Lineare Algebra I

Wintersemester 2016/2017

Universität Bayreuth

Michael Stoll

Inhaltsverzeichnis

Druckversion vom 24. Februar 2017, 14:30 Uhr.

1. Einige allgemeine Vorbemerkungen

Die meisten von Ihnen kommen mehr oder weniger direkt von der Schule (die prominenteste Ausnahme sind die zukünftigen Realschullehrer/innen, die die Lineare Algebra I erst im dritten Semester hören). Das Erste, das Sie sich zu Beginn Ihres Mathematik-Studiums klar machen müssen, ist, dass das, was Sie in der Schule unter der Bezeichnung „Mathematik“ kennen gelernt haben, nicht wirklich Mathematik ist. Das bedeutet, dass Sie hier an der Universität im Grunde auf völlig andere Art und Weise gefordert sein werden als an der Schule. Das heißt jetzt nicht, dass Sie die ganze Schulmathematik vergessen können — manches kann als Beispielmaterial noch nützlich sein, und es schadet auch nicht, wenn man eine gewisse Fertigkeit im Rechnen hat, wie man sie an der Schule lernt.

Was folgt aus diesem in Deutschland leider traditionellen Bruch zwischen Schule und Universität?

- Die meisten von Ihnen werden sich erst einmal sehr schwer tun. Das ist völlig normal und kein Grund zur Beunruhigung.
- Wenn Sie in der Schule in Mathe sehr gut waren, heißt das nicht, dass Ihnen die Mathematik an der Universität auch leicht fällt. Umgekehrt kann es sein, dass Ihnen die Mathematik an der Schule langweilig war und Sie dann hier auf den Geschmack kommen.
- Sie sollten nicht erwarten, den Stoff sofort während der Vorlesung zu verstehen. Das Nacharbeiten der Vorlesung ist sehr wichtig, da man Mathematik nur verstehen kann, wenn man darüber nachdenkt. (Das Modulhandbuch sieht drei Stunden pro Woche dafür vor.) Ganz wichtig ist auch, dass Sie die Übungsaufgaben bearbeiten, denn richtig versteht man den Stoff erst, wenn man ihn anwendet. (Das Modulhandbuch sieht dafür fünf Stunden pro Woche vor.) Dabei hilft es, gemeinsam in kleinen Gruppen zu arbeiten, denn für das Verständnis ist es ungemein förderlich, wenn man versucht, jemand anderem etwas zu erklären.
- Für diejenigen von Ihnen, die Lehrer/innen werden wollen, heißt das umgekehrt auch, dass Sie den größten Teil von dem, was Sie hier lernen, in der Schule nicht direkt verwenden können. Es ist zu hoffen, dass sich das bald einmal ändert und Sie die Möglichkeit haben werden, die „richtige“ Mathematik Ihren Schülern nahezubringen. In jedem Fall sollte Sie die Ausbildung, die Sie an der Universität erhalten, in die Lage versetzen, Ihren Unterricht innerhalb der Mathematik einzuordnen und weiter gehende Fragen Ihrer Schüler/innen souverän zu beantworten.

Lassen Sie sich von den Schwierigkeiten am Anfang nicht zu sehr frustrieren! Bei den meisten von Ihnen wird in den ersten beiden Semestern der Groschen fallen. Falls Sie aber nach zwei Semestern immer noch das Gefühl haben, nichts zu verstehen, dann kann es auch sein, dass das Mathematikstudium doch nicht das Richtige für Sie ist.

Ich habe in dieses Skript an manchen Stellen Links zu Webseiten eingebaut (anklickbar in der Bildschirmversion), die so aussehen (dieser Link führt auf meine Homepage). Die meisten davon verweisen auf die Wikipedia, die für den Zweck einer ersten Orientierung meistens gut geeignet ist. (Als Hauptquelle für Zitate in einer wissenschaftlichen Arbeit wie z.B. einer Bachelor- oder Masterarbeit ist die Wikipedia aber nicht geeignet. Da müssen Sie Lehrbücher oder Fachartikel zitieren.)

2. Die Sprache der Mathematik: Logik und Mengenlehre

Worum geht es nun in der Mathematik?

Die Wikipedia schreibt (Stand Oktober 2016):

> Für *Mathematik* gibt es keine allgemein anerkannte Definition; heute wird sie üblicherweise als eine Wissenschaft beschrieben, die durch logische Definitionen selbst geschaffene abstrakte Strukturen mittels der Logik auf ihre Eigenschaften und Muster untersucht.

Es geht also unter anderem um *Abstraktion*. Man abstrahiert von den speziellen Eigenschaften, die man in verschiedenen Situationen vorliegen hat, und zieht das Gemeinsame heraus. Dann versucht man auf der Grundlage nur dieser wesentlichen Merkmale möglichst viele Aussagen abzuleiten, die dann auf *alle* Situationen zutreffen, die diese Merkmale aufweisen. Dies geschieht durch den zentralen Vorgang aller mathematischen Tätigkeit, nämlich durch das Führen eines *mathematischen Beweises*. Zugespitzt kann man sagen, dass ein *Mathematiker* der- oder diejenige ist, die oder der in der Lage ist, einen solchen Beweis zu führen:

- Das wichtigste „Lernziel“ in den Grundvorlesungen besteht darin, dass Sie lernen, wie man mathematische Beweise führt.

Sie sollen hier natürlich auch und nicht zuletzt Ergebnisse und Methoden der Linearen Algebra kennen lernen, aber ohne die mathematische Grundfertigkeit des Beweisens würde Ihnen das kaum etwas nützen.

Bevor wir damit beginnen können, müssen wir die Vokabeln und Grammatik der Sprache der Mathematik lernen. Mathematische Aussagen und Beweise werden in der Sprache der *Logik* formuliert; die Objekte, von denen die Rede ist, in der Sprache der *Mengenlehre*. Beide werden wir hier kurz einführen (oder wiederholen, je nachdem wie viel Sie davon schon aus der Schule kennen). Es handelt sich um das „Handwerkszeug“, mit dem Sie täglich zu tun haben werden, also passen Sie gut auf!

2.A Aussagenlogik.

Die Aussagenlogik verknüpft mathematische *Aussagen* (die wahr oder falsch sein können) miteinander und untersucht, wie das Wahr- oder Falschsein einer zusammengesetzten Aussage von den beteiligten Aussagen abhängt. (In der Bildschirmversion des Skripts erscheint der folgende Text in grün. Diese Farbe kennzeichnet Definitionen.)

Die logischen Verknüpfungen sind:

DEF

$\neg A$
$A \wedge B$
$A \vee B$
$A \Rightarrow B$
$A \Leftrightarrow B$

(1) Die *Negation*: wir schreiben „nicht A“ oder „$\neg A$“ für die Verneinung der Aussage A. $\neg A$ ist genau dann wahr, wenn A falsch ist, und umgekehrt.

(2) Die *Konjunktion*: wir schreiben „A und B“ oder „$A \wedge B$“; diese Aussage ist genau dann wahr, wenn sowohl A als auch B wahr sind.

(3) Die *Disjunktion*: wir schreiben „A oder B“ oder „$A \vee B$“; diese Aussage ist genau dann wahr, wenn wenigstens eine der Aussagen A und B wahr ist.

(4) Die *Implikation*: wir schreiben „aus A folgt B“, „A impliziert B“ oder „$A \Rightarrow B$“; diese Aussage ist genau dann wahr, wenn A falsch oder B wahr ist (oder beides).

(5) Die *Äquivalenz*: wir schreiben „A genau dann, wenn B“, „A und B sind äquivalent“ oder „$A \Leftrightarrow B$“; diese Aussage ist genau dann wahr, wenn entweder A und B beide wahr oder A und B beide falsch sind.

Alle hier aufgeführten Schreibweisen sind möglich und erlaubt; die Schreibweise „$A \wedge B$“ ist zum Beispiel nicht besser oder schlechter als „A und B“ (nur kürzer). Bei verschachtelten Verknüpfungen werden Klammern gesetzt, um die Bedeutung klar zu machen: Bei „A und B oder C“ ist sonst nicht klar, ob $(A \wedge B) \vee C$ oder $A \wedge (B \vee C)$ gemeint ist.

Die Definition der logischen Verknüpfungen lässt sich übersichtlich durch die entsprechenden *Wahrheitstafeln* zusammenfassen. Wir schreiben W für *wahr* und F für *falsch*. Dann lässt sich die Negation wie folgt definieren:

A	$\neg A$
W	F
F	W

Die übrigen Verknüpfungen sind gegeben durch:

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

A	B	$A \Rightarrow B$
W	W	W
W	F	F
F	W	W
F	F	W

A	B	$A \Leftrightarrow B$
W	W	W
W	F	F
F	W	F
F	F	W

Die wichtigste (und gleichzeitig die am schwersten zu verstehende) dieser Verknüpfungen ist die Implikation. Sie ist wichtig, weil die große Mehrzahl aller mathematischen Sätze die Form einer Implikation haben: Wenn gewisse Voraussetzungen A gelten, dann folgt eine Aussage B. Sie ist ein wenig schwierig, weil mit ihr im täglichen Leben oft ungenau bis falsch umgegangen wird. Vor allem neigen viele Menschen dazu, zwischen „aus A folgt B“ und „aus B folgt A“ nicht sorgfältig zu unterscheiden. Diesen Unterschied zu begreifen, ist die erste wichtige Hürde für Sie als zukünftige Mathematiker. Machen Sie sich Folgendes klar:

- $A \Rightarrow B$ ist jedenfalls immer dann wahr, *wenn A falsch ist*.
- $A \Rightarrow B$ ist auch immer dann wahr, wenn B wahr ist.
- $A \Rightarrow B$ kann *nur dann* falsch sein, wenn A wahr, aber B falsch ist.

Wir verwenden manchmal die Schreibweise „$\perp$“ für das *Falsum*, also eine stets falsche Aussage oder einen Widerspruch. Analog gibt es die stets wahre Aussage „$\top$“. Dann können wir also schreiben

$$\perp \Rightarrow B \qquad \text{und} \qquad A \Rightarrow \top \qquad \text{gelten stets.}$$

Für die Lateiner unter Ihnen: Die erste dieser Tatsachen ist auch unter dem schönen Namen *Ex falso quodlibet* bekannt.

Folgende Schlussweise ist *nicht* erlaubt:

> Wir wollen A zeigen. Also nehmen wir einmal an, dass A stimmt. Dann müsste auch B gelten. B ist aber richtig, also muss auch A gelten.

(In der Bildschirmversion erscheint der folgende Text in blau. Diese Farbe wird für Beispiele verwendet.)

Als Beispiel: Wir wollen $0 = 1$ zeigen. Dazu formen wir um: Aus $0 = 1$ folgt durch Verdoppeln $0 = 2$, dann durch Subtraktion von 1 auf beiden Seiten $-1 = 1$,

schließlich durch Quadrieren $1 = 1$, was offensichtlich stimmt. Also gilt auch die ursprüngliche Gleichung $0 = 1$.

Hier ist alles korrekt bis auf das „Also“ im letzten Satz, denn der Schluss von $A \Rightarrow B$ und B auf A ist nicht möglich.

Der Schluss von $A \Rightarrow B$ und A auf B ist hingegen sehr wohl möglich und stellt eine der grundlegenden Schlussweisen in Beweisen dar. Häufig ist „$A \Rightarrow B$“ ein mathematischer Satz, der angewendet werden soll. Wir weisen nach, dass die Voraussetzung A gilt, und können dann auf B schließen. Die Korrektheit dieses Schlusses drückt sich darin aus, dass die Aussage

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

stets wahr ist. So eine Aussage heißt auch eine *Tautologie*. In den Tautologien

DEF Tautologie

$$(A \wedge B) \Rightarrow A\,, \qquad A \Rightarrow (A \vee B) \quad \text{und} \quad (A \Leftrightarrow B) \Leftrightarrow ((A \Rightarrow B) \wedge (B \Rightarrow A))$$

verbergen sich weitere Schlussregeln. Die letzte davon zeigt, dass man eine Äquivalenz $A \Leftrightarrow B$ dadurch beweisen kann, dass man die beiden Implikationen $A \Rightarrow B$ und $B \Rightarrow A$ nachweist. Das wird uns häufig begegnen.

Wie zeigt man, dass eine Verknüpfung von Aussagen eine Tautologie ist? Das kann man mit Hilfe von Wahrheitstafeln tun, indem man alle möglichen Kombinationen von Wahrheitswerten der beteiligten Grundaussagen ausprobiert. Zum Beispiel:

A	B	$A \Rightarrow B$	$(A \Rightarrow B) \wedge A$	$((A \Rightarrow B) \wedge A) \Rightarrow B$
$*$	W	W	$*$	W
W	F	F	F	W
F	F	W	F	W

Der Stern $*$ steht dabei für einen nicht festgelegten Wahrheitswert; wir nutzen aus, dass die Implikation $C \Rightarrow B$ immer wahr ist, wenn B wahr ist.

Weitere wichtige Schlussregeln kommen aus den Tautologien

$$\neg A \Leftrightarrow (A \Rightarrow \bot) \qquad \text{und} \qquad (A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)\,.$$

Die erste besagt, dass man die Negation von A dadurch beweisen kann, dass man die Annahme, dass A gilt, zum Widerspruch („$\bot$“) führt. Die zweite ist der klassische *Widerspruchsbeweis*: Um die Implikation $A \Rightarrow B$ zu zeigen, nehme ich A an und will B zeigen. Für den Widerspruchsbeweis nehme ich nun an, dass B *falsch* ist (also dass $\neg B$ gilt) und leite daraus den Widerspruch $\neg A$ zu A ab. Das zeigt, dass $\neg B$ unter der Annahme A nicht gelten kann, also muss B richtig sein. Die Implikation $\neg B \Rightarrow \neg A$ wird auch die *Kontraposition* der zu ihr äquivalenten Implikation $A \Rightarrow B$ genannt.

DEF Kontraposition

Hier sind ein paar weitere Tautologien, an denen Sie sich versuchen können. Sie zeigen, wie man eine Negation in andere Verknüpfungen „hineinziehen“ kann.

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

Die ersten beiden davon sind als *de Morgansche Regeln* bekannt.

Als ein weiteres Beispiel möchte ich Ihnen vorführen, dass

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

eine Tautologie ist.

A	B	C	$A \Rightarrow B$	$B \Rightarrow C$	$(A \Rightarrow B) \wedge (B \Rightarrow C)$	$A \Rightarrow C$	$(\ldots) \Rightarrow (A \Rightarrow C)$
F	*	*	W	*	*	W	W
*	*	W	*	W	*	W	W
W	W	F	W	F	F	F	W
W	F	F	F	W	F	F	W

Das entspricht einer Schlusskette: Wenn wir aus A folgern können, dass B gilt, und aus B, dass C gilt, dann ist es auch richtig, dass aus A die Richtigkeit von C folgt. Man kann also den Beweis von „$A \Rightarrow C$“ zerlegen in Beweise von „$A \Rightarrow B$“ und von „$B \Rightarrow C$“. Wenn man einen solchen Beweis aufschreibt, schreibt man dann auch einfach

$$A \Rightarrow B \Rightarrow C$$

oder

$$A \Rightarrow B_1 \Rightarrow B_2 \Rightarrow \ldots \Rightarrow B_n \Rightarrow C\,,$$

wenn es über mehrere Zwischenschritte geht.

Warnung. Die Notation „$A \Rightarrow B$“ kann zweierlei bedeuten:

- Die *Aussage* „aus A folgt B“, und
- den *Beweisschritt* „wir schließen von A auf B“, der die *als wahr bekannte* Aussage $A \Rightarrow B$ *verwendet*.

Eigentlich wäre es besser, dies auch in der Schreibweise zu unterscheiden, etwa indem man ein anderes Symbol (wie zum Beispiel $\curvearrowright$) für Beweisschritte verwendet. Allerdings sind beide Verwendungen von „$\Rightarrow$“ ziemlich üblich, und so werden wir hier auch beide benutzen.

Manche Teile dieses Skripts sind — wie hier — kleiner gedruckt. Dort wird Material behandelt, das über den eigentlichen Stoff der Vorlesung hinaus geht, aber vielleicht für den Einen oder die Andere von Ihnen interessant ist.

Hier geht es um die Frage, ob es schwierig ist, für eine gegebene aussagenlogische Formel zu entscheiden, ob sie eine Tautologie ist. Wir haben ja gesehen, dass man mit Hilfe einer Wahrheitstafel immer feststellen kann, ob eine Tautologie vorliegt oder nicht. Allerdings gibt es, wenn n verschiedene elementare Aussagen (wie A und B oben) beteiligt sind, 2^n mögliche Kombinationen von Wahrheitswerten, die überprüft werden müssen. Diese Zahl wächst sehr schnell mit n: $2^{100} = 1\,267\,650\,600\,228\,229\,401\,496\,703\,205\,376$ („exponentielles Wachstum“), so dass es praktisch unmöglich ist, alles durchzuprobieren. Auf der anderen Seite haben wir gesehen, dass man oft mehrere Möglichkeiten zusammenfassen kann, sodass man sich fragen kann, ob es auch eine einigermaßen effiziente (also mit vertretbarem Aufwand durchführbare) Methode gibt. Solche Fragen werden von der *Komplexitätstheorie* studiert, die im Bereich zwischen mathematischer Logik und theoretischer Informatik angesiedelt ist. Im vorliegenden Fall ist die Antwort „wahrscheinlich Nein“: Das eng verwandte Erfüllbarkeitsproblem ist *NP-vollständig* (eine Aussage ist genau dann nicht erfüllbar, wenn ihre Negation eine Tautologie ist), und für solche Probleme sind keine effizienten Lösungsverfahren (Algorithmen) *bekannt*. Die Frage danach, ob es tatsächlich keine *gibt*, ist der Inhalt des „P = NP?“-Problems, für dessen Lösung man eine Million Dollar bekommen würde.

2.B Mengen.

Ich setze voraus, dass Sie in der Schule gelernt haben, mit *Mengen* umzugehen. Daher werde ich mich auf eine kurze Wiederholung bzw. Einführung von Schreibweisen und grundlegenden Operationen und Rechenregeln beschränken.

Endliche Mengen können durch Aufzählung ihrer Elemente angegeben werden:

$$\{1, 2, 4, 8\}\,, \qquad \{\{1\}, \{1, 2\}\}\,.$$

DEF $\emptyset$, $x \in M$

Beachte: die Elemente einer Menge können selbst wieder Mengen sein. Sehr wichtig ist die *leere Menge*, die $\{\}$ geschrieben werden kann. Es ist aber die Schreibweise $\emptyset$ allgemein gebräuchlich; wir werden uns ebenfalls daran halten. Wir schreiben „$x \in M$“ für die Aussage „x ist Element der Menge M“ und „$x \notin M$“ für ihre Negation. Zum Beispiel ist $x \in \emptyset$ stets falsch, da die leere Menge keine Elemente hat. Zwei Mengen sind *gleich*, wenn sie dieselben Elemente haben. Insbesondere kommt es nicht darauf an, wie oft man ein Element aufführt:

DEF $M = N$

$$\{1, 1, 2, 2, 2, 3\} = \{1, 2, 3\}\,.$$

Man kann Mengen auch durch Angabe der Eigenschaften beschreiben, die ihre Elemente haben, wie zum Beispiel

$$\{n \mid n \text{ ist Primzahl}\}\,.$$

(Statt des senkrechten Strichs „|“ ist auch ein Doppelpunkt „:“ gebräuchlich.) Es gibt Symbole für gewisse häufig benötigte Mengen, wie

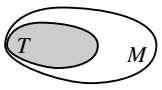
DEF $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$

- die Menge $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ der *natürlichen Zahlen*,
- die Menge $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ der *ganzen Zahlen*,
- die Menge $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}$ der *rationalen Zahlen* und
- die Menge $\mathbb{R}$ der *reellen Zahlen*.

Warnung. Die Definition von $\mathbb{N}$ ist in der Literatur nicht einheitlich; sehr häufig wird auch $\mathbb{N} = \{1, 2, 3, \ldots\}$ (also ohne die Null) gesetzt. Hier gibt es kein Richtig oder Falsch; letzten Endes ist das eine Geschmacksfrage. Für mich sind die natürlichen Zahlen gerade die Mächtigkeiten (die *Mächtigkeit* einer Menge ist die Anzahl ihrer Elemente) von endlichen Mengen, und da die leere Menge endlich ist, sollte auch die Null eine natürliche Zahl sein.

2.1. **Definition.** Eine Menge T heißt *Teilmenge* der Menge M, kurz $T \subset M$, wenn jedes Element von T auch ein Element von M ist. Man beachte, dass der Fall $T = M$ hier erlaubt ist. Um auszudrücken, dass T eine *echte* Teilmenge von M ist (also Teilmenge von M, aber nicht ganz M), schreiben wir $T \subsetneq M$. Statt $M \subset N$ kann man auch $N \supset M$ schreiben. Die Teilmengenbeziehung heißt auch *Inklusion*. ♢

DEF Teilmenge



Warnung. Die Schreibweise wird in der Literatur nicht einheitlich verwendet; oft findet man $T \subset M$ für echte Teilmengen und $T \subseteq M$ für beliebige Teilmengen. Machen Sie sich solche Unterschiede bewusst, wenn Sie Lehrbücher benutzen!

Einfache Beispiele von Teilmengen sind die leere Menge, die Teilmenge *jeder* Menge ist: $\emptyset \subset M$, und natürlich ist jede Menge Teilmenge von sich selbst: $M \subset M$. Für die oben eingeführten Zahlenmengen haben wir die Beziehungen $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

2.2. **Definition.** Die Menge aller Teilmengen von M heißt die *Potenzmenge* von M; wir schreiben

$$\mathcal{P}(M) = \{T \mid T \subset M\}$$

dafür. ♢

DEF
Potenzmenge
$\mathcal{P}(M)$

2.3. **Beispiel.** Zum Beispiel gilt

$$\mathcal{P}(\emptyset) = \{\emptyset\}\,, \quad \mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\} \quad \text{und} \quad \mathcal{P}(\{1,2\}) = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}\,. \quad ♣$$

BSP
Potenzmenge

An dieser Stelle gleich noch ein wichtiger Hinweis:

- Man muss sorgfältig zwischen Mengen und ihren Elementen unterscheiden. Zum Beispiel haben „$a \in M$“ und „$a \subset M$“ völlig verschiedene Bedeutungen.
- Besonders schwer fällt die Unterscheidung zwischen dem Element a und der Einermenge $\{a\}$. Es ist sehr wichtig, sich diese Unterschiede gleich zu Beginn klar zu machen!

Auf der anderen Seite ist $a \in M$ äquivalent zu $\{a\} \subset M$ — wenn man also beides falsch macht, wird es wieder richtig.

Mengen können miteinander verknüpft werden:

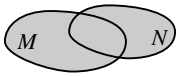
2.4. **Definition.** Seien M und N zwei Mengen.

DEF

(1) Die *Vereinigung* von M und N ist

$$M \cup N = \{x \mid x \in M \vee x \in N\}\,.$$

$M \cup N$

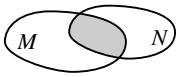


(2) Der *Durchschnitt* oder die *Schnittmenge* von M und N ist

$$M \cap N = \{x \mid x \in M \wedge x \in N\}\,.$$

Zwei Mengen, deren Durchschnitt leer ist ($M \cap N = \emptyset$), heißen *disjunkt*.

$M \cap N$

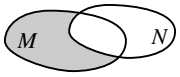


(3) Die *Mengendifferenz* von M und N ist

$$M \setminus N = \{x \mid x \in M \wedge x \notin N\}\,.$$

Sie besteht aus den Elementen von M, die keine Elemente von N sind. ♢

$M \setminus N$



Für diese Verknüpfungen gelten gewisse Rechenregeln und es gibt Beziehungen zum Begriff der Teilmenge. Ich werde hier einige davon angeben und auch beweisen; andere können Sie sich selbst herleiten, was eine gute erste Übung darstellt, einfache Beweise zu führen. Einige solche Aufgaben finden Sie auf dem Übungsblatt. (In der Bildschirmversion des Skripts erscheint der folgende Text in rot. Diese Farbe kennzeichnet mathematische Sätze oder Lemmata (Hilfssätze).)

2.5. **Satz.**

SATZ
Eigensch.
Mengen

(1) *Es gilt für alle Mengen M und N, dass M genau dann eine Teilmenge von N ist, wenn die Mengen $M \cup N$ und N übereinstimmen:*

$$M \subset N \iff M \cup N = N\,.$$

(2) *Für je zwei Mengen X und Y gilt das „Absorptionsgesetz“*

$$(X \cap Y) \cup Y = Y\,.$$

(3) *Für alle Mengen A, B, C gilt das „Distributivgesetz“*

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)\,.$$

Es gelten auch die analogen folgenden Aussagen, die wir hier aber nicht beweisen werden:

(1) Für alle Mengen M und N gilt $\quad M \subset N \iff M \cap N = M$.

(2) Für je zwei Mengen X und Y gilt $\quad (X \cup Y) \cap Y = Y$.

(3) Für alle Mengen A, B, C gilt $\quad (A \cup B) \cap C = (A \cap C) \cup (B \cap C)$.

(In der Bildschirmversion des Skripts erscheint der folgende Text in dunkelrot. Diese Farbe kennzeichnet Beweise.)

Beweis.

(1) Zu zeigen ist eine Äquivalenz $A \Leftrightarrow B$. In den meisten Fällen ist es am besten, den Beweis in zwei Teile, nämlich den Beweis von $A \Rightarrow B$ und den Beweis von $B \Rightarrow A$, zu zerlegen.

„$\Rightarrow$“: (D.h., wir beweisen die Richtung „von links nach rechts“, also die Aussage $M \subset N \Rightarrow M \cup N = N$.) Wir setzen voraus, dass M eine Teilmenge von N ist, und wir müssen zeigen, dass $M \cup N = N$ ist. Dies ist wiederum eine Äquivalenz, nämlich die Aussage $x \in M \cup N \Leftrightarrow x \in N$. Wir zerlegen den Beweis wieder in zwei Schritte:

„$M \cup N \subset N$“: Sei $x \in M \cup N$. (Das ist die übliche Formulierung dafür, dass man annimmt, dass $x \in M \cup N$ richtig ist.) Das bedeutet nach Definition $x \in M$ oder $x \in N$. Im ersten Fall ($x \in M$) folgt $x \in N$, da nach Voraussetzung M Teilmenge von N ist. Im zweiten Fall gilt $x \in N$ bereits.

„$N \subset M \cup N$“: Das gilt immer, denn aus $x \in N$ folgt $x \in M \vee x \in N$.

Damit ist die Gleichheit $M \cup N = N$ gezeigt und der Beweis der einen Richtung beendet.

„$\Leftarrow$“: (Jetzt beweisen wir die Richtung „von rechts nach links“, also die Aussage $M \cup N = N \Rightarrow M \subset N$.) Es gelte $M \cup N = N$. Zu zeigen ist $M \subset N$, also die Implikation $x \in M \Rightarrow x \in N$. Sei also $x \in M$. Dann ist auch $x \in M \cup N$, aber $M \cup N = N$, also folgt $x \in N$. Damit ist gezeigt, dass M eine Teilmenge von N ist.

(2) Es ist eine Gleichheit von Mengen zu beweisen. Wir haben eben schon gesehen, dass das eine Äquivalenz $a \in (X \cap Y) \cup Y \Leftrightarrow a \in Y$ ist, die man in der Regel am besten in zwei Implikationen aufspaltet. In diesem Fall verwenden wir häufig folgende Kurzschreibweise:

„$\subset$“: (Wir beweisen die Inklusion $(X \cap Y) \cup Y \subset Y$.) Sei $a \in (X \cap Y) \cup Y$. Das bedeutet $a \in X \cap Y$ oder $a \in Y$. Der zweite Fall ($a \in Y$) ist gerade, was wir zeigen wollen. Im ersten Fall gilt $a \in X$ und $a \in Y$, also insbesondere (ein Wort, das Mathematiker gerne verwenden) wieder $a \in Y$.

„$\supset$“: (Wir beweisen die Inklusion $(X \cap Y) \cup Y \supset Y$, also $Y \subset (X \cap Y) \cup Y$.) Sei $a \in Y$. Dann gilt die schwächere Aussage „$a \in X \cap Y$ oder $a \in Y$“, und das bedeutet gerade $a \in (X \cap Y) \cup Y$.

Alternativ kann man auch argumentieren, dass Aussage (2) aus Aussage (1) folgt, wenn man dort $M = X \cap Y$ und $N = Y$ setzt, denn die Inklusion $X \cap Y \subset Y$, die dann auf der linken Seite der Äquivalenz steht, ist immer richtig.

(3) Wie eben teilen wir den Beweis der Gleichheit zweier Mengen auf in die Teile „$\subset$“ und „$\supset$“.

„$\subset$“: Sei $s \in (A \cap B) \cup C$. Das bedeutet $s \in A \cap B$ oder $s \in C$. Im ersten Fall gilt $s \in A$ und $s \in B$, daraus folgt $s \in A \cup C$ und $s \in B \cup C$ und damit $s \in (A \cup C) \cap (B \cup C)$. Im zweiten Fall (also $s \in C$) gilt ebenfalls $s \in A \cup C$ und $s \in B \cup C$, also folgt auch in diesem Fall $s \in (A \cup C) \cap (B \cup C)$. Damit ist die Inklusion $(A \cap B) \cup C \subset (A \cup C) \cap (B \cup C)$ gezeigt.

„$\supset$“: Sei $s \in (A \cup C) \cap (B \cup C)$. Das bedeutet $s \in A \cup C$ und $s \in B \cup C$. Die erste dieser Aussagen heißt $s \in A$ oder $s \in C$. Wenn $s \in C$ ist, dann ist auch $s \in (A \cap B) \cup C$. Wenn $s \notin C$ ist, dann muss $s \in A$ und $s \in B$ sein, also ist $s \in A \cap B$ und damit auch $s \in (A \cap B) \cup C$. ❑

Beachten Sie, dass wir in einigen der obigen Beweise eine *Fallunterscheidung* benutzt haben. Dahinter stecken die Tautologien

$$((A \Rightarrow C) \wedge (B \Rightarrow C)) \Rightarrow ((A \vee B) \Rightarrow C) \ ,$$

die man benutzt, um eine Implikation der Form $(A \vee B) \Rightarrow C$ zu zeigen, und

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

die die klassische Fallunterscheidung darstellt: Wenn ich B sowohl unter der Annahme, dass A gilt, als auch unter der Annahme, dass A nicht gilt, zeigen kann, dann muss B richtig sein.

Da die Aussagen, die wir hier bewiesen haben, durch aussagenlogische Verknüpfungen aus endlich vielen „Elementaraussagen“ der Form $x \in M$ zusammengesetzt sind, könnten wir sie auch durch Aufstellen einer Wahrheitstafel beweisen. Zweck der Übung sollte aber sein, zu einer gewissen Fingerfertigkeit im logischen Schließen zu kommen, denn später wird es meistens nicht mehr möglich sein, Beweise rein aussagenlogisch zu führen.

2.C Prädikatenlogik.

Nun hat man es in der Mathematik nicht nur mit einfachen Aussagen zu tun, die man irgendwie verknüpft, sondern in aller Regel hängen die Aussagen noch von gewissen Parametern oder Variablen ab. Ein typisches Beispiel ist die Aussage „$x \in M$“, deren Wahrheitswert davon abhängt, wofür x und M stehen. (Man nennt solche parameterabhängigen Aussagen manchmal *Aussageformen*, weil sie erst dadurch zu einer Aussage mit festgelegtem Wahrheitswert werden, dass man den Parametern Werte zuweist. Auch der Begriff *Prädikat* ist gebräuchlich, was die Bezeichnung „Prädikatenlogik“ erklärt.) Um aus solchen von Variablen abhängigen Aussagen wiederum Aussagen zu machen, die nicht mehr von (einigen oder allen) Variablen abhängen, gibt es im Wesentlichen zwei Möglichkeiten. Sei dafür $A(x)$ eine (möglicherweise) von der Variablen x abhängige Aussage.

DEF Quantoren $\forall$, $\exists$

- Wir machen die Aussage „für alle x gilt $A(x)$“ oder kurz „$\forall x \colon A(x)$“.
- Wir machen die Aussage „es gibt ein x, sodass $A(x)$ gilt“ oder kurz „$\exists x \colon A(x)$“.

Im Fachjargon spricht man von *Quantifizierung*, da man eine Aussage darüber macht, für wie viele x (alle oder wenigstens eines) $A(x)$ stimmt. In diesem Zusammenhang heißen die Symbole $\forall$ und $\exists$ auch *Quantoren*, genauer *Allquantor* ($\forall$) und *Existenzquantor* ($\exists$).

In der Praxis kommen fast nur die Kombinationen

$$\forall x \colon x \in M \Rightarrow A(x) \qquad \text{und} \qquad \exists x \colon x \in M \wedge A(x)$$

vor, die man dann zu

$$\forall x \in M \colon A(x) \qquad \text{„für alle } x \in M \text{ gilt } A(x)\text{“}$$

und

$$\exists x \in M\colon A(x) \qquad \text{„es gibt ein } x \in M \text{ mit } A(x)\text{“}$$

abkürzt. An der ausführlicheren Form oben erkennt man, dass

$$\forall x \in \emptyset\colon A(x)$$

immer wahr ist, denn die Voraussetzung $x \in \emptyset$ in der Implikation „$x \in \emptyset \Rightarrow A(x)$“ ist falsch. Entsprechend ist

$$\exists x \in \emptyset\colon A(x)$$

immer falsch, denn es gibt ja kein Element der leeren Menge, also erst recht keines mit zusätzlichen Eigenschaften.

Für den Umgang mit den Quantoren sind folgende Regeln wichtig:

$$\neg\forall x \in M\colon A(x) \qquad \text{ist gleichbedeutend mit} \qquad \exists x \in M\colon \neg A(x)$$

und

$$\neg\exists x \in M\colon A(x) \qquad \text{ist gleichbedeutend mit} \qquad \forall x \in M\colon \neg A(x)\,.$$

Die erste zeigt, wie man eine „Allaussage“ widerlegt: *Man gibt ein Gegenbeispiel an.* Das macht auch verständlich, warum $\forall x \in \emptyset\colon A(x)$ wahr sein muss: Es gibt kein Gegenbeispiel! Das klingt jetzt vielleicht wie esoterische Spielerei, das ist es aber keineswegs: *Es ist sehr wichtig, Grenzfälle zu verstehen.* Die leere Menge ist ein typischer Grenzfall in vielen Situationen, und nur wenn Sie diesen Grenzfall verstehen, haben Sie die Situation wirklich verstanden! Zum Beispiel gilt die auf den ersten Blick offensichtliche Implikation

$$\big(\forall x \in M\colon A(x)\big) \Rightarrow \big(\exists x \in M\colon A(x)\big)$$

(„wenn alle Elemente von M die Eigenschaft A haben, dann hat wenigstens eines diese Eigenschaft“) *nur dann, wenn M nicht die leere Menge ist.* Das zu wissen kann einem manche Falltüren im Beweisgeschäft ersparen.

Eine wichtige Regel ist, dass es auf den Namen der Variablen nicht ankommt: Wenn y in $A(x)$ nicht vorkommt, dann sind

$$\forall x \in M\colon A(x) \qquad \text{und} \qquad \forall y \in M\colon A(y)$$

äquivalent („gebundene Umbenennung“); analog für Existenzaussagen. Sie kennen das vielleicht in ähnlicher Form von Integralen:

$$\int_a^b f(x)\,dx = \int_a^b f(y)\,dy\,.$$

Sehr wichtig ist auch, dass es auf die *Reihenfolge* der Quantoren ankommt: Die Aussagen

$$\forall x \in M\ \exists y \in N\colon A(x,y) \qquad \text{und} \qquad \exists y \in N\ \forall x \in M\colon A(x,y)$$

haben unterschiedliche Bedeutung — in der ersten Aussage kann das y, dessen Existenz behauptet wird, von x abhängen, in der zweiten Aussage gibt es *ein festes* y, das für alle x funktionieren muss. Deswegen sollte man sich auch angewöhnen, alle Quantoren systematisch *vor* die Aussage zu stellen, auf die sie sich beziehen. Etwas in der Art von

$$\exists a \in M\colon A(a,x), \forall x \in N$$

ist unbedingt zu vermeiden! Das gilt entsprechend auch für die Formulierung von Aussagen in Textform.

2.6. **Beispiel.** **BSP** ∀∃ und ∃∀

$$\forall x \in \mathbb{N}\ \exists y \in \mathbb{N}\colon y > x$$

(„für jede natürliche Zahl x gibt es eine größere natürliche Zahl y“) ist sicher richtig, während

$$\exists y \in \mathbb{N}\ \forall x \in \mathbb{N}\colon y > x$$

(„es gibt eine natürliche Zahl y, die größer ist als alle natürlichen Zahlen x“) falsch ist. Die Variante

$$\exists x \in \mathbb{N}\ \forall y \in \mathbb{N}\colon y > x$$

(„es gibt eine natürliche Zahl x, die kleiner ist als alle natürlichen Zahlen y“) ist ebenfalls falsch, aber sozusagen nur knapp, denn mit „$\geq$“ statt „$>$“ wäre sie richtig (mit $x = 0$). ♣

Auf der anderen Seite sind

$$\forall x \in M\ \forall y \in N\colon A(x,y) \qquad \text{und} \qquad \forall y \in N\ \forall x \in M\colon A(x,y)$$

äquivalent; deswegen kürzt man das auch gerne in der Form

$$\forall x \in M, y \in N\colon A(x,y)$$

oder, falls $M = N$,

$$\forall x, y \in M\colon A(x,y)$$

ab. Das gilt dann analog auch für zwei (oder mehr) Existenzquantoren.

Als Illustration dafür, wie man Beweise von quantifizierten Aussagen führt, zeige ich jetzt, dass es unendlich viele gerade natürliche Zahlen gibt. Dabei lernen Sie auch gleich eine mathematische „Redewendung“ kennen, mit der man ausdrücken kann, dass es unendlich viele natürliche Zahlen mit einer gewissen Eigenschaft gibt: Man sagt, dass es zu jeder gegebenen natürlichen Zahl eine größere gibt, die die Eigenschaft hat.

Behauptung: $\forall m \in \mathbb{N}\ \exists n \in \mathbb{N}\colon n > m$ und n ist gerade.

Beweis. Sei $m \in \mathbb{N}$ beliebig. Dann gilt für $n_0 = 2m + 2 \in \mathbb{N}$, dass $n_0 > m$ ist (denn $n_0 = m + (m + 2)$ und $m + 2 > 0$), und $n_0 = 2(m + 1)$ ist gerade. Also existiert ein $n \in \mathbb{N}$ (nämlich zum Beispiel n_0) mit $n > m$ und n gerade: $\exists n \in \mathbb{N}\colon n > m$ und n ist gerade. Da $m \in \mathbb{N}$ beliebig war, gilt diese Aussage für alle $m \in \mathbb{N}$. ❑

Man beweist also eine Allaussage $\forall x \in M\colon A(x)$, indem man ein nicht näher spezifiziertes $x \in M$ betrachtet („Sei $x \in M$“ — das Wort „beliebig“ steht oben nur zur Verdeutlichung und kann weggelassen werden) und für dieses x die Aussage $A(x)$ zeigt. Eine Existenzaussage $\exists x \in M\colon A(x)$ kann man zeigen, indem man für ein *bestimmtes* $x_0 \in M$ die Aussage $A(x_0)$ beweist.

Das nennt man dann auch einen *konstruktiven* Existenzbeweis, weil man ein geeignetes Element explizit angibt oder konstruiert. Alternativ kann man die äquivalente Aussage $\neg\forall x \in M\colon \neg A(x)$ beweisen, indem man die Annahme, dass kein $x \in M$ die Eigenschaft A hat, zum Widerspruch führt. Dabei muss kein Element von M angegeben werden, das tatsächlich die Eigenschaft A hat. Zum besseren Verständnis hier ein Beispiel: Sei $N > 1$ eine gegebene natürliche Zahl. Sie wollen beweisen, dass N einen echten Teiler $d > 1$ hat, also die Aussage

$$\exists d, m \in \mathbb{N}\colon d > 1 \ \wedge\ m > 1 \ \wedge\ N = d \cdot m\,.$$

Das kann man natürlich tun, indem man einen echten Teiler d findet. Man kann aber auch versuchen, die Negation der Aussage, nämlich „N ist Primzahl“ zum Widerspruch zu führen. Dazu kann man etwa den „kleinen Satz von Fermat“ verwenden, der aussagt,

dass für jede Primzahl p und jede ganze Zahl a die Zahl $a^p - a$ durch p teilbar ist. Wenn Sie also eine ganze Zahl a finden, sodass $a^N - a$ *nicht* durch N teilbar ist, dann folgt daraus, dass N keine Primzahl ist, also muss N einen echten Teiler haben, ohne dass Sie einen angeben können. Dieser Unterschied ist durchaus auch praktisch relevant. Es gibt nämlich effiziente Algorithmen, die feststellen, ob N eine Primzahl ist oder nicht, aber es sind bisher keine effizienten Algorithmen bekannt, die eine zusammengesetzte Zahl faktorisieren können.

Weitere Beispiele für Beweise werden in großer Zahl im Lauf der Vorlesung folgen.

An den bisherigen Beispielen von Beweisen können Sie jedenfalls schon sehen, dass *die Struktur der zu beweisenden Aussage die Struktur des Beweises vorgibt:* Bis zu einem gewissen Grad gibt es auch für das Beweisen Rezepte, die man anwenden kann!

2.D Geordnete Paare und kartesische Produkte.

Häufig möchte man mit zwei (oder vielleicht auch mehr, siehe unten) Elementen von verschiedenen Mengen gemeinsam arbeiten, wobei es auf die Reihenfolge ankommt. (Wenn die Reihenfolge keine Rolle spielt, also bei *ungeordneten* Paaren, kann man Zweiermengen $\{a, b\}$ verwenden.) Dazu führt man geordnete Paare ein:

DEF geordnetes Paar (a, b)

2.7. **Definition.** Sind a und b Elemente irgendwelcher Mengen, dann steht (a, b) für das daraus gebildete *geordnete Paar*. Die wesentliche Eigenschaft dieser geordneten Paare ist, dass zwei solche Paare genau dann gleich sind, wenn sie in beiden Komponenten übereinstimmen:

$$(a, b) = (x, y) \iff \big(a = x \text{ und } b = y\big)\,. \qquad \diamondsuit$$

Man kann geordnete Paare innerhalb der Mengenlehre definieren, indem man

$$(a, b) = \{\{a\}, \{a, b\}\}$$

setzt. (Man beachte den Sonderfall $(a, a) = \{\{a\}\}$.) Man muss dann zeigen, dass die so definierten Paare die obige Eigenschaft haben. Das sollten Sie als Aufforderung begreifen!

DEF $M \times N$

2.8. **Definition.** Sind M und N zwei Mengen, dann schreibt man

$$M \times N = \{(m, n) \mid m \in M, n \in N\}$$

(lies „M kreuz N“) für die Menge der geordneten Paare, deren erste Komponente aus M und deren zweite Komponente aus N kommt; die Menge $M \times N$ heißt das *kartesische Produkt* der Mengen M und N. $\diamondsuit$

R. Descartes (1596–1650)

(„Kartesisch“ leitet sich vom latinisierten Namen *Cartesius* des Mathematikers und Philosophen („cogito ergo sum“: „ich denke, also bin ich“) René Descartes ab.)

DEF Tripel, ... n-Tupel $M_1 \times \cdots \times M_n$ M^n

2.9. **Definition.** Analog zu Paaren kann man (geordnete) *Tripel* (a, b, c), *Quadrupel* (a, b, c, d), *Quintupel* (a, b, c, d, e), *Sextupel* (a, b, c, d, e, f) und ganz allgemein *n-Tupel* $(a_1, a_2, \ldots, a_n)$ einführen und kartesische Produkte mit mehr als zwei Faktoren definieren, zum Beispiel

$$A \times B \times C \times D = \{(a, b, c, d) \mid a \in A, b \in B, c \in C, d \in D\}\,.$$

Die Gleichheit zweier n-Tupel $(a_1, a_2, \ldots, a_n)$ und $(b_1, b_2, \ldots, b_n)$ definiert man analog zur Gleichheit von Paaren:

$$(a_1, a_2, \ldots, a_n) = (b_1, b_2, \ldots, b_n) \iff a_1 = b_1 \wedge a_2 = b_2 \wedge \ldots \wedge a_n = b_n\,.$$

Einen wichtigen Spezialfall des allgemeinen kartesischen Produkts erhalten wir, wenn alle beteiligten Mengen übereinstimmen. Dann schreibt man kurz M^2 für $M \times M$, M^3 für $M \times M \times M$ und allgemein

$$M^n = \{(m_1, m_2, \ldots, m_n) \mid m_1, m_2, \ldots, m_n \in M\}$$

für die Menge der n-Tupel, deren Komponenten aus der Menge M kommen. Für $n = 1$ haben wir "1-Tupel"; wir setzen $(a) = a$, damit ist dann $M^1 = M$. Für $n = 0$ gibt es das (eindeutig bestimmte) Nulltupel (), und $M^0 = \{()\}$ ist eine einelementige Menge. ♢

2.10. **Beispiele.** Zum Beispiel ist $\mathbb{R}^2$ die Menge der Paare reeller Zahlen. Wenn man die Komponenten als x- und y-Koordinate interpretiert, dann kann man $\mathbb{R}^2$ als die Menge der Punkte der Ebene auffassen, und entsprechend $\mathbb{R}^3$ als die Menge der Punkte des (dreidimensionalen) Raumes. Diese Mengen und ihre allgemeinere Form $\mathbb{R}^n$ werden uns bald wieder als Standardbeispiele von „Vektorräumen" begegnen. ♣

BSP
kartesische Produkte

2.E Abbildungen.

Der (vorläufig) letzte wichtige Begriff, den wir einführen müssen, ist der der Abbildung zwischen zwei Mengen.

2.11. **Definition.** Seien M und N zwei Mengen. Dann ist eine *Abbildung f von M nach N* eine Vorschrift, die jedem $x \in M$ ein eindeutig bestimmtes $y \in N$ zuordnet; für dieses y schreiben wir dann $f(x)$ („f von x"). Wir schreiben

$$f\colon M \longrightarrow N$$

oder, wenn wir die Abbildungsvorschrift angeben wollen,

$$f\colon M \longrightarrow N, \quad x \longmapsto f(x)\,,$$

wobei statt „$f(x)$" meistens eine konkrete Formel oder Ähnliches steht. Beachten Sie die beiden unterschiedlichen Pfeile „$\to$" und „$\mapsto$"! Der erste steht zwischen den *Mengen* M und N, der zweite zwischen den *Elementen* $x \in M$ und $f(x) \in N$. $f(x) \in N$ heißt dann das *Bild* von $x \in M$ unter f. Gilt $f(x) = y$ für ein $y \in N$, dann heißt x ein *Urbild* von y unter f. Man beachte: Es ist durchaus möglich, dass ein $y \in N$ kein Urbild oder viele verschiedene Urbilder unter f hat.

DEF
Abbildung

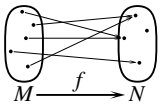


Bild
Urbild

Häufig (vor allem in der Analysis) verwendet man auch das Wort *Funktion* für *Abbildung* (was die häufig verwendete Bezeichnung „f" für Abbildungen erklärt). ♢

Mit „Vorschrift" ist hier nicht gemeint, dass das Bild von x unter f durch einen Rechenausdruck oder so etwas gegeben sein muss. Es kommt *nur* darauf an, dass *jedem $x \in M$ genau ein $f(x) \in N$* zugeordnet ist. Man kann sich f als eine „Black Box" vorstellen, die einem, wenn man ein $x \in M$ hineinsteckt, ein Element $f(x) \in N$ herausgibt (und zwar für dasselbe x immer dasselbe $f(x)$):

$$M \ni x \longrightarrow \boxed{f} \longrightarrow f(x) \in N$$

(Die Pfeile hier dienen nur der Illustration und sind nicht Teil der oben eingeführten Notation $f\colon M \to N$, $x \mapsto f(x)$.)

2.12. **Definition.** M heißt die *Definitionsmenge*, der *Definitionsbereich* oder die *Quelle* der Abbildung $f\colon M \to N$, N dementsprechend der *Wertebereich* oder das *Ziel* von f. Wichtig ist dabei, dass zur Angabe einer Abbildung immer auch Quelle und Ziel gehören; die Abbildungsvorschrift alleine genügt nicht.

DEF Definitions-, Wertebereich

Zwei Abbildungen f und g sind genau dann gleich (und man schreibt $f = g$), wenn ihre Definitions- und Wertebereiche übereinstimmen und für alle Elemente x des Definitionsbereichs gilt $f(x) = g(x)$: Abbildungen sind (bei gegebenem Definitions- und Wertebereich) durch ihre Werte festgelegt.

$f = g$

Wir schreiben $\mathrm{Abb}(M, N)$ für die Menge aller Abbildungen mit Definitionsbereich M und Wertebereich N. ♢

$\mathrm{Abb}(M, N)$

Der Wertebereich N von f ist von der *Wertemenge* oder *Bildmenge*

$$\{f(x) \mid x \in M\} \subset N$$

von f zu unterscheiden. Die Bildmenge kann eine echte Teilmenge des Wertebereichs sein.

DEF Bildmenge

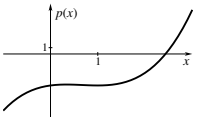
2.13. **Beispiele.** Beispiele von Abbildungen sind

BSP Abbildungen

$$n\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto 0$$

(die Nullfunktion; es gilt $n(x) = 0 \in \mathbb{R}$ für alle $x \in \mathbb{R}$),

$$p\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad x \longmapsto x^3 - 2x^2 + x - 5$$



(eine Polynomfunktion; es gilt zum Beispiel $p(1) = p(0) = -5$),

$$s\colon \mathbb{R} \longrightarrow \{-1, 0, 1\}, \quad x \longmapsto \begin{cases} 1 & \text{falls } x > 0, \\ 0 & \text{falls } x = 0, \\ -1 & \text{falls } x < 0 \end{cases}$$

(die Vorzeichenfunktion). Für eine beliebige Menge M gibt es die „Einermengenabbildung“

$$e\colon M \longrightarrow \mathcal{P}(M), \quad x \longmapsto \{x\}\,.$$

Zum kartesischen Produkt $M \times N$ gehören die *Projektionsabbildungen*

DEF Projektion

$$\mathrm{pr}_1\colon M \times N \longrightarrow M, \quad (a, b) \longmapsto a \qquad \text{und} \qquad \mathrm{pr}_2\colon M \times N \to N, \quad (a, b) \longmapsto b\,.$$

Ist T eine Teilmenge von M, dann hat man die *Inklusionsabbildung*

DEF Inklusionsabb.

$$i\colon T \longrightarrow M, \quad x \longmapsto x\,.$$

Für jede Menge X gibt es (als Spezialfall der Inklusionsabbildung) die *identische Abbildung* oder kurz *Identität*

DEF Identität id_X

$$\mathrm{id}_X\colon X \longrightarrow X, \quad x \longmapsto x\,,$$

die jedes Element von X auf sich selbst abbildet. Als Grenzfälle haben wir für jede Menge X genau eine Abbildung $\emptyset \to X$; eine Abbildung $X \to \emptyset$ gibt es jedoch nur dann, wenn X selbst die leere Menge ist, denn wenn X ein Element x hat, könnte es auf kein Element abgebildet werden (denn die leere Menge hat keine Elemente). ♣

Wenn Sie den Begriff „Vorschrift“, den wir oben verwendet haben, zu schwammig finden, dann erfahren Sie hier, wie man den Abbildungsbegriff auf eine solide Grundlage stellen kann. Man greift dazu auf die Mengenlehre zurück und identifiziert eine Abbildung $f\colon M \to N$ mit ihrem *Graphen*

$$\Gamma(f) = \{(x, f(x)) \mid x \in M\} \subset M \times N\,.$$

(Das verallgemeinert die Funktionsgraphen von Funktionen $\mathbb{R} \to \mathbb{R}$, die Sie aus der Schule kennen.) Dann kann man sagen, dass eine Teilmenge $F \subset M \times N$ genau dann einer Abbildung $f\colon M \to N$ entspricht, wenn die Bedingungen

$$\forall x \in M\ \exists y \in N\colon (x, y) \in F$$

und

$$\forall x \in M\ \forall y_1, y_2 \in N\colon\ ((x, y_1) \in F \wedge (x, y_2) \in F) \Rightarrow y_1 = y_2$$

erfüllt sind. Die erste Bedingung drückt aus, dass *jedes* $x \in M$ auf ein Element von N abgebildet werden muss, und die zweite Bedingung sagt, dass es *höchstens ein* solches Element von N gibt.

Es gibt gewisse wichtige Eigenschaften, die eine Abbildung haben kann oder nicht.

2.14. **Definition.** Sei $f\colon M \to N$ eine Abbildung. **DEF** injektiv surjektiv bijektiv

(1) f heißt *injektiv* oder eine *Injektion*, wenn f keine zwei verschiedenen Elemente von M auf dasselbe Element von N abbildet:

$$\forall x_1, x_2 \in M\colon f(x_1) = f(x_2) \Rightarrow x_1 = x_2$$

(So weist man auch nach, dass f injektiv ist: Man nimmt an, zwei Elemente hätten dasselbe Bild unter f und zeigt dann, dass diese beiden Elemente gleich sein müssen.)

(2) f heißt *surjektiv* oder eine *Surjektion* (das „sur“ ist französisch für „auf“, daher ist die korrekte Aussprache „ßür“), wenn jedes Element von N als Bild unter f eines Elements von M auftritt:

$$\forall y \in N\ \exists x \in M\colon f(x) = y$$

(3) f heißt *bijektiv* oder eine *Bijektion*, wenn f sowohl injektiv als auch surjektiv ist. ♢

Man kann das auch so ausdrücken:

- f ist genau dann injektiv, wenn jedes Element von N *höchstens* ein Urbild unter f hat.
- f ist genau dann surjektiv, wenn jedes Element von N *mindestens* ein Urbild unter f hat.
- f ist genau dann bijektiv, wenn jedes Element von N *genau* ein Urbild unter f hat.

2.15. **Definition.** Sei $f\colon M \to N$ bijektiv. Wir definieren dann eine Abbildung $f^{-1}\colon N \to M$ dadurch, dass wir für $f^{-1}(y)$ das eindeutig bestimmte $x \in M$ mit $f(x) = y$ nehmen. Diese Abbildung f^{-1} heißt die *Umkehrabbildung* oder *inverse Abbildung* von f. Eine bijektive Abbildung $f\colon X \to X$ heißt auch eine *Permutation* von X. ♢ **DEF** Umkehrabb. Permutation

2.16. **Beispiele.** Wir schreiben $\mathbb{R}_{\geq 0}$ für die Menge $\{x \in \mathbb{R} \mid x \geq 0\}$ der nichtnegativen reellen Zahlen. Dann gilt: **BSP** injektiv surjektiv

(1) $f_1\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$, ist weder injektiv noch surjektiv, denn es gilt zum Beispiel $f_1(1) = f_1(-1) = 1$ und $-1 \in \mathbb{R}$ hat kein Urbild.

(2) $f_2\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$, $x \mapsto x^2$, ist injektiv, aber nicht surjektiv.

(3) $f_3\colon \mathbb{R} \to \mathbb{R}_{\geq 0}$, $x \mapsto x^2$, ist surjektiv, aber nicht injektiv.

(4) $f_4\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $x \mapsto x^2$, ist bijektiv.
Die Umkehrabbildung ist $f_4^{-1}\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $x \mapsto \sqrt{x}$.

Daran sieht man auch sehr schön, dass Definitions- und Wertebereich wesentlich für eine Abbildung sind. Weitere allgemeine Beispiele sind:

(5) Für jede Menge M ist die identische Abbildung id_M bijektiv.

(6) Für jede Menge M ist die „leere Abbildung“ $\emptyset \to M$ injektiv.

(7) Jede Abbildung $\{a\} \to M$ ist injektiv.

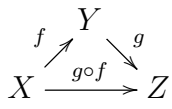
(8) Eine Abbildung $M \to \{a\}$ ist genau dann surjektiv, wenn M nicht leer ist.

(9) Die Einermengenabbildung $e\colon M \to \mathcal{P}(M)$ ist injektiv, aber nicht surjektiv (Letzteres, weil zum Beispiel die leere Menge kein Urbild hat). ♣

Abbildungen können verknüpft werden, indem man sie „hintereinanderschaltet“:

DEF Komposition

2.17. **Definition.** Sind $f\colon X \to Y$ und $g\colon Y \to Z$ Abbildungen, sodass der Wertebereich von f mit dem Definitionsbereich von g übereinstimmt, dann kann man die zusammengesetzte Abbildung $g \circ f\colon X \to Z$ bilden, die $x \in X$ auf $g(f(x)) \in Z$ abbildet:



$$x \longrightarrow \boxed{f} \longrightarrow f(x) \longrightarrow \boxed{g} \longrightarrow g(f(x)) \qquad \diamondsuit$$

Man muss sich merken, dass in $g \circ f$ die Abbildung f *zuerst* ausgeführt wird, obwohl sie hinter g steht. Die Sprechweise „g nach f“ für $g \circ f$ hilft dabei.

Diese Verknüpfung oder *Komposition* von Abbildungen hat einige wichtige Eigenschaften:

SATZ Eigensch. Abbildungen

2.18. **Satz.**

(1) *Sind $f\colon W \to X$, $g\colon X \to Y$ und $h\colon Y \to Z$ Abbildungen, dann gilt $(h \circ g) \circ f = h \circ (g \circ f)$. Man lässt deswegen meistens die Klammern weg und schreibt $h \circ g \circ f$.*

(2) *Ist $f\colon X \to Y$ eine Abbildung, dann gilt*

$$f \circ \mathrm{id}_X = f \qquad \text{und} \qquad \mathrm{id}_Y \circ f = f\,.$$

(3) *Sind $f\colon X \to Y$ und $g\colon Y \to Z$ injektive Abbildungen, dann ist auch $g \circ f\colon X \to Z$ injektiv.*

(4) *Sind $f\colon X \to Y$ und $g\colon Y \to Z$ surjektive Abbildungen, dann ist auch $g \circ f\colon X \to Z$ surjektiv.*

(5) *Ist $f\colon X \to Y$ bijektiv mit Umkehrabbildung $f^{-1}\colon Y \to X$, dann gilt*

$$f^{-1} \circ f = \mathrm{id}_X \qquad \text{und} \qquad f \circ f^{-1} = \mathrm{id}_Y\,.$$

(6) *Sind $f\colon X \to Y$ und $g\colon Y \to Z$ zwei Abbildungen, dann gilt*

$g \circ f$ injektiv $\Longrightarrow$ f injektiv und $g \circ f$ surjektiv $\Longrightarrow$ g surjektiv.

(7) *Ist $f\colon X \to Y$ eine Abbildung, dann ist f genau dann injektiv, wenn X leer ist oder es eine Abbildung $g\colon Y \to X$ mit $g \circ f = \mathrm{id}_X$ gibt.*

(8) *Ist $f\colon X \to Y$ eine Abbildung, dann ist f genau dann surjektiv, wenn es eine Abbildung $g\colon Y \to X$ gibt mit $f \circ g = \mathrm{id}_Y$.*

(9) *Ist $f\colon X \to Y$ eine Abbildung, dann ist f genau dann bijektiv, wenn es eine Abbildung $g\colon Y \to X$ gibt mit $g \circ f = \mathrm{id}_X$ und $f \circ g = \mathrm{id}_Y$.*

Man kann also auf (mindestens) zwei verschiedene Arten beweisen, dass eine Abbildung $f\colon X \to Y$ bijektiv ist:

- Man weist nach, dass f injektiv und surjektiv ist, oder
- man findet einen Kandidaten g für die Umkehrabbildung und rechnet nach, dass $g \circ f = \mathrm{id}_X$ und $f \circ g = \mathrm{id}_Y$ ist.

In vielen Fällen ist die zweite Methode einfacher durchzuführen.

Beweis.

(1) Erst einmal ist klar, dass die Abbildungen den gemeinsamen Definitionsbereich W und den gemeinsamen Wertebereich Z haben. Die Aussage „$(h \circ g) \circ f = h \circ (g \circ f)$“ bedeutet dann

$$\forall w \in W\colon \big((h \circ g) \circ f\big)(w) = \big(h \circ (g \circ f)\big)(w)\,.$$

Sei also $w \in W$. Dann ist

$$\big((h \circ g) \circ f\big)(w) = (h \circ g)(f(w)) = h\big(g(f(w))\big)$$

und ebenso

$$\big(h \circ (g \circ f)\big)(w) = h\big((g \circ f)(w)\big) = h\big(g(f(w))\big)\,,$$

also gilt die behauptete Gleichheit für w. Da $w \in W$ beliebig war, gilt die Gleichheit für alle $w \in W$.

$$w \longrightarrow \boxed{f} \longrightarrow f(w) \longrightarrow \boxed{g} \longrightarrow g(f(w)) \longrightarrow \boxed{h} \longrightarrow h(g(f(w)))$$

(2) In beiden Fällen haben alle beteiligten Abbildungen denselben Definitionsbereich X und denselben Wertebereich Y. Für $x \in X$ gilt

$$(f \circ \mathrm{id}_X)(x) = f(\mathrm{id}_X(x)) = f(x)\,,$$

also ist $f \circ \mathrm{id}_X = f$, und

$$(\mathrm{id}_Y \circ f)(x) = \mathrm{id}_Y(f(x)) = f(x)\,,$$

also ist auch $\mathrm{id}_Y \circ f = f$.

(3) Übung.

(4) Übung.

(5) Die Definitions- und Wertebereiche stimmen jeweils überein. Für $x \in X$ gilt $f^{-1}(f(x)) = x = \mathrm{id}_X(x)$ nach Definition der Umkehrabbildung, also ist $f^{-1} \circ f = \mathrm{id}_X$. Für $y \in Y$ gilt $f(f^{-1}(y)) = y = \mathrm{id}_Y(y)$ ebenfalls nach Definition der Umkehrabbildung, also ist $f \circ f^{-1} = \mathrm{id}_Y$.

(6) Wir nehmen an, dass $g \circ f$ injektiv ist; wir müssen zeigen, dass auch f injektiv ist. Seien dazu $x_1, x_2 \in X$ mit $f(x_1) = f(x_2)$. Dann folgt

$$(g \circ f)(x_1) = g(f(x_1)) = g(f(x_2)) = (g \circ f)(x_2)\,,$$

und weil $g \circ f$ injektiv ist, muss $x_1 = x_2$ sein. Damit ist gezeigt, dass f injektiv ist.
Jetzt nehmen wir an, dass $g \circ f$ surjektiv ist; wir müssen zeigen, dass auch g surjektiv ist. Sei dazu $z \in Z$. Da nach Voraussetzung $g \circ f$ surjektiv ist, gibt es $x \in X$ mit $(g \circ f)(x) = z$. Das heißt aber $g(f(x)) = z$, also gilt mit $y = f(x) \in Y$ auch $g(y) = z$. Das zeigt, dass g surjektiv ist.

(7) Zu zeigen ist die Äquivalenz

$$f\colon X \to Y \text{ injektiv} \iff (X = \emptyset \ \vee \ \exists g \in \mathrm{Abb}(Y, X)\colon g \circ f = \mathrm{id}_X)\,.$$

„⇒“: Wir nehmen an, f sei injektiv. Wenn X leer ist, dann gilt die rechte Seite. Wenn X nicht leer ist, dann sei $x_0 \in X$ irgendein Element. Wir konstruieren eine passende Abbildung $g\colon Y \to X$ wie folgt: Sei $y \in Y$. Wenn es ein $x \in X$ gibt mit $f(x) = y$, dann setzen wir $g(y) = x$. Da es (weil f injektiv ist) dann genau ein solches x gibt, ist $g(y)$ eindeutig bestimmt. Wenn es kein $x \in X$ gibt mit $f(x) = y$, dann setzen wir $g(y) = x_0$. Jetzt müssen wir nachprüfen, dass g die geforderte Eigenschaft $g \circ f = \mathrm{id}_X$ hat. Definitions- und Wertebereich beider Seiten stimmen überein, und für $x \in X$ gilt nach Definition von g, dass $(g \circ f)(x) = g(f(x)) = x = \mathrm{id}_X(x)$ ist. Damit ist die Gleichheit der Abbildungen gezeigt.
„⇐“: Wenn $X = \emptyset$ ist, dann ist f injektiv. Wenn es $g\colon Y \to X$ gibt mit $g \circ f = \mathrm{id}_X$, dann ist f ebenfalls injektiv nach Teil (6), denn id_X ist injektiv.

(8) „⇒“: Ist f surjektiv, dann können wir zu jedem $y \in Y$ ein $x_y \in X$ auswählen mit $f(x_y) = y$ (denn es gibt ja immer mindestens ein Urbild). Wir setzen dann $g(y) = x_y$ und es folgt $f \circ g = \mathrm{id}_Y$.
„⇐“: Das folgt aus Teil (6), denn id_Y ist surjektiv.

(9) „⇒“: Ist f bijektiv, dann hat $g = f^{-1}$ die verlangte Eigenschaft.
„⇐“: Nach Teil (7) ist f injektiv und nach Teil (8) auch surjektiv, also bijektiv. ❑

Wenn man Abbildungen definieren möchte, die von zwei (oder mehr) Elementen möglicherweise verschiedener Mengen abhängen, dann kann man dies unter Zuhilfenahme von kartesischen Produkten tun: Möchte man einem Element von M_1 und einem Element von M_2 ein Element von N zuordnen, so entspricht das einer Abbildung $M_1 \times M_2 \to N$. Zum Beispiel kann man die Addition reeller Zahlen als eine Abbildung $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $(x, y) \mapsto x + y$, auffassen. Ist $f\colon M_1 \times M_2 \to N$ eine Abbildung, dann schreibt man auch $f(m_1, m_2)$ für $f\big((m_1, m_2)\big)$. Eine Abbildung der Form $M \times M \to M$ heißt auch eine *Verknüpfung* auf M.

