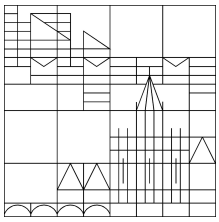
Claus Scheiderer

Lineare Algebra I

Universität Konstanz, WS 2016/17

Skript, Stand 18. Februar 2017



Inhaltsverzeichnis

Literaturverzeichnis

[1] G. Fischer: *Lineare Algebra*. 18. Aufl., Springer Spektrum, 2014.

[2] G. Fischer: *Lernbuch Lineare Algebra und Analytische Geometrie*. Das Wichtigste ausführlich für das Lehramts- und Bachelorstudium. 2. Aufl., Springer Vieweg, 2012.

[3] S. Bosch: *Lineare Algebra*. 5. Aufl., Springer, 2014.

[4] H-J. Kowalsky, G. Michler: *Lineare Algebra*. 12. Aufl., De Gruyter, 2003.

[5] E. Brieskorn: *Lineare Algebra und Analytische Geometrie I, II*. Vieweg + Teubner, 1983.

[6] G. Strang: *Lineare Algebra*. Springer, 2003.

[7] A. Beutelspacher: *Lineare Algebra*. 8. Aufl., Springer Spektrum, 2014.

KAPITEL I

Grundlagen

1. Mengen und Abbildungen

In der ersten Vorlesungswoche erläutern wir zunächst die grundlegenden Konzepte um Mengen und Abbildungen. Danach beginnen wir mit dem eigentlichen Stoff der linearen Algebra.

1.1 *Aussagenlogik:* Jede (mathematische) Aussage ist entweder wahr oder falsch. Zum Beispiel sind

$$4 = 4, \quad 3^2 > 5, \quad 1 + 1 \neq 3$$

wahre Aussagen,

$$1 = 2, \quad 2 \cdot 3 = 5, \quad 7 - 2 < 5$$

dagegen falsche. Die Operatoren der Aussagenlogik sind

$$\wedge \text{ (und)}, \ \vee \text{ (oder)}, \ \neg \text{ (nicht)},$$
$$\Rightarrow \text{ (impliziert)}, \ \Leftrightarrow \text{ (genau dann wenn, äquivalent zu)}.$$

Sind also A, B zwei Aussagen, so sind auch $\neg A$, $A \wedge B$, $A \vee B$, $A \Rightarrow B$, $A \Leftrightarrow B$ wieder Aussagen. Die Bedeutung dieser Operatoren ergibt sich aus den sogenannten Wahrheitstafeln: Steht 1 für wahr und 0 für falsch, so ist der Wahrheitswert der genannten Aussagen gegeben durch

A	$\neg A$
0	1
1	0

und

A	B	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Leftrightarrow B$
0	0	0	0	1	1
0	1	0	1	1	0
1	0	0	1	0	0
1	1	1	1	1	1

Beachte: $A \vee B$ bedeutet "A oder B oder beide". Das "sowohl... als auch" ist darin also eingeschlossen. In der Umgangssprache des Alltags wird das Wort "oder" dagegen meistens im Sinn von "entweder ... oder" verwendet. Das entspricht nicht dem mathematischen "oder"!

Aus einer falschen Aussage folgt stets jede beliebige (wahre oder falsche) Aussage. Eine wahre Aussage folgt stets aus jeder (wahren oder falschen) Aussage. Die Äquivalenz $A \Leftrightarrow B$ ist genau dann wahr, wenn sowohl $A \Rightarrow B$ als auch $B \Rightarrow A$ wahr sind.

Neben diesen Symbolen gibt es noch die Quantoren $\forall$ (für alle) und $\exists$ (es existiert), mit der Bedeutung

$\exists x\ P(x)$: es existiert (mindestens) ein x mit der Eigenschaft $P(x)$,
$\forall\, x\ P(x)$: für alle x gilt die Eigenschaft $P(x)$.

Für die negierten Quantoren gilt

$$\neg\Big(\exists\, x\ \mathcal{P}(x)\Big) \quad \text{ist äquivalent zu} \quad \forall\, x\ \neg\,\mathcal{P}(x),$$

$$\neg\Big(\forall\, x\ \mathcal{P}(x)\Big) \quad \text{ist äquivalent zu} \quad \exists\, x\ \neg\,\mathcal{P}(x).$$

1.2 *Mengen:* Der Mengenbegriff geht auf Cantor[1] zurück, und ist heute in allen Bereichen der Mathematik die Grundlage der mathematischen Sprache und Ausdrucksweise. Die Frage, wie man den Begriff der Menge präzise einführt, ist keineswegs leicht zu beantworten. Geht man zu naiv vor, erhält man unauflösliche Widersprüche. Mit solchen Grundlagenfragen beschäftigt sich ein eigenes Teilgebiet der Mathematik, die Mengenlehre. Für die allermeisten Teile der Mathematik ist es jedoch gar nicht nötig, dieses Problem mit letzter Strenge zu lösen, und es genügt ein naiver Zugang, wie er hier eingeführt wird. Wichtig ist vor allem, welche Beziehungen zwischen verschiedenen Mengen bestehen, und wie man aus gegebenen Mengen neue bilden kann.

Eine *Menge* besteht aus *Elementen*, und ist dadurch bestimmt, welche Elemente zu ihr gehören. Die Elemente können Objekte aller Art sein, z.B. Zahlen, Funktionen, mathematische Strukturen, oder auch selbst Mengen. Zum Beispiel hat die Menge

$$M := \{1, 2, 3, 4, 5, 6\}$$

genau sechs Elemente, nämlich die natürlichen Zahlen von 1 bis 6. Die Menge

$$N := \{\{1\}, \{2, 3\}, \{4, 5, 6\}\}$$

hat dagegen genau drei Elemente, nämlich die drei Mengen $\{1\}$, $\{2, 3\}$ und $\{4, 5, 6\}$.

Daß x ein Element der Menge X ist, drückt man aus durch die Schreibweise $x \in X$; entsprechend bedeutet $x \notin X$, daß x kein Element von X ist. Im obigen Beispiel ist also $1 \in M$ und $\{1, 2\} \notin M$, aber $1 \notin N$ und $\{1, 2\} \in N$.

Zwei Mengen X, Y heißen *gleich*, i. Z. $X = Y$, wenn sie genau dieselben Elemente enthalten. Eine Menge wird angegeben, indem man alle ihre Elemente angibt. Das kann durch Aufzählen geschehen, wie in den obigen Beispielen, oder wie in

$$\mathbb{N} := \{1, 2, 3, 4, 5, \dots\}$$

(die Menge der *natürlichen Zahlen*),

$$\mathbb{N}_0 := \{0, 1, 2, 3, 4, 5, \dots\}$$

(die Menge der natürlichen Zahlen zusammen mit der Null),

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

(die Menge der *ganzen Zahlen*). Die Pünktchen bedeuten "und so weiter". Diese Schreibweise darf nur verwendet werden, wenn das Gesetz, nach dem die Aufzählung weiter gehen soll, aus dem ausgeschriebenen Teil eindeutig erkennbar ist. Zum Beispiel ist eine Beschreibung wie

$$\{1, 3, \dots\}$$

[1] Georg Cantor (1845–1918)

zu knapp; es könnte die Menge

$$\{1, 3, 5, 7, 9, \dots\} \;=\; \{2n-1 \colon n \in \mathbb{N}\}$$

aller ungeraden natürlichen Zahlen gemeint sein, oder auch die Menge

$$\{1, 3, 7, 15, 31, \dots\} \;=\; \{2^n - 1 \colon n \in \mathbb{N}\},$$

usw. Hier sehen wir schon eine andere häufig gebrauchte Notation: Mengen werden oft in der Form

$$\{x \colon P(x)\} \quad (\text{oder} \quad \{x \mid P(x)\})$$

angegeben, die Menge aller Elemente x, welche die Eigenschaft P haben. Zum Beispiel kann man die Menge aller reellen Zahlen[2] zwischen 0 und 1 nicht in aufzählender Form hinschreiben, wohl aber in der Form

$$\{x \in \mathbb{R} \colon 0 \le x \le 1\}.$$

Mengen müssen wohldefiniert sein, d. h. es muß zumindest im Prinzip eindeutig entscheidbar sein, was ihre Elemente sind.

In der Beschreibung einer Menge brauchen die angegebenen Elemente nicht alle voneinander verschieden zu sein. Trotzdem werden sie "nur einmal gezählt". In

$$\mathbb{Q} \;:=\; \left\{ \frac{m}{n} \colon \; m,\, n \in \mathbb{Z},\; n \neq 0 \right\}$$

(der Menge der *rationalen Zahlen*) etwa kommt jede rationale Zahl in der Aufzählung sehr oft vor (z. B. $\frac{1}{2} = \frac{2}{4} = \frac{-7}{-14} \dots$). In der Menge $\mathbb{Q}$ dagegen kommt sie nur einmal vor. Auch spielt die Reihenfolge der Elemente keine Rolle. So ist etwa

$$\{1, 2\} \;=\; \{2, 1\} \;=\; \{2, 1, 1, 2, 1\},$$

und diese Menge enthält genau 2 Elemente.

Die Anzahl der (verschiedenen) Elemente in einer Menge X ist $|X|$, die *Mächtigkeit* (oder *Kardinalität*) von X. Dabei ist $|X|$ eine natürliche Zahl oder 0, oder (das Symbol) ∞. Eine Menge heißt *endlich*, falls $|X| \neq \infty$ ist, sonst *unendlich*. Die *leere Menge* $\varnothing$ ist die (einzige) Menge, die kein einziges Element enthält. Es ist also $|\varnothing| = 0$.

1.3 *Neue Mengen aus alten:* Aus gegebenen Mengen kann man auf verschiedene Weise neue bilden. Sind $X_1, \dots, X_n$ endlich viele Mengen (hier ist $n \in \mathbb{N}$), so ist

$$X_1 \cup \dots \cup X_n \;=\; \bigcup_{i=1}^{n} X_i \;=\; \{x \colon x \in X_1 \vee \dots \vee x \in X_n\}$$

die *Vereinigung* von $X_1, \dots, X_n$ und

$$X_1 \cap \dots \cap X_n \;=\; \bigcap_{i=1}^{n} X_i \;=\; \{x \colon x \in X_1 \wedge \dots \wedge x \in X_n\}$$

[2] die Menge $\mathbb{R}$ der reellen Zahlen ist zumindest intuitiv aus der Schule bekannt, und wird auch zu Beginn der Vorlesung Analysis I ausführlich eingeführt

der *Durchschnitt* von $X_1, \dots, X_n$. Solche Konstruktionen kann man durch Venn-Diagramme veranschaulichen. Die *Produktmenge* (auch genannt *kartesisches*[3] *Produkt*) von $X_1, \dots, X_n$ ist die Menge

$$X_1 \times \cdots \times X_n \;=\; \prod_{i=1}^{n} X_i \;=\; \{(x_1, \dots, x_n) \colon x_1 \in X_1, \dots, x_n \in X_n\}$$

aller *geordneten n-Tupel*, deren i-ter Eintrag in X_i liegt, für $i = 1, \dots, n$. Hierbei heißen zwei n-Tupel $(x_1, \dots, x_n)$ und $(y_1, \dots, y_n)$ gleich genau dann, wenn $x_i = y_i$ für alle $i = 1, \dots, n$ gilt. Bei Tupeln kommt es also, anders als bei Mengen, auf die Reihenfolge an! (Beachte die unterschiedliche Schreibweise, runde Klammern statt geschweifte.) Zum Beispiel hat die Produktmenge

$$\{1,2\} \times \{1,2,3\} \;=\; \Big\{(1,1),\, (1,2),\, (1,3),\, (2,1),\, (2,2),\, (2,3)\Big\}$$

genau 6 Elemente. Man schreibt auch $X^n := X \times \cdots \times X$ (mit n Faktoren). Zum Beispiel ist $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ die Menge aller (geordneten) Paare (x, y), wobei x, y reelle Zahlen sind, entsprechend $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ die Menge aller Tripel reeller Zahlen, usw.

Die Menge Y heißt eine *Teilmenge* der Menge X, i. Z. $Y \subseteq X$ oder $X \supseteq Y$, wenn jedes Element von Y auch Element von X ist, d.h. wenn gilt:

$$\forall\, y \quad \big(y \in Y \Rightarrow y \in X\big).$$

Hierbei ist Gleichheit $X = Y$ erlaubt. Will man Gleichheit ausschließen, so schreibt man $\subsetneq$:

$$Y \subsetneq X \quad :\Leftrightarrow \quad Y \subseteq X \;\wedge\; Y \neq X.$$

Für je zwei Mengen X, Y gilt also: $X = Y \Leftrightarrow X \subseteq Y \wedge Y \subseteq X$.

Sind X, Y Mengen, so heißt $X \setminus Y := \{x \colon\; x \in X,\, x \notin Y\}$ die *(mengentheoretische) Differenz* "X ohne Y". Für die Bildung von $X \setminus Y$ muß Y nicht notwendig eine Teilmenge von X sein. Ist jedoch $Y \subseteq X$, so heißt $X \setminus Y$ auch das *(relative) Komplement* von Y in X.

Die *Potenzmenge* $\mathcal{P}(X)$ von X ist die Menge aller Teilmengen von X. Also z.B. $\mathcal{P}(\varnothing) = \{\varnothing\}$ (das ist eine 1-elementige Menge!), $\mathcal{P}(\{1\}) = \{\varnothing,\, \{1\}\}$, $\mathcal{P}(\{1,2\}) = \{\varnothing,\, \{1\},\, \{2\},\, \{1,2\}\}$, usw. Man kann leicht beweisen: Ist $|X| = n < \infty$, so ist $|\mathcal{P}(X)| = 2^n$ (Aufgabe 4).

Es gibt eine Reihe von Regeln für die diversen mengentheoretischen Operationen. Viele davon sind unmittelbar klar, wie etwa $X \cup Y = Y \cup X$ oder $X \cap (Y \cap Z) = (X \cap Y) \cap Z$. Diese sollen hier nicht alle aufgezählt werden. Exemplarisch führen wir eine andere Identität ausführlich vor. Dabei sehen wir nebenbei, wie man generell die Gleichheit zweier Mengen beweist:

1.4 Lemma. *Seien $M_1, \dots, M_r$ und $N_1, \dots, N_s$ Mengen (mit $r, s \in \mathbb{N}$). Dann gelten die beiden* Distributivgesetze

$$\Big(\bigcap_{i=1}^{r} M_i\Big) \cup \Big(\bigcap_{j=1}^{s} N_j\Big) \;=\; \bigcap_{i=1}^{r} \bigcap_{j=1}^{s} (M_i \cup N_j).$$

[3] René Descartes (1596–1650)

und

$$\Big(\bigcup_{i=1}^{r} M_i\Big) \cap \Big(\bigcup_{j=1}^{s} N_j\Big) \;=\; \bigcup_{i=1}^{r}\bigcup_{j=1}^{s} (M_i \cap N_j).$$

BEWEIS. Wir beweisen die erste Gleichheit. Der Beweis der zweiten geht analog (führen Sie ihn zur Übung selbst durch!). Auf der rechten Seite der ersten Gleichheit steht ein Durchschnitt von $r{\cdot}s$ Mengen. Man könnte die rechte Menge auch schreiben als

$$\bigcap_{(i,j)\in\{1,\dots,r\}\times\{1,\dots,s\}} (M_i \cup N_j),$$

oder als

$$\bigcap_{\substack{i=1,\dots,r\\ j=1,\dots,s}} (M_i \cup N_j).$$

Eine Gleichheit $X = Y$ von zwei Mengen beweist man, indem man sowohl $X \subseteq Y$ als auch $Y \subseteq X$ zeigt.

"$\subseteq$": Sei $x \in (\bigcap_{i=1}^{r} M_i) \cup (\bigcap_{j=1}^{s} N_j)$. Fixiere $i \in \{1,\dots,r\}$ und $j \in \{1,\dots,s\}$, wir müssen zeigen $x \in M_i \cup N_j$. Ist $x \in M_1 \cap \dots \cap M_r$, so ist insbesondere $x \in M_i \subseteq M_i \cup N_j$, und wir sind schon fertig. Ist dagegen $x \notin M_1 \cap \dots \cap M_r$, so ist nach Voraussetzung $x \in N_1 \cap \dots \cap N_s$, also insbesondere $x \in N_j \subseteq M_i \cup N_j$. Für jedes Paar $(i,j) \in \{1,\dots,r\} \times \{1,\dots,s\}$ haben wir also gezeigt $x \in M_i \cup N_j$. Das bedeutet, wir haben bewiesen $x \in \bigcap_{i=1}^{r}\bigcap_{j=1}^{s}(M_i \cup N_j)$.

"$\supseteq$": Sei $x \in \bigcap_{i=1}^{r}\bigcap_{j=1}^{s}(M_i \cup N_j)$, wir müssen zeigen $x \in (\bigcap_{i=1}^{r} M_i) \cup (\bigcap_{j=1}^{s} N_j)$. Ist $x \in \bigcap_{i=1}^{r} M_i$, so sind wir fertig. Sei also $x \notin \bigcap_{i=1}^{r} M_i$. Behaupte, dann ist $x \in \bigcap_{j=1}^{s} N_j$. In der Tat, es gibt ein $i \in \{1,\dots,r\}$ mit $x \notin M_i$. Sei $j \in \{1,\dots,s\}$. Nach Voraussetzung ist $x \in M_i \cup N_j$. Wegen $x \notin M_i$ muß also $x \in N_j$ sein. Dies für alle j zeigt $x \in \bigcap_{j=1}^{s} N_j$. □

1.5 *Abbildungen:* Seien X, Y Mengen. Eine *Abbildung von X nach Y* ist eine Vorschrift, die jedem Element aus X ein Element aus Y zuordnet. Man schreibt sie typischerweise in der Form

$$f\colon X \to Y, \quad x \mapsto f(x) \quad (x \in X).$$

Das ist also die Abbildung, die jedem Element $x \in X$ das Element $f(x)$ zuordnet. Man nennt X die *Definitionsmenge* und Y die *Zielmenge* der Abbildung f.

Die Gesamtheit aller Abbildungen von X nach Y bildet selbst eine Menge. Sie wird mit $\mathrm{Abb}(X,Y)$ oder mit Y^X bezeichnet.

1.6 Beispiele.

1. Die Abbildung $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ $(x \in \mathbb{R})$, oder die Abbildung $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto e^{x-1} + \sin(x) - 5$, oder...

2. Ist X eine Menge, so haben wir zum Beispiel die Abbildung

$$\cap\colon \mathcal{P}(X) \times \mathcal{P}(X) \to \mathcal{P}(X), \quad (A,B) \mapsto A \cap B,$$

die jedem Paar (A,B) von Teilmengen von X den Durchschnitt $A \cap B$ zuordnet.

3. Abbildungen zwischen endlichen Mengen kann man durch Pfeildiagramme visualisieren.

4. Zwei Abbildungen $f\colon X \to Y$ und $g\colon X \to Y$ heißen gleich, i. Z. $f = g$, genau dann, wenn $f(x) = g(x)$ für alle $x \in X$ ist.

5. Für jede Menge X hat man die *identische Abbildung* $\mathrm{id}\colon X \to X$, $x \mapsto x$ ($x \in X$) von X. Allgemeiner hat man für jede Teilmenge $Y \subseteq X$ die *Inklusionsabbildung* $i : Y \to X$, $i(y) = y$ ($y \in Y$).

6. Sei $n \in \mathbb{N}$. Abbildungen $f\colon \{1, \dots, n\} \to X$ identifiziert man oft mit n-Tupeln von Elementen aus X, also mit Elementen von X^n: Man identifiziert $f\colon \{1, \dots, n\} \to X$ mit dem n-Tupel $(f(1), \dots, f(n))$. Analog identifiziert man Abbildungen $\mathbb{N} \to X$ oft mit unendlichen Folgen $(x_1, x_2, x_3, \dots)$ von Elementen aus X.

1.7 *Komposition:* Sind $f\colon X \to Y$ und $g\colon Y \to Z$ Abbildungen, so kann man sie *komponieren* (zusammensetzen). Nach Definition ist $g \circ f$ die Abbildung $X \to Z$, $x \mapsto g(f(x))$ ($x \in X$). *Vorsicht:* Die Komposition $g \circ f$ entsteht dadurch, daß man *zuerst f und dann g* ausführt (Reihenfolge!).

Für zwei Abbildungen $f_1\colon X_1 \to Y_1$ und $f_2\colon X_2 \to Y_2$ kann man die Komposition $f_2 \circ f_1\colon X_1 \to Y_2$ nur dann bilden, wenn $Y_1 = X_2$ (oder zumindest $Y_1 \subseteq X_2$) ist.

Das Komponieren von Abbildungen ist *assoziativ*: Ist $h\colon Z \to W$ eine dritte Abbildung, so gilt $(h \circ g) \circ f = h \circ (g \circ f)$. Beide sind die Abbildung $X \to W$, $x \mapsto h(g(f(x)))$ ($x \in X$).

1.8 Sei $f\colon X \to Y$ eine Abbildung. Ist $M \subseteq X$ eine Teilmenge, so heißt

$$f(M) := \{f(x)\colon\ x \in M\}$$

das *Bild* (oder die *Bildmenge*) von M unter f. Dies ist eine Teilmenge von Y. Ist $N \subseteq Y$ eine Teilmenge, so heißt

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

das *Urbild* (oder die *Urbildmenge*) von N; dies ist eine Teilmenge von X.

Ist $M \subseteq X$, so bezeichnet $f|_M\colon M \to Y$, $x \mapsto f(x)$ ($x \in M$) die *Restriktion* (oder *Einschränkung*) von f auf M. Im Vergleich zu f ist hier also nur die Definitionsmenge verkleinert worden.

1.9 Beispiel. Die Abbildung $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ hat als Bild die Teilmenge

$$f(\mathbb{R}) = \{x \in \mathbb{R}\colon x \geq 0\} = [0, \infty[$$

von $\mathbb{R}$. Für $y \in \mathbb{R}$ ist die Urbildmenge der einelementigen Menge $\{y\}$ gleich

$$f^{-1}(\{y\}) = \begin{cases} \varnothing & \text{falls } y < 0, \\ \{0\} & \text{falls } y = 0, \\ \{\sqrt{y}, -\sqrt{y}\} & \text{falls } y > 0 \end{cases}$$

Für die Mächtigkeit dieser Urbildmenge gilt also

$$\left|f^{-1}(\{y\})\right| = \begin{cases} 0 & \text{falls } y < 0, \\ 1 & \text{falls } y = 0, \\ 2 & \text{falls } y > 0. \end{cases}$$

Man kann diese Sachverhalte auch direkt am Graph von f ablesen.

1.10 Bemerkung. Der *Graph* einer Abbildung $f\colon X \to Y$ zwischen beliebigen Mengen ist definiert als die Teilmenge $\Gamma_f = \{(x, y) \in X \times Y\colon y = f(x)\}$ von $X \times Y$.

1.11 Die Abbildung $f\colon X \to Y$ heißt

(a) *injektiv*, wenn gilt:

$$\forall\, x_1,\, x_2 \in X\ \Big(f(x_1) = f(x_2)\ \Rightarrow\ x_1 = x_2\Big),$$

oder äquivalent, wenn gilt $|f^{-1}(\{y\})| \leq 1$ für jedes $y \in Y$;

(b) *surjektiv*, wenn gilt:

$$\forall\, y \in Y\ \exists\, x \in X\ f(x) = y,$$

oder äquivalent, wenn gilt $|f^{-1}(\{y\})| \geq 1$ für jedes $y \in Y$;

(c) *bijektiv*, wenn f injektiv und surjektiv ist, also wenn $|f^{-1}(\{y\})| = 1$ ist für jedes $y \in Y$.

Synonym verwendet man auch die Begriffe *Injektion, Surjektion, Bijektion.*

1.12 *Umkehrabbildung:* Sei jetzt $f\colon X \to Y$ eine bijektive Abbildung. Wir definieren (nur!) dann eine Abbildung $f^{-1}\colon Y \to X$ durch

$$f^{-1}(y)\ :=\ \text{das eindeutig bestimmte } x \in X \text{ mit } f(x) = y,$$

für $y \in Y$. Die so definierte Abbildung $f^{-1}\colon Y \to X$ heißt die *Umkehrabbildung* von f (oder die zu f *inverse Abbildung*). Es gilt $f^{-1} \circ f = \mathrm{id}_X$ und $f \circ f^{-1} = \mathrm{id}_Y$. Auch die Abbildung f^{-1} ist bijektiv, und $(f^{-1})^{-1} = f$. Beweisen Sie diese Aussagen!

Vorsicht: Die Notation für die Umkehrabbildung ist scheinbar dieselbe wie für die Urbildmenge. Aber das kann im Normalfall nicht zu Verwechslungen führen. Denn die Umkehrabbildung wird auf Elemente von Y angewandt und liefert Elemente von X. Die Operation Urbildmenge wird dagegen auf Teilmengen von Y angewandt und liefert Teilmengen von X. Insbesondere muß man für $y \in Y$ das Element $f^{-1}(y) \in X$ unterscheiden von der Teilmenge $f^{-1}(\{y\}) \subseteq X$. Beide hängen zusammen gemäß $f^{-1}(\{y\}) = \{f^{-1}(y)\}$.

1.13 Sei X eine Menge. Die in Beispiel 1.6.6 erwähnte Identifizierung von X^n mit $\mathrm{Abb}(\{1, \ldots, n\}, X)$ verallgemeinert sich wie folgt. Sei $I \neq \varnothing$ eine ("Index"-) Menge. Eine Abbildung $x\colon I \to X$ kann man auch als eine mit I indizierte *Familie* von Elementen aus X betrachten, oder als ein *I-Tupel* von Elementen aus X. Man schreibt x dann in der Form $(x_i)_{i\in I}$ oder $(x_i\colon i \in I)$, mit $x_i := x(i)$ für $i \in I$. Zwei I-Tupel $(x_i)_{i\in I}$ und $(y_i)_{i\in I}$ sind gleich genau dann, wenn gilt $x_i = y_i$ für alle $i \in I$ (das entspricht der Gleichheitsdefinition für Abbildungen).

Beachte, bei Familien kommt es auf die "Reihenfolge" der Elemente an, und Wiederholungen von Elementen werden unterschieden. Bei Mengen ist das nicht der Fall: Sei etwa $I = \{1, 2, 3, 4, 5\}$ und $X = \mathbb{N}$. Für

$$x\ =\ (2, 1, 1, 3, 2), \quad y\ =\ (2, 1, 3, 3, 2)\ \in X^I = \mathbb{N}^5$$

gilt $x \neq y$, denn $x_3 \neq y_3$. Für die beiden Teilmengen

$$A\ =\ \{2, 1, 1, 3, 2\}, \quad B\ =\ \{2, 1, 3, 3, 2\}\ \subseteq\ X$$

von X gilt dagegen $A = B = \{1, 2, 3\}$.

1.14 Bemerkung. Einige Mengenoperationen verallgemeinern sich von endlich vielen auf beliebig viele Mengen. Sei weiter $I \neq \varnothing$ eine (Index-) Menge, und sei

X eine Menge. Für jedes $i \in I$ sei eine Teilmenge $X_i \subseteq X$ gegeben. Dann heißt

$$\bigcap_{i\in I} X_i \;=\; \{x \in X\colon\ \forall\, i \in I\ x \in X_i\}$$

der Durchschnitt der X_i und

$$\bigcup_{i\in I} X_i \;=\; \{x \in X\colon\ \exists\, i \in I\ x \in X_i\}$$

die Vereinigung der X_i. Das kartesische Produkt $\prod_{i\in I} X_i$ ist definiert als Menge aller Abbildungen $x\colon I \to X$, $i \mapsto x_i$ $(i \in I)$ mit $x_i \in X_i$ für alle $i \in I$. Es ist also $\prod_{i\in I} X_i$ die Menge aller I-Tupel $(x_i)_{i\in I}$ mit $x_i \in X_i$ für alle $i \in I$.

2. Gruppen

Der Begriff der Gruppe zählt zu den elementarsten und zugleich wichtigsten algebraischen Strukturen in der Mathematik.

2.1 Definition. Eine *Gruppe* ist ein Paar $(G, \circ)$, bestehend aus einer Menge G und einer Abbildung

$$\circ\colon\ G \times G \to G, \quad (a, b) \mapsto a \circ b,$$

der *Gruppenverknüpfung*, so daß die folgenden Eigenschaften erfüllt sind:

(G1) Die Verknüpfung ist *assoziativ*:

$$\forall\, a, b, c \in G \quad (a \circ b) \circ c \;=\; a \circ (b \circ c);$$

(G2) es gibt ein Element $e \in G$, so daß gelten:
  (G2.1) $\forall\, a \in G \;\; e \circ a = a$;
  (G2.2) $\forall\, a \in G \;\; \exists\, a' \in G \;\; a' \circ a = e$.

Dabei heißt $a \circ b$ das *Produkt* von a und b. Ein Element e wie in (G2) heißt ein *neutrales Element* von G (wir werden gleich sehen, daß es nur eines gibt). Ein Element a' wie in (G2.2) heißt ein *inverses Element* zu a (auch hier wird die Eindeutigkeit gleich gezeigt).

2.2 Definition. Ist $(G, \circ)$ eine Gruppe, und gilt darüber hinaus $a \circ b = b \circ a$ für alle $a, b \in G$, so heißt die Gruppe $(G, \circ)$ *abelsch*[4] (oder *kommutativ*).

Hier sind erste Folgerungen aus den Gruppenaxiomen:

2.3 Lemma. *Sei $(G, \circ)$ eine Gruppe.*

(a) *Das neutrale Element e aus (G2) ist eindeutig bestimmt, und erfüllt auch $a \circ e = a$ für alle $a \in G$.*
(b) *Für jedes $a \in G$ ist das Element a' aus (G2.2) eindeutig bestimmt, und erfüllt auch $a \circ a' = e$.*

Beweis. Sei $e \in G$ ein Element, welches (G2) erfüllt. Wir zeigen zunächst:

$$\text{Sind } a, b \in G \text{ mit } a \circ b = e, \text{ so ist auch } b \circ a = e. \qquad (*)$$

[4]Niels Henrik Abel (1802–1829)

In der Tat, es gibt $a' \in G$ mit $a' \circ a = e$ (G2.2), und es folgt

$$\begin{aligned} b \circ a &= e \circ (b \circ a) = (a' \circ a) \circ (b \circ a) = a' \circ (a \circ (b \circ a)) \\ &= a' \circ ((a \circ b) \circ a) = a' \circ (e \circ a) = a' \circ a = e, \end{aligned}$$

womit $(*)$ bewiesen ist.

Nun zeigen wir (a). Dazu sei $f \in G$ ein Element mit $f \circ a = a$ für alle $a \in G$. Insbesondere (nimm $a := e$) ist $f \circ e = e$. Nach $(*)$ folgt daraus $e \circ f = e$. Andererseits ist $e \circ f = f$ (G2.1), also folgt $f = e$. Für $a \in G$ (und a' wie in (G2.2)) ist außerdem $a \circ e = a \circ (a' \circ a) = (a \circ a') \circ a =_{(*)} e \circ a = a$. Damit ist (a) gezeigt.

Für den Beweis von (b) sei $a \in G$, und seien a', $b \in G$ mit $a' \circ a = e = b \circ a$. Wir müssen $b = a'$ zeigen. Multipliziere die Gleichung $e = b \circ a$ von rechts mit a', das ergibt

$$a' = (b \circ a) \circ a' = b \circ (a \circ a') =_{(*)} b \circ e =_{(\mathrm{a})} b.$$

Das Lemma ist damit bewiesen. □

Wegen Lemma 2.3 können wir in einer Gruppe $(G, \circ)$ von *dem* neutralen Element e und *dem* zu $a \in G$ inversen Element a' sprechen.

2.4 Beispiele.

1. $(\mathbb{R}, +)$ ist eine Gruppe mit neutralem Element $e = 0$. Das zu $a \in \mathbb{R}$ inverse Element ist $a' = -a$, denn $(-a) + a = 0$. Dasselbe gilt für $(\mathbb{Q}, +)$, die additive Gruppe der *rationalen Zahlen*, und für $(\mathbb{Z}, +)$, die additive Gruppe der *ganzen Zahlen*. Alle diese Gruppen sind abelsch, denn $a + b = b + a$ für alle a, $b \in \mathbb{R}$.

2. Sei $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$. Dann ist $(\mathbb{R}^*, \cdot)$ eine abelsche Gruppe. Das neutrale Element ist $e = 1$, das zu $a \in \mathbb{R}^*$ inverse Element ist $a' = \frac{1}{a}$. Dagegen ist $(\mathbb{R}, \cdot)$ keine Gruppe (warum nicht?). Weiter sind auch $\mathbb{R}_+^* := \{x \in \mathbb{R} \colon x > 0\}$, $\mathbb{Q}^* := \mathbb{Q} \setminus \{0\}$ und $\mathbb{Q}_+^* := \{x \in \mathbb{Q} \colon x > 0\}$ abelsche Gruppen bezüglich Multiplikation.

3. Ist G eine endliche Menge, etwa $G = \{g_1, \ldots, g_n\}$, so kann man eine Verknüpfung (Abbildung) $G \times G \to G$, $(g, h) \mapsto g \circ h$ wie folgt in einer Tabelle zusammenfassen, der *Verknüpfungstafel*:

	g_1	g_2	$\cdots$	g_n
g_1	$g_1 \circ g_1$	$g_1 \circ g_2$	$\cdots$	$g_1 \circ g_n$
g_2	$g_2 \circ g_1$	$g_2 \circ g_2$	$\cdots$	$g_2 \circ g_n$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
g_n	$g_n \circ g_1$	$g_n \circ g_2$	$\cdots$	$g_n \circ g_n$

Zum Beispiel ist

	a	b	c
a	a	b	c
b	b	c	a
c	c	a	b

die Verknüpfungstafel einer Gruppe $G = \{a, b, c\}$ mit neutralem Element a, in der $b \circ b = c$ und $b \circ c = a$ gilt. Daß die Gruppenaxiome erfüllt sind, kann man in diesem Beispiel nachprüfen, aber der Beweis der Assoziativität ist schon etwas langwierig. Sobald $|G|$ größer ist, wird die Sache schnell unhandlich. Verknüpfungstafeln sind deshalb nicht wirklich geeignet, um Gruppen zu definieren oder zu studieren.

4. Wenn die Gruppenverknüpfung $\circ$ klar ist, erwähnt man sie meistens gar nicht und sagt einfach "G ist eine Gruppe", statt korrekter "$(G, \circ)$ ist eine Gruppe".

5. Wir geben jetzt Beispiele von nichtabelschen Gruppen. Sei $M \neq \varnothing$ eine Menge. Mit $\mathrm{Sym}(M)$ bezeichnet man die Menge aller bijektiven Abbildungen $\sigma\colon M \to M$. Für $\sigma, \tau \in \mathrm{Sym}(M)$ ist auch die Komposition $\sigma \circ \tau \in \mathrm{Sym}(M)$. Behaupte, $(\mathrm{Sym}(M), \circ)$ ist eine Gruppe mit neutralem Element $e = \mathrm{id}_M$ (identische Abbildung von M) und mit zu σ inversem Element $\sigma' = \sigma^{-1}$ (Umkehrabbildung von σ, 1.12). In der Tat, das Assoziativgesetz gilt für Komposition von Abbildungen ganz allgemein (1.7), und für alle $\sigma \in \mathrm{Sym}(M)$ gilt $\mathrm{id}_M \circ \sigma = \sigma$ und $\sigma^{-1} \circ \sigma = \mathrm{id}_M$.

Die Gruppe $\mathrm{Sym}(M)$ heißt die *symmetrische Gruppe von M*. Die Elemente von $\mathrm{Sym}(M)$ heißen *Permutationen* (oder *Vertauschungen*) der Menge M. Ist $M = \{1, \dots, n\}$ (mit $n \in \mathbb{N}$), so schreibt man kurz $S_n := \mathrm{Sym}(\{1, \dots, n\})$, die *symmetrische Gruppe von $\{1, \dots, n\}$*.

Eine Permutation $\sigma \in S_n$ notiert man symbolisch in der Form

$$\sigma \;=\; \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix}.$$

Mit dieser Notation kann man leicht die Komposition von Permutationen berechnen, z.B. gilt

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 1 & 5 \end{pmatrix}$$

in der Gruppe S_5.

2.5 Satz. *Sei S_n die symmetrische Gruppe von $\{1, \dots, n\}$ (mit $n \in \mathbb{N}$).*

(a) *Es ist $|S_n| = n! := 1 \cdot 2 \cdots n$ (sprich "n Fakultät").*

(b) *Ist $n \geq 3$, so ist die Gruppe S_n nicht abelsch.*

Beweis. (a) Wir zählen die Elemente von S_n. Um $\sigma \in S_n$ anzugeben, gibt es n Möglichkeiten für $\sigma(1)$, dann verbleiben noch $n-1$ für $\sigma(2)$, dann noch $n-2$ für $\sigma(3)$, usw., schließlich nur noch eine Möglichkeit für $\sigma(n)$. Für σ gibt es also genau $n(n-1)\cdots 2 \cdot 1 = n!$ Möglichkeiten.

(b) Sei $n \geq 3$. Betrachte die beiden wie folgt definierten Permutationen σ und τ in S_n:

$$\sigma(i) \;:=\; \begin{cases} 2 & i = 1, \\ 1 & i = 2, \\ i & 3 \leq i \leq n \end{cases}$$

und

$$\tau(i) \;:=\; \begin{cases} i+1 & 1 \leq i \leq n-1, \\ 1 & i = n \end{cases}$$

$(i = 1, \dots, n)$. In der zuvor eingeführten Schreibweise ist also

$$\sigma \;:=\; \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ 2 & 1 & 3 & \cdots & n \end{pmatrix}, \quad \tau \;:=\; \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ 2 & 3 & \cdots & n & 1 \end{pmatrix}.$$

Dann ist $\sigma \circ \tau(1) = 1$, aber $\tau \circ \sigma(1) = 3$. Also ist $\sigma \circ \tau \neq \tau \circ \sigma$. □

2.6 *Konvention:* In Zukunft schreiben wir allgemeine Gruppen multiplikativ als $(G, \cdot)$. Den Multiplikationspunkt läßt man dabei meistens weg und schreibt ab statt $a \cdot b$. Das zu $a \in G$ inverse Element wird mit a^{-1} bezeichnet, statt wie bisher mit a'. Es gilt also $ea = ae = a$ und $aa^{-1} = a^{-1}a = e$ für alle $a \in G$. Außerdem gelten folgende Rechenregeln:

2.7 Lemma. *Sei $G = (G, \cdot)$ eine Gruppe, seien $a, b \in G$.*
(a) $(a^{-1})^{-1} = a$.
(b) $(ab)^{-1} = b^{-1}a^{-1}$.

Beweis. Nach Lemma 2.3 ist $aa^{-1} = e$. Das bedeutet $(a^{-1})^{-1} = a$. Weiter ist

$$(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}(ab)) = b^{-1}((a^{-1}a)b) = b^{-1}(eb) = b^{-1}b = e,$$

und das bedeutet $(ab)^{-1} = b^{-1}a^{-1}$. □

Bisher haben wir bei Produkten in Gruppen immer sorgfältig Klammern gesetzt, um eindeutig zu machen, in welcher Reihenfolge das Produkt berechnet wird. Als Folge des Assoziativgesetzes zeigen wir jetzt, daß man sich das Setzen von Klammern generell sparen kann:

2.8 Satz. *Sei G eine Gruppe, seien $x_1, \ldots, x_n \in G$ ($n \in \mathbb{N}$). Dann ist das Produkt $x_1 \cdots x_n$ von der Klammerung unabhängig.*

Beweis. Vollständige Induktion nach der Anzahl n der Faktoren. Für $n \leq 2$ ist nichts zu zeigen, und für $n = 3$ gilt die Behauptung nach Axiom (G1). Sei $n > 3$, und sei der Satz für alle kleineren Werte von n schon bewiesen. Seien zwei Klammerungen des Produkts $x_1 \cdots x_n$ gegeben, und seien g, h die entsprechenden Produkte. Betrachte die Klammern auf der äußersten Ebene: Dann gibt es Indices $1 \leq i, j < n$, so daß $g = (x_1 \cdots x_i) \cdot (x_{i+1} \cdots x_n)$ und $h = (x_1 \cdots x_j) \cdot (x_{j+1} \cdots x_n)$ ist, mit jeweils geeigneten Klammerungen der Teilprodukte. Da diese weniger als n Faktoren haben, ist nach Induktionsvoraussetzung ihre Klammerung irrelevant. Ist $i = j$, so sind wir fertig. Wir können also $i < j$ annehmen. Dann ist (unter weiterer Benutzung der Induktionsvoraussetzung)

$$\begin{aligned} g &= (x_1 \cdots x_i) \cdot \big((x_{i+1} \cdots x_j) \cdot (x_{j+1} \cdots x_n)\big) \\ &= \big((x_1 \cdots x_i) \cdot (x_{i+1} \cdots x_j)\big) \cdot (x_{j+1} \cdots x_n) = h, \end{aligned}$$

und der Beweis ist fertig. □

Wegen Satz 2.8 werden wir in Zukunft bei Produkten in Gruppen meistens keine Klammern mehr setzen.

Eine neue Charakterisierung von Gruppen ergibt sich aus folgendem Lemma:

2.9 Lemma. *Sei G eine Menge, sei $\cdot\colon G \times G \to G$, $(a, b) \mapsto a \cdot b = ab$ eine Abbildung. Genau dann ist $(G, \cdot)$ eine Gruppe, wenn $G \neq \varnothing$ ist, das Assoziativgesetz (G1) gilt (siehe 2.1), und außerdem gilt: Für beliebige $a, b \in G$ gibt es genau ein $x \in G$ mit $ax = b$, und genau ein $y \in G$ mit $ya = b$.*

Beweis. "Nur dann": Sei G eine Gruppe. Dann ist $G \neq \varnothing$ (G2). Seien a, $b \in G$. Für $x := a^{-1}b$ gilt $ax = b$, und für $y := ba^{-1}$ gilt $ya = b$. Ist auch $\tilde{x} \in G$ mit $a\tilde{x} = b$, so folgt $\tilde{x} = a^{-1}a\tilde{x} = a^{-1}b = x$. Analog zeigt man die Eindeutigkeit von y.

"Dann": Für die Umkehrung starte mit einem beliebigen Element $g \in G$ (gibt es wegen $G \neq \varnothing$). Sei $e \in G$ das (nach Voraussetzung eindeutig existierende) Element mit $ge = g$. Dann ist e ein neutrales Element: Denn für $a \in G$ ist $ga = (ge)a = g(ea)$, und aus der Eindeutigkeit der Lösung x von $gx = ga$ folgt $a = ea$. Nach Voraussetzung gibt es auch zu jedem $a \in G$ (genau) ein $a' \in G$ mit $a'a = e$. Die Gruppenaxiome sind damit bewiesen. □

2.10 Korollar. *Ist G eine Gruppe, so gilt für alle a, x, $y \in G$:*

$$ax = ay \;\Rightarrow\; x = y \quad \textit{und} \quad xa = ya \;\Rightarrow\; x = y. \qquad \square$$

2.11 Bemerkung. 2.10 sagt, daß man Gleichungen in einer Gruppe kürzen darf. Aber Vorsicht: Man darf ein Element nur kürzen, wenn es auf beiden Seiten der Gleichung auf derselben Seite steht! Aus $ax = ya$ folgt im allgemeinen keineswegs $x = y$.

2.12 Definition. Sei $(G, \cdot)$ eine Gruppe. Eine Teilmenge H von G heißt eine *Untergruppe* von G, falls gilt:

(UG1) $\forall\, a, b \in H \quad ab \in H$;
(UG2) $(H, \cdot)$ ist selbst eine Gruppe.

Dabei ist in (UG2) die Verknüpfung $\cdot\colon H \times H \to H$ die nach (UG1) wohldefinierte Einschränkung der Gruppenmultiplikation von G.

2.13 Satz. *Sei G eine Gruppe und H eine Teilmenge von G. Genau dann ist H eine Untergruppe von G, wenn gelten:*

(1) $H \neq \varnothing$;
(2) $\forall\, a, b \in H \quad ab^{-1} \in H$.

(In (2) werden Inverses b^{-1} und Produkt ab^{-1} bezüglich der Gruppenstruktur von G gebildet.)

Beweis. Sei e das neutrale Element von G. Zunächst sei H Untergruppe von G, sei f das neutrale Element von H. Dann gilt $ef = f$ (wegen e neutral in G) und $ff = f$ (wegen f neutral in H). Kürzen in G (2.10) gibt $e = f$. Sei $a \in H$, sei a^{-1} bzw. c das inverse Element von a in G bzw. in H. Dann gilt $a^{-1}a = e = f = ca$, und Kürzen in G gibt $a^{-1} = c$. Aus $a \in H$ folgt also $a^{-1} \in H$. Damit ist (2) gezeigt.

Umgekehrt gelte (1) und (2). Wähle ein $h \in H$. Aus (2) folgt $hh^{-1} = e \in H$. Erneutes Anwenden von (2) gibt für jedes $b \in H$: $eb^{-1} = b^{-1} \in H$. Für alle a, $b \in H$ ist nach (2) also auch $ab = a(b^{-1})^{-1} \in H$. Jetzt ist klar, daß H eine Untergruppe von G ist. $\square$

2.14 Bemerkungen.

1. Der Beweis hat gezeigt: Ist H Untergruppe von G, so haben G und H dasselbe neutrale Element, und inverse Elemente von Elementen aus H sind in G und H dieselben.

2. Ist G Gruppe, so sind $\{e\}$ und G Untergruppen von G. Dabei heißt $\{e\}$ die *triviale Untergruppe*. Ist H Untergruppe von G und K Untergruppe von H, so ist auch K Untergruppe von G (Transitivität).

3. Für $m \in \mathbb{Z}$ sei $m\mathbb{Z} := \{ma\colon a \in \mathbb{Z}\}$. Es ist $m\mathbb{Z} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$ eine Kette von Untergruppen bezüglich Addition, und $\mathbb{Q}_+^* \subseteq \mathbb{Q}^* \subseteq \mathbb{R}^*$ ist eine Kette von Untergruppen bezüglich Multiplikation.

4. Sei $n \in \mathbb{N}$, sei S_n die symmetrische Gruppe. Für jedes $i \in \{1, \ldots, n\}$ ist $H_i := \{\sigma \in S_n\colon \sigma(i) = i\}$ eine Untergruppe von S_n. Weitere Beispiele und Nichtbeispiele von Untergruppen gibt es in Aufgabe 7.

2.15 Korollar. *Sei G eine Gruppe und $(H_i)_{i\in I}$ eine Familie von Untergruppen von G. Dann ist auch $\bigcap_{i\in I} H_i$ eine Untergruppe von G.*

Beweis. Sei $H := \bigcap_{i\in I} H_i$. Für alle $i \in I$ ist $e \in H_i$ (2.14.1), also ist $e \in H$. Damit ist $H \neq \varnothing$. Bedingung (2) aus Satz 2.13 ist für alle H_i erfüllt, also auch für H. Die Aussage folgt also aus Satz 2.13. □

Wir können deshalb definieren:

2.16 Definition. Sei G eine Gruppe und $X \subseteq G$ eine Teilmenge. Dann ist

$$\langle X\rangle := \bigcap\{H\colon\ H \text{ Untergruppe von } G \text{ mit } X \subseteq H\}$$

eine Untergruppe von G, und heißt die *von X erzeugte Untergruppe*.

Nach Definition ist $\langle X\rangle$ die (eindeutig bestimmte) kleinste Untergruppe von G, welche X enthält. Für weitere Beispiele von Untergruppen führen wir folgende Notation ein:

2.17 Definition. Sei $(G,\cdot)$ eine Gruppe, sei $a \in G$. Für $n \in \mathbb{N}$ schreibt man $a^n := a\cdots a$ (n Faktoren) und $a^{-n} := (a^{-1})^n$. Außerdem setzt man $a^0 := e$ (neutrales Element).

2.18 Lemma. *Für $a \in G$ und alle $m, n \in \mathbb{Z}$ gilt $a^m \cdot a^n = a^{m+n}$. Insbesondere ist $(a^n)^{-1} = a^{-n}$.*

Beweis. Einfache Übung (man muß mehrere Fälle unterscheiden). □

2.19 Beispiel. Sei $(G,\cdot)$ eine Gruppe. Für $a \in G$ ist $\langle\{a\}\rangle = \{a^n\colon n \in \mathbb{Z}\}$, nach Lemma 2.18. Man schreibt für diese Untegruppe auch einfach $\langle a\rangle$.

2.20 Bemerkung. Wir haben gesehen, daß die Verknüpfung in einer Gruppe verschieden notiert sein kann. Häufig sind vor allem $\cdot$ (multiplikative Notation) und $+$ (additive Notation), daneben auch noch $\circ$ oder $*$. Die additive Notation $+$ ist nur bei abelschen Gruppen erlaubt!

Wie erwähnt, werden allgemeine Gruppen G stets multiplikativ geschrieben, mit neutralem Element e oder 1. Ist $(G,+)$ eine additiv geschriebene (abelsche) Gruppe, so bezeichnet man das neutrale Element mit 0 und das zu $a \in G$ inverse Element mit $-a$. Statt $a+(-b)$ schreibt man einfach $a-b$, statt $a+\cdots+a$ (n-mal) schreibt man na, und schreibt $(-n)a := n(-a) = -(na)$ ($n \in \mathbb{N}$, a, $b \in G$).

Hier eine Synopse von multiplikativer und additiver Notation:

	$(G,\cdot)$	$(G,+)$
Verknüpfung	ab	$a+b$
inverses Element	a^{-1}	$-a$
neutrales Element	e, oder 1	0
Potenz/Vielfaches	a^n	na
	ab^{-1}	$a-b$
Regeln (2.7)	$(a^{-1})^{-1} = a$	$-(-a) = a$
	$(ab)^{-1} = b^{-1}a^{-1}$	$-(a+b) = -a-b$

3. Ringe und Körper

3.1 Definition. Ein *Ring* ist ein Tripel $(A, +, \cdot)$, bestehend aus einer Menge A und zwei Verknüpfungen $+\colon A \times A \to A$ (*Addition*) bzw. $\cdot\colon A \times A \to A$ (*Multiplikation*) derart, daß die folgenden Eigenschaften erfüllt sind:

(R1) $(A, +)$ ist eine abelsche Gruppe;
(R2) $\forall\, a, b, c \in A \;\; (a \cdot b) \cdot c = a \cdot (b \cdot c)$ (Assoziativität der Multiplikation);
(R3) $\forall\, a, b, c \in A \;\; a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ und $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$ (Distributivgesetze);
(R4) $\exists\, e \in A \;\; \forall\, a \in A \;\; e \cdot a = a \cdot e = a$ (Einselement).

Der Ring heißt *kommutativ*, wenn zusätzlich gilt:

(R5) $\forall\, a, b \in A \;\; a \cdot b = b \cdot a$.

3.2 *Konventionen*: $\cdot$ bindet stärker als $+$. Den Malpunkt läßt man meist weg. Das erste Gesetz in (R3) schreibt sich also $a(b+c) = ab+ac$. Das neutrale Element bezüglich $+$ wird mit 0 bezeichnet, das bezüglich $+$ zu $a \in A$ inverse Element mit $-a$.

