Nach diesen abstrakten Überlegungen werden wir das Matrixprodukt explizit für die Komponenten der Matrizen aufschreiben. Seien dazu $A = (a_{ij}) \in M(m \times n, K)$, $B = (B_{jk}) \in M(n \times r, K)$ und $A \cdot B = C = (c_{ik}) \in M(r \times s, K)$ die Matrizen, die den Abbildungen $F : K^n \to K^m$, $G : K^r \to K^n$ und $F \circ G : K^r \to K^m$ entsprechen. Dann haben wir 2.26

$$\begin{aligned}
(F \circ G)(e_k) &= \sum_{i=1}^{m} c_{ik} \cdot e_i && \text{(Satz 2.26 für } F \circ G\text{)} \\
&= F(G(e_k)) && \text{(Definition von } F \circ G\text{)} \\
&= F\Big(\sum_{j=1}^{n} b_{jk} \cdot e_j\Big) && \text{(Satz 2.26 für } G\text{)} \\
&= \sum_{j=1}^{n} b_{jk} \cdot F(e_j) && \text{(Linearität von } F\text{)} \\
&= \sum_{j=1}^{n} b_{jk} \cdot \Big(\sum_{i=1}^{m} a_{ij} \cdot e_i\Big) && \text{(Satz 2.26 für } F\text{)} \\
&= \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} b_{jk}\Big) \cdot e_i && \text{(Umordnen der Summen)}
\end{aligned}$$

Da die $(e_i)_{i=1,\dots,m}$ eine Basis bilden, folgt aus der Eindeutigkeit der Koeffizienten die *Multiplikationsregel für Matrizen*

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

Ein Spezialfall dieser Rechenregel ist das schon zuvor erklärte Matrixprodukt $A \cdot v \in K^m$ einer Matrix $A \in M(m \times n, K)$ mit einem Zeilenvektor $v \in K^n$.

Beispiele 2.7 Eine Drehung in der Ebene um den Ursprung mit Winkel α wird durch die Matrix $A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ beschrieben. Entsprechend beschreibt $B = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix}$ eine Drehung mit Winkel β. Die Hintereinanderausführung beider Drehungen ist durch das Matrixprodukt gegeben:

$$\begin{aligned} A \cdot B &= \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \cdot \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \\ &= \begin{pmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & -\sin\alpha\cos\beta - \cos\alpha\sin\beta \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & \cos\alpha\cos\beta - \sin\alpha\sin\beta \end{pmatrix} . \end{aligned}$$

Andererseits ist die Hintereinanderausführung beider Drehungen wieder eine Drehung mit Gesamtwinkel $\alpha + \beta$. Daraus folgen die Additionstheoreme für Sinus und Cosinus,

$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta \ , \qquad \cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$$

Satz 2.29 *Sind Matrizen* $A, A' \in M(m \times n, K)$, $B, B' \in M(n \times s, K)$ *und* $C \in M(r \times s, K)$ *gegeben sowie* $\lambda \in K$, *so gilt:*

i) $A \cdot (B + B') = A \cdot B + A \cdot B'$ *und* $(A + A') \cdot B = A \cdot B + A' \cdot B$ *(Distributivgesetze)*

ii) $A \cdot (\lambda \cdot B) = (\lambda \cdot A) \cdot B = \lambda \cdot (A \cdot B)$

iii) $(A \cdot B) \cdot C = A \cdot (B \cdot C)$ *(Assoziativgesetz)*

iv) $E_m \cdot A = A \cdot E_n = A$ *(Neutralität der Einheitsmatrix)*

Beweis. i) und ii) folgen aus der Summendarstellung des Matrixprodukts und iii) und iv) aus der Konstruktion über Homomorphismen der Standardräume. □

Im weiteren schreiben wir oft AB statt $A \cdot B$ für das Produkt der Matrizen.

Definition 2.11 Die Abbildung $t : M(m \times n, K) \to M(n \times m, K)$ definiert durch $t : A = (a_{ij}) \mapsto A^t := (a_{ji})$ heißt *Transposition von Matrizen.*

Satz 2.30 *Wenn das Matrixprodukt* AB *erklärt ist, dann gilt* $(A \cdot B)^t = B^t \cdot A^t$.

Beweis. Sei $A = (a_{ij}) \in M(m \times n, K)$, $B = (b_{jk}) \in M(n \times r, K)$ und $C = AB = (c_{ik}) \in M(m \times r, K)$. Dann ist $c_{ik} = \sum_{j=1}^n a_{ij} b_{jk}$ und $C^t = (c'_{ki})$ mit $c'_{ki} = c_{ik}$. Weiter ist $B^t = (b'_{kj})$ mit $b'_{kj} = b_{jk}$ und $A^t = (a'_{ji})$ mit $a'_{ji} = a_{ij}$. Dann ist $B^t A^t = D = (d_{ki})$ mit $d_{ki} = \sum_{j=1}^n b'_{kj} a'_{ji} = \sum_{j=1}^n b_{jk} a_{ij} = \sum_{j=1}^n a_{ij} b_{jk} = c_{ik} = c'_{ki}$ und somit $D = C^t$. □.

Satz 2.31 *Die Menge* $M(n, K) := M(n \times n, K)$ *der quadratischen Matrizen ist ein Ring.* □

Definition 2.12 Eine quadratische Matrix $A \in M(n,K)$ heißt *invertierbar*, wenn es ein $A^{-1} \in M(n,K)$ gibt mit $A \cdot A^{-1} = A^{-1} \cdot A = E_n$.

Satz 2.32 *Die Menge*

$$GL(n,K) = \{A \in M(n,K) \ : \ A \textit{ ist invertierbar}\}$$

der invertierbaren Matrizen ist eine Gruppe mit der Matrixmultiplikation als Verknüpfung und der Einheitsmatrix E_n als neutralem Element.

Diese Gruppe $GL(n,K)$ heißt die *allgemeine lineare Gruppe.*

Beweis. Da E_n ein neutrales Element ist, ist nur zu zeigen, daß aus $A, B \in GL(n,K)$ auch $AB \in GL(n,K)$ folgt. Die Matrix $B^{-1}A^{-1}$ erfüllt $(B^{-1}A^{-1})AB = E_n$ unter Verwendung des Assoziativgesetzes. Aus den allgemeinen Gruppeneigenschaften folgt, daß das neutrale Element und das zu A inverse Element eindeutig sind. □

Unter Verwendung des Isomorphismus zwischen Matrizen $A \in M(m \times n, K)$ und linearen Abbildungen $F : K^n \to K^m$ definieren wir den *Rang einer Matrix* $A \in M(m \times n, K)$ zu $\mathrm{rang}(A) = \dim(A \cdot K^n)$, wobei $A \cdot K^n = \{A \cdot v \ : \ v \in K^n\} \subset K^m$.

Satz 2.33

i) *Für* $A \in M(m \times n, K)$ *ist* $\mathrm{rang}(A) = \mathrm{rang}(A^t)$.

ii) $A \in M(n,K)$ *ist genau dann invertierbar, wenn* $\mathrm{rang}(A) = n$.

iii) *Für* $A \in GL(n,K)$ *gilt* $(A^t)^{-1} = (A^{-1})^t$.

Beweis. i) Die Definition der Dimension eines Vektorraumes ist unabhängig von der Wahl der Basis. Wir können nach Satz 2.28 angepaßte Basen $\mathcal{A}$ in K^n und $\mathcal{B}$ in K^m wählen, so daß $M^{\mathcal{A}}_{\mathcal{B}}(F) = \begin{pmatrix} E_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix}$. Die transponierte Matrix entspricht dann der Abbildung $F^t : K^m \to K^n$ mit $M^{\mathcal{B}}_{\mathcal{A}}(F^t) = \begin{pmatrix} E_r & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(r-r)} \end{pmatrix}$. Damit ist $\dim(\mathrm{im}(F^t)) = r = \dim(\mathrm{im}(F))$ und somit $\mathrm{rang}(A^t) = \mathrm{rang}(A)$.

ii) Ist $\mathrm{rang}(A) = \mathrm{rang}(F) = n$, dann ist F surjektiv nach Satz 2.19.ii) und dann bijektiv nach Satz 2.22.

iii) Wir haben $(A^{-1})^t A^t = (AA^{-1})^t = E_n^t = E_n$, so daß $(A^{-1})^t$ das Inverse zu A^t ist. □

2.7 Lineare Gleichungssysteme

Definition 2.13 Eine Teilmenge $X \subset V$ eines K-Vektorraumes V heißt *affiner Unterraum*, falls es ein $v \in V$ und einen Untervektorraum $W \subset V$ gibt, so daß

$$X = v + W := \{x \in V \ : \ \text{es gibt ein } w \in W \text{ mit } x = v + w\} \ .$$

Zur Veranschaulichung folgendes Beispiel. Ist $V = \mathbb{R}^2$ (also die Ebene), dann sind die eindimensionalen Untervektorräume W die Geraden durch den Nullpunkt. Ein aus W hervorgehender affiner Unterraum $X = v + W$ ist dann eine Gerade, die nicht durch 0, sondern durch v geht.

Satz 2.34 *Ist $X = v + W$ ein affiner Unterraum, dann gilt*

i) *Für beliebiges $v' \in X$ ist $X = v' + W$.*

ii) *Ist $v' \in V$ und $W' \subset V$ ein Untervektorraum mit $v' + W' = v + W$, so folgt $W = W'$ und $v' - v \in W$.*

Beweis. i) Wegen $v' \in X$ gibt es ein $w' \in W$ mit $v' = v + w'$. Sei $x = v + w \in X$ ein beliebiges Element, so schreibt sich $x = v' + w - w'$. Also ist $X \subset v' + W$. Durch Vertauschen von v, v' folgt $v' + W = X$.

ii) Seien $x, x' \in X = v + W$, dann ist $x - x' \in W$. Gäbe es eine zweite Darstellung $X = v' + W'$, dann ist $x - x' \in W'$ für alle x, x', also $W = W'$. Dann folgt $v' - v \in W$. $\square$

Da also in einem affinen Unterraum $X = v + W$ der Untervektorraum eindeutig ist, definieren wir die Dimension $\dim(X) := \dim(W)$ wenn $X = v + W$.

Definition 2.14 Für eine gegebene Matrix $A = (a_{ij}) \in M(m \times n, K)$, einen gegebenen Vektor $b = (b_i) \in K^m$ und einen unbekannten (und zu bestimmenden Vektor) $x = (x_j) \in K^n$ heißt

$$A \cdot x = b \tag{*}$$

ein *lineares Gleichungssystem aus m Gleichungen mit n Unbekannten*. Die aus (*) abgeleitete Gleichung

$$A \cdot x = 0 \tag{**}$$

heißt das zu (*) gehörende *homogene System*. Die Matrix $A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}$ heißt die *Koeffizientenmatrix* und die Matrix $(A, b) = \begin{pmatrix} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{pmatrix} \in M(m \times (n+1), K)$ heißt *erweiterte Koeffizientenmatrix*. Die Menge

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

heißt *Lösungsraum* des linearen Gleichungssystems.

In Komponentenschreibweise besteht also ein lineares Gleichungssystem aus m Gleichungen $\sum_{j=1}^{n} a_{ij}x_j = b_i$ zur Bestimmung der n unbekannten Koeffizienten $x_1, \dots, x_n$. Zunächst werden wir den Lösungsraum abstrakt charakterisieren und dann ein Verfahren (den Gaußschen Algorithmus) angeben, um den Lösungsraum in konkreten Beispielen zu bestimmen.

Ist $F : K^n \to K^m$ die lineare Abbildung, welche in der Standardbasis durch die Matrix $A \in M(m \times n, K)$ gegeben ist, so ist $\text{Lös}(A, b) = F^{-1}(b)$ und insbesondere $\text{Lös}(A, 0) = F^{-1}(0) = \ker(F)$. Damit gilt stets $0 \in \text{Lös}(A, 0)$, und diese Lösung heißt die *triviale Lösung* des homogenen Systems.

Satz 2.35 *Gegeben sei das lineare Gleichungssystem $Ax = b$ aus m Gleichungen mit n Unbekannten. Ist* $\text{rang}(A) = r$*, dann gilt für die Lösungsräume:*

i) $\text{Lös}(A, 0) \subset K^n$ *ist ein Untervektorraum der Dimension $n - r$.*

ii) $\text{Lös}(A, b) \subset K^n$ *ist entweder leer oder ein affiner Raum der Dimension $n - r$. Ist $v \in$* $\text{Lös}(A, b)$ *eine beliebige Lösung, dann gilt* $\text{Lös}(A, b) = v + \text{Lös}(A, 0)$.

Beweis. i) $\text{Lös}(A, 0) = \ker(F)$ ist nach Satz 2.19 ein Untervektorraum. Nach Satz 2.20 ist seine Dimension $\dim(\ker(F)) = \dim(K^n) - \dim(\text{im}(F)) = n - r$.

ii) Seien $v, v' \in \text{Lös}(A, b)$ zwei Lösungen des linearen Gleichungssystems, also $Av = b$ und $Av' = b$. Dann ist $A(v - v') = 0$, also $v - v' \in \text{Lös}(A, 0)$. $\square$

Der Satz besagt, daß man eine allgemeine Lösung des inhomogenen linearen Gleichungssystems erhält, indem man zu einer spezielen Lösung die allgemeine Lösung des homogenen Systems addiert. Sei also $(w_1, \dots, w_{n-r})$ eine Basis von $\text{Lös}(A, 0)$ und $v \in \text{Lös}(A, b)$ eine beliebige spezielle Lösung des inhomogenen Systems, so ist

$$\text{Lös}(A, b) = v + K \cdot w_1 + \dots + K \cdot w_{n-r} \ .$$

Satz 2.36 *Der Lösungsraum eines linearen Gleichungssystems $Ax = b$ ist genau dann nicht leer, wenn* $\text{rang}(A, b) = \text{rang}(A)$.

Beweis. Die Matrizen $A \in M(m \times n, K)$ und $(A, b) \in M(m \times (n + 1), K)$ beschreiben lineare Abbildungen $A : K^n \to K^m$ bzw. $A' : K^{n+1} \to K^m$ mit $A(e_j) = A'(e_j) = \sum_{i=1}^{m} a_{ij}e_i$ für $1 \leq j \leq n$ und $A'(e_{n+1}) = b$. Damit ist $\text{im}(A) \subset \text{im}(A')$, also $\text{rang}(A) \leq \text{rang}(A')$, sowie $b \in \text{im}(A')$. Ist nun $\text{rang}(A) = \text{rang}(A')$, dann ist $\text{im}(A) = \text{im}(A')$ und damit $b \in \text{im}(A)$, das Gleichungssystem ist also lösbar.

Ist $\text{rang}(A) < \text{rang}(A')$, dann muß es Vektoren $v \in \text{im}(A')$ geben, die nicht in $\text{im}(A)$ liegen. Da $\text{span}(A'(e_j))_{j=1,\dots,n} \subset \text{im}(A)$, bleiben nur die b enthaltenden Linearkombinationen, welche nicht in $\text{im}(A)$ liegen. Das bedeutet $b \notin \text{im}(A)$. $\square$

Insbesondere ist der Lösungsraum $\text{Lös}(A, b)$ für $A \in M(m \times n, K)$ nichtleer, wenn $\text{rang}(A) = m$, da dann die durch A beschriebene lineare Abbildung surjektiv ist. Das bedeutet, daß $Ax = b$ für jedes $b \in K^m$ lösbar ist. In diesem Fall heißt das lineare Gleichungssystem *universell lösbar*. Ist $\operatorname{rang}(A) < m$, so gibt es nicht für alle $b \in K^m$ eine Lösung. Es ist nur für jene $b \in K^m$ lösbar, für die die $\operatorname{rang}(A) = \operatorname{rang}(A, b)$ gilt.

Wir sagen, das lineare Gleichungssystem $Ax = b$ ist *eindeutig lösbar*, wenn $\text{Lös}(A, b)$ nur aus einem Element besteht.

Satz 2.37 *Das lineare Gleichungssystem $Ax = b$ für $A \in M(m \times n, K)$ und $b \in K^m$ ist genau dann eindeutig lösbar, wenn* $\operatorname{rang}(A) = \operatorname{rang}(A, b) = n$.

Beweis. Nach Satz 2.36 gibt es eine Lösung, und nach Satz 2.35 ist diese eindeutig. □

Ist speziell $A \in M(n, K)$ eine quadratische Matrix und gilt $\operatorname{rang}(A) = n$, dann ist A invertierbar nach Satz 2.33, und die eindeutige Lösung des linearen Gleichungssystems $Ax = b$ ist gegeben durch $x = A^{-1}b$.

2.8 Das Gaußsche Eliminationsverfahren

Definition 2.15 Eine Matrix $A \in M(m \times n, K)$ heißt in *Zeilenstufenform*, falls für alle $i = 2, \ldots, m$ gilt:

- Sind die ersten $(j - 1)$ Einträge der $(i - 1)$-ten Zeile gleich 0, so sind auch die ersten j Einträge der i-ten Zeile gleich 0, wobei $j = 1, \ldots, n$.

Im allgemeinen wird kein Eintrag $(j - 1 = 0)$ der ersten $(i - 1 = 1)$ Zeile mit 0 beginnen. Dann bedeutet die Definition, daß der 1. Eintrag $(j = 1)$ der 2. Zeile $(i = 2)$ gleich 0 ist. Ist der erste $(j - 1 = 1)$ Eintrag der ersten $(i - 1 = 1)$ Zeile gleich 0, so muß der 2. Eintrag $(j = 2)$ der 2. Zeile $(i = 2)$ gleich 0 sein, usw.

Beispiele 2.8 Ein Beispiel für eine Matrix in Zeilenstufenform ist also:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 9 & 8 & 7 & 6 & 5 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Satz 2.38 *Ist eine Matrix $A \in M(m \times n, K)$ in Zeilenstufenform gegeben, dann ist der Rang von A gleich der Anzahl der Zeilen, in denen nicht nur Nullen stehen.*

Beweis. Für $(e_j)_{j=1,\ldots,n} \in K^n$ ist $\operatorname{span}(A \cdot e_j)_{j=1,\ldots,n} = K^r$. Dabei ist K^r dargestellt als Untervektorraum von K^m, so daß bezüglich der Standardbasis von K^m die letzten $m - r$ Komponenten identisch Null sind. □

Satz 2.39 *Ist* $A \in M(m \times n, K)$ *in Zeilenstufenform gegeben und ist* $\mathrm{rang}(A) = r$*, dann ist das lineare Gleichungssystem* $Ax = b$ *genau dann lösbar, wenn* $b_{r+1} = \cdots = b_m = 0$*.*

Beweis. Folgt aus Satz 2.36. □

Wir geben nun ein Verfahren zur Lösung eines in Zeilenstufenform gebegenen linearen Gleichugssystems $Ax = b$ mit $\mathrm{rang}(A) = \mathrm{rang}(A, b)$ an. Zur Vereinfachung der Bezeichnungen werden die *Spalten* umnumeriert, so daß $a_{ii} \neq 0$ entsteht für $1 \leq i \leq r$.

In Komponenten schreibt sich das Gleichungssystem $\sum_{j=1}^{n} a_{ij} x_j = b_i$. Dabei ist j ein *Summationsindex*, und das geordnete n-Tupel $(1, \ldots, n)$ kann durch eine beliebige *Permutation* $(\pi(1), \pi(2), \ldots, \pi(n))$ ersetzt werden, wobei π eine bijektive Abbildung der Menge $1, \ldots, n$ auf sich selbst ist. Das geschieht wie folgt:

- Ist $a_{11} = 0$, dann tauschen wir den Summationsindex j in *zyklischer Weise*:
$$n \leftarrow 1 \leftarrow 2 \leftarrow \cdots \leftarrow n-1 \leftarrow n \,.$$
Das bedeutet, wir nehmen die erste Spalte weg und setzen sie ans Ende wieder an, alle anderen Spalten rücken um eine nach links. Damit sich das Gleichungssystem nicht ändert, sind ebenfalls die Unbekannten x_i an die entsprechende Stelle zu bringen. Dazu führen wir den Spaltenvektor $y \in K^n$, zunächst mit Komponenten $y_i = x_i$. Diese sind nun zyklisch umzubenennen in $y_j \mapsto y_{j-1}$ für $2 \leq j \leq n$ und $y_1 \mapsto y_n$. Dabei ändert sich nichts an den Indizes i, also an den Zeilen von A und dem Spaltenvektor b.

  Das Verfahren wird solange wiederholt, bis (für die neue Matrix A) $a_{11} \neq 0$. Die neue Matrix A bleibt bei diesen Umnumerierungen in Zeilenstufenform.