DEF Verknüpfung

Schließlich führen wir noch eine weitere Interpretation und Schreibweise für Abbildungen ein, die häufig vorkommt.

2.19. **Definition.** Wenn $a\colon I \to M$ eine Abbildung ist, dann schreibt man dafür auch $(a_i)_{i \in I}$ und nennt das eine *Familie* von Elementen von M mit der *Indexmenge* I. Dabei ist $a_i = a(i)$ der Wert der Abbildung a an der Stelle $i \in I$. Sie kennen das von Folgen $(a_n)_{n \in \mathbb{N}}$. Die n-Tupel, die wir vor einer Weile eingeführt haben, kann man als den Spezialfall $I = \{1, 2, \ldots, n\}$ einer solchen Familie betrachten. In Analogie zur Schreibweise M^n für die Menge der n-Tupel mit Komponenten aus M schreibt man auch M^I für die Menge der Familien von Elementen von M mit Indexmenge I. Das ist (bis auf die Schreibweise) nichts anderes als die Menge $\mathrm{Abb}(I, M)$ der Abbildungen von I nach M. ♢

DEF Familie M^I

Der Unterschied zwischen einer *Abbildung* $a\colon I \to M$ und einer *Familie* $(a_i)_{i \in I}$ von Elementen von M ist lediglich einer der Perspektive:

- Bei der Abbildung interessiert man sich mehr für den Vorgang der *Zuordnung*.
- Bei der Familie interessiert man sich mehr für die vorkommenden *Werte*.

An dieser Stelle bietet es sich an, etwas mehr zur Mengenlehre zu sagen. Was wir hier betreiben, ist „naive“ Mengenlehre; wir machen uns hier also keine Gedanken darüber,


B. Russell
(1872–1970)

welche Konstruktionen mit Mengen tatsächlich möglich oder erlaubt sind. Das führt normalerweise auch nicht zu Problemen. Sie sollten aber wissen, dass die Mengenlehre durchaus nicht so harmlos ist, wie sie einem zunächst erscheinen mag. Wenn man bei der Bildung von Mengen zu viel erlaubt, kommt man in Schwierigkeiten, wie die berühmte *Russellsche Antinomie* zeigt. Denn dann könnte man die „Menge aller Mengen, die sich nicht selbst als Element enthalten", also $M = \{x \mid x \notin x\}$ konstruieren. Die Frage, ob M ein Element von M ist, führt auf einen unauflösbaren Widerspruch. (In der Unterhaltungsmathematik gibt es die Variante mit dem Dorfbarbier, der genau die Männer im Dorf rasiert, die sich nicht selbst rasieren. Rasiert sich nun der Barbier oder nicht?) Um diesen Widerspruch zu vermeiden, muss man genaue Regeln formulieren, wie man Mengen konstruieren darf. Das führt zur *axiomatischen Mengenlehre*.

Die meisten der Axiome sind recht „harmlos"; sie besagen etwa, dass die leere Menge existiert, dass man Einer- und Zweiermengen bilden kann, dass man immer Teilmengen bilden kann, und dass Vereinigungsmengen und Potenzmengen existieren. Es gibt aber ein Axiom, das *Auswahlaxiom*, das von einigen Mathematikern abgelehnt wurde. Es besagt, dass „es zu jeder Familie nichtleerer Mengen eine Auswahlfunktion gibt". Genauer: Ist $(X_i)_{i\in I}$ eine Familie von Mengen mit $X_i \neq \emptyset$ für alle $i \in I$, dann gibt es eine *Auswahlfunktion* $f\colon I \to X$, wobei $X = \{x \mid \exists i \in I\colon x \in X_i\}$ die Vereinigung aller Mengen X_i ist (die nach einem der harmlosen Axiome existiert), sodass für jedes $i \in I$ das Bild $f(i)$ ein Element von X_i ist. Die Auswahlfunktion wählt also aus jeder Menge X_i ein Element aus. Wir haben dieses Auswahlaxiom im Beweis von Teil (8) von Satz 2.18 benutzt, als wir für jedes $y \in Y$ ein Urbild x_y ausgewählt haben. Der Grund für die Ablehnung des Auswahlaxioms liegt darin, dass es nicht „konstruktiv" ist: Es macht eine Existenzaussage („es gibt eine Auswahlfunktion"), sagt aber nicht, *wie* man eine Auswahlfunktion bekommt. Heutzutage vertreten die meisten Mathematiker den pragmatischen Standpunkt, dass das Auswahlaxiom nützlich ist und es deswegen seine Berechtigung hat. Vor allem in der Analysis käme man ohne das Auswahlaxiom nicht weit. Es ist bekannt, dass die Hinzunahme des Auswahlaxioms nicht zu einem Widerspruch in der Mengenlehre führt, wenn die Mengenlehre ohne das Auswahlaxiom widerspruchsfrei ist (allerdings gilt das auch für seine Verneinung).

Zum Abschluss dieses Abschnitts über Grundlagen gibt es hier noch eine Tabelle mit griechischen Buchstaben. Als Mathematiker gehen einem schnell die Buchstaben aus, um die verschiedenen Objekte zu bezeichnen, mit denen man es zu tun hat. Darum wird gerne auf das griechische Alphabet zurückgegriffen.

klein	groß	Name
α	A	Alpha
β	B	Beta
γ	Γ	Gamma
δ	Δ	Delta
ε, ϵ	E	Epsilon
ζ	Z	Zeta
η	H	Eta
θ, ϑ	Θ	Theta

klein	groß	Name
ι	I	Iota
κ	K	Kappa
λ	Λ	Lambda
μ	M	My
ν	N	Ny
ξ	Ξ	Xi
o	O	Omikron
π	Π	Pi

klein	groß	Name
ρ, ϱ	P	Rho
σ	Σ	Sigma
τ	T	Tau
υ	Υ	Ypsilon
ϕ, φ	Φ	Phi
χ	X	Chi
ψ	Ψ	Psi
ω	Ω	Omega

3. Algebraische Strukturen: Gruppen, Ringe, Körper

In diesem Abschnitt werden wir die wichtigsten algebraischen Strukturen einführen. Gruppen treten in vielen Zusammenhängen in der Mathematik auf, allerdings wird das hier in der Linearen Algebra noch nicht so deutlich werden. Für uns wichtig sind Körper (das sind Strukturen, in denen man die vier Grundrechenarten zusammen mit den üblichen Rechenregeln zur Verfügung hat), denn zu einem Vektorraum (das ist die Struktur, die in der Linearen Algebra hauptsächlich betrachtet wird) gehört immer ein Körper, aus dem die „Skalare“ kommen. Ringe sind gewissermaßen Körper ohne Division; sie sind als Zwischenschritt bei der Definition von Körpern praktisch und auch wichtig in der Algebra. Sie werden ausführlicher in der Vorlesung „Einführung in die Zahlentheorie und algebraische Strukturen“ untersucht.

Wir beginnen mit dem Minimum, das man für eine halbwegs interessante algebraische Struktur braucht.

DEF Halbgruppe

3.1. **Definition.** Eine *Halbgruppe* ist ein Paar $(H, *)$, bestehend aus einer Menge H und einer Verknüpfung $*\colon H \times H \to H$, $(a, b) \mapsto a*b$, die das *Assoziativgesetz* erfüllt:

$$\forall a, b, c \in H\colon (a * b) * c = a * (b * c)\,.$$

Die Halbgruppe heißt *kommutativ*, wenn zusätzlich das *Kommutativgesetz* gilt:

$$\forall a, b \in H\colon a * b = b * a\,.$$

Wenn die Verknüpfung $*$ aus dem Kontext klar ist, spricht man der Einfachheit halber meist von „der Halbgruppe H“. ♢

Eine Bemerkung zur Notation: Verknüpfungen in algebraischen Strukturen wie $*$ in obiger Definition werden gerne in „Infix-Notation“ geschrieben, also $a * b$ statt $*(a, b)$.

Das Assoziativgesetz bewirkt, dass es nicht darauf ankommt, wie Ausdrücke, die drei oder mehr Elemente miteinander verknüpfen, geklammert sind. Zum Beispiel gilt für beliebige Elemente a, b, c, d, e von H:

$$\begin{aligned} a * ((b * c) * d) &= a * (b * (c * d)) = (a * b) * (c * d) \\ &= ((a * b) * c) * d = (a * (b * c)) * d \qquad \text{und} \\ a * (b * (c * (d * e))) &= (a * b) * (c * (d * e)) = ((a * b) * (c * d)) * e = \dots\,. \end{aligned}$$

Man kann deswegen einfach $a * b * c * d$ bzw. $a * b * c * d * e$ schreiben.

Hier ergibt sich die interessante kombinatorische Frage, *wie viele* verschiedene Klammerungen es für eine Verknüpfung von n Elementen gibt. Wir schreiben C_n für diese Zahl. Dann gilt offenbar $C_1 = C_2 = 1$, $C_3 = 2$ und $C_4 = 5$. Wenn man sich überlegt, dass man n Elemente dadurch verknüpfen kann, dass man eine Verknüpfung von k Elementen (mit $1 \le k < n$) mit einer Verknüpfung von $n - k$ Elementen verknüpft, dann sieht man die folgende Rekursion für die Zahlen C_n:

$$C_n = \sum_{k=1}^{n-1} C_k C_{n-k} = C_1 C_{n-1} + C_2 C_{n-2} + \ldots + C_{n-2} C_2 + C_{n-1} C_1 \qquad \text{für alle } n \ge 2.$$

Damit kann man dann zum Beispiel $C_5 = 1 \cdot 5 + 1 \cdot 2 + 2 \cdot 1 + 5 \cdot 1 = 14$, $C_6 = 42$, $C_7 = 132$ usw. berechnen. Es gibt auch eine Formel für C_n, nämlich

$$C_n = \frac{1}{n}\binom{2n-2}{n-1} = \frac{1}{2n-1}\binom{2n-1}{n-1} = \frac{(2n-2)!}{(n-1)!n!}\,,$$

die aber direkt nicht so einfach zu beweisen ist (was Sie natürlich nicht von einem Versuch abhalten soll!). Die Zahlen C_n heißen *Catalan-Zahlen* (was die Bezeichnung erklärt; oft ist der Index verschoben und man fängt mit $C_0 = C_1 = 1$ an) und treten in der Kombinatorik in vielen verschiedenen Zusammenhängen auf.

Wenn die Halbgruppe kommutativ ist, dann kommt es auch nicht auf die Reihenfolge an:

$$a * b * c = b * a * c = b * c * a = c * b * a = c * a * b = a * c * b\,.$$

DEF $\mathbb{N}_{>0}$

Im Folgenden schreiben wir $\mathbb{N}_{>0} = \{n \in \mathbb{N} \mid n > 0\} = \{1, 2, 3, \ldots\}$ für die Menge der echt positiven natürlichen Zahlen.

BSP Halbgruppen

3.2. **Beispiele.** Das Trivialbeispiel einer Halbgruppe ist $(\emptyset, *)$, wobei $*\colon \emptyset \times \emptyset \to \emptyset$ die leere Abbildung ist (beachte: $\emptyset \times \emptyset = \emptyset$).

Beispiele von kommutativen Halbgruppen sind $(\mathbb{N}_{>0}, +)$, $(\mathbb{N}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{N}_{>0}, \cdot)$, $(\mathbb{N}, \cdot)$, $(\mathbb{Z}, \cdot)$. Die Halbgruppe $(\mathrm{Abb}(X, X), \circ)$ für eine beliebige Menge X, mit der Komposition von Abbildungen als Verknüpfung, ist im Allgemeinen nicht kommutativ. (Diese Halbgruppe ist genau dann kommutativ, wenn X höchstens ein Element hat — Übung!) ♣

Mit Halbgruppen kann man allerdings noch nicht allzu viel anfangen. Deshalb fordern wir zusätzliche Eigenschaften.

DEF Monoid

3.3. **Definition.** Ein *Monoid* ist ein Tripel $(M, *, e)$, bestehend aus einer Menge M, einer Verknüpfung $*\colon M \times M \to M$ und einem Element $e \in M$, sodass $(M, *)$ eine Halbgruppe mit *neutralem Element* e ist:

$$\forall a \in M\colon e * a = a = a * e\,.$$

Das Monoid heißt *kommutativ*, wenn die Halbgruppe $(M, *)$ kommutativ ist. ♢

Wenn es ein neutrales Element gibt, dann ist es eindeutig bestimmt, wie das folgende Lemma zeigt. (Ein *Lemma* ist eine Hilfsaussage oder ein weniger wichtiger mathematischer Satz. Plural: Lemmas oder Lemmata (wenn man seine klassische Bildung heraushängen lassen will).)

LEMMA Eindeutigkeit des neutralen Elements

3.4. **Lemma.** *Sei $(H, *)$ eine Halbgruppe. Ist e ein links- und e' ein rechtsneutrales Element in dieser Halbgruppe, also*

$$\forall a \in H\colon e * a = a \text{ und } a * e' = a\,,$$

dann gilt $e = e'$.

Beweis. Da e linksneutral ist, gilt $e * e' = e'$. Da e' rechtsneutral ist, gilt $e * e' = e$. Es folgt $e = e'$. ❑

Aus diesem Grund lässt man meistens die Angabe des neutralen Elements weg und spricht vom „Monoid $(M, *)$“ oder auch nur vom „Monoid M“, wenn die Verknüpfung aus dem Kontext klar ist.

Es ist allerdings möglich, dass es in einer Halbgruppe zum Beispiel mehrere linksneutrale Elemente (und dann natürlich kein rechtsneutrales Element) gibt. Wenn etwa M beliebig ist und man als Verknüpfung pr_2 wählt (also $a * b = b$), dann hat man eine Halbgruppe, in der *alle* Elemente linksneutral sind.

BSP Monoide

3.5. **Beispiele.** Da die Definition von „Monoid“ ein neutrales Element fordert, kann die leere Menge kein Monoid sein. Das triviale Monoid ist dann $(\{e\}, *, e)$, wobei $*$ die einzige Abbildung $\{e\} \times \{e\} \to \{e\}$ ist (es ist also $e * e = e$).

Bis auf $(\mathbb{N}_{>0}, +)$, wo es kein neutrales Element gibt, lassen sich alle Beispiele von Halbgruppen aus 3.2 als Monoide $(\mathbb{N}, +, 0)$, $(\mathbb{Z}, +, 0)$, $(\mathbb{N}_{>0}, \cdot, 1)$, $(\mathbb{N}, \cdot, 1)$, $(\mathbb{Z}, \cdot, 1)$ und $(\mathrm{Abb}(X, X), \circ, \mathrm{id}_X)$ betrachten. ♣

Noch schöner ist es, wenn sich die Verknüpfung mit einem Element durch die Verknüpfung mit einem (in der Regel) anderen Element wieder rückgängig machen lässt. Das führt auf den Begriff der Gruppe.

DEF Gruppe

3.6. **Definition.** Eine *Gruppe* ist ein Quadrupel $(G, *, e, i)$, bestehend aus einer Menge G, einer Verknüpfung $*\colon G \times G \to G$, einem Element $e \in G$ und einer Abbildung $i\colon G \to G$, sodass $(G, *, e)$ ein Monoid ist und für jedes $g \in G$ das Element $i(g) \in G$ ein *Inverses* von g ist:

$$\forall g \in G\colon i(g) * g = e = g * i(g)\,.$$

Die Gruppe heißt *kommutativ* oder *abelsch*, wenn das Monoid $(G, *, e)$ kommutativ ist. ♢

N.H. Abel
1802–1829

Die Bezeichnung „abelsch“ ehrt den norwegischen Mathematiker Niels Henrik Abel, nach dem auch der *Abelpreis* benannt ist, ein dem Nobelpreis vergleichbarer Preis für Mathematik, der seit 2003 jährlich verliehen wird.

Auch Inverse sind eindeutig bestimmt:

LEMMA Eindeutigkeit des Inversen

3.7. **Lemma.** *Sei $(M, *, e)$ ein Monoid und sei $a \in M$. Ist $b \in M$ ein Linksinverses und $c \in M$ ein Rechtsinverses von a, also*

$$b * a = e = a * c\,,$$

dann gilt $b = c$.

Beweis. Wir haben

$$b = b * e = b * (a * c) = (b * a) * c = e * c = c\,.$$ ❑

Analog zu Monoiden spricht man deshalb auch einfach von „der Gruppe $(G, *)$“ oder auch von „der Gruppe G“, wenn die Verknüpfung aus dem Kontext klar ist.

Gruppen schreibt man gerne „multiplikativ“, dann ist die Verknüpfung $a \cdot b$ oder kurz ab, das neutrale Element heißt 1 und das Inverse von a wird a^{-1} geschrieben. Kommutative Gruppen schreibt man auch häufig „additiv“, dann ist die Verknüpfung $a + b$, das neutrale Element heißt 0 und das Inverse von a wird als das Negative von a geschrieben: $-a$. Dann schreibt man auch kurz $a - b$ für $a + (-b)$.

BSP Gruppen

3.8. **Beispiele.** Das triviale Monoid lässt sich auch als Gruppe betrachten, denn das einzige Element e ist sein eigenes Inverses.

Von den übrigen Beispielen von Monoiden in 3.5 kann nur $(\mathbb{Z}, +, 0, -)$ auch als Gruppe betrachtet werden (und im letzten Beispiel $\mathrm{Abb}(X, X)$, wenn X höchstens ein Element hat; dann hat man eine triviale Gruppe). Ein weiteres Beispiel einer kommutativen Gruppe ist $(\mathbb{R}_{>0}, \cdot, 1, x \mapsto 1/x)$, wobei $\mathbb{R}_{>0}$ die Menge der positiven reellen Zahlen ist.

Wenn man sich bei den Abbildungen $X \to X$ auf die bijektiven Abbildungen beschränkt, dann erhält man eine Gruppe $(S(X), \circ, \mathrm{id}_X, f \mapsto f^{-1})$, die auch die *symmetrische Gruppe* von X heißt. Dabei ist

DEF Symmetrische Gruppe $S(X)$

$$S(X) = \{f\colon X \to X \mid f \text{ bijektiv}\}\,.$$

Diese Gruppe ist genau dann kommutativ, wenn X höchstens zwei Elemente enthält (Übung).

Gruppen treten häufig in der Mathematik als „Symmetriegruppen“ von irgendwelchen Objekten auf. Zum Beispiel bilden die Drehungen und Spiegelungen der Ebene, die ein regelmäßiges n-Eck auf sich abbilden, eine Gruppe, oder die Drehungen des dreidimensionalen Raumes, die ein reguläres Tetraeder, einen Würfel (oder ein reguläres Oktaeder) oder ein reguläres Dodekaeder (oder Ikosaeder) in sich abbilden, bilden jeweils eine Gruppe, die Tetraeder-, Oktaeder- und Ikosaedergruppe. In einem recht allgemeinen Sinn ist die symmetrische Gruppe $S(X)$ die Symmetriegruppe der Menge X ohne weitere Struktur. In der Algebra treten Symmetriegruppen als „Automorphismengruppen“ auf. Zum Beispiel bildet für eine Gruppe $(G, *)$ die Menge

$$\mathrm{Aut}(G) = \{f\colon G \to G \mid f \text{ bijektiv und } \forall g, g' \in G\colon f(g * g') = f(g) * f(g')\}$$

mit der Komposition von Abbildungen eine Gruppe, die *Automorphismengruppe* von G. Sie besteht aus den bijektiven Abbildungen $G \to G$, die mit der Struktur von G als Gruppe verträglich sind. ♣

Damit eine Halbgruppe sogar eine Gruppe ist, genügt es, die Existenz eines linksneutralen Elements e und für jedes Element x die Existenz eines Linksinversen $i(x)$ (also mit $i(x) * x = e$) zu fordern. Dann folgt zunächst, dass e auch rechtsneutral ist, denn es gilt

$$x*e = e*x*e = i(i(x))*i(x)*x*e = i(i(x))*e*e = i(i(x))*e = i(i(x))*i(x)*x = e*x = x\,.$$

Daraus ergibt sich auch $i(i(x)) = x$. Damit kann man dann zeigen, dass $i(x)$ auch Rechtsinverses von x ist:

$$x * i(x) = i(i(x)) * i(x) = e\,.$$

Ganz analog funktioniert das natürlich auch, wenn man „links“ jeweils durch „rechts“ ersetzt. Auf der anderen Seite gibt es aber Halbgruppen mit linksneutralen und rechtsinversen Elementen, die keine Gruppen sind. Finden Sie ein Beispiel!

Eine wichtige Eigenschaft von Gruppen ist, dass sich gewisse Gleichungen stets eindeutig lösen lassen. Zuerst beweisen wir aber eine Kürzungsregel.

3.9. **Lemma.** *Sei $(G, *, e, i)$ eine Gruppe und seien $a, b, c \in G$. Dann gilt*

LEMMA Kürzungsregel in Gruppen

$$a * c = b * c \iff a = b \iff c * a = c * b\,.$$

Beweis. Wir beweisen die erste Äquivalenz; der Beweis der zweiten ist analog.

„$\Leftarrow$“ ist klar. Für „$\Rightarrow$“ haben wir

$$\begin{aligned} a * c = b * c &\implies (a * c) * i(c) = (b * c) * i(c) \implies a * (c * i(c)) = b * (c * i(c)) \\ &\implies a * e = b * e \implies a = b\,. \end{aligned}$$

❑

LEMMA Gleichungen in Gruppen

3.10. **Lemma.** *Sei $(G, *, e, i)$ eine Gruppe und seien $a, b \in G$. Dann haben die Gleichungen*

$$a * x = b \qquad \text{und} \qquad x * a = b$$

jeweils eine eindeutige Lösung $x \in G$, nämlich $x = i(a) * b$ bzw. $x = b * i(a)$.

Beweis. Wir führen den Beweis exemplarisch für die erste Gleichung:

$$a * x = b \iff i(a) * a * x = i(a) * b \iff e * x = i(a) * b \iff x = i(a) * b\,.$$

Für die erste Äquivalenz haben wir Lemma 3.9 benutzt. ❑

Als Nächstes betrachten wir Strukturen mit zwei Verknüpfungen.

DEF Ring

3.11. **Definition.** Ein *Ring* ist ein Sextupel $(R, +, 0, -, \cdot, 1)$, bestehend aus einer Menge R, Verknüpfungen $+, \cdot\colon R \times R \to R$, Elementen $0, 1 \in R$ und einer Abbildung $-\colon R \to R$, sodass $(R, +, 0, -)$ eine kommutative Gruppe und $(R, \cdot, 1)$ ein Monoid ist und die *Distributivgesetze*

$$\forall a, b, c \in R\colon\ a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{und} \quad (a + b) \cdot c = a \cdot c + b \cdot c$$

gelten. Der Ring heißt *kommutativ*, wenn das Monoid $(R, \cdot, 1)$ kommutativ ist. ♢

Da die neutralen und inversen Elemente eindeutig bestimmt sind, spricht man oft nur vom „Ring $(R, +, \cdot)$“ oder sogar vom „Ring R“, wenn die Verknüpfungen aus dem Kontext klar sind. Ist der Ring kommutativ, dann genügt es, eines der beiden Distributivgesetze zu fordern (das andere folgt dann). Für das Produkt $a \cdot b$ zweier Elemente schreibt man auch kurz ab.

In einem Ring kann man also addieren, subtrahieren und multiplizieren, und die üblichen Rechenregeln gelten, wie zum Beispiel $0 \cdot a = a \cdot 0 = 0$, $-(a+b) = -a - b$, $(-a) \cdot (-b) = a \cdot b$. Was aber im Allgemeinen *nicht* gelten muss, ist die Implikation $a \cdot b = 0 \Rightarrow a = 0 \vee b = 0$.

BSP Ringe

3.12. **Beispiele.** Das Trivialbeispiel für einen Ring ist der sogenannte *Nullring* $(\{0\}, +, 0, -, \cdot, 0)$, in dem $0 = 1$ und $0 + 0 = -0 = 0 \cdot 0 = 0$ gelten. Jeder Ring R, in dem $0_R = 1_R$ gilt, ist so ein Nullring, denn für alle $r \in R$ gilt dann $r = 1_R \cdot r = 0_R \cdot r = 0_R$.

Das Standardbeispiel für einen (kommutativen) Ring ist der Ring $\mathbb{Z}$ der ganzen Zahlen mit der üblichen Addition und Multiplikation als Verknüpfungen. Ein etwas anders geartetes Beispiel ist $(\mathcal{P}(X), \triangle, \emptyset, \mathrm{id}_{\mathcal{P}(X)}, \cap, X)$ für eine beliebige Menge X; dabei ist $T_1 \triangle T_2 = (T_1 \setminus T_2) \cup (T_2 \setminus T_1)$ die „symmetrische Differenz“ der Mengen T_1 und T_2 (Übung).

Falls Sie aus der Schule Matrizen kennen und wissen, wie man sie addiert und multipliziert, dann können Sie nachprüfen, dass die Menge der 2×2-Matrizen mit Einträgen aus $\mathbb{R}$ zusammen mit der Addition und Multiplikation von Matrizen einen nicht-kommutativen Ring bildet. ♣

Schließlich kommen wir zu den Körpern. (Der Stern im linken Rand neben der Definition bedeutet, dass dies eine der Definitionen ist, die möglicherweise in der Klausur explizit abgefragt werden. Ähnliches gilt für Sätze „mit Stern“. Das heißt natürlich nicht, dass Sie die übrigen Definitionen und Sätze nicht kennen müssten!)

* 3.13. **Definition.** Ein *Körper* ist ein Septupel $(K, +, 0, -, \cdot, 1, i)$, bestehend aus einer Menge K, Verknüpfungen $+, \cdot\colon K \times K \to K$, Elementen $0, 1 \in K$ und Abbildungen $-\colon K \to K$, $i\colon K \setminus \{0\} \to K \setminus \{0\}$, sodass $(K, +, 0, -, \cdot, 1)$ ein kommutativer Ring und $(K \setminus \{0\}, \cdot, 1, i)$ eine (kommutative) Gruppe ist. Für $i(a)$ schreibt man a^{-1}. ♢

DEF Körper

Wie üblich spricht man meistens einfach von dem „Körper $(K, +, \cdot)$“ oder von dem „Körper K“. Aus der Definition folgt, dass 0 und 1 in einem Körper verschieden sein müssen, denn 1 soll das neutrale Element der Gruppe $K \setminus \{0\}$ sein. Für diese Gruppe $(K \setminus \{0\}, \cdot)$ schreibt man auch $K^\times$ und nennt sie die *multiplikative Gruppe* von K. (Häufig findet man auch die Schreibweise K^* dafür.)

DEF multiplikative Gruppe $K^\times$

Man kann natürlich auch ohne Rückgriff auf Ringe und Gruppen definieren, was ein Körper ist. Dann hat man für alle $a, b, c \in K$ die folgenden Axiome:

$$\begin{array}{cc} (a+b)+c = a+(b+c)\,, & a+b = b+a \\ a+0 = a\,, & a+(-a) = 0 \\ (a\cdot b)\cdot c = a\cdot(b\cdot c) & a\cdot b = b\cdot a \\ a\cdot 1 = a\,, & a \neq 0 \Rightarrow a\cdot a^{-1} = 1 \\ 0 \neq 1\,, & a\cdot(b+c) = a\cdot b + a\cdot c \end{array}$$

Für $a, b \in K$, $b \neq 0$, kann man die Division definieren durch $a/b = a \cdot b^{-1}$. Dann hat man die vier Grundrechenarten zur Verfügung und die üblichen Rechenregeln dafür gelten, denn man kann sie aus den Körperaxiomen ableiten. Zum Beispiel gilt in einem Körper stets, dass aus $a \cdot b = 0$ folgt, dass $a = 0$ oder $b = 0$ ist. (Denn ist $a \neq 0$, dann folgt $0 = a^{-1} \cdot 0 = a^{-1} \cdot a \cdot b = 1 \cdot b = b$.)

3.14. **Beispiele.** Das kleinste Beispiel für einen Körper hat nur die beiden Elemente 0 und 1, die in der Definition gefordert werden. Für die Addition und Multiplikation folgen $0 + 0 = 0$, $0 + 1 = 1 + 0 = 1$, $0 \cdot 0 = 0 \cdot 1 = 1 \cdot 0 = 0$ und $1 \cdot 1 = 1$ direkt aus der Definition; für die verbleibende Summe $1 + 1$ bleibt nur der Wert 0, da die Gleichung $a + 1 = 0$ lösbar sein muss. Man kann (einfach, aber länglich) nachprüfen, dass dieser Körper, der mit $\mathbb{F}_2$ bezeichnet wird, die Axiome erfüllt.

BSP Körper

Es gibt noch weitere endliche Körper: Zu jeder Potenz p^e einer Primzahl p (mit $e \geq 1$) gibt es im Wesentlichen („bis auf Isomorphie“) genau einen Körper $\mathbb{F}_{p^e}$ mit p^e Elementen, und es gibt keine anderen endlichen Körper. Das wird in der „Einführung in die Algebra“ genauer besprochen.

Standardbeispiele für Körper sind der Körper $\mathbb{Q}$ der rationalen Zahlen und der Körper $\mathbb{R}$ der reellen Zahlen, jeweils mit der bekannten Addition und Multiplikation. Im nächsten Abschnitt werden wir einen weiteren Körper konstruieren, den Körper $\mathbb{C}$ der komplexen Zahlen. ♣

4. Der Körper der komplexen Zahlen

Der Körper $\mathbb{R}$ der reellen Zahlen hat, wie Sie in der Analysis lernen, viele schöne Eigenschaften. Eine Eigenschaft allerdings fehlt ihm: Es sind nicht alle Gleichungen der Form

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

(mit $n \geq 1$ und $a_0, a_1, \ldots, a_{n-1} \in \mathbb{R}$) in $\mathbb{R}$ lösbar.

Für ungerades n folgt aus dem *Zwischenwertsatz*, dass es stets eine Lösung geben muss; das lernen Sie bald in der Analysis.

Die einfachste Gleichung dieser Art ohne Lösung ist $x^2 + 1 = 0$: Die linke Seite ist stets ≥ 1, kann also niemals null werden. Wir werden jetzt einen $\mathbb{R}$ umfassenden Körper konstruieren, in dem diese Gleichung eine Lösung hat.

Um zu sehen, wie man dabei vorgehen kann, stellen wir uns einfach einmal vor, dass wir schon so einen Körper hätten; wir nennen ihn $\mathbb{C}$. Dann haben wir eine Lösung $\boldsymbol{i}$ obiger Gleichung, also ein Element $\boldsymbol{i} \in \mathbb{C}$ mit $\boldsymbol{i}^2 = -1$. Wir haben natürlich auch die reellen Zahlen in $\mathbb{C}$. Mit $a, b \in \mathbb{R}$ können wir dann das Element $a + b\boldsymbol{i} \in \mathbb{C}$ erzeugen. Muss es noch weitere Elemente geben? Dazu müssen wir überprüfen, ob die vier Grundrechenarten aus der Menge der Elemente der Form $a + b\boldsymbol{i}$ herausführen. Seien $a, b, a', b' \in \mathbb{R}$. Dann gilt, wenn $\mathbb{C}$ ein Körper ist,

$$(a + b\boldsymbol{i}) + (a' + b'\boldsymbol{i}) = (a + a') + (b + b')\boldsymbol{i} \qquad \text{und}$$
$$(a + b\boldsymbol{i}) \cdot (a' + b'\boldsymbol{i}) = aa' + ab'\boldsymbol{i} + ba'\boldsymbol{i} + bb'\boldsymbol{i}^2 = (aa' - bb') + (ab' + ba')\boldsymbol{i}\,.$$

Dabei haben wir $\boldsymbol{i}^2 = -1$ benutzt. Offensichtlich ist das additive Inverse (also das Negative) von $a + b\boldsymbol{i}$ gerade $(-a) + (-b)\boldsymbol{i}$. Wie sieht es mit dem multiplikativen Inversen aus (also dem Kehrwert)? Dazu überlegen wir uns erst, dass genau dann $a + b\boldsymbol{i} = 0$ ist, wenn $a = b = 0$ gilt. Eine Richtung („$\Leftarrow$“) ist klar. Umgekehrt sei $a + b\boldsymbol{i} = 0$. Dann folgt

$$0 = (a - b\boldsymbol{i}) \cdot 0 = (a - b\boldsymbol{i}) \cdot (a + b\boldsymbol{i}) = a^2 + b^2\,.$$

Da a und b reelle Zahlen sind, ist das nur möglich, wenn $a = b = 0$ gilt. Seien also a und b nicht beide null. Dann sollte gelten (das ist der alte Trick, wie man „Quadratwurzeln aus dem Nenner entfernt“; man beachte, dass $\boldsymbol{i} = \text{„}\sqrt{-1}\text{“}$ ist):

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

Offenbar brauchen wir also keine zusätzlichen Elemente.

Um das Ganze formal auf eine solide Grundlage zu stellen, ersetzen wir einen Ausdruck der Form $a + b\boldsymbol{i}$ durch das Paar $(a, b) \in \mathbb{R} \times \mathbb{R}$. Wir schreiben $\mathbb{C}$ für $\mathbb{R} \times \mathbb{R}$ und definieren die folgenden Abbildungen:

$$\begin{aligned}
+_{\mathbb{C}} &\colon \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}, & \big((a,b),(a',b')\big) &\longmapsto (a + a', b + b')\\
\cdot_{\mathbb{C}} &\colon \mathbb{C} \times \mathbb{C} \longrightarrow \mathbb{C}, & \big((a,b),(a',b')\big) &\longmapsto (aa' - bb', ab' + ba')\\
-_{\mathbb{C}} &\colon \mathbb{C} \longrightarrow \mathbb{C}, & (a,b) &\longmapsto (-a, -b)\\
i_{\mathbb{C}} &\colon \mathbb{C} \setminus \{(0,0)\} \longrightarrow \mathbb{C} \setminus \{(0,0)\}, & (a,b) &\longmapsto \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2}\right)
\end{aligned}$$

Außerdem schreiben wir $0_{\mathbb{C}}$ und $1_{\mathbb{C}}$ für $(0, 0)$ und $(1, 0)$.

4.1. **Satz.** *Die Menge $\mathbb{C} = \mathbb{R}\times\mathbb{R}$ zusammen mit den oben definierten Abbildungen und Elementen bildet einen Körper.* **SATZ** Körper $\mathbb{C}$

Beweis. Es sind die verschiedenen Axiome nachzuprüfen. Für die additive Gruppe $(\mathbb{C}, +_\mathbb{C}, 0_\mathbb{C}, -_\mathbb{C})$ ist das sehr leicht; darum lassen wir das hier weg (es sei Ihnen aber als Übung empfohlen). Wir prüfen Assoziativität und Kommutativität der Multiplikation. Dabei benutzen wir, dass $\mathbb{R}$ ein Körper ist, dass also dort die bekannten Rechenregeln gelten.

$$\begin{aligned}\big((a,b)\cdot_\mathbb{C}(a',b')\big)\cdot_\mathbb{C}(a'',b'') &= (aa'-bb',ab'+a'b)\cdot_\mathbb{C}(a'',b'')\\ &= \big((aa'-bb')a''-(ab'+ba')b'', (aa'-bb')b''+(ab'+ba')a''\big)\\ &= (aa'a''-ab'b''-ba'b''-bb'a'', aa'b''+ab'a''+ba'a''-bb'b'')\end{aligned}$$

und dasselbe Resultat erhalten wir aus $(a,b)\cdot_\mathbb{C}\big((a',b')\cdot_\mathbb{C}(a'',b'')\big)$. Ebenso gilt

$$(a,b)\cdot_\mathbb{C}(a',b') = (aa'-bb',ab'+ba') = (a'a-b'b,ba'+ab') = (a',b')\cdot_\mathbb{C}(a,b)\,.$$

Dass $1_\mathbb{C} = (1,0)$ neutrales Element der Multiplikation ist, folgt aus

$$(1,0)\cdot_\mathbb{C}(a,b) = (1\cdot a - 0\cdot b, 1\cdot b + 0\cdot a) = (a,b)\,.$$

Wir rechnen nach, dass $i_\mathbb{C}\big((a,b)\big)$ das multiplikative Inverse von $(a,b)\neq(0,0)$ ist:

$$\begin{aligned}(a,b)\cdot_\mathbb{C} i_\mathbb{C}\big((a,b)\big) &= (a,b)\cdot_\mathbb{C}\left(\frac{a}{a^2+b^2},\frac{-b}{a^2+b^2}\right)\\ &= \left(\frac{a^2}{a^2+b^2}-\frac{-b^2}{a^2+b^2},\frac{-ab}{a^2+b^2}+\frac{ba}{a^2+b^2}\right)\\ &= (1,0) = 1_\mathbb{C}\,.\end{aligned}$$

$0_\mathbb{C}\neq 1_\mathbb{C}$ ist klar. Es bleibt das Distributivgesetz nachzuprüfen:

$$\begin{aligned}(a,b)\cdot_\mathbb{C}\big((a',b')+_\mathbb{C}(a'',b'')\big) &= (a,b)\cdot_\mathbb{C}(a'+a'',b'+b'')\\ &= \big(a(a'+a'')-b(b'+b''),a(b'+b'')+b(a'+a'')\big)\\ &= (aa'+aa''-bb'-bb'',ab'+ab''+ba'+ba'')\\ &= (aa'-bb'+aa''-bb'',ab'+ba'+ab''+ba'')\\ &= (aa'-bb',ab'+ba')+_\mathbb{C}(aa''-bb'',ab''+ba'')\\ &= (a,b)\cdot_\mathbb{C}(a',b')+_\mathbb{C}(a,b)\cdot_\mathbb{C}(a'',b'')\,.\end{aligned}$$

❑

Ist a eine reelle Zahl, dann haben wir das Element $a_\mathbb{C} = (a,0)\in\mathbb{C}$. Für $a,b\in\mathbb{R}$ gilt

$$a=b \iff a_\mathbb{C}=b_\mathbb{C}\,,\qquad (a+b)_\mathbb{C} = a_\mathbb{C}+_\mathbb{C} b_\mathbb{C}\qquad\text{und}\qquad (ab)_\mathbb{C} = a_\mathbb{C}\cdot_\mathbb{C} b_\mathbb{C}\,.$$

Mit den Elementen $a_\mathbb{C}$ rechnet man also genauso wie mit den zugehörigen reellen Zahlen a. Deswegen macht man keinen Unterschied zwischen a und $a_\mathbb{C}$ und betrachtet $\mathbb{R}$ als eine Teilmenge von $\mathbb{C}$. Wir schreiben also einfach a für das Element $a_\mathbb{C} = (a,0)$ von $\mathbb{C}$. Außerdem schreiben wir ab jetzt der Einfachheit halber meistens $+$, $\cdot$ und so weiter statt $+_\mathbb{C}$, $\cdot_\mathbb{C}$ etc.

4.2. **Definition.** Der in Satz 4.1 eingeführte Körper $\mathbb{C}$ heißt der *Körper der komplexen Zahlen*. Wir schreiben $\boldsymbol{i}$ für das Element $(0,1) \in \mathbb{C}$. Dann gilt $\boldsymbol{i}^2 = -1$, und jedes Element $z = (a,b) \in \mathbb{C}$ kann geschrieben werden als $z = a + b\boldsymbol{i}$ (oder $a + \boldsymbol{i}b$) mit $a, b \in \mathbb{R}$. Dann heißt a der *Realteil* $\operatorname{Re} z$ und b der *Imaginärteil* $\operatorname{Im} z$ von z. Gilt $\operatorname{Re} z = 0$, dann heißt z *rein imaginär*. ♢

DEF Körper der komplexen Zahlen

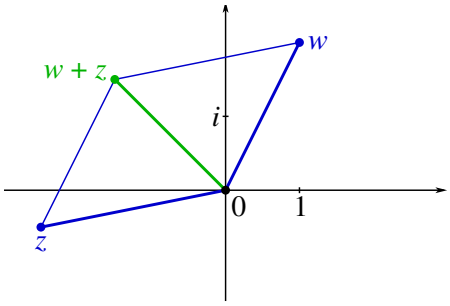
Die in der Definition gemachten Behauptungen sollten wir nachprüfen:

$$\boldsymbol{i}^2 = (0,1) \cdot_{\mathbb{C}} (0,1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) = (-1,0) = (-1)_{\mathbb{C}} = -1$$

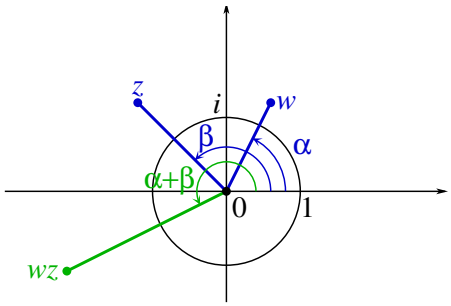
und

$$a + b\boldsymbol{i} = (a,0) +_{\mathbb{C}} (b,0) \cdot_{\mathbb{C}} (0,1) = (a,0) +_{\mathbb{C}} (0,b) = (a,b)\,.$$

Man kann sich die komplexen Zahlen ganz gut veranschaulichen, wenn man sich daran erinnert, dass $\mathbb{C} = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ der Menge der Punkte der Ebene entspricht. Wenn man die Ebene so interpretiert, spricht man auch von der *komplexen (Zahlen-)Ebene*. Die Addition entspricht dann dem, was Sie aus der Physik als „Kräfteparallelogramm“ kennen.



Addition $w + z$



Multiplikation $w \cdot z$

Auch die Multiplikation lässt sich geometrisch interpretieren. Wir betrachten dazu $z = a + b\boldsymbol{i} \in \mathbb{C}$. Dann ist $a^2 + b^2 \geq 0$; man setzt $|z| = \sqrt{a^2 + b^2}$ und nennt das den *Absolutbetrag* von z. Das entspricht dem Abstand des Punktes z in der komplexen Ebene vom Ursprung $0 \in \mathbb{C}$. Für $z \in \mathbb{R}$ (also $b = 0$) bekommt man den bekannten Absolutbetrag auf $\mathbb{R}$. Ist $z \neq 0$, dann hat $w = z/|z|$ den Absolutbetrag 1. Wenn wir $w = u + v\boldsymbol{i}$ schreiben, dann gilt $u^2 + v^2 = 1$, also liegt der Punkt (u,v) auf dem Einheitskreis. Es gibt dann $\alpha \in \mathbb{R}$ mit $u = \cos\alpha$, $v = \sin\alpha$. Dieser Winkel α heißt auch das *Argument* von w und von z (es ist aber nur bis auf Vielfache von $2\pi \mathrel{\hat{=}} 360°$ eindeutig bestimmt). Es gilt die Beziehung

DEF Absolutbetrag $|z|$

DEF Argument

$$(\cos\alpha + \boldsymbol{i}\sin\alpha) \cdot (\cos\beta + \boldsymbol{i}\sin\beta) = \cos(\alpha+\beta) + \boldsymbol{i}\sin(\alpha+\beta)$$

— das ist äquivalent zu den Additionstheoremen für Sinus und Cosinus:

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

Daher addieren sich die Winkel bei Multiplikation. Man kann das dann so formulieren: Multiplikation mit $z \neq 0$ bewirkt eine *Drehstreckung* der komplexen Ebene mit dem Drehwinkel α und dem Streckfaktor $|z|$.

In der Analysis werden Sie lernen, dass

$$\cos\alpha + \boldsymbol{i}\sin\alpha = e^{\alpha\boldsymbol{i}}$$

ist. Die Relation oben folgt dann aus $e^{x+y} = e^x \cdot e^y$.

Wir können jetzt immerhin zeigen, dass man quadratische Gleichungen in $\mathbb{C}$ stets lösen kann.

SATZ quadratische Gleichungen in $\mathbb{C}$

4.3. **Satz.** *Seien $a, b, c \in \mathbb{C}$ mit $a \neq 0$. Dann hat die Gleichung*

$$az^2 + bz + c = 0$$

mindestens eine Lösung $z \in \mathbb{C}$.

Beweis. Die Gleichung ist äquivalent zu $(2az + b)^2 = b^2 - 4ac$. Es genügt also zu zeigen, dass jede komplexe Zahl eine Quadratwurzel in $\mathbb{C}$ hat. (Ist $z' \in \mathbb{C}$ mit $z'^2 = b^2 - 4ac$, dann ist $z = (-b + z')/(2a)$ eine Lösung der Gleichung. Das ist die bekannte Lösungsformel für quadratische Gleichungen.) Sei also $w \in \mathbb{C}$. Wir wollen $z \in \mathbb{C}$ finden mit $z^2 = w$. Ist $w = 0$, dann ist $z = 0$ eine Lösung. Sonst können wir wie oben w als $w = |w|(\cos\alpha + \boldsymbol{i}\sin\alpha)$ schreiben. Dann ist $z = \sqrt{|w|}\big(\cos\frac{\alpha}{2} + \boldsymbol{i}\sin\frac{\alpha}{2}\big)$ eine Lösung. ❑

Analog zeigt man, dass es für jede komplexe Zahl w und jede natürliche Zahl $n \geq 1$ eine $n$te Wurzel $z \in \mathbb{C}$ von w gibt, also eine Lösung der Gleichung $z^n = w$. Es gilt aber sogar noch viel mehr.

SATZ Fundamentalsatz der Algebra

4.4. **Satz.** *Jede Gleichung*

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

mit $n \geq 1$ und $a_0, a_1, \ldots, a_{n-1} \in \mathbb{C}$ hat mindestens eine Lösung $z \in \mathbb{C}$.

Beweisen können wir diesen Satz hier nicht. Es gibt verschiedene Beweise; der wohl einfachste verwendet den *Satz von Liouville* aus der Funktionentheorie. Sie werden ihn in der Vorlesung „Funktionentheorie“ kennenlernen.

Zwar lassen sich Gleichungen der obigen Form mit $n \leq 4$ durch Ziehen von Quadrat- und Kubikwurzeln lösen (das wurde bereits im 16. Jahrhundert von del Ferro, Tartaglia und Ferrari entdeckt); Gleichungen mit $n \geq 5$ lassen sich jedoch im Allgemeinen nicht mehr mit Hilfe der vier Grundrechenarten und durch das Ziehen von beliebigen $m$ten Wurzeln lösen (Satz von Abel-Ruffini; erster vollständiger Beweis 1824 von Abel). Die Aussage von Satz 4.4 ist also viel stärker als die Existenz von $n$ten Wurzeln in $\mathbb{C}$.

Ein Körper K mit der Eigenschaft, dass jede Gleichung

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

mit $n \geq 1$ und $a_0, a_1, \ldots, a_{n-1} \in K$ eine Lösung $x \in K$ hat, heißt *algebraisch abgeschlossen*. Der „Fundamentalsatz der Algebra“ lässt sich also auch so formulieren:

DEF algebraisch abgeschlossen

Der Körper der komplexen Zahlen ist algebraisch abgeschlossen.

Demgegenüber ist der Körper der reellen Zahlen *nicht* algebraisch abgeschlossen, wie wir gesehen haben. In dieser Hinsicht ist $\mathbb{C}$ also „besser“ als $\mathbb{R}$. Auf der anderen Seite ist $\mathbb{C}$ kein angeordneter Körper mehr; man verliert also auch etwas beim Übergang von $\mathbb{R}$ zu $\mathbb{C}$. (In einem angeordneten Körper K gilt $x^2 \geq 0$ für $x \in K$. Damit müsste in $\mathbb{C}$ gelten, dass $-1 = \boldsymbol{i}^2 \geq 0$ ist, aber -1 ist in einem angeordneten Körper immer negativ, und wir haben einen Widerspruch.)

Da mit $\boldsymbol{i}$ auch $-\boldsymbol{i}$ eine Lösung von $x^2 + 1 = 0$ ist, könnte man überall $\boldsymbol{i}$ durch $-\boldsymbol{i}$ ersetzen und alles würde genauso funktionieren. Für $z = a + b\boldsymbol{i} \in \mathbb{C}$ setzen wir daher $\bar{z} = a - b\boldsymbol{i}$; die Abbildung $\mathbb{C} \to \mathbb{C}$, $z \mapsto \bar{z}$, heißt die *komplexe Konjugation*. Es gilt $\overline{w + z} = \bar{w} + \bar{z}$ und $\overline{wz} = \bar{w} \cdot \bar{z}$ (leichte Übung); außerdem $z\bar{z} = a^2 + b^2 = |z|^2$ (das haben wir schon benutzt). Daraus bekommt man die Formel $z^{-1} = \bar{z}/|z|^2$ für den Kehrwert einer komplexen Zahl $z \neq 0$; das ist derselbe Ausdruck, den wir bereits hergeleitet

hatten, in einer etwas abgekürzten Form. Außerdem hat die komplexe Konjugation noch die folgenden Eigenschaften:

$$z \in \mathbb{R} \iff z = \bar{z}\,, \qquad \operatorname{Re} z = \frac{z + \bar{z}}{2}\,, \qquad \operatorname{Im} z = \frac{z - \bar{z}}{2\boldsymbol{i}}\,.$$

5. Vektorräume: Definition und Beispiele

In diesem Abschnitt beginnen wir mit dem Studium der Linearen Algebra. Was ist „Lineare Algebra“? Die Lineare Algebra befasst sich mit „linearen Strukturen“, genauer mit *Vektorräumen* und *linearen Abbildungen* zwischen ihnen. Diese Begriffe sind zunächst einmal sehr abstrakt, aber darin liegt gerade die Stärke der Linearen Algebra: Vektorräume und lineare Abbildungen treten sehr häufig in der Mathematik in den unterschiedlichsten Zusammenhängen auf. Gerade weil man von den jeweils konkreten individuellen Umständen abstrahiert und sich auf die wesentlichen gemeinsamen Eigenschaften beschränkt, lassen sich die Ergebnisse der Linearen Algebra in all diesen unterschiedlichen Situationen anwenden. Es war, historisch gesehen, ein langwieriger Prozess, zu dieser Abstraktion zu gelangen, aber am Endpunkt dieser Entwicklung steht eine sehr leistungsfähige, allgemein anwendbare und erfolgreiche Theorie. Das hat dazu geführt, dass *lineare* Probleme als *einfach* gelten, während *nichtlineare* Probleme sehr häufig besonders *schwierig* sind. In Ausschreibungen für Mathematik-Professuren findet man zum Beispiel häufiger das Wort „nichtlinear“ (etwa im Kontext von „nichtlinearen partiellen Differentialgleichungen“), aber so gut wie niemals das Wort „linear“. Zwei Beispiele mit physikalischem Hintergrund: Die *Wärmeleitungsgleichung*, die die zeitliche Entwicklung der Temperaturverteilung in einem Körper beschreibt, ist eine *lineare* partielle Differentialgleichung. Die zugehörige Lösungstheorie wurde bereits von Jean-Baptiste-Joseph Fourier entwickelt („Théorie analytique de la chaleur“, 1822). Im Gegensatz dazu sind die *Navier-Stokes-Gleichungen*, die die Bewegung von Flüssigkeiten beschreiben, *nichtlineare* partielle Differentialgleichungen, und die Frage, ob sie für vernünftige Anfangsbedingungen im dreidimensionalen Raum immer eindeutig lösbar sind, ist eines der sieben Millenniumprobleme der Clay Foundation; für die Lösung bekommt man eine Million US-Dollar.


J.-B.-J. Fourier (1768–1830)

Was bedeutet nun „linear“? Dazu als Beispiele drei lineare Gleichungen (oder Gleichungssysteme):

BSP lineare Gleichungen

5.1. **Beispiele.**

(1) Wir suchen $w, x, y, z \in \mathbb{R}$ mit

$$w + x + y + z = 0 \qquad \text{und} \qquad x + 2y + 3z = 0\,.$$

Wahrscheinlich haben Sie in der Schule gelernt, wie man solche Gleichungssysteme löst (und in jedem Fall werden wir das auch in dieser Vorlesung besprechen). Als Lösungen erhält man

$$(w, x, y, z) = (a, -2a + b, a - 2b, b) \qquad \text{mit } a, b \in \mathbb{R}.$$

(2) Wir suchen Folgen $(a_n)_{n\in\mathbb{N}}$ reeller Zahlen, für die gilt

$$a_{n+2} = a_{n+1} + a_n \qquad \text{für alle } n \in \mathbb{N}.$$

Die Folge $(0, 1, 1, 2, 3, 5, 8, \ldots)$ der Fibonacci-Zahlen ist eine Lösung, aber es gibt noch mehr. Alle Lösungen lassen sich darstellen in der Form

$$a_n = a\Big(\frac{1+\sqrt{5}}{2}\Big)^n + b\Big(\frac{1-\sqrt{5}}{2}\Big)^n \qquad \text{mit } a, b \in \mathbb{R}.$$

(3) Wir suchen (zweimal differenzierbare) Funktionen $f\colon \mathbb{R} \to \mathbb{R}$, für die gilt

$$f''(x) + f(x) = 0 \qquad \text{für alle } x \in \mathbb{R}.$$

Hier sind die Lösungen gegeben durch

$$f(x) = a\cos x + b\sin x \qquad \text{mit } a, b \in \mathbb{R}. \qquad \clubsuit$$

Obwohl die betrachteten Objekte ganz unterschiedlich sind (Quadrupel von reellen Zahlen, Folgen reeller Zahlen, zweimal differenzierbare reelle Funktionen), ist die Struktur der Lösungsmenge in allen drei Fällen sehr ähnlich. Dass dies so sein muss, ist ein allgemeines Resultat über lineare Gleichungen. Etwas konkreter äußert sich die Linearität darin, dass die *Summe* zweier Lösungen wieder eine Lösung ist, und dass *Vielfache* einer Lösung wieder Lösungen sind. Diese beiden Operationen, also Addition und Vervielfachung, d.h. Multiplikation mit einem „Skalar“ (in den Beispielen ist das jeweils eine reelle Zahl), ergeben die lineare Struktur, die in der folgenden Definition formalisiert ist.

* 5.2. **Definition.** Sei K ein Körper. Ein *K-Vektorraum* oder *Vektorraum über K* oder *linearer Raum über K* ist ein Quintupel $(V, +, \mathbf{0}, -, \cdot)$, bestehend aus einer Menge V, einer Verknüpfung $+\colon V \times V \to V$ (genannt Addition), einem Element $\mathbf{0} \in V$, einer Abbildung $-\colon V \to V$ und einer Abbildung $\cdot\colon K \times V \to V$ (Skalarmultiplikation), sodass $(V, +, \mathbf{0}, -)$ eine kommutative Gruppe ist und die folgenden weiteren Bedingungen („Axiome“) erfüllt sind: **DEF** Vektorraum

(1) $\forall v \in V\colon 1 \cdot v = v$ (hier ist $1 \in K$ das Einselement des Körpers K).

(2) (Assoziativität der Skalarmultiplikation)
$\forall \lambda, \mu \in K\ \forall v \in V\colon \lambda \cdot (\mu \cdot v) = (\lambda\mu) \cdot v$.

(3) (Distributivgesetze)
$\forall \lambda, \mu \in K\ \forall v \in V\colon (\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v$ und
$\forall \lambda \in K\ \forall v, w \in V\colon \lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w$.

Statt $\lambda \cdot v$ schreibt man oft kurz λv. Die Elemente eines Vektorraums werden auch *Vektoren* genannt. $\mathbf{0} \in V$ heißt der *Nullvektor* des Vektorraums V.

Ein $\mathbb{R}$-Vektorraum heißt auch *reeller Vektorraum*, ein $\mathbb{C}$-Vektorraum *komplexer Vektorraum*. ♢

Machen Sie sich klar, dass „+“ in diesen Axiomen zwei verschiedene Bedeutungen hat: Es kann die Addition im Körper K gemeint sein oder die Addition im Vektorraum V!

Der Vollständigkeit halber und zur Erinnerung sind hier noch einmal die vier Axiome für eine kommutative Gruppe $(V, +, \mathbf{0}, -)$ angegeben:

(1) (Assoziativität der Addition) $\forall v_1, v_2, v_3 \in V\colon (v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$.

(2) (Kommutativität der Addition) $\forall v, w \in V\colon v + w = w + v$.

(3) (Nullelement) $\forall v \in V\colon v + \mathbf{0} = v$.

(4) (Negative Elemente) $\forall v \in V\colon v + (-v) = \mathbf{0}$.

Wir kürzen $v + (-w)$ zu $v - w$ ab.

Wie üblich kann man sich auf die Angabe von Addition und Skalarmultiplikation beschränken, da das Nullelement und die Negation eindeutig bestimmt sind. Wenn die Verknüpfungen aus dem Kontext klar sind, spricht man einfach nur vom „K-Vektorraum V“; wenn auch der Körper K aus dem Kontext klar ist, vom „Vektorraum V“.

Wir kommen zu einigen einfachen Eigenschaften.

LEMMA
Rechenregeln
Vektorraum

5.3. **Lemma.** *Sei $(V, +, \mathbf{0}, -, \cdot)$ ein K-Vektorraum. Dann gilt:*

(1) $\forall v \in V : 0 \cdot v = \mathbf{0}$.

(2) $\forall \lambda \in K : \lambda \cdot \mathbf{0} = \mathbf{0}$.

(3) $\forall v \in V : (-1) \cdot v = -v$.

(4) $\forall \lambda \in K \; \forall v \in V : \lambda \cdot v = \mathbf{0} \iff \lambda = 0 \quad \textit{oder} \quad v = \mathbf{0}$.

Beweis.

(1) Es ist (mit einem der beiden Distributivgesetze)

$$0 \cdot v + 0 \cdot v = (0 + 0) \cdot v = 0 \cdot v$$

und Addition von $-(0 \cdot v)$ auf beiden Seiten liefert

$$0 \cdot v = 0 \cdot v + 0 \cdot v - 0 \cdot v = 0 \cdot v - 0 \cdot v = \mathbf{0}\,.$$

(2) Das geht analog unter Verwendung des anderen Distributivgesetzes:

$$\mathbf{0} = \lambda \cdot \mathbf{0} - \lambda \cdot \mathbf{0} = \lambda \cdot (\mathbf{0} + \mathbf{0}) - \lambda \cdot \mathbf{0} = \lambda \cdot \mathbf{0} + \lambda \cdot \mathbf{0} - \lambda \cdot \mathbf{0} = \lambda \cdot \mathbf{0}\,.$$

(3) Es gilt

$$v + (-1) \cdot v = 1 \cdot v + (-1) \cdot v = (1 + (-1)) \cdot v = 0 \cdot v = \mathbf{0}\,,$$

also muss $(-1) \cdot v$ das eindeutig bestimmte Negative $-v$ von v sein.

(4) Seien $\lambda \in K$ und $v \in V$. Die Implikation „$\Leftarrow$“ wurde bereits in den ersten beiden Teilen des Lemmas bewiesen. Es gelte also $\lambda \cdot v = \mathbf{0}$. Ist $\lambda = 0$, dann gilt die rechte Seite. Anderenfalls gibt es $\lambda^{-1} \in K$ und es folgt (mit Teil (2) und der Assoziativität der Skalarmultiplikation)

$$\mathbf{0} = \lambda^{-1} \cdot \mathbf{0} = \lambda^{-1} \cdot (\lambda \cdot v) = (\lambda^{-1}\lambda) \cdot v = 1 \cdot v = v\,. \qquad \square$$

Hier sind einige Beispiele von Vektorräumen:

BSP
Vektorräume

5.4. **Beispiele.** Sei K ein Körper.

(1) Der kleinste K-Vektorraum besteht nur aus dem Nullvektor: $V = \{\mathbf{0}\}$ und es gilt $\mathbf{0} + \mathbf{0} = \mathbf{0}$ und $\lambda \cdot \mathbf{0} = \mathbf{0}$ für alle $\lambda \in K$. Dieser Vektorraum heißt der *Null-Vektorraum*. Er ist als Vektorraum nicht besonders interessant, spielt aber in der Linearen Algebra eine ähnliche Rolle wie die leere Menge in der Mengenlehre.

(2) Das nächste Beispiel ist der Körper K selbst mit seiner Addition und Multiplikation. Die Vektorraum-Axiome entsprechen einem Teil der Körper-Axiome.

(3) Sehr wichtig ist die folgende Klasse von Beispielen, denn es sind die Standardbeispiele für K-Vektorräume. Als Menge nimmt man K^n, die Menge der n-Tupel von Elementen von K, und die Verknüpfungen definiert man „komponentenweise“:

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \qquad \text{und}$$
$$\lambda \cdot (x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)\,.$$

Dann kann man die Axiome leicht nachprüfen. Wir führen das hier exemplarisch für eines der Distributivgesetze durch:

$$\begin{aligned}\lambda \cdot \big((x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n)\big) &= \lambda \cdot (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)\\ &= \big(\lambda(x_1 + y_1), \lambda(x_2 + y_2), \ldots, \lambda(x_n + y_n)\big)\\ &= (\lambda x_1 + \lambda y_1, \lambda x_2 + \lambda y_2, \ldots, \lambda x_n + \lambda y_n)\\ &= (\lambda x_1, \lambda x_2, \ldots, \lambda x_n) + (\lambda y_1, \lambda y_2, \ldots, \lambda y_n)\\ &= \lambda \cdot (x_1, x_2, \ldots, x_n) + \lambda \cdot (y_1, y_2, \ldots, y_n)\,.\end{aligned}$$

Man sieht, dass das direkt aus dem Distributivgesetz $\lambda(x + y) = \lambda x + \lambda y$ von K folgt. Für die übrigen Axiome geht das ganz analog. In diesem Beispiel sind die beiden vorigen Beispiele als Grenzfälle enthalten: Für $n = 0$ hat die Menge K^0 nur ein Element (das Nulltupel, das keine Komponenten hat) und ist somit ein Null-Vektorraum. Für $n = 1$ ist $K^1 = K$ und man bekommt K als Vektorraum über K. Für $K = \mathbb{R}$ und $K = \mathbb{C}$ haben wir den reellen Vektorraum $\mathbb{R}^n$ und den komplexen Vektorraum $\mathbb{C}^n$ für jedes $n \in \mathbb{N}$.

(4) Man kann das vorige Beispiel noch verallgemeinern: K^n kann als der Spezialfall $I = \{1, 2, \ldots, n\}$ der Menge K^I der Familien von Elementen von K mit Indexmenge I aufgefasst werden. (Zur Erinnerung: Familien $(x_i)_{i\in I}$ mit $x_i \in K$ sind nur eine andere Schreibweise für Abbildungen $I \to K$.) Man macht K^I zu einem K-Vektorraum, indem man Addition und Skalarmultiplikation wieder komponentenweise definiert:

$$(x_i)_{i\in I} + (y_i)_{i\in I} = (x_i + y_i)_{i\in I} \qquad \text{und}$$
$$\lambda \cdot (x_i)_{i\in I} = (\lambda x_i)_{i\in I}\,.$$

Wenn man statt mit Familien mit Abbildungen $I \to K$ arbeitet, dann sieht das so aus (in diesem Fall sagt man auch „punktweise“ statt „komponentenweise“):

$$f + g\colon I \longrightarrow K\,, \quad i \longmapsto f(i) + g(i)\,, \qquad \text{d.h.} \quad (f + g)(i) = f(i) + g(i) \qquad \text{und}$$
$$\lambda \cdot f\colon I \longrightarrow K\,, \quad i \longmapsto \lambda f(i)\,, \qquad \text{d.h.} \quad (\lambda \cdot f)(i) = \lambda f(i).$$

Das Nachprüfen der Axiome funktioniert im Wesentlichen genauso wie für die n-Tupel. Als Beispiel hier das andere Distributivgesetz (in der Abbildungs-Schreibweise): Seien $\lambda, \mu \in K$ und $f\colon I \to K$ eine Abbildung. Dann gilt für $i \in I$:

$$\begin{aligned}\big((\lambda + \mu) \cdot f\big)(i) &= (\lambda + \mu) f(i) = \lambda f(i) + \mu f(i)\\ &= (\lambda \cdot f)(i) + (\mu \cdot f)(i) = (\lambda \cdot f + \mu \cdot f)(i)\,,\end{aligned}$$

also folgt $(\lambda + \mu) \cdot f = \lambda \cdot f + \mu \cdot f$. Zum Beispiel können wir den reellen Vektorraum $\mathbb{R}^{\mathbb{R}} = \mathrm{Abb}(\mathbb{R}, \mathbb{R})$ aller reellen Funktionen betrachten oder den Vektorraum $\mathbb{R}^{\mathbb{N}}$ aller Folgen reeller Zahlen.

(5) Ein auf den ersten Blick ganz anders gearteter Vektorraum ist der folgende: Sei X eine Menge. Dann definieren wir eine Addition auf der Potenzmenge $\mathcal{P}(X)$ durch

$$A + B = A \triangle B = (A \setminus B) \cup (B \setminus A)$$

(symmetrische Differenz, siehe Beispiel 3.12) und eine Skalarmultiplikation mit Elementen des Körpers $\mathbb{F}_2 = \{0, 1\}$ in der einzig möglichen Form, nämlich durch $0 \cdot A = \mathbf{0} = \emptyset$ und $1 \cdot A = A$. Dann erhält man einen $\mathbb{F}_2$-Vektorraum. Man kann die Axiome wieder nachrechnen, aber man tut

sich etwas leichter, wenn man sich klar macht, dass die Potenzmenge $\mathcal{P}(X)$ und die Menge $\mathbb{F}_2^X$ der Abbildungen $X \to \mathbb{F}_2$ einander bijektiv entsprechen durch

$$\mathcal{P}(X) \longrightarrow \mathbb{F}_2^X\,, \qquad A \longmapsto \left(x \mapsto \left\{\begin{matrix} 0 & \text{falls } x \notin A, \\ 1 & \text{falls } x \in A, \end{matrix}\right\}\right)$$

$$\mathbb{F}_2^X \longrightarrow \mathcal{P}(X)\,, \qquad f \longmapsto \{x \in X \mid f(x) = 1\}\,.$$

Dann entsprechen sich auch Addition und Skalarmultiplikation auf beiden Seiten, also folgt die Gültigkeit der Axiome für $\mathcal{P}(X)$ aus ihrer Gültigkeit für $\mathbb{F}_2^X$.

