UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Lineare Algebra 1

Jochen Glück

Manuskript
Wintersemester 2021/22

Inhaltsverzeichnis

Vorwort

Aufbau des Manuskripts

Der Aufbau des Manuskripts ist einfach:

- Es ist unterteilt in Kapitel (nummeriert als 1, 2, ...), und jedes Kapitel ist unterteilt in Abschnitte (nummeriert als 1.1, 1.2, ...).

- Definitionen, Sätze, Bemerkungen, usw. sind innerhalb der Abschnitte mit einer gemeinsamen Nummer durchlaufend nummeriert.

- Am Anfang jedes Kapitels finden Sie eine Liste mit *Fragen zum Einstieg*; sie dienen zur Motivation des Stoffes in diesem Kapitel.

  Bitte lesen Sie diese Fragen jeweils in Ruhe durch und denken Sie ein wenig darüber nach. Dies wird Ihnen beim Verständnis und der Einordnung des Stoffes helfen.

- Am Ende jedes Kapitels steht ein nicht-nummerierter Abschnitt mit dem Titel *Literaturhinweise*. Dort sind stets einige Bücher angegeben, in denen die Inhalte des Kapitels ebenfalls behandelt werden. Wenn Sie Schwierigkeiten beim Verständnis eines Kapitels haben, kann es oft hilfreich sein, einen Blick in manche dieser Bücher zu werfen – verschiedene Autorinnen und Autoren stellen dasselbe Thema nämlich oft unterschiedlich dar, und eine weitere Perspektive kann Ihnen manchmal das Verständnis erleichtern.

  Alle Bücher, die in den Abschnitten mit Literaturhinweisen referenziert werden, sind im Literaturverzeichnis am Endes des Manuskripts aufgelistet.

  Neben verschiedenen Büchern enthalten die Abschnitte mit Literaturhinweisen manchmal auch noch Hinweise auf weitere Quellen – zum Beispiel Links auf Erklärungen auf bestimmten Internetseiten oder online verfügbare Videos.

- Vor dem Abschnitt mit Literaturhinweisen befindet sich in manchen Kapiteln ein Abschnitt mit dem Titel *Ergänzungen*. In diesen Abschnitten finden Sie Zusatzinformationen, die nicht Teil des Vorlesungsstoffes sind und auch nicht in der Vorlesung behandelt werden.

Diese Ergänzungen können Ihnen aber beim Verständnis helfen oder Ihre Neugier auf weitere verwandte Themen wecken. Bitte entscheiden Sie selbst, welche Ergänzungen Sie lesen möchten und welche nicht.

Vorsicht: Fehler!

Dieses Manuskript enthält mit an Sicherheit grenzender Wahrscheinlichkeit Fehler. Mit hoher Wahrscheinlichkeit enthält es sogar viele Fehler. Wenn Sie glauben, einen Fehler entdeckt zu haben: Geben Sie mir bitte Bescheid (wirklich!), am einfachsten per E-Mail.

E-Mail-Adresse (geändert seit März 2022): glueck@uni-wuppertal.de

Danksagung

Ich bedanke mich bei allen Personen, die zur Erstellung dieses Manuskripts beigetragen haben: insbesondere bei Isabella Habereder für ihre Unterstützung beim TeXen mehrere Teilabschnitte, bei allen Studierenden und allen Mitarbeiterinnen und Mitarbeitern, die mir Fehler und Verbesserungsvorschläge mitgeteilt haben, und bei Maximilian Reif, der sich lange nach Ende der Veranstaltung nochmals die Mühe gemacht hat, mir eine sehr ausführliche Liste mit Korrekturen zu schicken.

Kapitel 1

Grundlegende Konzepte in der Mathematik

Einstiegsfragen. (a) Schauen Sie sich die folgende Aussage an: „Alle Mathe-Dozenten sind Nerds.“

Wie lautet die Verneinung dieser Aussage? Können Sie die Verneinung formulieren, ohne die Worte “alle” und “jeder” (bzw. Synonyme dieser Worte) zu verwenden?

(b) Können Sie in einfachen Worten erklären, was die Vereinigung von zwei Mengen ist?

(c) Was sind eigentlich Variablen, und wozu braucht man sie?

(d) Finden Sie eine Gemeinsamkeit zwischen den folgenden fünf Dingen? (i) Ein Telefonbuch; (ii) die täglichen CoV19-Infektionszahlen über die vergangenen sechs Monate; (iii) der Sachindex dieses Manuskripts; (iv) eine aktuelle Tabelle der Fußballbundesliga; (v) ein Radiosignal, dass Sie mit einem Autoradio empfangen.

1.1 Exaktheit in der Mathematik

Ein zentrales und besonderes Merkmal der Wissenschaft *Mathematik* ist, dass mathematische Resultate nicht empirisch – also beruhigend auf realen Beobachtungen und Experimenten – gewonnen werden, sondern dass alle Behauptungen stets zweifelsfrei zu beweisen sind. Dies ist nur möglich, indem eine Reihe von Regeln konsequent beachtet wird:

- Wann immer man einen mathematischen Begriff verwendet, muss man präzise beschreiben, wie dieser Begriff zu verstehen ist. Dies ist der Sinn von *Definitionen*.

- Mathematische Resultate bestehen aus *Annahmen* und *Konklusionen*, und sind von der Form: Wenn die Annahmen alle erfüllt sind, dann sind auch die Konklusionen wahr.

  Je nach Kontext werden mathematische Resultate oft als *Theorem*, *Satz*, *Lemma*, *Proposition* oder *Korollar* bezeichnet.[1] Sie werden im Laufe der Vorlesung noch häufig sehen, welcher dieser Begriffe für welche Art von mathematischem Resultat verwendet wird.

- Mathematische Resultate werden stets sauber getrennt von Ihrer Begründung: Zunächst schreibt man das Resultat auf, anschließend folgt der Beweis des Resultats. Der Zweck des Beweises ist es, darzulegen, dass aus den Annahmen des Resultats die Konklusionen logisch folgen. Erst nach dem Beweis darf man davon ausgehen, dass das Resultat tatsächlich gültig ist.

- Wichtig: Das Wort „logisch“ im vorangehenden Punkt darf auf keinen Fall umgangssprachlich im Sinne von „plausibel“ verstanden werden. Mit dem Begriff *logisch folgen* meint man in der Mathematik stattdessen eine Situation, in der die Richtigkeit der Annahmen zwangsläufig und in jedem Fall dazu führt, dass auch die Konklusionen richtig sind.

  Damit es hier nicht zu Missverständnissen oder Ungenauigkeiten kommt, bedient man sich der *Aussagenlogik*, in der genau festgelegt ist, wie man mit mathematischen Aussagen umgehen darf, und wann eine Aussage logisch aus einer anderen folgt.

Wir werden uns in den nachfolgenden Abschnitten 1.2 und 1.4 mit der Aussagenlogik beschäftigen.

Außerdem werden Sie in den weiteren Abschnitten dieses Kapitels einige weitere Konzepte lernen, die sowohl in der Linearen Algebra als auch in allen anderen mathematischen Teilgebieten eine wichtige Rolle spielen: *Mengen und Tupel* (Abschnitt 1.3) dienen dazu, mehrere mathematische Objekte (z.B. mehrere Zahlen) zu einem größeren Objekt zusammenzufassen. *Funktionen* (Abschnitt 1.5) sind wichtig, um verschiedene mathematische Objekte zueinander in Beziehung setzen zu können.

1.2 Aussagenlogik

In der *Aussagenlogik* geht es darum, Aussagen über mathematische Objekte zu treffen und festzulegen, wie man mit diesen Aussagen arbeiten darf.

[1] Wobei die Unterscheidung zwischen diesen Begriffen recht subjektiv und häufig auch etwas willkürlich ist. Dies ist aber kein Problem, da es für den Inhalt eines mathematischen Resultats nicht von Bedeutung ist, ob man das Resultat z.B. als Theorem oder als Lemma bezeichnet.

Was ist eine Aussage?

In der Mathematik möchte man stets wahre Dinge über mathematische Objekte sagen – dazu muss man also stets sauber unterscheiden können, was wahr und was falsch ist. Formulierungen, bei denen diese Unterscheidung möglich ist, bezeichnen wir als *Aussagen*:

Definition 1.2.1 (Aussagen und Wahrheitswerte)**.** (a) Eine **Aussage** ist ein sprachlicher Ausdruck[2], von dem eindeutig feststeht, ob er wahr oder falsch ist.

(b) Wenn eine Aussage wahr ist, dann sagen wir, sie hat den **Wahrheitswert** „wahr“; wenn sie hingegen falsch ist, dann sagen wir, sie hat den **Wahrheitswert** „falsch“.

Es folgen sowohl einige alltägliche Beispiele als auch einige mathematische Beispiele.

Beispiele 1.2.2. (a) Der Ausdruck „*Friedrich Schiller*“ ist keine Aussage, denn es ergibt keinen Sinn zu sagen, dass eine Person war oder falsch sei.

(b) Der Ausdruck „*Friedrich Schiller hat das Drama ‘Der Götz von Berlichingen’ geschrieben.*“ ist eine Aussage. Sie ist falsch[3], hat also den Wahrheitswert „falsch“.

(c) Der Ausdruck „*Friedrich Schiller hat das Drama ‘Die Räuber’ geschrieben.*“ ist eine Aussage. Sie ist wahr, hat also den Wahrheitswert „wahr“.

(d) Der Ausdruck „*Friedrich Schiller hat vielleicht die Ballade ‘Die Bürgschaft’ geschrieben*“ ist keine Aussage[4], denn aufgrund des Wortes „vielleicht“ ist es nicht möglich zu sagen, ob der Ausdruck wahr oder falsch ist.

(e) Der Ausdruck „$2 + 2$“ ist keine Aussage, denn „$2 + 2$“ ist weder wahr noch falsch.

(f) Der Ausdruck „$2 + 2 = 5 - 1$“ ist eine Aussage. Sie ist wahr, hat also den Wahrheitswert „wahr“.

(g) Der Ausdruck „$2+2 = 4$“ ist eine Aussage. Sie hat ebenfalls den Wahrheitswert „wahr“.

(h) Der Ausdruck „$2 + 2 = 5 + 2$“ ist eine Aussage. Sie hat den Wahrheitswert „falsch“.

[2] Man könnte auch sagen: Eine sprachliche Formulierung.
[3] Das Drama ‘Der Götz von Berlichingen’ wurde von Goethe verfasst.
[4] Im mathematischen Sinne.

Inhalt der Mathematik ist es, wahre Aussagen über mathematische Objekte zu machen und zu beweisen. Deshalb hat man wenig Lust, zu jeder wahren Aussage stets explizit dazuzuschreiben, dass sie wahr ist. Denken Sie zum Beispiel an die Aussage aus Beispiel 1.2.2(c): Wenn Sie einem Kommilitonen mitteilen möchten, dass Schiller 'Die Räuber' geschrieben hat, dann sagen Sie nicht etwa

Die Aussage „*Friedrich Schiller hat das Drama 'Die Räuber' geschrieben.*" ist wahr.

Sondern Sie sagen einfach:

Friedrich Schiller hat das Drama 'Die Räuber' geschrieben.

Damit wollen Sie selbstverständlich zum Ausdruck bringen, dass diese Aussage wahr ist. Ebenso hält man es auch in der Mathematik:

Vereinbarung 1.2.3. Wenn wir eine Aussage machen ohne Ihren Wahrheitsgehalt zu spezifizieren, so wollen wir damit zum Ausdruck bringen, dass die Aussage wahr ist.

Verknüpfungen von Aussagen und Wahrheitstabellen

Einzelne Aussagen sind aus mathematischer Sicht recht unspektakulär. Interessant wird es erst, wenn man mehrere Aussagen verknüpft – das bedeutet, man baut aus mehreren gegebenen Aussagen eine neue. Ein einfaches Beispiel für solche eine Verknüpfung ist die *Verundung*[5] von zwei Aussagen. Sehen wir uns das zunächst an einem Beispiel an:

Beispiel 1.2.4. Die beiden Aussagen *Friedrich Schiller hat das Drama 'Die Räuber' geschrieben* und *Johann Wolfgang von Goethe hat die Ballade 'Der Erlkönig' geschrieben* können wir mit Hilfe des Wortes „und" zu einer neuen Aussage verknüpfen:

Friedrich Schiller hat das Drama 'Die Räuber' geschrieben und Johann Wolfgang von Goethe hat die Ballade 'Der Erlkönig'.

Diese neue Aussage ist wahr, weil die beiden Aussagen, mit denen wir begonnen hatten, wahr sind.

Dasselbe kann man freilich auch für beliebige andere Aussagen machen: Wenn A and B Aussagen sind, dann können wir daraus eine neue Aussage „A und B" konstruieren.[6]

An dieser Stelle sollten Sie kurz innehalten und an Abschnitt 1.1 zurückdenken: Als aller erstes haben wir besprochen, dass jeder Begriff, den wir verwenden, sauber definiert werden muss, damit wir stets genau wissen, wovon wir eigentlich sprechen:

[5] Häufig auch *Konjunktion* genannt.

[6] Hier sehen Sie bereits eine Vorgehensweise, die in der Mathematik sehr üblich und sehr nützlich ist: Wenn man etwas nicht nur für ein konkretes Beispiel tun möchte, sondern ganz allgemein, so verwendet man Platzhalter – häufig bezeichnet man diese Platzhalter mit Buchstaben, wie hier A und B.

Wenn also A und B zwei Aussagen sind, was genau ist dann mit der und-Verknüpfung „A und B“ gemeint?

Wir meinen mit „A und B“ eine neue Aussage, und damit keine Meinungsverschiedenheit darüber auftreten kann, ob diese Aussage wahr oder falsch ist, müssen wir *definieren*, welchen Wahrheitswert „A und B“ hat. Dies wird natürlich davon abhängen, welchen Wahrheitswert A und B jeweils haben. Allerdings gibt es für die Wahrheitswerte von A und B nur vier mögliche Fälle, und wenn wir für jeden dieser Fälle den Wahrheitswert von „A und B“ festlegen, dann ist immer eindeutig bestimmt, ob „A und B“ wahr oder falsch ist. Also tun wir genau das:

Definition 1.2.5 (und-Verknüpfung)**.** Seien A und B beliebige Aussagen. Wir definieren eine neue Aussage, die wir als $A \wedge B$ notieren und als „A und B“ aussprechen, in dem wir die folgenden Wahrheitswerte für $A \wedge B$ festlegen:

- Wenn A wahr ist und B wahr ist, dann ist $A \wedge B$ ebenfalls wahr.
- Wenn A wahr ist und B falsch ist, dann ist $A \wedge B$ falsch.
- Wenn A falsch ist und B wahr ist, dann ist $A \wedge B$ falsch.
- Wenn A falsch ist und B falsch ist, dann ist $A \wedge B$ ebenfalls falsch.

Die Aussage $A \wedge B$ wird als **Konjunktion** von A und B bezeichnet.

Diese Definition orientiert sich an der Bedeutung des Wortes „und“ in der Alltagssprache: „Es gilt A und B“ ist in der Alltagssprache gleichbedeutend mit „Es gilt sowohl A als auch B“. Diese Aussage ist wahr, wenn A, B beide wahr sind, sie ist jedoch falsch, wenn mindestens eine der beiden Aussagen A, B falsch ist.

Dies ist eine passende Stelle für einige allgemeine Kommentare zu Definitionen in der Mathematik:

Bemerkungen 1.2.6. (a) Grundsätzlich darf man definieren, was immer man möchte – solange die Definition sich nicht selbst widerspricht. Wenn man einen mathematischen Begriff oder ein mathematisches Symbol definiert, muss man sich also nicht zwangsläufig daran halten, wie dieser Begriff oder dieses Symbol in der alltäglichen Sprache verwendet werden.

Wichtig dabei ist natürlich: Jeder Begriff hat dann natürlich auch nur die Bedeutung, die ihm in seiner Definition zugewiesen wurde. Ob der Begriff in der Alltagssprache anders verwendet wird, spielt somit bei der Verwendung des Begriffs in der Mathematik keine Rolle – man hat sich einzig und allein an seine Definition zu halten.

(b) Trotzdem würde es vermutlich große Verwirrung stiften, wenn man einen Begriff so definiert, da seine mathematischen Definition der alltagssprachlichen Bedeutung des Begriffs stark zuwider läuft. Deshalb versucht man, Begriffe so zu definieren, dass sich die Definition eines Begriffs zumindest grob an seiner alltagssprachlichen Bedeutung orientiert. Wie oben erläutert, haben wir das auch in Definition 1.2.5 so gemacht.

(c) Beachten Sie trotzdem unbedingt: Ausschlaggebend für die Bedeutung eines mathematischen Begriffs ist letztlich immer seine mathematische Definition! Wenn Sie also unsicher sind, was ein mathematischer Begriff, den wir verwenden, bedeutet, dann lautet der erste Ratschlag immer: Blättern Sie in Ihren Unterlagen zurück schlagen Sie die Definition des Begriffs nach!

Definition 1.2.5 ist ziemlich ausladend formuliert und wenig übersichtlich. Denselben Inhalt kann man aber auch übersichtlicher darstellen, indem man einfach eine Tabelle verwendet: Definition 1.2.5 besagt, dass die Wahrheitswerte von $A \wedge B$ folgendermaßen lauten:

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

Hierbei haben wir „wahr“ mit dem Buchstaben „w“ abgekürzt und „falsch“ mit dem Buchstaben „f“. Eine Tabelle von obenstehender Form nennt man aus naheliegenden Gründen **Wahrheitstabelle**.

Wir definieren nun noch weitere Verknüpfungen von zwei Aussagen, sowie die Verneinung von Aussagen. Um die Definition dieser Begriffe übersichtlicher zu gestalten als in Definition 1.2.5, schreiben wir sie direkt mit Hilfe von Wahrheitstabellen auf.

Definition 1.2.7 (oder-Verknüpfungen und Negation)**.** Seien A und B beliebige Aussagen. Wir definieren drei weitere Aussagen $A \vee B$, $A \dot{\vee} B$, $\neg A$, deren Werte durch die folgenden Wahrheitstabellen festgelegt werden:

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

A	B	$A \dot{\vee} B$
w	w	f
w	f	w
f	w	w
f	f	f

A	$\neg A$
w	f
f	w

Für die so definierten Aussagen verwenden wir die folgenden Bezeichnungen:

(a) Die Aussage $A \vee B$ wird als „A inklusiv oder B“ ausgesprochen und als **inklusive Disjunktion** von A und B bezeichnet; oft sagt man anstelle „A inklusiv oder B“ auch einfach „A oder B“.[7]

(b) Die Aussage $A \dot{\vee} B$ wird als „A exklusiv oder B“ ausgesprochen und als **exklusive Disjunktion** von A und B bezeichnet.[8]

[7] Das bedeutet, das Wort „oder“ wird in der Mathematik, solange man nichts anderes dazu sagt, immer als „inklusives Oder“ verstanden.

[8] Neben $\dot{\vee}$ gibt es noch weitere gebräuliche Symbole für das exklusive Oder. Zum Beispiel ist es in der Informatik üblich anstelle von $A \dot{\vee} B$ die Notation $A \operatorname{xor} B$ zu verwenden.

(c) Die Aussage $\neg A$ wir als „nicht A“ ausgesprochen und als **Verneinung** oder **Negation** von A bezeichnet.

Bisher haben wir nur Begriffe und Symbole definiert. Als nächstes wollen wir ein paar mathematische Resultate aus der Aussagenlogik besprechen, und uns überzeugen, dass diese tatsächlich stimmen.

Proposition 1.2.8. *Sei A eine beliebige Aussage. Dann gilt immer (d.h. unabhängig vom Wahrheitswerte von A):*[9]

(a) *Die Aussage $A \vee (\neg A)$ ist wahr.*[10]

(b) *Die Aussage $A \wedge (\neg A)$ ist falsch.*[11]

(c) *Die Aussage $A \dot{\vee} (\neg A)$ ist wahr.*

Bevor Sie weiterlesen, sollten Sie sich auf jeden Fall überlegen, warum die Resultate in Proposition 1.2.8 intuitiv wirklich so erwarten würde. Wie Sie schon wissen, reicht Intuition aber in der Mathematik als Begründung nicht aus – wir wollen absolut sicher sein, dass Proposition 1.2.8 richtig ist. Deshalb **beweisen** wir die Proposition nun mit Hilfe von Wahrheitstabellen:

Beweis von Proposition 1.2.8. Die Aussage A hat genau einen der Wahrheitswerte „wahr“ oder „falsch“. Wenn wir also diese beiden Fälle betrachten, und in jedem der Fälle den Wahrheitswert der drei Aussagen $A \vee (\neg A)$, $A \wedge (\neg A)$, $A \dot{\vee} (\neg A)$ bestimmen, können wir sicherstellen, dass diese drei Aussagen wirklich immer den Wahrheitswert „wahr“ besitzen. Am übersichtlichsten lässt sich dies mit Hilfe einer Wahrheitstabelle darstellen:

A	$\neg A$	$A \vee (\neg A)$	$A \wedge (\neg A)$	$A \dot{\vee} (\neg A)$
w	f	w	f	w
f	w	w	f	w

Die Einträge in der Tabelle sind wie folgt zustande gekommen:

- Die Definition einer Aussage (Definition 1.2.1) besagt, dass A genau einen der beiden Wahrheitswerte „wahr“ und „falsch“ besitzt. Also können nur die beiden Fälle auftreten, die in der ersten Spalte aufgelistet sind.

- Die Definition der Negation (in Definition 1.2.7) sagt uns nun, welchen Wahrheitswert $\neg A$ in jedem der Fälle besitzt. Somit sind die Wahrheitswerte in der zweiten Spalte festgelegt.

[9] An der Notation dieser Aussagen sehen Sie, dass wir Klammern verwenden, um klarzumachen, in welcher Reihenfolge der Verknüpfungen ausgewertet werden.

[10] Dieses Resultat bezeichnet man manchmal auch als Satz vom ausgeschlossenen Dritten.

[11] Dieses Resultat bezeichnet man manchmal auch als Satz vom Widerspruch.

- Aus den Wahrheitswerten in den ersten beiden Spalten und der Definition der inklusiven Disjunktion (in Definition 1.2.7) erhalten wir nun die Wahrheitswerte in der dritten Spalte.

- Aus den Wahrheitswerten in den ersten beiden Spalten und der Definition der Konjunktion (in Definition 1.2.5) erhalten wir nun die Wahrheitswerte in der dritten Spalte.

- Aus den Wahrheitswerten in den ersten beiden Spalten und der Definition der exklusiven Disjunktion (in Definition 1.2.7) erhalten wir die Wahrheitswerte in der vierten Spalte.

In der dritten Spalte der Tabelle können Sie sehen, dass die Aussage $A \vee (\neg A)$ tatsächlich in jedem Fall den Wahrheitswert „wahr“ besitzt. Damit ist (a) bewiesen.

In der vierten Spalte können Sie sehen, dass die Aussage $A \wedge (\neg A)$ immer den Wahrheitswert „falsch“ besitzt. Damit ist (b) bewiesen.

Und laut der fünften Spalte besitzt die Aussage $A \dot{\vee} (\neg A)$ immer den Wahrheitswert „wahr“. Somit ist (c) bewiesen. □

Bemerkung 1.2.9. Am Ende des vorangehenden Beweises konnten Sie eine in der Mathematik sehr weit verbreitete Notation sehen: Das Ende eines Beweises wird meist mit einem kleinen Quadrat[12] markiert. Dies wird als symbolische Abkürzung für die Wendung „was zu zeigen war“ – auf lateinisch „quot erat demonstrandum“ – verstanden. Die Abkürzung „qed“ hingegen ist in der Mathematik heute sehr aus der Mode gekommen.

Mit Hilfe von Wahrheitstabellen kann man auch komplexere Resultate über Verknüpfungen von mehreren Aussagen beweisen. Die folgenden vier Propositionen enthalten einige solcher Aussagen.

Proposition 1.2.10 (De Morgansche Regeln für die Verneinung von Konjunktionen und Disjunktionen)**.** *Seien A, B beliebige Aussagen.*

(a) *Die Aussage $\neg(A \wedge B)$ hat immer denselben Wahrheitswert wie $(\neg A) \vee (\neg B)$.*

(b) *Die Aussage $\neg(A \vee B)$ hat immer denselben Wahrheitswert wie $(\neg A) \wedge (\neg B)$.*

Es ist wichtig, dass Sie sich in Ruhe überlegen, weshalb die Resultat in Proposition 1.2.10 stimmen. Unabhängig davon muss man die Proposition aber natürlich auch beweisen, wobei man nur die Definitionen der logischen Verknüpfungen verwendet. Weil dies mit Hilfe von Wahrheitstabellen sehr einfach möglich ist, besprechen wir hier nur exemplarisch den Beweis von Teil (a). Um mit dem Konzept der Wahrheitstabelle vertraut zu werden, sollten Sie selbst versuchen, Teil (b) zu beweisen.

[12] Manche Autorinnen und Autoren verwenden stattdessen z.B. auch eine Raute.

Beweis von Proposition 1.2.10(a). Für die möglichen Wahrheitswerte der Aussagen A, B gibt es insgesamt vier Möglichkeiten. Mit Hilfe der Definitionen 1.2.5 und 1.2.7 können wir somit die folgende Wahrheitstabelle ausfüllen:

A	B	$A \wedge B$	$\neg(A \wedge B)$	$\neg A$	$\neg B$	$(\neg A) \vee (\neg B)$
w	w	w	f	f	f	f
w	f	f	w	f	w	w
f	w	f	w	w	f	w
f	f	f	w	w	w	w

Die Wahrheitswerte für $A \wedge B$ sowie für die Aussage $\neg A$ und die Aussage $\neg B$ haben wir hier nur als Zwischenergebnisse eingefügt, um daraus die Wahrheitswerte der Aussage $\neg(A \wedge B)$ und der Aussage $(\neg A) \vee (\neg B)$ zu bestimmen, die uns eigentlich interessieren.

Wie Sie in der Tabelle erkennen können, hat die Aussage $\neg(A \wedge B)$ immer denselben Wahrheitswert wie die Aussage $(\neg A) \vee (\neg B)$. Somit ist die Behauptung von Proposition 1.2.10(a) bewiesen. □

Proposition 1.2.11 (Kommutativ-Gesetz für Konjunktion und Disjunktionen)**.** *Seien A, B beliebige Aussagen.*

(a) *Die Aussage $A \wedge B$ hat immer denselben Wahrheitswert wie $B \wedge A$.*

(b) *Die Aussage $A \vee B$ hat immer denselben Wahrheitswert wie $B \vee A$.*

(c) *Die Aussage $A \dot{\vee} B$ hat immer denselben Wahrheitswert wie $B \dot{\vee} A$.*

Genau wie Proposition 1.2.10 kann man auch Proposition 1.2.11 mit Hilfe von Wahrheitstabellen beweisen. Wir verzichten darauf, diese Tabellen hier alle anzugeben.

Proposition 1.2.12 (Assoziativ-Gesetz für Konjunktion und Disjunktionen)**.** *Seien A, B, C beliebige Aussagen.*

(a) *Die Aussage $(A \wedge B) \wedge C$ hat immer denselben Wahrheitswert wie $A \wedge (B \wedge C)$.*

(b) *Die Aussage $(A \vee B) \vee C$ hat immer denselben Wahrheitswert wie $A \vee (B \vee C)$.*

(c) *Die Aussage $(A \dot{\vee} B) \dot{\vee} C$ hat immer denselben Wahrheitswert wie $A \dot{\vee} (B \dot{\vee} C)$.*

Auch hier lässt sich der Beweis einfach mit Hilfe von Wahrheitstabellen führen. Weil in Proposition 1.2.12 drei Aussagen A, B, C vorkommen (statt wie bisher nur zwei) geben wir den Beweis für Teil (a) noch einmal explizit an. Die Beweise der anderen beiden Aussagen können Sie, wenn Sie möchten, selbst führen.

Beweis von Proposition 1.2.12(a). Es gibt insgesamt acht Möglichkeiten, wie die Wahrheitswerte von A, B, C lauten können. Wir gehen alle acht Möglichkeiten in der folgenden Tabelle durch:

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

Also besitzen $(A \wedge B) \wedge C$ und $A \wedge (B \wedge C)$ tatsächlich immer dieselben Wahrheitswerte. □

Weil $(A \wedge B) \wedge C$ und $A \wedge (B \wedge C)$ immer dieselben Wahrheitswerte besitzen[13] ist es also nicht wichtig, wo wir die Klammern setzen. Deshalb ist es Konvention, die Klammern auch einfach ganz wegzulassen, und stattdessen die Aussage mit den Wahrheitswerten von $(A \wedge B) \wedge C$ (bzw. $A \wedge (B \wedge C)$) als $A \wedge B \wedge C$ zu notieren.[14]

Proposition 1.2.13 (Distributivgesetze für Konjunktion und Disjunktion)**.** *Seien A, B, C beliebige Aussagen.*

(a) *Es hat $(A \vee B) \wedge C$ immer denselben Wahrheitswert wie $(A \wedge C) \vee (B \wedge C)$.*

(b) *Es hat $(A \wedge B) \vee C$ immer denselben Wahrheitswert wie $(A \vee C) \wedge (B \vee C)$.*

Auch hier erfolgt der Beweis, wie bei Proposition 1.2.12, über Wahrheitstabellen. Da ist nicht allzu sehr aufschlussreich ist, fertig ausgefüllte Tabellen zu lesen, verzichten wir darauf, die Tabellen hier alle aufzuführen. Wie oben gilt stattdessen: Wenn Sie wissen möchten, weshalb die Proposition stimmt, können Sie die Wahrheitstabellen selbst erstellen und somit die Proposition beweisen.[15]

Zum Abschluss dieses Abschnitts noch zwei Bemerkungen über die bisher geführten Beweise:

Bemerkungen 1.2.14. (a) Die bisher geführten Beweise fanden Sie vielleicht, obgleich überzeugend, nicht besonders aufschlussreich oder interessant – Tabellen auszufüllen ist eine nicht besonders kreative Tätigkeit. Sie können aber unbesorgt sein: Die große Mehrzahl der Beweise, die Sie ab sofort sehen werden, hat mit dem Ausfüllen von Tabellen nichts (oder sehr wenig) zu tun.

Genau gesprochen handelt es sich bei der oben verwendeten Methoden um eine bestimmte **Beweistechnik**, nämlich um **Fallunterscheidungen**: Wenn man

[13] Später werden wir für diese Eigenschaft übrigens noch einen speziellen Begriff einführen: Wir werden zwei Aussagen **äquivalent** nennen, wenn Sie dieselben Wahrheitswerte besitzen.

[14] Ebenso kann man dann auch für noch mehr als nur drei Aussagen vorgehen. Darauf gehen wir aber an dieser Stelle erst mal nicht weiter ein, weil es nicht wirklich zu neuen Einsichten führt.

[15] Versuchen Sie das ruhig einmal, zumindest für eine oder zwei der Propositionen! Sie sollen ja kritisch sein und nicht einfach alles glauben, was Ihnen im Manuskript oder in der Vorlesung erzählt wird!

weiß, dass nur eine bestimmte Anzahl an Fällen auftreten kann (z.B. können im Beweis von Proposition 1.2.12(a) nur acht Fälle für die Wahrheitswerte von A, B, C auftreten), dann kann man all diese Fälle einzeln durchgehen, und sehen, was in jedem Fall jeweils passiert. In den obigen Beweisen war es besonders leicht, in jedem Fall zu sehen, was passiert: Wenn man nämlich die Wahrheitswerte aller gegebenen Aussagen kennt, dann kann man daraus recht mechanisch auch die Wahrheitswerte von zusammengesetzten Aussagen bestimmen.[16]

Sie werden schon bald noch weitere nützliche Beweistechniken kennenlernen, und Fallunterscheidungen werden nur noch ab und zu vorkommen.

(b) Grundsätzlich werden alle Beweise, egal wie sie im Detail ausgestaltet sind, darauf beruhen, korrekte logische Schlussfolgerungen zu ziehen: Ein „korrekte logische Schlussfolgerung zu ziehen“ bedeutet hierbei, dass man eine Aussage gegeben hat, und im nächsten Schritt eine Aussage angibt, die immer richtig ist, falls die vorangehende Aussage richtig ist.

Um dies effizient und fehlerfrei zu tun, muss man sicher mit Aussagen und deren Verknüpfungen umgehen. Man kann den Inhalt dieses Abschnitts also, wenn man möchte, als ein wenig „meta-mathematisch“ betrachten: Wir möchten im Rest der Vorlesung (während des ganzen Semesters) gerne verschiedenste mathematische Resultate beweisen. Dazu müssen wir sicher mit Aussagen umgehen können, und deshalb haben wir hier zuerst einige Dinge über Aussagen selbst bewiesen.

In Abschnitt 1.4 bauen wir die Aussagenlogik noch weiter aus, indem wir dort über **quantifizierte Aussagen** sprechen.

1.3 Mengen und Tupel

Wie Sie schon wissen, möchten wir in der Mathematik Aussagen über mathematische Objekte beweisen. Die einfachsten mathematischen Objekte, die Ihnen vielleicht einfallen, sind vermutlich Zahlen (z.B., ganze Zahlen, rationale Zahlen, reelle Zahlen).

Es gibt aber noch zahlreiche weitere mathematische Objekte, und es ist ein wichtiges Grundprinzip in der Mathematik, dass man aus bekannten Objekten neue Objekte baut, in dem man zum Beispiel mehrere Objekte zu einem Objekt zusammenfasst. In diesem Abschnitt besprechen wir zwei Möglichkeiten, um dies zu tun: **Mengen** und **Tupel**.

[16] Dass wir die Fallunterscheidungen in Tabellenform aufgeschrieben haben, liegt einfach daran, dass dies für die Bestimmung der Wahrheitswerte von Aussagen am übersichtlichsten ist. Auch im weiteren Verlauf der Vorlesung und Ihres Studiums werden Ihnen immer wieder einmal Beweise per Fallunterscheidung begegnen – allerdings sind die einzelnen Fälle dann meist komplizierter als nur eine einzelne Zeile von Wahrheitswerten, und deshalb werden solche Beweise dann meist nicht mehr in Tabellenform aufgeschrieben.

Was ist eine Menge?

Wenn Sie im Supermarkt eine Gurke, eine Flasche Bier, eine Packung Chips und eine Melone kaufen, haben Sie vermutlich Schwierigkeiten, diese einzeln nach Hause zu tragen. Eine bewährte Möglichkeit, dies zu lösen, besteht darin, all Ihre Einkäufe in eine Tüte (oder einen Korb oder einen Rucksack) zu packen und somit nur noch ein Objekt transportieren zu müssen statt vier einzelne Objekte. Kurzum: Sie fassen mehrere Objekte zu einem Objekt zusammen, indem Sie eine Art von Hülle um die Objekte legen.

Dasselbe tun wir nun in der Mathematik: Wir fassen mehrere Objekte zu einem neuen Objekt zusammen. Das zusammengefasste Objekt bezeichnen wir dann als **Menge**, in die einzelnen Objekte, die wir „hineingelegt“ haben, bezeichnen wir als **Elemente** der Menge.

Wie in Ihrer Einkaufstüte ist es dabei nicht von Belang, in welcher Reihenfolge Sie die Objekte in die Tüte legen.[17] Es gibt allerdings einen entscheidenden Unterschied zwischen der Einkaufstüte und einer Menge: In einer Menge verlangt man, dass jedes Element der Menge dort höchstens einmal vorkommt,[18] während Sie in eine Einkaufstüte durchaus zwei Flaschen Bier derselben Sorte legen können.

Was wir oben beschrieben haben, wird in der folgenden klassischen Definition des Begriffs **Menge**, die von Georg Cantor[19] stammt, kanpp zusammengefasst:

Definition 1.3.1 (Mengendefinition nach Cantor)**.** Eine **Menge** M ist eine Zusammenfassung bestimmter, wohlunterschiedener Objekte unserer Anschauung oder unseres Denkens (welche die **Elemente** von M genannt werden) zu einem Ganzen.

Die Reihenfolge, in der die Objekte zusammengefasst werden, spielt hierbei keine Rolle.[20]

Das Wort „wohlunterschiedene“ in der Definition stellt klar, dass jedes Element nur einmal in der Menge M auftreten darf.

Ein Problem bei dieser Definition ist, dass nicht geklärt wird, war genau man unter einer „Zusammenfassung“ versteht, und ob es wirklich immer widerspruchsfrei möglich ist, beliebige Objekte zu einem neuen Objekt zusammenzufassen. Deshalb wurde später eine formale Mengenlehre entwickelt, die sich dieser Probleme annimmt und die viel präziser ist. Diese formale Mengenlehre ist aber viel zu abstrakt um sie im ersten Semester zu besprechen, und in vielen mathematischen Teilgebieten reicht obige Definition im Arbeitsalltag meist aus.

[17] Nun gut, vielleicht möchten Sie die Chips nicht ganz nach unten legen – aber lassen Sie uns dies hier einfach ignorieren.

[18] Manchmal kann es auch interessant sein, ein Objekt mehrmals zuzulassen – dafür gibt es das Konzept der sogenannten **Multimenge**. Allerdings kommen Multimengen im mathematischen Alltag viel seltener vor als Mengen, und in der Linearen Algebra 1 werden wir Multimengen gar nicht benötigen.

[19] Georg Cantor (1845 in St. Petersburg – 1918 in Halle) war ein deutscher Mathematiker, der als Begründer der Mengenlehre gilt.

[20] Genau genommen stammt der erste Absatz in der Definition von Cantor; den zweiten Absatz haben wir hier hinzugefügt, um Missverständnisse auszuschließen.

Um eventuelle Widersprüche auszuschließen, wollen wir Definition 1.3.1 noch um folgende Vereinbarung ergänzen:

Vereinbarung 1.3.2. Wir vereinbaren, dass die „Zusammenfassung bestimmter, wohlunterschiedener Objekte“ zu einer Menge nur dann möglich ist, wenn sichergestellt ist, dass von jedem mathematischen Objekt entschieden werden kann, ob es ein Element der Menge ist, oder nicht.

Um zusagen, dass ein Objekt x Element einer Menge M ist, benutzen wir die Notation

$$x \in M;$$

man spricht dies beim Vorlesen folgendermaßen aus: „x Element M“. Um hingegen zu sagen, dass ein Objekt x nicht Element einer Menge ist, benutzen wir die Notation

$$x \notin M;$$

dies spricht man beim Vorlesen folgendermaßen aus: „x nicht Element M“.

Wenn Sie sich nun Definition 1.2.1 noch einmal durchlesen, dann sehen Sie: Wenn M eine Menge ist und x irgendein Objekt, dann ist „$x \in M$“ eine Aussage. Ebenso ist „$x \notin M$“ eine Aussage – nämlich die Verneinung der vorangehenden Aussage; anders gesagt hat „$x \notin M$“ immer denselben Wahrheitswert wie „$\neg(x \in M)$“.

Einige Mengen, deren Elemente bestimmte Zahlen sind, kommen in der Mathematik so häufig vor, dass man eigene Symbole für sie einführt und durchgängig benutzt. Sie kennen viele dieser Mengen und deren Symbole vermutlich bereits aus der Schule – aber der Vollständigkeit halber wiederholen wir sie hier noch einmal kurz:

Notation 1.3.3.

(a) Mit $\mathbb{Z}$ bezeichnen wir die Menge der **ganzen Zahlen**.[21]

(b) Mit $\mathbb{N}$ bezeichnen wir die Menge aller ganzen Zahlen, die größer oder gleich 0 sind.

Wir nennen $\mathbb{N}$ die Menge der **natürlichen Zahlen**.

(c) Mit $\mathbb{N}^*$ bezeichnen wir die Menge der ganzen Zahlen, die größer oder gleich 1 sind.

[21] An dieser Stelle sehen Sie, dass wir mathematisch nicht komplett von vorne beginnen: An einigen Stellen setzen wir Wissen aus der Schule voraus, zum Beispiel ein intuitives Verständnis, was die ganzen Zahlen, die rationalen Zahlen und die reellen Zahlen sind, und was es bedeutet, dass eine Zahl größer oder größer gleich ist als eine andere Zahl.

Für all diese Zahlenmengen und deren Eigenschaft kann man präzise angeben, wie Sie definiert sind – dies erfordert jedoch einiges an Zusatzaufwand und gehört thematisch nicht zur Linearen Algebra 1. Möglicherweise werden Sie in der Vorlesung Analysis 1 ansprechen, wie man z.B. die reellen Zahlen konstruieren kann, und welche axiomatischen Eigenschaften sie besitzen.

(d) Mit $\mathbb{Q}$ bezeichnen wir die Menge aller **rationalen Zahlen**.[22]

(e) Mit $\mathbb{R}$ bezeichnen wir die Menge der **reellen Zahlen**.

Bemerkung 1.3.4. Bzgl. des Begriffs **natürliche Zahlen** herrscht Uneinheitlichkeit in der Literatur: Manche Autorinnen und Autoren verwenden dieselbe Nomenklatur wie in diesem Manuskript, d.h., sie setzen $\mathbb{N} \coloneqq \{0, 1, 2, \dots\}$ und $\mathbb{N}^* \coloneqq \{1, 2, \dots\}$, und nennen die erstgenannte Menge die **natürlichen Zahlen**. Andere hingegen setzen $\mathbb{N}_0 \coloneqq \{0, 1, 2, \dots\}$ and $\mathbb{N} \coloneqq \{1, 2, \dots\}$ und nennen die letztgenannte Menge die **natürlichen Zahlen**. Wenn Sie etwas in einem Buch oder Manuskript nachlesen, ist es also wichtig darauf zu achten, wie die natürlichen Zahlen dort definiert sind, und welches Symbol hierfür benutzt wird.

Methoden zur Beschreibung von Mengen

Definition 1.3.1 und Vereinbarung 1.3.2 sagen uns zwar, was wir unter einer Menge verstehen – aber um mit Mengen zu arbeiten, braucht man auch Möglichkeiten um Mengen effizient aufzuschreiben.

Dazu gibt es in der Mathematik mehrere Möglichkeiten, die im folgenden aufgezählt werden:

Vereinbarung 1.3.5 (Beschreibung von Mengen)**.** (a) Eine Menge, die nur endlich viele Elemente besitzt, kann man beschreiben, indem man alle Ihre Elemente aufzählt.[23]

(b) Man kann eine Menge (egal ob sie endlich oder unendlich viele Elemente besitzt) beschreiben, in dem man Bedingungen angibt, die ein Objekt erfüllen muss, um Element der Menge zu sein.

(c) Man kann eine Menge (egal ob sie endlich oder unendlich viele Elemente besitzt) beschreiben, in dem man eine Regel angibt, nach der alle ihre Elemente konstruiert werden.[24]

Einige Beispiele für die Möglichkeit (b) haben Sie bereits in Notation 1.3.3 gesehen, denn zum Beispiel kann man die dort verwendete Formulierung „Mit $\mathbb{Z}$ bezeichnen wir die Menge der ganzen Zahlen“ auch ewas ausführlicher schreiben als „Mit $\mathbb{Z}$ bezeichnen wir die Menge aller Objekte, die eine ganze Zahl sind.“

Jedes der drei Möglichkeiten in Vereinbarung 1.3.5 kann man auf verschiedene Weise umsetzen, zum Beispiel indem man eine geeignete natürlichsprachliche Formulierung verwendet oder indem man die in der Mathematik sehr übliche Notation

[22] Zur Erinnerung: Eine Zahl heißt rational, wenn man sie als Bruch schreiben kann, wobei der Zähler eine ganze Zahl und der Nenner eine von Null verschiedene ganze Zahl ist.

[23] Für Mengen mit sehr vielen Elementen ist das aber manchmal äußerst unpraktisch, und für Mengen mit unendlichen vielen Elementen geht es gar nicht.

[24] Im nächsten Abschnitt, 1.4, werden Sie sehen, dass Möglichkeit (c) im Grunde ein Spezialfall von Möglichkeit (b) ist.

mit geschweiften Klammern verwendet. Das lässt sich am einfachsten anhand eines Beispiels demonstrieren:

Beispiel 1.3.6. Die folgenden Formulierungen beschreiben jeweils dieselbe Menge M:

(a) *Aufzählen der Elemente*:

- Natürlichsprachliche Formulierung:

  „Es sei M die Menge, deren Elemente die Zahlen 2, 4 und 6 sind."

- Notation mit geschweiften Klammern:

  „Es sei $M = \{2, 4, 6\}$."

(b) *Angeben von Bedingungen, die Objekte erfüllen müssen, um Elemente von M zu sein:*

- Natürlichsprachliche Formulierung:

  „Es sei M die Menge aller ganzen Zahlen, die gerade sind, und die größer oder gleich 2 und kleiner oder gleich 6 sind."

- Natürlichsprachliche Formulierung, etwas knapper: Dieses Mal drücken wir uns etwas kürzer aus, indem wir eine Variable verwenden:[25]

  „Es sei M die Menge aller Objekte z, die die folgenden Bedingungen erfüllen: $z \in \mathbb{Z}$ und z ist gerade und $2 \leq z \leq 6$."

- Notation mit geschweiften Klammern:

  „Es sei $M = \{z \mid \ z \in \mathbb{Z}$ und z ist gerade und $2 \leq z \leq 6\}$."

  Den Strich „$\mid$ " liest man hier als „mit den Eigenschaften" oder „welche die folgenden Eigenschaften erfüllen". Die gesamte Beschreibung von M kann man z.B. folgendermaßen lesen:

  „Es sei M die Menge aller Objekte z, welche die folgenden Eigenschaften erfüllen: z ist in $\mathbb{Z}$, und z ist gerade, und $2 \leq z \leq 6$."

- Notation mit geschweiften Klammern, noch knapper: Wir drücken uns noch etwas kürzer aus, indem wir die Symbole zur Verknüpfungen von Aussagen verwenden, die wir im vorangehenden Abschnitt 1.2 besprochen haben:

  „Es sei $M = \{z \mid \ z \in \mathbb{Z} \ \wedge \ z$ ist gerade $\ \wedge \ 2 \leq z \leq 6\}$."

[25] Mehr Informationen zu Variablen folgen zu Beginn des Abschnitts 1.4.

- Notation mit geschweiften Klammern, nochmals effizienter: Wir verkürzen die Schreibweise ein wenig mehr, in dem wir die Bedingung „$z \in \mathbb{Z}$“ vor den Strich | schreiben:

  $$\text{„Es sei } M = \{z \in \mathbb{Z} \mid \quad z \text{ ist gerade } \quad \wedge \quad 2 \leq z \leq 6\}.\text{“}$$

  Anschaulich kann man diese Notation so lesen, dass die Elemente von M diejenigen Elemente aus einer anderen Menge – in diesem Fall $\mathbb{Z}$ – sind, welche die nach dem Strich | folgenden Bedingungen erfüllen.

  Man kann diese Beschreibung von M zum Beispiel folgendermaßen aussprechen: „Es sei M die Menge aller Elemente z von $\mathbb{Z}$, welche die folgenden Eigenschaften erfüllen: z ist gerade, und $2 \leq z \leq 6$.“

(c) *Angeben einer Regel um die Elemente von M zu konstruieren:*

- Natürlichsprachliche Formulierung:

  „Es sei M die Menge der Zahlen, die man erhält, indem man alle ganzen Zahlen, die größer oder gleich 1 und kleiner oder gleich 3 sind, mit 2 multipliziert.“

- Notation mit geschweiften Klammern:

  $$\text{„Es sei } M = \{2y \mid \quad y \in \mathbb{Z} \quad \wedge \quad 1 \leq y \leq 3\}\text{“}.$$

  Diese Notation kann man zum Beispiel folgendermaßen aussprechen: „Es sei M die Menge aller Objekte von der Form $2y$, wobei folgendes gilt: $y \in \mathbb{Z}$ and $1 \leq y \leq 3$.“

Das korrekte Beschreiben von Mengen ist für alles, was in dieser Vorlesung folgt – und auch für jede weitere Mathematikvorlesung – absolut essentiell. Beim Üben sollten Sie die folgende Bemerkung beachten.

Bemerkung 1.3.7. In Beispiel 1.3.6 ist für die verschiedenen Notationsmöglichkeiten jeweils angegeben, wie Sie diese Aussprechen können (und diese lässt sich natürlich auch auf verschiedene Weise variieren).

Üben Sie auf jeden Fall, Mengen nicht nur zu lesen und aufzuschreiben, sondern auf jeden Fall auch vorzulesen. Dies wird Ihre Intuition für den Umgang mit Mengen stärken, und es wird Ihnen regelmäßig ins Gedächtnis rufen, dass auf dem Blatt Papier oder Computerbildschirm vor Ihnen nicht irgendein Symbol-Salat steht, sondern Dinge mit einer ganz konkreten Bedeutung, die man auch natürlichsprachlich formulieren kann.[26]

Wichtig: Man kann – und soll! – die mathematische Beschreibung einer Menge immer so vorlesen, dass das, was man sagt, ein grammatisch sinnvoller Satz in der

[26] Wenn Sie mit Kommilitoninnen und Komilitonen über die Vorlesunginhalte und Übungsaufgaben diskutieren (was Sie auf jeden Fall tun sollten!), sollten Sie aus demselben Grund ebenfalls eine Kombination aus Aufschreiben und mündlicher Diskussion wählen.

verwendeten Sprache (in Ihrem Fall vermutlich: Deutsch) ist. Mathematische Notation so vorzulesen, dass Sie grammatisch sinnvolle Sätze sagen, ist wichtig, damit Ihr Gegenüber Sie verstehen kann, und damit Sie sicher sind, dass Sie das Vorgelesene selbst verstehen. Üben Sie deshalb das Vorlesen von Mengen unbedingt solange, bis Sie dies stets in vollständigen und grammatisch richtigen Sätzen tun!

Als ein weiteres wichtiges Beispiel erwähnen wir noch die leere Menge:

Beispiel 1.3.8 (Leere Menge)**.** Die Menge, die gar kein Element enthält, bezeichnet man als **leere Menge**. Man notiert sie häufig mit dem Symbol $\emptyset$.

Auch diese Menge kann man mithilfe von geschweiften Klammern aufschreiben, indem man alle Elemente dieser Menge zwischen den geschweiften Klammern aufzählt: Mit dieser Schreibeweise erhält man die Notation $\{\}$ für die leere Menge.[27]

Beziehungen zwischen Mengen

Weil wir sehr viel mit Mengen arbeiten werden und dabei keine Missverständnisse zur Bedeutung bestimmter Begriffe aufkommen lassen wollen, wollen wir im folgenden klären, wann genau wir zwei Menge als gleich auffassen:

Definition 1.3.9 (Gleichheit von Mengen)**.** Zwei Mengen M_1, M_2 heißen **gleich**, falls sie genau die gleichen Elemente haben.

Wir verwenden die Notation $M_1 = M_2$ um auszudrücken, dass die Mengen M_1 und M_2 gleich sind und die Notation $M_1 \neq M_2$ um auszudrücken, dass M_1 und M_2 nicht gleich sind.

Für zwei Mengen M_1, M_2 ist „$M_1 = M_2$“ also ein Aussage; ihre Verneinung „$\neg(M_1 = M_2)$“ ist genau die Aussage $M_1 \neq M_2$.

Beispiele 1.3.10. (a) Es gilt $\{7, -1, \pi\} = \{-1, \pi, 7\}$ (denn die Reihenfolge der Elemente spielt für eine Menge keine Rolle.[28]

(b) Es gilt $\{n \in \mathbb{N} \mid\ 5 \leq n \leq 8\} = \{5, 6, 7, 8\}$.

(c) Es gilt $\{k \in \mathbb{Z} \mid\ k^2 = 9\} \neq \{3\}$.[29]

(d) Es gilt $\{\} \neq \big\{\{\}\big\}$.[30]

Neben der Gleichheit der Mengen sind auch noch Teilmengenbeziehungen[31] zwischen Mengen von großer Bedeutung. Diese sind folgendermaßen definiert:

[27] Die beiden Schreibweise $\emptyset$ und $\{\}$ für die leere Menge sind beide gebräuchlich, allerdings ist $\emptyset$ erfahrungsgemäß verbreiteter.

[28] Siehe Definition 1.3.1.

[29] Denn die Menge auf der linken Seite besteht aus den Elementen 3 und -3.

[30] Denn die Menge auf der linken Seite hat kein Element, während die Menge auf der rechten Seite das Element $\{\}$ besitzt.

Wichtig um das zu verstehen, ist die Beobachtung, dass die leere Menge $\{\}$ nicht „nichts“ ist, sondern ein mathematisches Objekt: nämlich eine Menge, die nichts enthält.

[31] Die man oft auch **Inklusionen** nennt.

Definition 1.3.11 (Teilmengen und Obermengen)**.** Eine Menge M_1 heißt **Teilmenge** von M_2, falls jedes Element von M_1 auch ein Element von M_2 ist. In diesem Fall nennt man M_2 auch **Obermenge** von M_1.

Um auszudrücken, dass M_1 Teilmenge von M_2 ist, verwenden wir die Notation $M_1 \subseteq M_2$ oder, alternativ, die Notation $M_2 \supseteq M_1$.

Ähnlich wie zuvor schreiben wir $M_1 \not\subseteq M_2$ (oder $M_2 \not\supseteq M_1$) als Abkürzung für die Aussage $\neg(M_1 \subseteq M_2)$.

Bemerkung 1.3.12. Mit diesen Begriffsbildungen folgt für alle Mengen M_1, M_2 sofort die folgende Beobachtung: Die Aussage $M_1 = M_2$ ist gleichbedeutend mit der Aussage $M_1 \subseteq M_2 \ \wedge \ M_1 \supseteq M_2$.

Beispiele 1.3.13. (a) Es gilt $\{1, 3\} \subseteq \{3, 2, 1\}$, aber $\{3, 2, 1\} \not\subseteq \{1, 3\}$.

(b) Es gilt $\{\frac{3}{2}, 5\} \not\subseteq \{-3, 10\}$ und $\{-3, 10\} \not\subseteq \{\frac{3}{2}, 5\}$.

(c) Für jede Menge M gilt $M \supseteq \emptyset$.

(d) Für jede nicht-leere Menge M gilt $M \not\subseteq \emptyset$.

(e) Es gilt $\emptyset \subseteq \emptyset$.[32]

(f) Es gilt $\{2\} \subseteq \big\{2, \{2, 3\}\big\}$, aber $\{2, 3\} \not\subseteq \big\{2, \{2, 3\}\big\}$.[33]

Mengenoperationen

Aus gegebenen Mengen kann man neue Mengen bauen – unter anderem folgendermaßen:

Definition 1.3.14 (Durchschnitt, Vereinigung und Differenz zweier Mengen)**.** Seien L, M Mengen.

(a) Die Menge $L \cap M \coloneqq \{x \mid \ x \in L \ \wedge \ x \in M\}$ heißt der **Durchschnitt** von L und M.[34]

(b) Die Menge $L \cup M \coloneqq \{x \mid \ x \in L \ \vee \ x \in M\}$ heißt die **Vereinigung** von L und M.

(c) Die Menge $L \setminus M \coloneqq \{x \mid \ x \in L \ \wedge \ x \notin M\}$ heißt die **mengentheoretische Differenz** – oder kürzer die **Differenz** – von L und M.

[32] Weil ja sogar $\emptyset = \emptyset$ gilt.

[33] Denn die Menge $\big\{2, \{2, 3\}\big\}$ hat nur die beiden Elemente 2 und $\{2, 3\}$.

[34] Das Gleichheitszeichen mit Doppelpunkt auf einer Seite benutzt man in der Mathematik häufig; es bedeutet: Das Objekt auf der Seite des Gleichheitszeichen, wo der Doppelpunkt steht, wird **definiert** als das Objekt auf der anderen Seite des Gleichheitszeichens. Die Verwendung dieses Symbols ergibt natürlich nur Sinn, wenn das Objekt auf der Seite, wo der Doppelpunkt steht, bisher noch nicht definiert ist.

Die obenstehenden Mengenoperationen wurden alle mit Hilfe der logischen Verknüpfungen definiert, die Sie bereits aus Abschnitt 1.2 kennen. In diesem Abschnitt hatten wir mithilfe von Wahrheitstabellen verschiedene Eigenschaften für die Verknüpfungen von Aussagen bewiesen. Diese Eigenschaften können wir nun verwenden, um interessante Eigenschaften von Mengenoperationen zu beweisen.

Als ein erstes einfaches Beispiel zeigen wir, dass die Durchschnittsbildung zweier Mengen kommutativ ist:

Proposition 1.3.15. *Seien L, M Mengen. Dann gilt $L \cap M = M \cap L$.*

Beweis. Um die Proposition zu beweisen, verwenden wir die Beobachtung aus Bemerkung 1.3.12. Es genügt demnach, wenn wir die beiden Inklusionen „$L \cap M \subseteq M \cap L$“ und „$L \cap M \supseteq M \cap L$“ beweisen.

„$\subseteq$“ Um die Inklusion „$L \cap M \subseteq M \cap L$“ zu zeigen, müssen wir laut Definition 1.3.11 beweisen, dass jedes Element von $L \cap M$ auch ein Element von $M \cap L$ ist.

Sei also x ein beliebiges Element von $L \cap M$. Dann gilt laut Definition 1.3.14(a)

$$x \in L \ \wedge \ x \in M.$$

Aus Proposition 1.2.11(a) wissen wir, dass dann auch die Aussage

$$x \in M \ \wedge \ x \in L$$

wahr ist. Dies bedeutet laut Definition 1.3.14(a), dass $x \in M \cap L$ ist.

„$\supseteq$“ Der Beweis der umgekehrten Inklusion „$L \cap M \subseteq M \cap L$“ ist sehr ähnlich:[35] Laut Definition 1.3.11 müssen wir, um diese Inklusion zu beweisen, zeigen, dass jedes Element von $M \cap L$ auch ein Element von $L \cap M$ ist.

Sei also x ein beliebiges Element von $M \cap L$. Dann gilt laut Definition 1.3.14(a) die Aussage.

$$x \in M \ \wedge \ x \in L.$$

Somit ist laut Proposition 1.2.11(a) auch die Aussage

$$x \in L \ \wedge \ x \in M$$

wahr. Dies bedeutet gemäß Definition 1.3.14(a), dass $x \in L \cap M$ ist. □

Mithilfe der Resultate aus Abschnitt 1.2 kann man noch viele weitere Regeln für Durchschnitt, Vereinigung und Differenz von Mengen beweisen. Darauf kommen wir in den Übungen zurück.

[35] Wir werden aber schon bald Beweise von Mengengleichheiten führen, in denen die beiden Inklusionen auf sehr verschiedene Weise bewiesen werden.

Tupel und kartesische Produkte

Sie haben bereits gelernt, dass Mengen in der Mathematik genutzt werden, um mehrere Objekte zu einem Objekt zusammenzufassen – wobei jedes der Objekte, die zusammengefasst werden, nur einmal in der Menge auftauchen darf, und die Reihenfolge der Objekte beim Zusammenfassen keine Rolle spielt.

In diesem Abschnitt besprechen wir, die entgegengesetzte Situation: Wir fassen mehrere Objekte zu einem neuen Objekt zusammen, wobei wir aber die Reihenfolge beachten wollen und wobei ein Objekt auch mehrmals vorkommen darf. Außerdem wollen wir uns in diesem Abschnitt darauf beschränken, nur endlich viele Objekte zusammenzufassen.[36]

Definition 1.3.16 (Tupel)**.** (a) Die Zusammenfassung endlich vieler Objekte (wobei auch einige oder alle dieser Objekte gleich sein dürfen) in einer bestimmten Reihenfolge zu einem einzelnen Objekt bezeichnet man als **Tupel**. Die so zusammengefassten Objekte heißen **Einträge** des Tupels.

(b) Zur Notation eines Tupels verwenden wir runde Klammern, innerhalb derer die Einträge in entsprechender Reihenfolge aufgezählt werden. Ist x ein Tupel, so bezeichnen wir mit x_1 den ersten Eintrag des Tupels, mit x_2 den zweiten Eintrag, und so weiter (wobei der Index nicht größer sein darf als die Anzahl der Einträge des Tupels).

(c) Zwei Tupel x und y heißen **gleich**, falls die Anzahl Ihrer Einträge die gleiche Zahl $n \in \mathbb{N}$ ist, und falls für jede natürliche Zahl k zwischen 1 und n gilt: $x_k = y_k$.

Wir verwenden die Notation $x = y$ um zu sagen, dass zwei Tupel x und y gleich sind, und wir verwenden erneut die Notation $x \neq y$ als Abkürzung für die Aussage $\neg(x = y)$.

Konkret schreibt man die Objekte, aus denen ein Tupel besteht, innerhalb der runden Klammern oft von links nach rechts auf und trennt sie durch Kommata. Eine häufig gebrauchte Alternative, die manchmal übersichtlicher ist, besteht darin, die Einträge von oben nach unten aufzuzählen. Zum Beispiel könnten wir das Tupel $(5, \pi, 5)$ auch in der Form

$$\begin{pmatrix} 5 \\ \pi \\ 5 \end{pmatrix}$$

[36] Man kann natürlich auch unendlich viele Objekte auf diese Weise zusammenzufassen; es ist aber schwieriger, dies sauber aufzuschreiben – insbesondere, weil man darauf achten muss, was bei unendlich vielen Objekten mit dem Begriff „Reihenfolge“ überhaupt gemeint ist. Wir kommen in Abschnitt 1.5 hierauf zurück.

schreiben, und meinen damit dasselbe.[37]

Es folgen ein paar Beispiele:

Beispiel 1.3.17. Die Tupel

$$(5, \pi, 5), \qquad (5, 5, \pi), \qquad (5, 5, \pi, 5), \qquad ()$$

sind alle verschieden. Das letztgenannte ist Tupel mit Null Einträgen, das sogenannte **leere Tupel**. Es gilt beispielsweise

$$(5, \pi, 5)_1 = 5, \quad (5, \pi, 5)_2 = \pi, \quad (5, \pi, 5)_3 = 5,$$

während $(5, \pi, 5)_4$ nicht definiert ist, weil das Tupel $(5, \pi, 5)$ nur drei Einträge hat.

Tupel ermöglichen die äußerst nützliche Begriffsbildung des kartesischen Produktes von Mengen. Kartesische Produkte von $\mathbb{R}$ mit sich selbst werden später eines der häufigsten Beispiele in der Vorlesung sein.

Definition 1.3.18 (Kartesisches Produkt)**.** (a) Sei $n \in \mathbb{N}$ und seien $M_1, \ldots, M_n$ Mengen.[38] Die Menge

$$M_1 \times \cdots \times M_n := \big\{(x_1, \ldots, x_n) \mid \ x_1 \in M_1 \ \wedge \cdots \wedge \ x_n \in M_n\big\}$$

nennt man das *kartesische Produkt* der Mengen $M_1, \ldots M_n$.

(b) Sei $n \in \mathbb{N}$ und sei M eine Menge. Dann verwendet man die Abkürzung

$$M^n := \underbrace{M \times \cdots \times M}_{n \text{ Faktoren}}$$

für das n-fache kartesische Produkt mit sich selbst.

Die vorangehende Definition wirkt auf den ersten Blick sehr abstrakt; sie wird aber deutlich klarer, wenn man sich einige einfache Beispiele ansieht:

Beispiele 1.3.19. (a) Sei $L = \{5, 6\}$ und $M = \{-2, -1, 0\}$. Dann gilt

$$L \times M = \big\{(5, -2),\ (5, -1),\ (5, 0),\ (6, -2),\ (6, -1),\ (6, 0)\big\}.$$

(b) Es ist $\mathbb{R}^2 = \big\{(x_1, x_2) \mid \ x_1 \in \mathbb{R} \ \wedge \ x_2 \in \mathbb{R}\big\}$.

Die Menge $\mathbb{R}^2$ kann man geometrisch interpretieren; dies besprechen wir auf dem dritten Übungsblatt genauer.

[37] Hier aber schon einmal eine Vorwarnung: Die Sache wird etwas subtiler werden, wenn wir in Abschnitt 4.2 sogenannte *Matrizen* einführen – denn dann werden wir tatsächlich sauber unterscheiden müssen, ob wir die Einträge nebeneinander oder übereinander schreiben. Im Moment braucht uns das aber noch nicht zu kümmern.

[38] Von diesen Mengen dürfen auch mehrere gleich sein. Dies ist eine generelle Regel in der Mathematik: Wenn man mehrere Objekte mit verschiedenen Namen aufzählt, ist trotzdem zugelassen, dass manche dieser Objekte gleich sind – es sei denn, man sagt explizit dazu, dass die Objekte alle verschieden sein sollen.

1.4 Quantoren

Variablen

Variablen (oder **Platzhalter**) werden in der Mathematik verwendet, um nicht nur Aussagen über einzelne mathematische Objekte, sondern über viele (oder unendliche viele) mathematische Objekte zu treffen. Zur Einstimmung besprechen wir kurz einige sehr einfache Beispiele:

Beispiele 1.4.1. (a) Sehen Sie sich die folgenden drei (wahren) Aussagen an:

$$\begin{aligned} &\text{Es gilt } (1+1)^2 \leq 1^2 + 3 \cdot 1. \\ &\text{Es gilt } (2+1)^2 \leq 2^2 + 3 \cdot 2. \\ &\text{Es gilt } (3+1)^2 \leq 3^2 + 3 \cdot 3. \end{aligned} \tag{1.4.1}$$

Diese drei Aussagen können wir komprimierter aufschreiben, indem wir eine Variable – nennen wir sie zum Beispiel r – verwenden:

$$\text{Für jede Zahl } r \in \{1,2,3\} \text{ gilt } (r+1)^2 \leq r^2 + 3 \cdot r. \tag{1.4.2}$$

Beachten Sie unbedingt, dass der Beginn dieses Satzes, also die Formulierung „Für jede Zahl $r \in \{1,2,3\}$“, benötigt wird um zu erkennen, dass (1.4.2) eine Kurzfassung von genau den drei Aussagen in (1.4.1) ist.

(b) Mit Hilfe einer binomischen Formel können Sie sich überlegen, dass die Aussage $(r+1)^2 \leq r^2 + 3 \cdot r$ nicht nur für $r \in \{1,2,3\}$ stimmt, sondern sogar für alle reellen Zahlen r, die größer oder gleich 1 sind.[39] Um dies kurz und knapp zum Ausdruck zu bringen, können Sie schreiben:

$$\text{Für jede reelle Zahl } r \geq 1 \text{ gilt } (r+1)^2 \leq r^2 + 3 \cdot r.$$

Sie können hier einen großen Vorteil der Verwendung von Variablen erkennen: Mit ihrer Hilfe kann man in einem einzigen Satz Aussagen über unendlich viele Objekte formulieren.

(c) Variablen sind nicht nur nützlich um auszudrücken, dass eine Formel für mehrere (oder sogar unendlich viele) Objekte gilt; auch wenn Sie etwas ausdrücken möchten, wozu Sie gar keine Formel benötigen, können Sie – und sollten Sie häufig auch – Variablen verwenden, um sich möglichst verständlich auszudrücken. Betrachten Sie zum Beispiel die folgende Geschichte:

„Nehmen wir an, dass Adrian und Berta jeweils eine Geldsumme an Christina verschenken. Anschließend verschenkt Christina das Doppelte der von Adrian erhaltenen Summe an die Berta und die Hälfte der von Berta erhaltene

[39] Das können Sie so sehen: Wenn r eine reelle Zahl ist, die größer oder gleich 1 ist, dann gilt $(r+1)^2 = r^2 + 2r + 1 \leq r^2 + 2r + r = r^2 + 3r$.

Summe an Adrian. Ob Christina dabei insgesamt einen Gewinn oder Verlust gemacht hat, hängt davon ab, wieviel sie zu Beginn jeweils von Adrian und Berta erhalten hatte."

Sind Sie noch dabei? Lassen Sie uns dasselbe noch einmal formulieren, aber diedes Mal verwenden wir Variablen um die Geldummen zu bezeichnen:

„Nehmen wir an, dass Adrian und Berta Geldsummen s_A bzw. s_B an Christina verschenken. Anschließend schenkt Christina das Doppelte von s_A weiter an Berta und die Hälfte von s_B weiter an Adrian. Ob Christina dabei insgesamt einen Gewinn oder Verlust gemacht hat, hängt von der Größe der Beträge s_A und s_B ab."

Wir haben beides mal denselben Sachverhalt beschrieben. Der Vorteil der zweiten Formulierung besteht darin, dass man umständliche sprachliche Konstruktionen wie „das Doppelte der von Adrian erhaltenen Summe" durch einfachere Formulierungen wie „das Doppelte von s_A" ersetzen kann.[40]

Eine weitere Beobachtung: Sogar in der ersten Formulierung kommen genau genommen schon Variablen vor: Für den beschriebenen Sachverhalt ist es ja völlig irrelevant, ob die Personen tatsächlich Adrian, Berta und Christina heißen – wir verstehen intuitiv sofort, dass sich nichts ändert, wenn wir die Namen durch andere ersetzen. Solange wir uns nicht auf eine ganze konkrete Situation beziehen, sind also „Adrian", „Berta" und „Christina" auch nur Variablen, die für generische Personen stehen. Wir könnten stattdessen z.B. auch „Person P_1", „Person P_2" und „Person P_3" schreiben.

In Beispiel 1.4.1 kam die Aussage $(r+1)^2 \leq r^2 + 3 \cdot r$ vor. Laut Definition 1.2.1 können wir diesen Ausdruck nur dann als Aussage $(r+1)^2 \leq r^2 + 3 \cdot r$ bezeichnen, wenn feststeht, ob die Ungleichung wahr oder falsch ist. Ob die Ungleichung wahr oder falsch ist, hängt aber vom Wert von r ab – zum Beispiel ist sie für $r = 0$ falsch!

Deshalb ist es sehr wichtig, dass es sich bei der Ungleichung $(r+1)^2 \leq r^2 + 3 \cdot r$ für jede reelle Zahl r um eine eigene Aussage handelt! Für $r = 4$ handelt es sich beispielsweise um die (wahre) Aussage $(4+1)^2 \leq 4^2 + 3 \cdot 4$; für $r = 0$ handelt es sich um die (falsche) Aussage $(0+1)^2 \leq 0^2 + 3 \cdot 0$.

D.h. die Ungleich $(r+1)^2 \leq r^2 + 3 \cdot r$ ist eine Kurzform für unendlich viele Aussagen, von denen manche wahr und manche falsch sind. Um generell mit solche Situationen umgehen zu können, ist die folgende Notation nützlich:

Notation 1.4.2. Wenn wir mehrere Aussagen betrachten, die von einer Variablen abhängen, so ist die folgende Notation häufig nützlich: Wir fassen die Aussagen mit einem Buchstaben – zum Beispiel A – zusammen und bringen die Abhängigkeit

[40] Noch etwas einfacher wird es natürlich, wenn man auch Formeln verwendet. Dann kann man z.B. anstelle von „Das Doppelte von s_A" einfach „$2s_A$" schreiben.

von der Variablen zum Ausdruck, in dem wir anschließend die Variablen in runden Klammern angeben.

Um das zu erläutern, betrachten wir nochmals das zuvor besprochene Beispiel:

Beispiel 1.4.3. Für jede reelle Zahl r bezeichnen wir mit die Aussage $A(r)$ die Aussage „$(r+1)^2 \leq r^2 + 3 \cdot r$“.

Aus Beispiel 1.4.1(b) wissen Sie bereits, dass die Aussage $A(r)$ für jede reelle Zahl $r \geq 1$ wahr ist.

Andererseits ist die Aussage $A(r)$ für jede reelle Zahl $r < 1$ falsch.[41]

Folgende Bemerkung ist wichtig, um korrekt mit Aussagen umzugehen, die von einer Variablen abhängen:

Bemerkung 1.4.4. Wenn Sie über Aussagen sprechen, die von einer Variablen abhängen – nennen wir die Variable zum Beispiel x und die Aussage $A(x)$ –, ist es wichtig, klar zum Ausdruck zu bringen, welche Werte für x Sie betrachten. Dies ist erstens nötig, um der Leserin oder dem Leser den richtigen Kontext zu vermitteln; und zweitens um sicherzustellen, dass $A(x)$ überhaupt eine sinnvolle Aussage ist (egal, ob die Aussage wahr oder falsch ist, zunächst muss es überhaupt eine Aussage sein).

Lassen Sie uns drei einfache Beispiele hierzu ansehen:

(a) Stellen Sie sich vor, jemand schreibt folgendes auf:

$$\text{Mit } A(x) \text{ bezeichnen wir die Aussage „}\tfrac{1}{x} \leq x\text{“.}$$

Es ist im Prinzip gar nicht möglich, zu verstehen, was genau hier gemeint ist, denn Sie können als Leserin oder Leser gar nicht wissen, was x überhaupt sein soll.[42]

(b) Geringfügig besser wäre es, wenn jemand folgendes schreibt:

$$\text{Für jede reelle Zahl } x \text{ bezeichnen wir mit } A(x) \text{ die Aussage „}\tfrac{1}{x} \leq x\text{“.}$$

Nun wissen Sie beim Lesen zumindest, dass x eine reelle Zahl sein soll. Wirklich Sinn ergibt das ganze trotzdem noch nicht, denn was genau soll mit $A(0)$ gemeint sein?

Beachten Sie hier unbedingt, dass $A(0)$ nicht etwa falsch ist – es lässt sich gar nicht entscheiden, ob $A(0)$ wahr oder falsch ist, denn der Bruch $\frac{1}{0}$ in der

[41] Versuchen Sie sich – als eine kleine Aufgabe – herauszufinden, warum $A(r)$ für $r < 1$ falsch ist.

[42] In diesem einfachen Fall können Sie mit etwas Fantasie vielleicht noch erraten, dass mit x wohl eine reelle Zahl gemeint ist – aber selbst das ist nicht wirklich klar, und Sie werden schon bald so viele verschiedene mathematische Objekte kennenlernen, dass es unmöglich sein wird, einfach zu erraten, welcher Typ von Objekt mit einer bestimmten Variable gemeint ist.

Ungleichung „$\frac{1}{0} \leq 0$“ ist schlichtweg nicht definiert. Also handelt es sich bei $A(0)$ gar nicht um eine Aussage.[43]

(c) Sinnvoll ist es zum Beispiel, wenn jemand schreibt:

Für jede Zahl $x \in \mathbb{R} \setminus \{0\}$ bezeichnen wir mit $A(x)$ die Aussage „$\frac{1}{x} \leq x$.“

Für jede reelle Zahl x, die nicht 0 ist, ist $A(x)$ nun tatsächlich eine Aussage. Die Aussage $A(2)$ ist zum Beispiel wahr, die Aussage $A(\frac{1}{2})$ ist hingegen falsch.[44]

Und-Verknüpfung und oder-Verknüpfung beliebig vieler Aussagen: Der Allquantor und der Existenzquantor

In Definitionen 1.2.5 und 1.2.7 haben wir aus zwei Aussagen durch und-Verknüpfung bzw. oder-Verknüpfung eine neue Aussage konstruiert. Mit Hilfe von Variablen können wir, wie soeben beschrieben, auch über unendlich viele Aussagen sprechen. Man kann auch unendlich viele Aussagen mit einem „und“ bzw. einem „oder“ verknüpfen. Um dies sprachlich präzise zu machen, benutzt man den sogenannten **Allquantor** und den sogenannten *Existenzquantor*:

Definition 1.4.5 (Allquantor und Existenzquantor)**.** Sei M eine Menge, und für jedes $x \in M$ sei eine Aussage $A(x)$ gegeben.

(a) Wir definieren mit Hilfe der gegebenen Aussagen $A(x)$ eine neue Aussage

$$\forall x \in M : \ A(x)$$

folgendermaßen:[45] Sie hat den Wahrheitswert „wahr“, wenn $A(x)$ für jedes x in M wahr ist, und sie hat den Wahrheitswert „falsch“, wenn es mindestes ein $x \in M$ gibt, für welches $A(x)$ falsch ist.

Man liest diese Aussage vor als „Für alle x in M gilt A von x.“ Das Symbol $\forall$ bezeichnet man als **Allquantor**.[46]

[43] Für mathematische Behauptungen, die weder wahr noch falsch, sondern in Wirklichkeit gar keine mathematischen Aussagen sind, verwenden Zyniker manchmal die englische Beschreibung „not even false“, was sich in etwa mit „noch nicht einmal falsch“ übersetzen lässt.

[44] Überlegen Sie sich bei Gelegenheit einmal, für welche $x \in \mathbb{R} \setminus \{0\}$ die Aussage $A(x)$ wahr ist und für welche $x \in A(x)$ sie falsch ist.

[45] Den Doppelpunkt hinter „$\forall x \in M$“ kann man genau genommen auch weglassen. Im täglichen Umgang mit logischen Ausdrücken ist es aber oft nützlich, ihn zu verwenden, da komplizierte Aussagen hierdurch etwas übersichtlicher werden.

[46] Man kann es etwas unglücklich finden, dass als Allquantor das „nach oben geöffnete“ Symbol $\forall$ verwendet wird, während das logische und mit dem „nach unten geöffneten“ Symbol $\wedge$ bezeichnet wird. Allerdings haben sich diese Symbole eingebürgert, und man gewöhnt sich recht schnell daran.

Vielleicht hilft Ihnen am Anfang auch die folgende Eselsbrücke: Der Allquantor $\forall$ sieht aus wie ein umbedrehtes „A“ aus dem Wort „Alle“.

(b) Sei M eine Menge, und für jeds $x \in M$ sei eine Aussage $A(x)$ gegeben. Dann definieren wir eine neue Aussage

$$\exists x \in M : \ A(x)$$

folgendermaßen:[47] Sie hat den Wahrheitswert „wahr“, wenn es mindestes ein $x \in M$ gibt, für welches $A(x)$ wahr ist, und sie hat den Wahrheitswert „falsch“, wenn $A(x)$ für alle x in M falsch ist.

Man liest diese Aussage vor als „Es gibt x in M, für das A von x gilt“, oder als „Es gibt ein x in M, für das gilt: A von x.“[48] Das Symbol $\exists$ bezeichnet man als **Existenzquantor**.

Es ist wichtig, sich den Zusammenhang mit der und-Verknüpfung und der oder-Verknüpfung von Aussagen klarzumachen:

Beispiel 1.4.6. Lassen Sie uns mit $A(1)$ die Aussage

„Friedrich Schiller hat das Drama *Die Räuber* geschrieben“

bezeichnen, mit $A(2)$ die Aussage

„Conrad F. Meyer hat das Gedicht *Die Brück' am Tay* geschrieben“,

und mit $A(3)$ die Aussage

„Max Frisch hat *Herr Biedermann und die Brandstifter* geschrieben“.

Für jedes $n \in \{1, 2, 3\}$ haben wir hier also eine Aussage $A(n)$.[49] Die Aussage $A(1)$ ist wahr, die Aussage $A(2)$ is falsch, und die Aussage $A(3)$ ist wahr. Aus diesen Aussagen können wir neue Aussagen konstruieren:

(a) Die Aussage $A(1) \wedge A(2) \wedge A(3)$ kann man auch in der Form

$$\forall n \in \{1, 2, 3\} : \ A(n)$$

schreiben. Sie ist falsch (weil $A(2)$ falsch ist).

(b) Die Aussage $A(1) \wedge A(3)$ kann man auch in der Form

$$\forall n \in \{1, 3\} : \ A(n)$$

schreiben. Sie ist wahr (weil $A(1)$ und $A(3)$ beide wahr sind).

[47] Auch hier kann man den Doppelpunkt hinter „$\exists x \in M$“ genau genommen weglassen, aber es ist häufig übersichtlicher, ihn mit anzuschreiben.

[48] Wenn man sich noch etwas klarer und unmissverständlicher ausdrücken will, kann man zum Beispiel die Formulierung „Es gibt mindestens ein x in $M \ldots$“ anstelle von „Es gibt ein x in $M \ldots$“ verwenden.

[49] Hier können Sie übrigens beobachten: Variablen, die für Zahlen stehen, müssen nicht unbedingt verwendet werden, um Größen zu beschreiben, mit denen man rechnen möchte. Man kann sie auch schlicht benutzen, um Ding durchzunummerieren.

(c) Die Aussage $A(1) \vee A(2) \vee A(3)$ kann man auch in der Form

$$\exists n \in \{1, 2, 3\} : \ A(n)$$

schreiben. Sie ist wahr (weil z.B. $A(1)$ wahr ist).

(d) Die Aussage $A(2) \vee A(3)$ kann man auch in der Form

$$\exists n \in \{2, 3\} : \ A(n)$$

schreiben. Sie ist wahr (weil $A(3)$ wahr ist).

Sie fragen sich vermutlich, wozu der Allquantor und der Existenzquantor taugen, wenn wir doch auch einfach die Symbole $\wedge$ bzw. $\vee$ verwenden können. Hier sind einige Situationen, in denen der Allquantor (bzw. Existenzquantor) sehr nützlich ist:

- Wenn Sie sehr viele Aussagen mit einer und-Verknüpfung (bzw. einer oder-Verknüpfung) verknüpfen möchten.
- Wenn Sie sogar unendlich viele Aussagen mit einer und-Verknüpfung (bzw. einer oder-Verknüpfung) verknüpfen möchten.
- Wenn Sie die Menge, aus der die Variablen stammen, die Sie mit einer und-Verknüpfung (bzw. einer oder-Verknüpfung) verknüpfen möchten, nicht genauer spezifiziert ist.

Lassen Sie uns für die und- bzw. oder-Verknüpfung von unendlich vielen Aussagen mit Hilfe von Quantoren noch ein Beispiel besprechen:

Beispiele 1.4.7. Lassen Sie uns noch einmal die Ungleichung $(r+1)^2 \leq r^2 + 3r$ für reelle Zahlen r betrachten.

(a) Die Aussage

$$\forall r \in \mathbb{R} : \ (r+1)^2 \leq r^2 + 3r$$

ist falsch (weil die Ungleichung zum Beispiel für $r = 0$ falsch ist).