- Wir haben also $a_{11} \neq 0$ und $a_{21} = 0$, da A in Zeilenstufenform. Ist $a_{22} = 0$, so lassen wir die Spalte $j = 1$ unverändert und permutieren $j = 2, \ldots, n$ in zyklischer Weise, d.h. wir bringen die 2. Spalte ans Ende und benennen $y_j \mapsto y_{j-1}$ für $3 \leq j \leq n$ und $y_2 \mapsto y_n$.

  Das Verfahren wird solange wiederholt, bis $a_{22} \neq 0$.

- Dann betrachtet man a_{33} und erreicht schließlich durch wiederholte zyklische Permutationen des Summationsindex, daß alle $a_{ii} \neq 0$ sind für $1 \leq i \leq r$, während $a_{ij} = 0$ für $i > r$ und alle j.

In Beispiel 2.8 haben wir zunächst $a_{33} = 0$, was zu folgender äquivalenter Umformung führt:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 9 & 8 & 7 & 6 & 5 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 4 & 5 & 6 & 3 \\ 0 & 9 & 7 & 6 & 5 & 8 \\ 0 & 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_4 \\ x_5 \\ x_6 \\ x_3 \end{pmatrix} .$$

Nach Lösung des Gleichungssystems (im folgenden beschrieben) sind die y_i wieder durch die x_i auszudrücken.

Das Gleichungssystem $A \cdot y = b$ wird nun *von der r-ten Zeile aufwärts* gelöst. Die Komponenten

$$y_{r+1} = w_1 \in K \ , \quad y_{r+2} = w_2 \in K \ , \quad \ldots, \quad y_n = w_{n-r} \in K$$

sind *beliebig* zu wählen, da diese Vektoren gerade den $(n-r)$-dimensionalen Untervektorraum aufspannen, welcher nach Satz 2.35 die Lösung des homogenen Systems liefert. Damit schreibt sich die r-te Gleichung als $a_{rr}y_r + \sum_{j=r+1}^n a_{rj}w_{j-r} = b_r$. Da $a_{rr} \neq 0$, erhält man

$$y_r = \frac{1}{a_{rr}} \Big(b_r - \sum_{j=r+1}^{n} a_{rj} w_{j-r} \Big) \ .$$

Ist $n = r$, so haben wir natürlich $y_r = \frac{b_r}{a_{rr}}$. Wir schreiben diese Gleichung etwas allgemeiner als

$$y_r = \sum_{i=1}^{r} d_{ri} b_i + \sum_{l=1}^{k} c_{rl} w_l$$

mit $k = n - r$. Wir lesen $d_{rr} = \frac{1}{a_{rr}}$ und $c_{rl} = -\frac{a_{r,r+l}}{a_{rr}}$ ab. Alle d_{ir} mit $i < r$ verschwinden.

Damit ist die $(r-1)$-te Gleichung gegeben durch

$$a_{r-1,r-1} y_{r-1} + a_{r-1,r} y_r + \sum_{j=r+1}^{n} a_{r-1,j} y_j = b_{r-1} \ ,$$

aus der sich durch Einsetzen von y_r die Lösung

$$y_{r-1} = \sum_{i=1}^{r} d_{r-1,i} b_i + \sum_{l=1}^{k} c_{r-1,l} w_l$$

ergibt. Auf diese Weise geht man Schritt für Schritt aufwärts in den Zeilen der Matrix und bestimmt so alle $y_1, \dots, y_r$ in Abhängigkeit der b_i und der Parametrisierung $w_1, \dots, w_k$ des $k = (n-r)$-dimensionalen Untervektorraums:

$$\begin{aligned} y_j &= \sum_{i=1}^{r} d_{ji} b_i + \sum_{l=1}^{k} c_{jl} w_l \ , && 1 \le j \le r \ , \\ y_j &= w_{j-r} \ , && r < j \le n \ . \end{aligned}$$

In Matrixschreibweise bedeuten diese Gleichungen $y = D \cdot b + C \cdot w$ mit

$$\begin{aligned} D &= \begin{pmatrix} D' \\ 0_{k \times r} \end{pmatrix} \in M(n \times r, K) \ , && D' = (d_{ji}) \in M(r \times r, K) \ , \\ C &= \begin{pmatrix} C' \\ E_k \end{pmatrix} \in M(n \times k) \ , && C' = (c_{jl}) \in M(r \times k, K) \ . \end{aligned}$$

Wir erhalten also lineare Abbildungen

$$\begin{aligned} \varphi &: K^r \to K^n \ , && \varphi : b \mapsto D \cdot b \ , \\ \Phi_0 &: K^k \to K^n \ , && \Phi_0 : w \mapsto C \cdot w \ . \end{aligned}$$

sowie als Summe eine Abbildung

$$\Phi_b : K^k \to K^n \ , \qquad \Phi_b(w) := \varphi(b) + \Phi_0(w) \ .$$

Die Abbildung Φ_0 ist injektiv, denn $C \cdot w = \begin{pmatrix} C' \cdot w \\ w \end{pmatrix}$, so daß $\ker(\Phi_0) = \{0\}$. Aus Satz 2.20 folgt dann $\dim(\mathrm{im}(\Phi_0)) = k$. Wir wissen, daß $\Phi_0(K^n) \subset \mathrm{Lös}(A, 0)$, denn $y = C \cdot w$ ist nach Konstruktion eine Lösung von $A \cdot y = b$. Da $\dim(\mathrm{Lös}(A, 0)) = k$ nach Satz 2.35, folgt aus der Gleichheit der Dimensionen $\Phi_0(K^k) = \mathrm{Lös}(A, 0)$.

Da $y = \varphi(b)$ nach Konstruktion eine spezielle Lösung des inhomogenen Gleichungssystems $A \cdot y = b$ ist, gilt $\mathrm{Lös}(A, b) = \varphi(b) + \mathrm{Lös}(A, 0) = \Phi_b(K^k)$. Insgesamt ist damit der folgende Satz bewiesen:

Satz 2.40 *Sei $A \in M(m \times n, K)$ in (spezieller) Zeilenstufenform gegeben mit* $\mathrm{rang}(A, b) = \mathrm{rang}(A) = r$ *und $a_{ii} \neq 0$ für $1 \le i \le r$. Dann hat die zuvor konstruierte b-abhängige Parametrisierung $\Phi_b : K^{n-r} \to \mathrm{Lös}(A, b) \subset K^n$ folgende Eigenschaften:*

i) $\Phi_0 : K^{n-r} \to \mathrm{Lös}(A, b) \subset K^n$ *ist ein Isomorphismus von Vektorräumen.*

ii) $\Phi_b : K^{n-r} \to \mathrm{Lös}(A, b) \subset K^n$ *ist für jedes $b \in K^r$ bijektiv.*

iii) *Es gibt eine lineare Abbildung $\varphi : K^r \to K^n$, so daß für alle $b \in K^r$ gilt: $\Phi_b = \varphi(b) + \Phi_0$ und $Lös(A, b) = \varphi(b) + Lös(A, 0)$.*

Man erkennt hier die Bedeutung der Vektorraum-Techniken, die uns garantieren, daß das Verfahren wirklich *alle* Lösungen generiert.

Wir lösen Beispiel 2.8 für geeignetes b:

$$\left(\begin{array}{cccc|cc} 1 & 2 & 4 & 5 & 6 & 3 \\ 0 & 9 & 7 & 6 & 5 & 8 \\ 0 & 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \left(\begin{array}{c} x_1 \\ x_2 \\ x_4 \\ x_5 \\ \hline x_6 \\ x_3 \end{array}\right) = \left(\begin{array}{c} b_1 \\ b_2 \\ b_3 \\ b_4 \\ \hline b_5 \end{array}\right) .$$

Die Lösbarkeit (Rangbedingung) erfordert $b_5 = 0$. Dann ist $y_5 = \underline{x_6 = w_1}$ und $y_6 = \underline{x_3 = w_2}$ beliebig zu wählen. Die 4. Gleichung liefert $\underline{x_5} = b_4 - 2x_6 \underline{= b_4 - 2w_1}$. Die 3. Gleichung wird $x_4 + 2x_5 + 3x_6 = x_4 + (2b_4 - 4x_6) + 3x_6 = b_3$, also $\underline{x_4 = b_3 - 2b_4 + w_1}$. Nun ist die 2. Gleichung

$$\begin{aligned} & 9x_2 + 7x_4 + 6x_5 + 5x_6 + 8x_3 \\ & = 9x_2 + (7b_3 - 14b_4 + 7w_1) + (6b_4 - 12w_1) + 5w_1 + 8w_2 \\ & = 9x_2 + 7b_3 - 8b_4 + 8w_2 = b_2 , \end{aligned}$$

also $\underline{x_2 = \frac{1}{9}b_2 - \frac{7}{9}b_3 + \frac{8}{9}b_4 - \frac{8}{9}w_2}$. Schließlich entsteht für die erste Gleichung

$$\begin{aligned} & x_1 + 2x_2 + 4x_4 + 5x_5 + 6x_6 + 3x_3 \\ & = x_1 + (\tfrac{2}{9}b_2 - \tfrac{14}{9}b_3 + \tfrac{16}{9}b_4 - \tfrac{16}{9}w_2) + (4b_3 - 8b_4 + 4w_1) + (5b_4 - 10w_1) + 6w_1 + 3w_2 \\ & = x_1 + \tfrac{2}{9}b_2 + \tfrac{22}{9}b_3 - \tfrac{11}{9}b_4 + \tfrac{11}{9}w_2 = b_1 , \end{aligned}$$

und damit $\underline{x_1 = b_1 - \frac{2}{9}b_2 - \frac{22}{9}b_3 + \frac{11}{9}b_4 - \frac{11}{9}w_2}$. Damit erhalten wir in Matrixform

$$\left(\begin{array}{c} x_1 \\ x_2 \\ x_4 \\ x_5 \\ \hline x_6 \\ x_3 \end{array}\right) = \left(\begin{array}{cccc} 1 & -\frac{2}{9} & -\frac{22}{9} & \frac{11}{9} \\ 0 & \frac{1}{9} & -\frac{7}{9} & \frac{8}{9} \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \left(\begin{array}{c} b_1 \\ b_2 \\ b_3 \\ b_4 \end{array}\right) + \left(\begin{array}{cc} 0 & -\frac{11}{9} \\ 0 & -\frac{8}{9} \\ 1 & 0 \\ -2 & 0 \\ \hline 1 & 0 \\ 0 & 1 \end{array}\right) \left(\begin{array}{c} w_1 \\ w_2 \end{array}\right) .$$

Wir geben nun eine Strategie an, um ein lineares Gleichungssystem, dessen Matrix nicht in Zeilenstufenform gegeben ist, in ein dazu äquivalentes Gleichungssystem in Zeilenstufenform zu überführen. Dazu werden wir Umformungen der *Zeilen* vornehmen; die Spalten werden nicht verändert. Wir betrachten die folgenden *elementaren Zeilenumformungen der erweiterten Koeffizientenmatrix*:

Typ I Vertauschen der i-ten mit der k-ten Zeile.

Typ II Ersetzen der k-ten Zeile durch die Summe aus der k-ten Zeile und dem λ-fachen der i-ten Zeile sowie Beibehalten der i-ten Zeile an ihrer Stelle.

(Dabei sind Vielfache und Summen von Zeilen im Sinne der Linearkombinationen von Vektoren zu verstehen, d.h. ein λ-faches ist Multiplikation *jedes Eintrags der Zeile mit* λ und Addition von Zeilen bedeutet Addition der an gleicher Stelle stehenden Einträge der beiden Zeilen.)

Wir zeigen, daß sich die Lösungsmenge unter diesen elementaren Zeilenumformungen nicht ändert.

Satz 2.41 *Sei (A, b) die erweiterte Koeffizientenmatrix eines linearen Gleichungssystems und $(\tilde{A}, \tilde{b})$ aus (A, b) durch endlich viele elementare Zeilenumformungen entstanden. Dann haben die Systeme $A \cdot x = b$ und $\tilde{A} \cdot x = \tilde{b}$ die gleichen Lösungsräume,* $\text{Lös}(A, b) = \text{Lös}(\tilde{A}, \tilde{b})$.

Beweis. Es genügt zu beweisen, daß eine elementare Zeilenumformung den Lösungsraum invariant läßt. Typ I ist unproblematisch, da die Menge der Gleichungen gleich bleibt und sich nur die Reihenfolge der Gleichungen ändert.

Typ II: Die i-te und k-te Gleichung sind

$$\begin{aligned} a_{i1}x_1 + \cdots + a_{in}x_n &= b_i \\ a_{k1}x_1 + \cdots + a_{kn}x_n &= b_k \ . \end{aligned}$$

Die Zeilenumformung ersetzt dieses Paar durch

$$\begin{aligned} a_{i1}x_1 + \cdots + a_{in}x_n &= b_i \\ (a_{k1} + \lambda a_{i1})x_1 + \cdots + (a_{kn} + \lambda a_{in})x_n &= b_k + \lambda b_i \ . \end{aligned}$$

Wenn $x = (x_1, \ldots, x_n)$ das erste System erfüllt, so erfüllt es auch das zweite, und umgekehrt. □

Mittels elementarer Zeilenumformungen kann die erweiterte Koeffizientenmatrix (A, b) in Zeilenstufenform gebracht werden.

- Wir suchen die am weitesten links stehende *Spalte*, in der es einen Eintrag ungleich 0 gibt. Sei das die l-te Spalte und sei $a_{il} \neq 0$. Ist $i \neq 1$, so tauschen wit mit Typ I die *Zeilen* i und 1. (Ist $a_{1l} \neq 0$, so kann auf das Austauschen verzichtet werden.) In $(\tilde{A}, \tilde{b}) = (c_{ij})$ ist dann $c_{1l} \neq 0$ und $c_{ij} = 0$ für alle $j < l$ und alle i.
- Für alle Zeilen $i > 1$ wenden wir jetzt Typ II an, indem wir zur i-ten Zeile das $(-\frac{c_{il}}{c_{1l}})$-fache der 1. Zeile addieren. (Dabei ist nur für $c_{il} \neq 0$ wirklich etwas zu tun.) Bezeichnen wir die so entstehende erweiterte Koeffizientenmatrix wieder mit (c_{ij}), so ist $c_{1l} \neq 0$, $c_{il} = 0$ für alle $i > 1$ und $c_{ij} = 0$ für alle $j < l$ und alle i. Die erste Zeile hat also bereits die Zeilenstufeneigenschaft.

- Wir suchen die am weitesten links stehende Spalte der neuen erweiterten Koeffizientenmatrix (A, b), in der es Einträge $a_{il} \neq 0$ gibt mit $i > 1$. Sei das die l-te Spalte. Ist $a_{2l} = 0$ aber $a_{il} \neq 0$, so tauschen wir mit Typ I die i-te Zeile gegen die 2. aus. Dann ist $c_{2l} \neq 0$, und wir addieren die entsprechenden Vielfache der neuen 2. Zeile zu allen anderen, so daß $c_{il} = 0$ für alle $i > 2$ entsteht.
- Das Verfahren bricht ab, wenn die neue erweiterte Koeffizientenmatrix in Zeilenstufenform vorliegt. Das Gleichungssystem kann dann mit der zuvor beschriebenen Methode gelöst werden.

Die elementaren *Zeilen*umformungen bringen also die erweiterte Koeffizientenmatrix (A, b) in Zeilenstufenform $(\tilde{A}, \tilde{b})$ mit $Ax = b \Leftrightarrow \tilde{A}x = \tilde{b}$. Anschließend erreichen wir durch Permutieren der *Spalten* von $(\tilde{A}, \tilde{b}) = (a_{ij})$ und der Unbekannten x_j, daß $a_{ii} \neq 0$ für $1 \leq i \leq r$. Dabei bleibt $\tilde{b}$ unverändert. Zwar kann das Gleichungssystem nun leicht gelöst werden, wir können aber die elementaren Zeilenumformungen ein weiteres Mal anwenden, um die erweiterte Koeffizientenmatrix weiter zu vereinfachen.

- Beginnend mit $i = r$ und dann Schritt für Schritt aufsteigend bis $i = 2$ addieren wir zu jeder darüberliegenden Zeile $k < i$ das $(-\frac{a_{ki}}{a_{ii}})$-fache der i-ten Zeile. Im Ergebnis hat die erweiterte Koeffizientenmatrix die Form

$$\begin{pmatrix} D_{r\times r} & B_{r\times(n-r)} & b \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} & 0 \end{pmatrix},$$

  wobei D eine Diagonalmatrix ist, d.h. $d_{ij} = 0$ für $i \neq j$. (Es ist angenommen, daß $b \in K^r$, denn sonst ist der Lösungsraum leer.)

Das Lösen des Gleichungssystems wird für diese spezielle Zeilenstufenform wesentlich erleichtert. Noch übersichtlicher wird es, wenn wir eine weitere Zeilenumformung zulassen:

Typ III Multiplikation der i-ten Zeile mit $\lambda \neq 0$.

Mit Typ III läßt sich erreichen, daß die obige Diagonalmatrix in eine Einheitsmatrix umgewandelt wird. Nun läßt sich die Lösung des Gleichungssystems

$$\begin{pmatrix} E_r & B_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix} \begin{pmatrix} y \\ w \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix}$$

sofort ablesen: es ist $y = b - B \cdot w$ mit $w \in K^{n-r}$.

Beispiele 2.9 Wir werden das Verfahren an einem Beispiel demonstrieren. Wir bezeichnen mit I_{ik} und II_{ik} die Zeilenumformungen vom Typ I bzw. Typ II (mit

Angabe der betroffenen Zeilen) und mit π_n die Permutation, die die n-te Spalte an das Ende der Koeffizientenmatrix bringt, d.h. also *direkt vor die Spalte der* b_i.

$$\left(\begin{array}{rrrrr||l} 1 & 3 & 5 & 2 & 0 & b_1 \\ 3 & 9 & 10 & 1 & 2 & b_2 \\ 0 & 2 & 7 & 3 & -1 & b_3 \\ 2 & 8 & 12 & 2 & 1 & b_4 \end{array}\right) \xrightarrow{\mathrm{II}_{12},\mathrm{II}_{14}} \left(\begin{array}{rrrrr||r} 1 & 3 & 5 & 2 & 0 & b_1 \\ 0 & 0 & -5 & -5 & 2 & -3b_1+b_2 \\ 0 & 2 & 7 & 3 & -1 & b_3 \\ 0 & 2 & 2 & -2 & 1 & -2b_1+b_4 \end{array}\right)$$

$$\xrightarrow{\mathrm{I}_{13}} \left(\begin{array}{rrrrr||r} 1 & 3 & 5 & 2 & 0 & b_1 \\ 0 & 2 & 2 & -2 & 1 & -2b_1+b_4 \\ 0 & 2 & 7 & 3 & -1 & b_3 \\ 0 & 0 & -5 & -5 & 2 & -3b_1+b_2 \end{array}\right) \xrightarrow{\mathrm{II}_{23}} \left(\begin{array}{rrrrr||r} 1 & 3 & 5 & 2 & 0 & b_1 \\ 0 & 2 & 2 & -2 & 1 & -2b_1+b_4 \\ 0 & 0 & 5 & 5 & -2 & 2b_1+b_3-b_4 \\ 0 & 0 & -5 & -5 & 2 & -3b_1+b_2 \end{array}\right)$$

$$\xrightarrow{\mathrm{II}_{34}} \left(\begin{array}{rrrrr||r} 1 & 3 & 5 & 2 & 0 & b_1 \\ 0 & 2 & 2 & -2 & 1 & -2b_1+b_4 \\ 0 & 0 & 5 & 5 & -2 & 2b_1+b_3-b_4 \\ 0 & 0 & 0 & 0 & 0 & -b_1+b_2+b_3-b_4 \end{array}\right)$$

