

Skriptum

Lineare Algebra

Stefan Fredenhagen

16. Dezember 2018

Inhaltsverzeichnis

1	Grundbegriffe	6
1.1	Mengen	6
1.2	Abbildungen	15
1.3	Algebraische Strukturen: Gruppen, Körper, usw.	20
2	Geometrie im \mathbb{R}^n	27
2.1	Algebraisierung der Geometrie	27
2.2	Abstände und Winkel: Skalarprodukt	35
2.3	Kegelschnitte und Quadriken	44
3	Vektorräume	47
3.1	Allgemeine reelle Vektorräume	47
3.2	Ein Exkurs über lineare Gleichungssysteme	54
3.3	Basen und Dimension	62
4	Lineare Abbildungen und Matrizen	69
4.1	Lineare Abbildungen	69
4.2	Matrizenkalkül	79
4.2.1	Komplexe Matrizen	91
4.3	Basiswechsel und Normalform	94
5	Lineare Gleichungssysteme und affine Unterräume	102
5.1	Affine Unterräume	102
5.2	Lösung inhomogener linearer Gleichungssysteme	106
5.3	Quotientenvektorräume	110
6	Determinanten	115
6.1	Determinante und Volumen	115

6.2	Eigenschaften	123
6.3	Berechnungsverfahren und Anwendung	130
6.4	Vektorprodukt	142
7	Eigenwerte und Normalformen	146
7.1	Motivation	146
7.2	Eigenwerte und Eigenvektoren	150
7.3	Das charakteristische Polynom	155
7.4	Diagonalisierung und Trigonalisierung	163
7.5	Funktionen von Matrizen und Anwendungen	171
8	Euklidische und unitäre Vektorräume	180
8.1	Skalarprodukt	180
8.2	Orthonormalsysteme und Matrixdarstellungen	185
8.3	Orthogonale und unitäre Abbildungen	195
8.4	Symmetrische und hermitesche Abbildungen	200
8.5	Anwendungen: Quadriken und Matrixzerlegungen	209
9	Tensoren	219
9.1	Dualraum	219
9.2	Bilinearformen und Tensoren	224
9.3	Tensorprodukt	228

Herzlichen Dank an Thomas Zauner, der weite Teile des Skripts getippt und meine Zeichnungen in geeignete Abbildungen übersetzt hat!

Vorbemerkungen

Diese Vorlesung über lineare Algebra richtet sich an Studierende der Physik. Daher möchte ich ein paar Gedanken über das Verhältnis von Mathematik und Physik an den Anfang stellen. Dieses Verhältnis ist extrem komplex, und die folgenden, stark vereinfachenden Überlegungen werden dieser Komplexität natürlich nicht gerecht.

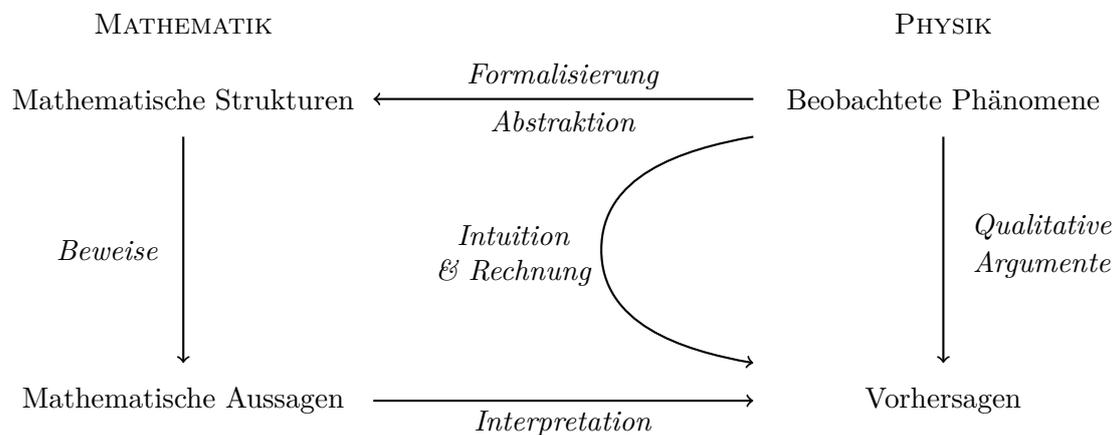
In der Mathematik geht man von wohldefinierten mathematischen Strukturen aus und versucht, mit Hilfe streng logischer Argumente mathematische Aussagen über diese Strukturen zu machen.

In der Physik geht man von den beobachteten Phänomenen aus und versucht, Gesetzmäßigkeiten zu finden und daraus Vorhersagen ableiten zu können. Dabei bedient man sich der Sprache der Mathematik.

Im optimalen Fall könnte das so aussehen: Die beobachteten Strukturen werden formalisiert, d.h. es wird eine mathematische Struktur definiert, die die beobachteten Gesetzmäßigkeiten erfasst. Durch mathematische Schlussfolgerungen können Aussagen abgeleitet werden, die wiederum als physikalische Vorhersagen interpretiert werden können. Die Überprüfung dieser Vorhersagen ist dann eine Überprüfung, ob der Prozess der Formalisierung gelungen ist.

In der Praxis sieht es meist etwas anders aus. Oft spielen auch rein anschauliche Argumente in der Physik eine Rolle, meist ist es eine Mischung aus Intuition und Rechnung, ohne dass die mathematische Beschreibung wirklich rigoros gemacht wurde.

Die verschiedenen Herangehensweisen haben alle ihre Berechtigung, aber es ist wichtig zu lernen, zwischen anschaulichen/intuitiven und rigorosen Argumenten zu unterscheiden.



In dieser Vorlesung geht es hauptsächlich um die Mathematik, auch wenn wir natürlich Anwendungen besprechen und versuchen wollen, mathematische Argumente und

Aussagen auch anschaulich zu verstehen. So wichtig diese Anschauung ist, zählt in der Mathematik letztlich nur der rein logische Beweis.

An dieser Stelle haben viele Studierende zu Beginn Schwierigkeiten, da in der Schulmathematik oft wenig Wert auf Beweise gelegt wird. Neben dem Einüben all der für die lineare Algebra relevanten Techniken, wird es in der Vorlesung daher auch darum gehen, das Verständnis der abstrakten Strukturen und Argumente zu schulen.

Betrachten wir ein einfaches Beispiel für einen mathematischen Satz und Beweis, der über 2000 Jahre alt ist, und den einige von Ihnen wahrscheinlich aus der Schule kennen.

Satz 0.1 (Satz von Euklid). *Es gibt unendlich viele Primzahlen.*

Wie Sie sich erinnern, ist eine Primzahl eine natürliche Zahl größer 1, die nur durch sich selbst und durch 1 teilbar ist.

Wie bei den meisten zahlentheoretischen Problemen nützt einem die Anschauung hier wenig und man muss sich ganz auf die mathematischen Argumente verlassen.

Beweis. Wie gehen wir vor? Ein direkter, konstruktiver Beweis scheint nicht so einfach — wir benötigen einen Algorithmus, der unendlich viele Primzahlen konstruiert. Stattdessen wählen wir wie Euklid einen Widerspruchsbeweis: Wir nehmen an, es gäbe nur endlich viele Primzahlen und zeigen, dass diese Annahme zu einem Widerspruch führt.

Angenommen, es gäbe nur endlich viele Primzahlen, die wir mit p_1, \dots, p_N bezeichnen wollen (N ist die Gesamtzahl aller Primzahlen). Betrachten wir dann $p = p_1 \cdot \dots \cdot p_N + 1$.

Bei Division von p durch jede Primzahl p_i bleibt immer Rest 1. p ist also durch keine Primzahl teilbar und damit durch überhaupt keine Zahl teilbar außer durch 1 und durch sich selbst. Dann müsste p eine weitere Primzahl sein, laut Annahme waren aber p_1, \dots, p_N schon alle Primzahlen. Damit sind wir zu einem Widerspruch gelangt, was bedeutet, dass unsere ursprüngliche Annahme falsch war. Es gibt also unendlich viele Primzahlen.

□

Das war jetzt ein Beispiel aus der Zahlentheorie, nicht aus der linearen Algebra (wobei es durchaus Beziehungen zwischen den Gebieten gibt), aber es ist eines der schönsten Beispiele für einen mathematischen Beweis.

Worum geht es jetzt in der linearen Algebra? Einiges davon werden Sie aus der Schule kennen: Vektoren, Systeme linearer Gleichungen, Matrizen ...

Grob gesagt geht es allgemein um lineare Strukturen, wobei linear einerseits die Proportionalität und andererseits die Additivität umfasst, d.h. im Beispiel des Newtonschen Gesetzes $\vec{F} = m\vec{a}$:

- $(2\vec{F}) = m(2\vec{a})$: doppelte Kraft führt zu doppelter Beschleunigung.

- $(\vec{F}_1 + \vec{F}_2) = m(\vec{a}_1 + \vec{a}_2)$: Die Summe von zwei Kräften führt zu einer Beschleunigung, die durch die Summe der Einzelbeschleunigungen gegeben ist.

Lineare Strukturen tauchen an vielen Stellen auf, insbesondere in der Physik: bei vektoriellen Gleichungen sowieso, aber approximativ praktisch überall, wenn eine Funktion näherungsweise durch ihre Tangente beschrieben werden kann.

Das Paradebeispiel für Anwendungen der linearen Algebra ist das Lösen von linearen Gleichungssystemen, wie z.B.

$$\begin{aligned}x + 2y &= 0 \\ -x + y &= 3.\end{aligned}$$

In diesem Fall ist die Lösung sehr einfach: Addieren beider Gleichungen ergibt $3y = 3$, also $y = 1$. Einsetzen davon in eine der ursprünglichen Gleichungen gibt $x = -2$.

Die lineare Algebra liefert eine komplette Lösungstheorie solcher Gleichungssysteme (die beliebig groß sein können) samt einem effizienten Algorithmus. Auch lineare Differentialgleichungen wie sie beispielsweise bei Stabilitätsanalysen auftreten (um z.B. die Resonanzfrequenz einer Brückenkonstruktion zu bestimmen) können mit Hilfe der linearen Algebra behandelt werden.

Alle Anwendungsgebiete aufzuzählen ist praktisch nicht möglich, aber eine gerade für die Physik nicht wegzudenkende Anwendung ist die *Quantenmechanik*. Die Formulierung der Quantenphysik baut auf der linearen Algebra auf — wenn Sie die Quantenmechanik verstehen wollen, dann widmen Sie sich jetzt erst einmal der linearen Algebra.

Literatur

Diese Vorlesung folgt nicht direkt einem einzelnen Buch, ist aber in vielen Teilen vom Buch *Lineare Algebra: Eine Einführung für Studienanfänger* von Gerd Fischer [F] inspiriert, insbesondere was die Notation betrifft. Der Inhalt der Vorlesung findet sich in den allermeisten Lehrbüchern über lineare Algebra wieder, eine relativ knappe Darstellung der für uns interessanten Konzepte kann man in dem Kapitel 7 über Lineare Algebra im Buch *Mathematik für Physiker* von Hans Kerner und Wolf von Wahl [KvW] nachlesen.

1 Grundbegriffe

In diesem Kapitel werden wir uns zunächst über einige Grundbegriffe wie Mengen und Abbildungen sowie algebraische Strukturen verständigen.

Vieles, aber vermutlich nicht alles aus diesem Kapitel wird Ihnen aus der Schule bekannt vorkommen und viele der Aussagen werden Ihnen intuitiv verständlich sein. Wir nutzen diese Vertrautheit, um auf halbwegs vertrautem Terrain die saubere mathematische Begriffsbildung und Argumentation einzuüben, die für die späteren abstrakten Konzepte unentbehrlich sind.

1.1 Mengen

Was ist eigentlich eine Menge? Üblicherweise lässt man an dieser Stelle zunächst Cantor (1845–1918) zu Wort kommen:

Unter einer Menge verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die Elemente von M genannt werden) zu einem Ganzen.

Wie Sie vielleicht bemerkt haben, ist dies keine Definition, die unsere höchsten mathematischen Rigorositätsansprüche erfüllt, fasst aber gut zusammen, was wir uns intuitiv unter einer Menge vorstellen. Unproblematisch sind sicherlich endliche Mengen, die wir durch Auflistung ihrer Elemente definieren können, z.B.

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

$$\{47, 33\}.$$

Selbst unter Mathematikern ist es üblich, in den Fällen, wo das Bildungsgesetz klar ist, manche Elemente mit Punkten abzukürzen,

$$\{1, 2, \dots, 10\},$$

$$\{\text{Montag, Dienstag}, \dots, \text{Sonntag}\}.$$

Eine Menge ist eindeutig dadurch charakterisiert, welche Elemente zu ihr gehören, die Zugehörigkeit wird mit dem Ihnen vermutlich vertrauten Symbol \in gekennzeichnet, z.B.

$$1 \in \{1, 2, 3\}. \quad (1.1)$$

Wir wollen die Menge $\{1, 2, 3\}$ für den Augenblick N_3 nennen, das schreiben wir so:

$$N_3 = \{1, 2, 3\}. \quad (1.2)$$

Da N_3 durch seine Elemente charakterisiert ist, können wir schreiben

$$x \in N_3 \iff x = 1 \text{ oder } x = 2 \text{ oder } x = 3 \quad (\text{bzw. } x = 1 \vee x = 2 \vee x = 3). \quad (1.3)$$

Insbesondere ist die Anordnung der Elemente unwichtig, z.B. ist auch $N_3 = \{3, 1, 2\}$.

Auch Mengen mit unendlich vielen Elementen sind Ihnen bekannt, z.B.

- \mathbb{N} , die Menge der natürlichen Zahlen,¹ $\mathbb{N} = \{1, 2, 3, \dots\}$,
- \mathbb{N}_0 , die Menge der natürlichen Zahlen einschließlich 0, $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$,
- \mathbb{Z} , die Menge der ganzen Zahlen, $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$,
- \mathbb{Q} , die Menge der rationalen Zahlen,
- \mathbb{R} , die Menge der reellen Zahlen,
- \mathbb{C} , die Menge der komplexen Zahlen.

Diese Zahlensysteme werden Sie genauer in der Analysis besprechen.

Bei Mengen mit unendlich vielen Elementen kann man die Menge natürlich nicht mehr durch eine vollständige Auflistung aller Elemente definieren. Eine andere Möglichkeit, eine Menge zu definieren, besteht darin, die Elementzugehörigkeit durch eine Eigenschaft zu erklären, z.B. können wir die oben eingeführte Menge N_3 auch beschreiben als

$$N_3 = \{n \mid n \in \mathbb{N} \text{ und } 1 \leq n \leq 3\}, \quad (1.4)$$

was wir üblicherweise schreiben würden als

$$N_3 = \{n \in \mathbb{N} \mid 1 \leq n \leq 3\}. \quad (1.5)$$

Wenn man nicht aufpasst, kann die Beschreibung von Mengen anhand der Eigenschaften der Elemente zu Widersprüchen führen. Nehmen wir an, wir könnten die folgende Menge bilden:

$$R = \{M \mid M \text{ Menge und } M \notin M\}, \quad (1.6)$$

also die Menge aller Mengen, die sich nicht selbst als Element enthalten. (Zunächst würde man vielleicht gar nicht auf die Idee kommen, dass eine Menge sich selbst als Element enthalten kann, aber warum nicht?) Ist jetzt R in R enthalten oder nicht?

Angenommen $R \in R$. Dann müsste R die Zugehörigkeitseigenschaft erfüllen, also $R \notin R$. Ist andersherum $R \notin R$, dann müsste R aufgrund seiner Definition R enthalten, also $R \in R$. Wir sehen also:

$$R \in R \iff R \notin R. \quad (1.7)$$

Das kann nicht sein! Bei unserer Annahme, wir könnten Mengen mittels beliebiger Eigenschaften bilden, waren wir zu unvorsichtig und haben einen Widerspruch konstruiert.

Ein Ausweg ist es, eine weitere Struktur zu definieren, sogenannte Klassen. In diesem Formalismus gibt es dann keine Menge aller Mengen, sondern nur die Klasse aller Mengen. Klassen können über Eigenschaften definiert werden und können nur Mengen als

¹Wir verwenden hier die Konvention, dass 0 kein Element der Menge \mathbb{N} ist, so wie es im Buch von Kerner und von Wahl [KvW] gemacht wird, aber abweichend von der Konvention im Buch von Fischer [F]

Elemente enthalten, aber keine (echten) Klassen (Mengen fasst man als spezielle Klassen auf). Dies führt zur sogenannten Zermelo-Fraenkel-Mengenlehre, in der keine solchen Widersprüche gefunden wurden, wie wir sie oben gesehen haben. Wir werden darauf hier nicht weiter eingehen (ich ermutige Sie aber, sich selbst einen Einblick in die axiomatische Mengenlehre zu verschaffen, z.B. in dem Buch von Friedrichsdorf und Prestel [FP] oder auch im Wikipedia-Eintrag unter „Zermelo-Fraenkel-Mengenlehre“).

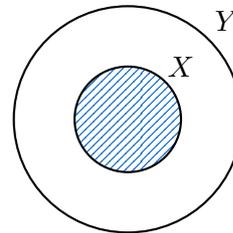
Wir wollen hier diese Axiomatik nicht behandeln (sozusagen das Kleingedruckte), sondern eher einen pragmatischen Standpunkt einnehmen, sollten aber im Hinterkopf haben, dass es bei der Einführung „zu großer“ Mengen, wie der vermeintlichen „Menge aller Mengen“, zu Subtilitäten kommen kann.

Nun wollen wir einige Symbole und Begrifflichkeiten einführen, die wir im Umgang mit Mengen benötigen:

- \emptyset bezeichnet die **leere Menge** (manchmal auch $\{\}$), die keine Elemente enthält.
- $X \subset Y$ bedeutet: X ist **Teilmenge** von Y , d.h. jedes Element von X ist auch ein Element von Y . Formaler können wir diesen Sachverhalt ausdrücken als

$$\forall x \in X : x \in Y, \quad (1.8)$$

„für jedes x aus X gilt: x ist Element von Y “. \forall ist der **Allquantor**, $\forall x$ bedeutet „für alle x “ oder besser „für jedes x “.



- $X \cap Y$ heißt der **Durchschnitt** von X und Y , also die Menge der Elemente, die sowohl in X als auch in Y enthalten sind. Formal können wir schreiben

$$X \cap Y = \{x \mid x \in X \overset{\text{und}}{\wedge} x \in Y\} \quad (1.9)$$

oder

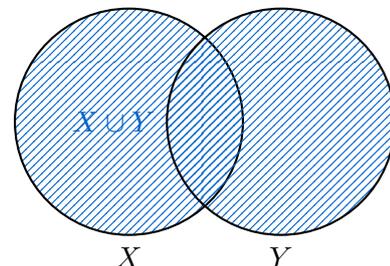
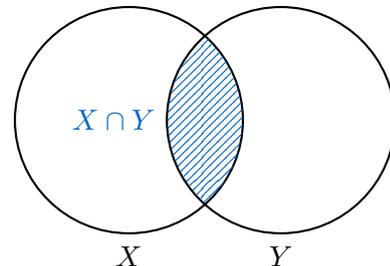
$$X \cap Y = \{x \in X \mid x \in Y\} \quad (1.10)$$

oder

$$X \cap Y = \{y \in Y \mid y \in X\}. \quad (1.11)$$

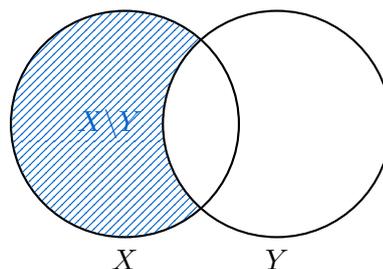
- $X \cup Y$ heißt **Vereinigung** von X und Y , also die Menge der Elemente, die in X oder in Y (kein exklusives oder) enthalten sind. Formal schreibt man

$$X \cup Y = \{x \mid x \in X \overset{\text{oder}}{\vee} x \in Y\}. \quad (1.12)$$



- $X \setminus Y$ heißt „ X ohne Y “ (manchmal auch geschrieben als $X - Y$), also die Menge der Elemente von X , die nicht in Y enthalten sind, wobei wir nicht voraussetzen, dass X Teilmenge von Y ist. Formal schreibt man

$$X \setminus Y = \{x \in X \mid x \notin Y\} \quad (1.13)$$



- $X \times Y$ bezeichnet das **(kartesische) Produkt² von X und Y** . Das ist die Menge der geordneten Paare (x, y) , wobei $x \in X$ und $y \in Y$, also

$$X \times Y = \{(x, y) \mid x \in X \wedge y \in Y\}. \quad (1.14)$$

Beispiel 1.1. Wir betrachten ein Kartenspiel. Die Menge der möglichen Werte der Karten nennen wir W ,

$$W = \{As, 2, 3, 4, 5, 6, 7, 8, 9, 10, Bube, Dame, König\}, \quad (1.15)$$

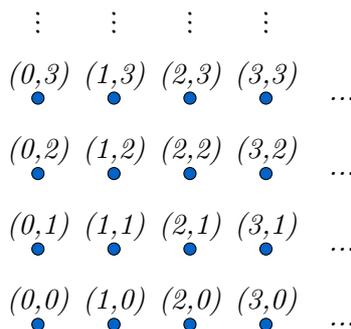
die Menge der Farben F ,

$$F = \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}. \quad (1.16)$$

Jeder Spielkarte ist eine Farbe und ein Wert zugeordnet, also können die Spielkarten durch Paare (Farbe, Wert) charakterisiert werden. Die Menge aller Spielkarten lässt sich somit durch die Produktmenge $F \times W$ beschreiben,

$$F \times W = \{(\heartsuit, As), (\clubsuit, 5), \dots\}. \quad (1.17)$$

Beispiel 1.2. Die Menge $\mathbb{N}_0 \times \mathbb{N}_0 = \{(0, 0), (0, 1), (1, 0), (0, 2), (2, 0), (1, 1), \dots\}$ kann als Gitter dargestellt werden.



²Im Buch von Fischer [F] heißt das Produkt *direktes Produkt*.

Ebenso kann man das Produkt von $n \in \mathbb{N}$ Mengen X_1, X_2, \dots, X_n definieren als die Menge der n -Tupel, oder Folgen der Länge n , (x_1, x_2, \dots, x_n) mit $x_i \in X_i$ für jedes $i = 1, 2, \dots, n$ und $n \geq 1$.

Beispiel 1.3. *Hat man Spielkarten mit verschiedenen Rückseiten, so kann man die Karten durch ein Tripel (Farbe, Wert, Rücken) kennzeichnen, also durch Elemente in $F \times W \times R$ mit der Menge $R = \{\text{blau kariert, rot gestreift, \dots}\}$ der möglichen Rückseiten.*

Auch Durchschnitte und Vereinigungen können wir von mehr als zwei Mengen bilden. Sei I eine nichtleere **Indexmenge** und für jedes $i \in I$ sei X_i eine Menge. Dann ist

$$\bigcap_{i \in I} X_i = \{x \mid \forall i \in I : x \in X_i\} \quad (1.18)$$

$$\bigcup_{i \in I} X_i = \{x \mid \underbrace{\exists i \in I : x \in X_i}_{\text{„es gibt ein } i \in I, \text{ sodass } x \in X_i\text{“}}\}. \quad (1.19)$$

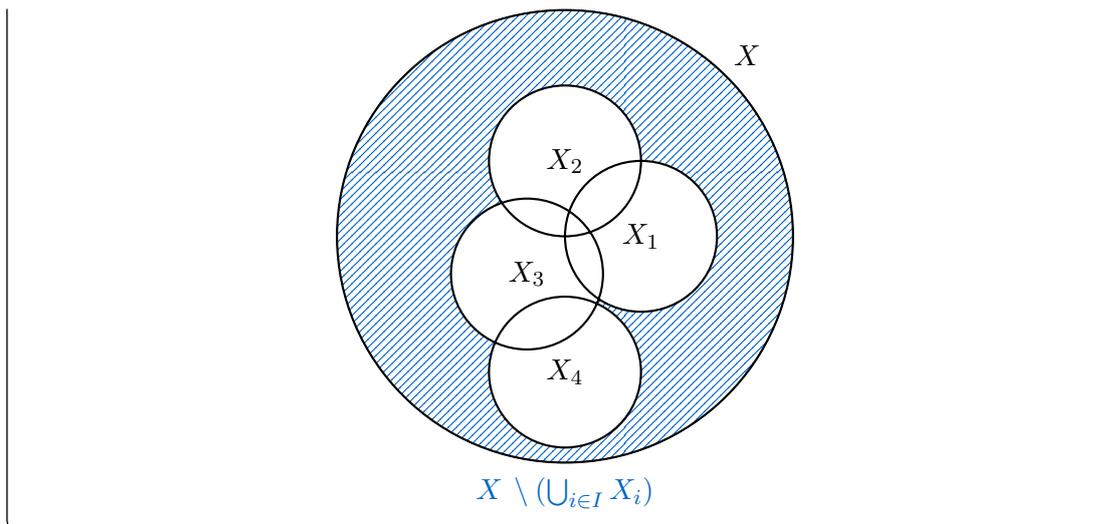
Das Zeichen \exists ist der **Existenzquantor**, $\exists x$ bedeutet „es gibt / es existiert (mindestens) ein x “. Des Weiteren führen wir ein:

$\exists! x$: „es gibt genau ein x “

$\nexists x$: „es gibt kein x “.

Beispiel 1.4. *Als kleine Anwendung der formalen Notation:*

$$X \setminus \left(\bigcup_{i \in I} X_i \right) = \{x \in X \mid \nexists i \in I : x \in X_i\} \quad (1.20)$$



Nun noch ein kleiner Satz:

$$X \setminus (Y \cup Z) = (X \setminus Y) \cap (X \setminus Z). \quad (1.21)$$

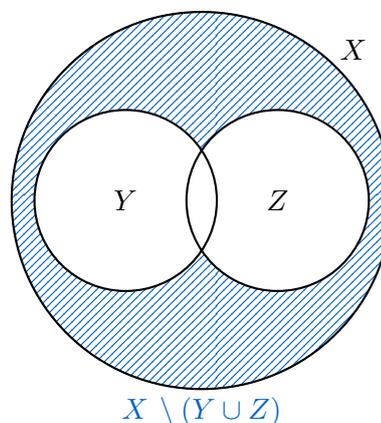
Anschaulich ist das wieder klar, aber zum Spaß wollen wir diese Aussage etwas formaler beweisen.

Wenn wir zeigen wollen, dass zwei Mengen M_1, M_2 gleich sind, müssen wir zeigen, dass gilt:

$$m \in M_1 \iff m \in M_2. \quad (1.22)$$

(Manchmal ist es einfacher, die zwei Aussagen $m \in M_1 \implies m \in M_2$ und $m \in M_2 \implies m \in M_1$ einzeln zu beweisen.)

Versuchen wir es im obigen Fall:



$$x \in X \setminus (Y \cup Z) \iff x \in X \wedge x \notin (Y \cup Z) \quad (1.23)$$

$$\iff x \in X \wedge x \notin Y \wedge x \notin Z \quad (1.24)$$

$$\iff x \in X \setminus Y \wedge x \in X \setminus Z \quad (1.25)$$

$$\iff x \in (X \setminus Y) \cap (X \setminus Z). \quad (1.26)$$

Als weiteres Konzept der Mengenlehre wollen wir die **Potenzmenge** $\mathcal{P}(M)$ einer Menge M einführen als die Menge aller Teilmengen von M ,

$$\mathcal{P}(M) = \{A \mid A \subset M\}. \quad (1.27)$$

Also $A \in \mathcal{P}(M) \iff A \subset M$.

Beispiel 1.5.

$$\mathcal{P}(N_3) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\} \quad (1.28)$$

Falls jemand besorgt ist, das wir schon wieder Mengen von Mengen bilden, so sei er/sie beruhigt: Die Potenzmenge ist ein harmloses Objekt und wirklich eine Menge.

Eine Menge ist eine ungeordnete Zusammenfassung bestimmter Objekte. Manchmal ist es nützlich, weitere Strukturen einzuführen, die eine gewisse Ordnung hineinbringen, z.B. kann man versuchen, bestimmte Objekte, die etwas gemeinsam haben, zusammenzufassen und auf diese Weise eine Menge M in paarweise diskunkte Teilmengen (also Teilmengen, deren paarweiser Durchschnitt leer ist) von „ähnlichen“ Objekten zu zerlegen. Was man genau mit „ähnlich“ meint, hängt davon ab, welche Eigenschaften eines Objekts einen interessieren. Beispielsweise können wir die ganzen Zahlen \mathbb{Z} in die geraden und die ungeraden Zahlen zerlegen,

$$\mathbb{Z} = \underbrace{\{0, 2, -2, 4, -4, \dots\}}_{\mathbb{Z}_{\text{gerade}}} \cup \underbrace{\{1, -1, 3, -3, \dots\}}_{\mathbb{Z}_{\text{ungerade}}} . \quad (1.29)$$

Eine Zerlegung einer Menge M in nichtleere, disjunkte Teilmengen X_i ($i \in I$) nennt man eine Zerlegung in **Äquivalenzklassen**³,

$$M = \bigcup_{i \in I} X_i \quad , \quad \text{für } i \neq j : X_i \cap X_j = \emptyset \quad , \quad \text{für alle } i \in I : X_i \neq \emptyset . \quad (1.30)$$

Man fasst also alle Elemente in einer Teilmenge X_i als *äquivalent* auf, wobei es von der konkreten Situation abhängt, was damit genau gemeint ist: Es gibt keinen universellen Äquivalenzbegriff für Elemente beliebiger Mengen. Um das etwas klarer zu machen, schauen wir uns ein Beispiel an.

Beispiel 1.6. *Interessiert man sich für die Teilbarkeit von Zahlen bezüglich 2, würde man zum einen alle geraden Zahlen und zum anderen alle ungeraden Zahlen als äquivalent ansehen, weil sie jeweils den gleichen Rest bei Division durch 2 haben, hier wäre also wie oben*

$$\mathbb{Z} = \mathbb{Z}_{\text{gerade}} \cup \mathbb{Z}_{\text{ungerade}} \quad (1.31)$$

die für die Fragestellung sinnvolle Zerlegung. Würden wir uns für die Teilbarkeit durch 3 interessieren, so scheint es sinnvoll, alle durch 3 teilbaren Zahlen in eine Menge X_0 zusammenzufassen, alle Zahlen, die bei Division durch 3 Rest 1 haben, in eine Menge X_1 und schließlich alle Zahlen, die bei Division durch 3 Rest 2 besitzen,

³Weiter unten werden wir den Begriff der Äquivalenzklasse ausgehend vom Begriff der Äquivalenzrelation definieren, so wie es üblich ist und beispielsweise im Buch von Fischer [F, Abschnitt 1.1.8] gemacht wird. An dieser Stelle nehmen wir die Zerlegung in Äquivalenzklassen vorweg als Motivation für den Begriff der Äquivalenzrelation.

in eine Menge X_2 :

$$\mathbb{Z} = X_0 \cup X_1 \cup X_2 \quad , \quad X_r = \{n \in \mathbb{Z} \mid n = r \pmod{3}\}. \quad (1.32)$$

Hier sehen wir ganze Zahlen als äquivalent an, wenn sie gleichen Rest bei Division durch 3 besitzen. Man sieht also: Was mit Äquivalenz gemeint ist, hängt von der Situation ab.

Zwei Elemente $a, b \in X_i$ in der gleichen Äquivalenzklasse nennen wir äquivalent und schreiben dafür $a \sim b$. Dies definiert eine Relation auf M , für $a, b \in M$ gilt entweder $a \sim b$ (d.h. sie liegen in der gleichen Äquivalenzklasse) oder $a \not\sim b$ (d.h. sie liegen in verschiedenen Äquivalenzklassen). Statt nun die Zerlegung einer Menge in Äquivalenzklassen anzugeben, ist es oft sinnvoller, die Äquivalenzrelation selbst festzulegen. Dafür müssen wir aber genauer sagen, was wir mit einer Relation meinen:

Definition 1.1. Sei M eine Menge. Wir nennen eine Teilmenge $R \subset M \times M$ eine **Relation** und schreiben

$$\text{für alle } a, b \in M : \quad a \sim b : \iff (a, b) \in R. \quad (1.33)$$

Mit anderen Worten, wir geben mit der Menge R einfach alle Paare an, die über die Relation in Beziehung stehen. Damit wir nicht nur irgend eine Relation, sondern eine Äquivalenzrelation beschreiben, muss diese Relation ein paar weitere Eigenschaften erfüllen:

Definition 1.2. Eine Relation \sim auf der Menge M heißt **Äquivalenzrelation** genau dann, wenn für alle $x, y, z \in M$ gilt

1. $x \sim x$ (d.h. \sim ist reflexiv),
2. $x \sim y \implies y \sim x$ (d.h. \sim ist symmetrisch),
3. $x \sim y$ und $y \sim z \implies x \sim z$ (d.h. \sim ist transitiv).

Wenn $x \sim y$, sagt man: x ist äquivalent zu y .

Wenn man an die Zerlegung in disjunkte Teilmengen denkt, ist es klar, dass eine Äquivalenzrelation diese Eigenschaften erfüllen muss: 1. x ist natürlich in der gleichen Teilmenge wie x , 2. wenn x in der gleichen Teilmenge wie y ist, dann ist y offensichtlich in der gleichen Teilmenge wie x und 3. wenn x und y in der gleichen Teilmenge sind und y und z auch in der gleichen Teilmenge sind, dann sind auch x und z in der gleichen Teilmenge, da ja alle Teilmengen paarweise disjunkt sind. Umgekehrt kann man von der obigen Definition einer Äquivalenzrelation ausgehen und die zugehörigen Äquivalenzklassen definieren, die zu einer disjunkten Zerlegung der Menge führen.

Definition 1.3. Sei \sim eine Äquivalenzrelation auf M . Eine Teilmenge $A \subset M$ heißt **Äquivalenzklasse** (bezüglich der Relation \sim) genau dann, wenn

1. $A \neq \emptyset$
2. $x, y \in A \implies x \sim y$
3. $x \in A, y \in M, x \sim y \implies y \in A$.

Man kann dann überprüfen, dass die Menge M tatsächlich in die so definierten Äquivalenzklassen zerlegbar ist (siehe z.B. [F, Ende von Abschnitt 1.1]), d.h. diese Definition passt mit unserer anfänglich gegebenen Charakterisierung zusammen.

Beispiel 1.7. Im Beispiel 1.6 haben wir \mathbb{Z} in die Mengen X_0, X_1 und X_2 zerlegt, die wir durch den Rest bei Division mit 3 charakterisiert hatten. Die zugehörige Äquivalenzrelation lässt sich einfach formulieren:

$$m \sim n \iff m - n \text{ ist durch } 3 \text{ teilbar.} \quad (1.34)$$

Typischerweise ist die Äquivalenzrelation einfacher zu charakterisieren als die konkrete Zerlegung der gegebenen Menge. Trotzdem ist es sinnvoll, immer im Kopf zu haben, dass die Eigenschaften einer Äquivalenzrelation gerade so gewählt sind, dass sie zu einer disjunkten Zerlegung der Ausgangsmenge führen.

Wenn wir eine Äquivalenzrelation definiert und die Ausgangsmenge in Äquivalenzklassen zerlegt haben, können wir auch die Menge aller Äquivalenzklassen betrachten,

$$M/\sim := \{A \subset M \mid A \text{ ist Äquivalenzklasse bezüglich } \sim\}. \quad (1.35)$$

Wir können uns das so vorstellen, dass wir uns nicht mehr für die einzelnen Elemente interessieren, sondern nur noch für eine bestimmte Eigenschaft der Elemente, und wir identifizieren Elemente, die in dieser Eigenschaft übereinstimmen. Das passiert beispielsweise schon, wenn ein Kleinkind mit Bausteinen spielt und diese nach den Farben sortiert: Alle blauen Steine kommen auf einen Haufen, alle gelben auf den zweiten und die roten auf den dritten Haufen. Die Haufen sind dann die Äquivalenzklassen (bezüglich der durch die Farbe definierten Äquivalenzrelation), und die Menge der Äquivalenzklassen hat dann drei Elemente — in ihr vergessen wir die Form der Bausteine und unterscheiden nur noch nach den Farben, es gibt also eine „blaue“ Klasse, eine „gelbe“ und eine „rote“.

1.2 Abbildungen

Seien X, Y zwei Mengen. Dann nennen wir f eine **Abbildung** von X nach Y , wenn f eine Vorschrift ist, die jedem $x \in X$ genau ein $f(x) \in Y$ zuordnet. Wir schreiben das als

$$f : X \rightarrow Y. \quad (1.36)$$

Die Zuordnungsvorschrift wird geschrieben als

$$f : x \mapsto f(x). \quad (1.37)$$

Man beachte den Extrastrich am Anfang des Pfeils. X heißt der **Definitionsbereich** von f und Y der **Bildbereich** von f .

Beispiel 1.8.

$$f_1 : \mathbb{R} \rightarrow \mathbb{R} \quad \text{Quadrieren}$$

$$x \mapsto x^2$$

$$id_X : X \rightarrow X \quad \text{„Identität“ oder „identische Abbildung“}$$

$$x \mapsto x$$

$$f_2 : \mathbb{N}_0 \times \mathbb{N}_0 \rightarrow \mathbb{N}_0 \quad \text{Addition}$$

$$(x, y) \mapsto x + y$$

$$f_3 : X \rightarrow Y \quad \text{für ein festes } y_0 \in Y : \text{konstante Abbildung}$$

$$x \mapsto y_0$$

Beispiel 1.9. Seien X_1, X_2 Mengen und $X_1 \times X_2$ die Produktmenge. Dann definiert man die Projektionen als Abbildungen

$$\pi_1 : X_1 \times X_2 \rightarrow X_1$$

$$(x_1, x_2) \mapsto x_1$$

$$\pi_2 : X_1 \times X_2 \rightarrow X_2$$

$$(x_1, x_2) \mapsto x_2.$$

Im Beispiel des Kartenspiels $F \times W$ ordnet die Projektion auf den ersten Faktor jeder Karte ihre Spielfarbe zu,

$$\pi_F : F \times W \rightarrow F$$

$$(\text{Farbe}, \text{Wert}) \mapsto \text{Farbe}.$$

$$\pi_F(\diamond, 3) = \diamond$$

Die Projektion auf den zweiten Faktor liefert den Wert,

$$\pi_W : F \times W \rightarrow W$$

$$(\text{Farbe}, \text{Wert}) \mapsto \text{Wert}$$

$$\pi_W(\diamond, 3) = 3$$

Nicht immer werden alle Elemente des Bildbereichs erreicht, die Menge aller Bilder $f(x)$ nennen wir das **Bild** von X , das wir mit $f(X)$ bezeichnen wollen:

$$f(X) := \{y \in Y \mid \exists x \in X : y = f(x)\} \quad (1.38)$$

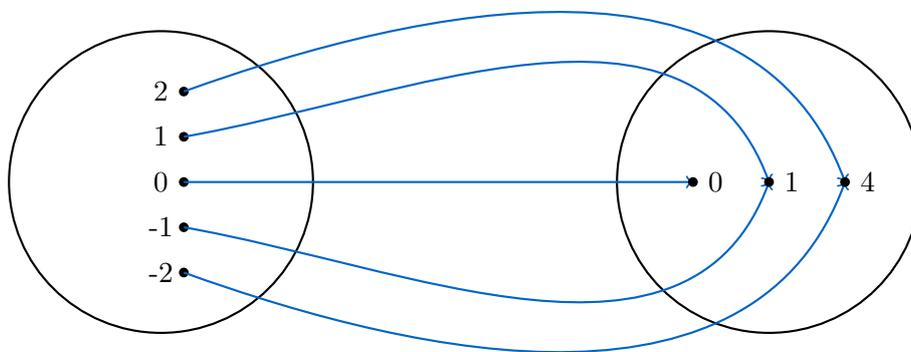
$$\text{oder kurz } f(X) := \{f(x) \mid x \in X\}. \quad (1.39)$$

Im ersten Beispiel ($f_1 : x \mapsto x^2$) ist $f_1(\mathbb{R}) = \{x \in \mathbb{R} \mid x \geq 0\} =: \mathbb{R}_+$.

Für ein gegebenes $y = f(x)$ nennen wir x ein **Urbild** von y (bei f). Für eine Teilmenge $N \subset Y$ definieren wir die Menge der Urbilder⁴ von N als

$$f^{-1}(N) = \{x \in X \mid f(x) \in N\}. \quad (1.40)$$

Im obigen Beispiel ist $f_1^{-1}(\{1, 4\}) = \{-2, -1, 1, 2\}$ und $f_1^{-1}(\{-1\}) = \emptyset$.



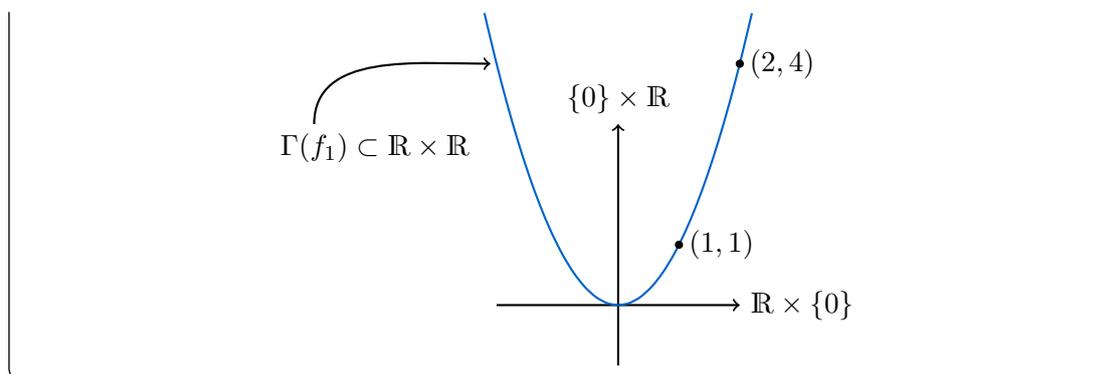
Statt durch eine Vorschrift können wir f auch durch seinen **Graphen** definieren. Dabei ist hierbei nicht unbedingt ein Bild gemeint, sondern eine Teilmenge $\Gamma(f)$ von $X \times Y$:

$$\Gamma(f) := \{(x, y) \in X \times Y \mid y = f(x)\}. \quad (1.41)$$

Beispiel 1.10. Für das Quadrieren sähe das so aus:

$$\Gamma(f_1) = \{(x, x^2) \in \mathbb{R} \times \mathbb{R}\}. \quad (1.42)$$

⁴Im Buch von Kerner und von Wahl [KvW] wird das Urbild von N bei f mit $f^{-1}(N)$ bezeichnet.



Eine Teilmenge G von $X \times Y$ ist genau dann ein Graph einer Abbildung, wenn gilt: Zu jedem $x \in X$ gibt es genau ein $y \in Y$, sodass $(x, y) \in G$, oder formaler:

$$G \subset X \times Y \text{ ist ein Graph} \iff \forall x \in X \exists! y \in Y : (x, y) \in G. \quad (1.43)$$

Abbildungen können sehr unterschiedliche Eigenschaften haben. Wir führen folgende Bezeichnungen für eine Abbildung $f : X \rightarrow Y$ ein:

- f heißt **surjektiv** genau dann, wenn $f(X) = Y$, d.h. wenn jedes Element von Y ein Urbild hat.
- f heißt **injektiv** genau dann, wenn jedes Bild $f(x)$ nur ein Urbild (nämlich x) hat, in anderen Worten, wenn gilt $f(x) = f(x') \implies x = x'$.
- f heißt **bijektiv** genau dann, wenn f injektiv und surjektiv ist, d.h. wenn es zu jedem $y \in Y$ genau ein $x \in X$ gibt mit $y = f(x)$.

Beispiel 1.11.

- $\mathbb{R} \rightarrow \mathbb{R}, x \mapsto x^2$ ist weder injektiv noch surjektiv.
- $\mathbb{R} \rightarrow \mathbb{R}_+, x \mapsto x^2$ ist surjektiv, aber nicht injektiv.
- $\mathbb{R}_+ \rightarrow \mathbb{R}_+, x \mapsto x^2$ ist bijektiv.

Wir sehen, dass die Eigenschaften einer Abbildung nicht nur von der Zuordnungsvorschrift, sondern auch vom Definitions- und Bildbereich abhängen. Man beachte also, dass zur Definition einer Abbildung immer **drei** Dinge gehören:

- Definitionsbereich,
- Bildbereich,
- Zuordnungsvorschrift (oder der Graph).

Beispiel 1.12. Die Projektion $\pi_1 : X_1 \times X_2 \rightarrow X_1, (x_1, x_2) \mapsto x_1$ ist surjektiv, wenn $X_2 \neq \emptyset$.

Wie würde so ein Surjektivitätsbeweis aussehen? Wir müssen zeigen, dass alle Ele-

mente des Bildbereichs erreicht werden, d.h. wir könnten versuchen zu jedem Element des Bildbereichs ein Urbild zu finden. Also so:

Falls $X_1 = \emptyset$ ist nichts zu beweisen. Sei also $x_1 \in X_1$ beliebig. Wähle ein $x_2 \in X_2$ (nach Voraussetzung ist X_2 nicht leer). Dann ist $(x_1, x_2) \in X_1 \times X_2$ ein Urbild von $x_1 \in X_1$ bezüglich $\pi_1: \pi_1(x_1, x_2) = x_1$. Also ist π_1 surjektiv. \square

Beispiel 1.13. Sei $f : X \rightarrow Y$ eine Abbildung. Dann ist die Abbildung

$$F : X \rightarrow X \times Y \quad (1.44)$$

$$x \mapsto (x, f(x)), \quad (1.45)$$

die jedem $x \in X$ den zugehörigen Punkt des Graphen zuordnet, injektiv. Wie würde man das beweisen? Injektivität bedeutet, dass jedes Element des Bildes nur ein Urbild hat. Man fängt also mit zwei Elementen des Definitionsbereichs an, die auf das gleiche Element abgebildet werden, und zeigt, dass sie gleich sind:

Seien $x_1, x_2 \in X$ mit $F(x_1) = F(x_2)$. Nach Definition von F ist dies gleichbedeutend mit

$$(x_1, f(x_1)) = (x_2, f(x_2)) \implies x_1 = x_2. \quad (1.46)$$

Also ist F injektiv. \square

Würde man den Bildbereich von F auf $\Gamma(f)$ einschränken, so wäre F sogar bijektiv.

Für eine bijektive Funktion $f : X \rightarrow Y$ können wir die **inverse Abbildung** (oder **Umkehrabbildung**) $f^{-1} : Y \rightarrow X$ definieren, die jedem y genau das $x \in X$ als Bild $f^{-1}(y)$ zuordnet, für das $y = f(x)$ gilt. Für diese Umkehrabbildung gilt

$$\forall x \in X : f^{-1}(f(x)) = x \quad \text{und} \quad \forall y \in Y : f(f^{-1}(y)) = y. \quad (1.47)$$

Für das Quadrieren $\mathbb{R}_+ \rightarrow \mathbb{R}_+, x \mapsto x^2$ ist offenbar die Umkehrabbildung gegeben durch $\mathbb{R}_+ \rightarrow \mathbb{R}_+, y \mapsto \sqrt{y}$.

Die Eigenschaft der Umkehrabbildung können wir noch eleganter schreiben als

$$f^{-1} \circ f = \text{id}_X \quad \text{und} \quad f \circ f^{-1} = \text{id}_Y, \quad (1.48)$$

wenn wir die **Komposition von Abbildungen** definieren:

Definition 1.4. Seien $f : X \rightarrow Y, g : Y \rightarrow Z$ Abbildungen. Dann definieren wir die **Komposition** von g und f als Abbildung $g \circ f : X \rightarrow Z$ durch die Vorschrift

$$g \circ f : x \mapsto g \circ f(x) = g(f(x)). \quad (1.49)$$

Um uns ein wenig in der Abstraktion zu üben, wollen wir kurz überlegen, was die Komposition für eine Struktur ist. Offensichtlich ist sie eine Vorschrift, die zwei Abbildungen eine neue zugeordnet. Damit ist sie selbst eine Abbildung:

$$\circ : \text{Abb}(Y, Z) \times \text{Abb}(X, Y) \rightarrow \text{Abb}(X, Z) \quad (1.50)$$

$$(g, f) \mapsto g \circ f \quad (1.51)$$

Eine wichtige Eigenschaft von \circ ist die **Assoziativität**:

Satz 1.1. *Die Komposition \circ ist assoziativ, d.h. für Mengen U, V, W, X und die Abbildungen $f : U \rightarrow V$, $g : V \rightarrow W$, $h : W \rightarrow X$ gilt*

$$(h \circ g) \circ f = h \circ (g \circ f). \quad (1.52)$$

Beweis. Sei $u \in U$. Dann ist

$$(h \circ g) \circ f(u) = (h \circ g)(f(u)) = h(g(f(u))) \quad (1.53)$$

und andererseits

$$h \circ (g \circ f)(u) = h(g \circ f(u)) = h(g(f(u))), \quad (1.54)$$

also gilt für jedes u : $(h \circ g) \circ f(u) = h \circ (g \circ f)(u)$. \square

Wenn alle Mengen gleich sind ($X = Y = Z$), definiert die Komposition eine **assoziative Verknüpfung** auf der Menge $\text{Abb}(X, X)$, ganz ähnlich wie beispielsweise die Multiplikation eine assoziative Verknüpfung auf \mathbb{N} darstellt. So wie die 1 das neutrale Element der Multiplikation ist, $1 \cdot n = n \cdot 1 = n$, gibt es auch bezüglich \circ auf $\text{Abb}(X, X)$ ein **neutrales Element**, nämlich die Identität id_X ,

$$\text{für jedes } f \in \text{Abb}(X, X) \text{ ist } f \circ \text{id}_X = f \text{ und } \text{id}_X \circ f = f. \quad (1.55)$$

Für bijektive Abbildungen gibt es sogar ein inverses Element, nämlich die Umkehrabbildung, genauso wie es beispielsweise in den rationalen Zahlen zu jeder Zahl $q \neq 0$ ein multiplikatives Inverses $1/q$ gibt.

Diese Ähnlichkeit in den Strukturen hat die Mathematiker bewogen, diese Strukturen abstrakt zu untersuchen anstelle ihrer konkreten Realisierungen. Dem wenden wir uns als Nächstes zu.

1.3 Algebraische Strukturen: Gruppen, Körper, usw.

Wir wollen uns hier mit den Strukturen beschäftigen, die mit Verknüpfungen zusammenhängen.

Definition 1.5. Eine *Verknüpfung* auf einer Menge X ist eine Abbildung

$$\star : X \times X \rightarrow X \quad (1.56)$$

$$(x, y) \mapsto x \star y. \quad (1.57)$$

Beispiele für Verknüpfungen sind Addition und Multiplikation in $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, Addition und Multiplikation von reellen und komplexen Matrizen (später mehr), Komposition von Abbildungen, Mittelwertbildung $((x, y) \mapsto (x + y)/2), \dots$: immer dann, wenn eine Operation aus zwei Elementen einer Menge ein neues bildet.

Definition 1.6. Eine Verknüpfung $\star : X \times X \rightarrow X$ heißt **assoziativ** genau dann, wenn

$$\forall x, y, z \in X : (x \star y) \star z = x \star (y \star z). \quad (1.58)$$

Die oben genannten Verknüpfungen sind alle assoziativ mit Ausnahme der Mittelwertbildung.

Definition 1.7. Eine Verknüpfung $\star : X \times X \rightarrow X$ heißt **kommutativ** genau dann, wenn

$$\forall x, y \in X : x \star y = y \star x. \quad (1.59)$$

Von den oben genannten Verknüpfungen sind die meisten kommutativ, die Ausnahmen sind Matrizenmultiplikation und Komposition von Abbildungen.

Beispiel 1.14. Sei $f : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto x + 1$ und $g : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto 2x$. Dann ist

$$(f \circ g)(x) = f(g(x)) = f(2x) = 2x + 1, \quad (1.60)$$

aber umgekehrt

$$(g \circ f)(x) = g(f(x)) = g(x + 1) = 2(x + 1). \quad (1.61)$$

Definition 1.8. Wir nennen $e \in X$ ein **neutrales Element** bezüglich der Verknüpfung \star auf X genau dann, wenn

$$\forall x \in X : e \star x = x \text{ und } x \star e = x. \quad (1.62)$$

- 0 ist neutrales Element bezüglich der Addition in $\mathbb{N}_0, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$,

- 1 ist neutrales Element bezüglich der Multiplikation in $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$.
- Entsprechend ist die Nullmatrix neutrales Element der Matrizenaddition, die Einheitsmatrix neutrales Element der Matrizenmultiplikation (das werden wir später noch besprechen).
- Die Identität ist neutrales Element der Komposition von Abbildungen,
- zur Mittelwertbildung gibt es kein neutrales Element.

Die Beispiele im Überblick:

	assoziativ	kommutativ	neutrales Element
Addition in $\mathbb{N}_0, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	✓	✓	0
Multiplikation in $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$	✓	✓	1
Addition von reellen Matrizen	✓	✓	Nullmatrix
Multiplikation von reellen Matrizen	✓	–	Einheitsmatrix
Komposition von Abbildungen	✓	–	Identität
Mittelwertbildung	–	✓	–

In den Beispielen gibt es immer nur ein neutrales Element. Ist das immer so? Ja, es gilt der folgende Satz.

Satz 1.2. *Sei X eine Menge und \star eine Verknüpfung auf X . Seien e und e' neutrale Elemente bezüglich \star . Dann gilt $e = e'$.*

Beweis. Nach Voraussetzung ist e ein neutrales Element, also gilt $e \star e' = e'$. Andererseits ist e' ein neutrales Element, also ist $e \star e' = e$ und somit $e = e'$. \square

Bemerkung

Eine Menge mit assoziativer Verknüpfung wird auch Halbgruppe genannt. Existiert überdies ein neutrales Element, so spricht man von einer **Halbgruppe mit Eins**.

Beispiele für Mengen mit assoziativer Verknüpfung und neutralem Element (Halbgruppe mit Eins) sind

- $\text{Abb}(X, X)$ mit \circ und neutralem Element id_X .
- \mathbb{N}_0 mit $+$ und neutralem Element 0.
- \mathbb{N} mit \cdot und neutralem Element 1.

Wenn es Halbgruppen gibt, liegt es nahe, dass man auch den Begriff der Gruppe definiert hat. Hier fordert man noch etwas mehr Struktur. Um das zu verstehen, schauen wir uns einmal die Lösbarkeit von Gleichungen an. Angenommen $a, b \in X$ sind gegeben und wir suchen $x \in X$, sodass

$$a \star x = b. \tag{1.63}$$

Wären dies gewöhnliche Zahlen mit der gewöhnlichen Multiplikation, dann würden wir

zum Lösen durch a teilen bzw. mit $1/a$ multiplizieren, wenn $a \neq 0$. Dies führt auch im Allgemeinen zum Konzept eines inversen Elements a^{-1} mit $a^{-1} \star a = e$. Wenn es ein solches a^{-1} gibt, können wir x angeben: $x = a^{-1} \star b$.

Dies führt uns zum Begriff der Gruppe, der Ihnen in der Physik immer wieder begegnet wird.

Definition 1.9. Ein Paar (G, \star) bestehend aus einer Menge G mit einer Verknüpfung \star heißt **Gruppe** genau dann, wenn folgende Eigenschaften erfüllt sind:

1. die Verknüpfung \star ist assoziativ,
2. es gibt ein neutrales Element e ,
3. zu jedem $a \in G$ existiert ein inverses Element $a^{-1} \in G$, sodass $a^{-1} \star a = e$ und $a \star a^{-1} = e$.

Ist die Verknüpfung überdies kommutativ, so nennen wir (G, \star) eine **kommutative Gruppe** oder eine **abelsche Gruppe**.

Beispiel 1.15.

- Sei $S(X) \subset \text{Abb}(X, X)$ die Menge der bijektiven Abbildungen. Dann ist $(S(X), \circ)$ eine Gruppe und das zu einer Abbildung f inverse Element ist die Umkehrung f^{-1} . $S(X)$ nennt man die symmetrische Gruppe von X .
- $(\mathbb{Z}, +)$ ist eine abelsche Gruppe.
- $(\mathbb{R} \setminus \{0\}, \cdot)$ ist eine abelsche Gruppe.

In einer Gruppe gibt es zu jedem Element genau ein inverses Element.

Satz 1.3. Sei (G, \star) eine Gruppe und $a \in G$ beliebig. Seien $a', a'' \in G$ zu a inverse Elemente. Dann gilt $a' = a''$.

Beweis. Da a' ein inverses Element zu a ist, gilt $a' \star a = e$ mit dem neutralen Element $e \in G$. Daraus folgt

$$(a' \star a) \star a'' = e \star a'' \quad (1.64)$$

$$\iff a' \star (a \star a'') = e \star a'' \quad \text{Assoziativität} \quad (1.65)$$

$$\iff a' \star e = e \star a'' \quad a'' \text{ ist inverses Element zu } a \quad (1.66)$$

$$\iff a' = a'' \quad e \text{ ist neutrales Element} \quad (1.67)$$

Daher ist es gerechtfertigt, von **dem** inversen Element zu a zu reden. \square

Bemerkung

Wenn wir uns die Definition der Gruppe noch einmal ansehen, so bemerken wir ein

grundsätzliches Schema, das wir häufiger antreffen werden. Zunächst führen wir die Strukturen ein, um die es in der Definition geht (hier die Menge G und die Abbildung $\star : G \times G \rightarrow G$). Das sind sozusagen die Zutaten. Dann folgen die Eigenschaften dieser Strukturen — die Axiome — die wir fordern wollen. Das ist gewissermaßen die Bauanleitung, die uns sagt, was wir mit den Zutaten machen wollen.

Was für Axiome wir fordern, da sind wir im Prinzip frei. Wenn wir uns ohne nachzudenken wilde Axiome ausdenken, kann es aber passieren, dass wir keine besonders interessante mathematische Struktur definieren. Die Definitionen, denen wir hier begegnen, basieren auf unzähligen Erfahrungen; es hat sich gezeigt, dass die so definierten Strukturen besonders interessant und nützlich sind — sie haben sich gewissermaßen bewährt.

Bemerkung

Für eine Gruppe (G, \star) kann man die Axiome reduzieren auf

1. \star ist assoziativ
2. es gibt ein Element $e \in G$, sodass $\forall a \in G$ gilt $e \star a = a$ (Existenz eines „linksneutralen“ Elements) für alle Elemente in G .
3. zu jedem $a \in G$ existiert ein $a' \in G$ mit $a' \star a = e$ (Existenz eines „linksinversen“ Elements) zu jedem Element in G .

Wer mag, kann versuchen zu zeigen, dass aus diesen Axiomen die Eigenschaften in unserer Definition einer Gruppe gefolgert werden können, dass also beide Definitionen äquivalent sind (siehe [F, Abschnitt 1.2.3.]).

Oft haben wir auf einer Menge mehr als eine Verknüpfung definiert, z.B. Addition und Multiplikation auf \mathbb{N} usw. Da auch diese Strukturen recht häufig auftreten, lohnt es, auch diese etwas abstrakter zu betrachten.

Als Prototyp für die nun zu untersuchende Struktur betrachten wir \mathbb{Z} mit der Addition $+$ und Multiplikation \cdot . Bezüglich $+$ bildet \mathbb{Z} eine abelsche Gruppe, bezüglich \cdot eine Halbgruppe. Zudem gibt es Beziehungen zwischen den Operationen,

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c) \tag{1.68}$$

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c), \tag{1.69}$$

die wir **Distributivgesetz** nennen.

Definition 1.10. *Ein Ring (mit Eins) ist ein Tripel $(R, +, \cdot)$ bestehend aus einer Menge R und zwei Verknüpfungen $+$ und \cdot auf R , sodass gilt*

1. $(R, +)$ ist eine abelsche Gruppe
2. \cdot ist eine assoziative Verknüpfung mit neutralem Element

3. es gelten die Distributivgesetze:

$$\forall a, b, c \in R: \quad a \cdot (b + c) = (a \cdot b) + (a \cdot c) \quad (1.70)$$

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c). \quad (1.71)$$

Beispiel 1.16. Ein Beispiel haben wir schon gesehen. $(\mathbb{Z}, +, \cdot)$ ist sogar ein **kommutativer Ring**, weil die Multiplikation kommutativ ist.

Ein anderes Beispiel ist der Polynomring. Sei $\mathbb{R}[t]$ die Menge der Polynome in einer Unbestimmten t mit reellen Koeffizienten,

$$\mathbb{R}[t] = \{a_0 + a_1t + a_2t^2 + \dots + a_nt^n \mid a_0, a_1, \dots, a_n \in \mathbb{R}, n \in \mathbb{N}_0\}. \quad (1.72)$$

Polynome können wir addieren und multiplizieren, d.h. wir können zwei Verknüpfungen, $+$ und \cdot , auf $\mathbb{R}[t]$ definieren. $(\mathbb{R}[t], +, \cdot)$ ist dann ein kommutativer Ring: Das neutrale Element der Addition ist das Nullpolynom, $p = 0$, das neutrale Element der Multiplikation ist das konstante Polynom mit Koeffizienten 1, $p = 1$.

Betrachten wir noch einmal einfache Gleichungen. Wenn wir zwei Verknüpfungen, $+$ und \cdot , zur Verfügung haben, können wir die Gleichung

$$a \cdot x + b = c \quad (1.73)$$

betrachten. a, b, c sind hier fest aus der Grundmenge gewählt und x ist gesucht. In einem ersten Schritt können wir das zu b bezüglich $+$ inverse Element $-b$ addieren,

$$a \cdot x = c + (-b). \quad (1.74)$$

Um nach x auflösen zu können, würden wir jetzt gerne mit dem zu a bezüglich \cdot inversen Element a^{-1} multiplizieren, wenn es existiert. Dann wäre die Lösung

$$x = a^{-1} \cdot (c + (-b)). \quad (1.75)$$

Zum Lösen von Gleichungen wäre es also wünschenswert, wenn jedes Element auch ein multiplikatives Inverses hätte, und wir könnten auf die Idee kommen, die Definition eines Rings so zu erweitern, dass die Menge auch bezüglich der Verknüpfung \cdot eine Gruppe bildet. Dies hätte aber unerwünschte Nebeneffekte, wie wir gleich einsehen werden.

Satz 1.4. Sei $(R, +, \cdot)$ ein Ring (mit Eins). Sei 0 das neutrale Element bezüglich $+$, dann gilt

$$\text{für alle } r \in R: \quad 0 \cdot r = 0 \text{ und } r \cdot 0 = 0. \quad (1.76)$$

Beweis. Wir nutzen das Distributivgesetz:

$$0 \cdot r + 0 \cdot r = (0 + 0) \cdot r = 0 \cdot r, \quad (1.77)$$

da 0 das neutrale Element bzgl. $+$ ist. Addition von $-(0 \cdot r)$ führt dann zu

$$0 \cdot r = 0. \quad (1.78)$$

Analog schließt man $r \cdot 0 = 0$. \square

Angenommen, es gäbe zu jedem Element ein multiplikatives Inverses, also auch 0^{-1} zu 0, dann wäre

$$0 \cdot 0^{-1} = 1, \quad (1.79)$$

aber wegen des obigen Satzes auch

$$0 \cdot 0^{-1} = 0, \quad (1.80)$$

und somit $0 = 1$, die neutralen Elemente würden also übereinstimmen. Sei $r \in R$ beliebig, dann gilt: $r = 1 \cdot r = 0 \cdot r = 0$. Ein Ring, in dem jedes Element ein multiplikatives Inverses hätte, besäße nur ein Element!

Damit hätten wir also keine besonders interessante Struktur definiert. Das Beste, was wir versuchen können, ist, dass alle Elemente mit Ausnahme von 0 ein multiplikatives Inverses besitzen (und das wäre uns aus \mathbb{Q} oder \mathbb{R} auch vertraut). Eine solche Struktur nennt man einen Körper, genauer:

Definition 1.11. Ein Tripel $(K, +, \cdot)$ aus einer Menge K und zwei Verknüpfungen $+$ und \cdot auf K heißt **Körper** genau dann, wenn gilt

1. $(K, +)$ ist eine abelsche Gruppe,
2. $(K \setminus \{0\}, \cdot)$ ist eine abelsche Gruppe (wobei 0 das neutrale Element bzgl. $+$ ist),
3. es gilt das Distributivgesetz: $\forall a, b, c \in K : a \cdot (b + c) = a \cdot b + a \cdot c$.

Bemerkung

Beachten Sie, dass wir für einen Körper voraussetzen, dass die Multiplikation **kommutativ** ist. Daher müssen wir auch nur eine Variante der Distributivgesetze fordern.

Beispiel 1.17. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ mit der gewöhnlichen Addition und Multiplikation sind Körper.

Der Körper ist die für alles Folgende grundlegende Struktur. Wir werden hauptsächlich mit dem vertrauten Körper \mathbb{R} (bzw. $(\mathbb{R}, +, \cdot)$, aber das werden wir nicht mehr ausschreiben) und später mit dem Körper \mathbb{C} der komplexen Zahlen arbeiten. Die meisten der von uns betrachteten Strukturen können aber leicht auf beliebige Körper ausgeweitet werden.

Zusammenfassung Kapitel 1

Nach diesem Exkurs in abstrakte Strukturen werden wir uns im Folgenden wieder mit sehr konkreten Strukturen beschäftigen, die später wieder abstrahiert werden. Wir wollen aber noch einmal kurz zusammenfassen, was Sie aus diesem Kapitel mitgenommen haben sollten:

- Sie sind mit dem Begriff der Menge und den Mengenoperationen ($\cup, \cap, \setminus, \times$) vertraut.
- Sie können mit dem Begriff der Abbildung umgehen, wissen, wie man eine Abbildung aufschreibt, und sind sich bewusst, dass zur Charakterisierung drei Dinge gehören (welche?).
- Sie wissen, was injektiv, surjektiv und bijektiv bedeutet (versuchen Sie es einmal aus dem Gedächtnis zu definieren). Welche Eigenschaften braucht man für eine Abbildung, damit eine Umkehrabbildung existiert?
- Sie kennen den Begriff der Verknüpfung und die wichtigsten Eigenschaften, die eine Verknüpfung haben kann: Assoziativität, Existenz eines neutralen Elements, Existenz eines Inversen, Kommutativität.
- Sie erinnern sich, welche drei Eigenschaften man braucht, um eine Gruppe zu definieren. Welche kommt bei einer abelschen Gruppe hinzu?
- Bei mehr als einer Verknüpfung kommt die Frage auf, inwieweit diese verträglich sind. Sie kennen das Distributivgesetz, das eine solche Verträglichkeit regelt.
- Sie wissen, dass bei einem Körper zwei Verknüpfungen $(+, \cdot)$ definiert sind, und können die Körperaxiome in drei Typen einteilen: die Eigenschaften von $+$, von \cdot und ihre Verträglichkeit.
- Wenn Sie nach einem Beispiel für einen Körper gefragt werden, können Sie ohne zu zögern antworten: z.B. „ \mathbb{R} !“ oder die Präzisen unter Ihnen „ $(\mathbb{R}, +, \cdot)$!“.

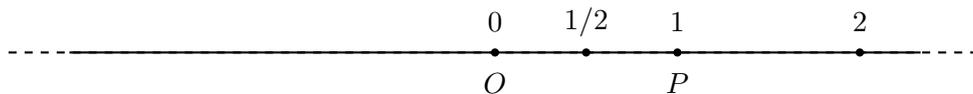
Und genau mit \mathbb{R} machen wir jetzt auch im folgenden Kapitel weiter.

2 Geometrie im \mathbb{R}^n

2.1 Algebraisierung der Geometrie

Für die Physik ist es wichtig, die beobachteten Strukturen in mathematische Konzepte zu übertragen, also zu formalisieren. Als Beispiel nehmen wir den uns umgebenden Raum, bzw. den idealisierten Anschauungsraum.

Der Einfachheit halber beginnen wir mit dem eindimensionalen Raum, den wir uns idealisiert als Linie unendlicher Länge vorstellen. Wie führen wir hierauf eine algebraische Struktur ein, mit der wir rechnen können? Als Physikerinnen und Physiker würden wir zunächst einmal einen Punkt markieren, unser Ausgangspunkt, den wir O (für „origin“) nennen wollen. Des Weiteren benötigen wir einen Maßstab, dafür markieren wir einen weiteren Punkt P , dem Abstand von O zu P geben wir den Wert 1. Damit sind wir in der Lage, jedem Punkt den Abstand zu O zuzuordnen. Physikalisch können wir uns das so vorstellen, dass wir einen Faden von O nach P spannen, unser Maßband. Wenn wir dieses Maßband an P anlegen, erreichen wir den Punkt, dessen Abstand zu O wir 2 nennen wollen. Halbieren wir den Faden (falten ihn in der Mitte) und legen ihn an O an, so erreichen wir den Punkt mit Abstand $1/2$. Durch immer genaueres Abmessen können wir (zumindest im idealisierten Raum) jedem Punkt eine reelle Zahl zuordnen, wobei negative Zahlen Punkte auf der zu P entgegengesetzten Seite bezeichnen.



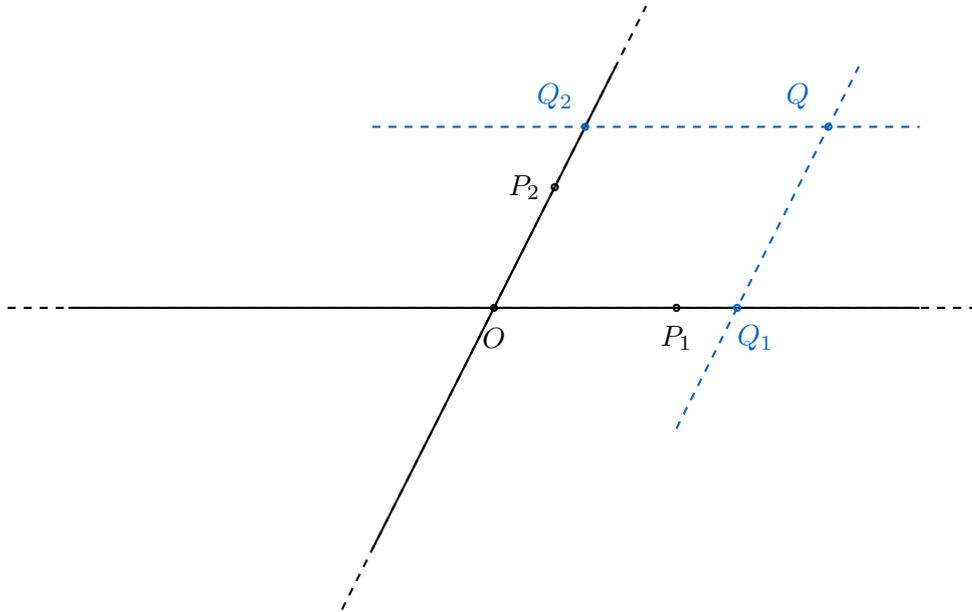
Damit haben wir unser Ziel erreicht, den eindimensionalen physikalischen Raum (bzw. Anschauungsraum) in eine algebraische Struktur zu übersetzen. Geometrische Operationen, wie das Verschieben oder das Aneinanderlegen von Strecken, können wir mit Hilfe der auf \mathbb{R} definierten Addition in algebraische Operationen übersetzen.

