Scriptum zur Vorlesung
Lineare Algebra I

Prof. W. Hoffmann
Universität Bielefeld
SS 2021

Index

Einleitung

Die lineare Algebra entstand aus zwei Quellen. Die eine ist die analytische Geometrie, die (im Unterschied zur axiomatischen Geometrie) Punkte durch Koordinaten beschreibt. Die andere Quelle sind die Lösungsverfahren für Systeme linearer Gleichungen, die (im Unterschied zu allgemeinen algebraischen Gleichungen) eine vollständige Beschreibung der Lösungsmenge erlauben. Angesichts eines solch schmalen Themengebietes mag es verwundern, dass die lineare Algebra zusammen mit der Analysis die fachliche Basis eines Mathematikstudiums bildet. Dies erklärt sich zum Einen daraus, dass viele allgemeine Begriffsbildungen der Algebra bereits im Teilgebiet der linearen Algebra als Spezialfälle in einer intuitiv fassbaren Form auftauchen. Zum anderen liegt es daran, dass allgemeine Probleme der Analysis, etwa das Lösen nichtlinearer Gleichungen, durch den Begriff der Ableitung näherungsweise auf lineare Probleme zurückgeführt werden.

1 Grundbegriffe der Logik

Die Mathematik stellt viel höhere Anforderungen an die Genaugkeit von Formulierungen, als es im Alltag üblich ist. Darum sollten wir uns zunächst mit den Anfangsgründen der Logik befassen.

1.1 Aussagen

Es werden nur Aussagen betrachtet, die einen der beiden Wahrheitswerte „wahr“ oder „falsch“, abgekürzt w oder f, haben. Aus gegebenen Aussagen kann man durch Verknüpfung neue Aussagen bilden. So ist die *Negation*[1] einer Aussage A, abgekürzt $\neg A$ und gelesen „nicht A“, gegeben durch die Wahrheitstafel

A	$\neg A$
w	f
f	w

Die Verneinung der Aussage

„Du hast immer Zeit für mich“

ist die Aussage

„Du hast nicht immer Zeit für mich“.

[1] auch Verneinung genannt

Verknüpfungen von zwei Aussagen A und B sind beispielsweise die *Konjunktion*[2] $A \wedge B$, gelesen „A und B“, sowie die *Disjunktion*[3] $A \vee B$, gelesen „A oder B“, die gegeben sind durch die Wahrheitstafeln

A	B	$A \wedge B$
w	w	w
w	f	f
f	w	f
f	f	f

A	B	$A \vee B$
w	w	w
w	f	w
f	w	w
f	f	f

Beispiele sind die Aussagen

> „Der Beschuldigte hatte ein Motiv und die Gelegenheit für die Tat.“
> „Dieser Neureiche hat eine Erbschaft gemacht oder im Lotto gewonnen.“

Man beachte, dass in der Logik das Wort „oder“ im einschließenden Sinne gebraucht wird.

Die *Implikation*[4] ist die Aussage $A \Rightarrow B$, gelesen „wenn A, dann B“, die gegeben ist durch

A	B	$A \Rightarrow B$
w	w	w
w	f	f
f	w	w
f	f	w

Sie drückt keine Ursache-Wirkung-Beziehung aus, z. B.:

> „Wenn du diese Aufgabe lösen kannst, dann bist du ein Genie.“

Schließlich haben wir noch die *Äquivalenz* $A \Leftrightarrow B$, gelesen „genau dann A, wenn B“, und die *Antivalenz* $A \not\Leftrightarrow B$, gelesen „entweder A oder B“, mit den Wahrheitstafeln

A	B	$A \Leftrightarrow B$
w	w	w
w	f	f
f	w	f
f	f	w

A	B	$A \not\Leftrightarrow B$
w	w	f
w	f	w
f	w	w
f	f	f

[2] auch Und-Verknüpfung genannt
[3] auch Oder-Verknüpfung genannt
[4] auch Subjunktion genannt

Im täglichen Leben werden diese oft durch die Worte „wenn“ bzw. „oder“ ausgedrückt, z. B.

> „Wenn du mir dein Fahrrad leihst, kannst du mit meinem Ball spielen.“
> „Du gibst mir jetzt den Ball zurück oder ich sag' es Mutti.“

Ein logisches Gesetz ist eine Verknüpfung von Variablen, die bei jeder Belegung der Variablen mit Aussagen zu einer wahren Aussage wird, z. B.

$$A \Rightarrow A \vee B, \qquad A \wedge B \Rightarrow B.$$

Viele logische Gesetze haben die Struktur einer Äquivalenz, z. B.

$$\neg\neg A \Leftrightarrow A, \qquad (A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A),$$
$$(A \Leftrightarrow B) \Leftrightarrow ((A \Rightarrow B) \wedge (B \Rightarrow A)).$$

(Die Aussage $\neg B \Rightarrow \neg A$ nennt man übrigens die *Kontraposition* der Implikation $A \Rightarrow B$). Weitere Beispiele sind die *Kommutativgesetze*

$$A \wedge B \Leftrightarrow B \wedge A, \qquad A \vee B \Leftrightarrow B \vee A,$$

die *Assoziativgesetze*

$$(A \wedge B) \wedge C \Leftrightarrow A \wedge (B \wedge C), \qquad (A \vee B) \vee C \Leftrightarrow A \vee (B \vee C),$$

die *Distributivgesetze*

$$A \vee (B \wedge C) \Leftrightarrow (A \vee B) \wedge (A \vee C), \quad A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$$

sowie die *de Morganschen Gesetze*

$$\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B, \qquad \neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B.$$

Um ein logisches Gesetz zu beweisen, müssen wir nachprüfen, dass es für alle möglichen Wahrheitswerte der vorkommenden Variablen wahr ist. Dabei sind die Wahrheitswerte von verschachtelten Verknüpfungen schrittweise zu bestimmen, z. B.

A	B	$A \Rightarrow B$	$B \Rightarrow A$	$(A \Rightarrow B) \wedge (B \Rightarrow A)$
w	w	w	w	w
w	f	f	w	f
f	w	w	f	f
f	f	w	w	w

Durch Vergleich mit der Wahrheitstafel für die Äquivalenz $A \Leftrightarrow B$ finden wir, dass diese zu der Aussage

$$(A \Rightarrow B) \wedge (B \Rightarrow A)$$

äquivalent ist.

1.2 Prädikate

Der Wahrheitswert des Satzes „Ich bin ein Berliner“ hängt von der Bedeutung des Wortes „ich“ ab. Ein *Prädikat* (auch *Aussageform* genannt) enthält Variablen und wird erst zu einer Aussage, wenn man diese durch Objekte ersetzt. Wird es zu einer wahren Aussage, so sagt man, dass es für diese Objekte *gilt*. Aussageformen mit einer Variablen nennt man auch Eigenschaften, z. B. „x ist ehrlich“, solche mit zwei Variablen nennt man *Relationen*, z. B. „x ist der Vater von y“. Wir vereinbaren manchmal Abkürzungen für Prädikate in der Form $P(x)$ oder $Q(x, y)$ usw.

Eine *Menge* ist, naiv gesprochen, eine Zusammenfassungen von Objekten. Manchmal kann man ihre Elemente aufzählen, wobei man sie in geschweifte Klammern einschließt und die Reihenfolge gleichgültig ist, also

$$\{\text{Sonne, Erde, Mond}\} = \{\text{Erde, Mond, Sonne}\}.$$

Wenn ein Objekt zu einer Menge gehört, nennt man es ein *Element* dieser Menge. Die Relation „x ist ein Element von M“ wird durch $x \in M$ abgekürzt. Sie wird zu einer wahren Aussage, wenn wir z. B. x durch J. F. Kennedy und M durch die Menge der Menschen ersetzen.

Manche Prädikate ergeben nur dann einen Sinn, wenn man die Variablen durch Elemente von bestimmten Mengen ersetzt. Wenn ein Prädikat $P(x)$ für alle Elemente einer Menge M gilt, so schreibt man

$$\forall\, x \in M\ P(x),$$

gelesen „für alle x in M (gilt) $P(x)$“. Wenn wenigstens ein Element der Menge M existiert, für das das Prädikat $P(x)$ gilt, so schreibt man

$$\exists\, x \in M\ P(x),$$

gelesen „(es) existiert ein x in M, so dass $P(x)$ (gilt)“. Manche Autoren schließen hier das Prädikat in Klammern ein. Die Zeichen $\forall$ und $\exists$ nennt man *Quantoren*.

Im Fall einer endlichen Menge kann man solche Aussagen auch ohne Quantoren formulieren, z. B.

$$\begin{aligned} \forall\, x \in \{\mathrm{a}, \mathrm{b}\}\ P(x) &\iff P(\mathrm{a}) \wedge P(\mathrm{b}), \\ \exists\, x \in \{\mathrm{a}, \mathrm{b}\}\ P(x) &\iff P(\mathrm{a}) \vee P(\mathrm{b}). \end{aligned}$$

Dies zeigt, dass die Gesetze

$$\begin{aligned} \neg\, \forall\, x \in M\ P(x) &\iff \exists\, x \in M\ \neg P(x), \\ \neg\, \exists\, x \in M\ P(x) &\iff \forall\, x \in M\ \neg P(x) \end{aligned}$$

Verallgemeinerungen der de Morganschen Gesetze sind. Gesetze der Prädikatenlogik kann man nicht mehr durch Wahrheitstafeln nachprüfen. Man legt einige als Axiome fest und gewinnt weitere durch sogenannte Schlussregeln, worauf wir aber nicht eingehen werden.

Beispiel 1.1. Betrachten wir das Schachspiel. Das Ziel des Spiels besteht eigentlich darin, den gegnerischen König zu schlagen. Wohl wegen der Etikette am persischen Hof verbot man Züge, bei denen der Antwortzug den König schlagen könnte.

Definition. *Ein am Zug befindlicher Spieler ist schachmatt, wenn es für jeden möglichen Zug und auch beim Auslassen des Zuges einen Antwortzug gibt, der seinen König schlagen würde.*

Wir erinnern daran, dass in einer *Definition* ein neuer Begriff erklärt wird, indem man ihn mit einem bereits bekannten Begriff gleichsetzt. Im vorliegenden Fall ist der bekannte Begriff eine Aussage mit der Struktur

$$\forall\, x \in M\ \exists\, y \in N(x)\ Q(x, y).$$

Man beachte, dass die Menge $N(x)$ der erlaubten Antwortzüge vom Zug x abhängt.

Betrachten wir zum Beispiel eine Schachaufgabe von W. Meredith aus dem Jahr 1886 (siehe Abbildung 1).

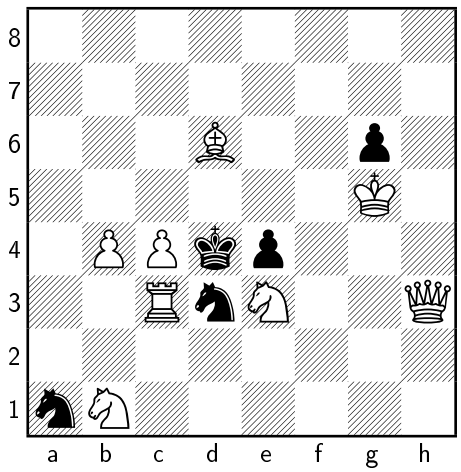


Abbildung 1: Weiß gewinnt in zwei Zügen

Nach der Konvention ist in solchen Aufgaben Weiß am Zug. Die Aufgabe besteht also darin, folgende Aussage zu beweisen:

Es gibt einen Zug von Weiß, so dass es für jeden Antwortzug von Schwarz wiederum einen Zug von Weiß gibt, so dass Schwarz matt ist.

Dies wird eine freiwillige Zusatzaufgabe auf dem ersten Übungszettel sein. Wir wollen hier nur die einfachere Aufgabe betrachten, dass Weiß in einem Zug gewinnt, wenn Schwarz am Zug ist.

In diesem Fall können die schwarzen Bauern nicht vorrücken, da weiße Figuren im Weg stehen, und wo sie schlagen könnten, stehen keine weißen Figuren. Würde der schwarze König auf ein freien Feld ziehen, so würde ihm von weißen Bauern oder dem weißen Läufer Schach geboten. Würde er den weißen Bauern, Turm oder Springer schlagen, so würde ihm vom weißen Turm, dem anderen weißen Springer bzw. der weißen Dame Schach geboten.

Schwarz kann also nur einen seiner Springer ziehen. Zieht er den vom Feld a1, so ist das Feld c2 nicht mehr gedeckt, und Weiß kann seinen Springer von e3 dorthin ziehen und dabei unter Umständen den schwarzen Springer schlagen. Auf jeden Fall wird er Schach bieten, und das freigewordene Feld e3 ist kein Ausweg für den schwarzen König, da es von der weißen Dame gedeckt ist.

Wenn Schwarz hingegen seinen Springer von d3 zieht, so kann der weiße Turm jetzt den Schutz des Springers auf e3 übernehmen. Wenn der schwarze Springer nicht auf e5 gezogen wurde, kann Weiß seine Dame nach h8 ziehen und Schach bieten. Das freigewordene Feld d3 ist kein Ausweg für den König, weil es vom weißen Turm gedeckt ist. Steht hingegen der schwarze Springer auf e5, so versperrt er dem König die Flucht auf dieses Feld, und der freigesetzte weiße Läufer kann nach c5 gezogen werden, wo er Schach bietet und von einem Bauern geschützt ist. ◁

Zwischen Mengen und Prädikaten besteht ein enger Zusammenhang. Ist ein Prädikat $P(x)$ für beliebige Objekte x sinnvoll, so bezeichnet man die Menge aller Objekte, für die $P(x)$ gilt, mit

$$\{x \mid P(x)\}.$$

Umgekehrt ist für jede Menge M das Prädikat „$x \in M$“ für beliebige Objekte x sinnvoll. Ist ein Prädikat $P(x)$ zumindest für die Elemente einer gegebenen Menge N sinnvoll, so wird durch

$$\{x \in N \mid P(x)\}$$

eine Teilmenge von N definiert.

Wenn zwei Prädikate für alle Objekte äquivalent sind, dann sind die zugehörigen Mengen gleich. Also gilt für Mengen M und N genau dann $M = N$, wenn für alle Objekte x gilt $x \in M \Leftrightarrow x \in N$.

Da Mengen auch wieder Objekte sind, führt dieser naive Ansatz leider schnell zu Widersprüchen. Am bekanntesten ist die *Russellsche Antinomie*: Wenn die Menge

$$\{x \mid x \notin x\}$$

Element ihrer selbst ist, dann ist sie es nach ihrer Definition nicht, und umgekehrt. Der Ausweg besteht in der axiomatischen Mengenlehre, die aber den Rahmen dieser kurzen Einführung sprengen würde.

2 Lineare Gleichungssysteme

Wie bereits in der Einleitung erwähnt, ist das Lösen linearer Gleichungssysteme einer der Ursprünge der linearen Algebra. Solche Gleichungssysteme treten in vielen Situationen auf.

2.1 Beispiele

Beispiel 2.1. Über die Kinder Xaver, Yvonne und Zoe einer Familie sei folgendes bekannt.

- Yvonne hat das Durchschnittsalter von Xaver und Zoe.
- Zoe ist so alt wie Xaver und Yvonne zusammen.
- Vor zwei Jahren war das Gesamtalter der Kinder drei Mal so groß wie das jetzige Alter von Xaver.

Bezeichnen wir die Alter in Jahren mit x, y und z, so bedeutet die erste Aussage

$$y = \frac{x+z}{2}$$

und die dritte Aussage

$$(x-2)+(y-2)+(z-2) = 3x.$$

Die drei Aussagen lassen sich also in dem Gleichungssystem

$$\begin{cases} x+z = 2y \\ x+y = z \\ y+z = 2x+6 \end{cases}$$

zusammenfassen.

Die zweite Gleichung besagt, dass wir die Variable z durch den Ausdruck $x+y$ substituieren dürfen. Damit können wir z aus den anderen Gleichungen eliminieren und erhalten das Gleichungssystem

$$\begin{cases} 2x = y \\ 2y = x+6 \end{cases}$$

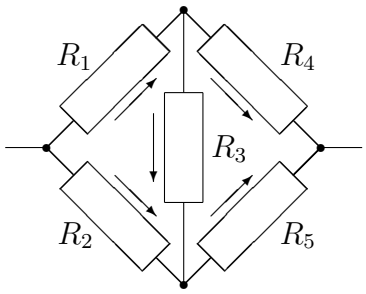
Eliminieren wir hier mit Hilfe der ersten Gleichung die Variable y aus der zweiten, so erhalten wir

$$3x = 6.$$

Somit ist $x = 2$. Mit der letzten Substitutionsgleichung $y = 2x$ erhalten wir $y = 4$, und mit der ersten Substitutionsgleichung $z = x + y$ folgt $z = 6$.

Allgemein gilt: Bei jeder Elimination ist das alte Gleichungssystem äquivalent zum neuen Gleichungssystem ergänzt um die Substitutionsgleichung, denn wir können durch Rücksubstitution das alte System zurückgewinnen. ◁

Beispiel 2.2. Wir betrachten folgende Schaltung, die nur aus Widerständen besteht:



Es sei I_k die Stärke des Stroms durch den Widerstand R_k, wobei das Vorzeichen positiv oder negativ ist, je nachdem ob der Strom in Pfeilrichtung oder dagegen fließt. Weiter sei U_k die Spannung zwischen den Kontakten von R_k mit der analogen Vorzeichenregel. Nach den Kirchhoffschen Regeln verschwindet die Summe der Stromstärken in jedem Knotenpunkt wie auch die Summe der Spannungen entlang jeder Masche, also

$$\begin{aligned} I_1 &= I_3 + I_4, & U_1 + U_3 &= U_2, \\ I_2 + I_3 &= I_5, & U_4 &= U_3 + U_5. \end{aligned}$$

Nach dem Ohmschen Gesetz gilt

$$U_k = R_k I_k.$$

Damit können wir die Spannungen durch die Stromstärken ausdrücken und erhalten ein lineares Gleichungssystem aus vier Gleichungen mit fünf Unbekannten $I_1, \ldots, I_5$, das wir aus Platzgründen nicht noch einmal hinschreiben.

Wir benutzen nun die linken Gleichungen, um I_4 und I_5 zu eliminieren. Die letzte Gleichung wird zu

$$R_4(I_1 - I_3) = R_3 I_3 + R_5(I_2 + I_3),$$

und es verbleibt das Gleichungssystem mit drei Unbekannten

$$\begin{cases} -R_1 I_1 + R_2 I_2 = R_3 I_3 \\ R_4 I_1 - R_5 I_2 = R_r I_3 \end{cases}$$

wobei wir die Bezeichnungen

$$R_l = R_1 + R_2 + R_3, \qquad R_r = R_3 + R_4 + R_5$$

für die Summe der Widerstände der linken bzw. rechten Masche einführen.

Ist $R_3 \neq 0$, so können wir die erste Gleichung nach I_3 auflösen und in die zweite einsetzen:

$$R_4 I_1 - R_5 I_2 = R_r \frac{R_2 I_2 - R_1 I_1}{R_3}.$$

Durch Multiplikation beider Seiten mit R_3 erhalten wir

$$R_3(R_4 I_1 - R_5 I_2) = R_r(R_2 I_2 - R_1 I_1),$$

was offenbar auch im Fall $R_3 = 0$ gilt. Man hätte dieses Ergebnis auch ohne Fallunterscheidung gewinnen können, indem man zunächst die zweite Gleichung mit R_3 multipliziert und dann die erste eingesetzt hätte. *So sollte man grundsätzlich verfahren.* Fassen wir wieder gleiche Variablen zusammen, so ergibt sich

$$(R_1 R_r + R_3 R_4) I_1 = (R_2 R_r + R_3 R_5) I_2.$$

Da nur noch eine Gleichung vorliegt, ist keine weitere Elimination möglich. Dies deckt sich mit der Erfahrung, dass es nicht nur eine Lösung gibt.

An Stelle von I_3 können wir aus unserem Gleichungssystem auch I_2 eliminieren. Ähnlich wie oben sollten wir dazu wir die zweite Gleichung mit R_2 multiplizieren:

$$R_2 R_4 I_1 - R_2 R_5 I_2 = R_2 R_r I_3$$

Anstatt die erste Gleichung nach $R_2 I_2$ aufzulösen und das Ergebnis hier einzusetzen, ist es geschickter, die erste Gleichung mit R_5 zu multiplizieren:

$$-R_1 R_5 I_1 + R_2 R_5 I_2 = R_3 R_5 I_3$$

Wenn man nun jeweils die linken und rechten Seiten addiert, so heben sich die Terme mit I_2 auf, und man erhält

$$(R_2 R_4 - R_1 R_5) I_1 = (R_2 R_r + R_3 R_5) I_3.$$

Bemerkung: Der Brückenstrom I_3 verschwindet, wenn $R_2 R_4 = R_1 R_5$ ist, was man im Fall $R_4 \neq 0$ und $R_5 \neq 0$ auch in der Form

$$\frac{R_1}{R_4} = \frac{R_2}{R_5}$$

schreiben kann. Dies liegt der Wheatstoneschen Messbrücke zu Grunde.

Nun betrachten wir Gesamtstrom und Gesamtspannung

$$I = I_1 + I_2, \qquad U = U_1 + U_4.$$

Wir wollen beide durch eine einzige Größe ausdrücken, beispielsweise durch I_1. Dazu multiplizieren wir die linke Gleichung mit $R_2R_r + R_3R_5$ und benutzen die bereits gewonnene Beziehung zwischen I_2 und I_1:

$$(R_2R_r + R_3R_5)I = (R_2R_r + R_3R_5)I_1 + (R_1R_r + R_3R_4)I_1.$$

Die rechte Seite kann man übrigens auch in der Form $(R_lR_r - R_3^2)I_1$ schreiben.

In der Formel für U kommt $U_4 = R_4I_4$ vor. Die Variable I_4 hatten wir eliminiert. Um auch sie durch I_1 auszudrücken, multiplizieren wir in der Substitutionsgleichung $I_4 = I_1 - I_3$ beide Seiten mit $R_2R_r + R_3R_5$ und benutzen die bereits gewonnene Beziehung zwischen I_3 und I_1:

$$(R_2R_r + R_3R_5)I_4 = (R_2R_r + R_3R_5)I_1 - (R_2R_4 - R_1R_5)I_1.$$

Schreiben wir R_r aus und fassen neu zusammen, so vereinfacht sich die rechte Seite:

$$(R_2R_r + R_3R_5)I_4 = (R_lR_5 + R_2R_3)I_1.$$

Da wir an U_4 interessiert sind, multiplizieren wir hier beide Seiten mit R_4 und erhalten

$$(R_2R_r + R_3R_5)U_4 = (R_lR_5 + R_2R_3)R_4I_1.$$

Multiplizieren wir beide Seiten der Gleichung $U_1 = R_1I_1$ mit dem Koeffizienten von U_4 in der letzten Gleichung, so folgt nach Addition beider Gleichungen

$$(R_2R_r + R_3R_5)U = (R_2R_r + R_3R_5)R_1I_1 + (R_lR_5 + R_2R_3)R_4I_1.$$

Wir haben somit U und I durch I_1 ausgedrückt, wobei beide Variablen mit dem selben Koeffizienten $R_2R_r + R_3R_5$ auftreten. Bilden wir auf beiden Seiten die Quotienten, so ergibt sich für den Gesamtwiderstand $R = U/I$ die Formel

$$R = \frac{R_1R_2R_r + R_lR_4R_5 + R_1R_3R_5 + R_2R_3R_4}{R_lR_r - R_3^2},$$

vorausgesetzt, die Größen $R_2R_r + R_3R_5$ und $R_lR_r - R_3^2$ sind von Null verschieden. Ist der Nenner $R_lR_r - R_3^2$ gleich Null, so verschwinden alle Widerstände entlang einer Masche (Übungsaufgabe). In diesem Fall sind die Ströme nicht eindeutig bestimmt, denn man kann einen beliebigen Kreisstrom entlang dieser Masche addieren, ohne die Gleichungen zu verletzen. Die Schaltung vereinfacht sich dann aber, und R lässt sich sogar leichter bestimmen. Gilt $R_2R_r + R_3R_5 = 0$, ohne dass der Nenner verschwindet, so verschwindet R_2 und einer der Widerstände R_3 oder R_5. Durch Vertauschung der Bezeichnungen sieht man, dass die Formel auch dann das richtige Ergebnis liefert. ◁

2.2 Das Gaußsche Eliminationsverfahren

Wir betrachten Gleichungen mit mehreren Unbekannten, auch Variablen genannt. Dabei beschränken wir uns auf lineare Gleichungen, das heißt solche, bei denen die Terme auf beiden Seiten als Summen von Vielfachen der Variablen und von Konstanten geschrieben werden können, wie zum Beispiel

$$5x + 3y - 7z - 1 = 2y + z + 6.$$

Die Vorfaktoren vor den Variablen nennt man *Koeffizienten*. Indem man auf beiden Seiten der Gleichung gleiche Terme addiert, gelangt man zu einer äquivalenten Gleichung, in der alle Terme mit Variablen auf die linke Seite stehen und alle Konstanten auf die rechte Seite, in unserem Beispiel also

$$5x + y - 8z = 7.$$

Ein *Gleichungssystem* besteht aus mehreren Gleichungen. Eine *Lösung* eines Gleichungssystems mit n Variablen besteht aus n Zahlen mit der Eigenschaft, dass beim Einsetzen alle Gleichungen erfüllt sind. Dabei kommt es natürlich auf die Reihenfolge der Zahlen an. Eine Lösung ist im Fall von zwei Variablen ein *geordnetes Paar*, im Fall von drei Variablen ein *Tripel*, im Fall von vier Variablen ein *Quadrupel* usw. Aus den weiteren lateinischen Wörtern Quintupel, Sextupel usw. bildete man den allgemeinen Begriff *n-Tupel*. Um bei der Aufzählung mehrerer Lösungen deutlich zu machen, welche Zahlen jeweils zu einem n-Tupel gehören, schließt man diese in Klammern ein. Innerhalb eines n-Tupels werden die Zahlen durch Kommata getrennt. In Beispiel 2.1 war das Tripel $(2, 4, 6)$ die einzige Lösung. Beim Vorkommen von Dezimalbrüchen benutzt man in deutschen Texten Semikola als Trennzeichen.

Nun beschreiben wir das *Eliminationsverfahren* (das oft nach Gauß benannt wird, obwohl es bereits vor ihm bekannt war) systematisch. Im Unterschied zu unserem bisherigen Vorgehen eliminiert man die Variablen in der allgemeinen Theorie meist in aufsteigender Reihenfolge. Wenn eine Variable nicht vorkommt, kann man natürlich gleich zur nächsten übergehen. Wir suchen also die erste Variable, die überhaupt im System vorkommt, wählen dann eine Gleichung, in der sie vorkommt, als *Substitutionsgleichung*, und addieren jeweils ein Vielfaches dieser Gleichung zu jeder anderen Gleichung, um dort die besagte Variable (die man in diesem Zusammenhang *Eliminationsvariable* nennt) zu eliminieren. Anstatt die Substitutionsgleichung getrennt aufzubewahren, belassen wir sie im Gleichungssystem, aber als erste Gleichung, damit das verbleibende Teilsystem, das eine Gleichung weniger besitzt, übersichtlich zusammengefasst ist. Auf dieses Teilsystem wird dann

das selbe Verfahren angewendet. Da das verbleibende System mit jedem Mal weniger Gleichungen besitzt, muss das Verfahren irgendwann zum Abschluss kommen.

Wir kommen dabei mit den folgenden zwei Arten von Operationen aus: Der Addition des Vielfachen einer Gleichung zu einer anderen und der Vertauschung von Gleichungen. Jede Lösung des ursprünglichen Systems ist offensichtlich auch eine des entstehenden Systems. Es gilt aber auch die Umkehrung, denn beide Operationen lassen sich mit der selben Art von Operationen rückgängig machen. Das entstehende System ist somit äquivalent zum ursprünglichen System.

Beispiel 2.3. Wir betrachten das Gleichungssystem

$$\begin{cases} v+2w- & x+3y- & z=2 \\ v+3w-3x+4y+ & & z=5 \\ 2v+3w & +5y-3z=3 \\ 3v+4w+ & x+6y-2z=4 \end{cases}$$

Da die erste Variable v in der ersten Gleichung vorkommt, können wir mit deren Hilfe diese Variable aus den anderen Gleichungen eliminieren. Dazu subtrahieren wir die erste Gleichung von der zweiten, ihr zweifaches von der dritten und ihr dreifaches von der vierten:

$$\begin{cases} v+2w- & x+3y- & z= & 2 \\ & w-2x+ & y+2z= & 3 \\ - & w+2x- & y- & z=-1 \\ & -2w+4x-3y+ & & z=-2 \end{cases}$$

Die erste Variable, die in dem Teilsystem vorkommt, das aus den letzten drei Gleichungen besteht, ist w, und wir können sie mit Hilfe der zweiten Gleichung aus den restlichen Gleichungen eliminieren, indem wir die zweite Gleichung zur dritten und ihr zweifaches zur vierten Gleichung addieren:

$$\begin{cases} v+2w- & x+3y- & z=2 \\ & w-2x+ & y+2z=3 \\ & & z=2 \\ & - & y+5z=4 \end{cases}$$

Die erste Variable, die in dem Teilsystem vorkommt, das aus den letzten beiden Gleichungen besteht, ist y. Sie kommt in der letzten Gleichung vor, die somit als Substitutionsgleichung zu verwenden und in diesem Teilsystem als oberste Zeile zu schreiben ist. In der verbleibenden Gleichung kommt y nicht mehr vor, also erübrigt sich die Elimination. Das Ergebnis des Verfahrens ist

das System

$$\begin{cases} v+2w- \;\; x+3y- \;\; z=2 \\ \quad\; w-2x+ \;\; y+2z=3 \\ \qquad\quad - \;\; y+5z=4 \\ \qquad\qquad\qquad\quad z=2 \end{cases}$$

das sich von dem vorigen nur durch die Vertauschung der letzten beiden Zeilen unterscheidet. ◁

Man macht sich leicht klar, dass jedes Gleichungssystem, das im Ergebnis des Gaußschen Eliminationsverfahrens entsteht, eine *Stufenform* hat, d. h., die Nummer der ersten Variablen, die in jeder Zeile vorkommt, nimmt von oben nach unten zu. Dies sind genau die Variablen, die im Verfahren substituiert wurden. Am Schluss stehen die Gleichungen, die gar keine Variablen enthalten, wo also auf der linken Seite nur eine Null steht. Wenn dort irgendwo auf der rechten Seite eine von Null verschiedene Zahl steht, so hat das System keine Lösung.

Andernfalls kann man die Gleichungen ohne Variablen ignorieren. Die Variablen, die in keiner Zeile an erster Stelle stehen, nennt man *freie Variablen*, denn man kann sie frei wählen und erhält jedes Mal eine Lösung, indem man die übrigen Variablen, beginnend mit der letzten, eine nach der anderen bestimmt. Man nennt sie darum *abhängige Variablen*. Für die freien Variablen führt man häufig neue Variablennamen ein, genannt *Parameter*. Man erhält eine Formel, die beim Einsetzen verschiedener Parameterwerte sämtliche Lösungen liefert und darum *allgemeine Lösung* genannt wird. Eine einzelne Lösung nennt man in diesem Zusammenhang auch *spezielle* oder *partikuläre Lösung*.

Beispiel 2.4. Kehren wir zum obigen Gleichungssystem zurück. Die einzige freie Variable ist x, die wir als Parameter t bezeichnen. Die letzte Gleichung des Systems in Stufenform liefert $z = 2$, aus der vorherigen erhalten wir

$$y = 5z - 4 = 6,$$

dann aus der vorangehenden Gleichung

$$w = 3 + 2x - y - 2z = 3 + 2t - 6 - 2 \cdot 2 = -7 + 2t$$

und schließlich aus der ersten Gleichung

$$v = 2 - 2w + x - 3y + z = 2 - 2(-7 + 2t) + t - 3 \cdot 6 + 2 = -3t.$$

Die allgemeine Lösung ist also das Quintupel

$$(-3t, -7 + 2t, t, 6, 2).$$

Insbesondere hat unser Gleichungssystem unendlich viele Lösungen. ◁

Im Vorgriff auf die Vektorrechnung sei erwähnt, dass man die allgemeine Lösung auch in der Form

$$(0, -7, 0, 6, 2) + t \cdot (-3, 2, 1, 0, 0)$$

schreiben kann, wobei die Variablen jetzt die Rolle von Koordinaten spielen. Dies ist die Parameterdarstellung einer Geraden im fünfdimensionalen Raum mit dem Stützpunkt $(0, -7, 0, 6, 2)$ und dem Richtungsvektor $(-3, 2, 1, 0, 0)$. Sie beschreibt die gleichförmige Bewegung eines Punktes, wenn man den Parameter als Zeit interpretiert, was die Verwendung der Variablen t erklärt. Man beachte, dass der Parameter in einer solchen Darstellung mit keiner der Koordinaten übereinstimmen muss.

Da man im Verlauf des Verfahrens aus mehreren Möglichkeiten wählen kann, verwundert es nicht, dass die Aufteilung in freie und abhängige Variablen nicht eindeutig ist, ebenso wenig wie die Formel für die allgemeine Lösung.

Eigentlich sind wir nur von einer Darstellung der Lösungsmenge (als Menge von n-Tupeln, die den vorgegebenen Gleichungen genügen) zu einer anderen (als Menge von n-Tupeln, die man für irgend eine Wahl der Parameter erhält) übergegangen, und jede von beiden hat ihre Vor- und Nachteile. Die allgemeine Lösung erlaubt es, einzelne Lösungen hinzuschreiben, nützt aber gar nichts bei der Frage, ob ein vorgegebenes n-Tupel eine Lösung ist. Letztere Frage kann man viel einfacher mit Hilfe des ursprünglichen Gleichungssystems beantworten, welches hingegen nicht unmittelbar erlaubt, auch nur eine Lösung hinzuschreiben.

Schaut man sich die Methode an, mit der wir die Lösungen des Systems bestimmt haben, nachdem es in Stufenform gebracht war, so erkennt man auch hier Substitutionen: Wir haben den für eine Variable gefunden Wert in die früheren Gleichungen eingesetzt. Um sich das Auflösen der Gleichungen zu ersparen, liegt es nahe, auch hierbei Gleichungen zueinander zu addieren wie schon beim Gauß-Verfahren. Man eliminiert dabei zunächst die letzte gebundene Variable aus allen vorherigen Gleichungen, dann die vorletzte usw. Das gesamte derartige Vorgehen zur Lösung linearer Gleichungssysteme nennt man auch *Gauß-Jordan-Verfahren*.

Beispiel 2.5. Betrachten wir noch einmal das obige System in Stufenform, so muss man zur Elimination von z die letzte Gleichung zur Ersten addieren, ihr doppeltes von der zweiten und ihr fünffaches von der dritten subtrahieren:

$$\begin{cases} v + 2w - x + 3y = 4 \\ w - 2x + y = -1 \\ -y = -6 \\ z = 2 \end{cases}$$

Als Nächstes addiert man zur Elimination von y die dritte Gleichung zur

zweiten und ihr dreifaches zur ersten:

$$\begin{cases} v+2w- \quad x & & = -14 \\ \quad w-2x & & = \quad -7 \\ & -y & = \quad -6 \\ & & z = \quad 2 \end{cases}$$

Schließlich subtrahiert man das doppelte der zweiten Gleichung von der ersten, um dort die Variable w zu eliminieren:

$$\begin{cases} v \quad +3x & & = \quad 0 \\ \quad w-2x & & = -7 \\ & -y & = -6 \\ & & z = \quad 2 \end{cases}$$

Aus diesem System kann man die Lösung leicht ablesen. ◁

Das gewählte Beispiel war besonders einfach, da die Substituionsvariablen in den Substitutionsgleichungen mit dem Koeffizienten 1 oder -1 vorkamen. Im Allgemeinen erreicht man das erst mit Hilfe einer weiteren Umformung, nämlich der Multiplikation beider Seiten einer Gleichung mit einer von Null verschiedenen Zahl. Auch diese Operation ist umkehrbar, verändert also die Lösungsmenge nicht.

Bei der Division durch betragsmäßig kleine Zahlen verursachen kleine Abweichungen in den Eingaben große Fehler im Ergebnis. Darum sollte man bei mehreren Möglichkeiten diejenige *Substitutionsvariable* und diejenige Substitutionsgleichung wählen, in der die Substitutionsvariable mit dem betragsmäßig größten Koeffizienten vorkommt. Um nicht durcheinander zu kommen, empfielt es sich, dabei die Reihenfolge der Variablen anzupassen.

2.3 Matrizen

Es ist lästig, die Gleichungen eines Systems, die sich nicht mehr ändern, immer wieder abzuschreiben. Da die Reihenfolge der Variablen festgelegt ist, kann man sich etwas Erleichterung verschaffen, indem man nur noch die Koeffizienten notiert. Kommt eine Variable in einer Gleichung nicht vor, so wird der Koeffizient 0 hingeschrieben, damit die Anordnung nicht durcheinanderkommt. Das entstehende Schema nennt man *Matrix*, man schließt es in Klammern ein. Wir werden Matrizen noch zu anderen Zwecken verwenden, und im konkreten Fall spricht man von der *Koeffizientenmatrix* eines Gleichungssystems, bzw. von der *erweiterten Koeffizientenmatrix*, wenn auch die Konstanten auf der rechten Seite aufgenommen werden. Man kann auch n-Tupel als einzeilige Matrizen auffassen. Im Unterschied zu ihnen trennt

man die Einträge von Matrizen nicht durch Kommata, sondern durch etwas größere Abstände.

Die bisherigen Umformungen von linearen Gleichungssystemen übersetzen sich in *Zeilenoperationen*, die mit Matrizen vorgenommen werden:

- die Addition des Vielfachen einer Zeile zu einer anderen,
- die Vertauschung von Zeilen,
- die Multiplikation einer Zeile mit einer von Null verschiedenen Zahl.

Damit lässt sich das Gauß-Jordan-Verfahren mit geringerem Schreibaufwand umsetzen und auch für die Ausführung durch einen Computer programmieren. Den Koeffizienten der Substitutionsvariablen in der Substitutionsgleichung nennt man auch *Pivotelement*.

Beispiel 2.6. Betrachten wir das Gleichungssystem

$$\begin{cases} x_1 + 2x_2 + 2x_3 + 3x_4 = -1 \\ -2x_1 + 2x_2 + x_3 + 3x_4 = 3 \\ 3x_1 - x_2 + x_3 - 3x_4 = 2 \end{cases}$$

Seine erweiterte Koeffizientenmatrix ist

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

Addieren wir das Doppelte der ersten Zeile zur zweiten und subtrahieren ihr Dreifaches von der dritten, so erhalten wir

$$\begin{pmatrix} 1 & 2 & 2 & 3 & -1 \\ 0 & 6 & 5 & 9 & 1 \\ 0 & -7 & -5 & -12 & 5 \end{pmatrix}$$

Als Nächstes ist das $\frac{7}{6}$-fache der zweiten Zeile zur dritten zu addieren. Das Ergebnis ist

$$\begin{pmatrix} 1 & 2 & 2 & 3 & -1 \\ 0 & 6 & 5 & 9 & 1 \\ 0 & 0 & \frac{5}{6} & -\frac{3}{2} & \frac{37}{6} \end{pmatrix}$$

Hier benötigt die Meisten schon einen Schmierzettel für Nebenrechnungen. Nach dem Gauß-Jordan-Verfahren müssen wir nun das Sechsfache der letzten Zeile von der vorletzten und ihr $\frac{12}{5}$-faches von der ersten Zeile subtrahieren. Dies ergibt

$$\begin{pmatrix} 1 & 2 & 0 & \frac{33}{5} & -\frac{79}{5} \\ 0 & 6 & 0 & 18 & -36 \\ 0 & 0 & \frac{5}{6} & -\frac{3}{2} & \frac{37}{6} \end{pmatrix}$$

Schließlich ist noch ein Drittel der zweiten Zeile von der ersten zu subtrahieren:

$$\begin{pmatrix} 1 & 0 & 0 & \frac{3}{5} & -\frac{19}{5} \\ 0 & 6 & 0 & 18 & -36 \\ 0 & 0 & \frac{5}{6} & -\frac{3}{2} & \frac{37}{6} \end{pmatrix}$$

Nun können wir noch die letzte Zeile mit $\frac{6}{5}$ zu multiplizieren und die vorletzte durch 6 zu dividieren:

$$\begin{pmatrix} 1 & 0 & 0 & \frac{3}{5} & -\frac{19}{5} \\ 0 & 1 & 0 & 3 & -6 \\ 0 & 0 & 1 & -\frac{9}{5} & \frac{37}{5} \end{pmatrix}$$

Unser Gleichungssystem ist also äquivalent zu dem System

$$\begin{cases} x_1 & & & + \frac{3}{5}x_4 = -\frac{19}{5} \\ & x_2 & & + 3x_4 = -6 \\ & & x_3 & - \frac{9}{5}x_4 = \frac{37}{5} \end{cases}$$

und wählen wir die freie Variable x_4 als Parameter t, so erhalten wir die allgemeine Lösung

$$\left(-\frac{19}{5} - \frac{3}{5}t, -6 - 3t, \frac{37}{5} + \frac{9}{5}t, t \right).$$

Anstatt beide Seiten einer Gleichung mit einer gebrochenen Zahl zu multiplizieren, um sie dann zu einer anderen Gleichung zu addieren, kann man vorher beide Seiten der Zielgleichung mit dem Nenner multiplizieren, um Brüche zu vermeiden. Wir werden das aktuelle Beispiel gleich in etwas anderem Zusammenhang auf die angedeutete Weise behandeln. ◁

Sind die Koeffizienten eines Gleichungssystems nur näherungsweise bekannt, so werden sie üblicherweise als Dezimalbrüche angegeben, und dann rechnet man mit diesen weiter. Sind hingegen die Koeffizienten wie im obigen Beispiel als gemeine Brüche gegeben, so wird in der Regel auch eine exakte Antwort erwartet, und dann verbieten sich Dezimalbrüche wegen der Rundungsfehler.

