Lineare Algebra I und II

Winter- und Sommersemester 2020/21

Universität Bayreuth

Michael Stoll

Inhaltsverzeichnis

1.	Einige allgemeine Vorbemerkungen	2
2.	Die Sprache der Mathematik: Logik und Mengenlehre	3
3.	Algebraische Strukturen: Gruppen, Ringe, Körper	21
4.	Der Körper der komplexen Zahlen	27
5.	Vektorräume: Definition und Beispiele	32
6.	Untervektorräume	37
7.	Erzeugendensysteme	40
8.	Lineare Unabhängigkeit	47
9.	Basis und Dimension	53
10.	Lineare Abbildungen	65
11.	Matrizen	80
12.	Der Normalformalgorithmus und Lineare Gleichungssysteme	85
13.	Matrizen und lineare Abbildungen	98
14.	Die Determinante	103
15.	Eigenwerte und Eigenvektoren	115
16.	Diagonalisierbarkeit und der Satz von Cayley-Hamilton	126
17.	Summen von Untervektorräumen, Komplemente, Kodimension	139
18.	Äquivalenzrelationen, Quotientenräume und affine Unterräume	149
19.	Der Dualraum	160
20.	Bilinearformen	167
21.	Euklidische und unitäre Vektorräume	178
22.	Orthogonale und unitäre Diagonalisierung	188
23.	Quadratische Formen	197
24.	Klassifikation von Quadriken	203
25.	Orthogonale Gruppen	210
26.	Die Jordansche Normalform	220
27.	Das Tensorprodukt	236
28.	Symmetrische und alternierende Potenzen	244

Druckversion vom 28. Juli 2021, 22:00 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 2020):

> 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*e *Mathematiker*in* die*derjenige ist, die*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"), sodass 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

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

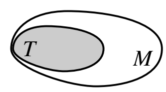
(Statt des senkrechten Strichs „|" ist auch ein Doppelpunkt „:" gebräuchlich.) Es gibt Symbole für gewisse häufig benötigte Mengen, wie

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.

DEF Teilmenge



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*. ◊

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 $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

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

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. $\diamondsuit$

BSP Potenzmenge

2.3. **Beispiel.** Zum Beispiel gilt

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

An dieser Stelle gleich noch ein wichtiger Hinweis:

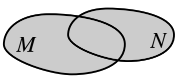
- Man muss sorgfältig zwischen Mengen und ihren Elementen unterscheiden. Zum Beispiel haben „$a \in M$“ und „$a \subset M$“ völlig verschiedene Bedeutungen.
- Besonders schwer fällt die Unterscheidung zwischen dem Element a und der Einermenge $\{a\}$. Es ist sehr wichtig, sich diese Unterschiede gleich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

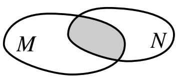
Mengen können miteinander verknüpft werden:

DEF

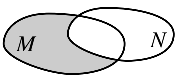
$M \cup N$



$M \cap N$



$M \setminus N$



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

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

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

(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*.

(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. $\diamondsuit$

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).)

SATZ Eigensch. Mengen

2.5. **Satz.**

(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

$$((A \Rightarrow C) \land (B \Rightarrow C)) \Rightarrow ((A \lor B) \Rightarrow C) \ ,$$

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

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

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

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

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

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

(„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:

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

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** $\forall\exists$ und $\exists\forall$

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

(„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:

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

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. ♢

BSP
kartesische Produkte

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. ♣

2.E Abbildungen.

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

DEF
Abbildung

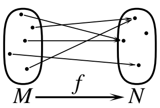


Bild
Urbild

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 „$\rightarrow$“ 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.

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 $f \colon M \to N$, $x \mapsto f(x)$.)

DEF Definitions-, Wertebereich $f = g$ $\mathrm{Abb}(M, N)$

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.

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.

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

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

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

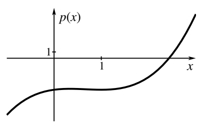
von f zu unterscheiden. Die Bildmenge kann eine echte Teilmenge des Wertebereichs sein.

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.

DEF
injektiv
surjektiv
bijektiv

2.14. **Definition.** Sei $f\colon M \to N$ eine Abbildung.

(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.

DEF
Umkehrabb.
Permutation

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. ♢

BSP
injektiv
surjektiv

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:

(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

$$\begin{tikzcd} & Y \arrow[dr, "g"] & \\ X \arrow[ur, "f"] \arrow[rr, "g\circ f"] & & Z \end{tikzcd}$$

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:

$$x \longrightarrow \boxed{f} \longrightarrow f(x) \longrightarrow \boxed{g} \longrightarrow g(f(x))$$ ♢

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

(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*
$$f^{-1} \circ f = \mathrm{id}_X \qquad \textit{und} \qquad f \circ f^{-1} = \mathrm{id}_Y\ .$$

(6) *Sind $f\colon X \to Y$ und $g\colon Y \to Z$ zwei Abbildungen, dann gilt*
$$g \circ f \textit{ injektiv} \Longrightarrow f \textit{ injektiv} \qquad \textit{und} \qquad g \circ f \textit{ surjektiv} \Longrightarrow g \textit{ surjektiv.}$$

(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

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

„$\Rightarrow$“: 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.
„$\Leftarrow$“: 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) „$\Rightarrow$“: 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 $f \circ g = \mathrm{id}_Y$.
„$\Leftarrow$“: Das folgt aus Teil (6), denn id_Y ist surjektiv.

(9) „$\Rightarrow$“: Ist f bijektiv, dann hat $g = f^{-1}$ die verlangte Eigenschaft.
„$\Leftarrow$“: 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.

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: **DEF** Halbgruppe

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

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:

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

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

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

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 $S(X)$

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

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

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

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.

LEMMA
Kürzungsregel in Gruppen

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

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

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*

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

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.

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* **DEF** Ring

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

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 $0 \cdot a = a \cdot 0 = 0$, $-(a+b) = -a-b$, $(-a) \cdot (-b) = a \cdot b$ (Übung). Was aber im Allgemeinen *nicht* gelten muss, ist die Implikation $a \cdot b = 0 \Rightarrow a = 0 \lor b = 0$. Wenn sowohl Summen als auch Produkte in einem Term vorkommen, gilt „Punkt vor Strich“, d.h., Multiplikation bindet stärker als Addition:

$$a \cdot b + c \cdot d = (a \cdot b) + (c \cdot d) \qquad \text{und nicht} \qquad a \cdot (b + c) \cdot d\,.$$

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 $r = 1_R \cdot r = 0_R \cdot r = 0_R$. **BSP** Ringe

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!)

DEF ∗
Körper

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
multiplikative Gruppe $K^\times$

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.)

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 $0 = a^{-1} \cdot 0 = a^{-1} \cdot a \cdot b = 1 \cdot b = b$.)

BSP
Körper

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.

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 $\boldsymbol{i} =$ „$\sqrt{-1}$“ 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.

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

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.

DEF Körper der komplexen Zahlen

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*. $\diamondsuit$

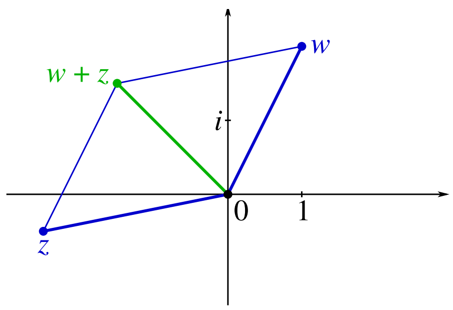
Die in der Definition gemachten Behauptungen sollten wir nachprüfen:

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

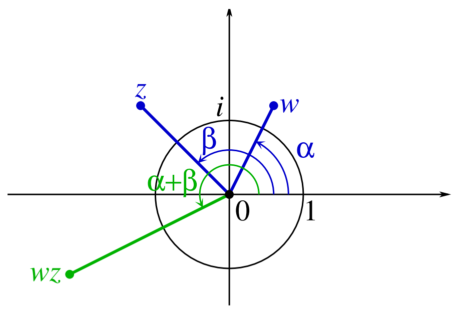
und

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

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.

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 $z \in \mathbb{C}$.

SATZ
quadratische Gleichungen 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 $z = \sqrt{|w|}\bigl(\cos\frac{\alpha}{2} + \boldsymbol{i}\sin\frac{\alpha}{2}\bigr)$ 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

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

DEF
algebraisch abgeschlossen

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:

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 ein *reeller Vektorraum*, ein $\mathbb{C}$-Vektorraum ein *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) $\forall v \in V\colon 0 \cdot v = \mathbf{0}$.

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

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

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

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)

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

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:

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

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"):

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

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

$$\begin{aligned} \mathcal{P}(X) &\longrightarrow \mathbb{F}_2^X\,, & A &\longmapsto \left(x \mapsto \begin{cases} 0 & \text{falls } x \notin A, \\ 1 & \text{falls } x \in A, \end{cases}\right) \\ \mathbb{F}_2^X &\longrightarrow \mathcal{P}(X)\,, & f &\longmapsto \{x \in X \mid f(x) = 1\}\,. \end{aligned}$$

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) $\forall u_1, u_2 \in U : u_1 + u_2 \in U$
(„U ist abgeschlossen unter der Addition"),

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

Wir zeigen gleich, dass diese Definition sinnvoll ist.

6.2. **Lemma.** *Seien 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 $+_U : U \times U \to U$, $(u_1, u_2) \mapsto u_1 + u_2$, $-_U$ für die auf U eingeschränkte Negationsabbildung $-_U : U \to U$, $u \mapsto -u$, und $\cdot_U$ für die auf U eingeschränkte Skalarmultiplikation $\cdot_U : K \times U \to U$, $(\lambda, u) \mapsto \lambda \cdot u$. Dann ist $(U, +_U, \mathbf{0}, -_U, \cdot_U)$ ein K-Vektorraum.

Beweis. Die erste Behauptung ist $\forall u \in U : -u \in U$. 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 $u_1, u_2, \ldots$ 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.

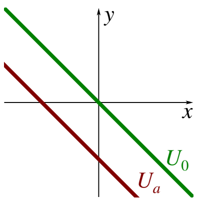
Die Schreibweise $+_U$ usw. für die auf U eingeschränkten Abbildungen diente nur der Verdeutlichung für die Formulierung des Lemmas. Wir schreiben normalerweise einfach $+$ usw. für die Addition usw. auf U.

BSP
triviale
Unter-VR

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

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

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 $\mathbb{R}^2$?



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.

BSP
Folgenräume

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

(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). ♣

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

BSP Funktionenräume

(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

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

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

Damit sind alle drei Bedingungen erfüllt. ❑

♣

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

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

BSP Lineare Codes

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

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

$$\begin{aligned} &U_1, U_2, \ldots, U_n \subset V \text{ Untervektorräume} \\ &\qquad\qquad \Longrightarrow \; U_1 \cap U_2 \cap \ldots \cap U_n \subset V \text{ Untervektorraum.} \end{aligned}$$

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 $\lambda u \in U$. ❑

7.4. **Beispiel.** Der Raum **BSP**

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

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.

DEF $\ast$
Lineare Hülle

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

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

(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 $A = \{v_1, v_2, \dots, v_n\}$ endlich, dann schreiben wir auch

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

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 $\langle \{v_i \mid i \in I\} \rangle$. $\diamondsuit$

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 $\langle 1 \rangle = \langle 1 \rangle_{\mathbb{C}} = \mathbb{C}$.

BSP
$\langle \emptyset \rangle = \{\mathbf{0}\}$

7.7. **Beispiel.** In Definition 7.6 können wir für A die leere Menge wählen. Was ist der von A erzeugte Untervektorraum?

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 $\langle \emptyset \rangle = \{\mathbf{0}\}$. $\clubsuit$

DEF $\ast$
Erzeugendensystem

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. $\diamondsuit$

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.

7.9. **Satz.** *Sei V ein K-Vektorraum.* **SATZ** Beschreibung von $\langle A\rangle$

(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:

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

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.

DEF
Linear-
kombination

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

(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!

DEF
$\sum_{i\in I} a_i$

7.11. **Definition.** Um Summen von beliebig vielen Elementen präzise und kurz hinschreiben zu können, führen wir folgende Summenschreibweise ein: Wir schreiben

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

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

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

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. ♢

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

BSP Erzeugendensystem von K^n

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

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:

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

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

BSP Viele Erzeugendensysteme

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

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

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

BSP Vektorraum der Polynomfunktionen

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

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.

LEMMA
Eigensch. lineare Hülle

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

(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 folgt aus Lemma 7.15: Wir müssen zeigen, dass jedes Element von E eine Linearkombination von Elementen von $E \setminus \{v_0\}$ ist. Für $v \in E$ mit $v \neq v_0$ können wir aber einfach $v = 1 \cdot v$ schreiben (denn $v \in E \setminus \{v_0\}$). Für v_0 haben wir die obige Darstellung.

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*. $\diamondsuit$

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!

BSP
$\emptyset$ ist linear unabhängig

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

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. ♣

BSP
Wann ist v linear unabhängig?

8.3. **Beispiel.** Wann ist ein einzelner Vektor 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. ♣

BSP
Lineare Unabhängigkeit von zwei Vektoren

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

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?

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. ♣

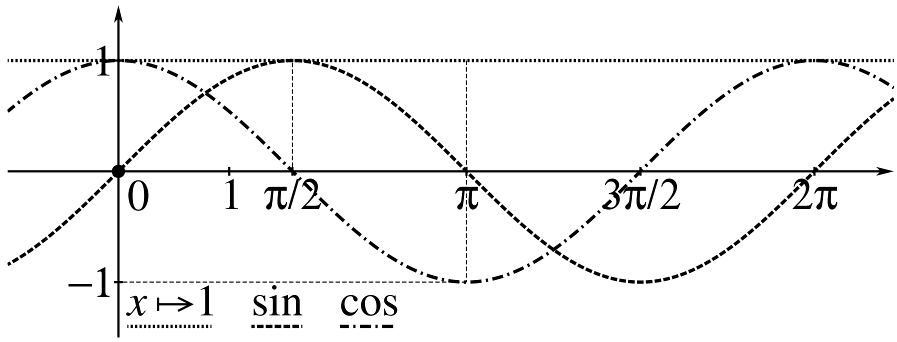


ABBILDUNG 1. Zu Beispiel 8.7

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

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

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

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 $v \in V$ gilt entsprechend:*

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

Beweis.

(1) „$\Rightarrow$“: $v \in \langle v_1, v_2, \ldots, v_n \rangle$ 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.

„$\Leftarrow$“: Da $(v, v_1, \dots, v_n)$ 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 $(v_1, v_2, \dots, v_n)$, 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 $v \in \langle v_1, v_2, \dots, v_n\rangle$ 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 $v_1, v_2, \dots, v_n \in V$. Wir definieren die zugehörige „Linearkombinationenabbildung“*

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

Dann gilt:

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

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

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

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

Wir beweisen Teil (2). „⇒“: Wir nehmen an, dass $v_1, v_2, \dots, v_n$ 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 $\lambda_1, \lambda_2, \dots, \lambda_n, \mu_1, \mu_2, \dots, \mu_n \in K$, 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 $v_1, v_2, \dots, v_n$ linear unabhängig sind, muss das die triviale Linearkombination sein, also folgt $\lambda_1 = \mu_1,\ \lambda_2 = \mu_2,\ \dots,\ \lambda_n = \mu_n$.

„⇐“: Wir nehmen an, dass jedes $v \in V$ höchstens auf eine Weise als Linearkombination von $v_1, v_2, \dots, v_n$ 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 $v_1, v_2, \dots, v_n$ 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

SATZ
Basisergänzungssatz

* 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*

(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 $(v_1, \ldots, v_n, w_1, \ldots, w_m)$ sukzessive Elemente w_j entfernt, bis man linear unabhängige Vektoren hat. Dazu fixieren wir das n-Tupel $(v_1, \ldots, v_n)$ 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 $(v_1, \ldots, v_n)$ und jedes m-Tupel $(w_1, \ldots, w_m)$ von Vektoren von V gilt.

Im Induktionsanfang ist $m = 0$. Dann ist nach Voraussetzung (2) $(v_1, \ldots, v_n)$ 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 $(w_1, \ldots, w_m)$. Um sie für $m + 1$ zu zeigen, betrachten wir ein $(m + 1)$-Tupel $(w_1, \ldots, w_{m+1})$ von Vektoren von V, sodass $(v_1, \ldots, v_n, w_1, \ldots, w_{m+1})$ ein Erzeugendensystem von V ist. Es gibt nun zwei Möglichkeiten:

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

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

  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 $(v_1, \ldots, v_n)$. Also gibt es ein $j_0 \in \{1, 2, \ldots, m + 1\}$ 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

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

  ist, also ist

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

  Voraussetzung (2) ist also für $(w_1, \ldots, w_m)$ 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 $i \in \{1, 2, \ldots, m\} \setminus \{j_0\}$ und $w'_{j_0} = w_{m+1}$ (falls $j_0 \neq m + 1$), und wendet die Induktionsannahme auf $(w'_1, \ldots, w'_m)$ an. Dann bekommt man $k \leq m$ und $j'_1, \ldots, j'_k$, sodass $(v_1, \ldots, v_n, w'_{j'_1}, \ldots, w'_{j'_k})$ 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 $(v_1, \ldots, v_n, w_{j_1}, \ldots, w_{j_k})$ eine Basis. ❑

Man kann alternativ den Beweis auch so formulieren, dass man nacheinander Vektoren w_j zu den $v_1, v_2, \ldots, v_n$ 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 $(v_1, v_2, \ldots, v_n)$ 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 $(v_1, v_2, \ldots, v_n)$ linear unabhängig und seien außerdem $w_1, w_2, \ldots, w_m, w_{m+1} \in V$, sodass $(v_1, v_2, \ldots, v_n, w_1, w_2, \ldots, w_m, w_{m+1})$ ein Erzeugendensystem von V ist. Wir unterscheiden zwei Fälle:

(1) $w_{m+1} \in \langle v_1, v_2, \dots, v_n \rangle$.
Dann ist $(v_1, v_2, \dots, v_n, w_1, w_2, \dots, w_m)$ auch schon ein Erzeugendensystem; die Behauptung folgt direkt aus der Induktionsannahme.

(2) $w_{m+1} \notin \langle v_1, v_2, \dots, v_n \rangle$.
Wir schreiben v_{n+1} für w_{m+1}. Dann ist $(v_1, v_2, \dots, v_n, v_{n+1})$ linear unabhängig (wir benutzen hier wieder Lemma 8.9) und

$$(v_1, v_2, \dots, v_n, v_{n+1}, w_1, w_2, \dots, w_m)$$

ist ein Erzeugendensystem (dasselbe wie vorher). Nach der Induktionsannahme gibt es $j'_1, \dots, j'_{k'} \in \{1, 2, \dots, m\}$, sodass

$$(v_1, v_2, \dots, v_n, v_{n+1}, w_{j'_1}, w_{j'_2}, \dots, w_{j'_{k'}}) = (v_1, \dots, v_n, w_{m+1}, w_{j'_1}, \dots, w_{j'_{k'}})$$

eine Basis von V ist. Wir setzen

$$k = k' + 1, \quad j_1 = m + 1 \quad \text{und} \quad j_2 = j'_1,\ j_3 = j'_2,\ \dots,\ j_k = j'_{k'}$$

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* **SATZ** Zornsches Lemma

$$\forall T_1, T_2 \in \mathcal{K} \colon T_1 \subset T_2 \quad \text{oder} \quad T_2 \subset T_1 \,.$$

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 gilt*

$$\forall T \in \mathcal{K} \colon T \subset S \,,$$

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 $\mathcal{M}$).

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 $K_j \in \mathcal{K}$ mit $v_j \in K_j$. 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 $E = V$):

FOLG
Existenz
von Basen

Folgerung. *Jeder Vektorraum hat eine Basis.*

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.

FOLG
Erweiterung
zu Basis

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.*

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. ❑

BSP
Basis

9.8. **Beispiel.** Wir finden eine Basis des Untervektorraums

$$U = \{(x, y, z) \in \mathbb{R}^3 \mid z = x + y\} \subset \mathbb{R}^3 .$$

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 $(v_1, v_2, \ldots, v_n)$ und $(w_1, w_2, \ldots, w_m)$ zwei Basen von V. Für jedes $i \in \{1, 2, \ldots, n\}$ gibt es ein $j \in \{1, 2, \ldots, m\}$, sodass $(v_1, \ldots, v_{i-1}, w_j, v_{i+1}, \ldots, v_n)$ 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 $v_1, v_2, \ldots, v_{n-1}$ und $w_1, w_2, \ldots, w_m$. 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 $j_1, \ldots, j_k \in \{1, 2, \ldots, m\}$, sodass $(v_1, \ldots, v_{n-1}, w_{j_1}, w_{j_2}, \ldots, w_{j_k})$ 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 $(v_1, v_2, \ldots, v_{n-1})$ ist kein Erzeugendensystem mehr ($(v_1, v_2, \ldots, v_{n-1}, v_n)$ ist ein minimales Erzeugendensystem, aus dem wir ein Element entfernt haben). Wir zeigen, dass $(v_1, v_2, \ldots, v_{n-1}, w_{j_1})$ ein Erzeugendensystem ist; daraus folgt die Behauptung.

Wir haben $w_{j_1} \in V = \langle v_1, v_2, \ldots, v_n \rangle$. Nach Lemma 8.9 bedeutet das, dass $(v_1, v_2, \ldots, v_{n-1}, v_n, w_{j_1})$ linear abhängig ist. Da $(v_1, v_2, \ldots, v_{n-1}, w_{j_1})$ als Teil der Basis $(v_1, \ldots, v_{n-1}, w_{j_1}, w_{j_2}, \ldots, w_{j_k})$ linear unabhängig ist, folgt dann wieder mit Lemma 8.9, dass $v_n \in \langle v_1, v_2, \ldots, v_{n-1}, w_{j_1} \rangle$ ist. Da natürlich auch $v_1, v_2, \ldots, v_{n-1}$ in diesem Untervektorraum enthalten sind, enthält er ein Erzeugendensystem von V; es folgt $\langle v_1, v_2, \ldots, v_{n-1}, w_{j_1} \rangle = V$ wie behauptet. ❑

9.10. **Folgerung.** *Sei V ein Vektorraum und seien $(v_1, \ldots, v_n)$ und $(w_1, \ldots, w_m)$ 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 $i = 1, 2, \ldots, n$) erhalten wir Indizes $j_1, j_2, \ldots, j_n \in \{1, 2, \ldots, m\}$, sodass $(w_{j_1}, w_{j_2}, \ldots, w_{j_n})$ eine Basis von V ist. Da m kleiner als n ist, müssen sich in diesem Tupel Vektoren wiederholen. Dann sind $w_{j_1}, w_{j_2}, \ldots, w_{j_n}$ 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: $\mathbb{N}_{<n} = \{m \in \mathbb{N} \mid m < n\} = \{0, 1, 2, \ldots, n-1\}$ 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 (nach der ersten Definition) 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.

DEF ∗
Dimension

9.12. **Definition.** Sei V ein Vektorraum. Wenn V eine endliche Basis $(v_1, \ldots, v_n)$ 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 $\dim V < \infty$.

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 $m \le n$. 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 $m \ge n$. 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 $m \le n$. 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 $m \ge n$. 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**:*

$$\dim\langle v_1, v_2, \ldots, v_m\rangle \le m\,.$$

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.

FOLG
$\dim = \infty$

9.15. **Folgerung.** *Sei V ein Vektorraum. Die folgenden Aussagen sind äquivalent:*

(1) *Es gibt in V eine (unendliche) Folge $(v_n)_{n \in \mathbb{N}}$ linear unabhängiger Vektoren.*

(2) $\dim V = \infty$.

Beweis. „(1) $\Rightarrow$ (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) $\Rightarrow$ (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.

DEF
Grad einer Polynomfkt.

9.16. **Definition.** Wir sagen, eine Polynomfunktion $f \in P$ habe *Grad* $\le n$ (und wir schreiben $\deg(f) \le n$), wenn sie in der Form

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

(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 $\le n-1$ hat. f hat *Grad* $< n$ ($\deg(f) < n$), wenn f Grad $\le n-1$ 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 $f(x) = a_0 + a_1x + \ldots$ 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 $\mathbb{R}^n$:

$$\begin{aligned}
v_0 &= (1, 1, 1, \ldots, 1) \\
v_1 &= (x_1, x_2, x_3, \ldots, x_n) \\
v_2 &= (x_1^2, x_2^2, x_3^2, \ldots, x_n^2) \\
\vdots &\qquad \vdots \\
v_{n-1} &= (x_1^{n-1}, x_2^{n-1}, x_3^{n-1}, \ldots, x_n^{n-1}) \\
v_n &= (y_1, y_2, y_3, \ldots, y_n)
\end{aligned}$$

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 $v_j = (x_1^j, \ldots, x_n^j)$ für $j \in \{0, 1, \ldots, n-1\}$ 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 $\dim U \le \dim V$. Ist V endlich-dimensional und gilt $\dim U = \dim V$, dann ist $U = V$.*

Dabei gelte $n \le \infty$ für alle $n \in \mathbb{N}$ und $\infty \le \infty$.

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 $\dim U = m \le n = \dim V$. 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 $P_g \ne P$. 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

$$\dim V < \infty \implies V \cong K^{\dim V} .$$ ♣

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
$$\begin{aligned}(\phi_2 \circ \phi_1)(v + w) &= \phi_2\bigl(\phi_1(v + w)\bigr) = \phi_2\bigl(\phi_1(v) + \phi_1(w)\bigr) \\ &= \phi_2\bigl(\phi_1(v)\bigr) + \phi_2\bigl(\phi_1(w)\bigr) = (\phi_2 \circ \phi_1)(v) + (\phi_2 \circ \phi_1)(w)\end{aligned}$$
und
$$(\phi_2 \circ \phi_1)(\lambda v) = \phi_2\bigl(\phi_1(\lambda v)\bigr) = \phi_2\bigl(\lambda\phi_1(v)\bigr) = \lambda\phi_2\bigl(\phi_1(v)\bigr) = \lambda(\phi_2 \circ \phi_1)(v) .$$

(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
$$\phi^{-1}(v + w) = \phi^{-1}\bigl(\phi(v') + \phi(w')\bigr) = \phi^{-1}\bigl(\phi(v' + w')\bigr) = v' + w' = \phi^{-1}(v) + \phi^{-1}(w)$$
und
$$\phi^{-1}(\lambda v) = \phi^{-1}\bigl(\lambda\phi(v')\bigr) = \phi^{-1}\bigl(\phi(\lambda v')\bigr) = \lambda v' = \lambda\phi^{-1}(v) .$$ ❑

Bevor wir weitere Eigenschaften untersuchen, führen wir noch eine Schreibweise ein.

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 **DEF** Bilder und Urbilder von Teilmengen

$$f(T) = \{f(x) \mid x \in T\} \subset Y$$

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*. ♢

Aus $T \subset T' \subset X$ folgt $f(T) \subset f(T')$, und aus $U \subset U' \subset Y$ folgt $f^{-1}(U) \subset f^{-1}(U')$.

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. Finden Sie Beispiele dafür! ⚠

Da der Nullvektor eine ausgezeichnete Rolle in einem Vektorraum spielt, ist die Menge seiner Urbilder unter einer linearen Abbildung ein wichtiges Datum.

* 10.6. **Definition.** Sei $\phi\colon V_1 \to V_2$ eine lineare Abbildung. Der *Kern* von ϕ ist die Menge der Urbilder von $\mathbf{0} \in V_2$: **DEF** Kern

$$\ker(\phi) = \phi^{-1}(\{\mathbf{0}\}) = \{v \in V_1 \mid \phi(v) = \mathbf{0}\} \subset V_1 \,. \qquad \diamondsuit$$

Nach Lemma 10.4 gilt stets $\mathbf{0} \in \ker(\phi)$.

10.7. **Beispiel.** Sei $V \subset \mathbb{R}^{\mathbb{N}}$ der Vektorraum der konvergenten Folgen. Dann ist **BSP** Limes ist linear

$$\lim\colon V \longrightarrow \mathbb{R}, \qquad (a_n)_{n\in\mathbb{N}} \longmapsto \lim_{n\to\infty} a_n$$

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:*

LEMMA
injektiv
$\iff$
$\ker = \{\mathbf{0}\}$

$$\phi \textit{ ist injektiv} \iff \ker(\phi) = \{\mathbf{0}\}\,.$$

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 $v = w$. ❑

DEF
Kern trivial

Wenn $\ker(\phi) = \{\mathbf{0}\}$ ist, sagt man auch, der Kern sei *trivial*.

Wie zu erwarten, vertragen sich lineare Abbildungen sehr gut mit Untervektorräumen.

SATZ
lin. Abb.
und UVR

10.9. **Satz.** *Sei $\phi\colon V_1 \to V_2$ eine K-lineare Abbildung.*

(1) *Ist $U_1 \subset V_1$ ein Untervektorraum, dann ist $\phi(U_1) \subset V_2$ wieder ein Untervektorraum. Insbesondere ist $\operatorname{im}(\phi) = \phi(V_1) \subset V_2$ 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*

$$M_1 = \{U_1 \mid U_1 \subset V_1 \textit{ Untervektorraum mit } \ker(\phi) \subset U_1\} \qquad \textit{und}$$
$$M_2 = \{U_2 \mid U_2 \subset V_2 \textit{ Untervektorraum mit } U_2 \subset \operatorname{im}(\phi)\}\,.$$

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 $\operatorname{im}(\phi)$ ein Untervektorraum von V_2 ist.
Dass $\phi|_{U_1}$ linear ist, folgt daraus, dass man $\phi|_{U_1}$ als Verkettung der (linearen) Inklusionsabbildung $U_1 \to V_1$ und ϕ schreiben kann.

(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

$$\begin{aligned} v \in \phi^{-1}\bigl(\phi(U_1)\bigr) &\iff \phi(v) \in \phi(U_1) \\ &\iff \exists v' \in U_1 \colon \phi(v) = \phi(v') \\ &\iff \exists v' \in U_1 \colon \phi(v - v') = \mathbf{0} \\ &\iff \exists v' \in U_1 \colon v - v' \in \ker(\phi) \\ &\iff v \in U_1 \,. \end{aligned}$$

(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 $\phi^{-1}\bigl(\phi(U_1)\bigr) = U_1$.

Sei jetzt $U_2 \subset \operatorname{im}(\phi)$ ein Untervektorraum von V_2. Dann gilt

$$\begin{aligned} v \in \phi\bigl(\phi^{-1}(U_2)\bigr) &\iff \exists v' \in \phi^{-1}(U_2) \colon \phi(v') = v \\ &\iff \exists v' \in V_1 \colon \phi(v') \in U_2 \text{ und } \phi(v') = v \\ &\iff v \in U_2 \quad \text{und} \quad v \in \operatorname{im}(\phi) \\ &\iff v \in U_2 \cap \operatorname{im}(\phi) \\ &\iff v \in U_2 \,. \end{aligned}$$

Das zeigt $\phi\bigl(\phi^{-1}(U_2)\bigr) = U_2$. ❑

Den Zusammenhang, der im dritten Teil dieses Satzes formuliert wird, kann man sich etwa (sehr schematisch) wie in Abbildung 2 veranschaulichen.

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*

BSP 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!)

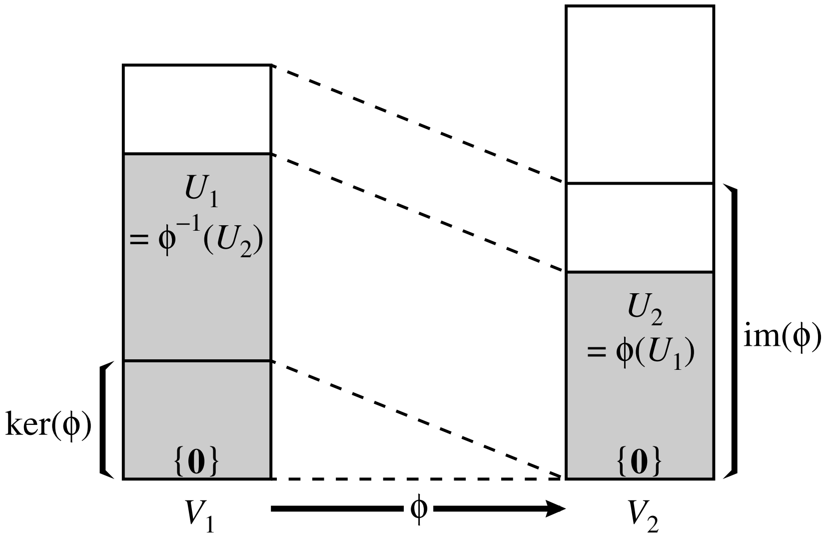


ABBILDUNG 2. Skizze zu Satz 10.9

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.

DEF Projektion

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:

$$\mathrm{pr}_j\colon K^n \longrightarrow K, \quad (a_1, a_2, \ldots, a_n) \longmapsto a_j$$

Sie sind also ebenfalls linear. ♣

Wir zeigen jetzt, dass eine lineare Abbildung dadurch festgelegt ist, was sie auf einer Basis macht.

SATZ ∗ Basen und lin. Abb.

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 $j \in \{1, 2, \ldots, n\}$.*

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.

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.* **SATZ** inj./surj. lin. Abb.

(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 $w_1 = \phi(b_1)$, $w_2 = \phi(b_2)$, $\ldots$, $w_n = \phi(b_n)$. 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.

FOLG endl.-dim. VR gleicher Dimension sind isomorph

10.13. **Folgerung.** *Sind V und W zwei K-Vektorräume derselben endlichen Dimension n, dann sind V und W 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. ❑

FOLG lin. Abb. bei gleicher Dimension

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:*

(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:

$$\begin{aligned} p_1(x) &= \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} = \frac{x(x-1)}{(-1)(-2)} = \tfrac{1}{2}x^2 - \tfrac{1}{2}x \\ p_2(x) &= \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} = \frac{(x+1)(x-1)}{1 \cdot (-1)} = -x^2 + 1 \\ p_3(x) &= \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)} = \frac{(x+1)x}{2 \cdot 1} = \tfrac{1}{2}x^2 + \tfrac{1}{2}x \end{aligned}$$

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 $\phi \circ \psi = \mathrm{id}_{\mathbb{R}^n}$:

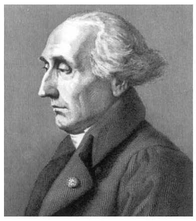
$$\phi\bigl(\psi(\mathbf{e}_j)\bigr) = \phi(p_j) = \bigl(p_j(x_1), \ldots, p_j(x_n)\bigr) = \mathbf{e}_j .$$

$\phi \circ \psi$ 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 $f = y_1 p_1 + y_2 p_2 + \ldots + y_n p_n$.*

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

$$f = \psi(y_1, y_2, \ldots, y_n) = y_1 p_1 + y_2 p_2 + \ldots + y_n p_n .$$


J.-L. Lagrange
1736–1813

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.

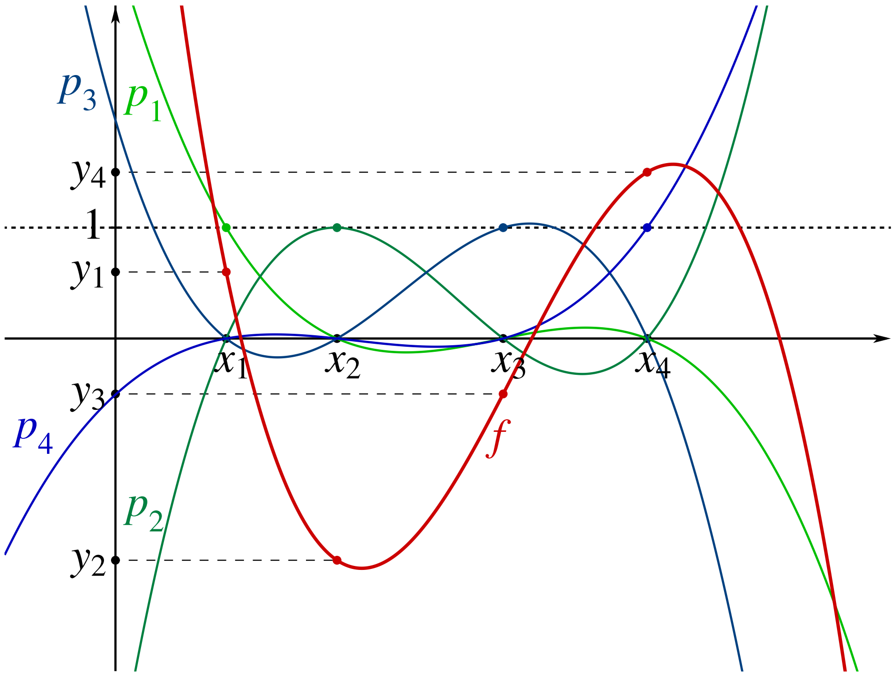


ABBILDUNG 3. Interpolationspolynom (zu Beispiel 10.15); hier ist der Fall mit vier Stützstellen $x_1, \ldots, x_4$ dargestellt.

BSP
lin. Abb. auf Polynomfkt.

Beispiele.

(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:

$$I_a \colon P \longrightarrow P, \quad f \longmapsto \Bigl(x \mapsto \int_a^x P(t)\,dt\Bigr).$$

Das ist die lineare Abbildung mit $I_a(f_n) = \bigl(x \mapsto \frac{1}{n+1}(x^{n+1} - a^{n+1})\bigr)$. 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.

$\ast$ 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.)

$\ast$ 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 $\phi(b_{k+1}) = b'_1$, $\phi(b_{k+2}) = b'_2$, $\ldots$, $\phi(b_{k+r}) = b'_r$. 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

$$\begin{aligned} \phi\bigl(v - (\mu_1 b_{k+1} + \ldots + \mu_r b_{k+r})\bigr) &= \phi(v) - \bigl(\mu_1 \phi(b_{k+1}) + \ldots + \mu_r \phi(b_{k+r})\bigr) \\ &= \phi(v) - (\mu_1 b'_1 + \ldots + \mu_r b'_r) = \mathbf{0}\,, \end{aligned}$$

  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. ❑

BSP
Dimension eines Kerns

10.18. **Beispiel.** Seien $V = K^n$ mit $n \geq 1$ und

$$\phi \colon V \longrightarrow K, \quad (x_1, x_2, \ldots, x_n) \longmapsto x_1 + x_2 + \ldots + x_n\,.$$

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 $\mathrm{im}(\phi) = K$:

$$\text{für } \lambda \in K \text{ ist } \phi\bigl((\lambda, 0, \ldots, 0)\bigr) = \lambda.$$

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.

DEF
Vektorraum
V^X

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

$$f + g \colon x \mapsto f(x) + g(x) \qquad \text{und} \qquad \lambda f \colon x \mapsto \lambda f(x)\,.$$

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*

$$\Phi \colon \mathrm{Hom}(V, W) \longrightarrow W^n, \quad \phi \longmapsto \big(\phi(b_1), \phi(b_2), \ldots, \phi(b_n)\big)$$

ein Isomorphismus. Insbesondere ist im Fall von $\dim W = m < \infty$

$$\dim \mathrm{Hom}(V, W) = \dim W^n = n \dim W = mn = (\dim V)(\dim W)\,.$$

SATZ
$\mathrm{Hom}(V, W)$
$\cong W^{\dim V}$

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 $b_1, \ldots, b_n$ 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. ❑

FOLG
Basis von $\mathrm{Hom}(V,W)$

10.23. **Folgerung.** *Seien V und W zwei K-Vektorräume; sei $(b_1, b_2, \ldots, b_n)$ eine Basis von V und $(b'_1, b'_2, \ldots, b'_m)$ eine Basis von W.*
Für $(i,j) \in \{1,2,\ldots,m\} \times \{1,2,\ldots,n\}$ sei $\phi_{ij} \colon V \to W$ die lineare Abbildung mit $\phi_{ij}(b_k) = \mathbf{0}$ für $k \neq j$ und $\phi_{ij}(b_j) = b'_i$. Dann ist $(\phi_{ij})_{(i,j)\in\{1,2,\ldots,m\}\times\{1,2,\ldots,n\}}$ eine 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 $k \in \{1,2,\ldots,n\}$. Einsetzen von b_k liefert dann

$$\mathbf{0} = \Bigl(\sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_{ij} \phi_{ij}\Bigr)(b_k) = \sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_{ij} \phi_{ij}(b_k) = \sum_{i=1}^{m} \lambda_{ik} b'_i$$

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 $(x_1, x_2, \ldots, x_n) \in K^n$ und steckt sie in die i-te Komponente des Resultats in K^m; die übrigen Komponenten sind null.

BSP
Basis von $\mathrm{Hom}(\mathbb{R}^3, \mathbb{R}^2)$

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:

$$\begin{aligned}
\phi_{11} &\colon (x,y,z) \longmapsto (x,0) \\
\phi_{12} &\colon (x,y,z) \longmapsto (y,0) \\
\phi_{13} &\colon (x,y,z) \longmapsto (z,0) \\
\phi_{21} &\colon (x,y,z) \longmapsto (0,x) \\
\phi_{22} &\colon (x,y,z) \longmapsto (0,y) \\
\phi_{23} &\colon (x,y,z) \longmapsto (0,z)
\end{aligned}$$

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

SATZ $\mathrm{End}(V)$ ist ein Ring

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.*

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 $\phi\colon \mathbb{R}^3 \to \mathbb{R}^2$

$$A = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}.$$ ♣

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

$$\Phi\colon\ \mathrm{Mat}(m\times n,K)=K^{\{1,2,\ldots,m\}\times\{1,2,\ldots,n\}}\to\mathrm{Hom}(K^n,K^m)$$

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

$$\phi(x_1,x_2,\ldots,x_n)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\phi_{ij}(x_1,x_2,\ldots,x_n)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_j\mathbf{e}'_i\,,$$

also ist die i-te Komponente von $\phi(x_1,x_2,\ldots,x_n)$ 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 $(x_1,\ldots,x_n)$ 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 $(x_1,\ldots,x_n)$ 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“!)

BSP

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:

$$\begin{pmatrix}a&b&c\\d&e&f\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}ax+by+cz\\dx+ey+fz\end{pmatrix}$$

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

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

LEMMA
Eigensch.
Matrixmult.

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:*

(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. ❑

DEF
Matrizen-
ring
invertierbare
Matrix

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. ♢

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.

BSP

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

$$\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\lambda \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$ ♣

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 mit $a_{ij} = 0$, 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 $\lambda_r = 0, \ldots, \lambda_2 = 0, \lambda_1 = 0$. 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:

$$\mathbf{0} = \sum_{j\in J} \lambda_j b_j = \sum_{j\in J} \lambda_j \mathbf{e}_j - \sum_{i=1}^{r} \Bigl(\sum_{j\in J} a_{ij}\lambda_j \Bigr) \mathbf{e}_{j_i}$$

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

$$\begin{array}{rcrcrcrcrcl} x_2 & + & 2x_3 & & & & & - & 2x_6 & = & 0 \\ & & & & x_4 & & & + & x_6 & = & 0 \\ & & & & & & x_5 & + & 5x_6 & = & 0 \end{array}$$

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) Zeilen-umformungen

(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 vom 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.

SATZ *
Normalform von Matrizen

12.7. Satz. *Seien K ein Körper, $m, n \in \mathbb{N}$ und $A \in \operatorname{Mat}(m \times n, K)$. Dann lässt sich A durch sukzessive elementare Zeilenumformungen in eine Matrix A' in reduzierter Zeilenstufenform überführen.*

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 $\mathbf{II}_{2,1}(-a_{2j_1}), \mathbf{II}_{3,1}(-a_{3j_1}), \ldots, \mathbf{II}_{m,1}(-a_{mj_1})$ 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 $k = 1, 2, \ldots, r$ und $i = 1, 2, \ldots, k-1$ 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.

12.8. **Beispiel.** Wir bestimmen die reduzierte Zeilenstufenform der folgenden Matrix über $\mathbb{R}$: **BSP** Umformung in ZSF

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix}$$

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 $M_{kl} = (\delta_{ik}\delta_{jl})_{1 \le i,j \le m}$ 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 $(\phi_{kl})_{1 \le k,l \le m}$ 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

$$M_{kl}A = \Bigl(\sum_{h=1}^{m} \delta_{ik}\delta_{hl}a_{hj}\Bigr)_{1 \le i \le m, 1 \le j \le n} = (\delta_{ik}a_{lj})_{1 \le i \le m, 1 \le j \le n}\,;$$

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 Elementarmatrizen erzeugen invertierbare Matrizen

Satz. *Jede invertierbare Matrix ist ein Produkt von Elementarmatrizen.*

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 $B = A\,Q$ 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 Eindeutigkeit der Zeilenstufenform

Satz. *Sind $A, B \in \mathrm{Mat}(m \times n, K)$ zwei Matrizen in reduzierter Zeilenstufenform mit demselben Zeilenraum, dann gilt $A = B$.*

Beweis. Sei $U \subset K^n$ der Zeilenraum von A und B. Für $0 \leq k \leq n$ sei

$$V_k = \{(x_1, \ldots, x_n) \in K^n \mid x_1 = x_2 = \ldots = x_k = 0\} \subset K^n$$

und $d_k = \dim(U \cap V_k)$. Dann ist $d_0 = r = \dim U$, $d_n = 0$ und $d_k - 1 \leq d_{k+1} \leq d_k$. Es gibt also genau r „Sprungstellen“ j_i mit $d_{j_i - 1} = r + 1 - i$ und $d_{j_i} = r - i$ für $i \in \{1, 2, \ldots, r\}$. 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

$$\phi \colon U \longrightarrow K^r, \qquad (x_1, x_2, \ldots, x_n) \longmapsto (x_{j_1}, x_{j_2}, \ldots, x_{j_r})$$

ist dann ein Isomorphismus, und die ersten r Zeilen von A und B müssen die Urbilder $\phi^{-1}(\mathbf{e}_1), \phi^{-1}(\mathbf{e}_2), \ldots, \phi^{-1}(\mathbf{e}_r)$ der Standard-Basisvektoren von K^r sein. Insbesondere sind A und B gleich. ❑

Wir kommen zu linearen Gleichungen und Gleichungssystemen.

DEF Lineare Gleichung

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*.

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 Unbekannten). ♢

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 $\ker(f)$.

BSP inhomogene lineare Diff.gleichung

12.11. **Beispiel.** Wir betrachten die folgende inhomogene lineare Differentialgleichung erster Ordnung:
$$y'(x) + y(x) = x\,.$$
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
$$z'(x) = e^x y(x) + e^x y'(x) = e^x\bigl(y(x) + y'(x)\bigr) = \mathbf{0}\,,$$
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 $\boldsymbol{b} = \mathbf{0}$ 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.

DEF erweiterte Matrix

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 $A' = (A \mid \boldsymbol{b})$ die erweiterte Matrix des linearen Gleichungssystems $A\boldsymbol{x} = \boldsymbol{b}$. Dann gilt*

SATZ inhom. LGS

$$\boldsymbol{b} \in \mathrm{im}(A) \iff \mathrm{rk}(A') = \mathrm{rk}(A)\,.$$

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

$$\tilde{A}\boldsymbol{x}_0 = \sum_{i=1}^{r} \tilde{b}_i \tilde{A} \mathbf{e}_{j_i} = \sum_{i=1}^{r} \tilde{b}_i \mathbf{e}'_i = \tilde{\boldsymbol{b}}\,.$$ ❑

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:

$$\begin{array}{rcrcrcrcr} x_1 & & & - & x_3 & & & = & 1 \\ & & x_2 & & & & & = & -1 \\ & & & & & & x_4 & = & -1 \end{array}$$

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 $1 \le j_1 < j_2 < \ldots < j_r \le n + 1$ 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 $j_1, j_2, \dots, j_r$ (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 $v_1, \dots, v_m \in K^n$ 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.**

BSP Transponierte Matrix

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

♣

Der Zeilenraum von A ist der Spaltenraum von $A^\top$ und umgekehrt. Satz 12.14 sagt also

$$\operatorname{rk}(A^\top) = \operatorname{rk}(A).$$

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:

$$A' \xrightarrow{\mathbf{I}_1(-1);\, \mathbf{II}_{2,1}(-2),\, \mathbf{II}_{3,1}(1),\, \mathbf{II}_{4,1}(-1)} \begin{pmatrix} \mathbf{1} & 2 & -2 & -2 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 3 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{\mathbf{II}_{1,2}(-2),\, \mathbf{II}_{3,2}(-1),\, \mathbf{II}_{4,2}(1)} \begin{pmatrix} \mathbf{1} & 0 & -2 & -2 & -5 & -2 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & -3 & -1 & 1 & 0 \\ 0 & 0 & 2 & 3 & 3 & 1 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{\mathbf{I}_3(-1);\, \mathbf{II}_{1,3}(2),\, \mathbf{II}_{4,3}(-2)} \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 1 & 0 & -2 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 1 & 3 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -3 & -1 & 2 & 1 \end{pmatrix}$$

$$\xrightarrow{\mathbf{II}_{3,4}(-1)} \begin{pmatrix} \mathbf{1} & 0 & 0 & 0 & 1 & 0 & -2 & 0 \\ 0 & \mathbf{1} & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & 6 & 2 & -3 & -1 \\ 0 & 0 & 0 & \mathbf{1} & -3 & -1 & 2 & 1 \end{pmatrix}$$

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.

$$\begin{tikzcd}
V \arrow[d, "f"'] & K^n \arrow[l, "\phi_B"'] \arrow[d, dashed, "\mathrm{Mat}_{B,B'}(f)"] \\
V' & K^m \arrow[l, "\phi_{B'}"]
\end{tikzcd}$$

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:

$$\begin{tikzcd}
(V,\tilde{B}) \arrow[r, "f"] \arrow[d, "\mathrm{id}_V"'] & (V',\tilde{B}') \\
(V,B) \arrow[r, "f"] & (V',B') \arrow[u, "\mathrm{id}_{V'}"']
\end{tikzcd}$$

❑

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. ❑

SATZ Gruppe der invertierbaren Matrizen

13.6. **Satz.** *Seien K ein Körper und $n \in \mathbb{N}$. Die Menge der invertierbaren Matrizen in $\operatorname{Mat}(n, K)$ bildet eine Gruppe unter der Matrixmultiplikation.*

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.

DEF $\operatorname{GL}(n, K)$

13.7. **Definition.** Die Gruppe der invertierbaren Matrizen in $\operatorname{Mat}(n, K)$ heißt *allgemeine lineare Gruppe* und wird mit $\operatorname{GL}(n, K)$ bezeichnet. ♢

Die Abkürzung „GL" kommt von englisch *general linear* group. Es ist auch die Notation $\operatorname{GL}_n(K)$ gebräuchlich.

SATZ ∗ Matrizen derselben lin. Abb.

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 $f : V \to V'$ linear. Seien weiter B eine Basis von V, B' eine Basis von V' und $A = \operatorname{Mat}_{B,B'}(f)$. Dann gilt: Die Menge der Matrizen von f bezüglich beliebiger Basen von V und V' ist genau*

$$\{PAQ \mid P \in \operatorname{GL}(m, K), Q \in \operatorname{GL}(n, K)\}\,.$$

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 $P \in \operatorname{GL}(m, K)$ und $Q \in \operatorname{GL}(n, K)$ Basen $\tilde{B}$ von V und $\tilde{B}'$ von V', sodass $P = \operatorname{Mat}_{B',\tilde{B}'}(\operatorname{id}_{V'})$ und $Q = \operatorname{Mat}_{\tilde{B},B}(\operatorname{id}_V)$. Dann ist

$$PAQ = \operatorname{Mat}_{B',\tilde{B}'}(\operatorname{id}_{V'})\operatorname{Mat}_{B,B'}(f)\operatorname{Mat}_{\tilde{B},B}(\operatorname{id}_V) = \operatorname{Mat}_{\tilde{B},\tilde{B}'}(f)$$

auch eine Matrix von f. ❑

DEF ∗ Äquivalenz von Matrizen

13.9. **Definition.** Seien K ein Körper, $m, n \in \mathbb{N}$ und $A, B \in \operatorname{Mat}(m \times n, K)$. Die Matrizen A und B heißen *äquivalent*, wenn es Matrizen $P \in \operatorname{GL}(m, K)$ und $Q \in \operatorname{GL}(n, K)$ gibt mit $PAQ = B$. ♢

Zwei Matrizen in $\operatorname{Mat}(m \times n, K)$ sind also genau dann äquivalent, wenn sie dieselbe lineare Abbildung (aber evtl. bezüglich verschiedener Basen) repräsentieren.

SATZ ∗
Klassifikation von Matrizen bis auf Äquivalenz

13.10. **Satz.** *Seien K ein Körper, $m, n \in \mathbb{N}$ und $A, B \in \operatorname{Mat}(m \times n, K)$. Die Matrizen A und B sind genau dann äquivalent, wenn $\operatorname{rk}(A) = \operatorname{rk}(B)$ ist. In diesem Fall sei $r = \operatorname{rk}(A)$; dann sind beide Matrizen äquivalent zur Matrix*

$$M_r = \left(\begin{array}{c|c} I_r & \mathbf{0}_{r,n-r} \\ \hline \mathbf{0}_{m-r,r} & \mathbf{0}_{m-r,n-r} \end{array}\right).$$

Dabei steht $\mathbf{0}_{k,l}$ für eine Nullmatrix mit k Zeilen und l Spalten.

Beweis. Sei $r = \operatorname{rk}(A)$. 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 $\operatorname{im}(f)$: Wir haben $\dim \operatorname{im}(f) = r$ und

$$\operatorname{im}(f) = \langle f(b_1), \ldots, f(b_r), f(b_{r+1}), \ldots, f(b_n)\rangle = \langle b'_1, \ldots, b'_r, \mathbf{0}, \ldots, \mathbf{0}\rangle = \langle b'_1, \ldots, b'_r\rangle;$$

$(b'_1, \ldots, b'_r)$ ist also ein Erzeugendensystem von $\operatorname{im}(f)$ der richtigen Länge r. Wir ergänzen diese Basis von $\operatorname{im}(f)$ zu einer Basis $B' = (b'_1, \ldots, b'_m)$ von K^m. Dann ist $\operatorname{Mat}_{B,B'}(f)$ gerade M_r; M_r ist damit äquivalent zu A.

Gilt auch $\operatorname{rk}(B) = r$, dann ist B ebenfalls äquivalent zu M_r. Es folgt, dass A und B äquivalent sind: Es gibt $P, P' \in \operatorname{GL}(m, K)$ und $Q, Q' \in \operatorname{GL}(n, K)$ mit $M_r = PAQ = P'BQ'$. Dann ist $B = (P'^{-1}P)A(QQ'^{-1})$.

Umgekehrt gilt $\operatorname{rk}(B) = r = \operatorname{rk}(A)$ 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 $A \in \operatorname{Mat}(m \times n, K)$ lässt sich durch Zeilen- und Spaltenumformungen in die Matrix M_r mit $r = \operatorname{rk}(A)$ 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 $\forall x \in X : x\,R\,x$,
- *symmetrisch*, wenn $\forall x, y \in X : x\,R\,y \Rightarrow y\,R\,x$, und
- *transitiv*, wenn $\forall x, y, z \in X : (x\,R\,y \ \wedge\ y\,R\,z) \Rightarrow x\,R\,z$.

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 $A, B \in \operatorname{Mat}(m \times n, K)$ schreiben wir $A \sim B$, wenn A und B äquivalent sind, wenn es also $P \in \operatorname{GL}(m, K)$ und $Q \in \operatorname{GL}(n, K)$ gibt mit $B = PAQ$.*

Die Relation $\sim$ ist eine Äquivalenzrelation auf $\operatorname{Mat}(m \times n, K)$.

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
Eigensch.
Äqu.rel.

Lemma. *Sei $\sim$ eine Äquivalenzrelation auf einer Menge X und sei $x \in X$. Dann sind für $y \in X$ die folgenden Aussagen äquivalent:*

(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:

$$\det\big((a_{ij})\big) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$ ♢

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.

$$\left|\begin{array}{ccc|cc} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{array}\right.$$

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 $d = \det\colon \operatorname{Mat}(n, K) \to K$ 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 $d\colon \operatorname{Mat}(n, K) \to K$ 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) *Es gilt $d(A) = \det(A)d(I_n)$ für alle $A \in \operatorname{Mat}(n, K)$. Insbesondere ist $\det\colon \operatorname{Mat}(n, K) \to K$ die einzige Abbildung, die (1), (2) und (3) erfüllt.*

Außerdem gilt

(8) $\det(A) \neq 0 \iff \operatorname{rk}(A) = n \iff A$ *invertierbar.*

Beweis. Wir beweisen zuerst (4)–(6).

(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 $1 \le i < j \le n$ 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

$$d(A') = d_{ij}(v_i + \lambda v_j, v_j) \stackrel{(1)}{=} d_{ij}(v_i, v_j) + \lambda d_{ij}(v_j, v_j) \stackrel{(2)}{=} d_{ij}(v_i, v_j) = d(A)\,.$$

(6) In der Notation des Beweises von Teil (5) haben wir

$$\begin{aligned} 0 &\stackrel{(2)}{=} d_{ij}(v_i + v_j, v_i + v_j) \\ &\stackrel{(1)}{=} d_{ij}(v_i, v_i) + d_{ij}(v_i, v_j) + d_{ij}(v_j, v_i) + d_{ij}(v_j, v_j) \\ &\stackrel{(2)}{=} d_{ij}(v_i, v_j) + d_{ij}(v_j, v_i)\,, \end{aligned}$$

also ist $d(A') = d_{ij}(v_j, v_i) = -d_{ij}(v_i, v_j) = -d(A)$.

Der Beweis der ersten drei Aussagen für $d = \det$ 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 $1 \le k < l \le n$ ist. 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 $a_{2j} = a_{1j}$)

$$\begin{aligned}
\det(A) &= \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \det(A_{1j}) \\
&= \sum_{j=1}^{n} (-1)^{j-1} a_{1j} \left(\sum_{m=1}^{j-1} (-1)^{m-1} a_{2m} d_{jm} + \sum_{m=j+1}^{n} (-1)^{m} a_{2m} d_{jm} \right) \\
&= \sum_{1 \le m < j \le n} (-1)^{j-m} a_{1j} a_{1m} d_{jm} + \sum_{1 \le j < m \le n} (-1)^{m-j-1} a_{1j} a_{1m} d_{jm} \\
&\overset{(*)}{=} \sum_{1 \le j < m \le n} (-1)^{m-j} a_{1j} a_{1m} d_{jm} + \sum_{1 \le j < m \le n} (-1)^{m-j-1} a_{1j} a_{1m} d_{jm} \\
&= \sum_{1 \le j < m \le n} \big((-1)^{m-j} + (-1)^{m-j-1} \big) a_{1j} a_{1m} d_{jm} \\
&= 0\,.
\end{aligned}$$

(Bei $(*)$ haben wir 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:

(7) Sei A' die reduzierte Zeilenstufenform von A. Aus (4), (5) und (6) 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.

(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

$$\begin{aligned} f(v_1, \ldots,&v_{i-1}, v_i + v_i', v_{i+1}, \ldots, v_m) \\ &= f(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_m) + f(v_1, \ldots, v_{i-1}, v_i', v_{i+1}, \ldots, v_m) \end{aligned}$$

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.**

$$\begin{vmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 2 & 0 & 2 \\ 0 & 3 & 3 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 6 & 6 \end{vmatrix}$$

$$= 2 \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 6 & 6 \end{vmatrix} = 2 \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 6 \end{vmatrix} = 2 \cdot 6 \cdot \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 12 \qquad ♣$$

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 sei $A = (a_{ij}) \in \mathrm{Mat}(n, K)$. 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.

(Alternativ könnte man auch die Zeilen 1 und i vertauschen und Satz 14.3 (6) verwenden.) ❑

14.6. **Beispiel.** Die Berechnung der Determinante in Beispiel 14.4 lässt sich vereinfachen, indem man nach der zweiten Zeile entwickelt: **BSP**

$$\begin{vmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{vmatrix} = - \begin{vmatrix} -1 & 1 & -1 \\ 1 & 1 & 1 \\ 2 & 4 & 8 \end{vmatrix} = \ldots = 12$$ ♣

* 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 sich ergibt als die k-te Zeile von A (als Zeilenvektor) multipliziert mit B. Insbesondere ist die k-te Zeile von AB eine lineare Funktion der k-ten Zeile von A und hängt nicht von den Einträgen in den anderen Zeilen von A ab. 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, 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 sei $A = (a_{ij}) \in \mathrm{Mat}(n, K)$. Mit der in Definition 14.1 eingeführten Schreibweise A_{ij} gilt für jedes $j \in \{1, \ldots, n\}$:* **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 $\det(A^\top) = \det(A)$. ❑

BSP

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?

Es gilt

$$1 = \det(I_n) = \det(AA^\top) = \det(A)\det(A^\top) = \det(A)^2 ,$$

also ist $\det(A) = \pm 1$. ♣

BSP Determinantenberechnung

14.11. **Beispiel.** Wir berechnen die Determinante aus Beispiel 14.4 noch einmal.

$$\begin{vmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \end{vmatrix} = - \begin{vmatrix} -1 & 1 & -1 \\ 1 & 1 & 1 \\ 2 & 4 & 8 \end{vmatrix} = - \begin{vmatrix} -1 & 1 & 0 \\ 1 & 1 & 0 \\ 2 & 4 & 6 \end{vmatrix} = -6 \cdot \begin{vmatrix} -1 & 1 \\ 1 & 1 \end{vmatrix}$$
$$= -6((-1) \cdot 1 - 1 \cdot 1) = 12$$

(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.

FOLG $\det(\lambda A)$

14.12. **Folgerung.** *Seien K ein Körper, $n \in \mathbb{N}$, $\lambda \in K$ und $A \in \mathrm{Mat}(n, K)$. Dann gilt*

$$\det(\lambda A) = \lambda^n \det(A) .$$

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.

DEF ∗ Adjungierte 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. ♢

SATZ ∗
Adjungierte Matrix

14.14. **Satz.** *Seien K ein Körper, $n > 0$ und $A \in \mathrm{Mat}(n, K)$. Dann gilt*

$$A\tilde{A} = \tilde{A}A = \det(A)I_n \,.$$

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

BSP
Inverse einer 2×2-Matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} .$$ ♣

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.

G.W. Leibniz (1646–1716)

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:

SATZ * Leibniz-Formel

14.17. **Satz.** *Sei $n \in \mathbb{N}$. Es gibt eine eindeutig bestimmte Abbildung*

$$\varepsilon\colon S_n \to \{-1, 1\}\,,$$

sodass für jeden Körper K und jede Matrix $A = (a_{ij}) \in \mathrm{Mat}(n, K)$ gilt

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

⚠

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.

DEF Signum einer Permutation

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. ◊

Um etwas über diese Vorzeichenfunktion herauszufinden, führen wir die zu σ gehörende Permutationsmatrix ein.

DEF Permutationsmatrix

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.* ♢

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.

LEMMA Eigensch. Permutationsmatrix

14.20. **Lemma.** *Seien K ein Körper und $n \in \mathbb{N}$.*

(1) *Für $\sigma, \tau \in S_n$ gilt $P(\sigma \circ \tau) = P(\sigma)P(\tau)$.*
(2) *Für $\sigma \in S_n$ gilt $\varepsilon(\sigma) = \det\bigl(P(\sigma)\bigr)$.*
(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 $\varepsilon(\sigma) = \pm 1$)

$$\det\bigl(P(\sigma)\bigr) = \det\bigl(P(\sigma)^{-1}\bigr)^{-1} = \det\bigl(P(\sigma^{-1})\bigr)^{-1} = \varepsilon(\sigma)^{-1} = \varepsilon(\sigma)\,.$$

(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:

DEF Fehlstand

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. ♢

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.

DEF * Orientierung einer Basis

14.21. **Definition.** Sei $(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$ eine Basis des $\mathbb{R}^n$ und $A = \big(\boldsymbol{b}_1 \mid \cdots \mid \boldsymbol{b}_n\big)$ 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. ♢

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.

DEF orientierungs-erhaltend, -umkehrend

14.22. **Definition.** Sei $f : \mathbb{R}^n \to \mathbb{R}^n$ 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. ♢

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:

DEF Parallelotop

14.23. **Definition.** Ein *Parallelotop* im $\mathbb{R}^n$ ist die Menge

$$P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) = \{t_1\boldsymbol{x}_1 + t_2\boldsymbol{x}_2 + \ldots + t_n\boldsymbol{x}_n \mid t_1, t_2, \ldots, t_n \in [0,1]\} \subset \mathbb{R}^n$$

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 orientiertes 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*

$$\operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) = \det(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\,,$$

wobei $\det(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ *für die Determinante der Matrix steht, deren Spalten die Vektoren* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ *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

$$\frac{1}{2}\Bigl|\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}\Bigr|\,.$$

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 $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^n$. Dann gilt*

$$\operatorname{vol} f\bigl(P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\bigr) = \det(A) \operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\,.$$

Die Determinante eines Endomorphismus gibt also an, mit welchem Faktor das orientierte Volumen bei seiner Anwendung multipliziert wird.

Beweis. Sei X die Matrix, deren Spalten die Vektoren $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ sind. Da f linear ist, gilt $f\bigl(P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\bigr) = P\bigl(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots, f(\boldsymbol{x}_n)\bigr)$. Es folgt

$$\begin{aligned} \operatorname{vol} f\bigl(P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\bigr) &= \det\bigl(f(\boldsymbol{x}_1), f(\boldsymbol{x}_2), \ldots, f(\boldsymbol{x}_n)\bigr) \\ &= \det(A\boldsymbol{x}_1, A\boldsymbol{x}_2, \ldots, A\boldsymbol{x}_n) \\ &= \det(AX) = \det(A)\det(X) \\ &= \det(A) \operatorname{vol} P(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\,. \end{aligned}$$

❑

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 $f: V \to W$ linear mit $\operatorname{rk}(f) = r$, 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* $f: V \to V$ 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 $f: V \to V$ 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 $\operatorname{Mat}_{B,B'}(\operatorname{id}_V)$. 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.

15.1. **Satz.** *Sei V ein K-Vektorraum mit Basis $B = (b_1, b_2, \ldots, b_n)$ und sei f ein Endomorphismus von V. Sei weiter $A = \operatorname{Mat}_{B,B}(f) \in \operatorname{Mat}(n, K)$ die Matrix von f bezüglich B. Dann ist* **SATZ** Matrizen eines Endomorphismus

$$\{\operatorname{Mat}_{B',B'}(f) \mid B' \text{ Basis von } V\} = \{PAP^{-1} \mid P \in \operatorname{GL}(n, K)\}.$$

Beweis. Es ist $\operatorname{Mat}_{B',B'}(f) = \operatorname{Mat}_{B,B'}(\operatorname{id}_V)\operatorname{Mat}_{B,B}(f)\operatorname{Mat}_{B',B}(\operatorname{id}_V) = PAP^{-1}$ mit $P = \operatorname{Mat}_{B,B'}(\operatorname{id}_V) \in \operatorname{GL}(n, K)$. Umgekehrt lässt sich jede Matrix $P \in \operatorname{GL}(n, K)$ in dieser Form schreiben (Folgerung 13.4). ❑

* 15.2. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$. Zwei Matrizen $A, A' \in \operatorname{Mat}(n, K)$ heißen *ähnlich*, wenn es eine Matrix $P \in \operatorname{GL}(n, K)$ gibt mit $A' = PAP^{-1}$. ◊ **DEF** Ähnlichkeit von Matrizen

Analog 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 $\lambda \in K\colon v \mapsto \lambda v$. Wir können uns fragen, ob es Elemente von V gibt, die sich unter f und dieser Abbildung gleich verhalten. Das führt auf folgende Definition.

DEF $\ast$ Eigenwert Eigenvektor

15.3. **Definition.** Seien V ein K-Vektorraum und $f \in \operatorname{End}(V)$. 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. $\diamondsuit$

Man beachte die Bedingung $v \neq \mathbf{0}$! Ohne sie wäre die Definition sinnlos, weil dann jedes λ ein Eigenwert wäre (denn $f(\mathbf{0}) = \mathbf{0} = \lambda\mathbf{0}$).

DEF $\ast$ Eigenraum geom. Vielfachheit

15.4. **Definition.** Seien V ein K-Vektorraum, $\lambda \in K$ und $f \in \operatorname{End}(V)$. Der Untervektorraum

$$E_\lambda(f) = \{v \in V \mid f(v) = \lambda v\} = \ker(\lambda \operatorname{id}_V - f)$$

von V heißt der λ-*Eigenraum* oder der *Eigenraum zum Eigenwert* λ von f.

Die Dimension $\dim E_\lambda(f)$ des λ-Eigenraums von f heißt die *geometrische Vielfachheit* des Eigenwerts λ von f. $\diamondsuit$

$E_\lambda(f)$ 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.

BSP Eigenwerte Eigenräume

15.5. **Beispiele.**

(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.) $\clubsuit$

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.

SATZ $\ast$ Lin. Unabh. von Eigenvektoren

15.6. **Satz.** *Seien V ein K-Vektorraum und $f \in \operatorname{End}(V)$. 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.*

Beweis. Wir verwenden Induktion über m. Der Fall $m = 0$ ist klar (null Vektoren sind stets linear unabhängig). Sei also $m > 0$, und die Behauptung für $m - 1$ sei schon bewiesen. 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. 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 $\alpha_m = 0$. ❑

15.7. **Folgerung.** *Seien V ein K-Vektorraum und $f \in \mathrm{End}(V)$. 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* **FOLG** Dimension von Eigenräumen

$$\dim E_{\lambda_1}(f) + \dim E_{\lambda_2}(f) + \ldots + \dim E_{\lambda_m}(f) \leq \dim V .$$

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 lineare Unabhängigkeit.

Im Fall, dass $\dim E_{\lambda_j}(f) = \infty$ ist für ein j, folgt die letzte Aussage daraus, dass $E_{\lambda_j}(f)$ ein Untervektorraum von V ist, woraus sich $\dim V = \infty$ ergibt. Anderenfalls können wir für jedes j das Tupel $(v_{j1}, v_{j2}, \ldots, v_{jn_j})$ als eine Basis von $E_{\lambda_j}(f)$ wählen. Wegen der linearen Unabhängigkeit der v_{ji} gilt dann

$$\begin{aligned}
\dim V &\geq \#\{v_{ji} \mid j \in \{1, 2, \ldots, m\}, i \in \{1, 2, \ldots, n_j\}\} \\
&= n_1 + n_2 + \ldots + n_m \\
&= \dim E_{\lambda_1}(f) + \dim E_{\lambda_2}(f) + \ldots + \dim E_{\lambda_m}(f) .
\end{aligned}$$ ❑

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.

