Lineare Algebra I-II

Skript zur Vorlesung
2023/2024

Daniel Plaumann
Rainer Sinn

Fassung vom 22. Januar 2024

Inhalt

Vorwort

It is idiotic to spend seven or eight months writing a novel, when you can buy one in a shop for two dollars.

Mark Twain

Die lineare Algebra ist neben der Analysis seit Jahrzehnten eine der beiden großen Vorlesungen am Anfang des Mathematikstudiums. Während man von der Analysis die Grundbegriffe der Differential- und Integralrechnung aus der Schule kennt, ist es bei der linearen Algebra vielleicht nicht so klar, worum es in dieser Vorlesung geht. Einige Themen möchten wir kurz vorstellen:

- **Lineare Gleichungssysteme.** Lineare Gleichungen sind die einfachsten, die es gibt, und fast der einzige Typ von Gleichungen, die sich im Prinzip immer lösen lassen. Die Theorie dazu steht am Anfang dieser Vorlesung.
- **Vektor- und Matrizenrechnung.** Die Rechentechniken der linearen Algebra werden wir die ganze Zeit üben. Besonders wichtig ist dabei das allgemeine Rechnen mit Symbolen, statt nur mit konkreten Zahlen.
- **Analytische Geometrie.** Als Vorlesung ist die lineare Algebra aus der analytischen Geometrie entstanden. Die Geometrie hat sowohl an der Schule als auch an der Universität sehr an Bedeutung verloren, was viele Leute als ästhetischen Verlust empfinden. Es ist aber wichtig, mit der Algebra zusammen auch die geometrische Anschauung zu entwickeln.
- **Strukturmathematik.** Ein Kennzeichen der modernen Mathematik ist der abstrakte Umgang mit ihren Objekten. So sind zum Beispiel Vektoren nicht einfach Listen von Zahlen, sondern werden zu 'Elementen eines Vektorraums' abstrahiert, die bestimmten Axiomen genügen. Diese Abstraktionen bereiten Anfängern erfahrungsgemäß Schwierigkeiten und sind mit einem hohen Lernaufwand verbunden. Es lohnt sich aber, weil grundlegende Ideen dadurch klarer werden und viele moderne Methoden, auch in der angewandten Mathematik, anders gar nicht zu verstehen sind.

Dieses Skript wird Ihnen vor der Vorlesung über Moodle zur Verfügung gestellt. Es kann trotzdem sinnvoll sein, in der Vorlesung mitzuschreiben, wenigstens in Stichpunkten.

Der Stoff der linearen Algebra ist weitgehend standardisiert und hat sich seit Jahrzehnten nur wenig verändert. Beim Schulstoff in der gymnasialen Oberstufe ist das völlig anders. An der Uni soll es im ersten Semester wieder *bei Null* losgehen, aber natürlich stimmt das nicht wirklich. Wir haben unter anderem deshalb den Einstieg in die Vorlesung etwas anders gestaltet als üblich. Wir erheben aber keinen Anspruch auf Originalität. (Das ist auch ein Unterschied zwischen einem Skript und einem Lehrbuch.) Neben den unten angegebenen Lehrbüchern haben wir auch Material aus anderen Skripten übernommen, von Rudolf Scharlau (Dortmund), Claus Scheiderer (Konstanz) und vor allem Wolf Barth (Erlangen).

Dieses Skript wird laufend überarbeitet, was leider fast sicher wieder Fehler aller Art mit sich bringt. Wenn Ihnen Fehler auffallen, machen Sie uns bitte darauf aufmerksam, in der Vorlesung oder per email.

Dortmund, Leipzig — Oktober 2023

Daniel Plaumann, Rainer Sinn

Literatur

Bücher über lineare Algebra gibt es wie Sand am Meer. Sie unterscheiden sich oft eher in der Präsentation und Reihenfolge des Stoffs als im eigentlichen Inhalt. Hier sind drei Empfehlungen. Diese Bücher können sie innerhalb des Universitätsnetzes auch als pdf-Dateien herunterladen:

[Fi] G. Fischer, *Lineare Algebra*. Springer Spektrum, Neunzehnte Auflage, 2020.
Das beliebteste Buch zur linearen Algebra in Deutschland.

[Be] A. Beutelspacher, *Lineare Algebra*. Springer Spektrum, Achte Auflage, 2014.
Der Autor ist seit Jahrzehnten in der Popularisierung und Darstellung von Mathematik in der Öffentlichkeit sehr aktiv. Das Buch gilt als besonders leicht verständlich.

[Kn-Ba] P. Knabner und W. Barth, *Lineare Algebra*. Springer Spektrum, 2013.
Ein sehr ausführliches Buch (fast 1000 Seiten!) mit vielen Beispielen und Anwendungen. Teile basieren auf Vorlesungen von Barth, an denen wir uns manchmal auch orientieren.

[Ax] S. Axler, *Linear Algebra Done Right*. Springer UTM, Dritte Auflage, 2015.
Ein fortgeschritteneres amerikanisches Lehrbuch, für eine zweite Vorlesung im amerikanischen System, welche die Rechentechniken der linearen Algebra weitgehend voraussetzt. Interessant trotzdem für einen anderen Blick, gerade auf den zweiten Teil der Vorlesung.

1

Einführung

L'algèbre n'est qu'une géométrie écrite;
la géométrie n'est qu'une algèbre figurée.

Algebra ist nichts als geschriebene Geometrie,
Geometrie nichts als gezeichnete Algebra.

Sophie Germain (1776–1831)

1 Lineare Gleichungssysteme

Die Algebra ist ursprünglich die Lehre vom Auflösen von Gleichungen, die lineare Algebra von *linearen* Gleichungen. Das sind die einfachsten Gleichungen, die es gibt. Ihre Lösungstheorie ist grundlegend für viele weitere Fragen und steht deshalb am Anfang dieser Vorlesung.

1.1 Zwei Beispiele

1.1 Beispiel *Der griechische Held Achilles läuft ein Wettrennen gegen eine Schildkröte. Die Schildkröte schafft* 200*m in einer Stunde. Achilles ist achtzig Mal so schnell, gibt der Schildkröte aber einen Kilometer Vorsprung. Wann wird Achilles die Schildkröte überholen?*

Wenn Achilles x Stunden unterwegs gewesen ist, hat die Schildkröte eine Distanz von $1 + \frac{1}{5}x$ Kilometer zurückgelegt, Achilles $16x$ Kilometer. Sie treffen sich, wenn beide Distanzen übereinstimmen. Das gibt uns eine lineare Gleichung

$$1 + \frac{1}{5}x = 16x.$$

Wir bringen x auf eine Seite und bekommen $1 = \frac{16\cdot 5-1}{5}x = \frac{79}{5}x$, also

$$x = \frac{5}{79}$$

Stunden. Das sind ungefähr 228 Sekunden, also etwas weniger als vier Minuten; siehe auch: *Achilles und die Schildkröte*. ◇

1.2 Beispiel Ein Beispiel aus dem chinesischen Altertum[1] (ca. 186 v. Chr.):

Erhält jeder zwei Münzen, bleiben zwei übrig. Erhält jeder drei, sind es zwei zu wenig. Um wieviele Personen und wieviele Münzen geht es?

Sei m die Anzahl der Münzen, p die der Personen. Dem Text entnehmen wir die beiden Gleichungen

$$m - 2p = 2$$
$$m - 3p = -2$$

[1] nach R.Hart. *Chinese Roots of Linear Algebra*, Johns Hopkins University Press (2010), S.53

Die erste Gleichung sagt also $m = 2 + 2p$. Einsetzen in die zweite ergibt

$$2 + 2p - 3p = -2$$

und damit also

$$p = 4 \quad \text{und} \quad m = 2 + 2 \cdot 4 = 10.$$

Es sind also 4 Personen und 10 Münzen. ◇

1.2 Lineare Gleichungen und Gleichungssysteme

Von diesen sehr einfachen Beispielen gehen wir direkt zum allgemeinen Fall.