3.3 Bemerkungen.

1. Für alle $a, b \in A$ gilt $0 \cdot a = a \cdot 0 = 0$ und $-(ab) = (-a)b = a(-b)$. (Denn $0 \cdot a + 0 \cdot a = (0+0)a = 0 \cdot a = 0 \cdot a + 0$, und Kürzen in $(A, +)$ gibt $0 \cdot a = 0$. Analog $a \cdot 0 = 0$. Die zweite Behauptung folgt damit aus $0 = (a + (-a))b = ab + (-a)b$ und $0 = a(b + (-b)) = ab + a(-b)$.)

2. Für $a_1, \ldots, a_n \in A$ ($n \in \mathbb{N}$) hängen Summe $a_1 + \cdots + a_n$ und Produkt $a_1 \cdots a_n$ nicht von der Klammerung ab. (Der in 2.8 gegebene Beweis hat nur das Assoziativgesetz benutzt und keine anderen Eigenschaften einer Gruppe, überträgt sich also auch auf Produkte in Ringen.) Man schreibt $a^n := a \cdots a$ (n Faktoren) für $n \in \mathbb{N}$, $a \in A$.

3. Das in (R4) geforderte Element $e \in A$ ist eindeutig bestimmt. Man nennt e das *Einselement*, oder die *Eins*, von A, und schreibt meist $1 := e$, entsprechend auch $n := n \cdot 1 \in A$ für $n \in \mathbb{Z}$. (Meistens geht aus dem Kontext hervor, daß diese Elemente in A aufzufassen sind. Falls nicht, muß man es dazu sagen.)

3.4 Beispiele.

1. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, jeweils mit der üblichen Addition und Multiplikation, sind kommutative Ringe.

2. $A = \{0\}$ (mit der einzig möglichen Addition und Multiplikation) ist ein kommutativer Ring, genannt der *Nullring*. Hier ist $0 = 1$. In jedem Ring $A \neq \{0\}$ ist $1 \neq 0$.

3. Sei X eine Menge und $\mathcal{P}(X)$ ihre Potenzmenge. Für $A, B \in \mathcal{P}(X)$ sei $A+B := (A \cup B) \smallsetminus (A \cap B)$. Dann ist $\big(\mathcal{P}(X),\, +,\, \cap\big)$ ein kommutativer Ring, siehe Aufgabe 9.

Jetzt arbeiten wir im Ring $\mathbb{Z}$.

3.5 Satz. (Division mit Rest) *Seien $a, b \in \mathbb{Z}$, $b \neq 0$. Dann existieren $q, r \in \mathbb{Z}$ mit $a = qb + r$ und $0 \leq r < |b|$. Dadurch sind q und r eindeutig bestimmt.*

BEWEIS. Für den Beweis der Existenz können wir $b > 0$ annehmen. Setze $q := \lfloor \frac{a}{b} \rfloor := \max\{n \in \mathbb{Z}\colon n \le \frac{a}{b}\}$ (*ganzzahliger Anteil* von $\frac{a}{b}$), dann ist $0 \le \frac{a}{b} - q < 1$, und Multiplikation mit b gibt $0 \le a - qb < b$. Wir können also $r = a - qb$ nehmen. Eindeutigkeit: Ist $a = qb + r = q'b + r'$ mit q, q', r, $r' \in \mathbb{Z}$ und $0 \le r$, $r' < |b|$, so folgt $(q - q')b = r' - r$ und $|r' - r| < |b|$. Also muß $r' - r = 0$ sein, und damit auch $q - q' = 0$. □

3.6 Definition.

(a) Seien a, $b \in \mathbb{Z}$. Man sagt a *teilt* b (oder a ist ein *Teiler* von b), und schreibt $a \mid b$, wenn es $c \in \mathbb{Z}$ gibt mit $b = ac$. Ist a kein Teiler von b, so schreibt man $a \nmid b$.

(b) $p \in \mathbb{N}$ heißt *Primzahl*, wenn $p > 1$ ist und p außer 1 und p keine Teiler in $\mathbb{N}$ hat.

3.7 Beispiel. (Wichtig!) Fixiere $n \in \mathbb{N}$. Wir definieren den Ring $\mathbb{Z}/n\mathbb{Z}$ der *ganzen Zahlen modulo* n. Für a, $b \in \mathbb{Z}$ schreiben wir dazu

$$a \equiv b \pmod{n},$$

(sprich: a *ist kongruent zu* b *modulo* n), falls $n \mid (b - a)$ gilt, und setzen

$$\overline{a} := a + n\mathbb{Z} := \{a + kn\colon k \in \mathbb{Z}\} = \{b \in \mathbb{Z}\colon b \equiv a \pmod{n}\}.$$

Es ist also $\overline{a}$ die Menge aller ganzen Zahlen, die bei Division durch n denselben Rest lassen wie a. Für a, $b \in \mathbb{Z}$ gilt nach Definition $\overline{a} = \overline{b} \Leftrightarrow a \equiv b \pmod{n}$. Man nennt $\overline{a}$ die *Restklasse von* a *modulo* n. Wir definieren

$$\mathbb{Z}/n\mathbb{Z} := \{\overline{a}\colon a \in \mathbb{Z}\}.$$

Nach 3.5 ist $\mathbb{Z}/n\mathbb{Z} = \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\}$, es ist also $|\mathbb{Z}/n\mathbb{Z}| = n$. Wir wollen Addition und Multiplikation auf $\mathbb{Z}/n\mathbb{Z}$ definieren durch

$$\overline{a} + \overline{b} := \overline{a + b}, \quad \overline{a} \cdot \overline{b} := \overline{ab} \quad (a, b \in \mathbb{Z}).$$

Dabei müssen wir jeweils zeigen, daß die rechte Seite *wohldefiniert* ist, also nicht von der Auswahl der Vertreter aus den Restklassen $\overline{a}$ und $\overline{b}$ abhängt. Für a, a', b, $b' \in \mathbb{Z}$ mit $a \equiv a' \pmod{n}$ und $b \equiv b' \pmod{n}$ müssen wir also zeigen, daß auch

$$a + b \equiv a' + b' \pmod{n}, \quad ab \equiv a'b' \pmod{n}$$

gilt. Dafür bemerken wir, daß es $r, s \in \mathbb{Z}$ gibt mit $a' = a + rn$ und $b' = b + sn$. Es folgt $a' + b' = (a + b) + (r + s)n$, also $\overline{a' + b'} = \overline{a + b}$, und ebenso $a'b' = ab + (as + br + rsn)n$, also $\overline{a'b'} = \overline{ab}$.

Damit sind Verknüpfungen $+$ und $\cdot$ auf $\mathbb{Z}/n\mathbb{Z}$ definiert. Man verifiziert ohne Problem, daß $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ ein kommutativer Ring ist, mit Null $\overline{0}$ und Eins $\overline{1}$, und mit $-\overline{a} = \overline{-a}$ für $a \in \mathbb{Z}$.

Bevor wir dieses und andere Beispiele genauer analysieren, müssen wir mehr über Teilbarkeit wissen:

3.8 Lemma. *Sei $a \in \mathbb{Z}$, sei p eine Primzahl, und es gelte $p \nmid a$. Dann gibt es m, $n \in \mathbb{Z}$ mit $am + pn = 1$.*

BEWEIS. Sei k die kleinste natürliche Zahl der Form $k = am + pn$ mit m, $n \in \mathbb{Z}$. Wir wollen zeigen $k = 1$. Division von a durch p mit Rest gibt $a = pq + r$ mit q, $r \in \mathbb{Z}$

und $0 \le r < p$. Wegen $p \nmid a$ ist dabei $r \ge 1$, und wegen $r = -a + pq$ sehen wir $k \le r < p$. Angenommen $k \ge 2$. Dividiere p durch k mit Rest, das gibt $p = kb + l$ mit $b, l \in \mathbb{Z}$ und $0 \le l < k$. Wegen $1 < k < p$ und p Primzahl ist $k \nmid p$, also folgt $l \ge 1$. Und wegen $l = p - kb = p - (am + pn)b = -abm + p(1 - bn)$ erhalten wir einen Widerspruch zur Minimalität von k. Also war die Annahme falsch, d.h. es ist $k = 1$. □

3.9 Satz. *Ist p eine Primzahl, sind a, $b \in \mathbb{Z}$, und gilt $p \mid ab$, so folgt $p \mid a$ oder $p \mid b$.*

Beweis. Angenommen falsch. Nach Lemma 3.8 können wir schreiben $1 = am + pn = br + ps$ mit $m, n, r, s \in \mathbb{Z}$. Es folgt

$$1 \;=\; 1 \cdot 1 \;=\; (am + pn)(br + ps) \;=\; ab \cdot mr + p(ams + bnr + pns).$$

Nach Voraussetzung ist ab durch p teilbar, also auch die ganze rechte Seite, Widerspruch. □

Induktiv sieht man, daß der Satz auch für mehr als zwei Faktoren gilt: Ist p Primzahl und gilt $p \mid a_1 \cdots a_n$ mit $a_1, \ldots, a_n \in \mathbb{Z}$, so folgt $p \mid a_i$ für mindestens ein $i \in \{1, \ldots, n\}$. Wir erhalten damit:

3.10 Theorem. (Fundamentalsatz der elementaren Zahlentheorie) *Jede natürliche Zahl $n > 1$ ist Produkt von endlich vielen Primzahlen, und die Produktzerlegung ist eindeutig bis auf die Reihenfolge der Faktoren.*

Die Eindeutigkeitsaussage besagt genauer: Ist $n = p_1 \cdots p_r = q_1 \cdots q_s$ mit r, $s \in \mathbb{N}$ und Primzahlen p_i, q_j, so gilt $r = s$, und es gibt eine Permutation $\sigma \in S_r$ mit $q_i = p_{\sigma(i)}$ für $i = 1, \ldots, r$.

Beweis. Wir zeigen zunächst durch vollständige Induktion, daß n ein Produkt von Primzahlen ist. Der Beginn $n = 2$ ist richtig, sei also $n > 2$. Ist n selbst eine Primzahl, so sind wir fertig. Andernfalls ist $n = n_1 n_2$ mit natürlichen Zahlen $1 < n_1, n_2 < n$. Nach Induktionsannahme sind n_1, n_2 Produkte von Primzahlen, also gilt dasselbe für n.

Für die Eindeutigkeit sei $n = p_1 \cdots p_r = q_1 \cdots q_s$ mit r, $s \in \mathbb{N}$ und Primzahlen p_i, q_j. Aus $p_r \mid n$ und Satz 3.9 folgt $p_r = q_j$ für ein $j \in \{1, \ldots, s\}$. Wir können jetzt $p_r = q_j$ kürzen. Per Induktion folgt so die Behauptung. □

3.11 Definition.

(a) Ein Ring K heißt ein *Schiefkörper*, wenn $1 \neq 0$ in K ist und außerdem gilt:

$$\text{(SK)} \quad \forall\, a \in K \smallsetminus \{0\} \;\; \exists\, b \in K \;\; ba = 1.$$

(b) Der Schiefkörper K heißt ein *Körper*, wenn K kommutativ (als Ring) ist.

3.12 Satz. *Sei $(K, +, \cdot)$ ein Ring. Genau dann ist K ein Schiefkörper, wenn $(K \smallsetminus \{0\}, \cdot)$ eine Gruppe ist.*

Vor dem Beweis zunächst eine Definition:

3.13 Definition. Ein Ring $K \neq \{0\}$ heißt *nullteilerfrei*, wenn gilt:

$$\forall\, x,\, y \in K\ \Big(xy = 0\ \Rightarrow\ x = 0\ \vee\ y = 0\Big),$$

also wenn $K \setminus \{0\}$ unter Multiplikation abgeschlossen ist.

Die Bedingung K nullteilerfrei gilt keineswegs von selbst, betrachte etwa $K = \mathbb{Z}/6\mathbb{Z}$. Hier ist $\overline{2} \cdot \overline{3} = \overline{0}$, aber $\overline{2} \neq \overline{0}$ und $\overline{3} \neq \overline{0}$. Der Ring $\mathbb{Z}/6\mathbb{Z}$ ist also nicht nullteilerfrei.

Beweis von Satz 3.12. Wir setzen $K^* := K \setminus \{0\}$. Zunächst sei K ein Schiefkörper. Dann ist K nullteilerfrei. Denn sind $x,\, y \in K$ mit $xy = 0$, und ist $x \neq 0$, so gibt es $b \in K$ mit $bx = 1$ (SK), und es folgt $y = 1 \cdot y = bx \cdot y = b \cdot xy = b \cdot 0 = 0$. Also ist $(x, y) \mapsto xy$ eine Abbildung $K^* \times K^* \to K^*$. Diese ist assoziativ, hat neutrales Element 1, und jedes Element aus K^* hat ein inverses Element wegen (SK). Also ist $(K^*, \cdot)$ eine Gruppe.

Umgekehrt sei $(K^*, \cdot)$ eine Gruppe. Es ist $1 \neq 0$, da sonst $K = \{0\}$, also $K^* = \varnothing$ wäre. Also ist 1 das neutrale Element der Gruppe $(K^*, \cdot)$. Axiom (SK) entspricht dem Gruppenaxiom (G2.2) (Existenz des inversen Elements, siehe 2.1). □

Notation: Ist K ein Schiefkörper, so schreibt man $K^* := K \setminus \{0\}$.

3.14 Beispiele.

1. Der Ring $\mathbb{Q}$ der rationalen Zahlen ist ein Körper, ebenso der Ring $\mathbb{R}$ der reellen Zahlen. Den Körper $\mathbb{C}$ der komplexen Zahlen kann man so einführen: Als Menge ist $\mathbb{C} := \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$, mit Addition

$$(x_1, y_1) + (x_2, y_2)\ :=\ (x_1 + x_2,\, y_1 + y_2)$$

und Multiplikation

$$(x_1, y_1) \cdot (x_2, y_2)\ :=\ (x_1x_2 - y_1y_2,\, x_1y_2 + x_2y_1).$$

($x_1,\, x_2,\, y_1,\, y_2 \in \mathbb{R}$). Mit diesen Definitionen ist $(\mathbb{C}, +, \cdot)$ ein Körper mit Null $(0, 0)$ und Eins $(1, 0)$, siehe Aufgabe 11. In der Regel identifiziert man $x \in \mathbb{R}$ mit $(x, 0) \in \mathbb{C}$ und schreibt $i := (0, 1)$. Mit diesen Identifizierungen ist $i^2 = -1$ und $\mathbb{C} = \{x + iy\colon x,\, y \in \mathbb{R}\}$.

2. Ist K ein (kommutativer) Körper, so schreibt man für $a,\, b \in K$ mit $b \neq 0$ alternativ auch $\frac{a}{b} := ab^{-1} = b^{-1}a$.

3. Sei $d \in \mathbb{N}$, sei $\sqrt{d} \in \mathbb{R}$ die positive Quadratwurzel aus d, und sei

$$\mathbb{Q}[\sqrt{d}]\ :=\ \{a + b\sqrt{d}\colon a,\, b \in \mathbb{Q}\},$$

eine Teilmenge von $\mathbb{R}$. Summe und Produkt von zwei Elementen aus $\mathbb{Q}[\sqrt{d}]$ liegen wieder in $\mathbb{Q}[\sqrt{d}]$, letzteres wegen

$$(a + b\sqrt{d})(a' + b'\sqrt{d})\ =\ (aa' + bb'd) + (ab' + a'b)\sqrt{d}.$$

Deshalb ist $\big(\mathbb{Q}[\sqrt{d}], +, \cdot\big)$ ein kommutativer Ring. Tatsächlich ist $\mathbb{Q}[\sqrt{d}]$ ein Körper. Zum Beweis können wir $\sqrt{d} \notin \mathbb{Z}$ annehmen (sonst ist $\mathbb{Q}[\sqrt{d}] = \mathbb{Q}$). Sei $x = a + b\sqrt{d}$ mit $a,\, b \in \mathbb{Q}$, und sei $x \neq 0$, wir zeigen $x^{-1} \in \mathbb{Q}[\sqrt{d}]$. Dazu können wir $b \neq 0$ annehmen. In $\mathbb{R}$ gilt

$$\frac{1}{x} = \frac{1}{a + b\sqrt{d}} = \frac{a - b\sqrt{d}}{(a + b\sqrt{d})(a - b\sqrt{d})} = \frac{a - b\sqrt{d}}{a^2 - b^2d} = \frac{a}{a^2 - b^2d} - \frac{b}{a^2 - b^2d}\sqrt{d}.$$

Daß diese Rechnung korrekt ist, ist nicht trivial. Denn wir haben mit $a - b\sqrt{d}$ erweitert, und dürfen das nur tun, wenn wir wissen $a - b\sqrt{d} \neq 0$, also $\sqrt{d} \neq \frac{a}{b}$. Das ist tatsächlich der Fall:

3.15 Lemma. *Für $d \in \mathbb{N}$ mit $\sqrt{d} \notin \mathbb{N}$ ist $\sqrt{d} \notin \mathbb{Q}$.*

Beweis. Angenommen $\sqrt{d} = \frac{m}{n}$ mit $m, n \in \mathbb{N}$, also $dn^2 = m^2$. Schreibt man beide Seiten als Produkte von Primzahlen, so sieht man aus der Eindeutigkeit der Darstellung (Theorem 3.10), daß jede Primzahl in d gerade oft als Faktor vorkommt. Damit ist aber $\sqrt{d} \in \mathbb{N}$, Widerspruch zur Voraussetzung. □

Beispiele von nicht kommutativen Schiefkörpern werden wir später sehen. In dieser Vorlesung kommen sonst fast nur (kommutative) Körper vor. Hier ist ein weiteres wichtiges Beispiel:

3.16 Satz. *Für jede Primzahl p ist der Ring $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$ ein Körper.*

Beweis. Wir wissen schon, daß $\mathbb{F}_p$ ein kommutativer Ring und $\mathbb{F}_p \neq \{0\}$ ist. Also ist nur Axiom (SK) zu zeigen. Dafür sei $\overline{a} \in \mathbb{F}_p \smallsetminus \{0\}$, also $a \in \mathbb{Z}$ mit $p \nmid a$. Nach Lemma 3.8 gibt es $m, n \in \mathbb{Z}$ mit $am + pn = 1$. Es folgt $am \equiv 1 \pmod{p}$, also $\overline{a} \cdot \overline{m} = \overline{1}$ in $\mathbb{F}_p$, womit die Behauptung bewiesen ist. □

3.17 Beispiel. Die Körper $\mathbb{F}_p$ haben eine ungewohnte Arithmetik. Für $p = 7$ etwa sind die inversen Elemente in $\mathbb{F}_7^*$ gegeben durch

a	1	2	3	4	5	6
a^{-1}	1	4	5	2	3	6

In $\mathbb{F}_7$ gilt also zum Beispiel $\frac{1}{2} = 4$ oder $\frac{3}{5} = 2$.

3.18 Definition. Für einen Körper K ist die *Charakteristik* $\operatorname{char}(K)$ von K wie folgt definiert. Gibt es ein $k \in \mathbb{N}$ mit $k \cdot 1 = 1 + \cdots + 1 = 0$ in K, und ist $n \in \mathbb{N}$ das kleinste solche k, so setzt man $\operatorname{char}(K) := n$. Ist dagegen $k \cdot 1 \neq 0$ in K für alle $k \in \mathbb{N}$, so setzt man $\operatorname{char}(K) := 0$.

3.19 Beispiele. Die Körper $K = \mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$, $\mathbb{Q}[\sqrt{d}]$ haben alle Charakteristik 0. Für p eine Primzahl ist $\operatorname{char}(\mathbb{F}_p) = p$.

3.20 Satz. *Sei K ein Körper. Dann ist* $\operatorname{char}(K)$ *gleich* 0 *oder eine Primzahl.*

Beweis. Sei $n := \operatorname{char}(K)$, o.E. $n > 0$. Wegen $1 \neq 0$ in K ist $n \geq 2$. Angenommen, n sei nicht prim, etwa $n = rs$ mit $1 \neq r$, $s \in \mathbb{N}$. Nach Voraussetzung gilt $rs = 0$ in K. Wegen K nullteilerfrei folgt daraus $r = 0$ oder $s = 0$ in K. Wegen $r, s < n$ ist das ein Widerspruch zur Minimalität von n. □

3.21 Definition. Sei A ein Ring. Eine Teilmenge B von A heißt ein *Teilring* von A, wenn B unter $+$ und $\cdot$ abgeschlossen in A und $(B, +, \cdot)$ ein Ring ist, und wenn B die Eins von A enthält. Ist dabei B ein Körper, so nennt man B auch einen *Teilkörper* von A.

3.22 Beispiele. $\mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{Q}[\sqrt{d}] \subseteq \mathbb{R} \subseteq \mathbb{C}$ ist eine Kette von Teilringen (für $d \in \mathbb{N}$). Dagegen ist $\{0\}$ kein Teilring von $\mathbb{Z}$, wegen $1 \notin \{0\}$. Aus ganz anderem Grund ist $\mathbb{F}_p$ (für p Primzahl) kein Teilring von $\mathbb{Z}$. Denn obwohl man $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$ als Teilmenge von $\mathbb{Z}$ auffassen kann, sind Addition und Multiplikation in $\mathbb{F}_p$ und in $\mathbb{Z}$ völlig verschieden.

4. Polynome

In diesem Abschnitt sei A stets ein *kommutativer* Ring.

4.1 Definition. Sei X eine Menge. Für $x, y \in X$ ist das *Kroneckersymbol*[5] definiert durch

$$\delta_{xy} \;:=\; \delta_{x,y} \;:=\; \begin{cases} 1 & \text{falls } x = y, \\ 0 & \text{falls } x \neq y. \end{cases}$$

4.2 Konstruktion. Ein *Polynom (in der Variable t) über A* ist ein formaler Ausdruck

$$f \;=\; \sum_{i \in \mathbb{N}_0} a_i t^i$$

mit $a_i \in A$ ($i \in \mathbb{N}_0$) derart, daß es $n \in \mathbb{N}_0$ gibt mit $a_i = 0$ für alle $i > n$. Statt $f = \sum_{i\in\mathbb{N}_0} a_i t^i$ schreibt man dann meist einfacher $f = \sum_{i=0}^{n} a_i t^i = a_0 + a_1 t + \cdots + a_n t^n$. Zwei Polynome $f = \sum_{i\in\mathbb{N}_0} a_i t^i$ und $g = \sum_{i\in\mathbb{N}_0} b_i t^i$ heißen gleich, i. Z. $f = g$, wenn $a_i = b_i$ für alle $i \geq 0$ gilt. Die Menge aller Polynome (in der Variablen t über A) wird mit $A[t]$ bezeichnet. Ist $f = \sum_{i=0}^{n} a_i t^i \in A[t]$ mit $a_0, \ldots, a_n \in A$ und $a_n \neq 0$, so heißt $\deg(f) := n$ der *Grad* von f, und a_n heißt der *Leitkoeffizient* von f. Für das Nullpolynom definiert man $\deg(0) := -\infty$ (es hat keinen Leitkoeffizient). Stets ist also $\deg(f) \in \mathbb{N}_0 \cup \{-\infty\}$.

Heißt die Polynomvariable statt t etwa $x, y, \ldots$, dann schreibt man entsprechend $A[x]$, $A[y], \ldots$ für den Polynomring. Für $f = \sum_{i\geq 0} a_i t^i$, $g = \sum_{i\geq 0} b_i t^i \in A[t]$ definiere

$$f + g \;:=\; \sum_{i\geq 0} (a_i + b_i) t^i, \quad fg \;:=\; \sum_{k\geq 0} c_k t^k \;\text{ mit }\; c_k \;=\; \sum_{i=0}^{k} a_i b_{k-i}$$

($k \geq 0$). Dann sind auch $f + g$ und fg Polynome. Denn ist $\deg(f) \leq m$ und $\deg(g) \leq n$, so ist $a_i + b_i = 0$ für $i > \max\{m, n\}$ und $c_k = 0$ für $k > m + n$.

Diese Definition macht $A[t]$ zu einem kommutativen Ring (Übung!). Die Null im Ring $A[t]$ ist das Nullpolynom 0 (alle Koeffizienten $a_i = 0$), die Eins in $A[t]$ ist das Polynom 1 (mit $a_0 = 1$ und $a_i = 0$ für $i \geq 1$). Der Ring A ist ein Teilring des Rings $A[t]$, und besteht aus den Polynomen vom Grad ≤ 0.

4.3 Satz. *Seien f, $g \in A[t]$.*

(a) $\deg(f + g) \leq \max\{\deg(f), \deg(g)\}$, *mit Gleichheit falls* $\deg(f) \neq \deg(g)$.
(b) $\deg(fg) \leq \deg(f) + \deg(g)$, *mit Gleichheit falls A nullteilerfrei ist.*
(c) *Ist der Ring A nullteilerfrei, so ist auch der Ring $A[t]$ nullteilerfrei.*

Wir verwenden hierbei folgende Regeln für das Symbol $-\infty$: $-\infty \leq n$ und $-\infty + n = n + (-\infty) = -\infty$ für alle $n \in \mathbb{N}_0 \cup \{-\infty\}$.

[5] Leopold Kronecker (1823–1891)

BEWEIS. (a) ist klar. (b) Seien $f = \sum_{i\geq 0} a_i t^i$ und $g = \sum_{i\geq 0} b_i t^i$ in $A[t]$. Ist $f = 0$ oder $g = 0$, so ist die Aussage richtig. Seien also $f,\ g \neq 0$, sei $m = \deg(f)$ und $n = \deg(g)$. Dann ist

$$fg \;=\; \Big(a_m t^m + \sum_{i=0}^{m-1} a_i t^i\Big)\Big(b_n t^n + \sum_{i=0}^{n-1} b_i t^i\Big) \;=\; a_m b_n t^{m+n} + \sum_{k=0}^{m+n-1} c_k t^k$$

mit geeigneten $c_k \in A$. Daraus sehen wir $\deg(fg) \leq m+n$. Wegen $a_m,\ b_n \neq 0$ ist $a_m b_n \neq 0$ falls A nullteilerfrei ist, dann gilt also $\deg(fg) = m+n$. (c) folgt aus (b): $f,\ g \neq 0 \Rightarrow \deg(f) \geq 0,\ \deg(g) \geq 0 \Rightarrow \deg(fg) \geq 0$ nach (b) $\Rightarrow fg \neq 0$. □

Jetzt sei $A = K$ ein Körper.

4.4 Satz. (Polynomdivision mit Rest) *Seien $f,\ g \in K[t]$, sei $g \neq 0$. Dann gibt es Polynome $q,\ r \in K[t]$ mit $f = qg + r$ und $\deg(r) < \deg(g)$. Dadurch sind q und r eindeutig bestimmt.*

Beachte die formale Analogie dieses Satzes zur Division mit Rest von ganzen Zahlen (Satz 3.5).

BEWEIS. Wir zeigen zunächst die Existenz von q und r durch Induktion nach $\deg(f)$. Ist $\deg(f) < \deg(g)$, so können wir $q = 0$ und $r = f$ nehmen. Sei also $m := \deg(f) \geq \deg(g) =: n$, sei $f = a_m t^m + \cdots + a_0$ und $g = b_n t^n + \cdots + b_0$ (mit $a_i,\ b_j \in K$, beachte $a_m,\ b_n \neq 0$). Betrachte das Polynom

$$\tilde{f} \;:=\; f - \frac{a_m}{b_n}\, t^{m-n} \cdot g \;=\; \Big(a_m t^m + \cdots + a_0\Big) - \frac{a_m}{b_n}\, t^{m-n} \cdot \Big(b_n t^n + \cdots + b_0\Big)$$

Es gilt $\deg(\tilde{f}) < m = \deg(f)$, da sich der Koeffizient von t^m in den beiden Summanden weghebt. Nach Induktionsvoraussetzung gibt es also Polynome $\tilde{q},\ r \in K[t]$ mit $\deg(r) < \deg(g)$ und $\tilde{f} = \tilde{q}g + r$, und es folgt

$$f \;=\; \tilde{f} + \frac{a_m}{b_n}\, t^{m-n} \cdot g \;=\; \Big(\tilde{q} + \frac{a_m}{b_n}\, t^{m-n}\Big) \cdot g + r.$$

Damit ist die Existenzaussage bewiesen.

Zur Eindeutigkeit: Seien auch $q_1,\ r_1 \in K[t]$ mit $f = qg + r = q_1 g + r_1$ und $\deg(r_1) < \deg(g)$. Dann ist $(q - q_1)g = r_1 - r$ und $\deg(r_1 - r) \leq \max\{\deg(r_1), \deg(r)\} < \deg(g)$. Wäre $q - q_1 \neq 0$, so wäre $\deg((q-q_1)g) \geq \deg(g)$ nach 4.3(b). Also muß $q - q_1 = 0$ sein, und somit auch $r = r_1$. □

4.5 Bemerkung. Der Beweis war konstruktiv, d.h. er liefert einen *Algorithmus* zur Berechnung von q und r. Beispiel $f = t^4 - 2t^3 - 7$, $g = t^2 + 3$, dann ergibt sich schrittweise

$$\begin{array}{l}
(t^4 - 2t^3 \qquad\quad - 7) \;:\; (t^2 + 3)\,? \\
\underline{\;t^4 \qquad + 3t^2 \qquad\quad\;\;} \\
\quad - 2t^3 - 3t^2 \qquad - 7 \\
\quad \underline{- 2t^3 \qquad - 6t \qquad\;\;} \\
\qquad\quad - 3t^2 + 6t - 7 \\
\qquad\quad \underline{- 3t^2 \qquad\quad - 9} \\
\qquad\qquad\quad\;\; 6t + 2
\end{array}$$

$q = t^2 - 2t - 3$ und $r = 6t + 2$.

4.6 Definition. Sei A ein kommutativer Ring. Für $f = \sum_{i=0}^{n} a_i t^i \in A[t]$ und $c \in A$ schreibt man

$$f(c) := \sum_{i=0}^{n} a_i c^i \in A,$$

man setzt also das Element c für die Variable t ein. Dabei gilt

$$(f+g)(c) = f(c) + g(c), \quad (fg)(c) = f(c)g(c)$$

für $f, g \in A[t]$. Das Element $c \in A$ heißt eine *Nullstelle* von f, wenn $f(c) = 0$ ist.

4.7 Korollar. *Sei K ein Körper, sei $f \in K[t]$.*

(a) *Ist $c \in K$ mit $f(c) = 0$, so gibt es $g \in K[t]$ mit $f = (t-c)g$.*

(b) *Ist $f \neq 0$ und $n = \deg(f)$, so hat f höchstens n verschiedene Nullstellen in K.*

Beachte, in (a) gilt trivialerweise auch die Umkehrung.

Beweis. (a) Dividiere f durch $t - c$ mit Rest. Das gibt $f = (t-c)g + r$ mit $g, r \in K[t]$ und $\deg(r) \leq 0$, also $r \in K$. Einsetzen von c zeigt $r = 0$.

(b) Induktion nach n, der Fall $n = 0$ ist klar. Sei $\deg(f) = n \geq 1$, und sei $c \in K$ eine Nullstelle von f (gibt es keine, so sind wir fertig). Nach (a) ist $f = (t-c)g$ mit $g \in K[t]$. Dabei ist $\deg(g) = n - 1$ (und $g \neq 0$), nach Induktionsvoraussetzung hat g also höchstens $n - 1$ Nullstellen. Jede Nullstelle $c' \neq c$ von f ist Nullstelle von g wegen $0 = f(c') = (c' - c)\, g(c')$, und da K als Körper nullteilerfrei ist. Daraus folgt die Behauptung. □

KAPITEL II

Vektorräume

1. Erste Definitionen

1.1 *Vorbemerkung*: Zur Analyse der Geometrie in der Ebene führte Descartes Koordinaten ein. Auf diese Weise lassen sich geometrische Frage in algebraische übersetzen, was sich als äußerst erfolgreich erwies. Zeichne einen Punkt O der Ebene als Nullpunkt (Ursprung) aus, wähle zwei verschiedene Geraden g_1, g_2 durch O, und dann für $i = 1, 2$ je einen Punkt $P_i \neq O$ auf der Gerade g_i, der als "Einheit" dient. Dann identifizieren sich die Punkte von g_i mit den reellen Zahlen, und über die Konstruktion von Parallelen die Punkte der Ebene mit Paaren $(a_1, a_2) \in \mathbb{R}^2$.

Ähnlich geht es im Raum, wo man 3 statt 2 Geraden durch O braucht und schon vorsichtiger sein muß: Es genügt nicht, daß die Geraden g_1, g_2, g_3 paarweise verschieden sind, vielmehr dürfen sie nicht in einer Ebene liegen. Wählt man wieder Punkte $P_i \neq O$ auf g_i $(i = 1, 2, 3)$ und geht ansonsten vor wie in der Ebene, so hat man die Punkte des Raums mit den Tripeln in $\mathbb{R}^3$ identifiziert.

Die Systematisierung des Ansatzes führt zu Vektorräumen. Jeder Vektorraum V über $\mathbb{R}$ hat eine Addition und eine Skalarmultiplikation. Im Beispiel der Ebene $\mathbb{R}^2$ entspricht die Addition der üblichen "Vektoraddition" mit Hilfe von Parallelogrammen. Die Multiplikation eines Vektors v mit einem Skalar a entspricht einer Streckung von v um den Faktor a, was für $a < 0$ auch eine Spiegelung am Nullpunkt O bedeutet. Völlig analog ist es im Raum $\mathbb{R}^3$.

Wir führen Vektorräume jetzt völlig abstrakt ein. Dabei beschränken wir uns nicht auf $\mathbb{R}^2$ oder $\mathbb{R}^3$, sondern lassen auch höhere (sogar unendliche) "Dimension" zu. Aus vielen Gründen ist es außerdem wichtig, auch Multiplikation von Vektoren mit Skalaren zuzulassen, die nicht reelle Zahlen sind. Vielmehr betrachten wir Vektorräume über beliebigen abstrakten Körpern K (oder später sogar über Ringen, dann benutzt man statt Vektorraum nur ein anderes Wort).

Sei stets K ein Körper.

1.2 Definition. Ein *K-Vektorraum* (oder *Vektorraum über K*) ist eine Menge V zusammen mit zwei Abbildungen

$$+\colon V \times V \to V, \qquad (v, w) \mapsto v + w$$

(die *(Vektor-) Addition*) und

$$\cdot\colon K \times V \to V, \qquad (a, v) \mapsto a \cdot v = av,$$

(die *Skalarmultiplikation*), so daß gelten:

(VR1) $(V, +)$ ist eine abelsche Gruppe;

(VR2) $(a + b) \cdot v = a \cdot v + b \cdot v$ und $a \cdot (v + w) = a \cdot v + a \cdot w$ für alle $a, b \in K$ und $v, w \in V$;

(VR3) $a \cdot (b \cdot v) = (ab) \cdot v$ für alle $a, b \in K$ und $v \in V$;
(VR4) $1 \cdot v = v$ für alle $v \in V$.

1.3 Bemerkung. Das neutrale Element der abelschen Gruppe $(V, +)$ bezeichnen wir vorläufig mit $\mathbf{0}$ (*Nullvektor*); das neutrale Element von $(K, +)$ wird mit 0 bezeichnet. Die Elemente von V bzw. K werden auch *Vektoren* bzw. *Skalare* genannt.

1.4 Lemma. *Sei V ein K-Vektorraum, seien $a \in K$ und $v \in V$.*
(a) $av = \mathbf{0} \;\Leftrightarrow\; a = 0$ *oder* $v = \mathbf{0}$.
(b) $(-a)v = -(av) = a(-v)$.

Beweis. (a) "$\Leftarrow$": $0 \cdot v + 0 \cdot v = (0 + 0) \cdot v = 0 \cdot v = 0 \cdot v + \mathbf{0}$, und Kürzen in $(V, +)$ gibt $0 \cdot v = \mathbf{0}$. Analog sieht man $a \cdot \mathbf{0} = \mathbf{0}$. "$\Rightarrow$": Sei $av = \mathbf{0}$. Ist $a \neq 0$, so folgt $v = 1 \cdot v = a^{-1}av = a\mathbf{0} = \mathbf{0}$ (letzte Gleichheit nach dem eben bewiesenen Schritt).
(b) Analog zur Argumentation in Ringen, siehe Bemerkung I.3.3.1. □

1.5 Beispiele. Hier sind Beispiele von Vektorräumen:

1. (Das Standardbeispiel) $K^n = \{x = (x_1, \ldots, x_n) \colon x_1, \ldots, x_n \in K\}$ wird ein K-Vektorraum durch

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

$$a \cdot (x_1, \ldots, x_n) \;:=\; (ax_1, \ldots, ax_n)$$

(für $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in K^n$ und $a \in K$). Axiome (VR1)–(VR4) folgen aus den Körperaxiomen von K. Insbesondere ist K selbst ein K-Vektorraum.

2. Der Polynomring $K[t]$ ist ein K-Vektorraum, wobei Addition und Skalarmultiplikation durch die Ringoperationen von $K[t]$ gegeben sind.

3. Ist $(V_i)_{i \in I}$ eine Familie von K-Vektorräumen, so wird das kartesische Produkt

$$\prod_{i \in I} V_i \;=\; \Big\{(v_i)_{i \in I} \colon\; v_i \in V_i \;(i \in I)\Big\}$$

selbst zu einem K-Vektorraum durch komponentenweise Definition von Addition und Skalarmultiplikation:

$$(x_i)_{i \in I} + (y_i)_{i \in I} \;:=\; (x_i + y_i)_{i \in I}, \quad a \cdot (x_i)_{i \in I} \;:=\; (ax_i)_{i \in I}.$$

Dieser Vektorraum heißt das *direkte Produkt* der Vektorräume V_i $(i \in I)$.

4. Der einfachste K-Vektorraum ist $V = \{\mathbf{0}\}$, der Nullvektorraum.

5. $\mathbb{C}$ läßt sich auffassen nicht nur als Vektorraum über $\mathbb{C}$, sondern z.B. auch als Vektorraum über $\mathbb{R}$ oder $\mathbb{Q}$. Allgemeiner: Ist L ein Körper und $K \subseteq L$ ein Teilkörper, so wird jeder L-Vektorraum W durch Einschränkung des Skalarenbereichs zu einem K-Vektorraum. Vektorräume können also oft als Vektorräume über verschiedenen Körpern aufgefaßt werden. Daher ist es wichtig, stets klar darüber zu sein, mit welchem Grundkörper man arbeitet.

1.6 Definition. Sei V ein K-Vektorraum. Ein *(K-) Untervektorraum* (oder *(K-) linearer Unterraum*, oder *(K-) Unterraum*) von V ist eine Untergruppe U von $(V, +)$, für die gilt:

$$\forall\, a \in K \;\forall\, u \in U \quad au \in U.$$

Ist U ein K-Unterraum von V, so ist U mit den Restriktionen $+\colon U \times U \to U$ und $\cdot\colon K \times U \to U$ der Strukturabbildungen von V selbst ein K-Vektorraum.

1.7 Lemma. *Sei V ein K-Vektorraum, sei $U \subseteq V$ eine Teilmenge. Genau dann ist U ein K-Unterraum von V, wenn gilt:*

(1) $U \neq \varnothing$;
(2) $\forall\, x,\, y \in U \quad x + y \in U$;
(3) $\forall\, a \in K\ \forall\, x \in U \quad ax \in U$.

Beweis. Ist U ein Unterraum, so gelten (1)–(3). Aus (2) und (3) folgt umgekehrt $x - y = x + (-1) \cdot y \in U$ für alle x, $y \in U$. Nach Satz I.2.13 ist also U Untergruppe von $(V, +)$. □

1.8 Satz. *Sei V ein K-Vektorraum, und sei $(U_i)_{i\in I}$ eine Familie von Unterräumen von V. Dann ist auch $\bigcap_{i\in I} U_i$ ein Unterraum von V.*

Beweis. Klar (siehe Korollar I.2.15). □

1.9 Beispiele.

1. Sei $V = K^n$, seien $a_1, \ldots, a_n \in K$. Dann ist

$$U \;:=\; \{x \in K^n\colon\ a_1x_1 + \cdots + a_nx_n = 0\}$$

ein Unterraum von K^n. In der Tat, Bedingungen (1)–(3) aus 1.7 sind erfüllt.

2. $U = \mathbb{Z}^n$ ist eine Untergruppe, aber kein $\mathbb{R}$-Unterraum, von $V = \mathbb{R}^n$.

3. Für jedes $d \in \mathbb{N}_0$ ist $K[t]_d := \{f \in K[t]\colon\ \deg(f) \leq d\}$ ein Unterraum des K-Vektorraums $K[t]$.

4. Für jeden Vektorraum V ist $\{\mathbf{0}\}$ ein Unterraum von V.

5. Vor dem nächsten Beispiel führen wir eine verbreitete Sprechweise ein. Sei I eine Menge, und sei $\mathcal{P}$ eine Eigenschaft, die für jedes $i \in I$ entweder wahr oder falsch ist. Man sagt, $\mathcal{P}$ gilt *für fast alle (f. f. a.)* $i \in I$, wenn die Menge $\{i \in I\colon \mathcal{P}(i)$ ist falsch$\}$ endlich ist. Beachte: Ist $|I| < \infty$, so ist jede Eigenschaft für fast alle $i \in I$ erfüllt.

Sei $(V_i)_{i\in I}$ eine Familie von Vektorräumen, und sei $V = \prod_{i\in I} V_i$ ihr direktes Produkt (1.5.3). Dann ist

$$U \;=\; \{(v_i)_i \in V\colon v_i = \mathbf{0} \text{ für fast alle } i \in I\}$$

ein Unterraum von V. Ist $|I| < \infty$, so ist $U = V$.

2. Lineare Abhängigkeit, Basen, Dimension

Sei K stets ein Körper. Falls nicht anders gesagt, ist V immer ein beliebiger K-Vektorraum.

2.1 Definition. Sei V ein K-Vektorraum, sei $\mathcal{F} = (v_i)_{i\in I}$ eine Familie von Elementen von V.

(a) Ein Vektor $v \in V$ heißt eine *(K-) Linearkombination* von $\mathcal{F}$, wenn es $n \in \mathbb{N}_0$ und $i_1, \ldots, i_n \in I$ und $a_1, \ldots, a_n \in K$ gibt mit $v = a_1v_{i_1} + \cdots + a_nv_{i_n}$.
(b) Die Menge aller K-Linearkombinationen von $\mathcal{F}$ wird mit $\operatorname{span}_K(\mathcal{F})$ oder $\operatorname{span}_K(v_i\colon i \in I)$ bezeichnet.

(Falls über K keine Unklarheit besteht, wird K meistens nicht erwähnt.)

2.2 Bemerkungen.

1. Der Nullvektor $\mathbf{0}$ ist stets Linearkombination von $\mathcal{F}$, auch für $I = \varnothing$: $\mathbf{0}$ ist die "leere Summe". Wenn das zu spitzfindig erscheint, kann man es als Teil der Definition betrachten.

2. Sei $V = K^n$. Für $1 \leq i \leq n$ sei

$$e_i \;:=\; (\delta_{ij})_{j=1,\dots,n} \;=\; (0,\dots,0,1,0,\dots,0) \;\in K^n$$

(mit 1 an der i-ten Stelle) der *i-te kanonische Basisvektor*. Es gilt $\operatorname{span}(e_1,\dots,e_n) = K^n$, denn für $x = (x_1,\dots,x_n) \in K^n$ ist $x = \sum_{i=1}^n x_i e_i$. Allgemeiner gilt für $1 \leq m \leq n$

$$\operatorname{span}(e_1,\dots,e_m) \;=\; \{x \in K^n \colon x_{m+1} = \dots = x_n = 0\}.$$

3. Definition 2.1 überträgt sich von *Familien* von Vektoren in V auf *Teilmengen* von V: Sei $M \subseteq V$ eine Teilmenge. Dann heißt $v \in V$ eine (K-) Linearkombination von M, wenn es $n \in \mathbb{N}_0$, $v_1,\dots,v_n \in M$ und $a_1,\dots,a_n \in K$ gibt mit $v = a_1v_1 + \dots + a_nv_n$. Analog zu vorher schreibt man $\operatorname{span}_K(M)$ für die Menge aller Linearkombinationen von M.

2.3 Satz. *Für jede Teilmenge $M \subseteq V$ ist* $\operatorname{span}(M)$ *ein Unterraum von V, und es gilt*

$$\operatorname{span}(M) \;=\; \bigcap \Big\{U \colon\; U \text{ Unterraum von } V,\; M \subseteq U\Big\}.$$

Man nennt $\operatorname{span}(M)$ *den von M* aufgespannten *(oder* erzeugten*)* Unterraum *von V, und nennt M ein* Erzeugendensystem *von* $\operatorname{span}(M)$.

Beweis. Aus Lemma 1.7 folgt sofort (beachte $\mathbf{0} \in \operatorname{span}(M)$), daß $\operatorname{span}(M)$ ein Unterraum von V ist. Für jeden Unterraum U von V mit $M \subseteq U$ ist $\operatorname{span}(M) \subseteq U$. Daraus folgt "$\subseteq$", und "$\supseteq$" ist klar, da $\operatorname{span}(M)$ selbst ein Unterraum ist. □

2.4 Bemerkungen.

1. $\operatorname{span}(\varnothing) = \{\mathbf{0}\}$, und $\operatorname{span}(v) = Kv := \{av \colon a \in K\}$ für $v \in V$.

2. Für Teilmengen $M' \subseteq M \subseteq V$ ist $\operatorname{span}(M') \subseteq \operatorname{span}(M)$. Ist dabei $M \subseteq \operatorname{span}(M')$, so gilt $\operatorname{span}(M') = \operatorname{span}(M)$. (Denn $\operatorname{span}(M') \subseteq \operatorname{span}(M)$ ist trivial. Ist $M \subseteq \operatorname{span}(M')$, so ist $\operatorname{span}(M')$ ein Unterraum von V, welcher M enthält. Nach 2.3 folgt dann also $\operatorname{span}(M) \subseteq \operatorname{span}(M')$.)

3. Konkretes Beispiel: Sei $V = K^3$, seien $v_1 = (-3,2,1)$, $v_2 = (-2,1,1)$. Dann ist

$$\operatorname{span}(v_1,v_2) \;=\; \Big\{x \in K^3 \colon\; x_1 + x_2 + x_3 = 0\Big\}.$$

Denn die rechte Menge ist ein Unterraum, welcher v_1, v_2 enthält, also gilt "$\subseteq$". Ist umgekehrt $x \in K^3$ mit $x_1 + x_2 + x_3 = 0$, so ist

$$x \;=\; (x_1 + 2x_2)v_1 - (2x_1 + 3x_2)v_2 \;\in\; \operatorname{span}(v_1,v_2)$$

(nachrechnen!). Wir werden bald Verfahren lernen, um solche Rechnungen zu systematisieren.

2.5 Definition. Ein K-Vektorraum V heißt *endlich erzeugt*, wenn es eine endliche Teilmenge $M \subseteq V$ mit $V = \operatorname{span}_K(M)$ gibt.

2.6 Beispiel. Der K-Vektorraum K^n ist endlich erzeugt (Bemerkung 2.2.2). Der K-Vektorraum $K[t]$ ist dagegen nicht endlich erzeugt. Denn sind $f_1, \dots, f_r \in K[t]$ und ist $d := \max\{\deg(f_i)\colon i = 1, \dots, r\}$, so ist $t^{d+1} \notin \operatorname{span}(f_1, \dots, f_r)$, und insbesondere $\operatorname{span}(f_1, \dots, f_r) \neq K[t]$.

2.7 Definition.

(a) Eine endliche Folge $(v_1, \dots, v_n)$ von Vektoren aus V heißt *linear abhängig (über K)*, wenn es Elemente $a_1, \dots, a_n \in K$ gibt mit $a_i \neq 0$ für mindestens ein $i \in \{1, \dots, n\}$ und mit $a_1 v_1 + \dots + a_n v_n = \mathbf{0}$.

(b) Eine beliebige Familie $(v_i)_{i \in I}$ von Vektoren aus V heißt *linear abhängig (über K)*, wenn es eine endliche Teilmenge $J \subseteq I$ gibt, so daß $(v_j)_{j \in J}$ linear abhängig (im Sinn von (a)) ist. Andernfalls heißt $(v_i)_{i \in I}$ *linear unabhängig* (über K).

Eine Familie ist also genau dann linear unabhängig, wenn man aus ihr den Nullvektor nur in trivialer Weise linear kombinieren kann (alle Koeffizienten gleich 0).

2.8 Beispiele.

1. Die leere Familie (Indexmenge $I = \varnothing$) ist linear unabhängig. Eine Familie (v) aus einem Vektor ist genau dann linear abhängig, wenn $v = \mathbf{0}$ ist (Lemma 1.4(a)). Eine Familie (v, w) aus zwei Vektoren ist genau dann linear abhängig, wenn $w \in Kv$ oder $v \in Kw$ ist (Übung!).

2. Betrachte die Vektoren $v_1 = (1, 2)$, $v_2 = (-1, 0) \in K^2$. Ist die Folge (v_1, v_2) linear unabhängig? Für $a_1, a_2 \in K$ ist $a_1 v_1 + a_2 v_2 = (a_1 - a_2,\, 2a_1)$. Aus $a_1 v_1 + a_2 v_2 = \mathbf{0}$ folgt also $a_1 = a_2$ und $2a_1 = 0$. Für $\operatorname{char}(K) \neq 2$ folgt daraus $a_1 = a_2 = 0$, also ist (v_1, v_2) linear unabhängig. Für $\operatorname{char}(K) = 2$ dagegen ist $v_1 + v_2 = \mathbf{0}$, also ist (v_1, v_2) linear abhängig.

3. Man sagt oft etwas schlampig "die Vektoren $v_1, \dots, v_n$ sind linear unabhängig", und meint damit "die Familie $(v_1, \dots, v_n)$ von Vektoren ist linear unabhängig".

Hier kommen nun zwei andere Charakterisierungen der linearen (Un-) Abhängigkeit:

2.9 Lemma. *Sei $(v_i)_{i \in I}$ eine Familie von Vektoren in V. Dann sind äquivalent:*

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

(ii) *jedes $v \in \operatorname{span}(v_i\colon i \in I)$ ist* in eindeutiger Weise *Linearkombination der v_i.*

Mit (ii) ist dabei gemeint: Aus $\sum_{i \in I} a_i v_i = \sum_{i \in I} b_i v_i$ (mit $a_i, b_i \in K$ und $a_i = b_i = 0$ für fast alle i) folgt $a_i = b_i$ für alle $i \in I$.