DEF Eigenwert etc. für Matrizen

15.8. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$ 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

$$E_\lambda(A) = \{\boldsymbol{x} \in K^n \mid A\boldsymbol{x} = \lambda\boldsymbol{x}\} = \ker(\lambda I_n - A)$$

heißt der λ-*Eigenraum* von A oder 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.

LEMMA Charakterisierung von Eigenwerten

15.9. **Lemma.** *Seien K ein Körper, $\lambda \in K$, $n \in \mathbb{N}$ 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 $\dim\ker(\lambda I_n - A) = n - \mathrm{rk}(\lambda I_n - A)$.*

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 und dem Rangsatz 10.17. ❑

BSP Bestimmung der Eigenwerte

15.10. **Beispiel.** Wir betrachten

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathrm{Mat}(2, \mathbb{R}).$$

(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

$$\lambda I_2 - A = I_2 - A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \xrightarrow{\mathbf{II}_{2,1}(1)} \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix};$$

das liefert die (einelementige) Basis $\big((1, 1)\big)$ von $E_1(A)$. Für $\lambda = -1$ sieht es wie folgt aus:

$$\lambda I_2 - A = -I_2 - A = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \xrightarrow{\mathbf{I}_1(-1)} \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \xrightarrow{\mathbf{II}_{2,1}(1)} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix};$$

die Basis ist $\big((-1, 1)\big)$. ♣

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.

15.11. **Definition.** Sei K ein Körper. Ein *Polynom* in der *Variablen* (oder *Unbestimmten*) X über K ist ein Ausdruck der Form **DEF** Polynom Polynomring Grad Leitkoeffizient normiert

$$p = a_n X^n + a_{n-1} X^{n-1} + \ldots + a_1 X + a_0$$

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 $q = b_m X^m + \ldots + b_1 X + b_0$ ein weiteres Polynom. 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 $\deg(p + q) \le \max\{\deg(p), \deg(q)\}$. Das *Produkt* von p und q ist

$$p \cdot q = \sum_{k=0}^{m+n} \Bigl(\sum_{i,j \colon i+j=k} a_i b_j \Bigr) X^k \,;$$

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 $K[X] \to \operatorname{Abb}(K, K)$, 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. Diese Aussage ist die Verallgemeinerung des Resultats von Beispiel 10.15 von $\mathbb{R}$ auf beliebige Körper. Der Beweis geht genauso wie dort. ❑

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 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 $K[X] \to \operatorname{Abb}(K, K)$, 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 $K[X] \to \operatorname{Abb}(K, K)$. 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 $\dim_K \operatorname{Abb}(K, K) = q$, 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 $K[X] \to \operatorname{Abb}(K, K)$ 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 $K[X]$ mit $f = qg + r$ und $\deg(r) < \deg(g)$.* **SATZ** Polynom-division

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 $\deg(r) < 1 = \deg(X - \lambda)$ ist. 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.

DEF Vielfachheit einer Nullstelle

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 $q(\lambda) \neq 0$. ♢

λ ist also genau dann eine Nullstelle, wenn die Vielfachheit von λ als Nullstelle positiv ist.

BSP Vielfachheiten

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).

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 $-\boldsymbol{i}$. ♣

BSP Nullstellen

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.

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

$$\begin{aligned} &(\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn}) + \text{Terme mit} \le n-2 \text{ Faktoren } \lambda - a_{jj} \\ &= \lambda^n - (a_{11} + \ldots + a_{nn})\lambda^{n-1} + \ldots + (-1)^n \det(A)\,. \end{aligned}$$

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. $\diamondsuit$

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:

$$\lambda_2 = 2 + \sqrt{3} \qquad \text{und} \qquad \lambda_3 = 2 - \sqrt{3}\,.$$ ♣

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 $f \in \operatorname{End}(V)$. Sei B eine beliebige Basis von V und sei $A = \operatorname{Mat}_{B,B}(f)$. 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. $\diamondsuit$

Die Definition ist sinnvoll, weil sie nicht von der Wahl der Basis B abhängt: Ist $A' = \operatorname{Mat}_{B',B'}(f)$ mit einer anderen Basis B' von V, dann gibt es eine Matrix $P \in \operatorname{GL}(n, K)$ (mit $n = \dim V$), 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 $\ast$
algebraische Vielfachheit

15.24. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$, $A \in \operatorname{Mat}(n, K)$ 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. $\diamondsuit$

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. $\neq$ geom. Vielfachheit

15.25. **Beispiel.** Sei

$$A = \begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix} \in \operatorname{Mat}(2, K)\,.$$

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 $\operatorname{rk}(\alpha I_2 - A) = 1$), also hat α die geometrische Vielfachheit 1. $\clubsuit$

Die Vielfachheiten können also verschieden sein. Eine Beziehung gilt jedoch.

SATZ $\ast$
geom. $\leq$ alg. Vielfachheit

15.26. **Satz.** *Seien K ein Körper, V ein endlich-dimensionaler K-Vektorraum, $f \in \operatorname{End}(V)$ 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 $A \in \operatorname{Mat}(n, K)$.

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

$$A = \operatorname{Mat}_{B,B}(f) = \left(\begin{array}{c|c} \lambda I_m & D \\ \hline \mathbf{0}_{n-m,m} & C \end{array}\right)$$

mit Matrizen $D \in \mathrm{Mat}(m \times (n-m), K)$ und $C \in \mathrm{Mat}(n-m, K)$, denn für $j \in \{1, 2, \dots, m\}$ 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.

DEF ∗
Diagonalmatrix

16.1. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$. Eine Matrix $A = (a_{ij}) \in \operatorname{Mat}(n, K)$ ist eine *Diagonalmatrix* oder *diagonal*, wenn $a_{ij} = 0$ ist für alle $i, j \in \{1, 2, \dots, n\}$ mit $i \neq j$. Ist $a_{ii} = d_i$ für $i \in \{1, 2, \dots, n\}$, dann schreiben wir $\operatorname{diag}(d_1, d_2, \dots, d_n)$ für A:

$$\operatorname{diag}(d_1, d_2, \dots, d_n) = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}. \qquad \diamondsuit$$

LEMMA
Diagonalisierbarkeit

16.2. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$, V ein n-dimensionaler K-Vektorraum und $f \in \operatorname{End}(V)$. Dann sind die folgenden Aussagen äquivalent:*

(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 $A = \operatorname{Mat}_{B',B'}(f)$ die Matrix von f bezüglich einer beliebigen Basis B' von V. Dann ist A ähnlich zu einer Diagonalmatrix.*

Beweis.

„(1) ⇒ (2)“: Seien $\lambda_1, \lambda_2, \dots, \lambda_k \in K$ 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) ⇒ (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) ⇒ (3)“: Sei $B = (b_1, b_2, \dots, b_n)$ und sei α_j der Eigenwert von f, sodass $f(b_j) = \alpha_j b_j$. Dann ist

$$\operatorname{Mat}_{B,B}(f) = \operatorname{diag}(\alpha_1, \alpha_2, \dots, \alpha_n)$$

eine Diagonalmatrix, und nach Satz 15.1 ist A zu ihr ähnlich.

„(3) ⇒ (1)“: Ist A ähnlich zu $D = \operatorname{diag}(\alpha_1, \alpha_2, \dots, \alpha_n)$, dann gibt es nach Satz 15.1 eine Basis $B = (b_1, b_2, \dots, b_n)$ von V, sodass $D = \operatorname{Mat}_{B,B}(f)$ 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. ❑

∗ **DEF** diagonalisierbar

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.

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.

DEF Zerlegung in Linearfaktoren

16.4. **Definition.** Seien 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

$$p = \prod_{j=1}^{n} (X - \alpha_j) = (X - \alpha_1)(X - \alpha_2) \cdots (X - \alpha_n)$$

ist. ♢

Das bedeutet also, dass die Summe der Vielfachheiten der Nullstellen von p in K gleich dem Grad n von p ist.

∗ **FOLG** Charakterisierung von Diagonalisierbarkeit

16.5. **Folgerung.** *Seien K ein Körper und f ein Endomorphismus eines endlich-dimensionalen K-Vektorraums V. Dann sind die folgenden Aussagen äquivalent:*

(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 $n_j \le m_j$ (Satz 15.26) und $m_1 + m_2 + \ldots + m_k \le \deg(\chi_f) = n$:

$$\begin{aligned}
f \text{ diagonalisierbar} &\iff n_1 + n_2 + \ldots + n_k = n \\
&\iff m_1 + m_2 + \ldots + m_k = n \quad \text{und} \quad \forall j\colon m_j = n_j \\
&\iff \chi_f \text{ zerfällt in Linearfaktoren über } K \text{ und } \forall j\colon m_j = n_j.
\end{aligned}$$

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 $n_1 + n_2 + \ldots + n_k = n$. ❑

BSP diagonalisierbare Matrix

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 Nicht 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). ♣

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:

FOLG Faktorisierung von Polynomen über $\mathbb{C}$

16.8. **Folgerung.** *Sei $p \in \mathbb{C}[X]$ normiert. Dann zerfällt p in Linearfaktoren 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 $q \in \mathbb{C}[X]$ normiert 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:

$$p = (X - \alpha_1)q = (X - \alpha_1)(X - \alpha_2) \cdots (X - \alpha_{n+1}).$$ ❑

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:

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 **DEF** Spur einer Matrix

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \ldots + a_{nn}\,.$$

(„Tr“ von englisch *trace*.) ♢

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.* **LEMMA** Spur und Determinante durch Eigenwerte

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.

∗ 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 $\mathrm{Tr}(A^\top) = \mathrm{Tr}(A)$.* **SATZ** Eigensch. der Spur

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:

$$E_1(A) = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle \quad \text{und} \quad E_{-1}(A) = \langle \begin{pmatrix} -1 \\ 1 \end{pmatrix} \rangle\,.$$

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 Grundkörpers in ein Polynom einsetzen kann (Definition 15.12). Man kann aber auch allgemeinere Objekte in Polynome einsetzen, zum Beispiel Matrizen oder Endomorphismen.

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

DEF Einsetzen von Matrizen und Endomorphismen in Polynome

$$p = a_0 + a_1 X + a_2 X^2 + \ldots + a_m X^m ,$$

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 notieren 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.

16.16. **Lemma.** *In der Situation von Definition 16.15 gilt für $p, q \in K[X]$:*

LEMMA Einsetzungshomomorphismus

$$(p + q)(A) = p(A) + q(A) \quad \textit{und} \quad (p \cdot q)(A) = p(A) \cdot q(A)$$

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 $\phi(1_R) = 1_{R'}$ ist 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.

LEMMA Existenz von $p \in K[X]$ mit $p(A) = \mathbf{0}$

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.*

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:

LEMMA Struktur von $P(A)$

16.19. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Sei*

$$P(A) = \{p \in K[X] \mid p(A) = \mathbf{0}\} \,.$$

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 $P(A)$.

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 kleinsten Grades in $P(A)$, und* **SATZ Minimalpolynom**

$$P(A) = \{q \cdot m_A \mid q \in K[X]\}$$

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 $0 \le \deg(p - m_A) < \deg(m_A)$. 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.

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. ♢ **DEF Minimalpolynom**

16.22. **Beispiele.** Seien $n \ge 1$ und $A \in \mathrm{Mat}(n, K)$. **BSP Minimalpolynome**

(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?

Es stellt sich heraus, dass beide Fragen mit „Ja“ beantwortet werden können. Das werden wir zu Beginn des Sommersemesters sehen.

Lineare Algebra II

Sommersemester 2021

Wir kommen zurück auf die beiden Fragen vom Ende des Wintersemesters:

- Für A diagonalisierbar gilt, dass χ_A ein Vielfaches von m_A ist. Ist das auch allgemein richtig?
- Gilt die Umkehrung der Beobachtung in Beispiel 16.22 (3): Wenn m_A in Linearfaktoren zerfällt und keine mehrfachen Nullstellen hat, muss dann A diagonalisierbar sein?

Dabei ist χ_A das charakteristische Polynom und m_A das Minimalpolynom von A.

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 $A \in \mathrm{Mat}(n, K)$. Die Matrix A ist 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. Die Richtung „⇒“ wurde in Beispiel 16.22 gezeigt. Wir zeigen noch „⇐“. 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 $\deg(p_j) \le m - 1$. 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:

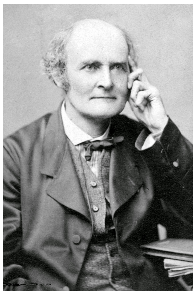
BSP
Cayley-Hamilton für $n = 2$

16.24. **Beispiel.** Sei $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{Mat}(2, K)$. Dann ist

$$\chi_A = X^2 - (a+d)X + ad - bc\,,$$

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

Aus $\chi_A(A) = \mathbf{0}$ folgt, dass χ_A ein Vielfaches von m_A ist. ♣

Wir beweisen diese Aussage, die als „Satz von Cayley-Hamilton“ bekannt ist, jetzt allgemein.

SATZ ∗
Cayley-Hamilton

16.25. **Satz.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Dann ist*

$$\chi_A(A) = \mathbf{0}\,.$$

Warum ist der folgende „Beweis“ nicht korrekt?

$$\chi_A(X) = \det(XI_n - A) \implies \chi_A(A) = \det(AI_n - A) = \det(\mathbf{0}) = 0\,.$$

W.R. Hamilton
1805–1865

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.

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

$$\tilde{B} = \tilde{B}_{n-1}X^{n-1} + \ldots + \tilde{B}_1 X + \tilde{B}_0$$

mit $\tilde{B}_k = (b_{ij}^{(k)})_{i,j} \in \mathrm{Mat}(n, K)$. Sei außerdem

$$\chi_A(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_1 X + a_0 \,.$$

Wir setzen in (16.1) ein und erhalten

$$\begin{aligned} &\tilde{B}_{n-1}X^n + (\tilde{B}_{n-2} - \tilde{B}_{n-1}A)X^{n-1} + \ldots + (\tilde{B}_1 - \tilde{B}_2 A)X^2 + (\tilde{B}_0 - \tilde{B}_1 A)X - \tilde{B}_0 A \\ &\quad = I_n X^n + a_{n-1}I_n X^{n-1} + \ldots + a_2 I_n X^2 + a_1 I_n X + a_0 I_n \,. \end{aligned} \tag{16.2}$$

Koeffizientenvergleich zeigt dann

$$\tilde{B}_{n-1} = I_n \,,\ \tilde{B}_{n-2} - \tilde{B}_{n-1}A = a_{n-1}I_n \,,\ \ldots\,,\ \tilde{B}_0 - \tilde{B}_1 A = a_1 I_n \,,\ -\tilde{B}_0 A = a_0 I_n \,.$$

Wir multiplizieren diese Gleichungen von rechts mit A^n, A^{n-1}, $\ldots$, A, I_n und summieren auf:

$$\begin{array}{rcl} \tilde{B}_{n-1}A^n \qquad\qquad\qquad\qquad\qquad & = & A^n \\ -\tilde{B}_{n-1}A^n + \tilde{B}_{n-2}A^{n-1} \qquad\qquad\qquad & = & a_{n-1}A^{n-1} \\ \ddots \qquad\quad \ddots \qquad\qquad\qquad & & \vdots \\ \ddots \qquad \ddots \qquad & & \vdots \\ -\tilde{B}_1 A^2 + \tilde{B}_0 A & = & a_1 A \\ -\tilde{B}_0 A & = & a_0 I_n \\ \hline \mathbf{0} & = & \chi_A(A) \end{array}$$

(Das läuft darauf hinaus, in (16.2) A für X einzusetzen. Dass man das auch „darf“ (d.h., dass dabei wieder eine gültige Gleichung herauskommt), müsste man aber begründen.) ❑

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.

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.* **FOLG** m_A teilt χ_A und hat dieselben Nullstellen

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
Minimal-
polynom

16.27. **Beispiel.** Wir bestimmen das Minimalpolynom von

$$A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix} \in \operatorname{Mat}(3, \mathbb{R})\,.$$

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). ♣

17. Summen von Untervektorräumen, Komplemente, Kodimension

Eines unserer Ziele in diesem Semester ist die Vervollständigung der Klassifikation der Endomorphismen eines endlich-dimensionalen Vektorraums V (oder äquivalent: der Klassifikation von $n \times n$-Matrizen bis auf Ähnlichkeit) im Fall, dass das charakteristische Polynom in Linearfaktoren zerfällt (was über einem algebraisch abgeschlossenen Körper wie $\mathbb{C}$ immer der Fall ist). Das wird auf die sogenannte „Jordan-Normalform" führen. In diesem Zusammenhang (aber auch an anderen Stellen) wird es hilfreich sein, den Vektorraum V zu „zerlegen", sodass der Endomorphismus auf den einzelnen „Teilen" von V ein leicht überschaubares Verhalten zeigt. Dafür brauchen wir den Begriff der „direkten Summe" von (Unter-)Vektorräumen.

Sei V ein Vektorraum. Wir erinnern uns daran, dass beliebige Durchschnitte von Untervektorräumen von V wieder Untervektorräume sind (Lemma 7.3), dass das im Allgemeinen aber nicht für Vereinigungen von Untervektorräumen gilt. Stattdessen können wir aber den kleinsten Untervektorraum betrachten, der alle betrachteten Untervektorräume (und damit ihre Vereinigung) enthält. Das führt auf folgende Definition.

* 17.1. **Definition.** Seien V ein Vektorraum und $(U_i)_{i\in I}$ eine Familie von Untervektorräumen von V. Dann heißt der von der Vereinigung $\bigcup_{i\in I} U_i$ erzeugte Untervektorraum von V die *Summe* der Untervektorräume U_i; wir schreiben

DEF Summe von Unter-VR

$$\sum_{i\in I} U_i = \Big\langle \bigcup_{i\in I} U_i \Big\rangle .$$

Im Fall $I = \{1, 2, \ldots, n\}$ schreibt man auch

$$U_1 + U_2 + \ldots + U_n \qquad \text{oder} \qquad \sum_{i=1}^{n} U_i$$

statt $\sum_{i\in I} U_i$. ♢

Die Schreibweise und der Name erklären sich durch die folgende Eigenschaft.

17.2. **Lemma.** *Sei V ein Vektorraum.*

LEMMA Elemente der Summe

(1) *Sind $U_1, U_2, \ldots, U_n$ Untervektorräume von V, dann ist*
$$U_1 + U_2 + \ldots + U_n = \{u_1 + u_2 + \ldots + u_n \mid u_1 \in U_1, u_2 \in U_2, \ldots, u_n \in U_n\} .$$

(2) *Ist $(U_i)_{i\in I}$ eine Familie von Untervektorräumen von V, dann ist*
$$\sum_{i\in I} U_i = \Big\{ \sum_{i\in J} u_i \mid J \subset I \text{ endlich, } u_i \in U_i \text{ für alle } i \in J \Big\} .$$

Beweis. Es ist klar, dass die jeweils rechts stehende Menge in der links stehenden enthalten ist, denn ihre Elemente sind Linearkombinationen von Elementen von $U_1 \cup \ldots \cup U_n$ bzw. $\bigcup_{i\in I} U_i$ (Satz 7.9). Wir zeigen, dass die rechts stehende Menge ein Untervektorraum von V ist. Dann folgt analog zum Beweis von Satz 7.9, dass sie das Erzeugnis der Vereinigung der U_i ist. Sei U die Menge auf der rechten Seite. Wir prüfen die drei Bedingungen für einen Untervektorraum nach (Definition 6.1). Dabei nutzen wir aus, dass die U_i Untervektorräume sind.

- $\mathbf{0} \in U$: Wir können alle $u_i = \mathbf{0}$ wählen (bzw. im zweiten Fall für J die leere Menge nehmen).

- Abgeschlossenheit unter der Addition: Im ersten Fall seien

$$u = u_1 + u_2 + \ldots + u_n \quad \text{und} \quad u' = u'_1 + u'_2 + \ldots + u'_n$$

zwei Elemente von U (mit $u_i, u'_i \in U_i$ für alle $i \in \{1, 2, \ldots, n\}$). Dann ist auch

$$u + u' = (u_1 + u'_1) + (u_2 + u'_2) + \ldots + (u_n + u'_n) \in U\,.$$

Im zweiten Fall seien $J, J' \subset I$ endlich und

$$u = \sum_{i\in J} u_i \qquad \text{und} \qquad u' = \sum_{i\in J'} u'_i$$

zwei Elemente von U. Wenn wir $u_i = \mathbf{0}$ (bzw. $u'_i = \mathbf{0}$) setzen für $i \in J' \setminus J$ (bzw. $i \in J \setminus J'$), dann gilt

$$u + u' = \sum_{i\in J} u_i + \sum_{i\in J'} u'_i = \sum_{i\in J\cup J'} u_i + \sum_{i\in J\cup J'} u'_i = \sum_{i\in J\cup J'} (u_i + u'_i) \in U\,.$$

- Abgeschlossenheit unter der Skalarmultiplikation: Seien λ ein Skalar und $u = u_1 + u_2 + \ldots + u_n$ bzw. $u = \sum_{i\in J} u_i$ ein Element von U. Dann ist

$$\lambda u = \lambda u_1 + \lambda u_2 + \ldots + \lambda u_n \quad \text{bzw.} \quad \lambda u = \sum_{i\in J} \lambda u_i$$

wieder ein Element von U. ❑

BSP
Summen von Unter-VR

17.3. **Beispiele.**

(1) Ist $I = \emptyset$, dann ist $\sum_{i\in I} U_i = \{\mathbf{0}\}$ der Null-Vektorraum.

(2) Ist $U \subset V$ ein Untervektorraum, dann gilt $U + U = U$.

(3) Ist V ein Vektorraum, I eine Menge und sind (für $i \in I$) $A_i \subset V$ beliebige Teilmengen, dann gilt

$$\sum_{i\in I} \langle A_i \rangle = \Big\langle \bigcup_{i\in I} A_i \Big\rangle\,.$$

(Beweis als Übung.) Das bedeutet: Ist A_i ein Erzeugendensystem von U_i (für alle $i \in I$), dann ist $\bigcup_{i\in I} A_i$ ein Erzeugendensystem von $\sum_{i\in I} U_i$. ♣

Was kann man über die Dimension von $U = U_1 + U_2$ sagen? Da $U_1 \subset U$ und $U_2 \subset U$ ist, gilt jedenfalls

$$\dim U \geq \max\{\dim U_1, \dim U_2\}$$

(und Gleichheit ist möglich, nämlich genau dann, wenn $U_1 \subset U_2$ oder $U_2 \subset U_1$ gilt). Wie groß kann $\dim U$ höchstens werden? Wenn B_1 eine Basis von U_1 und B_2 eine Basis von U_2 ist, dann ist nach dem obigen Beispiel $B_1 \cup B_2$ ein Erzeugendensystem von U, also gilt

$$\dim U \leq \#(B_1 \cup B_2) \leq \#B_1 + \#B_2 = \dim U_1 + \dim U_2\,.$$

Der folgende Satz gibt genauere Auskunft.

SATZ ∗
Dimension der Summe

17.4. **Satz.** *Sei V ein Vektorraum mit Untervektorräumen U_1 und U_2. Dann gilt*

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

Daran sieht man, dass $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$ genau dann gilt, wenn U_1 und U_2 den kleinstmöglichen Durchschnitt $\{\mathbf{0}\}$ haben (vorausgesetzt, alle Dimensionen sind endlich).

Beweis. Ist $\dim U_1 = \infty$ oder $\dim U_2 = \infty$, dann ist auch $\dim(U_1 + U_2) = \infty$ (denn $U_1, U_2 \subset U_1 + U_2$), und die Gleichung stimmt. Wir können also annehmen, dass U_1 und U_2 endlich-dimensional sind, etwa $\dim U_1 = n_1$ und $\dim U_2 = n_2$. Sei $m = \dim(U_1 \cap U_2) \le \min\{n_1, n_2\}$. Wir wählen eine Basis $(b_1, b_2, \ldots, b_m)$ von $U_1 \cap U_2$, die wir einerseits zu einer Basis $(b_1, \ldots, b_m, b'_{m+1}, b'_{m+2}, \ldots, b'_{n_1})$ von U_1 und andererseits zu einer Basis $(b_1, \ldots, b_m, b''_{m+1}, b''_{m+2}, \ldots, b''_{n_2})$ von U_2 ergänzen (Basisergänzungssatz mit Folgerung 9.7). Ich behaupte, dass

$$B = (b_1, \ldots, b_m, b'_{m+1}, b'_{m+2}, \ldots, b'_{n_1}, b''_{m+1}, b''_{m+2}, \ldots, b''_{n_2})$$

eine Basis von $U_1 + U_2$ ist. Daraus folgt die Gleichung im Satz, denn

$$\dim(U_1 + U_2) = \#B = m + (n_1 - m) + (n_2 - m) = n_1 + n_2 - m \,.$$

Es bleibt die Behauptung zu zeigen. Es ist klar, dass B ein Erzeugendensystem von $U_1 + U_2$ ist, denn B enthält Erzeugendensysteme von U_1 und von U_2. Wir müssen also noch nachweisen, dass B linear unabhängig ist. Seien also λ_i (für $i \in \{1, 2, \ldots, m\}$), λ'_i (für $i \in \{m+1, \ldots, n_1\}$) und λ''_i (für $i \in \{m+1, \ldots, n_2\}$) Skalare mit

$$\lambda_1 b_1 + \ldots + \lambda_m b_m + \lambda'_{m+1} b'_{m+1} + \ldots + \lambda'_{n_1} b'_{n_1} + \lambda''_{m+1} b''_{m+1} + \ldots + \lambda''_{n_2} b''_{n_2} = \mathbf{0} \,.$$

Wir schreiben diese Gleichung als

$$u = \underbrace{\lambda_1 b_1 + \ldots + \lambda_m b_m + \lambda'_{m+1} b'_{m+1} + \ldots + \lambda'_{n_1} b'_{n_1}}_{\in U_1} = \underbrace{-\lambda''_{m+1} b''_{m+1} - \ldots - \lambda''_{n_2} b''_{n_2}}_{\in U_2} \,.$$

Wir sehen, dass $u \in U_1 \cap U_2$ ist, also ist u eine Linearkombination von $b_1, \ldots, b_m$. Da $b_1, \ldots, b_m, b'_{m+1}, \ldots, b'_{n_1}$ und $b_1, \ldots, b_m, b''_{m+1}, \ldots, b''_{n_2}$ jeweils linear unabhängig sind (als Basen von U_1 und U_2), müssen

$$\lambda'_{m+1} = \ldots = \lambda'_{n_1} = \lambda''_{m+1} = \ldots = \lambda''_{n_2} = 0$$

sein; daraus folgt dann auch $u = \mathbf{0}$ und damit $\lambda_1 = \ldots = \lambda_m = 0$. ❑

Man kann sich die Aussage ganz gut mit Hilfe der analogen Aussage über Kardinalitäten von Mengen merken:

$$\#(M_1 \cup M_2) + \#(M_1 \cap M_2) = \#M_1 + \#M_2 \,.$$

Tatsächlich beruht obiger Beweis auf dieser Relation, wobei die Mengen Basen der vorkommenden Untervektorräume sind. Allerdings darf man diese Analogie auch nicht zu weit treiben: Die für Mengen gültige Relation

$$\begin{aligned} \#(M_1 \cup M_2 \cup M_3) + \#(M_1 \cap M_2) + \#(M_1 \cap M_3) + \#(M_2 \cap M_3) \\ = \#M_1 + \#M_2 + \#M_3 + \#(M_1 \cap M_2 \cap M_3) \end{aligned}$$

übersetzt sich nicht in eine analoge Dimensionsformel (Übung).

Besonders interessant ist der Fall $U_1 \cap U_2 = \{\mathbf{0}\}$.

LEMMA Summe direkt

17.5. **Lemma.** *Sei V ein Vektorraum mit Unterverktorräumen U_1 und U_2. Dann sind die folgenden Aussagen äquivalent:*

(1) $U_1 \cap U_2 = \{\mathbf{0}\}$.

(2) *Jedes Element $u \in U_1 + U_2$ lässt sich* ***eindeutig*** *schreiben als $u = u_1 + u_2$ mit $u_1 \in U_1$ und $u_2 \in U_2$.*

Ist $U_1 + U_2$ endlich-dimensional, dann sind beide Aussagen äquivalent zu

(3) $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$.

Beweis. Die Äquivalenz von (1) und (3) (unter der angegebenen Voraussetzung) folgt direkt aus der Dimensionsformel in Satz 17.4. Wir zeigen noch die Äquivalenz von (1) und (2).

Es gelte (1) und es sei $u = u_1 + u_2 = u_1' + u_2'$ mit $u_1, u_1' \in U_1$ und $u_2, u_2' \in U_2$. Daraus folgt $u_1 - u_1' = u_2' - u_2 \in U_1 \cap U_2 = \{\mathbf{0}\}$, also $u_1 = u_1'$ und $u_2 = u_2'$.

Jetzt gelte (2) und es sei $u \in U_1 \cap U_2$. Dann sind $\mathbf{0} = \mathbf{0} + \mathbf{0} = u + (-u)$ zwei Darstellungen des Nullvektors; aus der Eindeutigkeit der Summendarstellung folgt also $u = \mathbf{0}$. ❑

DEF direkte Summe zweier UVR

17.6. **Definition.** Wenn die Aussagen (1) und (2) in Lemma 17.5 gelten, dann heißt die Summe von U_1 und U_2 *direkt.* ♢

Die Eigenschaften (1) und (2) in Lemma 17.5 lassen sich auch so ausdrücken:

$$\forall u_1 \in U_1, u_2 \in U_2 \colon \big(u_1 + u_2 = \mathbf{0} \implies u_1 = u_2 = \mathbf{0}\big)\,.$$

In dieser Form lässt sich das verallgemeinern.

DEF ∗ direkte Summe

17.7. **Definition.** Seien V ein Vektorraum und $(U_i)_{i \in I}$ eine Familie von Untervektorräumen von V. Dann heißt die Summe der U_i *direkt*, wenn für jede endliche Teilmenge $J \subset I$ und beliebige Elemente $u_i \in U_i$ für $i \in J$ gilt

$$\sum_{i \in J} u_i = \mathbf{0} \implies \forall i \in J \colon u_i = \mathbf{0}\,.$$

Ist $V = \sum_{i \in I} U_i$ und die Summe direkt, dann schreiben wir auch

$$V = \bigoplus_{i \in I} U_i$$

bzw. $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$, wenn $I = \{1, 2, \ldots, n\}$ ist. ♢

Lemma 17.5 hat dann die folgende Verallgemeinerung.

LEMMA direkte Summe

17.8. **Lemma.** *Seien V ein Vektorraum und $(U_i)_{i \in I}$ eine Familie von Untervektorräumen von V. Dann sind die folgenden Aussagen äquivalent:*

(1) *Für jedes $i \in I$ gilt $U_i \cap \sum_{j \in I \setminus \{i\}} U_j = \{\mathbf{0}\}$.*

(2) *Die Summe der U_i ist direkt.*

Ist I endlich und $\sum_{i \in I} U_i$ endlich-dimensional, dann sind die Aussagen äquivalent zu

(3) $\dim \sum_{i \in I} U_i = \sum_{i \in I} \dim U_i$.

Beweis. „(1)⇒(2)“: Sei $J \subset I$ endlich und seien $u_i \in U_i$ für $i \in J$ mit $\sum_{i\in J} u_i = \mathbf{0}$. Sei $i_0 \in J$. Dann ist

$$u_{i_0} = \sum_{i\in J\setminus\{i_0\}} (-u_i) \in U_{i_0} \cap \sum_{i\in I\setminus\{i_0\}} U_i = \{\mathbf{0}\}\,,$$

also ist $u_{i_0} = \mathbf{0}$. Da $i_0 \in J$ beliebig war, müssen alle $u_i = \mathbf{0}$ sein; damit ist die Summe direkt.

„(2)⇒(1)“: Sei $i \in I$ und $u \in U_i \cap \sum_{j\in I\setminus\{i\}} U_j$. Dann gibt es $J \subset I \setminus \{i\}$ endlich und Elemente $u_j \in U_j$ für $j \in J$, sodass

$$\sum_{j\in J} u_j = u\,, \qquad \text{also} \qquad (-u) + \sum_{j\in J} u_j = \mathbf{0}$$

ist, wobei $-u$ als Element von U_i betrachtet wird. Definition 17.7 besagt dann, dass $u = \mathbf{0}$ sein muss.

Wir nehmen jetzt an, dass I endlich und $\sum_{i\in I} U_i$ endlich-dimensional ist.

„(2)⇒(3)“: Ist $\sum_{i\in I} U_i$ endlich-dimensional, dann gilt das auch für alle U_i (denn sie sind in der Summe enthalten). Für jedes $i \in I$ sei B_i eine Basismenge von U_i, dann ist $B = \bigcup_{i\in I} B_i$ ein (endliches) Erzeugendensystem von $\sum_{i\in I} U_i$.

B ist linear unabhängig: Sei $B_i = \{b_{i1}, b_{i2}, \ldots, b_{im_i}\}$ mit $m_i = \dim U_i$ und seien λ_{ij} Skalare mit

$$\sum_{i\in I} \underbrace{\sum_{j=1}^{m_i} \lambda_{ij} b_{ij}}_{\in U_i} = \mathbf{0}\,.$$

Weil die Summe direkt ist, folgt $\sum_{j=1}^{m_i} \lambda_{ij} b_{ij} = \mathbf{0}$ für alle $i \in I$ und dann $\lambda_{ij} = 0$ für alle $j \in \{1, 2, \ldots, m_i\}$, weil B_i eine Basis ist. Als linear unabhängiges Erzeugendensystem ist B eine Basis von $\sum_{i\in I} U_i$, also gilt

$$\dim \sum_{i\in I} U_i = \#B \overset{(*)}{=} \sum_{i\in I} \#B_i = \sum_{i\in I} \dim U_i\,.$$

Für $(*)$ braucht man noch, dass die B_i paarweise disjunkt sind, also $B_i \cap B_j = \emptyset$ ist für $i \neq j$. Das folgt aber aus der obigen Überlegung, denn mit $b \in B_i \cap B_j$ hätte man eine nichttriviale Linearkombination $1 \cdot b + (-1) \cdot b = \mathbf{0}$, wobei das erste b als Element von B_i und das zweite b als Element von B_j betrachtet wird.

„(3)⇒(1)“: Es gilt für jedes $i \in I$

$$\begin{aligned}\sum_{j\in I} \dim U_j &\overset{(3)}{=} \dim \sum_{j\in I} U_j = \dim\Bigl(U_i + \sum_{j\in I\setminus\{i\}} U_j\Bigr) \\ &\overset{(*)}{\leq} \dim U_i + \dim \sum_{j\in I\setminus\{i\}} U_j \leq \dim U_i + \sum_{j\in I\setminus\{i\}} \dim U_j = \sum_{j\in I} \dim U_j\,,\end{aligned}$$

also muss bei $(*)$ Gleichheit herrschen. Aus Satz 17.4 folgt dann

$$U_i \cap \sum_{j\in I\setminus\{i\}} U_j = \{\mathbf{0}\}\,. \qquad \square$$

17.9. **Beispiel.** Hier ist ein Beispiel, das zeigt, dass die vielleicht erst einmal näher liegende Version „$\forall i, j \in I : i \neq j \Rightarrow U_i \cap U_j = \{\mathbf{0}\}$“ für Bedingung (1) nicht ausreichend ist. **BSP**

Seien $V = \mathbb{R}^2$, $U_1 = \langle(1,0)\rangle$, $U_2 = \langle(0,1)\rangle$ und $U_3 = \langle(1,1)\rangle$. Dann gilt offenbar

$$U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{(0,0)\},$$

aber die Summe $U_1 + U_2 + U_3$ ist *nicht* direkt. Zum Beispiel gilt

$$(0,0) = (1,0) + (0,1) + (-1,-1)$$

als Summe je eines Elements von U_1, U_2 und U_3. ♣

Eine Zerlegung von V als direkte Summe, $V = U_1 \oplus U_2$, führt in natürlicher Weise zu zwei linearen Abbildungen $\pi_1 : V \to U_1$ und $\pi_2 : V \to U_2$. Wir erhalten sie wie folgt: Jedes $v \in V$ lässt sich eindeutig schreiben als $v = u_1 + u_2$ mit $u_1 \in U_1$ und $u_2 \in U_2$. Dann ist $\pi_1(v) = u_1$ und $\pi_2(v) = u_2$. Diese Abbildungen sind linear, weil $\lambda v = \lambda u_1 + \lambda u_2$ ist, und für $v' = u'_1 + u'_2$ gilt $v + v' = (u_1 + u'_1) + (u_2 + u'_2)$. Außerdem sind π_1 und π_2 surjektiv, denn $\pi_1|_{U_1} = \mathrm{id}_{U_1}$ und $\pi_2|_{U_2} = \mathrm{id}_{U_2}$.

DEF
Projektionen

17.10. **Definition.** Die Abbildungen π_1 und π_2 heißen die *Projektionen* von V auf U_1 bzw. U_2 bezüglich der Zerlegung $V = U_1 \oplus U_2$. ♢

Wenn wir mit $p_1, p_2 : V \to V$ die Abbildungen bezeichnen, die durch $p_i(v) = \pi_i(v)$ gegeben sind (sie unterscheiden sich von π_1 und π_2 nur durch den vergrößerten Wertebereich), dann gilt

$$p_1 \circ p_1 = p_1\,, \quad p_2 \circ p_2 = p_2\,, \quad \text{sowie} \quad p_1 \circ p_2 = p_2 \circ p_1 = \mathbf{0} \quad \text{und} \quad p_1 + p_2 = \mathrm{id}_V\,.$$

DEF
Projektor

Umgekehrt gilt: Ist $p : V \to V$ ein *Projektor*, d.h. eine lineare Abbildung mit $p \circ p = p$, dann gilt $V = \mathrm{im}(p) \oplus \ker(p)$, wobei $\ker(p) = \mathrm{im}(\mathrm{id}_V - p)$ ist (Übung). Mit $p' = \mathrm{id}_V - p$ gilt dann auch $p' \circ p' = p'$, $p \circ p' = p' \circ p = \mathbf{0}$ und $p + p' = \mathrm{id}_V$.

Wir führen jetzt noch zwei Begriffe ein, die manchmal nützlich sind.

DEF ∗
Komplement

17.11. **Definition.** Seien V ein Vektorraum und $U \subset V$ ein Untervektorraum. Ein weiterer Untervektorraum $U' \subset V$ heißt *komplementär* zu U oder ein *Komplement* von U in V, wenn $V = U \oplus U'$ gilt. ♢

Konkret bedeutet die Bedingung, dass $U \cap U' = \{\mathbf{0}\}$ und $U + U' = V$ gilt.

BSP
Komplemente

17.12. **Beispiele.**

(1) V ist das einzige Komplement von $\{\mathbf{0}\}$ in V und $\{\mathbf{0}\}$ ist das einzige Komplement von V in V.
(2) Normalerweise gibt es aber viele Komplemente. Sei zum Beispiel $V = \mathbb{R}^2$ und $U = \langle(1,0)\rangle \subset V$. Dann sind die Komplemente von U gerade alle Untervektorräume der Form $U' = \langle(a,1)\rangle$ mit $a \in \mathbb{R}$ beliebig. ♣

Gibt es immer ein Komplement?

SATZ Existenz von Komplementen

17.13. **Satz.** *Seien V ein endlich-dimensionaler Vektorraum und $U \subset V$ ein Untervektorraum. Dann gibt es ein Komplement U' von U in V. Für jedes Komplement U' von U gilt*

$$\dim U + \dim U' = \dim V \,.$$

Beweis. Sei $m = \dim U \leq \dim V = n$. Wir wählen eine Basis $(b_1, b_2, \ldots, b_m)$ von U und ergänzen sie zu einer Basis $(b_1, \ldots, b_m, b_{m+1}, \ldots, b_n)$ von V. Dann ist $U' = \langle b_{m+1}, \ldots, b_n \rangle$ ein Komplement von U:

$$U + U' = \langle b_1, \ldots, b_m, b_{m+1}, \ldots, b_n \rangle = V \qquad \text{und} \qquad U \cap U' = \{\mathbf{0}\} \,,$$

weil $b_1, \ldots, b_n$ linear unabhängig sind. Die Dimensionsformel folgt aus Lemma 17.5. ❑

Derselbe Beweis zeigt, dass es auch in beliebigen Vektorräumen stets Komplemente gibt, wenn man den Basisergänzungssatz für Mengen verwendet, der mit Hilfe des Zornschen Lemmas bewiesen wurde. Vergleiche die Diskussion nach Satz 9.5.

Wir sehen hier insbesondere, dass alle Komplemente von U dieselbe Dimension $\dim V - \dim U$ haben. Das gilt ganz allgemein, auch wenn V unendliche Dimension hat.

LEMMA Komplemente sind isomorph

17.14. **Lemma.** *Seien V ein Vektorraum und $U \subset V$ ein Untervektorraum. Seien weiter U'_1 und U'_2 zwei Komplemente von U in V. Dann sind U'_1 und U'_2 isomorph; insbesondere gilt* $\dim U'_1 = \dim U'_2$.

Beweis. Wir betrachten die lineare Abbildung $\phi = \pi \circ \iota \colon U'_1 \to U'_2$; dabei sei $\iota \colon U'_1 \to V$ die Inklusionsabbildung und $\pi \colon V \to U'_2$ die Projektion bezüglich der Zerlegung $V = U \oplus U'_2$. Dann gilt

$$\ker(\phi) = \ker(\pi) \cap U'_1 = U \cap U'_1 = \{\mathbf{0}\} \,,$$

also ist ϕ injektiv. (Die erste Gleichheit folgt daraus, dass ι injektiv ist.) Wir müssen noch zeigen, dass ϕ auch surjektiv ist; dann ist ϕ ein Isomorphismus und U'_1 und U'_2 sind isomorph. Sei dazu $u_2 \in U'_2$. Dann gibt es eindeutig bestimmte $u \in U$ und $u_1 \in U'_1$ mit $u_2 = u + u_1$. Wir können das auch als $u_1 = (-u) + u_2$ lesen, woraus $\phi(u_1) = u_2$ folgt. ❑

Damit ist folgende Definition sinnvoll.

DEF Kodimension

∗ 17.15. **Definition.** Seien V ein Vektorraum und $U \subset V$ ein Untervektorraum, der ein Komplement U' in V hat. Dann heißt

$$\operatorname{codim}_V U = \dim U'$$

die *Kodimension* von U in V. ♢

Es gilt $\dim U + \operatorname{codim}_V U = \dim V$: Ist die Kodimension klein, dann ist U „groß“, also nicht weit davon entfernt, ganz V zu sein.

17.16. **Beispiel.** Die Kodimension kann auch für unendlich-dimensionale Untervektorräume endlich sein. Sei zum Beispiel P der reelle Vektorraum der Polynomfunktionen, sei $a \in \mathbb{R}$ und sei $U_a = \{p \in P \mid p(a) = 0\} = \ker \mathrm{ev}_a$. Dann ist der eindimensionale Untervektorraum $C = \langle x \mapsto 1 \rangle$ der konstanten Funktionen ein Komplement von U_a in P (für $p \in P$ gilt eindeutig $p = (p - p(a)) + p(a) \in U_a + C$), also ist $\operatorname{codim}_P U_a = 1$. Dieselbe Überlegung zeigt, dass der Untervektorraum der in einem Punkt a verschwindenden Funktionen auch in anderen Funktionenräumen (alle Funktionen, stetige Funktionen, n-mal stetig differenzierbare Funktionen usw.) Kodimension 1 hat.

BSP
Kodimension

Mit Polynomdivision (Satz 15.16) sieht man analog: Sind $a_1, a_2, \ldots, a_n \in \mathbb{R}$ paarweise verschieden und ist $U_{a_1,\ldots,a_n} = U_{a_1} \cap U_{a_2} \cap \ldots \cap U_{a_n}$ der Untervektorraum der Polynomfunktionen, die in $a_1, \ldots, a_n$ verschwinden, dann ist der Untervektorraum $P_{<n}$ der Polynomfunktionen vom Grad $< n$ ein Komplement von $U_{a_1,\ldots,a_n}$ in P, also gilt $\operatorname{codim}_P U_{a_1,\ldots,a_n} = n$. Denn jedes Polynom p kann eindeutig geschrieben werden als

$$p(x) = q(x)(x - a_1)(x - a_2) \cdots (x - a_n) + r(x)$$

mit $q \in P$ und $r \in P_{<n}$, und die Polynomfunktionen, die in $a_1, \ldots, a_n$ verschwinden, sind von der Form $x \mapsto q(x)(x - a_1)(x - a_2) \cdots (x - a_n)$. ♣

Der Begriff der Kodimension erlaubt eine etwas genauere Formulierung des „Rangsatzes“ 10.17. Zur Erinnerung: Der Satz besagt, dass für eine lineare Abbildung $\phi\colon V \to W$ gilt

$$\dim \ker(\phi) + \operatorname{rk}(\phi) = \dim \ker(\phi) + \dim \operatorname{im}(\phi) = \dim V\,.$$

Wenn V unendlich-dimensional ist, dann ist das eine relativ schwache Aussage. Die folgende Version gibt zusätzliche Information, wenn der Rang von ϕ endlich ist:

SATZ
Rangsatz mit Kodimension

17.17. **Satz.** *Sei $\phi\colon V \to W$ eine lineare Abbildung mit $\operatorname{rk}(\phi) < \infty$. Dann gilt*

$$\operatorname{codim}_V \ker(\phi) = \dim \operatorname{im}(\phi) = \operatorname{rk}(\phi)\,.$$

Beweis. Wir wählen eine Basis $(b_1, b_2, \ldots, b_m)$ von $\operatorname{im}(\phi)$ (mit $m = \dim \operatorname{im}(\phi)$). Seien weiter $v_1, v_2, \ldots, v_m \in V$ Urbilder von $b_1, b_2, \ldots, b_m$ unter ϕ. Dann sind $v_1, v_2, \ldots, v_m$ linear unabhängig, denn aus

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_m v_m = \mathbf{0}$$

folgt

$$\mathbf{0} = \phi(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_m v_m) = \lambda_1 b_1 + \lambda_2 b_2 + \ldots + \lambda_m b_m$$

und damit $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$, weil $b_1, b_2, \ldots, b_m$ linear unabhängig sind. Wir setzen

$$U = \langle v_1, v_2, \ldots, v_m \rangle \subset V\,;$$

dann ist U ein Komplement von $\ker(\phi)$ in V und $\dim U = m$, woraus die Behauptung folgt.

- $U + \ker(\phi) = V$: Sei $v \in V$, dann gibt es Skalare $\lambda_1, \ldots, \lambda_m$ mit

  $$\phi(v) = \lambda_1 b_1 + \ldots + \lambda_m b_m\,.$$

  Sei $v' = v - (\lambda_1 v_1 + \ldots + \lambda_m v_m)$, dann ist

  $$\phi(v') = \phi(v) - (\lambda_1 b_1 + \ldots + \lambda_m b_m) = \mathbf{0}\,,$$

  also $v' \in \ker(\phi)$ und $v = v' + (\lambda_1 v_1 + \ldots + \lambda_m v_m) \in \ker(\phi) + U$.

- $U \cap \ker(\phi) = \{\mathbf{0}\}$: Sei $v \in U \cap \ker(\phi)$, dann gibt es Skalare $\lambda_1, \ldots, \lambda_m$ mit $v = \lambda_1 v_1 + \ldots + \lambda_m v_m$. Außerdem gilt
$$\mathbf{0} = \phi(v) = \lambda_1 b_1 + \ldots + \lambda_m b_m ,$$
woraus $\lambda_1 = \ldots = \lambda_m = 0$ und damit $v = \mathbf{0}$ folgt. ❑

Am Ende dieses Abschnitts wollen wir noch untersuchen, wann eine Zerlegung eines Vektorraums V als direkte Summe $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$ einer analogen Zerlegung eines Endomorphismus f von V entspricht. Dazu müssen wir erst einmal sagen, was Letzteres bedeutet.

17.18. **Lemma.** *Sei $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$ eine Zerlegung eines Vektorraums V als direkte Summe von Untervektorräumen. Seien weiter*
$$f_1 \colon U_1 \to U_1, \quad f_2 \colon U_2 \to U_2, \quad \ldots, \quad f_n \colon U_n \to U_n$$
lineare Abbildungen. Dann gibt es einen eindeutig bestimmten Endomorphismus f von V mit $f(v) = f_i(v)$ für alle $i \in \{1, 2, \ldots, n\}$ und alle $v \in U_i$.

LEMMA direkte Summe von Endomorphismen

Beweis. Jedes $v \in V$ lässt sich eindeutig schreiben als $v = u_1 + u_2 + \ldots + u_n$ mit $u_1 \in U_1$, $u_2 \in U_2$, …, $u_n \in U_n$. Wenn $f(u_i) = f_i(u_i)$ gelten soll, dann müssen wir f definieren durch
$$f(v) = f_1(u_1) + f_2(u_2) + \ldots + f_n(u_n) .$$
Dann gilt auch $f(v) = f_i(v)$ für $v \in U_i$; es bleibt nur noch zu zeigen, dass f linear ist. Das folgt aus
$$f = \iota_1 \circ f_1 \circ \pi_1 + \iota_2 \circ f_2 \circ \pi_2 + \ldots + \iota_n \circ f_n \circ \pi_n ,$$
wobei $\pi_i \colon V \to U_i$ die Projektion auf U_i und $\iota_i \colon U_i \to V$ die Inklusionsabbildung ist. (Alternativ kann man das auch wie vor Definition 17.10 nachrechnen.) ❑

17.19. **Definition.** Die Abbildung f in Lemma 17.18 heißt die *direkte Summe* von $f_1, f_2, \ldots, f_n$; wir schreiben
$$f = f_1 \oplus f_2 \oplus \ldots \oplus f_n . \qquad \diamondsuit$$

DEF direkte Summe von Endomorphismen

Wenn ein Endomorphismus f als direkte Summe von Endomorphismen f_i der U_i geschrieben werden kann, dann muss offenbar $f(U_i) \subset U_i$ gelten. Wir geben dieser Eigenschaft einen Namen:

17.20. **Definition.** Seien V ein Vektorraum, $U \subset V$ ein Untervektorraum und f ein Endomorphismus von V. Dann heißt U *f-invariant* oder *invariant unter f*, wenn $f(U) \subset U$ gilt. ♢

DEF invarianter Untervektorraum

17.21. **Satz.** *Sei $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$ eine Zerlegung eines Vektorraums V als direkte Summe von Untervektorräumen. Sei weiter $f \in End(V)$. f lässt sich genau dann als direkte Summe von Endomorphismen $f_i \in \operatorname{End}(U_i)$ schreiben, wenn alle Untervektorräume U_i invariant unter f sind.*

SATZ Zerlegung von Endomorphismen

Beweis. Wir hatten schon gesehen, dass die Bedingung notwendig ist. Wir müssen noch zeigen, dass sie auch hinreichend ist. Es gelte also $f(U_i) \subset U_i$ für alle $i \in \{1, 2, \ldots, n\}$. Dann können wir $f_i \in \operatorname{End}(U_i)$ definieren durch $f_i(u) = f(u)$ für $u \in U_i$; damit gilt $f = f_1 \oplus f_2 \oplus \ldots \oplus f_n$. ❑

BSP
Zerlegung bei Diagonalisierbarkeit

17.22. **Beispiel.** Ist f ein Endomorphismus von V und λ ein Skalar, dann ist der Eigenraum $E_\lambda(f)$ unter f invariant (denn $f(v) = \lambda v$ für $v \in E_\lambda(f)$).

Ist V endlich-dimensional und $f \in \mathrm{End}(V)$ diagonalisierbar mit den paarweise verschiedenen Eigenwerten $\lambda_1, \lambda_2, \ldots, \lambda_m$, dann gilt

$$V = E_{\lambda_1}(f) \oplus E_{\lambda_2}(f) \oplus \ldots \oplus E_{\lambda_m}(f)$$

(denn V hat eine Basis aus Eigenvektoren von f) und

$$f = f_1 \oplus f_2 \oplus \ldots \oplus f_m$$

mit $f_i = \lambda_i \, \mathrm{id}_{E_{\lambda_i}(f)}$. In diesem Fall lässt sich f also in besonders einfach gebaute Abbildungen zerlegen. ♣

Wir überlegen uns jetzt noch, wie die Matrix von $f = f_1 \oplus f_2 \oplus \ldots \oplus f_n$ bezüglich einer an die direkte Summenzerlegung angepassten Basis aussieht.

LEMMA
Matrix einer direkten Summe

17.23. **Lemma.** *Sei V ein endlich-dimensionaler Vektorraum mit einer Zerlegung $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$, seien $f_i \in \mathrm{End}(U_i)$ für $i \in \{1, 2, \ldots, n\}$ und sei $f = f_1 \oplus f_2 \oplus \ldots \oplus f_n$. Seien B_i Basen von U_i und B die durch Aneinanderhängen von $B_1, B_2, \ldots, B_n$ gegebene Basis von V. Dann ist*

$$\mathrm{Mat}_{B,B}(f) = \left(\begin{array}{c|c|c|c} M_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & M_2 & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \cdots & M_n \end{array}\right)$$

eine Block-Diagonalmatrix, wobei $M_i = \mathrm{Mat}_{B_i,B_i}(f_i)$ die Matrizen von f_i bezüglich der Basen B_i sind.

Beweis. Das folgt aus $f(b) = f_i(b) \in U_i$ für Elemente $b \in B_i$: in den den Elementen von B_i entsprechenden Spalten von $\mathrm{Mat}_{B,B}(f)$ können nur die den Elementen von B_i entsprechenden Zeilen von null verschiedene Einträge enthalten, und die sich daraus ergebende Untermatrix ist die Matrix von f_i bezüglich B_i. ❑

18. Äquivalenzrelationen, Quotientenräume und affine Unterräume

Wir erinnern uns daran, dass der Kern jeder linearen Abbildung $f\colon V \to V'$ ein Untervektorraum von V ist. In diesem Abschnitt wollen wir der umgekehrten Frage nachgehen: Ist jeder Untervektorraum der Kern einer linearen Abbildung?

Im Fall, dass V endlich-dimensional ist, können wir mit unseren Kenntnissen über direkte Summen und Komplemente recht leicht zeigen, dass die Antwort „Ja" lautet: Sei U ein Untervektorraum des endlich-dimensionalen Vektorraums V, dann hat U ein Komplement U' in V (Satz 17.13), es ist also $V = U \oplus U'$. Die zu dieser Zerlegung gehörende Projektion $\pi\colon V \to U'$ hat dann U als Kern.

Dieses Argument ist aus zwei Gründen etwas unbefriedigend. Zum Einen verwendet es die Existenz von Basen (genauer: den Basisergänzungssatz), die wir nur für endlich-dimensionale Vektorräume gezeigt haben. Zum Anderen ist das Komplement U' im Normalfall weit davon entfernt, eindeutig bestimmt zu sein; wir müssen bei der Konstruktion der linearen Abbildung also eine Wahl treffen.

In diesem Abschnitt werden wir eine Konstruktion kennen lernen, die diese Nachteile vermeidet: Sie funktioniert für jeden Untervektorraum jedes Vektorraums und erfordert keine Auswahlen. Die Art dieser Konstruktion des „Quotientenraums" und des zugehörigen „kanonischen Epimorphismus" ist recht typisch für die Methoden der Algebra und wird in sehr ähnlicher Form im Rahmen der Vorlesungen „Einführung in die Zahlentheorie und algebraische Strukturen" und „Einführung in die Algebra" wieder auftauchen, dann für andere algebraische Strukturen wie zum Beispiel Ringe und Gruppen.

Sei also V ein (beliebiger) K-Vektorraum und $U \subset V$ ein Untervektorraum. Wenn es eine lineare Abbildung $f\colon V \to V'$ gibt mit $\ker(f) = U$, dann gibt es auch eine surjektive solche Abbildung, denn wir können einfach die im Wertebereich eingeschränkte Abbildung $f\colon V \to \operatorname{im}(f)$ betrachten. Wir nehmen jetzt also an, dass wir so eine surjektive lineare Abbildung $f\colon V \to V'$ mit Kern U haben. Wie können wir dann den Vektorraum V' beschreiben?

- Die **Elemente** von V' können wir durch Elemente von V repräsentieren; dabei wird $v' \in V'$ durch jedes $v \in V$ mit $f(v) = v'$ repräsentiert (das ist möglich, weil f surjektiv ist). Zwei Elemente v_1 und v_2 von V stellen genau dann dasselbe Element von V' dar, wenn $f(v_1) = f(v_2)$ ist. Das ist äquivalent zu $f(v_1 - v_2) = \mathbf{0}$, also zu $v_1 - v_2 \in \ker(f) = U$.
- Die **Addition** und **Skalarmultiplikation** auf V' kann unter Zuhilfenahme der Linearität von f ebenfalls über die entsprechenden Operationen von V erfolgen: Sind v_1 und v_2 Repräsentanten von $v'_1 = f(v_1)$ und $v'_2 = f(v_2)$, dann ist $v_1 + v_2$ ein Repräsentant von $v'_1 + v'_2$, denn $f(v_1 + v_2) = f(v_1) + f(v_2)$. Ebenso ist für $\lambda \in K$ auch λv_1 ein Repräsentant von $\lambda v'_1$.

Wir schreiben $[v]$ (statt $f(v)$) für das von $v \in V$ repräsentierte Element von V'. Dann können wir unsere Überlegungen wie folgt zusammenfassen: Falls V' existiert, dann

(1) besteht V' aus allen $[v]$ mit $v \in V$;

(2) es gilt $[v_1] = [v_2] \iff v_1 - v_2 \in U$

(3) und $[v_1] + [v_2] = [v_1 + v_2]$, $\lambda[v] = [\lambda v]$.

Es liegt also nahe, V' auf diese Weise zu *definieren*; dann wäre $f: V \to V'$, $v \mapsto [v]$, die passende surjektive lineare Abbildung mit Kern U (denn $[v] = \mathbf{0}$ genau dann, wenn $v \in U$). Dafür müssen wir nachweisen, dass diese Vorgehensweise zu keinen Widersprüchen führt.

Der erste Punkt dabei ist, sich zu überlegen, dass die Gleichheit der Elemente von V' sinnvoll definiert ist. Eine sinnvolle Definition von Gleichheit muss sicher die folgenden Eigenschaften haben:

(1) Jedes Element ist gleich zu sich selbst.

(2) Wenn a und b gleich sind, dann sind auch b und a gleich.

(3) Wenn sowohl a und b als auch b und c gleich sind, dann sind auch a und c gleich.

Wir gießen das in eine formale Definition. Dafür brauchen wir den Begriff der *Relation*.

DEF
Relation

18.1. **Definition.** Seien X und Y beliebige Mengen. Eine *Relation zwischen X und Y* ist eine Teilmenge $R \subset X \times Y$. Man sagt, $x \in X$ steht in der Relation R zu $y \in Y$ (oder x und y stehen in der Relation R), wenn $(x, y) \in R$ gilt. Manchmal schreibt man dafür abkürzend $x \, R \, y$.

Im Fall $X = Y$ spricht man von einer *Relation auf X*. ♢

BSP
Relationen

18.2. **Beispiele.**

(1) Auf jeder Menge X gibt es die *Gleichheitsrelation* $\{(x, x) \mid x \in X\}$ und die *Allrelation* $X \times X$.

(2) Auf $\mathbb{R}$ gibt es die Vergleichsrelation $\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \leq y\}$; analog für $<$, $\geq$, $>$.

(3) Auf $\mathbb{Z}$ gibt es die *Teilbarkeitsrelation* $\{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid \exists c \in \mathbb{Z}: ac = b\}$, deren Bestehen als $a \mid b$ notiert wird.

(4) Zwischen einer Menge X und ihrer Potenzmenge $\mathcal{P}(X)$ gibt es die Element-Relation $\{(x, T) \in X \times \mathcal{P}(X) \mid x \in T\}$. ♣

DEF ∗
Äquivalenz-
relation

18.3. **Definition.** Seien X eine Menge und R eine Relation auf X. Dann heißt R eine *Äquivalenzrelation*, wenn folgende Bedingungen erfüllt sind:

(1) $\forall x \in X: x \, R \, x$ (Reflexivität).

(2) $\forall x, y \in X: (x \, R \, y \implies y \, R \, x)$ (Symmetrie).

(3) $\forall x, y, z \in X: (x \, R \, y \land y \, R \, z \implies x \, R \, z)$ (Transitivität). ♢

Die Gleichheitsrelation und die Allrelation sind Äquivalenzrelationen auf jeder Menge X. Dagegen sind die Vergleichsrelationen $\leq$, $<$, $\geq$, $>$ auf $\mathbb{R}$ und die Teilbarkeitsrelation auf $\mathbb{Z}$ keine Äquivalenzrelationen, denn (z.B.) aus $a \leq b$ folgt nicht unbedingt $b \leq a$ und aus $a \mid b$ folgt nicht unbedingt $b \mid a$.

BSP Äquivalenzrelationen

18.4. **Beispiele.** Beispiele für Äquivalenzrelationen, die wir schon kennen gelernt haben, sind die „Äquivalenz“ und die „Ähnlichkeit“ von Matrizen.

Zur Erinnerung: Zwei Matrizen $A, B \in \mathrm{Mat}(m \times n, K)$ sind *äquivalent*, wenn sie dieselbe lineare Abbildung $K^n \to K^m$ bezüglich i.A. verschiedener Basen (von K^n und K^m) darstellen. Das ist gleichbedeutend damit, dass es Matrizen $P \in \mathrm{GL}(m, K)$ und $Q \in \mathrm{GL}(n, K)$ gibt mit $B = PAQ$. Wir hatten in Satz 13.10 gezeigt, dass das genau dann der Fall ist, wenn $\mathrm{rk}(A) = \mathrm{rk}(B)$ gilt.

Zwei Matrizen $A, B \in \mathrm{Mat}(n, K)$ sind *ähnlich*, wenn sie denselben Endomorphismus von K^n bezüglich i.A. verschiedener Basen von K^n darstellen. Das ist gleichbedeutend damit, dass es $P \in \mathrm{GL}(n, K)$ gibt mit $B = P^{-1}AP$. Über einem algebraisch abgeschlossenen Körper wie $\mathbb{C}$ wird die Klassifikation von Matrizen bis auf Ähnlichkeit durch die *Jordan-Normalform* geleistet, die wir später in dieser Vorlesung studieren werden. ♣

Man kann eine Äquivalenzrelation als eine „vergröberte“ Version von Gleichheit verstehen: Man betrachtet Elemente als gleich, obwohl sie nicht unbedingt identisch sind, aber so, dass die wesentlichen Eigenschaften der Gleichheit erfüllt sind. Das führt zu einer Einteilung von X in Klassen als untereinander gleich betrachteter Elemente:

SATZ Äquivalenzklassen

18.5. **Satz.** *Sei X eine Menge und sei $\sim$ eine Äquivalenzrelation auf X. Ist $x \in X$, dann schreiben wir $[x]$ für die Menge $\{y \in X \mid y \sim x\}$ und nennen $[x]$ die Äquivalenzklasse von x (bezüglich $\sim$). Für $x, y \in X$ sind dann die folgenden Aussagen äquivalent:*

(1) $y \sim x$.

(2) $y \in [x]$.

(3) $[x] = [y]$.

(4) $[x] \cap [y] \neq \emptyset$.

Insbesondere sind je zwei Äquivalenzklassen entweder gleich oder disjunkt.

Beweis.
„(1) ⇔ (2)“: Das ist die Definition von $[x]$.

„(1) ⇒ (3)“: Für $z \in X$ gilt (unter der Voraussetzung $y \sim x$, also auch $x \sim y$):

$$z \in [x] \iff z \sim x \iff z \sim y \iff z \in [y]\,,$$

also sind $[x]$ und $[y]$ gleich. Die mittlere Äquivalenz benutzt die Transitivität von $\sim$.

„(3) ⇒ (4)“: Es ist $x \in [x] = [y]$.

„(4) ⇒ (1)“: Sei $z \in [x] \cap [y]$, dann gilt $z \sim x$ und $z \sim y$. Die Symmetrie von $\sim$ impliziert $y \sim z$, die Transitivität dann $y \sim x$. ❑

DEF Quotientenmenge

kanonische Surjektion

18.6. **Definition.** In der Situation von Satz 18.5 schreiben wir

$$X/{\sim} = \{[x] \mid x \in X\} \subset \mathcal{P}(X)$$

für die Menge der Äquivalenzklassen und nennen $X/{\sim}$ („X modulo $\sim$“) die *Quotientenmenge* von X bezüglich $\sim$. Die Abbildung $X \to X/{\sim}$, $x \mapsto [x]$, ist surjektiv; sie heißt die *kanonische Surjektion*. ♢

18.7. **Beispiel.** Brüche $\frac{a}{b} \in \mathbb{Q}$ sind Äquivalenzklassen bezüglich der Äquivalenzrelation auf $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ (oder auf $\mathbb{Z} \times \mathbb{Z}_{>0}$), die durch

BSP Rationale Zahlen

$$(a, b) \sim (c, d) \iff ad = bc$$

definiert ist. In diesem Sinn ist $\mathbb{Q} = \big(\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})\big)/\sim$ und $\frac{a}{b} = [(a, b)]$. ♣

Wir können diese Überlegungen auf unser Problem anwenden.

LEMMA ∗ Kongruenz modulo U ist Äquivalenzrelation

18.8. **Lemma.** *Seien V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Die wie folgt definierte Relation $\equiv_U$ auf V ist eine Äquivalenzrelation. Statt $v \equiv_U v'$ schreiben wir $v \equiv v' \bmod U$ (gesprochen „v ist kongruent zu v' modulo U“).*

$$v \equiv v' \bmod U \iff v - v' \in U\,.$$

Statt $V/{\equiv_U}$ schreiben wir V/U („V modulo U“) für die Quotientenmenge. Für die Äquivalenzklassen gilt

$$[v] = \{v' \in V \mid v' - v \in U\} = \{v + u \mid u \in U\} = v + U\,.$$

Beweis. Wir müssen zeigen, dass die so definierte Relation reflexiv, symmetrisch und transitiv ist:

- Für $v \in V$ gilt $v - v = \mathbf{0} \in U$, also $v \equiv v \bmod U$.
- Für $v, v' \in V$ gelte $v \equiv v' \bmod U$, das bedeutet $v - v' \in U$. Dann ist auch $v' - v = -(v - v') \in U$ und damit $v' \equiv v \bmod U$.
- Für $v, v', v'' \in V$ gelte $v \equiv v' \bmod U$ und $v' \equiv v'' \bmod U$, also $v - v', v' - v'' \in U$. Dann ist auch $v - v'' = (v - v') + (v' - v'') \in U$, also gilt $v \equiv v'' \bmod U$.

Die Aussage über die Gestalt der Äquivalenzklassen ist klar (mit $u = v' - v$). ❑

Wir wollen jetzt gerne $V' = V/U$ setzen, mit der kanonischen Surjektion als linearer Abbildung $V \to V'$. Dazu müssen wir auf V/U eine K-Vektorraum-Struktur definieren. Wenn die kanonische Surjektion linear sein soll, dann müssen $[v] + [v'] = [v + v']$ und $\lambda[v] = [\lambda v]$ gelten (für $v, v' \in V$ und $\lambda \in K$). Wir müssen also nachweisen, dass diese Definitionen der Addition und der Skalarmultiplikation sinnvoll sind. Das wird durch zusätzliche Eigenschaften von $\equiv_U$ sichergestellt.

LEMMA Kongruenz modulo U ist Kongruenzrelation

18.9. **Lemma.** *Die Relation $\equiv_U$ aus Lemma 18.8 hat folgende Eigenschaften:*

(1) *Für $v_1, v_2, v_1', v_2' \in V$ gilt:*
Aus $v_1 \equiv v_1' \bmod U$ und $v_2 \equiv v_2' \bmod U$ folgt $v_1 + v_2 \equiv v_1' + v_2' \bmod U$.

(2) *Für $v, v' \in V$ und $\lambda \in K$ gilt: Aus $v \equiv v' \bmod U$ folgt $\lambda v \equiv \lambda v' \bmod U$.*

DEF Kongruenzrelation

Eine Äquivalenzrelation auf einem Vektorraum mit diesen zusätzlichen Eigenschaften (also Verträglichkeit mit der Vektorraum-Struktur) wird auch als *Kongruenzrelation* bezeichnet.

Beweis. (1) Wir haben $v_1 - v_1', v_2 - v_2' \in U$, also auch
$(v_1 + v_2) - (v_1' + v_2') = (v_1 - v_1') + (v_2 - v_2') \in U$.

(2) Aus $v - v' \in U$ folgt $\lambda v - \lambda v' = \lambda(v - v') \in U$. ❑

SATZ ∗
Quotientenraum

18.10. **Satz.** *Seien V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Durch die Festlegungen*

$$[v] + [v'] = [v + v'] \qquad \textit{und} \qquad \lambda \cdot [v] = [\lambda v]$$

wird die Menge V/U zu einem K-Vektorraum.

Die kanonische Surjektion $\pi\colon V \to V/U$, $v \mapsto [v]$, ist dann eine lineare Abbildung mit $\ker(\pi) = U$.

Beweis. Zuerst ist zu zeigen, dass die Definitionen der Addition und Skalarmultiplikation sinnvoll („wohldefiniert“) sind: Da es im Allgemeinen viele Möglichkeiten gibt, ein Element von V/U in der Form $[v]$ zu schreiben, müssen wir nachprüfen, dass die Definitionen nicht von der Auswahl der Repräsentanten abhängen. Es gelte also $[v_1] = [v_1']$ und $[v_2] = [v_2']$, d.h. $v_1 \equiv v_1' \bmod U$ und $v_2 \equiv v_2' \bmod U$. Nach Lemma 18.9 folgt dann $v_1 + v_2 \equiv v_1' + v_2' \bmod U$, also $[v_1 + v_2] = [v_1' + v_2']$. Das zeigt, dass die Summe von $[v_1]$ und $[v_2]$ nicht von der Wahl der Repräsentanten abhängt. Auf analoge Weise zeigt man, dass auch die Definition der Skalarmultiplikation sinnvoll ist.