Definition Eine **lineare Gleichung** ist eine Gleichung der Form

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

wobei $a_1, \ldots, a_n, b$ gegebene und $x_1, \ldots, x_n$ unbekannte reelle Zahlen sind.

Bei einer allgemeinen Diskussion müssen wir genau darauf achten, dass wir keine Spezialfälle vergessen (und zum Beispiel nicht aus Versehen durch Null teilen). Schon bei einer Gleichung sind drei Fälle zu unterscheiden:

(A) Nicht alle Koeffizienten $a_1, \ldots, a_n$ sind 0. Dann sei etwa k der erste Index mit $a_k \neq 0$. Die Gleichung hat damit die Form

$$0 \cdot x_1 + \cdots + 0 \cdot x_{k-1} + a_k \cdot x_k + a_{k+1} \cdot x_{k+1} + \cdots + a_n \cdot x_n \;=\; b.$$

Wir können also die ersten $k-1$ Variablen $x_1, \ldots, x_{k-1}$ beliebig wählen; ihre Werte haben auf die Gleichung gar keinen Einfluss. Anschließend teilen wir die ganze Gleichung durch a_k und lösen nach x_k auf. Wir erhalten

$$x_k \;=\; \frac{b - (a_{k+1}x_{k+1} + \cdots + a_nx_n)}{a_k}.$$

Wir können also auch $x_{k+1}, \ldots, x_n$ beliebig wählen und erhalten für jede solche Wahl genau das passende x_k, das die Gleichung löst.

(B) Alle Koeffizienten $a_1, \ldots, a_n$ sind 0, aber $b \neq 0$. Die Gleichung lautet dann

$$0 \cdot x_1 + \cdots + 0 \cdot x_n \;=\; b.$$

Egal, wie wir $x_1, \ldots, x_n$ wählen, diese Gleichung ist unlösbar.

(C) Es gilt $a_1 = \cdots = a_n = 0$ und auch $b = 0$. Dann ist die Gleichung immer erfüllt, sie stellt an die Unbekannten $x_1, \ldots, x_n$ gar keine Bedingungen.

Definition Ein **lineares Gleichungssystem** ist ein System

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n &= b_2 \\ \vdots \qquad\quad \vdots \quad \vdots \qquad\quad \vdots & \\ a_{m,1}x_1 + a_{m,2}x_2 + \ldots + a_{m,n}x_n &= b_m. \end{aligned}$$

aus mehreren linearen Gleichungen. Eine Lösung des Systems ist eine gemeinsame Lösung aller dieser Gleichungen.

Wir können ein lineares Gleichungssystem folgendermaßen lösen: Nehmen wir zunächst an, dass der allererste Koeffizient nicht 0 ist, also $a_{1,1} \neq 0$. Wir können dann die erste Zeile wie in (A) nach x_1 auflösen und x_1 dann in allen übrigen Gleichungen ersetzen. Das Gleichungssystem nimmt damit die Form

$$\begin{aligned} x_1 + a'_{1,2}x_2 + \ldots + a'_{1,n}x_n &= b'_1 \\ a'_{2,2}x_2 + \ldots + a'_{2,n}x_n &= b'_2 \\ \vdots \quad \vdots \qquad\quad \vdots & \\ a'_{m,2}x_2 + \ldots + a'_{m,n}x_n &= b'_n \end{aligned}$$

an, mit neuen Koeffizienten $a'_{1,2}, \ldots, a'_{m,n}$ und neuen Konstanten $b'_1, \ldots, b'_n$. Jetzt lassen wir die erste Gleichung stehen und wiederholen das Prozedere mit den übrigen Gleichungen, in denen nur noch $x_2, \ldots, x_n$ vorkommen. In dieser Weise können wir das Gleichungssystem sukzessive lösen, es sei denn wir treffen irgendwann auf eine unlösbare Gleichung (Typ (B)), dann ist das ganze System unlösbar.

Wegen seiner grundlegenden Bedeutung werden wir dieses Eliminationsverfahren jetzt formalisierter beschreiben. Als erstes sparen wir uns Schreibarbeit, indem wir die folgende vereinfachende Notation einführen.

Definition Die **Koeffizientenmatrix** eines linearen Gleichungssystems wie oben ist das rechteckige Zahlenschema[2]

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

[2]Wir lassen das Komma zwischen den Indizes meistens weg und schreiben a_{ij} statt $a_{i,j}$.

Daraus entsteht die **erweiterte Koeffizientenmatrix**, indem wir die rechte Seite der Gleichungen als weitere Spalte hinzufügen.

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right)$$

1.3 Das Eliminationsverfahren

An der erweiterten Koeffizientenmatrix nehmen wir drei Typen von Umformungen vor, die wir **elementare Zeilenumformungen** nennen:

(I) Vertauschen zweier Zeilen;

(II) Multiplikation einer Zeile mit einer Konstanten ungleich 0;

(III) Addition des c-fachen einer Zeile ($c \in \mathbb{R}$) zu einer anderen.

Alle drei Typen von elementaren Zeilenumformungen ändern nichts an den Lösungen des zugehörigen linearen Gleichungssystems. (Für den dritten Typ sollte man sich das einmal klar machen!)

Definition Eine Matrix hat **Zeilenstufenform**, wenn sie von der Form

$$\left(\begin{array}{ccccccccccccc} \overbrace{0 & \cdots & 0}^{n_0} & 1 & \overbrace{* & \cdots & *}^{n_1} & * & \cdots & * & \overbrace{* & \cdots & *}^{n_r} \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & 1 & \cdots & * & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \cdots & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right)$$

ist. An den mit $*$ markierten Stellen darf dabei eine beliebige Zahl stehen. Jede Zeile beginnt also mit mehr Nulleinträgen als die vorige, und der erste Eintrag ungleich 0 ist immer[3] eine 1, außer in den Nullzeilen am Ende. Dabei ist r die Anzahl der *Stufen* und $n_0, \dots, n_r$ sind die *Stufenlängen*. Die Stufenlängen können auch 0 sein und r kann mit der Gesamtzahl der Zeilen übereinstimmen, so dass unten keine Nullzeile steht. Es darf auch $r = 0$ sein, nämlich für die Nullmatrix.

1.3 Satz (Eliminationsverfahren) *Jede Matrix kann durch eine endliche Abfolge von elementaren Zeilenumformungen in Zeilenstufenform gebracht werden.*

[3] Bei uns steht in der Zeilenstufenform am Anfang jeder Stufe eine 1. Das ist in der Literatur häufig anders. Insgesamt macht das aber keinen großen Unterschied.

1.4 Beispiel Bevor wir den Satz beweisen, rechnen wir ein weiteres Beispiel. Gegeben sei das lineare Gleichungssystem

$$\begin{aligned} -\ x_2 + \ 5x_3 &= 3 \\ 3x_1 + 4x_2 + 16x_3 &= 0 \\ -2x_1 - 3x_2 - \ 9x_3 &= 1 \end{aligned}$$

Wir bringen die erweiterte Koeffizientenmatrix durch Zeilenumformungen in Zeilenstufenform. Dabei deuten wir mit Pfeilen an, welche Umformung vorgenommen wurde.