(b) Wenn wir aber die Menge $L \coloneqq \{x \in \mathbb{R} \mid \ x \geq 1\}$ betrachten, dann ist die Aussage

$$\forall r \in L : \ (r+1)^2 \leq r^2 + 3r$$

wahr.

Übrigens ist es natürlich etwas umständlich, extra die Menge L einzuführen um auszudrücken, dass der Allquantor sich auf alle reellen Zahl, die größer oder gleich 1 sind, bezieht. Deshalb kombiniert man den Quantor häufig mit einfach

natürlich-sprachlichen Ausdrücken um dasselbe zum Ausdruck zu bringen; zum Beispiel so:

$$\forall r \in \mathbb{R} \text{ mit } r \geq 1 : \ (r+1)^2 \leq r^2 + 3r.$$

Diese Aussage kann man z.B. folgendermaßen vorlesen: „Für alle r in $\mathbb{R}$ mit $r \geq 1$ gilt $(r+1)^2 \leq r^2 + 3r$.“ Oder noch etwas ausführlicher: „Für alle r in $\mathbb{R}$ mit der Eigenschaft $r \geq 1$ gilt $(r+1)^2 \leq r^2 + 3r$.“

(c) Die Aussage

$$\exists r \in \mathbb{R} : \ (r+1)^2 \leq r^2 + 3r$$

ist wahr, weil z.B. Ungleichung $(4+1)^2 \leq 4^2 + 3 \cdot 4$ wahr ist.

Im Kontext der vorangehenden Beispiele hier nochmals ein wichtiger Hinweis: Bitte achten Sie, wie bereits früher erwähnt, darauf, dass alle Aussagen, die Sie vorlesen, grammatisch sinnvolle Sätze ergeben müssen. Wenn Sie eine Aussage vorlesen und den Eindruck haben, dass das, was Sie sagen, grammatisch keinen Sinn ergibt, dann halten Sie inne und versuchen Sie, das Vorgelesene zu korrigieren.

Bemerkung 1.4.8 (Quantifizierung über die leere Menge)**.** In Definition 1.4.5 ist auch der Fall zugelassen, dass die Menge M leer ist. In diesem Fall ist die Aussage

$$\forall x \in M : \ A(x)$$

wahr[50], und die Aussage

$$\exists x \in M : \ A(x)$$

falsch.[51]

In der formalen Logik wird übrigens sehr präzise beschrieben, wie genau die Quantoren $\forall$ und $\exists$ zu verwenden sind (wesentlich genauer, als wir dies hier tun). Für den alltäglichen Gebrauch in der Mathematik genügt es aber oft, sich das Symbol $\forall$ tatsächlich als Abkürzung für „für alle“ zu denken, und sich das Symbol $\exists$ als Abkürzung von „es gibt ein“ zu denken.

Entsprechend wird es häufig vorkommen, dass wir anstelle der Symbole $\forall$ und $\exists$ einfach die Worte „für alle“ und „es gibt ein“ ausschreiben.

Exklusiv-oder-Verknüpfung beliebig vieler Aussagen

Es gibt noch einen weiteren Quantor, der immer von Bedeutung ist, weil er eine die exklusiv-oder-Verknüpfung auf beliebig viele Aussagen verallgemeinert:

[50] Bitte überlegen Sie sich in Ruhe, weshalb.
[51] Bitte überlegen Sie sich auch hier in Ruhe, weshalb.

Definition 1.4.9 (Existenz- und Eindeutigkeitsquantor)**.** Sei M eine Menge, und für jedes $x \in M$ sei eine Aussage $A(x)$ gegeben. Dann definieren wir eine neue Aussage

$$\exists! x \in M : \ A(x)$$

folgendermaßen:[52] Sie hat den Wahrheitswert „wahr“, wenn es genau ein $x \in M$ gibt, für welches $A(x)$ wahr ist (und $A(x)$ somit für alle anderens $x \in M$ falsch ist). Sie hat den Wahrheitswert „falsch“, wenn $A(x)$ für alle $x \in M$ falsch ist oder wenn $A(x)$ für mindestens zwei verschiedene $x \in M$ wahr ist.

Man liest diese Aussage vor als „Es gibt genau ein x in M, für das A von x gilt“.

Beispiele 1.4.10. (a) Die Aussage

$$\exists! x \in \mathbb{R} : \quad x^2 = 4$$

ist falsch, denn es gibt zwei reelle Zahlen, deren Quadrat gleich 4 ist (nämlich 2 und -2).

(b) Die Aussage

$$\exists! x \in \mathbb{R} : \quad x^2 = 4 \ \wedge \ x \geq 0$$

ist hingegen wahr, denn es gibt genau eine reelle Zahl, deren Quadrat gleich 4 ist und die zugleich größer oder gleich 0 ist (nämlich die Zahl 2).

(c) Die Aussage

$$\exists! x \in \mathbb{R} : \quad x^2 = -1$$

ist falsch, denn es gibt gar keine reelle Zahl, deren Quadrat gleich -1 ist.

Verneinung und Verschachtelung von Quantoren

Sie werden sehr häufig in die Situation kommen (insbesondere in der Analysis, aber auch bereits im aktuellen Semester in der Linearen Algebra 1), die Sie Aussagen, die Quantoren enthalten, verneinen müssen. Anhand der Definition des All- und de Existenzquantors in Definition 1.4.5 kann man sich leicht überlegen, wie das funktioniert:

Bemerkung 1.4.11 (Verneinung von Aussagen, die Quantoren enthalten)**.** Sei M eine Menge und für jedes $x \in M$ sei eine Aussage $A(x)$ gegeben. Dann gilt:

(a) Die Aussage

$$\neg(\forall x \in M : \ A(x))$$

hat denselben Wahrheitswert wie die Aussage

$$\exists x \in M : \ \neg A(x).$$

[52] Manche Autorinnen und Autoren verwenden anstelle des Symboll $\exists!$ das Symbol $\exists_1$.

(b) Die Aussage

$$\neg(\exists x \in M : \ A(x))$$

hat denselben Wahrheitswert wie die Aussage

$$\forall x \in M : \ \neg A(x).$$

Richtig interessant wird es, wenn man Aussagen baut, in denen mehrere Quantoren verschachtelt sind. Auch dies kommt sehr häufig vor; hier ein Beispiel mit einer ersten Kostprobe:

Beispiel 1.4.12 (Verschachtelung von Quantoren)**.** Es bezeichne A die Aussage

$$\forall k \in \mathbb{N} : \ \exists n \in \mathbb{N} : \ n^2 \geq k + 1,$$

und es bezeichne B die Aussage

$$\exists n \in \mathbb{N} : \ \forall k \in \mathbb{N} : \ n^2 \geq k + 1.$$

Dann ist A wahr, B hingegen nicht.

Daran können Sie erkennen, dass man Quantoren nicht einfach vertauschen darf – es kommt auf die Reihenfolge an! Dieses Beispiel werden Sie in den Tutorien noch genauer besprechen.

Durchschnitte und Vereinigungen von Mengen: Noch einmal

In Definition 1.3.14 hatten wir mit Hilfe des logischen Unds und des logischen Oders den Durchschnitt und die Vereinigung von je zwei Mengen definiert. Ebenso kann man mit Hilfe des Allquantors- und mit Hilfe des Existenz-Quantors den Durchschnitt und die Vereinigung von beliebig vielen Mengen definieren:

Definition 1.4.13 (Durchschnitt und Vereinigung beliebig vieler Mengen)**.** Sei I eine nicht-leere Menge und für jedes $i \in I$ sei eine Menge M_i gegeben.

(a) Die Menge

$$\bigcap_{i \in I} M_i \coloneqq \{x \mid \ \forall i \in I : \ x \in M_i\}$$

heißt der **Durchschnitt** der Mengen M_i für $i \in I$.

(b) Die Menge

$$\bigcup_{i \in I} M_i \coloneqq \{x \mid \ \exists i \in I : \ x \in M_i\}$$

heißt die **Vereinigung** der Mengen M_i für $i \in I$.

Wenn man zwei Mengen M_1, M_2 gegeben hat (d.h., wenn $I = \{1, 2\}$ ist), kann man sehen, dass

$$\bigcap_{i\in\{1,2\}} M_i = \{x \mid \forall i \in \{1,2\} : x \in M_i\} = \{x \mid x \in M_1 \wedge x \in M_2\} = M_1 \cap M_2.$$

gilt.[53] Ebenso kann man sehen, dass in diesem Fall

$$\bigcup_{i\in\{1,2\}} M_i = M_1 \cup M_2$$

gilt.[54]

Implikationen und Äquivalenz

Nun kommen wir noch einmal zurück zur Verknüpfung von Aussagen (zunächst ganz ohne Variablen; in Abschnitt 1.2 hatten wir bereits mehrere Verknüpfungen von Aussagen eingeführt. Nun folgen noch drei weitere Verknüpfungen:

Definition 1.4.14 (Implikationen und Äquivalenz)**.** Seien A und B beliebige Aussagen. Wir definieren drei weitere Aussagen $A \Rightarrow B$, $A \Leftarrow B$, $A \Leftrightarrow B$, deren Werte durch die folgenden Wahrheitstabellen festgelegt werden:

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

A	B	$A \Leftarrow B$
w	w	w
w	f	w
f	w	f
f	f	w

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

Für die so definierten Aussagen verwenden wir die folgenden Sprechweisen:

(a) Die Aussage $A \Rightarrow B$ wird als „A impliziert B“ ausgesprochen, oder als „Wenn A, dann auch B“, oder als „Aus A folgt B“. Manchmal sagt man stattdessen auch „A ist hinreichend für B“ oder „B ist notwendig für A“.[55]

[53] Hier haben wir also die Mengengleichheit $\bigcap_{i\in\{1,2\}} M_i = M_1 \cap M_2$ bewiesen, indem wir die Menge $\bigcap_{i\in\{1,2\}} M_i$ solange anders dargestellt haben – ohne die Menge selbst dabei zu verändern – bis wir die Menge $M_1 \cap M_2$ erhalten haben. Selbstverständlich kann man die Mengengleichheit $\bigcap_{i\in\{1,2\}} M_i = M_1 \cap M_2$ aber auch zeigen, indem man die Methode verwendet, die im Beweis von Proposition 1.3.15 vorgestellt wurde – d.h., indem man die beiden Inklusionen „$\subseteq$“ und „$\supseteq$“ einzeln beweist.. Davon überzeugen Sie sich am besten, indem Sie es auf einem Blatt Papier (oder auf einem Tablet) selbst versuchen.

[54] Überprüfen Sie für hier die Details unbedingt auf einem Blatt Papier noch einmal selbst um sicherzustellen, dass Sie das richtig verstanden haben.

[55] Bei einem Feierabendbier können Sie sich ein wenig den Kopf darüber zerbrechen, weshalb man hier die Begriffe „hinreichend“ und „notwendig“ verwendet. Oder Sie bleiben lieber bei drei erst genannten Formulierungen, die intuitiv vermutlich etwas klarer sind.

(b) Weil die Aussage $A \Leftarrow B$ immer denselben Wahrheitswert wie $B \Rightarrow A$ hat, spricht man sie genauso aus wie $B \Rightarrow A$.[56]

(c) Die Aussage $A \Leftrightarrow B$ wird ausgesprochen als „A ist äquivalent zu B“ oder als „A genau dann, wenn B“ oder als „A dann und nur dann, wenn B“.

Bevor wir genauer darauf eingehen, was es mit den Implikationen $A \Rightarrow B$ und $B \Rightarrow A$ auf sich hat, sind zwei Bemerkungen sinnvoll:

Bemerkungen 1.4.15. (a) Die Äqujivalenz $A \Leftrightarrow B$ ist in genau denjenigen Fällen wahr, in denen A und B den gleichen Wahrheitswert haben. Zu sagen „Es gilt die Äquivalenz $A \Leftrightarrow B$“ ist somit eine andere Möglichkeit um zu sagen „A und B haben denselben Wahrheitswert.“

Auf diese Weise kann man einige der Resultate aus Abschnitt 1.2 formulieren, indem man Äquivalenzen verwendet. Zum Beispiel besagt Proposition 1.2.10(a) für jede Aussage A und jede Aussage B folgendes: die Aussage $\neg(A \wedge B)$ hat immer denselben Wahrheitswert wie die Aussage $(\neg A) \vee (\neg B)$. Genauso gut könnten wir auch sagen: es gilt stets

$$\Big(\neg(A \wedge B)\Big) \Leftrightarrow \Big((\neg A) \vee (\neg B)\Big).$$

(b) Mit Hilfe einer Wahrheitstabelle kann man sich leicht von folgendem überzeugen:[57] Die Aussage $A \Leftrightarrow B$ hat stets denselben Wahrheitswert wie die Aussage $(A \Rightarrow B) \wedge (A \Leftarrow B)$.[58]

Also bedeutet „A ist äquivalent zu B“ dasselbe wie „Aus A folgt B und aus B folgt A“.

Diese harmlos anmutende Beobachtung wird Sie den Rest Ihres Studiums verfolgen, den sehr viele Resultate in der Mathematik sind als Äquivalenzen formuliert, und diese werden meist bewiesen, indem man die beiden Implikationen einzeln beweist.

Wie versprochen besprechen wir nun, was es mit der Implikation $A \Rightarrow B$ auf sich hat.[59] Erfahrungsgemäß fällt es vielen Studierenden am Anfang schwer, die vorletzte Zeile in der Wahrheitstabelle von $A \Rightarrow B$ intuitiv nachzuvollziehen – warum sollte es richtig sein zu sagen, dass aus etwas Falschem etwas Wahres folgt? Am einfachsten können Sie dies vermutlich nachvollziehen, wenn Sie nicht nur einzelne Aussagen betrachten, sondern Aussagen, die von einer Variablen abhängen:

[56] Also z.B. als „B impliziert A“ oder „Aus B folgt A“.

[57] Und das sollten Sie sogleich auf einem Blatt Papier tun!

[58] Übrigens können Sie hier gleich testen, ob Sie Teil (a) der Bemerkung verstanden haben: Wie können Sie den Satz, der mit dieser Fußnote abschließt, stattdessen formulieren, wenn Sie anstelle der Worte “stets denselben Wahrheitswert” lieber einen Äquivalenzpfeil verwenden möchten?

[59] Sobald Sie diese Implikation wirklich verstanden haben, verstehen Sie automatisch auch die Bedeutung der Implikation $B \Rightarrow A$ (denn hierbei sind ja nur die Bezeichnungen der Aussagen vertauscht) und somit auch die Implikation $A \Leftarrow B$ (weil diese ja äquivalent zu $B \Rightarrow A$ ist).

Diskussion 1.4.16. Sei M eine Menge, und für jedes $x \in M$ seien Aussagen $A(x)$ und $B(x)$ gegeben. Wie Sie bereits wissen, kann es passieren, dass $A(x)$ für manche $x \in M$ wahr ist, und für andere $x \in M$ falsch; ebenso kann es passieren, dass $B(x)$ für manche x aus M wahr ist, für andere hingegen nicht. Diejenigen x, für die $A(x)$ wahr ist, müssen natürlich nicht unbedingt dieselben sein, für die auch $B(x)$ wahr ist.

Nun will man in konkreten Situationen häufig wissen, wie die Aussagen $A(x)$ und $B(x)$ zusammenhängen. Besonders interessant ist zum Beispiel die folgende Situation, die wir zunächst umgangssprachlich beschreiben:

$$\text{Für jedes } x \text{ in } M \text{ gilt: Wenn } A(x) \text{ gilt, muss auch } B(x) \text{ gelten.} \tag{1.4.3}$$

Beachten Sie, dass „Wenn $A(x)$ gilt, muss auch $B(x)$ gelten“, kein Aussage darüber macht, was passiert, wenn $A(x)$ falsch ist. Wir können (1.4.3) also auch folgendermaßen formulieren:

Für jedes x in M tritt einer der folgenden beiden Fälle auf:
(i) $A(x)$ and $B(x)$ sind beide wahr; (ii) $A(x)$ ist falsch;

oder etwas kürzer und formallastiger aufgeschrieben:

$$\forall x \in M : \ \big(A(x) \wedge B(x)\big) \vee \big(\neg A(x)\big).$$

Nun hat $\big(A(x) \wedge B(x)\big) \vee \big(\neg A(x)\big)$ stets denselben Wahrheitswert wie $A(x) \Rightarrow B(x)$ lauft Definition 1.4.14.[60,61]

Das heißt, die Wahrheitstablle für die Implikation in Definition 1.4.14 erlaubt es uns, die Aussage (1.4.3) in der Form

$$\forall x \in M : \ (A(x) \Rightarrow B(x))$$

zu schreiben[62] – und das ist ja durchaus intuitiv.

Zum Abschluss dieses Abschnitts wollen wir noch einmal demonstrieren, wie man Äquivalenzen verwenden kann um mathematische Resultate zu formulieren – und wie man Bemerkung 1.4.15(b) verwenden kann, um solche Resultate zu beweisen. Als Anschauungsobjekt verwenden wir das folgende Resultat:

Proposition 1.4.17. *Seien L, M Mengen. Dann gilt*

$$L \subseteq M \quad \Leftrightarrow \quad L \cap M = L.$$

[60] Achtung: Glauben Sie das nicht einfach! Seien Sie kritisch und überzeugen Sie sich selbst, indem Sie alle vier möglichen Fälle durchgehen.

[61] Der spannenste Fall ist hier natürlich derjenige, in dem uns die linksstehende Wahrheitstabelle aus Definition 1.4.14 auf den ersten Blick unintuitiv erscheint – also, wenn $A(x)$ falsch und $B(x)$ wahr ist. In diesem Fall ist $\big(A(x) \wedge B(x)\big) \vee \big(\neg A(x)\big)$ wahr, und dies erklärt, warum der Eintrag in der vorletzten Zeile der Wahrheitstabelle so gewählt wird, wie in Definition 1.4.14 beschrieben.

[62] Die Klammer um $A(x) \Rightarrow B(x)$ haben wir hier nur der einfacheren Lesbarkeit halber hinzugefügt.

Die Proposition besagt also in Worten: Für zwei Mengen M und L ist die Teilmengenbeziehung $L \subseteq M$ genau dann erfüllt, wenn der Durschnitt $L \cap M$ gleich L ist. Wir beweisen die Proposition in Kürze. Vorher aber ist folgende Bemerkung extrem wichtig:

Bemerkung 1.4.18. In Proposition 1.4.17 sehen Sie eine Sprechweise, die in mathematischen Resultaten sehr üblich ist: Man beginnt mit dem Wort „Seien“ und führt dann einige Objekte ein. Anschließend macht man eine Aussage über diese Objekt.

Diese Formulierung ist als eine Verwendung eines Allquantors zu verstehen, nur dass sie sprachlich über mehrere Sätze verteilt ist. Die komplette Aussage von Proposition 1.4.17 könnte man zum Beispiel ganz knapp und formal auch in folgender Form schreiben:

$$\forall \text{ Mengen } L, M: \qquad L \subseteq M \quad \Leftrightarrow \quad L \cap M = L.$$

Beweis von Proposition 1.4.17. Unser Ziel ist es, eine Äquivalenz zu beweisen. Laut Bemerkung 1.4.15(b): ist dies gleichbedeutend damit, die beiden Implikationen

$$L \subseteq M \quad \Rightarrow \quad L \cap M = L.$$

und

$$L \subseteq M \quad \Leftarrow \quad L \cap M = L.$$

zu beweisen. Dies tun wir im folgenden.

„$\Rightarrow$“ Es gelte $L \subseteq M$. Wir müssen $L \cap M = L$ zeigen, und dies tun wir wie üblich, indem wir beide Inklusionen zeigen.

- „$\subseteq$“ Sei $x \in L \cap M$ beliebig, aber fest. Dann gilt wegen der Definition des Durchschnitts automatisch $x \in L$. Somit ist gezeigt, dass $L \cap M \subseteq L$ gilt.
- „$\supseteq$“ Sei $x \in L$ beliebig, aber fest. Weil $L \subseteq M$ ist, gilt dann auch $x \in M$. Das heißt, insgesamt gilt $x \in L$ und $x \in M$, also $x \in L \cap M$. Somit haben wir $L \subseteq L \cap M$ gezeigt.

Insgesamt ist also $L \cap M = L$.

„$\Leftarrow$“ Sei nun $L \cap M = L$. Wir müssen $L \subseteq M$ zeigen.

Sie also $x \in L$ beliebig aber fest. Wegen der Voraussetzung $L \cap M = L$ gilt dann auch $x \in L \cap M$, und somit $x \in M$. Somit haben wir $L \subseteq M$ gezeigt. □

1.5 Funktionen

Was ist eine Funktion?

Bisher haben wir nicht nur über mathematisch Aussagen gesprochen, sondern auch über verschiedene mathematische Objekte: Zum Beispiel Zahlen, Mengen und Tupel. Um die Mathematik wirklich „zum Leben zu erwecken“, brauchen wir aber noch einen weiteren Typ von mathematischen Objekten: **Funktionen**.

Definition 1.5.1 (Funktion/Abbildung)**.** Seien X, Y Mengen. Eine **Funktion** (oder **Abbildung**) f von X nach Y ist eine Zuordnungsvorschrift, die jedem Element $x \in X$ ein eindeutig bestimmtes Element aus Y – welches wir mit $f(x)$ bezeichnen – zuweist.

Wir nennen X den **Definitionsbereich** von X und Y den **Wertebereich** von f.

Beachten Sie unbedingt: Um eine Funktion konkret anzugeben, müssen Sie auf jeden Fall auch den Definitions- und den Wertebereich angeben. Funktionen erhalten nicht auf magische Weise von selbst einen Werteebereich,[63] sondern der Definitionsbereich der Funktion muss explizit angegeben werden, wenn man eine Funktion angibt. Ohne Angabe des Definitionsbereich kann man eine Funktion nicht wirklich verstehen.

Wenn man keine Lust hat, Sätze wie „Sei f eine Funktion von X nach Y“ jedes mal auszuschreiben, ist folgende Notation nützlich:

Notation 1.5.2. Seien X, Y Mengen. Jede der folgenden beiden Notationen wird synonym mit dem Satz „Sei f eine Funktion von X nach Y“ verwendet:

(a) Sei $f : X \to Y$.

(b) Sei $X \xrightarrow{f} Y$.

Verschiedene Arten um Funktionen zu beschreiben

Bis jetzt haben wir lediglich abstrakt gesagt, was man unter einer Funktion versteht. Um mit diesem Begriff arbeiten zu können, brauchen man natürlich Möglichkeiten, um eine Funktion explizit anzugeben. Es gibt verschiedene solche Möglichkeiten; einige wichtige stellen wir im folgenden vor:

Beispiele 1.5.3. Seien X, Y Mengen. Es folgen einige Möglichkeiten um eine konkrete Funktion von X nach Y zu spezifieren.

(a) Wenn X nur endlich viele Elemente hat, kann man eine Funktion $f : X \to Y$ angeben, indem man alle Elemente x von X aufzählt und für jedes dieser Elemente den Wert $f(x)$ konkret angibt – zum Beispiel in einer Tabelle.

Sei zum Beispiel $X = \{1, 2, 3, 4\}$ und $Y = \mathbb{R}$. Wir definieren eine Funktion $f : X \to Y$ durch die folgende Wertetabelle, in der alle Elemente $x \in X$ aufgezählt sind:

x	1	2	3	4
$f(x)$	0	$-\frac{3}{2}$	0	π

[63] Auch, wenn in der Schule manchmal dieser Eindruck erweckt wird.

(b) Manche Funktionen lassen sich mit Hilfe von Formeln darstellen. Wir können zum Beispiel eine Funktion $g : \mathbb{R} \to \mathbb{R}$ durch die Formel

$$g(x) = x^3 - 7 \qquad \text{für alle } x \in \mathbb{R}$$

definieren.[64,65]

Anstatt „$g(x) = x^3 - 7$ für alle $x \in \mathbb{R}$“ zu schreiben, benutzt man in der Mathematik auch sehr häufig die Notation „$x \mapsto x^3 - 7$“.[66]

Wenn man eine Funktion sehr effizient definieren will, kann man die Notation für Definitions- und Wertebereich sowie die zugehörige Formel[67] auch direkt untereinander schreiben – zum Beispiel kann man eine Funkton h so folgendermaßen beschreiben:

$$\begin{aligned} h : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto 7x - 5. \end{aligned}$$

Übrigens können Funktionen natürlich auch kompliziertere Definitions- und Wertebereiche haben – betrachten Sie als Beispiel die Funktion

$$\begin{aligned} k : \mathbb{R}^2 &\to \mathbb{R}^3, \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\mapsto \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \\ 2x_1 \end{pmatrix}. \end{aligned}$$

(c) Eine weitere nützliche Möglichkeit um manche Funktionen zu beschreiben ist die Verwendung einer *Fallunterscheidung*. Hier ist ein Beispiel: Es sei[68]

$$\begin{aligned} \ell : \mathbb{R} &\to \mathbb{R}, \\ \ell(z) &= \begin{cases} 2z & \text{falls } z \geq 0, \\ -1 & \text{falls } z < 0. \end{cases} \end{aligned}$$

[64] Beachten Sie aber unbedingt, dass man nicht jede Funktion von $\mathbb{R}$ nach $\mathbb{R}$ mit solch einer einfachen Formel ausdrücken kann. Man kann sogar beweisen, dass es Funktionen von $\mathbb{R}$ nach $\mathbb{R}$ gibt, die sich gar nicht mit Hilfe einer Formel darstellen lassen!

[65] Übrigens haben wir uns hier ein wenig am eigenen Schopf aus dem Sumpf gezogen: Eine Funktion auf diese Weise zu definieren, ist uns nur deshalb möglich, weil wir bereits einige andere Funktionen kennen und von ihnen hier sehr freigiebig Gebrauch machen: Addition und Multiplikation.
Über Addition, Multiplikation und ähnliche Funktionen werden wir in Abschnitt 2 noch ausgiebig diskutieren.

[66] Beachten Sie hier den vertikalen Strich am Beginn des Pfeils $\mapsto$. Der Pfeil $\to$, mit dem angegeben wird, von wo nach wo die Funktion abbildet, hat diesen Strich nicht.

[67] Sofern sich die Funktion durch eine Formel beschreiben lässt.

[68] In diesem Beispiel sehen Sie übrigens, dass man die Variable keineswegs x nennen muss – auch jede andere Variable ist in Ordnung.

Natürlich könnte man dasselbe genauso gut mit Hilfe des Pfeils $\mapsto$ zum Ausdruck bringen, indem man stattdessen schreibt: Es sei

$$\begin{aligned} \ell : \mathbb{R} &\to \mathbb{R} \\ z &\mapsto \begin{cases} 2z & \text{falls } z \geq 0, \\ -1 & \text{falls } z < 0. \end{cases} \end{aligned}$$

Übrigens ist die Spezifizierung einer Funktion mit Hilfe einer Tabelle, die wir in Beispiel (a) besprochen haben, auch nur eine Kurzschreibweise für eine Fallunterscheidung. Beispielsweise kann man die Funktion f aus Beispiel (a) auch folgendermaßen beschreiben:

$$\begin{aligned} f : \{1,2,3,4\} &\to Y, \\ x &\mapsto \begin{cases} 0 & \text{falls } x = 1, \\ -\frac{3}{2} & \text{falls } x = 2, \\ 0 & \text{falls } x = 3, \\ \pi & \text{falls } x = 4. \end{cases} \end{aligned}$$

In diesem Fall ist die Tabelle aber vielleicht etwas übersichtlicher.

Funktion, Argument und Funktionswert

In der Mathematik ist die folgende Terminologie üblich:

Vereinbarung 1.5.4 (Argument und Funktionswert)**.** Seien X, Y Mengen und sei $f : X \to Y$. Für ein Element $x \in X$ betrachten wir die Notation „$f(x)$“: in dieser Notation heißt

- $\ldots x$ das **Argument**,
- und $f(x)$ der **Wert von f an dieser Stelle x**.[69]

Die folgende Bemerkung finden Sie auf den ersten Blick wahrscheinlich subtil[70], aber sie ist enorm wichtig um im Laufe der Vorlesung – und im Laufe Ihres restlichen Studiums – korrekt mit Funktionen umzugehen:

Bemerkung 1.5.5 (Funktion vs. Funktionswert)**.** Man muss auf jeden Fall eine *Funktion* von ihren *Funktionswerten* unterscheiden: Seien X, Y Mengen und sei $f : X \to Y$. Es ist *nicht* richtig zu sagen, „$f(x)$ ist eine Funktion“. Die Funktion heißt f, nicht $f(x)$. Mit $f(x)$ ist etwas anderes gemeint: Für ein Element $x \in X$ ist $f(x)$ – wie Sie der vorangehenden Vereinbarung entnehmen können – der Wert von f an

[69] Oft sagt man hier anstelle von „Wert“ auch etwas länger „Funktionswert“.

[70] Und womöglich steht die Bemerkung auch der Art entgegen, wie Sie in der Schule über Funktionen gedacht oder gesprochen haben.

der Stelle x. Es handelt sich bei $f(x)$ also um ein Element von Y, während es sich bei f um eine Zuordnung von X nach Y handelt.

Das lässt sich am besten mit Hilfe eines Beispiels veranschaulichen: Sei X die Menge aller Einwohner von Passau, die im Telefon verzeichnet sind, und sei Y die Menge aller in Passau vergebenen Telefonnummern. Dann ist das Telefonbuch von Passau schlicht und einfach diejenige Funktion f, die jeder Person $x \in X$ ihre Telefonnummer zuordnet. Hier sehen Sie den Unterscheid zwischen f und $f(x)$: Bei f handelt es sich um das gesamte Telefonbuch; bei $f(x)$ handelt es sich um eine einzelne Telefonnummer[71].

Hintereinanderausführung und Gleichheit von Funktionen

Wenn wir zwei Funktionen f und g gegeben haben und der Wertebereich von f mit dem Definitionsbereich von g übereinstimmt, dann können wir die beiden Funktionen zu einer neuen Funktion verknüpfen:

Definition 1.5.6 (Hintereinanderausführung)**.** Seien X, Y, Z Mengen und seien $f : X \to Y$ und $g : Y \to Z$. Dann definieren wir eine Funktion $g \circ f : X \to Z$ durch die Formel[72]

$$(g \circ f)(x) = g\big(f(x)\big) \quad \text{für alle } x \in X.$$

Die Funktion $g \circ f$ heißt die **Hintereinanderausführung** (oder **Komposition**) von f und g.

Wenn man anstelle der Notation $f : X \to Y$ und $g : Y \to Z$ die Notation $X \xrightarrow{f} Y$ und $Y \xrightarrow{g} Z$ verwenden[73], lässt sich die Komposition von $g \circ f$ besonders anschaulich in der Form

$$X \xrightarrow{f} Y \xrightarrow{g} Z$$

darstellen.

Wir wollen in Kürze das sogenannten *Assoziativgesetz* für die Hintereinanderausführung von Funktionen beweisen.[74] Damit wir dies sinnvoller Weise tun können,

[71] Welche Telefonnummer das konkret ist, kann man natürlich nur wissen, wenn man weiß, welche Person gerade mit x gemeint ist.

[72] Beachten Sie hierbei unbedingt die Reihenfolge: Anschaulich gesprochen wird f zuerst ausgeführt und dann erst g – man schreibt in der Notation $g \circ f$ allerdings g nach links.
Der Grund für diese Konvention ist einfach: Man möchte in den beiden Ausdrücken $(g \circ f)(x)$ und $g\big(f(x)\big)$ die Funktionen f und g gerne in derselben Reihenfolge anschreiben.

[73] Die wir ebenfalls in Notation 1.5.2 eingeführt hatten.

[74] Ein Assoziativgesetz für die Addition und die Multiplikation reeller Zahlen kennen Sie bereits aus der Schule; und einige Assoziativgesetze in der Aussagenlogik haben Sie bereits in Proposition 1.2.12 gesehen. Auf dieser Grundlagen können Sie vielleicht jetzt schon erraten, was sich hinter dem Begriff „Assoziativgesetz für die Hintereinanderausführung von Funktionen“ verbirgt.

müssen wir uns aber zunächst einmal darauf einigen, wann wir zwei Funktionen als **gleich** ansehen.[75]

Definition 1.5.7 (Gleichheit von Funktionen)**.** Zwei Funktionen f und g heißen **gleich**, wenn die drei folgenden Aussagen wahr sind:

(I) Die Funktionen f und g haben denselben Definitionsbereich.

(II) Die Funktionen f und g haben denselben Wertebereich.

(III) Für jedes x aus dem Definitionsbereich von f und g gilt $f(x) = g(x)$.

Wir verwenden die Notation $f = g$ um auszudrücken, dass f und g gleich sind.

Nun kommen wir zum bereits angekündigten Assoziativegesetz:

Proposition 1.5.8 (Assoziativgesetz für die Hintereinanderausführung von Funktionen)**.** *Seien W, X, Y, Z Mengen und seien*

$$W \xrightarrow{f} X, \qquad X \xrightarrow{g} Y, \qquad Y \xrightarrow{h} Z$$

Funktionen. Dann gilt

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

Beweis. Laut Proposition 1.5.7 müssen wir folgende Aussagen zeigen:

Gleichheit der Definitionsbereiche und Gleichheit der Wertebereiche: Weil f von W nach X abbildet und $h \circ g$ von X nach Z, bildet $(h \circ g) \circ f$ von W nach Z ab.

Ebenso gilt: Weil $g \circ f$ von W nach Y abbildet und h von Y nach Z, bildet $h \circ (g \circ f)$ von W nach Z ab. Also haben die Funktionen $(h \circ g) \circ f$ und $h \circ (g \circ f)$ beide den Definitionsbereich W und den Wertebereich Z.

Gleichheit der Funktionswerte an allen Elementen des Definitonsbereichs: Wir müssen die Aussage

$$\forall w \in W : \qquad \big((h \circ g) \circ f\big)(w) = \big(h \circ (g \circ f)\big)(w)$$

[75] Denken Sie an dieser Stelle zurück an die Einführung in die Mengentheorie in Abschnitt 1.3: Dort sind wir auch nicht einfach davon ausgegangen, dass schon irgendwie klar sein wird, wann zwei Mengen gleich sind, sondern haben in Definition 1.3.9 exakt festgelegt, wann zwei Mengen gleich heißen.

Um dies zu zeigen, betrachten wir ein beliebiges, aber festes Element $w \in W$.[76] Für dieses Element w gilt zum einen

$$\big((h \circ g) \circ f\big)(w) \overset{\text{Def. 1.5.6}}{=} (h \circ g)\big(f(w)\big) \overset{\text{Def. 1.5.6}}{=} h\Big(g\big(f(w)\big)\Big),$$

und andererseits

$$\big(h \circ (g \circ f)\big)(w) \overset{\text{Def. 1.5.6}}{=} h\big((g \circ f)(w)\big) \overset{\text{Def. 1.5.6}}{=} h\Big(g\big(f(w)\big)\Big).$$

Also sind $\big((h \circ g) \circ f\big)(w)$ und $\big(h \circ (g \circ f)\big)(w)$ tatsächlich gleich. □

Injektivität, Surjektivität und Bijektivität

Im Umgang mit Funktionen gehören die folgenden drei Begriffe zum täglichen Handwerkszeug. Sie müssen deshalb schon jetzt lernen, den Umgang mit desen Begriffen zu beherrschen.

Definition 1.5.9 (Injektive, surjektive und bijektive Funktionen)**.** Seien X, Y Mengen und sei $f : X \to Y$.

(a) Die Funktion f heißt **injektiv**,[77] wenn für alle $x, \tilde{x} \in X$ gilt: Falls $x \neq \tilde{x}$ ist, dann ist auch $f(x) \neq f(\tilde{x})$.

(b) Die Funktion f heißt **surjektiv**, falls es für jedes $y \in Y$ ein $x \in X$ mit der Eigenschaft $f(x) = y$ gibt.

(c) Die Funktion f heißt **bijektiv**, falls sie sowohl injektiv als auch surjektiv ist.

Lassen Sie uns alle drei Teile der Definition noch eimal kurz und knapp und aussagenlogischer Notation darstellen: Für eine Funktion $f : X \to Y$ besagt obige Defintion:

(a) Es ist f genau dann injektiv, wenn folgendes gilt:

$$\forall x, \tilde{x} \in X : \quad x \neq \tilde{x} \quad \Rightarrow \quad f(x) \neq f(\tilde{x}).$$

[76] Wichtig: Hier sehen Sie (nicht zum ersten mal in dieser Vorlesung) eine extrem wichtige Beweistechnik, die Sie in Ihrem Studium ständig benötigen werden: Wenn man eine Aussage für alle Elemente einer Menge – in der aktuellen Situation heißt sie W – zeigen will, dann betrachtet man hierzu ein einzelnes Element w der Menge; dieses Element fixiert man gedanklich für die Dauer des Beweises (damit man sicher ist, während des kompletten Beweises immer über dasselbe Element zu sprechen), aber man bestimmt *nicht* näher, um welches Element es sich konkret handelt. Dies wird mit der Floskel „Sei $w \in W$ beliebig, aber fest“ zu Beginn des Beweises zum Ausdruck gebracht; das Wort „beliebig“ ist hierbei also im Sinne von „nicht näher bestimmt“ gemeint. Hat man die gewünschte Aussage dann am Ende für dieses Element w bewiesen, so kann man sicher sein, dass die Aussage für alle Elemente von W stimmt – denn weil w ein nicht näher bestimmtes Element von W war, funktioniert der Beweis, den man angegeben hat, für jedes Element von W.

[77] Manchmal sagt man anstelle von injektiv auch **ein-eindeutig**.

(b) Es ist f genau dann surjektiv, wenn folgendes gilt:

$$\forall y \in Y \; \exists x \in X : \quad f(x) = y.$$

(c) Es ist f genau dann bijektiv, wenn folgendes gilt:

$$\forall y \in Y \; \exists! x \in X : \quad f(x) = y.$$

Beispiele 1.5.10. (a) Die Abbildung f aus Beispiel 1.5.3(a) ist nicht injektiv, denn es gilt $f(1) = f(3)$.

Die Abbildung

$$\begin{aligned} g : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x^2 \end{aligned}$$

ist ebenfalls nicht injektiv, denn es ist z.B. $g(-1) = g(1)$.

(b) Die Abbildung

$$\begin{aligned} k : \mathbb{R}^2 &\to \mathbb{R}^3, \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\mapsto \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \\ 2x_1 \end{pmatrix}. \end{aligned}$$

aus Beispiel 1.5.3(b) ist injektiv.

Beweis. Seien $x, \tilde{x} \in \mathbb{R}^2$ beliebig, aber fest. Wir müssen die Implikation

$$x \neq \tilde{x} \quad \Rightarrow \quad k(x) \neq k(\tilde{x}) \tag{1.5.1}$$

zeigen. Hierzu verwenden wir eine aussagenlogisches Resultat, dass Sie in Aufgabe 2(a) auf Tutoriumsblatt 2 bewiesen haben: Wenn A und B Aussagen sind, dann ist die Implikation $A \Rightarrow B$ gleichbedeutend mit der Implikation $\neg B \Rightarrow \neg A$.

Anstatt die Implikation 1.5.1 zu beweisen, können wir also auch einfach die Implikation

$$k(x) = k(\tilde{x}) \quad \Rightarrow \quad x = \tilde{x}$$

zeigen, und genau dies tun wir nun.[78,79]

[78] Dies ist ein sogenannter Beweis per **Kontraposition** (ein Spezialfall des sogenannten Widerspruchsbeweises).

[79] Diese Vorgehensweise – also anzunehmen, dass $k(x) = k(\tilde{x})$ gilt, und daraus $x = \tilde{x}$ zu folgern, ist in vielen Fällen gut geeignet um Injektivität einer Funktion k zu beweisen.

Sei also $k(x) = k(\tilde{x})$. Aufgrund der Definition von k gilt dann

$$\begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \\ 2x_1 \end{pmatrix} = \begin{pmatrix} \tilde{x}_1 + \tilde{x}_2 \\ \tilde{x}_1 - \tilde{x}_2 \\ 2\tilde{x}_1 \end{pmatrix}.$$

Wegen der Definition der Gleichheit von Tupeln (Definition 1.3.16(c)) folgt hieraus

$$\begin{aligned} & x_1 + x_2 = \tilde{x}_1 + \tilde{x}_2 \\ \wedge \quad & x_1 - x_2 = \tilde{x}_1 - \tilde{x}_2 \\ \wedge \quad & 2x_1 = 2\tilde{x}_1. \end{aligned}$$

Aus der dritten dieser Gleichungen erhalten wir $x_1 = \tilde{x}_1$, und wenn wir dann noch die erste der drei Gleichungen verwenden, folgt zudem $x_2 = \tilde{x}_2$.

Dies bedeutet – erneut wegen der Definition der Gleichheit von Tupeln –, dass $x = \tilde{x}$ ist. □

(c) Die Abbildung f aus Beispiel 1.5.3(a) ist nicht surjektiv, denn es gibt z.B. kein $x \in \{1, 2, 3, 4\}$ mit der Eigenschaft $f(x) = 10$.

Die Abbildung

$$\begin{aligned} g : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x^2 \end{aligned}$$

ist ebenfalls nicht surjektiv, denn es gibt kein $x \in \mathbb{R}$ mit $g(x) = -1$.

(d) Die Abbildung

$$\begin{aligned} h : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto 2x + 1 \end{aligned}$$

ist bijektiv.

Beweis. Wir müssen Injektivität und Surjektivität zeigen.

Injektivität: Seien $x, \tilde{x} \in \mathbb{R}$ beliebig, aber fest. Wir führen den Beweis der Injektivität erneut per Kontraposition, d.h. wir zeigen die Implikation

$$h(x) = h(\tilde{x}) \quad \Rightarrow \quad x = \tilde{x}.$$

Sei also $h(x) = h(\tilde{x})$. Wegen der Definition von h gilt somit $2x + 1 = 2\tilde{x} + 1$. Indem wir auf beiden Seiten der Gleichung zuerst 1 subtrahieren und dann durch zwei teilen, folgt hieraus $x = \tilde{x}$.

Surjektivität: Wir müssen folgendes zeigen:

$$\forall y \in \mathbb{R}\ \exists x \in \mathbb{R}: \qquad h(x) = y.$$

Sei also $y \in \mathbb{R}$ beliebig, aber fest. Unsere Aufgabe ist es zu beweisen, dass ein $x \in \mathbb{R}$ existiert, welches die Gleichung $h(x) = y$ erfüllt. In der vorliegenden Situation ist dies sehr einfach, denn wir können ein solches x konkret angeben: Wir wählen $x = \frac{y-1}{2}$. Dann ist x tatsächlich ein Element von $\mathbb{R}$, und es gilt, wie gewünscht, $h(x) = 2x + 1 = y$. □

Lassen Sie uns nun zeigen, dass die Hintereinanderausführung zweier injektiver Funktionen wieder injektiv ist, und dass die Hintereinanderausführung zweier surjektiver Funktionen wieder surjektiv ist.

Proposition 1.5.11. *Seien X, Y, Z Mengen, und seien $f : X \to Y$ und $g : Y \to Z$.*

(a) *Wenn f und g beide injektiv sind, dann ist auch $g \circ f$ injektiv.*

(b) *Wenn f und g beide surjektiv sind, dann ist auch $g \circ f$ surjektiv.*

Beweis. (a) Seien f und g injektiv. Wir müssen zeigen, dass $g \circ f$ injektiv ist.

Seien dazu $x, \tilde{x} \in X$. Wie Sie nun bereits zweimal gesehen haben, genügt es die Implikation

$$(g \circ f)(x) = (g \circ f)(\tilde{x}) \quad \Rightarrow \quad x = \tilde{x}$$

zu zeigen. Also gelte nun $(g \circ f)(x) = (g \circ f)(\tilde{x})$. Laut Definition der Hintereinanderausführung von Funktionen ist dann

$$g\big(f(x)\big) = g\big(f(\tilde{x})\big).$$

Weil g injektiv ist, folgt hieraus $f(x) = f(\tilde{x})$. Und weil f injektiv ist, folgt hieraus wiederum $x = \tilde{x}$.

(b) Seien f und g surjektiv. Wir müssen die Aussage

$$\forall z \in Z\ \exists x \in X: \qquad (g \circ f)(x) = z$$

zeigen. Sei also $z \in Z$ beliebig, aber fest.

Weil g surjektiv ist, gibt es ein $y \in Y$ mit der Eigenschaft $g(y) = z$. Und weil auch f surjektiv ist, gibt es ein $x \in X$ mit der Eigenschaft $f(x) = y$. Für dieses x gilt somit

$$(g \circ f)(x) = g\big(f(x)\big) = g(y) = z.$$

Also haben wir gezeigt, dass es tatsächlich ein $x \in X$ mit der Eigenschaft $(g \circ f)(x) = z$ gibt. □

Bijektive Funktionen sind deshalb besonders nützlich, weil man Sie umkehren kann:

Definition 1.5.12 (Umkehrfunktion)**.** Seien X, Y Mengen und sei $f : X \to Y$ eine bijektive Abbildung.[80]

[80] D.h., es gibt für jedes $y \in Y$ genau ein $x \in X$ mit der Eigenschaft $f(x) = y$.

Wir definieren eine Funktion $f^{-1} : Y \to X$, die jedem $y \in Y$ genau dasjenige $x \in X$ zuordnet, für welches $f(x) = y$ gilt. Die Funktion f^{-1} heißt die **Umkehrfunktion** (oder **Umkehrabbildung**) von f.

Man beachte: Die Definition der Umkehrabbildung f^{-1} besagt gerade, dass für jedes $y \in Y$ die Gleichung

$$f\big(f^{-1}(y)\big) = y$$

gilt. Bevor wir einige Eigenschaften der Umkehrfunktion beweisen, benötigen wir die folgende Terminologie:

Definition 1.5.13 (Identische Funktion)**.** Sei X eine Menge. Die Abbildung

$$\begin{aligned} \mathrm{id}_X : X &\to X, \\ x &\mapsto x \end{aligned}$$

heißt die **identische Abbildung** oder die **Identität** auf X.

Wenn die Menge X aus dem Kontext klar ist, lassen wir das X im Index manchmal auch weg und schreiben einfach nur id anstelle von id_X.

Wenn wir die Identität mit einer anderen Funktion verknüpfen, dann erhalten wir diesselbe Funktion. Genauer: Wenn $f : X \to Y$ eine Funktion zwischen zwei Mengen X, Y ist, dann gilt

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

Dies kann man sich leicht mit Hilfe der Definition der Identität und der Definition der Gleichheit von Funktionen überlegen.[81]

Proposition 1.5.14 (Eigenschaften der Umkehrfunktion)**.** *Seien X, Y, Z Mengen, und seien $f : X \to Y$ und $g : Y \to Z$ zwei bijektive Funktionen.*

(a) *Für alle $x \in X$ und alle $y \in Y$ gilt*

$$f^{-1}\big(f(x)\big) = x \qquad \textit{und} \qquad f\big(f^{-1}(y)\big) = y;$$

d.h. etwas kürzer ausgedrückt:

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

(b) *Die Umkehrabbildung $f^{-1} : Y \to X$ ist ebenfalls bijektiv, und es gilt*

$$(f^{-1})^{-1} = f.$$

[81] Erinnern Sie sich daran, was schon in einer vorherigen Fußnote stand: Das Wort „leicht“ ist hier als Zielmarke zu verstehen, und die Formulierung „Das kann man sich leicht überlegen“ bedeutet nicht, dass Sie das ohne weitere Überlegung glauben dürfen, sondern Sie bedeutet, dass Sie sich das jetzt noch einmal selbst im Detail überlegen müssen um zu sehen, dass es wirklich direkt aus den Definitionen folgt.

(c) *Die Hintereinanderausführung $g \circ f$ ist ebenfalls bijektiv, und es gilt*

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

Beweis. (a) Die Eigenschaft

$$\forall y \in Y : \qquad f\big(f^{-1}(y)\big) = y$$

ist exakt die Eigenschaft, durch die wir die Umkehrfunktion definiert haben (siehe Definition 1.5.12 und die Erläuterung direkt nach der Definition).

Wir müssen noch die zweite Eigenschaft zeigen. Sei also $x \in X$ beliebig, aber fest. Wir setzen $y := f(x)$. Für dieses y gilt aufgrund der soeben besprochenen Eigenschaft die Gleichheit $f\big(f^{-1}(y)\big) = y$, also

$$f\Big(f^{-1}\big(f(x)\big)\Big) = f(x).$$

Weil f injektiv ist, folgt daraus die Gleichheit

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

Wenn wir die Definition der Gleichheit von Funktionen (Definition 1.5.7) verwenden, können wir aus den bisher gezeigten Eigenschaften sofort folgern, dass

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

gilt.[82]

(b) *Injektivität:* Seien $y, \tilde{y} \in Y$ und sei $f^{-1}(y) = f^{-1}(\tilde{y})$. Dann ist

$$f\big(f^{-1}(y)\big) = f\big(f^{-1}(\tilde{y})\big).$$

[82] **Achtung:** Wenn in einer Vorlesung in einem Beweis gesagt wird, dass etwas „sofort folgt“, dann bedeutet dies lediglich, dass man die behauptete Aussage ohne großen Aufwand aus dem folgern kann, was soeben gesagt wurde. Das bedeutet aber nicht automatisch, dass Ihnen dies auch wirklich klar ist. Wann immer Sie eine Aussage von der Form „nun folgt sofort“ (oder eine ähnliche Formulierung wie „jetzt folgt leicht“) lesen, müssen Sie also noch einmal nachprüfen, ob Sie diese Folgerung wirklich bis ins Detail begründen können. Dies tun Sie am besten auf einem Blatt Papier (oder einem Tablet).

Nehmen Sie sich also sogleich einen Stift und versuchen Sie extrem detailliert aufzuschreiben, weshalb die Gleichungen $f^{-1} \circ f = \mathrm{id}_X$ und $f \circ f^{-1} = \mathrm{id}_Y$ an dieser Stelle im Beweis wirklich folgen. Beachten Sie dabei: Sie können dies natürlich nur dann korrekt begründen, wenn Sie verwenden, wie die Gleichheit von Funktionen definiert ist. Ihre Begründung kann also nur funktionieren, wenn Sie Definition 1.5.7 zu Rate ziehen.

Übrigens: Es besteht kein Grund zur Frustration, falls Sie für das Ausarbeiten der Details länger brauchen, als Sie aufgrund des Wortes „sofort“ erwarten würden: Was eine Person mit einiger Erfahrung innerhalb von Sekunden sehen kann, kann jemand ohne Erfahrung manchmal erst nach einer viertel oder halben Stunde sehen. Fassen Sie den hier verwendeten Begriff „sofort“ deshalb als Zielmarke auf: Sie müssen sich solche „einfachen“ Details solange ausführlich und mit viel Zeitaufwand überlegen, bis Sie soviel Übung, Verständnis und Erfahrung gesammelt haben, dass Sie solche Details selbst auch sofort verstehen.

Wie bereits in (a) festgestellt, ist die linke Seite dieser Gleichung gleich y und die rechte Seite gleich $\tilde{y}$. Es folgt also $y = \tilde{y}$.

Surjektivität: Sei $x \in X$. Wir müssen zeigen, dass ein $y \in Y$ existiert, für welches $f^{-1}(y) = x$ gilt. Hierzu wählen wür einfach $y = f(x)$. Für dieses y gilt tatsächlich

$$f^{-1}(y) = f^{-1}\big(f(x)\big) = x,$$

wobei die letzte Gleichheit aus (a) folgt.

Gleichheit $(f^{-1})^{-1} = f$*:* Hier zu müssen wir laut Definition 1.5.7 zeigen, dass beiden Funktionen f und $(f^{-1})^{-1}$ denselben Definitions- und Wertebereich haben, und dass sie an alle Elementen des Wertebereichs denselben Werte annehmen.

Weil f von X nach Y abbildet, bildet f^{-1} laut Definition der Umkehrfunktion von Y nach X ab. Erneut aufgrund der Definition der Umkehrfunktion folgt hieraus, dass $(f^{-1})^{-1}$ von X nach Y abbildet. Also haben beide Funktionen f und $(f^{-1})^{-1}$ den Definitionsbereich X und den Wertebereich Y.

Nun müssen wir noch zeigen, dass für alle $x \in X$ die Gleichheit $f(x) = (f^{-1})^{-1}(x)$ gilt. Sei also $x \in X$ beliebig, aber fest. Der Übersicht halber ist es nützlich, im folgenden die Notation $h := f^{-1}$ zu verwenden. Somit ist h eine bijektive Abbildung von Y nach X, und unsere Aufgabe ist es, $f(x) = h^{-1}(x)$ zu zeigen.

Indem wir die rechtsstehende Formel aus Teil (a) anwenden – allerdings nicht auf die Funktion $f : X \to Y$, sondern auf die Funktion $h : Y \to X$ – erhalten wir die Formel

$$h\big(h^{-1}(x)\big) = x, \qquad \text{d.h.} \qquad f^{-1}\big(h^{-1}(x)\big) = x.$$

Jetzt wenden wir auf letztgenannte Gleichheit noch die Funktion f an, und erhalten somit

$$f\Big(f^{-1}\big(h^{-1}(x)\big)\Big) = f(x).$$

Die linke Seite dieser Formel ist aber – erneut wegen der rechtsstehenden Formel in Teil (a) (die wir aber dieses Mal auf die Funktion $f : X \to Y$, und auf das Element $y := h^{-1}(x) \in Y$ anwenden) – gleich $h^{-1}(x)$. Somit haben wir, wie gewünscht, $h^{-1}(x) = f(x)$ gezeigt.

(c) Diesen Beweis lagern wir in die Übungen aus. □

1.6 Ergänzungen

Kardinalität von Mengen

In diesem Abschnitt erfahren Sie etwas mehr über ein Thema, das kurz in den Übungen angeschnitten wurde: Die Anzahl der Elemente von Mengen. Siehe hierzu zum Beispiel auch [Beu14, Abschnitt 1.4] oder [Fis11, Abschnitt 1.1.4].

Zur Motivation der nachfolgenden Begriffsbildung ist es sinnvoll, sich folgende Beobachtung klar zu machen:

Bemerkung 1.6.1. Seien X, Y Mengen.

(a) Wenn X endlich ist und es eine bijektive Abbildung $f : X \to Y$ gibt, dann ist Y ebenfalls endlich und hat diesselbe Anzahl an Elementen wie X.

(b) Wenn X und Y beide endlich sind und gleich viele Elemente haben, dann gibt es eine bijektive Abbildung $f : X \to Y$.

Definition 1.6.2 (Mächtigkeit und Gleichmächtigkeit)**.** Seien X, Y Mengen.

(a) Wenn X endlich ist, dann bezeichnet man die Anzahl der Elemente von X als **Mächtigkeit** oder **Kardinalität** von X.[83]

Die Mächtigkeit einer endlichen Menge X ist also eine natürliche Zahl. Wir notieren[84] sie mit dem Symbol $\#X$ oder $\#(X)$.

(b) Die Menge X heißt **gleichmächtig** zu Y, wenn eine bijektive Abbildung $f : X \to Y$ existiert.

Beachten Sie, dass in Teil (b) auch unendliche Mengen erlaubt sind. Für den Fall, dass X endlich ist, folgt aus Bemerkung 1.6.1: Die Menge X ist gleichmächtig zu Y genau dann, wenn Y ebenfalls endlich ist und $\#X = \#Y$ gilt.

Wir beweisen einige Eigenschaften des Konzeptes „Gleichmächtigkeit“:

Proposition 1.6.3 (Eigenschaften von Gleichmächtigkeit)**.** *Seien X, Y, Z Mengen. Es gelten folgende Eigenschaften des Gleichmächtigkeits-Begriffs:*

(a) Reflexivität: *Die Menge X ist gleichmächtig zu sich selbst.*

(b) Symmetrie: *Wenn X gleichmächtig zu Y ist, dann ist Y auch gleichmächtig zu X.*

(c) Transitivität: *Wenn X gleichmächtig zu Y ist und Y gleichmächtig zu Z ist, dann ist X ebenfalls gleichmächtig zu Z.*

Beweis. (a) Die identische Abbildung $\mathrm{id}_X : X \to X$ ist bijektiv[85] , also existiert eine bijektive Abbildung $X \to X$.

(b) Sei X gleichmächtig zu Y. Dann gibt es eine bijektive Abbildung $f : X \to Y$. Laut Proposition 1.5.14(b) ist auch $f^{-1} : Y \to X$ bijektiv. Somit gibt es eine bijektive Abbildung $Y \to X$, d.h., Y ist gleichmächtig zu X.

(c) Sei X gleichmächtig zu Y und Y gleichmächtig zu Z. Dann gibt es eine bijektive Abbildung $f : X \to Y$ und eine bijektive Abbildung $g : Y \to Z$. Laut Proposition 1.5.14(c) ist die Hintereinanderausführung $g \circ f : X \to Z$ ebenfalls bijektiv. Somit gibt es eine bijektive Abbildung von X nach Z, d.h. X ist gleichmächtig zu Z. □

[83] Man kann auch für unendliche Mengen einen Kardinalitätsbegriff definieren. Dieser ist aber subtiler und abstrakter, weshalb wir uns an dieser Stelle der Vorlesung nicht weiter damit beschäftigen.

[84] Häufig wird anstelle des Symbols $\#X$ auch das Symbol $|X|$ für die Mächtigkeit verwendet.

[85] Warum eigentlich?

Nun kommen wir zu einer Sache, die viele Studierende am Anfang überrascht: Wir sehen uns an, zwischen welchen unendlichen Mengen es im bijektive Abbildungen gibt – d.h., welche unendlichen Mengen gleichmächtig sind (und welche nicht).

Definition 1.6.4 (Abzählbare und überabzählbare Mengen)**.** Sei X eine Menge.

(a) Die Menge X heißt **abzählbar unendlich**, falls sie gleichmächtig zu $\mathbb{N}$ ist.

(b) Die Menge X heißt **höchstens abzählbar**, falls sie endlich oder abzählbar unendlich ist.

(c) Die Menge X heißt **überabzählbar**, falls sie nicht höchstens abzählbar ist.[86]

Die folgende Proposition bestimmt für einige interessante Mengen, ob Sie abzählbar sind:

Proposition 1.6.5 (Einige abzählbare und überabzählbare Mengen)**.** (a) *Die Menge $\mathbb{N}^*$ ist abzählbar unendlich.*

(b) *Die Menge $\mathbb{Z}$ ist abzählbar.*

(c) *Menge $\mathbb{Q}$ ist abzählbar.*

(d) *Menge $\mathbb{Q}$ ist abzählbar.*

Beweis. (a) Die Abbildung

$$\begin{aligned} f : \mathbb{N}^* &\to \mathbb{N}, \\ n &\mapsto n - 1 \end{aligned}$$

is bijektiv.[87] Somit ist $\mathbb{N}^*$ laut Definition 1.6.2(b) gleichmächtig zu $\mathbb{N}$.

(b) Lassen Sie uns die Funktion

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{N}, \\ n &\mapsto \begin{cases} 2n & \text{falls } n \geq 0, \\ -2n - 1 & \text{falls } n < 0. \end{cases} \end{aligned}$$

Diese ist bijektiv.[88] Somit ist $\mathbb{Z}$ gleichmächtig zu $\mathbb{N}$.

(c) Für den Beweis verweisen wir zum Beispiel auf [Fis11, Satz 1 und Seite 88].

(d) Siehe zum Beispiel [Fis11, Satz 2 und Seite 89]. □

[86] Anders ausgerückt: Falls sie unendlich, aber nicht abzählbar unendlich ist.

[87] Warum?

[88] Warum?

Literaturhinweise

Grundlegendes

- Ein einführendes Kapitel über Aussagenlogik und Quantoren finden Sie zum Beispiel in [TT08, Abschnitt 1.1].
- Eine Einführung zu Mengenlehre und Funktionen bieten beispielsweise in [TT08, Abschnitte 1.2 und 5.2] sowie viele Lehrbücher über Lineare Algebra – zum Beispiel [Beu14, Abschnitte 1.1 und 1.3], [Fis11, Abschnitt 1.1], [FS20, Abschnitt 2.1] und [Jän08, Kapitel 1].

Weitere Literatur

- Eine Einführung in Logik und Mengenlehre, die speziell für Studentinnen und Studenten der Informatik zugeschnitten ist, bietet beispielsweise das Buch [Wit13].
- In den USA ist es an vielen Hochschulen üblich, dass sich Mathematik-Kurse in den ersten Semestern nicht auf Beweise konzentrieren, sondern eher auf Anschauung und rechnerische Verfahren – ähnlich, wie es es vielleicht aus der Schule kurz vor dem Abitur gewohnt sind.[89] Aus diesem Grund gibt es an US-amerikansichen Hochschulen häufig spezielle Kurse, die zu einem gewissen Zeitpunkt im Studium den Umgang mit Aussagenlogik und Beweisen lehren sollen. Deshalb kann man zahlreiche englischsprachige Lehrbücher finden, die sich einzig und allein auf mathematische Grundlagen und das Erlernen von Beweisen konzentrieren.

  Wenn Sie einmal versuchen möchten, ob solch ein Buch hilfreich für Sie ist, können Sie zum Beispiel einen Blick in [Blo11] werfen.

Fortgeschrittene Themen

- Logik und Mengentheorie kann man auch auf einem deutlichen präziseren, höheren und abstrakteren Niveau betreiben, als wir es in diesem einführenden Kapitel getan haben. Für die Lineare Algebra 1 würde das deutlich zu weit führen – aber falls Sie auf den Geschmack gekommen sind und mal hineinschnuppern wollen, können Sie zum Beispiel die folgenden Bücher konsultieren:

  Die Bücher [Rau08] und [Zie17] bieten eine Einführung in die formale Logik.

[89] Dies hängt natürlich mit dem Bildungssystem der USA zusammen, welches sich deutlich von dem in Deutschland unterscheidet.

Kapitel 2

Wichtige algebraische Strukturen

Einstiegsfragen. (a) Erinnern Sie sich, was das Assoziativgesetz für die Addition reeller Zahlen besagt?

Kennen Sie noch weitere Rechenoperationen, die ebenfalls assoziativ sind? Kennen Sie auch Rechenoperationen, die nicht assoziativ sind?

(b) Wenn Sie zu einer gegebenen reellen Zahl r die Zahl 5 addieren – wie können Sie das wieder rückgängig machen?

Wenn Sie eine gegebene reelle Zahl r mit π multiplizieren – wie können Sie das wieder rückgängig machen?

Wenn Sie ein Blatt Papier auf Ihrem Tisch um 35 Grad gegen den Uhrzeigersinn drehen – wie können Sie das wieder rückgängig machen?

Was haben die vorangehenden drei Fragen miteinander zu tun?

(c) Legen Sie die folgenden vier Münzen in einer Reihe mit aufsteigenden Werten (von links nach rechts) vor sich auf den Tisch: 10 Cent, 20 Cent, 50 Cent, 1 Euro. Vertauschen Sie fünf mal hintereinander jeweils zwei Münzen, wobei Sie die Münzen, die Sie vertauschen, in jedem Schritt selbst aussuchen dürfen.

Zeigen Sie nun die Reihe einer Kommilitonin oder einem Komilitonen K. Schafft K es ohne Ihre Hilfe zu rekonstruieren, welche fünf Vertauschungen Sie in den fünf Schritten durchgeführt haben? Macht es für die Antwort einen Unterschied, ob K auch die Reihenfolge angeben muss, in der die fünf Schritte durchgeführt wurden?

(d) Was halten Sie von der Aussage $1 + 1 = 0$?

2.1 Assoziative Verknüpfungen und Halbgruppen

Halbgruppen

In diesem Abschnitt werden wir über sogenannte **assoziative binäre Verknüpfungen** sprechen. Um dies effizient tun zu können, besprechen wir zunächst kurz

verschiedene Notationen für Funktionen, die wir bisher noch nicht explizit diskutiert haben.

Wenn X and Y Mengen sind und $f : X \to Y$ eine Funktion ist, so wissen Sie bereits, dass man den Funktionswert von f an einem Element $x \in X$ üblicherweise mit $f(x)$ bezeichnet. Es gibt aber auch noch andere mögliche Notationen für eine Funktion:

Bemerkungen 2.1.1 (Postfix- und Infix-Notation für Funktionen)**.** (a) Für jede natürliche Zahl $n \in \mathbb{N}^*$ nennt man das Produkt aller Zahlen von 1 bis n die **Fakultät** von n. Man bezeichnet sie überlicherweise mit dem Symbol $n!$. Außerdem erweist es sich als nützlich, auch die Fakultät von 0 zu definieren, und zwar als 1 (d.h., $0! = 1$ per Definition).

Wie Sie sehen, ist die Fakultät eine Abbildung, nämlich die Abbildung

$$\begin{aligned} \mathbb{N} &\to \mathbb{N}^*, \\ n &\mapsto n!, \end{aligned}$$

wobei $n!$ definiert ist wie oben beschrieben.

Wenn man zur Notation einer Abbildung ein Symbol benutzt und dieser aber *hinter* das Argument setzt, so bezeichnet man dies als **Postfix-Notation** für diese Funktion.

(b) Eine der wichtigsten Abbildungen, die Sie kennen, ist die Addition reeller Zahlen. Es handelt sich dabei um eine Abbildung von $\mathbb{R}^2$ nach $\mathbb{R}$ (denn die Abbildung nimmt sich ein Tupel aus zwei reellen Zahlen und weist diesem Tupel die Summe der beiden Zahlen zu).

Das Symbol für diese Abbildung, nämlich „+“, schreibt man aber üblicherweise *zwischen* den beiden Zahlen, die man addiert. Dies bezeichnet man als **Infix-Notation**.

Neben der Addition reeller Zahlen gibt es noch viele weitere Funktionen, die je zwei Elementen aus einer gegebenen Menge X ein weiteres Element aus X zuweisen. Für solche Abbildung verwendet man oft die folgende recht intuitive Terminologie:

Sei X eine Menge. Eine **binäre**[1] Verknüpfung auf X ist eine Abbildung von X^2 nach X. Genau wie die Addition reeller Zahlen notiert man auch andere binäre Verknüpfungen häufig[2] mit Hilfe der Infix-Notation.

Besonders interessant sind binäre Verknüpfungen, die **assoziativ** sind:

Definition 2.1.2 (Halbgruppe)**.** (a) Unter einer **Halbgruppe** versteht man ein Tupel $(X, \circ)$, bestehend aus einer Menge X und einer Abbildung[3] $\circ$, die folgende Eigenschaften erfüllt:

[1] Oder auch: **zweistellige**.

[2] Aber nicht immer.

[3] Die man üblicherweise mit Infix-Notation notiert, d.h. also, man schreibt für $a, b \in X$ lieber $a \circ b$ anstelle von $\circ(a, b)$.

(HG0) *Binäre Verknüpfung auf X:* Es gilt $\circ : X^2 \to X$.

(HG1) *Assoziativgesetz*:

$$\forall a, b, c \in X : \quad (a \circ b) \circ c = a \circ (b \circ c).$$

(b) Eine Halbgruppe $(X, \circ)$ heißt **kommutativ**[4], falls sie das sogenannte **Kommutativgesetz** erfüllt:

$$\forall a, b \in X : \quad a \circ b = b \circ a.$$

Einige Beispiele für Halbgruppen kann man sofort erhalten, wenn man die Addition und Multiplikation reeller Zahlen verwendet; es gibt aber noch viele weitere Beispiele für Halbgruppen. Dies ist gerade der Reiz bei dieser Begriffsbildung: Mit de axiomatischen Definition einer Halbgruppe erfasst man zahlreiche verschiedene mathematische Objekte auf einmal – und alles, was man für allgemeine Halbgruppen beweist, stimmt somit automatisch für jedes Objekt, das eine Halbgruppe ist.

Lassen Sie uns zunächst einige einfache Beispiele von Halbgruppen aufzählen:[5]

Beispiele 2.1.3. (a) Es sind $(\mathbb{N}, +)$ und $(\mathbb{N}, \cdot)$ Halbgruppen (wobei $+$ die übliche Addition und $\cdot$ die übliche Multiplikation bezeichnet).

(b) Ebenso sind $(\mathbb{N}^*, +)$ und $(\mathbb{N}^*, \cdot)$ Halbgruppen.

(c) Es ist $(\mathbb{N} \cup \{-1\}, +)$ keine Halbgruppe, denn $+$ ist keine binäre Verknüpfung auf der Menge $\mathbb{N} \cup \{-1\}$ (weil nämlich $(-1) + (-1) = -2 \notin \mathbb{N} \cup \{-1\}$ ist).

(d) Sei $\star : \mathbb{R}^2 \to \mathbb{R}$ durch $a \star b = a^2 b$ für alle $(a, b) \in \mathbb{R}^2$ gegeben. Dann ist $(\mathbb{R}, \star)$ keine Halbgruppe, weil die binäre Verknüpfung $\star$ nicht assoziativ ist.

(e) Es ist $\big(\{-1, 0, 1\}, \cdot\big)$ eine Halbgruppe (wobei $\cdot$ hier wieder die übliche Multiplikation bezeichnet).

Ein etwas interessanteres Beispiel ist das folgende

Beispiel 2.1.4. Sei X eine Menge und bezeichne $\mathrm{Abb}(X; X)$ die Menge aller Abbildungen von X nach X. Außerdem bezeichnen wir mit $\circ$ die Hintereinanderausführung von Abbildungen.[6] Dann ist $(\mathrm{Abb}(X; X), \circ)$ eine Halbgruppe.

[4] Oder auch: **Abelsch** – benannt nach dem norwegischen Mathematiker Niels Henrik Abel (1802 in Frindöe, Norwegen, – 1829 in Froland, Norwegen).

[5] In den Übungen werden Sie noch weitere kennenlernen.

[6] Beachten Sie, dass man hier unbedingt dazu sagen muss, was gemeint ist: In Definition 1.5.6 haben wir mit $\circ$ tatsächlich die Hintereinanderausführung von Funktionen bezeichnet. In Definition 2.1.2 haben wir mit $\circ$ hingegen eine allgemeine binäre Verknüpfung bezeichnet.
Sie erkennen hier bereits eine Sache, die Ihnen bereits im Laufe Ihres ersten Semesters noch öfter begegnen wird: Viele Symbole in der Mathematik sind mit verschiedenen Bedeutungen „überladen“ (*überladen* ist übrigens tatsächlich ein technischer Begriff aus der Informatik, um solch einen Sachverhalt zu beschreiben). Man muss dann aus dem Kontext erkennen, was gemeint ist. Falls die

Beweis. Lassen Sie uns überprüfen, dass $(\mathrm{Abb}(X;X),\circ)$ die beiden Axiome aus Definition 2.1.2(a) erfüllt:

- (HG0): Für zwei Funktionen $f,g \in \mathrm{Abb}(X;X)$ ist auch $f\circ g \in \mathrm{Abb}(X;X)$, also ist $\circ$ tatsächlich eine binäre Verknüpfung auf $\mathrm{Abb}(X;X)$.[7]
- (HG1): Die Assoziativität dieser Verknüpfung haben wir bereits in Proposition 1.5.8 bewiesen. □

Die Halbgruppe $(\mathrm{Abb}(X;X),\circ)$ ist im Allgemeinen nicht kommutativ. Um das zu erkennen, können Sie zum Beispiel die Menge $X=\{1,2\}$ betrachten und sich zwei Funktionen $f,g\in\mathrm{Abb}(X;X)$ überlegen, für die

$$f\circ g \neq g\circ f$$

gilt.[8]

Neutrale Elemente

In manchen Halbgruppen gibt es Elemente, die keinerlei Wirkung haben, wenn man Sie mit anderen Elemente verknüpft. Dieses Verhalten präzisieren wir in der folgenden Begriffsbildung:

Definition 2.1.5 (Neutrales Element)**.** Sei $(X,\circ)$ eine Halbgruppe. Ein Element $e\in X$ heißt **neutrales Element** der Halbgruppe, falls

$$\forall a\in X:\quad a\circ e = a = e\circ a$$

gilt.[9]

Weil es sehr viele Halbgruppen gibt (einige Beispiele haben Sie oben bereits gesehen), können Sie sich vielleicht vorstellen, dass man auch viele Halbgruppen finden kann, in denen es eine neutrales Element gibt. Und nun eine kleine Denksportaufgabe: Glauben Sie, man kann auch eine Halbgruppe finden, in der es zwei verschiedene neutrale Elemente gibt?

Die folgende Proposition liefert die Antwort auf diese Frage.

Chance einer Unklarheit besteht, muss man explizit dazusagen, was mit einem Symbol gemeint ist. Auf den ersten Blick mag das so wirken, als wäre es unnötig fehleranfällig und schwer zu durchschauen. Aber im Laufe des Semesters werden wir noch viele weitere binäre Verknüpfungen diskutieren, und dann werden Sie die notationelle Einfachheit, die dadurch entsteht, verschiedene Dinge mit demselben Symbol zu bezeichnen, schnell zu schätzen lernen. (Ganz generell ist es übrigens eine große Kunst, mathematische Notation genauso schlank zu halten, dass man sie effizient benutzen kann, dass es aber zugleich nicht ständig zu Missverständnissen kommt.)

[7] Machen Sie sich bitte unbedingt klar, dass hier etwas „Wildes“ passiert: Der Satz vor dieser Fußnote besagt, dass $\circ$ eine Abbildung von $\big(\mathrm{Abb}(X;X)\big)^2 \to \mathrm{Abb}(X;X)$ ist – d.h., wir betrachten hier gerade eine Abbildung, die Tupel aus Abbildungen auf Abbildungen abbildet.

[8] Und das sollten Sie auch tatsächlich tun!

[9] Eine Halbgruppe, in der ein neutrales Element existiert, wird auch als **Monoid** bezeichnet.

Proposition 2.1.6 (Eindeutigkeit des neutralen Elemente)**.** *Sei* $(X, \circ)$ *eine Halbgruppe. Dann gibt es höchstens ein neutrales Element in* X*.*

Beweis. Seien $e_1, e_2 \in X$ neutrale Elemente. Dann gilt

$$e_2 = e_1 \circ e_2 = e_1;$$

hierbei gilt die erste Gleichheit, weil e_1 ein neutrales Element ist, und die zweite Gleichheit gilt, weil e_2 ein neutrales Element ist.[10] □

Auch für den Begriff des neutralen Elements geben wir einige Beispiele an:

Beispiele 2.1.7. (a) Die Halbgruppe $(\mathbb{N}^*, +)$ besitzt kein neutrales Element. Die Halbgruppe $(\mathbb{N}^*, \cdot)$ hingegen besitzt ein neutrales Element, nämlich die Zahl 1.

(b) Die Halbgruppe $\big(\{-1, 0, 1\}, \cdot\big)$ besitzt ein neutrales Element, nämlich die Zahl 1.

(c) Sei X eine Menge. Die Halbgruppe $\big(\mathrm{Abb}(X; X), \circ\big)$ (wobei $\circ$ die Hintereinanderausführung bezeichnet) besitzt ein neutrales Element, nämlich id_X.

2.2 Gruppen

Inverse Elemente und Gruppen

Nun geht's zur Sache: Wir betrachten jetzt Halbgruppen, in denen es nicht nur ein neutrales Element gibt, sondern auch noch sogenannte **inverse Elemente**:

Definition 2.2.1 (Gruppe)**.** Eine **Gruppe** ist ein Tupel $(G, \circ)$, wobei G eine Menge und $\circ$ eine Abbildung ist, und folgende Eigenschaften erfüllt sind:

(G0) *Binäre Verknüpfung auf* G*:* Es gilt $\circ\colon G^2 \to G$.

(G1) Assoziativität: Die binäre Verknüpfung $\circ$ erfüllt das Assoziativgesetz.[11]

(G2) Existenz eines neutralen Elementes: Es gibt ein neutrales Element e in der Halbgruppe $(G, \circ)$.[12]

[10] Hier sehen Sie eine allgemeine Beweistechnik um Eindeutigkeit zu zeigen: Wenn man beweisen will, dass es von einem bestimmten Typ von Objekt nur eines (oder höchstens eines) gibt, dann nimmt man sich zwei solcher Objekte (die man mit verschiedenen Variablen bezeichnet, von denen man aber nicht voraussetzt, dass sie verschieden sein müssen). Daraufhin beweist man dann, dass diese beiden Objekte unter den gegebenen Voraussetzungen automatisch gleich sind. Damit hat man dann gezeigt, dass es in Wirklichkeit höchstens ein solches Objekt geben kann.

[11] D.h. anders ausgedrückt: $(G, \circ)$ ist eine Halbgruppe.

[12] Beachten Sie, dass das neutrale Element dann wegen Proposition 2.1.6 eindeutig bestimmt ist.

(G3) Existenz inverser Elemente: Es gilt

$$\forall a \in G \ \exists b \in G : \quad a \circ b = e.$$

Wenn $a \in G$ ist, dann nennt man ein Element $b \in G$ mit der Eigenschaft $a \circ b$ ein **rechtsinverses Element von a**.

Jede Gruppe ist natürlich auch eine Halbgruppe. Entsprechend ist der Begriff **kommutativ**, den wir für Halbgruppen in Definition 2.1.2(b) eingeführt haben, auch für Gruppen definiert.[13]

Bevor wir Beispiele diskutieren, beweisen wir zunächst einige allgemeine Aussagen über Gruppen:

Proposition 2.2.2 (Eigenschaften von rechtsinversen Elementen in Gruppen)**.** *Sei $(G, \circ)$ eine Gruppe, und bezeichne e ihr neutrales Element.*

(a) *Sei $a \in G$ und sei $b \in G$ ein rechtsinverses Element von a. Dann ist b auch **linksinvers** zu a, d.h. es gilt $b \circ a = e$.*

(b) *Jedes Element in G besitzt genau ein rechtsinverses Element in G, d.h., es gilt*

$$\forall a \in G \ \exists! b \in G : \quad a \circ b = e.$$

Beweis. (a) Sei $a \in G$ und sei $b \in G$ rechtsinvers zu a. Laut Axiom (G3) in der Definition einer Gruppe besitzt jedes Element der Gruppe ein rechtsinverses Element – also besitzt auch b eine rechtsinverses Element in G, nennen wir es c. Es gilt

$$a = a \circ e = a \circ (b \circ c) = (a \circ b) \circ c = e \circ c = c,$$

und somit

$$b \circ a = b \circ c = e.$$

Dies beweist, dass b tatsächlich linksinvers zu a ist.

(b) Sei $a \in G$ beliebig, aber fest. Laut Axiom (G3) in der Definition einer Gruppe besitzt a ein rechtsinverses Element $b \in G$, also müssen wir nur die Eindeutigkeit zeigen.

Sei also $\hat{b} \in G$ ebenfalls ein rechtsinverses Element von a. Es gilt

$$b = b \circ e = b \circ (a \circ \hat{b}) = (b \circ a) \circ \hat{b} = e \circ \hat{b} = \hat{b};$$

für die vorletzte Gleichheit haben wir die bereits bewiesene Aussage (a) benutzt. □

[13] D.h., eine Gruppe heißt **kommutativ**, wenn $a \circ b = b \circ a$ for alle $a, b \in G$ gilt.

Laut der vorangehenden Proposition hat jedes Element a einer Gruppe also nur ein rechtsinverses Element – deshalb ist es sinnvoll, vom *dem* rechtsinversen Element von a zu sprechen. Zudem ist das rechtsinverse Element von a immer auch automatisch linksinvers zu a – deshalb ist es sinnvoll, das rechtsinverse Element von a einfach das **inverse Element von a** zu nennen. Es ist üblich die folgende Notation für das inverse Element zu verwenden:

Notation 2.2.3. Sei $(G, \circ)$ eine Gruppe und sei $a \in G$. Das rechtsinverse Element von a (das laut Proposition 2.2.2 eindeutig bestimmt ist und zugleich auch linksinvers zu a ist) wird mit a^{-1} notiert.

In vielen Fällen wird die binäre Verknüpfung einer Gruppe nicht mit dem Symbol $\circ$ bezeichnet, sondern mit dem Symbol $+$.[14] Dies ist zum Beispiel bei der Gruppe $(\mathbb{R}, +)$ der Fall, wobei $+$ hier die übliche Addition auf den reellen Zahlen beschreibt.

Wenn die Gruppenverknüpfung als $+$ geschrieben wird, bezeichnet man das inverse Element eines Elementes a aus der Gruppe üblicherweise nicht mit a^{-1}, sondern mit $-a$.

Wir werden von nun an häufig von der Termonologie Gebrauch machen, die direkt vor der vorangehenden Notation eingeführt wurde und a^{-1} somit als inverses Element von a bezeichnen. Es folgen einige Eigenschaften inverser Elemente:

Proposition 2.2.4 (Eigenschaften inverser Elemente)**.** *Sei $(G, \circ)$ eine Gruppe, und bezeichne e das neutrale Element der Gruppe.*

(a) *Es gilt $e^{-1} = e$.*

(b) *Für jedes $a \in G$ gilt $(a^{-1})^{-1} = a$.*

(c) *Für alle $a_1, a_2 \in G$ gilt $(a_1 \circ a_2)^{-1} = a_2^{-1} \circ a_1^{-1}$.*

Beweis. (a) Die Behauptung besagt nichts weiter, als dass e das rechtsinverse Element von e ist. Und dass dies wirklich stimmt, folgt aus der Gleichung $e \circ e = e$ (welche wiederum wahr ist, weil e das neutrale Element der Gruppe ist).

(b) Laut Proposition 2.2.2(a) gilt $a^{-1} \circ a = e$. Wenn wir nun diese Gleichung von links mit $(a^{-1})^{-1}$ verknüpfen, dann folgt

$$(a^{-1})^{-1} \circ a^{-1} \circ a = (a^{-1})^{-1} \circ e,$$

und somit $e \circ a = (a^{-1})^{-1}$. Das Element auf der linken Seite dieser Gleichung ist gleich a, womit die Behauptung gezeigt ist.