Als Nächstes müssen wir die Axiome für einen Vektorraum nachprüfen. Sobald klar ist, dass Addition und Skalarmultiplikation wohldefiniert sind, folgen diese aber direkt aus ihrer Gültigkeit für V, wobei man natürlich $\mathbf{0} = [\mathbf{0}]$ und $-[v] = [-v]$ setzt. Wir zeigen das am Beispiel eines der Distributivgesetze: Seien $v, v' \in V$ und $\lambda \in K$. Dann gilt

$$\lambda([v] + [v']) = \lambda[v + v'] = [\lambda(v + v')] = [\lambda v + \lambda v'] = [\lambda v] + [\lambda v'] = \lambda[v] + \lambda[v'] .$$

Die anderen Axiome zeigt man nach demselben Schema: Linke Seite als Restklasse eines Elements von V schreiben, dann das Axiom in V anwenden, dann in die rechte Seite umformen.

Dass die kanonische Surjektion π linear ist, folgt schließlich direkt aus der Definition von Addition und Skalarmultiplikation in V/U. Tatsächlich ist die Definition gerade so gemacht, *damit* π linear wird! Es gilt dann (man beachte $\mathbf{0}_{V/U} = [\mathbf{0}_V]$, also $[v] = \mathbf{0}_{V/U} \iff v \in U$)

$$\ker(\pi) = \{v \in V \mid [v] = \mathbf{0}\} = \{v \in V \mid v \in U\} = U . \qquad ❑$$

18.11. **Definition.** Der Vektorraum V/U („V modulo U“) heißt der *Quotientenraum* von V modulo U; die lineare Abbildung $\pi\colon V \to V/U$ heißt der zugehörige *kanonische Epimorphismus.* ♢

DEF
Quotientenraum

kanonischer Epimorphismus

Damit ist die eingangs gestellte Frage positiv beantwortet.

Hat U ein Komplement U' in V, dann ist die Einschränkung des kanonischen Epimorphismus $\pi\colon V \to V/U$ auf U' ein Isomorphismus $U' \to V/U$ (Übung). Es folgt $\operatorname{codim}_V U = \dim U' = \dim V/U$. Wir können also die Kodimension für beliebige Untervektorräume als $\operatorname{codim}_V U = \dim V/U$ definieren, ohne auf die Existenz eines Komplements angewiesen zu sein. Die Formel

DEF
Kodimension allgemein

$$\dim U + \operatorname{codim}_V U = \dim V$$

ist dann nichts anderes als der „Rangsatz“ $\dim V = \dim \ker(\pi) + \dim \operatorname{im}(\pi)$ für den kanonischen Epimorphismus π.

Wir beweisen jetzt noch einige Eigenschaften von Quotientenraum und kanonischem Epimorphismus.

Als Erstes beantworten wir die Frage, wann eine lineare Abbildung $f\colon V \to W$ eine lineare Abbildung $\phi\colon V/U \to W$ „induziert“, wann es also so ein ϕ gibt, sodass $\phi \circ \pi = f$ ist, wobei $\pi\colon V \to V/U$ der kanonische Epimorphismus ist. Anders gesagt: Gibt es ϕ, sodass das folgende Diagramm „kommutiert“?

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\pi"'] & W \\ V/U \arrow[ur, dashed, "\phi"'] & \end{tikzcd}$$

Da für jedes $u \in U$ gilt, dass $\pi(u) = \mathbf{0}$ ist, muss auch $f(u) = \phi\bigl(\pi(u)\bigr) = \mathbf{0}$ sein; das bedeutet $U \subset \ker(f)$. Wie der folgende Satz zeigt, ist diese Bedingung auch hinreichend.

SATZ lin. Abb. auf V/U

18.12. **Satz.** *Seien V ein Vektorraum mit Untervektorraum U und $\pi\colon V \to V/U$ der kanonische Epimorphismus. Sei weiter $f\colon V \to W$ linear. Es gibt genau dann eine lineare Abbildung $\phi\colon V/U \to W$ mit $f = \phi \circ \pi$, wenn $U \subset \ker(f)$ ist. ϕ ist eindeutig bestimmt und genau dann injektiv, wenn $U = \ker(f)$ ist.*

Beweis. Dass die Bedingung notwendig ist, hatten wir uns bereits überlegt. Für die Gegenrichtung nehmen wir $U \subset \ker(f)$ an. Wenn es ϕ gibt, dann muss gelten

$$\phi([v]) = \phi\bigl(\pi(v)\bigr) = f(v)\,;$$

das zeigt schon einmal die Eindeutigkeit; die Frage ist nur, ob wir ϕ tatsächlich so definieren können. Dazu müssen wir zeigen, dass $f(v)$ nicht vom Repräsentanten von $[v]$ abhängt. Es seien also $v, v' \in V$ mit $[v] = [v']$, also $v - v' \in U$. Dann ist

$$f(v) = f\bigl((v - v') + v'\bigr) = f(v - v') + f(v') = f(v')\,,$$

weil aus $v - v' \in U \subset \ker(f)$ folgt, dass $f(v - v') = \mathbf{0}$ ist. Damit ist ϕ durch $\phi([v]) = f(v)$ wohldefiniert, und es gilt jedenfalls $\phi \circ \pi = f$. Es bleibt zu zeigen, dass ϕ linear ist. Das folgt aber aus der Linearität von π und von f:

$$\phi([v] + [v']) = \phi([v + v']) = f(v + v') = f(v) + f(v') = \phi([v]) + \phi([v'])$$

und

$$\phi(\lambda[v]) = \phi([\lambda v]) = f(\lambda v) = \lambda f(v) = \lambda\phi([v])\,.$$

ϕ ist genau dann injektiv, wenn $\ker(\phi)$ trivial ist. Aus der Definition von ϕ folgt $\phi([v]) = \mathbf{0} \iff v \in \ker(f)$; außerdem gilt natürlich $[v] = \mathbf{0} \iff v \in U$. Beides zusammen zeigt $\ker(\phi) = \{\mathbf{0}\} \iff \ker(f) = U$. ❑

Satz 18.12 erlaubt uns, lineare Abbildungen mit Definitionsbereich V/U zu konstruieren.

Der folgende Satz ist wichtig. In den Algebra-Vorlesungen werden Sie ähnliche Resultate für Gruppen und Ringe (statt wie hier Vektorräume) kennen lernen.

SATZ * Homomorphiesatz für lineare Abb.

18.13. **Satz.** *Sei $f\colon V \to W$ eine lineare Abbildung. Seien $\pi\colon V \to V/\ker(f)$ der kanonische Epimorphismus und $\iota\colon \operatorname{im}(f) \to W$ die Inklusionsabbildung. Dann gibt es einen eindeutig bestimmten Isomorphismus $\phi\colon V/\ker(f) \to \operatorname{im}(f)$, sodass $f = \iota \circ \phi \circ \pi$ ist:*

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\pi"'] & W \\ V/\ker(f) \arrow[r, "\cong", "\phi"'] & \operatorname{im}(f) \arrow[u, "\iota"'] \end{tikzcd}$$

Insbesondere sind $V/\ker(f)$ *und* $\mathrm{im}(f)$ *isomorph.*

Eine wichtige Aussage des Homomorphiesatzes ist, dass Bilder von Homomorphismen mit Quelle V (bis auf Isomorphie) dasselbe sind wie Quotientenräume von V.

Beweis. Nach Satz 18.12 gibt es eine eindeutig bestimmte lineare Abbildung $\tilde{\phi}\colon V/\ker(f) \to W$ mit $f = \tilde{\phi} \circ \pi$. Es gilt $\mathrm{im}(\tilde{\phi}) = \mathrm{im}(f)$, also können wir $\tilde{\phi}$ im Wertebereich einschränken zu $\phi\colon V/\ker(f) \to \mathrm{im}(f)$; es folgt $f = \iota \circ \phi \circ \pi$. Es bleibt zu zeigen, dass ϕ ein Isomorphismus ist: ϕ ist injektiv nach Satz 18.12 und surjektiv wegen $\mathrm{im}(\phi) = \mathrm{im}(\tilde{\phi}) = \mathrm{im}(f)$. ❑

BSP
Konstruktion von $\mathbb{R}$ aus $\mathbb{Q}$

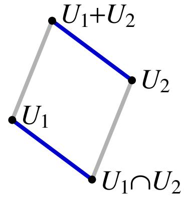
18.14. **Beispiel.** Die rationalen Cauchy-Folgen bilden einen Untervektorraum C des Vektorraums $\mathbb{Q}^{\mathbb{N}}$ der Folgen über $\mathbb{Q}$, denn Summen und skalare Vielfache von Cauchy-Folgen sind wieder Cauchy-Folgen. In C bilden die Nullfolgen einen Untervektorraum N. Jede Cauchy-Folge konvergiert in $\mathbb{R}$ und jede reelle Zahl ist Grenzwert einer rationalen Cauchy-Folge. Das liefert uns eine surjektive $\mathbb{Q}$-lineare Abbildung

$$\lim\colon C \longrightarrow \mathbb{R}, \quad (a_n)_{n\in\mathbb{N}} \longmapsto \lim_{n\to\infty} a_n$$

mit Kern N („Nullfolge" heißt ja gerade „Grenzwert null"). Aus dem Homomorphiesatz 18.13 folgt jetzt, dass C/N isomorph zu $\mathbb{R}$ ist (als $\mathbb{Q}$-Vektorraum). Dies ist eine der Möglichkeiten, wie man die reellen Zahlen aus den rationalen Zahlen konstruieren kann. In der „Einführung in die Zahlentheorie und algebraische Strukturen" werden wir lernen, dass die gleiche Konstruktion auch die Struktur von $\mathbb{R}$ als Körper mitliefert. ♣

Weitere Anwendungen des Homomorphiesatzes sind durch die folgenden „Isomorphiesätze" gegeben.

SATZ
Erster Isomorphiesatz



18.15. **Satz.** *Seien V ein Vektorraum und $U_1, U_2 \subset V$ zwei Untervektorräume. Dann ist die Abbildung*

$$\phi\colon U_1/(U_1 \cap U_2) \longrightarrow (U_1 + U_2)/U_2, \quad u + (U_1 \cap U_2) \longmapsto u + U_2$$

ein Isomorphismus.

(Wir verwenden hier die präzisere Schreibweise $v + U$ für die Äquivalenzklasse $[v]$, weil wir es mit zwei verschiedenen Quotientenräumen zu tun haben. In der Beschreibung von ϕ ist u ein Element von U_1.)

Beweis.

$$\begin{tikzcd} U_1 \arrow[rr] \arrow[d] \arrow[drr, "f"] & & U_1 + U_2 \arrow[d] \\ U_1/(U_1 \cap U_2) \arrow[rr, "\cong", "\phi"'] & & (U_1 + U_2)/U_2 \end{tikzcd}$$

Wir betrachten die Verknüpfung $f\colon U_1 \to (U_1 + U_2)/U_2$ der Inklusionsabbildung $U_1 \to U_1 + U_2$ mit dem kanonischen Epimorphismus $U_1 + U_2 \to (U_1 + U_2)/U_2$. Dann ist $\ker(f) = U_1 \cap U_2$. Außerdem ist f surjektiv: Sei $v + U_2 \in (U_1 + U_2)/U_2$ mit $v \in U_1 + U_2$, dann gibt es $u_1 \in U_1$ und $u_2 \in U_2$ mit $v = u_1 + u_2$. Es folgt $v + U_2 = u_1 + U_2 = f(u_1)$, da $v - u_1 = u_2 \in U_2$. Nach dem Homomorphiesatz 18.13 existiert der Isomorphismus ϕ wie angegeben. ❑

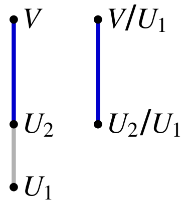
SATZ Zweiter Isomorphiesatz

18.16. **Satz.** *Seien V ein Vektorraum und $U_1 \subset U_2 \subset V$ Untervektorräume. Dann ist U_2/U_1 ein Untervektorraum von V/U_1 und die Abbildung*

$$\phi\colon V/U_2 \longrightarrow (V/U_1)/(U_2/U_1), \quad v + U_2 \longmapsto (v + U_1) + U_2/U_1$$

ist ein Isomorphismus.

Beweis.



$$\begin{tikzcd} V \arrow[r, "\pi"] \arrow[d] \arrow[dr, "f"] & V/U_1 \arrow[d] \\ V/U_2 \arrow[r, "\cong", "\phi"'] & (V/U_1)/(U_2/U_1) \end{tikzcd}$$

Sei $\pi\colon V \to V/U_1$ der kanonische Epimorphismus; dann ist $U_2/U_1 = \pi(U_2)$ ein Untervektorraum von V/U_1. Wir betrachten die Verknüpfung f von π mit dem kanonischen Epimorphismus $V/U_1 \to (V/U_1)/(U_2/U_1)$. Da beide Epimorphismen surjektiv sind, gilt das auch für f. Außerdem ist $\ker(f) = \pi^{-1}(U_2/U_1) = U_2$. Die Behauptung folgt dann wieder aus dem Homomorphiesatz 18.13. ❑

DEF Nebenklasse Restklasse

Die Äquivalenzklassen $[v] = v + U$, die in diesem Zusammenhang auch *Nebenklassen* (von U) oder *Restklassen* (modulo U) heißen, haben auch eine geometrische Interpretation als „verschobene Untervektorräume“ (man verschiebt nämlich U um den Vektor v). Dafür gibt es einen eigenen Namen.

DEF * Affiner Unterraum Gerade Ebene

18.17. **Definition.** Sei V ein Vektorraum. Ein *affiner Unterraum* von V ist entweder die leere Menge oder eine Menge der Form $v + U$ mit $v \in V$ und einem Untervektorraum U von V. Die *Dimension* von $v + U$ ist $\dim(v + U) = \dim U$; die Dimension des leeren affinen Unterraums wird als $-\infty$ definiert. Affine Unterräume der Dimension 1 heißen auch *(affine) Geraden*, affine Unterräume der Dimension 2 heißen auch *(affine) Ebenen*. ♢

Wir kennen affine Unterräume bereits als Lösungsmengen von linearen Gleichungen (siehe Satz 12.10): Die Lösungsmenge jeder linearen Gleichung $f(x) = b$ (wobei $f\colon V \to W$ eine lineare Abbildung ist) ist ein affiner Unterraum von V. Umgekehrt ist jeder affine Unterraum von V auch Lösungsmenge einer linearen Gleichung. Das ist klar für die leere Menge (wähle $f = \mathbf{0}\colon V \to K$ und $b = 1$); für $A = v + U$ ist $A = \pi^{-1}([v])$ für den kanonischen Epimorphismus $\pi\colon V \to V/U$.

Das liefert uns zwei verschiedene Möglichkeiten, einen nicht-leeren affinen Unterraum A des endlich-dimensionalen Standard-Vektorraums K^n zu beschreiben (wir betrachten die Elemente von K^n als Spaltenvektoren):

(1) Als verschobener Untervektorraum in der Form $A = \boldsymbol{x}_0 + \langle \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \rangle$, wobei $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ eine Basis des linearen Unterraums

$$A - A = \{v - v' \mid v, v' \in A\}$$

ist ($A - A$ ist U in Definition 18.17).

(2) Als Lösungsmenge eines Systems von linearen Gleichungen $M\boldsymbol{x} = \boldsymbol{b}$.

Von einer Beschreibung der Form (2) kommen wir zu einer Beschreibung der Form (1), indem wir den Lösungsalgorithmus für lineare Gleichungssysteme verwenden. Für die umgekehrte Richtung müssen wir $U = \langle \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \rangle$ als Kern einer Matrix M schreiben. Sei dafür M' die $(m \times n)$-Matrix, deren Zeilen durch $\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_m^\top$ gegeben sind. Dann folgt, dass der Zeilenraum der gesuchten Matrix M

genau der Kern von M' ist (Übung). Indem wir eine Basis des Kerns von M' als Zeilen in eine Matrix schreiben, bekommen wir also einen Kandidaten für M; die rechte Seite $\boldsymbol{b}$ ergibt sich dann als $\boldsymbol{b} = M\boldsymbol{x}_0$.

Welche Art der Beschreibung nützlicher ist, hängt davon ab, was wir mit dem affinen Unterraum anstellen wollen. Wenn wir zum Beispiel den Durchschnitt von zwei (oder mehr) affinen Unterräumen bestimmen wollen (der, wie wir bald sehen werden, wieder ein affiner Unterraum ist), dann ist Beschreibung (2) hilfreicher.

Der Untervektorraum U, der zu einem nicht-leeren affinen Unterraum A gehört, ist durch A eindeutig bestimmt, denn es ist $U = A - A = \{v - v' \mid v, v' \in A\}$. Dagegen ist der „Aufpunkt“ v nicht eindeutig bestimmt (außer im Fall $U = \{\mathbf{0}\}$), denn jedes $v \in A$ erfüllt $A = v + U$.

Wir können affine Unterräume durch eine Abgeschlossenheitseigenschaft charakterisieren.

∗ 18.18. **Satz.** *Seien V ein K-Vektorraum und $A \subset V$ eine Teilmenge. Dann sind äquivalent:*

SATZ Charakterisierung affiner Unterräume

(1) *A ist ein affiner Unterraum von V.*

(2) *A ist unter „affinen Linearkombinationen“ abgeschlossen: Für alle $a_1, a_2, \ldots, a_n \in A$ und $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$ mit $\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1$ gilt $\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n \in A$.*

Beweis. „(1)⇒(2)“: Wenn $A = \emptyset$ ist, ist nichts zu zeigen. Sei also $A = v + U$ mit $v \in V$ und einem Untervektorraum U. Dann ist $a_j = v + u_j$ mit $u_j \in U$ für alle $j \in \{1, 2, \ldots, n\}$, also erhalten wir

$$\begin{aligned}
\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n &= \lambda_1(v + u_1) + \lambda_2(v + u_2) + \ldots + \lambda_n(v + u_n) \\
&= (\lambda_1 + \lambda_2 + \ldots + \lambda_n)v + \lambda_1 u_1 + \lambda_2 u_2 + \ldots + \lambda_n u_n \\
&= v + (\lambda_1 u_1 + \lambda_2 u_2 + \ldots + \lambda_n u_n) \in v + U = A\,.
\end{aligned}$$

„(2)⇒(1)“: Wenn $A = \emptyset$ ist, dann ist A ein affiner Unterraum. Wir können also $A \neq \emptyset$ annehmen; sei $v \in A$ fest gewählt und $U = A - v = \{a - v \mid a \in A\} \subset V$. Wir zeigen, dass U ein Untervektorraum von V ist; dann folgt, dass $A = v + U$ ein affiner Unterraum ist.

- $\mathbf{0} \in U$, da $\mathbf{0} = v - v$ und $v \in A$ ist.
- U ist abgeschlossen unter der Addition: Seien $u = a - v$ und $u' = a' - v$ mit $a, a' \in A$. Nach (2) gilt $a + a' - v \in A$ (das ist eine affine Linearkombination), also ist $u + u' = (a + a' - v) - v \in U$.
- U ist abgeschlossen unter der Skalarmultiplikation: Seien $u = a - v$ mit $a \in A$ und $\lambda \in K$. Nach (2) gilt $\lambda a - \lambda v + v \in A$, also ist $\lambda u = (\lambda a - \lambda v + v) - v \in U$. ❑

Daraus folgt, dass Durchschnitte von affinen Unterräumen wieder affine Unterräume sind.

18.19. **Folgerung.** *Sei V ein Vektorraum und sei $(A_i)_{i \in I}$ eine Familie von affinen Unterräumen von V mit $I \neq \emptyset$. Dann ist $\bigcap_{i \in I} A_i$ ebenfalls ein affiner Unterraum von V.*

FOLG Durchschnitte von affinen Unterräumen

Beweis. Nach Satz 18.18 sind alle A_i abgeschlossen unter affinen Linearkombinationen. Seien jetzt $a_1, a_2, \ldots, a_n \in A = \bigcap_{i \in I} A_i$ und seien $\lambda_1, \lambda_2, \ldots, \lambda_n$ Skalare mit $\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1$. Dann ist $a = \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n \in A_i$ für alle $i \in I$, also ist $a \in A$. Damit ist A unter affinen Linearkombinationen abgeschlossen, also ist A wiederum nach Satz 18.18 ein affiner Unterraum von V. ❑

Ist $A_i = v_i + U_i$ und $\bigcap_{i \in I} A_i = v + U \neq \emptyset$, dann ist $U = \bigcap_{i \in I} U_i$ (Übung).

BSP
Affine Unterräume im $\mathbb{R}^3$

18.20. **Beispiel.** Welche affinen Unterräume gibt es im $\mathbb{R}^3$?

- Die leere Menge ist ein affiner Unterraum.
- Jede einelementige Menge $\{\boldsymbol{x}\}$ ist ein affiner Unterraum der Dimension 0.
- Jede Gerade (nicht unbedingt durch den Nullpunkt) ist ein affiner Unterraum der Dimension 1.
- Jede Ebene (nicht unbedingt durch den Nullpunkt) ist ein affiner Unterraum der Dimension 2.
- $\mathbb{R}^3$ selbst ist der einzige affine Unterraum der Dimension 3.

Zwei Geraden können zusammenfallen, sich in einem Punkt (affiner Unterraum der Dimension 0) schneiden oder disjunkt sein (dann sind sie parallel oder windschief). Für eine Gerade g und eine Ebene E gibt es die folgenden Möglichkeiten: $g \subset E$, $g \cap E = \{P\}$ oder $g \cap E = \emptyset$ (dann ist g parallel zu E). Zwei Ebenen können übereinstimmen, sich in einer Geraden schneiden oder disjunkt sein (dann sind sie parallel).

Man kann affine Unterräume wahlweise in der Form $A = v + U$ (wenn $A \neq \emptyset$) oder als Lösungsmenge eines linearen Gleichungssystems beschreiben. Eine (affine) Gerade im $\mathbb{R}^3$ kann also in der Form $g = \boldsymbol{x}_0 + \langle \boldsymbol{y} \rangle$ beschrieben werden (mit „Aufpunkt“ $\boldsymbol{x}_0$ und „Richtungsvektor“ $\boldsymbol{y} \neq \boldsymbol{0}$) oder in der Form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \end{aligned}$$

(mit linear unabhängigen Vektoren (a_{11}, a_{12}, a_{13}) und (a_{21}, a_{22}, a_{23})). Diese zweite Form kann man auch so interpretieren, dass man g als Schnitt zweier nicht paralleler Ebenen darstellt, denn jede der beiden Gleichungen beschreibt eine Ebene. ♣

Analog zur linearen Hülle kann man jetzt die *affine Hülle* einer Teilmenge $T \subset V$ definieren als den kleinsten affinen Unterraum, der T enthält (formal: als Durchschnitt aller affinen Unterräume, die T enthalten). Auf dieselbe Weise, wie wir gezeigt haben, dass die lineare Hülle von T genau aus allen Linearkombinationen von Elementen von T besteht, sieht man, dass die affine Hülle von T genau aus allen affinen Linearkombinationen von Elementen von T besteht. Es gilt dim(affine Hülle von T) $\leq \#T - 1$. Zum Beispiel ist die affine Hülle von drei verschiedenen Punkten im $\mathbb{R}^3$ entweder eine Gerade (wenn die drei Punkte auf einer Geraden liegen) oder eine Ebene, nämlich die durch die drei Punkte aufgespannte Ebene.

Ein anderes Beispiel ist die affine Hülle A der Vereinigung $g_1 \cup g_2$ zweier Geraden im $\mathbb{R}^3$. Im Fall $g_1 = g_2$ ist $A = g_1 = g_2$. Schneiden sich g_1 und g_2 in einem Punkt, dann spannen sie gemeinsam die Ebene A auf (die die Form $A = \boldsymbol{x}_0 + \langle \boldsymbol{y}_1, \boldsymbol{y}_2 \rangle$ hat, wobei $g_1 \cap g_2 = \{\boldsymbol{x}_0\}$ ist und $\boldsymbol{y}_1, \boldsymbol{y}_2$ Richtungsvektoren von g_1 und g_2 sind). Sind g_1 und g_2 parallel, dann spannen sie ebenfalls eine Ebene auf (finden Sie eine Beschreibung dieser Ebene!). Sind g_1 und g_2 schließlich windschief, dann ist $A = \mathbb{R}^3$.

Sind $A_1 = v_1 + U_1$ und $A_2 = v_2 + U_2$ endlich-dimensionale und nicht-leere affine Unterräume eines Vektorraums V und ist A die affine Hülle von $A_1 \cup A_2$, dann kann man folgende Dimensionsformel zeigen:

$$\dim A = \begin{cases} \dim A_1 + \dim A_2 - \dim(A_1 \cap A_2), & \text{falls } A_1 \cap A_2 \neq \emptyset; \\ \dim A_1 + \dim A_2 - \dim(U_1 \cap U_2) + 1, & \text{falls } A_1 \cap A_2 = \emptyset. \end{cases}$$

19. Der Dualraum

Wir hatten im ersten Semester schon gesehen, dass die Menge aller Homomorphismen $f\colon V \to W$ zwischen zwei K-Vektorräumen V und W selbst wieder die Struktur eines K-Vektorraums $\mathrm{Hom}(V, W)$ hat (Satz 10.20). Ein besonders wichtiger Spezialfall tritt auf, wenn $W = K$ ist.

DEF ∗
Linearform
Dualraum

19.1. **Definition.** Sei V ein K-Vektorraum. Eine lineare Abbildung $\phi\colon V \to K$ heißt auch eine *Linearform* auf V. Der Vektorraum $V^* = \mathrm{Hom}(V, K)$ heißt der *Dualraum* von V. ♢

Die Elemente von V^* sind also gerade die Linearformen auf V. Wir erinnern uns an die Definition der Vektorraumstruktur von V^*: Für Linearformen $\phi, \phi' \in V^*$ und $\lambda \in K$ ist $\phi + \phi'$ die Linearform $v \mapsto \phi(v) + \phi'(v)$ und $\lambda\phi$ ist die Linearform $v \mapsto \lambda\phi(v)$.

BSP
Linearformen

19.2. **Beispiele.**

(1) Auf dem Standardvektorraum K^n sind die *Koordinatenabbildungen* oder *Projektionen*

$$\mathrm{pr}_j\colon (x_1, x_2, \ldots, x_n) \longmapsto x_j, \qquad j \in \{1, 2, \ldots, n\},$$

Linearformen.

(2) Ist V ein Vektorraum von reellen Funktionen auf einer Menge X, dann ist für jedes $x \in X$ die *Auswertungsabbildung*

$$\mathrm{ev}_x\colon f \longmapsto f(x)$$

eine Linearform auf V.

(3) Ist V^* der Dualraum eines Vektorraums V, dann ist für jedes $v \in V$ die *Auswertungsabbildung*

$$\mathrm{ev}_v\colon \phi \longmapsto \phi(v)$$

DEF
Bidualraum

eine Linearform auf V^*, also ein Element des *Bidualraums* $V^{**} = (V^*)^*$. ♣

Wir erinnern uns daran, dass eine lineare Abbildung durch ihre Werte auf einer Basis eindeutig bestimmt ist und dass diese Werte beliebig vorgegeben werden können (Satz 10.11). Daraus ergibt sich der folgende wichtige Satz.

SATZ ∗
Existenz und Eindeutigkeit der dualen Basis

19.3. **Satz.** *Sei V ein endlich-dimensionaler Vektorraum mit Basis $(v_1, \ldots, v_n)$. Dann gibt es eine eindeutig bestimmte Basis $(v_1^*, \ldots, v_n^*)$ des Dualraums V^*, sodass für alle $i, j \in \{1, 2, \ldots, n\}$ gilt*

$$v_i^*(v_j) = \delta_{ij} = \begin{cases} 1, & \text{falls } i = j; \\ 0, & \text{falls } i \neq j. \end{cases}$$

DEF * duale Basis

19.4. Definition. In der Situation von Satz 19.3 heißt die Basis $(v_1^*, \dots, v_n^*)$ die *zur Basis* $(v_1, \dots, v_n)$ *duale Basis* von V^*. ♢

Man beachte, dass jedes Element v_i^* der dualen Basis von *allen* Elementen $v_1, \dots, v_n$ abhängt, nicht nur von v_i!

Beweis. Die Linearformen v_i^* sind durch die angegebene Bedingung eindeutig festgelegt, denn wir schreiben ihre Bilder auf einer Basis von V vor. Es bleibt zu zeigen, dass diese Elemente $v_1^*, v_2^*, \dots, v_n^* \in V^*$ eine Basis bilden.

Wir zeigen zuerst, dass sie linear unabhängig sind. Seien dazu $\lambda_1, \lambda_2, \dots, \lambda_n$ Skalare mit $\lambda_1 v_1^* + \lambda_2 v_2^* + \ldots + \lambda_n v_n^* = \mathbf{0}$. Wir werten die links stehende Linearform auf $v_1, v_2, \dots, v_n$ aus:

$$\begin{aligned} 0 = \mathbf{0}(v_j) &= (\lambda_1 v_1^* + \lambda_2 v_2^* + \ldots + \lambda_n v_n^*)(v_j) \\ &= \lambda_1 v_1^*(v_j) + \ldots + \lambda_{j-1} v_{j-1}^*(v_j) + \lambda_j v_j^*(v_j) + \lambda_{j+1} v_{j+1}^*(v_j) + \ldots + \lambda_n v_n^*(v_j) \\ &= \lambda_1 \cdot 0 + \ldots + \lambda_{j-1} \cdot 0 + \lambda_j \cdot 1 + \lambda_{j+1} \cdot 0 + \ldots + \lambda_n \cdot 0 \\ &= \lambda_j \,. \end{aligned}$$

Also sind alle $\lambda_j = 0$, und die lineare Unabhängigkeit ist bewiesen.

Wir müssen noch zeigen, dass die v_i^* ein Erzeugendensystem von V^* sind. Sei dazu $\phi \in V^*$ beliebig. Dann gilt

$$\phi = \phi(v_1) v_1^* + \phi(v_2) v_2^* + \ldots + \phi(v_n) v_n^* \,,$$

denn beide Seiten sind Linearformen, die auf der gegebenen Basis von V dieselben Werte annehmen: Wie eben gilt

$$\bigl(\phi(v_1) v_1^* + \phi(v_2) v_2^* + \ldots + \phi(v_n) v_n^*\bigr)(v_j) = \phi(v_j) \,.$$

Das zeigt, dass ϕ eine Linearkombination von $v_1^*, v_2^*, \dots, v_n^*$ ist.

(Alternativ könnten wir argumentieren, dass $\dim V^* = \dim \operatorname{Hom}(V, K) = \dim V = n$ ist, sodass die lineare Unabhängigkeit bereits ausreicht.) ❑

Wenn V nicht endlich-dimensional ist, dann kann man zu einer Basis $(b_i)_{i \in I}$ von V immer noch eine Familie $(b_i^*)_{i \in I}$ in V^* konstruieren, die $b_i^*(b_j) = \delta_{ij}$ erfüllt. Diese Familie ist linear unabhängig (mit demselben Beweis wie eben), aber kein Erzeugendensystem von V^*, denn jede Linearkombination (die ja immer nur endlich viele Vektoren involviert) der b_i^* nimmt nur auf endlich vielen Basiselementen b_j von null verschiedene Werte an. Es gibt aber zu *jeder* Wahl von Werten auf allen b_j eine zugehörige Linearform; zum Beispiel gibt es $\phi \in V^*$ mit $\phi(b_j) = 1$ für alle $j \in I$, aber $\phi \notin \langle \{b_i^* \mid i \in I\} \rangle$.

19.5. Beispiel. Die duale Basis zur Standardbasis $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ von K^n besteht gerade aus den Koordinatenabbildungen $(\mathrm{pr}_1, \mathrm{pr}_2, \dots, \mathrm{pr}_n)$. Wir bekommen einen Isomorphismus $f\colon K^n \to (K^n)^*$, indem wir $\mathbf{e}_j$ auf pr_j abbilden. Wenn wir die Elemente von K^n als Spaltenvektoren betrachten, dann können wir die Elemente des dualen K^n als Zeilenvektoren auffassen, denn sie sind lineare Abbildungen $K^n \to K$, die wir durch $(1 \times n)$-Matrizen darstellen können. Sind $\boldsymbol{x}, \boldsymbol{y} \in K^n$, dann ist die Auswertung von $\boldsymbol{x}^\top$ auf $\boldsymbol{y}$ gegeben durch die Multiplikation einer $(1 \times n)$-Matrix mit einer $(n \times 1)$-Matrix oder auch durch das Skalarprodukt: BSP duale Basis der Standardbasis

$$\boldsymbol{x}^\top(\boldsymbol{y}) = f(\boldsymbol{x})(\boldsymbol{y}) = \boldsymbol{x}^\top \boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + \ldots + x_n y_n \,.$$

Es ist nämlich

$$f(\boldsymbol{x})(\boldsymbol{y}) = (x_1 \,\mathrm{pr}_1 + \ldots + x_n \,\mathrm{pr}_n)(\boldsymbol{y}) = x_1 y_1 + \ldots + x_n y_n \,.$$ ♣

Aus dem Satz ergibt sich unmittelbar:

FOLG
$V \cong V^*$ für
V endl.-dim.

19.6. **Folgerung.** *Ist V ein endlich-dimensionaler Vektorraum, dann gilt*

$$\dim V = \dim V^* .$$

Insbesondere sind V und V^* isomorph.

Die Aussage von Folgerung 19.6 ist für unendlich-dimensionale Vektorräume *falsch.* Das liegt daran, dass die Dimension (als Mächtigkeit einer Basis definiert) des Dualraums V^* „unendlicher" ist als die Dimension von V selbst. Genauer bedeutet das: Es gibt zwar injektive, aber keine surjektiven Abbildungen von einer Basis von V in eine Basis von V^*. Diese Aussage ist verwandt mit dem Satz aus der Mengenlehre, dass die Potenzmenge $\mathcal{P}(X)$ einer Menge X stets echt mächtiger ist als X: Es gibt keine surjektive Abbildung $X \to \mathcal{P}(X)$. Zum Beweis sei $f\colon X \to \mathcal{P}(X)$ irgendeine Abbildung. Wir betrachten die Teilmenge

$$T = \{x \in X \mid x \notin f(x)\} \subset X .$$

(Als Element von $\mathcal{P}(X)$ ist $f(x)$ eine Teilmenge von X, also ist die Bedingung „$x \notin f(x)$" sinnvoll. Die Konstruktion ist ähnlich wie in der Russellschen Antinomie, die am Ende des Abschnitts über Mengenlehre in der Linearen Algebra I im Kleingedruckten erwähnt wird.) Dann ist $T \in \mathcal{P}(X)$ nicht im Bild von f. Denn wäre $T = f(x)$ für ein $x \in X$, dann erhielte man den Widerspruch

$$x \in T \iff x \notin f(x) \iff x \notin T$$

(die erste Äquivalenz ist die Definition von T, die zweite folgt aus $f(x) = T$). Der Zusammenhang ergibt sich so: Sei B eine Basis von V. Dann gibt es zu jeder Teilmenge T von B eine eindeutig bestimmte Linearform $\phi_T \in V^*$ mit $\phi_T(b) = 1$ für alle $b \in T$ und $\phi_T(b) = 0$ für alle $b \in B \setminus T$. Die Menge $\mathcal{T} = \{\phi_T \mid T \subset B\} \subset V^*$ hat die Mächtigkeit von $\mathcal{P}(B)$. Die ϕ_T sind zwar nicht linear unabhängig (zum Beispiel gilt $\phi_T + \phi_{T'} - \phi_{T\cup T'} - \phi_{T\cap T'} = \mathbf{0}$), aber man kann zeigen, dass $\mathcal{T}$ eine linear unabhängige Teilmenge gleicher Mächtigkeit enthält (das kommt daher, dass jede lineare Relation nur endlich viele ϕ_T enthält). Es folgt, dass jede Basis von V^* echt mächtiger sein muss als B.

Ein Isomorphismus $V \to V^*$ ist — nach Wahl einer Basis $(v_1, v_2, \dots, v_n)$ von V — dadurch gegeben, dass man v_i auf v_i^* abbildet. Der Isomorphismus hängt von der Wahl der Basis ab (man kann leicht Beispiele finden, die das belegen), er ist also nicht „natürlich" oder *kanonisch.* Im Unterschied dazu gibt es eine kanonische lineare Abbildung in den Bidualraum V^{**}.

Bevor wir das zeigen, formulieren wir noch eine Aussage, die wir später brauchen.

LEMMA
Fortsetzung
linearer
Abbildungen

19.7. **Lemma.**

(1) *Seien V und W Vektorräume und $U \subset V$ ein Untervektorraum. Ist außerdem $f\colon U \to W$ eine lineare Abbildung, dann kann man f zu einer linearen Abbildung $F\colon V \to W$ fortsetzen (es gilt also $F|_U = f$).*

(2) *Ist V ein Vektorraum und $\mathbf{0} \neq v \in V$, dann gibt es $\phi \in V^*$ mit $\phi(v) = 1$. Allgemeiner gilt: Ist $U \subset V$ ein Untervektorraum und $v \in V \setminus U$, dann gibt es $\phi \in V^*$ mit $\phi|_U = \mathbf{0}$ und $\phi(v) = 1$.*

Beweis.

(1) Wir verwenden, dass es ein Komplement U' von U in V gibt. Das haben wir nur für V endlich-dimensional bewiesen; es gilt jedoch auch allgemein. (Dafür braucht man den Basisergänzungssatz für unendliche Mengen und damit das Auswahlaxiom.) Jedes Element v von V lässt sich dann eindeutig schreiben als $v = u + u'$ mit $u \in U$ und $u' \in U'$; wir definieren F durch

$F(v) = f(u)$. F ist linear als Komposition der Projektion auf U (bezüglich der Zerlegung $V = U \oplus U'$) und der linearen Abbildung f; es ist klar, dass $F|_U = f$ gilt.

(2) Wir wenden Teil (1) an auf $U = \langle v \rangle$ und $f \colon U \to K$, $\lambda v \mapsto \lambda$. Für die allgemeinere Aussage betrachten wir $\mathbf{0} \neq [v] \in V/U$; dann gibt es eine Linearform $\bar{\phi} \colon V/U \to K$ mit $\bar{\phi}([v]) = 1$. Damit hat $\phi = \bar{\phi} \circ \pi$ die gewünschten Eigenschaften, wobei $\pi \colon V \to V/U$ der kanonische Epimorphismus ist. ❑

Die erste Aussage von Teil (2) in Lemma 19.7 kann man so interpretieren, dass man mithilfe von Linearformen auf V die Punkte (Elemente) von V „trennen" kann: Sind $v, v' \in V$ verschieden, dann ist $v - v' \neq \mathbf{0}$, also gibt es $\phi \in V^*$ mit $\phi(v - v') = 1$ und damit $\phi(v) \neq \phi(v')$. Daraus folgt, dass $v \in V$ durch die Abbildung $V^* \to K$, $\phi \mapsto \phi(v)$, eindeutig bestimmt ist. Das ist gleichbedeutend mit der Injektivität von α_V im folgenden Satz.

* 19.8. **Satz.** *Sei V ein K-Vektorraum. Dann ist die folgende Abbildung ein injektiver Homomorphismus:* **SATZ** kanon. Abb. in den Bidualraum

$$\alpha_V \colon V \longrightarrow V^{**}, \quad v \longmapsto \big(\mathrm{ev}_v \colon \phi \mapsto \phi(v)\big) .$$

Ist V endlich-dimensional, dann ist α_V ein Isomorphismus.

Beweis. Es sind verschiedene Aussagen zu zeigen.

- $\mathrm{ev}_v \in V^{**}$ (siehe Beispiel 19.2): ev_v ist eine Abbildung $V^* \to K$; wir müssen zeigen, dass ev_v linear ist:

$$\mathrm{ev}_v(\phi + \phi') = (\phi + \phi')(v) = \phi(v) + \phi'(v) = \mathrm{ev}_v(\phi) + \mathrm{ev}_v(\phi')$$

  und analog für die Skalarmultiplikation. (Hier benutzen wir die Definition der Vektorraumstruktur von V^*.)
- α_V ist linear: $\alpha_V(v + v') = \mathrm{ev}_{v+v'}$ bildet $\phi \in V^*$ auf $\phi(v + v') = \phi(v) + \phi(v') = \mathrm{ev}_v(\phi) + \mathrm{ev}_{v'}(\phi)$ ab, hat also denselben Effekt wie $\alpha_V(v) + \alpha_V(v')$. Analog sehen wir, dass $\alpha_V(\lambda v) = \mathrm{ev}_{\lambda v}$ die Abbildung $\phi \mapsto \phi(\lambda v) = \lambda\phi(v)$ ist und daher mit $\lambda\alpha_V(v)$ übereinstimmt. (Hier benutzen wir, dass die Elemente von V^* lineare Abbildungen sind, und die Definition der Vektorraumstruktur von V^{**}.)
- α_V ist injektiv: Wir zeigen $\ker(\alpha_V) = \{\mathbf{0}\}$. Sei $v \neq \mathbf{0}$. Nach Lemma 19.7 gibt es $\phi \in V^*$ mit $\big(\alpha_V(v)\big)(\phi) = \phi(v) = 1 \neq 0$, also ist $\alpha_V(v) \neq \mathbf{0}$ und damit $v \notin \ker(\alpha_V)$. Es bleibt also nur der Nullvektor als einzig mögliches Element von $\ker(\alpha_V)$.
- α_V ist Isomorphismus, falls $\dim V < \infty$: In diesem Fall gilt nach Folgerung 19.6 $\dim V = \dim V^* = \dim V^{**}$. Als injektive lineare Abbildung zwischen endlich-dimensionalen Vektorräumen derselben Dimension muss α_V dann ein Isomorphismus sein (Folgerung 10.14). ❑

Ist V endlich-dimensional, dann kann man also V und V^{**} durch den kanonischen Isomorphismus α_V miteinander identifizieren und damit V als den Dualraum von V^* betrachten. Das bedeutet, dass die Beziehung zwischen V und V^* symmetrisch ist. Das wird zum Beispiel durch die nächste Aussage illustriert.

FOLG Basis dual zu Basis von V^*

19.9. **Folgerung.** *Sei V ein endlich-dimensionaler Vektorraum; sei $(v_1^*, \ldots, v_n^*)$ eine Basis von V^*. Dann gibt es eine eindeutig bestimmte Basis $(v_1, v_2, \ldots, v_n)$ von V, sodass $(v_1^*, v_2^*, \ldots, v_n^*)$ die zu $(v_1, v_2, \ldots, v_n)$ duale Basis ist.*

Beweis. Sei $(v_1^{**}, v_2^{**}, \ldots, v_n^{**})$ die zu $(v_1^*, v_2^*, \ldots, v_n^*)$ duale Basis von V^{**}. Da α_V ein Isomorphismus ist, gibt es eindeutig bestimmte $v_1, v_2, \ldots, v_n \in V$ mit $\alpha_V(v_j) = v_j^{**}$ für alle $j \in \{1, 2, \ldots, n\}$; $(v_1, v_2, \ldots, v_n)$ ist eine Basis von V. Außerdem gilt für alle $i, j \in \{1, 2, \ldots, n\}$:

$$v_i^*(v_j) = \mathrm{ev}_{v_j}(v_i^*) = \bigl(\alpha_V(v_j)\bigr)(v_i^*) = v_j^{**}(v_i^*) = \delta_{ji} = \delta_{ij}\,,$$

also ist $(v_1^*, v_2^*, \ldots, v_n^*)$ die zu $(v_1, v_2, \ldots, v_n)$ duale Basis. Die Eindeutigkeit folgt aus der Eindeutigkeit der v_j^{**}. ❑

Sind $(v_1, v_2, \ldots, v_n)$ und $(v_1^*, v_2^*, \ldots, v_n^*)$ zueinander duale Basen von V und V^*, dann gilt:

$$\begin{aligned} \forall v^* \in V^*\colon &\quad v^* = v^*(v_1) \cdot v_1^* + v^*(v_2) \cdot v_2^* + \ldots + v^*(v_n) \cdot v_n^* \qquad \text{und} \\ \forall v \in V\colon &\quad v = v_1^*(v) \cdot v_1 + v_2^*(v) \cdot v_2 + \ldots + v_n^*(v) \cdot v_n\,. \end{aligned}$$

Die erste Aussage haben wir im Beweis von Satz 19.3 verwendet, die zweite folgt durch Vertauschen der Rollen von V und V^*. Diese zweite Relation zeigt, dass man die Elemente v_i^* der dualen Basis als Koordinatenabbildungen bezüglich der Basis $(v_1, \ldots, v_n)$ von V interpretieren kann.

BSP Interpolations-polynome als duale Basis

19.10. **Beispiel.** Sei $V = K[X]_{<n}$ der Vektorraum der Polynome über K vom Grad $< n$. Seien $a_1, a_2, \ldots, a_n \in K$ paarweise verschieden. Dann wissen wir, dass die Auswertungsabbildungen ev_{a_i} für $i \in \{1, 2, \ldots, n\}$ linear unabhängig sind; sie bilden also eine Basis von V^*. Welche Basis von V ist dazu dual? Wenn diese Basis $(p_1, p_2, \ldots, p_n)$ ist, dann muss gelten $p_i(a_j) = \delta_{ij}$, also ist

$$p_i = \prod_{j \neq i} \frac{X - a_j}{a_i - a_j}\,.$$

Die obige Relation liefert dann für $p \in V$ beliebig, dass

$$p = p(a_1) \cdot p_1 + p(a_2) \cdot p_2 + \ldots + p(a_n) \cdot p_n$$

ist — wir erhalten wieder die Lagrangesche Interpolationsformel, vergleiche Beispiel 10.15. ♣

Wir haben gesehen, wie man Vektorräume und Basen „dualisieren“ kann. Jetzt erweitern wir das auf lineare Abbildungen: Ist $f\colon V \to W$ linear und $\phi \in W^*$, dann ist $f^\top(\phi) = \phi \circ f\colon V \to K$ eine Linearform auf V:

$$\begin{tikzcd} V \arrow[d, "f"'] \arrow[dr, "f^\top(\phi)"] & \\ W \arrow[r, "\phi"] & K \end{tikzcd}$$

Wir erhalten eine Abbildung $f^\top\colon W^* \to V^*$.

DEF * duale lineare Abbildung

19.11. **Definition.** Ist $f\colon V \to W$ linear, dann heißt $f^\top\colon W^* \to V^*$, $\phi \mapsto \phi \circ f$, die zu f *duale* oder *transponierte* lineare Abbildung. ♢

Dass $f^\top$ tatsächlich linear ist, folgt aus der Definition der Vektorraumstruktur auf W^*:

$$f^\top(\phi + \phi') = (\phi + \phi') \circ f = \phi \circ f + \phi' \circ f = f^\top(\phi) + f^\top(\phi')$$

und

$$f^\top(\lambda\phi) = (\lambda\phi) \circ f = \lambda(\phi \circ f) = \lambda f^\top(\phi)\,.$$

Auch die Bezeichnung f^* ist gebräuchlich.

Die Notation $f^\top$ erklärt sich durch die folgende Aussage.

SATZ $f^\top$ und $A^\top$

19.12. **Satz.** *Seien V und W endlich-dimensionale K-Vektorräume, seien B und B' Basen von V und W und seien B^* und B'^* die dazu dualen Basen von V^* und W^*. Sei $f\colon V \to W$ linear und $A = \mathrm{Mat}_{B,B'}(f)$ die f bezüglich der Basen B und B' darstellende Matrix. Dann gilt*

$$\mathrm{Mat}_{B'^*,B^*}(f^\top) = A^\top\,.$$

Beweis. Seien $B = (b_1, b_2, \ldots, b_n)$, $B' = (b'_1, b'_2, \ldots, b'_m)$ und $A = (a_{ij})$. Wir schreiben $B^* = (b_1^*, b_2^*, \ldots, b_n^*)$ und $B'^* = (b_1'^*, b_2'^*, \ldots, b_m'^*)$. Dann ist

$$f(b_j) = a_{1j}b'_1 + a_{2j}b'_2 + \ldots + a_{mj}b'_m\,,$$

also ist

$$a_{ij} = b_i'^*\big(f(b_j)\big) = (b_i'^* \circ f)(b_j) = \big(f^\top(b_i'^*)\big)(b_j)\,.$$

Auf der anderen Seite gilt mit $\mathrm{Mat}_{B'^*,B^*}(f^\top) = (a'_{ij})$:

$$f^\top(b_i'^*) = a'_{1i}b_1^* + a'_{2i}b_2^* + \ldots + a'_{ni}b_n^*\,;$$

Anwenden auf b_j ergibt

$$a_{ij} = \big(f^\top(b_i'^*)\big)(b_j) = a'_{ji}\,,$$

also sind die beiden Matrizen zueinander transponiert. ❑

LEMMA $f \mapsto f^\top$

19.13. **Lemma.** *Sind V und W zwei K-Vektorräume, dann ist*

$$\Phi\colon\ \mathrm{Hom}(V, W) \longrightarrow \mathrm{Hom}(W^*, V^*), \quad f \longmapsto f^\top$$

eine injektive lineare Abbildung. Sind V und W beide endlich-dimensional, dann ist Φ ein Isomorphismus.

Beweis. Φ ist linear, denn für $\phi \in W^*$ und $f, g \in \mathrm{Hom}(V, W)$ und $\lambda \in K$ gilt

$$(f+g)^\top(\phi) = \phi \circ (f+g) = \phi \circ f + \phi \circ g = f^\top(\phi) + g^\top(\phi) = (f^\top + g^\top)(\phi)\,,$$

also ist $\Phi(f+g) = (f+g)^\top = f^\top + g^\top = \Phi(f) + \Phi(g)$, und

$$(\lambda f)^\top(\phi) = \phi \circ (\lambda f) = \lambda(\phi \circ f) = \lambda f^\top(\phi) = (\lambda f^\top)(\phi)\,,$$

also ist $\Phi(\lambda f) = (\lambda f)^\top = \lambda f^\top = \lambda\Phi(f)$.

Φ ist injektiv, denn für $f \in \mathrm{Hom}(V, W)$ mit $\Phi(f) = f^\top = \mathbf{0}$ gilt $\phi \circ f = \mathbf{0}$ für alle $\phi \in W^*$. Da es zu jedem $\mathbf{0} \neq w \in W$ ein $\phi \in W^*$ gibt mit $\phi(w) \neq 0$ (Lemma 19.7), folgt $f(v) = \mathbf{0}$ für alle $v \in V$, also $f = \mathbf{0}$.

Sind V und W beide endlich-dimensional, dann gilt (Satz 10.22)

$$\dim \mathrm{Hom}(W^*, V^*) = \dim W^* \cdot \dim V^* = \dim W \cdot \dim V = \dim \mathrm{Hom}(V, W)\,,$$

also ist Φ als injektive lineare Abbildung zwischen zwei endlich-dimensionalen Vektorräumen derselben Dimension ein Isomorphismus. ❑

Wir zeigen noch einige weitere einfache Eigenschaften der transponierten Abbildung.

LEMMA Eigenschaften von $f^\top$

19.14. **Lemma.**

(1) *Ist V ein Vektorraum, dann gilt* $\mathrm{id}_V^\top = \mathrm{id}_{V^*}$.

(2) *Sind V, V', V'' Vektorräume und $f\colon V \to V'$ und $g\colon V' \to V''$ lineare Abbildungen, dann gilt* $(g \circ f)^\top = f^\top \circ g^\top$.

(3) *Ist $f\colon V \to W$ ein Isomorphismus, dann ist auch $f^\top$ ein Isomorphismus und es gilt* $(f^\top)^{-1} = (f^{-1})^\top$.

Beweis.

(1) Für $\phi \in V^*$ ist $\mathrm{id}_V^\top(\phi) = \phi \circ \mathrm{id}_V = \phi$.

(2) Für $\phi \in (V'')^*$ ist

$$(g \circ f)^\top(\phi) = \phi \circ (g \circ f) = (\phi \circ g) \circ f = g^\top(\phi) \circ f = f^\top\big(g^\top(\phi)\big) = (f^\top \circ g^\top)(\phi)\,.$$

(3) Nach den beiden ersten Teilen gilt

$$f^\top \circ (f^{-1})^\top = (f^{-1} \circ f)^\top = \mathrm{id}_V^\top = \mathrm{id}_{V^*}$$

und

$$(f^{-1})^\top \circ f^\top = (f \circ f^{-1})^\top = \mathrm{id}_W^\top = \mathrm{id}_{W^*}\,,$$

woraus die Behauptungen folgen. ❑

Der Beweis der folgenden Aussagen, die Zusammenhänge zwischen $f^\top$ und der kanonischen Injektion α_V aufzeigen, ist eine Übungsaufgabe.

LEMMA $f^\top$ und α_V

19.15. **Lemma.**

(1) *Sei V ein Vektorraum. Dann gilt* $\alpha_V^\top \circ \alpha_{V^*} = \mathrm{id}_{V^*}$.

(2) *Sei $f\colon V \to W$ linear. Dann kommutiert das folgende Diagramm:*

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\alpha_V"'] & W \arrow[d, "\alpha_W"] \\ V^{**} \arrow[r, "f^{\top\top}"] & W^{**} \end{tikzcd}$$

es gilt also $f^{\top\top} \circ \alpha_V = \alpha_W \circ f$.

20. Bilinearformen

Nachdem wir im letzten Abschnitt Linearformen besprochen haben, kommen wir jetzt zu bilinearen Abbildungen und Bilinearformen.

∗ 20.1. **Definition.** Seien K ein Körper und V_1, V_2, W drei K-Vektorräume. Eine Abbildung $\beta\colon V_1 \times V_2 \to W$ heißt *(K-)bilinear*, wenn β in jedem der beiden Argumente K-linear ist, also wenn für alle $v_1, v_1' \in V_1$, $v_2, v_2' \in V_2$ und $\lambda \in K$ gilt

DEF
bilineare Abb.
Bilinearform

$$\begin{aligned} \beta(v_1 + v_1', v_2) &= \beta(v_1, v_2) + \beta(v_1', v_2)\,, & \beta(\lambda v_1, v_2) &= \lambda\beta(v_1, v_2) \\ \beta(v_1, v_2 + v_2') &= \beta(v_1, v_2) + \beta(v_1, v_2')\,, & \beta(v_1, \lambda v_2) &= \lambda\beta(v_1, v_2)\,. \end{aligned}$$

Ist $W = K$, dann heißt β eine *(K-)Bilinearform* oder auch *Paarung*. Gilt außerdem $V_1 = V_2 = V$, dann heißt β eine *(K-)Bilinearform auf V*. Wir bezeichnen den K-Vektorraum aller Bilinearformen $V_1 \times V_2 \to K$ mit $\operatorname{Bil}(V_1, V_2)$.

Ist $\beta\colon V \times V \to K$ eine Bilinearform auf V, dann heißt β *symmetrisch*, wenn für alle $v_1, v_2 \in V$ gilt, dass $\beta(v_2, v_1) = \beta(v_1, v_2)$ ist. β heißt *alternierend*, wenn für alle $v \in V$ gilt, dass $\beta(v, v) = 0$ ist. ♢

Ist $\beta\colon V \times V \to K$ eine alternierende Bilinearform, dann gilt $\beta(v_2, v_1) = -\beta(v_1, v_2)$. Das sieht man so:

$$\begin{aligned} 0 &= \beta(v_1 + v_2, v_1 + v_2) \\ &= \beta(v_1, v_1) + \beta(v_1, v_2) + \beta(v_2, v_1) + \beta(v_2, v_2) \\ &= \beta(v_1, v_2) + \beta(v_2, v_1)\,. \end{aligned}$$

Umgekehrt folgt aus $\beta(v_2, v_1) = -\beta(v_1, v_2)$ für alle $v_1, v_2 \in V$ die Gleichung $\beta(v, v) = -\beta(v, v)$, also $2\beta(v, v) = 0$ für alle $v \in V$. Kann man in K durch 2 teilen (im Körper $\mathbb{F}_2$ mit zwei Elementen ist $2 = 0$, dort geht das nicht, aber sonst praktisch immer), dann folgt, dass β alternierend ist. Im Normalfall sind alternierende Bilinearformen also dasselbe wie schief-symmetrische Bilinearformen.

Bilineare Abbildungen treten häufig in Gestalt einer Multiplikationsabbildung auf.

20.2. **Beispiele.**

BSP
bilineare Abb.

(1) Die Matrixmultiplikation

$$\operatorname{Mat}(l \times m, K) \times \operatorname{Mat}(m \times n, K) \longrightarrow \operatorname{Mat}(l \times n, K), \quad (A, B) \longmapsto AB$$

ist eine bilineare Abbildung (das folgt aus den Rechenregeln für Matrizen). Genauso ist die Multiplikation von Polynomen

$$K[X] \times K[X] \longrightarrow K[X], \quad (p, q) \longmapsto pq$$

eine bilineare Abbildung.

(2) Das *Standard-Skalarprodukt*

DEF
Standard-Skalarprodukt

$$K^n \times K^n \longrightarrow K, \quad \bigl((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)\bigr) \longmapsto x_1y_1 + x_2y_2 + \ldots + x_ny_n$$

ist eine symmetrische Bilinearform auf K^n.

(3) Die Abbildung

$$K^2 \times K^2 \longrightarrow K, \quad \bigl((x_1, x_2), (y_1, y_2)\bigr) \longmapsto x_1y_2 - x_2y_1$$

ist eine alternierende Bilinearform auf K^2.

(4) Die *Spurform*

$$\mathrm{Mat}(m \times n, K) \times \mathrm{Mat}(m \times n, K) \longrightarrow K, \quad (A, B) \longmapsto \mathrm{Tr}(A^\top B) = \mathrm{Tr}(AB^\top)$$

ist eine symmetrische Bilinearform auf $\mathrm{Mat}(m \times n, K)$. ♣

Allgemein gilt (leichte Übung): Ist $\beta\colon V_1 \times V_2 \to W$ bilinear und sind $f_1\colon V_1' \to V_1$, $f_2\colon V_2' \to V_2$ und $f\colon W \to W'$ linear, dann ist

$$f \circ \beta \circ (f_1, f_2)\colon V_1' \times V_2' \longrightarrow W', \qquad (v_1', v_2') \longmapsto f\bigl(\beta(f_1(v_1'), f_2(v_2'))\bigr)$$

ebenfalls bilinear. Ähnlich wie wir linearen Abbildungen zwischen endlich-dimensionalen Vektorräumen mit festgelegten Basen Matrizen zuordnen können, können wir auch Bilinearformen durch Matrizen beschreiben.

DEF Matrix einer Bilinearform

20.3. **Definition.** Seien V und W endlich-dimensionale K-Vektorräume. Wir setzen $\dim V = m$ und $\dim W = n$. Sei $\beta\colon V \times W \to K$ eine Bilinearform. Seien weiter $B = (b_1, b_2, \ldots, b_m)$ eine Basis von V und $B' = (b_1', b_2', \ldots, b_n')$ eine Basis von W. Dann heißt

$$\mathrm{Mat}_{B,B'}(\beta) = \bigl(\beta(b_i, b_j')\bigr)_{1 \le i \le m, 1 \le j \le n} = \begin{pmatrix} \beta(b_1, b_1') & \beta(b_1, b_2') & \cdots & \beta(b_1, b_n') \\ \beta(b_2, b_1') & \beta(b_2, b_2') & \cdots & \beta(b_2, b_n') \\ \vdots & \vdots & & \vdots \\ \beta(b_m, b_1') & \beta(b_m, b_2') & \cdots & \beta(b_m, b_n') \end{pmatrix}$$

die *Matrix von β* bezüglich B und B'. Im Fall $V = W$ und $B = B'$ schreiben wir auch $\mathrm{Mat}_B(\beta)$ statt $\mathrm{Mat}_{B,B}(\beta)$. ♢

Sind $v = x_1 b_1 + x_2 b_2 + \ldots + x_m b_m \in V$ und $v' = y_1 b_1' + y_2 b_2' + \ldots + y_n b_n' \in W$, dann ist $\beta(v, v') = \sum_{i=1}^m \sum_{j=1}^n x_i y_j \beta(b_i, b_j')$, was sich in folgende Matrixmultiplikation übersetzen lässt (rechts steht eine 1×1-Matrix, die wir mit ihrem einzigen Eintrag identifizieren):

$$\beta(v, v') = (x_1\, x_2\, \ldots\, x_m)\, \mathrm{Mat}_{B,B'}(\beta) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

Seien B_V, B_V' Basen von V und B_W, B_W' Basen von W, sowie $\beta\colon V \times W \to K$ eine Bilinearform. Mit $A = \mathrm{Mat}_{B_V,B_W}(\beta)$, $A' = \mathrm{Mat}_{B_V',B_W'}(\beta)$ und den Basiswechselmatrizen $P = \mathrm{Mat}_{B_V',B_V}(\mathrm{id}_V)$ und $Q = \mathrm{Mat}_{B_W',B_W}(\mathrm{id}_W)$ gilt dann (ähnlich wie für lineare Abbildungen)

$$A' = P^\top A Q\,.$$

Für Bilinearformen auf einem Vektorraum V, wo man nur eine Basis (von V) wählen kann, muss dabei $Q = P$ sein. In diesem Fall heißen zwei Matrizen A und A', die dieselbe Bilinearform (bezüglich zweier i.A. verschiedener Basen) beschreiben, auch *kongruent*. Das bedeutet also, dass es $P \in \mathrm{GL}(\dim V, K)$ gibt mit $A' = P^\top A P$; wir erhalten eine Äquivalenzrelation auf $\mathrm{Mat}(\dim V, K)$.

DEF kongruente Matrizen

Es gibt einen Zusammenhang zwischen Bilinearformen einerseits und Linearformen und Dualräumen andererseits.

LEMMA Linearformen aus einer Bilinearform

20.4. Lemma. *Seien V und W zwei K-Vektorräume.*

(1) *Sei $\beta\colon V \times W \to K$ eine Bilinearform. Dann ist für jedes $v \in V$ die Abbildung $W \to K$, $w \mapsto \beta(v,w)$, eine Linearform auf W, und für jedes $w \in W$ ist die Abbildung $V \to K$, $v \mapsto \beta(v,w)$, eine Linearform auf V.*

(2) *Die durch* (1) *gegebenen Abbildungen*

$$\beta_L\colon V \longrightarrow W^*, \quad v \longmapsto \bigl(w \mapsto \beta(v,w)\bigr)$$

und

$$\beta_R\colon W \longrightarrow V^*, \quad w \longmapsto \bigl(v \mapsto \beta(v,w)\bigr)$$

sind linear.

(3) *Die sich aus* (2) *ergebenden Abbildungen*

$$\mathrm{Bil}(V,W) \longrightarrow \mathrm{Hom}(V,W^*), \quad \beta \longmapsto \beta_L$$

und

$$\mathrm{Bil}(V,W) \longrightarrow \mathrm{Hom}(W,V^*), \quad \beta \longmapsto \beta_R$$

sind Isomorphismen. Insbesondere sind $\mathrm{Hom}(V,W^*)$ *und* $\mathrm{Hom}(W,V^*)$ *isomorph.*

Beweis.

(1) Das folgt unmittelbar aus der Definition von „Bilinearform“.

(2) Das folgt ebenfalls direkt aus der Definition von „Bilinearform“ und der Definition der Vektorraumstrukturen auf W^* und V^*.

(3) Die Linearität der Abbildungen ergibt sich aus der Definition der Vektorraumstrukturen der Abbildungsräume. Wir zeigen, dass die erste Abbildung bijektiv ist mit Inverser $f \mapsto \bigl((v,w) \mapsto (f(v))(w)\bigr)$: Einerseits wird $\beta \in \mathrm{Bil}(V,W)$ unter der Komposition der beiden Abbildungen wie folgt abgebildet:

$$\beta \longmapsto \beta_L \longmapsto \bigl((v,w) \mapsto \underbrace{(\beta_L(v))(w)}_{=\beta(v,w)}\bigr) = \beta\,;$$

andererseits haben wir für $f \in \mathrm{Hom}(V,W^*)$

$$f \longmapsto \bigl((v,w) \mapsto (f(v))(w)\bigr) \longmapsto \bigl(v \mapsto \underbrace{\bigl(w \mapsto (f(v))(w)\bigr)}_{=f(v)}\bigr) = f\,.$$

Die Bijektivität der zweiten Abbildung zeigt man analog. (Die Inverse ist $f \mapsto \bigl((v,w) \mapsto (f(w))(v)\bigr)$.) ❑

Ist B eine endliche Basis von V und B' eine endliche Basis von W, dann gilt

$$\mathrm{Mat}_{B,B'^*}(\beta_L) = \mathrm{Mat}_{B,B'}(\beta)^\top \qquad \text{und} \qquad \mathrm{Mat}_{B',B^*}(\beta_R) = \mathrm{Mat}_{B,B'}(\beta)\,,$$

wenn B^* und B'^* die zu B bzw. B' dualen Basen von V^* bzw. W^* sind.

DEF
Bilinearform nicht-ausgeartet

20.5. **Definition.** Eine Bilinearform $\beta\colon V \times W \to K$ heißt *nicht-ausgeartet*, wenn $\beta_L\colon V \to W^*$ und $\beta_R\colon W \to V^*$ Isomorphismen sind. Anderenfalls heißt β *ausgeartet*. ♢

Wenn man eine solche nicht-ausgeartete Bilinearform hat, dann kann man (via β_L und β_R) V als Dualraum von W und umgekehrt betrachten: Zu jeder Linearform ϕ auf V gibt es genau ein Element $w \in W$ mit $\phi = \beta_R(w)$ (also sodass $\phi(v) = \beta(v, w)$ ist für alle $v \in V$), und zu jeder Linearform ψ auf W gibt es genau ein Element $v \in V$ mit $\psi = \beta_L(v)$ (also sodass $\psi(w) = \beta(v, w)$ ist für alle $w \in W$).

Man kann zeigen, dass es eine nicht-ausgeartete Bilinearform auf $V \times W$ nur dann geben kann, wenn V und W endlich-dimensional sind.

Es folgt nämlich wegen $\beta_R^\top \circ \alpha_V = \beta_L$ (Übung), dass α_V ein Isomorphismus ist. Das ist aber nur für endlich-dimensionale Vektorräume V der Fall.

Dann müssen V und W dieselbe Dimension haben: $\dim W = \dim W^* = \dim V$.

BSP
nicht-ausg. Bilinearform

20.6. **Beispiel.** Sei V ein endlich-dimensionaler K-Vektorraum. Dann ist die *Auswertungspaarung*

$$\mathrm{ev}\colon V \times V^* \longrightarrow K, \quad (v, \phi) \longmapsto \phi(v)$$

nicht-ausgeartet. Für beliebiges V gilt nämlich

$$\mathrm{ev}_L = \alpha_V\colon V \longrightarrow V^{**} \qquad \text{und} \qquad \mathrm{ev}_R = \mathrm{id}_{V^*}\colon V^* \longrightarrow V^*,$$

und α_V ist ein Isomorphismus, wenn V endlich-dimensional ist. ♣

LEMMA
Kriterium für nicht-ausgeartet

20.7. **Lemma.** *Seien V und W zwei K-Vektorräume derselben endlichen Dimension n, seien B eine Basis von V, B' eine Basis von W und $\beta \in \mathrm{Bil}(V, W)$. Wir setzen $A = \mathrm{Mat}_{B,B'}(\beta)$. Dann sind folgende Aussagen äquivalent:*

(1) *β ist nicht-ausgeartet.*

(2) $\ker(\beta_L) = \{\mathbf{0}\}$.

(3) $\ker(\beta_R) = \{\mathbf{0}\}$.

(4) $\det(A) \neq 0$.

Beweis. Dass aus (1) die Aussagen (2) und (3) folgen, ist klar nach Definition 20.5. Umgekehrt folgt aus (2) zunächst, dass β_L ein Isomorphismus ist (denn $\dim V = n = \dim W^*$) und dann, dass $\beta_R = \beta_L^\top \circ \alpha_W$ ebenfalls ein Isomorphismus ist (denn $\beta_L^\top$ ist ein Isomorphismus nach Lemma 19.14). Genauso zeigt man „(3)⇒(1)“. Schließlich ist (3) äquivalent dazu, dass $\mathrm{Mat}_{B',B^*}(\beta_R)$ invertierbar ist. Diese Matrix ist aber genau A, und „A invertierbar“ ist äquivalent zu „$\det(A) \neq 0$“. ❑

Wir wollen jetzt symmetrische Bilinearformen auf einem Vektorraum V genauer betrachten.

LEMMA Matrix einer symm. Bilinearform

20.8. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$, V ein n-dimensionaler K-Vektorraum mit Basis B und $\beta \colon V \times V \to K$ eine Bilinearform auf V. Sei weiter $A = \mathrm{Mat}_B(\beta)$. Dann gilt:*

$$\beta \text{ ist symmetrisch} \quad \iff \quad A^\top = A$$

DEF symmetrische Matrix

Eine Matrix $A \in \mathrm{Mat}(n, K)$ heißt *symmetrisch*, wenn $A^\top = A$ ist. Damit gilt also, dass eine Bilinearform genau dann symmetrisch ist, wenn ihre Matrix (bezüglich einer beliebigen Basis) symmetrisch ist.

Beweis. Sei $B = (b_1, b_2, \ldots, b_n)$.

„$\Rightarrow$": Ist β symmetrisch, dann ist $\beta(b_i, b_j) = \beta(b_j, b_i)$; das bedeutet gerade $A^\top = A$.

„$\Leftarrow$": Sei $A^\top = A$. Dann gilt für Spaltenvektoren $\boldsymbol{x}, \boldsymbol{y} \in K^n$:

$$\boldsymbol{x}^\top A \boldsymbol{y} = (\boldsymbol{x}^\top A \boldsymbol{y})^\top = \boldsymbol{y}^\top A^\top \boldsymbol{x} = \boldsymbol{y}^\top A \boldsymbol{x}\,.$$

(Beachte: $\boldsymbol{x}^\top A \boldsymbol{y}$ ist eine 1×1-Matrix und damit gleich ihrer Transponierten.) Daraus folgt $\beta(v, v') = \beta(v', v)$ für alle $v, v' \in V$. ❑

Zum Beispiel sieht man so sehr leicht, dass das Standard-Skalarprodukt eine nichtausgeartete symmetrische Bilinearform auf K^n ist, denn die zugehörige Matrix bezüglich der Standardbasis ist die Einheitsmatrix I_n, die symmetrisch ist und $\det(I_n) = 1$ erfüllt.