(6) Der Körper $\mathbb{C}$ der komplexen Zahlen ist ein reeller Vektorraum: Die Addition ist die von $\mathbb{C}$, die Skalarmultiplikation ist die Multiplikation von $\mathbb{C}$, eingeschränkt auf $\mathbb{R} \times \mathbb{C}$. (Die *Einschränkung* einer Abbildung $f\colon X \to Y$ auf eine Teilmenge $T \subset X$ ist die Abbildung $f|_T\colon T \to Y$, $x \mapsto f(x)$, bei der der Definitionsbereich verkleinert (also eingeschränkt) wird, die Abbildungsvorschrift aber unverändert bleibt.) Wenn wir $\mathbb{C}$ als $\mathbb{R}^2$ betrachten, dann ist das derselbe reelle Vektorraum wie in (3) oben mit $K = \mathbb{R}$ und $n = 2$. ♣

DEF Einschränkung

Weitere Beispiele von Vektorräumen erhalten wir als *Untervektorräume* von anderen Vektorräumen; das werden wir im nächsten Abschnitt genauer betrachten.

In den Beispielen 5.1 für lineare Gleichungen vom Beginn dieses Abschnitts sind Lösungen in gewissen reellen Vektorräumen gesucht: Im ersten Beispiel in $\mathbb{R}^4$, im zweiten Beispiel in $\mathbb{R}^{\mathbb{N}}$ und im dritten Beispiel in einem Untervektorraum von $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$.

6. Untervektorräume

Häufig möchte man, wenn man einen Vektorraum V gegeben hat, nicht mit dem ganzen Vektorraum arbeiten, sondern mit einer Teilmenge. Damit stellt sich die Frage, wann so eine Teilmenge (wenn man die Addition und Skalarmultiplikation darauf einschränkt) selbst wieder ein Vektorraum ist. Damit diese Frage sinnvoll ist, müssen die Addition und Skalarmultiplikation auf der Teilmenge *wohldefiniert* sein, das heißt, dass Summen und Vielfache von Elementen der Teilmenge wieder in der Teilmenge liegen müssen. Außerdem brauchen wir natürlich das Nullelement. Das führt auf folgende Definition:

* 6.1. **Definition.** Seien K ein Körper, V ein K-Vektorraum und $U \subset V$ eine Teilmenge von V. Dann heißt U ein *Untervektorraum* oder *linearer Unterraum* von V, wenn U die folgenden Bedingungen erfüllt: **DEF** Untervektorraum

(1) $\mathbf{0} \in U$,

(2) $\forall u_1, u_2 \in U\colon u_1 + u_2 \in U$
(„U ist abgeschlossen unter der Addition“),

(3) $\forall \lambda \in K\ \forall u \in U\colon \lambda \cdot u \in U$
(„U ist abgeschlossen unter der Skalarmultiplikation“). ♢

Wir zeigen gleich, dass diese Definition sinnvoll ist.

6.2. **Lemma.** *Sei K ein Körper, V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Dann gilt für alle $u \in U$, dass auch $-u$ ein Element von U ist.* **LEMMA** Unter-VR ist Vektorraum

Wir schreiben $+_U$ für die auf U (auch im Wertebereich) eingeschränkte Addition $+_U\colon U \times U \to U$, $(u_1, u_2) \mapsto u_1 + u_2$, $-_U$ für die auf U eingeschränkte Negationsabbildung $-_U\colon U \to U$, $u \mapsto -u$, und $\cdot_U$ für die auf U eingeschränkte Skalarmultiplikation $\cdot_U\colon K \times U \to U$, $(\lambda, u) \mapsto \lambda \cdot u$. Dann ist $(U, +_U, \mathbf{0}, -_U, \cdot_U)$ ein K-Vektorraum.

Beweis. Die erste Behauptung ist $\forall u \in U\colon -u \in U$. Das folgt aber aus der Definition, denn $-u = (-1) \cdot u$, vgl. Lemma 5.3. Deshalb und nach der Definition können wir $+_U$, $-_U$ und $\cdot_U$ wie angegeben definieren (denn die Bilder liegen jeweils in U). Es bleiben die Vektorraum-Axiome für U nachzuprüfen. Diese haben aber alle die Form von „Allaussagen“, es wird also verlangt, dass eine Aussage für alle Elemente $u_1, u_2, \dots$ von U gilt. Da V ein Vektorraum ist, gelten diese Aussagen aber sogar für alle Elemente von V, also erst recht für alle Elemente von U. ❑

In der Literatur finden Sie meistens eine Definition von „Vektorraum“ (und analog für Gruppen, Ringe, Körper, ...), die von dem Tripel $(V, +, \cdot)$ ausgeht und dann die *Existenz* eines Nullelements und von Inversen bezüglich der Addition fordert. Im Gegensatz dazu haben wir hier das Nullelement und die Negationsabbildung mit in die „Daten“ des Vektorraums aufgenommen. Der Vorteil ist, dass die Axiome dann alle zu Allaussagen werden, die man leichter nachprüfen kann, wie im obigen Beweis. Auf der anderen Seite muss man sich aber vorher überlegen, was das Nullelement ist und wie die Negationsabbildung aussieht. Im gerade bewiesenen Lemma geschieht dies dadurch, dass wir zeigen, dass U auch unter der Negation abgeschlossen ist, sodass wir die Negationsabbildung $-_U$ definieren können. Wenn man die andere Formulierung der Axiome benutzt, dann muss man diesen Beweisschritt ausführen, wenn man die Existenz des zu u negativen Elements zeigt. Im Endeffekt muss man also das Gleiche tun, nur die Reihenfolge ist etwas anders.

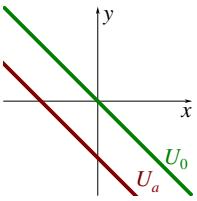
Die Schreibweise $+_U$ usw. für die auf U eingeschränkten Abbildungen diente nur der Verdeutlichung für die Formulierung des Lemmas. Wir schreiben normalerweise einfach $+$ usw. für die Addition usw. auf U.

6.3. **Beispiele.** Jeder Vektorraum V hat die Untervektorräume $U = \{\mathbf{0}\} \subset V$ (ein Null-Vektorraum) und $U = V$. ♣

BSP triviale Unter-VR

6.4. **Beispiel.** Sei $a \in \mathbb{R}$. Wir betrachten den reellen Vektorraum $V = \mathbb{R}^2$ und setzen $U_a = \{(x, y) \in \mathbb{R}^2 \mid x + y = a\}$. Für welche a ist U_a ein Untervektorraum von $\mathbb{R}^2$?

BSP Unter-VR von $\mathbb{R}^2$



Dazu müssen wir die Bedingungen in der Definition nachprüfen. Die erste davon sagt, dass der Nullvektor $\mathbf{0} = (0, 0)$ ein Element von U_a sein muss. Das bedeutet $0 + 0 = a$, also ist das nur für $a = 0$ möglich. Wir prüfen die beiden anderen Bedingungen:

- U_0 ist abgeschlossen unter der Addition, denn für Elemente $u_1 = (x_1, y_1)$ und $u_2 = (x_2, y_2)$ von U_0 gilt $u_1 + u_2 = (x_1 + x_2, y_1 + y_2)$ und
$$(x_1 + x_2) + (y_1 + y_2) = (x_1 + y_1) + (x_2 + y_2) = 0 + 0 = 0\,,$$
also ist $u_1 + u_2 \in U_0$.
- U_0 ist abgeschlossen unter der Skalarmultiplikation, denn für $u = (x, y) \in U_0$ und $\lambda \in \mathbb{R}$ ist $\lambda \cdot u = (\lambda x, \lambda y)$ und es gilt
$$\lambda x + \lambda y = \lambda(x + y) = \lambda \cdot 0 = 0\,,$$
also ist $\lambda \cdot u \in U_0$. ♣

Weitere interessante Beispiele sind „Folgenräume“ und „Funktionenräume“, die als Untervektorräume des Vektorraums $\mathbb{R}^{\mathbb{N}}$ der Folgen reeller Zahlen oder des Vektorraums $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ der reellen Funktionen auftreten.

6.5. **Beispiele.** Sei $V = \mathbb{R}^{\mathbb{N}}$ der reelle Vektorraum, dessen Elemente alle Folgen reeller Zahlen sind.

BSP Folgenräume

(1) Sei $U_b = \{(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid (a_n)_{n\in\mathbb{N}} \text{ ist beschränkt}\}$. Dann ist U_b ein Untervektorraum von $\mathbb{R}^{\mathbb{N}}$.

Beweis. Wir prüfen die Bedingungen nach. Die konstante Nullfolge (mit $a_n = 0$ für alle $n \in \mathbb{N}$) ist beschränkt, also gilt $\mathbf{0} \in U_b$. Seien $(a_n)_{n\in\mathbb{N}}$ und $(b_n)_{n\in\mathbb{N}}$ zwei beschränkte Folgen. Dann gibt es $A, B \in \mathbb{R}$ mit $|a_n| \le A$ und $|b_n| \le B$ für alle $n \in \mathbb{N}$. Es folgt $|a_n + b_n| \le A + B$, also ist auch die Summenfolge $(a_n)_{n\in\mathbb{N}} + (b_n)_{n\in\mathbb{N}} = (a_n + b_n)_{n\in\mathbb{N}}$ beschränkt. Ist zusätzlich $\lambda \in \mathbb{R}$, dann gilt $|\lambda a_n| \le |\lambda| A$, also ist auch die Folge $\lambda \cdot (a_n)_{n\in\mathbb{N}} = (\lambda a_n)_{n\in\mathbb{N}}$ beschränkt. ❑

(2) Sei $U_n = \{(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid (a_n)_{n\in\mathbb{N}} \text{ ist eine Nullfolge}\}$. Dann ist U_n ein Untervektorraum von $\mathbb{R}^{\mathbb{N}}$ (oder auch von U_b). (Übung.)

(3) Sei $U_k = \{(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid (a_n)_{n\in\mathbb{N}} \text{ konvergiert}\}$. Dann ist U_k ein Untervektorraum von $\mathbb{R}^{\mathbb{N}}$ (oder auch von U_b).

Beweis. Die konstante Nullfolge konvergiert (gegen 0), also ist sie in U_k. In der Analysis lernen Sie, dass die Summe zweier konvergenter Folgen wieder konvergiert und dass jedes Vielfache einer konvergenten Folge konvergiert. Damit sind die drei Bedingungen erfüllt. ❑

Für diese drei Untervektorräume gilt $U_n \subset U_k \subset U_b$ (denn jede Nullfolge konvergiert gegen 0 und jede konvergente Folge ist beschränkt, vgl. Analysis). ♣

BSP Funktionenräume

6.6. **Beispiele.** Sei $V = \mathrm{Abb}(\mathbb{R}, \mathbb{R})$ der reelle Vektorraum, dessen Elemente alle Funktionen $\mathbb{R} \to \mathbb{R}$ sind.

(1) Sei $\mathcal{C}(\mathbb{R}) = \{f \in \mathrm{Abb}(\mathbb{R}, \mathbb{R}) \mid f \text{ ist stetig}\}$. Dann ist $\mathcal{C}(\mathbb{R})$ ein Untervektorraum von V.

Beweis. Die Nullfunktion $x \mapsto 0$ ist stetig. In der Analysis lernen Sie, dass Summen und Vielfache stetiger Funktionen wieder stetig sind. ❑

(2) Sei $n \in \mathbb{N}$ und
$\mathcal{C}^n(\mathbb{R}) = \{f \in \mathrm{Abb}(\mathbb{R}, \mathbb{R}) \mid f \text{ ist } n\text{-mal differenzierbar und } f^{(n)} \text{ ist stetig}\}$
der Raum der n-mal stetig differenzierbaren Funktionen. Aus Ergebnissen der Analysis folgt, dass $\mathcal{C}^n(\mathbb{R})$ ein Untervektorraum von V ist.

(3) Sei $a > 0$ und $\mathcal{P}(a) = \{f \in \mathrm{Abb}(\mathbb{R}, \mathbb{R}) \mid \forall x \in \mathbb{R}\colon f(x + a) = f(x)\}$ die Menge der periodischen Funktionen mit Periode a (zum Beispiel sind sin und cos Elemente von $\mathcal{P}(2\pi)$). Dann ist $\mathcal{P}(a)$ ein Untervektorraum von V.

Beweis. Die Nullfunktion ist periodisch, also ein Element von $\mathcal{P}(a)$. Seien $f, g \in \mathcal{P}(a)$ und $\lambda \in \mathbb{R}$. Wir zeigen $f + g, \lambda f \in \mathcal{P}(a)$: Für alle $x \in \mathbb{R}$ gilt

$$(f+g)(x+a) = f(x+a) + g(x+a) \overset{f,g\in\mathcal{P}(a)}{=} f(x) + g(x) = (f+g)(x) \qquad \text{und}$$

$$(\lambda f)(x+a) = \lambda f(x+a) \overset{f\in\mathcal{P}(a)}{=} \lambda f(x) = (\lambda f)(x)\,.$$

Damit sind alle drei Bedingungen erfüllt. ❑

♣

Auch in der *Codierungstheorie* spielt der Begriff des Untervektorraums eine sehr wichtige Rolle.

BSP Lineare Codes

6.7. **Beispiel.** Seien F ein endlicher Körper (zum Beispiel $F = \mathbb{F}_2$) und $n \in \mathbb{N}$. Dann heißt ein Untervektorraum von F^n ein *linearer Code* der Länge n über F. Ein Beispiel ist der *Hamming-Code* der Länge 7 über $\mathbb{F}_2$, der gegeben ist durch

$$H = \{(x_1, x_2, x_3, x_4, x_1{+}x_2{+}x_4, x_1{+}x_3{+}x_4, x_2{+}x_3{+}x_4) \in \mathbb{F}_2^7 \mid x_1, x_2, x_3, x_4 \in \mathbb{F}_2\}\,.$$

In der Codierungstheorie interessiert man sich dann für die „Größe“ (genauer: die *Dimension*, die wir bald einführen werden) des Codes und dafür, wie viele Fehler er korrigieren kann. Dafür ist wichtig, dass je zwei verschiedene Codewörter (also Elemente des Codes) sich an möglichst vielen Stellen unterscheiden. Wegen der linearen Struktur kann man Differenzen bilden und daher annehmen, dass eines der Codewörter null ist. Dann ist die Frage, an mindestens wie vielen Stellen ein von $\mathbf{0}$ verschiedenes Codewort eine von 0 verschiedene Komponente hat. Für den Hamming-Code H ist diese „Minimaldistanz“ 3, was bedeutet, dass er „einen Fehler korrigieren“ kann. (Wenn ein Codewort an einer Stelle verändert wird, kann man es rekonstruieren, da sich jedes andere Codewort von dem veränderten Wort an mindestens zwei Stellen unterscheidet.) ♣

7. Erzeugendensysteme

Wir erinnern uns an die Beispiele von Funktionenräumen im letzten Abschnitt. Dort hatten wir gesehen, dass der Raum $\mathcal{C}(\mathbb{R})$ der stetigen reellen Funktionen und der Raum $\mathcal{P}(a)$ der a-periodischen reellen Funktionen beides Untervektorräume von $\mathrm{Abb}(\mathbb{R},\mathbb{R})$ sind. Wie sieht es mit stetigen periodischen Funktionen aus? Muss $\mathcal{C}(\mathbb{R})\cap\mathcal{P}(a)$ auch ein Untervektorraum sein?

LEMMA Schnitt von zwei UVR

7.1. **Lemma.** *Sei V ein K-Vektorraum mit zwei Untervektorräumen U_1 und U_2. Dann ist $U_1\cap U_2$ ebenfalls ein Untervektorraum von V.*

Beweis. Wir müssen die drei Bedingungen aus der Definition von „Untervektorraum“ nachprüfen.

(1) Da U_1 und U_2 Untervektorräume sind, ist $\mathbf{0}\in U_1$ und $\mathbf{0}\in U_2$, also auch $\mathbf{0}\in U_1\cap U_2$.

(2) Seien $u,u'\in U_1\cap U_2$. Dann gilt $u,u'\in U_1$ und $u,u'\in U_2$. Da U_1 und U_2 Untervektorräume sind, folgt $u+u'\in U_1$ und $u+u'\in U_2$, also auch $u+u'\in U_1\cap U_2$.

(3) Seien $\lambda\in K$ und $u\in U_1\cap U_2$. Dann ist $u\in U_1$ und $u\in U_2$. Da U_1 und U_2 Untervektorräume sind, folgt $\lambda u\in U_1$ und $\lambda u\in U_2$, also auch $\lambda u\in U_1\cap U_2$. ❑

Wir wollen diese Aussage jetzt auf Durchschnitte von beliebig (auch unendlich) vielen Untervektorräumen verallgemeinern. Dazu führen wir erst eine Schreibweise für Vereinigungen und Durchschnitte von vielen Mengen ein.

DEF
$\bigcup_{i\in I} A_i$
$\bigcap_{i\in I} A_i$
$\bigcup \mathcal{M}$
$\bigcap \mathcal{M}$

7.2. **Definition.** Ist $(A_i)_{i\in I}$ eine Familie von Mengen, dann schreiben wir

$$\bigcup_{i\in I} A_i = \{x \mid \exists i\in I\colon x\in A_i\}$$

für die Vereinigung aller Mengen A_i. (Ist I die leere Menge, dann ist diese Vereinigung ebenfalls leer.) Ist $I\neq\emptyset$, dann schreiben wir analog

$$\bigcap_{i\in I} A_i = \{x \mid \forall i\in I\colon x\in A_i\}$$

für den Durchschnitt aller Mengen A_i.

Ist $\mathcal{M}$ eine Menge, deren Elemente selbst Mengen sind, dann schreiben wir

$$\bigcup\mathcal{M} = \bigcup_{A\in\mathcal{M}} A = \{x \mid \exists A\in\mathcal{M}\colon x\in A\}$$

für die Vereinigung all dieser Mengen und, falls $\mathcal{M}$ nicht leer ist,

$$\bigcap\mathcal{M} = \bigcap_{A\in\mathcal{M}} A = \{x \mid \forall A\in\mathcal{M}\colon x\in A\}$$

für ihren Durchschnitt. ♢

Im Fall $I=\emptyset$ wäre die Bedingung $\forall i\in I\colon x\in A_i$ für alle x erfüllt und man bekäme die Menge, die alles enthält. Diese Menge kann es aber nicht geben, denn sie würde die Menge enthalten, die zur Russellschen Antinomie führt, siehe die Bemerkungen zur Mengenlehre am Ende von Abschnitt 2.

Damit können wir jetzt die Verallgemeinerung formulieren:

LEMMA Durchschnitt von Unter-VR

7.3. **Lemma.** *Sei V ein K-Vektorraum und sei $(U_i)_{i\in I}$ eine Familie von Untervektorräumen von V mit $I \neq \emptyset$. Dann ist*

$$U = \bigcap_{i\in I} U_i$$

ebenfalls ein Untervektorraum von V.

Für $I = \{1, 2, 3, \ldots, n\}$ (mit $n \geq 1$) haben wir den Spezialfall

$$U_1, U_2, \ldots, U_n \subset V \text{ Untervektorräume} \implies U_1 \cap U_2 \cap \ldots \cap U_n \subset V \text{ Untervektorraum.}$$

Beweis. Wir müssen die Bedingungen aus Definition 6.1 für U nachprüfen.

(1) Da jede Teilmenge U_i ein Untervektorraum von V ist, gilt $\forall i \in I\colon \mathbf{0} \in U_i$. Das bedeutet gerade $\mathbf{0} \in U$.

(2) Seien $u_1, u_2 \in U$. Nach Definition von U bedeutet das $\forall i \in I\colon u_1, u_2 \in U_i$. Da alle U_i Untervektorräume von V sind, folgt $\forall i \in I\colon u_1 + u_2 \in U_i$, also $u_1 + u_2 \in U$.

(3) Sei $\lambda \in K$ und $u \in U$. Dann gilt $\forall i \in I\colon u \in U_i$. Da alle U_i Untervektorräume von V sind, folgt $\forall i \in I\colon \lambda u \in U_i$, also $\lambda u \in U$. ❑

BSP

7.4. **Beispiel.** Der Raum

$$\mathcal{C}(\mathbb{R}) \cap \mathcal{P}(a) = \{f \in \mathrm{Abb}(\mathbb{R}, \mathbb{R}) \mid f \text{ ist stetig und } a\text{-periodisch}\}$$

ist ein Untervektorraum von $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$. ♣

Wie sieht es mit Vereinigungen von Untervektorräumen aus? Im Allgemeinen erhält man daraus *keinen* Untervektorraum. Die Vereinigung von zwei Untervektorräumen U_1 und U_2 zum Beispiel ist nur dann wieder ein Untervektorraum, wenn einer der beiden im anderen enthalten ist (Übung). Man hat aber immerhin das folgende Resultat.

LEMMA aufsteigende Vereinigung von Unter-VR

7.5. **Lemma.** *Sei V ein K-Vektorraum und sei $(U_n)_{n\in\mathbb{N}}$ eine aufsteigende Folge von Untervektorräumen von V (d.h. $U_n \subset U_{n+1}$ für alle $n \in \mathbb{N}$). Dann ist*

$$U = \bigcup_{n\in\mathbb{N}} U_n$$

ebenfalls ein Untervektorraum von V.

Beweis. Wir prüfen die Bedingungen für U.

(1) $\mathbf{0} \in U_0$, also ist auch $\mathbf{0} \in U$.

(2) Seien $u_1, u_2 \in U$. Dann gibt es $n_1, n_2 \in \mathbb{N}$ mit $u_1 \in U_{n_1}$ und $u_2 \in U_{n_2}$. Sei n die größere der beiden Zahlen n_1 und n_2. Da wir eine aufsteigende Folge von Untervektorräumen haben, gilt dann $U_{n_1} \subset U_n$ und $U_{n_2} \subset U_n$ und damit $u_1, u_2 \in U_n$. Da U_n ein Untervektorraum ist, folgt $u_1 + u_2 \in U_n \subset U$.

(3) Sei $\lambda \in K$ und $u \in U$. Dann gibt es $n \in \mathbb{N}$, sodass $u \in U_n$ ist. Da U_n ein Untervektorraum ist, folgt $\lambda u \in U_n \subset U$. ❑

Lemma 7.3 erlaubt es uns nun, den kleinsten Untervektorraum zu konstruieren, der eine gegebene Teilmenge eines Vektorraums V enthält.

∗ 7.6. **Definition.** Sei V ein K-Vektorraum und $A \subset V$ eine beliebige Teilmenge von V. Dann heißt der Untervektorraum **DEF** Lineare Hülle

$$\langle A\rangle = \langle A\rangle_K = \bigcap\{U \subset V \mid U \text{ Untervektorraum von } V \text{ und } A \subset U\}$$

(also der Durchschnitt aller A enthaltenden Untervektorräume von V) der *von A erzeugte* oder *aufgespannte* Untervektorraum von V, die *(K-)lineare Hülle von A* oder der *(K-)Spann von A*. Ist $A = \{v_1, v_2, \ldots, v_n\}$ endlich, dann schreiben wir auch

$$\langle v_1, v_2, \ldots, v_n\rangle \qquad \text{oder} \qquad \langle v_1, v_2, \ldots, v_n\rangle_K$$

statt $\langle A\rangle$ oder $\langle A\rangle_K$.

Wir werden statt Mengen häufiger auch Familien $(v_i)_{i\in I}$ von Elementen von V betrachten. Wir schreiben dann

$$\langle (v_i)_{i\in I}\rangle \qquad \text{oder auch} \qquad \langle v_i \mid i \in I\rangle$$

für den von allen v_i erzeugten Untervektorraum $\langle\{v_i \mid i \in I\}\rangle$. ♢

Lemma 7.3 garantiert uns, dass $\langle A\rangle$ tatsächlich ein Untervektorraum von V ist, denn $\langle A\rangle$ ist definitionsgemäß der Durchschnitt einer nichtleeren Menge von Untervektorräumen (nichtleer, weil V selbst immer ein A enthaltender Untervektorraum von V ist).

Wir benutzen die Schreibweise $\langle A\rangle_K$, um zu verdeutlichen, welcher Körper zugrunde gelegt wird. Zum Beispiel gilt im $\mathbb{R}$-Vektorraum $\mathbb{C}$, dass $\langle 1\rangle = \langle 1\rangle_{\mathbb{R}} = \mathbb{R}$ ist. Wird $\mathbb{C}$ aber als $\mathbb{C}$-Vektorraum betrachtet, dann haben wir $\langle 1\rangle = \langle 1\rangle_{\mathbb{C}} = \mathbb{C}$.

7.7. **Beispiel.** In Definition 7.6 können wir für A die leere Menge wählen. Was ist der von A erzeugte Untervektorraum? **BSP** $\langle\emptyset\rangle = \{\mathbf{0}\}$

Da *jeder* Untervektorraum von V die leere Menge enthält, müssen wir den Durchschnitt über *alle* Untervektorräume von V bilden. Da jeder Untervektorraum den Nullvektor enthält und $\{\mathbf{0}\}$ ein Untervektorraum ist, folgt $\langle\emptyset\rangle = \{\mathbf{0}\}$. ♣

∗ 7.8. **Definition.** Sei V ein K-Vektorraum und $E \subset V$ eine Teilmenge von V. Dann heißt E ein *(K-)Erzeugendensystem* von V, wenn $V = \langle E\rangle$ gilt. Analog heißt eine Familie $(v_i)_{i\in I}$ von Elementen von V ein *(K-)Erzeugendensystem* von V, wenn $V = \langle (v_i)_{i\in I}\rangle$ ist. ♢ **DEF** Erzeugendensystem

Zum Beispiel ist die leere Menge ein Erzeugendensystem des Null-Vektorraums.

Definition 7.6 ist sehr elegant, aber nicht besonders praktisch, weil sie uns nicht sagt, „wie die lineare Hülle von A aussieht“, also was ihre Elemente sind. In gewisser Weise ist es eine Definition „von oben“ — wir betrachten alle Untervektorräume, die mindestens so groß sind wie gewünscht, und wählen dann den kleinsten (im Sinne der Inklusion von Mengen) aus. (Das ist übrigens völlig analog zur Definition des Abschlusses $\bar{A}$ einer Menge A als Durchschnitt aller A enthaltenden abgeschlossenen Mengen oder auch zur Definition des Supremums einer Menge reeller Zahlen als kleinste obere Schranke.) Was wir aber gerne hätten, ist eine Definition „von unten“, die die Elemente von $\langle A\rangle$ aus den Elementen von A konstruiert.

Dafür betrachten wir als Beispiel ein Paar (v_1, v_2) von Vektoren in V. Welche Elemente muss $\langle v_1, v_2\rangle$ mindestens enthalten?

Wir wissen, dass v_1 und v_2 Elemente von $\langle v_1, v_2\rangle$ sind, außerdem ist $\langle v_1, v_2\rangle$ ein Untervektorraum, also unter Addition und Skalarmultiplikation abgeschlossen. Es

müssen also insbesondere Summen von Vielfachen von v_1 und v_2 in $\langle v_1, v_2\rangle$ enthalten sein:

$$\{\lambda_1 v_1 + \lambda_2 v_2 \mid \lambda_1, \lambda_2 \in K\} \subset \langle v_1, v_2\rangle\,.$$

Auf der anderen Seite überlegt man sich leicht, dass diese Menge selbst schon ein Untervektorraum von V ist. Da dieser Untervektorraum v_1 und v_2 enthält und gleichzeitig in allen v_1 und v_2 enthaltenden Untervektorräumen enthalten ist, muss er gleich $\langle v_1, v_2\rangle$ sein. (Das ist analog zu unserer Konstruktion des Körpers $\mathbb{C}$: Wir haben erst überlegt, welche Elemente er enthalten muss, und dann gezeigt, dass diese bereits ausreichen, da sie einen Körper bilden.) Diese Beobachtung lässt sich verallgemeinern.

SATZ
Beschreibung von $\langle A\rangle$

7.9. **Satz.** *Sei V ein K-Vektorraum.*

(1) *Sind $v_1, v_2, \ldots, v_n \in V$, dann gilt*

$$\langle v_1, v_2, \ldots, v_n\rangle = \{\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n \mid \lambda_1, \lambda_2, \ldots, \lambda_n \in K\}\,.$$

(2) *Ist $A \subset V$ beliebig, dann gilt*

$$\langle A\rangle = \{\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n \mid n \in \mathbb{N}, v_1, v_2, \ldots, v_n \in A, \lambda_1, \lambda_2, \ldots, \lambda_n \in K\}\,.$$

(3) *Ist $(v_i)_{i\in I}$ eine Familie von Elementen von V, dann gilt*

$$\langle v_i \mid i \in I\rangle = \{\lambda_1 v_{i_1} + \lambda_2 v_{i_2} + \ldots + \lambda_n v_{i_n} \mid n \in \mathbb{N}, i_1, \ldots, i_n \in I, \lambda_1, \ldots, \lambda_n \in K\}\,.$$

Für $n = 0$ setzen wir dabei $\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = \mathbf{0}$ („leere Summe“).

Beweis.

(1) Sei $U = \{\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n \mid \lambda_1, \lambda_2, \ldots, \lambda_n \in K\}$ die Menge auf der rechten Seite der Gleichung. Da $v_1, v_2, \ldots, v_n \in \langle v_1, v_2, \ldots, v_n\rangle$ und $\langle v_1, v_2, \ldots, v_n\rangle$ unter Skalarmultiplikation und Addition abgeschlossen ist, muss jedes Element der Form $\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$ ebenfalls in $\langle v_1, v_2, \ldots, v_n\rangle$ liegen. Es gilt also $U \subset \langle v_1, v_2, \ldots, v_n\rangle$.
Auf der anderen Seite gilt $v_1, v_2, \ldots, v_n \in U$ (wähle $\lambda_j = 1$ und $\lambda_i = 0$ für alle $i \in \{1, 2, \ldots, n\} \setminus \{j\}$, um zu sehen, dass $v_j \in U$ ist) und U ist ein Untervektorraum von V:
- $\mathbf{0} \in U$ (setze $\lambda_i = 0$ für alle i).
- U ist abgeschlossen unter der Addition, denn

$$\begin{aligned}(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n) + (\mu_1 v_1 + \mu_2 v_2 + \ldots + \mu_n v_n)\\ = (\lambda_1 + \mu_1) v_1 + (\lambda_2 + \mu_2) v_2 + \ldots + (\lambda_n + \mu_n) v_n\,.\end{aligned}$$

- U ist abgeschlossen unter der Skalarmultiplikation, denn

$$\lambda(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n) = (\lambda\lambda_1) v_1 + (\lambda\lambda_2) v_2 + \ldots + (\lambda\lambda_n) v_n\,.$$

Da U ein $v_1, v_2, \ldots, v_n$ enthaltender Untervektorraum von V ist, folgt (nach Definition 7.6) $\langle v_1, v_2, \ldots, v_n\rangle \subset U$; insgesamt erhalten wir die behauptete Gleichheit.

(2) Sei wieder U die Menge auf der rechten Seite der Gleichung. Wie in Teil (1) ist klar, dass $U \subset \langle A\rangle$ ist. Es gilt wieder, dass U ein A enthaltender Untervektorraum ist. Die einzige Schwierigkeit tritt beim Nachweis der Abgeschlossenheit unter der Addition auf, denn in den beiden zu addierenden Summen können verschiedene Elemente von A auftreten. Da aber nicht

vorausgesetzt ist, dass die auftretenden Elemente paarweise verschieden[1] sein müssen, können wir die beiden Summen einfach „formal“ addieren:

$$(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n) + (\mu_1 w_1 + \mu_2 w_2 + \ldots + \mu_m w_m) \\ = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n + \lambda_{n+1} v_{n+1} + \ldots + \lambda_{n+m} v_{n+m}\,,$$

wenn wir $\lambda_{n+j} = \mu_j$ und $v_{n+j} = w_j$ setzen für $j \in \{1, 2, \ldots, m\}$.

(3) Das folgt wegen $\langle v_i \mid i \in I\rangle = \langle\{v_i \mid i \in I\}\rangle$ aus der vorigen Aussage. ❑

Es ist eine gute Übung, sich zu überlegen, an welcher Stelle in diesem Beweis welche der Vektorraum-Axiome verwendet werden.

Weil die Ausdrücke der Form $\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$ so wichtig sind, haben sie einen eigenen Namen.

7.10. **Definition.** Seien K ein Körper und V ein K-Vektorraum. **DEF** Linear-kombination

(1) Sind $v_1, v_2, \ldots, v_n \in V$ und $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$, dann heißt

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$$

eine *(K-)Linearkombination von* $v_1, v_2, \ldots, v_n$.

Dabei heißt λ_j der *Koeffizient* von v_j in der Linearkombination.

(2) Ist $A \subset V$ eine beliebige Teilmenge von V, dann heißt jede K-Linearkombination von Elementen $v_1, v_2, \ldots, v_n \in A$ eine *(K-)Linearkombination von Elementen von A.*

(3) Ist $(v_i)_{i\in I}$ eine Familie von Elementen von V, dann heißt jede K-Linearkombination von Vektoren $v_{i_1}, v_{i_2}, \ldots, v_{i_n}$ mit $i_1, i_2, \ldots, i_n \in I$ eine *(K-) Linearkombination der Familie* $(v_i)_{i\in I}$. ♢

Satz 7.9 kann dann so formuliert werden:

Die lineare Hülle von $A \subset V$ besteht genau aus allen Linearkombinationen von Elementen von A. Die entsprechende Aussage gilt für Familien.

Eine Teilmenge $E \subset V$ ist genau dann ein Erzeugendensystem von V, wenn jedes Element von V eine Linearkombination von Elementen von E ist. Die analoge Aussage gilt für Familien als Erzeugendensysteme.

Warnung. In einer Linearkombination kommen immer nur **endlich viele** Elemente vor! In der Linearen Algebra gibt es (im Gegensatz zur Analysis) keine unendlichen Summen!

7.11. **Definition.** Analog zur in der Analysis eingeführten Summenschreibweise schreiben wir **DEF** $\sum_{i\in I} a_i$

$$\sum_{i\in I} a_i \qquad \text{bzw.} \qquad \sum_{i=1}^{n} a_i$$

für die Summe der Glieder der Familie $(a_i)_{i\in I}$ bzw. für die Summe der Komponenten des n-Tupels $(a_1, a_2, \ldots, a_n)$. Dabei sind die a_i aus einer kommutativen Gruppe (bei uns fast immer Elemente eines Vektorraums) und die Menge I ist

[1] „$v_1, v_2, \ldots, v_n$ sind paarweise verschieden“ bedeutet „$\forall i, j \in \{1, 2, \ldots, n\}\colon i \neq j \Rightarrow v_i \neq v_j$“.

endlich. Ist I leer (bzw. $n = 0$), dann ist der Wert dieser „leeren Summe“ das Nullelement der Gruppe. Eine Linearkombination kann dann in der Form

$$\sum_{i=1}^{n} \lambda_i v_i$$

geschrieben werden.

Für unendliche Indexmengen I verlangen wir, dass alle bis auf endlich viele Summanden a_i null sind, und setzen dann

$$\sum_{i\in I} a_i = \sum_{i\in\{j\in I | a_j\neq 0\}} a_i \,;$$

die rechts stehende Summe ist wieder endlich. In dieser Schreibweise sind dann die Linearkombinationen der Familie $(v_i)_{i\in I}$ gegeben durch

$$\sum_{i\in I} \lambda_i v_i \,,$$

wobei $(\lambda_i)_{i\in I}$ eine Familie von Skalaren (Elementen von K) ist, sodass nur endlich viele $\lambda_i \neq 0$ sind. ♢

BSP Erzeugendensystem von K^n

7.12. **Beispiel.** Sei K ein Körper und $n \in \mathbb{N}$. Im Standard-Vektorraum K^n haben wir die Elemente

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \quad \ldots, \quad \mathbf{e}_n = (0, 0, 0, \ldots, 0, 1)\,.$$

Dabei sind alle Komponenten von $\mathbf{e}_j$ null mit Ausnahme der j-ten, die den Wert 1 hat. $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ ist ein Erzeugendensystem von K^n, denn jedes Element von K^n ist eine Linearkombination dieser Elemente:

$$(x_1, x_2, \ldots, x_n) = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n\,. \qquad ♣$$

BSP Viele Erzeugendensysteme

7.13. **Beispiel.** Ein Vektorraum hat im Allgemeinen viele Erzeugendensysteme. Zum Beispiel sind

$$(\mathbf{e}_1, \mathbf{e}_2), \quad \big((1,1),(1,-1)\big), \quad \{(1,2),(2,3),(3,4)\}, \quad \mathbb{Z}\times\mathbb{Z} \quad \text{und} \quad \mathbb{R}\times\mathbb{R}$$

alles Erzeugendensysteme des reellen Vektorraums $V = \mathbb{R}^2$. ♣

BSP Vektorraum der Polynomfunktionen

7.14. **Beispiel.** Im reellen Vektorraum $V = \mathrm{Abb}(\mathbb{R}, \mathbb{R})$ betrachten wir die *Potenzfunktionen*

$$f_0\colon x \longmapsto 1\,, \quad f_1\colon x \longmapsto x\,, \quad f_2\colon x \longmapsto x^2\,, \quad \ldots\,, \quad f_n\colon x \longmapsto x^n\,, \quad \ldots\,.$$

Wie sieht der von $(f_0, f_1, f_2, \ldots) = (f_n)_{n\in\mathbb{N}}$ erzeugte Untervektorraum P von V aus?

Seine Elemente sind gerade die Linearkombinationen von endlich vielen der Potenzfunktionen. Indem wir eventuell Potenzfunktionen mit Koeffizient 0 hinzufügen (was am Wert der Linearkombination nichts ändert) und gleichartige Terme zusammenfassen, können wir annehmen, dass die Linearkombination die Form

$$f = a_0 f_0 + a_1 f_1 + \ldots + a_n f_n$$

hat mit $n \in \mathbb{N}$ und $a_0, a_1, \ldots, a_n \in \mathbb{R}$. Dann gilt

$$f(x) = a_0 f_0(x) + a_1 f_1(x) + \ldots + a_n f_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n\,.$$

Die Elemente von P sind also gerade die *Polynomfunktionen*. ♣

Wir notieren noch einige einfache Eigenschaften der linearen Hülle.

7.15. **Lemma.** *Sei V ein K-Vektorraum.*

LEMMA
Eigensch. lineare Hülle

(1) *Für Teilmengen $A \subset B \subset V$ gilt $\langle A\rangle \subset \langle B\rangle$.*

(2) *Sei E ein Erzeugendensystem von V. Eine Teilmenge $A \subset V$ ist genau dann ein Erzeugendensystem von V, wenn $E \subset \langle A\rangle$ gilt.*

Beweis. Übung. ❑

8. Lineare Unabhängigkeit

Wir haben gesehen, dass ein K-Vektorraum V sehr viele Erzeugendensysteme haben kann; eines davon ist zum Beispiel die Menge V selbst. Das erscheint aber ein wenig verschwenderisch, sodass sich die Frage stellt, ob es auch minimale Erzeugendensysteme gibt und wie sie gegebenenfalls charakterisiert werden können. Dazu überlegen wir Folgendes: Sei E ein Erzeugendensystem von V, das nicht minimal ist in dem Sinn, dass es ein Element $v_0 \in E$ gibt, sodass $E_0 = E \setminus \{v_0\}$ auch schon ein Erzeugendensystem von V ist. Dann können wir v_0 als Linearkombination von Elementen von E_0 schreiben:

$$v_0 = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$$

mit $v_1, v_2, \ldots, v_n \in E_0$ und $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$. Dabei können wir annehmen, dass $v_1, v_2, \ldots, v_n$ paarweise verschieden sind (sonst fassen wir die Terme entsprechend zusammen). Wenn wir $\lambda_0 = -1$ setzen, dann können wir das auch in symmetrischer Form schreiben als

$$\lambda_0 v_0 + \lambda_1 v_1 + \ldots + \lambda_n v_n = \mathbf{0}\,.$$

Es gibt also eine *nichttriviale* Linearkombination (das ist eine, in der nicht alle Koeffizienten null sind; hier ist $\lambda_0 = -1 \neq 0$) von paarweise verschiedenen Elementen von E, die den Nullvektor ergibt.

Umgekehrt gilt: Gibt es eine solche nichttriviale Linearkombination von Elementen von E, deren Wert der Nullvektor ist, etwa

$$\lambda_0 v_0 + \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = \mathbf{0}$$

mit $v_0, v_1, v_2, \ldots, v_n \in E$ paarweise verschieden, dann ist $\lambda_j \neq 0$ für wenigstens ein $j \in \{0, 1, 2, \ldots, n\}$. Wir können (falls nötig) die Nummerierung so ändern, dass $\lambda_0 \neq 0$ ist. Dann ist die Gleichung äquivalent zu

$$v_0 = (-\lambda_0^{-1}\lambda_1)v_1 + (-\lambda_0^{-1}\lambda_2)v_2 + \ldots + (-\lambda_0^{-1}\lambda_n)v_n\,.$$

Wir können also ein Element v_0 von E als Linearkombination von Elementen von $E \setminus \{v_0\}$ schreiben. Daraus folgt, dass $E_0 = E \setminus \{v_0\}$ immer noch ein Erzeugendensystem von V ist. Das sieht man so: Wir nehmen an, dass es eine Darstellung

$$v_0 = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$$

gibt mit $v_1, v_2, \ldots, v_n \in E_0$ und $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$ (oben hatten wir $-\lambda_0^{-1}\lambda_j$ statt λ_j). Jetzt müssen wir zeigen, dass jedes Element $v \in V$ als Linearkombination von Elementen von E_0 geschrieben werden kann. Wir können (da ja E ein Erzeugendensystem ist) v jedenfalls als Linearkombination von Elementen von E schreiben:

$$v = \mu_1 w_1 + \mu_2 w_2 + \ldots + \mu_m w_m$$

mit $\mu_1, \mu_2, \ldots, \mu_m \in K$ und $w_1, w_2, \ldots, w_m \in E$; durch Zusammenfassen gleichartiger Terme können wir erreichen, dass $w_1, w_2, \ldots, w_m$ paarweise verschieden sind. Kommt v_0 nicht unter diesen Elementen vor, dann haben wir bereits eine Linearkombination von Elementen von E_0. Wenn v_0 vorkommt, dann können wir (möglicherweise nach Änderung der Nummerierung) annehmen, dass $v_0 = w_m$ ist. Dann haben wir

$$\begin{aligned} v &= \mu_1 w_1 + \mu_2 w_2 + \ldots + \mu_{m-1} w_{m-1} + \mu_m w_m \\ &= \mu_1 w_1 + \mu_2 w_2 + \ldots + \mu_{m-1} w_{m-1} + \mu_m v_0 \\ &= \mu_1 w_1 + \mu_2 w_2 + \ldots + \mu_{m-1} w_{m-1} + \mu_m(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n) \\ &= \mu_1 w_1 + \mu_2 w_2 + \ldots + \mu_{m-1} w_{m-1} + (\mu_m\lambda_1)v_1 + (\mu_m\lambda_2)v_2 + \ldots + (\mu_m\lambda_n)v_n\,; \end{aligned}$$

dies ist eine Linearkombination von Elementen von E_0 (denn wir haben v_0 durch Elemente von E_0 ersetzt).

E ist also genau dann ein minimales Erzeugendensystem, wenn der Nullvektor *nicht* als nichttriviale Linearkombination von (paarweise verschiedenen) Elementen von E geschrieben werden kann. Diese Eigenschaft ist sehr wichtig und hat einen eigenen Namen.

* 8.1. **Definition.** Sei V ein K-Vektorraum.

DEF
Linear unabhängig

(1) Sei $n \in \mathbb{N}$. Die Vektoren $v_1, v_2, \ldots, v_n \in V$ heißen *(K-)linear unabhängig*, wenn gilt:

$$\forall \lambda_1, \lambda_2, \ldots, \lambda_n \in K\colon \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = \mathbf{0} \Rightarrow \lambda_1 = \lambda_2 = \ldots = \lambda_n = 0\,.$$

Anderenfalls heißen die Vektoren *(K-)linear abhängig.*

(2) Sei I eine Menge. Eine Familie $(v_i)_{i\in I}$ von Elementen von V heißt *(K-)linear unabhängig*, wenn für jede endliche Teilmenge $\{i_1, i_2, \ldots, i_n\} \subset I$ (mit $i_1, i_2, \ldots, i_n$ paarweise verschieden) die Vektoren $v_{i_1}, v_{i_2}, \ldots, v_{i_n}$ linear unabhängig sind. Anderenfalls heißt $(v_i)_{i\in I}$ *(K-)linear abhängig.*

(3) Eine Teilmenge $A \subset V$ heißt *(K-)linear unabhängig*, wenn die Familie $(v)_{v\in A}$ linear unabhängig ist, sonst heißt A *(K-)linear abhängig.* ♢

Eine Familie oder Menge von Vektoren ist also genau dann linear abhängig, wenn man den Nullvektor als nichttriviale Linearkombination von Vektoren aus der Familie oder der Menge schreiben kann.

Der Unterschied zwischen Familien und Mengen ist, dass die Elemente in einer Familie gewissermaßen durch die Indexmenge nummeriert sind und sich wiederholen können, während die Elemente einer Menge keine weitere Ordnung haben und nicht mehrfach vorkommen. Zum Beispiel ist im K-Vektorraum K die Familie $(1)_{i\in\{1,2\}}$ linear abhängig, weil $1\cdot 1 + (-1)\cdot 1 = 0$ eine nichttriviale Linearkombination ist, die den Nullvektor $0 \in K$ darstellt. (Allgemeiner ist jede Familie, in der ein Element mehrfach vorkommt, linear abhängig, wie man auf die gleiche Weise sehen kann.) Dagegen ist die Menge $\{1 \mid i \in \{1,2\}\} = \{1\}$ linear unabhängig, vergleiche Beispiel 8.3 unten.

Eine Menge A von Vektoren ist genau dann linear unabhängig, wenn jede *endliche* Teilmenge von A linear unabhängig ist.

Wie wir oben gesehen haben, ist ein Erzeugendensystem genau dann minimal, wenn es linear unabhängig ist. Aus unseren Überlegungen hat sich auch Folgendes ergeben:

$v_1, v_2, \ldots, v_n \in V$ sind genau dann linear abhängig, wenn sich einer der Vektoren als Linearkombination der übrigen schreiben lässt.

Wichtig: Die Definition der Linearen Unabhängigkeit ist zentral für die Lineare Algebra. Es ist äußerst wichtig, dass Sie sie verstehen!

8.2. **Beispiel.** Wir betrachten den Grenzfall: Ist die leere Menge linear unabhängig oder linear abhängig?

BSP
∅ ist linear unabhängig

Die einzige Linearkombination der leeren Menge ist die leere Summe mit dem Wert $\mathbf{0}$. Ist diese Linearkombination trivial oder nicht? Da „trivial“ bedeutet, dass alle Koeffizienten null sind, muss die leere Linearkombination trivial sein, denn da

es keine Koeffizienten gibt, ist jede Allaussage über die Koeffizienten wahr. Die leere Menge ist also linear unabhängig.

Das passt auch mit der obigen Beobachtung zusammen, dass ein Erzeugendensystem genau dann minimal ist, wenn es linear unabhängig ist, denn die leere Menge kann man ja nicht verkleinern. ♣

8.3. **Beispiel.** Wann ist ein einzelner Vektor v linear unabhängig?

BSP Wann ist v linear unabhängig?

Die Linear„kombinationen“ haben die Form λv mit λ aus dem jeweiligen Körper. Aus $\lambda v = \mathbf{0}$ folgt $\lambda = 0$ oder $v = \mathbf{0}$ (vgl. Lemma 5.3). Das zeigt, dass v linear unabhängig ist, wenn v nicht der Nullvektor ist. Auf der anderen Seite ist $1 \cdot \mathbf{0} = \mathbf{0}$ eine nichttriviale Linearkombination, die den Nullvektor ergibt, also ist $\mathbf{0}$ linear abhängig. ♣

8.4. **Beispiel.** Nach unseren Überlegungen vom Anfang dieses Abschnitts sind zwei Vektoren $v_1, v_2 \in V$ genau dann linear abhängig, wenn einer der beiden ein Vielfaches des anderen ist: $v_2 = \lambda v_1$ oder $v_1 = \lambda v_2$ für ein $\lambda \in K$. (Ist $v_1 = \mathbf{0}$, $v_2 \neq \mathbf{0}$, dann ist v_1 ein Vielfaches von v_2, aber nicht umgekehrt.) ♣

BSP Lineare Unabhängigkeit von zwei Vektoren

8.5. **Beispiel.** Hier ist ein sehr konkretes (und typisches) Beispiel. Sind die Vektoren $v_1 = (1, 1, 1, 1)$, $v_2 = (1, 2, 3, 4)$ und $v_3 = (1, 3, 5, 7)$ in $V = \mathbb{R}^4$ linear unabhängig oder nicht?

BSP 3 Vektoren im $\mathbb{R}^4$

Wir müssen die Bedingung überprüfen. Seien also $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ mit

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = \mathbf{0} = (0, 0, 0, 0)\,.$$

Die Frage ist, ob daraus zwingend $\lambda_1 = \lambda_2 = \lambda_3 = 0$ folgt. Ausgeschrieben lautet die Gleichung

$$(\lambda_1 + \lambda_2 + \lambda_3, \lambda_1 + 2\lambda_2 + 3\lambda_3, \lambda_1 + 3\lambda_2 + 5\lambda_3, \lambda_1 + 4\lambda_2 + 7\lambda_3) = (0, 0, 0, 0)\,;$$

das ist äquivalent zu den vier Gleichungen

$$\begin{aligned} \lambda_1 + \ \lambda_2 + \ \lambda_3 &= 0 \\ \lambda_1 + 2\lambda_2 + 3\lambda_3 &= 0 \\ \lambda_1 + 3\lambda_2 + 5\lambda_3 &= 0 \\ \lambda_1 + 4\lambda_2 + 7\lambda_3 &= 0 \end{aligned}$$

Dieses Gleichungssystem hat $(\lambda_1, \lambda_2, \lambda_3) = (1, -2, 1)$ als eine nichttriviale Lösung. Das bedeutet, dass die Vektoren linear abhängig sind. ♣

8.6. **Beispiel.** Das Erzeugendensystem $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ von K^n ist linear unabhängig, denn

BSP $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ ist linear unabhängig

$$\lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \ldots + \lambda_n \mathbf{e}_n = (\lambda_1, \lambda_2, \ldots, \lambda_n)$$

ist genau dann der Nullvektor, wenn alle Koeffizienten null sind. ♣

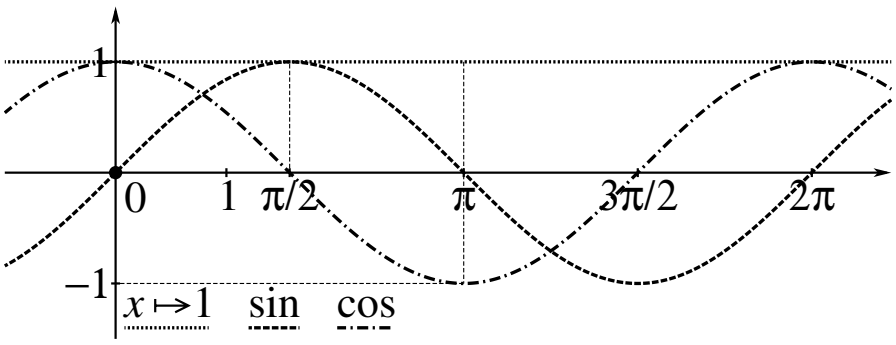


Abbildung 1. Zu Beispiel 8.7

BSP Lineare Unabhängigkeit in $\mathcal{C}(\mathbb{R})$

8.7. **Beispiel.** Die Funktionen $x \mapsto 1$, sin, cos, $\sin^2$, $\cos^2$ aus dem Raum $\mathcal{C}(\mathbb{R})$ der stetigen reellen Funktionen sind linear abhängig, denn es gilt

$$\forall x \in \mathbb{R}\colon\ \sin^2(x) + \cos^2(x) - 1 = 0\,,$$

also haben wir eine nichttriviale Linearkombination, die die Nullfunktion darstellt:

$$(-1)\cdot(x \mapsto 1) + 0\cdot\sin + 0\cdot\cos + 1\cdot\sin^2 + 1\cdot\cos^2 = \mathbf{0}\,.$$

Andererseits sind $x \mapsto 1$, sin und cos linear unabhängig:
Aus $\lambda_1 + \lambda_2\sin(x) + \lambda_3\cos(x) = 0$ für alle $x \in \mathbb{R}$ folgt durch Einsetzen von $x = 0, \pi, \pi/2$

$$\lambda_1 + \lambda_3 = \lambda_1 - \lambda_3 = \lambda_1 + \lambda_2 = 0$$

und damit $\lambda_1 = \lambda_2 = \lambda_3 = 0$. ♣

Im folgenden Beispiel wird vollständige Induktion verwendet. Die vollständige Induktion wurde in der Analysis behandelt. Für diejenigen, die die Analysis I erst nach der Linearen Algebra hören, ist hier ein kleiner Crashkurs.

Das Beweisprinzip der vollständigen Induktion beruht auf der rekursiven Definition der natürlichen Zahlen: 0 ist eine natürliche Zahl (oder 1, je nach Vorliebe...), mit n ist auch $n+1$ eine natürliche Zahl, und alle natürlichen Zahlen entstehen sukzessive auf diese Weise:

$$0 \to 1 \to 2 \to 3 \to \ldots \to n \to n+1 \to \ldots$$

Sei nun $A(n)$ eine Aussage über die natürliche Zahl n. Wenn wir zeigen können, dass $A(0)$ gilt und dass außerdem die Implikationen

$$A(0) \Rightarrow A(1), \quad A(1) \Rightarrow A(2), \quad \ldots, \quad A(n) \Rightarrow A(n+1), \quad \ldots$$

gelten, dann folgt $A(n)$ für alle $n \in \mathbb{N}$: $A(0)$ haben wir gezeigt, aus $A(0)$ und $A(0) \Rightarrow A(1)$ folgt $A(1)$, aus $A(1)$ und $A(1) \Rightarrow A(2)$ folgt $A(2)$, und so weiter. Daraus ergibt sich das folgende **Induktionsprinzip**:

$$\text{Aus} \quad A(0) \quad \text{und} \quad \forall n \in \mathbb{N}\colon A(n) \Rightarrow A(n+1) \qquad \text{folgt} \qquad \forall n \in \mathbb{N}\colon A(n).$$

Dabei heißt der Beweis von $A(0)$ der *Induktionsanfang* und der Beweis der Implikation $A(n) \Rightarrow A(n+1)$ der *Induktionsschritt*. Im Induktionsschritt heißt $A(n)$ meistens die *Induktionsannahme* oder *Induktionsvoraussetzung*.

Das folgende Standardbeispiel sollte das etwas klarer machen. Wir wollen zeigen, dass

$$\forall n \in \mathbb{N}\colon \sum_{j=0}^{n} j = \frac{n(n+1)}{2}$$

gilt (links in der Gleichung steht die Summe $0 + 1 + 2 + \ldots + (n-1) + n$).

Induktionsanfang: Wir setzen $n = 0$; die Aussage ist dann

$$\sum_{j=0}^{0} j = \frac{0(0+1)}{2};$$

beide Seiten sind null, also gilt die Aussage.

Induktionsschritt: Jetzt ist $n \in \mathbb{N}$ beliebig. Wir nehmen an (Induktionsvoraussetzung), dass die Aussage für n stimmt. Unter dieser Annahme müssen wir zeigen, dass sie auch für $n+1$ richtig ist. Das geht in diesem Fall zum Beispiel so:

$$\sum_{j=0}^{n+1} j = \sum_{j=0}^{n} j + (n+1) \overset{\text{IV}}{=} \frac{n(n+1)}{2} + n + 1 = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}$$

An der mit „IV“ markierten Stelle haben wir die Induktionsvoraussetzung benutzt.

Varianten des Induktionsprinzips ergeben sich, wenn man nicht bei 0 anfängt, sondern mit einer anderen ganzen Zahl n_0; dann folgt die Aussage für alle $n \geq n_0$. Manchmal braucht man die Voraussetzung für zwei Werte von n (dann zeigt man $A(0)$ und $A(1)$ als Induktionsanfang und $(A(n) \wedge A(n+1)) \Rightarrow A(n+2)$ als Induktionsschritt), oder man stützt sich gleich auf *alle* kleineren Fälle:

$$\text{Aus} \quad \forall n \in \mathbb{N}\colon \big(\forall k \in \mathbb{N}, k < n\colon A(k)\big) \Rightarrow A(n) \qquad \text{folgt} \qquad \forall n \in \mathbb{N}\colon A(n).$$

Die Induktionsvoraussetzung ist dann also, dass $A(k)$ für *alle* $k < n$ gilt; daraus muss $A(n)$ hergeleitet werden. Der Induktionsanfang ist dabei implizit eingeschlossen, denn für $n = 0$ ist die Voraussetzung trivialerweise erfüllt (die Menge der kleineren k ist leer), also zeigt man hier $A(0)$ ohne Zusatzannahmen.

Alternativ dazu kann man das **Prinzip des kleinsten Verbrechers** (*minimal criminal*) verwenden: Ist die Aussage $\forall n \in \mathbb{N}\colon A(n)$ falsch, dann gibt es ein *kleinstes* n, für das $A(n)$ nicht gilt. Damit muss aber $A(k)$ für alle $k < n$ gelten, woraus man dann einen Widerspruch ableitet. Die Annahme, dass es ein Gegenbeispiel gibt, ist also nicht haltbar; damit muss $A(n)$ für alle $n \in \mathbb{N}$ gelten.

BSP Potenzfunktionen

8.8. **Beispiel.** Die Potenzfunktionen $f_n\colon x \mapsto x^n$ für $n \in \mathbb{N}$ sind linear unabhängig. Das bedeutet

$$\forall n \in \mathbb{N}\, \forall a_0, a_1, \ldots, a_n \in \mathbb{R}\colon$$
$$(\forall x \in \mathbb{R}\colon a_0 + a_1 x + \ldots + a_n x^n = 0) \Rightarrow a_0 = a_1 = \ldots = a_n = 0\,.$$

Das kann man durch vollständige Induktion beweisen. Für $n = 0$ reduziert sich die Behauptung auf die triviale Aussage $a_0 = 0 \Rightarrow a_0 = 0$. Für den Induktionsschritt nehmen wir an, dass die Aussage für n gilt. Um die Aussage für $n+1$ zu beweisen, seien $a_0, a_1, \ldots, a_{n+1} \in \mathbb{R}$ mit

$$\forall x \in \mathbb{R}\colon a_0 + a_1 x + a_2 x^2 + \ldots + a_{n+1} x^{n+1} = 0\,.$$

Einsetzen von $x = 0$ liefert $a_0 = 0$, also haben wir

$$\forall x \in \mathbb{R}\colon x(a_1 + a_2 x + \ldots + a_{n+1} x^n) = 0\,,$$

was bedeutet

$$\forall x \in \mathbb{R} \setminus \{0\}\colon a_1 + a_2 x + \ldots + a_{n+1} x^n = 0\,.$$

Weil Polynomfunktionen stetig sind (das lernen Sie in der Analysis), gilt dies dann auch für $x = 0$, also

$$\forall x \in \mathbb{R}\colon a_1 + a_2 x + \ldots + a_{n+1} x^n = 0\,.$$

Aus der Induktionsvoraussetzung folgt dann $a_1 = a_2 = \ldots = a_{n+1} = 0$ wie gewünscht.

Man kann diese Aussage auch beweisen, indem man die (aus der Schule bekannte?) Tatsache verwendet, dass ein Polynom vom Grad n (also eine Polynomfunktion wie oben mit $a_n \neq 0$) höchstens n Nullstellen hat. Das bedeutet, dass es nicht die Nullfunktion sein kann (denn die hat unendlich viele Nullstellen). Die einzige Möglichkeit, die Nullfunktion zu bekommen, ist dann, dass man alle Koeffizienten null setzt. ♣

Wir schreiben noch eine einfache, aber nützliche Beobachtung auf, die unsere Überlegungen vom Beginn dieses Abschnitts formalisiert.

LEMMA
Erzeugnis einer linear unabhängigen Menge

8.9. **Lemma.** *Sei V ein Vektorraum.*

(1) *Sei $(v_1, v_2, \ldots, v_n)$ ein linear unabhängiges Tupel von Vektoren in V. Dann gilt für alle $v \in V$:*

$$v \in \langle v_1, v_2, \ldots, v_n\rangle \iff (v, v_1, v_2, \ldots, v_n) \text{ linear abhängig.}$$

(2) *Sei $(v_i)_{i\in I}$ eine linear unabhängige Familie von Vektoren in V. Sei $v \in V$ beliebig. Sei weiter $i_0 \notin I$ und $I' = I \cup \{i_0\}$; wir setzen $v_{i_0} = v$. Dann gilt:*

$$v \in \langle v_i \mid i \in I\rangle \iff (v_i)_{i\in I'} \text{ linear abhängig.}$$

(3) *Für Teilmengen $A \subset V$ und Vektoren $v \in V \setminus A$ gilt entsprechend:*

$$v \in \langle A\rangle \iff A \cup \{v\} \text{ linear abhängig.}$$

Beweis.

(1) „⇒“: $v \in \langle v_1, v_2, \ldots, v_n\rangle$ bedeutet, dass $v = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$ eine Linearkombination der Vektoren v_j ist. Dann ist

$$(-1)v + \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = \mathbf{0}$$

und diese Linearkombination ist nichttrivial. Also ist $(v, v_1, \ldots, v_n)$ linear abhängig.

„⇐“: Da $(v, v_1, \ldots, v_n)$ linear abhängig ist, gibt es eine nichttriviale Linearkombination

$$\lambda v + \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = \mathbf{0}\,.$$

Dabei kann λ nicht null sein, denn sonst hätten wir eine nichttriviale Linearkombination von $(v_1, v_2, \ldots, v_n)$, die den Nullvektor darstellt, im Widerspruch zur linearen Unabhängigkeit dieser Vektoren. Dann können wir die Gleichung aber nach v auflösen:

$$v = -\lambda^{-1}\lambda_1 v_1 - \lambda^{-1}\lambda_2 v_2 - \ldots - \lambda^{-1}\lambda_n v_n\,,$$

was $v \in \langle v_1, v_2, \ldots, v_n\rangle$ zeigt.

(2) Das folgt aus dem ersten Teil, da in den jeweils zu betrachtenden Linearkombinationen nur endlich viele Vektoren v_i vorkommen.

(3) Der Beweis für Teilmengen ist analog. ❑

9. Basis und Dimension

Linear unabhängige Erzeugendensysteme spielen eine fundamentale Rolle in der Linearen Algebra.

* 9.1. **Definition.** Sei V ein K-Vektorraum. Eine Familie $(v_i)_{i\in I}$ von Elementen von V heißt *(K-)Basis(familie)* von V, wenn sie linear unabhängig und gleichzeitig ein Erzeugendensystem von V ist. Eine Teilmenge $B \subset V$ heißt *(K-)Basis(menge)* von V, wenn sie ein linear unabhängiges Erzeugendensystem von V ist. ♢

DEF Basis

Manchmal ist es praktischer, mit Familien (also „nummerierten Mengen“) zu arbeiten, und manchmal ist es praktischer, mit Mengen zu arbeiten, darum haben wir den Begriff der Basis in beiden Versionen definiert. Der Unterschied ist gering, denn in einer linear unabhängigen Familie kann kein Element mehrfach auftreten.

9.2. **Beispiele.**

BSP Basen

(1) Ist V ein Vektorraum und $A \subset V$ linear unabhängig, dann ist A eine Basis von $\langle A\rangle$ (denn A ist ein linear unabhängiges Erzeugendensystem von $\langle A\rangle$).

(2) Die leere Menge ist Basis des Null-Vektorraums $\{\mathbf{0}\}$.

(3) Das Tupel $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ ist eine K-Basis von K^n, die sogenannte *Standardbasis* von K^n.

DEF Standardbasis von K^n

(4) Die Folge $(f_n)_{n\in\mathbb{N}}$ der Potenzfunktionen ist eine Basis des Vektorraums P der Polynomfunktionen. ♣

Wir hatten zu Beginn von §8 gesehen, dass ein Erzeugendensystem genau dann minimal ist, wenn es linear unabhängig (also eine Basis) ist. Wir formulieren das hier noch einmal und ergänzen es um eine ähnliche Aussage über linear unabhängige Mengen.

* 9.3. **Lemma.** *Seien V ein Vektorraum und $B = (b_i)_{i\in I}$ eine Familie von Vektoren in V. Dann sind die folgenden Aussagen äquivalent:*

LEMMA Charakterisierung von Basen

(1) *B ist eine Basis von V.*

(2) *B ist ein minimales Erzeugendensystem von V.*

(3) *B ist eine maximale linear unabhängige Familie in V.*

„Maximal“ heißt dabei, dass für jedes $v \in V$ die um v erweiterte Familie nicht mehr linear unabhängig ist.

Beweis. Nach Definition 9.1 ist eine Basis ein linear unabhängiges Erzeugendensystem.