2.4 Systeme mit variabler rechter Seite

In manchen Anwendungen ist ein lineares Gleichungssystem immer wieder mit verschiedenen rechten Seiten zu lösen, während die Koeffizienten auf der linken Seite unverändert bleiben. In diesem Fall ist es sinnvoll, die Konstanten auf der rechten Seite durch weitere Variablen zu ersetzen, um eine Lösungsformel zu finden.

Beispiel 2.7. Wir betrachten das Gleichungssystem

$$\begin{cases} x_1 + 2x_2 + 2x_3 + 3x_4 = y_1 \\ -2x_1 + 2x_2 + x_3 + 3x_4 = y_2 \\ 3x_1 - x_2 + x_3 - 3x_4 = y_3 \end{cases}$$

Wie üblich haben wir jeder Variablen eine Spalte zugewiesen, so dass sich die Zugehörigkeit von Koeffizienten aus ihrer Stellung ergibt und wir die Matrizenschreibweise verwenden können. Die erweiterte Koeffizientenmatrix nimmt nun folgende verallgemeinerte Form an:

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

Wie oben addieren wir das Doppelte der ersten Zeile zur zweiten und subtrahieren ihr Dreifaches von der dritten:

$$\begin{pmatrix} 1 & 2 & 2 & 3 & 1 & 0 & 0 \\ 0 & 6 & 5 & 9 & 2 & 1 & 0 \\ 0 & -7 & -5 & -12 & -3 & 0 & 1 \end{pmatrix}$$

Man könnte auch weiter die selben Zeilenoperationen wie im vorigen Abschnitt vornehmen, aber wir demonstrieren im Folgenden die schon erwähnte Methode, Brüche so lange wie möglich zu vermeiden. Dazu multiplizieren wir die dritte Zeile mit 6 und addieren das Siebenfache der zweiten:

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

Wir multiplizieren die erste Zeile mit 5 und subtrahieren die letzte von der vorletzten und ihr Doppeltes von der ersten:

$$\begin{pmatrix} 5 & 10 & 0 & 33 & 13 & -14 & -12 \\ 0 & 6 & 0 & 18 & 6 & -6 & -6 \\ 0 & 0 & 5 & -9 & -4 & 7 & 6 \end{pmatrix}$$

Nun dividieren wir die zweite Zeile durch 6 und subtrahieren ihr Zehnfaches von der ersten:

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

Schließlich dividieren wir noch die erste und letzte Zeile durch 5:

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

Unser Gleichungssystem ist also äquivalent zu dem System

$$\begin{cases} x_1 + \frac{3}{5}x_4 = \frac{3}{5}y_1 - \frac{4}{5}y_2 - \frac{2}{5}y_3 \\ x_2 + 3x_4 = y_1 - y_2 - y_3 \\ x_3 - \frac{9}{5}x_4 = -\frac{4}{5}y_1 + \frac{7}{5}y_2 + \frac{6}{5}y_3 \end{cases}$$

Die Lösung hängt nun von den auf der rechten Seite vorgegebenen Zahlen y_1, y_2 und y_3 ab sowie einem weiteren frei wählbaren Parameter, z. B. $t = x_4$. ◁

Bei der Anwendung des Gauß-Verfahrens auf ein System mit k Gleichungen können die linken Seiten mancher Gleichungen zu Null werden. Diese bilden ein Teilsystem, in dem nur die Variablen $y_1, \ldots, y_k$ vorkommen. Das ursprüngliche System ist offenbar genau für diejenigen Werte dieser Variablen lösbar, die dem besagten Teilsystem genügen.

Wir hatten auf Seite 14 erwähnt, dass die Bestimmung der allgemeinen Lösung nur der Übergang zu einer anderen Beschreibung der Lösungsmenge ist. Nun sind wir auch in der Lage, den umgekehrten Übergang vorzunehmen. Ist beispielsweise M die Menge der Tripel (x_1, x_2, x_3), die sich in der Form

$$\begin{aligned} x_1 &= 3 + t_1 - 4t_2 \\ x_2 &= 1 + 3t_1 + 2t_2 \\ x_3 &= 2 - 4t_1 - 5t_2 \end{aligned}$$

mit beliebigen Parametern t_1 und t_2 schreiben lassen, so können wir dies als Gleichungssystem mit den Unbekannten t_1 und t_2 sowie variabler rechter Seite umschreiben:

$$\begin{aligned} t_1 - 4t_2 &= x_1 - 3 \\ 3t_1 + 2t_2 &= x_2 - 1 \\ -4t_1 - 5t_2 &= x_3 - 2 \end{aligned}$$

Das Gaußverfahren führt, wenn wir Nenner vermeiden, auf

$$\begin{aligned} t_1 - 4t_2 &= x_1 - 3 \\ 14t_2 &= -3x_1 + x_2 + 8 \\ 0 &= -x_1 + 3x_2 + 2x_3 - 4 \end{aligned}$$

Somit ist M die Lösungsmenge der Gleichung

$$-x_1 + 3x_2 + 2x_3 = 4.$$

2.5 Die Cramersche Regel

Manchmal benötigt man eine Formel für die Lösung eines linearen Gleichungssystems mit variabler rechter Seite und variablen Koeffizienten. Das System in Beispiel 2.2 war von dieser Art, wobei dort allerdings die Lösung nicht eindeutig bestimmt war. Das ist, wenn es keine überflüssigen Gleichungen gibt, nur dann zu erwarten, wenn die Anzahl der Gleichungen gleich der Anzahl n der Variablen ist.

Eine lineare Gleichung mit n Variablen x_1, x_2, ..., x_n hat die Form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b,$$

wobei a_1, a_2, ..., a_n und b gegebene Zahlen sind. Sind mehrere solche Gleichungen gegeben so bezeichnen wir den Koeffizienten der j-ten Variablen in der i-ten Gleichung mit a_{ij}. (Eigentlich müsste man $a_{i,j}$ schreiben, aber aus Bequemlichkeit tut man das nur, wenn Verwechslungen drohen, etwa wenn für i oder j ein Term oder eine mehrstellige Zahl eingesetzt ist.) Im Fall $n = 2$ hat das System die Form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases}$$

Zunächst eliminieren wir x_1 aus der zweiten Gleichung. Um Nenner zu vermeiden, multiplizieren wir diese Gleichung mit a_{11}, bevor wir das a_{21}-fache der ersten Gleichung subtrahieren, und erhalten

$$(a_{11}a_{22} - a_{12}a_{21})x_2 = a_{11}b_2 - a_{21}b_1.$$

Bei Elimination von x_1 aus der ersten Gleichung hätten wir bis auf Multiplikation beider Seiten mit -1 das Selbe erhalten. Wir können auch x_2 eliminieren, indem wir z. B. die erste Gleichung mit a_{22} multiplizieren und davon das a_{12}-fache der zweiten Gleichung subtrahieren:

$$(a_{11}a_{22} - a_{21}a_{12})x_1 = b_1a_{22} - b_2a_{12}.$$

In beiden Gleichungen kommt auf der linken Seite der gleiche Koeffizient vor. Man nennt ihn die Determinante der Koeffizientenmatrix und definiert dafür die Schreibweise

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Wenn die Determinante verschwindet, so ist in den letzten Gleichungen die linke Seite gleich Null. Im Allgemeinen trifft das aber auf die rechte Seite nicht zu, und dann hat das System keine Lösung. Ist die Determinante von

Null verschieden, so erhalten wir durch Auflösen nach den Unbekannten die Cramersche Regel

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}, \qquad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}.$$

Wir weisen darauf hin, dass die Determinante bei Vertauschung der Spalten nur das Vorzeichen, aber nicht den Betrag wechselt.

Nun betrachten wir den Fall $n = 3$. Eliminieren wir in dem System

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 & (\mathrm{I}) \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 & (\mathrm{II}) \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 & (\mathrm{III}) \end{cases}$$

die Variable x_3 unter Benutzung der ersten und zweiten Gleichung, so erhalten wir

$$\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} x_1 + \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} x_2 = \begin{vmatrix} b_1 & a_{13} \\ b_2 & a_{23} \end{vmatrix}. \qquad (\mathrm{III}')$$

Wie wir bereits wissen, ist es dabei gleichgültig, welche von beiden Gleichungen wir als Substitutionsgleichung benutzen. Eliminieren wir die selbe Variable mit Hilfe der ersten und dritten Gleichung, so erhalten wir

$$\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} x_1 + \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} x_2 = \begin{vmatrix} b_1 & a_{13} \\ b_3 & a_{33} \end{vmatrix}. \qquad (\mathrm{II}')$$

Entsprechend dem Gaußverfahren würde man sich die erste Gleichung als Substitutionsgleichung merken und weiter das aus den beiden letzten Gleichungen bestehende Teilsystem betrachten. Wendet man darauf die obige Cramersche Regel an, so erhält man eine Formel für die Unbekannten in Form von kürzbaren Brüchen (Übungsaufgabe). Man kann aber sofort den gekürzten Bruch erhalten, wenn man sich die Symmetrien des Systems zu Nutze macht. Es gibt nämlich keinen Grund, zwei Gleichungspaare zu bevorzugen. Wir können auch die zweite und dritte Gleichung benutzen und erhalten

$$\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} x_1 + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} x_2 = \begin{vmatrix} b_2 & a_{23} \\ b_3 & a_{33} \end{vmatrix}. \qquad (\mathrm{I}')$$

Zunächt ist nicht klar, was es uns nützen soll, drei Gleichungen zu haben. Gewöhnlich bleibt ja nach der Elimination einer Variablen ein System übrig, das eine Gleichung weniger enthält. In der Tat kann man im Allgemeinen aus zweien unserer drei Gleichungen die dritte gewinnen. Wenn wir uns daran

erinnern, wie diese Gleichungen entstanden sind, nämlich in symbolischer Schreibweise

$$\mathrm{I}' = a_{33}\mathrm{II} - a_{23}\mathrm{III}, \qquad \mathrm{II}' = a_{33}\mathrm{I} - a_{13}\mathrm{III}, \qquad \mathrm{III}' = a_{23}\mathrm{I} - a_{13}\mathrm{II},$$

so ist offensichtlich, dass die Gleichung $a_{13}\mathrm{I}' - a_{23}\mathrm{II}' + a_{33}\mathrm{III}'$ nichts anderes als $0 = 0$ ist. Der Koeffizient von x_1 auf ihrer linken Seite ist

$$a_{13}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} - a_{23}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + a_{33}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}.$$

Er ist der Wert der linke Seite, wenn man $x_1 = 1$ und $x_2 = 0$ einsetzt, muss also verschwinden. Dies kann man auch unter Benutzung der Definition der Determinante nachrechnen.

Dieses Verschwinden bleibt erhalten, wenn wir in jeder der drei Determinanten die Spalten vertauschen, da dann der gesamte Term nur das Vorzeichen wechselt. Vertauschen wir dann noch die Rollen von x_1 und x_3, ersetzen also den rechten Index 1 durch 3 bzw. 3 durch 1, so folgt

$$a_{11}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} = 0.$$

Dies kann man natürlich einfach nachrechnen, aber uns ist klar geworden, wie man auf die Idee kam, einen solchen Term zu betrachten.

Die letzte Gleichung zeigt, dass wir die Variable x_1 eliminieren können, indem wir die Gleichungen I', II' und III' mit a_{11}, a_{21} bzw. a_{31} multiplizieren und dann mit abwechselnden (lateinisch: alternierenden) Vorzeichen addieren. Es ergibt sich

$$\left(a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}\right) x_2 = a_{11}\begin{vmatrix} b_2 & a_{23} \\ b_3 & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} b_1 & a_{13} \\ b_3 & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} b_1 & a_{13} \\ b_2 & a_{23} \end{vmatrix}.$$

Den Koeffizienten auf der linken Seite nennt man die *Determinante* der Koeffizientenmatrix des Gleichungssystems und definiert die Schreibweise

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$

Wenn diese Determinante verschwindet, so ist das System nicht für beliebige rechte Seiten lösbar. Ist sie aber von Null verschieden, so erhalten wir die

Cramersche Regel

$$x_2 = \frac{\begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}.$$

Es ist klar, dass die Determinante bei der Vertauschung der letzten beiden Spalten nur das Vorzeichen wechselt. Wenn wir also die Rollen von x_2 und x_3 vertauschen und dann in der Lösungsformel die letzten beiden Spalten sowohl im Zähler als auch im Nenner vertauschen, so folgt

$$x_3 = \frac{\begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}.$$

Wir können die Determinante der Koeffizientenmatrix vollständig ausmultiplizieren und neu zusammenfassen. Sie ist z. B. gleich

$$\begin{aligned} &a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{21}a_{32}a_{13} - a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} \\ &\qquad = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} a_{13} - \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} a_{23} + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} a_{33} \end{aligned}$$

Also wechselt die Determinante auch bei der Vertauschung der ersten beiden Spalten nur das Vorzeichen, und wie oben folgt

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}}.$$

Man nennt die obigen Formeln, die die Determinante einer dreireihigen Matrix durch zweireihige Determinanten ausdrücken, die *Entwicklung* nach der ersten bzw. dritten Spalte. Man findet leicht die Entwicklung nach der zweiten Spalte

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{32} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}.$$

Die Entwicklung nach einer Spalte kann wie folgt einheitlich beschreiben: Die Determinante ist die Summe aller Einträge der betreffenden Spalte multipliziert mit ihren jeweiligen Kofaktoren. Dabei ist der *Kofaktor* eines Matrixeintrages die Determinante der Untermatrix, die sich durch Streichen seiner Zeile und Spalte ergibt, multipliziert mit einem Vorzeichenfaktor. Diese Faktoren bilden ein schachbrettartiges Muster, beginnend mit einem Plus in der linken oberen Ecke.

Die beschriebene Lösungsformel für $n = 2$ und $n = 3$ wurde von Colin Maclaurin gefunden. Gabriel Cramer hat 1750 ohne Beweis eine Verallgemeinerung auf beliebige n beschrieben, die Gottfried Wilhelm Leibniz bereits früher gefunden und bewiesen, aber nicht veröffentlicht hatte. Pierre Frédéric Sarrus gab 1833 noch eine Merkregel für den Fall $n = 3$ an, die sich aber nicht auf größere n verallgemeinert.

3 Algebraische Strukturen

Bisher haben wir auf die Schulkenntnisse über Zahlbereiche aufgebaut. Um tiefer in die Theorie einzudringen, benötigen wir nun genauere Kenntnisse. Wir setzen die Grundbegriffe der Mengenlehre voraus, auf die in der Veranstaltung zur Analysis näher eingegangen wurde. Insbesondere bezeichnen wir die Menge aller geordneten Paare (x, y), wobei x ein Element einer Menge X und y ein Element einer Menge Y ist, mit $X \times Y$, genannt Produktmenge von X und Y. Die Menge aller n-Tupel, deren Einträge Elemente einer Menge X sind, bezeichnen wir mit X^n.

3.1 Äquivalenzrelationen und Ordnungen

Eine wichtige Methode der Wissenschaften ist die Klassifikation von Lebewesen, Objekten, Lauten oder Begriffen. Ein ähnlicher Prozess ist die Einteilung der Teilnehmer/innen eines jeden Tutoriums in Abgabegruppen. Solche Sachverhalte können wie folgt mathematisch wiedergegeben werden.

Definition 1. *Eine* Partition *einer Menge X ist eine Menge $\mathcal{P}$ von nichtleeren Teilmengen von X, so dass es für jedes Element x von X genau ein Element A von $\mathcal{P}$ mit der Eigenschaft $x \in A$ gibt.*

Bei einer Klassifikation, beispielsweise von Lebewesen, gibt es keine vorgefasste Liste von Klassen[5] oder Arten. Vielmehr gibt es Kriterien dafür, wann zwei Lebewesen zur selben Art gehören. So ein Kriterium ist logisch betrachtet eine Relation. Dabei sind beide Variablen, die in der Relation vorkommen, mit Elementen ein und der selben Menge zu belegen, und wir sprechen dann von einer Relation auf einer Menge.

Definition 2. *Eine Relation*[6] *„$\sim$“ auf einer Menge X wird* Äquivalenzrelation *genannt, wenn für beliebige Elemente x, y und z von X gilt:*

- $x \sim x$
- $x \sim y \;\Rightarrow\; y \sim x$
- $x \sim y \;\wedge\; y \sim z \;\Rightarrow\; x \sim z$

Diese drei Axiome der Äquivalenzrelation nennt man *Reflexivität*, *Symmetrie* und *Transitivität*.

[5] Das lateinische Wort *classis* kommt von dem Verb *clamare* (rufen, auch einberufen). Im alten Rom wurden die Bürger nach ihrem Stand in in Klassen eingeteilt, die jeweils bestimmte Teile des Heeres und der Flotte aufbieten mussten. Daraus gingen die Steuerklassen hervor.

[6] Das Zeichen „$\sim$“ wird „Tilde“ oder „Schlange“ gelesen, und wenn es eine Äquivalenzrelation ohne eigenen Namen bezeichnet, auch als „äquivalent“.

Satz 1. *Es sei X eine Menge.*

(i) *Ist $\mathcal{P}$ eine Partition von X und definieren wir für x, $y \in X$*

$$x \sim y \quad \Leftrightarrow \quad \exists\, A \in \mathcal{P}\,(x \in A \ \wedge \ y \in A),$$

so ist „$\sim$“ eine Äquivalenzrelation auf X.

(ii) *Ist „$\sim$“ eine Äquivalenzrelation auf X und setzen wir für jedes Element x von X*

$$[x] = \{y \in X \mid y \sim x\},$$

genannt Äquivalenzklasse *von x, so ist*[7]

$$\mathcal{P} = \{[x] \mid x \in X\}$$

eine Partition von X.

Die Gleichheitsrelation ist übrigens ein Beispiel einer Äquivalenzrelation. In diesem Fall besteht die zugehörige Partition aus den Einermengen in X.

Beweis. Ist eine Partition $\mathcal{P}$ gegeben und definieren wir eine Relation $\sim$ wie in (i), so ist sie offensichtlich reflexiv und symmetrisch. Gilt $x \sim y$ und $y \sim z$, so gibt es A, $B \in \mathcal{P}$, so dass x, $y \in A$ und y, $z \in B$ ist. Da aber y in nur einem Teil der Partition enthalten sein kann, gilt $A = B$, und die Transitivität folgt.

Nun betrachten wir eine Äquivalenzrelation „$\sim$“ und definieren $\mathcal{P}$ wie in Behauptung (ii). Wegen der Reflexivität gehört jedes Element x von X zu einem Element von $\mathcal{P}$, nämlich $[x]$.

Angenommen, für Elemente x, y von X haben $[x]$ und $[y]$ ein gemeinsames Element z. Dann gilt $z \sim x$ und $z \sim y$. Mit der Symmetrie folgt $x \sim z$, und mit der Transitivität folgt $x \sim y$. Für jedes Element u von $[x]$ gilt $u \sim x$, und mit der Transitivität folgt $u \sim y$, so dass $[x] \subseteq [y]$. Genauso zeigt man $[y] \subseteq [x]$, und somit folgt $[x] = [y]$. Verschiedene Äquivalenzklassen sind also disjunkt, und darum kann ein Element von X nur zu einer gehören. □

Eine andere wissenschaftliche Methode besteht darin, Objekte (etwa chemische Elemente, archäologische Funde) nach gewissen Kriterien in eine Reihenfolge zu bringen. Auch diese Kriterien sind Relationen.

Definition 3. *Eine Relation*[8] *„$\preceq$“ auf einer Menge X wird* Ordnung *genannt, wenn für beliebige Elemente x, y, z von X gilt*

- $x \preceq x$

[7] Exakter wäre $\mathcal{P} = \{A \mid \exists x \in X\ A = [x]\}$.

[8] oft gelesen „vor oder gleich“

- $x \preceq y \ \wedge \ y \preceq x \ \Rightarrow \ x = y$
- $x \preceq y \ \wedge \ y \preceq z \ \Rightarrow \ x \preceq z$
- $x \preceq y \ \vee \ y \preceq x$

Eine Menge zusammen mit einer Ordnung nennt man geordnete Menge.

Die zweite Eigenschaft nennt man *Antisymmetrie*, die letzte nennt man *Totalität*, die anderen sind uns bereits bekannt. Beispiele von Ordnungen sind die Kleiner-Gleich-Relation zwischen reellen Zahlen oder die Reihenfolge im Alphabet. Aus jeder Ordnung erhält man eine strikte Ordnung

$$x \prec y \quad \Leftrightarrow \quad x \preceq y \ \wedge \ x \neq y,$$

die man ebenfalls durch vier Axiome charakterisieren kann, und aus jeder strikten Ordnung erhält man eine nichtstrikte:

$$x \preceq y \quad \Leftrightarrow \quad x \prec y \ \vee \ x = y.$$

Durch Vertauschung der Argumente entsteht eine Ordnung, die man gern durch das gespiegelte Symbol abkürzt und *entgegengesetzte Ordnung* nennt, also

$$x \succ y \quad \Leftrightarrow \quad y \prec x.$$

Die Verknüpfung

$$x \prec y \ \wedge \ y \prec z$$

schreibt man oft in der platzsparenden Form

$$x \prec y \prec z.$$

Eine Relation, die reflexiv, antisymmetrisch und transitiv, aber nicht unbedingt total ist, nennt man eine *Halbordnung*. Ist beispielsweise „$<$“ eine strikte Ordnung auf einer Menge X, so erhalten wir eine strikte Halbordnung „$\lhd$“ auf X^n durch die Festlegung

$$(x_1, \ldots, x_n) \lhd (y_1, \ldots, y_n), \qquad \text{wenn} \quad x_1 < y_1, \ \ldots, \ x_n < y_n.$$

Ein Beispiel für eine Ordnung auf X^n ist die (aufsteigende, strikte) *lexikographische Ordnung* „$\prec$“, die wie folgt definiert ist:

> $(x_1, \ldots, x_n) \prec (y_1, \ldots, y_n)$, wenn für den kleinsten Index l mit der Eigenschaft $x_l \neq y_l$ gilt $x_l < y_l$.

Auf diese Weise werden Wörter der Länge n im Lexikon geordnet, wenn X das Alphabet ist. Geht man von der entgegengesetzten Ordnung auf X aus, so erhält man die absteigende lexikographische Ordnung.

3.2 Ringe

3.2.1 Definition und erste Beispiele

Die Bereiche der ganzen Zahlen, der rationalen Zahlen und der reellen Zahlen, die man nach Bourbaki mit $\mathbb{Z}$, $\mathbb{Q}$ bzw. $\mathbb{R}$ abkürzt, sind Beispiele für den folgenden Begriff.

Definition 4. *Ein* kommutativer Ring *ist eine Menge R mit zwei* Operationen *bzw.* Verknüpfungen*, genannt Addition und Multiplikation, die folgende Eigenschaften haben.*

(i) *Für alle a, $b \in R$ gilt*

$$a + b = b + a, \qquad a \cdot b = b \cdot a.$$

(ii) *Für alle a, b, $c \in R$ gilt*

$$(a + b) + c = a + (b + c), \qquad (a \cdot b) \cdot c = a \cdot (b \cdot c)$$

(iii) *Für alle a, b, $c \in R$ gilt*

$$(a + b) \cdot c = a \cdot c + b \cdot c, \qquad c \cdot (a + b) = c \cdot a + c \cdot b.$$

(iv) *Es gibt ein Element $0 \in R$, so dass für alle $a \in R$ gilt*

$$a + 0 = a, \qquad 0 + a = a.$$

(v) *Es gibt ein Element $1 \in R$, so dass für alle $a \in R$ gilt*

$$a \cdot 1 = a, \qquad 1 \cdot a = a.$$

(vi) *Für beliebige a, $b \in R$ ist die Gleichung*

$$a = x + b$$

in R lösbar.

Dabei verstehen wir unter einer Operation bzw. Verknüpfung auf einer Menge R eine Abbildung $R \times R \to R$. Die genannten Eigenschaften nennt man die Axiome eines kommutativen Ringes. Insbesondere nennt man die Eigenschaften (i), (ii) und (iii) das *Kommutativgesetz*, das *Assoziativgesetz* und das *Distributivgesetz*. Verlangt man nicht die Kommutativität der Multiplikation, so erhält man den Begriff eines *Ringes* ohne das Adjektiv „kommutativ“, wobei aber die Kommutativität der Addition auch hier vorausgesetzt wird.

Haben Elemente 0 und $0'$ beide die Eigenschaft (iv), so folgt

$$0' = 0' + 0 = 0.$$

Analog zeigt man, dass es nur ein Element mit der Eigenschaft (v) geben kann. Obwohl jeder Ring sein eigenes Null- bzw. Einselement besitzt, werden diese meist einheitlich mit 0 und 1 bezeichnet. Das Multiplikationszeichen wird oft weggelassen, wenn das nicht zu Verwechslungen führt. Manchmal werden sogar Ringe ohne Einselement betrachtet.

Sind c und c' Lösungen der Gleichung in (vi) im Fall $a = 0$, so folgt

$$c' = c' + 0 = c' + (b + c) = (c' + b) + c = 0 + c = c.$$

Man nennt das eindeutig bestimmte Element c mit der besagten Eigenschaft das *entgegengesetzte Element* von b. Ist nun a beliebig, so ist $x = a + c$ eine Lösung. Man sieht leicht, dass sie eindeutig bestimmt ist, und bezeichnet sie mit $a - b$. Statt $0 - b$ schreibt man auch $-b$.

Für jedes Element a erhalten wir

$$a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0.$$

Addieren wir auf beiden Seiten das entgegengesetzte Element von $a \cdot 0$, so folgt

$$0 = a \cdot 0. \tag{vii}$$

Ist in einem Ring das Nullelement gleich dem Einselement, so gilt für jedes Ringelement a, dass $a = a \cdot 1 = a \cdot 0 = 0$, also folgt $R = \{0\}$, und dies erfüllt in der Tat alle Anforderungen an einen Ring.

Der Bereich $\mathbb{N}$ der natürlichen Zahlen bildet keinen Ring, weil die Bedingung (vi) nicht erfüllt ist. Als weiteres Beispiel eines Ringes betrachten wir die Menge $\{\mathrm{g}, \mathrm{u}\}$ mit den Operationen

$+$	g	u
g	g	u
u	u	g

$\cdot$	g	u
g	g	g
u	g	u

In diesem Ring ist g das Nullelement und u das Einselement.

Auch die Menge aller Funktionen auf einer Menge X mit Werten in einem Ring R, die man manchmal mit R^X bezeichnet, wird zu einem Ring, wenn wir die Summe $f + g$ und das Produkt $f \cdot g$ von Funktionen f und g durch

$$(f + g)(x) = f(x) + g(x), \qquad (f \cdot g)(x) = f(x) \cdot g(x)$$

für $x \in X$ definieren.

In der Mathematik sind nicht nur Strukturen wichtig, sondern auch strukturerhaltende Abbildungen.

Definition 5. *Gegeben seien Ringe R und S. Eine Abbildung $h : R \to S$ wird Homomorphismus genannt, wenn sie folgende Eigenschaften hat.*

(i) *Es gilt $h(0) = 0$ und $h(1) = 1$.*

(ii) *Für alle Elemente a und b von R gilt*

$$h(a + b) = h(a) + h(b), \qquad h(a \cdot b) = h(a) \cdot h(b).$$

Ist zudem $R = S$, so nennt man h einen Endomorphismus. Ein umkehrbarer Homomorphismus heißt Isomorphismus, ein umkehrbarer Endomorphismus heißt Automorphismus. Zwei Ringe heißen isomorph, wenn es zwischen ihnen einen Isomorphismus gibt.

Man beachte, dass hier die Symbole 0, 1, $+$ und $\cdot$ auf verschiedenen Seiten der Gleichung verschiedene Bedeutung haben: Links beziehen sie sich auf den Ring R und rechts auf S. Man prüft leicht nach, dass die Umkehrabbildung eines umkehrbaren Homomorphismus wieder ein Homomorphismus ist. Hier sind einige Beispiele:

(i) Ist R ein Unterring von S, also eine Teilmenge, die das Null- und Einselement enthält, abgeschlossen unter Addition und Multiplikation ist und bezüglich dieser Operationen wieder ein Ring ist, so ist die Abbildung $R \to S$, die jedes Element auf sich selbst abbildet, ein Homomorphismus.

(ii) Wir erhalten einen Homomorphismus $\mathbb{Z} \to \{\mathrm{g}, \mathrm{u}\}$, indem wir jeder geraden Zahl das Element g und jeder ungeraden Zahl das Element u zuordnen.

3.2.2 Polynome

Das Rechnen mit Variablen funktioniert auch, wenn wir statt Zahlen Elemente eines kommutativen Ringes R betrachten, vorausgesetzt, wir benutzen nur die Operationen Addition und Multiplikation. *Terme* sind Zeichenketten, die nach folgenden Regeln gebildet werden:

- Jedes Element von R ist ein Term.
- Jede Variable ist ein Term.
- Schreiben wir zwei Terme hintereinander, getrennt durch das Zeichen „+“ oder „·“, so erhalten wir einen Term, genannt formale Summe bzw. formales Produkt der beiden Terme.

Im letzteren Fall sind die beiden Terme, zumindest bei Abweichung von den Vorrangregeln, in Klammern einzuschließen. Belegt man die Variablen mit Elementen von R, so kann man alle Rechenoperationen ausführen und erhält einen Wert in R. Dabei gehen die formale Summe und das formale Produkt zweier Terme in die Summe bzw. das Produkt ihrer Werte über. Als Maß für die Kompliziertheit eines Terms definiert man seinen *Grad* wie folgt:

- Das Nullelement 0 hat den Grad $-\infty$.
- Jedes Element von $R \setminus \{0\}$ hat den Grad 0.
- Jede Variable hat den Grad 1.
- Der Grad der Summer zweier Terme ist das Maximum ihrer Grade.
- Der Grad des Produkts zweier Terme ist die Summe ihrer Grade.

Definieren wir $(-\infty)+n = -\infty$ für alle natürlichen Zahlen n, so ist der Grad eines Terms ohne Variablen gleich dem Grad seines Wertes.

Definition 6. *Zwei Terme nennen wir äquivalent, wenn man durch formale Anwendung der Rechengesetze* (i)–(v) *und* (vii) *sowie durch die Ausführung von Rechenoperationen mit Ringelementen den einen in den anderen umformen kann.*

Die Äquivalenzklassen von Termen mit Elementen aus R und Variablen x, y, ... nennt man Polynome *in x, y, ... mit Koeffizienten in R.*

Der Grad *eines Polynoms ist das Minimum der Grade seiner Repräsentanten.*

Wir benutzen das Zeichen „$\equiv$“ für die Äquivalenz von Termen. Dies ist tatsächlich eine Äquivalenzrelation. Ihre Abhängigkeit von R kennzeichnen wir nicht in der Notation. Die Rechengesetze übersetzen sich in Äquivalenzen, z. B. die Kommutativgesetze in

$$x + y \equiv y + x, \qquad x \cdot y \equiv y \cdot x. \tag{i′}$$

Solche Äquivalenzen sind keine Prädikate, sondern Aussagen, und die Variablen sind hier keine freien, sondern so genannte *formale Variablen*. Man beachte, dass im Unterschied zu Gleichung (vii) die Äquivalenz

$$x \cdot 0 \equiv 0 \tag{vii′}$$

nicht aus den Äquivalenzen (i′)–(v′) folgt.

Beispiel 3.1. In jedem Ring gilt die erste *binomische Formel*

$$(x+y)\cdot(x+y) \equiv x\cdot x + (1+1)\cdot x\cdot y + y\cdot y.$$

Im Ring der ganzen Zahlen kann man $1+1$ durch 2 ersetzen, während im obigen Ring $\{\mathrm{g},\mathrm{u}\}$ dies das Nullelement g ist und man den mittleren Term weglassen kann.

Man kann auch Terme mit der Rechenoperation „$-$“ zulassen, wenn man zusätzlich die Relation $x-y \equiv x+(-1)\cdot y$ benutzt, wobei -1 hier als Element von R gemeint ist. Dann haben wir die zweite und die dritte binomische Formel

$$\begin{aligned}(x-y)\cdot(x-y) &\equiv x\cdot x - (1+1)\cdot x\cdot y + y\cdot y,\\ (x+y)\cdot(x-y) &\equiv x\cdot x - y\cdot y.\end{aligned}$$

Natürlich kann man die zweite binomische Formel als Spezialfall der ersten betrachten. ◁

Üblicherweise führt man für die Äquivalenzklassen von Termen keine neue Bezeichnung ein, sondern schreibt sie einfach als Terme. Manche Klammern werden aufgrund der Rechengesetze nun überflüssig. Der Kürze halber schreibt man Produkte einer Variablen mit sich selbst als Potenzen, aber allgemein ist die Operation des Potenzierens (mit variablen Exponenten) in Polynomen nicht erlaubt. Es ist üblich, ein Polynom durch einen Buchstaben abkürzen, hinter den man die vorkommenden Variablen in Klammern schreibt, z. B.

$$f(x,y) = (3x^2+2xy)(5x+y^2+4y).$$

Das Belegen der Variablen durch Elemente von R wird auch in dieser Schreibweise vorgenommen, beispielsweise

$$f(2,1) = (3\cdot 2^2 + 2\cdot 2\cdot 1)(5\cdot 2 + 1^2 + 4\cdot 1) = 16\cdot 15 = 240.$$

Bei den Äquivalenzumformungen haben wir Rechengesetze formal auf Terme angewendet. Diese werden beim Einsetzen von Ringelementen zu wahren Aussagen. Äquivalente Terme liefern somit das selbe Ergebnis. Jedes Polynom in n Variablen mit Koeffizienten in einem Ring R liefert also eine Funktion $f : R^n \to R$, für die man das selbe Formelzeichen benutzt. Funktionen, die auf diese Weise entstehen, nennt man *ganzrationale Funktionen*.

Mit Termen kann man mehrere Operationen ausführen: Man kann sie addieren, miteinander multiplizieren, und man kann eine Variable in einem Term durch einen anderen Term ersetzen. Substituieren wir etwa in dem obigen Beispiel die Variable y durch den Term $7x+z$, so erhalten wir

$$f(x,7x+z) = (3x^2+2x(7x+z))(5x+(7x+z)^2+4(7x+z)).$$

Wenn man einen der beteiligten Terme durch einen äquivalenten ersetzt, dann geht in all diesen Fällen das Ergebnis in einen äquivalenten Term über. Folglich kann man diese Operationen auch mit Polynomen ausführen. Insbesondere ist die Menge der Polynome in x und y mit Koeffizienten in einem gegebenen kommutativen Ring R wieder ein kommutativer Ring, den man mit $R[x, y]$ bezeichnet. Analog erhält man den *Polynomring* für eine beliebige Menge von Variablen. Außerdem ist eine Verkettung ganzrationaler Funktionen wieder eine ganzrationale Funktion.

Die Menge aller Terme ist schwer fassbar, aber man kann sie vereinfachen. Ein Term heißt *Monom*, wenn er ein Produkt eines Ringelementes und von Variablen ist, wobei Letztere auch fehlen dürfen. Wählt man eine Ordnung (Reihenfolge) auf der Menge der Variablen, so ist es üblich, die Faktoren in aufsteigender Reihenfolge zu ordnen. Ein Monom in den Variablen x_1, …, x_n lässt sich beispielsweise in der Form

$$a x_1^{i_1} \cdots x_n^{i_n}$$

schreiben. Unter seinem Exponenten verstehen wir das n-Tupel $(i_1, \ldots, i_n)$, und sein Grad ist $i_1 + \ldots + i_n$. Wir sagen, ein Term sei in *Normalform*, wenn er eine Summe von Monomen mit paarweise verschiedenen Exponenten ist. Man kann zeigen, dass jeder Term äquivalent zu einem Term in Normalform ist, z. B.

$$f(x, y) \equiv 15x^3 + 3x^2y^2 + 22x^2y + 2xy^3 + 8xy^2.$$

Üblicherweise zählt man die Monome entsprechend der absteigenden lexikographischen Ordnung ihrer Exponenten auf.

Satz 2. *Zwei Terme in Normalform sind genau dann äquivalent, wenn in beiden die Monome mit gleichen Exponenten jeweils die gleichen Koeffizienten haben.*

Der Grad eines Polynoms ist gleich dem Grad seiner Normalform.

Nicht vorkommende Monome kann man bei Bedarf mit dem Koeffizienten Null hinzufügen. Der Satz zeigt, dass man die Gleichheit zweier Polynome durch *Koeffizientenvergleich* in der Normalform feststellen kann.

Beispiel 3.2. Für den Ring $R = \{\mathrm{g}, \mathrm{u}\}$ gibt es nur vier Funktionen $R \to R$. Insbesondere ist

$$\forall\, x \in R\ (x^2 = x),$$

obwohl nach Satz 2 gilt[2]

$$x^2 \not\equiv x.$$

[2]Manche Autoren sprechen das Zeichen „≡“ als „identisch gleich“ aus und verstehen es als Synonym für die entsprechende Allaussage, aber wir benutzen es, um die beiden Aussagen zu unterscheiden.

Es können also unendlich viele Polynome die selbe ganzrationale Funktion ergeben. ◁

Zum Beweis von Satz 2, der nicht zum Pflichtstoff gehört, benötigen wir folgenden Hilfssatz.

Lemma 1. *Es sei R ein Unterring des kommutativen Ringes S und n eine natürliche Zahl. Des weiteren seien $x_1, \ldots, x_n$ Variable und $b_1, \ldots, b_n$ Elemente von S. Dann gibt es genau einen Homomorphismus $h : R[x_1, \ldots, x_n] \to S$, so dass $h(x_1) = b_1, \ldots, h(x_n) = b_n$ und $h(a) = a$ für alle $a \in R$.*

Zum Beweis des Lemmas nur soviel: Ist ein Polynom mit Koeffizienten in R gegeben, so können wir es als Polynom mit Koeffizienten in S betrachten und für die Variablen $x_1, \ldots, x_n$ die Elemente $b_1, \ldots, b_n$ einsetzen. Der entstehende Wert in S ist dann das Bild des gegebenen Polynoms unter der Abbildung h.

Die meisten Autoren definieren einen anderen Begriff des Polynoms, den wir zur Unterscheidung Normalpolynom nennen wollen. Ein Normalpolynom in n Variablen mit Koeffizienten in einem Ring R ist eine Abbildung a, die jedem n-Tupel $(i_1, \ldots, i_n)$ von natürlichen Zahlen ein Element $a_{i_1,\ldots,i_n}$ von R zuordnet, wobei nur endlich viele Werte von Null verschieden sind. Für den Moment bezeichnen wir die Menge dieser Normalpolynome mit R_n. Es sei $g : R_n \to R[x_1, \ldots, x_n]$ die Abbildung, die einem Normalpolynom a das Polynom

$$\sum_{(i_1,\ldots,i_n)} a_{i_1,\ldots,i_n} x_1^{i_1} \ldots x_n^{i_n}$$

zuordnet. Es gibt nur eine Möglichkeit, eine Addition und eine Multiplikation von Normalpolynomen zu definieren, so dass R_n zu einem Ring und g zu einem Homomorphismus wird. Für diesen Ring beweist man ein Analogon von Lemma 1, und g ist der darin vorkommende Homomorphismus im Fall $S = R[x_1, \ldots, x_n]$ und $b_1 = x_1, \ldots, b_n = x_n$. Nun können wir das ursprüngliche Lemma 1 im Fall $S = R_n$ anwenden und erhalten einen Homomorphismus $h : R[x_1, \ldots, x_n] \to R_n$. Die Verkettung $g \circ h$ erfüllt dann ebenso wie die identische Abbildung die Bedingungen von Lemma 1 im Fall $S = R[x_1, \ldots, x_n]$, also muss wegen der Eindeutigkeitsaussage $g \circ h = \mathrm{id}$ sein. Ebenso beweist man, dass $h \circ g = \mathrm{id}$. Somit sind g und h zueinander inverse Homomorphismen und folglich die Ringe $R[x_1, \ldots, x_n]$ und R_n isomorph, so dass die Bezeichnung R_n letztlich überflüssig wird. Nun folgt die erste Behauptung von Satz 2 aus der Bijektivität von g.

Die zweite folgt daraus, dass jeder Term äquivalent zu einem Term in Normalform von nicht höherem Grade ist.

In Anlehnung an das englische Wort „degree“ kürzt man den Grad eines Polynoms f oft mit $\deg f$ ab. Ein Polynom, in dessen Normalform alle Monome den gleichen Grad haben, nennt man *homogenes Polynom*. Man kann jedes Polynom als Summe von homogenen Polynomen schreiben, die man seine homogenen Komponenten nennt. So hat das obige Polynom f beispielsweise die homogenen Komponenten

$$3x^2y^2 + 2xy^3 \qquad \text{und} \qquad 15x^3 + 22x^2y + 8xy^2.$$

Lemma 2. *Für Polynome f und g mit Koeffizienten in einem nullteilerfreien kommutativen Ring R gilt*

$$\deg(f+g) \leq \max\{\deg f, \deg g\}, \qquad \deg(f \cdot g) = \deg f + \deg g$$

und, wenn f ein Polynom in einer Variablen ist,

$$\deg(f \circ g) = \deg f \cdot \deg g.$$

Dabei wird $-\infty \cdot n = -\infty$ festgelegt. Ein Ring heißt *nullteilerfrei*, wenn er keine von Null verschiedenen Elemente a und b hat, für die $ab = 0$ ist.