$$\left(\begin{array}{ccc|c} 0 & -1 & 5 & 3 \\ 3 & 4 & 16 & 0 \\ -2 & -3 & -9 & 1 \end{array}\right) \overset{\longleftarrow}{\longleftarrow} \rightsquigarrow \left(\begin{array}{ccc|c} 3 & 4 & 16 & 0 \\ 0 & -1 & 5 & 3 \\ -2 & -3 & -9 & 1 \end{array}\right) \mid \cdot \tfrac{1}{3} \rightsquigarrow \left(\begin{array}{ccc|c} 1 & \frac{4}{3} & \frac{16}{3} & 0 \\ 0 & -1 & 5 & 3 \\ -2 & -3 & -9 & 1 \end{array}\right) \overset{\cdot 2}{\underset{+}{\longleftarrow}} \rightsquigarrow$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & \frac{4}{3} & \frac{16}{3} & 0 \\ 0 & -1 & 5 & 3 \\ 0 & -\frac{1}{3} & \frac{5}{3} & 1 \end{array}\right) \mid \cdot -1 \rightsquigarrow \left(\begin{array}{ccc|c} 1 & \frac{4}{3} & \frac{16}{3} & 0 \\ 0 & 1 & -5 & -3 \\ 0 & -\frac{1}{3} & \frac{5}{3} & 1 \end{array}\right) \overset{\cdot \frac{1}{3}}{\underset{+}{\longleftarrow}} \rightsquigarrow \left(\begin{array}{ccc|c} 1 & \frac{4}{3} & \frac{16}{3} & 0 \\ 0 & 1 & -5 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Damit ist die Zeilenstufenform erreicht. Schreiben wir das wieder als Gleichungssystem hin:

$$\begin{aligned} x_1 + \frac{4}{3}x_2 + \frac{16}{3}x_3 &= \ 0 \\ x_2 - \ 5x_3 &= -3 \end{aligned}$$

Dieses System hat dieselben Lösungen wie das Ausgangssystem, und wir können alle Lösungen direkt durch *Rückeinsetzen* bestimmen: Aus der zweiten Gleichung bekommen wir $x_2 = 5x_3 - 3$ und Einsetzen in die erste ergibt

$$x_1 = -\frac{4}{3}x_2 - \frac{16}{3}x_3 = -\frac{4}{3}\Big(5x_3 - 3\Big) - \frac{16}{3}x_3 = -12x_3 + 4.$$

Wir erhalten also die Lösungen des Gleichungssystems dadurch, dass wir x_3 beliebig vorgeben und können dann x_2 und x_1 sofort berechnen:

$$x_1 = -12x_3 + 4, \quad x_2 = 5x_3 - 3, \quad x_3 \text{ beliebig}$$

Für zum Beispiel $x_3 = 0$ oder $x_3 = 1$ bekommen wir die beiden Lösungen

$$x_1 = 4,\ \ x_2 = -3,\ \ x_3 = 0 \qquad \text{oder} \qquad x_1 = -8,\ \ x_2 = 2,\ \ x_3 = 1$$

Diese Art Rückeinsetzung funktioniert immer, sobald die erweiterte Koeffizientenmatrix in Zeilenstufenform vorliegt. ◇

Beweis von Satz 1.3. Mit den drei Typen von Zeilenumformungen räumen wir von links nach rechts eine Spalte nach der anderen auf: Wenn die ganze Matrix nur lauter Nullen enthält, brauchen wir nichts zu tun. Ansonsten suchen wir die erste Spalte, in der ein Eintrag ungleich 0 steht. Steht dieser Eintrag an der Stelle (k, ℓ), also in der k-ten Zeile und der ℓ-ten Spalte, dann wenden wir als erstes Typ (I) an und vertauschen die k-te Zeile mit der ersten. An der Stelle $(1, \ell)$ steht dann also ein Eintrag $a \neq 0$, während alle Einträge in den ersten $\ell - 1$ Spalten 0 sind. Wir wenden Typ (II) an und teilen die erste Zeile durch a, so dass an der Stelle $(1, \ell)$ anschließend eine 1 steht. Nun verwenden wir Typ (III): Für jedes $j = 2, \dots, m$ sei $b_{j\ell}$ der Eintrag der jetzigen Matrix an der Stelle (j, ℓ). Dann addieren wir das $-b_{j\ell}$-fache der ersten Zeile zur j-ten Zeile, um dort den Eintrag an der Stelle (j, ℓ) zu 0 zu machen. (Das hat den gleichen Effekt, wie die Gleichung für x_ℓ, die sich aus der ersten Gleichung ergibt, in der j-ten Gleichung einzusetzen.) Das folgende Schema zeigt die bisherigen Schritte:

$$\begin{array}{c} \\ k \end{array}\!\!\begin{array}{c} \overset{\ell}{} \\ \begin{pmatrix} 0 & \cdots & 0 & * & * & \cdots & * \\ 0 & \cdots & 0 & * & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & a & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & * & * & \cdots & * \end{pmatrix} \end{array} \rightsquigarrow \begin{array}{c} \overset{\ell}{} \\ \begin{pmatrix} 0 & \cdots & 0 & a & * & \cdots & * \\ 0 & \cdots & 0 & * & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & * & * & \cdots & * \end{pmatrix} \end{array} \rightsquigarrow \begin{array}{c} \overset{\ell}{} \\ \begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & * & \cdots & * \end{pmatrix} \end{array}$$

Nun lassen wir die erste Zeile in Frieden und wenden auf die verbleibenden $m-1$ Zeilen die gleiche Prozedur an. Nach spätestens m Schritten erreichen wir so die Zeilenstufenform. ∎

Nachdem wir die erweiterte Koeffizientenmatrix eines linearen Gleichungssystems in Zeilenstufenform gebracht haben, wird es sehr einfach, die Lösungen zu bestimmen. Als erstes betrachten wir die letzte Stufe. Dort lässt sich direkt ablesen, ob das System lösbar ist oder nicht.

Lösbar

$$\left(\begin{array}{cccccccccccc|c} 0 & \cdots & 0 & 1 & * & \cdots & * & * & \cdots & * & * & \cdots & * & * \\ & & & 0 & 0 & \cdots & 0 & 1 & \cdots & * & * & \cdots & * & * \\ & & & & & & & 0 & \cdots & 1 & * & \cdots & * & * \\ & & & & & & & & & 0 & 0 & \cdots & 0 & 0 \end{array}\right)$$

Unlösbar

$$\left(\begin{array}{cccccccccccc|c} 0 & \cdots & 0 & 1 & * & \cdots & * & * & \cdots & * & * & \cdots & * & * \\ & & & 0 & 0 & \cdots & 0 & 1 & \cdots & * & * & \cdots & * & * \\ & & & & & & & 0 & \cdots & 1 & * & \cdots & * & * \\ & & & & & & & & & 0 & 0 & \cdots & 0 & 1 \\ & & & & & & & & & & & \cdots & 0 & 0 \end{array}\right)$$

Wenn in der letzten Spalte der erweiterten Koeffizientenmatrix noch einmal eine neue Stufe anfängt, wie im System auf der rechten Seite, dann ist das System unlösbar. Die entsprechende Gleichung lautet dann nämlich $0 \cdot x_1 + \cdots + 0 \cdot x_n = 1$. Und wenn eine Gleichung unlösbar ist, dann ist auch das ganze System unlösbar.

Im System auf der linken Seite dagegen können wir die Unbekannten, die nicht am Anfang einer Stufe stehen, beliebig vorgeben und dann die übrigen Unbekannten von unten nach oben durch Einsetzen ausrechnen, beginnend auf der letzten Stufe. Solche Unbekannten, die wir in der Zeilenstufenform beliebig wählen dürfen, nennen wir **freie Unbekannte**.

Dieses allgemeine Verfahren wird **Gaußsches Eliminationsverfahren** oder **Gauß-Algorithmus** genannt. Tatsächlich hat Gauß[4] es wohl als erster systematisch beschrieben, obwohl ähnliche Rechentechniken lange vorher bekannt waren. (Das sogenannte *Fangcheng* ist eine frühe Form der Matrizenrechnung, die schon im chinesischen Altertum beschrieben wurde.) Lineare Gleichungssysteme sind fast die einzigen Gleichungssysteme in der ganzen Mathematik, die sich immer exakt lösen lassen. Deshalb versucht man, möglichst viele andere Rechenprobleme auf das Lösen linearer Gleichungssysteme zurückzuführen.