„(1)⇒(2)“: Es ist noch zu zeigen, dass keine echte Teilfamilie von B ein Erzeugendensystem ist. Sei dafür $i_0 \in I$ und $I' = I \setminus \{i_0\}$. Die Familie $B' = (b_i)_{i\in I'}$ entsteht aus B durch Weglassen der Komponente b_{i_0}. Wir zeigen, dass B' kein Erzeugendensystem von V ist. Dazu müssen wir einen Vektor angeben, der nicht in $\langle B'\rangle$ liegt. Da B linear unabhängig ist, gilt das auch für die kleinere Familie B'. Wäre $b_{i_0} \in \langle B'\rangle$, dann würde aus Lemma 8.9, (2) folgen, dass B (nämlich B' zusammen mit b_{i_0}) linear abhängig sein müsste. Das ist jedoch nicht der Fall, also folgt, dass $b_{i_0} \notin \langle B'\rangle$ ist. Damit ist B' kein Erzeugendensystem von V.

„(2)⇒(1)“: B ist bereits ein Erzeugendensystem; es bleibt zu zeigen, dass B linear unabhängig ist. Wäre B linear abhängig, dann könnte eine Komponente b_{i_0} von B als Linearkombination der übrigen Komponenten B' geschrieben werden. Dann ist aber B' bereits ein Erzeugendensystem (vergleiche die Überlegungen zu Beginn von §8), damit wäre B nicht minimal gewesen, Widerspruch. Also muss B linear unabhängig sein.

„(1)⇒(3)“: Es ist zu zeigen, dass jede echt größere Familie linear abhängig ist. Das folgt aber unmittelbar aus Lemma 8.9, (2), da für alle $v \in V$ ja $v \in \langle B \rangle$ gilt.

„(3)⇒(1)“: B ist bereits linear unabhängig; es bleibt zu zeigen, dass B ein Erzeugendensystem ist. Dazu sei $v \in V$. Dann ist die um v erweiterte Familie B' linear abhängig; nach Lemma 8.9. (2) folgt $v \in \langle B \rangle$. Da $v \in V$ beliebig war, folgt $\langle B \rangle = V$, also ist B ein Erzeugendensystem von V. ❑

Wir können die Eigenschaften, ein Erzeugendensystem, linear unabhängig oder eine Basis zu sein, auch durch die Anzahl der Linearkombinationen ausdrücken, die ein gegebenes Element von V darstellen. Wir formulieren das hier für endlich viele Vektoren.

LEMMA
EZS/LU/Basis über Anzahl Lin.komb.

9.4. **Lemma.** *Sei V ein K-Vektorraum und seien $v_1, v_2, \ldots, v_n \in V$. Wir definieren die zugehörige „Linearkombinationenabbildung“*

$$\phi_{(v_1,v_2,\ldots,v_n)} \colon K^n \longrightarrow V, \quad (\lambda_1, \lambda_2, \ldots, \lambda_n) \longmapsto \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n \,.$$

Dann gilt:

(1) *$(v_1, v_2, \ldots, v_n)$ ist genau dann ein **Erzeugendensystem** von V, wenn jeder Vektor $v \in V$ auf **mindestens** eine Weise als Linearkombination von $(v_1, v_2, \ldots, v_n)$ geschrieben werden kann, also genau dann, wenn $\phi_{(v_1,v_2,\ldots,v_n)}$ **surjektiv** ist.*

(2) *$(v_1, v_2, \ldots, v_n)$ ist genau dann **linear unabhängig**, wenn jeder Vektor $v \in V$ auf **höchstens** eine Weise als Linearkombination von $(v_1, v_2, \ldots, v_n)$ geschrieben werden kann, also genau dann, wenn $\phi_{(v_1,v_2,\ldots,v_n)}$ **injektiv** ist.*

(3) *$(v_1, v_2, \ldots, v_n)$ ist genau dann eine **Basis** von V, wenn jeder Vektor $v \in V$ auf **genau** eine Weise als Linearkombination von $(v_1, v_2, \ldots, v_n)$ geschrieben werden kann, also genau dann, wenn $\phi_{(v_1,v_2,\ldots,v_n)}$ **bijektiv** ist.*

Beweis. Teil (1) folgt direkt aus Definition 7.8 und Satz 7.9.

Wir beweisen Teil (2). „⇒“: Wir nehmen an, dass $v_1, v_2, \ldots, v_n$ linear unabhängig sind. Sei $v \in V$. Wenn wir zwei Linearkombinationen haben, also

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = \mu_1 v_1 + \mu_2 v_2 + \ldots + \mu_n v_n$$

mit $\lambda_1, \lambda_2, \ldots, \lambda_n, \mu_1, \mu_2, \ldots, \mu_n \in K$, dann bilden wir die Differenz:

$$(\lambda_1 - \mu_1)v_1 + (\lambda_2 - \mu_2)v_2 + \ldots + (\lambda_n - \mu_n)v_n = \mathbf{0} \,.$$

Weil $v_1, v_2, \ldots, v_n$ linear unabhängig sind, muss das die triviale Linearkombination sein, also folgt $\lambda_1 = \mu_1$, $\lambda_2 = \mu_2$, …, $\lambda_n = \mu_n$.

„⇐“: Wir nehmen an, dass jedes $v \in V$ höchstens auf eine Weise als Linearkombination von $v_1, v_2, \ldots, v_n$ darstellbar ist. Das gilt dann auch für $v = \mathbf{0}$. Da die triviale Linearkombination $\mathbf{0}$ darstellt, muss es die einzige sein. Damit ist gezeigt, dass $v_1, v_2, \ldots, v_n$ linear unabhängig sind.

Teil (3) folgt dann aus (1) und (2). ❑

Um das vorstehende Lemma auch für beliebige Familien $(v_i)_{i\in I}$ von Vektoren formulieren zu können, definieren wir

$$K^{(I)} = \{(\lambda_i)_{i\in I} \in K^I \mid \{i \in I \mid \lambda_i \neq 0\} \text{ ist endlich}\}\,.$$

Das ist also die Menge derjenigen Familien von Elementen von K mit Indexmenge I, die nur endlich viele von null verschiedene Komponenten haben. Dann können wir die Linearkombinationenabbildung analog definieren als

$$\phi_{(v_i)_{i\in I}} \colon K^{(I)} \longrightarrow V, \quad (\lambda_i)_{i\in I} \longmapsto \sum_{i\in I} \lambda_i v_i$$

(vergleiche Definition 7.11). Dann gilt wieder:

(1) $(v_i)_{i\in I}$ Erzeugendensystem $\iff$ $\phi_{(v_i)_{i\in I}}$ surjektiv.

(2) $(v_i)_{i\in I}$ linear unabhängig $\iff$ $\phi_{(v_i)_{i\in I}}$ injektiv.

(3) $(v_i)_{i\in I}$ Basis $\iff$ $\phi_{(v_i)_{i\in I}}$ bijektiv.

$K^{(I)}$ ist genau der K-Untervektorraum von K^I, der durch die Familien $\mathbf{e}_i = (\delta_{ij})_{j\in I}$ für $i \in I$ erzeugt wird. Dabei ist $\delta_{ij} = 1$ für $i = j$ und $\delta_{ij} = 0$ für $i \neq j$ (das sogenannte *Kronecker-Delta*); die Familie $\mathbf{e}_i$ hat also als i-te Komponente eine Eins, alle anderen Komponenten sind null. Das verallgemeinert die Standardbasis von K^n auf den Vektorraum $K^{(I)}$.

Eine Basis $(v_1, v_2, \ldots, v_n)$ von V verhilft uns also zu einer bijektiven Abbildung $K^n \to V$. Damit können wir die Elemente von V durch ihr Koeffiziententupel $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in K^n$ beschreiben (und Addition und Skalarmultiplikation von V verhalten sich genauso wie die von K^n). Das ist natürlich eine schöne Sache. Es stellt sich dann die Frage, ob jeder Vektorraum eine Basis hat. Wir werden das hier für endlich erzeugte Vektorräume positiv beantworten. (Ein Vektorraum ist *endlich erzeugt*, wenn er ein endliches Erzeugendensystem hat.) Dafür beweisen wir sogar eine stärkere Aussage, die viele nützliche Anwendungen haben wird.

DEF endl. erzeugter Vektorraum

* 9.5. **Satz.** *Sei V ein Vektorraum und seien $v_1, v_2, \ldots, v_n$ und $w_1, w_2, \ldots, w_m$ Elemente von V, sodass*

SATZ Basisergänzungssatz

(1) *$(v_1, v_2, \ldots, v_n)$ linear unabhängig und*

(2) *$(v_1, v_2, \ldots, v_n, w_1, w_2, \ldots, w_m)$ ein Erzeugendensystem von V ist.*

Dann kann man $(v_1, v_2, \ldots, v_n)$ durch Hinzunahme geeigneter Vektoren w_j zu einer Basis von V ergänzen.

Genauer bedeutet das: Es gibt $k \in \mathbb{N}$ und Indizes $j_1, j_2, \ldots, j_k \in \{1, 2, \ldots, m\}$, sodass

$$(v_1, v_2, \ldots, v_n, w_{j_1}, w_{j_2}, \ldots, w_{j_k})$$

eine Basis von V ist.

Die natürlichen Zahlen n und m dürfen und k kann auch null sein. Wenn $m = 0$ ist, dann ist $(v_1, v_2, \ldots, v_n)$ schon eine Basis, und es ist nichts zu tun (dann ist auch $k = 0$). Das werden wir im Beweis als Induktionsanfang benutzen.

Wenn $n = 0$ ist, dann sagt der Satz, dass jedes endliche Erzeugendensystem eine Basis enthält. Das ist plausibel, denn man kann ja immer Elemente entfernen, solange das Erzeugendensystem nicht minimal ist. Irgendwann (nach spätestens m-maligem Entfernen eines Elements) muss man bei einem minimalen Erzeugendensystem ankommen; das ist dann eine Basis.

Wenn sich $k = 0$ ergibt, dann bedeutet das, dass $(v_1, v_2, \ldots, v_n)$ bereits eine Basis ist.

Beweis. Der Beweis benutzt vollständige Induktion nach m. Er basiert auf der anschaulichen Idee, dass man ausgehend von $(v_1, \dots, v_n, w_1, \dots, w_m)$ sukzessive Elemente w_j entfernt, bis man linear unabhängige Vektoren hat. Dazu fixieren wir das n-Tupel $(v_1, \dots, v_n)$ und nehmen an (Voraussetzung (1)), dass es linear unabhängig ist. Die Aussage $A(m)$, die wir durch Induktion beweisen wollen, ist dann, dass die Behauptung für das fest gewählte Tupel $(v_1, \dots, v_n)$ und jedes m-Tupel $(w_1, \dots, w_m)$ von Vektoren von V gilt.

Im Induktionsanfang ist $m = 0$. Dann ist nach Voraussetzung (2) $(v_1, \dots, v_n)$ ein Erzeugendensystem von V und nach Voraussetzung (1) linear unabhängig, also bereits eine Basis: Die Behauptung gilt mit $k = 0$.

Für den Induktionsschritt nehmen wir an, die Behauptung gelte für m-Tupel $(w_1, \dots, w_m)$. Um sie für $m + 1$ zu zeigen, betrachten wir ein $(m + 1)$-Tupel $(w_1, \dots, w_{m+1})$ von Vektoren von V, sodass $(v_1, \dots, v_n, w_1, \dots, w_{m+1})$ ein Erzeugendensystem von V ist. Es gibt nun zwei Möglichkeiten:

- Entweder ist $(v_1, \dots, v_n, w_1, \dots, w_{m+1})$ linear unabhängig. Dann haben wir eine Basis; die Behauptung gilt mit $k = m + 1$ und $j_1 = 1, \dots, j_{m+1} = m + 1$.
- Anderenfalls ist $(v_1, \dots, v_n, w_1, \dots, w_{m+1})$ linear abhängig. Dann gibt es eine nichttriviale Linearkombination

  $$\lambda_1 v_1 + \dots + \lambda_n v_n + \mu_1 w_1 + \dots + \mu_m w_m + \mu_{m+1} w_{m+1} = \mathbf{0}\,.$$

  Hier können nicht alle $\mu_j = 0$ sein, denn dann hätten wir eine nichttriviale Linearkombination nur der v_i, die den Nullvektor darstellt, im Widerspruch zur linearen Unabhängigkeit von $(v_1, \dots, v_n)$. Also gibt es ein $j_0 \in \{1, 2, \dots, m + 1\}$ mit $\mu_{j_0} \neq 0$. Falls nötig, vertauschen wir w_{j_0} und w_{m+1}; dann können wir annehmen, dass $j_0 = m + 1$ ist. Wir können obige Gleichung dann nach w_{m+1} auflösen. Das zeigt, dass

  $$w_{m+1} \in \langle v_1, \dots, v_n, w_1, \dots, w_m \rangle$$

  ist, also ist

  $$\langle v_1, \dots, v_n, w_1, \dots, w_m \rangle = \langle v_1, \dots, v_n, w_1, \dots, w_m, w_{m+1} \rangle = V\,.$$

  Voraussetzung (2) ist also für $(w_1, \dots, w_m)$ erfüllt, sodass wir die Induktionsannahme verwenden können, was uns die gewünschte Aussage liefert.

  Wenn man ganz genau sein will, dann setzt man $w'_i = w_i$ für $i \in \{1, 2, \dots, m\} \backslash \{j_0\}$ und $w'_{j_0} = w_{m+1}$ (falls $j_0 \neq m + 1$), und wendet die Induktionsannahme auf $(w'_1, \dots, w'_m)$ an. Dann bekommt man $k \leq m$ und $j'_1, \dots, j'_k$, sodass $(v_1, \dots, v_n, w'_{j'_1}, \dots, w'_{j'_k})$ eine Basis ist. Mit $j_i = j'_i$ für $j'_i \neq j_0$ und $j_i = m + 1$, falls $j'_i = j_0 \neq m + 1$, ist dann $(v_1, \dots, v_n, w_{j_1}, \dots, w_{j_k})$ eine Basis. ❑

Man kann alternativ den Beweis auch so formulieren, dass man nacheinander Vektoren w_j zu den $v_1, v_2, \dots, v_n$ hinzunimmt, solange das entstehende Tupel linear unabhängig ist. Ist das nicht mehr möglich, dann muss man eine Basis haben.

Man beachte, dass in diesem Beweis die Zahl n nicht fixiert ist. (Er ist deswegen etwas weniger leicht zu verstehen, weswegen ich zunächst den anderen Beweis formuliert habe.)

Der Induktionsanfang, also der Fall $m = 0$, ist klar, denn dann ist $(v_1, v_2, \dots, v_n)$ bereits ein linear unabhängiges Erzeugendensystem, also eine Basis. Die Behauptung gilt also mit $k = 0$.

Für den Induktionsschritt nehmen wir an, dass die Aussage für ein gegebenes m stimmt, und beweisen sie für $m + 1$. Sei also $(v_1, v_2, \dots, v_n)$ linear unabhängig und seien außerdem $w_1, w_2, \dots, w_m, w_{m+1} \in V$, sodass $(v_1, v_2, \dots, v_n, w_1, w_2, \dots, w_m, w_{m+1})$ ein Erzeugendensystem von V ist. Wir unterscheiden zwei Fälle:

(1) $w_{m+1} \in \langle v_1, v_2, \ldots, v_n\rangle$.
Dann ist $(v_1, v_2, \ldots, v_n, w_1, w_2, \ldots, w_m)$ auch schon ein Erzeugendensystem; die Behauptung folgt direkt aus der Induktionsannahme.

(2) $w_{m+1} \notin \langle v_1, v_2, \ldots, v_n\rangle$.
Wir schreiben v_{n+1} für w_{m+1}. Dann ist $(v_1, v_2, \ldots, v_n, v_{n+1})$ linear unabhängig (wir benutzen hier wieder Lemma 8.9) und

$$(v_1, v_2, \ldots, v_n, v_{n+1}, w_1, w_2, \ldots, w_m)$$

ist ein Erzeugendensystem (dasselbe wie vorher). Nach der Induktionsannahme gibt es $j'_1, \ldots, j'_{k'} \in \{1, 2, \ldots, m\}$, sodass

$$(v_1, v_2, \ldots, v_n, v_{n+1}, w_{j'_1}, w_{j'_2}, \ldots, w_{j'_{k'}}) = (v_1, \ldots, v_n, w_{m+1}, w_{j'_1}, \ldots, w_{j'_{k'}})$$

eine Basis von V ist. Wir setzen

$$k = k' + 1, \quad j_1 = m + 1 \quad \text{und} \quad j_2 = j'_1,\ j_3 = j'_2,\ \ldots,\ j_k = j'_{k'}$$

und erhalten die Behauptung.

9.6. **Folgerung.** *Jeder Vektorraum, der ein endliches Erzeugendensystem besitzt, hat eine Basis.*

FOLG Existenz einer Basis

Beweis. Das folgt aus Satz 9.5, wenn man $n = 0$ nimmt. Genauer erhalten wir die Aussage, dass man eine Basis finden kann, die aus Elementen eines gegebenen endlichen Erzeugendensystems besteht. (Der Beweis reduziert sich in diesem Fall darauf, dass man aus dem gegebenen Erzeugendensystem solange Elemente entfernen kann, bis es minimal, also linear unabhängig, ist.) ❑

Was passiert, wenn es kein endliches Erzeugendensystem gibt? Dann gibt es auch noch einen Basisergänzungssatz, den wir hier für Mengen formulieren:

Satz. *Sei V ein Vektorraum und seien A und E Teilmengen von V, sodass A linear unabhängig und $A \cup E$ ein Erzeugendensystem von V ist. Dann gibt es eine Teilmenge $B \subset E$, sodass $A \cup B$ eine Basismenge von V ist.*

SATZ Basisergänzungssatz

Den Beweis kann man jetzt natürlich nicht mehr durch vollständige Induktion führen. Man braucht ein anderes Werkzeug dafür, zum Beispiel das sogenannte *Zornsche Lemma*. Es besagt Folgendes.

Satz. *Seien X eine Menge und $\mathcal{M} \subset \mathcal{P}(X)$ eine Menge von Teilmengen von X. Eine **Kette** in $\mathcal{M}$ ist eine Teilmenge $\mathcal{K} \subset \mathcal{M}$, sodass je zwei Elemente von $\mathcal{K}$ vergleichbar sind, das heißt*

$$\forall T_1, T_2 \in \mathcal{K}\colon T_1 \subset T_2 \quad \textit{oder} \quad T_2 \subset T_1\,.$$

*Wenn jede solche Kette $\mathcal{K}$ eine **obere Schranke** in $\mathcal{M}$ hat, wenn es also zu $\mathcal{K}$ ein Element $S \in \mathcal{M}$ gibt, so dass*

$$\forall T \in \mathcal{K}\colon T \subset S\,,$$

*dann hat $\mathcal{M}$ **maximale** Elemente. Es gibt dann also (mindestens) ein $M \in \mathcal{M}$, sodass gilt*

$$\forall T \in \mathcal{M}\colon M \subset T \Rightarrow M = T$$

(d.h., es gibt keine echt größere Menge in $\mathcal{M}$).

SATZ Zornsches Lemma

Man kann zeigen, dass das Zornsche Lemma (wenn man die „harmlosen“ Axiome der Mengenlehre als gegeben annimmt) zum Auswahlaxiom (siehe die Diskussion im Kleingedruckten auf Seite 20) äquivalent ist.

Der Beweis des Basisergänzungssatzes geht dann so: E ist die Menge X im Zornschen Lemma und $\mathcal{M} = \{B \subset E \mid A \cup B \text{ linear unabhängig}\}$. Wir müssen die Voraussetzung

des Zornschen Lemmas nachprüfen. Sei dazu $\mathcal{K} \subset \mathcal{M}$ eine Kette. Wir setzen $S = \bigcup \mathcal{K}$ (das ist also die Vereinigung all der Teilmengen von E, die Elemente der Kette $\mathcal{K}$ sind). Es ist dann klar, dass $T \subset S$ für alle $T \in \mathcal{K}$ gilt. Wir müssen noch zeigen, dass $S \in \mathcal{M}$ ist, dass also $A \cup S$ linear unabhängig ist. Angenommen, das wäre falsch, dann gäbe es eine nichttriviale Linearkombination von Elementen von $A \cup S$, die den Nullvektor darstellt. In dieser Linearkombination kommen nur endlich viele Elemente $v_1, v_2, \ldots, v_n$ von S vor. Da $S = \bigcup \mathcal{K}$, gibt es für jedes v_j ein $T_j \in \mathcal{K}$ mit $v_j \in T_j$. Nach eventueller Umnummerierung können wir annehmen, dass $K_1 \subset K_2 \subset \ldots \subset K_n$ ist (hier wird verwendet, dass $\mathcal{K}$ eine Kette ist). Dann sind aber $v_1, v_2, \ldots, v_n \in K_n$, und es würde folgen, dass $A \cup K_n$ linear abhängig ist. Weil $K_n \in \mathcal{M}$ ist, ist das ein Widerspruch, also muss $A \cup S$ linear unabhängig sein. (Für dieses Argument ist die Endlichkeit von Linearkombinationen entscheidend!) Damit ist S eine obere Schranke von $\mathcal{K}$ in $\mathcal{M}$ und die Voraussetzung im Zornschen Lemma ist erfüllt. Es folgt, dass $\mathcal{M}$ ein maximales Element B hat. Da $B \in \mathcal{M}$ ist, ist $A \cup B$ linear unabhängig. Wäre $A \cup B$ kein Erzeugendensystem, dann gäbe es $v \in E$ mit $v \notin \langle A \cup B \rangle$. Dann wäre aber $A \cup (B \cup \{v\})$ ebenfalls linear unabhängig. Das würde $B \cup \{v\} \in \mathcal{M}$ bedeuten, aber das kann nicht sein, da B maximal ist (v kann kein Element von B sein, sonst wäre $v \in \langle A \cup B \rangle$). Also ist $A \cup B$ auch ein Erzeugendensystem und somit eine Basis.

Wir erhalten daraus sofort (mit $A = \emptyset$ und $E = V$):

Folgerung. *Jeder Vektorraum hat eine Basis.*

FOLG Existenz von Basen

Aus dem Auswahlaxiom folgt also zum Beispiel, dass $\mathbb{R}$ als $\mathbb{Q}$-Vektorraum (als den man $\mathbb{R}$ mit seiner Addition und der auf $\mathbb{Q} \times \mathbb{R}$ eingeschränkten Multiplikation betrachten kann) eine Basis hat. Gesehen hat so eine Basis aber noch niemand. Wie schon früher erwähnt ist das Auswahlaxiom (und damit auch das Zornsche Lemma) inhärent inkonstruktiv, sodass unser Beweis oben (im Gegensatz zum endlichen Fall) keinerlei Hinweis darauf gibt, wie die gesuchte Teilmenge B zu finden wäre.

Eine weitere wichtige Folgerung besagt, dass man (in einem endlich erzeugten Vektorraum) beliebige linear unabhängige Vektoren stets zu einer Basis ergänzen kann.

9.7. **Folgerung.** *Sei V ein Vektorraum mit endlichem Erzeugendensystem und sei $(v_1, v_2, \ldots, v_n) \in V^n$ linear unabhängig. Dann gibt es $k \in \mathbb{N}$ und Vektoren $v_{n+1}, v_{n+2}, \ldots, v_{n+k} \in V$, sodass $(v_1, v_2, \ldots, v_{n+k})$ eine Basis von V ist.*

FOLG Erweiterung zu Basis

Beweis. Sei $(w_1, w_2, \ldots, w_m)$ ein endliches Erzeugendensystem von V. Dann sind für $v_1, \ldots, v_n$ und $w_1, \ldots, w_m$ die Voraussetzungen von Satz 9.5 erfüllt. Die Aussage des Satzes liefert dann die Behauptung, wenn man $v_{n+1} = w_{j_1}, \ldots, v_{n+k} = w_{j_k}$ setzt. ❑

9.8. **Beispiel.** Wir finden eine Basis des Untervektorraums

BSP Basis

$$U = \{(x, y, z) \in \mathbb{R}^3 \mid z = x + y\} \subset \mathbb{R}^3 .$$

Dazu finden wir möglichst viele linear unabhängige Vektoren und prüfen dann, ob wir ein Erzeugendensystem haben. Zum Beispiel sind $(1, 0, 1)$ und $(0, 1, 1)$ linear unabhängige Elemente von U, denn

$$\lambda(1, 0, 1) + \mu(0, 1, 1) = \mathbf{0} \iff (\lambda, \mu, \lambda + \mu) = (0, 0, 0) \iff \lambda = \mu = 0 .$$

Diese beiden Vektoren bilden auch ein Erzeugendensystem, denn für $(x, y, z) \in U$ gilt $z = x + y$, also

$$(x, y, z) = (x, y, x + y) = x(1, 0, 1) + y(0, 1, 1) \in \langle (1, 0, 1), (0, 1, 1) \rangle .$$

Damit ist $\big((1,0,1),(0,1,1)\big)$ eine Basis von U. ♣

Eine weitere wichtige Konsequenz des Basisergänzungssatzes ist der *Basisaustauschsatz*.

* 9.9. **Satz.** *Sei V ein Vektorraum und seien $(v_1, v_2, \dots, v_n)$ und $(w_1, w_2, \dots, w_m)$ zwei Basen von V. Für jedes $i \in \{1, 2, \dots, n\}$ gibt es ein $j \in \{1, 2, \dots, m\}$, sodass $(v_1, \dots, v_{i-1}, w_j, v_{i+1}, \dots, v_n)$ ebenfalls eine Basis von V ist.* **SATZ** Basisaustauschsatz

Man tauscht also das Basiselement v_i der ersten Basis durch ein Element der zweiten Basis aus.

Beweis. Wir können ohne Einschränkung $i = n$ annehmen (sonst ändere man die Nummerierung entsprechend). Wir wenden den Basisergänzungssatz 9.5 an mit $v_1, v_2, \dots, v_{n-1}$ und $w_1, w_2, \dots, w_m$. Die Voraussetzungen sind erfüllt, da Teilfamilien von linear unabhängigen Familien immer linear unabhängig sind und die w_j schon alleine ein Erzeugendensystem bilden. Es gibt also $k \in \mathbb{N}$ und Indizes $j_1, \dots, j_k \in \{1, 2, \dots, m\}$, sodass $(v_1, \dots, v_{n-1}, w_{j_1}, w_{j_2}, \dots, w_{j_k})$ eine Basis von V ist. Die Behauptung bedeutet gerade, dass man $k = 1$ wählen kann; wir setzen dann $j = j_1$. Es ist klar, dass $k > 0$ sein muss, denn $(v_1, v_2, \dots, v_{n-1})$ ist kein Erzeugendensystem mehr ($(v_1, v_2, \dots, v_{n-1}, v_n)$ ist ein minimales Erzeugendensystem, aus dem wir ein Element entfernt haben). Wir zeigen, dass $(v_1, v_2, \dots, v_{n-1}, w_{j_1})$ ein Erzeugendensystem ist; daraus folgt die Behauptung.

Wir haben $w_{j_1} \in V = \langle v_1, v_2, \dots, v_n\rangle$. Nach Lemma 8.9 bedeutet das, dass $(v_1, v_2, \dots, v_{n-1}, v_n, w_{j_1})$ linear abhängig ist. Da $(v_1, v_2, \dots, v_{n-1}, w_{j_1})$ als Teil der Basis $(v_1, \dots, v_{n-1}, w_{j_1}, w_{j_2}, \dots, w_{j_k})$ linear unabhängig ist, folgt dann wieder mit Lemma 8.9, dass $v_n \in \langle v_1, v_2, \dots, v_{n-1}, w_{j_1}\rangle$ ist. Da natürlich auch $v_1, v_2, \dots, v_{n-1}$ in diesem Untervektorraum enthalten sind, enthält er ein Erzeugendensystem von V; es folgt $\langle v_1, v_2, \dots, v_{n-1}, w_{j_1}\rangle = V$ wie behauptet. ❑

9.10. **Folgerung.** *Sei V ein Vektorraum und seien $(v_1, \dots, v_n)$ und $(w_1, \dots, w_m)$ zwei (endliche) Basen von V. Dann ist $n = m$.* **FOLG** Größe von Basen

Je zwei Basen haben also gleich viele Elemente.

Beweis. Wir nehmen $n > m$ an und leiten einen Widerspruch her (der Fall $n < m$ geht genauso). Durch n-malige Anwendung von Satz 9.9 (mit $i = 1, 2, \dots, n$) erhalten wir Indizes $j_1, j_2, \dots, j_n \in \{1, 2, \dots, m\}$, sodass $(w_{j_1}, w_{j_2}, \dots, w_{j_n})$ eine Basis von V ist. Da m kleiner als n ist, müssen sich in diesem Tupel Vektoren wiederholen. Dann sind $w_{j_1}, w_{j_2}, \dots, w_{j_n}$ aber nicht linear unabhängig. Dies ist der gewünschte Widerspruch. ❑

Wir führen eine Schreibweise für die Anzahl der Elemente einer Menge ein.

9.11. **Definition.** Sei M eine Menge. Wir schreiben $\#M$ für die Anzahl der Elemente von M. Wenn M unendlich ist, dann setzen wir $\#M = \infty$. ♢ **DEF** $\#M$

Eine andere häufig anzutreffende Schreibweise ist $|M|$. Ich bevorzuge $\#M$, weil es dabei keine Verwechslungsgefahr gibt.

Wir setzen hier ein intuitives Verständnis davon voraus, was die „Anzahl der Elemente“ einer (endlichen) Menge ist. Wenn man das formal sauber definieren will, ist es aber gar nicht so einfach. Man kann zum Beispiel für eine Menge M und $n \in \mathbb{N}$ definieren:

$$\#M = n \iff \exists f\colon M \to \mathbb{N}_{<n} \quad \text{mit } f \text{ bijektiv}\,.$$

(Beachte: $\mathbb{N}_{<n} = \{m \in \mathbb{N} \mid m < n\} = \{0, 1, 2, \dots, n-1\}$ hat gerade n Elemente.) Das formalisiert die Vorstellung, dass man die Elemente von 0 bis $n-1$ (oder analog von 1 bis n) durchnummerieren kann. Man muss dann noch zeigen, dass eine solche Bijektion nicht mit zwei Mengen $\mathbb{N}_{<n}$ zu verschiedenen n möglich ist und dass eine Menge genau dann unendlich ist, wenn es für kein $n \in \mathbb{N}$ eine solche Bijektion gibt.

Das Erste lässt sich durch vollständige Induktion erledigen: Angenommen, es gäbe eine Bijektion $f\colon \mathbb{N}_{<n+m} \to \mathbb{N}_{<n}$ mit $m > 0$. Für $n = 0$ ist das unmöglich (es gibt keine Abbildung von der nichtleeren Menge $\mathbb{N}_{<m}$ in die leere Menge $\mathbb{N}_{<0}$). Für $n > 0$ konstruiert man aus f eine neue Bijektion $f'\colon \mathbb{N}_{<n-1+m} \to \mathbb{N}_{<n-1}$, die es nach Induktionsannahme aber nicht geben kann. Dabei ist $f'(k) = f(k)$, außer $f(k) = n-1$; dann setzt man $f'(k) = f(n-1+m)$.

Für das Zweite braucht man eine Definition, was eine „unendliche Menge“ ist. Eine Möglichkeit ist

$$M \text{ ist unendlich} \iff \exists f\colon \mathbb{N} \to M \quad \text{mit } f \text{ injektiv}.$$

Eine alternative Definition ist

$$M \text{ ist unendlich} \iff \exists f\colon M \to M \quad \text{mit } f \text{ injektiv, aber nicht surjektiv}.$$

Es ist eine interessante Übung, die Äquivalenz dieser beiden Definitionen zu zeigen. Man zeigt wieder durch vollständige Induktion mit der zweiten Definition, dass die Mengen $\mathbb{N}_{<n}$ nicht unendlich sind (eine injektive Abbildung $\mathbb{N}_{<n} \to \mathbb{N}_{<n}$ ist immer auch bijektiv). Es ist dann noch zu zeigen, dass eine nicht unendliche Menge M bijektiv zu einer der Mengen $\mathbb{N}_{<n}$ ist. Dazu definiert man eine injektive Abbildung f von einem Anfangsstück von $\mathbb{N}$ nach M: Falls M leer ist, dann ist die Abbildung leer; das Anfangsstück ist $\mathbb{N}_{<0}$ und $\#M = 0$. Sonst gibt es ein $m_0 \in M$; wir setzen $f(0) = m_0$ und $M_1 = M \setminus \{m_0\}$. Wenn f schon auf $\mathbb{N}_{<n}$ definiert und parallel M_n konstruiert ist, dann ist entweder M_n leer; in diesem Fall haben wir eine Bijektion $f\colon \mathbb{N}_{<n} \to M$ und $\#M = n$; oder es gibt ein $m_n \in M_n$, dann setzen wir $f(n) = m_n$ und $M_{n+1} = M_n \setminus \{m_n\}$. Damit ist f auf $\mathbb{N}_{<n+1}$ definiert. Wenn diese Konstruktion nicht abbricht, dann erhalten wir eine injektive Abbildung $f\colon \mathbb{N} \to M$, also muss M unendlich sein.

Allgemeiner kann man für beliebige (auch unendliche) Mengen M_1 und M_2 definieren

$$\#M_1 \le \#M_2 \iff \exists f\colon M_1 \to M_2 \quad \text{mit } f \text{ injektiv}.$$

Nach dem Satz von Cantor-Bernstein-Schröder folgt aus $\#M_1 \le \#M_2$ und $\#M_2 \le \#M_1$, dass es eine Bijektion zwischen M_1 und M_2 gibt; deshalb ist es sinnvoll zu definieren

$$\#M_1 = \#M_2 \iff \exists f\colon M_1 \to M_2 \quad \text{mit } f \text{ bijektiv}$$

(man sagt dann auch, M_1 und M_2 seien *gleichmächtig*). Damit kann man dann auch bei unendlichen Mengen zwischen verschiedenen Mächtigkeiten oder *Kardinalitäten* unterscheiden. Zum Beispiel gilt für jede unendliche Menge M, dass $\#\mathbb{N} \le \#M$ ist (das war eine der beiden Definitionen oben); es gibt aber auch „*überabzählbare*“ Mengen M mit $\#\mathbb{N} < \#M$ (also $\#\mathbb{N} \le \#M$ und $\#\mathbb{N} \ne \#M$). In der Analysis lernen Sie, dass $M = \mathbb{R}$ dafür ein Beispiel ist. Allgemein gilt für jede Menge M, dass $\#M < \#\mathcal{P}(M)$ ist; es gibt zu jeder Menge also eine noch „größere“.

Wir können jetzt die Dimension eines Vektorraums einführen.

* 9.12. **Definition.** Sei V ein Vektorraum. Wenn V eine endliche Basis $(v_1, \dots, v_n)$ hat, dann sagen wir, dass V *Dimension* n hat oder n*-dimensional* ist und schreiben $\dim V = n$. Hat V keine endliche Basis, dann sagen wir, dass V *unendlich-dimensional* ist und schreiben $\dim V = \infty$. Hat V Dimension n für ein $n \in \mathbb{N}$, dann heißt V *endlich-dimensional* und wir schreiben $\dim V < \infty$. **DEF** Dimension

Wenn wir betonen wollen, dass es um die Dimension von V als K-Vektorraum geht, dann schreiben wir genauer $\dim_K V$. ♢

Zum Beispiel ist $\dim_{\mathbb{C}} \mathbb{C} = 1$ ($\mathbb{C}$-Basis (1)), aber $\dim_{\mathbb{R}} \mathbb{C} = 2$ ($\mathbb{R}$-Basis $(1, \boldsymbol{i})$).

Folgerung 9.10 sagt uns, dass diese Definition sinnvoll ist, weil alle endlichen Basen von V (wenn es sie gibt) dieselbe Anzahl von Elementen haben.

9.13. **Beispiele.** **BSP** Dimension

(1) Die leere Menge ist Basis des Null-Vektorraums, also ist $\dim\{\mathbf{0}\} = 0$. Ist umgekehrt V ein Vektorraum mit $\dim V = 0$, dann hat V eine Basis aus null Vektoren, also ist $V = \{\mathbf{0}\}$.

(2) Für $n \in \mathbb{N}$ gilt $\dim K^n = n$, denn K^n hat die n-elementige Standardbasis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$.

(3) Für den Vektorraum der Polynomfunktionen gilt $\dim P = \infty$, denn er hat eine unendliche Basis und kann deswegen nicht endlich-dimensional sein (siehe Folgerung 9.15 unten). ♣

Die Dimension eines Vektorraums ist eine wichtige Größe, wie die folgenden Aussagen zeigen.

9.14. **Satz.** *Seien $m, n \in \mathbb{N}$, sei V ein n-dimensionaler Vektorraum und seien $v_1, v_2, \ldots, v_m \in V$.* **SATZ** Eigensch. Dimension

(1) *Wenn $(v_1, v_2, \ldots, v_m)$ linear unabhängig ist, dann ist $m \leq n$. Im Fall $m = n$ ist $(v_1, v_2, \ldots, v_m)$ eine Basis von V.*

(2) *Wenn $(v_1, v_2, \ldots, v_m)$ ein Erzeugendensystem von V ist, dann ist $m \geq n$. Im Fall $m = n$ ist $(v_1, v_2, \ldots, v_m)$ eine Basis von V.*

Beweis.

(1) Nach Folgerung 9.7 können wir $(v_1, v_2, \ldots, v_m)$ durch Hinzunehmen von geeigneten Vektoren von V zu einer Basis von V ergänzen. Diese Basis hat n Elemente, also gilt $m \leq n$. Wenn $m = n$ ist, dann werden keine Elemente hinzugefügt, also liegt bereits eine Basis vor.

(2) Nach dem Basisergänzungssatz 9.5 (mit $n = 0$ in der dortigen Notation) gibt es eine Basis, die durch Weglassen von geeigneten Vektoren v_j aus $(v_1, v_2, \ldots, v_m)$ entsteht. Diese Basis hat Länge n, also gilt $m \geq n$. Wenn $m = n$ ist, dann kann nichts weggelassen werden, also liegt bereits eine Basis vor. ❑

Weil dieser Satz so wichtig ist, gebe ich eine weitere Formulierung.

Man kann den ersten Teil der beiden Aussagen auch so ausdrücken:

(1) *In einem **n-dimensionalen** Vektorraum sind **mehr als n** Vektoren **immer linear abhängig**.*

(2) *Die **lineare Hülle** von **m** Vektoren hat **Dimension höchstens m**:*

$$\dim\langle v_1, v_2, \ldots, v_m\rangle \leq m\,.$$

Die erste dieser Aussagen ist eine starke *Existenzaussage*. Sie besagt nämlich Folgendes: *Sind $v_1, v_2, \ldots, v_m \in V$ mit $m > \dim V$, dann gibt es eine nichttriviale Linearkombination*

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_m v_m = \mathbf{0}\,.$$

Der zweite Teil der beiden Aussagen im Satz oben bedeutet:

(1) *In einem n-dimensionalen Vektorraum sind n linear unabhängige Vektoren immer schon eine Basis.*

(2) *In einem n-dimensionalen Vektorraum ist ein Erzeugendensystem mit n Elementen immer schon eine Basis.*

Linear unabhängige Familien geben also untere Schranken und Erzeugendensysteme geben obere Schranken für die Dimension. Es ist daher plausibel, dass wir unendlich-dimensionale Vektorräume wie folgt charakterisieren können.

9.15. **Folgerung.** *Sei V ein Vektorraum. Die folgenden Aussagen sind äquivalent:*

FOLG dim = ∞

(1) *Es gibt in V eine (unendliche) Folge $(v_n)_{n\in\mathbb{N}}$ linear unabhängiger Vektoren.*

(2) $\dim V = \infty$.

Beweis. „(1) ⇒ (2)“: Sei $(v_n)_{n\in\mathbb{N}}$ linear unabhängig. Wenn $\dim V = m < \infty$ wäre, dann müssten nach Satz 9.14 die $m+1$ Vektoren $v_0, v_1, \ldots, v_m$ linear abhängig sein, was aber der linearen Unabhängigkeit von $(v_n)_{n\in\mathbb{N}}$ widerspricht. Also muss V unendlich-dimensional sein.

„(2) ⇒ (1)“: Sei V unendlich-dimensional. Das bedeutet, dass V keine endliche Basis hat; damit kann eine endliche linear unabhängige Teilmenge von V kein Erzeugendensystem sein. Wir konstruieren rekursiv eine linear unabhängige Folge $(v_n)_{n\in\mathbb{N}}$ in V. Sei dazu $(v_0, v_1, \ldots, v_{n-1})$ bereits konstruiert und linear unabhängig. (Für $n = 0$ ist das das leere Tupel, das immer linear unabhängig ist.) Da $(v_0, v_1, \ldots, v_{n-1})$ kein Erzeugendensystem ist, gibt es $v_n \in V \setminus \langle v_0, v_1, \ldots, v_{n-1}\rangle$. Nach Lemma 8.9 ist dann $(v_0, \ldots, v_{n-1}, v_n)$ linear unabhängig. Die so konstruierte Folge $(v_n)_{n\in\mathbb{N}}$ ist linear unabhängig, weil das für alle endlichen Anfangsstücke gilt (in einer Linearkombination kommen nur endlich viele Vektoren v_n vor). ❑

Wir können also sagen:

- Die Dimension von V ist die **maximale** Anzahl **linear unabhängiger** Vektoren in V.
- Die Dimension von V ist die **minimale** Anzahl von **Erzeugern** von V.

Hier ist eine Anwendung der Aussage, dass $n+1$ Vektoren in einem n-dimensionalen Vektorraum linear abhängig sein müssen.

9.16. **Definition.** Wir sagen, eine Polynomfunktion $f \in P$ habe *Grad* $\leq n$ (und wir schreiben $\deg(f) \leq n$), wenn sie in der Form

DEF Grad einer Polynomfkt.

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

(mit $a_0, a_1, \ldots, a_n \in \mathbb{R}$) geschrieben werden kann. f hat *Grad* n ($\deg(f) = n$), wenn $a_n \neq 0$ ist, wenn also f nicht Grad $\leq n-1$ hat. f hat *Grad* $< n$ ($\deg(f) < n$), wenn f Grad $\leq n-1$ hat. ♢

Sie wissen aus der Schule, dass eine Polynomfunktion vom Grad n höchstens n reelle Nullstellen haben kann. Das kann man auch so ausdrücken:

9.17. **Lemma.** *Ist f eine Polynomfunktion mit $\deg(f) < n$, die mindestens n reelle Nullstellen hat, dann ist f die Nullfunktion.* **LEMMA** Polynom = 0

Beweisen kann man das zum Beispiel mit Mitteln der Analysis. Dazu verwendet man einerseits, dass die Ableitung einer Polynomfunktion vom Grad $< n$ eine Polynomfunktion vom Grad $< n-1$ ist, und andererseits den *Satz von Rolle:* Zwischen zwei Nullstellen einer differenzierbaren Funktion f liegt eine Nullstelle der Ableitung f'.

Der Beweis geht dann durch Induktion über n. Ist $n = 0$, dann ist f schon die Nullfunktion (denn „$\deg(f) < 0$“ bedeutet, dass in der Darstellung von f als $f(x) = a_0 + a_1 x + \ldots$ kein von null verschiedener Koeffizient a_j auftreten kann). Für den Induktionsschritt nehmen wir an, dass die Aussage für Polynomfunktionen vom Grad $< n$ gilt. Sei f eine Polynomfunktion vom Grad $< n+1$ mit mindestens $n+1$ reellen Nullstellen. Dann ist die Ableitung f' eine Polynomfunktion vom Grad $< n$ mit mindestens n reellen Nullstellen (nach dem Satz von Rolle liegt zwischen je zwei aufeinanderfolgenden Nullstellen von f eine Nullstelle von f'). Die Induktionsannahme liefert jetzt $f' = 0$. Dann muss f konstant sein; weil f aber mindestens eine Nullstelle hat ($n + 1 > 0$), muss $f = 0$ sein.

9.18. **Beispiel.** *Seien $x_1, \ldots, x_n \in \mathbb{R}$ paarweise verschieden und $y_1, \ldots, y_n \in \mathbb{R}$. Dann gibt es eine Polynomfunktion f mit $\deg(f) < n$, sodass $f(x_j) = y_j$ ist für alle $j \in \{1, 2, \ldots, n\}$.* **BSP** Interpolation

Beweis. Wir betrachten die folgenden $n+1$ Vektoren in $\mathbb{R}^n$:

$$\begin{aligned} v_0 &= (1, 1, 1, \ldots, 1) \\ v_1 &= (x_1, x_2, x_3, \ldots, x_n) \\ v_2 &= (x_1^2, x_2^2, x_3^2, \ldots, x_n^2) \\ &\vdots \qquad\qquad \vdots \\ v_{n-1} &= (x_1^{n-1}, x_2^{n-1}, x_3^{n-1}, \ldots, x_n^{n-1}) \\ v_n &= (y_1, y_2, y_3, \ldots, y_n) \end{aligned}$$

Dann wissen wir, dass $v_0, v_1, \ldots, v_n$ linear abhängig sein müssen, denn es ist $\dim \mathbb{R}^n = n < n+1$. Es gibt also $\lambda_0, \lambda_1, \ldots, \lambda_n \in \mathbb{R}$, nicht alle null, mit

$$\lambda_0 + \lambda_1 x_j + \lambda_2 x_j^2 + \ldots + \lambda_{n-1} x_j^{n-1} + \lambda_n y_j = 0$$

für alle $j \in \{1, 2, \ldots, n\}$. Ich behaupte, dass λ_n nicht null sein kann. Denn sonst hätte die Polynomfunktion

$$x \longmapsto \lambda_0 + \lambda_1 x + \ldots + \lambda_{n-1} x^{n-1}$$

vom Grad $< n$ mindestens die n Nullstellen $x_1, x_2, \ldots, x_n$, müsste also nach Lemma 9.17 die Nullfunktion sein, was $\lambda_0 = \lambda_1 = \ldots = \lambda_{n-1} = 0$ bedeuten würde. Dann wäre die obige Linearkombination aber trivial, ein Widerspruch. Also ist $\lambda_n \neq 0$. Wir setzen

$$a_0 = -\frac{\lambda_0}{\lambda_n}, \quad a_1 = -\frac{\lambda_1}{\lambda_n}, \quad \ldots, \quad a_{n-1} = -\frac{\lambda_{n-1}}{\lambda_n}$$

und

$$f(x) = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1};$$

dann gilt

$$f(x_j) = a_0 + a_1 x_j + \ldots + a_{n-1} x_j^{n-1} = y_j$$

wie gewünscht. ❑

Als Nebenprodukt unserer Überlegungen ergab sich, dass die Vektoren $v_j = (x_1^j, \dots, x_n^j)$ für $j \in \{0, 1, \dots, n-1\}$ linear unabhängig sind. ♣

Die Dimension ist ein Maß für die „Größe“ eines Vektorraums. Das wird deutlich, wenn man die Dimension eines Untervektorraums betrachtet.

SATZ Dimension von Unter-VR

9.19. **Satz.** *Seien V ein Vektorraum und $U \subset V$ ein Untervektorraum. Dann gilt $\dim U \leq \dim V$. Ist V endlich-dimensional und gilt $\dim U = \dim V$, dann ist $U = V$.*

Dabei gelte $n \leq \infty$ für alle $n \in \mathbb{N}$ und $\infty \leq \infty$.

Beweis. Im Fall $\dim V = \infty$ ist die Aussage trivialerweise richtig. Sei jetzt also $\dim V = n \in \mathbb{N}$. Wäre $\dim U = \infty$, dann gäbe es nach Folgerung 9.15 unendlich viele linear unabhängige Elemente in U und damit auch in V, ein Widerspruch. Also ist U endlich-dimensional mit $\dim U = m \in \mathbb{N}$. Eine Basis von U besteht aus m linear unabhängigen Vektoren von V. Nach Satz 9.14 folgt $\dim U = m \leq n = \dim V$. Gilt $m = n$, dann ist die Basis von U bereits eine Basis von V und es folgt $U = V$. ❑

BSP $\dim U = \dim V = \infty$ $U \subsetneq V$

9.20. **Beispiel.** Ein unendlich-dimensionaler Vektorraum kann durchaus echte Untervektorräume haben, die ihrerseits unendlich-dimensional sind. Zum Beispiel können wir im Vektorraum P der Polynomfunktionen den Untervektorraum P_g der geraden Polynomfunktionen betrachten:

$$P_g = \{f \in P \mid \forall x \in \mathbb{R}\colon f(-x) = f(x)\}\,.$$

(Prüfen Sie nach, dass P_g tatsächlich ein Untervektorraum von P ist!) Da die Funktion $x \mapsto x$, die ein Element von P ist, nicht in P_g liegt, gilt $P_g \neq P$. Auf der anderen Seite sind die geraden Potenzfunktionen $x \mapsto x^{2n}$ für $n \in \mathbb{N}$ alle linear unabhängig, also ist $\dim P_g = \infty$. ♣

10. Lineare Abbildungen

Sei V ein K-Vektorraum und seien $v_1, v_2, \ldots, v_n \in V$. Sei weiter $\phi = \phi_{(v_1,\ldots,v_n)}$ die zugehörige Linearkombinationenabbildung

$$\phi\colon K^n \longrightarrow V, \qquad (x_1, x_2, \ldots, x_n) \longmapsto x_1v_1 + x_2v_2 + \ldots + x_nv_n\,.$$

Dann gilt für $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in K^n$, $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in K^n$ und $\lambda \in K$:

$$\begin{aligned}
\phi(\boldsymbol{x} + \boldsymbol{y}) &= \phi\big((x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n)\big)\\
&= \phi(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)\\
&= (x_1 + y_1)v_1 + (x_2 + y_2)v_2 + \ldots + (x_n + y_n)v_n\\
&= (x_1v_1 + x_2v_2 + \ldots + x_nv_n) + (y_1v_1 + y_2v_2 + \ldots + y_nv_n)\\
&= \phi(x_1, x_2, \ldots, x_n) + \phi(y_1, y_2, \ldots, y_n)\\
&= \phi(\boldsymbol{x}) + \phi(\boldsymbol{y})
\end{aligned}$$

und

$$\begin{aligned}
\phi(\lambda\boldsymbol{x}) &= \phi\big(\lambda(x_1, x_2, \ldots, x_n)\big)\\
&= \phi(\lambda x_1, \lambda x_2, \ldots, \lambda x_n)\\
&= (\lambda x_1)v_1 + (\lambda x_2)v_2 + \ldots + (\lambda x_n)v_n\\
&= \lambda(x_1v_1 + x_2v_2 + \ldots + x_nv_n)\\
&= \lambda\phi(x_1, x_2, \ldots, x_n)\\
&= \lambda\phi(\boldsymbol{x})\,.
\end{aligned}$$

(Man beachte, dass Addition und Skalarmultiplikation hier einmal in K^n und einmal in V stattfinden.)

Die Abbildung ϕ ist also mit Addition und Skalarmultiplikation verträglich: Das Bild einer Summe ist die Summe der Bilder und das Bild eines skalaren Vielfachen ist das entsprechende Vielfache des Bildes. Solche mit der linearen Struktur verträgliche Abbildungen heißen *lineare Abbildungen*.

* 10.1. **Definition.** Sei K ein Körper und seien V_1 und V_2 zwei K-Vektorräume. Eine Abbildung $\phi\colon V_1 \to V_2$ heißt *(K-)linear* oder ein *Homomorphismus (von K-Vektorräumen)*, wenn sie die folgenden beiden Bedingungen erfüllt: **DEF** Lineare Abbildung Homomorphismus

(1) $\forall v, w \in V_1\colon \phi(v + w) = \phi(v) + \phi(w)$.

(2) $\forall \lambda \in K\ \forall v \in V_1\colon \phi(\lambda v) = \lambda\phi(v)$.

Eine lineare Abbildung heißt ein *Monomorphismus*, wenn sie injektiv ist, ein *Epimorphismus*, wenn sie surjektiv ist, und ein *Isomorphismus*, wenn sie bijektiv ist. Eine lineare Abbildung $\phi\colon V \to V$ heißt ein *Endomorphismus* von V; ϕ heißt ein *Automorphismus* von V, wenn ϕ außerdem bijektiv ist. Zwei Vektorräume V_1 und V_2 heißen (zueinander) *isomorph*, $V_1 \cong V_2$, wenn es einen Isomorphismus $\phi\colon V_1 \to V_2$ gibt. ♢ Mono-, Epi-, Iso-, Endo-, Automorph. isomorph

10.2. **Beispiele.** **BSP** lineare Abbildungen

(1) Für beliebige K-Vektorräume V_1 und V_2 ist die *Nullabbildung* $V_1 \to V_2$, $v \mapsto \mathbf{0}$, eine lineare Abbildung.

(2) Für jeden K-Vektorraum V ist die identische Abbildung $\mathrm{id}_V\colon V \to V$ ein Automorphismus von V.

(3) Ist V ein Vektorraum und $U \subset V$ ein Untervektorraum, dann ist die Inklusionsabbildung $U \to V$ linear. ♣

BSP Linearkomb.-Abbildung ist linear

10.3. **Beispiel.** Seien V, $v_1, v_2, \ldots, v_n$ und ϕ wie zum Beginn dieses Abschnitts. Dann ist ϕ ein Homomorphismus. Nach Lemma 9.4 gilt außerdem:

- $(v_1, v_2, \ldots, v_n)$ ist linear unabhängig $\iff$ ϕ ist ein Monomorphismus.
- $(v_1, v_2, \ldots, v_n)$ ist ein Erzeugendensystem von V $\iff$ ϕ ist ein Epimorphismus.
- $(v_1, v_2, \ldots, v_n)$ ist eine Basis von V $\iff$ ϕ ist ein Isomorphismus.

Aus dem letzten Punkt ergibt sich die Aussage

$$\dim V < \infty \implies V \cong K^{\dim V}. \qquad \clubsuit$$

Wir überzeugen uns noch davon, dass eine lineare Abbildung wirklich mit der gesamten Struktur verträglich ist und dass sich lineare Abbildungen bezüglich Komposition und Inversion gut verhalten.

LEMMA Lin. Abb.: Eigensch.

10.4. **Lemma.** *V_1, V_2 und V_3 seien K-Vektorräume.*

(1) *Sei $\phi\colon V_1 \to V_2$ eine lineare Abbildung. Dann gilt*
$$\phi(\mathbf{0}) = \mathbf{0} \qquad \text{und} \qquad \forall v \in V_1\colon \phi(-v) = -\phi(v).$$

(2) *Seien $\phi_1\colon V_1 \to V_2$ und $\phi_2\colon V_2 \to V_3$ lineare Abbildungen. Dann ist auch $\phi_2 \circ \phi_1\colon V_1 \to V_3$ linear.*

(3) *Sei $\phi\colon V_1 \to V_2$ ein Isomorphismus. Dann ist die Umkehrabbildung $\phi^{-1}\colon V_2 \to V_1$ ebenfalls ein Isomorphismus.*

Teil (3) zeigt, dass es in der Definition von „isomorph“ nicht darauf ankommt, ob man einen Isomorphismus $V_1 \to V_2$ oder einen Isomorphismus $V_2 \to V_1$ fordert.

Beweis.

(1) Es gilt $\phi(\mathbf{0}) = \phi(\mathbf{0} + \mathbf{0}) = \phi(\mathbf{0}) + \phi(\mathbf{0})$. Durch Addition von $-\phi(\mathbf{0})$ folgt $\phi(\mathbf{0}) = \mathbf{0}$.
Außerdem gilt für $v \in V_1$: $\phi(-v) = \phi((-1)v) = (-1)\phi(v) = -\phi(v)$.

(2) Wir müssen die beiden Eigenschaften aus Definition 10.1 für $\phi_2 \circ \phi_1$ nachweisen. Seien dazu $v, w \in V_1$ und $\lambda \in K$. Dann gilt
$$\begin{aligned}(\phi_2 \circ \phi_1)(v + w) &= \phi_2\big(\phi_1(v + w)\big) = \phi_2\big(\phi_1(v) + \phi_1(w)\big) \\ &= \phi_2\big(\phi_1(v)\big) + \phi_2\big(\phi_1(w)\big) = (\phi_2 \circ \phi_1)(v) + (\phi_2 \circ \phi_1)(w)\end{aligned}$$
und
$$(\phi_2 \circ \phi_1)(\lambda v) = \phi_2\big(\phi_1(\lambda v)\big) = \phi_2\big(\lambda\phi_1(v)\big) = \lambda\phi_2\big(\phi_1(v)\big) = \lambda(\phi_2 \circ \phi_1)(v).$$

(3) Wir weisen die Eigenschaften aus Definition 10.1 für ϕ^{-1} nach. Seien dazu $v, w \in V_2$ und $\lambda \in K$. Wir setzen $v' = \phi^{-1}(v)$ und $w' = \phi^{-1}(w)$, sodass $v = \phi(v')$ und $w = \phi(w')$. Dann gilt
$$\phi^{-1}(v + w) = \phi^{-1}\big(\phi(v') + \phi(w')\big) = \phi^{-1}\big(\phi(v' + w')\big) = v' + w' = \phi^{-1}(v) + \phi^{-1}(w)$$
und
$$\phi^{-1}(\lambda v) = \phi^{-1}\big(\lambda\phi(v')\big) = \phi^{-1}\big(\phi(\lambda v')\big) = \lambda v' = \lambda\phi^{-1}(v). \qquad \square$$

Bevor wir weitere Eigenschaften untersuchen, führen wir noch eine Schreibweise ein.

DEF Bilder und Urbilder von Teilmengen

10.5. **Definition.** Sei $f\colon X \to Y$ eine Abbildung zwischen beliebigen Mengen X und Y. Ist T eine Teilmenge von X, dann schreiben wir

$$f(T) = \{f(x) \mid x \in T\} \subset Y$$

für die Menge der Bilder der Elemente von T und nennen $f(T)$ das *Bild von T unter f*. Im Spezialfall $T = X$ schreiben wir auch $\operatorname{im}(f)$ für $f(X)$; $\operatorname{im}(f)$ heißt das *Bild* oder die *Bildmenge* von f. Ist U eine Teilmenge von Y, dann schreiben wir

$$f^{-1}(U) = \{x \in X \mid f(x) \in U\} \subset X$$

für die Menge der Urbilder der Elemente von U und nennen $f^{-1}(U)$ das *Urbild von U unter f*. ♢

Häufig wird auch $f^{-1}(y) = \{x \in X \mid f(x) = y\}$ für die Menge der Urbilder eines *Elements* $y \in Y$ geschrieben. Das kann zu Verwirrung führen, denn wenn f bijektiv ist, dann bedeutet $f^{-1}(y)$ auch das Bild von y unter der Umkehrabbildung, also das Urbild von y und nicht die Urbild*menge* von y. Wir werden die „Datentypen“ (Elemente bzw. Teilmengen) hier aber sorgfältig auseinanderhalten und immer $f^{-1}(\{y\})$ für diese Menge schreiben.

Wenn f bijektiv ist, dann hat $f^{-1}(U)$ zwei mögliche Bedeutungen: einerseits ausgehend von f wie oben definiert und andererseits ausgehend von der Umkehrfunktion f^{-1}. Zum Glück stimmen beide Versionen überein.

Noch eine **Warnung:** Die hier eingeführte Schreibweise kann einen dazu verführen zu denken, dass $f^{-1}(f(T)) = T$ und $f(f^{-1}(U)) = U$ sein muss. Das ist aber im Allgemeinen **falsch**! Es gilt immer $f^{-1}(f(T)) \supset T$ und $f(f^{-1}(U)) \subset U$; die Inklusionen können jedoch echt sein.

Da der Nullvektor eine ausgezeichnete Rolle in einem Vektorraum spielt, ist die Menge seiner Urbilder unter einer linearen Abbildung ein wichtiges Datum.

DEF Kern

* 10.6. **Definition.** Sei $\phi\colon V_1 \to V_2$ eine lineare Abbildung. Der *Kern* von ϕ ist die Menge der Urbilder von $\mathbf{0} \in V_2$:

$$\ker(\phi) = \phi^{-1}(\{\mathbf{0}\}) = \{v \in V_1 \mid \phi(v) = \mathbf{0}\} \subset V_1\,. \qquad \diamondsuit$$

Nach Lemma 10.4 gilt stets $\mathbf{0} \in \ker(\phi)$.

BSP Limes ist linear

10.7. **Beispiel.** Sei $V \subset \mathbb{R}^{\mathbb{N}}$ der Vektorraum der konvergenten Folgen. Dann ist

$$\lim\colon V \longrightarrow \mathbb{R}, \qquad (a_n)_{n\in\mathbb{N}} \longmapsto \lim_{n\to\infty} a_n$$

eine $\mathbb{R}$-lineare Abbildung. Das folgt aus den Rechenregeln für Grenzwerte.

Der Kern $\ker(\lim)$ ist gerade die Menge der Nullfolgen, denn das sind definitionsgemäß die Folgen mit Limes null. ♣

Eine wichtige Eigenschaft des Kerns ist die folgende:

10.8. **Lemma.** *Sei $\phi\colon V_1 \to V_2$ eine lineare Abbildung. Dann gilt:*

$$\phi \text{ ist injektiv} \iff \ker(\phi) = \{\mathbf{0}\}\,.$$

LEMMA
injektiv
$\iff$
$\ker = \{\mathbf{0}\}$

Man sagt in diesem Fall auch, der Kern sei *trivial*.

DEF
Kern
trivial

Beweis. „$\Rightarrow$“: Sei ϕ injektiv und $v \in \ker(\phi)$. Dann ist $\phi(v) = \mathbf{0} = \phi(\mathbf{0})$, also $v = \mathbf{0}$.
„$\Leftarrow$“: Es gelte $\ker(\phi) = \{\mathbf{0}\}$. Seien weiter $v, w \in V_1$ mit $\phi(v) = \phi(w)$. Dann folgt $\mathbf{0} = \phi(v) - \phi(w) = \phi(v - w)$, also ist $v - w \in \ker(\phi) = \{\mathbf{0}\}$ und damit $v - w = \mathbf{0}$; das bedeutet $v = w$. ❑

Wie zu erwarten, vertragen sich lineare Abbildungen sehr gut mit Untervektorräumen.

10.9. **Satz.** *Sei $\phi\colon V_1 \to V_2$ eine K-lineare Abbildung.*

SATZ
lin. Abb.
und UVR

(1) *Ist $U_1 \subset V_1$ ein Untervektorraum, dann ist $\phi(U_1) \subset V_2$ wieder ein Untervektorraum. Insbesondere ist $\mathrm{im}(\phi) = \phi(V_1) \subset V_2$ ein Untervektorraum von V_2. Außerdem ist die auf U_1 eingeschränkte Abbildung $\phi|_{U_1}\colon U_1 \to V_2$ ebenfalls linear.*

(2) *Ist $U_2 \subset V_2$ ein Untervektorraum, dann ist $\phi^{-1}(U_2) \subset V_1$ wieder ein Untervektorraum. Insbesondere ist $\ker(\phi) = \phi^{-1}(\{\mathbf{0}\}) \subset V_1$ ein Untervektorraum von V_1.*

(3) *Seien*

$$M_1 = \{U_1 \mid U_1 \subset V_1 \text{ Untervektorraum mit } \ker(\phi) \subset U_1\} \qquad \text{und}$$
$$M_2 = \{U_2 \mid U_2 \subset V_2 \text{ Untervektorraum mit } U_2 \subset \mathrm{im}(\phi)\}\,.$$

Die Abbildungen $M_1 \to M_2$, $U_1 \mapsto \phi(U_1)$, und $M_2 \to M_1$, $U_2 \mapsto \phi^{-1}(U_2)$, sind zueinander inverse Bijektionen.

Die Aussage, dass der Kern einer linearen Abbildung ein Untervektorraum ist, ist oft nützlich, wenn man zeigen möchte, dass eine Teilmenge eines Vektorraums ein Untervektorraum ist. Oft kann man nämlich Untervektorräume in natürlicher Weise als Kerne schreiben.

Beweis.

(1) Wir müssen die Bedingungen für einen Untervektorraum für $\phi(U_1)$ nachprüfen:
- $\mathbf{0} = \phi(\mathbf{0}) \in \phi(U_1)$, da $\mathbf{0} \in U_1$.
- Seien $v, w \in \phi(U_1)$. Dann gibt es $v', w' \in U_1$ mit $\phi(v') = v$, $\phi(w') = w$. Es folgt $v + w = \phi(v') + \phi(w') = \phi(v' + w') \in \phi(U_1)$, denn $v' + w' \in U_1$.
- Sei $v \in \phi(U_1)$ und $\lambda \in K$. Dann gibt es $v' \in U_1$ mit $\phi(v') = v$. Es folgt $\lambda v = \lambda\phi(v') = \phi(\lambda v') \in \phi(U_1)$, denn $\lambda v' \in U_1$.

Da V_1 ein Untervektorraum von V_1 ist, folgt, dass $\mathrm{im}(\phi)$ ein Untervektorraum von V_2 ist.
Dass $\phi|_{U_1}$ linear ist, folgt aus Definition 10.1, da die geforderten Eigenschaften die Form „für alle ...“ haben.

(2) Wir prüfen die Bedingungen für $\phi^{-1}(U_2)$:
- $\mathbf{0} \in \phi^{-1}(U_2)$, da $\phi(\mathbf{0}) = \mathbf{0} \in U_2$.

- Seien $v, w \in \phi^{-1}(U_2)$. Dann sind $\phi(v), \phi(w) \in U_2$. Es folgt $\phi(v+w) = \phi(v) + \phi(w) \in U_2$ und damit $v + w \in \phi^{-1}(U_2)$.
- Seien $v \in \phi^{-1}(U_2)$ und $\lambda \in K$. Dann ist $\phi(v) \in U_2$, also auch $\phi(\lambda v) = \lambda\phi(v) \in U_2$, und damit $\lambda v \in \phi^{-1}(U_2)$.