Beweis. Die erste Behauptung ist klar, da die Monome in der Normalform von $f + g$ durch Zusammenfassen der Monome von f und g mit gleichen Exponenten entstehen. Wenn sich dabei die Monome vom höchsten Grad wegkürzen, so ist die Ungleichung strikt.

Für die zweite Behauptung genügt es, die homogenen Komponenten höchsten Grades von f und g zu betrachten. Darum können wir annehmen, dass f und g selbst homogen sind. Wir ordnen die Exponenten lexikographisch. Kommt $(i_1, \ldots, i_n)$ vor $(i'_1, \ldots, i'_n)$, so kommt $(i_1 + j_1, \ldots, i_n + j_n)$ vor $(i'_1 + j_1, \ldots, i'_n + j_n)$. Somit ergibt sich der lexikographisch höchste Exponent von $f \cdot g$ nur einmal, nämlich als Summe der lexikographisch höchsten Exponenten von f und von g. Der zugehörige Koeffizient von f ist das Produkt der zugehörigen Koeffizienten in f und g, also wegen der Nullteilerfreiheit von R nicht Null.

Für die dritte Behauptung genügt es wieder, den Fall von homogenem f zu betrachten. Da f jetzt nur von einer Variablen abhängt, bedeutet das $f(x) = ax^n$. Es ist also zu zeigen, dass $\deg g^n = n \deg g$ ist. Für $n = -\infty$ gilt dies laut Festlegung. Für $n \geq 0$ verwenden wir die Methode der vollständigen Induktion, die in der Veranstaltung zur Analysis besprochen wurde. Im Fall $n = 0$ ist die Behauptung klar. Wegen $g^{n+1} = g^n \cdot g$ erhalten wir aus der bereits bewiesenen Formel für Produkte $\deg g^{n+1} = \deg g^n + \deg g$. Somit folgt aus der Gültigkeit der Formel für eine Zahl n ihre Gültigkeit für $n + 1$. □

In dieser Vorlesung geht es zwar vorrangig um die Lösung linearer Gleichungen, aber wir benötigen auch Aussagen über nichtlineare algebraische Gleichungen mit einer Variablen. Jede solche Gleichung kann man in der Form $f(x) = 0$ mit einem Polynom f schreiben, und ihre Lösungen im Ring R, aus dem die Koeffizienten von f stammen, nennt man die *Nullstellen* des Polynoms f.

Satz 3. *Ein Polynom vom Grad $n \geq 0$ in einer Variablen mit Koeffizienten in einem nullteilerfreien Ring hat höchstens n Nullstellen.*

Beweis. Ein solches Polynom hat die Normalform

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n.$$

Offensichtlich ist $f(0) = a_0$. Ist 0 eine Nullstelle von f, so gilt

$$f(x) = x(a_1 + a_2 x + \cdots + a_n x^{n-1}).$$

Wir behaupten allgemein: Ist c eine Nullstelle von f, so gibt es ein Polynom g, so dass $f(x) = (x - c)g(x)$. Zum Beweis bemerken wir, dass c genau dann eine Nullstelle von f ist, wenn 0 eine Nullstelle des Polynoms $f_c(y) = f(y+c)$ ist. Nach dem bereits beweisenen Spezialfall gibt es dann ein Polynom h, so dass $f_c(y) = yh(y)$. Nun folgt $f(x) = f_c(x - c) = (x - c)h(x - c)$, und wir können $g(x) = h(x - c)$ setzen.

Die Behauptung des Satzes folgt nun durch vollständige Induktion. Ist $n = 0$, so ist $f(x)$ eine Konstante, die nach Definition des Grades nicht Null ist, also hat f keine Nullstelle. Nun sei $n > 0$, und die Behauptung sei bereits für Polynome vom Grad $n - 1$ bewiesen. Hat f keine Nullstelle, so sind wir fertig. Hat f hingegen eine Nullstelle c, so ist $f(x) = (x - c)g(x)$, und nach Lemma 2 ist $\deg g = n - 1$. Wegen der Nullteilerfreiheit ist jede von c verschiedene Nullstelle von f eine Nullstelle von g, und nach Induktionsvoraussetzung gibt es davon höchstens $n - 1$. □

Es gibt zahlreiche Homomorphismen zwischen Polynomringen.

(i) Wir können jedes Polynom in den Variablen $x_1, \ldots, x_{n-1}$ auch als Polynom in den Variablen $x_1, \ldots, x_n$ auffassen und erhalten so einen Homomorphismus $R[x_1, \ldots, x_{n-1}] \to R[x_1, \ldots, x_n]$.

(ii) Für jedes Element a eines Ringes R erhalten wir einen Homomorphismus $R[x_1, \ldots, x_n] \to R[x_1, \ldots, x_{n-1}]$, indem wir a an Stelle von x_n einsetzen.

(iii) Allgemeiner erhalten wir für jedes Polynom $g \in R[y_1, \ldots, y_m]$ einen Homomorphismus $R[x_1, \ldots, x_n] \to R[x_1, \ldots, x_{n-1}, y_1, \ldots, y_m]$, indem wir x_n durch $g(y_1, \ldots, y_m)$ substituieren.

(iv) Indem wir die Variablen $x_1, \ldots, x_n$ durch die Variablen $y_1, \ldots, y_n$ ersetzen, erhalten wir einen Isomorphismus $R[x_1, \ldots, x_n] \to R[y_1, \ldots, y_n]$.

(v) Die Abbildung, die einem Polynom $f(x, y)$ das Polynom $f(y, x)$ zuordnet, ist ein Automorphismus des Ringes $R[x, y]$.

(vi) Ist X eine Menge und R ein kommutativer Ring, so erhalten wir einen Homomorphismus $R[x] \to R^R$, indem wir jedem Polynom die entsprechende ganzrationale Funktion zuordnen. Ist R unendlich, so ist dies ein Monomorphismus, d. h. ein injektiver Homomorphismus. Dies zeigt man durch vollständige Induktion nach dem Grad unter Verwendung von Satz 3.

3.3 Körper

3.3.1 Definition

Das Gaußsche Eliminationsverfahren ist nicht auf lineare Gleichungssysteme mit Koeffizienten in einem beliebigen kommutativen Ring R anwendbar. Manchmal muss man dabei lineare Gleichungen der Form

$$a = x \cdot b \tag{1}$$

für Elemente a und b von R lösen, wobei $b \neq 0$ ist. Wenn diese Gleichung in R lösbar ist, nennt man a ein *Vielfaches* von b im Ring R und b einen *Teiler* von a im Ring R. Nach dieser Definition wäre jedes Element ein Teiler des Nullelements, aber von einem Nullteiler spricht man nur dann, wenn eine vom Nullelement verschiedene Lösung existiert.

Definition 7. *Ein Körper ist ein kommutativer Ring K, der nicht nur aus dem Nullelement besteht und in dem die Gleichung (1) für beliebige Elemente a und $b \neq 0$ lösbar ist.*

Wie auch schon im Fall des Ringaxioms (vi) sieht man leicht, dass es im Fall $a = 1$ nur eine Lösung gibt. Man nennt sie das inverse Element von b. Ist nun a beliebig und sind c und c' Lösungen, so ist $bc = bc'$, und durch Multiplikation beider Seiten mit dem Inversen von b folgt $c = c'$. Somit ist die Lösung immer eindeutig bestimmt. Man bezeichnet sie mit $a : b$ und nennt sie den *Quotienten* von a und b.

Für die so definierte *Division* gelten die Rechenregeln

$$a : b = a \cdot (1 : b), \qquad (c \cdot a) : (c \cdot b) = a : b, \qquad (a + b) : c = a : c + b : c.$$

Die erste folgt aus der Tatsache, dass $a \cdot (1 : b) \cdot b = a \cdot 1 = a$, die zweite ergibt sich, wenn wir beide Seiten von (1) von links mit c multiplizieren, und die dritte folgt aus der Gleichung $(a : c + b : c) \cdot c = (a : c) \cdot c + (b : c) \cdot c = a + b$, die ihrerseits aus dem Distributivgesetz folgt.

Hat b das Inverse d, so setzt man für jede positive ganze Zahl n

$$b^n = \underbrace{b \cdots b}_{n}, \qquad b^{-n} = \underbrace{d \cdots d}_{n}, \qquad b^0 = 1.$$

Dann gelten für alle invertierbaren Elemente a und b und alle ganzzahligen Exponenten k und l die Potenzgesetze

$$a^{k+l} = a^k \cdot a^l, \qquad a^{kl} = (a^k)^l, \qquad (a \cdot b)^k = a^k \cdot b^k.$$

Als Bezeichung für das inverse Element von b benutzt man meist b^{-1}.

Jeder Körper ist nullteilerfrei, denn wenn $ab = 0$ ist, aber $a \neq 0$, so können wir beide Seiten der Gleichung mit dem Inversen von a multiplizieren und erhalten $b = 0$.

Der Ring $\{g, u\}$ mit zwei Elementen ist offenbar ein Körper, der Ring $\mathbb{Z}$ der ganzen Zahlen hingegen nicht.

Ein *Schiefkörper* ist ein (nicht notwendig kommutativer) Ring, der nicht nur aus dem Nullelement besteht und in dem für beliebige Elemente a und $b \neq 0$ die Gleichungen

$$a = x \cdot b, \qquad a = b \cdot y$$

lösbar sind. Da keine Kommutativität vorausgesetzt wird, brauchen die Lösungen nicht übereinzustimmen. Im Unterschied zum kommutativen Fall gibt es traditionell kein linkes bzw. rechtes Divisionszeichen, sinnvoll wären wohl die Bezeichnungen $x = a/b$ und $y = b\backslash a$. Die linken und rechten Inversen eines Elements stimmen aber überein, denn aus $1 = x \cdot b$ und $1 = b \cdot y$ folgt $x = x \cdot (b \cdot y) = (x \cdot b) \cdot y = y$. Somit kann man das Inverse von b auch hier mit b^{-1} bezeichnen.

3.3.2 Quotientenkörper

Es gibt eine allgemeine Methode, mit der man einen beliebigen nullteilerfreien kommutativen Ring in einen Körper einbetten kann, und im Fall des Ringes $\mathbb{Z}$ ergibt sie den Körper $\mathbb{Q}$ der rationalen Zahlen. Da dies in der Schule nicht klar herausgearbeitet wird, wollen wir noch einmal darauf eingehen.

Angenommen, wir hätten bereits einen Körper K, der den gegebenen Ring R enthält. Die Teilmenge aller Quotienten $a : b$ von Ringelementen mit $b \neq 0$ bildet einen Teilring von K, denn sie enthält $0 : 1$ und $1 : 1$ und für beliebige $a : b$ und $c : d$ auch

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

Zum Beweis der Gleichungen müssen wir nur die linke Seite mit bd multiplizieren und das Ergebnis mit Hilfe der Axiome in $ad + bc$ bzw. ac umformen. Außerdem ist $(-a) : b$ das entgegengesetzte Element zu $a : b$. Der besagte Teilring ist sogar ein Körper, denn das Inverse von $a : b$ ist $b : a$. Wenn wir also R in einen Körper einbetten wollten, so würde dieser Teilring genügen, und wir können annehmen, dass er gleich K ist. Mit anderen Worten, die Abbildung $R \times (R \setminus \{0\}) \to K$, die jedem Paar (a, b) den Quotienten $a : b$ zuordnet, ist surjektiv.

Diese Abbildung ist im Allgemeinen nicht injektiv. Ist nämlich

$$a : b = c : d,$$

so erhalten wir durch Mulitplikation beider Seiten mit $b \cdot d$, dass

$$a \cdot d = b \cdot c.$$

Es gilt aber auch die Umkehrung, wie wir durch Multiplikation mit dem Inversen von $b \cdot d$ feststellen, welches wegen $b \cdot d \neq 0$ existiert. Wir können somit innerhalb des Ringes R überprüfen, ob zwei Paare auf das selbe Körperelement abgebildet werden.

Ist der Körper K nicht gegeben, so konstruieren wir ihn als Menge von Äquivalenzklassen.

Definition 8. (i) *Wir nennen zwei Paare (a, b) und (c, d) aus $R \times (R \setminus \{0\})$ quotientengleich, abgekürzt $(a, b) \sim (c, d)$, wenn $a \cdot d = b \cdot c$ ist.*

(ii) *Wir definieren Operationen auf $R \times (R \setminus \{0\})$ wie folgt:*

$$(a, b) + (c, d) = (a \cdot d + b \cdot c, b \cdot d), \qquad (a, b) \cdot (c, d) = (a \cdot c, b \cdot d).$$

Damit die Menge $R \times (R \setminus \{0\})$ tatsächlich in disjunkte Äquivalenzklassen zerfällt, muss man prüfen, dass die Quotientengleichheit eine Äquivalenzrelation ist (Übungsaufgabe). Damit durch (ii) auch Operationen auf der Menge der Äquivalenzklassen entstehen, ist zu prüfen, dass

$$\begin{aligned}(a, b) \sim (a', b'),\ (c, d) \sim (c', d') \Rightarrow (a, b) + (c, d) &\sim (a', b') + (c', d'),\\ (a, b) \cdot (c, d) &\sim (a', b') \cdot (c', d').\end{aligned}$$

Dies überlassen wir den Lesern. Man bezeichnet die Äquivalenzklasse des Paares (a, b) mit $\frac{a}{b}$. Die Operationen mit diesen Klassen nehmen dann die gewohnte Form

$$\frac{a}{b} + \frac{c}{d} = \frac{a \cdot d + b \cdot c}{b \cdot d}, \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d}$$

an.

Satz 4. *Es sei R ein nullteilerfreier kommutativer Ring. Die Menge der Äquivalenzklassen in $R \times (R \setminus \{0\})$ mit den oben definierten Operationen ist ein Körper K. Die Abbildung $a \mapsto \frac{a}{1}$ ist ein injektiver Homomorphismus $R \to K$.*

Zum Beweis des Satzes muss man sämtliche Körperaxiome nachprüfen. Dies ist nicht schwer und soll hier nicht ausgeführt werden.

Man nennt K den Quotientenkörper von R. Im Fall des Ringes $\mathbb{Z}$ ist, wie gesagt, der Quotientenkörper gleich $\mathbb{Q}$. Meist „identifiziert“ man jedes Ringelement a mit seinem Bild $\frac{a}{1}$, um R als Teilring eines Körpers darzustellen (was logisch nicht ganz einwandfrei ist). Nach der genannten Identifikation bedeuten $a : b$ und $\frac{a}{b}$ für Elemente a und $b \neq 0$ von R das Selbe. Dies hat dazu geführt, dass der Bruchstrich als Synonym für das Divisionszeichen in einem beliebigen Körper „missbraucht“ wird, zumal damit Quotienten von umfangreichen Termen platzsparend geschrieben werden können.

In der Analysis betrachtet man angeordnete Körper. Eine Anordnung auf einem Körper K ist eine Ordnung „$\geq$“ mit der zusätzlichen Eigenschaft der Monotonie:

> Für alle x, y und $z \in K$ gilt: Ist $x \geq y$, so gilt $x + z \geq y + z$, und ist außerdem $z \geq 0$, so gilt $x \cdot z \geq y \cdot z$.

Man definiert für ganze Zahlen a, $b > 0$, c und $d > 0$:

$$\frac{a}{b} \geq \frac{c}{d}, \qquad \text{wenn} \qquad a \cdot d \geq b \cdot c.$$

Es ist nicht schwer nachzuprüfen, dass dies nicht von der Wahl der Vertreter abhängt und eine Anordnung auf dem Körper $\mathbb{Q}$ definiert.

Man kann die selbe Konstruktion auch auf den Ring $R[x_1, \ldots, x_n]$ der Polynome mit Koeffizienten in einem nullteilerfreien Ring R anwenden. Der entstehende Quotientenkörper enthält den Quotientenkörper K von R, und man bekommt das selbe Ergebnis, wenn man gleich mit dem Ring $K[x_1, \ldots, x_n]$ beginnt. Man nennt seinen Quotientenkörper den Körper der rationalen Funktionen mit Koeffizienten in K und bezeichnet ihn mit $K(x_1, \ldots, x_n)$. Genaugenommen muss man seine Elemente wieder von den Funktionen unterscheiden, aber dafür gibt es traditionell keine verschiedenen Wörter.

Dieser Körper ist uns schon begegnet: Die in der Cramerschen Regel angegebenen Lösungen sind rationale Funktionen der Koeffizienten des Gleichungssystems, und auch der Gesamtwiderstand in Beispiel 2.2 ist eine rationale Funktion der Teilwiderstände. Die Bezeichnung „ganzrationale Funktionen“ erklärt sich daraus, dass diese unter den rationalen Funktionen die selbe Rolle spielen wie die ganzen Zahlen unter den rationalen Zahlen. Man könnte die Elemente von $K(x_1, \ldots, x_n)$ auch als Äquivalenzklassen von Termen einführen, in denen alle vier Grundrechenarten vorkommen, wobei aber nicht durch Terme dividiert werden darf, die äquivalent zu Null sind.

3.3.3 Komplexe Zahlen

In der Analysis wird der Körper $\mathbb{R}$ der reellen Zahlen eingeführt. Er ist ein angeordneter Körper, für eine beliebige reelle Zahl x gilt also wegen der Totalität $x \geq 0$ oder $0 \geq x$. Im Fall $0 \geq x$ folgt mit der Monotonie $0+(-x) \geq x+(-x)$, also $-x \geq 0$, und schließlich $(-x) \cdot (-x) \geq 0 \cdot (-x)$. Für alle x gilt somit $x^2 \geq 0$, und wegen $1 \neq 0$ ist $1^2 > 0$. Die quadratische Gleichung

$$x^2 + 1 = 0$$

hat darum keine Lösung.

Wir nehmen einmal an, dass es einen Körper gibt, der $\mathbb{R}$ als Teilkörper enthält und in dem diese Gleichung eine Lösung i besitzt. Dann ist natürlich auch $-i$ eine Lösung, und mehr Lösungen kann eine quadratische Gleichung nach Satz 3 nicht haben. Nun gilt für alle reellen Zahlen x, y, u und v

$$\begin{aligned}(u+iv)+(x+iy) &= (u+x)+i(v+y),\\ (u+iv)\cdot(x+iy) &= (ux-vy)+i(uy+vx).\end{aligned}$$

Die Teilmenge aller Elemente der Form $x+iy$ ist also abgeschlossen unter Addition und Multiplikation. Könnte man ein Element auf zwei Arten darstellen, z. B.

$$u+iv = x+iy,$$

so wäre $u-x = i(y-v)$, also

$$(u-x)^2 + (v-y)^2 = 0$$

und somit $u = x$ und $v = y$. In dem Erweiterungskörper hat für gegebene u und v die Gleichung

$$(u+iv)+(x+iy) = 0$$

eine Lösung, nämlich $x = -u$, $y = -v$, und für $u+iv \neq 0$ hat auch die Gleichung

$$(u+iv)\cdot(x+iy) = 1$$

eine Lösung. Sie ist ja äquivalent zu dem Gleichungssystem

$$\begin{cases} ux - vy = 1\\ uy + vx = 0\end{cases}$$

und man findet z. B. mit Hilfe der Cramerschen Regel

$$x = \frac{u}{u^2+v^2}, \qquad y = \frac{-v}{u^2+v^2},$$

wobei $u^2 + v^2 > 0$, weil nach Voraussetzung u und v nicht beide gleich Null sind. Es folgt, dass die Teilmenge K der Elemente der Form $x + iy$ bereits einen Körper bildet, in dem die obige quadratische Gleichung lösbar ist, und dass die durch $h(x, y) = x + iy$ Abbildung $h : \mathbb{R} \times \mathbb{R} \to K$ bijektiv ist.

Haben wir einen weiteren solchen Erweiterungskörper K' und die entsprechende Abbildung $h' : \mathbb{R}\times\mathbb{R} \to K'$, so sind die Abbildungen $h' \circ h^{-1} : K \to K'$ und $h \circ h'^{-1} : K' \to K$ zueinander inverse Homomorphismen, also sind K und K' isomorph. Der Isomorphismus zwischen ihnen ist aber nicht eindeutig bestimmt, denn man könnte ihn z. B. mit einem Automoprhismus g von K verketten, der jedes Element des Teilkörpers $\mathbb{R}$ fixiert (d. h., auf sich selbst abbildet). Wir können alle solchen Automorphismen leicht bestimmen. Wegen $g(z^2 + 1) = g(z)^2 + 1$ ist auch $g(i)$ eine Lösung der eingangs betrachteten quadratischen Gleichung. Es gilt also $g(i) = i$ oder $g(i) = -i$. Wegen $g(x + iy) = g(x) + g(i)g(y)$ ist im ersten Fall für alle x, $y \in \mathbb{R}$

$$g(x + iy) = x + iy,$$

während im zweiten Fall für alle x, $y \in \mathbb{R}$ gilt

$$g(x + iy) = x - iy.$$

Einer dieser Automorphismen ist also die identische Abbildung, den anderen nennt man die Konjugation.

Alles Gesagte beruhte auf einer unbewiesenen Annahme, aber wir wissen nun, wie wir den fraglichen Körper konstruieren können.

Definition 9. *Es sei $\mathbb{C}$ die Menge aller geordneten Paare reeller Zahlen mit den Operationen*

$$\begin{aligned}(u, v) + (x, y) &= (u + x, v + y),\\ (u, v) \cdot (x, y) &= (ux - vy, uy + vx).\end{aligned}$$

Satz 5. *Die Menge $\mathbb{C}$ mit diesen Operationen ist ein Körper mit dem Nullelement $(0, 0)$ und dem Einselement $(1, 0)$. Die Abbildung $x \mapsto (x, 0)$ ist ein injektiver Homomorphismus des Körpers $\mathbb{R}$ in den Körper $\mathbb{C}$.*

Beweis. Wir führen die Nachprüfung der Körperaxiome nur am Beispiel des Assoziativgesetzes der Multiplikation vor:

$$\begin{aligned}(s, t) \cdot \big((u, v) \cdot (x, y)\big) &= (s, t) \cdot (ux - vy, uy + vx)\\ &= \big(s(ux - vy) - t(uy + vx), s(uy + vx) + t(ux - vy)\big),\\ \big((s, t) \cdot (u, v)\big) \cdot (x, y) &= (su - tv, sv + tu) \cdot (x, y)\\ &= \big((su - tv)x - (sv + tu)y, (su - tv)y + (sv + tu)x\big).\end{aligned}$$

Die geordneten Paare auf der rechten Seite stimmen überein.

Die motivierenden Betrachtungen vor der Definition lassen sich in die Sprache geordneter Paare umformulieren und zeigen die Existenz der entgegengesetzten Zahl und des Inversen, woraus die Existenz der Differenz und des Quotienten folgt. □

Wir nennen die Elemente des Körpers $\mathbb{C}$ komplexe Zahlen. Wir identifizieren $\mathbb{R}$ vermittels des im Satz erwähnten Homomorphismus mit einem Teilkörper von $\mathbb{C}$. Das Element $(0, 1)$ bezeichnen wir mit i und nennen es imaginäre Einheit. Die eingangs gemachte Annahme ist jetzt vollständig realisiert. Man veranschaulicht komplexe Zahlen z als Punkte mit Koordinaten $\operatorname{Re} z$ und $\operatorname{Im} z$ in einer Ebene (benannt nach Gauß oder Argand).

Jede komplexe Zahl lässt sich eindeutig in der Form $z = x + iy$ schreiben, und wir werden die Schreibweise als Paar nicht mehr benutzen. Man bezeichnet x als den Realteil und y als den Imaginärteil von z, abgekürzt $x = \operatorname{Re} z$ und $y = \operatorname{Im} z$, und mann nennt $\bar{z} = x - iy$ die zu z konjugierte Zahl.

Durch die Erweiterung des Körpers $\mathbb{R}$ zu $\mathbb{C}$ wurde nicht nur die eingangs genannte quadratische Gleichung lösbar, sondern alle algebraischen Gleichungen mit einer Unbekannten:

Satz 6 (Hauptsatz der Algebra)**.** *Jedes Polynom in einer Variablen mit komplexen Koeffizienten hat eine komplexe Nullstelle.*

Ungeachtet seines Namens kann dieser Satz nicht ohne Analysis bewiesen werden, wie ja schon die Konstruktion der reellen Zahlen analytische Begriffe erfordert.

Folgerung 1. *Ist f ein Polynom n-ten Grades in einer Variablen mit komplexen Koeffizienten, so gibt es komplexe Zahlen $a \neq 0$ und $z_1, \ldots, z_n$, so dass*

$$f(z) = a(z - z_1) \cdots (z - z_n).$$

Für $n = 0$ ist in der Tat $f(z) = a$. Für $n > 0$ hat f nach dem Hauptsatz wenigstens eine Nullstelle z_1, und laut einem Zwischenergebnis aus dem Beweis von Satz 3 gibt es dann ein Polynom g, so dass $f(z) = (z - z_1)g(z)$. Nun folgt die Behauptung durch vollständige Induktion.

3.4 Ringe (Fortsetzung)

Ausgestattet mit Kenntnissen über Körper kehren noch einmal zum vorigen Thema zurück.

3.4.1 Euklidische Ringe

In manchen Ringen kann man eine Division mit Rest einführen.

Definition A. *Ein* euklidischer Ring *ist ein kommutativer Ring R, so dass eine Funktion $\varphi : R \to \mathbb{N}$ mit folgenden Eigenschaften existiert:*

(i) *Für $a \in R$ gilt genau dann $a = 0$, wenn $\varphi(a) = 0$ ist.*

(ii) *Für $a, b \in R$ gilt $\varphi(a)\varphi(b) = \varphi(ab)$.*

(iii) *Für $a, b \in R$ und $b \neq 0$ gibt es $q, r \in R$, so dass*

$$a = qb + r, \qquad \varphi(r) < \varphi(b).$$

Man nennt r einen Rest bei der Division von a durch b. Natürlich ist q nur dann der Quotient, wenn $r = 0$ ist. Aus (i) und (ii) folgt die Nullteilerfreiheit von R und

(ii′) Für $a, b \in R$ und $b \neq 0$ gilt $\varphi(a) \leq \varphi(ab)$.

Viele Autoren verlangen nur diese schwächere Eigenschaft.

Lemma A. (i) *Der Ring $\mathbb{Z}$ der ganzen Zahlen ist ein euklidischer Ring mit der Funktion*

$$\varphi(a) = |a|.$$

Für positive a und b kann man in Eigenschaft (iii) sogar $0 \leq r < b$ verlangen, und dann sind q und r eindeutig bestimmt.

(ii) *Der Ring $R = K[x]$ der Polynome in einer Variablen mit Koeffizienten in einem Körper K ist ein euklidischer Ring mit der Funktion*

$$\varphi(f) = 2^{\deg f},$$

und in (iii) sind q und r eindeutig bestimmt.

In (ii) hätte man an Stelle von 2 jede natürliche Zahl außer 0 und 1 als Basis benutzen können. Es gilt natürlich genau dann $\varphi(f) \leq \varphi(g)$, wenn $\deg f \leq \deg g$, aber die Einführung von φ vereinheitlicht beide Fälle. Polynomringe in mehr als einer Variablen sind übrigens keine euklidischen Ringe.

Beweis. Eigenschaften (i) und (ii) sind offensichtlich erfüllt. Wir beweisen Eigenschaft (iii) bei festem b durch vollständige Induktion nach $\varphi(a)$. Für $\varphi(a) = 0$ ist $a = 0$, und wir können $q = 0$ und $r = 0$ setzen.

Im Fall $R = \mathbb{Z}$ genügt es, $a \geq 0$ und $b > 0$ zu betrachten. Angenommen, $a > 0$, und (iii) gilt bereits für kleinere Zahlen. Setzen wir $\tilde{a} = a - 1$, so gibt es nach Induktionsvoraussetzung $\tilde{q}$ und $\tilde{r}$, so dass $\tilde{a} = \tilde{q}b + \tilde{r}$ und $0 \leq \tilde{r} < b$. Ist $\tilde{r} + 1 < b$, so setzen wir $q = \tilde{q}$ und $r = \tilde{r} + 1$. Andernfalls ist $\tilde{r} = b$, uns wir setzen $q = \tilde{q} + 1$, $r = \tilde{r}$. Es folgt $a = qb + r$ und $r < b$.

Im Fall $R = K[x]$ ersetzen wir die Bezeichnungen a und b durch f und g. Angenommen, f ist nicht das Nullpolynom, und (iii) gilt bereits für Polynome von kleinerem Grad. Ist $\varphi(f) < \varphi(g)$, so setzen wir $q = 0$, $r = f$. Andernfalls ist $\deg f \geq \deg g$, also gibt es ein $n \in \mathbb{N}$ und $c \in K^\times$, so dass $f(x)$ und $cx^n g(x)$ die gleichen höchsten homogenen Komponenten haben. Setzen wir $\tilde{f}(x) = f(x) - cx^n g(x)$, so gilt $\varphi(\tilde{f}) < \varphi(f)$, und nach der Induktionsvoraussetzung gibt es $\tilde{q}$ und r, so dass $\tilde{f} = \tilde{q}g + r$ und $\varphi(r) < \varphi(g)$. Setzen wir nun $q(x) = cx^n + \tilde{q}(x)$, so folgt $f = qg + r$. □

Der Beweis des Lemmas im Fall von Polynomen ist der erste Schritt des Divisionsalgorithmus, auf den wir aber angesichts seiner Bekanntheit nicht näher eingehen.

Für Elemente a und $b \neq 0$ eines beliebigen kommutativen Ringes R schreiben wir $b \mid a$ (gelesen b *teilt* a), wenn a ein *Vielfaches* von b ist, d. h. wenn es ein $q \in R$ gibt, so dass $a = qb$. Die Einschränkung dieser Relation auf $R \setminus \{0\}$ ist eine Quasiordnung, d. h. transitiv und reflexiv, aber nicht antisymmetrisch. Elemente a und b mit den Eigenschaften $a \mid b$ und $b \mid a$ nennt man *assoziiert*, abgekürzt $a \sim b$. In diesem Fall haben wir $a = qb$ und $b = ua$, und mit der Nullteilerfreiheit folgt $qu = 1$, d. h. q und u sind invertierbar. Invertierbare Elemente teilen jedes Element von R. Die Assoziiertheit ist offensichtlich eine Äquivalenzrelation auf $R \setminus \{0\}$.

In den Beispielen aus Lemma A gibt es eine Teilmenge S von $R \setminus \{0\}$, die je einen Vertreter aus jeder Äquivalenzklasse enthält. Im Fall des Ringes $\mathbb{Z}$ ist S die Menge der positiven ganzen Zahlen, im Fall des Ringes $K[x]$ ist S die Menge der *normierten Polynome*, d. h. der Polynome mit höchstem Koeffizienten 1.

Ist M eine Teilmenge von R, so nennt man ein Element d von R einen *gemeinsamen Teiler* der Elemente von M, wenn für jedes $a \in M$ gilt $d \mid a$. Analog definiert man *gemeinsame Vielfache*. Obwohl die Teilerrelation keine Halbordnung ist, spricht man trotzdem von *größten gemeinsamen Teilern* und *kleinsten gemeinsamen Vielfachen*. Wenn sie existieren, sind sie nur bis auf Assoziiertheit bestimmt, aber innerhalb von S sind sie eindeutig bestimmt und im Fall der natürlichen Zahlen sind sie auch die größten bzw. kleinsten

im gewöhnlichen Sinne.

Meist wendet man diese Begriffe auf Teilmengen $\{a, b\}$ von zwei Elementen an. Man sagt, a und b seien *teilerfremd*, wenn ihre einzigen (also auch ihre größten) gemeinsamen Teiler die invertierbaren Elemente sind.

Lemma B. *Es sei R ein euklidischer Ring.*

(i) *Für beliebige Elemente a und b, die nicht beide Null sind, gibt es einen größten gemeinsamen Teiler d und ein kleinstes gemeinsames Vielfaches m, und es gibt x, $y \in R$ mit der Eigenschaft*

$$d = xa + yb.$$

(ii) *Sind a und b teilerfremd und ist $a \mid c$ und $b \mid c$, so ist $ab \mid c$.*

(iii) *Sind a und b teilerfremd und ist $a \mid bc$, so ist $a \mid c$.*

(iv) *Ist d ein größter gemeinsamer Teiler und m ein kleinstes gemeinsames Vielfaches von a und b, so gilt*

$$ab \sim dm.$$

Beweis. Wir beweisen (i) durch vollständige Induktion nach $\min(\varphi(a), \varphi(b))$, wobei wir annehmen können, dass $\varphi(a) \geq \varphi(b)$ ist. Sie gilt für $\varphi(b) = 0$, weil dann $b = 0$ ist und a ein größter gemeinsamer Teiler. Nun sei die Behauptung für kleinere Werte als $\varphi(b)$ beweisen. Wir wählen q und r wie in Definition A(iii). Jeder gemeinsame Teiler von a und b ist auch ein Teiler von r, also ein gemeinsamer Teiler von b und r. Umgekehrt ist jeder gemeinsame Teiler von b und r auch ein Teiler von a. Wegen $\min(\varphi(r), \varphi(b)) < \varphi(b)$ gibt es nach Induktionsvoraussetzung einen größten gemeinsamen Teiler d und Elemente x und z, so dass

$$d = xr + zb.$$

Es folgt

$$d = x(a - qb) + zb = xa + (z - xq)b,$$

und wir können $y = z - xq$ setzen. Die Existenz von m beweisen wir zusammen mit Behauptung (iv).

Sind a und b teilerfremd, so gibt es nach Teil (i) Elemente x und y, so dass

$$xa + yb = 1.$$

Multiplizieren wir beide Seiten mit c, so erhalten wir

$$xac + ybc = c.$$

In der Situation von (ii) ist jeder Summand auf der linken Seite ein Vielfaches von ab und in der Situation von (iii) ein Vielfaches von a.

In der Situation von (i) gibt es wegen $d \mid a$ und $d \mid b$ Elemente u und v, so dass $a = du$ und $b = dv$. Für jeden gemeinsamen Teiler w von u und v ist dw ein gemeinsamer Teiler von a und b. Da d der größte ist, muss dann $w \sim 1$ sein, also sind u und v teilerfremd.

Ist nun n ein gemeinsames Vielfaches von a und b, so gibt es Elemente x und y, so dass $n = ax$ und $n = by$. Es folgt $dux = dvy$ und wegen der Nullteilerfreiheit $ux = vy$. Mit Behauptung (iii) ergibt sich $v \mid x$ und $u \mid y$, also $duv \mid n$. Andererseits ist duv selbst ein gemeinsames Vielfaches von a und b. Somit existiert ein kleinstes gemeinsames Vielfaches duv. Jedes andere ist dazu assoziiert, d. h. $duv \sim m$. Multiplikation beider Seiten mit d ergibt die Behauptung. □

Die Bezeichnung euklidischer Ring rührt daher, dass man in einem solchen Ring mit Hilfe des euklidischen Algorithmus einen größten gemeinsamen Teiler von zwei Elementen a und b finden kann. Der erste Schritt des Algorithmus ist im obigen Beweis enthalten. Um das Verfahren zu iterieren, bezeichnen wir die gegebenen Elemente a und b mit a_0 und a_1. Durch Division mit Rest findet man nacheinander Elemente a_2, a_3, ... mit den Eigenschaften

$$\begin{aligned} a_0 &= q_1 a_1 + a_2, & \varphi(a_2) &< \varphi(a_1), \\ a_1 &= q_2 a_2 + a_3, & \varphi(a_3) &< \varphi(a_2), \\ &\dots \\ a_{k-2} &= q_{k-1} a_{k-1} + a_k, & \varphi(a_k) &< \varphi(a_{k-1}), \\ a_{k-1} &= q_k a_k. \end{aligned}$$

Die streng monoton fallende Folge der natürlichen Zahlen $\varphi(a_i)$ muss irgendwann bei Null ankommen, und aus $\varphi(a_{k+1}) = 0$ folgt $a_{k+1} = 0$. Das letzte von Null verschiedene Glied a_k ist dann ein größter gemeinsamer Teiler. Lösen wir die Gleichungen nach den Resten auf, also

$$\begin{aligned} a_2 &= a_0 - q_1 a_1, \\ a_3 &= a_1 - q_2 a_2, \\ &\dots \\ a_{k-1} &= a_{k-3} - q_{k-2} a_{k-2} \\ a_k &= a_{k-2} - q_{k-1} a_{k-1} \end{aligned}$$

und eliminieren dann in der letzten Gleichung nacheinander die Variablen a_{k-1}, a_{k-2},..., a_2, so erhalten wir die Darstellung des größten gemeinsamen Teilers a_k durch a_0 und a_1 wie in Lemma B(i).

Hat man einen größten gemeinsamen Teiler bestimmt, so erhält man ein kleinstes gemeinsames Vielfaches mit Lemma B(iv).

3.4.2 Primelemente

Ein Element p eines kommutativen Ringes heißt *Primelement*, wenn es nicht assoziiert zu 1 ist und jeder Teiler von p assoziiert zu p oder zu 1 ist. Zwei Primelemente sind entweder assoziiert oder teilerfremd. Die positiven Primelemente des Ringes der ganzen Zahlen nennt man auch *Primzahlen*, die Primelemente eines Polynomringes nennt man auch *irreduzible Polynome*.

Wenn ein irreduzibles Polynom eine Nullstelle hat, so hat es nach dem Beweis von Satz 3 einen Teiler vom Grad 1, muss also selbst den Grad 1 haben. Nach Lemma 2 ist jedes Polynom vom Grad 1 irreduzibel. Nach Satz 6 gibt es im Ring $\mathbb{C}[x]$ keine anderen irreduziblen Polynome. Wir behaupten, dass die weiteren irreduzible Polynome in $\mathbb{R}[x]$ die quadratischen Polynome ohne Nullstelle sind. Für jede imaginäre Nullstelle z_0 eines Polynoms $f(x) \in \mathbb{R}[x]$ ist nämlich auch $\bar{z}_0$ eine Nullstelle, also ist $f(x) = (x - z_0)(x - \bar{z}_0)h(x)$ mit $h \in \mathbb{C}[x]$. Das Gleiche gilt auch mit $\bar{h}$ an Stelle von h, und wegen der Nullteilerfreiheit ist $\bar{h} = h$, also $h(x) \in \mathbb{R}[x]$, d. h. $f(x)$ ist in $\mathbb{R}[x]$ durch ein quadratisches Polynom teilbar. Ist f irreduzibel, so ist es zu diesem assoziiert.

Satz A. *Es sei R ein euklidischer Ring, in dem jedes Element c mit der Eigenschaft $\varphi(c) = 1$ invertierbar ist. Dann gibt es für jedes Element $a \neq 0$ Primelemente $p_1, \ldots, p_k$ von R, so dass*

$$a \sim p_1 \cdots p_k.$$

Haben wir eine weitere derartige Zerlegung

$$a \sim q_1 \cdots q_l,$$

so ist $k = l$, und es gibt eine Permutation $\pi \in S_k$, so dass für alle i gilt

$$p_i \sim q_{\pi(i)}.$$

Das angegebene Produkt nennt man die *Primfaktorzerlegung* von a. Die Zusatzbedingung an den euklidischen Ring R ist in unseren beiden Beispielen erfüllt. Der Satz gilt auch ohne sie, ist dann aber schwerer zu beweisen.

Beweis. Wir beweisen die Existenz der Zerlegung durch vollständige Induktion nach $\varphi(a)$. Für $\varphi(a) = 1$ ist nichts zu beweisen, wobei wir ein Produkt mit null Faktoren als 1 interpretieren. Nun sei $\varphi(a) > 1$, und die Behauptung sei für Elemente mit kleineren Werten von φ bereits bewiesen. Ist a ein

Primelement, so setzen wir $k = 1$ und $a = p_1$. Andernfalls gibt es Elemente b und c, die weder zu a noch zu 1 assoziiert sind, so dass $a = bc$. Wäre $\varphi(b) = \varphi(a)$ oder $\varphi(c) = \varphi(a)$, so wäre wegen $\varphi(a) = \varphi(b)\varphi(c)$ dann $\varphi(c) = 1$ oder $\varphi(b) = 1$, also einer der Faktoren assoziiert zu a und der andere zu 1. Somit können wir auf b und c die Induktionsvoraussetzung anwenden und ihre Zerlegungen zu einer Zerlegung von a zusammenfügen.

Wir beweisen die Eindeutigkeit ebenfalls durch vollständige Induktion nach $\varphi(a)$. Für $\varphi(a) = 1$ muss $k = 0$ sein, weil ein Primelement p nicht invertierbar ist und mit unserer Voraussetzung $\varphi(p) > 1$ folgt. Nun sei $\varphi(a) > 1$ und die Behauptung für Elemente mit kleinerem Wert von φ bereits bewiesen. Angenommen, es gibt zwei Zerlegungen wie im Satz. Dann ist $q_l \mid a$, und durch mehrmalige Anwendung von Lemma B(iii) finden wir ein i, so dass $q_l \sim p_i$. Mit der Nullteilerfreiheit folgt

$$p_1 \cdots p_{i-1} \cdot p_{i+1} \cdots p_k \sim q_1 \cdots q_{l-1}.$$

Wegen $\varphi(p_i) > 1$ und $\varphi(q_l) > 1$ ist der Wert von φ auf beiden Seiten kleiner als $\varphi(a)$, und nach Induktionsvoraussetzung stimmen beide Zerlegungen bis auf die Reihenfolge und Assoziiertheit überein. □

Folgerung A. *Zwei Elemente eines euklidischen Ringes sind genau dann teilerfremd, die keinen gemeinsamen Primteiler haben.*

Dabei meinen wir mit *Primteiler* einen Teiler, der ein Primelement ist. Nach dem Satz hat nämlich jedes nicht invertierbare Element einen Primteiler.