Wir betrachten im Folgenden den Fall $K = \mathbb{R}$. Dann können wir zwischen positiven und negativen Elementen von $\mathbb{R}$ unterscheiden. Das führt zu folgender Definition.

* 20.9. **Definition.** Seien V ein $\mathbb{R}$-Vektorraum und $\beta \colon V \times V \to \mathbb{R}$ eine symmetrische Bilinearform auf V.

DEF positiv/negativ (semi)definit; indefinit

(1) β heißt *positiv semidefinit*, wenn $\beta(v, v) \geq 0$ ist für alle $v \in V$.
(2) β heißt *positiv definit*, wenn $\beta(v, v) > 0$ ist für alle $\mathbf{0} \neq v \in V$.
(3) β heißt *negativ semidefinit*, wenn $\beta(v, v) \leq 0$ ist für alle $v \in V$.
(4) β heißt *negativ definit*, wenn $\beta(v, v) < 0$ ist für alle $\mathbf{0} \neq v \in V$.
(5) β heißt *indefinit*, wenn es $v, v' \in V$ gibt mit $\beta(v, v) > 0$ und $\beta(v', v') < 0$.

Seien $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, \mathbb{R})$ symmetrisch, also $A^\top = A$. Im Folgenden betrachten wir $\boldsymbol{x} \in \mathbb{R}^n$ stets als Spaltenvektor.

(1) A heißt *positiv semidefinit*, wenn $\boldsymbol{x}^\top A \boldsymbol{x} \geq 0$ ist für alle $\boldsymbol{x} \in \mathbb{R}^n$.
(2) A heißt *positiv definit*, wenn $\boldsymbol{x}^\top A \boldsymbol{x} > 0$ ist für alle $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n$.
(3) A heißt *negativ semidefinit*, wenn $\boldsymbol{x}^\top A \boldsymbol{x} \leq 0$ ist für alle $\boldsymbol{x} \in \mathbb{R}^n$.
(4) A heißt *negativ definit*, wenn $\boldsymbol{x}^\top A \boldsymbol{x} < 0$ ist für alle $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n$.
(5) A heißt *indefinit*, wenn es $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ gibt mit $\boldsymbol{x}^\top A \boldsymbol{x} > 0$ und $\boldsymbol{y}^\top A \boldsymbol{y} < 0$. ◊

Daraus folgt im Fall $\dim V < \infty$, dass β genau dann positiv/negativ (semi-)definit bzw. indefinit ist, wenn das für $\mathrm{Mat}_B(\beta)$ mit irgendeiner Basis B von V gilt.

BSP positiv definite Bilinearformen

20.10. **Beispiele.**

(1) Das Standard-Skalarprodukt auf $\mathbb{R}^n$ ist positiv definit, denn es ist

$$x_1^2 + x_2^2 + \ldots + x_n^2 > 0\,,$$

wenn nicht alle x_j null sind.

(2) Die Spurform auf $\mathrm{Mat}(m \times n, \mathbb{R})$ ist ebenfalls positiv definit, denn für eine Matrix $A = (a_{ij}) \in \mathrm{Mat}(m \times n, \mathbb{R})$ ist

$$\mathrm{Tr}(A^\top A) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2\,.$$

(Wenn man $\mathrm{Mat}(m \times n, \mathbb{R})$ mit $\mathbb{R}^{mn}$ in der üblichen Weise identifiziert, dann ist die Spurform einfach das Standard-Skalarprodukt.) ♣

⚠ Die Matrix einer positiv definiten symmetrischen Bilinearform kann auch negative Einträge haben und eine symmetrische Matrix mit lauter positiven Einträgen braucht nicht positiv definit zu sein.

BSP

20.11. **Beispiele.** Seien

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Dann ist A positiv definit, denn

$$\begin{pmatrix} x & y \end{pmatrix} A \begin{pmatrix} x \\ y \end{pmatrix} = 2x^2 - 2xy + 2y^2 = 2(x - \tfrac{1}{2}y)^2 + \tfrac{3}{2}y^2\,,$$

und B ist nicht positiv semidefinit, denn

$$\begin{pmatrix} 1 & -1 \end{pmatrix} B \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1^2 + 4 \cdot 1 \cdot (-1) + (-1)^2 = -2\,.$$

Tatsächlich ist B indefinit, denn $\mathbf{e}_1^\top B \mathbf{e}_1 = 1$. ♣

LEMMA pos. def. ⇒ nicht-ausg.

20.12. **Lemma.** *Ist V ein endlich-dimensionaler reeller Vektorraum mit einer positiv definiten symmetrischen Bilinearform $\beta\colon V \times V \to \mathbb{R}$, dann ist β nicht-ausgeartet.*

Beweis. Nach Lemma 20.7 genügt es zu zeigen, dass $\ker(\beta_L) = \{\mathbf{0}\}$ ist. Sei also $v \in \ker(\beta_L)$. Dann ist

$$0 = \mathbf{0}(v) = \big(\beta_L(v)\big)(v) = \beta(v, v)\,.$$

Wäre $v \neq \mathbf{0}$, dann hätten wir $\beta(v, v) > 0$, also muss $v = \mathbf{0}$ sein. ❑

Unser nächstes Ziel wird es sein, ein relativ einfaches Kriterium herzuleiten, mit dem man entscheiden kann, ob eine symmetrische Matrix positiv (oder negativ) definit ist. Dies geschieht im Hinblick auf Anwendungen in der Analysis II (dort wird es um Kriterien gehen, wann eine Funktion in mehreren Variablen ein lokales Maximum oder Minimum hat).

Dafür werden wir folgende Aussage verwenden, die wir allerdings jetzt noch nicht beweisen können. Das werden wir bald nachholen. Zuerst noch eine Definition.

DEF orthogonale Matrix

20.13. **Definition.** Sei $n \in \mathbb{N}$. Eine Matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ heißt *orthogonal*, wenn $A^\top A = I_n$ ist. Wir schreiben $\mathrm{O}(n)$ für die Menge der orthogonalen Matrizen in $\mathrm{Mat}(n, \mathbb{R})$. ♢

Dann ist insbesondere A invertierbar (mit $A^{-1} = A^\top$). Man prüft ohne große Schwierigkeiten nach, dass $\mathrm{O}(n)$ eine Gruppe (mit der Matrixmultiplikation als Verknüpfung) ist.

SATZ Spektralsatz

20.14. **Satz.** *Ist $A \in \mathrm{Mat}(n, \mathbb{R})$ symmetrisch, dann ist A (über $\mathbb{R}$) orthogonal diagonalisierbar: Es gibt eine Matrix $P \in \mathrm{O}(n)$, sodass $P^\top AP = P^{-1}AP = D$ eine Diagonalmatrix ist.*

Daraus folgt leicht:

LEMMA Definitheit über Eigenwerte

20.15. **Lemma.** *Sei $A \in \mathrm{Mat}(n, \mathbb{R})$ symmetrisch. Dann gilt:*

(1) *A ist genau dann positiv semidefinit, wenn A keinen negativen Eigenwert hat.*

(2) *A ist genau dann positiv definit, wenn alle Eigenwerte von A positiv sind.*

(3) *A ist genau dann negativ semidefinit, wenn A keinen positiven Eigenwert hat.*

(4) *A ist genau dann negativ definit, wenn alle Eigenwerte von A negativ sind.*

(5) *A ist genau dann indefinit, wenn A positive und negative Eigenwerte hat.*

Beweis. Nach Satz 20.14 gibt es $P \in \mathrm{O}(n)$, sodass

$$P^\top AP = P^{-1}AP = D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

eine Diagonalmatrix ist; ihre Diagonaleinträge sind gerade die Eigenwerte von A. Nach Lemma 20.8 ist D die Matrix der A entsprechenden symmetrischen Bilinearform auf $\mathbb{R}^n$ bezüglich einer anderen Basis (gegeben durch die Spalten von P), also ist A genau dann positiv definit, wenn D positiv definit ist. Für einen Spaltenvektor $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$ gilt

$$\boldsymbol{x}^\top D\boldsymbol{x} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \ldots + \lambda_n x_n^2\,.$$

Sind alle $\lambda_j > 0$, dann ist das positiv für alle $\boldsymbol{x} \neq \mathbf{0}$, also ist D (und damit A) positiv definit. Ist hingegen $\lambda_j \leq 0$ für ein j, dann ist $\boldsymbol{x}^\top D\boldsymbol{x} \leq 0$ für $\boldsymbol{x} = \mathbf{e}_j$, und D (und damit A) ist nicht positiv definit. Die anderen Aussagen sieht man auf die gleiche Weise. ❑

Das Definitheitskriterium wird mit Hilfe von Determinanten geeigneter Untermatrizen formuliert, sogenannten Minoren.

DEF Untermatrix Minor Hauptminor

20.16. **Definition.** Seien K ein Körper, $m, n \in \mathbb{N}$, $A = (a_{ij}) \in \mathrm{Mat}(m \times n, K)$ und $0 \leq r \leq \min\{m, n\}$. Eine $r \times r$*-Untermatrix* von A ist eine Matrix der Form $(a_{i_k,j_l})_{1 \leq k,l \leq r}$, wobei $1 \leq i_1 < i_2 < \ldots < i_r \leq m$ und $1 < j_1 < j_2 < \ldots < j_r \leq n$ sind. Man wählt also r Zeilen und r Spalten von A aus und bildet die Matrix aus den Einträgen in diesen Zeilen und Spalten.

Ein r*-Minor* von A ist die Determinante einer $r \times r$-Untermatrix von A. Im Fall $m = n$ ist ein r*-Hauptminor* von A ein r-Minor von A, sodass in der obigen Notation $i_1 = j_1$, $i_2 = j_2$, …, $i_r = j_r$ gilt (man wählt also dieselben Zeilen- und Spaltenindizes aus). Der *führende r-Hauptminor* von A ist die Determinante der

Untermatrix $(a_{ij})_{1\le i,j\le r}$, die aus den ersten r Zeilen und Spalten von A gebildet wird. $\diamondsuit$

Minoren sind zum Beispiel nützlich, um den Rang einer Matrix zu beschreiben.

LEMMA
Rang über Minoren

20.17. **Lemma.** *Seien K ein Körper, $m, n \in \mathbb{N}$, $A \in \mathrm{Mat}(m \times n, K)$ und sei $1 \le r \le \min\{m, n\}$. Dann sind folgende Aussagen äquivalent:*

(1) $\mathrm{rk}(A) < r$.

(2) *Alle r-Minoren von A verschwinden.*

Beweis. „(1) $\Rightarrow$ (2)“: Sei A' eine $r \times r$-Untermatrix von A. Da je r Spalten von A linear abhängig sind, gilt das auch für die Spalten von A', also ist $\det A' = 0$.

„(2) $\Rightarrow$ (1)“: Wir nehmen an, dass $\mathrm{rk}(A) \ge r$ ist und zeigen, dass es einen nicht verschwindenden r-Minor gibt. Nach Voraussetzung gibt es r linear unabhängige Spalten in A; sei B die $m \times r$-Matrix, die aus diesen r Spalten besteht. Dann ist $\mathrm{rk}(B) = r$, also hat B auch r linear unabhängige Zeilen. Sei A' die Matrix, die aus diesen r Zeilen von B besteht; dann ist A' eine $r \times r$-Untermatrix von A. Außerdem ist $\mathrm{rk}(A') = r$, also ist der r-Minor $\det(A')$ von A nicht null. ❑

Mit Hilfe der Minoren lassen sich auch die weiteren Koeffizienten des charakteristischen Polynoms ausdrücken, denn es gilt für eine Matrix $A \in \mathrm{Mat}(n, K)$ mit charakteristischem Polynom $p \in K[X]$:

$$p = \sum_{k=0}^{n} (-1)^k s_k(A) X^{n-k},$$

wobei $s_k(A)$ die Summe der k-Hauptminoren von A ist. Für $k = 1$ ist das gerade die Spur von A, denn die 1-Hauptminoren sind genau die Einträge auf der Diagonalen; für $k = n$ ist das die Determinante von A (der einzige n-Hauptminor). Eine Möglichkeit das einzusehen besteht darin, die Multilinearität der Determinante als Funktion (z.B.) der Zeilen einer Matrix zu verwenden (vergleiche das Kleingedruckte auf Seite 106). Für eine Teilmenge T von $\{1, 2, \ldots, n\}$ sei A_T die $n \times n$-Matrix, deren j-te Zeile für $j \in T$ mit der j-ten Zeile von $-A$ und für $j \notin T$ mit der j-ten Zeile von I_n übereinstimmt. Dann ist $\det(A_T)$ ist gerade $(-1)^{\#T}$-mal der $\#T$-Minor von A, der zu den Zeilen und Spalten mit Nummern in T gehört (wie man durch Entwicklung nach den anderen Zeilen sieht). Es folgt

$$\sum_{T \subset \{1,2,\ldots,n\}, \#T = k} \det(A_T) = (-1)^k s_k(A)$$

und damit

$$p = \det(XI_n - A) = \sum_{T \subset \{1,2,\ldots,n\}} \det(A_T) X^{n-\#T} = \sum_{k=0}^{n} (-1)^k s_k(A) X^{n-k}.$$

Wir wollen die Minoren jetzt aber benutzen, um nachzuweisen, dass eine symmetrische Matrix positiv (oder negativ) definit ist. Dafür formulieren wir zunächst noch ein einfaches Lemma.

LEMMA Untermatrizen erben positive Definitheit

20.18. Lemma. *Sei $A = (a_{ij}) \in \mathrm{Mat}(n, \mathbb{R})$ eine positiv definite symmetrische Matrix. Dann ist für $r \in \{1, 2, \ldots, n\}$ die Untermatrix $A' = (a_{ij})_{1 \le i,j \le r}$ von A ebenfalls positiv definit.*

Beweis. Sei $\mathbf{0} \neq \boldsymbol{x}' = (x_1, \ldots, x_r)^\top \in \mathbb{R}^r$. Wir müssen zeigen, dass $(\boldsymbol{x}')^\top A' \boldsymbol{x}' > 0$ ist. Sei $\boldsymbol{x} = (x_1, x_2, \ldots, x_r, 0, \ldots, 0)^\top \in \mathbb{R}^n$ (wir fügen also $n - r$ Nullen an). Dann ist $\boldsymbol{x} \neq \mathbf{0}$, also nach Voraussetzung $\boldsymbol{x}^\top A \boldsymbol{x} > 0$. Es genügt also zu zeigen, dass $(\boldsymbol{x}')^\top A' \boldsymbol{x}' = \boldsymbol{x}^\top A \boldsymbol{x}$ ist. Mit $x_j = 0$ für $j \in \{r+1, r+2, \ldots, n\}$ gilt

$$\boldsymbol{x}^\top A \boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = \sum_{i=1}^{r} \sum_{j=1}^{r} a_{ij} x_i x_j = (\boldsymbol{x}')^\top A' \boldsymbol{x}' \,. \qquad \square$$

* **SATZ** Determinantenkriterium für positiv definit

20.19. Satz. *Seien $n \in \mathbb{N}$ und $A = (a_{ij}) \in \mathrm{Mat}(n, \mathbb{R})$ eine symmetrische Matrix. Für $r \in \{1, 2, \ldots, n\}$ sei $d_r(A) = \det(a_{ij})_{1 \le i,j \le r}$ der führende r-Hauptminor von A. Dann gilt:*

(1) *A ist positiv definit $\iff$ $d_r(A) > 0$ für alle $r \in \{1, 2, \ldots, n\}$.*

(2) *A ist negativ definit $\iff$ $(-1)^r d_r(A) > 0$ für alle $r \in \{1, 2, \ldots, n\}$.*

Die Bedingung für „negativ definit“ heißt also $d_1(A) < 0$, $d_2(A) > 0$, $d_3(A) < 0$ usw.: Die führenden Hauptminoren alternieren im Vorzeichen. Man merkt sich das am besten an den Vorzeichen der führenden Hauptminoren von $-I_n$.

Beweis. Wir beweisen zunächst Aussage (1). Die Richtung „$\Rightarrow$“ folgt aus Lemma 20.15, denn mit A sind nach Lemma 20.18 auch die Matrizen $A_r = (a_{ij})_{1 \le i,j \le r}$ positiv definit, und eine positiv definite Matrix hat positive Determinante (denn die ist das Produkt der (positiven) Eigenwerte). Die Richtung „$\Leftarrow$“ zeigen wir durch Induktion über n. Der Fall $n = 0$ (oder $n = 1$) ist klar. Für den Schritt von n auf $n + 1$ sei $A \in \mathrm{Mat}(n+1, \mathbb{R})$ symmetrisch mit positiven führenden Hauptminoren $d_r(A)$ für alle $r \in \{1, 2, \ldots, n+1\}$. Das gilt dann entsprechend auch für die Matrix $A_n \in \mathrm{Mat}(n, \mathbb{R})$ (denn $d_r(A_n) = d_r(A)$ für $r \le n$). Nach Induktionsvoraussetzung ist A_n positiv definit. Das heißt, dass für Spaltenvektoren $\mathbf{0} \neq \boldsymbol{x} \in \langle \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \rangle \subset \mathbb{R}^{n+1}$ stets $\boldsymbol{x}^\top A \boldsymbol{x} > 0$ ist. Wir zeigen jetzt, dass A höchstens einen (mit algebraischer Vielfachheit gerechnet) negativen Eigenwert haben kann: Nach Satz 20.14 gibt es $P \in \mathrm{O}(n+1)$ mit

$$P^\top A P = P^{-1} A P = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n+1})$$

diagonal. Wären wenigstens zwei Eigenwerte negativ, etwa λ_i und λ_j, mit zugehörigen Eigenvektoren $\boldsymbol{y}_i = P\mathbf{e}_i$ und $\boldsymbol{y}_j = P\mathbf{e}_j$ (als Spaltenvektoren), dann hätten wir für $(0,0) \neq (\alpha, \beta) \in \mathbb{R}^2$

$$(\alpha \boldsymbol{y}_i + \beta \boldsymbol{y}_j)^\top A (\alpha \boldsymbol{y}_i + \beta \boldsymbol{y}_j) = (\alpha \mathbf{e}_i + \beta \mathbf{e}_j)^\top P^\top A P (\alpha \mathbf{e}_i + \beta \mathbf{e}_j) = \lambda_i \alpha^2 + \lambda_j \beta^2 < 0 \,.$$

Da die $n + 2$ Vektoren $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n, \boldsymbol{y}_i, \boldsymbol{y}_j \in \mathbb{R}^{n+1}$ nicht linear unabhängig sein können, gibt es $(0,0) \neq (\alpha, \beta) \in \mathbb{R}^2$ mit $\mathbf{0} \neq \boldsymbol{x} = \alpha \boldsymbol{y}_i + \beta \boldsymbol{y}_j \in \langle \mathbf{e}_1, \ldots, \mathbf{e}_n \rangle$. Dann müsste aber sowohl $\boldsymbol{x}^\top A \boldsymbol{x} < 0$ als auch $\boldsymbol{x}^\top A \boldsymbol{x} > 0$ gelten, ein Widerspruch. Es kann also keine zwei negativen Eigenwerte geben. Da das Produkt aller Eigenwerte $d_{n+1}(A) = \det(A)$ positiv ist, kann es auch nicht genau einen negativen Eigenwert geben (und natürlich kann null kein Eigenwert sein), also sind alle Eigenwerte von A positiv; nach Lemma 20.15 ist A damit positiv definit.

Aussage (2) folgt aus Aussage (1): A ist genau dann negativ definit, wenn $-A$ positiv definit ist, und für die führenden Hauptminoren gilt $d_r(-A) = (-1)^r d_r(A)$. ❑

BSP

20.20. **Beispiele.** Wir betrachten wieder

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Die führenden Hauptminoren von A sind $d_1(A) = 2$, $d_2(A) = 2^2 - 1^2 = 3$, was bestätigt, dass A positiv definit ist. Hingegen sind die führenden Hauptminoren von B gegeben durch $d_1(B) = 1$ und $d_2(B) = 1^2 - 2^2 = -3$, was bestätigt, dass B nicht positiv definit ist (und auch nicht negativ definit, denn dafür haben beide Minoren das falsche Vorzeichen). ♣

Ist die symmetrische Matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ nur positiv semidefinit, dann folgt wie im Beweis von „$\Rightarrow$", dass die führenden Hauptminoren von A alle ≥ 0 sein müssen. Die Umkehrung gilt dann aber im Allgemeinen nicht.

BSP

20.21. **Beispiel.** Die Matrix $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ hat nichtnegative führende Hauptminoren (beide sind null), ist aber nicht positiv semidefinit (denn $\mathbf{e}_2^\top A \mathbf{e}_2 = -1$). ♣

Es gibt auch ein Determinanten-Kriterium für positive (oder negative) Semidefinitheit. Es lautet wie folgt.

SATZ
Determinantenkriterium für semidefinit

Satz. *Seien $n \in \mathbb{N}$ und $A = (a_{ij}) \in \mathrm{Mat}(n, \mathbb{R})$ eine symmetrische Matrix. Dann gilt:*

(1) *A ist positiv semidefinit $\iff$ $d \geq 0$ für alle Hauptminoren d von A.*

(2) *A ist negativ semidefinit $\iff$*
$\forall r \in \{1, 2, \ldots, n\}\colon\ (-1)^r d \geq 0$ für alle r-Hauptminoren d von A.

(3) *A ist indefinit $\iff$ es gibt einen $2r$-Hauptminor $d < 0$ von A, oder es gibt einen $(2r+1)$-Hauptminor $d > 0$ und einen $(2r'+1)$-Hauptminor $d' < 0$ von A.*

Beweis. Aussage (3) folgt formal-logisch aus (1) und (2) (A ist genau dann indefinit, wenn A weder positiv noch negativ semidefinit ist). Aussage (2) folgt aus (1) durch Anwendung von (1) auf $-A$. Es genügt also, die erste Aussage zu zeigen. Die Richtung „$\Rightarrow$" ist wieder klar: Jede Haupt-Untermatrix von A ist positiv semidefinit, hat also nichtnegative Eigenwerte und damit nichtnegative Determinante.

Zum Beweis von „$\Leftarrow$" nehmen wir an, dass alle Hauptminoren von A nichtnegativ sind. Wir bemerken zunächst, dass aus Lemma 20.15 folgt, dass eine symmetrische Matrix mit nicht verschwindender Determinante positiv oder negativ definit oder indefinit sein muss. Sei $K = \ker(A) \subset \mathbb{R}^n$ und $k = \dim K$. Wir wählen eine Basis $(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_k)$ von K. Wir können diese Basis durch Hinzunahme von $n-k$ Standard-Basisvektoren $\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \ldots, \mathbf{e}_{j_{n-k}}$ (mit $1 \leq j_1 < j_2 < \ldots < j_{n-k} \leq n$) zu einer Basis von $\mathbb{R}^n$ ergänzen (Basisergänzungssatz 9.5). Sei $V = \langle \mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \ldots, \mathbf{e}_{j_{n-k}} \rangle$; dann ist $V \cap K = \{\mathbf{0}\}$ und jeder Vektor $\boldsymbol{x} \in \mathbb{R}^n$ kann eindeutig geschrieben werden als $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{x}_1$ mit $\boldsymbol{x}_0 \in K$ und $\boldsymbol{x}_1 \in V$. Sei $\beta\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ die symmetrische Bilinearform, deren Matrix bezüglich der Standard-Basis A ist. Dann gilt $\beta(\boldsymbol{x}, \boldsymbol{x}_0) = \beta(\boldsymbol{x}_0, \boldsymbol{x}) = 0$ für alle $\boldsymbol{x} \in \mathbb{R}^n$ und $\boldsymbol{x}_0 \in K$. Für $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{x}_1$ wie oben gilt also $\beta(\boldsymbol{x}, \boldsymbol{x}) = \beta(\boldsymbol{x}_1, \boldsymbol{x}_1)$. Sei A' die $(n-k) \times (n-k)$-Untermatrix von A zu den Zeilen- und Spaltenindizes $j_1, j_2, \ldots, j_{n-k}$. Dann ist A' eine Matrix der Bilinearform $\beta' = \beta|_{V \times V}$, und nach der obigen Überlegung ist A genau dann positiv semidefinit, wenn das für A' gilt. Außerdem ist $\ker(A') = \{\mathbf{0}\}$ (wegen $K \cap V = \{\mathbf{0}\}$), also ist $\det(A') \neq 0$. Damit ist A' positiv oder negativ definit oder indefinit. Wie im Beweis von Satz 20.19 zeigt man induktiv, dass A' keine zwei negativen

Eigenwerte haben kann. Wegen $\det(A') > 0$ (hier verwenden wir die Voraussetzung) müssen alle Eigenwerte von A' positiv sein. Damit ist A' positiv definit, also ist A positiv semidefinit. ❑

Dieses Kriterium ist sehr viel weniger nützlich als Satz 20.19: Es gibt 2^n Hauptminoren, aber nur n führende Hauptminoren. Der Aufwand dafür, *alle* Hauptminoren zu testen, wird also schon für relativ kleine n zu groß, um praktikabel zu sein. Zum Glück gibt es bessere Möglichkeiten. Wir werden darauf noch genauer eingehen.

21. Euklidische und unitäre Vektorräume

Wir haben am Ende des letzten Abschnitts schon damit begonnen, von der „allgemeinen“ linearen Algebra über beliebigen Grundkörpern etwas wegzugehen und Resultate für den speziellen Körper $\mathbb{R}$ zu beweisen. Das setzen wir in diesem Abschnitt fort. Der Hintergrund dafür ist, dass wir *Geometrie* betreiben wollen: Wir wollen in der Lage sein, Abstände und Winkel zu messen. Dies wird in einem reellen Vektorraum durch eine positiv definite symmetrische Bilinearform ermöglicht.

DEF * euklidisches Skalarprod. euklidischer Vektorraum

21.1. **Definition.** Eine positiv definite symmetrische Bilinearform auf einem reellen Vektorraum heißt *euklidisches Skalarprodukt*. Ein reeller Vektorraum V zusammen mit einem euklidischen Skalarprodukt auf V ist ein *euklidischer Vektorraum*. Das Skalarprodukt in einem euklidischen Vektorraum wird häufig $(v, w) \mapsto \langle v, w \rangle$ (oder auch $v \cdot w$) geschrieben. $\diamondsuit$

„Euklidisch“ nach Euklid von Alexandria, da man in euklidischen Vektorräumen euklidische Geometrie betreiben kann.

Um Verwechslungen zu vermeiden, notieren wir den von einer Menge A erzeugten Untervektorraum als $\langle A \rangle_{\mathbb{R}}$.

BSP eukl. VR

21.2. **Beispiele.**

(1) Das *Standard-Skalarprodukt* $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y}$ auf $\mathbb{R}^n$ (mit Spaltenvektoren $\boldsymbol{x}$, $\boldsymbol{y}$) ist ein euklidisches Skalarprodukt. $\mathbb{R}^n$ mit diesem Skalarprodukt ist das Standardbeispiel für einen (endlich-dimensionalen) euklidischen Vektorraum.

(2) Seien $a, b \in \mathbb{R}$ mit $a < b$ und sei $V = \mathcal{C}([a, b])$ der Vektorraum der stetigen reellen Funktionen auf $[a, b]$. Dann definiert
$$\langle f, g \rangle = \int_a^b f(x)g(x)\,dx$$
ein euklidisches Skalarprodukt auf V. $\clubsuit$


C. Hermite 1822–1901

Wir wollen jetzt eine analoge Definition für komplexe (statt reelle) Vektorräume formulieren. Eine symmetrische Bilinearform (wie im reellen Fall) können wir nicht verwenden, denn eine symmetrische Bilinearform β auf einem komplexen Vektorraum kann nicht positiv definit sein (man kann nicht einmal erreichen, dass $\beta(v, v)$ stets reell ist): Es ist
$$\beta(\boldsymbol{i}v, \boldsymbol{i}v) = \boldsymbol{i}^2 \beta(v, v) = -\beta(v, v)\,.$$
Um das zu verhindern, modifizieren wir die Eigenschaften, die wir fordern.

DEF * Sesquilinearform hermitesch

21.3. **Definition.** Sei V ein $\mathbb{C}$-Vektorraum. Eine Abbildung $\beta\colon V \times V \to \mathbb{C}$ heißt eine *Sesquilinearform* auf V, wenn sie linear im ersten und konjugiert-linear im zweiten Argument ist: Für alle $v_1, v_1', v_2, v_2' \in V$ und alle $\lambda \in \mathbb{C}$ gilt
$$\beta(v_1 + v_1', v_2) = \beta(v_1, v_2) + \beta(v_1', v_2), \quad \beta(\lambda v_1, v_2) = \lambda \beta(v_1, v_2);$$
$$\beta(v_1, v_2 + v_2') = \beta(v_1, v_2) + \beta(v_1, v_2'), \quad \beta(v_1, \lambda v_2) = \bar{\lambda} \beta(v_1, v_2)\,.$$
Eine Sesquilinearform β auf V heißt *hermitesch*, wenn zusätzlich für alle $v_1, v_2 \in V$ gilt
$$\beta(v_2, v_1) = \overline{\beta(v_1, v_2)}\,. \qquad \diamondsuit$$

„Hermitesch" nach Charles Hermite.

„Sesqui-" bedeutet „$1\frac{1}{2}$-fach" (so wie „bi-" „zweifach" heißt); die konjugierte Linearität wird sozusagen halb gezählt (entsprechend heißt eine konjugiert-lineare Abbildung auch *semilinear*). Häufig wird in der Definition einer Sesquilinearform Linearität im zweiten und Semilinearität im ersten Argument gefordert (also umgekehrt wie in der Definition oben). Das macht keinen wesentlichen Unterschied; man muss nur beim Rechnen aufpassen, wann man Skalare konjugiert herausziehen muss.

Wir erinnern uns an die komplexe Konjugation: Für $z = x + y\boldsymbol{i} \in \mathbb{C}$ mit $x, y \in \mathbb{R}$ ist $\bar{z} = x - y\boldsymbol{i}$. Dann gilt für $z_1, z_2 \in \mathbb{C}$

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2 \,, \qquad \overline{z_1 z_2} = \bar{z}_1 \, \bar{z}_2 \,, \quad \text{und} \quad z\bar{z} = |z|^2 = x^2 + y^2 \,.$$

z ist genau dann reell, wenn $z = \bar{z}$ ist.

Für eine hermitesche Sesquilinearform β auf V gilt $\beta(v, v) \in \mathbb{R}$ für alle $v \in V$, denn

$$\beta(v, v) = \overline{\beta(v, v)} \,.$$

Daher ist die folgende Definition sinnvoll.

∗ 21.4. **Definition.** Sei V ein komplexer Vektorraum und sei β eine hermitesche Sesquilinearform auf V. β heißt *positiv definit*, wenn für alle $\mathbf{0} \neq v \in V$ gilt $\beta(v, v) > 0$. Eine positiv definite hermitesche Sesquilinearform auf V heißt auch ein *unitäres Skalarprodukt* auf V.

DEF
unitäres Skalarprod.

unitärer Vektorraum

Ein komplexer Vektorraum V zusammen mit einem unitären Skalarprodukt auf V heißt *unitärer Vektorraum*. Wie im reellen Fall schreiben wir das Skalarprodukt in einem unitären Vektorraum meistens in der Form $\langle v_1, v_2 \rangle$. ♢

21.5. **Beispiele.**

BSP
unitäre Vektorräume

(1) Das Standardbeispiel ist $\mathbb{C}^n$ mit dem Standard-Skalarprodukt

$$\big\langle (x_1, x_2, \ldots, x_n)^\top, (y_1, y_2, \ldots, y_n)^\top \big\rangle = x_1\bar{y}_1 + x_2\bar{y}_2 + \ldots + x_n\bar{y}_n \,.$$

(2) Auch das Beispiel aus der Analysis lässt sich übertragen:
Der Raum $\mathcal{C}([a, b], \mathbb{C})$ der stetigen Funktionen $[a, b] \to \mathbb{C}$ mit dem Skalarprodukt

$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}\, dx \,.$$

ist ein unitärer Vektorraum. ♣

Damit können wir jetzt Längen und Winkel einführen.

∗ 21.6. **Definition.** Sei V ein euklidischer oder unitärer Vektorraum. Für einen Vektor $v \in V$ heißt $\|v\| = \sqrt{\langle v, v \rangle}$ die *Länge* von v. Gilt $\|v\| = 1$, dann heißt v ein *Einheitsvektor*. ♢

DEF
Länge

Einheitsvektor

Es gilt dann $\|v\| \geq 0$ sowie $\|v\| = 0 \iff v = \mathbf{0}$.

Im Standardraum $\mathbb{R}^n$ ist $\|(x_1, \ldots, x_n)^\top\| = \sqrt{x_1^2 + \ldots + x_n^2}$ die übliche euklidische Länge eines Vektors. Im $\mathbb{C}^n$ gilt analog $\|(x_1, \ldots, x_n)^\top\| = \sqrt{|x_1|^2 + \ldots + |x_n|^2}$. Die Standardbasis besteht jeweils aus Einheitsvektoren.

Wir beweisen einige Eigenschaften der Länge.

SATZ $\ast$
Cauchy-Schwarzsche Ungleichung
Dreiecksungl.

21.7. **Satz.** *Sei V ein euklidischer oder unitärer Vektorraum.*

(1) *Für $v \in V$ und $\lambda \in \mathbb{R}$ bzw. $\mathbb{C}$ gilt $\|\lambda v\| = |\lambda|\|v\|$.*

(2) *(Cauchy-Schwarzsche Ungleichung) Für $v, w \in V$ gilt $|\langle v, w\rangle| \le \|v\|\|w\|$ mit Gleichheit genau dann, wenn v und w linear abhängig sind.*

(3) *(Dreiecksungleichung) Für $v, w \in V$ gilt $\|v+w\| \le \|v\|+\|w\|$ mit Gleichheit genau dann, wenn $v = \lambda w$ oder $w = \lambda v$ ist mit $\lambda \in \mathbb{R}_{\ge 0}$.*

A.-L. Cauchy
1789–1857

H.A. Schwarz
1843–1921

Beweis.

(1) $\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{\lambda\bar{\lambda}\langle v, v\rangle} = \sqrt{|\lambda|^2\langle v, v\rangle} = |\lambda|\sqrt{\langle v, v\rangle} = |\lambda|\|v\|$.

(2) Die Aussage ist klar für $w = \mathbf{0}$. Wir können also $w \neq \mathbf{0}$ annehmen. Sei

$$v' = v - \frac{\langle v, w\rangle}{\|w\|^2} w\,;$$

dann ist

$$\langle v', w\rangle = \langle v, w\rangle - \frac{\langle v, w\rangle}{\|w\|^2}\langle w, w\rangle = 0$$

und damit

$$0 \le \langle v', v'\rangle = \langle v', v\rangle = \langle v, v\rangle - \frac{\langle v, w\rangle}{\|w\|^2}\langle w, v\rangle = \|v\|^2 - \frac{|\langle v, w\rangle|^2}{\|w\|^2}\,,$$

was zur behaupteten Ungleichung äquivalent ist. Gleichheit gilt genau dann, wenn $v' = \mathbf{0}$ ist; daraus folgt, dass v ein skalares Vielfaches von w ist. Ist umgekehrt $v = \lambda w$, dann ist $v' = \mathbf{0}$ und es gilt Gleichheit in der Ungleichung.

(3) Es gilt unter Verwendung der Cauchy-Schwarzschen Ungleichung

$$\begin{aligned}
\|v + w\|^2 &= \langle v + w, v + w\rangle = \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle \\
&= \langle v, v\rangle + \langle v, w\rangle + \overline{\langle v, w\rangle} + \langle w, w\rangle = \|v\|^2 + 2\operatorname{Re}(\langle v, w\rangle) + \|w\|^2 \\
&\le \|v\|^2 + 2|\langle v, w\rangle| + \|w\|^2 \le \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2\,.
\end{aligned}$$

Die Ungleichung folgt. Gleichheit ist äquivalent zu $\langle v, w\rangle = \|v\|\|w\|$; dafür müssen v und w linear abhängig sein, und damit das Skalarprodukt links positiv ist (beachte $\langle \lambda w, w\rangle = \lambda\|w\|^2$), muss der Skalarfaktor reell und ≥ 0 sein. ❑

Die Eigenschaften (1) und (3) (zusammen mit $\|v\| = 0 \implies v = \mathbf{0}$) besagen, dass $\|\cdot\|$ eine *Norm* auf V ist. Daraus folgt insbesondere, dass

$$(v, w) \longmapsto d(v, w) = \|v - w\|$$

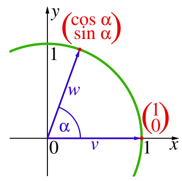
eine *Metrik* auf V ist. Damit wird V in natürlicher Weise zu einem metrischen Raum (diese Begriffe werden in der Analysis erklärt und studiert). Wir nennen $d(v, w)$ den *Abstand* zwischen v und w.

DEF ∗
Winkel
orthogonal
orthonormal
ONB

21.8. Definition. Sei V ein euklidischer oder unitärer Vektorraum.

(1) Im euklidischen Fall ist für zwei Vektoren $v, w \in V$ mit $v, w \neq \mathbf{0}$ der *Winkel* zwischen v und w die Zahl $\alpha = \angle(v, w) \in [0, \pi]$ mit $\|v\|\|w\| \cos\alpha = \langle v, w\rangle$ (oder $\cos\alpha = \left\langle \frac{v}{\|v\|}, \frac{w}{\|w\|} \right\rangle$).

(2) Zwei Vektoren $v, w \in V$ heißen *orthogonal* (oder *zueinander senkrecht*), wenn $\langle v, w\rangle = 0$ ist. Wir schreiben dafür $v \perp w$.



$\langle \binom{1}{0}, \binom{\cos\alpha}{\sin\alpha} \rangle = \cos\alpha$

(3) Sei $U \subset V$ ein Untervektorraum. Dann heißt

$$U^\perp = \{v \in V \mid \forall w \in U : v \perp w\}$$

das *orthogonale Komplement* von U in V.

(4) Eine Teilmenge $A \subset V$ heißt *orthogonal*, wenn ihre Elemente paarweise orthogonal sind ($\forall v, w \in A : v \neq w \Rightarrow v \perp w$). A heißt *orthonormal*, wenn zusätzlich alle Elemente von A Einheitsvektoren sind.

(5) Eine Basis B von V heißt eine *Orthonormalbasis* oder kurz *ONB* von V, wenn sie aus paarweise orthogonalen Einheitsvektoren besteht (wenn also die Menge der Vektoren in B orthonormal ist). ♢

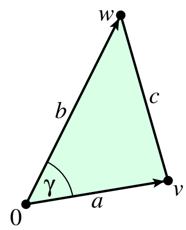
Die Cauchy-Schwarzsche Ungleichung stellt sicher, dass die Definition des Winkels sinnvoll ist, denn es gilt ja für jedes euklidische Skalarprodukt

$$-1 \le \frac{\langle v, w\rangle}{\|v\|\|w\|} \le 1 \,.$$

Zwei Vektoren in einem euklidischen Vektorraum sind genau dann orthogonal, wenn wenigstens einer der Nullvektor ist oder der Winkel zwischen ihnen $\pi/2$ (entsprechend $90°$) ist.

BSP
Cosinussatz
Pythagoras

21.9. Beispiel. Klassische Sätze über Dreiecke lassen sich elegant durch Rechnen in euklidischen Vektorräumen beweisen: Seien V ein euklidischer Vektorraum und $v, w \in V$; wir betrachten das Dreieck mit Eckpunkten $\mathbf{0}$, v und w; es hat Seitenlängen $a = \|v\|$, $b = \|w\|$, $c = \|v - w\|$; der Winkel bei $\mathbf{0}$ sei γ. Dann gilt $\langle v, w\rangle = \|v\|\|w\| \cos\gamma$, also



$$c^2 = \|v - w\|^2 = \|v\|^2 - 2\langle v, w\rangle + \|w\|^2 = a^2 - 2ab\cos\gamma + b^2 \,;$$

das ist der *Cosinussatz*. Gilt $v \perp w$ (dann ist γ ein rechter Winkel), dann vereinfacht sich das zum *Satz des Pythagoras*

$$c^2 = a^2 + b^2 \,.$$ ♣

Die Standardbasis ist eine Orthonormalbasis des Standardraums $\mathbb{R}^n$ oder $\mathbb{C}^n$.

LEMMA
orthogonal
⇒ lin.unabh.

21.10. Lemma. *Sei V ein euklidischer oder unitärer Vektorraum und seien die Vektoren $v_1, v_2, \ldots, v_n \in V$ paarweise orthogonal und von $\mathbf{0}$ verschieden. Dann ist $(v_1, v_2, \ldots, v_n)$ linear unabhängig.*

Beweis. Seien $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ bzw. $\mathbb{C}$ mit $\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = \mathbf{0}$. Es folgt für alle $j \in \{1, 2, \ldots, n\}$:

$$0 = \langle \mathbf{0}, v_j\rangle = \left\langle \sum_{i=1}^n \lambda_i v_i, v_j \right\rangle = \sum_{i=1}^n \lambda_i \langle v_i, v_j\rangle = \lambda_j \|v_j\|^2 \,,$$

und weil $v_j \neq \mathbf{0}$ ist, muss $\lambda_j = 0$ sein. ❑

Gibt es immer eine Orthonormalbasis? Der folgende wichtige Satz zeigt, dass man aus jeder endlichen Basis eine Orthonormalbasis konstruieren kann.

SATZ $\ast$
Gram-Schmidt-Orthonormalisierung

21.11. **Satz.** *Sei V ein euklidischer oder unitärer Vektorraum und seien $b_1, \ldots, b_n$ linear unabhängige Vektoren in V. Wir definieren sukzessive die folgenden Vektoren $v_1, \ldots, v_n \in V$:*

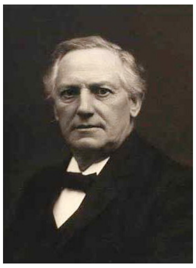
$$\begin{aligned} v_1 &= b_1 \\ v_2 &= b_2 - \frac{\langle b_2, v_1 \rangle}{\|v_1\|^2} v_1 \\ v_3 &= b_3 - \frac{\langle b_3, v_1 \rangle}{\|v_1\|^2} v_1 - \frac{\langle b_3, v_2 \rangle}{\|v_2\|^2} v_2 \\ \vdots \quad & \quad \vdots \\ v_n &= b_n - \sum_{j=1}^{n-1} \frac{\langle b_n, v_j \rangle}{\|v_j\|^2} v_j \end{aligned}$$

Dann sind $v_1, \ldots, v_n$ paarweise orthogonal, und es gilt $\langle v_1, \ldots, v_j \rangle_{\mathbb{K}} = \langle b_1, \ldots, b_j \rangle_{\mathbb{K}}$ für alle $j \in \{0, 1, \ldots, n\}$ (mit $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$).

Ist $(b_1, \ldots, b_n)$ eine Basis von V, dann ist mit

$$e_j = \frac{1}{\|v_j\|} v_j \qquad \text{für } j \in \{0, 1, \ldots, n\}$$

$(e_1, \ldots, e_n)$ eine ONB von V.



J.P. Gram
1850–1916

E. Schmidt
1876–1959
(© Konrad Jacobs)

Dieses Verfahren ist nach Jørgen Pedersen Gram und Erhard Schmidt benannt.

Beweis. Wir zeigen zunächst $\langle v_1, \ldots, v_j \rangle_{\mathbb{K}} = \langle b_1, \ldots, b_j \rangle_{\mathbb{K}}$ für alle $j \in \{0, 1, \ldots, n\}$. Wir beweisen das durch Induktion über j (bis $j = n$). Die Aussage ist klar für $j = 0$ (auf beiden Seiten steht der Null-Vektorraum). Für den Induktionsschritt nehmen wir an, dass sie für $j - 1$ richtig ist. Es ist dann

$$v_j = b_j - \sum_{i=1}^{j-1} \frac{\langle b_j, v_i \rangle}{\|v_i\|^2} v_i \in \langle v_1, \ldots, v_{j-1}, b_j \rangle_{\mathbb{K}} \overset{\text{IV}}{=} \langle b_1, \ldots, b_j \rangle_{\mathbb{K}} \,,$$

also auch $\langle v_1, \ldots, v_j \rangle_{\mathbb{K}} \subset \langle b_1, \ldots, b_j \rangle_{\mathbb{K}}$. Umgekehrt ist

$$b_j = \sum_{i=1}^{j-1} \frac{\langle b_j, e_i \rangle}{\|v_i\|^2} v_i + v_j \in \langle v_1, \ldots, v_j \rangle$$

und damit $\langle b_1, \ldots, b_j \rangle_{\mathbb{K}} \subset \langle v_1, \ldots, v_j \rangle_{\mathbb{K}}$. Insgesamt zeigt das die gewünschte Gleichheit.

Da wegen der linearen Unabhängigkeit der b_i die Dimension von $\langle b_1, \ldots, b_j \rangle_{\mathbb{K}}$ genau j ist, gilt das auch für $\langle v_1, \ldots, v_j \rangle_{\mathbb{K}}$. Insbesondere ist $(v_1, \ldots, v_j)$ linear unabhängig und damit $v_j \neq \mathbf{0}$, also ist v_{j+1} (und per Induktion alle weiteren v_i) wohldefiniert.

Wir zeigen jetzt durch Induktion, dass $\{v_1, \ldots, v_n\}$ orthogonal ist. Sei $1 \leq j \leq n$. Wir nehmen an, dass $\{v_1, \ldots, v_{j-1}\}$ orthogonal ist (für $j = 1$ gilt das trivialerweise). Dann ist auch $v_j \perp v_i$ für alle $i < j$, denn

$$\langle v_j, v_i \rangle = \langle b_j, v_i \rangle - \sum_{k=1}^{j-1} \frac{\langle b_j, v_k \rangle}{\|v_k\|^2} \langle v_k, v_i \rangle = \langle b_j, v_i \rangle - \frac{\langle b_j, v_i \rangle}{\|v_i\|^2} \langle v_i, v_i \rangle = 0 \,.$$

Mit $e_j = (1/\|v_j\|)v_j$ gilt dann in jedem Fall, dass $(e_1, \ldots, e_n)$ eine ONB von $\langle e_1, \ldots, e_n\rangle_{\mathbb{K}} = \langle b_1, \ldots, b_n\rangle_{\mathbb{K}}$ ist. Ist $(b_1, \ldots, b_n)$ eine Basis von V, dann ist dieser Untervektorraum ganz V, also ist $(e_1, \ldots, e_n)$ eine ONB von V. ❑

Man kann das von endlich vielen auf abzählbar unendlich viele Vektoren verallgemeinern (Übung).

Wenn man das Verfahren auf eine Basis des $\mathbb{R}^n$ anwendet, dann lässt sich die Beziehung zwischen der ursprünglichen Basis und der daraus konstruierten Orthonormalbasis in der Form $B = PT$ schreiben, wobei in den Spalten von B die gegebenen Basisvektoren $b_1, b_2, \ldots, b_n$ und in den Spalten von P die neuen Basisvektoren $e_1, e_2, \ldots, e_n$ stehen. Weil die neue Basis eine ONB ist, ist P eine orthogonale Matrix. Die Matrix T ist eine obere Dreiecksmatrix, deren Diagonaleinträge $\|v_1\|, \|v_2\|, \ldots, \|v_n\|$ positiv sind. Satz 21.11 impliziert also, dass jede invertierbare Matrix $B \in \mathrm{GL}(n, \mathbb{R})$ als Produkt PT geschrieben werden kann mit $P \in O(n)$ und einer oberen Dreiecksmatrix T mit positiven Diagonaleinträgen. Diese Zerlegung ist sogar eindeutig bestimmt; sie ist unter dem Namen QR-Zerlegung bekannt und wird in verschiedenen Algorithmen der numerischen Mathematik benötigt. (Wenn man die Basen in umgekehrter Reihenfolge in die Matrizen einträgt oder die Zeilen statt der Spalten verwendet oder beides, dann erhält man Versionen mit unteren Dreiecksmatrizen oder/und der umgekehrten Reihenfolge TP der Faktoren.)

21.12. **Beispiel.** Wir erzeugen eine ONB aus der Basis $B = (b_1, b_2, b_3)$ von $\mathbb{R}^3$ mit $b_1 = (1, 1, 1)$, $b_2 = (1, -1, 1)$, $b_3 = (1, 0, 0)$. Wir erhalten **BSP** ONB

$$\begin{aligned}
v_1 &= b_1 = (1, 1, 1) \\
v_2 &= b_2 - \frac{\langle b_2, v_1\rangle}{\|v_1\|^2} v_1 = (1, -1, 1) - \tfrac{1}{3}(1, 1, 1) = (\tfrac{2}{3}, -\tfrac{4}{3}, \tfrac{2}{3}) \\
v_3 &= b_3 - \frac{\langle b_3, v_1\rangle}{\|v_1\|^2} v_1 - \frac{\langle b_3, v_2\rangle}{\|v_2\|^2} v_2 = (1, 0, 0) - \tfrac{1}{3}(1, 1, 1) - \tfrac{1}{4}(\tfrac{2}{3}, -\tfrac{4}{3}, \tfrac{2}{3}) = (\tfrac{1}{2}, 0, -\tfrac{1}{2})
\end{aligned}$$

und somit

$$e_1 = \frac{1}{\sqrt{3}}(1, 1, 1), \quad e_2 = \frac{1}{\sqrt{6}}(1, -2, 1), \quad e_3 = \frac{1}{\sqrt{2}}(1, 0, -1).$$

♣

Wir rechtfertigen die Bezeichnung „orthogonales Komplement“ für $U^\perp$:

21.13. **Lemma.** *Seien V ein euklidischer oder unitärer Vektorraum und $U \subset V$ ein endlich-dimensionaler Untervektorraum. Dann ist $U^\perp$ ein Komplement von U in V (also $U + U^\perp = V$ und $U \cap U^\perp = \{\mathbf{0}\}$).* **LEMMA** $U^\perp$ ist Komplement von U

Beweis. Sei $v \in U \cap U^\perp$. Dann folgt aus der Definition von $U^\perp$, dass $\langle v, v\rangle = 0$ und damit $v = \mathbf{0}$ ist. Also ist $U \cap U^\perp = \{\mathbf{0}\}$.

Sei jetzt $v \in V$ beliebig und $(e_1, \ldots, e_n)$ eine ONB von U (die nach Satz 21.11 existiert). Wir setzen

$$v_1 = \langle v, e_1\rangle e_1 + \langle v, e_2\rangle e_2 + \ldots + \langle v, e_n\rangle e_n \in U \quad \text{und} \quad v_2 = v - v_1 .$$

Dann gilt jedenfalls $v = v_1 + v_2$. Es bleibt zu zeigen, dass $v_2 \in U^\perp$ ist. Sei dazu $u = \lambda_1 e_1 + \ldots + \lambda_n e_n \in U$. Dann gilt

$$\langle v_2, u\rangle = \sum_{j=1}^{n} \bar{\lambda}_j \langle v - v_1, e_j\rangle$$

und

$$\langle v - v_1, e_j \rangle = \langle v, e_j \rangle - \sum_{i=1}^{n} \langle v, e_i \rangle \langle e_i, e_j \rangle = \langle v, e_j \rangle - \langle v, e_j \rangle = 0 \,.$$

Also ist $v_2 \perp u$ für alle $u \in U$, damit $v_2 \in U^\perp$ und $v \in U + U^\perp$. ❑

Die Voraussetzung, dass U endlich-dimensional ist, ist hier wesentlich. Anderenfalls muss die Gleichung $V = U + U^\perp$ nicht gelten. Ein Beispiel ist

$$V = \ell^2 = \left\{ (a_n) \in \mathbb{R}^{\mathbb{N}} \mid \sum_{n=0}^{\infty} a_n^2 < \infty \right\}$$

mit dem Skalarprodukt

$$\langle (a_n), (b_n) \rangle = \sum_{n=0}^{\infty} a_n b_n \,.$$

Für den Untervektorraum

$$U = \{(a_n) \in V \mid \exists N \in \mathbb{N}\, \forall n > N \colon a_n = 0\}$$

der „endlichen“ Folgen gilt $U^\perp = \{\mathbf{0}\}$, denn die Folgen $\mathbf{e}_m = (\delta_{mn})_{n\in\mathbb{N}}$ sind in U, und $\langle (a_n), \mathbf{e}_m \rangle = a_m$, also impliziert $(a_n) \in U^\perp$, dass $(a_n) = \mathbf{0}$ ist. Auf der anderen Seite ist aber $U \neq V$, denn z.B. ist $(2^{-n})_{n\in\mathbb{N}} \in V \setminus U$.

Die Aussagen lassen sich noch etwas verfeinern.

SATZ ∗ Parsevalsche Gleichung

21.14. **Satz.** *Sei V ein euklidischer oder unitärer Vektorraum mit Orthonormalbasis $(e_1, e_2, \ldots, e_n)$.*

(1) *Für alle $v \in V$ gilt $v = \langle v, e_1 \rangle e_1 + \langle v, e_2 \rangle e_2 + \ldots + \langle v, e_n \rangle e_n$.*

(2) *Für alle $v \in V$ gilt $\|v\|^2 = |\langle v, e_1 \rangle|^2 + |\langle v, e_2 \rangle|^2 + \ldots + |\langle v, e_n \rangle|^2$.*

(3) *Für alle $v, w \in V$ gilt $\langle v, w \rangle = \langle v, e_1 \rangle \overline{\langle w, e_1 \rangle} + \ldots + \langle v, e_n \rangle \overline{\langle w, e_n \rangle}$.*

Beweis. Sei die rechte Seite in der ersten Gleichung u, dann ist wie im Beweis von Lemma 21.13 (mit $V = U$) $v - u \in V^\perp = \{\mathbf{0}\}$, also $v = u$.

Die beiden weiteren Aussagen folgen aus (1), da $\langle e_i, e_j \rangle = \delta_{ij}$. ❑

Die eigentliche Parsevalsche Gleichung ist eine analoge Aussage in sogenannten „Prä-Hilbert-Räumen“: Eine (auch unendliche) orthonormale Menge $A \subset V$ ist genau dann ein „vollständiges Orthonormalsystem“ (das bedeutet, dass der von A erzeugte Untervektorraum von V bezüglich der durch das Skalarprodukt definierten Topologie dicht in V ist), wenn Gleichung (2) oben in der Form $\|v\|^2 = \sum_{e\in A} |\langle v, e \rangle|^2$ für alle $v \in V$ gilt.

Die letzte Aussage im obigen Satz besagt also, dass die Matrix der Bilinearform (oder Sesquilinearform) $\langle \cdot, \cdot \rangle$ bezüglich einer ONB die Einheitsmatrix ist:

$$\langle x_1 e_1 + x_2 e_2 + \ldots + x_n e_n, y_1 e_1 + y_2 e_2 + \ldots + y_n e_n \rangle = x_1 \bar{y}_1 + x_2 \bar{y}_2 + \ldots + x_n \bar{y}_n \,.$$

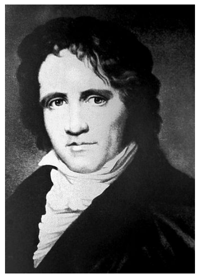
Für eine Sesquilinearform β ist die Matrix $A = \mathrm{Mat}_B(\beta)$ genauso definiert wie für Bilinearformen. Mit $B = (b_1, \ldots, b_n)$ gilt dann

$$\beta(x_1 b_1 + \ldots + x_n b_n, y_1 b_1 + \ldots + y_n b_n) = (x_1\ x_2\ \cdots\ x_n) A \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{pmatrix} .$$

Wenn $(e_1, \ldots, e_n)$ zwar orthonormal, aber nicht unbedingt eine Basis ist, dann gilt immerhin noch eine Ungleichung:

* **21.15. Satz.** *Sei V ein euklidischer oder unitärer Vektorraum und sei außerdem $\{e_1, e_2, \ldots, e_n\} \subset V$ eine orthonormale Menge. Dann gilt für alle $v \in V$*

SATZ
Besselsche Ungleichung



F.W. Bessel
1784–1846

$$\|v\|^2 \geq |\langle v, e_1\rangle|^2 + |\langle v, e_2\rangle|^2 + \ldots + |\langle v, e_n\rangle|^2$$

mit Gleichheit genau für $v \in \langle e_1, e_2, \ldots, e_n\rangle_{\mathbb{K}}$ (mit $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$).

Beweis. Sei $U = \langle e_1, e_2, \ldots, e_n\rangle_{\mathbb{K}}$. Wie im Beweis von Lemma 21.13 können wir $v \in V$ schreiben als $v = u + v'$ mit $u = \sum_{j=1}^n \langle v, e_j\rangle e_j \in U$ und $v' \in U^\perp$. Dann gilt $\|v\|^2 = \|u\|^2 + \|v'\|^2 = \sum_{j=1}^n |\langle v, e_j\rangle|^2 + \|v'\|^2$, wobei wir Satz 21.14 verwendet haben. Daraus folgt die Ungleichung; Gleichheit ist äquivalent mit $v' = \mathbf{0}$, also mit $v \in U$. ❑

Auch diese Ungleichung gilt allgemeiner für beliebige orthonormale Mengen oder Familien. Dies folgt aus dem oben formulierten endlichen Fall, indem man beliebige endliche Teilmengen oder -familien betrachtet.

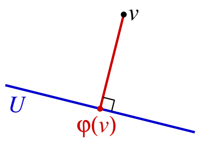
Die Metrik auf einem euklidischen oder unitären Vektorraum V hat die schöne Eigenschaft, dass es zu einem beliebigen Vektor $v \in V$ und einem beliebigen endlich-dimensionalen Untervektorraum $U \subset V$ stets ein eindeutig bestimmtes Element $u \in U$ mit minimalem Abstand zu v gibt.

21.16. Lemma. *Sei V ein euklidischer oder unitärer Vektorraum, sei $U \subset V$ ein endlich-dimensionaler Untervektorraum und sei $v \in V$. Dann gibt es ein eindeutig bestimmtes $u = \varphi(v) \in U$ mit*

LEMMA
orthogonale Projektion

$$d(v, u) = \|v - u\| = d(v, U) := \min\{d(v, u') \mid u' \in U\}.$$

Die Abbildung $\varphi\colon V \to U$ ist linear und erfüllt $\varphi|_U = \mathrm{id}_U$, also auch $\varphi|_U \circ \varphi = \varphi$.



Die lineare Abbildung $\varphi\colon V \to U$ heißt die *orthogonale Projektion* auf U.

DEF
orthogonale Projektion

Beweis. Nach Lemma 21.13 können wir v eindeutig in der Form $v = u + v'$ schreiben mit $u \in U$ und $v' \in U^\perp$. Dann hat $\varphi(v) = u$ die geforderte Eigenschaft. Sei dazu $u' \in U$ beliebig. Dann gilt

$$\begin{aligned}\|v - u'\|^2 &= \|(v - u) + (u - u')\|^2 = \|v' + (u - u')\|^2 \\ &= \|v'\|^2 + \|u - u'\|^2 \geq \|v'\|^2 = \|v - u\|^2\end{aligned}$$

(beachte, dass $v' \perp u - u'$ ist). Das zeigt, dass u den Abstand zu v minimiert und dass für jedes $u' \neq u$ der Abstand zu v größer ist.

Die Aussagen über φ folgen daraus, dass φ die Projektion $V \to U$ bezüglich der Zerlegung $V = U \oplus U^\perp$ ist. ❑

Die Aussage des Lemmas kann falsch sein, wenn U unendlich-dimensional ist. Seien zum Beispiel

$$V = \ell^2 = \Big\{(a_n) \in \mathbb{R}^{\mathbb{N}} \;\Big|\; \sum_{n=0}^{\infty} a_n^2 < \infty\Big\}$$

mit dem Skalarprodukt

$$\langle (a_n), (b_n)\rangle = \sum_{n=0}^{\infty} a_n b_n$$

und

$$U = \{(a_n) \in V \mid \exists N \in \mathbb{N}\, \forall n > N\colon a_n = 0\}$$

der Untervektorraum der „endlichen" Folgen. Für jede Folge $a \in V \setminus U$ gilt dann

$$d(a, U) := \inf\{d(a, u) \mid u \in U\} = 0\,,$$

aber wegen $a \notin U$ gibt es kein $u \in U$ mit $d(a,u) = 0$.

BSP
lineare Regression

C.F. Gauß
(1777–1855)

21.17. **Beispiel.** Es seien n Punkte $(x_j, y_j) \in \mathbb{R}^2$ gegeben. Wir suchen eine Gerade $y = ax+b$, die möglichst nahe an diesen Punkten vorbei läuft, wobei wir „nahe" durch den Unterschied in der y-Koordinate messen, also durch $|y_j - (ax_j + b)|$. Gauß hat dazu die *Methode der kleinsten Quadrate* eingeführt: Wir möchten $a, b \in \mathbb{R}$ so bestimmen, dass der „Fehler"

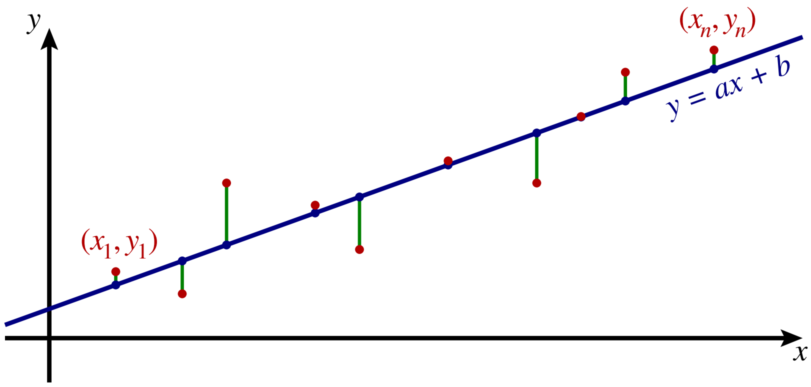
$$F(a,b) = |y_1 - (ax_1 + b)|^2 + |y_2 - (ax_2 + b)|^2 + \ldots + |y_n - (ax_n + b)|^2$$

minimal wird. Wir betrachten dazu die lineare Abbildung

$$\phi \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^n, \quad (a,b) \longmapsto (ax_1 + b, ax_2 + b, \ldots, ax_n + b);$$

wir setzen $U = \operatorname{im}(\phi)$. Dann ist, mit $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$,

$$F(a,b) = \|\phi(a,b) - \boldsymbol{y}\|^2 .$$



Der Fehler wird also genau dann minimal, wenn $\phi(a,b)$ das Bild von $\boldsymbol{y}$ unter der orthogonalen Projektion auf U ist. Falls ϕ injektiv ist (das ist der Fall, sobald man mindestens zwei verschiedene x_j hat), dann ist (a,b) dadurch eindeutig bestimmt. ♣

Isomorphismen zwischen euklidischen oder unitären Vektorräumen, die zusätzlich das Skalarprodukt erhalten, haben einen besonderen Namen.

DEF $*$
Isometrie

21.18. **Definition.** Seien V und W zwei euklidische oder zwei unitäre Vektorräume. Eine Abbildung $f \colon V \to W$ heißt *(lineare) Isometrie*, wenn f ein Isomorphismus ist und zusätzlich für alle $v, v' \in V$ gilt, dass $\langle f(v), f(v') \rangle = \langle v, v' \rangle$ ist. (Hier steht links das Skalarprodukt von W, rechts das von V.) Gibt es so eine Isometrie, dann heißen V und W *isometrisch*. ♢

Das Skalarprodukt $\langle v, w \rangle$ lässt sich durch Längen ausdrücken:

$$4\langle v, w \rangle = \|v + w\|^2 - \|v - w\|^2$$

im euklidischen Fall und

$$4\langle v, w \rangle = \|v + w\|^2 + \boldsymbol{i}\|v + \boldsymbol{i}w\|^2 - \|v - w\|^2 - \boldsymbol{i}\|v - \boldsymbol{i}w\|^2$$

im unitären Fall. Daher genügt es, statt der zweiten Bedingung nur zu fordern, dass $\|f(v)\| = \|v\|$ ist für alle $v \in V$.

Man kann das im euklidischen Fall so interpretieren, dass eine lineare Abbildung, die Längen erhält, auch Winkel erhalten muss. Das liegt daran, dass ein Dreieck durch die

Längen seiner drei Seiten bis auf Kongruenz eindeutig bestimmt ist: Die Winkel sind durch die Längen festgelegt.

Man kann sich auch überlegen, dass eine Abbildung $f\colon V \to W$ zwischen zwei euklidischen oder unitären Vektorräumen, die die Bedingung $\|f(v)\| = \|v\|$ für alle $v \in V$ erfüllt, linear sein muss.

21.19. **Beispiel.** Ist V ein euklidischer Vektorraum mit ONB $(e_1, e_2, \ldots, e_n)$, dann ist die Abbildung **BSP** Isometrie

$$\mathbb{R}^n \longrightarrow V, \quad (x_1, x_2, \ldots, x_n) \longmapsto x_1 e_1 + x_2 e_2 + \ldots + x_n e_n$$

eine Isometrie. Das ist gerade der Inhalt von Satz 21.14. Analog für einen unitären Vektorraum und die entsprechende Abbildung $\mathbb{C}^n \to V$.

So wie jeder n-dimensionale K-Vektorraum zum Standard-Vektorraum K^n isomorph ist, ist also jeder n-dimensionale euklidische Vektorraum zum euklidischen Standard-Vektorraum $\mathbb{R}^n$ isometrisch, und jeder n-dimensionale unitäre Vektorraum ist zum unitären Standard-Vektorraum $\mathbb{C}^n$ isometrisch. ♣

Allgemein gilt, analog zur Charakterisierung von Isomorphismen mittels des Bildes einer Basis:

21.20. **Lemma.** *Seien V und W zwei endlich-dimensionale euklidische oder unitäre Vektorräume und sei $(e_1, \ldots, e_n)$ eine Orthonormalbasis von V. Eine lineare Abbildung $f\colon V \to W$ ist genau dann eine Isometrie, wenn $(f(e_1), \ldots, f(e_n))$ eine Orthonormalbasis von W ist.* **LEMMA** Kriterium für Isometrie

Beweis. „⇒“: Wenn f eine Isometrie ist, dann gilt für $i, j \in \{1, \ldots, n\}$

$$\langle f(e_i), f(e_j)\rangle = \langle e_i, e_j\rangle = \delta_{ij}\,.$$

Das bedeutet gerade, dass $(f(e_1), \ldots, f(e_n))$ orthonormal ist. Da f nach Voraussetzung auch ein Isomorphismus ist, gilt $\dim W = \dim V = n$. Die n linear unabhängigen Vektoren $f(e_1), \ldots, f(e_n)$ bilden daher auch eine Basis von W.

„⇐“: Zunächst einmal folgt daraus, dass f eine Basis von V auf eine Basis von W abbildet, dass f ein Isomorphismus ist. Sei $v \in V$. Dann gilt nach Satz 21.14:

$$v = \langle v, e_1\rangle e_1 + \ldots + \langle v, e_n\rangle e_n\,.$$

Da $(f(e_1), \ldots, f(e_n))$ eine ONB von W ist, folgt ebenso

$$f(v) = \langle f(v), f(e_1)\rangle f(e_1) + \ldots + \langle f(v), f(e_n)\rangle f(e_n)\,.$$

Auf der anderen Seite ist

$$f(v) = f\bigl(\langle v, e_1\rangle e_1 + \ldots + \langle v, e_n\rangle e_n\bigr) = \langle v, e_1\rangle f(e_1) + \ldots + \langle v, e_n\rangle f(e_n)\,,$$

woraus (Koordinaten bezüglich einer Basis sind eindeutig bestimmt)

$$\langle f(v), f(e_j)\rangle = \langle v, e_j\rangle \qquad \text{für alle } j \in \{1, \ldots, n\}.$$

folgt. Ist $w = \lambda_1 e_1 + \ldots + \lambda_n e_n \in V$ ein weiterer Vektor, dann ergibt sich

$$\begin{aligned}\langle v, w\rangle &= \langle v, \lambda_1 e_1 + \ldots + \lambda_n e_n\rangle = \lambda_1\langle v, e_1\rangle + \cdots + \lambda_n\langle v, e_n\rangle \\ &= \lambda_1\langle f(v), f(e_1)\rangle + \cdots + \lambda_n\langle f(v), f(e_n)\rangle \\ &= \langle f(v), \lambda_1 f(e_1) + \ldots + \lambda_n f(e_n)\rangle = \langle f(v), f(w)\rangle\,.\end{aligned}$$

Damit ist f eine Isometrie. ❑

22. Orthogonale und unitäre Diagonalisierung

Sei V ein endlich-dimensionaler euklidischer Vektorraum. Wir hatten schon in Lemma 20.12 gesehen, dass eine positiv definite symmetrische Bilinearform auf einem endlich-dimensionalen reellen Vektorraum nicht-ausgeartet ist; dies lässt sich also auf das euklidische Skalarprodukt von V anwenden (und funktioniert analog im unitären Fall):

LEMMA Linearformen via $\langle \cdot, \cdot \rangle$

22.1. **Lemma.** *Sei V ein endlich-dimensionaler euklidischer oder unitärer Vektorraum und sei $\phi \in V^*$ eine Linearform auf V. Dann gibt es ein eindeutig bestimmtes Element $w \in V$ mit $\phi(v) = \langle v, w \rangle = \overline{\langle w, v \rangle}$ für alle $v \in V$.*

Beweis. Wir schreiben $\beta(v, w) = \langle v, w \rangle$ für das Skalarprodukt. Im euklidischen Fall verwenden wir Lemma 20.12: Da β nicht-ausgeartet ist, ist $\beta_R \colon V \to V^*$, $w \mapsto (v \mapsto \langle v, w \rangle)$, ein Isomorphismus. Dann ist klar, dass $w = \beta_R^{-1}(\phi)$ als einziges Element von V die gewünschte Eigenschaft hat.