Da $\{\mathbf{0}\}$ ein Untervektorraum von V_2 ist, folgt, dass $\ker(\phi)$ ein Untervektorraum von V_1 ist.

(3) Wir überlegen zunächst, dass die Abbildungen wohldefiniert sind: Für $U_1 \subset V_1$ gilt $\phi(U_1) \subset \phi(V_1) = \operatorname{im}(\phi)$ und für $U_2 \subset V_2$ gilt $\phi^{-1}(U_2) \supset \phi^{-1}(\{\mathbf{0}\}) = \ker(\phi)$. Nach den Teilen (1) und (2) werden Untervektorräume auf Untervektorräume abgebildet. Damit haben wir tatsächlich Abbildungen zwischen den beiden angegebenen Mengen.

Wir zeigen jetzt, dass die Abbildungen zueinander invers sind. Daraus folgt dann auch, dass sie bijektiv sind. Sei also $U_1 \subset V_1$ ein Untervektorraum mit $\ker(\phi) \subset U_1$. Dann gilt

$$\begin{aligned} v \in \phi^{-1}\big(\phi(U_1)\big) &\iff \phi(v) \in \phi(U_1) \\ &\iff \exists v' \in U_1 \colon \phi(v) = \phi(v') \\ &\iff \exists v' \in U_1 \colon \phi(v - v') = \mathbf{0} \\ &\iff \exists v' \in U_1 \colon v - v' \in \ker(\phi) \\ &\iff v \in U_1 \,. \end{aligned}$$

(Die letzte Äquivalenz sieht man so: „$\Leftarrow$“: wähle $v' = v$. „$\Rightarrow$“: Sei $v'' = v - v' \in \ker(\phi) \subset U_1$, dann ist $v = v' + v'' \in U_1$.) Das zeigt $\phi^{-1}\big(\phi(U_1)\big) = U_1$.

Sei jetzt $U_2 \subset \operatorname{im}(\phi)$ ein Untervektorraum von V_2. Dann gilt

$$\begin{aligned} v \in \phi\big(\phi^{-1}(U_2)\big) &\iff \exists v' \in \phi^{-1}(U_2) \colon \phi(v') = v \\ &\iff \exists v' \in V_1 \colon \phi(v') \in U_2 \text{ und } \phi(v') = v \\ &\iff v \in U_2 \quad \text{und} \quad v \in \operatorname{im}(\phi) \\ &\iff v \in U_2 \cap \operatorname{im}(\phi) \\ &\iff v \in U_2 \,. \end{aligned}$$

Das zeigt $\phi\big(\phi^{-1}(U_2)\big) = U_2$. ❑

Den Zusammenhang, der im dritten Teil dieses Satzes formuliert wird, kann man sich etwa (sehr schematisch) wie in Abbildung 2 veranschaulichen.

BSP Auswertungsabbildung

10.10. **Beispiel.** Seien K ein Körper, X eine Menge und V ein Untervektorraum von $K^X = \operatorname{Abb}(X, K)$ (zum Beispiel können wir $X = K = \mathbb{R}$ setzen und für V den Vektorraum der stetigen reellen Funktionen nehmen). Sei weiter $x \in X$. Dann ist die *Auswertungsabbildung*

DEF Auswertungsabbildung

$$\operatorname{ev}_x \colon V \longrightarrow K, \quad f \longmapsto f(x)$$

linear. Das ergibt sich direkt aus der Definition der Addition und der skalaren Multiplikation von Funktionen:

$$\begin{aligned} \operatorname{ev}_x(f+g) = (f+g)(x) = f(x) + g(x) = \operatorname{ev}_x(f) + \operatorname{ev}_x(g) \quad \text{und} \\ \operatorname{ev}_x(\lambda f) = (\lambda f)(x) = \lambda f(x) = \lambda \operatorname{ev}_x(f) \,. \end{aligned}$$

(Man kann sagen, dass die Addition und Skalarmultiplikation in K^X gerade so definiert sind, *damit* die Auswertungsabbildungen linear werden!)

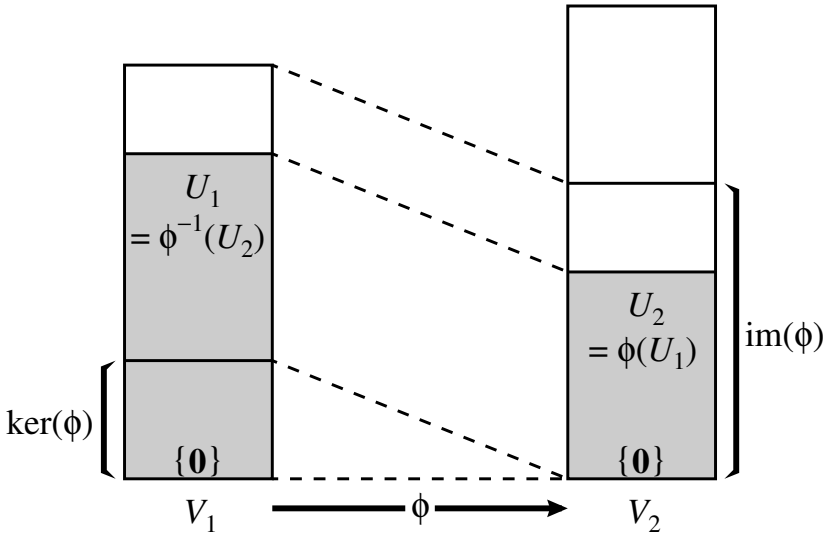


Abbildung 2. Skizze zu Satz 10.9

Sei T eine Teilmenge von X. Dann ist

$$\{f \in V \mid \forall x \in T\colon f(x) = 0\} = \bigcap_{x \in T} \ker(\mathrm{ev}_x)$$

ein Untervektorraum von V.

Im Spezialfall $X = \{1, 2, \ldots, n\}$ haben wir $K^X = K^n$; dann heißen die Abbildungen ev_j (für $j \in \{1, 2, \ldots, n\}$) *Projektionen* und werden pr_j geschrieben: **DEF** Projektion

$$\mathrm{pr}_j\colon K^n \longrightarrow K, \quad (a_1, a_2, \ldots, a_n) \longmapsto a_j$$

Sie sind also ebenfalls linear. ♣

Wir zeigen jetzt, dass eine lineare Abbildung dadurch festgelegt ist, was sie auf einer Basis macht.

* 10.11. **Satz.** *Sei V ein K-Vektorraum mit Basis $(b_1, b_2, \ldots, b_n)$ und sei W ein weiterer K-Vektorraum. Seien weiter $w_1, w_2, \ldots, w_n \in W$. Dann gibt es genau eine K-lineare Abbildung $\phi\colon V \to W$ mit $\phi(b_j) = w_j$ für alle $j \in \{1, 2, \ldots, n\}$.* **SATZ** Basen und lin. Abb.

Beweis. Wir beweisen zuerst die Eindeutigkeit. Seien also $\phi_1, \phi_2\colon V \to W$ lineare Abbildungen mit $\phi_1(b_j) = w_j = \phi_2(b_j)$ für alle $j \in \{1, 2, \ldots, n\}$. Sei $v \in V$ beliebig. Dann ist v eine Linearkombination der Basisvektoren:

$$v = \lambda_1 b_1 + \lambda_2 b_2 + \ldots + \lambda_n b_n\,.$$

Es folgt

$$\begin{aligned}
\phi_1(v) &= \phi_1(\lambda_1 b_1 + \lambda_2 b_2 + \ldots + \lambda_n b_n) \\
&= \lambda_1 \phi_1(b_1) + \lambda_2 \phi_1(b_2) + \ldots + \lambda_n \phi_1(b_n) \\
&= \lambda_1 w_1 + \lambda_2 w_2 + \ldots + \lambda_n w_n \\
&= \lambda_1 \phi_2(b_1) + \lambda_2 \phi_2(b_2) + \ldots + \lambda_n \phi_2(b_n) \\
&= \phi_2(\lambda_1 b_1 + \lambda_2 b_2 + \ldots + \lambda_n b_n) \\
&= \phi_2(v)\,,
\end{aligned}$$

also ist $\phi_1 = \phi_2$.

Dieser Eindeutigkeitsbeweis zeigt uns, wie wir die Existenz beweisen können: Wenn es eine lineare Abbildung $\phi\colon V \to W$ gibt mit $\phi(b_j) = w_j$ für alle $j \in \{1, 2, \ldots, n\}$, dann muss für $v \in V$ wie oben gelten

$$\phi(v) = \lambda_1 w_1 + \lambda_2 w_2 + \ldots + \lambda_n w_n\,.$$

Wir müssen prüfen,

(1) dass ϕ wohldefiniert ist, dass also $\phi(v)$ nicht davon abhängt, wie v als Linearkombination der b_j geschrieben wurde, und

(2) dass die so definierte Abbildung ϕ linear ist.

Die Wohldefiniertheit folgt daraus, dass v nur auf genau eine Weise als Linearkombination der Basisvektoren geschrieben werden kann (vgl. Lemma 9.4). Die Linearität rechnet man nach. Etwas eleganter ist es, wenn man bemerkt, dass $\phi = \phi_{(w_1,w_2,\ldots,w_n)} \circ \phi^{-1}_{(b_1,b_2,\ldots,b_n)}$ ist (mit den zu $(b_1, b_2, \ldots, b_n)$ und zu $(w_1, w_2, \ldots, w_n)$ gehörigen Linearkombinationenabbildungen $K^n \to V$ bzw. $K^n \to W$ — man beachte, dass $\phi_{(b_1,b_2,\ldots,b_n)}$ hier ein Isomorphismus ist). Die Linearität von ϕ folgt dann daraus, dass die Linearkombinationenabbildungen linear sind (Beispiel 10.3) und aus Lemma 10.4. ❑

Das analoge Resultat gilt auch für (nicht unbedingt endliche) Basismengen:

Sind V und W K-Vektorräume und ist $B \subset V$ eine Basis, dann gibt es zu jeder Abbildung $f\colon B \to W$ genau eine lineare Abbildung $\phi\colon V \to W$ mit $\phi(b) = f(b)$ für alle $b \in B$ (oder kurz: $\phi|_B = f$).

Der Beweis geht im Wesentlichen genauso unter Verwendung der allgemeinen Linearkombinationenabbildungen $K^{(B)} \to V$ bzw. $K^{(B)} \to W$.

Für praktische Zwecke lässt sich der Inhalt von Satz 10.11 wie folgt zusammenfassen. Seien dazu V ein K-Vektorraum mit Basis $(b_1, b_2, \ldots, b_n)$ und W ein weiterer K-Vektorraum.

- Wir können eine lineare Abbildung $V \to W$ definieren, indem wir die Bilder der Basisvektoren b_j beliebig festlegen.
- Zwei lineare Abbildungen $V \to W$ sind schon dann gleich, wenn die Bilder der Basisvektoren b_j unter beiden Abbildungen übereinstimmen.

Da eine lineare Abbildung also durch das Bild einer Basis eindeutig bestimmt ist, sollten sich auch Eigenschaften wie injektiv oder surjektiv zu sein durch das Bild der Basis ausdrücken lassen.

SATZ inj./surj. lin. Abb.

10.12. **Satz.** *Seien V und W K-Vektorräume und sei $\phi\colon V \to W$ linear. Sei weiter $(b_1, b_2, \ldots, b_n)$ eine Basis von V.*

(1) *ϕ ist genau dann injektiv, wenn $\big(\phi(b_1), \phi(b_2), \ldots, \phi(b_n)\big)$ linear unabhängig ist.*

(2) *ϕ ist genau dann surjektiv, wenn $\big(\phi(b_1), \phi(b_2), \ldots, \phi(b_n)\big)$ ein Erzeugendensystem von W ist.*

(3) *ϕ ist genau dann ein Isomorphismus, wenn $(\phi(b_1), \phi(b_2), \ldots, \phi(b_n))$ eine Basis von W ist.*

Beweis. Seien $w_1 = \phi(b_1)$, $w_2 = \phi(b_2)$, …, $w_n = \phi(b_n)$. Wie im Beweis von Satz 10.11 ist dann $\phi = \phi_{(w_1,w_2,\ldots,w_n)} \circ \phi^{-1}_{(b_1,b_2,\ldots,b_n)}$. Da $\phi_{(b_1,b_2,\ldots,b_n)}$ bijektiv ist, ist

ϕ injektiv bzw. surjektiv genau dann, wenn $\phi_{(w_1,w_2,\dots,w_n)}$ die entsprechende Eigenschaft hat (beachte dafür $\phi \circ \phi_{(b_1,b_2,\dots,b_n)} = \phi_{(w_1,w_2,\dots,w_n)}$). Die Behauptungen folgen dann sofort aus den Aussagen von Lemma 9.4 (siehe auch Beispiel 10.3). ❑

Daraus können wir gleich zwei wichtige Folgerungen ziehen.

10.13. **Folgerung.** *Sind V und W zwei K-Vektorräume derselben endlichen Dimension n, dann sind V und W isomorph.*

FOLG endl.-dim. VR gleicher Dimension sind isomorph

Beweis. Sei $(b_1, b_2, \dots, b_n)$ eine Basis von V und sei $(b'_1, b'_2, \dots, b'_n)$ eine Basis von W. Dann gibt es nach Satz 10.11 eine lineare Abbildung $\phi\colon V \to W$ mit $\phi(b_j) = b'_j$ für alle $j \in \{1, 2, \dots, n\}$. Nach Satz 10.12 ist ϕ ein Isomorphismus. ❑

10.14. **Folgerung.** *Seien V und W zwei K-Vektorräume derselben endlichen Dimension n und sei $\phi\colon V \to W$ eine lineare Abbildung. Dann sind die folgenden Aussagen äquivalent:*

FOLG lin. Abb. bei gleicher Dimension

(1) *ϕ ist ein Isomorphismus.*

(2) *ϕ ist injektiv.*

(3) *ϕ ist surjektiv.*

Beweis. Es ist klar, dass aus (1) die beiden Aussagen (2) und (3) folgen. Sei $(b_1, \dots, b_n)$ eine Basis von V. Nach Satz 10.12 ist ϕ genau dann injektiv, wenn $\big(\phi(b_1), \dots, \phi(b_n)\big)$ linear unabhängig ist. n linear unabhängige Vektoren bilden aber eine Basis (wegen $\dim W = n$, siehe Satz 9.14); nach Satz 10.12 ist ϕ dann ein Isomorphismus. Analog ist ϕ genau dann surjektiv, wenn $\big(\phi(b_1), \dots, \phi(b_n)\big)$ den Vektorraum W erzeugt. Ein Erzeugendensystem aus n Elementen ist aber wieder eine Basis, also ist ϕ dann ein Isomorphismus. ❑

Als Merkhilfe für diese wichtige Aussage kann die Analogie zu Abbildungen zwischen endlichen Mengen dienen. Es gilt nämlich (wie leicht einzusehen ist):

Seien X und Y zwei endliche Mengen mit $\#X = \#Y = n$ und sei $f\colon X \to Y$ eine Abbildung. Dann sind die folgenden Aussagen äquivalent:

(1) *f ist bijektiv.*

(2) *f ist injektiv.*

(3) *f ist surjektiv.*

Die im folgenden Beispiel verwendete Produktschreibweise

$$\prod_{i \in I} a_i$$

ist analog definiert wie die Schreibweise mit dem Summenzeichen $\sum$, nur dass die angegebenen Elemente multipliziert werden statt addiert. Ist $I = \emptyset$, dann hat das („leere“) Produkt den Wert 1.

BSP
Interpolation

10.15. **Beispiel.** Der Vektorraum $P_{<n}$ der Polynomfunktionen vom Grad $< n$ wird von den n Potenzfunktionen $x \mapsto x^j$ für $j \in \{0, 1, \ldots, n-1\}$ erzeugt. Diese Funktionen sind linear unabhängig, also hat $P_{<n}$ Dimension n.

Seien $x_1, x_2, \ldots, x_n \in \mathbb{R}$ paarweise verschieden. Wir definieren für $j \in \{1, 2, \ldots, n\}$ die Polynomfunktion $p_j \in P_{<n}$ durch

$$p_j(x) = \prod_{i \in \{1,\ldots,n\}\setminus\{j\}} \frac{x - x_i}{x_j - x_i} = \frac{x - x_1}{x_j - x_1} \cdots \frac{x - x_{j-1}}{x_j - x_{j-1}} \cdot \frac{x - x_{j+1}}{x_j - x_{j+1}} \cdots \frac{x - x_n}{x_j - x_n}\,.$$

Für $n = 3$ und $(x_1, x_2, x_3) = (-1, 0, 1)$ bedeutet das zum Beispiel:

$$p_1(x) = \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} = \frac{x(x-1)}{(-1)(-2)} = \tfrac{1}{2}x^2 - \tfrac{1}{2}x$$
$$p_2(x) = \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} = \frac{(x+1)(x-1)}{1 \cdot (-1)} = -x^2 + 1$$
$$p_3(x) = \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)} = \frac{(x+1)x}{2 \cdot 1} = \tfrac{1}{2}x^2 + \tfrac{1}{2}x$$

Es gilt dann für $i, j \in \{1, 2, \ldots, n\}$:

$$p_j(x_i) = \begin{cases} 1 & \text{falls } i = j; \\ 0 & \text{falls } i \neq j. \end{cases}$$

Wir definieren

$$\phi\colon P_{<n} \longrightarrow \mathbb{R}^n, \quad f \longmapsto (f(x_1), f(x_2), \ldots, f(x_n))$$

(ϕ ist aus Auswertungsabbildungen zusammengesetzt und daher linear) und eine lineare Abbildung $\psi\colon \mathbb{R}^n \to P_{<n}$ durch Festlegung der Bilder der Standardbasis:

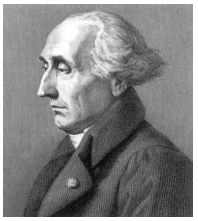
$$\psi(\mathbf{e}_j) = p_j \qquad \text{für alle } j \in \{1, 2, \ldots, n\}.$$

(ψ ist gerade die Linearkombinationenabbildung $\phi_{(p_1,\ldots,p_n)}$.) Dann gilt $\phi \circ \psi = \mathrm{id}_{\mathbb{R}^n}$:

$$\phi\big(\psi(\mathbf{e}_j)\big) = \phi(p_j) = \big(p_j(x_1), \ldots, p_j(x_n)\big) = \mathbf{e}_j\,.$$

$\phi \circ \psi$ und die identische Abbildung stimmen auf einer Basis überein, also sind sie gleich. Dann muss ψ injektiv sein und ϕ surjektiv. Nach Folgerung 10.14 sind wegen $\dim \mathbb{R}^n = n = \dim P_{<n}$ beide Abbildungen (zueinander inverse) Isomorphismen. Das bedeutet:

*Seien $x_1, x_2, \ldots, x_n \in \mathbb{R}$ paarweise verschieden und $y_1, y_2, \ldots, y_n \in \mathbb{R}$. Dann gibt es **genau eine** Polynomfunktion $f \in P_{<n}$ mit $f(x_j) = y_j$ für alle $j \in \{1, 2, \ldots, n\}$, nämlich $f = y_1 p_1 + y_2 p_2 + \ldots + y_n p_n$.*



J.-L. Lagrange
1736–1813

Das sieht man so: Die Bedingung an f bedeutet $f \in P_{<n}$ und $\phi(f) = (y_1, y_2, \ldots, y_n)$. Letzteres ist aber äquivalent zu

$$f = \psi(y_1, y_2, \ldots, y_n) = y_1 p_1 + y_2 p_2 + \ldots + y_n p_n\,.$$

Diese Formel für das Interpolationspolynom heißt Lagrangesche Interpolationsformel.

(Dass ϕ bijektiv sein muss, kann man alternativ auch so sehen: Aus Beispiel 9.18 wissen wir bereits, dass ϕ surjektiv ist. Da $\dim P_{<n} = n = \dim \mathbb{R}^n$ gilt, sagt Folgerung 10.14, dass ϕ schon bijektiv sein muss. Das sagt uns aber noch nicht, wie die Umkehrabbildung aussieht.) ♣

Wir bleiben bei den Polynomfunktionen und geben weitere Beispiele für lineare Abbildungen.

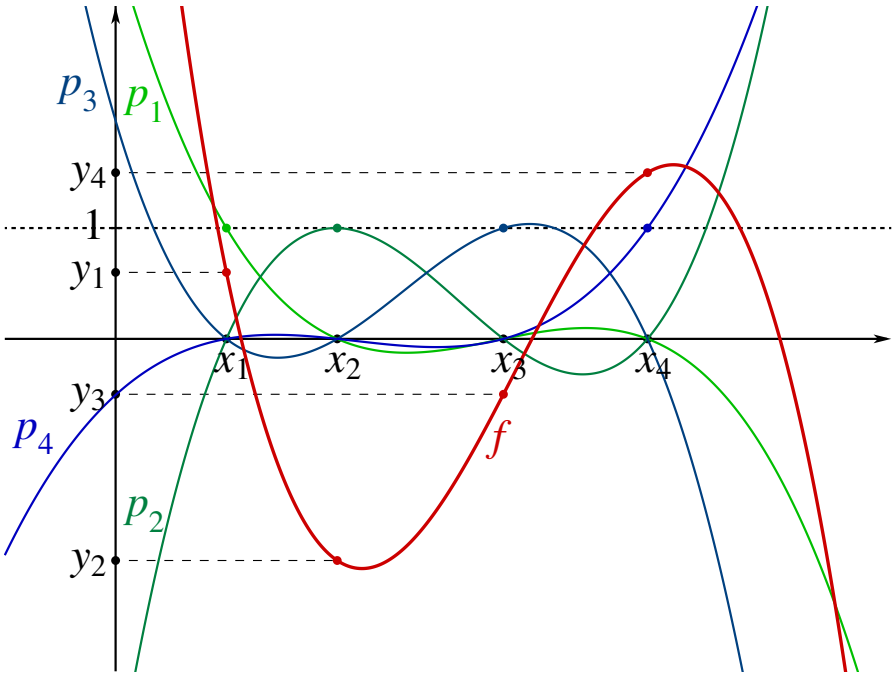


ABBILDUNG 3. Interpolationspolynom (zu Beispiel 10.15); hier ist der Fall mit vier Stützstellen $x_1, \ldots, x_4$ dargestellt.

Beispiele.

BSP
lin. Abb. auf Polynomfkt.

(1) Wir haben schon in Beispiel 10.10 gesehen, dass für $a \in \mathbb{R}$ die Auswertungsabbildung

$$\mathrm{ev}_a \colon P \longrightarrow \mathbb{R}, \quad f \longmapsto f(a)$$

linear ist.

(2) Die *Differentiation* von Polynomfunktionen ist linear:

$$D \colon P \longrightarrow P, \quad f \longmapsto f' .$$

Für $f(x) = a_0 + a_1 x + \ldots + a_n x^n$ gilt dabei $f'(x) = a_1 + 2a_2 x + \ldots + n a_n x^{n-1}$. Man könnte D also definieren als diejenige lineare Abbildung, die für $n > 0$ die Potenzfunktion $f_n \colon x \mapsto x^n$ auf $n f_{n-1}$ abbildet und f_0 auf die Nullfunktion.

D ist surjektiv (also ein Epimorphismus) und der Kern von D besteht genau aus den konstanten Funktionen. (An diesem Beispiel kann man sehen, dass die Aussage von Satz 10.12 für unendlich-dimensionale Vektorräume nicht gelten muss.)

(3) Die Berechnung des *bestimmten Integrals* von a bis b ist linear:

$$I_{a,b} \colon P \longrightarrow \mathbb{R}, \quad f \longmapsto \int_a^b f(x)\, dx .$$

Für die Potenzfunktionen f_n gilt $I_{a,b}(f_n) = \dfrac{b^{n+1} - a^{n+1}}{n+1}$.

(4) Auch die *unbestimmte Integration* mit Anfangspunkt $a \in \mathbb{R}$ ist linear:

$$I_a \colon P \longrightarrow P, \quad f \longmapsto \left(x \mapsto \int_a^x P(t)\, dt \right) .$$

Das ist die lineare Abbildung mit $I_a(f_n) = \left(x \mapsto \frac{1}{n+1}(x^{n+1} - a^{n+1})\right)$. Die Abbildung I_a ist injektiv, aber nicht surjektiv: ihr Bild ist gerade der Kern von ev_a (die Integralfunktionen verschwinden alle an der Stelle a).

(5) Die *Translation* (also Verschiebung) um $a \in \mathbb{R}$ ist linear:

$$T_a \colon P \longrightarrow P, \quad f \longmapsto (x \mapsto f(x-a)) .$$

T_a ist ein Automorphismus von P, der inverse Automorphismus ist T_{-a}.

Zwischen diesen Abbildungen bestehen eine Reihe von Relationen, wie zum Beispiel

$$\begin{gathered} \mathrm{ev}_b \circ I_a = I_{a,b}, \qquad D \circ I_a = \mathrm{id}_P, \qquad (I_a \circ D)(f) = f - \mathrm{ev}_a(f) f_0, \\ T_a \circ D = D \circ T_a, \qquad T_a \circ T_b = T_{a+b}, \qquad I_a \circ T_b = T_b \circ I_{a-b}, \\ I_{a,b} \circ T_c = I_{a-c,b-c}, \qquad \mathrm{ev}_a \circ T_b = \mathrm{ev}_{a-b} . \end{gathered}$$

Man kann sie leicht auf den Potenzfunktionen nachprüfen (das genügt, weil die Potenzfunktionen eine Basis von P sind).

In der Analysis lernen Sie, dass Differentiation und Integration ganz allgemein lineare Abbildungen sind. ♣

Kern und Bild einer linearen Abbildung sind wichtige Daten. Für die Dimension des Bildes gibt es sogar einen eigenen Namen.

* 10.16. **Definition.** Ist $\phi \colon V \to W$ eine lineare Abbildung, dann heißt **DEF** Rang einer linearen Abb.

$$\mathrm{rk}(\phi) = \dim \mathrm{im}(\phi)$$

der *Rang* von ϕ. ♢

Zwischen dem Rang und der Dimension des Kerns besteht ein einfacher Zusammenhang. (Siehe Abbildung 2 für eine Veranschaulichung.)

* 10.17. **Satz.** *Sei $\phi \colon V \to W$ eine lineare Abbildung. Dann gilt* **SATZ** Rangsatz

$$\dim \ker(\phi) + \mathrm{rk}(\phi) = \dim V .$$

Dabei sei $n + \infty = \infty + n = \infty + \infty = \infty$ für $n \in \mathbb{N}$.

Beweis. Ist $\dim \ker(\phi) = \infty$, dann muss auch $\dim V = \infty$ sein, denn $\ker(\phi)$ ist ein Untervektorraum von V (siehe Satz 9.19). Also ist die Behauptung in diesem Fall richtig. Ist $\mathrm{rk}(\phi) = \infty$, dann können wir unendlich viele linear unabhängige Vektoren $w_j \in \mathrm{im}(\phi)$ finden ($j \in \mathbb{N}$). Sei $v_j \in V$ ein Urbild von w_j; dann sind auch die v_j linear unabhängig. Denn sei $\lambda_0 v_0 + \lambda_1 v_1 + \ldots + \lambda_n v_n = \mathbf{0}$, dann folgt durch Anwenden von ϕ auch

$$\begin{aligned} \lambda_0 w_0 + \lambda_1 w_1 + \ldots + \lambda_n w_n &= \lambda_0 \phi(v_0) + \lambda_1 \phi(v_1) + \ldots + \lambda_n \phi(v_n) \\ &= \phi(\lambda_0 v_0 + \lambda_1 v_1 + \ldots + \lambda_n v_n) = \mathbf{0} . \end{aligned}$$

Weil $w_0, w_1, \ldots, w_n$ linear unabhängig sind, müssen alle Koeffizienten λ_j null sein, was zeigt, dass $v_0, v_1, \ldots, v_n$ linear unabhängig sind. Es gibt also unendlich viele linear unabhängige Vektoren in V; damit ist $\dim V = \infty$ und die Behauptung des Satzes stimmt. Wir können also jetzt annehmen, dass $\dim \ker(\phi)$ und $\mathrm{rk}(\phi)$ beide endlich sind.

Seien $k = \dim \ker(\phi)$, $r = \mathrm{rk}(\phi)$. Wir können eine Basis $(b_1, \ldots, b_k)$ von $\ker(\phi)$ und eine Basis $(b'_1, \ldots, b'_r)$ vom $\mathrm{im}(\phi)$ wählen. Es gibt dann Vektoren $b_{k+1}, \ldots, b_{k+r}$ mit $\phi(b_{k+1}) = b'_1$, $\phi(b_{k+2}) = b'_2$, …, $\phi(b_{k+r}) = b'_r$. Ich behaupte jetzt, dass $(b_1, \ldots, b_k, b_{k+1}, \ldots, b_{k+r})$ eine Basis von V ist. Daraus folgt $k + r = \dim V$, also die Behauptung des Satzes.

- Erzeugendensystem:
  Sei $v \in V$. Da $(b'_1, \ldots, b'_r)$ eine Basis von $\mathrm{im}(\phi)$ ist, gibt es Skalare $\mu_1, \ldots, \mu_r$ mit $\phi(v) = \mu_1 b'_1 + \ldots + \mu_r b'_r$. Dann ist
  $$\begin{aligned}\phi\big(v - (\mu_1 b_{k+1} + \ldots + \mu_r b_{k+r})\big) &= \phi(v) - \big(\mu_1 \phi(b_{k+1}) + \ldots + \mu_r \phi(b_{k+r})\big) \\ &= \phi(v) - (\mu_1 b'_1 + \ldots + \mu_r b'_r) = \mathbf{0}\,,\end{aligned}$$
  also ist $v - (\mu_1 b_{k+1} + \ldots + \mu_r b_{k+r}) \in \ker(\phi)$. Es gibt also Skalare $\lambda_1, \ldots, \lambda_k$ mit
  $$v - (\mu_1 b_{k+1} + \ldots + \mu_r b_{k+r}) = \lambda_1 b_1 + \ldots + \lambda_k b_k\,.$$
  Damit ist
  $$v = \lambda_1 b_1 + \ldots + \lambda_k b_k + \mu_1 b_{k+1} + \ldots + \mu_r b_{k+r}$$
  eine Linearkombination von $(b_1, \ldots, b_{k+r})$.
- linear unabhängig:
  Seien $\lambda_1, \ldots, \lambda_{k+r}$ Skalare mit
  $$\lambda_1 b_1 + \ldots + \lambda_k b_k + \lambda_{k+1} b_{k+1} + \ldots + \lambda_{k+r} b_{k+r} = \mathbf{0}\,.$$
  Dann ist
  $$\begin{aligned}\mathbf{0} &= \phi(\lambda_1 b_1 + \ldots + \lambda_k b_k + \lambda_{k+1} b_{k+1} + \ldots + \lambda_{k+r} b_{k+r}) \\ &= \lambda_1 \phi(b_1) + \ldots + \lambda_k \phi(b_k) + \lambda_{k+1} \phi(b_{k+1}) + \ldots + \lambda_{k+r} \phi(b_{k+r}) \\ &= \lambda_{k+1} \phi(b_{k+1}) + \ldots + \lambda_{k+r} \phi(b_{k+r}) \\ &= \lambda_{k+1} b'_1 + \ldots + \lambda_{k+r} b'_r\,,\end{aligned}$$
  denn $\phi(b_1) = \ldots = \phi(b_k) = \mathbf{0}$. Da $(b'_1, \ldots, b'_r)$ linear unabhängig ist, muss $\lambda_{k+1} = \ldots = \lambda_{k+r} = 0$ gelten. Eingesetzt in die ursprüngliche Relation liefert das
  $$\lambda_1 b_1 + \ldots + \lambda_k b_k = \mathbf{0}\,.$$
  Weil auch $(b_1, \ldots, b_k)$ linear unabhängig ist, folgt daraus $\lambda_1 = \ldots = \lambda_k = 0$, also war auch unsere ursprüngliche Linearkombination trivial. ❑

10.18. **Beispiel.** Seien $V = K^n$ mit $n \geq 1$ und

BSP Dimension eines Kerns

$$\phi\colon V \longrightarrow K, \quad (x_1, x_2, \ldots, x_n) \longmapsto x_1 + x_2 + \ldots + x_n\,.$$

Dann ist ϕ linear (leicht nachzurechnen). Damit ist

$$U = \{(x_1, x_2, \ldots, x_n) \in V \mid x_1 + x_2 + \ldots + x_n = 0\} = \ker(\phi) \subset V$$

ein Untervektorraum von V. Es gilt $\mathrm{im}(\phi) = K$:

$$\text{für } \lambda \in K \text{ ist } \phi\big((\lambda, 0, \ldots, 0)\big) = \lambda.$$

Damit ist $\mathrm{rk}(\phi) = \dim_K K = 1$. Es folgt

$$\dim U = \dim \ker(\phi) = \dim V - \mathrm{rk}(\phi) = n - 1\,.$$ ♣

In vielen Fällen ist es einfacher, den Kern und seine Dimension direkt zu bestimmen als den Rang. Mit Satz 10.17 kann man daraus dann den Rang berechnen.

Die Konstruktion des Vektorraums $K^X = \mathrm{Abb}(X, K)$ lässt sich verallgemeinern.

10.19. **Definition.** Seien K ein Körper und V ein K-Vektorraum. Sei weiter X eine Menge. Dann können wir auf $V^X = \mathrm{Abb}(X, V)$ eine Struktur als K-Vektorraum definieren durch

DEF Vektorraum V^X

$$f + g\colon x \mapsto f(x) + g(x) \qquad \text{und} \qquad \lambda f\colon x \mapsto \lambda f(x)\,.$$

Der Beweis ist analog zu dem für K^X.

Für $X = \{1, 2, \ldots, n\}$ identifizieren wir V^X mit V^n. ♢

Wir können also insbesondere zwei *lineare* Abbildungen $V \to W$ addieren oder eine solche Abbildung mit einem Skalar multiplizieren (da wir das sogar für beliebige Abbildungen $V \to W$ können).

10.20. **Satz.** *Seien V und W zwei K-Vektorräume. Die Menge der linearen Abbildungen $V \to W$ bildet einen K-Untervektorraum von* $\mathrm{Abb}(V, W)$.

SATZ Vektorraum der lin. Abb.

Beweis. Wir müssen die Bedingungen für einen Untervektorraum nachprüfen.

- Die Nullabbildung ist linear.
- Seien $\phi, \psi\colon V \to W$ linear. Wir müssen zeigen, dass $\phi + \psi$ ebenfalls linear ist. Seien dazu $v, v' \in V$, $\lambda \in K$. Dann haben wir

$$\begin{aligned}(\phi + \psi)(v + v') &= \phi(v + v') + \psi(v + v') = \phi(v) + \phi(v') + \psi(v) + \psi(v') \\ &= \phi(v) + \psi(v) + \phi(v') + \psi(v') = (\phi + \psi)(v) + (\phi + \psi)(v')\end{aligned}$$

  und

$$\begin{aligned}(\phi + \psi)(\lambda v) &= \phi(\lambda v) + \psi(\lambda v) = \lambda\phi(v) + \lambda\psi(v) \\ &= \lambda(\phi(v) + \psi(v)) = \lambda(\phi + \psi)(v)\,.\end{aligned}$$

- Sei $\phi\colon V \to W$ linear und $\lambda \in K$, Wir müssen zeigen, dass $\lambda\phi$ ebenfalls linear ist. Seien dazu $v, v' \in V$, $\mu \in K$. Dann haben wir

$$\begin{aligned}(\lambda\phi)(v + v') &= \lambda\phi(v + v') = \lambda(\phi(v) + \phi(v')) \\ &= \lambda\phi(v) + \lambda\phi(v') = (\lambda\phi)(v) + (\lambda\phi)(v')\end{aligned}$$

  und

$$\begin{aligned}(\lambda\phi)(\mu v) &= \lambda\phi(\mu v) = \lambda \cdot \mu\phi(v) \\ &= \mu \cdot \lambda\phi(v) = \mu(\lambda\phi)(v)\,.\end{aligned}$$

❑

10.21. **Definition.** Der Vektorraum der linearen Abbildungen $V \to W$ wird mit $\mathrm{Hom}(V, W)$ (oder $\mathrm{Hom}_K(V, W)$) bezeichnet. Im Fall $V = W$ schreiben wir auch $\mathrm{End}(V) = \mathrm{Hom}(V, V)$ (oder $\mathrm{End}_K(V)$) für den Vektorraum der Endomorphismen von V. ♢

DEF $\mathrm{Hom}(V, W)$ $\mathrm{End}(V)$

10.22. **Satz.** *Seien V und W zwei K-Vektorräume mit $\dim V = n < \infty$. Sei weiter $(b_1, b_2, \ldots, b_n)$ eine Basis von V. Dann ist*

SATZ $\mathrm{Hom}(V, W) \cong W^{\dim V}$

$$\Phi\colon \mathrm{Hom}(V, W) \longrightarrow W^n, \quad \phi \longmapsto \big(\phi(b_1), \phi(b_2), \ldots, \phi(b_n)\big)$$

ein Isomorphismus. Insbesondere ist im Fall von $\dim W = m < \infty$

$$\dim \mathrm{Hom}(V, W) = \dim W^n = n \dim W = mn = (\dim V)(\dim W)\,.$$

Beweis. Es ist klar, dass Φ linear ist (denn Φ setzt sich aus Auswertungsabbildungen zusammen; die Auswertungsabbildungen $\mathrm{Abb}(V,W) \to W$, $\phi \mapsto \phi(v)$, sind auch in diesem allgemeineren Kontext linear; man sieht das wie in Beispiel 10.10). Nach Satz 10.11 gibt es zu jeder Wahl der Bilder von $b_1, \dots, b_n$ in W genau eine lineare Abbildung; das bedeutet, dass Φ bijektiv ist. Isomorphe Vektorräume haben dieselbe Dimension; der Beweis von $\dim W^n = n \dim W$ ist eine Übungsaufgabe. ❑

10.23. **Folgerung.** *Ist $(b_1, b_2, \dots, b_n)$ eine Basis von V und ist $(b'_1, b'_2, \dots, b'_m)$ eine Basis von W, dann ist $(\phi_{ij})_{(i,j)\in\{1,2,\dots,m\}\times\{1,2,\dots,n\}}$ eine Basis von $\mathrm{Hom}(V,W)$, wobei $\phi_{ij}\colon V \to W$ die lineare Abbildung ist mit $\phi_{ij}(b_k) = \mathbf{0}$ für $k \neq j$ und $\phi_{ij}(b_j) = b'_i$.* **FOLG** Basis von $\mathrm{Hom}(V,W)$

Beweis. Nach Satz 10.11 existieren eindeutig bestimmte ϕ_{ij} wie angegeben. Wir zeigen, dass die $\phi_{ij} \in \mathrm{Hom}(V,W)$ linear unabhängig sind. Seien dazu λ_{ij} Skalare mit

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \lambda_{ij}\phi_{ij} = \mathbf{0}\,.$$

Sei $k \in \{1, 2, \dots, n\}$. Einsetzen von b_k liefert dann

$$\mathbf{0} = \Big(\sum_{i=1}^{m}\sum_{j=1}^{n} \lambda_{ij}\phi_{ij}\Big)(b_k) = \sum_{i=1}^{m}\sum_{j=1}^{n} \lambda_{ij}\phi_{ij}(b_k) = \sum_{i=1}^{m} \lambda_{ik} b'_i$$

Im letzten Schritt haben wir verwendet, dass $\phi_{ij}(b_k) = \mathbf{0}$ ist für $k \neq j$; in der inneren Summe bleibt dann nur $\lambda_{ik}\phi_{ik}(b_k) = \lambda_{ik}b'_i$ stehen. Da die b'_i linear unabhängig sind, folgt $\lambda_{ik} = 0$ für alle i. Da k beliebig war, sind also alle $\lambda_{ij} = 0$, was zu zeigen war. Nach Satz 10.22 ist $\dim \mathrm{Hom}(V,W) = nm$ gleich der Anzahl der linear unabhängigen Elemente $\phi_{ij} \in \mathrm{Hom}(V,W)$, nach Satz 9.14 sind die ϕ_{ij} dann bereits eine Basis von $\mathrm{Hom}(V,W)$. ❑

Im Fall $V = K^n$, $W = K^m$ mit den Standardbasen kann man das, was ϕ_{ij} bewirkt, so beschreiben: Man nimmt die j-te Komponente von $(x_1, x_2, \dots, x_n) \in K^n$ und steckt sie in die i-te Komponente des Resultats in K^m; die übrigen Komponenten sind null.

10.24. **Beispiel.** Als einfaches Beispiel betrachten wir $V = \mathbb{R}^3$, $W = \mathbb{R}^2$, jeweils mit der Standardbasis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ bzw. $(\mathbf{e}'_1, \mathbf{e}'_2)$. Die Basis von $\mathrm{Hom}(\mathbb{R}^3, \mathbb{R}^2)$ aus Folgerung 10.23 sieht in diesem Fall so aus: **BSP** Basis von $\mathrm{Hom}(\mathbb{R}^3, \mathbb{R}^2)$

$$\begin{aligned}
\phi_{11}&\colon (x,y,z) \longmapsto (x,0)\\
\phi_{12}&\colon (x,y,z) \longmapsto (y,0)\\
\phi_{13}&\colon (x,y,z) \longmapsto (z,0)\\
\phi_{21}&\colon (x,y,z) \longmapsto (0,x)\\
\phi_{22}&\colon (x,y,z) \longmapsto (0,y)\\
\phi_{23}&\colon (x,y,z) \longmapsto (0,z)
\end{aligned}$$

Jede lineare Abbildung $\phi\colon \mathbb{R}^3 \to \mathbb{R}^2$ lässt sich als Linearkombination dieser sechs Abbildungen schreiben; es gibt also $a, b, c, d, e, f \in \mathbb{R}$, sodass

$$\phi = a\phi_{11} + b\phi_{12} + c\phi_{13} + d\phi_{21} + e\phi_{22} + f\phi_{23},$$

also

$$\phi(x,y,z) = (ax+by+cz, dx+ey+fz)\,. \qquad \clubsuit$$

Die Endomorphismen eines Vektorraums V bilden sogar einen Ring, den *Endomorphismenring* von V:

DEF Endomorphismenring

10.25. **Satz.** *Sei V ein K-Vektorraum. Dann ist* $\mathrm{End}(V)$ *ein Ring mit der Addition des K-Vektorraums* $\mathrm{End}(V) = \mathrm{Hom}(V,V)$ *und der Komposition von Abbildungen als Multiplikation; das Einselement ist die identische Abbildung* id_V.

SATZ $\mathrm{End}(V)$ ist ein Ring

Beweis. Die Vektorraum-Axiome, die in $\mathrm{End}(V)$ gelten, liefern uns die Ring-Axiome für die Addition. Es bleibt noch zu zeigen, dass die Multiplikation assoziativ ist mit Einselement id_V und dass die beiden Ring-Distributivgesetze gelten. Seien also $f,g,h \in \mathrm{End}(V)$. Die Assoziativität $(f\circ g)\circ h = f\circ(g\circ h)$ gilt für Abbildungen ganz allgemein, ebenso wie $\mathrm{id}_V \circ f = f = f\circ \mathrm{id}_V$. Zum Nachweis der Distributivgesetze rechnen wir für $v\in V$:

$$\begin{aligned}\big((f+g)\circ h\big)(v) &= (f+g)\big(h(v)\big) = f\big(h(v)\big) + g\big(h(v)\big)\\ &= (f\circ h)(v) + (g\circ h)(v) = (f\circ h + g\circ h)(v)\,,\end{aligned}$$

also ist $(f+g)\circ h = f\circ h + g\circ h$, und

$$\begin{aligned}(f\circ(g+h))(v) &= f\big((g+h)(v)\big) = f\big(g(v)+h(v)\big)\\ &= f\big(g(v)\big) + f\big(h(v)\big) = (f\circ g)(v) + (f\circ h)(v)\\ &= (f\circ g + f\circ h)(v)\,,\end{aligned}$$

also ist $f\circ(g+h) = f\circ g + f\circ h$ (dabei haben wir verwendet, dass f linear ist). ❑

Der Endomorphismenring ist nicht kommutativ, wenn $\dim V \geq 2$ ist (Übung!). Für $\dim V = 1$ ist $\mathrm{End}(V) = K$, da alle Endomorphismen durch Multiplikation mit Skalaren gegeben sind; für $\dim V = 0$ ist $\mathrm{End}(V)$ der Nullring.

Die Automorphismen von V bilden eine Gruppe, die *Automorphismengruppe* $\mathrm{Aut}(V)$ von V (das ist auch die Gruppe der invertierbaren Elemente des Rings $\mathrm{End}(V)$).

11. Matrizen

Die Ergebnisse des letzten Abschnitts zeigen uns, dass wir lineare Abbildungen zwischen zwei endlich-dimensionalen K-Vektorräumen V und W der Dimensionen n und m durch mn Koeffizienten aus K beschreiben können. Dazu müssen wir Basen von V und W wählen; daraus bekommen wir eine Basis von $\mathrm{Hom}(V,W)$ wie in Folgerung 10.23 und die gesuchten Koeffizienten sind dann die Koeffizienten in der Darstellung der gegebenen linearen Abbildung als Linearkombination bezüglich dieser Basis. Für diese Koeffizienten führt man eine spezielle Form der Darstellung ein.

DEF Matrix

11.1. **Definition.** Sei K ein Körper und seien $m, n \in \mathbb{N}$. Eine $m \times n$-*Matrix mit Einträgen aus* K (oder kurz *über* K) ist ein rechteckiges Schema aus mn Elementen von K, das wie folgt notiert wird:

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

Zur Abkürzung schreiben wir auch $(a_{ij})_{1\le i\le m, 1\le j\le n}$ (oder auch $(a_{ij})_{i,j}$, falls die Zahlen m und n aus dem Kontext klar sind) für diese Matrix. Im Fall $m = n$ heißt die Matrix *quadratisch*. Für $i \in \{1, 2, \ldots, m\}$ heißt das n-Tupel $(a_{i1}, a_{i2}, \ldots, a_{in})$ die i*-te Zeile der Matrix*, für $j \in \{1, 2, \ldots, n\}$ heißt das m-Tupel $(a_{1j}, a_{2j}, \ldots, a_{mj})$ die j*-te Spalte* der Matrix.

Wir schreiben $\mathrm{Mat}(m \times n, K)$ für die Menge aller $m \times n$-Matrizen mit Einträgen aus K; im Fall $m = n$ auch kürzer $\mathrm{Mat}(n, K)$ für $\mathrm{Mat}(n \times n, K)$. ♢

Ebenso gebräuchlich ist die Notation $K^{m\times n}$. Matrizen werden auch (insbesondere in der englischsprachigen Literatur) mit eckigen Klammern notiert:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Im Grunde ist eine $m \times n$-Matrix über K nichts anderes als eine Familie von Elementen von K mit der Indexmenge $\{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$, also

$$\mathrm{Mat}(m \times n, K) = K^{\{1,2,\ldots,m\}\times\{1,2,\ldots,n\}}.$$

Da wir auf beliebigen Mengen der Form K^I eine Struktur als K-Vektorraum definiert haben, folgt sofort:

LEMMA Vektorraum der $m \times n$-Matrizen

11.2. **Lemma.** *Seien K ein Körper und $m, n \in \mathbb{N}$. Die Menge $\mathrm{Mat}(m \times n, K)$ mit komponentenweise definierter Addition und Skalarmultiplikation ist ein K-Vektorraum der Dimension mn.*

Ist $m = 0$ oder $n = 0$ (oder beides), dann ist $\mathrm{Mat}(m \times n, K)$ ein Null-Vektorraum; sein einziges Element ist eine leere Matrix (mit null Zeilen und n Spalten oder mit m Zeilen und null Spalten).

Matrizen (mit der gleichen Anzahl an Zeilen und Spalten) werden also wie folgt addiert und mit Skalaren multipliziert:

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

und

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

Wie zu Beginn dieses Abschnitts beschrieben, können wir linearen Abbildungen Matrizen zuordnen. Wir betrachten zunächst $V = K^n$ und $W = K^m$ mit den Standardbasen $B = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ von V und $B' = (\mathbf{e}'_1, \ldots, \mathbf{e}'_m)$ von W (wir schreiben hier $\mathbf{e}'_i$ für den i-ten Standard-Basisvektor in K^m zur Unterscheidung von den Basisvektoren $\mathbf{e}_j$ in K^n). Wir haben dann die Basis $(\phi_{ij})_{1\le i\le m, 1\le j\le n}$ von $\operatorname{Hom}(K^n, K^m)$ wie in Folgerung 10.23 mit $\phi_{ij}(\mathbf{e}_k) = \mathbf{0}$ für $k \neq j$ und $\phi_{ij}(\mathbf{e}_j) = \mathbf{e}'_i$. Ist $\phi\colon K^n \to K^m$ eine lineare Abbildung, dann schreiben wir ϕ als Linearkombination

$$\phi = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}\phi_{ij} \qquad \text{mit } a_{ij} \in K.$$

Die zugehörige Matrix ist dann $A = (a_{ij})_{1\le i\le m, 1\le j\le n}$.

11.3. **Beispiel.** Wie wir gesehen haben, hat eine lineare Abbildung $\phi\colon \mathbb{R}^3 \to \mathbb{R}^2$ die Form $\phi(x, y, z) = (ax + by + cz, dx + ey + fz)$ mit geeigneten $a, b, c, d, e, f \in \mathbb{R}$. Dann ist $\phi = a\phi_{11} + b\phi_{12} + c\phi_{13} + d\phi_{21} + e\phi_{22} + f\phi_{23}$ (vergleiche Beispiel 10.24), also ist die zugehörige Matrix

BSP Matrix für $\phi\colon \mathbb{R}^3 \to \mathbb{R}^2$

$$A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}. \qquad \clubsuit$$

Die j-te Spalte der zu $\phi\colon K^n \to K^m$ gehörigen $m \times n$-Matrix enthält gerade die Koeffizienten des Bildes des j-ten Standard-Basisvektors $\mathbf{e}_j$, denn

$$\phi(\mathbf{e}_j) = \sum_{i=1}^{m} \sum_{k=1}^{n} a_{ik}\phi_{ik}(\mathbf{e}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{e}'_i = (a_{1j}, a_{2j}, \ldots, a_{mj})$$

ähnlich wie im Beweis von Folgerung 10.23.

11.4. **Lemma.** *Die oben beschriebene Zuordnung definiert einen Isomorphismus* $\operatorname{Hom}(K^n, K^m) \to \operatorname{Mat}(m \times n, K)$. *Wenn man* $\operatorname{Mat}(m \times n, K)$ *mit* $K^{\{1,\ldots,m\}\times\{1,\ldots,n\}}$ *identifiziert, dann ist dieser Isomorphismus invers zu der Linearkombinationenabbildung* $K^{\{1,2,\ldots,m\}\times\{1,2,\ldots,n\}} \to \operatorname{Hom}(K^n, K^m)$, *die zur Basis* $(\phi_{ij})_{(i,j)\in\{1,\ldots,m\}\times\{1,\ldots,n\}}$ *von* $\operatorname{Hom}(K^n, K^m)$ *gehört.*

LEMMA $\operatorname{Mat}(m \times n, K) \cong \operatorname{Hom}(K^n, K^m)$

Beweis. Die erwähnte Linearkombinationenabbildung

$$\Phi\colon\ \mathrm{Mat}(m\times n,K)=K^{\{1,2,\dots,m\}\times\{1,2,\dots,n\}}\to\mathrm{Hom}(K^n,K^m)$$

bildet eine Matrix $(a_{ij})_{1\le i\le m,1\le j\le n}$ auf die Linearkombination $\sum_{i,j}a_{ij}\phi_{ij}$ ab; sie ist ein Isomorphismus, da $(\phi_{ij})_{1\le i\le m,1\le j\le n}$ eine Basis von $\mathrm{Hom}(K^n,K^m)$ ist (siehe Beispiel 10.3). Die Abbildung, die einer linearen Abbildung $\phi\colon K^n\to K^m$ ihre Matrix zuordnet, ist offenbar die Umkehrabbildung von Φ, insbesondere also ebenfalls ein Isomorphismus. ❑

Wie stellt sich die Anwendung der linearen Abbildung $\phi\colon K^n\to K^m$ dar, wenn wir die zugehörige Matrix $A=(a_{ij})_{i,j}$ verwenden? Es gilt

$$\phi(x_1,x_2,\dots,x_n)=\sum_{i=1}^m\sum_{j=1}^n a_{ij}\phi_{ij}(x_1,x_2,\dots,x_n)=\sum_{i=1}^m\sum_{j=1}^n a_{ij}x_j\mathbf{e}'_i\,,$$

also ist die i-te Komponente von $\phi(x_1,x_2,\dots,x_n)$ gegeben durch

$$\sum_{j=1}^n a_{ij}x_j=a_{i1}x_1+a_{i2}x_2+\ldots+a_{in}x_n\,.$$

Man schreibt das dann gerne als Multiplikation der Matrix A mit dem $(x_1,\dots,x_n)$ entsprechenden *Spaltenvektor*: Man identifiziert also K^n mit $\mathrm{Mat}(n\times 1,K)$ und K^m mit $\mathrm{Mat}(m\times 1,K)$. Dann haben wir für das Resultat der Anwendung von ϕ:

$$\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n\\a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n\\\vdots\\a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n\end{pmatrix}.$$

Das Ergebnis ist wieder ein Spaltenvektor, diesmal der Länge m. Seine i-te Komponente ergibt sich aus der i-ten Zeile der Matrix und dem Spaltenvektor zu $(x_1,\dots,x_n)$ als das *Skalarprodukt*

$$a_{i1}x_1+a_{i2}x_2+\ldots+a_{in}x_n\,.$$

(Das Skalarprodukt heißt so, weil sein Wert ein Skalar ist:

„Vektor mal Vektor = Skalar“.

Man beachte den Unterschied zur Skalarmultiplikation

„Skalar mal Vektor = Vektor“!)

11.5. **Beispiele.** 2×3-Matrizen mit Einträgen in $\mathbb{R}$ entsprechen linearen Abbildungen $\mathbb{R}^3\to\mathbb{R}^2$. In diesem Fall sieht obige Formel so aus: **BSP**

$$\begin{pmatrix}a&b&c\\d&e&f\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}ax+by+cz\\dx+ey+fz\end{pmatrix}$$

3×2-Matrizen über $\mathbb{R}$ entsprechen linearen Abbildungen $\mathbb{R}^2\to\mathbb{R}^3$. Dann haben wir:

$$\begin{pmatrix}a&b\\c&d\\e&f\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}ax+by\\cx+dy\\ex+fy\end{pmatrix}$$ ♣

Die Verknüpfung von linearen Abbildungen entspricht der Multiplikation von Matrizen.

11.6. **Definition.** Sei K ein Körper, seien weiter $l, m, n \in \mathbb{N}$. Für Matrizen $A \in \mathrm{Mat}(l \times m, K)$, $B \in \mathrm{Mat}(m \times n, K)$ ist das *Produkt* $A \cdot B \in \mathrm{Mat}(l \times n, K)$ definiert als die zu $f \circ g$ gehörende Matrix, wobei $f\colon K^m \to K^l$ und $g\colon K^n \to K^m$ die den Matrizen A und B entsprechenden linearen Abbildungen sind. Wie üblich schreibt man auch AB für $A \cdot B$. ♢

DEF Matrix-multiplikation

So wie man Abbildungen nur dann miteinander verknüpfen kann, wenn der Wertebereich der einen Abbildung mit dem Definitionsbereich der anderen übereinstimmt, kann man Matrizen nur dann miteinander multiplizieren, wenn sie in der Größe „zueinander passen“, wenn also die Spaltenanzahl des linken Faktors gleich der Zeilenanzahl des rechten Faktors ist.

Wie sieht diese Matrixmultiplikation konkret aus? Seien $A = (a_{ij})_{1\le i\le l, 1\le j\le m}$, $B = (b_{jk})_{1\le j\le m, 1\le k\le n}$ und $C = (c_{ik})_{1\le i\le l, 1\le k\le n} = AB$. Dann sollte c_{ik} die i-te Komponente von $f(g(\mathbf{e}_k))$ sein. Es ist

$$f(g(\mathbf{e}_k)) = f(b_{1k}\mathbf{e}'_1 + b_{2k}\mathbf{e}'_2 + \ldots + b_{mk}\mathbf{e}'_m) = b_{1k}f(\mathbf{e}'_1) + b_{2k}f(\mathbf{e}'_2) + \ldots + b_{mk}f(\mathbf{e}'_m)$$

und die i-te Komponente von $f(\mathbf{e}'_j)$ ist a_{ij}. Also ist

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{im}b_{mk} = \sum_{j=1}^{m} a_{ij}b_{jk}$$

das **Skalarprodukt der i-ten Zeile von A mit der k-ten Spalte von B**:

Die (i,k)-Komponente von AB ist „i-te Zeile von A mal k-te Spalte von B“.

Die oben eingeführte Multiplikation „Matrix mal Spaltenvektor“ ist dann also ein Spezialfall dieser allgemeinen Matrixmultiplikation.

11.7. **Beispiel.** Wir berechnen das Produkt zweier Matrizen über $\mathbb{R}$: **BSP**

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 1\cdot 1 + 2\cdot 3 + 3\cdot 5 & 1\cdot 2 + 2\cdot 4 + 3\cdot 6 \\ 4\cdot 1 + 5\cdot 3 + 6\cdot 5 & 4\cdot 2 + 5\cdot 4 + 6\cdot 6 \end{pmatrix} = \begin{pmatrix} 22 & 28 \\ 49 & 64 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1\cdot 1 + 2\cdot 4 & 1\cdot 2 + 2\cdot 5 & 1\cdot 3 + 2\cdot 6 \\ 3\cdot 1 + 4\cdot 4 & 3\cdot 2 + 4\cdot 5 & 3\cdot 3 + 4\cdot 6 \\ 5\cdot 1 + 6\cdot 4 & 5\cdot 2 + 6\cdot 5 & 5\cdot 3 + 6\cdot 6 \end{pmatrix} = \begin{pmatrix} 9 & 12 & 15 \\ 19 & 26 & 33 \\ 29 & 40 & 51 \end{pmatrix}$$

♣

Zur identischen Abbildung gehört eine spezielle Matrix.

11.8. **Definition.** Seien K ein Körper und $n \in \mathbb{N}$. Die Matrix $I_n \in \mathrm{Mat}(n, K)$, die der identischen Abbildung id_{K^n} entspricht, heißt die *Einheitsmatrix (der Größe n über K)*. ♢

DEF Einheits-matrix

In der j-ten Spalte von I_n muss der j-te Standard-Basisvektor stehen, also sieht I_n so aus:

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Man schreibt das auch $I_n = (\delta_{ij})_{1\le i,j\le n}$ mit dem *Kronecker-Delta*

DEF Kronecker-Delta

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

11.9. **Lemma.** *Sei K ein Körper. Die Matrixmultiplikation ist assoziativ und hat die Einheitsmatrix als neutrales Element; sie erfüllt die Distributivgesetze bezüglich der Matrixaddition:* **LEMMA** Eigensch. Matrixmult.

(1) *Für alle $A \in \mathrm{Mat}(k \times l, K)$, $B \in \mathrm{Mat}(l \times m, K)$, $C \in \mathrm{Mat}(m \times n, K)$ gilt $(AB)C = A(BC)$.*

(2) *Für alle $A \in \mathrm{Mat}(m \times n, K)$ gilt $I_m A = A = A I_n$.*

(3) *Für alle $A \in \mathrm{Mat}(l \times m, K)$ und $B, C \in \mathrm{Mat}(m \times n, K)$ gilt $A(B + C) = AB + AC$.*

(4) *Für alle $A, B \in \mathrm{Mat}(l \times m, K)$ und $C \in \mathrm{Mat}(m \times n, K)$ gilt $(A + B)C = AC + BC$.*

(5) *Außerdem gilt für alle $\lambda \in K$ und $A \in \mathrm{Mat}(l \times m, K)$, $B \in \mathrm{Mat}(m \times n, K)$: $\lambda(AB) = (\lambda A)B = A(\lambda B)$.*

Insbesondere ist $\mathrm{Mat}(n, K)$ mit der Matrixaddition und Matrixmultiplikation als Verknüpfungen ein Ring.

Beweis. Das ist eine unmittelbare Übersetzung der entsprechenden Aussagen für lineare Abbildungen, vergleiche den Beweis von Satz 10.25 (die Beweise etwa für die Distributivgesetze funktionieren auch in der etwas allgemeineren Situation, die hier vorliegt). Alternativ kann man das auch leicht direkt nachrechnen. ❑

11.10. **Definition.** Der Ring $\mathrm{Mat}(n, K)$ heißt der *Matrizenring (der Größe n über K)*. Eine Matrix $A \in \mathrm{Mat}(n, K)$ heißt *invertierbar*, wenn es eine Matrix $B \in \mathrm{Mat}(n, K)$ gibt mit $AB = I_n$. Dann gilt auch $BA = I_n$; wir schreiben A^{-1} für B und nennen B die *Inverse* von A. ♢ **DEF** Matrizenring invertierbare Matrix

Für die zu A und B gehörenden linearen Abbildungen $f, g\colon K^n \to K^n$ bedeutet $AB = I_n$, dass $f \circ g = \mathrm{id}_{K^n}$ ist. Dann ist f surjektiv, also ein Isomorphismus (siehe Folgerung 10.14) und $g = f^{-1}$, also ist auch $g \circ f = \mathrm{id}_{K^n}$, d.h., $BA = I_n$. Die Matrix $B = A^{-1}$ ist also die zu f^{-1} gehörende Matrix.

11.11. **Beispiel.** Die Matrix $A = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \in \mathrm{Mat}(2, K)$ (mit $\lambda \in K$ beliebig) ist invertierbar, denn **BSP**

$$\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\lambda \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \qquad ♣$$

Im nächsten Abschnitt werden wir lernen, wie wir Basen von Kern und Bild einer linearen Abbildung $f\colon K^n \to K^m$ anhand der zugehörigen Matrix berechnen können. Wir werden auch sehen, wie man feststellt, ob eine Matrix invertierbar ist, und wie man gegebenenfalls ihre Inverse findet.

12. Der Normalformalgorithmus und Lineare Gleichungssysteme

Wie können wir den Rang einer durch eine Matrix $A \in \mathrm{Mat}(m \times n, K)$ gegebenen linearen Abbildung $f\colon K^n \to K^m$ bestimmen und eine Basis ihres Kerns finden? Dazu überlegen wir uns, wie man die Matrix verändern kann, ohne dass sich der Kern ändert. Dann können wir versuchen, die Matrix in eine Form zu bringen, aus der sich zum Beispiel eine Basis des Kerns leicht ablesen lässt. Eine solche Form ist die *Zeilenstufenform*:

12.1. **Definition.** Seien K ein Körper, $m, n \in \mathbb{N}$ und $A = (a_{ij}) \in \mathrm{Mat}(m \times n, K)$. Die Matrix A ist in *Zeilenstufenform*, wenn sie folgende Form hat (ein Stern steht für einen beliebigen Eintrag): **DEF** Zeilenstufenform

$$A = \begin{pmatrix} 0\cdots0 & \mathbf{1} & *\cdots* & * & *\cdots* & * & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & \mathbf{1} & *\cdots* & * & *\cdots* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & \mathbf{1} & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \end{pmatrix}$$

Formal bedeutet das, dass es $0 \le r \le m$ und Indizes $1 \le j_1 < j_2 < \cdots < j_r \le n$ gibt, sodass $a_{ij} = 0$, wenn $i > r$ oder $i \le r$ und $j < j_i$, und $a_{ij_i} = 1$ für alle $i \in \{1, 2, \ldots, r\}$. (Dabei ist r die Anzahl der Zeilen, die nicht nur aus Nullen bestehen, und $j_1, j_2, \ldots, j_r$ sind die Nummern der Spalten, in denen die führenden Einsen der ersten r Zeilen stehen.)

A ist in *reduzierter Zeilenstufenform*, wenn zusätzlich $a_{ij_k} = 0$ ist für alle $1 \le i < k$ und alle $k \in \{1, 2, \ldots, r\}$:

$$A = \begin{pmatrix} 0\cdots0 & \mathbf{1} & *\cdots* & 0 & *\cdots* & 0 & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & \mathbf{1} & *\cdots* & 0 & *\cdots* \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & \mathbf{1} & *\cdots* \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 & 0 & 0\cdots0 \end{pmatrix}$$

Die „führenden Einsen“ der ersten r Zeilen (in den Spalten $j_1, j_2, \ldots, j_r$) sind durch Fettdruck hervorgehoben, die zugehörigen Spalten in der Bildschirmversion auch farblich abgesetzt. ♢

Zur Vereinfachung führen wir folgende Sprechweise ein:

12.2. **Definition.** Sei $A \in \mathrm{Mat}(m \times n, K)$. Dann ist das *Bild* von A, $\mathrm{im}(A)$, das Bild der zugehörigen linearen Abbildung $f\colon K^n \to K^m$, der *Rang* von A, $\mathrm{rk}(A)$, ist der Rang von f und der *Kern* von A, $\ker(A)$, ist der Kern von f. ♢ **DEF** Bild, Rang, Kern einer Matrix

Da die Bilder unter f der Standard-Basisvektoren von K^n gerade die Spalten von A sind, ist $\mathrm{im}(A)$ das Erzeugnis der Spalten von A. Deswegen heißt $\mathrm{im}(A)$ auch der *Spaltenraum von A*. Analog definiert man den *Zeilenraum von A* als den von den Zeilen von A erzeugten Untervektorraum von K^n. **DEF** Spaltenraum Zeilenraum

LEMMA Rang und Kern einer Matrix in ZSF

12.3. **Lemma.** *Sei $A \in \mathrm{Mat}(m\times n, K)$ in Zeilenstufenform mit r und $j_1, j_2, \ldots, j_r$ wie in Definition 12.1. Dann ist $\mathrm{rk}(A) = r$. Hat die Matrix sogar reduzierte Zeilenstufenform, dann erhalten wir eine Basis von $\ker(A)$ wie folgt:*
Sei $J = \{1, 2, \ldots, n\} \setminus \{j_1, j_2, \ldots, j_r\}$ die Menge der Indizes von Spalten ohne „führende Eins“ und sei für $j \in J$ der Vektor $b_j \in K^n$ definiert als

$$b_j = \mathbf{e}_j - \sum_{i=1}^{r} a_{ij}\mathbf{e}_{j_i}\,.$$

Dann ist $(b_j)_{j\in J}$ eine Basis von $\ker(A)$.