(c) Weil $(a_1 \circ a_2)^{-1}$ per Definition dieser Notation rechtsinvers zu $a_1 \circ a_2$ ist, gilt

$$a_1 \circ a_2 \circ (a_1 \circ a_2)^{-1} = e.$$

[14] Allerdings wird ein Plus meist nur dann zur Notation einer Gruppenverknüpfung verwendet, wenn die Gruppe kommutativ ist.

Wir benutzen nun erneut, dass rechtsinverse Elemente laut Proposition 2.2.2(a) auch linksinvers sind: Durch Verknüpfen der zuletzt angeschriebenen Gleichung von links mit a_1^{-1} erhält man

$$a_2 \circ (a_1 \circ a_2)^{-1} = a_1^{-1}.$$

Diese Gleichung verknüpfen wir nun noch von links mit a_2^{-1} und erhalten somit

$$(a_1 \circ a_2)^{-1} = a_2^{-1} \circ a_1^{-1}.$$

Dies ist gerade die Behauptung. □

Nun diskutieren wir wie angekündigt einige Beispiele.

Beispiele 2.2.5. (a) Sowohl $(\mathbb{Z}, +)$ als auch $(\mathbb{R}, +)$ ist eine Gruppe, und die Zahl 0 ist jeweils das neutrale Element. Außerdem ist für jedes Element a von $\mathbb{Z}$ bzw. $\mathbb{R}$ die Zahl $-a$ das inverse Element.

(b) Es ist $(\mathbb{R}, \cdot)$ eine Halbgruppe mit neutralem Element 1, aber keine Gruppe – denn das Element 0 hat kein rechtsinverses Element (wäre nämlich eine Zahl $r \in \mathbb{R}$ rechtsinvers zu 0, so würde $1 = 0 \cdot r = 0$ gelten).

(c) Sei $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$.[15] Im Gegensatz zum vorangehenden Beispiel ist $(\mathbb{R}^*, \cdot)$ eine Gruppe.

In den Übungen werden Sie noch einige deutlich überraschendere Beispiele für Gruppen kennenlernen. Sehr interessante Beispiele für Gruppen sind zudem sogenannten die Permutationsgruppen – diese besprechen wir kurz am Ende dieses Abschnitts.

Untergruppen

Ein wichtiges Konzept in der Mathematik besteht darin, aus gegebenen Objekten neue Objekte mit ähnlichen Eigenschaften zu bauen. Eine erste Demonstration dieses Konzeptes folgt nun anhand des Begriffs der **Untergruppe**.

Definition 2.2.6 (Untergruppe)**.** Sei $(G, \circ)$ eine Gruppe, und bezeichne e das neutrale Element von G. Eine **Untergruppe** von $(G, \circ)$ ist eine Teilmenge $U \subseteq G$, die die folgenden Eigenschaften erfüllt:

(UG1) Es ist $e \in U$.

(UG2) Abgeschlossenheit bzgl. der Gruppenvernüpfung:

$$\forall u_1, u_2 \in U : \quad u_1 \circ u_2 \in U.$$

[15] Diese Notation ist sehr üblich, und wir werden sie in der kompletten Linearen Algebra 1 verwenden.

(UG3) Abgeschlossenheit bzgl. der Bildung von Inversen:

$$\forall u \in U : \quad u^{-1} \in U.$$

Proposition 2.2.7. *Sie $(G, \circ)$ eine Gruppe, und sei $U \subseteq G$ eine Untergruppe. Dann ist $(U, \circ)$ selbst eine Gruppe.*

Bevor wir die Proposition beweisen, ist es sinnvoll folgenden Punkt zu besprechen: In der Formulierung der Proposition waren wir genau genommen etwas unpräzise in der Notation. Weil $(G, \circ)$ eine Gruppe ist, wissen wir ja, dass $\circ$ eine Abbildung von G^2 nach G ist. Somit kann aber $(U, \circ)$ genau genommen gar keine Gruppe sein, denn damit dies überhaupt möglich ist, müsste ja $\circ$ eine Abbildung von U^2 nach U sein (und nicht von G^2 nach G).

Was hier vor sich geht, ist folgendes: Genau genommen meint man, wenn man von der Gruppe $(U, \circ)$ spricht, mit $\circ$ nicht exakt diesselbe Abbildung wie in $(G, \circ)$ – sondern man meint die sogenannte **Einschränkung** von $\circ$ auf U^2 (d.h., man wendet $\circ$ nun nicht mehr auf Elemente a und b aus G an, sondern nur noch auf Elemente auf U). Aus Axiom (UG2) in der Definition einer Untergruppe folgt, dass man somit eine Abbildung enthält, deren Werte alle in U liegen – also tatsächlich ein binäre Verknüpfung auf U.[16]

Beweis von Proposition 2.2.7. Wir müssen nachweisen, dass $(U, \circ)$ die Axiome aus der Definition einer Gruppe erfüllt:

- (G0): Dies folgt, wie bereits vor dem Beweis diskutiert, aus Untergruppen-Axiom (UG2).
- (G1): Weil $(G, \circ)$ eine Gruppe ist, gilt dass Assoziativgesetz

  $$(a \circ b) \circ c = a \circ (b \circ c)$$

  für alle $a, b, c \in G$. Also gilt es insbesondere für alle $a, b, c \in U$.
- (G2): Bezeichne e das neutrale Element von G. Laut Untergruppen-Axiom (UG1) ist $e \in U$, und offensichtlich ist e auch in $(U, \circ)$ neutrales Element.
- (G3): Sei $u \in U$. Das inverse Element u^{-1} von u in der Gruppe $(G, \circ)$ ist laut Untergruppen-Axiom (UG3) ein Element von U. Somit besitzt u ein inverses – und somit insbesondere rechts-inverses – Element in U. □

[16] Für Einschränkungen von Abbildungen gibt es eigentlich eine spezielle Notation, die Sie auch schon in den Übungen (in Aufgabe 1(b) auf Hausaufgabenblatt 3) kennengelernt haben. Wenn man dieser Notation folgt, muss man der Genauigkeit halber eigentlich sagen, „$(U, \circ|_{U \times U})$ ist eine Gruppe“, anstelle von „$(U, \circ)$ ist eine Gruppe“. Die Erfahrung zeigt aber, dass diese genauere Notation beim Behandeln von Untergruppen keinen Mehrwert bringt, und lediglich die Notation verkomplizieren würde. Deshalb ist man hier meist etwas ungenau und spricht stattdessen einfach von der Gruppe $(U, \circ)$. In anderen Kontexten (wenn es nicht gerade um Untergruppen oder ähnliche Strukturen geht) ist es aber wichtig, dass man es notational deutlich zum Ausdruck bringt, wenn eine Funktion auf eine Teilmenge ihres Definitionsbereichs eingeschränkt wird.

Aufgrund von Proposition 2.2.7 ist das Konzept der Untergruppen sehr nützlich, um aus einer Gruppe neue, „kleinere“ Gruppen zu erhalten.

Wir schließen diesen Abschnitt mit folgenden Beispielen:

Beispiele 2.2.8. (a) Das Intervall $(0, \infty) \coloneqq \{x \in \mathbb{R} \mid x > 0\}$ ist eine Untergruppe von $(\mathbb{R}^*, \cdot)$. Dies kann man leicht nachrechen, indem man die Untergruppenaxiome überprüft.

(b) Das Intervall $[1, \infty) \coloneqq \{x \in \mathbb{R} \mid x \geq 1\}$ ist keine Untergruppe von $(\mathbb{R}^*, \cdot)$, denn zum Beispiel ist das inverse Element von 2 – also die Zahl $\frac{1}{2}$ – kein Element von $[1, \infty)$.

Permutationen

Hier ist eine weitere Klasse von Gruppen:

Beispiel 2.2.9. Sei X eine Menge. Es bezeichne $\mathcal{S}(X)$ die Menge aller bijektiven Abbildungen von X nach X,[17] und bezeichne $\circ$ die Hintereinanderausführung von Funktionen aus $\mathcal{S}(X)$.

Dann ist $(\mathcal{S}(X), \circ)$ eine Gruppe, und für jedes $f \in \mathcal{S}(X)$ ist die Umkehrfunktion von f zugleich das inverse Elemente von f in der Gruppe $(\mathcal{S}(X), \circ)$.[18]

Man bezeichnet sie als **symmetrische Gruppe** auf X.

Beweis. Lassen Sie uns überprüfen, dass $(\mathcal{S}(X), \circ)$ die Axiome einer Gruppe erfüllt:

- (G0): Wenn wir nur Funktionen aus $\mathcal{S}(X)$ verknüpfen, dann bildet $\circ$ tatsächlich von $\mathcal{S}(X) \times \mathcal{S}(X)$ nach $\mathcal{S}(X)$ ab, denn für $f, g \in \mathcal{S}(X)$ ist $f \circ g$ ebenfalls eine Abbildung von X nach X, und laut Proposition 1.5.14(c) auch bijektiv.

- (G1): Die Hintereinanderausführung von Funktionen ist laut Proposition 1.5.8 assoziativ.

- (G2): Die Identität id_X ist ein Element von $\mathcal{S}(X)$ und es gilt

$$\mathrm{id}_X \circ f = f \circ \mathrm{id}_X = f$$

  für jedes $f \in \mathcal{S}(X)$. Somit ist id_X neutrales Element in $(\mathcal{S}(X), \circ)$.

[17] D.h., $\mathcal{S}(X)$ ist eine Teilmenge von $\mathrm{Abb}(X; X)$.

[18] Beachten Sie, dass wir die Notation „hoch -1“ in zwei verschiedenen Kontexten eingeführt haben: Laut Definition 1.5.12 wird sie für die Umkehrfunktion benutzt, und laut Notation 2.2.3 wird sie zur Bezeichnung von inversen Elementen in einer Gruppe benutzt. Wenn f ein Element der symmetrischen Gruppe $\mathcal{S}(X)$ ist, ist auf den ersten Blick nicht klar, welche der beiden Bedeutungen mit der Notation f^{-1} gemeint ist. Es ist deshalb wichtig, dass Sie sich an der Stelle überlegen, weshalb die beiden Notation in diesem Fall dasgleiche bedeuten (und somit an dieser Stelle kein Notationskonflikt auftritt).

- (G3): Sei $f \in \mathcal{S}(X)$. Dann besitzt f aufgrund seiner Bijektivität eine Umkehrfunktion f^{-1}. Diese bildet ebenfalls von X nach X ab, und ist laut Proposition 1.5.14(b) ebenfalls bijektiv. Außerdem ist sie laut Proposition 1.5.14(a) rechts-invers zu f (womit auch gleich die letzte Behauptung im Beispiel bewiesen ist). □

Das vorangehende Beispiel ist besonders wichtig, wenn X eine endliche Menge ist.

Beispiel 2.2.10. Sei $n \in \mathbb{N}^*$. Dann notiert man die symmetrische Gruppe auf der Menge $\{1, \dots, n\}$ – also die Gruppe $\mathcal{S}\big(\{1, \dots, n\}\big)$ – oft kurz als $\mathcal{S}_n$ und bezeichnet sie auch als **symmetrische Gruppe auf n Elementen.** Die Elemente von $\mathcal{S}_n$ nennt man **Permutationen auf n Elementen**.

Jedes Element von $\mathcal{S}_n$ – d.h. jede Permutation von n Elementen – ist also eine bijektive Abbildungen von $\{1, \dots, n\}$ nach $\{1, \dots, n\}$. Wie Sie bereits aus Beispiel 1.5.3(a) wissen, kann man Funktionen, deren Definitionsbereich endlich ist, in Tabellenform angeben. Bei Permutationen ist dies sehr üblich und äußerst nützlich.

Hierbei sind die folgenden Konventionen üblich: Man schreibt die komplette Tabelle innerhalb von geschweiften Klammern auf, und man lässt oft sämtliche Trennstriche zwischen Zeilen und Spalten der Tabelle weg. Außerdem ist man meist sogar so dreist, die Tabelle selbst einfach als die entsprechende Funktion aufzufassen. Somit ist also die Permutation

$$\sigma \coloneqq \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$$

zum Beispiel ein Element aus $\mathcal{S}_4$. Für dieses σ gilt

$$\sigma(1) = 3, \qquad \sigma(2) = 1, \qquad \sigma(3) = 2, \qquad \sigma(4) = 4.$$

Das neutrale Element von $\mathcal{S}_3$ – also die Funktion $\mathrm{id}_{\{1,2,3\}}$ – lässt sich in dieser Notation schreiben als

$$\mathrm{id}_{\{1,2,3\}} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}.$$

Im folgenden und letzten Beispiel dieses Abschnitts zeigen wir, wie man die Hintereinanderausführung von Permutationen konkret berechnen kann:

Beispiel 2.2.11. Betrachten Sie die folgenden drei Permutationen aus $\mathcal{S}_4$:

$$\sigma_1 \coloneqq \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix}, \qquad \sigma_2 \coloneqq \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, \qquad \sigma_3 \coloneqq \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}.$$

Es gilt

$$\sigma_2 \circ \sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix},$$

und

$$\sigma_3 \circ \sigma_2 \circ \sigma_1 = \sigma_3 \circ (\sigma_2 \circ \sigma_1) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}.$$

Durch Permutationen lässt sich sehr gut die Umordnung von Objekten beschreiben, die in einer bestimmten Reihenfolge aufgestellt sind. Dies demonstrieren wir anschaulich in der Vorlesung (mit ganz konkreten Gegenständen, sodass sich die Demonstration nur schwer schriftlich darstellen lässt – deshalb wird sie hier im Manuskript nicht beschrieben).

2.3 Körper

Was ist ein Körper?

Wie Sie bereits wissen, ist eine Gruppe eine Menge zusammen mit einer binären Verknüpfung, welche bestimmte Eigenschaften erfüllt. Wenn Sie nun aber z.B. die reellen Zahlen betrachten, dann kennen Sie auf dieser Menge ja zwei Verknüpfungen mit interessanten Eigenschaften: Die Addition und die Multiplikation. Zudem hängen beide Verknüpfungen auch noch mittels des Distributivgesetzes zusammen.

Es gibt aber noch weitere Mengen, auf denen binäre Verknüpfungen mit ähnlichen Eigenschaften definiert sind. Um all diese zugleich behandeln zu können, führt man den folgenden Begriff ein:

Definition 2.3.1 (Körper)**.** Ein Körper ist ein Tupel $(\mathbb{K}, +, \cdot)$, wobei $\mathbb{K}$ eine Menge ist, und $+$ und $\cdot$ Abbildungen sind[19], welche die folgenden Eigenschaften erfüllen:[20]

(K0) *Binäre Verknüpfungen auf* $\mathbb{K}$*:* Es gilt $+ : \mathbb{K}^2 \to \mathbb{K}$ und $\cdot : \mathbb{K}^2 \to \mathbb{K}$.

(K1) *Axiome der Addition:*
Es ist $(\mathbb{K}, +)$ eine kommutative Gruppe.

Das neutrale Element dieser Gruppe bezeichnet man mit 0. Außerdem verwendet man für jedes $\alpha \in \mathbb{K}$ die Notation $-\alpha$ um das inverse Elemente von α in der Gruppe $(\mathbb{K}, +)$ zu bezeichnen.

(K2) *Axiome der Multiplikation, Teil 1:* Es ist $(\mathbb{K}, \cdot)$ eine kommutative Halbgruppe, die ein neutrales Element besitzt.[21] Das neutrale Element bezeichnen wir mit 1.

(K3) *Nicht-Trivialität des multiplikativ neutralen Elements:* Es ist $1 \neq 0$.

[19] Die wir mit Infix-Notation verwenden werden.

[20] Die Verknüpfung $+$ bezeichnet man üblicherweise als **Addition** und die Verknüpfung $\cdot$ als **Multiplikation**.

[21] Beachten Sie, dass das neutrale Element laut Proposition 2.1.6 automatisch eindeutig bestimmt ist.

(K4) *Axiome der Multiplikation, Teil 2:* Für jedes $\alpha \in \mathbb{K} \setminus \{0\}$ existiert ein $\beta \in \mathbb{K}$ mit der Eigenschaft $\alpha \cdot \beta = 1$.
Wir verwenden die Notation $\mathbb{K}^* \coloneqq \mathbb{K} \setminus \{0\}$.

(K5) *Distributivgesetz:* Für alle $\alpha, \beta, \lambda \in \mathbb{K}$ gilt: $(\alpha + \beta) \cdot \lambda = \alpha \cdot \lambda + \beta \cdot \lambda$.

Aus den Körperaxiomen folgen einige einfache, aber nützliche Konsequenzen, die im folgenden aufgezählt werden:

Proposition 2.3.2 (Rechenregeln in Körpern)**.** *Sei* $(\mathbb{K}, +, \cdot)$ *ein Körper.*

(a) *Für alle* $\alpha, \beta \in \mathbb{K}$ *gilt:*

$$\alpha \cdot \beta = 0 \qquad \Leftrightarrow \qquad \alpha = 0 \quad \vee \quad \beta = 0.$$

(b) *Für alle* $\alpha, \beta \in \mathbb{K}$ *gilt*

$$(-1) \cdot \beta = -\beta, \quad (-\alpha) \cdot \beta = -(\alpha \cdot \beta), \quad \textit{und} \quad (-\alpha) \cdot (-\beta) = \alpha \cdot \beta.$$

(c) *Für alle* $\alpha, \beta \in \mathbb{K}$ *gilt*

$$-(\alpha + \beta) = (-\alpha) + (-\beta).$$

(d) *Für jedes* $\alpha \in \mathbb{K}^*$ *gibt es genau ein* $\beta \in \mathbb{K}$ *mit der Eigenschaft* $\alpha \cdot \beta = 1$. *Außerdem ist dieses* β *nicht* 0*, d.h. ein Element von* $\mathbb{K}^*$.

(e) *Es ist* $(\mathbb{K}^*, \cdot)$ *eine kommutative Gruppe.*

Beweis. (a) Seien $\alpha, \beta \in \mathbb{K}$.
„$\Leftarrow$“ Wir nehmen zuerst an, dass $\alpha = 0$ gilt, und wir müssen $0 \cdot \beta = 0$ zeigen.
Hierzu verwenden wir das Distributivgesetz: Es gilt

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

Indem wir dass additiv inverse Elemente von $0 \cdot \beta$ auf beiden Seiten der Gleichung addieren, erhalten wir hieraus wie behauptet $0 = 0 \cdot \beta$.

Wenn stattdessen $\beta = 0$ ist, dann folgt mithilfe der Kommutativität von $\cdot$ und der bereits gezeigten Eigenschaft, dass $\alpha \cdot 0 = 0 \cdot \alpha = 0$ gilt.

„$\Rightarrow$“ Sei nun $\alpha \cdot \beta = 0$. Wir nehmen widerspruchshalber an, dass $\alpha \neq 0$ und $\beta \neq 0$ gilt. Dann gibt es laut Körperaxiomen Elemente $\tilde{\alpha}, \tilde{\beta} \in \mathbb{K}$ mit der Eigenschaft $\alpha\tilde{\alpha} = 1$ und $\beta\tilde{\beta} = 1$. Hieraus folgt

$$0 = \tilde{\alpha} \cdot 0 \cdot \tilde{\beta} = \tilde{\alpha} \cdot \alpha \cdot \beta \cdot \tilde{\beta} = 1 \cdot 1 = 1,$$

wobei wir für die erste Gleichheit die bereits gezeigte Implikation verwenden haben. Die Gleichheit $0 = 1$ ist aber ein Widerspruch zu den Körperaxiomen.

(b) Sei $\beta \in \mathbb{K}$. Wir zeigen zunächst, dass $(-1) \cdot \beta = -\beta$ gilt:

Mithilfe des Distributivgesetzes erhalten wir

$$\beta + (-1) \cdot \beta = 1 \cdot \beta + (-1) \cdot \beta = \big(1 + (-1)\big) \cdot \beta = 0 \cdot \beta = 0,$$

wobei wir für die letzte Gleichheit Aussage (a) benutzt haben. Indem wir nun auf beiden Seiten $-\beta$ addieren, folgt wie behauptet $(-1) \cdot \beta = -\beta$.

Die zweite Gleichheit in (b) folgt durch Umklammern aus der ersten.

Um die dritte Gleichheit zu zeigen, beobachten wir zuerst, dass aus der ersten Gleichheit die Eigenschaft

$$(-1) \cdot (-1) = -(-1) = 1$$

folgt. Für $\alpha, \beta \in \mathbb{K}$ folgt hieraus durch geeignetes Umklammern sofort die behauptete Gleichheit $(-\alpha) \cdot (-\beta) = \alpha \cdot \beta$

(c) Dies folgt aus der ersten Gleichheit in (b) und dem Distributivgesetz.

(d) Sei $\alpha \in \mathbb{K}^*$. Wir wissen bereits aus den Körperaxiomen, dass es ein $\beta \in \mathbb{K}$ mit der Eigenschaft $\alpha \cdot \beta = 1$ gibt. Wir müssen aber die Eindeutigkeit beweisen.

Seien also $\beta_1, \beta_2 \in \mathbb{K}$ derart, dass $\alpha \cdot \beta_1 = 1$ und $\alpha \cdot \beta_2 = 1$ gilt. Dann folgt

$$0 = \alpha \cdot \beta_1 + \big(- (\alpha \cdot \beta_2)\big) = \alpha \cdot \beta_1 + \alpha \cdot (-\beta_2) = \alpha \cdot \big(\beta_1 + (-\beta_2)\big).$$

Weil α nach Voraussetzung nicht 0 ist, folgt aus (a), dass $\beta_1 + (-\beta_2) = 0$ gilt. Durch Addition von β_2 auf beiden Seiten dieser Gleichung erhalten wir $\beta_1 = \beta_2$. Damit ist die Eindeutigkeit beweisen.

Dass das Element $\beta \in \mathbb{K}$, welches $\alpha \cdot \beta = 1$ erfüllt, nicht 0 sein kann (und somit in $\mathbb{K}^*$ liegt), folgt wegen $1 \neq 0$ aus (a).

(e) Dies folgt unmittelbar aus den Körperaxiomen und Aussage (d). □

In Körpern sind einige notationelle Konventionen üblich, die im folgenden aufgezählt werden:

Notation 2.3.3 (Übliche Notationen in Körpern)**.** Sei $(\mathbb{K}, +, \cdot)$ ein Körper

(a) Für zwei Elemente α, β verwendet man meist die Abkürzung $\alpha - \beta \coloneqq \alpha + (-\beta)$.

(b) Für jedes $\alpha \in \mathbb{K}^*$ notiert man das inverse Element von α in der Gruppe $(\mathbb{K}^*, \cdot)$ (d.h. das eindeutig bestimmte Element $\beta \in \mathbb{K}$ mit der Eigenschaft $\alpha \cdot \beta = 1$), wie in Gruppen üblich, als α^{-1}.

(c) Für alle $\alpha \in \mathbb{K}^*$ und alle $\beta \in \mathbb{K}$ folgt aus den Körperaxiomen und den bereits gezeigten Eigenschaften, dass es genau ein Element $x \in \mathbb{K}$ mit der Eigenschaft $\alpha \cdot x = \beta$ gibt. Dieses Element x bezeichent man üblicherweise mit der Notation $\frac{\beta}{\alpha}$.[22]

Zugleich kann man sofort nachrechen, dass $x = \alpha^{-1} \cdot \beta$ ist. D.h. $\frac{\beta}{\alpha}$ ist eine Kurzschreibweise für $\alpha^{-1} \cdot \beta$. Insbesondere ist somit $\frac{1}{\alpha} = \alpha^{-1}$.

[22] Manchmal schreibt man auch β/α anstelle von $\frac{\beta}{\alpha}$. Äußerst unüblich ist hingegen die Verwendung eines Doppelpunktes zur Notation einer Division.

Wir sprechen kurz einige klassische Beispiele an:

Beispiele 2.3.4. (a) Es ist $(\mathbb{R}, +, \cdot)$ ein Körper (wobei $+$ und $\cdot$ die übliche Addition und Multiplikation von reellen Zahlen bezeichen).

(b) Es ist auch $(\mathbb{Q}, +, \cdot)$ ein Körper (wobei $+$ und $\cdot$ ebenfalls die übliche Addition und Multiplikation bezeichnen).

(c) Es ist $(\mathbb{Z}, +, \cdot)$ (ebenfalls mit üblicher Addition und Multiplikation) *kein* Körper, denn beispielsweise gibt es für das Element $\alpha := 2 \in \mathbb{Z}$ keine $\beta \in \mathbb{Z}$ mit der Eigenschaft $\alpha \cdot \beta = 1$.

Körper sind also Verallgemeinerungen der Ihnen bereits bekannten Strukturen $\mathbb{R}$ bzw. $\mathbb{Q}$. Die Körperaxiome zusammen mit Proposition 2.3.2 und Notation 2.3.2 zeigen, dass Sie in allgemeinen Körpern im Grunde genauso rechnen dürfen, wie Sie es mit rationalen oder reellen Zahlen gewohnt sind.

Es gibt allerdings einen entscheidenden Unterschied: Zwei reelle Zahlen kann man immer der Größe nach vergleichen (und selbiges gilt für zwei Zahlen aus $\mathbb{Q}$). Dies ist für Elemente beliebiger Körper hingegen nicht richtig: Beachten Sie, dass wir in den Körperaxiomen nirgends gefordert haben, dass es eine Möglichkeit gibt, für zwei Elemente eines Körpers zu bestimmen, welches „größer“ ist. Wir haben noch nicht einmal definiert, was „größer“ in einem beliebigen Körper überhaupt heißen soll – und man kann sogar zeigen, dass es in allgemeinen Körpern gar nicht möglich ist, einen widerspruchsfreien „größer“-Begriff einzuführen, der unserem üblichen Verständnis dieses Wortes genügt.

Den Rest von Abschnitt 2.3 wollen wir nutzen, um zwei weitere wichtige Klassen von Körpern einzuführen.

Komplexe Zahlen

Der ursprüngliche Sinn sogenannter **komplexer Zahlen** bestand darin, für bestimmte Gleichungen, die in $\mathbb{R}$ keine Lösung besitzen,[23] dennoch Lösungen zu definieren, die sich in einem geeigneten Sinne „vernünftig“ verhalten. Aus heutiger Sicht geht der Nutzen komplexer Zahlen jedoch weit hierüber hinaus.

Wir beginnen zunächst damit, die Menge $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ durch geeignet definierte Operationen zu einem Körper zu machen:

Definition 2.3.5 (Der Körper der komplexen Zahlen)**.** Seien $+ : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ und $\cdot : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ definiert durch

$$\alpha + \beta = \begin{pmatrix} \alpha_1 + \beta_1 \\ \alpha_2 + \beta_2 \end{pmatrix}$$

und

$$\alpha \cdot \beta = \begin{pmatrix} \alpha_1\beta_1 - \alpha_2\beta_2 \\ \alpha_1\beta_2 + \alpha_2\beta_1 \end{pmatrix}$$

[23] Wie zum Beispiel $x^2 = -1$.

für alle $\alpha, \beta \in \mathbb{R}^2$.

Das Tupel $(\mathbb{R}^2, +, \cdot)$ nennt man den **Körper der komplexen Zahlen**. Wenn man $\mathbb{R}^2$ mit diesen beiden Verknüpfungen ausstattet, ist es üblicher, anstelle von $\mathbb{R}^2$ die Notation $\mathbb{C}$ zu verwenden. Somit notiert man den Körper der komplexen Zahlen dann mit $(\mathbb{C}, +, \cdot)$.

Man kann zeigen:

Proposition 2.3.6. *Der Körper der komplexen Zahlen ist tatsächlich ein Körper. Die neutralen Elemente der Addition und Multiplikation sind*

$$\begin{pmatrix}0\\0\end{pmatrix} \qquad \textit{und} \qquad \begin{pmatrix}1\\0\end{pmatrix}.$$

Beweis. Das Überprüfen der meisten Körperaxiome ist Routine. Interessant ist in erster Linie die Existenz von multiplikativ inverse Elementen – dies stellen wir als Übungsaufgabe. □

Der in Definition 2.3.1 eingeführten Notation für neutrale Element in Körpern folgend schreiben wir für das additiv bzw. multiplikativ neutrale Element

$$\begin{pmatrix}0\\0\end{pmatrix} \qquad \text{bzw.} \qquad \begin{pmatrix}1\\0\end{pmatrix}$$

kurz 0 bzw. 1.

Mit komplexen Zahlen in der obigen Darstellung zu rechnen, ist recht ungemütlich. Um zu einer besser handhabbaren Darstellung zu gelangen, ist die folgende Überlegung hilfreich:

Diskussion 2.3.7. (a) Wenn Sie $\mathbb{R}$ als Zahlenstrahl zeichnen, und $\mathbb{R}^2$ als Ebene, dann ist es naheliegend, den Zahlenstrahl als Rechtswertachse des Koordinatensystems von $\mathbb{R}^2$ einzuzeichnen. In diesem Sinne kann man $\mathbb{R}$ geometrisch als eine Teilmenge von $\mathbb{R}^2$ auffassen. Rechnerisch bedeutet dies, dass man jedes $r \in \mathbb{R}$ mit dem Tupel

$$\begin{pmatrix}r\\0\end{pmatrix}$$

identifiziert.[24] Wenn man dies tut, dann stellt man fest, dass die Rechenoperationen auf $\mathbb{R}^2 = \mathbb{C}$, die in Definition 2.3.5 eingeführt wurden, genau der

[24] Beachten Sie, dass r und das Tupel nicht wirklich gleich sind in dem Sinne, wie wir die Gleichheit zweier Tupel definiert haben. Wenn man das, was hier steht, völlig präzise aufschreiben möchte, muss man eine injektive Abbildung $\mathbb{R} \to \mathbb{R}^2$ betrachten, die jedem $r \in \mathbb{R}$ das entsprechende Tupel in $\mathbb{R}^2$ zuweist, und dann über sogenannte **Einbettungen** und **Isomorphismen** von Körpern sprechen – diesen terminologischen und konzeptuellen Aufwand wollen wir an dieser Stelle aber vermeiden, denn er bringt uns hier nicht wirklich weiter.
Deshalb sind wir lieber ein wenig ungenau und fassen, wann immer wir von den komplexen Zahlen sprechen – aber wirklich nur dann – $\mathbb{R}$ ab sofort als Teilmenge des $\mathbb{R}^2$ auf.

üblichen Addition und Multiplikation reeller Zahlen entsprechen, wenn man Sie auf $\mathbb{R}$ einschränkt: Für alle $r, s \in \mathbb{R}$ gilt nämlich

$$\begin{pmatrix} r \\ 0 \end{pmatrix} + \begin{pmatrix} s \\ 0 \end{pmatrix} = \begin{pmatrix} r+s \\ 0 \end{pmatrix}$$

und

$$\begin{pmatrix} r \\ 0 \end{pmatrix} \begin{pmatrix} s \\ 0 \end{pmatrix} = \begin{pmatrix} rs \\ 0 \end{pmatrix};$$

dies folgt direkt aus der Definition von $+$ und $\cdot$ auf $\mathbb{C} = \mathbb{R}^2$.

Man sagt deshalb auch, dass $\mathbb{R}$ ein **Teilkörper** von $\mathbb{C}$ ist.

(b) Nun führt man noch die folgende Notation und Terminologie ein: Man nennt das Element

$$\mathrm{i} \coloneqq \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \mathbb{C}$$

die **imaginäre Einheit**. Sie erfüllt die Gleichung $\mathrm{i}^2 = -1$ (das werden wir in den Übungen zeigen).

Mit Hilfe der imaginären Einheit kann man komplexe Zahlen in viel übersichtlicherer Form schreiben. Wenn wir nämlich wie oben erläutert jede reelle Zahl als ein Element von $\mathbb{R}$ auffassen, dann folgt aus der Definition der Multiplikation auf $\mathbb{C}$ für alle $\alpha \in \mathbb{C}$ die Formel

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} \alpha_2 \\ 0 \end{pmatrix} = \alpha_1 + \mathrm{i}\alpha_2.$$

Die reelle Zahl α_1 heißt der **Realteil** von α, und die reelle Zahl α_2 heißt der **Imaginärteil** von α; wir kürzen Sie mit der Notation

$$\operatorname{Re}\alpha \coloneqq \alpha_1 \qquad \text{und} \qquad \operatorname{Im}\alpha \coloneqq \alpha_2$$

ab.

Wir können – und werden – also ab sofort jede komplexe Zahl α in der Form

$$\alpha = \operatorname{Re}\alpha + \mathrm{i}\operatorname{Im}\alpha$$

schreiben. Indem man die üblichen Rechenregeln in Körpern verwendet und zudem von der Rechenregel $\mathrm{i}^2 = -1$ Gebrauch macht, kann man mit dieser Darstellung komplexer Zahlen bereits recht übersichtlich rechnen.

Außerdem ist es somit noch etwas intuitiver, die reellen Zahlen als Teilmenge der komplexen Zahlen aufzufassen, denn die reellen Zahlen sind somit einfach diejenigen komplexen Zahl, deren Imaginärteil gleich 0 ist. Oder anders ausgedrückt: Jede reelle Zahl r lässt sich in der Form

$$r = r + \mathrm{i}0$$

als komplexe Zahl auffassen.

Endliche Körper

Nun werden wir noch eine weitere Klasse von Körpern diskutieren – und zwar welche, die nur endlich viele Elemente besitzen. Als Vorbereitung sollten Sie sich an eine Rechenoperation erinnern, die Sie vermutlich schon in der Grundschule gelernt haben:

Bemerkung 2.3.8 (Teilen mit Rest)**.** Will man bei der Division natürlicher Zahlen ein ganzzahliges Ergebnis erhalten, so bleibt häufig ein Rest. Zum Beispiel:

(a) Die ganzzahlige Division von 10 durch 7 liefert 1, mit Rest 3.

(b) Die ganzzahlige Division von 7 durch 8 liefert 0, mit Rest 7.

(c) Die ganzzahlige Division von 5 durch 5 liefert 1, mit Rest 0.

(d) Die ganzzahlige Division von 20 durch 6 liefert 3, mit Rest 2.

Teilen mit Rest benötigen wir um die Rechenoperationen in den folgenden Körpern zu beschreiben:

Beispiel 2.3.9 (Körper mit einer primen Anzahl an Elementen)**.** Sei $p \in \mathbb{N}^*$ eine Primzahl.[25] Wir möchten gerne auf der Menge $\{0, \dots, p-1\}$ zwei Rechenoperationen definieren um einen Körper zu erhalten.

Bevor wir diese Rechenoperationen definieren, besprechen wir aber noch eine notationelle Besonderheit: Anstelle von $0, \dots, p-1$ wollen wir lieber die Symbole $[0], \dots, [p-1]$ für die Elemente dieser Menge benutzen. Für den Moment können Sie sich vorstellen, dass diese trotzdem einfach die natürlichen Zahlen $0, \dots, p-1$ sind, und dass wir mit den eckigen Klammern einfach zum Ausdruck bringen möchten, dass die Rechenoperationen auf diesen Zahlen nicht diejenigen sind, mit denen man üblicherweise auf den natürlichen Zahlen rechnet, sondern diejenigen, die wir nun besprechen werden.[26]

Wir bezeichnen die Menge $\{[0], \dots, [p-1]\}$ kurz mit $\mathbb{F}_p$, und wir führen die folgenden binären Verknüpfungen $+$ und $\cdot$ auf $\mathbb{F}_p$ ein: Für alle $[a], [b] \in \mathbb{F}_p$ setzen wir

$$[a] + [b] := [r],$$

wobei r diejenige natürliche Zahl von 0 bis $p-1$ ist, die man bei der ganzzahligen Division der natürlichen Zahl $a + b$ durch p als Rest erhält. Ebenso setzen wir für alle $[a], [b] \in \mathbb{F}_p$

$$[a] \cdot [b] := [s],$$

[25] Also eine Zahl ≥ 2, die nur durch 1 und sich selbst ohne Rest teilbar ist.

[26] Es gibt auch noch einen tieferliegenden Grund für die Verwendung der eckigen Klammern. Diesen können Sie nachvollziehen, falls Sie einmal in einer Algebra- oder Zahlentheorie-Vorlesung über sogenannte *Quotientenringe* sprechen werden. Für den Moment muss Sie das aber nicht unbedingt kümmern.

wobei s diejenige natürliche Zahl von 0 bis $p-1$ ist, die man bei der ganzzahligen Division der natürlichen Zahl $a \cdot b$ durch p als Rest erhält.

Man kann zeigen,[27] dass $(\mathbb{F}_p, +, \cdot)$ ein Körper ist.[28]

Zum Schluss geben wir noch ein paar konkrete Beispiele an:

Beispiele 2.3.10. (a) In $\mathbb{F}_7$ gilt: Es ist zum Beispiel

$$[3] + [6] = [2],$$

weil 9 geteilt durch 7 den Rest 2 ergibt. Und es ist

$$[3] + [4] = [0],$$

weil 7 geteilt durch 7 den Rest 0 ergibt. Außerdem ist beispielsweise

$$[3] \cdot [5] = [1],$$

weil 15 geteilt durch 7 den Rest 1 ergibt.

(b) In $\mathbb{F}_2$ gilt zum Beispiel

$$[1] + [1] = [0],$$

weil 2 geteilt durch 2 den Rest 0 ergibt.

2.4 Ergänzungen

Charakterisik von Körpern

Ein Begriff, der für das tieferergehende Studium von Körpern relevant ist, ist die **Charakteristik**. Um sie zu definieren, brauchen benötigen wir zunächst die folgenden Notation:

Notation 2.4.1. Sei $\mathbb{K}$ ein Körper und sei $n \in \mathbb{N}^*$. Für jedes $a \in \mathbb{K}$ definiert man

$$na \coloneqq a + \cdots + a,$$

wobei die Summe aus n Summanden besteht.

Nun ist die Charakteristik eines Körpers folgendermaßen definiert:

Definition 2.4.2 (Charakteristik von Körpern)**.** Sei $\mathbb{K}$ ein Körper.

(a) Falls es ein $n \in \mathbb{N}^*$ mit der Eigenschaft $n1 = 0$ gibt, dann nennt man die kleinste solche Zahl die **Charakteristik** von $\mathbb{K}$.

[27] Was wir in der Linearen Algebra 1 aber nicht tun werden.

[28] Es gibt übrigens auch noch andere Körper mit nur endlichen vielen Elementen. Diese werden wir in der Linearen Algebra 1 aber nicht besprechen.

(b) Falls es kein $n \in \mathbb{N}^*$ mit der Eigenschaft $n1 = 0$ gibt, dann definiert man die Charakteristik von $\mathbb{K}$ als 0.

Zum Beispiel haben die Körper $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ die Charakterisitik 0. Für jedes Primzahl $p \in \mathbb{N}$ hat der Körper $\mathbb{F}_p$ hingegen die Charakteristik p. Körper mit der Charakteristik 2 (zum Beispiel $\mathbb{F}_2$) haben die kuriose Eigenschaft, dass $1 + 1 = 0$ und somit $-1 = 1$ gilt. Es sind die einzigen Körper $\mathbb{K}$, in denen man für ein Element $a \in \mathbb{K}$ aus der Gleichung $-a = a$ nicht $a = 0$ folgern kann – was bereits andeutet, dass Körper mit Charakteristik 2 sich häufig etwas anders verhalten als andere Körper.

Man kann zeigen, dass die Charakteristik eines Körpers immer 0 oder eine Primzahl ist. Dies hängt mit dem Begriff der sogenannten **Teilkörper** zusammen: Ähnlich wie es für Untergruppen von Gruppen der Fall ist, ist ein Teilkörper eines Körpers $\mathbb{K}$ eine Teilmenge von $\mathbb{K}$, die bezüglich der algeraischen Operationen in $\mathbb{K}$ abgeschlossen ist, sodass sie selbst ebenfalls einen Körper bildet.

Indem man alle Teilkörper eines Körpers $\mathbb{K}$ schneidet, erhält man immer einen kleinsten Teilkörper von $\mathbb{K}$, und es stellt sich heraus, dass dieser eng mit der Charakteristik von $\mathbb{K}$ zusammenhängt:

- Wenn $\mathbb{K}$ die Charakteristik 0 hat, dann ist der kleinste Teilkörper von $\mathbb{K}$ der Körper $\mathbb{Q}$.[29]
- Wenn $\mathbb{K}$ die Charakteristik p für eine Primzahl $p \in \mathbb{N}^*$ hat, dann ist der kleinste Teilkörper von $\mathbb{K}$ der Körper $\mathbb{F}_p$.

Für mehr Details zur Charakteristik von Körpern verweisen wir zum Beispiel auf [Fis11, Abschnitt 1.3.7].

Literaturhinweise

Abschnitte über die wichtigsten algebraischen Strukturen finden Sie in den meisten Büchern über Lineare Algebra, die im Literaturverzeichnis dieses Manuskripts gelistet sind. Im folgenden führen wir beispielhaft einige Literaturstelle an, die sich ein wenig vom Inhalt dieses Manuskripts unterscheiden:

- In [Beu14, Kapitel 2] werden Körper direkt eingeführt, ohne vorher über Halbgruppen oder Gruppen zu sprechen.
- Die Körper $\mathbb{F}_p$ für p prim werden zum Beispiel in [Beu14, Abschnitt 2.2.3] im Detail besprochen (und dort als Z_p bezeichnet).
- In [Beu14, Abschnitt 2.2.4] wird ein Körper mit 4 Elementen angegeben. Diesen haben wir in der Vorlesung gar nicht erwähnt.

[29] Hier gibt es eigentlich eine kleine Subtilität zu beachten: Genau genommen, muss $\mathbb{Q}$ selbst gar nicht in $\mathbb{K}$ enthalten sein, sondern der kleinste Teilkörper von $\mathbb{K}$ ist lediglich **isomorph** zu $\mathbb{Q}$ – ein Begriff, den man zunächst definieren müsste, um hier weiter über Details zu sprechen.

- In [Bos14, Seite 20] finden Sie den Körper $\mathbb{Q}(\sqrt{2})$ (sprich „$\mathbb{Q}$ adjungiert $\sqrt{2}$“), dessen Eigenschaften Sie womöglich ein wenig an die komplexen Zahlen erinnern (obwohl es sich bei diesem Körper keineswegs um die komplexen Zahlen handelt).
- In [Beu14, Abschnitt 2.2.2] werden die **Quaterionen** besprochen, bei denen es sich nicht um einen Körper, sondern um die etwas allgemeinere Struktur eines **Schiefkörpers** handelt.
- In [Fis11, Abschnitt 1.3.2] wird besprochen, wie man die rationalen Zahlen mit Hilfe sogenannter **Äquivalenzrelationen** aus den ganzen Zahlen konstruieren kann.
- In [Fis11, Abschnitt 1.3.4] wird sogar eine Konstruktion der reellen Zahlen angegeben – dies ist allerdings ein Thema, das tendenziell eher zum Gebiet der Analysis gehört.[30]

[30] Womit natürlich nicht gesagt ist, dass es zwischen Linearer Algebra und Analysis keine Überschneidungen gebe – ganz im Gegenteil, spätestens in der Analysis mehrerer Veränderlichen („Analysis 2“) benötigt man ständig auch Lineare Algebra.

Kapitel 3

Nummerierungen, Rekursion und Induktion

Einstiegsfragen. (a) Sei n eine natürliche Zahl. Kennen Sie eine effiziente mathematische Notation für „die Summe der ersten n ungeraden natürlichen Zahlen“?

(b) A propos: Welcher Wert kommt eigentlich heraus, wenn man die Summe der ersten n ungeraden natürlichen Zahlen berechnet?

(c) Ein Gedankenspiel aus dem Märchenland: In den nächsten Semesterferien haben Sie bei einer guten Fee einen Wunsch frei, und Sie wünschen sich, dass immer und stets, egal was auch passiert, der erste Tag der Vorlesungszeit lieber noch ein Ferientag sei.

Die gute Fee erfüllt freilich Ihren Wunsch (sonst wäre sie ja keine gute Fee). Wann gehen Sie das nächste mal zur Vorlesung?

(d) Stellen Sie (in echt oder in Ihrer Fantasie) eine Webcam vor einem Computerbildschirm B auf, und zwar so, dass sie den Bildschirm und einen kleinen Ausschnitt des Raumes um den Bildschirm abfilmt. Verbinden Sie die Webcam mit dem Computer und lassen Sie live auf dem Bildschirm B anzeigen, was die Webcam filmt.

Was sehen Sie auf dem Bildschirm B?

3.1 Summen und Produkte

Indizierte Variablen, Tupel und Funktionen

Sie haben bisher in der Veranstaltung bereits mehrmals folgende Notation gesehen: Wenn man nicht nur zwei oder drei mathematische Objekte mit Variablen bezeichnen will, sondern mehr Objekte, so ist es oft am einfachsten, sie durchzunummerieren, z.B. in der Form

$$x_1, x_2, \ldots, x_n, \tag{3.1.1}$$

wobei $n \in \mathbb{N}$ ist.[1] Die folgenden beiden Bemerkungen sind wichtig, damit Sie dies richtig einordnen können:

Bemerkung 3.1.1 (Nummerierte Objekte, Tupel, Funktionen)**.** (a) Eine Durchnummerierung von Objekten mit Hilfe von Indizes wie in (3.1.1) ist konsistent mit der Notation, die wir beim Einführen von Tupeln in Definition 1.3.16(b) festgelegt haben: Wenn wir alle n Objekte aus (3.1.1) zu einem Tupel x zusammenfassen, dann wird laut Definition 1.3.16(b) der erste Eintrag des Tupels mit x_1 bezeichnet, der zwei mit x_2, und so weiter. Somit erhalten wir als ebenfalls die Notation $x_1, x_2, \ldots, x_n$ für die n Objekte, die in (3.1.1) aufgelistet sind.

(b) Wir wollen nun noch einen Schritt weiter gehen: Wenn Sie eine Liste von Objekten $x_1, \ldots, x_n$ (mit $n \in \mathbb{N}$) betrachten und all diese Objekte aus einer bestimmten Menge X stammen, so ist das Tupel

$$x = \begin{pmatrix} x_1 & \ldots & x_n \end{pmatrix}$$

nichts weiter als eine Zuordnung, die jeder Zahl von 1 bis n ein Element aus X zuordnet. Anders ausgedrückt: Ein Tupel aus X^n kann man auch als eine Funktion von $\{1, \ldots, n\}$ nach X auffassen.

Funktionswert dieser Funktion an der Stelle 1 ist dann x_1, der Funktionswert an der Stelle 2 ist x_2, und so weiter, und der Funktionswert an der Stelle n ist x_n.[2] Entsprechend benutzt man manchmal auch folgende Notation: Man bezeichnet die Funktion von $\{1, \ldots, n\}$, die zum Tupel x gehört, ebenfalls mit x – und somit ist dann $x(1) = x_1$, $x(2) = x_2$, …, $x(n) = x_n$.[3]

Wenn Sie, dem zweiten Teil der vorangehenden Bemerkung folgend, ein Tupel als Funktion auffassen, dann liegt es auch nahe, das man die verschiedenen Einträge eines Tupels oft durch Formeln spezifizieren kann. Zum Beispiel können wir schreiben „Sei $x \in \mathbb{R}^7$ das Tupel, dessen Einträge durch die Formel

$$x_k = k^2 \qquad \text{für alle } k \in \{1, \ldots, 7\}$$

[1] Hier kann man übrigens tatsächlich auch den Fall $n = 0$ zu lassen – in diesem Fall ist die Aufzählung $x_1, x_2, \ldots, x_n$ so zu verstehen, dass man überhaupt kein Objekt in dieser Liste hat.

[2] Hierin steckt übrigens schon wieder eine kleine Ungenauigkeit: Laut Definition 1.3.16(c) sind zwei Tupel x und y gleich, wenn Sie die gleiche Anzahl an Einträgen haben und die Einträge der beiden Tupel an jeder Stelle übereinstimmen. Wenn Sie die beiden Tupel x und y nun, wie hier beschrieben als Funktionen auffassen, dann müsste für deren Gleichheit aber laut Definition 1.5.7 auch noch deren Wertebereich gleich sein; dies hatten wir jedoch für die Gleichheit zweier Tupel nicht gefordert. Insofern ist es, wenn man allen Definitionen, die wir bisher gemacht haben, exakt folgt, nicht komplett richtig, Tupel als Funktionen aufzufassen. Häufig ist es aber dennoch eine hilfreiche Sichtweise, und wann immer der genaue Wertebereich einer Funktion nicht von Belang ist, entsteht dadurch auch kein Problem.

[3] Diese Schreibweise – mit dem Index in Klammern – wird zum Beispiel in der Software *Octave* verwendet. Das wird in den Übungen in einer Bonusaufgabe noch weiter besprochen.

gegeben sind.“ Oder noch etwas allgemeiner: „Sei $n \in \mathbb{N}$, und sei $x \in \mathbb{R}^n$ das Tupel, dessen Einträge durch die Formel

$$x_k = k^2 \qquad \text{für alle } k \in \{1, \dots, n\}$$

gegeben sind.“

Summen- und Produktschreibweise

Mit den vorausgehenden Bemerkungen gewappnet können wir nun die folgenden Summe- und Produktschreibweise in Körpern einführen:

Definition 3.1.2 (Summe- und Produktzeichen)**.** Sei $(\mathbb{K}, +, \cdot)$ ein Körper. Außerdem seien $m, n \in \mathbb{Z}$, und es seien Elemente $x_m, \dots, x_n \in \mathbb{K}$ gegeben.

(a) Wir definieren

$$\sum_{k=m}^{n} x_k \coloneqq \begin{cases} x_m + \cdots + x_n & \text{falls } n \geq m, \\ 0 & \text{falls } n < m. \end{cases}$$

(b) Wir definieren

$$\prod_{k=m}^{n} x_k \coloneqq \begin{cases} x_m \cdot \cdots \cdot x_n & \text{falls } n \geq m, \\ 1 & \text{falls } n < m. \end{cases}$$

Beispiele 3.1.3. (a) Sei $n \in \mathbb{N}$. Die Summe aller natürlichen Zahl von 0 bis n – also die Zahl $0 + 1 + \cdots + n$ – kann man mit der soeben eingeführten Notation auch kurz schreiben als

$$\sum_{k=0}^{n} k.$$

(b) Sei $n \in \mathbb{N}$. Dann kann man die Zahl $n!$ mit der soeben eingeführten Produktschreibweise auch schreiben als[4]

$$\prod_{k=1}^{n} k.$$

Die üblichen Rechenregeln auf Körpern, die Sie bereits aus den Körperaxiomen kennen – zum Beispiel das Distributivgesetz – übertragen sich ganz leicht auf endliche Summen. Ein paar weitere einfache Rechenregeln für Summen erhalten Sie direkt aus der Definition des Summenzeichens:

[4] An dieser Stelle ist es übrigens eine gute Idee, sich zu überlegen, warum man $n!$ hingegen nicht als $\prod_{k=0}^{n} k$ schreiben kann.

Proposition 3.1.4. *Sei $(\mathbb{K}, +, \cdot)$ ein Körper.*

(a) *Seien $m, n, s \in \mathbb{Z}$ und seien $x_m, \ldots, x_n \in \mathbb{K}$. Dann gilt*[5]

$$\sum_{k=m+s}^{n+s} x_{k-s} = \sum_{k=m}^{n} x_k.$$

(b) *Seien $\ell, m, n \in \mathbb{Z}$ mit $\ell \leq m \leq n$ und seien $x_\ell, \ldots, x_n \in \mathbb{K}$. Dann gilt*

$$\sum_{k=\ell}^{m-1} x_k + \sum_{k=m}^{n} x_k = \sum_{k=\ell}^{n} x_k.$$

Analoge Regeln gelten natürlich auch für das Produktzeichen.

3.2 Rekursion und Induktion

Rekursion

Zu Beginn von Abschnitt 3.1 haben wir diskutiert, wie man mehrere Zahlen mit durchnummerierten Variablen bezeichnen kann. Dasselbe geht auch mit unendlich vielen Variablen. Wir können zum Beispiel sagen: „Für jedes $n \in \mathbb{N}$ sei x_n eine reelle Zahl.“ Und wir können auf diese Wiese auch unendlich viele Zahl konkret spezifizieren, zum Beispiel so: „Für jedes $n \in \mathbb{N}$ sei $x_n \coloneqq \frac{n^3}{4}$“.

Interessanter wird es dann, wenn wir die x_n nicht alle explizit angeben, sondern z.B. nur x_0 angeben, und für alle $n \geq 1$ festlegen, wie man x_n aus der vorangehenden Zahl berechnen. Dieses Vorgehen bezeichnet man als **rekursive Definition** der Zahlen $x_0, x_1, x_2, \ldots$.

Beispiele 3.2.1. (a) Für jedes $n \in \mathbb{N}$ sei eine Zahl $x_n \in \mathbb{R}$ gegeben, die durch folgende Rekursionsvorschrift bestimmt ist:

$$\begin{aligned} & x_0 \coloneqq 0, \\ \forall n \in \mathbb{N}^* : \quad & x_n \coloneqq x_{n-1} + n. \end{aligned}$$

In diesem einfachen Beispiel kann man leicht sehen, dass $x_n = \sum_{k=0}^{n} k$ für alle $n \in \mathbb{N}$ gilt.

(b) Für jedes $n \in \mathbb{N}$ sei eine Zahl $x_n \in \mathbb{R}$ gegeben, die durch folgende Rekursionsvorschrift bestimmt ist:

$$\begin{aligned} & x_0 \coloneqq 1, \\ \forall n \in \mathbb{N}^* : \quad & x_n \coloneqq x_{n-1} \cdot n. \end{aligned}$$

In diesem leichten Beispiel kann man leicht sehen, dass $x_n = n!$ für alle $n \in \mathbb{N}$ gilt.

[5] Dies bezeichnet man manchmal auch als **Indexshift**.

Es ist natürlich auch möglich, Zahlen $x_0, x_1, x_2, \ldots$ auf eine Weise rekursiv zu definieren, bei der jede Zahlen nicht nur von ihrem Vorgänger, sondern zum Beispiel von ihren beiden Vorgängern abhängt. Damit das funktioniert, muss man dann natürlich zwei Werte am Anfang vorgeben, also x_0 und x_1. Hier ist ein sehr klassisches Beispiel hierfür:

Beispiel 3.2.2. Für jedes $n \in \mathbb{N}$ sei eine Zahl $f_n \in \mathbb{R}$ gegeben, die durch folgende Rekursionsvorschrift bestimmt ist:

$$\begin{aligned} f_0 &\coloneqq 1, \quad f_1 \coloneqq 1, \\ \forall n \in \mathbb{N}^* : \quad f_{n+1} &\coloneqq f_n + f_{n-1}. \end{aligned}$$

Man kann natürlich sehr leicht die ersten paar dieser Zahlen ausrechnen. Es ist zum Beispiel

$$\begin{aligned} f_0 &= 1, \\ f_1 &= 1, \\ f_2 &= 1 + 1 = 2, \\ f_3 &= 2 + 1 = 3, \\ f_4 &= 3 + 2 = 5, \\ f_5 &= 5 + 3 = 8, \\ f_6 &= 8 + 5 = 13, \\ f_7 &= 13 + 8 = 21, \\ &\vdots \end{aligned}$$

Allerdings ist es erstaunlich schwierig, eine explizite Formel für x_n zu finden, die für jedes $n \in \mathbb{N}$ gilt.[6]

Das Prinzip der vollständigen Induktion

In diesem Unterabschnitt wollen wir eine Beweismethode besprechen, mit der man Aussagen für alle natürlichen Zahlen beweisen kann. Es handelt sich um das sogenannten **Prinzip der vollständigen Induktion**. Dieses ist eng mit der soeben besprochenen Rekursion verwandt.

Wir benötigen zunächst die folgende Eigenschaft der natürlichen Zahlen:

Proposition 3.2.3. *Jede nichtleere Teilmenge von $\mathbb{N}$ besitzt ein kleinstes Element.*

Anschaulich ist die Eigenschaft klar: Wenn $M \subseteq \mathbb{N}$ nicht leer ist, können Sie bei 0 beginnend so lange die natürlichen Zahlen durchlaufen, bis Sie zum ersten Mal auf eine Zahl aus M treffen. Diese Zahl ist dann das kleinste Element von M. Allerdings gibt es hier ein Haken, den wir nach dem folgenden Theorem 3.2.4 erläutern werden.

[6] Man kann aber tatsächlich so eine Formel angeben; siehe Beispiel 3.2.5(b) weiter unten.

Theorem 3.2.4 (Prinzip der vollständigen Induktion)**.** *Für jedes $n \in \mathbb{N}$ sei eine Aussage $A(n)$ gegeben. Falls $A(0)$ wahr ist und falls für jedes $n \in \mathbb{N}$ die Implikation*

$$A(n) \Rightarrow A(n+1)$$

wahr ist, dann ist für jedes $n \in \mathbb{N}$ die Aussage $A(n)$ wahr.

Beweis. Lassen Sie uns widerspruchshalber annehmen, dass es ein $n \in \mathbb{N}$ gibt, für welches die Aussage $A(n)$ falsch ist. Dann gibt es laut Proposition 3.2.3 ein kleinstes solches n – nennen wir es n_1.[7]

Weil $A(0)$ laut Voraussetzung wahr ist, ist $n_1 \geq 1$. Damit besitzt n_1 einen Vorgänger in $\mathbb{N}$ – nennen wir ihn n_0.[8] Wegen der Minimalität von n_1 ist $A(n_0)$ wahr. Aber laut der zweiten Voraussetzung ist dann auch $A(n_1)$ wahr. Widerspruch. □

Es ist wichtig zu erkennen, dass man beim Prinzip der vollständigen Induktion nicht unbedingt bei 0 beginnen muss. Wenn man zum Beispiel weiß, dass $A(3)$ wahr ist und dass für alle $n \geq 3$ die Implikation

$$A(n) \Rightarrow A(n+1)$$

wahr ist, dann folgt freilich, dass $A(n)$ für alle $n \geq 3$ wahr ist.

Nun zum oben angekündigten Haken: Proposition 3.2.3 scheint zwar anschaulich klar zu sein und Theorem 3.2.4 folgt, wie Sie soeben gesehen haben, aus dieser Proposition. Wenn man aber die natürlichen Zahlen (und anschließend die ganzen Zahlen, die rationalen Zahlen und die reellen Zahlen) der Reihe nach konstruiert, kann man nicht so einfach anschaulich argumentieren wie wir es bei der Erläuterung direkt nach Proposition 3.2.3 getan haben – denn man muss ja zunächst einmal definieren, was die natürlichen Zahlen überhaupt sind.

Deswegen kommt man nicht umhin, irgendeine Aussage wie zum Beispiel Proposition 3.2.3 oder Theorem 3.2.4 bei der definition der natürlichen Zahlen als *Axiom* vorauszusetzen. Dies tut man in Form eines der Axiome unter den sogenannten **Peano-Axiomen**, die die natürlichen Zahlen beschreiben.

Wir demonstrieren nun an zwei Beispielen, wie man das Prinzip der vollständigen Induktion verwenden kann, um Aussagen für alle natürlichen Zahlen zu beweisen. Auf den Übungen sowie im Laufe der Veranstaltung werden Sie noch viele weitere Beispiele sehen.

Beispiele 3.2.5. (a) Für jedes $n \in \mathbb{N}$ gilt

$$\sum_{k=0}^{n} k = \sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

[7] Genauer: Man wendet Proposition 3.2.3 auf die Menge $M \coloneqq \{n \in \mathbb{N} \mid \neg A(n)\} \subseteq \mathbb{N}$ an.

[8] D.h. einfach, es ist $n_0 \coloneqq n_1 - 1$, bzw. $n_1 = n_0 + 1$.

(b) Für die in Beispiel 3.2.2 definierten Fibonacci-Zahlen $f_0, f_1, f_2, \ldots$ gilt:

$$\forall n \in \mathbb{N}: \quad f_n = \frac{1}{\sqrt{5}} \left(\Big(\frac{1+\sqrt{5}}{2}\Big)^{n+1} - \Big(\frac{1-\sqrt{5}}{2}\Big)^{n+1} \right)$$

Beweis. (a) Es genügt, für jedes $n \in \mathbb{N}$ die Gleichheit $\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$ zu zeigen.
Um dies zu tun, bezeichnen wir für jedes $n \in \mathbb{N}$ mit $A(n)$ die Aussage

„Es gilt $\sum_{k=0}^{n} k = \frac{n(n+1)}{2}$.“

Wir wollen zeigen, dass $A(n)$ für jedes $n \in \mathbb{N}$ wahr ist. Dazu genügt es laut Theorem 3.2.4 zwei Dinge zu zeigen:

- Dass $A(0)$ wahr ist, und
- dass für alle $n \in \mathbb{N}$ gilt: $A(n) \Rightarrow A(n+1)$.

Wir zeigen nun also diese beiden Behauptungen.

Behauptung: Es gilt $A(0)$.
Begründung: In der Tat ist einerseits $\sum_{k=0}^{0} k = 0$, und andererseits $\frac{0\cdot(0+1)}{2} = 0$, also ist $A(0)$ wahr.

Behauptung: Für jedes $n \in \mathbb{N}$ gilt $A(n) \Rightarrow A(n+1)$.
Begründung: Sei $n \in \mathbb{N}$ beliebig, aber fest. Wir setzen voraus, dass $A(n)$ wahr ist. Um zu zeigen, dass dann auch $A(n+1)$ war ist, berechnen wir

$$\begin{aligned} \sum_{k=0}^{n+1} k = \sum_{k=0}^{n} k \ + \ (n+1) \ &\overset{A(n) \text{ ist wahr}}{=} \ \frac{n(n+1)}{2} + (n+1) \\ &= (n+1)\big(\frac{n}{2} + 1\big) = \frac{(n+1)(n+2)}{2} = \frac{(n+1)\big((n+1)+1\big)}{2}. \end{aligned}$$

Also ist auch $A(n+1)$ wahr. Damit ist die Behauptung gezeigt.

(b) Den Beweises dieser Formel lagern wir in die Übungen aus. □

Literaturhinweise

- Zur Summennotation können Sie beispielsweise [Beu14, Abschnitt 1.5] konsultieren.
- Eine sehr ausführliche Besprechung von Induktion und Rekursion finden Sie zum Beispiel in [Blo11, Abschnitte 6.3 und 6.4].
- Zahlreiche Informationen zur mathematischen Rekursion aus Sicht der Informatik können Sie beispielsweise in [TT08, Kapitel 8] nachlesen.

Kapitel 4

Grundbegriffe der Linearen Algebra

Einstiegsfragen. (a) Wie können Sie die Position eines Punktes im Raum beschreiben?

Wie können Sie die Positionen von vier verschiedenen Punkten im Raum beschreiben?

(b) Zeichnen Sie fünf Punkte Ihrer Wahl in ein ebenes Koordinatensystem ein. Sie wollen nun alle fünf Punkte um 2 nach rechts und 1 nach unten verschieben.

Wie können Sie die Koordinaten der verschobenen Punkte aus den Koordinaten der ursprünglichen Punkte berechnen?

(c) Was verstehen Sie unter dem Begriff „linear“?

(d) Betrachten Sie eine Teilmenge M des $\mathbb{R}^3$. Was bedeutet es anschaulich zu sagen, dass M eine „Ursprungsebene“ ist?

Gibt es eine Möglichkeit, um mit Hilfe von Rechenoperationen auszudrücken, ob M eine Ursprungsebene ist?

(e) Können Sie reelle Zahlen x_1, x_2 und x_3 finden, die das folgende Gleichungssystem erfüllen?

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 1, \\ x_1 - x_2 + x_3 &= 0, \\ x_1 + x_2 - x_3 &= -4. \end{aligned}$$

4.1 Vektorräume und Untervektorräume

Addieren und Multiplizieren – noch einmal

Nun beginnen wir mit dem Teil der Vorlesung, in dem es tatsächlich um lineare Algebra geht. Lassen Sie uns zunächst kurz zur Motivation ein einfaches Beispiel

diskutieren, dass Sie vermutlich schon aus der Schule kennen:

Beispiel 4.1.1. Betrachten Sie in der Menge $\mathbb{R}^2$ – welche wir geometrisch als Ebene interpretieren können – die beiden Punkte

$$v \coloneqq \begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad \text{und} \qquad w \coloneqq \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Wenn wir vom Ursprung aus einen Pfeil zu beiden Punkten zeichnen, und den einen Pfeil anschließend an den anderen „anhängen“, so bedeutet das Aneinanderhängen der beiden Pfeile rechnerisch gerade, dass wir die Einträge der beiden Tupel v und w addieren und somit den Punkt

$$\begin{pmatrix} 3 \\ 3 \end{pmatrix}$$

erhalten.

Betrachten Sie nun zum Beispiel den Punkt

$$u \coloneqq \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Wie denken uns wieder einen Pfeil vom Ursprung zu diesem Punkt und wollen diesen Pfeil nun mit dem Faktor zwei strecken. Hierzu müssen wir beiden Einträge des Tupels u mit dem Faktor 2 multiplizieren; der gestreckte Pfeil verläuft somit vom Ursprung zum Punkt

$$\begin{pmatrix} 2 \\ -2 \end{pmatrix}.$$

Die beiden soeben beschriebenen Operationen – Addition sowie Multiplikation mit Zahlen – wollen wir nun verallgemeinern. Dies führt uns zum folgenden Konzept.

Vektorräume

Definition 4.1.2 (Vektorraum)**.** Sei $(\mathbb{K}, +, \cdot)$ ein Körper. Ein **Vektorraum über** $\mathbb{K}$ ist ein Tupel $(V, \oplus, \odot)$, wobei V eine Menge ist, und $\oplus$ und $\odot$ Abbildungen sind, welche die folgenden Axiome erfüllen:[1]

(VR0) *Definitions- und Wertebereich der Verknüpfungen:*[2] Es gilt $\oplus : V^2 \to V$ und $\odot : \mathbb{K} \times V \to V$.

(VR1) *Axiome der Addition:* Es ist $(V, \oplus)$ eine kommutative Gruppe.

[1] Die Abbildung $\oplus$ wird häufig als **Addition** bezeichnet, und die Abbildung $\odot$ häufig als **skalare Multiplikation**.

[2] Wir notieren $\oplus$ und $\odot$ ebenfalls in Infix-Notation.

(VR2) *Assoziativität der Multiplikation:*

$$\forall \alpha, \beta \in \mathbb{K} \quad \forall v \in V: \quad (\alpha \cdot \beta) \odot v = \alpha \odot (\beta \odot v)$$

(VR3) *Multiplikation ist distributiv über Vektoraddition:*[3]

$$\forall \alpha \in \mathbb{K} \quad \forall v, w \in V: \quad \alpha \odot (v \oplus w) = \alpha \odot v \ \oplus \ \alpha \odot w$$

(VR4) *Multiplikation ist distributiv über sklarer Addition:*

$$\forall \alpha, \beta \in \mathbb{K} \quad \forall v \in V: \quad (\alpha + \beta) \odot v = \alpha \odot v \ \oplus \ \beta \odot v.$$

(VR5) *Nicht-Trivialität der Multiplikation:*

$$\forall v \in V: \quad 1 \odot v = v.$$

Die Elemente aus V werden als **Vektoren** bezeichnet, und die Elemente aus $\mathbb{K}$ werden häufig als **Skalare** bezeichnet. Man nennt $\mathbb{K}$ auch den **Skalarkörper**, der dem Vektorraum zugrundeliegt.

Des neutrale Element der Gruppe $(V, \oplus)$ wird mit 0_V bezeichnet, und das inverse Elemente von einem Element $v \in V$ (bzgl. der Verknüpfung $\oplus$) mit $\ominus v$.

Ebenso wie in Körpern schreibt man Terme wie zum Beispiel $v \oplus (\ominus w)$ (für $v, w \in V$) etwas kürzer als $v \ominus w$.

Lassen Sie uns zunächst einige einfache Eigenschaften zeigen, die in allen Vektorräumen gelten:

Proposition 4.1.3. *Sei $(V, \oplus, \odot)$ ein Vektorraum über einem Körper $(\mathbb{K}, +, \cdot)$.*

(a) *Für alle $\alpha \in \mathbb{K}$ und alle $v \in V$ gilt:*

$$\alpha \odot v = 0_V \qquad \Leftrightarrow \qquad \alpha = 0 \quad \vee \quad v = 0_V.$$

(b) *Für alle $v \in V$ gilt $(-1) \odot v = \ominus v$.*

Beweis. (a) Dies lässt sich sehr ähnlich wie Proposition 2.3.2(a) zeigen.

(b) Sei $v \in V$ beliebig, aber fest. Dann gilt

$$(-1) \odot v \ \oplus \ v \overset{\text{(VR5)}}{=} (-1) \odot v \ \oplus \ 1 \odot v \overset{\text{(VR4)}}{=} \big((-1) + 1\big) \odot v = 0 \odot v \overset{\text{(a)}}{=} 0_V.$$

Also ist $(-1) \odot v$ tatsächlich das inverse Element von v bezüglich $\oplus$. □

Nun besprechen wir einige Beispiele von Vekorräumen.

Beispiele 4.1.4. Sei $(\mathbb{K}, +, \cdot)$ ein Körper.

[3] Hier und im Folgenden verwenden wir ebenfalls die Konvention „Punkt vor Strich".

(a) Sei $d \in \mathbb{N}^*$. Wir definieren Abbildungen $\oplus : \mathbb{K}^d \times \mathbb{K}^d \to \mathbb{K}^d$ und $\odot : \mathbb{K} \times \mathbb{K}^d \to \mathbb{K}^d$ durch

$$v \oplus w \coloneqq \begin{pmatrix} v_1 + w_1 \\ \vdots \\ v_d + w_d \end{pmatrix}, \qquad \alpha \odot v \coloneqq \begin{pmatrix} \alpha \cdot v_1 \\ \vdots \\ \alpha \cdot v_d \end{pmatrix}$$

für alle $v, w \in \mathbb{K}^d$ und alle $\alpha \in \mathbb{K}$.

Dann ist $(\mathbb{K}^d, \oplus, \odot)$ ein Vektorraum über $(\mathbb{K}, +, \cdot)$. Dies werden Sie in den Übungen beweisen.

(b) Für den Fall $d = 1$ gilt mit der vorangehenden Notation $(\mathbb{K}^1, \oplus, \odot) = (\mathbb{K}, +, \cdot)$. Somit ist also jeder Körper ein Vektorraum über sich selbst.

Beispiel 4.1.5. Sei $n \in \mathbb{N}^*$ und seien $(V_1, \oplus_1, \odot_1)$, …, $(V_n, \oplus_n, \odot_n)$ Vektorräume über einem Körper $(\mathbb{K}, +, \cdot)$. Wir setzen $V \coloneqq V_1 \times \cdots \times V_n$, und wir definieren Abbildungen $\oplus : V \times V \to V$ und $\odot : \mathbb{K} \otimes V \to V$ durch

$$v \oplus w \coloneqq \begin{pmatrix} v_1 \oplus_1 w_1 \\ \vdots \\ v_n \oplus_n w_n \end{pmatrix}, \qquad \alpha \odot v \coloneqq \begin{pmatrix} \alpha \odot_1 v_1 \\ \vdots \\ \alpha \odot_n v_n \end{pmatrix}$$

für alle $v, w \in V$ und alle $\alpha \in \mathbb{K}$. Dann ist $(V, \oplus, \odot)$ ein Vektorraum über $(\mathbb{K}, +, \cdot)$.

Beweis. Wir beweisen beispielhaft, dass das Vektorraum-Axiom (VR3) erfüllt ist. Der Beweis der anderen Axiome verläuft ähnlich.

(VR3): Seien $\alpha \in \mathbb{K}$ und $v, w \in V$ beliebig, aber fest. Dann gilt

$$\alpha \odot (v \oplus w) = \alpha \odot \begin{pmatrix} v_1 \oplus_1 w_1 \\ \vdots \\ v_n \oplus_n w_n \end{pmatrix} = \begin{pmatrix} \alpha \odot_1 \left(v_1 \oplus_1 w_1\right) \\ \vdots \\ \alpha \odot_n \left(v_n \oplus_n w_n\right) \end{pmatrix}$$

$$\overset{\text{(VR3) in } (V_1, \oplus_1, \odot_1), \ldots, (V_n, \oplus_n, \odot_n)}{=} \begin{pmatrix} \alpha \odot_1 v_1 \ \oplus_1 \ \alpha \odot_1 w_1 \\ \vdots \\ \alpha \odot_n v_n \ \oplus_n \ \alpha \odot_n w_n \end{pmatrix}$$

$$= \begin{pmatrix} \alpha \odot_1 v_1 \\ \vdots \\ \alpha \odot_n v_n \end{pmatrix} \oplus \begin{pmatrix} \alpha \odot_1 w_1 \\ \vdots \\ \alpha \odot_n w_n \end{pmatrix} = \alpha \odot v \ \oplus \ \alpha \odot w.$$

Damit ist bewiesen, dass $(V, \oplus, \odot)$ das Axiom (VR3) erfüllt. □

Beispiel 4.1.6. Sei X eine nicht-leere Menge und sei $(\mathbb{K}, +, \cdot)$ ein Körper. Wir setzen $V \coloneqq \mathrm{Abb}(X; \mathbb{K}) \coloneqq \{f \mid f : X \to \mathbb{K}\}$ und definieren zwei Abbildungen $\oplus : V \times V \to V$ und $\odot : \mathbb{K} \times V \to V$ durch

$$f \oplus g\colon \quad X \to \mathbb{K},$$

$$x \mapsto (f \oplus g)(x) \coloneqq f(x) + g(x)$$

und

$$\begin{aligned} \alpha \odot f\colon \quad & X \to \mathbb{K}, \\ & x \mapsto (\alpha \odot f)(x) \coloneqq \alpha \cdot f(x) \end{aligned}$$

für alle $f, g \in V$ und alle $\alpha \in \mathbb{K}$. Dann kann man nachrechnen, dass $(V, \oplus, \odot)$ ein Vektorraum über $(\mathbb{K}, +, \cdot)$ ist.

Untervektorräume

Bei der Einführung von Untergruppen haben wir in Abschnitt 2.2 bereits erwähnt, dass es in der Mathematik sehr üblich ist, für gegebene algebraische Stukturen jeweils Unterstrukturen zu betrachten. Dies tun wir nun auch für Vektorräume.

Definition 4.1.7 (Untervektorraum)**.** Sei $(V, \oplus, \odot)$ ein Vektorraum über einem Körper $(\mathbb{K}, +, \cdot)$. Eine Teilmenge $U \subseteq V$ heißt **Untervektorraum** von $(V, \oplus, \odot)$, falls die folgenden Eigenschaften erfüllt sind:

(UVR1) Es gilt $0_V \in U$.

(UVR2) Für alle $v, w \in U$ gilt $v \oplus w \in U$.

(UVR3) Für alle $\alpha \in \mathbb{K}$ und alle $u \in U$ gilt $\alpha \odot u \in U$.

Proposition 4.1.8. *Sei $(V, \oplus, \odot)$ ein Vektorraum über einem Körper $(\mathbb{K}, +, \cdot)$ und sei $U \subseteq V$ ein Untervektorraum von $(V, \oplus, \odot)$.*

Dann ist $(U, \oplus, \odot)$ selbst ein Vektorraum, und 0_V ist auch in U das neutrale Element der Addition.

Beweis. Der Beweis ist sehr ähnlich wie der Beweis von Proposition 2.2.7. Der einzige wesentliche Unterschied besteht darin, beim Beweis des Vektorraum-Axioms (VR1) des Existenz von inversen Elementen in U zu zeigen. Hierzu kann man folgendermaßen argumentieren:

Sei $u \in U$. Dann gilt laut Proposition (b), dass $\ominus u = (-1) \odot u$ ist, und letztgenanntes Produkt ist laut Untergruppen-Axiom (UVR3) ein Element von U. Somit besitzt u in U ein inverses Elemente bzgl. $\oplus$ – nämlich das Element $\ominus u$. □

Untervektorräume spielen in der Linearen Algebra eine zentrale Rolle. Im folgenden diskutieren wir aber zunächst einige sehr einfache Beispiele von Untervektorräumen.

Beispiele 4.1.9. (a) Sei $(V, \oplus, \odot)$ ein Vektorraum über einem Körper $(\mathbb{K}, +, \cdot)$. Dann sind V selbst und $\{0_V\}$ Untervektorräume von $(V, \oplus, \odot)$; man nennt sie die **trivialen Untervektorräume** von $(V, \oplus, \odot)$.

(b) Die Menge

$$U \coloneqq \left\{ \alpha \odot \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mid \alpha \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \mid \alpha \in \mathbb{R} \right\}$$

ist ein Untervektorraum des $\mathbb{R}^2$ (welcher den Verknüpfungen aus Beispiel (a) versehen ist).

Geometrisch ist U die winkelhalbierende Gerade des ersten und dritten Quadranten in der Ebene $\mathbb{R}^2$.

Untervektorräume haben eine Eigenschaft, die später noch sehr nützlich werden wir – sie sind stabil bezüglich Durchschnittsbildung:

Proposition 4.1.10. *Sei $(V, \oplus, \odot)$ ein Vektorraum über einem Körper $(\mathbb{K}, +, \odot)$. Sei J eine nicht-leere Menge, und für jedes $j \in J$ sei ein Untervektorraum U_j von $(V, \oplus, \odot)$ gegeben. Dann ist auch*

$$U \coloneqq \bigcap_{j \in J} U_j$$

ein Untervektorraum von $(V, \oplus, \odot)$.

Beweis. Wir müssen zeigen, dass U die Untervektorraum-Axiome erfüllt:

- (UVR1): Für jedes $j \in J$ gilt $0_V \in U_j$, da U_j nach Voraussetzung ein Untervektorraum von $(V, \oplus, \odot)$ ist. Somit folgt
$$0_V \in \bigcap_{j \in J} U_j = U.$$
- (UVR2): Seien $v, w \in U$ beliebig, aber fest. Für jedes $j \in J$ gilt dann $v, w \in U_j$ und somit, weil U_j ein Untervektorraum von $(V, \oplus, \odot)$ ist, auch $v \oplus w \in U_j$. Also folgt
$$v \oplus w \in \bigcap_{j \in J} U_j = U.$$
- (UVR3): Seien $\alpha \in \mathbb{K}$ und $u \in U$ beliebig, aber fest. Für jedes $j \in J$ gilt dann $u \in U_j$ und somit, weil U_j ein Untervektorraum von $(V, \oplus, \odot)$ ist, auch $\alpha \odot u \in U_j$. Folglich ist
$$\alpha \odot v \in \bigcap_{j \in J} U_j = U.$$

Also ist U tatsächlich ein Untervektorraum von $(V, \oplus, \odot)$. □

Zum Schluss dieses Abschnitts wollen wir nun die Notation etwas vereinfachen:

Bemerkungen 4.1.11. (a) Ab sofort schreiben wir in Vektorräumen die Addition nicht mehr als $\oplus$ sondern als $+$, und die Multiplikation nicht mehr als $\odot$ sondern als $\cdot$. Ebenso verwenden wir anstelle des Symbols $\ominus$ nun einfach das Symbol $-$.

Den Nullvektor in einem Vektorraum – d.h. das neutrale Element bezüglich der Addition – bezeichnen wir von nun an nicht mehr mit 0_V, sondern einfach mit 0.

(b) Wie in Körpern wird auch in Vektorräumen das Multiplikationszeichen $\cdot$ beim Rechnen oft weggelassen.

(c) Das Summenzeichen, das wir in Definition 3.1.2(a) für Elemente von Körpern eingeführt haben, kann man natürlich ebenso gut auch für Elemente von Vektorräumen verwenden.[4]

(d) Von nun an werden wir häufig etwas unpräziser in der Notation sein und die Rechenoperationen bei der Einführung eines Vektorraums oder eines Körpers nicht mehr explizit angeben. Wir sagen von nun an also zum Beispiel anstelle von

„Sei $(V, +, \cdot)$ ein Vektorraum über $(\mathbb{K}, +, \cdot)$."

nur noch

„Sei V ein Vektorraum über $\mathbb{K}$."

Dass die entsprechenden Operationen stets mit $+$ und $\cdot$ bezeichnet werden, denken wir uns dann einfach dazu.

4.2 Lineare Abbildungen und Matrizen

Abbildungen, die sich mit der Vektorraumstruktur vertragen

Wir betrachten nun Abbildungen zwischen Vektorräumen, die sich mit der Addition und der Multiplikation vertragen. Solche Abbildungen nennt man **linear**.

Definition 4.2.1. Seien V und W Vektorräume über einem Körper $\mathbb{K}$.

(a) Eine Abbildung[5] $T : V \to W$ heißt **linear**, falls sie die folgenden Eigenschaften erfüllt:

[4] Frage, anhand derer Sie Ihr Verständnis überprüfen können: Weshalb ist es nicht sinnvoll zu sagen, dass man auch das Produktzeichen, dass wir in Definition 3.1.2(a) für Elemente von Körpern eingeführt haben, ebenso auch für Elemente von Vektorräumen verwenden kann.

[5] Abbildungen zwischen Vektorräumen werden wir häufig mit Großbuchstaben bezeichnen. Dies hat keinen tieferen Grund, ist aber in manchen Bereichen so üblich.

(L1) *Verträglichkeit mit der Addition:*

$$\forall v, \tilde{v} \in V: \quad T(v + \tilde{v}) = T(v) + T(\tilde{v}).$$

(L2) *Verträglichkeit mit der Multiplikation:*

$$\forall \alpha \in \mathbb{K} \forall v \in V: \quad T(\alpha v) = \alpha T(v).$$

(b) Ein **Vektorraumisomorphismus** zwischen V und W ist eine Abbildung $T: V \to W$, die linear und bijektiv ist.