Das entspechende Gleichungssystem ist genau dann lösbar, wenn

$$b_4 = -b_1 + b_2 + b_3 \ .$$

Unter dieser Annahme liegt bereits spezielle Zeilenstufenform vor, es ist also keine Permutation der Spalten notwendig. Wir formen die erweiterte Koeffizientenmatrix nochmals um, um links oben eine Diagonalmatrix zu erreichen:

$$\left(\begin{array}{rrr|rr||r} 1 & 3 & 5 & 2 & 0 & b_1 \\ 0 & 2 & 2 & -2 & 1 & -3b_1+b_2+b_3 \\ 0 & 0 & 5 & 5 & -2 & 3b_1-b_2 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{\mathrm{II}_{31},\mathrm{II}_{32}} \left(\begin{array}{rrr|rr||r} 1 & 3 & 0 & -3 & 2 & -2b_1+b_2 \\ 0 & 2 & 0 & -4 & \frac{9}{5} & -\frac{21}{5}b_1+\frac{7}{5}b_2+b_3 \\ 0 & 0 & 5 & 5 & -2 & 3b_1-b_2 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{\mathrm{II}_{21}} \left(\begin{array}{rrr|rr||r} 1 & 0 & 0 & 3 & -\frac{7}{10} & \frac{43}{10}b_1-\frac{11}{10}b_2-\frac{3}{2}b_3 \\ 0 & 2 & 0 & -4 & \frac{9}{5} & -\frac{21}{5}b_1+\frac{7}{5}b_2+b_3 \\ 0 & 0 & 5 & 5 & -2 & 3b_1-b_2 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{\mathrm{III}_2,\mathrm{III}_3} \left(\begin{array}{rrr|rr||r} 1 & 0 & 0 & 3 & -\frac{7}{10} & \frac{43}{10}b_1-\frac{11}{10}b_2-\frac{3}{2}b_3 \\ 0 & 1 & 0 & -2 & \frac{9}{10} & -\frac{21}{10}b_1+\frac{7}{10}b_2+\frac{1}{2}b_3 \\ 0 & 0 & 1 & 1 & -\frac{2}{5} & \frac{3}{5}b_1-\frac{1}{5}b_2 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Damit ist die Lösung für $b_4 = -b_1 + b_2 + b_3$ gegeben durch

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \left(\begin{array}{ccc} \frac{43}{10} & -\frac{11}{10} & -\frac{3}{2} \\ -\frac{21}{10} & \frac{7}{10} & \frac{1}{2} \\ \frac{3}{5} & -\frac{1}{5} & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right) \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + \left(\begin{array}{cc} -3 & \frac{7}{10} \\ 2 & -\frac{9}{10} \\ -1 & \frac{2}{5} \\ \hline 1 & 0 \\ 0 & 1 \end{array}\right) \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} .$$

Lineare Gleichungssysteme treten insbesondere auf

- bei der Untersuchung, ob eine gegebene Familie von Vektoren linear unabhängig ist
- bei der Zerlegung eines Vektors $v \in V$ nach einer Basis von V
- bei der Inversion von quadratischen Matrizen

2.9 Elementarmatrizen

Die elementaren Zeilenumformungen, die eine Matrix $B \in M(m \times n, K)$ (z.B. die erweiterte Koeffizientenmatrix (A, b), welche ein ein lineares Gleichungssystem beschreibt) in Zeilenstufenform bringen, lassen sich durch Linksmultiplikation von B mit geeigneten $(m \times m)$-Matrizen beschreiben. Diese *Elementarmatrizen* sind:

Typ I: $B \xrightarrow{\mathrm{I}_{ik}} \tilde{B}$ ist Matrixmultiplikation $\tilde{B} = P_{ik} \cdot B$ mit

$$P_{ik} = \left(\begin{array}{ccccccccccc} 1 & & & | & & & & | & & & \\ & \ddots & & & & & & & & & \\ & & 1 & | & & & & | & & & \\ - & - & - & 0 & - & - & - & 1 & - & - & - \\ & & & | & 1 & & & | & & & \\ & & & & & \ddots & & & & & \\ & & & | & & & 1 & | & & & \\ - & - & - & 1 & - & - & - & 0 & - & - & - \\ & & & | & & & & | & 1 & & \\ & & & & & & & & & \ddots & \\ & & & | & & & & | & & & 1 \end{array}\right)$$
$$= E_m - E_{ii} - E_{kk} + E_{ik} + E_{ki}$$

(Angedeutet sind die i-te und k-te Zeile und Spalte. Alle nicht bezeichneten Einträge sind 0.)

Typ II: $B \xrightarrow{\mathrm{II}_{ik}} \tilde{B}$ ist Matrixmultiplikation $\tilde{B} = Q_{ik}(\lambda) \cdot B$ mit

$$Q_{ik}(\lambda) = \begin{pmatrix} 1 & & & \vert & & & & \vert & & & \\ & \ddots & & & & & & & & & \\ & & 1 & \vert & & & & \vert & & & \\ - & - & - & 1 & - & - & - & 0 & - & - & - \\ & & & \vert & 1 & & & \vert & & & \\ & & & & & \ddots & & & & & \\ & & & \vert & & & 1 & \vert & & & \\ - & - & - & \lambda & - & - & - & 1 & - & - & - \\ & & & \vert & & & & \vert & 1 & & \\ & & & & & & & & & \ddots & \\ & & & \vert & & & & \vert & & & 1 \end{pmatrix} = E_m + \lambda E_{ki}$$

(Angedeutet sind die i-te und k-te Zeile und Spalte.)

Typ III: $B \xrightarrow{\mathrm{III}_{i}} \tilde{B}$ ist Matrixmultiplikation $\tilde{B} = S_i(\lambda) \cdot B$ mit

$$S_i(\lambda) = \begin{pmatrix} 1 & & & \vert & & & \\ & \ddots & & & & & \\ & & 1 & \vert & & & \\ - & - & - & \lambda & - & - & - \\ & & & \vert & 1 & & \\ & & & & & \ddots & \\ & & & \vert & & & 1 \end{pmatrix} = E_m + (\lambda - 1)E_{ii}$$

Tatsächlich ist Typ I eine Kombination aus mehreren Umformungen vom Typ II und III:

$$P_{ik} = Q_{ik}(1)\, Q_{ki}(-1) Q_{ik}(1) S_k(-1) \,,$$

denn das bedeutet

$$(i,k) \xrightarrow{\mathrm{III}_k} (i,-k) \xrightarrow{\mathrm{II}_{ik}} (i, i-k) \xrightarrow{\mathrm{II}_{ki}} (k, i-k) \xrightarrow{\mathrm{II}_{ik}} (k,i) \,.$$

Die Elementarmatrizen sind invertierbar mit

$$P_{ik}^{-1} = P_{ik} \,, \qquad (Q_{ik}(\lambda))^{-1} = Q_{ik}(-\lambda) \,, \qquad (S_i(\lambda))^{-1} = S_i(\tfrac{1}{\lambda}) \,.$$

Eine wichtige Anwendung der Elementarmatrizen ist der folgende Satz:

Satz 2.42 *Jede invertierbare Matrix $A \in GL(n,K)$ ist ein endliches Produkt von Elementarmatrizen.*

Beweis. Da $\operatorname{rang}(A) = n$, führen elementare Zeilenumformungen mit Typ I und II erst zu einer Dreiecksmatrix mit Diagonalelementen $d_{ii} \neq 0$ für alle $1 \leq i \leq n$, nach weiterer Umformung mit Typ II dann zu einer Diagonalmatrix mit gleichen Diagonalelementen $d_{ii} \neq 0$, und schließlich mit Typ III zur Einheitsmatrix. Damit gilt $B_r \cdots B_2 \cdot B_1 \cdot A = E_n$, wobei jedes B_s eine Elementarmatrix $Q_{ik}(\lambda)$ oder $S_i(\lambda)$ ist. Folglich gilt $A = B_1^{-1} \cdot B_2^{-1} \cdots B_r^{-1}$ und $A^{-1} = B_r \cdots B_2 \cdot B_1$. □

Eine nützliche Anwendung dieses Beweises besteht in einer effizienten Berechnungsmethode für die *inverse Matrix*. Durch Vergleich von $B_r \cdots B_2 \cdot B_1 \cdot A = E_n$ mit $A^{-1} = B_r \cdots B_2 \cdot B_1 \cdot E_n$ lesen wir ab: Dieselben Zeilenumformungen, welche A in E_n überführen, überführen E_n in A^{-1}. Das soll an einem Beispiel erläutert werden

$$\left(\begin{array}{cc||cc} 9 & 4 & 1 & 0 \\ 11 & 5 & 0 & 1 \end{array}\right) \xrightarrow{Q_{12}(-\frac{11}{9})} \left(\begin{array}{cc||cc} 9 & 4 & 1 & 0 \\ 0 & \frac{1}{9} & -\frac{11}{9} & 1 \end{array}\right) \xrightarrow{Q_{21}(-36)} \left(\begin{array}{cc||cc} 9 & 0 & 45 & -36 \\ 0 & \frac{1}{9} & -\frac{11}{9} & 1 \end{array}\right)$$
$$\xrightarrow{S_1(\frac{1}{9})S_2(9)} \left(\begin{array}{cc||cc} 1 & 0 & 5 & -4 \\ 0 & 1 & -11 & 9 \end{array}\right)$$

Folglich ist $\begin{pmatrix} 9 & 4 \\ 11 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 5 & -4 \\ -11 & 9 \end{pmatrix}$, was durch Ausmultiplizieren leicht zu überprüfen ist.

Ein anderer Weg, diese Rechnung zu verstehen, interpretiert $A \cdot A^{-1} = E_n$ als n lineare Gleichungssysteme $A \cdot x^{(p)} = b^{(p)}$. Dabei sind die $x^{(p)}$ bzw. $b^{(p)}$ mit $1 \leq p \leq n$ gerade die n Spalten von A^{-1} bzw. E_n. Anstatt nun die Berechnung von $x^{(p)}$ durch Zeilenumformung jeder dieser erweiterten Koeffizientenmatrizen $(A, b^{(p)})$ separat durchzuführen, schreiben wir alle Spalten $b^{(p)}$ nebeneinander und formen die entstehende Matrix (A, E_n) gleichzeitig um. (Es ist stets dieselbe Rechnung, die A in E_n überführt, unabhängig von b).

Man muß zunächst nicht wissen, ob A invertierbar ist, um dieses Verfahren durchzuführen. Wir können sogar nichtquadratische Matrizen $A \in M(m \times n, K)$ zulassen. Sei also $\operatorname{rang}(A) = r$, dann führen die Zeilenumformungen zu $(m - r)$ Zeilen, in denen sämtliche Einträge gleich 0 sind.

Wir erreichen also bestenfalls die Blockdarstellung

$$(A, E_m) \mapsto \left(\begin{array}{cc||cc} E_r & B & L_1 & L_2 \\ 0 & 0 & L_3 & L_4 \end{array}\right)$$

mit $B \in M(r \times (n-r), K)$. Es gibt also Elementarmatrizen $B_1, \ldots, B_r \in GL(m, K)$ mit $B_r \cdots B_1 \cdot A = \begin{pmatrix} E_r & B \\ 0 & 0 \end{pmatrix}$ und $B_r \cdots B_1 \cdot E_m = L^{-1} = \begin{pmatrix} L_1 & L_2 \\ L_3 & L_4 \end{pmatrix}$. Wir können den Block $B \in M(r \times (m-r), K)$ jedoch beseitigen, indem wir die Matrix *transponieren* zu $\begin{pmatrix} E_r & B \\ 0 & 0 \end{pmatrix}^t = \begin{pmatrix} E_r & 0 \\ B^t & 0 \end{pmatrix} \in M(n \times m, K)$

und dann mittels elementarer Zeilenumformungen in $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ überführen. Es gibt also Elementarmatrizen $C_1, \ldots, C_s \in GL(n, K)$ mit

$$C_s \cdots C_2 \cdot C_1 \cdot \begin{pmatrix} E_r & B \\ 0 & 0 \end{pmatrix}^t = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

und somit nach Rücktransposition

$$\begin{pmatrix} E_r & B \\ 0 & 0 \end{pmatrix} \cdot (C_s \cdots C_2 \cdot C_1)^t = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} .$$

Das Produkt dieser Elementarmatrizen $\tilde{R} = C_s \cdots C_2 \cdot C_1$ erhalten wir wieder durch die korrespondierende Zeilenumformung der Einheitsmatrix:

$$\left(\begin{array}{cc||cc} E_r & 0 & E_r & 0 \\ B^t & 0 & 0 & E_{n-r} \end{array}\right) \mapsto \left(\begin{array}{cc||cc} E_r & 0 & E_r & 0 \\ 0 & 0 & R_3 & E_{n-r} \end{array}\right) , \qquad \tilde{R} = \begin{pmatrix} E_r & 0 \\ R_3 & E_{n-r} \end{pmatrix} .$$

Damit haben wir folgenden Satz bewiesen:

Satz 2.43 *Zu jeder Matrix $A \in M(m \times n, K)$ mit $\operatorname{rang}(A) = r$ gibt es invertierbare Matrizen $L \in GL(m, K)$ und $R \in GL(n, K)$, so daß $A = L \cdot \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} \cdot R^{-1}$. (Dabei ist $R = \tilde{R}^t$.)*

Beispiele 2.10

$$\left(\begin{array}{ccc||cc} 1 & 2 & 3 & 1 & 0 \\ 4 & 5 & 6 & 0 & 1 \end{array}\right) \xrightarrow{Q_{12}(-4)} \left(\begin{array}{cc|c||cc} 1 & 2 & 3 & 1 & 0 \\ 0 & -3 & -6 & -4 & 1 \end{array}\right)$$
$$\xrightarrow{Q_{21}(\frac{2}{3})} \left(\begin{array}{cc|c||cc} 1 & 0 & -1 & -\frac{5}{3} & \frac{2}{3} \\ 0 & -3 & -6 & -4 & 1 \end{array}\right) \xrightarrow{S_2(-\frac{1}{3})} \left(\begin{array}{cc|c||cc} 1 & 0 & -1 & -\frac{5}{3} & \frac{2}{3} \\ 0 & 1 & 2 & \frac{4}{3} & -\frac{1}{3} \end{array}\right) .$$

Damit ist $L^{-1} = \begin{pmatrix} -\frac{5}{3} & \frac{2}{3} \\ \frac{4}{3} & -\frac{1}{3} \end{pmatrix}$. Die übliche Rechnung für das Inverse liefert $L = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$.

Nach Transposition berechnen wir

$$\left(\begin{array}{cc||ccc} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ -1 & 2 & 0 & 0 & 1 \end{array}\right) \xrightarrow{Q_{13}(1)\, Q_{23}(-2)} \left(\begin{array}{cc||ccc} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{array}\right) ,$$

also $\tilde{R}^t = R = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$. Diese Matrix ist nun leicht zu invertieren: $R^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$.

Folglich gilt

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

Man kann Satz 2.43 auch abstrakt verstehen: Die Matrix $A = (a_{ij}) \in M(m \times n, K)$ entspricht einer linearen Abbildung $F : K^n \to K^m$ mit $F(e_j) = \sum_{i=1}^m a_{ij} \cdot e_i$. Ist $\mathrm{rang}(F) = \mathrm{rang}(A) = r$, dann existieren nach Satz 2.28 Basen $\mathcal{A} = (v_1, \dots, v_n)$ von K^n und $\mathcal{B}(w_1, \dots, w_m)$ von K^m, so daß $M_{\mathcal{B}}^{\mathcal{A}}(F) = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} = \tilde{A} = (\tilde{a}_{kl})$. Die mit L und R bewirkte Umrechnung von A in $\tilde{A}$ ist ein Beispiel für eine *Koordinatentransformation*.

Koordinatentransformationen werden hervorgerufen durch einen Basiswechsel in einem Vektorraum. Sind in einem Vektorraum V zwei Basen $\mathcal{A} = (v_1, \dots, v_n)$ und $\mathcal{B} = (w_1, \dots, w_n)$ gegeben, so existieren nach Satz 2.25 Isomorphismen $\Phi_{\mathcal{A}} : K^n \to V$ und $\Phi_{\mathcal{B}} : K^n \to V$ mit $\Phi_{\mathcal{A}}(e_i) = v_i$ und $\Phi_{\mathcal{B}}(e_i) = w_i$. Die identitische Abbildung auf V induziert dann einen Isomorphismus $\Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}} : K^n \to K^n$. Die zugehörige Matrix $\Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}}(e_j) = \sum_{i=1}^n (T_{\mathcal{B}}^{\mathcal{A}})_{ij} \cdot e_i$ ist invertierbar, $T_{\mathcal{B}}^{\mathcal{A}} = ((T_{\mathcal{B}}^{\mathcal{A}})_{ij}) \in GL(n, K)$. Sie ist die Matrix $T_{\mathcal{B}}^{\mathcal{A}} = M_{\mathcal{B}}^{\mathcal{A}}(\mathrm{id}_V)$ einer allgemeinen Koordinatentransformation.

Die Matrixmultiplikation war definiert als die Komposition linearer Abbildungen $F : U \to W$ und $G : V \to U$ bezüglich ihrer Basen,

$$M_{\mathcal{C}}^{\mathcal{A}}(F \circ G) = M_{\mathcal{C}}^{\mathcal{B}}(F) \cdot M_{\mathcal{B}}^{\mathcal{A}}(G) .$$

mit $\Phi_{\mathcal{B}}^{-1} \circ G \circ \Phi_{\mathcal{A}}(e_j) = \sum_i (M_{\mathcal{B}}^{\mathcal{A}}(G))_{ij} \cdot e_j$ Entsprechend ist für $F : V \to W$

$$M_{\mathcal{D}}^{\mathcal{C}}(F) = M_{\mathcal{D}}^{\mathcal{C}}(\mathrm{id}_W \circ F \circ \mathrm{id}_V) = M_{\mathcal{D}}^{\mathcal{B}}(\mathrm{id}_W) \cdot M_{\mathcal{B}}^{\mathcal{A}}(F) \cdot M_{\mathcal{A}}^{\mathcal{C}}(\mathrm{id}_V)$$

Sind nun $V = K^n$ und $W = K^m$ sowie $\mathcal{C}$ und $\mathcal{D}$ die Standardbasen von K^n bzw. K^m, so haben wir $\Phi_{\mathcal{C}} = \mathrm{id}_{K^n}$ und $\Phi_{\mathcal{D}} = \mathrm{id}_{K^m}$ und

$$M_{\mathcal{A}}^{\mathcal{C}}(\mathrm{id}_{K^n}) = R^{-1} , \qquad M_{\mathcal{D}}^{\mathcal{B}}(\mathrm{id}_{K^m}) = L$$

mit

$$\Phi_{\mathcal{A}}(e_j) = v_j = \sum_{l=1}^n R_{lj} e_l , \qquad \Phi_{\mathcal{B}}(e_i) = w_i = \sum_{k=1}^n L_{ki} e_k .$$

Mit $M_{\mathcal{D}}^{\mathcal{C}}(F) = A$ und $M_{\mathcal{B}}^{\mathcal{A}}(F) = \tilde{A}$ folgt $A = L \cdot \tilde{A} \cdot R^{-1}$.

3 Eigenwerte und Determinanten

3.1 Determinanten

Bei der Determinante hantelt es sich um eine wichtige Charakterisierung quadratischer Matrizen. Die Determinante ist ein Kriterium für die Invertierbarkeit einer Matrix. Sie tritt außerdem auf beim Eigenwertproblem für Matrizen.

Definition 3.1 Eine Abbildung

$$\det : M(n,K) \to K\ , \qquad \det : A \mapsto \det A$$

heißt *Determinante*, wenn folgendes gilt:

(D1) $\det$ ist linear in jeder Zeile, d.h. ist die i-te Zeile $a_i = \lambda' a_i' + \lambda'' a_i'' \in K^n$, so gilt

$$\det \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix} = \lambda' \det \begin{pmatrix} \vdots \\ a_i' \\ \vdots \end{pmatrix} + \lambda'' \det \begin{pmatrix} \vdots \\ a_i'' \\ \vdots \end{pmatrix} .$$

(Die mit $\vdots$ symbolisierten Zeilen $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$ sind in jeder der drei Matrizen identisch.)

Insbesondere gilt: Entsteht B aus A durch Multiplikation der i-ten Zeile mit $\lambda \neq 0$, ist also eine Zeilenumformung vom Typ III mit $B = S_i(\lambda) \cdot A$, so ist $\det B = \lambda \cdot \det A$.

(D2) $\det$ ist alternierend, d.h. hat A zwei gleiche Zeilen, so gilt $\det A = 0$.

(D3) $\det$ ist normiert auf $\det E_n = 1$.

Aus der Definition ergeben sich weitere Eigenschaften, die insbesondere eine Berechnungsmethode beinhalten:

Satz 3.1 *Die Determinante hat folgende weitere Eigenschaften:*

(D4) $\det(\lambda \cdot A) = \lambda^n \det A \quad \forall \lambda \in K$.

(D5) Ist eine Zeile von A identisch Null, so folgt $\det A = 0$.