Beweis. (i) $\Rightarrow$ (ii): Aus $\sum_{i \in I} a_i v_i = \sum_{i \in I} b_i v_i$ folgt $\sum_{i \in I} (a_i - b_i) v_i = \mathbf{0}$, und nach Voraussetzung (i) also $a_i - b_i = 0$ für alle i. Umgekehrt ist (i) ein Spezialfall von (ii) (nimm $b_i = 0$ für alle $i \in I$ in (ii)). □

2.10 Satz. *Eine Familie $(v_i)_{i \in I}$ ist genau dann linear abhängig, wenn es ein $j \in I$ gibt mit $v_j \in \operatorname{span}(v_i\colon\ i \in I \smallsetminus \{j\})$. Alsdann ist*

$$\operatorname{span}(v_i\colon i \in I) \;=\; \operatorname{span}(v_i\colon i \in I \smallsetminus \{j\}).$$

BEWEIS. Ist die Familie linear abhängig, so gibt es $a_i \in K$ $(i \in I)$, fast alle $= 0$, mit $\sum_{i\in I} a_i v_i = \mathbf{0}$ und $a_j \neq 0$ für ein $j \in I$. Auflösen nach v_j gibt $v_j = -\sum_{i\neq j} \frac{a_i}{a_j} v_i$, also $v_j \in \operatorname{span}(v_i \colon i \in I \smallsetminus \{j\}) = \operatorname{span}(v_i \colon i \in I)$. Es ist klar, daß damit auch die Zusatzbehauptung gilt (siehe auch Bemerkung 2.4.2). Umgekehrt folgt aus $v_j \in \operatorname{span}(v_i \colon i \neq j)$ die Existenz einer Identität $\sum_{i\in I} a_i v_i = \mathbf{0}$ mit $a_j = 1$, also die lineare Abhängigkeit von $(v_i)_{i\in I}$. □

2.11 Definition. Sei V ein K-Vektorraum. Eine Familie $\mathcal{B} = (v_i)_{i\in I}$ von Vektoren aus V heißt eine *(K-) Basis* von V, wenn $\mathcal{B}$ linear unabhängig (über K) und $\operatorname{span}_K(\mathcal{B}) = V$ ist.

2.12 Satz. *Eine Familie $\mathcal{B}$ in V ist genau dann eine Basis von V, wenn gilt: Jedes $v \in V$ ist in eindeutiger Weise (siehe 2.9) Linearkombination von $\mathcal{B}$.*

BEWEIS. Das ist klar nach Lemma 2.9. □

2.13 Beispiele.

1. Sei $V = K^n$, sei $e_i = (\delta_{ij})_{1\leq j\leq n}$ der i-te kanonische Basisvektor für $1 \leq i \leq n$ (siehe 2.2.2). Dann ist $(e_1, \ldots, e_n)$ eine Basis des K-Vektorraums K^n, die sogenannte *kanonische Basis* von K^n.

2. Die Familie $(t^n)_{n\in\mathbb{N}_0}$ (mit $t^0 := 1$) ist eine Basis des K-Vektorraums $K[t]$.

3. Für $V = \{\mathbf{0}\}$ ist die leere Familie die einzige Basis.

4. Fassen wir $\mathbb{C}$ als $\mathbb{R}$-Vektorraum auf, so ist $(1, i)$ eine $\mathbb{R}$-Basis von $\mathbb{C}$, denn jedes $z \in \mathbb{C}$ schreibt sich eindeutig als $z = x + yi$ mit $x, y \in \mathbb{R}$. Über dem Grundkörper $\mathbb{C}$ dagegen ist die Familie $(1, i)$ linear abhängig wegen $i \cdot 1 - 1 \cdot i = 0$, und ist deshalb keine Basis mehr. Die Begriffe lineare Unabhängigkeit, lineares Erzeugnis und Basis hängen also im allgemeinen vom Grundkörper ab.

2.14 Satz. *Sei V ein Vektorraum und $\mathcal{F} = (v_i)_{i\in I}$ eine Familie in V. Es sind äquivalent:*

(i) *$\mathcal{F}$ ist eine Basis von V;*
(ii) *$\mathcal{F}$ ist ein* minimales Erzeugendensystem *von V (d.h.* $\operatorname{span}(\mathcal{F}) = V$*, und* $\operatorname{span}(v_j \colon j \in J) \neq V$ *für jede echte Teilmenge $J \subsetneq I$);*
(iii) *$\mathcal{F}$ ist eine* maximale linear unabhängige Familie *in V (d.h. $\mathcal{F}$ ist linear unabhängig, und jede Familie $(v_j)_{j\in J}$ in V mit $I \subsetneq J$ ist linear abhängig).*

BEWEIS. (i) ⇒ (ii): Sei $j \in I$. Wäre $v_j \in \operatorname{span}(v_i \colon i \neq j)$, so wäre $\mathcal{F}$ linear abhängig nach 2.10, Widerspruch zu (i).

(ii) ⇒ (iii): Wäre $\mathcal{F}$ linear abhängig, so gäbe es nach 2.10 ein $j \in I$ mit $\operatorname{span}(v_i \colon i \in I \smallsetminus \{j\}) = \operatorname{span}(\mathcal{F}) = V$, im Widerspruch zu (ii). Also ist $\mathcal{F}$ linear unabhängig. Sei $\mathcal{F}' = (v_j \colon j \in J)$ eine Familie mit $I \subsetneq J$, wähle ein $k \in J \smallsetminus I$. Wegen $v_k \in \operatorname{span}(\mathcal{F}) = \operatorname{span}(v_j \colon k \in J \smallsetminus \{k\})$ ist $\mathcal{F}'$ linear abhängig (2.10).

(iii) ⇒ (i): Zu zeigen ist $V = \operatorname{span}(\mathcal{F})$. Angenommen, es gäbe $v \in V$ mit $v \notin \operatorname{span}(\mathcal{F})$. Sei $J := I \cup \{\infty\}$ mit $\infty \notin I$, und sei $v_\infty := v$. Die vergrößerte Familie $\mathcal{F}' := (v_j)_{j\in J}$ ist linear abhängig nach (iii), es gibt also eine Gleichung

$$av + \sum_{i\in I} a_i v_i \;=\; \mathbf{0},$$

mit a, $a_i \in K$ (fast alle $a_i = 0$) und mit $a \neq 0$ oder $a_i \neq 0$ für ein $i \in I$. Wegen $v \notin \operatorname{span}(\mathcal{F})$ ist $a = 0$, also ist $\mathcal{F}$ linear abhängig, Widerspruch zu (iii). □

Wir betrachten jetzt endlich erzeugte Vektorräume, und zeigen zunächst die Existenz einer Basis:

2.15 Satz. (Basisauswahlsatz) *Sei $V = \operatorname{span}(v_1, \ldots, v_n)$ mit $n \in \mathbb{N}_0$. Dann gibt es eine Teilmenge $I \subseteq \{1, \ldots, n\}$, so daß $(v_i \colon i \in I)$ eine Basis von V ist. Insbesondere hat jeder endlich erzeugte Vektorraum eine endliche Basis.*

Beweis. Sei $J = \{1, \ldots, n\}$, sei $\mathcal{F} = (v_i \colon i \in J)$. Ist $\mathcal{F}$ linear unabhängig, so ist $\mathcal{F}$ eine Basis von V. Andernfalls gibt es ein $j \in J$ mit $\operatorname{span}(v_i \colon i \in J \setminus \{j\}) = \operatorname{span}(\mathcal{F}) = V$ (Satz 2.10). Wir können also den gerade mit $\mathcal{F}$ durchgeführten Schritt für die Familie $\mathcal{F}' = (v_i \colon i \in J \setminus \{j\})$ wiederholen. Mache so weiter, dann hat man nach höchstens n Schritten eine Basis von V. □

Nun zeigen wir, daß je zwei Basen dieselbe Anzahl von Elementen haben. Zunächst ein Lemma:

2.16 Lemma. *Sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V, seien $a_1, \ldots, a_n \in K$, und sei $w = \sum_{i=1}^n a_i v_i$. Ist $j \in \{1, \ldots, n\}$ mit $a_j \neq 0$, so ist auch $\mathcal{B}' := (v_1, \ldots, v_{j-1}, w, v_{j+1}, \ldots, v_n)$ eine Basis von V.*

Beweis. Wegen $v_j = \frac{1}{a_j}\big(w - \sum_{i \neq j} a_i v_i\big)$ ist $v_j \in \operatorname{span}(\mathcal{B}')$, also $\operatorname{span}(\mathcal{B}') = V$ (2.4.2). Zu zeigen bleibt die lineare Unabhängigkeit von $\mathcal{B}'$. Sei

$$bw + \sum_{\substack{i=1 \\ i \neq j}}^{n} b_i v_i \;=\; \mathbf{0}$$

mit $b_i \in K$ $(i \neq j)$, $b \in K$. Einsetzen der Definition von w gibt

$$\mathbf{0} \;=\; b \sum_{i=1}^{n} a_i v_i + \sum_{\substack{i=1 \\ i \neq j}}^{n} b_i v_i.$$

Die rechte Seite ist eine Linearkombination von $\mathcal{B}$, und der Koeffizient von v_j ist ba_j. Wegen $\mathcal{B}$ linear unabhängig und $a_j \neq 0$ folgt $b = 0$, und damit sofort $b_i = 0$ für alle $i \neq j$, erneut wegen der linearen Unabhängigkeit von $\mathcal{B}$. Also ist $\mathcal{B}'$ linear unabhängig. □

2.17 Satz. (Austauschsatz von Steinitz[1]) *Sei $(v_1, \ldots, v_n)$ eine Basis von V, und sei $(w_1, \ldots, w_r)$ eine linear unabhängige Familie in V. Dann ist $r \leq n$, und es gibt $n - r$ Indices $i_{r+1}, \ldots, i_n \in \{1, \ldots, n\}$ so, daß auch $(w_1, \ldots, w_r, v_{i_{r+1}}, \ldots, v_{i_n})$ eine Basis von V ist.*

Man kann also r der v_i so auswählen, daß durch Austausch dieser r Vektoren gegen $w_1, \ldots, w_r$ wieder eine Basis entsteht.

Beweis. Induktion nach r. Für $r = 0$ ist nichts zu zeigen. Sei $r \geq 1$, sei der Satz für $r - 1$ schon bewiesen. Den schon bewiesenen Fall $r - 1$ können wir auf die (linear unabhängige) Familie $(w_1, \ldots, w_{r-1})$ anwenden. Es folgt $r - 1 \leq n$ und

[1] Ernst Steinitz (1871–1928)

nach geeigneter Umnumerierung der v_i, daß auch $\mathcal{B} := (w_1, \ldots, w_{r-1}, v_r, \ldots, v_n)$ eine Basis von V ist. Drücke w_r durch $\mathcal{B}$ aus, etwa

$$w_r = \sum_{i=1}^{r-1} a_i w_i + \sum_{i=r}^{n} a_i v_i$$

mit $a_i \in K$. Dann gibt es ein $i \in \{r, \ldots, n\}$ mit $a_i \neq 0$ (und insbesondere ist $r \leq n$). Andernfalls wäre nämlich $w_r \in \operatorname{span}(w_1, \ldots, w_{r-1})$, Widerspruch zur linearen Unabhängigkeit von $w_1, \ldots, w_r$ (2.10). Für dieses i können wir v_i in $\mathcal{B}$ durch w_r ersetzen und erhalten nach Lemma 2.16 wieder eine Basis. Damit ist der Induktionsschritt gezeigt und der Satz bewiesen. □

2.18 Korollar. *Sei V ein endlich erzeugter Vektorraum. Dann sind alle Basen von V endlich und haben dieselbe Länge.*

Die *Länge* $|\mathcal{F}|$ einer Familie $\mathcal{F} = (x_i)_{i \in I}$ ist dabei definiert als die Kardinalität von I, also $|\mathcal{F}| := |I| \in \mathbb{N}_0 \cup \{\infty\}$.

Beweis. Nach Satz 2.15 hat V eine endliche Basis $\mathcal{B}$. Sei $\mathcal{B}'$ eine weitere Basis von V. Aus Satz 2.16 folgt $|\mathcal{B}'| \leq |\mathcal{B}|$. Also sind alle Basen endlich. Wir können die Rollen von $\mathcal{B}$ und $\mathcal{B}'$ vertauschen und sehen $|\mathcal{B}| = |\mathcal{B}'|$. □

2.19 Korollar. (Basisergänzungssatz) *Sei V endlich erzeugt. Jede linear unabhängige Familie in V läßt sich zu einer Basis von V ergänzen.*

Beweis. Da V eine (endliche) Basis hat, folgt das aus dem Steinitzschen Austauschsatz (2.17). □

2.20 Bemerkung. Bei richtiger Formulierung gilt der Basisauswahlsatz 2.15 auch für beliebige (nicht notwendig endlich erzeugte) Vektorräume V: Zu jedem Erzeugendensystem $(v_i)_{i \in I}$ von V gibt es eine Teilmenge $J \subseteq I$ derart, daß $(v_j)_{j \in J}$ eine Basis von V ist. Insbesondere hat jeder Vektorraum eine Basis.

Ebenso gilt der allgemeine Basisergänzungssatz: Jede linear unabhängige Familie in V läßt sich zu einer Basis ergänzen.

In dieser Vorlesung werden wir die Aussagen in dieser allgemeinen Form nicht brauchen. Für die Beweise benötigt man stärkere Methoden der Mengenlehre, das *Zornsche Lemma*[2] oder das dazu äquivalente *Auswahlaxiom*.

Nach Korollar 2.18 können wir definieren:

2.21 Definition. Sei V ein K-Vektorraum. Ist V endlich erzeugt, so definiert man die *(K-) Dimension* $\dim_K(V) = \dim(V)$ von V als Länge einer beliebigen (K-) Basis von V. Ist V nicht endlich erzeugt, so setzt man $\dim(V) := \infty$.

2.22 Bemerkungen.

1. $\dim_K(K^n) = n$ $(n \in \mathbb{N})$, $\dim_K K[t] = \infty$, siehe 2.13. Es ist $\dim_{\mathbb{C}}(\mathbb{C}) = 1$, aber $\dim_{\mathbb{R}}(\mathbb{C}) = 2$ (2.13). Die Dimension hängt also vom Grundkörper ab.

2. $\dim(V) = 0 \Leftrightarrow V = \{\mathbf{0}\}$.

[2] Max Zorn (1906–1993)

3. Nach 2.14 ist auch

$$\dim(V) \;=\; \min\{|\mathcal{F}|\colon\ \mathcal{F} \text{ Erzeugendensystem von } V\}$$

und

$$\dim(V) \;=\; \max\{|\mathcal{F}|\colon\ \mathcal{F} \text{ linear unabhängige Familie in } V\}$$

(min und max jeweils in $\mathbb{N}_0 \cup \{\infty\}$ genommen).

4. Statt "V ist endlich erzeugt" sagen wir in Zukunft "V ist endlich-dimensional", oder "$\dim(V) < \infty$".

2.23 Korollar. *Sei $\dim(V) = n < \infty$, und sei $\mathcal{F} = (v_1, \ldots, v_n)$ eine Familie derselben Länge n in V. Dann gilt:*

$$\mathcal{F} \textit{ ist eine Basis von } V \;\Leftrightarrow\; \mathcal{F} \textit{ ist linear unabhängig} \;\Leftrightarrow\; \operatorname{span}(\mathcal{F}) = V.$$

Beweis. Ist $\mathcal{F}$ linear unabhängig, dann läßt sich $\mathcal{F}$ nach dem Basisergänzungssatz 2.19 zu einer Basis $\mathcal{B}$ von V erweitern. Wegen $|\mathcal{F}| = \dim(V) = |\mathcal{B}|$ folgt $\mathcal{B} = \mathcal{F}$, d.h. $\mathcal{F}$ ist bereits eine Basis. Ist andererseits $\operatorname{span}(\mathcal{F}) = V$, so enthält $\mathcal{F}$ eine Basis nach dem Basisauswahlsatz 2.15. Wie eben folgt, daß $\mathcal{F}$ bereits selbst eine Basis ist. □

Die Aussage wird falsch, wenn man $\dim(V) = \infty$ zuläßt (finde Gegenbeispiele mit $V = K[t]$).

2.24 Korollar. *Sei V ein Vektorraum und U ein Unterraum von V.*

(a) $\dim(U) \le \dim(V)$.
(b) *Ist* $\dim(U) = \dim(V) < \infty$, *so ist* $U = V$.

Beweis. Wir können annehmen $\dim(V) = n < \infty$ (für $\dim(V) = \infty$ ist nichts zu zeigen in (a)). Sei $\mathcal{F}$ eine Basis von U. Wegen $\mathcal{F}$ linear unabhängig (in U, also auch in V) läßt sich $\mathcal{F}$ zu einer Basis von V erweitern nach Satz 2.19. Daraus folgt $\dim(U) \le \dim(V)$. Ist $\dim(U) = \dim(V)$, so ist $\mathcal{F}$ schon eine Basis von V nach Korollar 2.23, also $U = V$. □

Aussage (b) wird wieder falsch, wenn $\dim(U) = \infty$ erlaubt wird.

2.25 Sei V ein K-Vektorraum. Wir haben bisher den Nullvektor $\mathbf{0}$ in der Notation von der Null 0 im Körper unterschieden. Im allgemeinen tut man das nicht, und bezeichnet auch den Nullvektor einfach mit 0. So werden wir es in Zukunft auch halten.

3. Summen von Unterräumen

Sei stets K ein fester Körper. Alle Vektorräume sind K-Vektorräume, sofern nicht anders gesagt.

3.1 Definition. Sei V ein Vektorraum, sei $(U_i)_{i\in I}$ eine Familie von Unterräumen von V. Dann heißt der Unterraum

$$\sum_{i\in I} U_i \;:=\; \operatorname{span}\Big(\bigcup_{i\in I} U_i\Big)$$

von V die *Summe* der U_i ($i \in I$). Für $I = \{1, \ldots, n\}$ endlich schreibt man auch einfach $U_1 + \cdots + U_n := \sum_{i=1}^{n} U_i$.

3.2 Lemma.

$$\sum_{i\in I} U_i \;=\; \Big\{\sum_{i\in I} u_i \colon u_i \in U_i \;(i\in I),\; u_i = 0 \text{ f. f. a. } i\in I\Big\}.$$

Die Elemente von $\sum_{i\in I} U_i$ sind also die endlichen Summen von Vektoren aus der Vereinigungsmenge $\bigcup_{i\in I} U_i$.

BEWEIS. "$\supseteq$" ist klar. "$\subseteq$": Die rechte Menge ist ein Unterraum von V und enthält $\bigcup_{i\in I} U_i$. Nach Satz 2.3 enthält sie also auch $\sum_{i\in I} U_i = \operatorname{span}(\bigcup_{i\in I} U_i)$. □

3.3 Beispiele.

1. Für jede Familie $(v_i)_{i\in I}$ von Vektoren in V ist $\sum_{i\in I} Kv_i = \operatorname{span}(v_i \colon i\in I)$. Insbesondere ist $\operatorname{span}(v_1,\dots,v_n) = Kv_1 + \cdots + Kv_n$.

2. Die eindimensionalen Unterräume von $V = \mathbb{R}^3$ sind die $\mathbb{R}v$ $(0 \neq v \in \mathbb{R}^3)$, die Geraden durch 0. Sind $U_1 = \mathbb{R}v_1$ und $U_2 = \mathbb{R}v_2$ zwei solche Geraden mit $U_1 \neq U_2$, so ist $U_1 + U_2 = \mathbb{R}v_1 + \mathbb{R}v_2$ die von U_1 und U_2 aufgespannte Ebene im $\mathbb{R}^3$. Ist $U_3 = \mathbb{R}v_3$ eine dritte Gerade durch 0, so ist $U_1 + U_2 + U_3$ gleich $U_1 + U_2$ falls $U_3 \subseteq U_1 + U_2$, und ist gleich $\mathbb{R}^3$ andernfalls.

Im allgemeinen hat ein Vektor aus $\sum_i U_i$ viele Darstellungen der Form $\sum_i u_i$ mit $u_i \in U_i$ $(i \in I)$. Falls diese Darstellung eindeutig ist, ist das ein wichtiger Spezialfall:

3.4 Satz. *Sei $(U_i)_{i\in I}$ eine Familie von Unterräumen von V. Es sind äquivalent:*

(i) *Jedes $u \in \sum_{i\in I} U_i$ hat eine* eindeutige *Summendarstellung $u = \sum_{i\in I} u_i$ mit $u_i \in U_i$ $(i \in I)$, $u_i = 0$ f. f. a. $i \in I$;*

(ii) *für alle $j \in I$ ist $U_j \cap \sum_{i\in I\setminus\{j\}} U_i = \{0\}$.*

BEWEIS. (i) ⇒ (ii): Sei $u = \sum_{i\neq j} u_i \in U_j \cap \sum_{i\neq j} U_i$, mit $u_i \in U_i$ $(i \neq j)$, fast alle $u_i = 0$. Aus $0 = u - \sum_{i\in I\setminus\{j\}} u_i$ und (i) folgt $u = 0$ wegen $u \in U_j$.

(ii) ⇒ (i): Hat u zwei verschiedene Summendarstellungen

$$u \;=\; \sum_{i\in I} u_i \;=\; \sum_{i\in I} u_i'$$

(mit u_i, $u_i' \in U_i$, fast alle $= 0$), so hat $0 = \sum_i (u_i - u_i')$ eine nichttriviale Summendarstellung. Es genügt deshalb, zu zeigen: Aus $\sum_{i\in I} u_i = 0$ mit $u_i \in U_i$ $(i \in I)$ und $u_i = 0$ f. f. a. $i \in I$ folgt $u_i = 0$ für alle i. Sei also $\sum_{i\in I} u_i = 0$, fixiere $j \in I$. Wegen $u_j = -\sum_{i\in I\setminus\{j\}} u_i$ ist $u_j \in U_j \cap \sum_{i\neq j} U_i$. Aus (ii) folgt also $u_j = 0$. Dies für jedes $j \in I$ zeigt die Behauptung. □

3.5 Definition. Sind die Bedingungen aus Satz 3.4 erfüllt, so heißt die Summe $\sum_{i\in I} U_i$ *direkt*, oder die *(interne) direkte Summe* der U_i. Man schreibt dann (und nur dann!) auch

$$\bigoplus_{i\in I} U_i \;:=\; \sum_{i\in I} U_i,$$

um die Direktheit der Summe auszudrücken.

Im Spezialfall $|I| = 2$ wird Satz 3.4 besonders einfach:

3.6 Korollar. *Seien U_1, U_2 Unterräume von V. Genau dann ist $V = U_1 \oplus U_2$, wenn $U_1 + U_2 = V$ und $U_1 \cap U_2 = \{0\}$ ist.* □

3.7 Beispiel. Sei $\mathcal{B} = (v_i)_{i\in I}$ eine Basis von V, sei die Indexmenge I zerlegt als $I = I_1 \cup I_2$ mit $I_1 \cap I_2 = \varnothing$. Dann ist $V = U_1 \oplus U_2$ mit $U_\nu = \operatorname{span}(v_i \colon i \in I_\nu)$ $(\nu = 1, 2)$. Denn $U_1 + U_2 = V$ ist klar, und $U_1 \cap U_2 = \{0\}$ wegen $\mathcal{B}$ linear unabhängig. Umgekehrt gilt:

3.8 Satz. *Sei $V = U_1 \oplus U_2$. Für $\nu = 1, 2$ sei $(v_i \colon i \in I_\nu)$ eine Basis von U_ν (mit $I_1 \cap I_2 = \varnothing$). Dann ist $(v_i \colon i \in I_1 \cup I_2)$ eine Basis von V. Insbesondere ist* $\dim(V) = \dim(U_1) + \dim(U_2)$.

Beweis. Wegen $U_1 + U_2 = V$ ist $\operatorname{span}(v_i \colon i \in I_1 \cup I_2) = V$. Sei $\sum_{i\in I_1\cup I_2} a_i v_i = 0$ mit $a_i \in K$ (und f. a. $a_i = 0$). Dann ist

$$\sum_{i\in I_1} a_i v_i \;=\; -\sum_{i\in I_2} a_i v_i \;\in U_1 \cap U_2 \;=\; \{0\}.$$

Wegen der linearen Unabhängigkeit von $(v_i \colon i \in I_\nu)$ für $\nu = 1, 2$ folgt $a_i = 0$ für alle $i \in I_1$ und alle $i \in I_2$. □

Induktiv folgt daraus auch $\dim\Big(\bigoplus_{i=1}^r U_i\Big) = \sum_{i=1}^r \dim(U_i)$ für alle $r \in \mathbb{N}$.

3.9 Satz. *Sei V ein Vektorraum mit $\dim(V) < \infty$, sei U ein Unterraum von V. Dann gibt es einen Unterraum W von V mit $V = U \oplus W$. Jedes solche W heißt ein* lineares Komplement *von U (oder ein zu U* komplementärer Unterraum*) in V, und erfüllt* $\dim(W) = \dim(V) - \dim(U)$.

Beweis. Ergänze eine Basis $(u_1, \dots, u_r)$ von U zu einer Basis $\mathcal{B} = (u_1, \dots, u_r, w_1, \dots, w_s)$ von V (Korollar 2.19) und nimm $W := \operatorname{span}(w_1, \dots, w_s)$. Nach Beispiel 3.7 ist dann $U \oplus W = V$. Die Dimensionsaussage gilt nach 3.8. □

3.10 Bemerkungen.

1. Im allgemeinen gibt es zu einem Unterraum U von V viele lineare Komplemente. Ist etwa $V = \mathbb{R}^2$ und $U \subseteq \mathbb{R}^2$ eine Gerade durch 0, so ist jede Gerade $U' \neq U$ durch 0 ein lineares Komplement zu U.

2. *Vorsicht* bei der direkten Summe: Sind $U_1, \dots, U_r$ Unterräume von V, und gilt $U_i \cap U_j = \{0\}$ für alle $i \neq j$, so folgt daraus für $r \geq 3$ noch *nicht*, daß die Summe $U_1 + \dots + U_r$ direkt ist! (Betrachte etwa r verschiedene Ursprungsgeraden in $\mathbb{R}^2$.)

3.11 Definition. Sei $(V_i)_{i\in I}$ eine Familie von Vektorräumen. Die *(externe) direkte Summe* der V_i ist der Unterraum

$$\bigoplus_{i\in I} V_i \;:=\; \Big\{(v_i)_{i\in I} \in \prod_{i\in I} V_i \colon\; v_i = 0 \text{ für fast alle } i \in I\Big\}$$

von $\prod_i V_i$, vergleiche Beispiel 1.9.5. (Es sind also Addition und Multiplikation mit Skalaren komponentenweise definiert.)

3.12 Bemerkungen.

1. Für endlich viele Vektorräume sind (externe) direkte Summe und direktes Produkt dasselbe:

$$V_1 \oplus \cdots \oplus V_n \;=\; V_1 \times \cdots \times V_n.$$

Für $|I| = \infty$ ist dagegen i.a.

$$\bigoplus_{i \in I} V_i \;\neq\; \prod_{i \in I} V_i.$$

2. Verwendung derselben Notation für interne und externe direkte Summe ist harmlos, wie wir bald sehen werden.

KAPITEL III

Lineare Abbildungen, Matrizen, lineare Gleichungssysteme

1. Matrizen

Im folgenden sei R stets ein *kommutativer* Ring.

1.1 Definition. Seien m, $n \in \mathbb{N}$. Eine $m \times n$-*Matrix über* R ist eine Abbildung $A\colon \{1,\dots,m\} \times \{1,\dots,n\} \to R$, $(i,j) \mapsto a_{ij}$. Man schreibt A als rechteckiges Schema aus m Zeilen und n Spalten:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Die Einträge a_{ij} heißen die *Koeffizienten* der Matrix A, das Element a_{ij} steht in der i-ten Zeile und der j-ten Spalte. Man schreibt die obige Matrix auch in der Form

$$A = (a_{ij})_{\substack{1\le i\le m \\ 1\le j\le n}}$$

(oder auch einfach als $A = (a_{ij})_{i,j} = (a_{ij})$, wenn das Format $m \times n$ klar ist). Die Menge aller $m \times n$-Matrizen über R wird mit $\mathrm{M}_{m\times n}(R)$ bezeichnet. Ist $m = n$, so schreibt man auch $\mathrm{M}_n(R) := \mathrm{M}_{n\times n}(R)$.

1.2 Bemerkungen.

1. Eine $1 \times n$-Matrix hat die Form $(a_1,\dots,a_n)$, und heißt auch ein *Zeilenvektor*. (Für bessere Lesbarkeit benutzt man bei Zeilenvektoren meistens Kommas.) Eine $m \times 1$-Matrix hat die Form

$$\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$$

und heißt auch ein *Spaltenvektor*. 1×1-Matrizen werden mit Elementen aus R identifiziert, man kann die Klammern weglassen.

2. Eine $m \times n$-Matrix heißt *quadratisch*, wenn $m = n$ ist. Eine quadratische Matrix $A = (a_{ij})$ heißt eine *obere* (bzw. *untere*) *Dreiecksmatrix*, falls $a_{ij} = 0$ ist für alle $i > j$ (bzw. für alle $i < j$):

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & * \end{pmatrix} \text{ obere,} \qquad \begin{pmatrix} * & 0 & \cdots & 0 \\ * & * & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{pmatrix} \text{ untere Dreiecksmatrix}$$

($*$ steht für irgendwelche Elemente im Ring R). Ist A sowohl obere wie untere Dreiecksmatrix, so heißt A eine *Diagonalmatrix*; man schreibt für $a_1, \ldots, a_n \in R$

$$\operatorname{diag}(a_1, \ldots, a_n) \;:=\; (\delta_{ij} a_i)_{1 \le i,j \le n} \;=\; \begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_n \end{pmatrix} \;\in\; \mathrm{M}_n(R).$$

1.3 Definition. Seien $A = (a_{ij})_{i,j}$, $B = (b_{ij})_{i,j} \in \mathrm{M}_{m\times n}(R)$ und $c \in R$. Man definiert

$$A + B \;:=\; (a_{ij} + b_{ij})_{i,j}, \quad cA \;:=\; (ca_{ij})_{i,j}.$$

Dies sind wieder Matrizen in $\mathrm{M}_{m\times n}(R)$.

Die Summe zweier Matrizen ist nur dann definiert, wenn beide dasselbe Format haben. Das folgende Lemma ist klar:

1.4 Lemma.

(a) $(\mathrm{M}_{m\times n}(R), +)$ *ist eine abelsche Gruppe.*

(b) *Ist $R = K$ ein Körper, so ist $\mathrm{M}_{m\times n}(K)$ mit den in 1.3 definierten Operationen ein K-Vektorraum von Dimension mn.* □

1.5 Definition. Seien m, n, $r \in \mathbb{N}$. Für

$$A \;=\; (a_{ij})_{\substack{1\le i\le m\\ 1\le j\le n}} \in \mathrm{M}_{m\times n}(R), \quad B \;=\; (b_{jk})_{\substack{1\le j\le n\\ 1\le k\le r}} \in \mathrm{M}_{n\times r}(R)$$

definiert man das *Produkt* $A \cdot B = AB$ als die $m \times r$-Matrix $AB = (c_{ik})_{1\le i\le m,\, 1\le k\le r}$ mit

$$c_{ik} \;:=\; \sum_{j=1}^{n} a_{ij} b_{jk} \quad (1 \le i \le m,\; 1 \le k \le r).$$

1.6 Bemerkungen.

1. Das Produkt AB zweier Matrizen A und B ist nur definiert, wenn B genau so viele Zeilen wie A Spalten hat.

2. Es ist

$$(x_1, \ldots, x_n) \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \;=\; \big(x_1 y_1 + \cdots + x_n y_n\big)$$

eine 1×1-Matrix, aber

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \cdot (y_1 \;\ldots\; y_n) \;=\; \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & & \vdots \\ x_n y_1 & \cdots & x_n y_n \end{pmatrix}$$

ist eine $n \times n$-Matrix.

3. Für $n \in \mathbb{N}$ heißt

$$I_n \;=\; \operatorname{diag}(1, \ldots, 1) \;=\; \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \;\in\; \mathrm{M}_n(R)$$

die $n \times n$ *Einheitsmatrix*.

1.7 Lemma. *Seien* A, $A' \in \mathrm{M}_{m\times n}(R)$, B, $B' \in \mathrm{M}_{n\times r}(R)$, $C \in \mathrm{M}_{r\times s}(R)$.
(a) $(AB)C = A(BC)$ *(Assoziativität der Multiplikation),*
(b) $A(B+B') = AB + AB'$, $(A+A')B = AB + A'B$ *(Distributivität),*
(c) $I_m \cdot A = A = A \cdot I_n$,
(d) $a(AB) = (aA)B = A(aB)$ *für* $a \in R$.

Beweis. Einfaches Nachrechnen. Wir zeigen (a): Sei $A = (a_{ij})$, $B = (b_{jk})$, $C = (c_{kl})$. Dann ist $(AB)C = (d_{il})$ die $m \times s$-Matrix mit

$$d_{il} = \sum_{k=1}^{r} \Big(\sum_{j=1}^{n} a_{ij} b_{jk} \Big) c_{kl},$$

andererseits $A(BC) = (d'_{il}) \in \mathrm{M}_{m\times s}(R)$ mit

$$d'_{il} = \sum_{j=1}^{n} a_{ij} \Big(\sum_{k=1}^{r} b_{jk} c_{kl} \Big).$$

Man sieht $d_{il} = d'_{il}$. Also gilt $(AB)C = A(BC)$. □

1.8 Notation. Für feste m, $n \in \mathbb{N}$ und für $1 \le i \le m$, $1 \le j \le n$ definieren wir die Matrix $E_{ij} \in \mathrm{M}_{m\times n}(R)$ durch

$$E_{ij} \;:=\; \Big(\delta_{ki}\delta_{lj} \Big)_{\substack{1\le k\le m \\ 1\le l\le n}}.$$

Es ist also $E_{ij} = (a_{kl})$ mit $a_{kl} = 1$ für $(k,l) = (i,j)$ und $a_{kl} = 0$ für $(k,l) \neq (i,j)$.

1.9 Bemerkungen.

1. Für $A = (a_{ij}) \in \mathrm{M}_{m\times n}(R)$ ist $A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} E_{ij}$. Ist $R = K$ ein Körper, so ist $(E_{ij})_{1\le i\le m,\ 1\le j\le n}$ eine Basis des K-Vektorraums $\mathrm{M}_{m\times n}(K)$.

2. Für Produkte der Matrizen E_{ij} gilt

$$E_{ij} \cdot E_{kl} \;=\; \delta_{jk}\, E_{il} \;=\; \begin{cases} E_{il} & \text{falls } j = k, \\ 0 & \text{falls } j \neq k. \end{cases}$$

1.10 Satz. *Für* $n \in \mathbb{N}$ *ist* $(\mathrm{M}_n(R), +, \cdot)$ *ein Ring. Ist dabei* $R \neq \{0\}$*, so ist dieser Ring für* $n \ge 2$ *nicht kommutativ und hat Nullteiler.*

(Erstes Beispiel eines nicht kommutativen Rings in dieser Vorlesung)

Beweis. Daß $\mathrm{M}_n(R)$ ein Ring ist, folgt sofort aus 1.7. Die Null ist die Nullmatrix 0, die Eins ist die Einheitsmatrix I_n. Ist $n \ge 2$, so ist $E_{11} \cdot E_{12} = E_{12}$, aber $E_{12} \cdot E_{11} = 0$, woraus man die beiden weiteren Behauptungen sieht. □

1.11 Definition. Für $A = (a_{ij})_{1\le i\le m,\ 1\le j\le n} \in \mathrm{M}_{m\times n}(R)$ heißt

$$A^t \;:=\; \big(a_{ji}\big)_{1\le j\le n,\ 1\le i\le m} \;\in \mathrm{M}_{n\times m}(R)$$

die zu A *transponierte Matrix.*

1.12 Satz. *Seien* A, A_1, $A_2 \in \mathrm{M}_{m\times n}(R)$ *und* $B \in \mathrm{M}_{n\times r}(R)$.
(a) $(a_1A_1 + a_2A_2)^t = a_1A_1^t + a_2A_2^t$ *für* a_1, $a_2 \in R$.
(b) $(A^t)^t = A$.

(c) $(AB)^t = B^t A^t$.

BEWEIS. (a), (b) sind klar, und (c) zeigt man durch Nachrechnen: Ist $A = (a_{ij})$, $B = (b_{kl})$, so ist der Eintrag von $(AB)^t$ an der Stelle (i, j) gleich

$$\sum_k a_{jk} b_{ki} \;=\; \sum_k b_{ki} a_{jk},$$

und das ist auch der Eintrag von $B^t A^t$ an dieser Stelle. □

Wir werden in dieser Vorlesung meistens mit Matrizen über Körpern arbeiten. Sei ab jetzt $R = K$ stets ein Körper.

1.13 Definition. Eine Matrix $A \in \mathrm{M}_n(K)$ heißt *invertierbar* (oder *regulär*), wenn es eine Matrix $A' \in \mathrm{M}_n(K)$ gibt mit $AA' = A'A = I_n$. (Andernfalls heißt A *singulär*.)

1.14 Satz und Definition. *Sei $A \in \mathrm{M}_n(K)$ invertierbar. Die Matrix $A' \in \mathrm{M}_n(K)$ mit $AA' = A'A = I_n$ ist eindeutig bestimmt und heißt die zu A* inverse Matrix*, i. Z. $A' =: A^{-1}$. Die Menge*

$$\mathrm{GL}_n(K) \;:=\; \{A \in \mathrm{M}_n(K) \colon\; A \textit{ ist invertierbar}\}$$

ist eine Gruppe (bezüglich Matrixprodukt), genannt die allgemeine lineare Gruppe *über K. Für $n \geq 2$ ist sie nicht abelsch.*

BEWEIS. Aus $A'A = I_n = AA''$ folgt $A'' = I_n A'' = (A'A)A'' = A'(AA'') = A'I_n = A'$, woraus die erste Aussage folgt. Für die anderen Aussagen siehe Aufgabe 22. □

1.15 Beispiel. Seien $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, $\widetilde{A} = \left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right) \in \mathrm{M}_2(K)$. Dann ist $A\widetilde{A} = \widetilde{A}A = (ad - bc)\, I_2$. Dabei heißt $\det(A) := ad - bc$ die *Determinante* von A. Für A, $B \in M_2(K)$ ist $\det(AB) = \det(A)\,\det(B)$ (nachrechnen!). (Wir werden die Aussage bald in größerer Allgemeinheit formulieren und beweisen.)

1.16 Korollar. *Die Matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{M}_2(K)$ ist genau dann invertierbar, wenn $\det(A) \neq 0$ ist. Alsdann ist*

$$A^{-1} \;=\; \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

BEWEIS. Für $\det(A) \neq 0$ folgt das aus 1.15. Für $\det(A) = 0$ und jede Matrix $B \in \mathrm{M}_2(K)$ ist auch $\det(AB) = 0$, nach 1.15, und insbesondere $AB \neq I_2$ wegen $\det(I_2) = 1$. Also ist A nicht invertierbar. □

2. Homomorphismen von Gruppen und Ringen

Wir betrachten im folgenden Gruppen, die wir meist multiplikativ schreiben, also $(G, \cdot)$, $(H, \cdot)$ usw., mit neutralem Element e.

2.1 Definition. Seien G, H Gruppen.

(a) Eine Abbildung $\varphi \colon G \to H$ heißt *(Gruppen-) Homomorphismus* (oder *Homomorphismus von Gruppen*), falls für alle g_1, $g_2 \in G$ gilt $\varphi(g_1 g_2) = \varphi(g_1)\,\varphi(g_2)$.

(b) Ein bijektiver Homomorphismus heißt auch ein *Isomorphismus*.
(c) G und H heißen *isomorph*, i. Z. $G \cong H$, wenn es einen Isomorphismus $G \to H$ gibt.

2.2 Beispiele.

1. Die Abbildung $\det\colon \mathrm{GL}_2(K) \to K^*$, $A \mapsto \det(A)$ ist ein Gruppenhomomorphismus nach Beispiel 1.15. Tatsächlich ist det surjektiv, z. B. wegen $\det \left(\begin{smallmatrix} a & 0 \\ 0 & 1 \end{smallmatrix}\right) = a$ für $a \in K$.

2. Die Exponentialfunktion $\exp\colon \mathbb{R} \to \mathbb{R}_+^*$, $x \mapsto e^x$ ist ein Gruppenhomomorphismus von $(\mathbb{R}, +)$ nach $(\mathbb{R}_+^*, \cdot)$, denn $e^{x+y} = e^x \cdot e^y$ gilt für alle $x,\ y \in \mathbb{R}$. Die Abbildung ist sogar bijektiv, also ein Gruppenisomorphismus.

3. Ist R ein Ring und $a \in R$ ein festes Element, so ist die Abbildung $R \to R$, $x \mapsto ax$ ein Homomorphismus der Gruppe $(R, +)$ in sich, denn $a(x + x') = ax + ax'$ gilt für alle x, $x' \in R$ (Distributivgesetz). Dieselbe Aussage gilt für die Abbildung $x \mapsto xa$. Analog lassen sich die Distributivgesetze für Vektorräume (II.1.2) deuten.

2.3 Lemma. *Sei $\varphi\colon G \to H$ ein Homomorphismus von Gruppen. Dann gilt:*

(a) $\varphi(e) = e$.
(b) $\varphi(g^{-1}) = \varphi(g)^{-1}$ *für alle* $g \in G$.
(c) *Für* $n \in \mathbb{N}$ *und* $g_1, \ldots, g_n \in G$ *ist* $\varphi(g_1 \cdots g_n) = \varphi(g_1) \cdots \varphi(g_n)$.

Beweis. (a) folgt aus $\varphi(e)^2 = \varphi(e^2) = \varphi(e) = \varphi(e) \cdot e$ durch Kürzen in H, und (b) folgt aus (a) wegen $\varphi(g)\varphi(g^{-1}) = \varphi(gg^{-1}) = \varphi(e) = e$. (c) folgt durch Induktion nach n. □

2.4 Lemma. *Seien G_1, G_2, G_3 Gruppen.*

(a) *Sind $\varphi\colon G_1 \to G_2$ und $\psi\colon G_2 \to G_3$ Homomorphismen, so ist auch $\psi \circ \varphi\colon G_1 \to G_3$ ein Homomorphismus.*
(b) *Ist $\varphi\colon G_1 \to G_2$ ein Isomorphismus, so ist auch die Umkehrabbildung $\varphi^{-1}\colon G_2 \to G_1$ ein Isomorphismus.*
(c) *Es gilt*

$$G_1 \cong G_1 \quad \textit{(Reflexivität)},$$
$$G_1 \cong G_2 \ \Rightarrow\ G_2 \cong G_1 \quad \textit{(Symmetrie)},$$
$$G_1 \cong G_2 \wedge G_2 \cong G_3 \ \Rightarrow\ G_1 \cong G_3 \quad \textit{(Transitivität)}.$$

Beweis. (a) $(\psi \circ \varphi)(gg') = \psi(\varphi(gg')) = \psi(\varphi(g)\varphi(g')) = \psi(\varphi(g))\,\psi(\varphi(g')) = \psi \circ \varphi(g)\,\psi \circ \varphi(g')$. (b) Sind h, $h' \in G_2$, und ist $g := \varphi^{-1}(h)$, $g' := \varphi^{-1}(h')$, so ist $\varphi(gg') = \varphi(g)\varphi(g') = hh'$, also $gg' = \varphi^{-1}(hh')$. (c) folgt aus (a) und (b). □

Ein Isomorphismus $G \to H$ ist nichts anderes als eine Umbenennung der Elemente von G, ohne daß dabei die Gruppenstruktur (die Multiplikationstabelle) geändert würde. Vom Standpunkt der Algebra sind isomorphe Gruppen daher im wesentlichen gleich, nämlich gleich bis auf Umbenennung der Elemente. Wir werden Isomorphismen vieler weiterer Strukturen (z. B. Ringe, Vektorräume) kennenlernen, stets gilt dann die analoge Bemerkung.

2.5 Definition. Sei $\varphi\colon G \to H$ ein Gruppenhomomorphismus. Der *Kern* von φ ist definiert als

$$\ker(\varphi) \;:=\; \varphi^{-1}(\{e\}) \;=\; \{x \in G\colon \varphi(x) = e\}.$$

2.6 Satz. *Sei $\varphi\colon G \to H$ ein Gruppenhomomorphismus.*

(a) $\ker(\varphi)$ *ist eine Untergruppe von G. Für alle $g \in G$ und $x \in \ker(\varphi)$ gilt $gxg^{-1} \in \ker(\varphi)$.*

(b) *φ ist injektiv* $\Leftrightarrow$ $\ker(\varphi) = \{e\}$.

Beweis. Wir setzen $N := \ker(\varphi)$. Es gilt $e \in N$, und für $x, y \in N$ ist $\varphi(xy^{-1}) = \varphi(x)\varphi(y)^{-1} = e$, also $xy^{-1} \in N$. Nach I.2.13 ist N eine Untergruppe von G. Für $x \in N$ und $g \in G$ ist weiter

$$\varphi(gxg^{-1}) \;=\; \varphi(g)\varphi(x)\varphi(g)^{-1} \;=\; \varphi(g)\,e\,\varphi(g)^{-1} \;=\; e,$$

also $gxg^{-1} \in N$. Damit ist (a) gezeigt. In (b) ist "$\Rightarrow$" klar. Umgekehrt sei $N = \{e\}$. Sind $x, y \in G$ mit $\varphi(x) = \varphi(y)$, so folgt $e = \varphi(x)\varphi(y)^{-1} = \varphi(xy^{-1})$, also $xy^{-1} \in N$, also $x = y$. Also ist φ injektiv. □

2.7 Definition. Sei G eine Gruppe. Eine Untergruppe N von G heißt ein *Normalteiler* (oder eine *normale Untergruppe*) von G, geschrieben $N \trianglelefteq G$, wenn gilt:

$$\forall\, x \in N\ \forall\, g \in G\ \ gxg^{-1} \in N.$$

2.8 Beispiele.

1. Für jeden Homomorphismus $\varphi\colon G \to H$ gilt $\ker(\varphi) \trianglelefteq G$ (Satz 2.6(a)).

2. Der Kern des Homomorphismus $\det\colon \mathrm{GL}_2(K) \to K^*$ (2.2.1) ist

$$\mathrm{SL}_2(K) \;:=\; \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in K,\ ad - bc = 1 \right\},$$

genannt die *spezielle lineare Gruppe* der 2×2-Matrizen. Es ist also $\mathrm{SL}_2(K) \trianglelefteq \mathrm{GL}_2(K)$.

3. Ist die Gruppe $(G, \cdot)$ abelsch, so ist jede Untergruppe von G ein Normalteiler. Denn $gxg^{-1} = x$ für alle $g, x \in G$.

4. Weitere Beispiele gibt es in den Übungen (Aufgabe 23).

2.9 Definition. Seien A, B Ringe. Eine Abbildung $\varphi\colon A \to B$ heißt ein *Ringhomomorphismus*, wenn $\varphi(1) = 1$ ist, und wenn für alle $a, b \in A$ gilt

$$\varphi(a + b) \;=\; \varphi(a) + \varphi(b), \quad \varphi(ab) \;=\; \varphi(a)\,\varphi(b).$$

Ist dabei φ bijektiv, so nennt man φ einen *Ringisomorphismus*. Zwei Ringe A und B heißen *isomorph*, i. Z. $A \cong B$, wenn es einen Ringisomorphismus $A \to B$ gibt.

2.10 Bemerkungen.

1. Sei K ein Körper, sei $c \in K$. Die Abbildung $K[t] \to K$, $f(t) \mapsto f(c)$ (Auswertung in c, siehe I.4.6) ist ein surjektiver Ringhomomorphismus. Dasselbe gilt für $n \in \mathbb{N}$ und die Abbildung $\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, $a \mapsto \overline{a} = a + n\mathbb{Z}$.

2. Die zu Lemma 2.4 analogen Aussagen gelten mit analogen Begründungen auch für Ringhomomorphismen.

3. Lineare Abbildungen

Sei K stets ein Körper.

3.1 Definition. Seien V, W K-Vektorräume. Eine Abbildung $f\colon V \to W$ heißt *(K-) linear*, oder ein *Homomorphismus von (K-) Vektorräumen*, wenn für alle $v, v' \in V$ und $a \in K$ gilt

(1) $f(v+v') = f(v) + f(v')$ (*Additivität*),
(2) $f(av) = af(v)$ (*Homogenität*).

Ist dabei f bijektiv (bzw. injektiv bzw. surjektiv), so heißt f auch ein *Isomorphismus* (bzw. *Monomorphismus* bzw. *Epimorphismus*). Gibt es einen Isomorphismus $V \to W$, so heißen die Vektorräume V und W *isomorph*, i. Z. $V \cong W$. Mit $\mathrm{Hom}_K(V,W) = \mathrm{Hom}(V,W)$ wird die Menge aller K-linearen Abbildungen $V \to W$ bezeichnet. Für $V = W$ heißen die linearen Abbildungen $V \to V$ auch *Endomorphismen* von V, und man schreibt $\mathrm{End}(V) = \mathrm{End}_K(V) := \mathrm{Hom}_K(V,V)$.

3.2 Bemerkungen und Beispiele.

1. Analog zu 2.4 ist die Komposition zweier linearer Abbildungen wieder linear, ebenso ist die Umkehrabbildung einer bijektiven linearen Abbildung wieder linear. Für beliebige Vektorräume V, W ist die Nullabbildung $V \to W$, $v \mapsto 0$ ($v \in V$) linear.

2. Sei $f\colon V \to W$ eine lineare Abbildung. Für jeden Unterraum $V' \subseteq V$ ist

$$f(V') = \{f(v)\colon v \in V'\} = \{w \in W\colon \exists\, v \in V'\ f(v) = w\},$$

die Bildmenge von V', ein Unterraum von W. Das folgt sofort aus dem Kriterium II.1.7 für Unterräume. Insbesondere ist $\mathrm{im}(f) := f(V)$ (das *Bild* von f) ein Unterraum von W. Für jeden Unterraum W' von W ist

$$f^{-1}(W') = \{v \in V\colon f(v) \in W'\},$$

die Urbildmenge von W', ein Unterraum von V. Insbesondere ist $\ker(f) := f^{-1}(\{0\})$ (der *Kern* von f) ein Unterraum von V. Genau dann ist f injektiv, wenn $\ker(f) = \{0\}$ ist (Satz 2.6(b)).

3. Sei V ein Vektorraum, seien Vektoren $v_1, \dots, v_n \in V$ fixiert. Die Abbildung

$$f\colon K^n \to V, \quad f(x_1, \dots, x_n) = \sum_{i=1}^{n} x_i v_i$$

ist linear. Genau dann ist f injektiv (bzw. surjektiv), wenn $(v_1, \dots, v_n)$ linear unabhängig (bzw. $\mathrm{span}(v_1, \dots, v_n) = V$) ist (Beweis!).

4. Für jedes Polynom $f \in \mathbb{R}[t]$ sei $f' = \frac{d}{dt} f(t)$ die *Ableitung* von f, also

$$\Big(\sum_{i=0}^{n} a_i t^i\Big)' = \sum_{i=1}^{n} i a_i t^{i-1}$$

($a_i \in \mathbb{R}$). Die Abbildung $\mathbb{R}[t] \mapsto \mathbb{R}[t]$, $f \mapsto f'$ ist $\mathbb{R}$-linear.

3.3 Satz. *Seien V, W K-Vektorräume.*

(a) *Die Menge $\mathrm{Hom}_K(V,W)$ wird durch*

$$(f+g)(v) := f(v) + g(v), \quad (af)(v) := af(v)$$

($f, g \in \mathrm{Hom}_K(V,W)$, $a \in K$, $v \in V$) selbst zu einem K-Vektorraum.

(b) $\big(\mathrm{End}(V),\, +,\, \circ\big)$ *ist ein Ring mit der Eins* id_V. *(Hier bedeutet* $\circ$ *die Komposition von Abbildungen.)*

BEWEIS. (a) Man muß sich zuerst vergewissern, daß die oben definierten Abbildungen $f+g$ und af wieder linear sind. Der Nachweis der Vektorraumaxiome für $\mathrm{Hom}(V,W)$ ist dann Routine. Dabei benutzt man die Vektorraumaxiome für V und für W. Die Null im Vektorraum $\mathrm{Hom}(V,W)$ ist die Nullabbildung $V \to W$, $v \mapsto 0$ $(v \in V)$.