Im unitären Fall haben wir auch die Abbildung $\beta_R \colon V \to V^*$, die $w \in V$ abbildet auf $v \mapsto \beta(v, w)$; β_R ist jetzt allerdings nicht linear, sondern semilinear, da $\beta_R(\lambda w) = \bar{\lambda}\beta_R(w)$ gilt (das kommt von der Semilinearität von β im zweiten Argument). Man kann das reparieren, indem man statt V^* den „konjugierten" Vektorraum $\overline{V^*}$ betrachtet, der dieselbe zugrunde liegende Menge und dieselbe Addition hat wie V^*, aber die geänderte Skalarmultiplikation $\lambda \cdot \phi = \bar{\lambda}\phi$ (links die Skalarmultiplikation von $\overline{V^*}$, rechts die von V^*). Dann ist $\beta_R \colon V \to \overline{V^*}$ eine lineare Abbildung zwischen komplexen Vektorräumen derselben endlichen Dimension. Wie im euklidischen Fall sieht man, dass $\ker(\beta_R) = \{\mathbf{0}\}$ ist; damit ist β_R ein Isomorphismus und wir können wie im euklidischen Fall schließen, dass $w = \beta_R^{-1}(\phi)$ das einzige Element mit der gewünschten Eigenschaft ist. ❑

Für β_L haben wir im unitären Fall einen Isomorphismus $V \to \bar{V}^*$ von V auf den Dualraum des konjugierten Vektorraums $\bar{V}$.

Für unendlich-dimensionale euklidische bzw. unitäre Vektorräume ist die Aussage des Lemmas im Allgemeinen falsch.

Für einen endlich-dimensionalen euklidischen Vektorraum gibt es also einen *kanonischen* Isomorphismus $V \to V^*$. Ist $(e_1, e_2, \ldots, e_n)$ eine Orthonormalbasis von V und $(e_1^*, e_2^*, \ldots, e_n^*)$ die dazu duale Basis von V^*, dann identifiziert dieser Isomorphismus e_j mit e_j^*, denn

$$\langle e_i, e_j \rangle = \delta_{ij} = e_j^*(e_i)\,.$$

Im unitären Fall ergibt sich ein Isomorphismus $V \to \overline{V^*}$.

BSP Vektorprodukt

22.2. **Beispiel.** Sei $V = \mathbb{R}^3$ (Elemente als Spaltenvektoren) mit dem Standard-Skalarprodukt. Seien $v_1, v_2 \in V$. Dann ist $v \mapsto \det(v_1, v_2, v)$ (das bezeichne die Determinante der Matrix mit den Spalten v_1, v_2, v) eine Linearform auf V, also gibt es nach Lemma 22.1 einen eindeutig bestimmten Vektor $v_1 \times v_2 \in V$ mit

$$\langle v_1 \times v_2, v \rangle = \det(v_1, v_2, v) \qquad \text{für alle } v \in V.$$

DEF Vektorprodukt

Dieser Vektor $v_1 \times v_2$ heißt das *Vektorprodukt* oder *Kreuzprodukt* von v_1 und v_2. Das Vektorprodukt hat folgende Eigenschaften:

(1) Die Abbildung $V \times V \longrightarrow V$, $(v_1, v_2) \longmapsto v_1 \times v_2$, ist bilinear.

Das folgt aus der Linearität der Determinante in jeder Spalte der Matrix.

(2) Sind $v_1, v_2 \in V$ linear abhängig, dann ist $v_1 \times v_2 = \mathbf{0}$.

Das folgt daraus, dass dann $\det(v_1, v_2, v) = 0$ ist für alle v.

(3) Für alle $v_1, v_2 \in V$ gilt $(v_1 \times v_2) \perp v_1$ und $(v_1 \times v_2) \perp v_2$.

$\langle v_1 \times v_2, v_1 \rangle = \det(v_1, v_2, v_1) = 0$; ebenso für v_2.

(4) Es gilt die explizite Formel

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix} .$$

Beweis als Übung.

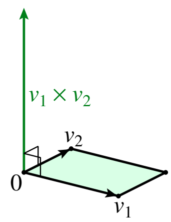
(5) Für alle $v_1, v_2 \in V \setminus \{\mathbf{0}\}$ gilt $\|v_1 \times v_2\| = \|v_1\|\|v_2\| \sin \angle(v_1, v_2)$.

Beweis als Übung. Zu zeigen ist $\|v_1 \times v_2\|^2 + \langle v_1, v_2 \rangle^2 = \|v_1\|^2\|v_2\|^2$.

(6) Sind v_1, v_2 linear unabhängig, dann ist $(v_1, v_2, v_1 \times v_2)$ eine positiv orientierte Basis von $\mathbb{R}^3$.

Ergänze (v_1, v_2) zu einer Basis (v_1, v_2, v); dann ist $\det(v_1, v_2, v) \neq 0$, also kann $v_1 \times v_2$ nicht der Nullvektor sein. Damit folgt

$$\det(v_1, v_2, v_1 \times v_2) = \langle v_1 \times v_2, v_1 \times v_2 \rangle = \|v_1 \times v_2\|^2 > 0\,.$$



Das lässt sich so interpretieren, dass $v_1 \times v_2$ der Nullvektor ist, wenn v_1 und v_2 linear abhängig sind; anderenfalls ist es ein Vektor, der auf der von v_1 und v_2 aufgespannten Ebene senkrecht steht und dessen Länge der Fläche des von v_1 und v_2 aufgespannten Parallelogramms entspricht. Dabei ist die Richtung so, dass $(v_1, v_2, v_1 \times v_2)$ eine positiv orientierte Basis bilden („Rechte-Hand-Regel“, siehe Definition 14.21). ♣

Der $\mathbb{R}^3$ zusammen mit dem Vektorprodukt ist ein Beispiel für eine sogenannte *Lie-Algebra* (nach Sophus Lie). Eine Lie-Algebra über einem Körper K ist ein K-Vektorraum V zusammen mit einer bilinearen Abbildung $V \times V \to V$, die üblicherweise $(v, w) \mapsto [v, w]$ geschrieben wird und die folgenden Bedingungen erfüllt:

- $\forall v \in V\colon [v, v] = \mathbf{0}$.
- $\forall v_1, v_2, v_3 \in V\colon [v_1, [v_2, v_3]] + [v_2, [v_3, v_1]] + [v_3, [v_1, v_2]] = \mathbf{0}$ (Jacobi-Identität).

Am einfachsten rechnet man die Jacobi-Identität für das Vektorprodukt nach, indem man sich überlegt, dass ihre Gültigkeit erhalten bleibt, wenn man einen Vektor skaliert oder ein skalares Vielfaches eines Vektors zu einem anderen addiert (also unter elementaren Spaltenumformungen an der aus den drei Vektoren als Spalten gebildeten Matrix). Damit kann man den Beweis auf die beiden trivialen Fälle $v_1 = \mathbf{0}$ und $(v_1, v_2, v_3) = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ reduzieren.

Ein anderes Beispiel einer Lie-Algebra ist $\mathrm{Mat}(n, K)$ mit $[A, B] = AB - BA$.

Wir schreiben weiterhin $\mathbb{K}$ für $\mathbb{R}$ (im euklidischen Kontext) oder $\mathbb{C}$ (im unitären Kontext).

22.3. **Folgerung.** *Seien V und W euklidische oder unitäre Vektorräume mit $\dim V < \infty$ und sei $f\colon V \to W$ linear. Dann gibt es eine eindeutig bestimmte lineare Abbildung $f^*\colon W \to V$ mit* **FOLG** adjungierte Abbildung

$$\langle f(v), w \rangle = \langle v, f^*(w) \rangle \qquad \textit{für alle } v \in V \textit{ und } w \in W.$$

Beweis. Sei zunächst $w \in W$ fest gewählt. Dann ist $v \mapsto \langle f(v), w\rangle$ eine Linearform auf V, also gibt es nach Lemma 22.1 ein eindeutig bestimmtes $f^*(w) \in V$ mit $\langle f(v), w\rangle = \langle v, f^*(w)\rangle$ für alle $v \in V$. Das liefert uns eine Abbildung $f^*\colon W \to V$. Es bleibt zu zeigen, dass f^* linear ist. Seien $w, w' \in W$. Dann gilt für alle $v \in V$:

$$\begin{aligned}\langle v, f^*(w + w')\rangle &= \langle f(v), w + w'\rangle = \langle f(v), w\rangle + \langle f(v), w'\rangle \\ &= \langle v, f^*(w)\rangle + \langle v, f^*(w')\rangle = \langle v, f^*(w) + f^*(w')\rangle\,;\end{aligned}$$

die Eindeutigkeitsaussage in Lemma 22.1 zeigt dann $f^*(w+w') = f^*(w)+f^*(w')$. Analog für die Skalarmultiplikation: Seien $\lambda \in \mathbb{K}$ und $w \in W$. Dann gilt für alle $v \in V$:

$$\langle v, f^*(\lambda w)\rangle = \langle f(v), \lambda w\rangle = \bar\lambda\langle f(v), w\rangle = \bar\lambda\langle v, f^*(w)\rangle = \langle v, \lambda f^*(w)\rangle\,,$$

also wie eben $f^*(\lambda w) = \lambda f^*(w)$. ❑

Die führt auf folgende Begriffsbildung.

DEF * (selbst-) adjungierte Abbildung normale Abbildung

22.4. **Definition.** Seien V und W euklidische oder unitäre Vektorräume.

(1) Sei $f\colon V \to W$ linear. Gibt es eine lineare Abbildung $f^*\colon W \to V$, sodass für alle $v \in V$ und $w \in W$ gilt $\langle f(v), w\rangle = \langle v, f^*(w)\rangle$, dann heißt f^* die *zu f adjungierte Abbildung.*

(2) Hat $f \in \operatorname{End}(V)$ eine adjungierte Abbildung f^* und gilt $f = f^*$, dann heißt f *selbst-adjungiert.* Das bedeutet also $\langle f(v), v'\rangle = \langle v, f(v')\rangle$ für alle $v, v' \in V$.

(3) Hat $f \in \operatorname{End}(V)$ eine adjungierte Abbildung f^* und gilt $f \circ f^* = f^* \circ f$, dann heißt f *normal.* ◊

Folgerung 22.3 besagt, dass es für V endlich-dimensional stets adjungierte Abbildungen gibt.

LEMMA Eigenschaften von f^*

22.5. **Lemma.** *Seien V_1, V_2 und V_3 euklidische oder unitäre Vektorräume und seien $f, g\colon V_1 \to V_2$ und $h\colon V_2 \to V_3$ linear, mit adjungierten Abbildungen f^*, g^* und h^*, und sei $\lambda \in \mathbb{K}$. Dann gilt:*

(1) *$f + g$ und λf haben adjungierte Abbildungen; es ist $(f + g)^* = f^* + g^*$ und $(\lambda f)^* = \bar\lambda f^*$.*

(2) *$h \circ f$ hat eine adjungierte Abbildung; es gilt $(h \circ f)^* = f^* \circ h^*$.*

(3) *f^* hat eine adjungierte Abbildung; es ist $(f^*)^* = f$.*

Außerdem gilt für beliebige lineare Abbildungen $f\colon V_1 \to V_2$:

(4) *f ist eine Isometrie*
$\iff$ *f ist ein Isomorphismus und hat f^{-1} als adjungierte Abbildung.*

Insbesondere sind selbst-adjungierte Abbildungen und Isometrien normal.

Beweis.

(1) Das geht ähnlich wie beim Nachweis der Linearität von f^* im Beweis von Folgerung 22.3: Für $v \in V_1$, $w \in V_2$ ist

$$\begin{aligned}\langle (f + g)(v), w\rangle &= \langle f(v) + g(v), w\rangle = \langle f(v), w\rangle + \langle g(v), w\rangle \\ &= \langle v, f^*(w)\rangle + \langle v, g^*(w)\rangle = \langle v, f^*(w) + g^*(w)\rangle\,,\end{aligned}$$

woraus $(f+g)^* = f^* + g^*$ folgt, und für $v \in V_1$, $\lambda \in \mathbb{K}$, $w \in V_2$ ist

$$\langle \lambda f(v), w\rangle = \lambda \langle f(v), w\rangle = \lambda \langle v, f^*(w)\rangle = \langle v, \bar{\lambda} f^*(w)\rangle\,,$$

und damit $(\lambda f)^* = \bar{\lambda} f^*$.

(2) Für $v \in V_1$, $w \in V_3$ gilt

$$\begin{aligned}\langle (h \circ f)(v), w\rangle &= \langle h(f(v)), w\rangle = \langle f(v), h^*(w)\rangle \\ &= \langle v, f^*(h^*(w))\rangle = \langle v, (f^* \circ h^*)(w)\rangle\,,\end{aligned}$$

also ist $(h \circ f)^* = f^* \circ h^*$.

(3) Für $v \in V_2$, $w \in V_1$ ist

$$\langle f^*(v), w\rangle = \overline{\langle w, f^*(v)\rangle} = \overline{\langle f(w), v\rangle} = \langle v, f(w)\rangle\,;$$

das bedeutet $(f^*)^* = f$.

(4) „$\Leftarrow$“: Sei f ein Isomorphismus mit $f^* = f^{-1}$. Für alle $v, w \in V_1$ gilt dann

$$\langle f(v), f(w)\rangle = \langle v, f^*(f(w))\rangle = \langle v, f^{-1}(f(w)\rangle = \langle v, w\rangle\,,$$

also ist f eine Isometrie.
„$\Rightarrow$“: Ist f eine Isometrie, dann ist f ein Isomorphismus und es gilt für alle $v \in V_1$, $w \in V_2$

$$\langle v, f^{-1}(w)\rangle = \langle f(v), f(f^{-1}(w))\rangle = \langle f(v), w\rangle\,,$$

woraus folgt, dass f^{-1} die zu f adjungierte Abbildung ist. ❑

Wir definieren analoge Begriffe für Matrizen. Die folgende Definition wiederholt im Wesentlichen Definition 20.13.

* 22.6. **Definition.** Eine Matrix $A \in \mathrm{GL}(n, \mathbb{R})$ heißt *orthogonal*, wenn sie die Gleichung $A^{-1} = A^\top$ erfüllt. Wir schreiben $\mathrm{O}(n)$ für die Gruppe (!) der orthogonalen $n \times n$-Matrizen; $\mathrm{O}(n)$ heißt die *orthogonale Gruppe*. **DEF** orthogonale Matrix

Die orthogonalen Matrizen mit Determinante 1 bilden ebenfalls eine Gruppe, die *spezielle orthogonale Gruppe* $\mathrm{SO}(n)$. ♢

Schreibt man die Bedingung $A^\top A = AA^\top = I_n$ aus, dann sieht man, dass A genau dann orthogonal ist, wenn die Spalten (Zeilen) von A eine Orthonormalbasis von $\mathbb{R}^n$ bilden.

* 22.7. **Definition.** Für eine Matrix $A = (a_{ij}) \in \mathrm{Mat}(m \times n, \mathbb{C})$ sei $\bar{A} = (\bar{a}_{ij}) \in \mathrm{Mat}(m \times n, \mathbb{C})$ und $A^* = \bar{A}^\top$. **DEF** hermitesche, unitäre, normale Matrix

Eine Matrix $A \in \mathrm{Mat}(n, \mathbb{C})$ heißt *hermitesch*, wenn $A = A^*$ ist. A heißt *unitär*, wenn $A^* A = I_n$ ist, und *normal*, wenn $AA^* = A^* A$ gilt.

Die Gruppe (!) der unitären $n \times n$-Matrizen heißt die *unitäre Gruppe* und wird mit $\mathrm{U}(n)$ bezeichnet. Die unitären Matrizen mit Determinante 1 bilden ebenfalls eine Gruppe, die *spezielle unitäre Gruppe* $\mathrm{SU}(n)$. ♢

Analog zum euklidischen Fall ist eine quadratische Matrix genau dann unitär, wenn ihre Spalten (oder Zeilen) eine ONB von $\mathbb{C}^n$ bilden. Hermitesche und unitäre Matrizen sind normal.

Für die Determinante einer unitären Matrix A gilt

$$\begin{aligned} 1 = \det(I_n) &= \det(A^* A) = \det(\bar{A}^\top) \det(A) \\ &= \det(\bar{A}) \det(A) = \overline{\det(A)} \det(A) \\ &= |\det(A)|^2 , \end{aligned}$$

also $|\det(A)| = 1$.

Die folgende Aussage zeigt, wie sich das Adjungieren auf die beschreibenden Matrizen auswirkt.

LEMMA
Matrix der adjungierten Abbildung

22.8. **Lemma.** *Seien V und W zwei endlich-dimensionale euklidische oder unitäre Vektorräume und seien $B = (e_1, e_2, \dots, e_n)$ und $B' = (e'_1, e'_2, \dots, e'_m)$ Orthonormalbasen von V bzw. W. Für eine lineare Abbildung $f \colon V \to W$ gilt dann*

$$\begin{aligned} \mathrm{Mat}_{B',B}(f^*) &= \mathrm{Mat}_{B,B'}(f)^\top \quad \textit{im euklidischen Fall bzw.} \\ \mathrm{Mat}_{B',B}(f^*) &= \mathrm{Mat}_{B,B'}(f)^* \quad \textit{im unitären Fall.} \end{aligned}$$

Beweis. Seien $A = (a_{ij}) = \mathrm{Mat}_{B,B'}(f)$ und $A' = (a'_{ij}) = \mathrm{Mat}_{B',B}(f^*)$. Es gilt nach Satz 21.14

$$\begin{aligned} f(e_j) &= \langle f(e_j), e'_1\rangle e'_1 + \langle f(e_j), e'_2\rangle e'_2 + \ldots + \langle f(e_j), e'_m\rangle e'_m \quad \text{und} \\ f^*(e'_i) &= \langle f^*(e'_i), e_1\rangle e_1 + \langle f^*(e'_i), e_2\rangle e_2 + \ldots + \langle f^*(e'_i), e_n\rangle e_n , \end{aligned}$$

also ist $a_{ij} = \langle f(e_j), e'_i\rangle$ und $a'_{ji} = \langle f^*(e'_i), e_j\rangle$ und damit für alle $i \in \{1, \dots, m\}$, $j \in \{1, \dots, n\}$

$$\bar{a}_{ij} = \overline{\langle f(e_j), e'_i\rangle} = \overline{\langle e_j, f^*(e'_i)\rangle} = \langle f^*(e'_i), e_j\rangle = a'_{ji} .$$

Es gilt also $A' = A^*$ ($= A^\top$ im euklidischen Fall) wie behauptet. ❑

FOLG
Matrix für selbst-adj. Endom.

22.9. **Folgerung.** *Sei V ein endlich-dimensionaler euklidischer oder unitärer Vektorraum mit Orthonormalbasis $B = (e_1, e_2, \dots, e_n)$ und sei $f \colon V \to V$ linear. Im euklidischen Fall gilt*

$$f \textit{ ist selbst-adjungiert} \iff \mathrm{Mat}_B(f) \textit{ ist symmetrisch}$$

und

$$f \textit{ ist eine Isometrie} \iff \mathrm{Mat}_B(f) \textit{ ist orthogonal.}$$

Im unitären Fall gilt entsprechend

$$\begin{aligned} f \textit{ ist selbst-adjungiert} &\iff \mathrm{Mat}_B(f) \textit{ ist hermitesch,} \\ f \textit{ ist normal} &\iff \mathrm{Mat}_B(f) \textit{ ist normal} \end{aligned}$$

und

$$f \textit{ ist eine Isometrie} \iff \mathrm{Mat}_B(f) \textit{ ist unitär.}$$

Beweis. Die Aussagen über selbst-adjungierte f folgen direkt aus Lemma 22.8. Die Aussagen über Isometrien folgen mit Lemma 22.5. ❑

Wir wollen uns jetzt mit der Frage beschäftigen, wann es für einen Endomorphismus f eines euklidischen oder unitären Vektorraums V eine Orthonormalbasis von V gibt, die aus Eigenvektoren von f besteht. Man sagt dann, f sei *orthogonal diagonalisierbar* bzw. *unitär diagonalisierbar*.

DEF
orthogonal/ unitär diagonalisierbar

Da sich die beiden Fälle hier doch stärker unterscheiden, behandeln wir zunächst den euklidischen Fall. Das Resultat wird sein, dass genau die selbst-adjungierten Endomorphismen orthogonal diagonalisierbar sind. Wir zeigen zuerst die einfachere Richtung.

22.10. **Satz.** *Sei V ein endlich-dimensionaler euklidischer Vektorraum und sei $f \in \operatorname{End}(V)$. Wenn f orthogonal diagonalisierbar ist, dann ist f selbst-adjungiert.*

SATZ
orthog. diag. $\Rightarrow$ selbst-adj.

Beweis. Nach Voraussetzung gibt es eine ONB B von V, sodass $A = \operatorname{Mat}_B(f)$ eine Diagonalmatrix ist. Dann gilt auch $A = A^\top$, also ist f nach Folgerung 22.9 selbst-adjungiert. ❑

Zum Beweis der Gegenrichtung brauchen wir eine Vorüberlegung, für die wir auf einige Grundtatsachen aus der Analysis zurückgreifen. Wir wissen, dass V zum Standardraum $\mathbb{R}^n$ isometrisch ist, also können wir ohne Einschränkung $V = \mathbb{R}^n$ (mit $n > 0$) betrachten. Die Menge $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| = 1\}$ (also die Oberfläche der n-dimensionalen Einheitskugel) ist eine abgeschlossene und beschränkte Teilmenge von $\mathbb{R}^n$, also ist S kompakt (Satz von Heine-Borel). Die Abbildung

$$h \colon \mathbb{R}^n \longrightarrow \mathbb{R}, \quad \boldsymbol{x} \longmapsto \langle f(\boldsymbol{x}), \boldsymbol{x} \rangle$$

ist stetig, denn $f \colon \mathbb{R}^n \to \mathbb{R}^n$ ist stetig und das Skalarprodukt ist ebenfalls stetig. Als stetige Funktion nimmt die Abbildung h auf der kompakten Menge S ihr Maximum an, etwa in $\boldsymbol{x}_0 \in S$.

22.11. **Lemma.** *Sei $n > 0$ und $f \colon \mathbb{R}^n \to \mathbb{R}^n$ eine selbst-adjungierte lineare Abbildung. Mit h und $\boldsymbol{x}_0$ wie oben ist dann $\boldsymbol{x}_0$ ein Eigenvektor von f zum Eigenwert $\lambda = h(\boldsymbol{x}_0)$.*

LEMMA
Maximum ist Eigenwert

Beweis. Wir bemerken zunächst, dass für alle $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n$ gilt

$$\langle f(\boldsymbol{x}), \boldsymbol{x} \rangle = \|\boldsymbol{x}\|^2 \Big\langle f\Big(\frac{1}{\|\boldsymbol{x}\|}\boldsymbol{x}\Big), \frac{1}{\|\boldsymbol{x}\|}\boldsymbol{x} \Big\rangle = \|\boldsymbol{x}\|^2 h\Big(\frac{1}{\|\boldsymbol{x}\|}\boldsymbol{x}\Big) \le \lambda \|\boldsymbol{x}\|^2 ,$$

denn $\|\boldsymbol{x}\|^{-1}\boldsymbol{x} \in S$. Für $\boldsymbol{x} = \mathbf{0}$ gilt die Ungleichung $\langle f(\boldsymbol{x}), \boldsymbol{x} \rangle \le \lambda \|\boldsymbol{x}\|^2$ ebenfalls.

Wir zeigen jetzt, dass $\boldsymbol{x}_0$ ein Eigenvektor ist. Wir können $f(\boldsymbol{x}_0) = \mu \boldsymbol{x}_0 + \boldsymbol{y}$ schreiben mit $\mu \in \mathbb{R}$ und $\boldsymbol{y} \perp \boldsymbol{x}_0$ (nach Lemma 21.13 mit $U = \langle \boldsymbol{x}_0 \rangle_{\mathbb{R}}$). Für $t \in \mathbb{R}$ betrachten wir den Vektor $\boldsymbol{x}_0 + t\boldsymbol{y}$. Es gilt

$$\langle f(\boldsymbol{x}_0 + t\boldsymbol{y}), \boldsymbol{x}_0 + t\boldsymbol{y} \rangle \le \lambda \|\boldsymbol{x}_0 + t\boldsymbol{y}\|^2 = \lambda(1 + t^2\|\boldsymbol{y}\|^2) = \lambda + t^2\lambda\|\boldsymbol{y}\|^2$$

(dabei haben wir $\boldsymbol{x}_0 \perp \boldsymbol{y}$ und den „Pythagoras“ benutzt). Auf der anderen Seite ist

$$\begin{aligned}
\langle f(\boldsymbol{x}_0 + t\boldsymbol{y}), \boldsymbol{x}_0 + t\boldsymbol{y} \rangle &= \langle f(\boldsymbol{x}_0) + t f(\boldsymbol{y}), \boldsymbol{x}_0 + t\boldsymbol{y} \rangle \\
&= \langle f(\boldsymbol{x}_0), \boldsymbol{x}_0 \rangle + t\big(\langle f(\boldsymbol{x}_0), \boldsymbol{y} \rangle + \langle f(\boldsymbol{y}), \boldsymbol{x}_0 \rangle\big) + t^2 \langle f(\boldsymbol{y}), \boldsymbol{y} \rangle \\
&= \lambda + t\big(\langle f(\boldsymbol{x}_0), \boldsymbol{y} \rangle + \langle \boldsymbol{y}, f(\boldsymbol{x}_0) \rangle\big) + t^2 \langle f(\boldsymbol{y}), \boldsymbol{y} \rangle \\
&= \lambda + 2t \langle f(\boldsymbol{x}_0), \boldsymbol{y} \rangle + t^2 \langle f(\boldsymbol{y}), \boldsymbol{y} \rangle \\
&= \lambda + 2t \langle \mu \boldsymbol{x}_0 + \boldsymbol{y}, \boldsymbol{y} \rangle + t^2 \langle f(\boldsymbol{y}), \boldsymbol{y} \rangle \\
&= \lambda + 2t \|\boldsymbol{y}\|^2 + t^2 \langle f(\boldsymbol{y}), \boldsymbol{y} \rangle .
\end{aligned}$$

Dabei haben wir verwendet, dass f selbst-adjungiert und $\boldsymbol{x}_0 \perp \boldsymbol{y}$ ist. Für $t > 0$ ergibt sich daraus die Ungleichung (nach Subtraktion von λ und Division durch t)

$$0 \leq 2\|\boldsymbol{y}\|^2 \leq t\big(\lambda\|\boldsymbol{y}\|^2 - \langle f(\boldsymbol{y}), \boldsymbol{y}\rangle\big)\,;$$

wenn wir t von oben gegen null gehen lassen, folgt daraus $\|\boldsymbol{y}\|^2 = 0$, also $\boldsymbol{y} = \boldsymbol{0}$ und damit $f(\boldsymbol{x}_0) = \mu\boldsymbol{x}_0$. Außerdem gilt

$$\lambda = \langle f(\boldsymbol{x}_0), \boldsymbol{x}_0\rangle = \langle \mu\boldsymbol{x}_0, \boldsymbol{x}_0\rangle = \mu\|\boldsymbol{x}_0\|^2 = \mu\,,$$

also ist λ der zu $\boldsymbol{x}_0$ gehörende Eigenwert. ❑

FOLG
selbst-adj. Abb. haben Eigenwert

22.12. **Folgerung.** *Sei V ein endlich-dimensionaler euklidischer Vektorraum mit $\dim V > 0$ und sei $f \in \operatorname{End}(V)$ selbst-adjungiert. Dann hat f einen (reellen) Eigenwert.*

Beweis. Wir wählen eine ONB von V; dann gibt es eine Isometrie $\phi\colon \mathbb{R}^n \to V$ (mit $n = \dim V$). Die Abbildung $\tilde{f} = \phi^{-1} \circ f \circ \phi \in \operatorname{End}(\mathbb{R}^n)$ ist ebenfalls selbst-adjungiert, denn nach Lemma 22.5 ist

$$\tilde{f}^* = (\phi^{-1} \circ f \circ \phi)^* = \phi^* \circ f^* \circ (\phi^{-1})^* = \phi^{-1} \circ f \circ \phi = \tilde{f}\,.$$

$\tilde{f}$ hat also nach Lemma 22.11 einen Eigenwert $\lambda \in \mathbb{R}$. Da $\tilde{f}$ und f dieselben Eigenwerte haben (ist $\boldsymbol{x} \in \mathbb{R}^n$ Eigenvektor von $\tilde{f}$ zum Eigenwert λ, dann ist $\phi(\boldsymbol{x})$ Eigenvektor von f zum selben Eigenwert), gilt das auch für f. ❑

SATZ ∗
Spektralsatz über $\mathbb{R}$

22.13. **Satz.** *Sei V ein endlich-dimensionaler euklidischer Vektorraum und sei $f \in \operatorname{End}(V)$. f ist genau dann orthogonal diagonalisierbar, d.h., V besitzt eine Orthonormalbasis, die aus Eigenvektoren von f besteht, wenn f selbst-adjungiert ist.*

Beweis. Die eine Richtung ist Satz 22.10. Die andere Richtung beweisen wir durch Induktion über $n = \dim V$. Für $n = 0$ ist nichts zu beweisen (die leere Familie ist eine ONB aus Eigenvektoren). Sei $n > 0$ und die Aussage für $\dim V = n-1$ richtig. Nach Folgerung 22.12 hat f einen Eigenwert $\lambda \in \mathbb{R}$; sei $e_n \in V$ ein zugehöriger Eigenvektor mit $\|e_n\| = 1$. Sei $U = \langle e_n\rangle_{\mathbb{R}}^{\perp} \subset V$ das orthogonale Komplement von $\langle e_n\rangle_{\mathbb{R}}$. Dann ist U ein f-invarianter Untervektorraum, denn für $u \in U$ gilt

$$\langle f(u), e_n\rangle = \langle u, f(e_n)\rangle = \langle u, \lambda e_n\rangle = \lambda\langle u, e_n\rangle = 0\,,$$

also ist $f(u) \in U$. U ist (mit dem auf $U \times U$ eingeschränkten Skalarprodukt von V) ein euklidischer Vektorraum mit $\dim U = n - 1$ (denn nach Lemma 21.13 ist U ein Komplement des eindimensionalen Unterraums $\langle e_n\rangle_{\mathbb{R}}$), und $f|_U$ ist ein selbst-adjungierter Endomorphismus von U. Nach der Induktionsannahme hat also U eine ONB $(e_1, e_2, \ldots, e_{n-1})$, die aus Eigenvektoren von f besteht. Dann ist $(e_1, \ldots, e_{n-1}, e_n)$ eine ONB von V aus Eigenvektoren von f. ❑

Für Matrizen lässt sich die interessante Richtung dieses Satzes auch so formulieren:

FOLG ∗ Spektralsatz für Matrizen über $\mathbb{R}$

22.14. **Folgerung.** *Sei $A \in \operatorname{Mat}(n, \mathbb{R})$ eine symmetrische Matrix. Dann gibt es eine orthogonale Matrix $P \in \mathrm{O}(n)$, sodass $P^\top AP = P^{-1}AP$ eine Diagonalmatrix ist.*

Das ist Satz 20.14, den wir bereits benutzt haben, um das Determinanten-Kriterium für positive Definitheit (Satz 20.19) zu beweisen.

Beweis. Sei $f\colon \boldsymbol{x} \mapsto A\boldsymbol{x}$, dann ist A die Matrix von $f \in \operatorname{End}(\mathbb{R}^n)$ bezüglich der Standardbasis E; da A symmetrisch ist, ist f selbst-adjungiert (Folgerung 22.9). Nach Satz 22.13 hat $\mathbb{R}^n$ eine Orthonormalbasis B aus Eigenvektoren von f, also ist $D = \operatorname{Mat}_B(f)$ eine Diagonalmatrix. Die Matrix $P = \operatorname{Mat}_{B,E}(\mathrm{id}_{\mathbb{R}^n})$ hat als Spalten die Vektoren von B und ist damit orthogonal (vergleiche die Bemerkung nach Definition 22.6). Außerdem ist

$$P^{-1}AP = \operatorname{Mat}_{E,B}(\mathrm{id}_{\mathbb{R}^n}) \operatorname{Mat}_E(f) \operatorname{Mat}_{B,E}(\mathrm{id}_{\mathbb{R}^n}) = \operatorname{Mat}_B(f) = D\,. \qquad \square$$

Wir wenden uns jetzt der Frage nach der unitären Diagonalisierbarkeit zu. Es wird sich herausstellen, dass genau die normalen Endomorphismen unitär diagonalisierbar sind. Wir beginnen mit einem Lemma.

22.15. **Lemma.** *Sei V ein unitärer Vektorraum und sei $f \in \operatorname{End}(V)$ normal.* **LEMMA** Eigenwerte und -vektoren von f^*

(1) *Es gilt $\|f^*(v)\| = \|f(v)\|$ für alle $v \in V$.*
(2) *Ist $v \in V$ ein Eigenvektor von f zum Eigenwert λ, dann ist v auch ein Eigenvektor von f^* zum Eigenwert $\bar\lambda$.*

Beweis. Die erste Aussage ergibt sich wie folgt:

$$\|f^*(v)\|^2 = \langle f^*(v), f^*(v)\rangle = \langle f(f^*(v)), v\rangle = \langle f^*(f(v)), v\rangle = \langle f(v), f(v)\rangle = \|f(v)\|^2$$

Die zweite Aussage folgt daraus: Zunächst einmal ist mit f auch $f - \lambda\,\mathrm{id}_V$ normal, denn

$$\begin{aligned}(f - \lambda\,\mathrm{id}_V) \circ (f - \lambda\,\mathrm{id}_V)^* &= (f - \lambda\,\mathrm{id}_V) \circ (f^* - \bar\lambda\,\mathrm{id}_V)\\ &= f \circ f^* - \lambda f^* - \bar\lambda f + |\lambda|^2\,\mathrm{id}_V\\ &= f^* \circ f - \lambda f^* - \bar\lambda f + |\lambda|^2\,\mathrm{id}_V\\ &= (f^* - \bar\lambda\,\mathrm{id}_V) \circ (f - \lambda\,\mathrm{id}_V)\\ &= (f - \lambda\,\mathrm{id}_V)^* \circ (f - \lambda\,\mathrm{id}_V)\,.\end{aligned}$$

Aus $f(v) = \lambda v$ und Aussage (1) ergibt sich dann

$$0 = \|(f - \lambda\,\mathrm{id}_V)(v)\| = \|(f - \lambda\,\mathrm{id}_V)^*(v)\| = \|(f^* - \bar\lambda\,\mathrm{id}_V)(v)\| = \|f^*(v) - \bar\lambda v\|\,;$$

damit ist $f^*(v) = \bar\lambda v$. $\square$

∗ 22.16. **Satz.** *Sei f ein Endomorphismus eines endlich-dimensionalen unitären Vektorraums V. f ist genau dann unitär diagonalisierbar, wenn f normal ist.* **SATZ** Spektralsatz über $\mathbb{C}$

Beweis. Sei zunächst f unitär diagonalisierbar. Dann ist die Matrix von f bezüglich einer ONB von V, die aus Eigenvektoren von f besteht, diagonal und damit normal. Es folgt, dass f ebenfalls normal ist.

Die umgekehrte Implikation beweisen wir durch Induktion über die Dimension n des Vektorraums V. Für $n = 0$ (oder $n = 1$) ist nichts zu zeigen. Sei also $n \geq 1$. Weil $\mathbb{C}$ algebraisch abgeschlossen ist, hat das charakteristische Polynom von f

eine Nullstelle, also hat f einen Eigenwert λ mit zugehörigem Eigenvektor v_n. Nach Skalieren können wir annehmen, dass $\|v_n\| = 1$ ist. Nach Lemma 22.15 ist $f^*(v_n) = \bar{\lambda} v_n$. Wir betrachten das orthogonale Komplement von $\langle v_n \rangle_{\mathbb{C}}$:

$$U = \{u \in V \mid \langle u, v_n \rangle = 0\}\,.$$

Dann ist U ein f-invarianter Untervektorraum von V, denn für $u \in U$ gilt

$$\langle f(u), v_n \rangle = \langle u, f^*(v_n) \rangle = \langle u, \bar{\lambda} v_n \rangle = \lambda \langle u, v_n \rangle = 0$$

und damit $f(u) \in U$. Analog sieht man, dass U ein f^*-invarianter Untervektorraum ist. Damit ist $f|_U$ ein normaler Endomorphismus von U (U ist ein unitärer Vektorraum mit dem eingeschränkten Skalarprodukt); außerdem gilt $V = \langle v_n \rangle_{\mathbb{C}} \oplus U$. Nach Induktionsannahme hat U eine ONB $(v_1, \ldots, v_{n-1})$ aus Eigenvektoren von f; dann ist $(v_1, \ldots, v_{n-1}, v_n)$ eine ONB von V aus Eigenvektoren von f. ❑

Für Matrizen lautet die interessante Richtung dieser Aussage wie folgt (der Beweis ist analog zum euklidischen Fall):

FOLG ∗ Spektralsatz für Matrizen über $\mathbb{C}$

22.17. **Folgerung.** *Ist $A \in \mathrm{Mat}(n, \mathbb{C})$ normal, dann gibt es eine unitäre Matrix $P \in \mathrm{U}(n)$, sodass $P^{-1}AP = P^*AP$ eine Diagonalmatrix ist.*

Die Verallgemeinerung des reellen Spektralsatzes auf den unitären Fall charakterisiert die selbst-adjungierten Endomorphismen auch in diesem Fall:

SATZ selbst-adj. über $\mathbb{C}$

22.18. **Satz.** *Ein normaler Endomorphismus eines endlich-dimensionalen unitären Vektorraums ist genau dann selbst-adjungiert, wenn er nur reelle Eigenwerte hat.*

Beweis. Sei V ein endlich-dimensionaler unitärer Vektorraum und sei $f\colon V \to V$ normal. Aus Satz 22.16 folgt, dass f unitär diagonalisierbar ist; sei also B eine ONB von V, die aus Eigenvektoren von f besteht. Dann ist $D = \mathrm{Mat}_B(f)$ eine Diagonalmatrix, deren Diagonaleinträge gerade die Eigenwerte von f sind; außerdem ist f genau dann selbst-adjungiert, wenn D hermitesch ist (Folgerung 22.9). Das bedeutet hier $D = D^* = \bar{D}^\top = \bar{D}$, was damit äquivalent ist, dass die Eigenwerte reell sind. ❑

Das liefert eine Möglichkeit, den reellen Spektralsatz aus der komplexen Version abzuleiten; damit kann man das Kompaktheitsargument im Beweis von Satz 22.13 (d.h., die Aussage von Lemma 22.11) umgehen: Satz 22.18 garantiert, dass alle Eigenwerte eines selbst-adjungierten Endomorphismus eines endlich-dimensionalen euklidischen Vektorraums reell sind (man kann eine zugehörige Matrix als Matrix über $\mathbb{C}$ betrachten). Das liefert die Aussage von Folgerung 22.12, die wir für den Beweis gebraucht haben. Allerdings haben wir für den komplexen Fall verwendet, dass $\mathbb{C}$ algebraisch abgeschlossen ist, was in gewisser Weise eine schwieriger zu zeigende Aussage ist als die Sätze über kompakte Mengen im $\mathbb{R}^n$. (Üblicherweise sehen Sie den ersten Beweis in der „Einführung in die Funktionentheorie“ im vierten Semester.)

23. Quadratische Formen

Aus einer bilinearen Abbildung kann man eine „quadratische“ Abbildung machen: Ist $\beta\colon V \times V \to W$ bilinear, dann hat die Abbildung

$$q\colon V \longrightarrow W, \quad v \longmapsto \beta(v,v)$$

folgende Eigenschaften:

$$\begin{aligned} q(\lambda v) &= \lambda^2 q(v) && \text{für alle } \lambda \in K,\, v \in V, \quad \text{und} \\ q(v+v') + q(v-v') &= 2q(v) + 2q(v') && \text{für alle } v, v' \in V \\ & && \text{(„Parallelogramm-Gleichung“).} \end{aligned}$$

Außerdem ist $q(v+v') - q(v) - q(v') = \beta(v,v') + \beta(v',v)$ eine (symmetrische) bilineare Abbildung. Wir untersuchen den Zusammenhang etwas genauer.

* 23.1. **Definition.** Sei V ein K-Vektorraum. Eine *quadratische Form* auf V ist eine Abbildung $q\colon V \to K$, sodass **DEF** quadratische Form

(1) $q(\lambda v) = \lambda^2 q(v)$ für alle $\lambda \in K$, $v \in V$, und
(2) $(v,w) \mapsto q(v+w) - q(v) - q(w)$ eine Bilinearform ist.

Die Menge aller quadratischen Formen auf V bildet in der üblichen Weise einen Vektorraum $\mathrm{Qu}(V)$.

Zwei quadratische Formen q und q' auf V heißen *äquivalent*, wenn es einen Automorphismus $f\colon V \to V$ gibt, sodass $q' = q \circ f$ ist. ♢

Die Parallelogramm-Gleichung folgt aus den beiden Eigenschaften in der Definition.

Analog zu symmetrischen Bilinearformen definiert man positive Definitheit usw. für quadratische Formen über $\mathbb{R}$.

23.2. **Definition.** Sei q eine quadratische Form auf einem reellen Vektorraum V. **DEF** pos./neg. definit für qu. Formen

(1) q heißt *positiv (negativ) definit*, wenn $q(v) > 0$ ($q(v) < 0$) für alle $\mathbf{0} \neq v \in V$ gilt.
(2) q heißt *positiv (negativ) semidefinit*, wenn $q(v) \geq 0$ ($q(v) \leq 0$) für alle $v \in V$ gilt.
(3) q heißt *indefinit*, wenn q weder positiv noch negativ semidefinit ist. ♢

Für das Folgende ist es wichtig, dass wir durch 2 teilen können. Deshalb noch eine Definition.

23.3. **Definition.** Sei K ein Körper. Ist $n \cdot 1_K \neq 0_K$ für alle $n \in \mathbb{Z}_{>0}$, dann hat K *Charakteristik* 0. Sonst ist die *Charakteristik* von K die kleinste positive ganze Zahl p mit $p \cdot 1_K = 0_K$. Wir schreiben $\mathrm{char}(K)$ für die Charakteristik von K. ♢ **DEF** Charakteristik

Die Charakteristik ist entweder null oder eine Primzahl: Wäre $\mathrm{char}(K) = n$ keine Primzahl, dann könnten wir schreiben $n = km$ mit $1 \leq k, m < n$. Aus $n \cdot 1_K = 0_K$ folgt $(k \cdot 1_K) \cdot (m \cdot 1_K) = 0_K$, also $k \cdot 1_K = 0_K$ oder $m \cdot 1_K = 0_K$, was ein Widerspruch dazu ist, dass n die kleinste solche Zahl ist.

Wenn K nicht Charakteristik 2 hat, dann ist $2 \neq 0$ in K und damit invertierbar. Ein Körper der Charakteristik 2 ist zum Beispiel $\mathbb{F}_2$; dort gilt ja $1 + 1 = 0$.

LEMMA
symm. bil. = quadr.

23.4. **Lemma.** *Sei V ein K-Vektorraum mit* $\operatorname{char}(K) \neq 2$. *Wir schreiben* $\operatorname{Sym}(V)$ *für den Vektorraum der symmetrischen Bilinearformen auf V. Dann ist*

$$\Phi\colon \operatorname{Sym}(V) \longrightarrow \operatorname{Qu}(V), \quad \beta \longmapsto \bigl(v \mapsto \beta(v,v)\bigr)$$

ein Isomorphismus.

Man kann also quadratische Formen mit den zugehörigen symmetrischen Bilinearformen identifizieren; insbesondere sind auch quadratische Formen durch symmetrische Matrizen beschrieben. Wir schreiben $\operatorname{Mat}_B(q)$ für diese Matrix; es gilt für $v = x_1 b_1 + \ldots + x_n b_n$ (wenn $B = (b_1, \ldots, b_n)$ ist)

$$q(v) = \boldsymbol{x}^\top \operatorname{Mat}_B(q) \boldsymbol{x}\,,$$

wobei $\boldsymbol{x}$ der Spaltenvektor $(x_1, \ldots, x_n)^\top$ ist.

Beweis. Dass die angegebene Abbildung wohldefiniert und linear ist, ist klar. Wir zeigen, dass sie bijektiv ist, indem wir die Umkehrabbildung angeben:

$$\Psi\colon \operatorname{Qu}(V) \longrightarrow \operatorname{Sym}(V), \quad q \longmapsto \Bigl((v,w) \mapsto \tfrac{1}{2}\bigl(q(v+w) - q(v) - q(w)\bigr)\Bigr)$$

(hier verwenden wir $\operatorname{char}(K) \neq 2$). Nach Definition 23.1 ist das Bild eine (symmetrische) Bilinearform, also ist die Abbildung wohldefiniert. Wir prüfen nach, dass das tatsächlich die Inverse ist: $(\Phi \circ \Psi)(q)$ ist die Abbildung, die v abbildet auf

$$\bigl(\Psi(q)\bigr)(v,v) = \tfrac{1}{2}\bigl(q(v+v) - q(v) - q(v)\bigr) = \tfrac{1}{2}\bigl(q(2v) - 2q(v)\bigr) = q(v)\,,$$

also ist $(\Phi \circ \Psi)(q) = q$, und $\bigl(\Psi \circ \Phi\bigr)(\beta)$ ist die Abbildung, die (v,w) abbildet auf

$$\begin{aligned}
\tfrac{1}{2}\Bigl(\bigl(\Phi(\beta)\bigr)(v+w) - \bigl(\Phi(\beta)\bigr)(v) - \bigl(\Phi(\beta)\bigr)(w)\Bigr)& \\
&= \tfrac{1}{2}\bigl(\beta(v+w, v+w) - \beta(v,v) - \beta(w,w)\bigr) \\
&= \tfrac{1}{2}\bigl(\beta(v,w) + \beta(w,v)\bigr) = \beta(v,w)
\end{aligned}$$

(denn β ist symmetrisch), also ist $(\Psi \circ \Phi)(\beta) = \beta$. ❑

Über Körpern der Charakteristik 2 ist die Aussage des Lemmas falsch. Zum Beispiel hat die quadratische Form $q(x,y) = xy$ auf $\mathbb{F}_2^2$ nicht die Form $q(\boldsymbol{x}) = \beta(\boldsymbol{x}, \boldsymbol{x})$ für eine symmetrische Bilinearform β.

Daraus folgt unmittelbar:

FOLG
qu. Formen auf K^n

23.5. **Folgerung.** *Ist K ein Körper mit* $\operatorname{char}(K) \neq 2$, *dann sind die quadratischen Formen auf K^n alle gegeben durch*

$$(x_1, x_2, \ldots, x_n)^\top \longmapsto \sum_{1 \le i \le j \le n} a_{ij} x_i x_j$$

mit $a_{ij} \in K$. Die zugehörige Matrix hat Diagonaleinträge a_{ii} und Einträge $a_{ij}/2$ an den Positionen (i,j) und (j,i), wenn $i < j$ ist.

Zwei quadratische Formen auf K^n sind genau dann äquivalent, wenn die zugehörigen symmetrischen Matrizen kongruent sind.

Die erste Aussage bleibt auch für Körper der Charakteristik 2 richtig; die Aussagen über die Matrizen haben in diesem Fall keinen Sinn.

Der Wert einer quadratischen Form q auf K^n bei einem Vektor $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ wird häufig zur Vereinfachung $q(x_1, \ldots, x_n)$ geschrieben statt $q\bigl((x_1, \ldots, x_n)^\top\bigr)$.

Äquivalenz von zwei quadratischen Formen q und q' auf K^n bedeutet dann ganz konkret, dass

$$\begin{aligned} q'(x_1, x_2, \ldots, x_n) = q(&a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n, \\ &a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n, \\ &\ldots, \\ &a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n) \end{aligned}$$

gilt mit einer Matrix $A = (a_{ij}) \in \mathrm{GL}(n, K)$.

23.6. **Definition.** Eine quadratische Form q auf K^n heißt *diagonal* oder eine *Diagonalform*, wenn sie die Form

$$q(x_1, x_2, \ldots, x_n) = a_1 x_1^2 + a_2 x_2^2 + \ldots + a_n x_n^2$$

hat mit geeigneten $a_1, a_2, \ldots, a_n \in K$. ♢

DEF diagonale qu. Form

Es treten also keine „gemischten Terme“ $x_i x_j$ (mit $i \neq j$) auf, und die zugehörige Matrix ist eine Diagonalmatrix.

Seien A und A' die symmetrischen Matrizen zweier äquivalenter quadratischer Formen q und q' auf K^n (mit $\mathrm{char}(K) \neq 2$), oder auch die Matrizen einer quadratischen Form auf V bezüglich zweier unterschiedlicher Basen. Dann gibt es $P \in \mathrm{GL}(n, K)$ mit $A' = P^\top AP$; insbesondere ist $\mathrm{rk}(A') = \mathrm{rk}(A)$ (denn Multiplikation mit einer invertierbaren Matrix ändert den Rang nicht). Das zeigt, dass folgende Definition sinnvoll ist.

23.7. **Definition.** Sei K ein Körper mit $\mathrm{char}(K) \neq 2$ und sei V ein endlich-dimensionaler K-Vektorraum. Sei weiter $q \in \mathrm{Qu}(V)$. Ist B eine Basis von V, dann heißt $\mathrm{rk}\big(\mathrm{Mat}_B(q)\big)$ der *Rang* von q. ♢

DEF Rang einer qu. Form

Wir wollen jetzt quadratische Formen auf endlich-dimensionalen komplexen und reellen Vektorräumen klassifizieren. Das ist dazu äquivalent, symmetrische Matrizen bis auf Kongruenz zu klassifizieren. Wir beginnen mit einem Resultat, das für (fast) beliebige Körper gilt.

* 23.8. **Satz.** *Sei K ein Körper mit $\mathrm{char}(K) \neq 2$. Dann ist jede quadratische Form q auf K^n äquivalent zu einer Diagonalform.*

SATZ Diagonalisierung von qu. Formen

Dazu äquivalent sind folgende Aussagen:

- Seien V ein endlich-dimensionaler K-Vektorraum und q eine quadratische Form auf V. Dann hat V eine Basis B, sodass $\mathrm{Mat}_B(q)$ eine Diagonalmatrix ist.
- Jede symmetrische Matrix $A \in \mathrm{Mat}(n, K)$ ist kongruent zu einer Diagonalmatrix.

Beweis. Der Beweis geht durch Induktion über n. Im Fall $n = 1$ (oder $n = 0$) ist nichts zu zeigen. Wir nehmen jetzt an, dass $n > 1$ ist und die Aussage für $n - 1$ gilt. Wir schreiben

$$q(x_1, x_2, \ldots, x_n) = q'(x_1, x_2, \ldots, x_{n-1}) + b_1 x_1 x_n + \ldots + b_{n-1} x_{n-1} x_n + a_n x_n^2$$

mit $b_1, \ldots, b_{n-1}, a_n \in K$ und einer quadratischen Form q' auf K^{n-1}. Ist $a_n \neq 0$, dann ist („quadratische Ergänzung“)

$$q\big(x_1, \ldots, x_{n-1}, x_n - \tfrac{1}{2a_n}(b_1 x_1 + \ldots + b_{n-1} x_{n-1})\big) = q''(x_1, x_2, \ldots, x_{n-1}) + a_n x_n^2$$

mit

$$q''(x_1, \ldots, x_{n-1}) = q'(x_1, \ldots, x_{n-1}) - \frac{1}{4a_n}(b_1 x_1 + \ldots + b_{n-1}x_{n-1})^2\,;$$

das ist eine quadratische Form auf K^{n-1}. Nach Induktionsannahme ist q'' äquivalent zu einer Diagonalform $a_1 x_1^2 + \ldots + a_{n-1}x_{n-1}^2$; damit ist q äquivalent zu $a_1 x_1^2 + \ldots + a_{n-1}x_{n-1}^2 + a_n x_n^2$.

Es bleibt der Fall $a_n = 0$ zu behandeln. Gilt $b_1 = \ldots = b_{n-1} = 0$, dann können wir die Induktionsannahme direkt auf q' anwenden und sind fertig. Sei also jetzt $b_m \neq 0$ für ein $m \in \{1, 2, \ldots, n-1\}$. Dann ist $q(x_1, \ldots, x_n) = \alpha x_m^2 + b_m x_m x_n + R$, wobei jeder Term, der in R vorkommt, eine Variable x_j mit $j \notin \{m, n\}$ enthält. Wir ersetzen x_m durch $x_m \pm x_n$ und erhalten

$$q(x_1, \ldots, x_m \pm x_n, \ldots, x_n) = \alpha x_m^2 + (b_m \pm 2\alpha)x_m x_n + (\alpha \pm b_m)x_n^2 + R'\,.$$

Da $b_m \neq 0$ ist, muss für wenigstens eine Wahl des Vorzeichens $\alpha \pm b_m \neq 0$ sein (hier benutzen wir wieder $\operatorname{char}(K) \neq 2$, also $1_K \neq -1_K$). Wir sehen, dass q zu einer Form mit $a_n \neq 0$ äquivalent ist; diesen Fall haben wir bereits behandelt. ❑

Auch diese Aussage ist i.A. falsch für $\operatorname{char}(K) = 2$: Die quadratische Form $q(x, y) = xy$ auf $\mathbb{F}_2^2$ ist nicht äquivalent zu einer Diagonalform (Übung).

Die Koeffizienten der Diagonalform sind keineswegs eindeutig bestimmt. Wir können die Reihenfolge beliebig ändern durch Permutation der Variablen. Durch Skalieren der Koordinaten können wir außerdem die Koeffizienten mit beliebigen Quadraten $\neq 0$ multiplizieren. Aber es gilt zum Beispiel auch, dass

$$2x_1^2 + 2x_2^2 \qquad \text{und} \qquad x_1^2 + x_2^2$$

über $\mathbb{Q}$ äquivalent sind, obwohl 2 kein Quadrat in $\mathbb{Q}$ ist:

$$(x_1 + x_2)^2 + (x_1 - x_2)^2 = 2x_1^2 + 2x_2^2\,.$$

BSP Diagonalisierung

23.9. **Beispiel.** Wie sieht eine zu $q(x_1, x_2, x_3) = x_1x_2 + x_1x_3 + x_2x_3$ äquivalente Diagonalform über $\mathbb{Q}$ aus? Da kein Term x_j^2 auftritt, müssen wir zunächst einen erzeugen:

$$q'(x_1, x_2, x_3) = q(x_1 + x_3, x_2, x_3) = x_1x_2 + x_1x_3 + 2x_2x_3 + x_3^2\,.$$

Jetzt können wir die quadratische Ergänzung durchführen:

$$q''(x_1, x_2, x_3) = q'\bigl(x_1, x_2, x_3 - \tfrac{1}{2}x_1 - x_2\bigr) = -\tfrac{1}{4}x_1^2 - x_2^2 + x_3^2\,.$$

Wir können x_1 noch mit 2 skalieren und erhalten die etwas hübschere Form

$$q'''(x_1, x_2, x_3) = q''(2x_1, x_2, x_3) = -x_1^2 - x_2^2 + x_3^2\,.$$ ♣

Im Körper $\mathbb{C}$ der komplexen Zahlen hat (nach Satz 4.3) jedes Element eine Quadratwurzel. Da wir die Diagonaleinträge mit beliebigen Quadraten multiplizieren können, erhalten wir den folgenden Klassifikationssatz.

SATZ ∗ Klassifikation qu. Formen über $\mathbb{C}$

23.10. Satz. *Jede quadratische Form $q \in \mathrm{Qu}(\mathbb{C}^n)$ ist äquivalent zu einer Form*

$$Q_r(x_1, \ldots, x_n) = x_1^2 + \ldots + x_r^2 .$$

Die Zahl $r \in \{0, 1, \ldots, n\}$ ist dabei eindeutig bestimmt.

Beweis. Nach Satz 23.8 ist q äquivalent zu einer Diagonalform q'. Wir können annehmen (nach eventueller Permutation der Variablen), dass in q' genau die Terme $x_1^2, \ldots, x_r^2$ vorkommen. Durch Skalieren können wir erreichen, dass die Koeffizienten $= 1$ sind; damit haben wir die gewünschte Form. Als Rang von q ist r eindeutig bestimmt. ❑

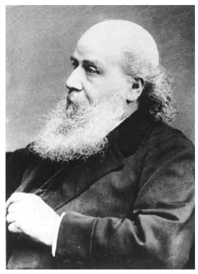
In $\mathbb{R}$ gilt nur noch, dass jede positive Zahl (und die Null) ein Quadrat ist. Das führt zum folgenden *Sylvesterschen Trägheitssatz* oder *Signatursatz*:

∗ **23.11. Satz.** *Jede quadratische Form $q \in \mathrm{Qu}(\mathbb{R}^n)$ ist äquivalent zu einer Form*

SATZ Klassifikation qu. Formen über $\mathbb{R}$

$$Q_{r,s}(x_1, \ldots, x_n) = x_1^2 + \ldots + x_r^2 - x_{r+1}^2 - \ldots - x_{r+s}^2 .$$

Die Zahlen $r, s \geq 0$ mit $r + s \leq n$ sind eindeutig bestimmt.



J.J. Sylvester 1814–1897

Beweis. Nach Satz 23.8 ist q äquivalent zu einer Diagonalform q'. Nach Permutation der Variablen können wir annehmen, dass $q'(x_1, \ldots, x_n) = \sum_{j=1}^{r+s} a_j x_j^2$ ist mit $a_1, \ldots, a_r > 0$ und $a_{r+1}, \ldots, a_{r+s} < 0$. Durch Skalieren können wir die positiven Koeffizienten durch 1 und die negativen Koeffizienten durch -1 ersetzen; damit haben wir die gewünschte Form.

Wie eben ist der Rang $r+s$ eindeutig durch q bestimmt. Die Zahl r ist die maximale Dimension eines Untervektorraums, auf dem q positiv definit ist und ist damit ebenfalls eindeutig bestimmt: Diese Dimension ist für äquivalente quadratische Formen offensichtlich gleich, also müssen wir diese Aussage nur für $Q_{r,s}$ zeigen. $Q_{r,s}$ ist auf dem r-dimensionalen Untervektorraum $\langle \mathbf{e}_1, \ldots, \mathbf{e}_r \rangle$ positiv definit. Ist $U \subset \mathbb{R}^n$ mit $\dim U > r$, dann gilt mit $U' = \langle \mathbf{e}_{r+1}, \ldots, \mathbf{e}_n \rangle$, dass

$$\dim U \cap U' = \dim U + \dim U' - \dim(U + U') \geq \dim U + (n - r) - n = \dim U - r > 0$$

ist, also gibt es einen Vektor $\mathbf{0} \neq v \in U \cap U'$. Es gilt dann aber $Q_{r,s}(v) \leq 0$, also ist $Q_{r,s}$ auf U nicht positiv definit. Damit ist r die maximale Dimension eines Untervektorraums, auf dem $Q_{r,s}$ positiv definit ist. ❑

23.12. Definition. In der Situation von Satz 23.11 heißt $r - s$ die *Signatur* von q. ♢

DEF Signatur

23.13. Beispiel. Die Signatur der quadratischen Form $x_1x_2 + x_1x_3 + x_2x_3$ aus Beispiel 23.9 ist -1, denn wir haben $r = 1$ und $s = 2$. ♣

BSP Signatur

23.14. **Folgerung.** *Sei $q \in \mathrm{Qu}(\mathbb{R}^n)$ mit r und s wie in Satz 23.11. Dann gilt:*

FOLG
Definitheit über r, s

(1) *q positiv definit* $\iff$ $r = n$.

(2) *q negativ definit* $\iff$ $s = n$.

(3) *q positiv semidefinit* $\iff$ $s = 0$.

(4) *q negativ semidefinit* $\iff$ $r = 0$.

(5) *q indefinit* $\iff$ $r, s > 0$.

Beweis. Es ist klar, dass diese Äquivalenzen für die zu q äquivalente Diagonalform $Q_{r,s}$ gelten. Weil die Definitheitseigenschaften unter Äquivalenz invariant sind, gelten sie dann auch für q. ❑

24. Klassifikation von Quadriken

Wir arbeiten in diesem Abschnitt im Standardraum $\mathbb{R}^n$.

Eine lineare Gleichung (mit $b_1, b_2, \ldots, b_n, c \in \mathbb{R}$ gegeben, nicht alle $b_j = 0$, und $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ gesucht)

$$b_1 x_1 + b_2 x_2 + \ldots + b_n x_n = c$$

hat als Lösungsmenge eine *affine Hyperebene* (also einen affinen Unterraum der Dimension $n-1$). Viel mehr gibt es dazu nicht zu sagen. Daher befassen wir uns jetzt mit *quadratischen* Gleichungen. Sie haben die allgemeine Form

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j + \sum_{j=1}^{n} b_j x_j = c$$

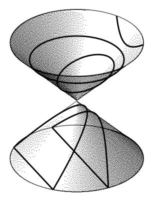
oder kurz

$$\langle \boldsymbol{x}, A\boldsymbol{x} \rangle + \langle \boldsymbol{b}, \boldsymbol{x} \rangle = c \qquad \text{bzw.} \qquad \boldsymbol{x}^\top A \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} = c\,;$$

dabei ist $\mathbf{0} \neq A = (a_{ij}) \in \mathrm{Mat}(n, \mathbb{R})$ eine symmetrische Matrix, $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$ ein Vektor und $c \in \mathbb{R}$. (Ist A die Nullmatrix, dann ist die Gleichung nicht wirklich quadratisch.)

* 24.1. **Definition.** Die Lösungsmenge einer quadratischen Gleichung wie oben heißt *Quadrik* im $\mathbb{R}^n$. Eine Quadrik im $\mathbb{R}^2$ heißt auch *Kegelschnitt*. $\diamondsuit$ **DEF** Quadrik Kegelschnitt

Die Bezeichnung „Kegelschnitt“ für Quadriken im $\mathbb{R}^2$ kommt daher, dass sich (fast) alle solchen Quadriken als Schnitt des Doppelkegels



$$x_1^2 + x_2^2 - x_3^2 = 0$$

im $\mathbb{R}^3$ mit einer Ebene realisieren lassen.

Man kann analog Quadriken auch über anderen Körpern (zum Beispiel $\mathbb{C}$) definieren und studieren.

24.2. **Beispiel.** Ein einfaches Beispiel für eine Quadrik im $\mathbb{R}^2$, also für einen Kegelschnitt, ist der *Einheitskreis*, der die Lösungsmenge der quadratischen Gleichung **BSP** Kreis

$$x_1^2 + x_2^2 = 1$$

ist. (Hier ist $A = I_2$ die Einheitsmatrix, $\boldsymbol{b} = \mathbf{0}$ und $c = 1$.) Allgemeiner ist ein Kreis mit Mittelpunkt (m_1, m_2) und Radius r ebenfalls ein Kegelschnitt; hier lautet die Gleichung

$$(x_1 - m_1)^2 + (x_2 - m_2)^2 = r^2\,,$$

was zu

$$x_1^2 + x_2^2 - 2m_1 x_1 - 2m_2 x_2 = r^2 - m_1^2 - m_2^2$$

äquivalent ist (also $A = I_2$, $\boldsymbol{b} = -2(m_1, m_2)^\top$ und $c = r^2 - m_1^2 - m_2^2$). $\clubsuit$

Eine Drehung, Spiegelung (allgemeiner eine Isometrie) oder Verschiebung ändert die geometrische Form einer Quadrik nicht. Deshalb sind wir an einer Normalform bzw. Klassifikation bis auf solche Geometrie erhaltenden Abbildungen interessiert. Wir geben diesen Abbildungen zuerst einen Namen. Vorher erinnern wir uns daran, dass die Determinante einer orthogonalen Matrix stets ± 1 ist:

$$A^\top A = I_n \implies 1 = \det(I_n) = \det(A^\top)\det(A) = \det(A)^2\,.$$

Die zugehörigen Isometrien $\mathbb{R}^n \to \mathbb{R}^n$ sind also orientierungserhaltend, wenn $\det(A) = 1$ ist, und orientierungsumkehrend, wenn $\det(A) = -1$ ist (vergleiche Definition 14.22). Die orthogonalen Matrizen mit Determinante 1 bilden eine

Untergruppe von $\mathrm{O}(n)$, die spezielle orthogonale Gruppe $\mathrm{SO}(n)$, siehe Definition 22.6.

DEF
Bewegung
Translation
Bewegungsgruppe

24.3. **Definition.** Sei V ein endlich-dimensionaler euklidischer Vektorraum. Eine Abbildung $f\colon V \to V$ heißt eine *(euklidische) Bewegung* von V, wenn es eine Isometrie $h\colon V \to V$ und einen Vektor $v_0 \in V$ gibt mit $f(v) = h(v) + v_0$ für alle $v \in V$. Im Fall $h = \mathrm{id}_V$ heißt f auch *Translation* um v_0.

Die Bewegung ist *orientierungserhaltend* bzw. *-umkehrend*, wenn h orientierungserhaltend (also $\det(h) > 0$) bzw. -umkehrend ($\det(h) < 0$) ist.

Die Menge aller Bewegungen von V bildet eine Gruppe (Übung), die *Bewegungsgruppe* von V. ♢

BSP
Klassifikation affiner Hyperebenen

24.4. **Beispiel.** Ist $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{b}^\top \boldsymbol{x} = c\}$ eine affine Hyperebene, dann gibt es eine Bewegung T mit $T(H) = \{\boldsymbol{x} \in \mathbb{R}^n \mid x_1 = 0\}$. Dafür ergänzen wir $\boldsymbol{b}$ zu einer Basis von $\mathbb{R}^n$ und wenden das Gram-Schmidt-Verfahren an. Wir erhalten eine ONB, deren erster Vektor ein skalares Vielfaches $\lambda\boldsymbol{b}$ ist. Sei P die zugehörige orthogonale Matrix (deren Spalten diese ONB bilden). Wir identifizieren P mit der Isometrie $\boldsymbol{x} \mapsto P\boldsymbol{x}$. Dann ist

$$\begin{aligned} P^{-1}(H) &= \{P^{-1}\boldsymbol{x} \mid \boldsymbol{x} \in H\} = \{\boldsymbol{x} \mid P\boldsymbol{x} \in H\} = \{\boldsymbol{x} \mid \boldsymbol{b}^\top P\boldsymbol{x} = c\} \\ &= \{\boldsymbol{x} \mid \boldsymbol{b}^\top (P^{-1})^\top \boldsymbol{x} = c\} = \{\boldsymbol{x} \mid (P^{-1}\boldsymbol{b})^\top \boldsymbol{x} = c\} = \{\boldsymbol{x} \mid x_1 = \lambda c\}\,, \end{aligned}$$

denn $P\mathbf{e}_1 = \lambda\boldsymbol{b}$, also ist $P^{-1}\boldsymbol{b} = \lambda^{-1}\mathbf{e}_1$. Translation um den Vektor $(\lambda c, 0, \ldots, 0)^\top$ ergibt schließlich die Hyperebene $x_1 = 0$.

Man kann die Aussage des Beispiels so interpretieren, dass alle affinen Hyperebenen „gleich aussehen“. Quadriken sind da komplizierter, wie das nächste Ergebnis zeigt. ♣

Das Hauptergebnis in diesem Abschnitt lautet wie folgt.

SATZ ∗
euklidische Normalform von Quadriken
Hauptachsentransformation

24.5. **Satz.** *Sei $Q = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, A\boldsymbol{x}\rangle + \langle \boldsymbol{b}, \boldsymbol{x}\rangle = c\}$ eine Quadrik im $\mathbb{R}^n$ (mit $n \ge 1$) und sei r der Rang von A. Dann gibt es eine orientierungserhaltende Bewegung T des $\mathbb{R}^n$, reelle Zahlen $a_1, \ldots, a_r > 0$ und Vorzeichen $\varepsilon_1, \ldots, \varepsilon_r \in \{\pm 1\}$ mit*

$$\begin{aligned} T(Q) &= \Big\{\boldsymbol{x} \in \mathbb{R}^n \;\Big|\; \varepsilon_1\Big(\frac{x_1}{a_1}\Big)^2 + \ldots + \varepsilon_r\Big(\frac{x_r}{a_r}\Big)^2 = 0\Big\} \qquad \textit{oder} \\ T(Q) &= \Big\{\boldsymbol{x} \in \mathbb{R}^n \;\Big|\; \varepsilon_1\Big(\frac{x_1}{a_1}\Big)^2 + \ldots + \varepsilon_r\Big(\frac{x_r}{a_r}\Big)^2 = 1\Big\} \qquad \textit{oder} \\ T(Q) &= \Big\{\boldsymbol{x} \in \mathbb{R}^n \;\Big|\; \varepsilon_1\Big(\frac{x_1}{a_1}\Big)^2 + \ldots + \varepsilon_r\Big(\frac{x_r}{a_r}\Big)^2 = x_{r+1}\Big\}\,. \end{aligned}$$

Im zweiten Fall sind die (a_j, ε_j) bis auf Permutation eindeutig bestimmt, im dritten Fall ist zusätzlich eine Umkehr aller Vorzeichen möglich, im ersten Fall ist auch noch eine zusätzliche gemeinsame Skalierung der a_j zugelassen.

Beweis. Nach dem Spektralsatz 22.14 gibt es $P \in \mathrm{O}(n)$ mit $P^\top AP = D$ diagonal. Dabei können wir annehmen, dass die $n - r$ Nullen auf der Diagonalen von D am Ende kommen. Falls $\det(P) = -1$ ist, können wir P ersetzen durch $P' = P \cdot \mathrm{diag}(-1, 1, \ldots, 1)$; dann gilt auch

$$P'^\top AP' = \mathrm{diag}(-1, 1, \ldots, 1) D \,\mathrm{diag}(-1, 1, \ldots, 1) = D$$

und $\det(P') = 1$. Wir können also $P \in \mathrm{SO}(n)$ annehmen. Wenn wir $P\boldsymbol{x}$ in die Gleichung von Q einsetzen, erhalten wir

$$c = \boldsymbol{x}^\top P^\top A P \boldsymbol{x} + \boldsymbol{b}^\top P \boldsymbol{x} = \boldsymbol{x}^\top D \boldsymbol{x} + (\boldsymbol{b}^\top P)\boldsymbol{x}\,.$$

Ausgeschrieben lautet das

$$\lambda_1 x_1^2 + \ldots + \lambda_r x_r^2 + b_1' x_1 + \ldots + b_n' x_n = c\,.$$

(Dabei sind $\lambda_1, \ldots, \lambda_r$ die von null verschiedenen Eigenwerte von A.) Die neue Quadrik ist das Bild der ursprünglichen unter der orientierungserhaltenden Isometrie P^{-1}.