(D6) Die Determinante ändert das Vorzeichen bei Zeilenumformungen von Typ I: Entsteht B aus A durch Zeilenvertauschung, also $B = P_{ik} \cdot A$, so gilt $\det B = -\det A$.

(D7) Zeilenumformungen von Typ II lassen die Determinante unverändert: Entsteht B aus A durch Addition der λ-fachen i-ten Zeile zur k-ten Zeile, also $B = Q_{ik}(\lambda) \cdot A$, so gilt $\det B = \det A$.

(D8) Ist $A = \begin{pmatrix} a_{11} & . & \dots & . \\ 0 & a_{22} & \dots & . \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix}$ *eine obere Dreiecksmatrix, so gilt* $\det A = a_{11}a_{22}\cdots a_{nn}$.

(D9) Sei $n \geq 2$ *und* $A \in M(n, K)$ *habe die folgende Blockdarstellung:*

$$A = \begin{pmatrix} A_1 & C \\ 0 & A_2 \end{pmatrix}, \qquad A_1 \in M(n_1, K)\ ,\ A_2 \in M(n_2, K)\ ,\ n_1 + n_2 = n\ .$$

Dann gilt $\det A = \det A_1 \cdot \det A_2$.

(D10) $\det A = 0 \quad \Leftrightarrow \quad \mathrm{rang}(A) < n$.

(D11) Es gilt der Determinantenmultiplikationssatz $\det(A \cdot B) = \det A \cdot \det B$ *für alle* $A, B \in M(n, K)$. *Insbesondere ist* $\det A^{-1} = \frac{1}{\det A}$ *für* $A \in GL(n, K)$. *Anders formuliert:* $\det : GL(n, K) \to K^*$ *ist ein Gruppenhomomorphismus.*

Beweis. D4) Nach (D1) gilt

$$\det \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix} = \lambda \det \begin{pmatrix} a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix} = \lambda^2 \det \begin{pmatrix} a_1 \\ a_2 \\ \lambda a_3 \\ \vdots \\ \lambda a_n \end{pmatrix} = \cdots = \lambda^n \det \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

D5) folgt aus (D1) mit $\lambda = 0$

D6) Ist $B = P_{ik}A$, so berechnen wir

$$\det A + \det B = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_k \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_i \\ \vdots \end{pmatrix}$$

$$\stackrel{(D2)}{=} \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_k \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_k \\ \vdots \end{pmatrix}$$

$$\overset{(D1)}{=} \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i + a_k \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_k \\ \vdots \\ a_i + a_k \\ \vdots \end{pmatrix} \overset{(D1)}{=} \det \begin{pmatrix} \vdots \\ a_i + a_k \\ \vdots \\ a_i + a_k \\ \vdots \end{pmatrix} \overset{(D2)}{=} 0$$

(D7) Ist $B = Q_{ik}(\lambda)A$, so gilt

$$\det B = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_k + \lambda a_i \\ \vdots \end{pmatrix} \overset{(D1)}{=} \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_k \\ \vdots \end{pmatrix} + \det \lambda \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i \\ \vdots \end{pmatrix} \overset{(D2)}{=} \det A$$

(D8) Für Diagonalmatrizen folgt durch wiederholte Anwendung von (D1)

$$\det \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} \overset{(D1)}{=} a_{11} \det \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & a_{nn} \end{pmatrix} \overset{(D1)}{=} \dots$$
$$\overset{(D1)}{=} a_{11}a_{22} \cdots a_{nn} \det E_n \overset{(D3)}{=} a_{11}a_{22} \cdots a_{nn}$$

Sind für eine allgemeine obere Dreiecksmatrix alle $a_{ii} \neq 0$, so bringen wir sie durch Zeilenumformung vom Typ II in Diagonalform mit den gleichen Diagonalelementen. Ist $a_{ii} = 0$ und $a_{kk} \neq 0$ für $k > i$, so führt die Zeilenumformung vom Typ II auf die Zeile $a_i = 0$.

(D9) Wir können Zeilenumformungen verwenden, die die beiden Blöcke nicht mischen. Durch Umformungen nur der ersten n_1 Zeilen und dann nur der letzten n_2 Zeilen erreicht man

$$\det A = (-1)^{s_1+s_2} \det \begin{pmatrix} D_1 & C' \\ 0 & D_2 \end{pmatrix} ,$$

wobei D_1, D_2 obere Dreiecksmatrizen sind und s_1, s_2 die Anzahl der Zeilenumformungen vom Typ I im oberen bzw unteren Block sind. Die Behauptung folgt nun aus (D8).

(D10) Durch elementare Zeilenumformung überführen wir A in eine obere Dreiecksmatrix mit Diagonalelementen $a_{11}, \dots, a_{nn}$. Dann ist $\det A = \pm a_{11} \cdots a_{nn}$. Aus $\det A = 0$ und der Zeilenstufenform folgt $a_{nn} = 0$ and damit $\det A = 0$ nach (D5), also $\operatorname{rang}(A) < n$. Ist umgekehrt $\operatorname{rang}(A) < n$, so führt elementare Zeilenumformung auf eine obere Dreiecksmatrix mit $a_{nn} = 0$, so daß $\det A = 0$.

(D11) Ist $\operatorname{rang}(A) < n$, so ist auch $\operatorname{rang}(AB) < n$ als Dimension des Bildes der Hintereinanderausführung linearer Abbildungen. Dann ist $\det(A) = \det(AB) = 0$. Ist $\operatorname{rang}(A) = n$, so ist A invertierbar und nach Satz 2.42 ein endliches Produkt von Elementarmatrizen. Also ist $\det(AB) = \det(C_1 \cdots C_r \cdot B)$, wobei $C_p = Q_{ik}(\lambda)$ oder $C_p = S_i(\lambda)$ für $1 \leq p \leq r$. Es gilt $\det(Q_{ik}(\lambda)B) = \det B$ nach (D7) und $\det(S_i(\lambda)B) = \lambda \det B$ nach (D1). Also ist $\det(AB) = \lambda_1 \cdots \lambda_s \det B$, wobei λ_p die Koeffizienten in den Umformungen vom Typ III sind. Andererseits ist $\det(A) = \det(AE_n) = \lambda_1 \cdots \lambda_s$, was die Behauptung liefert. □

Zusammengefaßt haben wir damit aus den Axiomen ein Berechnungsverfahren für die Determinante einer Matrix abgeleitet: Wir überführen mit elementaren Zeilenumformungen vom Typ I und II die Matrix $A \in M(n, K)$ in Zeilenstufenform $\tilde{A} = (a_{ij})$: Sind dabei s Vertauschungen von Zeilen (Typ I) notwendig, dann ist $\det A = (-1)^s a_{11} \cdots a_{nn}$. Das Verfahren beweist die *Eindeutigkeit* der Determinante: Gäbe es zwei Abbildungen det und $\widetilde{\det}$, die (D1), (D2) und (D3) erfüllen, so folgt aus beiden die Berechnungsvorschrift für die durch Zeilenumformung erhaltene Dreiecksmatrix $\tilde{A} = (a_{ij})$ mit $\det A = \widetilde{\det} A = (-1)^s a_{11} \cdots a_{nn}$.

In Beweisen und für speziell gewählte Matrizen sind auch rekursive und abstrakte Berechnungsformeln nützlich:

Satz 3.2 (Zeilenentwicklungssatz von Laplace) *Ist $A = (a_{ij}) \in M(n, K)$ und seien die Matrizen $A_{ij} \in M(n-1, K)$ aus A durch Streichen der i-ten Zeile und der j-ten Spalte erhalten. Dann gilt für beliebiges $1 \leq i \leq n$*

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} \ .$$

Beweis. Wir schreiben die i-te Zeile $a_i = (a_{i1}, a_{i2}, \ldots, a_{in})$ von $A = (a_{ij})$ als Linearkombination der kanonischen Basen

$$a_i = a_{i1}(1, 0, 0, \ldots, 0) + a_{i2}(0, 1, 0, \ldots, 0) + \cdots + a_{in}(0, 0, \ldots, 0, 1)$$

Anwenden von (D1) ergibt

$$\det A = \sum_{j=1}^{n} a_{ij} \det A'_{ij} \ , \quad A'_{ij} = \begin{pmatrix} a_{11} & \ldots & a_{1,j-1} & a_{1,j} & a_{1,j+1} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \ldots & a_{i-1,j-1} & a_{i-1,j} & a_{i-1,j+1} & \ldots & a_{i-1,n} \\ 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ a_{i+1,1} & \ldots & a_{i+1,j-1} & a_{i+1,j} & a_{i+1,j+1} & \ldots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{n,j-1} & a_{nj} & a_{n,j+1} & \ldots & a_{nn} \end{pmatrix} .$$

In jeder der Matrizen A'_{ij} bringen wir mit Zeilenumformungen vom Typ II (welche die Determinante nicht ändern) alle anderen Einträge der j-ten Spalte auf 0:

$$\det A'_{ij} = \det A''_{ij}\ , \quad A''_{ij} = \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & 0 & a_{1,j+1} & \dots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \dots & a_{i-1,n} \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ a_{i+1,1} & \dots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{n,j-1} & 0 & a_{n,j+1} & \dots & a_{nn} \end{pmatrix} .$$

Durch Vertauschen mit jeweils benachbarten Zeilen bringen wir die i-te Zeile von A''_{ij} an die j-te Stelle, wobei sich die Reihenfolge aller anderen Zeilen nicht ändert. Dazu sind $|i-j|$ Zeilenvertauschungen erforderlich. Die 1 steht nun auf der Diagonale: Für $i < j$ erhalten wir:

$$\det A''_{ij} = (-1)^{|i-j|} \det A'''_{ij}\ , \quad A'''_{ij} = \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & 0 & a_{1,j+1} & \dots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \dots & a_{i-1,n} \\ a_{i+1,1} & \dots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{j,1} & \dots & a_{j,j-1} & 0 & a_{j,j+1} & \dots & a_{j,n} \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ a_{j+1,1} & \dots & a_{j+1,j-1} & 0 & a_{j+1,j+1} & \dots & a_{j+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{n,j-1} & 0 & a_{n,j+1} & \dots & a_{nn} \end{pmatrix} .$$

Durch elementare Zeilenumformung vom Typ I und II überführen wir A'''_{ij} in eine obere Dreiecksmatrix und lesen die Determinante ab. Bei diesen Umformungen wird an der j-ten Zeile und Spalte von A'''_{ij} nichts geändert. Deshalb stimmt $\det A'''_{ij}$ mit der Determinante jener Matrix $A_{ij} \in M(n-1, K)$ überein, die aus A'''_{ij} durch Weglassen des Kreuzes aus j-ter Zeile und j-ter Spalte erhalten wird. Dieselbe Matrix A_{ij} entsteht aber auch aus A'_{ij} und damit auch aus A selbst durch Weglassen der i-ten Zeile und der j-ten Spalte. Unter Verwendung von $(-1)^{|i-j|} = (-1)^{i+j}$ folgt die Behauptung. □

Die Berechnung der Determinante von $A \in M(n, K)$ wird damit auf die Berechnung der Determinanten kleinerer Matrizen zurückgeführt. Beginnend mit $n = 1$ beweist man so die Existenz der Determinante $\det : M(n, K) \to K$.

Als Beispiel berechnen wir

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = (-1)^{1+1} a_{11} \det(a_{22}) + (-1)^{1+2} a_{12} \det(a_{21}) = a_{11}a_{22} - a_{12}a_{21}\ .$$

Einfache Rechenregel: Für 2×2-Matrizen ist die Determinante gleich dem Produkt der Hauptdiagonalelemente minus dem Produkt der Nebendiagonalelemente. Eine ähnliche graphische Rechenregel gibt es auch für 3×3-Matrizen (Regel von Sarrus).

Allgemein wird durch den Entwicklungssatz die Determinante von $A = (a_{ij})$ rekursiv durch Summen und Differenzen von Produkten $a_{i_1 j_1} \cdots a_{i_n j_n}$ der Einträge a_{ij} berechnet. Durch Entwicklung nach der jeweils obersten Zeile sieht man, daß in jedem dieser Produkte jeder Zeilenindex genau einmal vorkommt. Andererseits kommt auch jeder Spaltenindex genau einmal vor: Damit a_{ij} in einem Produkt vorkommen kann, darf zuvor kein a_{kj} aufgetreten sein, da sonst die j-te Spalte gestrichen wäre. Nach dem Auftreten von a_{ij} wird die j-te Spalte weggelassen und weitere a_{kj} können im Produkt nicht vorkommen. Eine solche injektive/surjektive/bijektive Zuordnung eines Spaltenindex $j = \sigma(i) \in \{1, \ldots, n\}$ zu jedem Zeilenindex $i \in \{1, \ldots, n\}$ heißt *Permutation*. Ein Beispiel ist durch folgende Tabelle gegeben:

i	1	2	3	4	5
$\sigma(i)$	2	5	3	1	4

Sei S_n die Menge (und Gruppe) aller Permutationen, dann gilt

$$\det A = \sum_{\sigma \in S_n} \lambda(\sigma)\; a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \;.$$

Dabei ist $\lambda(\sigma) \in \pm 1$, denn jede Permutation kommt vor und zwar mit Vorfaktor $+1$ oder -1 entsprechend dem Entwicklungssatz. Man kann zeigen, daß $\lambda(\sigma) = \operatorname{sign}(\sigma)$ das *Vorzeichen der Permutation* ist, das wie folgt berechnet werden kann: Eine Permutation τ heißt *Transposition*, wenn sie zwei Elemente aus $\{1, \ldots, n\}$ austauscht und alle anderen an ihrer Stelle beläßt. Jede Permutation kann als (nicht eindeutige) Hintereinanderausführung von Transpositionen erhalten werden. Dann ist

$$\operatorname{sign}(\sigma) = (-1)^{\text{Zahl der Transpositionen in } \sigma} \;.$$

Das Vorzeichen ist eindeutig, auch wenn die Zahl der Transpositionen selbst nicht eindeutig ist.

Alternativ kann man das Vorzeichen aus der Zahl der *Fehlstellen* der Permutation ablesen. Eine Fehlstelle ist ein Paar (i, j) mit $i < j$ und $\sigma(i) > \sigma(j)$. Dann ist

$$\operatorname{sign}(\sigma) = (-1)^{\text{Zahl der Fehlstellen in } \sigma} \;.$$

Damit gilt

Satz 3.3 (*Formel von Leibniz*)

$$\det A = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma)\; a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \;.$$

Es gibt $n!$ Permutationen $\sigma \in S_n$, so daß die Zahl der Produkte in der Formel von Leibniz mit wachsendem n sehr groß wird. Deshalb ist die Formel von Leibniz vor allem aus theoretischer Sicht bedeutsam; für praktische Berechnungen sind Zeilenumformungen geeigneter.

Wir untersuchen nun die Determinante der transponierten Matrix.

Satz 3.4 *Die Determinante ist linear in jeder Spalte.*

Beweis. Wir beweisen die Aussage mit dem Entwicklungssatz von Laplace und Induktion nach n. Die Aussage ist offenbar richtig für $n = 1$, da $\det(a) = a$ für $a \in M(1, K) = K$. Sei nun $A = (a_{ij}) \in M(n, K)$ gegeben mit $a_{il} = \lambda' a'_{il} + \lambda'' a''_{il}$ für einen Spaltenindex l und λ', λ'' unabhängig von i. Wir bezeichnen mit A' bzw. A'' die Matrizen, die durch Ersetzen von a_{kl} durch a'_{kl} bzw. a''_{kl} in A für $1 \leq k \leq n$ erhalten werden. Alle anderen Spalten $j \neq k$ von A, A', A'' sind identisch.

Wir entwickeln nach der i-ten Zeile:

$$\det A = \sum_{j=1}^{l-1} a_{ij} \det A_{ij} + a_{il} \det A_{il} + \sum_{j=l+1}^{n} a_{ij} \det A_{ij} \ .$$

Die Matrizen $A_{ij} \in M(n-1, K)$ für $j \neq l$ enthalten die ursprünglich l-te Spalte, so daß nach Induktion gilt

$$\det A_{ij} = \lambda' \det A'_{ij} + \lambda'' \det A''_{ij} \ , \quad j \neq l \ .$$

Da in A_{il} die modifizierte l-te Spalte weggelassen wird, gilt $A_{il} = A'_{il} = A''_{il}$. Mit $a_{il} = \lambda' a'_{il} + \lambda'' a''_{il}$ folgt nun die Behauptung $\det A = \lambda' \det A' + \lambda'' \det A''$. □

Satz 3.5 *Es gilt* $\det(A^t) = \det(A)$ *für alle* $A \in M(n, K)$.

Beweis. Wir definieren eine Abbildung $\widetilde{\det} : M(n, K) \to K$ durch $\widetilde{\det}(A) := \det(A^t)$ und beweisen, daß $\widetilde{\det}$ eine Determinante ist. Aus der Eindeutigkeit der Determinanten folgt dann die Behauptung. Es sind die Axiome zu überprüfen.

(D1) $\widetilde{\det}$ ist linear in jeder Zeile $\Leftrightarrow$ det ist linear in jeder Spalte, was nach Satz 3.4 zutrifft.

(D2) Habe A zwei gleiche Zeilen, dann hat A^t zwei gleiche Spalten und $\mathrm{rang}(A) = \mathrm{rang}(A^t) < n$. Dann ist $\widetilde{\det}(A) = \det(A^t) = 0$.

(D3) $\widetilde{\det}(E_n) = \det(E_n^t) = \det(E_n) = 1$. □

Damit erhalten wir sofort den

Satz 3.6 (Spaltenentwicklungssatz von Laplace) *Sei* $A = (a_{ij}) \in M(n, K)$ *und* $A_{ij} \in M(n-1, K)$ *durch Weglassen der* i*-ten Zeile und* j*-ten Spalte von* A *erhalten. Dann gilt für beliebiges* $1 \leq j \leq n$

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} \ .$$

Beweis. Wir verwenden $\det A = \det A^t$ und berechnen $\det A^t$ nach dem Zeilenentwicklungssatz von Laplace. Der Zeilenentwicklung von A^t entspricht aber genau die Spaltenentwicklung von A. $\square$

3.2 Einige Anwendungen

Wir ordnen einer Matrix $A = (a_{ij}) \in M(n, K)$ die folgende *komplementäre Matrix* $A^\sharp \in M(n, K)$ zu:

$$A^\sharp = (a^\sharp_{ij}) , \qquad a^\sharp_{ij} := (-1)^{i+j} \det A_{ji} .$$

Man beachte die Vertauschung der Reihenfolge von i, j: Die Matrix A_{ji} entsteht aus A durch Weglassen der j-ten Zeile und der i-ten Spalte.

Satz 3.7 *Ist $A \in M(n, K)$ und sei $A^\sharp$ die zu A komplementäre Matrix, dann gilt $A \cdot A^\sharp = A^\sharp \cdot A = (\det A)E_n$. Insbesondere gilt $A^{-1} = \frac{1}{\det A} A^\sharp$ für invertierbare Matrizen $A \in GL(n, K)$.*

Beweis. Wir berechnen die Komponenten von $A^\sharp \cdot A$:

$$(A^\sharp \cdot A)_{kj} = \sum_{i=1}^{n} a^\sharp_{ki} a_{ij} = \sum_{i=1}^{n} (-1)^{i+k} a_{ij} \det A_{ik} .$$

Ist $j = k$ beliebig, so entsteht gerade der Spaltenentwicklungssatz von Laplace: $\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}$. Für $j \neq k$ betrachten wir die Matrix $B = (b_{il}) \in M(n, K)$, die aus A entsteht, wenn man die k-te Spalte von A durch die j-te Spalte von A ersetzt. Es ist also $b_{il} = a_{il}$ für $l \neq k$ und $b_{ik} = a_{ij}$. Da B zwei gleiche Spalten besitzt, ist $\det B = 0$. Wir entwickeln $\det B$ nach der k-ten Spalte:

$$0 = \det B = \sum_{i=1}^{n} (-1)^{i+k} b_{ik} \det B_{ik} = \sum_{i=1}^{n} (-1)^{i+k} a_{ij} \det A_{ik}$$

wegen $A_{ik} = B_{ik}$. Folglich ist $(A^\sharp \cdot A)_{kj} = (\det A)\delta_{kj}$. $\square$

Mit $A^{-1} = \frac{1}{\det A} A^\sharp$ haben wir eine weitere Methode zur Berechnung der inversen Matrix kennengelernt. Für $n > 3$ ist diese Methode jedoch sehr aufwendig. Allerdings ist diese abstrakte Darstellung in der Analysis sehr nützlich, denn die Determinantenbildung hängt als Produkt der a_{ij} stetig und sogar differenzierbar von den Einträgen a_{ij} ab. Daraus folgt, daß für $A \in GL(n, K)$ die Abbildung $A \mapsto A^{-1}$ differenzierbar ist.