Lineare Abbildungen bilden immer den Nullvektor auf den Nullvektor ab:

Proposition 4.2.2. *Seien V und W Vektorräume über einem Körper $\mathbb{K}$ und sei $T: V \to W$ linear. Dann gilt $T(0) = 0$*

Beweis. Es gilt

$$T(0) = T(0 + 0) = T(0) + T(0).$$

Durch Addition des Vektors $-T(0)$ auf beiden Seiten erhalten wir hieraus $0 = T(0)$. □

Beispiele 4.2.3. (a) Die Abbildung $R: \mathbb{R}^2 \to \mathbb{R}^2$, die durch

$$R(x) = \begin{pmatrix} x_1 - 2x_2 \\ x_1 \end{pmatrix}$$

für alle $x \in \mathbb{R}^2$ gegeben ist, ist linear.

Beweis. Seien $x, y \in \mathbb{R}^2$. Dann gilt

$$\begin{aligned} R(x + y) &= R\left(\begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}\right) = \begin{pmatrix} x_1 + y_1 - 2(x_2 + y_2) \\ x_1 + y_1 \end{pmatrix} \\ &= \begin{pmatrix} x_1 - 2x_1 \\ x_1 \end{pmatrix} + \begin{pmatrix} y_1 - 2y_2 \\ y_1 \end{pmatrix} = R(x) + R(y). \end{aligned}$$

Ebenso zeigt man, dass $R(\alpha x) = \alpha R(x)$ für alle $\alpha \in \mathbb{R}$ und alle $x \in \mathbb{R}^2$ ist. □

(b) Die Abbildung $S: \mathbb{R}^2 \to \mathbb{R}^2$, die durch

$$S(x) = \begin{pmatrix} x_2 \\ x_1 - 1 \end{pmatrix}$$

für alle $x \in \mathbb{R}^2$ gegeben ist, ist nicht linear. Es gilt nämlich

$$S(0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

also kann S wegen Proposition 4.2.2 nicht linear sein.

(c) Sei $T : \mathbb{R}^2 \to \mathbb{R}^2$ durch

$$T(x) = \begin{pmatrix} x_1 x_2 \\ 0 \end{pmatrix}$$

für alle $x \in \mathbb{R}^2$ gegeben. Dann gilt zwar $T(0) = 0$, aber T ist trotzdem nicht linear, denn beispielsweise für

$$x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

ist

$$T(2x) = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \qquad \text{aber} \qquad 2T(x) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

Selbst das Nachrechnen der Linearität einer so simplen Abbildung wie in Beispiel 4.2.3(a) ist, wie sie oben sehen können, etwas unübersichtlich. Deshalb hat man zur Beschreibung von linearen Abbildungen zwischen $\mathbb{R}^n$ und $\mathbb{R}^m$ – oder allgemeiner, zwischen $\mathbb{K}^n$ und $\mathbb{K}^m$ – ein Konzept entwickelt, dass übersichtlicher und effizienter ist. Dieses Konzept behandeln wir als nächstes.

Matrizen und Rechenoperationen mit Matrizen

Definition 4.2.4. Sei $\mathbb{K}$ ein Körper und seien $m, n \in \mathbb{N}^*$.

(a) Eine $m \times n$-**Matrix mit Einträgen aus** $\mathbb{K}$ ist eine Tabelle A mit m Zeilen und n Spalten, die mit Elementen von $\mathbb{K}$ gefüllt ist. Dabei können Elemente auch mehrmals vorkommen.

Für jedes $j \in \{1, \ldots, m\}$ und jedes $k \in \{1, \ldots, n\}$ notieren wir den Eintrag von A in der j-ten Zeile und k-ten Spalte mit dem Symbol A_{jk}.[6]

Die Menge all dieser Matrizen bezeichnen wir mit $\mathbb{K}^{m \times n}$.

(b) Für jedes $A \in \mathbb{K}^{m \times n}$ definieren wir eine Matrix $A^{\mathrm{T}} \in \mathbb{K}^{n \times m}$, welche die Einträge

$$(A^{\mathrm{T}})_{kj} = A_{jk}$$

für alle $k \in \{1, \ldots, n\}$ und alle $j \in \{1, \ldots, m\}$ besitzt. Wir nennen A^{T} die **transponierte Matrix** von A.

(c) Zwei Matrizen mit Einträgen aus $\mathbb{K}$ heißen **gleich**, wenn sie gleichviele Spalten haben, gleich viele Zeilen haben, und ihre Einträge an jeder Stelle übereinstimmen.

[6] Beachten Sie, dass also $A_{jk} \in \mathbb{K}$ ist.

Wie Sie aus der Definition der Gleichheit von Matrizen sehen können, ist es wichtig, in welcher „Richtung“ man Matrizen aufschreibt. Für den kompletten Rest der Veranstaltung ist es sehr nützlich, Tupel auf $\mathbb{K}^n$ (für einen Körper $\mathbb{K}$ und eine Zahl $n \in \mathbb{N}^*$) als Matrizen in $\mathbb{K}^{n\times 1}$ aufzufassen. Deshalb werden wir Tupel auf $\mathbb{K}^n$ ab sofort immer in Spaltenform – d.h., von oben nach unten – aufschreiben und in eckige statt in runde Klammern setzen.[7]

Definition 4.2.5 (Rechnen mit Matrizen)**.** Sei $\mathbb{K}$ ein Körper und seien $\ell, m, n \in \mathbb{N}^*$.

(a) Für alle Matrizen $B, \tilde{B} \in \mathbb{K}^{m\times n}$ definieren wir eine Matrix $B + \tilde{B} \in \mathbb{K}^{m\times n}$ mit den Einträgen

$$(B + \tilde{B})_{jk} \coloneqq B_{jk} + \tilde{B}_{jk}$$

für alle $j \in \{1, \dots, m\}$ und alle $k \in \{1, \dots, n\}$.

(b) Für jedes Matrix $B \in \mathbb{K}^{m\times n}$ und jeden Skalar $\alpha \in \mathbb{K}$ definiert man eine Matrix $\alpha B \in \mathbb{K}^{m\times n}$ (die man manchmal auch als $\alpha \cdot B$ notiert) mit den Einträgen

$$(\alpha B)_{jk} \coloneqq \alpha \cdot B_{jk}$$

für alle $j \in \{1, \dots, m\}$ und alle $k \in \{1, \dots, n\}$.

(c) Für zwei Matrizen $A \in \mathbb{K}^{\ell\times m}$ und $B \in \mathbb{K}^{m\times n}$ ist das **Matrixprodukt** von A und B – dass wir als $A \cdot B$ oder kürzer als AB notieren – die Matrix $AB \in \mathbb{K}^{\ell\times n}$ mit den Einträgen

$$(AB)_{hk} \coloneqq \sum_{j=1}^{m} A_{hj} B_{jk}$$

für alle $h \in \{1, \dots, \ell\}$ und alle $k \in \{1, \dots, n\}$.

Beachten Sie unbedingt, dass das Matrixprodukt von A und B nur dann definiert ist, wenn die Anzahl der Spalten von A mit der Anzahl der Zeilen von B übereinstimmt.

Diskussion 4.2.6 (Zwei verschiedene Arten, die Matrixmultiplikation aufzufassen)**.** Sei $\mathbb{K}$ ein Körper, seien $\ell, m, n \in \mathbb{N}^*$ und seien $A \in \mathbb{K}^{\ell\times m}$ und $B \in \mathbb{K}^{m\times n}$.

Aus der Definition des Produkts von Matrizen erhält man verschiedene Möglichkeiten, wie man die Multiplikation zweier Matrizen verstehen kann. Es ist wichtig, beide Möglichkeiten zu kennen:[8]

[7] Man kann dies kurz ausdrücken, indem man sagt: „Wir identifizieren $\mathbb{K}^n$ mit $\mathbb{K}^{n\times 1}$.“

[8] In der Vorlesung werden diese zwei Möglichkeiten an der Tafel noch etwas plastischer dargestellt. Dies lässt sich hier im Manuskript aber kaum darstellen.

(a) *Die einzelnen Einträge von AB:* Direkt aus der Definition des Matrixprodukts erkennt man, dass, für $h \in \{1, \dots, \ell\}$ und $k \in \{1, \dots, n\}$, der Eintrag von AB an der Position (h, k) gegeben ist als das Produkt der h-ten Zeile von A mit der k-ten Spalte von B.[9]

(b) *Die Spalten von AB:*

Wenn wir die Spalten von A mit $A^{(1)}, \dots, A^{(m)} \in \mathbb{K}^\ell$ bezeichnen, dann kann man das Produkt AB auch folgendermaßen verstehen:

Für jedes $k \in \{1, \dots, n\}$ berechnet man die k-te Spalte von AB indem man die Spalten $A^{(1)}, \dots, A^{(m)}$ mit den m Einträgen der k-ten Spalte von B multipliziert und dann alle so entstandenen Vektoren in $\mathbb{K}^\ell$ aufsummiert. Oder in Formeln ausgedrückt: Die k-te Spalte von AB ist gleich der Summe[10]

$$B_{1k}A^{(1)} + B_{2k}A^{(2)} + \cdots + B_mA^{(m)}.$$

Rechenregeln für Matrizen

Für die soeben eingeführten Rechenoperationen gelten einige sehr nützliche Rechenregeln:

Proposition 4.2.7. *Sei $\mathbb{K}$ ein Körper, seien $\ell, m, n, p \in \mathbb{N}^*$. und seinen $A \in \mathbb{K}^{\ell \times m}$, $B, \tilde{B} \in \mathbb{K}^{m \times n}$ und $C \in \mathbb{K}^{n \times p}$. Außerdem sei $\alpha \in \mathbb{K}$.*

(a) *Mit den in Definition 4.2.5*(a) *und* (b) *Rechenoperationen $+$ und $\cdot$ gilt: $\mathbb{K}^{m \times n}$ ist ein Vektorraum über $\mathbb{K}$.*

(b) *Matrixmultiplikation ist assoziativ wann immer sie definiert ist, d.h. es gilt $(AB)C = A(BC)$.*

(c) *Matrixmultiplikation ist distributiv über der Matrixaddition, d.h. es gilt $A(B + \tilde{B}) = AB + A\tilde{B}$ und $(B + \tilde{B})C = BC + \tilde{B}C$.*

(d) *Es gilt $(\alpha B)C = \alpha(BC) = B(\alpha C)$.*

(e) *Rechenregeln für transponierte Matrizen: Es gilt $(B + \tilde{B})^{\mathrm{T}} = B^{\mathrm{T}} + \tilde{B}^{\mathrm{T}}$ sowie und $(\alpha B)^{\mathrm{T}} = \alpha(B^{\mathrm{T}})$ und $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$.*

Beweis. Die Beweise folgen alle direkt aus der Definition der entsprechenden Rechenoperationen und den Körperaxiomen von $\mathbb{K}$. Beispielhaft zeigen wir hier den Beweis der ersten Gleichung in (c):

Laut Definition des Matrixprodukts und der Addition von Matrizen ist sowohl $A(B + \tilde{B})$ als auch $AB + A\tilde{B}$ ein Element von $\mathbb{R}^{\ell \times n}$. Also müssen wir noch zeigen,

[9] Hierbei ist der Ausdruck „Produkt einer Zeile mit einer Spalte“ so gemeint, dass man an jeder Stelle den Eintrag der Zeile mit dem Eintrag der Spalte multiplizert – so erhält man also insgesamt n Produkte – und diese dann aufsummiert.

[10] Eine gute Übung an dieser Stelle: Können Sie diese Summe auch mit Hilfe des Summenzeichens Σ schreiben?

dass beide Matrizen diesselben Einträge haben. Für alle $h \in \{1, \dots, \ell\}$ und alle $k \in \{1, \dots, n\}$ gilt

$$\Big(A(B+\tilde{B})\Big)_{hk} = \sum_{j=1}^{m} A_{hj}(B+\tilde{B})_{jk}$$
$$= \sum_{j=1}^{m} A_{hj}(B_{jk}+\tilde{B}_{jk}) = \sum_{j=1}^{m}(A_{hj}B_{jk} + A_{hj}\tilde{B}_{jk})$$

□

Mit Hilfe von Matrizen kann man lineare Abbildungen von $\mathbb{K}^n$ nach $\mathbb{K}^m$ beschreiben:

Proposition 4.2.8. *Sei $\mathbb{K}$ ein Körper, seien $m, n \in \mathbb{N}^*$.*

(a) *Für jede Matrix $A \in \mathbb{K}^{m\times n}$ ist die Abbildung*

$$L_A : \mathbb{K}^n \to \mathbb{K}^m,$$
$$x \mapsto Ax$$

linear.

(b) *Und umgekehrt gilt: Für jede lineare Abbildung $T : \mathbb{K}^n \to \mathbb{K}^m$ gibt es eine Matrix $A \in \mathbb{K}^m$ mit der Eigenschaft*

$$\forall x \in \mathbb{K}^n : \quad T(x) = Ax.$$

Für den Beweis des zweiten Teils der Proposition ist die folgende Notation hilfreich: Für einen Körper $\mathbb{K}$, für $n \in \mathbb{N}^*$ und $k \in \{1, \dots, n\}$ verwenden wir die Notation

$$e_k \coloneqq \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{K}^n,$$

wobej die 1 an der k-ten Stelle steht. Der Vektor e_k heißt der **k-te kanonischen Einheitsvektor** in $\mathbb{K}^n$. Beachten Sie: Wenn $A \in \mathbb{K}^{m\times n}$ ist, dann folgt aus der Definition des Matrixprodukts, dass $Ae_k \in \mathbb{K}^m$ genau die k-te Spalte von A ist.

Beweis von Proposition 4.2.8. (a) Sei $A \in \mathbb{K}^{m\times n}$ beliebig, fest. Wegen Proposition 4.2.7(c) gilt für alle $x, y \in \mathbb{K}^n$

$$L_A(x+y) = A(x+y) = Ax + Ay = L_A(x) + L_A(y),$$

und wegen Proposition 4.2.7(d) gilt für alle $\alpha \in \mathbb{K}$ und alle $x \in \mathbb{K}^n$

$$L_A(\alpha x) = A(\alpha x) = \alpha(Ax) = \alpha L_A(x).$$

Also ist L_A tatsächlich linear.

(b) Für jedes $k \in \{1, \dots, n\}$ definieren wir $s_k \coloneqq T(e_k) \in \mathbb{K}^m$. Nun sei $A \in \mathbb{K}^{m \times n}$ die Matrix mit den Spalten $s_1, \dots, s_n$, d.h.,

$$A = \begin{bmatrix} s_1 & \dots & s_n \end{bmatrix}.$$

Für jedes $x \in \mathbb{K}^n$ gilt $x = \sum_{k=1}^n x_k e_k$, und somit

$$T(x) = T\Big(\sum_{k=1}^n x_k e_k\Big) = \sum_{k=1}^n x_k T(e_k) = \sum_{k=1}^n x_k s_k = Ax.$$

Dies zeigt die Behauptung. □

Sehen wir uns nun Beispiel 4.2.3(a) im Lichte von Matrizen noch einmal an:

Beispiel 4.2.9. Sei wieder $R : \mathbb{R}^2 \to \mathbb{R}^2$ durch

$$R(x) = \begin{bmatrix} x_1 - 2x_2 \\ x_1 \end{bmatrix}$$

für alle $x \in \mathbb{R}^2$ gegeben.[11] Aus der Definition der Matrixmultiplikation folgt dann sofort, dass

$$R(x) = \underbrace{\begin{bmatrix} 1 & -2 \\ 1 & 0 \end{bmatrix}}_{=:A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = Ax$$

für alle $x \in \mathbb{R}^2$ gilt. Somit ist R wegen Proposition 4.2.8(a) linear.[12]

Invertierbare Matrizen

Definition 4.2.10. (a) Sei $\mathbb{K}$ ein Körper und sei $m \in \mathbb{N}^*$. Dann bezeichnen wir mit $I_m \in \mathbb{K}^{m \times m}$ die Matrix

$$\begin{bmatrix} 1 & 0 & \dots & & 0 \\ 0 & 1 & & & \\ \vdots & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & & & 0 & 1 \end{bmatrix}$$

Wir nenne I_m die **Einheitsmatrix in** $\mathbb{K}^{m \times m}$ oder auch die m-**dimensionale Einheitsmatrix**.

[11] Beachten Sie, dass wir hier nun die neue Konvention benutzen und auch Vektoren aus $\mathbb{R}^2$ (und allgemeiner aus $\mathbb{K}^n$) wie Matrizen in eckige Klammern schreiben.

[12] Das wussten wir zwar genau genommen schon aus Beispiel 4.2.3(a) – dort hatten wir das aber mit einiger Mühe nachrechnen müssen, während wir nun einfach Proposition 4.2.8(a) verwenden können um die Linearität zu erhalten.

(b) Eine Matrix $A \in \mathbb{K}^{m\times m}$ heißt **invertierbar**, falls es eine Matrix $B \in \mathbb{K}^{m\times m}$ mit der Eigenschaft $AB = BA = I_m$ gibt.

Die Einheitsmatrix ist deshalb besonders wichtig, weil sie sich bzgl. der Multiplikation neutral verhält: Für jedes $C \in \mathbb{K}^{m\times n}$ und jedes $D \in \mathbb{K}^{\ell\times m}$ gilt[13]

$$I_m C = C \qquad \text{und} \qquad D I_m = D.$$

Wir listen einige Eigenschaften invertierbarer Matrizen auf:

Proposition 4.2.11. *Sei $\mathbb{K}$ ein Körper und sei $m \in \mathbb{N}^*$. Seien außerdem $A, \tilde{A} \in \mathbb{K}^{m\times m}$.*

(a) *Wenn A invertierbar ist, dann gibt es genau eine Matrix $B \in \mathbb{K}^{m\times m}$ mit der Eigenschaft $AB = BA = I_m$.*

*Wir nennen B dann die **inverse Matrix** von A, und bezeichnen sie mit A^{-1}.*

(b) *Wenn A und $\tilde{A}$ invertierbar sind, dann ist auch das Matrixprodukt $A\tilde{A}$ invertierbar, und es gilt $(A\tilde{A})^{-1} = \tilde{A}^{-1}A^{-1}$.*

(c) *Die Matrix A ist genau dann invertierbar, wenn A^{T} invertierbar ist, und in diesem Fall gilt $(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}$.*

(d) *Wenn A invertierbar ist, dann ist auch A^{-1} invertierbar, und es gilt $(A^{-1})^{-1} = A$.*

Beweis. (a) Seien $B, \tilde{B} \in \mathbb{K}^{m\times m}$ mit $AB = BA = I_m$ und $A\tilde{B} = \tilde{B}A = I_m$. Dann folgt

$$B = BI_m = B(A\tilde{B}) = (BA)\tilde{B} = I_m\tilde{B} = \tilde{B}.$$

(b) Es gilt

$$(\tilde{A}^{-1}A^{-1})(A\tilde{A}) = \tilde{A}^{-1}I_m\tilde{A} = I_m$$

und

$$(A\tilde{A})(\tilde{A}^{-1}A^{-1}) = AI_mA^{-1} = I_m.$$

Dies zeigt, dass $\tilde{A}^{-1}A^{-1}$ die inverse Matrix von $A\tilde{A}$ ist.

(c) Wir müssen eine Äquivalenz zeigen.
„⇒“ Sei A invertierbar. Dann gilt

$$A^T(A^{-1})^{\mathrm{T}} = (A^{-1}A)^{\mathrm{T}} = I_m^{\mathrm{T}} = I_m$$

[13] Warum diese beiden Gleichungen gelten, müssen Sie sich unbedingt im Detail überlegen. Das ist nicht schwer zu sehen, aber Sie können es nur verstehen, wenn Sie es sich bis ins letzte Detail selbst überlegt haben.

und

$$(A^{-1})^{\mathrm{T}} A^{\mathrm{T}} = (AA^{-1})^{\mathrm{T}} = I_m^{\mathrm{T}} = I_m.$$

Dies zeigt, dass A^{T} invertierbar ist, und dass ihre inverse Matrix gleich $(A^{-1})^{\mathrm{T}}$ ist.

„$\Leftarrow$“ Sei nun A^{T} invertierbar. Aufgrund der bereits gezeigten Implikation, angewendet auf die Matrix A^{T} anstelle von A, wissen wir, dass dann auch $(A^{\mathrm{T}})^{\mathrm{T}}$ invertierbar ist; letztere Matrix ist aber gleich A.

Die behauptete Formel $(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}$ wurde bereits am Ende des Beweises der Implikation „$\Rightarrow$“ gezeigt.[14]

(d) Wegen der Invertierbarkeit von A gilt $AA^{-1} = A^{-1}A = I_m$. Daraus folgt laut Definition der Invertierbarkeit von Matrizen, dass A^{-1} invertierbar ist, und laut Teil (a) dieser Proposition folgt hieraus zudem, dass die inverse Matrix von A^{-1} gleich A ist. □

4.3 Zeilenstufenform von Matrizen und das Gaußsche Eliminationsverfahren

Sie haben im vorangehenden Abschnitt unter anderem gelernt, was eine invertierbare Matrix ist. Um in der Praxis mit konkreten Matrizen arbeiten zu können, benötigt man Rechenverfahren, mit denen man überprüfen kann, ob eine gegebene Matrix invertierbar ist und mit denen man, falls ja, die inverse Matrix berechnen kann. Ein solches Rechenverfahren – den sogenannten **Gaußschen Eliminationsalgorithmus** – werden Sie in diesem Abschnitt kennenlernen. Wir benötigen zunächst einige Begriffe zur Vorbereitung.

Matrizen in Zeilenstufenform

Ziel des Gauß-Algorithmus ist es ganz allgemein gesprochen, eine Matrix in eine andere Matrix umzuformen, die eine bestimmte Gestalt hat.[15] Diese bestimme Gestalt von Matrizen wird im letzten Teil der nachfolgende Definition beschrieben:

Definition 4.3.1 ((Reduzierte) Zeilenstufenform)**.** Sei $\mathbb{K}$ ein Körper und seien $m, n \in \mathbb{N}^*$. Sei $A \in \mathbb{K}^{m \times n}$.

(a) Sei $k \in \{1, \ldots, n\}$. Wir definieren die **Treppentiefe** von A in der Spalte k als die kleinste Zahl $j \in \{0, \ldots, m\}$ mit der Eigenschaft, dass in den ersten k Spalten von A in den Zeilen $j+1, \ldots, m$ nur Nullen stehen.

(b) Sei $k \in \{1, \ldots, n\}$. Wir definieren die **Stufentiefe** von A in der k-ten Spalte folgendermaßen:

[14] Vorsicht! Wieso genau genügt es, wenn wir die behauptete Formel in einer der beiden Implikationen zeigen? Müssten wir Sie nicht in der anderen Implikation auch noch zeigen?

[15] Und Sie werden nachher noch sehen, dass das Invertieren einer Matrix ein Spezialfall hiervon ist.

- $k = 1$: Wir definieren die Stufentiefe in der ersten Spalte von A als die Treppentiefe von A in der ersten Spalte.
- $k \geq 2$: Wir setzen die Stufentiefe in der k-ten Spalte von A gleich der Treppentiefe von A in der k-ten Spalte minus der Treppentiefe von A in der $(k-1)$-ten Spalte.

(c) Die Matrix A heißt **in Zeilenstufenform**, wenn sie in jeder Spalte die Stufentiefe 0 oder 1 besitzt.

(d) Die Matrix A heißt **in reduzierter Zeilenstufenform**, wenn sie in Zeilenstufenform ist, und wenn zusätzlich folgendes gilt:

In jeder Spalte, in der A Stufentiefe 1 hat gilt: Wenn t die Treppentiefe dieser Spalte bezeichnet, steht im t-ten Eintrag der Spalte eine 1, und ansonsten stehen in der Spalte nur Nullen.[16]

Wir möchten nun eine gegebene Matrix A durch bestimmte Umformungen zu einer anderen Matrix „umbauen“, welche sich in der reduzierter Zeilenstufenform befindet. Alle Umformungen, die wir dabei verwenden werden, sind bestimmte Manipulationen der Zeilen der Matrix – und zwar folgende:

Definition 4.3.2 (Elementare Zeilenumformungen)**.** Sei $\mathbb{K}$ ein Körper und seien $m, n \in \mathbb{N}^*$. Sei $A \in \mathbb{K}^{m \times n}$. Unter einer **elementaren Zeilenumformung von A** versteht man einer der folgenden Veränderungen von A:

- Das Vertauschen von zwei Zeilen von A.
- Das Multiplizieren einer Zeile von A mit einem Skalar $\alpha \in \mathbb{K}^*$.
- Das Addieren des β-fachen einer Zeile von A (für ein $\beta \in \mathbb{K}$) zu einer anderen Zeile von A.

Elementare Zeilenumformungen kann man auch mit Hilfe der Multiplikation geeigneter Matrizen von links beschreiben. Dies wird in der folgenden Diskussion näher erläutert:

Diskussion 4.3.3. Sei $\mathbb{K}$ ein Körper und seien $m, n \in \mathbb{N}^*$. Sei $A \in \mathbb{K}^{m \times n}$.

(a) Zum Beispiel das Vertauschen der ersten und zweiten Zeile von A kann erreicht werden, indem man A von links mit der Matrix

$$E = \begin{bmatrix} 0 & 1 & & & \\ 1 & 0 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \in \mathbb{K}^{m \times m}$$

multipliziert (und analog für andere Zeilen- und Spaltennummern).

[16] Anders ausgedrückt: Die Spalte ist gleich dem kanonischen Einheitsvektor e_t.

(b) Zum Beispiel das Multiplizieren der zweiten Zeile von A mit einem Skalar $\alpha \in \mathbb{K}^*$ kann erreicht werden, indem man A von links mit der Matrix

$$E = \begin{bmatrix} 1 & & & & \\ & \alpha & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \in \mathbb{K}^{m \times m}$$

multipliziert (und analog für andere Zeilen).

(c) Zum Beispiel die Addition des β-fachen der dritten Zeile von A zur ersten Zeile (für ein $\beta \in \mathbb{K}$) kann erreicht werden, indem man A von links mit der Matrix

$$E = \begin{bmatrix} 1 & & \beta & & \\ & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \in \mathbb{K}^{m \times m}$$

multipliziert (und analog für andere Zeilen).

Es ist wichtig zu erkennen, dass die Matrix E in jedem der drei Fälle invertierbar ist.[17]

Nun können wir besprechen, wie man eine Matrix durch elementare Zeilenumformungen in eine andere Matrix umformen kann, die sich in Zeilenstufenform befindet.

Theorem 4.3.4. *Sei $\mathbb{K}$ ein Körper und seien $m, n, p \in \mathbb{N}^*$. Sei $A \in \mathbb{K}^{m \times n}$.*

(a) *Es gibt es genau eine Matrix Z in reduzierter Zeilenstufenform mit der Eigenschaft, dass man A durch elementare Zeilenumformungen zu Z umformen kann.*[18]

Die so erhaltene Matrix Z heißt **die reduzierte Zeilenstufenform von** A.

(b) *Sei $\tilde{A} \in \mathbb{K}^{m \times p}$. Bezeichne $[Z, \tilde{Z}] \in \mathbb{K}^{m \times (n+p)}$ (mit $Z \in \mathbb{K}^{m \times n}$ und $\tilde{Z} \in \mathbb{K}^{m \times p}$) die reduzierte Zeilenstufenform von $[A, \tilde{A}] \in \mathbb{K}^{m \times (n+p)}$. Dann ist Z die reduzierte Zeilenstufenform von A.*

(c) *Bezeichne $[Z, T] \in \mathbb{K}^{m \times (n+m)}$ (mit $Z \in \mathbb{K}^{m \times n}$ und $T \in \mathbb{K}^{m \times m}$) die reduzierte Zeilenstufenform von $[A, I_m] \in \mathbb{K}^{m \times (n+m)}$. Dann ist Z die reduzierte Zeilenstufenform von A, es gilt $TA = Z$, und T ist invertierbar.*

[17] Um sicherzustellen, ob Sie das verstanden haben, sollten Sie sich fragen, ob Sie in jedem der Fälle angeben können, wie die inverse Matrix von E aussieht.

[18] Beachten Sie aber, dass die elementaren Zeilenumformungen, mit denen das möglich ist, im Allgemeinen nicht eindeutig bestimmt sind.

Beweis. (a) *Existenz:* Die Existenz kann man konstruktiv sehen, indem man sich ein Verfahren überlegt, welche A durch elementare Zeilenumformungen in eine Matrix in reduzierter Zeilenstufenform umformt. Dieses Verfahren wird in Algorithmus 4.3.5 weiter unten im Manuskript beschrieben.

Eindeutigkeit: Weil die Eindeutigkeit für die nachfolgenden Argumente in der Vorlesung nur eine untergeordnete Rolle spielt, verzichten wir an dieser Stelle auf ihren Beweis und verweisen stattdessen auf die Literatur – zum Beispiel auf [Mey00, Seiten 134–135].

(b) Wenn $[Z, \tilde{Z}]$ sich in reduzierter Zeilenstufenform befindet, dann gilt dies auch für Z. Zudem gilt: Diejenigen elementaren Zeilemumformungen, die $[A, \tilde{A}]$ in $[Z, \tilde{Z}]$ umformen, formen natürlich auch A in Z um.

(c) Wie in Diskussion 4.3.3 besprochen, gibt es eine Zahl $s \in \mathbb{N}^*$ und invertierbare Matrizen $E_1, \ldots, E_s$ derart, dass

$$E_s \cdots E_1 [A, I_m] = [Z, T]$$

gilt. Die Matrix auf der linken Seite ist gleich

$$[E_s \cdots E_1 A,\ E_s \cdots E_1],$$

woraus $T = E_s \cdots E_1$ und $Z = E_s \cdots E_1 A = TA$ folgt. Insbesondere ist T als Produkt invertierbarer Matrizen invertierbar (siehe Proposition 4.2.11(b)). □

Beachten Sie: Das Theorem zeigt insbesondere, dass man jede Matrix durch elementare Zeilenumformungen in eine Zeilenstufenform bringen kann; letztere muss aber nicht eindeutig sein, sofern man nicht zusätzlich verlangt, dass die Zeilenstufenform reduziert ist.

Das Gaußsche Eliminationsverfahren

Im Beweis von Theorem 4.3.4(a) wurde versprochen, dass man die Existenz der reduzierten Zeilenstufenform konstruktiv sehen kann, indem man ein Verfahren angibt, mit dem sich jede beliebige Matrix durch elementare Zeilenumformungen auf Zeilenstufenform bringen lässt.

Bei diesem Verfahren handelt es sich um den sogenannten **Gaußschen Eliminationsalgorithmus**, den wir im folgenden Pseudocode-Listing exakt beschreiben:

Algorithmus 4.3.5 (Gaußsches Eliminationsverfahren)**.** Sei $\mathbb{K}$ ein Körper und seien $m, n \in \mathbb{N}^*$. Sei $A \in \mathbb{K}^{m \times n}$. Das folgende Verfahren bringt A durch elementare Zeilenumformungen auf reduzierte Zeilenstufenform:

```
   Eingabe: Matrix A
   Ausgabe: Reduzierte Zeilenstufenform Z von A
 1 Setze Z := A;
 2 Setze t := 0; % Treppentiefe der „nullten“ Spalte.
 3 für jedes k ∈ {1,...,n} tue
 4 |   Setze j_0 := t;
 5 |   für jedes j ∈ {t+1,...,m} tue
 6 |   |   wenn Z(j,k) ≠ 0 dann
 7 |   |   |   Setze j_0 = j;
 8 |   |   |   Breche die Schleife dann ab;
 9 |   |   Ende
10 |   Ende
   |   % Falls j_0 nun gleich t ist, ist die Stufentiefe in der
   |     aktuellen Spalte gleich 0. Gehe dann zur naechsten Spalte.
   |   % Falls aber j_0 ≥ t+1 ist, ist die Stufentiefe in der
   |     aktuellen Spalte mindestens 1. In diesem Fall aendern wir
   |     die Matrix nun so ab, dass die Stufentiefe zu 1 wird:
11 |   wenn j_0 ≥ t+1 dann
12 |   |   Setze t := t+1; % Zukuenftigen Wert der Treppentiefe schon
   |   |     mal speichern.
13 |   |   Vertausche die Zeilen t und j_0 von Z; % Somit ist nun Z_tk ≠ 0.
14 |   |   Teile die t-te Zeile von Z durch Z_tk; % Somit ist nun Z_tk = 1.
15 |   |   für jedes j ∈ {1,...,m} \ {t} tue
16 |   |   |   Addiere ein Vielfaches der t-ten Zeile von Z zur j-ten Zeile von Z
   |   |   |    um den Wert Z_jk zu 0 zu ändern;
17 |   |   Ende
18 |   Ende
19 Ende
```

Der Zusammenhang zwischen Zeilenstufenform und Invertierung

Lassen Sie uns nun besprechen, wie die reduzierte Zeilenstufenform einer Matrix verwendet werden kann, um eine Matrix zu invertieren.

Definition 4.3.6. Sei $\mathbb{K}$ ein Körper und $m \in \mathbb{N}^*$. Sei $A \in \mathbb{K}^{m\times m}$.

(a) Eine Matrix $A \in \mathbb{K}^{m\times m}$ heißt **rechtsinvertierbar**, falls es ein $B \in \mathbb{K}^{m\times m}$ mit der Eigenschaft $AB = I_m$ gibt.

(b) Eine Matrix $A \in \mathbb{K}^{m\times m}$ heißt **linksinvertierbar**, falls es ein $B \in \mathbb{K}^{m\times m}$ mit der Eigenschaft $BA = I_m$ gibt.

Es stellt sich heraus, dass eine Matrix, die linksinvertierbar (oder rechtsinvertierbar) ist, automatisch bereits invertierbar ist. Zudem kann man Invertierbarkeit

einer Matrix mit Hilfe ihrer reduzierten Zeilenstufenform charakterisieren. All diese Informationen sind Teil des nächsten Theorems:

Theorem 4.3.7. *Sei $\mathbb{K}$ ein Körper und $m \in \mathbb{N}^*$. Außerdem sei $A \in \mathbb{K}^{m\times m}$. Dann sind folgende Aussagen äquivalent:*

(i) *Die Matrix A ist invertierbar.*

(ii) *Die reduzierte Zeilenstufenform von A ist gleich der Einheitsmatrix I_m.*

(iii) *Die Matrix A ist rechtsinvertierbar.*

(iv) *Die Matrix A ist linksinvertierbar.*

Falls eine (gleichbedeutend damit: jede) dieser Aussagen erfüllt ist, dann gilt zudem: Wenn $[Z\ T]$ die reduzierte Zeilenstufenform von $[A\ I_m]$ bezeichnet, dann ist T die inverse Matrix von A.

Beweis. „(i) $\Rightarrow$ (iii)“ Klar nach Definition.

„(iii) $\Rightarrow$ (ii)“ Sei $Z \in \mathbb{K}^{m\times m}$ die reduzierte Zeilenstufenform von A. Gelte $\neg$(ii), d.h. $Z \neq I_m$. Dann besitzt Z eine Spalte mit Stufentiefe 0. Somit ist die letzte Zeile von Z gleich 0. Es gibt eine invertierbare Matrix $T \in \mathbb{K}^{m\times m}$ mit $TA = Z$ (Diskussion 4.3.3). Wegen (iii) gibt es ein $B \in \mathbb{K}^{m\times m}$ mit $AB = I_m$. Somit folgt

$$T = TAB = \underbrace{ZB}_{\text{letzte Zeile ist 0}}.$$

Also ist die letzte Zeile von T gleich 0. Aber T ist ein Produkt von Matrizen aus Diskussion 4.3.3, und damit kann die letzte Zeile von T nicht 0 sein. Widerspruch!

„(ii) $\Rightarrow$ (i)“ Die reduzierte Zeilenstufenform von $[A\ I_m]$ ist von der Form $[I_m\ T]$ (Theorem 4.3.4 (c)) und $TA = I_m$ mit invertierbarem T.
Also gilt $A = T^{-1}$ und damit ist A invertierbar.

„(i) $\Leftrightarrow$ (iv)“ Diese Äquivalenz zeigt man durch Verwendung transponierter Matrizen. □

Lineare Gleichungssysteme

Zum Schluss dieses Abschnitts diskutieren wir noch, wie die reduzierte Zeilenstufenform mit der Lösung linearer Gleichungssysteme zusammenhängt.

Diskussion 4.3.8. Sei $\mathbb{K}$ ein Körper, sei $m \in \mathbb{N}^*$. Sei $A \in \mathbb{K}^{m\times m}$ und $b \in \mathbb{K}^m$.

Das Gleichungssystem

$$\begin{array}{ccccc} A_{11}x_1+ & \dots & +A_{1m}x_m & = & b_1 \\ & \vdots & & \vdots & \vdots \\ A_{m1}x_1+ & \dots & +A_{1m}x_m & = & b_m \end{array}$$

wobei $x \in \mathbb{K}^m$ gesucht ist, ist äquivalent zur Gleichung

$$Ax = b.$$

Dies folgt direkt aus der Definition des Matrixprodukts.

Die eindeutige Lösbarkeit von linearen Gleichungssystemen für alle rechten Seiten lässt sich dadurch charakterisieren, dass die Matrix, welche die Koeffizienten des Gleichungssystems enthält, invertierbar ist:

Theorem 4.3.9. *Sei $\mathbb{K}$ ein Körper, seien $m \in \mathbb{N}^*$ und sei $A \in \mathbb{K}^{m\times m}$. Das sind folgende Aussagen äquivalent:*

(i) *Für jedes $b \in \mathbb{K}^m$ gibt es genau ein $x \in \mathbb{K}^m$, welches die Gleichung $Ax = b$ löst.*

(ii) *Die Matrix A ist invertierbar.*

Beweis. „(i) $\Rightarrow$ (ii)“ Gelte (i). Für jedes $k \in \{1, \dots, m\}$ gibt es ein $c_k \in \mathbb{K}^m$ mit

$$Ac_k = e_k.$$

Setze $C \coloneqq [c_1 \ c_2 \ \dots \ c_m] \in \mathbb{K}^{m\times m}$. Dann gilt

$$AC = [Ac_1 \ Ac_2 \ \dots \ Ac_m] = [e_1 \ e_2 \ \dots \ e_m] = I_m.$$

Also ist A rechtsinvertierbar und somit invertierbar.

„(ii) $\Rightarrow$ (i)“ Sei A invertierbar. Sei $b \in \mathbb{K}^m$ beliebig aber fest.
Existenz: Setze $x = A^{-1}b \in \mathbb{K}^m$. Dann ist

$$Ax = AA^{-1}b = I_m b = b$$

Eindeutigkeit: Seien $x, \tilde{x} \in \mathbb{K}^m$ mit $Ax = b$ und $A\tilde{x} = b$. Dann ist

$$Ax = A\tilde{x}, \text{ also } A^{-1}Ax = A^{-1}A\tilde{x}$$

und somit $x = \tilde{x}$. □

Wenn A invertierbar ist, kann man die Lösung der linearen Gleichung $Ax = b$ durch die Formel $x = A^{-1}b$ berechnen – und Sie wissen bereits, wie man A^{-1} mit Hilfe des Gaußschen Eliminiationsverfahrens bestimmen kann.

Man kann allerdings die Lösung x von $Ax = b$ auch direkt mit Hilfe des Gaußschen Eliminationsverfahrens berechnen. Dies ist der Inhalt des folgenden Korollars:

Korollar 4.3.10. *Sei $\mathbb{K}$ ein Körper, seien $m \in \mathbb{N}^*$, sei $A \in \mathbb{K}^{m\times m}$ invertierbar und sei $b \in \mathbb{K}^m$.*

Wenn $[I_m \ x] \in \mathbb{K}^{m\times(m+1)}$ (für ein $x \in \mathbb{K}^m$) die reduzierte Zeilenstufenform von $[A \ b] \in \mathbb{K}^{m\times(m+1)}$ ist, dann gilt $Ax = b$ (d.h. $x = A^{-1}b$).

Beweis. Die reduzierte Zeilenstufenform von $[A\ b\ I_m] \in \mathbb{K}^{m\times(2m+1)}$ hat die Form

$$[I_m\ x\ T] \in \mathbb{K}^{m\times(2m+1)}$$

mit $T \in \mathbb{K}^{m\times m}$ (Theorem 4.3.4(c)), und es gilt

$$T[A\ b] = [I_m\ x].$$

Also ist $TA = I_m$ und $Tb = x$, also $b = T^{-1}x = Ax$. □

Literaturhinweise

- Ein Einführung in Vektorräume und Untervektorräume finden Sie in den meisten Büchern über Lineare Algebra, die im Literaturverzeichnis aufgelistet sind. Manche Lehrbücher verzichten in einer Einführung in die Lineare Algebra auch darauf, über abstrakte Vektorräume zu sprechen und behandeln stattdessen zum Beispiel nur Untervektorräume von $\mathbb{R}^d$ und $\mathbb{C}^d$.

  Solche sehr konkreten Einführungen mit wenig abstrakter Theorie sind im deutschsprachigen Raum aber eher unüblich.

- Die Theorie von Matrizen spielt in konkreten Anwendungen eine große Rolle, weil man mit Ihnen explizit rechnen kann – insbesondere und gerade auch auf Computern. Zusätzlich zur Einführungsliteratur in Lineare Algebra gibt es deshalb insbesondere auch Bücher, die sich aus algorithmischer und numerischer Sicht mit Matrizen auseinandersetzen.

  Eines der bekanntesten dieser Bücher ist [GVL13]. Es beschreibt zahlreiche wichtige numerische Verfahren im Umgang mit Matrizen und ist eine der wichtigsten Standardreferenzen in der numerischen Linearen Algebra.

Kapitel 5

Darstellung von Vektoren und linearen Abbildungen

Einstiegsfragen. (a) Wie schafft es ein Computer mit Vektoren aus $\mathbb{R}^d$ zu rechnen?

Betrachten Sie nun einen beliebigen anderen Vektorraum V über $\mathbb{R}$. Können Sie einem Computer auch beibringen mit Elementen aus V zu rechnen?

(b) Welche Möglichkeiten fallen Ihnen ein um eine Ursprungsebene im $\mathbb{R}^3$ zu spezifizieren?

Welche Möglichkeiten fallen Ihnen ein um einen Untervektorraum des $\mathbb{R}^d$ zu spezifizieren?

(c) Sei E eine Ursprungsebene im $\mathbb{R}^3$, die außerdem durch die folgenden beiden Punkte verläuft:

$$\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Können Sie alle Elemente von E explizit beschreiben, z.B. mit Hilfe einer Formel?

(d) Was verstehen Sie unter dem Begriff „Dimension“?

(e) Ist es sinnvoll von einem 8-dimensionalen Raum zu sprechen? Und von einem 274-dimensionalen Raum?

Falls nein: Warum ergibt es keinen Sinn, über solche Räume zu sprechen? Falls ja: Warum sollte man sich für Räume von so hoher Dimension interessieren?

Und was ist in den vier vorangehenden Fragen überhaupt mit dem Wort „Raum“ gemeint?

5.1 Darstellung von Vektoren mittels Basen

Existenz von Darstellungen: Erzeugendensysteme und Aufspann

In diesem Abschnitt und im nachfolgenden Abschnitt wollen wir besprechen, wie man alle Vektoren in einem Vektorraum mit Hilfe von nur wenigen Vektoren darstellen kann. Um dies präzise zu machen, benötigen wir zunächst einiges an Terminologie.

Definition 5.1.1 (Linearkombinationen, Aufspann und Erzeugendensysteme)**.** Sei V ein Vektorraum über einem Körper $\mathbb{K}$ und sei $M \subseteq V$. Außerdem sei $n \in \mathbb{N}$ und es seien $v_1, \dots, v_n \in V$.

(a) Eine **Linearkombination** des Tupels $(v_1, \dots, v_n)$ ist ein Vektor der Form

$$\sum_{k=1}^{n} \alpha_k v_k,$$

wobei $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ sind.

(b) Der **Aufspann**[1] des Tupels $(v_1, \dots, v_n)$ ist definiert als die Menge

$$\operatorname{span}(v_1, \dots, v_n) \coloneqq \left\{ \sum_{k=1}^{n} \alpha_k v_k \mid \ \alpha_1, \dots, \alpha_n \in \mathbb{K} \right\},$$

und der Aufspann der Menge M ist definiert als die Menge

$$\operatorname{span} M \coloneqq \left\{ \sum_{k=1}^{m} \alpha_k w_k \mid \ m \in \mathbb{N},\ \alpha_1, \dots, \alpha_m \in \mathbb{K},\ w_1, \dots, w_m \in M \right\}.$$

(c) Das Tupel $(v_1, \dots, v_n)$ heißt ein **Erzeugendensystem** von V, falls

$$\operatorname{span}(v_1, \dots, v_n) = V$$

gilt.

Aus der Definition des Spanns erhält man leicht, dass für Vektoren $v_1, \dots, v_n \in V$ stets

$$\operatorname{span}(v_1, \dots, v_n) = \operatorname{span}\{v_1, \dots, v_n\}$$

gilt. In der folgenden Proposition zeigen wir, dass der Aufspann immer ein Untervektorraum ist; zudem geben wir eine bestimmte Darstellung des Spann einer Menge an.

[1] Oft sagt man anstelle von **Aufspann** auch **Spann** oder **Lineare Hülle**.

Proposition 5.1.2. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$ und sei $M \subseteq V$. Dann ist* $\operatorname{span} M$ *ein Untervektorraum von V, und es gilt*

$$M \quad \subseteq \quad \operatorname{span} M \quad = \bigcap_{\substack{U \text{ ist UVR von } V, \\ M \subseteq U}} U.$$

Beweis. Man kann direkt nachprüfen, dass $\operatorname{span} M$ die Untervektorraum-Axiome erfüllt. Somit bleibt noch die behauptete Inklusion und die behauptete Mengengleichheit zu zeigen. Zunächst zeigen wir die Inklusion $M \subseteq \operatorname{span} M$:

Sei $v \in M$. Dann ist $v = \sum_{k=1}^{1} \alpha_1 v_1$ mit $\alpha_1 \coloneqq 1 \in \mathbb{K}$ und $v_1 \coloneqq v \in M$, also $v \in \operatorname{span} M$.

Nun zeigen wir die behauptete Mengengleichheit:

„$\subseteq$“ Sei $v \in \operatorname{span} M$. Dann gibt es ein $m \in \mathbb{N}$ sowie Skalare $\alpha_1, \ldots, \alpha_m \in \mathbb{K}$ und Vektoren $v_1, \ldots, v_m \in M$ mit $v = \sum_{k=1}^{m} \alpha_k v_k$.

Jeder Untervektorraum U von V, der M als Teilmenge hat, enthält die Vektoren $v_1, \ldots, v_m$ und somit, aufgrund der Untervektorraum-Axiome auch v. Also ist

$$v \in \bigcap_{\substack{U \text{ ist UVR von } V, \\ M \subseteq U}} U.$$

„$\supseteq$“ Sei nun v im angegebenen Durchschnitt. Wie oben bereits bemerkt ist $\operatorname{span} M$ selbst ein Untervektorraum, und wie ebenfalls bereits bewiesen, enthält er M. Also ist $\operatorname{span} M$ selbst einer der Untervektorräume, über die geschnitten wird. Weil v in jedem dieser Untervektorräume liegt, gilt insbesondere $v \in \operatorname{span} M$. □

Wenn ein Tupel $(v_1, \ldots v_n)$ ein Erzeugendensystem eines Vektorraums V ist, dann bedeutet dies laut Definition des Begriffs Erzeugendensystem, dass man jeden Vektor aus V als Linearkombination des Tupels $(v_1, \ldots, v_n)$ darstellen kann. Im nächsten Unterabschnitt werden wir eine dazu komplementäre Eigenschaft betrachten: nämlich die Frage, ob jeder Vektor aus V höchstens eine Darstellung als Linearkombination des Tupels $(v_1, \ldots, v_n)$ besitzt.

Eindeutigkeit von Darstellungen: Lineare Unabhängigkeit

Definition 5.1.3 (Lineare Unabhängigkeit und Abhängigkeit)**.** Sei V ein Vektorraum über einen Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \ldots, v_n \in V$.

(a) Das Tupel $(v_1, \ldots, v_n)$ heißt **linear unabhängig**, falls es für jeden Vektor $v \in V$ höchstens ein Tupel $\alpha \in \mathbb{K}^n$ mit der Eigenschaft

$$v = \sum_{k=1}^{n} \alpha_k v_k$$

gibt.

(b) Das Tupel $(v_1, \ldots, v_n)$ heißt **linear abhängig**, falls es nicht linear unabhängig ist.

Wir zeigen in der folgenden Proposition, dass sich lineare Unabhängigkeit auf verschiedene Weise charakterisieren lässt:

Proposition 5.1.4. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \ldots, v_n \in V$. Dann sind folgende Aussagen äquivalent:*

(i) *Das Tupel $(v_1, \ldots, v_n)$ ist linear unabhängig.*

(ii) *Für jeden Vektor $v \in \operatorname{span}(v_1, \ldots, v_n)$ gibt es genau ein $\alpha \in \mathbb{K}^n$ mit der Eigenschaft*

$$v = \sum_{k=1}^{n} \alpha_k v_k.$$

(iii) *Für alle $\alpha \in \mathbb{K}^n$ gilt die folgende Implikation:*

$$\sum_{k=1}^{n} \alpha_k v_k = 0 \qquad \Rightarrow \qquad \alpha = 0.$$

Beweis. „(i) $\Leftrightarrow$ (ii)“ Diese Äquivalenz folgt unmittelbar aus der Definition der linearen Unabhängigkeit und des Aufspanns.

„(i) $\Rightarrow$ (iii)“ Es gelte (i). Sei $\alpha \in \mathbb{K}^n$ beliebig, aber fest. Wir müssen die in (iii) behauptete Implikation zeigen, also sei

$$\sum_{k=1}^{n} \alpha_k v_k = 0.$$

Dann gilt

$$\sum_{k=1}^{n} \alpha_k v_k = \sum_{k=1}^{n} 0 \cdot v_k.$$

Wegen der Definition der linearen Unabhängigkeit folgt daraus $\alpha_1 = 0, \ldots, \alpha_n = 0$, also ist $\alpha = 0$.

„(iii) $\Rightarrow$ (i)“ Gelte (iii). Wir müssen die lineare Unabhängigkeit von $(v_1, \ldots, v_n)$ zeigen, also seien $\alpha, \tilde{\alpha} \in \mathbb{K}^n$ mit

$$\sum_{k=1}^{n} \alpha_k v_k = \sum_{k=1}^{n} \tilde{\alpha}_k v_k.$$

Dann folgt

$$\sum_{k=1}^{n} (\alpha_k - \tilde{\alpha}_k) v_k = 0.$$

Wegen (iii) folgt hieraus $\alpha_1 - \tilde{\alpha}_1 = 0, \ldots \alpha_n - \tilde{\alpha}_n = 0$ und somit $\alpha = \tilde{\alpha}$. □

Wenn man lineare Unabhängigkeit für ein konkret gegebenes Tupel von Vektoren überprüfen will, ist es oft im einfachsten, die Aussage (iii) aus Proposition 5.1.4(iii) zu überprüfen. Mit Hilfe dieser Aussage können wir außerdem auch ein Verfahren angeben um zu überprüfen, ob ein Tupel von Vektoren aus $\mathbb{K}^m$ linear unabhängig ist:

Korollar 5.1.5. *Sei $\mathbb{K}$ ein Körper und seien $m, n \in \mathbb{N}^*$. Zudem seien $v_1, \ldots, v_n \in \mathbb{K}^m$.*

Das Tupel $(v_1, \ldots, v_n)$ ist genau dann linear unabhängig, wenn jede Spalte in der reduzierten Zeilenstufenform der Matrix $[v_1 \ \ldots \ v_n] \in \mathbb{K}^{m\times n}$ die Stufentiefe 1 besitzt.

Beweis. Sei $Z \in \mathbb{K}^{m\times n}$ die reduzierte Zeilenstufenform von $[v_1 \ \ldots \ v_n]$. Dann gibt es eine invertierbare Matrix $T \in \mathbb{K}^{m\times m}$ mit der Eigenschaft, dass $T[v_1 \ \ldots \ v_n] = Z$. (Theorem 4.3.4 (c))

„⇒“ Sei $(v_1, \ldots, v_n)$ linear unabhängig.
Angenommen, in der k-ten Spalte von Z für ein $k \in \{1, \ldots, n\}$ wäre die Stufentiefe gleich 0. Dann gibt es ein $\alpha \in \mathbb{K}^n$ mit $\alpha \neq 0$ und $Z\alpha = 0$.Um dies zu sehen, bezeichnen wir die Spalten von Z mit $Z_1, \ldots, Z_n$. Dann gibt es Skalare $\alpha_1, \ldots, \alpha_{k-1}$ mit $Z_k = \alpha_1 Z_1 + \cdots + \alpha_{k-1} Z_{k-1}$. Wenn wir $\alpha_k = -1$ und $\alpha_{k+1} = \alpha_{k+2} = \cdots = \alpha_n = 0$ setzen, dann gilt

$$Z\alpha = \sum_{k=1}^{n} \alpha_k Z_k = 0.$$

Hieraus folgt

$$\sum_{k=1}^{n} \alpha_k v_k = [v_1 \ \ldots \ v_n]\alpha = T^{-1} Z\alpha = 0.$$

Also ist $(v_1, \ldots, v_n)$ linear abhängig nach Proposition 5.1.4 (iii)

„⇐“ Zu zeigen ist die lineare Unabhängigkeit von $(v_1 \ \ldots \ v_n)$. Es genügt (iii) aus Proposition 5.1.4 zu zeigen. Sei $\alpha \in \mathbb{K}^n$ beliebig aber fest und gelte

$$\sum_{k=1}^{n} \alpha_k v_k = 0.$$

Es ist somit

$$0 = \sum_{k=1}^{n} \alpha_k v_k = [v_1 \ \ldots \ v_n]\alpha.$$

Hieraus folgt

$$0 = T0 = T[v_1 \ \ldots \ v_n]\alpha = Z\alpha = \begin{bmatrix} \alpha \\ 0 \end{bmatrix} \in \mathbb{K}^m.$$

Insbesondere folgt $\alpha = 0$. □

Ähnlich wie in Proposition 5.1.4 können wir auch eine Charakterisierung linearer Abhängigkeit angeben:

Proposition 5.1.6. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \dots, v_n \in V$. Dann sind folgende Aussagen äquivalent:*

(i) *Das Tupel $(v_1, \dots, v_n)$ ist linear abhängig.*

(ii) *Es gibt es $k \in \{1, \dots, n\}$ derart, dass*

$$\operatorname{span}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_n) = \operatorname{span}(v_1, \dots, v_n)$$

gilt.[2]

(iii) *Es gibt es $k \in \{1, \dots, n\}$ derart, dass*

$$v_k \in \operatorname{span}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_n)$$

gilt.

Beweis. „(i) $\Rightarrow$ (iii)“ Seien $(v_1, \dots, v_n)$ linear abhängig. Nach Proposition 5.1.4 (iii) $\exists \alpha \in \mathbb{K}^n$ mit $\alpha \neq 0$ und

$$\sum_{k=1}^{n} \alpha_k v_k = 0.$$

Wegen $\alpha \neq 0$, gibt es ein ein $k_0 \in \{1, \dots, n\}$ mit $\alpha_{k_0} \neq 0$. Damit gilt

$$\alpha_{k_0} v_{k_0} = \sum_{\substack{k=1 \\ k \neq k_0}}^{n} \alpha_k v_k,$$

d.h.

$$v_{k_0} = \sum_{\substack{k=1 \\ k \neq k_0}}^{n} -\frac{\alpha_k}{\alpha_{k_0}} v_k \in \operatorname{span}(v_1, \dots, v_{k_0-1}, v_{k_0+1}, \dots, v_n)$$

Also gilt (iii).

„(iii) $\Rightarrow$ (ii)“ Gelte (iii). Sei k wie in (iii). Dann ist

$$v_k \in \operatorname{span}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_n)$$

also gibt es $\alpha_1, \dots, \alpha_{k-1}, \alpha_{k+1}, \dots, \alpha_n$ mit

$$v_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} \alpha_j v_j$$

[2] D.h. also, dass der Aufspann von $(v_1, \dots, v_n)$ sich nicht verändert, wenn man v_k aus dem Tupel entfernt.

Wir zeigen eine Mengengleichheit.
„$\subseteq$“ Klar.
„$\supseteq$“ Sei

$$v \in \operatorname{span}(v_1, \ldots, v_n)$$

Dann gibt es $\beta_1, \ldots, \beta_n$ mit

$$v = \sum_{j=1}^{n} \beta_j v_j.$$

Somit folgt

$$v = \sum_{\substack{j=1 \\ j \neq k}}^{n} \beta_j v_j + \beta_k v_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} \beta_j v_j + \sum_{\substack{j=1 \\ j \neq k}}^{n} \beta_k \alpha_j v_j$$

$$= \sum_{\substack{j=1 \\ j \neq k}}^{n} (\beta_j - \beta_k \alpha_j) v_j \in \operatorname{span}(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n).$$

Also gilt (ii).
„(ii) $\Rightarrow$ (i)“ Gelte (ii). Dann gilt für das k aus (ii).

$$v_k \in \operatorname{span}(v_1, \ldots, v_n) \overset{(ii)}{=} \operatorname{span}(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n),$$

also gibt es $\alpha_1, \ldots, \alpha_{k-1}, \alpha_{k+1}, \ldots, \alpha_n$ mit der Eigenschaft

$$v_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} \alpha_j v_j.$$

Mit $\alpha_k \coloneqq -1$ folgt

$$0 = \sum_{j=1}^{n} \alpha_j v_j,$$

aber $\alpha \neq 0$ wegen $\alpha_k = -1 \neq 0$. Also ist $(v_1, \ldots, v_n)$ wegen Proposition 5.1.4 (iii) linear abhängig. □

Basen

Nun kommen wir zum entscheidenden Begriff, auf den die zuvor eingeführten Begriffe hingeführt haben:

Definition 5.1.7 (Basis)**.** Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \dots, v_n \in V$. Das Tupel $(v_1, \dots, v_n)$ heißt **Basis** von V, falls es für jedes $v \in V$ genau ein $\alpha \in \mathbb{K}^n$ mit der Eigenschaft

$$v = \sum_{k=1}^{n} \alpha_k v_k$$

gibt.

Ob ein gegebenes Tupel eine Basis ist, lässt sich mit Hilfe verschiedener Eigenschaften charakterisieren:

Proposition 5.1.8. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \dots, v_n \in V$. Dann sind folgende Aussagen äquivalent:*

(i) *Das Tupel $(v_1, \dots, v_n)$ ist eine Basis von V.*

(ii) *Das Tupel $(v_1, \dots, v_n)$ ist linear unabhängig und ein Erzeugendensystem von V.*

(iii) *Das Tupel $(v_1, \dots, v_n)$ ist ein maximales linear unabhängiges Tupel in V in dem Sinne, dass das Hinzufügen eines beliebigen Vektors aus V zum Tupel die lineare Unabhängigkeit zerstört.*

(iv) *Das Tupel $(v_1, \dots, v_n)$ ist ein minimales Erzeugendensystem von V in dem Sinne, dass das Wegnehmen irgendeines Vektors aus dem Tupel dafür sorgt, dass das verbleibende Tupel kein Erzeugendensystem mehr ist.*

Beweis. „(i) $\Leftrightarrow$ (ii)“ Diese Äquivalenz folgt sofort aus der Definition der Begriffe linear unabhängig, Erzeugendensystem und Basis.

„(ii) $\Rightarrow$ (iii)“ Gelte (ii). Dann ist $(v_1, \dots, v_n)$ linear unabhängig. Sei $v_{n+1} \in V$ beliebig aber fest. Weil $(v_1, \dots, v_n)$ auch ein Erzeugendensystem von V ist, gibt es $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ mit

$$v_{n+1} = \sum_{k=1}^{n} \alpha_k v_k.$$

Dann folgt mit $\alpha_{n+1} \coloneqq -1$

$$0 = \sum_{k=1}^{n+1} \alpha_k v_k$$

und

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{n+1} \end{bmatrix} \neq 0.$$

Also ist $(v_1, \dots, v_{n+1})$ **nicht** linear unabhängig nach Proposition 5.1.4 (iii).

„(iii) $\Rightarrow$ (ii)“ Gelte (iii). Dann ist $(v_1, \dots, v_n)$ linear unabhängig. Zu zeigen ist, dass $(v_1, \dots, v_n)$ auch ein Erzeugendensystem von V ist. Sei $v \in V$. Dann ist wegen (iii) $(v_1, \dots, v_n, v)$ linear abhängig. Wegen Proposition 5.1.4 (iii) gibt es ein $\alpha \in \mathbb{K}^{n+1}$ mit

$$\sum_{k=1}^{n+1} \alpha_k v_k = 0 \text{ und } \alpha \neq 0 \quad (\text{wobei } v_{n+1} \coloneqq v).$$

Es muss $\alpha_{n+1} \neq 0$ gelten, denn wäre $\alpha_{n+1} = 0$, dann wäre

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \neq 0.$$

und zudem

$$\sum_{k=1}^{n} \alpha_k v_k = 0$$

Dies widerspricht der linearen Unabhängigkeit von $(v_1, \dots, v_n)$. Nun folgt

$$\alpha_{n+1} v_{n+1} = -\sum_{k=1}^{n} \alpha_k v_k = 0$$

also

$$v = v_{n+1} = \sum_{k=1}^{n} \underbrace{-\frac{\alpha_k}{\alpha_{k+1}}}_{\in \mathbb{K}} v_k$$

Wir haben also $v \in \operatorname{span}(v_1, \dots, v_n)$ gezeigt, d.h. $(v_1, \dots, v_n)$ ist ein Erzeugendensystem von V. Also gilt (ii).

„(ii) $\Rightarrow$ (iv)“ Gelte (ii). Dann ist $(v_1, \dots, v_n)$ ein Erzeugendensystem von V. Sei $k \in \{1, \dots, n\}$. Wäre $(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_n)$ ein Erzeugendensystem von V, dann wäre

$$\operatorname{span}(v_1, \dots, v_n) = V = \operatorname{span}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_n).$$

Also wäre $(v_1, \dots, v_n)$ laut Proposition 5.1.6 linear abhängig, im Widerspruch zu (ii). Also ist $(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_n)$ kein Erzeugendensystem von V.

„(iv) $\Rightarrow$ (ii)“ Gelte (iv). Dann ist $(v_1, \dots, v_n)$ ein Erzeugendensystem von V. Wegen (iv) gilt für jedes $k \in \{1, \dots, n\}$, dass

$$\operatorname{span}(v_1, \dots, v_n) \neq \operatorname{span}(v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_n).$$

Dann ist nach Proposition 5.1.6 das Tupel $(v_1, \dots, v_n)$ linear unabhängig. □

Es ist eine sehr naheliegende Frage, wann ein Vektorraum eine Basis hat. Für Vektorräume, die ein endliches Erzeugendensystem besitzen, wird diese Frage durch den folgenden Satz beantwortet:

Theorem 5.1.9 (Basisauswahlsatz)**.** *Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \dots, v_n \in V$. Wenn $(v_1, \dots, v_n)$ ein Erzeugendensystem von V ist, dann kann man das Tupel durch Weglassen einiger Vektoren[3] zu einer Basis von V machen.*

Beweis. Die Aussage folgt aus Proposition 5.1.8(iv): Wir können einfach immer weiter geeignete Vektoren aus dem Tupel entfernen ohne den Aufspann zu ändern bis wir bei einem minimalen Erzeugendensystem angelangt sind; dieses ist dann eine Basis. □

Wir schließen diesen Abschnitt mit drei Beispielen dazu, wie man lineare Unabhängigkeit beziehungsweise lineare Abhängigkeit beweisen kann:

Beispiele 5.1.10. (a) Lassen Sie uns in $\mathbb{F}_3^3$ die beiden Vektoren

$$v_1 := \begin{bmatrix}[0]\\[2]\\[1]\end{bmatrix}, \qquad v_2 := \begin{bmatrix}[0]\\[1]\\[2]\end{bmatrix}$$

betrachten. Wir möchten wissen, ob (v_1, v_2) linear unabhängig ist. Dazu verwenden wir das Kriterium aus Korollar 5.1.5: Mit Hilfe des Gauß-Algorithmus kann man die Matrix

$$\begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix}[0] & [0]\\[2] & [1]\\[1] & [2]\end{bmatrix}$$

auf reduzierte Zeilenstufenform bringen. Nach kurzer Rechnung erhält man, dass diese gleich

$$\begin{bmatrix}[1] & [2]\\[0] & [0]\\[0] & [0]\end{bmatrix}$$

ist. Weil die Stufentiefe in der zweiten Spalte gleich 0 ist, folgt aus Korollar 5.1.5, dass das Tupel (v_1, v_2) linear abhängig ist.

Durch „scharfes Hinsehen“ kann man das in diesem konkreten Beispiel übrigens auch anders sehen: Es gilt nämlich

$$[2]v_1 = v_2,$$

[3] Dies beinhaltet den Fall, dass man eventuell gar keinen Vektor weglässt.

und somit

$$[2]v_1 - [1]v_2 = 0.$$

Also ist (v_1, v_2) wegen Proposition 5.1.4(iii) linear abhängig.

(b) Betrachten wir nun in $\mathbb{F}_7^3$ die beiden Vektoren

$$v_1 \coloneqq \begin{bmatrix}[0]\\[2]\\[1]\end{bmatrix}, \qquad v_2 \coloneqq \begin{bmatrix}[0]\\[1]\\[2]\end{bmatrix}$$

Wir verwenden wieder Korollar 5.1.5 und bringen dazu die Matrix

$$\begin{bmatrix} v_1 & v_2 \end{bmatrix} = \begin{bmatrix}[0] & [0]\\[2] & [1]\\[1] & [2]\end{bmatrix}$$

auf reduzierte Zeilenstufenform. Diese lautet[4]

$$\begin{bmatrix}[1] & [0]\\[0] & [1]\\[0] & [0]\end{bmatrix}$$

Diese Matrix hat in beiden Spalten die Stufentiefe 1, also ist (v_1, v_2) wegen Korollar 5.1.5 linear unabhängig.

(c) Nun betrachten wir noch eine etwas andere Situation: Wir betrachten den Vektorraum $\mathrm{Abb}([0,\infty);\mathbb{R})$ über $\mathbb{R}$, und hierin diejenigen Vektoren $f_1, f_2, f_3 : [0,\infty) \to \mathbb{R}$, die durch die Formeln

$$f_1(x) = \sin(x), \qquad f_2(x) = \cos(x), \qquad f_3(x) = x(x - \frac{\pi}{2})$$

für alle $x \in [0,\infty)$ gegeben sind.[5]

Lassen Sie uns zeigen, dass (f_1, f_2, f_3) linear unabhängig ist. Dieses mal können wir aber Korollar 5.1.5 nicht verwenden, denn das Korollar macht nur eine Aussage über lineare Unabhängigkeit in Vektorräumen der Form $\mathbb{K}^n$ – nicht aber in anderen Vektorräume wie zum Beispiel $\mathrm{Abb}([0,\infty);\mathbb{R})$.

Stattdessen verwenden wir die Charakterisierung linearer Unabhänigkeit aus Proposition 5.1.4(iii). Sei also $\alpha \in \mathbb{R}^3$ und sei

$$\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3 = 0.$$

[4]Es ist nicht allzu überraschend, dass wir dieses mal eine andere Matrix als reduzierte Zeilenstufenform erhalten – schließlich wird in $\mathbb{F}_7$ ja anders gerechnet als in $\mathbb{F}_3$.

[5]Das ist natürlich ein recht willkürlich gewähltes Beispiel; es ist vor allem so gewählt, dass die Zahlen nachher gut aufgehen.

Weil die Funktion gleich der Nullfunktion ist, ist sie an jedem Punkt $x \in [0, \infty)$ gleich 0 – beispielsweise an den Punkten 0, $\frac{\pi}{2}$ and π. Somit gilt also

$$0 = \big(\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3\big)(0) = \alpha_1 f_1(0) + \alpha_2 f_2(0) + \alpha_3 f_3(0) = \alpha_2,$$
$$0 = \big(\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3\big)(\frac{\pi}{2}) = \alpha_1 f_1(\frac{\pi}{2}) + \alpha_2 f_2(\frac{\pi}{2}) + \alpha_3 f_3(\frac{\pi}{2}) = \alpha_1,$$
$$0 = \big(\alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3\big)(\pi) = \alpha_1 f_1(\pi) + \alpha_2 f_2(\pi) + \alpha_3 f_3(\pi) = -\alpha_2 + \alpha_3 \frac{\pi^2}{2}.$$

Die ersten beiden dieser drei Gleichungen zeigen uns, dass $\alpha_1 = 0$ und $\alpha_2 = 0$ gilt, und wenn wir dies mit der dritten Gleichung kombinieren, erhalten wir auch $\alpha_3 = 0$. Somit ist $\alpha = 0$.

Laut Proposition 5.1.4(iii) zeigt dies, dass (f_1, f_2, f_3) linear unabhängig ist.

5.2 Der Dimensions-Begriff

Endlich-dimensionale Vektorräumen und die Koordinatenabbildung

Basierend auf dem Begriff der Basis eines Vektorraums können wir nun über den Dimensionsbegriff sprechen. Wir beginnen zunächst mit dem folgenden, noch recht unspezifischen Konzept:

Definition 5.2.1 (Endlich-dimensionaler Vektorraum)**.** Ein Vektorraum V über einem Körper $\mathbb{K}$ heißt **endlich-dimensional**, wenn es eine Zahl $n \in \mathbb{N}$ und Vektoren $v_1, \dots, v_n \in V$ gibt derart, dass $(v_1, \dots, v_n)$ eine Basis von V ist.

Aus dem Basisauswahlsatz folgt sofort die folgende Proposition:

Proposition 5.2.2. *Ein Vektorraum V über einem Körper $\mathbb{K}$ ist genau dann endlich-dimensional, wenn es eine Zahl $n \in \mathbb{N}$ und Vektoren $v_1, \dots, v_n \in V$ gibt derart, dass $(v_1, \dots, v_n)$ ein Erzeugendensystem von V ist.*

Endlich-dimensionale Vektorräume sind immer isomorph – im folgenden Sinne – zu $\mathbb{K}^n$ für ein geeignetes n:

Proposition 5.2.3. *Sei $V \neq \{0\}$ ein endlich-dimensionaler Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und sei $\mathcal{B} \coloneqq (v_1, \dots, v_n)$ eine Basis von V. Dann gilt $n \neq 0$ und die Abbildung*

$$\ell_{\mathcal{B}} : \mathbb{K}^n \to V,$$
$$\alpha \mapsto \sum_{k=1}^{n} \alpha_k v_k$$

ist ein Vektorraumisomorphismus.

Beweis. Dass die Abbildung $\ell_{\mathcal{B}}$ injektiv ist, folgt sofort daraus, dass $(v_1, \ldots, v_n)$ linear unabhängig ist. Und dass die Abildung $\ell_{\mathcal{B}}$ surjektiv ist, folgt sofort daraus, dass $(v_1, \ldots, v_n)$ ein Erzeugendensystem ist. Die Linearität von $\ell_{\mathcal{B}}$ wiederum folgt sofort aus der Definition von $\ell_{\mathcal{B}}$ und den Körperaxiomen. □

Definition 5.2.4 (Koordinatenabbildung)**.** Sei $V \neq \{0\}$ ein endlich-dimensionaler Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}^*$ und sei $\mathcal{B} \coloneqq (v_1, \ldots, v_n)$ eine Basis von V. Sei $\ell_{\mathcal{B}} : \mathbb{K}^n \to V$ der Vektorraumisomorphismus aus Proposition 5.2.3.

Dann heißt der Vektorraumisomorphismus $\kappa_{\mathcal{B}} \coloneqq \ell_{\mathcal{B}}^{-1} : V \to \mathbb{K}^n$ die **Koordinatenabbildung** von V bezüglich der Basis $\mathcal{B}$.

Beachten Sie, dass die Koordinatenabbildung tatsächlich ein Vektorraumisomorphismus ist, da die Umkehrabbildung eines Vektoraumisomorphismus laut Aufgabe 4 auf Hausaufgabenblatt 6 wieder ein Vektorraumisomorphismus ist.

Wir wollen nun als nächstes zeigen, dass in einem endlich-dimensionalen Vektorraum alle Basen gleich viele Elemente haben. Dies können wir aus dem folgenden Resultat folgern:

Lemma 5.2.5. *Sei $\mathbb{K}$ ein Körper und seien $m, n \in \mathbb{N}^*$. Wenn es einen Vektorraumisomorphismus $S : \mathbb{K}^n \to \mathbb{K}^m$ gibt, dann ist $m = n$.*

Beweis. Wir zeigen zunächst, dass $n \leq m$ gilt.

Das Tupel $(e_1, \ldots, e_n)$ ist eine Basis von $\mathbb{K}^n$ und somit linear unabhängig. Mit Hilfe der Injektivität von S kann man hieraus folgern, dass das Tupel

$$(S(e_1), \ldots, S(e_n))$$

von Vektoren in $\mathbb{K}^m$ ebenfalls linear unabhängig ist; dazu verwenden wir die Charakterisierung von linearer Unabhängigkeit aus Proposition (iii): Sein $\alpha \in \mathbb{K}^n$ derart, dass

$$\sum_{k=1}^{n} \alpha_k S(e_k) = 0$$

ist. Dann folgt wegen der Linearität von S

$$S(0) = 0 = S\Big(\sum_{k=1}^{n} \alpha_k e_k\Big).$$

Weil S injektiv ist, erhalten wir hieraus, dass

$$0 = \sum_{k=1}^{n} \alpha_k e_k$$

gilt. Da das Tupel $(e_1, \ldots, e_n)$ eine Basis von $\mathbb{K}^n$ und somit linear unabhängig ist, folgt aus dieser Gleichung wiederum, dass $\alpha = 0$ ist. Also ist $(S(e_1), \ldots, S(e_n))$ wie behauptet linear unabhängig.

Somit hat wegen Korollar 5.1.5 in der reduzierten Zeilenstufenform der Matrix

$$\begin{bmatrix} S(e_1) & \dots & S(e_n) \end{bmatrix} \in \mathbb{K}^{m \times n}$$

jede Spalte die Stufentiefe 1. Damit muss $n \leq m$ sein.

Weil $S^{-1} : \mathbb{K}^m \to \mathbb{K}^n$ ebenfalls ein Vektorraumisomorphismus ist, kann man dasselbe Argument auch auch S^{-1} anwenden und erhält somit auch $m \leq n$.[6] □

Aus dem vorangehenden Lemma können wir nun den folgenden Satz folgern:

Theorem 5.2.6. *Sei V ein endlich-dimensionaler Vektorraum über einem Körper $\mathbb{K}$. Dann haben alle Basen von V gleich viele Elemente.*

Beweis. Falls V nur den Nullvektor enthält, ist das leere Tupel die einzige Basis von V, also können wir von nun an annehmen, dass V nicht nur den Nullvektor enthält und dass jede Basis von V somit mindestes ein Element besitzt.

Seien nun $m, n \in \mathbb{N}^*$ und seien $\mathcal{B} \coloneqq (v_1, \dots, v_n)$ und $\mathcal{C} \coloneqq (v_1, \dots, v_m)$ Basen von V. Wir müssen $m = n$ zeigen. Dazu verwenden wir die Koordinatenabbildungen $\kappa_\mathcal{B} : V \to \mathbb{K}^n$ und $\kappa_\mathcal{C} : V \to \mathbb{K}^m$. Diese beiden Abbildungen sind, wie Sie bereits wissen, Vektorraumisomorphismen, und somit ist auch $\kappa_\mathcal{B}^{-1}$ ein Vektorraumisomorphismus (laut Aufgabe 4 auf Hausaufgabenblatt 6). Die Hintereinanderausführung

$$\mathbb{K}^n \xrightarrow{\kappa_\mathcal{B}^{-1}} V \xrightarrow{\kappa_\mathcal{C}} \mathbb{K}^m$$

ist somit laut Proposition 1.5.14(c) ebenfalls bijektiv, und es ist nicht schwer zu sehen, dass Sie auch linear ist.[7] Somit ist $\kappa_\mathcal{C} \circ \kappa_\mathcal{B}^{-1}$ ein Vektorraumisomorphismus von $\mathbb{K}^n$ nach $\mathbb{K}^m$. Also gilt laut Lemma 5.2.5 $m = n$. □

Aufgrund des vorangehenden Theorems ergibt die folgende Definition Sinn:

Definition 5.2.7 (Dimension eines Vektorraums)**.** Sei V ein endlich-dimensionaler Vektorraum über einem Körper $\mathbb{K}$. Die **Dimension von** V, welche wir mit $\dim V$ notieren, ist die Anzahl der Elemente einer (und somit jeder) Basis von V.

Für jeden endlich-dimensionalen Vektorraum V ist somit $\dim V \in \mathbb{N}$. Beachten Sie, dass die Dimension auch 0 sein kann.

Beispiele 5.2.8. (a) Ein Vektorraum, der nur aus dem Nullvektor 0 besteht, hat Dimension 0 (denn die einzige Basis dieses Vektorraums ist das leere Tupel.[8]

(b) Für einen Körper $\mathbb{K}$ und $n \in \mathbb{N}^*$ gilt $\dim \mathbb{K}^n = n$, denn man kann leicht überprüfen, dass $(e_1, \dots, e_n)$ eine Basis ist (wobei $e_1, \dots, e_n$ die kanonischen Einheitsvektoren bezeichnen).

[6] Frage: Wo im Beweis haben wir verwendet, dass S surjektiv ist?

[7] Allgemeiner gilt: Die Hintereinanderausführung zweier linearer Abbildungen ist immer linear. Versuchen Sie zu Ihrer Übung, dies zu beweisen! Sie benötigen dazu lediglich die Definition der Hintereinanderausführung und von linearen Abbildungen.

[8] Frage: Warum ist (0) keine Basis dieses Vektorraums?

(c) Für einen Körper $\mathbb{K}$ hat der Vektorraum $\mathbb{K}^{2\times 3}$ die Dimension $2 \cdot 3 = 6$. Wenn Sie nämlich die Matrizen

$$A_1 \coloneqq \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_3 \coloneqq \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad A_5 \coloneqq \begin{bmatrix} 0 & 0 & 5 \\ 0 & 0 & 0 \end{bmatrix},$$
$$A_2 \coloneqq \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad A_4 \coloneqq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_6 \coloneqq \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

betrachten, dann können Sie leicht überprüfen, dass das Tupel $(A_1, \ldots, A_6)$ eine Basis von $\mathbb{K}^{2\times 3}$ ist.

Als Konsequenz aus Theorem 5.2.6 erhalten wird folgende Beobachtung:

Korollar 5.2.9. *Sei V ein endlich-dimensionaler Vektorraum über $\mathbb{K}$ mit Dimension $\dim V = d \in \mathbb{N}$.*

(a) *Jedes Erzeugendensystem von V hat mindestens d Elemente.*

(b) *Wenn ein Erzeugendensystem von V genau d Elemente hat, dann ist es bereits eine Basis von V.*

(c) *Jedes linear unabhängige Tupel von Vektoren von V hat höchstens d Elemente.*

(d) *Wenn ein linear unabhängiges Tupel von Vektoren von V genau d Elemente hat, dann ist es bereits eine Basis von V.*

Beweis. Sei $n \in \mathbb{N}$ und seien $v_1, \ldots, v_n \in V$

(a) und (b): Sei $(v_1, \ldots, v_n)$ ein Erzeugendensystem von V. Wegen des Basisauswahlsatzes gilt: Wir können durch Weglassen von $p \in \{0, \ldots, n\}$ Vektoren eine Basis von V erhalten. Somit ist $n - p = d$. Also folgt:

- $n \geq n - p = d$. Dies zeigt (a).
- Wenn $n = d$ ist, dann gilt $p = 0$. D.h. $(v_1, \ldots, v_n)$ ist selbst schon eine Basis. Dies zeigt (b)

(c) und (d): Sei $(v_1, \ldots, v_n)$ linear unabhängig. Aussage (c) beweisen wir ähnlich wie in Theorem 5.2.6 und Lemma 5.2.5. Sei $(w_1, \ldots, w_d) = \mathcal{C}$ eine Basis von V. Sei

$$S : \mathbb{K}^n \to V$$
$$\alpha \mapsto \sum_{k=1}^{n} \alpha_k v_k$$

Dann ist S linear und injektiv (da $(v_1, \ldots, v_n)$ linear unabhängig ist). Die Hintereinanderausführung

$$\tilde{S} : \mathbb{K}^n \xrightarrow{S} V \xrightarrow{\kappa_{\mathcal{C}}} \mathbb{K}^d$$

ist linear und injektiv. Wie im Beweis von Lemma 5.2.5 folgt aus der Existenz einer injektiven Abbildung $\tilde{S} : \mathbb{K}^n \to \mathbb{K}^d$, dass $n \leq d$ ist. Dies zeigt (c).
Sei nun $n = d$. Wenn $(v_1, \ldots, v_n)$ keine Erzeugendensystem wäre, dann gäbe es einen Vektor $v_{n+1} \in V$, der nicht im Aufspann von $(v_1, \ldots, v_n)$ liegt. Aufgrund des nachfolgenden Lemmas wäre dann $(v_1, \ldots, v_{n+1})$ ebenfalls linear unabhängig. Wegen $n + 1 = d + 1 \geq d$ widerspricht dies (c). □

Lemma 5.2.10 (Ergänzungslemma)**.** *Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \ldots, v_n \in V$ derart, dass $(v_1, \ldots, v_n)$ linear unabhängig ist.*

Sei $v_{n+1} \in V$ ein Vektor, der nicht im Aufspann von $(v_1, \ldots, v_n)$ liegt. Dann ist auch $(v_1, \ldots, v_n, v_{n+1})$ linear unabhängig.