Etwas anschaulicher bekommen wir die Basis des Kerns so: Die Indizes in J, die den Spalten ohne führende Eins einer Zeile entsprechen, sind Positionen, für die wir die Komponenten frei wählen können. Wir setzen eine (die Position $j \in J$) davon auf 1, die anderen auf 0 und lösen die aus $Ab_j = \mathbf{0}$ entstehenden Gleichungen nach den übrigen Komponenten auf.

Beweis. Das Bild der zu A gehörenden linearen Abbildung f wird von den Spalten der Matrix erzeugt. Die Spalten mit den Nummern $j_1, j_2, \ldots, j_r$ sind linear unabhängig: Aus

$$\lambda_1(1,0,\ldots,0) + \lambda_2(*,1,0,\ldots,0) + \ldots + \lambda_r(*,\ldots,*,1,0,\ldots,0) = (0,\ldots,0)$$

folgt sukzessive $\lambda_r = 0$, …, $\lambda_2 = 0$, $\lambda_1 = 0$. Außerdem sind alle Spalten im Untervektorraum $\langle \mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_r\rangle$ der Dimension r von K^m enthalten. Also ist $\mathrm{im}(f) = \langle \mathbf{e}'_1, \mathbf{e}'_2, \ldots, \mathbf{e}'_r\rangle$; es folgt $\mathrm{rk}(A) = r$.

Wir nehmen jetzt an, dass die Matrix reduzierte Zeilenstufenform hat. Aus der Dimensionsformel in Satz 10.17 folgt, dass der Kern von A Dimension $n - r = \#J$ hat. Es genügt also zu zeigen, dass die b_j im Kern liegen und linear unabhängig sind. Wir schreiben A_j für die j-te Spalte von A. Dann ist $A_j = \sum_{i=1}^{r} a_{ij}\mathbf{e}'_i$ (dabei verwenden wir, dass $a_{ij} = 0$ ist für $i > r$), und $A_{j_i} = \mathbf{e}'_i$ für $1 \le i \le r$ (das ist Teil der Aussage, dass A reduzierte Zeilenstufenform hat). Es folgt

$$f(b_j) = A_j - \sum_{i=1}^{r} a_{ij}A_{j_i} = \sum_{i=1}^{r} a_{ij}\mathbf{e}'_i - \sum_{i=1}^{r} a_{ij}\mathbf{e}'_i = \mathbf{0}\,,$$

also ist b_j im Kern. Um zu zeigen, dass die b_j linear unabhängig sind, betrachten wir eine Linearkombination:

$$\mathbf{0} = \sum_{j\in J} \lambda_j b_j = \sum_{j\in J} \lambda_j \mathbf{e}_j - \sum_{i=1}^{r} \Big(\sum_{j\in J} a_{ij}\lambda_j\Big)\mathbf{e}_{j_i}$$

Da $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ eine Basis von K^n und $\{1, 2, \ldots, n\}$ die disjunkte Vereinigung von J und $\{j_1, j_2, \ldots, j_r\}$ ist, folgt $\lambda_j = 0$ für alle $j \in J$. ❑

BSP

12.4. **Beispiel.** Sei $K = \mathbb{R}$ und A die folgende Matrix über $\mathbb{R}$:

$$A = \begin{pmatrix} 0 & \mathbf{1} & 2 & 0 & 0 & -2 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 1 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Dann ist A in reduzierter Zeilenstufenform mit $r = 3$ (das ist die Anzahl der Zeilen, die keine Null-Zeilen sind) und $j_1 = 2$, $j_2 = 4$, $j_3 = 5$. Der Rang ist also 3, $J = \{1, 3, 6\}$ und eine Basis des Kerns ist gegeben durch

$$b_1 = (\mathbf{1}, 0, \mathbf{0}, 0, 0, \mathbf{0}), \quad b_3 = (\mathbf{0}, -2, \mathbf{1}, 0, 0, \mathbf{0}), \quad b_6 = (\mathbf{0}, 2, \mathbf{0}, -1, -5, \mathbf{1})\,.$$

Die frei wählbaren Komponenten (Positionen 1, 3, 6) sind durch Fettdruck hervorgehoben. Die restlichen Komponenten von b_j ergeben sich aus den Negativen der ersten r Einträge der j-ten Spalte von A.

Was hier passiert, wird vielleicht klarer, wenn man die Bedingung „$(x_1, x_2, \ldots, x_6) \in \ker(A)$“ explizit formuliert. Das führt auf das Gleichungssystem

$$\begin{array}{rcrcrcl} x_2 & + & 2x_3 & & & - \; 2x_6 & = 0 \\ & & & x_4 & & + \; x_6 & = 0 \\ & & & & x_5 & + \; 5x_6 & = 0 \end{array}$$

Dass A in reduzierter Zeilenstufenform ist, bedeutet, dass man dieses Gleichungssystem nach x_2, x_4, x_5 auflösen kann:

$$\begin{aligned} x_2 &= -2x_3 + 2x_6 \\ x_4 &= -x_6 \\ x_5 &= -5x_6 \end{aligned}$$

Man kann also x_1, x_3, x_6 beliebig vorgeben und daraus x_2, x_4, x_5 bestimmen. Wenn man (x_1, x_3, x_6) die Standardbasis von $K^3 = K^{n-j}$ durchlaufen lässt, dann bekommt man so eine Basis des Kerns. ♣

Wie bekommen wir nun eine Matrix in diese Zeilenstufenform, ohne ihren Kern zu ändern? Dazu gehen wir schrittweise vor und führen kleine Veränderungen durch, von denen man leicht einsehen kann, dass sie diese Eigenschaft haben.

12.5. **Definition.** Seien K ein Körper, $m, n \in \mathbb{N}$ und $A \in \mathrm{Mat}(m \times n, K)$.

DEF (elementare) Zeilenumformungen

(1) Eine *elementare Zeilenumformung vom Typ I* an der Matrix A besteht darin, die i-te Zeile von A mit λ zu multiplizieren. Dabei ist $i \in \{1, 2, \ldots, m\}$ und $\lambda \in K \setminus \{0\}$. Wir schreiben $\mathbf{I}_i(\lambda)$ für diese Umformung.

(2) Eine *elementare Zeilenumformung von Typ II* an der Matrix A besteht darin, das λ-fache der j-ten Zeile von A zur i-ten Zeile zu addieren. Dabei sind $i, j \in \{1, 2, \ldots, m\}$ mit $i \neq j$ und $\lambda \in K$. Wir schreiben $\mathbf{II}_{i,j}(\lambda)$ für diese Umformung.

(3) Eine *elementare Zeilenumformung vom Typ III* an der Matrix A besteht darin, in A die i-te und die j-te Zeile miteinander zu vertauschen. Dabei sind $i, j \in \{1, 2, \ldots, m\}$ mit $i \neq j$. Wir schreiben $\mathbf{III}_{i,j}$ für diese Umformung.

Eine *Zeilenumformung* an der Matrix A ist eine Abfolge von sukzessiven elementaren Zeilenumformungen, beginnend mit der Matrix A. ♢

Eine elementare Zeilenumformung vom Typ III kann durch eine Abfolge geeigneter Umformungen der Typen I und II erreicht werden (Übung). Diese Art der Umformung ist also eigentlich nicht nötig, stellt aber häufig eine praktische Abkürzung dar.

12.6. **Lemma.** *Seien K ein Körper, $m, n \in \mathbb{N}$ und $A \in \mathrm{Mat}(m \times n, K)$. Sei weiter A' eine Matrix, die aus A durch eine elementare Zeilenumformung hervorgeht. Dann ist $\ker(A') = \ker(A)$ und daher auch $\mathrm{rk}(A') = \mathrm{rk}(A)$. Außerdem haben A' und A denselben Zeilenraum.*

LEMMA Zeilenumf. erhalten Kern, Rang, Zeilenraum

Beweis. Ein Vektor $v = (x_1, x_2, \ldots, x_n) \in K^n$ ist genau dann im Kern von $A = (a_{ij})$, wenn für alle $i \in \{1, 2, \ldots, m\}$ gilt $\sum_{j=1}^{n} a_{ij}x_j = 0$. Eine elementare Zeilenumformung vom Typ I ersetzt eine dieser Gleichungen durch ihr λ-faches mit $\lambda \neq 0$, was ihre Gültigkeit nicht ändert. Bei einer elementaren Zeilenumformung vom Typ II wird zu einer der Gleichungen das λ-fache einer anderen Gleichung addiert, die neuen Gleichungen sind also gültig, wenn die alten es sind. Da man die Umformung rückgängig machen kann (durch Subtraktion des λ-fachen der j-ten Zeile von der i-ten), gelten die neuen Gleichungen genau dann, wenn die alten gelten. (Umformungen vom Typ III brauchen nicht extra betrachtet zu werden; da sie aber nur die Reihenfolge der Gleichungen ändern, ist klar, dass der Kern dabei erhalten bleibt.) Das zeigt, dass v genau dann im Kern von A ist, wenn v im Kern von A' ist. Die Gleichheit der Ränge folgt aus dem Rangsatz 10.17.

Elementare Zeilenumformungen ersetzen eine Zeile durch eine Linearkombination der Zeilen der Matrix; da diese Linearkombination im Zeilenraum der ursprünglichen Matrix liegt, ist der Zeilenraum der neuen Matrix jedenfalls im Zeilenraum der alten Matrix enthalten. Da sich elementare Zeilenumformungen rückgängig machen lassen, gilt auch die umgekehrte Inklusion. ❑

Durch Induktion (über die Anzahl der elementaren Zeilenumformungen) folgt dann sofort, dass Kern, Rang und Zeilenraum auch unter beliebigen Zeilenumformungen erhalten bleiben.

Der Rang bleibt unter Zeilenumformungen zwar erhalten, das Bild der Matrix kann sich jedoch ändern!

Wir zeigen jetzt, dass man jede Matrix durch Zeilenumformungen in Zeilenstufenform überführen kann.

* 12.7. **Satz.** *Seien K ein Körper, $m, n \in \mathbb{N}$ und $A \in \mathrm{Mat}(m \times n, K)$. Dann lässt sich A durch sukzessive elementare Zeilenumformungen in eine Matrix A' in reduzierter Zeilenstufenform überführen.*

SATZ Normalform von Matrizen

Beweis. Wir zeigen zuerst, dass sich A in (nicht notwendig reduzierte) Zeilenstufenform bringen lässt. Der Beweis dafür geht durch Induktion nach der Zeilenanzahl m. Im Fall $m = 0$ ist die Matrix (trivialerweise) bereits in Zeilenstufenform. Sei also $m > 0$ und die Behauptung für alle Matrizen mit weniger als m Zeilen schon gezeigt. Ist A die Nullmatrix, dann ist A in Zeilenstufenform und es ist nichts zu zeigen. Wir können also annehmen, dass A einen von null verschiedenen Eintrag hat. Sei j_1 der kleinste Index einer Spalte mit einem solchen Eintrag. Ist $a_{1j_1} = 0$, dann können wir durch eine Typ-III-Umformung erreichen, dass $a_{1j_1} \neq 0$ ist. Eine Umformung vom Typ I mit $\lambda = a_{1j_1}^{-1}$, angewandt auf die erste Zeile, ergibt $a_{1j_1} = 1$. Die Matrix hat jetzt die Form

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & a_{2j_1} & * & \cdots & * \\ & \vdots & & \vdots & & \vdots & \\ 0 & \cdots & 0 & a_{mj_1} & * & \cdots & * \end{pmatrix}$$

Durch die Umformungen $\mathbf{II}_{2,1}(-a_{2j_1}), \mathbf{II}_{3,1}(-a_{3j_1}), \dots, \mathbf{II}_{m,1}(-a_{mj_1})$ können wir die j_1-te Spalte unterhalb der ersten Zeile „ausräumen“, sodass wir nun die Form

$$\left(\begin{array}{cccc|ccc} 0 & \cdots & 0 & 1 & * & \cdots & * \\ \hline 0 & \cdots & 0 & 0 & & & \\ \vdots & & & \vdots & & \tilde{A} & \\ 0 & \cdots & 0 & 0 & & & \end{array}\right)$$

haben mit einer $(m-1) \times (n-j_1)$-Matrix $\tilde{A}$. Zeilenumformungen an $\tilde{A}$ können auch als Zeilenumformungen an dieser Matrix ausgeführt werden, ohne dass sich am linken Teil der Matrix etwas ändert. Nach Induktionsannahme kann nun $\tilde{A}$ durch Zeilenumformungen in Zeilenstufenform gebracht werden. Damit hat die gesamte Matrix ebenfalls Zeilenstufenform.

Wir führen jetzt noch für $k = 1, 2, \dots, r$ und $i = 1, 2, \dots, k-1$ die Umformungen $\mathbf{II}_{i,k}(-a_{ij_k})$ aus (mit dem jeweils aktuellen Wert des Eintrags a_{ij_k}) und räumen auf diese Weise auch noch den Teil der Spalten oberhalb der führenden Einsen aus. Wir erhalten so die reduzierte Zeilenstufenform. ❑

Dieser Beweis liefert uns sogar einen Algorithmus. Wir werden die Umformungen an einer Beispielmatrix durchführen.

BSP
Umformung in ZSF

12.8. **Beispiel.** Wir bestimmen die reduzierte Zeilenstufenform der folgenden Matrix über $\mathbb{R}$:

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}$$

Die erste Spalte ist keine Null-Spalte, also ist $j_1 = 1$. Der oberste Eintrag in der ersten Spalte ist bereits 1, also sind keine Umformungen vom Typ III oder I nötig. Wir räumen den Rest der Spalte aus, indem wir das Fünffache der ersten Zeile von der zweiten und das Neunfache der ersten Zeile von der dritten Zeile abziehen. Dann bekommen wir die neue Matrix

$$\begin{pmatrix} \mathbf{1} & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & -8 & -16 & -24 \end{pmatrix}$$

Wir machen mit der rechten unteren 2×3-Matrix weiter. Ihre erste Spalte $(-4, -8)$ ist keine Null-Spalte, also ist $j_2 = 2$. Wir multiplizieren die zweite Zeile der gesamten Matrix mit $-1/4$ und bekommen

$$\begin{pmatrix} \mathbf{1} & 2 & 3 & 4 \\ 0 & \mathbf{1} & 2 & 3 \\ 0 & -8 & -16 & -24 \end{pmatrix}.$$

Dann addieren wir das Achtfache der zweiten Zeile zur dritten:

$$\begin{pmatrix} \mathbf{1} & 2 & 3 & 4 \\ 0 & \mathbf{1} & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Damit haben wir die Zeilenstufenform erreicht (mit $r = 2$). Für die reduzierte Zeilenstufenform müssen wir noch das Doppelte der zweiten Zeile von der ersten abziehen; das liefert schließlich

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

Wir können jetzt eine Basis des Kerns von A (der gleich dem Kern von A' ist) ablesen, nämlich

$$b_3 = (1, -2, \mathbf{1}, \mathbf{0}) \qquad \text{und} \qquad b_4 = (2, -3, \mathbf{0}, \mathbf{1})\,. \qquad \clubsuit$$

Elementare Zeilenumformungen lassen sich durch Multiplikation mit gewissen invertierbaren Matrizen von links beschreiben. Diese Matrizen sind die sogenannten *Elementarmatrizen* $E_i(\lambda)$ mit $\lambda \in K^\times$ und $i \in \{1, 2, \ldots, m\}$ und $E_{ij}(\lambda)$ mit $\lambda \in K$ und $i, j \in \{1, 2, \ldots, m\}$, $i \neq j$. Um sie zu definieren, führen wir $M_{kl} = (\delta_{ik}\delta_{jl})_{1\leq i,j\leq m}$ ein; in dieser Matrix sind alle Einträge null bis auf den Eintrag in der k-ten Zeile und l-ten Spalte, der den Wert 1 hat. (Die Matrizen M_{kl} entsprechen der Basis $(\phi_{kl})_{1\leq k,l\leq m}$ von $\mathrm{Hom}(K^m, K^m)$ wie in Folgerung 10.23.) Dann ist

$$E_i(\lambda) = I_m + (\lambda - 1)M_{ii} \qquad \text{und} \qquad E_{ij}(\lambda) = I_m + \lambda M_{ij}\,.$$

$E_i(\lambda)$ unterscheidet sich von der Einheitsmatrix I_m dadurch, dass an der i-ten Position auf der Diagonalen statt 1 der Eintrag λ steht. In $E_{ij}(\lambda)$ steht außerhalb der Diagonalen an der Position (i, j) der Eintrag λ. Wegen

$$E_i(\lambda)E_i(\lambda^{-1}) = I_m \qquad \text{und} \qquad E_{ij}(\lambda)E_{ij}(-\lambda) = I_m$$

sind diese Elementarmatrizen invertierbar. Was bewirkt die Multiplikation von links mit so einer Elementarmatrix? Dazu überlegen wir, dass

$$M_{kl}A = \Big(\sum_{h=1}^{m} \delta_{ik}\delta_{hl}a_{hj}\Big)_{1\leq i\leq m, 1\leq j\leq n} = (\delta_{ik}a_{lj})_{1\leq i\leq m, 1\leq j\leq n}\,;$$

in dieser Matrix sind alle Zeilen null bis auf die k-te Zeile, in welcher sich die l-te Zeile von A befindet. Multiplikation von links mit M_{kl} setzt also die l-te Zeile von A in die k-te Zeile und löscht alle anderen Zeilen.

Damit ergibt sich, dass die Zeilen von $E_i(\lambda)A$ mit den entsprechenden Zeilen von A übereinstimmen bis auf die i-te Zeile, die mit λ multipliziert wird. Der Effekt ist also die elementare Zeilenumformung $\mathbf{I}_i(\lambda)$ vom Typ I. Ebenso stimmen die Zeilen von $E_{ij}(\lambda)A$ mit denen von A überein mit Ausnahme der i-ten Zeile, zu der das λ-fache der j-ten Zeile addiert wird. Der Effekt ist also die elementare Zeilenumformung $\mathbf{II}_{i,j}(\lambda)$ vom Typ II. Wir veranschaulichen das für die 2×3-Matrix $A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$:

$$\begin{aligned}
E_1(\lambda)A &= \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b & \lambda c \\ d & e & f \end{pmatrix} \\
E_2(\lambda)A &= \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} a & b & c \\ \lambda d & \lambda e & \lambda f \end{pmatrix} \\
E_{12}(\lambda)A &= \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} a + \lambda d & b + \lambda e & c + \lambda f \\ d & e & f \end{pmatrix} \\
E_{21}(\lambda)A &= \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} = \begin{pmatrix} a & b & c \\ d + \lambda a & e + \lambda b & f + \lambda c \end{pmatrix}
\end{aligned}$$

Der Inhalt von Satz 12.7 ist also, dass es zu jeder Matrix $A \in \mathrm{Mat}(m \times n, K)$ eine invertierbare Matrix $P \in \mathrm{Mat}(m, K)$ gibt, sodass PA reduzierte Zeilenstufenform hat, wobei P ein Produkt von Elementarmatrizen ist.

Sei jetzt $A \in \mathrm{Mat}(m, K)$ invertierbar. Wendet man diese Aussage an auf A^{-1} und beachtet, dass die reduzierte Zeilenstufenform einer invertierbaren $m \times m$-Matrix gerade die Einheitsmatrix I_m ist (siehe Lemma 12.18 unten), dann erhält man ein Produkt P von Elementarmatrizen mit $PA^{-1} = I_m$. Es folgt $A = P$. Wir haben bewiesen:

Satz. *Jede invertierbare Matrix ist ein Produkt von Elementarmatrizen.*

SATZ Elementarmatrizen erzeugen invertierbare Matrizen

Daraus folgt:

Zwei Matrizen $A, B \in \mathrm{Mat}(m \times n, K)$ lassen sich genau dann durch Zeilenumformungen ineinander überführen, wenn es eine invertierbare Matrix $P \in \mathrm{Mat}(m, K)$ gibt, sodass $B = PA$ ist.

Statt Zeilenumformungen kann man ganz analog *Spaltenumformungen* betrachten. Sie werden durch Multiplikation mit Elementarmatrizen von *rechts* bewirkt. Man hat dann die folgende analoge Aussage:

Zwei Matrizen $A, B \in \mathrm{Mat}(m \times n, K)$ lassen sich genau dann durch Spaltenumformungen ineinander überführen, wenn es eine invertierbare Matrix $Q \in \mathrm{Mat}(n, K)$ gibt, sodass $B = AQ$ ist.

Wir sprechen gelegentlich von „der“ reduzierten Zeilenstufenform einer Matrix. Tatsächlich ist das Ergebnis eines Verfahrens, das eine Matrix in reduzierte Zeilenstufenform überführt, eindeutig bestimmt, wie der folgende Satz zeigt. Es ist also letztlich ganz egal, welche Zeilenumformungen man in welcher Reihenfolge macht, um zur reduzierten Zeilenstufenform zu gelangen.

Satz. *Sind $A, B \in \mathrm{Mat}(m \times n, K)$ zwei Matrizen in reduzierter Zeilenstufenform mit demselben Zeilenraum, dann gilt $A = B$.*

SATZ Eindeutigkeit der Zeilenstufenform

Beweis. Sei $U \subset K^n$ der Zeilenraum von A und B. Für $0 \le k \le n$ sei

$$V_k = \{(x_1, \dots, x_n) \in K^n \mid x_1 = x_2 = \dots = x_k = 0\} \subset K^n$$

und $d_k = \dim(U \cap V_k)$. Dann ist $d_0 = r = \dim U$, $d_n = 0$ und $d_k - 1 \le d_{k+1} \le d_k$. Es gibt also genau r „Sprungstellen“ j_i mit $d_{j_i - 1} = r + 1 - i$ und $d_{j_i} = r - i$ für $i \in \{1, 2, \dots, r\}$. Aus der Definition der Zeilenstufenform ergibt sich, dass j_i genau die Position der führenden Eins in der i-ten Zeile von A und von B ist. Die lineare Abbildung

$$\phi \colon U \longrightarrow K^r, \qquad (x_1, x_2, \dots, x_n) \longmapsto (x_{j_1}, x_{j_2}, \dots, x_{j_r})$$

ist dann ein Isomorphismus, und die ersten r Zeilen von A und B müssen die Urbilder $\phi^{-1}(\mathbf{e}_1), \phi^{-1}(\mathbf{e}_2), \dots, \phi^{-1}(\mathbf{e}_r)$ der Standard-Basisvektoren von K^r sein. Insbesondere sind A und B gleich. ❑

Wir kommen zu linearen Gleichungen und Gleichungssystemen.

12.9. **Definition.** Seien V und W zwei K-Vektorräume und $f \colon V \to W$ eine lineare Abbildung. Ist $b \in W$ ein gegebener Vektor, dann heißt die Gleichung $f(x) = b$, deren Lösungen $x \in V$ gesucht sind, eine *lineare Gleichung*. Die Gleichung heißt *homogen*, wenn $b = \mathbf{0}$ ist, sonst *inhomogen*.

DEF Lineare Gleichung

Ist $V = K^n$ und $W = K^m$, dann kann die Gleichung unter Benutzung der zu f gehörenden Matrix $A \in \mathrm{Mat}(m \times n, K)$ auch geschrieben werden als $A\boldsymbol{x} = \boldsymbol{b}$ mit Spaltenvektoren $\boldsymbol{x} \in K^n$ und $\boldsymbol{b} \in K^m$. In diesem Fall spricht man auch von einem *linearen Gleichungssystem* (mit m Gleichungen in n Unbestimmten). ♢

Wir können schon recht genau sagen, welche Struktur die Lösungsmenge einer linearen Gleichung hat.

* 12.10. **Satz.** *Seien V und W zwei K-Vektorräume und $f\colon V \to W$ eine lineare Abbildung.*

SATZ Lösungsmenge einer linearen Gleichung

(1) *Die Lösungsmenge der homogenen linearen Gleichung $f(x) = \mathbf{0}$ ist ein Untervektorraum von V, nämlich der Kern von f.*

(2) *Sei $\mathbf{0} \neq b \in W$. Ist $b \notin \operatorname{im}(f)$, dann hat die inhomogene lineare Gleichung $f(x) = b$ keine Lösung. Anderenfalls sei $x_0 \in V$ mit $f(x_0) = b$. Dann ist die Lösungsmenge gegeben durch $x_0 + \ker(f) = \{x_0 + v \mid v \in \ker(f)\}$.*

Beweis. Die erste Aussage folgt direkt aus der Definition des Kerns und der Tatsache, dass $\ker(f)$ ein Untervektorraum von V ist. In der zweiten Aussage ist klar, dass es genau dann Lösungen gibt, wenn $b \in \operatorname{im}(f)$ ist (das ist die Definition von $\operatorname{im}(f)$). Es bleibt die letzte Behauptung zu zeigen. Sei dazu $x \in V$. Dann gilt

$$\begin{aligned} f(x) = b &\iff f(x) = f(x_0) \iff f(x - x_0) = \mathbf{0} \\ &\iff x - x_0 \in \ker(f) \iff x \in x_0 + \ker(f)\,. \end{aligned}$$ ❑

Das allgemeine Rezept für die Lösung einer linearen Gleichung $f(x) = b$ lautet also:

(1) Prüfe, ob $b \in \operatorname{im}(f)$. Falls nein, dann gibt es keine Lösung.

(2) Bestimme eine „spezielle Lösung“ $x_0 \in V$.

(3) Bestimme $\ker(f)$.

(4) Die Lösungsmenge ist $x_0 + \ker(f)$. Ist $\ker(f)$ endlich-dimensional mit Basis $(x_1, x_2, \ldots, x_n)$, dann ist die „allgemeine Lösung“

$$x = x_0 + \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n$$

mit $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$.

Die ersten beiden Schritte wird man im Regelfall zusammen ausführen, denn wenn man feststellt, dass $b \in \operatorname{im}(f)$ ist, dann wird man meistens auch ein Urbild gefunden haben.

Im homogenen Fall (also $b = \mathbf{0}$) gilt stets $b \in \operatorname{im}(f)$ und wir können $x_0 = \mathbf{0}$ nehmen; die Lösungsmenge ist dann $\ker(f)$.

12.11. **Beispiel.** Wir betrachten die folgende inhomogene lineare Differentialgleichung erster Ordnung:

BSP inhomogene lineare Diff.gleichung

$$y'(x) + y(x) = x\,.$$

Dabei sei $y \in \mathcal{C}^1(\mathbb{R})$. Hier ist $K = \mathbb{R}$, $V = \mathcal{C}^1(\mathbb{R})$, $W = \mathcal{C}(\mathbb{R})$ und $f\colon y \mapsto y' + y$. Die Gleichung ist $f(y) = \mathrm{id}_{\mathbb{R}}$. Wir suchen nach einer speziellen Lösung. Mit etwas Probieren finden wir $y_0(x) = x - 1$. (In der Vorlesung über *Gewöhnliche Differentialgleichungen* werden Sie lernen, wie man solche Lösungen systematisch findet.) Jetzt müssen wir den Kern von f bestimmen, also die Menge aller Funktionen y mit $y' + y = \mathbf{0}$. Ich behaupte, dass $\ker(f) = \langle x \mapsto e^{-x} \rangle$ ist; die Funktionen y mit $y' + y = \mathbf{0}$ haben also die Form $y(x) = Ce^{-x}$ mit $C \in \mathbb{R}$. Zum Beweis betrachten wir $z(x) = e^x y(x)$; dann gilt

$$z'(x) = e^x y(x) + e^x y'(x) = e^x \big(y(x) + y'(x)\big) = \mathbf{0}\,,$$

also ist $z(x) = C$ konstant und damit $y(x) = Ce^{-x}$. Umgekehrt sind diese Funktionen auch Lösungen von $y' + y = \mathbf{0}$. Die allgemeine Lösung ist also

$$y(x) = x - 1 + Ce^{-x}\,, \qquad C \in \mathbb{R}\,.$$ ♣

Wie sieht das obige Rezept konkret aus, wenn wir ein lineares Gleichungssystem lösen wollen? Sei $A\boldsymbol{x} = \boldsymbol{b}$ ein lineares Gleichungssystem mit $A \in \mathrm{Mat}(m \times n, K)$. Im homogenen Fall $\boldsymbol{b} = \boldsymbol{0}$ müssen wir eine Basis von $\ker(A)$ bestimmen. Dazu bringen wir A in reduzierte Zeilenstufenform und lesen eine Basis des Kerns ab wie in Lemma 12.3. Im inhomogenen Fall sei $A' = (A \mid \boldsymbol{b})$ die *erweiterte Matrix* des Systems; wir erhalten sie, indem wir an die Matrix A den Spaltenvektor $\boldsymbol{b}$ als $(n+1)$-te Spalte anfügen.

SATZ inhom. LGS

12.12. **Satz.** *Sei K ein Körper, seien $m, n \in \mathbb{N}$, sei $A \in \mathrm{Mat}(m \times n, K)$ und sei $\boldsymbol{b} \in K^m$ ein Spaltenvektor. Sei weiter $A' = (A \mid \boldsymbol{b})$. Dann gilt*

$$\boldsymbol{b} \in \mathrm{im}(A) \iff \mathrm{rk}(A') = \mathrm{rk}(A)\,.$$

Dies kann geprüft werden, indem A' in reduzierte Zeilenstufenform $\tilde{A}'$ gebracht wird. $\mathrm{rk}(A') = \mathrm{rk}(A)$ *ist dann dazu äquivalent, dass die letzte Spalte von $\tilde{A}'$ keine führende Eins einer Zeile enthält (das bedeutet $j_r \leq n$ in der Notation von Definition 12.1). In diesem Fall kann eine spezielle Lösung von $A\boldsymbol{x} = \boldsymbol{b}$ aus $\tilde{A}'$ wie folgt abgelesen werden: Die letzte Spalte von $\tilde{A}'$ sei $(\tilde{b}_1, \ldots, \tilde{b}_r, 0, \ldots, 0)$. Dann ist*

$$\boldsymbol{x}_0 = \sum_{i=1}^{r} \tilde{b}_i \mathbf{e}_{j_i}$$

eine Lösung des Gleichungssystems.

Beweis. Seien $A_1, \ldots, A_n$ die Spalten von A. Es gilt

$$\begin{aligned} \boldsymbol{b} \in \mathrm{im}(A) = \langle A_1, \ldots, A_n \rangle &\iff \langle A_1, \ldots, A_n, \boldsymbol{b} \rangle = \langle A_1, \ldots, A_n \rangle \\ &\iff \mathrm{im}(A') = \mathrm{im}(A)\,. \end{aligned}$$

Die letzte Aussage impliziert $\mathrm{rk}(A') = \mathrm{rk}(A)$. Es gilt immer $\mathrm{im}(A) \subset \mathrm{im}(A')$, also folgt aus $\mathrm{rk}(A') = \mathrm{rk}(A)$ auch die Gleichheit von $\mathrm{im}(A')$ und $\mathrm{im}(A)$. Damit ist die erste Behauptung gezeigt.

Sei nun $\tilde{A}'$ die reduzierte Zeilenstufenform von A'. Dann bilden die ersten n Spalten von $\tilde{A}'$ die reduzierte Zeilenstufenform $\tilde{A}$ von A. Der Rang von A' ist genau dann größer als der Rang von A, wenn $\tilde{A}'$ mehr Nichtnull-Zeilen hat als $\tilde{A}$. Das bedeutet aber gerade, dass die letzte Spalte von $\tilde{A}'$ eine führende Eins enthalten muss. Das zeigt die zweite Aussage. Für die letzte Aussage beachten wir, dass die Zeilenumformungen, die im Zuge der Herstellung der reduzierten Zeilenstufenform durchgeführt werden, die ursprünglichen Gleichungen durch äquivalente Gleichungen ersetzen. Mit $\tilde{A}' = (\tilde{A} \mid \tilde{\boldsymbol{b}})$ hat also das lineare Gleichungssystem $\tilde{A}\boldsymbol{x} = \tilde{\boldsymbol{b}}$ dieselben Lösungen wie das ursprüngliche Gleichungssystem. Da die j_i-te Spalte von $\tilde{A}$ gerade der Standard-Basisvektor $\mathbf{e}'_i$ ist, ergibt sich

$$\tilde{A}\boldsymbol{x}_0 = \sum_{i=1}^{r} \tilde{b}_i \tilde{A}\mathbf{e}_{j_i} = \sum_{i=1}^{r} \tilde{b}_i \mathbf{e}'_i = \tilde{\boldsymbol{b}}\,. \qquad \square$$

BSP LGS

12.13. **Beispiel.** Wir lösen das folgende lineare Gleichungssystem (mit $K = \mathbb{Q}$ oder $\mathbb{R}$):

$$\begin{array}{rcrcrcrcr} x_1 & & & - & x_3 & - & 2x_4 & = & 3 \\ -x_1 & + & x_2 & + & x_3 & & & = & -2 \\ & & x_2 & & & - & x_4 & = & 0 \end{array}$$

oder, in Matrixschreibweise,

$$\begin{pmatrix} 1 & 0 & -1 & -2 \\ -1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \\ 0 \end{pmatrix} .$$

Die erweiterte Matrix ist

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

Ihre reduzierte Zeilenstufenform ergibt sich als

$$\tilde{A}' = \begin{pmatrix} \mathbf{1} & 0 & -1 & 0 & 1 \\ 0 & \mathbf{1} & 0 & 0 & -1 \\ 0 & 0 & 0 & \mathbf{1} & -1 \end{pmatrix} .$$

Das zugehörige lineare Gleichungssystem ist:

$$\begin{array}{rcrcrcl} x_1 & & & - & x_3 & & = & 1 \\ & & x_2 & & & & = & -1 \\ & & & & & x_4 & = & -1 \end{array}$$

Als spezielle Lösung erhalten wir daraus $\boldsymbol{x}_0 = (\mathbf{1}, -\mathbf{1}, 0, -\mathbf{1})$. Außerdem lesen wir ab: $\operatorname{rk}(A) = 3$, $\dim \ker(A) = 4 - 3 = 1$, und eine Basis von $\ker(A)$ ist gegeben durch $\boldsymbol{x}_1 = (1, 0, \mathbf{1}, 0)$. Die allgemeine Lösung des Gleichungssystems ist also

$$\boldsymbol{x} = \boldsymbol{x}_0 + \lambda_1 \boldsymbol{x}_1 = (1 + \lambda_1, -1, \lambda_1, -1)$$

mit $\lambda_1 \in K$. ♣

Hier ist das Rezept noch einmal ganz konkret:

(1) Die erweiterte Matrix A' aufstellen.

(2) A' in reduzierte Zeilenstufenform $\tilde{A}'$ bringen. Sei $r = \operatorname{rk}(\tilde{A}')$ und seien $1 \leq j_1 < j_2 < \ldots < j_r \leq n+1$ die Positionen der führenden Einsen der ersten r Zeilen von $\tilde{A}'$.

(3) $j_r = n + 1 \Rightarrow$ keine Lösung. Anderenfalls:

(4) Sei $J = \{1, 2, \ldots, n\} \setminus \{j_1, j_2, \ldots, j_r\}$ die Menge der „freien“ Positionen. Setze $x_j = \lambda_j \in K$ beliebig für $j \in J$ und löse das der Matrix $\tilde{A}'$ entsprechende Gleichungssystem nach x_{j_i}, $i \in \{1, 2, \ldots, r\}$ auf. Das ergibt die allgemeine Lösung.

C.F. Gauß
(1777–1855)

Im Beispiel oben ist $r = 3$, $j_1 = 1$, $j_2 = 2$, $j_3 = 4 < 5 = n + 1$, $J = \{3\}$. Wir setzen also $x_3 = \lambda$ und lösen das System nach x_1, x_2, x_4 auf.

Diese Methode für die Lösung linearer Gleichungssysteme (und ihre Varianten) heißt *gaußsches Eliminationsverfahren* oder kürzer *Gauß-Elimination*. Eine Variante besteht darin, statt der reduzierten Zeilenstufenform nur die Zeilenstufenform herzustellen und dann das System schrittweise „von unten her“ durch Einsetzen zu lösen. Diese Version ist etwas effizienter im Hinblick auf die Zahl der nötigen Rechenoperationen, dafür aber auch etwas umständlicher durchzuführen.

Wir haben den Rang einer Matrix als den Rang der zugehörigen linearen Abbildung definiert, also als die Dimension ihres Spaltenraums. Man sollte also eigentlich genauer vom „Spaltenrang“ sprechen, denn man könnte genauso gut die

Dimension des Zeilenraums, also den „Zeilenrang“ betrachten. Zum Glück macht das keinen Unterschied, wie wir jetzt zeigen werden.

12.14. **Satz.** *Sei K ein Körper, seien $m, n \in \mathbb{N}$ und $A \in \mathrm{Mat}(m \times n, K)$. Dann ist die Dimension des Zeilenraums von A gleich der Dimension des Spaltenraums von A.* **SATZ** Zeilenrang = Spaltenrang

Beweis. Nach Lemma 12.6 und Satz 12.7 können wir annehmen, dass A reduzierte Zeilenstufenform hat. Sei $r = \mathrm{rk}(A)$ die Dimension des Spaltenraums von A. Dann hat A genau r Zeilen, die keine Null-Zeilen sind, und diese Zeilen sind linear unabhängig, denn das gilt bereits, wenn man nur die Spalten $j_1, j_2, \ldots, j_r$ (Notation wie in Definition 12.1) betrachtet — die Matrix $(a_{i,j_k})_{1 \le i,k \le r}$ ist die Einheitsmatrix I_r. Da diese Zeilen den Zeilenraum erzeugen, bilden sie also eine Basis; damit ist die Dimension des Zeilenraums $r = \mathrm{rk}(A)$. ❑

Der Normalformalgorithmus aus Satz 12.7 berechnet demnach auch die Dimension und eine Basis des Zeilenraums der gegebenen Matrix. Wenn man also die Dimension und eine Basis des von Vektoren $v_1, \ldots, v_m \in K^n$ erzeugten Untervektorraums bestimmen möchte, dann schreibt man diese Vektoren als Zeilen in eine Matrix und bestimmt ihre (reduzierte) Zeilenstufenform. Die von null verschiedenen Zeilen der resultierenden Matrix bilden dann eine Basis.

Man kann den Satz kurz und elegant formulieren, wenn man folgende Definition verwendet.

∗ 12.15. **Definition.** Sei K ein Körper, seien $m, n \in \mathbb{N}$ und sei $A = (a_{ij}) \in \mathrm{Mat}(m \times n, K)$. Die *Transponierte* von A oder die *zu A transponierte Matrix* ist $A^\top = (a_{ji})_{1 \le i \le n, 1 \le j \le m} \in \mathrm{Mat}(n \times m, K)$. ♢ **DEF** Transponierte Matrix

Da es leicht zu Verwirrung führt: Die Schreibweise

$$A^\top = (a_{ji})_{1 \le i \le n, 1 \le j \le m}$$

bedeutet Folgendes: Der erste Index unten hinter der Klammer (hier i) ist der Zeilenindex und der zweite (hier j) ist der Spaltenindex. Die Matrix $A^\top$ hat also n Zeilen und m Spalten. Der Eintrag in Zeile i und Spalte j ist a_{ji} und damit derselbe Eintrag wie in *Spalte i* und *Zeile j* der Matrix A. Gleichbedeutend könnte man auch

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

schreiben. In diesem Fall wäre j der Zeilen- und i der Spaltenindex.

Die Matrix wird also „an der Hauptdiagonale gespiegelt“.

12.16. **Beispiel.**

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^\top = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$ ♣ **BSP** Transponierte Matrix

Der Zeilenraum von A ist der Spaltenraum von $A^\top$ und umgekehrt. Satz 12.14 sagt also

$$\mathrm{rk}(A^\top) = \mathrm{rk}(A)\,.$$

Hier sind die wichtigsten Rechenregeln für transponierte Matrizen:

LEMMA Rechenregeln für $A^\top$

12.17. **Lemma.** *Sei K ein Körper und seien $l, m, n \in \mathbb{N}$.*

(1) *Die Abbildung $\mathrm{Mat}(m \times n, K) \to \mathrm{Mat}(n \times m, K)$, $A \mapsto A^\top$ ist ein Isomorphismus (es gilt also $(A+B)^\top = A^\top + B^\top$ und $(\lambda A)^\top = \lambda A^\top$ für $A, B \in \mathrm{Mat}(m \times n, K)$, $\lambda \in K$; die Bijektivität ist klar).*

(2) *Für $A \in \mathrm{Mat}(l \times m, K)$ und $B \in \mathrm{Mat}(m \times n, K)$ gilt $(AB)^\top = B^\top A^\top$.*

(3) *Für $A \in \mathrm{Mat}(m \times n, K)$ gilt $(A^\top)^\top = A$.*

Beweis. Übung. ❑

Abschließend wollen wir noch Antworten auf die folgenden zwei Fragen überlegen:

Frage 1: Wie kann man feststellen, ob eine Matrix $A \in \mathrm{Mat}(n, K)$ invertierbar ist?

Frage 2: Wie kann man zu einer invertierbaren Matrix $A \in \mathrm{Mat}(n, K)$ die Inverse A^{-1} berechnen?

Die erste Frage wird durch das folgende Lemma beantwortet:

LEMMA ZSF einer invertierbaren Matrix

12.18. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$ eine quadratische Matrix. A ist genau dann invertierbar, wenn ihre reduzierte Zeilenstufenform die Einheitsmatrix I_n ist.*

Beweis. A ist genau dann invertierbar, wenn die zugehörige lineare Abbildung $f\colon K^n \to K^n$ ein Isomorphismus ist. Da Definitions- und Wertebereich dieselbe Dimension haben, ist das dazu äquivalent, dass f surjektiv ist, also Rang n hat (siehe Folgerung 10.14). Das bedeutet, dass es in der reduzierten Zeilenstufenform von A keine Null-Zeile gibt, also ist $r = n$ und $j_1 = 1, j_2 = 2, \ldots, j_n = n$. In der j-ten Spalte steht also der j-te Standard-Basisvektor, und die Matrix ist die Einheitsmatrix. ❑

Jetzt zur zweiten Frage. Dazu beachten wir, dass ein lineares Gleichungssystem $A\boldsymbol{x} = \boldsymbol{b}$ für jedes $\boldsymbol{b}$ eine eindeutige Lösung hat, wenn A invertierbar ist; diese Lösung ist $\boldsymbol{x} = A^{-1}\boldsymbol{b}$. (Die Umkehrung gilt ebenfalls — gibt es für jedes $\boldsymbol{b}$ eine eindeutige Lösung, dann ist A invertierbar — Übung.) Wenn wir für $\boldsymbol{b}$ den Standard-Basisvektor $\mathbf{e}_j$ einsetzen, dann bekommen wir als Lösung gerade die j-te Spalte von A^{-1}. Wir können also A^{-1} finden, indem wir die linearen Gleichungssysteme $A\boldsymbol{x} = \mathbf{e}_j$ für $j \in \{1, 2, \ldots, n\}$ alle lösen. Dies geht im Wesentlichen in einem Rutsch, wie im nächsten Satz beschrieben wird.

* **SATZ** Berechnung von A^{-1}

12.19. **Satz.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$ eine quadratische Matrix. Sei weiter $A' = (A \mid I_n) \in \mathrm{Mat}(n \times 2n, K)$ und $\tilde{A}'$ ihre reduzierte Zeilenstufenform. A ist genau dann invertierbar, wenn $\tilde{A}'$ die Form $(I_n \mid B)$ hat; in diesem Fall ist $B = A^{-1}$.*

Beweis. Sei $\tilde{A}' = (\tilde{A} \mid B)$, dann ist $\tilde{A}$ die reduzierte Zeilenstufenform von A. Nach Lemma 12.18 ist A genau dann invertierbar, wenn $\tilde{A} = I_n$ ist. Die Matrix A' repräsentiert das Gleichungssystem $A(\boldsymbol{x}_1 \mid \boldsymbol{x}_2 \mid \cdots \mid \boldsymbol{x}_n) = (\mathbf{e}_1 \mid \mathbf{e}_2 \mid \cdots \mid \mathbf{e}_n)$ oder kurz $AX = I_n$ mit $X = (\boldsymbol{x}_1 \mid \boldsymbol{x}_2 \mid \cdots \mid \boldsymbol{x}_n) \in \mathrm{Mat}(n, K)$. Die Zeilenumformungen, die zur reduzierten Zeilenstufenform führen, ergeben das dazu äquivalente Gleichungssystem $I_n X = B$, also ist $X = B$ die Lösung von $AX = I_n$; damit ist $B = A^{-1}$. ❑

BSP
Berechnung von A^{-1}

12.20. **Beispiel.** Sei $K = \mathbb{Q}$ und

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

Wir überführen

$$A' = \left(\begin{array}{cccc|cccc} -1 & -2 & 2 & 2 & 1 & 0 & 0 & 0 \\ 2 & 5 & -4 & -4 & 0 & 1 & 0 & 0 \\ -1 & -1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{array}\right)$$

in reduzierte Zeilenstufenform:

$$A' \longrightarrow \left(\begin{array}{cccc|cccc} \mathbf{1} & 2 & -2 & -2 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 3 & 1 & 0 & 0 & 1 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|cccc} \mathbf{1} & 0 & -2 & -2 & -5 & -2 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & -3 & -1 & 1 & 0 \\ 0 & 0 & 2 & 3 & 3 & 1 & 0 & 1 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|cccc} \mathbf{1} & 0 & 0 & 0 & 1 & 0 & -2 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 1 & 3 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -3 & -1 & 2 & 1 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|cccc} \mathbf{1} & 0 & 0 & 0 & 1 & 0 & -2 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 6 & 2 & -3 & -1 \\ 0 & 0 & 0 & \mathbf{1} & -3 & -1 & 2 & 1 \end{array}\right)$$

Es folgt

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

♣

13. Matrizen und lineare Abbildungen

Wir haben bisher Matrizen als zu linearen Abbildungen $K^n \to K^m$ gehörend betrachtet. Dabei war es aber eigentlich nur wichtig, dass wir in Definitions- und Wertebereich jeweils eine bestimmte Basis betrachten, in diesem Fall die Standard-Basis. Ganz analog können wir einer K-linearen Abbildung $f\colon V \to V'$ eine Matrix zuordnen, wenn wir Basen $B = (b_1, b_2, \ldots, b_n)$ von V und $B' = (b'_1, b'_2, \ldots, b'_m)$ von V' fixieren. Es gibt dann nämlich eindeutig bestimmte Skalare $a_{ij} \in K$, sodass

$$f(b_j) = a_{1j}b'_1 + a_{2j}b'_2 + \ldots + a_{mj}b'_m$$

für alle $j \in \{1, 2, \ldots, n\}$ gilt.

DEF $\mathrm{Mat}_{B,B'}(f)$

13.1. **Definition.** In der oben beschriebenen Situation heißt

$$\mathrm{Mat}_{B,B'}(f) = (a_{ij})_{1\le i\le m, 1\le j\le n} \in \mathrm{Mat}(m \times n, K)$$

die *Matrix von f bezüglich der Basen B und B'.* ♢

Wie vorher auch enthält die j-te Spalte der Matrix die Koeffizienten des Bildes $f(b_j)$ des j-ten Basisvektors in B, wenn es als Linearkombination der Basisvektoren in B' geschrieben wird.

Man kann $\mathrm{Mat}_{B,B'}(f)$ auch beschreiben als die Matrix, die zu der linearen Abbildung $\phi_{B'}^{-1} \circ f \circ \phi_B\colon K^n \to K^m$ gehört.

$$\begin{array}{ccc} V & \xleftarrow{\phi_B} & K^n \\ {\scriptstyle f}\downarrow & & \vdots \\ & & \vdots\ {\scriptstyle \mathrm{Mat}_{B,B'}(f)} \\ & & \downarrow \\ V' & \xleftarrow{\phi_{B'}} & K^m \end{array}$$

Dabei ist $\phi_B\colon K^n \to V$ (und analog $\phi_{B'}\colon K^m \to V'$) die zu B gehörende Linearkombinationenabbildung:

$$\phi_B(\lambda_1, \lambda_2, \ldots, \lambda_n) = \lambda_1 b_1 + \lambda_2 b_2 + \ldots + \lambda_n b_n\,.$$

Da B und B' Basen sind, sind die Linearkombinationenabbildungen ϕ_B und $\phi_{B'}$ Isomorphismen; insbesondere gibt es die Umkehrabbildung $\phi_{B'}^{-1}$.

BSP Matrix einer lin. Abb.

13.2. **Beispiel.** Wir betrachten $V = P_{<3}$, den $\mathbb{R}$-Vektorraum der Polynomfunktionen vom Grad < 3 und die lineare Abbildung $D\colon V \to V$, $f \mapsto f'$. Seien weiter $B = (x \mapsto 1, x \mapsto x, x \mapsto x^2)$ und $B' = (x \mapsto 1, x \mapsto x-1, x \mapsto (x-1)(x-2))$ zwei Basen von V. Dann ist

$$\mathrm{Mat}_{B,B}(D) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathrm{Mat}_{B',B}(D) = \begin{pmatrix} 0 & 1 & -3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathrm{Mat}_{B,B'}(D) = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{und} \quad \mathrm{Mat}_{B',B'}(D) = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}. \qquad ♣$$

Wie wir sehen, kann ein und dieselbe lineare Abbildung durch viele verschiedene Matrizen beschrieben werden. Wie hängen diese miteinander zusammen? Dazu erst eine einfache Aussage über Verknüpfungen von linearen Abbildungen.

LEMMA Matrix von $f \circ g$

13.3. **Lemma.** *Seien $g\colon V \to V'$ und $f\colon V' \to V''$ zwei K-lineare Abbildungen zwischen endlich-dimensionalen Vektorräumen. Seien weiter B eine Basis von V, B' eine Basis von V' und B'' eine Basis von V''. Dann gilt*

$$\mathrm{Mat}_{B,B''}(f \circ g) = \mathrm{Mat}_{B',B''}(f)\,\mathrm{Mat}_{B,B'}(g)\,.$$

Beweis. Das folgt aus der Definition der Matrixmultiplikation. ❑

Daraus ergibt sich sofort:

FOLG Basiswechsel

13.4. **Folgerung.** *Sei $f\colon V \to V'$ eine K-lineare Abbildung zwischen endlich-dimensionalen Vektorräumen. Seien B und $\tilde{B}$ zwei Basen von V und B' und $\tilde{B}'$ zwei Basen von V'. Dann ist*

$$\mathrm{Mat}_{\tilde{B},\tilde{B}'}(f) = \mathrm{Mat}_{B',\tilde{B}'}(\mathrm{id}_{V'})\,\mathrm{Mat}_{B,B'}(f)\,\mathrm{Mat}_{\tilde{B},B}(\mathrm{id}_V)\,.$$

Beweis. Das folgt aus Lemma 13.3 und $f = \mathrm{id}_{V'} \circ f \circ \mathrm{id}_V$. Skizze:

$$\begin{array}{ccc} (V,\tilde{B}) & \xrightarrow{f} & (V',\tilde{B}') \\ {\scriptstyle \mathrm{id}_V}\downarrow & & \uparrow{\scriptstyle \mathrm{id}_{V'}} \\ (V,B) & \xrightarrow{f} & (V',B') \end{array}$$

❑

DEF Basiswechselmatrix

Da id_V und $\mathrm{id}_{V'}$ Isomorphismen sind, sind die *Basiswechselmatrizen* $\mathrm{Mat}_{\tilde{B},B}(\mathrm{id}_V)$ und $\mathrm{Mat}_{B',\tilde{B}'}(\mathrm{id}_{V'})$ invertierbar.

Umgekehrt kann jede invertierbare Matrix als eine Basiswechselmatrix auftreten, wobei eine der beiden Basen beliebig vorgegeben werden kann.

LEMMA Basiswechselmatrizen

13.5. **Lemma.** *Sei K ein Körper, sei $n \in \mathbb{N}$, sei V ein K-Vektorraum mit Basis $B = (b_1, b_2, \ldots, b_n)$. Sei weiter $A \in \mathrm{Mat}(n, K)$ invertierbar. Dann gibt es Basen B' und B'' von V, sodass*

$$A = \mathrm{Mat}_{B,B'}(\mathrm{id}_V) = \mathrm{Mat}_{B'',B}(\mathrm{id}_V)$$

ist.

Beweis. Seien $A = (a_{ij})$ und $B'' = (b_1'', b_2'', \ldots, b_n'')$. Die Aussage $A = \mathrm{Mat}_{B'',B}(\mathrm{id}_V)$ bedeutet $b_j'' = a_{1j}b_1 + \ldots + a_{nj}b_n$. Wir definieren b_j'' durch diese Gleichung für $j \in \{1, 2, \ldots, n\}$; dann gilt die gewünschte Aussage (B'' ist eine Basis, weil A invertierbar ist: die b_i lassen sich als Linearkombinationen der b_j'' ausdrücken, deren Koeffizienten die Einträge von A^{-1} sind).

Es gibt dann auch eine Basis B', sodass $A^{-1} = \mathrm{Mat}_{B',B}(\mathrm{id}_V)$ ist; damit folgt $A = \mathrm{Mat}_{B,B'}(\mathrm{id}_V)$, denn

$$\mathrm{Mat}_{B,B'}(\mathrm{id}_V)\,\mathrm{Mat}_{B',B}(\mathrm{id}_V) = \mathrm{Mat}_{B',B'}(\mathrm{id}_V) = I_n$$

nach Lemma 13.3. ❑

13.6. **Satz.** *Seien K ein Körper und $n \in \mathbb{N}$. Die Menge der invertierbaren Matrizen in* $\mathrm{Mat}(n, K)$ *bildet eine Gruppe unter der Matrixmultiplikation.*

SATZ Gruppe der invertierbaren Matrizen

Beweis. Die Matrixmultiplikation ist assoziativ, die (invertierbare) Einheitsmatrix I_n ist neutrales Element. Jede invertierbare Matrix hat definitionsgemäß eine (selbst invertierbare) Inverse. Es bleibt zu zeigen, dass die Verknüpfung wohldefiniert ist, d.h., dass das Produkt zweier invertierbarer Matrizen wieder invertierbar ist. Das folgt aus

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I_n = (B^{-1}A^{-1})(AB)\,;$$

die Inverse von AB ist also $B^{-1}A^{-1}$. ❑

In jedem Monoid M gilt, dass die Menge der invertierbaren Elemente ein Untermonoid bildet, das eine Gruppe ist. Der Beweis ist identisch.

13.7. **Definition.** Die Gruppe der invertierbaren Matrizen in $\mathrm{Mat}(n, K)$ heißt *allgemeine lineare Gruppe* und wird mit $\mathrm{GL}(n, K)$ bezeichnet. ♢

DEF $\mathrm{GL}(n, K)$

Die Abkürzung „GL“ kommt von englisch *general linear* group. Es ist auch die Notation $\mathrm{GL}_n(K)$ gebräuchlich.

* 13.8. **Satz.** *Sei K ein Körper, seien $m, n \in \mathbb{N}$, seien V ein n-dimensionaler und V' ein m-dimensionaler K-Vektorraum und sei $f\colon V \to V'$ linear. Seien weiter B eine Basis von V, B' eine Basis von V' und $A = \mathrm{Mat}_{B,B'}(f)$. Dann gilt: Die Menge der Matrizen von f bezüglich beliebiger Basen von V und V' ist genau*

SATZ Matrizen derselben lin. Abb.

$$\{PAQ \mid P \in \mathrm{GL}(m, K), Q \in \mathrm{GL}(n, K)\}\,.$$

Beweis. Nach Folgerung 13.4 und der nachfolgenden Diskussion hat jede Matrix von f die Form PAQ mit invertierbaren Matrizen P und Q. Nach Lemma 13.5 gibt es zu beliebig vorgegebenen invertierbaren Matrizen $P \in \mathrm{GL}(m, K)$ und $Q \in \mathrm{GL}(n, K)$ Basen $\tilde{B}$ von V und $\tilde{B}'$ von V', sodass $P = \mathrm{Mat}_{B',\tilde{B}'}(\mathrm{id}_{V'})$ und $Q = \mathrm{Mat}_{\tilde{B},B}(\mathrm{id}_V)$. Dann ist

$$PAQ = \mathrm{Mat}_{B',\tilde{B}'}(\mathrm{id}_{V'})\,\mathrm{Mat}_{B,B'}(f)\,\mathrm{Mat}_{\tilde{B},B}(\mathrm{id}_V) = \mathrm{Mat}_{\tilde{B},\tilde{B}'}(f)$$

auch eine Matrix von f. ❑

* 13.9. **Definition.** Seien K ein Körper, $m, n \in \mathbb{N}$ und $A, B \in \mathrm{Mat}(m \times n, K)$. Die Matrizen A und B heißen *äquivalent*, wenn es Matrizen $P \in \mathrm{GL}(m, K)$ und $Q \in \mathrm{GL}(n, K)$ gibt mit $PAQ = B$. ♢

DEF Äquivalenz von Matrizen

Zwei Matrizen in $\mathrm{Mat}(m \times n, K)$ sind also genau dann äquivalent, wenn sie dieselbe lineare Abbildung (aber evtl. bezüglich verschiedener Basen) repräsentieren.

* 13.10. **Satz.** *Seien K ein Körper, $m, n \in \mathbb{N}$ und $A, B \in \mathrm{Mat}(m \times n, K)$. Die Matrizen A und B sind genau dann äquivalent, wenn $\mathrm{rk}(A) = \mathrm{rk}(B)$ ist. In diesem Fall sei $r = \mathrm{rk}(A)$; dann sind beide Matrizen äquivalent zur Matrix*

SATZ Klassifikation von Matrizen bis auf Äquivalenz

$$M_r = \left(\begin{array}{c|c} I_r & \mathbf{0}_{r,n-r} \\ \hline \mathbf{0}_{m-r,r} & \mathbf{0}_{m-r,n-r} \end{array}\right).$$

Dabei steht $\mathbf{0}_{k,l}$ für eine Nullmatrix mit k Zeilen und l Spalten.

Beweis. Sei $r = \mathrm{rk}(A)$. Wir zeigen, dass A zu M_r äquivalent ist. Sei $f\colon K^n \to K^m$ die zugehörige lineare Abbildung; sie hat Rang r, also ist $\dim \ker(f) = n - r$. Wir wählen eine Basis $B = (b_1, \ldots, b_n)$ von K^n, sodass $(b_{r+1}, \ldots, b_n)$ eine Basis von $\ker(f)$ ist. Mit $b'_i = f(b_i)$ für $i \in \{1, 2, \ldots, r\}$ ist dann $(b'_1, \ldots, b'_r)$ eine Basis von $\mathrm{im}(f)$: Wir haben $\dim \mathrm{im}(f) = r$ und

$$\mathrm{im}(f) = \langle f(b_1), \ldots, f(b_r), f(b_{r+1}), \ldots, f(b_n)\rangle = \langle b'_1, \ldots, b'_r, \mathbf{0}, \ldots, \mathbf{0}\rangle = \langle b'_1, \ldots, b'_r\rangle;$$

$(b'_1, \ldots, b'_r)$ ist also ein Erzeugendensystem von $\mathrm{im}(f)$ der richtigen Länge r. Wir ergänzen diese Basis von $\mathrm{im}(f)$ zu einer Basis $B' = (b'_1, \ldots, b'_m)$ von K^m. Dann ist $\mathrm{Mat}_{B,B'}(f)$ gerade M_r; M_r ist damit äquivalent zu A.

Gilt auch $\mathrm{rk}(B) = r$, dann ist B ebenfalls äquivalent zu M_r. Es folgt, dass A und B äquivalent sind: Es gibt $P, P' \in \mathrm{GL}(m, K)$ und $Q, Q' \in \mathrm{GL}(n, K)$ mit $M_r = PAQ = P'BQ'$. Dann ist $B = (P'^{-1}P)A(QQ'^{-1})$.

Umgekehrt gilt $\mathrm{rk}(B) = r = \mathrm{rk}(A)$ für jede zu A äquivalente Matrix B, denn der Rang einer Matrix ist gleich dem Rang jeder von ihr repräsentierten linearen Abbildung. ❑

Mit den Resultaten aus dem Kleingedruckten von Seite 91 ergibt sich aus Satz 13.10:

Folgerung. *Jede Matrix $A \in \mathrm{Mat}(m \times n, K)$ lässt sich durch Zeilen- und Spaltenumformungen in die Matrix M_r mit $r = \mathrm{rk}(A)$ überführen.*

FOLG Zeilen- und Spaltenumf.

Die Äquivalenz von Matrizen ist ein Beispiel einer *Äquivalenzrelation*. Eine *Relation* R zwischen Mengen X und Y ist formal eine Teilmenge $R \subset X \times Y$. Ist für $x \in X$ und $y \in Y$ das Paar (x, y) ein Element von R, dann sagt man, x und y *stehen in der Relation R zueinander* und schreibt auch $x\,R\,y$ oder ähnlich. Im Fall $X = Y$ spricht man auch von einer *Relation auf X*. Eine solche Relation heißt

- *reflexiv*, wenn $\forall x \in X\colon x\,R\,x$,
- *symmetrisch*, wenn $\forall x, y \in X\colon x\,R\,y \Rightarrow y\,R\,x$, und
- *transitiv*, wenn $\forall x, y, z \in X\colon (x\,R\,y \ \wedge\ y\,R\,z) \Rightarrow x\,R\,z$.

Eine Relation auf X, die reflexiv, symmetrisch und transitiv ist, ist eine *Äquivalenzrelation auf X*. Beispiele sind die Gleichheitsrelation $x = y$ (das ist die „feinste“ Äquivalenzrelation auf X) oder auch die „Allrelation“ $R = X \times X$ (die „gröbste“ Äquivalenzrelation auf X).

Lemma. *Für Matrizen $A, B \in \mathrm{Mat}(m \times n, K)$ schreiben wir $A \sim B$, wenn A und B äquivalent sind, wenn es also $P \in \mathrm{GL}(m, K)$ und $Q \in \mathrm{GL}(n, K)$ gibt mit $B = PAQ$.*

Die Relation $\sim$ ist eine Äquivalenzrelation auf $\mathrm{Mat}(m \times n, K)$.

LEMMA Äquivalenz von Matrizen ist Äqu.rel.

Beweis. Wir müssen die drei Eigenschaften nachprüfen.

- Reflexivität: $A \sim A$, denn man kann $P = I_m$, $Q = I_n$ wählen.

- Symmetrie: Es gelte $A \sim B$; dann gibt es $P \in \mathrm{GL}(m,K)$ und $Q \in \mathrm{GL}(n,K)$ mit $B = PAQ$. Dann sind auch $P^{-1} \in \mathrm{GL}(m,K)$ und $Q^{-1} \in \mathrm{GL}(n,K)$ und es gilt $B = P^{-1}AQ^{-1}$, also $B \sim A$.
- Transitivität: Es gelte $A \sim B$ und $B \sim C$. Dann gibt es $P_1, P_2 \in \mathrm{GL}(m,K)$ und $Q_1, Q_2 \in \mathrm{GL}(n,K)$ mit $B = P_1AQ_1$ und $C = P_2BQ_2$. Es sind $P_2P_1 \in \mathrm{GL}(m,K)$ und $Q_1Q_2 \in \mathrm{GL}(n,K)$ und es gilt $C = (P_2P_1)A(Q_1Q_2)$, also ist $A \sim C$. ❑

Die wichtigste Eigenschaft einer Äquivalenzrelation auf einer Menge X ist, dass sie zu einer Einteilung von X in sogenannte Äquivalenzklassen führt. Ist $\sim$ eine Äquivalenzrelation auf X und $x \in X$, dann schreiben wir $[x]$ für die Menge $\{y \in X \mid x \sim y\}$ der zu x äquivalenten Elemente von X und nennen $[x]$ die *Äquivalenzklasse von x*. Jedes Element von $[x]$ heißt ein *Repräsentant* der Äquivalenzklasse.

Lemma. *Sei $\sim$ eine Äquivalenzrelation auf einer Menge X und sei $x \in X$. Dann sind für $y \in X$ die folgenden Aussagen äquivalent:* **LEMMA** Eigensch. Äqu.rel.

(1) $x \sim y$.

(2) $y \in [x]$.

(3) $[y] \cap [x] \neq \emptyset$.

(4) $[y] = [x]$.

Insbesondere sind zwei Äquivalenzklassen $[x]$ und $[y]$ entweder gleich oder disjunkt.

Beweis. Die Äquivalenz von (1) und (2) folgt aus der Definition von $[x]$.

„(2) $\Rightarrow$ (3)“: Wegen der Reflexivität von $\sim$ ist $y \in [y]$, also folgt aus $y \in [x]$, dass $y \in [y] \cap [x]$.

„(3) $\Rightarrow$ (4)“: Sei $z \in [y] \cap [x]$ und $w \in [y]$. Dann gilt $y \sim w$, $y \sim z$ und $x \sim z$; mit Symmetrie und Transitivität von $\sim$ folgt daraus $x \sim w$, also $w \in [x]$. Da w beliebig war, gilt $[y] \subset [x]$. Genauso erhalten wir $[x] \subset [y]$.