Es sei auch bemerkt, daß

$$SL(n, K) := \{A \in GL(n, K) \ : \ \det A = 1\}$$

eine Untergruppe von $GL(n,K)$ ist, die *spezielle lineare Gruppe*. Denn für $A, B \in SL(n,K)$ folgt aus den Eigenschaften der Determinante $\det(AB) = 1$ und $\det A^{-1} = 1$.

Eine weitere Anwendung der Determinanten besteht in einem Lösungsverfahren für lineare Gleichungssysteme mit invertierbaren Matrizen:

Satz 3.8 (Cramersche Regel) *Seien $A = (a_{ij}) \in GL(n,K)$ und $b \in K^n$ gegeben und sei $x = (x_1, \ldots, x_n)^t \in K^n$ die eindeutig bestimmte Lösung des linearen Gleichungssystems $A \cdot x = b$. Bezeichnen wir mit $a_1, \ldots, a_n$ die Spalten von A, also $a_j = (a_{1j}, \ldots, a_{nj})^t$, dann gilt*

$$x_j = \frac{\det(a_1, \ldots, a_{j-1}, b_j, a_{j+1}, \ldots, a_n)}{\det A} .$$

Beweis. Die endeutig bestimmte Lösung des linearen Gleichungssystems ist durch $x = A^{-1} \cdot b$ gegeben. In Komponenten gilt damit

$$x_j = \sum_{i=1}^{n} (A^{-1})_{ji} b_i = \frac{1}{\det A} \sum_{i=1}^{n} (-1)^{i+j} b_i (\det A_{ij}) .$$

Nach dem Spaltenentwicklungssatz von Laplace ist $\sum_{i=1}^{n} (-1)^{i+j} b_i (\det A_{ij})$ gerade die Determinante einer Matrix $B = (b_{il})$, deren Einträge auf der j-ten Spalte durch $b_{ij} = b_i$ gegeben sind und deren andere Spalten identisch sind mit den Spalten von A. □

Wieder liegt die Bedeutung der Cramerschen Regel in theoretischen Betrachtungen wie der Schlußfolgerung, daß die Lösung des linearen Gleichungssystems $Ax = b$ für $A \in GL(n,K)$ differenzierbar von der rechten Seite $b \in K^n$ sowie den Einträgen der matrix $A \in GL(n,K)$ abhängt.

Die Determinantenbildung für Matrizen läßt sich sofort auf die Determinate für Endomorphismen endlich-dimensionaler Vektorräume verallgemeinern. Sei V ein n-dimensionaler Vektorraum über K und $F : V \to V$ eine lineare Abbildung (ein Endomorphismus). Wir wählen eine beliebige Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V und stellen F in dieser Basis dar:

$$F(v_j) = \sum_{i=1}^{n} a_{ij} \cdot v_i , \qquad a_{ij} \in K .$$

Dann ist $A = (a_{ij}) = M_{\mathcal{B}}^{\mathcal{B}}(F)$ die darstellende Matrix von F bezüglich der Basis $\mathcal{B}$. Wir definieren $\det F := \det(M_{\mathcal{B}}^{\mathcal{B}}(F))$. Diese Definition ist sinnvoll, da sie nicht von der Wahl der Basis abhängt: Sei $\mathcal{A}$ eine andere Basis von V, dann gilt

$$M_{\mathcal{A}}^{\mathcal{A}}(F) = M_{\mathcal{A}}^{\mathcal{A}}(\mathrm{id}_V \circ F \circ \mathrm{id}_V) = M_{\mathcal{A}}^{\mathcal{B}}(\mathrm{id}_V) \cdot M_{\mathcal{B}}^{\mathcal{B}}(F) \cdot M_{\mathcal{B}}^{\mathcal{A}}(\mathrm{id}_V) .$$

Offenbar ist $M^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V) = (M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V))^{-1}$, so daß

$$\det(M^{\mathcal{A}}_{\mathcal{A}}(F)) = \det(M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V))\det(M^{\mathcal{B}}_{\mathcal{B}}(F)) \cdot \frac{1}{\det(M^{\mathcal{B}}_{\mathcal{A}}(\mathrm{id}_V))} = \det(M^{\mathcal{B}}_{\mathcal{B}}(F)) \,.$$

In direkter Verallgemeinerung der Determinanteneigenschaften gilt:

Satz 3.9 *Sind $F, G \in \mathrm{End}(V)$, so gilt*

i) $\det F \neq 0 \quad \Leftrightarrow \quad F \in \mathrm{Aut}(V)$

ii) $F \in \mathrm{Aut}(V) \quad \Rightarrow \quad \det F^{-1} = \frac{1}{\det F}$

iii) $\det(F \circ G) = \det F \cdot \det G$

Eine weitere nützliche Anwendung ist die Definition der *Orientierung* von Automorphismen und von Basen.

Definition 3.2 Sei V ein endlich-dimensionaler Vektorraum über K. Ein Automorphismus $F \in \mathrm{Aut}(V)$ heißt *orientierungstreu*, falls $\det F > 0$, ansonsten *orientierngsuntreu*.

Zwei Basen $\mathcal{A}$ und $\mathcal{B}$ von V heißen *gleich orientiert*, wenn für die darstellende Matrix der Identität gilt $\det(M^{\mathcal{A}}_{\mathcal{B}}(\mathrm{id}_V)) > 0$, ansonsten *ungleich orientiert*.

Offenbar gilt:

Satz 3.10 *Die Menge*

$$\mathrm{Aut}^+(V) := \{F \in \mathrm{Aut}(V) \;:\quad \det F > 0\}$$

der orientierungstreuen Automorphismen von V ist eine Untergruppe von $\mathrm{Aut}(V)$. *Insbesondere ist*

$$GL^+(m, K) := \{A \in GL(n, K) \;:\quad \det A > 0\}$$

eine Untergruppe von $GL(n, K)$.

3.3 Eigenwerte

Wir hatten im Abschnitt über Matrizen (Satz 2.28) gesehen, daß zu jeder linearen Abbildung $F : V \to W$ angepaßte Basen $\mathcal{A}$ von V und $\mathcal{B}$ von W existieren, so daß die darstellende Matrix die Form $M^{\mathcal{A}}_{\mathcal{B}}(F) = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ hat. Natürlich können wir $W = V$ wählen und dann entsprechende Basen $\mathcal{A}, \mathcal{B}$ von V finden. Die Frage ist: Gibt es zu einem Endomorphismus $F : V \to V$ auch *eine* Basis $\mathcal{B}$ von V, so daß für die darstellende Matrix gilt: $M^{\mathcal{B}}_{\mathcal{B}}(F) = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$? Die Antwort ist: Nein!

Es geht nun darum, zu gegebenen Endomorphismus die Basis so sinnvoll wie möglich zu wählen.

Definition 3.3 Sei $F : V \to V$ ein Endomorphismus eines Vektorraumes V über K. Ein Skalar $\lambda \in K$ heißt *Eigenwert* von F, wenn es einen Vektor $v \neq 0$ von V gibt, so daß $F(v) = \lambda \cdot v$. Jeder Vektor $v \neq 0$ von V mit $F(v) = \lambda \cdot v$ heißt *Eigenvektor* von F zum Eigenwert λ.

Zu beachten ist, daß es zu einem Eigenwert λ mehrere Eigenvektoren geben kann.

Definition 3.4 Ein Endomorphismus $F \in \mathrm{End}(V)$ heißt *diagonalisierbar*, wenn es eine Basis von V aus Eigenvektoren von F gibt.

In diesem Fall gilt:

Satz 3.11 *Ist* $\dim(V) = n$, *so ist* $F \in \mathrm{End}(V)$ *genau dann diagonalisierbar, wenn es eine Basis* $\mathcal{B} = (v_1, \ldots, v_n)$ *von* V *gibt, so daß* $M_{\mathcal{B}}^{\mathcal{B}}(F) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$.

Beweis. Für die darstellende Matrix $M_{\mathcal{B}}^{\mathcal{B}}(F) = (a_{ij})$ gilt $F(v_j) = \sum_{i=1}^n a_{ij} \cdot v_i$. Damit sind die v_i die Eigenvektoren zu den Eigenwerten λ_i. $\square$

Das ist die optimalste Situation für eine Basis zu gegebenem Endomorphismus. Jedoch ist nicht klar, daß jeder Endomorphismus auch diagonalisierbar ist. Zunächst untersuchen wir, ob die Eigenvektoren linear unabhängig sind:

Satz 3.12 *Sei* $F \in \mathrm{End}(V)$ *ein Endomorphismus und seien* $v_1, \ldots, v_m$ *Eigenvektoren zu paarweise verschiedenen Eigenwerten* $\lambda_1, \ldots, \lambda_m$ *von* V. *Dann sind die* $(v_1, \ldots, v_m)$ *linear unabhängig.*

Beweis. Wir beweisen den Satz durch Induktion nach der Anzahl m linear unabhängiger Eigenvektoren. Für $m = 1$ ist nichts zu zeigen. Angenommen, $v_2, \ldots, v_m$ seien linear unabhängig. Wir betrachten

$$\mu_1 v_1 + \cdots + \mu_m v_m = 0$$

Anwenden von F einerseits und Subtraktion des λ_1-fachen dieser Gleichung ergibt

$$0 = \mu_1(\lambda_1 - \lambda_1)v_1 + \mu_2(\lambda_2 - \lambda_1)v_2 + \cdots + \mu_m(\lambda_m - \lambda_1)v_2 \; .$$

Da $\lambda_1 \neq \lambda_i$ für $2 \leq i \leq m$ und $(v_2, \ldots, v_m)$ linear unabhängig, folgt $\mu_j = 0$ für alle $1 \leq j \leq m$. Damit ist $(v_1, \ldots, v_m)$ linear unabhängig. $\square$

Als direkte Konsequenz ergibt sich:

Satz 3.13 *Seien* $\lambda_1, \ldots, \lambda_m$ *paarweise verschiedene Eigenwerte eines Endomorphismus* $F \in \mathrm{End}(V)$ *und sei* $\dim(V) = n$. *Dann gilt:*

i) $m \leq n$

ii) *Ist* $m = n$*, dann ist* F *diagonalisierbar.*

Beweis. i) Für $m > n$ wären nach dem vorigen Satz mehr als n Vektoren aus V, nämlich Eigenvektoren zu $\lambda_1, \ldots, \lambda_m$, linear unabhängig. Das ist durch die Dimension ausgeschlossen.

ii) Ist $\dim(V) = n$, dann bilden n linear unabhängige Vektoren eine Basis. Damit ist $(v_1, \ldots, v_n)$ eine Basis aus Eigenvektoren, und F ist diagonalisierbar. $\square$

Es gibt natürlich Endomorphismen mit weniger als $\dim(V)$ paarweise verschiedenen Eigenwerten, die trotzdem diagonalisierbar sind. Ein Beispiel ist $\mathrm{id}_V \in \mathrm{End}(V)$. Es gibt nur einen Eigenwert $\lambda = 1$, aber jede Basis von V diagonalisiert id_V. Die Untersuchung der Eigenräume zu gegebenem Eigenwert ist also entscheidend:

Definition 3.5 Ist $F \in \mathrm{End}(V)$ und ist $\lambda \in K$ Eigenwert von F, dann heißt

$$\mathrm{Eig}(F; \lambda) := \{v \in V \; : \; F(v) = \lambda \cdot v\}$$

der *Eigenraum* von F zum Eigenwert λ.

Zu beachten ist, daß der Nullvektor im Eigenraum liegt, $0 \in \mathrm{Eig}(F; \lambda)$, obwohl er kein Eigenvektor ist. Der Grund ist i) im folgenden Satz:

Satz 3.14 *Sei* $\mathrm{Eig}(F; \lambda)$ *der Eigenraum von* $F \in \mathrm{End}(V)$ *zum Eigenwert* λ. *Dann gilt:*

i) $\mathrm{Eig}(F; \lambda) \subset V$ *ist Untervektorraum*

ii) λ *ist Eigenwert von* $F \quad \Leftrightarrow \quad \mathrm{Eig}(F; \lambda) \neq \{0\} \quad \Leftrightarrow \quad \dim(\mathrm{Eig}(F; \lambda)) > 0$

iii) $\mathrm{Eig}(F; \lambda) \setminus \{0\}$ *ist die Menge der Eigenvektoren von* F *zum Eigenwert* λ

iv) $\mathrm{Eig}(F; \lambda) = \ker(F - \lambda \cdot \mathrm{id}_V)$

v) *Ist* $\lambda_1 \neq \lambda_2$*, so folgt* $\mathrm{Eig}(F; \lambda_1) \cap \mathrm{Eig}(F; \lambda_2) = \{0\}$

Beweis. i) Sind $v_1, v_2 \in \mathrm{Eig}(F; \lambda)$ und $\mu_1, \mu_2 \in K$, so folgt

$$F(\mu_1 v_1 + \mu_2 v_2) = \mu_1 F(v_1) + \mu_2 F(v_2) = \mu_1(\lambda v_1) + \mu_2(\lambda v_2) = \lambda(\mu_1 v_1 + \mu_2 v_2)$$

und damit $\mu_1 v_1 + \mu_2 v_2 \in \mathrm{Eig}(F; \lambda)$.

ii) Es gibt ein $v \neq 0$ mit $F(v) = \lambda v$. Dann ist $v \in \mathrm{Eig}(F; \lambda) \neq \{0\}$.

iii) ist klar

iv) Sei $v \in \mathrm{Eig}(F; \lambda)$. Dann gilt:

$$(F - \lambda \cdot \mathrm{id}_V)(v) = \lambda v - \lambda v = 0 \; ,$$

also $v \in \ker(F - \lambda \cdot \mathrm{id}_V)$. Ebenso folgt die Umkehrung.

v) Ist $\lambda_1 \neq \lambda_2$ und $v \in \mathrm{Eig}(F; \lambda_1)$, so ist

$$(F - \lambda_2 \cdot \mathrm{id}_V)(v) = (\lambda_1 - \lambda_2)v \ .$$

Damit ist $v \in \ker(F - \lambda_2 \cdot \mathrm{id}_V)$ genau dann, wenn $v = 0$. □

3.4 Das charakteristische Polynom

Es geht nun um ein Verfahren zur Bestimmung der Eigenwerte und Eigenvektoren eines Endomorphismus, wobei die Determinante eine entscheidende Rolle spielt.

Satz 3.15 *Sei $F \in \mathrm{End}(F)$ Endomorphismus eines endlich-dimensionalen Vektorraumes V. Ein Skalar $\lambda \in K$ ist genau dann Eigenwert von F, wenn*

$$\det(F - \lambda \cdot \mathrm{id}_V) = 0 \ .$$

Beweis. Nach Satz 3.14.ii) und Satz 3.14.iv) ist λ genau dann Eigenwert von F, wenn $\dim(\ker(F - \lambda \cdot \mathrm{id}_V)) > 0$. Das ist gleichbedeutend mit

$$\dim(\mathrm{im}(F - \lambda \cdot \mathrm{id}_V)) = \mathrm{rang}(F - \lambda \cdot \mathrm{id}_V) < \dim(V) \ .$$

Nach der Determinanteneigenschaft (D10) ist diese Eigenschaft äquivalent zu $\det(F - \lambda \cdot \mathrm{id}_V) = 0$. □

Die Determinante eines Endomorphismus $F \in \mathrm{End}(V)$ berechnet sich als die Determinante der darstellenden Matrix $M_{\mathcal{B}}^{\mathcal{B}}(F) = A = (a_{ij}) \in M(n, K)$ bezüglich einer beliebigen Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V. Dann gilt

$$M_{\mathcal{B}}^{\mathcal{B}}(F - \lambda \cdot \mathrm{id}_V) = A - \lambda \cdot E_n \ ,$$

denn $M_{\mathcal{B}}^{\mathcal{B}}$ ist eine lineare Abbildung (Satz 2.27), und $M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V) = E_n$ unabhängig von der Wahl der Basis. Damit ist die Suche nach Eigenwerten von $F \in \mathrm{End}(V)$ zurückgeführt auf die Bestimmung der *Nullstellen* der Abbildung

$$P_A : K \to K \ : \qquad P_A : \lambda \mapsto \det(A - \lambda \cdot E_n) \ .$$

Nach der Formel von Leibniz gilt

$$\det(A - \lambda \cdot E_n) = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda) + Q \ ,$$

wobei in Q aus Summen und Differenzen von Produkten der Matrixelemente besteht, in denen mindestens zwei Nichtdiagonalelemente a_{ij} mit $i \neq j$ auftreten (welche kein λ beinhalten): Wenn in einem solchen Produkt a_{ij} auftritt, dann berechnen sich nach dem Entwicklungssatz von Laplace die weiteren Faktoren zu

$(-1)^{i+j} \det A_{ij}$, aber A_{ij} enthält *nicht* die Diagonalelemente $a_{ii} - \lambda$ und $a_{jj} - \lambda$. Ausmultiplikation der Produkte und Ordnen nach Potenzen von λ ergibt:

$$\det(A - \lambda \cdot E_n) = \sum_{k=0}^{n} \alpha_k \lambda^k \;,$$
$$\alpha_n = (-1)^n \;,\; \alpha_{n-1} = (a_{11} + \cdots + a_{nn}) \;,\; \ldots,\; \alpha_0 = \det A \;.$$

Dabei heißt im zweithöchsten Term $\mathrm{tr}(A) := a_{11} + \cdots + a_{nn}$ die *Spur* der Matrix $A \in M(n, K)$. Der niedrigste Term ist unabhängig von λ und stimmt damit mit der Rechnung für $\lambda = 0$ überein, was gerade die Determinante von A ergibt. Die anderen Koeffizienten $\alpha_1, \ldots, \alpha_{n-2}$ sind schwieriger zu charakterisieren. Ein Ausdruck

$$\sum_{k=0}^{n} \alpha_k \lambda^k \;, \qquad \alpha_k \in K$$

heißt ein *Polynom in λ mit Koeffizienten im Körper K*. Das spezielle Polynom $P_A(\lambda) = \det(A - \lambda \cdot E_n)$ heißt das *charakteristische Polynom* der Matrix $A \in M(n, K)$.

3.5 Polynome und ihre Nullstellen

Definition 3.6 Ein Polynom in t mit Koeffizienten in K ist ein formaler Ausdruck der Form

$$f(t) = \alpha_0 + \alpha_1 t + \cdots + \alpha_n t^n = \sum_{i=0}^{n} \alpha_i t^i \;, \qquad \alpha_i \in K \;.$$

Die Menge aller solcher Polynomome werde mit $K[t]$ bezeichnet.

Dabei ist t eine *formale Variable*. Ein Beispiel ist die Wahl $t \mapsto \lambda \in K$; in diesem Fall definiert das Polynom eine Abbildung $f : K \to K$, $f : \lambda \mapsto f(\lambda)$.

Es gibt eine offensichtliche Addition

$$\Big(\sum_{i=0}^{n} \alpha_i t^i\Big) + \Big(\sum_{i=0}^{m} \beta_i t^i\Big) := \sum_{i=0}^{\max(m,n)} (\alpha_i + \beta_i) t^i \;,$$

wobei $\alpha_i = 0$ für $i > n$ und $\beta_i = 0$ für $i > m$ gesetzt wird. Das Nullpolynom $f(t) = 0$ (alle α_i verschwinden) ist das neutrale Element. Außerdem erhält man durch Ausmultiplizieren und Ordnen nach Potenzen von t eine Multiplikation

$$\Big(\sum_{i=0}^{n} \alpha_i t^i\Big) \cdot \Big(\sum_{i=0}^{m} \beta_i t^i\Big) := \sum_{i=0}^{m+n} \gamma_i t^i \;, \qquad \gamma_i = \sum_{j=0}^{i} \alpha_{i-j} \beta_j \;.$$

(Es wird wieder $\alpha_i = 0$ für $i > n$ und $\beta_i = 0$ für $i > m$ gesetzt.) Dadurch wird $K[t]$ zu einem kommutativen Ring, dem *Polynomring über K*. Der maximale

Exponent von t, für den der Koeffizient in $f(t)$ ungleich Null ist, heißt der *Grad des Polynoms* und wird mit $\deg(f)$ bezeichnet. Genauer ist für $f(t) = \sum_{i=0}^{n} \alpha_i t^i$

$$\deg(f) = \begin{cases} -\infty & \text{für } f = 0 \\ \max\{i \in \mathbb{N} \; : \; \alpha_i \neq 0\} & \text{sonst} \end{cases}$$

Es gilt

$$\deg(f+g) \leq \max(\deg(f), \deg(g)) \;, \qquad \deg(f \cdot g) = \deg(f) + \deg(g)$$

(mit $(-\infty) + n = -\infty$ und $(-\infty) + (-\infty) = -\infty$).