Diese Algebraisierung der Geometrie können wir in höheren Dimensionen fortsetzen. Betrachten wir zwei Dimensionen. Wieder markieren wir einen Ursprung O . Wir markieren einen weiteren Punkt P_1 und können dann wie im obigen Beispiel eine Gerade durch reelle Zahlen beschreiben. Dies ist offenbar nicht genug, und wir führen einen weiteren Punkt P_2 ein (der nicht auf der durch O und P_1 bestimmten Geraden liegt). Damit definieren wir eine weitere Gerade durch O und P_2 , und wir können auch Punkten auf dieser Gerade eine Zahl zuordnen (wobei wir im Moment keine Aussage darüber machen wollen, ob der physikalische Abstand zwischen O und P_1 etwas mit dem physikalischen Abstand zwischen O und P_2 zu tun hat).

Um auch anderen Punkten Zahlen zuordnen zu können, nutzen wir aus, dass wir in der Geometrie den Begriff der Parallelität haben, wir also sagen können, wann zwei Geraden parallel sind. Das ist aus dem physikalischen Raum her anschaulich, wobei auch klar ist, dass man hier idealisiert und die Erfahrung, die man auf endlichem Raum

mit endlicher Messgenauigkeit gewonnen hat, stillschweigend auf den unendlichen Raum ausdehnt (und in der Tat wird in der Allgemeinen Relativitätstheorie der Raum nicht mehr durch eine flache Geometrie beschrieben). Das soll uns hier nicht davon abhalten, die Geometrie im zweidimensionalen Anschauungsraum voranzutreiben, aber man sollte sich immer darüber klar werden, welche Annahmen bzw. Idealisierungen man macht.

Mit dem Begriff der Parallelität können wir wie folgt vorgehen



Zu einem Punkt Q der Ebene zeichnen wir die zwei Geraden, die parallel liegen zu den beiden Geraden, die durch O und P_1 und durch O und P_2 festgelegt sind. Den Schnittpunkten Q_1 und Q_2 können wir auf gewohnte Weise jeweils eine reelle Zahl, x_1 und x_2 , zuordnen. Damit haben wir eine eindeutige Zuordnung zwischen Punkten auf der Ebene und reellen Zahlenpaaren gewonnen, die wir **Koordinaten** nennen. Unsere Algebraisierung führt also zu einer Identifikation zwischen der Ebene und dem

$$\mathbb{R}^2 := \mathbb{R} \times \mathbb{R} = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mid x_1, x_2 \in \mathbb{R} \right\}, \quad (2.1)$$

wobei wir die Koordinaten in diesem Fall senkrecht anordnen wollen.

Auch geometrische Operationen wollen wir in algebraische Operationen übersetzen. Ein Punkt Q mit dem Zahlenpaar $\begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$ definiert eine gerichtete Strecke \overrightarrow{OQ} von O nach Q . Durch Parallelverschiebung können wir diese Strecke an jedem beliebigen Punkt P ansetzen lassen, wobei der Endpunkt jetzt einen neuen Punkt P' definiert. Diese mit Hilfe der Strecke \overrightarrow{OQ} definierte Operation, $P \mapsto P'$, hat folgende algebraische Charakterisierung: Hat der Punkt P die Koordinaten $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, so hat der Punkt P' die Koordinaten $\begin{pmatrix} x_1 + q_1 \\ x_2 + q_2 \end{pmatrix}$. Das Anlegen von gerichteten Strecken entspricht der natürlichen additiven Struktur auf \mathbb{R}^2 , die wir gleich allgemeiner für \mathbb{R}^n definieren wollen.

Definition 2.1. Wir nennen die durch

$$+ : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad (2.2)$$

definierte Verknüpfung die **Addition** von reellen n -Tupeln.

Die gerichteten Strecken bzw. die zugehörigen reellen n -Tupel wollen wir auch Vektoren nennen; die obige Verknüpfung heißt dann auch Vektoraddition.

Satz 2.1. $(\mathbb{R}^n, +)$ ist eine abelsche Gruppe.

Beweis. Der Beweis sei Ihnen überlassen. □

Das neutrale Element $0 = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ nennen wir auch **Nullvektor**.

Reelle Zahlen können wir auch multiplizieren. Welche Bedeutung hat dies geometrisch? Schon im eindimensionalen Fall haben wir diskutiert, wie wir einer Strecke die doppelte Strecke oder auch die halbe Strecke zuordnen können. Dies lässt sich erweitern auf eine beliebige Skalierung einer Strecke unter Beibehaltung der Richtung.

Zurück zur Ebene: Bezeichnet $\begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$ die Strecke von O nach Q , so bezeichnet $\begin{pmatrix} \lambda q_1 \\ \lambda q_2 \end{pmatrix}$ die mit $\lambda \in \mathbb{R}$ skalierte Strecke (mit gleichem Ursprung). Einem negativen λ ordnen wir die entsprechend an O gespiegelte Strecke zu.

Damit haben wir eine Operation eingeführt, die wir gleich auf \mathbb{R}^n verallgemeinern wollen.

Definition 2.2. Die durch

$$\cdot : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \lambda \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix} \quad (2.3)$$

definierte Operation nennen wir **Skalarmultiplikation**.

Ein paar Eigenschaften folgen direkt aus den Eigenschaften der Addition und Multipli-

kation von reellen Zahlen:

$$\forall \lambda, \mu \in \mathbb{R}, v \in \mathbb{R}^n : \underbrace{(\lambda + \mu)}_{+ \text{ in } \mathbb{R}} \cdot v = (\lambda \cdot v) \underbrace{+}_{+ \text{ in } \mathbb{R}^n} (\mu \cdot v) \quad (2.4)$$

$$\forall \lambda \in \mathbb{R}, v, w \in \mathbb{R}^n : \lambda \cdot \underbrace{(v + w)}_{+ \text{ in } \mathbb{R}^n} = (\lambda \cdot v) \underbrace{+}_{+ \text{ in } \mathbb{R}^n} (\lambda \cdot w) \quad (2.5)$$

$$\forall \lambda, \mu \in \mathbb{R}, v \in \mathbb{R}^n : \lambda \cdot (\mu \cdot v) = (\lambda \cdot \mu) \cdot v \quad (2.6)$$

$$\forall v \in \mathbb{R}^n : 1 \cdot v = v \quad (2.7)$$

$$\forall v \in \mathbb{R}^n : \underbrace{0}_{\text{reelle Zahl } 0} \cdot v = \underbrace{0}_{\text{Nullvektor}} \quad (2.8)$$

$$\forall v \in \mathbb{R}^n : (-1) \cdot v = -v. \quad (2.9)$$

Bemerkung

Wir bezeichnen den Nullvektor mit dem gleichen Symbol 0 wie die reelle Zahl 0, in der Physik findet man auch manchmal die Notation $\vec{0}$. Die Addition in \mathbb{R}^n und in \mathbb{R} bezeichnen wir beide mit dem gleichen Symbol +, und auch die Multiplikation in \mathbb{R} und die Skalarmultiplikation auf \mathbb{R}^n werden mit dem gleichen Symbol \cdot gekennzeichnet. Das vereinfacht einerseits die Notation und ist andererseits praktisch, da die Eigenschaften (2.4) und (2.5) die gleiche Form wie ein Distributivgesetz bekommen. Man sollte sich dennoch immer bewusst sein, was +, \cdot oder 0 im Kontext jeweils bedeuten.

Wir werden auch manchmal das Symbol \cdot für die Skalarmultiplikation ganz weglassen, und wir wollen die Ihnen vertraute Konvention einführen, dass die multiplikative Operation vor der additiven Operation ausgeführt wird, sodass wir ein paar Klammern weglassen können; wir schreiben z.B. $\lambda v + w$ statt $(\lambda \cdot v) + w$.

Die Skalarmultiplikation erlaubt es, den geometrischen Begriff der Gerade in ein algebraisches Konzept zu übersetzen.

Definition 2.3. Seien $v, t \in \mathbb{R}^2$ und $t \neq 0$. Dann nennen wir

$$G_{v,t} := \{w \in \mathbb{R}^2 \mid \exists \lambda \in \mathbb{R} : w = v + \lambda \cdot t\} =: v + \mathbb{R} \cdot t \quad (2.10)$$

eine **Gerade** durch v mit Richtungsvektor t .

Haben wir die geometrischen Objekte algebraisiert, können wir auf rein algebraischem Wege Aussagen über diese Objekte ableiten. Diese Aussagen können wir dann wieder geometrisch interpretieren. In der Physik muss man die so erhaltenen Aussagen dann natürlich wieder an der Beobachtung überprüfen, es könnte ja sein, dass wir bei der Formalisierung zu unvorsichtig waren bzw. bei der Idealisierung zu übereifrig.

Wir wollen jetzt folgende geometrisch sehr offensichtliche Aussage rein algebraisch beweisen:

Satz 2.2. Sind $v_1, v_2 \in \mathbb{R}^2$ verschiedene Punkte, so gibt es genau eine Gerade durch v_1 und v_2 .

Beweis. v_1 und v_2 liegen auf der Geraden $G_{v_1, v_2 - v_1}$, denn $v_1 = v_1 + 0 \cdot (v_2 - v_1)$ und $v_2 = v_1 + 1 \cdot (v_2 - v_1)$.

Sei $G_{v,t}$ eine weitere Gerade durch v_1 und v_2 , d.h. es gibt $\lambda_1, \lambda_2 \in \mathbb{R}$ mit $v_1 = v + \lambda_1 \cdot t$, $v_2 = v + \lambda_2 \cdot t$, also auch $v_1 - v_2 = (\lambda_1 - \lambda_2) \cdot t$ und damit $t = \mu \cdot (v_2 - v_1)$ mit $\mu = 1/(\lambda_2 - \lambda_1)$ (da v_1 und v_2 verschieden sind, ist $\lambda_1 - \lambda_2 \neq 0$).

Sei $v + \lambda \cdot t$ ein beliebiges Element von $G_{v,t}$. Dann gilt (man beachte, dass $v = v_1 - \lambda_1 \cdot t$)

$$v + \lambda \cdot t = v_1 + (\lambda - \lambda_1) \cdot t = v_1 + (\lambda - \lambda_1)\mu \cdot (v_2 - v_1) \in G_{v_1, v_2 - v_1}. \quad (2.11)$$

Sei andersherum $v_1 + \lambda \cdot (v_2 - v_1)$ ein beliebiges Element von $G_{v_1, v_2 - v_1}$. Dann ist

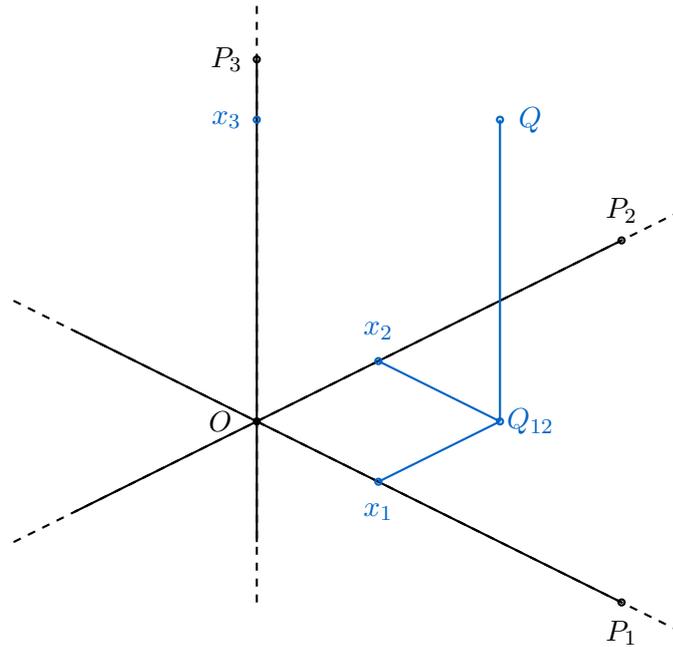
$$v_1 + \lambda \cdot (v_2 - v_1) = (v + \lambda_1 \cdot t) + \lambda \cdot \frac{1}{\mu} \cdot t = v + \left(\lambda_1 + \frac{\lambda}{\mu}\right) \cdot t \in G_{v,t}. \quad (2.12)$$

Damit gilt also $G_{v,t} = G_{v_1, v_2 - v_1}$. □

Diese Aussage gilt natürlich analog in jedem \mathbb{R}^n .

Nachdem wir uns eine Weile mit der (algebraischen) Ebene beschäftigt haben, wenden wir uns jetzt dem dreidimensionalen Raum zu. Wie vorher wählen wir einen Ursprung O und nun drei Punkte P_1, P_2, P_3 , sodass die Geraden durch $\overrightarrow{OP_1}, \overrightarrow{OP_2}$ und $\overrightarrow{OP_3}$ die Koordinatenachsen werden. Hierbei müssen wir nicht nur darauf achten, dass diese Geraden nicht übereinander liegen, sondern auch, dass z.B. die Gerade durch $\overrightarrow{OP_3}$ nicht in der durch $\overrightarrow{OP_1}$ und $\overrightarrow{OP_2}$ definierten Ebene liegt.

Auf analoge Art wie vorher ordnen wir einem Punkt Q drei reelle Zahlen zu, seine Koordinaten $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$. Dazu verschieben wir die durch $\overrightarrow{OP_3}$ bestimmte Gerade so, dass sie durch Q läuft. Sie schneidet dann die durch $\overrightarrow{OP_1}$ und $\overrightarrow{OP_2}$ festgelegte Ebene in einem Punkt Q_{12} , dessen Koordinaten in der Ebene x_1 und x_2 sind. Die Strecke $\overrightarrow{Q_{12}Q}$ wiederum ist parallel zu $\overrightarrow{OP_3}$ und kann daher auf die durch $\overrightarrow{OP_3}$ laufende Gerade verschoben werden mit O als Startpunkt. Dem Endpunkt wird dann die Koordinate x_3 zugeordnet. Die folgende Illustration soll das verdeutlichen:



Damit ist der dreidimensionale Raum mit \mathbb{R}^3 identifiziert.

Zusätzlich zu Geraden können wir nun auch Ebenen im \mathbb{R}^3 einführen.

Definition 2.4. Seien $v, t_1, t_2 \in \mathbb{R}^3$ mit $t_1 \neq 0$, $t_2 \neq 0$ und t_1 und t_2 nicht parallel (d.h. es gibt kein $\lambda \in \mathbb{R}$ mit $t_1 = \lambda t_2$). Dann nennen wir

$$E_{v,t_1,t_2} := \{w \mid \exists \lambda_1, \lambda_2 \in \mathbb{R} : w = v + \lambda_1 t_1 + \lambda_2 t_2\} = v + \mathbb{R} \cdot t_1 + \mathbb{R} \cdot t_2 \quad (2.13)$$

die durch t_1 und t_2 aufgespannte **Ebene** durch v .

Wir bemerken in der Definition einige Verrenkungen, die richtigen Forderungen an t_1 und t_2 zu stellen, damit wirklich eine Ebene herauskommt und keine Gerade oder womöglich nur ein Punkt. Wir wollen, dass die Vektoren unabhängige Richtungen definieren, und in der Tat nennt man zwei Vektoren, von denen keiner ein reelles Vielfaches des anderen ist, **linear unabhängig**.

Auch wenn unsere Anschauung bei vier- oder höherdimensionalen Räumen versagt, so spricht in der algebraischen Beschreibung nichts dagegen, auch den \mathbb{R}^4 , den \mathbb{R}^5 , usw. zu betrachten. Im \mathbb{R}^4 können wir dann — neben Geraden und Ebenen — auch dreidimensionale Unterräume betrachten. Diese werden nun durch drei Vektoren t_1, t_2, t_3 bestimmt, die wieder unabhängige Richtungen beschreiben sollen: Keiner der Vektoren soll in der durch die anderen aufgespannten Ebene liegen.

Angenommen, dies ist nicht gegeben und z.B. t_3 liegt in der durch t_1 und t_2 aufgespann-

ten Ebene (durch O). Dann gibt es λ_1, λ_2 , sodass

$$t_3 = \lambda_1 t_1 + \lambda_2 t_2. \quad (2.14)$$

Einen Ausdruck der Form $\lambda_1 t_1 + \lambda_2 t_2$ nennt man **Linearkombination** der Vektoren t_1 und t_2 . Die Vektoren t_1, t_2, t_3 definieren also unabhängige Richtungen, wenn sich keiner der Vektoren als Linearkombination der anderen schreiben lässt. Den Satz der Vektoren t_1, t_2, t_3 nennt man dann **linear unabhängig** und in dieser Form lässt sich lineare Unabhängigkeit auch für beliebig viele Vektoren in beliebigen Räumen \mathbb{R}^n definieren: Ein Satz von Vektoren ist linear unabhängig, wenn sich keiner der Vektoren als Linearkombination der anderen ausdrücken lässt.

Wir wollen die Definition in etwas anderer Form (äquivalent, aber einfacher anzuwenden) festhalten:

Definition 2.5. Eine Familie (v_1, \dots, v_m) von Vektoren im \mathbb{R}^n heißt genau dann **linear unabhängig**, wenn die Gleichung

$$\lambda_1 v_1 + \dots + \lambda_m v_m = 0 \quad (2.15)$$

nur die Lösung $\lambda_1 = \dots = \lambda_m = 0$ hat. Sie heißt **linear abhängig** genau dann, wenn es auch eine andere Lösung gibt.

Definition 2.6. Ein Vektor $v \in \mathbb{R}^n$ heißt Linearkombination der Vektoren $v_1, \dots, v_r \in \mathbb{R}^n$ genau dann, wenn es Koeffizienten $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ gibt mit $v = \lambda_1 v_1 + \dots + \lambda_r v_r = \sum_{i=1}^r \lambda_i v_i$.

Bemerkung

in einer linear abhängigen Familie (v_1, \dots, v_m) lässt sich mindestens einer der Vektoren v_i als Linearkombination der anderen ausdrücken: Sei $(\lambda_1, \dots, \lambda_m)$ eine nichttriviale Lösung von $\lambda_1 v_1 + \dots + \lambda_m v_m = 0$, d.h. mindestens ein λ_i ist von Null verschieden. Dann ist

$$v_i = -\frac{1}{\lambda_i} (\lambda_1 v_1 + \dots + \lambda_{i-1} v_{i-1} + \lambda_{i+1} v_{i+1} + \dots + \lambda_m v_m), \quad (2.16)$$

also ist v_i eine Linearkombination der anderen Vektoren.

Ist andersherum ein Vektor (z.B. v_1) eine Linearkombination der anderen, $v_1 = \mu_2 v_2 + \dots + \mu_m v_m$, dann folgt $(-1) \cdot v_1 + \mu_2 v_2 + \dots + \mu_m v_m = 0$, d.h. die Vektoren sind linear abhängig nach der obigen Definition.

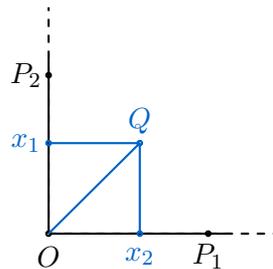
Das bedeutet auch, dass die Definition linearer Unabhängigkeit mit unserer ersten Charakterisierung („keiner der Vektoren ist eine Linearkombination der anderen“) übereinstimmt.

Wir werden den Begriff der linearen Unabhängigkeit später bei der Behandlung allgemeiner Vektorräume noch einmal aufgreifen.

In diesem ersten Abschnitt dieses Kapitels haben wir gesehen, wie wir zum einen die anschaulich vertraute Geometrie in algebraische Konzepte übersetzen können, und wie wir zum anderen mit Hilfe dieser Konzepte spielend Räume beliebiger Dimension erfassen können, wo unsere Anschauung typischerweise versagt. Wir werden jetzt diskutieren, wie wir weitere geometrische Konzepte in algebraische übersetzen können.

2.2 Abstände und Winkel: Skalarprodukt

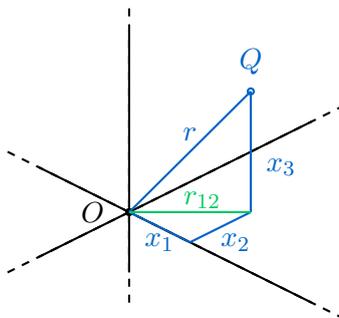
Wir haben im letzten Abschnitt den Anschauungsraum mit dem \mathbb{R}^n identifiziert; dabei haben wir ausgenutzt, dass wir Strecken parallel verschieben können. Wir können aber noch mehr: Insbesondere können wir Strecken auch drehen und damit können wir Abstände in verschiedenen Richtungen miteinander in Beziehung setzen. Wir können dann z.B. in der Ebene nach Wahl von O und P_1 verlangen, dass P_2 den gleichen Abstand zu O hat wie P_1 . Wir können sogar verlangen, dass $\overrightarrow{OP_2}$ senkrecht auf $\overrightarrow{OP_1}$ steht, beispielsweise können wir fordern, dass P_2 gleichen Abstand hat zu P_1 und zu dem an O gespiegelten P_1 . Wir erhalten dann ein rechtwinkliges, normiertes Koordinatensystem.



Aus der Schule wissen Sie auch, wie in einem so gewählten Koordinatensystem der Abstand r eines Punktes Q mit Koordinaten (x_1, x_2) vom Ursprung O bestimmt wird, nämlich mit Hilfe des Satzes von Pythagoras:

$$r^2 = x_1^2 + x_2^2. \quad (2.17)$$

Das funktioniert analog im dreidimensionalen Raum:



$$r^2 = x_3^2 + r_{12}^2 \quad (2.18)$$

$$r_{12}^2 = x_1^2 + x_2^2 \quad (2.19)$$

$$\implies r^2 = x_1^2 + x_2^2 + x_3^2 \quad (2.20)$$

Damit haben wir eine weitere Struktur im geometrischen Anschauungsraum definiert: Wir können Abstände zwischen Punkten bestimmen, insbesondere können wir einem Vektor eine Länge zuordnen, ein Konzept, das wir gleich auf den \mathbb{R}^n erweitern:

Definition 2.7. Die Länge (oder **Norm**) eines Vektors $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ ist defi-

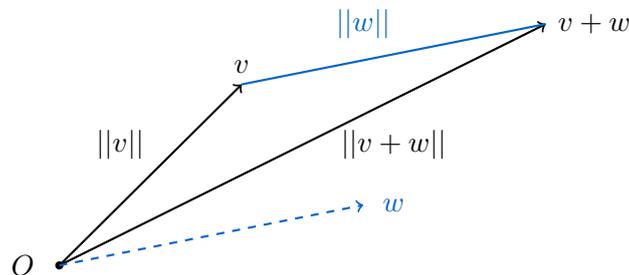
niert durch

$$\|v\| := \sqrt{\sum_{i=1}^n x_i^2}. \quad (2.21)$$

Eigenschaften der Norm sind

- $\|v\| \geq 0$ und $\|v\| = 0 \iff v = 0$
- $\|\lambda \cdot v\| = |\lambda| \cdot \|v\|$
- $\|v + w\| \leq \|v\| + \|w\|$ (**Dreiecksungleichung**).

Die ersten beiden sind leicht zu zeigen. Die Dreiecksungleichung ist geometrisch einsichtig, wenn man bedenkt, dass $\|w\|$ der Abstand zwischen den Vektoren v und $v + w$ ist.



Noch klarer wird es, wenn wir den Abstand zwischen zwei Punkten im \mathbb{R}^n definieren. Dies ist eine Abbildung d , die je zwei Punkten eine nichtnegative reelle Zahl zuordnet:

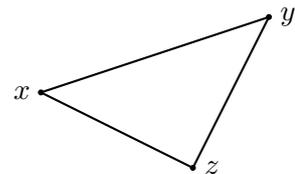
$$d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+ \quad (2.22)$$

$$(v, w) \mapsto d(v, w) = \|v - w\|. \quad (2.23)$$

Ein solcher Abstandsbegriff auf einer Menge wird **Metrik** genannt.

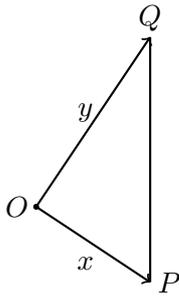
Aus der Definition und den Eigenschaften der Norm folgt sofort

- $d(x, y) = 0 \iff x = y$
- $d(x, y) = d(y, x)$ (Symmetrie)
- $d(x, y) \leq d(x, z) + d(y, z)$ (Dreiecksungleichung).



Wir werden die geometrisch sehr anschauliche Dreiecksungleichung etwas später algebraisch beweisen.

Bisher haben wir nur über Abstände gesprochen, aber nicht über Winkel. Betrachten wir zwei n -Tupel x und y , die den Strecken \overrightarrow{OP} und \overrightarrow{OQ} entsprechen sollen. Dann stehen die beiden Strecken senkrecht aufeinander, wenn das von O , P und Q gebildete Dreieck den Satz des Pythagoras erfüllt.



$$d(0, x)^2 + d(0, y)^2 = d(x, y)^2 \tag{2.24}$$

$$\iff \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i^2 = \sum_{i=1}^n (y_i - x_i)^2 \tag{2.25}$$

$$= \sum_{i=1}^n (y_i^2 - 2x_i y_i + x_i^2) \tag{2.26}$$

$$\iff \sum_{i=1}^n x_i \cdot y_i = 0. \tag{2.27}$$

Die durch x und y gegebenen Vektoren stehen also genau dann senkrecht aufeinander, wenn das sogenannte Skalarprodukt $\langle x, y \rangle$ Null ist.

Definition 2.8. Die Abbildung

$$\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \tag{2.28}$$

$$(x, y) \mapsto \langle x, y \rangle = \sum_{i=1}^n x_i y_i \tag{2.29}$$

heißt das kanonische **Skalarprodukt** auf \mathbb{R}^n .

Die Norm ergibt sich aus dem Skalarprodukt als $\|x\|^2 = \langle x, x \rangle$.

Aus der obigen Rechnung folgt

$$\langle x, y \rangle = \frac{1}{2} \left(\|x\|^2 + \|y\|^2 - \|y - x\|^2 \right), \tag{2.30}$$

das Skalarprodukt misst also in gewisser Weise, wie sehr das von x und y definierte Dreieck vom Satz des Pythagoras abweicht und damit wie sehr der Winkel zwischen x und y sich vom rechten Winkel unterscheidet.

Wir bemerken folgende Eigenschaften des Skalarprodukts:

- $\forall \lambda \in \mathbb{R}, \forall x, y \in \mathbb{R}^n : \langle \lambda x, y \rangle = \lambda \langle x, y \rangle$
- $\forall x, y, z \in \mathbb{R}^n : \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

Zusammen besagen diese Eigenschaften, dass das Skalarprodukt linear ist im ersten Eintrag. Das Skalarprodukt ist natürlich auch im zweiten Eintrag linear — das folgt auch aus der folgenden Eigenschaft

- $\forall x, y \in \mathbb{R}^n : \langle x, y \rangle = \langle y, x \rangle$ (Symmetrie).

Eine Abbildung $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ mit diesen Eigenschaften heißt **symmetrische Bilinearform**. Das kanonische Skalarprodukt ist außerdem **positiv definit**, d.h.

- $\forall x \in \mathbb{R}^n \setminus \{0\} : \langle x, x \rangle > 0$.

Das Skalarprodukt ist also eine positiv definite symmetrische Bilinearform.

Später werden wir Verallgemeinerungen des kanonischen Skalarprodukts begegnen, die auch diese Eigenschaften erfüllen.

Eine wichtige Eigenschaft des Skalarprodukts ist die folgende:

Satz 2.3 (Cauchy-Schwarzsche Ungleichung). $\forall v, w \in \mathbb{R}^n : |\langle v, w \rangle| \leq \|v\| \cdot \|w\|$.

Während sich die bisher angegebenen Eigenschaften sehr einfach aus der Definition des kanonischen Skalarprodukts ableiten lassen, ist es bei dieser Ungleichung schwieriger. Wir werden aber sehen, dass wir für ihren Beweis nur die oben aufgelisteten Eigenschaften benötigen: Linearität, Symmetrie, Positivität.

Anstatt diesen Satz direkt zu beweisen, betrachten wir die Cauchy-Schwarzsche Ungleichung in einem Spezialfall und beweisen den folgenden Hilfssatz:

Lemma 2.4. Seien $v, w \in \mathbb{R}^n$ mit $\|v\| = 1$ und $\|w\| = 1$. Dann gilt

$$-1 \leq \langle v, w \rangle \leq 1. \tag{2.31}$$

Beweis des Lemmas. Das Normquadrat ist nichtnegativ (**Positivität**), also

$$0 \leq \|v - w\|^2 = \langle v - w, v - w \rangle \tag{2.32}$$

$$= \langle v, v \rangle - 2 \langle v, w \rangle + \langle w, w \rangle \quad (\text{Linearität \& Symmetrie}) \tag{2.33}$$

$$= 2 - 2 \langle v, w \rangle \tag{2.34}$$

$$\implies \langle v, w \rangle \leq 1. \tag{2.35}$$

Analog folgt

$$0 \leq \|v + w\|^2 = 2 + 2 \langle v, w \rangle \tag{2.36}$$

$$\implies \langle v, w \rangle \geq -1. \tag{2.37}$$

□

Nun folgt der Beweis der allgemeinen Cauchy-Schwarzsche Ungleichung aus dem soeben bewiesenen Lemma.

Beweis der Cauchy-Schwarzschen Ungleichung. Seien $v, w \in \mathbb{R}^n$. Ist $v = 0$ oder $w = 0$ folgt die Behauptung sofort. Seien also $v \neq 0$ und $w \neq 0$, dann führen wir die normierten Vektoren $\hat{v} = \frac{1}{\|v\|}v$ und $\hat{w} = \frac{1}{\|w\|}w$ ein, für die laut dem Lemma gilt

$$|\langle \hat{v}, \hat{w} \rangle| \leq 1 \iff \left| \left\langle \frac{1}{\|v\|}v, \frac{1}{\|w\|}w \right\rangle \right| \leq 1 \tag{2.38}$$

$$\iff \frac{1}{\|v\|} \cdot \frac{1}{\|w\|} \cdot |\langle v, w \rangle| \leq 1 \quad (\text{Linearität}) \tag{2.39}$$

$$\iff |\langle v, w \rangle| \leq \|v\| \cdot \|w\|. \tag{2.40}$$

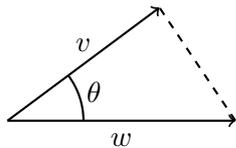
□

Bemerkung

Wie wir an der Herleitung sehen, gilt die Cauchy-Schwarzsche Ungleichung für jede positive symmetrische Bilinearform.

Aus der Cauchy-Schwarzschen Ungleichung folgt für $v, w \neq 0$:

$$-1 \leq \frac{\langle v, w \rangle}{\|v\| \cdot \|w\|} \leq 1. \quad (2.41)$$



Wir erhalten dann

$$\langle v, w \rangle = \|v\| \cdot \|w\| \cdot \cos \theta. \quad (2.42)$$

Dieser so definierte Winkel stimmt natürlich mit dem im zwei- bzw. dreidimensionalen Anschauungsraum definierten Winkel überein. Man betrachte zwei Vektoren v, w , dann gilt für das von $0, v$ und w gebildete Dreieck der Kosinussatz der Trigonometrie:

$$\|v - w\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\|\|w\|\cos \theta, \quad (2.43)$$

andererseits ist

$$\|v - w\|^2 = \|v\|^2 + \|w\|^2 - 2\langle v, w \rangle. \quad (2.44)$$

Zu beachten ist aber, dass wir in der Linearen Algebra den Winkel nicht als unabhängiges Konzept eingeführt haben, somit ist für uns die Aussage $\langle v, w \rangle = \|v\|\|w\|\cos \theta$ kein Satz, den wir beweisen könnten, sondern lediglich eine Definition eines Winkels.

Mit Hilfe der Cauchy-Schwarzschen Ungleichung können wir jetzt die Dreiecksungleichung von weiter oben beweisen:

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2\langle v, w \rangle \quad (2.45)$$

$$\leq \|v\|^2 + \|w\|^2 + 2\|v\|\|w\| = (\|v\| + \|w\|)^2 \quad (2.46)$$

$$\implies \|v + w\| \leq \|v\| + \|w\| \quad (\text{die Quadratwurzelfunktion ist monoton}). \quad (2.47)$$

Jetzt, da wir den Längenbegriff algebraisiert haben, können wir fortfahren, geometrische in algebraische Operationen zu übersetzen und uns kurz den Drehungen zuwenden.

Eine Drehung (der Einfachheit halber im \mathbb{R}^2) ist zunächst eine Abbildung $f_D : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, da sie die Ebene in sich selbst überführt. Wir wollen Drehungen um den Ursprung

betrachten, also $f_D(0) = 0$. Die Drehung einer Summe von Vektoren führt zur Summe der gedrehten Vektoren, also soll gelten

$$f_D(v + w) = f_D(v) + f_D(w). \quad (2.48)$$

Die Skalierung vertauscht mit der Drehung: erst drehen, dann skalieren führt zum gleichen Ergebnis wie erst skalieren und dann drehen, also verlangen wir

$$f_D(\lambda \cdot v) = \lambda \cdot f_D(v). \quad (2.49)$$

Eine Abbildung mit den gerade angegebenen zwei Eigenschaften heißt **lineare Abbildung**.

Man kann sich überlegen, dass eine lineare Abbildung $f_D : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ von der Form

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} d_{11}x_1 + d_{12}x_2 \\ d_{21}x_1 + d_{22}x_2 \end{pmatrix} \quad (2.50)$$

mit reellen Koeffizienten $d_{i,j}$ sein muss.

Dies kann man wie folgt einsehen: $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ werden durch f_D auf bestimmte Vektoren abgebildet, die wir $\begin{pmatrix} d_{11} \\ d_{21} \end{pmatrix}$ und $\begin{pmatrix} d_{12} \\ d_{22} \end{pmatrix}$ nennen wollen. Dann ist aufgrund der Linearität

$$f_D\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = f_D\left(x_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \quad (2.51)$$

$$= x_1 \cdot f_D\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + x_2 \cdot f_D\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \quad (2.52)$$

$$= x_1 \cdot \begin{pmatrix} d_{11} \\ d_{21} \end{pmatrix} + x_2 \cdot \begin{pmatrix} d_{12} \\ d_{22} \end{pmatrix} \quad (2.53)$$

$$= \begin{pmatrix} d_{11}x_1 + d_{12}x_2 \\ d_{21}x_1 + d_{22}x_2 \end{pmatrix}. \quad (2.54)$$

Die Koeffizienten fasst man in einer Matrix $\begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$ zusammen, d.h. die linearen Abbildungen $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ lassen sich durch eine solche Koeffizientenmatrix charakterisieren,

$$f_D \text{ linear} \iff \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} =: D \quad (2.55)$$

Wir schreiben dann auch

$$f_D\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = D \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad (2.56)$$

wobei die Wirkung einer Matrix D auf einen Vektor durch

$$D \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} := \begin{pmatrix} d_{11}x_1 + d_{12}x_2 \\ d_{21}x_1 + d_{22}x_2 \end{pmatrix} \quad (2.57)$$

definiert ist.

Nun fordern wir, dass die Abbildung f_D Längen und Winkel erhält:

$$\text{für alle } v, w \in \mathbb{R}^2 : \langle f_D(v), f_D(w) \rangle = \langle v, w \rangle . \quad (2.58)$$

Es genügt, dies für die kanonischen Einheitsvektoren $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ zu fordern, dann gilt die Eigenschaft auch für alle Vektoren, die aus e_1 und e_2 kombiniert werden können (also für alle Vektoren im \mathbb{R}^2).

Wir fordern daher

$$\langle f_D(e_1), f_D(e_1) \rangle = \langle e_1, e_1 \rangle , \quad (2.59)$$

also

$$\| \begin{pmatrix} d_{11} \\ d_{21} \end{pmatrix} \|^2 = \| \begin{pmatrix} 1 \\ 0 \end{pmatrix} \|^2 = 1 \iff d_{11}^2 + d_{21}^2 = 1 . \quad (2.60)$$

Damit ist $|d_{11}| \leq 1$ und es gibt ein $\alpha \in \mathbb{R}$, sodass $d_{11} = \cos \alpha$. Dann ist $d_{21}^2 = 1 - \cos^2 \alpha = \sin^2 \alpha$ und durch geeignete Wahl von $\alpha \in [0, 2\pi)$ können wir $d_{21} = \sin \alpha$ erreichen.

Ebenso fordern wir

$$\langle f_D(e_2), f_D(e_2) \rangle = \langle e_2, e_2 \rangle , \quad (2.61)$$

also $d_{12}^2 + d_{22}^2 = 1$. Analog zur obigen Diskussion können wir ein reelles $\beta \in [0, 2\pi)$ finden mit $d_{22} = \cos \beta$, $d_{12} = -\sin \beta$ (die Wahl des Vorzeichens wird sich gleich als praktisch erweisen).

Nun fordern wir noch

$$\langle f_D(e_1), f_D(e_2) \rangle = \langle e_1, e_2 \rangle = 0 , \quad (2.62)$$

also

$$d_{11} \cdot d_{12} + d_{21} \cdot d_{22} = 0 . \quad (2.63)$$

Setzen wir unsere oben gefundene Parametrisierung durch α und β ein, so erhalten wir die Bedingung

$$-\cos \alpha \sin \beta + \sin \alpha \cos \beta = 0 . \quad (2.64)$$

Diese Gleichung hat die Lösungen

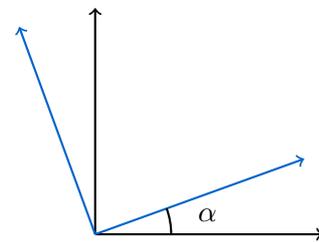
1. $\beta = \alpha$
2. $\beta = \alpha \pm \pi$ (das obere Vorzeichen gilt für $\alpha < \pi$, das untere für $\alpha \geq \pi$).

Es gibt also zwei Arten von längen- und winkelerhaltenden linearen Abbildungen im \mathbb{R}^2 :

1. $D_\alpha = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}$ ($0 \leq \alpha < 2\pi$) mit der Wirkung

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} -\sin \alpha \\ \cos \alpha \end{pmatrix} .$$

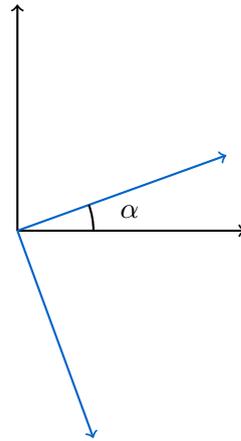
Dies beschreibt eine Drehung um den Winkel α .



2. Zweitens, $D'_\alpha = \begin{pmatrix} \cos \alpha & \sin \alpha \\ \sin \alpha & -\cos \alpha \end{pmatrix}$ ($0 \leq \alpha < 2\pi$) mit der Wirkung

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} \sin \alpha \\ -\cos \alpha \end{pmatrix}.$$

Dies beschreibt eine Spiegelung an der x_1 -Achse, gefolgt von einer Drehung um α .



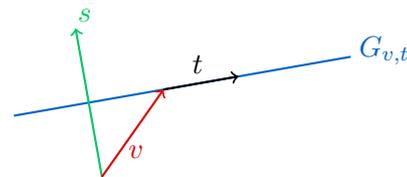
Damit haben wir jetzt auch die geometrische Operation der Drehung und der Spiegelung in die algebraische Formulierung übersetzt.

Mithilfe des Skalarprodukts können wir eine alternative Beschreibung von Geraden angeben:

Sind $v, s, t \in \mathbb{R}^2$, $s \neq 0, t \neq 0$, s und t orthogonal ($\langle s, t \rangle = 0$), so ist

$$G_{v,t} = \{u \in \mathbb{R}^2 \mid \langle s, u - v \rangle = 0\}, \quad (2.65)$$

d.h. wir haben eine Gerade $G_{v,t}$ statt durch eine Parametrisierung als Lösungsmenge einer linearen Gleichung geschrieben.



Den Vektor s , der orthogonal zur Geradenrichtung steht, nennen wir **Normalenvektor** der Geraden.

Bemerkung

Ist $\|s\| = 1$, so nennt man die obige Darstellung einer Geraden im \mathbb{R}^2 die *Hessesche Normalform* der Geraden.

Analog lässt sich eine Ebene im \mathbb{R}^3 als Lösungsmenge einer linearen Gleichung schreiben: Seien $v, s \in \mathbb{R}^3$, $s \neq 0$, dann ist

$$\{u \in \mathbb{R}^3 \mid \langle s, u - v \rangle = 0\}$$

eine Ebene durch v . Die Ebene wird aufgespannt von zwei linear unabhängigen Vektoren $t_1, t_2 \in \mathbb{R}^3$, die orthogonal sind zu s , $\langle s, t_1 \rangle = 0$, $\langle s, t_2 \rangle = 0$. Auch hier heißt s Normalenvektor der Ebene.

Auch Geraden lassen sich im \mathbb{R}^3 als Lösungsräume beschreiben. Wir können die Gerade als Schnittmenge zweier Ebenen realisieren, die wiederum durch zwei (linear unabhängige) Normalenvektoren gegeben sind: Sei $G_{v,t} \subset \mathbb{R}^3$ eine Gerade, s_1, s_2 linear unabhängige

Vektoren orthogonal zu t . Dann ist

$$G_{v,t} = \{u \in \mathbb{R}^3 \mid \langle s_1, u - v \rangle = 0 \text{ und } \langle s_2, u - v \rangle = 0\}. \quad (2.66)$$

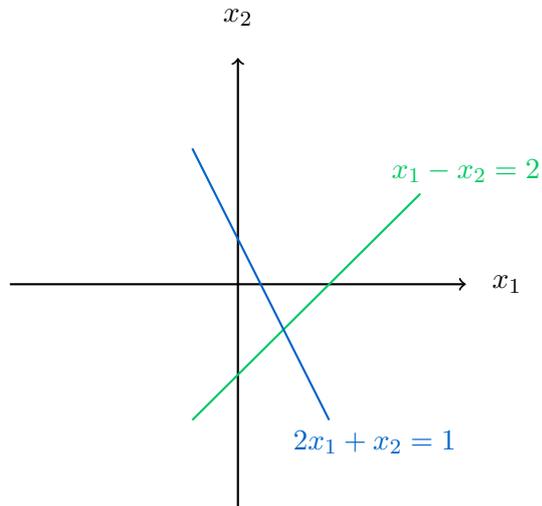
Wir sehen hieran zweierlei: Zum einen können wir geometrische Objekte als Lösungsräume linearer Gleichungssysteme beschreiben. Zum anderen können wir versuchen, gegebene lineare Gleichungssysteme geometrisch zu deuten.

Zum Beispiel

$$2x_1 + x_2 = 1 \quad (2.67)$$

$$x_1 - x_2 = 2. \quad (2.68)$$

Die erste Gleichung beschreibt eine Gerade mit Normalenvektor $s_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, die zweite eine mit Normalenvektor $s_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$; der Schnittpunkt $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ gibt die gemeinsame Lösung der beiden Gleichungen an.



Hieran kann man schon den engen Zusammenhang zwischen linearen Gleichungssystemen und den einfachen Objekten der Geometrie (Geraden, Ebenen, etc.) erkennen.

Im letzten Abschnitt dieses Kapitels werden wir diesen Zusammenhang auf bestimmte quadratische Gleichungen ausdehnen.

2.3 Kegelschnitte und Quadriken

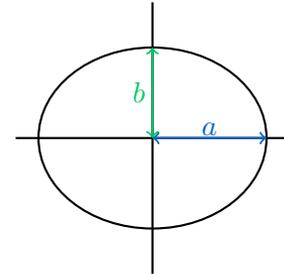
Betrachten wir statt linearer Gleichungen nun eine einfache quadratische Gleichung im \mathbb{R}^2 :

$$x_1^2 + x_2^2 = a^2. \quad (2.69)$$

Die Lösungsmenge sind alle Vektoren der Länge a , also haben wir einen Kreis mit Radius a beschrieben.

Betrachten wir stattdessen die Gleichung ($a, b > 0$)

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1. \quad (2.70)$$

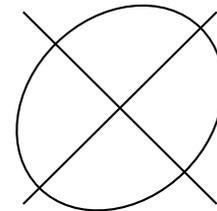


Der Lösungsraum ist eine Ellipse mit Halbachsen a und b .

Schwieriger zu entscheiden ist es beispielsweise bei der Gleichung $x_1^2 - x_1x_2 + x_2^2 = 1$. Dies ist wieder eine Ellipse, und zwar mit den Halbachsen $a = \sqrt{2}$ und $b = \sqrt{\frac{2}{3}}$, aber diese Ellipse ist um 45° gedreht. Dies kann man einsehen, indem man ein um 45° gedrehtes Koordinatensystem einführt:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} \quad (2.71)$$

$$= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} x'_1 - x'_2 \\ x'_1 + x'_2 \end{pmatrix}. \quad (2.72)$$



Dann haben wir

$$x_1^2 - x_1x_2 + x_2^2 = \frac{1}{2} \left((x'_1 - x'_2)^2 - (x'_1 - x'_2)(x'_1 + x'_2) + (x'_1 + x'_2)^2 \right) \quad (2.73)$$

$$= \frac{1}{2} (x_1'^2 + 3x_2'^2), \quad (2.74)$$

also wird die Gleichung in den neuen Koordinaten zu

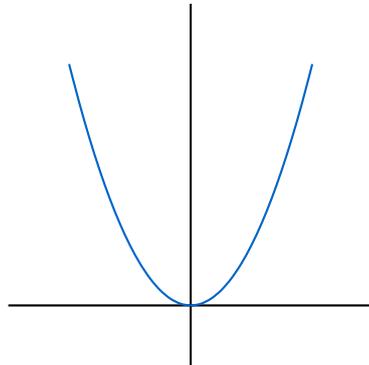
$$\frac{x_1'^2}{a^2} + \frac{x_2'^2}{b^2} = 1 \quad \text{mit} \quad a = \sqrt{2}, \quad b = \sqrt{\frac{2}{3}}. \quad (2.75)$$

Die Drehung, die wir angewandt haben, um eine Ellipsengleichung in Standardform zu erhalten, nennt man auch **Hauptachsentransformation**.

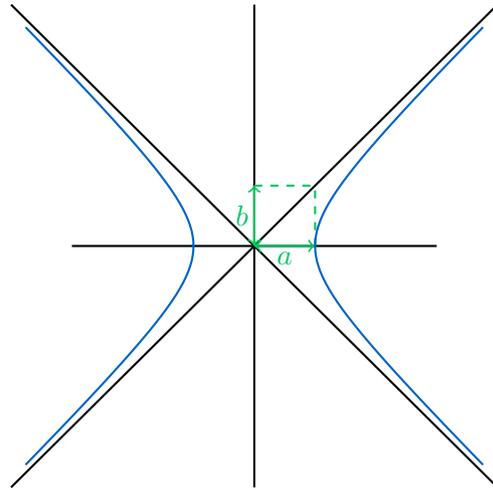
Weitere geometrische Objekte im \mathbb{R}^2 , die durch eine quadratische Gleichung beschrieben werden, sind die anderen Kegelschnitte.

Mithilfe der linearen Algebra lässt sich bei jedem durch eine quadratische Gleichung beschriebenen Objekt (abgesehen von bestimmten ausgearteten Spezialfällen wie z.B.

$x_1^2 + x_2^2 = 0$) eine Verschiebung und eine Drehung finden, die das Objekt auf einen der Standardkegelschnitte zurückführt.



Parabel $x_1^2 - 4ax_2 = 0$



Hyperbel $\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$

So etwas funktioniert auch im \mathbb{R}^n , allgemein nennt man Objekte, die durch eine quadratische Gleichung in den Koordinaten beschrieben sind, **Quadriken**.

Die in der Geometrie so nützlichen und anschaulichen Hauptachsentransformationen spielen aber auch in abstrakten Räumen (wie z.B. in der Quantenmechanik) eine wichtige Rolle. Wir werden sie später genauer diskutieren. Dafür benötigen wir aber erst einmal weiteres Rüstzeug, das wir uns in den nächsten Kapiteln zulegen wollen. Vorher aber noch eine kurze Zusammenfassung von dem, was Sie aus diesem Kapitel mitgenommen haben sollten:

Zusammenfassung Kapitel 2

- Sie haben eine Vorstellung bekommen, warum wir Geraden, Ebenen und den dreidimensionalen Raum mit den algebraischen Objekten \mathbb{R} , \mathbb{R}^2 und \mathbb{R}^3 identifizieren und wieso eine Verallgemeinerung auf den \mathbb{R}^n sinnvoll ist.
- Sie wissen, was die Vektoraddition ist und können Sie geometrisch deuten. Ähnliches gilt für die Skalarmultiplikation.
- Sie haben eine Vorstellung davon, was lineare Unabhängigkeit bedeutet und können die Definition angeben.
- Sie kennen die Definitionen vom kanonischen Skalarprodukt und von der Norm und ihre wichtigsten Eigenschaften (Dreiecksungleichung, Cauchy-Schwarzsche Ungleichung, ...) und können damit rechnen.
- Sie haben verstanden, wie eine Drehung im \mathbb{R}^2 mittels einer Matrix implementiert werden kann.
- Sie können die Orthogonalitätsbedingung für zwei Vektoren angeben und können Geraden im \mathbb{R}^2 und Ebenen im \mathbb{R}^3 durch ihre Normalenvektoren beschreiben.
- Sie haben einen Eindruck vom Zusammenhang zwischen geometrischen Objekten und Lösungsräumen von Gleichungen bekommen, z.B. im \mathbb{R}^3 :
 - lineare Gleichung \longleftrightarrow Gerade
 - quadratische Gleichung \longleftrightarrow Kegelschnitt

3 Vektorräume

3.1 Allgemeine reelle Vektorräume

Im letzten Kapitel haben wir die Räume \mathbb{R}^n untersucht. Wir wollen die dort beobachteten Strukturen nach und nach abstrahieren. Zunächst haben wir uns mit den Strukturen beschäftigt, die durch Vektoraddition und Skalarmultiplikation beschrieben werden. Wir wollen nun allgemeine reelle Vektorräume definieren als Mengen, auf denen genau solche Strukturen (mit den entsprechenden Eigenschaften) gegeben sind.

Wir führen also folgende Definition ein:

Definition 3.1. Ein Tripel $(V, +, \cdot)$ bestehend aus einer Menge V , einer Verknüpfung $+: V \times V \rightarrow V$ (genannt Addition), und einer Abbildung $\cdot: \mathbb{R} \times V \rightarrow V$ heißt **reeller Vektorraum** genau dann, wenn folgende Axiome gelten:

1. $(V, +)$ ist eine abelsche Gruppe (das neutrale Element $0 \in V$ nennen wir Nullvektor).
2. a) $\forall \lambda, \mu \in \mathbb{R}, v \in V: \lambda \cdot (\mu \cdot v) = (\lambda\mu) \cdot v$
 b) $\forall v \in V: 1 \cdot v = v$
3. a) $\forall \lambda \in \mathbb{R}, v, w \in V: \lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w$
 b) $\forall \lambda, \mu \in \mathbb{R}: (\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v$

Bemerkung

Wir haben die (acht) Axiome wieder in drei Gruppen eingeteilt:

1. Eigenschaften von $+$.
2. Verträglichkeit der Skalarmultiplikation mit der Multiplikation auf \mathbb{R} .
3. Verträglichkeit der Skalarmultiplikation mit der Addition auf \mathbb{R} und der Addition auf V .

Warum haben wir die Bedingung $1 \cdot v = v$ zu Punkt 2 sortiert? 1 ist das neutrale Element der Multiplikation in \mathbb{R} , und es zählt daher auch zu den Verträglichkeitsbedingungen, wie sich die ausgezeichneten Elemente von \mathbb{R} mit den neuen Strukturen verhalten. Warum hat man dann nicht auch z.B. $0 \cdot v = 0$ gefordert? Wie wir jetzt sehen werden, folgt diese Eigenschaft schon aus den anderen Axiomen:

Satz 3.1. Sei $(V, +, \cdot)$ ein reeller Vektorraum. Für alle $v \in V$ gilt dann $0 \cdot v = 0$.

Beweis. Es gilt

$$0 \cdot v = (0 + 0) \cdot v \quad (3.1)$$

$$= 0 \cdot v + 0 \cdot v. \quad (3.2)$$

Durch Addition von $-(0 \cdot v)$ folgt

$$0 = 0 \cdot v. \quad (3.3)$$

□

Geht das nicht auch für $1 \cdot v = v$? Folgt diese Bedingung vielleicht auch aus den anderen sieben Axiomen? Versuchen wir mal, es daraus abzuleiten:

Angenommen $1 \cdot v \neq v$. Multiplikation mit 1 liefert

$$1 \cdot (1 \cdot v) \neq 1 \cdot v \quad (3.4)$$

$$\iff (1 \cdot 1) \cdot v \neq 1 \cdot v \quad (3.5)$$

$$\iff 1 \cdot v \neq 1 \cdot v \quad \text{Widerspruch!} \quad (3.6)$$

Wo ist der Fehler? Gleich im ersten Schritt: Aus $v \neq w$ können wir nicht $1 \cdot v \neq 1 \cdot w$ schließen, solange wir nichts über die Eigenschaften der Multiplikation mit 1 wissen. Wenn wir allerdings voraussetzen, dass $v \mapsto 1 \cdot v$ injektiv ist, dann ist der obige Schluss erlaubt. Ohne diese Extra-Annahme müssen wir $1 \cdot v = v$ aber bei den Axiomen belassen.

Jetzt, wo wir die Eigenschaften von $+$ und \cdot aus dem \mathbb{R}^n abgekupfert haben, können wir uns fragen, ob es außer dem \mathbb{R}^n weitere interessante Beispiele für Vektorräume gibt.

Beispiel 3.1.

- Sei X eine Menge, $V = \text{Abb}(X, \mathbb{R})$ die Menge der Abbildungen von X nach \mathbb{R} . Dann ist $(V, +, \cdot)$ ein reeller Vektorraum, wenn $+$ und \cdot folgendermaßen definiert werden:

$$+ : V \times V \rightarrow V$$

$$(f, g) \mapsto f + g \text{ mit } (f + g)(x) := f(x) + g(x)$$

$$\cdot : \mathbb{R} \times V \rightarrow V$$

$$(\lambda, f) \mapsto \lambda \cdot f \text{ mit } (\lambda \cdot f)(x) := \lambda f(x)$$

Der Nullvektor ist die konstante Abbildung auf $0 : 0(x) = 0$.

- Sei $l = \{(a_1, a_2, \dots) \mid a_i \in \mathbb{R}\}$ die Menge der reellen Folgen. Mit der Addition

der Komponenten, bzw. Multiplikation der Komponenten mit einer reellen Zahl wird l zu einem reellen Vektorraum.

Bemerkung

Auf $\text{Abb}(X, \mathbb{R})$ ist auch eine multiplikative Verknüpfung definiert, indem das Produkt zweier Funktionen über das Produkt der Funktionswerte definiert wird,

$$(f \cdot g)(x) = f(x)g(x). \quad (3.7)$$

Damit wird $(\text{Abb}(X, \mathbb{R}), +, \cdot)$ ein Ring mit Eins (das neutrale Element bezüglich der Multiplikation ist die konstante Abbildung $x \mapsto 1$). Beachten Sie, dass \cdot hier in zwei Bedeutungen auftritt: als innere Verknüpfung und als Skalarmultiplikation. Einen Ring, der gleichzeitig ein reeller Vektorraum ist (mit entsprechenden Verträglichkeitsbedingungen — die Ringmultiplikation soll bilinear sein) nennt man eine **reelle Algebra**.

Oft kommt man in die Situation, bestimmte Teilmengen eines Vektorraums zu betrachten, beispielsweise die Menge aller reellen Funktionen auf \mathbb{R} , die eine vorgegebene Gleichung lösen. Dann stellt sich die Frage, ob diese Teilmenge mit den vererbten Strukturen wieder ein Vektorraum ist. Dies führt zum Begriff des Untervektorraums:

Definition 3.2. Sei $(V, +, \cdot)$ ein reeller Vektorraum. Eine Teilmenge $W \subset V$ heißt **Untervektorraum** von V genau dann, wenn

1. $W \neq \emptyset$
2. $\forall x, y \in W : x + y \in W$
3. $\forall x \in W, \lambda \in \mathbb{R} : \lambda \cdot x \in W$.

Wir fordern also, dass die beiden Strukturen, $+$ und \cdot , sich auf W einschränken lassen und wir nicht z.B. bei der Addition zweier Elemente von W außerhalb von W landen. Außerdem fordern wir, dass W nicht leer ist, sonst gäbe es in W auch kein neutrales Element 0 und W könnte kein Vektorraum sein. So aber gilt:

Satz 3.2. Sei $(V, +, \cdot)$ ein reeller Vektorraum und W ein Untervektorraum von V . Dann ist $(W, +, \cdot)$ mit den auf W eingeschränkten Operationen $+$ und \cdot ein reeller Vektorraum.

Beweis. Wir müssen die Vektorraumaxiome überprüfen. Die meisten sind offensichtlich erfüllt, da wir wissen, dass $+$ und \cdot die Eigenschaften haben, die $(V, +, \cdot)$ zum Vektorraum machen. Die einzigen Axiome, bei denen es nicht sofort klar ist, sind die Existenzforderungen: Existenz eines neutralen Elements und zu jedem $w \in W$ die Existenz eines inversen Elements $-w \in W$.

Sei $w \in W$ (da $W \neq \emptyset$, gibt es mindestens ein Element). Dann ist $0 \cdot w \in W$ (Abgeschlossenheit von W bezüglich der Skalarmultiplikation). Wie wir vorhin gesehen haben, ist $0 \cdot w = 0$, also $0 \in W$, somit gibt es in W ein neutrales Element.

Andererseits ist auch $(-1) \cdot w \in W$ und es gilt

$$(-1) \cdot w + w = (-1) \cdot w + 1 \cdot w = (-1 + 1) \cdot w = 0 \cdot w = 0, \quad (3.8)$$

also ist $-w = (-1) \cdot w \in W$. Mit w ist also auch das inverse Element $-w$ in W enthalten. \square

Beispiel 3.2.

- In jedem reellen Vektorraum $(V, +, \cdot)$ sind $\{0\}$ und V Untervektorräume.
- Eine Gerade $G_{v,t}$ im \mathbb{R}^n ($t \neq 0$) ist ein Untervektorraum genau dann, wenn $0 \in G_{v,t}$, wenn also die Gerade durch 0 geht. Wenn $0 \in G_{v,t}$, dann gibt es ein $\lambda \in \mathbb{R}$ mit $v = \lambda \cdot t$, also ist

$$G_{v,t} = \mathbb{R} \cdot t = \{\mu \cdot t \mid \mu \in \mathbb{R}\}. \quad (3.9)$$

$G_{v,t}$ ist nicht leer und offensichtlich abgeschlossen bezüglich Addition und Skalarmultiplikation.

Das Gleiche gilt auch für eine Ebene $\mathbb{R} \cdot t_1 + \mathbb{R} \cdot t_2$ durch 0 und allgemeiner für die Menge aller Linearkombinationen gegebener Vektoren t_1, \dots, t_m .

- Wir betrachten eine Teilmenge von $\text{Abb}(\mathbb{R}, \mathbb{R})$,

$$U = \{q \in \text{Abb}(\mathbb{R}, \mathbb{R}) \mid q \text{ zweimal differenzierbar und } \ddot{q} = -\omega^2 q\}, \quad (3.10)$$

wobei wir q als zeitabhängige Ortskoordinate, $t \mapsto q(t)$, interpretieren wollen und \ddot{q} stellt die zweite Ableitung von q nach der Zeit t dar. U ist also die Lösungsmenge der Bewegungsgleichung, die die Schwingung einer Masse an einer Feder beschreibt. Statt uns nun konkret auf die Suche nach Lösungen zu begeben, wollen wir die Struktur von U untersuchen; in der Tat können wir einsehen, dass U ein Untervektorraum von $(\text{Abb}(\mathbb{R}, \mathbb{R}), +, \cdot)$ ist:

- U ist nicht leer, da $t \mapsto 0$ die Gleichung löst.
- U ist abgeschlossen bezüglich $+$ und \cdot , da mit q und p auch $q+p$ und $\lambda \cdot q$ zweimal differenzierbar sind und

$$\underbrace{(\ddot{q} + \ddot{p})}_{\text{zweite Zeitableitung von } (q+p)} + \omega^2(q+p) = (\ddot{q} + \omega^2 q) + (\ddot{p} + \omega^2 p), \quad (3.11)$$

also ist $q+p$ eine Lösung, wenn q und p Lösungen sind. (Genauso ist $\lambda \cdot q$ eine Lösung, wenn q eine Lösung ist.)

Der Grund dafür, dass die Lösungsmenge U ein Untervektorraum ist, liegt darin, dass die Gleichung **linear** in q ist; würde beispielsweise ein quadratischer Term q^2 in der Gleichung auftreten, wäre die Lösungsmenge kein Untervektorraum.

Wenn wir mehrere Untervektorräume des gleichen reellen Vektorraums betrachten, können wir fragen, ob wir aus diesen Untervektorräumen neue konstruieren können.

Satz 3.3. Sei $(V, +, \cdot)$ ein reeller Vektorraum und U_1, U_2 Untervektorräume. Dann gilt:

- $U_1 \cap U_2$ ist ein Untervektorraum.
- $U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1, u_2 \in U_2\}$ ist ein Untervektorraum.

Bevor wir diesen Satz beweisen, können wir versuchen, uns anschaulich klar zu machen, was diese Operationen bewirken. Stellen wir uns beispielsweise zwei Ebenen durch 0 im \mathbb{R}^3 vor. Die Schnittmenge ist — sofern die Ebenen nicht gleich sind — eine Gerade durch 0, also wieder ein Untervektorraum.

Stellen wir uns andererseits zwei verschiedene Geraden G_1 und G_2 durch 0 vor. Ihre Summe $G_1 + G_2$ besteht aus allen Vektoren, die als Summe zweier Vektoren aus G_1 und G_2 beschreibbar sind: Das Resultat ist die von G_1 und G_2 aufgespannte Ebene durch 0, wieder ein Untervektorraum.

Die Vereinigung zweier Untervektorräume ist dagegen typischerweise kein Untervektorraum. Seien z.B. $G_1 = G_{0,e_1}$ und $G_2 = G_{0,e_2}$ die Geraden durch 0 im \mathbb{R}^2 mit Richtungsvektoren $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Dann sind insbesondere e_1 und e_2 in $G_1 \cup G_2$, aber $e_1 + e_2$ nicht.

Aber nun zum Beweis.

Beweis.

1. Da $0 \in U_1$ und $0 \in U_2$, ist $0 \in U_1 \cap U_2$, also ist $U_1 \cap U_2$ nicht leer.

Seien $u, v \in U_1 \cap U_2$, also $u, v \in U_1$ und $u, v \in U_2$. Da U_1 und U_2 Untervektorräume sind, gilt auch $u + v \in U_1$ und $u + v \in U_2$, also $u + v \in U_1 \cap U_2$, d.h. $U_1 \cap U_2$ ist abgeschlossen bezüglich der Addition.

Seien $\lambda \in \mathbb{R}$ und $u \in U_1 \cap U_2$, also $u \in U_1$ und $u \in U_2$. Da U_1 und U_2 Untervektorräume sind, gilt $\lambda \cdot u \in U_1$ und $\lambda \cdot u \in U_2$, also $\lambda \cdot u \in U_1 \cap U_2$. $U_1 \cap U_2$ ist also auch abgeschlossen bezüglich der Skalarmultiplikation.

2. $U_1 + U_2$ ist nicht leer, da $0 \in U_1$ und $0 \in U_2$ und somit $0 = 0 + 0 \in U_1 + U_2$.

Seien $u, v \in U_1 + U_2$, also gibt es $u_1, v_1 \in U_1$ und $u_2, v_2 \in U_2$ mit $u = u_1 + u_2$ und $v = v_1 + v_2$. Dann ist

$$u + v = (u_1 + u_2) + (v_1 + v_2) = (u_1 + v_1) + (u_2 + v_2) \in U_1 + U_2, \quad (3.12)$$

also ist $U_1 + U_2$ abgeschlossen bezüglich der Addition.

Seien $\lambda \in \mathbb{R}$ und $u \in U_1 + U_2$ mit $u = u_1 + u_2$, $u_1 \in U_1$ und $u_2 \in U_2$. Dann ist

$$\lambda \cdot u = \lambda \cdot (u_1 + u_2) = \lambda \cdot u_1 + \lambda \cdot u_2 \in U_1 + U_2, \quad (3.13)$$

also ist $U_1 + U_2$ abgeschlossen bezüglich der Skalarmultiplikation. □

Noch einmal zum Begriff der Linearkombination. Diesen hatten wir im \mathbb{R}^n eingeführt, er kann aber analog in beliebigen Vektorräumen formuliert werden. Da allgemeinere Räume unendlich viele unabhängige „Richtungen“ haben können, wollen wir den Begriff der Linearkombination für beliebige, möglicherweise unendlich große Familien definieren.

Definition 3.3. Sei $(V, +, \cdot)$ ein reeller Vektorraum und $(v_i)_{i \in I}$ eine Familie von Vektoren in V mit einer beliebigen Indexmenge I . Dann heißt $v \in V$ Linearkombination von $(v_i)_{i \in I}$, wenn es endlich viele $i_1, \dots, i_m \in I$ gibt und reelle Zahlen $\lambda_1, \dots, \lambda_m$ mit

$$v = \lambda_1 \cdot v_{i_1} + \dots + \lambda_m \cdot v_{i_m}. \quad (3.14)$$

m darf dabei auch Null sein, und die Summe von 0 Elementen ist dann als 0 definiert. Die leere Familie $I = \emptyset$ hat somit genau eine Linearkombination, nämlich den Nullvektor.

Definition 3.4. Für eine Familie $(v_i)_{i \in I}$ von Vektoren in einem reellen Vektorraum $(V, +, \cdot)$ nennen wir

$$\text{span}_{\mathbb{R}}(v_i)_{i \in I} = \{v \in V \mid v \text{ Linearkombination von } (v_i)_{i \in I}\} \quad (3.15)$$

den von $(v_i)_{i \in I}$ aufgespannten Raum. Den Index \mathbb{R} lassen wir meist weg.

Bemerkung

- $\text{span}(v_i)_{i \in I} \subset V$ ist ein Untervektorraum.
- Ist $W \subset V$ ein Untervektorraum und gilt $\forall i \in I : v_i \in W$, so ist $\text{span}(v_i)_{i \in I} \subset W$.

Also ist $\text{span}(v_i)_{i \in I}$ der kleinste Untervektorraum von V , der alle v_i enthält.

Beispiel 3.3.

- Für einen Vektor $t \in \mathbb{R}^n$, $t \neq 0$, ist $\text{span}(t) = \mathbb{R} \cdot t = G_{0,t}$ die Gerade durch 0 mit Richtungsvektor t .
- Wir betrachten wieder den Vektorraum l der reellen Folgen und darin die Vek-

toren

$$e_1 = (1, 0, 0, 0, \dots) \quad (3.16)$$

$$e_2 = (0, 1, 0, 0, \dots) \quad (3.17)$$

$$e_3 = (0, 0, 1, 0, \dots) \quad (3.18)$$

Der Untervektorraum, der von der Familie $(e_i)_{i \in \mathbb{N}}$ aufgespannt wird, ist dann

$$l_0 = \{(a_1, a_2, \dots) \mid a_i \in \mathbb{R}, \text{ nur endlich viele } a_i \neq 0\}. \quad (3.19)$$

Statt sich zu überlegen, welcher Untervektorraum von einer gegebenen Familie erzeugt wird, kann man sich bei einem gegebenen Raum auch fragen, was ein geeigneter Satz von Vektoren ist, der diesen Raum aufspannt. Man ist hierbei vor allem an einem möglichst kleinen Satz interessiert, um den Raum effizient beschreiben zu können. Beispielsweise bilden $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ und $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ einen geschickten Satz, um den \mathbb{R}^2 aufzuspannen: Mehr Vektoren benötigt man nicht, aber weglassen kann man auch keinen. Man sagt, (e_1, e_2) bilden eine **Basis** des \mathbb{R}^2 . Anscheinend ist es ungeschickt, eine Familie von Vektoren zu betrachten, die linear abhängig ist, denn das bedeutet ja gerade, dass sich wenigstens einer der Vektoren durch die anderen ausdrücken lässt und damit überflüssig ist. Aus diesem Grund werden wir uns, bevor wir uns genauer mit dem Begriff der Basis beschäftigen, noch einmal dem Konzept der linearen Unabhängigkeit zuwenden.

Wir verallgemeinern die bisherige Definition auf eine beliebige (möglicherweise unendliche) Familie von Vektoren in einem beliebigen Vektorraum V .

Wir werden von jetzt an der Einfachheit halber statt $(V, +, \cdot)$ nur noch V schreiben; es sollte aber trotzdem klar sein, dass zu einem Vektorraum immer eine Addition und eine Skalarmultiplikation gehören.

Definition 3.5. Eine endliche Familie (v_1, \dots, v_n) von Vektoren in einem reellen Vektorraum V heißt **linear unabhängig** genau dann, wenn gilt: Sind $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ und ist

$$\lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n = 0, \quad (3.20)$$

so folgt $\lambda_1 = 0, \dots, \lambda_n = 0$.

Eine beliebige Familie $(v_i)_{i \in I}$ heißt **linear unabhängig** genau dann, wenn jede endliche Teilfamilie linear unabhängig ist. (Dabei wird die leere Familie auch als linear unabhängig angesehen.)

Es stellt sich die Frage, wie man die lineare Unabhängigkeit in der Praxis überprüft. Offenbar muss man Aussagen über die Lösungen der entsprechenden Gleichung für die λ_i machen können. Das werden wir uns im folgenden Abschnitt am Beispiel des \mathbb{R}^n genauer ansehen.

3.2 Ein Exkurs über lineare Gleichungssysteme

Seien n Vektoren v_1, \dots, v_n im \mathbb{R}^m gegeben. Wir benennen die reellen Komponenten mit a_{ij} , wobei $j = 1, \dots, n$ die Vektoren kennzeichnet und $i = 1, \dots, m$ die Koordinaten des \mathbb{R}^m ,

$$v_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, v_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}. \quad (3.21)$$

Zum Test der linearen Unabhängigkeit betrachten wir die Vektorgleichung

$$\lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n = 0 \quad (3.22)$$

$$\Leftrightarrow \lambda_1 \cdot \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \dots + \lambda_n \cdot \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = 0, \quad (3.23)$$

die äquivalent ist zum Gleichungssystem

$$a_{11} \cdot \lambda_1 + \dots + a_{1n} \cdot \lambda_n = 0 \quad (3.24)$$

$$\vdots$$

$$a_{m1} \cdot \lambda_1 + \dots + a_{mn} \cdot \lambda_n = 0. \quad (3.25)$$

Dies nennt man ein **homogenes lineares** Gleichungssystem in den Variablen $\lambda_1, \dots, \lambda_n$: Linear bedeutet dabei, dass die λ_i nur linear auftreten (keine quadratischen Terme etc.), homogen bedeutet, dass auch keine konstanten Terme auftreten.

Um Aussagen über die lineare Unabhängigkeit machen zu können, müssen wir also solche homogenen linearen Gleichungssysteme verstehen.

Beispiel 3.4. Zu testen ist, ob die drei Vektoren $\begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix}$ linear unabhängig sind.