Hier können wir in den ersten r Variablen quadratisch ergänzen: Wir ersetzen x_j durch $x_j - \frac{b_j'}{2\lambda_j}$ und erhalten die neue Gleichung

$$\lambda_1 x_1^2 + \ldots + \lambda_r x_r^2 + b_{r+1}' x_{r+1} + \ldots + b_n' x_n = c'$$

mit

$$c' = c + \frac{1}{4}\Big(\frac{(b_1')^2}{\lambda_1} + \ldots + \frac{(b_r')^2}{\lambda_r}\Big)\,.$$

Diese Quadrik entsteht aus der vorigen durch eine Translation. Jetzt gibt es drei mögliche Fälle:

- $b_{r+1}' = \ldots = b_n' = c' = 0$. Dann hat die Gleichung die gewünschte (erste) Form mit $\varepsilon_j = \operatorname{sign}\lambda_j$ und $a_j = 1/\sqrt{|\lambda_j|}$ für $j \in \{1, 2, \ldots, r\}$.
- $b_{r+1}' = \ldots = b_n' = 0$ und $c' \neq 0$. Dann teilen wir die Gleichung durch c'; mit $\varepsilon_j = \operatorname{sign}(\lambda_j/c')$ und $a_j = \sqrt{|c'/\lambda_j|}$ für $j \in \{1, 2, \ldots, r\}$ haben wir die zweite Form.
- $b_j' \neq 0$ für ein $j \in \{r+1, \ldots, n\}$. Wie in Beispiel 24.4 kann der „lineare Teil" der Gleichung auf die Form $-\mu x_{r+1} = 0$ gebracht werden mit $\mu \in \mathbb{R}^\times$. Analog zum zweiten Fall setzen wir $\varepsilon_j = \operatorname{sign}(\lambda_j/\mu)$ und $a_j = \sqrt{|\mu/\lambda_j|}$, um die gewünschte (dritte) Form zu erhalten.

Die Eindeutigkeit ergibt sich daraus, dass die Koeffizienten der x_j^2 zueinander im Verhältnis der Eigenwerte $\neq 0$ von A stehen müssen; im zweiten Fall wird die Skalierung dadurch fixiert, dass die Konstante auf der rechten Seite 1 ist; im dritten Fall wird die Skalierung bis auf ein Vorzeichen durch den Koeffizienten von x_{r+1} festgelegt (denn das Vorzeichen von x_{r+1} in der Gleichung kann durch die orientierungserhaltende Isometrie, die nur die Vorzeichen von x_1 und x_{r+1} ändert, umgedreht werden). ❑

Für die Bestimmung der Form und der Parameter der euklidischen Normalform ist es unerheblich, ob die Bewegung T, die zu ihr führt, orientierungserhaltend oder -umkehrend ist. (Denn die Normalform ändert sich nicht, wenn man die orientierungsumkehrende Isometrie $(x_1, \ldots, x_n) \mapsto (-x_1, x_2, \ldots, x_n)$ anwendet.)

24.6. **Definition.** Die Form von $T(Q)$ in Satz 24.5 heißt die *euklidische Normalform* von Q. **DEF** eukl. Normalform

In der Situation des zweiten Falls von Satz 24.5 mit $r = n$ heißen die Zahlen $a_1, a_2, \ldots, a_n$ die *Halbachsen* von Q. Die Geraden $T^{-1}(\langle \mathbf{e}_j \rangle_{\mathbb{R}})$ heißen die *Hauptachsen* von Q, der Punkt $T^{-1}(\mathbf{0})$ der *Mittelpunkt* oder das *Zentrum* von Q. ♢ Halbachsen Hauptachsen Mittelpunkt

Das erklärt auch die Bezeichnung *Hauptachsentransformation* für die Bewegung, die eine Quadrik in Normalform bringt. Hauptachsen-transformation

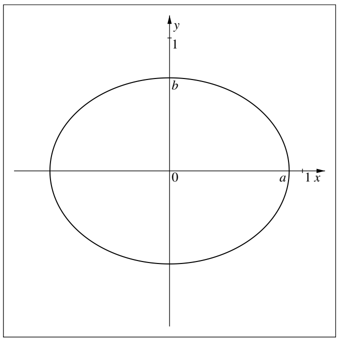
Je nach Verteilung der Vorzeichen erhalten wir verschiedene Typen von Quadriken. Für Kegelschnitte mit $\mathrm{rk}(A) = 2$ gibt es drei Möglichkeiten:

Vorz.		rechte Seite	
ε_1	ε_2	0	1
$+$	$+$	Punkt	Ellipse
$+$	$-$	Geradenpaar	Hyperbel
$-$	$-$	Punkt	leer

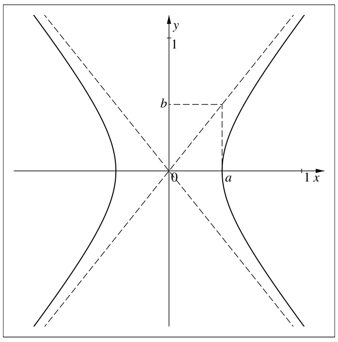
Kegelschnitte vom Rang 1 sind

Vorz.	rechte Seite		
ε_1	0	1	x_2
$+$	Doppelgerade	parallele Geraden	Parabel
$-$	Doppelgerade	leer	Parabel

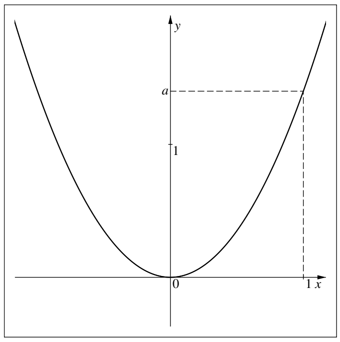
Einige Quadriken im $\mathbb{R}^2$



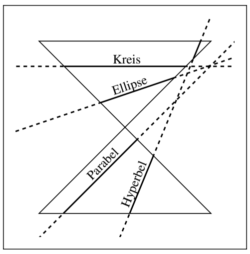
Ellipse:
$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

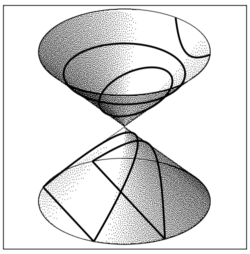


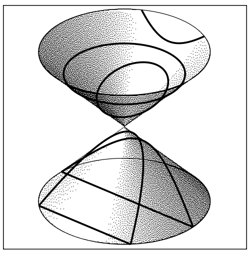
Hyperbel:
$$\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 1$$

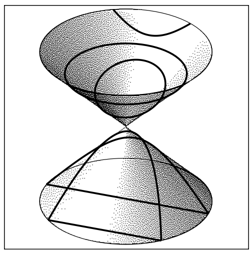


Parabel:
$$ax^2 = y$$









BSP Hauptachsentransformation

24.7. **Beispiel.** Als einfaches Beispiel bestimmen wir die euklidische Normalform des Kegelschnitts (wir schreiben x, y statt x_1, x_2 für die Koordinaten)

$$5x^2 + 4xy + 2y^2 = 1\,.$$

Die zugehörige Matrix ist

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\,;$$

der Vektor $\boldsymbol{b}$ ist der Nullvektor und $c = 1$. Das charakteristische Polynom von A ist

$$\chi_A = \begin{vmatrix} X-5 & -2 \\ -2 & X-2 \end{vmatrix} = (X-5)(X-2) - 4 = X^2 - 7X + 6 = (X-1)(X-6)\,,$$

also sind die Eigenwerte 1 und 6; die zugehörigen Eigenvektoren berechnet man als

$$\begin{pmatrix} 1 \\ -2 \end{pmatrix} \qquad \text{und} \qquad \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Wie es sein muss, sind diese Vektoren zueinander orthogonal. Um eine ONB zu erhalten, müssen wir noch skalieren; das liefert die Transformationsmatrix

$$P = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \in \mathrm{SO}(2)$$

mit

$$P^{-1}AP = P^{\top}AP = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

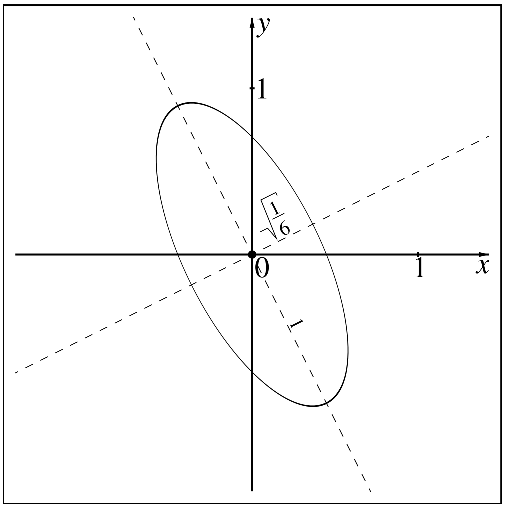
Die transformierte Gleichung lautet also

$$x^2 + 6y^2 = 1$$

oder

$$\left(\frac{x}{1}\right)^2 + \left(\frac{y}{1/\sqrt{6}}\right)^2 = 1\,;$$

das ist eine Ellipse mit Halbachsen 1 und $1/\sqrt{6}$.



♣

Bei Quadriken im $\mathbb{R}^3$ vom Rang 3 sind es vier Möglichkeiten für die Vorzeichen:

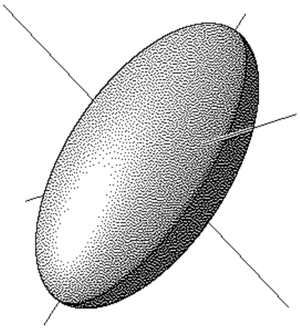
Vorz.			rechte Seite	
ε_1	ε_2	ε_3	0	1
+	+	+	Punkt	Ellipsoid
+	+	−	Doppelkegel	einschaliges Hyperboloid
+	−	−	Doppelkegel	zweischaliges Hyperboloid
−	−	−	Punkt	leer

Man sieht, dass der Typ in der Form $\boldsymbol{x}^{\top}A\boldsymbol{x} = 1$ durch die Vorzeichen der Eigenwerte λ_j von A bestimmt ist; die Halbachsen sind durch $1/\sqrt{|\lambda_j|}$ gegeben.

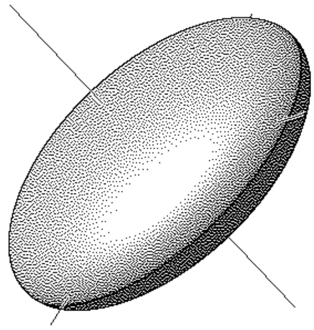
Quadriken im $\mathbb{R}^3$ vom Rang 2 sind

Vorz.		rechte Seite		
ε_1	ε_2	0	1	x_3
+	+	Doppelgerade	ellipt. Zylinder	ellipt. Paraboloid
+	−	Ebenenpaar	hyperbol. Zylinder	hyperbol. Paraboloid
−	−	Doppelgerade	leer	ellipt. Paraboloid

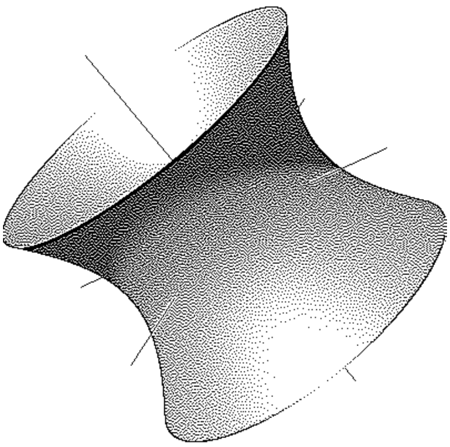
Einige Quadriken im $\mathbb{R}^3$



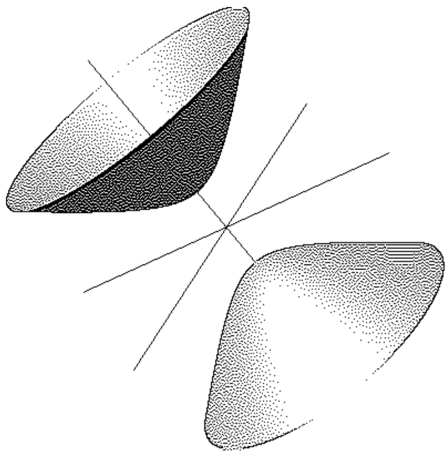
Prolates Rotationsellipsoid:
$5\,x^2 + y^2 + 5\,z^2 = 70$



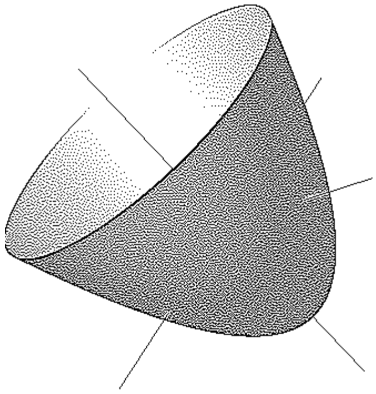
Oblates Rotationsellipsoid:
$x^2 + y^2 + 5\,z^2 = 70$



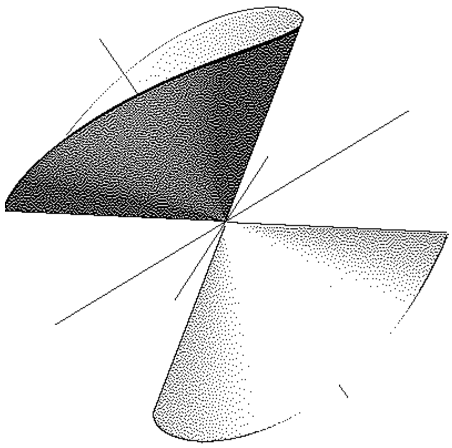
Einschaliges Rotationshyperboloid:
$x^2 + y^2 - z^2 = 30$



Zweischaliges Rotationshyperboloid:
$-x^2 - y^2 + z^2 = 5$



Rotationsparaboloid:
$x^2 + y^2 + 5\,z = 30$



Kegel:
$x^2 + 3\,y^2 - 2\,z^2 = 0$

und vom Rang 1:

Vorz. ε_1	rechte Seite 0	1	x_2
$+$	doppelte Ebene	parallele Ebenen	parabol. Zylinder
$-$	doppelte Ebene	leer	parabol. Zylinder

Neben der euklidischen Normalform gibt es auch die *affine Normalform.* Dabei sind statt Bewegungen *Affinitäten* erlaubt; eine Affinität ist eine Abbildung der Form

$$\boldsymbol{x} \mapsto A\boldsymbol{x} + \boldsymbol{b} \quad \text{mit} \quad A \in \mathrm{GL}(n, \mathbb{R})$$

(also ein Automorphismus, gefolgt von einer Translation). Der Unterschied zu Bewegungen ist, dass zusätzlich eine Skalierung der Koordinaten möglich ist. Das hat den Effekt, dass in der Normalform aus Satz 24.5 alle $a_j = 1$ gewählt werden können. Die affine Normalform legt bereits den Typ der Quadrik fest, da dieser nur von der Form der Gleichung und den Vorzeichen der quadratischen Terme abhängt. Für Kegelschnitte hat man also die folgenden affinen Normalformen:

Typ	Gleichung
Ellipse	$x^2 + y^2 = 1$
Hyperbel	$x^2 - y^2 = 1$
leere Menge	$-x^2 - y^2 = 1$
Punkt	$x^2 + y^2 = 0$
sich schneidende Geraden	$x^2 - y^2 = 0$
Parabel	$x^2 = y$
parallele Geraden	$x^2 = 1$
leere Menge	$-x^2 = 1$
Doppelgerade	$x^2 = 0$

Für Quadriken im $\mathbb{R}^3$ sieht der „Zoo“ so aus:

Typ	Gleichung
Ellipsoid	$x^2 + y^2 + z^2 = 1$
einschaliges Hyperboloid	$x^2 + y^2 - z^2 = 1$
zweischaliges Hyperboloid	$x^2 - y^2 - z^2 = 1$
leere Menge	$-x^2 - y^2 - z^2 = 1$
Punkt	$x^2 + y^2 + z^2 = 0$
Doppelkegel	$x^2 + y^2 - z^2 = 0$
elliptisches Paraboloid	$x^2 + y^2 = z$
hyperbolisches Paraboloid	$x^2 - y^2 = z$
elliptischer Zylinder	$x^2 + y^2 = 1$
hyperbolischer Zylinder	$x^2 - y^2 = 1$
leere Menge	$-x^2 - y^2 = 1$
Gerade	$x^2 + y^2 = 0$
sich schneidende Ebenen	$x^2 - y^2 = 0$
parabolischer Zylinder	$x^2 = y$
parallele Ebenen	$x^2 = 1$
leere Menge	$-x^2 = 1$
Doppelebene	$x^2 = 0$

25. Orthogonale Gruppen

Wir wollen uns jetzt die (speziellen) orthogonalen und unitären Gruppen $\mathrm{O}(n)$, $\mathrm{SO}(n)$, $\mathrm{U}(n)$ und $\mathrm{SU}(n)$ in kleinen Dimensionen n genauer ansehen und eine allgemeine Aussage über die Elemente von $\mathrm{O}(n)$ beweisen.

Ein ganz trivialer Spezialfall ist $n = 0$, dann sind alle Matrixgruppen triviale Gruppen (sie bestehen nur aus dem neutralen Element, das hier die leere Matrix ist).

Im Fall $n = 1$ haben wir es mit 1×1-Matrizen zu tun, die wir mit Elementen von $\mathbb{R}$ oder $\mathbb{C}$ identifizieren können. Transposition ist hier die Identität, ebenso die Determinante, also erhalten wir

$$\begin{aligned} \mathrm{O}(1) &= \{\lambda \in \mathbb{R} \mid \lambda^2 = 1\} = \{\pm 1\} \subset \mathbb{R}^\times , \\ \mathrm{SO}(1) &= \{\lambda \in \mathbb{R} \mid \lambda = 1\} = \{1\} \subset \mathbb{R}^\times , \\ \mathrm{U}(1) &= \{\lambda \in \mathbb{C} \mid |\lambda|^2 = 1\} = S^1 \subset \mathbb{C}^\times , \\ \mathrm{SU}(1) &= \{\lambda \in \mathbb{C} \mid \lambda = 1\} = \{1\} \subset \mathbb{C}^\times . \end{aligned}$$

Dabei ist

$$S^1 = \bigl\{ z \in \mathbb{C} \bigm| |z| = 1 \bigr\} \subset \mathbb{C}^\times$$

die *Kreisgruppe*. Für $z = x + y\boldsymbol{i} \in S^1$ gilt $1 = |z|^2 = x^2 + y^2$, also gibt es $\alpha \in \mathbb{R}$ mit $x = \cos\alpha$ und $y = \sin\alpha$, also $z = \cos\alpha + \boldsymbol{i}\sin\alpha$. Dieser Winkel α ist bis auf Addition eines ganzzahligen Vielfachen von 2π eindeutig bestimmt. Den Ausdruck für z kann man auch so schreiben:

$$e^{\boldsymbol{i}\alpha} = \sum_{n=0}^{\infty} \frac{(\boldsymbol{i}\alpha)^n}{n!} = \sum_{m=0}^{\infty} (-1)^m \frac{\alpha^{2m}}{(2m)!} + \boldsymbol{i} \sum_{m=0}^{\infty} (-1)^m \frac{\alpha^{2m+1}}{(2m+1)!} = \cos\alpha + \boldsymbol{i}\sin\alpha ;$$

es gilt

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

(durch Ausmultiplizieren der linken Seite und Vergleich von Real- und Imaginärteil erhält man die Additionstheoreme für Sinus und Cosinus).

Das führt zur *Polarkoordinatendarstellung* der komplexen Zahlen: Jede komplexe Zahl $z \in \mathbb{C}$ kann geschrieben werden als $z = re^{\boldsymbol{i}\alpha}$ mit $r = |z| \in \mathbb{R}_{\geq 0}$ eindeutig bestimmt; für $z \neq 0$ ist α wie oben eindeutig bestimmt bis auf Addition eines ganzzahligen Vielfachen von 2π, für $z = 0$ ist α beliebig. Diese Darstellung eignet sich besonders gut zum Multiplizieren:

$$re^{\boldsymbol{i}\alpha} \cdot r'e^{\boldsymbol{i}\alpha'} = (rr')e^{\boldsymbol{i}(\alpha+\alpha')} .$$

Wir betrachten als nächstes $\mathrm{SO}(2)$ und $\mathrm{O}(2)$. Wir erinnern uns daran, dass eine orthogonale Matrix die Eigenschaft hat, dass ihre Spalten (oder Zeilen) eine Orthonormalbasis bilden. Die erste Spalte einer Matrix $A \in \mathrm{O}(2)$ hat also die Form $(\cos\alpha, \sin\alpha)^\top$, denn ihre Länge muss 1 sein. Die zweite Spalte muss ebenfalls Länge 1 haben und auf der ersten senkrecht stehen; dafür gibt es genau die beiden Möglichkeiten

$$A_\alpha^+ = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad \text{und} \qquad A_\alpha^- = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} .$$

Die erste Matrix hat Determinante 1, ist also in $\mathrm{SO}(2)$, die zweite hat Determinante -1, ist also in $\mathrm{O}(2) \setminus \mathrm{SO}(2)$.

A^+_α beschreibt eine Drehung um den Winkel α gegen den Uhrzeigersinn (denn die beiden Standard-Basisvektoren $\mathbf{e}_1$ und $\mathbf{e}_2$ werden auf entsprechend gedrehte Vektoren abgebildet), während A^-_α eine Spiegelung ist: Das charakteristische Polynom ist

$$\begin{vmatrix} X - \cos\alpha & -\sin\alpha \\ -\sin\alpha & X + \cos\alpha \end{vmatrix} = X^2 - \cos^2\alpha - \sin^2\alpha = X^2 - 1 = (X-1)(X+1)\,;$$

also gibt es einen Eigenvektor, der fest bleibt (er spannt die Spiegelungsgerade auf) und senkrecht dazu (das folgt aus dem Spektralsatz, denn die Matrix ist symmetrisch) einen, der das Vorzeichen wechselt. Ist $\alpha = 2\beta$, dann wird die Gerade, an der gespiegelt wird, erzeugt von $(\cos\beta, \sin\beta)^\top$. Daraus, dass A^+_α eine Drehung um den Winkel α beschreibt, folgt auch $A^+_\alpha A^+_\beta = A^+_{\alpha+\beta}$ (das kann man mit den Additionstheoremen auch direkt nachrechnen). Es folgt:

25.1. **Satz.** *Die Abbildung* **SATZ** $\mathrm{U}(1) \cong \mathrm{SO}(2)$

$$\Phi\colon\ \mathrm{U}(1) \longrightarrow \mathrm{SO}(2), \qquad e^{\boldsymbol{i}\alpha} \longmapsto A^+_\alpha\,, \qquad \text{bzw.} \quad x + y\boldsymbol{i} \longmapsto \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

ist ein Gruppenisomorphismus.

Man definiert *Gruppenhomomorphismen* (analog zu Homomorphismen von Vektorräumen) als Abbildungen, die mit der Gruppenstruktur verträglich sind. Konkret ist ein Gruppenhomomorphismus von einer Gruppe $(G, *_G, 1_G, i_G)$ in eine weitere Gruppe $(H, *_H, 1_H, i_H)$ (zu Gruppen siehe Definition 3.6) eine Abbildung $f\colon G \to H$ mit der Eigenschaft

DEF Gruppenhomomorphismus

Gruppenisomorphismus

$$f(g *_G g') = f(g) *_H f(g') \qquad \text{für alle } g, g' \in G.$$

Es folgt dann (im Wesentlichen genauso wie bei linearen Abbildungen hinsichtlich der additiven Gruppenstruktur) $f(1_G) = 1_H$ und $f(i_G(g)) = i_H(f(g))$ für die Inversen. Ein *Gruppenisomorphismus* ist ein bijektiver Gruppenhomomorphismus; in diesem Fall ist die Umkehrabbildung ebenfalls ein Gruppenhomomorphismus.

Gruppen und Gruppenhomomorphismen werden ausführlich in der „Einführung in die Zahlentheorie und algebraische Strukturen“ und in der „Einführung in die Algebra“ besprochen.

Beweis. Dass beide angegebenen Abbildungsvorschriften dieselbe Abbildung definieren, ergibt sich aus

$$A^+_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad \text{und} \qquad e^{\boldsymbol{i}\alpha} = \cos\alpha + \boldsymbol{i}\sin\alpha\,.$$

Die zweite Abbildungsvorschrift zeigt, dass Φ wohldefiniert ist; Φ ist bijektiv, denn auf beiden Seiten ist α durch das Gruppenelement genau bis auf Addition von ganzzahligen Vielfachen von 2π eindeutig bestimmt. Die Abbildung ist auch ein Gruppenhomomorphismus, da gilt

$$\Phi(e^{\boldsymbol{i}\alpha})\Phi(e^{\boldsymbol{i}\beta}) = A^+_\alpha A^+_\beta = A^+_{\alpha+\beta} = \Phi(e^{\boldsymbol{i}(\alpha+\beta)}) = \Phi(e^{\boldsymbol{i}\alpha} \cdot e^{\boldsymbol{i}\beta})\,. \qquad \square$$

Ein weiterer Zugang zu Φ geht über die Struktur von $\mathbb{C}$ als reeller Vektorraum mit der kanonischen Basis $(1, \boldsymbol{i})$. Für $z = x + y\boldsymbol{i} \in \mathbb{C}$ ist die Multiplikation mit z eine $\mathbb{R}$-lineare Abbildung $\mathbb{C} \to \mathbb{C}$, $w \mapsto zw$ (diese Abbildung ist natürlich tatsächlich sogar $\mathbb{C}$-linear). Die Matrix dieses Endomorphismus bezüglich der Basis $(1, \boldsymbol{i})$ ist gerade

$$\Phi(z) = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}.$$

Dies definiert $\Phi\colon \mathbb{C} \to \mathrm{Mat}(2,\mathbb{R})$ mit der Eigenschaft $\Phi(wz) = \Phi(w)\Phi(z)$. Der Isomorphismus Φ im Satz oben ist dann gerade die Einschränkung auf S^1.

Es gilt auch $\Phi(w+z) = \Phi(w) + \Phi(z)$ und $\Phi(1) = I_2$: Φ ist ein injektiver Ringhomomorphismus, der $\mathbb{C}$ isomorph auf den Unterring von $\mathrm{Mat}(2,\mathbb{R})$ abbildet, dessen Elemente alle Matrizen der Form $\left(\begin{smallmatrix} x & -y \\ y & x \end{smallmatrix}\right)$ sind.

Da $\Phi(re^{i\alpha}) = r\Phi(e^{i\alpha})$ ist, sieht man, dass Multiplikation mit $z = re^{i\alpha}$ eine *Drehstreckung* der komplexen Ebene $\mathbb{C} \cong \mathbb{R}^2$ bewirkt: eine Drehung um den Winkel α zusammen mit einer Streckung um den Faktor r.

Sie erinnern sich vielleicht aus dem Schulunterricht, dass sich aus der Verknüpfung zweier Spiegelungen der Ebene eine Drehung ergibt (um den Schnittpunkt der Spiegelachsen; der Drehwinkel ist das Doppelte des orientierten Winkels zwischen den Achsen). Wir wollen das jetzt präzisieren und verallgemeinern.

DEF
orthogonal ähnlich

25.2. **Definition.** Zwei Matrizen $A, B \in \mathrm{Mat}(n,\mathbb{R})$ heißen *orthogonal ähnlich*, wenn es $P \in \mathrm{O}(n)$ gibt mit $B = P^{-1}AP$. $\diamondsuit$

Satz 22.13 kann dann so ausgedrückt werden:

Jede symmetrische reelle Matrix ist orthogonal ähnlich zu einer Diagonalmatrix.

Zwei Matrizen sind genau dann orthogonal ähnlich, wenn sie denselben Endomorphismus bezüglich zweier (möglicherweise) verschiedener ONBen beschreiben.

DEF
orthogonale direkte Summe

25.3. **Definition.** Sei V ein euklidischer (oder unitärer) Vektorraum. Eine Zerlegung $V = \bigoplus_{i\in I} U_i$ als direkte Summe von Untervektorräumen heißt *orthogonal*, wenn die U_i paarweise orthogonal sind:

$$\forall i, j \in I\colon \Big(i \neq j \Rightarrow \forall u_i \in U_i, u_j \in U_j\colon u_i \perp u_j\Big) \qquad \diamondsuit$$

Haben wir eine orthogonale direkte Summe $V = U \oplus U'$ und sind B und B' Orthonormalbasen von U und U', dann ist die Aneinanderhängung von B und B' eine Orthonormalbasis von V. Außerdem gilt $U' = U^\perp$ und $U = U'^\perp$.

LEMMA

25.4. **Lemma.** *Seien V ein endlich-dimensionaler euklidischer Vektorraum und $V = U \oplus U'$ eine orthogonale Zerlegung. Ist $f\colon V \to V$ eine Isometrie und U unter f invariant, dann ist auch U' unter f invariant; insbesondere zerlegt sich f als $f = f|_U \oplus f|_{U'}$.*

Beweis. Für alle $u \in U$ und $u' \in U'$ gilt (wir benutzen, dass f eine Isometrie ist)

$$\langle f(u'), f(u)\rangle = \langle u', u\rangle = 0\,.$$

Da f bijektiv ist, gilt $f(U) = U$ (das folgt aus $f(U) \subset U$ und $\dim f(U) = \dim U$), damit folgt $f(u') \in U^\perp = U'$. ❑

SATZ Normalform von orthogonalen Matrizen

25.5. **Satz.** *Jede orthogonale Matrix* $A \in \mathrm{O}(n)$ *ist orthogonal ähnlich zu einer Block-Diagonalmatrix, deren Blöcke die Form* (1), (-1) *oder* $A^+_\varphi \in \mathrm{SO}(2)$ *mit* $0 < \varphi < \pi$ *haben. Die Blöcke sind bis auf ihre Reihenfolge eindeutig bestimmt.*

Beweis. Wir zeigen die äquivalente Aussage, dass jede lineare Isometrie eines n-dimensionalen euklidischen Vektorraums bezüglich einer geeigneten ONB durch eine Blockmatrix der angegebenen Gestalt beschrieben wird, und zwar durch Induktion über n. Im Fall $n = 0$ ist nichts zu zeigen. Sei also $n > 0$ und $f \in \mathrm{End}(V)$ eine Isometrie, $\dim V = n$.

Hat f einen reellen Eigenwert λ mit Eigenvektor e, den wir auf Länge 1 skalieren können, dann zerlegt sich f als $f = \lambda \,\mathrm{id}_U \oplus f|_{U'}$ mit $U = \langle e\rangle_{\mathbb{R}}$ und $U' = U^\perp$. Wegen $1 = \|e\| = \|f(e)\| = \|\lambda e\| = |\lambda|$ ist dann $\lambda = \pm 1$. Nach Induktionsvoraussetzung gibt es eine ONB B von U', sodass $f|_{U'}$ bezüglich B durch eine Matrix der gewünschten Form beschrieben wird. Dann ist (e, B) eine ONB von V und die Matrix von f bezüglich dieser Basis entsteht durch Ergänzen eines 1×1-Blocks der Form $(\lambda) = (\pm 1)$; sie hat damit ebenfalls die gewünschte Form.

Wir betrachten jetzt den Fall, dass f keinen reellen Eigenwert hat. Wir identifizieren für einen Moment V mit $\mathbb{R}^n$ und f mit einer orthogonalen Matrix A. Wir können A auch als Matrix mit komplexen Einträgen betrachten, dann ist A unitär und der zugehörige Automorphismus $f_{\mathbb{C}}$ von $\mathbb{C}^n$ ist ebenfalls eine Isometrie. Sei λ ein Eigenwert von $f_{\mathbb{C}}$, dann folgt wie im reellen Fall, dass $|\lambda| = 1$ ist, also ist $\lambda = e^{\boldsymbol{i}\varphi}$ für $-\pi < \varphi < \pi$, $\varphi \neq 0$. Mit λ ist auch $\bar\lambda = e^{-\boldsymbol{i}\varphi}$ eine Nullstelle des charakteristischen Polynoms von f (denn das hat reelle Koeffizienten: Aus $\chi_f(\lambda) = 0$ folgt $0 = \overline{\chi_f(\lambda)} = \chi_f(\bar\lambda)$). Wenn wir eventuell λ und $\bar\lambda$ vertauschen, können wir $0 < \varphi < \pi$ annehmen. Ist $e \in \mathbb{C}^n$ ein Eigenvektor zum Eigenwert $\bar\lambda$, dann haben wir

$$\begin{aligned} f\bigl(\mathrm{Re}(e)\bigr) + \boldsymbol{i} f\bigl(\mathrm{Im}(e)\bigr) &= f_{\mathbb{C}}(e) = \bar\lambda e = (\cos\varphi - \boldsymbol{i}\sin\varphi)(\mathrm{Re}(e) + \boldsymbol{i}\,\mathrm{Im}(e)) \\ &= \bigl(\cos\varphi \cdot \mathrm{Re}(e) + \sin\varphi \cdot \mathrm{Im}(e)\bigr) + \boldsymbol{i}\bigl(-\sin\varphi \cdot \mathrm{Re}(e) + \cos\varphi \cdot \mathrm{Im}(e)\bigr)\,. \end{aligned}$$

Außerdem sind $\mathrm{Re}(e)$ und $\mathrm{Im}(e)$ orthogonal und von gleicher Länge, denn im unitären Vektorraum $\mathbb{C}^n$ gilt $e \perp \bar e$ (e und $\bar e$ sind Eigenvektoren zu verschiedenen Eigenwerten des normalen Endomorphismus $f_{\mathbb{C}}$, also nach Satz 22.16 bis auf Skalierung Teil einer ONB); daraus folgt

$$\begin{aligned} 0 = \langle e, \bar e\rangle &= \langle \mathrm{Re}(e) + \boldsymbol{i}\,\mathrm{Im}(e), \mathrm{Re}(e) - \boldsymbol{i}\,\mathrm{Im}(e)\rangle \\ &= \langle \mathrm{Re}(e), \mathrm{Re}(e)\rangle + \boldsymbol{i}\langle \mathrm{Im}(e), \mathrm{Re}(e)\rangle + \boldsymbol{i}\langle \mathrm{Re}(e), \mathrm{Im}(e)\rangle - \langle \mathrm{Im}(e), \mathrm{Im}(e)\rangle \\ &= \|\mathrm{Re}(e)\|^2 - \|\mathrm{Im}(e)\|^2 + 2\boldsymbol{i}\langle \mathrm{Re}(e), \mathrm{Im}(e)\rangle \end{aligned}$$

(beachte, dass $\langle \mathrm{Re}(e), \mathrm{Im}(e)\rangle$ als Skalarprodukt zweier reeller Vektoren reell ist; außerdem ist $\langle x, -\boldsymbol{i}y\rangle = \boldsymbol{i}\langle x, y\rangle$); es folgt $\|\mathrm{Re}(e)\| = \|\mathrm{Im}(e)\|$ und $\langle \mathrm{Re}(e), \mathrm{Im}(e)\rangle = 0$. Bei geeigneter Skalierung bilden also $\mathrm{Re}(e)$ und $\mathrm{Im}(e)$ eine ONB eines zweidimensionalen Untervektorraums U von V und bezüglich dieser Basis ist $f|_U$ gegeben durch die Drehmatrix

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} = A^+_\varphi \in \mathrm{SO}(2)\,.$$

Wir zerlegen $V = U \oplus U^\perp$ und wenden wie im ersten Fall die Induktionsvoraussetzung auf $f|_{U^\perp}$ an. In diesem Fall wird die Matrix durch den Block A^+_φ ergänzt.

Die Eindeutigkeit folgt aus dem Vergleich der (komplexen) Eigenwerte. ❑

Als Spezialfall erhalten wir folgende Aussage:

FOLG
Elemente von SO(3) sind Drehungen

25.6. **Folgerung.** *Sei $A \in \mathrm{SO}(3)$. Dann hat A einen Eigenvektor e_1 zum Eigenwert 1 mit $\|e_1\| = 1$. Ist (e_1, e_2, e_3) eine positiv orientierte Orthonormalbasis mit erstem Element e_1 und $P \in \mathrm{SO}(3)$ die Matrix mit Spalten e_1, e_2, e_3, dann ist*

$$P^{-1}AP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$$

mit einem $\varphi \in \mathbb{R}$. Ist $A \in \mathrm{O}(3) \setminus \mathrm{SO}(3)$, dann hat A den Eigenwert -1, und mit einer geeigneten Matrix $P \in \mathrm{SO}(3)$ hat man

$$P^{-1}AP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$$

mit $\varphi \in \mathbb{R}$.

Beweis. Das folgt aus Satz 25.5 und seinem Beweis. Man beachte, dass man für $\varphi = 0$ zwei Diagonalblöcke (1) und für $\varphi = \pi$ zwei Diagonalblöcke (-1) erhält. ❑

Im Fall $A \in \mathrm{SO}(3)$ beschreibt A also eine Drehung um die Achse $\mathbb{R}e_1$ mit dem Drehwinkel φ (gemessen in der von e_2 und e_3 aufgespannten Ebene in der Orientierung von e_2 nach e_3). Ist $A \in \mathrm{O}(3) \setminus \mathrm{SO}(3)$, dann kommt zur Drehung noch eine Spiegelung an der Ebene $\langle e_2, e_3\rangle_{\mathbb{R}}$ hinzu.

Wir verallgemeinern die Begriffe „Spiegelung“ und „Drehung“ auf höhere Dimensionen.

DEF
Spiegelung
Drehung

25.7. **Definition.** Sei $n \in \mathbb{N}$. Ein Element $A \in \mathrm{O}(n)$ heißt *Spiegelung*, wenn $\dim E_{-1}(A) = 1$ und $\dim E_1(A) = n - 1$ ist.

A heißt *Drehung*, wenn $\dim E_1(A) = n - 2$ und $A \in \mathrm{SO}(n)$ ist. ♢

Ist A eine Spiegelung, dann kann man $\mathbb{R}^n$ als orthogonale direkte Summe zerlegen in $\mathbb{R}^n = \langle e_{-1}\rangle_{\mathbb{R}} \oplus E_1(A)$; dabei sei e_{-1} ein Eigenvektor der Länge 1 zum Eigenwert -1. Man kann also jedes $\boldsymbol{x} \in \mathbb{R}^n$ schreiben als $\boldsymbol{x} = \lambda e_{-1} + \boldsymbol{y}$ mit $\boldsymbol{y} \perp e_{-1}$, und dann ist $A\boldsymbol{x} = -\lambda e_{-1} + \boldsymbol{y}$, was genau eine Spiegelung an der Hyperebene $E_1(A)$ beschreibt. Wir haben oben schon gesehen, dass jedes Element von $\mathrm{O}(2) \setminus \mathrm{SO}(2)$ eine Spiegelung ist. Für eine Spiegelung A gilt $A^2 = I_n$.

Ist A eine Drehung, dann hat die Normalform von A eine Drehmatrix als Block (oder zwei Blöcke (-1)) und sonst nur Blöcke (1).

LEMMA
Sp. ∘ Sp. = Drehung

25.8. **Lemma.** *Das Produkt zweier verschiedener Spiegelungen ist eine Drehung. Jede Drehung lässt sich als Produkt zweier Spiegelungen schreiben.*

Beweis. Seien A und B zwei verschiedene Spiegelungen in $\mathrm{O}(n)$. Wegen $A \neq B$ ist $\dim(E_1(A) \cap E_1(B)) = n - 2$ (aus der Dimensionsformel für Summen und Durchschnitte folgt, dass diese Dimension $n-2$ oder $n-1$ sein muss; wäre sie $n-1$, dann wäre $E_1(A) = E_1(B)$, also auch $E_{-1}(A) = E_1(A)^{\perp} = E_1(B)^{\perp} = E_{-1}(B)$, was $A = B$ bedeuten würde). Wir wählen eine ONB von $\mathbb{R}^n$, deren letzte $n - 2$ Elemente eine Basis von $U = E_1(A) \cap E_1(B)$ bilden. Bezüglich dieser ONB haben die Spiegelungen die Form $A' \oplus I_{n-2}$ und $B' \oplus I_{n-2}$ mit $A', B' \in \mathrm{O}(2) \setminus \mathrm{SO}(2)$. Dann ist aber $A'B' \in \mathrm{SO}(2)$ (denn $\det(A'B') = \det(A')\det(B') = (-1)^2 = 1$) und $A'B' \neq I_2$; es folgt, dass AB eine Drehung ist.

Sei jetzt umgekehrt A eine Drehung. Durch eine geeignete Zerlegung von $\mathbb{R}^n$ als orthogonale direkte Summe zerlegt sich A in der Form $A' \oplus I_{n-2}$ mit $A' \in \mathrm{SO}(2)$. Es genügt also, die Behauptung für A' zu zeigen. Ist 2φ der Drehwinkel, dann gilt

$$\begin{aligned} A' &= \begin{pmatrix} \cos 2\varphi & -\sin 2\varphi \\ \sin 2\varphi & \cos 2\varphi \end{pmatrix} = \begin{pmatrix} \cos^2\varphi - \sin^2\varphi & -2\sin\varphi\cos\varphi \\ 2\sin\varphi\cos\varphi & \cos^2\varphi - \sin^2\varphi \end{pmatrix} \\ &= \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

und alle Matrizen im letzten Produkt sind in $\mathrm{O}(2)$, wobei das Produkt der ersten drei und auch die letzte Matrix in $\mathrm{O}(2) \setminus \mathrm{SO}(2)$, also Spiegelungen sind. ❑

* 25.9. **Folgerung.** *Jedes Element von* $\mathrm{O}(n)$ *ist ein Produkt von höchstens* n *Spiegelungen. Für Elemente von* $\mathrm{SO}(n)$ *ist die Anzahl der Spiegelungen gerade, sonst ungerade.*

FOLG Spiegelungen erzeugen $\mathrm{O}(n)$

Beweis. Sei $A \in \mathrm{O}(n)$. Nach Satz 25.5 gibt es $P \in \mathrm{O}(n)$ mit

$$P^{-1}AP = \mathrm{diag}\big(\underbrace{(1), \ldots, (1)}_{r}, \underbrace{(-1), \ldots, (-1)}_{s}, A^+_{\varphi_1}, \ldots, A^+_{\varphi_t}\big)$$

mit $r+s+2t = n$. Es ist klar, dass die rechte Seite geschrieben werden kann als ein Produkt von s Spiegelungen und t Drehungen (man ersetze jeweils alle Blöcke bis auf einen durch die entsprechende Einheitsmatrix). Jede Drehung ist ein Produkt von zwei Spiegelungen; insgesamt hat man ein Produkt von $s + 2t \leq n$ Spiegelungen $S_1, \ldots, S_{s+2t}$. Dann ist $A = (PS_1P^{-1})(PS_2P^{-1}) \cdots (PS_{s+2t}P^{-1})$ ebenfalls ein Produkt von $s + 2t$ Spiegelungen.

Die zweite Aussage folgt aus einem Vergleich der Determinanten, denn eine Spiegelung hat Determinante -1. ❑

Um eine Drehung im Raum zu beschreiben, braucht man eine Matrix mit neun reellen Einträgen. Auf der anderen Seite zeigt Folgerung 25.6, dass so eine Drehung durch den Einheitsvektor e_1 (der die Drehachse beschreibt) und den Winkel α eindeutig beschrieben werden kann. Da $e_1 \in S^2 = \{\boldsymbol{x} \in \mathbb{R}^3 \mid \|\boldsymbol{x}\| = 1\}$ sich in etwas Zweidimensionalem und α oder äquivalent $e^{i\alpha} \in S^1$ sich in etwas Eindimensionalem bewegt, gibt es eigentlich nur drei „Freiheitsgrade" und nicht neun. Eine Möglichkeit, Drehungen mit drei Parametern zu beschreiben, ist die Darstellung durch die „Eulerschen Winkel". Dabei wird eine beliebige Drehung durch eine Abfolge von (höchstens) drei Drehungen um vorgegebene Achsen erhalten, die entweder Koordinatenachsen oder von den schon erfolgten Drehungen mitgedrehte Koordinatenachsen sind. Eine Variante (von insgesamt 24, je nach Auswahl der drei Drehachsen und Wahl eines starren oder mitbewegten Koordinatensystems) ist wie folgt.

Definition. Sei $\rho \in \mathrm{SO}(3)$ eine Drehung. Dann sind die *Eulerschen Winkel* von ρ wie folgt definiert. Wir setzen $\mathbf{e}'_j = \rho(\mathbf{e}_j)$.

DEF Eulersche Winkel

- $\vartheta = \angle(\mathbf{e}_3, \mathbf{e}'_3)$ (nicht orientiert) mit $0 \leq \vartheta \leq \pi$.
- Falls $0 < \vartheta < \pi$, dann sei $g = \langle \mathbf{e}_3, \mathbf{e}'_3 \rangle^\perp$ die sogenannte *Knotenlinie*. Es gibt genau einen Vektor $v \in g$ der Länge 1, sodass $(\mathbf{e}_3, \mathbf{e}'_3, v)$ positiv orientiert ist. Wir setzen $\varphi = \angle(\mathbf{e}_1, v)$ (orientierter Winkel in $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$) und $\psi = \angle(v, \mathbf{e}'_1)$ (orientierter Winkel in $\langle \mathbf{e}'_1, \mathbf{e}'_2 \rangle$) mit $0 \leq \varphi, \psi < 2\pi$.
- Falls $\vartheta = \pi$, dann muß ρ die Drehung um den Winkel π um eine in $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$ liegende Achse g sein, die dann *Knotenlinie* heißt. Wähle $v \in g$ der Länge 1, sodass $0 \leq \varphi = \angle(\mathbf{e}_1, v) \leq \pi$ (als orientierter Winkel in $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$) und setze $\psi = -\varphi$.

- Falls $\vartheta = 0$, dann ist ρ eine Drehung um $\langle \mathbf{e}_3 \rangle$ um einen Winkel φ. Wir setzen $\psi = 0$. $\diamondsuit$

Wir setzen

$$S_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \qquad \text{und} \qquad S_3(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$S_1(\alpha)$ ist die Drehung um die x-Achse $\langle \mathbf{e}_1 \rangle$ um den Winkel α; $S_3(\alpha)$ die Drehung um die z-Achse $\langle \mathbf{e}_3 \rangle$ um den Winkel α.

SATZ
Eulersche Winkel

Satz. *Sei $\rho \in \mathrm{SO}(3)$ mit Eulerschen Winkeln ϑ, φ, ψ. Dann ist*

$$\rho = S_3(\varphi) \circ S_1(\vartheta) \circ S_3(\psi).$$

Beweis. Übung. ❑

Eine andere und für manche Anwendungen besser geeignete Darstellung von Drehungen kann mit Hilfe der Quaternionen gegeben werden.

DEF
Schiefkörper der Quaternionen

Definition. Sei $\mathbb{H} = \mathbb{R}^4$, wobei wir die Elemente der Standardbasis mit $1, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ bezeichnen. Dann wird $\mathbb{H}$ zu einem Schiefkörper durch die Multiplikation, die die Skalarmultiplikation fortsetzt, die Distributivgesetze erfüllt und auf der Basis durch

$$\begin{array}{llll} 1\boldsymbol{i} = \boldsymbol{i} = \boldsymbol{i}1, & 1\boldsymbol{j} = \boldsymbol{j} = \boldsymbol{j}1, & 1\boldsymbol{k} = \boldsymbol{k} = \boldsymbol{k}1, & \boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1 \\ \boldsymbol{ij} = \boldsymbol{k}, \quad \boldsymbol{ji} = -\boldsymbol{k}, & \boldsymbol{jk} = \boldsymbol{i}, & \boldsymbol{kj} = -\boldsymbol{i}, & \boldsymbol{ki} = \boldsymbol{j}, \quad \boldsymbol{ik} = -\boldsymbol{j} \end{array}$$

festgelegt ist. Mit dieser Struktur heißt $\mathbb{H}$ der Schiefkörper der *Quaternionen*, ein Element von $\mathbb{H}$ heißt eine *Quaternion*.

Ist $\alpha = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \in \mathbb{H}$ mit $a, b, c, d \in \mathbb{R}$, dann ist $\bar{\alpha} = a - b\boldsymbol{i} - c\boldsymbol{j} - d\boldsymbol{k}$ die zu α *konjugierte* Quaternion. Ist $\bar{\alpha} = \alpha$, dann heißt α *reell*; ist $\bar{\alpha} = -\alpha$, dann heißt α eine *reine Quaternion*. Der dreidimensionale Untervektorraum der reinen Quaternionen wird mit $\operatorname{Im}\mathbb{H}$ bezeichnet. $\operatorname{Re}(\alpha) = a \in \mathbb{R}$ heißt der *Skalarteil* von α, $\operatorname{Im}(\alpha) = b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \in \operatorname{Im}\mathbb{H}$ der *Vektorteil* von α. $\diamondsuit$

Die Bezeichnung $\mathbb{H}$ ehrt Sir William Rowan Hamilton, der die Quaternionen (wieder-)entdeckte, ihnen ihren Namen gab und sie intensiv studierte.

Hier ist natürlich noch Einiges zu zeigen.

- $\mathbb{H}$ ist ein Ring:
  Das geht wohl am einfachsten dadurch, dass man die Elemente von $\mathbb{H}$ mit gewissen komplexen 2×2-Matrizen identifiziert:

$$\Psi \colon \mathbb{H} \longrightarrow \mathrm{Mat}(2, \mathbb{C}), \quad a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \longmapsto \begin{pmatrix} a + b\boldsymbol{i} & c + d\boldsymbol{i} \\ -c + d\boldsymbol{i} & a - b\boldsymbol{i} \end{pmatrix};$$

  das Bild von Ψ besteht aus allen Matrizen der Form $\left(\begin{smallmatrix} z & w \\ -\bar{w} & \bar{z} \end{smallmatrix}\right)$ mit $z, w \in \mathbb{C}$. Es ist klar, dass $\Psi(\alpha + \beta) = \Psi(\alpha) + \Psi(\beta)$ ist, und man rechnet nach, dass auch $\Psi(\alpha\beta) = \Psi(\alpha)\Psi(\beta)$ gilt (es genügt, das auf der Basis zu prüfen). Da die Ring-Axiome im Matrizenring gelten, ist auch $\mathbb{H}$ ein Ring (den wir mit einem Unterring von $\mathrm{Mat}(2, \mathbb{C})$ identifizieren können).
- $\mathbb{H}$ ist ein Schiefkörper:
  Es ist klar, dass die Multiplikation in $\mathbb{H}$ nicht kommutativ ist. Ist $\alpha \in \mathbb{H}$ nicht null, dann ist (mit α wie oben und $z = a + b\boldsymbol{i}$, $w = c + d\boldsymbol{i}$)

$$\det(\Psi(\alpha)) = z\bar{z} + w\bar{w} = |z|^2 + |w|^2 = a^2 + b^2 + c^2 + d^2 > 0,$$

  also ist $\Psi(\alpha)$ invertierbar, und weil

$$\begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix}^{-1} = \frac{1}{|z|^2 + |w|^2} \begin{pmatrix} \bar{z} & -w \\ \bar{w} & z \end{pmatrix}$$

wieder im Bild von Ψ liegt, ist auch α invertierbar mit $\alpha^{-1} = \Psi^{-1}\big(\Psi(\alpha)^{-1}\big)$.

Man sieht leicht, dass $\Psi(\bar{\alpha}) = \Psi(\alpha)^*$ ist; daraus folgt

$$\overline{\alpha + \beta} = \bar{\alpha} + \bar{\beta} \qquad \text{und} \qquad \overline{\alpha\beta} = \bar{\beta} \cdot \bar{\alpha}\,.$$

Man beachte die Vertauschung der Faktoren! Außerdem gilt für $\alpha = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$

$$\alpha\bar{\alpha} = \bar{\alpha}\alpha = a^2 + b^2 + c^2 + d^2 = |\alpha|^2\,;$$

so definieren wir $|\alpha| \in \mathbb{R}_{\geq 0}$. Aus den Eigenschaften der Konjugation oder aus $\det(\Psi(\alpha)) = |\alpha|^2$ folgt $|\alpha\beta| = |\alpha| \cdot |\beta|$. Man kann zum Beispiel so argumentieren:

$$|\alpha\beta|^2 = (\alpha\beta)\overline{(\alpha\beta)} = \alpha\beta\bar{\beta}\bar{\alpha} = \alpha|\beta|^2\bar{\alpha} = \alpha\bar{\alpha}\,|\beta|^2 = |\alpha|^2\,|\beta|^2\,.$$

Wenn man für $\xi = x_1 + x_2\boldsymbol{i} + x_3\boldsymbol{j} + x_4\boldsymbol{k}$ und $\eta = y_1 + y_2\boldsymbol{i} + y_3\boldsymbol{j} + y_4\boldsymbol{k}$ die Gleichung $|\xi|^2|\eta|^2 = |\bar{\xi}\eta|^2$ ausschreibt (beachte $|\xi| = |\bar{\xi}|$), erhält man eine Formel, die ein Produkt von Summen von vier Quadraten wieder als Summe von vier Quadraten darstellt:

$$\begin{aligned}
&(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) \\
&\qquad = (x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4)^2 + (x_1y_2 - x_2y_1 - x_3y_4 + x_4y_3)^2 \\
&\qquad\quad + (x_1y_3 + x_2y_4 - x_3y_1 - x_4y_2)^2 + (x_1y_4 - x_2y_3 + x_3y_2 - x_4y_1)^2
\end{aligned}$$

Daraus folgt zum Beispiel, dass das Produkt zweier natürlicher Zahlen, die Summen von vier Quadratzahlen sind, wieder eine Summe von vier Quadratzahlen ist. Dies ist ein wichtiger Schritt im Beweis des *Vier-Quadrate-Satzes* von Lagrange. Der Satz besagt, dass *jede* natürliche Zahl n Summe von vier Quadratzahlen ist (dabei ist null als Summand erlaubt); die eben gemachte Beobachtung erlaubt es, sich auf den Fall zu beschränken, dass n eine Primzahl ist.

Die obige Gleichung ist analog zur entsprechenden Gleichung für zwei Quadrate, die man aus der Multiplikativität des komplexen Absolutbetrags erhält:

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 + x_2y_2)^2 + (x_1y_2 - x_2y_1)^2\,.$$

Wenn wir reine Quaternionen mit Vektoren im $\mathbb{R}^3$ identifizieren, dann lässt sich das Produkt zweier reiner Quaternionen recht elegant schreiben als

$$\xi \cdot \eta = -\langle \xi, \eta \rangle + \xi \times \eta\,;$$

der Skalarteil des Produkts ist also bis auf das Vorzeichen das Skalarprodukt und der Vektorteil ist das Vektorprodukt der beiden Vektoren. Aus der Multiplikativität des Betrags folgt dann

$$|\xi|^2|\eta|^2 = \langle \xi, \eta \rangle^2 + |\xi \times \eta|^2 = |\xi|^2|\eta|^2 \cos^2 \angle(\xi, \eta) + |\xi \times \eta|^2$$

und damit $|\xi \times \eta| = |\xi||\eta| \sin \angle(\xi, \eta)$ (der Sinus ist positiv).

Aus der Multiplikativität des Absolutbetrags folgt auch, dass

$$S^3 = \big\{\alpha \in \mathbb{H} \bigm| |\alpha| = 1\big\}$$

eine (nicht-kommutative) Gruppe unter der Multiplikation von $\mathbb{H}$ ist. Die Matrizen im Bild von Ψ haben die Eigenschaft, dass ihre beiden Spalten (oder auch Zeilen) dieselbe Länge haben und zueinander orthogonal sind (bezüglich des unitären Skalarprodukts auf $\mathbb{C}^2$). Die Länge der Spalten von $\Psi(\alpha)$ ist gerade $|\alpha|$. Daraus folgt, dass $\Psi(S^3) \subset \mathrm{SU}(2)$ ist (denn für $|\alpha| = 1$ ist $\Psi(\alpha)$ unitär und die Determinante ist $\det(\Psi(\alpha)) = |\alpha|^2 = 1$). Umgekehrt liegt jedes Element von $\mathrm{SU}(2)$ im Bild von Ψ (denn die erste Zeile hat die Form (z, w) mit $|z|^2 + |w|^2 = 1$, dann muss die zweite Zeile die Form $\lambda(-\bar{w}, \bar{z})$ haben mit $|\lambda| = 1$, und da die Determinante dann λ ist, muss $\lambda = 1$ sein). Es folgt:

SATZ $S^3 \cong \mathrm{SU}(2)$

Satz. *Die Einschränkung von Ψ liefert einen Gruppenisomorphismus*

$$S^3 \longrightarrow \mathrm{SU}(2)\,.$$

Multiplikation mit einer Quaternion von links oder von rechts ergibt einen Endomorphismus von $\mathbb{H}$ als reeller Vektorraum. Wir können auch von links *und* rechts mit jeweils einer fest gewählten Quaternion multiplizieren.

LEMMA Invarianz von $\operatorname{Im}\mathbb{H}$

Lemma. *Sei $\alpha \in \mathbb{H}$ und $m_\alpha \colon \mathbb{H} \to \mathbb{H}$, $\xi \mapsto \alpha\xi\bar{\alpha}$. Dann ist $\operatorname{Im}\mathbb{H}$ ein unter m_α invarianter reeller Untervektorraum von $\mathbb{H}$.*

Ist $\alpha \in S^3$, dann ist die Einschränkung von m_α auf $\operatorname{Im}\mathbb{H}$ eine orientierungserhaltende Isometrie (also eine Drehung), und alle Drehungen von $\operatorname{Im}\mathbb{H}$ haben diese Form.

Beweis. Sei $\xi \in \operatorname{Im}\mathbb{H}$, also $\bar{\xi} = -\xi$. Dann gilt

$$\overline{m_\alpha(\xi)} = \overline{\alpha\xi\bar{\alpha}} = \bar{\bar{\alpha}}\bar{\xi}\bar{\alpha} = \alpha(-\xi)\bar{\alpha} = -\alpha\xi\bar{\alpha} = -m_\alpha(\xi)\,,$$

also ist $m_\alpha(\xi) \in \operatorname{Im}\mathbb{H}$. Weiter gilt für $\alpha \in S^3$

$$|m_\alpha(\xi)| = |\alpha\xi\bar{\alpha}| = |\alpha||\xi||\bar{\alpha}| = |\alpha|^2|\xi| = |\xi|\,,$$

also ist m_α eine Isometrie.

Alle Links- oder Rechts-Multiplikationen mit festen Quaternionen $\beta \neq 0$ haben Determinante $|\beta|^4 > 0$, sind also orientierungserhaltend. Damit ist m_α als Automorphismus von $\mathbb{H}$ orientierungserhaltend. Wegen $m_\alpha(1) = 1$ (für $\alpha \in S^3$) hat die Einschränkung von m_α auf $\operatorname{Im}\mathbb{H}$ dieselbe Determinante 1 wie m_α. Also ist auch die Einschränkung auf $\operatorname{Im}\mathbb{H}$ orientierungserhaltend.

Eine Drehung um die vom Einheitsvektor $\varepsilon \in \operatorname{Im}\mathbb{H}$ erzeugte Gerade mit dem Winkel 2φ bekommt man als m_α mit $\alpha = \cos\varphi + \varepsilon\sin\varphi$: Es gilt dann $\alpha\varepsilon = \varepsilon\alpha$, also

$$m_\alpha(\varepsilon) = \alpha\varepsilon\bar{\alpha} = \alpha\varepsilon\alpha^{-1} = \varepsilon\alpha\alpha^{-1} = \varepsilon$$

und für $\xi \in \operatorname{Im}\mathbb{H}$ mit $\langle\varepsilon, \xi\rangle = 0$ gilt

$$\varepsilon\xi = -\xi\varepsilon = \varepsilon \times \xi = \xi \text{ um } \pi/2 \text{ um die Achse } \mathbb{R}\varepsilon \text{ gedreht}$$

und $\varepsilon\xi\varepsilon = \xi$. Es folgt

$$m_\alpha(\xi) = (\cos^2\varphi - \sin^2\varphi)\xi + 2\cos\varphi\sin\varphi(\varepsilon \times \xi) = \cos(2\varphi)\xi + \sin(2\varphi)(\varepsilon \times \xi)\,,$$

was genau eine Drehung um den Winkel 2φ in der zu ε senkrechten Ebene in $\operatorname{Im}\mathbb{H}$ beschreibt. ❑

Analog zu linearen Abbildungen definiert man den *Kern* eines Gruppenhomomorphismus $f \colon G \to H$ als $\ker(f) = \{g \in G \mid f(g) = 1_H\}$.

SATZ $\mathrm{SO}(3) \cong \mathrm{SU}(2)/\{\pm I\}$

Satz. *Die Abbildung*

$$S^3 \longrightarrow \mathrm{SO}(3), \quad \alpha \longmapsto m_\alpha|_{\operatorname{Im}\mathbb{H}}$$

ist ein surjektiver Gruppenhomomorphismus mit Kern $\{\pm 1\}$.

Die Verknüpfung $\mathrm{SU}(2) \stackrel{\Psi^{-1}}{\to} S^3 \to SO(3)$ liefert demnach einen surjektiven Gruppenhomomorphismus $\mathrm{SU}(2) \to \mathrm{SO}(3)$ mit Kern $\{\pm I_2\}$.

Ähnlich wie wir für einen Vektorraum V und einen Untervektorraum U den Quotientenvektorraum V/U definiert haben, kann man für eine Gruppe G und eine Untergruppe H (die eine zusätzliche Eigenschaft haben muss — sie muss ein sogenannter *Normalteiler* sein) die Quotientengruppe G/H definieren. Der Kern eines Gruppenhomomorphismus $\varphi \colon G \to G'$ ist stets ein Normalteiler, und man hat wieder einen *Homomorphiesatz* $G/\ker(\varphi) \cong \operatorname{im}(\varphi)$.

Beweis. Dass die Abbildung wohldefiniert und surjektiv ist, wurde in Lemma 25 gezeigt. Dass es sich um einen Gruppenhomomorphismus handelt, sieht man so:

$$m_{\alpha\beta}(\xi) = (\alpha\beta)\xi(\overline{\alpha\beta}) = \alpha\beta\xi\bar{\beta}\bar{\alpha} = \alpha m_\beta(\xi)\bar{\alpha} = m_\alpha\big(m_\beta(\xi)\big) = (m_\alpha \circ m_\beta)(\xi)\,.$$

Ist $\alpha \in S^3$ im Kern, dann gilt $\alpha\xi\bar{\alpha} = \xi$, oder äquivalent (wegen $\bar{\alpha}\alpha = 1$) $\alpha\xi = \xi\alpha$ für alle $\xi \in \operatorname{Im}\mathbb{H}$ und damit auch für alle $\xi \in \mathbb{H}$. Schreibt man $\alpha = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$ und setzt $\xi = \boldsymbol{i}, \boldsymbol{j}$ ein, dann sieht man, dass $b = c = d = 0$ sein müssen. Es folgt $\alpha = a = \pm 1$. Umgekehrt ist klar, dass diese beiden Elemente im Kern liegen. ❑

Wenn man es vermeiden möchte, die Quaternionen, mit denen man rechnet, auf Länge 1 zu bringen, dann kann man auch für beliebiges $\alpha \in \mathbb{H}^\times$ die Abbildung

$$\xi \longmapsto \frac{1}{|\alpha|^2}\,\alpha\xi\bar{\alpha} = \alpha\xi\alpha^{-1}$$

(denn es ist $\alpha\bar{\alpha} = |\alpha|^2$, also $\bar{\alpha}/|\alpha|^2 = \alpha^{-1}$) betrachten. Das ist gleichbedeutend mit $m_{\alpha/|\alpha|}$, hat aber den Vorteil, dass man Quadratwurzeln vermeidet. Das ergibt dann einen surjektiven Gruppenhomomorphismus $\mathbb{H}^\times \to \mathrm{SO}(3)$ mit Kern $\mathbb{R}^\times$.

In jedem Fall sieht man, dass eine Drehung im $\mathbb{R}^3 \cong \operatorname{Im}\mathbb{H}$ durch eine Quaternion (bis auf reelle Skalierung), also durch ein Quadrupel reeller Zahlen, beschrieben werden kann. Verknüpfung von Drehungen entspricht der Multiplikation von Quaternionen. Das bedeutet 16 reelle Multiplikationen, während die Multiplikation zweier reeller 3×3-Matrizen 27 reelle Multiplikationen benötigt. (In beiden Fällen lässt sich die Anzahl der Multiplikationen durch geschicktes Umformen auf Kosten von zusätzlichen Additionen verringern; trotzdem bleibt die Version mit Quaternionen vorteilhaft.) Wegen der effizienteren Darstellung und Verknüpfung werden Quaternionen daher in Anwendungen wie zum Beispiel in der Computergrafik eingesetzt. Abgesehen davon bilden sie aber natürlich auch an sich ein interessantes mathematisches Objekt!

26. Die Jordansche Normalform

C. Jordan
1838–1922

Unser nächstes Ziel wird die Klassifikation von Endomorphismen durch die Jordansche Normalform sein. Wir setzen voraus, dass f ein Endomorphismus eines endlichdimensionalen K-Vektorraums ist, dessen charakteristisches Polynom in Linearfaktoren zerfällt. Ist f diagonalisierbar, dann zerlegt sich V in die direkte Summe der nichttrivialen Eigenräume von f. Als ersten Schritt verallgemeinern wir diese Zerlegung.

Dazu erst noch ein Lemma über Polynome.

LEMMA

26.1. **Lemma.** *Seien K ein Körper, $\lambda \in K$ und $p \in K[X]$ ein Polynom mit $p(\lambda) \neq 0$. Sei weiter $n \in \mathbb{Z}_{>0}$. Dann gibt es ein Polynom $q \in K[X]$ mit $\deg(q) < n$, sodass $q(X)p(X) - 1$ durch $(X - \lambda)^n$ teilbar ist.*

Beweis. Induktion über n. Im Fall $n = 1$ setzen wir $q(X) = p(\lambda)^{-1}$ (das ist ein konstantes Polynom). Für den Induktionsschritt nehmen wir an, dass die Aussage für n gilt und zeigen sie für $n+1$. Sei dazu q_1 ein Polynom mit $\deg(q_1) < n$, sodass $q_1(X)p(X) - 1 = (X - \lambda)^n r(X)$ ist mit einem Polynom r; q_1 existiert nach der Induktionsannahme. Wir machen den Ansatz $q(X) = q_1(X) + a \cdot (X - \lambda)^n$ (mit $a \in K$), dann ist

$$q(X)p(X) - 1 = (q_1(X)p(X) - 1) + a \cdot (X - \lambda)^n p(X) = (X - \lambda)^n \big(r(X) + ap(X)\big) .$$

Wenn wir $a = -r(\lambda)p(\lambda)^{-1}$ setzen, dann verschwindet die letzte Klammer für $X = \lambda$, ist also durch $X - \lambda$ teilbar, und $q(X)p(X) - 1$ ist wie gewünscht durch $(X - \lambda)^{n+1}$ teilbar. ❑

Jetzt beweisen wir die Existenz einer passenden Zerlegung.

SATZ
Zerlegung von Endomorphismen

26.2. **Satz.** *Seien K ein Körper, V ein K-Vektorraum und $f \in \operatorname{End}(V)$. Seien weiter $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$ paarweise verschieden und $e_1, e_2, \ldots, e_n \geq 1$; wir setzen*

$$p(X) = (X - \lambda_1)^{e_1}(X - \lambda_2)^{e_2} \cdots (X - \lambda_n)^{e_n} \in K[X] .$$

Wir nehmen an, dass $p(f) = \mathbf{0}$ ist und definieren $U_j = \ker\big((f - \lambda_j \operatorname{id}_V)^{\circ e_j}\big)$. Dann sind die U_j invariant unter f und wir haben die Zerlegung $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$.

Das verallgemeinert die (interessante Richtung in der) Aussage von Satz 16.23: Sind alle Exponenten $e_j = 1$, dann erhält man die Zerlegung in Eigenräume wie dort.

Beweis. Für $j \in \{1, 2, \ldots, n\}$ sei p_j das Produkt der Faktoren $(X - \lambda_i)^{e_i}$ für alle $i \in \{1, 2, \ldots, n\} \setminus \{j\}$ (also ohne den $j$ten Faktor). Dann ist $p_j(\lambda_j) \neq 0$, also gibt es nach Lemma 26.1 Polynome $q_j, r_j \in K[X]$ mit

$$\deg(q_j) < e_j \qquad \text{und} \qquad q_j(X)p_j(X) - 1 = (X - \lambda_j)^{e_j} r_j(X) .$$

Da jedes $p_i(X)$ für $i \neq j$ ebenfalls durch $(X - \lambda_j)^{e_j}$ teilbar ist, gilt für die Summe $s(X) = q_1(X)p_1(X) + \ldots + q_n(X)p_n(X)$, dass $s(X) - 1$ durch $(X - \lambda_j)^{e_j}$ teilbar ist. Da das für jedes j richtig ist, muss $s(X) - 1$ durch das Produkt aller $(X - \lambda_j)^{e_j}$ teilbar sein, also durch $p(X)$. Da aber $\deg(s) < \deg(p)$ ist, muss $s(X) = 1$ sein.

Wir zeigen jetzt, dass U_j unter f invariant ist. Sei

$$g_j = (f - \lambda_j \operatorname{id}_V)^{\circ e_j} = (X - \lambda_j)^{e_j}(f) ;$$

dann kommutiert g_j mit f:

$$f \circ g_j = (X(X-\lambda_j)^{e_j}))(f) = ((X-\lambda_j)^{e_j}X)(f) = g_j \circ f\,.$$

Ist $v \in U_j$, also $g_j(v) = \mathbf{0}$, dann folgt $g_j(f(v)) = f(g_j(v)) = f(\mathbf{0}) = \mathbf{0}$, also ist $f(v) \in U_j$.