Satz 3.16 *Sind $f, g \in K[t]$ und ist $g \neq 0$, dann gibt es eindeutig bestimmte Polynome $q, r \in K[t]$ mit*

$$f = q \cdot g + r \;, \qquad \deg(r) < \deg(g) \;.$$

Beweis. (1) Man beweist zunächst die Eindeutigkeit. Sei $f = q \cdot g + r = q' \cdot g + r'$ mit $\deg(r) < \deg(g)$ und $\deg(r') < \deg(g)$, dann folgt

$$(q - q') \cdot g = r' - r \;.$$

Wäre $q \neq q'$, dann ist der Grad der linken Seite $\deg((q-q') \cdot g) > \deg(g)$, während der Grad der rechten Seite $\deg(r' - r) < \deg(g)$ ist. Dann kann die Gleichung aber nicht erfüllt sein, und deshalb ist $q = q'$ und dann $r = r'$.

(2) Die Existenz der Polynome beweist man durch explizite Konstruktion mittels "Division mit Rest". Sei $f = \sum_{i=0}^{n} \alpha_i t^i$ und $g = \sum_{i=0}^{m} \beta_i t^i$, mit $\alpha_n, \beta_m \neq 0$. Für $n < m$ ist $q = 0$ und $r = f$. Sei also $n \geq m$. Wir setzen

$$q_{(1)} := \frac{\alpha_n}{\beta_m} t^{n-m} \;, \qquad f_{(1)} := f - q_{(1)} \cdot g \;.$$

Da der höchste Koeffizient von f durch Subtraktion entfernt wurde, gilt $\deg(f_{(1)}) < \deg(f)$. Auf diese Weise konstruiert man eine Folge von Monomen $q_{(k)} = \gamma_k t^{i(k)}$, so daß für $f_{(k)} := f_{(k-1)} - q_{(k)} \cdot g$ gilt $\deg(f_{(k)}) < \deg(f_{(k-1)})$. Das Verfahren bricht im l-ten Schritt ab, wenn $\deg(f_{(l)}) < \deg(g)$. Dann ist $r := f_{(l)}$ und $q := q_{(1)} + \cdots + q_{(l)}$. □

Definition 3.7 Ein $\lambda \in K$ heißt *Nullstelle* eines Polynoms $f \in K[t]$, wenn $f(\lambda) = 0$.

Satz 3.17 *Ist $\lambda \in K$ eine Nullstelle von $f \in K[t]$, so gibt es ein eindeutig bestimmtes Polynom $g \in K[t]$ mit*

i) $f = (t - \lambda) \cdot g$

ii) $\deg(g) = \deg(f) - 1$

Beweis. i) Nach Satz 3.16 gibt es eindeutige Polynome $g, r \in K[t]$ mit $f = g \cdot (t - \lambda) + r$ und $\deg(r) < \deg(t - \lambda) = 1$. Also ist $r = \alpha_0 \in K$ ein Polynom nullten Grades, also ein Skalar. Für $t = \lambda$ ist $0 = f(\lambda) = (\lambda - \lambda) \cdot g(\lambda) + r = r$.

ii) ist kar. □

Ist $\lambda \in K$ Nullstelle von $f \in K[t]$ und sei $g \in K[t]$ durch $f = (t - \lambda) \cdot g$ definiert, dann kann das gleiche λ auch Nullstelle von g sein. Wir sagen, daß λ eine vielfache Nullstelle ist:

Definition 3.8 Ist $f \in K[t]$ verschieden vom Nullpolynom und $\lambda \in K$, so heißt

$$\mu(f; \lambda) := \max\{r \in \mathbb{N} \ : \ f = (t - \lambda)^r \cdot g \text{ für } g \in K[t]\}$$

die *Vielfachkeit* der Nullstelle λ.

Ist $f = (t - \lambda)^r \cdot g$ und $r = \mu(f; \lambda)$, so ist $g(\lambda) \neq 0$. Durch wiederholtes Abdividieren der Nullstellen läßt sich jedes Polynom $f \in K[t]$ darstellen als

$$f = (t-\lambda_1)^{r_1}(t-\lambda_2)^{r_2} \cdots (t-\lambda_k)^{r_k} \cdot g \,, \qquad \deg(g) = \deg(f) - r_1 - r_2 - \cdots - r_k \geq 0 \,,$$

wobei g ein Polynom *ohne Nullstellen* ist. Insbesondere besitzt jedes Polynom n-ten Grades höchstens n mit Vielfachheit gezählte Nullstellen (damit auch höchstens n paarweise verschiedene Nullstellen). Ist $\deg(g) = 0$, dann sagen wir, daß f *in Linearfaktoren zerfällt.* Von größter Bedeutung ist

Theorem 3.1 (Fundamentalsatz der Algebra) *Jedes Polynom $f \in \mathbb{C}[t]$ zerfällt in Linearfaktoren, d.h. es gibt $a \in \mathbb{C}$ und (nicht notwendig verschiedene) $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ mit $n = \deg(f)$, so daß*

$$f = a(t - \lambda_1) \cdots (t - \lambda_n) \,.$$

Der einfachste Beweis erfordert Hilfsmittel aus der Funktionentheorie.

Im reellen Fall $K = \mathbb{R}$ faktorisiert ein Polynom im allgemeinen *nicht* in Linearfaktoren. Das einfachste Beispiel ist $f(t) = t^2 + 1$, welches keine reelle Nullstelle besitzt. Man kann aber $\mathbb{R} \subset \mathbb{C}$ ausnutzen und somit für jedes reelle Polynom $f \in \mathbb{R}[t]$ mit $\deg(f) = n$ genau n *komplexe Nullstellen* $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ (gezählt mit Vielfachheit) finden. Es gibt zwei Möglichkeiten: Ist $\lambda_i \in \mathbb{R} \subset \mathbb{C}$, dann ist λ_i auch eine Nullstelle von $f \in \mathbb{R}[t]$. Ist $\lambda_i \notin \mathbb{R}$, dann ist auch $\bar{\lambda}_i$ eine komplexe Nullstelle, was durch komplexe Konjugation folgt:

$$f(t) = \sum_{i=0}^{n} \alpha_i t^i \quad \text{mit } \alpha_i \in \mathbb{R} \text{ und } f(\lambda) = 0 \text{ für } \lambda \notin \mathbb{R}$$

$$\Rightarrow \quad \bar{f}(\bar{t}) = \sum_{i=0}^{n} \alpha_i \bar{t}^i = f(\bar{t}) \qquad \text{hat Nullstelle } \bar{t} = \lambda, \text{ also } t = \bar{\lambda} \text{ Nullstelle von } f \,.$$

Außerdem ist die Vielfachheit der komplexen Nullstellen $\lambda, \bar{\lambda}$ gleich, d.h.

$$\mu(f;\lambda) = \mu(f,\bar{\lambda}) \qquad \forall\ f \in \mathbb{R}[t]\ .$$

Angenommen, wir hätten $f(t) = (t-\lambda)^r (t-\bar{\lambda})^{r'} g(t)$ mit $r \neq r'$ und $\lambda, \bar{\lambda}$ sind keine Nullstellen von $g \in \mathbb{C}[t]$. Dann sind $\lambda, \bar{\lambda}$ auch keine Nullstellen von $\bar{g} \in \mathbb{C}[t]$, und es gilt

$$\overline{f(t)} = f(\bar{t}) = (\bar{t}-\bar{\lambda})^r (\bar{t}-\lambda)^{r'} \overline{g(t)}\ .$$

Durch Austausch der formalen Variable $t \mapsto \bar{t}$ folgt, daß nun $\mu(f;\bar{\lambda}) = r$ und $\mu(f;\lambda) = r'$, also $r = r'$. Aus

$$(t-\lambda)(t-\bar{\lambda}) = t^2 - 2\mathrm{Re}(\lambda)t + |\lambda|^2 \in \mathbb{R}[t]$$

folgt nun, daß für jedes reelle Polynom $f \in \mathbb{R}[t]$ gilt:

$$\begin{aligned} &f = a(t-\lambda_1)\cdots(t-\lambda_{n-2r})\cdot g_1 \cdots g_r\ , \qquad g_j = (t-\alpha_j)^2 + \beta_j^2 > 0\ , \\ &\quad a, \lambda_i, \alpha_j, \beta_j \in \mathbb{R}\ , \quad \beta_j \neq 0\ , \qquad r, n-2r \geq 0\ . \end{aligned}$$

Die analytische Berechnung (zunächst komplexer Nullstellen) ist im allgemeinen nur für Polynome vom Grad ≤ 4 möglich. Für Polynome mit höherem Grad ist man auf numerische Näherungsverfahren angewiesen. Der Fundamentalsatz der Algebra und für reelle Polynome zusätzlich die Gleichheit der Vielfachheit komplexer Nullstellen ist dann eine wichtige Kontrolle, ob man wirklich alle Nullstellen numerisch gefunden hat!

3.6 Diagonalisierbarkeit

Wir kommen nun zurück auf Eigenwerte und Diagonalisierbarkeit.

Satz 3.18 *Sei $F \in \mathrm{End}(V)$ und $A = M_{\mathcal{B}}^{\mathcal{B}}(F)$ die darstellende Matrix bezüglich einer beliebigen Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V. Dann gilt*

$$\mathrm{Eig}(F;\lambda) = \Phi_{\mathcal{B}}(\mathrm{Lös}(A - \lambda \cdot E_n, 0))\ .$$

Dabei ist

$$\Phi_{\mathcal{B}} : K^n \to V\ , \qquad \Phi_{\mathcal{B}}(e_i) = v_i\ ,$$

der Isomorphismus, der die Standardbasis des K^n in die Basis $\mathcal{B}$ überführt, und

$$\mathrm{Lös}(A - \lambda \cdot E_n, 0) = \{x \in K^n\ :\ (A - \lambda \cdot E_n)x = 0\}\ .$$

Beweis. Sei $v \in \mathrm{Eig}(F;\lambda)$, dann ist

$$F(v) = \lambda v \quad \Leftrightarrow \quad \Phi_{\mathcal{B}}^{-1} \circ F \circ \Phi_{\mathcal{B}} \circ \Phi_{\mathcal{B}}^{-1}(v) = \lambda \Phi_{\mathcal{B}}^{-1}(v)\ .$$

Wir setzen $x = \Phi_{\mathcal{B}}^{-1}(v)$. Mit $A = M_{\mathcal{B}}^{\mathcal{B}}(F) := \Phi_{\mathcal{B}}^{-1} \circ F \circ \Phi_{\mathcal{B}} \in M(n,K)$ folgt

$$v \in \mathrm{Eig}(F;\lambda) \quad \Leftrightarrow \quad A \cdot x = \lambda \cdot x \quad \Leftrightarrow \quad (A - \lambda \cdot E_n)x = 0 \qquad \square$$

Damit haben wir ein Verfahren zur Bestimmung von Eigenwerten und Eigenvektoren von $F \in \mathrm{End}(V)$ entwickelt:

1. Wähle eine beliebige Basis $\mathcal{A} = (w_1, \ldots, w_n)$ von V. Bestimme die darstellende Matrix $A = (a_{ij})$ durch Zerlegen der n Vektoren $F(v_j)$ nach der Basis $\mathcal{A}$,
$$F(w_j) = \sum_{i=1}^{n} a_{ij} \cdot w_i \ .$$
2. Berechne das charakteristische Polynom $P_A(t) = \det(A - t \cdot E_n) \in K[t]$ und bestimme seine Nullstellen $\lambda_1, \ldots, \lambda_m \in K$ (paarweise verschieden) und ihre Vielfachheit $r_i := \mu(P_A; \lambda_i)$ aus der Darstellung
$$P_A(t) = (t - \lambda_1)^{r_1} \cdots (t - \lambda_m)^{r_m} \cdot g(t) \ ,$$
wobei $g \in K[t]$ keine Nullstelle in K besitzt. Die $\lambda_1, \ldots, \lambda_m$ sind dann genau die Eigenwerte von F. Für $K = \mathbb{C}$ gilt $n = r_1 + \cdots + r_m$.
3. Löse zu jeder Nullstelle λ_i des charakteristischen Polynoms das lineare Gleichungssystem $(A - \lambda_i \cdot E_n)x^{(i)} = 0$. Der Lösungsraum ist ein s_i-dimensionaler Untervektorraum von K^n. Ist $x^{(i)} = \sum_{j=1}^{n} x_j^{(i)} e_j \in \text{Lös}(A - \lambda_i \cdot E_n, 0)$, dann ist $v^{(i)} = \Phi_{\mathcal{B}}(x^{(i)}) = \sum_{j=1}^{n} x_j^{(i)} v_j \in \text{Eig}(F; \lambda_i)$ mit $\dim(\text{Eig}(F; \lambda_i)) = s_i$.

Sei $F \in \text{End}(V)$ mit $\dim(V) = n$ und $A = M_{\mathcal{A}}^{\mathcal{A}}(F) \in M(n, K)$ die darstellende Matrix bezüglich einer beliebigen Basis $\mathcal{A}$ von V. Wir wissen:

i) Ist F diagonalisierbar, dann ist $P_A(t) = (-1)^n (t - \lambda_1)^{r_1} \cdots (t - \lambda_m)^{r_m}$, d.h. das charakteristische Polynom zerfällt in Linearfaktoren (Wähle eine Basis aus Eigenvektoren).

ii) Ist $P_A = (-1)^n (t - \lambda_1) \cdots (t - \lambda_n)$ und alle λ_i sind paarweise verschieden, dann ist F diagonalisierbar.

Es verbleibt also zu untersuchen, wann F diagonalisierbar ist im Falle von Vielfachheiten der Eigenwerte.

Satz 3.19 *Sei $F \in \text{End}(V)$ und A seine darstellende Matrix bezüglich einer beliebigen Basis von V. Ist λ ein Eigenwert von F, dann gilt*
$$1 \leq \dim(\text{Eig}(F; \lambda)) \leq \mu(P_A; \lambda) \ .$$

Beweis. Sei $(v_1, \ldots, v_s)$ eine Basis von $\text{Eig}(F; \lambda)$. Dann ist $s \geq 1$, da λ Eigenwert. Wir ergänzen $(v_1, \ldots, v_s)$ zu einer Basis $\mathcal{A} = (v_1, \ldots, v_s, v_{s+1}, \ldots, v_n)$ von V. Für die darstellende Matrix $A = M_{\mathcal{A}}^{\mathcal{A}}(F) = (a_{ij})$ gilt
$$A = \left(\begin{array}{ccc|c} \lambda & & 0 & \\ & \ddots & & B \\ 0 & & \lambda & \\ \hline & 0 & & C \end{array}\right) ,$$

wobei oben links der Block λE_s steht. Dann gilt für das charakteristische Polynom

$$P_A(t) := \det(A - tE_n) = (-1)^s (t-\lambda)^s \det(C - tE_{n-s}) \ .$$

Folglich ist $s \leq \mu(P_A; \lambda)$. □

Der Fall $\dim(\mathrm{Eig}(F;\lambda)) = \mu(P_A;\lambda)$ ist von besonderem Interesse:

Satz 3.20 *Sei V ein n-dimensionaler Vektorraum, $F \in \mathrm{End}(V)$ und $A = M^{\mathcal{B}}_{\mathcal{B}}(F) \in M(n,K)$ die darstellende Matrix bezüglich einer beliebigen Basis $\mathcal{B}$ von V. Dann sind die folgenden Eigenschaften äquivalent:*

i) *F ist diagonalisierbar.*

ii) *Das charakteristische Polynom zerfällt in Linearfaktoren und es gilt $\dim(\mathrm{Eig}(F;\lambda)) = \mu(P_A;\lambda)$ für jeden Eigenwert λ von F.*

iii) *Sind $\lambda_1,\dots,\lambda_k$ die paarweise verschiedenen Eigenwerte von F, dann gilt $V = \mathrm{Eig}(F;\lambda_1) \oplus \cdots \oplus \mathrm{Eig}(F;\lambda_k)$.*

Beweis. i)⇒ii) Ist F diagonalisierbar, so ordnen wir die zugehörige Basis aus Eigenvektoren wie folgt:

$$\mathcal{B} = (v_1^{(1)},\dots,v_{s_1}^{(1)}, v_1^{(2)},\dots,v_{s_2}^{(2)},\dots,v_1^{(k)},\dots,v_{s_k}^{(k)}) \ .$$

Dabei ist $(v_1^{(i)},\dots,v_{s_i}^{(i)})$ eine Basis von $\mathrm{Eig}(F;\lambda_i)$, und insbesondere gilt $F(v_j^{(i)}) = \lambda_i v_j^{(i)}$ für $1 \leq j \leq s_i$. In dieser Basis gilt für das charakteristische Polynom

$$P_A(t) = (\lambda_1 - t)^{s_1} (\lambda_2 - t)^{s_2} \cdots (\lambda_k - t)^{s_k} \ ,$$

welches somit die Eigenschaften ii) besitzt.

ii)⇒iii) Durch $W = \mathrm{Eig}(F;\lambda_i) + \cdots + \mathrm{Eig}(F;\lambda_k)$ werde ein Untervektorraum von V definiert. Da Eigenvektoren zu verschiedenen Eigenwerten linear unabhängig sind, gilt $W = \mathrm{Eig}(F;\lambda_1) \oplus \cdots \oplus \mathrm{Eig}(F;\lambda_k)$. Dann folgt $\dim(W) = s_1 + \cdots + s_k = n$ und somit $W = V$.

iii)⇒i) Sei $\mathcal{B}_i = (v_1^{(i)},\dots,v_{s_i}^{(i)})$ eine Basis von $\mathrm{Eig}(F;\lambda_i)$. Dann ist

$$\mathcal{B} = (v_1^{(1)},\dots,v_{s_1}^{(1)}, v_1^{(2)},\dots,v_{s_2}^{(2)},\dots,v_1^{(k)},\dots,v_{s_k}^{(k)})$$

eine Basis von V. Wegen $F(v_j^{(i)}) = \lambda_i v_j^{(i)}$ für $1 \leq j \leq s_i$ ist $\mathcal{B}$ eine Basis aus Eigenvektoren von F, d.h. F ist diagonalisierbar. □

Wir sehen uns ein Beispiel zur Diagonalisierung an:

Beispiele 3.1 Es sei $F \in \mathrm{End}(\mathbb{R}^3)$ gegeben durch

$$F(x,y,z) := (y - z,\ 3x + 2y - 3z,\ 2x + 2y - 3z) \ .$$

1. Schritt: Bestimmung der darstellenden Matrix bezüglich einer Basis. Die Vektoren $v = (x, y, z)$ und $w = (y - z,\ 3x + 2y - 3z,\ 2x + 2y - 3z)$ sind bereits in der Standardbasis $\mathcal{E}_3 = (e_1, e_2, e_3)$ des $\mathbb{R}^3$ dargestellt. Daraus lesen wir

$$F(e_1) = 3e_2 + 2e_3\ , \qquad F(e_2) = e_1 + 2e_2 + 2e_3\ , \qquad F(e_3) = -e_1 - 3e_2 - 3e_3$$

ab. Dann ist die darstellende Matrix bezüglich der Standardbasis $A = (a_{ij}) = M_{\mathcal{E}_3}^{\mathcal{E}_3}(F)$ gegeben durch $F(e_j) = \sum_{i=1}^{3} a_{ij} \cdot e_i$. Wir lesen ab:

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

(Die Bilder der Basisvektoren ergeben die Spalten von A.)