1.5 Beispiel Wir betrachten noch einmal das lineare Gleichungssystem aus Beispiel 1.4, aber mit beliebiger rechter Seite:

$$\begin{aligned} -\ x_2 + \ 5x_3 &= b_1 \\ 3x_1 + 4x_2 + 16x_3 &= b_2 \\ -2x_1 - 3x_2 - \ 9x_3 &= b_3 \end{aligned}$$

Wir bringen wieder die erweiterte Koeffizientenmatrix durch elementare Zeilenumformungen in Zeilenstufenform, mit denselben Rechenschritten wie zuvor.

$$\left(\begin{array}{ccc|c} 0 & -1 & 5 & b_1 \\ 3 & 4 & 16 & b_2 \\ -2 & -3 & -9 & b_3 \end{array}\right) \leftarrow\!\!\lrcorner \rightsquigarrow \left(\begin{array}{ccc|c} 3 & 4 & 16 & b_2 \\ 0 & -1 & 5 & b_1 \\ -2 & -3 & -9 & b_3 \end{array}\right) \mid \cdot \tfrac{1}{3} \rightsquigarrow \left(\begin{array}{ccc|c} 1 & \frac{4}{3} & \frac{16}{3} & \frac{1}{3}b_2 \\ 0 & -1 & 5 & b_1 \\ -2 & -3 & -9 & b_3 \end{array}\right) \begin{array}{l} \cdot 2 \\ \lrcorner\!\!\rightarrow + \end{array}$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & \frac{4}{3} & \frac{16}{3} & \frac{1}{3}b_2 \\ 0 & -1 & 5 & b_1 \\ 0 & -\frac{1}{3} & \frac{5}{3} & b_3 + \frac{2}{3}b_2 \end{array}\right) \mid \cdot(-1) \rightsquigarrow \left(\begin{array}{ccc|c} 1 & \frac{4}{3} & \frac{16}{3} & \frac{1}{3}b_2 \\ 0 & 1 & -5 & -b_1 \\ 0 & -\frac{1}{3} & \frac{5}{3} & b_3 + \frac{2}{3}b_2 \end{array}\right) \begin{array}{l} \\ \lrcorner\!\!\rightarrow + \end{array}^{\cdot \frac{1}{3}}$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & \frac{4}{3} & \frac{16}{3} & \frac{1}{3}b_2 \\ 0 & 1 & -5 & -b_1 \\ 0 & 0 & 0 & b_3 + \frac{2}{3}b_2 - \frac{1}{3}b_1. \end{array}\right)$$

[4] Carl Friedrich Gauss (1777–1855)

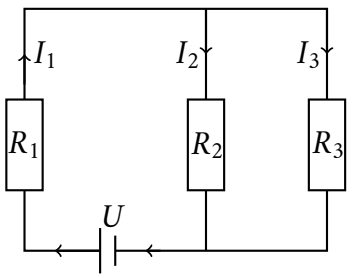
Damit das Gleichungssystem eine Lösung in x_1, x_2, x_3 hat, muss $b_3 + \frac{2}{3}b_2 - \frac{1}{3}b_1 = 0$ gelten. Das ist wieder eine lineare Gleichung in b_1, b_2, b_3. Für unsere ursprüngliche rechte Seite $b_1 = 3$, $b_2 = 0$ und $b_3 = 1$ war das zum Beispiel der Fall. ◇

1.4 Noch ein Beispiel

Lineare Gleichungssysteme haben eine riesige Bedeutung für die angewandte Mathematik. Viele Probleme in Wissenschaft und Technik sind linear. Noch viel mehr Probleme sind zwar nicht von Natur aus linear, lassen sich aber auf die eine oder andere Art *linearisieren* und dann als lineare Gleichungssysteme lösen.

Bei vielen Schulaufgaben bis hin zum Abitur besteht der schwierigste Teil darin, aus der Aufgabenstellung die richtigen Gleichungen zu ermitteln (Textaufgaben). In manchen Anwendungsproblemen ist das im Prinzip genauso. Im Mathematikstudium, und gerade jetzt am Anfang, steht dieser Aspekt aber überhaupt nicht im Vordergrund. Wir interessieren uns erst einmal nur für die Gleichungen und nicht dafür, wo sie herkommen. Deshalb werden echte Anwendungsbeispiele in dieser Vorlesung eine Ausnahme bleiben.

1.6 Beispiel Wir betrachten das elektrische Netzwerk[5]



Gegeben sind die Widerstände R_1, R_2, R_3 und die Spannung U (alles reelle Zahlen, wobei $R_1, R_2, R_3 > 0$ gelte). Gesucht sind die Stromstärken I_1, I_2, I_3. Nun muss man wissen, was die Gesetze der Physik hier sagen. Das ist der interessanteste Teil der Aufgabe: Die beteiligten Größen erfüllen die drei Gleichungen

$$\begin{aligned} I_1 &= I_2 + I_3 \\ R_2 I_2 &= R_3 I_3 \\ R_1 I_1 + R_2 I_2 &= U. \end{aligned}$$

[5] aus Knaber-Barth [Kn-Ba,§1.1.1]

Das kann man »zu Fuß« lösen, in dem man $I_1 = I_2 + I_3$ in den übrigen Gleichungen einsetzt usw. Wir gehen zur Übung streng nach Gauß-Algorithmus vor und bilden die erweiterte Koeffizientenmatrix in den Unbekannten I_1, I_2, I_3:

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & R_2 & -R_3 & 0 \\ R_1 & R_2 & 0 & U \end{array}\right)$$

Nun bringen wir diese Matrix in Zeilenstufenform. Dabei muss man nur aufpassen, das man mit den symbolischen Größen richtig rechnet:

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & R_2 & -R_3 & 0 \\ R_1 & R_2 & 0 & U \end{array}\right) \begin{array}{l} \\ \\ \cdot(-R_1) \\ \longleftarrow + \end{array} \rightsquigarrow \left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & R_2 & -R_3 & 0 \\ 0 & R_1+R_2 & R_1 & U \end{array}\right) \begin{array}{l} \\ \\ \cdot(-R_1) \\ \longleftarrow + \end{array}$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & R_2 & -R_3 & 0 \\ 0 & R_1+R_2 & R_1 & U \end{array}\right) \mid \cdot \tfrac{1}{R_2} \rightsquigarrow \left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & 1 & -\frac{R_3}{R_2} & 0 \\ 0 & R_1+R_2 & R_1 & U \end{array}\right) \begin{array}{l} \\ \\ \cdot(-(R_1+R_2)) \\ \longleftarrow + \end{array}$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & 1 & -\frac{R_3}{R_2} & 0 \\ 0 & 0 & R_1 + \frac{(R_1+R_2)R_3}{R_2} & U \end{array}\right) = \left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & 1 & -\frac{R_3}{R_2} & 0 \\ 0 & 0 & \frac{R_1R_2+R_1R_2+R_2R_3}{R_2} & U \end{array}\right) \mid \cdot \tfrac{R_2}{R_1R_2+R_1R_2+R_2R_3}$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ 0 & 1 & -\frac{R_3}{R_2} & 0 \\ 0 & 0 & 1 & \frac{R_2U}{R_1R_2+R_1R_2+R_2R_3} \end{array}\right)$$

Die Zeilenstufenform ist erreicht. Es ist wichtig, dass wir hier nirgends durch 0 geteilt haben, denn da $R_1, R_2, R_3 > 0$ gilt, ist auch $R_1R_2 + R_1R_3 + R_2R_3 > 0$. Wir bestimmen die Lösungen durch Rückeinsetzen. Setzen wir $R = R_1R_2 + R_1R_3 + R_2R_3$, dann bekommen wir

$$I_3 = \frac{R_2U}{R}, \quad I_2 = \frac{R_3}{R_2}I_3 = \frac{R_3U}{R}, \quad I_1 = I_2 + I_3 = \frac{(R_2+R_3)U}{R}$$

und haben die gesuchten Stomstärken gefunden. ◇

Beim Rechnen mit konkreten Zahlen ergeben sich in der Praxis, vor allem bei großen Systemen, zusätzliche Schwierigkeiten durch Rundungsfehler oder durch Ungenauigkeiten in den Ausgangsdaten. Wie man mit solchen Schwierigkeiten umgeht, lernt man in der numerischen Mathematik.