Da $p(f) = \mathbf{0}$ ist, gilt $\operatorname{im}(p_j(f)) \subset U_j$: Ist $v \in \operatorname{im}(p_j(f))$, dann gibt es $w \in V$ mit $v = (p_j(f))(w)$; es folgt

$$\mathbf{0} = (p(f))(w) = (f - \lambda_j \operatorname{id}_V)^{\circ e_j}\big((p_j(f))(w)\big) = g_j(v)\,,$$

also ist $v \in \ker(g_j) = U_j$.

Wir zeigen jetzt, dass $V = U_1 + U_2 + \ldots + U_n$ ist. Sei dazu $v \in V$. Dann ist

$$\begin{aligned} v &= \operatorname{id}_V(v) = (s(f))(v) \\ &= (p_1(f))\big((q_1(f))(v)\big) + (p_2(f))\big((q_2(f))(v)\big) + \ldots + (p_n(f))\big((q_n(f))(v)\big) \\ &= v_1 + v_2 + \ldots + v_n \end{aligned}$$

mit $v_j = (p_j(f))\big((q_j(f))(v)\big) \in \operatorname{im}(p_j(f)) \subset U_j$.

Wir zeigen noch, dass die Summe direkt ist. Seien $v_j \in V_j$ mit $v_1 + \ldots + v_n = \mathbf{0}$; wir müssen zeigen, dass alle $v_j = \mathbf{0}$ sind. Es gilt aber

$$\begin{aligned} \mathbf{0} &= (q_j(f))\big((p_j(f))(v_1 + \ldots + v_n)\big) = ((q_j p_j)(f))(v_j) \\ &= v_j + (r_j(f))\big((f - \lambda_j \operatorname{id}_V)^{\circ e_j}(v_j)\big) = v_j\,. \end{aligned}$$

Beachte dabei, dass $(p_j(f))(v_i) = \mathbf{0}$ ist für $i \neq j$ (denn p_j enthält den Faktor $(X - \lambda_i)^{e_i}$ und $v_i \in \ker\big((f - \lambda_i \operatorname{id}_V)^{\circ e_i}\big)$). ❑

Das legt folgende Definition nahe. Wir bemerken zuerst, dass für zwei lineare Abbildungen $V_1 \xrightarrow{f} V_2 \xrightarrow{g} V_3$ gilt $\ker(f) \subset \ker(g \circ f)$ (denn aus $f(v) = \mathbf{0}$ folgt $(g \circ f)(v) = g(\mathbf{0}) = \mathbf{0}$). Insbesondere ist für $f \in \operatorname{End}(V)$

$$\{\mathbf{0}\} = \ker(\operatorname{id}_V) \subset \ker(f) \subset \ker(f^{\circ 2}) \subset \ker(f^{\circ 3}) \subset \ldots$$

eine aufsteigende Kette von Untervektorräumen von V.

* 26.3. **Definition.** Sei f ein Endomorphismus eines endlich-dimensionalen K-Vektorraums V und sei $\lambda \in K$. Dann heißt der Untervektorraum **DEF** Hauptraum

$$H_\lambda(f) = \bigcup_{m \in \mathbb{N}} \ker\big((f - \lambda \operatorname{id}_V)^{\circ m}\big) \subset V$$

der *Hauptraum* oder *verallgemeinerte Eigenraum* von f zum Eigenwert λ. Analog definieren wir $H_\lambda(A)$ für Matrizen $A \in \operatorname{Mat}(n, K)$. ♢

Nach Lemma 7.5 ist $H_\lambda(f)$ als aufsteigende Vereinigung von Untervektorräumen ein Untervektorraum von V.

26.4. **Lemma.** *In der Situation von Satz 26.2 mit V endlich-dimensional gilt $U_j = H_{\lambda_j}(f)$ für alle $j \in \{1, 2, \ldots, n\}$.* **LEMMA** Zerlegung in Haupträume

Beweis. Da $U_j = \ker\big((f - \lambda_j \operatorname{id}_V)^{\circ e_j}\big)$ ist, müssen wir $\ker\big((f - \lambda_j \operatorname{id}_V)^{\circ m}\big) = U_j$ für $m \geq e_j$ zeigen. Wir können Satz 26.2 aber auch mit m statt e_j anwenden und erhalten die Zerlegung

$$V = U_1 \oplus \ldots \oplus U_{j-1} \oplus U'_j \oplus U_{j+1} \oplus \ldots \oplus U_n$$

mit U_i wie vorher für $i \neq j$ und $U_j' = \ker\big((f - \lambda_j \operatorname{id}_V)^{\circ m}\big)$. Da $U_j \subset U_j'$ und da wegen der Dimensionsformel für direkte Summen auch $\dim U_j = \dim U_j'$ gilt, folgt wie gewünscht $U_j' = U_j$. ❑

Indem man einen beliebigen Endomorphismus f eines endlich-dimensionalen Vektorraums auf die Summe seiner Haupträume einschränkt und auf die Einschränkung Satz 26.2 und Lemma 26.4 anwendet, sieht man, dass die Summe der Haupträume stets direkt ist.

Da V endlich-dimensional ist, kann die aufsteigende Kette von Untervektorräumen $\ker\big((f - \lambda \operatorname{id}_V)^{\circ m}\big)$ nicht unendlich oft echt aufsteigen, also muss es ein $m \in \mathbb{N}$ geben mit

$$\begin{aligned}\ker\big((f - \lambda \operatorname{id}_V)^{\circ m}\big) &= \ker\big((f - \lambda \operatorname{id}_V)^{\circ (m+1)}\big) = \ker\big((f - \lambda \operatorname{id}_V)^{\circ (m+2)}\big) = \ldots \\ &= H_\lambda(f)\,.\end{aligned}$$

(In der Situation von Satz 26.2 ist $m \le e_j$.) Sei $g = f|_{H_\lambda(f)} - \lambda \operatorname{id}_{H_\lambda(f)}$. Dann gilt $f|_{H_\lambda(f)} = \lambda \operatorname{id}_{H_\lambda(f)} + g$ und $g^{\circ m} = \mathbf{0}$. Diese (zweite) Eigenschaft von g hat einen Namen.

DEF ∗
nilpotent

26.5. **Definition.** Sei f ein Endomorphismus eines Vektorraums V. f heißt *nilpotent*, wenn es $m \in \mathbb{N}$ gibt mit $f^{\circ m} = \mathbf{0}$. Analog heißt eine Matrix $A \in \operatorname{Mat}(n, K)$ *nilpotent*, wenn es $m \in \mathbb{N}$ gibt mit $A^m = \mathbf{0}$. ♢

Noch allgemeiner heißt ein Element r eines Rings R *nilpotent*, wenn es $m \in \mathbb{N}$ gibt mit $r^m = 0$.

LEMMA

26.6. **Lemma.** *Sei $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$ eine Zerlegung des Vektorraums V als direkte Summe und seien für $j \in \{1, 2, \ldots, n\}$ jeweils $f_j, g_j \in \operatorname{End}(U_j)$. Seien weiter $f = f_1 \oplus f_2 \oplus \ldots \oplus f_n$ und $g = g_1 \oplus g_2 \oplus \ldots \oplus g_n$.*

(1) *Es gilt $f + g = (f_1 + g_1) \oplus (f_2 + g_2) \oplus \ldots \oplus (f_n + g_n)$.*
(2) *Es gilt $f \circ g = (f_1 \circ g_1) \oplus (f_2 \circ g_2) \oplus \ldots \oplus (f_n \circ g_n)$.*
(3) *Sind alle g_j nilpotent, so ist auch g nilpotent.*
(4) *Gilt $f_j \circ g_j = g_j \circ f_j$ für alle $j \in \{1, 2, \ldots, n\}$, dann gilt auch $f \circ g = g \circ f$.*
(5) *$\ker(f) = \ker(f_1) \oplus \ker(f_2) \oplus \ldots \oplus \ker(f_n)$.*

Beweis.

(1) Für $u_j \in U_j$ gilt $(f+g)(u_j) = f(u_j)+g(u_j) = f_j(u_j)+g_j(u_j) = (f_j+g_j)(u_j)$, also ist $f + g$ die direkte Summe der $f_j + g_j$.
(2) Für $u_j \in U_j$ gilt
$$(f \circ g)(u_j) = f\big(g(u_j)\big) = f\big(g_j(u_j)\big) = f_j\big(g_j(u_j)\big) = (f_j \circ g_j)(u_j)\,,$$
also ist $f \circ g$ die direkte Summe der $f_j \circ g_j$.
(3) Nach Voraussetzung gibt es zu jedem j ein $m_j \in \mathbb{N}$ mit $g_j^{\circ m_j} = \mathbf{0}$. Wir setzen $m = \max\{m_1, m_2, \ldots, m_n\}$, dann gilt $g_j^{\circ m} = \mathbf{0}$ für alle j. Damit ist $g^{\circ m}(u_j) = g_j^{\circ m}(u_j) = \mathbf{0}$ für $u_j \in U_j$ und es folgt für $v = u_1+u_2+\ldots+u_n \in V$, dass $g^{\circ m}(v) = \mathbf{0}$ ist. (Beachte $g^{\circ m} = g_1^{\circ m} \oplus \ldots \oplus g_n^{\circ m}$ nach Teil (2).) Also ist $g^{\circ m} = \mathbf{0}$ und g ist nilpotent.
(4) Das folgt aus Teil (2).

(5) Zunächst einmal ist klar, dass die Summe der Kerne rechts direkt ist: Sind $u_1 \in \ker(f_1) \subset U_1$, $u_2 \in \ker(f_2) \subset U_2$, $\ldots$, $u_n \in \ker(f_n) \subset U_n$ mit $u_1 + u_2 + \ldots + u_n = \mathbf{0}$, dann folgt $u_1 = u_2 = \ldots = u_n = \mathbf{0}$, weil die Summe der U_i direkt ist. Es ist also nur zu zeigen, dass

$$\ker(f) = \ker(f_1) + \ker(f_2) + \ldots + \ker(f_n)$$

ist. Die Inklusion „$\supset$“ folgt aus $\ker(f_j) \subset \ker(f)$ für alle $j \in \{1, 2, \ldots, n\}$ (denn für $u_j \in U_j$ gilt $f(u_j) = f_j(u_j)$), damit enthält der Untervektorraum $\ker(f)$ auch die Summe der $\ker(f_j)$. Für die Inklusion „$\subset$“ sei jetzt $v \in \ker(f)$. Wir können v (eindeutig) schreiben als $v = u_1 + u_2 + \ldots + u_n$ mit $u_j \in U_j$, dann ist

$$\begin{aligned}\mathbf{0} &= f(v) = f(u_1 + u_2 + \ldots + u_n) \\ &= f(u_1) + f(u_2) + \ldots + f(u_n) = f_1(u_1) + f_2(u_2) + \ldots + f_n(u_n)\,.\end{aligned}$$

Aus $f_j(u_j) \in U_j$ und daraus, dass die Summe der U_j direkt ist, folgt dann $f_1(u_1) = f_2(u_2) = \ldots = f_n(u_n) = \mathbf{0}$, also $u_j \in \ker(f_j)$ für alle j. Das bedeutet $v \in \ker(f_1) + \ker(f_2) + \ldots + \ker(f_n)$. ❑

26.7. **Beispiel.** Wenn V unendlich-dimensional ist, dann braucht die aufsteigende Kette der Kerne nicht „stationär“ zu werden. Sei zum Beispiel $V = K[X]$ der Polynomring und $f \in \mathrm{End}(V)$ die „Division durch X ohne Rest“, gegeben durch $f(p) = (p - p(0))/X$. Dann gilt $\ker(f^{\circ m}) = K[X]_{<m} = \langle 1, X, X^2, \ldots, X^{m-1}\rangle$, also werden diese Kerne immer größer. ♣

BSP unendlich aufsteigende Kerne

Wir können jetzt eine erste Version des Satzes von der Jordan-Normalform formulieren.

* 26.8. **Satz.** *Sei V ein endlich-dimensionaler K-Vektorraum und sei $f \in \mathrm{End}(V)$ ein Endomorphismus, sodass das charakteristische Polynom $\chi_f \in K[X]$ in Linearfaktoren zerfällt. Dann gibt es $d, g \in \mathrm{End}(V)$ mit $f = d + g$, $d \circ g = g \circ d$, d diagonalisierbar und g nilpotent.*

SATZ Jordansche Normalform (schwach)

Die analoge Aussage gilt für Matrizen $A \in \mathrm{Mat}(n, K)$: Zerfällt χ_A in Linearfaktoren, dann gibt es Matrizen $D, N \in \mathrm{Mat}(n, K)$ mit $A = D + N$, $DN = ND$, D diagonalisierbar und N nilpotent.

Beweis. Wir beweisen die Version für Endomorphismen; der Beweis für Matrizen ist analog. Sei

$$\chi_f = (X - \lambda_1)^{e_1}(X - \lambda_2)^{e_2} \cdots (X - \lambda_m)^{e_m}$$

mit $\lambda_1, \ldots, \lambda_m \in K$ paarweise verschieden und $e_1, \ldots, e_m \in \mathbb{Z}_{\geq 0}$. Nach Satz 16.25 gilt $\chi_f(f) = \mathbf{0}$. Satz 26.2 und Lemma 26.4 liefern dann f-invariante Untervektorräume $U_j = \ker\big((f - \lambda_j \,\mathrm{id}_V)^{\circ e_j}\big) = H_{\lambda_j}(f)$ und eine direkte Summenzerlegung $V = U_1 \oplus U_2 \oplus \ldots \oplus U_m$. Die Einschränkung f_j von f auf U_j (als Endomorphismus von U_j) hat also die Form $f_j = d_j + g_j$ mit $d_j = \lambda_j \,\mathrm{id}_{U_j}$ und $g_j^{\circ e_j} = \mathbf{0}$; es gilt $d_j \circ g_j = \lambda_j g_j = g_j \circ d_j$. Mit $d = d_1 \oplus d_2 \oplus \ldots \oplus d_m$ und $g = g_1 \oplus g_2 \oplus \ldots \oplus g_m$ gilt dann $f = d + g$, und nach Lemma 26.6 gilt $d \circ g = g \circ d$ und g ist nilpotent. Schließlich ist d diagonalisierbar, weil $U_j = E_{\lambda_j}(d)$ ist; damit ist V die direkte Summe der Eigenräume von d. ❑

Man kann auch zeigen, dass d und g eindeutig bestimmt sind: Man überlegt sich zunächst, dass die λ-Eigenräume von d genau die λ-Haupträume von f sind; daraus folgt leicht, dass d (und damit dann auch g) eindeutig bestimmt ist (Übung).

Um aus Satz 26.8 eine stärkere Version abzuleiten, die eine Normalform für Matrizen ähnlich den Diagonalmatrizen ergibt, müssen wir uns die Struktur von nilpotenten Endomorphismen noch genauer ansehen. Sei also $g \in \operatorname{End}(V)$ nilpotent und sei $m \in \mathbb{N}$ die kleinste natürliche Zahl mit $g^{\circ m} = \mathbf{0}$. Dann haben wir die Kette

$$\{\mathbf{0}\} \subsetneq \ker(g) \subsetneq \ker(g^{\circ 2}) \subsetneq \ldots \subsetneq \ker(g^{\circ(m-1)}) \subsetneq \ker(g^{\circ m}) = V$$

von Untervektorräumen von V. Dass die Inklusionen echt sind, ergibt sich aus dem folgenden Lemma.

LEMMA

26.9. **Lemma.** *Seien V ein Vektorraum, $g \in \operatorname{End}(V)$ und $m \in \mathbb{N}$. Dann folgt aus $\ker(g^{\circ(m+1)}) = \ker(g^{\circ m})$, dass $\ker(g^{\circ n}) = \ker(g^{\circ m})$ ist für alle $n \geq m$.*

Beweis. Übung. ❑

Der Begriff „Dreiecksmatrix“ kam schon vor; wir definieren ihn noch „offiziell“:

DEF ∗
Dreiecksmatrix

26.10. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$ und $A = (a_{ij}) \in \operatorname{Mat}(n, K)$. Dann heißt A eine *obere Dreiecksmatrix*, wenn $a_{ij} = 0$ ist für alle $1 \leq j < i \leq n$, und eine *strikte obere Dreiecksmatrix*, wenn $a_{ij} = 0$ ist für alle $1 \leq j \leq i \leq n$. Analog definiert man *(strikte) untere Dreiecksmatrizen* (mit $j > i$ bzw. $j \geq i$). ♢

In einer oberen Dreiecksmatrix sind also alle Einträge echt unterhalb der Diagonalen null, in einer strikten oberen Dreiecksmatrix sind zusätzlich die Einträge auf der Diagonalen null.

SATZ
Trigonalisierung nilpotenter Endomorphismen

26.11. **Satz.** *Seien V ein endlich-dimensionaler K-Vektorraum und $g \in End(V)$ nilpotent. Dann gibt es eine Basis B von V, sodass $\operatorname{Mat}_B(g)$ eine strikte obere Dreiecksmatrix ist.*

Analog gilt: Ist $A \in \operatorname{Mat}(n, K)$ nilpotent, dann ist A zu einer strikten oberen Dreiecksmatrix ähnlich.

Umgekehrt ist jede strikte obere Dreiecksmatrix $A \in \operatorname{Mat}(n, K)$ nilpotent; genauer gilt $A^n = \mathbf{0}$.

Beweis. Sei $U_j = \ker(g^{\circ j})$, dann gilt $\{\mathbf{0}\} = U_0 \subsetneq U_1 \subsetneq \ldots \subsetneq U_m = V$, wobei m minimal ist mit $g^{\circ m} = \mathbf{0}$. Außerdem ist $g(U_j) \subset U_{j-1}$ für alle $j > 0$ (denn aus $g^{\circ j}(v) = \mathbf{0}$ folgt $g^{\circ(j-1)}(g(v)) = \mathbf{0}$). Wir wählen eine Basis von U_1, die wir sukzessive zu Basen von U_2, U_3, …, U_m erweitern; sei $B = (b_1, b_2, \ldots, b_n)$ die resultierende Basis von V. Dann ist $(b_1, b_2, \ldots, b_{\dim U_j})$ eine Basis von U_j. Sei $\dim U_{j-1} < k \leq \dim(U_j)$; dann ist $g(b_k) \in U_{j-1}$ und damit eine Linearkombination von Basiselementen b_i mit $i \leq \dim U_{j-1} < k$. Da es zu jedem $1 \leq k \leq n$ ein passendes j gibt, gilt $g(b_k) \in \langle b_1, \ldots, b_{k-1} \rangle$ für alle $k \in \{1, \ldots, n\}$. Die k-te Spalte von $A = \operatorname{Mat}_B(g)$ hat also höchstens in den ersten $k - 1$ Positionen von oben Einträge ungleich null; das bedeutet gerade, dass A eine strikte obere Dreiecksmatrix ist.

Die Aussage über nilpotente Matrizen folgt aus der Aussage über Endomorphismen, indem man A als Endomorphismus von K^n betrachtet.

Die letzte Behauptung folgt aus Satz 16.25, denn für eine strikte Dreiecksmatrix $A \in \operatorname{Mat}(n, K)$ gilt $\chi_A = X^n$. ❑

* **26.12. Folgerung.** *Seien V ein endlich-dimensionaler K-Vektorraum und f ein Endomorphismus von V, sodass das charakteristische Polynom $\chi_f \in K[X]$ in Linearfaktoren zerfällt. Dann gibt es eine Basis B von V, sodass $\mathrm{Mat}_B(f)$ eine obere Dreiecksmatrix ist.* **FOLG** Trigonalisierung von Endomorphismen

Analog gilt: Ist $A \in \mathrm{Mat}(n, K)$ eine Matrix, sodass χ_A in Linearfaktoren zerfällt, dann ist A ähnlich zu einer oberen Dreiecksmatrix.

Beweis. Wie im Beweis von Satz 26.8 haben wir eine Zerlegung

$$V = U_1 \oplus U_2 \oplus \ldots \oplus U_m \qquad \text{mit } U_j = H_{\lambda_j}(f)$$

in die Haupträume von f, sodass f auf U_j die Form $\lambda_j \,\mathrm{id} + g_j$ hat mit g_j nilpotent. Wir wählen Basen B_j von U_j, sodass $\mathrm{Mat}_{B_j}(g_j)$ eine strikte obere Dreiecksmatrix ist (das ist möglich nach Satz 26.11). Die Matrix bezüglich B_j der Einschränkung von f auf U_j ist dann $M_j = \lambda_j I_{\dim U_j} + \mathrm{Mat}_{B_j}(g_j)$; dies ist eine obere Dreiecksmatrix. Wir setzen die Basis B von V aus den Basen B_j zusammen. Nach Lemma 17.23 ist dann $\mathrm{Mat}_B(f)$ eine Block-Diagonalmatrix mit Blöcken M_j; da die M_j obere Dreiecksmatrizen sind, gilt das auch für $\mathrm{Mat}_B(f)$. Die Aussage für Matrizen folgt in der üblichen Weise. ❑

Die Voraussetzung, dass das charakteristische Polynom in Linearfaktoren zerfällt, ist stets erfüllt, wenn der Körper K algebraisch abgeschlossen ist, wie zum Beispiel $K = \mathbb{C}$.

Da die Determinante einer Dreiecksmatrix das Produkt der Diagonalelemente ist, sehen wir, dass wir die Relation

$$\chi_f = (X - \lambda_1)^{\dim H_{\lambda_1}(f)} (X - \lambda_2)^{\dim H_{\lambda_2}(f)} \cdots (X - \lambda_m)^{\dim H_{\lambda_m}(f)}$$

haben. (Insbesondere zerfällt das charakteristische Polynom einer Dreiecksmatrix in Linearfaktoren; die Voraussetzung an χ_f bzw. χ_A in Folgerung 26.12 ist also notwendig.) Das bedeutet:

26.13. Lemma. *Seien V ein endlich-dimensionaler K-Vektorraum und f ein Endomorphismus von V, sei weiter $\lambda \in K$. Dann ist die Dimension von $H_\lambda(f)$ gleich der algebraischen Vielfachheit von λ als Eigenwert von f (also gleich der Vielfachheit von λ als Nullstelle von χ_f).* **LEMMA** Dimension des Hauptraums

Streng genommen haben wir diese Aussage nur in dem Fall bewiesen, dass das charakteristische Polynom in Linearfaktoren zerfällt. Durch Betrachtung über einem größeren Körper (z.B. $\mathbb{C}$ statt $\mathbb{Q}$ oder $\mathbb{R}$) kann man das aber immer erreichen.

Da der Eigenraum $E_\lambda(f)$ stets im Hauptraum $H_\lambda(f)$ enthalten ist, liefert dies auch einen weiteren Beweis der Aussage, dass die geometrische Vielfachheit eines Eigenwerts nicht größer als die algebraische Vielfachheit sein kann.

Für viele Anwendungen sind die Aussagen von Satz 26.8 oder Folgerung 26.12 ausreichend. Manchmal möchte man aber eine im Wesentlichen eindeutige Normalform von Matrizen bis auf Ähnlichkeit haben. Dazu betrachten wir noch einmal nilpotente Endomorphismen und daran angepasste Basen.

Zur Vorbereitung formulieren wir ein Lemma zur Struktur von nilpotenten Endomorphismen. Darin leisten wir die Hauptarbeit für den Beweis der (starken) Jordanschen Normalform.

26.14. **Lemma.** *Sei $V \neq \{\mathbf{0}\}$ ein endlich-dimensionaler K-Vektorraum und sei $f \in \operatorname{End}(V)$ nilpotent mit $f^{\circ m} = \mathbf{0}$ und $m \in \mathbb{N}$ minimal mit dieser Eigenschaft. Sei weiter $v \in V$ mit $f^{\circ(m-1)}(v) \neq \mathbf{0}$ und $U = \langle v, f(v), f^{\circ 2}(v), \ldots, f^{\circ(m-1)}(v)\rangle \subset V$. Dann gilt:*

LEMMA Struktur nilpotenter Endomorphismen

(1) *$\dim U = m$ (d.h., $\big(v, f(v), f^{\circ 2}(v), \ldots, f^{\circ(m-1)}(v)\big)$ ist linear unabhängig).*

(2) *Es gibt ein f-invariantes Komplement U' von U in V.*

Beweis.

(1) Wir zeigen, dass $\big(v, f(v), f^{\circ 2}(v), \ldots, f^{\circ(m-1)}(v)\big)$ linear unabhängig ist. Seien dazu $\lambda_0, \lambda_1, \ldots, \lambda_{m-1}$ Skalare mit

$$\lambda_0 v + \lambda_1 f(v) + \lambda_2 f^{\circ 2}(v) + \ldots + \lambda_{m-1} f^{\circ(m-1)}(v) = \mathbf{0}\,.$$

Wenn wir $f^{\circ(m-1)}$ auf diese Gleichung anwenden und beachten, dass $f^{\circ m} = \mathbf{0}$ ist, dann erhalten wir $\lambda_0 f^{\circ(m-1)}(v) = \mathbf{0}$, wegen $f^{\circ(m-1)}(v) \neq \mathbf{0}$ also $\lambda_0 = 0$. Durch Anwenden von $f^{\circ(m-2)}$ bekommen wir dann analog $\lambda_1 = 0$, und in der gleichen Art dann nacheinander $\lambda_2 = 0$, $\ldots$, $\lambda_{m-1} = 0$. Also sind die Vektoren linear unabhängig.

(2) Wir zeigen zunächst die folgende Aussage:

Sei $\tilde{U} \subset V$ ein f-invarianter Untervektorraum mit $U \cap \tilde{U} = \{\mathbf{0}\}$ und $U + \tilde{U} \neq V$. Dann gibt es einen Vektor $w \in V \setminus (U + \tilde{U})$ mit $f(w) \in \tilde{U}$.

Nach Annahme gibt es einen Vektor $w' \in V \setminus (U + \tilde{U})$. Sei $k \in \mathbb{Z}_{\geq 0}$ minimal mit $f^{\circ k}(w') \in U + \tilde{U}$. Wegen $w' \notin U + \tilde{U}$ ist $k \geq 1$ und wegen $f^{\circ m} = \mathbf{0}$ existiert so ein k. Sei $w'' = f^{\circ(k-1)}(w')$. Dann gilt $w'' \notin U + \tilde{U}$, aber $f(w'') \in U + \tilde{U}$. Wir können also schreiben

$$f(w'') = \lambda_0 v + \lambda_1 f(v) + \ldots + \lambda_{m-1} f^{\circ(m-1)}(v) + \tilde{u}$$

mit $\lambda_0, \ldots, \lambda_{m-1} \in K$ und $\tilde{u} \in \tilde{U}$. Es ist wegen $f^{\circ m} = \mathbf{0}$

$$\mathbf{0} = f^{\circ m}(w'') = \lambda_0 \underbrace{f^{\circ(m-1)}(v)}_{\neq \mathbf{0}} + \underbrace{f^{\circ(m-1)}(\tilde{u})}_{\in \tilde{U}};$$

weil die Summe $U + \tilde{U}$ nach Voraussetzung direkt ist, folgt daraus $\lambda_0 = 0$. Wir setzen nun $w = w'' - \lambda_1 v - \ldots - \lambda_{m-1} f^{\circ(m-2)}(v)$; dann ist $f(w) = \tilde{u} \in \tilde{U}$ wie gewünscht.

Sei $\tilde{U}' = \tilde{U} + \langle w\rangle$. Dann ist $\tilde{U}'$ ein f-invarianter Untervektorraum von V, denn $f(\tilde{U}') = f(\tilde{U}) + \langle f(w)\rangle \subset \tilde{U} \subset \tilde{U}'$. Die Summe $U + \tilde{U} + \langle w\rangle$ ist direkt: Seien $u \in U$, $\tilde{u} \in \tilde{U}$ und $\lambda \in K$ mit $u + \tilde{u} + \lambda w = \mathbf{0}$; dann folgt wegen $w \notin U + \tilde{U}$, dass $\lambda = 0$ ist, und wegen $U \cap \tilde{U} = \{\mathbf{0}\}$ folgt dann auch $u = \tilde{u} = \mathbf{0}$. Das impliziert $U \cap \tilde{U}' = \{\mathbf{0}\}$ und $\dim \tilde{U}' = \dim \tilde{U} + 1$.

Wir setzen nun $\tilde{U}_0 = \{\mathbf{0}\}$. Sei für $n \in \mathbb{Z}_{\geq 0}$ der f-invariante Untervektorraum $\tilde{U}_n$ von V mit $U \cap \tilde{U}_n = \{\mathbf{0}\}$ bereits konstruiert. Gilt $U + \tilde{U}_n = V$, dann ist $U' = \tilde{U}_n$ ein Untervektorraum mit den gewünschten Eigenschaften. Anderenfalls setzen wir $\tilde{U}_{n+1} = \tilde{U}'_n$ wie im vorigen Absatz; dann ist $\tilde{U}_{n+1}$ ebenfalls f-invariant und es gilt $U \cap \tilde{U}_{n+1} = \{\mathbf{0}\}$. Da $\dim \tilde{U}_n = n$ ist, muss die Konstruktion mit $n = \dim V - m$ abbrechen. ❑

DEF nilzyklisch

26.15. **Definition.** Sei V ein endlich-dimensionaler K-Vektorraum. Ein Endomorphismus f von V heißt *nilzyklisch*, wenn f nilpotent ist und es $v \in V$ gibt, sodass $V = \langle v, f(v), f^{\circ 2}(v), \ldots \rangle$ ist.

Nach Lemma 26.14 ist $\bigl(v, f(v), \ldots, f^{\circ(m-1)}(v)\bigr)$ linear unabhängig und damit eine Basis von V, wobei $m \in \mathbb{N}$ die kleinste Zahl ist mit $f^{\circ m} = \mathbf{0}$. Wir ordnen um und setzen $B = \bigl(f^{\circ(m-1)}(v), \ldots, f(v), v\bigr)$; dann ist

$$\operatorname{Mat}_B(f) = J_m = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \in \operatorname{Mat}(m, K)\,.$$

Die Matrizen J_m heißen ebenfalls *nilzyklisch*. ♢

SATZ Struktur nilpotenter Endomorphismen

26.16. **Satz.** *Sei f ein nilpotenter Endomorphismus eines endlich-dimensionalen Vektorraums V. Dann gibt es eine Zerlegung $V = U_1 \oplus U_2 \oplus \ldots \oplus U_m$ in f-invariante Untervektorräume, sodass $f|_{U_j}$ nilzyklisch ist für alle $j \in \{1, 2, \ldots, m\}$.*

Beweis. Induktion über $\dim V$. Im Fall $\dim V = 0$, also $V = \{\mathbf{0}\}$, ist nichts zu zeigen ($m = 0$). Sei also $\dim V > 0$. Dann gibt es nach Lemma 26.14 eine Zerlegung $V = U_1 \oplus V'$ in f-invariante Untervektorräume mit $U_1 \neq \{\mathbf{0}\}$, sodass $f|_{U_1}$ nilzyklisch ist (U_1 ist U in Lemma 26.14, V' ist U'). Nach Induktionsannahme (beachte $\dim V' = \dim V - \dim U_1 < \dim V$) gibt es eine Zerlegung $V' = U_2 \oplus \ldots \oplus U_m$ in f-invariante Untervektorräume, sodass $f|_{U_j}$ nilzyklisch ist für alle $j \in \{2, \ldots, m\}$. Insgesamt erhalten wir die gewünschte Zerlegung von V. ❑

FOLG Normalform für nilpotente Endomorphismen

26.17. **Folgerung.** *Sei V ein endlich-dimensionaler Vektorraum und sei f ein nilpotenter Endomorphismus von V. Dann gibt es eine Basis B von V und Zahlen $m_1, m_2, \ldots, m_k \in \mathbb{Z}_{>0}$ mit $m_1 + m_2 + \ldots + m_k = \dim V$, sodass*

$$\operatorname{Mat}_B(f) = \left(\begin{array}{c|c|c|c} J_{m_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & J_{m_2} & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \cdots & J_{m_k} \end{array}\right)$$

ist. Die Zahlen $m_1, m_2, \ldots, m_k$ sind bis auf ihre Reihenfolge eindeutig bestimmt.

Analog gilt: Jede nilpotente Matrix $A \in \operatorname{Mat}(n, K)$ ist zu einer Matrix der obigen Form ähnlich. Diese Matrix ist bis auf die Reihenfolge der J_{m_i} eindeutig bestimmt.

Beweis. Nach Satz 26.16 gibt es eine Zerlegung

$$V = U_1 \oplus U_2 \oplus \ldots \oplus U_k$$

in f-invariante Untervektorräume mit $f|_{U_i}$ nilzyklisch. Sei B_i für $i \in \{1, 2, \ldots, k\}$ eine Basis von U_i wie in Definition 26.15; sei B die durch Aneinanderhängen von $B_1, B_2, \ldots, B_k$ gegebene Basis von V. Die erste Behauptung folgt dann mit Lemma 17.23; dabei ist $m_i = \dim U_i$.

Zur Eindeutigkeit: Für $j \in \mathbb{N}$ gilt

$$\dim \ker(f^{\circ j}) = \sum_{i=1}^{k} \dim \ker(f|_{U_i}^{\circ j}) = \sum_{i=1}^{k} \min\{j, m_i\}$$

(die erste Gleichung folgt aus $\ker(f^{\circ j}) = \bigoplus_{i=1}^{k} \ker(f|_{U_i}^{\circ j})$, siehe Lemma 26.6(5), die zweite gilt, weil $\ker(f|_{U_i}^{\circ j}) = \langle f^{\circ(m_i-j)}(v_i), f^{\circ(m_i-j+1)}(v_i), \ldots, f^{\circ(m_i-1)}(v_i)\rangle$ ist für $j \le m_i$ und $\ker(f|_{U_i}^{\circ j}) = U_i$ für $j \ge m_i$, wenn $U_i = \langle v_i, f(v_i), \ldots, f^{\circ(m_i-1)}(v_i)\rangle$ ist) und damit

$$\begin{aligned}
\dim \ker(f^{\circ(j+1)}) - \dim \ker(f^{\circ j}) &= \sum_{i=1}^{k} \bigl(\min\{j+1, m_i\} - \min\{j, m_i\}\bigr) \\
&= \sum_{i=1}^{k} \left\{ \begin{array}{ll} 1, & \text{falls } m_i > j \\ 0, & \text{falls } m_i \le j \end{array} \right\} \\
&= \#\bigl\{i \in \{1, 2, \ldots, k\} \mid m_i > j\bigr\}\,,
\end{aligned}$$

also auch (für $j \ge 1$)

$$\begin{aligned}
&\bigl(\dim \ker(f^{\circ j}) - \dim \ker(f^{\circ(j-1)})\bigr) - \bigl(\dim \ker(f^{\circ(j+1)}) - \dim \ker(f^{\circ j})\bigr) \\
&\qquad = \#\bigl\{i \in \{1, 2, \ldots, k\} \mid m_i \ge j\bigr\} - \#\bigl\{i \in \{1, 2, \ldots, k\} \mid m_i > j\bigr\} \\
&\qquad = \#\bigl\{i \in \{1, 2, \ldots, k\} \mid m_i = j\bigr\}\,;
\end{aligned}$$

damit sind die Zahlen m_i bis auf ihre Reihenfolge eindeutig festgelegt.

Die Aussage für Matrizen folgt in der üblichen Weise. ❑

Man kann sich also einen nilpotenten Endomorphismus f wie in der folgenden Skizze vorstellen:

$$\begin{array}{lccccccccccc}
U_1\colon & \bullet & \xmapsto{f} & \bullet & \xmapsto{f} & \cdots & \xmapsto{f} & \bullet & \xmapsto{f} & \bullet & \xmapsto{f} & \mathbf{0} \\
U_2\colon & \bullet & \xmapsto{f} & \bullet & \xmapsto{f} & \cdots & \xmapsto{f} & \bullet & \xmapsto{f} & \bullet & \xmapsto{f} & \mathbf{0} \\
U_3\colon & & & \bullet & \xmapsto{f} & \cdots & \xmapsto{f} & \bullet & \xmapsto{f} & \bullet & \xmapsto{f} & \mathbf{0} \\
\vdots & & & & & & & \vdots & & & & \vdots \\
U_k\colon & & & & & & & \cdots & \xmapsto{f} & \bullet & \xmapsto{f} & \mathbf{0}
\end{array}$$

Die Punkte stehen dabei für die Basiselemente. Der Kern von $f^{\circ j}$ wird dann erzeugt von den Basiselementen in den letzten j Spalten (dabei zählt die Nullspalte am Ende nicht mit). Daraus kann man Folgendes ablesen:

Sei m die kleinste Zahl mit $f^{\circ m} = \mathbf{0}$. Für $j \in \{1, 2, \ldots, m\}$ sei V_j ein Komplement von $\ker(f^{\circ(j-1)}) + f\bigl(\ker(f^{\circ(j+1)})\bigr)$ in $\ker(f^{\circ j})$. Sei B_j eine Basis von V_j. Für $b \in B_j$ ist dann $B'_b = (f^{\circ(j-1)}(b), \ldots, f(b), b)$ eine Basis des f-invarianten Untervektorraums $\langle b, f(b), \ldots\rangle$; Hintereinanderhängen dieser Basen B'_b für alle $b \in B_1 \cup B_2 \cup \ldots \cup B_m$ liefert eine Basis B von V, bezüglich derer f durch eine Matrix wie in Folgerung 26.17 gegeben ist: Die Elemente von B_j entsprechen den Punkten im Diagramm in der $j$ten Spalte von rechts (außer der Nullspalte), an denen eine Kette $\bullet \mapsto \bullet \mapsto \ldots$ beginnt.

BSP Normalform für nilpotenten Endomorphismus

26.18. **Beispiel.** Sei V ein Vektorraum mit Basis $(v_1, v_2, \ldots, v_7)$ und $f \in \mathrm{End}(V)$ gegeben durch

$$\begin{tikzcd}
v_1 \arrow[dr, mapsto, "f"] & & & \\
& v_5 \arrow[ddr, mapsto, "f"] & & \\
v_2 \arrow[ur, mapsto, "f"'] & & & \\
& & v_7 \arrow[r, mapsto, "f"] & \mathbf{0} \\
v_3 \arrow[dr, mapsto, "f"] & & & \\
& v_6 \arrow[uur, mapsto, "f"'] & & \\
v_4 \arrow[ur, mapsto, "f"'] & & &
\end{tikzcd}$$

Dann ist $f^{\circ 3} = \mathbf{0}$ (und 3 ist die kleinste Zahl m mit $f^{\circ m} = \mathbf{0}$), also ist f nilpotent. Wir finden eine Basis von V wie in Folgerung 26.17. Dazu bestimmen wir erst einmal die „höheren Kerne“ $K_j = \ker(f^{\circ j})$ für $j = 0, 1, 2, 3, \ldots$. Offenbar ist $K_0 = \{\mathbf{0}\}$. Aus

$$f(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_7 v_7) = (\lambda_1 + \lambda_2)v_5 + (\lambda_3 + \lambda_4)v_6 + (\lambda_5 + \lambda_6)v_7$$

folgt

$$K_1 = \langle v_1 - v_2, v_3 - v_4, v_5 - v_6, v_7 \rangle .$$

Weiter ist

$$f^{\circ 2}(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_7 v_7) = (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)v_7$$

und damit

$$K_2 = \langle v_1 - v_2, v_3 - v_4, v_1 - v_3, v_5, v_6, v_7 \rangle .$$

Schließlich ist $K_j = V$ für $j \geq 3$. Wir müssen Komplemente V_j von $K_{j-1} + f(K_{j+1})$ in K_j wählen. Für $j = 1$ ist

$$K_0 + f(K_2) = f(K_2) = \langle v_5 - v_6, v_7 \rangle ;$$

ein Komplement in K_1 ist zum Beispiel gegeben durch

$$V_1 = \langle v_1 - v_2, v_3 - v_4 \rangle .$$

Für $j = 2$ ist

$$K_1 + f(K_3) = \langle v_1 - v_2, v_3 - v_4, v_5, v_6, v_7 \rangle ;$$

ein Komplement in K_2 ist etwa

$$V_2 = \langle v_1 - v_3 \rangle .$$

Für $j = 3$ schließlich ist

$$K_2 + f(K_4) = K_2 + f(V) = K_2 ;$$

ein Komplement ist zum Beispiel

$$V_3 = \langle v_1 \rangle .$$

Nach dem oben beschriebenen Rezept können wir als Basis wählen:

$$\begin{aligned}
b_7 &= v_1 - v_2 && \text{mit } f(b_7) = \mathbf{0} \\
b_6 &= v_3 - v_4 && \text{mit } f(b_6) = \mathbf{0} \\
b_5 &= v_1 - v_3 && \\
b_4 &= f(b_5) = v_5 - v_6 && \text{mit } f(b_4) = \mathbf{0} \\
b_3 &= v_1 && \\
b_2 &= f(b_3) = v_5 && \\
b_1 &= f(b_2) = v_7 && \text{mit } f(b_1) = \mathbf{0}
\end{aligned}$$

Mit $B = (b_1, b_2, b_3, b_4, b_5, b_6, b_7)$ ist dann

$$\mathrm{Mat}_B(f) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \left(\begin{array}{c|c|c|c} J_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & J_2 & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & J_1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & J_1 \end{array}\right)$$

Das zugehörige Diagramm sieht so aus:

$$\begin{array}{rcl} b_3 \stackrel{f}{\longmapsto} b_2 \stackrel{f}{\longmapsto} b_1 \stackrel{f}{\longmapsto} \mathbf{0} \\ b_5 \stackrel{f}{\longmapsto} b_4 \stackrel{f}{\longmapsto} \mathbf{0} \\ b_6 \stackrel{f}{\longmapsto} \mathbf{0} \\ b_7 \stackrel{f}{\longmapsto} \mathbf{0} \end{array}$$

Die Dimensionen von $K_0, K_1, K_2, K_3, \ldots$ sind $0, 4, 6, 7, 7, \ldots$, die Zuwächse der Dimensionen also $4, 2, 1, 0, 0, \ldots$ und die Differenzen der Zuwächse sind $2, 1, 1, 0, 0, \ldots$. Diese Zahlen geben die Häufigkeiten der nilzyklischen Kästchen $J_1, J_2, J_3, \ldots$ in der Matrix an, vergleiche den Beweis der Eindeutigkeitsaussage in Folgerung 26.17. ♣

DEF ∗
Jordan-Kästchen

26.19. **Definition.** Sei K ein Körper, seien $\lambda \in K$ und $m \in \mathbb{Z}_{>0}$. Die Matrix

$$J_m(\lambda) = \lambda I_m + J_m$$

heißt das *Jordan-Kästchen* oder der *Jordan-Block* der Größe m zum Eigenwert λ. ♢

SATZ ∗
Jordansche Normalform (stark)

26.20. **Satz.** *Seien V ein endlich-dimensionaler K-Vektorraum und f ein Endomorphismus von V mit in Linearfaktoren zerfallendem charakteristischem Polynom χ_f. Dann gibt es eine Zerlegung*

$$V = U_1 \oplus U_2 \oplus \ldots \oplus U_k$$

in f-invariante Untervektorräume, sodass $f|_{U_i} = \lambda_i \, \mathrm{id}_{U_i} + g_i$ ist mit g_i nilzyklisch, für alle $i \in \{1, 2, \ldots, k\}$. Insbesondere gibt es eine Basis B von V, sodass $\mathrm{Mat}_B(f)$ eine Block-Diagonalmatrix ist, deren Blöcke die Jordan-Kästchen $J_{\dim U_i}(\lambda_i)$ sind. Die Jordan-Kästchen sind bis auf ihre Reihenfolge eindeutig bestimmt.

Analog gilt: Ist $A \in \mathrm{Mat}(n, K)$ mit zerfallendem charakteristischem Polynom χ_A, dann ist A ähnlich zu einer Block-Diagonalmatrix, deren Blöcke Jordan-Kästchen sind, und die Jordan-Kästchen sind bis auf ihre Reihenfolge eindeutig bestimmt.

Beweis. Nach Satz 26.2 können wir V als direkte Summe in die verschiedenen Haupträume $H_\lambda(f)$ zerlegen; auf $H_\lambda(f)$ ist $f = \lambda \, \mathrm{id} + g$ mit g nilpotent. Nach Satz 26.16 gibt es eine Zerlegung von $H_\lambda(f)$ als direkte Summe von Untervektorräumen, auf denen g nilzyklisch ist, dort ist also $f = \lambda \, \mathrm{id} + \text{nilzyklisch}$. Wir erhalten die gewünschte Zerlegung von V, indem wir die Zerlegungen der Haupträume kombinieren. Wie in Definition 26.15 können wir Basen B_i der U_i so wählen, dass g_i auf U_i durch $J_{\dim U_i}$ gegeben ist, dann ist

$$\mathrm{Mat}_{B_i}(f|_{U_i}) = \lambda_i I_{\dim U_i} + J_{\dim U_i} = J_{\dim U_i}(\lambda_i)\,.$$

Setzen wir diese Basen zu einer Basis B von V zusammen, erhalten wir die Block-Diagonalmatrix für f wie angegeben. Die Eindeutigkeit folgt daraus, dass die

Summe der Größen der Jordan-Kästchen zu einem gegebenen Eigenwert λ gleich der algebraischen Vielfachheit von λ als Eigenwert von f sein muss (vergleiche Lemma 26.13), und aus der Eindeutigkeitsaussage in Folgerung 26.17.

Die Aussagen für Matrizen folgen wie üblich aus denen für Endomorphismen: Wir wenden die Aussage auf den Endomorphismus $f\colon K^n \to K^n$, $\boldsymbol{x} \mapsto A \cdot \boldsymbol{x}$, an (wobei die Elemente von K^n als Spaltenvektoren betrachtet werden). Dann ist $A = \mathrm{Mat}_{E,E}(f)$ mit der Standard-Basis E von K^n, und mit $P = \mathrm{Mat}_{B,E}(\mathrm{id}_{K^n})$ hat dann $\mathrm{Mat}_B(f) = P^{-1}AP$ die Form wie im Satz. ❑

* 26.21. **Definition.** Die Matrix in Satz 26.20 heißt die *Jordansche Normalform* von f bzw. A. ♢

DEF Jordansche Normalform

Die Jordansche Normalform liefert also eine vollständige Klassifikation der Matrizen mit zerfallendem charakteristischem Polynom bis auf Ähnlichkeit. Zum Beispiel gibt es genau drei Ähnlichkeitsklassen von Matrizen in $\mathrm{Mat}(3, K)$ mit charakteristischem Polynom $(X - \lambda)^3$, denn die Jordan-Normalform kann die Jordan-Kästchen $J_3(\lambda)$ oder $J_2(\lambda), J_1(\lambda)$ oder $J_1(\lambda), J_1(\lambda), J_1(\lambda)$ haben.

Wie kann man die Jordansche Normalform einer gegebenen Matrix $A \in \mathrm{Mat}(n, K)$ bestimmen?

Wenn man nur wissen möchte, wie die Jordansche Normalform aussieht, dann bestimmt man zuerst die Eigenwerte (indem man das charakteristische Polynom faktorisiert) und berechnet dann für jeden Eigenwert λ die Dimensionen der Kerne von $(A - \lambda I_n)^m$ für $m = 1, 2, 3, \ldots$. Aus diesen Dimensionen ergeben sich die Größen der vorkommenden Jordan-Kästchen $J_k(\lambda)$ wie im Beweis von Folgerung 26.17. Ist die algebraische Vielfachheit von λ höchstens 3, dann kann man die Jordan-Kästchen, die zu λ gehören, aus der algebraischen und der geometrischen Vielfachheit von λ bestimmen. Man überlegt sich nämlich leicht (Übung), dass jedes Kästchen $J_m(\lambda)$ genau 1 zur geometrischen Vielfachheit beiträgt: Die geometrische Vielfachheit von λ ist genau die Anzahl der Jordan-Kästchen $J_m(\lambda)$. Für algebraische Vielfachheit (also Gesamtgröße dieser Kästchen) ≤ 3 legt die Anzahl bereits die Größe der Kästchen fest.

Braucht man zusätzlich die Matrix $P \in \mathrm{GL}(n, K)$, sodass $P^{-1}AP$ in Jordan-Normalform ist, dann muss man für jeden Hauptraum $H_\lambda(A)$ eine Basis wie in Folgerung 26.17 bestimmen (zu $g = (f - \lambda\,\mathrm{id})|_{H_\lambda(A)}$) und diese Basen dann zu einer Basis von K^n zusammensetzen. Die Basiselemente bilden dann die Spalten von P.

Wir führen das in einem Beispiel durch:

26.22. **Beispiel.** Wir betrachten

BSP Jordansche Normalform

$$A = \begin{pmatrix} 5 & -1 & 0 & 0 & -10 \\ 4 & -1 & 0 & 0 & -8 \\ 1 & 2 & 1 & -2 & -1 \\ 5 & 0 & 0 & -1 & -9 \\ 2 & -1 & 0 & 0 & -3 \end{pmatrix} \in \mathrm{Mat}(5, \mathbb{R})\,.$$

Das charakteristische Polynom ist

$$\begin{aligned}
\chi_A &= \begin{vmatrix} X-5 & 1 & 0 & 0 & 10 \\ -4 & X+1 & 0 & 0 & 8 \\ -1 & -2 & X-1 & 2 & 1 \\ -5 & 0 & 0 & X+1 & 9 \\ -2 & 1 & 0 & 0 & X+3 \end{vmatrix} \\
&= (X-1)(X+1) \begin{vmatrix} X-5 & 1 & 10 \\ -4 & X+1 & 8 \\ -2 & 1 & X+3 \end{vmatrix} \\
&= (X-1)(X+1)\cdot \\
&\qquad \cdot \big((X-5)(X+1)(X+3) - 16 - 40 - 8(X-5) + 20(X+1) + 4(X+3)\big) \\
&= (X-1)(X+1)(X^3 - X^2 - X + 1) = (X-1)^3(X+1)^2 \,.
\end{aligned}$$

Es sind also die beiden Haupträume $H_1(A)$ (Dimension 3) und $H_{-1}(A)$ (Dimension 2) zu betrachten. Für die Kerne erhalten wir (wir schreiben die Elemente von $\mathbb{R}^5$ als Spaltenvektoren):

$$\begin{aligned}
\ker(A - I_5) &= \big\langle (0,0,1,0,0)^\top, (3,2,0,3,1)^\top \big\rangle \\
\ker\big((A - I_5)^2\big) &= \big\langle (0,0,1,0,0)^\top, (1,1,0,1,0)^\top, (0,-1,0,0,1)^\top \big\rangle \\
\ker(A + I_5) &= \big\langle (0,0,1,1,0)^\top \big\rangle \\
\ker\big((A + I_5)^2\big) &= \big\langle (0,0,1,1,0)^\top, (2,2,-2,0,1)^\top \big\rangle
\end{aligned}$$

Für die höheren Potenzen bleiben die Dimensionen gleich, da sie bereits den Dimensionen der Haupträume entsprechen. Daraus ergeben sich die Größen der Jordan-Kästchen:

$$\begin{aligned}
\lambda = 1\colon &\qquad 0,2,3,3,\ldots \longrightarrow 2,1,0,\ldots \longrightarrow 1,1,0,\ldots \longrightarrow J_1(1), J_2(1) \\
\lambda = -1\colon &\qquad 0,1,2,2,\ldots \longrightarrow 1,1,0,\ldots \longrightarrow 0,1,0,\ldots \longrightarrow J_2(-1)
\end{aligned}$$

(In diesem Fall hätte es gereicht, die geometrischen Vielfachheiten 2 bzw. 1 zu kennen, um die Jordan-Kästchen zu bestimmen.) Die Jordansche Normalform von A hat also die folgende Gestalt:

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

Um eine Transformationsmatrix P zu finden, gehen wir analog zu Beispiel 26.18 vor: Ein Komplement von $\ker(A + I_5)$ in $\ker\big((A + I_5)^2\big)$ wird zum Beispiel erzeugt von $v_1 = (2,2,-2,0,1)^\top$. Ein Komplement von $\ker(A - I_5)$ in $\ker\big((A - I_5)^2\big)$ wird erzeugt von $v_2 = (1,1,0,1,0)^\top$, und ein Komplement von

$$(A - I_5)\big(\ker((A - I_5)^2)\big) = \big\langle (3,2,1,3,1)^\top \big\rangle$$

in $\ker(A - I_5)$ wird erzeugt von $v_3 = (0, 0, 1, 0, 0)^\top$. Eine geeignete Basis ist damit $B = \bigl(v_3, (A - I_5) \cdot v_2, v_2, (A + I_5) \cdot v_1, v_1\bigr)$, entsprechend der Matrix

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

♣

26.23. **Beispiel.** Als eine Anwendung der Klassifikationsaussage von Satz 26.20 wollen wir untersuchen, welche der Matrizen

BSP
Testen auf Ähnlichkeit

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad \text{und} \qquad B(x, y) = \begin{pmatrix} -1 & -1 & x \\ 1 & 1 & y \\ 0 & 0 & 3 \end{pmatrix}$$

(für $x, y \in \mathbb{R}$) in $\mathrm{Mat}(3, \mathbb{R})$ ähnlich zueinander sind. Dazu berechnen wir zunächst die charakteristischen Polynome; das ergibt

$$\chi_A = \chi_{B(x,y)} = X^2(X - 3).$$

Das schließt noch keine Ähnlichkeitsrelationen aus. Deshalb bestimmen wir die Jordansche Normalform der Matrizen. Für den Eigenwert 3 der algebraischen Vielfachheit 1 gibt es nur die Möglichkeit eines Jordan-Kästchens $J_1(3)$. Für den Eigenwert 0 gibt es die beiden Möglichkeiten $J_2(0)$ und $J_1(0), J_1(0)$. Um sie zu unterscheiden, bestimmen wir die Dimension des Kerns von A bzw. von $B(x, y)$:

$$\dim \ker(A) = 3 - \mathrm{rk}(A) = 3 - 1 = 2$$

und

$$\dim \ker(B(x, y)) = 3 - \mathrm{rk}(B(x, y)) = 3 - 2 = 1.$$

Daraus ergibt sich, dass die Jordansche Normalform von A die Jordan-Kästchen $J_1(0), J_1(0), J_1(3)$ hat (insbesondere ist A diagonalisierbar), während die Jordan-Normalform der Matrizen $B(x, y)$ die Kästchen $J_2(0), J_1(3)$ hat (insbesondere sind die $B(x, y)$ nicht diagonalisierbar). Wir sehen also, dass A zu keiner der Matrizen $B(x, y)$ ähnlich ist, dass aber alle $B(x, y)$ zueinander ähnlich sind. ♣

Die hauptsächliche praktische Anwendung der Jordanschen Normalform besteht darin, dass sie die Berechnung von Potenzen einer Matrix vereinfacht: Sei etwa $J = P^{-1}AP$ die Jordan-Normalform einer Matrix A, dann ist $A = PJP^{-1}$ und $A^k = PJ^kP^{-1}$ für alle $k \in \mathbb{N}$. Da J eine Block-Diagonalmatrix ist, ist J^k ebenfalls eine Block-Diagonalmatrix, deren Blöcke die k-ten Potenzen der Blöcke $J_m(\lambda)$ sind. Nun ist $J_m(\lambda) = \lambda I_m + J_m$, also

$$J_m(\lambda)^k = \lambda^k I_m + \binom{k}{1}\lambda^{k-1} J_m + \binom{k}{2}\lambda^{k-2} J_m^2 + \ldots + J_m^k,$$

und die Potenzen J_m^k haben eine sehr einfache Gestalt. Zum Beispiel ist

$$J_3(\lambda)^2 = \begin{pmatrix} \lambda^2 & 2\lambda & 1 \\ 0 & \lambda^2 & 2\lambda \\ 0 & 0 & \lambda^2 \end{pmatrix}$$

und allgemeiner

$$J_3(\lambda)^k = \begin{pmatrix} \lambda^k & k\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{pmatrix}.$$

Eine Anwendung, die Sie in der „Einführung in die gewöhnlichen Differentialgleichungen“ kennenlernen werden, ist die Berechnung von e^{tA} für Matrizen $A \in \mathrm{Mat}(n, \mathbb{R})$ und $t \in \mathbb{R}$. Die Exponentialfunktion für Matrizen ist definiert wie für Zahlen:

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k ;$$

man kann zeigen, dass die Reihe stets konvergiert (die Partialsummen sind Matrizen mit der Eigenschaft, dass die Folge der Einträge an jeder gegebenen Position konvergiert). Ist $A = PJP^{-1}$ wie oben, dann gilt $e^{tA} = Pe^{tJ}P^{-1}$, und e^{tJ} ist eine Block-Diagonalmatrix mit Blöcken der Form

$$e^{tJ_m(\lambda)} = e^{\lambda t} e^{tJ_m} = e^{\lambda t}\Big(I_m + tJ_m + \frac{t^2}{2} J_m^2 + \ldots\Big).$$

(Dabei benutzt man die Funktionalgleichung $e^{A+B} = e^A \cdot e^B$ für Matrizen A, B mit $AB = BA$, die aus der Potenzreihenentwicklung und dem Binomialsatz folgt.) Die Wichtigkeit der Funktion $t \mapsto e^{tA}$ kommt aus folgendem Resultat:

SATZ
Systeme linearer Differentialgleichungen mit konstanten Koeffizienten

Satz. *Ist $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top : \mathbb{R} \to \mathbb{R}^n$ ein n-Tupel differenzierbarer Funktionen, das das System von Differentialgleichungen*

$$\boldsymbol{x}'(t) = \begin{pmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{pmatrix} = A \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = A \cdot \boldsymbol{x}(t)$$

erfüllt, dann gilt $\boldsymbol{x}(t) = e^{tA} \cdot \boldsymbol{x}(0)$.

Mit der Exponentialfunktion e^{tA} kann man also solche Differentialgleichungssysteme lösen. Für die Matrix A aus dem Beispiel 26.22 oben etwa hat man

$$\begin{aligned} e^{tA} &= P \cdot \begin{pmatrix} e^t & 0 & 0 & 0 & 0 \\ 0 & e^t & te^t & 0 & 0 \\ 0 & 0 & e^t & 0 & 0 \\ 0 & 0 & 0 & e^{-t} & te^{-t} \\ 0 & 0 & 0 & 0 & e^{-t} \end{pmatrix} \cdot P^{-1} \\ &= \begin{pmatrix} 3e^t - 2e^{-t} & (3t-2)e^t + 2e^{-t} & 0 & 0 & -(6t+2)e^t + 2e^{-t} \\ 2e^t - 2e^{-t} & (2t-1)e^t + 2e^{-t} & 0 & 0 & -(4t+2)e^t + 2e^{-t} \\ e^t - (t+1)e^{-t} & te^t + te^{-t} & e^t & -e^t + e^{-t} & -2te^t + te^{-t} \\ 3e^t - (t+3)e^{-t} & (3t-2)e^t + (t+2)e^{-t} & 0 & e^{-t} & -(6t+2)e^t + (t+2)e^{-t} \\ e^t - e^{-t} & (t-1)e^t + e^{-t} & 0 & 0 & -2te^t + e^{-t} \end{pmatrix}. \end{aligned}$$

Wie sieht die Klassifikation von Matrizen bis auf Ähnlichkeit aus über $\mathbb{R}$, wo ja nicht jedes Polynom in Linearfaktoren zerfällt?

Sei $A \in \mathrm{Mat}(n, \mathbb{R})$. Dann zerfällt χ_A jedenfalls in $\mathbb{C}[X]$ in Linearfaktoren. Diese können die Form $X - \lambda$ haben mit $\lambda \in \mathbb{R}$ oder die Form $X - (\lambda + \mu\boldsymbol{i})$ mit $\lambda, \mu \in \mathbb{R}$ und $\mu \neq 0$. Dann ist neben $\lambda + \mu\boldsymbol{i}$ auch $\lambda - \mu\boldsymbol{i}$ eine Nullstelle von χ_A, denn sei

$$\chi_A = X^n + a_{n-1}X^{n-1} + \ldots + a_1 X + a_0 \quad \text{mit } a_0, a_1, \ldots, a_{n-1} \in \mathbb{R},$$

dann ist für $z = a + b\boldsymbol{i} \in \mathbb{C}$ mit komplex konjugierter Zahl $\bar{z} = a - b\boldsymbol{i}$

$$\overline{\chi_A(z)} = \overline{z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0} = \bar{z}^n + a_{n-1}\bar{z}^{n-1} + \ldots + a_1\bar{z} + a_0 = \chi_A(\bar{z}).$$

(Dabei haben wir benutzt, dass für $w, z \in \mathbb{C}$ gilt $\overline{w+z} = \bar{w} + \bar{z}$ und $\overline{wz} = \bar{w}\,\bar{z}$, sowie $\bar{a} = a$ für $a \in \mathbb{R}$.) Aus $\chi_A(\lambda + \mu\boldsymbol{i}) = 0$ folgt daher $\chi_A(\lambda - \mu\boldsymbol{i}) = \overline{\chi_A(\lambda + \mu\boldsymbol{i})} = 0$. Man kann den Faktor

$$(X - \lambda - \mu\boldsymbol{i})(X - \lambda + \mu\boldsymbol{i}) = X^2 - 2\lambda X + \lambda^2 + \mu^2$$

abdividieren; eine einfache Induktion zeigt dann, dass $\lambda+\mu\boldsymbol{i}$ und $\lambda-\mu\boldsymbol{i}$ dieselbe Vielfachheit als Nullstelle von χ_A haben. Damit haben die zugehörigen Haupträume $H_{\lambda+\mu\boldsymbol{i}}(A)$ und $H_{\lambda-\mu\boldsymbol{i}}(A)$ in $\mathbb{C}^n$ dieselbe Dimension. Sei $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$ eine Basis von $H_{\lambda+\mu\boldsymbol{i}}(A)$, sodass $A \cdot \boldsymbol{x}_j = (\lambda+\mu\boldsymbol{i})\boldsymbol{x}_j$ oder $(\lambda+\mu\boldsymbol{i})\boldsymbol{x}_j + \boldsymbol{x}_{j-1}$ ist (also eine Basis, die zu den Jordan-Kästchen für den Eigenwert $\lambda + \mu\boldsymbol{i}$ gehört). Für einen Vektor $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathbb{C}^n$ sei $\bar{\boldsymbol{y}} = (\bar{y}_1, \ldots, \bar{y}_n)$. Dann ist $(\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, \ldots, \bar{\boldsymbol{x}}_m)$ eine Basis von $H_{\lambda-\mu\boldsymbol{i}}(A)$, und

$$\bigl(\boldsymbol{x}_1 + \bar{\boldsymbol{x}}_1, \boldsymbol{i}^{-1}(\boldsymbol{x}_1 - \bar{\boldsymbol{x}}_1), \boldsymbol{x}_2 + \bar{\boldsymbol{x}}_2, \boldsymbol{i}^{-1}(\boldsymbol{x}_2 - \bar{\boldsymbol{x}}_2), \ldots \boldsymbol{x}_m + \bar{\boldsymbol{x}}_m, \boldsymbol{i}^{-1}(\boldsymbol{x}_m - \bar{\boldsymbol{x}}_m)\bigr)$$

ist eine Basis von $H_{\lambda+\mu\boldsymbol{i}}(A) \oplus H_{\lambda-\mu\boldsymbol{i}}(A)$, deren Elemente in $\mathbb{R}^n$ liegen. Die Matrix bezüglich dieser Basis des durch A gegebenen Endomorphismus dieses Untervektorraums ist dann eine Block-Diagonalmatrix mit Blöcken der Form

$$J_{2m}(\lambda, \mu) = \begin{pmatrix} \lambda & -\mu & 1 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \mu & \lambda & 0 & 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \lambda & -\mu & 1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \mu & \lambda & 0 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \ddots & \lambda & -\mu & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & \mu & \lambda & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & \lambda & -\mu \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & \mu & \lambda \end{pmatrix} \in \operatorname{Mat}(2m, \mathbb{R}).$$

Diese Matrix entsteht aus $J_m(\lambda + \mu\boldsymbol{i})$, indem jeder Eintrag $a + b\boldsymbol{i}$ (mit $a, b \in \mathbb{R}$) durch die 2×2-Matrix $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ ersetzt wird (dies ist die Matrix des $\mathbb{R}$-linearen Endomorphismus $z \mapsto (a + b\boldsymbol{i})z$ von $\mathbb{C}$ bezüglich der $\mathbb{R}$-Basis $(1, \boldsymbol{i})$ von $\mathbb{C}$).

Daraus ergibt sich der folgende Satz (formuliert für Matrizen):

Satz. *Sei $A \in \operatorname{Mat}(n, \mathbb{R})$. Dann ist A ähnlich zu einer Block-Diagonalmatrix, deren Blöcke die Form $J_m(\lambda)$ (mit $\lambda \in \mathbb{R}$) oder $J_{2m}(\lambda, \mu)$ (mit $\lambda \in \mathbb{R}$, $\mu \in \mathbb{R}_{>0}$) haben. Diese Blöcke sind bis auf ihre Reihenfolge eindeutig bestimmt.* **SATZ** Reelle JNF

27. Das Tensorprodukt

Die Idee des Tensorprodukts ist es, bilineare Abbildungen $b\colon V_1 \times V_2 \to W$ zu „linearisieren", indem man den Definitionsbereich $V_1 \times V_2$ durch einen geeigneten Vektorraum V und b durch eine lineare Abbildung $V \to W$ ersetzt. Das führt auf folgende Definition:

DEF * Tensorprodukt

27.1. **Definition.** Seien V_1 und V_2 zwei K-Vektorräume. Ein K-Vektorraum V zusammen mit einer bilinearen Abbildung $\beta\colon V_1 \times V_2 \to V$ heißt ein *Tensorprodukt* von V_1 und V_2, wenn V und β die folgende „universelle Eigenschaft" erfüllen: Für jeden K-Vektorraum W und jede bilineare Abbildung $b\colon V_1 \times V_2 \to W$ gibt es genau eine lineare Abbildung $f\colon V \to W$ mit $b = f \circ \beta$. ♢

Wir schreiben $\mathrm{Bil}(V_1, V_2; W)$ für den Vektorraum der bilinearen Abbildungen von $V_1 \times V_2$ nach W; dann besagt die Definition, dass (V, β) genau dann ein Tensorprodukt von V_1 und V_2 ist, wenn die Abbildung

$$\mathrm{Hom}(V, W) \longrightarrow \mathrm{Bil}(V_1, V_2; W), \quad f \longmapsto f \circ \beta$$

für alle W bijektiv (und damit ein Isomorphismus) ist.

Es folgt, dass Tensorprodukte bis auf eindeutigen Isomorphismus eindeutig bestimmt sind:

SATZ Eindeutigkeit des Tensorprodukts

27.2. **Satz.** *Seien V_1 und V_2 zwei K-Vektorräume und seien (V, β) und (V', β') zwei Tensorprodukte von V_1 und V_2. Dann gibt es einen eindeutig bestimmten Isomorphismus $\varphi\colon V \to V'$ mit $\beta' = \varphi \circ \beta$.*

Beweis. Wir wenden die universelle Eigenschaft von V an auf die bilineare Abbildung β', das liefert eine eindeutig bestimmte lineare Abbildung $\varphi\colon V \to V'$ mit $\beta' = \varphi \circ \beta$. Analog gibt es $\varphi'\colon V' \to V$ mit $\beta = \varphi' \circ \beta'$. Die Eindeutigkeit in der universellen Eigenschaft liefert $\varphi' \circ \varphi = \mathrm{id}_V$ und $\varphi \circ \varphi' = \mathrm{id}_{V'}$; damit ist φ ein Isomorphismus. ❑

Wir schreiben $V_1 \otimes V_2$ (oder $V_1 \otimes_K V_2$, wenn es auf den Körper ankommt) für ein Tensorprodukt von V_1 und V_2; die bilineare Abbildung β wird dann in der Form $\beta(v_1, v_2) = v_1 \otimes v_2$ notiert.

Was ist eine „universelle Eigenschaft"? Wir haben gewisse „Objekte" (im Fall des Tensorprodukts sind es Paare (W, b) aus einem Vektorraum W und einer bilinearen Abbildung $b\colon V_1 \times V_2 \to W$), zwischen denen es „Morphismen" gibt (beim Tensorprodukt ist ein Morphismus $(W, b) \to (W', b')$ eine lineare Abbildung $\varphi\colon W \to W'$ mit $\varphi \circ b = b'$). Die universelle Eigenschaft besagt dann, dass es von dem betreffenden „universellen" Objekt genau einen Morphismus zu jedem anderen Objekt gibt (dann hat man ein *initiales* Objekt) oder auch, dass es von jedem Objekt genau einen Morphismus zum universellen Objekt gibt (dann hat man ein *finales* Objekt). Man kann sehr abstrakt formulieren, welche Eigenschaften die Objekte und Morphismen haben müssen (sie bilden dann eine sogenannte *Kategorie*); der Teil der Mathematik, der sich damit beschäftigt, heißt *Kategorientheorie* und wird gerne liebevoll als „abstract nonsense" bezeichnet.

Ein einfaches, aber triviales Beispiel erhalten wir, wenn wir als Objekte K-Vektorräume und als Morphismen lineare Abbildungen betrachten. Ein universelles Objekt U hat dann die Eigenschaft, dass es immer *genau eine* lineare Abbildung $U \to V$ gibt, für jeden K-Vektorraum V. Es ist dann leicht zu sehen, dass hier der Null-Vektorraum ein universelles (initiales) Objekt ist. Er ist übrigens auch ein finales Objekt in dieser Kategorie.

Die Eindeutigkeitsaussage führt auf die Frage nach der Existenz des Tensorprodukts. Zuerst noch eine Definition.

27.3. **Definition.** Seien X eine Menge und K ein Körper. Sei

$$K^{(X)} = \{(\lambda_x)_{x \in X} \mid \lambda_x = 0 \text{ für alle bis auf endlich viele } x \in X\}.$$

Dann ist $K^{(X)}$ ein Untervektorraum von K^X mit Basis $(\mathbf{e}_x)_{x \in X}$, wobei (analog zum Standardvektorraum K^n) $\mathbf{e}_x = (\delta_{x,y})_{y \in X}$ die Familie ist, deren Komponenten alle null sind bis auf die x-te Komponente, die den Wert 1 hat. $\diamondsuit$

DEF Vektorraum mit gegebener Basis

Der Beweis der letzten beiden Aussagen ist eine Übungsaufgabe.