Folgerung B. *Es sei R ein euklidischer Ring wie oben und K sein Quotientenkörper. Für jedes Element $x \neq 0$ von K gibt es paarweise nicht assoziierte Primelemente $p_1, \ldots, p_k$ und ganze Zahlen $n_1, \ldots, n_k$, so dass*

$$x \sim p_1^{n_1} \cdots p_k^{n_k}.$$

Das Element x liegt genau dann in R, wenn alle n_i nichtnegativ sind. Haben wir eine weitere derartige Zerlegung

$$x \sim q_1^{m_1} \cdots q_l^{m_l},$$

so ist $k = l$, und es gibt eine Permutation $\pi \in S_k$, so dass für alle i gilt

$$p_i \sim q_{\pi(i)}, \qquad n_i = m_{\pi(i)}.$$

Dabei bedeutet $x \sim y$ jetzt, dass $x/y \in R^\times$ ist.

Beweis. In Satz A können wir assoziierte Primelemente unter den p_i durch das selbe Primelement ersetzen und erhalten die Zerlegung für Elemente von $R \setminus \{0\}$ mit nichtnegativen Exponenten. Wenden dies auf zwei Elemente a und b an, so erhalten wir eine Zerlegung von $x = \frac{a}{b}$.

Haben wir zwei Zerlegungen des selben Elements x, so können wir annehmen, dass in beiden die selben Primelemente vorkommen, da man weitere mit dem Exponenten 0 hinzufügen kann. Schreiben wir beide Seiten als Brüche und wenden Definition 8 an, so erhalten wir eine Gleichheit in R, und die Eindeutigkeit der Zerlegung bis auf Vertauschung und Assoziiertheit folgt aus Satz A. Aus dieser Eindeutigkeit folgt auch das Kriterium der Zugehörigkeit zu R. □

Ist $R = \mathbb{Z}$ oder $R = K[x]$ und bezeichnet man mit P die Menge der Primelemente in der Menge S der Repräsentanten von Äquivalenzklassen, dann definiert jedes Element $x \neq 0$ von K eine Abbildung $n : P \to \mathbb{Z}$, die nur auf endlich vielen Elementen von Null verschieden ist, so dass

$$x \sim \prod_{p \in P} p^{n_p}.$$

Satz B. *Es sei R ein euklidischer Ring wie oben und K sein Quotientenkörper. Dann lässt sich jedes Element von K in der Form*

$$x = q + \sum_{i=1}^{k} \sum_{j=1}^{n_i} \frac{r_{ij}}{p_i^j}$$

schreiben, wobei $q \in R$ ist, $p_1, \ldots, p_k$ paarweise nicht assoziierte Primelemente von R sind und die r_{ij} Elemente von R mit der Eigenschaft $\varphi(r_{ij}) < \varphi(p_i)$ sind.

Die Brüche auf der rechten Seite nennt man die *Partialbrüche* des gegebenen Elements von K. Sie sind übrigens eindeutig bestimmt, wenn man verlangt, dass die p_i in S liegen und die r_{ij} in einem Repräsentantensystem von Resten bezüglich p_i wie in Lemma A liegen.

Beweis. Im Fall $x = 0$ setzen wir $q = 0$ und $k = 0$. Andernfalls ist $x = \frac{a}{b}$, wobei $a, b \in R \setminus \{0\}$. Haben wir eine Zerlegung $b = cd$ mit teilerfremden c und d, so gibt es nach Lemma B(i) Elemente u, v, so dass $1 = uc + vd$, und somit

$$x = \frac{va}{c} + \frac{ua}{d}.$$

Indem wir den gegebenen Bruch eventuell mit einem invertierbaren Element erweitern, erhalten wir mit Folgerung B eine Darstellung

$$b = p_1^{n_1} \cdots p_k^{n_k}$$

mit positiven Exponenten, wobei die Elemente $b_i = p_i^{n_i}$ nach Folgerung A paarweise teilerfremd sind. Durch mehrmalige Anwendung des obigen Arguments erhalten wir

$$x = \frac{a_1}{b_1} + \ldots + \frac{a_k}{b_k}$$

mit $a_i \in R$. Es genügt also, den Fall zu betrachten, dass $b = p^n$ für ein Primelement p ist.

Nach Definition A(iii) gibt es Elemente q_1, r_1, q_2, r_2, ... in R, so dass

$$\begin{aligned} a &= q_1 p + r_1, & \varphi(r_1) &< \varphi(p), \\ q_1 &= q_2 p + r_2, & \varphi(r_2) &< \varphi(p), \\ &\ldots \\ q_{n-1} &= q_n p + r_n, & \varphi(r_n) &< \varphi(p). \end{aligned}$$

Im Unterschied zum euklidischen Algorithmus dividieren wir hier immer durch die selbe Zahl (mit Rest) und brechen nach n Schritten ab. Dividieren wir beide Seiten der entstandenen Gleichungen im Körper K durch geeignete Potenzen von p, so erhalten wir

$$\begin{aligned} \frac{a}{p^n} &= \frac{r_1}{p^n} + \frac{q_1}{p^{n-1}}, \\ \frac{q_1}{p^{n-1}} &= \frac{r_2}{p^{n-1}} + \frac{q_2}{p^{n-2}}, \\ &\ldots \\ \frac{q_{n-1}}{p} &= \frac{r_n}{p} + q_n. \end{aligned}$$

Nach Umnummerierung der r_i folgt die Behauptung. □

3.5 Gruppen

3.5.1 Definition und Beispiele

Es gibt einfachere algebraische Strukturen als Ringe, die nicht zwei Verknüpfungen besitzen, sondern nur eine. So ist beispielsweise die Menge aller umkehrbaren Abbildungen einer Menge in sich selbst mit einer Verknüpfung versehen, nämlich der Verkettung. Oft sind nur diejenigen Abbildungen von Interesse, die gewisse Eigenschaften unverändert lassen. Manchmal nennt man sie Symmetrien. Die Teilmenge dieser Selbstabbildungen ist dann abgeschlossen unter der Verkettung. Diese Idee führte auf den folgenden allgemeinen Begriff.

Definition 10. *Eine* Gruppe *ist eine Menge G mit einer Verknüpfung „$\cdot$“, die folgende Eigenschaften hat:*

(i) *Für alle Elemente g, h und k gilt*

$$(g \cdot h) \cdot k = g \cdot (h \cdot k).$$

(ii) *Es gibt ein Element e, so dass für alle Elemente g gilt*

$$e \cdot g = g, \qquad g \cdot e = g.$$

(iii) *Für jedes Element g gibt es ein Element h, so dass*

$$g \cdot h = e, \qquad h \cdot g = e.$$

Gilt zusätzlich $g \cdot h = h \cdot g$ für alle Elemente g und h, so nennt man G eine kommutative oder abelsche Gruppe. Ist G endlich, so nennt man die Anzahl der Elemente die Ordnung von G.

Eigentlich müsste man für jede Verknüpfung ein eigenes Zeichen benutzen. Das Zeichen + ist nur bei kommutativen Gruppe üblich. Wie schon für Ringe zeigt man, dass die Elemente, deren Existenz in (ii) und (iii) postuliert wird, eindeutig bestimmt sind. Man nennt sie das Einselement von G und das inverse Element von g.

Beispiel 3.3. Ist R ein Ring, so ist R mit der Verknüpfung + eine abelsche Gruppe. Die Menge der invertierbaren Elemente eines (kommutativen) Ringes R ist eine (kommutative) Gruppe, die man auch mit $R^\times$ bezeichnet, beispielsweise $\mathbb{Z}^\times = \{1, -1\}$. Für einen Körper K ist $K^\times = K \setminus \{0\}$, und für jeden kommutativen nullteilerfreien Ring R gilt $R[x]^\times = R^\times$, denn

das Produkt zweier Polynome kann wegen Lemma 2 nur dann gleich dem Einselement sein, wenn beide den Grad 0 haben. ◁

Beispiel 3.4. Ist X eine Menge, so ist die Menge aller bijektiven Abbildungen von X auf X eine Gruppe mit der Verkettung „$\circ$“ als Verknüpfung. ◁

Beispiel 3.5. Ist R ein Ring, so ist die Menge $\mathrm{Aut}(R)$ aller Automorphismen von R eine Gruppe bezüglich der Verkettung (Übungsaufgabe). ◁

Beispiel 3.6. Es sei X eine Menge mit einer Relation „$\sim$“. Dann ist

$$G = \{g : X \to X \mid g \text{ ist invertierbar, } \forall\, x\, \forall\, y \left(x \sim y \Leftrightarrow g(x) \sim g(y)\right)\}$$

eine Gruppe bezüglich der Verkettung.

Betrachten wir zum Beispiel die Menge $X = \{p, q, r, s, t, u, v, w\}$ und die Relation mit der Wahrheitstafel

$\sim$	p	q	r	s	t	u	v	w
p	f	w	w	f	w	f	f	f
q	w	f	f	w	f	w	f	f
r	w	f	f	w	f	f	w	f
s	f	w	w	f	f	f	f	w
t	w	f	f	f	f	w	w	f
u	f	w	f	f	w	f	f	w
v	f	f	w	f	w	f	f	w
w	f	f	f	w	f	w	w	f

Die zugehörige Gruppe G enthält dann z. B. die Abbildung g mit der Wertetabelle

x	p	q	r	s	t	u	v	w
$g(x)$	q	u	p	t	s	w	r	v

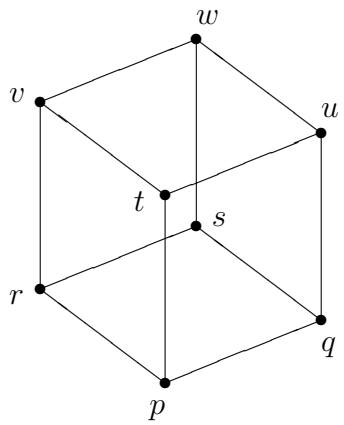
Ist die Relation symmetrisch, so genügt es zu ihrer Beschreibung, die Menge

$$E = \big\{\{x, y\} \subseteq X \mid x \sim y\big\}$$

vorgeben. Ist die Relation außerdem *irreflexiv*, das heißt mit der Eigenschaft $x \not\sim x$ für alle $x \in X$, dann besteht E aus zweielementigen Teilmengen der Menge X. Man nennt das geordnete Paar (X, E) dann einen *Graphen*, die Elemente von X seine Ecken und die Elemente von E seine Kanten. Die Elemente von G nennt man Automorphismen des Graphen. Im obigen Beispiel ist

$$E = \big\{\{p,q\},\{p,r\},\{p,t\},\{q,s\},\{q,u\},\{r,s\},\\ \{r,v\},\{s,w\},\{t,u\},\{t,v\},\{u,w\},\{v,w\}\big\}.$$

Man veranschaulicht einen Graphen, indem man X als Teilmenge der Ebene oder des Raumes realisiert und Elemente, die in der Relation „$\sim$“ stehen, miteinander verbindet. Die obige Relation entspricht zum Beispiel dem unten abgebildeten Graphen. Damit fällt es leichter nachzuprüfen, dass g ein Automorphismus ist. Man kann nichtsymmetrische Relationen ähnlich behandeln, indem man E als Teilmenge von X^2 definiert und ihre Elemente durch Pfeile veranschaulicht. Man nennt (X, E) dann einen gerichteten Graphen. ◁



In Analogie zu Definition 5 führt man folgende Begriffe ein.

Definition 11. *Eine Abbildung φ von einer Gruppe G in eine Gruppe G' wird* Homomorphismus *genannt, wenn $\varphi(e) = e'$ und $\varphi(g \cdot h) = \varphi(g) \cdot \varphi(h)$ für alle Elemente g und h von G gilt. Ein umkehrbarer Homomorphismus heißt* Isomorphismus*, im Fall $G = G'$* Automorphismus *der Gruppe.*

Ist beispielsweise $h : R \to R'$ ein Homomorphismus von Ringen, so bildet er $R^\times$ in $R'^\times$ ab, und seine Einschränkung ist ein Homorphismus von Gruppen.

3.5.2 Permutationen

Dem obigen Beispiel 3.4 widmen wir einen ganzen Abschnitt. Die umkehrbaren Abbildungen einer endlichen Menge X in sich selbst nennt man *Permutationen* von X. Bezüglich der Verkettung bilden sie eine Gruppe, genannt die *symmetrische Gruppe* von X. Im Fall der Menge $\{1, \ldots, n\}$ bezeichnet man diese Gruppe mit S_n, im allgemeinen Fall manchmal mit S_X.

Beispiel 3.7. Durch die Wertetabellen

x	a	b	c	d	e	f	g
$\pi(x)$	c	g	b	d	a	e	f

x	a	b	c	d	e	f	g
$\sigma(x)$	d	f	b	e	g	c	a

werden Permutationen π und σ der Menge $\{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}, \mathrm{f}, \mathrm{g}\}$ definiert, und durch

x	a	b	c	d	e	f	g
$\pi \circ \sigma(x)$	d	e	g	a	f	b	c

ist ihre Verkettung gegeben (die man auch Produkt nennt). ◁

Eine Permutation, die nur zwei Elemente vertauscht und jedes andere Element auf sich selbst abbildet, nennt man *Transposition*. Jede Transposition ist ihr eigenes Inverses.

Homomorphismen zwischen Permutationsgruppen entstehen u. a. auf folgende Weise.

Beispiel 3.8. Ist $\alpha : Y \to X$ eine injektive Abbildung, so erhalten wir einen injektiven Homomorphismus $\varphi : S_Y \to S_X$ durch die Festlegung

$$\varphi(\sigma)(x) = \begin{cases} \alpha \circ \sigma(y), & \text{falls es ein } y \in Y \text{ mit } x = \alpha(y) \text{ gibt,} \\ x & \text{andernfalls.} \end{cases}$$

Ist $\beta : Z \to Y$ eine weitere injektive Abbildung und $\psi : S_Z \to S_Y$ der zugehörige Homomorphismus, so ist auch $\alpha \circ \beta : Z \to X$ injektiv, und der dazu gehörige Homomorphismus ist $\varphi \circ \psi$. Ist z. B. die Abbildung α bijektiv und β ihre Umkehrabbildung, so ist $\varphi \circ \psi$ die identische Abbildung, und durch Vertauschung der Rollen von α und β sehen wir, dass auch $\psi \circ \varphi$ die identische Abbildung ist. Es folgt, dass φ ein Isomorphismus ist. Insbesondere ist also jede symmetrische Gruppe isomorph zur Gruppe S_n für ein geeignetes n. ◁

Lemma 3. *Ist X eine Mengen mit n Elementen, so hat die symmetrische Gruppe S_X die Ordnung $n!$. Ist $n \geq 1$, so ist jede Permutation von X ein Produkt von höchstens $n - 1$ Transpositionen.*

Beweis. Es gibt nur eine Abbildung $\varnothing \to \varnothing$, und die Behauptung für $n = 0$ folgt. Für $n = 1$ ist id die einzige Permutation, und diese kann man als Produkt von 0 Transpositionen auffassen. Außerdem ist $1 = 0!$. Nun sei $|X| = n > 1$, und die Behauptung sei bereits für kleinere n bewiesen. Dann können wir ein Element a von X wählen und $Y = X \setminus \{a\}$ setzen. Ist $\pi \in S_X$ und $b = \pi(a)$, so betrachten wir die Permutation τ von X, welche a auf b und b auf a abbildet sowie alle von a und b verschiedenen Elemente auf sich selbst. Die Abbildung $\sigma = \tau \circ \pi$ bildet Y in sich ab, und ihre Einschränkung auf Y ist ein Element von S_Y. Damit haben wir eine Abbildung

$$S_X \to X \times S_Y$$

definiert. Wir erhalten ihre Umkehrabbildung, indem wir jedem Paar $(b, \sigma) \in X \times S_Y$ die Permutation $\pi = \tau \circ \sigma$ von X zuordnen, wobei wir σ durch die Festlegung $\sigma(a) = a$ zu einer Permutation von X fortsetzen.

Nach Induktionsvoraussetzung ist $|S_Y| = (n-1)!$, und die erste Behauptung folgt. Ebenfalls nach Induktionsvoraussetzung ist σ ein Produkt von höchstens $n-2$ Transpositionen der Menge Y, die sich zu Transpositionen von X fortsetzen. Da τ eine Transposition oder die identische Abbildung ist, folgt die zweite Behauptung. □

Die folgende Definition gehört eigentlich in den vorigen Abschnitt.

Definition 12. (i) *Eine* Untergruppe *einer Gruppe G ist eine Teilmenge H, die bezüglich der Einschränkung der Verknüpfung von G ebenfalls eine Gruppe ist.*

(ii) *Die von einer Teilmenge von G* erzeugte *Untergruppe ist die kleinste Untergruppe, die diese Teilmenge enthält.*

(iii) *Wir sagen, dass eine Gruppe G von einer Teilmenge erzeugt wird, wenn sie gleich der von dieser Teilmenge erzeugten Untergruppe ist.*

Damit eine Teilmenge H einer Gruppe G eine Untergruppe ist, ist es notwendig und hinreichend, dass sie das Einselement von G enthält, für je zwei ihrer Elemente deren Produkt sowie für jedes ihrer Elemente dessen Inverses enthält. Die letzten beiden Bedingungen drückt man mit den Worten aus, dass H bezüglich Verknüpfung und Inversenbildung abgeschlossen sein muss.

In Teil (ii) der Definition ist die kleinste Untergruppe bezüglich der Enthaltenseinsrelation gemeint. Die von einer Teilmenge M erzeugte Untergruppe K ist also charakterisiert durch die Bedingungen $M \subseteq K$ sowie $K \subseteq H$ für jede Untergruppe H mit der Eigenschaft $M \subseteq H$. Eine solche Untergruppe H existiert, denn der Durchschnitt aller Untergruppen H mit der besagten Eigenschaft ist eine Untergruppe, die die genannten Bedingungen erfüllt. Man kann sie auch explizit beschreiben. Die Menge aller (auch mehrfachen) Produkte von Elementen von M bzw. ihrer Inversen ist nämlich in jeder der obigen Untergruppen H enthalten und ist selbst eine Gruppe, also gleich K.

Aus der zweiten Aussage von Lemma 3 folgt, dass die symmetrische Gruppe von den Transpositionen erzeugt wird. Für die Gruppe S_n lässt sich zeigen, dass sie schon von den Transpositionen aufeinanderfolgender Zahlen erzeugt wird (Übungsaufgabe).

3.5.3 Vorzeichen von Permutationen

Wir führen nun einen Gruppenhomomorphismus ein, dessen Existenz nicht offensichtlich ist.

Satz 7. *Für jede endliche Menge X gibt es genau einen Homomorphismus* $\operatorname{sgn} : S_X \to \{1, -1\}$*, der jede Transposition auf* -1 *abbildet.*

Man nennt $\operatorname{sgn}(\pi)$ das *Vorzeichen* von π. Zum Beweis wählen wir eine strikte Ordnung „$\prec$“ auf X. Um sie anzugeben, muss man für jede zweielementige Teilmenge festlegen, welches ihrer Elemente vor dem anderen kommt. Eine Menge mit n Elementen besitzt bekanntlich $\binom{n}{2}$ zweielementige Teilmengen. Der Kürze halber werden wir zweielementige Teilmengen auch (ungeordnete) Paare nennen.

Definition 13. *Wir sagen, dass eine zweielementige Teilmenge $\{a, b\}$ einer Menge X durch eine Permutation π dieser Menge* fehlgestellt *wird, wenn gilt*

$$a \prec b \quad \textit{und} \quad \pi(a) \succ \pi(b)$$

oder

$$a \succ b \quad \textit{und} \quad \pi(a) \prec \pi(b).$$

Paare, die von einer Permutation π fehlgestellt werden, nennen wir Fehlstellungen *von π. Man nennt eine Permutation gerade bzw. ungerade, wenn die Anzahl ihrer Fehlstellungen gerade bzw. ungerade ist.*

Beispiel 3.9. Die zweielementigen Teilmengen der Menge aus Beispiel 3.7 sind

$$\begin{gathered}\{\mathrm{a},\mathrm{b}\}, \underline{\{\mathrm{a},\mathrm{c}\}}, \{\mathrm{a},\mathrm{d}\}, \underline{\{\mathrm{a},\mathrm{e}\}}, \{\mathrm{a},\mathrm{f}\}, \{\mathrm{a},\mathrm{g}\}, \underline{\{\mathrm{b},\mathrm{c}\}},\\ \underline{\{\mathrm{b},\mathrm{d}\}}, \underline{\{\mathrm{b},\mathrm{e}\}}, \underline{\{\mathrm{b},\mathrm{f}\}}, \underline{\{\mathrm{b},\mathrm{g}\}}, \{\mathrm{c},\mathrm{d}\}, \underline{\{\mathrm{c},\mathrm{e}\}}, \{\mathrm{c},\mathrm{f}\},\\ \{\mathrm{c},\mathrm{g}\}, \underline{\{\mathrm{d},\mathrm{e}\}}, \{\mathrm{d},\mathrm{f}\}, \{\mathrm{d},\mathrm{g}\}, \{\mathrm{e},\mathrm{f}\}, \{\mathrm{e},\mathrm{g}\}, \{\mathrm{f},\mathrm{g}\},\end{gathered}$$

wobei wir die Paare unterstrichen haben, die von π bezüglich der alphabetischen Ordnung fehlgestellt werden. Man sieht, dass π bezüglich dieser Ordnung eine ungerade Permutation ist. ◁

Lemma 4. (i) *Sind π und σ Permutationen der geordneten Menge X mit k bzw. l Fehlstellungen und ist m die Anzahl der Fehlstellungen von $\pi \circ \sigma$, so ist $k + l - m$ gerade.*

(ii) *Jede Transposition einer geordneten Menge ist ungerade.*

Beweis. **(i)** Es sei n die Anzahl der Paare $\{a, b\}$, die von σ fehlgestellt werden und für die auch das Paar $\{\sigma(a), \sigma(b)\}$ von π fehlgestellt wird. Diese Paare werden von $\pi \circ \sigma$ nicht fehlgestellt, weil π die von σ vertauschte Reihenfolge wieder richtig stellt. Die Verkettung stellt also nur die Paare fehl, bei denen entweder $\{a, b\}$ von σ oder $\{\sigma(a), \sigma(b)\}$ von π fehlgestellt wird. Ihre Anzahl ist

$$m = (k - n) + (l - n),$$

und durch Umstellung erhalten wir

$$k + l - m = 2n.$$

(ii) Es sei τ die Transposition zweier Elemente a und b, wobei wir annehmen können, dass $a \prec b$ ist. Eine zweielementige Teilmenge, die weder a noch b enthält, wird von τ nicht fehlgestellt. Eine Teilmenge der Form $\{a, x\}$, wobei $x \neq b$ ist, wird fehlgestellt, wenn

$$a \prec x \prec b$$

ist, denn dann gilt $\tau(a) = b \succ x = \tau(x)$, während für $x \prec a$ bzw. $x \succ b$ keine Fehlstellung vorliegt. Das Gleiche gilt für Teilmengen der Form $\{b, x\}$, wobei $x \neq a$ ist. Die Teilmenge $\{a, b\}$ wird natürlich von τ fehlgestellt. Die Anzahl der Fehlstellungen ist folglich

$$2\left|\{x \in X \mid a \prec x \prec b\}\right| + 1,$$

also ungerade. □

Beweis von Satz 7. Wenn π genau k Fehlstellungen erzeugt, so setzen wir

$$\operatorname{sgn}(\pi) = (-1)^k.$$

Dann gilt

$$\operatorname{sgn}(\mathrm{id}) = (-1)^0 = 1$$

und, in den Bezeichnungen von Lemma 4(i), $(-1)^m = (-1)^k(-1)^l$, also

$$\operatorname{sgn}(\pi \circ \sigma) = \operatorname{sgn}(\pi) \cdot \operatorname{sgn}(\sigma).$$

Somit ist sgn ein Homomorphismus, und nach Lemma 4(ii) hat jede Transposition das Vorzeichen -1.

Ist $\operatorname{sgn}' : S_X \to \{1, -1\}$ ein beliebiger Homomorphismus mit den angegebenen Eigenschaften, so gilt für eine Permutation π, die sich als Produkt von j Transpositionen darstellen lässt,

$$\operatorname{sgn}'(\pi) = (-1)^j. \tag{2}$$

Da sich nach Lemma 3 jedes Element auf diese Weise für eine geeignete Anzahl j darstellen lässt, ist sgn' eindeutig bestimmt, also $\operatorname{sgn}' = \operatorname{sgn}$. □

Insbesondere ist der Homomorphismus sgn unabhängig von der Wahl der Ordnung auf X. Man hätte aber die Gleichung (2) nicht ohne weitere Begründung als Definition des Vorzeichens benutzen können, weil j durch π

nicht eindeutig bestimmt ist. So benötigt man in Beispiel 3.7 zur Darstellung von π bzw. σ jeweils 5 Transpositionen. Daraus erhält man eine Darstellung von $\pi \circ \sigma$ als Produkt von 10 Transpositionen, aber tatsächlich genügen 4.

Beispiel 3.10. Wir betrachten das Polynom

$$\Delta(x_1, \ldots, x_n) = \prod_{i<j} (x_j - x_i),$$

wobei das Produkt über alle Paare (i, j) von Zahlen $i, j \in \{1, \ldots, n\}$ mit der Eigenschaft $i < j$ erstreckt wird. Dann gilt

$$\Delta(x_{\pi(1)}, \ldots, x_{\pi(n)}) = \operatorname{sgn}(\pi)\Delta(x_1, \ldots, x_n),$$

weil links bis auf das Vorzeichen die selben Faktoren wie rechts stehen, wobei das Vorzeichen des Faktors mit x_i und x_j genau dann geändert ist, wenn π die Menge $\{i, j\}$ fehlstellt. $\triangleleft$

3.5.4 Zykel

Ist π eine Permutation der endlichen Menge X und $a \in X$, so muss in der Folge a, $\pi(a)$, $\pi^2(a)$ irgendwann ein Element auftauchen, das schon vorher vorkam, d. h. $\pi^k(a) = \pi^{k+l}(a)$. Durch Anwendung von π^{-k} folgt $\pi^l(a) = a$, d. h., die erste Wiederholung ist eine von a selbst. Man nennt die Menge

$$\{\pi^k(a) \mid k \in \mathbb{Z}\} = \{a, \pi(a), \ldots, \pi^{l-1}(a)\}$$

einen *Zykel* von π und l seine Länge. Diese Zykel sind die Äquivalenzklassen bezüglich der Relation

$$a \sim_\pi b \quad \Leftrightarrow \quad \exists\, k \in \mathbb{Z} \quad \pi^k(a) = b.$$

Dies motiviert die sogenannte Zykelschreibweise von Permutationen: Ist a_1, a_2, …, a_l eine Folge verschiedener Elemente von X, so bezeichnet man mit

$$(a_1, a_2, \ldots, a_l)$$

diejenige Permutation von X, welche jeweils a_i auf a_{i+1} abbildet, das Element a_l auf a_1 und jedes nicht aufgeführte Element auf sich selbst. Eine solche Permutation wird auch oft als Zykel bezeichnet. Beim Produkt lässt man in dieser Schreibweise das Verkettungszeichen $\circ$ üblicherweise weg.

Wendet man die Methode aus dem Beweis von Lemma 3 an, so erhält man die Darstellung

$$(a_1, a_2, \ldots, a_l) = (a_1, a_2)(a_2, a_3) \ldots (a_{l-1}, a_l)$$

als Produkt von Transpositionen, so dass ein Zykel der Länge l das Vorzeichen $(-1)^{l-1}$ hat.

Nach dem Gesagten und Satz 1 kann man jede Permutation als Produkt disjunkter Zykeln schreiben. Die Permutationen aus Beipiel 3.7 erscheinen dann als

$$\pi = (\mathrm{a}, \mathrm{c}, \mathrm{b}, \mathrm{g}, \mathrm{f}, \mathrm{e})(\mathrm{d}), \qquad \sigma = (\mathrm{a}, \mathrm{d}, \mathrm{e}, \mathrm{g})(\mathrm{b}, \mathrm{f}, \mathrm{c}).$$

3.5.5 Gruppenoperationen

Das Urbeispiel einer Gruppe ist eine Menge G von bijektiven Abbildungen einer Menge X auf sich selbst, die abgeschlossen unter Verkettung und Inversenbildung ist. In diesem Fall kann man Elemente von G auf Elemente von X anwenden. Für abstrakte Gruppen wird Ähnliches durch den folgenden Begriff beschrieben.

Definition 14. *Eine* Operation *einer Gruppe G auf einer Menge X ist eine Abbildung $G \times X \to X$, geschrieben $(g, x) \mapsto g \cdot x$, so dass für alle g, $h \in G$ und $x \in X$ gilt*

$$e \cdot x = x, \qquad (g \cdot h) \cdot x = g \cdot (h \cdot x).$$

Eine Teilmenge Y von X heißt invariant *unter dieser Operation, wenn für jedes $g \in G$ und $y \in Y$ gilt $g \cdot y \in Y$.*

Eine Operation einer Gruppe G auf einer Menge X heißt transitiv*, wenn es für beliebige x, $y \in X$ wenigstens ein $g \in G$ gibt, so dass $g \cdot x = y$, und die Wirkung heißt* frei*, wenn es für beliebige x, $y \in X$ höchstens ein $g \in G$ gibt, so dass $g \cdot x = y$.*

Genaugenommen müsste man jede Operation mit einem eigenen Symbol bezeichnen, z. B. $\psi : G \times X \to X$. Wir benutzen der Übersichtlichkeit halber das Symbol $\cdot$ oder, wenn die Gruppe G additiv notiert wird, das Symbol $+$. Der Sinn dieses Begriffs besteht darin, dass man aus gegebenen Gruppenoperationen neue erzeugen kann.

Beispiel 3.11. Wie in Beispiel 3.6 sei X eine Menge mit irreflexiver symmetrischer Relation und G die Untergruppe bestehend aus denjenigen Elementen von S_X, die die Relation erhalten. Dann operiert G auch auf der Kantenmenge E des zugehörigen Graphen durch $g \cdot \{x, y\} = \{g(x), g(y)\}$. ◁

Beispiel 3.12. Es sei R ein Teilring des nullteilerfreien kommutativen Ringes S und f ein Polynom in einer Variablen mit Koeffizienten in R. Die Automorphismen des Ringes S, die jedes Element von R auf sich selbst abbilden, bilden eine Gruppe G. Die endliche Menge Y der Nullstellen von f in S ist invariant unter dieser Operation, und durch Einschränkung erhalten wir eine

Operation von G auf Y. In diesem Zusammenhang wurde der Begriff der Gruppe zuerst von Galois eingeführt. $\triangleleft$

Beispiel 3.13. Operiert die Gruppe G auf der Menge X und ist Y eine beliebige Menge, so operiert G auch auf der Menge aller Abbildungen $f : X \to Y$ nach der Festlegung

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

Die Inversenbildung sorgt dafür, dass die zweite Bedingung an eine Gruppenoperation erfüllt ist, denn für g, $h \in G$ und $x \in X$ gilt

$$\begin{aligned}((g \cdot h) \cdot f)(x) &= f((g \cdot h)^{-1} \cdot x) = f((h^{-1} \cdot g^{-1}) \cdot x) \\ &= f(h^{-1} \cdot (g^{-1} \cdot x)) = (h \cdot f)(g^{-1} \cdot x) = (g \cdot (h \cdot f))(x).\end{aligned}$$

Interpretiert man ein n-Tupel $(x_1, \ldots, x_n) \in R^n$ als eine Abbildung $x : \{1, \ldots, n\} \to R$, so erhalten wir eine Operation von S_n auf R^n, nämlich

$$\pi \cdot x = (x_{\pi^{-1}(1)}, \ldots, x_{\pi^{-1}(n)}).$$

Betrachten wir dann die zugehörige Operation der Gruppe S_n auf Funktionen $R^n \to R$, so hebt sich die Inversenbildung wieder auf, d. h.

$$(\pi \cdot f)(x_1, \ldots, x_n) = f(x_{\pi(1)}, \ldots, x_{\pi(n)}).$$

Die selbe Gleichung definiert eine Operation der Gruppe S_n auf der Menge der Polynome in den Variablen x_1, …, x_n, wenn wir sie als Vorschrift zur Substitution auffassen. Man nennt ein Polynom f in n Variablen *symmetrisch*, wenn $\pi \cdot f = f$ für alle $\pi \in S_n$ ist. Dies erklärt den Namen symmetrische Gruppe. Man nennt ein Polynom (wie in Beispiel 3.10) *alternierend*, wenn sich der Wert bei der Vertauschung zweier Argumente mit -1 multipliziert, wohl in Anlehnung an das analoge Verhalten der Determinante, die als Summe mit abwechselnden Vorzeichen geschrieben werden kann. Ein alternierendes Polynom ist lediglich unter der Untergruppe

$$A_n = \{\pi \in S_n \mid \operatorname{sgn}(\pi) = 1\}$$

invariant, die man *alternierende Gruppe* nennt. $\triangleleft$

4 Vektorräume

Vektoren[3] wurden zur Modellierung physikalischer Begriffe wie Geschwindigkeit und Kraft eingeführt, die nicht nur einen Betrag, sondern auch eine Richtung haben. Der Begriff des Vektors für sich genommen ist kein mathematischer Begriff, der eines Vektorraums aber schon.

4.1 Definitionen und Beispiele

Definition 15. *Ein Vektorraum über einem Körper K ist eine Menge V mit einer Verknüpfung $V \times V \to V$, genannt Addition und bezeichnet mit „+“, und einer Verknüpfung $K \times V \to V$, genannt Skalarmultiplikation und bezeichnet mit „·“, mit folgenden Eigenschaften.*

(i) *Die Menge V mit der Addition ist eine abelsche Gruppe.*

(ii) *Für beliebige a, $b \in K$ und v, $w \in V$ gilt*

$$\begin{aligned} 1 \cdot v &= v, & (a+b) \cdot v &= a \cdot v + b \cdot v, \\ (a \cdot b) \cdot v &= a \cdot (b \cdot v), & a \cdot (v+w) &= a \cdot v + a \cdot w. \end{aligned}$$

Wir benutzen die Zeichen + und · auch für die Verknüpfungen in K, ihre Bedeutung ergibt sich jeweils aus dem Zusammenhang. Die Elemente von K nennt man Skalare, die von V nennt man Vektoren und das Nullelement von V den Nullvektor. Wir bezeichnen ihn mit 0_V oder meist einfach mit 0. Aus der zweiten und vierten Eigenschaft in (ii) im Fall $a = b = 0$ bzw. $v = w = 0_V$ folgt $0 \cdot v = 0 \cdot v + 0 \cdot v$ bzw. $a \cdot 0_V = a \cdot 0_V + a \cdot 0_V$, also wegen der Eindeutigkeit des Nullelements

$$0 \cdot v = 0_V, \qquad a \cdot 0_V = 0_V$$

für alle $v \in V$ und $a \in K$. Vektorräume über den Körpern $\mathbb{Q}$, $\mathbb{R}$ bzw. $\mathbb{C}$ bezeichnet man auch als rationale, reelle bzw. komplexe Vektorräume.

Beispiel 4.1. Die Menge K^n aller n-Tupel von Elementen eines Körpers K ist ein Vektorraum bezüglich der Verknüpfungen

$$\begin{aligned} (v_1, \ldots, v_n) + (w_1, \ldots, w_n) &= (v_1 + w_1, \ldots, v_n + w_n), \\ a \cdot (v_1, \ldots, v_n) &= (a \cdot v_1, \ldots, a \cdot v_n). \end{aligned}$$

Im Fall $n = 0$ setzt man $K^0 = \{0\}$. ◁

[3] von dem lateinischen Verb *vehere* (führen, ziehen, tragen)

Beispiel 4.2. Ist V ein Vektorraum über einem Körper K und X eine Menge, so wird die Menge aller Abbildungen $X \to V$, genannt vektorwertige Funktionen auf X, ebenfalls zu einem Vektorraum über K, wenn wir definieren

$$(f+g)(x) = f(x) + g(x), \qquad (a \cdot f)(x) = a \cdot f(x).$$

Im Spezialfall $X = \{1, \ldots, n\}$ und $V = K$ erhalten wir den Vektorraum K^n aus dem vorigen Beispiel. ◁

Beispiel 4.3. Ist R ein Ring und K ein Unterring, der gleichzeitig ein Körper ist, so wird R durch Einschränkung der Multiplikation auf $K \times R$ zu einem Vektorraum über K. So ist z. B. der Polynomring $K[x_1, \ldots, x_n]$ ein Vektorraum über K, der Körper $\mathbb{C}$ ist ein Vektorraum über $\mathbb{R}$ und der Körper $\mathbb{R}$ ein Vektorraum über $\mathbb{Q}$. ◁

Beispiel 4.4. Sind V und W Vektorräume über K, so wird die Produktmenge $V \times W$ mit den durch

$$(v', w') + (v'', w'') = (v' + v'', w' + w''), \qquad a \cdot (v, w) = (a \cdot v, a \cdot w)$$

definierten Verknüpfungen zu einem Vektorraum über K, den man die *direkte Summe* von V und W nennt und meist mit $V \oplus W$ bezeichnet. ◁

Definition 16. *Ein* Untervektorraum *(auch* linearer Unterraum *oder kurz Unterraum genannt) eines Vektorraums V ist eine Teilmenge, die mit der Einschränkung der beiden Verknüpfungen von V wieder ein Vektorraum ist.*

Man kann sich leicht davon überzeugen, dass eine Teilmenge U eines Vektorraums genau dann ein Untervektorraum ist, wenn sie den Nullvektor enthält und abgeschlossen unter der Addition von Vektoren und der Multiplikationen mit beliebigen Skalaren ist. Insbesondere ist U dann eine Untergruppe, wobei die Abgeschlossenheit unter dem Übergang zum entgegengesetzten Vektor nicht extra verlangt werden muss, da sich dieser durch Multiplikation mit dem Entgegengesetzten des Einselements ergibt.

Beispiel 4.5. Im Vektorraum der Funktionen auf einer Menge X mit Werten in einem Vektorraum V ist die Teilmenge der Funktionen, die nur an endlich vielen (bzw. abzählbar vielen) Stellen von Null verschieden sind, ein Untervektorraum. Ist $X = \mathbb{N}$ und $V = \mathbb{R}$ oder $V = \mathbb{C}$, so haben wir den Vektorraum der unendlichen Folgen, und die Teilmengen der beschränkten Folgen, der konvergenten Folgen oder der Nullfolgen sind Untervektorräume. ◁

In der Definition eines Vektorraums ist es eigentlich nicht von Belang, dass die Multiplikation im Skalarbereich eine Umkehroperation hat. Ersetzt man den Körper K durch

einen Ring R, so erhält man die Definition eines Moduls[4] über R. Ist R nicht kommutativ, so spricht man von einem Linksmodul. Betrachtet man statt dessen eine Skalarmultiplikation $V \times R \to V$, bei der das Assoziativgesetz die Form

$$v \cdot (a \cdot b) = (v \cdot a) \cdot b$$

hat, so erhält man den Begriff eines Rechtsmoduls. Bereits bei Schiefkörpern muss man zwischen linken und rechten Vektorräumen unterscheiden.

In Zukunft werden wir auch das Multiplikationszeichen für die Skalarmultiplikation wie üblich weglassen.

4.2 Affine Räume

Der Name „Vektorraum“ legt nahe, dass dieser etwas mit dem physikalischen Begriff des Raumes zu tun hat. In unserem Anschauungsraum ergibt aber die Addition von Punkten keinen Sinn; er wird durch einen anderen mathematischen Begriff modelliert. Dieser ist eine Konkretisierung des Begriffs der Gruppenoperation aus Definition 14.

Definition 17. *Ein* affiner Raum *über dem Körper K ist eine Menge X mit einer Gruppenoperation $X \times V \to X$ eines Vektorraums V über K, die wir mit $+$ bezeichnen, die folgende Eigenschaft hat: Für beliebige Elemente x, $y \in X$ gibt es genau einen Vektor $v \in V$, so dass $x + v = y$.*

Man bezeichnet diesen Vektor v mit $\overrightarrow{xy}$. Im Sinne von Definition 14 ist die Operation von V auf X also transitiv und frei. Die Elemente von X nennt man *Punkte*, und für jeden Vektor $v \in V$ nennt man die durch $x \mapsto x + v$ definierte Abbildung $X \to X$ eine *Verschiebung*. (Dass wir die Gruppenoperation als $X \times V \to X$ und $(x, v) \mapsto x + v$ notieren anstatt als $V \times X \to X$ wie in Definition 14, ist allein der Tradition geschuldet und wegen der Kommutativität der Vektoraddition ohne Belang.)

Halten wir einen Punkt $o \in X$ fest, so ist die durch $v \mapsto o + v$ gegebene Abbildung $V \to X$ nach Definition bijektiv. Ihre Umkehrabbildung ordnet dann einem Punkt x den Vektor $\overrightarrow{ox}$ zu, den man den *Ortsvektor* von x bezüglich des *Ursprungs*[5] o nennt. Angesichts der besagten bijektiven Abbildung kann man auch mit V an Stelle von X arbeiten, weshalb affine Räume oft stiefmütterlich behandelt werden.

Beispiel 4.6. Jeder Vektorraum V über einem Körper K wird zu einem affinen Raum über K, wenn wir V auf sich selbst durch die Vektoraddition operieren lassen. Für beliebige x, $y \in V$ ist dann $\overrightarrow{xy} = y - x$. ◁

[4] wie beim lateinischen Wort *modulus* („Maß(stab)“) Betonung auf der ersten Silbe, Mehrzahl „Moduln“

[5] von dem lateinischen Wort *origo*

Definition 18. *Ist X ein affiner Raum unter der Operation des Vektorraums V, so versteht man unter einem* affinen Unterraum *von X eine Teilmenge Y, für die es einen linearen Unterraum U von V gibt, so dass Y mit der auf $Y \times U$ eingeschränkten Gruppenoperation wieder ein affiner Raum ist.*

In diesem Fall gilt für jeden Punkt $x \in Y$

$$Y = \{x + u \mid u \in U\}.$$

Die Menge auf der rechten Seite wird auch mit $x + U$ bezeichnet. Ist U ein beliebiger linearer Unterraum und $x \in X$ ein beliebiger Punkt, so ist $x + U$ ein affiner Unterraum.