Dies führt zu dem Gleichungssystem

$$\lambda_1 - \lambda_2 + \lambda_3 = 0 \quad I \quad (3.26)$$

$$3\lambda_1 - 2\lambda_2 + 4\lambda_3 = 0 \quad II \quad (3.27)$$

$$2\lambda_2 + 2\lambda_3 = 0 \quad III \quad (3.28)$$

Wir ziehen die erste Gleichung mit drei multipliziert von der zweiten ab:

$$\lambda_1 - \lambda_2 + \lambda_3 = 0 \quad I \quad (3.29)$$

$$\lambda_2 + \lambda_3 = 0 \quad II' = II - 3 \cdot I \quad (3.30)$$

$$2\lambda_2 + 2\lambda_3 = 0 \quad III \quad (3.31)$$

Nun ziehen wir die zweite Gleichung multipliziert mit 2 von der dritten ab:

$$\lambda_1 - \lambda_2 + \lambda_3 = 0 \quad I \quad (3.32)$$

$$\lambda_2 + \lambda_3 = 0 \quad II' \quad (3.33)$$

$$0 = 0 \quad III' = III - 2 \cdot II' \quad (3.34)$$

Jetzt können wir ganz einfach eine Lösung ablesen. Wir wählen $\lambda_3 = \mu$ beliebig, dann ist

$$\lambda_2 = -\lambda_3 = -\mu \quad \text{wegen } II' \quad (3.35)$$

$$\lambda_1 = \lambda_2 - \lambda_3 = -2\mu \quad \text{wegen } I. \quad (3.36)$$

Die allgemeine Lösung ist damit

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} -2\mu \\ -\mu \\ \mu \end{pmatrix} = \mu \cdot \begin{pmatrix} -2 \\ -1 \\ 1 \end{pmatrix} \quad \text{mit } \mu \in \mathbb{R}, \quad (3.37)$$

wenn wir die Koeffizienten als Vektor schreiben.

Was schließen wir daraus?

1. Die Vektoren sind linear abhängig, denn es gibt nichttriviale Lösungen. Setzen wir $\mu = 1$, so schließen wir, dass gilt

$$-2 \cdot \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix} - 1 \cdot \begin{pmatrix} -1 \\ -2 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad (3.38)$$

was wir leicht nachrechnen können.

2. Da wir die allgemeine Lösung gefunden haben, sehen wir sofort, dass je zwei der Vektoren linear unabhängig sind: Wenn wir z.B. $\lambda_3 = 0$ setzen, folgt sofort $\lambda_1 = \lambda_2 = 0$, d.h. v_1 und v_2 sind linear unabhängig.

Wir wollen nun allgemeine homogene lineare Gleichungssysteme betrachten, d.h. Gleichungssysteme von der Form

$$a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n = 0 \quad (3.39)$$

$$\vdots \quad (3.40)$$

$$a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n = 0, \quad (3.41)$$

wobei wir die Unbestimmten jetzt x_i nennen wollen.

Die Strategie, die wir oben im Beispiel angewendet haben, müssen wir noch begründen. Zunächst machen wir die folgenden elementaren Beobachtungen:

1. $\left(\begin{array}{l} a = 0 \\ \text{und } b = 0 \end{array} \right) \iff \left(\begin{array}{l} b = 0 \\ \text{und } a = 0 \end{array} \right).$
2. Sei $\mu \in \mathbb{R}$. Dann gilt: $\left(\begin{array}{l} a = 0 \\ \text{und } b = 0 \end{array} \right) \iff \left(\begin{array}{l} a = 0 \\ \text{und } b + \mu a = 0 \end{array} \right).$
3. Sei $\mu \in \mathbb{R} \setminus \{0\}$. Dann gilt: $a = 0 \iff \mu a = 0.$

Hierbei sind a und b beliebige Ausdrücke, z.B.

$$a = a_{i1}x_1 + \dots + a_{in}x_n \quad (3.42)$$

$$b = a_{j1}x_1 + \dots + a_{jn}x_n. \quad (3.43)$$

Das bedeutet, dass wir ein äquivalentes Gleichungssystem erhalten, wenn wir folgende Operationen durchführen:

1. Vertauschung zweier Gleichungen,
2. Addition eines reellen Vielfachen einer Gleichung zu einer anderen,
3. Multiplikation einer Gleichung mit einer von Null verschiedenen reellen Zahl.

Mit diesen Operationen lassen sich lineare Gleichungssysteme systematisch lösen. Bevor wir das allgemeine Verfahren erläutern, wollen wir die Gleichungssysteme noch mit Hilfe von Matrizen geschickter aufschreiben.

Schon bei der Diskussion von Drehungen im \mathbb{R}^2 hatten wir gesehen, dass wir eine lineare Abbildung mit der Matrixschreibweise günstig notieren können. Wir definieren daher allgemein

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = A x := \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix} \quad (3.44)$$

mit der Matrix A und dem Vektor x .

Das Schema ist derart, dass jeder Zeilenvektor der Matrix mit dem Vektor x mit Hilfe des kanonischen Skalarprodukts multipliziert wird — dies ergibt die Einträge des resultierenden Vektors. Zu beachten ist, dass der Vektor x die Länge m hat, der Ergebnisvektor aber die Länge n .

Unser homogenes lineares Gleichungssystem wird dann zu

$$\underbrace{A}_{m \times n\text{-Matrix}} \underbrace{x}_{\text{Vektor der Länge } n} = \underbrace{0}_{\text{Nullvektor im } \mathbb{R}^m}. \quad (3.45)$$

Die Menge aller Lösungen dieser Gleichung bezeichnen wir mit

$$\text{Lös}(A, 0) := \{x \in \mathbb{R}^n \mid Ax = 0\}. \quad (3.46)$$

Später werden wir auch inhomogene Gleichungssysteme betrachten, $Ax = b$, die Lösungsmenge notieren wir dann als $\text{Lös}(A, b)$.

Bemerkung

Wir können die Lösung des linearen Gleichungssystems auch folgendermaßen deuten: Die Matrix A definiert eine Abbildung

$$f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m \tag{3.47}$$

$$x \mapsto Ax, \tag{3.48}$$

und die Menge der Lösungen ist das Urbild des Nullvektors im \mathbb{R}^m ,

$$\text{Lös}(A, 0) = f_A^{-1}(\{0\}). \tag{3.49}$$

Diese Sichtweise werden wir später wieder aufgreifen.

Alle Informationen über das Gleichungssystem stecken also jetzt in der Matrix A . Die Transformation des Gleichungssystems mit den diskutierten Operationen überträgt sich dann auf die folgenden elementaren Umformungen der Matrix:

1. Vertauschung von zwei Zeilen,
2. Addition eines reellen Vielfachen einer Zeile zu einer anderen Zeile,
3. Multiplikation einer Zeile mit einer von Null verschiedenen reellen Zahl.

Entsteht aus A durch diese **elementaren Zeilenumformungen** die Matrix \tilde{A} , dann gilt

$$\text{Lös}(A, 0) = \text{Lös}(\tilde{A}, 0), \tag{3.50}$$

d.h. die zugehörigen linearen Gleichungssysteme sind äquivalent.

Die Strategie ist nun, die Matrix A in eine Form \tilde{A} zu bringen, in der man das Gleichungssystem einfach lösen kann.

Die Form, die wir erreichen wollen, ist die sogenannte **Zeilenstufenform**:

$$\tilde{A} = \begin{pmatrix} \star & \cdots & & & \\ & \star & & & \\ & & \ddots & & \\ & & & \star & \\ & 0 & & & \star \end{pmatrix}. \tag{3.51}$$

Alle Einträge unter den Strichen sollen Null sein, die Sterne \star stehen für beliebige reelle Einträge ungleich Null (die Pivots genannt werden). Beispiele für Matrizen in Zeilenstufenform sind:

$$\begin{pmatrix} \textcircled{1} & 7 & 3 & 0 & -2 \\ 0 & 0 & \textcircled{2} & -1 & -1 \\ 0 & 0 & 0 & \textcircled{5} & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & \textcircled{1} & 6 & 2 \\ 0 & 0 & \textcircled{2} & 7 \\ 0 & 0 & 0 & \textcircled{4} \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{3.52}$$

Die umkreisten Einträge sind die Pivots.

Hieraus sollte die Definition der Zeilenstufenform klar werden. Wie man die Definition präzise aufschreiben kann, findet man z.B. im Buch von Fischer [F, Abschnitt 0.4.3].

Angenommen, wir haben eine Matrix \tilde{A} in Zeilenstufenform gefunden. Dann ist die Lösung des Gleichungssystems

$$\tilde{A}x = 0 \tag{3.53}$$

sehr einfach.

Betrachten wir das Beispiel von oben:

Für

$$\tilde{A} = \begin{pmatrix} \textcircled{1} & 7 & 3 & 0 & -2 \\ 0 & 0 & \textcircled{2} & -1 & -1 \\ 0 & 0 & 0 & \textcircled{5} & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.54}$$

wird $\tilde{A}x = 0$ zu

$$x_1 + 7x_2 + 3x_3 - 2x_5 = 0 \tag{3.55}$$

$$2x_3 - x_4 - x_5 = 0 \tag{3.56}$$

$$5x_4 + 3x_5 = 0 \tag{3.57}$$

$$0 = 0 \tag{3.58}$$

Wir unterscheiden zwischen zwei Arten von Variablen:

- Solche, die am Anfang einer Gleichung stehen: x_1, x_3, x_4 (diese entsprechen also den Spalten, in denen die Pivots stehen)
- Solche, die nicht am Anfang einer Gleichung auftauchen: x_2, x_5

x_2 und x_5 nennt man **freie Variablen**: Sie können beliebige Werte annehmen. Dadurch sind die anderen Variablen durch rekursives Einsetzen in die Gleichungen von unten nach oben festgelegt, sie werden **gebundene Variablen** genannt. Setzen wir $x_2 = \mu_2 \in \mathbb{R}$, $x_5 = \mu_5 \in \mathbb{R}$ mit zwei Parametern μ_2, μ_5 , so folgt

$$x_4 = -\frac{1}{5} \cdot 3x_5 = -\frac{3}{5}\mu_5 \tag{3.59}$$

$$x_3 = -\frac{1}{2} \cdot (-x_4 - x_5) = -\frac{1}{2} \cdot \left(\frac{3}{5} - 1\right) \cdot \mu_5 = \frac{1}{5} \cdot \mu_5 \tag{3.60}$$

$$x_1 = -(7x_2 + 3x_3 - 2x_5) = -7\mu_2 - 3 \cdot \left(\frac{1}{5}\mu_5\right) + 2\mu_5 = -7\mu_2 - \frac{7}{5}\mu_5. \tag{3.61}$$

Damit haben wir die Lösungsmenge eindeutig bestimmt:

$$\text{Lös}(\tilde{A}, 0) = \left\{ \mu_2 \cdot \begin{pmatrix} -7 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \frac{1}{5}\mu_5 \cdot \begin{pmatrix} -7 \\ 0 \\ 1 \\ -3 \\ 5 \end{pmatrix} \mid \mu_2, \mu_5 \in \mathbb{R} \right\} = \mathbb{R} \cdot \begin{pmatrix} -7 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mathbb{R} \cdot \begin{pmatrix} -7 \\ 0 \\ 1 \\ -3 \\ 5 \end{pmatrix}. \tag{3.62}$$

Im letzten Schritt haben wir die Parametrisierung in μ_5 durch einen Faktor 5 geändert, um die Brüche loszuwerden.

Der Lösungsraum ist in diesem Fall durch eine Ebene im \mathbb{R}^5 (durch 0) gegeben. Allgemein gilt, dass der Lösungsraum eines homogenen linearen Gleichungssystems ein Untervektorraum des \mathbb{R}^n ist. Darauf werden wir später noch zurückkommen.

Aus diesem Beispiel sollte klar sein, wie man ganz allgemein ein homogenes lineares Gleichungssystem löst, das in Zeilenstufenform gegeben ist: Für die freien Variablen werden beliebige reelle Parameter eingeführt, die Gleichungen erlauben es dann, die gebundenen Variablen durch diese Parameter ausdrücken.

Nun müssen wir noch verstehen, wie wir von einer beliebigen Matrix A zu einer Matrix \tilde{A} in Zeilenstufenform mit Hilfe der elementaren Zeilenumformungen gelangen.

Wir erläutern das Verfahren anhand eines Beispiels:

Sei

$$A = \begin{pmatrix} 0 & 0 & 0 & 2 & -1 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & -1 & 2 & 1 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}. \quad (3.63)$$

Im ersten Schritt wird die erste Spalte von A gesucht, die nicht nur aus Nullen besteht: In unserem Fall ist dies die zweite.

Durch Zeilenvertauschung wird ein von Null verschiedenes Element dieser Spalte in die erste Zeile gebracht (das Pivot):

$$\begin{pmatrix} 0 & \textcircled{1} & -2 & 1 & 0 \\ 0 & 0 & 0 & 2 & -1 \\ 0 & -1 & 2 & 1 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}. \quad (3.64)$$

Im zweiten Schritt werden durch Addition eines geeigneten Vielfachen der ersten Zeile zu den anderen Zeilen alle restlichen Einträge in der betrachteten Spalte zu Null gemacht. In unserem Beispiel wird dies durch Addition der ersten Zeile zur dritten Zeile erreicht:

$$\begin{pmatrix} 0 & \textcircled{1} & -2 & 1 & 0 \\ 0 & 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix} \quad (3.65)$$

Das gleiche Verfahren können wir nun für die blau gefärbte Teilmatrix anwenden, also die Matrix, die beim Element rechts unterhalb des ersten Pivots beginnt. In unserem Beispiel führt dies auf

$$\begin{pmatrix} 0 & \textcircled{1} & -2 & 1 & 0 \\ 0 & 0 & 0 & \textcircled{2} & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5/2 \end{pmatrix} \quad (3.66)$$

und durch nochmalige Anwendung auf die verbleibende **blaue** Teilmatrix schließlich auf

$$\tilde{A} = \begin{pmatrix} 0 & \textcircled{1} & -2 & 1 & 0 \\ 0 & 0 & 0 & \textcircled{2} & -1 \\ 0 & 0 & 0 & 0 & \textcircled{5/2} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (3.67)$$

Dieses am Beispiel illustrierte Verfahren kann auf allgemeine Matrizen angewandt werden. Die allgemeine Formulierung ist etwas mühsam, konzeptionell entstehen dabei aber keine Schwierigkeiten (siehe z.B. das Buch von Fischer [F, Abschnitt 0.4.7]). Was auch aus der Vorgehensweise ersichtlich sein sollte, ist, dass dieses Verfahren ohne grundsätzliche Schwierigkeiten auf einem Computer programmierbar ist.

Wir fassen das Ergebnis unserer Überlegungen als Satz zusammen:

Satz 3.4. *Jede Matrix A kann durch elementare Zeilenumformungen in eine Matrix \tilde{A} in Zeilenstufenform überführt werden.*

Beweis. Siehe Konstruktion oben. □

Damit sind wir nun am Ziel: Wir haben ein allgemeines Verfahren, ein homogenes lineares Gleichungssystem zu lösen. Es wird das **Eliminationsverfahren von Gauß** genannt. Wir fassen es noch einmal zusammen:

1. Das Gleichungssystem wird in der Form $Ax = 0$ geschrieben.
2. A wird auf Zeilenstufenform \tilde{A} gebracht.
3. Das äquivalente Gleichungssystem $\tilde{A}x = 0$ wird durch Parametrisierung der freien Variablen und rekursives Auflösen nach den gebundenen Variablen gelöst.

Eine Folgerung aus dem allgemeinen Verfahren ist:

Korollar 3.5. *Ein homogenes lineares Gleichungssystem in n Variablen bestehend aus $m < n$ Gleichungen besitzt nichttriviale Lösungen (also andere Lösungen als $x_1 = \dots = x_n = 0$).*

Dies folgt einfach daraus, dass es bei m Gleichungen, also m Zeilen der entsprechenden Matrix, höchstens m Pivots und also höchstens m gebundene Variablen gibt. Somit gibt es mindestens $n - m > 0$ freie Variablen und damit nichttriviale Lösungen.

Nach diesem Exkurs in lineare Gleichungssysteme wollen wir noch einmal auf die Frage der linearen Unabhängigkeit zurückkommen.

Unser Ausgangspunkt war die Frage, ob gegebene Vektoren $v_1, \dots, v_n \in \mathbb{R}^m$ linear unabhängig sind, d.h. ob die Gleichung

$$\lambda_1 \cdot v_1 + \dots + \lambda_n \cdot v_n = 0 \quad (3.68)$$

nichttriviale Lösungen besitzt. Indem wir die Vektoren v_1, \dots, v_n als Spaltenvektoren einer Matrix A schreiben, ist die Gleichung äquivalent zum Gleichungssystem

$$A\lambda = 0 \quad \text{mit} \quad \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}. \quad (3.69)$$

Wir wenden dann das Gaußsche Eliminationsverfahren an.

Wenn es am Ende keine freien Variablen λ_i gibt, ist die einzige Lösung $\lambda_1 = \dots = \lambda_n = 0$: Die Familie (v_1, \dots, v_n) ist linear unabhängig.

Angenommen, es gibt $f > 0$ freie Variablen $\lambda_{i_1}, \dots, \lambda_{i_f}$ und $n - f$ gebundene Variablen $\lambda_{j_1}, \dots, \lambda_{j_{n-f}}$. Dann ist (v_1, \dots, v_n) linear abhängig.

Es folgt aber auch, dass $(v_{j_1}, \dots, v_{j_{n-f}})$ linear unabhängig ist und

$$\text{span}(v_{j_1}, \dots, v_{j_{n-f}}) = \text{span}(v_1, \dots, v_n). \quad (3.70)$$

Die zu den freien Variablen gehörenden Vektoren v_{i_1}, \dots, v_{i_f} können also weggelassen werden, sie können aus den übrigen erzeugt werden.

Seien z.B. drei Vektoren (v_1, v_2, v_3) gegeben und bei der Lösung des Gleichungssystems $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0$ stellt sich heraus, dass wir λ_3 als freie Variable wählen können und λ_1, λ_2 dann durch λ_3 festgelegt sind. Setzen wir $\lambda_3 = -1$ so erhalten wir

$$\lambda_1 v_1 + \lambda_2 v_2 - v_3 = 0, \quad (3.71)$$

also lässt sich v_3 als Linearkombination der anderen schreiben. (v_1, v_2) ist dann linear unabhängig, denn

$$\lambda_1 v_1 + \lambda_2 v_2 = 0 \quad (3.72)$$

entspricht dem ursprünglichen System mit $\lambda_3 = 0$; da λ_1, λ_2 durch λ_3 eindeutig festgelegt sind, folgt auch $\lambda_1 = \lambda_2 = 0$.

Aus dem Korollar folgt übrigens noch das anschauliche Resultat: Im \mathbb{R}^m sind $n > m$ Vektoren immer linear abhängig.

3.3 Basen und Dimension

Wir kommen jetzt zu den allgemeinen reellen Vektorräumen zurück. Wie wir vorher schon angesprochen haben, ist es nützlich, zu einem Vektorraum einen effizienten Satz von Vektoren zu haben, die den Raum aufspannen, so wie die Einheitsvektoren im \mathbb{R}^n .

Dazu führen wir zunächst den Begriff des Erzeugendensystems ein.

Definition 3.6. Eine Familie $(v_i)_{i \in I}$ von Vektoren aus einem reellen Vektorraum $(V, +, \cdot)$ heißt **Erzeugendensystem** von V genau dann, wenn gilt

$$\text{span}(v_i)_{i \in I} = V, \quad (3.73)$$

d.h. wenn jeder Vektor $v \in V$ als Linearkombination endlich vieler v_i darstellbar ist.

Ein effizientes Erzeugendensystem ist eines, in dem es keine überflüssigen Vektoren gibt, die aus den restlichen kombiniert werden können, d.h. es ist eine linear unabhängige Menge. Das führt zu dem Begriff der Basis.

Definition 3.7. Eine Familie $(v_i)_{i \in I}$ von Vektoren in einem reellen Vektorraum $(V, +, \cdot)$ heißt **Basis** von V genau dann, wenn $(v_i)_{i \in I}$ ein **linear unabhängiges Erzeugendensystem** von V ist.

Es gibt mehrere äquivalente Charakterisierungen einer Basis:

Satz 3.6. Sei $(v_i)_{i \in I}$ eine Familie von Vektoren in einem reellen Vektorraum $(V, +, \cdot)$. Dann sind die folgenden Aussagen äquivalent:

1. $(v_i)_{i \in I}$ ist eine Basis von V .
2. $(v_i)_{i \in I}$ ist ein unverkürzbares Erzeugendensystem von V .
3. $(v_i)_{i \in I}$ ist eine maximale Familie linear unabhängiger Vektoren in V .
4. $(v_i)_{i \in I}$ ist ein Erzeugendensystem von V , und jeder Vektor $v \in V$ lässt sich als **eindeutige** Linearkombination von $(v_i)_{i \in I}$ schreiben.

Beweis. Wir beweisen hier nur die Äquivalenz von 1. und 2. Für einen Beweis der Äquivalenz von 3. und 4. kann man im Buch von Fischer [F, Abschnitt 1.5.2] nachlesen.

1. \implies 2.

Sei $(v_i)_{i \in I}$ eine Basis, dann ist es per Definition ein Erzeugendensystem. Angenommen, es wäre verkürzbar, d.h. wir könnten einen Vektor v_j herausnehmen und die Familie bliebe ein Erzeugendensystem. Dann könnten wir insbesondere v_j aus den übrigen als Linearkombination erzeugen. Das steht aber im Widerspruch zur vorausgesetzten linearen Unabhängigkeit, also muss das Erzeugendensystem unverkürzbar gewesen sein.

2. \implies 1.

Sei andersherum $(v_i)_{i \in I}$ ein unverkürzbares Erzeugendensystem. Wäre $(v_i)_{i \in I}$ linear abhängig, so könnte man einen Vektor v_j darin als Linearkombination der übrigen schreiben. Dann könnte man v_j aus der Familie herausnehmen und es bliebe ein Erzeugendensystem, wäre also verkürzbar, was zu einem Widerspruch führt. Also muss $(v_i)_{i \in I}$ linear unabhängig sein. \square

Als Nächstes beschäftigen wir uns mit der Frage, wie viele lineare unabhängige Vektoren es in einem gegebenen Vektorraum überhaupt geben kann. Wie wir am Ende des letzten Abschnitts schon gesehen hatten, ist im \mathbb{R}^m jede Familie mit mehr als m Vektoren immer linear abhängig. Eine ähnliche Aussage ist die folgende:

Lemma 3.7. *Sei V ein von einer endlichen Familie (v_1, \dots, v_m) erzeugter Vektorraum. Dann ist jede Familie mit mehr als m Vektoren linear abhängig.*

Beweis. Sei (w_1, \dots, w_n) eine Familie von $n > m$ Vektoren. Jedes w_i ist eine Linearkombination der v_j , also gibt es reelle Zahlen b_{ij} sodass $w_i = b_{i1}v_1 + \dots + b_{im}v_m$ mit $i = 1, \dots, n$. Die Gleichung

$$\lambda_1 \cdot w_1 + \dots + \lambda_n \cdot w_n = 0 \tag{3.74}$$

wird zu

$$\begin{aligned} & (b_{11}\lambda_1 + b_{12}\lambda_2 + \dots + b_{1n}\lambda_n) \cdot v_1 \\ & + (b_{21}\lambda_1 + b_{22}\lambda_2 + \dots + b_{2n}\lambda_n) \cdot v_2 \\ & + \dots \\ & + (b_{m1}\lambda_1 + b_{m2}\lambda_2 + \dots + b_{mn}\lambda_n) \cdot v_m = 0. \end{aligned}$$

Diese Vektorgleichung wird insbesondere dann gelöst, wenn alle Koeffizienten der v_j verschwinden (falls (v_1, \dots, v_m) linear abhängig ist, würde es noch weitere Lösungen geben). Das System

$$\begin{aligned} b_{11}\lambda_1 + b_{12}\lambda_2 + \dots + b_{1n}\lambda_n &= 0 \\ &\vdots \\ b_{m1}\lambda_1 + b_{m2}\lambda_2 + \dots + b_{mn}\lambda_n &= 0 \end{aligned}$$

ist ein homogenes lineares Gleichungssystem in n Variablen mit $m < n$ Gleichungen. Also gibt es immer eine nichttriviale Lösung und damit ist (w_1, \dots, w_n) linear abhängig. \square

Wir kommen jetzt zum wichtigsten Satz dieses Abschnitts:

Satz 3.8.

1. Jeder endlich erzeugte reelle Vektorraum V hat eine endliche Basis.
2. Alle Basen eines endlich erzeugten reellen Vektorraums V sind endlich und haben die gleiche Länge (d.h. gleich viele Basisvektoren).

Endlich erzeugt bedeutet, dass der Vektorraum ein endliches Erzeugendensystem besitzt, also von einer endlichen Familie erzeugt wird.

Beweis.

1. Der Vektorraum V besitzt nach Voraussetzung eine endliche Familie, die V erzeugt: $\text{span}(v_1, \dots, v_n) = V$. Ist (v_1, \dots, v_n) unverkürzbar, so haben wir eine Basis gefunden.

Ist es verkürzbar, so können wir einen Vektor weglassen, sodass die restliche Familie aus $n - 1$ Vektoren immer noch ein Erzeugendensystem ist. Ist es unverkürzbar, so haben wir eine Basis gefunden. Das Verfahren muss irgendwann enden, weil wir mit endlich vielen Vektoren angefangen haben.

2. Sei (v_1, \dots, v_n) eine endliche Basis von V und $(w_i)_{i \in I}$ eine weitere Basis. Da (v_1, \dots, v_n) insbesondere ein Erzeugendensystem ist, sind alle Familien mit mehr als n Vektoren sicherlich linear abhängig. Da $(w_i)_{i \in I}$ als Basis linear unabhängig ist, können wir sie als endliche Familie (w_1, \dots, w_m) mit $m \leq n$ annehmen. Da (w_1, \dots, w_m) andererseits V erzeugt, sind mehr als m Vektoren linear abhängig, also folgt $n \leq m$ und somit $m = n$.

□

Da also je zwei Basen die gleiche Länge haben, können wir definieren:

Definition 3.8. Sei V ein reeller Vektorraum mit einer endlichen Basis der Länge n . Dann sagen wir, V habe die **Dimension n** , $\dim V = n$.

Sei V ein Vektorraum ohne endliches Erzeugendensystem. Dann sagen wir, V habe die Dimension ∞ , $\dim V = \infty$.

Es stellt sich noch die Frage, ob alle reellen Vektorräume, also auch die der Dimension ∞ , eine Basis besitzen. Dies ist der Fall, allerdings benötigt der Beweis deutlich mehr Arbeit und einen Ausflug in die Grundlagen der Mathematik (man benötigt das sogenannte Auswahlaxiom oder das äquivalente Zornsche Lemma). Wer Interesse daran hat, kann z.B. beim Wikipedia-Eintrag zum Auswahlaxiom beginnen und von dort zum Zornschen Lemma weitergehen. Ein Beweis, dass jeder Vektorraum eine Basis besitzt, findet sich z.B. im Buch von Kowalsky [K, Satz 6.2 im zweiten Kapitel].

Manchmal sucht man nicht nach irgendeiner Basis, sondern man hat bereits eine Familie (v_1, \dots, v_m) und möchte diese zu einer Basis erweitern. Zunächst einmal stellen wir fest,

dass es — sofern (v_1, \dots, v_m) keine Basis ist — eine linear unabhängige Erweiterung von (v_1, \dots, v_m) gibt:

Lemma 3.9. *Sei V ein reeller Vektorraum, (v_1, \dots, v_m) eine linear unabhängige Familie mit $m < \dim V$. Dann gibt es ein $v_{m+1} \in V$, sodass (v_1, \dots, v_{m+1}) linear unabhängig ist.*

Beweis. Sei $U = \text{span}(v_1, \dots, v_m)$ der von (v_1, \dots, v_m) erzeugte Untervektorraum von V . Dann ist $U \neq V$, sonst wäre (v_1, \dots, v_m) Basis von V , also $m = \dim V$. Also ist $V \setminus U \neq \emptyset$.

Sei $v_{m+1} \in V \setminus U$. Dann ist (v_1, \dots, v_{m+1}) linear unabhängig: Denn seien $\lambda_1, \dots, \lambda_{m+1} \in \mathbb{R}$ mit $\lambda_1 v_1 + \dots + \lambda_{m+1} v_{m+1} = 0$. Angenommen $\lambda_{m+1} \neq 0$. Dann ließe sich v_{m+1} als Linearkombination von (v_1, \dots, v_m) schreiben, was $v_{m+1} \notin U$ widerspräche. Also ist $\lambda_{m+1} = 0$, also $\lambda_1 v_1 + \dots + \lambda_m v_m = 0$. Da (v_1, \dots, v_m) als linear unabhängig vorausgesetzt wurde, ist $\lambda_1 = \dots = \lambda_m = 0$. \square

Aus diesem Lemma folgt für endlichdimensionale Vektorräume der Basisergänzungssatz:

Satz 3.10 (Basisergänzungssatz). *Sei V ein reeller Vektorraum der endlichen Dimension n . Sei (v_1, \dots, v_m) eine linear unabhängige Familie von Vektoren in V . Dann gibt es Vektoren v_{m+1}, \dots, v_n , sodass (v_1, \dots, v_n) eine Basis von V ist.*

Beweis. Durch wiederholtes Anwenden des Lemmas erweitern wir (v_1, \dots, v_m) zu einer linearen unabhängigen Familie der Länge n . \square

Basen sind unter anderem deshalb so wichtig, weil wir Vektoren dann mithilfe von reellen Zahlen, den Koordinaten bezüglich der Basis, darstellen können — so wie wir auch am Anfang von Kapitel 2 reelle Koordinaten im Anschauungsraum eingeführt haben.

Bei gegebener Basis $\mathcal{B} = (v_1, \dots, v_n)$ eines n -dimensionalen reellen Vektorraums wollen wir also jedem Vektor $v \in V$ seine Koordinaten zuordnen: nämlich jene Koeffizienten λ_i , die auftauchen, wenn wir v als Linearkombination von (v_1, \dots, v_n) schreiben:

$$v = \lambda_1 v_1 + \dots + \lambda_n v_n. \quad (3.75)$$

Diese Koeffizienten sind eindeutig bestimmt.

Andererseits kann man einem Satz von Koordinaten den entsprechenden Vektor zuordnen.

Die Basis \mathcal{B} gibt also Anlass zu folgender **bijektiver** Abbildung:

$$\Phi_{\mathcal{B}}: \mathbb{R}^n \rightarrow V \quad (3.76)$$

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \mapsto \lambda_1 v_1 + \dots + \lambda_n v_n. \quad (3.77)$$

Die Umkehrabbildung $\Phi_{\mathcal{B}}^{-1}$ ordnet jedem Vektor seine Koordinaten bezüglich \mathcal{B} zu.

Damit sehen wir, dass sich reelle Vektorräume der Dimension n im Wesentlichen wie der \mathbb{R}^n verhalten.

Beispiel 3.5. *Wir wollen nun noch einmal auf unser Beispiel mit dem harmonischen Oszillator zurückkommen,*

$$U = \{q \in \text{Abb}(\mathbb{R}, \mathbb{R}) \mid q \text{ zweimal differenzierbar und } \ddot{q} = -\omega^2 q\}. \quad (3.78)$$

Wir hatten schon festgestellt, dass dieser Lösungsraum der Bewegungsgleichung ein Vektorraum ist.

Wie sieht dieser Raum aus?

Vielleicht kennen Sie die Lösung schon aus der Physik: Die allgemeine Lösung ist

$$q(t) = q_0 \cos \omega t + \frac{1}{\omega} v_0 \sin \omega t, \quad (3.79)$$

dabei ist $q(0) = q_0$ die Anfangsauslenkung und $\dot{q}(0) = v_0$ die Anfangsgeschwindigkeit.

Wir können also $\cos \omega t$ und $\sin \omega t$ als eine Basis von U nehmen, jedes andere Element in U ist eine Linearkombination der Standardlösung, wobei die Koeffizienten eine ganz klare physikalische Interpretation als Anfangsauslenkung und Anfangsgeschwindigkeit (durch ω) haben. U ist also ein zweidimensionaler reeller Vektorraum.

Bemerkung

Abschließend noch eine Bemerkung. Alles, was wir in diesem Kapitel für allgemeine reelle Vektorräume gemacht haben, können wir auch für Vektorräume über beliebigen Körpern K machen. In der Definition eines Vektorraums ersetzen wir dann die reelle Skalarmultiplikation durch eine entsprechende Abbildung ($\cdot : K \times V \rightarrow V$). Wir werden davon später Gebrauch machen, wenn wir komplexe Vektorräume betrachten, also $K = \mathbb{C}$.

Folgende Bemerkung geht über den normalen Vorlesungsstoff hinaus und richtet sich vor allem an die an abstrakten Konzepten Interessierten:

Bemerkung

Man kann auch versuchen, statt eines Körpers einen Ring mit Eins zu verwenden, die Axiome lassen sich dafür genauso verwenden das führt auf den Begriff eines Moduls über einen Ring. Da es in einem Ring nicht notwendigerweise zu jedem von 0 verschiedenen Element ein multiplikatives Inverses gibt, können wir aber nicht mehr alles, was wir für Vektorräume gemacht haben, auf Moduln übertragen, da wir nicht mehr durch die Koeffizienten teilen können.

Beispiel 3.6. $(\mathbb{Z}, +, \cdot)$ ist ein Ring mit Eins, $\mathbb{Z} = \mathbb{Z}^1$ ist dann auch ein Modul über \mathbb{Z} (so wie $\mathbb{R}^1 = \mathbb{R}$ ein reeller Vektorraum ist). Es gilt $\mathbb{Z} = 2\mathbb{Z} + 3\mathbb{Z}$, da jede ganze Zahl als Summe einer geraden und einer durch 3 teilbaren Zahl geschrieben werden kann. $(2, 3)$ ist also ein Erzeugendensystem von \mathbb{Z} , sogar ein unverkürzbares ($\mathbb{Z} \neq 2\mathbb{Z}$, $\mathbb{Z} \neq 3\mathbb{Z}$). Andererseits ist $(2, 3)$ linear abhängig, da $3 \cdot 2 - 2 \cdot 3 = 0$, in einem Modul kann es also unverkürzbare Erzeugendensysteme geben, die linear abhängig sind. Die einzigen Basen von \mathbb{Z} (d.h. linear unabhängige Erzeugendensysteme) sind (1) und (-1) .

Zusammenfassung Kapitel 3

Folgende Konzepte sollten Sie aus diesem Kapitel mitgenommen haben:

- Sie haben die allgemeine Definition eines reellen Vektorraums kennengelernt: Sie enthält drei Strukturen (die zugrundeliegende Menge, die Addition, die Skalarmultiplikation) und acht Axiome, die ihre Eigenschaften beschreiben.
- Sie kennen den Begriff des Untervektorraums und können beispielsweise angeben, welche Untervektorräume es im \mathbb{R}^2 gibt.
- Sie wissen, dass der Durchschnitt und die Summe zweier Untervektorräume wieder ein Untervektorraum ist.
- Sie haben eine Vorstellung folgender Begriffe: linear unabhängig, Linearkombination, Erzeugendensystem, Basis, Dimension.
- Sie können entscheiden, ob eine gegebene Matrix in Zeilenstufenform ist.
- Sie können ein gegebenes homogenes lineares Gleichungssystem mithilfe des Gaußschen Eliminationsverfahren lösen.
- Sie wissen, dass es dabei immer die triviale Lösung gibt und dass alle weiteren Lösungen mithilfe der beim Verfahren auftretenden freien Variablen parametrisiert werden können.

Im nächsten Kapitel werden wir uns mit linearen Abbildungen zwischen Vektorräumen beschäftigen. Dabei werden Matrizen eine große Rolle spielen, und wir werden uns daher mit ihnen und ihren Eigenschaften auseinander setzen.

4 Lineare Abbildungen und Matrizen

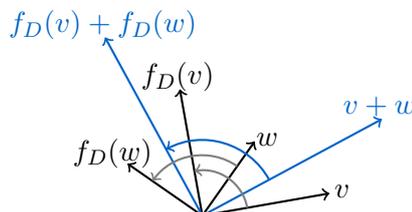
4.1 Lineare Abbildungen

Lineare Abbildungen sind uns schon begegnet: Die Drehungen im \mathbb{R}^2 um den Ursprung erfüllen

$$f_D(v+w) = f_D(v) + f_D(w) \quad (4.1)$$

und

$$f_D(\lambda \cdot v) = \lambda \cdot f_D(v). \quad (4.2)$$



Wir haben dann gesehen, dass eine solche lineare Abbildung immer mithilfe von Matrizen geschrieben werden kann. Dies werden wir nun genauer untersuchen. Zunächst definieren wir allgemein, was eine lineare Abbildung ist.

Definition 4.1. Seien V, W zwei reelle Vektorräume. Eine Abbildung

$$F : V \rightarrow W \quad (4.3)$$

heißt **linear** genau dann, wenn für alle $v, w \in V$ und alle $\lambda \in \mathbb{R}$ gilt:

1. $F(v+w) = F(v) + F(w)$
2. $F(\lambda \cdot v) = \lambda \cdot F(v)$.

Bemerkung

Die obigen Bedingungen 1. und 2. garantieren gerade, dass F mit den auf V und W definierten Strukturen verträglich ist. Eine solche strukturerhaltende Abbildung nennt man auch **Homomorphismus** oder in diesem Fall *Homomorphismus von reellen Vektorräumen*.

Man beachte, dass der Nullvektor von V immer auf den Nullvektor in W abgebildet wird:

$$F(\underbrace{0}_{0 \in V}) = F(\underbrace{0}_{0 \in \mathbb{R}} \cdot \underbrace{0}_{0 \in V}) = \underbrace{0}_{0 \in \mathbb{R}} \cdot F(\underbrace{0}_{0 \in V}) = \underbrace{0}_{0 \in W}. \quad (4.4)$$

Beispiel 4.1.

- Zunächst die trivialen Beispiele:

– Die Nullabbildung

$$\begin{aligned} 0 : V &\rightarrow W \\ v &\mapsto 0 \end{aligned}$$

– Die Identität

$$\begin{aligned} id_V : V &\rightarrow V \\ v &\mapsto v \end{aligned}$$

- Die Skalierung von Vektoren: Sei $\mu \in \mathbb{R}$, dann ist

$$\begin{aligned} F_\mu : V &\rightarrow V \\ v &\mapsto \mu \cdot v \end{aligned}$$

linear.

- Sei $\mathcal{C}^\infty(\mathbb{R}) = \{f \in \text{Abb}(\mathbb{R}, \mathbb{R}) \mid f \text{ ist beliebig oft differenzierbar}\}$ der Vektorraum der beliebig oft differenzierbaren Funktionen. Die Ableitung definiert eine lineare Abbildung:

$$\begin{aligned} D : \mathcal{C}^\infty(\mathbb{R}) &\rightarrow \mathcal{C}^\infty(\mathbb{R}) \\ f &\mapsto f'. \end{aligned}$$

- Sei $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine lineare Abbildung. Seien e_1, \dots, e_n die kanonischen Einheitsvektoren im \mathbb{R}^n ($e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots$), deren Bilder $F(e_j)$ sind Vektoren im \mathbb{R}^m , die wir $F(e_1) = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, F(e_n) = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$ nennen wollen. Dann ist für einen beliebigen Vektor $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \cdot e_1 + \dots + x_n \cdot e_n$

$$F(x) = F(x_1 \cdot e_1 + \dots + x_n \cdot e_n) \quad (4.5)$$

$$= x_1 \cdot F(e_1) + \dots + x_n \cdot F(e_n) \quad (4.6)$$

$$= \begin{pmatrix} a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n \\ \vdots \\ a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n \end{pmatrix} = \underbrace{\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}}_A \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_x = Ax. \quad (4.7)$$

Jede lineare Abbildung $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ hat diese Form.

Für zwei reelle Vektorräume V, W bezeichnen wir die Menge aller linearen Abbildungen von V nach W mit

$$\text{Hom}(V, W) = \{F : V \rightarrow W \mid F \text{ linear}\}. \quad (4.8)$$

Dieser Raum der Homomorphismen hat selbst eine Vektorraumstruktur, lineare Abbildungen können addiert werden und mit einem Skalar multipliziert werden.

Wie wir gerade im Beispiel gesehen haben, werden lineare Abbildungen von \mathbb{R}^n in den \mathbb{R}^m durch Matrizen vermittelt, und auch die Menge der reellen $m \times n$ -Matrizen, $M(m \times n, \mathbb{R})$, wird durch komponentenweise Addition und Skalarmultiplikation zu einem reellen Vektorraum.

Die Abbildung, die einer Matrix die zugehörige lineare Abbildung zuordnet, ist selbst linear,

$$f : M(m \times n, \mathbb{R}) \rightarrow \text{Hom}(\mathbb{R}^n, \mathbb{R}^m) \quad (4.9)$$

$$A \mapsto f_A \quad \text{mit} \quad f_A(x) = Ax, \quad (4.10)$$

denn $f_{A+B} = f_A + f_B$ und $f_{\lambda A} = \lambda f_A$.

Wie schon besprochen sichern die Eigenschaften einer linearen Abbildung, dass diese verträglich mit den Vektorraumstrukturen ist. Diese Verträglichkeit spiegelt sich in folgendem Satz wider:

Satz 4.1. *Seien V, W reelle Vektorräume, $F : V \rightarrow W$ eine lineare Abbildung. Seien $U_1 \subset V, U_2 \subset W$ Untervektorräume von V bzw. W . Dann gilt*

1. *Das Bild $F(U_1)$ von U_1 ist Untervektorraum von W .*
2. *Das Urbild $F^{-1}(U_2)$ von U_2 ist Untervektorraum von V .*

Beweis.

1. Wir überprüfen die Untervektorraumkriterien für $F(U_1)$:
 - Da $U_1 \neq \emptyset$ ist $F(U_1) \neq \emptyset$.
 - Seien $w_1, w_2 \in F(U_1)$, also $w_1 = F(u_1)$ und $w_2 = F(u_2)$ mit geeigneten $u_1, u_2 \in U_1$. Dann ist $w_1 + w_2 = F(u_1) + F(u_2) = F(u_1 + u_2) \in F(U_1)$.
 - Sei $w \in F(U_1)$, also $w = F(u)$ für $u \in U_1$. Sei $\lambda \in \mathbb{R}$. Dann ist $\lambda w = \lambda F(u) = F(\lambda u) \in F(U_1)$.
2. Wir überprüfen, dass $F^{-1}(U_2)$ ein Untervektorraum ist.
 - Es ist $F(0) = 0 \in U_2$, also $0 \in F^{-1}(U_2)$.
 - Seien $v_1, v_2 \in F^{-1}(U_2)$, also $F(v_1), F(v_2) \in U_2$. Dann ist $F(v_1 + v_2) = F(v_1) + F(v_2) \in U_2$, also $v_1 + v_2 \in F^{-1}(U_2)$.
 - Sei $v \in F^{-1}(U_2)$, also $F(v) \in U_2$. Sei $\lambda \in \mathbb{R}$. Dann ist $F(\lambda v) = \lambda F(v) \in U_2$, also $\lambda v \in F^{-1}(U_2)$.

□

Zwei spezielle Untervektorräume, die auf diese Weise entstehen, bekommen eigene Bezeichnungen:

Definition 4.2. • Wir nennen $\text{Im } F := F(V) \subset W$ das **Bild von F** .

• Wir nennen $\text{Ker } F := F^{-1}(\{0\}) \subset V$ den **Kern von F** .

Wenn F surjektiv ist, so gilt $\text{Im } F = W$, das Bild ist der gesamte Vektorraum W . Für Abbildungen zwischen endlichdimensionalen Vektorräumen ist $\text{Im } F$ natürlich auch endlichdimensional und man definiert:

Definition 4.3. Die Dimension von $\text{Im } F$ nennen wir den **Rang von F** :

$$\text{rang } F := \dim(\text{Im } F). \quad (4.11)$$

Ist $\text{rang } F = \dim W$, so ist F surjektiv.

Beispiel 4.2. Sei

$$f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m \quad (4.12)$$

$$x \mapsto Ax \quad (4.13)$$

eine durch eine $m \times n$ -Matrix A vermittelte lineare Abbildung. Das Bild $\text{Im } f_A$ ist der durch die Spaltenvektoren v_1, \dots, v_n von A aufgespannte Raum:

$$\text{Im } f_A = \{Ax \mid x \in \mathbb{R}^n\} \quad (4.14)$$

$$= \{x_1 \cdot v_1 + \dots + x_n \cdot v_n \mid x_1, \dots, x_n \in \mathbb{R}\} \quad (4.15)$$

$$= \text{span}(v_i)_{i \in \{1, \dots, n\}}. \quad (4.16)$$

Die Dimension von $\text{Im } f_A$ ergibt sich aus der Maximalzahl linear unabhängiger Vektoren in $(v_i)_{i \in \{1, \dots, n\}}$. Diese wiederum ergibt sich aus der Analyse der Gleichung $\lambda_1 v_1 + \dots + \lambda_n v_n = 0$: Hierbei ergeben sich eine Anzahl r gebundener Variablen (die Zahl der Pivot-Elemente der Zeilenstufenform von A) und $n - r$ freie Variablen. Wir hatten gesehen, dass die zu den $n - r$ freien Variablen gehörenden Vektoren aus den übrigen erzeugt werden können. Lässt man diese $n - r$ Vektoren weg, verbleibt eine Familie von r Vektoren, die linear unabhängig ist. Es gilt also

$$\text{rang } f_A = \dim \text{Im } f_A = r. \quad (4.17)$$

Dies nennt man auch den **Rang der Matrix A** .

Wir sehen daraus auch, dass die Zahl der Pivot-Elemente, die bei der Umformung von A in Zeilenstufenform auftritt, nicht von der gewählten Prozedur abhängt, also beispielsweise von der Wahl, welche Zeilen in einem bestimmten Schritt vertauscht wurden.

So wie das Bild $\text{Im } F$ uns Aufschluss darüber gibt, ob F surjektiv ist, so gibt uns der Kern Aufschluss über die Injektivität:

Lemma 4.2. *Seien V, W reelle Vektorräume, $F : V \rightarrow W$ eine lineare Abbildung. Dann gilt: F ist injektiv genau dann, wenn $\text{Ker } F = \{0\}$.*

Beweis. Sei F injektiv und $v \in V$ mit $F(v) = 0$ also $F(v) = F(0)$. Wegen der Injektivität folgt $v = 0$, also ist $\text{Ker } F = \{0\}$.

Sei andersherum $\text{Ker } F = \{0\}$ vorausgesetzt, und seien $v_1, v_2 \in V$ mit $F(v_1) = F(v_2)$, also wegen der Linearität $F(v_1 - v_2) = 0$ und somit $v_1 - v_2 \in \text{Ker } F$. Da $\text{Ker } F = \{0\}$ folgt $v_1 - v_2 = 0$, also $v_1 = v_2$. Somit ist F als injektiv erkannt. \square

Beispiel 4.3. *Sei*

$$f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m \tag{4.18}$$

$$x \mapsto Ax \tag{4.19}$$

wie im letzten Beispiel gegeben. Der Kern von f_A ist die Lösungsmenge der linearen Gleichung $Ax = 0$:

$$\text{Ker } f_A = \text{Lös}(A, 0) = \{x \in \mathbb{R}^n \mid Ax = 0\}. \tag{4.20}$$

Der Lösungsraum wird durch die $n - r$ freien Variablen ($r = \text{rang } f_A$) parametrisiert und hat also Dimension $n - r$. (Wir hatten gesehen, dass der Lösungsraum von $n - r$ linear unabhängigen Vektoren aufgespannt wird.)

Wir sehen im Beispiel der durch eine Matrix gegebenen linearen Abbildung den folgenden Zusammenhang:

$$\dim \text{Ker } f_A + \text{rang } f_A = (n - r) + r = n = \dim \mathbb{R}^n, \tag{4.21}$$

die Summe von Dimension von Bild und Kern von f_A ist gleich der Dimension des Definitionsbereichs \mathbb{R}^n . Diese **Dimensionsformel** gilt allgemein:

Satz 4.3. *Seien V, W endlichdimensionale reelle Vektorräume, $F : V \rightarrow W$ eine lineare Abbildung. Dann gilt*

$$\dim \text{Ker } F + \text{rang } F = \dim V. \tag{4.22}$$

Beweis. Sei $r = \text{rang } F$ und (w_1, \dots, w_r) eine Basis von $\text{Im } F$. Dann gibt es Vektoren $v_1, \dots, v_r \in V$ mit $F(v_1) = w_1, \dots, F(v_r) = w_r$. Sei $f = \dim \text{Ker } F$ und (u_1, \dots, u_f) eine Basis von $\text{Ker } F$. Wir wollen zeigen, dass $\mathcal{B} = (v_1, \dots, v_r, u_1, \dots, u_f)$ eine Basis von V ist. Dann folgt auch $r + f = \dim V$ und somit die Behauptung.

Wir zeigen zunächst, dass jeder Vektor $v \in V$ als Linearkombination von \mathcal{B} darstellbar ist.

Da $F(v) \in \text{Im } F$, gibt es $\lambda_1, \dots, \lambda_r \in \mathbb{R}$ mit

$$F(v) = \lambda_1 v_1 + \dots + \lambda_r v_r = \lambda_1 F(v_1) + \dots + \lambda_r F(v_r) = F(\lambda_1 v_1 + \dots + \lambda_r v_r), \quad (4.23)$$

also $v - \lambda_1 v_1 - \dots - \lambda_r v_r \in \text{Ker } F$. Somit gibt es $\mu_1, \dots, \mu_f \in \mathbb{R}$ mit

$$v - \lambda_1 v_1 - \dots - \lambda_r v_r = \mu_1 u_1 + \dots + \mu_f u_f \quad (4.24)$$

$$\implies v = \lambda_1 v_1 + \dots + \lambda_r v_r + \mu_1 u_1 + \dots + \mu_f u_f. \quad (4.25)$$

Es ist also $V = \text{span } \mathcal{B}$.

\mathcal{B} ist aber auch linear unabhängig: Seien $\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_f \in \mathbb{R}$ mit

$$\lambda_1 v_1 + \dots + \lambda_r v_r + \mu_1 u_1 + \dots + \mu_f u_f = 0. \quad (4.26)$$

Anwenden von F führt auf $\lambda_1 w_1 + \dots + \lambda_r w_r = 0$, also ist $\lambda_1 = \dots = \lambda_r = 0$, da (w_1, \dots, w_r) linear unabhängig ist. Die obige Gleichung reduziert sich daher auf

$$\mu_1 u_1 + \dots + \mu_f u_f = 0. \quad (4.27)$$

Da (u_1, \dots, u_f) linear unabhängig ist, folgt $\mu_1 = \dots = \mu_f = 0$. Also ist \mathcal{B} linear unabhängig. \square

Beispiel 4.4. Sei $A = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$. Dann ist $\text{rang } A = 1$, da A sofort auf Zeilenstufenform $\tilde{A} = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}$ gebracht werden kann. Es ist

$$\text{Ker } A = \text{Lös}(A, 0) = \text{Lös}(\tilde{A}, 0) = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mid x_1 + 2x_2 = 0 \right\} = \mathbb{R} \cdot \begin{pmatrix} -2 \\ 1 \end{pmatrix}. \quad (4.28)$$

Andererseits ist

$$\text{Im } A = \{Ax \mid x \in \mathbb{R}^2\} = \left\{ \frac{1}{5} \begin{pmatrix} x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{pmatrix} \mid x_1, x_2 \in \mathbb{R}^2 \right\} \quad (4.29)$$

$$= \left\{ \frac{1}{5} (x_1 + 2x_2) \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} \mid x_1, x_2 \in \mathbb{R}^2 \right\} \quad (4.30)$$

$$= \mathbb{R} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \quad (4.31)$$

$w = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ ist eine Basis von $\text{Im } A$. In diesem Fall können wir den gleichen Vektor $v = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ wählen mit

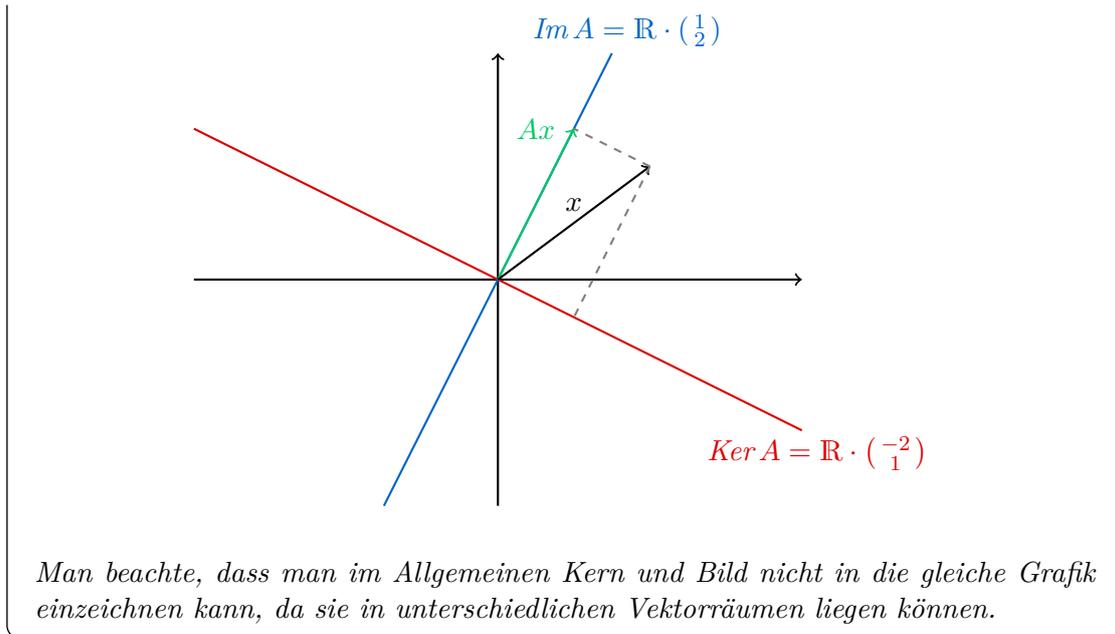
$$Av = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 5 \\ 10 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = w. \quad (4.32)$$

$u = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$ ist eine Basis von $\text{Ker } A$. (u, v) ist eine Basis von \mathbb{R}^2 (sie sind sogar orthogonal). Es gilt in diesem Fall (da $v = w$)

$$A(\lambda v + \mu u) = \lambda v. \quad (4.33)$$

Da $A(\lambda v) = \lambda v$, gilt also auch $f_A \circ f_A = f_A$.

A ist eine orthogonale Projektion auf die Gerade $\mathbb{R} \cdot v$:



Wir wollen noch einmal eine Methode zusammenfassen, wie wir Bild und Kern einer Matrix A (bzw. der zugehörigen Abbildung $f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$) bestimmen und eine Basis dafür angeben:

- Wir bringen A auf Zeilenstufenform \tilde{A} .
- Durch Parametrisierung der freien Variablen und rekursives Auflösen nach den gebundenen Variablen erhalten wir eine Darstellung des Lösungsraums $\text{Lös}(A, 0) = \text{Ker } A$.
- Seien μ_1, \dots, μ_f die freien Parameter, dann können wir aus der gewonnenen Darstellung

$$\text{Ker } A = \{ \mu_1 u_1 + \dots + \mu_f u_f \mid \mu_1, \dots, \mu_f \in \mathbb{R} \} \tag{4.34}$$

die Basis (u_1, \dots, u_f) des Kerns ablesen.

- Das Bild wird von den Spaltenvektoren v_1, \dots, v_n von A aufgespannt. Die Positionen (Spalten) i_1, \dots, i_r der Pivot-Elemente in \tilde{A} sagen uns, dass die entsprechenden Spaltenvektoren v_{i_1}, \dots, v_{i_r} von A (nicht von \tilde{A} !) eine Basis von $\text{Im } f_A = \text{Im } A$ bilden.

Wir haben nun einige Eigenschaften von linearen Abbildungen besprochen. Wir haben insbesondere gesehen, dass für eine lineare Abbildung $F : V \rightarrow W$ gilt

$$F \text{ surjektiv} \iff \text{Im } F = W \tag{4.35}$$

$$F \text{ injektiv} \iff \text{Ker } F = \{0\}. \tag{4.36}$$

Ist F **bijektiv**, so nennt man F einen **Isomorphismus**, bzw. Isomorphismus von reellen Vektorräumen.

Für Abbildungen zwischen endlichdimensionalen Räumen V, W haben wir den Rang als Dimension des Bildes eingeführt mit der Dimensionsformel

$$\text{rang } F + \dim \text{Ker } F = \dim V. \quad (4.37)$$

Für einen Isomorphismus gilt $\text{Ker } F = \{0\}$, also $\dim \text{Ker } F = 0$ und

$$\dim W = \dim \text{Im } F = \text{rang } F = \dim V. \quad (4.38)$$

Lineare Abbildungen können auch nacheinander ausgeführt werden. Wie verhält sich der Rang bei der Komposition?

Zunächst stellen wir fest, dass ganz allgemein die Komposition linearer Abbildungen wieder linear ist.

Satz 4.4. Seien V, W, Z reelle Vektorräume und $F : V \rightarrow W$, $G : W \rightarrow Z$ lineare Abbildungen. Dann ist

$$G \circ F : V \rightarrow Z \quad (4.39)$$

linear.

Beweis. Seien $v_1, v_2 \in V$ und $\lambda \in \mathbb{R}$. Dann ist

$$G \circ F(v_1 + v_2) = G(F(v_1 + v_2)) \quad (\text{Def. von } G \circ F) \quad (4.40)$$

$$= G(F(v_1) + F(v_2)) \quad (\text{Linearität von } F) \quad (4.41)$$

$$= G(F(v_1)) + G(F(v_2)) \quad (\text{Linearität von } G) \quad (4.42)$$

$$= G \circ F(v_1) + G \circ F(v_2). \quad (4.43)$$

Analog folgt $G \circ F(\lambda \cdot v) = \lambda \cdot G \circ F(v)$. □

Die Komposition linearer Abbildungen ist also linear. Außerdem ist sie verträglich mit der Addition und Skalarmultiplikation:

Satz 4.5. Für lineare Abbildungen $F_1, F_2 : V \rightarrow W$, $G_1, G_2 : W \rightarrow Z$ und $\lambda \in \mathbb{R}$ gilt

- $G_1 \circ (F_1 + F_2) = G_1 \circ F_1 + G_1 \circ F_2$ und $(G_1 + G_2) \circ F_1 = G_1 \circ F_1 + G_2 \circ F_1$,
- $(\lambda \cdot G_1) \circ F_1 = G_1 \circ (\lambda \cdot F_1) = \lambda \cdot (G_1 \circ F_1)$.

Beweis. Exemplarisch zeigen wir die erste Eigenschaft: Sei $v \in V$. Dann ist

$$(G_1 \circ (F_1 + F_2))(v) = G_1((F_1 + F_2)(v)) \quad \text{Def. von } \circ \quad (4.44)$$

$$= G_1(F_1(v) + F_2(v)) \quad \text{Def. der Addition lin. Abb.} \quad (4.45)$$

$$= G_1(F_1(v)) + G_1(F_2(v)) \quad \text{Linearität von } G_1 \quad (4.46)$$

$$= (G_1 \circ F_1)(v) + (G_1 \circ F_2)(v) \quad \text{Def. von } \circ. \quad (4.47)$$

□

Wie verhält sich nun der Rang unter Komposition?

Der Rang einer komponierten linearen Abbildung kann nie größer sein als der Rang einer der Komponenten:

Satz 4.6. Seien $F : V \rightarrow W$, $G : W \rightarrow Z$ lineare Abbildungen. Dann ist

$$\text{rang } G \circ F \leq \min(\text{rang } G, \text{rang } F). \quad (4.48)$$

Beweis. Es ist $\text{Im } G \circ F \subset \text{Im } G$, also $\text{rang } G \circ F \leq \text{rang } G$. Andererseits ist $\text{Ker } F \subset \text{Ker } G \circ F$, also $\dim \text{Ker } G \circ F \geq \dim \text{Ker } F$, und somit

$$\text{rang } G \circ F = \dim V - \dim \text{Ker } G \circ F \leq \dim V - \dim \text{Ker } F = \text{rang } F. \quad (4.49)$$

□

Der Rang einer Abbildung verändert sich nicht, wenn sie von links mit einem injektiven Homomorphismus oder von rechts mit einem surjektiven Homomorphismus verknüpft wird:

Satz 4.7. Seien $F : V \rightarrow W$, $G : W \rightarrow Z$ lineare Abbildungen. Dann gilt

$$G \text{ injektiv} \implies \text{rang } G \circ F = \text{rang } F \quad (4.50)$$

$$F \text{ surjektiv} \implies \text{rang } G \circ F = \text{rang } G. \quad (4.51)$$

Beweis. Sei $G : W \rightarrow Z$ injektiv. Dann ist

$$\text{Ker } G \circ F = \{v \in V \mid G(F(v)) = 0\} \quad (4.52)$$

$$= \{v \in V \mid F(v) = 0\} \quad (G \text{ injektiv}) \quad (4.53)$$

$$= \text{Ker } F \quad (4.54)$$

$$\implies \text{rang } G \circ F = \text{rang } F. \quad (4.55)$$

Sei $F : V \rightarrow W$ surjektiv. Dann ist

$$\text{Im } G \circ F = \{z \in Z \mid \exists v \in V : G(F(v)) = z\} \quad (4.56)$$

$$= \{z \in Z \mid \exists w \in W : G(w) = z\} \quad (F \text{ surjektiv}) \quad (4.57)$$

$$= \text{Im } G \quad (4.58)$$

$$\implies \text{rang } G \circ F = \text{rang } G \quad (4.59)$$

□

Komposition mit Isomorphismen verändert den Rang also nicht:

$$\text{rang } (H \circ G \circ F) = \text{rang } G \quad \text{für Isomorphismen } H, F. \quad (4.60)$$

Ein wichtiger Isomorphismus ist der Basisisomorphismus, der einen endlichdimensionalen Vektorraum V der Dimension n mit dem \mathbb{R}^n mithilfe einer gewählten Basis $\mathcal{B} = (v_1, \dots, v_n)$ in Beziehung setzt:

$$\Phi_B : \mathbb{R}^n \rightarrow V \tag{4.61}$$

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto x_1 v_1 + \dots + x_n v_n. \tag{4.62}$$

Mit dessen Hilfe kann die Untersuchung linearer Abbildungen zwischen endlichdimensionalen Vektorräumen auf die Untersuchung von Matrizen zurückgeführt werden.

4.2 Matrizenkalkül

Wir haben gesehen, dass lineare Abbildungen von \mathbb{R}^n nach \mathbb{R}^m immer von der Form $x \mapsto Ax$ sind.

Matrizen entsprechen linearen Abbildungen. Abbildungen können hintereinander ausgeführt werden. Was bedeutet diese Komposition für die Matrizen?

Wir betrachten also die Abbildungen

$$f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m \qquad f_B : \mathbb{R}^m \rightarrow \mathbb{R}^p \qquad (4.63)$$

$$x \mapsto Ax \qquad x \mapsto Bx. \qquad (4.64)$$

A habe die Einträge a_{ij} , B die Einträge b_{ij} . Wir schreiben dafür $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ oder kurz $A = (a_{ij})$ (und analog $B = (b_{ij})$).

Dann ist $f_B \circ f_A : \mathbb{R}^n \rightarrow \mathbb{R}^p$ wieder linear, also wieder durch eine Matrix darstellbar:

$$f_B \circ f_A(x) = Cx =: (BA)x. \qquad (4.65)$$

Die so entstandene Matrix nennen wir das **Produkt der Matrizen B und A** .

Es wird wie folgt bestimmt:

$$f_B \circ f_A(x) = B \begin{pmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \dots + a_{mn}x_n \end{pmatrix} \qquad (4.66)$$

$$= \begin{pmatrix} b_{11}(a_{11}x_1 + \dots + a_{1n}x_n) + \dots + b_{1m}(a_{m1}x_1 + \dots + a_{mn}x_n) \\ \vdots \\ b_{p1}(a_{11}x_1 + \dots + a_{1n}x_n) + \dots + b_{pm}(a_{m1}x_1 + \dots + a_{mn}x_n) \end{pmatrix} \qquad (4.67)$$

$$= \begin{pmatrix} (b_{11}a_{11} + \dots + b_{1m}a_{m1})x_1 + \dots + (b_{11}a_{1n} + \dots + b_{1m}a_{mn})x_n \\ \vdots \\ (b_{p1}a_{11} + \dots + b_{pm}a_{m1})x_1 + \dots + (b_{p1}a_{1n} + \dots + b_{pm}a_{mn})x_n \end{pmatrix} \qquad (4.68)$$

$$= \underbrace{\begin{pmatrix} (b_{11}a_{11} + \dots + b_{1m}a_{m1}) & \dots & (b_{11}a_{1n} + \dots + b_{1m}a_{mn}) \\ \vdots & & \vdots \\ (b_{p1}a_{11} + \dots + b_{pm}a_{m1}) & \dots & (b_{p1}a_{1n} + \dots + b_{pm}a_{mn}) \end{pmatrix}}_{BA} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \qquad (4.69)$$

Ausgeschrieben folgt also

$$\begin{pmatrix} b_{11} & \dots & b_{1m} \\ \vdots & & \vdots \\ b_{p1} & \dots & b_{pm} \end{pmatrix} \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = \begin{pmatrix} (b_{11}a_{11} + \dots + b_{1m}a_{m1}) & \dots & (b_{11}a_{1n} + \dots + b_{1m}a_{mn}) \\ \vdots & & \vdots \\ (b_{p1}a_{11} + \dots + b_{pm}a_{m1}) & \dots & (b_{p1}a_{1n} + \dots + b_{pm}a_{mn}) \end{pmatrix}. \quad (4.70)$$

Das Element von $C = (BA) = (c_{ij})$ in der i -ten Zeile und der j -ten Spalte erhält man als das kanonische Skalarprodukt des i -ten Zeilenvektors von A mit dem j -ten Spaltenvektor von B (die beide m -Tupel sind):

$$c_{ij} = b_{i1}a_{1j} + \dots + b_{im}a_{mj} = \sum_{k=1}^m b_{ik}a_{kj}. \quad (4.71)$$

Diese Operation ordnet also einer $p \times m$ -Matrix und einer $m \times n$ -Matrix eine $p \times n$ -Matrix zu.

Bemerkung

Das Produkt einer Matrix mit einem Spaltenvektor (interpretiert als Matrix mit nur einer Spalte) reproduziert unsere Konvention für den Ausdruck Ax .

Die Matrizenmultiplikation erfüllt eine Reihe von Eigenschaften, die sich direkt anhand der Definition zeigen lassen oder aus den entsprechenden Eigenschaften der Komposition linearer Abbildungen ergeben.

Satz 4.8.

1. Für alle $m \times n$ Matrizen A , $n \times p$ -Matrizen B und $p \times q$ -Matrizen C gilt:

$$(AB)C = A(BC) \quad (4.72)$$

d.h. Matrizenmultiplikation ist assoziativ.

2. Für alle $m \times n$ -Matrizen A_1, A_2 und $n \times p$ -Matrizen B_1, B_2

$$A_1(B_1 + B_2) = A_1B_1 + A_1B_2 \quad , \quad (A_1 + A_2)B_1 = A_1B_1 + A_2B_1, \quad (4.73)$$

d.h. für Matrizenaddition und -multiplikation gelten Distributivgesetze.

3. Für alle $m \times n$ -Matrizen A und $n \times p$ -Matrizen B und alle $\lambda \in \mathbb{R}$ gilt

$$\lambda(AB) = (\lambda A)B = A(\lambda B). \quad (4.74)$$

Beweis. Alle Eigenschaften kann man direkt nachrechnen, die erste folgt aber auch aus der Assoziativität der Komposition von Abbildungen (siehe Satz 1.1) und die beiden anderen folgen direkt aus den entsprechenden Eigenschaften linearer Abbildungen (siehe Satz 4.5). \square

Bemerkung

Die Menge der quadratischen Matrizen $M(n \times n, \mathbb{R})$ bildet bezüglich der Addition und der Matrizenmultiplikation einen Ring mit Eins; zusammen mit der Skalarmultiplikation sogar eine reelle Algebra.

Das neutrale Element der Multiplikation ist die Einheitsmatrix

$$E_n := \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}. \quad (4.75)$$

Für die Einträge führt man das sogenannte **Kronecker-Symbol** ein:

$$\delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases}. \quad (4.76)$$

Dann ist $E_n = (\delta_{ij})$.

Die Multiplikation von Matrizen ist **nicht kommutativ**, z.B. ist

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (4.77)$$

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \quad (4.78)$$

Wir sehen aus dem Beispiel auch, dass das Produkt zweier Matrizen Null ergeben kann, selbst wenn beide Matrizen von Null verschieden sind.

Zu einer Matrix $A \in M(m \times n, \mathbb{R})$ führt man die **transponierte Matrix** A^t ein, in der die Rolle von Zeilen und Spalten vertauscht sind.

$$\begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 1 \end{pmatrix}^t = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 3 & 1 \end{pmatrix}, \quad (4.79)$$

oder allgemein

$$A^t = (a'_{ij}) \quad (4.80)$$

mit $a'_{ij} = a_{ji}$.

Wir werden in Kürze sehen, warum das Transponieren eine sinnvolle und praktische Operation ist, und wir werden auch einen Eindruck davon bekommen, was die Transposition für lineare Abbildungen bedeutet.

Für das Transponieren gelten folgenden Regeln:

Satz 4.9. Für $A, C \in M(m \times n, \mathbb{R})$, $B \in M(n \times r, \mathbb{R})$, $\lambda \in \mathbb{R}$ gilt

1. $(A + C)^t = A^t + C^t$

2. $(\lambda A)^t = \lambda A^t$

3. $(A^t)^t = A$

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

Beweis. Die Regeln 1. bis 3. folgen sofort aus den Definitionen. Die letzte Relation rechnen wir nach:

Sei $A^t = (a'_{ij})$, $B^t = (b'_{ij})$, $C = (AB) = (c_{ij})$, $C^t = (c'_{ij})$. Dann gilt

$$c_{ik} = \sum_{j=1}^n a_{ij} b_{jk}, \quad (4.81)$$

also $c'_{ik} = c_{ki} = \sum_{j=1}^n a_{kj} b_{ji} = \sum_{j=1}^n b'_{ij} a'_{jk}$ und somit

$$\implies C^t = B^t A^t. \quad (4.82)$$

□

Mit der Transposition können wir jetzt das kanonische Skalarprodukt wie folgt beschreiben: Seien $v, w \in \mathbb{R}^n$, dann ist

$$\langle v, w \rangle = \sum_{i=1}^n v_i w_i = v^t w, \quad (4.83)$$

wenn wir den letzten Ausdruck als Matrixprodukt einer einzeiligen Matrix mit einer einspaltigen Matrix verstehen.

Wir hatten gesagt, dass wir Matrizen als lineare Abbildungen auf den Räumen \mathbb{R}^n ansehen können. Zu was für einer Abbildung gehört v^t für $v \in \mathbb{R}^n$? Dies entspricht einer linearen Abbildung

$$f_{v^t} : \mathbb{R}^n \rightarrow \mathbb{R} \quad (4.84)$$

$$w \mapsto v^t w = \langle v, w \rangle. \quad (4.85)$$

Den Raum $\text{Hom}(\mathbb{R}^n, \mathbb{R})$ der linearen Abbildungen auf \mathbb{R} nennt man auch den Raum der Linearformen (oder der linearen Funktionalen) oder den **Dualraum** $(\mathbb{R}^n)^*$ von \mathbb{R}^n .

Jeder Zeilenvektor führt zu einem solchen linearen Funktional, und jedes lineare Funktional wird durch einen Zeilenvektor beschrieben. Andererseits hat man mithilfe der Transposition eine bijektive lineare Abbildung zwischen Zeilen- und Spaltenvektoren. Dies führt somit zu einem Isomorphismus von \mathbb{R}^n und $\text{Hom}(\mathbb{R}^n, \mathbb{R})$:

$$\begin{aligned} \mathbb{R}^n &\rightarrow \text{Hom}(\mathbb{R}^n, \mathbb{R}) \\ v &\mapsto f_{v^t} \end{aligned}$$

Für $f_{v^t} : w \mapsto v^t w = \langle v, w \rangle$ findet man auch die Notation $\langle v, \cdot \rangle$, wobei der Punkt \cdot als Platzhalter für einen einzusetzenden Vektor steht.