2 Vektoren und Geraden

In dieser Vorlesung geht es um die Grundlagen der Vektorrechnung. Parallel dazu werden einige Mengennotationen und allgemeine Beweistechniken vorgestellt, am Beispiel von Geraden im Raum.

2.1 Vektoren

Ein **Vektor** ist eine Liste

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

von reellen Zahlen $x_1, \dots, x_n$. Jede dieser Zahlen heißt ein **Eintrag** von $\vec{x}$. Die Reihenfolge der Einträge ist nicht egal: Es gilt beispielsweise $\left(\begin{smallmatrix} 2 \\ -1 \end{smallmatrix}\right) \neq \left(\begin{smallmatrix} -1 \\ 2 \end{smallmatrix}\right)$. Deshalb ist dieser Vektor auch nicht dasselbe wie die Menge $\{-1, 2\} = \{2, -1\}$. Meistens schreibt man Vektoren als Spalten und nicht als Zeilen, aber im Moment spielt das keine große Rolle.

Die Menge aller Vektoren mit n reellen Einträgen ist der **n-dimensionale Zahlenraum** $\mathbb{R}^n$. Wir schreiben

$$\vec{x} \in \mathbb{R}^n$$

um auszudrücken, dass $\vec{x}$ ein Element der Menge $\mathbb{R}^n$ ist. Wir stellen uns $\mathbb{R}^1 = \mathbb{R}$ als Zahlenstrahl vor, $\mathbb{R}^2$ als Ebene und $\mathbb{R}^3$ als dreidimensionalen Raum. Die Vektoren sind die **kartesischen Koordinaten**[1] von Punkten (Abb. 2.1).
Der **Nullvektor** ist

$$\vec{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Im Zahlenraum ist der Nullvektor der **Ursprung** oder **Nullpunkt**.

[1] Das Wort »kartesisch« ist aus dem Namen des französischen Philosophen und Mathematikers René Descartes (1596–1650) abgeleitet. Die kartesischen Koordinaten sind in seiner Zeit entstanden, gehen aber nicht auf Descartes selbst zurück.

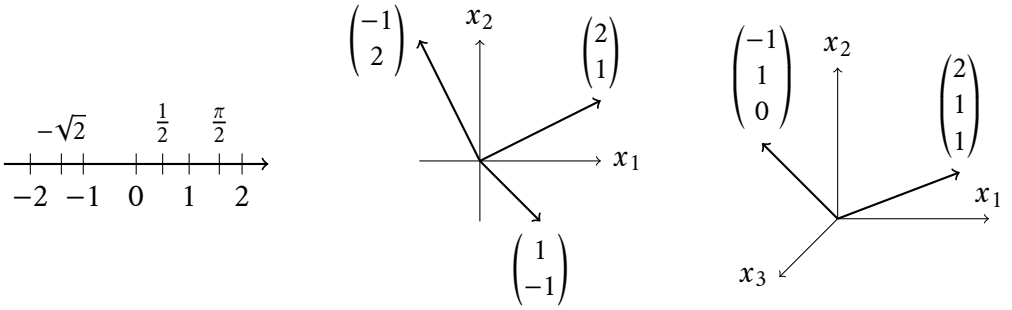


Abb. 2.1: Vektoren als Punkte in $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^3$

Für Vektoren gibt es zwei Rechenoperationen: Bei der **Vektoraddition** werden die Einträge addiert, also

$$\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}+\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}=\begin{pmatrix}x_1+y_1\\x_2+y_2\\\vdots\\x_n+y_n\end{pmatrix} \qquad \text{für } \vec{x},\vec{y}\in\mathbb{R}^n.$$

Zweitens gibt es die **Skalarmultiplikation**

$$a\cdot\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}a\cdot x_1\\a\cdot x_2\\\vdots\\a\cdot x_n\end{pmatrix} \qquad \text{für } \vec{x}\in\mathbb{R}^n \text{ und } a\in\mathbb{R}.$$

Hier wird jeder Eintrag eines Vektors mit derselben reellen Zahl multipliziert[2] (Abb. 2.2). Eine einzelne reelle Zahlen wird, zur Unterscheidung von einem Vektor, auch **Skalar** genannt.

Bei der Vektoraddition $\vec{x}+\vec{y}$ setzt man den Anfang des Vektors (Pfeils) $\vec{x}$ ans Ende des Vektors $\vec{y}$, oder umgekehrt. Die Skalarmultiplikation $a\cdot\vec{x}$ entspricht einer Streckung um den Faktor a, falls $a>0$ ist. Für $a<0$ entspricht $a\cdot\vec{x}$ der Spiegelung von $\vec{x}$ am Ursprung, gefolgt von einer Streckung um den positiven Faktor $|a|$. Für $a=0$ gilt immer $0\cdot\vec{x}=\vec{0}$.

[2]Natürlich könnte man auch die Einträge von zwei Vektoren multiplizieren statt addieren. Das spielt aber in der linearen Algebra keine Rolle.

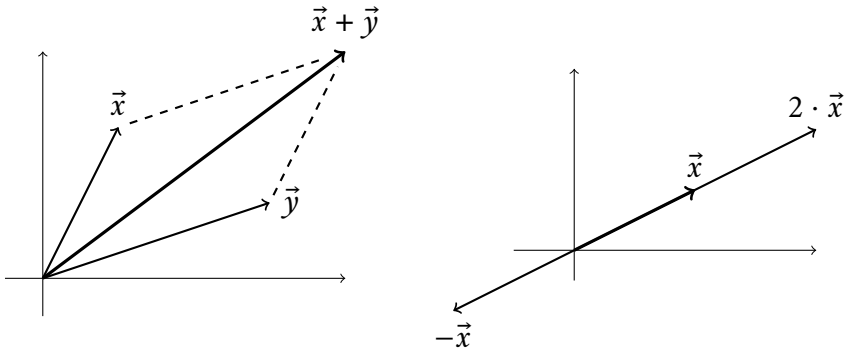


Abb. 2.2: Vektoraddition und Skalarmultiplikation in $\mathbb{R}^2$

Jeder Vektor $\vec{x} \neq \vec{0}$ hat damit *zwei verschiedene geometrische Interpretationen*[3]:

- Als **Punkt** mit den kartesischen Koordinaten $(x_1, \ldots, x_n)$.
- Als **Richtung** und **Länge**, was wir uns als Pfeil denken, der vom Ursprung zum Punkt $\vec{x}$ zeigt, oder auch von einem Punkt $\vec{y}$ zum Punkt $\vec{x} + \vec{y}$.

Übliche Rechenregeln für reelle Zahlen, die wir als bekannt voraussetzen, übertragen sich sofort auf die Vektoraddition und die Skalarmultiplikation. Für alle Vektoren $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ und alle Skalare $a, b \in \mathbb{R}$ gelten:

(1) $\vec{x} + \vec{y} = \vec{y} + \vec{x}$. (*Kommutativität der Addition*)

(2) $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$. (*Assoziativität der Addition*)

(3) $(a \cdot b) \cdot \vec{x} = a \cdot (b \cdot \vec{x})$. (*Assoziativität der Skalarmultiplikation*)

(4) $a \cdot (\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}$ und $(a + b)\vec{x} = a\vec{x} + b\vec{x}$. (*Distributivität*)

(5) $1 \cdot \vec{x} = \vec{x}$ und $0 \cdot \vec{x} = \vec{0}$.