Beispiele von Unterräumen ergeben sich aus der Betrachtung linearer Gleichungssysteme. Die i-te Gleichung in einem solchen System mit n Variablen hat bekanntlich die Form

$$a_{i1}x_1 + \ldots + a_{in}x_n = b_i. \tag{3}$$

Man kann dies auch in der Form $f_i(x_1, \ldots, x_n) = 0$ schreiben, wobei f_i ein Polynom vom Grad 1 ist. Ersetzt man jeweils f_i durch seine homogene Komponente vom Grad 1, ersetzt man also alle b_i durch Null, so erhält man wieder ein lineares Gleichungssystem, welches man das zugehörige *homogene lineare Gleichungssystem* nennt. Das ursprüngliche Gleichungssystem nennt man zur Unterscheidung meist das *inhomogene lineare Gleichungssystem*, obwohl alle Argumente gültig bleiben, falls auch dieses homogen sein sollte.

Lemma 5. *Die Lösungsmenge Y eines linearen Gleichungssystems mit n Variablen und mit Koeffizienten in einem Körper K ist ein affiner Unterraum von K^n. Der zugehörige lineare Unterraum U ist die Lösungsmenge des zugehörigen homogenen linearen Gleichungssystems.*

Beweis. Ist $x = (x_1, \ldots, x_n)$ eine Lösung des Gleichungssystems, so gilt für alle i die Gleichung (3). Ist $v = (v_1, \ldots, v_n)$ eine Lösung des zugehörigen homogenen Gleichungssystems, so gilt für alle i

$$a_{i1}v_1 + \ldots + a_{in}v_n = 0.$$

Durch Addition beider Gleichungen erhalten wir

$$a_{i1}(x_1 + v_1) + \ldots + a_{in}(x_n + v_n) = b_i,$$

also ist das n-Tupel $x + v$ eine Lösung des ursprünglichen inhomogenen Gleichungssystems. Die Lösungsmenge Y ist somit abgeschlossen unter der Einschränkung der Operation von K^n auf die Untergruppe U.

Zunächst hätte man natürlich nachprüfen sollen, dass V eine Untergruppe ist. Die Abgeschlossenheit unter der Addition ist aber der Spezialfall der eben vorgenommenen Rechnung, wenn alle b_i gleich Null sind, und die Abgeschlossenheit unter der Multiplikation mit einem Skalar $a \in K$ folgt analog durch Multiplikation beider Seiten der i-ten homogenen Gleichung mit a.

Sind schließlich x und y Lösungen des inhomogenen Gleichungssystems, so sieht man leicht, dass das n-Tupel $v = y - x$ eine Lösung des homogenen Gleichungssystems ist. Da es also für beliebige x, $y \in Y$ ein $v \in U$ mit der Eigenschaft $x+v = y$ gibt, ist Y unter der Operation von U ein affiner Raum, denn die Eindeutigkeit von v galt ja bereits in K^n. □

Beispiel 4.7. In dem Vektorraum V der Folgen in einem Körper $K \in \{\mathbb{R}, \mathbb{C}\}$ haben wir bereits den linearen Unterraum U der Nullfolgen betrachtet. Die Menge Y der Folgen, die gegen eine vorgegebene Zahl $a \in K$ konvergieren, ist nach den Rechenregeln für Grenzwerte ein affiner Unterraum mit dem zugehörigen linearen Unterraum U. ◁

4.3 Analytische Geometrie

Im griechischen Altertum wurde die Geometrie als Wissenschaft begründet. Dabei ging man von Grundbegriffen wie Punkt, Gerade und Ebene aus, die gewissen Axiomen genügen sollen. In der Neuzeit wurde von Descartes eine andere Beschreibung der Geometrie eingeführt, die die Kluft zwischen der Geometrie und der Analysis beseitigte. Man kann sie dadurch charakterisieren, dass der Raum nun als affiner Raum im Sinne der obigen Definition betrachtet wird.

Wir führen nun weitere geometrisch motivierte Begriffe ein. Es seien X und Y affine Umterräume eines affinen Raums Z, und es seien U, V und W die zugehörigen Vektorräume. Wir sagen, dass X und Y *parallel* sind, wenn $U \subseteq V$ oder $V \subseteq U$ gilt.

Die einfachsten affinen Räume sind die leere Menge und einzelne Punkte. Ein affiner Raum, der nicht von dieser Art ist, aber dessen einzige affine Unterräume Punkte und die leere Menge sind, nennt man *Gerade*. Ist p ein Punkt eines affinen Raumes X und $v \neq 0$ ein Vektor im zugehörigen Vektorraum V, so ist

$$L = \{p + tv \mid t \in K\}$$

eine Gerade in X, und jede Gerade in X entsteht auf diese Weise. Man nennt v einen *Richtungsvektor* von L. Ist K ein angeordneter Körper, so kann man die *Strecke* mit den Endpunkten p und q als

$$\overline{pq} = \{p + tv \mid t \in K,\, 0 \leq t \leq 1\}$$

definieren.

An Stelle von „$p \in L$“ sagt man traditionell, der Punkt p liege auf der Geraden L bzw. L gehe durch p. Für zwei verschiedene Punkte p, q eines affinen Raumes X gibt es genau eine Gerade durch diese beiden Punkte. Man bezeichnet sie traditionell mit pq, und $\overrightarrow{pq}$ ist einer ihrer Richtungsvektoren. Mehrere Punkte heißen *kollinear*, wenn sie auf einer Geraden liegen, und mehrere Geraden heißen *konkurrent*, wenn sie durch einen gemeinsamen Punkt gehen.

Beispiel 4.8. Sind p und q Punkte in einem affinen Raum, so nennt man traditionell einen Punkt m den Mittelpunkt der Strecke $\overline{pq}$, wenn gilt

$$\overrightarrow{pm} = \overrightarrow{mq}.$$

Besser wäre es, einfach vom Mittelpunkt der beiden Punkte p und q zu sprechen, weil dazu keine Anordnung auf K benötigt wird. Wegen

$$\overrightarrow{pm} + \overrightarrow{mq} = \overrightarrow{pq}$$

erhalten wir $2_K \cdot \overrightarrow{pm} = \overrightarrow{pq}$, wobei wir für natürliche Zahlen n definieren

$$n_K = \underbrace{1_K + \ldots + 1_K}_{n}.$$

Der Mittelpunkt zweier verschiedener Punkte existiert also genau dann, wenn $2_K \neq 0_K$ ist, und dann ist

$$m = p + 2_K^{-1} \cdot \overrightarrow{pq}.$$

Nun seien p, q und r nicht kollineare Punkte, und es sei l der Mittelpunkt von r und p, außerdem m der Mittelpunkt von p und q sowie n der Mittelpunkt von q und r. Die Geraden np, lq und mr nennt man traditionell die *Seitenhalbierenden* des Dreiecks pqr.

Wir wollen feststellen, ob und wo sich die Geraden np und mr schneiden. Durch Gleichsetzen der Parameterdarstellungen erhalten wir für den Schnittpunkt

$$n + s \cdot \overrightarrow{np} = m + t \cdot \overrightarrow{mr}.$$

Es liegt nahe, alle Vektoren durch die Kantenvektoren

$$u = \overrightarrow{rp}, \qquad v = \overrightarrow{pq}, \qquad w = \overrightarrow{qr}$$

des Dreiecks auszudrücken. Allerdings genügen wegen $u + v + w = 0_V$ zwei von ihnen, sagen wir v und w. Nun gilt

$$\begin{aligned} p &= q - v, \qquad & m &= q - 2_K^{-1} v, \\ r &= q + w, \qquad & n &= q + 2_K^{-1} w, \end{aligned}$$

also

$$\overrightarrow{np} = -v - 2_K^{-1}w, \qquad \overrightarrow{mr} = w + 2_K^{-1}v,$$

und die obige Gleichung wird zu

$$(q + 2_K^{-1}w) - s(v + 2_K^{-1}w) = (q - 2_K^{-1}v) + t(w + 2_K^{-1}v).$$

Da die Operation des Vektorraums auf dem affinen Raum frei ist, können wir q herauskürzen, und durch Zusammenfassen erhalten wir

$$(s + 2_K^{-1}t - 2_K^{-1})v + (2_K^{-1}s + t - 2_K^{-1})w = 0_V.$$

Wäre einer der Vektoren v und w ein Vielfaches des anderen, dann wären p, q und r kollinear. Da dem nicht so ist, müssen hier die Koeffizienten der beiden Vektoren verschwinden, d. h.

$$\begin{cases} s + 2_K^{-1}t = 2_K^{-1} \\ 2_K^{-1}s + t = 2_K^{-1} \end{cases}$$

Dies ist ein lineares Gleichungssystem. Multiplizieren wir beide Gleichungen mit 2_K, so ergibt sich

$$\begin{cases} 2_K s + t = 1_K \\ s + 2_K t = 1_K \end{cases}$$

Multiplizieren wir jeweils eine Gleichung mit 2_K und subtrahieren die andere, so erhalten wir

$$3_K t = 1_K, \quad 3_K s = 1_K.$$

Zwei Seitenhalbierende schneiden sich also genau dann, wenn $3_K \neq 0_K$ ist, und dann ist der Schnittpunkt von np und mr gleich

$$n + 3_K^{-1} \cdot \overrightarrow{np} = m + 3_K^{-1} \cdot \overrightarrow{mr}.$$

Analog ergibt sich der Schnittpunkt der Geraden np und lq als

$$n + 3_K^{-1} \cdot \overrightarrow{np} = l + 3_K^{-1} \cdot \overrightarrow{lq}.$$

Beide Schnittpunkte stimmen überein, also sind die Seitenhalbierenden konkurrent. $\triangleleft$

Die nächst komplizierteren affinen Räume nach Punkten und Geraden sind Ebenen. Anstatt aber schrittweise vorzugehen, werden wir im Folgenden sofort den allgemeinen Fall betrachten, wobei wir uns anstatt affiner Räume auf Vektorräume konzentrieren.

4.4 Lineare Hülle und lineare Abhängigkeit

In Analogie zu Definition 12 liegt es nahe, den folgenden Begriff einzuführen.

Definition 19. (i) *Den kleinsten Unterraum eines Vektorraums V, der eine gegebene Teilmenge M von V enthält, bezeichnen wir als die* lineare Hülle *von M oder als den von der Menge M erzeugten Unterraum oder als den von M aufgespannten Unterraum.*

(ii) *Wir sagen, dass eine Teilmenge M den Vektorraum V* erzeugt*, wenn V gleich dem von M erzeugten Unterraum ist.*

(iii) *Wir sagen, dass ein Vektorraum* endlich erzeugt *ist, wenn er eine endliche erzeugende Teilmenge besitzt.*

Die lineare Hülle von M existiert, denn sie ist gleich dem Durchschnitt aller Unterräume, die M enthalten (vgl. Aufgabe 29b). Es gibt wenigstens einen solchen Unterraum, nämlich V.

Die Darstellung der allgemeinen Lösung eines homogenen linearen Gleichungssystems legt folgende Begriffe nahe.

Definition 20. *Es sei V ein Vektorraum über einem Körper K und M eine Teilmenge von V.*

(i) *Wir sagen, der Vektor v sei von der Menge M* linear abhängig*, wenn es eine natürliche Zahl k, Elemente $a_1, \ldots, a_k$ von K und Vektoren $v_1, \ldots, v_k \in M$ gibt, so dass*

$$v = a_1 v_1 + \ldots + a_k v_k.$$

(ii) *Ist kein Element der Menge M linear abhängig von den übrigen Elementen von M, so sagen wir, die Menge M sei* (in sich) linear unabhängig.

Die Terme in (i) nennt man *Linearkombinationen* von Elementen von M. Ist M endlich, so kann man jede Linearkombination als Linearkombination sämtlicher Elemente von M schreiben, indem man die übrigen Elemente mit dem Koeffizienten 0 hinzufügt.

Praktisch alle Autoren (und auch wir nach einer Eingewöhnungsphase) lassen die Worte „in sich“ weg und sprechen von linear unabhängigen Vektoren $v_1, \ldots, v_n$. Man beachte, dass die in (ii) definierte lineare Unabhängigkeit nicht das Gegenteil der linearen Abhängigkeit aus (i) ist.

Lemma 6. *Es sei V ein Vektorraum über einem Körper K und M eine Teilmenge von V.*

(i) *Die lineare Hülle von M ist die Menge aller Vektoren, die von M linear abhängig sind.*

(ii) *Eine Menge M ist genau dann (in sich) linear unabhängig, wenn für alle natürlichen Zahlen $k \leq |M|$, alle $a_1, \ldots, a_k \in K$ und paarweise verschiedene $v_1, \ldots, v_k \in M$ mit der Eigenschaft*

$$a_1 v_1 + \ldots + a_k v_k = 0$$

gilt, dass $a_1 = 0, \ldots, a_k = 0$ ist.

Beweis. Ein Unterraum, der die Menge M enthält, enthält auch alle von ihr linear abhängigen Elemente. Die Menge aller dieser Elemente ist also in der linearen Hülle enthalten. Wir prüfen nun, dass diese Menge selbst ein Unterraum ist. Sind nämlich zwei Linearkombinationen $a_1 v_1 + \ldots + a_k v_k$ und $b_1 w_1 + \ldots b_l w_l$ gegeben, so ist auch ihre Summe

$$a_1 v_1 + \ldots + a_k v_k + b_1 w_1 + \ldots b_l w_l$$

eine Linearkombination von Elementen von M. Ebenso gilt für jedes $c \in K$

$$c(a_1 v_1 + \ldots + a_k v_k) = (ca_1)v_1 + \ldots + (ca_k)v_k,$$

und auch der Nullvektor ist eine Linearkombination von Elementen von M. Die Menge aller Linearkombinationen von Elementen von M enthält also die lineare Hülle von M, und somit sind beide gleich.

Wir beweisen nun die Kontrapositionen der beiden Implikationen in Aussage (ii). Angenommen, es gibt eine verschwindende Linearkombination wie in (ii), in der ein Koeffizient a_i nicht Null ist. Wir erhalten dann

$$v_i = -a_i^{-1}(a_1 v_1 + \ldots + a_{i-1} v_{i-1} + a_{i+1} v_{i+1} + \ldots + a_k v_k),$$

wobei die Vektoren auf der rechten Seite wegen der paarweisen Verschiedenheit in $M \backslash \{v_i\}$ liegen. Somit ist v_i linear abhängig von den übrigen Elementen von M, und die Menge M nicht (in sich) linear unabhängig.

Ist umgekehrt M nicht linear unabhängig, so gibt es ein Element v von M, das linear abhängig von den übrigen Elementen von M ist, also

$$v = a_1 v_1 + \ldots + a_k v_k$$

mit $a_1, \ldots, a_k \in K$ und Elementen $v_1, \ldots, v_k \in M \backslash \{v\}$, wobei wir annehmen können, dass die v_i paarweise verschieden sind, weil wir sonst Beiträge von gleichen Vektoren zusammenfassen können. Nun haben wir die verschwindende Linearkombination

$$1v - a_1 v_1 - \ldots - a_k v_k = 0,$$

wobei der Koeffizient von v nicht verschwindet. □

Bei der Arbeit mit den eben eingeführten Begriffen sind folgende Eigenschaften nützlich.

Lemma 7. *Für Teilmengen eines Vektorraums gelten folgende Aussagen.*

(i) *Ist jedes Element einer Menge N linear abhängig von einer Menge M und jedes Element von M linear abhängig von einer Menge L, so ist jedes Element von N linear abhängig von L.*

(ii) *Jede Teilmenge einer (in sich) linear unabhängigen Menge ist (in sich) linear unabhängig.*

(iii) *Ist M (in sich) linear unabhängig und v nicht linear abhängig von M, so ist $M \cup \{v\}$ (in sich) linear unabhängig.*

Aussage (i) zeigt man durch Einsetzen der entsprechenden Linearkombinationen in einander, und Aussage (ii) folgt direkt aus der Definition. Der Beweis von Aussage (iii) ist eine Übungsaufgabe.

Der folgende Satz wird oft irrtümlich Steinitz zugeschrieben.

Satz 8 (Austauschsatz)**.** *Es sei V ein Vektorraum, M eine endliche linear unabhängige Teilmenge und N eine erzeugende Teilmenge. Dann gibt es eine zu M gleichmächtige Teilmenge M' von N, so dass auch $M \cup (N \setminus M')$ den Vektorraum V erzeugt.*

Beweis. Wir verwenden die vollständige Induktion nach der Anzahl der Elemente von M. Für $M = \varnothing$ ist nichts zu beweisen. Nun sei die Behauptung für Mengen mit m Elementen bereits bewiesen und eine Menge M mit $m+1$ Elementen gegeben. Wählen wir $v \in M$, so ist $M \setminus \{v\}$ nach Lemma 7(ii) linear unabhängig. Nach Induktionsvoraussetzung gibt es eine Teilmenge M'' von N mit m Elementen, so dass die Menge $(M \setminus \{v\}) \cup (N \setminus M'')$ den Raum V erzeugt. Nach Lemma 6(i) ist also v von dieser Menge linear abhängig, das heißt, es gibt $a_1, \ldots, a_m, b_1, \ldots, b_n \in K$ und $v_1, \ldots, v_m \in M \setminus \{v\}$ sowie $w_1, \ldots, w_n \in N \setminus M''$, so dass

$$v = a_1 v_1 + \ldots + a_m v_m + b_1 w_1 + \ldots + b_n w_n.$$

Wären alle b_j gleich Null, so wäre v linear abhängig von $M \setminus \{v\}$ im Widerspruch zur linearen Unabhängigkeit von M. Also gibt es ein j, so dass b_j nicht Null ist, und wir erhalten

$$\begin{aligned} w_j = b_j^{-1}(v - a_1 v_1 - \ldots - a_m v_m \\ - b_1 w_1 - \ldots - b_{j-1} w_{j-1} - b_{j+1} w_{j+1} - \ldots - b_n w_n). \end{aligned}$$

Setzen wir $M' = M'' \cup \{w_j\}$, so ist w_j linear abhängig von $M \cup (N \setminus M')$. Das Gleiche gilt für alle übrigen Elemente von $(M \setminus \{v\}) \cup (N \setminus M'')$, da sie in $M \cup (N \setminus M')$ enthalten sind. Nach Lemma 7(i) ist jedes Element von V linear abhängig von $M \cup (N \setminus M')$, was zu beweisen war. □

Da eine Menge mindestens so mächtig wie eine beliebige ihrer Teilmengen ist, erhalten wir:

Folgerung 2. *Ist M eine endliche linear unabhängige Teilmenge und N eine erzeugende Teilmenge eines Vektorraums V, dann ist N mindestens so mächtig wie M.*

Abschließend bemerken wir, dass man den Begriff der linearen Unabhängigkeit nicht nur für Mengen, sondern auch für Familien von Vektoren definieren kann:

Definition 21. *Eine* Familie von Vektoren *in einem Vektorraum V ist eine Abbildung von einer Menge I nach V. Die Elemente von I nennt man Indizes, und das Bild von i schreibt als v_i. Eine Familie $(v_i)_{i \in I}$ von Vektoren heißt* linear unabhängig*, wenn für keinen Index i der Vektor v_i von der Menge $\{v_j \mid j \in I \setminus \{i\}\}$ linear abhängig ist.*

Ähnlich wie in Lemma 6 beweist man: Eine Familie $(v_i)_{i \in I}$ von Vektoren in einem Vektorraum über K ist genau dann linear unabhängig, wenn für jede Familie $(a_i)_{i \in I}$ von Skalaren in K, von denen nur endlich viele von Null verschieden sind, gilt: Wenn

$$\sum_{i \in I} a_i v_i = 0$$

ist, dann sind alle Koeffizienten a_i gleich Null.

In einer Familie kann ein Vektor v mehrmals vorkommen. In diesem Fall ist die Familie offensichtlich nicht linear unabhängig, egal, ob die Menge $\{v_i \mid i \in I\}$ linear unabhängig ist oder nicht.

Der Begriff der Familie verallgemeinert den Begriff der Folge bzw. des n-Tupels, denn

$$(v_1, \ldots, v_n) = (v_i)_{i \in \{1,\ldots,n\}}.$$

4.5 Basen

Definition 22. *Eine Teilmenge M eines Vektorraums V heißt* Basis *von V, wenn sie V erzeugt und linear unabhängig ist.*

Man sieht leicht, dass jede maximale linear unabhängige Teilmenge und jede minimale erzeugende Teilmenge eine Basis ist (Übungsaufgabe). Dabei heißt ein Element a einer halbgeordneten Menge X maximal, wenn es kein Element $x \in X$ mit der Eigenschaft $a \prec x$ gibt (oder im Fall einer nichtstrengen Ordnung, wenn für jedes Element x mit der Eigenschaft $a \preceq x$ gilt $a = x$). Im vorliegenden Fall ist natürlich die Enthaltenseinsrelation gemeint.

Aus Folgerung 2 ergibt sich sofort

Folgerung 3. *Jede Basis eines endlich erzeugten Vektorraums ist endlich.*

Eine Basis M erzeugt den Vektorraum, also kann man jeden Vektor v als Linearkombination der Elemente von M ausdrücken. Sind zwei solche Darstellungen gegeben, so können wir nach Hinzufügung von Termen mit dem Koeffizienten 0 und Zusammenfassen annehmen, dass in beiden die selben paarweise verschiedenen Elemente $v_1, \ldots, v_n$ von M vorkommen, also

$$v = a_1 v_1 + \ldots + a_n v_n, \qquad v = b_1 v_1 + \ldots + b_n v_n.$$

Durch Subtraktion erhalten wir

$$0 = (a_1 - b_1)v_1 + \ldots + (a_n - b_n)v_n.$$

Mit Lemma 6(ii) folgt $a_1 = b_1, \ldots, a_n = b_n$, weil M als Basis linear unabhängig ist. Die Koeffizienten sind also eindeutig bestimmt, und nur endlich viele sind von Null verschieden. Man nennt sie die *Koordinaten* von v bezüglich der Basis M. Genaugenommen sind sie durch eine Abbildung $M \to K$ gegeben.

Beispiel 4.9. Es sei V der Vektorraum der Funktionen auf einer Menge X mit Werten in einem Körper K, die nur an endlich vielen Stellen von Null verschieden sind. Für jedes Element i von X definieren wir eine Funktionen f_i durch die Festlegung $f_i(i) = 1$ und $f_i(x) = 0$ für $x \neq i$. Die Funktionen f_i für alle $i \in X$ bilden eine Basis von V, genannt *Standardbasis*. Die i-te Koordinate einer Funktion f bezüglich dieser Basis ist $f(i)$. Mit Folgerung 3 sehen wir, dass dieser Vektorraum genau dann endlich erzeugt ist, wenn X endlich ist. Im Fall $X = \{1, 2, \ldots, n\}$ können wir V als K^n interpretieren, und die Standardbasis besteht aus den Vektoren $v_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, bei denen die i-te Koordinate gleich 1 ist und alle anderen gleich Null sind. ◁

Beispiel 4.10. Die Menge der Monome mit Koeffizienten 1 (genauer: ihrer Äquivalenzklassen) ist nach Satz 2 eine Basis des Raumes der Polynome. ◁

Satz 9 (Basisergänzungssatz)**.** *Ist N eine erzeugende Teilmenge des endlich erzeugten Vektorraums V und L eine linear unabhängige Teilmenge von N, so gibt es eine Basis M von V, so dass $L \subseteq M \subseteq N$.*

Wenden wir dies mit $N = V$ und $L = \varnothing$ an, so sehen wir, dass jeder endlich erzeugte Vektorraum eine Basis besitzt. Mit Hilfe des Auswahlaxioms lässt sich dieser Satz übrigens ohne die Voraussetzung der endlichen Erzeugtheit beweisen.

Beweis. Nach Voraussetzung existiert eine endliche erzeugende Teilmenge N' von V, und nach der Folgerung aus Satz 8 ist $|L| \leq |N'|$. Wir beweisen die Behauptung durch vollständige Induktion nach $|N'| - |L|$. Im Fall $|L| = |N'|$ folgt aus Satz 8, dass L eine erzeugende Menge ist. Da L bereits linear unabhängig ist, können wir dann $M = L$ setzen.

Nun sei $|N'| - |L| > 0$, und die Behauptung sei im Fall einer kleineren Differenz bereits bewiesen. Wenn jeder Vektor aus N linear abhängig von L ist, so ist L nach Lemma 7(i) eine erzeugende Menge, und wir sind wieder fertig. Andernfalls gibt es ein $w \in N$, das nicht linear abhängig von L ist. Nach Lemma 7(iii) ist dann $L' = L \cup \{w\}$ eine linear unabhängige Teilmenge von N, und wegen $|N'| - |L'| < |N'| - |L|$ gibt es nach Induktionsvoraussetzung eine Basis M, so dass $L' \subseteq M \subseteq N$. □

Wenn wir Folgerung 2 auf zwei Basen anwenden und noch einmal mit vertauschten Rollen anwenden, so ergibt sich:

Folgerung 4. *Alle Basen eines endlich erzeugten Vektorraums sind gleichmächtig.*

Auch dies lässt sich übrigens ohne die Annahme der endlichen Erzeugtheit beweisen. Die Anzahl der Elemente einer Basis nennt man die *Dimension* von V, abgekürzt

$$\dim V,$$

und statt „endlich erzeugter Vektorraum“ sagt man „endlichdimensionaler Vektorraum“.

Folgerung 5. *Ist U ein Unterraum eines endlichdimensionalen Vektorraumes V, so ist $\dim U \leq \dim V$, und im Fall $\dim U = \dim V$ gilt $U = V$.*

Für jede linear unabhängige Teilmenge L von U gilt $|L| \leq \dim V$ nach Folgerung 2. Hat L die maximal mögliche Zahl von Elementen, so ist L nach Aufgabe 36(a) eine Basis von U. Somit ist U endlich erzeugt, und die erste Behauptung folgt. Nach Satz 9 können wir L zu einer Basis M von V ergänzen. Sind L und M gleichmächtig, so folgt wegen der Endlichkeit, dass $L = M$ ist, und die zweite Behauptung folgt. Ohne die Voraussetzung der Endlichdimensionalität braucht die letzte Behauptung aber nicht zu gelten.

Als Basen endlichdimensionaler Vektorräume bevorzugt man Folgen von Vektoren, da die Koordinaten den Vektoren dann durch ihre Nummerierung zugeordnet sind.

4.6 Summen und Durchschnitte

Sind U und W Unterräume eines Vektorraums V über einem Körper K, so ist der Durchschnitt $U \cap W$ wieder ein Unterraum von V, aber die Vereinigung $U \cup W$ ist im Allgemeinen kein Unterraum (Aufgabe 29). Darum liegt es nahe, die lineare Hülle dieser Vereinigung zu betrachten. Man nennt sie die *Summe der beiden Unterräume*, abgekürzt $U+W$. Natürlich ist diese Summe nicht das Gleiche wie die direkte Summe $U \oplus W$ aus Beispiel 4.4. So ist beispielsweise $U + U = U$. Wie in Lemma 6(i) sieht man, dass

$$U + W = \{u + w \mid u \in U,\, w \in W\}.$$

Satz 10. *Es seien U und W endlichdimensionale Unterräume eines Vektorraums. Dann gilt*

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

Beweis. Wir wählen eine Basis $\{v_1, \ldots, v_m\}$ von $U \cap W$ und ergänzen sie mit Hilfe von Satz 9 zu einer Basis $\{v_1, \ldots, v_m, u_1, \ldots, u_l\}$ von U sowie zu einer Basis $\{v_1, \ldots, v_m, w_1, \ldots, w_n\}$ von W. Die Vereinigung beider Basen erzeugt offensichtlich $U + W$. Wir wollen zeigen, dass sie linear unabhängig ist. Es seien $a_1, \ldots, a_l, b_1, \ldots, b_m, c_1, \ldots, c_n$ Konstanten, so dass

$$a_1u_1 + \ldots + a_lu_l + b_1v_1 + \ldots + b_mv_m + c_1w_1 + \ldots + c_nw_n = 0.$$

Der Vektor $c_1w_1 + \ldots + c_nw_n$ ist dann linear abhängig von der oben gewählten Basis von U, und laut Definition liegt er in W, also liegt er in $U \cap W$. Somit gibt es Konstanten $d_1, \ldots, d_m$, so dass

$$c_1w_1 + \ldots + c_nw_n = d_1v_1 + \ldots + d_mv_m.$$

Wegen der linearen Unabhängigkeit der hier vorkommenden Vektoren ist $c_1 = \ldots = c_n = 0$. Die ursprüngliche Linearkombination enthält nun nur noch linear unabhängige Vektoren, und es folgt $a_1 = \ldots = a_l = b_1 = \ldots = b_m = 0$.

Damit haben wir eine Basis von $U + W$ gefunden, und es ergibt sich $\dim(U + W) = l + m + n$. Wegen $\dim U = l + m$, $\dim W = m + n$ und $\dim(U \cap W) = m$ folgt die Behauptung. □

4.7 Affine Hülle

Analog zur linearen Hülle definiert man für eine Teilmenge M eines affinen Raumes X die affine Hülle von M als kleinsten affinen Unterraum von X, der M enthält (vgl. Aufgabe 32). Man nennt sie auch den von M erzeugten affinen Unterraum. Ein Punkt von X heißt affin abhängig von M, wenn er zur affinen Hülle von M gehört. Eine Teilmenge heißt affin unabhängig, wenn keiner ihrer Punkte von den anderen affin abhängig ist.

Schließlich kann man zeigen, dass alle affin unabhängigen und gleichzeitig affin erzeugenden Teilmengen eines affin endlich erzeugten affinen Raumes gleichviele Elemente haben. Man bezeichnet die um 1 verringerte Anzahl als Dimension von X. Der leere affine Raum hat also die Dimension -1. Ist X nicht leer und V der zugehörige Vektorraum, der durch Verschiebungen auf X operiert, so ist $\dim X = \dim V$. Eindimensionale Räume nennt man Geraden, zweidimensionale nennt man Ebenen, und einen Unterraum, dessen Dimension um 1 kleiner ist als die des umgebenden Raums, nennt man Hyperebene. Es gilt ein Analogon von Satz 10, dass nämlich für zwei affine Unterräume eines affinen Raumes die Summe ihrer Dimensionen gleich der Summe der Dimensionen ihres Durchschnitts und der affinen Hülle ihrer Vereinigung ist.

4.8 Lineare Abbildungen

4.8.1 Definitionen

Genau wie für Gruppen und Ringe gibt es auch für Vektorräume und affine Räume den zugehörigen Abbildungsbegriff. Als Motivation mag der Schatten dienen, der von einem Fensterbild durch parallele Lichtstrahlen auf den Boden geworfen wird.

Definition 23. *Es seien V und W Vektorräume über einem Körper K. Eine Abbildung $f : V \to W$ heißt* linear*, wenn für alle u, $v \in V$ und $a \in K$ gilt*

$$f(u + v) = f(u) + f(v), \qquad f(a \cdot v) = a \cdot f(v).$$

Weiter seien X und Y affine Räume und V bzw. W die zugehörigen Vektorräume. Eine Abbildung $h : X \to Y$ heißt affin*, wenn es eine lineare Abbildung $h' : V \to W$ gibt, so dass für alle $x \in X$ und $v \in V$ gilt*

$$h(x + v) = h(x) + h'(v).$$

Lange Zeit wurde zwischen linearen und affinen Abbildungen nicht klar unterschieden. Insbesondere werden affine Funktionen $V \to K$ meist als *lineare Funktionen* bezeichnet.

Man prüft leicht nach, dass die Verkettung linearer Abbildungen linear und dass die Verkettung affiner Abbildungen affin ist. Das Bild eines Untervektorraums unter einer linearen Abbildung ist wieder ein Untervektorraum, und analoges gilt für affine Unterräume und Abbildungen. Man nennt lineare Abbildungen auch *Homomorphismen von Vektorräumen* und solche von einem Vektorraum in sich selbst *Endomorphismen von Vektorräumen*. Eine umkehrbare lineare Abbildung nennt man einen *Isomorphismus von Vektorräumen*, symbolisch angedeutet als $V \xrightarrow{\sim} W$, die entsprechenden Vektorräume heißen dann isomorph, symbolisch $V \cong W$. Die Isomorphie von Vektorräumen ist eine Äquivalenzrelation.

Man bezeichnet die Menge aller linearen Abbildungen von V nach W mit $\mathrm{Hom}(V, W)$. Dies ist ein linearer Unterraum des Vektorraums aller Abbildungen von V nach W, den wir in Beispiel 4.2 eingeführt haben. Statt $\mathrm{Hom}(V, V)$ schreibt man auch $\mathrm{End}(V)$. Für Vektorräume über Schiefkörpern ist $\mathrm{Hom}(V, W)$ übrigens nur eine kommutative Untergruppe, denn beim Nachweis, dass für eine lineare Abbildung f und einen Skalar $a \in K$ auch die Abbildung $a \cdot f$ linear ist, wird die Kommutativität von K benutzt.

Ist $(v_1, \ldots, v_n)$ eine Folge von Vektoren in einem Vektorraum V, so ist die Abbildung, die jedem n-Tupel $(x_1, \ldots, x_n)$ den Vektor $x_1v_1 + \ldots + x_nv_n$ zuordnet, ein Homomorphismus $K^n \to V$. Ist die Folge linear unabhängig, so ist diese Abbildung injektiv, und erzeugt die Folge den Vektorraum V, so ist die Abbildung surjektiv. Ist die Folge eine Basis, so erhalten wir also einen Isomorphismus $K^n \xrightarrow{\sim} V$. Somit sind zwei endlichdimensionale Vektorräume genau dann isomorph, wenn sie die gleiche Dimension haben. Wir werden später sehen, dass dies auch ohne die Voraussetzung der Endlichdimensionalität gilt.

Geometrische Figuren (d. h. Teilmengen von affinen Räumen), die durch einen Isomorphismus zwischen affinen Räumen aufeinander abgebildet werden, nennt man zueinander affin (lat. für „angrenzend“, „verschwägert“). Dies ist der Ursprung des Wortes „affin“ in der Mathematik. Affinität ist eine schwächere Verwandtschaft als Kongruenz und Ähnlichkeit, auf die wir später zu sprechen kommen werden.

4.8.2 Abbildungsmatrizen

Wir halten einmal Basen $(v_1, \ldots, v_n)$ von V und $(w_1, \ldots, w_m)$ von W fest. Ist $f : V \to W$ eine lineare Abbildung, dann gibt es Elemente a_{ij} von K, so dass

$$\begin{cases} f(v_1) = a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m \\ f(v_2) = a_{12}w_1 + a_{22}w_2 + \ldots + a_{m2}w_m \\ \qquad \vdots \\ f(v_n) = a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{mn}w_m \end{cases}$$

Ist nun ein Vektor $x \in V$ durch seine Koordinaten $x_1, \ldots, x_n \in K$ bezüglich der gewählten Basis von V gegeben und hat sein Bild unter f die Zerlegung

$$f(x) = y_1w_1 + \ldots + y_mw_m$$

bezüglich der gewählten Basis von W, so können wir die Koordinaten y_i durch die Koordinaten x_j ausdrücken. Wegen der Linearität von f erhalten

wir nämlich

$$\begin{aligned} f(x) &= x_1 f(v_1) + \ldots + x_n f(v_n) \\ &= x_1(a_{11}w_1 + \ldots + a_{m1}w_m) + \ldots + x_n(a_{1n}w_1 + \ldots + a_{mn}w_m), \end{aligned}$$

und wegen der Eindeutigkeit der Koordinaten folgt

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \quad \vdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{cases} \tag{4}$$

Man fasst die Zahlen a_{ij} zu einer Matrix

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

zusammen und nennt sie die *Abbildungsmatrix* (oder auch Darstellungsmatrix) von f bezüglich der gegebenen Basen von V und W. Dabei betrachtet man Basen als Folgen und nicht Mengen von Vektoren, damit die Reihenfolge der Spalten und Zeilen festgelegt ist. Umgekehrt bestimmt jede $m \times n$-Matrix A nach Formel (4) eine lineare Abbildung f. Die obige Rechnung sollte korrekterweise mit Hilfe des Summenzeichens aufgeschrieben werden. Mit

$$f(x) = \sum_{j=1}^{n} x_j f(v_j) = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} w_i = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j w_i$$

ergibt sich dann

$$y_i = \sum_{j=1}^{n} a_{ij} x_j.$$

Interpretieren wir $m \times n$-Matrizen vermittels dieser Formel als Abbildungen $K^n \to K^m$, so können wir alle relevanten Abbildungen in dem *Diagramm*

$$\begin{array}{ccc} K^n & \xrightarrow{A} & K^m \\ {\scriptstyle (v_1,\ldots,v_n)}\downarrow\wr & & \wr\downarrow{\scriptstyle (w_1,\ldots,w_m)} \\ V & \xrightarrow{f} & W \end{array}$$

zusammenfassen, in dem die vertikalen Pfeile die Isomorphismen aus Abschnitt 4.8.1 darstellen, die durch die Festlegung der Basen entstehen. Für

jeden Weg, der immer der Pfeilrichtung folgt, ergibt sich eine verkettete Abbildung. Die so entstehende Abbildung von der linken oberen Ecke in die rechte untere Ecke hängt nicht von dem gewählten Weg ab. Solche Diagramme nennt man *kommutativ*.

Ordnen wir jeder linearen Abbildung $V \to W$ ihre Abbildungsmatrix zu, so erhalten wir einen Isomorphismus von $\mathrm{Hom}(V, W)$ auf den Vektorraum der $m \times n$-Matrizen, wenn wir die Addition und Skalarmultiplikation von Matrizen komponentenweise definieren. Wir bezeichnen diesen Vektorraum von Matrizen mit $K^{m\times n}$, denn er ist seinerseits isomorph zu K^{mn}. Sein Nullelement ist die Matrix $0_{m,n}$, deren sämtliche Koeffizienten gleich Null sind.

Beim Übergang zum Gleichungssystem (4) haben wir die Reihenfolge der Faktoren vertauscht, weil man traditionell die Koeffizienten vor den Variablen schreibt. Im Fall von Vektorräumen über Schiefkörpern wäre das nicht richtig, vielmehr gilt das oben Gesagte nur für rechte Vektorräume. Im Fall von linken Vektorräumen muss man die Koordinaten von x zu einem Zeilenvektor zusammenfügen und von rechts mit der Abbildungsmatrix multiplizieren.

4.8.3 Verkettung und Matrizenmultiplikation

Nun sei ein weiterer Vektorraum U und eine lineare Abbildung $g : U \to V$ gegeben. Ihre Abbildungsmatrix bezüglich einer Basis $(u_1, \ldots, u_p)$ von U und der obigen Basis von V sei

$$B = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1p} \\ b_{21} & b_{22} & \ldots & b_{2p} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \ldots & b_{np} \end{pmatrix}.$$

Hat nun ein Vektor $t \in U$ die Koordinaten $t_1, \ldots, t_p$, also

$$t = \sum_{k=1}^{p} t_k u_k,$$

so ergeben sich die Koordinaten von $x = g(t)$ auf die von der Abbildung f bekannte Weise als

$$x_j = \sum_{k=1}^{p} b_{jk} t_k.$$

Hat das Ergebnis der Verkettung $y = f(g(t))$ die Koordinaten $y_1, \ldots, y_m$, so ist

$$y_i = \sum_{j=1}^{n} a_{ij} \sum_{k=1}^{p} b_{jk} t_k = \sum_{k=1}^{p} c_{ik} t_k,$$

wobei

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

Die Abbildungsmatrix C der verketteten Abbildung $f \circ g$ nennt man die *Produktmatrix* von A und B, abgekürzt

$$C = AB.$$

Man beachte, dass sie nur definiert ist, wenn die Anzahl der Spalten von A gleich der Anzahl der Zeilen von B ist. Der Eintrag c_{ik} von C ergibt sich aus der i-ten Zeile von A und der k-ten Spalte von B nach der obigen Formel.

Man kann das Gleichungssystem (4) mit Hilfe der Matrizenmultiplikation auch in der Form

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

schreiben. Aus diesem Grunde verwendet man meist *Spaltenvektoren* (also $n \times 1$-Matrizen) anstelle von n-Tupeln, um die Elemente von K^n zu notieren. Wir erhalten nun das kommutative Diagramm von Abbildungen

$$\begin{array}{ccccc} & & C & & \\ & \overset{}{\curvearrowright} & & \overset{}{\curvearrowright} & \\ K^p & \xrightarrow{B} & K^n & \xrightarrow{A} & K^m \\ \downarrow \wr & & \downarrow \wr & & \downarrow \wr \\ U & \xrightarrow{g} & V & \xrightarrow{f} & W \\ & \underset{}{\curvearrowright} & & \underset{}{\curvearrowright} & \\ & & f \circ g & & \end{array} \tag{5}$$

in dem die Abbildungen im oberen Teil durch Linksmultiplikation mit den angegebenen Matrizen entstehen.