Beweis. Wir nehmen widerspruchshalber an, dass $(v_1, \ldots, v_{n+1})$ linear abhängig ist. Dann gibt es ein $\alpha \in \mathbb{K}^n \setminus \{0\}$ derart, dass $\sum_{k=1}^{n+1} \alpha_k v_k = 0$ ist. Nun kann der Skalar α_{n+1} nicht 0 sein: Wäre er nämlich gleich 0, so würde nämlich einerseits folgen, dass mindestens einer der Skalare $\alpha_1, \ldots, \alpha_n$ nicht 0 ist, und andererseits wäre $\sum_{k=1}^{n} \alpha_k v_k = 0$. Dies widerspricht der linearen Unabhängigkeit von $(v_1, \ldots, v_n)$.
Also ist tatsächlich $\alpha_{n+1} \neq 0$. Daraus folgt aber

$$v_{n+1} = \sum_{k=1}^{n} -\frac{\alpha_k}{\alpha_{n+1}} v_k,$$

was der Annahme $v_{n+1} \notin \operatorname{span}(v_1, \ldots, v_n)$ widerspricht. □

Aus dem vorangehenden Lemma können wir zudem das folgende Theorem folgern, welches in gewissem Sinne Komplementär zum Basisauswahlsatz ist:

Theorem 5.2.11 (Basisergänzungssatz)**.** *Sei V ein endlich-dimensionaler Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \ldots, v_n \in V$ derart, dass $(v_1, \ldots, v_n)$ linear unabhängig ist. Dann gibt es eine Zahl $p \in \mathbb{N}$ und Vektoren $v_{n+1}, \ldots, v_{n+p}$ derart, dass $(v_1, \ldots, v_{n+p})$ eine Basis von V ist.*

Beweis. Sei $d \coloneqq \dim V$. Dann ist $d \geq n$. Wenn $n = d$ ist, ist $(v_1, \ldots, v_n)$ bereits eine Basis von V. Wenn hingegen $n < d$ ist, ist das Tupel keine Basis von V und somit kein Erzeugendensystem. Somit ist der Aufspann dieses Tupels nicht gleich V, also gibt es einen Vektor $v_{n+1} \in V$, der nicht im Aufspann von V liegt. Laut Lemma 5.2.10 ist deshalb $(v_1, \ldots, v_{n+1})$ linear unabhängig. Wenn aber $n+1 = d$ ist, sind wir fertig, denn dann ist letztgenanntes Tupel sogar eine Basis von V. Wenn hingegen $n + 1 < d$ ist, iterieren wir den zuvor genannten Schritt solange, bis wir d erreichen. □

Zum Abschluss besprechen wir noch ein Beispiel um die bisher besprochenenen Konzepte zu veranschaulichen:

Beispiel 5.2.12. Betrachten wir im $\mathbb{R}^2$ die Vektoren

$$v_1 \coloneqq \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad v_2 \coloneqq \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \text{und} \qquad v \coloneqq \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

Dann kann man leicht nachrechnen, dass (v_1, v_2) linear unabhängig ist; wegen $\dim \mathbb{R}^2 = 2$ ist also $\mathcal{B} \coloneqq (v_1, v_2)$ sogar eine Basis von $\mathbb{R}^2$.

Wenn wir $\alpha \coloneqq \kappa_{\mathcal{B}}(v)$ setzen, dann gilt $v = \alpha_1 v_1 + \alpha_2 v_2$, also

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}.$$

Dieses lineare Gleichungssystem können wir mit Hilfe des Gaußalgorithmus lösen, wobei wir $\alpha_1 = 2$ und $\alpha_2 = 1$ erhalten. Also ist $v = 2 \cdot v_1 + 1 \cdot v_2$.

Dimension von Untervektorräumen

Da jeder Untervektorraum eines Vektorraums selbst wieder ein Vektorraum ist, kann man auch von der Dimension eines Untervektorraums sprechen.

Proposition 5.2.13. *Sei V ein endlich-dimensionaler Vektorraum über einem Körper $\mathbb{K}$.*

(a) *Für jeden Untervektorraum U von V gilt: Es ist U ebenfalls endlich-dimensional, und $\dim U \leq \dim V$.*

(b) *Seien $U_1, U_2 \subseteq V$ Untervektorräume. Wenn $U_1 \subseteq U_2$ und $\dim U_1 = \dim U_2$ gilt, dann ist $U_1 = U_2$.*

Beweis. (a) Sei $d \coloneqq \dim V \in \mathbb{N}$. Dann hat jedes linear unabhängige Tupel von Vektoren in V höchstens d Elemente. Insbesondere gilt dies dann natürlich auch für linear unabhängige Tupel von Vektoren aus U.

Unter allen linearen unabhängigen Tupeln von Vektoren in U können wir also eines – sagen wir $(u_1, \ldots, u_k)$ auswählen, dass maximal viele Elemente hat; es gilt dann natürlich $k \leq d$. Aus Proposition 5.1.8(iii) – angewendet auf den Vektorraum U statt V – folgt nun, dass $(u_1, \ldots, u_k)$ eine Basis von U ist. Also ist U endlich-dimensional und es gilt $\dim U = k \leq d = \dim V$.

(b) Sei $U_1 \subseteq U_2$ und $\dim U_1 = \dim U_2$. Wir müssen zeigen, dass auch $U_1 \supseteq U_2$ gilt.

Sei $(v_1, \ldots, v_n)$ für ein $n \in \mathbb{N}$ eine Basis von U_1. Laut Basisergänzungssatz können wir diese zu einer Basis $(v_1, \ldots, v_n, v_{n+1}, \ldots, v_{n+p})$ von U_2 (mit einem $p \in \mathbb{N}$) ergänzen. Damit ist $n = \dim U_1 = \dim U_2 = n + p$, also $p = 0$. Das heißt, $(v_1, \ldots, v_n)$ ist selbst bereits eine Basis von U_2.

Wenn also $u \in U_2$ ist, dann lässt sich u als Linearkombination von $(v_1, \ldots, v_n)$ schreiben, und weil diese Vektoren in U_1 liegen und U_1 ein Untervektorraum ist, folgt hieraus, dass auch $u \in U_1$ ist. □

Lassen Sie uns zunächst einige sehr einfache Beispiele für die Dimension von Untervektorräumen besprechen:

Beispiele 5.2.14. (a) Betrachten wir den Vektorraum $\mathbb{R}^2$ (über dem Körper $\mathbb{R}$). Er hat die Dimension 2,[9] also folgt aus Proposition (a), dass jeder Untervektorraum von $\mathbb{R}^2$ Dimension 0, 1, oder 2 besitzt. Lassen Sie uns diese Fälle etwas mehr im Detail besprechen:

- Es gibt nur einen 0-dimensionalen Untervektorraum von $\mathbb{R}^2$, nämlich $\{0\}$.
- Es gibt zahlreiche 1-dimensionale Untervektorräume von $\mathbb{R}^2$. Sie können sich anschaulich leicht überlegen, dass dies genau die Ursprungsgeraden im $\mathbb{R}^2$ sind.
- Weil $\mathbb{R}^2$ die Dimension 2 besitzt, folgt aus Proposition 5.2.13(b), dass $\mathbb{R}^2$ selbst der einzige 2-dimensionale Untervektorraum von $\mathbb{R}^2$ ist.

(b) Betrachten wir nun den Vektorraum $\mathbb{R}^3$ (über dem Körper $\mathbb{R}$). Er besitzt die Dimension 3, also hat jeder Untervektorraum von $\mathbb{R}^3$ Dimension 0, 1, 2, oder 3 (wegen Proposition 5.2.13(a)). Lassen Sie uns wieder alle Fälle durchgehen:

- Der einzige 0-dimensionale Untervektorraum von $\mathbb{R}^3$ ist die Menge $\{0\}$.[10]
- Es gibt zahlreiche 1-dimensionale Untervektorräumen von $\mathbb{R}^3$. Dies sind anschaulich gesprochen gerade die Ursprungsgeraden in $\mathbb{R}^3$.
- Ebenso gibt es zahlreiche 2-dimensionale Untervektorräume von $\mathbb{R}^3$. Anschaulich sind dies gerade die Ursprungsebenen in $\mathbb{R}^3$.
- Es gibt genau einen 3-dimensionalen Untervektorraum von $\mathbb{R}^3$, nämlich $\mathbb{R}^3$ selbst. Dies folgt wie im vorangehenden Beispiel auf Proposition 5.2.13(b).

(c) Zuletzt betrachten wir noch den Vektorraum $\mathbb{R}^1 = \mathbb{R}$ (über dem Körper $\mathbb{R}$). Er hat die Dimension 1. Also besitzt er wegen Proposition 5.2.13 genau zwei Untervektorräume, nämlich den 0-dimensionalen Raum $\{0\}$,[11] und den 1-dimensionalen Untervektorraum $\mathbb{R}$ (also sich selbst).

Für den Durchschnitt von zwei Untervektorräumen hat man die folgende Formel für die Dimension:

Theorem 5.2.15 (Dimensionsformel für den Durschnitt von Untervektorräumen)**.** *Sei V ein endlich-dimensionaler Vektorraum über einem Körper $\mathbb{K}$ und seien $U_1, U_2 \subseteq V$ Untervektorräume. Dann ist die Menge*

$$U_1 + U_2 \coloneqq \{u_1 + u_2 \mid \ u_1 \in U_1,\ u_2 \in U_2\}$$

ebenfalls ein Untervektorraum von V, und es gilt die Formel

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

Den Beweis lagern wir in die Ergänzungen in Abschnitt 5.8 am Ende dieses Kapitels aus; wenn Sie Lust haben, können Sie ihn gerne dort durchlesen.

[9] Warum nochmal?

[10] Wobei 0 nun natürlich den Nullvektor im $\mathbb{R}^3$ bezeichnet, während dieses Symbol im Beispiel zuvor den Nullvektor im $\mathbb{R}^2$ bezeichnet hatte.

[11] Was bezeichnet das Symbol 0 dieses Mal?

5.3 Explizite Darstellung von Untervektorräumen

Spezifikation eines Untervektorraumes durch Angabe einer Basis

Aus den Beispielen 5.2.14 können Sie bereits erkennen, dass Untervektorräume von Vektorräumen häufig eine geometrische Interpretation haben. Entsprechend ist es wichtig, Möglichkeiten zu finden um einen Untervektorraum konkret zu beschreiben.

Weil wir bereits wissen, dass der Aufspann einer Teilmenge – oder eines Tupels – eines Vektorraumes immer ein Untervektorraum ist (Proposition 5.1.2), können wir einen Untervektorraum angeben, indem wir ein Tupel (oder eine Menge) aufschreiben, das (bzw. die) den Untervektorraum aufspannt.

Um dabei mit möglichst wenigen Daten auszukommen, ist es sinnvoll, ein solches Tupel minimal zu wählen. Wir wollen also zum Beispiel in einem Vektorraum V aus einem Tupel von Vektoren $(v_1, \dots, v_n)$, welches einen Untervektorraum

$$U \coloneqq \operatorname{span}(v_1, \dots, v_n)$$

von V aufspannt, ein minimales Tupel mit derselben Eigenschaft auswählen. Laut Proposition 5.1.8(iv) bedeutet dies gerade, dass wir aus dem Tupel eine Basis von U auswählen.

Falls der Vektorraum V gleich $\mathbb{K}^m$ ist geht dies – wieder einmal – mit dem Gauß-Algorithmus:

Theorem 5.3.1. *Sei $\mathbb{K}$ ein Körper, seien $m, n \in \mathbb{N}^*$ und seien $v_1, \dots, v_n \in \mathbb{K}^m$. Es bezeichne $Z \in \mathbb{K}^{m \times n}$ die reduzierte Zeilenstufenform – oder allgemeiner irgendeine Zeilenstufenform – der Matrix*

$$\begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \in \mathbb{K}^{m \times n}.$$

Wenn $1 \leq k_1 < \dots < k_\ell \leq n$ (für ein $\ell \in \mathbb{N}$) diejenigen Spalten von Z bezeichnen, in denen Z Stufentiefe 1 hat, dann ist $(v_{k_1}, \dots, v_{k_\ell})$ eine Basis des Untervektorraums $\operatorname{span}(v_1, \dots, v_n)$.[12,13]

Beweis. Die Stufentiefen der einzelnen Spalten sind in jeder Zeilenstufenform gleich wie in der reduzierten Zeilenstufenform – somit genügt es das Theorem zu beweisen, wenn Z die reduzierte Zeilenstufenform von $\begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}$ ist; sei dies also nun der Fall.

Laut Theorem 4.3.4(c) gibt es eine invertierbare Matrix $T \in \mathbb{K}^{m \times m}$ mit der Eigenschaft

$$T \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} = Z.$$

[12] Wichtig: Es ist nicht etwa so, dass die Spalten von Z mit den Nummern $k_1, \dots, k_\ell$ eine Basis von $\operatorname{span}(v_1, \dots, v_n)$ bilden – sondern man nimmt tatsächlich die ursprünglichen Vektoren $v_1, \dots, v_n$ und wählt aus ihnen diejenigen mit den Indizes $k_1, \dots, k_\ell$ aus.

[13] Übrigens ist dies im Allgemeinen nicht die einzige Möglichkeit, aus den Vektoren $v_1, \dots, v_n$ eine Basis ihres Aufspanns auszuwählen – aber es ist eine, die immer funktioniert.

Wir bezeichnen nun mit $z_1, \ldots, z_n \in \mathbb{K}^m$ die Spalten von Z. Dann bedeutet die vorangehende Gleichheit, dass

$$Tv_1 = z_1, \quad \ldots \quad Tv_n = z_n$$

gilt. Es bezeichne nun $V \subseteq \mathbb{K}^m$ den Aufspann von $(v_1, \ldots, v_n)$, und es bezeichne $W \subseteq \mathbb{K}^m$ den Aufspann von $(z_1, \ldots, z_n)$. Weil sich Z in reduzierter Zeilenstufenform befindet, kann man leicht sehen, dass $(z_{k_1}, \ldots, z_{k_\ell})$ linear unabhängig ist und zudem denselben Aufspann hat wie $(z_1, \ldots, z_n)$. Das bedeutet also, $(z_{k_1}, \ldots, z_{k_l})$ ist eine Basis von W.

Nun können wir zeigen, dass $(v_{k_1}, \ldots, v_{k_\ell})$ wie behauptet eine Basis von V ist:

- *Erzeugendensystem:* Wir müssen $\operatorname{span}(v_{k_1}, \ldots, v_{k_\ell}) = V$ zeigen.

  „$\subseteq$“ Diese Inklusion ist klar.

  „$\supseteq$“ Sei $v \in V$. Dann gilt $Tv \in W$, also gibt es Skalare $\alpha_{k_1}, \ldots, \alpha_{k_\ell} \in \mathbb{K}$ mit der Eigenschaft

  $$Tv = \sum_{j=1}^{\ell} \alpha_{k_j} z_{k_j},$$

  und somit ist

  $$v = \sum_{j=1}^{\ell} \alpha_{k_j} T^{-1} z_{k_j} = \sum_{j=1}^{\ell} \alpha_{k_j} v_{k_j} \in \operatorname{span}(v_{k_1}, \ldots, v_{k_\ell}).$$

- *Lineare Unabhängigkeit:* Seien $\alpha_{k_1}, \ldots, \alpha_{k_\ell} \in \mathbb{K}$ derart, dass

  $$\sum_{j=1}^{\ell} \alpha_{k_j} v_{k_j} = 0$$

  gilt. Dann folgt

  $$0 = T0 = \sum_{j=1}^{\ell} \alpha_{k_j} T v_{k_j} = \sum_{j=1}^{\ell} \alpha_{k_j} z_{k_j}.$$

  Wegen der linearen Unabhängigkeit von $(z_{k_1}, \ldots, z_{k_\ell})$ impliziert dies, dass $\alpha_{k_1} = 0, \ldots, \alpha_{k_\ell} = 0$ gilt. □

Es gibt auch noch eine weitere Möglichkeit, eine Basis des Aufspanns eines Tupels zu berechnen. Der Vollständigkeit halber geben wir diese hier ohne Beweis an:[14]

[14] Der Beweis ist nicht etwas besonders schwer – aber wir wollen uns nicht zulange damit aufhalten verschiedene Verfahren zu beweisen um aus einem Tupel eine Basis seines Aufspanns zu berechnen.

Theorem 5.3.2. *Sei $\mathbb{K}$ ein Körper, seien $m, n \in \mathbb{N}^*$ und seien $v_1, \ldots, v_n \in \mathbb{K}^m$. Es bezeichne $Y \in \mathbb{K}^{n\times m}$ die reduzierte Zeilenstufenform – oder allgemeiner irgendeine Zeilenstufenform – der Matrix*

$$\begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ \vdots \\ v_n^{\mathrm{T}} \end{bmatrix} \in \mathbb{K}^{n\times m}.$$

Wenn man diejenigen Zeilen von Y, die nicht 0 sind, wieder transponiert, dann bilden die so erhaltenen Spaltenvektoren eine Basis von $\operatorname{span}(v_1, \ldots, v_n)$.

Einige Beispiele

Beispiele 5.3.3. (a) Betrachten Sie die drei Vektoren

$$v_1 := \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \qquad v_2 := \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \qquad v_3 := \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

in $\mathbb{R}^3$. Wir wollen eine Basis des Untervektorraums $U := \operatorname{span}(v_1, v_2, v_3)$ von $\mathbb{R}^3$ bestimmen. Dazu geben wir drei Möglichkeiten an:

(i) Wir verwenden Theorem 5.3.1. Dazu betrachten wir die Matrix

$$\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$$

im $\mathbb{R}^3$. Mit Hilfe des Gaußalgorithmus können wir die Matrix auf die Zeilenstufenform

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

bringen. (Diese ist noch nicht reduziert, aber das ist laut Theorem 5.3.1 auch nicht nötig.) In den Spalten 1 und 2 hat diese Matrix die Stufentiefe 1. Also ist laut Theorem 5.3.1 das Tupel (v_1, v_2) eine Basis von U.

(ii) Als Alternative wir nun Theorem 5.3.2 um diesselbe Aufgabe zu lösen: Die Matrix

$$\begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ v_3^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 2 \\ 2 & 1 & 3 \end{bmatrix}$$

kann man mit Hilfe des Gaußverfahrens auf eine Zeilenstufenform bringen, zum Beispiel auf die Form

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

Somit ist wegen Theorem 5.3.2 auch das Tupel

$$\left(\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -3 \\ 1 \end{bmatrix} \right)$$

eine Basis von U.[15]

(iii) Zuletzt zeigen wir noch eine etwas veränderte Version des soeben vorgestellten Vorgehens: Laut Theorem 5.3.2 können wir *irgendeine* Zeilenstufenform verwenden – zum Beispiel auch die reduzierte. Wenn wir die Matrix

$$\begin{bmatrix} v_1^{\mathrm{T}} \\ v_2^{\mathrm{T}} \\ v_3^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 2 \\ 2 & 1 & 3 \end{bmatrix}$$

deshalb – anders als soeben – auf reduzierte Zeilenstufenform bringen, erhalten wir die Matrix

$$\begin{bmatrix} 1 & 0 & \frac{5}{3} \\ 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix}.$$

Also ist auch

$$\left(\begin{bmatrix} 1 \\ 0 \\ \frac{5}{3} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -\frac{1}{3} \end{bmatrix} \right)$$

eine Basis von U. Wir haben somit mit drei verschiedenen Vorgehensweise drei verschiedene Basen von U gefunden.[16] Man kann übrigens noch viele weitere Basen von U angeben. Aber wir belassen es an dieser Stelle dabei, denn wir wollten ja ursprünglich einfach eine Basis von U finden, was uns schon lange gelungen ist.

[15] Beachten Sie aber, dass dieses Tupel kein Auswahl der Vektoren v_1, v_2, v_3 enthält – Theorem 5.3.2 garantiert lediglich, dass man *irgendeine* Basis von $U = \operatorname{span}(v_1, v_2, v_3)$ findet. Wenn man explizit aus dem Tupel (v_1, v_2, v_3) eine Basis auswählen will (sozusagen um die Situation im Basisauswahlsatz nachzustellen), ist deshalb Theorem 5.3.1 das Mittel der Wahl.

[16] Dies zeigt sehr schön auf, dass eine Vektorraum im Allgemeinen viele verschiedene Basen hat. Das sollte man immer im Hinterkopf behalten.

(b) Betrachten Sie das folgende Tupel, welches aus zwei Vektoren in $\mathbb{F}_2^2$ besteht:

$$\left(\begin{bmatrix}[1]\\[1]\end{bmatrix}, \begin{bmatrix}[0]\\[1]\end{bmatrix}\right)$$

Wenn wir die Matrix

$$\begin{bmatrix}[1] & [0]\\[1] & [1]\end{bmatrix}$$

auf eine Zeilenstufenform bringen, erhalten wir zum Beispiel[17]

$$\begin{bmatrix}[1] & [0]\\[0] & [1]\end{bmatrix}.$$

Diese Matrix hat in jeder Spalte die Stufentiefe 1, also ist das ursprünglich gegeben Tupel bereits eine Basis seines Aufspanns.

Übrigens: Das wiederum bedeutet, dass der Aufspann die Dimension 2 hat, und somit ist er wegen Proposition 5.2.13(b) gleich $\mathbb{F}_2^2$. D.h. wir haben somit erkannt, dass

$$\left(\begin{bmatrix}[1]\\[1]\end{bmatrix}, \begin{bmatrix}[0]\\[1]\end{bmatrix}\right)$$

eine Basis von $\mathbb{F}_2^2$ ist.

5.4 Direkte Summen

Zuletzt behandeln wir hier noch kurz den Begriff der **direkten Summe** von Untervektorräumen. Er lehnt sich an den Begriff der Basis an, verwendet allerdings zur Darstellung von Vektoren nicht ein Tupel aus endlich vielen Vektoren, sondern stattdessen endlich viele Untervektorräume:

Definition 5.4.1 (Direkte Summe)**.** Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $U_1, \ldots, U_n \subseteq V$ Untervektorräume. Man sagt, dass V die **direkte Summe** von $U_1, \ldots, U_n$ ist, falls es für jeden Vektor $v \in V$ genau ein Tupel $(u_1, \ldots, u_n) \in U_1 \times \cdots \times U_n$ mit der Eigenschaft

$$v = \sum_{k=1}^{n} u_k$$

gibt.

Man schreibt $V = U_1 \oplus \cdots \oplus U_n$ um auszudrücken, dass V die direkte von $U_1, \ldots, U_n$ ist.

[17] Weshalb nochmal stehen hier die beiden Worte „Zum Beispiel“?

Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}$ und seien $v_1, \ldots, v_n$ Vektoren. Man kann sich überlegen, dass $(v_1, \ldots, v_n)$ genau dann eine Basis von V ist, wenn alle Vektoren $v_1, \ldots, v_n$ ungleich 0 sind und V die direkte Summe der ein-dimensionalen Untervektorräume $\operatorname{span}(v_1), \ldots, \operatorname{span}(v_n)$ ist.

Wir betrachten nun ein etwas konkreteres Beispiel:

Beispiel 5.4.2. Betrachten wir nun den Vektorraum $\mathbb{R}^3$ über $\mathbb{R}$. Wir setzen

$$U_1 \coloneqq \operatorname{span}(e_1) = \left\{ \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} \mid x_1 \in \mathbb{R} \right\}$$

und

$$U_2 \coloneqq \operatorname{span}(e_2, e_3) \coloneqq \left\{ \begin{bmatrix} 0 \\ x_2 \\ x_3 \end{bmatrix} \mid x_2, x_3 \in \mathbb{R} \right\}.$$

Anschaulich ist U_1 also die x_1-Achse in $\mathbb{R}^3$, und U_2 ist die x_2-x_3-Ebene.

Dann gilt $\mathbb{R}^3 = U_1 \oplus U_2$, d.h. $\mathbb{R}^3$ ist die direkte Summe von U_1 und U_2. Für jedes $y \in \mathbb{R}^3$ gilt nämlich:

- Wir können y in der Form

  $$y = \begin{bmatrix} y_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y_2 \\ y_3 \end{bmatrix}$$

  als Summe von Vektoren aus U_1 und U_2 schreiben.

- Wir müssen nun noch die Eindeutigkeit zeigen: Seien $(u_1, u_2) \in U_1 \times U_2$ und $(\tilde{u}_1, \tilde{u}_2) \in U_1 \times U_2$ derart, dass $u_1 + u_2 = y$ und $\tilde{u}_1 + \tilde{u}_2 = y$ gilt.

  Wegen $u_1, \tilde{u}_1 \in U_1$ und $u_2, \tilde{u}_2 \in U_2$ gibt es reelle Zahlen $x_1, \tilde{x}_1$ sowie $x_2, x_3, \tilde{x}_2, \tilde{x}_3$ mit

  $$u_1 = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} \quad \text{und} \quad u_2 = \begin{bmatrix} 0 \\ x_2 \\ x_3 \end{bmatrix}, \tilde{u}_1 = \begin{bmatrix} \tilde{x}_1 \\ 0 \\ 0 \end{bmatrix} \quad \text{und} \quad \tilde{u}_2 = \begin{bmatrix} 0 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}$$

  Daraus folgt wegen $u_1 + u_2 = \tilde{u}_1 + \tilde{u}_2$, dass

  $$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix}$$

  ist, d.h. es gilt $x_1 = \tilde{x}_1$ und somit $u_1 = \tilde{u}_1$, sowie $x_2 = \tilde{x}_2$ und $x_3 = \tilde{x}_3$, und somit $u_2 = \tilde{u}_2$. Also haben wir insgesamt, wie gewünscht, $(u_1, u_2) = (\tilde{u}_1, \tilde{u}_2)$ gezeigt.

Wenn man nur zwei Untervektorräume hat, kann man auf folgende Weise charakterisieren, ob der komplette Raum die direkte Summe der beiden gegebenen Untervektorräume ist:[18]

Proposition 5.4.3. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$ und seien $U_1, U_2 \subseteq V$ Untervektorräume von V. Dann sind die folgenden Aussagen äquivalent:*

(i) *Es ist $V = U_1 \oplus U_2$.*

(ii) *Es gilt $V = U_1 + U_2$ und $U_1 \cap U_2 = \{0\}$.*

Den Beweis dieses Resultats lagern wir in die Übungen aus.

5.5 Intermezzo: Relationen

Bevor wir im nächsten Abschnitt über sogenannte **ähnliche Matrizen** sprechen werden, benötigen wir einen kurzen Einschub über sogenannte **Relationen**.

Was ist eine Relation?

Definition 5.5.1 (Relation)**.** Seien X, Y Mengen.

(a) Eine **Relation** R zwischen X und Y ist eine Teilmenge $R \subseteq X \times Y$.

Wenn $X = Y$ ist, dann ist hierfür die Formulierung „R ist eine Relation auf X“ gebräuchlich.

(b) Sei R eine Relation zwischen X und Y, und sei $x \in X$ und $y \in Y$.

Wir sagen, dass x **in der Relation** R **zu** y **steht**, wenn $(x, y) \in R$ gilt. Häufig verwenden wir die Schreibweise $x\ R\ y$ synonym mit $(x, y) \in R$.

Man kann übrigens auch Teilmengen von kartesischen Produkten von n Mengen (für ein $n \in \mathbb{N}^*$) als Relationen – genauer sogenannte n-stellige Relationen – auffassen. Diese tauchen zum Beispiel in der Theorie von sogenannten relationen Datenbanken in der Informatik auf.

Wir werden in dieser Vorlesung aber nur zweistellige Relationen betrachten (so wie wir sie in vorangehender Definition eingeführt haben). Im folgenden diskutieren wir mehrere wichtige Klassen von Relationen: **Partielle und lineare Ordnungen** und **Äquivalenzrelationen**.

[18] Aber Vorsicht: Für mehr als zwei Untervektorräume funktioniert dieselbe Charakterisierung nicht mehr!

Lineare Ordnungen

Definition 5.5.2 (Partielle und lineare Ordnungen)**.** Seien X eine Menge und $R \subseteq X \times X$ eine Relation auf X. Man bezeichnet X als eine **partielle Ordnung**, wenn Sie die folgenden Axiome erfüllt:

(O1) *Reflexivität:* Für alle $x \in X$ gilt $x \; R \; x$.

(O2) *Antisymmetrie:* Für alle $x, y \in X$ gilt die folgende Implikation:

$$(x \; R \; y \quad \wedge \quad y \; R \; x) \qquad \Rightarrow \qquad x = y.$$

(O3) *Transitivität:* Für alle $x, y, z \in X$ gilt die folgende Implikation:

$$(x \; R \; y \quad \wedge \quad y \; R \; z) \qquad \Rightarrow \qquad x \; R \; z.$$

Die Relation R heißt *totale Ordnung* oder *lineare Ordnung*[19], wenn sie eine partielle Ordnung ist und zudem das folgende Axiom erfüllt:

(O4) *Totalität:* Für alle $x, y \in Y$ gilt $x \; R \; y$ oder $y \; R \; x$.

Hier sind einige Beispiele für partielle und totale Ordnungen:

Beispiele 5.5.3. (a) Die übliche Ordnung $\leq$ auf $\mathbb{R}$ ist eine totale Ordnung.[20]

(b) Betrachten Sie die folgende Relation R auf $\mathbb{R}^2$ (d.h. R ist eine Teilmenge von $\mathbb{R}^2 \times \mathbb{R}^2$): Für $x, y \in \mathbb{R}^2$ gelte $x \; R \; y$ genau dann, wenn $x_1 \leq x_2$ und $y_1 \leq y_2$ ist. Anders ausgedrückt heißt dies, dass wir

$$R \coloneqq \{(x, y) \in \mathbb{R}^2 \times \mathbb{R}^2 \mid \; x_1 \leq x_1 \text{ und } y_1 \leq y_2\}$$

definieren. Dann ist R eine partielle Ordnung, aber keine totale Ordnung.

Beweis. Wir müssen zeigen, dass R reflexiv, antisymmetrisch und transitiv ist, dass R aber nicht total ist.

Reflexivititität: Sei $x \in \mathbb{R}^2$. Dann gilt $x_1 \leq x_1$ und $x_2 \leq x_2$, und somit $x \leq x$.

Antisymmetrie: Seien $x, y \in \mathbb{R}^2$, und sei $x \; R \; y$ und $y \; R \; x$. Dann gelten die Ungleichungen

$$x_1 \leq y_1, \quad x_2 \leq y_2 \qquad \text{und} \qquad y_1 \leq x_1, \quad y_2 \leq x_2.$$

[19] Hier ist es wichtig, diesen Begriff nicht mit dem Begriff der linearen Abbildung zu verwechseln. Die beiden Begriffe haben nichts Unmittelbares miteinander zu tun, und das Wort „linear“ kommt lediglich deshalb in beiden Begriffen vor, weil sich dies historisch so entwickelt hat.

[20] Beachten Sie, dass $\leq$ somit – formale betrachtet – eine Teilmenge von $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ ist. Für reellen Zahlen x, y ist $x \leq y$ ein andere Schreibweise für $(x, y) \in \leq$, und diese bedeutet, dass die Zahl x kleiner oder gleich als die Zahl y ist.

Weil die Relation $\leq$ auf $\mathbb{R}$ antisymmetrisch ist, folgt daraus, dass $x_1 = y_1$ und $x_2 = y_2$ ist. Somit gilt $x = y$.

Transitivität: Seien $x, y, z \in \mathbb{R}^2$, und sei $x \ R \ y$ und $y \ R \ z$. Dann gelten die Ungleichungen

$$x_1 \leq y_1, \quad x_2 \leq y_2 \qquad \text{und} \qquad y_1 \leq z_1, \quad y_2 \leq z_2.$$

Weil die Relation $\leq$ auf $\mathbb{R}$ transitiv ist, folgt daraus, dass $x_1 \leq z_1$ und $x_2 \leq z_2$ ist. Somit gilt $x \ R \ z$.

Totalität ist nicht erfüllt: Betrachten wir zum Beispiel die beiden Vektoren

$$x \coloneqq \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \text{und} \qquad y \coloneqq \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Dann gilt wegen $x_1 \not\leq y_1$ nicht $x \ R \ y$, und wegen $y_2 \not\leq x_2$ gilt nicht $y \ R \ x$. Also ist die Ordnung R nicht total. □

(c) Nun führen wir noch eine andere Relation $L \subseteq \mathbb{R}^2 \times \mathbb{R}^2$ auf $\mathbb{R}^2$ ein: Für $x, y \in \mathbb{R}^2$ definieren wir

$$x \ L \ y \qquad :\Leftrightarrow \qquad \big(x_1 < y_1 \quad \vee \quad (x_1 = y_1 \ \wedge \ x_2 \leq y_2)\big).$$

In den Übungen werden Sie zeigen, dass L eine totale Ordnung auf $\mathbb{R}^2$ ist. Man nennt sie die **lexikographische Ordnung**[21] auf $\mathbb{R}^2$.

Folgende Notation ist recht üblich, wenn man mit partiellen oder totalen Ordnungen arbeitet:

Notation 5.5.4. Für eine partielle (oder sogar totale) Ordnung auf einer Menge X verwendet man häufig das Symbol $\leq$ (das Sie bereits von den reellen Zahlen kennen) anstatt eines Buchstabens wie zum Beispiel R. Außerdem verwendet man für $x, y \in X$ dann die Notation $y \geq x$ synonym mit $x \leq y$.

Außerdem ist die folgende Notation, die Sie ebenfalls bereits für reelle Zahlen kennen, insbesondere dann üblich, wenn $\leq$ eine totale Ordnung ist auf X. Für $x, y \in X$ schreibt man häufig $x < y$ als Abkürzung für die Aussage „$x \leq y$ and $x \neq y$“. Außerdem verwendet man die Notation $y > x$ synonym mit $x < y$.

Äquivalenrelationen

Nun führen wir noch eine weitere wichtige Klasse von Relationen ein:

Definition 5.5.5 (Äquivalenzrelationen)**.** Eine Relation $R \subseteq X \times X$ auf einer Menge X heißt **Äquivalenzrelation**, falls sie die folgenden Axiome erfüllt:

(Ä1) *Reflexivität:* Für alle $x \in X$ gilt $x \ R \ x$.

[21] Überlegen Sie sich, weshalb diese Bezeichnung angemessen ist.

(Ä2) *Symmetrie:* Für alle $x, y \in X$ gilt die folgende Implikation:

$$x \, R \, y \qquad \Rightarrow \qquad y \, R \, x.$$

(Ä3) *Transitivität:* Für alle $x, y, z \in X$ gilt die folgende Implikation:

$$(x \, R \, y \quad \wedge \quad y \, R \, z) \qquad \Rightarrow \qquad x \, R \, z.$$

Für Äquivalenzrelationen benutzt man häufiger das Symbol $\sim$ (oder leichte Modifiktionen hiervon) als Buchstaben wie zum Beispiel R. Beachten Sie, dass aus der Symmetrie folgt, dass sogar die folgende Eigenschaft folgt: Für alle $x, y \in X$ gilt die Äquivalenz

$$x \, R \, y \qquad \Leftrightarrow \qquad y \, R \, x.$$

Beispiele für Äquivalenzrelationen besprechen wir in den Übungen und in Abschnitt 5.6.

Zusammenhang zwischen Funktionensgraphen und Relationen

Zum Abschluss dieses Abschnitts besprechen wir noch kurz den Zusammenhang zwischen Relationen und Funktionen bzw. deren Graphen:

Diskussion 5.5.6 (Funktionsgraphen und Relationen)**.** Seien X, Y Mengen und sei $f : X \to Y$ eine Funktion. Wie in Aufgabe 3 auf Tutoriumsblatt 3 nennen wir die Menge

$$G_f \coloneqq \{(x, f(x)) \mid \ x \in X\}$$

den **Graphen** – oder genauer den **Funktionsgraphen** von f.[22] Es gilt $G_f \subseteq X \times Y$, d.h. der Graph von f ist eine Relation zwischen X und Y. Somit kann man sagen: Funkionsgraphen sind Spezialfälle von Relationen.[23] Umgekehrt ist aber nicht jede Relation auch der Graph einer Funktion.

5.6 Ähnlichkeit und Diagonalisierbarkeit von Matrizen

Vermittelung zwischen linearen Abbildung und Ähnlichkeit

Manchmal kommt es vor, dass zwei lineare Abbildungen enger zusammenhängen, als es auf den ersten Blick scheint; dasselbe gilt auch für Matrizen. Dieses Phänomen wird durch die folgenden Begriffe formalisiert:

[22] Die Notation für Funktionsgraphen ist nicht besonders einheitlich. Andere naheliegende Notationen sind zum Beispiel $\mathrm{G}(f)$ oder $\mathrm{Gr}(f)$ oder $\mathrm{Graph}(f)$.

[23] Man kann sogar noch etwas weitergehen und gar nicht zwischen Funktion und Funktionsgraph unterscheiden, sondern sagen, dass eine Funktion schlichtweg ihr Funktionsgraph ist. Das ist eine nützliche Sichtweise, wenn man versucht, die Mathematik von Beginn an komplett formal aufzubauen, denn dabei möchte man möglichst nur mit Mengen auskommen. Man definiert dann zunächst Relationen mit Hilfe von Mengen, und legt dann fest, dass eine Funktion eine Relation ist, die bestimmte zusätzliche Eigenschaften erfüllt.

Definition 5.6.1 (Ähnlichkeit)**.** Sei $\mathbb{K}$ ein Körper.

(a) Seien V, W Vektorräume über $\mathbb{K}$ und seien $R : V \to V$ und $S : W \to W$ lineare Abbildungen. Die beiden Abbildungen R und S heißen **ähnlich**, falls es einen Vektorraumisomorphismus $J : W \to V$ gibt derart, dass

$$R \circ J = J \circ S$$

gilt. Etwas graphischer ausgedrückt bedeutet dies, dass das folgende Diagramm **kommutiert**:[24]

$$\begin{array}{ccc} W & \xrightarrow{S} & W \\ \downarrow{\scriptstyle J} & & \downarrow{\scriptstyle J} \\ V & \xrightarrow{R} & V \end{array}$$

In diesem Fall sagt man auch, dass der Vektorraumisomorphismus J **zwischen** R **und** S **vermittelt**.

(b) Seien $n \in \mathbb{N}^*$ und seien $A, B \in \mathbb{K}^{n \times n}$. Wir sagen, dass A **ähnlich zu** B ist, wenn es eine invertierbare Matrix $T \in \mathbb{K}^{n \times n}$ gibt derart, dass

$$AT = TB$$

gilt.

Wir schreiben $A \sim B$ um auszudrücken, dass A ähnlich zu B ist.

Dass wir Ähnlichkeit von Matrizen untersuchen wollen, ist einer der Gründe, weshalb wir im vorangehenden Abschnitt den Begriff der Äquivalenzrelation eingeführt haben.

Proposition 5.6.2. *Sei $\mathbb{K}$ ein Körper und sei $n \in \mathbb{N}^*$.*
Dann ist die Relation $\sim$ auf $\mathbb{K}^{n \times n}$ eine Äquivalenzrelation.

Beweis. Wir müssen die Eigenschaften einer Äquivalenzrelation nachprüfen:

- *Reflexivität:* Sei $A \in \mathbb{K}^{n \times n}$. Die Einheitsmatrix I_n ist invertierbar, und es gilt

$$AI_n = I_n A.$$

Also ist $A \sim A$.

- *Symmetrie:* Seien $A, B \in \mathbb{K}^{n \times n}$ und sei $A \sim B$. Dann gibt es eine invertierbare Matrix $T \in \mathbb{K}^{n \times n}$ mit der Eigenschaft $AT = TB$. Daraus folgt $BT^{-1} = T^{-1}A$. Weil T^{-1} laut Proposition 4.2.11(d) ebenfalls invertierbar ist, folgt daraus, dass $B \sim A$ gilt.

[24] Man sagt, dass ein Diagramm aus Abbildungen **kommutiert**, wenn verschiedene Pfade im Diagramm, die denselben Startpunkt und denselben Endpunkt haben, stets diesselbe Abbildung beschreiben.

- *Transitivität:* Seien $A, B, C \in \mathbb{K}^{n\times n}$, und sei $A \sim B$ und $B \sim C$. Dann gibt es invertierbare Matrizen $T_1, T_2 \in \mathbb{K}^{n\times n}$ derart, dass

$$AT_1 = T_1B \qquad \text{und} \qquad BT_2 = T_2C$$

gilt. Wie definieren nun $T \coloneqq T_1T_2$. Dies ist ebenfalls eine Matrix in $\mathbb{K}^{n\times n}$ und laut Proposition (c) ist sie ebenfalls invertierbar. Außerdem gilt

$$AT = (AT_1)T_2 = (T_1B)T_2 = T_1(BT_2) = T_1(T_2C) = TC.$$

Also gilt $A \sim C$.

□

Lassen Sie uns kurz zwei einfache Beispiel besprechen.

Beispiele 5.6.3. (a) Die Matrizen

$$A \coloneqq \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad B \coloneqq \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix}$$

in $\mathbb{R}^{2\times 2}$ sind ähnlich. Sei nämlich

$$T \coloneqq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Die Matrix T ist invertierbar, weil $T \cdot T = I_2$ gilt, und man kann direkt nachrechen, dass

$$AT = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix} = TB$$

gilt.

(b) Am vorangehenden Beispiel kann man gut erkennen, was Ähnlichkeit von Matrizen mit der Vermittelung zwischen linearen Abbildungen zu tun hat:

Sei $\mathbb{K} = \mathbb{R}$ und $V = W = \mathbb{R}^2$. Zudem seien A, B und T diesselbe Matrizen wie in Beispiel (a). Für jede Matrix $M \in \mathbb{R}^{2\times 2}$ bezeichnen wir, wie in Proposition 4.2.8(a), mit $L_M : \mathbb{R}^2 \to \mathbb{R}^2$ die lineare Abbildung, die durch $L_M(x) = Mx$ für alle $x \in \mathbb{R}^2$ gegeben ist.

Dann folgt aus der Invertierbarkeit von T, dass L_T ein Vektorraumisomorphismus zwischen $\mathbb{R}^2$ und $\mathbb{R}^2$ ist. Außerdem sind die beiden linearen Abbildungen L_A und L_B ähnlich, denn L_T vermittelt zwischen ihnen; für alle $x \in \mathbb{R}^2$ gilt nämlich

$$(L_A \circ L_T)(x) = L_A(L_T(x)) = ATx = TBx = L_T(L_B(x)) = (L_T \circ L_B)(x),$$

d.h. es ist $L_A \circ L_T = L_T \circ L_B$. Anders ausgedrückt bedeutet das, dass das Diagramm

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{x\mapsto Bx} & \mathbb{R}^2 \\ {\scriptstyle x\mapsto Tx}\downarrow & & \downarrow{\scriptstyle x\mapsto Tx} \\ \mathbb{R}^2 & \xrightarrow{x\mapsto Ax} & \mathbb{R}^2 \end{array}$$

kommutiert.

Diagonalmatrizen

Es ist im Umgang mit Matrizen sehr hilfreich, wenn eine gegebene quadratische Matrix A ähnlich zu einer anderen Matrix ist, welche besonders einfache Gestalt hat. Deshalb wollen wir als nächstes eine Klasse von Matrizen mit solch einfacher Gestalt im Detail besprechen.

Definition 5.6.4 (Diagonalmatrix)**.** Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}^*$. Eine Matrix $D \in \mathbb{K}^{n\times n}$ heißt **Diagonalmatrix**, falls all ihre Einträge, die außerhalb der Diagonalen stehen, gleich 0 sind – d.h., falls

$$A_{jk} = 0$$

für alle $j, k \in \{1, \dots, n\}$ mit $j \neq k$ gilt.

Hier sind ein paar einfache Beispiele:

Beispiele 5.6.5. (a) Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}^*$. Dann sind die Nullmatrix $0 \in \mathbb{K}^{n\times n}$ und die Einheitsmatrix $I_n \in \mathbb{K}^{n\times n}$ Diagonalmatrizen.

(b) Die Matrix

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1+2i \end{bmatrix} \in \mathbb{C}^{3\times 3}$$

ist eine Diagonalmatrix.

(c) Die Matrizen A, B und T aus Beispiel 5.6.3 sind keine Diagonalmatrizen.

Der Grund, weshalb Diagonalmatrizen besonders einfach zu handhaben sind, besteht darin, dass für sie besonders einfache Rechenregeln gelten.

Proposition 5.6.6 (Multiplikation mit Diagonalmatrizen)**.** *Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}^*$. Seien $D, \tilde{D}, \hat{D} \in \mathbb{K}^{n\times n}$ Diagonalmatrizen.*

(a) *Auch $\tilde{D}\hat{D}$ ist eine Diagonalmatrix, und ihre Einträge auf der Diagonalen sind gleich den Produkten der entsprechenden Diagonaleinträge von $\tilde{D}$ und $\hat{D}$, d.h. es gilt*

$$(\tilde{D}\hat{D})_{kk} = \tilde{D}_{kk}\hat{D}_{kk}$$

für alle $k \in \{1, \dots, n\}$.

(b) *Sei $m \in \mathbb{N}^*$ und $A \in \mathbb{K}^{m\times n}$ eine Matrix mit den Spalten $a_1, \dots, a_n \in \mathbb{K}^m$. Dann gilt*

$$AD = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} D = \begin{bmatrix} D_{11}a_1 & \dots & D_{nn}a_n \end{bmatrix}.$$

(c) *Sei $p \in \mathbb{N}^*$ und $B \in \mathbb{K}^{n\times p}$ eine Matrix mit den Zeilen $b_1, \dots, b_n \in \mathbb{K}^{1\times p}$. Dann gilt*

$$DB = D \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} D_{11}b_1 \\ \vdots \\ D_{nn}b_n \end{bmatrix}.$$

Den Beweis kann man einfach durch Nachrechnen führen; wir verzichten deshalb darauf, ihn hier explizit anzugeben. Besser für Ihr Verständnis ist es aber vermutlich ohnehin, wenn Sie die Aussagen der Proposition anhand einiger Beispiele oder konkreter Fälle nachvollziehen. Die erste Aussage kann man zum Beispiel für 2×2-Matrizen gut überprüfen: Es gilt

$$\tilde{D}\hat{D} = \begin{bmatrix} \tilde{D}_{11} & 0 \\ 0 & \tilde{D}_{22} \end{bmatrix} \begin{bmatrix} \hat{D}_{11} & 0 \\ 0 & \hat{D}_{22} \end{bmatrix} = \begin{bmatrix} \tilde{D}_{11}\hat{D}_{11} & 0 \\ 0 & \tilde{D}_{22}\hat{D}_{22} \end{bmatrix}.$$

Als nächstes besprechen wir anhand eines Beispiels, wie man Diagonalmatrizen geometrisch interpretieren kann:

Beispiel 5.6.7. Lassen Sie uns den Vektorraum $\mathbb{R}^2$ über $\mathbb{R}$ betrachten, und sei

$$K \coloneqq \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\} \subseteq \mathbb{R}^2$$

Die Menge K ist der sogenannte **Einheitskreis** in $\mathbb{R}^2$ (d.h. die Kreislinie mit Radius 1, deren Mittelpunkt der Ursprung ist).

Seien nun $\alpha, \beta > 0$ zwei reelle Zahlen, und sei $D \in \mathbb{R}^{2\times 2}$ die Diagonalmatrix mit den Diagonaleinträgen α und β, d.h.

$$D = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}.$$

Lassen Sie uns besprechen, was passiert, wenn wir D mit einem Vektor $x \in \mathbb{R}^2$ multiplizieren: Es gilt

$$Dx = D \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \beta x_2 \end{bmatrix},$$

d.h. D streckt die erste Komponente von x mit dem Faktor α und die zweite Komponente von x mit dem Vektor β.

Wenn wir nun jeden Vektor $x \in K$ mit der Matrix D multiplizieren, dann erhalten wir also eine Ellipse; ihre eine Halbachse liegt auf der x_1-Achse und hat die Länge α,

und ihre andere Halbachse liegt auf der x_2-Achse und hat die Länge β. In Formeln ist diese Ellipse gleich der Menge

$$\{Dx \mid x \in \mathbb{R}^2,\ x_1^2 + x_2^2 = 1\} = \left\{ \begin{bmatrix} \alpha x_1 \\ \beta x_2 \end{bmatrix} \mid x_1^2 + x_2^2 = 1 \right\}$$
$$= \{y \in \mathbb{R}^2 \mid \left(\frac{y_1}{\alpha}\right)^2 + \left(\frac{y_2}{\beta}\right)^2 = 1\}.$$

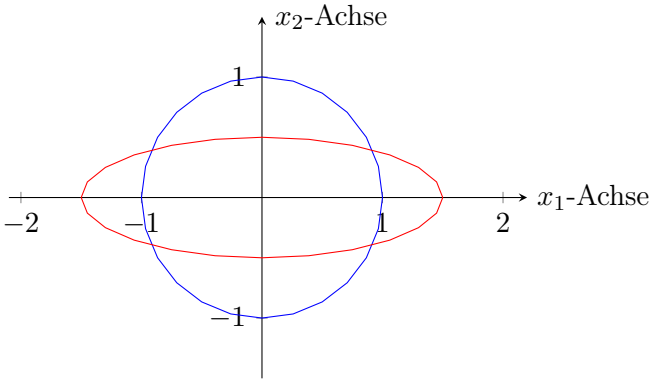

Abbildung 5.6.1: Der blaue Einheitskreis wird hier mit $\alpha = \frac{3}{2}$ und $\beta = \frac{1}{2}$ entlang der Koordinatenachsen gestreckt bzw. gestaucht.

Ähnliches kann man auch im $\mathbb{R}^3$ machen, indem man mit der Einheitssphäre startet und diese in drei Richtungen mit Faktoren $\alpha, \beta, \gamma > 0$ streckt. Die Fläche, die man so erhält ist ein sogenannter Ellipsoid.

Wenn dabei $\alpha = \beta$ wählt – d.h., wenn die Streckung in x_1-Richtung gleich der Streckung in x_2-Richtung ist, und lediglich die Streckung in x_3-Richtung, die durch die Zahl γ beschrieben wird, anders ist, dann erhält man einen sogenannten **Rotationsellipsoiden**; er heißt so, weil er symmetrisch bezüglich Rotation um die x_3-Achse ist.

Übrigens lässt sich die Oberfläche der Erde recht gut durch einen Rotationsellipsoiden annähern.[25]

An obigem Beispiel können Sie einerseits sehen, dass Diagonalmatrizen eine sehr anschauliche Interpretation haben. Andererseits sehen Sie aber auch einen großen Nachteil von Diagonalmatrizen: Sie können zum Beispiel nur solche Streckungen beschreiben, die parallel zu den Koorindatenachsen verlaufen. Den Rotationsellipsoiden, der die Erden beschreibt, kann man zum Beispiel nur dann durch Multiplikation mit einer Diagonalmatrix aus der Einheitssphäre erhalten, wenn man das Koordinatensystem so legt, dass einer der Achsen (im Beispiel war diese konkret die x_3-Achse)

[25] Der Äquatordurchmesser ist ca. 43 Kilometer größer als der Durchmesser zwischen den Polen. In der Notation des Beispiels wäre also, wenn man alle Längen in Kilometern angibt, $\alpha = \beta = \gamma + 21,5$.

die Rotationsachse der Erde ist (und wenn der Urpsrung zugleich der Mittelpunkt der Erde ist).

Diaognalmatrizen werden deshalb nützlicher – und interessanter –, wenn man ihren Zusammenhang mit anderen Matrizen untersucht. Dies tun wir im Folgenden.

Diagonalisierbarkeit

Definition 5.6.8 (Diagonalisierbare Matrizen)**.** Sei $\mathbb{K}$ ein Körper und sei $n \in \mathbb{K}$. Eine Matrix $A \in \mathbb{K}^{n\times n}$ heißt **diagonalisierbar**, wenn es eine Diagonalmatrix $D \in \mathbb{K}^{n\times n}$ gibt derart, dass A und D ähnlich sind.

Lassen sich uns zunächst zwei Beispiele besprechen:

Beispiele 5.6.9. (a) Die Matrix

$$A \coloneqq \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

ist diagonalisierbar. Seien nämlich

$$D \coloneqq \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{und} \qquad T \coloneqq \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Dann kann man nachrechnen, dass T invertierbar ist,[26] und dass

$$AT = TD$$

gilt. Also ist A tatsächlich diagonalisierbar.

Übrigens werden Sie sich vielleicht wundern, wie man denn die passenden Kandidaten für Matrizen D und T finden kann. Dies ist auf den ersten Blick gar nicht so einfach – wir werden aber im gegen Ende des Semesters ein Verfahren besprechen, um geeignete Matrizen D und T zu finden.[27]

(b) Betrachten wir nun die Matrix

$$A \coloneqq \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

Lassen Sie uns zeigen, dass die Matrix A nicht diaognalisierbar ist.

[26] Das kann man zum Beispiel mit Hilfe des Gaußalgorithmus tun. Alternativ kann man auch versuchen, die inverse Matrix zu „raten“ und nachzurechnen, dass es sich tatsächlich um die die inverse Matrix handelt. Man erhält jedenfalls, dass in diesem Beispiel $T^{-1} = \frac{1}{2}T$ gilt.

[27] Beachten Sie übrigens, dass die Matrizen D und T im Allgemeinen nicht eindeutig bestimmt sind. Auch darauf werden wir noch zurück kommen.

Beweis. Wir nehmen widerspruchshalber an, A wäre diaognalisierbar. Dann gibt es eine Diagonalmatrix $D \in \mathbb{R}^{2\times 2}$ und eine invertierbare Matrix $T \in \mathbb{R}^{2\times 2}$ derart, dass $AT = TD$ gilt; anders geschrieben bedeutet das, dass

$$T^{-1}AT = D$$

ist. Nun benutzen wir folgenden Trick: Es ist, wie man direkt nachrechnen kann $A^2 = 0$,[28] und somit erhält man

$$\begin{bmatrix} D_{11}^2 & 0 \\ 0 & D_{22}^2 \end{bmatrix} = D^2 = (T^{-1}AT)\,(T^{-1}AT) = T^{-1}A^2T = 0.$$

Dies zeigt, dass $D_{11} = 0$ und $D_{22} = 0$ gilt, und somit ist $D = 0$. Dann folgt allerdings

$$A = TDT^{-1} = 0,$$

und dies ist ein Widerspruch. □

Auf den Trick im Beweis zu kommen ist nicht einfach, weil Sie die Theorie, die dem Trick zugrunde liegt, noch nicht kennen.[29] Diese Theorie wird in der Linearen Algebra 2 behandelt – es handelt sich um die sogenannte **Jordan-Normalform** von Matrizen, die eng mit dem Konzept der sogenannten **nilpotenten Matrizen** zusammenhängt.

Als nächstes wollen wir an einem Beispiel diskutieren, wie die Formel $AT = TD$, die für diagonalisierbare Matrizen eine wichtige Rolle spielt, geometrisch interpretiert werden kann. Dazu benötigen wir zunächst noch den folgenden Begriff:

Definition 5.6.10 (Drehmatrizen)**.** Eine Matrix $T \in \mathbb{R}^{2\times 2}$ heißt **Drehmatrix**, wenn es eine Zahl $\theta \in \mathbb{R}$ gibt derart, dass

$$T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

gilt.[30]

Beachten Sie: Wenn $T \in \mathbb{R}^{2\times 2}$ eine Drehmatrix ist, d.h. die Form

$$T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

[28] Hierbei ist für $n \in \mathbb{N}^*$ einfach $A^n := A \cdots A$ definiert (mit n Faktoren). Ebenso definiert man übrigens A^0 als Einheitsmatrix.

[29] Aber trotzdem können Sie den Beweis mit Ihrem jetzigen Wissen natürlichen nachvollziehen und auf Richtigkeit überprüfen. Nur fällt der Trick halt ein wenig vom Himmel.

[30] Beachten Sie, dass dieses θ nicht eindeutig bestimmt ist – wenn Sie zum Beispiel 2π zu θ hinzuzählen, erhalten Sie dieselbe Matrix (weil ja die Funktionen cos und sin beide 2π-periodisch sind).

hat für ein $\theta \in \mathbb{R}$, dann gilt

$$Te_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \qquad \text{und} \qquad Te_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix},$$

d.h. Sie sehen, dass T anschaulich eine Drehung im $\mathbb{R}^2$ um den Winkel θ beschreibt.[31]

Es ist wenig überraschend, dass man eine Drehmatrix invertieren kann, indem man in die andere Richtung dreht:

Proposition 5.6.11. *Sei $\theta \in \mathbb{R}$. Dann ist die Drehmatrix*

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

invertierbar, und ihre inverse Matrix ist gleich

$$\begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Beweis. Die behauptete Gleichheit am Ende der Proposition ist klar, weil $\sin(-\theta) = -\sin\theta$ und $\cos(-\theta) = \cos\theta$ ist. Zudem gilt

$$\begin{aligned} &\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} (\cos\theta)^2 + (\sin\theta)^2 & \cos\theta\sin\theta - \sin\theta\cos\theta \\ \sin\theta\cos\theta - \cos\theta\sin\theta & (\sin\theta)^2 + (\cos\theta)^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Also ist

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

rechtsinvertierbar und somit laut Theorem 4.3.7 auch invertierbar. Durch Multiplikation mit der inversen Matrix von links folgt, dass

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

ist. □

Obige Proposition zeigt insbesondere, dass man eine Drehmatrix invertieren kann, indem man sie transponiert.

Mit Hilfe von Drehmatrizen können wir nun das angekündigte Beispiel besprechen, in dem wir Ellipsen konstruieren, deren Halbachsen nicht parallel zu den Koordinatenachsen sind.

[31] Im mathematisch positiven Sinne, d.h. gegen den Uhrzeigersinn.

Beispiele 5.6.12. (a) Wir möchten gerne eine Streckung im $\mathbb{R}^2$ beschreiben, wobei die eine Streckung entlang der ersten Winkelhalbierenden mit Faktor α erfolgen sollen, die die andere Streckung entlang der zweiten Winkelhalbierenden mit Faktor β.

Dies können wir erreichen, indem wir die folgenden drei Schritte hintereinander ausführen:

- Zunächst drehen wir um den Winkel $-\frac{\pi}{4}$; dadurch wird die erste Winkelhalbierende auf die x_1-Achse gedreht, und die zweite Winkelhalbierende wird auf die x_2-Achse gedreht.
- Dann strecken wir mit Hilfe einer Diagonalmatrix mit dem Faktor α ind Richtung der x_1-Achse und mit dem Faktor β in Richtung der x_2-Achse.
- Anschließend drehen wir wieder zurück.

Sei also

$$D \coloneqq \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \qquad \text{sowie} \qquad T \coloneqq \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Die Matrix T beschreibt die Drehung um $\frac{\pi}{4}$ (also die Drehung, die wir im letzten Schritt brauchen), und die Matrix

$$T^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

beschreibt die Drehung um $-\frac{\pi}{4}$.

Die gesamte Streckung, die wir durchführen wollen, wird somit durch Multiplikation mit der Matrix

$$A \coloneqq TDT^{-1}$$

beschrieben.

(b) Nun wenden wir das soeben Besprochene an, um für zwei Zahlen $\alpha, \beta > 0$ eine Ellipse E in $\mathbb{R}^2$ zu beschreiben, die um den Nullpunkt zentriert ist, und die die folgenden Eigenschaften hat:

- Ein Halbachse liegt auf auf der ersten Winkelhalbierenden und hat die Länge α.
- Die zweite Halbachse liegt auf der weiten Winkelhalbierenden und hat die Länge β.

Hierzu sei wieder $K \coloneqq \{x \in \mathbb{R}^2 \mid \ x_1^2 + x_2^2 = 1\}$ die Einheitskreislinie. Unsere Ellipse E ist somit gleich der Menge

$$E = AK \coloneqq \{Ax \mid \ x_1^2 + x_2^2 = 1\},$$

d.h. wir haben eine Beschreibung der Menge E mit Hilfe einer Formel gefunden.

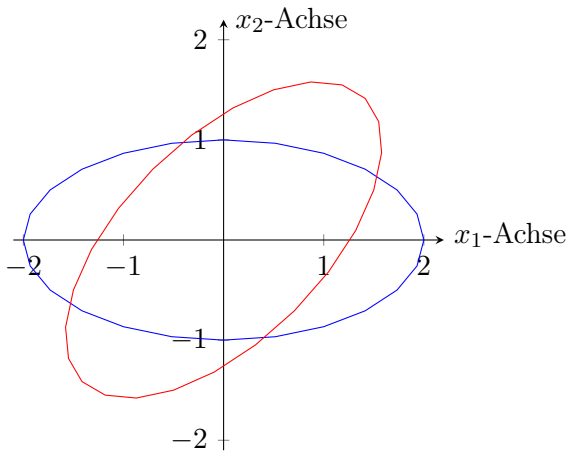

Abbildung 5.6.2: Veranschaulichung der Drehung der blauen Ellipse um $\frac{\pi}{4}$.

Bemerkung 5.6.13. Übrigens kann man die Matrix A sowie die Ellipse E in obigem Beispiel auch etwas expliziter darstellen: Man kann sofort ausrechnen, dass

$$A = TDT^{-1} = \frac{1}{2}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}\alpha & 0\\ 0 & \beta\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}\alpha+\beta & \alpha-\beta\\ \alpha-\beta & \alpha+\beta\end{bmatrix}$$

gilt, und somit ist

$$\begin{aligned} E = AK &\coloneqq \{Ax \mid \ x_1^2 + x_2^2 = 1\} \\ &= \left\{\frac{1}{2}\begin{bmatrix}\alpha(x_1+x_2)+\beta(x_1-x_2)\\ \alpha(x_1+x_2)+\beta(-x_1+x_2)\end{bmatrix} \mid \ x_1^2+x_2^2 = 1.\right\}. \end{aligned}$$

Für viele Anwendungen – zum Beispiel, wenn Sie geometrische Berechungen auf einem Computer durchführen wollen – nützen Ihnen diese expliziten Formeln aber nicht wirklich etwas.[32] Tatsächlich ist es oft viel einfacher und übersichtlicher, abstrakt nur mit den entsprechenden Matrizen zu rechnen, und das konkrete Berechnen der Einträge einfach dem Computer zu überlassen, sobald man explizite Zahlen einsetzt (denn wie man bereits in den Octave-Aufgaben auf den Hausaufgabenblättern sehen konnte, ist es nicht schwierig einem Computer beizubringen, wie man zum Beispiel Matrizen multipliziert).

5.7 Darstellung von linearen Abbildungen mittels Matrizen

Eine Schwierigkeit beim Arbeiten mit linearen Abbildungen besteht darin, dass diese oft abstrakte Objekte sind, mit denen zum Beispiel Computer nur schwer konkret

[32] Zumal Sie noch viel komplizierter werden, wenn man etwas kompliziertere geometrische Situationen betrachtet.

rechnen können. Matrizen hingegen sind sehr konkrete Objekte, mit denen man sehr explizit rechnen kann und für die es zahlreiche Verfahren gibt.

Nun wissen Sie für lineare Abbildungen von $\mathbb{K}^n$ nach $\mathbb{K}^m$ bereits, dass diese sich mit Hilfe von Matrizen beschreiben lassen – dies haben wir in Proposition 4.2.8 besprochen. In diesem Abschnitt besprechen wir, dass man auch lineare Abbildungen zwischen zwei ganz allgemeinen endlich-dimensionalen Vektorräumen mit Hilfe von Matrizen darstellen kann. Dazu muss man sich allerdings sowohl im Definitionsbereich als auch im Bildbereich eine Basis aussuchen.

Spezifikation von linearen Abbildungen mittels Basen

Als Motivation beginnen zuerst damit, dass man eine lineare Abbildung definieren kann, indem man festlegt, wie sie auf einer Basis ihres Definitionsbereichs operiert:

Proposition 5.7.1. *Seien V, W Vektorräume über einem Körper $\mathbb{K}$. Es sei V endlich-dimensional mit $n \coloneqq \dim V \in \mathbb{N}$.*

Wenn $(v_1, \dots, v_n)$ eine Basis von V ist und $w_1, \dots, w_n \in W$ sind, dann gibt es genau eine lineare Abbildung $T : V \to W$ mit der Eigenschaft

$$T(v_k) = w_k$$

für alle $k \in \{1, \dots, n\}$.

Beweis. Existenz: Wir verwenden die Abkürzung $\mathcal{B} \coloneqq (v_1, \dots, v_n)$. Wie bereits früher in der Vorlesung notieren wir mit $\kappa_{\mathcal{B}} : V \to \mathbb{K}^n$ die Koordinatenabbildung bezüglich der Basis $\mathcal{B}$. Außerdem sei $S : \mathbb{K}^n \to W$ gegeben durch

$$S(x) \coloneqq \sum_{j=1}^{n} x_j w_j \qquad \text{für alle } x \in \mathbb{K}^n.$$

Dann ist S linear, also ist auch die Abbildung $T \coloneqq S \circ \kappa_{\mathcal{B}} : V \to W$ linear. Für jedes $k \in \{1, \dots, n\}$ gilt zudem $\kappa_{\mathcal{B}}(v_k) = e_k$, und somit

$$T(v_k) = S(\kappa_{\mathcal{B}}(v_k)) = S(e_k) = w_k.$$

Also erfüllt T die gewünschte Eigenschaft

Eindeutigkeit: Seien nun $T, \tilde{T} : V \to W$ zwei lineare Abbildungen mit der gewünschten Eigenschaft. Wir müssen zeigen, dass $T = \tilde{T}$ gilt. Weil T und $\tilde{T}$ denselben Definitions- und Wertebereich haben, müssen wir lediglich $T(x) = \tilde{T}(x)$ für alle $x \in V$ beweisen. Sei also $x \in V$.

Dann gibt es Skalare $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ mit der Eigenschaft $x = \sum_{k=1}^{n} \alpha_k v_k$. Hieraus folgt

$$T(x) = \sum_{k=1}^{n} \alpha_k T(v_k) = \sum_{k=1}^{n} \alpha_k w_k = \sum_{k=1}^{n} \alpha_k \tilde{T}(v_k) = \tilde{T}(x).$$

Dies zeigt, dass wie behauptet $\tilde{T} = T$ gilt. □

Darstellungsmatrizen

Das folgende Theorem zeigt, wie man eine lineare Abbildung mit Hilfe einer Matrix beschreiben kann, wenn man je eine Basis des Defintionsbereichs und des Wertebereichs fixiert hat.

Theorem 5.7.2. *Seien V, W von Null verschiedene endlich-dimensionale Vektorräume über einem Körper $\mathbb{K}$ mit Basen $\mathcal{B} = (v_1, \dots, v_n)$ von V und $\mathcal{C} = (w_1, \dots, w_m)$ von W (wobei $n, m \in \mathbb{N}^*$ sind). Sei $T : V \to W$ linear. Dann gibt es genau eine Matrix ${}_{\mathcal{C}}T_{\mathcal{B}} \in \mathbb{K}^{m\times n}$ derart, dass das folgende Diagramm kommutiert:*

$$\begin{array}{ccc} V & \xrightarrow{\quad T \quad} & W \\ {\scriptstyle \kappa_{\mathcal{B}}}\downarrow & & \downarrow{\scriptstyle \kappa_{\mathcal{C}}} \\ \mathbb{K}^n & \xrightarrow{\quad x \mapsto {}_{\mathcal{C}}T_{\mathcal{B}}\, x \quad} & \mathbb{K}^m \end{array}$$

Für jedes $k \in \{1, \dots, n\}$ gilt: Die Einträge der k-ten Spalte von ${}_{\mathcal{C}}T_{\mathcal{B}}$ sind diejenigen Skalare $\alpha_1, \dots, \alpha_m \in \mathbb{K}$, für die

$$T(v_k) = \sum_{j=1}^{m} \alpha_j w_j$$

gilt.

Die Matrix ${}_{\mathcal{C}}T_{\mathcal{B}}$ aus dem Theorem nennt man die **Darstellungsmatrix** von T bezüglich der Basen $\mathcal{B}$ und $\mathcal{C}$.

Beweis von Theorem 5.7.2. Eine Matrix ${}_{\mathcal{C}}T_{\mathcal{B}} \in \mathbb{K}^{m\times n}$ hat genau dann die gewünschte Eigenschaft, wenn

$${}_{\mathcal{C}}T_{\mathcal{B}} x = (\kappa_{\mathcal{C}} \circ T \circ \kappa_{\mathcal{B}}^{-1})(x)$$

für alle $x \in \mathbb{K}^n$ gilt. Laut Proposition 4.2.8 gibt es eine solche Matrix. Die Eindeutigkeit kann man sofort durch Anwendungen auf die kanonischen Einheitsvektoren erkennen.

Wir müssen nun noch die behauptete Formel zeigen. Sei also $k \in \{1, \dots, n\}$ und sei $\alpha \in \mathbb{K}^m$ die k-te Spalte von ${}_{\mathcal{C}}T_{\mathcal{B}}$. Dann gilt wegen $e_k = \kappa_{\mathcal{B}}(v_k)$

$$T(v_k) = (T \circ \kappa_{\mathcal{B}}^{-1})(e_k) = \kappa_{\mathcal{C}}^{-1}\big({}_{\mathcal{C}}T_{\mathcal{B}}\, e_k\big) = \kappa_{\mathcal{C}}^{-1}(\alpha) = \sum_{j=1}^{m} \alpha_j w_j.$$

Dies zeigt die Behauptung. □

Lassen Sie uns das Theorem anhand zweier Beispiele etwas anschaulicher besprechen:

Beispiel 5.7.3. Lassen Sie uns den Untervektorraum $V \coloneqq \operatorname{span}(\cos, \sin)$ des Vektorraums $\operatorname{Abb}(\mathbb{R}; \mathbb{R})$ (über $\mathbb{R}$) betrachten. Es ist $\mathcal{B} \coloneqq (\cos, \sin)$ eine Basis von V (die lineare Unabhängigkeit kann man leicht überprüfen, indem man in beide Funktionen die Zahlen 0 und π einsetzt).

Sei $T : V \to V$ gegeben durch $T(f) = f'$ für alle $f \in V$, wobei f' die Ableitung von f bezeichnet (die Sie bereits aus der Schule kennen). Beachten Sie, dass T tatsächlich nach V abbildet, denn: Jedes $f \in V$ lässt sich schreiben als $f = \alpha_1 \cos + \alpha_2 \sin$ für zwei reelle Zahlen α_1, α_2, und somit gilt

$$T(f) = -\alpha_1 \sin + \alpha_2 \cos \in V.$$

Lassen Sie uns nun die Darstellungsmatrix ${}_{\mathcal{B}}T_{\mathcal{B}}$ bestimmen:

- Es gilt $T(\cos) = \cos' = -\sin = 0 \cdot \cos + (-1) \cdot \sin$. Also besteht die erste Spalte der Darstellungsmatrix aus den Zahlen 0 und -1.
- Es gilt $T(\sin) = \sin' = \cos = 1 \cdot \cos + 0 \cdot \sin$. Also besteht die zweite Spalte der Darstellungsmatrix aus den Zahlen 1 und 0.

Insgesamt gilt also

$${}_{\mathcal{B}}T_{\mathcal{B}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Beispiel 5.7.4. Sei V ein von Null verschiedener endlich-dimensionaler Vektorraum über einem Körper $\mathbb{K}$, und seien $\mathcal{B} = (v_1, \ldots, v_n)$ und $\tilde{\mathcal{B}} = (\tilde{v}_1, \ldots, \tilde{v}_n)$ zwei Basen von V (mit einem $n \in \mathbb{N}^*$). Dann ist[33]

$${}_{\tilde{\mathcal{B}}}(\mathrm{id}_V)_{\mathcal{B}} = \begin{bmatrix} \tilde{v}_1 & \ldots & \tilde{v}_n \end{bmatrix}^{-1} \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix}.$$

Beweis. Sei $k \in \{1, \ldots, n\}$ und bezeichne $c_k \in \mathbb{K}^n$ die k-te Spalte von ${}_{\tilde{\mathcal{B}}}(\mathrm{id}_V)_{\mathcal{B}}$. Dann gilt

$$v_k = \mathrm{id}_V(v_k) = \sum_{j=1}^{n} (c_k)_j \tilde{v}_j = \begin{bmatrix} \tilde{v}_1 & \ldots & \tilde{v}_n \end{bmatrix} c_k,$$

und somit

$$c_k = \begin{bmatrix} \tilde{v}_1 & \ldots & \tilde{v}_n \end{bmatrix}^{-1} v_k.$$

Dies zeigt die Behauptung. □

[33] Warum ist die inverse Matrix in dieser Formel überhaupt definiert?

Hintereinanderausführung und Basiswechsel

Die Hintereinanderausführung von linearen Abbildungen korrespondiert zur Matrixmultiplikation der zugehörigen Darstellungsmatrizen, sofern die Basen so gewählt wurden, dass sie „in der Mitte“ zueinander passen:

Theorem 5.7.5. *Seien U, V, W endlich-dimensionale und von Null verschiedene Vektorräume über einem Körper $\mathbb{K}$ mit Dimensionen $p, n, m \in \mathbb{N}^*$. Seien $\mathcal{A} = (u_1, \ldots, u_p)$, $\mathcal{B} = (v_1, \ldots, v_n)$ und $\mathcal{C} = (w_1, \ldots, w_m)$ Basen von U, V und W. Wenn $S : U \to V$ und $T : V \to W$ lineare Abbildungen sind, dann gilt*

$$ {}_{\mathcal{C}}(T \circ S)_{\mathcal{A}} = {}_{\mathcal{C}}T_{\mathcal{B}} \; {}_{\mathcal{B}}S_{\mathcal{A}}. $$

Beweis. Weil das Diagramm

$$\begin{array}{ccccc} U & \xrightarrow{\quad S \quad} & V & \xrightarrow{\quad T \quad} & W \\ {\scriptstyle \kappa_{\mathcal{A}}}\downarrow & & {\scriptstyle \kappa_{\mathcal{B}}}\downarrow & & {\scriptstyle \kappa_{\mathcal{C}}}\downarrow \\ \mathbb{K}^p & \xrightarrow{x \mapsto {}_{\mathcal{B}}S_{\mathcal{A}}\, x} & \mathbb{K}^n & \xrightarrow{x \mapsto {}_{\mathcal{C}}T_{\mathcal{B}}\, x} & \mathbb{K}^m \end{array}$$

kommutiert, kommutiert auch das folgende Diagram:

$$\begin{array}{ccc} U & \xrightarrow{\quad T \circ S \quad} & W \\ {\scriptstyle \kappa_{\mathcal{A}}}\downarrow & & \downarrow{\scriptstyle \kappa_{\mathcal{C}}} \\ \mathbb{K}^p & \xrightarrow{x \mapsto {}_{\mathcal{C}}T_{\mathcal{B}} \; {}_{\mathcal{B}}S_{\mathcal{A}}\, x} & \mathbb{K}^m \end{array}$$

Wegen der Eindeutigkeit von ${}_{\mathcal{C}}(T \circ S)_{\mathcal{A}}$ (Theorem 5.7.2) folgt daraus, dass

$$ {}_{\mathcal{C}}(T \circ S)_{\mathcal{A}} = {}_{\mathcal{C}}T_{\mathcal{B}} \; {}_{\mathcal{B}}S_{\mathcal{A}} $$

ist. □

Mit Hilfe der Formel aus dem vorangehenden Theorem kann man auch angeben, wie man bei einer Darstellungsmatrix die Basis austauschen kann:

Korollar 5.7.6. *Seien V, W endlich-dimensionale und von Null verschiedene Vektorräume über einem Körper $\mathbb{K}$ mit Dimensionen $n, m \in \mathbb{N}^*$. Seien $\mathcal{B} = (v_1, \ldots, v_n)$ und $\tilde{\mathcal{B}} = (\tilde{v}_1, \ldots, \tilde{v}_n)$ Basen von V, sowie $\mathcal{C} = (w_1, \ldots, w_m)$ und $\tilde{\mathcal{C}} = (\tilde{w}_1, \ldots, \tilde{w}_m)$ Basen von W. Sei $T : V \to W$ linear. Dann gilt*

$$\begin{aligned} {}_{\tilde{\mathcal{C}}}T_{\tilde{\mathcal{B}}} &= {}_{\tilde{\mathcal{C}}}(\mathrm{id}_W)_{\mathcal{C}} \; {}_{\mathcal{C}}T_{\mathcal{B}} \; {}_{\mathcal{B}}(\mathrm{id}_V)_{\tilde{\mathcal{B}}} \\ &= \begin{bmatrix} \tilde{w}_1 & \ldots & \tilde{w}_m \end{bmatrix}^{-1} \begin{bmatrix} w_1 & \ldots w_m \end{bmatrix} \; {}_{\mathcal{C}}T_{\mathcal{B}} \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix}^{-1} \begin{bmatrix} \tilde{v}_1 & \ldots \tilde{v}_n \end{bmatrix}. \end{aligned}$$

Beweis. Die erste Gleichheit folgt aus Theorem 5.7.5, und die zweite Gleichheit aus Beispiel 5.7.4. □

5.8 Ergänzungen

Beweis von Theorem 5.2.15

Den Beweis von Theorem 5.2.15 hatten wir in der Vorlesung nicht behandelt. Wenn Sie möchten, können Sie ihn stattdessen hier nachlesen:

Beweis von Theorem 5.2.15. Dass $U_1 + U_2$ tatsächlich ein Untervektorraum von V ist, haben Sie auf Hausaufgabenblatt 7 in Aufgabe 3(f) bewiesen.

Um die Dimensionsformel zu beweisen verwenden wir die Abkürzungen

$$\begin{aligned} s &\coloneqq \dim(U_1 + U_2) & d &\coloneqq \dim(U_1 \cap U_2) \\ d_1 &\coloneqq \dim U_1 & d_2 &\coloneqq \dim U_2. \end{aligned}$$

Es gilt sei $(u_1, \dots, u_s)$ eine Basis von $U_1 \cap U_2$. Wir können diese nun einerseits zu einer Basis

$$(u_1, \dots, u_s, v_1, \dots, v_m)$$

von U_1 mit $m = d_1 - s \in \mathbb{N}$ ergänzen, und andererseits zu einer Basis

$$(u_1, \dots, u_s, w_1, \dots, w_n)$$

mit $n = d_2 - s \in \mathbb{N}$. Der Aufspann des Tupels

$$(u_1, \dots, u_s, v_1, \dots, v_m, w_1, \dots, w_n)$$

ist gleich $U_1 + U_2$. Wenn wir also zeigen können, dass dieses Tupel linear unabhängig ist, dann ist es eine Basis von $U_1 + U_2$ und somit sind wir fertig, weil dann

$$\dim(U_1 + U_2) + \dim(U_1 \cap U_2) = s + m + n + s = (m + s) + (n + s) = d_1 + d_2$$

gilt.

Um die gewünschte lineare Unabhängigkeit zu zeigen, seien $\alpha_1, \dots, \alpha_s$ sowie $\beta_1, \dots, \beta_m$ und $\gamma_1, \dots, \gamma_n$ Skalare in $\mathbb{K}$, für welche

$$\sum_{k=1}^{s} \alpha_k v_k + \sum_{k=1}^{m} \beta_k v_k + \sum_{k=1}^{n} \gamma_k w_k = 0$$

gilt. Der Vektor $\sum_{k=1}^{n} \gamma_k w_k$ liegt in U_2. Wenn wir ihn in obiger Gleichung isolieren, sehen wir, dass er auch in U_1 – und somi in $U_1 \cap U_2$ liegt. Somit gibt es Skalare $\tilde{\alpha}_1, \dots, \tilde{\alpha}_s \in \mathbb{K}$ derart, dass

$$\sum_{k=1}^{n} \gamma_k w_k = \sum_{k=1}^{s} \tilde{\alpha}_k u_k$$

gilt. Wegen der linearen Unabhängigkeit von $(u_1, \ldots, u_s, w_1, \ldots, w_n)$ folgt hieraus, dass alle $\tilde{\alpha}_k$ und alle γ_k gleich 0 sind. Damit ist also

$$\sum_{k=1}^{s} \alpha_k v_k + \sum_{k=1}^{m} \beta_k v_k = 0,$$

und wegen der linearen Unabhängigkeit von $(u_1, \ldots, u_s, v_1, \ldots, v_m)$ folgt hieraus, dass auch alle Skalare α_k und β_k gleich 0 sind. Dies zeigt die behauptete lineare Unabhängigkeit von

$$(u_1, \ldots, u_s, v_1, \ldots, v_m, w_1, \ldots, w_n),$$

womit der Beweis vollständig ist. □

Literaturhinweise

Wir haben zahlreiche Resultate in endlich-dimensionalen Vektorräumen mit Hilfe von Koordinatenabbildungen bewiesen – hier haben wir uns zunutze gemacht, dass man in $\mathbb{K}^n$ viele Konzepte der Linearen Algebra mithilfe von Matrizen ausdrücken kann, und dass jede Matrix eine (reduzierte) Zeilenstufenform besitzt.

Mit Hilfe dieses Ansatzes haben wir uns einige technische Resultate gespart, die man auch direkt in abstrakten Vektorräumen zeigen könnte. Dazu gehört insbesondere der sogenannte **Austauschsatz von Steinitz**, der manchmal nützlich ist. Sie können ihn zum Beispiel in [Beu14, Abschnitt 3.3.2] nachlesen.

Kapitel 6

Theorie linearer Gleichungen

Einstiegsfragen. (a) Kennen Sie eine lineare Abbildung von $\mathbb{R}^4$ nach $\mathbb{R}^4$, die injektiv, aber nicht surjektiv ist? Und eine, die surjektiv, aber nicht injektiv ist?

(b) Sie wissen bereits, dass die Ursprungsebenen im $\mathbb{R}^3$ genau die zwei-dimensionalen Untervektorräume des $\mathbb{R}^3$ sind. Finden Sie eine Möglichkeit, um auch diejenigen Ebenen zu beschreiben, die nicht durch den Ursprung verlaufen?

(c) Wie sieht die Menge aller Lösungen eines linearen Gleichungssystem aus?

(d) Was verstehen Sie unter den Begriffen „Existenz“ und „Eindeutigkeit“ der Lösung(en) einer linearen Gleichung?