Für $n \geqslant 4$ können wir den »n-dimensionalen Raum« (früher *Hyperraum* genannt) nicht mehr visuell erfassen. Vektoren sind einfach Listen von Zahlen und können sehr allgemeine **Daten** repräsentieren, womöglich mit Tausenden von Einträgen. Es erscheint zunächst unsinnig, sie sich als Pfeilchen im Raum vorzustellen. Die geometrische Vorstellung ist aber in der linearen Algebra trotzdem oft sehr hilfreich, um abstrakte Aussagen besser zu verstehen.

[3] In der Literatur (besonders der älteren) wird manchmal zwischen *Ortsvektoren* und *Richtungsvektoren* unterschieden. Das sind für uns dieselben Objekte, nur verschieden interpretiert.

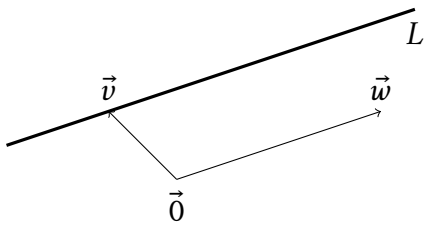
2.2 Geraden

Geraden in der Ebene kennt man aus der Schule. Wir geben die folgende Definition einer Geraden, die in beliebiger Dimension sinnvoll ist.

Definition Eine **Gerade** in $\mathbb{R}^n$ ist gegeben durch zwei Vektoren $\vec{v}, \vec{w} \in \mathbb{R}^n$ mit $\vec{w} \neq \vec{0}$ als die Menge

$$L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}.$$

Sie ist bestimmt durch einen **Fußpunkt** $\vec{v}$ und eine **Richtung** $\vec{w}$. Damit wirklich eine Gerade herauskommt, setzen wir $\vec{w} \neq \vec{0}$ voraus. (Für $\vec{w} = \vec{0}$ bestünde die Menge L nur aus dem Fußpunkt $\vec{v}$).



Die Notation $\{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$ ist eine *Mengenschreibweise*: Im Ausdruck $\vec{v} + t\vec{w}$ ist t ein **Parameter**, der alle reellen Werte durchläuft. Das wird durch das $t \in \mathbb{R}$ hinter dem Querstrich gekennzeichnet. Die beiden Vektoren $\vec{v}$ und $\vec{w}$ sind fest gewählt und verändern sich nicht.

2.1 Beispiel Wir betrachten die Vektoren

$$\vec{v} = \begin{pmatrix} -3 \\ 1 \end{pmatrix}, \quad \vec{w} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Die Gerade $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$ enthält zum Beispiel die Punkte

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \end{pmatrix} + 2 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} -5 \\ 0 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \end{pmatrix} + (-1) \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix},$$
$$\begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} -3 + 2\sqrt{2} \\ 1 + \sqrt{2} \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \end{pmatrix} + \sqrt{2} \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

und so weiter. Dagegen liegt der Punkt $\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ nicht auf der Geraden L: Es gibt keinen Wert des Parameters $t \in \mathbb{R}$, der

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \end{pmatrix} + t \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

erfüllt, denn die beiden Gleichungen

$$1 = -3 + 2t \quad \text{und} \quad 2 = 1 + t,$$

die sich für die erste und die zweite Koordinate ergeben, haben keine gemeinsame Lösung in $t \in \mathbb{R}$. Aus der zweiten Gleichung ergibt sich nämlich $t = 1$ als einzige Lösung, was keine Lösung der ersten Gleichung ist. ◇

Die Gerade $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$ enthält die Vektoren $\vec{v} + t\vec{w}$ für alle Werte des Parameters t. Das können wir mit einem **Allquantor** $\forall$ schreiben:

$$\forall t \in \mathbb{R}\colon \vec{v} + t\vec{w} \in L.$$

Andererseits liegt ein Vektor $\vec{x}$ genau dann auf der Geraden L, wenn eine reelle Zahl t mit $\vec{x} = \vec{v} + t\vec{w}$ existiert, was wir mit einem **Existenzquantor** $\exists$ schreiben können:

$$L = \left\{\vec{x} \in \mathbb{R}^n \mid \exists t \in \mathbb{R}\colon \vec{x} = \vec{v} + t\vec{w}\right\}.$$

Die Bezeichnungen innerhalb der Mengenklammern spielen übrigens keine Rolle, außer für die vorher festgelegten Vektoren $\vec{v}$ und $\vec{w}$. Es gilt zum Beispiel

$$\left\{\vec{v}+t\vec{w} \mid t \in \mathbb{R}\right\} = \left\{\vec{x} \in \mathbb{R}^n \mid \exists s \in \mathbb{R}\colon \vec{x} = \vec{v}+s\vec{w}\right\} = \left\{\vec{y} \in \mathbb{R}^n \mid \exists t \in \mathbb{R}\colon \vec{y} = \vec{v}+t\vec{w}\right\}.$$

Durch die Definition haben wir dem Wort »Gerade« eine eindeutige mathematische Bedeutung gegeben. Nun leiten wir aus der Definition einige Eigenschaften von Geraden her. Als erstes wollen wir wissen, wann zwei durch Fußpunkte und Richtungen gegebene Geraden übereinstimmen.

2.2 Beispiel Wir betrachten die Vektoren

$$\vec{v} = \begin{pmatrix} -3 \\ 1 \end{pmatrix}, \ \vec{w} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \ \vec{v}' = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \ \vec{w}' = \begin{pmatrix} -4 \\ -2 \end{pmatrix}$$

in $\mathbb{R}^2$ und die durch sie bestimmten Geraden

$$L = \left\{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\right\} \quad \text{und} \quad L' = \left\{\vec{v}' + t\vec{w}' \mid t \in \mathbb{R}\right\}.$$

Obwohl die vier Vektoren verschieden sind, stimmen die beiden Geraden L und L' überein (Abb. 2.3). Das liegt daran, dass die Richtungsvektoren sich nur um einen skalaren Faktor unterscheiden, nämlich $\vec{w}' = -2 \cdot \vec{w}$, und die zwei Fußpunkte jeweils auf beiden Geraden liegen. ◇

Anstatt die Gleichheit der Geraden in diesem Beispiel nachzurechnen, beweisen wir gleich die folgende allgemeine Aussage:

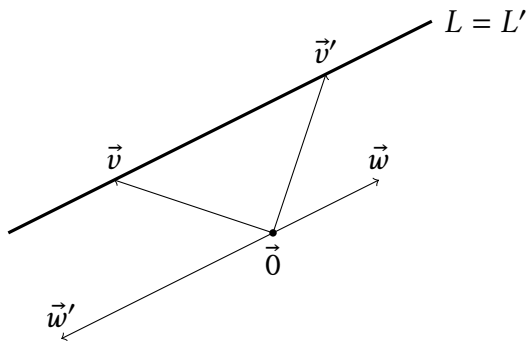


Abb. 2.3: Eine Geraden mit zwei verschiedenen Parametrisierungen

2.3 Satz *Es seien $\vec{v}, \vec{v}', \vec{w}, \vec{w}' \in \mathbb{R}^n$ mit $\vec{w}, \vec{w}' \neq \vec{0}$ und betrachte die Geraden*

$$L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\} \quad \text{und} \quad L' = \{\vec{v}' + t\vec{w}' \mid t \in \mathbb{R}\}.$$

Genau dann gilt $L = L'$, wenn $\vec{v}' \in L$ gilt und es $c \in \mathbb{R}$ gibt mit $\vec{w}' = c\vec{w}$.