Die Verkettung von Abbildungen ist assoziativ. Im Fall linearer Abbildungen hat sie die zusätzlichen Eigenschaften

$$\begin{aligned} (f_1 + f_2) \circ g &= f_1 \circ g + f_2 \circ g, & (a \cdot f) \circ g &= a \cdot (f \circ g), \\ f \circ (g_1 + g_2) &= f \circ g_1 + f \circ g_2, & f \circ (a \cdot g) &= a \cdot (f \circ g) \end{aligned}$$

Insbesondere ist $\mathrm{End}(V)$ ein Ring, genannt *Endomrophismenring* von V. Sein Einselement ist die identische Abbildung. Auch die $n \times n$-Matrizen bilden

einen Ring $K^{n\times n}$. Für $n > 1$ enthält er übrigens Nullteiler und ist nicht kommutativ. Sein Einselement ist die *Einheitsmatrix*

$$1_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

für die auch die Bezeichnungen E_n und I_n üblich sind. Dies ist die Matrix der identischen Abbildung bezüglich einer beliebigen Basis von V. Ihre Koeffizienten bezeichnet man mit dem *Kroneckersymbol*

$$\delta_{ij} = \begin{cases} 1, & \text{wenn } i = j, \\ 0 & \text{andernfalls.} \end{cases}$$

Ordnen wir jedem Endomorphismus seine Abbildungsmatrix zu, wobei wir für V als Definitions- und Zielbereich die selbe Basis festhalten, so erhalten wir einen Isomorphismus von Ringen $\mathrm{End}(V) \to K^{n\times n}$.

Im Allgemeinen brauchen aber bei der Definition der Abbildungsmatrix im Spezialfall $V = W$ die Basen nicht notwendigerweise übereinzustimmen. Die Abbildungsmatrix der identischen Abbildung nennt man in diesem Fall die *Übergangsmatrix* von der ersten Basis zur zweiten. Ist z. B. $(v'_1 \dots, v'_n)$ eine weitere Basis von V und

$$v_j = \sum_{i=1}^{n} r_{ij} v'_i,$$

so erhalten wir das kommutative Diagramm

$$\begin{array}{ccc} K^n & \xrightarrow{R} & K^n \\ {\scriptstyle (v_1,\dots,v_n)}\downarrow\wr & & \wr\downarrow{\scriptstyle (v'_1,\dots,v'_n)} \\ V & \xrightarrow{\mathrm{id}} & V \end{array}$$

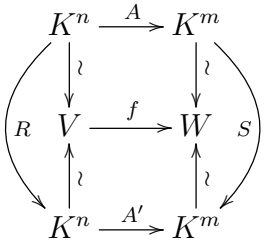
Mit Hilfe von Übergangsmatrizen kann man die Abhängigkeit einer beliebigen Abbildungsmatrix von der Wahl der Basen beschreiben.

Lemma 8. *Ist A' die Abbildungsmatrix von f bezüglich einer neuen Basis $(v'_1, \dots, v'_n)$ von V und einer neuen Basis $(w'_1, \dots, w'_m)$ von W, so ist*

$$A'R = SA,$$

wobei R die Übergangsmatrix von der alten zur neuen Basis von V und S die Übergangsmatrix von der alten zur neuen Basis von W bezeichnet.

Beweis. Das Diagramm



ist kommutativ, weil jede Masche in einem früher betrachteten kommutativen Diagramm vorkommt. Man kann den Weg $A'R$ schrittweise in den Weg SA umformen, indem man jeweils die Kommutativität einer Masche benutzt. □

Natürlich hätte man den Beweis auch in Formeln führen können, wenn man Bezeichnungen für die Isomorphismen $K^n \xrightarrow{\sim} V$ usw. eingeführt hätte.

Folgerung 6. *In der Situation von Lemma 8 sei T die Übergangsmatrix von der neuen zur alten Basis in V. Dann gilt*

$$A' = SAT.$$

Diagramm (5), angewendet auf $f = g = \mathrm{id}$, ergibt nämlich $RT = 1_n$.

Das Gesagte lässt sich auf affine Räume übertragen. Ist X ein n-dimensionaler affiner Raum, so können wir einen Koordinatenursprung $o \in X$ und eine Basis $(v_1, \ldots, v_n)$ des zugehörigen Vektorraums V wählen. Dann ist jeder Punkt

$$x = o + x_1 v_1 + \ldots + x_n v_n$$

eindeutig durch seine Koordinaten $x_1, \ldots, x_n$ bestimmt.

Wählen wir in einem zweiten affinen Raum Y einen Ursprung p und im zugehörigen Vektorraum eine Basis $(w_1, \ldots, w_m)$, so haben wir eine analoge Darstellung

$$y = p + y_1 w_1 + \ldots + y_m w_m$$

für die Punkte von Y. Ist nun $h : X \to Y$ eine affine Abbildung und $h' : V \to W$ die zugehörige lineare Abbildung, so gilt für

$$y = h(x)$$

wegen

$$h(o + v) = h(o) + h'(v)$$

dann

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$$

mit der Abbildungsmatrix von h' und

$$\overrightarrow{ph(o)} = a_1 w_1 + \ldots + a_m w_m.$$

Dies lässt sich zusammenfassen zu

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & a_1 \\ a_{21} & a_{22} & \dots & a_{2n} & a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & a_m \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{pmatrix},$$

wobei wir die $(m+1)\times(n+1)$-Matrix auf der rechten Seite als affine Abbildungsmatrix von h bezüglich der gewählten Ursprünge und Basen bezeichnen. Die afine Abbildungsmatrix einer Verkettung ist dann das Produkt der affinen Abbildungsmatrizen der Faktoren, und wenn wir die Abbildungsmatrix der identischen Abbildung als affine Übergangsmatrix bezeichnen, so gilt ein Analogon von Lemma 8.

4.8.4 Kerne und Bilder

Will man feststellen, ob eine Abbildung surjektiv bzw. injektiv ist, so muss man für jedes Element des Zielbereiches prüfen, ob sein Urbild wenigstens ein bzw. nicht mehr als ein Element hat. Für lineare Abbildungen lässt sich die Injektivität bzw. Surjektivität einfacher nachprüfen. Dabei helfen folgende Begriffe.

Definition 24. *Es sei $f : V \to W$ eine lineare Abbildung. Dann nennt man*

$$\operatorname{Ker} f = \{v \in V \mid f(v) = 0\}, \qquad \operatorname{Im} f = \{f(v) \mid v \in V\}$$

den Kern *bzw. das* Bild *von f. Die Dimension von $\operatorname{Im} f$ nennt man auch den* Rang *der Abbildung f, abgekürzt $\operatorname{rg} f$.*

Der Kern ist also nichts anderes als das Urbild des Nullvektors, und das Bild (dessen Abkürzung für das lateinische Wort „imago“ steht und nicht mit dem Imaginärteil einer komplexen Zahl verwechselt werden sollte) ist der Wertevorrat von f.

Lemma 9. *Es sei $f : V \to W$ eine lineare Abbildung.*

(i) *Die Mengen $\operatorname{Ker} f$ und $\operatorname{Im} f$ sind Untervektorräume von V bzw. W.*

(ii) *Die Abbildung f ist genau dann injektiv, wenn $\operatorname{Ker} f = \{0\}$ ist.*

(iii) *Die Abbildung f ist genau dann surjektiv, wenn $\operatorname{Im} f = W$ ist.*

Beweis. Aussage (i) folgt leicht aus der Linearität von f. Wegen der Linearität von f ist die Aussage $f(u) = f(v)$ äquivalent zu $f(u - v) = 0$, also zu $u - v \in \operatorname{Ker} f$. Nun folgt Aussage (ii). Aussage (iii) ist offensichtlich. □

Für die Surjektivität einer Abbildung $f : V \to W$ genügt es angesichts von Folgerung 5 also, dass $\operatorname{rg} f = \dim W$ ist, vorausgesetzt, W ist endlichdimensional. In diesem Zusammenhang ist manchmal folgendes Ergebnis nützlich.

Satz 11. *Ist $f : V \to W$ eine lineare Abbildung, wobei V endlichdimensional ist, so gilt*

$$\dim V = \dim \operatorname{Ker} f + \dim \operatorname{Im} f.$$

Beweis. Es sei M eine Basis von $\operatorname{Ker} f$. Nach Satz 9 können wir M zu einer Basis N von V ergänzen. Wir wollen zeigen, dass

$$L = \{f(v) \mid v \in N \setminus M\}$$

eine Basis von $\operatorname{Im} f$ ist.

Jedes Element x von V ist eine Linearkombination von N. Wegen der Linearität von f ist $f(x)$ eine Linearkombination der Vektoren $f(v)$ mit $v \in N$. Da man die Nullvektoren weglassen kann, ist $f(x)$ linear abhängig von L.

Ist $a_1 f(v_1) + \ldots + a_l f(v_l) = 0$ mit $a_1, \ldots, a_l \in K$ und paarweise verschiedenen $v_1, \ldots, v_l \in N \setminus M$, so folgt, dass $a_1 v_1 + \ldots + a_l v_l \in \operatorname{Ker} f$, also gibt es $b_1, \ldots b_k \in K$ und $u_1, \ldots, u_k \in M$, so dass

$$a_1 v_1 + \ldots + a_l v_l = b_1 u_1 + \ldots + b_k u_k.$$

Wegen der linearen Unabhängigkeit von N ist $a_1 = \ldots = a_l = 0$, also ist die Folge $(f(v_1), \ldots, f(v_l))$ linear unabhängig. Da die gewählte Teilmenge aus $N \setminus M$ beliebig war, ist L linear unabhängig, und der Fall $l = 2$ zeigt, dass die Einschränkung von f auf $N \setminus M$ injektiv ist. Somit ist $|N| = |M| + |L|$. □

Auch hier kann man zeigen, dass die Endlichdimensionalität von V nicht notwendig ist. Ergänzen wir die Basis L aus dem Beweis des Satzes zu einer Basis von W, so erhalten wir:

Folgerung 7. *Ist $f : V \to W$ eine lineare Abbildung zwischen endlichdimensionalen Vektorräumen, so gibt es Basen von V und W, bezüglich derer f die Abbildungsmatrix*

$$\begin{pmatrix} 1_k & 0_{k,n-k} \\ 0_{m-k,k} & 0_{m-k,n-k} \end{pmatrix}$$

hat.

Angesichts der Bemerkung vor dem Satz erhalten wir:

Folgerung 8. *Haben V und W gleiche endliche Dimension, so ist eine lineare Abbildung $f : V \to W$ genau dann injektiv, wenn sie surjektiv ist.*

Letzteres gilt im unendlichdimensionalen Fall allerdings nicht:
Beispiel 4.11. Es sei V der Vektorraum der unendlichen Folgen in einem Körper K, bei denen nur endlich viele Glieder nicht Null sind. Die durch

$$f(x_1, x_2, x_3, \dots) = (0, x_1, x_2, \dots), \qquad g(x_1, x_2, \dots) = (x_2, x_3, \dots)$$

gegebenen Abbildungen f, $g : V \to V$ sind linear. Die Abbildung f ist injektiv, aber nicht surjektiv, und die Abbildung g ist surjektiv, aber nicht injektiv. Übrigens gilt $g \circ f = \mathrm{id}$, aber $f \circ g \neq \mathrm{id}$. Man nennt f eine rechtsinverse Abbildung von g sowie g eine linksinverse Abbildung von f. $\triangleleft$

4.8.5 Direkte Summen und Projektionen

Die Definition der direkten Summe zweier Vektorräume verallgemeinert sich auf Familien von Vektorräumen: Sind für alle Elemente i einer Indexmenge I Vektorräume V_i über einem Körper K gegeben, so bezeichnet man mit $\bigoplus_{i \in I} V_i$ die Menge aller Familien x von Vektoren $x_i \in V_i$, von denen nur endlich viele nicht Null sind. Im Fall $I = \{1, \dots, r\}$ schreibt man die direkte Summe auch als $\bigoplus_{i=1}^{r} V_i$ und ihre Elemente als $x = (x_1, \dots, x_r)$. Sie ist isomorph zu $V_1 \oplus \dots \oplus V_r$ mit beliebiger Klammersetzung. Ist ein weiteres Element y von $\bigoplus_{i \in I} V_i$ gegeben, so setzen wir $(x + y)_i = x_i + y_i$, und für $a \in K$ setzen wir $(a \cdot x)_i = a \cdot x_i$. Damit wird die direkte Summe zu einem Vektorraum.

In dem Fall, dass all die Vektorräume V_i Unterräume eines Vektorraums V sind, haben wir eine lineare Abbildung

$$\bigoplus_{i \in I} V_i \to V,$$

die jeder Familie x von Vektoren die Summe $\sum_{i \in I} x_i$ ihrer Mitglieder zuordnet. Im Fall $I = \{1, \dots, r\}$ ist diese Abbildung also gegeben durch

$$(x_1, \dots, x_r) \mapsto x_1 + \dots + x_r.$$

Wenn diese Abbildung ein Isomorphismus ist, so sagt man, der Vektorraum V sei die *direkte Summe* seiner Unterräume V_i.

Dies ist offensichtlich genau dann der Fall, wenn sich jeder Vektor aus V auf eindeutige Weise als Summe von Vektoren aus den Unterräumen V_i darstellen lässt, wobei natürlich nur endlich viele Summanden nicht Null sein dürfen. Im Fall zweier Unterräume gibt es eine weitere Umformulierung:

Der Vektorraum V ist genau dann die direkte Summe seiner Unterräume U und W, wenn $U + W = V$ und $U \cap W = \{0\}$ ist. Manche Autoren schreiben dann $U \dot{+} W = V$, was allerdings eine Verknüpfung mit einer Aussage vermischt.

Im Fall von mehr als zwei Summanden genügt es für das Vorliegen einer direkten Summe nicht, dass V die Summe der Unterräume V_i ist und die V_i paarweise den Durchschnitt $\{0\}$ haben. Statt dessen muss man verlangen, dass jeder der Unterräume V_i den Durchschnitt $\{0\}$ mit der Summe $\sum_{j \neq i} V_j$ der übrigen Unterräume hat.

Um Verwechslungen zu vermeiden, spricht man hier manchmal von einer *inneren direkten Summe* und nennt den Begriff aus Definition 4.4 die *äußere direkte Summe*.

Ist ein Vektorraum V die direkte Summe seiner Unterräume U und W, so lässt sich, wie gesagt, jedes Element v von V auf eindeutige Weise in der Form $v = u + w$ mit $u \in U$ und $w \in W$ schreiben. Die Abbildung $V \to W$, die einem solchen Vektor v die Komponente w zuordnet, nennt man die *Projektion* von V auf W längs U. Sie ist eine lineare Abbildung.

Ist in dieser Situation X ein affiner Raum mit dem zugehörigen Vektorraum V sowie Y ein affiner Unterraum mit den zugehörigen Untervektorraum W, so lässt sich jedes Element x von X auf eindeutige Weise in der Form $x = y + u$ mit $y \in Y$ und $u \in U$ schreiben. Die Abbildung $X \to Y$, die einem solchen Punkt x den Punkt y zuordnet, nennt man die Projektion von X auf Y längs U. Sie ist eine affine Abbildung. Der am Anfang von Abschnitt 4.8.1 erwähnte Schattenwurf ist ein Beispiel, in dem die Richtung der Lichtstrahlen durch einen eindimensionalen Unterraum U vorgegeben wird.

4.9 Der Dualraum

In der Physik werden Geschwindigkeiten durch Vektoren modelliert, Impulse hingegen durch Kovektoren. So nennt man die Elemente des Dualraums eines Vektorraums, den wir jetzt einführen wollen. Wir erinnern daran, dass für Vektorräume V, W über einem Körper die Menge $\mathrm{Hom}(V, W)$ ein Untervektorraum des Vektorraums aller Abbildungen $V \to W$ aus Beispiel 4.2 ist.

Definition 25. (i) *Ist V ein Vektorraum über dem Körper K, so nennt man* $\mathrm{Hom}(V, K)$ *den* Dualraum *von V und bezeichnet ihn auch mit V^*.*

(ii) *Ist M eine Teilmenge von V, so definieren wir ihren* Annullator *als*

$$M^0 = \{l \in V^* \mid \forall\, v \in M\ (l(v) = 0)\}.$$

Obwohl für linke Vektorräume V, W über einem Schiefkörper K die Untergruppe $\mathrm{Hom}(V, W)$ des Raumes aller Abbildungen $V \to W$ im Allgemeinen kein Untervektorraum

ist, kann man im Fall $W = K$ definieren $(l \cdot a)(v) = l(a) \cdot a$, wodurch V^* zu einem rechten Vektorraum wird.

Ist $l \in V^*$ und $(v_1, \ldots, v_n)$ eine Basis von V, so gilt für einen beliebigen Vektor $x \in V$ mit den Koordinaten $x_1, \ldots, x_n$

$$l(x) = a_1 x_1 + \ldots + a_n x_n,$$

falls wir $a_i = l(v_i)$ setzen. Somit ist l durch ein homogenes Polynom der Ordnung 1 in den Koordinaten gegeben. Homogene Polynome nannte man früher Formen. Daher nennt man die Elemente l von V^* auch *Linearformen* auf V. Die Abbildungsmatrix von l bezüglich der gegebenen Basis von V und der Basis von K, die nur aus dem Einselement besteht, ist der *Zeilenvektor* $\begin{pmatrix} a_1 & \ldots & a_n \end{pmatrix}$, den man meist mit dem n-Tupel $(a_1, \ldots, a_n)$ identifiziert.

Die Bezeichnung U^0 ist eigentlich unvollständig, da der Annullator auch vom umgebenden Raum V abhängt.

Beispiel 4.12. Es sei $K \in \{\mathbb{R}, \mathbb{C}\}$ und V der K-Vektorraum der konvergenten Folgen $x = (x_1, x_2, \ldots)$, wobei für alle i gilt $x_i \in K$. Nach den Rechenregeln für Grenzwerte wird durch

$$l(x) = \lim_{i \to \infty} x_i$$

eine Linearform l auf V definiert. Der Kern von l besteht aus den Nullfolgen. ◁

Der Dualraum V^{**} des Dualraums besteht aus Linearformen auf V^*. Jedem Element v von V können wir eine solche Linearform zuordnen: Sie bildet ein beliebiges Element l von V^* auf $l(v)$ ab. Auf diese Weise erhalten wir eine lineare Abbildung $i_V : V \to V^{**}$. Man nennt sie eine natürliche Abbildung, da sie nicht von der Wahl einer Basis abhängt.

Satz 12. *Es sei V ein endlichdimensionaler Vektorraum.*

(i) *Der Annullator M^0 einer Teilmenge M von V ist ein Unterraum von V^*. Ist U ein Unterraum von V, so gilt*

$$\dim U + \dim U^0 = \dim V = \dim V^*.$$

(ii) *Der natürliche Homomorphismus i_V ist ein Isomorphismus, der einen beliebigen Unterraum U auf den Annullator U^{00} seines Annullators abbildet.*

(iii) *Sind V und W Vektorräume, so sind die Vektorräume $(V \oplus W)^*$ und $V^* \oplus W^*$ auf natürliche Weise isomorph.*

(iv) *Sind T und U Unterräume von V, so gilt*

$$(T+U)^0 = T^0 \cap U^0, \qquad (T \cap U)^0 = T^0 + U^0.$$

Beweis. Sind l und m Linearformen auf V und $a \in K$, so folgt für jedes $v \in M$ aus $l(v) = 0$ und $m(v) = 0$, dass $(l+m)(v) = 0$ und $(a \cdot l)(v) = 0$ ist. Somit ist M^0 ein Untervektorraum.

Eine Linearform ist wie jede lineare Abbildung durch ihre Werte auf den Vektoren einer Basis vollständig bestimmt. Ist eine Basis von V wie oben gegeben, so definieren wir Linearformen $l_i \in V^*$ durch

$$l_i(v_j) = \delta_{ij}.$$

Ist nun l eine beliebige Linearform und ist $a_i = l(v_i)$, so ist

$$l = a_1 l_1 + \ldots + a_n l_n,$$

denn beide Seiten haben die selben Werte auf allen Basisvektoren. Die Linearformen l_i erzeugen also den Dualraum. Haben wir eine verschwindende Linearkombination $a_1 l_1 + \ldots + a_n l_n = 0$, so erhalten wir durch Anwendung der rechten Seite auf den Basisvektor v_i, dass $a_i = 0$ ist. Somit sind die Linearformen l_i linear unabhängig. Man nennt $(l_1, \ldots, l_n)$ die zu $(v_1, \ldots, v_n)$ *duale Basis*, und es folgt $\dim V = \dim V^*$.

Ist ein Unterraum U gegeben, so können wir eine Basis $(v_1, \ldots, v_k)$ von U zu einer Basis $(v_1, \ldots, v_n)$ von V ergänzen. Ist $l \in U^0$, so gilt $l(v_1) = 0$, $\ldots$, $l(v_k) = 0$, also ist l eine Linearkombination von l_{k+1}, $\ldots$, l_n. Diese Elemente der Dualbasis erzeugen also U^0. Da sie linear unabhängig sind, bilden sie eine Basis von U^0, und die linke Gleichung in Aussage (i) folgt.

Der natürliche Homomorphismus i_V ist injektiv, denn jeder von Null verschiedene Vektor v_1 lässt sich zu einer Basis eränzen, und die zugehörige Linearform l_1 verschwindet nicht auf v_1. Durch zweimalige Anwendung von (i) erhalten wir $\dim V = \dim V^* = \dim V^{**}$, und aus Folgerung 8 ergibt sich die erste Aussage in (ii). Für $u \in U$ und $l \in U^0$ ist $i_V(u)(l) = l(u) = 0$. Der Isomorphismus i_V bildet also den Unterraum U von V in den Unterraum U^{00} von V^{**} ab, und seine Einschränkung auf U ist injektiv. Wegen (i) gilt

$$\dim U^{00} = \dim V^* - \dim U^0 = \dim V^* - (\dim V - \dim U) = \dim U.$$

Mit Folgerung 8 ergibt sich die Richtigkeit des Wortes „auf“ in Aussage (ii) .

Wir haben lineare Abbildungen der Vektorräume V und W in die direkte Summe $V \oplus W$, die gegeben sind durch $v \mapsto (v, 0)$ und $w \mapsto (0, w)$. Ihre Verkettungen mit einer Linearform auf $V \oplus W$ liefern ein Paar von Linearformen

auf V bzw. W, also ein Element von $V^* \oplus W^*$. Umgekehrt können wir für Linearformen $l \in V^*$ und $k \in W^*$ durch $(v, w) \mapsto l(v) + k(w)$ eine Linearform auf $V \oplus W$ definieren. Beide Abbildungen sind linear und zueinander invers, was Aussage (iii) beweist.

Für beliebige Teilmengen L und M von V folgt aus der Definition des Durchschnitts, dass

$$(L \cup M)^0 = L^0 \cap M^0.$$

Eine Linearform verschwindet offenbar genau dann auf der linearen Hülle einer Teilmenge von V, wenn sie auf dieser Menge selbst verschwindet. In Anwendung auf $T \cup U$ liefert das die erste Behauptung in (iv).

Für beliebige Teilmengen M und N von V gilt

$$M \subseteq N \quad \Rightarrow \quad M^0 \supseteq N^0,$$

also für Unterräume T und U

$$(T \cap U)^0 \supseteq T^0 \cup U^0.$$

Da der Annullator von $T \cap U$ nach Aussage (i) ein Unterraum ist, folgt

$$(T \cap U)^0 \supseteq T^0 + U^0.$$

Mit der Gleichung aus Behauptung (i), Satz 10 und dem schon bewiesenen Teil von Behauptung (iv) erhalten wir

$$\begin{aligned} \dim(T \cap U)^0 &= \dim V - \dim(T \cap U) \\ &= \dim V - \dim T - \dim U + \dim(T + U) \\ &= \dim T^0 + \dim U^0 - \dim(T + U)^0 \\ &= \dim T^0 + \dim U^0 - \dim(T^0 \cap U^0). = \dim(T^0 + U^0) \end{aligned}$$

Nun folgt die zweite Gleichung in Aussage (iv) mit Folgerung 5. □

Der Satz gilt nicht für unendlichdimensionale Vektorräume. So hat z. B. der Vektorraum V der Folgen in K mit nur endlich vielen von Null verschiedenen Gliedern die Standardbasis, die aus den Folgen

$$e_i = (0, \dots, 0, 1, 0, \dots)$$

besteht, in denen das i-te Glied gleich 1 und alle anderen gleich 0 sind. Eine Linearform l ist durch die Folge der Werte $a_i = l(e_i)$ festgelegt. Die Linearformen l_i bilden keine Basis von V^*, denn sie erzeugen den echten Unterraum der Linearformen, bei denen nur endlich viele dieser Werte nicht Null sind. Man kann zeigen, dass V^* keine abzählbare Basis besitzt.

4.10 Duale Abbildungen

Die im vorigen Abschnitt eingeführte Dualität bezieht sich nicht nur auf Vektorräume, sondern auch auf lineare Abbildungen.

Definition 26.

Ist $f : U \to V$ eine lineare Abbildung zwischen Vektorräumen, so definieren wir die zu f duale Abbildung *$f^* : V^* \to U^*$ durch*

$$f^*(l) = l \circ f \qquad \text{für } l \in V^*.$$

Beispiel 4.13. Es sei W der Vektorraum der Folgen $x = (x_1, x_2, \dots)$ im Körper $K \in \{\mathbb{R}, \mathbb{C}\}$, und für jede solche Folge sei $s(x) = (y_1, y_2, \dots)$ die Folge der Partialsummen

$$y_n = \sum_{k=1}^{n} x_k.$$

Dies definiert eine lineare Abbildung $s : W \to W$. Auf dem Unterraum V der konvergenten Folgen haben wir die Linearform l aus Beispiel 4.12, die jeder Folge ihren Grenzwert zuordnet. Das Urbild

$$U = \{x \in W \mid s(x) \in V\}$$

von V unter s ist ein linearer Unterraum von W, seine Elemente sollte man konvergente Reihen nennen. Durch Einschränkung von s erhalten wir eine lineare Abbildung $f : U \to V$. Die Linearform

$$f^*(l) \in U^*$$

ordnet jeder konvergenten Reihe ihre Reihensumme zu. ◁

Duale Abbildungen haben folgende allgemeine Eigenschaften.

Satz 13. (i) *Die duale Abbildung f^* einer linearen Abbildung $f : V \to W$ ist linear, und es gilt*

$$\operatorname{Ker} f^* = (\operatorname{Im} f)^0, \qquad f^{**} \circ i_V = i_W \circ f.$$

(ii) *Sind außerdem $h : V \to W$ und $g : U \to V$ weitere lineare Abbildungen und ist $a \in K$, so gilt*

$$(f + h)^* = f^* + h^*, \qquad (af)^* = af^*, \qquad (f \circ g)^* = g^* \circ f^*.$$

(iii) *Sind V und W endlichdimensional, so gilt*

$$\operatorname{rg} f = \operatorname{rg} f^*, \qquad \operatorname{Im} f^* = (\operatorname{Ker} f)^0.$$

Beweis. Die Linearität von f^* folgt aus den Rechenregeln für lineare Abbildungen. Für alle $v \in V$ und $k \in W^*$ gilt nach Definition von f^*

$$f^*(k)(v) = k(f(v)). \tag{6}$$

Die Aussage $k \in \operatorname{Ker} f^*$ bedeutet, dass die linke Seite für alle v verschwindet, während die Aussage $k \in (\operatorname{Im} f)^0$ bedeutet, dass die rechte Seite für alle v verschwindet. Damit ist die erste Formel bewiesen.

Die zweite Formel bedeutet, dass für alle $v \in V$ gilt

$$f^{**}(i_V(v)) = i_W(f(v)),$$

und nach Definition der dualen Abbildung ist die linke Seite gleich

$$i_V(v) \circ f^*.$$

Es ist also zu zeigen, dass für alle $v \in V$ und $k \in W^*$ gilt

$$i_V(v)(f^*(k)) = i_W(f(v))(k),$$

was nach Definition von i_V und i_W äquivalent zu Gleichung (6) ist.

Die Formeln in Teil (ii) folgen aus den Rechenregeln für lineare Abbildungen.

Aus Satz 11, dem in (i) Bewiesenen und Satz 12(i) folgt

$$\dim \operatorname{Im} f^* = \dim W^* - \dim \operatorname{Ker} f^* = \dim W - \dim(\operatorname{Im} f)^0 = \dim \operatorname{Im} f$$

wie in (iii) behauptet. Mit Satz 11 und Satz 12(i) können wir weiter umformen:

$$\dim \operatorname{Im} f = \dim V - \dim \operatorname{Ker} f = \dim(\operatorname{Ker} f)^0.$$

Aus Gleichung (6) folgt die Inklusion $\operatorname{Im} f^* \subseteq (\operatorname{Ker} f)^0$, und nach Folgerung 5 gilt sogar Gleichheit. □

Später werden wir die letzte Behauptung auch ohne die Endlichdimensionalität von V und W zeigen. Oft identifiziert man jeden endlichdimensionalen Vektorraum V vermittels des Isomorphismus i_V mit seinem zweiten Dualraum V^{**} und schreibt $f^{**} = f$.

Lemma 10. *Hat die lineare Abbildung $f : V \to W$ zwischen endlichdimensionalen Vektorräumen bezüglich gewisser Basen die Abbildungsmatrix A mit den Koeffizienten a_{ij}, so hat die duale Abbildung f^* bezüglich der dualen Basen die Abbildungsmatrix A^* mit den Koeffizienten $a^*_{ij} = a_{ji}$.*

Man nennt A^* die transponierte Matrix von A. Traditionelle Schreibweisen sind A^t oder, zur Vermeidung von Verwechslungen mit der t-ten Potenz, auch ${}^tA$ oder $A^\top$.

Beweis. Es seien $(v_1, \ldots, v_n)$ und $(w_1, \ldots, w_m)$ Basen von V bzw. W sowie $(l_1, \ldots, l_n)$ und $(h_1, \ldots, h_m)$ die entsprechenden dualen Basen von V^* bzw. W^*. Dann gilt nach Definition der Abbildungsmatrix

$$f(v_i) = \sum_{k=1}^{m} a_{ki} w_k, \qquad f^*(h_j) = \sum_{k=1}^{n} a^*_{kj} l_k.$$

Wir erhalten einerseits

$$f^*(h_j)(v_i) = h_j(f(v_i)) = \sum_{k=1}^{m} a_{ki} h_j(w_k) = \sum_{k=1}^{m} a_{ki} \delta_{jk} = a_{ji}$$

und andererseits

$$f^*(h_j)(v_i) = \sum_{k=1}^{n} a^*_{kj} l_k(v_i) = \sum_{k=1}^{n} a^*_{kj} \delta_{ki} = a^*_{ij}. \qquad \square$$

Folgerung 9. *Für Matrizen A, A_1, $A_2 \in K^{m\times n}$, $B \in K^{n\times p}$ und Skalare $a \in K$ gilt*

$$(A_1 + A_2)^* = A_1^* + A_2^*, \qquad (aA)^* = aA^*, \qquad (AB)^* = B^* A^*.$$

Dies folgt aus Satz 13 zunächst für Abbildungsmatrizen. Eine beliebige $(m \times n)$-Matrix interpretiert man als Abbildungsmatrix der linearen Abbildung $K^n \to K^m$, die durch Linksmultiplikation der Spaltenvektoren mit dieser Matrix gegeben ist, bezüglich der Standardbasen. Man kann die Formeln aber auch direkt nachprüfen.

4.11 Der Quotientenraum

Neben direkten Summen und Dualräumen gibt es eine weitere wichtige Konstruktion von Vektorräumen.

Lemma 11. *Es sei V ein Vektorraum über K und U ein Unterraum.*

(i) *Die durch*

$$v \sim_U w \quad \Longleftrightarrow \quad v - w \in U$$

definierte Relation „$\sim_U$“ auf V ist eine Äquivalenzrelation.

(ii) *Für alle v, v', w, $w' \in V$ und $a \in K$ gilt*

$$\begin{aligned} v \sim_U w, \quad v' \sim_U w' \quad &\Longrightarrow \quad v + v' \sim_U w + w', \\ v \sim_U w \quad &\Longrightarrow \quad a \cdot v \sim_U a \cdot w. \end{aligned}$$

(iii) *Auf der mit V/U bezeichneten Menge der Äquivalenzklassen bezüglich „$\sim_U$“ gibt es genau eine Struktur eines Vektorraums über K derart, dass die so genannte kanonische Abbildung $V \to V/U$, die jedem Vektor seine Äquivalenzklasse zuordnet, linear ist.*

Die Äquivalenzklasse eines Elements v ist

$$v + U = \{v + u \mid u \in U\}.$$

Dies ist natürlich ein affiner Unterraum, aber das ist hier ohne Belang, denn wir betrachten Äquivalenzklassen als Elemente von V/U ohne Beachtung ihrer inneren Struktur.

Beweis. (i) Wegen $v - v = 0 \in U$ ist die Relation reflexiv. Ist $v - w \in U$, so ist $w - v = -(v - w) \in U$, also ist die Relation symmetrisch. Ist $v - w \in U$ und $w - x \in U$, so folgt $v - x = (v - w) + (w - x) \in U$, also ist die Relation transitiv.
(ii) Ist $v - w \in U$ und $v' - w' \in U$, so ist $(v+v') - (w+w') = (v-w) + (v'-w') \in U$. Ist $v - w \in U$ und $a \in K$, so ist $a \cdot v - a \cdot w = a \cdot (v - w) \in U$.
(iii) Wenn besagte Abbildung linear sein soll, so müssen wir die Addition und die Skalarmultiplikation auf V/U durch

$$(v + U) + (w + U) = v + w + U, \qquad a \cdot (v + U) = a \cdot v + U$$

definieren. Nach dem in (ii) Bewiesenen ist die rechte Seite jeweils unabhängig von der Wahl der Repräsentanten, also ist diese Definition korrekt. Die Rechengesetze für die so definierten Verknüpfungen folgen sofort aus denen für die entsprechenden Verknüpfungen von V, wobei die Äquivalenzklassen des Null- und Einselements von V die Rolle des Null- bzw. Einselements in V/U spielen. Die Äquivalenzklasse von $-v$ ist das entgegengesetzte Element der Äquivalenzklasse von v, also sind alle Axiome eines Vektorraums erfüllt. □

Die kanonische Abbildung $V \to V/U$ ist offensichtlich surjektiv und hat den Kern U. Mit Satz 11 erhalten wir also:

Folgerung 10. *Ist V endlichdimensional, so gilt*

$$\dim(V/U) = \dim V - \dim U.$$

Ist V endlichdimensional, so gibt es zu jedem Unterraum U einen *komplementären Unterraum*, d. h. einen Unterraum T, so dass V die direkte Summe von T und U ist (Spezialfall von Aufgabe 37). Verketten wir die Einbettungsabbildung $T \to V$ mit der kanonischen Abbildung $V \to V/U$, so erhalten wir eine lineare Abbildung

$$T \to V/U, \qquad t \mapsto t + U.$$

Diese Abbildung ist ein Isomorphismus. Jeder Vektor $v \in V$ lässt sich nämlich in der Form $v = t + u$ schreiben, und $v + U$ ist dann das Bild von t, woraus die Surjektivität folgt. Ist $t + U = 0 + U$, so ist $t \in T \cap U = \{0\}$, und die Injektivität folgt. Mit der Isomorphie von T und V/U ergibt sich Folgerung 10 auch aus Satz 10. Für viele Zwecke genügt der komplementäre Unterraum, er ist allerdings im Unterschied zum Quotientenraum im Allgemeinen nicht eindeutig bestimmt. Später werden wir die Folgerung und die Existenz des komplementären Unterraums auch ohne die Voraussetzung der Endlichdimensionalität zeigen.

Die Bedeutung des Quotientenraums liegt in folgender Eigenschaft.

Satz 14. *Es sei V ein Vektorraum, U ein Unterraum und $h : V \to V/U$ die natürliche Abbildung. Für jede lineare Abbildung $f : V \to W$ mit der Eigenschaft $U \subseteq \operatorname{Ker} f$ gibt es genau eine lineare Abbildung $g : V/U \to W$, so dass $f = g \circ h$ ist.*

Diese Eigenschaft drückt man auch mit den Worten aus, die Abbildung h sei universell unter allen auf V definierten linearen Abbildungen f mit der Eigenschaft $U \subseteq \operatorname{Ker} f$.

Beweis. Damit die Abbildung g die geforderte Eigenschaft hat, muss für alle $v \in V$ gelten $g(v + U) = f(v)$. Damit ist g eindeutig bestimmt, falls diese Definition korrekt ist, d. h. falls $f(v)$ nicht von der Wahl des Repräsentanten v in $v + U$ abhängt. Ist $v \sim_U w$, so ist $v - w \in U \subseteq \operatorname{Ker} f$, also $f(v - w) = 0$ und somit $f(v) = f(w)$, wie verlangt. □

Folgerung 11. *Es sei $f : V_1 \to V_2$ eine lineare Abbildung, U_1 ein Unterraum von V_1 und U_2 ein Unterraum von V_2, so dass für alle $u \in U_1$ gilt $f(u) \in U_2$. Dann gibt es genau eine lineare Abbildung $g : V_1/U_1 \to V_2/U_2$, so dass das Diagramm*

$$\begin{array}{ccc} V_1 & \xrightarrow{f} & V_2 \\ {\scriptstyle h_1}\downarrow & & \downarrow{\scriptstyle h_2} \\ V_1/U_1 & \xrightarrow{g} & V_2/U_2 \end{array}$$

kommutativ ist, wobei $h_i : V_i \to V_i/U_i$ die natürliche Abbildung bezeichnet.

Dazu wenden wir einfach Satz 14 auf die Abbildung $h_2 \circ f$ an. Man nennt g die von f *induzierte Abbildung* zwischen den Quotientenräumen. Aus dem Diagramm ergibt sich

$$g(v + U_1) = f(v) + U_2.$$

4.12 Beziehung zu Gleichungssystemen

Wir setzen nun die Theorie der Vektorräume und linearen Abbildungen in Beziehung zur Theorie linearer Gleichungssysteme. Diese stellt uns Algorithmen zur Verfügung, mit denen allgemein definierte Objekte in Beispielen berechnet werden können.

4.12.1 Urbilder und Lösungräume

Eine lineare Abbildung $f : V \to W$ zwischen endlichdimensionalen Vektorräumen über einem Körper K kann man durch eine Abbildungsmatrix A beschreiben, wenn man Basen dieser Räume wählt. Letztere definieren Isomorphismen $K^n \cong V$ und $K^m \cong W$. Gilt hier sogar Gleichheit von Vektorräumen, so ist einfach

$$f(x) = Ax,$$

wobei $x \in K^n$ als Spaltenvektor geschrieben ist. (Man kommt in diese Situation, indem man V mit K^n und W mit K^m identifiziert, was natürlich logisch nicht ganz korrekt ist.) Nun ist das Urbild

$$f^{-1}(y) = \{x \in K^n \mid f(x) = y\}$$

eines gegebenen Spaltenvektors $y \in K^m$ gleich der Lösungsmenge des linearen Gleichungssystems

$$Ax = y$$

mit gegebener rechter Seite y. Insbesondere ist der Kern von f gleich der Lösungsmenge des zugehörigen homogenen linearen Gleichungssystems

$$Ax = 0,$$

die man wegen Lemma 5 als *Lösungsraum* bezeichnet.

4.12.2 Bilder und Spaltenräume

Definition 27. *Eine Spalte einer Matrix $A \in K^{m \times n}$, betrachtet als Element von K^m, bezeichnet man als Spaltenvektor von A. Die lineare Hülle der Familie der Spaltenvektoren von A bezeichnet man als* Spaltenraum *von A und seine Dimension als* Spaltenrang *von A.*

Die j-te Spalte von A ist das Bild des Vektors e_j der Standardbasis von K^n, und das Bild eines beliebigen Vektors x ist eine Linearkombination dieser Spalten. Somit ist $\operatorname{Im} f$ gleich dem Spaltenraum von A, und $\operatorname{rg} f$ ist gleich dem Spaltenrang von A.

4.12.3 Dualräume und duale Abbildungen

Ist unser Vektorraum V gleich K^n (oder haben wir das erreicht, indem wir V vermittels einer Basis mit K^n identifizieren), so kann der Wert einer Linearform $l \in V^*$ geschrieben werden als

$$l(x) = c_1 x_1 + \ldots + c_n x_n.$$

Wir können dies als Matrizenprodukt

$$l(x) = cx$$

der zugehörigen Koeffizientenmatrix $c = \begin{pmatrix} c_1 & \ldots & c_n \end{pmatrix} \in K^{1\times n}$ mit dem Spaltenvektor $x \in K^n$ schreiben. Der Dualraum $(K^n)^*$ steht also in Bijektion mit dem Raum $K^{1\times n}$ von Zeilenvektoren und wird oft mit diesem identifiziert. Die Dualbasis der Standardbasis besteht aus den Vektoren e_i^*, die wir durch Transponierung aus den Spaltenvektoren e_i erhalten.

Definition 28. *Eine Zeile einer Matrix $A \in K^{m\times n}$, betrachtet als Element von $(K^n)^*$, bezeichnet man als Zeilenvektor von A. Die lineare Hülle der Familie der Zeilenvektoren von A bezeichnet man als* Zeilenraum *von A und seine Dimension als* Zeilenrang *von A.*

Nun bezeichnen wir wieder die von A dargestellte lineare Abbildung mit $f : K^n \to K^m$. Die duale Abbildung f^*, angewendet auf eine Linearform $l \in (K^m)^*$, die wir mit einem Zeilenvektor $c \in K^{1\times n}$ identifizieren, ist definiert durch

$$f^*(l)(x) = l(f(x)) = c(Ax).$$

Wegen der Assoziativität des Matrizenprodukts ergibt sich

$$f^*(l) = cA.$$

Die zu f duale Abbildung ist also nach Identifikation von Linearformen mit Zeilenvektoren durch Rechtsmultiplikation mit A gegeben. Das Bild der dualen Abbildung f^* ist somit der Zeilenraum von A, und $\operatorname{rg} f^*$ ist gleich dem Zeilenrang von A, also dem Spaltenrang der transponierten Matrix A^*.