(e) Können Sie jede Ebene im $\mathbb{R}^3$ als Lösung eines linearen Gleichungssystems schreiben?

6.1 Kern und Bild

In diesem Abschnitt geht es um zwei Mengen, die zu einer linearen Abbildung gehören: den sogenannten *Kern* und das *Bild*. Bevor wir diese beiden Objekte definieren, besprechen wir zuerst zwei allgemeinere Begriffe, die man für beliebige Abbildungen definieren kann und die somit nicht direkt mit linearer Algebra zu tun haben.

Vorbemerkungen: Bild und Urbild

Definition 6.1.1 (Bild und Urbild)**.** Seien X, Y Mengen und sei $f : X \to Y$.

(a) Für eine Menge $A \subseteq X$ nennt man

$$f(A) \coloneqq \{f(a) \mid \ a \in A\} = \{y \in Y \mid \ \exists a \in A : \ y = f(a)\}$$

das **Bild von A unter f**.

Das Bild von X unter f – also die Menge $f(X)$ – bezeichnet man manchmal auch einfach als das **Bild von** f.

(b) Für eine Menge $B \subseteq Y$ nennt man

$$f^{-1}(B) \coloneqq \{x \in X \mid f(x) \in B\}$$

das **Urbild von** B **unter** f.

Beispiele 6.1.2. (a) Sei $X = \{1, 2, 3, 4\}$ und $Y = \{3, 4, 5, 6\}$. Wir definieren $f : X \to Y$ durch die Tabelle

x	1	2	3	4
$f(x)$	6	3	6	4.

Dann gilt zum Beispiel

$$f(\emptyset) = \emptyset, \quad f(\{1\}) = (\{f\}), \quad f(\{1, 2\}) = \{3, 6\},$$
$$f(\{1, 2, 3\}) = \{3, 6\}, \quad f(\{1, 2, 3, 4\}) = \{3, 4, 6\}.$$

Zudem gilt zum Beispiel

$$f^{-1}(\emptyset) = \emptyset, \quad f^{-1}(\{5\}) = \emptyset, \quad f^{-1}(\{3\}) = \{2\},$$
$$f^{-1}(\{6\}) = \{1, 3\}, \quad f^{-1}(\{5, 6\}) = \{1, 3\}.$$

(b) Sei $g : \mathbb{R} \to \mathbb{R}$ gegeben durch $g(x) = x^2$ für alle $x \in \mathbb{R}$. Dann gilt zum Beispiel

$$g(\mathbb{R}) = [0, \infty), \quad g([-1, 1]) = g([0, 1]) = [0, 1], \quad q([2, 3]) = [4, 9],$$

sowie

$$g^{-1}((0, 4]) = [-2, 0) \cup (0, 2], \quad g^{-1}(\{9\}) = \{-3, 3\}, \quad g^{-1}([-1, 0]) = \{0\}.$$

Aus der Definition der Begriffe Bild und Urbild folgt sofort die folgende Proposition:

Proposition 6.1.3. *Seien X, Y Mengen und sei $f : X \to Y$*

(a) *Die Abbildung f ist genau dann injektiv, wenn für jedes $y \in Y$ das Urbild $f^{-1}(\{y\})$ höchstens ein Element hat.*

(b) *Die Abbildung f ist surjektiv genau dann, wenn ihr Bild $f(X)$ gleich Y ist, genau dann, wenn für jedes $y \in Y$ das Urbild $f^{-1}(\{y\})$ mindestens ein Element hat.*

Bemerkung 6.1.4. Sei $f : X \to Y$ eine Abbildung von einer Menge X in eine Menge Y.

Beachten Sie unbedingt, dass das Urbild $f^{-1}(B)$ einer Menge $B \subseteq Y$ auch dann definiert ist, wenn f nicht bijektiv ist (und f somit keine Umkehrfunktion besitzt).

Zur Unterscheidung der Notationen für Urbild und Umkehrfunktion muss man den Kontext verwenden:

- Wenn $y \in Y$ ist, ist mit $f^{-1}(y)$ die Umkehrfunktion f, ausgewertet an der Stelle y, gemeint. Diese Notation ergibt somit nur Sinn, wenn f bijektiv ist.
- Wenn aber $B \subseteq Y$ ist, ist mit $f^{-1}(B)$ das Urbild von B unter f gemeint. Diese Notation ergibt immer Sinn, egal ob f bijektiv ist.

Sie sollten sich nun die folgenden beiden Fragen stellen um zu überprüfen, ob Sie das richtig verstanden haben:

- Wenn $y \in Y$ ist – was ist dann mit $f^{-1}(\{y\})$ gemeint?
- Sei $B \subseteq Y$ und sei außerdem f bijektiv ist. Oben haben wir gesagt, dass mit $f^{-1}(B)$ tatsächlich das Urbild von B unter f gemeint ist. Aber da $f^{-1} : Y \to X$ ja auch die Umkehrfunktion von f bezeichnet, könnte mit der Notation $f^{-1}(B)$ ja auch das Bild von B unter der Funktion f^{-1} gemeint sein – woher weiß man nun, welches von beiden wirklich gemeint ist?

Der Kern einer linearen Abbildung

Nun kommen wir zu linearen Abbildungen.

Definition 6.1.5 (Kern)**.** Seien V, W Vektorräume über einem Körper $\mathbb{K}$ und sei $T : V \to W$ linear. Die Menge

$$\ker T \coloneqq \{v \in V \mid T(v) = 0\}$$

heißt der **Kern** von T.

Anders gesprochen ist der Kern von T also gleich der Menge $T^{-1}(\{0\})$ (also gleich dem Urbild der Menge $\{0\}$ unter T). Hier sind einige sehr nützliche Eigenschaften des Kerns einer linearen Abbildung:

Proposition 6.1.6. *Seien V, W Vektorräume über einem Körper $\mathbb{K}$ und sei $T : V \to W$ linear.*

(a) *Es ist $\ker T$ ein Untervektorraum von V.*

(b) *Die Abbildung T ist genau dann injektiv, wenn $\ker T = \{0\}$ gilt.*

Beweis. (a) Wegen $T(0) = 0$ gilt $0 \in \ker T$.

Seien nun $v_1, v_2 \in \ker T$. Dann gilt $T(v_1) = 0$ und $T(v_2) = 0$ und somit

$$T(v_1 + v_2) = T(v_1) + T(v_2) = 0 + 0 = 0.$$

Also ist auch $v_1 + v_2 \in \ker T$.

Zuletzt sei $v \in \ker T$ und $\alpha \in \mathbb{K}$. Dann gilt $T(v) = 0$ und somit

$$T(\alpha v) = \alpha T(v) = \alpha \cdot 0 = 0.$$

Also ist auch $\alpha v \in U$.

(a) „$\Rightarrow$“ Sei T injektiv. Um $\ker T = \{0\}$ zu beweisen, müssen wir zwei Inklusionen zeigen:

- „$\supseteq$“ Es gilt $0 \in \ker T$ wegen Aussage (a). Also ist $\ker T \supseteq \{0\}$.
- „$\subseteq$“ Sei $v \in \ker T$. Dann gilt $T(v) = 0 = T(0)$. Wegen der Injektivität von T folgt daraus $v = 0$, also $v \in \{0\}$.

„$\Leftarrow$“ Gelte nun $\ker T = \{0\}$. Seien $v_1, v_2 \in V$ mit $T(v_1) = T(v_2)$. Wir müssen $v_1 = v_2$ zeigen.

Wegen der Linearität von T gilt

$$0 = T(v_2) - T(v_1) = T(v_1 - v_2),$$

also $v_1 - v_2 \in \ker T$. Wegen $\ker T = \{0\}$ folgt daraus $v_1 - v_2 = 0$, also $v_1 = v_2$. □

Beispiele 6.1.7. (a) Betrachten Sie die Abbildungen $T : \mathbb{C}^2 \to \mathbb{C}$, die durch

$$T(x) = \begin{bmatrix} i & 2 \end{bmatrix} x = ix_1 + 2x_2$$

für alle $x \in \mathbb{C}^2$ gegeben ist.

Für einen Vektor $x \in \mathbb{C}^2$ gilt $x \in \ker T$ genau dann, wenn $T(x) = 0$ ist; letztere Aussage ist äquivalent zu $ix_1 + 2x_2 = 0$, und dies wiederum ist äquivalent zu $x_1 = 2ix_2$. Also ist

$$\ker T = \left\{ x \in \mathbb{C}^2 \mid \ x_1 = 2ix_2 \right\} = \left\{ \begin{bmatrix} 2i\gamma \\ \gamma \end{bmatrix} \mid \ \gamma \in \mathbb{C} \right\}.$$

Weil es in $\ker T$ Elemente gibt, die nicht der Nullvektor sind, ist T nicht injektiv.

(b) Sei $\mathbb{K}$ ein Körper, $m, n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{m \times n}$. Wir betrachten die lineare Abbildungen $L_A : \mathbb{K}^n \to \mathbb{K}^m$, die jedes $x \in \mathbb{K}^n$ auf Ax schickt. Dann ist

$$\ker L_A = \{x \in \mathbb{K}^n \mid \ L_A(x) = 0\} = \{x \in \mathbb{K}^n \mid \ Ax = 0\}.$$

Diese Menge wird häufig auch einfach als der **Kern der Matrix** A bezeichnet und mit $\ker A$ notiert.

Wenn $m = n$ ist und die Matrix A außerdem invertierbar ist, dann ist für $x \in \mathbb{K}^n$ die Gleichung $Ax = 0$ äquivalent zu $x = A^{-1}0$, also zu $x = 0$. Somit ist $\ker L_A$ in diesem Fall gleich 0, d.h., L_A ist in diesem Fall injektiv.[1]

(c) Lassen Sie uns (nur für dieses Beispiel) mit $\mathcal{D} \subseteq \mathrm{Abb}(\mathbb{R};\mathbb{R})$ die Menge aller Funktionen von $\mathbb{R}$ nach $\mathbb{R}$ bezeichnen, deren Ableitung an jedem Punkt existiert.[2] Man kann zeigen, dass $\mathcal{D}$ ein Untervektorraum von $\mathrm{Abb}(\mathbb{R};\mathbb{R})$ ist.[3] Sei nun

$$\begin{aligned} T : \mathcal{D} &\to \mathrm{Abb}(\mathbb{R};\mathbb{R}), \\ f &\mapsto f' \end{aligned}$$

diejenige Abbildung, die jeder Funktion $f \in \mathcal{D}$ ihre Ableitung zuordnet. Wie Sie aus der Schule wissen,[4] gilt $(f+g)' = f' + g'$ für alle Funktionen f, g, deren Ableitung existiert; ebenso gilt $(\alpha f)' = \alpha f'$ für alle Funktionen f, deren Ableitung existiert, und für alle $\alpha \in \mathbb{R}$. Somit ist T linear.

Eine Funktion $f \in \mathcal{D}$ liegt genau dann in $\ker T$, wenn $f' = 0$ gilt, d.h. wenn die Ableitung f' von f die konstante Nullfunktion ist. Dies bedeutet, dass f überall die Steigung 0 besitzt, also eine konstante Funktion ist.[5] Somit gilt

$$\ker T = \{f : \mathbb{R} \to \mathbb{R} \mid \ f \text{ ist konstant}\}.$$

Weil es Elemente von $\ker T$ gibt, die nicht 0 sind, ist T also nicht injektiv.

(d) Betrachten Sie die Abbildung

$$\begin{aligned} T : \mathrm{Abb}(\mathbb{R};\mathbb{R}) &\to \mathbb{R}^2, \\ f &\mapsto \begin{bmatrix} f(-1) \\ f(1) \end{bmatrix}. \end{aligned}$$

Die Abbildung T ist linear,[6] und eine Funktion $f \in \mathrm{Abb}(\mathbb{R};\mathbb{R})$ liegt genau dann in $\ker T$, wenn $f(-1) = 0$ und $f(1) = 0$ gilt. Das heißt, es ist

$$\ker T = \{f \in \mathrm{Abb}(\mathbb{R};\mathbb{R}) \mid \ f(-1) = f(1) = 0\}.$$

Weil zum Beispiel die Funktion $g : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2 - 1$, in $\ker T$ liegt, aber nicht 0 ist, ist die Abbildung T also nicht injektiv.

[1] Übrigens ist es sehr leicht zu sehen, dass in diesem Fall L_A auch surjektiv ist. Können Sie das ebenfalls beweisen?

[2] Aus der Schule wissen Sie bereits, was man unter der Ableitung einer Funktion bezeichnet. An dieser Stelle sei daran erinnert, dass nicht jede Funktion eine Ableitung besitzt – d.h. betrachten wir hier nur die Menge $\mathcal{D}$ von Funktionen, die man tatsächlich (auf ganz $\mathbb{R}$) ableiten kann.

[3] Das folgt aus grundlegenden Eigenschaften der Ableitung, die Sie in der Analysis 1 lernen werden.

[4] Oder wissen sollten.

[5] Dieser Satz beschreibt natürlich nur eine geometrische Intuition. In der Analysis 1 werden Sie einen exakten Beweis für diese Aussage kennenlernen.

[6] Es sollte mittlerweile ein Leichtes für Sie sein zu zeigen, dass T tatsächlich linear ist. Probieren Sie es am besten gleich aus! Falls Sie dabei Schwierigkeiten haben, können Sie zum Beispiel in eine der Sprechstunden kommen.

Bild und Rang einer linearen Abbildung

Nach dem Kern besprechen wir nun das Bild einer linearen Abbildung. Hierfür führen wir keine eigenen Notation ein.[7]

Genau wie der Kern ist auch das Bild einer linearen Abbilungen immer ein Untervektorraum.

Proposition 6.1.8. *Seien V, W Vektorräume über einem Körper $\mathbb{K}$ und sei $T : V \to W$ linear. Dann ist das Bild $T(V)$ ein Untervektorraum von W.*

Beweis. Wegen $0 = T(0) \in T(V)$ ist der Nullvektor in $T(V)$ enthalten.

Seien nun $w_1, w_2 \in T(V)$. Dann gibt es $v_1, v_2 \in V$ mit der Eigenschaft $T(v_1) = w_1$ und $T(v_2) = w_2$. Somit folgt

$$w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2) \in T(V),$$

also ist auch $w_1 + w_2$ im Bild von T enthalten.

Zuletzt sei $w \in T(V)$ sowie $\alpha \in \mathbb{K}$. Dann gibt es ein $v \in V$ mit $T(v) = w$, und somit gilt

$$\alpha w = \alpha T(v) = T(\alpha v) \in T(V);$$

also ist auch αw im Bild von T enthalten. □

Die folgende Terminologie ist häufig praktisch:

Definition 6.1.9 (Rang einer linearen Abbildung)**.** Seien V, W Vektorräume über einem Körper $\mathbb{K}$ und sei $T : V \to W$ linear. Falls $T(V)$ endlich-dimensional ist, dann sagt man T **besitzt endlichen Rang** und nennt die Zahl

$$\operatorname{rang}(T) \coloneqq \dim T(V)$$

den **Rang von** T.

Lassen Sie Bild und Rang in einem Beispiel besprechen:

Beispiel 6.1.10. Betrachten wir wieder die lineare Abbildung

$$\begin{aligned} T : \operatorname{Abb}(\mathbb{R}; \mathbb{R}) &\to \mathbb{R}^2, \\ f &\mapsto \begin{bmatrix} f(-1) \\ f(1) \end{bmatrix}. \end{aligned}$$

Weil das Bild von T ein Untervektorraum von $\mathbb{R}^2$ ist, besitzt es endliche Dimension. Also besitzt T endlichen Rang.

[7] Manche Leuten bezeichnen das Bild einer linearen Abbildung T aber zum Beispiel mit dem Symbol $\operatorname{bild}(T)$, oder mit $\operatorname{rg}(T)$ (für Englisch „range“), oder mit $\operatorname{im}(T)$ (für Englisch „image“.

Lassen Sie uns das Bild von T bestimmen: Betrachten wir dazu die Funktionen $f, g : \mathbb{R} \to \mathbb{R}$, die durch

$$f(x) = x - 1 \qquad \text{und} \qquad g(x) = x + 1$$

für alle $x \in \mathbb{R}$ gegeben sind. Dann gilt

$$T(f) = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \qquad \text{und} \qquad T(g) = \begin{bmatrix} 0 \\ 2 \end{bmatrix},$$

also ist $\big(T(f), T(g)\big)$ ein linear unhängiges Tupel, das aus zwei Elementen des Bildes von T besteht. Damit hat das Bild von T mindestens Dimension 2; es hat aber auch höchstens Dimensions 2, weil es ein Untervektorraum von $\mathbb{R}^2$ ist. Daraus folgt $T(\mathrm{Abb}(\mathbb{R}; \mathbb{R})) = \mathbb{R}^2$, also ist T surjektiv und hat Rang 2.

Generell besteht der folgenden Zusammenhang zwischen dem Rang und der Surjektivität einer linearen Abbildung.

Proposition 6.1.11. *Seien V, W Vektorräume über einem Körper $\mathbb{K}$ und sei $T : V \to W$ linear. Außerdem sei W endlich-dimensional.*

Dann ist das Bild $T(V)$ ebenfalls endlich dimensional, und T ist genau dann surjektiv, wenn der Rang von T gleich $\dim W$ *ist.*

Beweis. Als Untervektorraum des endlich-dimensionalen Vektorraums W ist $T(V)$ selbst endlich-dimensional. Außerdem ist T genau dann surjektiv, wenn $T(V) = W$ gilt, und letzteres ist laut Proposition 5.2.13(b) äquivalent dazu, dass $\dim T(V) = \dim W$ gilt. □

Der Rangsatz

Die Dimension des Definitionsbereichs einer linearen Abbildung lässt sich schreiben als die Summe der Dimension des Kerns und des Bildes; dies ist der Inhalt des folgendes Theorems:

Theorem 6.1.12 (Rangsatz)**.** *Seien V, W Vektorräume über einem Körper $\mathbb{K}$ und sei $T : V \to W$ linear. Außerdem sei V endlich-dimensional. Dann ist auch $T(V)$ endlich-dimensional*[8] *und es gilt*

$$\dim V = \dim \ker T + \dim T(V).$$

Beweis. Weil $\ker T$ ein Untervektorraum des endlich-dimensionalen Vektorraums V ist, ist $\ker T$ selbst endlich-dimensional. Sei $(v_1, \ldots, v_k)$ eine Basis von $\ker T$, mit $k := \dim \ker T \in \mathbb{N}$. Laut Basisergänzungssatz (Theorem 5.2.11) können wir diese zu einer Basis $(v_1, \ldots, v_k, u_1, \ldots, u_\ell)$ von V ergänzen, wobei $\ell \in \mathbb{N}$ ist.

[8] Anders ausgedrückt: T besitzt endlichen Rang.

Mit diesen Bezeichnungen gilt also $\dim V = k + \ell$ und $\dim \ker T = k$. Lassen Sie und nun zeigen, dass $\mathcal{C} \coloneqq (T(u_1), \dots, T(u_\ell))$ eine Basis von $T(V)$ ist; dies zeigt $\dim T(V) = \ell$ und somit die Behauptung.

Wir zeigen zuerst, dass $\mathcal{C}$ ein Erzeugendensystem von $T(V)$ ist, d.h., dass $\operatorname{span} \mathcal{C} = T(V)$ gilt

- „$\subseteq$“ Sei $w \in \operatorname{span} \mathcal{C}$; dann gibt es Skalare $\beta_1, \dots, \beta_\ell \in \mathbb{K}$ mit der Eigenschaft
$$w = \sum_{j=1}^{\ell} \beta_j T(u_j) = T\Big(\sum_{j=1}^{\ell} \beta_j u_j \Big) \in T(V).$$
- „$\supseteq$“ Sei $w \in T(V)$. Dann gibt es ein $v \in V$ mit $w = T(v)$. Weil das Tupel $(v_1, \dots, v_k, u_1, \dots, u_\ell)$ eine Basis von V ist, gibt es Skalare $\alpha_1, \dots, \alpha_k \in \mathbb{K}$ und $\beta_1, \dots, \beta_\ell \in \mathbb{K}$ mit der Eigenschaft
$$v = \sum_{j=1}^{k} \alpha_j v_j + \sum_{j=1}^{\ell} \beta_j u_j.$$
Indem wir auf diese Gleichung T anwenden, erhalten wir
$$T(v) = \sum_{j=1}^{\ell} \beta_j T(u_j) \in \operatorname{span} \mathcal{C};$$
hierbei haben wir verwendet, dass T linear ist und dass $T(v_j) = 0$ für alle $j \in \{1, \dots, k\}$ gilt.

Nun zeigen wir noch, dasss $\mathcal{C}$ linear unabhängig ist. Seien hierzu $\beta_1, \dots, \beta_\ell \in \mathbb{K}$ mit der Eigenschaft $\sum_{j=1}^{\ell} \beta_j T(u_j) = 0$. Dann folgt
$$T\Big(\sum_{j=1}^{\ell} \beta_j u_j \Big) = 0,$$
also ist $\sum_{j=1}^{\ell} \beta_j u_j \in \ker T$. Weil $(v_1, \dots, v_k)$ eine Basis von $\ker T$ ist, gibt es somit Koeffizienten $\alpha_1, \dots, \alpha_k \in \mathbb{K}$ mit der Eigenschaft
$$\sum_{j=1}^{\ell} \beta_j u_j = \sum_{j=1}^{k} \alpha_j v_j,$$
und dies wiederum impliziert
$$\sum_{j=1}^{k} -\alpha_j v_j + \sum_{j=1}^{\ell} \beta_j u_j = 0.$$
Weil $(v_1, \dots, v_k, u_1, \dots, u_\ell)$ linear unabhängig ist, folgt daraus, dass die Skalare $\alpha_1, \dots, \alpha_k$ und $\beta_1, \dots, \beta_\ell$ alle gleich 0 sind.

Insbesondere sind alle die Skalare $\beta_1, \dots, \beta_\ell$ gleich 0, was die behauptete lineare Unabhängigkeit von $\mathcal{C}$ zeigt. □

Ein wichtige Konsequenz das Rangsatzes ist die folgende:

Korollar 6.1.13. *Seien V, W Vektorräume über einem Körper $\mathbb{K}$ und sei $T : V \to W$ linear. Wenn V und W endlich-dimensional sind und $\dim V = \dim W$ gilt, dann sind die folgenden Aussagen äquivalent:*

(i) *Die Abbildung T ist bijektiv.*

(ii) *Die Abbildung T ist injektiv.*

(iii) *Die Abbildung T ist surjektiv.*

Beweis. Es genügt offenbar, die Äquivalenz „(ii) $\Leftrightarrow$ (iii)“ zu beweisen.

„(ii) $\Rightarrow$ (iii)“ Sei T injektiv. Dann ist $\dim \ker T = 0$, und somit folgt aus dem Rangsatz, dass

$$\dim V = \dim T(V)$$

gilt. Wegen $\dim V = \dim W$ ist also $\dim W = \dim T(V)$, und dies impliziert, dass T surjektiv ist (Proposition 6.1.11).

„(iii) $\Rightarrow$ (ii)“ Sei nun T surjektiv. Dann gilt $\dim T(V) = \dim W = \dim V$, und somit folgt aus dem Rangsatz, dass $\dim \ker T = 0$ ist. Also ist $\ker T = \{0\}$, und T ist somit injektiv. □

Lassen Sie uns kurz anhand zweier Beispiele diskutieren, dass die Aussage des vorangehenden Korollars im Allgemeinen nicht mehr stimmt, wenn man die Annahme $\dim V = \dim W$ weglässt:

Beispiele 6.1.14. (a) Betrachten Sie die Matrix

$$A := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{4\times 3}.$$

Dann ist die lineare Abbildung

$$\begin{aligned} L_A\colon \qquad & \mathbb{R}^3 \to \mathbb{R}^4, \\ & x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto Ax = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 0 \end{bmatrix} \end{aligned}$$

injektiv (denn offensichtlich hat sie Kern $\{0\}$), aber nicht surjektiv (denn der Vektor $e_4 \in \mathbb{R}^4$ liegt nicht im Bild von L_A).

(b) Betrachten Sie die Matrix

$$B \coloneqq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 4}.$$

Dann ist die lineare Abbildung

$$\begin{aligned} L_B\colon \qquad & \mathbb{R}^4 \to \mathbb{R}^3, \\ & x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \mapsto Bx = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned}$$

surjektiv,[9] aber nicht injektiv (denn zum Beispiel liegt der Vektor $e_4 \in \mathbb{R}^4$ in ihrem Kern).

Bild und Rang einer Matrix

In Beispiel 6.1.7(b) hatten wir kurz angesprochen, dass man den Begriff Kern nicht nur für linearen Abbildungen, sondern auch für Matrizen definieren kann. Dasselbe gilt auch für die Begriffe Bild und Rang; dies besprechen wir im folgenden kurz:

Diskussion 6.1.15 (Bild und Rang einer Matrix)**.** Sei $\mathbb{K}$ ein Körper, seien $m, n \in \mathbb{N}^*$ und sei $A \in \mathbb{K}^{m\times n}$. Lassen Sie uns wieder die lineare Abbildung

$$\begin{aligned} L_A : \mathbb{K}^n &\to \mathbb{K}^m, \\ x &\mapsto Ax \end{aligned}$$

betrachten. Den Rang von L_A bezeichnet man häufig auch einfach als den **Rang der Matrix** A, abgekürzt als $\operatorname{rang}(A)$. Ebenso bezeichnet man das Bild von L_A häufig als das **Bild der Matrix** A. Hierzu sind ein paar Erläuterungen notwendig:

(a) Wenn wir mit $a_1, \ldots, a_n \in \mathbb{K}^m$ die Spalten von A bezeichnen, dann ist das Bild von A also die Menge

$$\begin{aligned} \{L_A(x) \mid\ x \in \mathbb{K}^n\} &= \{Ax \mid\ x \in \mathbb{K}^n\} \\ &= \{\sum_{k=1}^{n} x_j a_j \mid\ x_1, \ldots, x_n \in \mathbb{K}\} = \operatorname{span}(a_1, \ldots, a_n). \end{aligned}$$

Deshalb wird das Bild von A manchmal auch als **Spaltenraum von** A bezeichnet.

(b) Eine Basis des Bildes von A zu bestimmen bedeutet also nichts weiter als eine Basis von $\operatorname{span}(a_1, \ldots, a_n)$ zu bestimmen. Wie das geht, haben wir in Abschnitt 5.3 besprochen.

[9] Warum?

(c) Der Rang von A ist somit gleich der Dimension von $\operatorname{span}(a_1, \ldots, a_n)$ – und diese ist gleich der maximalen Länge aller linear unabhängigen Tupel, die man aus den Spalten von A auswählen kann.[10] Manchmal nennt man den Rang von A deshalb auch den **Spaltenrang von** A.

(d) Zur Berechnung des Rangs von A ist Theorem 5.3.1 sehr hilfreich: Wenn $Z \in \mathbb{K}^{m\times n}$ eine Zeilenstufenform von A ist, dann folgt aus dem Theorem, dass der Rang von A gleich der Anzahl der Spalten in Z mit Stufentiefe 1 ist. Dies ist aber offensichtlich gleich der Anzahl der von Null verschiedenen Zeilen von Z.

Also muss mann nur eine Zeilenstufenform Z von A berechnen und zählen, wieviele Zeilen von Z ungleich 0 sind, und hat damit den Rang von A berechnet.

(e) Indem wir auf die Abbildung L_A den Rangsatz anwenden, erhalten wir zudem die Formel

$$n = \dim \ker A + \operatorname{rang}(A)$$

für den Rang von A und die Dimension des Kerns von A.

Beispiel 6.1.16. Lassen Sie uns die Matrix

$$A := \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

betrachten, die bereits in Beispiel 5.3.3(a) vorkam. Wie in diesem Beispiel bereits angegeben, ist eine Zeilenstufenform von A gleich der Matrix

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

Diese hat zwei Zeilen, die nicht Null sind (oder anders formuliert: zwei Spalten mit Stufentiefe 1), und somit ist $\operatorname{rang}(A) = 2$.

Wegen $3 = \dim \ker A + \operatorname{rang}(A)$ folgt daraus, dass der Kern von A die Dimension 1 hat.

Man kann übrigens – analog zum Rang einer Matrix, der sich mit Hilfe der Spalten beschreiben lässt – auch den sogenannten **Zeilenrang** einer Matrix definieren. Aus Theorem 5.3.1 kann man leicht folgern, dass der Zeilenrang immer gleich dem Spaltenrang ist.

[10] Dies folgt aus den Resultaten in den Abschnitten 5.1 und 5.2.

6.2 Lösungsstruktur linearer Gleichungssysteme

Affine Unterräume

Sie wissen bereits, was ein Untervektorraum eines Vektorraums ist, und dass zum Beispiel im $\mathbb{R}^3$ die Untervektorräume der Dimension 1 und 2 genau die Ursprungsgeraden und die Ursprungsebenen sind. Im folgenden führen wir ein etwas allgemeineres Konzept ein. Wir benötigen hierfür die folgenden Notation: Wenn V ein Vektorraum über einem Körper $\mathbb{K}$ ist und $M \subseteq V$ sowie $a \in V$ ist, dann definiert man die Teilmenge

$$M + a \coloneqq \{v + a \mid \ v \in M\}$$

von V.

Definition 6.2.1 (Affiner Unterraum)**.** Sei V ein Vektorraum über einem Körper $\mathbb{K}$. Eine Teilmenge $L \subseteq V$ heißt ein **affiner Unterraum von** V, falls es ein $a \in V$ und einen Untervektorraum $U \subseteq V$ mit der Eigenschaft $L = U + a$ gibt.[11]

Anders ausgedrückt bedeutet das also, dass die affinen Unterräume eines Vektorraums genau diejenigen Teilmengen sind, die entstehen, wenn man einen Untervektorraum nimmt und um einen festen Vektor a verschiebt.

Lassen Sie uns nun kurz besprechen, wie affine Unterräume im $\mathbb{R}^2$ und im $\mathbb{R}^3$ geometrisch aussehen:

Beispiele 6.2.2. (a) Sei $L = U + a$ ein affiner Unterraum von $\mathbb{R}^2$, wobei a ein Punkt in $\mathbb{R}^2$ ist und $U \subseteq \mathbb{R}^2$ ein Untervektorraum. Lassen Sie uns diskutieren, wie L aussieht, je nachdem, welche Dimension U hat:

(0) $\dim U = 0$: In diesem Fall ist $U = \{0\}$ und somit ist $L = U + a = \{u + a \mid u \in \{0\}\} = \{a\}$ die Menge, die nur dem Punkt a besteht.

(1) $\dim U = 1$: In diesem Fall ist U eine Ursprungsgerade und $L = U + a$ somit eine Gerade, die durch den Punkt a verläuft.

(2) $\dim U = 2$: In diesem Fall ist $U = \mathbb{R}^2$ und somit $L = \mathbb{R}^2 + a = \mathbb{R}^2$, d.h., L ist gleich der kompletten Ebene $\mathbb{R}^2$.

(b) Sei $L = U + a$ ein affiner Unterraum von $\mathbb{R}^3$, wobe a ein Punkt in $\mathbb{R}^3$ ist und $U \subseteq \mathbb{R}^3$ ein Untervektorraum. Wir diskutieren erneut, wie L geometrisch aussieht, je nachdem, welche Dimension U besitzt:

(0) $\dim U = 0$: In dem Fall ist wieder $U = \{0\}$ und somit ist $L = U + a = \{u + a \mid \ u \in \{0\}\} = \{a\}$ die Menge, die einzig dem Punkt a besteht.

(1) $\dim U = 1$: In diesem Fall ist U eine Ursprungsgerade, also ist $L = U + a$ somit eine Gerade, die durch den Punkt a verläuft.

[11] Wenn dem so ist, dann gilt übrigens automatisch $a \in L$. Warum?

(2) $\dim U = 2$: In dem Fall ist U eine Ursprungsebene, und folglich ist $L = U + a$ eine Ebene, die durch den Punkt a verläuft.

(3) $\dim U = 3$: In diesem Fall ist $U = \mathbb{R}^3$, also ist $L = \mathbb{R}^3 + a = \mathbb{R}^3$, d.h., L ist der ganze Raum $\mathbb{R}^3$.

Das folgende Resultat zeigt, dass in der Darstellung $L = U + a$ eines affinen Untervektorraumes der Punkt a als ein beliebiger Punkt in A gewählt werden kann, während U ein eindeutiger bestimmter Untervektorraum ist.

Proposition 6.2.3. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$ und sei $L = U+a \subseteq V$ ein affiner Unterraum, wobei $a \in V$ ist und $U \subseteq V$ ein Untervektorraum ist.*

(a) *Für jedes $\tilde{a} \in L$ gilt $L = U + \tilde{a}$.*

(b) *Wenn für ein $\tilde{a} \in V$ und einen Untervektorraum $\tilde{U} \subseteq V$ die Gleichheit $L = \tilde{U} + \tilde{a}$ gilt, dann ist $\tilde{U} = U$.*

Beweis. (a) Sei $\tilde{a} \in L$ beliebig, aber fest. Dann gibt es ein $u_0 \in U$ mit $\tilde{a} = u_0 + a$. Wir müssen $U + \tilde{a} = U + a$ zeigen.

„$\subseteq$“ Sei $x \in U + \tilde{a}$. Dann gibt es ein $u \in U$ mit $x = u + \tilde{a}$. Folglich ist

$$x = u + \tilde{a} = u + (u_0 + a) = (u + u_0) + a \in U + a,$$

denn weil U ein Untervektorraum ist, gilt $u + u_0 \in U$.

„$\supseteq$“ Sei $x \in U + a$. Dann gibt es ein $u \in U$ mit $x = u + a$. Somit ist

$$x = u + a = u + (\tilde{a} - u_0) = (u - u_0) + \tilde{a} \in U + \tilde{a},$$

denn weil U ein Untervektorraum ist, gilt $u - u_0 \in U$.

(b) Sei $\tilde{a} \in V$ und sei $\tilde{U} \subseteq V$ ein Untervektorraum mit der Eigenschaft $L = \tilde{U}+\tilde{a}$. Dann ist $U + a = \tilde{U} + \tilde{a}$. Wir müssen $U = \tilde{U}$ zeigen, und aus Symmetriegründen genügt es, die Inklusion „$\subseteq$“ zu beweisen.

Sei also $u \in U$ beliebig, aber fest. Dann gilt $u + a \in U + a = \tilde{U} + \tilde{a}$, d.h., es gibt es $\tilde{u} \in \tilde{U}$ mit $u + a = \tilde{u} + \tilde{a}$. Somit ist also

$$u = \tilde{u} + (\tilde{a} - a),$$

d.h., wenn wir $\tilde{a} - a \in \tilde{U}$ zeigen können, dann folgt bereits, wie gewünscht, $u \in \tilde{U}$.

Um $\tilde{a} - a \in \tilde{U}$ zu zeigen, beachten wir zunächst, dass $a = 0 + a \in U + a = \tilde{U} + \tilde{a}$ gilt. Also gibt es ein $\tilde{w} \in \tilde{U}$ mit der Eigenschaft $a = \tilde{a} + \tilde{w}$, und somit ist $\tilde{a} - a = -\tilde{w} \in \tilde{U}$. □

Die vorangehende Proposition zeigt, dass der Untervektorraum U, der zu einem affinen Unterraum L gehört, eindeutig bestimmt ist. Aus diesem Grund ergibt es auch Sinn, die **Dimension** des affinnen Unterraums L zu definieren: Man setzt einfach $\dim L := \dim U$.

Demnach haben zum Beispiel in $\mathbb{R}^3$ Punkte die Dimension 0, Geraden die Dimension 1, Ebenen die Dimension 2 und der Raum $\mathbb{R}^3$ selbst die Dimension 3.

Der Lösungsraum linearer Gleichungen

Affine Unterräume von Vektorräumen hängen eng mit der Lösungsmenge von linearen Gleichungen zusammen:

Theorem 6.2.4. *Seien V, W Vektorräume über einem Körper $\mathbb{K}$, sei $T : V \to W$ linear und sei $b \in W$. Wir setzen*

$$L \coloneqq \{x \in V \mid T(x) = b\} \qquad \text{und} \qquad K \coloneqq \{x \in V \mid T(x) = 0\} = \ker T.$$

Dann gilt:

(a) *Es ist K ein Untervektorraum von V.*

(b) *Die Menge L ist entweder leer oder ein affiner Unterraum von V. In letzterem Fall gilt $L = K + x_0$ für jedes $x_0 \in L$.*

Beweis. (a) Aus Proposition 6.1.6(a) wissen wir bereits, dass der Kern einer linearen Abbildung ein Untervektorraum ist.

(b) Sei L nicht leer und $x_0 \in L$. Wir zeigen, dass $L = K + x_0$ gilt.[12]

„$\subseteq$“ Sei $x \in L$. Dann gilt $T(x) = b = T(x_0)$ und somit ist $x - x_0 \in \ker T = K$. Also folgt $x = (x - x_0) + x_0 \in K + x_0$.

„$\supseteq$“ Sei $x \in K + x_0$. Dann gibt es ein $y \in K$ mit $x = y + x_0$, und hieraus folgt

$$T(x) = T(y) + T(x_0) = 0 + b = b,$$

also gilt $x \in L$. □

Das vorangehende Theorem kann man folgendermaßen interpretieren: Die Menge K ist die Menge aller Lösungen x der sogenannten **homogenen Gleichung** $T(x) = 0$, und die Menge L ist die Menge der Lösungen x der sogenannten **inhomogenen Gleichung** $T(x) = b$. Teil (b) des Theorems sagt somit, dass man alle Lösungen der inhomogenen Gleichung finden kann, indem man eine Lösung der inhomogenen Gleichung sowie alle Lösungen der homogenen Gleichung findet und letztere zur ersteren addiert.

Beispiele 6.2.5. Es bezeichen $\mathcal{P} \subseteq \mathrm{Abb}(\mathbb{R}; \mathbb{R})$ den Untervektorraum, der aus allen Polynomfunktionen besteht. Wir betrachten die lineare Abbildung

$$\begin{aligned} T : \mathcal{P} &\to \mathrm{Abb}(\mathbb{R}; \mathbb{R}), \\ f &\mapsto f', \end{aligned}$$

die jede Polynomfunktion auf ihre Ableitung abbildet. Außerdem betrachten wir die Sinusfunktion $\sin \in \mathrm{Abb}(\mathbb{R}; \mathbb{R})$.

[12] Weil K laut (a) ein Untervektorraum ist, ist dann auch gezeigt, dass L ein affiner Unterraum ist.

(a) Dann ist die Menge L der Lösungen f der Gleichung $T(f) = \sin$, d.h. die Menge

$$L = \{f \in \mathcal{P} \mid \ T(f) = \sin\},$$

leer – denn die Ableitung einer Polynomfunktion ist stets eine Polynomfunktion und kann somit niemals gleich der Sinusfunktion sein.[13]

(b) Die Menge $\tilde{L}$ der Lösungen f der Gleichung $T(f) = m_2$, wobei $m_2 \in \mathrm{Abb}(\mathbb{R};\mathbb{R})$ durch $m_2(x) = x^2$ durch alle $x \in \mathbb{R}$ gegeben sei, ist hingegen ein ein-dimensionaler affiner Unterraum von $\mathcal{P}$; das kann man folgendermaßen sehen: Die Menge

$$K = \{f \in \mathcal{P} \mid \ T(f) = 0\}$$

der Lösungen der homogenen Gleichung besteht aus den Vielfachen der konstanten $1-$Funktion $\mathbb{1} \in \mathrm{Abb}(\mathbb{R};\mathbb{R})$, d.h. es ist $K = \mathrm{span}\{\mathbb{1}\}$.

Wenn außerdem $m_3 \in \mathcal{P}$ die Monomfunktion bezeichnet, die durch $m_3(x) = x^3$ für alle $x \in \mathbb{R}$ gegeben ist, dann dann ist $\frac{1}{3}m_3$ eine Lösung der inhomogenen Gleichung $T(f) = m_2$. Also gilt wegen Theorem 6.2.4(b)

$$L = \mathrm{span}\{\mathbb{1}\} + \frac{1}{3}m_3 = \{\alpha\,\mathbb{1} + \frac{1}{3}m_3 \mid \ \alpha \in \mathbb{R}\}.$$

Anders ausgedrückt besteht L genau es den Funktionen der Form $x \mapsto \alpha + \frac{1}{3}x^3$ für Koeffizienten $\alpha \in \mathbb{R}$.

Berechnung des Lösungsraums von Gleichungssystemen

Nun wollen wir nochmal konkret zu linearen Gleichungssystem zurückkommen. Denken Sie daran, dass man diese in der Form $Ax = b$ für einen Spaltenvektor $b \in \mathbb{K}^m$, eine Matrix $A \in \mathbb{K}^{m\times n}$ und einen gesuchten Vektor $x \in \mathbb{K}^n$ schreiben kann.

Wir geben nun zunächst an, wie man den Kern einer Matrix A – d.h., die Menge $\ker A = \{x \in \mathbb{K}^n \mid \ Ax = 0\}$ mit Hilfe des Gaußalgorithmus bestimmen kann. Anschließend wenden wir uns der inhomogenen Gleichung zu.

Theorem 6.2.6. *Sei $\mathbb{K}$ ein Körper, seien $m, n \in \mathbb{N}^*$ und sei $A \in \mathbb{K}^{m\times n}$. Es bezeichne $Z \in \mathbb{K}^{m\times n}$ die reduzierte Zeilenstufenform von A und seien $1 \le j_1 < \cdots < j_p \le n$ die Indizes derjenigen Spalten von Z, die Stufentiefe 0 haben. Dann gilt $\ker A = \ker Z$, dieser Raum besitzt Dimension p, und man hat die folgenden beiden Möglichkeiten um alle Vektoren in $\ker A = \ker Z$ zu bestimmen:*

[13] Beachten Sie aber unbedingt, dass L nur deshalb leer ist, weil T nur auf der Menge der Polynomfunktionen definiert haben. Wenn wir stattdessen einen größeren Definitionsbereich für T wählen – sagen wir zum Beispiel, den Untervektorraum von $\mathrm{Abb}(\mathbb{R};\mathbb{R})$, der aus allen differenzierbaren Funktionen besteht – dann gibt durchaus Lösungen der Gleichung $T(f) = \sin$, beispielsweise die Funktion $-\cos$.

(a) *Man kann eine Basis von* $\ker A = \ker Z$ *mit folgendem Verfahren bestimmen: Für jedes* $h \in \{1, \dots, p\}$ *betrachte man denjenigen Vektor* $x^{(h)} \in \mathbb{K}^n$*, der an der Stelle* j_h *den Eintrag* 1 *besitzt und an den Indizes aus* $\{j_1, \dots, j_p\} \setminus \{j_h\}$ *den Eintrag* 0 *besitzt, und dessen verbleibende Einträge so bestimmt werden, dass die Gleichung* $Zx^{(h)} = 0$ *erfüllt ist. Dann bildet* $(x^{(1)}, \dots, x^{(p)})$ *eine Basis von* $\ker A = \ker Z$*.*

(b) *Man kann alle Vektoren* $x \in \ker A = \ker Z$ *bestimmen, in dem man folgendermaßen vorgeht: Man lässt für die Einträge von* x *mit den Indizes* $j_1, \dots, j_p$ *beliebige Werte aus* $\mathbb{K}$ *zu und bestimmt die verbleibenden Einträge von* x *dann so, dass die Gleichung* $Zx = 0$ *erfüllt ist.*[14]

Man kann das Theorem beweisen, indem man die Rangformel sowie die Struktur der reduzierten Zeilenstufenform Z benutzt. Es ist aber vermutlich klarer, wenn wir das Theorem einfach in einem Beispiel veranschaulichen:

Beispiel 6.2.7. Lassen Sie uns noch einmal die Matrix

$$A \coloneqq \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

betrachten, die wir uns unter anderem schon in Beispiel 6.1.16 angesehen haben. Mit Hilfe des Gaußalgorithmus kann man ihre reduzierte Zeilenstufenform Z bestimmen; sie lautet

$$Z = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Diese Matrix besitzt nur in der letzten Spalte die Stufentiefe 0. Also hat $\ker A = \ker Z$ die Dimension 1, und besitzt somit eine Basis, die aus nur einem Vektor besteht $x^{(1)}$. Laut Theorem 6.2.6(a) können wir solch einen Vektor in $x^{(1)} \in \mathbb{R}^3$ bestimmen, indem wir den dritten Eintrag auf 1 setzen und die anderen Einträge aus der Gleichung $Zx^{(1)} = 0$ bestimmen. Daraus ergibt sich

$$x_2^{(1)} = -x_3^{(1)} = -1 \qquad \text{und} \qquad x_1^{(1)} = -x_3^{(1)} = -1,$$

also

$$x^{(1)} = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}.$$

[14] Das mag auf den ersten Blick trivial wirken; aber der Witz dabei ist erstens, dass das Verfahren immer funktioniert – d.h. man kann die verbleibenden Einträge von x immer so bestimmen, dass $Zx = 0$ gilt, und zwar auf eindeutige Weise – und zweitens, dass man auf diese Weise immer alle Lösungen erhält.

Alternativ kann man auch das Verfahren aus Theorem (b) verwenden um zumselben Ergebnis zu gelangen: Wir bestimmen die Lösungen $x \in \mathbb{R}^3$ von $Zx = 0$, indem wir den Eintrag x_3 beliebig lassen, und die beiden Einträge x_1 und x_2 dann so bestimmen, dass $Zx = 0$ gilt; d.h. wir müssen

$$x_2 = -x_3 \qquad \text{und} \qquad x_1 = -x_3$$

wählen, und erhalten somit den Lösungsraum

$$\begin{aligned}\ker A = \ker Z &= \{x \in \mathbb{R}^3 \mid Zx = 0\} \\ &= \left\{ \begin{bmatrix} -x_3 \\ -x_3 \\ x_3 \end{bmatrix} \mid x_3 \in \mathbb{R} \right\} = \left\{ x_3 \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \mid x_3 \in \mathbb{R} \right\} = \operatorname{span} \left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\}.\end{aligned}$$

Nun wenden wir uns, wie angekündigt, der Berechnung des Lösungsraums von inhomogenen linearen Gleichungssystemen zu:

Theorem 6.2.8. *Sei $\mathbb{K}$ ein Körper, seien $m, n \in \mathbb{N}^*$ und sei $A \in \mathbb{K}^{m\times n}$ sowie $b \in \mathbb{K}^m$. Es bezeichne $[Z\ c] \in \mathbb{K}^{m\times(n+1)}$ die reduzierte Zeilenstufenform von $[A\ b] \in \mathbb{K}^{m\times(n+1)}$.*

Dann sind die Lösungen $x \in \mathbb{K}^n$ der Gleichung $Ax = b$ genau die Lösungen der Gleichung $Zx = c$, und es tritt genau einer der folgenden drei Alternativen ein:

(A) *Die letzte Spalte von $[Z\ c]$ hat Stufentiefe 1.*

In diesem Fall gibt es kein $x \in \mathbb{K}^n$, das die Gleichung $Ax = b$ erfüllt.

(B) *Die letzte Spalte von $[Z\ c]$ hat Stufentiefe 0 und alle anderen Spalten haben Stufentiefe 1.*

In diesem Fall gibt es genau ein $x \in \mathbb{K}^n$, das die Gleichung $Ax = b$ erfüllt. Außerdem ist $n \leq m$ und die Einträge der Lösung $x \in \mathbb{K}^n$ sind genau die ersten n Einträge von c.

(C) *Die letzte Spalte von $[Z\ c]$ hat Stufentiefe 0 und manche der vorangehenden Spalten haben Stufentiefe 0.*

In diesem Fall hat die Gleichung $Ax = b$ eine Lösung, und für jede beliebige Lösung $\hat{x} \in \mathbb{K}^n$ ist der Lösungsraum $\{x \in \mathbb{K}^n \mid Ax = b\}$ gleich dem affinen Unterraum $\hat{x} + \ker A$, wobei $\ker A$ sich wie in Theorem 6.2.6 angegeben aus Z bestimmen lässt.

Außerdem lassen die Lösungen x von $Ax = b$ analog zu Theorem 6.2.6(b) *bestimmen.*[15]

[15] Wobei man natürlich die Gleichung $Zx = 0$ durch $Zx = c$ ersetzt.

Beweis. Laut Theorem 4.3.4(c) gibt es eine invertierbare Matrix $T \in \mathbb{K}^{m\times m}$ mit der Eigenschaft $T[A\ b] = [Z\ c]$, und somit ist

$$TA = Z \qquad \text{und} \qquad Tb = c.$$

Für $x \in \mathbb{K}^n$ ist die Gleichung $Ax = b$ äquivalent zur Gleichung $TAx = Tb$, also zur Gleichung $Zx = c$. Somit müssen wir für den Rest des Beweises nur noch die Gleichung $Zx = c$ studieren.

Offensichtlich tritt bezüglich der Stufentiefen in der Matrix $[Z\ c]$ genau einer der drei genannten Fälle ein. Lassen Sie uns nun zeigen, dass in jedem dieser drei Fälle die behaupteten Eigenschaften gelten.

(A) Bezeichne k die Treppentiefe von $[Z\ c]$ in der letzten Spalte. Wegen der Voraussetzung an die Stufentiefe gilt $c_k \neq 0$, aber die k-te Zeile von Z ist gleich 0 und somit ist für jedes $x \in \mathbb{K}^n$ der k-te Eintrag von Zx gleich 0. Also gibt es kein $x \in \mathbb{K}^n$ mit $Zx = c$.

(B) Aus den angegebenen Bedingungen an die Stufentiefen folgt sofort, dass $n \leq m$ ist und dass $Z \in \mathbb{K}^{m\times n}$ von der Form

$$Z = \begin{bmatrix} I_n \\ 0 \end{bmatrix}$$

ist. Hieraus ergeben sich die behaupteten Eigenschaften.

(C) Dass die Lösungen x von $Zx = c$ sich durch sukzessives Auflösen nach $x_n, \ldots, x_1$ bestimmen lassen, ist klar, da Z sich in reduzierter Zeilenstufenform befindet. Somit sieht man auch, dass die Gleichung eine Lösung besitzt. Die behauptete Darstellung des Lösungsraums in der Form $\ker A + \hat{x}$ folgt aus Theorem 6.2.4(b). □

Auch hier wollen wir zwei Beispiele diskutieren:

Beispiele 6.2.9. Wie in Beispiel 6.2.7 sei

$$A := \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

(a) Wir wollen die Lösungen $x \in \mathbb{R}^3$ der Gleichung $Ax = b$ finden für

$$b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Dazu bringen wir mit Hilfe des Gaußalgorithmus die Matrix $[A\ b]$ auf reduzierte Zeilenstufenform $[Z\ c]$ und erhalten so

$$[Z\ c] = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Weil die letzte Spalte Stufentiefe 1 besitzt, besitzt die Gleichung $Ax = b$ laut Theorem 6.2.8(A) keine Lösung $x \in \mathbb{R}^3$.

(b) Lassen Sie uns nun stattdessen den Vektor

$$b = \begin{bmatrix} 3 \\ 0 \\ 5 \end{bmatrix}$$

betrachten und wieder die Lösungen $x \in \mathbb{R}^3$ der Gleichung $Ax = b$ bestimmen: Die reduzierte Zeilenstufenform $[Z\ c]$ von $[A\ b]$ lautet

$$[Z\ c] = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Weil Z in der dritten Spalte die Stufentiefe 0 besitzt, können wir, um eine Lösung $\hat{x} \in \mathbb{R}^3$ zu bestimmen, den dritten Eintrag von $\hat{x}$ frei wählen. Wir setzen ihn zum Beispiel gleich 1 und erhalten dann durch sukzessives Auflösen der Gleichung $Z\hat{x} = c$ nach den anderen Variablen, dass

$$\hat{x}_2 = 2 - \hat{x}_3 = 1 \qquad \text{und} \qquad \hat{x}_1 = 1 - \hat{x}_3 = 0$$

und somit

$$\hat{x} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

gilt. Insgesamt ist somit laut Theorem 6.2.8(C) die Menge der Lösungen x von $Ax = b$ gegeben durch

$$\ker A + \hat{x} = \operatorname{span}\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \right\} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = \left\{ \alpha \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \mid \alpha \in \mathbb{R} \right\}.$$

6.3 Implizite Darstellung von affinen Unterräumen des $\mathbb{K}^n$

Darstellung von Untervektorräumen als Lösungsraum von Gleichungen

In der folgenden Proposition zeigen wir, dass sich jeder affine Unterraum des $\mathbb{K}^n$ als Lösungsmenge eines linearen Gleichungssystems schreiben lässt.

Proposition 6.3.1. *Sei $\mathbb{K}$ ein Körper, sei $n \in \mathbb{N}^*$ und sei $L \subseteq \mathbb{K}^n$ ein affiner Unterraum der Dimension $d \leq n-1$. Dann gibt es eine Matrix $A \in \mathbb{K}^{(n-d)\times n}$ sowie ein $b \in \mathbb{K}^{n-d}$ mit der Eigenschaft*

$$L = \{x \in \mathbb{K}^n \mid Ax = b\}.$$

Sie sollten sich kurz überlegen, dass es auch im Fall $d = n$ eine Matrix A und einen Vektor b gibt derart, dass $L = \{x \in \mathbb{K}^n \mid Ax = b\}$ gilt. Weshalb haben wir diesen Fall dann in der Proposition nicht zugelassen?

Proposition 6.3.1 kann man mit Hilfe des Basisergänzungssatzes beweisen. Der Beweis ist nicht überdurchschnittlich kompliziert, aber aus Zeitgründen lagern wir ihn in die Ergänzungen am Ende dieses Kapitels aus.

Explizite vs. implizite Darstellung

Mit Hilfe der vorangehenden Proposition können wir nun besprechen, wie man zwischen **impliziten** und **expliziten** Darstellungen von affinen Unterräumen des $\mathbb{K}^n$ wechseln kann:

Diskussion 6.3.2. Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$, und sei L ein affiner Unterraum von $\mathbb{K}^n$, d.h. wir können L schreiben als $L = U + a$ für einen Punkt $a \in \mathbb{K}^n$ und einen Untervektorraum $U \subseteq \mathbb{K}^n$.

(a) Unter einer **expliziten Darstellung von** L verstehen wir die Angabe eines Vektors $a \in \mathbb{K}^n$ mit $L = U + a$ und zusammen mit der Angabe einer Basis von U.

Der Begriff "explizit" ist hier sinnvoll, weil man mit diesen Informationen sofort sämtliche Vektoren aus L bestimmen kann: Man nimmt einfach alle Linearkombinationen der Basis und addiert jeweils a.

(b) Unter einer **impliziten Darstellung von** L verstehen wir die Angabe einer Zahl $m \in \mathbb{N}^*$ und einer Matrix $A \in \mathbb{K}^{m\times n}$ sowie eines Vektors $b \in \mathbb{K}^m$ derart, dass

$$L = \{x \in \mathbb{K}^n \mid Ax = b\}$$

gilt. Laut Proposition 6.3.1 besitzt jeder affine Unterraum eine implizite Darstellung.[16]

Der Begriff "implizit" ist hier sinnvoll, weil die Angabe von m, A und b einerseits den affinen Unterraum L eindeutig bestimmt, aber die Information, welche Vektoren in L liegen, mit Hilfe der Daten m, A und b nicht unmittelbar zugänglich ist – um diese Information zu erhalten, muss man die Gleichung $Ax = b$ lösen, was einen erheblichen Rechenaufwand bedeuten kann.

(c) Mit Hilfe der soeben eingeführten Terminologie kann man also sagen:

Das Lösen eines linearen Gleichungssystems bedeutet, von einer impliziten Darstellung eines affinen Unterraums zu einer expliziten Darstellung zu wechseln.

[16] Beachten Sie aber unbedingt, dass die Daten m, A und b nicht eindeutig bestimmt sind! Insbesondere ist man, wenn man nur m, A und b kennt, nicht automatisch in der Situation von Proposition 6.3.1, in der sich aus der Größe von A bereits die Dimension von L ablesen lässt!

Es ist sehr wichtig, sich klar zu machen, was das soeben Besprochene für Geraden und Ebenen im $\mathbb{R}^2$ und $\mathbb{R}^3$ bedeutet:

Beispiele 6.3.3. (a) Sei $L \subseteq \mathbb{R}^2$ eine Gerade, d.h. ein affiner Unterraum der Dimension 1. Dann kann man L schreiben als $L = U + a$ für einen eindimensionalen Untervektorraum $U \subseteq \mathbb{R}^2$ und einen Punkt $a \in \mathbb{R}^2$.

Eine explizite Darstellung von L anzugeben, bedeutet einen solchen Punkt a sowie eine Basis (v) von U (welche automatisch aus genau einem Vektor $v \in \mathbb{R}^2$ besteht) anzugeben. Wenn man b und a kennt, erhält man also

$$L = \{\alpha v + a \mid \alpha \in \mathbb{R}\}.$$

In der Schule haben Sie a möglicherweise unter dem Begriff **Aufpunkt** oder **Stützpunkt** der Gerade L kennen gelernt.

Man kann aber auch eine implizite Darstellung von L angeben: Laut Proposition 6.3.1 gibt es ein $A \in \mathbb{R}^{1\times 2}$ und ein $b \in \mathbb{R}^1 = \mathbb{R}$ mit der Eigenschaft

$$L = \{x \in \mathbb{R}^2 \mid Ax = b\} = \{x \in \mathbb{R}^2 \mid A_{11}x_1 + A_{12}x_2 = b\}.$$

Falls $A_{12} \neq 0$ ist, ist die Gleichung $A_{11}x_1 + A_{12}x_2 = b$ äquivalent zu

$$x_2 = \frac{b}{A_{12}} - \frac{A_{11}}{A_{12}}x_1,$$

und diese Form um Geraden um $\mathbb{R}^2$ zu beschreiben, kennen Sie vermutlich aus der Schule (wobei Sie vermutlich andere Variablennamen benutzt haben).[17]

(b) Sei nun L eine Ebene im $\mathbb{R}^3$. Dann ist L von der Form $L = U + a$ für ein $a \in \mathbb{R}^3$ und einen zwei-dimensionalen Untervektorraum $U \subseteq \mathbb{R}^2$.

Eine explizite Darstellung von L anzugeben bedeutet, solch einen Vektor a sowie eine Basis (v_1, v_2) von U anzugeben. Wenn man diese Vektoren zur Verfügung hat, dann gilt einfach

$$L = \{\alpha_1 v_1 + \alpha_2 v_2 + a \mid \alpha_1, \alpha_2 \in \mathbb{R}\}.$$

Auch diese Darstellung von Ebenen kennen Sie vielleicht schon aus der Schule, wobei a wiederum ein **Aufpunkt** oder **Stützpunkt** der Ebene ist.

Auch für eine solche Ebene L kann man laut Proposition 6.3.1 eine implizite Darstellung angeben: Es gibt ein $A \in \mathbb{R}^{1\times 3}$ und ein $b \in \mathbb{R}^1 = \mathbb{R}$ derart, dass

$$L = \{x \in \mathbb{R}^3 \mid Ax = b\}$$

gilt. Den Spaltenvektor A^{T} bezeichnet man häufig auch als den **Normalenvektor der Ebene** L. Dies hängt eng mit dem Konzept der **Orthogonalität** zusammen, dass in der Linearen Algebra 2 näher besprochen wird.

(c) Jetzt sind Sie dran: Was kann man über Geraden L in $\mathbb{R}^3$ sagen?

[17] Und was ist im Fall $A_{12} = 0$ los?

6.4 Ergänzungen

Ein Beweis von Proposition 6.3.1

Den folgenden Beweis hatten wir in der Vorlesung aus Zeitgründen ausgespart.

Beweis von Proposition 6.3.1. Sei $L = U + a$ für einen Untervektorraum $U \subseteq \mathbb{K}^n$ und ein $a \in \mathbb{K}^n$, und $(x_1, \dots, x_d)$ eine Basis von U. Laut Basisergänzungssatz können wir diese zu einer Basis $(x_1, \dots, x_d, x_{d+1}, \dots, x_n)$ von $\mathbb{K}^n$ ergänzen.

Laut Proposition 5.7.1 gibt es eine lineare Abbildung $T : \mathbb{K}^n \to \mathbb{K}^{n-d}$, die $x_1, \dots, x_d$ auf 0 abbildet, und $x_{d+1}, \dots, x_n$ auf die kanonischen Einheitsvektoren in $\mathbb{K}^{n-d}$. Somit ist $U = \ker T$, denn: die Vektoren $x_1, \dots, x_d$ liegen offensichtlich im Kern von T; außerdem ist T surjektiv, also folgt aus dem Rangsatz

$$\dim \ker T = \dim \mathbb{K}^n - \dim T(\mathbb{K}^n) = n - (n - d) = d;$$

dies zeigt, wie behauptet, $U = \ker T$.

Nun definieren wir einfach $b \coloneqq T(a)$, und wählen $A \in \mathbb{K}^{m \times n}$ als diejenige Matrix, welche $Ax = T(x)$ für alle $x \in \mathbb{K}^n$ erfüllt. Dann erhalten wir wie gewünscht

$$\{x \in \mathbb{K}^n \mid \ Ax = b\} = \{x \in \mathbb{K}^n \mid \ T(x) = b\} = \ker T + a = U + a = L,$$

wobei die zweite Gleichheit aus Theorem 6.2.4(b) folgt. □

Kapitel 7

Determinanten

Einstiegsfragen. (a) Zeichnen Sie im $\mathbb{R}^2$ das Viereck mit den folgenden Eckpunkten:

$$v_1 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 5 \end{bmatrix}, \quad v_3 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \quad v_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Können Sie die Fläche des Vierecks berechnen? Und den Umfang?

(b) Haben Sie eine Idee, wie man für einen beliebigen (nicht regelmäßigen) Tetraeder im $\mathbb{R}^3$ die Oberfläche berechnen kann?

(c) Wie nochmal können Sie herausfinden, ob eine gegebene Matrix $A \in \mathbb{C}^{3\times 3}$ invertierbar ist?

7.1 Axiome der Determinante

Zur Motivation: Flächen im $\mathbb{R}^2$

Lassen Sie uns darüber sprechen, wie man Flächen von Parallelogrammen im $\mathbb{R}^2$ berechnen kann. Dabei wollen wir aber nicht einfach eine Formel angeben, sondern axiomatisch vorgehen:

Diskussion 7.1.1 (Ein axiomatischer Zugang zu Parallelogrammflächen im $\mathbb{R}^2$)**.** Lassen Sie uns, innerhalb dieser Diskussion, für alle $v_1, v_2 \in \mathbb{R}^2$ die Fläche des Paralellogramms, das von v_1 und v_2 aufgespannt wird, mit $F(v_1, v_2)$ bezeichnen. Es ist also F eine Funktion von $\mathbb{R}^2 \times \mathbb{R}^2$ nach $\mathbb{R}$, und aus geometrischer Sicht sollte diese Funktion zumindest die folgenden Eigenschaft erfüllen:

(a) Die Fläche des **Einheitsquadrats** ist 1, d.h. $F(e_1, e_2) = 1$.

(b) Es gilt **positive Homogenität**, d.h. für alle $v_1, v_2 \in \mathbb{R}^2$ und alle $\alpha \in [0, \infty)$ ist

$$F(\alpha v_1, v_2) = \alpha F(v_1, v_2) \qquad \text{und} \qquad F(v_1, \alpha v_2) = \alpha F(v_1, v_2).$$

(c) Es gilt **Additivität** in beiden Argumenten, d.h. für alle $v_1, v_2, v_3 \in \mathbb{R}^2$ gilt

$$\begin{aligned} & F(v_1 + v_2, v_3) = F(v_1, v_3) + F(v_2, v_3) \\ \text{und} \qquad & F(v_1, v_2 + v_3) = F(v_1, v_2) + F(v_1, v_3). \end{aligned}$$

(d) Es gilt **Scherungsinvarianz**, d.h. für alle $v_1, v_2 \in \mathbb{R}^2$ gilt

$$F(v_1, v_2) = F(v_1, v_1 + v_2) \qquad \text{und} \qquad F(v_1, v_2) = F(v_1 + v_2, v_2).$$

Aus der Additivität und der positiven Homogenität kann man folgende bemerkenswerte Folgerung ziehen: Für alle $v_1, v_2 \in \mathbb{R}$ gilt:

$$F(v_1, v_2) + F(-v_1, v_2) = F(v_1 - v_1, v_2) = F(0, v_2) = 0 \cdot F(v_1, v_2) = 0,$$

und somit

$$F(-v_1, v_2) = -F(v_1, v_2).$$

Das heißt also: Wenn $F(v_1, v_2)$ nicht Null ist,[1] dann ist also eine der beiden Zahlen $F(v_1, v_2)$ und $F(-v_1, v_2)$ negativ. Das ist auf den ersten Blick merwürdig, denn F soll ja Flächen messen.

Der Grund für diese Verhalten liegt anschaulich darin, dass die Addition eines weiteren Vektors zu v_1 eben auch dazu führen kann, dass das von v_1 und v_2 aufgespannte Parallelogramm kleiner wird – wenn man also die Eigenschaft der Additivität in dieser Allgemeinheit behalten will, muss man auch negative Flächenwerte zulassen.[2]

Wenn man sich hiermit abgefunden hat, dann erkennt man, dass die Homogenität auch für negative Skalare stimmt: Für $\alpha \in [0, \infty)$ und $v_1, v_2 \in \mathbb{R}^2$ folgt nämlich aus dem soeben gezeigten Vorzeichenverhalten

$$F(-\alpha v_1, v_2) = -F(\alpha v_1, v_2) = -\alpha F(v_1, v_2).$$

Mit denselben Schritten wie oben kann man die entsprechende Formel auch im zweiten Argument zeigen.

Deswegen kann man bei der Homogenität auch genauso gut fordern, dass

$$F(\alpha v_1, v_2) = \alpha F(v_1, v_2) \qquad \text{und} \qquad F(v_1, \alpha v_2) = \alpha F(v_1, v_2)$$

sogar für alle $\alpha \in \mathbb{R}$ gilt.

[1] Wann hat das Parallelogramm, das von v_1 und v_2 aufgespannt wird, übrigens die Fläche 0?

[2] Dies klingt auf den ersten Blick sehr ungewohnt, erweist sich aber letztlich als sehr, sehr praktisch: Es stellt sich heraus, dass $F(v_1, v_2)$ somit nicht nur die Fläche des Parallelogramms zwischen v_1 und v_2 misst, sondern auch noch die Orientierung – also die Frage, ob v_2 mit v_1 einen positiven oder einen negativen Winkel aufspannt. Das besprechen wir an dieser Stelle aber vorerst nicht weiter.

Definition der Determinante

Die Erkenntnisse aus obiger Diskussion werden wir im folgenden verwenden um axiomatisch eine Funktion zu beschreiben, die in der Lage ist, im $\mathbb{R}^2$ Flächen von Parallelogrammen, im $\mathbb{R}^3$ Volumen vom Parallelepipeden, usw. (jeweils mit einem Vorzeichen) zu bestimmen. Wir führen solch eine Funktion aber nicht nur über $\mathbb{R}$ ein, sondern gleich über beliebigen Körpern (aber Achtung: auf anderen Körpern als $\mathbb{R}$ ergibt die oben genannten geometrische Interpretation keinen Sinn!).

Um das Sprechen (und Schreiben) über alle nötigen Konzepte zu erleichtern, führen wir zunächst noch eine allgemeine Terminologie für die oben verwendeten Begriffe ein:

Definition 7.1.2. Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}^*$ und sei $F : V^n \to \mathbb{K}$.

(a) Die Funktion F heißt **scherungsinvariant**, falls für alle $v_1, \dots, v_n \in V$ und alle $j, k \in \{1, \dots, n\}$ mit $j \neq k$ gilt:

$$F(v_1, \dots, v_n) = F(v_1, \dots, v_{k-1}, v_k + v_j, v_{k+1}, \dots, v_n).$$

(b) Die Funktion F heißt n-**linear**, falls sie in jedem Argument linear ist, d.h., falls für alle $k \in \{1, \dots, n\}$ und alle $v_1, \dots, v_{k-1}, v_{k+1}, \dots, v_n \in V$ ist die Abbildung

$$\begin{aligned} V &\to \mathbb{K}, \\ v_k &\mapsto F(v_1, \dots, v_{k-1}, v_k, v_{k+1}, \dots, v_n) \end{aligned}$$

linear ist.

(c) Die Funktion F heißt **anti-symmetrisch**, falls für alle $v_1, \dots, v_n \in V$ und alle $j, k \in \{1, \dots, n\}$ mit $j < k$ die Gleichheit

$$F(v_1, \dots, v_n) = -F(v_1, \dots, v_{j-1}, v_k, v_{j+1}, \dots, v_{k-1}, v_j, v_{k+1}, \dots, v_n)$$

gilt (d.h. falls das Vertauschen von zwei Vektoren in den Argumenten denselben Effekt hat wie den Funktionswert mit -1 zu multiplizieren).

Hier folgt nun die angekündigte Definition:

Definition 7.1.3 (Determinante)**.** Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}^*$. Ein Abbildung $\det : \mathbb{K}^{n \times n} \to \mathbb{K}$ heißt *Determinante*, falls $\det(I_n) = 1$ gilt, und falls die Abbildung in den Spalten ihres Arguments n-linear und scherungsinvariant ist (d.h. also falls zusätzlich zur Bedingung $\det(I_n) = 1$ die Abbildung

$$\begin{aligned} (\mathbb{K}^n)^n &\to \mathbb{K}, \\ (v_1, \dots, v_n) &\mapsto \det\big([v_1 \ \dots \ v_n]\big) \end{aligned}$$

n-linear und scherungsinvariant ist).

Bitte beachten Sie, dass wir bisher noch gar nicht geklärt haben, ob überhaupt eine Determinante existiert. Dies werden wir in Theorem 7.2.7 tun.

Zunächst leiten wir aber einige Folgerungen aus den Axiomen der Determinante her und beweisen darauf aufbauend die Eindeutigkeit.

Einige Eingeschaften der Determinante

Wir beweisen diese Aussagen unter etwas allgemeineren Voraussetzungen, nämlich für scherungsinvariante n-lineare Abbildungen auf beliebigen Vektorräumen. Indem man als Vektoren in der folgenden Proposition die Spalten einer Matrix verwendet, erhält man unmittelbar auch Aussagen über Determinanten.

Proposition 7.1.4. *Sei V ein Vektorraum über einem Körper $\mathbb{K}$, sei $n \in \mathbb{N}^*$ und sei $F : V^n \to \mathbb{K}$ eine n-lineare und scherungsinvariante Abbildung. Dann gilt:*

(a) *Für alle $v_1, , \ldots, v_n \in V$, alle $\alpha \in \mathbb{K}$ und alle $j, k \in \{1, \ldots, n\}$ mit $j \neq k$ gilt*

$$F(v_1, \ldots, v_n) = F(v_1, \ldots, v_{k-1}, v_k + \alpha v_j, v_{k+1}, \ldots, v_n).$$

(b) *Die Funktion F ist anti-symmetrisch.*

(c) *Seien $v_1, \ldots, v_n \in V$ derart, dass das Tupel $(v_1, \ldots, v_n)$ linear abhängig ist. Dann gilt*

$$F(v_1, \ldots, v_n) = 0.$$

Inbesondere ist also $F(v_1, \ldots, v_n) = 0$, falls einer der Vektoren $v_1, \ldots, v_n$ gleich 0 ist oder falls zwei der Vektoren gleich sind.

Beweis. Um (a) und (b) zu beweisen, genügt es, diese Aussagen für den Fall $n = 2$ zu beweisen – denn die Aussage betreffen jeweils nur zwei der Vektoren $v_1, \ldots, v_n$, und indem wir alle bis auf die zwei jeweils betroffenen Vektoren festhalten, erhalten wir eine 2-lineare Abbildung $V^2 \to \mathbb{K}$

(a) Sei, wie so eben erläutert $n = 2$. Seien $v_1, v_2 \in V$ und $\alpha \in \mathbb{K}$. Im Falle $\alpha = 0$ ist nichts zu zeigen, also sei $\alpha \neq 0$. Dann gilt

$$F(v_1, v_2) = \frac{1}{\alpha} F(\alpha v_1, v_2) = \frac{1}{\alpha} F(\alpha v_1, v_2 + \alpha v_1) = F(v_1, v_2 + \alpha v_1),$$

wobei die erste und die dritte Gleichung aus der Linearität im ersten Argument folgen und die zweite Gleichung aus der Scherungsinvarianz.

Analog dazu kann man $F(v_1, v_2) = F(v_1 + \alpha v_2, v_2)$ zeigen.

(b) Sei erneut $n = 2$. Zunächst bemerken wir, dass wegen der Scherungsinvarianz für alle v_1, v_2 gilt

$$F(v_1, v_2 - v_1) = F(v_1, v_2 - v_1 + v_1) = F(v_1, v_2).$$

Somit gilt für alle $v_1, v_2 \in V$

$$\begin{aligned} &F(v_1, v_2) = F(v_1, v_2 - v_1) = F(v_1 + (v_2 - v_1), v_2 - v_1) \\ &= F(v_2, v_2 - v_1) = F(v_2, (v_2 - v_1) - v_2) = F(v_2, -v_1) = -F(v_2, v_1), \end{aligned}$$

wobei wir für die letzte Gleichheit die Linearität im zweiten Argument benutzt haben.

(c) Diese Aussage hängt nicht nur von zwei der Vektoren $v_1, \dots, v_n$ ab, also können wir sie nicht auf den Fall $n = 2$ reduzieren, sondern müssen sie für jedes n direkt zeigen.

Seien also $v_1, \dots, v_n \in V$ derart, dass $(v_1, \dots, v_n)$ linear abhängig ist. Wenn $n = 1$ ist, bedeutet dies, dass $v_1 = 0$ ist und hieraus folgt in diesem Fall $F(v_1) = F(0 \cdot 0) = 0 \cdot F(0) = 0$.

Sei von nun an also $n \geq 2$. Aus der linearen Abhängigkeit folgt, dass es ein $k \in \{1, \dots, n\}$ gibt derart, dass v_k eine Linearkombination der anderen Vektoren ist, d.h. es gilt

$$v_k = \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_j v_j$$

für geeignete Skalare $\lambda_j \in \mathbb{K}$ ist. Durch $n-1$-fache Anwendung von Aussage (a) erhalten wir hieraus, dass

$$\begin{aligned} F(v_1, \dots, v_n) &= F(v_1, \dots, v_{k-1}, v_k - \sum_{\substack{j=1 \\ j \neq k}}^{n} \lambda_j v_j, v_{k+1}, \dots, v_n) \\ &= F(v_1, \dots, v_{k-1}, 0, v_{k+1}, \dots, v_n) = 0 \cdot F(v_1, \dots, v_{k-1}, 0, v_{k+1}, \dots, v_n) = 0. \end{aligned}$$

gilt. □

Eindeutigkeit der Determinante

Nun können wir zeigen, dass es auf $\mathbb{K}^{n \times n}$ höchstens eine Determinante geben kann.

Theorem 7.1.5 (Eindeutigkeit der Determinante)**.** *Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}^*$. Dann gibt es höchstens eine Determinante* $\det : \mathbb{K}^{n \times n} \to \mathbb{K}$.

Beweis. Sei $\det : \mathbb{K}^{n \times n} \to \mathbb{K}$ eine Determinante; dann erfüllt sie die Axiome aus der Definition von Determinanten sowie die Eigenschaften aus Proposition 7.1.4.

Sei nun $A \in \mathbb{K}^{n \times n}$. Es genügt zu zeigen, dass sich mit Hilfe der soeben genannten Axiome und Eigenschaften bereits der Wert von $\det(A)$ bestimmen lässt.

Wir unterscheiden nun zwei Fälle:

- *1. Fall:* Die Spalten von A sind linear abhängig. Laut Proposition (c) gilt dann $\det(A) = 0$.
- *2. Fall:* Die Spalten von A sind linear unabhängig. In diesem Fall ist A invertierbar. Somit ist auch die transponierte Matrix A^{T} invertierbar, und A^{T} hat somit die reduzierte Zeilenstufenform I_n. Dies bedeutet, dass wir A^{T} durch elementare Zeilenumformungen auf die Form I_n bringen können. Wenn wir stattdessen A selbst betrachten und anstelle elementarer Zeilenumformungen die analogen Umformungen mit den Spalten von A durchführen, erhalten wir durch diese Spaltenumformungen also die Matrix $I_n^{\mathrm{T}} = I_n$.

Weil det in jeder Spalte linear ist, in allen Spalten anti-symmetrisch ist, und außerdem die Eigenschaft (a) aus Proposition 7.1.4 erfüllt, wissen wir aber für jede dieser Spaltenumformungen, wie sie den Wert der Determinante ändert.[3]

Also ist $\det(A)$ bereits durch den Werte $\det(I_n)$ eindeutig bestimmt – und dieser Werte laut Definition von det gleich 1. □

Der oben stehende Beweis zeigt übrigens auch, dass die Determinante einer Matrix genau dann ungleich Null ist, wenn die Matrix invertierbar ist – allerdings können wir diese Beobachtung natürlich nur dann folgern, wenn wir gezeigt haben, dass es tatsächlich eine Determinante gibt. Deshalb kommen wir erst in Korollar 7.2.11 unten nochmal auf diese Beobachtung zurück, nachdem wir dann die Existenz der Determinante bewiesen haben.

Beachten Sie, dass der obenstehende Beweis auch gleich ein Verfahren zur Berechnung von $\det(A)$ liefert (sofern die Abbildung det denn überhaupt existiert – was wir ja noch nicht bewiesen haben). Lassen Sie uns dies an einem einfachen 3×3-Beispiel zeigen, in welchem wir für den Moment einfach so tun, also wüssten wir bereits, dass eine Abbildung det existiert, die die Axiome der Determinanten erfüllt.

Beispiel 7.1.6. Lassen Sie uns die Matrix

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}$$

betrachten. Wie im Beweis von Theorem 7.1.5 bringen wir A^{T} zunächst mit Hilfe des Gaußalgorithmus auf reduzierte Zeilenstufenform:

$$A^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & 1 \\ 2 & 2 & 0 \\ 1 & 1 & 1 \end{bmatrix} \overset{I\leftrightarrow II}{\rightsquigarrow} \begin{bmatrix} 2 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \overset{I/2}{\rightsquigarrow} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$\overset{III-I}{\rightsquigarrow} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \overset{II-III}{\rightsquigarrow} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{I-II}{\rightsquigarrow} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3.$$

Also ist A^{T} und somit auch A invertierbar.

Nun führen wir dieselben Zeilenoperationen wie oben mit A^{T} stattdessen als Spaltenoperationen mit A durch. Wir müssen im Prinzip also nur die obenstehenden Matrizen transponiert abschreiben. Allerdings wollen wir dabei $\det(A)$ berechnen und deshalb in jedem Schritt mit berücksichtigen, wie die jeweilige Spaltenoperation die Determinante ändert. Es gilt

$$\det(A) = \det \begin{bmatrix} 0 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} = -\det \begin{bmatrix} 2 & 0 & 1 \\ 2 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = -2\det \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

[3] Nämlich: Jede solche Spaltenumformung sorgt dafür, dass die Determinante gleichbleibt oder mit einem Wert ungleich 0 multipliziert wird.

$$= -2\det\begin{bmatrix}1&0&0\\1&1&0\\0&1&1\end{bmatrix} = -2\det\begin{bmatrix}1&0&0\\1&1&0\\0&0&1\end{bmatrix} = -2\det\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} = -2.$$

In der ersten Gleichheit haben wir einfach A eingesetzt; für die zweite Gleichheit haben wir verwendet, dass die Determinante anti-symmetrisch ist; für die dritte Gleichheit haben wir verwenden, dass die Determinanten linear in der ersten Spalte ist; für die nächsten drei Gleichheiten haben wir Proposition 7.1.4(a) verwendet, und für die letzte Gleichheit haben wir benutzt, dass laut Definition der Determinanten $\det(I_3) = 1$ gilt.

Die Rechnung wirkt ein wenig gewöhnungsbedürftig, weil mit Spalten operiert wird statt mit Zeilen, wie Sie es vom Gaußalgorithmus gewohnt sind. Glücklicherweise brauchen Sie sich daran aber gar nicht zu gewöhnen: Wir werden nämlich in diesem Kapitel noch zeigen, dass stets $\det(A) = \det(A^{\mathrm{T}})$ gilt (siehe Korollar 7.2.9 weiter unten), und somit können sie zur Berechnung der Determinanten auch direkt elementare Zeilenoperationen verwenden. Dies veranschaulichen wir mit derselben Matrix A wie oben noch einmal in Beispiel 7.2.10.

7.2 Leibnizformel und Existenz der Determinante

Wir wollen nun die Existenz einer Determinante zeigen, sowie eine explizite Formel für die Determinante angeben. Hierzu benötigen wir zuerst zwei andere Konzept zur Vorbereitung, nämlich das Gruppenhomomorphismus, sowie Vorzeichen von Permutationen.

Gruppenhomomorphismen

Ähnlich wie lineare Abbildungen zwischen Vektorräumen die Addition und die skalare Multiplikation respektieren, führen wir nun Abbildungen zwischen Gruppen ein, die die Gruppenverknüpfung respektieren:

Definition 7.2.1 (Gruppenhomomorphismen und -isomorphismen)**.** Seien $(G, \circ_G)$ und $(H, \circ_H)$ Gruppen und sei $\varphi : G \to H$.

(a) Man bezeichnet die Abbildung φ als *Gruppenhomomorphismus*, falls sie die folgende Eigenschaft erfüllt:

$$\forall a, b \in G: \quad \varphi(a \circ_G b) = \varphi(a) \circ_H \varphi(b).$$

(b) Man bezeichnet die Abbildung φ als *Gruppenisomorphismus*, falls φ bijektiv und ein Gruppenhomomorphismus ist.