Beweis. Dieser Satz ist unsere erste *Äquivalenz* (»genau dann, wenn«). Der Beweis hat deshalb zwei *Beweisrichtungen*, die wir getrennt behandeln:

($\Longrightarrow$): Es gelte $L = L'$. Wir haben $\vec{v}' \in L'$ (nämlich für $t = 0$) und wegen $L = L'$ dann auch $\vec{v}' \in L$. Es gibt also ein $t_0 \in \mathbb{R}$ mit $\vec{v}' = \vec{v} + t_0\vec{w}$. Ebenso folgt $\vec{v}' + \vec{w}' \in L$ (für $t = 1$). Also gibt es $t_1 \in \mathbb{R}$ mit $\vec{v}' + \vec{w}' = \vec{v} + t_1\vec{w}$. Daraus folgt

$$\vec{v} + t_0\vec{w} + \vec{w}' = \vec{v}' + \vec{w}' = \vec{v} + t_1\vec{w}.$$

Wir ziehen $\vec{v}$ auf beiden Seiten ab, bringen $t_0\vec{w}$ auf die rechte Seite und erhalten

$$\vec{w}' = t_1\vec{w} - t_0\vec{w} = (t_1 - t_0)\vec{w}.$$

Also folgt die Behauptung mit $c = t_1 - t_0$.

($\Longleftarrow$): Für die andere Beweisrichtung gelte $\vec{v}' \in L$ und sei $c \in \mathbb{R}$ mit $\vec{w}' = c\vec{w}$. Dann müssen wir daraus die Gleichheit $L = L'$ folgern.

Erst einmal gilt nach Voraussetzung $\vec{v}' \in L$. Es gibt also $t_0 \in \mathbb{R}$ mit $\vec{v}' = \vec{v} + t_0\vec{w}$. Aus $\vec{w}' = c\vec{w}$ folgt wegen $\vec{w}' \neq \vec{0}$ außerdem $c \neq 0$. Wir haben also

$$\begin{aligned} & \vec{v}' = \vec{v} + t_0\vec{w}, && \vec{w}' = c\vec{w} \\ \Longrightarrow \quad & \vec{v} = \vec{v}' - t_0\vec{w}, && \vec{w} = \tfrac{1}{c} \cdot \vec{w}'. \end{aligned}$$

Wir beweisen $L = L'$ wieder in zwei Schritten: Erst zeigen wir die Inklusion $L \subset L'$. Dazu nehmen wir einen Vektor in L und zeigen, dass er auch in L' liegen muss. Anschließend zeigen wir entsprechend die umgekehrte Inklusion $L' \subset L$. Sei also $\vec{v} + t\vec{w}$ ein beliebiger Vektor in L, dann gilt

$$\vec{v} + t\vec{w} = \vec{v}' - t_0\vec{w} + t\vec{w} = \vec{v}' + (t - t_0)\vec{w} = \vec{v}' + \frac{t - t_0}{c} \cdot \vec{w}' \in L'.$$

Das zeigt $L \subset L'$. Ist umgekehrt $\vec{v}' + t\vec{w}'$ ein Vektor in L', dann gilt

$$\vec{v}' + t\vec{w}' = \vec{v} + t_0\vec{w} + t\vec{w}' = \vec{v} + t_0\vec{w} + tc\vec{w} = \vec{v} + (t_0 + tc)\vec{w} \in L.$$

Damit ist alles bewiesen. ■

2.4 Korollar[4] *Es seien $\vec{v}, \vec{w} \in \mathbb{R}^n$ mit $\vec{w} \neq \vec{0}$ und sei $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$. Dann gilt $L = \{\vec{x} + t\vec{w} \mid t \in \mathbb{R}\}$ für jedes $\vec{x} \in L$. In Worten: Jeder Punkt auf der Geraden L kann als Fußpunkt von L verwendet werden.*

Beweis. Übung. ■

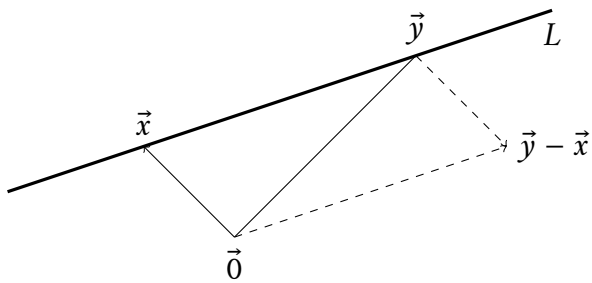
2.5 Satz (Verbindungsgerade) *Sind $\vec{x}, \vec{y} \in \mathbb{R}^n$ zwei verschiedene Vektoren, dann gibt es genau eine Gerade in $\mathbb{R}^n$, die $\vec{x}$ und $\vec{y}$ enthält.*

Beweis. Auch diese Aussage hat eine wiederkehrende logische Form: Sie behauptet *Existenz* (»es gibt«) und *Eindeutigkeit* (»genau eine«). Wieder führen wir den Beweis deshalb getrennt in zwei Teilen.

Existenz. Wir setzen $\vec{w} = \vec{y} - \vec{x}$. Da $\vec{x}$ und $\vec{y}$ verschieden sind, gilt $\vec{w} \neq \vec{0}$ und die Gerade

$$L = \{\vec{x} + t\vec{w} \mid t \in \mathbb{R}\}$$

enthält sowohl $\vec{x}$ (für $t = 0$) als auch $\vec{y} = \vec{x} + \vec{w}$ (für $t = 1$).



[4] Ein Korollar ist eine Aussage, die sich direkt oder mit einem kurzen Beweis aus einer anderen ergibt. Das Wort kommt aus dem Lateinischen und bedeutet »Kränzchen«. Es steht damit für so etwas wie ein kleines Geschenk.

Eindeutigkeit. Sei

$$L' = \{\vec{v}' + t\vec{w}' \mid t \in \mathbb{R}\}$$

irgendeine Gerade, die $\vec{x}$ und $\vec{y}$ enthält. Dann definieren wir L wie oben und zeigen, dass $L' = L$ gilt. Nach Kor. 2.4 gilt wegen $\vec{x} \in L'$ auch

$$L' = \{\vec{x} + t\vec{w}' \mid t \in \mathbb{R}\}.$$

Wegen $\vec{y} \in L'$ gibt es nun $t_0 \in \mathbb{R}$ mit $\vec{y} = \vec{x} + t_0\vec{w}'$, wobei $t_0 \neq 0$ wegen $\vec{x} \neq \vec{y}$. Es folgt

$$\vec{w}' = \frac{1}{t_0} \cdot (\vec{y} - \vec{x}) = \frac{1}{t_0} \cdot \vec{w}$$

und damit $L = L'$ nach Satz 2.3. ∎

Die eindeutig bestimmte Gerade L in Satz 2.5 heißt die **Verbindungsgerade** von $\vec{x}$ und $\vec{y}$. Es gilt

$$L = \{\vec{x}+t(\vec{y}-\vec{x}) \mid t \in \mathbb{R}\} = \{(1-t)\vec{x}+t\vec{y} \mid t \in \mathbb{R}\} = \{s\vec{x}+t\vec{y} \mid s,t \in \mathbb{R}, s+t = 1\}.$$

Die erste Gleichheit haben wir im Beweis gesehen, die anderen beiden sollte man sich klar machen.

2.6 Korollar *Es seien $\vec{x}, \vec{y} \in \mathbb{R}^n$ zwei verschiedene Vektoren, und es gelte $\vec{y} \neq \vec{0}$. Die folgenden Aussagen sind äquivalent:*

(1) Die Verbindungsgerade von $\vec{x}$ und $\vec{y}$ geht durch den Ursprung $\vec{0}$.

(2) Es gibt $c \in \mathbb{R}$ mit $\vec{x} = c\vec{y}$.

Wenn diese Bedingungen erfüllt sind, heißen die beiden Vektoren **kollinear**.