4.12.4 Lösungsräume und Annullatoren

Ist

$$M = \{a_1, \ldots, a_m\} \subseteq (K^n)^*$$

eine Menge von Zeilenvektoren und fassen wir diese zu einer Matrix A zusammen, so kann man das Gleichungssystem $Ax = 0$ auch in der Form

$$\begin{cases} a_1 x = 0 \\ \quad \vdots \\ a_m x = 0 \end{cases}$$

schreiben. Betrachten wir also den Raum K^n der Spaltenvektoren als Dualraum des Raums $(K^n)^*$ der Zeilenvektoren, so ist der Lösungsraum des obigen Gleichungssystems gerade der Annullator von M.

Die analoge Aufgabe für eine Menge von Spaltenvektoren führt man durch Transponierung auf die vorige Aufgabe zurück.

4.12.5 Rang und Gaußverfahren

Wir stellen uns die Frage, wie man den Zeilen- bzw. Spaltenrang einer Matrix explizit bestimmen kann.

In Abschnitt 2.3 haben wir das Gaußverfahren betrachtet, das mittels einer Folge von elementaren Zeilenoperationen eine gegebene $m \times n$-Matrix A in Zeilenstufenform bringt. Es ist für Matrizen mit Koeffizienten in einem Körper anwendbar. Setzen wir für $i \in \{1, \ldots, m\}$

$$j_i = \inf\{j \in \{1, \ldots, n\} \mid a_{ij} \neq 0\}$$

(wobei $\inf \varnothing = \infty$ ist), so ist eine Matrix genau dann in Zeilenstufenform, wenn die Folge $(1_1, \ldots, i_m)$ streng monoton wachsend ist. Indem man die Rolle von Zeilen und Spalten vertauscht, definiert man elementare Spaltenoperationen, mit denen man eine Matrix in Spaltenstufenform bringen kann.

Satz 15. (i) *Bei elementaren Zeilenoperationen bleibt der Zeilenraum und somit der Zeilenrang unverändert.*

(ii) *Bei einer Matrix in Zeilenstufenform sind die von Null verschiedenen Zeilen linear unabhängig.*

(iii) *Bei elementaren Zeilenoperationen bleibt der Spaltenrang unverändert.*

Analoges gilt bei Vertauschung der Rollen von Zeilen und Spalten.

Beweis. Addieren wir ein Vielfaches einer Zeile zu einer anderen Zeile, so ist die entstehende Zeile eine Linearkombination der bisherigen Zeilen und liegt somit in deren linearer Hülle. Der entstehende Zeilenraum ist also in dem Vorherigen enthalten. Da man die Zeilenoperation umkehren kann, gilt auch die umgekehrte Inklusion, also sind beide Zeilenräume gleich. Dies gilt offensichtlich auch für die anderen elementaren Zeilenoperationen, nämlich die Vertauschung zweier Zeilen und die Multiplikation einer Zeile mit einer von Null verschiedenen Konstanten.

Wir beweisen Behauptung (ii) durch vollständige Induktion nach der Anzahl der von Null verschiedenen Zeilen. Für die Nullmatrix ist nichts zu beweisen. Nun sei A eine von der Nullmatrix verschiedene Matrix in Zeilenstufenform. Die erste von Null verschiedene Eintrag der ersten Zeile sei der j-te. Da A in Zeilenstufenform ist, gilt $a_{ij} = 0$ für $i \neq 1$. Ist eine Linearkombination der nichtverschwindenden Zeilenvektoren mit Koeffizienten c_i gegeben, so ist ihre j-te Koordinate gleich $c_1 a_{1j}$. Ist die Linearkombination gleich Null, so ist also $c_1 = 0$, und es verbleibt eine verschwindende Linearkombination der übrigen nichtverschwindenden Zeilen. Durch Streichung der ersten Zeile entsteht wieder eine Matrix in Stufenform. Nach Induktionsvoraussetzung und Lemma 6 gilt $c_i = 0$ auch für $i > 1$.

Aussage (iii) folgt daraus, dass Zeilenoperationen einen Automorphismus des Raumes der Spaltenvektoren erzeugen. Wenden wir sie auf eine Matrix, also simultan auf alle ihre Spalten an, so ist der Spaltenraum der entstehenden Matrix das Bild des ursprünglichen Spaltenraums unter besagtem Automorphismus. □

Folgerung 12. *Das Gaußverfahren in Zeilen- bzw. Spaltenform erzeugt eine Basis des Zeilen- bzw. Spaltenraumes, wenn man am Ende die verschwindenden Zeilen bzw. Spalten der Matrix in Stufenform streicht.*

Hieran kann man dann den Zeilen- bzw. Spaltenrang ablesen.

4.12.6 Lösungsräume als Spaltenräume

Die allgemeine Lösung eines homogenen linearen Gleichungssystems hängt von Parametern $t_1, \ldots, t_p$ ab. Man kann sie als Linearkombination

$$t_1 v_1 + \ldots + t_p v_p$$

von Spaltenvektoren schreiben.

Satz 16. *Entstehen die Vektoren $v_1, \ldots, v_p$ im Ergebnis des Gaußverfahrens, so bilden sie eine Basis des Lösungsraums.*

Beweis. Laut Definition der allgemeinen Lösung erzeugen sie den Lösungsraum. Fasst man nun diese Spaltenvektoren zu einer $n \times p$-Matrix C zusammen, so kann man den Lösungsraum als ihren Spaltenraum interpretieren. Ordnet man bei dieser Matrix die Zeilen in umgekehrter Reihenfolge an und ordnet bei der entstehenden Matrix die Spalten in umgekehrter Reihenfolge an, so entsteht eine Matrix C' mit den Einträgen $c'_{jk} = c_{n+1-j,p+1-k}$. Die Lösungsmethode für ein Gleichungssystem mit Koeffizientenmatrix A in Zeilenstufenform zeigt, dass C' eine Matrix in Spaltenstufenform ohne Nullspalten ist. Das lässt sich auch als Eigenschaft von C ausdrücken: Setzt man

$$j_k = \sup\{j \in \{1, \dots, n\} \mid c_{jk} \neq 0\},$$

so ist die Folge $(j_1, \dots, j_p)$ streng monoton wachsend. Dies sind nämlich die Nummern der freien Variablen in aufsteigender Reihenfolge.

Nach Satz 15(ii) sind die Spalten von C' linear unabhängig, und nach Satz 15(iii) sind auch die Spalten von C linear unabhängig. □

Beispiel 4.14. Wir betrachten das homogene lineare Gleichungssystem mit der Koeffizientenmatrix in Zeilenstufenform

$$A = \begin{pmatrix} 0 & 1 & 3 & 1 & -2 & 4 \\ 0 & 0 & 0 & 2 & 1 & -2 \end{pmatrix}$$

Bezeichnen wir die freien Variablen x_1, x_3, x_5 und x_6 in dieser Reihenfolge mit t_1, t_2, t_3 und t_4, so ergeben sich die anderen Variablen wie folgt:

$$\begin{aligned} x_4 &= -\frac{1}{2}x_5 + x_6 = -\frac{1}{2}t_3 + t_4 \\ x_2 &= -3x_3 - x_4 + 2x_5 - 4x_6 \\ &= -3t_2 - \left(-\frac{1}{2}t_3 + t_4\right) + 2t_3 - 4t_4 \end{aligned}$$

Die allgemeine Lösung ist also

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & \frac{5}{2} & -5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{pmatrix},$$

was man als Linearkombination der Spalten der der rechts stehenden 6×4-Matrix C interpretieren kann. ◁

4.12.7 Umkehrmatrizen und Systeme mit variabler rechter Seite

Definition 29. *Eine Matrix A heißt* invertierbar*, wenn sie die Abbildungsmatrix einer umkehrbaren Abbildung ist, oder anders ausgedrückt, wenn es eine Matrix B gibt, so dass AB und BA Einheitsmatrizen sind.*

Man nennt B dann die *Umkehrmatrix* oder *inverse Matrix* von A. Wir sehen z. B. aus Folgerung 7, dass dies nur möglich ist, wenn Spalten- und Zeilenzahl übereinstimmen. Invertierbare Matrizen der Ordnung n sind also die invertierbaren Elemente im Ring der $n \times n$-Matrizen. Nach Folgerung 8 ist eine $n \times n$-Matrix A genau dann invertierbar, wenn der Lösungsraum des Gleichungssystems $Ax = 0$ der Nullraum ist. Ob das der Fall ist, sieht man bereits im Ergebnis des Gaußverfahrens: Die streng monotone Folge der j_i muss alle Zahlen von 1 bis n enthalten, d. h. es muss eine *Dreiecksmatrix* ohne Nullen auf der Diagonalen entstehen. Mit Satz 15(ii) folgt, dass solche Matrizen in der Tat invertierbar sind.

Ist $Ax = y$ für Spaltenvektoren x und y, so erhalten wir durch Lösen dieses Gleichungssystems mit variabler rechter Seite ein Gleichungssystem der Form $x = By$. Die Methode aus Abschnitt 2.4 liefert also ein Verfahren zur Bestimmung der Umkehrmatrix: Man wendet Zeilenoperationen auf die Matrix $(A\ 1_n)$ an, bis man eine Matrix der Form $(1_n\ B)$ erhält. Dann ist B die Umkehrmatrix von A, die man auch mit A^{-1} bezeichnet.

4.12.8 Rang und Untermatrizen

Unter einer Untermatrix einer $m \times n$-Matrix A verstehen wir eine Matrix, die durch Streichung einer beliebigen Teilfamilie von Zeilen und Spalten entsteht. Auch die Matrix A selbst und die 0×0-Matrix sind Untermatrizen von A.

Satz 17. *Der Zeilenrang wie auch der Spaltenrang einer Matrix A ist gleich der größten natürlichen Zahl k derart, dass A eine invertierbare Untermatrix der Ordnung k hat.*

Beweis. Der Zeilenrang k ist nach Satz 9 gleich der Mächtigkeit einer maximalen linear unabhängigen Teilfamilie von Zeilen. Auf die von ihnen gebildete Untermatrix können wir das Gaußverfahren anwenden und erhalten eine Matrix in Zeilenstufenform, die nach Satz 15(i) keine Nullzeile hat. Streichen wir die Spalten, die den freien Variablen entsprechen, so erhalten wir eine Dreiecksmatrix ohne Nullen auf der Diagonalen, also eine invertierbare Matrix der Ordnung k. Diese enstand durch Zeilenoperationen aus der entsprechenden Untermatrix der Ordnung k von A, die nach Satz 15(i) ebenfalls den Rang k hat.

Hat umgekehrt eine Untermatrix der Ordnung k von A linear unabhängige Zeilen, so sind auch die entsprechenden vollständigen Zeilen von A linear unabhängig.

Das Gleiche gilt für Spalten an Stelle von Zeilen. □

Aus Lemma 10 und Satz 13 wissen wir, dass Zeilenrang und Spaltenrang einer Matrix übereinstimmen. Der obige Satz liefert einen davon unabhängigen Beweis.

4.12.9 Elementare Operationen und Übergangsmatrizen

Wendet man eine elementare Zeilenoperation auf eine Produktmatrix AB an, so erhält man das gleiche Ergebnis, wie wenn man die selbe Operation vor der Multiplikation auf die Matrix A anwendet. Analog braucht man elementare Spaltenoperationen nur auf die Matrix B anzuwenden. Zum Beweis bezeichnen wir die i-te Zeile der Matrix A mit a_i und die k-te Spalte der Matrix B mit b_k. Dann gilt nach Aufgabe 45

$$\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} B = \begin{pmatrix} a_1 B \\ \vdots \\ a_m B \end{pmatrix}, \qquad A \begin{pmatrix} b_1 & \dots & b_p \end{pmatrix} = \begin{pmatrix} Ab_1 & \dots & Ab_p \end{pmatrix}.$$

Ist insbesondere B gleich der Einheitsmatrix, so sieht man, dass sich die elementaren Spaltenoperationen durch Rechtsmultiplikation mit geeigneten $n \times n$-Matrizen erzeugen lassen, die man *Elementarmatrizen* nennt:

- Die Addition des b-Fachen der i-ten Spalte zur j-ten Spalte durch Rechtsmultiplikation mit der Matrix

  $$1_n + bE_{ij},$$

  wobei der Koeffizient der Matrix E_{ij} mit den Indizes (i, j) gleich 1 ist und alle anderen Koeffizienten gleich Null,

- die Multiplikation der i-ten Spalte mit dem Skalar $a \neq 0$ durch Rechtsmultiplikation mit der Diagonalmatrix, deren Koeffizient an der Stelle (i, i) gleich a ist und an allen anderen Stellen auf der Diagonalen gleich 1,

- die Vertauschung der i-ten und j-ten Spalte durch Rechtsmultiplikation mit der Matrix, die aus der Einheitsmatrix durch die gleiche Spaltenvertauschung entsteht.

Die analogen Zeilenoperationen entstehen natürlich durch Linksmultiplikation mit den zugehörigen transponierten Matrizen.

Die Nacheinanderausführung mehrerer elementarer Spalten- oder Zeilenoperationen entspricht der Multiplikation mit dem Produkt der jeweiligen Elementarmatrizen. Will man am Ende einer Folge von elementaren Zeilenoperationen diese Matrix S wissen, so sollte man gleich zu Anfang die Matrix A rechts durch eine Einheitsmatrix erweitern und diese in die Operationen einbeziehen, denn nach Aufgabe 45 gilt

$$S(A \;\; 1_m) = (SA \;\; S).$$

Analog kann man die Gesamtmatrix für eine Folge von elementaren Spaltenoperationen ablesen, wenn man zu Beginn die Matrix unten durch eine Einheitsmatrix erweitert und diese in die Operationen einbezieht, denn

$$\begin{pmatrix} A \\ 1_n \end{pmatrix} T = \begin{pmatrix} AT \\ T \end{pmatrix}.$$

Ist A eine beliebige Abbildungsmatrix, so gewinnen wir mit Hilfe des Gauß-Verfahrens und anschließenden Spaltenoperationen eine Matrix $A' = SAT$ der in Folgerung 7 angegebenen Form. Wenden wir dabei die Zeilen- und Spaltenoperationen wie eben auf geeignet erweiterte Matrizen an, so können wir die Übergangsmatrizen S und T ablesen. In den Bezeichnungen von Lemma 8 ist $T = R^{-1}$ die Übergangsmatrix von der neuen zur alten Basis im Definitionsbereich und S die Übergangsmatrix von der alten zur neuen Basis im Zielbereich.

Beispiel 4.15. Wir betrachten die Koeffizientenmatrix des Gleichungssystems aus Beispiel 2.6

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

Die Zeilenoperationen erstrecken wir auf die erweiterte Matrix

$$\begin{pmatrix} A & 1_3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 2 & 3 & 1 & 0 & 0 \\ -2 & 2 & 1 & 3 & 0 & 1 & 0 \\ 3 & -1 & 1 & -3 & 0 & 0 & 1 \end{pmatrix}.$$

Addieren wir das Doppelte der ersten Zeile zur zweiten und subtrahieren ihr Dreifaches von der dritten, so erhalten wir

$$\begin{pmatrix} 1 & 2 & 2 & 3 & 1 & 0 & 0 \\ 0 & 6 & 5 & 9 & 2 & 1 & 0 \\ 0 & -7 & -5 & -12 & -3 & 0 & 1 \end{pmatrix}$$

Als Nächstes multiplizieren wir die zweite Zeile mit $\frac{1}{6}$ und addieren das 7-Fache der resultierenden Zeile zur dritten:

$$\begin{pmatrix} 1 & 2 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & \frac{5}{6} & \frac{3}{2} & \frac{1}{3} & \frac{1}{6} & 0 \\ 0 & 0 & \frac{5}{6} & -\frac{3}{2} & -\frac{2}{3} & \frac{7}{6} & 1 \end{pmatrix}$$

Als letzte Zeilenoperation multiplizieren wir die dritte Zeile mit $\frac{6}{5}$ und erhalten

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

Hier lesen wir die Matrizen

$$\tilde{A} = \begin{pmatrix} 1 & 2 & 2 & 3 \\ 0 & 1 & \frac{5}{6} & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{9}{5} \end{pmatrix}, \qquad S = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{6} & 0 \\ -\frac{4}{5} & \frac{7}{5} & \frac{6}{5} \end{pmatrix}$$

ab, wobei $\tilde{A}$ das Zwischenergebnis nach Durchführung der Zeilenoperationen ist und all diese Operationen durch Linksmultiplikation mit S auf einen Schlag durchgeführt werden.

Zur Anwendung von Spaltenoperationen erweitern wir $\tilde{A}$ nach unten. Wir subtrahieren das Doppelte der ersten Spalte von der zweiten und dritten sowie das Dreifache der ersten Spalte von der vierten:

$$\begin{pmatrix} \tilde{A} \\ 1_4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 2 & 3 \\ 0 & 1 & \frac{5}{6} & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{9}{5} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \frac{5}{6} & \frac{3}{2} \\ 0 & 0 & 1 & -\frac{9}{5} \\ 1 & -2 & -2 & -3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Nun subtrahieren wir das $\frac{5}{6}$-Fache der zweiten Spalte von der dritten sowie ihr $\frac{3}{2}$-Faches von der vierten. Im letzten Schritt addieren wir das $\frac{9}{5}$-Fache der

dritten Spalte zur vierten:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\frac{9}{5} \\ 1 & -2 & -\frac{1}{3} & 0 \\ 0 & 1 & -\frac{5}{6} & -\frac{3}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & -2 & -\frac{1}{3} & -\frac{3}{5} \\ 0 & 1 & -\frac{5}{6} & -3 \\ 0 & 0 & 1 & \frac{9}{5} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Hier lesen wir die Matrizen

$$A' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & -2 & -\frac{1}{3} & -\frac{3}{5} \\ 0 & 1 & -\frac{5}{6} & -3 \\ 0 & 0 & 1 & \frac{9}{5} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

ab, wobei die Rechtsmultiplikation mit T all unsere Spaltenoperationen auf einen Schlag durchführt und die Gleichheit

$$A' = SAT$$

gilt. $\triangleleft$

4.12.10 Basen von Summen und Durchschnitten

Sind Unterräume U und W von K^n als Lösungsräume homogener linearer Gleichungssysteme mit Koeffizientenmatrix A bzw. B gegeben, so ist $U \cap W$ der Lösungsraum des homogenen Gleichungssystems mit der Koeffizientenmatrix $\begin{pmatrix} A \\ B \end{pmatrix}$. Sind U und W hingegen als lineare Hüllen endlicher Mengen von Vektoren gegeben, also als Spaltenräume von Matrizen A bzw. B, so ist $U + W$ der Spaltenraum der Matrix $\begin{pmatrix} A & B \end{pmatrix}$. Will man im ersten Fall $U + W$ oder im zweiten Fall $U \cap W$ bestimmen, so kann man das Problem mit Hilfe von Satz 12(iv) und Abschnitt 4.12.4 auf das eben Betrachtete zurückführen. Im zweiten Fall geht es direkter mit dem *Zassenhaus-Algorithmus*.

Satz 18 (Zassenhaus)**.** *Es seien Matrizen A und B von gleicher Spaltenzahl n gegeben. Wir bezeichnen den Zeilenraum von A mit U und den von B mit W. Bringt man die Matrix*

$$\begin{pmatrix} A & A \\ B & 0_{m,n} \end{pmatrix}$$

durch Zeilenoperationen in Zeilenstufenform und schreibt diese als

$$\begin{pmatrix} C & E \\ 0_{k,n} & D \\ 0_{l,n} & 0_{l,n} \end{pmatrix},$$

wobei C und D keine Nullzeilen haben, dann bilden die Zeilen von C eine Basis von $U+W$ und die Zeilen von D eine Basis von $U \cap W$.

Beweis. Es sei $V = (K^n)^*$ der Raum der Zeilenvektoren. Wir betrachten den Unterraum

$$H = \{(u,u) \mid u \in U\} + \{(w,0) \mid w \in W\}$$

von $V \oplus V$ und definieren die Abbildung $p : V \oplus V \to V$ durch $p(v_1, v_2) = v_1$. Offensichtlich ist

$$\operatorname{Ker} p = \{0\} \oplus V, \qquad \operatorname{Ker}(p|_H) = H \cap (\{0\} \oplus V).$$

Man prüft leicht nach, dass

$$\operatorname{Im}(p|_H) = U + W, \qquad H \cap (\{0\} \oplus V) = \{0\} \oplus (U \cap W).$$

Mit Satz 11 folgt

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

Der Unterraum $U+W$ ist der Zeilenraum von $\binom{A}{B}$, der Unterraum H ist der Zeilenraum der gesamten Ausgangsmatrix. Nach Satz 15(i) ist $U+W$ gleich dem Zeilenraum von C und H gleich dem Zeilenraum der resultierenden Gesamtmatrix. Folglich enthält $H \cap (\{0\} \oplus V)$ den Zeilenraum von $\begin{pmatrix} 0_{k,n} & D \end{pmatrix}$, d. h. $U \cap W$ enthält den Zeilenraum von D. Da die Anzahl der Zeilen von C und D zusammen gleich der Dimension von H ist, muss die Anzahl der Zeilen von D gleich der Dimension von $U \cap W$ sein. □

5 Determinanten

Wir wollen nun Determinanten von Matrizen beliebiger Ordnung einführen. Dabei betrachten wir Matrizen mit Koeffizienten in einem beliebigen kommutativen Ring R. Ihre Multiplikation ist wie bei Matrizen über Körpern definiert und ist weiterhin assoziativ und distributiv. Man kann solche Matrizen übrigens auch als Abbildungsmatrizen interpretieren, wobei man den Begriff des Vektorraums über einem Körper durch den allgemeineren Begriff eines Moduls über einem Ring ersetzt. Allerdings muss man sich auf endlich erzeugte Moduln beschränken, die eine Basis besitzen (sogenannte freie Moduln), die also isomorph zu R^n für eine natürliche Zahl n sind. Auch die Menge der $m \times n$-Matrizen mit Koeffizienten in R wird zu einem freien Modul $R^{m \times n}$, dessen Standardbasis von den Matrizen E_{ij} gebildet wird.

5.1 Definition und Eigenschaften

In Abschnitt 2.5 haben wir Determinanten der Ordnung 2 und 3 eingeführt. Dabei haben wir gesehen, dass sie linear von jeder Spalte abhängen und bei der Vertauschung zweier Spalten das Vorzeichen wechseln. Es stellt sich heraus, dass sie durch diese Eigenschaften fast schon charakterisiert sind, und das ist der Schlüssel für die Verallgemeinerung.

Satz 19 (Weierstraß)**.** *Es sei R ein kommutativer Ring, n eine natürliche Zahl und $c \in R$. Dann gibt es genau eine Abbildung $f : R^{n \times n} \to R$ mit folgenden Eigenschaften:*

(i) *Es gilt $f(1_n) = c$.*

(ii) *Halten wir alle Spalten von A bis auf eine fest, so ist $f(A)$ als Funktion dieser Spalte eine lineare Abbildung $R^n \to R$.*

(iii) *Stimmen zwei Spalten von A überein, so ist $f(A) = 0$.*

Für diese Abbildung f gilt dann

(iv) *Bei Vertauschung zweier Spalten von A geht $f(A)$ in das entgegengesetzte Element über.*

Außerdem ist $f(A) = f(A^*)$, und die obigen Eigenschaften gelten auch für Zeilen an Stelle von Spalten.

Schreiben wir eine Matrix A als Blockmatrix von Spaltenvektoren $v_j \in R^n$, das heißt

$$A = (v_1, \ldots, v_n)$$

(wobei wir zur Platzersparnis Kommata statt Lerräume setzen), so bedeutet Eigenschaft (ii), dass für jeden Index $k \in \{1, \dots, n\}$ und jedes $a \in R$ gilt

$$f(v_1, \dots, v_{k-1}, a \cdot v_k, v_{k+1}, \dots, v_n) = a \cdot f(v_1, \dots, v_{k-1}, v_k, v_{k+1}, \dots, v_n)$$

sowie für weitere Spaltenvektoren v'_j, $v''_j \in R^n$

$$\begin{aligned} &f(v_1, \dots, v_{k-1}, v'_k + v''_k, v_{k+1}, \dots, v_n) \\ &\qquad = f(v_1, \dots, v_{k-1}, v'_k, v_{k+1}, \dots, v_n) + f(v_1, \dots, v_{k-1}, v''_k, v_{k+1}, \dots, v_n). \end{aligned}$$

Wenn man einen Index k sowie Spaltenvektoren v_j für $j \neq k$ festhält und dann die Bezeichnung

$$g(v) = f(v_1, \dots, v_{k-1}, v, v_{k+1}, \dots, v_n)$$

einführt, so bedeutet Eigenschaft (ii), dass $g : R^n \to R$ linear ist, wobei wir Definition 23 sinngemäß auf den Fall von R-Moduln verallgemeinern.

Im Fall $n = 0$ ist die einzige lineare Abbildung $R^0 \to R^0$ die identische Abbildung. Darum legt man fest, dass die einzige 0×0-Matrix die Einheitsmatrix 1_0 ist.

Beweis. Angenommen, die Abbildung f exsitiert. Wir zeigen zunächst Eigenschaft (iv), d. h., wenn k, $l \in \{1, \dots, n\}$ verschiedene Indizes sind, sagen wir $k < l$, und wir Spaltenvektoren v_j für $j \notin \{k, l\}$ festhalten sowie

$$h(v', v'') = f(v_1, \dots, v_{k-1}, v', v_{k+1}, \dots, v_{l-1}, v'', v_{l+1}, \dots, v_n)$$

setzen, so gilt

$$h(v'', v') = -h(v', v'').$$

Durch zweimalige Anwendung von (ii) erhalten wir nämlich

$$h(v' + v'', v' + v'') = h(v', v') + h(v', v'') + h(v'', v') + h(v'', v'').$$

Nach (iii) verschwindet die linke Seite sowie der erste und letzte Term auf der rechten Seite.

Hat A die Spalten v_j wie oben, so gilt

$$v_j = \sum_{i=1}^{n} a_{ij} e_i,$$

wobei $(e_1, \dots, e_n)$ wie üblich die Standardbsis bezeichnet. Nun folgt durch mehrfache Anwendung von Eigenschaft (ii) (streng genommen, durch vollständige Induktion)

$$f(A) = \sum_{i_1=1}^{n} \dots \sum_{i_n=1}^{n} a_{i_1,1} \cdots a_{i_n,n} f(e_{i_1}, \dots, e_{i_n}),$$

wobei wir für verschiedene Summationen verschiedene Indexvariablen wählen müssen. In den einzelnen Summanden ist jedem zweiten Index j ein erster Indes i_j zugeordnet, so dass durch $\sigma(j) = i_j$ jeweils eine Abbildung π der Menge $\{1, \ldots, n\}$ in sich selbst gegeben ist. Nach Eigenschaft (iii) sind nur diejenigen Summanden nicht Null, für die die Zahlen $i_1, \ldots, i_n$ paarweise verscheiden sind, d. h. für die die Abbildung σ injektiv ist. Da ihr Definitionsbereich und ihr Zielbereich die gleiche endliche Mächtigkeit haben, sind diese σ bijektiv, das heißt, Elemente der symmetrischen Gruppe S_n. Wir erhalten also

$$f(A) = \sum_{\sigma \in S_n} f(e_{\sigma(1)}, \ldots, e_{\sigma(n)}) \prod_{j=1}^{n} a_{\sigma(j),j}.$$

Da die Addition in R kommutativ ist, brauchen wir auf S_n keine Ordnung festzulegen. Ist $\tau \in S_n$ eine Transposition und $\pi = \sigma \circ \tau$, so folgt aus Eigenschaft (iv)

$$f(e_{\pi(1)}, \ldots, e_{\pi(n)}) = -f(e_{\sigma(1)}, \ldots, e_{\sigma(n)}).$$

Nach Lemma 3 können wir σ als Verkettung von Transpositionen darstellen, und mit Satz 7 folgt

$$f(e_{\sigma(1)}, \ldots, e_{\sigma(n)}) = \operatorname{sgn}(\sigma) f(e_1, \ldots, e_n),$$

wobei wir -1 als entgegengesetztes Element des Einselements von R auffassen oder $(-1)c = -c$ setzen. Mit Eigenschaft (i) folgt

$$f(A) = c \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j}.$$

Damit ist die Eindeutigkeit von f bewiesen.

Um die Existenz zu zeigen, müssen wir nur nachprüfen, dass die durch die letzte Formel definierte Abbildung die Eigenschaften (i)–(iii) hat. Bei (i) ist das offensichtlich, und (ii) folgt leicht aus dem Distributivgesetz. Stimmen nun die k-te und die l-te Spalte von A überein und ist τ die Transposition von k und l, so gilt für $\pi = \sigma \circ \tau$ und alle j

$$a_{\sigma(k),k} = a_{\pi(l),l}, \qquad a_{\sigma(l),l} = a_{\pi(k),k}, \qquad a_{\sigma(j),j} = a_{\pi(j),j} \quad (j \notin \{k, l\}).$$

Sach Satz 7 gilt $\operatorname{sgn}(\sigma) + \operatorname{sgn}(\pi) = 0$, also heben sich die zu σ und π gehörigen Summanden auf. Da S_n in disjunkte Teilmengen der Form $\{\sigma, \sigma \circ \tau\}$ zerfällt, folgt Eigenschaft (iii).

Analog zeigt man die Existenz und Eindeutigkeit einer Abbildung $g : R^{n \times n} \to R$, die die Eigenschaften (i)–(iii) für Zeilen an Stelle von Spalten

hat, wobei sich ergibt

$$g(A) = c \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i,\pi(i)}.$$

Ist $\pi = \sigma^{-1}$, so gilt nach Satz 7, dass $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\pi)$ ist, während aus der Kommutativität der Multiplikation in R folgt

$$\prod_{j=1}^{n} a_{\sigma(j),j} = \prod_{i=1}^{n} a_{i,\pi(i)}.$$

Da die Abbildung, die jeder Permutation ihr Inverses zuordnet, eine bijektive Abbildung von S_n auf sich selbst ist, folgt $f(A) = g(A)$. □

Definition 30. *Für einen kommutativen Ring R definieren wir die Abbildung*

$$\det : R^{n \times n} \to R$$

als die Abbildung mit den Eigenschaften (i)–(iii) aus Satz 19 für $c = 1$. Ist A eine $n \times n$-Matrix mit Koeffizienten in R, so nennt man $\det A$ *die* Determinante *von A. Man schreibt sie wie die Matrix als Zahlenschema, wobei man die Klammern durch senkrechte Striche ersetzt.*

Folgerung 13. *Die Determinante einer $n \times n$-Matrix mit Koeffizienten in einem kommutativen Ring ist gegeben durch*

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i,\pi(i)}.$$

Dies gilt insbesondere als Gleichheit im Ring der Polynome in den Variablen a_{11}, a_{12}, …, a_{nn} mit Koeffizienten in einem beliebigen Ring. Man kann die Determinante also als Polynom auffassen, aus dem sich die Funktion det im Sinne der Definition durch Einsetzen ergibt.

Beispiel 5.1. Zählt man alle 4! Elemente der Gruppe S_4 auf, so ergibt sich

$$\begin{aligned}\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{1n} \\ a_{21} & a_{22} & a_{23} & a_{2n} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} = \quad & a_{11}a_{22}a_{33}a_{44} \\ & - a_{12}a_{21}a_{33}a_{44} - a_{13}a_{22}a_{31}a_{44} - a_{14}a_{22}a_{33}a_{41} \\ & - a_{11}a_{23}a_{32}a_{44} - a_{11}a_{24}a_{33}a_{42} - a_{11}a_{22}a_{34}a_{43} \\ & + a_{12}a_{23}a_{31}a_{44} + a_{12}a_{24}a_{33}a_{41} + a_{13}a_{22}a_{34}a_{41} + a_{11}a_{23}a_{34}a_{42} \\ & + a_{13}a_{21}a_{32}a_{44} + a_{14}a_{21}a_{33}a_{42} + a_{14}a_{23}a_{31}a_{43} + a_{11}a_{24}a_{32}a_{43} \\ & - a_{12}a_{23}a_{34}a_{41} - a_{12}a_{24}a_{31}a_{43} - a_{13}a_{24}a_{32}a_{41} \\ & - a_{13}a_{21}a_{34}a_{42} - a_{14}a_{21}a_{32}a_{43} - a_{14}a_{23}a_{31}a_{42} \\ & + a_{12}a_{21}a_{34}a_{43} + a_{13}a_{24}a_{31}a_{42} + a_{14}a_{23}a_{32}a_{41}.\end{aligned}$$

Die Terme entsprechen der Identität, den Zweierzykeln, den Dreierzykeln, den Viererzykeln und den Paaren disjunkter Zweierzykeln. ◁

Aus Satz 19 ergibt sich

Folgerung 14. (i) *Bei einer Permutation von Spalten multipliziert sich die Determinante mit dem Vorzeichen der Permutation.*

(ii) *Bei Addition eines Vielfachen einer Spalte zu einer anderen bleibt die Determinante unverändert.*

Aussage (i) ergibt sich nämlich mit Satz 7 durch mehrfache Anwendung von Satz 19(iv). Aus Satz 19(ii) folgt für $k \neq l$

$$\begin{aligned}\det(v_1, \ldots, v_{k-1}, v_k + av_l, v_{k+1}, \ldots, v_n) = \det(v_1, \ldots, v_{k-1}, v_k, v_{k+1}, \ldots, v_n) \\ + a \det(v_1, \ldots, v_{k-1}, v_l, v_{k+1}, \ldots, v_n),\end{aligned}$$

und der letzte Summand verschwindet nach Satz 19(iii).

Wegen dieser Eigenschaften können wir Determinanten mit Koeffizienten in einem nullteilerfreien Ring R mit Hilfe des Gauß-Jordan-Verfahrens berechnen. Dazu ist es am Einfachsten, wenn man R mit der Konstruktion aus Abschnitt 3.3.2 in einen Körper einbettet (der natürlich mit R übereinstimmt, falls R bereits ein Körper ist). Teil (i) des folgenden Satzes zeigt, dass das Gaußverfahren ausreicht. Angesichts von Aufgabe 51 ist auch Aussage (ii) wichtig.

Satz 20. (i) *Die Determinante einer Dreiecksmatrix ist das Produkt ihrer Diagonalkoeffizienten.*

(ii) *Für Matrizen $A \in R^{m\times m}$, $B \in R^{m\times n}$ sowie $D \in R^{n\times n}$ gilt*

$$\det \begin{pmatrix} A & B \\ 0_{n,m} & D \end{pmatrix} = \det A \cdot \det D.$$

Beispiel 5.2. Im Körper der rationalen Zahlen gilt

$$\begin{aligned}\begin{vmatrix}2&3&7\\3&4&2\\5&2&3\end{vmatrix} &= \frac{1}{2}\begin{vmatrix}2&3&7\\6&8&4\\5&2&3\end{vmatrix} = \frac{1}{2}\begin{vmatrix}2&3&7\\0&-1&-17\\5&2&3\end{vmatrix}\\ &= \frac{1}{2\cdot 2}\begin{vmatrix}2&3&7\\0&-1&-17\\10&4&6\end{vmatrix} = \frac{1}{2\cdot 2}\begin{vmatrix}2&3&7\\0&-1&-17\\0&-11&-29\end{vmatrix}\\ &= \frac{1}{2\cdot 2}\begin{vmatrix}2&3&7\\0&-1&-17\\0&0&158\end{vmatrix} = -79.\end{aligned}$$

Wollte man die Rechnung im Ring der ganzen Zahlen durchführen, so müsste man bei jedem Bedürfnis zur Division durch eine Zahl c die Ausgangsdeterminante mit c multiplizieren. Im vorliegenden Fall erhielte man so das $2{\cdot}2$-Fache der Determinante und müsste dann die Nullteilerfreiheit benutzen. ◁

Beweis von Satz 20. **(i)** Eine obere Dreiecksmatrix ist charakterisiert durch die Eigenschaft

$$i > j \quad \Rightarrow \quad a_{ij} = 0.$$

Es tragen also nur diejenigen Permutationen π zur Determinante bei, die die Eigenschaft $\pi(i) \geq i$ für alle i haben. Wegen der Injektivität von π folgt $\pi(n) = n$, $\pi(n-1) = n-1$ usw., also $\pi = \mathrm{id}$. Wir erhalten

$$\det A = \operatorname{sgn}(\mathrm{id}) \prod_{i=1}^{n} a_{ii}.$$

Analog behandelt man eine untere Dreiecksmatrix.

(ii) Ist C die Gesamtmatrix und $\pi \in S_{m+n}$, so gilt für $i > m$ und $\pi(i) \leq m$, dass $c_{i,\pi(i)} = 0$ ist. Es tragen also nur diejenigen Permutationen π zu $\det C$ bei, die die Menge $\{m+1, \ldots, m+n\}$ in sich abbilden. Wegen der Bijektivität müssen sie dann auch $\{1, \ldots, m\}$ in sich abbilden, und die Einschränkungen auf diese beiden Mengen liefern Permutationen $\sigma \in S_m$ und $\varphi \in S_n$ mit $\sigma(i) = \pi(i)$ für $i \leq m$ und $\pi(m+i) = m + \varphi(i)$ für $i \leq n$. Umgekehrt liefert jedes Paar (σ, φ) eine Permutation π. Da π keine Zweiermenge $\{i, j\}$ mit $i \leq m$ und $j > m$ fehlstellt, ist die Anzahl der Fehlstellungen von π gleich der Summe der Anzahlen der Fehlstellungen von σ und φ. Wir erhalten $\operatorname{sgn}(\pi) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\varphi)$ und

$$\det C = \sum_{\sigma \in S_m} \sum_{\varphi \in S_n} \operatorname{sgn}(\sigma)\operatorname{sgn}(\varphi) \prod_{i=1}^{m} c_{i,\sigma(i)} \prod_{i=1}^{n} c_{m+i,m+\varphi(i)}.$$

Die Faktoren, die nicht von φ abhängen, kann man aus der inneren Summe ausklammern. Dann hängt die innere Summe nicht mehr von σ ab, und man kann sie aus der äußeren Summe ausklammern. Mit $c_{i,j} = a_{ij}$ für $i, j \in \{1, \ldots, m\}$ und $c_{m+i,m+j} = d_{ij}$ für $i, j \in \{1, \ldots, n\}$ folgt

$$\det C = \sum_{\sigma \in S_m} \operatorname{sgn}(\sigma) \prod_{i=1}^{m} a_{i,\sigma(i)} \cdot \sum_{\varphi \in S_n} \operatorname{sgn}(\varphi) \prod_{i=1}^{n} d_{i,\varphi(i)}. \qquad \square$$

Es folgt eine weitere wichtige Eigenschaft der Determinante.

Satz 21 (Determinantenproduktsatz)**.** *Für $n \times n$-Matrizen A und B mit Koeffizienten in einem kommutativen Ring gilt*

$$\det(AB) = \det(A) \det(B).$$

Beweis. Halten wir A fest und definieren

$$f(B) = \det(AB), \qquad g(B) = \det(A) \det(B),$$

so gilt für Spaltenvektoren $w_1, \ldots, w_n$

$$A(w_1, \ldots, w_n) = (Aw_1, \ldots, Aw_n), \quad f(w_1, \ldots, w_n) = \det(Aw_1, \ldots, Aw_n).$$

Laut Definition haben f und g die Eigenschaften (i)–(iii) aus Satz 19 mit $c = \det A$, und die Behauptung folgt aus der Eindeutigkeit. $\square$

Man kann zeigen, dass die Behauptung auch im Ring der Polynome in den Variablen $a_{11}, \ldots, a_{nn}, b_{11}, \ldots, b_{nn}$ gilt.

5.2 Entwicklung von Determinanten

Wir wollen nun zeigen, dass sich Determinanten beliebiger Ordnung nach Zeilen und auch nach Spalten entwickeln lassen.

Satz 22 (Laplace)**.** *Es sei A eine $n \times n$-Matrix und A_{ij} die Untermatrix, die durch Streichen der i-ten Zeile und der j-ten Spalte entsteht. Dann gilt für jedes $i \in \{1, \ldots, n\}$*

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

und für jedes $j \in \{1, \ldots, n\}$

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

Beweis. Angesichts von Satz 19 genügt es, den Beweis für die Entwicklung nach einer Zeile zu führen, sagen wir, der i-ten. Wir fassen für jedes j die Beiträge der Permutationen mit der Eigenschaft $\pi(i) = j$ zusammen. Daraus können wir den Faktor a_{ij} ausklammern:

$$\det A = \sum_{j=1}^{n} a_{ij} \sum_{\substack{\pi \in S_n \\ \pi(i)=j}} \operatorname{sgn}(\pi) \prod_{k=1}^{i-1} a_{k,\pi(k)} \prod_{k=i+1}^{n} a_{k,\pi(k)}.$$

Es bleibt zu zeigen, dass der nach a_{ij} folgende Term gleich $(-1)^{i+j} \det A_{ij}$ ist. Dazu halten wir i und j fest und bezeichnen die Koeffizienten von A_{ij} mit a'_{kl}. Das Streichen der i-ten Zeile und der j-ten Spalte bewirkt eine Umnummerierung der nachfolgenden Zeilen bzw. Spalten. Es gilt nämlich für $k \neq i$ und $l \neq j$

$$a_{k,l} = a'_{\alpha(k),\beta(l)},$$

wobei

$$\alpha(k) = \begin{cases} k, & \text{wenn } k < i, \\ k-1, & \text{wenn } k > i. \end{cases}, \qquad \beta(l) = \begin{cases} l, & \text{wenn } l < j, \\ l-1, & \text{wenn } l > j. \end{cases}.$$

Nun gibt es für jedes π mit der Eigenschaft $\pi(i) = j$ genau eine Permutation $\sigma \in S_{n-1}$, so dass das Diagramm

$$\begin{array}{ccc} \{1,\dots,n\} \setminus \{i\} & \xrightarrow{\pi} & \{1,\dots,n\} \setminus \{j\} \\ \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} \\ \{1,\dots,n-1\} & \xrightarrow{\sigma} & \{1,\dots,n-1\} \end{array}$$

kommutativ ist, und umgekehrt liefert jedes $\sigma \in S_{n-1}$ ein $\pi \in S_n$ mit besagter Eigenschaft.