(b) Nach (a) ist $(\mathrm{End}(V), +)$ eine abelsche Gruppe, und Komposition ist eine assoziative Verknüpfung auf $\mathrm{End}(V)$ (vgl. 3.2.1). Es gelten auch beide Distributivgesetze: Denn $(f+g)\circ h = f\circ h + g\circ h$ ist trivial nach Definition der Summe, und $f\circ(g+h) = f\circ g + f\circ h$ folgt aus der Additivität von f. □

3.4 Satz. *Seien* V, W K*-Vektorräume, sei* $(v_i)_{i\in I}$ *eine Basis von* V *und* $(w_i)_{i\in I}$ *eine beliebige Familie von Vektoren in* W *(indiziert mit derselben Indexmenge* I*). Dann gibt es genau eine lineare Abbildung* $f\colon V \to W$ *mit* $f(v_i) = w_i$ *für alle* $i \in I$.

Eine lineare Abbildung $V \to W$ ist also durch ihre Werte auf einer Basis von V eindeutig bestimmt, und umgekehrt kann man diese Werte beliebig vorschreiben. Insbesondere kennt man eine lineare Abbildung $V \to W$, sobald man ihre Werte auf einer Basis von V kennt.

BEWEIS. Jeder Vektor $v \in V$ hat eine eindeutige Darstellung $v = \sum_{i\in I} a_i v_i$ (mit $a_i \in K$ und $a_i = 0$ f. f. a. $i \in I$). Wir *müssen* also definieren

$$f(v) \;:=\; \sum_{i\in I} a_i w_i.$$

Umgekehrt ist klar, daß die hierdurch definierte Abbildung f linear ist. □

3.5 Bemerkung. (Wichtig!) Jede Matrix $A = (a_{ij}) \in \mathrm{M}_{m\times n}(K)$ definiert eine K-lineare Abbildung $F_A\colon\ K^n \to K^m$ durch $F_A(y) = Ay$ für $y \in K^n$, also

$$F_A:\ \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \;\mapsto\; \begin{pmatrix} a_{11}y_1 + \cdots + a_{1n}y_n \\ \vdots \\ a_{m1}y_1 + \cdots + a_{mn}y_n \end{pmatrix}$$

Die Linearität der Abbildung F_A folgt aus den Regeln 1.7. Wir fassen hier die Elemente von K^m oder K^n als *Spaltenvektoren* auf. Das werden wir in Zukunft immer tun, d. h. wir machen folgende generelle Konvention:

Vektoren aus K^n sind Spaltenvektoren!

Die Spalten der Matrix A sind gerade $F_A(e_1), \ldots, F_A(e_n)$, d. h. es ist

$$F_A(e_j) \;=\; \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \;=\; j\text{-te Spalte von } A \quad (j = 1, \ldots, n).$$

Die Abbildung $\mathrm{M}_{m\times n}(K) \to \mathrm{Hom}_K(K^n, K^m)$ ist K-linear, d.h. für $A,\ A' \in \mathrm{M}_{m\times n}(K)$ und $a \in K$ gilt $F_{A+A'} = F_A + F_{A'}$ und $F_{cA} = cF_A$. Außerdem gilt:

3.6 Lemma. *Seien $A \in \mathrm{M}_{m\times n}(K)$, $B \in \mathrm{M}_{n\times r}(K)$, seien $F_A\colon K^n \to K^m$ und $F_B\colon K^r \to K^n$ die zugehörigen linearen Abbildungen. Dann ist $F_{AB} = F_A \circ F_B$ (als lineare Abbildungen $K^r \to K^m$).*

Beweis. Sei $A = (a_{ij})$, $B = (b_{jk})$. Für $k = 1, \dots, r$ ist

$$F_{AB}(e_k) \;=\; \sum_{i=1}^{m} (AB)_{ik}\, e_i \;=\; \sum_{i=1}^{m} \Big(\sum_{j=1}^{n} a_{ij} b_{jk} \Big)\, e_i,$$

andererseits

$$(F_A \circ F_B)(e_k) \;=\; F_A\Big(\sum_{j=1}^{n} b_{jk} e_j \Big) \;=\; \sum_{j=1}^{n} b_{jk} \Big(\sum_{i=1}^{m} a_{ij} e_i \Big).$$

Beide sind gleich, also folgt $F_{AB} = F_A \circ F_B$ aus Satz 3.4. □

Wir fassen die Diskussion zusammen:

3.7 Satz.

(a) *Die Abbildung $\mathrm{M}_{m\times n}(K) \to \mathrm{Hom}_K(K^n, K^m)$, $A \mapsto F_A$ (siehe 3.5) ist ein Isomorphismus der Vektorräume.*

(b) *Für $m = n$ ist $\mathrm{M}_n(K) \to \mathrm{End}_K(K^n)$, $A \mapsto F_A$ außerdem ein Ringisomorphismus.*

Beweis. Nach 3.5 gilt $F_{A+B} = F_A + F_B$ und $F_{aA} = aF_A$, also ist die Abbildung linear. Die Spalten von A sind die Vektoren $F_A(e_1), \dots, F_A(e_n)$ (3.5). Nach Satz 3.4 ist die Abbildung also bijektiv. Das beweist (a). Wegen Lemma 3.6 folgt damit auch (b). □

3.8 Beispiel. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ die Drehung um $\binom{0}{0}$ um den Winkel ϑ. Das ist eine $\mathbb{R}$-lineare Abbildung. (Geometrisch kann man die Additivität so sehen: Jedes Parallelogramm in der Ebene mit $\binom{0}{0}$ als Eckpunkt wird durch f wieder in ein solches Parallelogramm überführt. Das bedeutet, daß f additiv ist.) Was ist die Matrix A mit $f = F_A$? Mit Elementargeometrie (rechtwinkliges Dreieck) sieht man $f(\binom{1}{0}) = \binom{\cos\vartheta}{\sin\vartheta}$, $f(\binom{0}{1}) = \binom{-\sin\vartheta}{\cos\vartheta}$, also ist $f = F_A$ mit der Matrix

$$A \;=\; \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}.$$

3.9 Definition. Sei V ein K-Vektorraum. Man nennt $f \in \mathrm{End}_K(V)$ einen *(linearen) Automorphismus* von V, wenn f bijektiv (also ein Isomorphismus) ist. Die Menge

$$\mathrm{GL}(V) \;:=\; \mathrm{GL}_K(V) \;:=\; \{f \in \mathrm{End}_K(V)\colon f \text{ bijektiv}\}$$

aller Automorphismen von V ist eine Gruppe unter Komposition (Bemerkung 3.2.1) und heißt die *allgemeine lineare Gruppe* von V.

3.10 Korollar. *Für $A \in \mathrm{M}_n(K)$ gilt $A \in \mathrm{GL}_n(K) \Leftrightarrow F_A \in \mathrm{GL}(K^n)$. Die Abbildung $\mathrm{GL}_n(K) \to \mathrm{GL}(K^n)$, $A \mapsto F_A$ ist ein Gruppenisomorphismus.*

In Worten: Eine quadratische Matrix A ist genau dann invertierbar, wenn die durch sie beschriebene lineare Abbildung F_A bijektiv ist.

BEWEIS. Aus Satz 3.7 folgt: $F_A \in \mathrm{GL}(K^n) \Leftrightarrow \exists\, g \in \mathrm{End}(K^n)$ mit $F_A \circ g = \mathrm{id} = g \circ F_A \Leftrightarrow \exists\, B \in \mathrm{M}_n(K)$ mit $AB = I_n = BA \Leftrightarrow A \in \mathrm{GL}_n(K)$. Damit folgt auch die zweite Aussage (erneut mit Satz 3.7). □

Der folgende Satz ist zentral und wird ständig verwendet:

3.11 Satz. *Sei $f\colon V \to W$ eine lineare Abbildung, sei $\dim(V) < \infty$. Dann ist*

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

BEWEIS. Sei $n := \dim(V)$ und $r := \dim\ker(f)$. Mit dem Basisergänzungssatz II.2.19 finden wir eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V derart, daß $(v_1, \ldots, v_r)$ eine Basis von $\ker(f)$ ist. Wegen $\mathrm{im}(f) = \mathrm{span}(f(v_1), \ldots, f(v_n))$ und $f(v_i) = 0$ für $i = 1, \ldots, r$ ist $\mathrm{im}(f) = \mathrm{span}(f(v_{r+1}), \ldots, f(v_n))$. Wir zeigen, daß die Vektoren $f(v_{r+1}), \ldots, f(v_n) \in W$ linear unabhängig sind, dann folgt $\dim\mathrm{im}(f) = n - r$, also die Behauptung. Sei also $\sum_{i=r+1}^{n} a_i f(v_i) = 0$ mit $a_i \in K$. Dann ist $\sum_{i=r+1}^{n} a_i v_i \in \ker(f) = \mathrm{span}(v_1, \ldots, v_r)$. Wegen $\mathcal{B}$ linear unabhängig folgt daraus $a_i = 0$ für $i = r+1, \ldots, n$. □

Jetzt kommen wichtige Folgerungen aus Satz 3.11:

3.12 Korollar. *Seien V, W Vektorräume mit $\dim(V) = \dim(W) < \infty$, sei $f\colon V \to W$ eine lineare Abbildung. Dann gilt:*

$$f \text{ injektiv} \;\Leftrightarrow\; f \text{ surjektiv} \;\Leftrightarrow\; f \text{ bijektiv.}$$

BEWEIS. Folgt sofort aus 3.11:

$$f \text{ injektiv} \;\Leftrightarrow\; \dim\ker(f) = 0 \;\Leftrightarrow\; \dim\mathrm{im}(f) = \dim(V) \;\Leftrightarrow\; f \text{ surjektiv}$$

(siehe II.2.24(b) für die letzte Äquivalenz). □

3.13 Korollar. *Sind V, W Vektorräume mit $\dim(V) = \dim(W) = n < \infty$, so ist $V \cong W$. Insbesondere ist $V \cong K^n$.*

BEWEIS. Sei $(v_1, \ldots, v_n)$ eine Basis von V und $(w_1, \ldots, w_n)$ eine Basis von W. Die lineare Abbildung $f\colon V \to W$ mit $f(v_i) = w_i$ $(i = 1, \ldots, n$, siehe 3.4) ist surjektiv, also nach 3.12 ein Isomorphismus. □

3.14 Korollar. *Sei $n \in \mathbb{N}$, seien $A, B \in \mathrm{M}_n(K)$ mit $AB = I_n$. Dann ist auch $BA = I_n$.*

Es sind also A und B invertierbar, und es gilt $B = A^{-1}$. (Die Aussage ist nicht trivial, siehe Aufgabe G.)

BEWEIS. Nach Voraussetzung ist $\mathrm{id}_{K^n} = F_{AB} = F_A \circ F_B$. Deshalb ist $F_B\colon K^n \to K^n$ injektiv, also nach Korollar 3.12 auch bijektiv. Nach Korollar 3.10 ist deshalb B invertierbar, also folgt $BA = B(AB)B^{-1} = BB^{-1} = I_n$. □

Mit Hilfe von linearen Abbildungen können wir jetzt den Zusammenhang zwischen interner und externer direkter Summe präzise formulieren (siehe Bemerkung II.3.12):

3.15 Satz. *Sei V ein Vektorraum, sei $(U_i)_{i\in I}$ eine Familie von Unterräumen von V, und sei $U := \sum_{i\in I} U_i$ ihre (interne) Summe (II.3.1), ein Unterraum von V. Sei $U' := \bigoplus_{i\in I} U_i$ die externe direkte Summe der U_i (II.3.11). Dann gilt:*

(a) *Die lineare Abbildung $f\colon U' \to U$, $(u_i)_{i\in I} \mapsto \sum_{i\in I} u_i$ ist surjektiv.*
(b) *Genau dann ist f ein Isomorphismus, wenn die interne Summe U der Unterräume U_i direkt ist (II.3.5).*

Beweis. Nach Definition ist

$$U' \;=\; \Big\{(u_i)_i \in \prod_{i\in I} U_i\colon\; u_i = 0 \text{ f. f. a. } i \in I\Big\}.$$

Es ist klar, daß f linear ist, und f ist surjektiv nach Lemma II.3.2. Genau dann ist f ein Isomorphismus, wenn $\ker(f) = \{0\}$ ist. Dazu äquivalent ist

$$\forall\, i \in I \quad U_i \cap \sum_{j\in I\smallsetminus\{i\}} U_j \;=\; \{0\}$$

Diese Bedingung ist nach Satz II.3.4 äquivalent zur Direktheit der internen Summe $\sum_i U_i$. □

Wegen Satz 3.15 ist die Unterscheidung zwischen interner und externer direkter Summe unwesentlich.

3.16 Korollar. (Dimensionsformel) *Sei V ein Vektorraum, seien U_1, U_2 endlichdimensionale Unterräume von V. Dann gilt*

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

Beweis. Sei $U_1 \oplus U_2$ die externe direkte Summe. Der Kern der surjektiven linearen Abbildung (3.15) $f\colon U_1 \oplus U_2 \to U_1 + U_2$, $f(u_1, u_2) = u_1 + u_2$ ist $\ker(f) = \big\{(u, -u)\colon u \in U_1 \cap U_2\big\}$, und die Abbildung $U_1 \cap U_2 \to \ker(f)$, $u \mapsto (u, -u)$ ist ein Isomorphismus. Also ist $\dim\ker(f) = \dim(U_1 \cap U_2)$. Wegen $\dim(U_1 \oplus U_2) = \dim(U_1) + \dim(U_2)$ (II.3.8) folgt nun die Behauptung durch Anwenden von Satz 3.11 auf f. □

Wir erklären jetzt, was das Lösen von linearen Gleichungssystemen mit linearen Abbildungen zu tun hat. Dafür zunächst etwas Terminologie.

3.17 Definition. Sei V ein K-Vektorraum. Eine Teilmenge $A \subseteq V$ heißt ein *affiner Unterraum* von V, wenn entweder $A = \varnothing$ ist, oder wenn ein Untervektorraum U von V und ein Vektor $v \in V$ existieren mit

$$A \;=\; \{v + u\colon u \in U\} \;=:\; v + U.$$

Ein nichtleerer affiner Unterraum ist also die Verschiebung eines Untervektorraums um einen festen Vektor.

3.18 Lemma. *Sei $A \subseteq V$ ein nichtleerer affiner Unterraum, etwa $A = v + U$ mit $v \in V$ und einem linearen Unterraum $U \subseteq V$. Dann ist*

$$U \;=\; \{x - y\colon x, y \in A\},$$

und $A = x + U$ für jedes $x \in A$. Insbesondere ist U durch A bestimmt. Man nennt U den Translationsraum *von A und schreibt $T(A) := U$.*

BEWEIS. Für $x \in A$, etwa $x = v + u$ mit $u \in U$, ist $x + U = \{(v+u)+u' \colon u' \in U\} = v + U = A$. Zeige die erste Behauptung. "$\subseteq$": Für $u \in U$ ist $u = (v+u) - v \in A - A$. "$\supseteq$": Für $x = v + u$, $y = v + w$ mit $u, w \in U$ ist $x - y = u - w \in U$. □

3.19 Definition. Die *Dimension* eines affinen Unterraums $A \subseteq V$ ist definiert durch $\dim(A) := \dim_K T(A)$ (Vektorraumdimension) falls $A \neq \varnothing$ ist, und $\dim(\varnothing) := -1$.

3.20 Beispiel. Für beliebige a, b, $c \in \mathbb{R}$ ist $A := \{(x,y) \in \mathbb{R}^2 \colon ax + by = c\}$ ein affiner Unterraum von $\mathbb{R}^2$ mit $T(A) = \{(x,y) \colon ax + by = 0\}$ falls $A \neq \varnothing$. Für $(a,b) \neq (0,0)$ ist $\dim(A) = 1$, für $a = b = c = 0$ ist $A = \mathbb{R}^2$, also $\dim(A) = 2$, für $(a,b) = (0,0)$ und $c \neq 0$ ist $A = \varnothing$, also $\dim(A) = -1$.

Affine Unterräume treten in der linearen Algebra auf als Urbildmengen eines Vektors unter einer linearen Abbildung.

3.21 Satz. *Sei $f \colon V \to W$ eine lineare Abbildung. Für jedes $w \in W$ ist die Urbildmenge $f^{-1}(\{w\})$ ein affiner Unterraum von V, mit Translationsraum* $\ker(f)$ *für $w \in f(V)$.*

BEWEIS. Gibt es $v \in V$ mit $f(v) = w$, so ist $f^{-1}(\{w\}) = v + \ker(f)$. Andernfalls ist $f^{-1}(\{w\}) = \varnothing$. □

3.22 Bemerkungen.

1. Sei $f \colon V \to W$ linear und $w \in W$. Die Bestimmung der Urbildmenge $f^{-1}(\{w\})$ zerfällt also in zwei Schritte:

   (1) Entscheide, ob $w \in \mathrm{im}(f)$ ist. Falls ja, finde ein $v \in V$ mit $f(v) = w$.
   (2) (falls $w \in \mathrm{im}(f)$) Bestimme den linearen Unterraum $\ker(f)$ von V.

Dann ist $f^{-1}(\{w\}) = v + \ker(f)$.

2. Betrachte das Beispiel $f = F_A \colon K^n \to K^m$, mit $A = (a_{ij}) \in \mathrm{M}_{m\times n}(K)$. Sei $w = (w_1, \ldots, w_m)^t \in K^m$, schreibe $\mathcal{L}(A,w) := F_A^{-1}(\{w\}) = \{x \in K^n \colon Ax = w\}$. Das ist die Lösungsmenge des *linearen Gleichungssystems (LGS)*

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & w_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & w_2 \\ \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & w_m \end{array} \qquad (*)$$

in den Unbekannten $x_1, \ldots, x_n$. Um $\mathcal{L}(A,w)$ zu bestimmen, müssen wir die beiden obigen Schritte ausführen:

   (1) Entscheide, ob $(*)$ lösbar ist, und finde gegebenenfalls eine spezielle Lösung $v = (v_1, \ldots, v_n)$ von $(*)$;
   (2) löse das zugehörige *homogene* Gleichungssystem (mit $w = 0$), d. h., bestimme den linearen Unterraum $\mathcal{L}(A,0)$ von K^n.

Dann ist $\mathcal{L}(A,w) = v + \mathcal{L}(A,0)$.

Das Matrixkalkül erlaubt es, beide Schritte auf systematische Weise durchzuführen und damit jedes lineare Gleichungssystem zu lösen. Wie man das macht, werden wir demnächst diskutieren.

4. Quotienten von Gruppen und Vektorräumen

4.1 Definition. Sei X eine Menge. Eine *Äquivalenzrelation auf X* ist eine Teilmenge $R \subseteq X \times X$, so daß für alle x, y, $z \in X$ gilt:

(1) $(x,x) \in R$ (*Reflexivität*);
(2) $(x,y) \in R \;\Rightarrow\; (y,x) \in R$ (*Symmetrie*);
(3) $(x,y) \in R \;\wedge (y,z) \in R \;\Rightarrow (x,z) \in R$ (*Transitivität*).

Statt $(x,y) \in R$ schreibt man meist $x \sim_R y$ oder $x \sim y$ und sagt, x ist (bezüglich R) *äquivalent* zu y. Mit dieser Notation schreiben sich die Bedingungen (1)–(3) als

(1) $x \sim x$,
(2) $x \sim y \;\Rightarrow\; y \sim x$,
(3) $x \sim y \;\wedge\; y \sim z \;\Rightarrow\; x \sim z$

für alle x, y, $z \in X$.

4.2 Beispiele.

1. Für festes $n \in \mathbb{N}$ ist $R = \{(a, a+kn) \colon a,\, k \in \mathbb{Z}\}$ eine Äquivalenzrelation auf $\mathbb{Z}$, und für a, $b \in \mathbb{Z}$ gilt $(a,b) \in R \Leftrightarrow a \equiv b \pmod{n}$ (siehe I.3.7).

2. Jede Abbildung $f \colon X \to Y$ zwischen Mengen definiert eine Äquivalenzrelation R_f auf X durch $R_f = \{(x_1, x_2) \in X \times X \colon f(x_1) = f(x_2)\}$.

4.3 Konstruktion. Sei $R \subseteq X \times X$ eine Äquivalenzrelation auf einer Menge X. Für x, $y \in X$ schreibe $x \sim y$ statt $(x,y) \in R$. Für $x \in X$ heißt die Teilmenge

$$[x] \;:=\; \{y \in X \colon\; x \sim y\}$$

von X die *Äquivalenzklasse* von x (bezüglich $\sim$). Behaupte, für je zwei Elemente x, $y \in X$ gilt:

$$[x] \;=\; [y] \quad \vee \quad [x] \cap [y] \;=\; \varnothing. \tag{$*$}$$

Denn ist $[x] \cap [y] \neq \varnothing$, etwa $z \in [x] \cap [y]$, und ist $w \in [x]$, so folgt $y \sim z \sim x \sim w$, also $w \in [y]$. Aus $[x] \cap [y] = \varnothing$ folgt also $[x] \subseteq [y]$; symmetrisch dazu folgt ebenso $[y] \subseteq [x]$, und daher $[x] = [y]$, womit die Behauptung bewiesen ist. Die Menge X ist somit die *disjunkte Vereinigung* der verschiedenen Äquivalenzklassen. Die Menge aller Äquivalenzklassen

$$X/\sim \;:=\; \{[x] \colon\; x \in X\}$$

("X modulo $\sim$") heißt die *Quotientenmenge* von X nach $\sim$, und ist eine Teilmenge der Potenzmenge $\mathcal{P}(X)$ von X. Die Abbildung

$$\pi \colon X \to X/\sim, \quad \pi(x) \;=\; [x] \quad (x \in X)$$

heißt die *Quotientenabbildung*. Die Abbildung π ist surjektiv, und nach Konstruktion gilt für alle x, $y \in X$:

$$x \sim y \quad \Leftrightarrow \quad \pi(x) = \pi(y).$$

4.4 Satz und Definition. *Sei G eine Gruppe und $H \subseteq G$ eine Untergruppe. Durch*

$$x \sim y \quad :\Leftrightarrow \quad x^{-1}y \in H \quad (x,\, y \in G)$$

*wird eine Äquivalenzrelation $\sim$ auf G definiert. Die Äquivalenzklasse von $x \in G$ ist $xH := \{xh \colon h \in H\}$ und heißt die H-*Linksnebenklasse *von x. Die Menge aller H-Linksnebenklassen wird mit $G/H := \{xH \colon x \in G\}$ bezeichnet.*

Es ist also G/H die Quotientenmenge von G nach der Äquivalenzrelation $\sim$.

BEWEIS. Wir müssen (1)–(3) aus 4.1 nachweisen. Für x, y, $z \in G$ gilt:

(1) Wegen $x^{-1}x = e \in H$ ist $x \sim x$;
(2) aus $x \sim y$, also $x^{-1}y \in H$, folgt $y^{-1}x = (x^{-1}y)^{-1} \in H$ (Lemma I.2.7), also $y \sim x$;
(3) aus $x \sim y$ und $y \sim z$, also $x^{-1}y \in H$ und $y^{-1}z \in H$, folgt $x^{-1}z = (x^{-1}y)(y^{-1}z) \in H$, also $x \sim z$.

Für x, $y \in G$ gilt: $y \in [x] \Leftrightarrow x \sim y \Leftrightarrow \exists\, h \in H\ x^{-1}y = h \Leftrightarrow \exists\, h \in H\ y = xh \Leftrightarrow y \in xH$. Also ist $[x] = xH$. □

4.5 Korollar. (Satz von Lagrange[1]) *Sei G eine endliche Gruppe, sei $H \subseteq G$ eine Untergruppe. Dann ist $|G| = |H| \cdot |G/H|$. Insbesondere ist $|H|$ ein Teiler von $|G|$.*

BEWEIS. Für jedes $x \in G$ ist die Abbildung $H \to xH$, $h \mapsto xh$ bijektiv (Kürzen I.2.10), also gilt $|xH| = |H|$. Da G die disjunkte Vereinigung der $|G/H|$ verschiedenen Nebenklassen xH ist, folgt die Behauptung. □

4.6 Bemerkung. Das in 4.4 und 4.5 Gesagte kann man ebenso mit der Äquivalenzrelation $x \sim y :\Leftrightarrow xy^{-1} \in H$ auf G und mit den *Rechtsnebenklassen*

$$Hx := \{hx \colon h \in H\} \quad (x \in G)$$

durchführen, sowie mit der Menge

$$H\backslash G := \{Hx \colon x \in G\}$$

aller solchen. Für $|G| < \infty$ folgt analog $|G| = |H| \cdot |H\backslash G|$.

4.7 Konstruktion. Sei G eine Gruppe und $N \trianglelefteq G$ ein Normalteiler von G (2.7). Für alle $x \in G$ gilt dann $xN = Nx$, denn für $n \in N$ ist $xn = xnx^{-1} \cdot x \in Nx$, und ebenso $nx = x \cdot x^{-1}nx \in xN$. Es ist also $G/N = N\backslash G$. Auf dieser Menge definieren wir eine Verknüpfung $\cdot$ durch

$$xN \cdot yN := (xy)N \quad (x, y \in G). \tag{$*$}$$

Diese Multiplikation von Nebenklassen ist also vertreterweise definiert, und wir müssen zeigen, daß die Definition nicht von der Auswahl der Vertreter abhängt. Für x, $y \in G$ und $x' = xn_1$, $y' = yn_2$ mit n_1, $n_2 \in N$ ist

$$x'y' = xn_1 \cdot yn_2 = xy \cdot y^{-1}n_1y \cdot n_2 \in (xy)N,$$

also ist $(x'y')N = (xy)N$. Die Regel $(*)$ ist also wohldefiniert und definiert eine Verknüpfung $(G/N) \times (G/N) \to G/N$.

4.8 Satz. *Sei G eine Gruppe, sei $N \subseteq G$ ein Normalteiler. Dann wird G/N durch 4.7 zu einer Gruppe, genannt die* Quotientengruppe G modulo N. *Die Quotientenabbildung*

$$\pi \colon G \to G/N, \quad \pi(x) = xN \quad (x \in G)$$

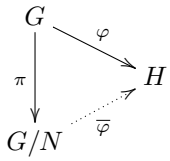
ist ein surjektiver Gruppenhomomorphismus, und $\ker(\pi) = N$.

[1] Joseph-Louis LAGRANGE (1736–1813)

BEWEIS. Die Gruppenaxiome für $(G/N, \cdot)$ folgen sofort aus den Gruppenaxiomen für G: Das neutrale Element ist $eN = N$, das zu xN inverse Element ist $(xN)^{-1} = x^{-1}N$. Die Homomorphie von π ist gerade $(*)$, gilt also nach Definition der Multiplikation in G/N, und $\ker(\pi) = N$ wegen $x \in \ker(\pi) \Leftrightarrow \pi(x) = eN = N$ $\Leftrightarrow x \in N$. □

4.9 Satz. (Homomorphiesatz für Gruppen) *Sei $\varphi\colon G \to H$ ein Gruppenhomomorphismus, sei $N \trianglelefteq G$ mit $N \subseteq \ker(\varphi)$, und sei $\pi\colon G \to G/N$ die Quotientenabbildung. Dann gibt es einen eindeutig bestimmten Homomorphismus $\overline{\varphi}\colon G/N \to H$ mit $\overline{\varphi} \circ \pi = \varphi$. Für diesen gilt* $\ker(\overline{\varphi}) = \ker(\varphi)/N$.

Die Existenzaussage kann man auch mit der folgenden suggestiven Sprechweise formulieren: Es gibt genau einen Homomorphismus $\overline{\varphi}\colon G/N \to H$, für den das Dreieck



kommutiert.

BEWEIS. Aus der geforderten Eigenschaft folgt: Wir *müssen* $\overline{\varphi}$ definieren durch $\overline{\varphi}(xN) := \varphi(x)$ $(x \in G)$, und müssen zunächst die Wohldefiniertheit zeigen. Aus $xN = yN$ folgt $y = xn$ mit $n \in N$, also ist $\varphi(y) = \varphi(x)\varphi(n) = \varphi(x)$ wegen $\varphi(n) = e$. Also ist $\overline{\varphi}$ wohldefiniert. Weiter ist $\overline{\varphi}$ ein Homomorphismus:

$$\overline{\varphi}(xN \cdot yN) = \overline{\varphi}(xyN) = \varphi(xy) = \varphi(x)\varphi(y) = \overline{\varphi}(xN) \cdot \overline{\varphi}(yN).$$

Es gilt $\overline{\varphi}(xN) = e \Leftrightarrow \varphi(x) = e \Leftrightarrow x \in N$, also $\ker(\overline{\varphi}) = \{xN\colon x \in \ker(\varphi)\} = \ker(\varphi)/N$. □

Besonders wichtig ist der folgende Spezialfall:

4.10 Korollar. *Ist $\varphi\colon G \to H$ ein surjektiver Gruppenhomomorphismus, so ist $H \cong G/\ker(\varphi)$.*

BEWEIS. $\ker(\varphi)$ ist ein Normalteiler von G (2.8.1). Wende 4.9 an mit $N = \ker(\varphi)$, sei $\overline{\varphi}\colon G/\ker(\varphi) \to H$ der dort konstruierte Homomorphismus. Nach 4.9 ist $\ker(\overline{\varphi})$ trivial, also ist $\overline{\varphi}$ injektiv. Wegen φ surjektiv ist $\overline{\varphi}$ auch surjektiv, also ein Isomorphismus. □

4.11 Beispiele.

1. Für jeden Körper K ist $\mathrm{GL}_2(K)/\mathrm{SL}_2(K) \cong K^*$. Denn $\mathrm{SL}_2(K)$ ist der Kern des surjektiven Homomorphismus $\det\colon \mathrm{GL}_2(K) \to K^*$ (siehe 2.8.2), die Behauptung folgt also aus 4.10. Weitere Beispiele gibt es in den Übungen (Aufgabe 27).

2. Stets gilt $G/\{e\} \cong G$ und $G/G \cong \{e\}$.

3. Ist $(G, +)$ eine additiv geschriebene abelsche Gruppe und $H \subseteq G$ eine Untergruppe (also automatisch $H \trianglelefteq G$), so schreibt man die Nebenklassen additiv als $x + H = \{x + h\colon h \in H\}$ $(x \in G)$, und schreibt auch die Quotientengruppe G/H additiv, also $(x + H) + (y + H) = (x + y) + H$ $(x, y \in G)$.

4. Ist $H \subseteq G$ eine Untergruppe, die kein Normalteiler ist, so haben wir auf der Menge G/H keine Gruppenstruktur definiert. Man kann zeigen, daß das tatsächlich nicht in "vernünftiger" Weise geht und muß dabei nur überlegen, was genau mit "vernünftig" gemeint sein soll.

4.12 Konstruktion. Jetzt konstruieren wir Quotienten von Vektorräumen. Sei V ein K-Vektorraum, sei U ein Untervektorraum von V, und sei $(V/U, +)$ die Quotientengruppe der additiven Gruppen. Definiere die Abbildung

$$K \times (V/U) \to V/U, \quad (a, v+U) \mapsto a(v+U) := av+U \quad (a \in K,\ v \in V).$$

Wegen $Ku \subseteq U$ für alle $u \in U$ ist dies wohldefiniert, und wir erhalten:

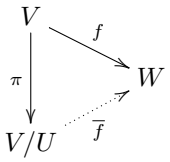
4.13 Satz. *Durch 4.12 wird V/U ein K-Vektorraum, genannt der* Quotientenraum V modulo U. *Die Quotientenabbildung*

$$\pi\colon V \to V/U, \quad \pi(v) = v+U \quad (v \in V)$$

ist K-linear und surjektiv und erfüllt $\ker(\pi) = U$. □

Wie bei Gruppen besteht auch für Vektorräume ein Homomorphiesatz:

4.14 Satz. (Homomorphiesatz) *Sei $f\colon V \to W$ eine lineare Abbildung zwischen K-Vektorräumen, sei $U \subseteq V$ ein Unterraum mit $U \subseteq \ker(f)$. Dann gibt es eine eindeutig bestimmte lineare Abbildung $\overline{f}\colon V/U \to W$ mit $\overline{f} \circ \pi = f$*



und dabei ist $\ker(\overline{f}) = \ker(f)/U$.

Beweis. $\overline{f}$ ist die Abbildung $\overline{f}(v+U) = f(v)$ $(v \in V)$ aus dem Homomorphiesatz für die additiven Gruppen 4.9. Zu zeigen ist nur noch, daß $\overline{f}$ homogen ist, also $\overline{f}(av+U) = a\overline{f}(v+U)$ $(a \in K,\ v \in V)$. Das ist aber klar wegen f homogen. □

4.15 Korollar. *Sei $f\colon V \to W$ eine lineare Abbildung. Dann ist*

$$V/\ker(f) \cong \operatorname{im}(f).$$

Beweis. Die lineare Abbildung $g\colon V \to \operatorname{im}(f)$, $g(v) := f(v)$ $(v \in V)$ ist surjektiv, und $\ker(g) = \ker(f)$. Die in 4.14 konstruierte lineare Abbildung $\overline{g}\colon V/\ker(f) \to \operatorname{im}(f)$ mit $\overline{g}(v + \ker(f)) = g(v) = f(v)$ ist ein Isomorphismus. □

4.16 Definition. Sei V ein K-Vektorraum und $U \subseteq V$ ein Untervektorraum. Dann heißt $\operatorname{codim}_V(U) := \dim(V/U)$ die *Kodimension* von U in V.

Die Kodimension ist also, im Gegensatz zur Dimension, eine relative Größe, nämlich relativ zu einem Obervektorraum. Die Bezeichnung Kodimension wird durch das nächste Korollar erklärt:

4.17 Korollar. *Ist* $\dim(V) < \infty$, *so gilt* $\operatorname{codim}_V(U) = \dim(V) - \dim(U)$.

BEWEIS. Wende die Dimensionsformel 3.11 auf die lineare Quotientenabbildung $\pi\colon V \to V/U$, $\pi(v) = v + U$ an: Diese ist surjektiv und hat $\ker(\pi) = U$, also ist $\dim(V) = \dim(U) + \dim(V/U)$. □

4.18 Beispiele.

1. Die Elemente des Quotientenvektorraums V/U sind genau die affinen Unterräume $A \neq \varnothing$ von V mit Translationsraum $T(A) = U$ (3.18).

2. $V/\{0\} \cong V$, $V/V \cong \{0\}$.

3. Ist $\operatorname{codim}_V(U) = 1$, so heißt U eine *(lineare) Hyperebene* in V. Beispiel: Sei $(0,\dots,0) \neq (a_1,\dots,a_n) \in K^n$, dann ist die Abbildung

$$f\colon K^n \to K, \quad f(x) = \sum_{i=1}^{n} a_i x_i \quad (x \in K^n)$$

linear und surjektiv. Für den Unterraum $U := \{x \in K^n\colon a_1x_1 + \cdots + a_nx_n = 0\}$ von K^n gilt $U = \ker(f)$, also $K^n/U \cong K$ nach dem Homomorphiesatz 4.15. Also ist U eine Hyperebene in K^n.

Als Anwendung des Homomorphiesatzes beweisen wir zwei Isomorphiesätze für Quotientenvektorräume:

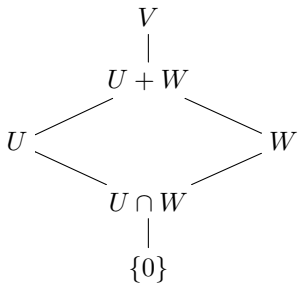
4.19 Satz. *Sei V ein Vektorraum, und seien U, W zwei lineare Unterräume von V.*

(a) *$U/U \cap W \;\cong\; (U+W)/W$ (1. Isomorphiesatz).*

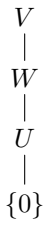
(b) *Ist $U \subseteq W$, so gilt $(V/U)\big/(W/U) \;\cong\; V/W$ (2. Isomorphiesatz).*

Beachte, daß W/U in (b) ein linearer Unterraum von V/U ist. Die Inklusionsverhältnisse zwischen den beteiligten Unterräumen werden durch die folgenden *Hasse-Diagramme*[2] veranschaulicht:

(a)



(b)



BEWEIS. In beiden Fällen zeigen wir tatsächlich, daß es nicht nur irgendwelche Isomorphismen gibt, sondern daß diese ganz kanonisch sind.

(a) Die lineare Abbildung

$$f\colon U \to (U+W)/W, \quad f(u) \;=\; u + W \quad (u \in U)$$

ist surjektiv, denn jedes Element in $(U+W)/W$ hat die Form $u+w+W = u+W = f(u)$ mit $u \in U$, $w \in W$. Für $u \in U$ gilt $f(u) = 0 \Leftrightarrow u + W = W \Leftrightarrow u \in W$. Also ist $\ker(f) = U \cap W$. Die nach dem Homomorphiesatz induzierte lineare Abbildung $\overline{f}\colon U/U \cap W \to (U+W)/W$ ist also ein Isomorphismus (Korollar 4.15).

[2] Helmut HASSE (1898–1979)

(b) Die lineare Abbildung

$$f\colon V/U \to V/W, \quad f(v+U) := v+W \quad (v \in V)$$

ist wohldefiniert wegen $U \subseteq W$ (siehe auch 4.13, angewandt auf $V \to V/W$ und den Unterraum U von V), und ist surjektiv mit $\ker(f) = \{v+U \in V/U\colon v+W = W\} = W/U$. Die induzierte lineare Abbildung $\overline{f}\colon (V/U)/(W/U) \to V/W$ ist also ein Isomorphismus, wieder nach 4.15. □

Wir können lineare Komplemente von Untervektorräumen auch wie folgt charakterisieren:

4.20 Korollar. *Seien U, W Untervektorräume von V. Genau dann ist W ein lineares Komplement von U in V (also $U \oplus W = V$, siehe II.3.9), wenn die lineare Abbildung*

$$f\colon W \to V/U, \quad f(w) = w+U \quad (w \in W)$$

ein Isomorphismus ist. Insbesondere ist jedes lineare Komplement von U in V zu V/U isomorph.

Beweis. Es ist $\mathrm{im}(f) = \{w+U\colon w \in W\} = (U+W)/U$, also ist die Surjektivität von f äquivalent zu $U+W = V$. Es ist $\ker(f) = U \cap W$, also ist die Injektivität von f äquivalent zu $U \cap W = \{0\}$. Genau dann ist also f bijektiv, wenn $V = U \oplus W$ ist. □

5. Koordinaten

K sei stets ein Körper. Alle Vektorräume in diesem Abschnitt sind K-Vektorräume von *endlicher Dimension* ≥ 1. Im folgenden sei V stets ein K-Vektorraum und $\mathcal{B} = (v_1, \ldots, v_n)$ eine feste Basis von V (es ist also $n = \dim(V) \in \mathbb{N}$).

5.1 Definition. Die lineare Abbildung

$$\Phi_{\mathcal{B}}\colon K^n \to V, \quad (x_1, \ldots, x_n)^t \mapsto \sum_{i=1}^{n} x_i v_i$$

ist ein Isomorphismus und heißt das zu $\mathcal{B}$ gehörende *Koordinatensystem* von V. Für $v \in V$ heißt $\Phi_{\mathcal{B}}^{-1}(v) \in K^n$ der *Koordinatenvektor von v* bezüglich der Basis $\mathcal{B}$.

Der Koordinatenvektor $(x_1, \ldots, x_n)^t \in K^n$ von v bezüglich $\mathcal{B}$ ist also charakterisiert durch die Gleichung $v = \sum_{i=1}^{n} x_i v_i$.

5.2 Definition. Sei W ein weiterer Vektorraum mit Basis $\mathcal{C} = (w_1, \ldots, w_m)$, und sei $f\colon V \to W$ eine lineare Abbildung. Wir drücken die Bildvektoren $f(v_j) \in W$ durch die Basis $\mathcal{C}$ aus, d.h. schreiben

$$f(v_j) = \sum_{i=1}^{m} a_{ij} w_i \qquad (j = 1, \ldots, n)$$

mit $a_{ij} \in K$. Dann heißt die Matrix

$$\mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f) := (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \in \mathrm{M}_{m \times n}(K)$$

die *f beschreibende Matrix* (bezüglich der gewählten Basen $\mathcal{B}$ und $\mathcal{C}$).

5.3 Satz. *In der Situation von 5.2 sei* $A := \mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f)$. *Dann ist* A *die eindeutig bestimmte Matrix, für die*

$$f \circ \Phi_{\mathcal{B}} \;=\; \Phi_{\mathcal{C}} \circ F_A$$

(als lineare Abbildungen $K^n \to W$*) gilt, also für die das Quadrat*

$$\begin{array}{ccc} K^n & \xrightarrow{F_A} & K^m \\ {\scriptstyle \Phi_{\mathcal{B}}}\downarrow & & \downarrow{\scriptstyle \Phi_{\mathcal{C}}} \\ V & \xrightarrow{f} & W \end{array}$$

kommutiert. Die Abbildung $\mathrm{M}_{\mathcal{C}}^{\mathcal{B}} \colon \mathrm{Hom}(V,W) \to \mathrm{M}_{m\times n}(K)$, $f \mapsto \mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f)$ *ist ein Vektorraum-Isomorphismus. Insbesondere ist* $\dim \mathrm{Hom}(V,W) = \dim(V)\cdot\dim(W)$.

Beweis. Für $j = 1,\dots,n$ ist $(f\circ\Phi_{\mathcal{B}})(e_j) = f(v_j)$ und

$$(\Phi_{\mathcal{C}} \circ F_A)(e_j) \;=\; \Phi_{\mathcal{C}}\Big(\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}\Big) \;=\; \sum_{i=1}^{m} a_{ij} w_i \;=\; f(v_j).$$

Also kommutiert das Quadrat wie behauptet. Es ergibt sich $F_A = \Phi_{\mathcal{C}}^{-1} \circ f \circ \Phi_B$, also ist F_A (und damit auch A, siehe 3.7(a)) durch f eindeutig bestimmt. Außerdem ist die Abbildung $f \mapsto F_A$ linear und bijektiv. Also ist auch die Abbildung $\mathrm{M}_{\mathcal{C}}^{\mathcal{B}}$ linear und bijektiv (siehe erneut 3.7(a)). □

5.4 Bemerkungen.

1. Anders gesagt: Ist $x = (x_1,\dots,x_n)^t$ der Koordinatenvektor von $v \in V$ beüglich $\mathcal{B}$ und $y = (y_1,\dots,y_m)^t$ der Koordinatenvektor von $f(v) \in W$ bezüglich $\mathcal{C}$, so ist

$$y \;=\; \mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f)\, x.$$

Merkregel: In der j-ten Spalte von $\mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f)$ stehen die $\mathcal{C}$-Koordinaten des Bildes des j-ten Vektors aus $\mathcal{B}$ unter f.

2. Für $n \in \mathbb{N}$ sei $\mathcal{K}_n = (e_1,\dots,e_n)$ die kanonische Basis von K^n. Das Koordinatensystem $\Phi_{\mathcal{K}_n} \colon K^n \to K^n$ ist die Identität. Für jede $m\times n$-Matrix $A \in \mathrm{M}_{m\times n}(K)$ ist

$$A \;=\; \mathrm{M}_{\mathcal{K}_m}^{\mathcal{K}_n}(F_A).$$

3. (Beispiel) Betrachte den $\mathbb{R}$-Vektorraum $\mathbb{R}[x]_d = \{f \in \mathbb{R}[x] \colon \deg(f) \le d\}$ und die Basis $\mathcal{B} = (1, x, \dots, x^d)$ von $\mathbb{R}[x]_d$. Sei $D \colon \mathbb{R}[x]_d \to \mathbb{R}[x]_d$ die Ableitung, also

$$D\Big(\sum_{i=0}^{d} a_i x^i\Big) \;=\; \sum_{i=1}^{d} i a_i\, x^{i-1}.$$

Die Abbildung D ist linear, und die darstellende Matrix bezüglich der Basis $\mathcal{B}$ ist

$$M_{\mathcal{B}}^{\mathcal{B}}(D) \;=\; \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ & 0 & 2 & \cdots & 0 \\ & & \ddots & \ddots & \vdots \\ & & & 0 & d \\ & & & & 0 \end{pmatrix} \;\in\; \mathrm{M}_{d+1}(\mathbb{R}).$$

Der Koordinatenvektor von $f = \sum_{i=0}^{d} a_i x^i$ bezüglich $\mathcal{B}$ ist $\Phi_{\mathcal{B}}^{-1}(f) = (a_0,\dots,a_d)^t$.

Wir diskutieren nun zunächst, wie sich die beschreibende Matrix einer linearen Abbildung mit Komposition von linearen Abbildungen verträgt.

5.5 Satz. *In der Situation von 5.2 sei $g\colon U \to V$ eine weitere lineare Abbildung und $\mathcal{A}$ eine Basis von U. Dann gilt*

$$\mathrm{M}_{\mathcal{C}}^{\mathcal{A}}(f \circ g) \;=\; \mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f) \cdot \mathrm{M}_{\mathcal{B}}^{\mathcal{A}}(g).$$

Beweis. Sei $r = \dim(U)$, $A = \mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f)$ und $B = \mathrm{M}_{\mathcal{B}}^{\mathcal{A}}(g)$. Betrachte das Diagramm

$$\begin{array}{ccccc} K^r & \xrightarrow{F_B} & K^n & \xrightarrow{F_A} & K^m \\ {\scriptstyle \Phi_{\mathcal{A}}}\downarrow & & {\scriptstyle \Phi_{\mathcal{B}}}\downarrow & & {\scriptstyle \Phi_{\mathcal{C}}}\downarrow \\ U & \xrightarrow{g} & V & \xrightarrow{f} & W \end{array}$$

Das linke und das rechte Quadrat kommutiert jeweils nach 5.3. Also kommutiert auch das äußere Quadrat, d.h. es gilt $\Phi_{\mathcal{C}} \circ F_A \circ F_B = f \circ g \circ \Phi_{\mathcal{A}}$. Nach 5.3 ist $\mathrm{M}_{\mathcal{C}}^{\mathcal{A}}(f \circ g)$ die eindeutig bestimmte Matrix $C \in \mathrm{M}_{m\times r}(K)$ mit $\Phi_{\mathcal{C}} \circ F_C = f \circ g \circ \Phi_{\mathcal{A}}$. Wegen $F_A \circ F_B = F_{AB}$ ist $C = AB$ eine solche Matrix. □

5.6 Korollar. *(Bezeichnungen wie zuvor) Die lineare Abbildung $f\colon V \to W$ ist genau dann ein Isomorphismus, wenn die Matrix $\mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f)$ invertierbar ist. In diesem Fall gilt für die Umkehrabbildung f^{-1} von f:*

$$\mathrm{M}_{\mathcal{B}}^{\mathcal{C}}(f^{-1}) \;=\; \mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f)^{-1}.$$

Beweis. Das folgt aus Satz 5.5. □

5.7 Korollar. *Sei $\dim(V) = n$, sei $\mathcal{B}$ eine Basis von V. Dann ist die Abbildung $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}\colon \mathrm{End}(V) \to \mathrm{M}_n(K)$ ein Ringisomorphismus.*

Beweis. Die Abbildung ist ein Vektorraum-Isomorphismus gemäß 5.3, und ist ein Ringhomomorphismus nach 5.5. □

5.8 Wenn wir die Basis wechseln, erhalten wir auch andere Koordinaten. Die Aufgabe, die Koordinaten bezüglich verschiedener Basen ineinander umzurechnen, tritt in der Praxis ständig auf. Betrachte also eine zweite Basis $\mathcal{B}' = (v'_1, \ldots, v'_n)$ von V und das zugehörige Koordinatensystem $\Phi_{\mathcal{B}'}\colon K^n \to V$.

5.9 Lemma und Definition. *Die Matrix $\mathrm{T}_{\mathcal{B}'}^{\mathcal{B}} := \mathrm{M}_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{id}_V)$ ist invertierbar und heißt die* Basiswechselmatrix *von $\mathcal{B}$ nach $\mathcal{B}'$. Sie ist charakterisiert durch $\mathrm{T}_{\mathcal{B}'}^{\mathcal{B}} = (t_{ij})_{1\le i,j\le n}$ mit $v_j = \sum_{i=1}^n t_{ij} v'_i$ $(j = 1, \ldots, n)$. Für jedes $v \in V$ mit Koordinatenvektoren $x = \Phi_{\mathcal{B}}^{-1}(v)$ bzgl. $\mathcal{B}$ und $x' = \Phi_{\mathcal{B}'}^{-1}(v)$ bzgl. $\mathcal{B}'$ ist $x' = \mathrm{T}_{\mathcal{B}'}^{\mathcal{B}} \cdot x$.*

In den Spalten von $T_{\mathcal{B}'}^{\mathcal{B}}$ stehen also die Elemente von $\mathcal{B}$, ausgedrückt durch $\mathcal{B}'$ (genauer gesagt, die zugehörigen Koordinatenvektoren).

Beweis. $T := T_{\mathcal{B}'}^{\mathcal{B}}$ ist invertierbar nach 5.6, angewandt auf id_V, und die Charakterisierung von T ist die Definition von T (5.2). Aus $\Phi_{\mathcal{B}} = \Phi_{\mathcal{B}'} \circ F_T$ folgt $v = \Phi_{\mathcal{B}}(x) = \Phi_{\mathcal{B}'}(Tx)$, also $Tx = \Phi_{\mathcal{B}'}^{-1}(v) = x'$. □

Nach 5.5 und 5.6 folgt also:

5.10 Korollar. *Für je drei Basen $\mathcal{B}_1$, $\mathcal{B}_2$, $\mathcal{B}_3$ von V gilt*

$$\mathrm{T}_{\mathcal{B}_3}^{\mathcal{B}_1} \;=\; \mathrm{T}_{\mathcal{B}_3}^{\mathcal{B}_2} \cdot T_{\mathcal{B}_2}^{\mathcal{B}_1}$$

und

$$\mathrm{T}_{\mathcal{B}_1}^{\mathcal{B}_2} \;=\; \big(\mathrm{T}_{\mathcal{B}_2}^{\mathcal{B}_1}\big)^{-1}. \qquad \square$$

Es ist jetzt auch leicht zu sagen, wie sich die beschreibende Matrix einer linearen Abbildung ändert, wenn man die Basen der beteiligten Vektorräume ändert:

5.11 Korollar. *Sei $f\colon V \to W$ eine lineare Abbildung, und seien $\mathcal{B}$, $\mathcal{B}'$ Basen von V und $\mathcal{C}$, $\mathcal{C}'$ Basen von W. Dann ist*

$$\mathrm{M}_{\mathcal{C}'}^{\mathcal{B}'}(f) \;=\; \mathrm{T}_{\mathcal{C}'}^{\mathcal{C}} \cdot \mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f) \cdot \big(\mathrm{T}_{\mathcal{B}'}^{\mathcal{B}}\big)^{-1}.$$

Beweis. Das folgt sofort aus 5.5 und 5.6, angewandt auf die Komposition

$$V \xrightarrow{\mathrm{id}} V \xrightarrow{f} W \xrightarrow{\mathrm{id}} W$$

und die Basen $\mathcal{B}'$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{C}'$ der vier Vektorräume. $\square$

5.12 Beispiel. Sei $f \in \mathrm{End}(\mathbb{R}^2)$ die Drehung um $(0,0)$ um den Winkel ϑ. Nach 3.8 ist $f = F_A$ mit

$$A \;=\; \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}$$

Sei $\mathcal{K} = (e_1, e_2)$ die kanonische Basis von $\mathbb{R}^2$, es ist also $\mathrm{M}_{\mathcal{K}}^{\mathcal{K}}(f) = A$. Angenommen, wir sollen f bezüglich der Basis $\mathcal{B} = (v_1, v_2)$ von $\mathbb{R}^2$ beschreiben mit $v_1 = (0,1)^t$, $v_2 = (-1,2)^t$. Es ist

$$T_{\mathcal{K}}^{\mathcal{B}} \;=\; \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix},$$

also folgt nach 5.10

$$T_{\mathcal{B}}^{\mathcal{K}} \;=\; \big(T_{\mathcal{K}}^{\mathcal{B}}\big)^{-1} \;=\; \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$$

und der Basiswechsel ergibt nach 5.11

$$\begin{aligned} \mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) &= T_{\mathcal{B}}^{\mathcal{K}} \cdot \mathrm{M}_{\mathcal{K}}^{\mathcal{K}}(f) \cdot T_{\mathcal{K}}^{\mathcal{B}} \\ &= \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} \cos\vartheta - 2\sin\vartheta & -5\sin\vartheta \\ \sin\vartheta & \cos\vartheta + 2\sin\vartheta \end{pmatrix} \end{aligned}$$

6. Rang

Sei K ein Körper. Alle Vektorräume seien endlich-dimensionale K-Vektorräume, alle linearen Abbildungen seien K-linear.