Es folgen einige Eigenschaften von Gruppenhomomorphismen.

Proposition 7.2.2. *Seien $(G, \circ_G)$ und $(H, \circ_H)$ Gruppen, deren neutrale Elemente wir mit e_G und e_H bezeichnen, und sei $\varphi : G \to H$ ein Gruppenhomomorphismus.*

(a) *Es gilt* $\varphi(e_G) = e_H$.

(b) *Für alle* $a \in G$ *gilt* $\varphi(a^{-1}) = \big(\varphi(a)\big)^{-1}$.

(c) *Wenn* φ *bijektiv (also ein Gruppenisomorphismus) ist, dann ist auch die Umkehrabbildung* $\varphi^{-1} : H \to G$ *eine Gruppenisomorphismus.*

Beweis. (a) und (c) Den Beweis von (a) and (c) stellen wir als Übungsaufgabe.
(b) Es gilt

$$\varphi(a) \circ_H \varphi(a^{-1}) = \varphi(a \circ_G a^{-1}) = \varphi(e_G) = e_H;$$

für die erste Gleichung haben wir verwendet, dass φ ein Gruppenhomomorphismus ist, und für die letzte Gleichheit haben wir Aussage (a) benutzt. Indem wir die soeben bewiesene Gleichheit von links mit $\big(\varphi(a)\big)^{-1}$ verknüpfen, erhalten wir die Behauptung. □

Ein sehr einfaches Beispiel eines Gruppenhomomorphismus ist durch das Vorzeichen von Null verschiedener reeller Zahl gegeben:

Beispiel 7.2.3. Für jedes $r \in \mathbb{R}^*$ definieren wir das **Vorzeichen** oder **Signum** sgn von r durch die Formel

$$\operatorname{sgn}(r) = \begin{cases} 1, & \text{falls } r > 0, \\ -1, & \text{falls } r < 0. \end{cases}$$

Anders ausgedrück ist $\operatorname{sgn}(r) = \frac{r}{|r|}$ für alle $r \in \mathbb{R}^*$.

Es ist sgn ein Gruppenhomomorphismus von $(\mathbb{R}^*, \cdot)$ nach $(\{-1, 1\}, \cdot)$, denn für alle $r, s \in \mathbb{R}^*$ gilt

$$\operatorname{sgn}(r \cdot s) = \frac{r \cdot s}{|r \cdot s|} = \frac{r}{|r|}\frac{s}{|s|} = \operatorname{sgn}(r) \cdot \operatorname{sgn}(s),$$

was die Behauptung zeigt.

Das Vorzeichen von Permutationen

Nun geben wir ein etwas interessanteres Beispiel eines Gruppenhomomorphismus an: Das Vorzeichen von Permutationen. Zunächst sollten Sie sich aus Abschnitt 2.2 in Erinnerung rufen, was man unter Permutationen und unter der symmetrischen Gruppen $\mathcal{S}_n$ versteht.

Definition 7.2.4 (Vorzeichen von Permutationen)**.** Sei $n \in \mathbb{N}^*$ und sei $\sigma \in \mathcal{S}_n$. Dann definieren wir das **Vorzeichen** oder auch **Signum** von σ als

$$\operatorname{sgn}(\sigma) \coloneqq \prod_{1 \le j < k \le n} \operatorname{sgn}\big(\sigma(k) - \sigma(j)\big) \in \{-1, 1\}.$$

Das bedeutet also: Für eine Permutation $\sigma \in \mathcal{S}_n$ sieht man sich an, wo es vorkommt, dass für zwei Zahlen $j, k \in \{1, \ldots, n\}$ mit $j < k$ die Ungleichung $\sigma(j) > \sigma(k)$ gilt; jedes solche Vorkommnis ist ein sogenannter **Fehlstand** der Permutation σ. Für solche Fehlstände ist $\operatorname{sgn}\big(\sigma(k) - \sigma(j)\big)$ gleich -1, und sonst ist diese Zahl gleich 1. Wenn die Anzahl der Fehlstände in σ also ungerade ist, dann ist $\operatorname{sgn}(\sigma) = -1$, und wenn die Anzahl der Fehlstände gerade ist, dann gilt $\operatorname{sgn}(\sigma) = 1$.

Hier ein konkretes und ein allgemeines Beispiel:

Beispiele 7.2.5. (a) Lassen Sie uns die Permutation

$$\sigma \coloneqq \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

betrachten. Sie hat einen Fehlstand an der Position $(1, 2)$ (wegen $\sigma(1) > \sigma(2)$) und an der Position $(1, 3)$ (wegen $\sigma(1) > \sigma(3)$). An der Position $(2, 3)$ hat Sie hingegen keinen Fehlstand (wegen $\sigma(2) < \sigma(3)$). Also besitzt σ genau zwei Fehlstände, d.h. es ist $\operatorname{sgn}(\sigma) = 1$.

Man kann dies auch sehen, indem man direkt die Formel aus Definition 7.2.4 anschreibt: Es gilt

$$\begin{aligned} \operatorname{sgn}(\sigma) &= \prod_{1 \le j < k \le 3} \operatorname{sgn}\big(\sigma(k) - \sigma(j)\big) \\ &= \operatorname{sgn}\big(\sigma(2) - \sigma(1)\big) \cdot \operatorname{sgn}\big(\sigma(3) - \sigma(1)\big) \cdot \operatorname{sgn}\big(\sigma(3) - \sigma(2)\big) \\ &= \operatorname{sgn}\big(1 - 3\big) \cdot \operatorname{sgn}\big(2 - 3\big) \cdot \operatorname{sgn}\big(2 - 1\big) = (-1) \cdot (-1) \cdot 1 = 1. \end{aligned}$$

(b) Sei $n \in \mathbb{N}^*$. Eine Permutation $\sigma \in \mathcal{S}_n$, die nur zwei Zahlen vertauscht und alle anderen gleich lässt, nennt man eine **Transposition**. Man kann sich überlegen, dass die Anzahl der Fehlstände einer Transposition immer ungerade ist, und somit jedes Transposition das Vorzeichen -1 hat.

Die Abbildung, die jeder Permutation aus $\mathcal{S}_n$ (für ein festes $n \in \mathbb{N}^*$) ihr Vorzeichen zuordnet, ist ein Gruppenhomomorphismus:

Proposition 7.2.6. *Sei $n \in \mathbb{N}^*$. Es ist $\operatorname{sgn} : S_n \to \{-1, 1\}$ eine Gruppenhomomorphismus (wobei $\{-1, 1\}$ mit der Multiplikation als Verknüpfung ausgestattet ist).*

Den Beweis lagern wir aus Zeitgründen in die Ergänzungen am Ende des Kapitels aus.

Existenz von Determinanten und die Leibnizformel

Nun beweisen wir, wie angekündigt, die Existenz von Determinanten – und zwar, indem wir sie mit Hilfe einer expliziten Formel einfach hinschreiben. Beachten Sie, dass wir die Eindeutigkeit schon in Theorem 7.1.5 gezeigt haben – somit gibt es dem folgenden Theorem nach also genau eine Determinante $\det : \mathbb{K}^{n \times n} \to \mathbb{K}$, und es ist deshalb gerechtfertigt von **der** Determinante zu sprechen.

Theorem 7.2.7 (Existenz und Leibnizformel)**.** *Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}^*$. Es gibt genau eine Determinante*[4] $\det : \mathbb{K}^{n\times n} \to \mathbb{K}$*, und diese ist durch die Formel*

$$\det A = \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} A_{\sigma(j),j}$$

gegeben.

Den Beweis dieses Theorems sparen wir an dieser Stelle aus und verweisen stattdessen auf die Literatur, zum Beispiel auf [Beu14, Abschnitt 7.4]. Als nächstes geben wir in den Dimensionen 1 und 2 eine einfache explizite Formel für die Determinante an; hierbei handelt es sich um Spezialfälle der Leibnizformel.

Korollar 7.2.8. *Sei $\mathbb{K}$ ein Körper.*

(a) *Für jedes $A \in \mathbb{K}^{1\times 1} = \mathbb{K}$ gilt $\det(A) = A$.*

(b) *Für jedes $A \in \mathbb{K}^{2\times 2}$ gilt*

$$\det(A) = A_{11}A_{22} - A_{21}A_{12}.$$

Beweis. (a) Sei $A \in \mathbb{K}$. Unmittelbar aus der Definition der Determinanten folgt

$$\det(A) = \det(A \cdot I_1) = A \det(I_1) = A \cdot 1 = A.$$

(b) Dies folgt aus der Leibnizformel in Theorem 7.2.7, indem man die Summe und die Produkte für $n = 2$ ausschreibt. □

Wie bereits erwähnt, können Sie den Beweis von Theorem 7.2.7 bei Interesse am Ende dieses Kapitels nachlesen. An dieser Stelle sei noch angemerkt, dass es lehrreich ist, zumindest für den Fall $n = 2$ (der in Korollar 7.2.8(b) explizit ausgeschrieben ist) nachrzurechnen, dass diese Formel alle Axiome der Determinante erfüllt.

Eine weitere Konsequenz der Leibnizformel ist folgender Zusammenhang zwischen der Determinanten einer Matrix und ihrer transponierten Matrix:

Korollar 7.2.9. *Sei $\mathbb{K}$ und Körper und $n \in \mathbb{N}^*$. Für jede Matrix $A \in \mathbb{K}^{n\times n}$ gilt*

$$\det(A) = \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} A_{j,\sigma(j)} = \det(A^{\mathrm{T}}).$$

Beweis. Sei $A \in \mathbb{K}^{n\times n}$. Wir rechnen die behauptete Gleichheit von rechts nach links nach: Laut Theorem 7.2.7, angewendet auf die Matrix A^{T}, gilt

$$\det(A^{\mathrm{T}}) = \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} (A^{\mathrm{T}})_{\sigma(j),j} = \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} A_{j,\sigma(j)}$$

[4] Die wir deshalb von nun an als **die** Determinante auf $\mathbb{K}^{n\times n}$ bezeichnen und mit det notieren.

$$= \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma^{-1}) \prod_{j=1}^{n} A_{\sigma^{-1}(j),j} = \sum_{\sigma\in\mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} A_{\sigma(j),j} = \det(A);$$

für die Gleichheit zwischen erster und zweiter Zeile haben wir einerseits verwendet, dass $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$ gilt, und andererseits haben wir die Faktoren im Produkt $\prod_{j=1}^{n} A_{j,\sigma(j)}$ umsortiert (indem wir den j-ten Faktor an die $\sigma^{-1}(j)$-te Stelle im Produkt verschoben haben); für die mittlere Gleichung in der zweiten Zeile haben wir benutzt, dass $\mathcal{S}_n \ni \sigma \mapsto \sigma^{-1} \in \mathcal{S}_n$ eine Bijektion ist; und für die letzte Gleichheit haben wir erneut Theorem 7.2.7 verwendet, aber dieses Mal für die Matrix A selbst. □

In Beispiel 7.1.6 hatten wir verwendet, dass wir verstehen, wie eine Determinante sich durch Umformungen der Spalten verändert, wenn die Spaltenumformungen so sind, wie wir es aus dem Gauß-Algorithmus eigentlich für Zeilenumformungen gewohnt sind.

Aus Korollar 7.2.9 folgt nun unmittelbar, dass die Determinante sich bei elementaren Zeilenumformungen einer Matrix genauso verhält wie bei den entsprechden Spaltenumformungen. Das macht das Berechnen von Determinanten einfacher, denn wir können jetzt wie gewohnt den Gauß-Algorithmus verwenden und müssen in jedem Schritt lediglich darauf achten, wie sich die Determinante ändert. Lassen Sie uns das anhand der Matrix aus Beispiel 7.1.6 noch einmal demonstrieren:

Beispiel 7.2.10. Wir betrachten erneut die Matrix

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

Anstatt nun den Gaußalgorithmus auf A^{T} anzuwenden, wenden wir ihn direkt auf die Matrix A selbst an und beachten dabei in jedem Schritt, wie sich die Determinante ändert. Es gilt

$$\begin{aligned}
\det A = \det \begin{bmatrix} 0 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} &\overset{I\leftrightarrow II}{=} -\det \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \\
&\overset{III-I}{=} -\det \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \\ 0 & -2 & 0 \end{bmatrix} \overset{III+II;\, I-II}{=} -\det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \\
&\overset{II-III}{=} -\det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{II\cdot\frac{1}{2}}{=} -2\det \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = -2.
\end{aligned}$$

Also erhalten wir – wie es ja auch sein muss – dasselbe Ergebnis wie in Beispiel 7.1.6.

Nach dem wir nun aus Theorem 7.2.7 wissen, dass es genau eine Determinante gibt, lohnt es sich, noch die folgenden Beobachtung aus dem Beweis von Theorem 7.1.5 explizit festzuhalten:

Korollar 7.2.11. *Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$. Es gilt $\det(A) \neq 0$ genau dann, wenn A invertierbar ist.*

7.3 Weitere Eigenschaften von Determinanten

In diesem Abschnitt besprechen wir noch einige weitere Eigenschaften von Determinanten.

Der Determinaten-Multiplikationssatz

Die folgende Eigenschaft von Determinanten ist auf den ersten Blick vielleicht etwas überraschend.

Theorem 7.3.1 (Determinanten-Multiplikationssatz)**.** *Sei $\mathbb{K}$ ein Körper, sei $n \in \mathbb{N}^*$ und seien $A, B \in \mathbb{K}^{n\times n}$. Dann gilt*

$$\det(AB) = \det(A)\det(B).$$

Beweis. Wenn A nicht invertierbar ist, dann ist auch AB nicht invertierbar,[5] und somit ist laut Korollar 7.2.11 sowohl $\det(AB) = 0$ als auch $\det(A) = 0$.

Sei nun also A invertierbar, und somit $\det(A) \neq 0$ laut Korollar 7.2.11. Wir betrachten die Abbildung

$$\begin{aligned}\widehat{\det} : \mathbb{K}^{n\times n} &\to \mathbb{K},\\ C &\mapsto \frac{1}{\det(A)}\det(AC).\end{aligned}$$

Weil die Abbildung det in den Spalten ihres Arguments n-linear und scherungsinvariant ist, kann man leicht nachrechnen, dass $\widehat{\det}$ ebenfalls diese Eigenschaften hat. Außerdem gilt

$$\widehat{\det}(I_n) = \frac{1}{\det(A)}\det(A) = 1.$$

Somit ist $\widehat{\det}$ eine Determinante. Laut Theorem 7.1.5 gibt es aber nur eine Determinante, d.h. es folgt $\widehat{\det} = \det$, und somit

$$\det(B) = \widehat{\det}(B) = \frac{1}{\det(A)}\det(AB),$$

was die Behauptung zeigt. □

[5] Warum nicht?

Entwicklung von Determinanten

Nun kommen wir noch zu einer weiteren Eigenschaft von Determinante, dem sogenannten Entwicklungssatz. Er liefert eine Möglichkeit die Determinante einer Matrix rekursiv über die Dimension zu berechnen.[6]

Theorem 7.3.2 (Entwicklungssatz)**.** *Sei $\mathbb{K}$ ein Körper und sei $n \in \mathbb{N}^*$. Dann gilt:*

(a) *Falls $n = 1$ ist, gilt $\det A = A$ für alle $A \in \mathbb{K}^{1\times 1} = \mathbb{K}$.*

(b) *Falls $n \geq 2$ ist, gilt für jedes $j \in \{1, \ldots, n\}$ und alle $A \in \mathbb{K}^{n\times n}$ die Formel*

$$\det A = \sum_{k=1}^{n} (-1)^{j+k} A_{jk} \det A^{(jk)},$$

wobei $A^{(jk)} \in \mathbb{K}^{(n-1)\times(n-1)}$ jeweils diejenige Matrix bezeichnet, die aus A entsteht, indem man die j-Zeile und die k-te Spalte streicht.

(c) *Falls $n \geq 2$ ist, gilt analog auch für jedes $k \in \{1, \ldots, n\}$ und alle $A \in \mathbb{K}^{n\times n}$ die Formel*

$$\det A = \sum_{j=1}^{n} (-1)^{j+k} A_{jk} \det A^{(jk)}$$

(mit derselben Notation wie in (b)*).*

Beweis. Für den Beweis verweisen wir auf die Literatur, zum Beispiel auf [Beu14, Seiten 225–228]. □

Die Formel in (b) nennt man **Entwicklung nach der j-ten Zeile**, und die Formel in (c) nennt man **Entwicklung nach der k-ten Spalte**.

Wir demonstrieren die Entwicklung nach der ersten Zeile an einem Beispiel:

Beispiele 7.3.3. Lassen Sie uns nochmals die Matrix

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

betrachten, die Sie nun schon aus mehreren Beispielen kennen. Indem wir die Determinante nach der ersten Zeile entwicklen erhalten wir

$$\begin{aligned} \det A &= 0 \cdot \det \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} - 2 \cdot \det \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + 1 \cdot \det \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix} \\ &= -2(1 \cdot 1 - 1 \cdot 1) + 1(1 \cdot 0 - 1 \cdot 2) = -2; \end{aligned}$$

für die letzte Gleichheit haben wir die Formel für die Determinante von 2×2-Matrizen aus Korollar 7.2.8(b) verwendet.

[6] Trotzdem ist der Satz eher von theoretischem Interesse, denn für die praktische Berechnung der Determinante einer Matrix ist der Satz meist nur dann geeignet, wenn man eine Matrix vorliegen hat, die in einer Zeile oder Spalte sehr viele Nulleinträge hat.

7.4 Flächen und Volumen

Nun kommen wir zur geometrischen Fragestellung vom Anfang des Kapitel zurück: Wir wollen Flächen und Volumina.

Um in Gefilden mit klarer geometrischer Intuition zu bleiben, sehen wir uns nur die Dimensionen 2 und 3 an.

Theorem 7.4.1. (a) *Für zwei Vektoren $v_1, v_2 \in \mathbb{R}^2$ ist*

$$\left|\det \begin{bmatrix} v_1 & v_2 \end{bmatrix}\right| \in [0, \infty)$$

die Fläche des Parallelogramms, das von v_1 und v_2 aufgespannt wird. Für $v_1, v_2, v_3 \in \mathbb{R}^2$ ist somit

$$\frac{1}{2}\left|\det \begin{bmatrix} v_2 - v_1, & v_3 - v_1 \end{bmatrix}\right| \in [0, \infty)$$

die Fläche des Dreiecks mit den Eckpunkten v_1, v_2 und v_3.

(b) *Für drei Vektoren $v_1, v_2, v_3 \in \mathbb{R}^3$ ist*

$$\left|\det \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}\right| \in [0, \infty)$$

das Volumen des Parallelepipeds, das von v_1, v_2 und v_3 aufgespannt wird. Für $v_1, v_2, v_3, v_4 \in \mathbb{R}^3$ ist somit

$$\frac{1}{6}\left|\det \begin{bmatrix} v_2 - v_1, & v_3 - v_1, & v_4 - v_1 \end{bmatrix}\right| \in [0, \infty)$$

das Volument des Tetraeders mit den Eckpunkten v_1, v_2, v_3 und v_4.

Einen exakten Beweis dieses Theorems können wir im Rahmen dieser Vorlesung nicht geben – schon allein deshalb nicht, weil es gar nicht so klar ist, was eigentlich eine präzise mathematische Definition der Größen **Fläche** und **Volumen** ist (einen sehr allgemeinen axiomatischen Zugang zu diesen Fragen liefert die **Maßtheorie**, in welcher man das sogenannte **Lebesgue-Maß** einführt). Intuitiv sollten die Formeln für die Parallelogrammfläche und das Volumen eines Parallepipeds Sie aber nicht überraschen, denn wir haben die Determinante ja so definiert, dass Ihre Eigenschaften zu den Eigenschaften von Parallelogrammflächen passen, die wir in Diskussion 7.1.1 besprochen haben.[7]

Die Formeln für Dreiecksflächen und Tetraedervolumen im Theorem folgen wiederum aus den entsprechenden Formeln für Paralleogramme und Parallelepipede.

Lassen Sie uns als einfache Anwendung noch einmal auf die erste Einstiegsfrage am Beginn des Kapitels zurückkommen:

[7] Für das Volumen von Parallelepipeden im $\mathbb{R}^3$ erwartet man natürlich analoge Eigenschaften.

Beispiel 7.4.2. Wir betrachten das Viereck im $\mathbb{R}^2$ mit den vier Eckpunkten

$$v_1 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 5 \end{bmatrix}, \quad v_3 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \quad v_4 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

und wollen seine Fläche berechnen. Hierzu unterteilen wir es an das Dreick mit den Eckpunkten v_1, v_2, v_3 – nennen wir es Δ_1 – und in das Dreieck mit den Eckpunkten v_1, v_3, v_4 – nennen wir es Δ_2.

Die Fläche von Δ_1 ist laut Theorem 7.4.1(a) gleich

$$\frac{1}{2}\left|\det\begin{bmatrix} v_2 - v_1, & v_3 - v_1 \end{bmatrix}\right| = \frac{1}{2}\left|\det\begin{bmatrix} -4 & -7 \\ 5 & 1 \end{bmatrix}\right| = \frac{31}{2},$$

und die Fläche von Δ_2 ist gleich

$$\frac{1}{2}\left|\det\begin{bmatrix} v_3 - v_1, & v_4 - v_1 \end{bmatrix}\right| = \frac{1}{2}\left|\det\begin{bmatrix} -7 & -3 \\ 1 & -1 \end{bmatrix}\right| = 5.$$

Also besitzt das gesamte Viereck die Fläche

$$\frac{31}{2} + 5 = 20{,}5.$$

Berechnung von Flächen im $\mathbb{R}^3$

Nun wollen wir noch besprechen, wie man Flächen von Parellogrammen und Dreiecken im $\mathbb{R}^3$ berechnen kann.

Diskussion 7.4.3 (Parallelogrammflächen im $\mathbb{R}^3$)**.** Seien $v, w \in \mathbb{R}^3$. Wenn wir die Fläche F des Parallelogramms berechnen wollen, das von v und w aufgespannt wird, können wir weder die Determinanten von 3×3-Matrizen noch die Determinante von 2×2-Matrizen berechnen – denn $[v\ w]$ ist ja eine 3×2-Matrix, und somit ist für diese Matrix gar keine Determinante definiert.[8]

Wir können aber trotzdem eine Möglichkeit finden die Fläche zu berechnen. Lassen Sie uns dies anhand mehrerer Schritte diskutieren:

- *Schritt 1:* Zunächst betrachten wir den Spezialfall, dass die beiden Vektoren v und w in der x_1x_2-Ebene liegen; dies bedeutet, dass ihre letzten Einträge jeweils 0 sind.

  Wir könnten nun einfach die letzten Einträge weglassen, würden somit zwei Vektoren im $\mathbb{R}^2$ erhalten und könnten dann die Fläche des aufgespannten Parallelogramms mithilfe der Determinante berechnen. Dies liefert das richtige Ergebnis. Es gibt aber noch ein alternatives Vorgehen, bei dem man nicht einzelne Einträge der Vektoren weglassen muss, und dieses Vorgehen stellt sich als praktischer heraus, wenn man eine allgemeine Formel herleiten möchte:

[8] Es ist wichtig, dass Sie sich dies noch einmal bewusst machen: Determinanten sind nur für quadratische Matrizen definiert!

Wir überlegen uns, dass die Fläche F des Parallelogramms natürlich genauso groß ist, wie das Volumen V des Parallelepipeds, das entsteht, wenn wir das Parallelogramm in die dritte Raumrichtung mit Länge 1 ausdehnen. Es gilt also

$$F = V = \left|\det \begin{bmatrix} v & w & e_3 \end{bmatrix}\right|$$

(hierbei ist es aber wesentlich, dass v und w beide wirklich in der x_1x_2-Ebene liegen).

- *Schritt 2*: Als nächstes beachten wir, dass für jede reelle 3×3-Matrix A die Gleicheit

$$|\det(A)| = \big(\det(A) \cdot \det(A)\big)^{1/2} = \big(\det(A^{\mathrm{T}}) \cdot \det(A)\big)^{1/2} = \big(\det(A^{\mathrm{T}}A)\big)^{1/2}$$

gilt; für die zweite Gleichheit haben wir Korollar 7.2.9 verwendet, und für die dritte Gleichheit den Determinanten-Multiplikationssatz.

Wenn wir dies auf die Formel aus Schritt 1 anwenden, dann erhalten wir, dass F^2 gleich der Determinante von

$$\begin{bmatrix} v & w & e_3 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} v & w & e_3 \end{bmatrix} = \begin{bmatrix} v^{\mathrm{T}} \\ w^{\mathrm{T}} \\ e_3^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} v & w & e_3 \end{bmatrix} = \begin{bmatrix} v^{\mathrm{T}}v & v^{\mathrm{T}}w & 0 \\ w^{\mathrm{T}}v & w^{\mathrm{T}}w & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

ist. Durch Entwicklung nach der dritten Zeile erhalten wir somit

$$F^2 = \det \begin{bmatrix} v^{\mathrm{T}}v & v^{\mathrm{T}}w \\ w^{\mathrm{T}}v & w^{\mathrm{T}}w \end{bmatrix} = \det \left(\begin{bmatrix} v^{\mathrm{T}} \\ w^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} v & w \end{bmatrix} \right) = \det \left(\begin{bmatrix} v & w \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} v & w \end{bmatrix} \right).$$

- *Schritt 3:* Zuletzt zeigen wir nun, dass die soeben hergeleitete Formel für F (bzw. F^2) sogar für alle Vektoren $v, w \in \mathbb{R}^3$ gilt – auch dann, wenn sie nicht in irgendeiner Koordinatenebene liegen.

Hierzu drehen wir den kompletten $\mathbb{R}^3$ so, dass v und w nach der Drehung in der x_1x_2-Ebene zum Liegen kommen. Solche eine Drehung ist linear und wird somit durch Multiplikation mit einer Matrix $U \in \mathbb{R}^{3\times 3}$ beschrieben. Weil man eine Drehung rückgängig machen kann, ist U invertierbar, und eine Drehung erhält natürlich die Fläche unseres Parallelogramms. Somit ist F^2 gleich der quadrierten Fläche des Parallelogramm, welches von Uv und Uw augespannt wird, und weil die beiden letztgenannten Vektoren in der x_1x_2-Ebenen liegen, folgt somit aus der Formel, die wir uns in Schritt 2 überlegt haben,

$$\begin{aligned} F^2 &= \det \left(\begin{bmatrix} Uv & Uw \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} Uv & Uw \end{bmatrix} \right) = \det \left(\left(U \begin{bmatrix} v & w \end{bmatrix} \right)^{\mathrm{T}} \left(U \begin{bmatrix} v & w \end{bmatrix} \right) \right) \\ &= \det \left(\begin{bmatrix} v & w \end{bmatrix}^{\mathrm{T}} U^{\mathrm{T}} U \begin{bmatrix} v & w \end{bmatrix} \right). \end{aligned}$$

Zuletzt benötigen wir nun noch ein Resultat, welches in der Linearen Algebra 2 behandelt wird:[9] Eine Matrix U, die eine Drehung beschreibt, hat immer die Eigenschaft, dass ihre inverse Matrix gleich ihrer transponierten Matrix ist, d.h. es gilt $U^{\mathrm{T}} = U^{-1}$ und somit $U^{\mathrm{T}}U = I_3$. Somit folgt also

$$F^2 = \det\left(\begin{bmatrix} v & w \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} v & w \end{bmatrix}\right).$$

Man kann die Matrixmultiplikation auch noch explizit durchführen und dann die Determinante der so erhaltenen 2×2-Matrix bestimmen: Damit erhält man

$$F^2 = \det\begin{bmatrix} v^{\mathrm{T}}v & v^{\mathrm{T}}w \\ w^{\mathrm{T}}v & w^{\mathrm{T}}w \end{bmatrix} = (v^{\mathrm{T}}v)(w^{\mathrm{T}}w) - (v^{\mathrm{T}}w)^2;$$

für die letzte Gleichheit haben wir verwendet, dass $(w^{\mathrm{T}}v) = (w^{\mathrm{T}}v)^{\mathrm{T}} = v^{\mathrm{T}}w$ gilt, weil $w^{\mathrm{T}}v \in \mathbb{K}^{1\times 1}$ ist.

Indem man die Wurzel zieht, hat man somit eine einfache und kompakte Formel für die Fläche F gefunden.

Lassen Sie uns die Formel, die wir soeben hergeleitet haben, in einem Theorem festhalten:

Theorem 7.4.4. *Seien $v, w \in \mathbb{R}^3$. Das Parallogramm, das von v und w aufgespannt wird, hat die Fläche*

$$\sqrt{\det\left(\begin{bmatrix} v & w \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} v & w \end{bmatrix}\right)} = \sqrt{(v^{\mathrm{T}}v)(w^{\mathrm{T}}w) - (v^{\mathrm{T}}w)^2}.$$

Beachten Sie, dass die Ausdrücke $v^{\mathrm{T}}v$, $w^{\mathrm{T}}w$ und $v^{\mathrm{T}}w$ jeweils Zahlen sind. Indem man die entsprechende Matrixmultiplikation in allen drei Ausdrücken explizit ausführt, erhält man

$$\begin{aligned} v^{\mathrm{T}}v &= v_1^2 + v_2^2 + v_3^2, \\ v^{\mathrm{T}}v &= w_1^2 + w_2^2 + w_3^2, \\ v^{\mathrm{T}}w &= v_1w_1 + v_2w_2 + v_3w_3. \end{aligned}$$

Diese Ausdrücke haben ebenfalls eine tieferliegende geometrischen Bedeutung, die in der Linearen Algebra 2 erklärt wird.

Aus dem Theorem kann man natürlich auch sofort eine Formel zur Berechnung von Dreiecksflächen angeben: Wenn das betrachtete Dreieck die Eckpunkte u, v, w besitzt, so kann man das Theorem einfach auf die Vektoren $v-u$ und $w-u$ anwenden und das Ergebnis durch 2 teilen.

Wir demonstrieren dies anhand eines einfachen Beispiels:

[9] Das heißt, wenn man diesen letzten Schritt der Herleitung verstehen will, muss man zuerst die Lineare Algebra 2 (oder Teile davon) hören. Es erscheint aber trotzdem sinnvoll, diese Herleitung hier schon einmal zu zeigen, damit Sie sehen, wie Determinanten auch zur Berechnung von Parallelogrammflächen im $\mathbb{R}^3$ verwendet werden können.

Beispiele 7.4.5. Betrachten wir im $\mathbb{R}^3$ das Dreieck mit den drei Eckpunkten e_1, e_2, e_3 (dies ist das sogenannten **Einheitssimplex** oder auch **Wahrscheinlichkeitsimplex**). Indem wir

$$v \coloneqq e_2 - e_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \qquad \text{und} \qquad w \coloneqq e_3 - e_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

setzen, erhalten wir für die Fläche des Einheitssimplex den Wert

$$\frac{1}{2}\sqrt{(v^{\mathrm{T}}v)(w^{\mathrm{T}}w) - (v^{\mathrm{T}}w)^2} = \frac{1}{2}\sqrt{2 \cdot 2 - 1^2} = \frac{\sqrt{3}}{2}.$$

Mit demselben Vorgehen kann man zum Beispiel die Oberfläche von Tetraedern berechnen, indem man für jede der vier Oberflächendreiecke die Fläche berechnet und alle vier Flächen addiert.

7.5 Ergänzungen

Das Vorzeichen von Permutationen

In der Vorlesung hatten wir den Beweis von Proposition 7.2.6 (welche besagt, dass das Vorzeichen von Permutationen ein Gruppenhomomorphismus ist) ausgespart. Bei Interesse können Sie ihn im Folgenden nachlesen:

Beweis von Proposition 7.2.6. Seien $\sigma, \tau \in \mathcal{S}_n$. Dann gilt

$$\begin{aligned}
\frac{\operatorname{sgn}(\sigma \circ \tau)}{\operatorname{sgn}(\tau)} &= \prod_{1 \le j < k \le n} \frac{\operatorname{sgn}\big(\sigma(\tau(k)) - \sigma(\tau(j))\big)}{\operatorname{sgn}\big(\tau(k) - \tau(j)\big)} \\
&= \prod_{\substack{1 \le j < k \le n \\ \tau(k) > \tau(j)}} \frac{\operatorname{sgn}\big(\sigma(\tau(k)) - \sigma(\tau(j))\big)}{\operatorname{sgn}\big(\tau(k) - \tau(j)\big)} \cdot \prod_{\substack{1 \le j < k \le n \\ \tau(k) < \tau(j)}} \frac{\operatorname{sgn}\big(\sigma(\tau(k)) - \sigma(\tau(j))\big)}{\operatorname{sgn}\big(\tau(k) - \tau(j)\big)} \\
&= \prod_{\substack{1 \le j < k \le n \\ \tau(k) > \tau(j)}} \operatorname{sgn}\big(\sigma(\tau(k)) - \sigma(\tau(j))\big) \cdot \prod_{\substack{1 \le j < k \le n \\ \tau(k) < \tau(j)}} \operatorname{sgn}\big(\sigma(\tau(j)) - \sigma(\tau(k))\big) \\
&= \prod_{\substack{1 \le j < k \le n \\ \tau(k) > \tau(j)}} \operatorname{sgn}\big(\sigma(\tau(k)) - \sigma(\tau(j))\big) \cdot \prod_{\substack{1 \le k < j \le n \\ \tau(k) > \tau(j)}} \operatorname{sgn}\big(\sigma(\tau(k)) - \sigma(\tau(j))\big) \\
&= \prod_{\substack{1 \le j, k \le n \\ \tau(k) > \tau(j)}} \operatorname{sgn}\big(\sigma(\tau(k)) - \sigma(\tau(j))\big) \\
&= \prod_{1 \le j < k \le n} \operatorname{sgn}\big(\sigma(k) - \sigma(j)\big) = \operatorname{sgn}(\sigma);
\end{aligned}$$

für die vierte Gleichheit haben wir im rechtsstehenden Produkt die Benennung der Indizies j und k vertauscht. □

Die Cramersche Regel

Man kann Determinanten auch verwenden, um eine Formel für die inverse Matrix anzugeben – die sogenannten **Cramersche Regel**. In höherer Dimension ist sie zur tatsächlichen Berechnung von inversen Matrizen äußerst ungeeignet, aber Sie ist von theoretischem Nutzen.

Sie können die Regel in der Literatur nachlesen, zum Beispiel in [Bos14, Abschnitt 4.4].

Inhalte im $\mathbb{R}^n$

Man kann Längen, Flächen und Volumen auf höher-dimensionale Räumen verallgemeinern und erhält dann auf $\mathbb{R}^n$ einen n-dimensionalen Inhalt. Wenn man nun im $\mathbb{R}^n$ ein geometrisches Objekt hat, welches Teilmenge eines affinen Unterraums von Dimension höchstens $n-1$ ist, dann ist intuitiv klar, dass dieses den Inhalt 0 haben wird – genauso wie zum Beispiel Linien und Oberflächen im $\mathbb{R}^3$ das Volumen 0 haben.

Allerdings haben Oberflächen im $\mathbb{R}^3$ natürlich eine Fläche, und Linien haben eine Länge. Dies kann man abstrakt erfassen, indem man für $d \in \{1, \dots, n\}$ d-dimensionale Inhalte im $\mathbb{R}^n$ einführt. Für $n = 3$ und $d = 2$ misst man damit also Flächen im $\mathbb{R}^3$. Die Erkenntnisse aus Diskussion 7.4.3 kann man auf diese Situation verallgemeinern und erhält somit das folgende Resultat:

Theorem 7.5.1. *Seien $1 \leq d \leq n$ natürliche Zahlen und seien $v_1, \dots, v_d \in \mathbb{R}^n$. Dann ist der d-dimensionale Inhalt I des Parallelepipeds, das von $v_1, \dots, v_d$ aufgespannt wird, gegeben durch*

$$I = \sqrt{\det\left(\begin{bmatrix} v_1 & \dots & v_d \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} v_1 & \dots & v_d \end{bmatrix}\right)}.$$

Es ist interessant, einige Spezialfälle der Formel zu disktuieren:

- Für $d = n$ erhält man mit Hilfe des Determinantenmultiplikationssatzes erwartungsgemäßg die Formel

$$I = \left|\det \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}\right|.$$

- Für $d = 2$ landet man wieder bei der Formel aus Theorem 7.4.4, nämlich

$$I = \sqrt{(v_1^{\mathrm{T}} v_1)(v_2^{\mathrm{T}} v_2) - (v_1^{\mathrm{T}} v_2)^2};$$

der einzige Unterschied besteht darin, dass die Vektoren nun n Einträge haben statt 3 Einträge (und dass die Vektoren nun v_1 und v_2 heißen anstelle von v und w).

- Für $d = 1$ beschreibt I den eindimensionalen Inhalt des eindimensionalen Parallelepipeds, das vom Vektor v_1 aufgespannt wird – also einfacher ausgedrückt die Länge des Vektors $v_1 \in \mathbb{R}^n$. Und in der Tat erhalten wir aus der Formel in Theorem 7.5.1, dass

$$I = \sqrt{\det v_1^{\mathrm{T}} v_1} = \sqrt{v_1^{\mathrm{T}} v_1} = \sqrt{v_{11}^2 + \cdots + v_{1n}^2}$$

  gilt – was genau die Formel ist, die man aufgrund des Satzes von Pythagoras erwarten würde.

Der Witz an Theorem 7.5.1 ist natürlich, dass es auch für alle anderen Werte von d gilt.

Beispiel 7.5.2. Lassen Sie uns in $\mathbb{R}^4$ das Einheitssimplex betrachten – dies ist ein 3-dimensionales geometrisches Objekt in $\mathbb{R}^4$, welche durch die Eckpunkte e_1, e_2, e_3, e_4 begrenzt wird.

Mit Theorem 7.5.1 kann man den 3-dimensionalen Inhalt des 3-dimensionalen Parallelepipeds bestimmen, welches von den Vektoren

$$e_2 - e_1, \quad e_3 - e_1, \quad e_4 - e_1$$

aufgespannt wird. Dieser Inhalt hat den Wert 2.

Wenn man lediglich den 3-dimensionalen Inhalt des Einheitssimplex wissen möchte, muss man noch durch $3! = 6$ teilen (um dies sauber zu begründen, sind zum Beispiel Mittel der Integralrechnung nützlich, die Sie in den Vorlesungen Analysis 1 und 2 lernen werden). Also hat das Einheitssimplex in $\mathbb{R}^4$ den 3-dimensionalen Inhalt $\frac{2}{6} = \frac{1}{3}$.

Allgemeiner kann man sich mit ähnlicher Vorgehensweise überlegen, dass das Einheitssimplex in $\mathbb{R}^n$ den $(n-1)$-dimensionalen Inhalt

$$\frac{\sqrt{n}}{(n-1)!}$$

besitzt.

Literaturhinweise

Hier einige Hinweise zu Literaturstellen, in denen Determinanten aus anderer Perspektive dargestellt werden, als wir es hier getan haben:

- Neben dem axiomatischen Zugang, den wir verwendet haben, wählen manche Autorinnen und Autoren zum Beispiel den Weg, die Determinante direkt über die Leibnizformel zu definieren; dies ist zum Beispiel in [Mey00, Kapitel 6] der Fall.

  Alternativ kann man die Determinante auch rekursiv über den Entwicklungssatz *definieren*; dies wird zum Beispiel in [TT08, Abschnitt 11.3] und [Lan86, Kapitel VII] getan.

- Ein anderer Beweis des Determinantenmultiplikationssatzes (der darauf basiert, invertierbare Matrizen als Produkt möglichst einfacher Matrizen darzustellen), kann zum Beispiel in [Beu14, Abschnitt 7.6] nachgelesen werden.
- Man kann große Teile der Linearen Algebra auch ganz ohne Determinanten entwickeln. Dieser Zugang wird zum Beispiel vom US-amerikanischen Mathematiker Sheldon Axler stark propagiert und in seinem Buch *Linear Algebra Done Right* [Axl97] dargestellt (Determinanten kommen in dem Buch trotzdem vor, aber erst ganz am Ende).

Kapitel 8

Polynome und Ringe

Einstiegsfragen. (a) Ist $\mathbb{R}^{3\times 3}$ mit der üblichen Addition und der Matrixmultiplikation eigentlich ein Körper? Und $\mathbb{R}^{2\times 2}$? Und $\mathbb{R}^{1\times 1}$?

(b) Sei $x \in \mathbb{R}$. Multiplizieren Sie den Ausdruck

$$(3x^2 - x + 1)(5x^3 - 4x^2 + x + 2)$$

aus.

(c) Und jetzt allgemeiner: Sei $x \in \mathbb{R}$, und seien auch $\alpha_0, \alpha_1, \alpha_2$ sowie $\beta_0, \ldots, \beta_3$ relle Zahlen. Multiplizieren Sie den Ausdruck

$$(\alpha_2 x^2 + \alpha_1 x + \alpha_0)(\beta_3 x^3 + \beta_2 x^2 + \beta_1 x + \beta_0)$$

aus.

(d) Und jetzt noch allgemeiner: Sei $x \in \mathbb{R}$, seien $m, n \in \mathbb{N}$, und seien $\alpha_0, \ldots, \alpha_m$ sowie $\beta_0, \ldots, \beta_n$ reelle Zahlen. Multiplizieren Sie den Ausdruck

$$(\alpha_m x^m + \ldots \alpha_1 x^1 + \alpha_0 x^0)(\beta_n x^n + \cdots + \beta_1 x^1 + \beta_0 x^0)$$

aus.

(e) Könnnen Sie die Ausdrücke $\alpha_m x^m + \ldots \alpha_1 x^1 + \alpha_0 x^0$ und $\beta_n x^n + \cdots + \beta_1 x^1 + \beta_0 x^0$ aus der vorangehenden Frage auch mit Hilfe des Summenzeichens $\sum$ schreiben?

Übrigens: Die Zahl x^0, die in beiden Ausdrücken vorkommt, müsste man nicht extra dazu schreiben. Wieso nicht? Und was hat man davon, sie trotzdem in dieser Form hinzuschreiben?

(f) Sei $A \in \mathbb{R}^{2\times 2}$. Was ist nochmal mit dem Ausdruck A^4 gemeint?

8.1 Ringe

Das Ziel von Kapitel 8 ist es, Polynome und Polynomfunktionen – denen Sie in den Übungen schon kurz begegnet sind – mehr im Detail zu betrachten. Dazu führen wir in diesem Abschnitt zunächste eine weitere algebraische Struktur ein, nämlich **Ringe**. Dieses Konzept ist etwas allgemeiner ist der Begriff des **Körpers**, den Sie bereits aus Kapitel 2 kennen.

Ringe und Ringhomomorphismen

Definition 8.1.1 (Ring)**.** Ein **Ring** ist ein Tupel $(R, +, \cdot)$, wobei R eine nicht-leere Menge ist, und $+$ und $\cdot$ Abbildungen sind mit den folgenden Eigenschaften:

(R0) *Binäre Verknüpfungen auf R:* Es gilt $+ : R^2 \to R$ und $\cdot : R^2 \to R$.

(R1) *Axiome der Addition:*
Es ist $(R, +)$ eine kommutative Gruppe.

Das neutrale Element dieser Gruppe bezeichnet man mit 0. Außerdem verwendet man für jedes $r \in R$ die Notation $-r$ um das inverse Elemente von r in der Gruppe $(R, +)$ zu bezeichnen.

(R2) *Assoziativität der Multiplikation:* Es ist $(R, \cdot)$ eine Halbgruppe mit neutralem Element.

Das neutrale Element bezeichnet man mit 1.

(R3) *Distributivgesetze:* Für alle $r, s, t \in R$ gilt

$$(r + s) \cdot t = r \cdot t + s \cdot t \qquad \text{und} \qquad t \cdot (r + s) = t \cdot r + t \cdot s.$$

Man nennt einen Ring $(R, +, \cdot)$ **kommutativ**, falls $\cdot$ kommutativ ist. [1]

Wie in Körpern vereinbart man auch in Ringen die Konvention **Punkt-vor-Strich** – d.h., das Zeichen $\cdot$ bindet stärker als das Zeichen $+$ –, und man lässt häufig das Zeichen $\cdot$ weg, d.h. man schreibt für $r, s \in R$ einfach rs anstelle von $r \cdot s$.

Ähnlich wie bei Körpern und Vektorräumen spricht man häufig kurz von einem „Ring R“ anstellen von einem „Ring $(R, +, \cdot)$“ – und denkt sich dann implizit dazu, dass die beiden Verknüpfungen wie üblich mit $+$ und $\cdot$ bezeichnet werden.

Genauso wie man zwischen Vektorräumen lineare Abbildungen studiert und man zwischen Gruppen Gruppenhomomorphismen untersucht, studiert man zwischen Ringen sogenannten **Ringhomomorphismen**:

Definition 8.1.2 (Ringhomomorphismus)**.** Seien R und S Ringe. Eine Abbildung $\varphi : R \to S$ heißt **Ringhomomorphismus**, wenn Sie die folgenden Eigenschaften erfüllt:

[1] D.h. falls $r \cdot s = s \cdot r$ für alle $r, s \in R$ gilt.

(RH0) Für alle $r_1, r_2 \in R$ gilt $\varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2)$.

(RH1) Für alle $r_1, r_2 \in R$ gilt $\varphi(r_1 r_2) = \varphi(r_1)\varphi(r_2)$.

(RH2) Es gilt $\varphi(1) = 1$.

Einige Beispiele

Lassen Sie uns einige Beispiele aufzählen:

Beispiele 8.1.3. (a) Jeder Körper ist eine kommutativer Ring.

(b) Es ist $(\mathbb{Z}, +, \cdot)$ ein kommutativer Ring.

(c) Es ist $(\mathbb{N}, +, \cdot)$ kein Ring, denn $(\mathbb{N}, +)$ ist keine Gruppe.

(d) Die Menge $2\mathbb{Z}$ aller geraden Zahlen, zusammen mit der üblichen Addition und Multiplikation, erfüllt alle Axiome eines Ringes mit Ausnahme einer Eigenschaft: Die Multiplikation besitzt kein neutrales Element.

(e) Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}^*$. Dann ist $\mathbb{K}^{n\times n}$ zusammen mit der komponentenweise Addition und der Matrixmultiplikation eine Ring.[2] Falls $n \geq 2$ ist, ist dieser Ring nicht kommutativ.

Zum Schluss dieses Abschnitts geben wir noch zwei etwas interessantere Beispiele an:

Beispiele 8.1.4. (a) Wir statten $\mathbb{Z}[\sqrt{2}] \coloneqq \{a + b\sqrt{2} \mid \ a, b \in \mathbb{Z}\} \subseteq \mathbb{R}$ mit der üblichen Addition + und Multiplikation $\cdot$ reeller Zahlen aus. Dann ist $(\mathbb{Z}[\sqrt{2}], +, \cdot)$ ein kommutativer Ring. Man kann alle Ringaxiome leicht nachprüfen; die einzige Eigenschaft, die vielleicht etwas überraschend ist, ist, dass die Multiplikation tatsächlich eine innere Verknüpfung auf $\mathbb{Z}[\sqrt{2}]$ ist. Das sieht man folgendermaßen: Seien $a_1 + b_1\sqrt{2}$ und $a_2 + b_2\sqrt{2}$ in $\mathbb{Z}[\sqrt{2}]$ (mit $a_1, b_1, a_2, b_2 \in \mathbb{Z}$). Dann gilt

$$(a_1 + b_1\sqrt{2})(a_2 + b_2\sqrt{2}) = a_1a_2 + 2b_1b_2 + (a_1b_2 + b_1a_2)\sqrt{2} \in \mathbb{Z}[\sqrt{2}].$$

Man kann sich überlegen, dass $(\mathbb{Z}[\sqrt{2}, +, \cdot)$ aber kein Körper ist. Das machen wir in den Übungen.

(b) Wir statten $\mathbb{Q}[\sqrt{2}] \coloneqq \{a + b\sqrt{2} \mid \ a, b \in \mathbb{Q}\} \subseteq \mathbb{R}$ mit der üblichen Addition + und Multiplikation $\cdot$ reeller Zahlen aus. Dann ist $(\mathbb{Q}[\sqrt{2}], +, \cdot)$ ebenfalls ein kommutativer Ring – und im Gegensatz zu $(\mathbb{Z}[\sqrt{2}], +, \cdot)$ sogar ein Körper.[3]

Ein Beispiel eines Ringhomomorphismus werden wir weiter unten sehen.

[2] Frage: Was ist das neutrale Element der Multiplikation in diesem Ring?

[3] Überlegen Sie sich bitte in Ruhe auf einem Blatt Papier, weshalb jedes Element in $\mathbb{Q}[\sqrt{2}] \setminus \{0\}$ ein multiplikativ inverses Element hat.

8.2 Polynome und Polynomfunktionen

Polynome und Multiplikation von Polynomen

Um Polynome einzuführen müssen wir zunächst kurz darüber reden, was man unter einer Folge versteht. Sei M eine Menge. Eine **Folge in** M ist eine Aufzählung von Elementen

$$(a_0, a_1, a_2, \dots)$$

von M – d.h. für jedes $n \in \mathbb{N}$ sei ein Elemente $a_n \in M$ gegeben. Häufig notiert man solche eine Folge in der Kurzform $(a_n)_{n\in\mathbb{N}}$.[4] Zwei Folgen in M heißen **gleich**, wenn an jeder Stelle ihre Komponenten gleich sind.

Folgen in Körpern, die nur endlich viele von Null verschiedene Glieder besitzen, kann man auf folgende Weise verknüpfen:

Diskussion 8.2.1 (Faltung endlicher Folgen)**.** Sei $\mathbb{K}$ ein Körper.

(a) Wir nennen eine Folge $(a_n)_{n\in\mathbb{N}}$ in $\mathbb{K}$ **endlich**, falls höchstens endlich viele der Komponenten a_n ungleich 0 sind – d.h., falls es ein $n_0 \in \mathbb{N}$ gibt mit der Eigenschaft $a_n = 0$ für alle $n \geq n_0$.

(b) Seien $a = (a_n)_{n\in\mathbb{N}}$ und $b = (b_n)_{n\in\mathbb{N}}$ zwei endliche Folgen in $\mathbb{K}$ und sei $\alpha \in \mathbb{K}$. Dann definiert man die Folgen $a + b$ und $\alpha \cdot a$ komponentenweise, d.h.

$$a + b = (a_n + b_n)_{n\in\mathbb{N}} \qquad \text{und} \qquad \alpha \cdot a = (\alpha a_n)_{n\in\mathbb{N}}.$$

Man kann dann nachrechnen, dass die Menge der endlichen Folgen in $\mathbb{K}$ mit den beiden Verknüpfungen $+$ und $\cdot$ ein Vektorraum über $\mathbb{K}$ ist.

(c) Seien wieder $a = (a_n)_{n\in\mathbb{N}}$ und $b = (b_n)_{n\in\mathbb{N}}$ zwei endliche Folgen in $\mathbb{K}$. Dann definieren wir ihre **Faltung** als die Folge

$$a \star b = (\sum_{k=0}^{n} a_k b_{n-k})_{n\in\mathbb{N}}$$

in $\mathbb{K}$. Man beachte, dass $a \star b$ selbst wieder eine endliche Folge ist.

Nun kann man durch nachprüfen der Axiome zeigen, dass die Menge aller endlichen Folgen in $\mathbb{K}$ zusammen mit den beiden Verknüpfungen $+$ und $\star$ ein kommutativer Ring ist. Das multiplikativ neutrale Element 1 ist hierbei die Folge

$$(1, 0, 0, \dots).$$

[4] Beachten Sie, dass eine Folge also im Grunde nichts weiter ist als ein Tupel – allerdings mit unendlich vielen Komponenten, die über $\mathbb{N}$ indiziert sind.

(d) Die Faltung von endlichen Folgen kann man etwas handlicher aufschreiben, wenn man die folgende Notation einführt: Die Folge

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

die an der Stelle 1 den Skalar 1 stehen hat und ansonsten nur aus dem Skalar 0 besteht, bezeichnen wir fortan mit dem Symbol X. Außerdem verwenden wir für endliche Folgen a in $\mathbb{K}$ und $n \in \mathbb{N}_0$ die Potenznotation $a^n = 1$ für $n = 0$ und $a^n = a \star \cdots \star a$ für $n \geq 1$, wobei a hier n-mal auftaucht. Es gilt natürlich $a^{m+n} = a^m \star a^n$ für alle $m, n \in \mathbb{N}$.

Mit dieser Notation gilt

$$X^n = (0, \dots, 0, 1, 0, 0, \dots),$$

wobei die 1 genau an der Stelle n steht.

Wenn nun $a = (a_n)_{n\in\mathbb{N}}$ und $b = (b_n)_{n\in\mathbb{N}}$ endliche Folgen in $\mathbb{K}$ sind, und $n_0, n_1 \in \mathbb{N}$ sind mit der Eigenschaft $a_n = 0$ für alle $n \geq n_0$ und $b_n = 0$ für alle $n \geq n_1$, dann gilt

$$a = \sum_{n=0}^{n_0} a_n X^n \qquad \text{und} \qquad b = \sum_{n=0}^{n_1} b_n X^n.$$

Damit kann man endliche Folgen in $\mathbb{K}$ also als Summe über X^n ausdrücken. Dies hat den großen Vorteil, dass man die Faltung von a und b nun ganz leicht ausdrücken kann, indem man die Summen ausmultipliziert[5] und die Rechenregel $X^{m+n} = X^m \star X^n$ für alle $n, m \in \mathbb{N}$ benutzt. Mit obenstehender Notation erhalten wir somit

$$a \star b = (\sum_{n=0}^{n_0} a_n X^n) \star (\sum_{n=0}^{n_1} b_n X^n) = \sum_{n=0}^{n_0+n_1} \Big(\sum_{k=0}^{n} a_k b_{n-k} \Big) X^n.$$

Im Grunde wissen wir natürlich schon, dass wir dieses Ergebnis erhalten müssen – denn genauso war ja die Faltung definiert. Allerdings hat diese Schreibweise den Vorteil, dass sie konkrete Rechnungen sehr erleichtert. Dies illustrieren wir im Folgenden anhand eines einfachen Beispiels.

Beispiel 8.2.2. Lassen Sie uns den Körper $\mathbb{R}$ betrachten und die beiden endlichen Folgen

$$a = (1, 2, 1, 0, 0, \dots) \qquad \text{und} \qquad b = (1, 1, 0, 0, \dots)$$

in $\mathbb{R}$. Wenn wir direkt die Definition der Faltung verwenden, erhalten wir für die Komponenten $(a \star b)_n$ die folgenden Werte:

$$(a \star b)_0 = \sum_{k=0}^{0} a_k b_{0-k} = a_0 b_0 = 1,$$

[5] Beachten Sie, dass Ausmultiplizieren tatsächlich möglich ist, weil in einem Ring das Distributivgesetz gilt!

$$
\begin{aligned}
(a \star b)_1 &= \sum_{k=0}^{1} a_k b_{1-k} = a_0 b_1 + a_1 b_0 = 3, \\
(a \star b)_2 &= \sum_{k=0}^{2} a_k b_{2-k} = a_0 b_1 + a_1 b_1 + a_2 b_0 = 3, \\
(a \star b)_3 &= \sum_{k=0}^{3} a_k b_{3-k} = a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0 = 1, \\
(a \star b)_4 &= \sum_{k=4}^{4} a_k b_{4-k} = a_0 b_4 + a_1 b_3 + a_2 b_2 + a_3 b_1 + a_4 b_0 = 0, \\
&\vdots \\
(a \star b)_n &= 0 \qquad \text{für alle weiteren } n.
\end{aligned}
$$

Also ist insgesamt

$$a \star b = (1, 3, 3, 1, 0, 0, \dots).$$

Dieselbe Rechnung kann man stattdessen aber auch übersichtlicher durchführen, indem man

$$a \star b = (1 + 2X + X^2) \star (1 + X) = 1 + 3X + 3X^3 + X^3 = (1, 3, 3, 1, \dots)$$

berechnet.

Definition 8.2.3 (Polynomringe)**.** Sei $\mathbb{K}$ ein Körper.

(a) Die Menge aller endlichen Folgen in $\mathbb{K}$ bezeichnen wir mit $\mathbb{K}[X]$; wir statten sie mit der Addition, der skalaren Multiplikation und der Faltung aus Diskussion 8.2.1 aus – wodurch die Menge zu einem Vektorraum über $\mathbb{K}$ und zu einem kommutativen Ring wird – und nennen Sie dann den **Polynomring über** $\mathbb{K}$.

(b) Die Elemente von $\mathbb{K}[X]$ bezeichnet man als **Polynome mit Koeffizienten aus** $\mathbb{K}$.[6] Man schreibt solch ein Element wie in Diskussion 8.2.1 in der Form

$$a_0 + a_1 X^1 + \dots a_n X^n$$

für geeignete Skalare $a_0, \dots, a_n \in \mathbb{K}$; diese Skalare heißen die **Koeffizienten** des Polynoms.[7]

Man kürzt Polynome häufig mit Symbolen wie $p(X)$ oder $q(X)$ ab.

[6] Oder manchmal auch kürzer als **Polynome über** $\mathbb{K}$.

[7] Genau genommen hat das Polynom natürlich unendlich viele Koeffizienten – aber die Koeffizienten für alle Indizes $> n$ sind 0. Beachten Sie, dass zwei Polynome genau dann gleich sind, wenn all ihre Koeffizienten gleich sind.

(c) Wenn man endliche Folgen in $\mathbb{K}$ in der gerade beschriebenen Weise notiert, dann bezeichnet man die Faltung üblicherweise einfach als **Multiplikation** und notiert sie mit $\cdot$ anstelle von $\star$.

(d) Sei $p(X) = \sum_{k=0}^{n} a_k X^k \in \mathbb{K}[X]$ mit $a_0, \ldots, a_n \in \mathbb{K}$. Falls mindestens einer der Koeffizienten $a_0, \ldots, a_n$ ungleich 0 ist, dann nennt man das größte $d \geq 0$ mit $a_d \neq 0$ den **Grad** von $p(X)$; man notiert ihn als $\deg(p(X))$.[8] Falls alle Koeffizienten des Polynoms $p(X)$ gleich Null sind, dann setzt man $\deg(p(X)) = -\infty$.

Wir zeigen kurz einige einfache Beispiele:

Beispiele 8.2.4. Lassen Sie uns den Körper $\mathbb{Q}$ betrachten.

(a) Das Polynom

$$(1 + X)(2 + X)$$

ist, wie man durch Ausmultiplizieren sehen kann, gleich dem Polynom

$$2 + 3X + X^2;$$

es hat somit die Koeffizienten $2, 3, 1$ und den Grad 2.

(b) Das sogenannte **Nullpolynom** 0 hingegen – welches das additiv neutrale Element in $\mathbb{Q}[X]$ ist – hat keine von Null verschiedenen Koeffizienten; es hat den Grad $-\infty$ (und es ist natürlich das einzige Polynom in $\mathbb{Q}[X]$ mit dieser Eigenschaft).

(c) Das Polynom 1 – welches das multiplikativ neutrale Element in $\mathbb{Q}[X]$ ist – hat den Koeffizienten 1 (an der Stelle 0), und ansonsnten nur Nullen als Koeffizienten. Es gilt $\deg(1) = 0$.[9]

Beachten Sie unbedingt, dass Polynome also keine Funktionen sind![10] Trotzdem kann man Polynome ineinander einsetzen:

Diskussion 8.2.5 (Einsetzen von Polynomen)**.** Sei $\mathbb{K}$ ein Körper und seien $p(X), q(X) \in \mathbb{K}[X]$. Wir schreiben $q(X)$ in der Form $q(X) = b_0 X^0 + \cdots + b_n X^n$ für ein $n \in \mathbb{N}$ und Skalare $b_0, \ldots, b_n \in \mathbb{K}$. Dann definiert man

$$q\big(p(X)\big) \coloneqq \sum_{k=0}^{n} b_k p(X)^k \in \mathbb{K}[X].$$

Man kann die folgenden Eigenschaften nachrechnen:[11]

[8] Die Abkürzung deg steht für das englische Wort „degree“.

[9] Beachten Sie, dass man 1 etwas ausführlicher in der Form $1X^0$ schreiben kann.

[10] Wir werden allerdings in Kürze noch sogenannte **Polynomfunktionen** einführen – diese sind tatsächlich Funktionen und hängen eng mit Polynomen zusammen.

[11] Was wir an dieser Stelle aber nicht im Detail ausführen.

(a) Das ineinander Einsetzen von Polynomen ist assoziativ.

(b) Für jedes $p(X) \in \mathbb{K}[X]$ ist die Abbildung

$$\begin{aligned} \mathbb{K}[X] &\to \mathbb{K}[X], \\ q(X) &\mapsto q\big(p(X)\big) \end{aligned}$$

ein Ringhomomorphismus.

Letztgenannte Eigenschaft demonstrieren wir noch anhand eines einfachen Beispiels:

Beispiel 8.2.6. Betrachen wir die Polynome

$$p(X) = X + 1, \qquad q(X) = X + 2, \qquad r(X) = X^2$$

in $\mathbb{R}[X]$. Laut Diskussion 8.2.5(b) gilt $(qr)(p(X)) = q(p(X))\, r(p(X))$.[12]

Lassen Sie uns dies der Anschauung halber in diesem Beispiel noch einmal konkret nachrechnen: Einerseits ist

$$(qr)(X) = q(X)r(X) = (X+2)X^2 = X^3 + 2X^2,$$

und somit

$$\begin{aligned} (qr)(p(X)) &= (X+1)^3 + 2(X+1)^2 \\ &= X^3 + 3X^2 + 3X + 1 + 2X^2 + 4X + 2 = X^3 + 5X^2 + 7X + 3. \end{aligned}$$

Andererseits ist

$$\begin{aligned} q(p(X)) + r(p(X)) &= \big((X+1)+2\big)\,\big((X+1)^2\big) \\ &= (X+3)(X^2+2X+1) = X^3 + 5X^2 + 7X + 3. \end{aligned}$$

Polynomfunktionen

Nachdem wir nun abstrakt geklärt haben, was ein Polynom ist, wollen wir als nächstes über den verwandten Begriff der **Polynomfunktion** sprechen.

Definition 8.2.7 (Polynomfunktionen)**.** Sei $\mathbb{K}$ ein Körper. Eine Funktion $f : \mathbb{K} \to \mathbb{K}$ heißt eine **Polynomfunktion**, falls es ein $n \in \mathbb{N}$ und Skalare $a_0, \ldots, a_n \in \mathbb{K}$ gibt mit der Eigenschaft

$$f(x) = \sum_{k=0}^{n} a_k x^k \qquad \text{für alle } x \in \mathbb{K}.$$

[12] Wobei wir $(qr)(X)$ als Abkürzung von $q(X)r(X)$ verwenden.

Während es sich bei einem Polynom also nicht um eine Funktion handelt, sondern lediglich um eine – speziell geschrieben – endliche Folge, ist eine Polynomfunktion tatsächlich eine Funktion.

Man kann allerdings jedem Polynom eine Polynomfunktion zuordnen. Dies besprechen wir in der folgenden Diskussion:

Diskussion 8.2.8. Sei $\mathbb{K}$ ein Körper.

(a) Die Menge $\mathrm{Abb}(\mathbb{K};\mathbb{K})$ ist mit der üblichen Addition und der durch

$$(f \cdot g)(x) = f(x)g(x) \qquad \text{für alle } x \in \mathbb{K}$$

für alle $f, g \in \mathrm{Abb}(\mathbb{K};\mathbb{K})$ gegebenen Multiplikation ein kommutativer Ring.[13]

(b) Für jedes Polynom $p(X) = \sum_{k=0}^{n} a_k X^k \in \mathbb{K}[X]$ (mit $n \in \mathbb{N}$ und $a_0, \ldots, a_n \in \mathbb{K}$) definieren wir eine Polynomfunktion $\hat{p} \in \mathrm{Abb}(\mathbb{K};\mathbb{K})$ durch

$$\hat{p}(x) = \sum_{k=0}^{n} a_k x^k \qquad \text{für alle } x \in \mathbb{K}.$$

Dann kann man direkt nachrechnen, dass die Abbildung

$$\begin{aligned} \varphi : \mathbb{K}[X] &\to \mathrm{Abb}(\mathbb{K};\mathbb{K}) \\ p(X) &\mapsto \hat{p} \end{aligned}$$

ein Ringhomomorphismus ist.

Außerdem ist φ verträglich mit der Hintereinanderausführung, d.h., für alle $p(X), q(X) \in \mathbb{K}[X]$ gilt

$$\varphi\big(q(p(X))\big) = \hat{q} \circ \hat{p}.$$

Bevor wir einige Eigenschaften der Abbildung $p(X) \mapsto \hat{p}(X)$ genauer studieren, wollen wir zunächst einige Beispiele betrachten.

Beispiele 8.2.9. (a) Betachten wir das Polynom

$$p(X) = X^2 + iX + 2 \in \mathbb{C}[X].$$

Die Funktion $\hat{p}$ ist dann die Abbildung $\mathbb{C} \to \mathbb{C}$, die durch

$$\hat{p}(x) = x^2 + ix + 2 \qquad \text{für alle } x \in \mathbb{C}$$

gegeben ist. Es gilt also beispielsweise

$$\hat{p}(0) = 2, \qquad \hat{p}(1) = 3 + i, \qquad \hat{p}(i) = 0.$$

[13] Was ist das multiplikativ neutrale Element ist in diesem Ring? Und warum handelt es sich nicht sogar um einen Körper?

(b) Lassen Sie uns nun die beiden Polynome

$$p(X) = X^2 + [1] \qquad \text{und} \qquad q(X) = X^3 + X^2 + [2]X + [1]$$

in $\mathbb{F}_3[X]$ betrachten. Die beiden Polynome haben verschiedene Koeffizienten und somit verschieden.

Betrachten wir nun die zugehörigen Polynomfunktionen $\hat{p}, \hat{q} : \mathbb{F}_3 \to \mathbb{F}_3$. Es erfüllen

$$\begin{aligned} & \hat{p}([0]) = [1], \qquad \hat{p}([1]) = [2], \qquad \hat{p}([2]) = [2], \\ \text{und} \quad & \hat{q}([0]) = [1], \qquad \hat{q}([1]) = [2], \qquad \hat{q}([2]) = [2]. \end{aligned}$$

Es gilt also $\hat{p} = \hat{q}$ – d.h. obwohl die beiden Polynome $p(X)$ und $q(X)$ verschieden sind, sind die zugehörigen Polynomfunktionen gleich. Oder anders ausgedrückt: Wenn die beiden Polynomfunktionen übereinstimmen, kann man trotzdem nicht schließen, dass die zugehörigen Koeffizienten bzw. die zugehörigen Polynome gleich sind.

Dies ist der Grund, weshalb man zwischen Polynomen und Polynomfunktionen unterscheidet.

In Theorem 8.2.17 werden wir charakterisieren, wann genau das soeben beobachtete Verhalten – das die Polynomfunktionen zweier verschiedener Polynome übereinstimmen – auftreten kann. Hierzu brauchen wir aber zunächst noch ein paar Begriffe.

Nullstellen und Faktorisierung

Definition 8.2.10 (Nullstelle)**.** Sei $\mathbb{K}$ ein Körper und $p(X) \in \mathbb{K}[X]$. Ein Element $x_0 \in \mathbb{K}$ heißt eine **Nullstelle** von $p(X)$, wenn $\hat{p}(x_0) = 0$ gilt.

Nullstellen von Polynomen kann man in folgender Weise nutzen um das Polynom zu faktorisieren:

Proposition 8.2.11. *Sei $\mathbb{K}$ ein Körper, $p(X) \in \mathbb{K}[X]$ und sei $x_0 \in \mathbb{K}$ eine Nullstelle von $p(X)$. Dann gibt es ein Polynom $q(X) \in \mathbb{K}[X]$ mit der Eigenschaft $p(X) = (X - x_0)q(X)$.*

Beweis. Lassen Sie uns das Polynom

$$r(X) \coloneqq p(X + x_0)$$

betrachten. Wir können es in der Form $r(X) = \sum_{k=0}^{n} a_k X^k$ für geeignetes $n \in \mathbb{N}$ und geeignete $a_0, \ldots, a_n \in \mathbb{K}$ schreiben.

Es gilt $\hat{r}(0) = \hat{p}(x_0) = 0$, und somit $a_0 = 0$. Also ist

$$r(X) = X \underbrace{\sum_{k=0}^{n-1} a_{k+1} X^k}_{=:s(X)},$$

und somit folgt $p(X) = r(X - x_0) = (X - x_0)s(X - x_0)$. Also folgt die Behauptung mit $q(X) \coloneqq s(X - x_0)$. □

Der Beweis liefert offenbar auch ein Verfahren um das Polynom $q(X)$ zu berechnen. Lassen Sie uns dies anhand eines Beispiels demonstrieren:

Beispiel 8.2.12. Sei

$$p(X) = X^3 + X^2 - 2 \in \mathbb{R}[X].$$

Es gilt $\hat{p}(1) = 0$. Lassen Sie uns nun das Verfahren aus dem Beweis von Proposition 8.2.11 verwenden um ein Polynom $q(X) \in \mathbb{R}[X]$ mit $p(X) = (X - 1)q(X)$ zu finden:

Wir definieren

$$\begin{aligned} r(X) \coloneqq p(X + 1) &= (X + 1)^3 + (X + 1)^2 - 2 \\ &= X^3 + 3X^2 + 3X + 1 + X^2 + 2X + 1 - 2 \\ &= X^3 + 4X^2 + 5X = X(X^2 + 4X + 5). \end{aligned}$$

Wir definieren also $s(X) = X^2 + 4X + 5$ und

$$\begin{aligned} q(X) \coloneqq s(X - 1) &= (X - 1)^2 + 4(X - 1) + 5 \\ &= X^2 - 2X + 1 + 4X - 4 + 5 = X^2 + 2X + 2. \end{aligned}$$

Laut des Beweises von Proposition 8.2.11 muss somit $p(X) = (X - 1)q(X)$ gelten.[14]

Bemerkung 8.2.13 (Polynomdivision)**.** Es gibt übrigens ein direkteres Verfahren um für ein Polynom $p(X)$ mit Nullstelle x_0 ein Polynom $q(X)$ mit der Eigenschaft $p(X) = (X - x_0)q(X)$ zu finden: Die sogenannte **Polynomdivision**.

Diese kennen Sie vielleicht schon aus der Schule. An dieser Stelle werden wir sie vorerst nicht weiter besprechen. Sie spielt aber in der Ringtheorie eine wichtige Rolle, weshalb sie vermutlich in Vorlesungen über Algebra (oder vielleicht auch in der Linearen Algebra 2) besprochen wird.

Lassen Sie uns für einen Körper $\mathbb{K}$ und ein Polynom $p(X) \in \mathbb{K}[X]$ betrachten, dass nicht das Nullpolynom ist und eine Nullstelle $x_0 \in \mathbb{K}$ besitzt. Aus Proposition 8.2.11 wissen wir, dass wir $p(X)$ in der Form $p(X) = (X - x_0)q(X)$ für ein Polynom $q(X)$ schreiben können. Durch Ausmultplizieren sehen wir außerdem, dass $\deg(p(X)) = \deg(q(X)) + 1$ gilt. Nun gibt es zwei Möglichkeiten: Entweder $q(X)$ hat keine Nullstelle in $\mathbb{K}$ – oder $q(X)$ hat eine Nullstelle, die wir dann wiederum absplitten können. Indem wir diesen Schritt endlich oft wiederholen, erhalten wir die folgende Aussage:

[14] Falls Sie skeptisch sind, können Sie das Ergebnis natürlich sofort nachprüfen, indem Sie einfach ausmultiplizieren.

Proposition 8.2.14. *Sei $\mathbb{K}$ ein Körper und sei $p(X) \in \mathbb{K}[X]$ ein Polynom mit Grad $d := \deg(p(X)) \in \mathbb{N}$ (also nicht das Nullpolynom). Dann gibt es ein $n \in \{0, \dots, d\}$, Skalare $x_1, \dots, x_n$ und ein Polynom $q(X) \in \mathbb{K}[X]$ vom Grad $\deg(q(X)) = d - n$ mit den folgenden Eigenschaften:*

(a) *Die Skalare $x_1, \dots, x_n$ sind genau die Nullstellen von $p(X)$.*[15]

(b) *Es gilt $p(X) = (X - x_1) \cdots (X - x_n) \cdot q(X)$.*

(c) *Das Polynom $q(X)$ hat keine Nullstelle in $\mathbb{K}$.*

Insbesondere ist die Anzahl der verschiedenen Nullstellen von $p(X)$ nicht größer als der Grad von $p(X)$.

Definition 8.2.15 (Vielfachheit von Nullstellen)**.** In der Situation von Proposition 8.2.14 verwenden wir die folgende Terminologie:

(a) Sei $x_0 \in \mathbb{K}$ eine Nullstelle von $p(X)$. Unter der **Vielfachheit** von x_0 versteht man die Anzahl, die angibt, wie häufig x_0 in der Liste $x_1, \dots, x_n$ vorkommt.[16]

(b) Man sagt, dass das Polynom $p(X)$ **über $\mathbb{K}$ in Linearfaktoren zerfällt**, wenn $\deg(q(X)) = 0$ gilt.

Wenn also $p(X)$ über $\mathbb{K}$ in Linearfaktoren zerfällt, dann ist $q(X)$ einfach nur eine (von Null verschiedene) Konstante. Es ist dann $n = d$ und $p(X) = (X - x_1) \cdots (X - x_d)q(X)$, also kann man durch Ausmultiplizieren sehen, dass die Konstante $q(X)$ gleich dem führenden Koeffizienten von $p(X)$ sein muss.[17]

Lassen Sie uns zur Illustration zwei Beispiele besprechen:

Beispiele 8.2.16. (a) Wir betrachten das Polynom $p(X) = X^3 + X \in \mathbb{R}[X]$. Es besitzt den Grad 3 und die Nullstelle $x_1 = 0$. Wir können es schreiben als

$$p(X) = X(X^2 + 1).$$

Das Polynom $X^2 + 1$ besitzt hingegen keine Nullstelle in $\mathbb{R}$ (warum übrigens nicht?), also sind wir mit $q(X) = X^2+1$ in der Situation von Proposition 8.2.14.

(b) Wir betrachten nochmals dasselbe Polynom, aber dieses mal über dem Körper $\mathbb{C}$, d.h. wir betrachten $p(X) = X^3 + X \in \mathbb{C}[X]$. Es gilt

$$p(X) = X(X^2) = X(X - i)(X + i),$$

d.h. wir sind in der Situation von Proposition 8.2.14 mit $q(X) = 1$, $n = 3$ und den Nullstellen $x_1 = 0$, $x_2 = i$ und $x_3 = -i$.

[15] Beachten Sie hierbei, dass manche der Skalare in der Liste $x_1, \dots, x_n$ auch mehrfach vorkommen können.

[16] Es handelt sich bei der Vielfachheit also um eine natürliche Zahl ungleich 0.

[17] Unter dem **führenden Koeffizienten** eines Polynoms versteht man den von 0 verschiedenen Koeffizienten mit dem größten Index.

Nun können wir wie angekündigt klären, wann ein Polynom durch seine zugehörige Polynomfunktion eindeutig bestimmt ist:

Theorem 8.2.17. *Sei $\mathbb{K}$ ein Körper. Der Ringhomomorphismus*

$$\begin{aligned}\varphi : \mathbb{K}[X] &\to \mathrm{Abb}(\mathbb{K};\mathbb{K})\\ p(X) &\mapsto \hat{p}\end{aligned}$$

aus Diskussion 8.2.8(b) *ist genau dann injektiv, wenn $\mathbb{K}$ unendlich viele Elemente hat.*

Beweis. „⇒“ Wir zeigen die Kontraposition der behaupteten Implikation: Besitze $\mathbb{K}$ also nur endlich viele Elemente $x_1, \dots, x_n$. Dann ist jedes Elemente von $\mathbb{K}$ eine Nullstelle des Polynoms

$$p(X) \coloneqq \prod_{k=1}^{n} (X - x_k),$$

also ist $\hat{p}$ die konstante Nullfunktion – d.h. $\varphi(p(X)) = \varphi(0)$. Allerdings ist p selbst nicht das Nullpolynom, denn es besitzt Grad n. Dies zeigt, dass φ nicht injektiv ist.

„⇐“ Es besitzt $\mathbb{K}$ unendlich viele Elemente, und es seien $p(X), q(X) \in \mathbb{K}[X]$ mit $\hat{p} = \hat{q}$. Für das Polynom $r(X) \coloneqq p(X) - q(X)$ gilt dann $\hat{r} = 0$.

Also verschwindet $\hat{r}$ an jedem Element aus $\mathbb{K}$, d.h. $r(X)$ besitzt unendlich viele Nullstellen. Laut Proposition 8.2.14 ist dies aber nur möglich, wenn $r(X)$ das Nullpolynom ist – also gilt $r(X) = 0$ und somit $p(X) = q(X)$.

Dies zeigt die Injektivität von φ. □

Zum Schluss vereinbaren wir noch die folgende notationelle Vereinfachung:

Vereinbarung 8.2.18. Sei $\mathbb{K}$ ein Körper, sei $p(X) = \sum_{k=0}^{n} a_k X^k \in \mathbb{K}[X]$ (mit $n \in \mathbb{N}$ und $a_0, \dots, a_n \in \mathbb{K}$) und sei $x_0 \in \mathbb{K}$. Dann verwenden wir anstelle der Notation $\hat{p}(x_0)$ auch einfach kurz die Notation $p(x_0)$ – d.h. wir benutzen die Notation

$$p(x_0) = \sum_{k=0}^{n} a_k x_0^k.$$

Im praktischen Umgang mit Polynomen unterscheidet man also oft nicht zwischen den Notationen $\hat{p}(x_0)$ und $p(x_0)$ – allerdings sollte man sich im Hinterkopf trotzdem stets bewusst sein, dass ein Polynom nicht das gleiche wie eine Polynomfunktion ist.

8.3 Der polynomielle Funktionalkalkül

Einsetzen von Matrizen in Polynome

Nun gehen wir noch einen Schritt weiter als zuvor: Wir setzen in Polynome nicht nur mehr Elemente des zugrunde liegenden Körpers ein, sondern sogar (quadratische Matrizen):

Zur Erinnerung: Für $k \in \mathbb{N}$ und $A \in \mathbb{K}^{n\times n}$ (wobei $\mathbb{K}$ ein Körper ist und $n \in \mathbb{N}^*$) verwenden wir für die Notation

$$A^k = \begin{cases} I_n & \text{falls } k = 0, \\ \underbrace{A \cdots A}_{n \text{ Faktoren}} & \text{falls } k \geq 1. \end{cases}$$

Damit können wir nun erklären, wie man eine quadratische Matrix in ein Polynom einsetzt:

Definition 8.3.1. Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und sei $A \in \mathbb{K}^{n\times n}$. Für jedes $p(X) = \sum_{k=0}^{m} a_k X^k \in \mathbb{K}[X]$ (mit $m \in \mathbb{N}$ und $a_0, \ldots, a_m \in \mathbb{K}$) definieren wir

$$p(A) \coloneqq \sum_{k=0}^{m} a_k A^k \in \mathbb{K}^{m\times m}.$$

Die Abbildung

$$\begin{aligned} \mathbb{K}[X] &\to \mathbb{K}^{n\times n}, \\ p(X) &\mapsto p(A) \end{aligned}$$

heißt der **polynomielle Funktionalkalkül** von A.

Die folgenden Eigenschaften des polynomiellen Funktionalkalküls kann man wieder direkt nachrechnen.

Proposition 8.3.2. *Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und sei $A \in \mathbb{K}^{n\times n}$.*

(a) *Der polynomielle Funktionalkalkül*

$$\begin{aligned} \mathbb{K}[X] &\to \mathbb{K}^{n\times n}, \\ p(X) &\mapsto p(A) \end{aligned}$$

von A ist ein Ringhomomorphismus.

(b) *Der polynomielle Funktionalkalkül ist verträglich mit dem ineinander Einsetzen von Polynomen, d.h. für $p(X), q(X) \in \mathbb{K}[X]$ liefert einsetzen von A in $q(p(X))$ dieselbe Matrix wie Einsetzen von $p(A)$ in $q(X)$.*[18]

Das charakteristische Polynom

Nun kommen wir zu einem äußerst wichtigen Zusammenhang zwischen Polynomen und Matrizen:

Definition 8.3.3 (Charakteristisches Polynom)**.** Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und sei $A \in \mathbb{K}^{n\times n}$. Das **charakteristische Polynom** $p_A(X) \in \mathbb{K}[X]$ von A ist definiert als

$$p_A(X) = \det(X \cdot I_n - A).$$

[18] Und diese bezeichnet man natürlich mit $q(p(A))$.

Genau genommen muss man sich fragen, was mit dem Ausdruck $\det(X \cdot I_n - A)$ eigentlich gemeint ist – denn X ist ja genau genommen gar kein Element von $\mathbb{K}$ (auch wenn wir häufig Elemente von $\mathbb{K}$ für X einsetzen). Wenn man ganz präzise sein möchte, kann man diesen Ausdruck einfach mit Hilfe der Leibnizformel erklären als

$$p_A(X) = \det(X \cdot I_n - A) = \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} (X \cdot I_n - A)_{\sigma(j),j},$$

wobei

$$(X \cdot I_n - A)_{h,j} = \begin{cases} -A_{h,j} & \text{falls } h \neq j, \\ X - A_{h,j} & \text{falls } h = j \end{cases}$$

für $h, j \in \{1, \ldots, n\}$ ist. Anhand dieser Formel sieht man auch, dass das charakteristische Polynom den Grad n besitzt.

Neben der Leibnizformel kann man zur Berechnung des charakteristischen Polynoms auch den Entwicklungssatz für Determinanten verwenden. Was hingegen im Allgemeinen nicht gut funktioniert, ist die Verwendung des Gauß-Algorithmus – denn dadurch, dass nun die Variable X in der Matrix vorkommt, führt die Division, die manchmal im Gauß-Algorithmus nötig ist, zu äußerst komplizierten Rechnungen und Fallunterscheidungen.