Bemerkung

In der Quantenphysik ist es auch üblich, die sogenannte Diracsche Bra-Ket-Notation einzuführen: Vektoren v eines Vektorraums werden Ket-Vektoren $|v\rangle$ genannt, das zugehörige lineare Funktional wird Bra-Vektor $\langle v|$ genannt. Die Wirkung eines Bra-Vektors $\langle v|$ auf einen Ket-Vektor $|w\rangle$ ist dann $\langle v|w\rangle := \langle v, w \rangle$. Man beachte aber, dass in der Quantenphysik typischerweise andere Räume als der \mathbb{R}^n auftauchen, d.h. auch das Skalarprodukt ist ein anderes.

Wir erkennen also einen Teil der Bedeutung der Transposition darin, dass mit ihrer Hilfe bzw. mit Hilfe des Skalarprodukts ein Isomorphismus zwischen \mathbb{R}^n und dem Dualraum $\text{Hom}(\mathbb{R}^n, \mathbb{R})$ definiert wird.

Was bedeutet Transposition nun für allgemeine Matrizen A ?

A^t vermittelt eine zu $x \mapsto Ax$ duale Abbildung auf dem Dualraum:

Sei $A \in M(m \times n, \mathbb{R})$, d.h. A vermittelt die Abbildung

$$f_A : \mathbb{R}^n \rightarrow \mathbb{R}^m \tag{4.86}$$

$$x \mapsto Ax. \tag{4.87}$$

Sei $y \in \mathbb{R}^m$. Dann ist

$$\langle y, Ax \rangle = y^t (Ax) = (y^t A) x = (A^t y)^t x = \langle A^t y, x \rangle \tag{4.88}$$

oder in Bra-Ket-Notation

$$\langle y | A | x \rangle = \langle A^t y | x \rangle. \tag{4.89}$$

Mithilfe von A^t können wir also aus dem Funktional $\langle y|$ ein Funktional $\langle A^t y|$ erzeugen, das auf $|x\rangle$ genauso wirkt wie $\langle y|$ auf $A|x\rangle$.

Wenn wir quadratische Matrizen betrachten, so gibt es unter ihnen solche, die $A = A^t$, also $\langle y, Ax \rangle = \langle Ay, x \rangle$ erfüllen. Diese nennt man **symmetrische** Matrizen. Bei einer symmetrischen Matrix $A = (a_{ij})$ gilt für alle $i, j : a_{ij} = a_{ji}$.

Folgt aus $\langle y, Ax \rangle = \langle Ay, x \rangle$ für alle $x, y \in \mathbb{R}^n$ schon $A = A^t$? Ja, denn $\langle Ay, x \rangle = (Ay)^t x = y^t A^t x = \langle y, A^t x \rangle$, und es gilt der Satz:

Satz 4.10. Seien $A, B \in M(m \times n, \mathbb{R})$, sodass für alle $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ gilt $\langle y, Ax \rangle = \langle y, Bx \rangle$. Dann ist $A = B$.

Beweis. Wir setzen $x = e_j$ und $y = e_i$ ein, dann ist

$$\langle e_j, Ae_i \rangle = \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix} A \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = a_{ij}. \quad (4.90)$$

Analog ist $\langle e_j, Be_i \rangle = b_{ij}$. Da i, j beliebig gewählt werden können, gilt für alle i, j $a_{ij} = b_{ij}$ und somit $A = B$. □

Wir haben uns schon mit Drehungen im \mathbb{R}^2 beschäftigt. Dort haben wir Drehungen und Drehspiegelungen aus der Forderung gefunden, dass Längen und Winkel erhalten sein sollen, was wir zu der Forderung umformuliert haben, dass das Skalarprodukt erhalten sein soll.

Analog suchen wir jetzt nach linearen Abbildungen im $\mathbb{R}^n, x \mapsto Ox$, mit

$$\forall x, y \in \mathbb{R}^n : \langle Oy, Ox \rangle = \langle y, x \rangle. \quad (4.91)$$

Nun ist $\langle Oy, Ox \rangle = \langle y, O^t O x \rangle$, d.h. wir suchen Matrizen O mit

$$O^t O = E_n. \quad (4.92)$$

Eine quadratische $n \times n$ -Matrix mit dieser Eigenschaft nennt man eine **orthogonale Matrix**.

Wir erwarten, dass die Hintereinanderausführung von solchen Operationen wieder das Skalarprodukt erhält. In der Tat ist das Produkt von orthogonalen Matrizen wieder orthogonal:

$$(O_1 O_2)^t O_1 O_2 = O_2^t \underbrace{O_1^t O_1}_{E_n} O_2 = O_2^t O_2 = E_n. \quad (4.93)$$

Wir sind versucht zu vermuten, dass die Menge aller orthogonalen $n \times n$ -Matrizen eine Gruppe bildet. Wie wir gleich sehen werden, ist dies auch so, sie bilden **die orthogonale Gruppe**:

$$\mathcal{O}(n) = \{O \in M(n \times n, \mathbb{R}) \mid O^t O = E_n\}. \quad (4.94)$$

Wir haben schon gesehen, dass durch die Matrizenmultiplikation eine assoziative Verknüpfung auf dieser Menge definiert ist. Ein neutrales Element gibt es auch, nämlich E_n , da

$$E_n^t E_n = E_n E_n = E_n. \quad (4.95)$$

Wie steht es um das inverse Element? Zu $O \in \mathcal{O}(n)$ gibt es per Definition ein linksinverses Element O^t , da $O^t O = E_n$. Können wir daraus schließen, dass $O O^t = E_n$? Wenn wir wüssten, dass O^t selbst wieder orthogonal ist, also $O^t \in \mathcal{O}(n)$, wären wir auch fertig, da dann $(O^t)^t O^t = O O^t = E_n$, aber das wollen wir ja gerade zeigen.

Man ist womöglich versucht, einfach den Ausdruck $O^t O = E_n$ zu transponieren, aber das führt nicht weiter, denn

$$(O^t O)^t = O^t (O^t)^t = O^t O \tag{4.96}$$

ist. (Wir bemerken nebenbei, dass jede Matrix der Form $A^t A$ symmetrisch ist.)

Wir müssen also etwas mehr arbeiten und überlegen, unter welchen Umständen eine Matrix ein multiplikatives Inverses besitzt, also invertierbar ist.

Seien $A, B \in M(n \times n, \mathbb{R})$ mit $AB = E_n$. Was wissen wir dann über A und B ? Wir erinnern uns, dass der Rang bei der Komposition von linearen Abbildungen nicht größer werden kann. Da $\text{rang } E_n = \text{rang id}_{\mathbb{R}^n} = n$, folgt $\text{rang } A \geq n$ und damit $\text{rang } A = n$ und ebenso $\text{rang } B = n$.

Aus $AB = E_n$ folgt also $\text{rang } A = \text{rang } B = n$.

Eine lineare Abbildung $F : V \rightarrow V$ mit $\text{rang } F = \dim V$ ist aber injektiv und surjektiv, also bijektiv (ein Isomorphismus) und somit invertierbar, d.h. es existiert eine eindeutige Umkehrabbildung $F^{-1} : V \rightarrow V$ (man mache sich klar, dass diese wieder linear ist) mit $F \circ F^{-1} = F^{-1} \circ F = \text{id}_V$.

Übertragen auf Matrizen bedeutet dies, dass es zu einer linearen Abbildung auf dem \mathbb{R}^n , $x \mapsto Ax$, mit $\text{rang } A = n$, eine Umkehrabbildung $x \mapsto Bx$ gibt mit $AB = E_n$ und $BA = E_n$.

Wir fassen zusammen:

Satz 4.11.

1. Sei $A \in M(n \times n, \mathbb{R})$ mit $\text{rang } A = n$. Dann existiert eine eindeutige Matrix $B \in M(n \times n, \mathbb{R})$ mit $BA = AB = E_n$. Wir nennen B die **zu A inverse Matrix** und schreiben

$$A^{-1} := B. \tag{4.97}$$

2. Seien $A, B \in M(n \times n, \mathbb{R})$ mit $AB = E_n$. Dann ist $B = A^{-1}$ und somit gilt auch $BA = E_n$.

Matrizen, die eine inverse Matrix besitzen, nennen wir **invertierbar**. Offenbar sind die invertierbaren Matrizen genau diejenigen quadratischen Matrizen mit maximalem Rang.

Damit haben wir auch gezeigt, dass aus $O^t O = E_n$ folgt, dass $O O^t = E_n$. Für $O \in \mathcal{O}(n)$ ist also O^t das inverse Element:

$$O \text{ orthogonal} \iff O^t O = E_n \iff O O^t = E_n \iff O^{-1} = O^t, \tag{4.98}$$

insbesondere haben wir nachgewiesen, dass $\mathcal{O}(n)$ eine Gruppe ist.

Die Menge aller invertierbaren Matrizen nennt man die **allgemeine lineare Gruppe** (general linear group):

$$GL(n, \mathbb{R}) = \{A \in M(n \times n, \mathbb{R}) \mid A \text{ invertierbar}\}. \quad (4.99)$$

Wieder sollten wir kurz innehalten, um zu verstehen, warum diese Menge eine Gruppe ist. Zunächst scheint es offensichtlich, da wir wissen dass

- die Matrizenmultiplikation assoziativ ist,
- das neutrale Element E_n in $GL(n, \mathbb{R})$ enthalten ist ($E_n^{-1} = E_n$)
- und per Definition alle Elemente invertierbar sind.

Wir müssen aber noch feststellen, ob die Matrizenmultiplikation überhaupt eine Verknüpfung auf $GL(n, \mathbb{R})$ darstellt, dass also das Produkt zweier invertierbarer Matrizen wieder invertierbar ist.

Das ist der Fall, da für $A, B \in GL(n, \mathbb{R})$ gilt

$$(B^{-1}A^{-1})(AB) = B^{-1}(\underbrace{A^{-1}A}_{E_n})B = B^{-1}B = E_n, \quad (4.100)$$

also ist (AB) invertierbar und es gilt

$$(AB)^{-1} = B^{-1}A^{-1}. \quad (4.101)$$

(Aus der Sicht der linearen Abbildungen ist dies auch klar, da die Komposition bijektiver Abbildungen wieder bijektiv ist.)

Die Transposition ist auch als Operation auf $GL(n, \mathbb{R})$ definiert: Zu $A \in GL(n, \mathbb{R})$ ist auch A^t invertierbar, da

$$(A^{-1})^t A^t = (A A^{-1})^t = E_n^t = E_n, \quad (4.102)$$

also ist

$$(A^t)^{-1} = (A^{-1})^t. \quad (4.103)$$

Die Regeln für die Inversenbildung fassen wir noch einmal zusammen:

Satz 4.12. *Für $A, B \in GL(n, \mathbb{R})$ gilt*

- $(A^{-1})^{-1} = A$,
- $(AB)^{-1} = B^{-1}A^{-1}$,
- $(A^t)^{-1} = (A^{-1})^t$,
- $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$ für $\lambda \in \mathbb{R} \setminus \{0\}$.

$S_a(\lambda)$ ist invertierbar mit $S_a(\lambda)S_a(\frac{1}{\lambda}) = E_n$.

Diese Elementarmatrizen erzeugen $GL(n, \mathbb{R})$ in folgendem Sinn:

Satz 4.13. *Zu jeder Matrix $A \in GL(n, \mathbb{R})$ gibt es Elementarmatrizen A_1, \dots, A_s der obigen Form mit $A = A_1 \dots A_s$.*

Beweis. Zunächst bringen wir A auf Zeilenstufenform. Der Rang ist durch die Zahl der Pivot-Elemente gegeben. Wenn A invertierbar ist, ist der Rang maximal und wir erreichen eine obere Dreiecksmatrix mit von Null verschiedenen Diagonalelementen:

$$\tilde{A}_1 = \begin{pmatrix} \tilde{a}_{11} & & \star \\ & \ddots & \\ 0 & & \tilde{a}_{nn} \end{pmatrix} = \underbrace{B_{s_1} \dots B_1}_{\text{Elementarmatrizen}} A \quad (4.109)$$

mit beliebigen Einträgen \star . \tilde{A}_1 geht aus A durch Multiplikation mit Elementarmatrizen hervor.

Durch Umformungen vom Typ 3 erreichen wir die Form

$$\tilde{A}_2 = \begin{pmatrix} 1 & & \star \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \underbrace{B_{s_2} \dots B_{s_1+1}}_{\substack{\text{Elementarmatrizen} \\ \text{vom Typ 3}}} \tilde{A}_1 = B_{s_2} \dots B_1 A. \quad (4.110)$$

Durch weitere elementare Umformungen können wir \tilde{A}_2 zur Einheitsmatrix umformen, indem wir die 1 an Position i, i benutzen, um durch geeignete Addition der i -ten Zeile zu den Zeilen j mit $j < i$ die Einträge in der i -ten Spalte oberhalb von i, i zu Null zu machen. Wir erreichen schließlich:

$$\tilde{A}_3 = E_n = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = B_s \dots B_{s_2+1} \tilde{A}_2 = B_s \dots B_1 A \quad (4.111)$$

$$\implies A = B_1^{-1} B_2^{-1} \dots B_s^{-1}. \quad (4.112)$$

Da die Inversen der Elementarmatrizen wieder Elementarmatrizen sind, ist die Behauptung bewiesen. \square

In dem eben geführten Beweis steckt ein Verfahren zur Bestimmung der inversen Matrix, denn nach den dort erfolgten Umformungen ergibt sich

$$A^{-1} = B_s \dots B_1. \quad (4.113)$$

Um A^{-1} zu erhalten, müssen wir also alle Operationen, die wir auf A anwenden, um zu E_n zu gelangen, parallel auf E_n anwenden:

$$E_n = B_s \dots B_1 A \iff A^{-1} = B_s \dots B_1 E_n \quad (4.114)$$

Während des Verfahrens überprüft man, ob A überhaupt invertierbar ist: Wenn die Zeilenstufenform erreicht ist, erkennt man die Invertierbarkeit daran, dass alle Diagonalelemente von Null verschieden sind.

Wir führen das Verfahren an einem Beispiel durch:

Beispiel 4.5.

$$\begin{array}{l}
 A = \begin{pmatrix} 0 & 1 & -4 \\ 1 & 2 & -1 \\ 1 & 1 & 2 \end{pmatrix} \quad \left| \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_3 \\
 \\
 \begin{array}{l} II \\ I \end{array} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -4 \\ 1 & 1 & 2 \end{pmatrix} \quad \left| \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\
 \\
 III - I \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -4 \\ 0 & -1 & 3 \end{pmatrix} \quad \left| \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix} \\
 \\
 III + II \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -4 \\ 0 & 0 & -1 \end{pmatrix} \quad \left| \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{pmatrix} \quad \Rightarrow A \text{ ist invertierbar} \\
 \\
 -III \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix} \quad \left| \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 1 & -1 \end{pmatrix} \\
 \\
 I - 2II \begin{pmatrix} 1 & 0 & 7 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix} \quad \left| \quad \begin{pmatrix} -2 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 1 & -1 \end{pmatrix} \\
 \\
 I - 7III \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix} \quad \left| \quad \begin{pmatrix} 5 & -6 & 7 \\ 1 & 0 & 0 \\ -1 & 1 & -1 \end{pmatrix} \\
 \\
 II + 4III \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \left| \quad \begin{pmatrix} 5 & -6 & 7 \\ -3 & 4 & -4 \\ -1 & 1 & -1 \end{pmatrix} = A^{-1}
 \end{array}
 \right.
 \end{array}$$

Durch Multiplikation kann man leicht testen, ob man richtig gerechnet hat:

$$AA^{-1} = E_3.$$

Spaltentransformationen

Statt elementarer Zeilenumformungen können wir entsprechende elementare **Spaltenumformungen** durchführen. Aus der bisherigen Diskussion sollte klar sein, dass elementare Spaltenumformungen durch Multiplikation mit Elementarmatrizen von rechts realisiert werden können:

$$A \xrightarrow{\substack{\text{elementare} \\ \text{Spaltenumf.}}} \tilde{A} = A \underbrace{\tilde{E}_n}_{\text{Elementarmatrix}} \quad (4.115)$$

Das Verfahren zur Inversion könnte man also genauso gut mit Spaltenumformungen durchführen (aber nicht durcheinander mit Zeilenumformungen verwenden!).

Ein praktischerer Nutzen liegt in der eleganten Bestimmung einer Basis des Bildes, da sich das Bild unter einer Spaltenumformung nicht ändert:

$$\text{Im } A = \text{Im } (AB) \quad (B \text{ invertierbar}). \quad (4.116)$$

Bringt man A auf Spaltenstufenform, kann man eine Basis leicht ablesen als

$$\left(\begin{array}{cccc} & & & \\ \star & & & \\ & \star & & \\ & & \star & \\ & & & 0 \end{array} \right).$$

Basis des Bildes

Das geht nicht durch Zeilenumformungen, da im Allgemeinen $\text{Im } A \neq \text{Im } (BA)$. (Hier kann man nur, wie wir gesehen hatten, indirekt eine Basis des Bildes bestimmen: Man nehme die Spalten von A , die den Spalten der in der Zeilenstufenform \tilde{A} auftretenden Pivot-Elemente entsprechen.)

Die Spaltenstufenform kann auch erreicht werden dadurch, dass man die Matrix zuerst transponiert, dann auf Zeilenstufenform bringt und dann wieder transponiert:

$$A \xrightarrow{\text{Transponieren}} A^t \xrightarrow{\text{Zeilenumformung}} \underbrace{B}_{\text{Zusammenfassung aller Zeilenumf.}} A^t \xrightarrow{\text{Transponieren}} A \underbrace{B^t}_{\text{Zusammenfassung aller Spaltenumf.}}$$

4.2.1 Komplexe Matrizen

Alles, was wir bisher gemacht haben, lässt sich auf Matrizen mit komplexen Einträgen übertragen mit der Ausnahme der Dinge, die wir mit Bezug auf das kanonische Skalarprodukt besprochen haben, da dieses im \mathbb{C}^n anders definiert ist. Das heißt

- Matrizenaddition

- Matrizenmultiplikation
- Transposition
- Invertierbarkeit
- Bestimmung der Inversen Matrix
- Begriff des Bildes und des Kerns, etc.

funktionieren ganz genauso, nur dass bei der Skalarmultiplikation nun immer komplexe Zahlen verwendet werden. Dagegen werden wir sehen, dass die Begriffe symmetrische und orthogonale Matrix durch neue Konzepte abgelöst (bzw. ergänzt) werden.

Das kanonische Skalarprodukt auf dem \mathbb{C}^n sieht wie folgt aus:

$$\langle x, y \rangle_c = \sum_{i=1}^n \bar{x}_i y_i = \bar{x}^t y, \quad (4.117)$$

wobei der erste Eintrag komplex konjugiert wird.

Warum wählt man diese Definition? Man will Vektoren wieder eine reelle, nichtnegative Länge zuordnen. In $\mathbb{C} = \mathbb{C}^1$ ist dies der Absolutbetrag, $z \mapsto |z| = \sqrt{\bar{z}z}$. Im \mathbb{C}^n ist das Analogon $\|z\|^2 = \sum_{i=1}^n |z_i|^2 = \sum_{i=1}^n \bar{z}_i z_i$. Diese Länge bekommt man als Quadrat eines Skalarprodukts $\langle \cdot, \cdot \rangle_c$, wenn man es wie oben definiert:

$$\|z\|^2 = \langle z, z \rangle_c. \quad (4.118)$$

Wir notieren ein paar Eigenschaften des Skalarprodukts auf \mathbb{C}^n :

1. $\forall x, y, z \in \mathbb{C}^n, \lambda \in \mathbb{C}$:

$$\left. \begin{aligned} \langle x, y+z \rangle_c &= \langle x, y \rangle_c + \langle x, z \rangle_c \\ \langle x, \lambda y \rangle_c &= \lambda \cdot \langle x, y \rangle_c \end{aligned} \right\} \text{linear im 2. Eintrag}$$

$$\left. \begin{aligned} \langle x+y, z \rangle_c &= \langle x, z \rangle_c + \langle y, z \rangle_c \\ \langle \lambda x, y \rangle_c &= \bar{\lambda} \cdot \langle x, y \rangle_c \end{aligned} \right\} \text{semilinear im 1. Eintrag}$$

2. $\forall x, y \in \mathbb{C}^n$: $\langle x, y \rangle_c = \langle y, x \rangle_c$ hermitesch
3. $\forall x \in \mathbb{C}^n \setminus \{0\}$: $\langle x, x \rangle_c > 0$ ($\langle 0, 0 \rangle_c = 0$) positiv definit

Diese Eigenschaften weisen $\langle \cdot, \cdot \rangle_c$ als eine positiv definite, hermitesche Sesquilinearform (oder kürzer: positiv definite hermitesche Form) aus.

Mit diesem Skalarprodukt ist

$$\langle x, Ay \rangle_c = \bar{x}^t Ay = \overline{(\bar{A}^t x)}^t y = \langle \bar{A}^t x, y \rangle_c = \langle A^* x, y \rangle_c, \quad (4.119)$$

wobei wir hier einen weiteren Begriff eingeführt haben:

Für $A \in M(m \times n, \mathbb{C})$ nennen wir

$$A^* := \bar{A}^t \quad (4.120)$$

die **adjungierte Matrix**, also $A^* = (a'_{ij})$ mit $a'_{ij} = \bar{a}_{ji}$.

Quadratische Matrizen mit $A^* = A$, also $\langle Ax, y \rangle_c = \langle x, Ay \rangle_c$, nennen wir **hermitesche Matrizen**.

Matrizen U , die das komplexe Skalarprodukt erhalten, erfüllen

$$\langle Ux, Uy \rangle_c = \langle x, y \rangle_c, \quad (4.121)$$

und da $\langle Ux, Uy \rangle_c = \langle x, U^*Uy \rangle_c$ gilt, folgt $U^*U = E_n$, also

$$U^* = U^{-1}. \quad (4.122)$$

Solche Matrizen nennt man **unitäre Matrizen**.

Die Menge aller solcher $n \times n$ -Matrizen bilden die **unitäre Gruppe**

$$U(n) = \{U \in M(n \times n, \mathbb{C}) \mid U^*U = E_n\}. \quad (4.123)$$

Unitäre Abbildungen spielen in der Quanten- und Teilchenphysik eine große Rolle.

Im folgenden Abschnitt wollen wir uns noch einmal genauer dem Zusammenhang zwischen Matrizen und linearen Abbildungen widmen.

4.3 Basiswechsel und Normalform

Seien V, W endlichdimensionale Vektorräume und $F : V \rightarrow W$ eine lineare Abbildung. Wir wollen F durch eine Matrix darstellen.

Damit wir überhaupt eine Chance haben, dies zu tun, müssen wir Basen definieren. Seien also

$$\mathcal{A}_V = (v_1, \dots, v_n) \quad \text{und} \quad \mathcal{A}_W = (w_1, \dots, w_m) \quad (4.124)$$

Basen von V und W .

Die Bilder $F(v_j)$ der Basisvektoren haben eine eindeutige Zerlegung bezüglich \mathcal{A}_W , d.h. es gibt reelle Parameter a_{ij} mit

$$F(v_j) = a_{1j} \cdot w_1 + \dots + a_{mj} \cdot w_m = \sum_{i=1}^m a_{ij} \cdot w_i. \quad (4.125)$$

Die Koeffizienten legen die Abbildung komplett fest, denn

$$F(x_1 \cdot v_1 + \dots + x_n \cdot v_n) = F\left(\sum_{j=1}^n x_j \cdot v_j\right) = \sum_{j=1}^n x_j \cdot F(v_j) \quad (4.126)$$

$$= \sum_{j=1}^n \sum_{i=1}^m a_{ij} \cdot x_j \cdot w_i = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} \cdot x_j\right) \cdot w_i. \quad (4.127)$$

Jede lineare Abbildung ist von dieser Form, andererseits ist jede so definierte Abbildung linear.

Bei gegebenen Basen \mathcal{A}_V und \mathcal{A}_W haben wir also eine bijektive Abbildung zwischen den Matrizen $M(m \times n, \mathbb{R})$ und den linearen Abbildungen $\text{Hom}(V, W)$ gefunden.

Genauer: Für endlichdimensionale reelle Vektorräume V, W (mit $\dim V = n$, $\dim W = m$) mit Basen $\mathcal{A}_V, \mathcal{A}_W$ gibt es einen Isomorphismus

$$\begin{array}{ccc} M_{\mathcal{A}_W}^{\mathcal{A}_V} : \text{Hom}(V, W) & \xrightarrow{\text{Isomorphismus}} & M(m \times n, \mathbb{R}) \\ & \begin{array}{c} \sim \\ \longmapsto \end{array} & \\ & F & M_{\mathcal{A}_W}^{\mathcal{A}_V}(F), \end{array} \quad (4.128)$$

sodass für $M_{\mathcal{A}_W}^{\mathcal{A}_V}(F) = (a_{ij})$ gilt

$$F\left(\sum_{j=1}^n x_j \cdot v_j\right) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} \cdot x_j\right) \cdot w_i. \quad (4.129)$$

Wir können diesen Zusammenhang auch mit Hilfe der Basisisomorphismen $\Phi_{\mathcal{A}_V}, \Phi_{\mathcal{A}_W}$ schreiben

$$\Phi_{\mathcal{A}_V} : \mathbb{R}^n \longrightarrow V \quad \Phi_{\mathcal{A}_W} : \mathbb{R}^m \longrightarrow W \quad (4.130)$$

$$x \longmapsto \sum_{j=1}^n x_j \cdot v_j \quad y \longmapsto \sum_{i=1}^m y_i \cdot w_i \quad (4.131)$$

Dann gilt:

$$F' : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

$$x \longmapsto M_{\mathcal{A}_W}^{\mathcal{A}_V}(F) \cdot x$$

ist gegeben durch

$$F' = \Phi_{\mathcal{A}_W}^{-1} \circ F \circ \Phi_{\mathcal{A}_V}, \quad (4.132)$$

denn x verhält sich unter der zusammengesetzten Abbildung wie folgt:

- erst ordnen wir einem n -Tupel x von Koordinaten einen Vektor zu

$$\Phi_{\mathcal{A}_V}(x) = \sum_{j=1}^n x_j \cdot v_j, \quad (4.133)$$

- dieser Vektor wird durch F nach W abgebildet,

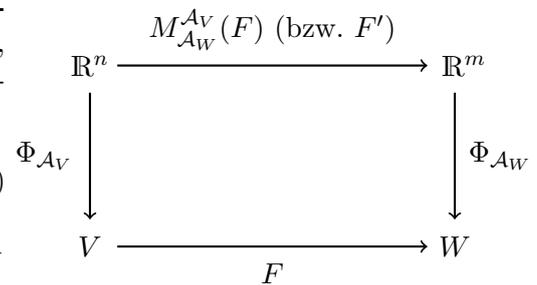
$$F(\Phi_{\mathcal{A}_V}(x)) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} \cdot x_j \right) \cdot w_i, \quad (4.134)$$

- dort ordnen wir dem resultierenden Vektor wieder seine Koordinaten bezüglich der gewählten Basis zu.

$$\Phi_{\mathcal{A}_W}^{-1}(F(\Phi_{\mathcal{A}_V}(x))) = \left(\sum_{j=1}^n a_{ij} \cdot x_j \right)_i = M_{\mathcal{A}_W}^{\mathcal{A}_V}(F) \cdot x. \quad (4.135)$$

Das können wir in einem **kommutativen Diagramm** ausdrücken — kommutativ bedeutet dabei, dass auf verschiedene Weise zusammengesetzte Abbildungen übereinstimmen, hier also

$$\Phi_{\mathcal{A}_W} \circ F' = F \circ \Phi_{\mathcal{A}_V}, \quad (4.136)$$



was äquivalent ist zu $F' = \Phi_{\mathcal{A}_W}^{-1} \circ F \circ \Phi_{\mathcal{A}_V}$, da $\Phi_{\mathcal{A}_W}$ ein Isomorphismus ist.

Beispiel 4.6. Wir betrachten den Vektorraum der Lösungen der Bewegungsgleichungen des harmonischen Oszillators mit Frequenz $\omega \in \mathbb{R}$, den wir erkannt hatten als

$$U_\omega = \{q \in \text{Abb}(\mathbb{R}, \mathbb{R}) \mid q : t \mapsto q(t) = a \cos \omega t + b \sin \omega t, \ a, b \in \mathbb{R}\}. \quad (4.137)$$

Auf U_ω betrachten wir die Abbildung $D : q \mapsto \dot{q}$. Nun ist

$$\dot{q}(t) = -\omega a \sin \omega t + \omega b \cos \omega t, \quad (4.138)$$

d.h. bezüglich der Basis $\mathcal{A}_{U_\omega} = (t \mapsto \cos \omega t, t \mapsto \sin \omega t)$ hat D die folgende Wirkung

auf die Koordinaten

$$\begin{pmatrix} a \\ b \end{pmatrix} \mapsto \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}. \quad (4.139)$$

D wird also dargestellt durch die Matrix

$$M_{\mathcal{A}_{U_\omega}}^{\mathcal{A}_{U_\omega}}(D) = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}. \quad (4.140)$$

Die Wirkung von D^2 (zweite Ableitung) wird in dieser Basis zu

$$M_{\mathcal{A}_{U_\omega}}^{\mathcal{A}_{U_\omega}}(D^2) = (M_{\mathcal{A}_{U_\omega}}^{\mathcal{A}_{U_\omega}}(D))^2 = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}^2 = -\omega^2 \cdot E_2, \quad (4.141)$$

d.h. $M_{\mathcal{A}_{U_\omega}}^{\mathcal{A}_{U_\omega}}(D^2 + \omega^2 \cdot id_{U_\omega}) = 0$.

Das muss gelten, da alle Funktionen $q \in U_\omega$ die Bewegungsgleichung $\ddot{q} = -\omega^2 q$ erfüllen.

$M_{\mathcal{A}_W}^{\mathcal{A}_V}(F)$ liefert uns eine Darstellung von F bezüglich der Basen \mathcal{A}_V und \mathcal{A}_W . Wie verändert sich diese Matrixdarstellung, wenn wir die Basis wechseln?

Wählen wir also eine andere Basis in V , $\mathcal{B}_V = (v'_1, \dots, v'_n)$. Die alte Basis können wir auf eindeutige Weise bezüglich der neuen darstellen:

$$v_k = \sum_{l=1}^n t_{lk} \cdot v'_l. \quad (4.142)$$

Andererseits können wir die neue durch die alte Basis ausdrücken:

$$v'_l = \sum_{j=1}^n t'_{jl} \cdot v_j. \quad (4.143)$$

Setzen wir diesen Ausdruck oben ein, so erhalten wir

$$v_k = \sum_{l=1}^n \sum_{j=1}^n t_{lk} \cdot t'_{jl} \cdot v_j. \quad (4.144)$$

Da die v_j linear unabhängig sind, folgt

$$\sum_{l=1}^n t_{lk} \cdot t'_{jl} = \delta_{jk}, \quad (4.145)$$

also $T'T = E_n$, wobei $T = (t_{lk})$, $T' = (t'_{jl})$, d.h. $T' = T^{-1}$.

Ein gegebener Vektor $\sum_{l=1}^n x_l \cdot v_l$ in V mit Koordinaten $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ hat andere Koordinaten bezüglich der neuen Basis:

$$\sum_{l=1}^n x_l \cdot v_l = \sum_{j=1}^n x'_j \cdot v'_j = \sum_{j=1}^n \sum_{l=1}^n x'_j \cdot t'_{lj} \cdot v_l = \sum_{l=1}^n \left(\sum_{j=1}^n t'_{lj} \cdot x'_j \right) \cdot v_l. \quad (4.146)$$

Also $x = T^{-1}x'$ und $x' = Tx$.

Bemerkung

Wenn wir uns darauf einlassen, dass Vektoren auch vektorwertige Einträge haben dürfen und wir Skalarmultiplikation auch von rechts definieren, können wir den Basisisomorphismus auch formal schreiben als

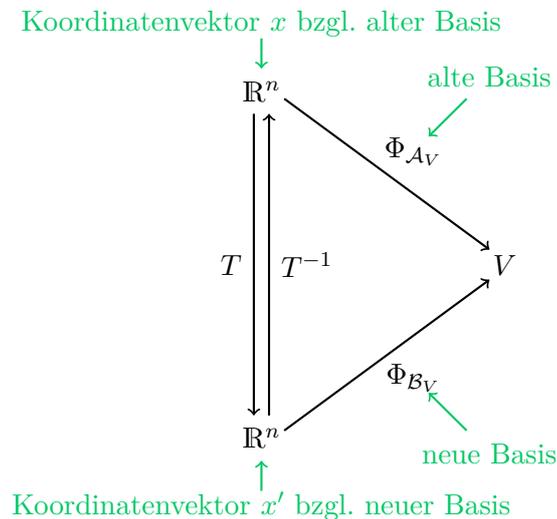
$$\Phi_{\mathcal{A}_V}(x) = \mathcal{A}_V \cdot x = (v_1, \dots, v_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 \cdot v_1 + \dots + x_n \cdot v_n. \quad (4.147)$$

Der Zusammenhang zwischen den Basen ist dann $\mathcal{A}_V = \mathcal{B}_V \cdot T$ und aus $\Phi_{\mathcal{A}_V}(x) = \Phi_{\mathcal{B}_V}(x')$ folgt dann $\mathcal{A}_V \cdot x = \mathcal{B}_V \cdot T \cdot x = \mathcal{B}_V \cdot x'$ und somit $x' = T \cdot x$.

Das können wir uns wieder mit Hilfe eines kommutierenden Diagramms klarmachen.

$$\Phi_{\mathcal{B}_V}^{-1} \circ \Phi_{\mathcal{A}_V} : x \mapsto T \cdot x$$

$$\Phi_{\mathcal{A}_V}^{-1} \circ \Phi_{\mathcal{B}_V} : x' \mapsto T^{-1} \cdot x'$$



Bemerkung

Wenn wir an einen Abbildungspfeil zwischen Vektorräumen \mathbb{R}^n eine Matrix schreiben, meinen wir immer die durch diese Matrix vermittelte lineare Abbildung.

Für eine Basistransformation von \mathcal{A}_W zu \mathcal{B}_W im Vektorraum W wollen wir die zugehörige Matrix, welche die Koordinatenvektoren transformiert, mit S bezeichnen,

$$\Phi_{\mathcal{B}_W}^{-1} \circ \Phi_{\mathcal{A}_W} : y \mapsto S \cdot y. \quad (4.148)$$

Nun können wir auch verstehen, wie sich die darstellende Matrix $A = M_{\mathcal{A}_W}^{\mathcal{A}_V}(F)$ unter einer Basistransformation verändert, denn die neue Matrix $B = M_{\mathcal{B}_W}^{\mathcal{B}_V}(F)$ gehört zur

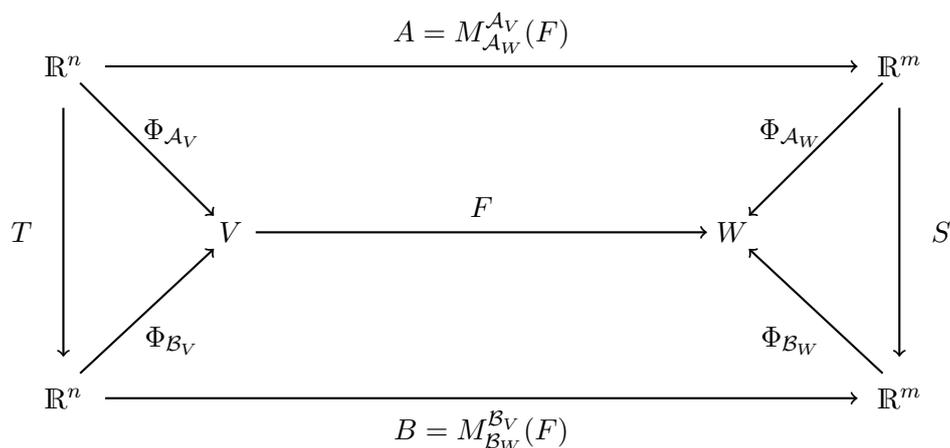
linearen Abbildung

$$\Phi_{\mathcal{B}_W}^{-1} \circ F \circ \Phi_{\mathcal{B}_V} = (\Phi_{\mathcal{B}_W}^{-1} \circ \Phi_{\mathcal{A}_W}) \circ (\Phi_{\mathcal{A}_W}^{-1} \circ F \circ \Phi_{\mathcal{A}_V}) \circ (\Phi_{\mathcal{A}_V}^{-1} \circ \Phi_{\mathcal{B}_V}), \quad (4.149)$$

d.h.

$$B = M_{\mathcal{B}_W}^{\mathcal{B}_V}(F) = S \cdot A \cdot T^{-1}. \quad (4.150)$$

Als kommutatives Diagramm stellt sich das wie folgt dar:



Zu einer gegebenen Abbildung möchte man nun gerne Basen finden, sodass die zugehörige Matrix möglichst einfach wird.

Satz 4.14. Sei $F : V \rightarrow W$ eine lineare Abbildung mit $\text{rang } F = r$. Dann gibt es Basen $\mathcal{B}_V, \mathcal{B}_W$ von V und W mit

$$M_{\mathcal{B}_W}^{\mathcal{B}_V}(F) = \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \vdots \\ & & & 0 & \ddots \\ 0 & \dots & & & 0 \end{pmatrix} \quad (4.151)$$

mit einer $r \times r$ -Einheitsmatrix in der linken oberen Ecke der Matrix $M_{\mathcal{B}_W}^{\mathcal{B}_V}(F)$. Diese Matrixdarstellung von F nennen wir **Normalform** von F .

Beweis. Seien $\mathcal{A}_V, \mathcal{A}_W$ Basen von V und W mit $M_{\mathcal{A}_W}^{\mathcal{A}_V}(F) = A$. Durch elementare

Zeilenumformungen bringe man A auf die normierte Zeilenstufenform

$$\tilde{A}_1 = \begin{pmatrix} 1 & \star & \cdots & & \\ & & & 1 & \\ & & & & \ddots \\ 0 & & & & & 1 \end{pmatrix} = SA \quad (4.152)$$

mit $S \in GL(m, \mathbb{R})$. Durch Vertauschung der Spalten bringe man \tilde{A}_1 zunächst zu

$$\tilde{A}_2 = \begin{pmatrix} 1 & \star & & \star \\ & \ddots & \star & \\ & & 1 & \star \\ & \cdots & 0 & 0 \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \quad (4.153)$$

und dann durch Addition von Vielfachen von Spalten zu

$$\tilde{A}_3 = \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \vdots \\ & & 1 & & \\ & & & 0 & \\ 0 & & \cdots & \ddots & 0 \end{pmatrix} = SAT^{-1} \quad \text{mit } T \in GL(n, \mathbb{R}). \quad (4.154)$$

Da S und T invertierbar sind, ist der Rang von \tilde{A}_3 gleich dem Rang von A , also $\text{rang } \tilde{A}_3 = r$, d.h. die Zahl der auftretenden Einsen ist r .

S und T entsprechen einem Wechsel der Basen, sodass SAT^{-1} die Darstellung von F bezüglich der neuen Basis ist. \square

Der Beweis liefert auch gleich ein konstruktives Verfahren zur Bestimmung der nötigen Basistransformationen: Man beginne mit den drei Matrizen E_n, A, E_n und führe alle Zeilenumformungen, die wir an A ausführen, auch an der ersten Einheitsmatrix durch

$$E_n, A, E_n \mapsto SE_n, SA, E_n. \quad (4.155)$$

Dann führt man alle Spaltenumformungen an SA auch an der zweiten Einheitsmatrix durch und erhält

$$S, \underbrace{SAT^{-1}}_{\text{Normalform}}, T^{-1}. \quad (4.156)$$

Der Beweis kann auch statt durch Umformung der entsprechenden Matrizen direkt über die Wahl einer geschickten Basis geführt werden. Nach dem Basisergänzungssatz gibt es zu einer Basis (w_1, \dots, w_r) von $\text{Im } F$ eine Basis $\mathcal{B}_W = (w_1, \dots, w_m)$ von W . Dann gibt es eine Basis $\mathcal{B}_V = (v_1, \dots, v_n)$, sodass $F(v_1) = w_1, \dots, F(v_r) = w_r$, und (v_{r+1}, \dots, v_n) bildet eine Basis von $\text{Ker } F$. Da die Spaltenvektoren von $M_{\mathcal{B}_W}^{\mathcal{B}_V}(F)$ die Koordinatenvektoren der Bilder der Basisvektoren sind, folgt

$$M_{\mathcal{B}_W}^{\mathcal{B}_V}(F) = \begin{pmatrix} 1 & & & & 0 \\ & \ddots & & & \\ & & 1 & & \vdots \\ & & & 0 & \\ 0 & \dots & & \ddots & 0 \end{pmatrix}. \quad (4.157)$$

Bemerkung

Wir sehen aus dem Satz auch sofort, dass $\text{rang } A = \text{rang } A^t$ für beliebige Matrizen A gilt, denn sei SAT^{-1} in Normalform wie oben, dann ist

$$\begin{aligned} \text{rang } A &= \text{rang } SAT^{-1} && \text{(der Rang ändert sich nicht)} \\ &= \text{rang } (SAT^{-1})^t && \text{(offensichtlich, da } SAT^{-1} \text{ in Normalform)} \\ &= \text{rang } ((T^{-1})^t A^t S^t) \\ &= \text{rang } A^t && \text{(Multiplikation mit invertierbaren Matrizen).} \end{aligned}$$

Wir haben den Rang einer linearen Abbildung als Dimension des Bildes definiert, für eine Matrix führt dies auf die Dimension des von den Spaltenvektoren erzeugten Raums bzw. auf die maximale Zahl linear unabhängiger Spaltenvektoren. Genauso könnte man den *Zeilenrang* als maximale Zahl linear unabhängiger Zeilen einführen. Aus dem obigen Resultat ist aber klar, dass diese Zahl identisch mit dem Rang der Matrix ist.

Zu jeder linearen Abbildung zwischen endlichdimensionalen Vektorräumen gibt es also Basen, sodass die Matrixdarstellung sehr einfach wird. Komplizierter und interessanter wird es bei Abbildungen $V \rightarrow V$ (sogenannte Endomorphismen). Ein Basiswechsel führt dann von einer Matrixdarstellung A zu einer neuen Darstellung SAS^{-1} . Mit den dabei auftretenden Normalformen werden wir uns später beschäftigen.

Zusammenfassung Kapitel 4

Wir haben uns mit linearen Abbildungen (Homomorphismen) zwischen (reellen) Vektorräumen beschäftigt.

Dabei haben wir eine Reihe von Konzepten eingeführt, u.a.:

- Kern, Bild
- Matrizenaddition, -multiplikation, -transposition, -inversion
- Symmetrische, hermitesche, orthogonale, unitäre Matrizen
- Basisisomorphismus, Matrixdarstellung einer linearen Abbildung
- Basiswechsel

Wichtige Aussagen in diesem Zusammenhang sind u.a. ($F : V \rightarrow W$ linear):

- Bilder bzw. Urbilder von Untervektorräumen sind Untervektorräume
- Injektivität $\iff \dim \text{Ker } F = 0$, Surjektivität $\iff \text{rang } F = \dim W$
- Dimensionsformel: $\dim \text{Ker } F + \text{rang } F = \dim V$
- Invertierbarkeit einer Matrix $A \iff A$ hat maximalen Rang
- Gruppenstrukturen: invertierbare Matrizen, orthogonale bzw. unitäre Matrizen
- Zeilenumformungen \leftrightarrow Multiplikation mit Elementarmatrizen von links
- Spaltenumformungen \leftrightarrow Multiplikation mit Elementarmatrizen von rechts

Wir haben dabei auch ein paar konkrete Methoden besprochen:

- Inversion durch Zeilenumformung A, E zu E, A^{-1}
- Rangbestimmung aus Zeilenstufenform (oder Spaltenstufenform)

Insgesamt sollten Sie eine gewisse Vertrautheit mit linearen Abbildungen und ihren Eigenschaften gewonnen haben und in der Lage sein, mit Matrizen sicher umzugehen.

5 Lineare Gleichungssysteme und affine Unterräume

Dieses Kapitel beschäftigt sich mit inhomogenen linearen Gleichungssystemen

$$Ax = b \quad (x \in \mathbb{R}^n, b \in \mathbb{R}^m, A \in M(m \times n, \mathbb{R})) \quad (5.1)$$

und ihren Lösungen, also den Lösungsmengen

$$\text{Lös}(A, b) = \{x \in \mathbb{R}^n \mid Ax = b\}. \quad (5.2)$$

Für $b = 0$ (homogene lineare Gleichungssysteme) ist die Lösungsmenge ein Untervektorraum (der Kern von A), für $b \neq 0$ ist die Lösungsmenge ein affiner Unterraum des \mathbb{R}^n , und mit solchen Räumen wollen wir uns zunächst beschäftigen.

5.1 Affine Unterräume

Definition 5.1. Sei V ein reeller Vektorraum. $B \subset V$ heißt **affiner Unterraum** von V , wenn es einen Untervektorraum $U \subset V$ gibt, so dass gilt

$$\text{für alle } v, w \in B \text{ ist } v - w \in U, \quad (5.3)$$

$$\text{für alle } v \in B, u \in U \text{ ist } v + u \in B. \quad (5.4)$$

Mit anderen Worten: Wir fordern, dass es einen zugehörigen Untervektorraum U gibt, sodass zum einen die Differenz von je zwei Vektoren in B in U liegt und man zum anderen nichtaus B hinausgelangt, wenn man einen Vektor in B durch einen Vektor aus U verschiebt.

Beispiel 5.1.

- Eine Gerade $G_{v,t} = v + \mathbb{R} \cdot t$ im \mathbb{R}^n ($t \neq 0$). Der zugehörige Untervektorraum ist die parallele Gerade $\mathbb{R} \cdot t$ durch den Ursprung: Differenzen von Vektoren in $G_{v,t}$ liegen in $\mathbb{R} \cdot t$, und wenn man einen Vektor $v + \lambda \cdot t$ in $G_{v,t}$ um ein Element $\mu \cdot t$ von $\mathbb{R} \cdot t$, also entlang der Geraden, verschiebt, ist der resultierende Vektor $v + \lambda \cdot t + \mu \cdot t = v + (\lambda + \mu) \cdot t$ wieder Element der Geraden $G_{v,t}$.
- Eine Ebene $E_{v,t_1,t_2} = v + \mathbb{R} \cdot t_1 + \mathbb{R} \cdot t_2$, der zugehörige Untervektorraum ist die parallele Ebene $\mathbb{R} \cdot t_1 + \mathbb{R} \cdot t_2$ durch den Ursprung.
- Ein einzelner Vektor $\{v\} \subset V$, der zugehörige Untervektorraum ist der 0-dimensionale Raum $\{0\}$.
- Die Definition lässt auch die leere Menge $\{\} \subset V$ zu, der zugehörige Untervektorraum ist dabei unbestimmt.
- Jede Menge von der Form $v + U$ mit U Untervektorraum von V .

Satz 5.1. *Jeder nichtleere affine Unterraum B von einem Vektorraum V ist von der Form $B = v + U$ für ein $v \in V$.*

Beweis. Sei $B \subset V$ ein affiner Unterraum mit zugehörigem Untervektorraum U . Da $B \neq \emptyset$, gibt es ein $v \in B$. Wir zeigen nun, dass $B = v + U$.

Sei $w \in B$. Dann ist $u := w - v \in U$, also $w = v + u \in v + U$.

Sei andersherum $w \in v + U$, also $w = v + u$ für ein $u \in U$. Da B affiner Unterraum ist, ist $v + u \in B$ also $w \in B$. □

Affine Unterräume sind also (wenn sie nicht leer sind) einfach verschobene Untervektorräume. Sie sind im Allgemeinen nicht gegenüber Addition und Skalarmultiplikation abgeschlossen. Die Rolle des Untervektorraums ist die, dass jeder Vektor $u \in U$ eine Translation t_u auf B definiert:

$$t_u : B \rightarrow B \tag{5.5}$$

$$v \mapsto v + u \tag{5.6}$$

(Auf ähnliche Weise kann man auch abstrakt das Konzept eines affinen Raumes (statt eines Unterraumes) einführen: eine Menge, auf der ein Vektorraum durch eine Art Translation mit bestimmten Eigenschaften operiert.)

Der Vektorraum V selbst ist natürlich auch ein affiner Unterraum von sich selbst. Es ist sozusagen der Vektorraum, in dem wir die Lage des Ursprungs vergessen haben und uns nur noch für Differenzen von Vektoren interessieren. Der dreidimensionale Anschauungsraum ist in dem Sinne eigentlich eher ein affiner Raum, in dem die Translationen bzw. Verschiebungen einen Vektorraum bilden. Zum Vektorraum wird er erst, wenn wir einen Ursprung wählen.

Zu der affinen Struktur kann man dann einen passenden Abbildungsbegriff definieren, der diese Struktur erhält. Dies führt zu den sogenannten **affinen Abbildungen**:

Definition 5.2. *Seien V, W reelle Vektorräume. Eine Abbildung $\Phi : V \rightarrow W$ heißt affine Abbildung genau dann, wenn es einen Vektor $w_0 \in W$ und eine lineare Abbildung $F : V \rightarrow W$ gibt, sodass für alle $v \in V$ gilt:*

$$\Phi(v) = w_0 + F(v). \tag{5.7}$$

Das passt zu unserer anschaulichen Charakterisierung eines affinen Raums als einen Vektorraum ohne ausgezeichneten Ursprung: Zusätzlich zu linearen Abbildungen können wir die Punkte auch um einen konstanten Vektor w_0 verschieben.

Die Hintereinanderausführung von affinen Abbildungen ist wieder affin: Seien $\Phi : V \rightarrow W$ und $\Psi : W \rightarrow Z$ affine Abbildungen von der Form

$$\Phi(v) = w_0 + F(v) \quad \Psi(w) = z_0 + G(w). \tag{5.8}$$

Dann wirkt $\Psi \circ \Phi : V \rightarrow Z$ als

$$(\Psi \circ \Phi)(v) = \Psi(\Phi(v)) = \Psi(w_0 + F(v)) \quad (5.9)$$

$$= z_0 + G(w_0 + F(v)) \quad (5.10)$$

$$= z_0 + G(w_0) + G(F(v)) \quad (5.11)$$

$$= \underbrace{(z_0 + G(w_0))}_{\text{konstant}} + \underbrace{G \circ F(v)}_{\text{linear}}. \quad (5.12)$$

Bemerkung

Im \mathbb{R}^n mit dem kanonischen Skalarprodukt $\langle \cdot, \cdot \rangle$ gibt es spezielle affine Abbildungen, die alle Abstände erhalten. Eine solche Abbildung Φ von der Form $\Phi(x) = x_0 + Ax$ ist also charakterisiert durch die Eigenschaft

$$\|\Phi(x_1) - \Phi(x_2)\| = \|x_1 - x_2\| \iff \|A(x_1 - x_2)\| = \|x_1 - x_2\| \quad (5.13)$$

Die zugehörige Matrix A erhält die Länge von Vektoren und muss damit orthogonal sein: $A^t A = E_n$. Denn

$$\langle Ax_1, Ax_2 \rangle = \frac{1}{4}(\|A(x_1+x_2)\|^2 - \|A(x_1-x_2)\|^2) = \frac{1}{4}(\|x_1+x_2\|^2 - \|x_1-x_2\|^2) = \langle x_1, x_2 \rangle. \quad (5.14)$$

Eine solche affine Abbildung, $\Phi(x) = x_0 + Ox$ mit O orthogonal, heißt **Euklidische Bewegung**. Mit der Hintereinanderausführung von Abbildungen bildet die Menge der euklidischen Bewegungen eine Gruppe, die **Euklidische Gruppe** $E(n)$ (manchmal auch $ISO(n)$ genannt).

Kommen wir zurück zu den Gleichungssystemen und betrachten zunächst den allgemeineren Fall der Gleichung $F(v) = w$.

Satz 5.2. Sei $F : V \rightarrow W$ eine lineare Abbildung zwischen reellen Vektorräumen. Sei $w \in W$. Dann ist $F^{-1}(\{w\})$ ein affiner Unterraum von V mit zugehörigem Untervektorraum $U = \text{Ker } F$.

Beweis. Falls $F^{-1}(\{w\}) = \emptyset$, es also keine Lösung gibt, so ist $F^{-1}(\{w\})$ ein trivialer affiner Unterraum. Ansonsten sei $v \in F^{-1}(\{w\})$, also $F(v) = w$. Für jedes weitere $v' \in F^{-1}(\{w\})$ gilt $F(v') = w$, also $F(v' - v) = 0$. Dann ist

$$F^{-1}(\{w\}) = \{v' \in V \mid F(v') = w\} \quad (5.15)$$

$$= \{v' \in V \mid F(v' - v) = 0\} \quad (5.16)$$

$$= \{v + u \in V \mid F(u) = 0\} \quad (5.17)$$

$$= v + \text{Ker } F. \quad (5.18)$$

Da $\text{Ker } F$ ein Untervektorraum ist, ist $F^{-1}(\{w\})$ ein affiner Unterraum.

□

Damit ist insbesondere auch klar, dass der Lösungsraum eines inhomogenen linearen Gleichungssystems ein affiner Unterraum ist: Falls zu $Ax = b$ überhaupt eine Lösung v existiert (also $Av = b$), so ist

$$\text{Lös}(A, b) = v + \text{Ker } A \tag{5.19}$$

$$= v + \text{Lös}(A, 0). \tag{5.20}$$

Die allgemeine Lösung eines inhomogenen linearen Gleichungssystems ergibt sich also als Summe einer speziellen Lösung des inhomogenen Systems und der allgemeinen Lösung des zugehörigen homogenen Gleichungssystems.

5.2 Lösung inhomogener linearer Gleichungssysteme

Jetzt, wo wir die Struktur der Lösungsmenge verstanden haben, wenden wir uns der konkreten Lösungsstrategie zu.

Die Strategie ist ähnlich wie bei den homogenen Systemen: Wir formen die Gleichungen so um, dass wir sie rekursiv lösen können (oder erkennen, dass keine Lösung existiert).

Sei also die Gleichung $Ax = b$ gegeben; ausgeschrieben ergibt sich das Gleichungssystem

$$a_{11} \cdot x_1 + \dots + a_{1n} \cdot x_n = b_1 \quad (5.21)$$

$$\vdots \quad \quad \quad \vdots \quad (5.22)$$

$$a_{m1} \cdot x_1 + \dots + a_{mn} \cdot x_n = b_n. \quad (5.23)$$

Die Vertauschung zweier Gleichungen, die Addition eines Vielfachen einer Gleichung zu einer anderen und die Multiplikation einer Gleichung mit einer von Null verschiedenen Zahl ändern die Lösungsmenge nicht und überführen das Gleichungssystem in ein äquivalentes System.

Wir können also die bekannte Strategie anwenden. Die Operationen übersetzen sich in elementare Zeilentransformationen der Koeffizientenmatrix A , aber gleichzeitig müssen wir die Transformationen am Vektor b durchführen. Dies können wir auch wie folgt verstehen:

Durch Zeilentransformationen bringen wir A auf Zeilenstufenform $\tilde{A} = SA$. Wir kommen dann zu der Gleichung

$$\tilde{A}x = SAx = Sb = \tilde{b}, \quad (5.24)$$

wobei \tilde{b} aus b durch die gleichen Zeilenumformungen hervorgeht.

Da S invertierbar ist, ist $Ax = b$ äquivalent zu $SAx = Sb$, also

$$\text{Lös}(A, b) = \text{Lös}(SA, Sb) = \text{Lös}(\tilde{A}, \tilde{b}). \quad (5.25)$$

Es ist daher praktisch, die erweiterte Matrix $A' = (A \mid b)$ mit b als Extra-Spalte einzuführen und dann A' auf Zeilenstufenform $SA' = (SA \mid Sb) = (\tilde{A} \mid \tilde{b})$ zu bringen. Dann ist \tilde{A} automatisch in Zeilenstufenform und b ist auch gleich mittransformiert worden.

In der Zeilenstufenform lässt sich die Gleichung dann leicht behandeln:

$$\tilde{A} = \left(\begin{array}{cccc} \star & \dots & & \\ & \star & & \\ & & \dots & \\ & 0 & & \star \end{array} \right) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \tilde{b}_1 \\ \vdots \\ \tilde{b}_r \\ \tilde{b}_{r+1} \\ \vdots \\ \tilde{b}_m \end{pmatrix}. \quad (5.26)$$

Offensichtlich hat die Gleichung höchstens dann eine Lösung, wenn $\tilde{b}_{r+1} = \dots = \tilde{b}_m = 0$ (wobei r der Rang von \tilde{A} bzw. A ist), d.h. wenn gilt $\text{rang } A = \text{rang } (A \mid b)$.

Wenn dies gegeben ist, dann kann man eine Lösung wie im homogenen Fall rekursiv konstruieren: Man parametrisiert die $f = n - r$ freien Variablen und drückt die r gebundenen Variablen (die den Spalten der Pivot-Elemente entsprechen) durch die freien Variablen und die \tilde{b}_i aus.

Die Lösung hat dann die Form

$$x = v + \lambda_1 u_1 + \dots + \lambda_f u_f, \quad (5.27)$$

wobei v die Lösung ist, die man erhält, wenn man alle freien Variablen zu Null setzt. Die Vektoren u_1, \dots, u_f bilden eine Basis von $\text{Ker } A$, dem Lösungsraum des zugehörigen homogenen Systems $Ax = 0$.

(u_1, \dots, u_f) heißt **Fundamentalsystem von Lösungen des homogenen Systems**.

Der Lösungsraum ist damit explizit von der Form eines affinen Raums

$$\text{Lös}(A, b) = v + \mathbb{R} \cdot u_1 + \dots + \mathbb{R} \cdot u_f = v + \text{Lös}(A, 0), \quad (5.28)$$

und er hat die Dimension

$$\dim \text{Lös}(A, b) = \dim \text{Lös}(A, 0) = \dim \text{Ker } A = n - \text{rang } A. \quad (5.29)$$

Wir fassen das Verfahren noch einmal kurz zusammen:

- Umschreiben des linearen Gleichungssystems als $Ax = b$.
- Überführung der erweiterten Matrix $(A \mid b)$ in Zeilenstufenform $(\tilde{A} \mid \tilde{b})$. Wenn $\text{rang } (A \mid b) = \text{rang } A = r$, dann existiert mindestens eine Lösung.
- Wir parametrisieren die $f = n - r$ freien Variablen durch Parameter $\lambda_1, \dots, \lambda_f$ und lösen rekursiv nach den gebundenen Variablen auf.

Beispiel 5.2.

$$\begin{aligned} x_1 + x_2 + x_3 &= 4 \\ x_1 + 2x_2 - x_3 &= 5 \\ x_1 - x_2 + 5x_3 &= 2 \end{aligned} \iff Ax = b \quad \text{mit } A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 5 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ 5 \\ 2 \end{pmatrix} \quad (5.30)$$

Wir bringen die erweiterte Matrix $(A|b)$ auf Zeilenstufenform:

$$(A|b) = \left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 1 & 2 & -1 & 5 \\ 1 & -1 & 5 & 2 \end{array} \right)$$

$$\begin{array}{l} II - I \\ III - I \end{array} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & -2 & 1 \\ 0 & -2 & 4 & -2 \end{array} \right)$$

$$III + 2 \cdot II \left(\begin{array}{ccc|c} 1 & 1 & 1 & 4 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right) = (\tilde{A}|\tilde{b})$$

Also haben A und $(A|b)$ beide Rang 2. Das Gleichungssystem ist somit lösbar, und der Lösungsraum hat Dimension $3 - 2 = 1$.

Für die konkrete Lösung schreiben wir uns $\tilde{A}x = \tilde{b}$ explizit hin:

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

(Die dritte Zeile von $(\tilde{A}|\tilde{b})$ führt zu einer trivialen Gleichung.) Wir parametrisieren x_3 durch $x_3 = \lambda$ und lösen rekursiv nach den anderen Variablen auf:

$$\begin{aligned} x_2 &= 1 + 2x_3 = 1 + 2\lambda \\ x_1 &= 4 - x_2 - x_3 = 4 - (1 + 2\lambda) - \lambda = 3 - 3\lambda. \end{aligned}$$

Die allgemeine Lösung ist also

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 - 3\lambda \\ 1 + 2\lambda \\ \lambda \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + \lambda \cdot \begin{pmatrix} -3 \\ 2 \\ 1 \end{pmatrix}, \quad (5.31)$$

und der Lösungsraum ist

$$\text{Lös}(A|b) = \underbrace{\begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}}_v + \mathbb{R} \cdot \underbrace{\begin{pmatrix} -3 \\ 2 \\ 1 \end{pmatrix}}_u, \quad (5.32)$$

also eine Gerade im \mathbb{R}^3 .

Am Ende bietet es sich an, die Lösung noch einmal zu testen: Die spezielle Lösung v erfüllt:

$$Av = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 5 \end{pmatrix} \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 2 \end{pmatrix} = b. \quad (5.33)$$

Der Vektor u soll im Kern von A liegen (in diesem Fall sogar eine Basis bilden) und erfüllt in der Tat

$$Au = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 5 \end{pmatrix} \begin{pmatrix} -3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \quad (5.34)$$

Die Dimension des Lösungsraums (falls er nicht leer ist) ist durch den Rang bestimmt, $\dim \text{Lös}(A, b) = n - \text{rang } A$. Also ist die Lösung eindeutig, wenn $\text{rang } A = n$.

Wenn $\text{rang } A = m$, ist $x \mapsto Ax$ surjektiv, und dann ist die Gleichung $Ax = b$ immer lösbar: $\text{rang}(A | b)$ kann nicht größer als die Zahl m der Zeilen werden.

Wir fassen diese Lösungskriterien zusammen:

Satz 5.3. Sei $Ax = b$ ein lineares Gleichungssystem, $A \in M(m \times n, \mathbb{R})$.

1. Ist $\text{rang } A = \text{rang}(A | b) = r$, dann existiert eine Lösung und der affine Lösungsraum ist $(n - r)$ -dimensional.
2. Für $\text{rang } A = n$ existiert höchstens eine Lösung.
3. Für $\text{rang } A = m$ existiert zu jedem $b \in \mathbb{R}^m$ eine Lösung.
4. Für $\text{rang } A = m = n$ gibt es zu jedem $b \in \mathbb{R}^n = \mathbb{R}^m$ eine eindeutige Lösung. Sie ist gegeben durch $x = A^{-1}b$.

5.3 Quotientenvektorräume

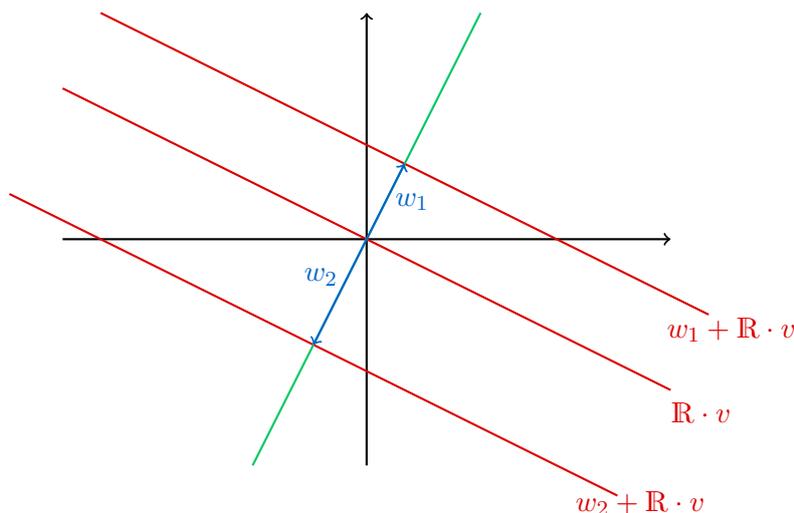
Wir haben gesehen, dass Lösungsräume inhomogener linearer Gleichungssysteme $Ax = b$ affine Unterräume sind. Schauen wir uns ein weiteres Beispiel an: Der Lösungsraum der Gleichung (wobei a_1 und a_2 nicht beide Null sein sollen)

$$\begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = b \quad (5.35)$$

ist eine Gerade, die im \mathbb{R}^2 orthogonal zu $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ verläuft (bezüglich des kanonischen Skalarprodukts auf \mathbb{R}^2), d.h. die Richtung der Geraden ist durch den Vektor $v = \begin{pmatrix} -a_2 \\ a_1 \end{pmatrix}$ gegeben. Der Abstand der Geraden vom Ursprung wird dabei durch die Zahl $b \in \mathbb{R}$ festgelegt, variiert man b , so erhält man verschiedene, zueinander parallel verschobene Geraden. Man kann nun auch die Menge aller dieser Lösungsräume betrachten, also die Menge in Richtung v zeigender Geraden im \mathbb{R}^2 :

$$\mathbb{R}^2 / (\mathbb{R} \cdot v) := \{w + \mathbb{R} \cdot v \mid w \in \mathbb{R}^2\}. \quad (5.36)$$

Die Notation der linken Seite wird in Kürze erklärt, im Moment sehen wir das einfach als Bezeichnung für die Menge der zu v parallelen Geraden in \mathbb{R}^2 an.



Dieser Menge kann man nun selbst wieder eine Vektorraumstruktur geben: Wir können Geraden addieren, indem wir die Menge aller Summen von Vektoren bilden,

$$(w_1 + \mathbb{R} \cdot v) + (w_2 + \mathbb{R} \cdot v) = (w_1 + w_2 + \mathbb{R} \cdot v), \quad (5.37)$$

und das Ergebnis ist wieder eine Gerade; wir können eine Gerade mit einem Skalar multiplizieren, indem wir jeden in der Gerade enthaltenen Vektor mit diesem Skalar multiplizieren,

$$\lambda \cdot (w + \mathbb{R} \cdot v) = (\lambda \cdot w + \mathbb{R} \cdot v), \quad (5.38)$$

und wir erhalten wieder eine Gerade, solange $\lambda \neq 0$. Für $\lambda = 0$ wäre das Ergebnis $\{0\}$, also die Menge, die nur den Ursprung enthält. Damit wir wieder eine Gerade erhalten, müssen wir die Multiplikation mit $\lambda = 0$ definieren durch

$$0 \cdot (w + \mathbb{R} \cdot v) := \mathbb{R} \cdot v.$$

Eigentlich wäre es hier sinnvoll, ein neues Symbol für die so definierte Skalarmultiplikation einzuführen, um sie von der Skalarmultiplikation auf \mathbb{R}^2 zu unterscheiden. Um die Notation einfach zu halten, werden wir aber bei dem gleichen Symbol bleiben. Man kann nun überprüfen, dass alle Vektorraumaxiome erfüllt sind.

Etwas anschaulicher kann man diese Vektorraumstruktur verstehen, wenn man jede Gerade durch den Vektor minimaler Länge charakterisiert, also den Vektor, der orthogonal zu v und damit in unserem Beispiel in Richtung $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ liegt. Durch jeden Vektor auf der (durch den Ursprung laufenden) Geraden $\mathbb{R} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ verläuft genau eine zu v parallele Gerade, d.h. wir haben eine Eins-zu-Eins-Beziehung zwischen Vektoren in $\mathbb{R} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ und Geraden in $\mathbb{R}^2 / (\mathbb{R} \cdot v)$. Der Raum $\mathbb{R} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ ist ein eindimensionaler Vektorraum (im obigen Bild als grüne Gerade eingezeichnet), und wir können den Vektorraum $\mathbb{R}^2 / (\mathbb{R} \cdot v)$ mit diesem eindimensionalen Vektorraum identifizieren (genauer gesagt definiert die oben erwähnte bijektive Abbildung einen Isomorphismus).

Die gerade eingeführte Konstruktion wollen wir jetzt verallgemeinern:

Satz 5.4. Sei V ein reeller Vektorraum, und sei $U \subset V$ ein Untervektorraum von V . Dann ist

$$V/U := \{v + U \mid v \in V\} \tag{5.39}$$

mit der Addition

$$\begin{aligned} + : V/U \times V/U &\longrightarrow V/U \\ (v_1 + U, v_2 + U) &\longmapsto v_1 + v_2 + U \end{aligned} \tag{5.40}$$

und der skalaren Multiplikation

$$\begin{aligned} \cdot : \mathbb{R} \times V/U &\longrightarrow V/U \\ (\lambda, v + U) &\longmapsto \lambda \cdot v + U \end{aligned} \tag{5.41}$$

ein Vektorraum. Ist V endlich-dimensional, so ist

$$\dim V/U = \dim V - \dim U. \tag{5.42}$$

Den Raum V/U nennt man **Quotientenvektorraum**.

Beweis. Zunächst muss man sich davon überzeugen, dass die oben eingeführten Abbildungen (Addition und Skalarmultiplikation) wohldefiniert sind. Es ist ja $v + U = (v + u) + U$ für jedes $u \in U$, d.h. man muss aufpassen, dass die Abbildungen das Gleiche

geben, wenn man den affinen Unterraum anders darstellt. Schauen wir uns das im Beispiel mit der Skalarmultiplikation an: Wir definieren $\lambda \cdot (v + U)$ als $\lambda \cdot v + U$. Hätten wir $v + U$ durch $(v + u) + U$ geschrieben, wäre das Resultat $\lambda \cdot (v + u + U) = \lambda \cdot (v + u) + U$. Da aber $\lambda \cdot u \in U$, gilt $\lambda \cdot (v + u) + U = \lambda \cdot v + U$ und wir erhalten das gleiche Ergebnis: Unsere Vorschrift ist also sinnvoll (wir sagen, sie ist *wohldefiniert*). Eine ähnliche Überlegung zeigt, dass die Addition wohldefiniert ist; sie stimmt auch mit der von V induzierten elementweisen Addition von Teilmengen von V überein:

$$(v_1 + U) + (v_2 + U) = \{w_1 + w_2 \mid w_1 \in v_1 + U, w_2 \in v_2 + U\} \quad (5.43)$$

$$= \{v_1 + u_1 + v_2 + u_2 \mid u_1, u_2 \in U\} \quad (5.44)$$

$$= \{v_1 + v_2 + u \mid u \in U\} \quad (5.45)$$

$$= v_1 + v_2 + U. \quad (5.46)$$

Man kann schnell überprüfen, dass diese Verknüpfung kommutativ und assoziativ ist. $0 + U = U$ ist das neutrale Element, und zu $v + U$ ist $-v + U$ das inverse Element.