6.1 Definition. Sei $f\colon V \to W$ eine lineare Abbildung. Der *Rang* von f ist definiert als $\mathrm{rk}(f) := \dim \mathrm{im}(f) = \dim f(V)$.

6.2 Lemma. *Sei $f\colon V \to W$ linear. Dann gilt*

$$\mathrm{rk}(f) \;\le\; \min\{\dim(V), \dim(W)\},$$

und $\mathrm{rk}(f) = \dim(V) \Leftrightarrow f$ *ist injektiv,* $\mathrm{rk}(f) = \dim(W) \Leftrightarrow f$ *ist surjektiv.*

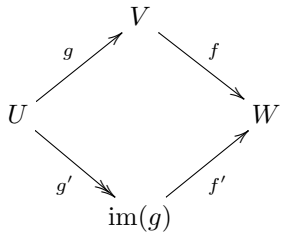
BEWEIS. Der Vergleich mit $\dim(V)$ folgt aus $\operatorname{rk}(f) = \dim(V) - \dim\ker(f)$ (Satz 3.11). Der Vergleich mit $\dim(W)$ ist klar. □

Daraus folgt:

6.3 Satz. *Seien $g\colon U \to V$ und $f\colon V \to W$ lineare Abbildungen.*

(a) $\operatorname{rk}(f \circ g) \leq \min\{\operatorname{rk}(f), \operatorname{rk}(g)\}$,
(b) *f injektiv* $\Rightarrow$ $\operatorname{rk}(f \circ g) = \operatorname{rk}(g)$,
(c) *g surjektiv* $\Rightarrow$ $\operatorname{rk}(f \circ g) = \operatorname{rk}(f)$.

BEWEIS. Sei $g'\colon U \to \operatorname{im}(g)$, $g'(u) := g(u)$ $(u \in U)$ die durch g induzierte lineare Abbildung, sei $f' := f|_{\operatorname{im}(g)}\colon \operatorname{im}(g) \to W$ die Restriktion von f. Dann ist $f \circ g = f' \circ g'$, und g' ist surjektiv:



Also ist $\operatorname{im}(f \circ g) = \operatorname{im}(f' \circ g') = \operatorname{im}(f')$, und somit nach 6.2

$$\operatorname{rk}(f \circ g) \;=\; \operatorname{rk}(f') \;\leq\; \min\{\dim \operatorname{im}(g), \dim \operatorname{im}(f)\} \;=\; \min\{\operatorname{rk}(g), \operatorname{rk}(f)\}.$$

Ist f injektiv, so auch f', also $\operatorname{rk}(f \circ g) = \dim \operatorname{im}(g) = \operatorname{rk}(g)$ nach 6.2. Ist g surjektiv, so ist $f' = f$. □

6.4 Satz. (Rangsatz) *Sei $f\colon V \to W$ eine lineare Abbildung.*

(a) *Es gibt Basen $\mathcal{B}$ von V und $\mathcal{C}$ von W sowie $r \in \mathbb{N}_0$ mit*

$$\mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f) \;=\; \sum_{i=1}^{r} E_{ii} \;=\; \left(\begin{array}{ccc|c} 1 & & 0 & \\ & \ddots & & 0 \\ 0 & & 1 & \\ \hline & 0 & & 0 \end{array}\right) \;=\; \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right). \qquad (*)$$

(b) *Für $\mathcal{B}$, $\mathcal{C}$, r wie in (a) ist $r = \operatorname{rk}(f)$.*

BEWEIS. Gibt es solche Basen $\mathcal{B} = (v_1, \ldots, v_n)$ und $\mathcal{C} = (w_1, \ldots, w_m)$, so ist

$$f(v_j) \;=\; \begin{cases} w_j & j = 1, \ldots, r, \\ 0 & j = r+1, \ldots, n, \end{cases}$$

also $\operatorname{im}(f) = \operatorname{span}(w_1, \ldots, w_r)$, und somit $\operatorname{rk}(f) = r$. Umgekehrt sei $r = \operatorname{rk}(f)$. Ergänze eine Basis $(w_1, \ldots, w_r)$ von $\operatorname{im}(f)$ zu einer Basis $\mathcal{C} = (w_1, \ldots, w_m)$ von W (II.2.19). Für $j = 1, \ldots, r$ wähle $v_j \in V$ mit $f(v_j) = w_j$, dann ist $(v_1, \ldots, v_r)$ linear unabhängig. Setze $U := \operatorname{span}(v_1, \ldots, v_r)$. Behaupte, es ist $V = U \oplus \ker(f)$. In der

Tat, für $v \in V$ gibt es $a_1, \dots, a_r \in K$ mit $f(v) = \sum_{i=1}^r a_i w_i$. Für $u := \sum_{i=1}^r a_i v_i$ ist also $u \in U$ und $f(v) = f(u)$, also $v - u \in \ker(f)$, und somit

$$v \;=\; u + (v-u) \in U + \ker(f).$$

Die Summe ist auch direkt: Ist $u = \sum_{j=1}^r a_j v_j \in U \cap \ker(f)$ (mit $a_j \in K$), so ist $0 = f(u) = \sum_{i=1}^r a_i w_i$, also $a_j = 0$ $(j = 1, \dots, r)$ wegen $(w_1, \dots, w_r)$ linear unabhängig, also $u = 0$.

Sei $(v_{r+1}, \dots, v_n)$ eine Basis von $\ker(f)$. Dann ist $\mathcal{B} := (v_1, \dots, v_r, v_{r+1}, \dots, v_n)$ eine Basis von V, und $\mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f)$ hat genau die Form $(*)$. □

Hier ist eine Folgerung für Matrizen:

6.5 Satz. *Zu jeder Matrix $A \in \mathrm{M}_{m\times n}(K)$ gibt es invertierbare Matrizen $T \in \mathrm{GL}_m(K)$ und $S \in \mathrm{GL}_n(K)$, so daß $T^{-1}AS$ die Gestalt $(*)$ mit $r = \mathrm{rk}(F_A)$ hat.*

Beweis. Wende Satz 6.4 an auf die lineare Abbildung $F_A \colon K^n \to K^m$. Das gibt Basen $\mathcal{B}$ von K^n und $\mathcal{C}$ von K^m, so daß $\mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(F_A)$ die Form $(*)$ hat. Es ist

$$\mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(F_A) \;=\; \mathrm{T}_{\mathcal{C}}^{\mathcal{K}_m} \cdot A \cdot \mathrm{T}_{\mathcal{K}_n}^{\mathcal{B}}\,,$$

(5.5), mit $\mathcal{K}_n$ bzw. $\mathcal{K}_m$ die kanonischen Basen, wir können also $T = \mathrm{T}_{\mathcal{K}_m}^{\mathcal{C}}$ und $S = \mathrm{T}_{\mathcal{K}_n}^{\mathcal{B}}$ nehmen. Das bedeutet, in den Spalten von S (bzw. von T) stehen die Vektoren aus $\mathcal{B}$ (bzw. aus $\mathcal{C}$). □

Wir übertragen nun die Definition des Rangs auf Matrizen.

6.6 Definition. Sei $A \in \mathrm{M}_{m\times n}(K)$. Sei $\mathrm{span}_s(A) \subseteq K^m$ der von den Spalten von A aufgespannte Unterraum von K^m (*Spaltenraum* von A), und sei $\mathrm{srk}(A) := \dim \mathrm{span}_s(A)$, der *Spaltenrang* von A. Weiter setzen wir $\mathrm{span}_z(A) := \mathrm{span}_s(A^t)$ und $\mathrm{zrk}(A) := \dim \mathrm{span}_z(A) = \mathrm{srk}(A^t)$, der *Zeilenraum* bzw. *Zeilenrang* von A. (Die Notationen srk und zrk werden gleich wieder abgeschafft.)

6.7 Bemerkung. Nach Definition ist $\mathrm{span}_s(A) = \mathrm{im}(F_A)$, also folgt $\mathrm{srk}(A) = \mathrm{rk}(F_A)$.

6.8 Lemma. *Sei $A \in \mathrm{M}_{m\times n}(K)$, seien $U \in \mathrm{GL}_m(K)$ und $V \in \mathrm{GL}_n(K)$. Dann gilt*

(a) $\mathrm{span}_s(AV) = \mathrm{span}_s(A)$ *und* $\mathrm{span}_z(UA) = \mathrm{span}_z(A)$,
(b) $\mathrm{zrk}(UAV) = \mathrm{zrk}(A)$ *und* $\mathrm{srk}(UAV) = \mathrm{srk}(A)$.

Beweis. Wir zeigen zunächst die Spaltenaussagen. Es ist

$$\mathrm{span}_s(AV) \;=\; \mathrm{im}(F_{AV}) \;=\; \mathrm{im}(F_A \circ F_V) \;=\; \mathrm{im}(F_A) \;=\; \mathrm{span}_s(A),$$

die vorletzte Gleichheit wegen F_V surjektiv. Ebenso ist

$$\mathrm{srk}(UAV) \;=\; \mathrm{rk}(F_{UAV}) \;=\; \mathrm{rk}(F_U \circ F_A \circ F_V) \;=\; \mathrm{rk}(F_A) \;=\; \mathrm{srk}(A),$$

die vorletzte Gleichheit wegen F_V surjektiv und F_U injektiv. Die Zeilenaussagen folgen aus den Spaltenaussagen durch Transposition. □

6.9 Satz und Definition. *Für jede Matrix $A \in \mathrm{M}_{m\times n}(K)$ ist* $\mathrm{zrk}(A) = \mathrm{srk}(A) = \mathrm{rk}(F_A)$. *Man nennt diese Zahl den* Rang *von A, i. Z.* $\mathrm{rk}(A)$.

BEWEIS. Nach Satz 6.5 gibt es $S \in \mathrm{GL}_n(K)$ und $T \in \mathrm{GL}_m(K)$, so daß $B := T^{-1}AS$ die Gestalt $(*)$ aus dem Rangsatz 6.4 hat, mit $r = \mathrm{rk}(F_A) = \mathrm{srk}(A)$. Offensichtlich ist auch $r = \mathrm{zrk}(B)$. Aus 6.8(b) folgt somit auch $r = \mathrm{zrk}(A)$. □

Für den Rang von Matrizen gelten die zu Satz 6.3 analogen Aussagen:

6.10 Korollar. *Sind $A \in \mathrm{M}_{m\times s}(K)$ und $B \in \mathrm{M}_{s\times n}(K)$, so ist*

$$\mathrm{rk}(AB) \;\le\; \min\{\mathrm{rk}(A),\, \mathrm{rk}(B)\},$$

und es gilt:

$$\mathrm{rk}(AB) \;=\; \begin{cases} \mathrm{rk}(A) & \textit{falls } \mathrm{rk}(B) = s, \\ \mathrm{rk}(B) & \textit{falls } \mathrm{rk}(A) = s. \end{cases}$$

BEWEIS. Das folgt sofort aus 6.3 (unter Benutzung von $\mathrm{rk}(C) = \mathrm{rk}(F_C)$ für jede Matrix C). □

Wir erhalten folgende alternative Charakterisierungen der invertierbaren Matrizen:

6.11 Korollar. *Für jede quadratische Matrix $A \in \mathrm{M}_n(K)$ sind äquivalent:*

(i) *A ist invertierbar (also $A \in \mathrm{GL}_n(K)$);*
(ii) $\mathrm{rk}(A) = n$;
(iii) *die Spalten von A sind linear unabhängig;*
(iv) *die Zeilen von A sind linear unabhängig.*

BEWEIS. Die Äquivalenz von (ii)–(iv) folgt aus Satz 6.9. Andererseits gilt $A \in \mathrm{GL}_n(K) \Leftrightarrow F_A$ bijektiv (3.10) $\Leftrightarrow F_A$ surjektiv (3.12) $\Leftrightarrow \mathrm{rk}(F_A) = n$, also (i) $\Leftrightarrow$ (ii). □

6.12 Satz und Definition. *Für je zwei Matrizen $A, B \in \mathrm{M}_{m\times n}(K)$ sind gleichwertig:*

(i) *Es gibt $S \in \mathrm{GL}_n(K)$ und $T \in \mathrm{GL}_m(K)$ mit $B = T^{-1}AS$;*
(ii) $\mathrm{rk}(A) = \mathrm{rk}(B)$.

Sind diese Bedingungen erfüllt, so heißen die Matrizen A und B äquivalent*, i. Z. $A \sim B$. Dies ist eine Äquivalenzrelation auf $\mathrm{M}_{m\times n}(K)$.*

BEWEIS. (i) $\Rightarrow$ (ii) folgt aus Lemma 6.8(b). Für (ii) $\Rightarrow$ (i) sei $\mathrm{rk}(A) = \mathrm{rk}(B) = r$. Nach Satz 6.5 gibt es S, $S' \in \mathrm{GL}_n(K)$ und T, $T' \in \mathrm{GL}_m(K)$ mit

$$T^{-1}AS \;=\; \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \;=\; T'^{-1}BS'.$$

Es folgt $B = T'T^{-1}ASS'^{-1} = (TT'^{-1})^{-1} \cdot A \cdot (SS'^{-1})$. □

6.13 Wir geben eine Anwendung des Rangs auf lineare Gleichungssysteme $Ax = u$ mit $A = (a_{ij}) \in \mathrm{M}_{m\times n}(K)$ und $u \in K^m$ (siehe 3.22):

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= u_1 \\ \vdots \qquad\qquad\quad \vdots \quad &\;\;\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= u_m \end{aligned}$$

Wie in 3.22 bezeichnet $\mathcal{L}(A,u) = \{x \in K^n \colon Ax = u\}$ die Lösungsmenge des Gleichungssystems. Die $m \times (n+1)$-Matrix

$$(A,u) \;:=\; \begin{pmatrix} a_{11} & \cdots & a_{1n} & u_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & u_m \end{pmatrix}$$

(bei der zu A als $(n+1)$-te Spalte der Spaltenvektor u hinzugefügt wurde) heißt die *erweiterte Koeffizientenmatrix* des Gleichungssystems $Ax = u$. Da zu A eine Spalte hinzugefügt wurde, ist $\operatorname{rk}(A,u) = \operatorname{rk}(A)$ oder $\operatorname{rk}(A,u) = \operatorname{rk}(A)+1$.

6.14 Satz. *Sei $A \in \mathrm{M}_{m\times n}(K)$.*

(a) *Sei $u \in K^m$. Dann ist $\mathcal{L}(A,u)$ ein affiner Unterraum von K^n, von Dimension $n - \operatorname{rk}(A)$ falls $\mathcal{L}(A,u) \neq \varnothing$. Genau dann ist $\mathcal{L}(A,u) \neq \varnothing$, wenn $\operatorname{rk}(A,u) = \operatorname{rk}(A)$ ist.*

(b) *$Ax = u$ ist lösbar für jedes $u \in K^m \;\Leftrightarrow\; \operatorname{rk}(A) = m \;\Leftrightarrow$ die Zeilen von A sind linear unabhängig.*

Beweis. (a) Ist $\mathcal{L}(A,u) \neq \varnothing$, so ist $\mathcal{L}(A,u)$ ein affiner Unterraum von Dimension $\dim\ker(F_A) = n - \dim\operatorname{im}(F_A) = n - \operatorname{rk}(A)$ (3.11, 3.21). Dabei gilt $\mathcal{L}(A,u) \neq \varnothing$ $\Leftrightarrow u \in \operatorname{im}(F_A) = \operatorname{span}_s(A) \Leftrightarrow \operatorname{span}_s(A,u) = \operatorname{span}_s(A) \Leftrightarrow \operatorname{rk}(A,u) = \operatorname{rk}(A)$. — (b) $\mathcal{L}(A,u) \neq \varnothing$ für jedes $u \Leftrightarrow F_A$ ist surjektiv $\Leftrightarrow \operatorname{rk}(A) = m$. □

7. Der Gauß-Algorithmus

Sei stets K ein Körper. Gegeben sei eine Matrix $A = (a_{ij}) \in \mathrm{M}_{m\times n}(K)$ und ein Vektor $u = (u_1, \ldots, u_m)^t \in K^m$. Wir wollen das lineare Gleichungssystem $Ax = u$ lösen, also die Lösungsmenge

$$\mathcal{L}(A,u) \;=\; \{x \in K^n \colon Ax = u\}$$

bestimmen. Diese ist leer oder ein affiner Unterraum von Dimension $n - \operatorname{rk}(A)$ von K^n (6.13). Wir beginnen mit einem Fall, wo die Lösung leicht gelingt.

7.1 Definition. Die Matrix $A = (a_{ij}) \in \mathrm{M}_{m\times n}(K)$ hat *Zeilenstufenform*, wenn A so aussieht

$$\begin{pmatrix} 0 & \cdots & 0 & * & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdots & \cdots & 0 & * & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdots & \cdots & \cdots & 0 & * & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ & & & & & \ddots & & & & & & & \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & * & \cdot & \cdot & & & \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & & & \\ \vdots & & & & & & & & & \vdots & & & \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & & & \end{pmatrix}$$

und dabei die Elemente $*$ nicht Null sind (und rechts oben beliebige Elemente stehen). Präziser: A hat Zeilenstufenform, wenn es $0 \le r \le m$ und $1 \le k_1 < \cdots < k_r \le n$ so gibt, daß gilt:

- $a_{ij} = 0$ für $1 \le i \le r$ und $1 \le j < k_i$;

- $a_{ik_i} \neq 0$ für $1 \leq i \leq r$;
- $a_{ij} = 0$ für $r+1 \leq i \leq m$ und $1 \leq j \leq n$.

Das Element a_{ik_i} heißt auch das *Pivotelement* der i-ten Zeile, und die k_i-te Spalte heißt die i-te *Pivotspalte*. Die Matrix A in Zeilenstufenform hat

- *normierte Zeilenstufenform*, wenn $a_{ik_i} = 1$ $(1 \leq i \leq r)$ gilt (alle Pivotelemente sind 1);
- *strikte Zeilenstufenform*, wenn $a_{lk_i} = 0$ ist für $1 \leq i \leq r$ und alle $l \in \{1, \ldots, m\} \smallsetminus \{i\}$ (die i-te Pivotspalte ist ein Vielfaches von e_i, für $i = 1, \ldots, r$).

7.2 Bemerkung. Es habe A Zeilenstufenform wie in 7.1. Die ersten r Zeilen $z_1, \ldots, z_r$ von A sind linear unabhängig, denn aus $a_1 z_1 + \cdots + a_r z_r = 0$ folgt sukzessive $a_1 = 0$, $a_2 = 0, \ldots$, $a_r = 0$. Also ist $\operatorname{rk}(A) = r$, und die ersten r Zeilen von A sind eine Basis von $\operatorname{span}_z(A)$. Für $u \in K^m$ läßt sich die Lösungsmenge $\mathcal{L}(A, u)$ wie folgt finden:

1. Fall: Ist $u_i \neq 0$ für ein $i \in \{r+1, \ldots, m\}$, so ist $\mathcal{L}(A, u) = \varnothing$.

2. Fall: Ist $u_{r+1} = \cdots = u_m = 0$, so kann man $x_j \in K$ für alle $j \in \{1, \ldots, n\} \smallsetminus \{k_1, \ldots, k_r\}$ beliebig vorgeben (also alle Nichtpivot-Variablen). Nach Fixierung solcher x_j gibt es dann eindeutig bestimmte $x_{k_1}, \ldots, x_{k_r} \in K$ mit $(x_1, \ldots, x_n) \in \mathcal{L}(A, u)$. Diese findet man rekursiv: Zunächst bestimmt man x_{k_r} aus der r-ten Zeile (beachte $a_{r,k_r} \neq 0$), dann $x_{k_{r-1}}$ aus der $(r-1)$-ten Zeile usw., bis man schließlich x_{k_1} aus der ersten Zeile findet.

Hat A sogar strikte Zeilenstufenform, so kann man sich das rekursive Vorgehen sparen und erhält x_{k_i} direkt aus der i-ten Zeile $(1 \leq i \leq r)$.

Falls also überhaupt eine Lösung existiert (2. Fall), so hat die Lösungsmenge $\mathcal{L}(A, u)$ $n - r$ "freie Parameter", nämlich die x_j für $j \in \{1, \ldots, n\} \smallsetminus \{k_1, \ldots, k_r\}$. Das entspricht gerade der Tatsache $\dim \mathcal{L}(A, u) = n - r$ (Satz 6.14(a)).

7.3 Beispiel. Als Beispiel betrachten wir das lineare Gleichungssystem

$$\begin{pmatrix} 0 & 2 & 0 & 5 & 0 \\ 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}$$

über $K = \mathbb{R}$. Die Matrix hat strikte Zeilenstufenform (mit $r = 3$ und $k_1 = 2$, $k_2 = 3$, $k_3 = 5$). Für $u_4 \neq 0$ ist $\mathcal{L} = \varnothing$. Für $u_4 = 0$ können wir x_1, $x_4 \in \mathbb{R}$ frei wählen und erhalten dann x_2, x_3, x_5 aus den drei Zeilen

$$\begin{aligned} 2x_2 + 5x_4 &= u_1, \\ -x_3 + 2x_4 &= u_2, \\ 3x_5 &= u_3 \end{aligned}$$

Die Lösungsmenge ist für $u_4 = 0$ also gleich

$$\mathcal{L} = \left\{ \begin{pmatrix} x_1 \\ \frac{1}{2}(u_1 - 5x_4) \\ -u_2 + 2x_4 \\ x_4 \\ \frac{1}{3}u_3 \end{pmatrix} : x_1, x_4 \in \mathbb{R} \right\} = \begin{pmatrix} 0 \\ \frac{1}{2}u_1 \\ -u_2 \\ 0 \\ \frac{1}{3}u_3 \end{pmatrix} + \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -5 \\ 4 \\ 2 \\ 0 \end{pmatrix} \right\}.$$

Um ein beliebiges Gleichungssystem $Ax = u$ zu lösen, addiert man so lange geeignete Vielfache einer Zeile zu einer anderen und vertauscht Zeilen, bis Zeilenstufenform erreicht ist. Dadurch wird $\mathcal{L}(A, u)$ nicht verändert. Wir systematisieren das Vorgehen durch folgende Definition:

7.4 Definition. Sei $A \in \mathrm{M}_{m\times n}(K)$. Für $k, l \in \{1, \dots, m\}$ mit $k \neq l$ und für $\lambda \in K$ sei

$$\text{(a)}\ \ \pi_{kl}(A), \quad \text{(b)}\ \ q_{kl}^{\lambda}(A), \quad \text{(c)}\ \ \delta_k^{\lambda}(A)$$

die Matrix, die aus A entsteht durch

(a) Vertauschen der k-ten und der l-ten Zeile,
(b) Addition des λ-fachen der l-ten Zeile zur k-ten Zeile,
(c) Multiplikation der k-ten Zeile mit λ.

Eine *elementare Zeilentransformation* ist der Übergang von A zu einer der Matrizen $\pi_{kl}(A)$, $q_{kl}^{\lambda}(A)$, $\delta_k^{\lambda}(A)$ mit $0 \neq \lambda \in K$.

7.5 Bemerkung. Elementare Zeilentransformationen von A sind Multiplikation von A mit geeigneten Matrizen von links. Um das zu erklären, seien $k, l \in \{1, \dots, m\}$ mit $k \neq l$. Für

$$\Pi_{kl} \ := \ \pi_{kl}(I_m) \ = \ E_{kl} + E_{lk} + \sum_{\substack{i=1\\ i\neq k,l}}^{m} E_{ii} \ \in \mathrm{M}_m(K)$$

gilt $\Pi_{kl}^2 = I_m$, insbesondere $\Pi_{kl} \in \mathrm{GL}_m(K)$. Für $\lambda \in K$ sei weiter

$$Q_{kl}(\lambda) \ := \ q_{kl}^{\lambda}(I_m) \ = \ I_m + \lambda E_{kl} \ \in \ \mathrm{M}_m(K).$$

Dann ist $Q_{kl}(0) = I_m$ und $Q_{kl}(\lambda) \cdot Q_{kl}(\mu) = Q_{kl}(\lambda + \mu)$ für $\lambda, \mu \in K$. Also ist auch $Q_{kl}(\lambda) \in \mathrm{GL}_m(K)$ (siehe auch Aufgabe 22). Schließlich sei

$$D_k(\lambda) \ := \ \delta_k^{\lambda}(I_m) \ = \ \mathrm{diag}\big(1, \dots, 1, \lambda, 1, \dots, 1\big) \ \in \ \mathrm{M}_m(K)$$

(mit λ an der k-ten Stelle). Für $\lambda \neq 0$ ist $D_k(\lambda) \in \mathrm{GL}_m(K)$. Es gilt:

7.6 Lemma. *Seien $k, l \in \{1, \dots, m\}$ mit $k \neq l$, und sei $\lambda \in K^*$. Für $A \in \mathrm{M}_{m\times n}(K)$ ist*

$$\pi_{kl}(A) = \Pi_{kl} A, \quad q_{kl}^{\lambda}(A) = Q_{kl}(\lambda)A, \quad \delta_k^{\lambda}(A) = D_k(\lambda)A.$$

Elementare Zeilenumformungen lassen $\mathrm{span}_z(A)$ *(und damit auch* $\mathrm{rk}(A)$*) unverändert.*

Beweis. Die erste Aussage sieht man sofort, die zweite folgt daraus und aus Lemma 6.8(a). □

7.7 Satz. (Gauß[3]-Algorithmus) *Jede Matrix $A \in \mathrm{M}_{m\times n}(K)$ kann durch eine endliche Folge von elementaren Zeilentransformationen π_{kl}, q_{kl}^{λ} in strikte Zeilenstufenform $\widetilde{A}$ gebracht werden. Die von Null verschiedenen Zeilen von $\widetilde{A}$ bilden eine Basis von* $\mathrm{span}_z(A)$*, ihre Anzahl ist also gleich* $\mathrm{rk}(A)$*. Verwendet man auch Operationen δ_k^{λ} (mit $\lambda \neq 0$), so kann man sogar normierte strikte Zeilenstufenform erreichen.*

[3]Carl Friedrich Gauss (1777–1855)

BEWEIS. Der Beweis ist wichtiger als die Aussage, denn er gibt einen Algorithmus, mit dem man eine solche Folge von Zeilentransformationen findet. Die erste von $(0,\dots,0)^t$ verschiedene Spalte sei die k_1-te. Durch eine Zeilenvertauschung bringen wir ein von 0 verschiedenes Element dieser Spalte in die erste Zeile, d. h. erreichen $a'_{1k_1} \neq 0$ für die neue Matrix $A' = (a'_{ij})$. Seien $z_1,\dots,z_m$ die Zeilen von A'. Für $i = 2,\dots,m$ ersetzen wir jetzt die i-te Zeile z_i durch $z_i - \frac{a'_{ik_1}}{a'_{1k_1}} z_1$, d. h. wenden q^λ_{i1} an mit

$$\lambda = -\frac{a'_{ik_1}}{a'_{1k_1}}.$$

Dann haben wir eine Matrix der Gestalt

$$\begin{pmatrix} 0 & \cdots & 0 & a'_{1k_1} & * & \cdots & * \\ 0 & \cdots & 0 & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & * & \cdots & * \end{pmatrix}$$

(mit irgendwelchen Elementen $*$ in K). Nun fahren wir auf dieselbe Weise mit dem rechten unteren Kästchen fort und erreichen so Zeilenstufenform nach endlich vielen Schritten. Aus dieser können wir bereits $\mathrm{rk}(A)$ und eine Basis des Zeilenraums $\mathrm{span}_z(A)$ von A ablesen. Durch weitere Operationen $q^\lambda_{lk_i}$ (für $1 \leq l < i \leq r$ und geeignete $\lambda \in K$) machen wir jetzt die Einträge über den Pivotelementen zu Null, erreichen also strikte Zeilenstufenform. Durch geeignete δ^λ_i $(1 \leq i \leq r)$ kann man die Pivotelemente auch noch zu 1 normieren. Wegen $\mathrm{span}_z(\widetilde{A}) = \mathrm{span}_z(A)$ (Lemma 7.6) bilden die von Null verschiedenen Zeilen von $\widetilde{A}$ eine Basis dieses Unterraums (siehe 7.2). □

Ein konkretes Beispiel dazu diskutieren wir gleich.

7.8 Korollar. *Zu jeder Matrix $A \in \mathrm{M}_{m\times n}(K)$ gibt es eine invertierbare Matrix $S \in \mathrm{GL}_m(K)$, so daß die Matrix SA normierte strikte Zeilenstufenform hat. Ein Algorithmus zum Auffinden eines solchen S wurde in 7.7 gegeben.*

BEWEIS. Das folgt aus dem Gauß-Algorithmus 7.7 und aus Lemma 7.6. □

7.9 Bemerkung. Mit dem Gauß-Algorithmus löst man beliebige lineare Gleichungssysteme $Ax = u$, $A \in \mathrm{M}_{m\times n}(K)$, $u \in K^m$: Bringe A durch elementare Zeilentransformationen π_{kl}, q^λ_{kl} auf (strikte) Zeilenstufenform und führe simultan dieselben Transformationen am Spaltenvektor u durch. So erhält man ein neues System $\widetilde{A}x = \widetilde{u}$ mit $\widetilde{A}$ in (strikter) Zeilenstufenform. Wegen $\widetilde{A} = SA$ und $\widetilde{u} = Su$ mit $S \in \mathrm{GL}_m(K)$ ist $\mathcal{L}(\widetilde{A},\widetilde{u}) = \mathcal{L}(A,u)$, das neue System hat dieselbe Lösungsmenge wie das alte. Gemäß 7.2 läßt es sich direkt explizit lösen. Falls gewünscht, kann man auch die Matrix S explizit finden, indem man die durchgeführten Transformationen gleichzeitig an einer zweiten Matrix durchführt, beginnend mit I_m.

Als nächstes zeigen wir, wie der Gauß-Algorithmus eine algorithmische Version des Basisauswahlsatzes gibt.

7.10 Notation. Zu $u_1, \dots, u_n \in K^m$ sei $M(u_1, \dots, u_n) \in \mathrm{M}_{m\times n}(K)$ die Matrix mit den Spalten $u_1, \dots, u_n$ (in dieser Reihenfolge). Entsprechend sei

$$M(u_1^t, \dots, u_n^t) \;:=\; M(u_1, \dots, u_n)^t \;\in \mathrm{M}_{n\times m}(K)$$

die Matrix mit den Zeilen $u_1^t, \dots, u_n^t$.

7.11 Satz. *Sei $A = M(u_1, \dots, u_n)$ mit $u_1, \dots, u_n \in K^m$. Die Zeilenstufenmatrix $\widetilde{A}$ entstehe aus A durch elementare Zeilenumformungen. Sind $k_1, \dots, k_r \in \{1, \dots, n\}$ die Indices der Pivotspalten von $\widetilde{A}$ (mit $r = \mathrm{rk}(A)$), so ist $(u_{k_1}, \dots, u_{k_r})$ eine Basis von* $\mathrm{span}_s(A) = \mathrm{span}(u_1, \dots, u_n)$.

Beweis. Sei $\widetilde{A} = M(\widetilde{u}_1, \dots, \widetilde{u}_n)$. Die Folge $(\widetilde{u}_{k_1}, \dots, \widetilde{u}_{k_r})$ der r Pivotspalten von $\widetilde{A}$ ist linear unabhängig, ist also wegen $\dim \mathrm{span}_s(\widetilde{A}) = r$ eine Basis von $\mathrm{span}_s(\widetilde{A})$. Es gibt $S \in \mathrm{GL}_m(K)$ mit $\widetilde{A} = SA = M(Su_1, \dots, Su_n)$, d.h. es ist $\widetilde{u}_j = Su_j$ für $j = 1, \dots, n$, und die lineare Abbildung F_S ist ein Isomorphismus von $\mathrm{span}_s(A)$ auf $\mathrm{span}_s(\widetilde{A})$. Daraus folgt die Behauptung. □

7.12 Bemerkung. Satz 7.11 gibt einen algorithmischen Zugang zum Basisauswahlsatz: Sind $u_1, \dots, u_n \in K^m$, so finden wir eine Teilmenge $I \subseteq \{1, \dots, n\}$, so daß $(u_i)_{i\in I}$ eine Basis von $\mathrm{span}(u_1, \dots, u_n)$ ist. Dazu bringen wir die Matrix $M(u_1, \dots, u_n)$ durch elementare Zeilenumformungen in Zeilenstufenform und nehmen für I die Menge der Indices der Pivotspalten.

7.13 Beispiel. Hier ist ein konkretes Beispiel. Sei

$$A \;=\; \begin{pmatrix} 0 & -1 & 5 & 3 \\ 3 & 4 & 16 & -3 \\ -2 & -3 & -9 & 3 \end{pmatrix} \;\in\; \mathrm{M}_{3\times 4}(\mathbb{R}),$$

sei $u = (u_1, u_2, u_3)^t \in \mathbb{R}^3$. Um das Gleichungssystem $Ax = u$ zu lösen, bringe die erweiterte Matrix (A, u) in Zeilenstufenform (linke und mittlere Spalte):

$$\begin{array}{rrrrcccc}
0 & -1 & 5 & 3 & u_1 & 1 & 0 & 0 \\
3 & 4 & 16 & -3 & u_2 & 0 & 1 & 0 \\
-2 & -3 & -9 & 3 & u_3 & 0 & 0 & 1 \\
\hline
3 & 4 & 16 & -3 & u_2 & 0 & 1 & 0 \\
0 & -1 & 5 & 3 & u_1 & 1 & 0 & 0 \\
-2 & -3 & -9 & 3 & u_3 & 0 & 0 & 1 \\
\hline
1 & 1 & 7 & 0 & u_2 + u_3 & 0 & 1 & 1 \\
0 & -1 & 5 & 3 & u_1 & 1 & 0 & 0 \\
-2 & -3 & -9 & 3 & u_3 & 0 & 0 & 1 \\
\hline
1 & 1 & 7 & 0 & u_2 + u_3 & 0 & 1 & 1 \\
0 & -1 & 5 & 3 & u_1 & 1 & 0 & 0 \\
0 & -1 & 5 & 3 & 2u_2 + 3u_3 & 0 & 2 & 3 \\
\hline
\end{array}
\qquad
\begin{array}{l}
\\ \\ \\ \Pi_{12} \\ \\ \\ Q_{13}(1) \\ \\ \\ Q_{31}(2) \\ \\ \\ Q_{32}(-1)
\end{array}$$

$$\begin{array}{cccc|c|ccc|l}
1 & 1 & 7 & 0 & u_2+u_3 & 0 & 1 & 1 & \\
0 & -1 & 5 & 3 & u_1 & 1 & 0 & 0 & \\
0 & 0 & 0 & 0 & -u_1+2u_2+3u_3 & -1 & 2 & 3 & \\
\hline
1 & 0 & 12 & 3 & u_1+u_2+u_3 & 1 & 1 & 1 & \\
0 & -1 & 5 & 3 & u_1 & 1 & 0 & 0 & \\
0 & 0 & 0 & 0 & -u_1+2u_2+3u_3 & -1 & 2 & 3 &
\end{array} \qquad Q_{12}(1)$$

Sei $\widetilde{A}$ die letzte Matrix in der linken Spalte (in strikter Zeilenstufenform). Man liest ab:

(1) Für $-u_1+2u_2+3u_3 \neq 0$ ist $\mathcal{L}(A,u) = \varnothing$.

(2) Für $-u_1+2u_2+3u_3 = 0$ kann man x_3, x_4 frei wählen und erhält daraus x_1, x_2. Es ergibt sich

$$\mathcal{L}(A,u) = \begin{pmatrix} u_1+u_2+u_3 \\ -u_1 \\ 0 \\ 0 \end{pmatrix} + \operatorname{span}\left\{ \begin{pmatrix} -12 \\ 5 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 3 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Dabei ist der rechte Span gleich $\mathcal{L}(A,0) = \ker(F_A)$, und $(-12,5,1,0)^t$, $(-3,3,0,1)^t$ bilden eine Basis von $\ker(F_A)$.

(3) Gleichzeitig sieht man, daß $\operatorname{rk}(A) = 2$ ist, daß die beiden ersten Zeilen von $\widetilde{A}$ eine Basis von $\operatorname{span}_z(A)$ bilden, und daß die beiden ersten Spalten von A eine Basis von $\operatorname{span}_s(A)$ bilden (Satz 7.11).

(4) Um eine explizite Matrix $S \in \mathrm{GL}_3(\mathbb{R})$ mit $SA = \widetilde{A}$ zu finden, führt man die obigen Zeilentransformationen noch einmal durch, beginnend mit I_3. Das gibt die rechte Spalte, und somit

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

Gemäß der obigen Buchführung ist

$$S = Q_{12}(1) \cdot Q_{32}(-1) \cdot Q_{31}(2) \cdot Q_{13}(1) \cdot \Pi_{12}.$$

7.14 Bemerkung. Es ist bei den elementaren Zeilentransformationen nicht erlaubt, für zwei Indizes $k \neq l$ *gleichzeitig* die Transformationen

$$z_k' := z_k + a z_l \quad \text{und} \quad z_l' := z_l + b z_k$$

mit a, $b \in K$ zu machen (beliebter Fehler!). Das führt im allgemeinen zu falschen Ergebnissen, wie folgendes Beispiel zeigt: Sei $A = \begin{pmatrix} 1 & 2 & 1 \\ -2 & -4 & 0 \end{pmatrix} \in \mathrm{M}_{2\times 3}(\mathbb{R})$, es ist $\operatorname{rk}(A) = 2$. Die nicht erlaubte Operation

$$z_1' := z_1 + \frac{1}{2} z_2, \quad z_2' := z_2 + 2z_1$$

gibt die neue Matrix $A' = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ mit $\operatorname{rk}(A') = 1$.

Der Gauß-Algorithmus leistet tatsächlich noch wesentlich mehr. Im folgenden zeigen wir einige wichtige weitere Anwendungen.

7.15 Bemerkung. Hier ist eine algorithmische Form des Basisergänzungssatzes. Seien $u_1, \dots, u_s \in K^n$, und sei $U = \operatorname{span}(u_1, \dots, u_s)$. Um ein lineares Komplement von U in K^n zu finden, also einen Unterraum $W \subseteq K^n$ mit $K^n = U \oplus W$, betrachte die Matrix

$$A \;=\; M(u_1, \dots, u_s, e_1, \dots, e_n) \;\in\; \mathrm{M}_{n\times(s+n)}(K)$$

(mit den Spalten $u_1, \dots, u_s, e_1, \dots, e_n$). Bringe A durch elementare Zeilenumformungen in Zeilenstufenform $\widetilde{A}$. Wegen $\operatorname{rk}(A) = n$ gibt es n Pivotspalten in $\widetilde{A}$, deren Indices seien $1 \le k_1 < \cdots < k_n \le n + s$. Sei dabei r der Index mit $k_r \le s < k_{r+1}$. Für $i = r + 1, \dots, n$ schreibe $k_i = s + l_i$ mit $l_i \ge 1$. Dann ist $(u_{k_1}, \dots, u_{k_r})$ eine Basis von U (Satz 7.11), also $\dim(U) = r$. Behaupte, für

$$W \;=\; \operatorname{span}(e_{l_{r+1}}, \dots, e_{l_n})$$

gilt $U \oplus W = K^n$.

In der Tat, die Pivotspalten von $\widetilde{A}$ sind eine Basis von $\operatorname{span}_s(\widetilde{A}) = K^n$. Ist $\widetilde{A} = SA$ mit $S \in \mathrm{GL}_n(K)$, so sind das die Spalten

$$Su_{k_1}, \dots, Su_{k_r}, Se_{l_{r+1}}, \dots, Se_{l_n}$$

von $\widetilde{A}$. Da $F_S \colon K^n \to K^n$ ein Isomorphismus ist, ist auch $(u_{k_1}, \dots, u_{k_r}, e_{l_{r+1}}, \dots, e_{l_n})$ eine Basis von K^n, woraus $U \oplus W = K^n$ folgt.

7.16 Bemerkung. Zu einem durch Erzeuger gegebenen Unterraum U von K^n kann man eine Matrix $A \in \mathrm{M}_{s\times n}(K)$ (für ein $s \in \mathbb{N}$) finden mit $\ker(F_A) = U$, also ein homogenes lineares Gleichungssystem mit Lösungsraum U. Das geht so:

Sei $U = \operatorname{span}(u_1, \dots, u_m)$, sei $B = M(u_1^t, \dots, u_m^t)$ (Matrix mit den Zeilen $u_1^t, \dots, u_m^t$). Finde mit dem Gauß-Algorithmus eine Basis $v_1, \dots, v_s$ von $\ker(F_B) = \mathcal{L}(B, 0)$ und bilde die Matrix $A = M(v_1^t, \dots, v_s^t)$. Dann ist $U = \ker(F_A)$. Denn für $i = 1, \dots, m$ ist

$$Au_i \;=\; \begin{pmatrix} v_1^t u_i \\ \vdots \\ v_s^t u_i \end{pmatrix} \;=\; \begin{pmatrix} u_i^t v_1 \\ \vdots \\ u_i^t v_s \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

wegen $v_j \in \ker(B)$, also gilt $U \subseteq \ker(F_A)$. Andererseits ist $\dim(U) = \operatorname{rk}(B)$ und

$$\operatorname{rk}(A) \;=\; s \;=\; \dim\ker(F_B) \;=\; n - \operatorname{rk}(B)$$

(3.11), also $\dim(U) = n - \operatorname{rk}(A) = \dim\ker(F_A)$, woraus $U = \ker(F_A)$ folgt.

Für ein Beispiel sei $U = K(3, 2, 1, 0)^t + K(0, 1, 2, 3)^t \subseteq K^4$. Für die Matrix $B = \begin{pmatrix} 3 & 2 & 1 & 0 \\ 0 & 1 & 2 & 3 \end{pmatrix}$ findet man

$$\ker(F_B) \;=\; K(1, -2, 1, 0)^t + K(2, -3, 0, 1)^t.$$

Also ist $U = \ker(F_A)$ mit $A = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 2 & -3 & 0 & 1 \end{pmatrix}$.

7.17 Bemerkung. Mit Bemerkung 7.16 kann man den Durchschnitt von zwei durch Erzeuger gegebenen Unterräumen von K^n berechnen. Sind nämlich

$$U \;=\; \operatorname{span}(u_1, \dots, u_r), \quad W \;=\; \operatorname{span}(w_1, \dots, w_s)$$

Unterräume von K^n, so finde mit 7.16 Matrizen A, B mit jeweils n Spalten und mit $\ker(F_A) = U$, $\ker(F_B) = W$. Das homogene lineare Gleichungssystem

$$\begin{pmatrix} A \\ B \end{pmatrix} x \;=\; 0$$

($x = (x_1, \ldots, x_n)^t$) hat dann den genauen Lösungsraum $U \cap W$. Mit dem Gaußschen Algorithmus erhält man also eine Basis von $U \cap W$.

7.18 Bemerkung. Für eine gegebene quadratische Matrix $A \in \mathrm{M}_n(K)$ können wir entscheiden, ob A invertierbar ist, und gegebenenfalls die inverse Matrix A^{-1} berechnen. Dazu betrachte Paare (X, Y) von $n \times n$-Matrizen und starte mit $X = A$ und $Y = I_n$. An beiden Matrizen führe jeweils dieselben Zeilentransformationen durch und bringe so X in Zeilenstufenform. Ist $\mathrm{rk}(A) < n$, so ist A nicht invertierbar. Ist $\mathrm{rk}(A) = n$, so kann man X durch weitere Zeilentransformationen in $X' = I_n$ überführen. Behaupte, dann ist $Y' = A^{-1}$.

In der Tat, man ersetzt das Paar $(X, Y) = (A, I_n)$ durch $(X', Y') = (SA, S)$ mit $S \in \mathrm{GL}_n(K)$. Aus $X' = I_n$ folgt also $S = A^{-1}$.

7.19 Beispiel. Um

$$A \;=\; \begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \\ 2 & 1 & 2 \end{pmatrix} \;\in\; \mathrm{M}_3(\mathbb{R})$$

zu invertieren, können wir z. B. so vorgehen:

$$\begin{array}{rrr|rrr|l}
1 & 2 & 0 & 1 & 0 & 0 & \\
1 & 1 & 0 & 0 & 1 & 0 & \\
2 & 1 & 2 & 0 & 0 & 1 & \\
\hline
1 & 2 & 0 & 1 & 0 & 0 & Q_{21}(-1),\ Q_{31}(-2) \\
0 & -1 & 0 & -1 & 1 & 0 & \\
0 & -3 & 2 & -2 & 0 & 1 & \\
\hline
1 & 0 & 0 & -1 & 2 & 0 & Q_{12}(2),\ Q_{32}(-3) \\
0 & -1 & 0 & -1 & 1 & 0 & \\
0 & 0 & 2 & 1 & -3 & 1 & \\
\hline
1 & 0 & 0 & -1 & 2 & 0 & \mathrm{diag}(1, -1, \tfrac{1}{2}) \\
0 & 1 & 0 & 1 & -1 & 0 & \\
0 & 0 & 1 & \tfrac{1}{2} & -\tfrac{3}{2} & \tfrac{1}{2} &
\end{array}$$

Die Matrix A ist also invertierbar, und die inverse Matrix ist

$$A^{-1} \;=\; \frac{1}{2} \begin{pmatrix} -2 & 4 & 0 \\ 2 & -2 & 0 \\ 1 & -3 & 1 \end{pmatrix}.$$

7.20 Bemerkungen.

1. Statt elementarer Zeilenumformungen kann man ebenso gut elementare Spaltenumformungen an der Matrix A durchführen:

- Vertauschen zweier Spalten,
- Addition eines Vielfachen einer Spalte zu einer anderen Spalte,
- Multiplikation einer Spalte mit einem Skalar $\lambda \neq 0$.

Diese entsprechen der Multiplikation von A mit geeigneten invertierbaren Matrizen von rechts. Alle bisher gemachten Bemerkungen gelten analog für Spalten- statt Zeilenumformungen, wenn man dabei jeweils die Rolle von Zeilen und Spalten vertauscht.

2. Zur Bestimmung etwa von $\mathrm{rk}(A)$ sind Spaltenumformungen genau so gut geeignet wie Zeilenumformungen, und man kann beide auch abwechselnd vornehmen. Zum Lösen von linearen Gleichungssystemen dagegen wird man in der Regel nur Zeilenumformungen zulassen, da Spaltenumformungen Veränderungen an den gesuchten Größen x_j bedeuten.

3. Um A^{-1} zu berechnen, kann man entweder Zeilen- oder Spaltenumformungen vornehmen, aber man darf (bei der in 7.18 beschriebenen Methode) die beiden nicht mischen. Denn aus $SAT = I$ folgt im allgemeinen nicht $A^{-1} = ST$.

7.21 Bemerkung. Das Gaußsche Eliminationsverfahren war schon lange vor Gauß bekannt, insbesondere schon in China um etwa 150 v. C., viel später dann auch in Europa (Newton).

KAPITEL IV

Determinanten

1. Vorzeichen von Permutationen

1.1 Zunächst erinnern wir an die symmetrische Gruppe (I.2.4.5): Für $n \in \mathbb{N}$ ist $S_n = \mathrm{Sym}(\{1, \ldots, n\})$ die symmetrische Gruppe der Menge $\{1, \ldots, n\}$. Die Elemente von S_n sind die Permutationen von $\{1, \ldots, n\}$, also die bijektiven Abbildungen von $\{1, \ldots, n\}$ in sich. Eine solche Abbildung σ schreiben wir als

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

mit $a_i = \sigma(i)$ $(i = 1, \ldots, n)$. Die Gruppenverknüpfung in S_n ist die Komposition der Abbildungen: Für $\rho, \sigma \in S_n$ ist $\rho \circ \sigma = \rho\sigma \in S_n$ definiert durch

$$(\rho \circ \sigma)(i) = \rho(\sigma(i)) \quad (i = 1, \ldots, n).$$

(Beachte, die rechts stehende Abbildung wird immer zuerst ausgeführt.) Das neutrale Element der Gruppe S_n ist die identische Abbildung $\mathrm{id} = \binom{1\ 2\ \cdots\ n}{1\ 2\ \cdots\ n}$. Die Ordnung (Mächtigkeit) von S_n ist $|S_n| = n!$.

1.2 Definition.

(a) Für $\sigma \in S_n$ sei $\mathrm{Fix}(\sigma) := \{i \in \{1, \ldots, n\} \colon \sigma(i) = i\}$, die Menge der *Fixpunkte* von σ.

(b) $\tau \in S_n$ heißt eine *Transposition*, wenn $|\mathrm{Fix}(\tau)| = n - 2$ ist.

1.3 Bemerkung. Für $i, j \in \{1, \ldots, n\}$ mit $i \neq j$ bezeichne $\tau_{ij} \in S_n$ diejenige Transposition, welche i und j vertauscht, also

$$\tau_{ij}(k) = \begin{cases} j, & k = i, \\ i, & k = j, \\ k, & k \notin \{i, j\} \end{cases} \qquad (k = 1, \ldots, n).$$

Es gibt genau $\binom{n}{2} = \frac{n}{2}(n-1)$ verschiedene Transpositionen in S_n, nämlich die $\tau_{ij} = \tau_{ji}$ für $i \neq j$ in $\{1, \ldots, n\}$. Für jede Transposition τ ist $\tau^2 = \tau \circ \tau = \mathrm{id}$.

1.4 Satz. *Jedes $\sigma \in S_n$ ist Produkt von höchstens $n - 1$ Transpositionen in S_n.*

Beweis. Der Satz ist richtig für $\sigma = \mathrm{id}$, denn id ist Produkt von 0 Transpositionen. Für $\sigma \neq \mathrm{id}$ zeigen wir genauer: σ ist Produkt von höchstens $n - 1 - r$ Transpositionen mit $r := |\mathrm{Fix}(\sigma)|$. Der Beweis geschieht durch absteigende Induktion nach r.