2. Schritt: Berechnung des charakteristischen Polynoms. Zu berechnen ist $\det(A - t \cdot E_3)$, sinnvollerweise durch elementare Zeilenumformungen in eine obere Dreiecksmatrix:

$$\det(A - t \cdot E_3) = \det \begin{pmatrix} -t & 1 & -1 \\ 3 & 2-t & -3 \\ 2 & 2 & -3-t \end{pmatrix} \overset{P_{13}}{=} -\det \begin{pmatrix} 2 & 2 & -3-t \\ 3 & 2-t & -3 \\ -t & 1 & -1 \end{pmatrix}$$

$$\overset{Q_{12}(-\frac{3}{2})}{=} -\det \begin{pmatrix} 2 & 2 & -3-t \\ 0 & -1-t & \frac{3}{2} + \frac{3}{2}t \\ -t & 1 & -1 \end{pmatrix} \overset{Q_{13}(\frac{t}{2})}{=} -\det \begin{pmatrix} 2 & 2 & -3-t \\ 0 & -1-t & \frac{3}{2} + \frac{3}{2}t \\ 0 & 1+t & -1 - \frac{3}{2}t - \frac{1}{2}t^2 \end{pmatrix}$$

$$\overset{Q_{23}(1)}{=} -\det \begin{pmatrix} 2 & 2 & -3-t \\ 0 & -1-t & \frac{3}{2} + \frac{3}{2}t \\ 0 & 0 & \frac{1}{2} - \frac{1}{2}t^2 \end{pmatrix} = -2(-1-t)(\tfrac{1}{2} - \tfrac{1}{2}t^2) = (1+t)(1-t^2)$$

$$= -(t-1)(t-(-1))^2\ .$$

Damit zerfällt das charakteristische Polynom in Linearfaktoren. Seine Nullstellen sind $\lambda_1 = 1$ mit Vielfachheit $\mu(P_A(t), 1) = 1$ und $\lambda_2 = -1$ mit Vielfachheit $\mu(P_A(t), -1) = 2$.

3. Schritt: Bestimmen der Eigenräume. Zum Eigenwert $\lambda_1 = 1$ lösen wir das lineare Gleichungssystem $(A - \lambda_1 \cdot E_3)x = 0$ durch elementare Zeilenumformungen der erweiterten Koeffizientenmatrix

$$\left(\begin{array}{ccc||c} -1 & 1 & -1 & 0 \\ 3 & 1 & -3 & 0 \\ 2 & 2 & -4 & 0 \end{array}\right) \xrightarrow{Q_{12}(3)\,Q_{13}(2)} \left(\begin{array}{ccc||c} -1 & 1 & -1 & 0 \\ 0 & 4 & -6 & 0 \\ 0 & 4 & -6 & 0 \end{array}\right) \xrightarrow{Q_{23}(-1)} \left(\begin{array}{ccc||c} -1 & 1 & -1 & 0 \\ 0 & 4 & -6 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{Q_{12}(-\frac{1}{4})} \left(\begin{array}{cc|c||c} -1 & 0 & \frac{1}{2} & 0 \\ 0 & 4 & -6 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{S_1(-1)\,S_2(\frac{1}{4})} \left(\begin{array}{cc|c||c} 1 & 0 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{3}{2} & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) .$$

Damit gilt $\dim(\operatorname{Eig}(F;1)) = 1$, und das zu $(A - \lambda_1 \cdot E_3)x = 0$ äquivalente Problem $\begin{pmatrix} E_r & B \\ 0 & 0 \end{pmatrix} x = \begin{pmatrix} \tilde{b} \\ 0 \end{pmatrix}$ besitzt die Lösung $x = \begin{pmatrix} \tilde{b} \\ 0 \end{pmatrix} + \begin{pmatrix} -B \\ E_{n-r} \end{pmatrix} \cdot w$ mit $w \in K^{n-r}$. In unserem Fall erhalten wir also

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \hline 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ \frac{3}{2} \\ \hline 1 \end{pmatrix} \cdot w = \begin{pmatrix} \frac{1}{2}w \\ \frac{3}{2}w \\ w \end{pmatrix} .$$

Ein Basisvektor ist also $v_1^{(1)} = (1, 3, 2)$. Wir überprüfen:

$$\begin{aligned} F(v_1^{(1)}) &= F(e_1 + 3e_2 + 2e_3) = F(e_1) + 3F(e_2) + 2F(e_3) \\ &= (3e_2 + 2e_3) + 3(e_1 + 2e_2 + 2e_3) + 2(-e_1 - 3e_2 - 3e_3) = e_1 + 3e_2 + 2e_3 \\ &= v_1^{(1)} . \end{aligned}$$

Wir bestimmen nun den Eigenraum zu $\lambda_2 = -1$: Zu lösen ist das lineare Gleichungssystem $(A - \lambda_2 \cdot E_3)x = 0$ durch elementare Zeilenumformungen der erweiterten Koeffizientenmatrix

$$\left(\begin{array}{ccc||c} 1 & 1 & -1 & 0 \\ 3 & 3 & -3 & 0 \\ 2 & 2 & -2 & 0 \end{array}\right) \xrightarrow{Q_{12}(-3)\, Q_{13}(-2)} \left(\begin{array}{c|cc||c} 1 & 1 & -1 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) .$$

Damit gilt $\dim(\operatorname{Eig}(F;1)) = 2$, und die Lösungsvektoren sind

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \hline 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ \hline 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} -w_1 + w_2 \\ w_1 \\ w_2 \end{pmatrix} .$$

Damit finden wir die beiden Basisvektoren $v_1^{(2)} = (-1, 1, 0)$ und $v_2^{(2)} = (1, 0, 1)$. Wir überprüfen

$$\begin{aligned} F(v_1^{(2)}) &= F(-e_1 + e_2) = -F(e_1) + F(e_2) = -(3e_2 + 2e_3) + (e_1 + 2e_2 + 2e_3) \\ &= e_1 - e_2 = -v_1^{(2)} \end{aligned}$$

und

$$\begin{aligned} F(v_2^{(2)}) &= F(e_1 + e_3) = F(e_1) + F(e_3) = (3e_2 + 2e_3) + (-e_1 - 3e_2 - 3e_3) \\ &= -e_1 - e_3 = -v_2^{(2)} . \end{aligned}$$

Also sind die Dimensionen der Eigenräume gleich der Vielfachheit der Nullstellen, und F ist diagonalisierbar. Eine Basis von V, welche F diagonalisiert, ist also

$$\mathcal{B} = \big((1, 3, 2),\ (-1, 1, 0),\ (1, 0, 1)\big) .$$

In dieser Basis gilt also

$$M^{\mathcal{B}}_{\mathcal{B}}(F) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} .$$

Für diagonalisierbare Endomorphismen $F \in \mathrm{End}(V)$ ist auch die folgende Betrachtung interessant. Sei $\mathcal{A} = (w_1, \ldots, w_n)$ die beliebig gewählte Basis, bezüglich der wir die darstellende Matrix $A = M^{\mathcal{A}}_{\mathcal{A}}(F)$ und dann das charakteristische Polynom $P_A(t) = \det(A - t \cdot E_n)$ bestimmen. Sei $\mathcal{B} = (v_1, \ldots, v_n)$ die Basis aus Eigenvektoren. Dann können wir die Eigenvektoren wieder in der Basis $\mathcal{A}$ darstellen:

$$v_j = \sum_{i=1}^{n} c_{ij} \cdot w_i .$$

Wie die Schreibweise bereits andeutet, sehen wir c_{ij} als Komponenten einer Matrix $C = (c_{ij})$ an, deren j-te Spalte gerade die Komponenten des j-ten Eigenvektors bezüglich der Basis $\mathcal{A}$ sind. Wir berechnen das Matrixprodukt

$$\begin{aligned} A \cdot C &= A \cdot (v_1, \ldots, v_n) = (A \cdot v_1, \ldots, A \cdot v_n) = (\lambda_1 v_1, \ldots, \lambda_n v_n) \\ &= (v_1, \ldots, v_n) \cdot \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = C \cdot \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} . \end{aligned}$$

Durch Multiplikation mit C^{-1} erhalten wir die beiden Beziehungen

$$A = C^{-1} \cdot \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \cdot C , \qquad \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} = C \cdot A \cdot C^{-1} .$$

Wir führen die entsprechende Rechnung für das Beispiel 3.1 durch. Die Rechnung wurde in der Standardbasis $\mathcal{A} = \mathcal{E}_3$ durchgeführt. Die Spalten der Matrix C sind dann die Komponenten der Eigenvektoren in der Standardbasis,

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

Wir berechnen die inverse Matrix:

$$\left(\begin{array}{ccc||ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow{Q_{12}(-3)\, Q_{13}(-2)} \left(\begin{array}{ccc||ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 4 & -3 & -3 & 1 & 0 \\ 0 & 2 & -1 & -2 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{Q_{23}(-\frac{1}{2})} \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 4 & -3 & -3 & 1 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 1 \end{array}\right) \xrightarrow{Q_{32}(6)\, Q_{31}(-2)} \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 2 & 1 & -2 \\ 0 & 4 & 0 & -6 & -2 & 6 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 1 \end{array}\right)$$

$$\xrightarrow{Q_{21}(\frac{1}{4})} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 4 & 0 & -6 & -2 & 6 \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 1 \end{array}\right) \xrightarrow{S_2(\frac{1}{4})\, S_3(2)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & -\frac{3}{2} & -\frac{1}{2} & \frac{3}{2} \\ 0 & 0 & 1 & -1 & -1 & 2 \end{array}\right) .$$

Somit gilt

$$C^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{3}{2} & -\frac{1}{2} & \frac{3}{2} \\ -1 & -1 & 2 \end{pmatrix} .$$

Damit ergibt sich $A = C \cdot \mathrm{diag}(\lambda_i) \cdot C^{-1}$ zu

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

Zum Abschluß der Betrachtungen zur Diagonalisierbarkeit untersuchen wir folgendes Problem: Gegeben seinen zwei diagonalisierbare Endomorphismen $F, G \in \mathrm{End}(V)$. Unter welchen Bedingungen sind F, G *simultan diagonalisierbar*, d.h. es gibt eine Basis von V aus Eigenvektoren von F und G gleichzeitig?

Satz 3.21 *Zwei diagonalisierbare Endomorphismen $F, G \in \mathrm{End}(V)$ sind genau dann simultan diagonalisierbar, wenn sie miteinander* kommutieren, *d.h. wenn $F \circ G = G \circ F$.*

Beweis. ($\Rightarrow$) Sind F, G simultan diagonalisierbar, dann existiert eine Basis $\mathcal{B} = (v_1, \dots, v_n)$ von V mit $F(v_i) = \lambda_i v_i$ und $G(v_i) = \mu_i v_i$. Dann gilt für einen beliebigen Vektor $v = \sum_{i=1}^n \kappa_i v_i \in V$

$$(F \circ G)(v) = F(G(\sum_{i=1}^n \kappa_i v_i)) = \sum_{i=1}^n \lambda_i \kappa_i \mu_i v_i = (G \circ F)(v) .$$

($\Leftarrow$) Wir zerlegen den Vektorraum V in die Eigenräume:

$$\begin{aligned} V &= \mathrm{Eig}(F; \lambda_1) \oplus \cdots \oplus \mathrm{Eig}(F; \lambda_k) \\ &= \mathrm{Eig}(G; \mu_1) \oplus \cdots \oplus \mathrm{Eig}(G; \mu_l) . \end{aligned}$$

Kommutieren F und G, dann gilt $F(\mathrm{Eig}(G; \mu_j)) \subset \mathrm{Eig}(G; \mu_j)$ für alle $1 \le j \le l$, denn für $w_j \in \mathrm{Eig}(G; \mu_j)$ folgt

$$G(F(w_j)) = F(G(w_j)) = F(\mu_j w_j) = \mu_j F(w_j) .$$

Sei nun $W_{ij} := \mathrm{Eig}(F;\lambda_i) \cap \mathrm{Eig}(G;\mu_j)$. Dann ist $W_{ij} \subset V$ ein Untervektorraum, und es gilt $\mathrm{Eig}(F;\lambda_i) = W_{i1} \oplus \cdots \oplus W_{il}$. Denn sei $v_i \in \mathrm{Eig}(F;\lambda_i)$, dann gibt es $w_1 \in \mathrm{Eig}(G;\mu_1), \dots, w_l \in \mathrm{Eig}(G;\mu_l)$ mit $v_i = w_1 + \cdots + w_l$ (verwende Basis von V aus Eigenvektoren von G). Anwenden von F liefert

$$F(v_i) = F(w_1) + \cdots + F(w_l) = \lambda v_i = \lambda w_1 + \cdots + \lambda w_l \,.$$

Nun ist $F(w_j) \subset \mathrm{Eig}(G;\mu_j)$, und da Vektoren w_j, w'_j linear unabhängig sind für $j \neq j'$, folgt $F(w_j) = \lambda w_j$ für alle $1 \leq j \leq l$. Damit ist $w_j \subset \mathrm{Eig}(F;\lambda_i)$, also $\mathrm{Eig}(F;\lambda_i) = W_{i1} + \cdots + W_{il}$, und aus der linearen Unabhängigkeit folgt die Behauptung. Sei $\mathcal{B}_{ij} = (v_1^{(ij)}, \dots v_{s_{ij}}^{(ij)})$ eine Basis von W_{ij}, dann ist

$$(\mathcal{B}_{11}, \dots, \mathcal{B}_{1l}, \mathcal{B}_{21}, \dots, \mathcal{B}_{2l}, \dots, \mathcal{B}_{k1}, \dots, \mathcal{B}_{kl})$$

eine Basis von V, in der F und G simultan diagonalisierbar sind, mit $F(v_r^{(ij)}) = \lambda_i v_r^{(ij)}$ und mit $G(v_r^{(ij)}) = \mu_j v_r^{(ij)}$. □

Simultane Diagonalisierbarkeit (in verallgemeinerter Form) ist wichtig in der Quantenmechanik, wo man in einem System zwei physikalische Größen nur dann gleichzeitig messen kann, wenn die entsprechenden Endomorphismen (des Hilbert-Raumes) miteinander kommutieren. Im Wasserstoffatom sind das die Energie, der Gesamtdrehimpuls, eine Komponente des Drehimpulses (üblicherweise die z-Komponente) und der Spin. Entsprechend schreibt sich der Hilbert-Raum als direkte Summe von Eigenunterräumen der linearen Abbildungen, welche diesen physikalischen Größen entsprechen.

3.7 Trigonalisierung

Diagonalisierbarkeit eines Endomorphismus erfordert zwei Bedingungen:
(1) Das charakteristische Polynom zerfällt in Linearfaktoren.
(2) Die Vielfachheit der Nullstellen ist gleich der Dimension der Eigenräume.
Wir werden nun sehen, daß für Endomorphismen, die nur (1) erfüllen, eine Basis exisitiert, so daß die darstellende Matrix eine obere Dreiecksmatrix ist. Das genügt zur Lösung von Gleichungssystemen.

Definition 3.9 Sei $F \in \mathrm{End}(V)$. Ein Untervektorraum $W \subset V$ heißt *F-invariant*, wenn $F(W) \subset W$.

Offenbar sind die Eigenräume $\mathrm{Eig}(F;\lambda)$ automatisch F-invariant. Für die Triagonalisierung sind invariante Unterräume interessant, die keine Eigenräume sind.

Satz 3.22 *Sei $W \subset V$ ein F-invarianter Unterraum und $F|_W : W \to W$ die Einschränkung von F auf W. Sei $A|_W$ die darstellende Matrix von F_W und A die darstellende Matrix von F bezüglich beliebiger Basen. Dann ist das charakteristische Polynom $P_{A|_W}(t)$ ein Teiler von $P_A(t)$.*

Beweis. Sei $\dim(V) = n$ und $\dim(W) = r \leq n$. Wir ergänzen eine Basis $\mathcal{B}|_W$ von W zu einer Basis $\mathcal{B} = (\mathcal{B}|_W, \mathcal{B}')$ von V. Sei $A|_W := M^{\mathcal{B}|_W}_{\mathcal{B}|_W}(F|_W)$, dann gilt

$$A = M^{\mathcal{B}}_{\mathcal{B}}(F) = \begin{pmatrix} A|_W & * \\ 0 & A' \end{pmatrix} .$$

Damit ist $P_A(t) = \det(A - t \cdot E_n) = P_{A|_W}(t) \cdot \det(A' - t \cdot E_{n-r})$. □

Sei $F \in \mathrm{End}(K^n)$ in der Standardbasis durch eine obere Dreiecksmatrix $A \in M(n, K)$, d.h. $a_{ij} = 0$ für $i > j$, dargestellt. Definieren wir $W_i := \mathrm{span}(e_1, \ldots, e_r)$, dann gilt $F(W_r) \subset W_r$, d.h. alle W_r mit $1 \leq r \leq n$ sind F-invariant. Abstrakter formuliert:

Definition 3.10 Eine *Fahne* (V_r) in einem n-dimensionalen Vektorraum V ist eine Kette

$$\{0\} = V_0 \subset V_1 \subset V_2 \cdots \subset V_n = V$$

von Untervektorräumen mit $\dim(V_r) = r$. Ist $F \in \mathrm{End}(V)$, dann heißt die Fahne F-invariant, wenn $F(V_r) \subset V_r$ für alle $0 \leq r \leq n$.

Jede Basis von V definiert eine Fahne. Entscheidend ist, daß in einer F-invarianten Fahne gilt $F(V_1) \subset V_1$ mit $\dim(V_1) = 1$, so daß es einen Eigenvektor von F geben muß. Aus der Definition folgt direkt:

Satz 3.23 *Für $F \in \mathrm{End}(V)$ sind folgende Bedingungen äquivalent:*

i) *Es gibt eine F-invariante Fahne von V.*

ii) *Es gibt eine Basis $\mathcal{B}$ von V, so daß $M^{\mathcal{B}}_{\mathcal{B}}(F)$ eine obere Dreiecksmatrix ist.*

Ist das der Fall, dann heißt F trigonalisierbar.

Nun der entscheidende Satz:

Satz 3.24 *Für $F \in \mathrm{End}(V)$ mit $\dim(V) = n$ sind folgende Bedingungen äquivalent:*

i) *F ist trigonalisierbar.*

ii) *Das charakteristische Polynom von F zerfällt in Linearfaktoren, d.h.*

$$P_A(t) = (-1)^n (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n) , \qquad \lambda_1, \ldots, \lambda_n \in K ,$$

wobei A die darstellende Matrix von F in einer beliebigen Basis von V ist.

Insbesondere gilt: Jeder Endomorphismus eines endlich-dimensionalen komplexen Vektorraumes ist trigonalisierbar.

Beweis. i)⇒ii) ist klar, denn ist die darstellende Matrix A von F eine obere Dreiecksmatrix so gilt $P_A(t) = \det(A - t \cdot E_n) = (a_{11} - t)\cdots(a_{nn} - t)$, d.h. $\lambda_i = a_{ii}$.

ii)⇒i) durch Induktion in $n = \dim(V)$. Der Fall $n = 1$ ist klar. Sei also $n \geq 2$, dann wählen wir einen Eigenvektor v_1 zum Eigenwert (Nullstelle) λ_1 und ergänzen v_1 zu einer Basis $\mathcal{B} = (v_1, w_1, \ldots, w_{n-1})$ von V. Dann ist

$$V = V_1 \oplus W \qquad \text{mit} \quad V_1 := \mathrm{span}(v_1)\ ,\ W = \mathrm{span}(w_1, \ldots, w_{n-1})\ .$$

Nun ist V_1 ein F-invarianter Untervektorraum, W im allgemeinen aber nicht:

$$A = M_{\mathcal{B}}^{\mathcal{B}}(F) = \left(\begin{array}{c|ccc} \lambda_1 & a_1 & \ldots & a_{n-1} \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right) .$$

Diese Darstellung definiert zwei lineare Abbildungen

$$\begin{aligned} G &: W \to W\ , & G(w_j) &= \sum_{i=1}^{n-1} B_{ij} \cdot w_i\ , \\ H &: W \to V_1\ , & H(w_j) &= a_j \cdot v_1\ . \end{aligned}$$

Nun gilt für das charakteristische Polynom $P_A(t) = -(t-\lambda_1)P_B(t)$. Da $P_A(t)$ nach Voraussetzung in Linearfaktoren zerfällt, zerfällt auch $P_B(t)$ in Linearfaktoren. Nach Induktionsvoraussetzung ist die durch B definierte lineare Abbildung $G \in \mathrm{End}(W)$ trigonalisierbar. Es gibt also eine G-invariante Fahne $\{0\} = W_0 \subset W_1 \subset \cdots \subset W_{n-1} = W$. Wir setzen $V_r := V_1 + W_{r-1}$ für $1 \leq r \leq n$. Ist $v = \mu v_1 + w \in V_r$, also $w \in W_{r-1}$, dann gilt

$$F(v) = \mu F(v) + F(w) = \mu\lambda_1 v_1 + G(w) + H(w) \subset V_1 + W_{r-1}$$

wegen $H(w) \in V_1$ und $G(w) \in W_{r-1}$. Somit ist $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = V$ eine F-invariante Fahne, und F ist trigonalisierbar. $\square$

Beispiele 3.2 Eine gedämpfte Schwingung wird durch die Differentialgleichung

$$\ddot{x}\ (t) + 2\mu\dot{x}(t) + \omega^2 x(t) = 0\ , \qquad x(0) = x_0\ ,\ \dot{x}(0) = v_0$$

beschrieben. Wir setzen $y_1(t) = x(t)$, $y_2(t) = \dot{x}(t)$, dann ergibt sich ein System von zwei gekoppelten lineren Differentialgleichungen erster Ordnung

$$\dot{y} = A \cdot y \qquad \text{mit} \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\ , \quad A = \begin{pmatrix} 0 & 1 \\ -\omega^2 & -2\mu \end{pmatrix}\ , \quad y(0) = \begin{pmatrix} x_0 \\ v_0 \end{pmatrix}\ .$$

Da A unabhängig von t ist, ist die formale (und korrekte) Lösung gegeben durch

$$y(t) = \exp(At) \cdot y(0) \quad \text{mit} \quad \exp(At) = E_2 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \underbrace{A \cdots A}_{n} ,$$

jedoch lassen sich diese Produkte so nicht leicht berechnen. Der Ausweg besteht in der Trigonalisierung von A.