Die definierte Skalarmultiplikation erfüllt die Verträglichkeitsbedingungen mit der Addition: Für $\lambda \in \mathbb{R}$ und $v_1, v_2 \in V$ ist

$$\lambda \cdot ((v_1 + U) + (v_2 + U)) = \lambda \cdot (v_1 + v_2 + U) = \lambda \cdot (v_1 + v_2) + U = \lambda \cdot v_1 + \lambda \cdot v_2 + U \quad (5.47)$$

und

$$(\lambda \cdot (v_1 + U)) + (\lambda \cdot (v_2 + U)) = (\lambda \cdot v_1 + U) + (\lambda \cdot v_2 + U) = \lambda \cdot v_1 + \lambda \cdot v_2 + U. \quad (5.48)$$

Die übrigen Verträglichkeitsbedingungen zeigt man analog. V/U ist also ein Vektorraum. Um etwas über die Dimension auszusagen, betrachten wir nun die lineare Abbildung

$$\begin{aligned} \pi : V &\longrightarrow V/U \\ v &\longmapsto v + U. \end{aligned} \quad (5.49)$$

Sie ist offensichtlich surjektiv, also $\text{Im } \pi = V/U$. Der Kern von π ,

$$\text{Ker } \pi = \{v \in V \mid \pi(v) = 0 + U = U\}, \quad (5.50)$$

ist die Menge aller Vektoren in V , die auf das Nullelement $0 + U = U$ von V/U abgebildet werden. $\pi(v) = v + U = U$ bedeutet insbesondere, dass jedes Element von $v + U$ und daher auch $v + 0$ in U enthalten ist, $\pi(v) = U \Rightarrow v \in U$. Andererseits ist für $v \in U$ $\pi(v) = v + U = U$, also ist

$$\text{Ker } \pi = U. \quad (5.51)$$

Nach der Dimensionsformel ist dann

$$\dim V = \dim \text{Ker } \pi + \dim \text{Im } \pi = \dim U + \dim V/U. \quad (5.52)$$

□

Bemerkung

Die Dimension von V/U können wir auch durch Betrachtung einer Basis bestimmen: Sei V ein Vektorraum mit Dimension $\dim V = d$, $U \subset V$ habe die Dimension $\dim U = r$. Sei (u_1, \dots, u_r) eine Basis von U , die wir zu einer Basis $(u_1, \dots, u_r, v_1, \dots, v_{d-r})$ von V ergänzen. Man kann sich dann davon überzeugen, dass $(v_1 + U, \dots, v_{d-r} + U)$ eine Basis von V/U bildet.

Zusammenfassung Kapitel 5

- Sie sind in der Lage, ein inhomogenes Gleichungssystem mit Hilfe des Gaußschen Eliminationsverfahrens zu lösen, d.h. Sie können entscheiden, ob eine Lösung existiert ($\text{rang}(A | b) = \text{rang } A$), und falls ja, können Sie eine Parametrisierung der allgemeinen Lösung angeben.
- Sie verstehen die Bedeutung des Satzes: Die allgemeine Lösung des inhomogenen Systems erhält man durch die Addition einer speziellen Lösung des inhomogenen Systems und der allgemeinen Lösung des zugehörigen homogenen Systems.
- Wenn das Gleichungssystem überhaupt lösbar ist, ergibt sich damit der Lösungsraum als ein verschobener Untervektorraum vom \mathbb{R}^n , also z.B. eine Gerade oder Ebene, die nicht durch den Ursprung geht (außer wenn $b = 0$). Solche Räume heißen affine Unterräume.
- Die Menge aller affinen Unterräume $v + U$ von V mit dem gleichen zugehörigen Untervektorraum U bildet selbst wieder einen Vektorraum, den Quotientenvektorraum V/U .

Eine abschließende Bemerkung zur Bedeutung der aus praktischen Gründen eingeführten erweiterten Matrix $(A | b)$:

Inhomogene Gleichungssysteme lassen sich auf homogene Systeme zurückführen, wenn wir schreiben

$$Ax = b \iff Ax - b = 0 \iff (A | b) \begin{pmatrix} x \\ -1 \end{pmatrix} = 0, \quad (5.53)$$

also ergibt sich der Lösungsraum des inhomogenen Systems als der Schnitt

$$\text{Ker}(A | b) \cap \{x' \in \mathbb{R}^{n+1} \mid x'_{n+1} = -1\} \quad (5.54)$$

des Untervektorraums $\text{Ker}(A | b)$ mit einer affinen Hyperebene durch Projektion auf die ersten n Koordinaten.

6 Determinanten

6.1 Determinante und Volumen

Matrizen vermitteln lineare Abbildungen im \mathbb{R}^n . Im \mathbb{R}^n haben wir aber auch einen Längenbegriff und darauf aufbauend Begriffe von Flächeninhalt und Volumen (und entsprechende n -dimensionale Verallgemeinerungen). Da ist die Frage naheliegend, wie sich Flächen und Volumina unter linearen Abbildungen verhalten.

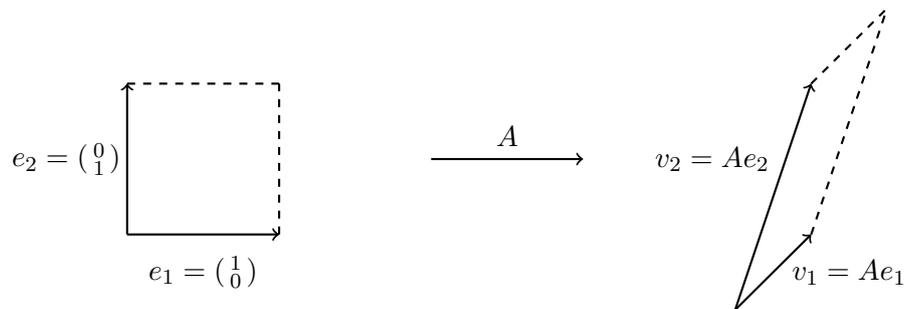
Schauen wir uns als Beispiel zwei Dimensionen an. Eine einfache geometrische Figur, die ihre Form unter linearen Abbildungen erhält, ist ein Parallelogramm,

$$P(v_1, v_2) = \{\lambda_1 v_1 + \lambda_2 v_2 \mid 0 \leq \lambda_1, \lambda_2 \leq 1\} \quad (6.1)$$

das von zwei Vektoren $v_1, v_2 \in \mathbb{R}^2$ aufgespannt wird.

Winkel und Längen mögen sich ändern unter einer linearen Abbildung, aber das Bild ist wieder ein (im schlimmsten Fall entartetes) Parallelogramm.

Als Beispiel betrachten wir das Bild des von e_1, e_2 aufgespannten Parallelogramms (Quadrats):



Der Flächeninhalt des von e_1, e_2 aufgespannten Parallelogramms setzen wir sinnvollerweise zu

$$\text{vol}_2(P(e_1, e_2)) = 1, \quad (6.2)$$

↑

2-dimensionales Volumen

der Flächeninhalt von $P(v_1, v_2) = P(Ae_1, Ae_2)$ ist durch die Eigenschaften von A bestimmt und durch die (noch zu definierende) Determinante von A gegeben. Nennen wir also

$$|\det A| = \text{vol}_2(P(Ae_1, Ae_2)). \quad (6.3)$$

Was erwarten wir, wenn wir mit einem beliebigen Parallelogramm starten (aufgespannt von Vektoren u_1, u_2)? Wieder wird sich der Flächeninhalt möglicherweise ändern. Wir

wollen kurz (ohne Anspruch auf Rigorosität) uns veranschaulichen, dass diese Änderung wieder durch Multiplikation mit $|\det A|$ gegeben ist.

Man stelle sich vor, wir nähern das neue Parallelogramm $P(u_1, u_2)$ durch eine große Zahl N von kleinen Quadraten mit Seitenlängen λ (λ stellen wir uns klein vor), d.h. der Flächeninhalt von $P(u_1, u_2)$ ist ungefähr $N \cdot \lambda^2$.

Jedes Quadrat wird auf ein Parallelogramm der Fläche $\lambda^2 \cdot |\det A|$ abgebildet (der Flächeninhalt verhält sich quadratisch unter Skalierung), d.h. das Parallelogramm $P(Au_1, Au_2)$ hat ungefähr die Fläche $N \cdot \lambda^2 \cdot |\det A|$. Der Flächeninhalt hat sich also um den Faktor $|\det A|$ geändert. Man könnte dieses Argument präziser machen durch entsprechende Limesbetrachtungen und kann es dann sogar sinnvoll auf beliebige „Figuren“ (in einem zu präzisierenden Sinn) angewendet werden.

Wir können daraus zwei Dinge folgern:

1. $|\det A|$ beschreibt ganz allgemein die multiplikative Änderung des Flächeninhalts unter der Abbildung A .
2. Betrachten wir eine zweite Abbildung gegeben durch eine Matrix B : Sie bildet $P(e_1, e_2)$ auf $P(Be_1, Be_2)$ ab mit Fläche $\text{vol}_2(P(Be_1, Be_2)) = |\det B|$. Bilden wir dieses Parallelogramm ab durch A auf $P(ABe_1, ABe_2)$, bekommen wir die Fläche

$$\text{vol}_2(P(ABe_1, ABe_2)) = |\det A| \cdot \text{vol}_2(P(Be_1, Be_2)) \quad (6.4)$$

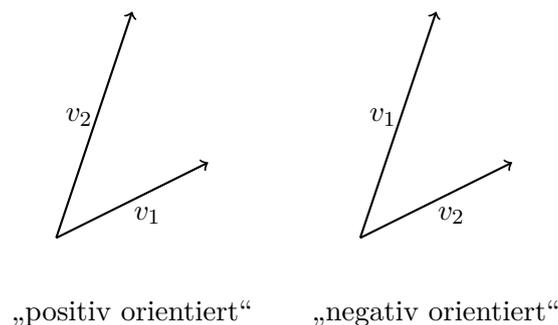
$$= |\det A| \cdot |\det B|. \quad (6.5)$$

Andererseits ist die Fläche gegeben durch $|\det(AB)|$. Wir erwarten also von diesem Faktor $|\det(\cdot)|$, dass er ein einfaches Verhalten unter Multiplikation von Matrizen hat:

$$|\det(AB)| = |\det A| \cdot |\det B|. \quad (6.6)$$

Das Argument können wir analog für ein Parallelepipid im \mathbb{R}^3 („Spat“) führen und gelangen zur gleichen Schlussfolgerung.

Um zur Determinante zu kommen, müssen wir kurz über Orientierung reden. Wenn wir eine Basis (v_1, v_2) von \mathbb{R}^2 haben, können wir zwei Fälle unterscheiden:



(v_1, v_2) ist positiv orientiert, wenn die kürzeste Rotation von der v_1 -Richtung zur v_2 -Richtung im mathematisch positiven Sinn (gegen den Uhrzeigersinn) verläuft.

Ebenso können wir im \mathbb{R}^3 eine Orientierung einführen: (v_1, v_2, v_3) ist positiv orientiert („Rechtssystem“), wenn die drei Vektoren wie Daumen, Zeigefinger und Mittelfinger der rechten Hand angeordnet sind (Daumen und Zeigefinger ausgestreckt).

Eine anschauliche Definition (eine präzise folgt später) der Determinante ist wie folgt:

$$\det A = \begin{cases} +\text{vol}_n(P(Ae_1, \dots, Ae_n)) & \text{wenn } (Ae_1, \dots, Ae_n) \text{ genauso orientiert} \\ & \text{ist wie } (e_1, \dots, e_n) \end{cases} \quad (6.7)$$

$$\begin{cases} -\text{vol}_n(P(Ae_1, \dots, Ae_n)) & \text{sonst} \end{cases}$$

Für diese Größe erwarten wir die Eigenschaft

$$\det(AB) = \det A \cdot \det B \quad (6.8)$$

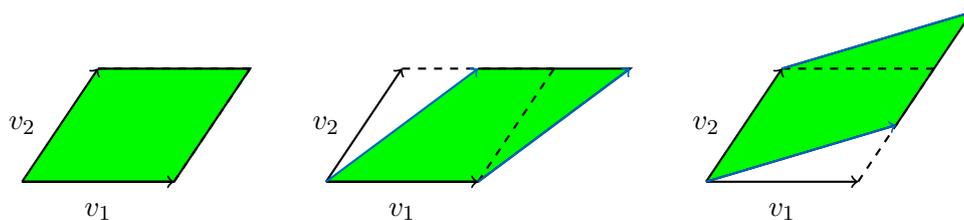
(für den Betrag hatten wir das schon diskutiert und man kann sich leicht überlegen, dass auch die Orientierung mit dieser Regel einverstanden ist).

In einer Dimension, im \mathbb{R}^1 , ist das sehr einfach. Die Länge eines Vektors (x) , $x \in \mathbb{R}$, ist $|x|$. Eine lineare Abbildung ist durch eine 1×1 -Matrix $A = (a)$ gegeben. Der Vektor $e_1 = (1)$ wird auf (a) abgebildet, also ist $|\det A| = a$. Ist $a > 0$, so zeigt Ae_1 in die gleiche Richtung wie e_1 (ist also gleich orientiert); ist hingegen $a < 0$, so zeigt Ae_1 in die entgegengesetzte Richtung. Also ist $\det A = a$.

Gehen wir zum \mathbb{R}^2 mit einer Matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Wir müssen den Flächeninhalt des Parallelogramms $P(v_1, v_2)$ mit Vektoren $v_1 = Ae_1 = \begin{pmatrix} a \\ c \end{pmatrix}$ und $v_2 = Ae_2 = \begin{pmatrix} b \\ d \end{pmatrix}$ bestimmen.

Der Flächeninhalt ändert sich nicht, wenn wir die Seite v_2 um ein Vielfaches von v_1 verschieben (und auch umgekehrt):

$$\text{vol}_2(P(v_1, v_2)) = \text{vol}_2(P(v_1, v_2 + \lambda \cdot v_1)) = \text{vol}_2(P(v_1 + \lambda \cdot v_2, v_2)) \quad (6.9)$$

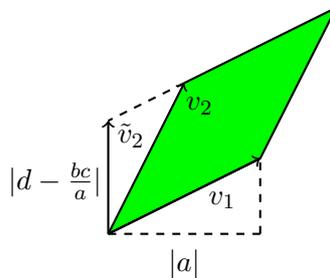


Ist $a \neq 0$ (d.h. $v_1 = \begin{pmatrix} a \\ c \end{pmatrix}$ liegt nicht parallel zu e_2), können wir v_2 so entlang v_1 verschieben, dass $v_2 + \lambda \cdot v_1$ parallel zu e_2 liegt:

$$\tilde{v}_2 = v_2 - \frac{b}{a} \cdot v_1 = \begin{pmatrix} b \\ d \end{pmatrix} - \frac{b}{a} \cdot \begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ d - \frac{bc}{a} \end{pmatrix}. \quad (6.10)$$

Der Flächeninhalt ist jetzt gegeben durch die Länge einer Grundseite $|d - \frac{bc}{a}|$ multipliziert mit der Höhe, also dem zu \tilde{v}_2 orthogonalen Teil $|a|$ von v_1 :

$$\text{vol}_2(P(v_1, v_2)) = \text{vol}_2(P(v_1, \tilde{v}_2)) = \left|d - \frac{bc}{a}\right| \cdot |a| = |ad - bc|. \quad (6.11)$$



Falls $a = 0$, können wir stattdessen $v_1 = \begin{pmatrix} 0 \\ c \end{pmatrix}$ als Grundseite (mit Länge $|c|$) nehmen, die Höhe ist dann der zu v_1 orthogonale Anteil $|b|$ von v_2 , sodass die Fläche $|bc|$ ist. Die Formel $|ad - bc|$ ist also auch in diesem Fall richtig.

Also ist $\left| \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \right| = |ad - bc|$.

Das Vorzeichen von $\det A$ können wir uns überlegen durch die Forderung $\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1$, d.h.

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc. \quad (6.12)$$

Für eine Matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ wird dies zu

$$\det A = a_{11}a_{22} - a_{12}a_{21} = \sum_{i_1, i_2} \epsilon_{i_1 i_2} a_{i_1 1} a_{i_2 2} \quad (6.13)$$

mit dem zweidimensionalen ϵ -Symbol

$$\epsilon_{12} = 1, \quad \epsilon_{21} = -1, \quad \epsilon_{11} = \epsilon_{22} = 0. \quad (6.14)$$

Das Symbol ϵ_{ij} ist antisymmetrisch bei Vertauschung der Indizes:

$$\epsilon_{ij} = -\epsilon_{ji}. \quad (6.15)$$

Gehen wir nun zu drei Dimensionen und betrachten die Matrix

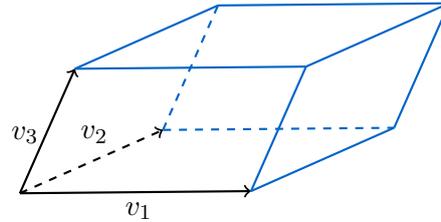
$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \quad (6.16)$$

Sie bildet die drei Einheitsvektoren e_1, e_2, e_3 auf die Vektoren

$$v_1 = Ae_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \quad v_2 = Ae_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}, \quad v_3 = Ae_3 = \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix} \quad (6.17)$$

ab.

Sie bilden ein Parallelepiped (einen Spat) $P(v_1, v_2, v_3)$:



Das Volumen ändert sich nicht, wenn ein Vektor, z.B. v_3 , entlang eines anderen Vektors verschoben wird,

$$\text{vol}_3(P(v_1, v_2, v_3)) = \text{vol}_3(P(v_1, v_2, v_3 + \lambda \cdot v_1)). \quad (6.18)$$

Anschaulich ergibt sich das daraus, dass das Volumen aus dem Flächeninhalt einer Seite (z.B. das von v_1 und v_2 gebildete Parallelogramm) multipliziert mit der Höhe (in dem Beispiel der zu v_1 und v_2 orthogonale Anteil von v_3) ergibt. (Die Höhe ändert sich nicht, wenn v_3 entlang von v_1 oder v_2 verschoben wird.)

Wir setzen der Einfachheit halber $a_{11} \neq 0$ voraus. Dann betrachten wir

$$P(v_1, \underbrace{v_2 - \frac{a_{12}}{a_{11}}v_1}_{\tilde{v}_2}, \underbrace{v_3 - \frac{a_{13}}{a_{11}}v_1}_{\tilde{v}_3}) = P\left(\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \begin{pmatrix} 0 \\ a_{22} - \frac{a_{12}}{a_{11}}a_{21} \\ a_{32} - \frac{a_{12}}{a_{11}}a_{31} \end{pmatrix}, \begin{pmatrix} 0 \\ a_{23} - \frac{a_{13}}{a_{11}}a_{21} \\ a_{33} - \frac{a_{13}}{a_{11}}a_{31} \end{pmatrix}\right). \quad (6.19)$$

Wir betrachten das von \tilde{v}_2, \tilde{v}_3 aufgespannte Parallelogramm in der x_2 - x_3 -Ebene mit Flächeninhalt

$$\begin{aligned} \text{vol}_2(P(\tilde{v}_2, \tilde{v}_3)) &= \left| \left(a_{22} - \frac{a_{12}}{a_{11}}a_{21} \right) \left(a_{33} - \frac{a_{13}}{a_{11}}a_{31} \right) - \left(a_{23} - \frac{a_{13}}{a_{11}}a_{21} \right) \left(a_{32} - \frac{a_{12}}{a_{11}}a_{31} \right) \right| \\ &= \left| a_{22}a_{33} - \frac{a_{12}}{a_{11}}a_{21}a_{33} - a_{22}\frac{a_{13}}{a_{11}}a_{31} + \frac{a_{12}}{a_{11}}a_{21}\frac{a_{13}}{a_{11}}a_{31} \right. \\ &\quad \left. - \left(a_{23}a_{32} - \frac{a_{13}}{a_{11}}a_{21}a_{32} - a_{23}\frac{a_{12}}{a_{11}}a_{31} + \frac{a_{13}}{a_{11}}a_{21}\frac{a_{12}}{a_{11}}a_{31} \right) \right|. \quad (6.20) \end{aligned}$$

Das Volumen ergibt sich aus diesem Flächeninhalt multipliziert mit der Höhe $|a_{11}|$:

$$\text{vol}_3(P(v_1, v_2, v_3)) = |a_{11}| \cdot \text{vol}_2(P(\tilde{v}_2, \tilde{v}_3)) \quad (6.21)$$

$$\begin{aligned} &= |a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ &\quad - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32}|. \quad (6.22) \end{aligned}$$

Dies liefert auch das richtige Resultat für $a_{11} = 0$.

Wir können dies wieder einfacher schreiben als

$$\text{vol}_3(P(v_1, v_2, v_3)) = \left| \sum_{i_1, i_2, i_3=1}^3 \epsilon_{i_1 i_2 i_3} a_{i_1 1} a_{i_2 2} a_{i_3 3} \right|. \quad (6.23)$$

Hierbei ist ϵ_{ijk} der total antisymmetrische ϵ -Tensor (Levi-Civita-Symbol), der definiert ist als

$$\epsilon_{123} = \epsilon_{312} = \epsilon_{231} = 1 \quad (\text{„gerade“ Permutation}) \quad (6.24a)$$

$$\epsilon_{213} = \epsilon_{321} = \epsilon_{132} = -1 \quad (\text{„ungerade“ Permutation}) \quad (6.24b)$$

$$\epsilon_{ijk} = 0 \quad (\text{falls zwei Indizes den gleichen Wert haben}). \quad (6.24c)$$

Das Levi-Civita-Symbol ist antisymmetrisch unter Vertauschung zweier Indizes:

$$\epsilon_{ijk} = -\epsilon_{jik} \quad (6.25a)$$

$$\epsilon_{ijk} = -\epsilon_{ikj} \quad (6.25b)$$

$$\epsilon_{ijk} = -\epsilon_{kji}. \quad (6.25c)$$

Diese Eigenschaft zusammen mit der „Normierung“ $\epsilon_{123} = 1$ bestimmt ϵ_{ijk} vollständig.

Die Einheitsmatrix E_3 überführt eine positive Orientierung in die gleiche positive Orientierung, also fordern wir $\det E_3 = +1$. Daher definieren wir die Determinante für eine 3×3 -Matrix als

$$\det A = \sum_{i_1, i_2, i_3=1}^3 \epsilon_{i_1 i_2 i_3} a_{i_1 1} a_{i_2 2} a_{i_3 3}. \quad (6.26)$$

Fassen wir unsere Überlegungen zusammen. Für die Determinante gilt:

- in einer Dimension: $\det(a_{11}) = a_{11} = \epsilon_1 a_{11}$ für $\epsilon_1 = 1$
- in zwei Dimensionen: $\det(a_{ij}) = \sum_{i_1, i_2=1}^2 \epsilon_{i_1 i_2} a_{i_1 1} a_{i_2 2}$
- in drei Dimensionen: $\det(a_{ij}) = \sum_{i_1, i_2, i_3=1}^3 \epsilon_{i_1 i_2 i_3} a_{i_1 1} a_{i_2 2} a_{i_3 3}$.

Dies legt folgende allgemeine Definition nahe:

Definition 6.1. Für eine reelle (oder komplexe) $n \times n$ -Matrix $A = (a_{ij})$ definieren wir die Determinante von A als

$$\det A = \sum_{i_1, \dots, i_n=1}^n \epsilon_{i_1 \dots i_n} a_{i_1 1} a_{i_2 2} \dots a_{i_n n}. \quad (6.27)$$

Dabei ist das ϵ -Symbol definiert als

$$\epsilon_{i_1 \dots i_n} = \begin{cases} 1 & \text{wenn } i_1, \dots, i_n \text{ durch eine gerade Permutation aus } 1, \dots, n \text{ hervorgeht} \\ -1 & \text{wenn } i_1, \dots, i_n \text{ durch eine ungerade Permutation aus } 1, \dots, n \text{ hervorgeht} \\ 0 & \text{wenn zwei (oder mehr) Indizes gleich sind.} \end{cases} \quad (6.28)$$

Dabei nennen wir eine bijektive Abbildung

$$\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\} \quad (6.29)$$

eine **Permutation** der Zahlen $1, \dots, n$. i_1, \dots, i_n geht also durch eine Permutation aus $1, \dots, n$ hervor, wenn es eine solche Abbildung π gibt mit $\pi(1) = i_1, \dots, \pi(n) = i_n$.

Die einfachsten Permutationen sind die, bei der einfach zwei Zahlen vertauscht werden (eine **Transposition**). Jede Permutation lässt sich durch Hintereinanderausführung von Transpositionen erreichen. Wir nennen eine Permutation **gerade**, wenn dazu eine gerade Zahl von Vertauschungen benötigt wird, und **ungerade**, wenn die Permutation aus einer ungeraden Anzahl von Vertauschungen hervorgeht.

Für eine Permutation π (im obigen Sinn) definieren wir dann das sogenannte Signum von π als

$$\text{sign } \pi = \begin{cases} +1 & \text{wenn } \pi \text{ eine gerade Permutation ist} \\ -1 & \text{wenn } \pi \text{ eine ungerade Permutation ist.} \end{cases} \quad (6.30)$$

Das Signum ist wohldefiniert: Es hängt nicht davon ab, wie man eine Permutation aus Vertauschungen zusammengesetzt hat.

Das sieht man über eine explizite Formel:

$$\text{sign } \pi = \prod_{1 \leq i < j \leq n} \frac{\pi(j) - \pi(i)}{j - i} \quad (6.31)$$

Da im Zähler und Nenner insgesamt die gleichen Paare vorkommen, gilt $|\text{sign } \pi| = 1$. Man kann sich dann überlegen, dass die so definierte Funktion bei einer Transposition das Vorzeichen ändert (siehe Übungen).

Es gilt für jede Permutation

$$\epsilon_{\pi(1) \dots \pi(n)} = \text{sign } \pi. \quad (6.32)$$

Der ϵ -Tensor kann dann auch dadurch charakterisiert werden, dass man fordert

$$\epsilon_{1 \dots n} = 1 \quad (6.33)$$

$$\epsilon_{\dots i \dots j \dots} = -\epsilon_{\dots j \dots i \dots} \quad (\text{antisymmetrisch unter Vertauschung beliebiger Indizes}). \quad (6.34)$$

Wir können die Determinante jetzt auch so schreiben:

$$\det A = \sum_{i_1, \dots, i_n=1}^n \epsilon_{i_1 \dots i_n} a_{i_1 1} a_{i_2 2} \dots a_{i_n n} = \quad (6.35)$$

$$= \sum_{\pi \in S_n} \text{sign}(\pi) a_{\pi(1)1} a_{\pi(2)2} \dots a_{\pi(n)n} \quad (6.36)$$

mit S_n als der Menge aller Permutationen $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. Wir bemerken kurz, dass die Permutationen mit der Hintereinanderausführung eine Gruppe, die symmetrische Gruppe von $\{1, \dots, n\}$, bildet (siehe Beispiel 1.15).

Bemerkung

Oft sieht man diese Formel mit vertauschten Indizes:

$$\det A = \sum_{\pi \in S_n} \text{sign}(\pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)}. \quad (6.37)$$

Diese Formeln sind äquivalent, denn durch Umordnen der Koeffizienten a_{ij} kann man den letzten Ausdruck schreiben als

$$\begin{aligned} & \sum_{\pi \in S_n} \text{sign}(\pi) a_{\pi^{-1}(1)1} a_{\pi^{-1}(2)2} \cdots a_{\pi^{-1}(n)n} \\ &= \sum_{\pi' \in S_n} \text{sign}((\pi')^{-1}) a_{\pi'(1)1} a_{\pi'(2)2} \cdots a_{\pi'(n)n} \end{aligned} \quad (6.38)$$

$$= \sum_{\pi' \in S_n} \text{sign}(\pi') a_{\pi'(1)1} a_{\pi'(2)2} \cdots a_{\pi'(n)n}. \quad (6.39)$$

Der letzte Schritt gilt, da $\text{sign} \pi = \text{sign} \pi^{-1}$, denn wenn π aus r Transpositionen aufgebaut ist, kann man π auch durch r Transpositionen rückgängig machen, d.h. π^{-1} kann aus r Transpositionen gebildet werden.

Mit dieser Erkenntnis folgt dann auch

$$\det A = \det A^t, \quad (6.40)$$

wofür wir später noch einen anderen Beweis sehen werden.

6.2 Eigenschaften

Jetzt, wo wir die Determinante — motiviert aus unseren geometrischen Überlegungen — definiert haben, wollen wir ihre Eigenschaften herausarbeiten. Wir werden dann sehen, wie sie darüber berechnet werden kann und dass die Determinante durch einige ihrer Eigenschaften schon eindeutig bestimmt ist.

Wir schreiben im Folgenden die Matrix mithilfe ihrer Spaltenvektoren, $A = (v_1 \cdots v_n)$, $v_i \in \mathbb{R}^n$.

Satz 6.1. *Die Determinante hat die folgenden Eigenschaften:*

Für alle $A = (v_1 \cdots v_n) \in M(n \times n, \mathbb{R})$, $\lambda \in \mathbb{R}$, gilt

1. $\det(v_1 \cdots v_i \cdots v_j \cdots v_n) = -\det(v_1 \cdots v_j \cdots v_i \cdots v_n)$, ($i, j \in \{1, \dots, n\}, i < j$), d.h. die Determinante ändert das Vorzeichen bei Vertauschung zweier Spalten.
2. $\det(v_1 \cdots v_i + \lambda \cdot v_j \cdots v_n) = \det(v_1 \cdots v_n)$, ($i, j \in \{1, \dots, n\}, i \neq j$), d.h. die Determinante ändert sich nicht, wenn ein Vielfaches einer Spalte zu einer anderen Spalte addiert wird.
3. $\det(v_1 \cdots \lambda \cdot v_i \cdots v_n) = \lambda \cdot \det(v_1 \cdots v_n)$, ($i \in \{1, \dots, n\}$), d.h. die Determinante ist homogen unter Skalierung jeder Spalte.
4. $\det(e_1 \cdots e_n) = 1$.
5. Für alle $w \in \mathbb{R}^n$ gilt

$$\det(v_1 \cdots v_i + w \cdots v_n) = \det(v_1 \cdots v_i \cdots v_n) + \det(v_1 \cdots w \cdots v_n) ,$$

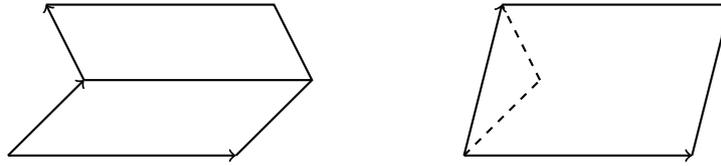
($i \in \{1, \dots, n\}$) mit w an der i -ten Position, d.h. die Determinante ist additiv in jeder Spalte.

Bemerkung

Bevor wir diese Eigenschaften beweisen, wollen wir kurz ihre geometrische Bedeutung beleuchten:

1. Umm Nummerierung der Kanten (Vertauschung zweier Indizes) führt zu Wechsel der Orientierung.
2. Invarianz des Volumens eines Parallelotops unter Verschiebung eines Seitenvektors parallel zu einer anderen Kante.
3. Skalierung des Volumens unter Skalierung einer Kante ($\lambda < 0$ bedeutet Spiegelung des Kantenvektors und damit Wechsel der Orientierung).
4. Das (orientierte) Volumen des Einheitskubus setzen wir zu 1.
5. Das Gesamtvolumen zweier Parallelotope mit gleicher „Grundfläche“ ist gegeben

durch das Volumen eines Parallelotops mit dieser Grundfläche und einer Höhe, die gleich der Summe der einzelnen Höhen ist.



Nun zum Beweis.

Beweis.

1. Offensichtlich aus der Antisymmetrie von $\epsilon_{i_1 \dots i_n}$. Es folgt aber auch aus 2. und 3., die wir gleich beweisen werden, denn die Vertauschung zweier Spalten kann man zusammensetzen aus Operationen von Typ 2. und 3.:

$$v_i, v_j \xrightarrow{2.} v_i, v_j + v_i \xrightarrow{2.} \underbrace{v_i - (v_j + v_i)}_{-v_j}, v_j + v_i \quad (6.41)$$

$$\xrightarrow{2.} -v_j, \underbrace{(v_j + v_i) - v_j}_{v_i} \xrightarrow{3. \text{ mit } \lambda=-1} v_j, v_i \quad (6.42)$$

2. Sei der Einfachheit halber $i = 1, j = 2$.

$$\det(v_1 + \lambda \cdot v_2, v_2, \dots, v_n) = \sum_{i_1, \dots, i_n=1}^n \epsilon_{i_1 \dots i_n} (a_{i_1 1} + \lambda \cdot a_{i_1 2}) a_{i_2 2} \dots a_{i_n n} \quad (6.43)$$

$$= \det(v_1, \dots, v_n) + \lambda \sum_{i_1, \dots, i_n=1}^n \underbrace{\epsilon_{i_1 \dots i_n}}_{\substack{\text{antisymm.} \\ \text{in } i_1, i_2}} \underbrace{a_{i_1 2} a_{i_2 2} a_{i_3 3} \dots a_{i_n n}}_{\substack{\text{symm.} \\ \text{in } i_1, i_2}} \quad (6.44)$$

$$= \det(v_1, \dots, v_n) \quad (6.45)$$

Zur Verdeutlichung des letzten Schritts ein einfaches Beispiel: Sei $s_{ij} = s_{ji}$ ein symmetrisches Objekt. Dann ist

$$s = \sum_{ij} \epsilon_{ij} s_{ij} = - \sum_{ij} \epsilon_{ji} s_{ij} = - \sum_{i'j'} \epsilon_{i'j'} s_{j'i'} = - \sum_{i'j'} \epsilon_{i'j'} s_{i'j'} = -s, \quad (6.46)$$

also $s = 0$.

3. Offensichtlich aus der Definition von \det .

4. $\det(e_1, \dots, e_n) = \sum_{i_1, \dots, i_n=1}^n \epsilon_{i_1 \dots i_n} \delta_{i_1 1} \dots \delta_{i_n n} = \epsilon_{1 \dots n} = 1$.

5. Offensichtlich aus der Definition von \det .

□

Bemerkung

Die Eigenschaften 3. und 5. besagen, dass die Determinante als Funktion der n Spaltenvektoren linear in jedem Eintrag ist, also können wir die Determinante als **Multilinearform** auf \mathbb{R}^n ansehen:

$$\det : \mathbb{R}^n \times \dots \times \mathbb{R}^n \longrightarrow \mathbb{R} \quad (6.47)$$

$$(v_1, \dots, v_n) \longmapsto \det(v_1 \cdots v_n). \quad (6.48)$$

Wir erinnern uns, dass wir eine lineare Abbildung $\mathbb{R}^n \rightarrow \mathbb{R}$ eine Linearform nennen, das Skalarprodukt $\langle \cdot | \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ ist eine Bilinearform.

Eine Multilinearform mit der Eigenschaft 1. (Vorzeichenwechsel bei Vertauschung zweier Einträge) heißt **alternierend**. Die Determinante definiert also eine **alternierende Multilinearform** auf \mathbb{R}^n . (Alternierend bedeutet zunächst, dass die Multilinearform verschwindet, wenn zwei Einträge gleich sind — über \mathbb{R} ist dies aber äquivalent zur Antisymmetrie unter Vertauschung.)

Die Eigenschaften 1., 2. und 3. beschreiben das Verhalten der Determinante unter Spaltentransformationen. Das können wir zum einen später für ein Verfahren zur Berechnung verwenden. Zum anderen können wir daraus noch ein paar wichtige Eigenschaften ableiten.

Zunächst folgt aus 1., 2. und 3., dass für eine Elementarmatrix S (die durch Multiplikation von rechts die Vertauschung zweier Spalten / die Addition eines Vielfachen einer Spalte zu einer anderen / die Skalierung einer Spalte bewirkt) gilt:

Für alle $A \in M(n \times n, \mathbb{R})$ ist

$$\det(AS) = \det A \cdot \det S. \quad (6.49)$$

Beweis. Sei S eine Elementarmatrix, die eine der Transformationen 1., 2. oder 3. bewirkt. Dann folgt aus den Eigenschaften der Determinante für alle $A \in M(n \times n, \mathbb{R})$

$$\det(AS) = c \cdot \det A, \quad (6.50)$$

wobei c unabhängig ist von A ($c = -1$ für Vertauschung, $c = 1$ für Spaltenaddition, $c = \lambda$ für Spaltenmultiplikation mit λ).

Also gilt insbesondere für $A = E_n$

$$\det S = \det(E_n S) = c \cdot \det E_n = c \quad (6.51)$$

und somit

$$\det(AS) = \det A \cdot \det S. \quad (6.52)$$

□

Die Determinanten der Elementarmatrizen sind dann

$$\begin{aligned} \bullet \det P_a^b &= \det \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 0 & & 1 & \\ & & & \ddots & & \\ & 1 & & & 0 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} = -1 \\ \bullet \det Q_a^b(\lambda) &= \det \begin{pmatrix} 1 & & \lambda \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \det \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ \lambda & & 1 \end{pmatrix} = 1 \\ \bullet \det Q_a(\lambda) &= \det \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} = \lambda. \end{aligned}$$

Wir wissen, dass wir A durch elementare Spaltentransformationen S_1, \dots, S_r in Spaltenstufenform bringen können, $\tilde{A} = AS_1 \dots S_r$, also

$$\det \tilde{A} = \det(AS_1 \dots S_r) \tag{6.53}$$

$$= \det(AS_1 \dots S_{r-1}) \cdot \det S_r \tag{6.54}$$

$$= \det(AS_1 \dots S_{r-2}) \cdot \det S_{r-1} \cdot \det S_r \tag{6.55}$$

$$= \dots \tag{6.56}$$

$$= \det A \cdot \det S_1 \cdot \dots \cdot \det S_r. \tag{6.57}$$

Wenn A nicht maximalen Rang hat, so hat \tilde{A} einen Spaltenvektor der Null ist, also $\det \tilde{A} = 0$. (Denn $\lambda \cdot \det(\dots 0) = \det(\dots \lambda \cdot 0) = \det(\dots 0)$ für alle $\lambda \in \mathbb{R}$.) Da $\det S_i \neq 0$ für jede Elementarmatrix, ist dann auch $\det A = 0$.

Wenn A maximalen Rang hat, können wir A sogar durch Spaltentransformationen in die Einheitsmatrix verwandeln:

$$E_n = AS_1 \dots S_r, \tag{6.58}$$

also $1 = \det A \cdot \det S_1 \cdot \dots \cdot \det S_r$ und somit $\det A \neq 0$.

Daher gilt:

Satz 6.2. Sei $A \in M(n \times n, \mathbb{R})$. Dann gilt

$$\text{rang } A = n \iff \det A \neq 0. \tag{6.59}$$

Wir können auch sofort den Determinantenmultiplikationssatz, den wir schon zu Beginn diskutiert hatten, beweisen:

Satz 6.3. Für alle $A, B \in M(n \times n, \mathbb{R})$ gilt

$$\det(AB) = \det A \cdot \det B. \quad (6.60)$$

Beweis. Wenn $\text{rang } B < n$, dann ist auch $\text{rang}(AB) < n$ und beide Seiten sind Null.

Wenn $\text{rang } B = n$, so gibt es Elementarmatrizen S_1, \dots, S_r mit $B = S_1 \dots S_r$, sodass

$$\det(AB) = \det(AS_1 \dots S_r) = \det A \cdot \det S_1 \cdot \dots \cdot \det S_r. \quad (6.61)$$

Gleichzeitig gilt

$$\det B = \det(S_1 \dots S_r) = \det S_1 \cdot \dots \cdot \det S_r \quad (6.62)$$

und daher gilt

$$\det(AB) = \det A \cdot \det B. \quad (6.63)$$

□

Wir haben den Determinantenmultiplikationssatz rein über die Eigenschaften der Determinante bewiesen ohne die explizite Formel zu verwenden.

Wir haben dabei die Eigenschaft 5 (Additivität in den Spalten) nie benutzt. Da Eigenschaft 1 (Vorzeichenwechsel bei Vertauschung von Spalten) aus den anderen folgt, ist die Determinante durch die Eigenschaften 2, 3 und 4 eindeutig festgelegt (2, 3, und 4 legen zunächst die Determinante der Einheitsmatrix und der Elementarmatrizen fest, wegen des Determinantenmultiplikationssatzes ist dann die Determinante eindeutig bestimmt).

Satz 6.4.

1. Sei $f : \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}$ eine Abbildung mit den Eigenschaften ($\forall v_1, \dots, v_n \in \mathbb{R}^n, \lambda \in \mathbb{R}, i, j \in \{1, \dots, n\}, i \neq j$)

- $f(v_1, \dots, v_i + \lambda \cdot v_j, \dots, v_n) = f(v_1, \dots, v_n)$
- $f(v_1, \dots, \lambda \cdot v_i, \dots, v_n) = \lambda \cdot f(v_1, \dots, v_n)$
- $f(e_1, \dots, e_n) = 1$.

Dann gilt für alle $v_1, \dots, v_n \in \mathbb{R}^n$: $f(v_1, \dots, v_n) = \det(v_1 \dots v_n)$.

2. Sei $F : \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}$ eine Abbildung mit den Eigenschaften ($\forall v_1, \dots, v_n \in \mathbb{R}^n, \lambda \in \mathbb{R}, i, j \in \{1, \dots, n\}, i \neq j$)

- $F(v_1, \dots, v_i + \lambda \cdot v_j, \dots, v_n) = F(v_1, \dots, v_n)$
- $F(v_1, \dots, \lambda \cdot v_i, \dots, v_n) = |\lambda| \cdot F(v_1, \dots, v_n)$
- $F(e_1, \dots, e_n) = 1$.

Dann gilt für alle $v_1, \dots, v_n \in \mathbb{R}^n$: $F(v_1, \dots, v_n) = |\det(v_1 \cdots v_n)|$.

3. Sei $f : \mathbb{R}^n \times \dots \times \mathbb{R}^n \rightarrow \mathbb{R}$ eine alternierende Multilinearform mit $f(e_1, \dots, e_n) = 1$.
 1. Dann ist $\forall v_1, \dots, v_n \in \mathbb{R}^n$: $f(v_1, \dots, v_n) = \det(v_1 \cdots v_n)$.

Teil 1 haben wir oben schon begründet, dieser Satz ermöglicht eine Charakterisierung der Determinante über ihre Eigenschaften. Teil 2 folgt analog und charakterisiert das Volumen über die Eigenschaften. Eine weitere Charakterisierung der Determinante erlaubt Teil 3 (über diese wird die Determinante in der Weierstraßschen Axiomatik definiert). Dieser Teil des Satzes folgt daraus, das aus der Linearität (3 und 5) und der Antisymmetrie unter Vertauschung (1) auch sofort folgt, dass jede alternierende Multilinearform invariant ist, wenn das Vielfache eines (Vektor-)Eintrags zu einem anderen Vektor addiert wird:

$$f(v_1, \dots, v_i + \lambda \cdot v_j, \dots, v_j, \dots, v_n) = f(\dots) + \lambda \cdot f(\dots, v_j, \dots, v_j, \dots) \quad (6.64)$$

Das heißt, dass jede (normierte) alternierende Multilinearform die Voraussetzungen von Teil 1 erfüllt und damit mit der Determinantenabbildung übereinstimmt.

Wie wir schon erwähnt hatten, gilt $\det A = \det A^t$. Das können wir nun auch über die Eigenschaften beweisen.

Satz 6.5. Für jede reelle $n \times n$ -Matrix A gilt $\det A = \det A^t$.

Beweis. Wie man überprüfen kann, ist diese Eigenschaft für eine Elementarmatrix S_i erfüllt, $\det S_i = \det S_i^t$. Ist $\text{rang } A < n$, so ist auch $\text{rang } A^t < n$ und beide Seiten sind Null. Ist $\text{rang } A = n$, so gibt es Elementarmatrizen S_1, \dots, S_r mit $A = S_1 \dots S_r$, also

$$\det A^t = \det(S_1 \dots S_r)^t \quad (6.65)$$

$$= \det(S_r^t \dots S_1^t) \quad (6.66)$$

$$= \det S_r^t \cdot \dots \cdot \det S_1^t \quad (6.67)$$

$$= \det S_r \cdot \dots \cdot \det S_1 \quad (6.68)$$

$$= \det(S_1 \dots S_r) = \det A. \quad (6.69)$$

□

Aus dem Determinantenmultiplikationssatz folgt sofort die Beziehung zwischen der Determinante einer Matrix und der Determinante der inversen Matrix.

Satz 6.6. Sei $A \in GL(n, \mathbb{R})$ (Menge der invertierbaren $n \times n$ -Matrizen). Dann ist

$$\det(A^{-1}) = (\det A)^{-1}. \quad (6.70)$$

Beweis. Da $AA^{-1} = E_n$, folgt

$$1 = \det E_n = \det(AA^{-1}) = \det A \cdot \det(A^{-1}). \quad (6.71)$$

□

Wir hatten in Kapitel 4 die Matrixdarstellung von linearen Abbildungen mit Hilfe von Basisisomorphismen besprochen. Für einen Endomorphismus $F : V \rightarrow V$, also eine Abbildung von einem Vektorraum V in sich selbst, ergibt dies bezüglich einer Basis \mathcal{A} eine quadratische $n \times n$ -Matrix $M_{\mathcal{A}}^{\mathcal{A}}(F)$ wenn $\dim V = n$.

Damit haben wir die Möglichkeit, einem Endomorphismus F auf einem endlichdimensionalen Vektorraum eine Determinante zuzuordnen:

$$\det F := \det(M_{\mathcal{A}}^{\mathcal{A}}(F)). \quad (6.72)$$

Das ist aber natürlich nur dann sinnvoll, wenn diese Definition nicht von der Basis abhängt. Sei $A = M_{\mathcal{A}}^{\mathcal{A}}(F)$, und sei $B = M_{\mathcal{B}}^{\mathcal{B}}(F)$ die Matrixdarstellung bezüglich einer zweiten Basis \mathcal{B} . Wie wir gesehen hatten, gilt dann $B = SAS^{-1}$, wobei S den Basiswechsel sowohl im Definitionsraum V als auch im Bildraum V beschreibt.

Dann gilt

$$\det B = \det(SAS^{-1}) \quad (6.73)$$

$$= \det S \cdot \det A \cdot \det S^{-1} = \det A, \quad (6.74)$$

d.h. die Determinante $\det(M_{\mathcal{A}}^{\mathcal{A}}(F))$ ist unabhängig von der Basis.

Damit haben wir die wichtigsten Eigenschaften der Determinante besprochen. Wir wenden uns nun ihrer Berechnung und Anwendung zu.

6.3 Berechnungsverfahren und Anwendung

Die Determinante kann auf verschiedene Weisen berechnet werden. Für die Determinante einer 2×2 -Matrix bietet sich die direkte Methode an:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc. \quad (6.75)$$

Auch für eine 3×3 -Matrix kann man noch den direkten Ausdruck verwenden. Dieser lässt sich in der Regel von Sarrus zusammenfassen:

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = (\text{Summe der Produkte der grünen Linien}) \quad (6.76)$$

$$- (\text{Summe der Produkte der roten Linien}) \quad (6.77)$$

$$= (a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}) - \quad (6.78)$$

$$- (a_{12}a_{21}a_{33} + a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32}) \quad (6.79)$$

Bei größeren Matrizen wird diese Methode immer aufwendiger. Bei 4×4 -Matrizen muss man 24 Terme aufaddieren, bei 5×5 -Matrizen schon 120 Terme. Daher sucht man nach praktischeren Verfahren.

Ein sehr effizientes Verfahren beruht auf den Zeilen- und Spaltentransformationen. Da $\det A = \det A^t$, gilt natürlich alles, was wir über das Verhalten der Determinante bei Spaltentransformationen gesagt haben, auch für Zeilentransformationen. Da uns diese von der Behandlung von Gleichungssystemen vertrauter sind, werden wir das Verfahren mit Zeilentransformationen formulieren.

Die Strategie ist simpel: Wir versuchen die Matrix durch Zeilentransformationen auf die Einheitsmatrix zu bringen und merken uns die dabei auftretenden Faktoren. Wenn wir in dem Verfahren merken, dass die Matrix nicht maximalen Rang hat, ist die Determinante Null.

Beispiel 6.1.

$$\det A = \det \begin{pmatrix} 1 & -1 & 5 \\ 3 & 1 & 9 \\ 2 & -4 & 2 \end{pmatrix} \quad (6.80)$$

$$= \det \begin{pmatrix} 1 & -1 & 5 \\ 0 & 4 & -6 \\ 0 & -2 & -8 \end{pmatrix} \begin{array}{l} II - 3 \cdot I \\ III - 2 \cdot I \end{array} \quad (6.81)$$

$$= 2 \cdot \det \begin{pmatrix} 1 & -1 & 5 \\ 0 & 2 & -3 \\ 0 & -2 & -8 \end{pmatrix} \frac{1}{2} \cdot II \quad (6.82)$$

$$= 2 \cdot \det \begin{pmatrix} 1 & -1 & 5 \\ 0 & 2 & -3 \\ 0 & 0 & -11 \end{pmatrix} III + II \quad (6.83)$$

Jetzt können wir weitermachen, aber von nun an werden wir nur noch Transformationen vom Typ 2 (Addition eines Vielfachen einer Zeile zu einer anderen) benötigen, um die Matrix in eine Diagonalmatrix zu überführen, die aber die Determinante nicht ändern. Also

$$\det A = 2 \cdot \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -11 \end{pmatrix} \quad (6.84)$$

$$= 2 \cdot 2 \cdot (-11) \cdot \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = -44 \quad (6.85)$$

Wir sehen somit, dass es genügt, die Matrix in Dreiecksform zu bringen, dann ist die verbleibende Determinante durch das Produkt der Diagonalelemente gegeben.

Das Verfahren ist also wie folgt:

- Wir überführen die Matrix in Zeilenstufenform. Dabei berücksichtigen wir, dass das Vertauschen von Zeilen das Vorzeichen der Determinante ändert und dass das Multiplizieren einer Zeile mit einer Zahl λ die Determinante um λ^{-1} ändert.
- Wenn der Rang nicht maximal ist, ist die Determinante Null. Ansonsten hat die Matrix nun obere Dreiecksform, deren Determinante durch das Produkt der Einträge auf der Diagonale gegeben ist.

Dass die Determinante einer oberen Dreiecksmatrix durch das Produkt der Diagonalelemente gegeben ist, folgt auch aus dem folgenden Kästchensatz:

Satz 6.7. Sei $A \in M(m \times m, \mathbb{R})$ eine Matrix von der Form

$$A = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix} \quad (6.86)$$

mit quadratischen Matrizen $A_1 \in M(r \times r, \mathbb{R})$, $A_2 \in M((m-r) \times (m-r), \mathbb{R})$, ($1 \leq r \leq m-1$) und $B \in M(r \times (m-r), \mathbb{R})$.

Dann ist

$$\det A = \det A_1 \cdot \det A_2. \quad (6.87)$$

Beweis. Die Determinante von A verschwindet, wenn A_1 und A_2 nicht maximalen Rang haben, da dann auch A nicht maximalen Rang hat. Wir setzen voraus, dass A_1 und A_2 maximalen Rang haben. Sei S eine $r \times r$ -Matrix, die eine elementare Zeilentransformation $A_1 \rightsquigarrow \tilde{A}_1 = SA_1$ bewirkt mit $\det \tilde{A}_1 = \det S \cdot \det A_1$. Die entsprechenden Zeilentransformation in den ersten r Zeilen von A wird bewirkt durch

$$\begin{pmatrix} S & 0 \\ 0 & E_{m-r} \end{pmatrix}. \quad (6.88)$$

Also gilt

$$\det \left(\begin{pmatrix} S & 0 \\ 0 & E_{m-r} \end{pmatrix} \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix} \right) = \det \begin{pmatrix} SA_1 & SB \\ 0 & A_2 \end{pmatrix} = \det S \cdot \det \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}. \quad (6.89)$$

Sei T eine $(m-r) \times (m-r)$ -Matrix, die eine Zeilentransformation von A_2 bewirkt. Dann ist

$$\det \left(\begin{pmatrix} E_r & 0 \\ 0 & T \end{pmatrix} \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix} \right) = \det \begin{pmatrix} A_1 & B \\ 0 & TA_2 \end{pmatrix} = \det T \cdot \det \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix}. \quad (6.90)$$

Wenn $A_1 = S_1 \dots S_{r_1}$, $A_2 = T_1 \dots T_{r_2}$ mit Elementarmatrizen S_i , T_i , dann folgt

$$\det \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix} = \det \begin{pmatrix} S_1 \dots S_{r_1} & B \\ 0 & T_1 \dots T_{r_2} \end{pmatrix} \quad (6.91)$$

$$= \det(S_1 \dots S_{r_1}) \cdot \det(T_1 \dots T_{r_2}) \cdot \det \begin{pmatrix} E_r & A_1^{-1}B \\ 0 & E_{m-r} \end{pmatrix} \quad (6.92)$$

$$= \det A_1 \cdot \det A_2 \cdot \det \begin{pmatrix} E_r & 0 \\ 0 & E_{m-r} \end{pmatrix} \quad (6.93)$$

$$= \det A_1 \cdot \det A_2, \quad (6.94)$$

wobei im Schritt von der zweiten zur dritten Zeile die Matrix durch Typ 2 Umformungen verändert wurde. \square

also

$$\det A = (-1)^{i+j} \cdot \begin{pmatrix} 1 & a_{i1} & \cdots & a_{ij-1} & a_{ij+1} & \cdots & a_{in} \\ 0 & a_{11} & \cdots & a_{1j-1} & a_{1j+1} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_{i-11} & \cdots & a_{i-1j-1} & a_{i-1j+1} & \cdots & a_{i-1n} \\ 0 & a_{i+11} & \cdots & a_{i+1j-1} & a_{i+1j+1} & \cdots & a_{i+1n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & a_{n1} & \cdots & a_{nj-1} & a_{nj+1} & \cdots & a_{nn} \end{pmatrix} \quad (6.99)$$

$$= (-1)^{i+j} \cdot \det 1 \cdot \det A'_{ij} = (-1)^{i+j} \cdot \det A'_{ij}, \quad (6.100)$$

wobei A'_{ij} die $(n-1) \times (n-1)$ -Matrix ist, die aus A durch Streichen der i -ten Zeile und j -ten Spalte entsteht. (Diese Notation wollen wir dabei ganz allgemein einführen, nicht nur für Matrizen, deren j -ter Spaltenvektor e_i ist.)

Damit können wir die Determinante einer beliebigen Matrix auch wie folgt ausrechnen: Wir wählen eine Spalte, z.B. die j -te Spalte v_j . Diese können wir als Linearkombination der kanonischen Einheitsvektoren schreiben,

$$v_j = \sum_{i=1}^n a_{ij} e_i. \quad (6.101)$$

Mit Hilfe der Linearität der Determinante ergibt sich dann

$$\det A = \det(v_1 \cdots v_{j-1} v_j v_{j+1} \cdots v_n) \quad (6.102)$$

$$= \det(v_1 \cdots v_{j-1} \sum_{i=1}^n a_{ij} e_i v_{j+1} \cdots v_n) \quad (6.103)$$

$$= \sum_{i=1}^n a_{ij} \det(v_1 \cdots v_{j-1} e_i v_{j+1} \cdots v_n) \quad (6.104)$$

$$= \sum_{i=1}^n (-1)^{i+j} \cdot a_{ij} \cdot \det A'_{ij}. \quad (6.105)$$

Dies ist der **Laplacesche Entwicklungssatz**:

Satz 6.8. Sei A eine (reelle oder komplexe) $n \times n$ -Matrix ($n \geq 2$). Dann gilt für jedes $j \in \{1, \dots, n\}$,

$$\det A = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A'_{ij}, \quad (6.106)$$

sowie für jedes $i \in \{1, \dots, n\}$

$$\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det A'_{ij}. \quad (6.107)$$

Die erste Gleichung beschreibt die Entwicklung nach der j -ten Spalte, die zweite die analoge Entwicklung nach der i -ten Zeile.

Zur konkreten Berechnung einer Determinante ist dieser Satz nur nützlich, wenn in einer Zeile oder Spalte viele Nullen stehen:

Beispiel 6.2.

$$\det \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} = -1 \cdot \det \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} + 1 \cdot \det \begin{pmatrix} 0 & 2 \\ 3 & 1 \end{pmatrix} - 0 \cdot \det \begin{pmatrix} 0 & 1 \\ 3 & 2 \end{pmatrix} \quad (6.108)$$

Entw. nach der 2. Zeile

$$= -1 \cdot (1 - 4) + 1 \cdot (-6) = -3 \quad (6.109)$$

Inverse Matrix und Cramersche Regel

Der Laplacesche Entwicklungssatz ermöglicht es, explizite Ausdrücke für die inverse Matrix und für die Lösung linearer Gleichungssysteme (Cramersche Regel) anzugeben, die wir bisher nur algorithmisch bestimmt hatten.

Dazu betrachten wir die Matrix $\hat{A}^{(i,k)}$, die aus A entsteht, indem die k -te Zeile durch die i -te ersetzt wird, die i -te aber unverändert bleibt.

$$\hat{A}^{(i,k)} = \begin{matrix} & i & & k \\ & \begin{pmatrix} \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \end{pmatrix} & & \end{matrix} \quad (6.110)$$

Für $i = k$ stimmt diese Matrix mit A überein. Für $i \neq k$ ist der Rang von $\hat{A}^{(i,k)}$ nicht maximal, also gilt

$$\det \hat{A}^{(i,k)} = \begin{cases} 0 & \text{für } i \neq k \\ \det A & \text{für } i = k \end{cases} = \delta_{ik} \det A. \quad (6.111)$$

Die Entwicklung nach der k -ten Zeile liefert dann

$$\delta_{ik} \det A = \det \hat{A}^{(i,k)} = \sum_{j=1}^n (-1)^{k+j} a_{ij} \det A'_{kj} \quad (6.112)$$

$$= \sum_{j=1}^n a_{ij} \underbrace{(-1)^{k+j} \det A'_{kj}}_{=: c_{jk}}. \quad (6.113)$$

Die Matrix mit den Komponenten c_{jk} nennen wir die **Adjunkte** von A ,

$$\text{Adj } A = (c_{ij}) = \begin{pmatrix} \det A'_{11} & -\det A'_{12} & \dots & (-1)^{n+1} \det A'_{1n} \\ -\det A'_{21} & & & (-1)^{n+2} \det A'_{2n} \\ \vdots & & & \vdots \\ (-1)^{n+1} \det A'_{n1} & \dots & \dots & \det A'_{nn} \end{pmatrix}^t. \quad (6.114)$$

Man beachte, dass auf der rechten Seite die Matrix noch transponiert wird. Wie wir gesehen haben, gilt $(\det A) E_n = A (\text{Adj } A)$, also

$$A^{-1} = \frac{1}{\det A} \cdot \text{Adj } A, \quad (6.115)$$

sofern A^{-1} existiert.

Beispiel 6.3.

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ 4 & 0 & 5 \end{pmatrix} \quad (6.116)$$

$$\text{Adj } A = \begin{pmatrix} 15 & 0 & -12 \\ 0 & -3 & 0 \\ -6 & 0 & 3 \end{pmatrix}^t = \begin{pmatrix} 15 & 0 & -6 \\ 0 & -3 & 0 \\ -12 & 0 & 3 \end{pmatrix} \quad (6.117)$$

$$\det A = 3 \cdot \det \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} = -9 \quad (6.118)$$

$$\Rightarrow A^{-1} = -\frac{1}{9} \cdot \begin{pmatrix} 15 & 0 & -6 \\ 0 & -3 & 0 \\ -12 & 0 & 3 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -5 & 0 & 2 \\ 0 & 1 & 0 \\ 4 & 0 & -1 \end{pmatrix} \quad (6.119)$$

Test:

$$\frac{1}{3} \begin{pmatrix} -5 & 0 & 2 \\ 0 & 1 & 0 \\ 4 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ 4 & 0 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (6.120)$$

Für eine 2×2 -Matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ergibt sich

$$\text{Adj } A = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}^t = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \quad (6.121)$$

also

$$A^{-1} = \frac{1}{ad - bc} \cdot \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad (6.122)$$

für invertierbare Matrizen A .

Für große Matrizen ist dies kein praktisches Berechnungsverfahren. Andererseits ist eine explizite Formel nützlich, um etwas über die Stetigkeit oder Differenzierbarkeit der Inversenbildung auszusagen,

$$\begin{aligned} \cdot^{-1} : GL(n, \mathbb{R}) &\longrightarrow GL(n, \mathbb{R}) \\ A &\longmapsto A^{-1}. \end{aligned}$$

Für Gleichungssysteme $Ax = b$ mit quadratischer invertierbarer Matrix A ist die eindeutige Lösung gegeben durch

$$x = A^{-1}b. \quad (6.123)$$

Verwenden wir den obigen Ausdruck für A^{-1} , so erhalten wir

$$x = \frac{1}{\det A} (\text{Adj } A) b, \quad (6.124)$$

also

$$x_j = \frac{1}{\det A} \sum_{i=1}^n \underbrace{(-1)^{i+j} \det A'_{ij}}_{c_{ji}} b_i. \quad (6.125)$$

Die auftretende Summe hat die Form einer Entwicklung nach der j -ten Spalte der Matrix

$$B_j = \begin{pmatrix} a_{11} & \cdots & a_{1j-1} & b_1 & a_{1j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj-1} & b_1 & a_{nj+1} & \cdots & a_{nn} \end{pmatrix}. \quad (6.126)$$

Also gilt die **Cramersche Regel**

$$x_j = \frac{\det B_j}{\det A}. \quad (6.127)$$

Zur Berechnung der Lösung x ist diese Regel wiederum nur von begrenztem Nutzen für praktische Belange, außer bei kleinen Matrizen oder wenn man sich nicht für alle x_i interessiert, sondern nur für wenige Komponenten.

Theoretisch ist so ein expliziter Ausdruck aber wieder interessant, weil er z.B. zeigt, dass die Lösung x stetig von den Koeffizienten von A und b abhängt.

Volumen

Eingangs haben wir die Determinante eingeführt bzw. motiviert als einen Faktor, der angibt, wie sich das Volumen durch eine Abbildung $x \mapsto Ax$ verändert. Die Frage nach der Volumensänderung kann man nun auch für Abbildungen $\mathbb{R}^n \rightarrow \mathbb{R}^m$ mit $n \leq m$ diskutieren. Hat die Abbildung $x \mapsto Ax$ Rang n , so ist das Bild ein n -dimensionaler Untervektorraum von \mathbb{R}^m . Ein „Einheitskubus“ in \mathbb{R}^n wird dann auf ein n -dimensionales Parallelotop im \mathbb{R}^m abgebildet, dessen n -dimensionales Volumen wir bestimmen können.

Im \mathbb{R}^n ist das Volumen eines von v_1, \dots, v_n aufgespannten Parallelotops gegeben durch

$$\text{vol}_n P(v_1, \dots, v_n) = |\det(v_1 \cdots v_n)|. \quad (6.128)$$

Dieser Ausdruck ergibt keinen Sinn mehr, wenn die $v_i \in \mathbb{R}^m$ mit $m \neq n$. Der obige Ausdruck hängt von den Koeffizienten der v_i bezüglich der kanonischen Basis e_1, \dots, e_n ab. Wir würden erwarten, dass es eine Darstellung gibt, die von der Wahl der Basis nicht abhängt, sondern lediglich von den Längen der Kantenvektoren und den Winkeln. Diese Information ist in den Skalarprodukten $\langle v_i, v_j \rangle$ enthalten.

Für $A = (v_1 \cdots v_n)$ gilt $A^t A = (\langle v_i, v_j \rangle)_{i,j=1,\dots,n}$. Im \mathbb{R}^n gilt also

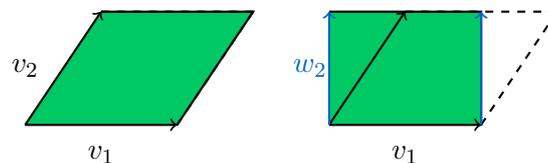
$$\text{vol}_n(v_1, \dots, v_n) = |\det A| = \sqrt{(\det A)^2} = \sqrt{\det(A^t A)} = \sqrt{\det(\langle v_i, v_j \rangle)}, \quad (6.129)$$

d.h. es gibt tatsächlich einen Ausdruck für das Volumen, der nur von den Längen und Winkeln abhängt.

Diesem Ausdruck ist es nun egal, in welchem Raum die v_i leben, daher kann er auch für n -dimensionale Parallelotope im \mathbb{R}^m angewendet werden.

Anschaulich sollte das klar sein: Der Flächeninhalt eines Parallelogramms im dreidimensionalen Raum hängt nicht von seiner Lage im Raum ab, sondern nur von den Kantenlängen und dem Winkel.

Beispiel 6.4. Zwei Vektoren $v_1, v_2 \in \mathbb{R}^n$ spannen ein Parallelogramm auf. Die Fläche ergibt sich aus der Länge des einen Vektors, $\|v_1\|$, multipliziert mit dem zu v_1 orthogonalen Anteil von v_2 (der Höhe des Parallelogramms über v_1), also der Länge eines Vektors w_2 von der Form $w_2 = v_2 + \lambda \cdot v_1$ mit $\langle w_2, v_1 \rangle = 0$.



Wir können also jetzt den Flächeninhalt des von v_1, v_2 aufgespannten Parallelogramms bestimmen:

- Zuerst bestimmen wir w_2 :

$$0 = \langle w_2, v_1 \rangle = \langle v_2 + \lambda \cdot v_1, v_1 \rangle = \langle v_2, v_1 \rangle + \lambda \cdot \langle v_1, v_1 \rangle \quad (6.130)$$

$$\implies \lambda = -\frac{\langle v_1, v_2 \rangle}{\|v_1\|^2}. \quad (6.131)$$

- Dann ergibt sich der Flächeninhalt zu

$$\text{vol}_2(v_1, v_2) = \text{vol}_2(v_1, w_2) \quad (6.132)$$

$$= \|v_1\| \|w_2\| \quad (6.133)$$

$$= \|v_1\| \cdot \left\| v_2 - \frac{\langle v_1, v_2 \rangle}{\|v_1\|^2} \cdot v_1 \right\| \quad (6.134)$$

$$= \|v_1\| \cdot \sqrt{\|v_2\|^2 + \frac{\langle v_1, v_2 \rangle^2}{\|v_1\|^2} - 2 \cdot \frac{\langle v_1, v_2 \rangle^2}{\|v_1\|^2}} \quad (6.135)$$

$$= \sqrt{\|v_1\|^2 \cdot \|v_2\|^2 - \langle v_1, v_2 \rangle^2} \quad (6.136)$$

$$= \sqrt{\det \begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_2, v_1 \rangle \\ \langle v_1, v_2 \rangle & \langle v_2, v_2 \rangle \end{pmatrix}}, \quad (6.137)$$

im Einklang mit unserer allgemeinen Formel.

Um den Ausdruck im Allgemeinen zu beweisen, müssten wir das n -dimensionale Volumen eines Parallelotops im \mathbb{R}^m zunächst einführen; z.B. als eine Funktion, die invariant ist unter gemeinsamer Rotation aller Kanten und die im Untervektorraum $\{(x_1, \dots, x_n, 0, \dots, 0) \in \mathbb{R}^m\}$ mit dem aus \mathbb{R}^n übernommenen Volumenbegriff übereinstimmt.

Man kann das Volumen eines n -dimensionalen Parallelotops auch wieder über seine Eigenschaften definieren als Abbildung

$$F : \mathbb{R}^m \times \dots \times \mathbb{R}^m \rightarrow \mathbb{R} \quad (6.138)$$

für die für alle $v_1, \dots, v_n \in \mathbb{R}^m$, $\lambda \in \mathbb{R}$, $i, j \in \{1, \dots, n\}$ ($i \neq j$) gilt:

- $F(\dots, v_i + \lambda \cdot v_j, \dots) = F(\dots)$
- $F(\dots, \lambda \cdot v_i, \dots) = |\lambda| \cdot F(\dots)$
- $F(v_1, \dots, v_n) = 1$, wenn für alle $i, j \in \{1, \dots, n\}$ gilt $\langle v_i, v_j \rangle = \delta_{ij}$, wenn also v_1, \dots, v_n ein Orthonormalsystem bilden.

Man kann zeigen, dass

$$F(v_1, \dots, v_n) = \sqrt{\det(\langle v_i, v_j \rangle)} \quad (6.139)$$

die einzige Abbildung mit diesen Eigenschaften ist.

(Das folgt aus dem später zu diskutierenden Orthonormalisierungsverfahren: Wir können zu (v_1, \dots, v_n) durch Bilden von Linearkombinationen ein Orthonormalsystem (w_1, \dots, w_n)

konstruieren. Durch Verschieben und Skalieren führen wir das durch (v_1, \dots, v_n) gegebene Parallelotop dabei in einen Einheitskubus über, dessen Volumen zu 1 gesetzt wird.)

6.4 Vektorprodukt

Wir schließen das Kapitel mit einer kurzen Diskussion des Vektorprodukts im \mathbb{R}^3 ab. Aus der expliziten Form von $\det A$ für eine 3×3 -Matrix sehen wir, dass

$$\det A = \sum_{i_1, i_2, i_3=1}^3 \epsilon_{i_1 i_2 i_3} a_{i_1 1} a_{i_2 2} a_{i_3 3}. \quad (6.140)$$

Wir bezeichnen Spaltenvektoren von A mit v_1, v_2, v_3 . $a_{i_1 1}$ sind die Koeffizienten des Spaltenvektors v_1 , $a_{i_2 2}$ von v_2 , $a_{i_3 3}$ von v_3 . Wir können aus v_2, v_3 einen neuen Vektor $u := v_2 \times v_3$ bilden, wobei wir das **Kreuzprodukt** bzw. **Vektorprodukt** $c = a \times b$ definieren durch

$$c_i = \sum_{j,k=1}^3 \epsilon_{ijk} a_j b_k. \quad (6.141)$$

Dann ist

$$u_i = \sum_{j,k=1}^3 \epsilon_{ijk} a_{j2} a_{k3} \quad (6.142)$$

und somit

$$\det A = \det(v_1, v_2, v_3) = \langle v_1, u \rangle = \langle v_1, v_2 \times v_3 \rangle. \quad (6.143)$$

Das Volumen $|\det A|$ des von v_1, v_2, v_3 aufgespannten Parallelotops ist das Produkt der Fläche des von v_2 und v_3 aufgespannten Parallelogramms mit der Höhe, also dem auf dem Parallelogramm senkrechten Anteil von v_1 . Aus dieser Interpretation sieht man schon, dass $v_2 \times v_3$ ein Vektor sein muss, der senkrecht auf v_2 und v_3 steht und dessen Länge gerade der Flächeninhalt des von v_2 und v_3 aufgespannten Parallelogramms ist.

Wir untersuchen jetzt die Eigenschaften des Vektorprodukts:

1. Das Vektorprodukt ist explizit gegeben durch

$$(u \times v) = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}. \quad (6.144)$$

2. Es definiert eine Verknüpfung auf dem \mathbb{R}^3 :

$$\times : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3 \quad (6.145)$$

$$(u, v) \mapsto u \times v. \quad (6.146)$$

3. Diese Verknüpfung ist offensichtlich linear in jedem Eintrag, d.h. für alle $u, u', v, v' \in \mathbb{R}^3, \lambda \in \mathbb{R}$ gilt

$$(u + u') \times v = u \times v + u' \times v \quad (6.147)$$

$$(\lambda \cdot u) \times v = \lambda \cdot (u \times v) \quad (6.148)$$

$$u \times (v + v') = u \times v + u \times v' \quad (6.149)$$

$$u \times (\lambda \cdot v) = \lambda \cdot (u \times v). \quad (6.150)$$

4. Das Vektorprodukt ist antisymmetrisch, d.h. für alle $u, v \in \mathbb{R}^3$ gilt

$$u \times v = -v \times u. \quad (6.151)$$

5. Die durch das Vektorprodukt definierte Verknüpfung ist aber nicht assoziativ. Stattdessen gilt

$$(u \times v) \times w = \langle u, w \rangle \cdot v - \langle v, w \rangle \cdot u. \quad (6.152)$$

Das letzte Resultat kann man z.B. mit den Eigenschaften des ϵ -Tensors beweisen. Insbesondere ist folgende Identität nützlich:

$$\text{für alle } j, k, l, m \in \{1, 2, 3\} : \sum_{i=1}^3 \epsilon_{ijk} \epsilon_{ilm} = \delta_{jl} \delta_{km} - \delta_{jm} \delta_{kl}. \quad (6.153)$$

Beweis der obigen Identität. Seien j, k gegeben. Wenn $j = k$, sind beide Seiten gleich Null. Ansonsten gibt es genau ein i , sodass (i, j, k) eine Permutation der Zahlen $(1, 2, 3)$ ist. Dann ist die linke Seite nur dann ungleich Null, wenn auch (i, l, m) eine Permutation von $(1, 2, 3)$ ist, d.h. wenn $(i, l, m) = (i, j, k)$ oder wenn $(i, m, l) = (i, j, k)$.