Ergänzen wir die Definitionen noch durch

$$\alpha(i) = n, \qquad \beta(j) = n, \qquad \sigma(n) = n,$$

so werden α, β und σ zu Elementen von S_n, und es gilt weiterhin $\beta \circ \pi = \sigma \circ \alpha$. Die Abbildung α stellt genau die Zweiermengen $\{k, n\}$ mit $i \leq k < n$ fehl und die Abbildung β genau die Zweiermengen $\{l, n\}$ mit $j \leq l < n$. Deren Anzahl ist $n - i$ bzw. $n - j$, und wir erhalten

$$(-1)^{n-j} \operatorname{sgn}(\pi) = \operatorname{sgn}(\sigma)(-1)^{n-i}.$$

Es ist also

$$\operatorname{sgn}(\pi) = (-1)^{i+j} \operatorname{sgn}(\sigma), \qquad a_{k,\pi(k)} = a'_{\alpha(k),\sigma(\alpha(k))},$$

und die Behauptung folgt mit der Definition von $\det A_{ij}$. □

Beispiel 5.3. Auf vielen Gebieten der Mathematik (Interpolationspolynome, Körpererweiterungen, Differentialgleichungen) trifft man auf die so genannte *Vandermondsche Determinante*

$$\Delta_n(x_1,\dots,x_n) = \begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^{n-2} & x_2^{n-2} & \dots & x_n^{n-2} \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{vmatrix}.$$

Um das Gaußverfahren anzuwenden, würde man für alle $i > 1$ das x^{i-1}-fache der ersten Zeile von der i-ten Zeile subtrahieren. Geschickter ist es, zunächst das x_1-fache der vorletzten Zeile von der Letzten zu subtrahieren, dann das x_1-fache der Vorvorletzten von der Vorletzten usw. Wir erhalten

$$\Delta_n(x_1,\dots,x_n) = \begin{vmatrix} 1 & 1 & \dots & 1 \\ 0 & x_2 - x_1 & \dots & x_n - x_1 \\ 0 & (x_2 - x_1)x_2 & \dots & (x_n - x_1)x_n \\ \vdots & \vdots & & \vdots \\ 0 & (x_2 - x_1)x_2^{n-2} & \dots & (x_n - x_1)x_n^{n-2} \end{vmatrix}.$$

Entwickeln wir dies nach der ersten Spalte und ziehen dann die Faktoren aus den übrigen Spalten heraus, so folgt

$$\Delta_n(x_1,\dots,x_n) = (x_2 - x_1)\dots(x_n - x_1)\Delta_{n-1}(x_2,\dots,x_n).$$

Durch vollständige Induktion folgt nun

$$\Delta_n(x_1,\dots,x_n) = \prod_{1\le i<j\le n} (x_j - x_i).$$

Die Vandermondsche Determinante stimmt also mit dem Polynom aus Beispiel 3.10 überein. ◁

Definition 31. *Die* Kofaktoren *einer $n \times n$-Matrix A sind die Skalare*

$$a_{ij}^\sharp = (-1)^{i+j} \det A_{ji}$$

(man beachte die vertauschten Indizes). Die von ihnen gebildete Matrix $A^\sharp$ nennt man die Komplementärmatrix *oder* Adjunkte *von A.*

Satz 23. *Für $n \times n$-Marizen gilt*

$$A^\sharp \cdot A = A \cdot A^\sharp = (\det A) \cdot 1_n.$$

Beweis. Die beiden Formeln bedeuten ausgeschrieben

$$\sum_{j=1}^{n}(-1)^{i+j}a_{kj}\det A_{ij} = \delta_{ik}\det A, \qquad \sum_{i=1}^{n}(-1)^{i+j}a_{il}\det A_{ij} = \delta_{jl}\det A.$$

Der Fall $i = k$ bzw $j = l$ ist genau die Behauptung von Satz 22. Wenden wir den gleichen Satz auf die Matrix an, die aus A durch Ersetzung der i-ten Zeile durch die k-te Zeile bzw. durch Ersetzung der j-ten Spalte durch die l-te Spalte entsteht, so bleiben die Untermatrizen A_{ij} unverändert, und die Behaptung für $i \neq k$ bzw. $j \neq l$ folgt mit Eigenschaft (iii) aus Satz 19. □

5.3 Anwendungen

5.3.1 Die Umkehrmatrix

Ist A invertierbar im Ring $R^{n\times n}$, d. h. gibt es eine Matrix B, so dass $AB = 1_n$, so folgt mit Satz 21, dass $\det(A)\det(B) = 1$, also ist $\det B$ das Inverse von $\det A$ im Ring R. Ist umgekehrt $\det A$ invertierbar, so ist nach Satz 23 auch A invertierbar.

Folgerung 15. *Eine $n\times n$-Matrix A ist genau dann invertierbar, wenn* $\det A$ *invertierbar ist, und dann gilt*

$$A^{-1} = (\det A)^{-1}A^\sharp.$$

Ist R ein Körper, so lautet die Bedingung für die Invertierbarkeit natürlich einfach $\det A \neq 0$.

Die Menge aller invertierbaren $n \times n$-Matrizen mit Koeffizienten in einem Ring R bezeichnet man als *allgemeine lineare Gruppe* (englisch „general linear group“). An Stelle von $(R^{n\times n})^\times$ benutzt man meist die Bezeichnung

$$\mathrm{GL}_n(R) = \{A \in R^{n\times n} \mid \det A \in R^\times\}.$$

Die Einschränkung $\det : \mathrm{GL}_n(R) \to R^\times$ ist nach Satz 21 ein Gruppenhomomorphismus. Sein Kern ist die so genannte *spezielle lineare Gruppe*

$$\mathrm{SL}_n(R) = \{A \in R^{n\times n} \mid \det A = 1\}.$$

Ihre Elemente nennt man *unimodulare Matrizen*.

5.3.2 Rang und Minoren

Aus Folgerung 15 ergibt sich eine Charakterisierung des Ranges einer Matrix durch Determinanten von quadratischen Untermatrizen, die wir kurz als *Unterdeterminanten* oder auch *Minoren* bezeichnen, wobei wir einer 0×0-Matrix die Determinante 1 geben.

Folgerung 16. *Der Rang einer Matrix A mit Koeffzienten in einem Körper ist gleich der größten Zahl k derart, dass A eine nichtverschwindende Unterdeterminante der Ordnung k hat.*

5.3.3 Lineare Gleichungssysteme

Wir können nun die *Cramersche Regel* für Gleichungssysteme beliebiger Ordnung formulieren und beweisen.

Satz 24. *Für jede Lösung des linearen Gleichungssystems*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \qquad\qquad\qquad\qquad\qquad\quad \vdots \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n \end{cases}$$

gilt

$$\det(A)x_i = \det(A_i),$$

wobei die $n \times n$-Matrix A_i aus der Koeffizientenmatrix A durch Ersetzung ihrer i-ten Spalte durch den Koeffizientenvektor der rechten Seite entsteht.

Ist A invertierbar, so können wir beide Seiten durch $\det A$ dividieren und erhalten die Cramersche Regel.

Beweis. Schreiben wir das Gleichungssystem in der Form

$$Ax = b,$$

so ergibt sich mit Satz 23

$$\det(A)x = A^\sharp Ax = A^\sharp b.$$

Vergleichen wir die i-ten Zeilen, so erhalten wir

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

Mit Satz 22 ergibt sich die Behauptung durch Entwicklung von $\det(A_i)$ nach der i-ten Spalte. □

Man kann die Lösung eines linearen Gleichungssystems $Ax = b$, deren $m \times n$-Matrix A Koeffizienten in einem Körper hat, vollständig durch Determinanten beschreiben (was allerdings bei numerischen Verfahren wegen des zu hohen Rechenaufwandes nichts nützt). Das Gleichungssystem ist genau dann lösbar, wenn b im Spaltenraum der Matrix A liegt, also genau dann, wenn der Spaltenrang der Koeffizientenmatrix A gleich dem Spaltenrang der erweiterten Koeffizientenmatrix $(A\ b)$ ist. Satz 16 drückt dies durch Determinanten aus. Wählt man nun eine maximale Untermatrix von A mit nichtverschwindender Determinante, so kann man alle darin nicht vorkommenden Zeilen aus dem System weglassen. Alle darin nicht vorkommenden Spalten gehören zu freien Variablen. Bringt man sie auf die rechte Seite, so erhält man die allgemeine Lösung mit Hilfe der Cramerschen Regel (vgl. Aufgabe 12).

5.3.4 Das charakteristische Polynom

Wir betrachten die Aufgabe, beliebige Potenzen einer quadratischen Matrix zu berechnen. Ist K ein Körper, so ist der Ring $K^{n\times n}$ der quadratischen Matrizen der Ordnung n ein Vektorraum der Dimension n^2. Für eine Matrix $A \in K^{n\times n}$ kann die unendliche Folge $1_n = A^0$, $A = A^1$, A^2, $\ldots$ nicht linear unabhängig sein. Es gibt eine kleinste natürliche Zahl $k \leq n^2$, so dass A^k von den kleineren Potenzen linear abhängig ist, d. h. es gibt Skalare c_1, $\ldots$, $c_n \in K$, so dass

$$A^k = c_1 A^{k-1} + \ldots + c_{k-1} A + c_k 1_n.$$

Hat man diese erst einmal gefunden, so kann man noch höhere Potenzen mit Hilfe der Rekursionsformel

$$A^{k+l} = c_1 A^{k+l-1} + \ldots + c_{k-1} A^{l+1} + c_k A^l$$

ohne weitere Matrizenmultiplikation berechnen.

Dieses Ergebnis kann man umformulieren, wodurch weitere Zusammenhänge ans Tageslicht kommen. Ist allgemein eine $n \times n$-Matrix A mit Koeffizienten in einem kommutativen Ring R gegeben, so können wir jedem Term, in dem Elemente R und eine Variable t durch Addition und Multiplikation verknüpft sind, eine Matrix zuordnen, indem wir A an Stelle von t einsetzen. Äquivalente Terme führen auf die selbe Matrix, weil der von den Skalaren und A erzeugte Unterring von $R^{n\times n}$ kommutativ ist. (Wenn wir zwei gegebene Matrizen in Terme mit zwei Variablen einsetzten, wäre das nicht der Fall.) Auf diese Weise erhalten wir einen Homomorphismus von Ringen

$$R[t] \to R^{n\times n}.$$

Hat z. B. ein Polynom $q(t)$ die Standardform

$$q(t) = b_0 t^k + b_1 t^{k-1} + \ldots + b_{k-1} t + b_k,$$

so erhalten wir die Matrix

$$q(A) = b_0 A^k + b_1 A^{k-1} + \ldots + b_{k-1} A + b_k 1_n.$$

Nun wird klar, warum es so wichtig war, Polynome im Unterschied zu ganzrationalen Funktionen als Äquivalenzklassen von Termen zu definieren. Unsere obige Diskussion zeigt, dass es für jede $n \times n$-Matrix A mit Koeffizienten in einem Körper K ein Polynom $q(t) \in K[t]$ vom Grad höchstens n^2 gibt, so dass $q(A) = 0$ ist, aber seine Bestimmung ist ziemlich umständlich. Überraschender Weise kann man ein solches Polynom sofort hinschreiben.

Definition 32. *Ist A eine $n \times n$-Matrix mit Koeffizienten in einem kommutativen Ring R, so nennen wir*

$$\det(t \cdot 1_n - A) \in R[t]$$

das charakteristische Polynom *von A.*

Dieses Polynom hat den Grad n, und sein höchster Koeffizient ist 1 (Übungsaufgabe). In der älteren Literatur betrachtet man statt dessen das Polynom

$$\det(A - t \cdot 1_n),$$

das sich von unserem um den Faktor $(-1)^n$ unterscheidet. Jetzt erkennt man, warum es wichtig war, in diesem Kapitel Matrizen mit Koeffizienten in einem kommutativen Ring wie z. B. $R[t]$ und nicht nur in Körpern zuzulassen.

Beispiel 5.4. Die Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

hat das charakteristische Polynom

$$p(t) = (t-a)(t-d) - bc = t^2 - (a+d)t + (ad - bc),$$

und wir erhalten

$$p(A) = \begin{pmatrix} 0 & b \\ c & d-a \end{pmatrix} \begin{pmatrix} a-d & b \\ c & 0 \end{pmatrix} - \begin{pmatrix} bc & 0 \\ 0 & bc \end{pmatrix},$$

was gleich der Nullmatrix ist. ◁

Satz 25 (Cayley-Hamilton)**.** *Ist A eine $n \times n$-Matrix mit Koeffizienten in einem kommutativen Ring R und $p(t)$ ihr charakteristisches Polynom, so gilt*

$$p(A) = 0_{n,n}.$$

Beweis. Die Matrix $B(t) = t \cdot 1_n - A \in R[t]^{n \times n}$ hat die Koeffizienten

$$b_{ij}(t) = \delta_{ij} t - a_{ij},$$

und es gilt $p(t) = \det B(t)$. Nach Satz 23 ist $B(t)B(t)^\sharp = p(t) \cdot 1_n$, d. h. in $R[t]$ gilt für alle i und k

$$\delta_{ik} p(t) = \sum_{j=1}^{n} b_{ij}(t) c_{jk}(t) = \sum_{j=1}^{n} c_{jk}(t) b_{ij}(t)$$

mit gewissen Polynomen $c_{jk}(t)$. Setzen wir A an Stelle der Variablen t ein, so erhalten wir eine Gleichheit im Matrizenring $R^{n \times n}$. Wenden wir nun beide Seiten auf den Spaltenvektor e_i der Standardbasis an und summieren über i, so erhalten wir

$$p(A)e_k = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{jk}(A) b_{ij}(A) e_i = \sum_{j=1}^{n} c_{jk}(A) \sum_{i=1}^{n} b_{ij}(A) e_i,$$

wobei wir im letzten Schritt benutzt haben, dass $c_{jk}(t)$ nicht von i abhängt. Nach Definition des Kronecker-Symbols und der Matrizenmultiplikation ist

$$\sum_{i=1}^{n} b_{ij}(A) e_i = \sum_{i=1}^{n} \delta_{ij} A e_i - \sum_{i=1}^{n} a_{ij} e_i = A e_j - A e_j = 0,$$

denn Ae_j ist die j-te Spalte von A, und es folgt

$$p(A)e_k = 0.$$

Da k beliebig war, ist $p(A)$ gleich der Nullmatrix. □

5.3.5 Determinante, Spur und charakteristisches Polynom eines Endomorphismus

Man kann auch die Determinante einer linearen Abbildung eines Vektorraumes in sich selbst definieren.

Lemma 12. *Ist V ein Vektorraum über einem Körper K und $f : V \to V$ eine lineare Abbildung, so hängt die Determinante der Abbildungsmatrix von f bezüglich einer Basis von V nicht von der Wahl der Basis ab.*

Beweis. Nach Lemma 8 ist die Abbildungsmatrix bezüglich einer anderen Basis gleich $A' = SAS^{-1}$, wobei S die Übergangsmatrix bezeichnet, und nach Satz 21 gilt $\det A' = \det A$. □

Wir bezeichnen die Determinante einer Abbildungsmatrix von f mit $\det f$. Nun können wir auch das *charakteristische Polynom*

$$p(t) = \det(t \cdot \mathrm{id}_V - f)$$

von f definieren. Es stimmt natürlich mit dem charakteristischen Polynom der Abbildungsmatrix von f bezüglich irgendeiner Basis überein. Schreiben wir

$$p(t) = t^n + a_1 t^{n-1} + \ldots + a_{n-1} t + a_n,$$

so ist $a_n = (-1)^n \det f$. Man nennt $-a_1$ die *Spur* (engl. *trace*) von f, abgekürzt $\operatorname{tr} f$. Sie spielt eine Rolle bei der Ableitung der Determinante. Ist A eine Abbildungsmatrix von f, so gilt $\operatorname{tr} f = \operatorname{tr} A$, wobei

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$$

die Spur der Matrix A bezeichnet (Übungsaufgabe). Allgemein lässt sich a_i durch ein homogenes Polynom vom Grad i in den Koeffizienten der Matrix A darstellen.

Ein Unterraum U von V heißt *invariant* unter dem Endomorphismus f von V (oder kurz f-invariant), wenn für alle $u \in U$ gilt $f(u) \in U$.

Lemma 13. *Es sei f ein Endomorphismus des endlichdimensionalen Vektorraums V und U ein f-invarianter Unterraum. Weiter sei g der durch $g(v + U) = f(v) + U$ definierte Endomorphismus von V/U sowie h die Einschränkung von f zu einem Endomorphismus von U. Dann ist das charakteristische Polynom von f das Produkt der charakteristischen Polynome von g und h.*

Beweis. Wir wählen eine Basis $(v_1, \ldots, v_m)$ von U und ergänzen sie zu einer Basis $(v_1, \ldots, v_n)$ von V. Nach Aufgabe 51 hat die Abbildungsmatrix von f die Form

$$\begin{pmatrix} A & B \\ 0_{n-m,m} & D \end{pmatrix},$$

wobei A und D die Abbildungsmatrizen von h und g sind. Die Abbildungsmatrix von $t \cdot \mathrm{id} - f$ ist dann

$$\begin{pmatrix} t \cdot 1_m - A & -B \\ 0_{n-m,m} & t \cdot 1_{n-m} - D \end{pmatrix},$$

und die Behauptung folgt mit Satz 20(ii). □

5.3.6 Orientierungen

In der Elektrodynamik kommt die Rechte-Hand-Regel vor. Sie dient dazu, eine Klasse von Basen des Anschauungraumes zu charaktrisieren: Bei ihnen ist es möglich, mit dem abgewinkelten Daumen in Richtung des ersten Basisvektors zu zeigen, mit dem ausgestreckten Zeigefinger in Richtung des zweiten Basisvektors und mit dem Mittelfinger in Richtung des Dritten. Solche Basen sind aber nicht von vornherein ausgezeichnet; es wäre z. B. unmöglich, einem entfernten Außerirdischen in Worten mitzuteilen, was wir mit rechts meinen.

Definition 33. *Eine* Orientierung *auf einem endlichdimensionalen Vektorraum V über einem angeordneten Körper ist eine Abbildung o, die jeder Basis von V den Wert 1 oder -1 zuordnet, so dass für zwei beliebige Basen gilt*

$$o(w_1, \dots, w_n) = \operatorname{sgn}(\det S) o(v_1, \dots, v_n),$$

wobei S die Übergangsmatrix von einer der Basen zur anderen bezeichnet.
Wir nennen $-o$ die zu o entgegengesetzte Orientierung.

Dabei betrachten wir Basen als Folgen von Vektoren, also als Abbildungen $\{1, \dots, n\} \to V$. Entsteht eine Basis aus der Anderen durch eine Permutation der Vektoren (oder richtiger, der Indizes), so ist S eine Permutationsmatrix, und $\det S$ ist gleich dem Vorzeichen der Permutation. Die einzigen in dieser Vorlesung behandelten angeordneten Körper sind der Körper der reellen Zahlen und der Körper der rationalen Zahlen. Eine Orientierung auf einem eindimensionalen Vektorraum V bestimmt einen Strahl $\{v \in V \setminus \{0\} \mid o(v) = 1\}$.

Lemma 14. (i) *Auf jedem endlichdimensionalen Vektorraum über einem angeordneten Körper gibt es genau zwei Orientierungen.*

(ii) *Ist U ein Unterraum von V, so kann man von drei Orientierungen*

$$o \text{ auf } V, \qquad o' \text{ auf } U, \qquad o'' \text{ auf } V/U$$

zwei Beliebige vorgeben, und dann ist die Dritte dadurch bestimmt, dass für jede Basis $(v_1, \dots, v_n)$ von V derart, dass $(v_1, \dots, v_k)$ eine Basis von U ist, gelten soll

$$o(v_1, \dots, v_n) = o'(v_1, \dots, v_k) o''(v_{k+1} + U, \dots, v_n + U).$$

Aussage (i) gilt übrigens auch für nulldimensionale Vektorräume $V \cong \{0\}$. In diesem Fall gibt es nur eine Basis, nämlich die Abbildung $\varnothing \to V$. Eine Orientierung ist dann durch ihren Wert auf dieser Basis gegeben, also durch eine Zahl aus $\{1, -1\}$.

Ist $\dim V/U$ eindimensional, so definiert o'' einen Halbraum $\{v \in V \mid o''(v+U) = 1\}$ (von dem aus gewissermaßen die Vorderseite von U zu sehen ist). Aussage (ii) kann man auf den Spezialfall anwenden, dass V die direkte Summe von Unterräumen U und W ist, wobei jeder Orientierung auf W über den Isomorphismus $W \to V/U$ eine Orientierung auf V/U entspricht.

Beweis. Es ist zu zeigen, dass es für eine vorgegebene Basis $(v_1, \dots, v_n)$ und eine vorgegebene Zahl $\varepsilon \in \{1, -1\}$ genau eine Orientierung o gibt, so dass $o(v_1, \dots, v_n) = \varepsilon$. Die Formel in der Definition schreibt vor, wie sich der Wert von o auf jeder anderen Basis ergeben muss. Angenommen, es seien Basen $(u_1, \dots, u_n)$ und $(w_1, \dots, w_n)$ gegeben, und die Übergangsmatrizen von $(v_1, \dots, v_n)$ zu diesen Basen seien R und S. Die Übergangsmatrix von $(u_1, \dots, u_n)$ zu $(w_1, \dots, w_n)$ ist dann $T = SR^{-1}$. Nach Satz 21 gilt $\det(T) = \det(S)\det(R)^{-1}$, und es folgt

$$o(w_1, \dots, w_n) = \operatorname{sgn}\det(T) \cdot o(u_1, \dots, u_n).$$

Da die beiden Basen beliebig waren, ist o eine Orientierung.

Ist ein Unterraum U gegeben und gehen wir zu einer anderen Basis $(w_1, \dots, w_n)$ des Vektorraums V über, so dass $(w_1, \dots, w_k)$ wieder eine Basis von U ist, so hat die Übergangsmatrix zwischen den Basen von V nach Aufgabe 48 die Form

$$S = \begin{pmatrix} S' & R \\ 0_{n-k,k} & S'' \end{pmatrix},$$

wobei S' die Übergangsmatrix zwischen den Basen von U ist und S'' die Übergangsmatrix zwischen den Basen von V/U. Behauptung (ii) folgt nun aus Satz 20(ii). □

Wie der Beweis zeigt, ist eine Orientierung durch ihren Wert auf einer einzigen Basis bestimmt. Die *Standardorientierung* auf dem Vektorraum K^n ist dadurch charakterisiert, dass sie auf der Standardbasis den Wert 1 annimmt.

Ist V ein endlichdimensionaler komplexer Vektorraum, so erhalten wir aus jeder Basis $(v_1, \dots, v_n)$ von V eine Basis $(v_1, iv_1, \dots, v_n, iv_n)$ des unterliegenden reellen Vektorraums. Ist S die Übergangsmatrix zu einer anderen Basis $(w_1, \dots, w_n)$, so hat die Übergangsmatrix zwischen den zugehörigen reellen Basen nach Aufgabe 59(b) die Determinante $|\det S|^2$. Somit gibt es genau eine Orientierung auf dem unterliegenden reellen Vektorraum, die auf allen solchen reellen Basen den Wert 1 hat. Man nennt sie die *kanonische Orientierung*. Ist U einem komplexer Unterraum von V, so sind (wegen der von Aufgabe 59 abweichenden Reihenfolge der Vektoren) die kanonischen Orientierungen auf den unterliegenden reellen Vektorräumen von V, U und V/U miteinander verträglich im Sinne von Lemma 14(ii).

6 Lineare Selbstabbildungen

Wir richten nun unser Augenmerk auf Endomorphismen von Vektorräumen, also lineare Abbildungen f eines Vektorraumes V in sich selbst. Ordnet man jedem Endomorphismus seine Abbildungsmatrix zu, so erhält man nur dann einen Homomorphismus vom Ring der Endomorphismen auf den Ring der quadratischen Matrizen, wenn man ein und die selbe Basis von V sowohl in seiner Eigenschaft als Definitionsbereich als auch als Zielbereich von f wählt. Dann ist aber Folgerung 7 nicht anwendbar, und es stellt sich die Frage, wie man die Basis wählen muss, damit die Abbildungsmatrix von f eine möglichst einfache Form annimmt.

6.1 Physikalische Motivation

Ein Massenpunkt der Masse m sei an einer Feder aufgehängt. Er bewege sich nur in senkrechter Richtung, und x sei seine Auslenkung vom Ruhepunkt als Funktion der Zeit t. In der Ruhelage wird die Schwerkraft durch die Grundspannung der Feder ausgeglichen, und wenn das Hooksche Gesetz für unsere Feder gilt, so ist bei einer Auslenkung x die resultierende Kraft gleich $-kx$, wobei man k als Federkonstante bezeichnet. Wird aber der Aufhängungspunkt auf und ab bewegt, wobei seine Auslenkung durch eine Funktion h von t gegeben ist, so ist die auf den Massenpunkt einwirkende Kraft gleich $k(h-x)$, und nach dem *Newtonschen Gesetz* gilt

$$m\ddot{x} = k(h-x),$$

wobei die Ableitung in Newtonscher Schreibweise durch einen übergesetzten Punkt gekennzeichnet wird. Dies ist eine *lineare Differentialgleichung* mit gegebener Funktion h und gesuchter Funktion x. Wie in Lemma 5 zeigt man, dass ihr Lösungsraum ein affiner Unterraum X des Vektorraums $C^\infty(\mathbb{R},\mathbb{R})$ der beliebig oft differenzierbaren Funktionen $x : \mathbb{R} \to \mathbb{R}$ ist.

Ist beispielsweise $h(t) = b\sin\varphi t$ mit einer *Kreisfrequenz* $\varphi > 0$, so liegt der Ansatz

$$x(t) = c\sin\varphi t$$

nahe, und bei Einsetzen in die Differentialgleichung ergibt sich

$$-mc\varphi^2\sin\varphi t = k(b-c)\sin\varphi t.$$

Wir erhalten also tatsächlich eine Lösung, wenn wir

$$c = \frac{kb}{k-m\varphi^2}$$

setzen, vorausgesetzt,

$$\varphi \neq \sqrt{\frac{k}{m}}.$$

Der Massenpunkt führt in diesem Fall eine *erzwungene Schwingung* aus.

Der zum affinen Lösungsraum X gehörige Vektorraum V ist die Lösungsmenge der zugehörigen homogenen Gleichung

$$m\ddot{x} = -kx.$$

Er besteht aus Lösungen, die bei festgehaltenem Aufhängungspunkt auftreten und deshalb als *Eigenschwingungen* bezeichnet werden. Der Ansatz

$$x(t) = \sin \omega t \qquad \text{bzw.} \qquad x(t) = \cos \omega t$$

mit unbestimmtem $\omega > 0$ führt durch Einsetzen auf die Bedingung $m\omega^2 = k$, also erhalten wir zwei Lösungen mit der sogenannten Eigenfrequenz

$$\omega = \sqrt{\frac{k}{m}}.$$

Man kann zeigen, dass sie eine Basis von V bilden. Die allgemeine Lösung der inhomogenen Gleichung (ein beliebiges Element von X) ergibt sich dann wie in Lemma 5 als Überlagerung einer erzwungenen Schwingung und einer Eigenschwingung.

Es gibt noch mehr Beziehungen zur linearen Algebra. Definieren wir einen Endomorphismus d des Vektorraums $C^\infty(\mathbb{R}, \mathbb{R})$ durch

$$d(x) = \dot{x},$$

so können wir die obige homogene Differentialgleichung in der Form

$$d^2(x) = -\frac{k}{m}x$$

schreiben, und der Lösungsraum $V = \operatorname{Ker}\left(d^2 + \frac{k}{m}\mathrm{id}\right)$ ist invariant unter d. Suchen wir Lösungen mit dem Ansatz

$$x(t) = e^{\lambda t},$$

so gilt

$$d(x) = \lambda x,$$

und die Bedingung an die Lösung lautet

$$\lambda^2 = -\frac{k}{m}.$$

Solche Lösungen existieren also nur im größeren Raum $\mathbb{C}^\infty(\mathbb{R},\mathbb{C})$ der komplexwertigen Funktionen $x : \mathbb{R} \to \mathbb{C}$, und dann ist $\lambda = \pm i\omega$. Die obige Basis von V ist auch eine Basis des komplexen Lösungsraums W, aber die Funktionen

$$e^{i\omega t} = \cos\omega t + i\sin\omega t, \qquad e^{-i\omega t} = \cos\omega t - i\sin\omega t$$

bilden ebenfalls eine Basis von W.

Eine gedämpfte Eigenschwingung, bei der die Reibungskraft proportional zur Geschwindigkeit $\dot{x}$ ist, gehorcht der Differentialgleichung

$$m\ddot{x} + l\dot{x} + kx = 0,$$

und die Funktionen der Form $x(t) = e^{\lambda t}$ sind Lösungen, wenn λ eine Nullstelle des Polynoms

$$m\lambda^2 + l\lambda + k$$

ist. Im Fall einer doppelten Nullstelle ist auch $te^{\lambda t}$ eine Lösung, man findet also immer eine Basis des Lösungsraums.

Der Vektorraum in Aufagbe 42 ist übrigens ebenfalls ein Lösungsraum einer homogenen linearen Differentialgleichung, nämlich der Kern von

$$d \circ (d^2 + \mathrm{id}) \circ (d^2 + 4\,\mathrm{id}).$$

6.2 Eigenwerte

Nun haben wir die Motivation für den folgenden mathematischen Begriff.

Definition 34. *Gegeben sei ein Vektorraum V über einem Körper K und eine lineare Abbildung $f : V \to V$. Ein Skalar $\lambda \in K$ heißt* Eigenwert *von f, wenn es einen von Null verschiedenen Vektor $x \in V$ gibt, so dass*

$$f(x) = \lambda x.$$

Ein solcher Vektor heißt dann Eigenvektor *zum Eigenwert λ.*

Analog definiert man Eigenwerte von quadratischen Matrizen. Die Eigenwerte einer Abbildung f sind dann die gleichen wie die ihrer Abbildungsmatrix bezüglich einer beliebigen Basis.

Lemma 15. *Ein Skalar λ ist genau dann Eigenwert des Endomorphismus f eines endlichdimensionalen Vektorraums, wenn er Nullstelle des charakteristischen Polynoms von f ist.*

Beweis. Die Bedingung an λ, Eigenwert zu sein, ist die Existenz eines Vektors $x \neq 0$, so dass

$$(f - \lambda \cdot \mathrm{id})(x) = 0.$$

Dies ist also genau dann der Fall, wenn $\mathrm{Ker}(f - \lambda \cdot \mathrm{id})$ nicht der Nullraum ist, also nach Folgerung 8 genau dann, wenn $f - \lambda \cdot \mathrm{id}$ nicht invertierbar ist. Nach Folgerung 15 ist dies genau dann der Fall, wenn $\det(f - \lambda \cdot \mathrm{id}) = 0$ ist. □

Man nennt $\mathrm{Ker}(f - \lambda \cdot \mathrm{id})$ den *Eigenraum* von f zum Eigenwert λ und seine Dimension die *Vielfachheit* des Eigenwertes λ. Die von Null verschiedenen Elemente des Eigenraumes sind die Eigenvektoren zum Eigenwert λ.

6.3 Diagonalisierung und Trigonalisierung

Wir wollen feststellen, wann ein endlichdimensionaler Vektorraum eine Basis besitzt, bezüglich derer die Abbildungsmatrix eines gegebenen Endomorphismus eine Diagonalmatrix oder eine Dreiecksmatrix ist.

Satz 26. *Die Summe der Eigenräume eines Endomorphismus ist direkt.*

Wenden wir dies z. B. auf die Eigenfunktionen $e^{\lambda t}$ der Abbildung d aus dem obigen Beispiel an, so sehen wir, dass sie über $\mathbb{C}$ linear unabhängig sind.

Beweis. Wir müssen zeigen, dass die Summenabbildung von der (äußeren) direkten Summe der Eigenräume in den Vektorraum V injektiv ist. Dazu beweisen wir durch vollständige Induktion nach k, dass für beliebige Vektoren $v_1, \ldots, v_k$ in verschiedenen Eigenräumen mit der Eigenschaft

$$v_1 + \ldots + v_k = 0$$

gilt, dass sie alle gleich Null sind. Diese Behauptung gilt für $k = 1$. Angenommen, sie ist für weniger als k Eigenvektoren bewiesen. Es sei $f(v_i) = \lambda_i v_i$. Nun folgt durch Anwendung von $f - \lambda_k \cdot \mathrm{id}$

$$(\lambda_1 - \lambda_k)v_1 + \ldots + (\lambda_{k-1} - \lambda_k)v_{k-1} = 0.$$

Die Summanden sind Eigenvektoren zu $k-1$ verschiedenen Eigenwerten, und nach Induktionsvoraussetzung sind sie alle gleich Null. Wegen der Verschiedenheit der Eigenwerte folgt $v_1 = 0, \ldots, v_{k-1} = 0$. Damit ist auch $v_k = 0$. □

Wählen wir Basen in den einzelnen Eigenräumen, so ist ihre Vereinigung also eine Basis der Summe dieser Räume, und wir erhalten:

Folgerung 17. *Ein endlichdimensionalen Vektorraums V besitzt genau dann eine Basis, die aus Eigenvektoren eines Endomorphismus f besteht, wenn die Summe der Vielfachheiten der Eigenwerte von f gleich der Dimension von V ist.*

Dies lässt sich praktisch realisieren, wenn die Abbildungsmatrix A von f bezüglich einer Basis gegeben ist. Der Eigenraum zu λ ist ja als Kern der linearen Abbildung $f - \lambda \cdot \mathrm{id}$ definiert, also sind die n-Tupel der Koordinaten von Eigenvektoren die Lösungen des Gleichungssystems mit der Koeffizientenmatrix $A - \lambda \cdot 1_n$.

Da die Eigenräume mindestens eindimensional sind, ergibt sich:

Folgerung 18. *Hat das charakteristische Polynom eines Endomorphismus eines n-dimensionalen Vektorraums n verschiedene Nullstellen, so gibt es eine Basis aus Eigenwerten.*

Hat der Eigenwert λ von f die Vielfachheit m, so ist das charakteristische Polynom der Einschränkung von f auf den Eigenraum zu λ gleich $(t - \lambda)^m$. Dies ist nach Lemma 13 ein Teiler des charakteristischen Polynoms von f. Die höchste Potenz von $t - \lambda$, durch die ein Polynom $p(t)$ teilbar ist, heißt Vielfachheit der Nullstelle λ von $p(t)$. Wir erhalten also:

Folgerung 19. *Die Vielfachheit von λ als Eigenwert eines Endomorphismus f ist nicht größer als die Vielfachheit von λ als Nullstelle des charakteristischen Polynoms von f.*

Wenn es eine Basis aus Eigenvektoren gibt, so muss das charakteristische Polynom also in Linearfaktoren zerfallen. Die Umkehrung gilt zwar im Spezialfall von Folgerung 18, aber im Allgemeinen haben wir nur:

Satz 27. *Das charakteristische Polynom eines Endomorphismus f eines endlichdimensionalen Vektorraums V zerfällt genau dann in Linearfaktoren, wenn es f-invariante Unterräume*

$$\{0\} = V_0 \subset V_1 \subset \ldots \subset V_n = V$$

gibt, so dass für alle i gilt $\dim V_i = i$.

Man nennt eine streng wachsende Folge von Unterräumen eine *Fahne*, wobei man sich V_0 als Befestigungspunkt, V_1 als Stange und V_2 als Tuch vorstellt. Eine Fahne heißt *vollständig*, wenn man keine weiteren Unterräume mehr einfügen kann, so dass wieder eine Fahne entsteht. Nach Aufgabe 37 und Folgerung 5 ist das genau der Fall, wenn sämtliche Dimensionen zwischen 0 und $\dim V$ bereits vorkommen.

Beweis. Angenommen, es gibt eine vollständige Fahne von f-invarianten Unterräumen. Für jedes i können wir dann einen Endomorphismus f_i des Quotientenraums V_i/V_{i-1} definieren, indem wir setzen

$$f_i(v + V_{i-1}) = f(v) + V_{i-1}.$$

Es sei p das charakteristische Polynom von f und p_i das von f_i. Durch mehrmalige Anwendung von Lemma 13 sehen wir, dass

$$p(t) = p_1(t) \cdots p_n(t).$$

Nach Folgerung 10 ist $\dim V_i/V_{i-1} = 1$, also $\deg p_i = 1$, das heißt, $p(t)$ zerfällt in Linearfaktoren.

Wir beweisen die Umkehrung durch vollständige Induktion. Für $n = 0$ ist nichts zu beweisen, weil dann $p(t) = 1$ und $V = \{0\}$ ist. Wir nehmen an, dass die Behauptung für Räume der Dimension $n - 1$ bereits bewiesen ist, und betrachten einen n-dimensionalen Vektorraum V, wobei $n \geq 1$ ist. Zerfällt nun das charakteristische Polynom $p(t)$ von $f \in \mathrm{End}(V)$ in Linearfaktoren, so wählen wir einen von ihnen, sagen wir $p_1(t) = t - \lambda_1$. Nach Lemma 15 ist λ_1 ein Eigenwert von f, und wir wählen einen zugehörigen Eigenvektor v_1. Die lineare Hülle V_1 von v_1 ist dann ein eindimensionaler f-invarianter Unterraum. Nach Folgerung 10 ist $\dim V/V_1 = n - 1$, und f induziert einen Endomorphismus g von V/V_1. Ist q das charakteristische Polynom von g, so gilt nach Lemma 13

$$p(t) = p_1(t)q(t).$$

Wegen der Nullteilerfreiheit von $K[t]$ ist $q(t)$ das Produkt der übrigen Linearfaktoren von $p(t)$. Nach Induktionsvoraussetzung gibt es eine g-invariante Fahne

$$\{0\} = W_1 \subset W_2 \subset \ldots \subset W_n = V/V_1,$$

wobei $\dim W_i = i - 1$ ist. Wir bezeichnen das Urbild von W_i unter der kanonischen Abbildung $V \to V/V_1$ mit V_i. Nach Folgerung 10 ist $\dim V_i = i$, diese Unterräume bilden also zusammen mit $V_0 = \{0\}$ eine vollständige Fahne in V. Für $v \in V_i$ ist $v + V_1 \in W_i$, also

$$f(v) + V_1 = g(v + V_1) \in W_i$$

und somit $f(v) \in V_i$. Die besagte Fahne ist also f-invariant. □

Traditionell werden die obigen Ergebnisse in der Sprache der Matrizen formuliert.

Definition 35. (i) *Zwei quadratische Matrizen A und A' heißen* ähnlich, *wenn es eine umkehrbare Matrix T gibt, so dass $A' = T^{-1}AT$ ist.*

(ii) *Eine Matrix heißt* diagonalisierbar*, wenn sie ähnlich zu einer Diagonalmatrix ist.*

(iii) *Eine Matrix heißt* trigonalisierbar*, wenn sie ähnlich zu einer Dreiecksmatrix ist.*

Die Abbildungsmatrizen eines Endomorphismus bezüglich verschiedener Basen sind nach Lemma 8 ähnlich zueinander. Offensichtlich ist die Abbildungsmatrix von f genau dann diagonalisierbar, wenn es eine Basis aus Eigenvektoren von f gibt oder, was dazu äquivalent ist, wenn es eine Basis des Raumes der Spaltenvektoren aus Eigenvektoren von A gibt. Wie man diese aus den allgemeinen Lösungen der homogenen Gleichungssysteme mit den Koeffizientenmatrizen $A - \lambda 1_n$ findet, wurde oben bereits beschrieben. Hat man n linear unabhängige Spaltenvektoren v_i mit der Eigenschaft $Av_i = \lambda_i v_i$ und fügt man diese Spalten zu einer invertierbaren Matrix

$$T = \begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix}$$

zusammen, so ergibt sich mit Aufgabe 45

$$AT = \begin{pmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \dots & \lambda_n v_n \end{pmatrix} = TA',$$

wobei A' die Diagonalmatrix mit den Diagonaleinträgen $\lambda_1, \dots, \lambda_n$ ist.

Folgerung 20. *Eine Matrix ist genau dann trigonalisierbar, wenn ihr charakteristisches Polynom in Linearfaktoren zerfällt (was nach Satz 6 im Fall $K = \mathbb{C}$ immer erfüllt ist).*

Nach Satz 27 müssen wir uns nur überzeugen, dass die Existenz einer vollständigen Fahne von f-invarianten Unterräumen äquivalent zur Trigonalisierbarkeit der Abbildungsmatrix ist. Wählt man einen Basisvektor von V_1 und ergänzt ihn nacheinander zu Basen von V_2, V_3 usw., so erhält man eine Basis, bezüglich derer die Abbildungsmatrix eine obere Dreiecksmatrix ist. (Kehren wir die Reihenfolge der Basisvektoren um, so erhalten wir eine untere Dreiecksmatrix.) Ist umgekehrt die Abbildungsmatrix eine obere bzw. untere Dreiecksmatrix, so ist für jedes i die lineare Hülle der ersten bzw. letzten i Basisvektoren invariant unter f.