„(4) $\Rightarrow$ (2)“: Aus $y \in [y]$ und $[y] = [x]$ folgt $y \in [x]$. ❑

Wir können die *Menge der Äquivalenzklassen* $X/\sim = \{[x] \mid x \in X\}$ bilden. Dann gibt es eine natürliche (oder „kanonische“) surjektive Abbildung $f\colon X \to X/\sim$, $x \mapsto [x]$. Die Urbildmenge $f^{-1}(\{[x]\})$ ist nach dem gerade bewiesenen Lemma genau $[x]$. Umgekehrt führt jede surjektive Abbildung $f\colon X \to M$ zu einer Äquivalenzrelation auf X (man sagt auch, f *induziert* eine Äquivalenzrelation) durch $x \sim y \iff f(x) = f(y)$.

Die Aussage von Satz 13.10 bedeutet dann, dass die Äquivalenz von $m \times n$-Matrizen mit der durch $\mathrm{Mat}(m \times n, K) \to \{1, 2, \ldots, \min\{m,n\}\}$, $A \mapsto \mathrm{rk}(A)$, induzierten Äquivalenzrelation übereinstimmt und dass M_r ein Repräsentant der durch $\mathrm{rk}(A) = r$ gegebenen Äquivalenzklasse ist.

Wenn man nur Zeilenumformungen betrachtet, dann wird man zwei Matrizen A und B als äquivalent betrachten, wenn $B = PA$ ist mit einer invertierbaren Matrix P, denn Zeilenumformungen entsprechen der Multiplikation von links mit einer invertierbaren Matrix; vergleiche die Diskussion im Kleingedruckten auf Seite 91. Die dortigen Ergebnisse lassen sich so interpretieren, dass jede Äquivalenzklasse dieser Äquivalenzrelation einen eindeutig bestimmten Repräsentanten in reduzierter Zeilenstufenform hat.

14. Die Determinante

In diesem Abschnitt führen wir die *Determinante* einer quadratischen Matrix ein. Das ist ein Skalar, der darüber Auskunft gibt, ob die Matrix invertierbar ist oder nicht. Wir definieren die Determinante rekursiv.

∗ 14.1. **Definition.** Seien K ein Körper und $A = (a_{ij}) \in \mathrm{Mat}(n, K)$ mit $n \in \mathbb{N}$. Wir definieren die *Determinante* von A, $\det(A)$, rekursiv wie folgt: **DEF** Determinante einer Matrix

(1) Im Fall $n = 0$ ist $\det(A) = 1$.

(2) Im Fall $n > 0$ sei $A_{ij} \in \mathrm{Mat}(n-1, K)$ (für $i, j \in \{1, 2, \ldots, n\}$) die Matrix, die aus A entsteht, wenn man die i-te Zeile und die j-te Spalte entfernt. Wir definieren

$$\begin{aligned}\det(A) &= \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \det(A_{1j}) \\ &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \ldots + (-1)^{n-1} a_{1n}\det(A_{1n})\,.\end{aligned}$$

Für die Determinante ist auch folgende Schreibweise üblich:

$$\det\big((a_{ij})\big) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \qquad \diamondsuit$$

14.2. **Beispiele.** Für kleine positive Werte von n erhalten wir folgende Formeln: **BSP** Determinante

$$\begin{aligned} \det\big((a)\big) &= a \\ \begin{vmatrix} a & b \\ c & d \end{vmatrix} &= ad - bc \\ \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= aei - afh + bfg - bdi + cdh - ceg \end{aligned}$$

Die Formel für die 3×3-Determinante lässt sich mit Hilfe der „Sarrus-Regel“ merken: Man schreibt die ersten beiden Spalten noch einmal hinter die Matrix und bildet die Summe der Produkte über die nach rechts fallenden Diagonalen minus die Summe der Produkte über die nach rechts steigenden Diagonalen.

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \begin{matrix} a & b \\ d & e \\ g & h \end{matrix}$$

Für größere Determinanten gibt es allerdings keine solche Merkregel! ♣

Welche Eigenschaften hat die Determinante?

SATZ Eigensch. der Determinante

14.3. **Satz.** *Seien K ein Körper und $n \in \mathbb{N}$. Für $d = \det\colon \mathrm{Mat}(n, K) \to K$ gilt:*

(1) *$d(A)$ ist linear als Funktion jeder Zeile von A (dabei werden die Einträge der übrigen Zeilen als fest angesehen).*

(2) *Hat A zwei gleiche Zeilen, dann ist $d(A) = 0$.*

(3) *$d(I_n) = 1$.*

Ist $d\colon \mathrm{Mat}(n, K) \to K$ eine Abbildung, die (1) und (2) erfüllt, dann gilt:

(4) *Geht A' aus A durch eine elementare Zeilenumformung $\mathbf{I}_i(\lambda)$ hervor, dann gilt $d(A') = \lambda d(A)$.*

(5) *Geht A' aus A durch eine elementare Zeilenumformung $\mathbf{II}_{i,j}(\lambda)$ hervor, dann gilt $d(A') = d(A)$.*

(6) *Geht A' aus A durch Vertauschen zweier Zeilen hervor, dann gilt $d(A') = -d(A)$.*

(7) *Ist $d\colon \mathrm{Mat}(n, K) \to K$ eine Abbildung, die die Eigenschaften (1) und (2) hat, dann gilt $d(A) = \det(A) d(I_n)$ für alle $A \in \mathrm{Mat}(n, K)$. Insbesondere ist $\det\colon \mathrm{Mat}(n, K) \to K$ die einzige Abbildung, die (1), (2) und (3) erfüllt.*

Außerdem gilt

(8) $\det(A) \neq 0 \iff \mathrm{rk}(A) = n \iff$ *A invertierbar.*

Beweis. Wir beweisen zuerst (4)–(7).

(4) Das ist ein Spezialfall von Eigenschaft (1).

(5) Wir schreiben $d(v_1, \ldots, v_n)$ für $d(A)$, wenn A die Zeilen $v_1, \ldots, v_n \in K^n$ hat; für $1 \leq i < j \leq n$ sei $d_{ij}(v_i, v_j) = d(v_1, \ldots, v_n)$, wobei die v_k mit $k \notin \{i, j\}$ fest gewählt sind. Dann gilt

$$d(A') = d_{ij}(v_i + \lambda v_j, v_j) \overset{(1)}{=} d_{ij}(v_i, v_j) + \lambda d_{ij}(v_j, v_j) \overset{(2)}{=} d_{ij}(v_i, v_j) = d(A)\,.$$

(6) In der Notation des Beweises von Teil (5) haben wir

$$\begin{aligned} 0 &\overset{(2)}{=} d_{ij}(v_i + v_j, v_i + v_j) \\ &\overset{(1)}{=} d_{ij}(v_i, v_i) + d_{ij}(v_i, v_j) + d_{ij}(v_j, v_i) + d_{ij}(v_j, v_j) \\ &\overset{(2)}{=} d_{ij}(v_i, v_j) + d_{ij}(v_j, v_i)\,, \end{aligned}$$

also ist $d(A') = d_{ij}(v_j, v_i) = -d_{ij}(v_i, v_j) = -d(A)$.

(7) Aus (1) und (2) folgen (4), (5) und (6). Sei A' die reduzierte Zeilenstufenform von A. Es folgt, dass $d(A) = d_0(A) d(A')$ ist, wobei $d_0(A) \neq 0$ nur von A und nicht von d abhängt: $d_0(A)$ ist das Produkt der Skalare λ^{-1} und -1, die aus den elementaren Zeilenumformungen $\mathbf{I}_i(\lambda)$ und $\mathbf{III}_{i,j}$ herrühren, die man ausführt, um von A zu A' zu gelangen. Wenn A invertierbar ist, dann ist $A' = I_n$ (Lemma 12.18), also $d(A) = d_0(A) d(I_n)$. Für $d = \det$ ergibt sich $d_0(A) = \det(A)$. Ist A nicht invertierbar, dann hat A' eine Null-Zeile, womit (aus Eigenschaft (1)) $d(A') = 0$ folgt. Für $d = \det$ hat man $\det(A') = 0$. In beiden Fällen erhalten wir $d(A) = \det(A) d(I_n)$ wie behauptet. ❑

Der Beweis der ersten drei Aussagen erfolgt durch Induktion über n. Im Fall $n = 0$ sind die Aussagen trivialerweise richtig. Sei also $n > 0$ und die Aussagen seien für kleinere Werte von n gezeigt.

(1) $\det(A)$ ist linear in der ersten Zeile von A, denn nach Definition ist $\det(A)$ eine Linearkombination der Einträge der ersten Zeile, deren Koeffizienten nicht von der ersten Zeile abhängen. Sei $k \in \{1, 2, \ldots, n-1\}$. Nach Induktionsannahme sind alle $\det(A_{1j})$ linear in der k-ten Zeile von A_{1j} und damit linear in der $(k+1)$-ten Zeile von A. $\det(A)$ ist somit eine Linearkombination von Abbildungen, die linear als Funktion der $(k+1)$-ten Zeile von A sind (mit Koeffizienten, die nicht von der $(k+1)$-ten Zeile abhängen) und somit ebenfalls linear in der $(k+1)$-ten Zeile von A.

(2) Sei A eine Matrix, in der die k-te und die l-te Zeile übereinstimmen, wobei $1 \leq k < l \leq n$. Ist $k > 1$, dann stimmt in jeder Matrix A_{1j} die $(k-1)$-te mit der $(l-1)$-ten Zeile überein; nach Induktionsannahme gilt $\det(A_{1j}) = 0$ für alle j, also auch $\det(A) = 0$. Es bleibt der Fall $k = 1$ zu betrachten. Falls $l > 2$ ist, dann vertauschen wir die l-te mit der zweiten Zeile. Nach Induktionsannahme und der damit für $d = \det$ auf Matrizen der Größe $n-1$ geltenden Aussage (6) bewirkt das einen Vorzeichenwechsel in allen $\det(A_{1j})$, ändert also nichts daran, ob $\det(A) = 0$ ist oder nicht. Wir können daher annehmen, dass die beiden ersten Zeilen von A gleich sind. Wir schreiben $d_{jm} = d_{mj}$ für die Determinante der Matrix, die aus A durch Streichen der ersten beiden Zeilen und der Spalten j und m entsteht. Dann gilt (unter Beachtung von $a_{2j} = a_{1j}$)

$$\begin{aligned}
\det(A) &= \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \det(A_{1j}) \\
&= \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \left(\sum_{m=1}^{j-1} (-1)^{m-1} a_{2m} d_{jm} + \sum_{m=j+1}^{n} (-1)^{m} a_{2m} d_{jm} \right) \\
&= \sum_{1 \leq m < j \leq n} (-1)^{j-m} a_{1j} a_{1m} d_{jm} + \sum_{1 \leq j < m \leq n} (-1)^{m-j-1} a_{1j} a_{1m} d_{jm} \\
&= \sum_{1 \leq j < m \leq n} (-1)^{m-j} a_{1j} a_{1m} d_{jm} + \sum_{1 \leq j < m \leq n} (-1)^{m-j-1} a_{1j} a_{1m} d_{jm} \\
&= \sum_{1 \leq j < m \leq n} \big((-1)^{m-j} + (-1)^{m-j-1} \big) a_{1j} a_{1m} d_{jm} \\
&= 0 \,.
\end{aligned}$$

(Wir haben in der ersten Summe j und m vertauscht und dabei ausgenutzt, dass $a_{1m} a_{1j} d_{mj} = a_{1j} a_{1m} d_{jm}$ ist.)

(3) Nach der rekursiven Definition ist $\det(I_n) = 1 \cdot \det(I_{n-1}) = 1$.

Es verbleibt:

(8) Aus den Teilen (4), (5) und (6) folgt, dass $\det(A)$ genau dann null ist, wenn $\det(A')$ null ist, wobei A' die reduzierte Zeilenstufenform von A ist. Gilt $\operatorname{rk}(A) = n$, dann ist A invertierbar, und nach Lemma 12.18 ist $A' = I_n$ und damit $\det(A') = \det(I_n) = 1 \neq 0$ nach Teil (3). Gilt $\operatorname{rk}(A) < n$, dann hat A' eine Null-Zeile und damit ist $\det(A') = 0$ nach Teil (1). Die zweite Äquivalenz folgt daraus, dass eine lineare Abbildung $K^n \to K^n$ genau dann surjektiv ist, wenn sie ein Isomorphismus ist, vgl. Folgerung 10.14.

Wenn wir die Determinante einer $n \times n$-Matrix A als Funktion der n Zeilen von A betrachten, die selbst Vektoren in K^n sind, dann erhalten wir eine sogenannte *alternierende Multilinearform*. Das ist ein Spezialfall einer multilinearen Abbildung.

DEF multilineare Abbildung, alternierende Multilinearform

Definition. Sei K ein Körper und seien $V_1, V_2, \ldots, V_m$ und W K-Vektorräume. Eine Abbildung $f\colon V_1 \times V_2 \times \ldots \times V_m \to W$ heißt *multilinear*, wenn f in jedem Argument K-linear ist, wenn also gilt

$$f(v_1, \ldots, v_{i-1}, \lambda v_i, v_{i+1}, \ldots, v_m) = \lambda f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_m)$$

und

$$\begin{aligned} &f(v_1, \ldots, v_{i-1}, v_i + v_i', v_{i+1}, \ldots, v_m) \\ &\qquad = f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_m) + f(v_1, \ldots, v_{i-1}, v_i', v_{i+1}, \ldots, v_m) \end{aligned}$$

für alle $i \in \{1, 2, \ldots, m\}$, $v_j \in V_j$, $\lambda \in K$, $v_i' \in V_i$. Ist $W = K$, dann heißt f auch eine *Multilinearform*.

Eine Multilinearform $f\colon V^m = V \times V \times \cdots \times V \to K$ heißt *alternierend*, wenn stets $f(v_1, \ldots, v_m) = 0$ ist, sobald $v_i = v_j$ ist für gewisse $1 \le i < j \le m$. ♢

Aussagen (1) und (2) in Satz 14.3 besagen gerade, dass $\det(A)$ eine alternierende Multilinearform der Zeilen von A ist. Da man einen K-Vektorraum V mit Basis $(b_1, \ldots, b_n)$ mit K^n identifizieren kann, hat Aussage (7) in Satz 14.3 die folgende Interpretation:

SATZ Existenz u. Eindeutigkeit alternierender Multilinearformen

Satz. *Sei V ein K-Vektorraum mit Basis $(b_1, b_2, \ldots, b_n)$. Dann gibt es genau eine alternierende Multilinearform $d\colon V^n \to K$ mit $d(b_1, b_2, \ldots, b_n) = 1$.*

Für praktische Zwecke wichtig sind die Aussagen in Satz 14.3, die zeigen, wie sich die Determinante unter elementaren Zeilenumformungen verhält. Das liefert ein praktisches Verfahren zur Berechnung auch größerer Determinanten.

BSP Determinantenberechnung

14.4. **Beispiel.**

$$\begin{vmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 2 & 0 & 2 \\ 0 & 3 & 3 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 6 & 6 \end{vmatrix}$$

$$= 2 \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 6 \end{vmatrix} = 2 \cdot 6 \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 12 \qquad ♣$$

Die Eindeutigkeitsaussage (7) in Satz 14.3 ist wichtig, weil sie weitere Eigenschaften der Determinante zur Folge hat.

* **SATZ** Entwicklung der Det. nach der i-ten Zeile

14.5. **Satz.** *Sei K ein Körper, sei $n > 0$ und $A = (a_{ij}) \in \mathrm{Mat}(n, K)$. Mit der in Definition 14.1 eingeführten Schreibweise A_{ij} gilt für jedes $i \in \{1, 2, \ldots, n\}$:*

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

Beweis. Wie im Beweis von Satz 14.3 zeigt man, dass die rechte Seite die Eigenschaften (1), (2) und (3) hat. Wegen der Eindeutigkeit folgt, dass die rechte Seite gleich $\det(A)$ sein muss. ❑

14.6. **Beispiel.** Die Berechnung der Determinante in Beispiel 14.4 lässt sich vereinfachen, indem man nach der zweiten Zeile entwickelt: **BSP**

$$\begin{vmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{vmatrix} = - \begin{vmatrix} -1 & 1 & -1 \\ 1 & 1 & 1 \\ 2 & 4 & 8 \end{vmatrix} = \ldots = 12 \qquad \clubsuit$$

* 14.7. **Satz.** *Sei K ein Körper, sei $n \in \mathbb{N}$ und seien $A, B \in \mathrm{Mat}(n, K)$. Dann gilt* **SATZ** Multiplikativität der Det.

$$\det(AB) = \det(A) \det(B) .$$

Ist A invertierbar, dann ist $\det(A^{-1}) = \det(A)^{-1}$.

Beweis. Wir fixieren B und betrachten A als variabel. Sei $d_B \colon \mathrm{Mat}(n, K) \to K$, $A \mapsto \det(AB)$. Aus den Eigenschaften der Matrixmultiplikation folgt, dass die k-te Zeile von AB nur von der k-ten Zeile von A abhängt und zwar linear. Es folgt, dass d_B linear in den Zeilen von A ist. Ebenso gilt, dass aus der Gleichheit der k-ten und der l-ten Zeile von A die entsprechende Aussage für AB folgt. Damit erfüllt d_B auch die Eigenschaft (2) in Satz 14.3. Die Eindeutigkeitsaussage in Satz 14.3 liefert nun $\det(AB) = d_B(A) = \det(A) d_B(I_n) = \det(A) \det(B)$. Die letzte Aussage ergibt sich aus $\det(A) \det(A^{-1}) = \det(I_n) = 1$. ❑

* 14.8. **Satz.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Dann gilt* **SATZ** Symmetrie der Det.

$$\det(A^\top) = \det(A) .$$

Beweis. Wir müssen zeigen, dass $\det(A^\top)$ die Eigenschaften (1), (2) und (3) aus Satz 14.3 hat. $\det(I_n^\top) = \det(I_n) = 1$ ist klar. Die beiden anderen Aussagen sind dazu äquivalent, dass $\det(A)$ linear in den *Spalten* von A ist und verschwindet, wenn A zwei gleiche *Spalten* hat. Die erste Aussage folgt leicht mit Induktion aus der rekursiven Definition der Determinante, denn für festes k ist jeder Term in der Summe linear in der k-ten Spalte von A (entweder durch a_{1k} oder durch $\det(A_{1j})$). Die zweite Aussage kann wie folgt gezeigt werden: Wenn A zwei gleiche Spalten hat, dann ist $\mathrm{rk}(A) < n$, also $\det(A) = 0$ nach Satz 14.3, Teil (8). ❑

Daraus folgt zum Beispiel, dass man auch *Spalten*umformungen bei der Berechnung der Determinante verwenden kann, auch mit Zeilenumformungen gemischt. Ebenso ergibt sich eine Formel zur Entwicklung der Determinante nach einer Spalte.

* 14.9. **Folgerung.** *Sei K ein Körper, sei $n > 0$ und $A = (a_{ij}) \in \mathrm{Mat}(n, K)$. Mit der in Definition 14.1 eingeführten Schreibweise A_{ij} gilt für jedes $j \in \{1, 2, \ldots, n\}$:* **FOLG** Entwicklung der Det. nach der j-ten Spalte

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

Beweis. Das folgt aus Satz 14.5, angewandt auf $A^\top$ und aus $\det(A^\top) = \det(A)$. ❑

14.10. **Beispiel.** Eine Matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ mit $AA^\top = I_n$ heißt *orthogonal*. Was kann man über $\det(A)$ sagen? **BSP**

Es gilt

$$1 = \det(I_n) = \det(AA^\top) = \det(A)\det(A^\top) = \det(A)^2\,,$$

also ist $\det(A) = \pm 1$. ♣

14.11. **Beispiel.** Wir berechnen die Determinante aus Beispiel 14.4 noch einmal. **BSP** Determinantenberechnung

$$\begin{vmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{vmatrix} = -\begin{vmatrix} -1 & 1 & -1 \\ 1 & 1 & 1 \\ 2 & 4 & 8 \end{vmatrix} = -\begin{vmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 2 & 4 & 6 \end{vmatrix} = -6 \cdot \begin{vmatrix} -1 & 1 \\ 1 & 1 \end{vmatrix}$$

$$= -6((-1)\cdot 1 - 1 \cdot 1) = 12$$

(Entwicklung nach der zweiten Zeile, elementare Spaltenumformung $\mathbf{II}_{3,1}(-1)$, Entwicklung nach der dritten Spalte, Formel für 2×2-Determinante). ♣

Der Vollständigkeit halber halten wir noch fest, wie sich die Determinante bei Skalierung der ganzen Matrix ändert.

14.12. **Folgerung.** *Seien K ein Körper, $n \in \mathbb{N}$, $\lambda \in K$ und $A \in \mathrm{Mat}(n, K)$. Dann gilt* **FOLG** $\det(\lambda A)$

$$\det(\lambda A) = \lambda^n \det(A)\,.$$

Beweis. Das folgt daraus, dass $\det(A)$ in *jeder* Spalte von A linear ist:
Mit $A = (\boldsymbol{a}_1 \mid \boldsymbol{a}_2 \mid \cdots \mid \boldsymbol{a}_n)$ ergibt sich

$$\begin{aligned} \det(\lambda A) &= \det(\lambda \boldsymbol{a}_1 \mid \lambda \boldsymbol{a}_2 \mid \cdots \mid \lambda \boldsymbol{a}_n) \\ &= \lambda \det(\boldsymbol{a}_1 \mid \lambda \boldsymbol{a}_2 \mid \cdots \mid \lambda \boldsymbol{a}_n) = \ldots = \lambda^n \det(\boldsymbol{a}_1 \mid \boldsymbol{a}_2 \mid \cdots \mid \boldsymbol{a}_n)\,. \end{aligned}$$ ❑

Die Entwicklung der Determinante nach Zeilen und Spalten führt zu folgender „Formel“ für die Inverse einer Matrix.

∗ 14.13. **Definition.** Seien K ein Körper, $n > 0$ und $A \in \mathrm{Mat}(n, K)$. Die Matrix $\tilde{A} \in \mathrm{Mat}(n, K)$, deren Eintrag in der i-ten Zeile und j-ten Spalte durch $(-1)^{i-j}\det(A_{ji})$ (nicht A_{ij}!) gegeben ist, heißt die *adjungierte* (oder auch *adjunkte*) *Matrix* zu A. ♢ **DEF** Adjungierte Matrix

∗ 14.14. **Satz.** *Seien K ein Körper, $n > 0$ und $A \in \mathrm{Mat}(n, K)$. Dann gilt* **SATZ** Adjungierte Matrix

$$A\tilde{A} = \tilde{A}A = \det(A) I_n\,.$$

Ist A invertierbar, dann ist $A^{-1} = \det(A)^{-1}\tilde{A}$.

Beweis. Der Eintrag an der Stelle (i, k) im Produkt $A\tilde{A}$ ist

$$\sum_{j=1}^{n} a_{ij}(-1)^{j-k}\det(A_{kj})\,.$$

Im Fall $k = i$ ergibt das $\det(A)$ nach dem Satz 14.5 über die Entwicklung der Determinante nach der i-ten Zeile. Im Fall $k \neq i$ ergibt sich analog die Entwicklung nach der k-ten Zeile der Determinante der Matrix, die aus A entsteht, wenn man die k-te Zeile durch die i-te ersetzt. Da diese Matrix zwei gleiche Zeilen hat, ist

ihre Determinante null. Das zeigt $A\tilde{A} = \det(A)I_n$. Die Aussage $\tilde{A}A = \det(A)I_n$ sieht man analog unter Verwendung von Folgerung 14.9. Die letzte Aussage folgt durch Multiplikation mit $\det(A)^{-1}A^{-1}$. ❑

14.15. **Beispiel.** Für $n = 2$ bekommen wir die Formel

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} .$$

♣

BSP Inverse einer 2×2-Matrix

Hier ist ein Beispiel für eine allgemeine Formel für eine spezielle Determinante:

14.16. **Satz.** *Sei K ein Körper, sei $n \in \mathbb{N}$, seien $a_1, a_2, \ldots, a_n \in K$ und sei A folgende „Vandermonde-Matrix“ zu $a_1, a_2, \ldots, a_n$:*

SATZ Vandermonde-Determinante

$$A = (a_i^{j-1})_{1 \le i,j \le n} = \begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{pmatrix} \in \mathrm{Mat}(n, K) .$$

Dann ist

$$\det(A) = \prod_{1 \le i < j \le n} (a_j - a_i) .$$

Beweis. Durch Induktion nach n. Der Fall $n = 0$ ist klar ($\det(A) = 1$ und das Produkt ist leer). Sei also $n > 0$ und die Behauptung für $n - 1$ bewiesen. Wir subtrahieren die erste Zeile von den übrigen und entwickeln nach der ersten Spalte; das liefert

$$\det(A) = \begin{vmatrix} a_2 - a_1 & a_2^2 - a_1^2 & \cdots & a_2^{n-1} - a_1^{n-1} \\ \vdots & \vdots & & \vdots \\ a_n - a_1 & a_n^2 - a_1^2 & \cdots & a_n^{n-1} - a_1^{n-1} \end{vmatrix} .$$

Nun ist $x^m - y^m = (x - y)(x^{m-1} + x^{m-2}y + \ldots + xy^{m-2} + y^{m-1})$. Wir können aus der ersten, …, $(n-1)$-ten Zeile also jeweils einen Faktor $(a_2 - a_1)$, …, $(a_n - a_1)$ herausziehen:

$$\det(A) = \prod_{j=2}^{n} (a_j - a_1) \cdot \begin{vmatrix} 1 & a_2 + a_1 & \cdots & a_2^{n-2} + a_1 a_2^{n-3} + \ldots + a_1^{n-2} \\ \vdots & \vdots & & \vdots \\ 1 & a_n + a_1 & \cdots & a_n^{n-2} + a_1 a_n^{n-3} + \ldots + a_1^{n-2} \end{vmatrix} .$$

Durch Subtraktion von a_1-mal der vorletzten Spalte von der letzten, dann Subtraktion von a_1-mal der drittletzten Spalte von der vorletzten, …, a_1-mal der ersten Spalte von der zweiten erhält man aus der verbliebenen Matrix die Vandermonde-Matrix zu $a_2, a_3, \ldots, a_n$. Die Behauptung folgt dann aus der Induktionsvoraussetzung. ❑

Die Vandermonde-Matrix ist genau die Matrix der linearen Abbildung

$$\phi \colon P_{<n} \longrightarrow \mathbb{R}^n, \quad f \longmapsto \big(f(a_1), f(a_2), \ldots, f(a_n)\big)$$

(wo $P_{<n}$ der $\mathbb{R}$-Vektorraum der Polynomfunktionen vom Grad $< n$ ist), die bei der Interpolation von gegebenen Werten durch Polynome eine Rolle spielt, bezüglich der Basis $(x \mapsto 1, x \mapsto x, x \mapsto x^2, \ldots, x \mapsto x^{n-1})$ von $P_{<n}$ und der Standardbasis von $\mathbb{R}^n$, siehe Beispiele 9.18 und 10.15.

Nach Vandermonde ist auch die „Vandermonde-Identität“

$$\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} = \binom{m+n}{k}$$

benannt.

Wir werden uns jetzt mit einer Verallgemeinerung der Formeln für die Determinante wie in Beispiel 14.2 beschäftigen. Diese Formeln erhält man aus der rekursiven Definition der Determinante wie in Definition 14.1. Das Resultat ist eine Summe von Termen der Form $\pm a_{1,\sigma(1)}a_{2,\sigma(2)}\cdots a_{n,\sigma(n)}$, wobei die Spaltenindizes $\sigma(1),\sigma(2),\ldots,\sigma(n)$ paarweise verschieden sind (denn jede „verbrauchte“ Spalte wird in der weiteren Entwicklung entfernt). Die Abbildung $\sigma\colon \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ ist demnach bijektiv, also eine Permutation. Wie man sich leicht überlegt, kommt auch jede Permutation in der Entwicklung der Determinante vor. Wir erinnern uns daran, dass die Permutationen von $\{1,2,\ldots,n\}$ die Elemente der *symmetrischen Gruppe* S_n sind; die Verknüpfung in dieser Gruppe ist die Komposition von Abbildungen. Damit haben wir Folgendes gezeigt:

G.W. Leibniz (1646–1716)

* 14.17. **Satz.** *Seien K ein Körper und $n \in \mathbb{N}$. Dann gibt es eine Abbildung $\varepsilon\colon S_n \to \{-1,1\}$, sodass für alle $A = (a_{ij}) \in \mathrm{Mat}(n,K)$ gilt*

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

SATZ Leibniz-Formel

Diese Formel hat $\#S_n = n! = 1\cdot 2\cdot 3\cdots(n-1)\cdot n$ Terme und taugt damit außer für sehr kleine Werte von n nicht zur praktischen Berechnung der Determinante! Sie ist aber nützlich für theoretische Überlegungen. Zum Beispiel folgt sofort, dass die Determinante einer Matrix mit ganzzahligen Einträgen eine ganze Zahl ist.

14.18. **Definition.** $\varepsilon(\sigma) \in \{-1,1\}$ heißt das *Signum* oder *Vorzeichen* der Permutation $\sigma \in S_n$. σ heißt *gerade*, wenn $\varepsilon(\sigma) = 1$ und *ungerade*, wenn $\varepsilon(\sigma) = -1$ ist. ♢

DEF Signum einer Permutation

Um etwas über diese Vorzeichenfunktion herauszufinden, führen wir die zu σ gehörende Permutationsmatrix ein.

14.19. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$ und $\sigma \in S_n$. Dann bezeichnen wir mit $P(\sigma)$ die Matrix $(\delta_{i,\sigma(j)})_{1\le i,j\le n} \in \mathrm{Mat}(n,K)$ und nennen $P(\sigma)$ die *zu σ gehörende Permutationsmatrix*. ♢

DEF Permutationsmatrix

Die Einträge von $P(\sigma)$ sind 1 an Positionen der Form $(\sigma(j),j)$ und sonst 0. Das bedeutet, dass in der j-ten Spalte von $P(\sigma)$ gerade der Standard-Basisvektor $\mathbf{e}_{\sigma(j)}$ steht; die zu P gehörende lineare Abbildung $K^n \to K^n$ ist also genau die, die die Standardbasis entsprechend der Permutation σ vertauscht: $P(\sigma)\mathbf{e}_j = \mathbf{e}_{\sigma(j)}$, wenn wir $\mathbf{e}_j$ als Spaltenvektor auffassen.

14.20. **Lemma.** *Seien K ein Körper und $n \in \mathbb{N}$.*

LEMMA Eigensch. Permutationsmatrix

(1) *Für $\sigma,\tau \in S_n$ gilt $P(\sigma\circ\tau) = P(\sigma)P(\tau)$.*
(2) *Für $\sigma \in S_n$ gilt $\varepsilon(\sigma) = \det\big(P(\sigma)\big)$.*
(3) *Für $\sigma,\tau \in S_n$ gilt $\varepsilon(\sigma\circ\tau) = \varepsilon(\sigma)\varepsilon(\tau)$.*
(4) *Ist σ eine Transposition (also eine Permutation, die zwei Elemente vertauscht und alle anderen nicht ändert), dann ist $\varepsilon(\sigma) = -1$.*

Beweis.

(1) Für alle $j \in \{1, 2, \ldots, n\}$ gilt
$$P(\sigma)P(\tau)\mathbf{e}_j = P(\sigma)\mathbf{e}_{\tau(j)} = \mathbf{e}_{\sigma(\tau(j))} = \mathbf{e}_{(\sigma\circ\tau)(j)} = P(\sigma\circ\tau)\mathbf{e}_j\,,$$
also ist $P(\sigma\circ\tau) = P(\sigma)P(\tau)$.

(2) Der einzige von null verschiedene Term in der Formel 14.17 für $\det(P(\sigma^{-1}))$ ist $\varepsilon(\sigma)\delta_{1,1}\cdots\delta_{n,n} = \varepsilon(\sigma)$. Es folgt (wegen $P(\sigma)^{-1} = P(\sigma^{-1})$ nach Teil (1) und $\varepsilon(\sigma) = \pm 1$)
$$\det\big(P(\sigma)\big) = \det\big(P(\sigma)\big)^{-1} = \det\big(P(\sigma)^{-1}\big) = \det\big(P(\sigma^{-1})\big) = \varepsilon(\sigma)\,.$$

(3) Das folgt aus (1) und (2) und der Multiplikativität der Determinante.

(4) In diesem Fall erhält man $P(\sigma)$ aus I_n durch Vertauschen zweier Spalten (oder Zeilen), also ist $\varepsilon(\sigma) = \det(P(\sigma)) = -\det(I_n) = -1$. ❑

Da sich (wie man sich leicht überlegen kann) jede Permutation als Komposition von Transpositionen schreiben lässt, ist ε durch die Eigenschaften (3) und (4) in Lemma 14.20 eindeutig festgelegt: Ist σ Komposition einer geraden Anzahl von Transpositionen, dann ist σ gerade, sonst ungerade.

Es gibt eine Art Formel für $\varepsilon(\sigma)$. Dazu eine kleine Definition:

Definition. Sei $\sigma \in S_n$. Ein Paar (i, j) mit $1 \le i < j \le n$ heißt *Fehlstand* von σ, wenn $\sigma(i) > \sigma(j)$ ist. ♢ **DEF** Fehlstand

Dann gilt der folgende Satz.

Satz. *Sei $\sigma \in S_n$ und sei m die Anzahl der Fehlstände von σ. Dann ist $\varepsilon(\sigma) = (-1)^m$.* **SATZ** Signum und Fehlstände

Beweis. Sei $\varepsilon'(\sigma)$ die durch $(-1)^{\text{Anzahl Fehlstände von } \sigma}$ definierte Funktion. Die Transposition τ, die k und l vertauscht (mit $k < l$), hat genau $m = 1 + 2(l - k - 1)$ Fehlstände (nämlich (k, l) sowie (k, j) und (j, l) für alle $k < j < l$). Da m ungerade ist, ist $\varepsilon'(\tau) = (-1)^m = -1 = \varepsilon(\tau)$.

Außerdem gilt
$$\varepsilon'(\sigma) = \prod_{1\le i<j\le n} \frac{\sigma(j) - \sigma(i)}{j - i}\,,$$
denn das rechts stehende Produkt hat Betrag 1 (jeder Faktor im Nenner tritt bis aufs Vorzeichen auch im Zähler auf) und $\sigma(j) - \sigma(i)$ ist genau dann negativ, wenn (i, j) ein Fehlstand von σ ist. Es folgt
$$\begin{aligned}
\varepsilon'(\sigma\circ\tau) &= \prod_{1\le i<j\le n} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{j - i} \\
&= \prod_{1\le i<j\le n} \frac{\sigma(\tau(j)) - \sigma(\tau(i))}{\tau(j) - \tau(i)} \prod_{1\le i<j\le n} \frac{\tau(j) - \tau(i)}{j - i} \\
&= \prod_{1\le k<l\le n} \frac{\sigma(l) - \sigma(k)}{l - k} \prod_{1\le i<j\le n} \frac{\tau(j) - \tau(i)}{j - i} = \varepsilon'(\sigma)\varepsilon'(\tau)\,.
\end{aligned}$$
Die Funktion ε' hat also die Eigenschaften (3) und (4) aus Lemma 14.20 und muss daher mit ε übereinstimmen. ❑

Zum Abschluss dieses Kapitels wollen wir noch die geometrische Bedeutung der Determinante untersuchen. Wir betrachten den $\mathbb{R}^n$ und definieren erst einmal den Begriff der (positiven oder negativen) Orientierung einer Basis.

* 14.21. **Definition.** Sei $(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$ eine Basis des $\mathbb{R}^n$ und $A = \left(\boldsymbol{b}_1 \,\middle|\, \cdots \,\middle|\, \boldsymbol{b}_n\right)$ die Matrix, deren Spalten die Basisvektoren sind. Wir sagen, die Basis sei *positiv orientiert*, wenn $\det(A) > 0$ ist, und *negativ orientiert*, wenn $\det(A) < 0$ ist. ♢ **DEF** Orientierung einer Basis

Die Standardbasis ist positiv orientiert. Im Fall $n = 2$ ist eine Basis genau dann positiv orientiert, wenn der gegen den Uhrzeigersinn gemessene Winkel vom ersten zum zweiten Basisvektor kleiner ist als π $(= 180°)$.

Der Vergleich der Orientierung einer Basis und ihres Bildes führt zum Begriff der orientierungserhaltenden bzw. -umkehrenden linearen Abbildung.

14.22. **Definition.** Sei $f\colon \mathbb{R}^n \to \mathbb{R}^n$ ein Automorphismus (also ein invertierbarer Endomorphismus). Dann heißt f *orientierungserhaltend*, wenn f positiv orientierte Basen auf positiv orientierte Basen abbildet, und *orientierungsumkehrend*, wenn f positiv orientierte Basen auf negativ orientierte Basen abbildet. ♢ **DEF** orientierungserhaltend, -umkehrend

Man sieht leicht, dass f genau dann orientierungserhaltend (-umkehrend) ist, wenn $\det(A) > 0$ (< 0) ist, wobei A die zu f gehörende Matrix ist: Ist B die Matrix, deren Spalten die Vektoren einer positiv orientierten Basis bilden, dann sind die Spalten von AB die Bilder der Basisvektoren. Wegen $\det(B) > 0$ gilt dann

$$\det(AB) = \det(A)\det(B) > 0 \iff \det(A) > 0\,.$$

Wir wollen jetzt das Volumen von „linear verzerrten Würfeln“ betrachten. In der Ebene $\mathbb{R}^2$ sind das Parallelogramme. Allgemeiner definieren wir:

14.23. **Definition.** Ein *Parallelotop* im $\mathbb{R}^n$ ist die Menge **DEF** Parallelotop

$$P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) = \{t_1\boldsymbol{x}_1 + t_2\boldsymbol{x}_2 + \ldots + t_n\boldsymbol{x}_n \mid t_1, t_2, \ldots, t_n \in [0,1]\} \subset \mathbb{R}^n$$

für ein n-Tupel $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ von Vektoren im $\mathbb{R}^n$. Das Parallelotop heißt *ausgeartet*, wenn die es aufspannenden Vektoren $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ linear abhängig sind (dann ist $P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ im echten Untervektorraum $\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \rangle$ von $\mathbb{R}^n$ enthalten). ♢

Wir wollen jetzt untersuchen, wie man das „orientierte Volumen“ solcher Parallelotope definieren kann. Es sollte folgende Eigenschaften haben:

(1) $\operatorname{vol} P(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = 1$ (der n-dimensionale Einheitswürfel hat Volumen 1).

(2) $\operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ ist positiv (bzw. negativ), wenn $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ eine positiv (bzw. negativ) orientierte Basis von $\mathbb{R}^n$ ist.

(3) $\operatorname{vol} P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{j-1}, \lambda\boldsymbol{x}_j, \boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_n) = \lambda \operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ für $\lambda \in \mathbb{R}$.

(4) $\operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) = 0$, wenn $P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ ausgeartet ist.

(5) $$\begin{aligned}&\operatorname{vol} P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_j + \boldsymbol{x}'_j, \boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_n)\\ &\quad = \operatorname{vol} P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{j-1}, \boldsymbol{x}_j, \boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_n)\\ &\qquad + \operatorname{vol} P(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{j-1}, \boldsymbol{x}'_j, \boldsymbol{x}_{j+1}, \ldots, \boldsymbol{x}_n).\end{aligned}$$

Diese letzte Eigenschaft kann man sich plausibel machen, wenn man an die Formel „Grundfläche mal Höhe“ denkt: Die Höhe von $\boldsymbol{x}_j + \boldsymbol{x}'_j$ über der „Grundfläche“, die durch das von den übrigen Vektoren aufgespannte Parallelotop gegeben ist, ist die Summe der (orientierten) Höhen von $\boldsymbol{x}_j$ und $\boldsymbol{x}'_j$.

SATZ Determinante ist Volumen

14.24. **Satz.** *Die einzige Abbildung* vol *von der Menge der Parallelotope im* $\mathbb{R}^n$ *nach* $\mathbb{R}$*, die die obigen Eigenschaften hat, ist*

$$\operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n) = \det(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)\,,$$

wobei $\det(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$ *für die Determinante der Matrix steht, deren Spalten die Vektoren* $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ *sind.*

Beweis. Die Determinante hat jedenfalls die geforderten Eigenschaften (Satz 14.3 und Satz 14.8, sowie Definition 14.21). Aus der Eindeutigkeitsaussage in Satz 14.3, zusammen mit Satz 14.8, der besagt, dass die analoge Aussage auch für Spalten statt Zeilen gilt, folgt, dass die Determinante die einzige Abbildung ist, die die Eigenschaften (1), (3), (4) und (5) hat. ❑

Man kann also mit Hilfe der Determinante Volumen messen.

BSP Fläche eines Dreiecks

14.25. **Beispiel.** Die Fläche des Dreiecks mit den Eckpunkten (x_1, y_1), (x_2, y_2) und (x_3, y_3) ist

$$\frac{1}{2}\Big|\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}\Big|\,.$$

Wir können das Dreieck so verschieben, dass die erste Ecke im Ursprung zu liegen kommt. Dann ist die gesuchte Fläche die Hälfte der Fläche des von (x_2-x_1, y_2-y_1) und (x_3-x_1, y_3-y_1) aufgespannten Parallelogramms. Die orientierte Fläche dieses Parallelogramms ist

$$\det\begin{pmatrix} x_2-x_1 & x_3-x_1 \\ y_2-y_1 & y_3-y_1 \end{pmatrix}\,.$$

Die obige Determinante lässt sich durch die Spaltenoperationen $\mathbf{II}_{2,1}(-1)$ und $\mathbf{II}_{3,1}(-1)$ und nachfolgender Entwicklung nach der dritten Zeile auf diese Form bringen. Durch den Absolutbetrag erhalten wir die Fläche statt der orientierten Fläche. ♣

Aus der Multiplikativität der Determinante folgt eine Interpretation der Determinante eines Endomorphismus, die für Anwendungen in der Analysis (z.B. die Transformationsformel für mehrdimensionale Integrale) relevant ist.

SATZ Determinante ist Skalierung des Volumens

∗ 14.26. **Satz.** *Sei* $f\colon \mathbb{R}^n \to \mathbb{R}^n$ *linear mit zugehöriger Matrix* A *und seien außerdem* $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n \in \mathbb{R}^n$*. Dann gilt*

$$\operatorname{vol} f\big(P(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)\big) = \det(A) \operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)\,.$$

Die Determinante eines Endomorphismus gibt also an, mit welchem Faktor das Volumen bei seiner Anwendung multipliziert wird.

Beweis. Sei X die Matrix, deren Spalten die Vektoren $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$ sind. Da f linear ist, gilt $f\big(P(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)\big) = P\big(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \dots, f(\boldsymbol{x}_n)\big)$. Es folgt

$$\begin{aligned} \operatorname{vol} f\big(P(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)\big) &= \det\big(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \dots, f(\boldsymbol{x}_n)\big) \\ &= \det(A\boldsymbol{x}_1, A\boldsymbol{x}_2, \dots, A\boldsymbol{x}_n) \\ &= \det(AX) = \det(A)\det(X) \\ &= \det(A) \operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)\,. \end{aligned}$$

❑

15. Eigenwerte und Eigenvektoren

Im vorletzten Abschnitt haben wir einen Klassifikationssatz bewiesen (Satz 13.10). Man kann ihn so interpretieren, dass die einzige Eigenschaft, die lineare Abbildungen zwischen zwei gegebenen endlich-dimensionalen Vektorräumen voneinander unterscheidet, der *Rang* ist: Ist $f\colon V \to W$ linear mit $\mathrm{rk}(f) = r$, dann können wir Basen von V und W so wählen, dass f durch die Matrix M_r gegeben ist, und M_r hängt nur von r (und den Dimensionen von V und W) ab.

Wir werden jetzt statt linearen Abbildungen zwischen verschiedenen Vektorräumen V und W *Endomorphismen* $f\colon V \to V$ betrachten. Da wir nur einen Vektorraum haben, können wir auch nur *eine* Basis wählen. Wir haben also deutlich weniger Spielraum, was sich in einem erheblich schwierigeren Klassifikationsproblem niederschlägt.

Natürlich kann man lineare Abbildungen $f\colon V \to V$ auch als Spezialfall von linearen Abbildungen $V \to W$ betrachten, wo „zufällig“ $W = V$ ist. Dann wird man die Wahl von verschiedenen Basen von V auf der Quell- und der Zielseite zulassen. Zum Beispiel erhält man so die Basiswechselmatrizen $\mathrm{Mat}_{B,B'}(\mathrm{id}_V)$. Auf der anderen Seite geht so aber die Information verloren, dass es sich wirklich auf beiden Seiten um *denselben* Vektorraum handelt und nicht um zwei Vektorräume, die zufällig isomorph sind (d.h., dieselbe Dimension haben). Für das Klassifikationsproblem, das wir in diesem Abschnitt (und dann weiter in der *Linearen Algebra II*) studieren wollen, ist es aber wesentlich, dass f als Endomorphismus von V betrachtet wird. Anderenfalls wäre eine Aussage der Form $f(v) = \lambda v$ (siehe Definition 15.3 unten) nicht sinnvoll, bzw. sie würde sich nicht auf die f beschreibenden Matrizen übertragen.

Wir schreiben erst einmal auf, wie die Matrizen von f bezüglich verschiedener Basen von V miteinander zusammenhängen.

SATZ Matrizen eines Endomorphismus

15.1. **Satz.** *Sei V ein K-Vektorraum mit Basis $B = (b_1, b_2, \dots, b_n)$ und sei f ein Endomorphismus von V. Sei weiter $A = \mathrm{Mat}_{B,B}(f) \in \mathrm{Mat}(n, K)$ die Matrix von f bezüglich B. Dann ist*

$$\{\mathrm{Mat}_{B',B'}(f) \mid B' \text{ Basis von } V\} = \{PAP^{-1} \mid P \in \mathrm{GL}(n, K)\}\,.$$

Beweis. Es ist $\mathrm{Mat}_{B',B'}(f) = \mathrm{Mat}_{B,B'}(\mathrm{id}_V)\,\mathrm{Mat}_{B,B}(f)\,\mathrm{Mat}_{B',B}(\mathrm{id}_V) = PAP^{-1}$ mit $P = \mathrm{Mat}_{B,B'}(\mathrm{id}_V) \in \mathrm{GL}(n, K)$. Umgekehrt lässt sich jede Matrix $P \in \mathrm{GL}(n, K)$ in dieser Form schreiben (Folgerung 13.4). ❑

* **DEF** Ähnlichkeit von Matrizen

15.2. **Definition.** Sei K ein Körper und $n \in \mathbb{N}$. Zwei Matrizen $A, A' \in \mathrm{Mat}(n, K)$ heißen *ähnlich*, wenn es eine Matrix $P \in \mathrm{GL}(n, K)$ gibt mit $A' = PAP^{-1}$. ♢

Ähnlich wie für die Äquivalenz von Matrizen zeigt man, dass die Ähnlichkeit von Matrizen eine Äquivalenzrelation ist.

Wenn A eine Matrix eines Endomorphismus f von V ist, dann sind die Matrizen von f bezüglich beliebiger Basen von V also gerade die zu A ähnlichen Matrizen.

Die Klassifikation von Matrizen bis auf Ähnlichkeit (und damit die Klassifikation der Endomorphismen endlich-dimensionaler Vektorräume) ist relativ kompliziert. Sie wird durch die *Jordan-Normalform* geleistet, die wir im nächsten Semester besprechen werden. Hier werden wir uns erst einmal auf die Diskussion einfacherer „Invarianten“ (also Daten, die nur von f und nicht von der Basis abhängen) beschränken.

Die Idee ist, den Endomorphismus $f\colon V \to V$ mit anderen besonders einfachen Endomorphismen zu vergleichen. Die einfachsten Endomorphismen sind sicher die Multiplikationen mit einem Skalar $\lambda \in K$: $v \mapsto \lambda v$. Wir können uns fragen, ob es Elemente von V gibt, die sich unter f und dieser Abbildung gleich verhalten. Das führt auf folgende Definition.

* 15.3. **Definition.** Seien V ein K-Vektorraum und $f \in \mathrm{End}(V)$. Ein Skalar $\lambda \in K$ heißt *Eigenwert* von f, wenn es einen Vektor $\mathbf{0} \neq v \in V$ gibt, sodass $f(v) = \lambda v$ ist. Jeder solche Vektor heißt ein *Eigenvektor* von f zum Eigenwert λ. ♢

DEF
Eigenwert
Eigenvektor

Man beachte die Bedingung $v \neq \mathbf{0}$! Ohne sie wäre die Definition sinnlos, weil dann jedes λ ein Eigenwert wäre (denn $f(\mathbf{0}) = \mathbf{0} = \lambda\mathbf{0}$).

* 15.4. **Definition.** Seien V ein K-Vektorraum, $\lambda \in K$ und $f \in \mathrm{End}(V)$. Der Untervektorraum

$$E_\lambda(f) = \{v \in V \mid f(v) = \lambda v\} = \ker(\lambda\,\mathrm{id}_V - f)$$

von V heißt der λ-*Eigenraum* von f.

Die Dimension $\dim E_\lambda(f)$ des λ-Eigenraums von f heißt die *geometrische Vielfachheit* des Eigenwerts λ von f. ♢

DEF
Eigenraum
geom. Vielfachheit

$E_\lambda(f)$ besteht also aus dem Nullvektor und den Eigenvektoren zum Eigenwert λ. λ ist genau dann ein Eigenwert von f, wenn $E_\lambda(f) \neq \{\mathbf{0}\}$, also die geometrische Vielfachheit positiv ist.

15.5. **Beispiele.**

BSP
Eigenwerte
Eigenräume

(1) $E_0(f)$ ist gerade der Kern von f. Null ist also genau dann ein Eigenwert von f, wenn f nicht injektiv ist. Ist V endlich-dimensional, dann ist das auch dazu äquivalent, dass f kein Isomorphismus ist.

(2) Sei $f\colon \mathbb{R}^2 \to \mathbb{R}^2$, $(x, y) \mapsto (y, x)$. Dann hat f die Eigenwerte 1 und -1, denn $v_1 = (1, 1) \in E_1(f)$ und $v_{-1} = (1, -1) \in E_{-1}(f)$. Man sieht leicht, dass beide Eigenräume eindimensional sind.

(3) Sei $f\colon \mathcal{C}^\infty(\mathbb{R}) \to \mathcal{C}^\infty(\mathbb{R})$, $h \mapsto h'$. Dann hat f jedes $\lambda \in \mathbb{R}$ als Eigenwert, und es gilt $E_\lambda(f) = \langle x \mapsto e^{\lambda x}\rangle$. (Beweis wie in Beispiel 12.11.) ♣

Wir werden sehen, dass die Situation von Beispiel (2) oben für Endomorphismen von endlich-dimensionalen Vektorräumen recht typisch ist. Wir zeigen erst einmal, dass es nicht zu viele Eigenwerte geben kann.

* 15.6. **Satz.** *Seien V ein K-Vektorraum und $f \in \mathrm{End}(V)$. Seien $\lambda_1, \ldots, \lambda_m \in K$ paarweise verschieden und für $j \in \{1, 2, \ldots, m\}$ sei $v_j \in V$ ein Eigenvektor von f zum Eigenwert λ_j. Dann ist $(v_1, v_2, \ldots, v_m)$ linear unabhängig.*

SATZ
Lin. Unabh. von Eigenvektoren

Beweis. Seien $\alpha_1, \alpha_2, \ldots, \alpha_m \in K$ mit $\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_m v_m = \mathbf{0}$. Wir müssen zeigen, dass alle $\alpha_j = 0$ sind. Dazu verwenden wir Induktion über m. Der Fall

$m = 0$ ist klar (null Vektoren sind stets linear unabhängig). Sei also $m > 0$ und die Behauptung für $m - 1$ schon bewiesen. Es ist

$$\begin{aligned}\mathbf{0} &= \lambda_m \mathbf{0} - f(\mathbf{0}) \\ &= \lambda_m(\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_m v_m) - f(\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_m v_m) \\ &= \lambda_m \alpha_1 v_1 + \lambda_m \alpha_2 v_2 + \ldots + \lambda_m \alpha_m v_m - \alpha_1 \lambda_1 v_1 - \alpha_2 \lambda_2 v_2 - \ldots - \alpha_m \lambda_m v_m \\ &= \alpha_1(\lambda_m - \lambda_1) v_1 + \alpha_2(\lambda_m - \lambda_2) v_2 + \ldots + \alpha_{m-1}(\lambda_m - \lambda_{m-1}) v_{m-1}\,.\end{aligned}$$

Aus der Induktionsannahme folgt

$$\alpha_1(\lambda_m - \lambda_1) = \alpha_2(\lambda_m - \lambda_2) = \ldots = \alpha_{m-1}(\lambda_m - \lambda_{m-1}) = 0\,.$$

Weil $\lambda_m \neq \lambda_1, \lambda_2, \ldots, \lambda_{m-1}$ ist, ergibt sich daraus $\alpha_1 = \alpha_2 = \ldots = \alpha_{m-1} = 0$. Die ursprüngliche Gleichung reduziert sich also auf $\alpha_m v_m = \mathbf{0}$. Weil $v_m \neq \mathbf{0}$ ist, folgt auch $\alpha_m = 0$. ❑

FOLG Dimension von Eigenräumen

15.7. **Folgerung.** *Seien V ein K-Vektorraum und $f \in \operatorname{End}(V)$. Seien weiter $\lambda_1, \lambda_2, \ldots, \lambda_m$ paarweise verschiedene Elemente von K und für $j \in \{1, 2, \ldots, m\}$ seien $v_{j1}, v_{j2}, \ldots, v_{jn_j}$ (mit $n_j \in \mathbb{N}$) linear unabhängige Elemente von $E_{\lambda_j}(f)$. Dann sind die v_{ji} (mit $j \in \{1, 2, \ldots, m\}$ und $i \in \{1, 2, \ldots, n_j\}$) linear unabhängig. Insbesondere gilt*

$$\dim E_{\lambda_1}(f) + \dim E_{\lambda_2}(f) + \ldots + \dim E_{\lambda_m}(f) \leq \dim V\,.$$

Ist V endlich-dimensional, dann kann f also höchstens $\dim V$ Eigenwerte haben. Genauer gilt, dass die Summe der geometrischen Vielfachheiten der Eigenwerte höchstens $\dim V$ sein kann.

Beweis. Sei

$$\sum_{j=1}^{m} \sum_{i=1}^{n_j} \alpha_{ji} v_{ji} = \mathbf{0}$$

mit $\alpha_{ji} \in K$. Sei $v_j = \sum_{i=1}^{n_j} \alpha_{ji} v_{ji} \in E_{\lambda_j}(f)$, dann gilt $v_1 + v_2 + \ldots + v_m = \mathbf{0}$. Aus Satz 15.6 folgt dann $v_1 = v_2 = \ldots = v_m = \mathbf{0}$, denn eventuell vorkommende Vektoren $\neq \mathbf{0}$ müssten linear unabhängig sein und könnten sich also nicht zum Nullvektor addieren. Da $(v_{j1}, v_{j2}, \ldots, v_{jn_j})$ linear unabhängig ist, folgt dann aus $v_j = \mathbf{0}$ auch $\alpha_{ji} = 0$ für alle $i \in \{1, 2, \ldots, n_j\}$. Da das für jedes $j \in \{1, 2, \ldots, m\}$ gilt, sind alle $\alpha_{ji} = 0$. Das zeigt die Behauptung. Die letzte Aussage folgt, wenn man für $(v_{j1}, v_{j2}, \ldots, v_{jn_j})$ eine Basis von $E_{\lambda_j}(f)$ wählt, aus

$$\begin{aligned}\dim V &\geq \#\big\{v_{ji} \mid j \in \{1, 2, \ldots, m\}, i \in \{1, 2, \ldots, n_j\}\big\} \\ &= n_1 + n_2 + \ldots + n_m \\ &= \dim E_{\lambda_1}(f) + \dim E_{\lambda_2}(f) + \ldots + \dim E_{\lambda_m}(f)\,.\end{aligned}$$ ❑

Wie können wir die Eigenwerte (und dann die Eigenräume) finden? Dazu wählen wir eine Basis und bestimmen die Matrix A von f bezüglich dieser Basis. Wir übertragen die Begriffe Eigenwert usw. auf Matrizen.

15.8. **Definition.** Seien K ein Körper und $A \in \mathrm{Mat}(n, K)$. Sei $\lambda \in K$. Dann heißt λ ein *Eigenwert* von A, wenn es einen Spaltenvektor $\mathbf{0} \neq \boldsymbol{x} \in K^n$ gibt mit $A\boldsymbol{x} = \lambda\boldsymbol{x}$. In diesem Fall heißt $\boldsymbol{x}$ ein *Eigenvektor* von A zum Eigenwert λ. Der Untervektorraum

DEF Eigenwert etc. für Matrizen

$$E_\lambda(A) = \{\boldsymbol{x} \in K^n \mid A\boldsymbol{x} = \lambda\boldsymbol{x}\} = \ker(\lambda I_n - A)$$

heißt der *Eigenraum* von A zum Eigenwert λ; seine Dimension heißt die *geometrische Vielfachheit* des Eigenwerts λ von A. ♢

Die Eigenwerte von A und ihre geometrischen Vielfachheiten entsprechen dann denen von f.

Der Schlüssel zur Bestimmung der Eigenwerte ist folgende einfache Beobachtung.

15.9. **Lemma.** *Seien K ein Körper, $\lambda \in K$ und $A \in \mathrm{Mat}(n, K)$. λ ist genau dann ein Eigenwert von A, wenn $\det(\lambda I_n - A) = 0$ ist. Die geometrische Vielfachheit des Eigenwerts λ ist $\dim\ker(\lambda I_n - A) = n - \mathrm{rk}(\lambda I_n - A)$.*

LEMMA Charakterisierung von Eigenwerten

Beweis. Wir haben folgende Kette von Äquivalenzen:

$$\begin{aligned} \lambda \text{ ist Eigenwert von } A &\iff E_\lambda(A) \neq \{\mathbf{0}\} \\ &\iff \ker(\lambda I_n - A) \neq \{\mathbf{0}\} \\ &\iff \det(\lambda I_n - A) = 0 \end{aligned}$$

Die letzte Aussage folgt aus der Definition der geometrischen Vielfachheit. ❑

15.10. **Beispiel.** Wir betrachten

BSP Bestimmung der Eigenwerte

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathrm{Mat}(2, \mathbb{R})\,.$$

(Das ist die Matrix zu $f\colon (x, y) \mapsto (y, x)$ wie in Beispiel 15.5 (2).) Dann ist (für $\lambda \in \mathbb{R}$)

$$\det(\lambda I_2 - A) = \begin{vmatrix} \lambda & -1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 - 1 = (\lambda - 1)(\lambda + 1)\,.$$

Das verschwindet genau für $\lambda = 1$ und $\lambda = -1$, also sind das die Eigenwerte von A. Wir können Basen der Eigenräume $E_\lambda(A)$ mit dem Zeilenstufenform-Algorithmus, angewandt auf $\lambda I_2 - A$, berechnen. Für $\lambda = 1$ haben wir

$$\lambda I_2 - A = I_2 - A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix};$$

das liefert die Basis $(1, 1)$ für $E_1(A)$. Für $\lambda = -1$ sieht es so aus:

$$\lambda I_2 - A = -I_2 - A = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix};$$

die Basis ist $(-1, 1)$. ♣

An diesem Beispiel sieht man, dass die Determinante, deren Verschwinden anzeigt, dass λ ein Eigenwert ist, ein *Polynom* in λ (mit Koeffizienten in K) ist. Wir müssen daher etwas ausholen und ein wenig über Polynome sprechen.

Exkurs: Polynome.

DEF
Polynom
Polynomring

15.11. **Definition.** Sei K ein Körper. Ein *Polynom* in der *Variablen* (oder *Unbestimmten*) X über K ist ein Ausdruck der Form

$$p = a_n X^n + a_{n-1} X^{n-1} + \ldots + a_1 X + a_0$$

mit $n \in \mathbb{N}$ und $a_0, a_1, \ldots, a_n \in K$. a_j heißt der *j-te Koeffizient* von p oder der *Koeffizient von X^j* in p. Wir setzen $a_j = 0$ für $j > n$. Ist $a_n \neq 0$, dann hat das Polynom *Grad n*: $\deg(p) = n$ (englisch „degree“). In diesem Fall heißt a_n der *Leitkoeffizient* von p. Ist $a_n = 1$, dann heißt p *normiert*. Sind alle $a_j = 0$, dann ist p das *Nullpolynom*; sein Grad ist definiert als $\deg(\mathbf{0}) = -\infty$. Ist $n = 0$, dann heißt p *konstant* (d.h., $p = \mathbf{0}$ oder $\deg(p) = 0$). Wir schreiben $K[X]$ für die Menge der Polynome in X über K.

Sei $q = b_m X^m + \ldots + b_1 X + b_0$. Die Polynome p und q sind genau dann *gleich*, $p = q$, wenn ihre Koeffizienten übereinstimmen: $a_j = b_j$ für alle $j \in \mathbb{N}$ (mit der Konvention $a_j = 0$ für $j > n$ und $b_j = 0$ für $j > m$). Die *Summe* von p und q ist

$$p + q = \sum_{j=0}^{\max\{m,n\}} (a_j + b_j) X^j \,;$$

es gilt $\deg(p+q) \leq \max\{\deg(p), \deg(q)\}$, und das *Produkt* von p und q ist

$$p \cdot q = \sum_{k=0}^{m+n} \Big(\sum_{i,j\colon i+j=k} a_i b_j \Big) X^k \,.$$

Es gilt $\deg(pq) = \deg(p) + \deg(q)$. Wir identifizieren K mit der Teilmenge der konstanten Polynome: $K \subset K[X]$. Die Menge $K[X]$ wird mit der eben definierten Addition und Multiplikation ein kommutativer Ring; er heißt der *Polynomring* in X über K. Die Einschränkung der Multiplikation auf $K \times K[X]$ macht $K[X]$ zu einem unendlich-dimensionalen K-Vektorraum mit Basis $(1, X, X^2, X^3, \ldots)$. ♢

Wenn es Sie stört, dass in der Definition von einem „Ausdruck der Form . . . “ gesprochen wird, ohne dass gesagt wird, was das eigentlich „ist“, dann lesen Sie hier weiter.

Formal kann man die Definition auf stabile Füße stellen, indem man setzt

$$K[X] = \{(a_n)_{n\in\mathbb{N}} \in K^{\mathbb{N}} \mid \exists N \in \mathbb{N}\, \forall n > N \colon a_n = 0\} \,.$$

Das sind also die endlichen Folgen von Elementen von K, in dem Sinne, dass alle bis auf endlich viele Folgenglieder null sind. Man setzt weiter $X = (0, 1, 0, 0, 0, \ldots) \in K[X]$ und definiert die Abbildung $i\colon K \to K[X]$, $a \mapsto (a, 0, 0, 0, \ldots)$. Die Addition in $K[X]$ wird komponentenweise definiert, die Multiplikation mit X durch

$$X \cdot (a_0, a_1, a_2, \ldots) = (0, a_0, a_1, a_2, \ldots)$$

und die mit $i(a)$ durch

$$i(a) \cdot (a_0, a_1, a_2, \ldots) = (aa_0, aa_1, aa_2, \ldots) \,.$$

Dann ist

$$(a_0, a_1, \ldots, a_n, 0, 0, 0, \ldots) = i(a_0) + i(a_1)X + i(a_2)X^2 + \ldots + i(a_n)X^n \,,$$

und die Multiplikation damit wird so definiert, dass das Assoziativ- und das Distributivgesetz gelten. Mittels der Abbildung i wird K mit seinem Bild in $K[X]$ identifiziert; man schreibt also einfach a statt $i(a)$. Die Ringaxiome muss man dann noch nachprüfen. Die K-Vektorraum-Struktur von $K[X]$ ist einfach die als Untervektorraum von $K^{\mathbb{N}}$.

Das funktioniert auch dann noch, wenn man die Endlichkeitsbedingung in der Definition weglässt. Man erhält dann den Ring $K[\![X]\!]$ der *formalen Potenzreihen* in X über K. Für

$K = \mathbb{R}$ oder $\mathbb{C}$ spielen diese Potenzreihen eine wichtige Rolle in der Analysis (bzw. Funktionentheorie).

In Polynome kann man einsetzen:

DEF Werte und Nullstellen von Polynomen

15.12. **Definition.** Seien K ein Körper und $p = a_n X^n + \ldots + a_1 X + a_0 \in K[X]$ ein Polynom. Für $\lambda \in K$ ist der *Wert von p bei λ* gegeben durch

$$p(\lambda) = a_n \lambda^n + \ldots + a_1 \lambda + a_0 \,.$$

λ heißt eine *Nullstelle* von p, wenn $p(\lambda) = 0$ ist. Für $p, q \in K[X]$ und $\lambda \in K$ gilt dann $(p+q)(\lambda) = p(\lambda) + q(\lambda)$ und $(p \cdot q)(\lambda) = p(\lambda) \cdot q(\lambda)$. ♢

Ein Polynom $p \in K[X]$ führt also zu einer *Polynomfunktion* $K \to K$, $\lambda \mapsto p(\lambda)$. Die Abbildung $K[X] \to \mathrm{Abb}(K, K)$, die einem Polynom die zugehörige Polynomfunktion zuordnet, ist injektiv, wenn der Körper K unendlich ist. Das ergibt sich aus dem folgenden Satz.