Im ersten Fall ist $l = j$ und $m = k$ und die linke Seite ergibt 1. Im zweiten Fall ist $l = k$ und $m = j$ und das Resultat ist -1 . Die rechte Seite hat die gleichen Eigenschaften. \square

Beweis von (6.152).

$$((u \times v) \times w)_i = \sum_{j,k=1}^3 \epsilon_{ijk} (u \times v)_j w_k \quad (6.154)$$

$$= \sum_{j,k,l,m=1}^3 \epsilon_{ijk} \epsilon_{jlm} u_l v_m w_k \quad (6.155)$$

$$= \sum_{k,l,m=1}^3 (\delta_{kl} \delta_{im} - \delta_{km} \delta_{il}) u_l v_m w_k \quad (\epsilon_{ijk} = \epsilon_{jki}) \quad (6.156)$$

$$= \sum_{k=1}^3 (u_k w_k v_i - v_k w_k u_i) \quad (6.157)$$

$$= \langle u, w \rangle v_i - \langle v, w \rangle u_i. \quad (6.158)$$

\square

Eine weitere wichtige Eigenschaft ist

$$\langle u \times v, w \times z \rangle = \sum_i (u \times v)_i (w \times z)_i \quad (6.159)$$

$$= \sum_{i,j,k,j',k'} \epsilon_{ijk} u_j v_k \epsilon_{ij'k'} w_{j'} z_{k'} \quad (6.160)$$

$$= \sum_{j,k,j',k'} (\delta_{jj'} \delta_{kk'} - \delta_{jk'} \delta_{kj'}) u_j v_k w_{j'} z_{k'} \quad (6.161)$$

$$= \sum_{j,k} (u_j w_j v_k z_k - u_j z_j v_k w_k) \quad (6.162)$$

$$= \langle u, w \rangle \langle v, z \rangle - \langle u, z \rangle \langle v, w \rangle . \quad (6.163)$$

Aus der letzten Relation folgt auch

$$\|u \times v\|^2 = \langle u \times v, u \times v \rangle \quad (6.164)$$

$$= \langle u, u \rangle \langle v, v \rangle - \langle u, v \rangle \langle v, u \rangle \quad (6.165)$$

$$= \|u\|^2 \|v\|^2 - \langle u, v \rangle^2 . \quad (6.166)$$

Also stimmt die Länge von $u \times v$ mit dem Flächeninhalt des von u, v gebildeten Parallelogramms überein.

Außerdem ist $u \times v$ wie gewünscht orthogonal zu u und v :

$$\langle u \times v, u \rangle = \sum_i (u \times v)_i u_i = \sum_{ijk=1}^3 \underbrace{\epsilon_{ijk}}_{\text{antisym.}} \underbrace{u_j v_k u_i}_{\text{sym. in } i, j} = 0 . \quad (6.167)$$

In höheren Dimensionen können wir etwas Ähnliches machen. Die Determinante einer Matrix A mit Spaltenvektoren v_1, \dots, v_n ist

$$\det A = \sum_{i_1, \dots, i_n} \epsilon_{i_1 \dots i_n} a_{i_1 1} \dots a_{i_n n} \quad (6.168)$$

$$= \sum_{i_1=1}^n a_{i_1 1} \underbrace{\left(\sum_{i_2, \dots, i_n=1}^n \epsilon_{i_1 \dots i_n} a_{i_2 2} \dots a_{i_n n} \right)}_{=: w_{i_1}} \quad (6.169)$$

$$= \langle v_1, w \rangle . \quad (6.170)$$

Der so definierte Vektor w steht senkrecht auf allen v_2, \dots, v_n und hat als Länge das Volumen des von v_2, \dots, v_n aufgespannten Parallelotops. Das so definierte verallgemeinerte Vektorprodukt ordnet $n - 1$ Vektoren im \mathbb{R}^n einen Vektor zu.

Zusammenfassung Kapitel 6

- Wir haben die Determinante als Maß für die Volumensänderung unter der Abbildung $x \mapsto Ax$ motiviert.
- Ausgehend von der konkreten Definition mittels Permutationen bzw. dem ϵ -Symbol haben wir die Eigenschaften der Determinante untersucht (und auch gesehen, dass schon wenige dieser Eigenschaften genügen, um die Determinante eindeutig zu charakterisieren).
- Die Eigenschaften erlauben es, die Determinante mit Hilfe von Zeilen- und Spaltentransformationen zu berechnen.
- Wichtige Ergebnisse über Determinanten sind:
 - Multiplikationssatz: $\det(AB) = \det A \cdot \det B$
 - Determinante der Transponierten: $\det A^t = \det A$
 - Determinante der Inversen: $\det A^{-1} = (\det A)^{-1}$
 - Kästchensatz: $\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \det A \cdot \det C$
 - Laplacescher Entwicklungssatz
- Anwendungen:
 - Formel für A^{-1} über Adjunkte
 - Cramersche Regel
 - Volumen von Parallelotopen: $\text{vol}_n P(v_1, \dots, v_n) = \sqrt{\det(\langle v_i, v_j \rangle)}$

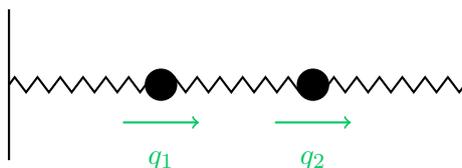
7 Eigenwerte und Normalformen

7.1 Motivation

Wir wollen uns zunächst ein Beispiel aus der Physik anschauen, in dem die in diesem Kapitel behandelten Konzepte auftauchen.

Wir hatten schon die Bewegungsgleichung des harmonischen Oszillators betrachtet, $\ddot{q} = -\omega^2 q$ (wir erinnern uns: Der Lösungsraum ist ein zweidimensionaler Vektorraum).

Wir betrachten nun zwei gekoppelte Oszillatoren: Die Auslenkungen der gleichen Massen



bezeichnen wir mit q_1 und q_2 .

Ohne die mittlere Feder hätten wir zwei harmonische Oszillatoren, $\ddot{q}_1 = -\omega^2 q_1$, $\ddot{q}_2 = -\omega^2 q_2$, mit gleicher Schwingungsfrequenz (bei gleichen Federn und gleichen Massen).

Die mittlere Feder übt eine zusätzliche Kraft auf die beiden Massen aus, die proportional zur Differenz $q_1 - q_2$ ist, d.h. wir bekommen die gekoppelten Differentialgleichungen

$$\ddot{q}_1 = -\omega^2 q_1 - \alpha \omega^2 (q_1 - q_2) \quad (7.1)$$

$$\ddot{q}_2 = -\omega^2 q_2 + \alpha \omega^2 (q_1 - q_2). \quad (7.2)$$

Die Konstante $\alpha > 0$ hängt von den Eigenschaften der Federn ab (sie bestimmt sich aus dem Verhältnis der Federkonstanten).

Wie lösen wir diese Bewegungsgleichungen? Zunächst schreiben wir sie in Vektorschreibweise um:

$$\begin{pmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{pmatrix} = \begin{pmatrix} -(1 + \alpha)\omega^2 q_1 & \alpha\omega^2 q_2 \\ \alpha\omega^2 q_1 & -(1 + \alpha)\omega^2 q_2 \end{pmatrix} \quad (7.3)$$

$$= -\omega^2 \cdot \underbrace{\begin{pmatrix} 1 + \alpha & -\alpha \\ -\alpha & 1 + \alpha \end{pmatrix}}_{=:A} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}. \quad (7.4)$$

Schreiben wir $q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$, so lautet die Gleichung

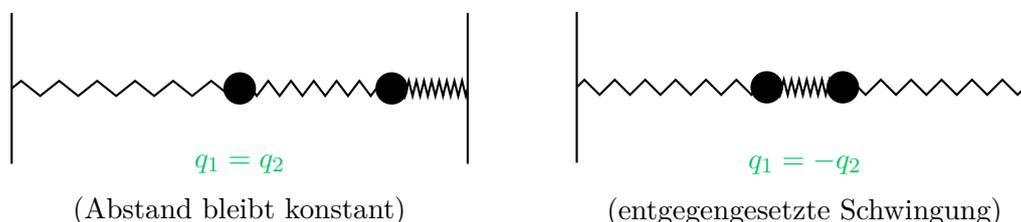
$$\ddot{q} = -\omega^2 \cdot Aq. \quad (7.5)$$

Formal könnten wir Lösungen von der Form

$$„ q(t) = \cos(\omega\sqrt{A}t) \cdot v + \sin(\omega\sqrt{A}t) \cdot w “ \quad (7.6)$$

erwarten, aber wie geben wir einem Ausdruck wie $\cos(\omega\sqrt{At})$ einen Sinn? Das wollen wir jetzt nicht weiter verfolgen, werden aber später sehen, dass dies unter bestimmten Umständen möglich ist.

Wir werden erst einmal unserer physikalischen Intuition folgen. Aus Symmetriegründen erwarten wir, dass es zwei einfache Typen von Lösungen gibt:



Daher liegt es nahe, neue Koordinaten einzuführen,

$$q_+ = \frac{1}{\sqrt{2}}(q_1 + q_2) \quad (7.7)$$

$$q_- = \frac{1}{\sqrt{2}}(q_1 - q_2), \quad (7.8)$$

wobei der Faktor $1/\sqrt{2}$ zur zweckmäßigen Normierung dient.

In diesen Koordinaten erhalten wir

$$\ddot{q}_+ = \frac{1}{\sqrt{2}}(\ddot{q}_1 + \ddot{q}_2) = \frac{1}{\sqrt{2}}(-\omega^2 q_1 - \omega^2 q_2) = -\omega^2 q_+ \quad (7.9)$$

und

$$\ddot{q}_- = \frac{1}{\sqrt{2}}(\ddot{q}_1 - \ddot{q}_2) = \frac{1}{\sqrt{2}}(-(1 + 2\alpha)\omega^2 q_1 + (1 + 2\alpha)\omega^2 q_2) = -(1 + 2\alpha)\omega^2 q_-, \quad (7.10)$$

die Gleichungen zweier ungekoppelter Oszillatoren mit Frequenzen ω und $\sqrt{1 + 2\alpha}\omega$.

Es gibt also zwei Grundtypen von Schwingungen: eine mit Frequenz ω ,

$$q_+(t) = \cos \omega t \cdot a + \sin \omega t \cdot b \quad q_-(t) = 0, \quad (7.11)$$

und eine mit Frequenz $\sqrt{1 + 2\alpha}\omega$,

$$q_+(t) = 0 \quad q_-(t) = \cos(\sqrt{1 + 2\alpha}\omega t) \cdot c + \sin(\sqrt{1 + 2\alpha}\omega t) \cdot d. \quad (7.12)$$

Diese einfachen Lösungen nennt man **Eigenschwingungen**, die zugehörigen Frequenzen heißen **Eigenfrequenzen**.

Die allgemeine Lösung erhält man durch eine beliebige Wahl der vier Parameter a, b, c, d . Der Lösungsraum (der wegen der Linearität der Bewegungsgleichungen ein Vektorraum ist) ist ein vierdimensionaler Raum.

Die allgemeine Bewegung ist also eine Überlagerung von zwei regelmäßigen Schwingungen. In den ursprünglichen Koordinaten war diese Lösung aber nicht offensichtlich.

Was haben wir beim Koordinatenwechsel gemacht? Wir haben einen Basiswechsel im zweidimensionalen Konfigurationsraum durchgeführt, sodass die neuen Koordinaten $q' = \begin{pmatrix} q_+ \\ q_- \end{pmatrix}$ mit den alten $q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$ über eine invertierbare Matrix S zusammenhängen:

$$q' = \begin{pmatrix} q_+ \\ q_- \end{pmatrix} = \frac{1}{\sqrt{2}} \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_S \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = Sq. \quad (7.13)$$

Die inverse Matrix ist $S^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ (in diesem Fall die gleiche Matrix S), und es gilt

$$\begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = S^{-1}q' = \frac{1}{\sqrt{2}} \begin{pmatrix} q_+ + q_- \\ q_+ - q_- \end{pmatrix}. \quad (7.14)$$

Die Bewegungsgleichung für q' lautet

$$\ddot{q}' = (\ddot{S}q) = S\ddot{q} = -\omega^2 SAq = -\omega^2 SAS^{-1}q' \quad (7.15)$$

mit

$$SAS^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 + \alpha & -\alpha \\ -\alpha & 1 + \alpha \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (7.16)$$

$$= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 + 2\alpha & -(1 + 2\alpha) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (7.17)$$

$$= \frac{1}{2} \begin{pmatrix} 2 & 0 \\ 0 & 2(1 + 2\alpha) \end{pmatrix} \quad (7.18)$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 + 2\alpha \end{pmatrix}. \quad (7.19)$$

Nach dem Basiswechsel (Koordinatentransformation mit S) hat die transformierte Matrix Diagonalgestalt.

Matrizen A , die mit Hilfe einer invertierbaren Matrix S in Diagonalgestalt SAS^{-1} gebracht werden können, heißen **diagonalisierbar**. Die entsprechenden Diagonalelemente heißen **Eigenwerte** von A .

Die Diagonalisierbarkeit hängt offensichtlich damit zusammen, dass es spezielle Vektoren gibt, auf welche die Matrix besonders einfach wirkt. Die speziellen Lösungen mit $q_1 = q_2$ gehören zu einem Vektor $q = q_1 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = q_+ \cdot v_+$ mit $v_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Auf diesen wirkt die Matrix A als

$$Av_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 + \alpha & -\alpha \\ -\alpha & 1 + \alpha \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = v_+. \quad (7.20)$$

Lösungen mit $q_1 = -q_2$ gehören zu einem Vektor $q = q_1 \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = q_- \cdot v_-$ mit $v_- = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ und daher

$$Av_- = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 + \alpha & -\alpha \\ -\alpha & 1 + \alpha \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}(1 + 2\alpha) \begin{pmatrix} 1 \\ -1 \end{pmatrix} = (1 + 2\alpha)v_-. \quad (7.21)$$

Vektoren, auf welche die Matrix A eine derartig einfache Wirkung hat, heißen **Eigenvektoren**. Das Aufspüren solcher einfachen Richtungen hat eine große praktische Bedeutung (nicht nur für die Lösung von gekoppelten Differentialgleichungen).

7.2 Eigenwerte und Eigenvektoren

Nach dem einführenden Beispiel wollen wir die dort aufgetretenen Begrifflichkeiten allgemein definieren.

Definition 7.1. Sei $F : V \rightarrow V$ eine lineare Abbildung des reellen (oder komplexen) Vektorraums V in sich selbst (also ein Endomorphismus von V). Eine Zahl $\lambda \in \mathbb{R}$ (oder \mathbb{C}) heißt **Eigenwert** von F genau dann, wenn es ein $v \in V$, $v \neq 0$, gibt mit

$$F(v) = \lambda \cdot v. \quad (7.22)$$

Ein solcher Vektor v heißt **Eigenvektor** von F zum Eigenwert λ .

Bemerkung

Für $V = \mathbb{R}^n$ (\mathbb{C}^n) ist ein Endomorphismus F durch eine $n \times n$ -Matrix A gegeben. Die Eigenwertgleichung wird dann zu

$$Ax = \lambda \cdot x. \quad (7.23)$$

Man beachte, dass $v = 0$ nicht als Eigenvektor zugelassen wird, der Eigenwert $\lambda = 0$ ist aber erlaubt: Eigenvektoren von F zum Eigenwert 0 sind also Vektoren $v \neq 0$ mit

$$F(v) = 0, \quad (7.24)$$

also genau die nichttrivialen Elemente des Kerns von F .

Da die Null nicht enthalten ist, ist die Menge der Eigenvektoren zu einem gegebenen Eigenwert λ kein Untervektorraum; sobald man den Nullvektor aber dazu nimmt, wird es einer:

Definition 7.2. Sei F wie oben, $\lambda \in \mathbb{R}$ (oder \mathbb{C}). Der Untervektorraum

$$\text{Eig}(F, \lambda) = \{v \in V \mid F(v) = \lambda \cdot v\} = \text{Ker}(F - \lambda \cdot \text{id}_V) \quad (7.25)$$

heißt **Eigenraum** von F zu λ .

Ist λ kein Eigenwert von F , so ist $\text{Eig}(F, \lambda) = \{0\}$, ansonsten gilt

$$\text{Eig}(F, \lambda) \setminus \{0\} = \{\text{alle Eigenvektoren von } F \text{ zum Eigenwert } \lambda\}. \quad (7.26)$$

Für eine lineare Abbildung f_A auf \mathbb{R}^n (\mathbb{C}^n), die durch eine $n \times n$ -Matrix A gegeben ist, $f_A : x \mapsto Ax$, schreiben wir auch

$$\text{Eig}(A, \lambda) = \text{Ker}(A - \lambda \cdot E_n) \quad (7.27)$$

für den Eigenraum.

Ein Eigenvektor kann natürlich nicht zu verschiedenen Eigenwerten gehören, also gilt

$$\text{Eig}(F, \lambda_1) \cap \text{Eig}(F, \lambda_2) = \{0\} \quad \text{für } \lambda_1 \neq \lambda_2. \quad (7.28)$$

Zwei Eigenvektoren mit verschiedenen Eigenwerten sind also linear unabhängig. Das gilt allgemeiner:

Satz 7.1. *Seien v_1, \dots, v_m Eigenvektoren des Endomorphismus $F : V \rightarrow V$, deren Eigenwerte paarweise verschieden sind. Dann sind v_1, \dots, v_m linear unabhängig.*

Beweis. Wir bemerken zunächst, dass der Endomorphismus $F - \lambda_j \cdot \text{id}_V$ auf die Eigenvektoren von F wie folgt wirkt:

$$(F - \lambda_j \cdot \text{id}_V)(v_i) = \begin{cases} (\lambda_i - \lambda_j)v_i & i \neq j \\ 0 & i = j \end{cases}. \quad (7.29)$$

Um die lineare Abhängigkeit zu prüfen, setzen wir an:

$$\mu_1 \cdot v_1 + \dots + \mu_m \cdot v_m = 0. \quad (7.30)$$

Wir wählen ein beliebiges $i \in \{1, \dots, m\}$ und wirken auf diese Gleichung nacheinander mit allen $(F - \lambda_j \cdot \text{id}_V)$ für $i \neq j$. Nur der Term mit dem Vektor v_i bleibt übrig und wir finden

$$\mu_i \cdot (\lambda_i - \lambda_1) \cdot \dots \cdot (\lambda_i - \lambda_{i-1})(\lambda_i - \lambda_{i+1}) \cdot \dots \cdot (\lambda_i - \lambda_m) \cdot v_i = 0. \quad (7.31)$$

Da $v_i \neq 0$ und alle Eigenwerte als paarweise verschieden vorausgesetzt sind, folgt $\mu_i = 0$. Da i beliebig gewählt war, gilt $\mu_1 = \dots = \mu_m = 0$: (v_1, \dots, v_m) ist linear unabhängig. \square

Beispiel 7.1.

- $A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ bewirkt eine Drehung im \mathbb{R}^2 .

Es gibt keine invariante Richtung, also keine Eigenvektoren, es sei denn

– $\theta = 0$: Dann ist $A = E_2$ und alle Vektoren (außer 0) sind Eigenvektoren zum Eigenwert 1.

– $\theta = \pi$: Das beschreibt eine Drehung um 180 deg und $A = -E_2$, und alle Vektoren (außer 0) sind Eigenvektoren zum Eigenwert -1 .

- $A = \begin{pmatrix} \mu & 1 \\ 0 & \mu \end{pmatrix}$ wirkt als $A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mu x_1 + x_2 \\ \mu x_2 \end{pmatrix}$ auf \mathbb{R}^2 .

$Ax = \lambda x$ führt zu

$$\begin{pmatrix} \mu x_1 + x_2 \\ \mu x_2 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}. \quad (7.32)$$

x_2 muss Null sein: Wäre $x_2 \neq 0$, dann würde $\lambda = \mu$ aus der zweiten Zeile folgen und dann $x_2 = 0$ aus der ersten. Da wir nach einem Eigenvektor suchen,

darf x_1 nicht auch Null sein; dann ist $\lambda = \mu$ und jedes $x_1 \neq 0$ ist erlaubt.

A besitzt also einen Eigenwert $\lambda = \mu$ und

$$\text{Eig}(A, \mu) = \left\{ \begin{pmatrix} x_1 \\ 0 \end{pmatrix}, x_1 \in \mathbb{R} \right\}. \quad (7.33)$$

- $A = \begin{pmatrix} 1 + \alpha & -\alpha \\ -\alpha & 1 + \alpha \end{pmatrix}$ wie im Beispiel der gekoppelten Oszillatoren. Wie wir gesehen haben, gibt es zwei Eigenwerte 1 und $1 + 2\alpha$ mit Eigenräumen

$$\text{Eig}(A, 1) = \left\{ \begin{pmatrix} x \\ x \end{pmatrix} \mid x \in \mathbb{R} \right\} \quad (7.34)$$

$$\text{Eig}(A, 1 + 2\alpha) = \left\{ \begin{pmatrix} x \\ -x \end{pmatrix} \mid x \in \mathbb{R} \right\}. \quad (7.35)$$

In diesem letzten Beispiel gibt es also genügend viele Eigenvektoren, dass man eine Basis aus Eigenvektoren wählen kann. Durch die entsprechende Koordinatentransformation wird die Matrix in eine Diagonalmatrix überführt.

Das funktioniert allgemein: Gibt es zu einem Endomorphismus $F : V \rightarrow V$ eine Basis $\mathcal{A} = (v_1, \dots, v_n)$ von V bestehend aus Eigenvektoren von F , $F(v_i) = \lambda_i \cdot v_i$ ($i \in \{1, \dots, n\}$), dann wird F bezüglich dieser Basis durch eine Diagonalmatrix dargestellt:

$$M_{\mathcal{A}}^{\mathcal{A}}(F) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}. \quad (7.36)$$

Definition 7.3. Ein Endomorphismus $F : V \rightarrow V$ ($\dim V < \infty$) heißt **diagonalisierbar** genau dann, wenn es eine Basis gibt, bezüglich der F durch eine Diagonalmatrix dargestellt wird.

Bemerkung

Ein Endomorphismus $F : V \rightarrow V$ ist genau dann diagonalisierbar, wenn V eine Basis bestehend aus Eigenvektoren von F besitzt.

Da Eigenvektoren zu verschiedenen Eigenwerten linear unabhängig sind, finden wir folgendes (hinreichendes) Kriterium für die Diagonalisierbarkeit:

Satz 7.2. Sei $F : V \rightarrow V$ ein Endomorphismus mit $n = \dim V$ paarweise verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_n$. Dann ist F diagonalisierbar.

Dieses Kriterium ist hinreichend, aber nicht notwendig: Beispielsweise ist die Identität auf \mathbb{R}^2 sicherlich diagonalisierbar, obwohl es nur einen Eigenwert 1 gibt.

Wir wollen nun sehen, was wir dann aussagen können, wenn es weniger Eigenwerte als die Dimension von V gibt.

Ist F diagonalisierbar, kann jeder Vektor in V als Linearkombination von Eigenvektoren und damit (da jedes skalare Vielfache eines Eigenvektors wieder ein Eigenvektor mit gleichem Eigenwert ist) als Summe von Eigenvektoren dargestellt werden. Also gilt:

Satz 7.3. *Ein Endomorphismus $F : V \rightarrow V$ (mit $\dim V = n$) ist diagonalisierbar genau dann, wenn für die paarweise verschiedenen Eigenwerte $\lambda_1, \dots, \lambda_r$ gilt*

$$V = \text{Eig}(F, \lambda_1) + \dots + \text{Eig}(F, \lambda_r). \quad (7.37)$$

In dieser Summe von Untervektorräumen ist die Zerlegung jedes Vektors $v = v_1 + \dots + v_r$ wegen der linearen Unabhängigkeit der v_i eindeutig. Man spricht dann auch von einer **direkten Summe** und schreibt

$$V = \text{Eig}(F, \lambda_1) \oplus \dots \oplus \text{Eig}(F, \lambda_r). \quad (7.38)$$

Hat man zu jedem Eigenraum $\text{Eig}(F, \lambda_i)$ eine Basis $(v_1^{(i)}, \dots, v_{n_i}^{(i)})$, so ergeben alle diese Vektoren zusammen eine Basis

$$(v_1^{(1)}, \dots, v_{n_1}^{(1)}, v_1^{(2)}, \dots, v_{n_2}^{(2)}, \dots, v_1^{(r)}, \dots, v_{n_r}^{(r)}) \quad (7.39)$$

von V . Also gilt dann $n_1 + \dots + n_r = n = \dim V$.

Satz 7.4. *$F : V \rightarrow V$ ($\dim V < \infty$) ist diagonalisierbar genau dann, wenn für die (paarweise verschiedenen) Eigenwerte $\lambda_1, \dots, \lambda_r$ gilt*

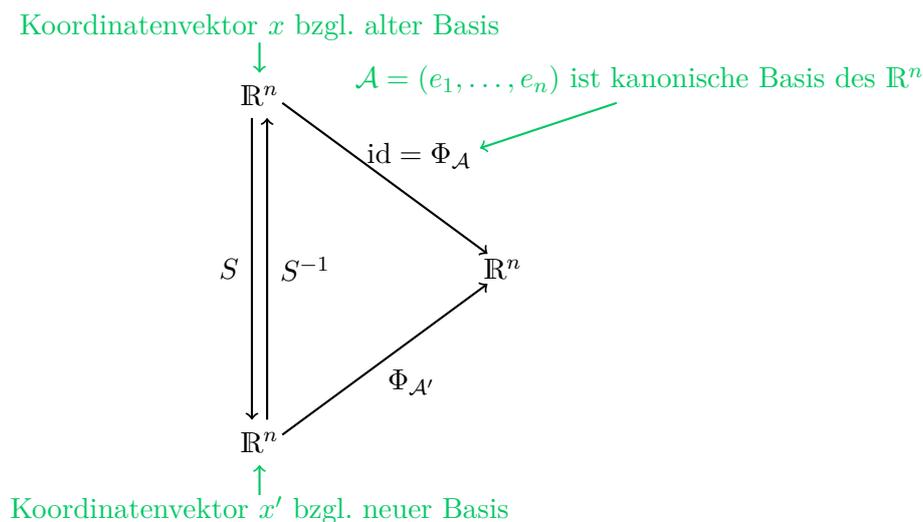
$$\dim \text{Eig}(F, \lambda_1) + \dots + \dim \text{Eig}(F, \lambda_r) = \dim V. \quad (7.40)$$

Die Diagonalisierbarkeit können wir also überprüfen, indem wir alle Eigenwerte von F bestimmen und dann die Dimensionen der zugehörigen Eigenräume ermitteln.

Bevor wir uns genauer anschauen, wie man das machen kann, wollen wir uns den Fall von Endomorphismen auf \mathbb{R}^n genauer anschauen.

Eine $n \times n$ -Matrix A nennen wir diagonalisierbar, wenn die lineare Abbildung $f_A : x \mapsto Ax$ diagonalisierbar ist. Wenn wir zu A eine linear unabhängige Familie von n Eigenvektoren (v_1, \dots, v_n) gefunden haben, dann wird der Basiswechsel von $\mathcal{A} = (e_1, \dots, e_n)$ zu $\mathcal{A}' = (v_1, \dots, v_n)$ durch eine Matrix S beschrieben mit Spaltenvektoren $S^{-1} = (v_1 v_2 \cdots v_n)$.

Das können wir wie folgt einsehen. Ein Basiswechsel von der kanonischen Basis $\mathcal{A} = (e_1, \dots, e_n)$ vom \mathbb{R}^n (für \mathbb{C}^n funktioniert alles analog) zu einer Basis \mathcal{A}' wird durch das folgende Diagramm beschrieben:



Dabei ist die Abbildung von den neuen zu den alten Koordinaten gegeben durch

$$x' \mapsto \Phi_{\mathcal{A}'}(x') = \sum_{i=1}^n x'_i \cdot v_i = \underbrace{(v_1 v_2 \dots v_n)}_{S^{-1}} \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}. \quad (7.41)$$

Die Matrixdarstellung bezüglich der neuen Basis ist dann

$$B = S A S^{-1}. \quad (7.42)$$

B ist diagonal: Wenn $A v_i = \lambda_i \cdot v_i$, dann ist

$$B = S A S^{-1} = S A (v_1 \dots v_n) \quad (7.43)$$

$$= S (\lambda_1 \cdot v_1 \dots \lambda_n \cdot v_n) \quad (7.44)$$

$$= S (v_1 \dots v_n) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \quad (7.45)$$

$$= S S^{-1} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}. \quad (7.46)$$

A ist also genau dann diagonalisierbar, wenn es eine invertierbare Matrix S gibt, sodass $B = S A S^{-1}$ diagonal ist. Die Spaltenvektoren von S^{-1} sind dann Eigenvektoren von A , die Diagonaleinträge von B sind die Eigenwerte.

Wie bestimmen wir jetzt Eigenwerte und Eigenvektoren einer Matrix A ? Wenn wir erst einmal die Eigenwerte λ_i von A kennen, können wir durch Lösung des homogenen Gleichungssystems

$$A x = \lambda_i \cdot x \quad \implies \quad (A - \lambda_i E_n)x = 0 \quad (7.47)$$

die Eigenvektoren bestimmen. Wir müssen uns also jetzt damit beschäftigen, wie wir etwas über die Eigenwerte herausfinden.

7.3 Das charakteristische Polynom

Wie gerade besprochen, ergeben sich die Eigenvektoren aus der Lösung des Systems

$$(A - \lambda \cdot E_n)x = 0. \quad (7.48)$$

Wir müssen also wissen, zu welchen λ es überhaupt nichttriviale Lösungen dieser Gleichung gibt. Dies ist genau dann der Fall, wenn die Matrix $A - \lambda \cdot E_n$ nicht maximalen Rang hat, also wenn

$$\det(A - \lambda E_n) = 0. \quad (7.49)$$

Die linke Seite ist ein Polynom in λ mit maximaler Potenz n .

Definition 7.4. Zu einer (reellen oder komplexen) $n \times n$ -Matrix A definieren wir das **charakteristische Polynom**

$$P_A(t) = \det(A - t E_n). \quad (7.50)$$

Bemerkung

Für einen Endomorphismus $F : V \rightarrow V$ auf einem n -dimensionalen Vektorraum V können wir auch das charakteristische Polynom definieren als

$$P_F(t) = \det(F - t \cdot \text{id}_V), \quad (7.51)$$

da wir die Determinante auch für Endomorphismen definieren können. In einer Basis \mathcal{A} ist dann die darstellende Matrix

$$M_{\mathcal{A}}^{\mathcal{A}}(F - t \text{id}_V) = M_{\mathcal{A}}^{\mathcal{A}}(F) - t E_n. \quad (7.52)$$

Dass die Determinante unabhängig ist von der Basis, bezüglich der wir den Endomorphismus darstellen, hatten wir schon gesehen. Wir wollen dies trotzdem noch einmal explizit zeigen:

Sei $A = M_{\mathcal{A}}^{\mathcal{A}}(F)$ und $B = SAS^{-1} = M_{\mathcal{B}}^{\mathcal{B}}(F)$ die Matrixdarstellung bezüglich einer anderen Basis \mathcal{B} . Dann ist

$$P_B(t) = \det(B - t E_n) \quad (7.53)$$

$$= \det(SAS^{-1} - t E_n) \quad (7.54)$$

$$= \det(S(A - t E_n)S^{-1}) \quad (7.55)$$

$$= \det(A - t E_n) \quad (7.56)$$

$$= P_A(t). \quad (7.57)$$

Wir wollen uns nun die Struktur des charakteristischen Polynoms genauer ansehen:

$$P_A(t) = \det(A - t E_n) \tag{7.58}$$

$$= \det \begin{pmatrix} a_{11} - t & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - t & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} - t \end{pmatrix} \tag{7.59}$$

$$= (a_{11} - t) \cdot (a_{22} - t) \cdot \dots \cdot (a_{nn} - t) + \underbrace{\dots}_{\text{höchste Potenz } t^{n-2}}. \tag{7.60}$$

Hier haben wir die Determinante über die explizite Formel

$$\det A = \sum_{i_1, \dots, i_n} \epsilon_{i_1 \dots i_n} a_{i_1 1} \cdot \dots \cdot a_{i_n n} \tag{7.61}$$

ausgewertet. Für $i_1 = 1, \dots, i_n = n$ bekommen wir als Beitrag das Produkt aller Diagonalelemente. Für jede andere Permutation i_1, \dots, i_n der Zahlen $1, \dots, n$ sind mindestens zwei Zahlen nicht an ihrem Platz (also $i_a \neq a$ und $i_b \neq b$), also enthält jeder andere Beitrag höchstens $n - 2$ Diagonalelemente und trägt somit höchstens eine Potenz t^{n-2} bei.

Durch Ausmultiplizieren der Diagonalterme ergibt sich

$$P_A(t) = \det(A - t \cdot E_n) = (-t)^n + (-t)^{n-1}(a_{11} + a_{22} + \dots + a_{nn}) + \underbrace{\dots}_{\text{höchste Potenz } t^{n-2}}. \tag{7.62}$$

Weiter wissen wir, dass der Term mit Potenz t^0 , der also kein t enthält, mit der Determinante $\det A$ übereinstimmt, $P_A(0) = \det A$.

Wir finden also

$$P_A(t) = \sum_{i=0}^n \alpha_i t^i \tag{7.63}$$

$$= (-1)^n \cdot t^n + (-1)^{n-1} (\text{tr } A) t^{n-1} + \alpha_{n-2} t^{n-2} + \dots + \alpha_1 t^1 + \det A, \tag{7.64}$$

wobei wir die sogenannte **Spur** („trace“) von A definiert haben als die Summe der Diagonalelemente,

$$\text{tr } A := a_{11} + \dots + a_{nn}. \tag{7.65}$$

Die anderen Koeffizienten α_i lassen sich auch aus der Matrix A bestimmen, allerdings durch kompliziertere Ausdrücke.

Wir halten das Erreichte in einem Satz fest:

Satz 7.5. Sei A eine (reelle oder komplexe) $n \times n$ -Matrix und sei $P_A(t) = \det(A - t E_n)$ das charakteristische Polynom.

Dann gilt

1. $P_A(\lambda) = 0 \iff \lambda$ ist Eigenwert von A .

2. $P_{SAS^{-1}}(t) = P_A(t)$ für invertierbare Matrizen S .

3. $P_A(t)$ hat Grad n und ist von der Form

$$P_A(t) = (-1)^n t^n + (-1)^{n-1} (\operatorname{tr} A) t^{n-1} + \alpha_{n-2} \cdot t^{n-2} + \dots + \alpha_1 \cdot t^1 + \det A. \quad (7.66)$$

Bemerkung

Da das charakteristische Polynom invariant ist unter $A \mapsto SAS^{-1}$, haben auch alle Koeffizienten diese Eigenschaft, insbesondere gilt

$$\operatorname{tr}(SAS^{-1}) = \operatorname{tr}(A). \quad (7.67)$$

Es gilt sogar die weitergehende Eigenschaft

$$\operatorname{tr}(AB) = \operatorname{tr}(BA), \quad (7.68)$$

die auch wieder für das gesamte charakteristische Polynom gilt:

$$P_{AB}(t) = P_{BA}(t) \quad (7.69)$$

Den Beweis führen wir in einer Übungsaufgabe.

Beispiel 7.2. Wie im Beispiel mit den gekoppelten Oszillatoren sei $A = \begin{pmatrix} 1+\alpha & -\alpha \\ -\alpha & 1+\alpha \end{pmatrix}$. Dann ist

$$P_A(t) = \det \begin{pmatrix} 1 + \alpha - t & -\alpha \\ -\alpha & 1 + \alpha - t \end{pmatrix} \quad (7.70)$$

$$= (1 + \alpha - t)^2 - \alpha^2 \quad (7.71)$$

$$= t^2 - 2(1 + \alpha)t + (1 + 2\alpha). \quad (7.72)$$

Wir bestimmen die Nullstellen:

$$P_A(t) = 0 \quad (7.73)$$

$$\iff (t - (1 + \alpha))^2 = \alpha^2 \quad (7.74)$$

$$\iff t = 1 + \alpha \pm \alpha, \quad (7.75)$$

also sind die Nullstellen und damit die Eigenwerte $\lambda_1 = 1$ und $\lambda_2 = 1 + 2\alpha$.

Bevor wir diese Methode der Eigenwertbestimmung weiter vertiefen, wollen wir uns noch ein paar wichtige Eigenschaften von Polynomen anschauen.

Wir hatten schon den Polynomring mit reellen Koeffizienten

$$\mathbb{R}[t] = \{a_0 + a_1 t + \dots + a_n t^n \mid a_i \in \mathbb{R}\} \quad (7.76)$$

eingeführt (analog $\mathbb{C}[t]$). Wir können Polynome addieren und multiplizieren und auch mit reellen (bzw. komplexen) Zahlen multiplizieren ($\mathbb{R}[t]$ ist also nicht nur ein Ring, sondern eine reelle Algebra). Die höchste auftretende Potenz von t in einem Polynom heißt **Grad** des Polynoms.

Ein wichtiges Konzept ist die sogenannte **Polynomdivision**, die wir hier für $\mathbb{R}[t]$ erklären, sie gilt aber analog auch für $\mathbb{C}[t]$. Zu $P, U \in \mathbb{R}[t]$ gibt es eindeutige Polynome $Q, R \in \mathbb{R}[t]$ mit

$$P = UQ + R \quad (7.77)$$

mit $\text{grad } R < \text{grad } U$.

Das funktioniert ähnlich wie beim Dividieren mit Rest bei den ganzen Zahlen: Q gibt an, wie „oft“ U in P hineinpasst, R ist der verbleibende Rest.

Der Algorithmus funktioniert wie folgt:

- Ist $n_P = \text{grad } P < \text{grad } U = n_U$, so ist $Q = 0$, $R = P$.
- Ansonsten bestimmt man zunächst den Beitrag mit der höchsten Potenz in Q :

$$Q = \underbrace{p_{n_P} \frac{1}{u_{n_U}} \cdot t^{n_P - n_U}}_{Q^{(1)}} + \dots \quad (7.78)$$

mit $P = p_{n_P} t^{n_P} + \dots$ und $U = u_{n_U} t^{n_U} + \dots$.

- Dann hat $P' = P - UQ^{(1)}$ den Grad $n_{P'} = n_P - 1$. Ist $n_{P'} < n_U$, dann ist $R = P'$, $Q = Q^{(1)}$ und man ist fertig.
- Ansonsten wendet man nun den Algorithmus auf P' an, usw.
- Da sich der Grad jedes Mal verringert, ist man irgendwann fertig.

Wie man sieht, hat Q den Grad $n_P - n_U$.

Beispiel 7.3. $P = 5t^3 + 3t^2 + 1$ und $U = t^2 - 2t$. Wir suchen Q, R mit

$$P = QU + R \quad , \quad \text{deg } R < \text{deg } U = 2. \quad (7.79)$$

Aus den Koeffizienten der höchsten Potenzen von P und U bestimmen wir

$$Q^{(1)} = 5t. \quad (7.80)$$

Dann ist

$$P' = P - Q^{(1)}U \quad (7.81)$$

$$= 5t^3 + 3t^2 + 1 - 5t(t^2 - 2t) \quad (7.82)$$

$$= 13t^2 + 1 \quad (7.83)$$

Aus den Koeffizienten der höchsten Potenzen von P' und U bestimmen wir

$$Q^{(2)} = 13. \quad (7.84)$$

Wir erhalten dann

$$P'' = P' - Q^{(2)}U \quad (7.85)$$

$$= 13t^2 + 1 - 13(t^2 - 2t) \quad (7.86)$$

$$= 26t + 1 \quad (\text{Grad} < 2) \quad (7.87)$$

$$= R \quad (7.88)$$

Also ist $Q = Q^{(1)} + Q^{(2)} = 5t + 13$ und $R = 26t + 1$.

Genauer zur Polynomdivision finden Sie im Buch von G. Fischer [F] in Abschnitt 1.3.7.

Lemma 7.6. Ist λ eine Nullstelle von $P \in \mathbb{R}[t]$ (oder $\mathbb{C}[t]$), so gibt es (genau) ein Polynom $Q \in \mathbb{R}[t]$ (bzw. $\mathbb{C}[t]$) mit

$$P(t) = (t - \lambda) \cdot Q(t). \quad (7.89)$$

Für Q gilt $\text{grad} Q = \text{grad} P - 1$ (falls $P \neq 0$).

Lemma. Wir führen die Polynomdivision von P bezüglich $(t - \lambda)$ durch,

$$P(t) = (t - \lambda) \cdot Q(t) + R(t), \quad (7.90)$$

wobei $\text{grad} R < 1$, also ist $R = r_0$ konstant. Einsetzen von λ ergibt nach der Voraussetzung

$$0 = P(\lambda) = (\lambda - \lambda) \cdot Q(\lambda) + r_0 = r_0, \quad (7.91)$$

also ist $R = 0$. □

Sobald wir eine Nullstelle λ gefunden haben, können wir also einen linearen Faktor $(t - \lambda)$ aus dem Polynom herausziehen.

Manchmal hat das verbleibende Polynom wieder λ als Nullstelle, dann können wir weitere Linearfaktoren $(t - \lambda)$ herausziehen. Die maximale Anzahl solcher Linearfaktoren nennen wir die **Vielfachheit** $\mu(\lambda, P)$ der Nullstelle λ :

$$P(t) = (t - \lambda)^{\mu(\lambda, P)} \cdot Q(t) \quad (7.92)$$

mit $Q(\lambda) \neq 0$.

Unter welchen Umständen aber hat ein Polynom überhaupt Nullstellen? Darauf gibt Antwort der

Satz 7.7 (Fundamentalsatz der Algebra). *Jedes komplexe Polynom $P \in \mathbb{C}[t]$ mit $\text{grad} P > 0$ hat eine Nullstelle.*

Diesen Satz beweisen wir hier nicht. Einen Beweis kann man später elegant im Rahmen der Funktionentheorie führen.

Durch wiederholte Anwendung des Lemmas ergibt sich sofort, dass jedes komplexe Polynom in Linearfaktoren zerfällt:

$$P(t) = a(t - \lambda_1)^{\mu(\lambda_1, P)} \cdots (t - \lambda_s)^{\mu(\lambda_s, P)} \quad (7.93)$$

mit $a, \lambda_1, \dots, \lambda_s \in \mathbb{C}$.

Für reelle Polynome gilt das nicht (z.B. hat $x^2 + 1$ keine reelle Nullstelle), aber da wir \mathbb{R} in \mathbb{C} einbetten können, definiert jedes reelle Polynom $P(t)$ auch ein Polynom in $\mathbb{C}[t]$, das über \mathbb{C} faktorisiert werden kann:

$$P(t) = a \cdot (t - \lambda_1)^{r_1} \cdots (t - \lambda_s)^{r_s} \quad (7.94)$$

mit $a \in \mathbb{R}, \lambda_1, \dots, \lambda_s \in \mathbb{C}$.

Wenn P nur reelle Koeffizienten hat, gilt

$$\overline{P(\lambda)} = P(\bar{\lambda}), \quad (7.95)$$

wenn $\lambda \in \mathbb{C}$ eine Nullstelle ist, so ist daher auch $\bar{\lambda} \in \mathbb{C}$ eine Nullstelle. Nichtreelle Nullstellen kommen also immer als komplex-konjugierte Paare; die entsprechenden Linearfaktoren kombinieren sich zu einem quadratischen Faktor

$$(t - \lambda)(t - \bar{\lambda}) = t^2 - 2 \operatorname{Re}(\lambda)t + \lambda^2 \quad (7.96)$$

mit reellen Koeffizienten. Ein reelles Polynom zerfällt also immer in Linearfaktoren (die den reellen Nullstellen entsprechen) und quadratische Faktoren (die den echt komplexen Nullstellen entsprechen).

Was können wir daraus über Eigenwerte lernen?

Zunächst einmal können wir schließen, dass *jede komplexe Matrix mindestens einen komplexen Eigenwert hat* — reelle Matrizen haben nicht unbedingt einen reellen Eigenwert, haben aber als Abbildung auf \mathbb{C}^n wenigstens einen komplexen Eigenwert.

Im Komplexen faktorisiert das charakteristische Polynom in Linearfaktoren,

$$P(t) = a \cdot (t - \lambda_1)^{r_1} \cdots (t - \lambda_s)^{r_s}. \quad (7.97)$$

Das Auftreten eines Faktors $(t - \lambda)^r$ signalisiert uns die Existenz mindestens eines Eigenvektors zum Eigenwert λ . Was sagt uns der Exponent, also die Multiplizität $\mu(\lambda, P_A)$?

Satz 7.8. Für einen Eigenwert λ einer (reellen oder komplexen) $n \times n$ -Matrix A ist

$$\mu(\lambda, P_A) \geq \dim \operatorname{Eig}(A, \lambda). \quad (7.98)$$

Mit anderen Worten: Bei höherer Multiplizität kann es mehrere unabhängige Eigenvektoren zum gleichen Eigenwert geben, muss es aber nicht.

Beweis. Sei (v_1, \dots, v_r) eine Basis von $\operatorname{Eig}(A, \lambda)$. Wir können (v_1, \dots, v_r) zu einer Basis $\mathcal{B} = (v_1, \dots, v_r, w_1, \dots, w_{n-r})$ vom gesamten Raum ergänzen. Bezüglich dieser Basis hat die zu A gehörige Abbildung $f_A : x \mapsto Ax$ eine darstellende Matrix der Form

$$B = M_{\mathcal{B}}^{\mathcal{B}}(f_A) = \begin{pmatrix} \lambda & & & \\ & \ddots & & \\ & & \lambda & \star \\ 0 & & & B' \end{pmatrix}, \quad (7.99)$$

wobei B' eine $(n-r) \times (n-r)$ -Teilmatrix bezeichnet, \star steht für beliebige Einträge in dem $r \times (n-r)$ -Block oberhalb von B' . Nach dem Kästchensatz ist das charakteristische Polynom

$$P_A(t) = P_B(t) = \det(B - t E_n) \quad (7.100)$$

$$= \det \begin{pmatrix} \lambda E_r - t E_r & \star \\ 0 & B' - t E_{n-r} \end{pmatrix} \quad (7.101)$$

$$= (\lambda - t)^r \det(B' - t E_{n-r}) \quad (7.102)$$

$$= (\lambda - t)^r P_{B'}(t), \quad (7.103)$$

d.h. $\mu(\lambda, P_A) \geq r = \dim \operatorname{Eig}(A, \lambda)$. □

Die Dimension des Eigenraums kann echt kleiner sein, wie das folgende Beispiel zeigt:

Beispiel 7.4. Sei $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, dann ist das charakteristische Polynom

$$P_A = \det \begin{pmatrix} -t & 1 \\ 0 & -t \end{pmatrix} = t^2, \quad (7.104)$$

also $\mu(0, A) = 2$. $\lambda = 0$ ist also der einzige Eigenwert und es gilt

$$\operatorname{Eig}(A, 0) = \operatorname{Ker} A = \mathbb{R} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (7.105)$$

also $\dim \operatorname{Eig}(A, 0) = 1 < 2 = \mu(0, A)$.

Die Vielfachheit ist also immer größer oder gleich der Dimension des Eigenraums. Da sich für eine diagonalisierbare Matrix die Dimensionen der Eigenräume zu n addieren müssen, die Summe der Vielfachheiten aber durch n begrenzt ist, folgt:

Satz 7.9. *Eine $n \times n$ -Matrix (reell oder komplex) ist genau dann diagonalisierbar, wenn*

1. P_A in Linearfaktoren zerfällt und
2. für alle Eigenwerte λ_i von A , $i = 1, \dots, s$, gilt

$$\mu(\lambda_i, P_A) = \dim \operatorname{Eig}(A, \lambda_i). \quad (7.106)$$

Im Komplexen ist das erste Kriterium immer erfüllt, dies alleine ist aber nicht ausreichend. Da zu einem Eigenwert λ_i der Eigenraum wenigstens eindimensional ist, folgt das Korollar:

Korollar 7.10. *Zerfällt P_A in Linearfaktoren mit Vielfachheit 1, so ist A diagonalisierbar.*

Im folgenden Abschnitt werden wir uns konkreter damit beschäftigen, wie man eine Matrix diagonalisiert, bzw. was man noch machen kann, falls sie nicht diagonalisierbar ist.

7.4 Diagonalisierung und Trigonalisierung

Matrizen A und SAS^{-1} , die durch einen Basiswechsel auseinander hervorgehen, nennt man **ähnlich**. Das charakteristische Polynom, und damit insbesondere die Spur und die Determinante, sind invariant unter Ähnlichkeitstransformationen.

Die Ähnlichkeitsbeziehung definiert eine Äquivalenzrelation auf der Menge der $n \times n$ -Matrizen. Eine Äquivalenzrelation \sim auf einer Menge M hatten wir in Definition 1.2 eingeführt; zur Erinnerung listen wir die Eigenschaften noch einmal auf. Eine Relation \sim heißt Äquivalenzrelation, wenn für alle $x, y, z \in M$ gilt:

1. $x \sim x$ (reflexiv)
2. $x \sim y \implies y \sim x$ (symmetrisch)
3. $x \sim y$ und $y \sim z$ (transitiv)

Die Ähnlichkeit von Matrizen hat genau diese Eigenschaften, z.B. 3.:

$$\text{Wenn } B = SAS^{-1}, C = TBT^{-1}, \text{ dann ist } C = TSAS^{-1}T^{-1} = (TS)A(TS)^{-1}.$$

Man kann dann Matrizen in **Äquivalenzklassen** bezüglich der Ähnlichkeitsbeziehung zusammenfassen: Die Äquivalenzklasse von A besteht dann aus allen Matrizen, die ähnlich sind zu A . Bis auf einen Basiswechsel beschreiben sie alle den gleichen Endomorphismus.

Man kann sich nun die Frage stellen, ob es in einer Äquivalenzklasse besonders einfache Vertreter (Repräsentanten) dieser Klasse gibt, sogenannte Normalformen. Anhand solcher Normalformen kann man dann auch die auftretenden Äquivalenzklassen charakterisieren.

Beispiel 7.5.

- Auf \mathbb{Z} sei eine Äquivalenzrelation wie folgt definiert:

$$n \sim m : \iff n - m \text{ ist gerade.} \quad (7.107)$$

Dann gibt es genau zwei Äquivalenzklassen, die Menge der geraden Zahlen und die Menge der ungeraden Zahlen. Einfache Repräsentanten dieser Klassen sind z.B. 0 und 1 (siehe auch Beispiel 1.6).

- Auf der Menge $M(m \times n, \mathbb{C})$ der $m \times n$ -Matrizen sei eine Äquivalenzrelation definiert durch

$$A \sim B : \iff \text{es gibt } S, T \text{ invertierbar mit } B = SAT^{-1}. \quad (7.108)$$

(A und B beschreiben also bis auf Basiswechsel (unabhängig im Definitionsbereich und Bildbereich) den gleichen Homomorphismus von \mathbb{C}^n nach \mathbb{C}^m ; man nennt solche Matrizen auch äquivalent.) Ein einfacher Repräsentant, wie wir in Ka-

pitel 4.3 gesehen haben, ist

$$\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}, \quad (7.109)$$

wobei r den Rang der Matrizen dieser Klasse angibt. Diesen Repräsentanten der Äquivalenzklasse nennen wir Normalform. Da $r \leq m, n$, ist die Zahl der Äquivalenzklassen

$$\min(m, n) + 1, \quad (7.110)$$

es gibt also nur endlich viele.

Anders als im letzten Beispiel gilt dies für die Ähnlichkeitsbeziehung nicht mehr: Es gibt unendlich viele „Ähnlichkeitsklassen“. Auch die Suche nach Normalformen gestaltet sich schwieriger, wie wir sehen werden.

Ein Typ von Matrizen, die eine einfache Normalform haben, sind die diagonalisierbaren: Für sie ist die Diagonalgestalt die Normalform. Bist auf eine Permutation der Diagonaleinträge (der Eigenwerte) ist sie eindeutig bestimmt (da die Eigenwerte beliebige reelle Zahlen sein können, erkennt man schon, dass es unendlich viele Ähnlichkeitsklassen gibt).

Bevor wir uns mit den nichtdiagonalisierbaren Matrizen beschäftigen, wollen wir uns ganz praktisch mit der Frage beschäftigen, wie man feststellt, ob eine Matrix A diagonalisierbar ist, und wenn ja, wie man die Diagonalisierung durchführt.

Diagonalisierungsverfahren

1. Eigenwerte bestimmen

- Finde die Nullstellen λ_i des charakteristischen Polynoms P_A und ihre Vielfachheit $\mu(\lambda_i, A)$. (Das ist im Allgemeinen natürlich nicht mehr exakt möglich und man muss die Nullstellen numerisch bestimmen — in diesem Fall bietet sich ein anderes Verfahren an, dass auch die Eigenvektoren in einem Näherungsverfahren konstruiert. Hier tun wir so, als könnten wir die Nullstellen exakt bestimmen.)
- Wenn P_A nicht in Linearfaktoren zerfällt, ist A nicht diagonalisierbar.
- Wenn es in Linearfaktoren zerfällt (über \mathbb{C} immer!), geht es weiter:

2. Eigenräume bestimmen

- Bestimme die Eigenräume $\text{Eig}(A, \lambda_i)$ durch Lösen der linearen Gleichungssysteme $(A - \lambda_i E_n)x = 0$.
- Gilt $\dim \text{Eig}(A, \lambda_i) = \mu(\lambda_i, A)$ für alle Nullstellen, so ist A diagonalisierbar (ansonsten nicht). In diesem Fall bestimmen wir zu jedem Eigenraum eine Basis.
- Zusammen ergeben diese n Vektoren eine Basis (v_1, \dots, v_n) des gesamten Raums \mathbb{R}^n oder \mathbb{C}^n .

3. Transformationsmatrix S bestimmen

- Wir definieren $S^{-1} = (v_1 \dots v_n)$ mit v_i als Spaltenvektoren.
- Wir bestimmen S als inverse Matrix zu S^{-1} mit dem bekannten Invertierungsverfahren.

Dann ist SAS^{-1} eine Diagonalmatrix mit den Eigenwerten als Einträgen.

Beispiel 7.6. Sei $A = \begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix}$.

$$P_A(t) = \det(A - tE_n) \quad (7.111)$$

$$= \begin{pmatrix} 1-t & 2 \\ 8 & 1-t \end{pmatrix} \quad (7.112)$$

$$= (1-t)^2 - 16 \quad (7.113)$$

$$P_A(t) = 0 \iff (1-t)^2 = 16 \quad (7.114)$$

$$\iff t = 1 \pm 4. \quad (7.115)$$

Nullstellen sind also $\lambda_1 = 5$ und $\lambda_2 = -3$. Da es $n = 2$ verschiedene Nullstellen gibt, ist A diagonalisierbar.

$Eig(A, \lambda_1)$:

$$(A - \lambda_1 E_2)x = 0 \quad (7.116)$$

$$\iff \begin{pmatrix} -4 & 2 \\ 8 & -4 \end{pmatrix} x = 0 \quad (7.117)$$

$$\implies Eig(A, \lambda_1) = \left\{ \mu \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} \mid \mu \in \mathbb{R} \right\}. \quad (7.118)$$

Der Eigenraum ist eindimensional mit Basisvektor $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

$Eig(A, \lambda_2)$:

$$(A - \lambda_2 E_2)x = 0 \quad (7.119)$$

$$\iff \begin{pmatrix} 4 & 2 \\ 8 & 4 \end{pmatrix} x = 0 \quad (7.120)$$

$$\implies Eig(A, \lambda_2) = \left\{ \mu \cdot \begin{pmatrix} 1 \\ -2 \end{pmatrix} \mid \mu \in \mathbb{R} \right\}. \quad (7.121)$$

Der Eigenraum ist eindimensional mit Basisvektor $v_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

Definiere nun

$$S^{-1} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \quad (7.122)$$

$$S = \frac{1}{-4} \cdot \begin{pmatrix} -2 & -1 \\ -2 & 1 \end{pmatrix} = \frac{1}{4} \cdot \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix}. \quad (7.123)$$

Dann ist

$$SAS^{-1} = \frac{1}{4} \cdot \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \quad (7.124)$$

$$= \frac{1}{4} \cdot \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 5 & -3 \\ 10 & 6 \end{pmatrix} \quad (7.125)$$

$$= \frac{1}{4} \cdot \begin{pmatrix} 20 & 0 \\ 0 & -12 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & -3 \end{pmatrix} \quad (7.126)$$

Bemerkung

Manchmal sieht man in den Büchern die Diagonalisierung mit einer Matrix T aus Eigenvektoren, sodass dann $T^{-1}AT$ diagonal wird, d.h. statt die inverse Matrix S^{-1} aus den Eigenvektoren definiert, gibt man der Matrix einen eigenen Namen (hier: T) und muss dann nur aufpassen, dass dann bei der Anwendung auf A die Matrix T rechts steht und T^{-1} links. Wierum man es macht, ist egal, und man kann sich leicht merken, welche Form dann richtig für die Diagonalisierung ist: Die Matrix aus Eigenvektoren steht immer rechts von A : A „möchte“ auf ihre Eigenvektoren wirken.

Manchmal taucht die Frage auf, ob zwei Endomorphismen $F, G : V \rightarrow V$ simultan diagonalisierbar sind, ob es also eine Basis gibt, bezüglich der beide durch Diagonalmatrizen dargestellt werden. Das ist offensichtlich dann der Fall, wenn es eine Basis gibt, in der jeder Vektor sowohl Eigenvektor von F als auch von G ist. Es gilt folgender Satz:

Satz 7.11. *Zwei diagonalisierbare Endomorphismen $F, G : V \rightarrow V$ ($\dim V < \infty$) sind genau dann simultan diagonalisierbar, wenn*

$$F \circ G = G \circ F. \quad (7.127)$$

Beweis. Wir zeigen zunächst, dass aus der simultanen Diagonalisierbarkeit (7.127) folgt. Sei $\mathcal{A} = (v_1, \dots, v_n)$ eine Basis aus Eigenvektoren von F und G mit Eigenwerten $\lambda_1, \dots, \lambda_n$ und μ_1, \dots, μ_n . Für $i \in \{1, \dots, n\}$ ist

$$F \circ G(v_i) = F(G(v_i)) = F(\mu_i \cdot v_i) = \lambda_i \cdot \mu_i \cdot v_i = \dots = G \circ F(v_i). \quad (7.128)$$

Da $F \circ G$ mit $G \circ F$ auf jedem Basisvektor übereinstimmt, gilt

$$F \circ G = G \circ F. \quad (7.129)$$

Sei andersherum $F \circ G = G \circ F$ vorausgesetzt. Seien $\lambda_1, \dots, \lambda_r$ die paarweise verschiedenen Eigenwerte von F , μ_1, \dots, μ_s die paarweise verschiedenen Eigenwerte von G .

Sei $v \in \text{Eig}(G, \mu_j)$. Dann ist

$$G(F(v)) = F(G(v)) = F(\mu_j \cdot v) = \mu_j \cdot F(v), \quad (7.130)$$

d.h. $F(v) \in \text{Eig}(G, \mu_j)$, also

$$F(\text{Eig}(G, \mu_j)) \subset \text{Eig}(G, \mu_j). \quad (7.131)$$

F lässt also die Eigenräume von G invariant (und andersherum). Wären alle Eigenräume eindimensional, wären wir fertig (für $v \in \text{Eig}(G, \mu_j)$ wäre $F(v)$ ein Vielfaches von v , also v auch Eigenvektor von F).

Wenn die Eigenräume nicht eindimensional sind, müssen wir mehr arbeiten. Wir wissen, dass wir jeden Vektor in V als Summe von Eigenvektoren von F und ebenso als Summe von Eigenvektoren von G schreiben können. Wir zeigen zunächst, dass sich jeder Eigenvektor von F als Summe von Vektoren schreiben lässt, die sowohl Eigenvektoren von F als auch Eigenvektoren von G sind.

Sei also $F(v) = \lambda \cdot v$ und $v = v_1 + \dots + v_s$ die Zerlegung von v in Eigenvektoren von G ($v_j \in \text{Eig}(G, \mu_j)$). Wir wollen jetzt zeigen, dass $F(v_j) = \lambda \cdot v_j$ gilt. Dazu setzen wir an

$$F(v) = \lambda \cdot v = \lambda \cdot v_1 + \dots + \lambda \cdot v_s \quad (7.132)$$

und andererseits

$$F(v) = F(v_1) + \dots + F(v_s). \quad (7.133)$$

Da F die Eigenräume von G invariant lässt, ist $F(v_j) \in \text{Eig}(G, \mu_j)$. Die Zerlegung eines Vektors in Eigenvektoren von G ist eindeutig, daher folgt $F(v_j) = \lambda \cdot v_j$ für $j = 1, \dots, s$.

Jeder Eigenvektor von F lässt sich also als Summe gemeinsamer Eigenvektoren von F und G schreiben. Da sich jeder Vektor in V aus Eigenvektoren von F zusammensetzen lässt, lässt sich jeder Vektor in V als Summe gemeinsamer Eigenvektoren schreiben. Somit gibt es eine Zerlegung

$$V = \bigoplus_{i,j} \text{Eig}(F, \lambda_i) \cap \text{Eig}(G, \mu_j). \quad (7.134)$$

Wählt man in jedem dieser Summanden eine Basis, lässt sich aus ihnen eine Basis von V aus gemeinsamen Eigenvektoren gewinnen. \square

Die Aussage gilt natürlich entsprechend für Matrizen:

Korollar 7.12. *Zwei diagonalisierbare Matrizen A und B sind genau dann simultan diagonalisierbar, wenn $AB = BA$.*

Simultan diagonalisierbar bedeutet dabei, dass es eine invertierbare Matrix S gibt, sodass sowohl SAS^{-1} als auch SBS^{-1} Diagonalmatrizen sind.

Wir wenden uns jetzt dem Fall zu, dass eine Matrix nicht diagonalisierbar ist.

Wenn eine Matrix nicht diagonalisierbar ist, kann man wenigstens versuchen, sie auf Dreiecksgestalt zu bringen.

Definition 7.5. Eine $n \times n$ -Matrix A heißt trigonalisierbar genau dann, wenn es eine invertierbare Matrix S gibt, sodass SAS^{-1} eine obere Dreiecksmatrix ist.

Ein Endomorphismus $F : V \rightarrow V$ ($\dim V < \infty$) heißt trigonalisierbar genau dann, wenn es eine Basis \mathcal{A} von V gibt, sodass $M_{\mathcal{A}}^{\mathcal{A}}(F)$ eine obere Dreiecksmatrix ist.

Die Diagonaleinträge einer oberen Dreiecksmatrix sind die Nullstellen des charakteristischen Polynoms, denn

$$\det \left(\begin{pmatrix} a_{11} & & \star \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} - t E_n \right) = (a_{11} - t) \cdot \dots \cdot (a_{nn} - t). \quad (7.135)$$

d.h. eine Matrix kann höchstens dann auf Dreiecksgestalt gebracht werden, wenn das charakteristische Polynom in Linearfaktoren zerfällt.

Das ist sogar ausreichend:

Satz 7.13. Sei $F : V \rightarrow V$ ein Endomorphismus, $1 \leq \dim V < \infty$. Dann gilt:

$$F \text{ ist trigonalisierbar} \iff P_F \text{ zerfällt in Linearfaktoren} \quad (7.136)$$

Korollar 7.14. Jede komplexe $n \times n$ -Matrix ist trigonalisierbar.

Satz. Wir beweisen den Satz durch vollständige Induktion über n . Offensichtlich ist die Aussage richtig für $n = 1$.

Sei nun $n \geq 1$. Wir setzen voraus, dass die Aussage für $\dim V = n$ richtig ist, und beweisen, dass sie dann auch für $\dim V = n + 1$ gilt.

Die Aussage ist äquivalent dazu, dass die darstellende Matrix $A = M_{\mathcal{A}}^{\mathcal{A}}(F)$ bezüglich einer Basis \mathcal{A} durch eine Ähnlichkeitstransformation in eine obere Dreiecksmatrix SAS^{-1} transformiert werden kann.

Sei $F : V \rightarrow V$ ein Endomorphismus, $\dim V = n + 1$, sodass $P_F(t)$ in Linearfaktoren zerfällt. Dann gibt es einen Eigenvektor $w_1 \neq 0$ mit $F(w_1) = \lambda \cdot w_1$. Wir wählen nun

eine Basis $\mathcal{A} = (w_1, v_1, \dots, v_n)$ von V . Bezüglich \mathcal{A} hat F die Darstellung

$$A = M_{\mathcal{A}}^{\mathcal{A}}(F) = \begin{pmatrix} \lambda & \star \\ 0 & \\ \vdots & A_2 \\ 0 & \end{pmatrix} \quad (7.137)$$

und es gilt

$$P_F(t) = P_A(t) = \det(A - tE_n) = (\lambda - t) \det(A_2 - tE_{n-1}) = (\lambda - t) P_{A_2}(t). \quad (7.138)$$

Wenn P_F in Linearfaktoren zerfällt, zerfällt auch P_{A_2} in Linearfaktoren. A_2 ist eine $n \times n$ -Matrix, deren Polynom zerfällt, also lässt sich nach Induktionsvoraussetzung A_2 trigonalisieren: Es gibt also eine invertierbare Matrix S , sodass SA_2S^{-1} eine obere Dreiecksmatrix ist. Dann ist $(\star, \star', \star''$ stehen dabei für beliebige Einträge)

$$\begin{pmatrix} 1 & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \lambda & \star \\ 0 & A_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & S^{-1} \end{pmatrix} = \begin{pmatrix} \lambda & \star' \\ 0 & SA_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & S^{-1} \end{pmatrix} = \begin{pmatrix} \lambda & \star'' \\ 0 & SA_2S^{-1} \end{pmatrix} \quad (7.139)$$

eine obere Dreiecksmatrix. □

Reelle Matrizen, die nicht trigonalisierbar sind, können übrigens auf eine „Beinahe-Dreiecksmatrix“ gebracht werden, in der nur noch in der ersten unteren Nebendiagonale nichttriviale Einträge verbleiben (siehe 4.5.4 im Buch von G. Fischer [F]).

Was ist für trigonalisierbare Matrizen eine adäquate Normalform? Man kann zeigen, dass die Dreiecksgestalt noch deutlich vereinfacht werden kann in die **Jordansche Normalform**. Ihre Bausteine sind sogenannte Jordan-Matrizen von der Form

$$J = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ & \lambda & 1 & \ddots & 0 \\ & & \lambda & \ddots & 0 \\ & & & \ddots & 1 \\ 0 & & & & \lambda \end{pmatrix}. \quad (7.140)$$

Satz 7.15 (Existenz der Jordanschen Normalform). *Sei A eine trigonalisierbare $n \times n$ -Matrix. Dann gibt es eine invertierbare Matrix S , sodass*

$$SAS^{-1} = \begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_r \end{pmatrix} \quad (7.141)$$

mit Jordan-Matrizen J_1, \dots, J_r . Die Jordan-Matrizen sind bis auf ihre Reihenfolge

eindeutig bestimmt.

Bemerkung

Es können zu einem Eigenwert λ mehrere Jordanblöcke auftreten. Die Gesamtzahl der Jordanblöcke ist durch die maximale Anzahl linear unabhängiger Eigenvektoren gegeben. Man kann sich leicht überlegen, dass eine Jordan-Matrix mit Diagonaleintrag λ bis auf Skalierung genau einen Eigenvektor hat,

$$\begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ & \lambda & 1 & \ddots & 0 \\ & & \lambda & \ddots & 0 \\ & & & \ddots & 1 \\ 0 & & & & \lambda \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} = \lambda \cdot \begin{pmatrix} 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix}. \quad (7.142)$$

Entsprechend hat eine Matrix mit r Jordanblöcken r linear unabhängige Eigenvektoren (und nicht mehr).

Da der Beweis dieses Satzes etwas aufwendig ist, werden wir ihn hier nicht führen. Einen Beweis finden Sie im Buch von G. Fischer [F] in Abschnitt 4.6.

7.5 Funktionen von Matrizen und Anwendungen

Endomorphismen können wir mehrmals hintereinander ausführen. Wir schreiben dafür dann

$$F^2 = F \circ F$$

$$F^3 = F \circ F \circ F$$

usw.

Für $n \times n$ -Matrizen funktioniert natürlich das Gleiche, wir können Potenzen einer Matrix bilden:

$$A^2 = A \cdot A$$

$$A^3 = A \cdot A \cdot A$$

usw.

Da wir Endomorphismen und Matrizen auch addieren und sie mit reellen bzw. komplexen Zahlen multiplizieren können, können wir auch polynomiale Ausdrücke aus ihnen bilden, z.B. $A^2 + 3A^3$. Wir können sie also in Polynome einsetzen, d.h. wenn

$$p(t) = a_0 + a_1 t + \dots + a_m t^m \quad (7.143)$$

ein Polynom ist, so erklären wir

$$p(A) := a_0 E_n + a_1 A + \dots + a_m A^m, \quad (7.144)$$

wobei der konstante Term mit der Einheitsmatrix versehen wird: Wir setzen sozusagen $A^0 = E_n$ für eine $n \times n$ -Matrix A . Das Einsetzen von A ist dabei verträglich mit der Multiplikation von Polynomen:

Wenn $p(t) = p_1(t) p_2(t)$, dann ist

$$p(A) = p_1(A) p_2(A). \quad (7.145)$$

Da $p_1(t) p_2(t) = p_2(t) p_1(t)$, gilt insbesondere auch

$$p_1(A) p_2(A) = p_2(A) p_1(A). \quad (7.146)$$

Matrizen, die aus polynomialen Ausdrücken in der gleichen Matrix A entstanden sind, vertauschen miteinander!

Das Gleiche funktioniert natürlich auch mit Endomorphismen $F : V \rightarrow V$, wobei wir $F^0 := \text{id}_V$ definieren, d.h. zu einem Polynom $p(t)$ können wir einen Endomorphismus $p(F) : V \rightarrow V$ bilden.

Bemerkung

Für einen fest gewählten Endomorphismus $F : V \rightarrow V$ definiert dies eine Abbildung

$$\begin{aligned} E_F : \mathbb{R}[t] &\rightarrow \text{End}(V) \\ p(t) &\mapsto p(F), \end{aligned}$$

die mit Addition und Multiplikation verträglich ist (also einen Ringhomomorphismus — sogar einen Algebromorphismus, da die Abbildung auch mit der skalaren Multiplikation verträglich ist).

Man kann sich fragen, ob es auch andere Funktionen gibt, in die man Matrizen oder Endomorphismen einsetzen kann, z.B. $\cos A$ oder e^A . Damit werden wir uns in Kürze beschäftigen.

Zunächst aber wollen wir uns mit Folgendem befassen: Der Vektorraum $M(n \times n, \mathbb{R})$ (entsprechende Aussagen gelten für komplexwertige Matrizen) der $n \times n$ -Matrizen hat Dimension n^2 (Matrizen können als n^2 -Tupel aufgefasst werden). Das heißt, dass nicht alle Potenzen E_n, A, A^2, \dots linear unabhängig sein können. Insbesondere liegen sie alle im Untervektorraum der mit A kommutierenden Matrizen, und wir erwarten, dass dessen Dimension deutlich kleiner als n^2 ist.

Es wird also für jede Matrix A einen Exponenten m_A geben, sodass A^{m_A} durch eine Linearkombination der kleineren Potenzen von A ausgedrückt werden kann, aber $(E_n, A, \dots, A^{m_A-1})$ linear unabhängig ist. Das entsprechende Polynom $t^{m_A} + a_{m_A-1}t^{m_A-1} + \dots + a_0 = M_A(t)$ mit $M_A(A) = 0$, das diese lineare Abhängigkeit ausdrückt, heißt **Minimalpolynom**.