Zunächst hat $P_A(t) = t^2 + 2\mu t + \omega^2$ die komplexen Nullstellen $\lambda_{1,2} = -\mu \pm \sqrt{\mu^2 - \omega^2}$. Für $\mu = \omega$ hat die Nullstelle $\lambda = -\mu$ die Vielfachheit 2, die uns näher interessieren wird. Für $\mu = \omega$ bestimmen wir den Eigenraum zu $\lambda = -\omega$:

$$\begin{aligned}
(A + \omega \cdot E_2) \cdot v &= \begin{pmatrix} \omega & 1 \\ -\omega^2 & -\omega \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
\Rightarrow \left(\begin{array}{cc|c} \omega & \omega & 1 \\ -\omega^2 & -\omega & 0 \end{array} \right) &\longrightarrow \left(\begin{array}{cc|c} \omega & 1 & 0 \\ 0 & 0 & 0 \end{array} \right) \qquad \Rightarrow v = \begin{pmatrix} 1 \\ -\omega \end{pmatrix} \cdot w \, .
\end{aligned}$$

Damit ist $\dim(\mathrm{Eig}(A; -\omega)) = 1$, aber die Nullstelle hat Vielfachheit 2. Folglich ist A nicht diagonalisierbar.

Zur Trigonalisierung wählen wie die Basis $\mathcal{B} = (v, e_2)$, dann ist

$$A = \underbrace{\begin{pmatrix} 1 & 0 \\ -\omega & 1 \end{pmatrix}}_{C} \begin{pmatrix} -\omega & 1 \\ 0 & -\omega \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 0 \\ \omega & 1 \end{pmatrix}}_{C^{-1}} .$$

Somit gilt

$$\begin{aligned}
A^n &= C \begin{pmatrix} -\omega & 1 \\ 0 & -\omega \end{pmatrix}^n C^{-1} = C \begin{pmatrix} (-\omega)^n & n\omega^{n-1} \\ 0 & (-\omega)^n \end{pmatrix}^n C^{-1} \\
\Rightarrow \exp(At) &= C \begin{pmatrix} e^{-\omega t} & -\frac{1}{\omega}(e^{-\omega t} - 1) \\ 0 & e^{-\omega t} \end{pmatrix} C^{-1}
\end{aligned}$$

und damit insbesondere

$$(C^{-1}y)(t) = \begin{pmatrix} e^{-\omega t} & -\frac{1}{\omega}(e^{-\omega t} - 1) \\ 0 & e^{-\omega t} \end{pmatrix} C^{-1} y(0)$$

Unter Verwendung von $C^{-1}y = \begin{pmatrix} x \\ \omega x + \dot{x} \end{pmatrix}$ erhalten wir für die 2. Komponente

$$\begin{aligned}
&& \omega x + \dot{x} &= \mathrm{e}^{-\omega t}(\omega x_0 + v_0) \\
\Rightarrow && \mathrm{e}^{-\omega t} \frac{d}{dt}(\mathrm{e}^{\omega t} x) &= \mathrm{e}^{-\omega t}(\omega x_0 + v_0) \\
\Rightarrow && (\mathrm{e}^{\omega t} x) &= const + t(\omega x_0 + v_0) \, , \quad t = 0 \;\Rightarrow\; const = x_0 \\
\Rightarrow && x &= \mathrm{e}^{-\omega t}(x_0 + t\omega x_0 + t v_0) \, .
\end{aligned}$$

3.8 Das Minimalpolynom

Sei $F \in \operatorname{End}(V)$ Endomorphismus eines endlich-dimensionalen Vektorraumes V. Jedes Polynom $f(t) \in K[t]$, $f(t) = \sum_{i=0}^{k} \alpha_i t^i$ definiert eine Abbildung $f : \operatorname{End}(V) \to \operatorname{End}(V)$ durch

$$f(F) = \sum_{i=1}^{k} \alpha_i F^i , \qquad F^i = \underbrace{F \circ \cdots \circ F}_{i \text{ mal}} , \;\; F^0 = \mathrm{id}_V .$$

Zu gegebenem $F \in \operatorname{End}(V)$ ist die Menge $K[F]$ aller Polynome in F mit Koeffizienten in K ein *kommutativer Unterring* von $\operatorname{End}(V)$. Interessant sind nun die "Nullstellen" solcher Polynome: Wir bezeichnen mit

$$\mathcal{I}_F := \{f \in K[t] \;:\; f(F) = 0 \in \operatorname{End}(F)\} \subset K[t]$$

die Menge aller Polynome $f \in K[t]$ mit $f(F) = 0$ (rechts steht der Nullendomorphismus, nicht $0 \in K$). Klar ist, daß $\mathcal{I}_F$ ein Unterring von $K[t]$ ist. Außerdem gilt

$$\mathcal{I}_F \cdot K[t] \subset \mathcal{I}_F , \qquad \text{d.h. } g(F) \cdot f(F) \subset \mathcal{I}_F \quad \forall g \in \mathcal{I}_F , \; f \in K[t] .$$

Eine Teilmenge eines Ringes mit solchen Eigenschaften heißt *Ideal*:

Definition 3.11 Eine Teilmenge $\mathcal{I}$ eines Ringes R heißt Linksideal bzw. Rechtsideal, wenn

(I1) $P, Q \in \mathcal{I} \quad \Rightarrow \quad P - Q \in \mathcal{I}$

(I1) $P \in \mathcal{I}$, $Q \in R \quad \Rightarrow \quad P \cdot Q \in \mathcal{I}$ bzw. $Q \cdot P \in \mathcal{I}$.

Ein Linksideal, das gleichzeitig Rechtsideal ist, heißt beidseitiges Ideal.

In kommutativen Ringen sind Links-, Rechts- und beidseitige Ideale identisch.

Die Frage ist nun: Was ist zu gegebenem $F \in \operatorname{End}(V)$ der *minimale Grad d* eines Polynoms $f \in \mathcal{I}_F$? Eine Schranke gibt der

Satz 3.25 (Satz von Cayley-Hamilton) *Sei V ein n-dimensionaler Vektorraum, $F \in \operatorname{End}(V)$ ein Endomorphismus und $P_F(t)$ sein charakteristisches Polynom. Dann gilt $P_F(F) = 0 \in \operatorname{End}(V)$. Insbesondere gilt für jede Matrix $A \in M(n, K)$ die Identität $P_A(A) = 0 \in M(n, K)$.*

Beweis. Wir beweisen den Satz nur für die (physikalisch interessanten Fälle) $K = \mathbb{R}$ und $K = \mathbb{C}$ und beginnen mit dem komplexen Fall. Dann ist F trigonalisierbar, und es gibt eine Basis $\mathcal{B} = (v_1, \dots, v_n)$, so daß

$$A = M_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} .$$

Folglich gilt $P_A(t) = P_F(t) = (\lambda_1 - t) \cdots (\lambda_n - t)$ und dann

$$P_F(F) = (\lambda_1 \cdot \mathrm{id}_V - F) \circ \cdots \circ (\lambda_n \cdot \mathrm{id}_V - F) \ .$$

Wir nutzen, daß $\{0\} = V_0 \subset V_1 \subset \cdots \subset V_n = V$ mit $V_i = \mathrm{span}(v_1, \ldots, v_i)$ eine invariante Fahne ist, d.h. $F(V_i) \subset V_i$ und $\dim(V_i) = i$. Sei entsprechend

$$G_i := (\lambda_1 \cdot \mathrm{id}_V - F) \circ \cdots \circ (\lambda_i \cdot \mathrm{id}_V - F) \ .$$

Wir zeigen mittels vollständiger Induktion, daß $G_i(V_i) = 0 \in V$ gilt für alle $1 \leq i \leq n$. Der Induktionsanfang ist klar: Ein beliebiger Vektor aus V_1 hat die Form $\mu_1 v_1$ und ist für $\mu_1 \neq 0$ ein Eigenvektor von F. Dann ist

$$G_1(\mu_1 v_1) = (\lambda_1 \cdot \mathrm{id}_V - F)(\mu_1 v_1) = \mu_1(\lambda_1 v_1 - F(v_1)) = 0 \ .$$

Angenomen, $G_{i-1}(V_{i-1}) = 0$ ist gezeigt. Dann hat ein beliebiger Vektor $v \in V_i$ die Gestalt $v = \mu v_i + w$ mit $w \in V_{i-1}$. Es gilt

$$\lambda_i w - F(w) \in V_{i-1} \ , \qquad \lambda_i v_i - F(v_i) \in V_{i-1} \ ,$$

(die erste Beziehung aus der Invarianz von V_i, die zweite aus der Darstellung über die obere Dreiecksmatrix A) und somit

$$G_i(\mu v_i + w) = G_{i-1} \circ (\lambda_i \cdot \mathrm{id}_V - F)(\mu v_i + w) \subset G_{i-1}(V_{i-1}) = 0 \ .$$

Im reellen Fall können wir dennoch jede Matrix komplex trigonalisieren, d.h. zu $A \in M(n, \mathbb{R})$ gibt es $C \in M(n, \mathbb{C})$ und eine obere Dreiecksmatrix $D \in M(n, \mathbb{C})$ mit $A = CDC^{-1}$. Dann ist $P_A(t) = \det(A - tE_n) = \det(D - tE_n) = P_D(t)$ und somit

$$P_D(A) = P_D(CAC^{-1}) = CP_D(D)C^{-1} = 0 \ . \qquad \square$$

Satz 3.26 *Zu jedem Ideal $\mathcal{I} \subset K[t]$ mit $\mathcal{I} \neq \{0\}$ gibt es ein eindeutiges Polynom M mit folgenden Eigenschaften:*

i) *M ist normiert, d.h. $M = \sum_{i=0}^{d} \alpha_i t^i$ mit $\alpha_d = 1$.*

ii) *Für jedes $P \subset \mathcal{I}$ gibt es ein $Q \in K[t]$ mit $P = Q \cdot M$.*

Dieses Polynom M heißt das Minimalpolynom von $\mathcal{I}$. Ist $\mathcal{I} = \mathcal{I}_F$, so heißt M das Minimalpolynom von F.

Beweis. Sei $d \geq 0$ der minimale Grad eines Polynoms $M = \sum_{i=0}^{r} \alpha_i t^i \in \mathcal{I}$. Durch Division durch den höchsten Koeffizienten $\alpha_d \neq 0$ kann stets $\alpha_d = 1$ erreicht werden. Die weiteren Koeffizienten $\alpha_0, \ldots, \alpha_{d-1}$ sind zunächst noch unbestimmt. Sei $P \in \mathcal{I}$ ein beliebiges Polynom aus dem Ideal. Dann gilt $\deg(P) \geq d$. Wir dividieren P mit Rest durch M:

$$P = Q \cdot M + R \ , \qquad Q, R \in K[t] \ , \ \deg(R) < d \ .$$

Wäre $R \neq 0$, so wäre $R = P - Q \cdot M \subset \mathcal{I}$ im Widerspruch zur Minimalität des Grades d von Polynomen in $\mathcal{I}$. Insbesondere kann es nur ein normiertes Polynom minimalen Grades geben. □

Der Satz ist eine reine Existenzaussage. Er liefert keinen Hinweis, wie das Minimalpolynom wirklich ausgerechnet werden kann. Es gibt jedoch eine nützliche Beziehung zwischen dem Minimalpolynom und dem charakteristischen Polynom:

Satz 3.27 *Für einen beliebigen Endomorphismus $F \in \mathrm{End}(V)$ haben das charakteristische Polynom $P_F(t) \in K[t]$ und das Minimalpolynom $M_F(t) \in K[t]$ dieselben Nullstellen.*

Beweis. Wir betrachten das Ideal $\mathcal{I}_F = \{P \in K[t] \;:\; P(F) = 0\}$. Nach dem Satz von Cayley-Hamilton ist $P_F \in \mathcal{I}_F$, so daß es auf Grund der Eigenschaften des Minimalpolynoms ein Polynom $g \in K[t]$ gibt mit $P_F(t) = g(t) \cdot M_F(t)$. Damit ist jede Nullstelle von M_F auch eine Nullstelle von P_F.

Sei umgekehrt λ eine Nullstelle von P_F. Dann ist λ ein Eigenwert von F. Wir wählen einen Eigenvektor $v \in V$ mit $v \neq 0$ von F zum Eigenwert λ. Sei $M_F(t) = t^d + \sum_{i=0}^{d-1} \alpha_i t^i$ das Minimalpolynom. Dann gilt

$$(M_F(F))(v) = F^d(v) + \sum_{i=1}^{d-1} \alpha_i F^i(v) + \alpha_0 v = \Big(\lambda^d + \sum_{i=1}^{d-1} \alpha_i \lambda^i + \alpha_0\Big) \cdot v = M_F(\lambda) \cdot v \;.$$

Andererseits ist $M_F(F)$ der Nullendomorphismus, also $(M_F(F))(v) = 0 \in V$ für alle $v \in V$, so daß $M_F(\lambda) = 0 \in K$. □

Satz 3.28 *Sei $F \in \mathrm{End}(V)$. Wenn das charakteristische Polynom von F in Linearfaktoren zerfällt, $P_F(t) = (-1)^n (t - \lambda_1)^{\mu_1} \cdots (t - \lambda_r)^{\mu_r}$ mit $\lambda_1, \dots, \lambda_r \in K$ paarweise verschieden, dann zerfällt auch das Minimalpolynom in Linearfaktoren, und es gibt natürliche Zahlen $\gamma_1, \dots, \gamma_r$ mit $1 \leq \gamma_i \leq \mu_i$, so daß*

$$M_F(t) = (t - \lambda_1)^{\gamma_1} \cdots (t - \lambda_r)^{\gamma_r} \;.$$

Beweis. Nach dem Satz von Cayley-Hamilton gibt es ein Polynom $g \in K[t]$ mit $P_F(t) = g(t) \cdot M_F(t)$. Ist λ_1 Nullstelle von P_F, dann ist λ Nullstelle von g oder von M_F. Nach Satz 3.17 gilt dann $g = (t - \lambda_1) g_1$ oder $M_F = (t - \lambda_1) M_1$ für Polynome $g_1, M_1 \in K[t]$. Wir könnnen dann das Polynom $t - \lambda_1$ ohne Rest abdividieren und wiederholen das Verfahren für alle Linearfaktoren $(t - \lambda_i)$ entsprechend der Vielfachheit μ_i. Das Ergebnis ist eine Gleichung der Form

$$1 = g_{\mu_1 + \dots + \mu_r} \cdot M_{\mu_1 + \dots + \mu_r} \;, \qquad \begin{aligned} g &= g_{\mu_1 + \dots + \mu_r} (t - \lambda_1)^{\mu_1 - \gamma_1} \cdots (t - \lambda_r)^{\mu_1 - \gamma_r} \;, \\ M_F &= M_{\mu_1 + \dots + \mu_r} (t - \lambda_1)^{\gamma_1} \cdots (t - \lambda_r)^{\gamma_r} \;, \end{aligned}$$

mit $\gamma_i > 0$ und $\mu_i - \gamma_i > 0$. Nach dem vorigen Satz ist in Wirklichkeit $\gamma_i > 1$. Zunächst sind $g_{\mu_1 + \dots + \mu_r}$ und $M_{\mu_1 + \dots + \mu_r}$ beliebige Polynome, aber aus

$\deg(g_{\mu_1+\cdots+\mu_r}) = \deg(M_{\mu_1+\cdots+\mu_r}) = 0$ folgt $g_{\mu_1+\cdots+\mu_r}, M_{\mu_1+\cdots+\mu_r} \in K$ und dann die Behauptung. □

Damit haben wir in einem wichtigen Spezialfall folgende Strategie zur Bestimmung des Minimalpolynoms von $F \in \mathrm{End}(V)$ gefunden: Wenn das charakteristische Polynom von F in Linearfaktoren zerfällt mit $P_F(t) = (-1)^n(t-\lambda_1)^{\mu_1}\cdots(t-\lambda_r)^{\mu_r}$, dann probiere für alle Mengen natürlicher Zahlen $\{\gamma_1,\dots,\gamma_r\}$ mit $1 \leq \gamma_i \leq \mu_i$, ob $(F-\lambda_1\mathrm{id}_V)^{\gamma_1} \circ \dots (F-\lambda_r\mathrm{id}_V)^{\gamma_r} = 0$ ist. Die Lösung mit minimalem $\gamma_1+\cdots+\gamma_r$ ist das Minimalpolynom. Starten würde man mit $\gamma_1 = \cdots = \gamma_r = 1$. Ist damit bereits das Minimalpolynom gefunden, so ist F diagonalisierbar:

Satz 3.29 *Ein Endomorphismus $F \in \mathrm{End}(V)$ ist genau dann diagonalisierbar, wenn das charakteristische Polynom in Linearfaktoren zerfällt mit paarweise verschiedenen Nullstellen $\lambda_1,\dots,\lambda_r$ und $M_F = (t-\lambda_1)\cdots(t-\lambda_r)$ das Minimalpolynom von F ist.*

Beweis. (⇒) Ist F diagonalisierbar, so zerfällt das charakterisitsche Polynom in Linearfaktoren mit paarweise verschiedenen Nullstellen $\lambda_1,\dots,\lambda_r$, und für die Eigenräume gilt

$$V = \mathrm{Eig}(F;\lambda_1) \oplus \cdots \mathrm{Eig}(F;\lambda_r)\ .$$

Damit hat ein beliebiger Vektor aus $v \in V$ die Darstellung $v = v_1+\cdots+v_r$ mit $v_i \in \mathrm{Eig}(F;\lambda_i)$. Wir wenden den Endomorphismus $(F-\lambda_1\cdot\mathrm{id}_V)\cdots(F-\lambda_r\cdot\mathrm{id}_V)$ auf v an. Wegen der Kommutativität der Faktoren können wir zu jedem v_i den Faktor $(F-\lambda_i\cdot\mathrm{id}_V)$ nach rechts bringen, für den $(F-\lambda_i\cdot\mathrm{id}_V)(v_i) = 0$ gilt. Damit ist $M_F(t) = (t-\lambda_1)\cdots(t-\lambda_r)$ das Minimalpolynom von F.

(⇐) Aus Zeitgründen verzichten wir auf den Beweis der Umkehrung... □

Stattdessen soll die Diagonalisierbarkeit für die bisherigen beiden Beispiele durch Kontrolle des Minimalpolynoms untersucht werden. Im Beispiel zur gedämpften Schwingung hatte wir die Matrix

$$A = \begin{pmatrix} 0 & 1 \\ -\omega^2 & -2\mu \end{pmatrix}$$

betrachtet. Das charakteristische Polynom hatte die Nullstellen $\lambda_{1,2} = -\mu \pm \sqrt{\mu^2-\omega^2}$. Im Fall $\mu \neq \lambda$ ist A (zumindest komplex) diagonalisierbar. Für $\mu = \omega$ hat die Nullstelle die Vielfachheit 2, und damit ist A genau dann diagonalisierbar, wenn $(t+\mu)$ das Minimalpolynom ist. Wir haben aber

$$A + \mu\cdot E_n = \begin{pmatrix} \mu & 1 \\ -\mu^2 & -\mu \end{pmatrix} \neq 0\ ,$$

also ist A nicht diagonalisierbar.

Das erste Beispiel war

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

mit charakteristischem Polynom $P_A(t) = (1-t)(t+1)^2$. Wir testen also, ob $(t-1)(t+1) = t^2 - 1$ das Minimalpolynom ist, d.h. $A^2 = E_3$ gilt:

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

Damit ist A diagonalisierbar.