Lassen Sie uns nun zur Veranschaulichung einige Beispiele betrachten:

Beispiele 8.3.4. (a) Betrachten wir die Matrix

$$A = \pi \in \mathbb{R}^{1\times 1}.$$

Dann gilt

$$p_A(X) = \det(X - \pi) = X - \pi.$$

(b) Betrachten wir die Matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{Q}^{2\times 2}.$$

Dann ist

$$p_A(X) = \det(XI_2 - A) = \det \begin{bmatrix} X & 1 \\ -1 & X \end{bmatrix} = X^2 + 1.$$

(c) Betrachten wir zuletzt noch die Matrix

$$A = \begin{bmatrix} [1] & [0] & [2] \\ [0] & [0] & [0] \\ [2] & [1] & [0] \end{bmatrix} \in \mathbb{F}_3^{3\times 3}.$$

Dann ist

$$p_A(X) = \det \begin{bmatrix} X+[2] & [0] & [1] \\ [0] & X & [0] \\ [1] & [2] & X \end{bmatrix} = X \det \begin{bmatrix} X+[2] & [1] \\ [1] & X \end{bmatrix}$$
$$= X\Big((X+[2])X - [1]\Big) = X^3 + [2]X^2 + [2]X.$$

Der Satz von Cayley–Hamilton

Das charakteristische Polynom einer Matrix hat eine ganz spezielle Eigenschaft, die wir im folgenden Theorem beschreiben wollen. Als Vorbereitung und Motivation bemerken wir zunächst:

Bemerkung 8.3.5. Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$. Dann gibt es ein von Null verschiedenes Polynom $p(X) \in \mathbb{K}[X]$ mit der Eigenschaft $p(A) = 0$.

Um dies zu sehen, beachte man zunächst, dass der Vektorraum $\mathbb{K}^{n\times n}$ über $\mathbb{K}$ die Dimension n^2 hat. Somit ist jedes Tupel bestehend aus n^2+1 Vektoren in $\mathbb{K}^{n\times n}$ automatisch linear abhängig. Inbesondere ist also

$$(A^0, A^1, \ldots, A^{n^2})$$

linear abhängig, d.h. es gibt Skalare $a_0, \ldots, a_{n^2} \in \mathbb{K}$, von denen mindestes einer ungleich 0 ist, derart, dass

$$\sum_{k=0}^{n^2} a_k A^k = 0$$

gilt. Also ist $p(X) \coloneqq \sum_{k=0}^{n^2} a_k X^k \in \mathbb{K}[X]$ ein von Null verschiedenes Polynom mit der Eigenschaft $p(A) = 0$.

Das soeben ausgeführte Argument zeigt zudem, dass man $p(X)$ so wählen kann, dass $\deg p(X) \leq n^2$ gilt. Es gilt aber noch eine viel stärkere Aussage – für deren Beweis wir allerdings auf die Vorlesung *Lineare Algebra 2* verweisen.

Theorem 8.3.6 (Cayley–Hamilton)**.** *Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$. Dann gilt $p_A(A) = 0$.*[19]

Anwendung: Lineare Rekurrenzen

Zum Abschluss dieses Abschnitts wollen wir eine Anwendung des Satzes von Cayley–Hamilton zeigen: Die Lösung **linearen Rekurrenzen**. In Beispiel 3.2.5(b) hatten wir bereits ein Formel für die Fibonacci-Zahlen angegeben. Diese Formel kann man per Rekursion beweisen, aber es ist damit nicht klar, wie man überhaupt zu dieser

[19] D.h., wenn wir eine Matrix in ihr eigens charakteristisches Polynom einsetzen, erhalten wir die Nullmatrix.

Formel gelangt. Im folgenden zeigen wir – in einem allgemeineren Rahmen – einen Weg hierzu auf.

Seien $\alpha, \beta \in \mathbb{R}$ mit $\beta \neq 0$.[20] Wir betrachten eine Folge $(x_n)_{n\in\mathbb{N}}$, welche die Rekursionsvorschrift

$$x_{n+1} = \alpha x_n + \beta x_{n-1} \tag{8.3.1}$$

für alle $n \in \mathbb{N}^*$ erfüllt.[21] Um die Lösung der Gleichung (8.3.1) (in Abhängigkeit von x_0 und x_1) zu finden, gehen wir in fünf Schritten vor:

Schritt 1: Zunächst schreiben wir die Gleichung (8.3.1) um in eine Matrixgleichung: Dazu definieren wir für jedes $n \in \mathbb{N}$ einen Vektor

$$y_n \coloneqq \begin{bmatrix} x_n \\ x_{n+1} \end{bmatrix} \in \mathbb{C}^2.$$

Dann gilt für jedes $n \in \mathbb{N}$ die Gleichung

$$y_{n+1} = \begin{bmatrix} x_{n+1} \\ x_{n+2} \end{bmatrix} = \begin{bmatrix} x_{n+1} \\ \alpha x_{n+1} + \beta x_n \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ \beta & \alpha \end{bmatrix}}_{=:A} \begin{bmatrix} x_n \\ x_{n+1} \end{bmatrix} = Ay_n.$$

Schritt 2: Aus $y_{n+1} = Ay_n$ für alle $n \in \mathbb{N}$ folgt sofort $y_n = A^n y_0$ für alle $n \in \mathbb{N}$. Also müssen wir nun eine Methode finden, um die Potenzen A^n für alle $n \in \mathbb{N}$ zu berechnen.

Schritt 3: Sei $n \in \mathbb{N}$ fest. Um A^n zu berechnen, verwenden wir nun den Satz von Cayley–Hamilton. Hierzu benötigen wir folgende Proposition über Polynome, die sich zum Beispiel mit Hilfe von Polynomdivision zeigen lässt (vgl. Proposition 8.2.13).

Proposition 8.3.7. *Sei $\mathbb{K}$ ein Körper und seien $q(X), p(X) \in \mathbb{K}[X]$ mit $p(X) \neq 0$. Dann gibt es ein eindeutig bestimmte Polynome $r(X), f(X) \in \mathbb{K}[X]$ derart, dass*

$$\deg r(X) < \deg p(X) \qquad \textit{und} \qquad q(X) = p(X)f(X) + r(X)$$

gilt.

Wir wenden diese Proposition nun auf die Polynome $q_n(X) \coloneqq X^n$ und auf $p(X) \coloneqq p_A(X)$ an. Es ist

$$p(X) = p_A(X) = \det(XI_2 - A) = \det \begin{bmatrix} X & -1 \\ -\beta & X - \alpha \end{bmatrix} = X^2 - \alpha X - \beta.$$

[20] Das folgende Verfahren funktioniert übrigens auch über $\mathbb{C}$, und sogar über allgemeineren Körpern. Hierzu benötigt man aber mehr wissen über Nullstellen von Polynomen, und insbesondere über den sogenannten **algebraischen Abschluss** eines Körper – dies wird üblicherweise in Algebravorlesungen besprochen.

[21] Beachten Sie: Um alle Werte x_n eindeutig festzulegen, muss man natürlich – genau wie im Spezialfall der Fibonacci-Zahlen – noch die Werte x_0 und x_1 vorgeben.

Also gibt es laut der Proposition Polynome $f_n(X)$ und $r_n(X)$ mit $\deg r_n(X) \leq 1$ derart, dass

$$X^n = p_A(X)f_n(X) + r_n(X). \tag{8.3.2}$$

Aufgrund des Satzes von Caley–Hamilton gilt $p_A(A) = 0$, und somit folgt

$$A^n = 0 \cdot f_n(A) + r_n(A) = r_n(A).$$

Weil $r_n(X)$ höchstens Grad 1 hat, gibt es Koeffizienten $b_n, c_n \in \mathbb{C}$ mit $r_n(X) = b_nX + c_n$. Damit ist also

$$A^n = b_nA + c_nI_2.$$

Mit Hilfe des Satzes von Caley–Hamilton ist es uns also gelungen, die Potenz A^n als eine Linearkombination von A und I_2 darzustellen. Wir müssen nun lediglich die Koeffizienten b_n und c_n berechnen.

Schritt 4: Nach wie vor sei $n \in \mathbb{N}$ fest. Wir wollen die Koeffizienten b_n, c_n aus dem vorangehenden Schritt berechnen. Hierzu benötigen wir ein weiteres Resultat – den sogenannten **Fundamentalsatz der Algebra**:

Theorem 8.3.8 (Fundamentalsatz der Algebra)**.** *Jedes Polynom $p(X) \in \mathbb{C}[X]$ zerfällt über $\mathbb{C}$ in Linearfaktoren.*[22]

Obwohl des Resultat Fundamentalsatz der *Algebra* heißt, handelt es sich in gewissem Sinne um ein analytisches Resultat: Für den Beweis benötigt man einerseits, wie die komplexen Zahlen mit den reellen Zahlen zusammenhängen, und andererseits bestimmte Eigenschaften der reellen Zahlen. Genauer: Eine wesentliche Zutat des Beweises ist die Erkenntnis, dass jedes Polynom in $\mathbb{R}[X]$ von ungeradem Grad mindestens eine Nullstelle in $\mathbb{R}$ hat – dies wiederum ist anschaulich überhaupt nicht überraschend, wenn man sich das Verhalten der zugehörigen Polynomfunktion für $x \to \pm\infty$ ansieht, aber um den Beweis präzise zu führen benötigt man eine Eigenschaft der reellen Zahlen (die sogenannte **Ordnungsvollständigkeit**), die zum Bereich der Analysis und zur Konstruktion der reellen Zahlen gehört. Deshalb gehen wir an dieser Stelle nicht weiter darauf ein.[23]

Aufgrund des Fundamentalsatzes der Algebra gibt es also zwei komplexe Zahlen $\lambda_0, \lambda_1 \in \mathbb{C}$ mit der Eigenschaft $p_A(X) = (X - \lambda_0)(X - \lambda_1)$, und eine kurze Rechnung liefert somit

$$(X - \lambda_0)(X - \lambda_1) = p_A(X) = X^2 - \alpha X - \beta = (X - \frac{\alpha}{2})^2 - \frac{\alpha^2}{4} - \beta.$$

Es folgt $\lambda_0\lambda_1 = \beta$, und wegen der Voraussetzung $\beta \neq 0$ sind somit die beiden Nullstellen λ_0, λ_1 beide ungleich 0.

Nun unterscheiden wir zwei Fälle:

[22] Man sagt hierzu manchmal auch, der Körper $\mathbb{C}$ ist **algebraisch abgeschlossen**.

[23] Übrigens gibt es auch noch einen anderen Beweis des Fundamentalsatzes der Algebra, welcher die sogenannten **Funktionentheorie** benutzt – dabei handelt es sich ebenfalls um eine analytische Theorie.

- *1. Fall:* Es ist $\lambda_0 \neq \lambda_1$.

  Dann erhalten wir, indem wir λ_0 und λ_1 in die Gleichung (8.3.2) einsetzen, zwei Gleichungen für die Koeffizienten b_n und c_n, nämlich

  $$\lambda_0^n = b_n\lambda_0 + c_n,$$
  $$\lambda_1^n = b_n\lambda_1 + c_n.$$

  Dies können wir umschreiben in die Matrix-Vektor-Gleichung

  $$\underbrace{\begin{bmatrix}\lambda_0 & 1\\ \lambda_1 & 1\end{bmatrix}}_{=:\Lambda}\begin{bmatrix}b_n\\ c_n\end{bmatrix} = \begin{bmatrix}\lambda_0^n\\ \lambda_1^n\end{bmatrix}.$$

  Die Matrix $\Lambda \in \mathbb{C}^{2\times 2}$ hat Determinante $\det \Lambda\lambda_0 - \lambda_1 \neq 0$ und ist somit invertierbar; ihre inverse Matrix ist gleich[24]

  $$\Lambda^{-1} = \frac{1}{\lambda_0 - \lambda_1}\begin{bmatrix}1 & -1\\ -\lambda_1 & \lambda_0\end{bmatrix}.$$

  Somit erhalten wir insgesamt

  $$\begin{bmatrix}b_n\\ c_n\end{bmatrix} = \Lambda^{-1}\begin{bmatrix}\lambda_0^n\\ \lambda_1^n\end{bmatrix} = \begin{bmatrix}\frac{\lambda_0^n - \lambda_1^n}{\lambda_0 - \lambda_1}\\ \\ \frac{\lambda_0\lambda_1^n - \lambda_1\lambda_0^n}{\lambda_0 - \lambda_1}\end{bmatrix}.$$

- *2. Fall:* Es ist $\lambda_0 = \lambda_1$.

  In diesem Fall setzen wir zunächst wieder λ_0 und λ_1 in die Gleichung (8.3.2) ein, und erhalten somit die Gleichung

  $$\lambda_0^n = b_n\lambda_0 + c_n.$$

  Auch die Zahl λ_1 in die Gleichung einzusetzen, liefert allerdings wegen $\lambda_1 = \lambda_0$ lediglich dieselbe Gleichung und somit keine neue Information.

  Stattdessen können wir nun aber folgendermaßen vorgehen: In dem wir komplexe Zahlen $x \in \mathbb{C}$ in die Gleichung (8.3.2) einsetzen, erhalten wir für alle $x \in \mathbb{C}$

  $$x^n = p_A(x)f(x) + b_n x + c_n.$$

  Nun borgen wir uns eine Technik aus der Analysis: Wir leiten beide Seiten der Gleichung nach x ab[25] und erhalten somit

  $$nx^{n-1} = p_A'(x)f(x) + p_A(x)f'(x) + b_n.$$

[24] Das kann man übrigens ganz leicht sehen – wir werden in den Übungen noch eine sehr einfache Formel zur Invertierung von 2×2-Matrizen kennenlernen.

[25] Wenn Sie genau sind, sollten Sie an dieser Stelle natürlich nachfragen, ob das für komplexen Zahlen tatsächlich genauso funktioniert, wie für reelle. Zumindest für Polynomfunktionen (übrigens auch noch für eine allgemeinere Klasse von Funktionen – den sogenannten **holomorphen Funktionen**) ist die tatsächlich der Fall – die Details hierzu lernt man üblicherweise in einer Vorlesung über **Funktionentheorie** (und die Grundlagen hierzu manchmal auch schon in einer Analysis-Vorlesung).

Sowohl $p_A(x) = (x - \lambda_0)^2$ als auch $p'_A(x) = 2(x - \lambda_0)$ verschwindet, wenn wir für x die Zahl λ_0 einsetzen – und somit folgt

$$n\lambda_0^{n-1} = b_n.$$

Für b_n erhalten wir somit die Formel

$$c_n = \lambda_0^n - b_n\lambda_0 = -(n-1)\lambda_0^n.$$

Insgesamt ist in diesem Fall also

$$\begin{bmatrix} b_n \\ c_n \end{bmatrix} = \begin{bmatrix} n\lambda_0^{n-1} \\ -(n-1)\lambda_0^n \end{bmatrix}.$$

Schritt 5: Wir fassen zusammen: Es gilt für jedes $n \in \mathbb{N}$ wegen Schritt 3

$$A^n = b_n A + c_n I_2 = \begin{bmatrix} c_n & b_n \\ \beta b_n & \alpha b_n + c_n \end{bmatrix}.$$

Wegen Schritt 2 ist somit für jedes $n \in \mathbb{N}$

$$x_n = (y_n)_1 (A^n y_0)_1 = \Big(\begin{bmatrix} c_n & b_n \\ \beta b_n & \alpha c_n \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \Big)_1 = c_n x_0 + b_n x_1,$$

wobei die Zahlen b_n, c_n durch die Formeln in Schritt 4 gegeben sind. Somit erhalten wir also insgesamt:

Proposition 8.3.9. *Seien $x_0, x_1 \in \mathbb{C}$, seien $\alpha, \beta \in \mathbb{C}$ mit $\beta \neq 0$.*[26] *Außerdem seien λ_0, λ_1 die beiden*[27] *Nullstellen des Polynoms $X^2 - \alpha X - \beta \in \mathbb{C}[X]$. Die Folge $(x_n)_{n\in\mathbb{N}}$, die durch die Rekursionsbedingung*

$$x_{n+1} = \alpha x_n + \beta x_{n-1} \qquad \text{für alle } n \in \mathbb{N}^*$$

definiert ist, ist durch folgende explizite Formel gegeben:[28]

(a) *Falls $\lambda_0 \neq \lambda_1$ ist, gilt für alle $n \in \mathbb{N}$*

$$x_n = \frac{\lambda_0\lambda_1^n - \lambda_1\lambda_0^n}{\lambda_0 - \lambda_1} x_0 + \frac{\lambda_0^n - \lambda_1^n}{\lambda_0 - \lambda_1} x_1.$$

(b) *Falls $\lambda_0 = \lambda_1$ ist, gilt für alle $n \in \mathbb{N}$*

$$x_n = -(n-1)\lambda_0^n x_0 + n\lambda_0^{n-1} x_1.$$

[26] Was passiert übrigens im Fall $\beta = 0$?

[27] Mit Vielfachheit gezählt, d.h. sie können auch gleich sein

[28] Es ist übrigens sehr lehrreich, noch einmal nachzurechnen, warum die folgenden Formeln für $n = 0$ und $n = 1$ tatsächlich wieder die Anfangswerte x_0 bzw. x_1 liefern.

Lassen Sie uns als Beispiel noch einmal auf die Fibonacci-Zahlen zurückkommen:

Beispiel 8.3.10. In der Notation von Proposition 8.3.9 sei $x_0 = x_1 = 1$ und $\alpha = \beta = 1$. Dann sind die Zahlen x_n gleich den Fibonacci-Zahlen f_n, die wir in Beispiel 3.2.2 eingeführt hatten. Die beiden Nullstellen λ_0 und λ_1 des Polynomes $X - X - 1$ kann man leicht ausrechnen,[29] sie lauten

$$\lambda_0 = \frac{1-\sqrt{5}}{2} \qquad \text{und} \qquad \lambda_1 = \frac{1+\sqrt{5}}{2}.$$

Somit sind wir in Fall (a) von Proposition 8.3.9, d.h. für alle $n \in \mathbb{N}$ gilt die Formel

$$\begin{aligned} x_n &= \frac{\lambda_0\lambda_1^n - \lambda_1\lambda_0^n}{\lambda_0 - \lambda_1} + \frac{\lambda_0^n - \lambda_1^n}{\lambda_0 - \lambda_1} \\ &= \frac{\lambda_0^n(\lambda_1 - 1) + \lambda_1^n(1 - \lambda_0)}{\lambda_1 - \lambda_0} = \frac{\lambda_1^{n+1} - \lambda_0^{n+1}}{\sqrt{5}}, \end{aligned}$$

wobei wir im letzten Schritt die – leicht nachprüfbaren – Gleichungen $\lambda_1 - \lambda_0 = \sqrt{5}$ sowie $\lambda_1 - 1 = -\lambda_0$ und $1 - \lambda_0 = \lambda_1$ verwendet haben. Dies ist genau die Formel, die wir in Beispiel 3.2.5(b) angegeben hatten (und die in Bonusaufgabe 4(c) auf Hausaufgabenblatt 5 bereits per Induktion bewiesen wurde).

Literaturhinweise

- Polynome und Ringe spielen eine äußerst zentrale Rolle in weiterführenden Themen der Algebra. Einen (durchaus anspruchsvollen) Einstieg in die Algebra bietet zum Beispiel das klassische Lehrbuch von Bosch [Bos20].
- Für einen Zugang zur Linearen Algebra, der sich stark auf Polynome stützt verweisen wir auf das Buch [Fuh12].

[29] Zum Beispiel mit quadratischer Ergänzung, oder mit der Lösungsformel für quadratische Gleichungen.

Kapitel 9

Eigenwerte und Eigenvektoren – Eine Einführung

Einstiegsfragen. (a) Können Sie

$$\begin{pmatrix} 0 & 1 \\ -3 & 4 \end{pmatrix}^{2022}$$

explizit berechnen? Können Sie erläutern, was das mit dem Satz von Cayley–Hamilton zu tun hat? Kennen Sie sonst noch eine Möglichkeit zur Berechnung solcher Matrixpotenzen?

(b) Denken Sie an Abschnitt 5.6 zurück: Wann noch gleich heißt eine Matrix **diagonalisierbar**? Warum sind diagonalisierbare Matrizen interessant?

(c) Wie kann man herausfinden, ob eine gegebene Matrix diagonalisierbar ist?

9.1 Grundbegriffe

Eigenwerte und Eigenvektoren

Der folgende Begriff (und Verallgemeinerungen dieses Begriffs, die Sie womöglich in Veranstaltungen später im Studium lernen werden) ist zentral für verschiedenste mathematische Anwendungen.[1]

Definition 9.1.1 (Eigenwerte und Eigenvektoren)**.** Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$. Man nennt einen Skalar $\lambda \in \mathbb{K}$ einen **Eigenwert** von A, falls es ein $x \in \mathbb{K}^n \setminus \{0\}$ mit der Eigenschaft $Ax = \lambda x$ gibt.

[1] Zum Beispiel, um nur ein paar zu nennen, für die sogenannte **Hauptachsentransformation** in der Geometrie (Lineare Algebra 2), die **Hauptkomponentenanalyse** in Statistik und Data Science, die Analyse von **Markovketten** in der Stochastik, die Untersuchung von **Extremwertaufgaben** in der Analysis, für die **Fouriertransformation** in der Signalanalyse, für die **Quantenmechanik** in der Physik (und somit inbesondere auch für **Quanteninformationstheorie**), sowie für die Analyse von sogenannten **partiellen Differentialgleichungen**.

In diesem Fall heißt jedes solche x ein **Eigenvektor** von A zum Eigenwert λ.

Ob ein gegebener Skalar ein Eigenwert ist, kann man folgendermaßen charakterisieren:

Theorem 9.1.2. *Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$; sei $\lambda \in \mathbb{K}$. Dann sind die folgenden Aussagen äquivalent:*

(i) *Es ist λ ein Eigenwert von A.*

(ii) *Der Untervektorraum $\ker(\lambda I_n - A)$ von $\mathbb{K}^n$ besteht nicht nur aus dem Nullvektor.*

(iii) *Die Matrix $\lambda I_n - A \in \mathbb{K}^{n\times n}$ ist nicht invertierbar.*

(iv) *Es ist λ eine Nullstelle des charakteristischen Polynoms $p_A(X)$ von A.*

Beweis. „(i) $\Leftrightarrow$ (ii)“ Aus der Mengengleichheit

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

folgt sofort die behauptete Äquivalenz.

„(ii) $\Leftrightarrow$ (iii)“ Indem wir den Rangsatz auf die Abbildung

$$\begin{aligned} \mathbb{K}^n &\to \mathbb{K}^n, \\ x &\mapsto (\lambda I_n - A)x \end{aligned}$$

anwenden, sehen wir, dass $\ker(\lambda I_n - A) = \{0\}$ äquivalent zu $\operatorname{rang}(A) = n$ und somit zur Invertierbarkeit von A ist.

„(iii) $\Leftrightarrow$ (iv)“ Diese Äquivalenz folgt sofort aus Korollar 7.2.11. □

Aus dem vorangehenden Theorem erhalten wir die folgende einfache und nützliche Konsequenz:

Korollar 9.1.3. *Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$. Dann besitzen A und A^{T} diesselben Eigenwerte.*

Beweis. Wegen Korollar 7.2.9 haben A und A^{T} dasselbe charakteristische Polynom, und somit laut Theorem 9.1.2(i) und (iv) auch diesselben Eigenwerte. □

Lassen Sie uns anhand eines Beispiels besprechen, wie man Korollar 9.1.3 und Theorem 9.1.2 verwenden kann, um zu zeigen, dass eine gegebene Zahl ein Eigenwert einer gegebenen Matrix ist:

Beispiel 9.1.4. Lassen Sie uns die Matrix

$$A = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

betrachten. Es handelt sich um eine sogenannte **spaltenstoachstische Matrix** – so nennt man Matrizen in $\mathbb{R}^{n\times n}$ (für $n \in \mathbb{N}^*$), deren Einträge alle ≥ 0 sind und deren Spalten sich jeweils zu 1 summieren. Lassen Sie uns zeigen, dass 1 ein Eigenwert von A ist. Wir geben mehrere Möglichkeiten an, wie man das beweisen kann:

- *1. Möglichkeit:* Sei $\mathbb{1} \in \mathbb{R}^3$ derjenige Vektor, dessen Einträge alle gleich 1 sind. Weil jede Spalte von A sich zu 1 summiert, gilt $\mathbb{1}^{\mathrm{T}} A = \mathbb{1}^{\mathrm{T}}$, und somit $A^{\mathrm{T}} \mathbb{1} = \mathbb{1}$. Also ist 1 ein Eigenwert von A^{T} und somit laut Korollar 9.1.3 auch von Eigenwert von A.

- *2. Möglichkeit:* Man kann die Determinante von $1 \cdot I_3 - A$ direkt ausrechnen und erhält dabei, dass diese 0 ist. Also folgt aus Theorem 9.1.2(i) und (iv), dass 1 ein Eigenwert von A ist.

- *3. Möglichkeit:* Man kann die Matrix $1 \cdot I_3 - A$ auf eine Zeilstufenform bringen und sieht dann, dass sie nicht invertierbar ist. Aus Theorem 9.1.2(i) und (iii) folgt somit, dass 1 ein Eigenwert von A ist.

Als nächstes führen wir einige weitere Begriffe ein, die für ein tieferes Verständnis von Eigenwerten wichtig sind.

Definition 9.1.5 (Eigenräume und Vielfachheiten)**.** Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$; sei $\lambda \in \mathbb{K}$ ein Eigenwert von A.

(a) Der Untervektorraum $\ker(\lambda I_n - A)$ heißt der **Eigenraum** von A zum Eigenwert λ.[2]

(b) Die Dimension $\dim \ker(\lambda I_n - A)$ nennt man die **geometrischen Vielfachheit** des Eigenwerts λ.

(c) Die Vielfachheit von λ als Nullstelle des charakteristischen Polynoms $p_A(X)$ nennt man die **algebraische Vielfachheit** des Eigenwerts λ.[3]

Wenn also λ ein Eigenwert von A ist, dann besteht besteht der Eigenraum $\ker(\lambda I_n - A)$ genau aus den Eigenvektoren von A und aus dem Nullvektor[4].

Der Fundamentalsatz der Algebra, Theorem 8.3.8, besagt, dass jedes Polynom mit komplexen Koeffizienten über $\mathbb{C}$ in Linaerfaktoren zerfällt. Hieraus erhalten wir das folgende Korollar für Matrizen mit komplexen Einträgen:

Korollar 9.1.6. *Sei $n \in \mathbb{N}^*$ und sei $A \in \mathbb{C}^{n\times n}$.*

(a) *Die Matrix A besitzt einen Eigenwert.*

[2] Beachten Sie, dass dieser Untervektorraum laut Theorem 9.1.2(i) und (ii) mindestens die Dimension 1 hat.

[3] Beachten Sie, dass λ laut Theorem 9.1.2(i) und (iv) tatsächliche eine Nullstelle von $p_A(X)$ ist.

[4] Welcher definitionsgemäß kein Eigenvektor von A ist.

(b) *Es gilt sogar noch mehr: Wenn man die Eigenwert mit ihrer algebraischen Vielfachheit zählt, dann hat A genau n Eigenwerte $\lambda_1, \dots, \lambda_n$, und es gilt* $\det(A) = \lambda_1 \cdots \lambda_n$.

Beweis. (a) Dies folgt sofort aus dem Fundamentalsatz der Algebra, Theorem 8.3.8.

(b) Da das charakterisitische Polynom $p_A(X)$ den Grad n hat, gibt es laut Fundamentalsatz der Algebra komplexe Zahlen $\lambda_1, \dots, \lambda_n$ mit der Eigenschaft

$$p_A(X) = (X - \lambda_1) \cdots (X - \lambda_n).$$

Laut Theorem 9.1.2(iv) sind $\lambda_1, \dots, \lambda_n$ genau die Eigenwerte von A, wobei diese hier laut Definition 9.1.5(c) entsprechend ihrer algebraischen Vielfachheit oft vorkommen.

Wir müssen also nur noch zeigen, dass das Produkt der Eigenwerte gleich $\det A$ ist. Dazu beachten wir zunächst, dass

$$\det(-A) = p_A(0) = (-\lambda_1) \cdots (-\lambda_n) = (-1)^n \lambda_1 \cdots \lambda_n$$

gilt. Weil die Determinante n-linear in den Spalten ihres Arguments ist, gilt zudem $\det(-A) = (-1)^n \det A$, also folgt tatsächlich

$$\det(A) = \lambda_1 \cdots \lambda_n,$$

wie behauptet. □

Lassen Sie uns die bisherigen Resultate und Begriffe anhand zweier weiterer Beispiele besprechen:

Beispiele 9.1.7. (a) Die Matrix $A \coloneqq \pi \in \mathbb{R}^{1\times 1}$ besitzt genau einen Eigenwert, nämlich π – weil nämlich π die einzige Nullstelle des charakteristischen Polynoms $p_A(X) = \det(XI_1 - A) = X - \pi$ ist.

(b) Lassen Sie uns die Matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

betrachten. Ihr charakteristisches Polynom ist

$$p_A(X) = \det(XI_2 - A) = \det \begin{bmatrix} X & 1 \\ -1 & X \end{bmatrix} = X^2 + 1.$$

Da dieses Polynom keine reelle Nullstelle hat,[5] besitzt A keinen Eigenwert in $\mathbb{R}$.

Die Situation ändert sich allerdings, wenn wir A als Matrix in $\mathbb{C}^{2\times 2}$ auffassen – was problemlos möglich ist, da ja jede reelle Zahl auch eine komplexe Zahl ist. Im Körper $\mathbb{C}$ hat $p_A(X) = X^2 + 1$ zwei einfache Nullstellen, nämlich i und

[5] Warum nicht?

$-i$ Also besitzt A die beiden Eigenwerte i und $-i$ in $\mathbb{C}$, und beide haben die algebraische Vielfachheit 1.

Auch Korollar 9.1.6(b) lässt sich anhand dieses Beispiels gut veranschaulichen: Es ist A eine 2×2-Matrix, und sie besitzt genau zwei Eigenwerte (wobei jeder mit seiner algebraischen Vielfachheit 1 gezählt wird). Zudem ist das Produkt der beiden Eigenwerte gleich

$$i \cdot (-i) = 1 = \det A.$$

Um die zu den Eigenwerten i und $-i$ zugehörenden Eigenräume zu bestimmen, müssen wir den Kern von $iI_2 - A$ und $-iI_2 - A$ berechnen. Zum Beispiel mit Hilfe des Gauß-Algorithmus erhält man

$$\ker(iI_2 - A) = \operatorname{span} \begin{bmatrix} 1 \\ -i \end{bmatrix} \qquad \text{und} \qquad \ker(-iI_2 - A) = \operatorname{span} \begin{bmatrix} 1 \\ i \end{bmatrix}.$$

Somit ist zum Beispiel $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ ein Eigenvektor von A zum Eigenwert i und $\begin{bmatrix} 1 \\ i \end{bmatrix}$ ein Eigenwert von A zum Eigenwert $-i$.

Lineare Unabhängigkeit von Eigenvektoren

Theorem 9.1.8. *Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n \times n}$; außerdem seien $\lambda_1, \ldots, \lambda_d$ die Eigenwerte von A, wobei hier aber jeder verschiedene Eigenwert von A nur einmal aufgezählt sei, unabhängig von seiner algebraischen oder geometrischen Vielfachheit.*[6]

(a) *Wir bezeichnen mit $V \coloneqq \ker(\lambda_1 I_n - A) + \cdots + \ker(\lambda_d I_n - A)$ die Summe der Eigenräume von A. Dann gilt sogar $V = \ker(\lambda_1 I_n - A) \oplus \cdots \oplus \ker(\lambda_d I_n - A)$, d.h., die Summe ist direkt.*[7]

(b) *Inbesondere gilt: Wenn $v_1, \ldots, v_d \in \mathbb{K}^n \setminus \{0\}$ jeweils Eigenvektoren von A zu den Eigenwerten $\lambda_1, \ldots, \lambda_d$ sind, dann ist das Tupel $(v_1, \ldots, v_d)$ linear unabhängig.*

Beweis. (a) Sei $v \in V$ und seien $u_k, v_k \in \ker(\lambda_k I_n - A)$ für $k = 1, \ldots, d$ mit

$$v = \sum_{k=1}^{d} u_k \qquad \text{und} \qquad v = \sum_{k=1}^{d} v_k.$$

Wir müssen $u_1 = v_1, \ldots, u_d = v_d$ zeigen. Indem wir die Gleichung

$$\sum_{k=1}^{d} u_k = \sum_{k=1}^{d} v_k$$

[6] Beachten Sie, dass somit $d \leq n$ gilt. Weshalb?

[7] Zur Erinnerung: Direkte Summen hatten wir in Abschnitt 5.4 besprochen.

von links mit der Matrix $(\lambda_2 I_n - A)(\lambda_3 I_n - A) \dots (\lambda_d I_n - A)$ multiplizieren, erhalten wir

$$\prod_{j=2}^{d} (\lambda_1 - \lambda_j) u_1 = \prod_{j=2}^{d} (\lambda_1 - \lambda_j) v_1$$

und somit $u_1 = v_1$. Ebenso zeigt man, dass auch $u_2 = v_2$ gilt, und so weiter.

(b) Dies folgt direkt aus (a). □

Eigenwerte von Dreiecksmatrizen und Diagonalmatrizen

Den Begriff der **oberen Dreiecksmatrix** kennen Sie bereits aus Aufgabe 1 auf Tutoriumsblatt 13. Wir wiederholen ihn im folgenden und führen zudem noch einen weiteren analogen Begriff ein:

Definition 9.1.9 (Dreickecksmatrizen)**.** Sei $\mathbb{K}$ ein Körper und $n \in \mathbb{N}^*$; sei $A \in \mathbb{K}^{n \times n}$.

(a) Man nennt A eine **obere Dreiecksmatrix**, wenn unterhalb der Diagonalen von A nur Nullen stehen – d.h., wenn $A_{jk} = 0$ für alle $j, k \in \{1, \dots, n\}$ mit $j > k$ gilt.

(b) Man nennt A eine **untere Dreiecksmatrix**, wenn oberhalb der Diagonalen von A nur Nullen stehen – d.h., wenn $A_{jk} = 0$ für alle $j, k \in \{1, \dots, n\}$ mit $j < k$ gilt.

(c) Man nennt A eine **Dreiecksmatrix**, wenn A eine obere oder eine untere Dreiecksmatrix ist.

Man beachte, dass A genau dann eine obere Dreiecksmatrix ist, wenn A^{T} eine untere Dreiecksmatrix ist. Außerdem ist A genau dann eine Diagonalmatrix, wenn A zugleich eine obere und eine untere Dreieckmatrix ist.

Proposition 9.1.10 (Determinante und Eigenwerte von Dreiecksmatrizen)**.** *Sei $\mathbb{K}$ ein Körper, sei $n \in \mathbb{N}^*$ und sei $A \in \mathbb{K}^{n \times n}$ eine Dreiecksmatrix.*

(a) *Es ist* $\det A = \prod_{j=1}^{n} A_{jj}$ *– d.h. die Determinante von A ist gleich dem Produkt aller Diagonaleinträge von A.*

(b) *Die Eigenwerte von A (gezählt mit algebraischer Vielfachheit) sind genau die Diagonaleinträge von A. Für jeden Eigenwert ist außerdem die algebraische Vielfachheit gleich der Anzahl seiner Vorkommens auf der Diagonalen.*

Beweis. (a) Für obere Dreiecksmatrizen wurde das bereits in Aufgabe 1 auf Tutoriumsblatt 13 gezeigt. Für untere Dreiecksmatrizen folgt die Behauptung somit daraus, dass eine (quadratische) Matrix stets diesselbe Determinante hat wie ihre transponierte Matrix (Korollar 7.2.9).

(b) Aus (a) folgt, dass das charakteristische Polynom von A gleich

$$p_A(X) = \prod_{j=1}^{n} (X - A_{jj})$$

ist. Dies zeigt die Behauptung. □

Wir veranschaulichen die vorangehenden Konzepte wiederum kurz anhand einiger Beispiele:

Beispiele 9.1.11. (a) Die obere Dreiecksmatrix

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \in \mathbb{C}^{2\times 2}$$

besitzt laut Proposition 9.1.10(b) lediglich den Eigenwert 1, und dieser hat die algebraische Vielfachheit 2. Außerdem kann man mit dem Gauß-Algorithmus den Eigenraum $\ker(I_2 - A)$ bestimmen: Er ist gleich $\operatorname{span}(e_1)$ und hat somit die Dimension 1. Also besitzt der Eigenwert 1 von A die geometrische Vielfachheit 1.

(b) Die obere Dreiecksmatrix

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$$

hat die beiden Eigenwerte 1 und 2, jeweils mit algebraischer Vielfachheit 1. Die zugehörigen Eigenräumen kann man wieder mit dem Gauß-Algorithmus bestimmen: Es ist

$$\ker(1 \cdot I_2 - A) = \ker \begin{bmatrix} 0 & -1 \\ 0 & -1 \end{bmatrix} = \operatorname{span}(e_1)$$

und

$$\ker(2 \cdot I_2 - A) = \ker \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} = \operatorname{span} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Insbesondere haben die beiden Eigenwerte 1 und -1 jeweils die geometrische Vielfachheit 1.

(c) Die Diagonalmatrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1+i \end{bmatrix}$$

besitzt den Eigenwert 1 mit algebraischer Vielfachheit 2 und den Eigenwert $1 + i$ mit algebraischer Vielfachheit 1. Die zugehörigen Eigenräume lauten

$$\ker(1 \cdot I_2 - A) = \operatorname{span}(e_1, e_2) \qquad \text{und} \qquad \ker((1+i) \cdot I_2 - A) = \operatorname{span}(e_3).$$

Insbesondere ist die geometrischen Vielfachheit des Eigenwerts 1 gleich 2 und die geometrische Vielfachheit des Eigenwerts $1 + i$ gleich 1.

Man kann übrigens zeigen, dass die geometrische Vielfachheit eines Eigenwerte immer kleiner oder gleich seiner algebraischen Vielfachheit ist. Weil wir hier nur eine kurze Einführung in Eigenwerte geben, diskutieren wir dies an dieser Stelle nicht weiter – dies gehört aber zum Stoff der Linearen Algebra 2.

9.2 Diagonalisierung

Zum Abschluss kommen wir noch einmal auf den Begriff der Diagonalisierbarkeit zurück, den wir in Abschnitt 5.6 eingeführt hatten. Ob eine Matrix diagonalisierbar ist, lässt sich mit Hilfe von Eigenwerten und Eigenvektoren charakterisieren:

Theorem 9.2.1. *Sei $\mathbb{K}$ ein Körper, $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$. Dann sind die folgenden Aussagen äquivalent:*

(i) *Die Matrix A ist diagonalisierbar.*

(ii) *Es gibt eine Basis von $\mathbb{K}^n$, die aus Eigenvektoren von A besteht.*

(iii) *Die Summe der geometrischen Vielfachheiten der Eigenwerte von A ist gleich n.*

Für den Beweis des Theorems verweisen wir entweder auf die handschriftlichen Aufzeichnungen der letzten Vorlesung im Semester oder auf die Literatur. Besonders wichtig ist es, die Äquivalenz der Aussagen (i) und (ii) gut zu verstehen. Diese Äquivalenz wird zum Beispiel auch in [Mey00, Seiten 506–507] erklärt.

Bevor wir Bespiele diskutieren, fügen wir noch einige Bemerkungen an.

Bemerkung 9.2.2. (a) Wie weiter oben bereits erwähnt, kann man zeigen, dass die geometrische Vielfachheit eines Eigenwertes niemals größer als die algebraische Vielfachheit sein kann.

(b) Für Matrizen über $\mathbb{C}$ folgt aus dem Fundementalsatz der Algebra, dass die algebraischen Vielfachheiten der Eigenwerte einer Matrix $n \times n$-Matrix sich immer zu n summieren. Somit erhält man aus Theorem 9.2.1: Eine Matrix $A \in \mathbb{C}^{n\times n}$ (mit $n \in \mathbb{N}^*$) ist genau dann diagonalisierbar, wenn für jeden ihrer Eigenwerte die geometrische Vielfachheit gleich der algebraischen Vielfachheit ist.[8]

Lassen Sie uns nun zwei Beispiel besprechen.

Beispiele 9.2.3. (a) Wir betrachten die Matrix

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \in \mathbb{C}^{2\times 2}.$$

[8] Solche Eigenwerte nennt man auch **semi-simpel** oder **halbeinfach**.

Laut Beispiel 9.1.11(a) besitzt diese nur den Eigenwert 1, und dieser hat geometrische Vielfachheit 1. Also ist die Summe der geometrischen Vielfachheiten aller Eigenwerte echt kleiner als 2, und somit ist A laut Theorem 9.2.1 nicht diagonalisierbar.

(b) Lassen Sie uns nun die Matrix

$$A = \begin{bmatrix} 0 & 1 \\ -3 & 4 \end{bmatrix} \in \mathbb{C}^{2\times 2}.$$

betrachten. Indem man die Nullstellen des charakteristischen Polynoms berechnet, sieht man, dass die Eigenwerte von A die Zahlen 1 und 3 sind. Zum Beispiel mit dem Gaußverfahren kann man sehen, dass ihre Eigenräume jeweils gleich ein-dimensional sind und von den beiden Eigenvektoren

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{bzw.} \qquad \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

aufgespannt werden (wobei der linksstehende Vektor zum Eigenwert 1 gehört und der rechtsstehende Eigenvektoren zum Eigenwert 3.

Weil die beiden Eigenwerte die geometrische Vielfachheit 1 haben, summieren sich ihre geometrischen Vielfachheiten zu 2 auf. Deshalb ist A laut Theorem 9.2.1 diagonalisierbar. Wenn man dem Beweis von Theorem 9.2.1 folgt, kann man leicht eine invertierbare Matrix T und eine Diagonalmatrix D bestimmen, welche $AT = TD$ bzw. $A = TDT^{-1}$ erfüllt. Dazu schreibt man einfach auf die Diagonale von D die beiden Eigenwerte von A und wählt als Spalten von T die zugehörigen Eigenvektoren, d.h. man setzt

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \qquad \text{und} \qquad T = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}.$$

Zum Abschluss wollen wir zeigen, wie man die Diagonalisierung von Matrizen benutzen kann, um explizite Formeln für Matrixpotenzen zu berechnen.[9]

Beispiel 9.2.4 (Eine Matrix-Potenz)**.** Lassen Sie uns erneut die Matrix

$$A = \begin{bmatrix} 0 & 1 \\ -3 & 4 \end{bmatrix} \in \mathbb{C}^{2\times 2}.$$

aus dem vorangehenden Beispiel 9.2.3(b) betrachten. Lassen Sie uns aus Spaß versuchen, die Matrixpotenz A^{2022} zu berechnen.

Mit der Notation aus Beispiel 9.2.3(b) gilt $A = TDT^{-1}$, und somit

$$A^{2022} = TDT^{-1}\, TDT^{-1}\, \cdots TDT^{-1}\, TDT^{-1} = TD\, D\, \cdots\, D\, DT^{-1} = TD^{2022}T^{-1}.$$

[9] Sir kennen hierfür bereits eine Methode: Am Ende von Abschnitt 8.3 haben wir besprochen, wie man diese mit Hilfe des Satzes von Caley–Hamilton tun kann. Hier lernen Sie nun noch eine weitere Methoden kennen.

Wir haben unsere Aufgabe also im Wesentlichen darauf reduziert, D^{2022} zu bestimmen – was aber viel leichter ist, weil man Potenzen von Diagonalmatrizen ganz einfach berechnen kann: Es gilt

$$D^{2022} = \begin{bmatrix} 1^{2022} & 0 \\ 0 & 3^{2022} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 3^{2022} \end{bmatrix}.$$

Hieraus folgt nun

$$A^{2022} = TD^{2022}T^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3^{2022} \end{bmatrix} \frac{1}{2} \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 3 - 3^{2022} & 3^{2022} - 1 \\ 3 - 3^{2023} & 3^{2023} - 1 \end{bmatrix}.$$

Wir haben somit unser Ziel erreicht, eine explizite Formel für A^{2022} zu finden.

Appendices

Anhang A

Eine Einführung in Octave und Matlab

A.1 Überblick

Was sind Octave und Matlab?

Die beiden Computerprogramme *Matlab* und *Octave* werden hauptsätzlich dazu verwendet numerische Berechnungen und Simulationen durchzuführen. Weil Methoden aus der Linearen Algebra eine wichtige Rolle in der Numerik spielen, sind beide Programme auch sehr gut dazu geeignet, um Konzepte und Verfahren aus der Linearen Algebra zu veranschaulichen.

In der Vorlesung *Lineare Algebra 1* beschäftigen wir uns so gut wie gar nicht mit numerischen Fragestellungen (hierzu können Sie später im Studium spezielle Veranstaltungen belegen). Wir werden aber *Octave* in vielen Bonusaufgaben auf den Übungsblättern benutzen um Konzepte aus der Vorlesung am Computer zu implementieren und zu visualisieren.

Octave vs. Matlab

Matlab ist ein proprietäres Programm, das vom Unternehmen *The MathWorks* entwickelt und vertrieben wird. Es ist in manchen Branchen in der Industrie sehr verbreitet. *Octave* hingegen ist ein Freeware-Programm, welches Matlab stark nachempfunden ist. Es verwendet zur Bedienung dieselbe Programmiersprache wie Matlab, ist allerdings komplett unabhängig von Matlab implementiert.

In diesem Anhang werden wir die Verwendung von Octave beschreiben – allerdings lässt sich Matlab in vielerlei Hinsicht genauso bedienen. Insbesondere läuft Code, der für Octave geschrieben wurde, meist auch in Matlab, und umgekehrt.[1]

Octave können Sie von der folgenden Internetseite kostenfrei herunterladen und anschließend installieren:[2]

[1] Selbstverständlich gibt es aber auch Ausnahmen von dieser Regel.

[2] Wobei die Installation of Apple-Endgeräten womöglich etwas umständlicher sein kann als auf

https://octave.org

Matlab wiederum ist auf den Rechnern in einigen PC-Pools die Universität Passau installiert.

Programmieren

Die Steuerung von Octave und Matlab erfolgt hauptsächlich über eine Programmiersprache, die speziell für Matlab entwickelt wurde (und deshalb manchmal als *Matlab-Programmiersprache* bezeichnet wird). In diesem Anhang wird erklärt, wie sie Octave mit Hilfe der Matlab-Programmiersprache bedienen können.

Somit gibt dieser Anhang auch automatisch eine kurze Einführung in einige Programmiergrundlagen. Der Anhang ist so konzipiert, dass Sie ihn auch dann gut verstehen können, wenn Sie noch nie zuvor programmiert haben und auch nicht nebenbei in einer anderen Veanstaltung programmieren lernen.[3]

Wichtiger Hinweis:
Falls Sie Lehramt Gymnasium studieren und Ihr zweites Fach nicht Informatik ist, gibt es in Ihrem Studium keine Pflichtveranstaltung, bei der Sie laut Modulkatalog Programmieren lernen müssen. Deshalb ist es auf jeden Fall empfehlenswert, diese Chance zu nutzen, um freiwillig einen Einblick in die Programmierung zu erhalten – Prorgammieren ist eine Fähigkeit, die Ihnen in verschiedenen Situationen sehr nützlich sein kann, und mit Octave ist der Einstieg sehr einfach.

Graphische Version vs. Kommandozeilen-Variante

Nach der Installation können Sie Octave in zwei Varianten öffenen:

- Wenn Sie die graphische Variante (*GNU Octave GUI*[4]) starten, öffnet sich ein Fenster mit einigen verschiedenen Bedienelementen. Das wichtigste Element ist das *Befehlsfenster* (das manchmal auch *Konsole* oder *Kommandozeile* genannt wird).

- Wenn Sie die Kommandozeilen-Variante (*GNU Octave CLI*[5]) starten, öffnet sich ein sehr farbloses Fenster, das nur aus dem Befehlsfenster besteht.

Wir werden zwar im Wesentlichen mit dem Befehlsfenster arbeiten (für welches es egal ist, welche Variante Sie verwenden) – aber falls Sie noch keine Erfahrung mit der Verwendung von Kommandozeilen haben, ist es trotzdem empfehlenswert, dass Sie die graphische Variante benutzen. Im Laufe der Veranstaltung werden wir

Windows- oder Linuxrechnern. Aber auch auf Apple-Geräten sollte es (mit etwas Mehraufwand bei der Installation) gut funktionieren.

[3] Falls Sie bereits Programmiererfahrung haben oder nebenbei in einer anderen Veranstaltung programmieren lernen, macht das die Sache für Sie natürlich noch einfacher.

[4] *GUI* steht für *Graphical User Interface*.

[5] *CLI* steht für *Command Line Interface*.

nämlich auch ein paar wenige Funktionen verwenden, die in der graphischen Variante etwas zugänglicher sind.

Erste Eingaben im Befehlsfenster

Nachdem Sie (die graphische Variante von) Octave geöffnet haben, können Sie zum Einstieg ein paar Rechnungen ins Befehlsfenster eingeben. Die Syntax ist ähnlich, wie Sie es zum Beispiel von einem modernen Taschenrechner erwarten würden. Wenn Sie eine Rechnung eingeben und Enter drücken, erhalten Sie das Ergebnis.

Ein paar Hinweise:

- Das Multiplikationszeichen ist der Stern *.

  Das Divisionszeichen ist der Schrägstrich /. Ein Doppelpunkt als Divisionszeichen wird vom Programm nicht erkannt.[6]

- Kommazahlen werden im englischsprachigen Format eingegeben, d.h. anstelle eines Kommas wird ein Punkt verwendet.

- Um Potenzen zu berechnen können Sie das Symbol ^ benutzen.

  Zum Beispiel wird 3^4 in Octave eingegeben als 3^4. Als Ergebnis erhalten Sie, wie erwartet, 81.

- Octave respektiert die Konvention „Punkt vor Strich“, und Sie können auch Klammern verwenden.

- Es ist in Octave hingegen nicht möglich, das Multiplikationszeichen wegzulassen. Die Eingabe $(3 + 1)2$ – die man in der Mathematik als $(3 + 1) \cdot 2$ lesen würde – führt in Octave zu einem Syntaxfehler.[7]

  Probieren Sie es ruhig einmal aus – als Ausgabe in der nächsten Zeile erfolgt dann eine Fehlermeldung, und Sie können anschließend weiterarbeiten, als hätten Sie die Zeile „$(3 + 1)2$“ nie eingegeben.

Machen Sie sich mit dem Befehlsfenster vertraut, indem Sie einige Rechnungen eingeben und ein wenig herumprobieren.

Variablen

Nun wird es etwas interessanter: Sie können im Befehlsfenster in Octave auch *Variablen* definieren und ihnen einen Wert zuweisen. Hierzu wird in Octave das Symbol

[6] Der Doppelpunkt als Divisionszeichen ist nämlich sowohl in der Mathematik als auch in der Informatik äußerst unüblich.

[7] Das ist übrigens keine Eigenheit von Octave, sondern in vielen weiteren Programmiersprachen genauso.

= verwendet.[8] Wenn Sie zum Beispiel

$$a = 2$$

eingeben und auf Enter drücken, wird dadurch eine Variable mit dem Namen a definiert und ihr der Wert 2 zugewiesen.[9,10]

Sie können eine Variable auch jederzeit überschreiben. Nachdem Sie wie oben beschrieben der Variable a den Wert 2 zugewiesen haben, können Sie zum Beispiel

$$a = -1$$

eingeben (und anschließend nicht vergessen, auf Enter zu drücken). Der Wert der Variablen wird dadurch überschrieben und die Variable hat ab sofort den Wert -1.

Mit Variablen können Sie rechnen wie mit Zahlen. Geben Sie zum Beispiel als nächstes

$$a * 2 + 7$$

ein, und Sie erhalten (wie immer nach Drücken der Enter-Taste) erwartungsgemäß das Ergebnis 5.

Bemerkungen:

- Natürlich können Sie auch mehrere Variablen einführen (und mit ihnen rechnen).
- Variablennamen müssen nicht unbedingt aus nur einem Buchstaben bestehen, sondern können auch ganze Wörter sein.
- Übrigens: Wenn es Sie nervt, dass nach der Eingabe jeder Zeile eine weitere Zeile mit dem Ergebnis eingeblendet wird, dann können Sie eine Zeile, die Sie eingeben, einfach mit einem Semikolon abschließen – dadurch wird die Ausgabe des Ergebnisses unterdrückt.

Probieren Sie ein wenig herum, indem Sie ein paar Variablen eingeben und mit ihnen einige Rechnungen durchführen!

In Abbildung A.1.1 (auf der nächsten Seite) sehen Sie einen Screenshot des Befehlsfenster, in dem einige (einfache und willkürliche) Rechungen durchgeführt wurden.

[8] Das ist am Anfang ein bisschen gewöhnungsbedürftig: Das = wird in dieser Software also nicht verwendet um mathematische Aussagen – wie zum Beispiel „$3 + 4 = 7$“ – zu beschreiben, sondern es wird verwendet um einer Variable einen Wert zuzuweisen.

[9] Hier unterscheidet sich Octave von Sprachen von C++ oder Java: Sie müssen eine Variable in Octave nicht erst deklarieren oder irgendwie anderweitig einführen – sondern die Variable wird automatisch angelegt, wenn Sie ihr einen Wert zuweisen.

[10] Falls Sie schon einmal in einer Sprache wie C++ oder Java programmiert haben (falls nicht, ignorieren Sie den Rest dieser Fußnote bitte!), wird Ihnen vermutlich noch etwas anderes auf den ersten Blick auffallen: Sie müssen eine Variable in Octave nicht typisieren um Sie zu verwenden. (Was sehr bequem ist, aber auch große Nachteile mit sich bringt.)

```
Befehlsfenster
>>
>>
>> 3+4
ans =  7
>>
>>
>> x = -3
x = -3
>>
>> y = 17.2
y =  17.200
>>
>> (x+y)/2
ans =  7.1000
>>
>>
>> VariableMitLangemNamen = 4.3;
>> a = 19;
>>
>> VariableMitLangemNamen - a
ans = -14.700
>>
```

Abbildung A.1.1: Screenshot: Einige einfache Eingaben im Befehlsfenster von Octave.

A.2 Logische Ausdrücke

Wahrheitswerte

Sie können in Octave auch mit Wahrheitswerten rechnen. Die beiden Wahrheitswerte heißen in Octave *true* und *false*. Das sind aber nur Schlüsselwörter, die in Octave in Wirklichkeit als 1 (für *true*) und 0 (für *false*) dargestellt werden. Wenn Sie zum Beispiel

$$A = \text{false}$$

im Befehlsfenster eingeben, sehen Sie, dass der Variablen A einfach der Wert 0 zugewiesen wird (Sie hätten stattdessen also genauso gut auch „$A = 0$“ eingeben können). Und wenn Sie

$$B = \text{true}$$

eingeben, dann wird B der Wert 1 zugewiesen (d.h., Sie hätten genauso gut auch „$B = 1$“ eingeben können).

Es gibt die üblichen logischen Operationen auch in Octave, allerdings mit etwas anderer Schreibweise. Die folgende Tabelle listet die zugehörigen Befehle für zwei Variablen A und B auf:[11]

Logische Verknüpfung	und	oder	exkl. oder	nicht
Mathematische Notation	$A \wedge B$	$A \vee B$	$A \dot{\vee} B$	$\neg A$
Notation in Octave	$\mathrm{and}(A, B)$	$\mathrm{or}(A, B)$	$\mathrm{xor}(A, B)$	$\mathrm{not}(A)$
Alternative Notation in Octave	$A \,\&\, B$	$A \,\vert\, B$		$\sim A$

Machen Sie sich mit den Schreibweisen vertraut, indem Sie sie im Befehlsfenster ausprobieren.

Skripte

Wenn Sie etwas längere Rechnungen durchführen, ist es sinnvoll diese nicht einfach im Befehlsfester einzugeben, sondern die entsprechenden Befehle alle in eine Datai zu schreiben. Dies hat den Vorteil, dass Sie die komplette Berechnung jederzeit mit wenig Aufwand ändern und dann wieder komplett ausführen können. Eine Datei, die eine Abfolge von Befehlen enthält, nennt man **Skript**. Skripte funktionieren in Octave folgendermaßen:

- Jedes Skript ist schlicht eine Datei, die die Dateiendung .m hat und deren Inhalt aus Text besteht. Jede Zeile des Skriptes muss ein Kommando sein, dass Sie genauso gut auch im Befehlsfenster eingeben könnten.

  Leerzeilen sind in Skripten auch erlaubt und werden beim Ausführen ignoriert.

- Oben links in der graphischen Benutzeroberfläche von Octave sehen Sie ein kleines Fenster, in dem Sie erkennen, auf welchen Ordner auf Ihrem Computer Octave gerade zugreift. Sie können wie gewohnt durch Ihre Ordnerstruktur navigieren.

  Wichtig ist vorerst nur: Wenn Sie ein Skript ausführen wollen, muss das Skript in dem Ordner liegen, der in diesem Fenster gerade geöffnet ist.

- Zum Erstellen von Skripten finden Sie ganz oben links (in der Nähe des Reiters *Datei*) einen Button. Wenn Sie daraufklicken, wird ein leeres Skript erstellt. Sie können in jede Zeile ein Kommando eingeben, genau wie im Befehlsfenster. Vergessen Sie nicht, die Datei zu speichern, bevor Sie sie ausführen möchten (und achten Sie darauf, sie in dem Ordner zu speichern, in dem Sie sie ausführen möchten).

[11] Falls Sie Programmierkenntnisse z.B. in C++ oder in Java haben (falls nicht, ignorieren Sie diese Fußtnote bitte), sollten Sie beim Lesen der Tabelle folgendes beachten: Das kaufmännische und-Zeichen & bzw. der einfache Strich | werden in Octave – anders als in C++ und in Java – nicht als bitweise *und* bzw. *oder* verwendet, sondern einfach nur als logisches *und* bzw. *oder*.

- Zum Ausführen des Skripts geben Sie im Befehlsfenster von Octave den Namen der Datei ein – allerdings lassen Sie dabei die Dateiendung .m weg.

  (Denken Sie daran, dass Sie, wie oben bereits gesagt, zuerst in den Ordner navigieren müssen, in dem das Skript liegt, damit das Ausführen funktioniert.)

- Oft ist es eine gute Idee, ein bisschen zu erklären, was der eigenen Code tut. Dazu kann man sogenannte **Kommentare** verwenden:

  Wenn Sie irgendwo im Skript ein Prozentzeichen verwenden, dann wird der Rest dieser Zeile (ab dem Prozentzeichen) von Octave beim Ausführen des Skripts ignoriert. Alles, was hinter einem Prozenzeichen steht, bezeichnet man als **Kommentar**. Weil Kommentare die Ausführung des Skriptes nicht beeinflusst, können Sie sie nutzen, um Erläuterungen im Skript einzufügen.

  Hinweis: Nutzen Sie Kommentare sehr großzügig! Das erleichtert das Lesen Ihrer Skripte.

Das war's im Wesentlichen. In Abbildung A.2.1 auf der nächsten Seite finden Sie einen Screenshot[12] eines Skripts, mit dem man in Octave nachprüfen kann, dass die de Morgansche Regel aus Proposition 1.2.10(a) tatsächlich stimmt.[13] Die Ausgabe, die Sie erhalten, wenn Sie das Skript im Befehlsfenster ausführen, können Sie in Abbildung A.2.2 auf der übernächsten Seite sehen.

[12] Der Grund, weshalb hier nur Screenshots abgebildet werden statt das Skript als Text einzubinden (was deutlich besser aussehen würde), ist folgender: Viele Leute neigen dazu, zu „programmieren", indem sie eine „Vorlage" irgendwo rauskopieren und dann nur an ihre Bedürfnisse anpassen. Diese Methode ist aber denkbar schlecht geeignet, um Programmieren allgemein und eine Programmiersprache im Speziellen wirklich zu lernen. Die Verwendung der Screenshots soll verhindern, dass Sie beim Bearbeiten der Übungsaufgaben diese Copy- & Paste-Methode verwenden.

[13] D.h. anstatt die Werte in der Wahrheitstabelle per Hand zu bestimmen, lässt man sie von Octave berechnen. Am besten vergleichen Sie die Wahrheitswerte, die vom Skript ausgegeben werden, mit denen, die wir im Beweis der Proposition „per Hand" in die Wahrheitstabelle eingefügt haben.

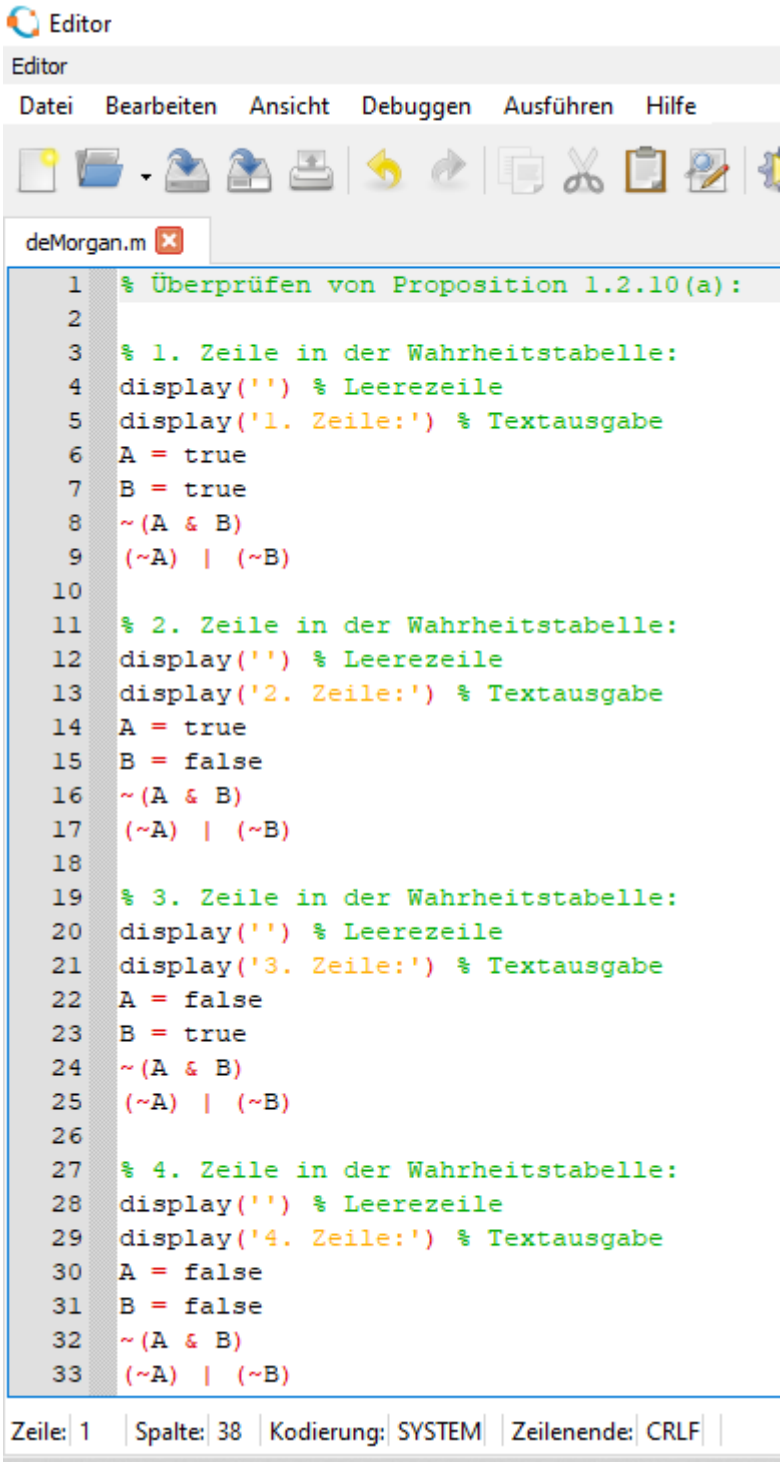

Abbildung A.2.1: Screenshot des Skripts *deMorgan.m*, mit dem eine der aussagenlogischen de Morganschen Regeln am Computer überprüft werden kann. Die Ausgabe beim Ausführen des Skripts ist in Abbildung A.2.2 gezeigt.

```
Befehlsfenster
>> deMorgan

1. Zeile:
A = 1
B = 1
ans = 0
ans = 0

2. Zeile:
A = 1
B = 0
ans = 1
ans = 1

3. Zeile:
A = 0
B = 1
ans = 1
ans = 1

4. Zeile:
A = 0
B = 0
ans = 1
ans = 1
>>
```

Abbildung A.2.2: Screenshot der Ausgabe im Befehlsfenster beim Ausführen des Skripts *deMorgan.m* (welches in Abbildung A.2.1 gelistet ist).

A.3 Fallunterscheidungen und Funktionen

In diesem Abschnitt besprechen wir zwei Themen, die auf den ersten Blick wenig mit einander zu tun haben, die aber beide gut zu Abschnitt 1.5 passen.

Fallunterscheidungen

Im Alltag sind Sie häufig mit Situationen konfrontiert, in denen Sie Entscheidungen treffen müssen; wie Sie sich entscheiden, kann von einer Vielzahl von Einflussfaktoren abhängen. Man könnte durchaus behaupten: Das Leben wird vor allem dadurch interessant und facettenreich, dass Sie Entscheidungen treffen müssen.

Bei Computerprogrammen ist es ganz ähnlich: Interessant (und nützlich) werden Programme vor allem dadurch, dass an manchen Stellen im Programmverlauf eine Entscheidung getroffen werden kann, was das Programm nun als nächstes tun soll.[14]

[14] Genau das passiert übrigens ständig, wenn Sie einen Computer oder ein Smartphone oder ein

fallunterscheidung.m

```
1
2  zahl1 = 3;
3  zahl2 = -2;
4  zahl3 = 9;
5
6  if (zahl1 <= zahl2)
7    a = 2;
8  elseif (zahl1 < zahl3)
9    a = 5;
10 elseif (zahl2 == zahl3)
11   a = 4.3;
12 else
13   a = 0;
14 end
15
16 a
```

Abbildung A.3.1: Beispiel-Listing einer sehr einfachen Fallunterscheidung in Octave.

Alle Programmiersprachen verfügen über Befehle, mit denen solche Entscheidungen des Programms modelliert werden können. Das einfachste solche Konstrukt ist eine **Fallunterscheidung**. In Octave kann man Fallunterscheidungen mit Hilfe der Schlüsselwörter *if* and *else* programmieren. Dies lässt sich am einfachsten mit Hilfe eines kurzen Videos (ca. $7\frac{1}{2}$ Minuten) erklären, das Sie sowohl im Stud.IP (in der Übungs-Veranstaltung) als auch unter folgendem Link finden:

Link zum Video über Fallunterscheidungen in Octave (ca. $7\frac{1}{2}$ Minuten)

Um Ihnen das Nachgucken der verwendeten Befehle zu erleichtern, können Sie den Screenshot einer sehr einfachen Fallunterscheidung zudem in Abbildung A.3.1 ansehen.

Funktionen

Funtionen funktionieren in Programmiersprachen ähnlich wie in der Mathematik (auch, wenn die Details sich häufig unterscheiden): Es handelt sich im Teiles eines Programms, denen man Eingaben – die sogenannten **Argumente** – übergeben

anderes elektronisches Gerät benutzen: Je nachdem was Sie tun, reagiert das Programm anders: Klicken Sie in Ihrem Browser zum Beispiel auf einen Link, dann wird die verlinkte Seite geöffnet. Klicken Sie hingegen auf den Button zum Schließen des Browsers, so wird er tatsächlich geschlossen. Das Programm *Browser* reagiert also auf unterschiedliche Situationen („Wohin wurde geklickt?“) mit einer unterschiedlichen Fortsetzung des Programmverlaufs – in den genannten Beispielen lädt der Browser entweder die Seite die unter dem angeklickten Link hinterlegt ist, oder er beendet sich selbst.

meineErsteFunktion.m

```
1  function rueckgabeWert = meineErsteFunktion(z)
2
3    if (z >= 0)
4      rueckgabeWert = 2*z;
5    else
6      rueckgabeWert = -1;
7    end
8
9  endfunction
```

Abbildung A.3.2: Octave-Implementierung der Funktion ℓ aus Beispiel 1.5.3(c).

kann, und die diesen Eingaben Ausgaben – die man, wie Sie bereits wissen, in der Mathematik **Funtionswerte** nennt – zuordnen.

Ein wichtiger Unterscheid ist, dass Funktionen in Programmiersprachen wie z.B. Octave Teile des Programms sind, die wirklich ausgeführt werden können. Sie können eine Funktion also zum Beispiel im Befehlsfenster ausführen (wobei Sie ihr konkrete Werte als Argumente übergeben müssen), und der Computer berechnet dann den zugehörigen Funktionswert.

Damit dies funktionieren kann, muss eine Funktion natürlich so programmiert werden, dass sie das Gewünschte tut. Wie man das in Octave macht, lässt sich ebenfalls am besten in einem kurzen Video (ca. $6\frac{1}{2}$ Minuten) erklären:

Link zum Video über Funktionen in Octave (ca. $6\frac{1}{2}$ Minuten)

Die im Video implementierte Funktion (die aus Beispiel 1.5.3(c) stammt) finden Sie auch in Abbildung A.3.2 nochmals als Screenshot.

A.4 Komplexe Zahlen in Octave

Octave kann nicht nur mit reellen Zahlen rechnen, sondern genauso gut auch mit komplexen Zahlen. Die imaginäre Einheit wird in Octave ebenfalls einfach mit i bezeichnet.

Wenn Sie im Befehlsfenster also zum Beispiel

$$(1+2i)*(-3+i)$$

eingeben, berechnet Octave das Ergebnis für Sie; es lautet[15]

$$-5-5i.$$

[15]Das können Sie natürlich ganz leicht auch per Hand nachrechnen.

Sie können komplexe Zahlen in Octave auch addieren, subtrahieren, dividieren oder potenzieren. Wenn Sie zum Beispiel

$$1/(1+i)$$

eingeben, erhalten Sie als Ergebnis $0.5 - 0.5i$, and wenn Sie zum Beispiel

$$i^6$$

eingeben, erhalten Sie -1 als Ergebnis.[16]

Zwei nützliche Funkionen, die Sie auf eine komplexe Zahl z anwenden können, sind *conj* und *abs*:

Octave-Kommando	Ergebnis	Beschreibung
conj(z)	$\overline{z}$	konjugierte komplexe Zahl von z
abs(z)	$\|z\|$	Betrag von z

A.5 For-Schleifen

Damit ein Computer größere Aufgaben verrichten kann, benötigt man eine Möglichkeit, dem Computer sehr viele Kommandos auf einmal mitzuteilen. Hierzu kann man unter anderem **Schleifen** verwenden – diese sorgen dafür, dass ein bestimmtes Stück Code sehr oft durchlaufen wird.

Es gibt verschiedene Varianten von Schleifen. Besonders einfach zu verwenden ist in Octave die *for-Schleife*. Sie besteht aus mehreren Zeilen: Die erste Zeile ist zum Beispiel von der Form

$$for\ k = 1:100$$

und die letzte Zeile lautet

$$endfor$$

und zeigt somit das Ende der for-Schleife an. Sobald die Schleife ausgeführt wird, passiert folgendes:

- Zuerst wird k auf den Wert 1 gesetzt. Dann werden alle Code-Zeilen ausgeführt, die vor $endfor$ stehen.
- Nun springt der Computer zurück an den Anfang der Schleife, setzt k auf 2, und führt wieder alle Code-Zeilen aus, die vor $endfor$ stehen.
- Dies geht so weiter, bis k auf den Wert 100 gesetzt wird. Auch dann werden noch einmal alle Code-Zeilen bis $endfor$ ausgeführt. Anschließend ist die Ausführung der Schleife beendet.

[16] Übrigens: Warum eigentlich?

gaussSumme.m

```
1  function ergebnis = gaussSumme(n)
2    % Berechnet die Summe der ganzen Zahlen von 1 bis n
3
4
5    % Lege eine Variable an, auf die schrittweise alle Zahlen
6    % von 1 bis n addiert werden sollen:
7    zwischensumme = 0;
8
9    % Starte eine Schleife, die insgesamt n mal durchlaufen wird:
10   for k=1:n
11     % Addiere im k-ten Schritt die Zahl k
12     % zur Zwischensumme:
13     zwischensumme = zwischensumme + k;
14   endfor
15
16
17   % Uebergebe die insgesamt berechnete Summe an die Variable,
18   % die von der Funktion zurueckgegeben wird:
19   ergebnis = zwischensumme;
20
21 endfunction
```

Abbildung A.5.1: Octave-Implementierung einer Funktion names *gaussSumme*. Sie berechnet die Summe $\sum_{k=1}^{n} k$. Diesselbe Summe könnte man aber auch ohne Schleife berechnen, indem man die Gaußsche Summenformel auf Beispiel 3.2.5(a) verwendet.

Selbstverständlich muss die Variable, die von 1 bis 100 durchgezählt wird, nicht k heißen – Sie können ihr auch einen beliebigen anderen Namen geben. Und selbstverständlich muss die Schleife auch nicht von 1 bis 100 zählen; sie können auch andere Grenzen angeben.

Wenn Sie Schleifen in einem Skript oder in einer Funktion benutzen, ist es zudem nützlich zu wissen, dass die Grenzen selbst auch Variablen sein dürfen – so kann man erreichen, dass erst zur Laufzeit entschieden wird, aus wievielen Durchläufen die Schleife besteht.

In den Abbildungen A.5.1 und A.5.2 sehen Sie die Screenshots zweier Octave-Funktionen, mit denen die Summe aller Zahlen von 1 bis n sowie die n-te Fibonacci-Zahl berechnet werden kann.

fibonacci.m

```
function ergebnis = fibonacci(n)
  % Berechnet die n-te Fibonacci-Zahl.

  % Die Fibonacci-Zahlen f_0 und f_1 sind beide 1.
  % Wir speichern diese Werte in Variablen f und fNext:
  f = 1;
  fNext = 1;

  % Wir starten nun eine Schleife, die n-1 mal durchlaufen wird:
  for k=2:n
    % An dieser Stelle hat die Variable f den Wert f_(k-2),
    % und die Variable fNext hat den Wert f_(k-1)
    % Im Folgenden werden beide Variablen so geaendert, dass Sie
    % jeweils die naechsthoehere Fibonacci-Zahl als Wert haben.

    % Berechne den Wert von f_k:
    fk = f + fNext;

    % Ueberschreibe jetzt f und fNext mit den
    % jeweils nachfolgenden Fibonacci-Zahlen:
    f = fNext;  % Nun hat f den Wert f_(k-1)
    fNext = fk; % Nun hat fNext den Wert f_k.

  endfor
  % Wenn die Schleife zu Ende ist, hatte k zuletzt den Wert n,
  % also hat fNext nun den Wert f_n.
  % Wir uebergeben diesen Wert an die Variable,
  % die von der Funktion zurueckgegeben wird:
  ergebnis = fNext;

endfunction
```

Abbildung A.5.2: Octave-Implementierung einer Funktion names *fibonacci*. Sie berechnet die n-te Fibonacci-Zahl aus Beispiel 3.2.2.

A.6 Vektoren und Matrizen

In Octave kann man Vektoren und Matrizen sehr einfach handhaben.[17] Grundsätzlich gilt hierbei:

- Jeder Vektor ist ebenfalls einfach eine Matrix.
- Um Matrizen explizit anzugeben, benutzt man eckige Klammern, innerhalb derer die Einträge der Matrix dann zeilenweise aufgelistet werden. Zeilen werden durch Strichpunkte getrennt; die Einträge innerhalb einer Zeile werden durch Kommata[18] getrennt.
- Auf den k-ten Eintrag eines Vektors x greifen Sie durch die Notation $x(k)$ zu,
- Auf den Eintrag einer Matrix A an der Stelle (j, k) greifen Sie durch die Notation $A(j, k)$ zu.[19] [20]
- Mit dem Befehl $size(A)$ erhalten Sie einen Zeilenvektor mit zwei Einträgen zurück; sein erster Eintrag gibt die Anzahl der Zeilen von A und, und sein zweiter Eintrag die Anzahl der Spalten von A.

Das Ganze lässt sich am Besten anhand eines Beispiel demonstrieren. Sehen Sie sich hierzu einfach das Skript *matrizen.m* an, welches Sie im Stud.IP in der Vorlesungsveranstaltung unter dem Dateiordner *Weitere Materialien* finden.

Im selben Ordner finden Sie außerdem eine Datei *matrixSumme.m*; sie enthält eine Funktion, in der „per Hand“ die Summe von zwei Matrizen mit Hilfe von for-Schleifen berechnet wird.

[17] Vektoren in Octave verhalten sich in vielerlei – aber nicht in jeder – Hinsicht ähnlich wie Arrays in vielen anderen Programmiersprachen.

[18] Oder durch Leerzeichen – aber bei Verwendung von Leerzeichen zur Trennung der Einträge können manchmal Subtilitäten auftreten, wenn man zugleich noch Minuszeichen verwendet.

[19] Wie in der Mathematik durchgehend üblich, bezeichnet j dabei den Zeilenindex und k den Spaltenindex.

[20] Also bedeutet $A(j, k)$ in Octave das, was wir in der Vorlesung als A_{jk} notieren würden.

Anhang B

Invertierbare Matrizen – Ein Überblick zum Abschluss

Wir hatten im Laufe der Vorlesung zahlreiche verschiedene Charakterisierungen der Invertierbarkeit von Matrizen besprochen. Des besseren Überblicks halber fassen wir viele von Ihnen hier noch einmal zusammen:

Theorem B.0.1 (Charakterisierung invertierbarer Matrizen)**.** *Sei $\mathbb{K}$ ein Körper, sei $n \in \mathbb{N}^*$ und $A \in \mathbb{K}^{n\times n}$. Dann sind folgende Aussagen äquivalent:*

(i) *A ist invertierbar.*

(ii) *A^{T} ist invertierbar.*

(iii) *A ist rechtsinvertierbar.*

(iv) *A ist linksinvertierbar.*

(v) *Die reduzierte Zeilenstufenform von A ist gleich I_n.*

(vi) *Es gilt* $\ker A = \{0\}$.

(vii) *Der Spaltenraum von A ist gleich $\mathbb{K}^n$.*

(viii) *Es gilt* $\operatorname{rang}(A) = n$.

(ix) *Die Spalten von A sind linear unabhängig.*

(x) *Die Zeilen von A sind linear unabhängig.*

(xi) *Für jedes $b \in \mathbb{K}^n$ besitzt das Gleichungssystem $Ax = b$ mindestens eine Lösung $x \in \mathbb{K}^n$.*

(xii) *Für jedes $b \in \mathbb{K}^n$ besitzt das Gleichungssystem $Ax = b$ höchstens eine Lösung $x \in \mathbb{K}^n$.*

(xiii) *Für jedes* $b \in \mathbb{K}^n$ *besitzt das Gleichungssystem* $Ax = b$ *genau eine Lösung* $x \in \mathbb{K}^n$.

(xiv) *Es gilt* $\det(A) \neq 0$.

(xv) *Das Element* $0 \in \mathbb{K}$ *ist kein Eigenwert von* A.

Literaturverzeichnis

[Axl97] Sheldon Axler. *Linear algebra done right.* New York, NY: Springer, 1997.

[Beu14] Albrecht Beutelspacher. *Lineare Algebra. Eine Einführung in die Wissenschaft der Vektoren, Abbildungen und Matrizen.* Heidelberg: Springer Spektrum, 2014.

[Blo11] Ethan D. Bloch. *Proofs and fundamentals. A first course in abstract mathematics.* Berlin: Springer, 2011.

[Bos14] Siegfried Bosch. *Lineare Algebra.* Heidelberg: Springer Spektrum, 2014.

[Bos20] Siegfried Bosch. *Algebra.* Berlin: Springer Spektrum, 9te auflage edition, 2020.

[Fis11] Gerd Fischer. *Lernbuch Lineare Algebra und Analytische Geometrie. Das Wichtigste ausführlich für das Lehramts- und Bachelorstudium.* Wiesbaden: Vieweg+Teubner, 2011.

[FS20] Gerd Fischer and Boris Springborn. *Lineare Algebra. Eine Einführung für Studienanfänger.* Berlin: Springer Spektrum, 2020.

[Fuh12] Paul A. Fuhrmann. *A polynomial approach to linear algebra.* New York, NY: Springer, 2012.

[GVL13] Gene H. Golub and Charles F. Van Loan. *Matrix computations.* Baltimore, MD: The Johns Hopkins University Press, 4th ed. edition, 2013.

[Jän08] Klaus Jänich. *Lineare Algebra.* Berlin: Springer, 2008.

[Lan86] Serge Lang. *Introduction to linear algebra. 2nd ed.* Springer, Cham, 1986.

[Mey00] Carl D. Meyer. *Matrix analysis and applied linear algebra (incl. CD-ROM and solutions manual).* Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2000.

[Rau08] Wolfgang Rautenberg. *Einführung in die mathematische Logik. Ein Lehrbuch.* Wiesbaden: Vieweg+Teubner, 2008.

[TT08] Gerald Teschl and Susanne Teschl. *Mathematik für Informatiker. Band 1: Diskrete Mathematik und lineare Algebra.* Berlin: Springer, 2008.

[Wit13] Kurt-Ulrich Witt. *Mathematische Grundlagen für die Informatik. Mengen, Logik, Rekursion.* Wiesbaden: Springer Vieweg, 2013.

[Zie17] Martin Ziegler. *Mathematische Logik.* Basel: Birkhäuser/Springer, 2017.