Die Zahl m_A und das entsprechende Polynom werden von der Matrix A abhängen. Es gilt aber allgemein $m_A \leq n$ und man kann immer ein entsprechendes Polynom n -ter Ordnung angeben, das bei Einsetzen der Matrix verschwindet: das charakteristische Polynom!

Satz 7.16 (Cayley-Hamilton). *Sei $P_F(t)$ das charakteristische Polynom eines Endomorphismus $F : V \rightarrow V$ ($\dim V = n < \infty$). Dann gilt*

$$P_F(F) = 0. \tag{7.147}$$

Vor dem Beweis wollen wir uns diesen Satz für Diagonalmatrizen klar machen. Wir hatten einmal in den Übungen gesehen, dass eine typische Diagonalmatrix (also eine mit paarweise verschiedenen Diagonaleinträgen) nur mit anderen Diagonalmatrizen vertauscht. Der Raum der Diagonalmatrizen ist n -dimensional, d.h. wir erwarten, dass es ein Polynom n -ter Ordnung gibt, das — die Matrix eingesetzt — verschwindet.

Das charakteristische Polynom für eine Diagonalmatrix D mit (paarweise verschiedenen) Einträgen $\lambda_1, \dots, \lambda_n$ ist

$$P_D(t) = (\lambda_1 - t) \cdots (\lambda_n - t). \tag{7.148}$$

Setzen wir D ein, so erhalten wir

$$P_D(D) = (\lambda_1 E_n - D) \cdots (\lambda_n E_n - D), \quad (7.149)$$

also ein Produkt von Diagonalmatrizen, wobei der i -te Faktor den i -ten Eintrag Null besitzt. Also ist $P_D(D) = 0$.

Das gilt dann auch für beliebige diagonalisierbare Matrizen und Endomorphismen, denn für $B = S A S^{-1}$ gilt $p(B) = S p(A) S^{-1}$ für ein beliebiges Polynom $p(t)$. Ist also $P_B(B) = 0$, so ist

$$P_A(A) = P_B(A) = S^{-1} P_B(B) S = 0. \quad (7.150)$$

Beweis vom Satz von Cayley-Hamilton. Sei $v \in V$ beliebig. Wir zeigen, dass $P_F(F)(v) = 0$.

Man betrachte den Untervektorraum U , der von $v, F(v), F^2(v), \dots$ aufgespannt wird mit Dimension $\dim U = r \leq n$. Dann ist $(v, F(v), \dots, F^{r-1}(v))$ linear unabhängig aber $F^r(v)$ wird aus den anderen linear erzeugt, d.h. es gibt Koeffizienten $\alpha_0, \dots, \alpha_{r-1}$ mit

$$F^r(v) = \alpha_0 \cdot v + \alpha_1 \cdot F(v) + \dots + \alpha_{r-1} \cdot F^{r-1}(v). \quad (7.151)$$

Alle höheren Terme $F^{r+1}(v), \dots$ werden dann auch aus den ersten r Termen erzeugt, da z.B.

$$F^{r+1}(v) = F(F^r(v)) = F(\alpha_0 \cdot v + \dots + \alpha_{r-1} \cdot F^{r-1}(v)) \quad (7.152)$$

$$= \alpha_0 \cdot F(v) + \dots + \alpha_{r-2} \cdot F^{r-1}(v) + \alpha_{r-1} \cdot F^r(v) \quad (7.153)$$

$$= \alpha_0 \cdot F(v) + \dots + \alpha_{r-2} \cdot F^{r-1}(v) + \alpha_{r-1} \cdot (\alpha_0 \cdot v + \dots + \alpha_{r-1} \cdot F^{r-1}(v)). \quad (7.154)$$

Wir ergänzen $(v, F(v), \dots, F^{r-1}(v))$ zu einer Basis \mathcal{A} von V . Bezüglich \mathcal{A} hat F die Matrixdarstellung

$$M_{\mathcal{A}}^{\mathcal{A}}(F) = \begin{pmatrix} A' & \star \\ 0 & A'' \end{pmatrix} \quad (7.155)$$

mit

$$A' = \begin{pmatrix} 0 & 0 & \dots & 0 & \alpha_0 \\ 1 & 0 & & & \alpha_1 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & \alpha_{r-1} \end{pmatrix}. \quad (7.156)$$

Nach dem Kästchensatz ist $P_F(t) = P_{A'}(t) P_{A''}(t)$. Es gilt (siehe Übung):

$$P_{A'}(t) = (-1)^{r-1} \cdot (\alpha_0 + \alpha_1 \cdot t + \dots + \alpha_{r-1} \cdot t^{r-1} - t^r). \quad (7.157)$$

Dann ist

$$P_{A'}(F)(v) = (-1)^{r-1} \cdot (\alpha_0 \cdot v + \alpha_1 \cdot F(v) + \dots + \alpha_{r-1} \cdot F^{r-1}(v) - F^r(v)) = 0, \quad (7.158)$$

also auch

$$P_F(F)(v) = P_{A''}(F) \circ P_{A'}(F)(v) = 0. \quad (7.159)$$

□

Wir haben uns jetzt Polynome von Matrizen bzw. Endomorphismen angesehen und haben mit dem Satz von Cayley-Hamilton gesehen, dass sich die n -te Potenz von einer $n \times n$ -Matrix A (und damit alle höheren Potenzen) durch kleinere Potenzen von A darstellen lässt.

Der nächste Schritt zu allgemeineren Funktionen von Matrizen ist die Betrachtung von Potenzreihen.

Sei $f(z) = \sum_{n=0}^{\infty} a_n z^n$ eine Potenzreihe mit Konvergenzradius R . Können wir dann $f(A)$ bilden?

Wenn wir unendlich viele Matrizen aufsummieren, müssen wir uns Gedanken über die Konvergenz machen. Wenn wir die Folgen der Partialsummen, $\sum_{n=0}^N a_n A^n$, betrachten, verlangen wir, dass die sich in jedem der n^2 Einträge ergebende Folge konvergiert.

Statt jeden dieser Einträge einzeln zu untersuchen, wäre es gut, eine geeignete Abschätzung zu haben, eine Art Abstandsbegriff, damit wir diskutieren können, wie „weit“ zwei Matrizen voneinander entfernt sind. Dann kann man auch leicht diskutieren, ob sich eine Folge von Matrizen einer bestimmten Matrix annähert, also gegen diese Matrix konvergiert.

Wir wollen nun einen Abstandsbegriff einführen, der über ein Skalarprodukt definiert ist. Dabei stellen wir uns eine $n \times n$ -Matrix $A = (v_1 \cdots v_n)$ als einen Vektor mit n^2 Komponenten vor, und bilden das gewöhnliche Skalarprodukt im \mathbb{R}^{n^2} oder \mathbb{C}^{n^2} . Wir werden im Folgenden komplexe Matrizen betrachten, alle Aussagen über reelle Matrizen ergeben sich daraus.

Definition 7.6. Das **Hilbert-Schmidt-Skalarprodukt** auf $M(n \times n, \mathbb{C})$ ist definiert als

$$\langle A, B \rangle_{HS} := \operatorname{tr}(A^* B). \quad (7.160)$$

Schreiben wir $A = (v_1 \cdots v_n)$ und $B = (w_1 \cdots w_n)$, so ist

$$\operatorname{tr}(A^* B) = \operatorname{tr}((\bar{v}_i^t w_j)) = \operatorname{tr}(\langle v_i, w_j \rangle_C) = \langle v_1, w_1 \rangle_C + \dots + \langle v_n, w_n \rangle_C, \quad (7.161)$$

also das kanonische Skalarprodukt im \mathbb{C}^{n^2} .

Mithilfe des Skalarprodukts können wir eine entsprechende Länge definieren:

Definition 7.7. Die **Hilbert-Schmidt-Norm** (auch **Frobenius-Norm**) auf $M(n \times$

n, \mathbb{C}) ist definiert als

$$\|A\|_{HS} := \sqrt{\operatorname{tr}(A^*A)}. \quad (7.162)$$

Das Hilbert-Schmidt-Skalarprodukt erfüllt die gleichen Eigenschaften, die wir schon vom kanonischen Skalarprodukt gewohnt sind:

Satz 7.17. *Das Hilbert-Schmidt-Skalarprodukt hat die folgenden Eigenschaften (wobei $A, B, C \in M(n \times n, \mathbb{C})$ beliebige Matrizen sind und $\lambda \in \mathbb{C}$ eine beliebige komplexe Zahl ist):*

1. es ist sesquilinear (für reelle Matrizen ist das Skalarprodukt bilinear):

$$\langle A + B, C \rangle_{HS} = \langle A, C \rangle_{HS} + \langle B, C \rangle_{HS} \quad (7.163a)$$

$$\langle \lambda \cdot A, B \rangle_{HS} = \bar{\lambda} \cdot \langle A, B \rangle_{HS} \quad (7.163b)$$

$$\langle A, B + C \rangle_{HS} = \langle A, B \rangle_{HS} + \langle A, C \rangle_{HS} \quad (7.163c)$$

$$\langle A, \lambda \cdot B \rangle_{HS} = \lambda \cdot \langle A, B \rangle_{HS}, \quad (7.163d)$$

2. es ist hermitesch (für reelle Matrizen: symmetrisch):

$$\langle A, B \rangle_{HS} = \overline{\langle B, A \rangle_{HS}}, \quad (7.164)$$

3. es ist positiv definit: Für alle $A \neq 0$ ist

$$\langle A, A \rangle_{HS} > 0. \quad (7.165)$$

4. Es gilt die Cauchy-Schwarzsche Ungleichung:

$$|\langle A, B \rangle_{HS}| \leq \|A\|_{HS} \|B\|_{HS}. \quad (7.166)$$

5. Die Hilbert-Schmidt-Norm ist submultiplikativ:

$$\|AB\|_{HS} \leq \|A\|_{HS} \cdot \|B\|_{HS}. \quad (7.167)$$

6. Die Norm ist selbstadjungiert:

$$\|A^*\|_{HS} = \|A\|_{HS}. \quad (7.168)$$

Die beiden letzten Eigenschaften sind für uns neu, sie haben kein Analogon für Vektoren. Die Eigenschaften 1.–4. folgen automatisch aus den Eigenschaften des Skalarprodukts auf \mathbb{C}^{n^2} . Eigenschaft 6. ergibt sich aus der Beobachtung

$$\|A\|_{HS}^2 = \operatorname{tr}(A^*A) = \sum_{i,k} \bar{a}_{ik} a_{ik} = \sum_{i,k} |a_{ik}|^2, \quad (7.169)$$

$\|A\|_{HS}^2$ ist also als Summe der Betragsquadrate aller Einträge gegeben und stimmt daher mit $\|A^*\|_{HS}^2$ überein.

Die zusätzliche Eigenschaft 5. macht $\|\cdot\|_{HS}$ zu einer sogenannten **Algebranorm**. Der Beweis von 5. geht wie folgt:

Beweis der Eigenschaft 5. Seien $A = (v_1 \cdots v_n)$ und $B = (w_1 \cdots w_n)$. Dann erfüllt der Koeffizient von A^*B in der i -ten Zeile und k -ten Spalte

$$|(A^*B)_{ik}| = |v_i^* w_k| = |\langle v_i, w_k \rangle| \leq \|v_i\| \cdot \|w_k\|, \quad (7.170)$$

also

$$\operatorname{tr}((A^*B)^*(A^*B)) = \sum_{i,k} |(A^*B)_{ik}|^2 \leq \|v_i\|^2 \|w_k\|^2 \leq \|A\|_{HS}^2 \|B\|_{HS}^2. \quad (7.171)$$

Daraus folgt

$$\|A^*B\|_{HS} \leq \|A\|_{HS} \|B\|_{HS} = \|A^*\|_{HS} \|B\|_{HS} \quad (7.172)$$

für beliebige Matrizen A, B und damit die Eigenschaft 5. □

Damit können wir jetzt Potenzreihen abschätzen:

Satz 7.18. Sei $f(z) = \sum_{m=0}^{\infty} a_m z^m$ eine komplexe Potenzreihe mit Konvergenzradius R . Für $A \in M(n \times n, \mathbb{C})$ mit $\|A\|_{HS} < R$ ist die Folge

$$\left(\sum_{m=0}^M a_m A^m \right)_M \quad (7.173)$$

in jedem Eintrag konvergent, die Matrix der entsprechenden Grenzwerte nennen wir $f(A)$.

Beweis. Für jeden Summanden $C_m = a_m A^m$ der Reihe gilt ($i, j \in \{1, \dots, n\}$)

$$|(C_m)_{ij}| \leq \|C_m\|_{HS} = |a_m| \cdot \|A^m\|_{HS} \leq |a_m| \cdot \|A\|_{HS}^m. \quad (7.174)$$

Für $\|A\|_{HS} < R$ konvergiert $\sum_{m=0}^{\infty} |a_m| \cdot \|A\|_{HS}^m$, also hat die Reihe $\sum_{m=0}^{\infty} |(C_m)_{ij}|$ eine konvergente Majorante. Daher konvergiert für jeden Eintrag $\sum_{m=0}^{\infty} (C_m)_{ij}$ absolut, und wir können eine Matrix B von Grenzwerten bilden,

$$B_{ij} = \sum_{m=0}^{\infty} (C_m)_{ij} = \sum_{m=0}^{\infty} a_m \cdot (A^m)_{ij}. \quad (7.175)$$

□

Beispiel 7.7. $e^A = \exp(A)$ ist für jede $n \times n$ -Matrix A definiert.

Bemerkung

Die Aussage gilt für jede Algebrannorm auf $M(n \times n, \mathbb{C})$, insbesondere für die sogenannte **Operatornorm** (oder *Spektralnorm*)

$$\|A\|_{ON} := \max_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}. \quad (7.176)$$

Da für alle Matrizen A gilt $\|A\|_{ON} \leq \|A\|_{HS}$, liefert sie für Potenzreihen eine bessere Abschätzung als die Hilbert-Schmidt-Norm.

Nur zu wissen, dass eine Potenzreihe $\sum a_m A^m$ konvergiert, genügt meist nicht, um sie zu berechnen. Für diagonalisierbare Matrizen ist dies aber einfach: Sei $D = SAS^{-1}$ eine Diagonalmatrix. Es gilt dann für alle Partialsummen

$$\sum_{m=0}^M a_m A^m = S^{-1} \left(\sum_{m=0}^M a_m D^m \right) S. \quad (7.177)$$

Wenn also $\sum_{m=0}^M a_m D^m$ konvergiert, konvergiert auch $\sum_{m=0}^M a_m A^m$ und umgekehrt.

Für Diagonalmatrizen können wir die Potenzreihe aber leicht ausrechnen:

$$f(D) = \sum_{m=0}^{\infty} a_m D^m = \sum_{m=0}^{\infty} a_m \begin{pmatrix} \lambda_1^m & & \\ & \ddots & \\ & & \lambda_n^m \end{pmatrix} = \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix}. \quad (7.178)$$

Für diagonalisierbare Matrizen $A = S^{-1} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} S$ können wir also $f(A)$ erklären, wenn $|\lambda_i| < R$ für alle Eigenwerte λ_i gilt, als

$$f(A) = S^{-1} \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix} S. \quad (7.179)$$

Das eröffnet auch die Möglichkeit, Funktionen, die nicht durch Potenzreihen darstellbar sind, für diagonalisierbare Matrizen zu erklären.

Beispiel 7.8. Sei $q : \mathbb{R} \rightarrow \mathbb{C}^n$, $t \mapsto q(t)$ eine vektorwertige Funktion auf \mathbb{R} , welche die Differentialgleichung

$$\dot{q}(t) = Aq(t) \quad (7.180)$$

mit einer Matrix $A \in M(n \times n, \mathbb{C})$ erfüllt. Dann ist die allgemeine Lösung

$$q(t) = e^{At} q_0 \quad (7.181)$$

mit einem konstanten Vektor $q_0 = q(0)$, der die Anfangsbedingungen beschreibt.

Für eine diagonalisierbare Matrix $A = S^{-1}DS$ sieht man sofort, dass dies die allgemeine Lösung ist, denn dann wird die Gleichung in neuen Koordinaten $q' = Sq$ zu

$$\dot{q}' = SAS^{-1}q' = Dq', \quad (7.182)$$

was n entkoppelte Differentialgleichungen $\dot{q}'_i(t) = \lambda_i q'_i$ ergibt. Entsprechend ist die Lösung

$$q'(t) = e^{Dt} q'_0 \quad (7.183)$$

und damit

$$q(t) = S^{-1}q'(t) = S^{-1}e^{Dt}S \underbrace{S^{-1}q'_0}_{q_0} = e^{At} q_0. \quad (7.184)$$

Beispiel 7.9. In der Physik trifft man oft auf Gleichungen der Form

$$\ddot{q} = Aq, \quad (7.185)$$

z.B. bei Stabilitätsanalysen, wo q die kleine Störung aus einer statischen Konfiguration beschreibt.

Ist A diagonalisierbar, $SAS^{-1} = D = \begin{pmatrix} \lambda_1 & & \\ & \dots & \\ & & \lambda_n \end{pmatrix}$, so wird die Gleichung in Koordinaten $q' = Sq$ zu

$$\ddot{q}' = \begin{pmatrix} \lambda_1 & & \\ & \dots & \\ & & \lambda_n \end{pmatrix} q'. \quad (7.186)$$

Für negative reelle λ_i sind alle n Gleichungen ganz gewöhnliche Schwingungsgleichungen, die durch Sinus-/ Kosinusfunktionen gelöst werden: Das System oszilliert also um seine Ruhelage.

Treten aber andere von Null verschiedene Eigenwerte auf, so gibt es Lösungen, die exponentiell anwachsen: Das System ist dann nicht stabil. Zum Beispiel:

$\ddot{q} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} q$ hat die oszillierenden Lösungen $\begin{pmatrix} 0 \\ \cos t \end{pmatrix}$ und $\begin{pmatrix} 0 \\ \sin t \end{pmatrix}$, aber auch die exponentiell anwachsenden bzw. unterdrückten Lösungen $\begin{pmatrix} e^t \\ 0 \end{pmatrix}$ und $\begin{pmatrix} e^{-t} \\ 0 \end{pmatrix}$.

Zusammenfassung Kapitel 7

- Wir haben die Begriffe Eigenwert, Eigenvektor und Eigenraum kennengelernt.
- Existiert eine Basis aus Eigenvektoren, so kann ein Endomorphismus diagonalisiert werden.
- Zum Aufspüren der Eigenwerte suchen wir Nullstellen des charakteristischen Polynoms.
- Haben wir die Eigenwerte von A gefunden, bestimmen wir Basen der Eigenräume. Stimmen die Dimensionen der Eigenräume mit den Vielfachheiten der Nullstellen im charakteristischen Polynom überein, dann ist die Matrix diagonalisierbar: SAS^{-1} ist diagonal, mit $S^{-1} = (v_1 \dots v_n)$ bestehend aus der Basis aus Eigenvektoren.
- Wenn eine Matrix nicht diagonalisierbar ist, so ist sie wenigstens trigonalisierbar (und kann auf Jordan-Normalform gebracht werden), wenn das charakteristische Polynom in Linearfaktoren zerfällt (in \mathbb{C} funktioniert das immer).
- Wir können unter bestimmten Umständen Matrizen in Funktionen einsetzen:
 - Polynome
 - Potenzreihen (wenn $\|A\|_{HS} < \text{Konvergenzradius } R$ bzw. wenn $|\lambda_i| < R$ für diagonalisierbare A)
 - andere Funktionen für diagonalisierbare Matrizen als

$$f(A) = S^{-1} \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix} S. \quad (7.187)$$

8 Euklidische und unitäre Vektorräume

8.1 Skalarprodukt

In Kapitel 2 haben wir das Skalarprodukt auf \mathbb{R}^n eingeführt als Objekt, das uns Auskunft über Längen und Winkel gibt. Dieses Produkt

$$\langle x, y \rangle = x^t y \quad (8.1)$$

ist mit Hilfe der Koordinaten x_i und y_j bezüglich der kanonischen Einheitsvektoren definiert.

Nun können wir den Raum \mathbb{R}^n auch völlig anders parametrisieren mit neuen Koordinaten

$$x' = Sx. \quad (8.2)$$

Die geometrischen Strukturen (Längen, Winkel) sollen sich unter einer solchen Reparametrisierung nicht ändern, aber ausgedrückt durch die neuen Koordinaten werden sie eine andere Form bekommen. Da $x = S^{-1}x'$, finden wir

$$x^t y = (S^{-1}x')^t (S^{-1}y') \quad (8.3)$$

$$= x'^t ((S^{-1})^t S^{-1})y' \quad (8.4)$$

$$= x'^t A y'. \quad (8.5)$$

Bezüglich der neuen Basis berechnen wir das Skalarprodukt also nicht mehr als $x'^t y'$, sondern wir fügen eine symmetrische Matrix

$$A = (S^{-1})^t S^{-1} \quad (8.6)$$

ein.

Wenn wir das Skalarprodukt nicht als Rechenvorschrift für Zahlentupel sehen, sondern als geometrische Struktur, die unabhängig von einer Basis besteht, dann können wir auch für abstrakte Vektorräume formulieren, was wir mit einem Skalarprodukt meinen. Es ist eine Abbildung, die zwei Vektoren eine Zahl zuordnet (mit gewissen Eigenschaften, die wir gleich besprechen), die uns damit in die Lage versetzt, über Längen und Winkel zu sprechen.

Ein gegebener abstrakter Vektorraum hat eine solche Struktur erst einmal nicht; es ist eine zusätzliche Struktur, die wir einführen können. Das setzt also unser Abstraktionsprogramm fort: Erst haben wir uns die Addition und Skalarmultiplikation vom \mathbb{R}^n abgeschaut und dadurch allgemein den Begriff eines reellen Vektorraums eingeführt, jetzt schauen wir uns die durch Längen und Winkel gegebene Struktur ab.

Wir definieren jetzt, was wir mit einem Skalarprodukt auf einem allgemeinen (auch ∞ -dimensionalen) reellen oder komplexen Vektorraum verstehen, indem wir die Eigenschaften des kanonischen Skalarprodukts auf \mathbb{R}^n und \mathbb{C}^n übertragen:

Definition 8.1. Sei V ein reeller Vektorraum.

Ein **Skalarprodukt** auf V ist eine Abbildung

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$$

mit den Eigenschaften (für alle $\alpha \in \mathbb{R}$, $u, v, w \in V$)

1. $\langle \cdot, \cdot \rangle$ ist bilinear, d.h.

$$\langle v, \alpha w \rangle = \alpha \langle v, w \rangle$$

$$\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$$

$$\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$$

$$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$$

2. $\langle \cdot, \cdot \rangle$ ist symmetrisch, d.h.

$$\langle v, w \rangle = \langle w, v \rangle$$

3. $\langle \cdot, \cdot \rangle$ ist positiv definit, d.h.

$$\langle v, v \rangle > 0 \text{ für } v \neq 0.$$

Definition 8.2. Sei V ein komplexer Vektorraum.

Ein **Skalarprodukt** auf V ist eine Abbildung

$$\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$$

mit den Eigenschaften (für alle $\alpha \in \mathbb{C}$, $u, v, w \in V$)

1. $\langle \cdot, \cdot \rangle$ ist sesquilinear, d.h.

$$\langle v, \alpha w \rangle = \alpha \langle v, w \rangle$$

$$\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$$

$$\langle \alpha v, w \rangle = \bar{\alpha} \langle v, w \rangle$$

$$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$$

2. $\langle \cdot, \cdot \rangle$ ist hermitesch, d.h.

$$\langle v, w \rangle = \overline{\langle w, v \rangle}$$

3. $\langle \cdot, \cdot \rangle$ ist positiv definit, d.h.

$$\langle v, v \rangle > 0 \text{ für } v \neq 0.$$

Der Unterschied zwischen dem reellen und dem komplexen Skalarprodukt besteht also nur darin, dass

- bei Multiplikation des ersten Eintrags mit einem Skalar der Skalar komplex konjugiert wird und
- bei Vertauschung der Einträge das Skalarprodukt komplex konjugiert wird.

Die dritte Eigenschaft (positive Definitheit) ermöglicht es, eine „Länge“ der Vektoren zu definieren: Wir nennen

$$\|v\| = \sqrt{\langle v, v \rangle} \tag{8.7}$$

die (durch $\langle \cdot, \cdot \rangle$ induzierte) **Norm** auf V .

Beispiel 8.1. Sei $\mathcal{C}([-1, 1], \mathbb{C})$ der komplexe Vektorraum der komplexwertigen, stetigen Funktionen auf dem reellen Intervall $[-1, 1]$. Dann definiert

$$\langle f, g \rangle := \int_{-1}^1 \overline{f(x)} g(x) dx \tag{8.8}$$

ein Skalarprodukt.

Die Sesquilinearität kann man leicht überprüfen, ebenso die Hermitizität. Die Be-

dingung der positiven Definitheit ist auch erfüllt:

$$\langle f, f \rangle = \int_{-1}^1 |f(x)|^2 dx > 0 \text{ für } f \neq 0, \text{ wenn } f \text{ stetig ist.} \quad (8.9)$$

Für reelle und komplexe Skalarprodukte gelten (wie auf \mathbb{R}^n und \mathbb{C}^n) die folgenden Ungleichungen

Satz 8.1. Sei V ein reeller oder komplexer Vektorraum mit Skalarprodukt $\langle \cdot, \cdot \rangle$. Dann gilt:

1. für alle $v, w \in V$: $|\langle v, w \rangle| \leq \|v\| \|w\|$ (Cauchy-Schwarzsche Ungleichung)
2. für alle $v, w \in V$: $\|v + w\| \leq \|v\| + \|w\|$ (Dreiecksungleichung)

Der Beweis der Cauchy-Schwarzschen Ungleichung im reellen Fall, wie wir früher gesehen haben (Seite 38), beruht nur auf der Bilinearität, der Symmetrie und der Positivität, d.h. er lässt sich direkt für ein allgemeines reelles Skalarprodukt führen. Für das komplexe Skalarprodukt muss man den Beweis etwas anpassen, aber er funktioniert im Wesentlichen genauso. Die Dreiecksungleichung ist, wie wir auch früher gesehen haben (Seite 39), eine Folgerung aus der Cauchy-Schwarzschen Ungleichung.

Beispiel 8.2. Sei $\ell^2(\mathbb{C})$ der Raum der komplexen quadratsummierbaren Folgen $(a_n)_{n \in \mathbb{N}}$,

$$\ell^2(\mathbb{C}) = \{(a_n)_{n \in \mathbb{N}} \mid \forall n \in \mathbb{N} : a_n \in \mathbb{C}, \sum_{n=0}^{\infty} |a_n|^2 < \infty\}. \quad (8.10)$$

Wir definieren auf $\ell^2(\mathbb{C})$ ein Skalarprodukt durch

$$\langle (a_n), (b_n) \rangle := \sum_{n=1}^{\infty} \overline{a_n} b_n. \quad (8.11)$$

Wir müssen erst einmal sicherstellen, dass die Reihe existiert. Wegen der Cauchy-Schwarzschen Ungleichung im \mathbb{C}^N gilt

$$\sum_{n=1}^N |a_n| |b_n| \leq \sqrt{\sum_{n=1}^N |a_n|^2} \sqrt{\sum_{n=1}^N |b_n|^2} \quad (8.12)$$

für beliebige $N \in \mathbb{N}$. Da die rechte Seite für $N \rightarrow \infty$ konvergiert, ist auch $\sum_{n=1}^{\infty} |a_n| |b_n|$ endlich, daher konvergiert $\sum_{n=1}^{\infty} \overline{a_n} b_n$ absolut.

Bemerkung

1. Einen reellen Vektorraum mit Skalarprodukt nennt man auch einen **Euklidischen Vektorraum**.

2. Einen komplexen Vektorraum mit Skalarprodukt nennt man auch einen **Unitären Vektorraum** oder **Prä-Hilbertraum**.

3. Die durch ein Skalarprodukt definierte Norm $\|v\| = \sqrt{\langle v, v \rangle}$ ermöglicht es, Konvergenz von Folgen zu diskutieren, weil wir einen Begriff haben, der uns sagt, wie verschieden zwei Vektoren sind: Eine Folge $(v_n)_{n \in \mathbb{N}}$ heißt konvergent gegen $w \in V$, wenn

$$\lim_{n \rightarrow \infty} \|v_n - w\| = 0. \quad (8.13)$$

4. Ebenso kann man den Begriff einer Cauchy-Folge definieren: $(v_n)_{n \in \mathbb{N}}$ ist eine Cauchy-Folge, wenn es zu jedem $\epsilon > 0$ ein $N \in \mathbb{N}$ gibt, sodass für alle $m, n \geq N$ gilt, dass $\|v_m - v_n\| < \epsilon$.

5. Ein Vektorraum mit Skalarprodukt heißt vollständig genau dann, wenn alle Cauchy-Folgen gegen einen Grenzwert konvergieren. Einen vollständigen komplexen Vektorraum mit Skalarprodukt nennt man **Hilbertraum**. Beispiele für Hilberträume sind \mathbb{C}^n und $\ell^2(\mathbb{C})$. $\mathcal{C}([-1, 1], \mathbb{C})$ mit dem oben eingeführten Skalarprodukt ist nicht vollständig und daher kein Hilbertraum.

Wir haben das Skalarprodukt als eine Operation eingeführt, die zwei Vektoren eine Zahl zuordnet. Wir wollen nun das Ganze aus einem anderen Blickwinkel betrachten, wobei wir jetzt annehmen wollen, dass V ein komplexer Vektorraum ist (für einen reellen Vektorraum funktioniert alles analog). Wenn wir einen Vektor $v \in V$ fixieren, liefert $\langle v, \cdot \rangle$ eine lineare Abbildung von V in die komplexen Zahlen,

$$\langle v, \cdot \rangle : V \rightarrow \mathbb{C} \quad (8.14)$$

$$w \mapsto \langle v, w \rangle, \quad (8.15)$$

d.h. jedem Vektor w wird das Skalarprodukt mit dem fest gewählten Vektor v zugeordnet. Dank des Skalarprodukts können wir also jedem Vektor $v \in V$ eine lineare Abbildung $\langle v, \cdot \rangle$, also ein Element in $\text{Hom}(V, \mathbb{C})$ zuordnen. Diese Zuordnung ist wieder semi-linear (da das Skalarprodukt semi-linear im ersten Eintrag ist): Das Skalarprodukt beschert uns also eine semi-lineare Abbildung

$$S : V \rightarrow \text{Hom}(v, \mathbb{C}) \quad (8.16)$$

$$v \mapsto \langle v, \cdot \rangle. \quad (8.17)$$

(Semi-linear wegen $\lambda v \mapsto \langle \lambda v, \cdot \rangle = \bar{\lambda} \langle v, \cdot \rangle$, im reellen Fall ist die entsprechende Abbildung linear.)

Die Abbildung S ist injektiv, denn $S(v) = 0$ genau dann, wenn für alle $w \in V$ gilt $\langle v, w \rangle = 0$; das ist aber nur für $v = 0$ erfüllt. S ist aber im Allgemeinen nicht surjektiv: Auf $\mathcal{C}([-1, 1], \mathbb{C})$ ist

$$\begin{aligned} \delta : \mathcal{C}([-1, 1], \mathbb{C}) &\rightarrow \mathbb{C} \\ f &\mapsto f(0) \end{aligned}$$

eine Linearform (also $\delta \in \text{Hom}(\mathcal{C}([-1, 1], \mathbb{C}), \mathbb{C})$), aber es gibt keine stetige Funktion g , sodass für jedes f gilt

$$\langle g, f \rangle = \int_{-1}^1 \overline{g(x)} f(x) dx = f(0). \quad (8.18)$$

(δ ist das sogenannte *Delta-Funktional* auf $\mathcal{C}([-1, 1], \mathbb{C})$.)

Für endlich-dimensionale Vektorräume ist S aber bijektiv, da $\dim \text{Hom}(V, \mathbb{C}) = \dim V$. (Mit Hilfe einer Basis von V können wir den Raum der linearen Abbildungen von V ($\dim V = m$) nach \mathbb{C} ($\dim \mathbb{C} = 1$) mit dem Raum der $1 \times m$ -Matrizen $M(1 \times m, \mathbb{C})$ (also der Zeilenvektoren) identifizieren.)

Für endlich-dimensionale Vektorräume mit Skalarprodukt gilt also: Zu jeder Linearform $F : V \rightarrow \mathbb{C}$ gibt es ein $v \in V$, sodass für alle $w \in V$: $F(w) = \langle v, w \rangle$.

Mit dem Raum der Linearformen werden wir uns in Kapitel 9 noch genauer auseinander setzen. Im Folgenden wollen wir den mit einem Skalarprodukt einhergehenden Begriff der Orthogonalität diskutieren.

8.2 Orthonormalsysteme und Matrixdarstellungen

Wir haben schon im Fall des \mathbb{R}^n gesehen, dass es nützlich ist, eine Basis aus orthogonalen Vektoren der Länge 1 zu betrachten: die kanonischen Einheitsvektoren e_1, \dots, e_n . Sobald wir ein Skalarprodukt auf einem Vektorraum V definiert haben, können wir etwas Ähnliches tun.

Definition 8.3. Sei V ein (reeller oder komplexer) Vektorraum mit Skalarprodukt.

1. Zwei Vektoren $v, w \in V$ heißen **orthogonal** genau dann, wenn

$$\langle v, w \rangle = 0. \quad (8.19)$$

2. Eine Familie $(v_i)_{i \in I}$ von Vektoren in V heißt **Orthonormalsystem** genau dann, wenn

$$\text{für alle } i, j \in I, i \neq j \text{ gilt } \langle v_i, v_j \rangle = 0 \quad (8.20a)$$

$$\text{für alle } i \in I \text{ gilt } \|v_i\| = 1. \quad (8.20b)$$

(Oder kurz: $\langle v_i, v_j \rangle = \delta_{ij}$.)

Ein Orthonormalsystem ist immer linear unabhängig. Denn sei

$$\alpha_1 v_{i_1} + \dots + \alpha_m v_{i_m} = 0. \quad (8.21)$$

Bilden wir vom obigen Ausdruck das Skalarprodukt mit v_{i_1} , erhalten wir $\alpha_1 = 0$, analog folgt $\alpha_2 = 0, \dots, \alpha_m = 0$.

Beispiel 8.3. In $\mathcal{C}([-1, 1], \mathbb{C})$ mit dem oben definierten Skalarprodukt (siehe (8.8)) seien die Funktionen f_n ($n \in \mathbb{Z}$) mit

$$f_n(x) = \frac{1}{\sqrt{2}} e^{i\pi n x} \quad (8.22)$$

gegeben. $(f_n)_{n \in \mathbb{Z}}$ ist ein Orthonormalsystem:

$$\langle f_m, f_n \rangle = \int_{-1}^1 \overline{f_m(x)} f_n(x) dx \quad (8.23)$$

$$= \frac{1}{2} \int_{-1}^1 e^{i\pi(n-m)x} dx \quad (8.24)$$

$$= \begin{cases} 1 & \text{für } m = n \\ 0 & \text{für } m \neq n. \end{cases} \quad (8.25)$$

$(f_n)_{n \in \mathbb{Z}}$ ist aber keine Basis von $\mathcal{C}([-1, 1], \mathbb{C})$, da eine allgemeine stetige Funktion sich nicht als Linearkombination der f_n wird schreiben lassen können. (In einem gewissen Sinne, den Sie in der Analysis lernen, können wir alle Funktionen durch

unendliche Kombinationen der f_n erhalten — das führt zum Begriff der Fourier-Reihen.)

Für jeden endlich-dimensionalen Untervektorraum können wir eine **Orthonormalbasis** finden, also ein Orthonormalsystem, das gleichzeitig eine Basis ist (für einen unendlich-dimensionalen Vektorraum ist dies typischerweise nicht möglich; daher führt man in dem Fall einen neuen Basisbegriff (die sogenannte *Hilbert-Basis*) ein — das geht aber über die lineare Algebra hinaus).

Wir lernen nun ein praktisches Verfahren kennen, mit dem man zu einem gegebenen Satz von Vektoren, die einen Untervektorraum U aufspannen, eine Orthonormalbasis von U finden kann.

Orthonormalisierungsverfahren nach Gram und Schmidt

Sei (v_1, \dots, v_n) eine linear unabhängige Familie von Vektoren in einem (reellen oder komplexen) Vektorraum V (der endlich- oder unendlich-dimensional sein kann), die einen n -dimensionalen Untervektorraum U aufspannen. Wir konstruieren nun eine Orthonormalbasis (w_1, \dots, w_n) von U :

1. Setze

$$w_1 := \frac{1}{\|v_1\|} v_1. \quad (8.26)$$

Dann ist w_1 normiert, $\|w_1\| = 1$. Die Operation ist möglich, da die Vektoren als linear unabhängig vorausgesetzt waren, d.h. insbesondere ist $v_1 \neq 0$.

2. Setze zunächst

$$\hat{w}_2 = v_2 - \langle w_1, v_2 \rangle w_1. \quad (8.27)$$

Dann ist \hat{w}_2 orthogonal zu w_1 :

$$\langle w_1, \hat{w}_2 \rangle = \langle w_1, v_2 \rangle - \langle w_1, \langle w_1, v_2 \rangle w_1 \rangle \quad (8.28)$$

$$= \langle w_1, v_2 \rangle - \langle w_1, v_2 \rangle \underbrace{\langle w_1, w_1 \rangle}_{=1} = 0. \quad (8.29)$$

2'. Setze

$$w_2 := \frac{1}{\|\hat{w}_2\|} \hat{w}_2. \quad (8.30)$$

Dann ist $\|w_2\| = 1$. Die Operation ist möglich, da $\hat{w}_2 \neq 0$: Denn wäre $\hat{w}_2 = 0$, so wäre $v_2 = \langle w_1, v_2 \rangle w_1 = \frac{\langle w_1, v_2 \rangle}{\|v_1\|} v_1$, also wären v_1 und v_2 linear abhängig — sie waren aber als linear unabhängig vorausgesetzt worden.

3. Setze

$$\hat{w}_3 := v_3 - \langle w_1, v_3 \rangle w_1 - \langle w_2, v_3 \rangle w_2. \quad (8.31)$$

Dann ist \hat{w}_3 orthogonal zu w_1 und w_2 . Da v_1, v_2, v_3 linear unabhängig sind, ist $\hat{w}_3 \neq 0$.

3'. Setze

$$w_3 := \frac{1}{\|\hat{w}_3\|} \hat{w}_3. \quad (8.32)$$

⋮

n . Setze

$$\hat{w}_n := v_n - \sum_{i=1}^{n-1} \langle w_i, v_n \rangle w_i. \quad (8.33)$$

n' . Setze

$$w_n := \frac{1}{\|\hat{w}_n\|} \hat{w}_n. \quad (8.34)$$

Dann ist (w_1, \dots, w_n) die gewünschte Orthonormalbasis von U .

Bemerkung

Wenn man dieses Verfahren auf eine linear abhängige Familie (v_1, \dots, v_n) anwendet, so erhält man in irgend einem Schritt $\hat{w}_i = 0$. Solche Vektoren lässt man dann aus und man erhält wieder eine Orthonormalbasis des von (v_1, \dots, v_n) erzeugten Untervektorraums.

Beispiel 8.4. Sei

$$V = \mathbb{C}^4 \quad \text{und} \quad v_1 = \begin{pmatrix} 1 \\ i \\ 0 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -i \\ -1 \\ 1 \\ i \end{pmatrix} \quad (8.35)$$

gegeben. Wir wenden das Orthonormalisierungsverfahren an:

1.

$$\|v_1\|^2 = 3 \quad \implies \quad w_1 = \frac{1}{\|v_1\|} v_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ i \\ 0 \\ 1 \end{pmatrix} \quad (8.36)$$

2.

$$\langle w_1, v_2 \rangle = \frac{1}{\sqrt{3}} (1 \quad -i \quad 0 \quad 1) \begin{pmatrix} -i \\ -1 \\ 1 \\ i \end{pmatrix} = \frac{1}{\sqrt{3}} i \quad (8.37)$$

$$\implies \hat{w}_2 = v_2 - \langle w_1, v_2 \rangle w_1 = \begin{pmatrix} -i \\ -1 \\ 1 \\ i \end{pmatrix} - \frac{i}{3} \begin{pmatrix} 1 \\ i \\ 0 \\ 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -4i \\ -2 \\ 3 \\ 2i \end{pmatrix}. \quad (8.38)$$

2'.

$$\|\hat{w}_2\|^2 = \frac{1}{9} (16 + 4 + 9 + 4) = \frac{33}{9} \quad (8.39)$$

$$\implies w_2 = \frac{1}{\sqrt{33}} \begin{pmatrix} -4i \\ -2 \\ 3 \\ 2i \end{pmatrix}. \quad (8.40)$$

Die Vektoren

$$w_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ i \\ 0 \\ 1 \end{pmatrix}, \quad w_2 = \frac{1}{\sqrt{33}} \begin{pmatrix} -4i \\ -2 \\ 3 \\ 2i \end{pmatrix} \quad (8.41)$$

bilden eine Orthonormalbasis des von v_1 und v_2 erzeugten Untervektorraums.

Beispiel 8.5. Sei $V = \mathbb{R}[t]$ der Raum der reellen Polynome in t mit Skalarprodukt

$$\langle p_1, p_2 \rangle = \int_{-1}^1 p_1(t) p_2(t) dt. \quad (8.42)$$

Seien $q_n(t) = t^n$ (für $n \in \mathbb{N}_0$) die Monome in t . Auch wenn das eine unendliche Familie ist, können wir das Gram-Schmidtsche Orthonormalisierungsverfahren darauf anwenden, um ein unendliches Orthonormalsystem (p_0, p_1, \dots) von Polynomen zu konstruieren.

0.

$$\langle q_0, q_0 \rangle = \int_{-1}^1 dt = 2 \quad \Longrightarrow \quad p_0 = \frac{1}{\sqrt{2}} q_0 = \frac{1}{\sqrt{2}} \quad (8.43)$$

1.

$$\langle p_0, q_1 \rangle = \frac{1}{\sqrt{2}} \int_{-1}^1 t dt = \frac{1}{\sqrt{2}} \left[\frac{1}{2} t^2 \right]_{-1}^1 = 0 \quad \Longrightarrow \quad \hat{p}_1 = q_1 = t \quad (8.44)$$

1'.

$$\|\hat{p}_1\|^2 = \int_{-1}^1 t^2 dt = \frac{2}{3} \quad \Longrightarrow \quad p_1 = \sqrt{\frac{3}{2}} t \quad (8.45)$$

2.

$$\langle p_0, q_2 \rangle = \frac{1}{\sqrt{2}} \int_{-1}^1 t^2 dt = \frac{\sqrt{2}}{3} \quad (8.46)$$

$$\langle p_1, q_2 \rangle = \sqrt{\frac{3}{2}} \int_{-1}^1 t^3 dt = 0 \quad (8.47)$$

$$\Longrightarrow \quad \hat{p}_2 = t^2 - \frac{\sqrt{2}}{3} \frac{1}{\sqrt{2}} = t^2 - \frac{1}{3} \quad (8.48)$$

2'.

$$\|\hat{p}_2\|^2 = \int_{-1}^1 \left(t^2 - \frac{1}{3}\right)^2 dt = \int_{-1}^1 \left(t^4 - \frac{2}{3}t^2 + \frac{1}{9}\right) dt \quad (8.49)$$

$$= 2 \left(\frac{1}{5} - \frac{2}{3} \frac{1}{3} + \frac{1}{9}\right) = \frac{8}{45} \quad (8.50)$$

$$\Rightarrow p_2 = \sqrt{\frac{45}{8}} \left(t^2 - \frac{1}{3}\right) = \sqrt{\frac{5}{2}} \left(\frac{3}{2}t^2 - \frac{1}{2}\right) \quad (8.51)$$

usw.

Bis auf die Normierung erhält man auf diese Weise das System der sogenannten **Legendre-Polynome** P_n , die Ihnen in Physik und Mathematik wieder begegnen werden (die Beziehung zu den eben konstruierten Polynomen ist $p_n = \frac{2n+1}{2} P_n$). Die Legendre-Polynome sind Eigenvektoren eines Differentialoperators,

$$\left[(1-x^2)\frac{d^2}{dx^2} - 2x\frac{d}{dx}\right]P_n(x) = -n(n+1)P_n(x). \quad (8.52)$$

Der Orthogonalitätsbegriff kann auch auf Mengen von Vektoren ausgedehnt werden. Wir nennen zwei Untervektorräume $U_1, U_2 \subset V$ orthogonal genau dann, wenn

$$\text{für alle } u_1 \in U_1, u_2 \in U_2 \text{ gilt: } \langle u_1, u_2 \rangle = 0. \quad (8.53)$$

Zu einem gegebenen Untervektorraum $U \subset V$ definieren wir das **orthogonale Komplement** U^\perp als den Raum aller Vektoren, die orthogonal sind zu allen Vektoren in U , d.h.

$$U^\perp = \{v \in V \mid \text{für alle } u \in U : \langle v, u \rangle = 0\}. \quad (8.54)$$

Die Untervektorräume U und U^\perp sind unabhängig in dem Sinne, dass

$$U \cap U^\perp = \{0\}. \quad (8.55)$$

Die Summe der Untervektorräume ist also eine direkte Summe $U \oplus U^\perp$.

Für einen endlich-dimensionalen Vektorraum V gilt für jeden Untervektorraum $U \subset V$

$$V = U \oplus U^\perp, \quad (8.56)$$

d.h. ein gegebener Untervektorraum spannt zusammen mit seinem orthogonalen Komplement den gesamten Vektorraum auf. Geometrisch ist dies auch anschaulich: Stellt man sich z.B. eine Gerade (durch den Ursprung) im \mathbb{R}^3 vor, so ist das orthogonale Komplement die Ebene (durch den Ursprung), die orthogonal auf der Gerade steht. Zusammen spannen sie den gesamten dreidimensionalen Raum auf. Beweisen kann man die Aussage mit dem Gram-Schmidtschen Orthonormalisierungsverfahren: Man beginne mit einer Basis (u_1, \dots, u_r) von U und ergänze sie zu einer Basis $(u_1, \dots, u_r, u_{r+1}, \dots, u_n)$ von V . Das Orthonormalisierungsverfahren liefert eine Orthonormalbasis $(v_1, \dots, v_r, v_{r+1}, \dots, v_n)$ von

V , wobei (v_1, \dots, v_r) eine Orthonormalbasis von U ist, während — wie man sich überzeugen kann — (v_{r+1}, \dots, v_n) eine Orthonormalbasis von U^\perp ergibt.

Orthonormalbasen erweisen sich als besonders praktisch, wenn man das Skalarprodukt in Koeffizienten auswertet. Bevor wir dazu kommen, wollen wir uns ansehen, wie die Koordinatendarstellung in einer allgemeinen Basis aussieht.

Sei also $\mathcal{A} = (v_1, \dots, v_n)$ eine Basis des endlich-dimensionalen Vektorraums V (reell oder komplex) mit Skalarprodukt $\langle \cdot, \cdot \rangle$. Sei $\Phi_{\mathcal{A}} : x \mapsto \sum_{i=1}^n x_i v_i$ der gewohnte Basisisomorphismus. In Koordinaten wird das Skalarprodukt zu

$$\langle \Phi_{\mathcal{A}}(x), \Phi_{\mathcal{A}}(y) \rangle = \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j \right\rangle \quad (8.57)$$

$$= \sum_{i,j=1}^n \bar{x}_i y_j \langle v_i, v_j \rangle \quad (8.58)$$

$$= \bar{x}^t A y \quad (8.59)$$

mit der Matrix $A = (\langle v_i, v_j \rangle)$.

Wenn wir eine Orthonormalbasis gewählt haben, ist $A = E_n$ die Einheitsmatrix und das Skalarprodukt wird zu

$$\langle \Phi_{\mathcal{A}}(x), \Phi_{\mathcal{A}}(y) \rangle = \bar{x}^t y, \quad (8.60)$$

also das kanonische Skalarprodukt auf \mathbb{C}^n (oder \mathbb{R}^n wenn wir reelle Vektorräume betrachten).

Wir sehen daraus zwei Dinge: Zum einen — wie angekündigt — dass die Orthonormalbasis besonders praktisch für die Beschreibung des Skalarprodukts auf dem Koordinatenraum ist. Zum anderen, dass wir durchaus auch auf dem \mathbb{C}^n (bzw. \mathbb{R}^n) andere Skalarprodukte definieren können, nämlich von der Form

$$\langle x, y \rangle = \bar{x}^t A y \quad (8.61)$$

mit einer Matrix A . Diese Matrix ist nicht beliebig: Sie erfüllt

- $A^* = A$ (d.h. sie ist symmetrisch bzw. hermitesch)
- für alle $x \neq 0$: $\bar{x}^t A x > 0$.

Eine solche Matrix heißt **positiv definit**. Man kann nachprüfen, dass eine als $A = (\langle v_i, v_j \rangle)$ definierte Matrix diese Eigenschaften hat, aber auch, dass jede positiv definite Matrix A mittels (8.61) ein Skalarprodukt auf \mathbb{C}^n (bzw. auf \mathbb{R}^n , wenn A reell ist) definiert.

Die Matrixdarstellung ändert sich bei einem Basiswechsel. Sei $\mathcal{B} = (w_1, \dots, w_n)$ eine weitere Basis und $x' = S x$ beschreibe den Basiswechsel (x' sind die Koordinaten bezüglich \mathcal{B}). Sei $B = (\langle w_i, w_j \rangle)$ die Matrixdarstellung des Skalarprodukts bezüglich \mathcal{B} . Dann

ist

$$\langle \Phi_B(x'), \Phi_B(y') \rangle = \overline{x'}^t B y' \tag{8.62}$$

$$= \overline{(Sx)}^t B S y \tag{8.63}$$

$$= \overline{x}^t S^* B S y. \tag{8.64}$$

Andererseits wissen wir, dass das Skalarprodukt in den Koeffizienten x und y durch $\overline{x}^t A y$ gegeben ist, also muss gelten

$$A = S^* B S \quad \text{und} \quad B = (S^{-1})^* A S^{-1}. \tag{8.65}$$

Matrizen A und B , die in einer solchen Beziehung stehen, nennt man **kongruent**. Kongruenz definiert eine Äquivalenzrelation. Wieder stellt sich dann die Frage, wie viele Äquivalenzklassen es gibt und was geeignete Normalformen sind.

Für positiv definite Matrizen können wir die Frage direkt beantworten: Jede positiv definite Matrix A definiert ein Skalarprodukt, und durch Wahl einer Orthonormalbasis (die immer existiert) wird das Skalarprodukt durch die Einheitsmatrix dargestellt. Also gilt

Satz 8.2. *Zu jeder positiv definiten $n \times n$ -Matrix A gibt es eine invertierbare Matrix S , sodass*

$$S^* A S = E_n. \tag{8.66}$$

Ist A reell, so kann S reell gewählt werden.

Dies ist ein Spezialfall des Sylvesterschen Trägheitssatzes, der Auskunft über die Normalform (bezüglich Kongruenz) beliebiger symmetrischer bzw. hermitescher Matrizen gibt:

Satz 8.3 (Sylvesterscher Trägheitssatz). *Sei A eine hermitesche $n \times n$ -Matrix. Dann gibt es eine invertierbare Matrix S , sodass*

$$S^* A S = \left(\begin{array}{cccccccc} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & -1 & & & & \\ & & & & \ddots & & & \\ & & & & & -1 & & \\ & & & & & & 0 & \\ & & & & & & & \ddots \\ & & & & & & & & 0 \end{array} \right) \left. \begin{array}{l} \} \\ \} \\ \} \\ \} \\ \} \\ \} \\ \} \\ \} \\ \} \end{array} \right\} \begin{array}{l} r_1 \\ r_2 \\ r_3 \end{array}, \tag{8.67}$$

wobei die Anzahl r_1 der Einsen, die Zahl r_2 der negativen Einträge (-1) und die Zahl r_3 der Nullen auf der Diagonalen eindeutig ist (d.h. wenn es S und T gibt, sodass

S^*AS und T^*AT von der obigen Form sind, dann sind in dieser Darstellung auch r_1, r_2, r_3 gleich). Ist A reell, so kann S auch reell gewählt werden.

Hier müssen wir zwei Dinge beweisen: Erstens, dass es eine solche Matrix S gibt — das werden wir später bei der Diagonalisierung hermitescher Matrizen sehen — und zweitens, dass die Zahlen r_1, r_2, r_3 eindeutig sind. Bevor wir uns das ansehen, wollen wir noch kurz diskutieren, wie man solche hermiteschen Matrizen, die typischerweise nicht positiv definit sind, mit Hilfe einer Art verallgemeinerten Skalarprodukts interpretieren kann.

Jede hermitesche Matrix definiert eine hermitesche Sesquilinearform

$$\mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C} \tag{8.68}$$

$$(x, y) \mapsto \bar{x}^t A y. \tag{8.69}$$

Unter einem Basiswechsel wird die Sesquilinearform mit einer neuen Matrix B beschrieben mit $A = S^*BS$. Die Kongruenzklasse von A beschreibt also alle Matrixdarstellungen einer gegebenen hermiteschen Sesquilinearform. Analog definiert jede reelle symmetrische Matrix eine Bilinearform auf \mathbb{R}^n .

Beispiel 8.6. Eine in der Physik wichtige Bilinearform ist die **Minkowski-Metrik** im \mathbb{R}^4 , wobei wir die Koordinaten $\begin{pmatrix} x_0 \\ \vdots \\ x_3 \end{pmatrix}$ verwenden wollen und $x_0 = ct$ als Zeitkoordinate zu interpretieren ist (c ist die Lichtgeschwindigkeit). Dann ist

$$\langle x, y \rangle_{\text{Minkowski}} = x_0y_0 - x_1y_1 - x_2y_2 - x_3y_3. \tag{8.70}$$

Diese Bilinearform ist nicht positiv definit, daher handelt es sich — auch wenn die Notation das suggeriert — nicht um ein Skalarprodukt. Die zugehörige Matrixdarstellung ist

$$\eta := A_{\text{Minkowski}} = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}. \tag{8.71}$$

Nun zum teilweisen Beweis des Sylvesterschen Satzes:

Beweis der Eindeutigkeit der Zahlen r_1, r_2, r_3 . Sei $s(x, y) = \bar{x}^t A y$ die durch A definierte hermitesche Bilinearform. Sei S^*AS von der im Satz gegebenen Form. Das bedeutet, es gibt eine Basis $(p_1, \dots, p_{r_1}, m_1, \dots, m_{r_2}, n_1, \dots, n_{r_3})$ mit

$$s(p_i, p_j) = \delta_{ij}, \quad s(m_i, m_j) = -\delta_{ij} \tag{8.72}$$

und alle anderen „Produkte“ (bezüglich s) der Basisvektoren ergeben Null,

$$s(n_i, n_j) = 0, \quad s(n_i, p_j) = 0, \quad s(n_i, m_j) = 0, \quad s(p_i, m_j) = 0. \tag{8.73}$$

Sei T^*AT eine weitere Darstellung der gewünschten Form mit der Verteilung (s_1, s_2, s_3) der Einsen, Minus-Einsen und Nullen. Zu dieser Darstellung gehört eine Basis mit analogen Eigenschaften wie oben, wir notieren sie als $(p'_1, \dots, p'_{s_1}, m'_1, \dots, m'_{s_2}, n'_1, \dots, n'_{s_3})$.

Wir zeigen nun, dass die Familie $(p'_1, \dots, p'_{s_1}, m_1, \dots, m_{r_2}, n_1, \dots, n_{r_3})$, in der wir die ersten s_1 Basisvektoren der zweiten Basis und die letzten $r_2 + r_3$ Basisvektoren der ersten Basis zusammengestellt haben, linear unabhängig ist. Da es höchstens n linear unabhängige Vektoren geben kann, folgt daraus, dass $s_1 \leq r_1$. Durch Vertauschen der Rollen der ersten und der zweiten Basis folgt dann auch $r_1 \leq s_1$, also $s_1 = r_1$. Durch ein analoges Argument (man wählt im ersten Schritt die mittleren s_2 Vektoren der zweiten Basis und die ersten r_1 und die letzten r_3 der zweiten Basis und zeigt, dass sie linear unabhängig sind; danach vertauscht man die Rollen der ersten und zweiten Basis) erhält man auch $s_2 = r_2$. Da s_3 und r_3 durch die Größe der Matrix festgelegt sind, folgt auch $s_3 = r_3$.

Es fehlt also nur noch der Beweis der linearen Unabhängigkeit. Dazu setzen wir die Gleichung

$$v := \alpha_1 p'_1 + \dots + \alpha_{s_1} p'_{s_1} = \beta_1 m_1 + \dots + \beta_{r_2} m_{r_2} + \gamma_1 n_1 + \dots + \gamma_{r_3} n_{r_3} =: w \quad (8.74)$$

an; $v - w = 0$ ist dann die Gleichung, von der wir zeigen wollen, dass sie nur eine Lösung hat, nämlich wenn alle Koeffizienten $\alpha_i, \beta_i, \gamma_i$ Null sind. Wir setzen v bzw. w in s ein und erhalten

$$s(v, v) = |\alpha_1|^2 + \dots + |\alpha_{s_1}|^2 \geq 0 \quad \text{und} \quad s(w, w) = -|\beta_1|^2 - \dots - |\beta_{r_2}|^2 \leq 0, \quad (8.75)$$

wegen $v = w$ ist also $s(v, v) = s(w, w) = 0$. Somit sind alle $\alpha_i = 0$ und alle $\beta_i = 0$, und da (n_1, \dots, n_{r_3}) linear unabhängig ist, sind auch alle $\gamma_i = 0$. Wie gewünscht, haben wir also gezeigt, dass $(p'_1, \dots, p'_{s_1}, m_1, \dots, m_{r_2}, n_1, \dots, n_{r_3})$ linear unabhängig ist. \square

Damit haben wir die Eindeutigkeit der Normalform kongruenter Matrizen gezeigt. Dass die Normalform immer erreicht werden kann, wird sich später aus der Diagonalisierung hermitescher Matrizen ergeben.

8.3 Orthogonale und unitäre Abbildungen

Orthogonale Matrizen hatten wir über die Forderung eingeführt, dass wir lineare Abbildungen auf \mathbb{R}^n finden wollten, die Längen erhalten. Analog definieren wir

Definition 8.4. Sei V ein reeller Vektorraum mit Skalarprodukt. Eine invertierbare lineare Abbildung $F : V \rightarrow V$ heißt **orthogonal** genau dann, wenn für alle $v, w \in V$ gilt

$$\langle Fv, Fw \rangle = \langle v, w \rangle. \quad (8.76)$$

Bemerkung

Eine lineare Abbildung mit der Eigenschaft 8.76 ist automatisch injektiv,⁵ d.h. wenn V endlich-dimensional ist, muss aufgrund der Dimensionsformel F auch surjektiv, also invertierbar sein. Für endlich-dimensionale Vektorräume kann man also die Einschränkung „invertierbar“ in der obigen Definition weglassen.

Für einen endlich-dimensionalen Vektorraum mit Orthonormalbasis \mathcal{A} wird F durch eine Matrix $A = M_{\mathcal{A}}(F)$ dargestellt; das Skalarprodukt wird in den Koordinaten zum kanonischen Skalarprodukt auf dem \mathbb{R}^n . Die obige Bedingung wird also zu

$$(Ax)^t Ay = x^t y, \quad (8.77)$$

was genau dann erfüllt ist, wenn $A^t A = E_n$, das heißt, wenn die darstellende Matrix orthogonal ist:

Satz 8.4. Sei V ein endlich-dimensionaler reeller Vektorraum mit Skalarprodukt und einer Orthonormalbasis \mathcal{A} . Dann ist eine Abbildung $F : V \rightarrow V$ genau dann orthogonal, wenn die darstellende Matrix $A = M_{\mathcal{A}}(F)$ orthogonal ist.

Die orthogonalen Abbildungen bilden eine Gruppe. Für endlich-dimensionale Vektorräume (der Dimension n) lässt sie sich mit der Matrixgruppe $O(n)$ identifizieren.

Bemerkung

Wir hatten im letzten Abschnitt auch Bilinearformen (bzw. Sesquilinearformen) diskutiert, die nicht positiv definit waren. Auch hier kann man sich fragen, welche Abbildungen die Bilinearformen erhalten. Betrachten wir den \mathbb{R}^4 mit der Minkowski-Metrik, den wir in Beispiel 8.6 eingeführt haben,

$$\langle x, y \rangle_{\text{Minkowski}} = x^t \eta y \quad \text{mit} \quad \eta = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}. \quad (8.78)$$

⁵denn sei $Fv = 0$, dann folgt aus (8.76) $\langle v, v \rangle = \langle Fv, Fv \rangle = 0$, also $v = 0$

Wenn wir für eine lineare Abbildung $x \mapsto Lx$ fordern, dass für alle $x, y \in \mathbb{R}^4$

$$\langle x, y \rangle_{\text{Minkowski}} = \langle Lx, Ly \rangle_{\text{Minkowski}}, \quad (8.79)$$

so folgt daraus, dass für alle $x, y \in \mathbb{R}^4$

$$x^t \eta y = (Lx)^t \eta (Ly) = x^t (L^t \eta L) y, \quad (8.80)$$

und somit

$$L^t \eta L = \eta. \quad (8.81)$$

Matrizen L mit dieser Eigenschaft bilden wieder eine Gruppe unter Matrixmultiplikation: Sie heißt **Lorentz-Gruppe**,

$$O(1, 3) = \{L \in GL(4, \mathbb{R}) \mid L^t \eta L = \eta\}. \quad (8.82)$$

Solche Matrizen L bewirken sogenannte Lorentz-Transformationen, die in der speziellen Relativitätstheorie eine wichtige Rolle spielen: Sie beschreiben die Transformation zwischen den Bezugssystemen zweier zueinander geradlinig gleichförmig bewegter Beobachter.

Was orthogonale Abbildungen für reelle Vektorräume sind, sind unitäre Abbildungen für komplexe Vektorräume:

Definition 8.5. Eine invertierbare lineare Abbildung $F : V \rightarrow V$ auf einem komplexen Vektorraum mit Skalarprodukt heißt **unitär** genau dann, wenn für alle $v, w \in V$

$$\langle Fv, Fw \rangle = \langle v, w \rangle. \quad (8.83)$$

Bemerkung

Für endlich-dimensionale Vektorräume kann man den Zusatz „invertierbar“ in der obigen Definition weglassen, da — analog zu der Bemerkung zur Definition 8.4 der orthogonalen Abbildung — eine lineare Abbildung mit der Eigenschaft (8.83) injektiv ist und für einen endlich-dimensionalen Vektorraum V somit invertierbar sein muss.

Auf einem endlich-dimensionalen Vektorraum V mit orthonormaler Basis \mathcal{A} ist das Skalarprodukt durch das kanonische Skalarprodukt auf \mathbb{C}^n dargestellt, daher wird eine unitäre Abbildung F durch eine Matrix $A = M_{\mathcal{A}}(F)$ mit

$$\bar{x}^t y = \overline{(Ax)}^t (Ay) = \bar{x}^t A^* A y \quad (8.84)$$

dargestellt, für die also $A^* A = E_n$ gilt, d.h. A ist unitär:

Satz 8.5. Sei V ein endlich-dimensionaler unitärer Vektorraum mit Orthonormalbasis \mathcal{A} . Eine lineare Abbildung $F : V \rightarrow V$ ist genau dann unitär, wenn die darstellende Matrix $A = M_{\mathcal{A}}(F)$ unitär ist.

Bemerkung

Diese einfache Eins-zu-Eins-Beziehung zwischen unitären Abbildungen und unitären Matrizen gilt nur in einer Orthonormalbasis; wählt man eine andere Basis, so sind die darstellenden Matrizen einer unitären Abbildung nicht unitär (eine analoge Bemerkung gilt für orthogonale Abbildungen und Matrizen).

Bemerkung

Unitäre Abbildungen von V nach V bilden die **unitäre Gruppe** $U(V)$. Für endlich-dimensionale Räume wird diese dargestellt durch die Gruppe der unitären Matrizen

$$U(n) = \{A \in GL(n, \mathbb{C}) \mid A^*A = E_n\}. \quad (8.85)$$

Orthogonale und unitäre Abbildungen sind die strukturerhaltenden Abbildungen auf Räumen mit Skalarprodukt. Da sie das Skalarprodukt erhalten, werden durch sie Orthonormalbasen (v_1, \dots, v_n) in Orthonormalbasen (Av_1, \dots, Av_n) überführt:

$$\langle Av_i, Av_j \rangle = \langle v_i, v_j \rangle = \delta_{ij}. \quad (8.86)$$

Wir wollen nun die Eigenschaften dieser Abbildungen besser verstehen.

Was sind die möglichen Eigenwerte? Sei $F : V \rightarrow V$ eine unitäre Abbildung auf dem unitären Vektorraum V , und sei $v \in V \setminus \{0\}$ mit

$$F(v) = \lambda \cdot v \quad , \quad \lambda \in \mathbb{C}. \quad (8.87)$$

Dann ist

$$\langle v, v \rangle = \langle F(v), F(v) \rangle = \langle \lambda v, \lambda v \rangle = \bar{\lambda} \lambda \langle v, v \rangle, \quad (8.88)$$

also $|\lambda|^2 = 1$. Analog können wir orthogonale Abbildungen auf reellen Vektorräumen betrachten mit dem gleichen Resultat. Wir halten fest:

Satz 8.6.

- *Eigenwerte λ unitärer Abbildungen erfüllen $|\lambda|^2 = 1$.*
- *(Reelle) Eigenwerte λ orthogonaler Abbildungen erfüllen $\lambda^2 = 1$, also $\lambda = 1$ oder $\lambda = -1$.*

Sind orthogonale bzw. unitäre Abbildungen auf endlich-dimensionalen Vektorräumen diagonalisierbar? Für orthogonale Abbildungen wissen wir, dass das im Allgemeinen nicht möglich ist: Eine Drehung im \mathbb{R}^2 hat keine invariante Richtung, wenn es nicht die triviale Drehung oder die Drehung um 180 Grad ist.

Unitäre Abbildungen lassen sich dagegen immer diagonalisieren und zwar sogar mit Hilfe einer unitären Basistransformation.

wobei R_1, \dots, R_s orthogonale 2×2 -Matrizen sind.

Diesen Satz werden wir hier nicht beweisen. Das Auftreten der 2×2 -Blöcke hängt damit zusammen, dass im Reellen ein Polynom nicht notwendigerweise in Linearfaktoren zerfällt, sondern auch quadratische Faktoren auftreten.

8.4 Symmetrische und hermitesche Abbildungen

Eine wichtige Klasse von Matrizen, die wir kennengelernt haben, ist die der hermiteschen Matrizen ($A = A^*$), bzw. im reellen Fall die der symmetrischen. Sie spielen in der Physik eine ebenso große Rolle wie die orthogonalen und unitären Matrizen.

Es gibt sogar einen Zusammenhang: Wenn A eine hermitesche Matrix ist, dann ist $U = e^{iA}$ unitär:

$$U^* = (e^{iA})^* = e^{-iA^*} = e^{-iA} = U^{-1}. \quad (8.92)$$

Das wollen wir an dieser Stelle nicht vertiefen, sondern uns stattdessen zunächst anschauen, wie man symmetrische bzw. hermitesche Abbildungen charakterisiert, wenn man ein allgemeines Skalarprodukt betrachtet.

Definition 8.6. *Seien V ein reeller bzw. komplexer Vektorraum mit Skalarprodukt und $F : V \rightarrow V$ ein Endomorphismus. Dann heißt F **symmetrisch** bzw. **hermitesch** genau dann, wenn für alle $v, w \in V$ gilt*

$$\langle F(v), w \rangle = \langle v, F(w) \rangle. \quad (8.93)$$

Bemerkung

Man beachte den Unterschied aber auch die Gemeinsamkeiten mit den orthogonalen bzw. unitären Abbildungen. In beiden Fällen betrachtet man lineare Abbildungen, die mit der durch das Skalarprodukt gegebenen Struktur in gewisser Weise kompatibel sind, in einem Fall ist das Skalarprodukt unter der Abbildung erhalten, im anderen Fall erfüllt die Abbildung eine bestimmte Symmetrieeigenschaft bezüglich des Skalarprodukts. Das setzt einen Prozess fort, den wir schon vorher beobachtet haben und der typisch für mathematische Begriffsbildungen ist: Man führt eine Struktur ein, z.B. die Vektorraumstruktur auf einer Menge V , und betrachtet dann Abbildungen, die mit dieser Struktur auf bestimmte Weise verträglich sind. Im Vektorraumbeispiel führt das auf die linearen Abbildungen, und im Beispiel des Skalarprodukts führt dies in natürlicher Weise auf den Begriff der unitären/orthogonalen und hermiteschen/symmetrischen Abbildungen.

In einer Orthonormalbasis \mathcal{A} in einem endlich-dimensionalen Vektorraum V wird eine symmetrische bzw. hermitesche Abbildung F durch eine symmetrische bzw. hermitesche Matrix $A = M_{\mathcal{A}}(F)$ dargestellt.

Wir wollen nun die Eigenschaften dieser Abbildungen bzw. Matrizen besser verstehen, und insbesondere die Frage nach der Existenz von Eigenwerten untersuchen. Zunächst stellen wir fest, dass alle Eigenwerte einer hermiteschen Abbildung reell sind.

Satz 8.11. *Seien V ein unitärer Vektorraum und $F : V \rightarrow V$ eine hermitesche Abbildung, sowie λ ein Eigenwert von F . Dann ist λ reell.*

Beweis. Sei $v \in V \setminus \{0\}$ mit $F(v) = \lambda \cdot v$. Dann ist

$$\langle F(v), v \rangle = \langle v, F(v) \rangle \tag{8.94}$$

$$\iff \langle \lambda \cdot v, v \rangle = \langle v, \lambda \cdot v \rangle \tag{8.95}$$

$$\iff \bar{\lambda} \langle v, v \rangle = \lambda \langle v, v \rangle. \tag{8.96}$$

Da $\langle v, v \rangle > 0$ ist, folgt $\lambda = \bar{\lambda}$. □

Eine weitere, ebenso einfach zu beweisende Aussage ist, dass Eigenvektoren zu verschiedenen Eigenwerten orthogonal sind:

Satz 8.12. *Sei V ein unitärer Vektorraum. Seien $F : V \rightarrow V$ eine hermitesche Abbildung und $v_1, v_2 \in V \setminus \{0\}$ mit*

$$F(v_1) = \lambda_1 \cdot v_1 \quad , \quad F(v_2) = \lambda_2 \cdot v_2 \quad , \quad \text{wobei } \lambda_1 \neq \lambda_2. \tag{8.97}$$

Dann ist $\langle v_1, v_2 \rangle = 0$.

Beweis.

$$\langle F(v_1), v_2 \rangle = \langle v_1, F(v_2) \rangle \tag{8.98}$$

$$\iff \langle \lambda_1 \cdot v_1, v_2 \rangle = \langle v_1, \lambda_2 \cdot v_2 \rangle \tag{8.99}$$

$$\iff \lambda_1 \langle v_1, v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle \quad (\lambda_1 \text{ ist reell}). \tag{8.100}$$

Wegen $\lambda_1 \neq \lambda_2$ ist $\langle v_1, v_2 \rangle = 0$. □

Satz und Beweis gelten analog auch für symmetrische Abbildungen auf Euklidischen Vektorräumen.