SATZ Eindeutigkeit von Polynomen

15.13. **Satz.** *Seien K ein Körper, $n \in \mathbb{N}$ und $x_1, x_2, \ldots, x_n \in K$ paarweise verschieden. Seien weiter $y_1, y_2, \ldots, y_n \in K$. Dann gibt es ein eindeutig bestimmtes Polynom $p \in K[X]$ mit $\deg(p) < n$, sodass $p(x_j) = y_j$ ist für alle $j \in \{1, 2, \ldots, n\}$.*

Beweis. Die Existenz folgt mit der Lagrangeschen Interpolationsformel wie in Beispiel 10.15. Damit ist die lineare Abbildung

$$\phi\colon \{p \in K[X] \mid \deg(p) < n\} \longrightarrow K^n, \quad p \longmapsto \big(p(x_1), p(x_2), \ldots, p(x_n)\big)$$

surjektiv. Da die beiden beteiligten Vektorräume dieselbe Dimension n haben (der Vektorraum der Polynome vom Grad $< n$ hat Basis $(1, X, X^2, \ldots, X^{n-1})$), folgt aus der Surjektivität die Bijektivität. ❑

FOLG Nullstellen von Polynomen

15.14. **Folgerung.** *Ein Polynom $p \in K[X]$ mit $\deg(p) = n \in \mathbb{N}$ kann nicht mehr als n Nullstellen in K haben.*

Beweis. Angenommen, p hat $n+1$ Nullstellen $x_1, x_2, \ldots, x_{n+1} \in K$. Dann muss p das eindeutig bestimmte Polynom von Grad $< n+1$ sein, dass $p(a_j) = 0$ erfüllt für alle $j \in \{1, 2, \ldots, n+1\}$. Das Nullpolynom hat aber diese Eigenschaft, also muss $p = \mathbf{0}$ sein. Das ist aber ein Widerspruch zur Voraussetzung $\deg(p) = n$. ❑

FOLG Polynome und Polynomfunktionen

15.15. **Folgerung.** *Hat der Körper K unendlich viele Elemente, dann ist die Abbildung $K[X] \to \mathrm{Abb}(K, K)$, die einem Polynom die zugehörige Polynomfunktion zuordnet, injektiv.*

Ein Polynom $p \in K[X]$ ist dann also durch seine Werte $p(\lambda)$ für $\lambda \in K$ eindeutig bestimmt. Wir können also zum Beispiel den Vektorraum P der Polynomfunktionen mit $\mathbb{R}[X]$ identifizieren.

Beweis. Wir schreiben Φ für die Abbildung $K[X] \to \mathrm{Abb}(K, K)$. Φ ist linear, also genügt es zu zeigen, dass $\ker(\Phi) = \{\mathbf{0}\}$ ist. Sei also $p \in \ker(\Phi)$. Dann ist $p(\lambda) = 0$ für alle $\lambda \in K$, also hat (da K unendlich ist) das Polynom p unendlich viele Nullstellen in K. Nach Folgerung 15.14 muss p das Nullpolynom sein. ❑

Für endliche Körper K ist die Aussage falsch: Ist $\#K = q < \infty$, dann ist $\dim_K \mathrm{Abb}(K,K) = q$, denn eine Abbildung $f\colon K \to K$ ist durch die q Werte $f(a)$ für $a \in K$ eindeutig bestimmt. Auf der anderen Seite ist $\dim_K K[X] = \infty$, und damit kann es keine injektive lineare Abbildung $K[X] \to \mathrm{Abb}(K,K)$ geben.

Der Kern von Φ besteht in diesem Fall aus allen Polynomen, die alle Elemente von K als Nullstellen haben. Man kann zeigen, dass

$$\prod_{a\in K}(X-a) = X^q - X$$

ist; der Kern besteht demnach genau aus den Vielfachen von $X^q - X$.

So wie man ganze Zahlen mit Rest durcheinander dividieren kann, gibt es auch für Polynome eine Division mit Rest („Polynomdivision“).

15.16. **Satz.** *Sei K ein Körper und seien $f, g \in K[X]$ mit g normiert. Dann gibt es eindeutig bestimmte Polynome q („Quotient“) und r („Rest“) in $K[X]$, sodass $f = qg + r$ und $\deg(r) < \deg(g)$.* **SATZ** Polynomdivision

Beweis. Wir beweisen zunächst die Existenz. Sei $\deg(g) = m$, also

$$g = X^m + b_{m-1}X^{m-1} + \ldots + b_1 X + b_0\,.$$

Wir betrachten g als fest und führen den Beweis durch Induktion über $\deg(f)$. Ist $\deg(f) < m$, dann erfüllen $q = 0$ und $r = f$ die Bedingungen. Wir können also annehmen, dass $n = \deg(f) \geq m$ ist; die Existenzaussage sei für $\deg(f) < n$ bereits bewiesen. Es ist $f = a_n X^n + a_{n-1}X^{n-1} + \ldots + a_0$, also ist

$$\tilde f = f - a_n X^{n-m} g = (a_{n-1} - a_n b_{m-1})X^{n-1} + \ldots\,;$$

der Grad von $\tilde f$ ist damit kleiner als n. Nach Induktionsvoraussetzung gibt es $\tilde q, r \in K[X]$ mit $\tilde f = \tilde q g + r$ und $\deg(r) < m$. Wir setzen $q = a_n X^{n-m} + \tilde q$; dann ist

$$f = a_n X^{n-m} g + \tilde f = a_n X^{n-m} g + \tilde q g + r = qg + r$$

wie gewünscht.

Zur Eindeutigkeit: Seien $q_1, q_2, r_1, r_2 \in K[X]$ mit $f = q_1 g + r_1 = q_2 g + r_2$ und $\deg(r_1), \deg(r_2) < \deg(g)$. Es folgt

$$(q_1 - q_2)g = r_2 - r_1\,.$$

Die rechte Seite hat Grad $< \deg(g)$. Wäre $q_1 \neq q_2$, dann hätte die linke Seite Grad $\deg(q_1 - q_2) + \deg(g) \geq \deg(g)$, ein Widerspruch. Also ist $q_1 = q_2$ und damit auch $r_1 = r_2$. ❑

Der Beweis übersetzt sich direkt in den üblichen Algorithmus zur Polynomdivision: Man subtrahiert geeignete Vielfache von g solange von f, bis man ein Polynom zurückbehält, dessen Grad kleiner als der von g ist.

15.17. **Folgerung.** *Seien K ein Körper, $p \in K[X]$ und $\lambda \in K$. Wir schreiben* **FOLG** Nullstellen

$$p = q \cdot (X - \lambda) + r$$

wie in Satz 15.16. Dann ist $r = p(\lambda)$ konstant. Insbesondere ist λ genau dann eine Nullstelle von p, wenn $r = 0$ ist.

Beweis. r ist konstant, da $\deg(r) < 1 = \deg(X - \lambda)$. Außerdem gilt

$$p(\lambda) = q(\lambda)(\lambda - \lambda) + r = r\,.$$ ❑

Ist λ eine Nullstelle von p, dann ist demnach $p = (X - \lambda) \cdot q$ mit einem Polynom $q \in K[X]$. Ist λ auch eine Nullstelle von q, dann ist $p = (X - \lambda)^2 \cdot \tilde{q}$ und so fort. Das führt zu folgender Definition.

15.18. **Definition.** Seien K ein Körper, $\mathbf{0} \neq p \in K[X]$ ein Polynom und $\lambda \in K$. Die *Vielfachheit* der Nullstelle λ von p ist die größte Zahl $n \in \mathbb{N}$, sodass man p in der Form $p = (X - \lambda)^n \cdot q$ schreiben kann mit einem Polynom $q \in K[X]$. In diesem Fall ist $q(\lambda) \neq 0$. ♢

DEF Vielfachheit einer Nullstelle

λ ist also genau dann eine Nullstelle, wenn die Vielfachheit von λ als Nullstelle positiv ist.

15.19. **Beispiele.** Wir betrachten $K = \mathbb{R}$. Für $p = X^3 - X^2 - X + 1 \in \mathbb{R}[X]$ gilt

$$p = (X - 1)^2(X + 1)\,,$$

also hat p die Nullstellen 1 (mit Vielfachheit 2: eine „doppelte“ Nullstelle) und -1 (mit Vielfachheit 1: eine „einfache“ Nullstelle).

BSP Vielfachheiten

Für $q = X^3 + X^2 + X + 1 \in \mathbb{R}[X]$ gilt dagegen

$$q = (X + 1)(X^2 + 1)\,,$$

also hat q nur die (einfache) Nullstelle -1 in $\mathbb{R}$, denn der zweite Faktor $X^2 + 1$ nimmt nur positive Werte an und hat daher keine reelle Nullstelle. Wenn wir q aber als Polynom in $\mathbb{C}[X]$ betrachten, dann haben wir

$$q = (X + 1)(X + \boldsymbol{i})(X - \boldsymbol{i})\,;$$

q hat also die drei komplexen (einfachen) Nullstellen -1, $\boldsymbol{i}$ und $-\boldsymbol{i}$. ♣

15.20. **Beispiel.** Wie kann man die Nullstellen eines Polynoms p finden? Das ist einfach im Fall $\deg(p) = 1$, und im Fall $\deg(p) = 2$ kann man die Lösungsformel für quadratische Gleichungen verwenden. Für allgemeine Polynome ist die Bestimmung der Nullstellen allerdings ein schwieriges Problem. In einfach gelagerten Fällen (wie üblicherweise etwa bei Klausuraufgaben) kann man häufig eine Nullstelle α durch Probieren erraten. Dann kann man p durch $X - \alpha$ teilen, um ein Polynom kleineren Grades zu erhalten, mit dem man dann weitermacht.

BSP Nullstellen

Als Beispiel betrachten wir

$$p = X^4 + X^3 - 6X^2 - 2X + 4 \in \mathbb{R}[X]\,.$$

Wir probieren, ob 0, 1, -1 Nullstellen sind, und finden $p(-1) = 0$. Division von p durch $X + 1$ ergibt

$$p = (X + 1)(X^3 - 6X + 4)\,,$$

und wir müssen noch die Nullstellen von $p_1 = X^3 - 6X + 4$ finden. Weiteres Probieren liefert die Nullstelle 2. Division von p_1 durch $X - 2$ ergibt

$$p_1 = (X - 2)(X^2 + 2X - 2)\,.$$

Auf den verbleibenden Faktor können wir jetzt die Lösungsformel für quadratische Gleichungen anwenden; sie liefert die weiteren Nullstellen

$$\frac{-2 \pm \sqrt{2^2 - 4 \cdot (-2)}}{2} = -1 \pm \sqrt{3}\,.$$ ♣

Zurück zu Eigenwerten und Eigenräumen. Wir haben gesehen, dass der Ausdruck

$$\det(\lambda I_n - A)$$

darüber entscheidet, ob λ ein Eigenwert von A ist oder nicht. Ist $A = (a_{ij})$, dann hat diese Determinante die folgende Form:

$$\begin{vmatrix} \lambda - a_{11} & -a_{12} & -a_{13} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & -a_{23} & \cdots & -a_{2n} \\ -a_{31} & -a_{32} & \lambda - a_{33} & \cdots & -a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & -a_{n3} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

Wenn wir das in die Leibniz-Formel einsetzen, dann bekommen wir

$$\begin{aligned} &(\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn}) + \text{Terme mit} \leq n-2 \text{ Faktoren } \lambda - a_{jj} \\ &\qquad = \lambda^n - (a_{11} + \ldots + a_{nn})\lambda^{n-1} + \ldots + (-1)^n \det(A)\,. \end{aligned}$$

Das hat die Form $p(\lambda)$ mit einem normierten Polynom $p \in K[X]$ vom Grad n.

* 15.21. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Das Polynom $\chi_A = \det(XI_n - A) \in K[X]$ heißt das *charakteristische Polynom* von A. ♢

DEF Charakteristisches Polynom

Die Leibniz-Formel zeigt, dass man die Determinante allgemeiner für Matrizen mit Einträgen in einem kommutativen Ring R definieren kann; die Determinante ist dann ebenfalls ein Element von R. Die Multiplikativität und die Symmetrie der Determinante gelten auch in diesem erweiterten Kontext. In der obigen Definition wenden wir das für den Ring $K[X]$ und die Matrix $XI_n - A \in \mathrm{Mat}(n, K[X])$ an.

In der Literatur findet man auch häufig die Definition $\det(A - XI_n)$ für das charakteristische Polynom von A. Diese Definition unterscheidet sich von der hier gegebenen nur durch einen Faktor $(-1)^n$. Das ändert nichts an den Nullstellen, sodass es für den Zweck der Eigenwert-Berechnung oder für die Definition der algebraischen Vielfachheit (siehe unten) keinen Unterschied macht. Der Nachteil der anderen Variante ist aus meiner Sicht, dass das Polynom für ungerades n dann nicht normiert ist (sondern Leitkoeffizient -1 hat).

Wir haben gesehen, dass die Eigenwerte von A genau die Nullstellen des charakteristischen Polynoms von A sind.

15.22. **Beispiel.** Was sind die Eigenwerte der Matrix

BSP Eigenwerte

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \in \mathrm{Mat}(3, \mathbb{R})\ ?$$

Wir bestimmen das charakteristische Polynom:

$$\begin{aligned} \det(XI_3 - A) &= \begin{vmatrix} X-1 & 0 & 0 \\ -1 & X-1 & -1 \\ -1 & -2 & X-3 \end{vmatrix} \\ &= (X-1)\begin{vmatrix} X-1 & -1 \\ -2 & X-3 \end{vmatrix} \\ &= (X-1)\big((X-1)(X-3) - 1 \cdot 2\big) \\ &= (X-1)(X^2 - 4X + 1) \end{aligned}$$

Ein Eigenwert ist $\lambda_1 = 1$, die anderen beiden finden wir mit Hilfe der Lösungsformel für quadratische Gleichungen:

$$\lambda_2 = 2 + \sqrt{3} \qquad \text{und} \qquad \lambda_3 = 2 - \sqrt{3}\,. \qquad \clubsuit$$

Wir können die Definitionen von Determinante und charakteristischem Polynom auch auf Endomorphismen übertragen.

DEF Determinante, char. Pol. von Endomorphismen

15.23. **Definition.** Seien K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}(V)$. Sei B eine beliebige Basis von V und sei $A = \mathrm{Mat}_{B,B}(f)$. Dann ist die *Determinante* von f definiert als $\det(f) = \det(A)$ und das *charakteristische Polynom* χ_f von f ist das charakteristische Polynom von A. ♢

Die Definition ist sinnvoll, weil sie nicht von der Wahl der Basis B abhängt: Ist $A' = \mathrm{Mat}_{B',B'}(f)$ mit einer anderen Basis B' von V, dann gibt es eine Matrix $P \in \mathrm{GL}(n, K)$ (wenn $\dim V = n$), sodass $A' = PAP^{-1}$ ist. Dann ist

$$\begin{aligned}\det(A') &= \det(PAP^{-1}) = \det(P)\det(A)\det(P^{-1}) \\ &= \det(A)\det(PP^{-1}) = \det(A)\det(I_n) = \det(A)\,.\end{aligned}$$

Es ist auch $P(XI_n - A)P^{-1} = XPI_nP^{-1} - PAP^{-1} = XI_n - A'$, und die gleiche Rechnung wie eben zeigt, dass A und A' dasselbe charakteristische Polynom haben.

DEF algebraische Vielfachheit

∗ 15.24. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$, $A \in \mathrm{Mat}(n, K)$ und $\lambda \in K$. Die *algebraische Vielfachheit* von λ als Eigenwert von A ist die Vielfachheit von λ als Nullstelle des charakteristischen Polynoms von A. Entsprechend definieren wir die *algebraische Vielfachheit* von λ als Eigenwert eines Endomorphismus f eines endlich-dimensionalen K-Vektorraums V. ♢

Wir haben jetzt also zwei Vielfachheiten von Eigenwerten definiert, die geometrische und die algebraische. In welcher Beziehung stehen sie zueinander? Wir wissen bisher Folgendes:

$$\text{geom. Vielfachheit} > 0 \iff \text{Eigenwert} \iff \text{alg. Vielfachheit} > 0$$

Müssen die beiden Vielfachheiten immer gleich sein?

BSP alg. ≠ geom. Vielfachheit

15.25. **Beispiel.** Sei

$$A = \begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix} \in \mathrm{Mat}(2, K)\,.$$

Das charakteristische Polynom von A ist $(X - \alpha)^2$, also hat α die algebraische Vielfachheit 2. Auf der anderen Seite ist $E_\alpha(A) = \langle (1, 0) \rangle$ (denn $\mathrm{rk}(\alpha I_2 - A) = 1$), also hat α die geometrische Vielfachheit 1. ♣

Die Vielfachheiten können also verschieden sein. Eine Beziehung gilt jedoch.

SATZ geom. ≤ alg. Vielfachheit

∗ 15.26. **Satz.** *Seien K ein Körper, V ein endlich-dimensionaler K-Vektorraum, $f \in \mathrm{End}(V)$ und $\lambda \in K$. Dann ist die geometrische Vielfachheit von λ als Eigenwert von f nicht größer als seine algebraische Vielfachheit.*

Die analoge Aussage gilt dann natürlich auch für Matrizen $A \in \mathrm{Mat}(n, K)$.

Beweis. Sei $m = \dim E_\lambda(f)$ die geometrische Vielfachheit, sei $n = \dim V$ und sei $(b_1, b_2, \ldots, b_m)$ eine Basis von $E_\lambda(f)$. Wir können diese Basis zu einer Basis $B = (b_1, b_2, \ldots, b_n)$ von V erweitern. Dann ist

$$A = \mathrm{Mat}_{B,B}(f) = \left(\begin{array}{c|c} \lambda I_m & D \\ \hline \mathbf{0}_{n-m,m} & C \end{array}\right)$$

mit Matrizen $D \in \mathrm{Mat}(m \times (n-m), K)$ und $C \in \mathrm{Mat}(n-m, K)$, denn für $j \in \{1, 2, \ldots, m\}$ ist $f(b_j) = \lambda b_j$; in der j-ten Spalte von A kommt also das λ-fache des j-ten Standard-Basisvektors zu stehen. Das charakteristische Polynom von f ist dann

$$\det(XI_n - A) = \det\left(\begin{array}{c|c} (X-\lambda)I_m & -D \\ \hline \mathbf{0}_{n-m,m} & XI_{n-m} - C \end{array}\right) = (X-\lambda)^m \det(XI_{n-m} - C)\,.$$

Dabei haben wir die Determinante m-mal jeweils nach der ersten Spalte entwickelt. Das zeigt, dass die Vielfachheit von λ als Nullstelle des charakteristischen Polynoms mindestens m ist. ❑

Man kann die obige Formel auf allgemeinere Blockmatrizen ausdehnen (Beweis als Übung): Sind $A \in \mathrm{Mat}(m, K)$, $B \in \mathrm{Mat}(m \times n, K)$ und $C \in \mathrm{Mat}(n, K)$, dann ist

$$\det\left(\begin{array}{c|c} A & B \\ \hline \mathbf{0}_{n,m} & C \end{array}\right) = \det(A)\det(C)\,.$$

16. Diagonalisierbarkeit und der Satz von Cayley-Hamilton

Wir haben in Folgerung 15.7 gesehen, dass die Summe der geometrischen Vielfachheiten der Eigenwerte eines Endomorphismus f eines n-dimensionalen Vektorraums (oder einer $n \times n$-Matrix) höchstens n ist. Das macht den Fall interessant, in dem diese Schranke erreicht wird. Wir formulieren zunächst eine Definition und werden dann sehen, was sie mit dieser Frage zu tun hat.

* 16.1. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$. Eine Matrix $A = (a_{ij}) \in \mathrm{Mat}(n, K)$ ist eine *Diagonalmatrix* oder *diagonal*, wenn $a_{ij} = 0$ ist für alle $i, j \in \{1, 2, \ldots, n\}$ mit $i \neq j$. Ist $a_{ii} = d_i$ für $i \in \{1, 2, \ldots, n\}$, dann schreiben wir $\mathrm{diag}(d_1, d_2, \ldots, d_n)$ für A: **DEF** Diagonalmatrix

$$\mathrm{diag}(d_1, d_2, \ldots, d_n) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}. \qquad \diamondsuit$$

16.2. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$, V ein n-dimensionaler K-Vektorraum und $f \in \mathrm{End}(V)$. Dann sind die folgenden Aussagen äquivalent:* **LEMMA** Diagonalisierbarkeit

(1) *V hat eine Basis B, die aus Eigenvektoren von f besteht.*

(2) *Die Summe der geometrischen Vielfachheiten der Eigenwerte von f ist n.*

(3) *Sei $A = \mathrm{Mat}_{B',B'}(f)$ die Matrix von f bezüglich einer beliebigen Basis B' von V. Dann ist A ähnlich zu einer Diagonalmatrix.*

Beweis.

„(1) $\Rightarrow$ (2)“: Seien $\lambda_1, \lambda_2, \ldots, \lambda_k \in K$ die (paarweise verschiedenen) Eigenwerte von f, sei $n_j = \dim E_{\lambda_j}(f)$ die geometrische Vielfachheit von λ_j und sei m_j die Anzahl der Basisvektoren in B, die Eigenvektoren zum Eigenwert λ_j sind. Dann haben wir jeweils m_j linear unabhängige Vektoren in $E_{\lambda_j}(f)$, also ist $m_j \leq n_j$. Auf der anderen Seite ist

$$n = \#B = m_1 + m_2 + \ldots + m_k \leq n_1 + n_2 + \ldots + n_k \leq n\,,$$

also haben wir Gleichheit; insbesondere ist $n_1 + n_2 + \ldots + n_k = n$.

„(2) $\Rightarrow$ (1)“: Wir behalten die Bezeichnungen bei. Sei B_j eine Basis von $E_{\lambda_j}(f)$. Nach Folgerung 15.7 bilden die B_j zusammen eine linear unabhängige Familie B in V. Wegen

$$\#B = \#B_1 + \#B_2 + \ldots + \#B_k = n_1 + n_2 + \ldots + n_k = n = \dim V$$

ist B dann eine Basis von V, die nach Konstruktion aus Eigenvektoren von f besteht.

„(1) $\Rightarrow$ (3)“: Sei $B = (b_1, b_2, \ldots, b_n)$ und sei α_j der Eigenwert von f, sodass $f(b_j) = \alpha_j b_j$. Dann ist

$$\mathrm{Mat}_{B,B}(f) = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$$

eine Diagonalmatrix, und nach Satz 15.1 ist A zu ihr ähnlich.

„(3) $\Rightarrow$ (1)“: Ist A ähnlich zu $D = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$, dann gibt es nach Satz 15.1 eine Basis $B = (b_1, b_2, \ldots, b_n)$ von V, sodass $D = \mathrm{Mat}_{B,B}(f)$ ist. Daran liest man ab, dass $f(b_j) = \alpha_j b_j$ ist (und $b_j \neq \mathbf{0}$), also besteht B aus Eigenvektoren von f. ❑

* 16.3. **Definition.** Sei f ein Endomorphismus eines endlich-dimensionalen K-Vektorraums V. Dann heißt f *diagonalisierbar*, wenn V eine Basis hat, die aus Eigenvektoren von f besteht. **DEF** diagonalisierbar

Eine Matrix $A \in \mathrm{Mat}(n, K)$ heißt *diagonalisierbar*, wenn sie zu einer Diagonalmatrix ähnlich ist, wenn es also eine Diagonalmatrix $D \in \mathrm{Mat}(n, K)$ und eine invertierbare Matrix $P \in \mathrm{GL}(n, K)$ gibt mit $PAP^{-1} = D$. ♢

Aus dem Lemma ergibt sich, dass ein Endomorphismus genau dann diagonalisierbar ist, wenn die zugeordnete Matrix (bezüglich irgendeiner Basis) diagonalisierbar ist.

Aus dem Beweis ergibt sich auch, dass die Einträge auf der Diagonalen der Diagonalmatrix gerade die Eigenwerte sind; sie kommen so oft vor, wie es ihrer geometrischen Vielfachheit entspricht.

Da die geometrische Vielfachheit eines Eigenwerts höchstens so groß ist wie seine algebraische Vielfachheit, ist eine *notwendige* Bedingung für die Diagonalisierbarkeit, dass die Summe der *algebraischen* Vielfachheiten der Eigenwerte n ist. Das ist eine Eigenschaft des charakteristischen Polynoms.

16.4. **Definition.** Sei K ein Körper und $p \in K[X]$ ein normiertes Polynom vom Grad n. Wir sagen, p *zerfällt in Linearfaktoren* über K, wenn es $\alpha_1, \alpha_2, \ldots, \alpha_n \in K$ gibt, sodass **DEF** Zerlegung in Linearfaktoren

$$p = \prod_{j=1}^{n} (X - \alpha_j) = (X - \alpha_1)(X - \alpha_2) \cdots (X - \alpha_n)$$

ist. ♢

Das bedeutet also, dass die Summe der Vielfachheiten der Nullstellen von p in K gleich dem Grad n von p ist.

* 16.5. **Folgerung.** *Seien K ein Körper und f ein Endomorphismus eines endlich-dimensionalen K-Vektorraums V. Dann sind die folgenden Aussagen äquivalent:* **FOLG** Charakterisierung von Diagonalisierbarkeit

(1) *f ist diagonalisierbar.*

(2) *Das charakteristische Polynom von f zerfällt in Linearfaktoren über K und für jeden Eigenwert von f stimmt die geometrische mit der algebraischen Vielfachheit überein.*

Insbesondere gilt: Hat f genau $n = \dim V$ verschiedene Eigenwerte, dann ist f diagonalisierbar.

Wichtig: Die Umkehrung des letzten Satzes gilt nicht. Zum Beispiel sind die Einheitsmatrix oder die Nullmatrix diagonalisierbar (weil schon diagonal), sie haben aber jeweils nur einen Eigenwert (nämlich 1 bzw. 0).

Beweis. Seien $\lambda_1, \lambda_2, \ldots, \lambda_k \in K$ die verschiedenen Eigenwerte von f und seien n_j bzw. m_j ihre geometrischen bzw. algebraischen Vielfachheiten. Sei $n = \dim V$. Dann gilt wegen $n_j \leq m_j$ (Satz 15.26) und $m_1 + m_2 + \ldots + m_k \leq \deg(\chi_f) = n$:

$$\begin{aligned} f \text{ diagonalisierbar} &\iff n_1 + n_2 + \ldots + n_k = n \\ &\iff m_1 + m_2 + \ldots + m_k = n \quad \text{und} \quad \forall j \colon m_j = n_j \\ &\iff \chi_f \text{ zerfällt in Linearfaktoren über } K \text{ und } \forall j \colon m_j = n_j. \end{aligned}$$

Hat f n verschiedene Eigenwerte, dann ist $k = n$ und $m_j = 1$. Dann muss auch die geometrische Vielfachheit $n_j = 1$ sein, also ist $n_1 + n_2 + \ldots + n_k = n$. ❑

16.6. **Beispiel.** Die Matrix aus Beispiel 15.22 ist diagonalisierbar, weil sie die drei verschiedenen Eigenwerte 1, $2 + \sqrt{3}$ und $2 - \sqrt{3}$ hat. ♣

BSP diagonalisierbare Matrix

16.7. **Beispiel.** Eine Matrix muss nicht diagonalisierbar sein, wenn ihr charakteristisches Polynom in Linearfaktoren zerfällt. Das hatten wir (siehe Beispiel 15.25) an Hand der Matrix

$$\begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix}$$

gesehen, deren charakteristisches Polynom $(X - \alpha)^2$ in Linearfaktoren zerfällt, für die aber die geometrische Vielfachheit von α (nämlich 1) kleiner ist als die algebraische Vielfachheit (nämlich 2). ♣

BSP Nicht diagonalisierbare Matrix

Die Gleichheit der geometrischen und algebraischen Vielfachheit ist also eine wesentliche Bedingung. Die Bedingung, dass das charakteristische Polynom in Linearfaktoren zerfällt, können wir hingegen erfüllen, wenn unser Körper „groß genug“ ist. Wir erinnern uns daran (Satz 4.4), dass der Körper $\mathbb{C}$ der komplexen Zahlen *algebraisch abgeschlossen* ist. Das bedeutet, dass jedes nicht-konstante Polynom $p \in \mathbb{C}[X]$ eine Nullstelle in $\mathbb{C}$ hat. Daraus folgt, dass jedes normierte Polynom über $\mathbb{C}$ in Linearfaktoren zerfällt:

16.8. **Folgerung.** *Sei $p \in \mathbb{C}[X]$ normiert. Dann zerfällt p in Linearfaktoren über $\mathbb{C}$.*

FOLG Faktorisierung von Polynomen über $\mathbb{C}$

Beweis. Wir beweisen die Aussage durch Induktion über den Grad von p. Im Fall $\deg(p) = 0$ ist $p = 1$ und damit gleich dem leeren Produkt (anders ausgedrückt, p hat genau $\deg(p) = 0$ Nullstellen in $\mathbb{C}$), also zerfällt p trivialerweise in Linearfaktoren. Die Aussage gelte für Polynome vom Grad n, und p habe Grad $n + 1$. Nach dem Fundamentalsatz der Algebra 4.4 hat p eine Nullstelle $\alpha_1 \in \mathbb{C}$. Dann ist $p = (X - \alpha_1)q$ mit $q \in \mathbb{C}[X]$ und $\deg(q) = n$. Nach der Induktionsvoraussetzung zerfällt q in Linearfaktoren:

$$q = (X - \alpha_2)(X - \alpha_3) \cdots (X - \alpha_{n+1}) \,,$$

also gilt das auch für p:

$$p = (X - \alpha_1)q = (X - \alpha_1)(X - \alpha_2) \cdots (X - \alpha_{n+1}) \,. \qquad ❑$$

Für den Beweis haben wir nur verwendet, dass $\mathbb{C}$ algebraisch abgeschlossen ist; die Aussage gilt also entsprechend für jeden algebraisch abgeschlossenen Körper. Außerdem kann man ganz allgemein zeigen, dass es zu jedem Körper K einen algebraisch abgeschlossenen Körper $\bar{K}$ gibt, der K als Teilkörper enthält (also sodass die Addition und Multiplikation in K die Einschränkungen derjenigen von $\bar{K}$ sind). Durch Übergang von K zu $\bar{K}$ kann man dann also immer erreichen, dass das charakteristische Polynom einer Matrix (oder eines Endomorphismus) in Linearfaktoren zerfällt. Die Bedingung an die Gleichheit der geometrischen und algebraischen Vielfachheiten der Eigenwerte ist also die eigentlich entscheidende für die Diagonalisierbarkeit.

Wir erinnern uns daran, dass das charakteristische Polynom χ_A einer Matrix $A = (a_{ij}) \in \mathrm{Mat}(n, K)$ die Form

$$\chi_A = X^n - (a_{11} + a_{22} + \ldots + a_{nn})X^{n-1} + \ldots + (-1)^n \det(A)$$

hat; der Koeffizient von X^0 ergibt sich dabei aus

$$\chi_A(0) = \det(0 \cdot I_n - A) = \det(-A) = (-1)^n \det(A)\,.$$

Wenn χ_A in Linearfaktoren zerfällt:

$$\begin{aligned}\chi_A &= (X-\lambda_1)(X-\lambda_2)\cdots(X-\lambda_n)\\ &= X^n - (\lambda_1+\lambda_2+\ldots+\lambda_n)X^{n-1} + \cdots + (-1)^n\lambda_1\lambda_2\cdots\lambda_n\,,\end{aligned}$$

dann sehen wir durch Vergleich der beiden Darstellungen von χ_A, dass wir Summe und Produkt der Eigenwerte einfach von der Matrix ablesen können. Bevor wir das als Lemma formulieren, ist hier noch eine Definition:

DEF Spur einer Matrix

16.9. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$ und $A = (a_{ij}) \in \mathrm{Mat}(n, K)$. Die *Spur* von A ist

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \ldots + a_{nn}\,.$$

(„Tr“ von englisch *trace*.) ♢

LEMMA Spur und Determinante durch Eigenwerte

16.10. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Das charakteristische Polynom von A zerfalle in Linearfaktoren. Dann ist* $\mathrm{Tr}(A)$ *die Summe und* $\det(A)$ *das Produkt der Eigenwerte von A, jeweils entsprechend ihrer algebraischen Vielfachheit gezählt.*

Da die charakteristischen Polynome ähnlicher Matrizen gleich sind, folgt

$$\mathrm{Tr}(PAP^{-1}) = \mathrm{Tr}(A)\,.$$

Die Spur erfüllt aber sogar noch eine etwas stärkere Aussage.

* **SATZ** Eigensch. der Spur

16.11. **Satz.** *Die Spur ist eine K-lineare Abbildung* $\mathrm{Mat}(n, K) \to K$. *Für alle $A \in \mathrm{Mat}(n, K)$ gilt* $\mathrm{Tr}(A^\top) = \mathrm{Tr}(A)$.

Sind $A \in \mathrm{Mat}(m \times n, K)$ und $B \in \mathrm{Mat}(n \times m, K)$, dann gilt

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA)\,.$$

Man beachte, dass AB eine $m \times m$-Matrix und BA eine $n \times n$-Matrix ist.

Beweis. Die erste Aussage ist klar (die Spur ist eine Linearkombination der Matrix-Einträge), die zweite ebenfalls, da $A^\top$ und A dieselben Diagonaleinträge haben. Für die dritte Aussage seien $A = (a_{ij})_{1\le i\le m, 1\le j\le n}$ und $B = (b_{kl})_{1\le k\le n, 1\le l\le m}$. Der Diagonaleintrag in der i-ten Zeile und Spalte von $C = AB$ ist

$$c_{ii} = \sum_{j=1}^{n} a_{ij}b_{ji} \qquad (i \in \{1, 2, \ldots, m\})$$

und der Diagonaleintrag in der j-ten Zeile und Spalte von $C' = BA$ ist

$$c'_{jj} = \sum_{i=1}^{m} b_{ji}a_{ij} \qquad (j \in \{1, 2, \ldots, n\})\,.$$

Es folgt

$$\begin{aligned}\operatorname{Tr}(AB) = \operatorname{Tr}(C) &= \sum_{i=1}^{m} c_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji} \\ &= \sum_{j=1}^{n}\sum_{i=1}^{m} b_{ji}a_{ij} = \sum_{j=1}^{n} c'_{jj} = \operatorname{Tr}(C') = \operatorname{Tr}(BA)\,. \end{aligned}$$ ❑

Die Gleichung $\operatorname{Tr}(A) = \operatorname{Tr}(PAP^{-1})$ folgt daraus:

$$\operatorname{Tr}(PAP^{-1}) = \operatorname{Tr}\big(P(AP^{-1})\big) = \operatorname{Tr}\big((AP^{-1})P\big) = \operatorname{Tr}\big(A(P^{-1}P)\big) = \operatorname{Tr}(A)\,.$$

Analog zu Definition 15.23 können wir daher auch die Spur eines Endomorphismus definieren.

DEF Spur eines Endomorphismus

16.12. **Definition.** Seien K ein Körper, V ein endlich-dimensionaler K-Vektorraum und $f \in \operatorname{End}(V)$. Seien B eine beliebige Basis von V und $A = \operatorname{Mat}_{B,B}(f)$. Dann ist die *Spur* von f definiert als $\operatorname{Tr}(f) = \operatorname{Tr}(A)$. ♢

Die Aussage, dass Spur und Determinante die Summe und das Produkt der Eigenwerte sind (mit algebraischer Vielfachheit gezählt), gilt dann entsprechend auch für Endomorphismen.

Nun wollen wir uns noch überlegen, wie man, wenn die Matrix $A \in \operatorname{Mat}(n, K)$ diagonalisierbar ist, eine Matrix $P \in \operatorname{GL}(n, K)$ findet, die A diagonalisiert, also sodass $PAP^{-1} = D$ eine Diagonalmatrix ist.

LEMMA Diagonalisierung

16.13. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \operatorname{Mat}(n, K)$ eine diagonalisierbare Matrix. Sei $(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$ eine Basis von K^n, die aus Eigenvektoren von A besteht, mit $A\boldsymbol{b}_j = \lambda_j \boldsymbol{b}_j$ (wir betrachten $\boldsymbol{b}_j$ als Spaltenvektor). Sei $Q \in \operatorname{GL}(n, K)$ die Matrix, deren j-te Spalte $\boldsymbol{b}_j$ ist für $j \in \{1, 2, \ldots, n\}$. Dann ist*

$$Q^{-1}AQ = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

eine Diagonalmatrix.

Man kann dann also $P = Q^{-1}$ nehmen.

Beweis. Sei $\mathbf{e}_j$ der j-te Standard-Basisvektor von K^n als Spaltenvektor. Dann gilt $Q\mathbf{e}_j = \boldsymbol{b}_j$ und damit auch $Q^{-1}\boldsymbol{b}_j = \mathbf{e}_j$. Es folgt

$$Q^{-1}AQ\mathbf{e}_j = Q^{-1}A\boldsymbol{b}_j = \lambda_j Q^{-1}\boldsymbol{b}_j = \lambda_j \mathbf{e}_j\,.$$

Das zeigt, dass die j-te Spalte von $Q^{-1}AQ$ gerade $\lambda_j\mathbf{e}_j$ ist, also ist

$$Q^{-1}AQ = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)\,.$$ ❑

BSP Diagonalisieren einer Matrix

16.14. **Beispiel.** Sei $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) \in \operatorname{Mat}(2, \mathbb{R})$. Dann hat A die beiden Eigenwerte 1 und -1 und ist daher diagonalisierbar. Wir hatten in Beispiel 15.10 Basen der beiden Eigenräume gefunden:

$$E_1(A) = \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle \quad \text{und} \quad E_{-1}(A) = \left\langle \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\rangle.$$

Eine geeignete Matrix Q ist demnach $Q = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$; mit $P = Q^{-1} = \frac{1}{2}\left(\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right)$ ist dann $PAP^{-1} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. ♣

Wir haben gesehen, dass man Elemente des zu Grunde liegenden Körpers in ein Polynom einsetzen kann (Definition 15.12). Man kann aber auch allgemeinere Objekte in Polynome einsetzen, zum Beispiel Matrizen oder Endomorphismen.

DEF Einsetzen von Matrizen und Endomorphismen in Polynome

16.15. **Definition.** Sei K ein Körper, sei $n \in \mathbb{N}$ und seien $p \in K[X]$ ein Polynom und $A \in \mathrm{Mat}(n, K)$. Wie üblich setzen wir $A^0 = I_n$ (das Einselement des Matrizenrings $\mathrm{Mat}(n, K)$) und $A^{k+1} = A \cdot A^k$ für $k \in \mathbb{N}$. Ist

$$p = a_0 + a_1 X + a_2 X^2 + \ldots + a_m X^m \,,$$

dann definieren wir

$$p(A) = \sum_{j=0}^{m} a_j A^j = a_0 I_n + a_1 A + a_2 A^2 + \ldots + a_m A^m \in \mathrm{Mat}(n, K) \,.$$

Ist V ein K-Vektorraum und f ein Endomorphismus von V, dann setzen wir analog $f^{\circ 0} = \mathrm{id}_V$ und $f^{\circ(k+1)} = f \circ f^{\circ k}$ und definieren

$$p(f) = \sum_{j=0}^{m} a_j f^{\circ j} = a_0 \,\mathrm{id}_V + a_1 f + a_2 f^{\circ 2} + \ldots + a_m f^{\circ m} \in \mathrm{End}(V) \,. \qquad \diamondsuit$$

(Die Schreibweise $f^{\circ k}$ anstatt von f^k verwenden wir, um Verwechslungen mit der punktweise definierten Potenz von (z.B.) reellen Funktionen zu vermeiden.)

Dann ist Folgendes klar: Ist A die Matrix von f bezüglich einer Basis B von V, dann ist $p(A)$ die Matrix von $p(f)$ bezüglich B.

Wir zeigen noch, dass die Abbildungen $K[X] \to \mathrm{Mat}(n, K)$, $p \mapsto p(A)$, und $K[X] \mapsto \mathrm{End}(V)$, $p \mapsto p(f)$, schöne Eigenschaften haben.

LEMMA Einsetzungshomomorphismus

16.16. **Lemma.** *In der Situation von Definition 16.15 gilt für $p, q \in K[X]$:*

$$(p + q)(A) = p(A) + q(A) \quad \textit{und} \quad (p \cdot q)(A) = p(A) \cdot q(A)$$

bzw.

$$(p + q)(f) = p(f) + q(f) \quad \textit{und} \quad (p \cdot q)(f) = p(f) \circ q(f) \,.$$

Außerdem ist $1(A) = I_n$ und $1(f) = \mathrm{id}_V$, wobei 1 das konstante Polynom 1 bezeichnet.

Beweis. Das folgt aus den Rechenregeln für Matrizen bzw. Endomorphismen (also daraus, dass $\mathrm{Mat}(n, K)$ und $\mathrm{End}(V)$ Ringe sind), zusammen mit

$$\lambda A = (\lambda I_n) A = A(\lambda I_n) \quad \text{und} \quad \lambda f = (\lambda \,\mathrm{id}_V) \circ f = f \circ (\lambda \,\mathrm{id}_V) \,. \qquad \square$$

Sind R und R' zwei Ringe, dann heißt eine Abbildung $\phi\colon R \to R'$ ein *Ringhomomorphismus*, wenn $\phi(1_R) = 1_{R'}$ und für alle $r_1, r_2 \in R$ gilt $\phi(r_1 + r_2) = \phi(r_1) + \phi(r_2)$ und $\phi(r_1 r_2) = \phi(r_1)\phi(r_2)$ (d.h., ϕ bildet die Einselemente aufeinander ab und ist mit Addition und Multiplikation verträglich). Die Abbildungen $K[X] \to \mathrm{Mat}(n, R)$, $p \mapsto p(A)$, und $K[X] \to \mathrm{End}(V)$, $p \mapsto p(f)$, sind also Ringhomomorphismen. Weil sie dadurch gegeben sind, dass man etwas in die Polynome einsetzt, heißen sie auch *Einsetzungshomomorphismen.*

Mehr über Ringe im Allgemeinen und Polynomringe im Besonderen gibt es in der „Einführung in die Zahlentheorie und algebraische Strukturen“.

16.17. **Beispiel.** Seien $A \in \mathrm{Mat}(n,K)$ und $\lambda \in K$, seien weiter $v \in E_\lambda(A)$ und $p \in K[X]$. Dann gilt

BSP
$Av = \lambda v \Longrightarrow$
$p(A)v = p(\lambda)v$

$$p(A) \cdot v = p(\lambda)v \ :$$

Ist $p = a_0 + a_1X + a_2X^2 + \ldots + a_mX^m$, dann rechnet man

$$\begin{aligned} p(A) \cdot v &= (a_0 I_n + a_1 A + a_2 A^2 + \ldots + a_m A^m) \cdot v \\ &= a_0 v + a_1 A \cdot v + a_2 A^2 \cdot v + \ldots + a_m A^m \cdot v \\ &= a_0 v + a_1 \lambda v + a_2 \lambda^2 v + \ldots + a_m \lambda^m v \\ &= (a_0 + a_1\lambda + a_2\lambda^2 + \ldots + a_m\lambda^m)v \\ &= p(\lambda)v\,. \end{aligned}$$

♣

Sei eine Matrix $A \in \mathrm{Mat}(n,K)$ gegeben. Dann kann man sich fragen, ob es stets ein Polynom $p \in K[X]$ gibt mit $p(A) = \mathbf{0}$ (und $p \neq 0$), bzw. was man über die Menge solcher Polynome mit „Nullstelle“ A aussagen kann.

Die erste Frage kann man leicht beantworten.

16.18. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n,K)$. Dann gibt es ein normiertes Polynom $p \in K[X]$ mit $\deg(p) \leq n^2$, sodass $p(A) = \mathbf{0}$ ist.*

LEMMA
Existenz von $p \in K[X]$ mit $p(A) = \mathbf{0}$

Beweis. Der Beweis ist eine schöne Anwendung grundlegender Resultate der Linearen Algebra. $\mathrm{Mat}(n,K)$ ist ein K-Vektorraum der Dimension n^2, also müssen die n^2+1 Elemente $A^0, A^1, A^2, \ldots, A^{n^2} \in \mathrm{Mat}(n,K)$ linear abhängig sein. Es gibt also $\lambda_0, \lambda_1, \ldots, \lambda_{n^2} \in K$, nicht alle null, mit $\sum_{j=0}^{n^2} \lambda_j A^j = \mathbf{0}$. Wir setzen $m = \max\{j \mid \lambda_j \neq 0\}$. Nach eventueller Multiplikation mit λ_m^{-1} können wir $\lambda_m = 1$ annehmen. Die Behauptung folgt dann mit $p = \sum_{j=0}^{m} \lambda_j X^j$. ❑

Man beachte, dass der Satz, dass ein Polynom vom Grad n höchstens n Nullstellen hat, nicht für Matrizen gilt. Zum Beispiel gilt für jede Matrix $A = \left(\begin{smallmatrix} a & b \\ c & -a \end{smallmatrix}\right)$ mit $a^2 + bc = -1$, dass $A^2 + I_2 = \mathbf{0}$ ist. Alle diese Matrizen (und davon gibt es unendlich viele, wenn K unendlich ist) sind also „Nullstellen“ von $X^2 + 1$.

Die zweite Frage kann man wie folgt beantworten:

16.19. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n,K)$. Sei*

LEMMA
Struktur von $P(A)$

$$P(A) = \{p \in K[X] \mid p(A) = \mathbf{0}\}\,.$$

Dann gilt:

(1) $\mathbf{0} \in P(A)$.

(2) *Aus $p, q \in P(A)$ folgt $p + q \in P(A)$.*

(3) *Aus $p \in P(A)$, $q \in K[X]$ folgt $q \cdot p \in P(A)$.*

Beweis. Die erste Aussage ist klar. Die beiden anderen sieht man so:

$$(p+q)(A) = p(A)+q(A) = \mathbf{0}+\mathbf{0} = \mathbf{0} \quad \text{und} \quad (q\cdot p)(A) = q(A)\cdot p(A) = q(A)\cdot\mathbf{0} = \mathbf{0}\,.$$

❑

Sei R ein kommutativer Ring. Eine Teilmenge $I \subset R$ heißt ein *Ideal* von R, wenn I die obigen Eigenschaften hat:

$$0 \in I\,, \quad r, r' \in I \Rightarrow r + r' \in I\,, \quad r \in R, r' \in I \Rightarrow rr' \in I\,.$$

Ist $\phi : R \to R'$ ein Ringhomomorphismus, dann zeigt derselbe Beweis wie oben, dass sein *Kern* $\ker(\phi) = \{r \in R \mid \phi(r) = 0\}$ ein Ideal von R ist.

Die Tatsache, dass es im Polynomring eine „Division mit Rest“ gibt, führt zu einer einfachen Beschreibung von $P(A)$.

SATZ Minimalpolynom

16.20. **Satz.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Dann gibt es ein eindeutig bestimmtes normiertes Polynom $m_A \neq \mathbf{0}$ kleinsten Grades in $P(A)$, und*

$$P(A) = \{q \cdot m_A \mid q \in K[X]\}$$

besteht genau aus den Vielfachen von m_A.

Beweis. Nach Lemma 16.18 gibt es normierte Polynome in $P(A)$. Sei m_A ein solches mit minimalem Grad. Ist p ein weiteres normiertes Polynom in $P(A)$ mit $\deg(p) = \deg(m_A)$, dann ist $p - m_A \in P(A)$ entweder das Nullpolynom (und damit $p = m_A$) oder $0 \leq \deg(p - m_A) < \deg(m_A)$. Nach Division durch den Leitkoeffizienten würde man dann ein normiertes Polynom in $P(A)$ mit kleinerem Grad als $\deg(m_A)$ erhalten, im Widerspruch zur Wahl von m_A. Also ist m_A eindeutig bestimmt. Wir sehen auch, dass $\deg(m_A) = \min\{\deg(p) \mid \mathbf{0} \neq p \in P(A)\}$ ist.

Nach Lemma 16.19 ist klar, dass $\{q \cdot m_A \mid q \in K[X]\} \subset P(A)$ ist. Sei umgekehrt $p \in P(A)$. Dann gibt es $q, r \in K[X]$ mit $p = q \cdot m_A + r$ und $\deg(r) < \deg(m_A)$ (Satz 15.16). Es folgt $r = p - q \cdot m_A \in P(A)$, und wegen $\deg(r) < \deg(m_A)$ muss (mit demselben Argument wie oben) $r = \mathbf{0}$ sein. Damit ist $p = q \cdot m_A$, also hat man auch die umgekehrte Inklusion. ❑

Derselbe Beweis zeigt, dass jedes Ideal I von $K[X]$ die Form $I = \{p \cdot a \mid p \in K[X]\}$ hat mit einem geeigneten $a \in K[X]$ ($a = \mathbf{0}$ ist möglich, dann ist $I = \{\mathbf{0}\}$). So ein Ideal, dessen Elemente genau die Vielfachen eines Elements a sind, heißt ein *Hauptideal*, und ein Ring, in dem jedes Ideal ein Hauptideal ist, ist ein *Hauptidealring*. Wir haben also gezeigt, dass $K[X]$ ein solcher Hauptidealring ist. Ein anderes Beispiel für einen Hauptidealring ist der Ring $\mathbb{Z}$ der ganzen Zahlen. Der Beweis ist im Wesentlichen derselbe, nur dass man Division mit Rest von ganzen Zahlen verwendet statt der Polynomdivision.

DEF Minimalpolynom

16.21. **Definition.** Das Polynom m_A in Satz 16.20 heißt das *Minimalpolynom* von A. Ist f ein Endomorphismus eines endlich-dimensionalen Vektorraums, der bezüglich einer geeigneten Basis durch A beschrieben wird, dann heißt $m_f = m_A$ das *Minimalpolynom* von f. ♢

BSP Minimalpolynome

16.22. **Beispiele.** Seien $n \geq 1$ und $A \in \mathrm{Mat}(n, K)$.

(1) Sei $A = \mathbf{0}$ die Nullmatrix. Dann ist $m_A = X$.

(2) Ist $A = \lambda I_n$, dann ist $m_A = X - \lambda$. Die Umkehrung gilt ebenfalls.

(3) Ist A diagonalisierbar und hat die (paarweise verschiedenen) Eigenwerte $\lambda_1, \ldots, \lambda_m$, dann ist

$$m_A = (X - \lambda_1)(X - \lambda_2) \cdots (X - \lambda_m)\,.$$

Das Minimalpolynom zerfällt also in Linearfaktoren und hat keine mehrfachen Nullstellen. Das sieht man so: Jedes Element v von K^n lässt sich

schreiben als $v = v_1 + \ldots + v_m$ mit $v_j \in E_{\lambda_j}(A)$. Wegen $Av_j = \lambda_j v_j$ ist dann mit $p = (X - \lambda_1)(X - \lambda_2)\cdots(X - \lambda_m)$ und Beispiel 16.17

$$\begin{aligned} p(A) \cdot v &= p(A) \cdot v_1 + p(A) \cdot v_2 + \ldots + p(A) \cdot v_m \\ &= p(\lambda_1)v_1 + p(\lambda_2)v_2 + \ldots + p(\lambda_m)v_m = \mathbf{0}\,, \end{aligned}$$

also ist p ein normiertes Polynom in $P(A)$. Auf der anderen Seite gibt es für jedes $j \in \{1, 2, \ldots, m\}$ ein $\mathbf{0} \neq v_j \in E_{\lambda_j}(A)$, und es muss gelten

$$\mathbf{0} = m_A(A) \cdot v_j = m_A(\lambda_j)v_j\,, \quad \text{also ist} \quad m_A(\lambda_j) = 0\,.$$

Das Minimalpolynom m_A muss also alle λ_j als Nullstellen haben. Es folgt $\deg(m_A) \geq m = \deg(p)$. Nach der Definition des Minimalpolynoms ist demnach $m_A = p$.

Da das charakteristische Polynom von A die Form

$$(X - \lambda_1)^{e_1}(X - \lambda_2)^{e_2} \cdots (X - \lambda_m)^{e_m}$$

hat, wobei die $e_j \geq 1$ (für $j \in \{1, 2, \ldots, m\}$) die algebraischen Vielfachheiten der Eigenwerte sind, sehen wir, dass das charakteristische Polynom ein Vielfaches des Minimalpolynoms ist (also $\chi_A = q \cdot m_A$ mit einem Polynom $q \in K[X]$). Hat A n verschiedene Eigenwerte, dann sind die beiden Polynome gleich. ♣

Das dritte Beispiel oben wirft zwei Fragen auf:

- Für A diagonalisierbar gilt, dass χ_A ein Vielfaches von m_A ist. Ist das auch allgemein richtig?
- Gilt die Umkehrung der Beobachtung im Beispiel: Wenn m_A in Linearfaktoren zerfällt und keine mehrfachen Nullstellen hat, muss dann A diagonalisierbar sein?

Die erste Frage wird durch den Satz von Cayley-Hamilton (mit „Ja“) beantwortet, den wir bald beweisen werden. Die zweite können wir gleich behandeln.

* 16.23. **Satz.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Dann ist A genau dann diagonalisierbar, wenn das Minimalpolynom m_A in Linearfaktoren zerfällt und keine mehrfachen Nullstellen hat.*

SATZ Kriterium für Diagonalisierbarkeit

Die analoge Aussage gilt natürlich auch für Endomorphismen endlich-dimensionaler Vektorräume.

Beweis. Es ist nur noch die Rückrichtung zu zeigen. Es gelte also

$$m_A = (X - \lambda_1)(X - \lambda_2)\cdots(X - \lambda_m)$$

mit paarweise verschiedenen $\lambda_1, \lambda_2, \ldots, \lambda_m \in K$. Für $j \in \{1, 2, \ldots, m\}$ sei

$$p_j = \prod_{i \neq j} \frac{X - \lambda_i}{\lambda_j - \lambda_i} \in K[X]\,.$$

Dann gilt $p_j(\lambda_i) = \delta_{ij}$ und $\deg(p_j) \leq m - 1$. Für die Summe $p = p_1 + p_2 + \ldots + p_m$ gilt also $p(\lambda_j) = 1$ für alle j und $\deg(p) < m$. Es folgt $p = 1$, denn das Polynom $p - 1$ vom Grad $< m$ hat mindestens m Nullstellen. Außerdem ist $(X - \lambda_j)p_j$ ein (konstantes) Vielfaches von m_A, also ist $(A - \lambda_j I_n)p_j(A) = \mathbf{0}$.

Sei $U_j = \mathrm{im}(p_j(A)) \subset K^n$. Dann gelten die folgenden Aussagen:

(1) $U_j \subset E_{\lambda_j}(A)$: Sei $u \in U_j$, also $u = p_j(A) \cdot v$ für ein $v \in K^n$. Es folgt

$$\mathbf{0} = (A - \lambda_j I_n) p_j(A) \cdot v = (A - \lambda_j I_n) \cdot u\,, \quad \text{also} \quad A \cdot u = \lambda_j u\,.$$

(2) $\langle U_1 \cup U_2 \cup \ldots \cup U_m \rangle = K^n$: Sei $v \in K^n$. Dann gilt

$$v = p(A) \cdot v = p_1(A) \cdot v + p_2(A) \cdot v + \ldots + p_m(A) \cdot v \in \langle U_1 \cup U_2 \cup \ldots \cup U_m \rangle\,.$$

Für $j \in \{1, 2, \ldots, m\}$ sei B_j eine Basis von U_j und B die durch Aneinanderhängen von $B_1, B_2, \ldots, B_m$ entstehende Familie. Punkt (1) oben zeigt, dass B aus Eigenvektoren von A besteht; außerdem ist B nach Folgerung 15.7 linear unabhängig. Punkt (2) zeigt, dass B ein Erzeugendensystem von K^n ist; insgesamt ist also B eine Basis von K^n, die aus Eigenvektoren von A besteht. Damit ist A diagonalisierbar. ❑

Um Satz 16.23 anwenden zu können, müssen wir in der Lage sein, das Minimalpolynom zu berechnen. Wir hatten bereits beobachtet, dass für Diagonalmatrizen das Minimalpolynom stets ein Teiler des charakteristischen Polynoms ist. Für 2×2-Matrizen gilt das auch allgemein:

16.24. **Beispiel.** Sei $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{Mat}(2, K)$. Dann ist

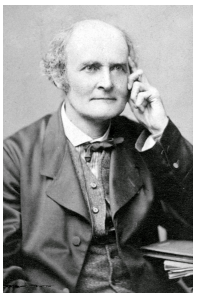
BSP Cayley-Hamilton für $n = 2$

$$\chi_A = X^2 - (a+d)X + ad - bc\,,$$

und damit

$$\begin{aligned} \chi_A(A) &= A^2 - (a+d)A + (ad-bc)I_2 \\ &= \begin{pmatrix} a^2+bc & (a+d)b \\ (a+d)c & bc+d^2 \end{pmatrix} - \begin{pmatrix} (a+d)a & (a+d)b \\ (a+d)c & (a+d)d \end{pmatrix} + \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

♣



A. Cayley 1821–1895

Wir beweisen diese Aussage, die als „Satz von Cayley-Hamilton“ bekannt ist, jetzt allgemein.

* 16.25. **Satz.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Dann ist*

SATZ Cayley-Hamilton

$$\chi_A(A) = \mathbf{0}\,.$$

Warum ist der folgende „Beweis“ nicht korrekt?

$$\chi_A(X) = \det(XI_n - A) \implies \chi_A(A) = \det(AI_n - A) = \det(\mathbf{0}) = 0\,.$$

Wenn wir an dieser Stelle die richtigen Hilfsmittel zur Verfügung hätten, dann könnten wir so argumentieren: Die Aussage gilt für diagonalisierbare Matrizen nach Beispiel 16.22(3). Diagonalisierbare Matrizen sind „dicht“ (in einem geeigneten Sinn) in allen $n \times n$-Matrizen; die Aussage folgt dann, weil die Abbildung $\mathrm{Mat}(n, A) \to \mathrm{Mat}(n, A)$, $A \mapsto \chi_A(A)$ stetig ist (wiederum in einem geeigneten Sinn). Das wird im Kleingedruckten unten genauer erklärt.

W.R. Hamilton 1805–1865

Beweis. Wir gehen hier anders vor. Wir stellen erst einmal fest, dass wir auch mit Matrizen über kommutativen Ringen (statt über Körpern) rechnen können, solange wir nicht durch Ringelemente dividieren müssen. Insbesondere können wir Determinanten bilden und damit auch die adjungierte Matrix (siehe Definition 14.13). Wir wenden das an auf den Ring $K[X]$, für den wir schon mit Determinanten gearbeitet haben, denn χ_A ist ja definiert als $\det(XI_n - A)$. Sei also $B = XI_n - A \in \mathrm{Mat}(n, K[X])$ und $\tilde{B}$ die adjungierte Matrix zu B (deren Einträge bis aufs Vorzeichen Determinanten von $(n-1) \times (n-1)$-Untermatrizen von B sind). Dann gilt (Satz 14.14)

$$\tilde{B}(XI_n - A) = \tilde{B}B = \det(B)I_n = \chi_A(X)I_n\,. \tag{16.1}$$

Die Einträge von $\tilde{B}$ sind Polynome in X vom Grad $< n$, wie man sich leicht überlegt; wir können also schreiben

$$\tilde{B} = \left(b_{ij}^{(n-1)}X^{n-1} + \ldots + b_{ij}^{(1)}X + b_{ij}^{(0)}\right)_{i,j}$$

mit geeigneten Koeffizienten $b_{ij}^{(k)} \in K$. (Wir schreiben den oberen Index in Klammern, um eine Verwechslung mit Potenzen zu vermeiden.) Nach den Rechenregeln für Matrizen können wir das auch schreiben als

$$\tilde{B} = \tilde{B}_{n-1}X^{n-1} + \ldots + \tilde{B}_1X + \tilde{B}_0$$

mit $\tilde{B}_k = (b_{ij}^{(k)})_{i,j} \in \mathrm{Mat}(n, K)$. Sei außerdem

$$\chi_A(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_1X + a_0\,.$$

Wir setzen in (16.1) ein und erhalten

$$\begin{aligned}\tilde{B}_{n-1}X^n + (\tilde{B}_{n-2} - \tilde{B}_{n-1}A)X^{n-1} + \ldots + (\tilde{B}_1 - \tilde{B}_2A)X^2 + (\tilde{B}_0 - \tilde{B}_1A)X - \tilde{B}_0A\\ = I_nX^n + a_{n-1}I_nX^{n-1} + \ldots + a_2I_nX^2 + a_1I_nX + a_0I_n\,.\end{aligned}$$

Koeffizientenvergleich zeigt dann

$$\tilde{B}_{n-1} = I_n\,,\ \tilde{B}_{n-2} - \tilde{B}_{n-1}A = a_{n-1}I_n\,,\ \ldots\,,\ \tilde{B}_0 - \tilde{B}_1A = a_1I_n\,,\ -\tilde{B}_0A = a_0I_n\,.$$

Wir multiplizieren diese Gleichungen von rechts mit A^n, A^{n-1}, …, A, I_n und summieren auf:

$$\begin{array}{rrrrrl}
\tilde{B}_{n-1}A^n & & & & & = A^n\\
-\tilde{B}_{n-1}A^n & + \tilde{B}_{n-2}A^{n-1} & & & & = a_{n-1}A^{n-1}\\
\ddots & \ddots & & & & \vdots\\
& \ddots & \ddots & & & \vdots\\
& & - \tilde{B}_1A^2 & + \tilde{B}_0A & & = a_1A\\
& & & - \tilde{B}_0A & & = a_0I_n\\
\hline
& & & & \mathbf{0} & = \chi_A(A)
\end{array}$$

❑

Ein Beweis, wie er oben angedeutet wurde, könnte etwa wie folgt aussehen. Wir zeigen die Aussage erst einmal für Matrizen über $\mathbb{C}$. Sei also $A \in \mathrm{Mat}(n, \mathbb{C})$. Ist A diagonalisierbar, dann ist $\chi_A(A) = \mathbf{0}$ und wir sind fertig. Sonst können wir A beliebig wenig stören, sodass das charakteristische Polynom der gestörten Matrix A' keine mehrfachen Nullstellen hat. Dann ist A' diagonalisierbar, also gilt $\chi_{A'}(A') = \mathbf{0}$. Da $\chi_A(A)$ eine stetige Funktion von A ist (d.h., die Einträge dieser Matrix hängen stetig von den Einträgen von A ab), folgt $\chi_A(A) = \mathbf{0}$.

Nun gilt ganz allgemein, dass die Einträge von $\chi_A(A)$ Polynome in den Einträgen von A mit ganzzahligen Koeffizienten sind (diese Polynome haben Grad n in dem Sinne, dass sie

ganzzahlige Linearkombinationen von Produkten von jeweils n Einträgen von A sind, vergleiche Beispiel 16.24) — daraus folgt auch die oben schon verwendete Stetigkeit. Wir haben gesehen, dass diese Polynome stets den Wert null annehmen, wenn man beliebige komplexe Zahlen einsetzt. Daraus folgt aber, dass die Polynome null sind (als Polynome). Das zeigt dann, dass $\chi_A(A) = \mathbf{0}$ für Matrizen über beliebigen Körpern (oder sogar kommutativen Ringen) gilt.

FOLG m_A teilt χ_A und hat dieselben Nullstellen

16.26. **Folgerung.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Das charakteristische Polynom χ_A von A ist ein Vielfaches des Minimalpolynoms m_A von A. Jede Nullstelle von χ_A ist auch eine Nullstelle von m_A.*

Beweis. Jedes Polynom $p \in K[X]$ mit $p(A) = \mathbf{0}$ ist ein Vielfaches von m_A, vgl. Satz 16.20, und $\chi_A(A) = \mathbf{0}$ nach Satz 16.25.

Ist λ eine Nullstelle von χ_A, dann ist λ ein Eigenwert von A, es gibt also einen Eigenvektor $\mathbf{0} \neq v \in K^n$ mit $A \cdot v = \lambda v$. Es folgt (vergleiche Beispiel 16.17)

$$\mathbf{0} = m_A(A) \cdot v = m_A(\lambda) v \,,$$

wegen $v \neq \mathbf{0}$ also $m_A(\lambda) = 0$. ❑

Folgerung 16.26 hilft uns, das Minimalpolynom zu bestimmen: Wir können χ_A berechnen und hoffentlich in Linearfaktoren zerlegen als

$$\chi_A = (X - \lambda_1)^{e_1} (X - \lambda_2)^{e_2} \cdots (X - \lambda_m)^{e_m}$$

mit paarweise verschiedenen λ_j und Exponenten (also algebraischen Vielfachheiten) $e_j \geq 1$. Die Folgerung sagt uns, dass

$$m_A = (X - \lambda_1)^{f_1} (X - \lambda_2)^{f_2} \cdots (X - \lambda_m)^{f_m}$$

sein muss mit $1 \leq f_j \leq e_j$. Es gibt also nur endlich viele Möglichkeiten für m_A. Wir können diese Polynome p dann, aufsteigend sortiert nach ihrem Grad, testen, ob sie $p(A) = \mathbf{0}$ erfüllen. Das erste Polynom, das den Test besteht, ist das Minimalpolynom.

BSP Minimalpolynom

16.27. **Beispiel.** Wir bestimmen das Minimalpolynom von

$$A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix} \in \mathrm{Mat}(3, \mathbb{R}) \,.$$

Es ist

$$\chi_A = \begin{vmatrix} X & -1 & 0 \\ 1 & X-2 & 0 \\ -1 & 1 & X-1 \end{vmatrix} = (X-1)\big(X(X-2)+1\big) = (X-1)^3 \,,$$

also ist $m_A \in \{X-1, (X-1)^2, (X-1)^3\}$. Wir probieren die Möglichkeiten der Reihe nach durch:

$$(X-1)(A) = A - I_3 = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}$$

ist nicht die Nullmatrix, aber $(X-1)^2(A) = (A - I_3)^2 = \mathbf{0}$, also ist $m_A = (X-1)^2$ (und A ist nicht diagonalisierbar nach Satz 16.23, denn das Minimalpolynom hat eine mehrfache Nullstelle). ♣