* 27.4. **Satz.** *Seien B_1 und B_2 Basen der K-Vektorräume V_1 und V_2. Dann ist $V = K^{(B_1 \times B_2)}$ zusammen mit*

SATZ Existenz des Tensorprodukts

$$\beta\colon V_1 \times V_2 \longrightarrow V, \quad \Bigl(\sum_{v \in B_1} \lambda_v v, \sum_{v' \in B_2} \mu_{v'} v'\Bigr) \longmapsto (\lambda_v \mu_{v'})_{(v,v') \in B_1 \times B_2}$$

ein Tensorprodukt von V_1 und V_2. (In den Summen sind alle bis auf endlich viele Koeffizienten λ_v bzw. $\mu_{v'}$ null.)

Beweis. Wir schreiben $\mathbf{e}_{(v,v')}$ wie in Definition 27.3 für die Elemente der „Standard-Basis“ von V. Sei W ein weiterer K-Vektorraum und $b\colon V_1 \times V_2 \to W$ bilinear. Eine lineare Abbildung $f\colon V \to W$ mit $f \circ \beta = b$ muss dann für alle $v \in B_1$ und $v' \in B_2$ die Gleichung

$$f(\mathbf{e}_{(v,v')}) = f\bigl(\beta(v, v')\bigr) = b(v, v')$$

erfüllen. Es gibt genau eine lineare Abbildung f, die auf der Standard-Basis von V diese Werte annimmt (Satz 10.11). Es bleibt zu zeigen, dass für diese lineare Abbildung tatsächlich $f \circ \beta = b$ gilt: Seien $v_1 = \sum_{v \in B_1} \lambda_v v \in V_1$ und $v_2 = \sum_{v' \in B_2} \mu_{v'} v' \in V_2$. Dann gilt

$$\begin{aligned} f\bigl(\beta(v_1, v_2)\bigr) &= f\bigl((\lambda_v \mu_{v'})_{(v,v') \in B_1 \times B_2}\bigr) = f\Bigl(\sum_{(v,v') \in B_1 \times B_2} \lambda_v \mu_{v'} \mathbf{e}_{(v,v')}\Bigr) \\ &= \sum_{(v,v') \in B_1 \times B_2} \lambda_v \mu_{v'} f(\mathbf{e}_{(v,v')}) = \sum_{(v,v') \in B_1 \times B_2} \lambda_v \mu_{v'} b(v, v') \\ &= b\Bigl(\sum_{v \in B_1} \lambda_v v, \sum_{v' \in B_2} \mu_{v'} v'\Bigr) = b(v_1, v_2). \end{aligned}$$

❑

Damit ist die Existenz des Tensorprodukts jedenfalls für endlich-dimensionale Vektorräume gezeigt. Da (unter Verwendung des Auswahlaxioms) jeder Vektorraum eine Basis hat, gilt die Existenzaussage auch allgemein.

Es gibt auch eine Basis-freie Konstruktion des Tensorprodukts, die allerdings ziemlich „brutal“ und „verschwenderisch“ anmutet. Wir setzen $\mathcal{V} = K^{(V_1 \times V_2)}$ (das ist also ein Vektorraum, der für *jedes Element* $(v_1, v_2) \in V_1 \times V_2$ ein Basiselement $\mathbf{e}_{(v_1,v_2)}$ hat) und definieren $\mathcal{U}$ als den Untervektorraum von $\mathcal{V}$, der von allen Elementen einer der Formen

$$\mathbf{e}_{(\lambda v_1, v_2)} - \lambda \mathbf{e}_{(v_1,v_2)}, \quad \mathbf{e}_{(v_1, \lambda v_2)} - \lambda \mathbf{e}_{(v_1,v_2)},$$
$$\mathbf{e}_{(v_1+v_1', v_2)} - \mathbf{e}_{(v_1,v_2)} - \mathbf{e}_{(v_1',v_2)}, \quad \mathbf{e}_{(v_1, v_2+v_2')} - \mathbf{e}_{(v_1,v_2)} - \mathbf{e}_{(v_1,v_2')}$$

mit $v_1, v_1' \in V_1$, $v_2, v_2' \in V_2$ und $\lambda \in K$ erzeugt wird. Dann setzen wir $V = \mathcal{V}/\mathcal{U}$ und $\beta(v_1, v_2) = [\mathbf{e}_{(v_1,v_2)}]$. Man rechnet nach, dass β bilinear ist (das kommt direkt aus der Definition von $\mathcal{U}$). Ist $b\colon V_1 \times V_2 \to W$ bilinear, dann definiert man zunächst eine lineare Abbildung $F\colon \mathcal{V} \to W$ durch $F(\mathbf{e}_{(v_1,v_2)}) = b(v_1, v_2)$ (eindeutige Festlegung durch Bild der Basis). Aus der Bilinearität von b folgt, dass $\mathcal{U}$ im Kern von F enthalten ist; es gibt dann (das ist die universelle Eigenschaft des Quotientenraums, siehe Satz 18.12) eine eindeutig bestimmte lineare Abbildung $f\colon V \to W$ mit $F = f \circ \pi$, wobei $\pi\colon \mathcal{V} \to V$ der kanonische Epimorphismus ist. Dann gilt $f \circ \beta = b$. Die Eindeutigkeit von f ist auch leicht zu sehen — das Bild von β erzeugt V, also gibt es höchstens eine lineare Abbildung, die auf dem Bild von β gegebene Werte annimmt.

BSP
Matrizenraum als Tensorprodukt

27.5. **Beispiel.** Seien K ein Körper und $m, n \in \mathbb{N}$. Dann ist der Vektorraum $\mathrm{Mat}(m \times n, K)$ isomorph zum Tensorprodukt $K^m \otimes K^n$. Dabei ist die bilineare Abbildung $\beta\colon K^m \times K^n \to \mathrm{Mat}(m \times n, K)$ gegeben durch $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} \cdot \boldsymbol{y}^\top$ („Spaltenvektor mal Zeilenvektor"); das Bild von β besteht genau aus der Nullmatrix und den Matrizen vom Rang 1. Daran sieht man sehr schön, dass **keineswegs alle Elemente von $V_1 \otimes V_2$ die Form $v_1 \otimes v_2$ haben!** Es gibt ja auch Matrizen von höherem Rang (jedenfalls, wenn m und n größer als 1 sind). ♣

Das Tensorprodukt $V_1 \otimes V_2$ wird von den Elementen der Form $v_1 \otimes v_2$ (also dem Bild von β) erzeugt. (Das folgt aus der Konstruktion in Satz 27.4 oder auch direkt aus der universellen Eigenschaft: Wäre $U = \langle \mathrm{im}(\beta) \rangle$ nicht ganz $V_1 \otimes V_2$, dann könnte man ein Komplement $U' \neq \{\mathbf{0}\}$ von U wählen und darauf die lineare Abbildung aus der universellen Eigenschaft beliebig definieren, was der Eindeutigkeit widerspräche.) Man kann dann fragen, wie viele solche Elemente man höchstens braucht, um ein beliebiges Element darzustellen.

SATZ
Elemente des Tensorprodukts

27.6. **Satz.** *Seien V und V' zwei K-Vektorräume und sei $(b_1, \ldots, b_n)$ eine Basis von V'. Dann lässt sich jedes Element w von $V \otimes V'$ eindeutig schreiben als*

$$w = v_1 \otimes b_1 + v_2 \otimes b_2 + \ldots + v_n \otimes b_n$$

mit $v_1, v_2, \ldots, v_n \in V$.

Man kann das so interpretieren, dass man beim Übergang von V' zu $V \otimes V'$ die skalaren Koeffizienten der Basis $(b_1, \ldots, b_n)$ durch „Koeffizienten" aus V ersetzt.

Beweis. Wir haben die lineare Abbildung

$$\psi\colon V^n \longrightarrow V \otimes V', \quad (v_1, v_2, \ldots, v_n) \longmapsto v_1 \otimes b_1 + v_2 \otimes b_2 + \ldots + v_n \otimes b_n$$

(wobei die Vektorraumstruktur von V^n komponentenweise definiert ist); zu zeigen ist, dass ψ ein Isomorphismus ist. Dazu konstruieren wir die Umkehrabbildung: Wir betrachten folgende bilineare Abbildung $b\colon V \times V' \to V^n$:

$$b\Bigl(v, \sum_{j=1}^{n} \lambda_j b_j\Bigr) = (\lambda_1 v, \lambda_2 v, \ldots, \lambda_n v)\,.$$

Wegen der universellen Eigenschaft gibt es dann eine eindeutig bestimmte lineare Abbildung $\varphi\colon V \otimes V' \to V^n$ mit $\varphi(v \otimes b_j) = (\delta_{ij} v)_{1 \le i \le n}$. Es gilt offensichtlich $\varphi \circ \psi = \mathrm{id}_{V^n}$ und $(\psi \circ \varphi)(v \otimes b_j) = v \otimes b_j$ für alle $v \in V$ und $j \in \{1, 2, \ldots, n\}$. Da $V \otimes V'$ von allen $v \otimes b_j$ erzeugt wird, folgt daraus $\psi \circ \varphi = \mathrm{id}_{V \otimes V'}$, also ist ψ wie gewünscht ein Isomorphismus. (Dass die $v \otimes b_j$ Erzeuger von $V \otimes V'$ sind, kommt daher, dass jedes $v \otimes v'$ eine Linearkombination dieser spezielleren Elemente ist: $v' = \sum_j \lambda_j b_j$ impliziert $v \otimes v' = \sum_j \lambda_j (v \otimes b_j)$.) ❑

BSP Matrizen

27.7. **Beispiel.** Für eine $m \times n$-Matrix A bedeutet das, dass A Summe von höchstens n Matrizen vom Rang 1 ist. (Genauer gilt, dass A Summe von genau $r = \operatorname{rk}(A)$ Matrizen vom Rang 1 ist.) ♣

Für das Tensorprodukt gelten die folgenden Rechenregeln.

SATZ „Rechenregeln“ für $\otimes$

27.8. **Satz.** *Seien V_1, V_2, V_3 drei K-Vektorräume. Dann gibt es kanonische Isomorphismen*

$$\begin{aligned}(V_1 \otimes V_2) \otimes V_3 &\cong V_1 \otimes (V_2 \otimes V_3)\\ V_1 \otimes V_2 &\cong V_2 \otimes V_1\\ K \otimes V_1 &\cong V_1\\ \{\mathbf{0}\} \otimes V_1 &\cong \{\mathbf{0}\}\end{aligned}$$

Wegen der Assoziativität des Tensorprodukts schreibt man auch einfach $V_1 \otimes V_2 \otimes V_3$ (und analog mit mehr als drei „Faktoren“).

Beweis. Wir zeigen hier exemplarisch nur eine der Aussagen; die übrigen Beweise sollten Sie als Übungsaufgaben betrachten. Die Beweis-Struktur ist immer dieselbe: Man konstruiert natürliche lineare Abbildungen in beiden Richtungen und zeigt unter Verwendung der universellen Eigenschaft, dass sie zueinander invers sind.

Wir beweisen das „Assoziativgesetz“ $(V_1 \otimes V_2) \otimes V_3 \cong V_1 \otimes (V_2 \otimes V_3)$. Dazu fixieren wir erst einmal $v_3 \in V_3$. Die Abbildung

$$V_1 \times V_2 \longrightarrow V_1 \otimes (V_2 \otimes V_3), \quad (v_1, v_2) \longmapsto v_1 \otimes (v_2 \otimes v_3)$$

ist offensichtlich bilinear und führt daher zu einer linearen Abbildung

$$f_{v_3} \colon V_1 \otimes V_2 \longrightarrow V_1 \otimes (V_2 \otimes V_3) \quad \text{mit} \quad f_{v_3}(v_1 \otimes v_2) = v_1 \otimes (v_2 \otimes v_3)\,.$$

Da f_{v_3} linear von v_3 abhängt (d.h., $V_3 \to \operatorname{Hom}\bigl(V_1 \otimes V_2, V_1 \otimes (V_2 \otimes V_3)\bigr)$, $v_3 \mapsto f_{v_3}$, ist linear), ist die Abbildung

$$b \colon (V_1 \otimes V_2) \times V_3 \longrightarrow V_1 \otimes (V_2 \otimes V_3), \quad b(v, v_3) = f_{v_3}(v)$$

bilinear. Deshalb gibt es eine lineare Abbildung

$$\varphi \colon (V_1 \otimes V_2) \otimes V_3 \longrightarrow V_1 \otimes (V_2 \otimes V_3) \quad \text{mit} \quad \varphi\bigl((v_1 \otimes v_2) \otimes v_3\bigr) = v_1 \otimes (v_2 \otimes v_3)\,.$$

Analog gibt es eine lineare Abbildung

$$\psi \colon V_1 \otimes (V_2 \otimes V_3) \longrightarrow (V_1 \otimes V_2) \otimes V_3 \quad \text{mit} \quad \psi\bigl(v_1 \otimes (v_2 \otimes v_3)\bigr) = (v_1 \otimes v_2) \otimes v_3\,.$$

Da die angegebenen Elemente jeweils ein Erzeugendensystem bilden, folgt, dass φ und ψ zueinander inverse Isomorphismen sind. ❑

Die wichtigsten Prinzipien beim Umgang mit Tensorprodukten sind:

- Es gibt genau dann eine (dann auch eindeutig bestimmte) lineare Abbildung $f \colon V_1 \otimes V_2 \to W$ mit $f(v_1 \otimes v_2) = b(v_1, v_2)$ für alle $v_1 \in V_1$, $v_2 \in V_2$, wenn $b \colon V_1 \times V_2 \to W$ bilinear ist.
- Die Abbildung $(v_1, v_2) \mapsto v_1 \otimes v_2$ ist selbst bilinear.
- Die Elemente der Form $v_1 \otimes v_2$ erzeugen $V_1 \otimes V_2$, aber im Allgemeinen hat nicht jedes Element von $V_1 \otimes V_2$ diese Form.

Zum Beispiel ist die Auswertung von linearen Abbildungen

$$\mathrm{Hom}(V,W)\times V\longrightarrow W,\quad (f,v)\longmapsto f(v)$$

bilinear und führt daher zu einer linearen Abbildung

$$\mathrm{Hom}(V,W)\otimes V\longrightarrow W\quad\text{mit}\quad f\otimes v\longmapsto f(v)\,.$$

Auch die Abbildung

$$V^*\times W\longrightarrow \mathrm{Hom}(V,W),\quad (\phi,w)\longmapsto \big(v\mapsto \phi(v)w\big)$$

ist bilinear und führt zu einer kanonischen linearen Abbildung

$$V^*\otimes W\longrightarrow \mathrm{Hom}(V,W)\,.$$

SATZ $\mathrm{Hom}(V,W) \cong V^*\otimes W$

27.9. **Satz.** *Seien V und W zwei K-Vektorräume. Ist V oder W endlich-dimensional, dann ist die kanonische lineare Abbildung $\Phi\colon V^*\otimes W\to \mathrm{Hom}(V,W)$ ein Isomorphismus.*

Beweis. Sei zunächst W endlich-dimensional und $(b_1,\dots,b_m)$ eine Basis von W. Jede lineare Abbildung $f\colon V\to W$ hat dann die Form

$$f(v)=f_1(v)b_1+f_2(v)b_2+\ldots+f_m(v)b_m$$

mit eindeutig bestimmten Linearformen $f_1,f_2,\dots,f_m\in V^*$. Wir definieren

$$\Psi\colon\ \mathrm{Hom}(V,W)\longrightarrow V^*\otimes W,\quad f\longmapsto f_1\otimes b_1+f_2\otimes b_2+\ldots+f_m\otimes b_m\,;$$

dann gilt $\Phi\circ\Psi=\mathrm{id}_{\mathrm{Hom}(V,W)}$ (klar) und $\Psi\circ\Phi=\mathrm{id}_{V^*\otimes W}$: Sei $w=\lambda_1b_1+\ldots+\lambda_mb_m$, dann ist

$$\begin{aligned}(\Psi\circ\Phi)(\phi\otimes w)&=(\Psi\circ\Phi)(\lambda_1\phi\otimes b_1+\ldots+\lambda_m\phi\otimes b_m)\\&=\Psi\big(v\mapsto\lambda_1\phi(v)b_1+\ldots+\lambda_m\phi(v)b_m\big)\\&=\lambda_1\phi\otimes b_1+\ldots+\lambda_m\phi\otimes b_m\\&=\phi\otimes(\lambda_1b_1+\ldots+\lambda_mb_m)\\&=\phi\otimes w\,.\end{aligned}$$

Damit gilt $\Psi\circ\Phi=\mathrm{id}$ auf einem Erzeugendensystem von $V^*\otimes W$, also ist $\Psi\circ\Phi=\mathrm{id}$. Also ist Φ ein Isomorphismus.

Sei jetzt V endlich-dimensional mit Basis $(b_1,\dots,b_n)$ und dualer Basis $(b_1^*,\dots,b_n^*)$ von V^*. Eine lineare Abbildung $f\colon V\to W$ ist eindeutig festgelegt durch die beliebig wählbaren Bilder $f(b_1),\dots,f(b_n)$. Wir definieren

$$\Psi\colon\ \mathrm{Hom}(V,W)\to V^*\otimes W,\quad f\longmapsto b_1^*\otimes f(b_1)+\ldots+b_n^*\otimes f(b_n)\,.$$

Dann sind Φ und Ψ wieder invers zueinander, denn

$$\begin{aligned}(\Psi\circ\Phi)(\phi\otimes w)&=\Psi\big(v\mapsto\phi(v)w\big)\\&=b_1^*\otimes\big(\phi(b_1)w\big)+\ldots+b_n^*\otimes\big(\phi(b_n)w\big)\\&=\phi(b_1)b_1^*\otimes w+\ldots+\phi(b_n)b_n^*\otimes w\\&=\big(\phi(b_1)b_1^*+\ldots+\phi(b_n)b_n^*\big)\otimes w\\&=\phi\otimes w\end{aligned}$$

(also $\Psi \circ \Phi = \mathrm{id}$ auf einem Erzeugendensystem, damit gilt $\Psi \circ \Phi = \mathrm{id}$) und

$$\begin{aligned}(\Phi \circ \Psi)(f) &= \Phi\bigl(b_1^* \otimes f(b_1) + \ldots + b_n^* \otimes f(b_n)\bigr) \\ &= \Phi\bigl(b_1^* \otimes f(b_1)\bigr) + \ldots + \Phi\bigl(b_n^* \otimes f(b_n)\bigr) \\ &= \bigl(v \mapsto b_1^*(v) f(b_1) + \ldots + b_n^*(v) f(b_n)\bigr) \\ &= f\,,\end{aligned}$$

denn die Abbildung in der vorletzten Zeile bildet für alle $j \in \{1, 2, \ldots, n\}$ das Basiselement b_j auf $f(b_j)$ ab. ❑

Sind V und W beide unendlich-dimensional, dann ist Φ zwar noch injektiv, aber nicht mehr surjektiv — das Tensorprodukt ist „zu klein", um alle Homomorphismen zu spezifizieren: Sei B eine (unendliche) Basis von W, dann hat jedes Element des Tensorprodukts $V^* \otimes W$ die Form $t = \sum_{b \in B'} \phi_b \otimes b$ mit einer endlichen Teilmenge $B' \subset B$ (das beweist man ähnlich wie in Satz 27.6). Die lineare Abbildung $\Phi(t)$ bildet $v \in V$ auf $\sum_{b \in B'} \phi_b(v) b$ ab, das Bild von $\Phi(t)$ ist also im endlich-dimensionalen Untervektorraum $\langle B' \rangle$ von W enthalten. Ähnlich wie in Satz 27.9 sieht man, dass jede lineare Abbildung mit endlich-dimensionalem Bild (also mit endlichem Rang) im Bild von Φ liegt. Es gibt aber stets lineare Abbildungen $V \to W$, deren Bild unendlich-dimensional ist. Im Fall $V = W$ liegt zum Beispiel id_V nicht im Bild von Φ.

Im Fall $V = W$ endlich-dimensional haben wir dann einen Isomorphismus

$$\mathrm{End}(V) = \mathrm{Hom}(V, V) \xrightarrow{\cong} V^* \otimes V$$

(die Umkehrabbildung von Φ in Satz 27.9) und wir können folgende Komposition bilden:

$$\mathrm{End}(V) \xrightarrow{\cong} V^* \otimes V = \mathrm{Hom}(V, K) \otimes V \xrightarrow{\mathrm{ev}} K\,,$$

wobei die letzte Abbildung die von der Auswertung induzierte Abbildung ist. Diese Abbildung $V^* \otimes V \to K$ (oder entsprechend $V \otimes V^* \to K$) heißt auch *Kontraktion*. **DEF** Kontraktion

27.10. **Satz.** *Die so definierte Abbildung* $\mathrm{End}(V) \to K$ *ist die Spur* $f \mapsto \mathrm{Tr}(f)$. **SATZ** Spur Basis-frei

Beweis. Da die Spur über Matrizen definiert ist, müssen wir zuerst eine Basis $B = (b_1, \ldots, b_n)$ von V wählen; sei $(b_1^*, \ldots, b_n^*)$ die duale Basis von V^*. Dann ist für $f \in \mathrm{End}(V)$

$$\mathrm{Mat}_B(f) = \bigl(b_i^*(f(b_j))\bigr)_{i,j}\,,$$

also

$$\begin{aligned}\mathrm{Tr}(f) &= \mathrm{Tr}\bigl(\mathrm{Mat}_B(f)\bigr) \\ &= b_1^*\bigl(f(b_1)\bigr) + \ldots + b_n^*\bigl(f(b_n)\bigr) \\ &= \mathrm{ev}\bigl(b_1^* \otimes f(b_1) + \ldots + b_n^* \otimes f(b_n)\bigr) \\ &= \mathrm{ev}\bigl(\Phi^{-1}(f)\bigr)\end{aligned}$$

(vergleiche den Beweis von Satz 27.9 für $\dim V < \infty$). Das ist genau die Behauptung. ❑

Wir betrachten jetzt das Zusammenspiel von linearen Abbildungen mit dem Tensorprodukt.

LEMMA Tensorprodukt von Abb.

27.11. **Lemma.** *Seien V, V', W, W' vier K-Vektorräume und seien $f: V \to W$ und $f': V' \to W'$ lineare Abbildungen. Dann gibt es genau eine lineare Abbildung $f \otimes f': V \otimes V' \to W \otimes W'$ mit $(f \otimes f')(v \otimes v') = f(v) \otimes f'(v')$ für alle $v \in V$ und $v' \in V'$.*

Beweis. Die Abbildung $b: V \times V' \to W \otimes W'$, $(v, v') \mapsto f(v) \otimes f'(v')$, ist bilinear; nach der universellen Eigenschaft von $V \otimes V'$ existiert also eine eindeutig bestimmte lineare Abbildung $f \otimes f'$ wie angegeben. ❑

Es gilt dann $\mathrm{id}_V \otimes \mathrm{id}_{V'} = \mathrm{id}_{V \otimes V'}$, und wenn $V_1 \xrightarrow{f} V_2 \xrightarrow{g} V_3$ und $V_1' \xrightarrow{f'} V_2' \xrightarrow{g'} V_3'$ lineare Abbildungen sind, dann gilt $(g \circ f) \otimes (g' \circ f') = (g \otimes g') \circ (f \otimes f')$, wie man leicht auf Elementen der Form $v \otimes v'$ nachprüft.

BSP Komposition als Kontraktion

27.12. **Beispiel.** Seien V_1, V_2 und V_3 drei endlich-dimensionale K-Vektorräume. Wir betrachten

$$\begin{aligned} \mathrm{Hom}(V_1, V_2) \times \mathrm{Hom}(V_2, V_3) &\cong (V_1^* \otimes V_2) \times (V_2^* \otimes V_3) \\ &\longrightarrow V_1^* \otimes (V_2 \otimes V_2^*) \otimes V_3 \xrightarrow{\mathrm{id} \otimes \mathrm{ev} \otimes \mathrm{id}} V_1^* \otimes K \otimes V_3 \\ &\cong V_1^* \otimes V_3 \cong \mathrm{Hom}(V_1, V_3)\,. \end{aligned}$$

Diese Abbildung ist dieselbe wie die Komposition $(g, f) \mapsto f \circ g$. Es genügt, das für $g: v_1 \mapsto \phi(v_1)w$ und $f: v_2 \mapsto \phi'(v_2)w'$ nachzuweisen, wobei $\phi \in V_1^*$, $\phi' \in V_2^*$ und $w \in V_2$, $w' \in V_3$ seien. Wir erhalten

$$\begin{aligned} (g, f) &\longmapsto (\phi \otimes w, \phi' \otimes w') \longmapsto \phi \otimes (w \otimes \phi') \otimes w' \\ &\longmapsto \phi \otimes \phi'(w) \otimes w' \longmapsto \phi'(w)\, \phi \otimes w' \\ &\longmapsto \big(v_1 \mapsto \phi'(w)\phi(v_1)w' = \phi'(\phi(v_1)w)w' = (f \circ g)(v_1)\big) \\ &= f \circ g \end{aligned}$$

Im Fall $V_1 = V_3$ erhält man dann auch sehr leicht die Beziehung $\mathrm{Tr}(f \circ g) = \mathrm{Tr}(g \circ f)$ (Übung). ♣

BSP Kroneckerprodukt von Matrizen

27.13. **Beispiel.** Ist $B = (b_1, \ldots, b_n)$ eine Basis von V und $B' = (b_1', \ldots, b_m')$ eine Basis von W, dann ist

$$B'' = (b_1 \otimes b_1', \ldots, b_1 \otimes b_m', b_2 \otimes b_1', \ldots, b_2 \otimes b_m', \ldots\ldots, b_n \otimes b_1', \ldots, b_n \otimes b_m')$$

DEF Kronecker-Produkt

eine Basis von $V \otimes W$. Ist f ein Endomorphismus von V und g ein Endomorphismus von W, dann ist $f \otimes g$ ein Endomorphismus von $V \otimes W$. Die zugehörige Matrix $A'' = \mathrm{Mat}_{B''}(f \otimes g)$ heißt das *Kronecker-Produkt* von $A = (a_{ij}) = \mathrm{Mat}_B(f)$ und $A' = \mathrm{Mat}_{B'}(g)$. Es gilt dann

$$A'' = \left(\begin{array}{c|c|c|c} a_{11}A' & a_{12}A' & \cdots & a_{1n}A' \\ \hline a_{21}A' & a_{22}A' & \cdots & a_{2n}A' \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline a_{n1}A' & a_{n2}A' & \cdots & a_{nn}A' \end{array}\right).$$

Man schreibt dafür auch $A'' = A \otimes A'$.

Es gilt $\mathrm{Tr}(A \otimes A') = \mathrm{Tr}(A)\,\mathrm{Tr}(A')$ und $\det(A \otimes A') = \det(A)^m \det(A')^n$ (Übung). ♣

Der folgende Satz gehört eigentlich in das Kapitel über euklidische Vektorräume.

SATZ
Hadamardsche Ungleichung

Satz. *Sei* $A = (\boldsymbol{x}_1|\boldsymbol{x}_2|\cdots|\boldsymbol{x}_n) \in \mathrm{Mat}(n, \mathbb{R})$. *Dann gilt*

$$|\det(A)| \le \|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|\cdots\|\boldsymbol{x}_n\|$$

mit Gleichheit genau dann, wenn eine Spalte null ist oder die Spalten $\boldsymbol{x}_j \in \mathbb{R}^n$ *paarweise orthogonal sind.*

Beweis. Wir beweisen die Aussage durch Induktion. Der Fall $n = 1$ (und auch der Fall $n = 0$) ist klar. Sei also $n \ge 2$. Dass Gleichheit gilt, wenn eine Spalte null ist, ist offensichtlich. Wir können also annehmen, dass $\boldsymbol{x}_j \ne 0$ ist für alle $j \in \{1, 2, \ldots, n\}$. Mit dem Gram-Schmidt-Verfahren konstruieren wir eine ONB B von $\mathbb{R}^n$, deren erstes Element ein skalares Vielfaches von $\boldsymbol{x}_1$ ist; sei P die Matrix, deren Spalten die Vektoren in B sind. Dann ist

$$P^{-1}A = \left(\begin{array}{c|c} \|\boldsymbol{x}_1\| & \boldsymbol{y}^\top \\ \hline \mathbf{0} & A' \end{array}\right)$$

mit $\boldsymbol{y} = (y_2, \ldots, y_n)^\top \in \mathbb{R}^{n-1}$. Es gilt (beachte $P \in \mathrm{O}(n)$, damit $\det(P) = \pm 1$)

$$|\det(A)| = |\det(P^{-1}A)| = \|\boldsymbol{x}_1\||\det(A')|\,.$$

Wir schreiben $A' = (\boldsymbol{x}'_2|\cdots|\boldsymbol{x}'_n)$. Da P^{-1} orthogonal ist, haben die Spalten von $P^{-1}A$ dieselbe Länge wie die Spalten von A, also gilt $\|\boldsymbol{x}_j\|^2 = y_j^2 + \|\boldsymbol{x}'_j\|^2$. Aus der Induktionsvoraussetzung ergibt sich

$$|\det(A')| \le \|\boldsymbol{x}'_2\|\cdots\|\boldsymbol{x}'_n\| \le \|\boldsymbol{x}_2\|\cdots\|\boldsymbol{x}_n\|\,.$$

Daraus folgt die behauptete Ungleichung. In der zweiten Ungleichung oben gilt Gleichheit genau dann, wenn $\boldsymbol{y} = \mathbf{0}$ ist, und das bedeutet gerade, dass $\boldsymbol{x}_1 \perp \boldsymbol{x}_j$ ist für alle $j \in \{2, \ldots, n\}$. Nach Induktionsvoraussetzung gilt Gleichheit in der ersten Ungleichung genau dann, wenn ein $\boldsymbol{x}'_j = \mathbf{0}$ ist oder alle $\boldsymbol{x}'_j$ paarweise orthogonal sind. Beide Bedingungen zusammen gelten genau dann, wenn alle $\boldsymbol{x}_j$ paarweise orthogonal sind (den Fall $\boldsymbol{x}_j = \mathbf{0}$ hatten wir ja ausgeschlossen). ❑

Eine $n \times n$-Matrix A mit Einträgen ± 1, für die $|\det(A)| = n^{n/2}$ gilt (das bedeutet gerade, dass die Spalten (oder Zeilen) paarweise orthogonal sind; äquivalent ist also die Bedingung $A^\top A = nI_n$), heißt *Hadamard-Matrix*. Man überlegt sich relativ leicht, dass es solche Matrizen nur für $n = 1$, $n = 2$ und $n = 4m$ mit $m \ge 1$ geben kann. Es ist ein offenes Problem, ob es für *alle* diese n Hadamard-Matrizen gibt. Da man leicht zeigen kann, dass das Kronecker-Produkt zweier Hadamard-Matrizen wieder eine Hadamard-Matrix ist, folgt jedenfalls, dass die Menge der natürlichen Zahlen n, für die es eine $n \times n$-Hadamard-Matrix gibt, multiplikativ abgeschlossen ist. Zum Beispiel ist

$$A_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

eine Hadamard-Matrix, also gibt es für jedes $n = 2^k$ Hadamard-Matrizen. Mit Hilfe von zahlentheoretischen Konstruktionen erhält man Hadamard-Matrizen für

$$n = 4m = q + 1 \qquad \text{oder} \qquad n = 8m + 4 = 2(q + 1)\,,$$

wenn q die Potenz einer Primzahl ist (damit bekommt man

$$n = 4, 8, 12, 20, 24, 28, 32, 36, 44, 48, 52, 60, 68, 72, 76, 80, 84, \ldots\,;$$

die weiteren Werte

$$n = 16, 40, 56, 64, 88, \ldots$$

bekommt man aus der Multiplikativität; für $n = 92$ muss man sich schon etwas anderes überlegen). Die bekannten Konstruktionen decken nicht alle Fälle ab. Stand Juli 2021 ist $n = 668$ der kleinste ungelöste Fall.

28. Symmetrische und alternierende Potenzen

Wir erweitern die Definition von bilinearen Abbildungen auf Abbildungen mit (möglicherweise) mehr als zwei Argumenten.

DEF * multilineare Abbildung symmetrisch alternierend

28.1. **Definition.** Seien $V_1, V_2, \dots, V_n$ und W K-Vektorräume. Eine Abbildung $m\colon V_1 \times V_2 \times \dots \times V_n \to W$ heißt *(K-)multilinear*, wenn sie in jedem Argument linear ist, d.h., für jedes $j \in \{1, 2, \dots, n\}$ und für alle $v_i \in V_i$ ist die Abbildung

$$\begin{array}{ccccc} V_j & \longrightarrow & V_1 \times V_2 \times \dots \times V_n & \xrightarrow{m} & W \\ v & \longmapsto & (v_1, \dots, v_{j-1}, v, v_{j+1}, \dots, v_n) & & \end{array}$$

linear. Im Fall $W = K$ heißt m eine *Multilinearform.*

Eine multilineare Abbildung $m\colon V^n \to W$ (also mit $V_1 = V_2 = \ldots = V_n = V$) heißt *symmetrisch*, wenn für alle $v_1, \dots, v_n \in V$ und alle Permutationen $\sigma \in S_n$ gilt

$$m(v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}) = m(v_1, v_2, \dots, v_n)\,.$$

(Es kommt also nicht auf die Reihenfolge der Argumente an.)

Eine multilineare Abbildung $m\colon V^n \to W$ heißt *alternierend*, wenn

$$m(v_1, \dots, v_n) = \mathbf{0}$$

ist, sobald es $i \neq j$ gibt mit $v_i = v_j$. Daraus folgt

$$m(v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}) = \varepsilon(\sigma) m(v_1, v_2, \dots, v_n)$$

für alle $\sigma \in S_n$. ♢

Um Letzteres zu sehen, genügt es eine Transposition σ zu betrachten (denn jede Permutation ist Produkt von Transpositionen und das Vorzeichen ε ist multiplikativ). Wenn σ zum Beispiel 1 und 2 vertauscht, dann betrachten wir

$$\begin{aligned} \mathbf{0} &= m(v_1 + v_2, v_1 + v_2, v_3, \dots, v_n) - m(v_1, v_1, v_3, \dots, v_n) - m(v_2, v_2, v_3, \dots, v_n) \\ &= m(v_1, v_2, v_3, \dots, v_n) + m(v_2, v_1, v_3, \dots, v_n) \\ &= m(v_1, v_2, v_3, \dots, v_n) + m(v_{\sigma(1)}, v_{\sigma(2)}, v_{\sigma(3)}, \dots, v_{\sigma(n)})\,, \end{aligned}$$

woraus mit $\varepsilon(\sigma) = -1$

$$m(v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}) = \varepsilon(\sigma) m(v_1, v_2, \dots, v_n)$$

folgt.

BSP multilineare Abbildungen

28.2. **Beispiele.**

(1) Die Abbildung $K[X]^n \to K[X]$, $(p_1, \dots, p_n) \mapsto p_1 \cdots p_n$, ist eine symmetrische K-multilineare Abbildung.

(2) Die Determinante $(K^n)^n \to K$, $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \mapsto \det(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$, ist eine alternierende Multilinearform.

(3) Das Vektorprodukt $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} \times \boldsymbol{y}$, ist eine alternierende bilineare Abbildung. ♣

Analog zum Tensorprodukt von zwei Vektorräumen kann man das Tensorprodukt von n Vektorräumen definieren.

DEF Tensorprodukt

28.3. **Definition.** Seien $V_1, V_2, \ldots, V_n$ K-Vektorräume. Ein *Tensorprodukt* von $V_1, V_2, \ldots, V_n$ ist ein K-Vektorraum V zusammen mit einer multilinearen Abbildung $\mu\colon V_1 \times \cdots \times V_n \to V$ mit der folgenden universellen Eigenschaft: Für jeden K-Vektorraum W und jede multilineare Abbildung $m\colon V_1 \times \cdots \times V_n \to W$ gibt es genau eine lineare Abbildung $f\colon V \to W$ mit $f \circ \mu = m$. ♢

Wie üblich ist dieses Tensorprodukt eindeutig bis auf eindeutigen Isomorphismus. Die Existenz sieht man wie folgt:

LEMMA Existenz des allg. Tensorprodukts

28.4. **Lemma.** *Ist (V, μ) ein Tensorprodukt von $V_1, \ldots, V_n$, dann ist $V \otimes V_{n+1}$ mit der Abbildung $(v_1, \ldots, v_n, v_{n+1}) \mapsto \mu(v_1, \ldots, v_n) \otimes v_{n+1}$ ein Tensorprodukt von $V_1, \ldots, V_{n+1}$.*

Beweis. Wir müssen die universelle Eigenschaft nachprüfen. Sei dazu W ein Vektorraum und $m\colon V_1 \times \cdots \times V_n \times V_{n+1} \to W$ multilinear. Für einen zunächst fest gewählten Vektor $v_{n+1} \in V_{n+1}$ ist die Abbildung

$$m_{v_{n+1}}\colon V_1 \times \cdots \times V_n \longrightarrow W, \quad (v_1, \ldots, v_n) \longmapsto m(v_1, \ldots, v_n, v_{n+1})$$

multilinear, also gibt es (weil (V, μ) ein Tensorprodukt von $V_1, \ldots, V_n$ ist) eine eindeutig bestimmte lineare Abbildung $f_{v_{n+1}}\colon V \to W$ mit $f_{v_{n+1}} \circ \mu = m_{v_{n+1}}$. Dann ist die Abbildung $b\colon V \times V_{n+1} \to W$, $(v, v_{n+1}) \mapsto f_{v_{n+1}}(v)$, bilinear, also gibt es eine eindeutig bestimmte lineare Abbildung $f\colon V \otimes V_{n+1} \to W$ mit

$$\begin{aligned} m(v_1, \ldots, v_n, v_{n+1}) &= m_{v_{n+1}}(v_1, \ldots, v_n) = f_{v_{n+1}}\bigl(\mu(v_1, \ldots, v_n)\bigr) \\ &= b\bigl(\mu(v_1, \ldots, v_n), v_{n+1}\bigr) = f\bigl(\mu(v_1, \ldots, v_n) \otimes v_{n+1}\bigr). \end{aligned}$$

Damit ist die universelle Eigenschaft nachgewiesen. ❑

Induktion über die Anzahl n der betrachteten Vektorräume zeigt dann, dass immer ein Tensorprodukt existiert.

Man schreibt $V_1 \otimes V_2 \otimes \cdots \otimes V_n$ für „das" Tensorprodukt von $V_1, V_2, \ldots, V_n$ und $v_1 \otimes \cdots \otimes v_n$ für $\mu(v_1, \ldots, v_n)$. Das Lemma zeigt, dass diese Schreibweise mit der früher eingeführten (Weglassen von Klammern bei sukzessiven Tensorprodukten von je zwei Vektorräumen) kompatibel ist.

DEF Tensorpotenz

28.5. **Definition.** Ist $n \geq 1$ und V ein K-Vektorraum, dann schreiben wir $V^{\otimes n}$ für das Tensorprodukt $\underbrace{V \otimes V \otimes \cdots \otimes V}_{n \text{ Faktoren}}$; außerdem setzen wir $V^{\otimes 0} = K$. $V^{\otimes n}$ heißt die *n-te Tensorpotenz* von V. ♢

Wir interessieren uns nun dafür, durch einen geeigneten Vektorraum die symmetrischen bzw. alternierenden multilinearen Abbildungen $V^n \to W$ zu klassifizieren, analog dazu, wie das Tensorprodukt beliebige multilineare Abbildungen klassifiziert. Wegen der Eindeutigkeit von universellen Objekten verwenden wir im Folgenden den bestimmten Artikel („die" statt „eine").

DEF symmetrische Potenz

28.6. **Definition.** Sei V ein K-Vektorraum und sei $n \in \mathbb{N}$. Die *n-te symmetrische Potenz* von V ist ein K-Vektorraum $S^n V$ zusammen mit einer symmetrischen multilinearen Abbildung $\sigma\colon V^n \to S^n V$ (oft $\sigma(v_1, \ldots, v_n) = v_1 \cdot v_2 \cdots v_n$ geschrieben), sodass die folgende universelle Eigenschaft gilt: Zu jedem K-Vektorraum W und jeder symmetrischen multilinearen Abbildung $s\colon V^n \to W$ gibt es genau eine lineare Abbildung $f\colon S^n V \to W$ mit $f \circ \sigma = s$. ♢

DEF
alternierende Potenz

28.7. **Definition.** Sei V ein K-Vektorraum und sei $n \in \mathbb{N}$. Die *n-te alternierende Potenz* (oder *äußere Potenz*) von V ist ein K-Vektorraum $\bigwedge^n V$ zusammen mit einer alternierenden multilinearen Abbildung $\alpha\colon V^n \to \bigwedge^n V$ (die meist $\alpha(v_1, \ldots, v_n) = v_1 \wedge v_2 \wedge \cdots \wedge v_n$ geschrieben wird), sodass die folgende universelle Eigenschaft gilt: Zu jedem K-Vektorraum W und jeder alternierenden multilinearen Abbildung $a\colon V^n \to W$ gibt es genau eine lineare Abbildung $f\colon \bigwedge^n V \to W$ mit $f \circ \alpha = a$. $\diamondsuit$

Da es um spezielle multilineare Abbildungen geht, sollten sich $S^n V$ und $\bigwedge^n V$ irgendwie aus der Tensorpotenz $V^{\otimes n}$ konstruieren lassen. Das geht wie folgt:

SATZ
Konstruktion der symm. Potenz

28.8. **Satz.** *Sei V ein K-Vektorraum und sei $n \in \mathbb{N}$. Sei $U \subset V^{\otimes n}$ der Untervektorraum, der von allen Elementen der Form*

$$v_1 \otimes v_2 \otimes \cdots \otimes v_n - v_{\tau(1)} \otimes v_{\tau(2)} \otimes \cdots \otimes v_{\tau(n)}$$

erzeugt wird; dabei sind $v_1, v_2, \ldots, v_n \in V$ und $\tau \in S_n$. Sei $\pi\colon V^{\otimes n} \to V^{\otimes n}/U$ der kanonische Epimorphismus und $\mu\colon V^n \to V^{\otimes n}$ die kanonische multilineare Abbildung. Dann ist $(S^n V, \sigma) = (V^{\otimes n}/U, \pi \circ \mu)$ die n-te symmetrische Potenz von V.

Beweis. Wir überlegen uns erst einmal, dass $\pi \circ \mu$ tatsächlich eine symmetrische multilineare Abbildung ist. Die Multilinearität folgt daraus, dass μ multilinear und π linear ist. Die Symmetrie sieht man wie folgt: Seien $v_1, \ldots, v_n \in V$ und $\tau \in S_n$. Dann ist

$$\begin{aligned}(\pi \circ \mu)(v_{\tau(1)}, \ldots, v_{\tau(n)}) &= \pi(v_{\tau(1)} \otimes \ldots \otimes v_{\tau(n)}) \\ &= \pi(v_1 \otimes \ldots \otimes v_n) - \pi(\underbrace{v_1 \otimes \ldots \otimes v_n - v_{\tau(1)} \otimes \ldots \otimes v_{\tau(n)}}_{\in U}) \\ &= \pi(v_1 \otimes \ldots \otimes v_n) = (\pi \circ \mu)(v_1, \ldots, v_n)\,.\end{aligned}$$

Dabei haben wir verwendet, dass π auf U verschwindet.

Wir müssen noch die universelle Eigenschaft nachprüfen. Sei also W ein K-Vektorraum und sei $s\colon V^n \to W$ multilinear und symmetrisch. Aus der universellen Eigenschaft von $V^{\otimes n}$ folgt, dass es eine eindeutig bestimmte lineare Abbildung $\varphi\colon V^{\otimes n} \to W$ gibt mit $\varphi \circ \mu = s$.

$$\begin{tikzcd} V^n \arrow[d, "\mu"'] \arrow[dr, "s"] & \\ V^{\otimes n} \arrow[r, "\varphi"] \arrow[d, "\pi"'] & W \\ V^{\otimes n}/U \arrow[ur, dashed, "f"'] & \end{tikzcd}$$

Da s symmetrisch ist, gilt

$$\begin{aligned}&\varphi(v_1 \otimes v_2 \otimes \cdots \otimes v_n - v_{\tau(1)} \otimes v_{\tau(2)} \otimes \cdots \otimes v_{\tau(n)}) \\ &\qquad = s(v_1, v_2, \ldots, v_n) - s(v_{\tau(1)}, v_{\tau(2)}, \ldots, v_{\tau(n)}) = \mathbf{0}\end{aligned}$$

für alle Erzeuger von U, also ist $U \subset \ker(\varphi)$. Deshalb gibt es eine (dann auch eindeutig bestimmte) lineare Abbildung $f\colon V^{\otimes n}/U \to W$ mit $\varphi = f \circ \pi$, also $f \circ (\pi \circ \mu) = \varphi \circ \mu = s$. ❑

Für die alternierende Potenz funktioniert das analog.

28.9. **Satz.** *Sei V ein K-Vektorraum und sei $n \in \mathbb{N}$. Sei $U \subset V^{\otimes n}$ der Untervektorraum, der von allen Elementen der Form*

$$v_1 \otimes v_2 \otimes \cdots \otimes v_n \qquad \text{mit } v_i = v_j \text{ für zwei Indizes } i \neq j$$

erzeugt wird; dabei sind $v_1, v_2, \ldots, v_n \in V$. Sei $\pi\colon V^{\otimes n} \to V^{\otimes n}/U$ der kanonische Epimorphismus und $\mu\colon V^n \to V^{\otimes n}$ die kanonische multilineare Abbildung. Dann ist $(\bigwedge^n V, \alpha) = (V^{\otimes n}/U, \pi \circ \mu)$ die n-te alternierende Potenz von V.

SATZ
Konstruktion der alt. Potenz

Beweis. Wie im Beweis von Satz 28.8 überlegt man sich, dass $\pi \circ \mu$ eine alternierende multilineare Abbildung ist.

$$\begin{tikzcd}
V^n \arrow[d, "\mu"'] \arrow[dr, "a"] & \\
V^{\otimes n} \arrow[r, "\varphi"] \arrow[d, "\pi"'] & W \\
V^{\otimes n}/U \arrow[ur, dashed, "f"'] &
\end{tikzcd}$$

Wir müssen die universelle Eigenschaft nachprüfen. Sei also W ein K-Vektorraum und sei $a\colon V^n \to W$ multilinear und alternierend. Aus der universellen Eigenschaft von $V^{\otimes n}$ ergibt sich, dass es eine eindeutig bestimmte lineare Abbildung $\varphi\colon V^{\otimes n} \to W$ gibt mit $\varphi \circ \mu = a$. Da a alternierend ist, gilt

$$\varphi(v_1 \otimes v_2 \otimes \cdots \otimes v_n) = a(v_1, v_2, \ldots, v_n) = \mathbf{0}$$

für alle Erzeuger von U, also ist $U \subset \ker(\varphi)$. Deshalb gibt es eine (dann auch eindeutig bestimmte) lineare Abbildung $f\colon V^{\otimes n}/U \to W$ mit $\varphi = f \circ \pi$, also $f \circ (\pi \circ \mu) = \varphi \circ \mu = a$. ❑

Als nächstes überlegen wir uns, wie eine Basis von $S^n V$ bzw. $\bigwedge^n V$ aussieht, wenn wir eine Basis von V kennen. Als ersten Schritt leiten wir ein Kriterium dafür her, wann eine Multilinearform symmetrisch bzw. alternierend ist.

28.10. **Lemma.** *Sei V ein K-Vektorraum und $\phi\colon V^n \to K$ eine Multilinearform; sei weiter $B = (b_1, \ldots, b_m)$ eine Basis von V.*

LEMMA
Kriterium für symm. bzw. alt.

(1) *ϕ ist genau dann symmetrisch, wenn für alle $1 \le i_1 \le \ldots \le i_n \le m$ und alle $\sigma \in S_n$ gilt*

$$\phi(b_{i_{\sigma(1)}}, b_{i_{\sigma(2)}}, \ldots, b_{i_{\sigma(n)}}) = \phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n}).$$

(2) *ϕ ist genau dann alternierend, wenn für alle $1 \le i_1 < \ldots < i_n \le m$ und alle $\sigma \in S_n$ gilt*

$$\phi(b_{i_{\sigma(1)}}, b_{i_{\sigma(2)}}, \ldots, b_{i_{\sigma(n)}}) = \varepsilon(\sigma)\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n})$$

und außerdem

$$\phi(b_{j_1}, b_{j_2}, \ldots, b_{j_n}) = 0$$

ist, wenn zwei der Indizes übereinstimmen.

Beweis. Dass die Bedingungen notwendig sind, folgt unmittelbar aus den Definitionen. Es bleibt zu zeigen, dass sie auch hinreichend sind. Es ist erst einmal klar, dass aus den angegebenen Bedingungen dieselben Aussagen folgen ohne die Voraussetzung, dass die i_k monoton wachsend sind. Zur Vereinfachung schreiben wir $(v_1, \ldots, v_n)^\sigma$ für $(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$, wobei $v_1, \ldots, v_n \in V$ und $\sigma \in S_n$ sind.

(1) Wir müssen zeigen, dass $\phi(\underline{v}^\sigma) = \phi(\underline{v})$ gilt für alle $\underline{v} = (v_1, \ldots, v_n) \in V^n$ und alle $\sigma \in S_n$. Wir schreiben $v_j = \sum_{i=1}^m \lambda_{ij} b_i$, dann ist

$$\phi(\underline{v}) = \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1,1}\lambda_{i_2,2} \cdots \lambda_{i_n,n}\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n}).$$

Es folgt

$$\begin{aligned}
\phi(\underline{v}^\sigma) &= \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1,\sigma(1)}\lambda_{i_2,\sigma(2)}\cdots\lambda_{i_n,\sigma(n)}\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n}) \\
&\stackrel{(*)}{=} \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_{\sigma(1)},\sigma(1)}\lambda_{i_{\sigma(2)},\sigma(2)}\cdots\lambda_{i_{\sigma(n)},\sigma(n)}\phi(b_{i_{\sigma(1)}}, b_{i_{\sigma(2)}}, \ldots, b_{i_{\sigma(n)}}) \\
&\stackrel{(**)}{=} \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1,1}\lambda_{i_2,2}\cdots\lambda_{i_n,n}\phi(b_{i_{\sigma(1)}}, b_{i_{\sigma(2)}}, \ldots, b_{i_{\sigma(n)}}) \\
&= \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1,1}\lambda_{i_2,2}\cdots\lambda_{i_n,n}\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n}) \\
&= \phi(\underline{v})\,.
\end{aligned}$$

Bei $(*)$ haben wir die Indizes $i_1, \ldots, i_n$ mittels σ vertauscht, was nur einer Umordnung der Summe entspricht. Bei $(**)$ haben wir die Faktoren $\lambda_{i_k,k}$ in die „richtige“ Reihenfolge gebracht, was ihr Produkt nicht ändert. Am Schluss haben wir die Voraussetzung verwendet.

(2) Wir müssen zeigen, dass $\phi(\underline{v}) = 0$ ist, wenn in $\underline{v} = (v_1, \ldots, v_n)$ zwei Komponenten übereinstimmen. Wir nehmen an, dass $v_1 = v_2$ ist (der allgemeine Fall geht genauso). Wie oben schreiben wir die v_j als Linearkombination der Basis; es gilt dann $\lambda_{i1} = \lambda_{i2} = \lambda_i$. Sei $\tau \in S_n$ die Transposition, die 1 und 2 vertauscht; es ist $\varepsilon(\tau) = -1$. Damit erhalten wir

$$\begin{aligned}
\phi(\underline{v}) &= \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1}\lambda_{i_2}\lambda_{i_3,3}\cdots\lambda_{i_n,n}\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n}) \\
&= \sum_{i=1}^m \sum_{i_3=1}^m \cdots \sum_{i_n=1}^m \lambda_i^2\lambda_{i_3,3}\cdots\lambda_{i_n,n}\underbrace{\phi(b_i, b_i, b_{i_3}, \ldots, b_{i_n})}_{=0} \\
&\qquad + \sum_{1\le i_1<i_2\le m} \sum_{i_3=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1}\lambda_{i_2}\lambda_{i_3,3}\cdots\lambda_{i_n,n}\,\cdot \\
&\qquad\qquad \cdot \big(\underbrace{\phi(b_{i_1}, b_{i_2}, b_{i_3}, \ldots, b_{i_n}) + \phi(b_{i_2}, b_{i_1}, b_{i_3}, \ldots, b_{i_n})}_{=0}\big) \\
&= 0\,.
\end{aligned}$$

❑

SATZ
Basen von S^nV, $\bigwedge^n V$

28.11. **Satz.** *Sei V ein Vektorraum mit Basis $B = (b_1, b_2, \ldots, b_m)$.*

(1) *$S^nB = (b_{i_1} \cdot b_{i_2} \cdots b_{i_n})_{1\le i_1\le i_2\le\ldots\le i_n\le m}$ ist eine Basis von S^nV.*

(2) *$\bigwedge^n B = (b_{i_1} \wedge b_{i_2} \wedge \cdots \wedge b_{i_n})_{1\le i_1<i_2<\ldots<i_n\le m}$ ist eine Basis von $\bigwedge^n V$.*

Insbesondere ist

$$\dim S^nV = \binom{m+n-1}{n} \qquad \textit{und} \qquad \dim \textstyle\bigwedge^n V = \dbinom{m}{n}.$$

Beweis. Wie in Satz 27.4 sieht man, dass

$$B^{\otimes n} = (b_{i_1} \otimes \cdots \otimes b_{i_n})_{1\le i_1,\ldots,i_n\le m}$$

eine Basis von $V^{\otimes n}$ ist. Nach Satz 28.8 bzw. Satz 28.9 sind die kanonischen linearen Abbildungen

$$V^{\otimes n} \longrightarrow S^n V \qquad \text{bzw.} \qquad V^{\otimes n} \longrightarrow \bigwedge^n V$$

surjektiv. Da sie die Basis $B^{\otimes n}$ auf $S^n B$ bzw. $\pm \bigwedge^n B \cup \{\mathbf{0}\}$ abbilden, bilden $S^n B$ bzw. $\bigwedge^n B$ jedenfalls ein Erzeugendensystem von $S^n V$ bzw. $\bigwedge^n V$. Es bleibt zu zeigen, dass sie auch linear unabhängig sind. Dafür überlegen wir uns, dass es zu jedem Basiselement b eine Linearform ϕ_b auf $S^n V$ bzw. $\bigwedge^n V$ gibt, die auf b den Wert 1 und auf allen anderen Basiselementen den Wert 0 annimmt. Daraus folgt die lineare Unabhängigkeit: Seien nämlich λ_b Skalare mit $\sum_{b \in S^n B} \lambda_b b = \mathbf{0}$ (bzw. $\sum_{b \in \bigwedge^n B} \lambda_b b = \mathbf{0}$), dann folgt

$$0 = \phi_{b'}(\mathbf{0}) = \phi_{b'}\Bigl(\sum_b \lambda_b b\Bigr) = \sum_b \lambda_b \phi_{b'}(b) = \lambda_{b'}$$

für alle $b' \in S^n B$ (bzw. $b' \in \bigwedge^n B$).

Im Fall $S^n V$ seien $1 \le i_1 \le \ldots \le i_n \le m$ und $b \in S^n B$ das zugehörige Element. Wir definieren eine Linearform ϕ auf $V^{\otimes n}$, indem wir die Bilder der Elemente von $B^{\otimes n}$ festlegen:

$$\phi(b_{j_1} \otimes b_{j_2} \otimes \cdots \otimes b_{j_n}) = 1, \quad \text{falls es } \sigma \in S_n \text{ gibt mit } j_k = i_{\sigma(k)} \text{ für alle } k$$

und $= 0$ sonst. Dann ist $(v_1, \ldots, v_n) \mapsto \phi(v_1 \otimes \cdots \otimes v_n)$ nach Lemma 28.10 eine symmetrische Multilinearform auf V, induziert also eine Linearform ϕ_b auf $S^n V$, die die gewünschten Eigenschaften hat.

Im Fall $\bigwedge^n V$ seien $1 \le i_1 < \ldots < i_n \le m$ und $b \in \bigwedge^n B$ das zugehörige Element. Wir definieren wieder eine Linearform ϕ auf $V^{\otimes n}$, indem wir die Bilder der Elemente von $B^{\otimes n}$ festlegen:

$$\phi(b_{j_1} \otimes b_{j_2} \otimes \cdots \otimes b_{j_n}) = \varepsilon(\sigma), \quad \text{falls es } \sigma \in S_n \text{ gibt mit } j_k = i_{\sigma(k)} \text{ für alle } k$$

und $= 0$ sonst. Wieder nach Lemma 28.10 ist $(v_1, \ldots, v_n) \mapsto \phi(v_1 \otimes \cdots \otimes v_n)$ eine alternierende Multilinearform auf V, induziert also eine Linearform ϕ_b auf $\bigwedge^n V$, die die gewünschten Eigenschaften hat.

Die Formeln für die Dimensionen ergeben sich daraus, dass die Elemente von $\bigwedge^n B$ genau den n-elementigen Teilmengen der Menge $\{1, 2, \ldots, m\}$ entsprechen, und aus der Überlegung, dass die schwach monoton wachsenden Tupel $(i_1, \ldots, i_n)$ mittels der Abbildung

$$(i_1, \ldots, i_n) \longmapsto \{i_1, i_2 + 1, i_3 + 2, \ldots, i_n + n - 1\}$$

genau den n-elementigen Teilmengen von $\{1, 2, \ldots, m + n - 1\}$ entsprechen. ❑

Für $n > \dim V$ ist also $\bigwedge^n V = \{\mathbf{0}\}$, und $\bigwedge^{\dim V} V$ ist eindimensional.

28.12. **Beispiele.** Das Vektorprodukt $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} \times \boldsymbol{y}$, ist alternierend und bilinear und induziert deshalb eine lineare Abbildung $\bigwedge^2 \mathbb{R}^3 \longrightarrow \mathbb{R}^3$. Diese bildet $\mathbf{e}_1 \wedge \mathbf{e}_2$ auf $\mathbf{e}_3$, $\mathbf{e}_1 \wedge \mathbf{e}_3$ auf $-\mathbf{e}_2$ und $\mathbf{e}_2 \wedge \mathbf{e}_3$ auf $\mathbf{e}_1$ ab, ist also ein Isomorphismus. **BSP** Vektorprodukt Determinante

Die Determinante liefert entsprechend eine lineare Abbildung $\bigwedge^n(K^n) \to K$. Dabei wird der Erzeuger $\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \cdots \wedge \mathbf{e}_n$ von $\bigwedge^n(K^n)$ auf $1 \in K$ abgebildet (denn $\det(I_n) = 1$), also ist diese Abbildung ebenfalls ein Isomorphismus. ♣

Eine wichtige Eigenschaft des alternierenden Produkts $v_1 \wedge \cdots \wedge v_n$ ist, dass man damit lineare Unabhängigkeit testen kann.

SATZ lin. Unabh. über $\bigwedge^n V$

28.13. **Satz.** *Seien V ein Vektorraum und $v_1, \ldots, v_n \in V$. Dann ist $(v_1, \ldots, v_n)$ genau dann linear unabhängig, wenn $v_1 \wedge \cdots \wedge v_n \neq \mathbf{0}$ in $\bigwedge^n V$ ist.*

Beweis. Ist $(v_1, \ldots, v_n)$ linear abhängig, dann ist einer der Vektoren eine Linearkombination der übrigen. Sei etwa $v_n = \lambda_1 v_1 + \ldots + \lambda_{n-1} v_{n-1}$; dann ist

$$v_1 \wedge \cdots \wedge v_{n-1} \wedge v_n = \sum_{i=1}^{n-1} \lambda_i v_1 \wedge \cdots \wedge v_{n-1} \wedge v_i = \mathbf{0}\,,$$

da in jedem Summanden im alternierenden Produkt zwei gleiche Argumente stehen. (Das Argument funktioniert für jede alternierende multilineare Abbildung.)

Ist $(v_1, \ldots, v_n)$ linear unabhängig und ist V endlich-dimensional, dann können wir $(v_1, \ldots, v_n)$ zu einer Basis von V ergänzen; nach Satz 28.11 ist dann $v_1 \wedge \cdots \wedge v_n$ Element einer Basis von $\bigwedge^n V$ und damit insbesondere nicht null.

Im allgemeinen Fall zeigt diese Überlegung, dass $v_1 \wedge \cdots \wedge v_n \neq \mathbf{0}$ ist in $\bigwedge^n U$ mit $U = \langle v_1, \ldots, v_n \rangle$. Dann gibt es eine alternierende Multilinearform ϕ auf U^n mit $\phi(v_1, \ldots, v_n) \neq 0$. Sei U' ein Komplement von U in V und $p\colon V \to U$ die zugehörige Projektion; dann ist $m\colon V^n \xrightarrow{p^n} U^n \xrightarrow{\phi} K$ eine alternierende Multilinearform auf V^n mit $m(v_1, \ldots, v_n) \neq 0$. Daraus folgt $v_1 \wedge \cdots \wedge v_n \neq \mathbf{0}$ in $\bigwedge^n V$. ❑

Lineare Abbildungen $V \to W$ induzieren lineare Abbildungen zwischen den symmetrischen und alternierenden Potenzen.

LEMMA $S^n f$, $\bigwedge^n f$

28.14. **Lemma.** *Seien V und W zwei K-Vektorräume und $f\colon V \to W$ eine lineare Abbildung. Für jedes $n \in \mathbb{N}$ gibt es dann eindeutig bestimmte lineare Abbildungen $S^n f\colon S^n V \to S^n W$ und $\bigwedge^n f\colon \bigwedge^n V \to \bigwedge^n W$ mit*

$$\big(S^n f\big)(v_1 \cdots v_n) = f(v_1) \cdots f(v_n) \quad \textit{bzw.} \quad \big(\textstyle\bigwedge^n f\big)(v_1 \wedge \cdots \wedge v_n) = f(v_1) \wedge \cdots \wedge f(v_n)$$

für alle $v_1, \ldots, v_n \in V$.

Beweis. Wir beweisen die Aussage über $\bigwedge^n f$, die andere zeigt man analog. Da f linear ist, ist

$$V^n \longrightarrow \textstyle\bigwedge^n W, \quad (v_1, \ldots, v_n) \longmapsto f(v_1) \wedge \cdots \wedge f(v_n)$$

eine alternierende multilineare Abbildung. Die universelle Eigenschaft von $\bigwedge^n V$ liefert dann die gewünschte Abbildung $\bigwedge^n f$ und zeigt, dass sie eindeutig bestimmt ist. ❑

Wie beim Tensorprodukt gilt dann natürlich auch $S^n(g \circ f) = (S^n g) \circ (S^n f)$ und $\bigwedge^n (g \circ f) = (\bigwedge^n g) \circ (\bigwedge^n f)$.

Als (krönenden?) Abschluss dieses Kapitels zeigen wir, wie man eine Basis-freie Definition der Determinante bekommen kann.

SATZ Determinante Basis-frei

28.15. **Satz.** *Sei V ein K-Vektorraum mit $\dim V = n < \infty$. Für $f \in \mathrm{End}(V)$ gilt dann*

$$\textstyle\bigwedge^n f = \det(f)\, \mathrm{id}_{\bigwedge^n V}\ .$$

Beweis. Wir erinnern uns daran, dass $\dim \bigwedge^n V = 1$ ist. Also ist jeder Endomorphismus von $\bigwedge^n V$ durch Multiplikation mit einem Skalar gegeben. Wir müssen zeigen, dass für $\bigwedge^n f \in \operatorname{End}(\bigwedge^n V)$ dieser Skalar $\lambda(f)$ gerade $\det(f)$ ist. Sei $(b_1, \ldots, b_n)$ eine Basis von V. Dann ist $b_1 \wedge \cdots \wedge b_n$ ein von null verschiedenes Element von $\bigwedge^n V$ (Satz 28.11), also ist $\lambda(f)$ durch

$$f(b_1) \wedge \cdots \wedge f(b_n) = \bigl(\textstyle\bigwedge^n f\bigr)(b_1 \wedge \cdots \wedge b_n) = \lambda(f)\, b_1 \wedge \cdots \wedge b_n\,.$$

eindeutig festgelegt. Daran sieht man, dass $\lambda(f)$ linear in jedem $f(b_j)$ ist und verschwindet, wenn $f(b_i) = f(b_j)$ ist für $i \neq j$. Außerdem ist $\lambda(\mathrm{id}_V) = 1$. Wenn man $f(b_j)$ über die Basisdarstellung mit der j-ten Spalte von $\mathrm{Mat}_B(f)$ identifiziert, dann sind das gerade die Eigenschaften, die die Determinante charakterisieren (vergleiche Satz 14.3). Also muss $\lambda(f) = \det(f)$ sein. ❑

Als Folgerung erhalten wir einen ganz schmerzlosen Beweis der Multiplikativität der Determinante.

28.16. **Folgerung.** *Sei V ein endlich-dimensionaler Vektorraum und seien f und g Endomorphismen von V. Dann gilt* $\det(g \circ f) = \det(g)\det(f)$. **FOLG** det ist multiplikativ

Beweis. Sei $\dim V = n$ und $\mathbf{0} \neq w \in \bigwedge^n V$. Es gilt nach Satz 28.15:

$$\begin{aligned}\det(g \circ f)w &= \bigl(\textstyle\bigwedge^n (g \circ f)\bigr)(w) = \bigl(\textstyle\bigwedge^n g\bigr)\bigl((\textstyle\bigwedge^n f)(w)\bigr)\\ &= \bigl(\textstyle\bigwedge^n g\bigr)(\det(f)w) = \det(g)\det(f)w\,.\end{aligned}$$

Aus $w \neq \mathbf{0}$ folgt $\det(g \circ f) = \det(g)\det(f)$. ❑

Die universelle Eigenschaft von $\bigwedge^n V$ besagt, dass der Raum der alternierenden multilinearen Abbildungen $V^n \to K$ kanonisch isomorph zu $(\bigwedge^n V)^*$ ist. Im Fall $V = \mathbb{R}^m$ kann man $\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_n \in \bigwedge^n \mathbb{R}^m$ als ein Objekt auffassen, dass das von $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ im $\mathbb{R}^m$ aufgespannte Parallelotop repräsentiert. Eine Linearform in $(\bigwedge^n \mathbb{R}^m)^*$ weist dann jedem solchen (orientierten) Parallelotop eine Zahl zu, die man als zum Beispiel im Fall $n = 2$ und $m = 3$ als den darauf entfallenden Durchfluss einer strömenden Flüssigkeit interpretieren kann (ob man den positiv oder negativ zählt, hängt von der Orientierung des Parallelogramms und der Richtung des Flusses ab). Will man zum Beispiel den Gesamtfluss durch ein Flächenstück F im $\mathbb{R}^3$ bestimmen, so muss man den Durchfluss durch viele kleine Flächenstücke zusammenzählen (und einen geeigneten Grenzübergang durchführen). An jedem Punkt $\boldsymbol{x}$ von $\mathbb{R}^3$ braucht man ein $\phi_{\boldsymbol{x}} \in (\bigwedge^2 \mathbb{R}^3)^*$, das sagt, welchen Durchfluss (kleine) Parallelogramme in der Nähe von $\boldsymbol{x}$ haben sollen. So eine Zuordnung $\phi\colon \boldsymbol{x} \mapsto \phi_{\boldsymbol{x}}$ heißt eine *Differentialform* (hier wäre es eine 2-Form ($n = 2$) auf dem $\mathbb{R}^3$). Den Fluss berechnet man, indem man „ϕ über F integriert".

Für endlich-dimensionale Vektorräume V gilt $(\bigwedge^n V)^* \cong \bigwedge^n (V^*)$ (Übung), sodass man diese beiden Räume identifizieren kann. Die zu $(\mathbf{e}_1, \ldots, \mathbf{e}_m)$ duale Basis von $(\mathbb{R}^m)^*$ schreibt man in diesem Zusammenhang üblicherweise $(dx_1, \ldots, dx_m)$. Eine 2-Form ϕ wie oben hat dann etwa die Gestalt

$$\phi_{\boldsymbol{x}} = f_{12}(\boldsymbol{x})\, dx_1 \wedge dx_2 + f_{13}(\boldsymbol{x})\, dx_1 \wedge dx_3 + f_{23}(\boldsymbol{x})\, dx_2 \wedge dx_3$$

mit Funktionen $f_{12}, f_{13}, f_{23}\colon \mathbb{R}^3 \to \mathbb{R}$. Man kann den Vektor $\bigl(f_{23}(\boldsymbol{x}), -f_{13}(\boldsymbol{x}), f_{12}(\boldsymbol{x})\bigr)$ als Geschwindigkeit mal Massendichte der Strömung an der Stelle $\boldsymbol{x}$ interpretieren: Die $\mathbf{e}_1$-Komponente der Strömung macht sich nicht auf den parallel dazu ausgerichteten Parallelogrammen $\mathbf{e}_1 \wedge \mathbf{e}_2$ und $\mathbf{e}_1 \wedge \mathbf{e}_3$ bemerkbar, sondern nur auf $\mathbf{e}_2 \wedge \mathbf{e}_3$; deswegen entspricht sie f_{23}. Das andere Vorzeichen bei f_{13} hat mit der Orientierung zu tun: Sind $\boldsymbol{x}_1, \boldsymbol{x}_2$ und der „Flussvektor" $\boldsymbol{y}$ in dieser Reihenfolge positiv orientiert, dann soll der Fluss positiv sein (und zwar soll gerade $\det(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ herauskommen).

Eine 3-Form
$$\eta_{\boldsymbol{x}} = g(\boldsymbol{x})\, dx_1 \wedge dx_2 \wedge dx_3$$
liefert eine Massen- oder Ladungsdichte. Zum Beispiel erhält man (für $g = 1$) als Integral von $dx_1 \wedge dx_2 \wedge dx_3$ über eine Teilmenge $A \subset \mathbb{R}^3$ gerade das Volumen von A. Näheres dazu gibt es in der Vektoranalysis.