Wegen $\sigma \neq \mathrm{id}$ ist $r \leq n - 2$, der Induktionsbeginn ist also $r = n - 2$. Dann ist σ eine Transposition, und das ist schon die Behauptung in diesem Fall. Sei jetzt

$|\mathrm{Fix}(\sigma)| = r$ mit $0 \le r < n-2$, und sei die Aussage für größeres r schon bewiesen. Wähle $i \in \{1,\dots,n\}$ mit $\sigma(i) \neq i$, und setze $\tau := \tau_{i,\sigma(i)}$ und $\rho := \tau \circ \sigma$. Dann ist $\rho(i) = \tau(\sigma(i)) = i$. Andererseits ist $\mathrm{Fix}(\sigma) \subseteq \{1,\dots,n\} \smallsetminus \{i,\sigma(i)\} = \mathrm{Fix}(\tau)$. Für $j \in \mathrm{Fix}(\sigma)$ ist also $\rho(j) = \tau(\sigma(j)) = \tau(j) = j$. Das zeigt $\mathrm{Fix}(\sigma) \subseteq \mathrm{Fix}(\rho)$, also ist $|\mathrm{Fix}(\rho)| \ge r+1$. Wäre $\rho = \mathrm{id}$, so wäre $\sigma = \tau^{-1} = \tau$, Widerspruch zu $r < n-2$. Also können wir die Induktionsvoraussetzung auf ρ anwenden. Danach ist ρ ein Produkt von höchstens $n-1-|\mathrm{Fix}(\rho)| \le n-1-(r+1) = n-2-r$ Transpositionen. Somit ist $\sigma = \tau \circ \rho$ ein Produkt von höchstens $n-1-r$ Transpositionen, wie behauptet. □

1.5 Bemerkung. Der Beweis war konstruktiv, liefert also zu gegebenem $\sigma \in S_n$ eine konkrete Darstellung von σ als Produkt von Transpositionen. Will man etwa

$$\sigma \;=\; \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix}$$

als Produkt von Transpositionen schreiben, so findet man nacheinander (etwa)

$$\sigma = \tau_{13} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 4 & 2 \end{pmatrix} = \tau_{13} \circ \tau_{23} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 5 & 4 & 3 \end{pmatrix} = \tau_{13} \circ \tau_{23} \circ \tau_{35}.$$

Die Darstellung als Produkt von Transpositionen ist aber nicht eindeutig, zum Beispiel ist auch

$$\sigma \;=\; \tau_{12}\tau_{15}\tau_{13} \;=\; \tau_{15}\tau_{13}\tau_{25} \;=\; \tau_{25}\tau_{14}\tau_{13}\tau_{23}\tau_{24} \;=\; \cdots$$

1.6 Definition. Sei $\sigma \in S_n$.

(a) Ein *Fehlstand* von σ ist ein Paar (i,j) mit $1 \le i < j \le n$ und $\sigma(i) > \sigma(j)$.

(b) Das *Signum* (oder *Vorzeichen*) von σ ist definiert als

$$\mathrm{sgn}(\sigma) \;:=\; (-1)^{f(\sigma)} \;\in\; \{1,\,-1\},$$

wobei $f(\sigma)$ die Anzahl der Fehlstände von σ ist.

(c) σ heißt eine *gerade* oder *ungerade Permutation*, je nachdem ob $\mathrm{sgn}(\sigma) = 1$ oder $\mathrm{sgn}(\sigma) = -1$ ist.

1.7 Beispiele.

1. Es ist $f(\mathrm{id}) = 0$, also $\mathrm{sgn}(\mathrm{id}) = 1$.

2. Ist $\tau = \tau_{ij}$ eine Transposition mit $i < j$, so hat τ genau die folgenden Fehlstände:

$$(i,j), \;\text{ sowie alle } (i,k) \text{ und } (k,j) \text{ für } i < k < j.$$

Das sind genau $1 + 2(j-i-1) = 2j-2i-1$ Stück. Jede Transposition ist damit eine ungerade Permutation.

3. $\sigma = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{smallmatrix}\right)$ hat die Fehlstände $(1,3)$ und $(2,3)$, $\sigma^2 = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{smallmatrix}\right)$ hat die Fehlstände $(1,2)$ und $(1,3)$. Also sind σ und σ^2 gerade.

1.8 Satz. *Für beliebige ρ, $\sigma \in S_n$ ist* $\mathrm{sgn}(\rho \circ \sigma) = \mathrm{sgn}(\rho) \cdot \mathrm{sgn}(\sigma)$.

Beweis. Sei $\sigma \in S_n$, sei f die Anzahl der Fehlstände von σ. Dann ist

$$\prod_{1 \le i < j \le n} \big(\sigma(j) - \sigma(i)\big) \;=\; (-1)^f \cdot \prod_{1 \le i < j \le n} (j-i).$$

Denn durchläuft (i,j) alle $\binom{n}{2}$ Paare mit $1 \le i < j \le n$, so durchläuft auch die Menge $\{\sigma(i),\sigma(j)\}$ alle 2-elementigen Teilmengen von $\{1,\dots,n\}$. Beide Seiten

haben also denselben Absolutbetrag, und der Vorzeichenunterschied ist $(-1)^f$. Es folgt

$$\operatorname{sgn}(\sigma) \;=\; \prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i}.$$

Ist auch $\rho \in S_n$, so ist also

$$\operatorname{sgn}(\rho\circ\sigma) = \prod_{i<j} \frac{\rho\circ\sigma(j)-\rho\circ\sigma(i)}{j-i} = \prod_{i<j} \frac{\rho(\sigma(j))-\rho(\sigma(i))}{\sigma(j)-\sigma(i)} \cdot \prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i}.$$

Das zweite Produkt auf der rechten Seite ist $\operatorname{sgn}(\sigma)$. Das erste Produkt auf der rechten Seite ist $\operatorname{sgn}(\rho)$, denn mit $\{i,j\}$ durchläuft auch $\{\sigma(i),\sigma(j)\}$ alle 2-elementigen Teilmengen von $\{1,\dots,n\}$, und der Quotient

$$\frac{\rho(b)-\rho(a)}{b-a}$$

bleibt bei Vertauschen von a und b unverändert. Der Satz ist bewiesen. □

Wir sammeln einige Folgerungen aus Satz 1.8:

1.9 Korollar. *Die Abbildung* $\operatorname{sgn}\colon S_n \to \{1,-1\}$ *ist ein Gruppenhomomorphismus von* S_n *in die multiplikative Gruppe* $\{1,-1\}$*, surjektiv für* $n \geq 2$. □

1.10 Korollar. *Ist* $\sigma = \tau_1 \circ \cdots \circ \tau_r$ *mit Transpositionen* $\tau_1,\dots,\tau_r$*, so ist* $\operatorname{sgn}(\sigma) = (-1)^r$. □

Auch wenn die Zerlegung einer Permutation σ in Transpositionen nicht eindeutig ist, sehen wir hieraus, daß die Anzahl der Transpositionen entweder stets gerade oder stets ungerade ist, je nachdem ob σ gerade oder ungerade ist.

1.11 Korollar. *Für* $\sigma \in S_n$ *ist* $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$. □

1.12 Definition. Die Menge

$$A_n \;:=\; \big\{\sigma \in S_n\colon \operatorname{sgn}(\sigma) = 1\big\} \;=\; \ker\big(\operatorname{sgn}\colon S_n \to \{1,-1\}\big)$$

aller geraden Permutationen ist ein Normalteiler in S_n und heißt die *alternierende Gruppe* (von $\{1,\dots,n\}$).

1.13 Korollar. *Für* $n \geq 2$ *ist* $S_n/A_n \cong \{1,-1\}$ *und* $|A_n| = \frac{1}{2}n!$.

Beweis. Das folgt aus dem Homomorphiesatz (Korollar III.4.10), angewandt auf $\operatorname{sgn}\colon S_n \to \{1,-1\}$, und dem Satz von Lagrange (Korollar III.4.5). □

Insbesondere sehen wir $A_2 = \{\mathrm{id}\}$ und $A_3 = \{\mathrm{id},\sigma,\sigma^2\} = \langle\sigma\rangle$ mit $\sigma = \left(\begin{smallmatrix}1&2&3\\2&3&1\end{smallmatrix}\right)$ (Beispiel 1.7.3).

Wir führen nun eine viel bequemere Notation für Permutationen ein.

1.14 Definition. Sei $2 \leq r \leq n$. Eine Permutation $\sigma \in S_n$ heißt ein *r-Zykel* (oder *Zykel der Länge r*), wenn es paarweise verschiedene $i_1,\dots,i_r \in \{1,\dots,n\}$ gibt mit $\sigma(i_k) = i_{k+1}$ $(k = 1,\dots,r-1)$, $\sigma(i_r) = i_1$ und $\sigma(j) = j$ für alle $j \in \{1,\dots,n\} \setminus \{i_1,\dots,i_r\}$. Man notiert σ dann in der Form

$$\sigma \;=:\; (i_1\ i_2\ \cdots\ i_r).$$

Zwei Zykel $(i_1 \cdots i_r)$ und $(j_1 \cdots j_s)$ heißen *disjunkt*, wenn gilt

$$\{i_1, \ldots, i_r\} \cap \{j_1, \ldots, j_s\} = \varnothing.$$

1.15 Bemerkungen.

1. 2-Zykel sind dasselbe wie Transpositionen.
2. Einen r-Zykel kann man auf r verschiedene Weisen hinschreiben:

$$(i_1\ i_2\ \cdots\ i_r) = (i_2\ i_3\ \cdots\ i_r\ i_1) = \cdots = (i_r\ i_1\ \cdots\ i_{r-1}),$$

d.h. man kann die Einträge zyklisch vertauschen, ohne die Permutation zu ändern.

1.16 Satz.

(a) *Ist σ ein r-Zykel, so ist* $\operatorname{sgn}(\sigma) = (-1)^{r-1}$.
(b) *Sind ρ, σ disjunkte Zykel, so ist $\rho \circ \sigma = \sigma \circ \rho$.*
(c) *Jedes $\sigma \in S_n$ ist ein Produkt von paarweise disjunkten Zykeln, und diese Darstellung ist eindeutig bis auf die Reihenfolge der Faktoren.*

Beweis. (a) Seien $i_1, \ldots, i_r$ paarweise verschieden ($2 \le r \le n$). Dann ist

$$(i_1\ i_2\ \cdots\ i_r) = (i_1\ i_r) \cdots (i_1\ i_3)\,(i_1\ i_2)$$

ein Produkt von $r-1$ Transpositionen, das Signum ist also $(-1)^{r-1}$ nach 1.10. (b) ist klar. Für (c) zeigen wir durch absteigende Induktion nach $|\mathrm{Fix}(\sigma)|$: σ ist ein Produkt von paarweise disjunkten Zykeln, welche alle zu $\mathrm{Fix}(\sigma)$ disjunkt sind. Der Induktionsbeginn ist $\sigma = \mathrm{id}$, in diesem Fall ist σ das leere Produkt. Sei $\sigma \neq \mathrm{id}$, fixiere $i \in \{1, \ldots, n\}$ mit $\sigma(i) \neq i$. Es gibt $k, l \in \mathbb{N}$ mit $k < l$ und $\sigma^k = \sigma^l$, also $\sigma^{l-k} = \mathrm{id}$. Insbesondere gibt es ein minimales $r > 1$ mit $\sigma^r(i) = i$. Dann sind i, $\sigma(i), \ldots, \sigma^{r-1}(i)$ paarweise verschieden, denn

$$\sigma^\mu(i) = \sigma^\nu(i) \quad \Rightarrow \quad \sigma^{\nu-\mu}(i) = \sigma^{-\mu}(\sigma^\nu(i)) = i.$$

Für den r-Zykel $\rho := \big(i\ \sigma(i)\ \cdots\ \sigma^{r-1}(i)\big)$ gilt einerseits $\mathrm{Fix}(\sigma) \subseteq \mathrm{Fix}(\rho)$, und andererseits $(\rho^{-1} \circ \sigma)(\sigma^j(i)) = \sigma^j(i)$ für alle j. Also ist

$$\mathrm{Fix}(\rho^{-1} \circ \sigma) = \mathrm{Fix}(\sigma) \cup \{i, \sigma(i), \ldots, \sigma^{r-1}(i)\},$$

und insbesondere $|\mathrm{Fix}(\rho^{-1} \circ \sigma)| > |\mathrm{Fix}(\sigma)|$. Die nach Induktion in der Zerlegung von $\rho^{-1} \circ \sigma$ vorkommenden Zykel sind also disjunkt zu ρ. Wegen $\sigma = \rho \circ (\rho^{-1} \circ \sigma)$ folgt die Behauptung für σ jetzt aus der Induktionsvoraussetzung für $\rho^{-1} \circ \sigma$.

Die Eindeutigkeit der Zerlegung zeigt man mit einer naheliegenden Induktion (Übung). □

1.17 Bemerkung. Ein Vorteil der Zykelzerlegung ist die kompaktere Notation: Die 6 Elemente der Gruppe S_3 sind etwa

$$\mathrm{id},\ (1\,2),\ (1\,3),\ (2\,3),\ (1\,2\,3),\ (1\,3\,2).$$

Vor allem aber wird in der Zykelzerlegung die Struktur der Permutation sichtbar. Es ist sehr einfach, die Zykelzerlegung einer gegebenen Permutation zu finden, der Beweis von 1.16 hat ein Verfahren gegeben. Beispiel:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 3 & 8 & 10 & 2 & 6 & 4 & 5 & 1 & 7 \end{pmatrix} = (1\,9)\,(2\,3\,8\,5)\,(4\,10\,7).$$

Hat man $\sigma \in S_n$ in Zykel zerlegt, so kann man $\operatorname{sgn}(\sigma)$ sofort ablesen, ohne Fehlstände zählen oder in Transpositionen zerlegen zu müssen. Für obiges $\sigma \in S_{10}$ ist etwa $\operatorname{sgn}(\sigma) = (-1)(-1)(+1) = 1$.

Das Inverse einer in Zykelzerlegung gegebenen Permutation läßt sich sofort hinschreiben: Ist

$$\sigma \;=\; (i_1\ i_2\ \cdots\ i_r)\,(j_1\ j_2\ \cdots\ j_s)\cdots,$$

so ist

$$\sigma^{-1} \;=\; (i_r\ \cdots\ i_2\ i_1)\,(j_s\ \cdots\ j_2\ j_1)\cdots$$

2. Determinante einer quadratischen Matrix

2.1 Bemerkung. Seien Punkte $v = (v_1, v_2)$ und $w = (w_1, w_2)$ in der Ebene $\mathbb{R}^2$ gegeben, und sei

$$P \;=\; \big\{av + bw\colon\ a,\ b \in [0,1]\big\},$$

das von v und w aufgespannte Parallelogramm. Für den Flächeninhalt $F(P)$ von P gilt dann

$$F(P) \;=\; |v_1w_2 - v_2w_1| \;=\; \left|\det\begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}\right|.$$

Das sieht man leicht durch elementargeometrische Überlegungen. Auch das Vorzeichen der Determinante läßt sich geometrisch interpretieren: Ist $\alpha = \measuredangle(v, w)$ der (orientierte) Winkel zwischen v und w, so ist $\det\left(\begin{smallmatrix} v_1 & w_1 \\ v_2 & w_2 \end{smallmatrix}\right) > 0$ für $0 < \alpha < \pi$, und < 0 für $-\pi < \alpha < 0$. Man kann also die 2×2-Determinante auffassen als "orientierter Flächeninhalt" des von den Spalten aufgespannten Parallelogramms. Analoges gilt auch in höheren Dimensionen, wenn die allgemeine Determinante und das n-dimensionale Volumen eingeführt sind.

2.2 Sei im folgenden R stets ein beliebiger kommutativer Ring, und sei

$$R^n \;=\; \{(x_1, \ldots, x_n)\colon x_1, \ldots, x_n \in R\},$$

eine abelsche Gruppe bezüglich komponentenweiser Addition

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

Außerdem können wir mit Elementen $a \in R$ multiplizieren:

$$a \cdot (x_1, \ldots, x_n) \;:=\; (ax_1, \ldots, ax_n),$$

und dabei gelten das Assoziativgesetz $a(bx) = (ab)x$ und die Distributivgesetze $a(x + y) = ax + ay$ und $(a + b)x = ax + bx$ ($a, b \in R$, $x, y \in R^n$), analog zu Vektorräumen. Wie bei Körpern schreiben wir

$$e_i \;=\; (\delta_{ij})_{j=1,\ldots,n} \;=\; (0, \ldots, 1, \ldots, 0) \;\in R^n$$

mit 1 an der i-ten Stelle ($i = 1, \ldots, n$). Für $v_1, \ldots, v_n \in R^m$ sei $M(v_1, \ldots, v_n)$ die $m \times n$-Matrix mit den Spalten $v_1, \ldots, v_n$ (siehe III.7.10).

Wir führen die Determinante "abstrakt" ein, indem wir von ihr gewisse Eigenschaften fordern und dann zeigen, daß es genau eine Definition von $\det(A)$ gibt, die diese Eigenschaften erfüllt.

2.3 Definition. Eine Abbildung $\delta\colon \mathrm{M}_n(R) \to R$ heißt eine *Determinantenabbildung*, wenn sie die folgenden Eigenschaften hat:

(D1) δ ist *multilinear* bezüglich der Spalten:

$$\begin{aligned}\delta\big(M(v_1,\dots,\,av_k+a'v'_k,\,\dots,v_n)\big)\\ =\ a\,\delta\big(M(v_1,\dots,v_k,\dots,v_n)\big)+a'\,\delta\big(M(v_1,\dots,v'_k,\dots,v_n)\big)\end{aligned}$$

für alle $k=1,\dots,n$, alle $v_1,\dots,v_n$, $v'_k\in R^n$ und alle a, $a'\in R$.

(D2) δ ist *alternierend*: $\delta(A)=0$, falls A zwei gleiche Spalten hat.

(D3) δ ist *normiert*: $\delta(I_n)=1$.

2.4 Lemma. *Sei $\delta\colon \mathrm{M}_n(R)\to R$ eine Abbildung mit (D1) und (D2), seien $v_1,\dots,v_n\in R^n$, und sei $\sigma\in S_n$. Dann ist*

$$\delta\big(M(v_{\sigma(1)},\dots,v_{\sigma(n)})\big)\ =\ \mathrm{sgn}(\sigma)\cdot\delta\big(M(v_1,\dots,v_n)\big).$$

Beweis. Es genügt, dies für Transpositionen zu zeigen, denn σ ist Produkt von Transpositionen (1.4) und das Vorzeichen sgn ist multiplikativ (1.8). Für $i<j$ ist (wir notieren ab der 2. Zeile nur die Stellen i und j, alle anderen seien jeweils fest):

$$\begin{aligned}0\ &=_{(D2)}\ \delta\big(M(v_1,\dots,v_{i-1},v_i+v_j,v_{i+1},\dots,v_{j-1},v_i+v_j,v_{j+1},\dots,v_n)\big)\\ &=_{(D1)}\ \delta\big(M(\dots,v_i,\dots,v_i,\dots)\big)+\delta\big(M(\dots,v_i,\dots,v_j,\dots)\big)\\ &\qquad\quad+\delta\big(M(\dots,v_j,\dots,v_i,\dots)\big)+\delta\big(M(\dots,v_j,\dots,v_j,\dots)\big)\\ &=_{(D2)}\ \delta\big(M(\dots,v_i,\dots,v_j,\dots)\big)+\delta\big(M(\dots,v_j,\dots,v_i,\dots)\big).\end{aligned}$$

Beim Vertauschen zweier Spalten wechselt δ also das Vorzeichen, wie behauptet. □

2.5 Lemma. *Sei $\delta\colon \mathrm{M}_n(R)\to R$ eine Abbildung mit (D1) und (D2). Für $A=(a_{ij})\in \mathrm{M}_n(R)$ gilt dann*

$$\delta(A)\ =\ \delta(I_n)\cdot\sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)\prod_{j=1}^{n}a_{\sigma(j),j}.$$

Insbesondere gibt es höchstens eine Determinantenabbildung $\delta\colon \mathrm{M}_n(R)\to R$.

Beweis. Für $j=1,\dots,n$ ist $\sum_{i=1}^n a_{ij}e_i$ die j-te Spalte von A. Aus (D1) folgt

$$\begin{aligned}\delta(A)\ &=\ \delta\Big(M\Big(\sum_{i_1=1}^{n}a_{i_1,1}e_{i_1},\ \dots,\ \sum_{i_n=1}^{n}a_{i_n,n}e_{i_n}\Big)\Big)\\ &=\ \sum_{i_1=1}^{n}\cdots\sum_{i_n=1}^{n}a_{i_1,1}\cdots a_{i_n,n}\cdot\delta\big(M(e_{i_1},\dots,e_{i_n})\big).\end{aligned}$$

Wegen (D2) ist $\delta\big(M(e_{i_1},\dots,e_{i_n})\big)\neq 0$ nur dann, wenn $i_1,\dots,i_n$ paarweise verschieden sind, also wenn es ein $\sigma\in S_n$ gibt mit $i_j=\sigma(j)$ für $j=1,\dots,n$. Nach Lemma 2.4 ist dann

$$\delta\big(M(e_{\sigma(1)},\dots,e_{\sigma(n)})\big)\ =\ \mathrm{sgn}(\sigma)\cdot\delta\big(M(e_1,\dots,e_n)\big)\ =\ \mathrm{sgn}(\sigma)\cdot\delta(I_n).$$

Es folgt die Behauptung. □

Wir *definieren* jetzt die Determinante durch die Formel aus 2.5:

2.6 Definition. Für eine Matrix $A = (a_{ij}) \in \mathrm{M}_n(R)$ ist die *Determinante* von A definiert durch

$$\det(A) := \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j}.$$

Die Formel für $\det(A)$ in 2.6 heißt auch die *Leibniz-Formel*[1]. Wir haben gezeigt: Wenn es überhaupt eine Determinantenabbildung gibt, dann ist sie gleich $A \mapsto \det(A)$ wie in 2.6 definiert. Die Umkehrung, nämlich daß 2.6 auch tatsächlich eine Determinantenabbildung ist, ist richtig, muß aber noch bewiesen werden:

2.7 Satz. *Für $\delta = \det$ wie in 2.6 gelten (D1), (D2) und (D3). Damit ist $\det\colon \mathrm{M}_n(R) \to R$ eine Determinantenabbildung, und ist die einzige solche.*

Beweis. Die Multilinearität (D1) ist klar, da jeder Summand in der Leibnizformel linear von jeder Spalte abhängt. Ebenso ist (D3) klar: $\det(I_n) = 1$. Wir beweisen (D2).

Sei $A = (a_{ij}) \in \mathrm{M}_n(R)$, seien $v_1, \ldots, v_n \in R^n$ die Spalten von A, und sei dabei $v_k = v_l$ mit $k \neq l$. Betrachte die Transposition $\tau := (k\ l) \in S_n$. Wegen $S_n = A_n \cup A_n\tau$ (disjunkte Vereinigung der Rechtsnebenklassen) ist

$$\det(A) = \sum_{\sigma \in A_n} \Big(a_{\sigma(1),1} \cdots a_{\sigma(n),n} - a_{\sigma\tau(1),1} \cdots a_{\sigma\tau(n),n} \Big).$$

Dabei ist jeder Summand $\big(\cdots\big) = 0$ (und damit $\det(A) = 0$). Denn sei $\sigma \in A_n$. Für $\nu \notin \{k, l\}$ ist $a_{\sigma(\nu),\nu} = a_{\sigma\tau(\nu),\nu}$, und andererseits ist

$$a_{\sigma(k),k} \cdot a_{\sigma(l),l} - a_{\sigma\tau(k),k} \cdot a_{\sigma\tau(l),l} = a_{\sigma(k),k} \cdot a_{\sigma(l),l} - a_{\sigma(l),k} \cdot a_{\sigma(k),l} = 0,$$

denn wegen $v_k = v_l$ gilt $a_{\sigma(k),k} = a_{\sigma(k),l}$ und $a_{\sigma(l),k} = a_{\sigma(l),l}$. □

2.8 Bemerkungen.

1. Für 2×2-Matrizen stimmt die alte *ad hoc* Definition (III.1.15, über Körpern) mit der neuen überein:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

2. Für 3×3-Matrizen gibt es die Merkregel von *Sarrus*:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{array}{c} a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12} \end{array}$$

3. Die Determinante einer Dreiecksmatrix ist das Produkt der Diagonalelemente:

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22} \cdots a_{nn}.$$

[1] Gottfried Wilhelm Leibniz (1646–1716)

Denn für $\sigma \neq \mathrm{id}$ ist $\prod_{j=1}^{n} a_{\sigma(j),j} = 0$, denn es gibt einen Index j mit $\sigma(j) > j$. Analog für untere Dreiecksmatrizen.

4. Für $a \in R$ und $A \in \mathrm{M}_n(R)$ ist $\det(aA) = a^n \det(A)$.

Transposition der Matrix läßt die Determinante unverändert:

2.9 Satz. *Für $A \in \mathrm{M}_n(R)$ ist $\det(A) = \det(A^t)$.*

Beweis. Mit der Substitution $\rho = \sigma^{-1}$ ist

$$\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j} \;=\; \sum_{\rho \in S_n} \operatorname{sgn}(\rho) \prod_{i=1}^{n} a_{i,\rho(i)}$$

wegen $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$. Die linke Seite ist $\det(A)$, die rechte Seite ist $\det(A^t)$. □

Folglich ist die Determinante auch multilinear und alternierend bezüglich der Zeilen, und sie ist die einzige normierte solche Abbildung.

Weniger offensichtlich und von zentraler Wichtigkeit ist die Multiplikativität der Determinante:

2.10 Theorem. *Für A, $B \in \mathrm{M}_n(R)$ ist* $\det(AB) = \det(A) \cdot \det(B)$.

Schon im Fall $n = 2$ ist der direkte Beweis dieser Identität etwas langwierig (siehe III.1.15). Für größeres n wäre es sehr aufwendig, den Beweis durch direktes Nachrechnen zu führen. Unsere abstrakte Charakterisierung der Determinante ermöglicht dagegen ein ganz simples Argument:

Beweis. Fixiere A und betrachte B als variabel, d. h. betrachte die Abbildung

$$\delta \colon \mathrm{M}_n(R) \to R, \quad \delta(B) \;:=\; \det(AB).$$

Hat B die Spalten $v_1, \ldots, v_n$, so hat AB die Spalten $Av_1, \ldots, Av_n$, d.h. es ist

$$A \cdot M(v_1, \ldots, v_n) \;=\; M(Av_1, \ldots, Av_n).$$

Die Abbildung δ erfüllt also (D1) und (D2). Nach Lemma 2.5 ist daher $\delta(B) = \delta(I_n) \cdot \det(B) = \det(A) \cdot \det(B)$ für alle $B \in \mathrm{M}_n(R)$. □

2.11 Bemerkung. Wie ändert sich die Determinante bei elementaren Umformungen? Beim Vertauschen von zwei Spalten wechselt $\det(A)$ das Vorzeichen (Lemma 2.4). Addiert man ein Vielfaches einer Spalte zu einer anderen Spalte, so ändert sich $\det(A)$ nicht, und Multiplikation einer Spalte mit $\lambda \in R$ multipliziert $\det(A)$ mit λ. Denn die beiden letzten Operationen ersetzen A durch SA mit $S = Q_{ij}(\lambda)$ bzw. $S = D_i(\lambda)$ (siehe III.7.5), und es ist $\det Q_{ij}(\lambda) = 1$, $\det D_i(\lambda) = \lambda$. Analoges gilt für Zeilenoperationen.

2.12 Bemerkung. Konkrete Determinanten mit vier oder mehr Zeilen rechnet man normalerweise nicht mit der Leibnizformel aus. Zumindest wenn K ein Körper ist, bringt man vielmehr A durch elementare Zeilen- oder Spaltenumformungen in Dreiecksform A' und liest $\det(A')$ direkt ab (Bemerkung 2.8.3). Zeilen- und Spaltenoperationen darf man dabei beliebig mischen. Bemerkung 2.11 zeigt, wie $\det(A)$ und $\det(A')$ zusammenhängen. Man muß also über die vorgenommenen Schritte Buch führen, um $\det(A)$ aus $\det(A')$ korrekt zu erhalten.

Zum Beispiel kann man für

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

so vorgehen (verwende die Bezeichnungen aus III.7.4 und schreibe π'_{ij} für das Vertauschen der Spalten i und j):

$$\begin{array}{rrrr|l} 1 & 5 & 0 & 2 & \\ -4 & 0 & 1 & 1 & \\ 1 & 3 & 2 & 2 & \\ 2 & 0 & -1 & 4 & \\ \hline & & & & q_{21}(4),\ q_{31}(-1),\ q_{41}(-2) \\ 1 & 5 & 0 & 2 & \\ 0 & 20 & 1 & 9 & \\ 0 & -2 & 2 & 0 & \\ 0 & -10 & -1 & 0 & \\ \hline & & & & \pi_{23} \\ 1 & 5 & 0 & 2 & \\ 0 & -2 & 2 & 0 & \\ 0 & 20 & 1 & 9 & \\ 0 & -10 & -1 & 0 & \\ \hline & & & & q_{32}(10),\ q_{42}(-5) \\ 1 & 5 & 0 & 2 & \\ 0 & -2 & 2 & 0 & \\ 0 & 0 & 21 & 9 & \\ 0 & 0 & -11 & 0 & \\ \hline & & & & \pi'_{34} \\ 1 & 5 & 2 & 0 & \\ 0 & -2 & 0 & 2 & \\ 0 & 0 & 9 & 21 & \\ 0 & 0 & 0 & -11 & \end{array}$$

Da insgesamt zwei Vertauschungen vorgenommen wurden, ist

$$\det(A) = (-1)^2 \cdot 1 \cdot (-2) \cdot 9 \cdot (-11) = 198.$$

Wir diskutieren nun den Zusammenhang zwischen Invertierbarkeit einer Matrix und ihrer Determinante, und verallgemeinern zunächst Definition III.1.13 auf Matrizen über kommutativen Ringen:

2.13 Definition. Sei R ein kommutativer Ring. Eine Matrix $A \in \mathrm{M}_n(R)$ heißt *invertierbar*, wenn es eine Matrix $A' \in \mathrm{M}_n(R)$ gibt mit $AA' = A'A = I_n$.

Wie über Körpern gilt:

2.14 Satz. *Sei R ein kommutativer Ring.*

(a) *Ist* $A \in \mathrm{M}_n(R)$ *invertierbar, so ist die Matrix* $A' \in \mathrm{M}_n(R)$ *mit* $AA' = A'A = I_n$ *eindeutig bestimmt. Sie heißt die zu* A *inverse Matrix und wird mit* A^{-1} *bezeichnet.*
(b) $\mathrm{GL}_n(R) := \{A \in \mathrm{M}_n(R)\colon A \text{ ist invertierbar}\}$ *ist eine Gruppe unter Multiplikation, genannt die* allgemeine lineare Gruppe.
(c) *Insbesondere ist* $R^* := \{u \in R\colon \exists\, u' \in R\ uu' = 1\}$ *eine Gruppe unter Multiplikation, genannt die* Einheitengruppe *des Rings* R*. Die Elemente von* R^* *heißen die* Einheiten *von* R*.*

Beweis. (a) und (b) beweist man genauso wie für Körper (Satz III.1.14). Aussage (c) ist der Spezialfall $n = 1$ von (b). □

2.15 Beispiele. Ist $R = K$ ein Körper, so ist die Einheitengruppe $K^* = K \smallsetminus \{0\}$, in Übereinstimmung mit der schon bisher verwendeten Notation (I.3.12). Für $R = K[t]$ mit K ein Körper ist $K[t]^* = \{f \in K[t]\colon \deg(f) = 0\}$, die Menge der von 0 verschiedenen konstanten Polynome. (Beweis?) Für $R = \mathbb{Z}$ ist $\mathbb{Z}^* = \{\pm 1\}$.

2.16 Satz. *Sei* R *ein kommutativer Ring. Eine quadratische Matrix* $A \in \mathrm{M}_n(R)$ *ist genau dann invertierbar, wenn* $\det(A)$ *eine Einheit in* R*, also* $\det(A) \in R^*$ *ist. Alsdann ist* $\det(A^{-1}) = \det(A)^{-1}$*.*

Beweis. Ist A invertierbar, so folgt $\det(A)\cdot\det(A^{-1}) = \det(I_n) = 1$ nach 2.10. Also ist dann $\det(A)$ eine Einheit in R und $\det(A)^{-1} = \det(A^{-1})$. Die Umkehrung beweisen wir vorerst nur im Fall, wo $R = K$ ein Körper ist: Ist A nicht invertierbar, so ist $\mathrm{rk}(A) = r < n$ (III.6.11), und es gibt S, $T \in \mathrm{GL}_n(K)$ mit

$$A \;=\; T^{-1} \cdot \mathrm{diag}(\underbrace{1,\dots,1}_{r},\underbrace{0,\dots,0}_{n-r}) \cdot S$$

(Rangsatz III.6.5). Wegen $n - r \geq 1$ folgt dann $\det(A) = 0$. Siehe 3.8 unten für die Umkehrung im allgemeinen Fall. □

Insbesondere gilt also für $R = K$ ein Körper:

2.17 Korollar. *Sei* K *ein Körper.*
(a) *Eine Matrix* $A \in \mathrm{M}_n(K)$ *ist genau dann invertierbar, wenn* $\det(A) \neq 0$ *ist.*
(b) *Für* $v_1, \dots, v_n \in K^n$ *ist* $(v_1, \dots, v_n)$ *linear unabhängig genau dann, wenn* $\det M(v_1, \dots, v_n) \neq 0$ *ist.*

Beweis. (a) wurde in 2.16 gezeigt. Daraus folgt (b), denn $(v_1, \dots, v_n)$ ist genau dann linear unabhängig, wenn $M(v_1, \dots, v_n)$ invertierbar ist (Korollar III.6.11). □

3. Spezielle Determinanten, Komplementärmatrix, Minoren

Weiterhin sei R stets ein kommutativer Ring.

3.1 Satz. (Kästchensatz) *Sei* $A \in \mathrm{M}_n(R)$*. Hat* A *die Gestalt*

$$A \;=\; \left(\begin{array}{c|c} A' & * \\ \hline 0 & A'' \end{array}\right) \quad \textit{oder} \quad A \;=\; \left(\begin{array}{c|c} A' & 0 \\ \hline * & A'' \end{array}\right)$$

mit $1 \le r < n$ *und* $A' \in \mathrm{M}_r(R)$, $A'' \in \mathrm{M}_{n-r}(R)$, *so ist*

$$\det(A) \;=\; \det(A') \cdot \det(A'').$$

BEWEIS. Es genügt, den ersten Fall zu betrachten (der zweite folgt daraus durch Transposition). Sei also $A = (a_{ij})$, und sei $a_{ij} = 0$ für alle (i,j) mit $1 \le j \le r < i \le n$. In der Leibnizformel

$$\det(A) \;=\; \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j}$$

brauchen wir nur über die $\sigma \in S_n$ mit $\sigma(\{1,\dots,r\}) = \{1,\dots,r\}$ zu summieren, denn alle anderen Summanden sind 0. (Für jedes andere σ gibt es ein $j \in \{1,\dots,r\}$ mit $\sigma(j) > r$, also mit $a_{\sigma(j),j} = 0$.) Es gelte also $\sigma(\{1,\dots,r\}) = \{1,\dots,r\}$, und damit auch $\sigma(\{r+1,\dots,n\}) = \{r+1,\dots,n\}$. Schreibe $\sigma = (\sigma', \sigma'')$, wobei $\sigma' \in S_r$ und $\sigma'' \in S_{n-r}$ definiert sind durch

$$\sigma'(j) \;:=\; \sigma(j) \quad (j = 1,\dots,r),$$
$$\sigma''(k) + r \;:=\; \sigma(k+r) \quad (k = 1,\dots,n-r).$$

Für die Zahl der Fehlstände gilt $f(\sigma) = f(\sigma') + f(\sigma'')$. Also ist $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma') \cdot \operatorname{sgn}(\sigma'')$, und es folgt

$$\begin{aligned} \det(A) &= \sum_{\substack{\sigma' \in S_r \\ \sigma'' \in S_{n-r}}} \operatorname{sgn}(\sigma')\operatorname{sgn}(\sigma'') \cdot \prod_{j=1}^{r} a_{\sigma'(j),j} \prod_{k=1}^{n-r} a_{\sigma''(k)+r,k+r} \\ &= \det(A') \cdot \det(A''). \end{aligned}$$

□

3.2 Bemerkung. Induktiv verallgemeinert man den Kästchensatz: Ist $n = n_1 + \cdots + n_r$ mit $n_i \in \mathbb{N}$, und sind $A_i \in \mathrm{M}_{n_i}(R)$ $(i = 1,\dots,r)$, so ist

$$\det \begin{pmatrix} \boxed{A_1} & * & \cdots & \boxed{*} \\ \boxed{0} & \boxed{A_2} & \cdots & \boxed{*} \\ \vdots & \vdots & \ddots & \vdots \\ \boxed{0} & \boxed{0} & \cdots & \boxed{A_r} \end{pmatrix} \;=\; \det(A_1) \cdots \det(A_r).$$

Als Anwendung berechnen wir eine wichtige Determinante:

3.3 Satz. (Vandermonde-Determinante[2]) *Für* $a_1,\dots,a_n \in R$ *ist*

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_n \\ a_1^2 & a_2^2 & \cdots & a_n^2 \\ \vdots & \vdots & & \vdots \\ a_1^{n-1} & a_2^{n-1} & \cdots & a_n^{n-1} \end{vmatrix} \;=\; \prod_{1 \le i < j \le n} (a_j - a_i).$$

Insbesondere folgt: Ist R ein Körper (oder ein nullteilerfreier Ring), und sind $a_1,\dots,a_n \in R$ paarweise verschieden, so ist die Vandermonde-Determinante nicht Null.

[2]Alexandre Théophile VANDERMONDE (1735-1796)

BEWEIS. Sei $D(a_1,\dots,a_n)$ die gesuchte Determinante. Für $i = n, n-1, \dots, 2$ (in dieser Reihenfolge!) ziehe das a_1-fache der $(i-1)$-ten Zeile von der i-ten Zeile ab, das gibt

$$D(a_1,\dots,a_n) \;=\; \begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & a_2 - a_1 & \cdots & a_n - a_1 \\ 0 & a_2^2 - a_1 a_2 & \cdots & a_n^2 - a_1 a_n \\ \vdots & \vdots & & \vdots \\ 0 & a_2^{n-1} - a_1 a_2^{n-2} & \cdots & a_n^{n-1} - a_1 a_n^{n-2} \end{vmatrix}.$$

Nach dem Kästchensatz und nach Herausziehen des Faktors $a_j - a_1$ aus der j-ten Spalte für $j = 2,\dots,n$ ergibt sich

$$\begin{aligned} D(a_1,\dots,a_n) &= \begin{vmatrix} 1 & \cdots & 1 \\ a_2 & \cdots & a_n \\ \vdots & & \vdots \\ a_2^{n-2} & \cdots & a_n^{n-2} \end{vmatrix} \cdot \prod_{j=2}^{n} (a_j - a_1) \\ &= D(a_2,\dots,a_n) \cdot \prod_{j=2}^{n} (a_j - a_1). \end{aligned}$$

Die Behauptung folgt also durch Induktion nach n. □

3.4 Definition. Sei $A = (a_{ij}) \in \mathrm{M}_n(R)$. Für feste $i, j \in \{1,\dots,n\}$ sei $A_{ij} \in \mathrm{M}_n(R)$ die Matrix, die aus A entsteht, indem man das Element a_{ij} durch 1 und alle anderen Elemente der i-ten Zeile und der j-ten Spalte durch 0 ersetzt:

$$A_{ij} \;=\; \begin{pmatrix} a_{11} & \cdots & a_{1,j-1} & 0 & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & 0 & a_{n,j+1} & \cdots & a_{nn} \end{pmatrix}$$

Die Matrix $A^\sharp := (a_{ij}^\sharp) \in \mathrm{M}_n(R)$ mit

$$a_{ij}^\sharp \;:=\; \det(A_{ji}) \quad (i, j = 1,\dots,n)$$

(sic!) heißt die zu A *komplementäre Matrix*.

3.5 Bemerkung. Sei $n > 1$. Für $i, j \in \{1,\dots,n\}$ bezeichne $A'_{ij} \in \mathrm{M}_{n-1}(R)$ die Matrix, die aus A entsteht, indem man i-te Zeile und j-te Spalte streicht:

$$A'_{ij} \;=\; \begin{pmatrix} a_{11} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{nn} \end{pmatrix}$$

Dann ist

$$\det(A_{ij}) \;=\; (-1)^{i+j}\det(A'_{ij}).$$

In der Tat, durch $i-1$ Zeilenvertauschungen und $j-1$ Spaltenvertauschungen bringt man A_{ij} in die Form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & A'_{ij} & \\ 0 & & & \end{pmatrix},$$

und mit dem Kästchensatz 3.1 folgt

$$\det(A_{ij}) \;=\; (-1)^{i+j-2}\,\det(A'_{ij}) \;=\; (-1)^{i+j}\,\det(A'_{ij}).$$

Für die zu A komplementäre Matrix $A^\sharp$ gilt also

$$A^\sharp \;=\; \begin{pmatrix} +\det(A'_{11}) & -\det(A'_{21}) & +\det(A'_{31}) & \cdots \\ -\det(A'_{12}) & +\det(A'_{22}) & -\det(A'_{32}) & \cdots \\ +\det(A'_{13}) & -\det(A'_{23}) & +\det(A'_{33}) & \cdots \\ \cdots & \cdots & \cdots & \ddots \end{pmatrix}$$

(Vorzeichen nach der "Schachbrettregel")

3.6 Beispiel. Für $n=2$ und $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ ist $A^\sharp = \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$.
Für $n=3$ und $A=(a_{ij})$ ist

$$A^\sharp = \begin{pmatrix} a_{22}a_{33}-a_{23}a_{32} & -a_{12}a_{33}+a_{13}a_{32} & a_{12}a_{23}-a_{13}a_{22} \\ -a_{21}a_{33}+a_{23}a_{31} & a_{11}a_{33}-a_{13}a_{31} & -a_{11}a_{23}+a_{13}a_{21} \\ a_{21}a_{32}-a_{22}a_{31} & -a_{11}a_{32}+a_{12}a_{31} & a_{11}a_{22}-a_{12}a_{21} \end{pmatrix}.$$

Die wesentliche Eigenschaft der komplementären Matrix ist:

3.7 Satz. *Sei $A \in \mathrm{M}_n(R)$, sei $A^\sharp$ die komplementäre Matrix zu A. Dann ist*

$$A \cdot A^\sharp \;=\; A^\sharp \cdot A \;=\; \det(A)\cdot I_n.$$

Beweis. Sei $A=(a_{ij})$, sei $v_j = \sum_{k=1}^n a_{kj}e_k$ die j-te Spalte von A ($j = 1,\dots,n$), also $A = M(v_1,\dots,v_n)$, und sei $A^\sharp = (a^\sharp_{ij})$. Nach Definition ist $a^\sharp_{ik} = \det(A_{ki})$ mit

$$A_{ki} \;=\; M\big(v_1 - a_{k1}e_k,\; \dots,\; e_k,\; \dots,\; v_n - a_{kn}e_k\big);$$

beachte, hier steht e_k an der i-ten Stelle. Der (i,j)-Koeffizient von $A^\sharp A$ ist also gleich

$$\begin{aligned}
\sum_{k=1}^n a^\sharp_{ik} a_{kj} &= \sum_{k=1}^n \det(A_{ki}) \cdot a_{kj} \\
&= \sum_k a_{kj} \cdot \det M\big(v_1 - a_{k1}e_k,\ \ldots,\ e_k, \ldots,\ v_n - a_{kn}e_k\big) \\
&= \sum_k a_{kj} \cdot \det M\big(v_1,\ \ldots,\ v_{i-1},\ e_k,\ v_{i+1},\ \ldots,\ v_n\big) \\
&= \det M\Big(v_1,\ \ldots,\ v_{i-1},\ \sum_{k=1}^n a_{kj}e_k,\ v_{i+1},\ \ldots,\ v_n\Big) \\
&= \det M\big(v_1,\ \ldots,\ v_{i-1},\ v_j,\ v_{i+1},\ \ldots,\ v_n\big) \\
&= \delta_{ij} \cdot \det(A).
\end{aligned}$$

(Beim dritten Gleichheitszeichen wurde (D1) und (D2) benutzt, beim vierten (D1), und beim letzten (D2).) Also ist $A^\sharp \cdot A = \det(A) \cdot I_n$. Um die andere Gleichung zu zeigen, beachte zunächst $(A^t)_{ij} = (A_{ji})^t$ für alle i, j. Daraus folgt $(A^\sharp)^t = (A^t)^\sharp$, und somit

$$(A \cdot A^\sharp)^t = (A^\sharp)^t \cdot A^t = (A^t)^\sharp \cdot A^t = \det(A^t) \cdot I_n = \det(A) \cdot I_n,$$

wobei das dritte Gleichheitszeichen nach dem ersten Beweisteil gilt. □

Aus Satz 3.7 folgt sofort:

3.8 Korollar. *Sei $A \in \mathrm{M}_n(R)$. Ist $\det(A)$ eine Einheit von R, so ist A invertierbar, und*

$$A^{-1} \;=\; \det(A)^{-1} \cdot A^\sharp. \qquad \square$$

Das ist eine *geschlossene Formel* für die Inverse einer invertierbaren Matrix. Damit ist Satz 2.16 jetzt auch für einen beliebigen kommutativen Ring R vollständig bewiesen. Hier sind weitere Folgerungen aus Satz 3.7:

3.9 Korollar. (Entwicklungssatz von Laplace[3]) *Sei R ein kommutativer Ring, sei $A = (a_{ij}) \in \mathrm{M}_n(R)$. Für jedes $j = 1, \ldots, n$ ist*

$$\det(A) \;=\; \sum_{i=1}^n (-1)^{i+j}\, a_{ij}\, \det(A'_{ij})$$

("Entwicklung von $\det(A)$ nach der j-ten Spalte"), für jedes $i = 1, \ldots, n$ ist

$$\det(A) \;=\; \sum_{j=1}^n (-1)^{i+j}\, a_{ij}\, \det(A'_{ij})$$

("Entwicklung von $\det(A)$ nach der i-ten Zeile").

Beweis. Sei j fest. Für den (j,j)-Koeffizient von $A^\sharp A$ gilt wegen $A^\sharp A = \det(A) I_n$:

$$\det(A) \;=\; \sum_i a^\sharp_{ji} a_{ij} = \sum_i a_{ij} \cdot (-1)^{i+j} \det(A'_{ij}),$$

[3]Pierre-Simon Laplace (1749–1827)

siehe 3.5. Die zweite Behauptung beweist man analog, oder folgert sie aus der ersten durch Transposition. □

3.10 Beispiel. Wieder hat man Vorzeichen nach der Schachbrettregel. Zum Beispiel gibt Entwicklung einer 3×3-Determinante nach der 2. Zeile:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{23} \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}.$$

Die Formeln von Laplace führen die Berechnung einer $n \times n$-Determinante auf die Berechnung von $(n-1) \times (n-1)$-Determinanten zurück, können also auch als induktive Beschreibung der Determinante angesehen werden. Jedoch werden im Laplaceschen Entwicklungssatz die $n!$ Summanden der Leibnizformel nur neu gruppiert; er bietet also eigentlich nichts Neues. Trotzdem ist er manchmal nützlich, wenn man nämlich die Existenz von Zeilen oder Spalten mit vielen Nullen ausnützen will:

3.11 Beispiel. Sei $A = \begin{pmatrix} 1 & 5 & 0 & 2 \\ -4 & 0 & 1 & 1 \\ 1 & 3 & 2 & 2 \\ 2 & 0 & -1 & 4 \end{pmatrix}$ (vgl. Beispiel 2.12). Entwicklung nach der 2. Spalte (dort gibt es zwei Nullen!) gibt

$$\det(A) = -5 \begin{vmatrix} -4 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & -1 & 4 \end{vmatrix} - 3 \begin{vmatrix} 1 & 0 & 2 \\ -4 & 1 & 1 \\ 2 & -1 & 4 \end{vmatrix} = -5 \cdot (-45) - 3 \cdot 9 = 198.$$

3.12 Satz. (Cramersche Regel[4]) *Sei K ein Körper, sei $A = M(v_1, \ldots, v_n)$ invertierbar (mit $v_1, \ldots, v_n \in K^n$), und sei $u \in K^n$. Dann ist die eindeutige Lösung $x \in K^n$ des linearen Gleichungssystems $Ax = u$ gegeben durch*

$$x_j = \frac{\det M(v_1, \ldots, v_{j-1}, u, v_{j+1}, \ldots, v_n)}{\det M(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_n)}$$

(der Nenner ist $\det(A)$*!), also*

$$x_j = \frac{1}{\det(A)} \cdot \begin{vmatrix} a_{11} & \cdots & a_{1,j-1} & u_1 & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & u_n & a_{n,j+1} & \cdots & a_{nn} \end{vmatrix} \quad (j = 1, \ldots, n).$$

Beweis. Sei $A = (a_{ij})$. Es genügt nachzurechnen, daß der angegebene Vektor $(x_1, \ldots, x_n)$ eine Lösung von $Ax = u$ ist. Dazu sei $i \in \{1, \ldots, n\}$ fixiert. Durch

[4]Gabriel Cramer (1704–1752)

Entwickeln der Zähler-Determinante nach der j-ten Spalte (für jedes j) erhält man

$$\begin{aligned}
\sum_{j=1}^{n} a_{ij}x_j &= \frac{1}{\det(A)} \sum_{j=1}^{n} a_{ij} \sum_{k=1}^{n} (-1)^{j+k} u_k \det(A'_{kj}) \\
&= \frac{1}{\det(A)} \sum_{k=1}^{n} u_k \sum_{j=1}^{n} a_{ij} (-1)^{j+k} \det(A'_{kj}) \\
&= \frac{1}{\det(A)} \sum_{k=1}^{n} u_k \sum_{j=1}^{n} a_{ij} a^{\sharp}_{jk} \\
&= \frac{1}{\det(A)} \sum_{k=1}^{n} u_k \cdot \delta_{ik} \det(A) \\
&= u_i.
\end{aligned}$$

□

Ebenso wie die Formel für die inverse Matrix (3.8) ist die Cramersche Regel für die praktische Rechnung ungeeignet. Die Bedeutung dieser Aussagen ist vielmehr eine theoretische. Aus 3.7 etwa folgt, daß die Abbildung $A \mapsto A^{-1}$ (von $\mathrm{GL}_n(\mathbb{R})$ in sich) stetig und sogar beliebig oft differenzierbar ist. Aus 3.12 folgt, daß die Lösung eines eindeutig lösbaren linearen Gleichungssystems stetig (sogar differenzierbar) von den Koeffizienten des Systems abhängt.

3.13 Definition. Sei R kommutativer Ring, sei $A \in \mathrm{M}_{m\times n}(R)$, und sei $1 \leq r \leq \min\{m,n\}$. Jede Determinante

$$\begin{vmatrix} a_{i_1 j_1} & a_{i_1 j_2} & \cdots & a_{i_1 j_r} \\ a_{i_2 j_1} & a_{i_2 j_2} & \cdots & a_{i_2 j_r} \\ \vdots & \vdots & & \vdots \\ a_{i_r j_1} & a_{i_r j_2} & \cdots & a_{i_r j_r} \end{vmatrix} \qquad (*)$$

mit $1 \leq i_1 < \cdots < i_r \leq m$ und $1 \leq j_1 < \cdots < j_r \leq n$ heißt ein *r-(reihiger) Minor* (oder eine *r-reihige Unterdeterminante*) von A.