Wir kommen nun zum wichtigsten Satz dieses Abschnitts:

Satz 8.13. *Sei $F : V \rightarrow V$ eine hermitesche Abbildung auf einem endlich-dimensionalen, unitären Vektorraum V . Dann gibt es eine Orthonormalbasis von V , die aus Eigenvektoren von F besteht.*

Beweis. Das komplexe charakteristische Polynom $P_F(t)$ zerfällt wegen des Fundamentalsatzes der Algebra in Linearfaktoren, somit gibt es eine Nullstelle und damit einen Eigenwert λ von F mit Eigenvektor $v \in V$ (es sei denn, V ist 0-dimensional, $V = \{0\}$, aber dann gibt es nichts zu beweisen). Wir setzen

$$v_1 = \frac{1}{\|v\|} v \tag{8.101}$$

und haben damit den ersten Vektor der zu konstruierenden Orthonormalbasis gefunden.

Wenn V 1-dimensional ist, sind wir schon fertig. Ansonsten betrachten wir den zu v_1 orthogonalen Untervektorraum

$$W_1 = \{w \in V \mid \langle v_1, w \rangle = 0\}. \quad (8.102)$$

Die wesentliche Beobachtung ist nun, dass W_1 von F invariant gelassen wird: Sei $w \in W_1$. Dann ist

$$\langle v_1, F(w) \rangle = \langle F(v_1), w \rangle = \langle \lambda \cdot v_1, w \rangle = \lambda \langle v_1, w \rangle, \quad (8.103)$$

also ist $F(w)$ orthogonal zu v_1 und somit $F(w) \in W_1$. Daher können wir F auf den Raum W_1 einschränken und definieren

$$\begin{aligned} F_1 : W_1 &\longrightarrow W_1 \\ w &\longmapsto F(w). \end{aligned} \quad (8.104)$$

F_1 definiert wieder eine hermitesche Abbildung auf W_1 (das Skalarprodukt können wir natürlich auch auf W_1 einschränken).

Jetzt können wir das Spiel von vorne beginnen: F_1 besitzt wieder mindestens einen Eigenwert λ_2 mit normiertem Eigenvektor $v_2 \in W_1 \subset V$. v_2 ist dann auch Eigenvektor von F . Wir bilden den zu v_2 orthogonalen Untervektorraum $W_2 \subset W_1$; F_1 lässt sich wieder darauf einschränken zu einer hermiteschen Abbildung F_2 usw. Bei einem endlich-dimensionalen Vektorraum ist man irgendwann fertig: (v_1, \dots, v_n) ist dann die gewünschte Basis. □

Auf Matrizen bezogen lautet der Satz:

Satz 8.14. *Sei A eine hermitesche Matrix. Dann gibt es eine unitäre Matrix S , sodass SAS^{-1} eine Diagonalmatrix ist.*

Die Diagonalelemente sind dann natürlich die Eigenwerte. Wir sagen auch: Eine hermitesche Matrix ist **unitär diagonalisierbar**.

Wir können analoge Aussagen für symmetrische Abbildungen auf Euklidischen Vektorräumen bzw. für symmetrische Matrizen machen. Die einzige Stelle, an der wir explizit ausgenutzt haben, dass wir über den komplexen Zahlen gerechnet haben, war die Feststellung, dass die Abbildung überhaupt einen Eigenwert besitzt. Das ist über \mathbb{R} nicht immer richtig, da reelle Polynome nicht unbedingt eine reelle Nullstelle haben.

Für symmetrische Matrizen ist dies aber immer der Fall: Eine reelle symmetrische Matrix A können wir als hermitesche Matrix mit reellen Einträgen ansehen. Also besitzt A auf jeden Fall einen komplexen Eigenwert, der aber wegen der Hermitizität von A reell sein muss: Das charakteristische Polynom von A besitzt also eine reelle Nullstelle, und somit besitzt A auch als Abbildung auf \mathbb{R}^n einen reellen Eigenwert. Der Beweis lässt sich daher direkt übertragen, und wir finden:

Satz 8.15.

1. Sei $F : V \rightarrow V$ eine symmetrische Abbildung auf einem endlich-dimensionalen reellen Vektorraum V mit Skalarprodukt. Dann gibt es eine Orthonormalbasis von V bestehend aus Eigenvektoren.
2. Sei A eine reelle symmetrische Matrix. Dann gibt es eine orthogonale Matrix S , sodass $S A S^{-1}$ eine Diagonalmatrix ist. (Wir sagen auch: Eine symmetrische Matrix ist **orthogonal diagonalisierbar**.)

Die eben formulierten Sätze sind die wichtigsten Feststellungen, die wir über symmetrische und hermitesche Abbildungen und Matrizen treffen werden. Diese (be-)merkwürdigen Aussagen sollten Ihnen sofort (neben der Definition) einfallen, wenn Sie an solche Abbildungen denken.

Sollte Ihnen also in Ihrem späteren Leben eine symmetrische oder hermitesche Matrix über den Weg laufen (und das wird der Fall sein, wenn Sie bei der Physik bleiben), denken Sie bitte sofort: „DIAGONALISIERBAR!“

So wichtig die Sätze auch sind, werden sie natürlich erst richtig nützlich, wenn wir eine gegebene Matrix auch unitär (bzw. orthogonal) diagonalisieren können. Dafür haben wir aber bereits alle Hilfsmittel beisammen, und wir wollen uns jetzt klarmachen, wie wir diese Diagonalisierung konkret durchführen.

Verfahren zur unitären Diagonalisierung einer hermiteschen Matrix

Sei A eine hermitesche $n \times n$ -Matrix. Das Verfahren ist sehr ähnlich zum bereits bekannten Diagonalisierungsverfahren:

1. Bilde das charakteristische Polynom $P_A(t)$ und bestimme alle Nullstellen λ_i . Dies sind die Eigenwerte von A . (Alle Eigenwerte werden reell sein!)
2. Bestimme zu jeder Nullstelle λ_i den Eigenraum

$$\text{Eig}(A, \lambda_i) = \text{Ker}(A - \lambda_i E_n) \quad (8.105)$$

durch Lösen des homogenen linearen Gleichungssystems und bestimme eine Basis. Mit Hilfe des Orthonormalisierungsverfahrens finde eine Orthonormalbasis zu jedem Eigenraum.

3. Alle Orthonormalbasen zusammen ergeben eine Orthonormalbasis (v_1, \dots, v_n) von V . Setze

$$S^* = \underbrace{(v_1 \cdots v_n)}_{\text{Spaltenvektoren}}, \quad \text{also} \quad S = \begin{pmatrix} \overline{v_1^t} \\ \vdots \\ \overline{v_n^t} \end{pmatrix}. \quad (8.106)$$

Dann ist S der gesuchte Basiswechsel, d.h. $S A S^*$ ist diagonal.

Die Abweichungen vom einfachen Diagonalisierungsverfahren sind also:

- Wir müssen in jedem Eigenraum das Orthonormalisierungsverfahren durchführen,
- dafür ersparen wir uns das Invertieren im letzten Schritt, da die Matrix S — wenn wir alles richtig gemacht haben — unitär ist und wir statt zu invertieren einfach die adjungierte Matrix bilden können.

Bemerkung

Manchmal gibt es Verwirrungen, ob die Eigenvektoren jetzt in S eingetragen werden oder in S^* bzw. S^{-1} , da in manchen Büchern auch die Konvention verwendet wird, dass die Diagonalmatrix als $S^* A S$ gebildet wird. Die Merkregel ist einfach: Die Matrix A will auf die Eigenvektoren wirken, d.h. die Eigenvektoren stehen als Spaltenvektoren in der Matrix rechts von A .

Das Verfahren funktioniert analog im Reellen, wie wir im folgenden Beispiel sehen:

Beispiel 8.7. Sei

$$A = \begin{pmatrix} 4 & 3 \\ 3 & -4 \end{pmatrix}. \quad (8.107)$$

Dann ist das charakteristische Polynom

$$P_A(t) = (4-t)(-4-t) - 9 \quad (8.108)$$

$$= t^2 - 25 \quad (8.109)$$

mit Nullstellen $\lambda_1 = 5$ und $\lambda_2 = -5$. Dann ist

$$\text{Eig}(A, 5) = \text{Ker} \begin{pmatrix} -1 & 3 \\ 3 & -9 \end{pmatrix} = \mathbb{R} \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \text{ setze } v_1 = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 \\ 1 \end{pmatrix} \quad (8.110)$$

$$\text{Eig}(A, -5) = \text{Ker} \begin{pmatrix} 9 & 3 \\ 3 & 1 \end{pmatrix} = \mathbb{R} \cdot \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \text{ setze } v_2 = \frac{1}{\sqrt{10}} \begin{pmatrix} -1 \\ 3 \end{pmatrix}. \quad (8.111)$$

Die normierten Eigenvektoren werden Einträge der Matrix S^t , d.h.

$$S^t = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} \implies S = \frac{1}{\sqrt{10}} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix}. \quad (8.112)$$

Zum Schluss können wir überprüfen, dass wir alles richtig gemacht haben:

$$S A S^t = \frac{1}{\sqrt{10}} S \begin{pmatrix} 4 & 3 \\ 3 & -4 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 3 \end{pmatrix} \quad (8.113)$$

$$= \frac{1}{10} \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 15 & 5 \\ 5 & -15 \end{pmatrix} \quad (8.114)$$

$$= \begin{pmatrix} 5 & 0 \\ 0 & -5 \end{pmatrix}. \quad (8.115)$$

Mit diesen Diagonalisierungsmethoden sind wir übrigens auch in der Lage, den noch ausstehenden Teil des Beweises des Sylvesterschen Trägheitssatzes zu führen:

Beweis des Existenzteils des Sylvesterschen Trägheitssatzes. Sei A eine hermitesche $n \times n$ -Matrix und S die unitäre Matrix, die sie diagonalisiert. Die diagonale Matrix $D = S A S^*$ hat reelle Diagonalelemente, und wir nennen die Zahl der positiven Elemente r_1 , die Zahl der negativen Elemente r_2 und die Zahl der Nullen $r_3 = n - r_1 - r_2$. Wir nehmen an, dass wir die Diagonalisierung so durchgeführt haben, dass die Diagonalelemente so angeordnet sind, dass zuerst die r_1 positiven Eigenwerte $\lambda_1, \dots, \lambda_{r_1}$ stehen, dann die r_2 negativen μ_1, \dots, μ_{r_2} und zum Schluss die r_3 Nullen (das können wir bei der Diagonalisierung leicht erreichen, indem wir die Eigenvektoren entsprechend anordnen,

Satz 8.16. *Seien $F : V \rightarrow V$, $G : V \rightarrow V$ hermitesche Abbildungen auf dem endlich-dimensionalen unitären Vektorraum V . Dann gibt es eine Orthonormalbasis von V , deren Elemente sowohl Eigenvektoren von F als auch von G sind, genau dann, wenn*

$$F \circ G = G \circ F. \quad (8.119)$$

Beweis. Einen ähnlichen Satz haben wir schon bewiesen: dass diagonalisierbare Endomorphismen simultan diagonalisierbar sind genau dann, wenn sie miteinander vertauschen. Hier müssen wir also nur noch zeigen, dass, wenn $F \circ G = G \circ F$, es nicht nur eine Basis aus gemeinsamen Eigenvektoren gibt, sondern sogar eine Orthonormalbasis. Wie wir schon früher gesehen hatten, lässt F die Eigenräume von G invariant,

$$F(\text{Eig}(G, \lambda)) \subset \text{Eig}(G, \lambda) \quad \text{wenn } F \circ G = G \circ F. \quad (8.120)$$

Also lässt sich F auf jeden Eigenraum von G einschränken und definiert dort eine hermitesche Abbildung, sodass es zu jedem Eigenwert λ von G eine Orthonormalbasis von $\text{Eig}(G, \lambda)$ gibt, die aus Eigenvektoren von F besteht. Alle diese Basen zusammen ergeben die gewünschte Basis von V . \square

Wir haben früher behauptet, dass auch unitäre Matrizen selbst unitär diagonalisierbar sind. Wir müssen den Beweis dazu noch nachtragen, wollen uns aber gleich die umfassendere Frage stellen, welche Matrizen sonst noch diese Eigenschaft haben.

Sei A unitär diagonalisierbar, es gibt also eine unitäre Matrix S , sodass $S A S^* = D$ eine Diagonalmatrix ist. Dann gilt $D D^* = D^* D$ und daher auch

$$A A^* = (S^* D S)(S^* D S)^* \quad (8.121)$$

$$= S^* D D^* S \quad (8.122)$$

$$= S^* D^* D S \quad (8.123)$$

$$= (S^* D S)^*(S^* D S) \quad (8.124)$$

$$= A^* A. \quad (8.125)$$

Matrizen mit dieser Eigenschaft heißen normal:

Definition 8.7. *Eine komplexe $n \times n$ -Matrix A mit $A A^* = A^* A$ heißt **normale Matrix**.*

Wir kennen schon Beispiele für normale Matrizen: hermitesche Matrizen sind offensichtlich normal, und unitäre Matrizen auch, denn $U U^* = E_n = U^* U$.

Wenn A unitär diagonalisierbar ist, dann gilt, wie wir gerade gesehen haben, $A A^* = A^* A$, A ist also normal. Die Umkehrung gilt auch:

Satz 8.17. *Sei A eine komplexe $n \times n$ -Matrix. Dann ist A unitär diagonalisierbar genau dann, wenn A normal ist.*

Beweis. Wenn A normal ist, dann sind

$$A_+ := A + A^* \quad \text{und} \quad A_- := i(A - A^*) \quad (8.126)$$

zwei hermitesche Matrizen, die miteinander vertauschen: $A_+ A_- = A_- A_+$. Also sind A_+ und A_- simultan unitär diagonalisierbar, und somit ist auch $A = \frac{1}{2}(A_+ - i A_-)$ unitär diagonalisierbar. \square

Die Diagonalisierung einer hermiteschen oder symmetrischen Matrix nennt man auch **Hauptachsentransformation**. Um den geometrischen Ursprung dieser Bezeichnung zu verstehen, wenden wir uns im folgenden Abschnitt zunächst den im zweiten Kapitel schon einmal angesprochenen Quadriken zu.

8.5 Anwendungen: Quadriken und Matrixzerlegungen

Wir wollen in diesem Abschnitt zwei Themen ansprechen, bei denen wir die gewonnenen Erkenntnisse verwenden können. Zunächst wenden wir uns der Klassifikation der Quadriken zu.

Eine Quadrik im \mathbb{R}^n (man könnte das Ganze auch im Komplexen betrachten, aber wir beschränken uns auf den reellen Fall) ist eine Teilmenge, die durch eine quadratische Gleichung in den Koordinaten definiert ist,

$$Q = \left\{ x \in \mathbb{R}^n \mid \sum_{i,j=1}^n a_{ij} x_i x_j + \sum_{i=1}^n b_i x_i + c = 0 \right\}. \quad (8.127)$$

Beispiele sind die Kreislinie $\{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 - 1 = 0\}$, Ellipsen, Parabeln, Hyperbeln, etc.

Wir können nun unser ganzes erlerntes Wissen anwenden, um die Quadrik auf eine möglichst einfache Form zu bringen. Dabei erlauben wir alle Koordinatentransformationen, die die Abstände nicht verändern, also Verschiebungen um einen konstanten Vektor und orthogonale Abbildungen.

Zunächst schreiben wir die Quadrik in einer für die lineare Algebra angemessenen Weise,

$$Q = \{x \in \mathbb{R}^n \mid x^t A x + b^t x + c = 0\}, \quad (8.128)$$

wobei $A = (a_{ij})$ und $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$. Wir können A als symmetrisch annehmen: In der Summe $\sum_{i,j=1}^n a_{ij} x_i x_j$ kommt für $i \neq j$ der Term $x_i x_j$ mit dem Koeffizienten $(a_{ij} + a_{ji})$ vor. Da a_{ij} und a_{ji} sonst nicht in der Gleichung auftauchen, können wir sie immer gleich wählen, $a_{ij} = a_{ji}$.

Eine symmetrische Matrix? Die können wir diagonalisieren! Es gibt also eine orthogonale Matrix S , sodass $S A S^t = D$ diagonal ist. Wir führen neue Koordinaten $x' = S x$ ein, also $x = S^t x'$, dann wird die Quadrik in den neuen Koordinaten zu

$$Q' = \{x' \in \mathbb{R}^n \mid x'^t D x' + b'^t x' + c = 0\}, \quad (8.129)$$

wobei wir $b' = S b$ eingeführt haben. Damit ist die Quadrik schon deutlich einfacher beschrieben, da keine gemischten Terme $x'_i x'_j$ mit $i \neq j$ auftauchen.

Wir können durch eine Verschiebung des Ursprungs die Gleichung womöglich weiter vereinfachen. Seien $x'' = x' + d$ die Koordinaten nach einer Verschiebung um den Vektor d . Die Quadrikgleichung lautet dann

$$(x'' - d)^t D (x'' - d) + b'^t (x'' - d) + c = 0 \quad (8.130)$$

$$\iff x''^t D x'' - d^t D x'' - \underbrace{x''^t D d}_{=d^t D x''} + d^t D d + b'^t x'' - b'^t d + c = 0 \quad (8.131)$$

$$\iff x''^t D x'' + (b' - 2 D d)^t x'' + (c - b'^t d + d^t D d) = 0. \quad (8.132)$$

Falls D keine Null auf der Diagonalen hat, können wir den linearen Term durch Wahl von $d = \frac{1}{2}D^{-1}b'$ eliminieren und erhalten mit $c'' := c - b'^t d + d^t D d$

$$Q'' = \{x'' \in \mathbb{R}^n \mid x''^t D x'' + c'' = 0\}. \quad (8.133)$$

Für $D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ ist dies explizit

$$Q'' = \{x'' \in \mathbb{R}^n \mid \lambda_1 x_1^2 + \dots + \lambda_n x_n^2 + c'' = 0\}. \quad (8.134)$$

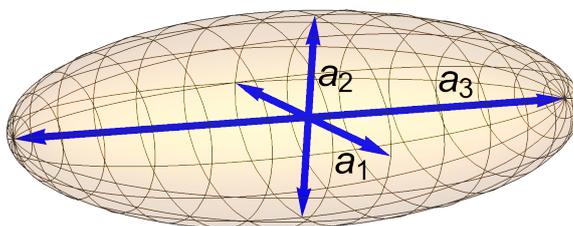
Sind alle λ_i positiv und $c'' < 0$, so erhält man eine höher-dimensionale Verallgemeinerung einer Ellipse, und man kann die Längen der Halbachsen sofort ablesen:

$$a_1 = \sqrt{\frac{c''}{\lambda_1}}, \dots, a_n = \sqrt{\frac{c''}{\lambda_n}}, \quad (8.135)$$

sodass die Gleichung für Q'' lautet:

$$\left(\frac{x_1''}{a_1}\right)^2 + \dots + \left(\frac{x_n''}{a_n}\right)^2 = 1. \quad (8.136)$$

Die Koordinaten x_i'' liegen entlang der Hauptachsen des n -dimensionalen Ellipsoids, die sich aus den Eigenvektoren der Matrix A ergeben haben.



Treten positive und negative Eigenwerte auf, so erhält man verschiedenartige Verallgemeinerungen von Hyperbeln.

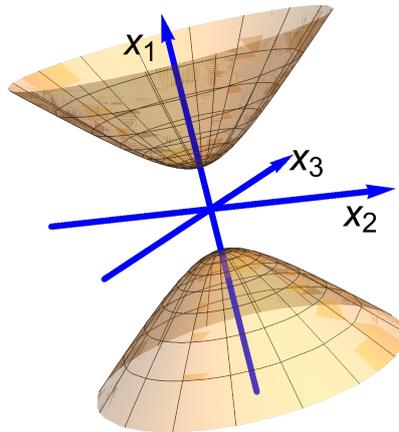
Beispiel 8.8.

$$Q = \{x \in \mathbb{R}^3 \mid x_1^2 - x_2^2 - x_3^2 = 1\} \quad (8.137)$$

Zu jedem $x_1 \geq 1$ und jedem $x_1 \leq -1$ liegen x_2, x_3 auf einer Kreislinie,

$$x_2^2 + x_3^2 = x_1^2 - 1. \quad (8.138)$$

Die Quadrik zerfällt also in zwei Hyperboloide:

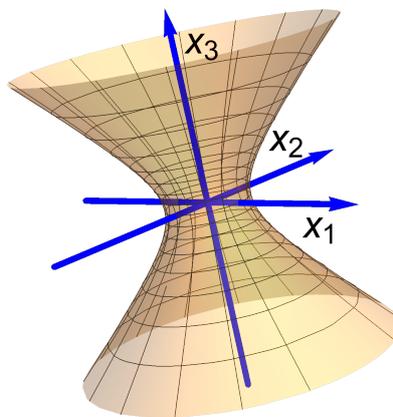


Beispiel 8.9.

$$Q = \{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 - x_3^2 = 1\} \quad (8.139)$$

Zu jedem $x_3 \in \mathbb{R}$ beschreiben die Koordinaten x_1, x_2 eine Kreislinie,

$$x_1^2 + x_2^2 = 1 + x_3^2. \quad (8.140)$$



Treten Nullen als Eigenwerte auf, so lässt sich der letzte Schritt nur teilweise ausführen.

Sei

$$D = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}. \quad (8.141)$$

Dann wählen wir

$$d = \frac{1}{2} \begin{pmatrix} \lambda_1^{-1} b'_1 \\ \vdots \\ \lambda_r^{-1} b'_r \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (8.142)$$

und erhalten

$$Q'' = \{x'' \in \mathbb{R}^n \mid \sum_{i=1}^r \lambda_i x_i''^2 + \sum_{i=r+1}^n b'_i x_i'' + c'' = 0\}. \quad (8.143)$$

Falls $b' = 0$, dann hängt die Gleichung nur noch von den x_i'' für $1 \leq i \leq r$ ab: Man erhält eine ellipsoid- oder hyperboloidartige Quadrik in einer niedrigeren Dimension, die trivial in den weiteren $n - r$ Dimensionen fortgesetzt ist — z.B. einen Zylinder, der ein in der dritten Dimension fortgesetzter Kreis ist,

$$\{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 = 1\}. \quad (8.144)$$

Wenn b' nicht verschwindet, können wir durch Drehung in den letzten $n - r$ Koordinaten und einer Verschiebung immer erreichen, dass in den neuen Koordinaten (die wir weiterhin x'' nennen wollen) die Quadrik die Form hat

$$\{x'' \in \mathbb{R}^n \mid \sum_{i=1}^r \lambda_i x_i''^2 + \|b'\| x_{r+1}'' = 0\}. \quad (8.145)$$

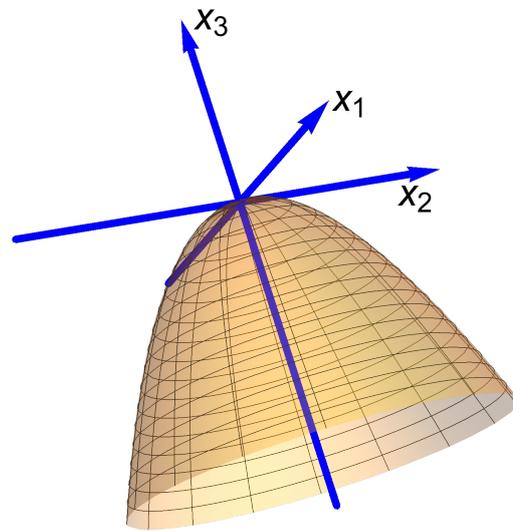
Dies ist eine höher-dimensionale Verallgemeinerung einer Parabel.

Beispiel 8.10.

$$Q = x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 + x_3 = 0 \quad (8.146)$$

Für jedes $x_3 \leq 0$ beschreiben x_1, x_2 eine Kreislinie,

$$x_1^2 + x_2^2 = -x_3. \quad (8.147)$$



Polar- und Singulärwertzerlegung

Zum Abschluss dieses Kapitels wollen wir noch auf zwei Typen von Zerlegungen von Matrizen eingehen, die sehr nützlich sind. Eine ist die **Polarzerlegung**. Die Idee ist, die Zerlegung einer komplexen Zahl z als Produkt einer nicht-negativen Zahl $r > 0$ und einer Phase $e^{i\alpha}$ als $z = r e^{i\alpha}$ für Matrizen nachzuahmen.

Satz 8.18. *Sei A eine komplexe $n \times n$ -Matrix. Dann gibt es eine positive hermitesche Matrix B und eine unitäre Matrix U , sodass $A = UB$.*

Positivität von B bedeutet dabei, dass für alle $x \in \mathbb{C}^n$ gilt: $\bar{x}^t B x \geq 0$. Berücksichtigt man, dass B in unserem Fall hermitesch ist und daher diagonalisiert werden kann, bedeutet dies, dass alle Eigenwerte positiv (oder Null) sind.

Die positive Matrix B spielt die Rolle des Absolutbetrags. In \mathbb{C} bilden wir den Betrag als $r = \sqrt{z \bar{z}}$. Für die Matrizen versuchen wir etwas Ähnliches:

$$B = \sqrt{A^* A}. \tag{8.148}$$

Was bedeutet das? $A^* A$ ist hermitesch ($(A^* A)^* = A^* A$), kann also diagonalisiert werden. Andererseits ist $A^* A$ positiv:

$$\bar{x}^t A^* A x = (Ax)^*(Ax) \geq 0. \tag{8.149}$$

Also hat $A^* A$ nur positive Eigenwerte (oder Nullen), es gibt also eine unitäre Matrix S mit

$$A^* A = S^{-1} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} S, \quad \lambda_i \geq 0. \tag{8.150}$$

Wir definieren die Wurzel von $A^* A$, indem wir die Wurzeln der Eigenwerte nehmen:

$$\sqrt{A^* A} := S^{-1} \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} S \tag{8.151}$$

und setzen $B = \sqrt{A^* A}$. Dann ist

$$B^2 = B B = S^{-1} \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} S S^{-1} \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} S \tag{8.152}$$

$$= S^{-1} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} S = A^* A, \tag{8.153}$$

d.h. wir können B wirklich als Quadratwurzel von A^*A auffassen.

Die Behauptung ist nun, dass es ein unitäres U gibt mit

$$A = U B. \quad (8.154)$$

Wir zeigen das hier nur für invertierbare Matrizen A . B ist dann auch invertierbar und wir setzen

$$U := A B^{-1}. \quad (8.155)$$

Jetzt überprüfen wir, dass U unitär ist:

$$U^*U = (A B^{-1})^* A B^{-1} \quad (8.156)$$

$$= (B^{-1})^* A^* A B^{-1} \quad (8.157)$$

$$= B^{-1} B^2 B^{-1} \quad (B \text{ und damit } B^{-1} \text{ ist hermitesch}) \quad (8.158)$$

$$= E_n. \quad (8.159)$$

Eine weitere wichtige Produktzerlegung ist die **Singulärwertzerlegung**:

Satz 8.19. *Sei A eine komplexe $m \times n$ -Matrix. Dann lässt sich A schreiben als*

$$A = U_1 D U_2 \quad (8.160)$$

wobei

- D eine $m \times n$ -Diagonalmatrix mit nichtnegativen Einträgen,
- U_1 eine unitäre $m \times m$ -Matrix und
- U_2 eine unitäre $n \times n$ -Matrix ist.

Beweis nur für $m = n$. Eine Polarzerlegung von A liefert $A = U B$. B ist eine positive hermitesche Matrix, also gibt es ein unitäres U_2 mit

$$U_2 B U_2^{-1} = D, \quad (8.161)$$

wobei die Diagonalmatrix D nur nichtnegative Einträge hat. Also ist

$$A = U U_2^{-1} D U_2 \quad (8.162)$$

und mit $U_1 = U U_2^{-1}$ folgt die Behauptung. □

Die Singulärwertzerlegung findet in der Numerik und den Computerwissenschaften Anwendung, z.B. bei der Datenkompression/-reduktion (aber auch in der Quantenphysik ...).

Bemerkung

Als Ergänzung sei hier für die Interessierten noch der Beweis der Existenz der Singulärwertzerlegung für den allgemeinen Fall gegeben: Sei A also eine $m \times n$ -Matrix. Dann ist A^*A eine positive hermitesche $n \times n$ -Matrix, es gibt also eine unitäre Matrix U_2 mit

$$U_2 A^* A U_2^{-1} = D_2, \quad (8.163)$$

wobei die Diagonalmatrix D_2 nur nicht-negative Einträge $\lambda_1, \dots, \lambda_n$ besitzt. Wir nehmen an, dass die Einträge so geordnet sind, dass $\lambda_1, \dots, \lambda_r$ positiv sind und $\lambda_{r+1} = \dots = \lambda_n = 0$. Die Spaltenvektoren von $U_2^{-1} = (v_1 \cdots v_n)$ sind Eigenvektoren von A^*A .

Wir betrachten nun die Vektoren $A v_i$. Bezüglich des kanonischen Skalarprodukts auf \mathbb{C}^n erfüllen sie

$$\langle A v_i, A v_j \rangle = \bar{v}_i^t A^* A v_j \quad (8.164)$$

$$= \lambda_j \bar{v}_i^t v_j \quad (8.165)$$

$$= \lambda_j \delta_{ij}, \quad (8.166)$$

wobei wir im letzten Schritt ausgenutzt haben, dass (v_1, \dots, v_n) eine Orthonormalbasis ist. Für $i > r$ ist $A v_i = 0$, für $i = 1, \dots, r$ setzen wir

$$w_i := \frac{1}{\sqrt{\lambda_i}} A v_i \quad (8.167)$$

und erhalten ein Orthonormalsystem (w_1, \dots, w_r) . Wenn $r < m$, ergänzen wir es zu einer Orthonormalbasis $(w_1, \dots, w_r, w_{r+1}, \dots, w_m)$ in \mathbb{C}^m . Dann setzen wir

$$U_1 = (w_1 \ \cdots \ w_m) \quad \text{und} \quad D = \begin{pmatrix} \sqrt{\lambda_1} & & & \\ & \ddots & & \\ & & \sqrt{\lambda_r} & \\ & & & 0 \end{pmatrix}, \quad (8.168)$$

wobei D eine $m \times n$ -Matrix ist, die bis auf die ersten r Diagonalelemente überall Null ist. Wir zeigen nun, dass $A U_2^{-1} = U_1 D$:

$$A U_2^{-1} = A (v_1 \ \cdots \ v_n) \quad (8.169)$$

$$= (A v_1 \ \cdots \ A v_n) \quad (8.170)$$

$$= (\sqrt{\lambda_1} w_1 \ \cdots \ \sqrt{\lambda_r} w_r \ 0 \ \cdots \ 0) \quad (8.171)$$

und

$$U_1 D = (w_1 \ \cdots \ w_m) \begin{pmatrix} \sqrt{\lambda_1} & & & \\ & \ddots & & \\ & & \sqrt{\lambda_r} & \\ & & & 0 \end{pmatrix} \quad (8.172)$$

$$= (\sqrt{\lambda_1} w_1 \ \cdots \ \sqrt{\lambda_r} w_r \ 0 \ \cdots \ 0). \quad (8.173)$$

Also ist $AU_2^{-1} = U_1D$, und nach Multiplikation mit U_2 von rechts erhalten wir wie gewünscht $A = U_1DU_2$.

Zusammenfassung Kapitel 8

- Wir haben den Begriff des Skalarprodukts auf allgemeinen Vektorräumen eingeführt, z.B. auf Funktioneneräumen.
- In Vektorräumen mit Skalarprodukt spielen Orthonormalbasen eine wichtige Rolle. Dank des Orthonormalisierungsverfahrens können wir auf jedem endlich-dimensionalen Vektorraum eine solche Basis angeben.
- Orthogonale bzw. unitäre Endomorphismen sind strukturerhaltenden Abbildungen: Sie erhalten das Skalarprodukt. In einer Orthonormalbasis werden sie durch orthogonale bzw. unitäre Matrizen dargestellt. Alle Eigenwerte solcher Matrizen haben Betrag 1, und Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal.
- Für hermitesche Abbildungen gilt:
 - sie haben nur reelle Eigenwerte,
 - Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal,
 - im endlich-dimensionalen Fall existiert eine Orthonormalbasis aus Eigenvektoren.
- Hermitesche Matrizen lassen sich also mit Hilfe eines unitären Basiswechsels diagonalisieren. Das Gleiche gilt für alle normalen Matrizen ($A^* A = A A^*$), insbesondere für unitäre.
- Symmetrische Matrizen lassen sich mit Hilfe eines orthogonalen Basiswechsels diagonalisieren. Mit dieser *Hauptachsentransformation* lassen sich z.B. Quadriken in eine einfache Form bringen.

9 Tensoren

In der Physik und Mathematik findet man oft unterschiedliche Herangehensweisen an den Begriff des Tensors. Ursprünglich kommt der Begriff aus der Physik — und da man in der Physik häufiger und lieber als in der Mathematik in konkreten Basen rechnet, steht in der Physik oft das Transformationsverhalten unter einem Basiswechsel im Vordergrund. Tensoren sind hier meist assoziiert mit dem physikalischen Ortsraum oder der Raumzeit.

Dieser etwas engere Tensorbegriff wird in der Mathematik auf Tensoren als Elemente allgemeiner Tensorprodukte ausgedehnt, welche wiederum in der Physik — insbesondere auch in der Quantenphysik — eine wichtige Rolle spielen.

Wir werden uns hier dem Begriff des Tensors nähern, indem wir zunächst, wie in der Physik üblich, das Transformationsverhalten betrachten; später stoßen wir auf die mathematische Definition des Tensorprodukts.

Im gesamten Kapitel beschränken wir uns auf endlich-dimensionale, reelle Vektorräume. Die Aussagen lassen sich direkt auf komplexe Vektorräume übertragen, die Übertragung auf den unendlich-dimensionalen Fall ist zwar teilweise auch möglich, hier muss man allerdings sehr vorsichtig sein.

9.1 Dualraum

Bevor wir uns den Tensoren zuwenden, wollen wir noch einmal den Begriff des Dualraums genauer betrachten. In gewisser Weise ist es intuitiver, Tensoren im Dualraum einzuführen, wie wir noch sehen werden.

Sei V ein endlich-dimensionaler, reeller Vektorraum. Den **Dualraum** V^* von V hatten wir eingeführt als den Raum aller linearen Abbildungen von V nach \mathbb{R} (lineare Funktionale oder Linearformen):

$$V^* = \text{Hom}(V, \mathbb{R}). \quad (9.1)$$

Als Physikerinnen und Physiker rechnen wir gerne in Koordinaten und führen daher mit Vorliebe irgend welche Basen ein. Da V^* mit V eng verwandt ist, stellt sich die Frage, ob die Einführung einer bestimmten Basis auf V vielleicht die Einführung einer bestimmten Basis auf V^* nahelegt.

Sei $\mathcal{A} = (v_1, \dots, v_n)$ eine Basis von V , und sei $F \in \text{Hom}(V, \mathbb{R})$ ein lineares Funktional auf V . F ist charakterisiert durch seine Wirkung auf die Basisvektoren v_i : Denn sei $v = \sum_{i=1}^n x^i v_i$ (warum wir den Index bei x^i hochstellen wird in Kürze erläutert) ein allgemeiner Vektor in V , dann ist

$$F(v) = F\left(\sum_{i=1}^n x^i v_i\right) = \sum_{i=1}^n x^i F(v_i) = \sum_{i=1}^n x^i f_i. \quad (9.2)$$

Hierbei ist $f_i = F(v_i) \in \mathbb{R}$ das Resultat von F bei Anwendung auf v_i . F ist durch Festlegung der f_i eindeutig charakterisiert. Wir können die f_i daher als Koordinaten von F in V^* ansehen, und haben somit eine Parametrisierung von V^* gefunden.

Bemerkung

Kurz als Erinnerung und zur Anknüpfung an das, was wir bereits früher diskutiert haben: Der Zeilenvektor $(f_1 \cdots f_n)$ ist genau die $1 \times n$ -Matrix $M_{\mathcal{A}}^{\{e_1\}}(F)$ der Matrixdarstellung von $F : V \rightarrow \mathbb{R}$ bezüglich der Basis \mathcal{A} von V und der Standardbasis $e_1 = 1$ von \mathbb{R} .

Wenn die f_i die Koordinaten von F sind, was ist dann die Basis? Die Basisvektoren v_*^i in V^* sind die linearen Funktionale, die in der Koordinatendarstellung die kanonischen Einheitsvektoren $(1 \ 0 \ \cdots \ 0)$, $(0 \ 1 \ 0 \ \cdots \ 0)$, usw. ergeben, also der erste Basisvektor

$$v_*^1 \text{ wirkt als } v_*^1(v_i) = \begin{cases} 1 & \text{für } i = 1 \\ 0 & \text{sonst,} \end{cases} \quad (9.3)$$

und allgemein

$$v_*^i(v_j) = \delta_j^i, \quad (9.4)$$

wobei δ_j^i das Kronecker-Symbol ist, nur dass wir die Indizes nicht beide nach unten geschrieben haben.

Definition 9.1. Sei V ein endlich-dimensionaler, reeller Vektorraum mit Basis $\mathcal{A} = (v_1, \dots, v_n)$. Seien $v_*^1, \dots, v_*^n \in V^*$ die linearen Funktionale, sodass für alle $i, j \in \{1, \dots, n\}$ gilt: $v_*^i(v_j) = \delta_j^i$. Dann heißt $\mathcal{A}^* = (v_*^1, \dots, v_*^n)$ die zu \mathcal{A} **duale Basis** von V^* .

Bemerkung

Es gibt also die Möglichkeit, jeder Basis \mathcal{A} von V in natürlicher Weise eine Basis \mathcal{A}^* von V^* zuzuordnen; es gibt aber für einen Vektorraum ohne weitere Struktur keine ausgezeichnete Abbildung, die einem Vektor aus V einen Vektor in V^* zuordnet. Da V und V^* die gleiche Dimension haben, gibt es natürlich Isomorphismen zwischen V und V^* , aber unter diesen lässt sich keiner (ohne Wahl einer weiteren Struktur) gegenüber den anderen auszeichnen.

Da wir nun gesehen haben, dass eine Basis von V die Wahl einer Basis von V^* nahelegt, können wir auch fragen, welche Auswirkungen ein Basiswechsel in V auf die duale Basis hat. Vorher wollen wir uns aber noch anschauen, ob wir weitere Strukturen finden, wenn wir den Dualraum vom Dualraum betrachten.

Satz 9.1. Sei V ein endlich-dimensionaler, reeller Vektorraum. Die Abbildung

$$\begin{aligned} \psi : V &\longrightarrow (V^*)^* \\ v &\longmapsto \psi_v \quad \text{wobei für } F \in V^* : \psi_v(F) = F(v) \end{aligned} \quad (9.5)$$

ist ein Isomorphismus.

Was macht die Abbildung ψ ? Sie ordnet jedem Vektor $v \in V$ in natürlicher Weise (ohne Bezugnahme auf die Wahl einer Basis oder irgend einer anderen Struktur) ein Element

$\psi_v \in V^{**}$ zu, indem sie den Spieß umkehrt: Ein $F \in V^*$ ordnet jedem Vektor $v \in V$ eine Zahl $F(v)$ zu — andersherum ordnet ein Vektor $v \in V$ jedem $F \in V^*$ auch eine Zahl $F(v)$ zu; der Vektor v kann also als Linearform auf V^* gedeutet werden. Zur Unterscheidung, ob wir v gerade als Element von V oder als Linearform auf V^* (also als Element von V^{**}) sehen wollen, haben wir die Notation $\psi_v \in V^{**}$ eingeführt.

Wenn man sich das klargemacht hat, ist auch der Beweis nicht schwierig:

Beweis. Man kann sich schnell davon überzeugen, dass ψ linear ist. Da V und V^* und damit auch V^{**} die gleiche Dimension haben, genügt es (wegen der Dimensionsformel) zu zeigen, dass ψ injektiv ist. Sei also $\psi_v = 0$, d.h. für alle $F \in V^*$ ist $\psi_v(F) = F(v) = 0$. Dann ist aber $v = 0$ (ansonsten könnte man $v_1 := v$ zu einer Basis (v_1, \dots, v_n) von V ergänzen und v_*^1 aus der dualen Basis würde $v_*^1(v_1) = 1 \neq 0$ erfüllen: ein Widerspruch). \square

Nun aber zum Basiswechsel. Wir wechseln von einer Basis \mathcal{A} zu einer Basis \mathcal{A}' von V , sodass die Koordinatenvektoren

$$x' = S x \tag{9.6}$$

erfüllen. Die duale Basis \mathcal{A}^* wird dann zu $(\mathcal{A}')^*$, dementsprechend transformieren die Koordinaten (f_i) zu Koordinaten (f'_i) . Sei $v \in V$ der Vektor mit Koordinaten (x^i) bzw. (x'^i) , und sei $F \in V^*$ die Linearform mit Koordinaten (f_i) bzw. (f'_i) . $F(v)$ können wir bezüglich beider Basiswahlen in Koordinaten ausdrücken, daher gilt

$$F(v) = \sum_{i=1}^n f_i x^i = \sum_{i=1}^n f'_i x'^i. \tag{9.7}$$

Interpretieren wir $f = (f_1 \cdots f_n)$ als Zeilenvektor, so wird daraus

$$F(v) = f x = f' x', \tag{9.8}$$

wobei rechts die übliche Multiplikation eines Zeilenvektors mit einem Spaltenvektor steht. Da $x' = S x$, folgt

$$\text{für alle } x \in \mathbb{R}^n : f x = f' S x \implies f = f' S \implies f' = f S^{-1}. \tag{9.9}$$

Die Koordinaten im Dualraum transformieren mit der inversen Matrix, sie verhalten sich also anders als die Koordinaten-Tupeln der Vektoren in V . Elemente von V^* nennt man auch **Kovektoren**. Zu einem geeigneten Vektorraum (z.B. dem Anschauungsraum \mathbb{R}^3) gibt es also zwei verschiedene Typen von n -Tupeln, die sich durch ihr Transformationsverhalten unterscheiden. Dies drückt man durch die Stellung der Indizes aus:

- (x^i) : hochgestellte Indizes für Koordinaten von Vektoren,
- (f_i) : tiefgestellte Indizes für Koordinaten von Kovektoren.

Wir können das auch pragmatischer zusammenfassen als:

- hochgestellte Indizes für Spaltenvektoren und

- tiefgestellte Indizes für Zeilenvektoren.

Die Transformation sieht dann wie folgt aus:

$$x'^i = \sum_{j=1}^n S^i_j x^j \quad (9.10)$$

$$f'_i = \sum_{j=1}^n f_j (S^{-1})^j_i. \quad (9.11)$$

Da bei Multiplikation von Matrizen bzw. Vektoren immer Zeilen mit Spalten verknüpft werden, treten bei dieser Notation in den Summen die Summationsindizes immer einmal unten und einmal oben auf.

Ein Beispiel für einen Kovektor kennen Sie vermutlich alle, auch wenn Sie ihn vielleicht noch nicht als solchen wahrgenommen haben: den Gradienten einer Funktion. Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine reellwertige differenzierbare Funktion auf $V = \mathbb{R}^n$. Der Einfachheit halber schauen wir uns den Gradienten nur an der Stelle 0 an. Wir nennen $df(0) \in V^*$ die Linearform, die einem Vektor $x \in \mathbb{R}^n$ die Richtungsableitung von f (an der Stelle 0) in Richtung x zuordnet:

$$\begin{aligned} df(0) : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \left. \frac{d}{dt} f(tx) \right|_{t=0}. \end{aligned} \quad (9.12)$$

Wir setzen für den Moment voraus, dass wir mit partiellen Ableitungen umgehen können. Dann ist

$$df(0)(x) = \left. \frac{d}{dt} f(tx) \right|_{t=0} = \sum_{i=1}^n x^i \frac{\partial f}{\partial x^i}(0). \quad (9.13)$$

Die Darstellung ist richtig, egal, welche Basis wir gewählt haben. Wählen wir im \mathbb{R}^n eine andere Basis mit Koordinaten $x' = Sx$, so ist die Funktion f in den neuen Koordinaten gegeben durch $f'(x') = f(S^{-1}x')$. Dementsprechend ist

$$\frac{\partial f'}{\partial x'^i}(0) = \sum_{j=1}^n \frac{\partial f}{\partial x^j}(0) (S^{-1})^j_i. \quad (9.14)$$

Das n -Tupel $\left(\frac{\partial f'}{\partial x'^i}(0)\right)$ transformiert also wie ein Kovektor, damit ist

$$\sum_{i=1}^n \frac{\partial f'}{\partial x'^i}(0) x'^i = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{\partial f}{\partial x^j}(0) \underbrace{(S^{-1})^j_i S^i_k}_{\delta_k^j \text{ nach Summation über } i} x^k \quad (9.15)$$

$$= \sum_{j=1}^n \frac{\partial f}{\partial x^j}(0) x^j. \quad (9.16)$$

An diesem Beispiel merken wir, dass es praktisch ist, eine **Summenkonvention** einzuführen: Tritt ein Index einmal oben und einmal unten auf, so wird über ihn summiert. (Dies ist eine Variante der *Einsteinschen Summenkonvention*, die manchmal *Riccische Summenkonvention* genannt wird.)

Warum wird der Gradient dann in der Physik oft wie ein Vektor behandelt? Das wollen wir uns genauer anschauen. Ein Kovektor mit Koordinatenzeilenvektor f transformiert zu $f' = f S^{-1}$. Als Spaltenvektor wird dies zu

$$(f')^t = (S^{-1})^t f^t. \quad (9.17)$$

Ist S eine orthogonale Matrix, so ist $(S^{-1})^t = S$ und f^t transformiert genau wie ein Vektor. Orthogonale Koordinatentransformationen treten auf, wenn wir Vektorräume mit Skalarprodukt und Transformationen von Orthonormalbasen betrachten. Im Anschauungsraum \mathbb{R}^3 beispielsweise haben wir das kanonische Skalarprodukt, und wir können den Gradienten durch einen Spaltenvektor darstellen, der sich unter Drehungen und Spiegelungen (also orthogonalen Transformationen) genauso verhält wie ein gewöhnlicher Vektor.

Wenn wir den Gradienten als Vektor darstellen, haben wir anscheinend den Dualraum V^* irgendwie mit V identifiziert. Es ist gerade das Skalarprodukt, das uns einen ausgezeichneten Isomorphismus $\sigma : V \rightarrow V^*$ liefert:

Satz 9.2. *Sei V ein endlich-dimensionaler, reeller Vektorraum mit Skalarprodukt $\langle \cdot, \cdot \rangle$. Dann ist*

$$\begin{aligned} \sigma : V &\longrightarrow V^* \\ v &\longmapsto \langle v, \cdot \rangle \end{aligned} \quad (9.18)$$

ein Isomorphismus.

Beweis. σ ist offenbar linear, und da V und V^* die gleiche Dimension haben, genügt es zu zeigen, dass σ injektiv ist, dass also der Kern von σ trivial ist. Sei $v \in V$ ein Vektor, der durch σ auf $0 \in V^*$ abgebildet wird, d.h.

$$\text{für alle } w \in V : \quad \langle v, w \rangle = 0. \quad (9.19)$$

Das muss auch für $w = v$ gelten, also $\langle v, v \rangle = 0$. Folglich muss $v = 0$ sein, also ist σ injektiv. \square

Da wir im Anschauungsraum \mathbb{R}^3 immer stillschweigend das kanonische Skalarprodukt verwenden, können wir Vektoren und Kovektoren munter identifizieren. In einer orthonormalen Basis geschieht die Identifikation der entsprechenden Koordinatentupel einfach durch Transposition von Zeilenvektoren zu Spaltenvektoren.

Trotzdem ist dieses Beispiel eines Objekts, das sich anders transformiert als ein Vektor, sehr lehrreich, wenn wir uns allgemeineren Tensoren zuwenden wollen.

9.2 Bilinearformen und Tensoren

Ausgehend von einem Vektorraum V mit Basis \mathcal{A} haben wir uns die Frage gestellt, wie wir in dem daraus abgeleiteten Raum $V^* = \text{Hom}(V, \mathbb{R})$ eine verwandte Basis \mathcal{A}^* bilden können. Wir betrachten nun einen weiteren abgeleiteten Vektorraum, nämlich den Raum der Bilinearformen

$$\text{Bil}(V \times V, \mathbb{R}) = \{s : V \times V \rightarrow \mathbb{R} \mid s \text{ linear in beiden Einträgen}\}. \quad (9.20)$$

Dies ist ein Vektorraum, da wir auf natürliche Weise Bilinearformen addieren und mit reellen Zahlen multiplizieren können.

Sei $\mathcal{A} = (v_1, \dots, v_n)$ eine Basis von V . Seien

$$v = x^i v_i \quad \text{und} \quad w = y^j v_j \quad (9.21)$$

zwei Vektoren in V und $s \in \text{Bil}(V \times V, \mathbb{R})$. Dann ist

$$s(v, w) = s(x^i v_i, y^j v_j) = x^i y^j s(v_i, v_j). \quad (9.22)$$

s ist somit durch die n^2 Zahlen $s_{ij} := s(v_i, v_j)$ festgelegt. Da andersherum jede Wahl von Zahlen s_{ij} eine Bilinearform definiert, können wir die s_{ij} als Koordinaten des n^2 -dimensionalen Vektorraums $\text{Bil}(V \times V, \mathbb{R})$ ansehen. Die entsprechende Basis wird gebildet von den Bilinearformen v^{ij} mit

$$v^{ij}(v_k, v_l) = \delta_k^i \delta_l^j. \quad (9.23)$$

Dann ist $s = s_{ij} v^{ij}$:

$$s(v, w) = s(x^k v_k, y^l v_l) \quad (9.24)$$

$$= s_{ij} v^{ij}(x^k v_k, y^l v_l) \quad (9.25)$$

$$= s_{ij} x^k y^l \delta_k^i \delta_l^j \quad (9.26)$$

$$= s_{ij} x^i y^j. \quad (9.27)$$

Die Wahl einer Basis \mathcal{A} von V legt also in natürlicher Weise eine Basis von $\text{Bil}(V \times V, \mathbb{R})$ nahe. Wieder kann man sich die Frage stellen, was bei einem Basiswechsel passiert. Wenn x und y zu

$$x'^k = S^k_i x^i \quad \text{und} \quad y'^l = S^l_j y^j \quad (9.28)$$

transformieren und $s_{ij} x^i y^j = s'_{kl} x'^k y'^l$ invariant bleiben soll, müssen die Koeffizienten s_{ij} transformiert werden zu

$$s'_{kl} = s_{ij} (S^{-1})^i_k (S^{-1})^j_l. \quad (9.29)$$

Damit haben wir einen weiteren Typ von Objekten mit einem charakteristischen Transformationsverhalten identifiziert. Dies ist ein Beispiel für einen Tensor 2. Stufe. Den Tensorbegriff werden wir noch genauer diskutieren, aber man sieht an dieser Stelle schon

einen gewissen Unterschied zum Begriff der Matrix: Natürlich lassen sich die Koeffizienten s_{ij} in ein quadratisches Zahlenschema (eine Matrix) schreiben, die wesentlichen Eigenschaften stecken aber in der durch die s_{ij} definierten Bilinearform, und diese Interpretation als Bilinearform legt das Transformationsverhalten fest.

Bevor wir zu weiteren von V abgeleiteten Vektorräumen und damit allgemeineren Tensoren kommen, wollen wir die Struktur von $\text{Bil}(V \times V, \mathbb{R})$ noch besser verstehen.

Die Basis-Bilinearformen v^{ij} wirken wie eine Art Produkt von v_*^i und v_*^j :

$$v^{ij}(v, w) = v_*^i(v) v_*^j(w). \quad (9.30)$$

Wir schreiben daher $v^{ij} = v_*^i \otimes v_*^j$, wobei das Tensorproduktsymbol \otimes definiert wird als die folgende bilineare Abbildung:

$$\begin{aligned} \otimes : V^* \times V^* &\longrightarrow \text{Bil}(V \times V, \mathbb{R}) \\ (F, G) &\longmapsto F \otimes G \end{aligned} \quad (9.31)$$

$$\text{wobei für alle } (v, w) \in V \times V : (F \otimes G)(v, w) = F(v) G(w). \quad (9.32)$$

Der Raum der Bilinearformen auf $V \times V$ ist also eine Art Produktraum der Linearformen auf V , und wir definieren das Tensorprodukt der Räume V^* und V^* als

$$V^* \otimes V^* := \text{Bil}(V \times V, \mathbb{R}). \quad (9.33)$$

Entsprechend können wir $V^* \otimes V^* \otimes V^*$ als Raum der Trilinearformen auf $V \times V \times V$ definieren. Zu einer Basis (v_1, \dots, v_n) von V gehört dann eine Basis bestehend aus $v_*^i \otimes v_*^j \otimes v_*^k$ von $V^* \otimes V^* \otimes V^*$. Die Koordinatendarstellung von Elementen von $V^* \otimes V^* \otimes V^*$ ist dann gegeben durch Tensoren dritter Stufe t_{ijk} mit dem Transformationsverhalten

$$t'_{ijk} = t_{lmn} (S^{-1})_i^l (S^{-1})^m_j (S^{-1})^n_k. \quad (9.34)$$

Die Verallgemeinerung zu Tensoren höherer Stufe ist offensichtlich.

Nun gibt es weitere von V abgeleitete Vektorräume, z.B. $\text{Hom}(V, V)$: den Raum der Endomorphismen von V . Ein Endomorphismus $F : V \rightarrow V$ wirkt auf $v = x^i v_i$ als

$$F(x^i v_i) = x^i F(v_i) = x^i (f^j_i v_j) = (f^j_i x^i) v_j, \quad (9.35)$$

wobei $F(v_i) = \sum_{j=1}^n f^j_i v_j$ die Zerlegung von $F(v_i) \in V$ in die Basisvektoren beschreibt. Die Matrix (f^j_i) ist die Matrixdarstellung $M_{\mathcal{A}}(F)$ von F bezüglich der Basis \mathcal{A} , nur jetzt mit einem hochgestellten Index. Dies passt zum Transformationsverhalten von $M_{\mathcal{A}}(F)$, das wir schon kennen:

$$(f') = S(f) S^{-1}, \text{ also} \quad (9.36)$$

$$(f')^j_i = S^j_k f^k_l (S^{-1})^l_i, \quad (9.37)$$

d.h. hochgestellte Indizes transformieren mit S , tiefgestellte mit S^{-1} .

Die Koeffizienten f^j_i sind Koordinaten von F bezüglich einer Basis von $\text{Hom}(V, V)$, nämlich der Basis v_j^i mit

$$v_j^i(v_k) = \delta_k^i v_j. \quad (9.38)$$

Denn dann erfüllt $F = f^j_i v_j^i$

$$F(v_k) = f^j_i v_j^i(v_k) = f^j_i \delta_k^i v_j = f^j_k v_j. \quad (9.39)$$

Diese Basis sieht aus wie ein Produkt aus v_*^i und v_j , und es liegt wieder nahe, ein Produkt $v_*^i \otimes v_j$ einzuführen. Wir hatten vorher $v_*^i \otimes v_*^j$ als Produkt von Linearformen definiert. Wir können auf analoge Weise verfahren, indem wir die v_j auch als Linearformen interpretieren, nämlich als Linearformen auf V^* (wir erinnern uns, dass V^{**} und V auf natürliche Weise isomorph sind).

Damit wir so verfahren können, müssen wir zunächst $\text{Hom}(V, V)$ mit $\text{Bil}(V \times V^*, \mathbb{R})$ identifizieren. Dies gelingt mittels des Isomorphismus

$$\begin{aligned} \phi : \text{Hom}(V, V) &\longrightarrow \text{Bil}(V \times V^*, \mathbb{R}) \\ F &\longmapsto \phi(F) \end{aligned} \quad (9.40)$$

$$\text{wobei für alle } (v, G) \in V \times V^* : \phi(F)(v, G) = G(F(v)). \quad (9.41)$$

Die Linearität von ϕ ist offensichtlich. Da die Dimension beider Räume gleich ist (nämlich n^2), muss ϕ ein Isomorphismus sein, wenn wir zeigen können, dass ϕ injektiv ist: Sei $\phi(F) = 0$, d.h. für alle $G \in V^*$, $v \in V$ ist $G(F(v)) = 0$. Dann ist $F(v) = 0$ für alle $v \in V$ und somit $F = 0$.

Wir wollen nun $\text{Bil}(V \times V^*, \mathbb{R})$ als Tensorprodukt der Räume $V^* = \text{Hom}(V, \mathbb{R})$ und $V \cong \text{Hom}(V^*, \mathbb{R})$ definieren mit der Tensorabbildung

$$\begin{aligned} \otimes : V^* \times V &\longrightarrow V^* \otimes V := \text{Bil}(V \times V^*, \mathbb{R}) \\ (G, v) &\longmapsto G \otimes v \end{aligned} \quad (9.42)$$

$$\text{wobei für alle } (w, F) \in V \times V^* : (G \otimes v)(w, F) = G(w) F(v). \quad (9.43)$$

Allgemein können wir damit beliebige Tensorprodukte von V und V^* bilden:

$$\begin{aligned} &\underbrace{V \otimes \cdots \otimes V}_r \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_s \\ &= \{\text{Multilineare Abbildungen } \underbrace{V^* \times \cdots \times V^*}_r \times \underbrace{V \times \cdots \times V}_s \rightarrow \mathbb{R}\}. \end{aligned} \quad (9.44)$$

Bezüglich der Standardbasis bestehend aus

$$v_{i_1} \otimes \cdots \otimes v_{i_r} \otimes v_*^{j_1} \otimes \cdots \otimes v_*^{j_s} \quad (9.45)$$

werden Elemente dieses Raums dargestellt durch Koeffizienten

$$T^{i_1 \cdots i_r}_{j_1 \cdots j_s}. \quad (9.46)$$

Unter einem Basiswechsel transformieren diese zu

$$T^{k_1 \dots k_r}_{l_1 \dots l_s} = S^{k_1}_{i_1} \dots S^{k_r}_{i_r} T^{i_1 \dots i_r}_{j_1 \dots j_s} (S^{-1})^{j_1}_{l_1} \dots (S^{-1})^{j_s}_{l_s}. \quad (9.47)$$

Man nennt solche Objekte Tensoren der Stufe (r, s) oder r -fach kontravariante und s -fach kovariante Tensoren.

Bemerkung

Betrachtet man einen Vektorraum mit Skalarprodukt und schränkt sich auf Orthonormalbasen ein, dann wird ein Basiswechsel durch eine orthogonale Matrix S beschrieben und der Unterschied zwischen kovarianten und kontravarianten Indizes verschwindet.

9.3 Tensorprodukt

Wir haben im letzten Abschnitt die Tensorprodukte $V \otimes V$ etc. eingeführt. Dies ist mathematisch aus verschiedenen Gründen noch nicht ganz befriedigend.

Zum einen sind unsere Definitionen streng genommen nicht konsistent, was wir aber leicht reparieren können. Wir haben geschrieben

$$V \otimes V := \text{Bil}(V^* \times V^*, \mathbb{R}). \quad (9.48)$$

Da V^* selbst natürlich ein Vektorraum ist, können wir ihn auch in diese Definition einsetzen und erhalten

$$V^* \otimes V^* = \text{Bil}(V^{**} \times V^{**}, \mathbb{R}) \quad (9.49)$$

im Gegensatz zu $\text{Bil}(V \times V, \mathbb{R})$, das in unserer ursprünglichen Definition aufgetaucht ist. Da V und V^{**} kanonisch isomorph zueinander sind, würde diese konsistentere Definition nichts wesentliches ändern (und hätte uns bei der Einführung von $V^* \otimes V^*$ eher verwirrt).

Zum anderen könnten wir weitere Wünsche an unser Tensorprodukt haben:

- Formulierung für das Produkt $V \otimes W$ verschiedener Vektorräume,
- Formulierung ohne auf den Dualraum zurückgreifen zu müssen,
- Charakterisierung anhand der Eigenschaften.

Den ersten Punkt können wir leicht abhaken und definieren

$$V \otimes W := \text{Bil}(V^* \times W^*, \mathbb{R}) \quad (9.50)$$

zusammen mit der Abbildung

$$\begin{aligned} \otimes : V \times W &\longrightarrow V \otimes W \\ (v, w) &\longmapsto v \otimes w \quad , \quad (v \otimes w)(F, G) = F(v)G(w). \end{aligned} \quad (9.51)$$

Um zu verstehen, wie wir $V \otimes W$ allein aus V und W aufbauen können, müssen wir diese Abbildung genauer verstehen.

- \otimes ist bilinear: für alle $v_1, v_2 \in V, w_1, w_2 \in W, \lambda \in \mathbb{R}$:

$$(v_1 + v_2) \otimes w_1 = v_1 \otimes w_1 + v_2 \otimes w_1 \quad (9.52)$$

$$v_1 \otimes (w_1 + w_2) = v_1 \otimes w_1 + v_1 \otimes w_2 \quad (9.53)$$

$$(\lambda v_1) \otimes w_1 = v_1 \otimes (\lambda w_1) = \lambda (v_1 \otimes w_1). \quad (9.54)$$

- \otimes ist im Allgemeinen weder injektiv (z.B. $0 \otimes v = 0 \otimes 0$) noch surjektiv: Sind $v_1, v_2 \in V$ linear unabhängig und ebenso $w_1, w_2 \in W$, dann lässt sich $v_1 \otimes w_1 + v_2 \otimes w_2$ nicht als $v \otimes w$ darstellen; $(v_1 + v_2) \otimes (w_1 + w_2)$ würde z.B. nicht funktionieren,

$$(v_1 + v_2) \otimes (w_1 + w_2) = v_1 \otimes w_1 + v_1 \otimes w_2 + v_2 \otimes w_1 + v_2 \otimes w_2 \neq v_1 \otimes w_1 + v_2 \otimes w_2. \quad (9.55)$$

Diese Eigenschaft ist übrigens wesentlich in der Quantenmechanik bei der Beschreibung verschränkter Zustände.

Sind (v_1, \dots, v_n) und (w_1, \dots, w_m) Basen von V bzw. von W , dann ist $(v_i \otimes w_j)$ eine Basis von $V \otimes W$, d.h. alle Elemente in $V \otimes W$ lassen sich als Linearkombinationen von Produkten $v_i \otimes w_j$ schreiben. Auch wenn die Abbildung \otimes nicht surjektiv ist, werden genügend viele Elemente erreicht, sodass alle anderen durch Linearkombinationen gebildet werden können.

Wir könnten nun versuchen, das Tensorprodukt $V \otimes W$ zu erklären, indem wir den Vektorraum

$$\{(a_{ij})_{\substack{i=1 \dots m \\ j=1 \dots n}} \mid a_{ij} \in \mathbb{R}\} \quad (9.56)$$

betrachten, also den mn -dimensionalen Raum der reellen $m \times n$ -Matrizen, und die Abbildung von $V \times W$ in diesen Raum über die Koeffizienten bezüglich der gewählten Basen definieren:

$$(x_i), (y_j) \mapsto (x_i y_j). \quad (9.57)$$

Diese Definition hätte den Vorteil, dass wir uns nicht des Dualraums bedienen müssten, aber den großen Nachteil, dass unsere Definition basisabhängig ist.

Alternativ können wir wie folgt vorgehen: Wir wissen, dass $V \otimes W$ durch Linearkombinationen von Produktvektoren $v \otimes w$ gebildet wird. Wenn wir keine Basis auszeichnen wollen, lassen wir formal zunächst alle (endlichen) Linearkombinationen von Paaren (v, w) zu und bilden den Vektorraum

$$\mathbb{R}^{V \times W} := \{(a_{(v,w)})_{(v,w) \in V \times W} \mid \text{für alle } v \in V, w \in W : a_{v,w} \in \mathbb{R}, \\ \text{nur endlich viele } a_{v,w} \neq 0\}. \quad (9.58)$$

Dieser Vektorraum (der *freie Vektorraum über $V \times W$*) ist sicherlich zu groß (er ist unendlich-dimensional). Eine Basis ist $(e_{(v,w)})_{(v,w) \in V \times W}$ mit

$$e_{(v,w)} = (\delta_{v',v} \delta_{w',w})_{(v',w') \in V \times W}. \quad (9.59)$$

Elemente von $\mathbb{R}^{V \times W}$ sind dann von der Form

$$\sum_{i=1}^N \lambda_i e_{(v_i, w_i)}, \quad (9.60)$$

wobei die v_i, w_i nun beliebige Vektoren sind (keine Auszeichnung einer Basis).

Dieser Vektorraum ist deshalb zu groß, weil wir Vektoren unterscheiden, die in $V \otimes W$ eigentlich gleich sein sollen, z.B. ist in $\mathbb{R}^{V \times W}$ für $\lambda \neq 1$

$$e_{(\lambda v, w)} \neq \lambda e_{(v, w)} \neq e_{(v, \lambda w)}. \quad (9.61)$$

Der Trick ist nun, dass wir in $\mathbb{R}^{V \times W}$ einfach alle Vektoren gleich setzen, die gleich sein sollen: Wir führen eine Äquivalenzrelation auf $\mathbb{R}^{V \times W}$ ein, die die gewünschten Rechenregeln implementiert:

$$e_{(v_1+v_2, w_1)} \sim e_{(v_1, w_1)} + e_{(v_2, w_1)} \quad (9.62)$$

$$e_{(v_1, w_1+w_2)} \sim e_{(v_1, w_1)} + e_{(v_1, w_2)} \quad (9.63)$$

$$e_{(\lambda v_1, w_1)} \sim \lambda e_{(v_1, w_1)} \sim e_{(v_1, \lambda w_1)}. \quad (9.64)$$

Diese Äquivalenzen generieren weitere, wenn wir fordern, dass \sim mit Addition und skalarer Multiplikation verträglich ist. Wir betrachten dann den Vektorraum

$$V \otimes W := \mathbb{R}^{V \times W} / \sim, \quad (9.65)$$

der aus den Äquivalenzklassen gebildet wird.

Bemerkung

Vektoren von der Form $e_{(v_1+v_2, w_1)} - e_{(v_1, w_1)} - e_{(v_2, w_1)}$ usw. generieren einen Untervektorraum $U \subset \mathbb{R}^{V \times W}$, die Äquivalenzrelation $a \sim b$ besagt dann, dass $a - b \in U$. Die Äquivalenzklassen sind also die affinen Räume $a + U$, und der Vektorraum dieser affinen Räume ist dann der Quotientenvektorraum,

$$V \otimes W = \mathbb{R}^{V \times W} / U. \quad (9.66)$$

Zuletzt noch ein mathematischer Leckerbissen: die Charakterisierung des Tensorprodukts durch seine Eigenschaften.

Das Tensorprodukt $V \otimes W$ ist gewissermaßen eine Linearisierung bilinearer Ausdrücke: $V \otimes W$ ist ein linearer Raum (also ein Vektorraum), wird aber aus Produkten aus zwei linearen Räumen gebildet. Bilineare Abbildungen $V \times W \rightarrow Z$ können dann über lineare Abbildungen $V \otimes W \rightarrow Z$ beschrieben werden, was Inhalt des folgenden Satzes ist:

Satz 9.3. *Seien V, W endlich-dimensionale, reelle Vektorräume. Zu jeder bilinearen Abbildung*

$$\xi : V \times W \rightarrow Z \quad (9.67)$$

gibt es genau eine lineare Abbildung

$$\xi_{\otimes} : V \otimes W \rightarrow Z, \quad (9.68)$$

sodass für alle $(v, w) \in V \times W$ gilt

$$\xi(v, w) = \xi_{\otimes}(v \otimes w). \quad (9.69)$$

Mit anderen Worten: Das folgende Diagramm kommutiert:

$$\begin{array}{ccc} V \times W & & \\ \otimes \downarrow & \searrow \xi & \\ V \otimes W & \xrightarrow{\xi_{\otimes}} & Z \end{array}$$

Beweis. Wir wählen Basen (v_1, \dots, v_n) , (w_1, \dots, w_m) , $(v_i \otimes w_j)$ von V , W und $V \otimes W$. Die obige Eigenschaft (9.69) von ξ_{\otimes} ausgewertet auf den Basisvektoren ergibt

$$\xi_{\otimes}(v_i \otimes w_j) = \xi(v_i, w_j). \quad (9.70)$$

Da ξ_{\otimes} eine lineare Abbildung ist, ist es durch die Wirkung auf Basisvektoren eindeutig festgelegt. Man kann sich davon überzeugen, dass die so definierte Abbildung die geforderte Eigenschaft (9.69) für beliebige v, w erfüllt (in den folgenden Ausdrücken wird über i und j summiert):

$$\xi_{\otimes}((x^i v_i) \otimes (y^j v_j)) = x^i y^j \xi_{\otimes}(v_i \otimes w_j) \quad (9.71)$$

$$= x^i y^j \xi(v_i, w_j) \quad (9.72)$$

$$= \xi(x^i v_i, y^j v_j). \quad (9.73)$$

□

Man kann nun zeigen, dass das Tensorprodukt $V \otimes W$ zusammen mit der bilinearen Abbildung $\otimes : V \times W \rightarrow V \otimes W$ bis auf Isomorphie durch diese Eigenschaft eindeutig bestimmt ist. Man nennt sie die **universelle Eigenschaft** des Tensorprodukts: Mit ihr haben wir eine Charakterisierung durch die Eigenschaften gefunden.

Damit sind wir am Ende des Kapitels und der Vorlesung angelangt. Wir haben die wichtigsten Grundlagen der linearen Algebra besprochen, und ich hoffe, Sie haben genug mitgenommen, um die Anwendungen der linearen Algebra, die Sie im Studium erwarten, meistern zu können.

Wie immer wird das Kapitel durch eine kurze Zusammenfassung geschlossen:

Zusammenfassung Kapitel 9

- Wir haben den Dualraum V^* näher kennen gelernt und die Konstruktion einer dualen Basis besprochen.
- Wir haben gesehen, dass sich die Koeffizienten einer Linearform bezüglich der dualen Basis unter einem Basiswechsel gerade entgegengesetzt zu den Koordinaten eines Vektors verhalten.
- Wir haben dann gesehen, dass auch andere mit V verwandte Räume wie $\text{Bil}(V \times V, \mathbb{R})$ eine natürliche Basis besitzen, sobald die Basis von V festgelegt ist. Dies impliziert wieder ein charakteristisches Verhalten unter Basistransformationen.
- Der Raum $\text{Bil}(V \times V, \mathbb{R})$ hat eine Struktur, die ähnlich aussieht wie das Produkt zweier Räume $V^* = \text{Hom}(V, \mathbb{R})$. Daher haben wir das Tensorprodukt $V^* \otimes V^* := \text{Bil}(V \times V, \mathbb{R})$ definiert.
- Ausgehend davon haben wir Produkte $\underbrace{V \otimes \cdots \otimes V}_r \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_s$ als Raum von Multilinearformen auf $\underbrace{V^* \times \cdots \times V^*}_r \times \underbrace{V \times \cdots \times V}_s$ definiert. Die entsprechenden Koeffizienten verhalten sich wie r -fach kontravariante und s -fach kovariante Tensoren.
- Zum Schluss haben wir den Begriff des Tensorprodukts näher beleuchtet. Wir haben die universelle Eigenschaft des Tensorprodukts kennen gelernt, die zeigt, dass mit Hilfe des Tensorprodukts bilineare Abbildungen auf $V \times W$ auf lineare Abbildungen auf $V \otimes W$ zurückgeführt werden können.

Literatur

- [F] Gerd Fischer, *Lineare Algebra, Eine Einführung für Studienanfänger*, Springer, 2014
- [FP] Ulf Friedrichsdorf und Alexander Prestel, *Mengenlehre für den Mathematiker*, Vieweg & Sohn, 1985
- [KvW] Hans Kerner und Wolf von Wahl, *Mathematik für Physiker*, Springer 2013
- [K] Hans-Joachim Kowalsky, *Lineare Algebra*, Walter de Gruyter, 1977