3.14 Beispiel. Sei $A = (a_{ij}) \in \mathrm{M}_{m\times n}(R)$. Die 1-Minoren von A sind die a_{ij}; es gibt genau mn davon. Die 2-Minoren von A sind die $\begin{vmatrix} a_{ik} & a_{il} \\ a_{jk} & a_{jl} \end{vmatrix}$ mit $1 \leq i < j \leq m$ und $1 \leq k < l \leq n$. Es gibt also $\binom{m}{2}\binom{n}{2}$ davon. Allgemein ist die Anzahl der r-Minoren von A gleich $\binom{m}{r}\binom{n}{r}$.

3.15 Satz. *Sei K ein Körper, und sei $A \in \mathrm{M}_{m\times n}(K)$, $A \neq 0$. Dann ist* $\mathrm{rk}(A)$ *gleich dem maximalen r, für das ein r-reihiger Minor $\neq 0$ von A existiert.*

Beweis. Sei $z_i(A)$ die i-te Zeile und $s_j(A)$ die j-te Spalte von A. Zunächst sei etwa der durch $(*)$ gegebene r-Minor von A von Null verschieden. Die Zeilen der Matrix in $(*)$ sind also linear unabhängig (2.17). Daher sind auch die entsprechenden Zeilen $z_{i_1}(A), \ldots, z_{i_r}(A)$ von A linear unabhängig. Also ist $\mathrm{rk}(A) \geq r$.

Umgekehrt sei $\mathrm{rk}(A) \geq r$. Dann gibt es $1 \leq i_1 < \cdots < i_r \leq m$, so daß $z_{i_1}(A), \ldots, z_{i_r}(A)$ linear unabhängig sind. Sei A' die aus diesen Zeilen bestehende $r \times n$-Matrix. Wegen $\mathrm{rk}(A') = r$ gibt es $1 \leq j_1 < \cdots < j_r \leq n$, so daß die Spalten

$s_{j_1}(A'), \dots, s_{j_r}(A')$ linear unabhängig sind. Die $r \times r$-Matrix A'' mit diesen Spalten hat also $\det(A'') \neq 0$ (2.17), und $\det(A'')$ ist der r-Minor von A zu den Zeilen $i_1, \dots, i_r$ und den Spalten $j_1, \dots, j_r$. □

Anders formuliert besagt Satz 3.15: Genau dann ist $\mathrm{rk}(A) \leq r$, wenn alle $(r+1)$-Minoren von A verschwinden.

4. Ähnlichkeit von Matrizen, Determinante und Spur von Endomorphismen, Orientierung

Sei K stets ein Körper. Auch in diesem Abschnitt sind alle Vektorräume endlichdimensional.

4.1 Satz. *Sei V ein K-Vektorraum mit $\dim(V) = n < \infty$, sei $f \in \mathrm{End}(V)$. Sei $\mathcal{B}$ eine Basis von V und $A = \mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) \in \mathrm{M}_n(K)$. Für $B \in \mathrm{M}_n(K)$ sind äquivalent:*

(i) *Es gibt eine Basis $\mathcal{C}$ von V mit $B = \mathrm{M}_{\mathcal{C}}^{\mathcal{C}}(f)$;*
(ii) *es gibt $S \in \mathrm{GL}_n(K)$ mit $B = SAS^{-1}$.*

Beweis. (i) ⇒ (ii) gilt wegen $\mathrm{M}_{\mathcal{C}}^{\mathcal{C}}(f) = T_{\mathcal{C}}^{\mathcal{B}} \cdot \mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) \cdot (T_{\mathcal{C}}^{\mathcal{B}})^{-1}$ (siehe Korollar III.5.11). (ii) ⇒ (i): Definiere die Basis $\mathcal{C}$ von V durch $T_{\mathcal{B}}^{\mathcal{C}} = S^{-1}$, das heißt: Ist $S^{-1} = (c_{ij})$ und $\mathcal{B} = (v_1, \dots, v_n)$, so sei $\mathcal{C} = (w_1, \dots, w_n)$ mit $w_j := \sum_i c_{ij} v_i$ $(j = 1, \dots, n)$. Dann ist $\mathcal{C}$ eine Basis von V und

$$B \;=\; SAS^{-1} \;=\; T_{\mathcal{C}}^{\mathcal{B}} \cdot \mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) \cdot T_{\mathcal{B}}^{\mathcal{C}} \;=\; \mathrm{M}_{\mathcal{C}}^{\mathcal{C}}(f).$$

□

4.2 Definition. Zwei (quadratische) Matrizen $A, B \in \mathrm{M}_n(K)$ heißen *ähnlich*, i. Z. $A \approx B$, wenn es $S \in \mathrm{GL}_n(K)$ gibt mit $B = SAS^{-1}$.

4.3 Satz. *Seien $A, B \in \mathrm{M}_n(K)$.*

(a) *Ähnlichkeit ($\approx$) ist eine Äquivalenzrelation auf $\mathrm{M}_n(K)$.*
(b) *$A \approx B \;\Rightarrow\; A \sim B$ (Ähnlichkeit impliziert Äquivalenz).*
(c) *$A \approx B \;\Rightarrow\; \det(A) = \det(B)$ (ähnliche Matrizen haben dieselbe Determinante).*

Beweis. (a) ist klar, (b) ist klar nach Definition der Äquivalenz (siehe III.6.12), und (c) gilt wegen $\det(SAS^{-1}) = \det(S)\det(A)\det(S)^{-1} = \det(A)$ für $S \in \mathrm{GL}_n(K)$. □

4.4 Bemerkung. Ist $\dim(V) = n$ und $f \in \mathrm{End}(V)$, so ist die Menge

$$\big\{\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) \colon\ \mathcal{B} \text{ Basis von } V\big\}$$

nach Satz 4.1 eine volle Ähnlichkeitsklasse in $\mathrm{M}_n(K)$. Wir können dem Endomorphismus f nach 4.3(c) eine wohlbestimmte Determinante zuordnen:

4.5 Definition. Sei V ein K-Vektorraum mit $\dim(V) < \infty$. Die *Determinante* von $f \in \mathrm{End}(V)$ ist definiert durch

$$\det(f) \;:=\; \det\Big(\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)\Big),$$

falls $V \neq \{0\}$ und $\mathcal{B}$ eine beliebige Basis von V ist. Für $V = \{0\}$ setzt man $\det(f) := 1$.

4.6 Korollar. *Sei* $\dim(V) < \infty$*, seien* $f, g \in \mathrm{End}(V)$.

(a) f *bijektiv* $\Leftrightarrow$ $\det(f) \neq 0$.
(b) $\det(f \circ g) = \det(f) \cdot \det(g)$. □

Beweis. Das ist klar: (a) folgt aus Korollar III.5.6 und Satz IV.2.17, (b) aus Satz III.5.5 und Theorem IV.2.10. □

Neben der Determinante gibt es eine andere wichtige Invariante einer quadratischen Matrix, die nur von ihrer Ähnlichkeitsklasse abhängt:

4.7 Definition. Für $A = (a_{ij}) \in \mathrm{M}_n(K)$ ist die *Spur* (engl.: trace) von A definiert als $\mathrm{tr}(A) := \sum_{i=1}^{n} a_{ii}$.

4.8 Satz.

(a) *Die Abbildung* $\mathrm{tr}\colon \mathrm{M}_n(K) \to K$ *ist* K*-linear, d. h.* $\mathrm{tr}(aA+bB) = a\,\mathrm{tr}(A) + b\,\mathrm{tr}(B)$ *für* $a, b \in K$ *und* $A, B \in \mathrm{M}_n(K)$.
(b) $\mathrm{tr}(A^t) = \mathrm{tr}(A)$ *für* $A \in \mathrm{M}_n(K)$.
(c) *Für* $A \in \mathrm{M}_{m\times n}(K)$ *und* $B \in \mathrm{M}_{n\times m}(K)$ *gilt* $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

Beweis. (a) und (b) sind sofort klar. Beweis von (c): Mit $A = (a_{ij})$ und $B = (b_{jk})$ ist

$$\mathrm{tr}(AB) \;=\; \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji} \;=\; \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij} \;=\; \mathrm{tr}(BA).$$

□

4.9 Korollar. *Für* $A, B \in \mathrm{M}_n(K)$ *mit* $A \approx B$ *gilt* $\mathrm{tr}(A) = \mathrm{tr}(B)$.

Beweis. Sei $B = SAS^{-1}$ mit $S \in \mathrm{GL}_n(K)$, dann ist $\mathrm{tr}(B) = \mathrm{tr}(SA \cdot S^{-1}) = \mathrm{tr}(S^{-1} \cdot SA) = \mathrm{tr}(A)$ nach 4.8(c). □

Deshalb können wir, analog zur Determinante, die Spur auch für Endomorphismen definieren:

4.10 Definition. Sei V ein K-Vektorraum mit $\dim(V) < \infty$. Für $f \in \mathrm{End}(V)$ ist die *Spur* von f definiert durch

$$\mathrm{tr}(f) \;:=\; \mathrm{tr}\Big(\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)\Big),$$

wobei $\mathcal{B}$ eine beliebige Basis von V ist (Fall $V \neq \{0\}$). Für $V = \{0\}$ setzt man $\mathrm{tr}(f) := 0$. Die Spur von f ist wohldefiniert nach Satz 4.1 und Korollar 4.9.

4.11 Bemerkungen.

1. Es gelten die zu Satz 4.8 analogen Regeln: Für $f, g \in \mathrm{End}(V)$ und $a, b \in K$ ist $\mathrm{tr}(af + bg) = a\,\mathrm{tr}(f) + b\,\mathrm{tr}(g)$ und $\mathrm{tr}(g \circ f) = \mathrm{tr}(f \circ g)$.

2. Ein Endomorphismus $p \in \mathrm{End}(V)$ heißt eine *Projektion*, wenn $p^2 = p$ gilt (mit $p^2 = p \circ p$). Für jede Projektion p ist $V = \ker(p) \oplus \mathrm{im}(p)$ (Aufgabe 26) und

$p(v) = v$ für alle $v \in \mathrm{im}(p)$. Bezüglich einer geeigneten Basis $\mathcal{B}$ von V ist also $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(p) = \mathrm{diag}(0, \ldots, 0, 1, \ldots, 1)$. Somit ist $\mathrm{tr}(p) = \mathrm{rk}(p)$.

3. Die Frage, ob zwei Matrizen äquivalent sind, ist leicht zu entscheiden: Man berechnet ihre Ränge. Die Frage, ob zwei quadratische Matrizen A, B ähnlich sind, ist viel schwieriger. Für $A \approx B$ ist notwendig, daß Rang, Spur, Determinante von A und B gleich sind. Diese Bedingungen sind aber nicht hinreichend für $A \approx B$, wie man an $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ und $B = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ sieht.

Sei jetzt $K = \mathbb{R}$.

4.12 Definition. Sei V ein $\mathbb{R}$-Vektorraum, sei $n = \dim(V) < \infty$.

(a) Ein Automorphismus $f \in \mathrm{GL}(V)$ heißt *orientierungstreu* (bzw. *orientierungsumkehrend*), wenn $\det(f) > 0$ (bzw. $\det(f) < 0$) ist.

(b) Zwei Basen $\mathcal{B} = (v_1, \ldots, v_n)$, $\mathcal{C} = (w_1, \ldots, w_n)$ von V heißen *gleich orientiert*, wenn der Automorphismus $f \in \mathrm{GL}(V)$ mit $f(v_i) = w_i$ $(i = 1, \ldots, n)$ orientierungstreu ist. Andernfalls heißen $\mathcal{B}$, $\mathcal{C}$ *entgegengesetzt orientiert*.

4.13 Bemerkungen.

1. Zwei Basen $\mathcal{B} = (v_1, \ldots, v_n)$, $\mathcal{C} = (w_1, \ldots, w_n)$ von V sind genau dann gleich orientiert, wenn $\det(T_{\mathcal{C}}^{\mathcal{B}}) > 0$ ist. Denn sei $f \in \mathrm{GL}(V)$ mit $f(v_i) = w_i$ $(i = 1, \ldots, n)$ wie in 4.12(b). Dann ist

$$I_n \;=\; \mathrm{M}_{\mathcal{C}}^{\mathcal{B}}(f) \;=\; \mathrm{M}_{\mathcal{C}}^{\mathcal{C}}(f) \cdot T_{\mathcal{C}}^{\mathcal{B}}$$

(erste Gleichheit nach Definition von f, zweite Gleichheit nach Satz III.5.5), woraus die Behauptung folgt.

2. Ist $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V, und ist $\mathcal{B}' = (-v_1, v_2, \ldots, v_n)$, $\mathcal{B}'' = (v_2, v_1, v_3, \ldots, v_n)$ (für $n \geq 2$), so sind $\mathcal{B}$, $\mathcal{B}'$ entgegengesetzt orientiert, ebenso $\mathcal{B}$, $\mathcal{B}''$. Dagegen sind $\mathcal{B}'$, $\mathcal{B}''$ gleich orientiert.

3. Sei $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ die Drehung um $(0,0)$ um den Winkel ϑ. Dann ist f orientierungstreu, denn wegen

$$\mathrm{M}_{\mathcal{K}_2}^{\mathcal{K}_2}(f) \;=\; \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}$$

(III.3.8) ist $\det(f) = 1$. Sind dagegen $v_1, v_2 \in \mathbb{R}^2$ linear unabhängig und ist f die Spiegelung an der Geraden $\mathbb{R}v_1$ längs der Geraden $\mathbb{R}v_2$, so ist $\det(f) = -1$, also f orientierungsumkehrend. Denn bezüglich der Basis $\mathcal{B} = (v_1, v_2)$ ist

$$\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) \;=\; \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

4.14 Satz und Definition. *Sei V ein $\mathbb{R}$-Vektorraum mit $\dim(V) < \infty$.*

(a) *Gleichorientiertheit ist eine Äquivalenzrelation auf der Menge aller Basen von V.*

(b) *Eine* Orientierung *von V ist eine Äquivalenzklasse von Basen von V bezüglich der Äquivalenzrelation (a).*

(c) *Für $V \neq \{0\}$ hat V genau zwei verschiedene Orientierungen.*

Beweis. (a) folgt aus Bemerkung 4.13.1 (transitiv wegen $T_{\mathcal{B}''}^{\mathcal{B}} = T_{\mathcal{B}''}^{\mathcal{B}'} \cdot T_{\mathcal{B}'}^{\mathcal{B}}$). Nach 4.13.2 hat $V \neq \{0\}$ genau zwei Orientierungen. □

Für einen $\mathbb{R}$-Vektorraum V mit $\dim(V) < \infty$ ist im allgemeinen keine der beiden Orientierungen vor der anderen ausgezeichnet. Eine Orientierung von V legt man fest, indem man eine zugehörige Basis angibt. Für $V = \mathbb{R}^n$ hat man die kanonische Orientierung, gegeben durch die kanonische Basis $(e_1, \ldots, e_n)$.

KAPITEL V

Strukturtheorie von Endomorphismen

1. Eigenwerte und Eigenvektoren

Stets sei K ein Körper und V ein K-Vektorraum. A priori ist $\dim(V)$ beliebig, später setzen wir fast immer $\dim(V) < \infty$ voraus.

1.1 Definition. Sei $f \in \mathrm{End}(V)$. Ein $\lambda \in K$ heißt ein *Eigenwert* von f, wenn es $v \in V$ gibt mit $v \neq 0$ und $f(v) = \lambda v$. Jedes solche $v \neq 0$ heißt ein *Eigenvektor* von f zum Eigenwert λ. Für jedes $\lambda \in K$ ist die Menge

$$\mathrm{Eig}(f, \lambda) \;:=\; \{v \in V \colon f(v) = \lambda v\} \;=\; \ker(\lambda \,\mathrm{id}_V - f)$$

ein Untervektorraum von V und heißt der *Eigenraum* von f zum Parameter λ.

Die Eigenwerte von f sind also die $\lambda \in K$ mit $\mathrm{Eig}(f, \lambda) \neq \{0\}$. Für $\dim(V) < \infty$ läßt sich das auch anders sagen:

1.2 Satz. *Sei* $\dim(V) < \infty$*, sei* $f \in \mathrm{End}(V)$*. Genau dann ist* $\lambda \in K$ *ein Eigenwert von* f*, wenn* $\det(\lambda \,\mathrm{id}_V - f) = 0$ *ist.*

Beweis. λ ist Eigenwert von f genau dann, wenn $\ker(\lambda \,\mathrm{id}_V - f) \neq \{0\}$ ist. Nach IV.4.6(a) ist das äquivalent zu $\det(\lambda \,\mathrm{id}_V - f) = 0$. □

1.3 Beispiele.

1. Ist $p \in \mathrm{End}(V)$ eine Projektion, d. h. gilt $p^2 = p$ (siehe IV.4.11), und ist $p \neq 0$, $p \neq \mathrm{id}$, so hat p genau die beiden Eigenwerte 0 und 1. Begründe dies im Fall $\dim(V) < \infty$ (siehe Aufgabe 42(a) für $\dim(V) = \infty$): Bezüglich einer geeigneten Basis $\mathcal{B}$ von V ist $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(p) = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$ (*loc. cit.*), also ist

$$\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}\big(\lambda \,\mathrm{id} - p\big) \;=\; \mathrm{diag}(\lambda - 1, \ldots, \lambda - 1, \lambda, \ldots, \lambda).$$

Diese Matrix ist invertierbar für $\lambda \notin \{0, 1\}$, und wegen $f \neq 0$, $f \neq \mathrm{id}$ kommen 0 und 1 beide auf der Diagonale vor.

2. Sei f die Drehung der Ebene $\mathbb{R}^2$ um den Ursprung mit Drehwinkel $\vartheta \in \mathbb{R}$. Bezüglich der kanonischen Basis wird f beschrieben durch

$$A \;=\; \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}$$

(siehe III.3.8). Für $\vartheta \notin \{k\pi \colon k \in \mathbb{Z}\}$ hat f keine Eigenwerte. Denn für solches ϑ ist $\sin(\vartheta) \neq 0$, also gilt

$$\det(\lambda \,\mathrm{id} - f) \;=\; (\lambda - \cos(\vartheta))^2 + \sin^2(\vartheta) \;>\; 0$$

für alle $\lambda \in \mathbb{R}$. (Welche Eigenwerte und Eigenvektoren hat f für $\vartheta = 0$, $\vartheta = \pi$?)

3. Sei V der $\mathbb{R}$-Vektorraum der beliebig oft differenzierbaren Funktionen $\mathbb{R} \to \mathbb{R}$ (es ist $\dim(V) = \infty$). Die Ableitung $D\colon V \to V$, $D(f) = f'$ ist ein linearer Endomorphismus von V. Jedes $\lambda \in \mathbb{R}$ ist ein Eigenwert von D. Denn für $f(x) = e^{\lambda x}$ ist $f'(x) = \lambda e^{\lambda x}$, also $D(f) = \lambda f$. (Man kann zeigen, daß $\dim \operatorname{Eig}(D, \lambda) = 1$ ist.)

1.4 Satz. *Sei $f \in \operatorname{End}(V)$, seien $\lambda_1, \dots, \lambda_r \in K$ paarweise verschieden. Dann ist die Summe $\sum_{i=1}^{r} \operatorname{Eig}(f, \lambda_i)$ (von Untervektorräumen von V) direkt.*

Beweis. Zu zeigen ist (siehe Satz III.3.15): Sind $v_i \in \operatorname{Eig}(f, \lambda_i)$ (für $i = 1, \dots, r$) mit $v_1 + \cdots + v_r = 0$, so ist $v_i = 0$ für alle i. Angenommen falsch, seien $\lambda_1, \dots, \lambda_r$ und $v_1, \dots, v_r$ ein Gegenbeispiel, und sei dabei $r \geq 1$ minimal. Aus der Minimalität folgt $v_i \neq 0$ für alle i, und aus

$$v_1 + \cdots + v_r \;=\; 0$$

folgt durch Anwenden von f

$$\lambda_1 v_1 + \cdots + \lambda_r v_r \;=\; 0.$$

Ziehe von der zweiten Gleichung das λ_1-fache der ersten Gleichung ab, das gibt

$$(\lambda_2 - \lambda_1)v_2 + \cdots + (\lambda_r - \lambda_1)v_r \;=\; 0.$$

Die Vektoren $w_i := (\lambda_i - \lambda_1)v_i$ erfüllen $w_i \in \operatorname{Eig}(f, \lambda_i)$ und $w_i \neq 0$ $(i = 2, \dots, r)$ sowie $w_2 + \cdots + w_r = 0$, Widerspruch zur minimalen Wahl von r. □

Daraus folgt sofort:

1.5 Korollar. *Für $\dim(V) = n < \infty$ hat $f \in \operatorname{End}(V)$ höchstens n verschiedene Eigenwerte.*

Beweis. Sind $\lambda_1, \dots, \lambda_r \in K$ verschiedene Eigenwerte von f, so folgt

$$n \;=\; \dim(V) \;\geq\; \dim\Big(\sum_{i=1}^{r} \operatorname{Eig}(f, \lambda_i)\Big) \underset{1.4}{=} \sum_{i=1}^{r} \dim \operatorname{Eig}(f, \lambda_i) \;\geq\; r$$

wegen $\dim \operatorname{Eig}(f, \lambda_i) \geq 1$ für $i = 1, \dots, r$. □

1.6 Satz und Definition. *Sei $\dim(V) < \infty$. Für $f \in \operatorname{End}(V)$ sind äquivalent:*

(i) *V hat eine Basis aus Eigenvektoren von f,*
(ii) *es gibt eine Basis $\mathcal{B}$ von V, so daß $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)$ eine Diagonalmatrix ist.*

Gelten (i) und (ii), so heißt f diagonalisierbar. □

Offensichtlich gilt:

1.7 Korollar. *Sei $\dim(V) = n < \infty$. Genau dann ist f diagonalisierbar, wenn*

$$\sum_{\lambda \in K} \operatorname{Eig}(f, \lambda) \;=\; V$$

ist. Hat f n verschiedene Eigenwerte, so tritt dieser Fall ein.

Beweis. Die Äquivalenz ist klar, und der Zusatz folgt mit dem Argument aus dem Beweis von Korollar 1.5. □

1.8 Beispiel. Sei $\dim(V) < \infty$. Jede Projektion $p \in \mathrm{End}(V)$ (d.h. $p^2 = p$) ist diagonalisierbar (Beispiel 1.3.1). Jedes $f \in \mathrm{End}(V)$ mit $f^2 = \mathrm{id}_V$ ist diagonalisierbar, falls $\mathrm{char}(K) \neq 2$ ist. (Übung, siehe Aufgabe 42) Eine Drehung in der Ebene um einen Winkel $\alpha \notin \mathbb{Z}\pi$ ist dagegen nicht diagonalisierbar (Beispiel 1.3.2).

1.9 Bemerkung. Diese Konzepte machen auch Sinn für quadratische Matrizen, und sind für $A \in \mathrm{M}_n(K)$ definiert als die entsprechenden Begriffe für den Endomorphismus F_A von K^n. Für $\lambda \in K$ heißt also

$$\mathrm{Eig}(A, \lambda) \;=\; \{v \in K^n \colon Av = \lambda v\}$$

der Eigenraum von A zum Parameter λ. Ist $\mathrm{Eig}(A, \lambda) \neq \{0\}$, so heißt λ Eigenwert von A, bzw. die Vektoren in $\mathrm{Eig}(A, \lambda) \smallsetminus \{0\}$ heißen die Eigenvektoren von A zum Eigenwert λ. Die Matrix A heißt diagonalisierbar, wenn K^n eine Basis aus Eigenvektoren von A hat.

1.10 Satz. *Für $A \in \mathrm{M}_n(K)$ sind äquivalent:*

(i) *A ist diagonalisierbar;*
(ii) *A ist ähnlich zu einer Diagonalmatrix, d.h. es gibt $S \in \mathrm{GL}_n(K)$ mit $S^{-1}AS$ Diagonalmatrix.*

Explizit: Ist $(v_1, \ldots, v_n)$ eine Basis von K^n mit $Av_i = \lambda_i v_i$ ($\lambda_i \in K$, $i = 1, \ldots, n$), so ist $S^{-1}AS = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ mit $S = M(v_1, \ldots, v_n) \in \mathrm{GL}_n(K)$.

Beweis. Zunächst die letzte Aussage: Für $i = 1, \ldots, n$ ist $ASe_i = Av_i = \lambda_i v_i = \lambda_i Se_i$, also folgt $S^{-1}ASe_i = \lambda_i e_i$ für alle i, also $S^{-1}AS = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

(i) $\Rightarrow$ (ii): Wegen A diagonalisierbar gibt es v_i und λ_i wie im Zusatz. Also folgt (ii).

(ii) $\Rightarrow$ (i): Ist $S \in \mathrm{GL}_n(K)$ mit $S^{-1}AS = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, und ist $S = M(v_1, \ldots, v_n)$, so ist $(v_1, \ldots, v_n)$ eine Basis von K^n (III.6.11) und $Av_i = \lambda_i v_i$ ($i = 1, \ldots, n$), also A diagonalisierbar. □

2. Das charakteristische Polynom

Alle Vektorräume seien jetzt endlich-dimensional.

2.1 Definition. Für $A \in \mathrm{M}_n(K)$ heißt das Polynom

$$\chi_A \;:=\; \det(tI_n - A) \;\in K[t]$$

das *charakteristische Polynom* von A.

Die $n \times n$-Matrix $tI_n - A$ hat Koeffizienten im Polynomring $K[t]$. Wir verwenden hier also die Determinante über dem kommutativen Ring $K[t]$.

2.2 Beispiele.

1. Für $n = 2$ und $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ ist

$$\chi_A \;=\; \begin{vmatrix} t-a & -b \\ -c & t-d \end{vmatrix} \;=\; t^2 - (a+d)t + (ad - bc) \;=\; t^2 - \mathrm{tr}(A) \cdot t + \det(A).$$

2. Für eine Dreiecksmatrix

$$A \;=\; \begin{pmatrix} a_1 & & * \\ & \ddots & \\ 0 & & a_n \end{pmatrix}$$

ist $\chi_A = (t-a_1)\cdots(t-a_n)$ (siehe IV.2.8.3).

Nach Satz 1.2 sind die Eigenwerte einer Matrix genau die Nullstellen ihres charakteristischen Polynoms:

2.3 Korollar. *Sei $A \in \mathrm{M}_n(K)$. Genau dann ist $\lambda \in K$ ein Eigenwert von A, wenn $\chi_A(\lambda) = 0$ ist.* □

2.4 Satz. *Für das charakteristische Polynom von $A \in \mathrm{M}_n(K)$ gilt*

$$\chi_A \;=\; t^n + a_1\, t^{n-1} + \cdots + a_{n-1}\, t + a_n$$

mit $a_1, \ldots, a_n \in K$ und $a_1 = -\operatorname{tr}(A)$, $a_n = (-1)^n \det(A)$.

Beweis. Sei $A = (a_{ij})$. Es ist $\chi_A = \prod_{j=1}^n (t - a_{jj}) + g$ mit

$$g \;=\; \sum_{\substack{\sigma \in S_n \\ \sigma \neq \mathrm{id}}} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} \bigl(\delta_{\sigma(j),j} t - a_{\sigma(j),j}\bigr).$$

Für das Polynom $g \in K[t]$ gilt dabei $\deg(g) \leq n-2$, denn für $\sigma \neq \mathrm{id}$ ist $|\mathrm{Fix}(\sigma)| \leq n-2$. Ausmultiplizieren gibt

$$\chi_A \;=\; t^n - \Bigl(\sum_{i=1}^{n} a_{ii}\Bigr)\, t^{n-1} + \tilde{g}$$

mit $\deg(\tilde{g}) \leq n-2$. Die letzte Behauptung gilt wegen $a_n = \chi_A(0) = \det(-A) = (-1)^n \det(A)$. □

Ähnliche Matrizen haben dasselbe charakteristische Polynom:

2.5 Satz. *Für A, $B \in \mathrm{M}_n(K)$ mit $A \approx B$ gilt $\chi_A = \chi_B$.*

Beweis. Ist $B = SAS^{-1}$ mit $S \in \mathrm{GL}_n(K)$, so folgt

$$\begin{aligned} \chi_B &= \det(tI_n - SAS^{-1}) \;=\; \det\bigl(S\,(tI_n - A)\,S^{-1}\bigr) \\ &= \det(S) \cdot \chi_A \cdot \det(S)^{-1} \;=\; \chi_A. \end{aligned}$$

□

2.6 Bemerkung. Wir haben damit eine neue für $A \approx B$ notwendige Bedingung gefunden, nämlich $\chi_A = \chi_B$. Wegen Satz 2.4 enthält diese die schon früher gefundenen Bedingungen, daß Spur und Determinante von A und B gleich sind (Bemerkung IV.4.11.3).

Wegen Satz 2.5 können wir für jeden Endomorphismus ein charakteristisches Polynom definieren:

2.7 Definition. Sei V ein Vektorraum mit $\dim(V) < \infty$, und sei $f \in \mathrm{End}(V)$. Das *charakteristische Polynom* von f ist definiert als

$$\chi_f \;:=\; \chi_A$$

falls $V \neq \{0\}$, für $A = \mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)$ und $\mathcal{B}$ eine beliebige Basis von V. Für $V = \{0\}$ setzt man $\chi_f := 1$.

Es ist also $\chi_f \in K[t]$ ein Polynom mit $\deg(\chi_f) = \dim(V)$, und für χ_f gelten die zu 2.3 und 2.4 analogen Aussagen. Hier ist eine erste Anwendung für reelle Vektorräume:

2.8 Satz. *Sei V ein $\mathbb{R}$-Vektorraum von ungerader (endlicher) Dimension. Jeder Endomorphismus von V hat einen Eigenwert in $\mathbb{R}$.*

Beweis. Wegen $\deg \chi_f$ ungerade (und χ_f normiert) ist $\lim\limits_{\lambda\to-\infty} \chi_f(\lambda) = -\infty$ und $\lim\limits_{\lambda\to+\infty} \chi_f(\lambda) = +\infty$. Nach dem Zwischenwertsatz aus der Analysis hat χ_f eine reelle Nullstelle. □

Für gerade Dimension ist Satz 2.8 dagegen falsch, siehe Beispiel 1.3.2.

2.9 Definition. Sind p, $q \in K[t]$ Polynome, so sagt man p *teilt* q und schreibt $p \mid q$, wenn es ein Polynom $h \in K[t]$ gibt mit $q = ph$.

Aus $p \mid q$ und $q \neq 0$ folgt $\deg(p) \leq \deg(q)$ nach Satz I.4.3.

2.10 Definition. Sei $p \in K[t]$, $p \neq 0$.

(a) Für $\lambda \in K$ heißt

$$\mu(p, \lambda) \;:=\; \max\bigl\{k \geq 0 \colon (t-\lambda)^k \mid p\bigr\}$$

die *Vielfachheit* der Nullstelle λ von p. (Man setzt $(t-\lambda)^0 := 1$.)

(b) p *zerfällt (über K) in Linearfaktoren*, wenn es c, $a_1, \ldots, a_n \in K$ gibt mit $p = c \cdot \prod_{i=1}^{n} (t - a_i)$.

Nach der Bemerkung zuvor ist $\mu(p, \lambda) \leq \deg(p)$. Es gilt $\mu(p, \lambda) \geq 1 \Leftrightarrow p(\lambda) = 0$ nach Korollar I.4.7.

2.11 Lemma. *Sei $p = (t-\lambda)^m \cdot q$ mit $m \geq 0$, $q \in K[t]$ und $q(\lambda) \neq 0$. Dann ist $\mu(p, \lambda) = m$.*

Beweis. Nach Definition ist $\mu(p, \lambda) \geq m$. Gäbe es ein Polynom $h \in K[t]$ mit $p = (t-\lambda)^{m+1} \cdot h$, so folgte $q = (t-\lambda) \cdot h$ nach Kürzen durch $(t-\lambda)^m$, also $q(\lambda) = 0$, Widerspruch zur Voraussetzung. □

2.12 Definition. Für $f \in \mathrm{End}(V)$ und $\lambda \in K$ heißt

$$\mu_g(f, \lambda) \;:=\; \dim \mathrm{Eig}(f, \lambda)$$

die *geometrische Vielfachheit* und

$$\mu_a(f, \lambda) \;:=\; \mu(\chi_f, \lambda)$$

die *algebraische Vielfachheit* des Eigenwerts λ. Für $A \in \mathrm{M}_n(K)$ definiert man $\mu_g(A,\lambda)$ und $\mu_a(A,\lambda)$ analog als die entsprechenden Vielfachheiten von $F_A \in \mathrm{End}(K^n)$.

2.13 Satz. *Sei $f \in \mathrm{End}(V)$, und sei U ein Untervektorraum von V mit $f(U) \subseteq U$. Dann induziert f Endomorphismen $f|_U$ von U und $\bar f$ von V/U. Für die charakteristischen Polynome gilt*

$$\chi_f \;=\; \chi_{f|U} \cdot \chi_{\bar f}.$$

Insbesondere ist $\chi_{f|U}$ ein Teiler von χ_f.

Beweis. $\bar f \in \mathrm{End}(V/U)$ ist definiert durch $\bar f(v+U) = f(v)+U$ $(v \in V)$; das ist wohldefiniert wegen $f(U) \subseteq U$. Ergänze eine Basis $(v_1,\dots,v_m)$ von U zu einer Basis $\mathcal{B} = (v_1,\dots,v_n)$ (mit $n \geq m$) von V. Dann ist

$$\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) \;=\; \begin{pmatrix} B & * \\ 0 & C \end{pmatrix}$$

mit $B \in \mathrm{M}_m(K)$ und $C \in \mathrm{M}_{n-m}(K)$. Bezüglich der Basis $(v_1,\dots,v_m)$ von U hat $f|_U$ die Matrix B, bezüglich der Basis $(v_{m+1}+U,\dots,v_n+U)$ von V/U hat $\bar f$ die Matrix C. Also ist

$$\chi_f \;=\; \det\Big(tI_n - \begin{pmatrix} B & * \\ 0 & C \end{pmatrix}\Big) \;=\; \det(tI_m - B)\,\det(tI_{n-m} - C) \;=\; \chi_{f|U} \cdot \chi_{\overline{f}}$$

nach dem Kästchensatz IV.3.1. □

2.14 Bemerkung. Sei $f \in \mathrm{End}(V)$, sei $U \subseteq V$ ein Untervektorraum mit $f(U) \subseteq U$. Im Beweis von Satz 2.13 haben wir gesehen:

Sei $\mathcal{B} = (v_1,\dots,v_n)$ eine Basis von V derart, daß $\mathcal{B}' = (v_1,\dots,v_m)$ (mit $m = \dim(U)$) eine Basis von U ist. Dann ist $\mathcal{B}'' = (v_{m+1}+U,\dots,v_n+U)$ eine Basis von V/U, und es ist

$$\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) \;=\; \begin{pmatrix} A' & * \\ 0 & A'' \end{pmatrix}$$

mit

$$A' \;=\; \mathrm{M}_{\mathcal{B}'}^{\mathcal{B}'}(f|_U) \in \mathrm{M}_m(K), \quad A'' \;=\; \mathrm{M}_{\mathcal{B}''}^{\mathcal{B}''}(\bar f) \in \mathrm{M}_{n-m}(K).$$

Dabei sind $f|_U \in \mathrm{End}(U)$ und $\bar f \in \mathrm{End}(V/U)$ die von f induzierten Endomorphismen.

2.15 Satz. *Für $f \in \mathrm{End}(V)$ und $\lambda \in K$ gilt:*

(a) $\mu_g(f,\lambda) + \mathrm{rk}(\lambda\,\mathrm{id} - f) = \dim(V)$;
(b) $\mu_a(f,\lambda) \geq \mu_g(f,\lambda)$;
(c) *λ ist Eigenwert von f* $\Leftrightarrow$ $\mu_a(f,\lambda) \geq 1$ $\Leftrightarrow$ $\mu_g(f,\lambda) \geq 1$.

Beweis. Satz III.3.11 sagt $\dim\ker(g) + \mathrm{rk}(g) = \dim(V)$ für $g \in \mathrm{End}(V)$. Für $g := \lambda\mathrm{id} - f$ gibt das Behauptung (a). (b) Sei $U = \mathrm{Eig}(f,\lambda)$. Es ist $f(U) \subseteq U$ und $\chi_{f|U} = (t-\lambda)^m$ mit $m = \dim(U) = \mu_g(f,\lambda)$. Nach 2.13 folgt $(t-\lambda)^m \mid \chi_f$, also $\mu_a(f,\lambda) \geq m$. (c) Es gilt: λ Eigenwert von $f \Rightarrow \mu_g(f,\lambda) \geq 1 \Rightarrow \mu_a(f,\lambda) \geq 1$ (nach (b)) $\Rightarrow \chi_f(\lambda) = 0 \Rightarrow \lambda$ Eigenwert von f (nach 2.3). □

2.16 Satz. *Für $f \in \mathrm{End}(V)$ sind äquivalent:*

(i) *f ist diagonalisierbar;*
(ii) *χ_f zerfällt in Linearfaktoren, und für alle $\lambda \in K$ gilt $\mu_g(f,\lambda) = \mu_a(f,\lambda)$.*

In Bedingung (ii) kann man sich dabei auf die Eigenwerte λ von f mit $\mu_a(f,\lambda) \geq 2$ beschränken.

BEWEIS. (i) $\Rightarrow$ (ii) ist klar: Ist $V = \bigoplus_{i=1}^r \mathrm{Eig}(f,\lambda_i)$, so ist $\chi_f = \prod_{i=1}^r (t-\lambda_i)^{m_i}$ mit $m_i = \mu_g(f,\lambda_i)$. Umgekehrt gelte $\mu_g(f,\lambda) = \mu_a(f,\lambda)$ für alle λ mit $\mu_a(f,\lambda) \geq 2$. Dann gilt diese Gleichheit auch für alle $\lambda \in K$, nach Satz 2.15. Weiter zerfalle χ_f in Linearfaktoren, etwa $\chi_f = \prod_{i=1}^r (t-\lambda_i)^{m_i}$ mit paarweise verschiedenen $\lambda_1, \ldots, \lambda_r \in K$ und mit $m_1, \ldots, m_r \in \mathbb{N}$. Der Unterraum $U := \sum_{i=1}^r \mathrm{Eig}(f,\lambda_i)$ von V hat dann die Dimension

$$\sum_i \mu_g(f,\lambda_i) \;=\; \sum_i \mu_a(f,\lambda_i) \;=\; \sum_i m_i \;=\; \deg(\chi_f) \;=\; \dim(V),$$

denn die Summe ist direkt nach 1.4. Also ist $U = V$. Somit ist f diagonalisierbar (Korollar 1.7). □

2.17 Beispiel. Ist die Matrix

$$A \;=\; \begin{pmatrix} -2 & 3 & -3 \\ 3 & -2 & 3 \\ 3 & -3 & 4 \end{pmatrix}$$

diagonalisierbar? Es ist

$$\chi_A \;=\; t^3 - 3t + 2 \;=\; (t-1)^2(t+2).$$

Also hat A die Eigenwerte 1 und -2, und es ist $\mu_a(A,1) = 2$ und $\mu_a(A,-2) = 1$ (wir setzen $\mathrm{char}(K) \neq 3$ voraus, das bedeutet $1 \neq -2$). Ohne Rechnung folgt bereits $\mu_g(A,-2) = 1$ (2.15). Um $\mu_g(A,1)$ zu bestimmen, berechne

$$\mathrm{rk}(I-A) \;=\; \mathrm{rk} \begin{pmatrix} 3 & -3 & 3 \\ -3 & 3 & -3 \\ -3 & 3 & -3 \end{pmatrix} \;=\; 1.$$

Nach 2.15(a) ist also $\mu_g(A,1) = 3 - 1 = 2$. Damit ist A diagonalisierbar nach 2.16, also $A \approx \mathrm{diag}(1,1,-2)$. Um auch eine diagonalisierende Matrix anzugeben, müssen wir Basen der Eigenräume bestimmen. Man findet

$$\mathrm{Eig}(A,1) \;=\; K(1,1,0)^t + K(0,1,1)^t, \quad \mathrm{Eig}(A,-2) \;=\; K(-1,1,1)^t.$$

Die Matrix

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

mit diesen Spalten erfüllt also $S^{-1}AS = \mathrm{diag}(1,1,-2)$.

2.18 Bemerkung. Das Beispiel zeigt, wie man allgemein die Diagonalisierbarkeit von $A \in \mathrm{M}_n(K)$ (oder $f \in \mathrm{End}(V)$) entscheidet und gegebenenfalls eine entsprechende Basiswechselmatrix findet:

1. Prüfe, ob χ_A in Linearfaktoren zerfällt. Falls nicht, ist A nicht diagonalisierbar. Sei also χ_A Produkt von Linearfaktoren.

2. Berechne $\mu_g(A,\lambda) = n - \mathrm{rk}(\lambda I - A)$ für jeden Eigenwert λ von A mit $\mu_a(A,\lambda) \geq 2$. Genau dann ist A diagonalisierbar, wenn $\mu_g(A,\lambda) = \mu_a(A,\lambda)$ für alle diese λ ist.
3. Um $S \in \mathrm{GL}_n(K)$ mit $S^{-1}AS =$ Diagonalmatrix zu finden, bestimme zu jedem Eigenwert λ von A eine Basis von $\mathrm{Eig}(A,\lambda)$, und nimm für S die Matrix mit allen diesen Eigenvektoren als Spalten.

Ist f nicht diagonalisierbar, so kann man immer noch versuchen, f in Dreiecksform zu bringen:

2.19 Definition.

(a) Ein Endomorphismus f von V heißt *trigonalisierbar*, wenn es eine Basis $\mathcal{B}$ von V gibt, so daß $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)$ eine Dreiecksmatrix ist.
(b) $A \in \mathrm{M}_n(K)$ heißt *trigonalisierbar*, wenn $F_A \in \mathrm{End}(K^n)$ trigonalisierbar ist, d.h. wenn es $S \in \mathrm{GL}_n(K)$ gibt mit $S^{-1}AS$ Dreiecksmatrix.

Jede obere Dreiecksmatrix ist ähnlich zu einer unteren Dreiecksmatrix (mache Basiswechsel von $\mathcal{K} = (e_1,\ldots,e_n)$ nach $(e_n,\ldots,e_1)$). Deshalb spielt es keine Rolle, ob die Dreiecksmatrizen in 2.19 obere oder untere sind.

Ist f trigonalisierbar, so zerfällt χ_f über K in Linearfaktoren (2.2.2). Davon gilt auch die Umkehrung:

2.20 Satz. *$f \in \mathrm{End}(V)$ ist genau dann trigonalisierbar, wenn χ_f in Linearfaktoren zerfällt.*

Beweis. Beweis von "$\Leftarrow$" durch Induktion nach $\dim(V)$. Für $\dim(V) = 1$ ist nichts zu zeigen. Sei $\dim(V) = n > 1$, sei $\chi_f = (t-\lambda_1)\cdots(t-\lambda_n)$ mit $\lambda_1,\ldots,\lambda_n \in K$, und sei $0 \neq u \in \mathrm{Eig}(f,\lambda_1)$. Für $U := Ku$ gilt $f(U) \subseteq U$. Sei $\bar{f} \in \mathrm{End}(V/U)$ der durch f induzierte Endomorphismus (siehe Satz 2.13). Wegen

$$\chi_f \;=\; (t-\lambda_1)\cdot\chi_{\bar{f}}$$

(2.13) folgt $\chi_{\bar{f}} = (t-\lambda_2)\cdots(t-\lambda_n)$ durch Kürzen von $t-\lambda_1$, d.h. auch $\chi_{\bar{f}}$ zerfällt in Linearfaktoren. Nach Induktionsvoraussetzung ist $\bar{f}$ also trigonalisierbar. Also gibt es $v_2,\ldots,v_n \in V$, so daß $\mathcal{C} := (v_2+U,\ldots,v_n+U)$ eine Basis von V/U und $\mathrm{M}_{\mathcal{C}}^{\mathcal{C}}(\bar{f}) =: C$ eine obere Dreiecksmatrix ist. Damit ist $\mathcal{B} := (u,v_2,\ldots,v_n)$ eine Basis von V, und es ist $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) = \begin{pmatrix}\lambda_1 & * \\ 0^t & C\end{pmatrix}$ (siehe 2.14), eine obere Dreiecksmatrix. □

2.21 Beispiel. Der Beweis von Satz 2.20 hat ein Verfahren geliefert, um eine Matrix mit zerfallendem charakteristischem Polynom zu trigonalisieren. Als Beispiel betrachte die 3×3-Matrix

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

mit charakteristischem Polynom $\chi_A = t^3 - 2t^2 + t = t(t-1)^2$. Die Matrix A ist also trigonalisierbar und hat die Eigenwerte 0 und 1. Wir testen zunächst, ob A

diagonalisierbar ist, und müssen dazu den Eigenwert 1 betrachten. Es ist

$$\operatorname{rk}(I-A) \;=\; \operatorname{rk}\begin{pmatrix} 0 & 0 & -1 \\ -2 & 2 & -3 \\ -1 & 1 & -1 \end{pmatrix} \;=\; 2$$

also $\mu_g(A,1) = 3-2 = 1 < 2 = \mu_a(A,1)$. Damit ist A nicht diagonalisierbar. Um A zu trigonalisieren, beginne mit einem Eigenvektor von A, z.B. mit $u = (-1,1,1)^t \in \operatorname{Eig}(A,0)$. Ergänze u zu einer Basis von K^3, etwa zu $\mathcal{B} = (u, e_2, e_3)$. Transformation von A auf diese Basis gibt wegen

$$T^{\mathcal{B}}_{\mathcal{K}} \;=\; \begin{pmatrix} -1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \;=\; T^{\mathcal{K}}_{\mathcal{B}}$$

das Resultat

$$\mathrm{M}^{\mathcal{B}}_{\mathcal{B}}(F_A) \;=\; \mathrm{T}^{\mathcal{K}}_{\mathcal{B}} \cdot A \cdot \mathrm{T}^{\mathcal{B}}_{\mathcal{K}} \;=\; \begin{pmatrix} 0 & 0 & -1 \\ 0 & -1 & 4 \\ 0 & -1 & 3 \end{pmatrix}.$$

Das 2×2-Kästchen rechts unten ist die Matrix von $\overline{F}_A \in \operatorname{End}(K^3/Ku)$ bezüglich der Basis $(\bar{e}_2, \bar{e}_3)$. Hiervon ist etwa $2\bar{e}_2 + \bar{e}_3$ ein Eigenvektor (zum Eigenwert $\lambda = 1$), den wir durch $\bar{e}_3$ zu einer Basis von K^3/Ku ergänzen können. Gehe also von $\mathcal{B}$ über zur Basis $\mathcal{C} = (u, 2e_2 + e_3, e_3)$ von K^3. Dann ist $\mathrm{M}^{\mathcal{C}}_{\mathcal{C}}(F_A)$ eine Dreiecksmatrix, nämlich

$$\mathrm{M}^{\mathcal{C}}_{\mathcal{C}}(F_A) \;=\; S^{-1}AS \;=\; \begin{pmatrix} 0 & -1 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

mit

$$S \;=\; \mathrm{T}^{\mathcal{C}}_{\mathcal{K}} \;=\; \begin{pmatrix} -1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Ein anderes Verfahren zur Trigonalisierung von Matrizen werden wir später sehen.

2.22 Satz und Definition. *Ein Körper K heißt* algebraisch abgeschlossen, *wenn die beiden folgenden äquivalenten Bedingungen gelten:*

(i) *Jedes Polynom $p \in K[t]$ mit $\deg(p) \geq 1$ hat eine Nullstelle in K;*
(ii) *jedes Polynom $p \in K[t]$ mit $\deg(p) \geq 1$ zerfällt über K in Linearfaktoren.*

Beweis. (ii) $\Rightarrow$ (i) ist klar. Umgekehrt gelte (i), und sei $\deg(p) \geq 1$. Wähle eine Nullstelle λ von p, dann ist $p = (t-\lambda)\,q$ mit $q \in K[t]$ (I.4.7) und $\deg(q) = \deg(p) - 1$ (I.4.3). Nun macht man induktiv mit q weiter. □

2.23 Beispiel. Die Körper $\mathbb{Q}$ oder $\mathbb{R}$ sind nicht algebraisch abgeschlossen, zum Beispiel hat $p = t^2 + 1$ keine Nullstelle. Ein endlicher Körper K ist niemals algebraisch abgeschlossen, denn das Polynom

$$p \;=\; 1 + \prod_{a \in K} (t-a) \;\in K[t]$$

erfüllt $p(a) = 1$ für jedes $a \in K$, hat also keine Nullstelle in K.

2.24 Theorem. (Fundamentalsatz der Algebra, Gauß) *Der Körper $\mathbb{C}$ der komplexen Zahlen ist algebraisch abgeschlossen.*

Es gibt mehrere Beweise, z. B. mit Methoden der Funktionentheorie oder der Topologie. In der Vorlesung Algebra (3. Semester) sehen wir einen Beweis mittels Galoistheorie. Mit den Mitteln des ersten Semesters läßt sich der Satz noch nicht beweisen.

Für Endomorphismen bzw. Matrizen bedeutet Theorem 2.24:

2.25 Korollar. *Jede Matrix aus* $\mathrm{M}_n(\mathbb{C})$*, oder jeder Endomorphismus eines endlich-dimensionalen* $\mathbb{C}$*-Vektorraums, ist trigonalisierbar.* □

Über $\mathbb{R}$ und für $n > 1$ ist die entsprechende Aussage falsch, wie schon gesehen.

3. Minimalpolynom, Satz von Hamilton-Cayley

Weiter seien alle Vektorräume endlich-dimensional.

3.1 Sei $A \in \mathrm{M}_n(K)$. Die Potenzen von A sind definiert durch $A^0 = I_n$ und $A^i = A \cdots A$ (i Faktoren) für $i \in \mathbb{N}$. Ist $p = \sum_{i=0}^m a_i t^i \in K[t]$ ein Polynom, so kann man A in p einsetzen. Man definiert also

$$p(A) \;:=\; \sum_{i=0}^m a_i A^i \;=\; a_m A^m + \cdots + a_1 A + a_0 I_n \;\in\; \mathrm{M}_n(K).$$

Dasselbe kann man für jeden K-Vektorraum V und jeden Endomorphismus f von V machen: Die Potenzen von f sind definiert durch $f^0 = \mathrm{id}_V$ und $f^i = f \circ \cdots \circ f$ (i Faktoren, $i \in \mathbb{N}$), und man definiert

$$p(f) \;:=\; \sum_{i=0}^m a_i f^i \;=\; a_m f^m + \cdots + a_1 f + a_0 \mathrm{id}_V \;\in\; \mathrm{End}(V).$$

3.2 Lemma. *Sei* $A \in \mathrm{M}_n(K)$*.*

(a) *Die Einsetzabbildung* $K[t] \to \mathrm{M}_n(K)$*,* $p \mapsto p(A)$ *ist ein Ringhomomorphismus. Dessen Bild*

$$K[A] \;:=\; \{p(A) \colon p \in K[t]\} \;=\; \mathrm{span}(A^i \colon i \geq 0)$$

ist ein Untervektorraum und ein kommutativer Teilring von $\mathrm{M}_n(K)$*.*

(b) *Für* $S \in \mathrm{GL}_n(K)$ *und* $p \in K[t]$ *ist* $p(SAS^{-1}) = S \cdot p(A) \cdot S^{-1}$*. Insbesondere ist* $K[SAS^{-1}] = \{SBS^{-1} \colon B \in K[A]\} =: S \cdot K[A] \cdot S^{-1}$*.*

Die analogen Aussagen gelten für $f \in \mathrm{End}(V)$ *und* $g \in \mathrm{GL}(V)$ *(in (b)).*

Beweis. (a) sagt insbesondere: Für $p, q \in K[t]$ gilt $(p+q)(A) = p(A) + q(A)$ und $(pq)(A) = p(A)q(A)$, bzw. $(p+q)(f) = p(f) + q(f)$ und $(pq)(f) = p(f) \circ q(f)$ für $f \in \mathrm{End}(V)$. Diese Behauptungen sind klar. (b) Für $i \geq 0$ ist

$$(SAS^{-1})^i \;=\; (SAS^{-1}) \cdot (SAS^{-1}) \cdots (SAS^{-1}) \;=\; SA^iS^{-1}.$$

Für $p = \sum_{i=0}^m a_i t^i \in K[t]$ ist also

$$\begin{aligned} p(SAS^{-1}) = \sum_{i=0}^m a_i (SAS^{-1})^i \;&=\; \sum_{i=0}^m a_i \cdot SA^iS^{-1} \\ &=\; S\Big(\sum_{i=0}^m a_i A^i\Big)S^{-1} \;=\; S \cdot p(A) \cdot S^{-1}. \end{aligned}$$

Analog für $f \in \mathrm{End}(V)$. □

3.3 Bemerkung. Ist eine Matrix $A \in \mathrm{M}_n(K)$ diagonalisiert, etwa

$$SAS^{-1} \;=\; \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

mit $S \in \mathrm{GL}_n(K)$ und $\lambda_1, \ldots, \lambda_n \in K$, so gilt für jedes Polynom $p \in K[t]$:

$$p(A) \;=\; S^{-1}\,\mathrm{diag}\big(p(\lambda_1), \ldots, p(\lambda_n)\big)\,S.$$

Zum Beispiel kann man dann die Potenzen A^i $(i \geq 0)$ geschlossen hinschreiben. Siehe Aufgaben 45 und K.

3.4 Definition. Ein Polynom $p \in K[t]$ heißt *normiert*, wenn $p \neq 0$ und der Leitkoeffizient von p (siehe I.4.2) gleich 1 ist.

3.5 Satz und Definition. *Sei $A \in \mathrm{M}_n(K)$ (bzw. sei V ein K-Vektorraum mit $\dim(V) < \infty$, und sei $f \in \mathrm{End}(V)$). Es gibt ein eindeutig bestimmtes normiertes Polynom $q \in K[t]$ kleinsten Grades mit $q(A) = 0$ (bzw. mit $q(f) = 0$). Dieses Polynom heißt das* Minimalpolynom *von A (bzw. von f) und wird mit q_A (bzw. q_f) bezeichnet.*

Beweis. Vorbemerkung: Gibt es ein Polynom $q \in K[t]$ mit $q(A) = 0$ und $\deg(q) = d \geq 0$, so gibt es auch ein normiertes Polynom $\tilde{q} \in K[t]$ mit $\tilde{q}(A) = 0$ und $\deg(\tilde{q}) = d$, nämlich $\tilde{q} = \frac{1}{c}\,q$, wobei $c \in K^*$ der Leitkoeffizient von q ist.

Wegen $\dim \mathrm{M}_n(K) = n^2$ sind die Potenzen $A^0 = I$, A, A^2, …, A^{n^2} von A linear abhängig. Es gibt also ein normiertes Polynom $q \in K[t]$ mit $q(A) = 0$ (tatsächlich mit $\deg(q) \leq n^2$). Also gibt es auch ein solches q von kleinstem Grad.

Seien $q_1, q_2 \in K[t]$ normiert mit $\deg(q_1) = \deg(q_2) = d$ und $q_1(A) = q_2(A) = 0$, und sei d minimal. Dann ist $\deg(q_1 - q_2) < d$, und es ist auch $(q_1 - q_2)(A) = 0$. Wäre $q_1 \neq q_2$, so gäbe es nach der Vorbemerkung ein normiertes $\tilde{q} \in K[t]$ mit $\deg(\tilde{q}) < d$ und $\tilde{q}(A) = 0$, Widerspruch zur minimalen Wahl von d. Also ist $q_1 = q_2$. Für $f \in \mathrm{End}(V)$ geht der Beweis völlig analog. □

3.6 Beispiele.

1. Ist $f \in \mathrm{End}(V)$, ist $\mathcal{B}$ eine Basis von V und $A = \mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)$, so ist $q_f = q_A$. Denn für jedes $p \in K[t]$ ist $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(p(f)) = p(\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)) = p(A)$, da $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}} \colon \mathrm{End}(V) \to \mathrm{M}_n(K)$ $(n = \dim(V))$ ein Ringhomomorphismus ist (III.5.7).

2. Ist $f = c\,\mathrm{id}_V$ mit $c \in K$, so ist $q_f = t - c$. Ist f eine Projektion (d. h. gilt $f^2 = f$), so ist $q_f = t^2 - t$, außer wenn $f = 0$ oder $f = \mathrm{id}$ ist.

3. Was ist das Minimalpolynom von $A = \left(\begin{smallmatrix} a & c \\ 0 & b \end{smallmatrix}\right) \in \mathrm{M}_2(K)$? Für $q = (t-a)(t-b)$ ist

$$q(A) \;=\; \begin{pmatrix} 0 & c \\ 0 & b-a \end{pmatrix} \begin{pmatrix} a-b & c \\ 0 & 0 \end{pmatrix} \;=\; 0.$$

Für $a \neq b$ oder für $c \neq 0$ ist also $q_A = (t-a)(t-b)$. Für $a = b$ und $c = 0$ ist $q_A = t - a$.

3.7 Satz. *Sei $A \in \mathrm{M}_n(K)$.*

(a) *Ist $p \in K[t]$ ein beliebiges Polynom mit $p(A) = 0$, so gilt $q_A \mid p$.*
(b) *Ist $B \in \mathrm{M}_n(K)$ mit $A \approx B$, so ist $q_B = q_A$.*

BEWEIS. (a) Division von p durch q_A mit Rest (Satz I.4.4) gibt Polynome g, r mit $p = q_A g + r$ und $\deg(r) < \deg(q_A)$. Einsetzen von A gibt

$$0 \;=\; p(A) \;=\; q_A(A)g(A) + r(A) \;=\; r(A).$$

Wegen $\deg(r) < \deg(q_A)$ folgt $r = 0$. Also gilt $q_A \mid p$.

(b) Für $B = SAS^{-1}$ mit $S \in \mathrm{GL}_n(K)$ und jedes Polynom $q \in K[t]$ ist $q(B) = S\,q(A)\,S^{-1}$ (3.2(b)), also gilt $q(B) = 0$ genau dann, wenn $q(A) = 0$ ist. Daraus folgt $q_A = q_B$. □

3.8 Bemerkung. Folgende Bedingungen sind notwendig für $A \approx B$:

$$\mathrm{rk}(A) \;=\; \mathrm{rk}(B), \quad \chi_A \;=\; \chi_B, \quad q_A \;=\; q_B.$$

Auch die Kombination aus allen dreien ist jedoch noch nicht hinreichend (siehe nächster Abschnitt).

3.9 Theorem. (Hamilton-Cayley[1]) *Für $f \in \mathrm{End}(V)$ gilt $\chi_f(f) = 0$. Für $A \in \mathrm{M}_n(K)$ gilt $\chi_A(A) = 0$.*

Setzt man also die Matrix A in ihr charakteristisches Polynom ein, so ergibt sich die Nullmatrix. Zum Satz von Hamilton-Cayley ist nach Satz 3.7(a) die folgende Formulierung äquivalent:

3.10 Korollar. *Für $f \in \mathrm{End}(V)$ gilt $q_f \mid \chi_f$. Für $A \in \mathrm{M}_n(K)$ gilt $q_A \mid \chi_A$.* □

BEWEIS VON 3.9. Wir beweisen $\chi_f(f) = 0$ und müssen dafür $\chi_f(f)(v) = 0$ zeigen für jeden Vektor $v \in V$. Sei $v \neq 0$ fixiert. Es gibt eine kleinste Zahl $k \in \mathbb{N}$ derart, daß die Folge $(v, f(v), \ldots, f^k(v))$ in V linear abhängig ist. Wegen k minimal ist die Folge $(v, \ldots, f^{k-1}(v))$ linear unabhängig, und es gibt (eindeutige) $a_0, \ldots, a_{k-1} \in K$ mit

$$f^k(v) \;=\; a_0 v + a_1 f(v) + \cdots + a_{k-1} f^{k-1}(v).$$

Ergänze die Folge $(v, \ldots, f^{k-1}(v))$ zu einer Basis $\mathcal{B} = (v, \ldots, f^{k-1}(v), w_1, \ldots, w_m)$ von V. Dann gilt

$$\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) \;=\; \begin{pmatrix} A & * \\ 0 & B \end{pmatrix}$$

mit

$$A \;=\; \begin{pmatrix} 0 & & & & a_0 \\ 1 & 0 & & & a_1 \\ & 1 & 0 & & a_2 \\ & & \ddots & \ddots & \vdots \\ & & & 0 & a_{k-2} \\ & & & 1 & a_{k-1} \end{pmatrix} \;\in\; \mathrm{M}_n(K).$$

Es folgt $\chi_f = \chi_A \cdot \chi_B = \chi_B \cdot \chi_A$. Insbesondere ist

$$\chi_f(f) \;=\; \chi_B(f) \circ \chi_A(f).$$

[1] Sir William Rowan HAMILTON (1805–1865), Arthur CAYLEY (1821–1895)

Nach Aufgabe 43 ist $\chi_A = t^k - \sum_{i=0}^{k-1} a_i t^i$. Es folgt also insbesondere

$$\chi_A(f)(v) \;=\; f^k(v) - \sum_{i=0}^{k-1} a_i f^i(v) \;=\; 0,$$

und daher $\chi_f(f)(v) = \chi_B(f)\big(\chi_A(f)(v)\big) = 0$. □

3.11 Korollar. *Für $f \in \mathrm{End}(V)$ haben die Polynome χ_f und q_f dieselben Nullstellen in K. Analog für $A \in \mathrm{M}_n(K)$.*

Beweis. Sei $\lambda \in K$ mit $\chi_f(\lambda) = 0$. Nach 2.3 ist λ ein Eigenwert von f, also gibt es $0 \neq v \in V$ mit $f(v) = \lambda v$. Für jedes $p \in K[t]$ ist $p(f)(v) = p(\lambda) \cdot v$. Wegen $q_f(f)(v) = 0$ folgt daraus $q_f(\lambda) = 0$. Also ist jede Nullstelle von χ_f auch eine von q_f. Die Umkehrung folgt aus $q_f \mid \chi_f$ (3.10). □

3.12 Korollar. *Für $A \in \mathrm{M}_n(K)$ ist $\deg(q_A) \leq n$. Der Teilring $K[A]$ von $\mathrm{M}_n(K)$ hat als K-Vektorraum die Dimension $\deg(q_A)$. Analog für $f \in \mathrm{End}(V)$ und $K[f] \subseteq \mathrm{End}(V)$.*

Beweis. Aus $q_A \mid \chi_A$ (3.10) folgt $\deg(q_A) \leq \deg(\chi_A) = n$. Die Einsetzabbildung $K[t] \to K[A]$, $p \mapsto p(A)$ ist linear und surjektiv. Ihr Kern ist der Unterraum $U := \{p \in K[t] \colon q_A \mid p\}$ von $K[t]$ (Satz 3.7(a)). Also ist $K[A] \cong K[t]/U$ (als Vektorraum). Andererseits ist $\dim K[t]/U = \deg(q_A)$ gemäß Polynomdivision mit Rest (I.4.4, siehe auch Aufgabe 44). □

3.13 Bemerkungen.

1. Im allgemeinen kann $\deg(q_A) < n$, also $q_A \neq \chi_A$, sein, wie man für $A = \lambda I_n$ und $n \geq 2$ sieht.

2. Für $A \in \mathrm{GL}_n(K)$ erhält man aus dem Satz von Hamilton-Cayley eine Formel für A^{-1} als Polynom in A: Ist

$$\chi_A \;=\; t^n + a_1 t^{n-1} + \cdots + a_{n-1} t + a_n$$

mit $a_i \in K$, so gilt $0 = \chi_A(A) = A \cdot \big(A^{n-1} + a_1 A^{n-2} + \cdots + a_{n-1} I_n\big) + a_n I_n$ nach Theorem 3.9. Wegen $a_n = (-1)^n \det(A) \neq 0$ (2.4) folgt daraus

$$A^{-1} \;=\; \frac{(-1)^{n+1}}{\det(A)} \Big(A^{n-1} + a_1 A^{n-2} + \cdots + a_{n-2} A + a_{n-1} I_n \Big).$$

Für $A \in \mathrm{GL}_2(K)$ etwa ergibt das die Formel

$$A^{-1} \;=\; \frac{1}{\det(A)} \Big(\mathrm{tr}(A) \cdot I_2 - A \Big).$$

4. Die Jordansche Normalform

Im ganzen Abschnitt ist V ein K-Vektorraum mit $\dim(V) < \infty$. Das Ziel ist es, einen gegebenen Endomorphismus von V (oder eine gegebene quadratische Matrix) durch Wahl einer geeigneten Basis in möglichst einfacher Form zu beschreiben. Man wünscht sich dabei eine Gestalt, aus der sich möglichst alle wichtigen Eigenschaften und Invarianten von f direkt ablesen lassen, wie etwa Eigenwerte, Dimension der Eigenräume, Minimalpolynom usw. Die Jordansche Normalform existiert für alle f mit zerfallendem charakteristischem Polynom und genügt allen

diesen Anforderungen. Gleichzeitig erhalten wir eine vollständige Klassifizierung der Ähnlichkeitsklassen von Matrizen, deren charakteristisches Polynom in Linearfaktoren zerfällt. Ist der Körper K algebraisch abgeschlossen, etwa $K = \mathbb{C}$, so werden also alle Ähnlichkeitsklassen von Matrizen klassifiziert.

4.1 Definition. Sei X eine Menge, seien $\mathcal{F} = (x_1, \ldots, x_m)$ und $\mathcal{G} = (y_1, \ldots, y_n)$ zwei (endliche) Folgen von Elementen von X. Die *Konkatenation* von $\mathcal{F}$ und $\mathcal{G}$ ist die Folge

$$\mathcal{F} \sqcup \mathcal{G} := (x_1, \ldots, x_m, y_1, \ldots, y_n).$$

Wir beginnen mit nilpotenten Endomorphismen, deren Studium eine Schlüsselrolle spielt.

4.2 Definition. Sei $f \in \mathrm{End}(V)$. Ein Untervektorraum U von V heißt *f-invariant*, wenn $f(U) \subseteq U$ ist. Analoge Definition für $A \in \mathrm{M}_n(K)$: Ein Untervektorraum $U \subseteq K^n$ heißt *A-invariant*, wenn U invariant unter F_A ist.

4.3 Bemerkung. Jeder Eigenraum von f, oder auch jeder Untervektorraum eines Eigenraums von f, ist f-invariant.

4.4 Lemma. *Für $f, g \in \mathrm{End}(V)$ mit $f \circ g = g \circ f$ sind die Unterräume* $\ker(g)$ *und* $\mathrm{im}(g)$ *von V f-invariant. Alle Eigenräume von g sind f-invariant.*

Beweis. Für $v \in \ker(g)$ ist $g(f(v)) = f(g(v)) = 0$, also $f(v) \in \ker(g)$. Für $v \in \mathrm{im}(g)$, etwa $v = g(w)$ mit $w \in V$, ist $f(g(v)) = g(f(v)) \in \mathrm{im}(g)$. Der Zusatz folgt aus $\mathrm{Eig}(g, \lambda) = \ker(g - \lambda \,\mathrm{id})$ und $f \circ (g - \lambda \,\mathrm{id}) = (g - \lambda \,\mathrm{id}) \circ f$. □

4.5 Definition. Ein Endomorphismus f von V heißt *nilpotent*, wenn es ein $k \in \mathbb{N}$ gibt mit $f^k = 0$. Das kleinste solche k heißt der *Nilpotenzindex* von f. Analoge Definitionen für $A \in \mathrm{M}_n(K)$.

4.6 Bemerkungen.

1. $A = \left(\begin{smallmatrix} 6 & -4 \\ 9 & -6 \end{smallmatrix}\right)$ erfüllt $A^2 = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, also ist A nilpotent vom Nilpotenzindex 2.

2. Für jede Basis $\mathcal{B}$ von V gilt: $f \in \mathrm{End}(V)$ ist genau dann nilpotent, wenn die Matrix $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)$ nilpotent ist. Denn $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f^k) = \mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f)^k$ für alle k (III.5.7).

3. Sind A, $B \in \mathrm{M}_n(K)$ ähnlich, so ist A genau dann nilpotent, wenn B nilpotent ist, und alsdann haben beide denselben Nilpotenzindex.

4.7 Satz. *Sei* $\dim(V) = n$*, und sei $f \in \mathrm{End}(V)$. Genau dann ist f nilpotent, wenn es eine Basis $\mathcal{B}$ von V so gibt, daß für* $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) = (a_{ij})$ *gilt $a_{ij} = 0$ für alle $i \geq j$, also*

$$\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) = \begin{pmatrix} 0 & * & \cdots & * \\ & 0 & \cdots & * \\ & & \ddots & \vdots \\ & & & 0 \end{pmatrix}.$$

Alsdann ist $f^n = 0$.

Beweis. "⇐": Sei $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) = A = (a_{ij})$ mit $a_{ij} = 0$ für $i \geq j$. Dann ist $\chi_A = t^n$ (2.2.2), also $A^n = \chi_A(A) = 0$ nach Hamilton-Cayley (3.9).

"⇒": Sei f nilpotent. Ist $\dim(V) = 1$, so ist $f = 0$. Beweis des allgemeinen Falls durch Induktion nach $\dim(V)$. Sei $f^k = 0$ und $f^{k-1} \neq 0$. Der Unterraum $U := \ker(f^{k-1})$ von V ist f-invariant (Lemma 4.4), und es ist $U \neq V$ wegen $f^{k-1} \neq 0$. Weiter gilt $f(V) \subseteq U$, denn $f^{k-1}(f(v)) = 0$ für alle $v \in V$. Da auch $f|_U \in \mathrm{End}(U)$ nilpotent ist, gibt es nach Induktionsvoraussetzung eine Basis $\mathcal{F}$ von U, so daß $\mathrm{M}_{\mathcal{F}}^{\mathcal{F}}(f|_U) =: B$ eine obere Dreiecksmatrix mit Nullen auf der Diagonalen ist. Ergänze $\mathcal{F}$ zu einer Basis $\mathcal{B} = \mathcal{F} \sqcup \mathcal{G}$ von V. Dann ist $\mathrm{M}_{\mathcal{B}}^{\mathcal{B}}(f) = \begin{pmatrix} B & * \\ 0 & 0 \end{pmatrix}$, hat also die gewünschte Form. □

4.8 Korollar. *Sei* $\dim(V) = n < \infty$, *sei* $f \in \mathrm{End}(V)$. *Dann gilt:* f *ist nilpotent* $\Leftrightarrow\ \chi_f = t^n\ \Leftrightarrow\ f^n = 0$.

Beweis. Aus f nilpotent folgt $\chi_f = t^n$ nach 4.7. Aus $\chi_f = t^n$ folgt $f^n = 0$ nach Hamilton-Cayley, und insbesondere ist dann f nilpotent. □

Folgender Satz ist der erste von zwei Hauptschritten im Beweis der Jordanschen Normalform:

4.9 Satz. *Sei* $\dim(V) = n$ *und* $f \in \mathrm{End}(V)$, *sei* $r := \mu_a(f, 0)$. *Es gibt eine Zahl* d *mit* $0 \leq d \leq r$, *so daß gelten:*

(1) $\{0\} = \ker(f^0) \subsetneq \ker(f) \subsetneq \cdots \subsetneq \ker(f^d)$ *und* $\ker(f^d) = \ker(f^{d+i})$ *für alle* $i \geq 0$;
(2) $V = \mathrm{im}(f^0) \supsetneq \mathrm{im}(f) \supsetneq \cdots \supsetneq \mathrm{im}(f^d)$ *und* $\mathrm{im}(f^d) = \mathrm{im}(f^{d+i})$ *für alle* $i \geq 0$.

Wir setzen $U := \ker(f^d)$ *und* $W := \mathrm{im}(f^d)$. *Dann gilt weiter:*

(a) $\dim(U) = r$ *und* $\dim(W) = n - r$.
(b) $V = U \oplus W$.
(c) U *und* W *sind* f*-invariant,* $f|_U \in \mathrm{End}(U)$ *ist nilpotent, und* $f|_W \in \mathrm{End}(W)$ *ist bijektiv.*

Beweis. Setze $U_i := \ker(f^i)$ für $i \geq 0$. Dann gilt

$$\{0\} = U_0 \subseteq U_1 \subseteq U_2 \subseteq \cdots$$

Wegen $\dim(V) < \infty$ kann $\neq$ nur an endlich vielen Stellen gelten. Für alle $i \geq 0$ ist

$$U_{i+1} = f^{-1}(U_i).$$

Denn für $v \in V$ gilt: $v \in U_{i+1} \Leftrightarrow 0 = f^{i+1}(v) = f^i(f(v)) \Leftrightarrow f(v) \in \ker(f^i) = U_i$. Ist $i \geq 1$ mit $U_{i-1} = U_i$, so folgt also auch $U_i = U_{i+1}$. Daher gibt es ein $d \geq 0$ mit (1). Für die Unterräume $W_i := \mathrm{im}(f^i)$ $(i \geq 0)$ von V gilt

$$V = W_0 \supseteq W_1 \supseteq W_2 \supseteq \cdots$$

Für alle $i \geq 0$ ist $\dim(U_i) + \dim(W_i) = n$ (III.3.11). Deshalb gilt auch (2), mit demselben d wie in (1).

Setze also $U = U_d$ und $W = W_d$. Beide Unterräume sind f-invariant (4.4), und $f|_U$ ist nilpotent wegen $(f|_U)^d = (f^d)|_U = 0$. Wegen $f(W) = W_{d+1} = W$ ist $f|_W \in \mathrm{End}(W)$ surjektiv, also auch bijektiv (III.3.12). Also ist $U \cap W = \ker(f^d) \cap W = \{0\}$. Andererseits ist $\dim(U) + \dim(W) = n$, und es folgt $V = U \oplus W$.

Aus (1) folgt $\dim(U) = \dim(U_d) \geq d$. Zu zeigen bleibt noch $\dim(U) = r = \mu_a(f, 0)$ (daraus folgt dann auch $r \geq d$). Es ist $\chi_f = \chi_{f|U} \cdot \chi_{f|W}$ (2.13). Wegen

$\chi_{f|U} = t^{\dim(U)}$ (4.8) und $\chi_{f|W}(0) = \pm\det(f|_W) \neq 0$ (wegen $f|_W$ injektiv) folgt $\mu_a(f,0) = \dim(U)$ aus Lemma 2.11. Der Satz ist bewiesen. □

4.10 Korollar. *Sei* $\dim(V) = n < \infty$*, sei* $f \in \mathrm{End}(V)$ *und* $\lambda \in K$ *sowie* $r := \mu_a(f,\lambda)$*. Für die Unterräume* $U(\lambda) := \ker((f-\lambda\,\mathrm{id})^n)$*,* $W(\lambda) := \mathrm{im}((f-\lambda\,\mathrm{id})^n)$ *von* V *gilt:*

(a) $U(\lambda) \oplus W(\lambda) = V$*;*
(b) $U(\lambda)$ *und* $W(\lambda)$ *sind* f*-invariant;*
(c) $\dim U(\lambda) = r$*,* $\dim W(\lambda) = n - r$*;*
(d) $\chi_{f|U(\lambda)} = (t-\lambda)^r$*, und* λ *ist kein Eigenwert von* $f|_{W(\lambda)}$*.*

Beweis. Setze $g := f - \lambda\,\mathrm{id}$. Dann ist

$$\chi_g(t) \;=\; \det(t\,\mathrm{id} - g) \;=\; \det\big((t+\lambda)\,\mathrm{id} - f\big) \;=\; \chi_f(t+\lambda).$$

Es ist $\chi_f = (t-\lambda)^r\, p$ mit $p \in K[t]$ und $p(\lambda) \neq 0$. Das gibt $\chi_g = t^r\,\tilde{q}$ mit $\tilde{q}(t) = q(t+\lambda)$ und $\tilde{q}(0) \neq 0$. Somit ist $r = \mu_a(g,0)$. Anwendung von Satz 4.9 auf g gibt (a), (c) und (d) sowie die g-Invarianz von $U(\lambda)$ und $W(\lambda)$, also auch ihre f-Invarianz wegen $f = g + \lambda\mathrm{id}$. □

4.11 Definition. Sei $\dim(V) = n < \infty$, und sei $f \in \mathrm{End}(V)$. Für $\lambda \in K$ heißt

$$\mathrm{Hau}(f,\lambda) \;:=\; \ker\big((f-\lambda\,\mathrm{id})^n\big)$$

der *Hauptraum* (oder *verallgemeinerte Eigenraum*) von f zum Parameter λ.

Es ist also $\mathrm{Hau}(f,\lambda)$ ein f-invarianter Unterraum von V, und mit $r := \mu_a(f,\lambda)$ gilt $\dim \mathrm{Hau}(f,\lambda) = r$ und $\mathrm{Hau}(f,\lambda) = \ker((f-\lambda\,\mathrm{id})^k)$ für alle $k \geq r$ (Korollar 4.10).

4.12 Satz. (Hauptraumzerlegung) *Sei* $f \in \mathrm{End}(V)$*, sei* $\chi_f = \prod_{i=1}^k (t-\lambda_i)^{r_i}$ *mit* $r_i \in \mathbb{N}$ *und* $\lambda_1, \ldots, \lambda_k \in K$ *paarweise verschieden. Dann gilt:*

(a) $V \;=\; \mathrm{Hau}(f,\lambda_1) \oplus \cdots \oplus \mathrm{Hau}(f,\lambda_k)$*;*
(b) $\dim \mathrm{Hau}(f,\lambda_i) \;=\; r_i \;=\; \mu_a(f,\lambda_i)$ *für* $i = 1, \ldots, k$*;*
(c) $\mathrm{Hau}(f,\lambda_i)$ *ist* f*-invariant, und* $(f-\lambda_i\,\mathrm{id})|_{\mathrm{Hau}(f,\lambda_i)}$ *ist nilpotent für* $i = 1, \ldots, k$*.*

Aus Satz 4.12 und Satz 4.7 folgt also: Bezüglich einer geeigneten Basis von V hat f eine Matrix der Form

$$\begin{pmatrix} \boxed{\begin{matrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_1 \end{matrix}} & & & \\ & \boxed{\begin{matrix} \lambda_2 & & * \\ & \ddots & \\ 0 & & \lambda_2 \end{matrix}} & & \\ & & \ddots & \\ & & & \boxed{\begin{matrix} \lambda_k & & * \\ & \ddots & \\ 0 & & \lambda_k \end{matrix}} \end{pmatrix} \qquad (*)$$

wobei das i-te Kästchen die Größe $r_i \times r_i$ hat ($i = 1, \ldots, k$). Außerhalb der k Kästchen stehen Nullen.

Aussagen (b) und (c) von Satz 4.12 wurden in Korollar 4.10 bewiesen. Nach (b) ist $\sum_{i=1}^{k} \dim \mathrm{Hau}(f, \lambda_i) = \dim(V)$. Deshalb folgt (a) aus folgendem Lemma:

4.13 Lemma. *Sei $f \in \mathrm{End}(V)$ beliebig, und seien $\lambda_1, \ldots, \lambda_k \in K$ paarweise verschieden. Dann ist die Summe $\sum_{i=1}^{k} \mathrm{Hau}(f, \lambda_i)$ (von Unterräumen von V) direkt.*

Beweis. Der Beweis geht analog wie bei Satz 1.4. Angenommen falsch, dann wählen wir ein Gegenbeispiel mit minimalem $k \geq 2$. Es gibt dann Vektoren $v_i \in \mathrm{Hau}(f, \lambda_i)$ mit $v_i \neq 0$ für $i = 1, \ldots, k$ und mit $v_1 + \cdots + v_k = 0$. Anwenden von $(f - \lambda_1 \,\mathrm{id})^n$ (mit $n = \dim(V)$) gibt $w_2 + \cdots + w_k = 0$ mit $w_i := (f - \lambda_1 \,\mathrm{id})^n(v_i)$ ($i = 2, \ldots, k$). Für $i = 2, \ldots, k$ ist $w_i \in \mathrm{Hau}(f, \lambda_i)$, da die Haupträume f-invariant sind, und $w_i \neq 0$, da λ_1 kein Eigenwert von $f|_{\mathrm{Hau}(f,\lambda_i)}$ ist (4.10). Das ist ein Widerspruch zur minimalen Wahl von k. □

4.14 Korollar. (Jordan-Chevalley Zerlegung[2]) *Sei $A \in \mathrm{M}_n(K)$ derart, daß χ_A in Linearfaktoren zerfällt. Dann gibt es eine diagonalisierbare Matrix D und eine nilpotente Matrix N mit $A = D + N$ und $DN = ND$.*

Beweis. Nach 4.12 gibt es $S \in \mathrm{GL}_n(K)$, so daß $\tilde{A} := SAS^{-1}$ die Form ($*$) aus 4.12 hat. Definiere (mit den dortigen Bezeichnungen)

$$\tilde{D} := \mathrm{diag}\Big(\underbrace{\lambda_1, \ldots, \lambda_1}_{r_1}, \ldots, \underbrace{\lambda_k, \ldots, \lambda_k}_{r_k}\Big)$$

und $\tilde{N} := \tilde{A} - \tilde{D}$. Dann kommutieren $\tilde{D}$ und $\tilde{N}$, also auch $D := S^{-1}\tilde{D}S$ und $N := S^{-1}\tilde{N}S$. Es ist $D + N = S^{-1}\tilde{A}S = A$, und D ist diagonalisierbar, N ist nilpotent. □

4.15 Bemerkungen.

1. Man kann zeigen, daß die Matrizen D und N in der Jordan-Chevalley Zerlegung 4.14 eindeutig bestimmt sind.

2. Um die Jordan-Chevalley Zerlegung einer Matrix A konkret durchzuführen, muß man Basen der Haupträume von A bestimmen und A auf eine aus diesen Basen zusammengesetzte Basis von K^n transformieren. Es genügt nicht, A irgendwie zu trigonalisieren, sondern man muß die speziellere Gestalt ($*$) aus 4.12 herstellen. Betrachte die Dreiecksmatrix

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

Für $D = \mathrm{diag}(1, 1, 2)$ und $N = A - D = E_{12} + E_{13}$ ist

$$DN = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = ND.$$

[2]Camille Jordan (1838–1922), Claude Chevalley (1909–1984)

Um die Jordan-Chevalley Zerlegung von A zu bestimmen, berechnet man vielmehr die Haupträume und findet $\mathrm{Hau}(A,1) = \ker((A-I)^2) = \mathrm{span}(e_1, e_2)$ sowie $\mathrm{Hau}(A,2) = \mathrm{Eig}(A,2) = \mathrm{span}(e_1 + e_3)$. Für $S = M(e_1, e_2, e_1 + e_3)$, die Matrix mit diesen Spalten, hat

$$S^{-1}AS \;=\; \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

die Form $(*)$. Daher ist die Jordan-Chevalley Zerlegung von A gegeben als $A = D + N$ mit

$$D \;=\; S \cdot \mathrm{diag}(1,1,2) \cdot S^{-1} \;=\; \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

und $N = A - D = E_{12}$.

4.16 Korollar. *Sei $f \in \mathrm{End}(V)$ derart, daß χ_f in Linearfaktoren zerfällt. Dann gibt es Endomorphismen f_d, f_n von V mit $f = f_d + f_n$ und $f_d \circ f_n = f_n \circ f_d$ derart, daß f_d diagonalisierbar und f_n nilpotent ist.* □

Wir verfeinern jetzt die Hauptraumzerlegung, indem wir die Restriktion von f auf seine Haupträume genauer analysieren und dadurch zur Jordanschen Normalform kommen.

4.17 Notation. Für $m \in \mathbb{N}$ und $\lambda \in K$ heißt die $m \times m$-Matrix

$$J_m(\lambda) \;:=\; \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix} \;=\; (a_{ij})_{1 \le i,j \le m}$$

mit $a_{ij} = \lambda$ für $i = j$, $a_{ij} = 1$ für $j - i = 1$ und $a_{ij} = 0$ sonst der *Jordanblock* der Größe m zum Parameter λ.

Die wesentliche Arbeit liegt im folgenden Satz:

4.18 Satz. *Sei $f \in \mathrm{End}(V)$ mit $\chi_f = (t-\lambda)^n$ für ein $\lambda \in K$. Bezüglich einer geeigneten Basis von V wird f beschrieben durch eine Matrix der Form*

$$\begin{pmatrix} \boxed{J_{m_1}(\lambda)} & & & \\ & \boxed{J_{m_2}(\lambda)} & & \\ & & \ddots & \\ & & & \boxed{J_{m_r}(\lambda)} \end{pmatrix}$$

mit $m_1 \ge \cdots \ge m_r \ge 1$ (und $m_1 + \cdots + m_r = n$). Unter dieser Nebenbedingung sind r und $m_1, \ldots, m_r$ sogar eindeutig bestimmt.

BEWEIS. Es genügt, den Satz für nilpotentes f (also für $\lambda = 0$) zu beweisen. Der allgemeine Fall folgt dann durch Anwendung des nilpotenten Falls auf $f - \lambda\,\mathrm{id}$.

Sei also $f \in \mathrm{End}(V)$ nilpotent mit $f^d = 0$ und $f^{d-1} \neq 0$. Wie im Beweis von 4.9 sei $U_i = \ker(f^i)$ für $i \geq 0$. Es gilt

$$\{0\} \;=\; U_0 \;\subsetneq\; U_1 \;\subsetneq\; \cdots \;\subsetneq\; U_{d-1} \;\subsetneq\; U_d = V,$$

die U_i sind f-invariant, und es ist $f^{-1}(U_{i-1}) = U_i$ für $i \geq 1$ (siehe Beweis von 4.9), insbesondere $f(U_i) \subseteq U_{i-1}$. Sei L_d ein Unterraum von V mit

$$U_d \;=\; U_{d-1} \oplus L_d.$$

Es folgt $f(L_d) \subseteq U_{d-1}$ und $L_d \cap f^{-1}(U_{d-2}) = L_d \cap U_{d-1} = \{0\}$, und daraus insbesondere $f(L_d) \cap U_{d-2} = \{0\}$. Daher gibt es einen Unterraum L_{d-1} mit

$$f(L_d) \;\subseteq\; L_{d-1} \;\text{ und }\; U_{d-1} \;=\; U_{d-2} \oplus L_{d-1}.$$

Induktiv findet man so für jedes $i = d, d-1, \ldots, 1$ einen Unterraum L_i mit $U_i = U_{i-1} \oplus L_i$ und mit $f(L_{i+1}) \subseteq L_i$ für $i < d$. Aus $U_1 = U_0 \oplus L_1$ und $U_0 = \{0\}$ folgt $L_1 = U_1$. Somit ist

$$\begin{aligned} V \;&=\; U_d \\ &=\; L_d \oplus U_{d-1} \\ &=\; L_d \oplus L_{d-1} \oplus U_{d-2} \\ &\quad\cdots \\ &=\; L_d \oplus L_{d-1} \oplus \cdots \oplus L_2 \oplus U_1 \\ &=\; L_d \oplus L_{d-1} \oplus \cdots \oplus L_2 \oplus L_1 \end{aligned}$$

Wegen $L_1 = \ker(f)$ ist die Restriktion von f auf $L_d \oplus \cdots \oplus L_2$ injektiv. Für $2 \leq i \leq d$ ist also $f|_{L_i}$ eine injektive lineare Abbildung $L_i \to L_{i-1}$.

Konstruiere jetzt eine an diese Zerlegung adaptierte Basis von V. Sei $\mathcal{B}_d$ eine Basis von L_d. Für $d > 1$ ist $f(\mathcal{B}_d)$ eine linear unabhängige Familie in L_{d-1}, es gibt also eine Familie $\mathcal{B}_{d-1}$, so daß $f(\mathcal{B}_d) \sqcup \mathcal{B}_{d-1}$ eine Basis von L_{d-1} ist. So fortfahrend erhalten wir folgende Basen der Unterräume L_i:

$$\begin{array}{ll} L_d: & \mathcal{B}_d \\ L_{d-1}: & f(\mathcal{B}_d) \sqcup \mathcal{B}_{d-1} \\ L_{d-2}: & f^2(\mathcal{B}_d) \sqcup f(\mathcal{B}_{d-1}) \sqcup \mathcal{B}_{d-2} \\ \cdots & \quad\cdots \\ L_1: & f^{d-1}(\mathcal{B}_d) \sqcup f^{d-2}(\mathcal{B}_{d-1}) \sqcup \cdots \sqcup f(\mathcal{B}_2) \sqcup \mathcal{B}_1 \end{array} \qquad (**)$$

Die Gesamtheit *aller* dieser Vektoren bildet eine Basis von V. Für jedes $i = 1, \ldots, d$ und jeden Vektor v aus $\mathcal{B}_i$ ist

$$\mathcal{C}_v \;:=\; \big(f^{i-1}(v),\, f^{i-2}(v),\, \ldots,\, f(v),\, v\big)$$

eine Basis eines f-invarianten Teilraums $W(v)$ von V, denn $f^i(v) = 0$ wegen $v \in L_i \subseteq U_i = \ker(f^i)$. Bezüglich dieser Basis hat $f|_{W(v)}$ die Matrix $J_i(0)$. Wir können die Gesamtheit der Vektoren $(**)$ also so sortieren, daß f bezüglich dieser Basis von V eine Matrix der gewünschten Form bekommt: Beginne mit den Basen $\mathcal{C}_v$ für v in $\mathcal{B}_d$, nimm dann die $\mathcal{C}_v$ für v in $\mathcal{B}_{d-1}$ usw., bis zu den $\mathcal{C}_v$ für v in $\mathcal{B}_1$. Ist also

$$\mathcal{B}_d \sqcup \mathcal{B}_{d-1} \sqcup \cdots \sqcup \mathcal{B}_1 \;=:\; (v_1, \ldots, v_r),$$

so ist

$$\mathcal{C} \;:=\; \mathcal{C}_{v_1} \sqcup \mathcal{C}_{v_2} \sqcup \cdots \sqcup \mathcal{C}_{v_r}$$

eine Basis von V der gewünschten Art. Die so erhaltene Matrix $\mathrm{M}_{\mathcal{C}}^{\mathcal{C}}(f)$ von f enthält genau $s_i := |\mathcal{B}_i|$ viele Jordanblöcke $J_i(0)$, für $i = 1, \ldots, d$. Nach Konstruktion $(**)$ ist

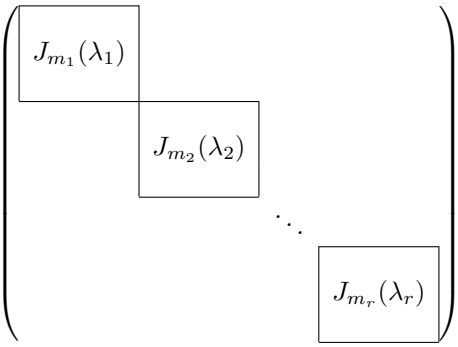
$$\dim(L_i) \;=\; \dim(U_i/U_{i-1}) \;=\; s_d + s_{d-1} + \cdots + s_i$$

für $i = 1, \ldots, d$. Wir können damit die s_i durch die $\dim(U_j)$ ausdrücken, und zwar ergibt sich $s_i = \dim(L_i) - \dim(L_{i+1})$, also

$$s_i \;=\; \dim(U_i/U_{i-1}) - \dim(U_{i+1}/U_i) \;=\; 2\dim(U_i) - \dim(U_{i-1}) - \dim(U_{i+1})$$

für $i = 1, \ldots, d$. Insbesondere sind die Zahlen s_i durch f bestimmt. Damit ist Satz 4.18 vollständig bewiesen. □

4.19 Theorem. (Jordansche Normalform) *Sei $f \in \mathrm{End}(V)$ derart, daß das charakteristische Polynom χ_f in Linearfaktoren zerfällt. Dann gibt es eine Basis von V, bezüglich welcher f durch eine Matrix der Gestalt*



beschrieben wird, mit r, $m_1, \ldots, m_r \in \mathbb{N}$ und $\lambda_1, \ldots, \lambda_r \in K$. Die Folge der Paare

$$(m_1, \lambda_1),\; \ldots,\; (m_r, \lambda_r)$$

ist dabei bis auf Permutation eindeutig durch f bestimmt. Die obige Matrix heißt eine Jordansche Normalform *von f.*

Beweis. Sind $\mu_1, \ldots, \mu_k \in K$ die paarweise verschiedenen Eigenwerte von f, so haben wir die f-invariante Hauptraumzerlegung $V = H_1 \oplus \cdots \oplus H_k$, mit $H_i := \mathrm{Hau}(f, \mu_i)$ für $i = 1, \ldots, k$ (4.12). Für jedes $i = 1, \ldots, k$ können wir Satz 4.18 auf die Restriktion $f|_{H_i} \in \mathrm{End}(H_i)$ anwenden, nach 4.12(c). □

4.20 Zusatz. *Für $f \in \mathrm{End}(V)$ wie in Theorem 4.19 und $\lambda \in K$ sei*

$$r_i(\lambda) \;:=\; \mathrm{rk}\big((f - \lambda\,\mathrm{id})^i\big)$$

für $i \geq 0$. Dann ist für $m \geq 1$ die Anzahl der Jordankästchen $J_m(\lambda)$ gleich

$$s_m(\lambda) \;=\; r_{m-1}(\lambda) - 2r_m(\lambda) + r_{m+1}(\lambda).$$

Beweis. Siehe Beweis von Satz 4.18: Mit $U_i = \ker((f - \lambda\,\mathrm{id})^i)$ $(i \geq 0)$ ist

$$s_m(\lambda) \;=\; 2\dim(U_m) - \dim(U_{m-1}) - \dim(U_{m+1}).$$

Wegen $\dim(U_i) = n - r_i(\lambda)$ (Satz III.3.11, mit $n := \dim(V)$) folgt die Behauptung. □

4.21 Bemerkungen.

1. Für die Aussage in 4.19 ist das Zerfallen von χ_f in Linearfaktoren natürlich auch notwendig. Über $K = \mathbb{C}$, oder jedem anderen algebraisch abgeschlossenen Körper, ist diese Voraussetzung stets von selbst erfüllt.

2. An der Jordanschen Normalform 4.19 kann man das Minimalpolynom direkt ablesen, und kann genau verstehen, was der Unterschied zwischen Minimalpolynom und charakteristischem Polynom ist. Ebenso liest man sofort Basen der Eigenräume ab:

Sei $A \in \mathrm{M}_n(K)$ eine Matrix mit zerfallendem charakteristischem Polynom χ_A, und sei $\lambda \in K$ ein fester Eigenwert von A. Es seien $J_{m_1}(\lambda), \ldots, J_{m_r}(\lambda)$ mit $m_1 \geq m_2 \geq \cdots \geq m_r \geq 1$ die Jordankästchen zum Eigenwert λ in einer Jordanschen Normalform von A. Dann gilt (die einfachen Beweise als Übung):

- $\mu_a(A, \lambda) = m_1 + \cdots + m_r$;
- $\mu_g(A, \lambda) = r$;
- die jeweils ersten Basisvektoren zu den r Jordankästchen geben zusammen eine Basis von $\mathrm{Eig}(A, \lambda)$;
- $\mu(q_A, \lambda) = m_1$. (Grund: Das Minimalpolynom von $J_m(\lambda)$ ist $(t - \lambda)^m$.)

Für $A \in \mathrm{M}_n(K)$ mit zerfallendem charakteristischem Polynom gilt also: Genau dann ist $\chi_A = q_A$, wenn je zwei Jordankästchen von A zu verschiedenen Eigenwerten gehören, oder äquivalent, wenn $\mu_g(A, \lambda) \leq 1$ für jedes $\lambda \in K$ ist.

3. Ist $A \in \mathrm{M}_n(K)$ eine Matrix mit zerfallendem charakteristischem Polynom, und will man eine Jordansche Normalform von A bestimmen, so hat man folgendes zu tun:

1. Zerlege χ_A in Linearfaktoren;
2. bestimme für jeden Eigenwert λ von A die Ränge $r_i(\lambda) = \mathrm{rk}\big((A - \lambda\, I_n)^i\big)$ für $i = 1, 2, \ldots, \mu_a(A, \lambda)$. (Es genügt, dies im Fall $\mu_a(A, \lambda) \geq 2$ zu tun.)

Daraus lassen sich dann die Größen der Jordankästchen ablesen, siehe Zusatz 4.20. Will man auch eine zugehörige Transformationsmatrix bestimmen, muß man Basen der Haupträume bestimmen und dabei wie im Beweis von Satz 4.18 vorgehen. Alle diese Rechnungen können algorithmisch durchgeführt werden, da die Beweise konstruktiv waren.

4.22 Beispiel. Für ein konkretes Beispiel betrachte die Matrix

$$A \;=\; \begin{pmatrix} 2 & 1 & 3 & 1 \\ 1 & 2 & -1 & 1 \\ 0 & 0 & 1 & 0 \\ -2 & -2 & 1 & -1 \end{pmatrix} \;\in\; \mathrm{M}_4(K).$$

Man findet

$$\chi_A \;=\; (t-1)\begin{vmatrix} t-2 & -1 & -1 \\ -1 & t-2 & -1 \\ 2 & 2 & t+1 \end{vmatrix} \;=\; (t-1)^4$$

und

$$A - I \;=\; \begin{pmatrix} 1 & 1 & 3 & 1 \\ 1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ -2 & -2 & 1 & -2 \end{pmatrix},$$

also $\operatorname{rk}(A-I)=2$ (Spalten!), und

$$(A-I)^2 \;=\; \begin{pmatrix} 0 & 0 & 3 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -6 & 0 \end{pmatrix},$$

also $\operatorname{rk}((A-I)^2)=1$, und somit $(A-I)^3=0$. Daraus folgt schon, daß die Jordansche Normalform von A gleich

$$B \;=\; \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \;=\; \Big(J_3(1),\; J_1(1)\Big)$$

ist, z.B. mit 4.20. Das Minimalpolynom ist also $q_A=(t-1)^3$.

Um eine konkrete Transformationsmatrix $S\in \mathrm{GL}_4(K)$ mit $S^{-1}AS=B$ zu finden, gehe vor wie im Beweis von 4.18: Sei $g:=F_{A-I}$, sei $U_i=\ker(g^i)$ $(i\geq 0)$. Es ist

$$U_1 \;=\; \operatorname{span}(e_1-e_2,\; e_2-e_4), \quad U_2 \;=\; \operatorname{span}(e_1,\, e_2,\, e_4), \quad U_3 \;=\; K^4,$$

also $d=3$. Kann nehmen $L_3=\operatorname{span}(v)$ mit $v=e_3$, also $\mathcal{B}_3=(v)$. Dann ist

$$\mathcal{C}_v \;=\; (g^2(v),\, g(v),\, v) \;=\; \Big(3e_1+3e_2-6e_4,\; 3e_1-e_2+e_4,\; e_3\Big),$$

was ein Kästchen $J_3(1)$ gibt. Es bleibt noch ein von $g^2(v)$ linear unabhängiger Vektor in U_1 zu finden, man kann etwa $w=e_1-e_2$ nehmen. Bezüglich der Basis $\mathcal{C}=(g^2(v),\, g(v),\, v,\, w)$ von K^4 hat dann F_A die Matrix B, d.h. es gilt $S^{-1}AS=B$ mit

$$S \;=\; M\Big((A-I)^2v,\, (A-I)v,\, v,\, w\Big) \;=\; \begin{pmatrix} 3 & 3 & 0 & 1 \\ 3 & -1 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ -6 & 4 & 0 & 0 \end{pmatrix}.$$

Unter der Voraussetzung, daß das charakteristische Polynom zerfällt, haben wir auch die Frage nach der Ähnlichkeit von Matrizen gelöst. Insbesondere haben wir es im Fall $K=\mathbb{C}$ für alle Matrizen gelöst:

4.23 Korollar. *Seien A, $B\in \mathrm{M}_n(K)$ zwei Matrizen mit zerfallendem charakteristischem Polynom. Es sind äquivalent:*

(i) *$A\approx B$;*
(ii) *bis auf Permutation der Jordanblöcke haben A und B dieselbe Jordansche Normalform;*
(iii) *für alle $\lambda\in K$ und alle $i=1,\ldots,n$ ist $\operatorname{rk}\big((A-\lambda I_n)^i\big)=\operatorname{rk}\big((B-\lambda I_n)^i\big)$.*

Beweis. $A\approx B$ bedeutet, daß A und B denselben Endomorphismus von K^n bezüglich zweier geeigneter Basen beschreiben (Satz IV.4.1). Die Implikation (i) $\Rightarrow$ (ii) folgt also aus Theorem 4.19, und (ii) $\Rightarrow$ (i) ist klar. Die Äquivalenz von (ii) und (iii) folgt daraus, daß die Jordanblöcke der Jordanschen Normalform von A durch die $r_i(\lambda,A)=\operatorname{rk}((A-\lambda\,\mathrm{id})^i)$ ausgedrückt werden können (siehe 4.20). □

4.24 Bemerkung. Seien $n\in\mathbb{N}$. Die Ähnlichkeitsklassen von nilpotenten Matrizen in $\mathrm{M}_n(K)$ stehen in Bijektion zur Menge aller Tupel $(m_1,\ldots,m_r)$ mit r,

$m_1, \ldots, m_r \in \mathbb{N}$, $m_1 \geq \cdots \geq m_r$ und $m_1 + \cdots + m_r = n$. Ein solches Tupel heißt eine *Partition* der Zahl n. Die Zahl $p(n)$ der Partitionen von n hat viele bemerkenswerte Eigenschaften. Kleine Werte von n:

n	$p(n)$	Partitionen von n
1	1	(1)
2	2	$(2),\ (1,1)$
3	3	$(3),\ (2,1),\ (1,1,1)$
4	5	$(4),\ (3,1),\ (2,2),\ (2,1,1),\ (1,1,1,1)$

usw. Es ist weiter $p(5) = 7$, $p(6) = 11$, $p(7) = 15$, $p(8) = 22$, $\ldots$, $p(100) = 190569292$ usw.

Im Sommersemester werden wir einen anderen Beweis der Jordanschen Normalform sehen, der konzeptionell viel einfacher ist, aber mehr Theorie erfordert (Struktur von endlich erzeugten $K[t]$-Moduln).