Beweis. Übung. ∎

2.7 Beispiel Wir wollen noch einen Zusammenhang zu den linearen Gleichungssystemen aus der ersten Vorlesung herstellen. Betrachten wir dazu folgende Frage: Gegeben seien die Vektoren

$$\vec{v} = \begin{pmatrix}2\\0\\1\end{pmatrix},\ \vec{w} = \begin{pmatrix}1\\-1\\0\end{pmatrix}\ ,\ \vec{v}' = \begin{pmatrix}0\\1\\2\end{pmatrix},\ \vec{w}' = \begin{pmatrix}-1\\2\\-1\end{pmatrix}$$

in $\mathbb{R}^3$, welche die beiden Geraden

$$L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\} \qquad \text{und} \qquad L' = \{\vec{v}' + t\vec{w}' \mid t \in \mathbb{R}\}$$

bestimmen. Wir wollen wissen, ob sich L und L' in einem Punkt schneiden und, wenn ja, den Schnittpunkt ausrechnen. Die Frage ist also, ob es Skalare s und t gibt, für welche die Gleichheit

$$\vec{v} + s\vec{w} = \vec{v}' + t\vec{w}'$$

erfüllt ist. (Warum wäre es falsch, auf beiden Seiten t zu verwenden?) Das bedeutet

$$\begin{pmatrix} 2+s \\ -s \\ 1 \end{pmatrix} = \begin{pmatrix} -t \\ 1+2t \\ 2-t \end{pmatrix}$$

Daraus bekommen wir für die drei Einträge die drei Gleichungen

$$\begin{aligned} 2 + s &= -t \\ -s &= 1 + 2t \\ 1 &= 2 - t \end{aligned}$$

in den Unbekannten s und t. Wir stellen die erweiterte Koeffizientenmatrix auf und lösen das System mit dem Eliminationsverfahren:

$$\left(\begin{array}{cc|c} 1 & 1 & -2 \\ -1 & -2 & 1 \\ 0 & 1 & 1 \end{array}\right) \overset{\text{⌐}\!\!\text{⌙}+}{\rightsquigarrow} \left(\begin{array}{cc|c} 1 & 1 & -2 \\ 0 & -1 & -1 \\ 0 & 1 & 1 \end{array}\right) \overset{\text{⌐}\!\!\text{⌙}+}{\rightsquigarrow} \left(\begin{array}{cc|c} 1 & 1 & -2 \\ 0 & -1 & -1 \\ 0 & 0 & 0 \end{array}\right)$$

Die zweite Gleichung sagt nun $t = 1$ und Einsetzen in die erste ergibt $s = -3$. Es gibt also einen Schnittpunkt von L und L', nämlich

$$L \cap L' = \{\vec{x}\}, \quad \vec{x} = \vec{v} - 3\vec{w} = \vec{v}' + \vec{w}' = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix} .$$

Wenn wir den Fußpunkt $\vec{v}$ von L stattdessen zum Beispiel durch den Vektor

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

ersetzen, ändert sich die letzte Gleichung zu $-1 = 2 - t$, was $t = 3$ ergibt. Da aus den ersten beiden Gleichungen $t = 1$ folgt, wird das Gleichungssystem unlösbar. Die beiden Geraden in $\mathbb{R}^3$ gehen aneinander vorbei, es gilt $L \cap L' = \emptyset$. ◇

2.3 Mengen

Wir erklären noch einmal die Konzepte und Notationen der Mengenlehre, die wir schon in der Vektorrechnung verwendet haben. Dieser Abschnitt ist zum Nachlesen zu Hause gedacht und wird in der Vorlesung nicht behandelt.

Mengen sind die Bausteine der Mathematik, aus denen alles andere aufgebaut wird.[5] Daraus ergibt sich paradoxerweise, dass man den Begriff »Menge« selbst nicht ohne Weiteres definieren kann, denn dazu müsste man ja wieder andere Begriffe verwenden. Solche formalen Probleme spielen für uns aber im Moment keine Rolle. Es genügt zu verstehen, wie man mit Mengen richtig umgeht. Eine Menge ist zusammengesetzt aus Elementen. Ein »Element« kann dabei praktisch alles sein, eine Zahl, ein Vektor, eine Funktion, oder sogar eine andere Menge.

Um eine Menge zu spezifizieren, gibt man die Elemente in Mengenklammern $\{\dots\}$ an. Einige vertraute Mengen sind:

- $\emptyset$ – die leere Menge
- $\mathbb{N} = \{1, 2, 3, \dots\}$ – die Menge der natürlichen Zahlen
- $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ – die Menge der natürlichen Zahlen mit Null
- $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ – die Menge der ganzen Zahlen
- $\mathbb{Q}$ – die Menge aller rationalen Zahlen (Brüche) $\frac{a}{b}$
- $\mathbb{R}$ – die Menge der reellen Zahlen

Ist ein Objekt x in einer Menge A enthalten, dann schreiben wir $x \in A$ und sagen, dass x ein **Element von** A ist. Ist x kein Element von A, dann schreiben wir $x \notin A$. Eine Menge ist vollständig dadurch beschrieben, welche Elemente sie enthält. Mit anderen Worten, wenn zwei Mengen A und B dieselben Elemente enthalten, dann gilt $A = B$. Die Elemente einer Menge haben insbesondere **keine Reihenfolge**. Es gilt also zum Beispiel $\{1, 2, 3, 4\} = \{2, 1, 4, 3\}$.

Definition Eine Menge B heißt eine **Teilmenge** einer Menge A, wenn jedes Element von B auch ein Element von A ist. Wir schreiben in diesem Fall[6]

$$B \subset A$$

andernfalls $B \not\subset A$. Gilt $B \subset A$ aber $A \neq B$, dann schreiben wir manchmal kurz $B \subsetneq A$ und nennen B eine **echte Teilmenge** von A.

[5] Die **Mengenlehre** wurde in der zweiten Hälfte des 19. Jahrhunderts entwickelt und geht wesentlich auf den deutschen Mathematiker Georg Cantor (1845–1918) zurück.

[6] Häufig findet man auch die Notation $B \subseteq A$, die das gleiche bedeutet wie $B \subset A$.

Aus Mengen kann man auf verschiedene Weisen neue Mengen bilden.

Definition Es seien A und B zwei Mengen.

- Die **Vereinigung** $A \cup B$ von A und B besteht aus allen Elementen, die in A *oder* in B enthalten sind.
- Der **Durchschnitt** (oder die *Schnittmenge*) $A \cap B$ von A und B besteht aus allen Elementen, die in A *und* in B enthalten sind. Die Mengen A und B heißen **disjunkt**, wenn sie keine gemeinsamen Elemente enthalten, wenn also ihr Durchschnitt leer ist.
- Das **Komplement** $A \setminus B$ von B in A besteht aus allen Elementen, die in A enthalten sind, aber nicht in B.[7]

Mengen, die aus anderen Mengen durch Vereinigung, Durchschnitt und Komplementbildung entstehen, werden auch als *Boolesche Kombinationen*[8] bezeichnet. Das geht natürlich auch mit mehr als zwei Mengen. Dabei muss man gegebenenfalls Klammern setzen, denn zum Beispiel sind $(A \cup B) \cap C$ und $A \cup (B \cap C)$ im allgemeinen nicht dasselbe. Genauer gilt für drei Mengen A, B und C immer

(1) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

(2) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

Solche Regeln kann man durch *Venn-Diagramme* visualisieren[9] (Abb. 2.4).

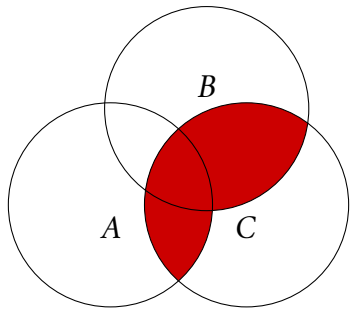


Abb. 2.4: Der rot markierte Bereich entspricht der Menge $(A \cup B) \cap C$

[7] Das Komplement $A \setminus B$ darf man auch bilden, wenn B keine Teilmenge von A ist. Allerdings gilt immer $A \setminus B = A \setminus (A \cap B)$.

[8] George Boole (1815–1864) war ein britischer Mathematiker, Philosoph und Logiker.

[9] nach dem britischen Philosophen und Logiker John Venn (1834–1923).

Häufig möchte man zu einer gegebenen Menge die Teilmenge aller Elemente bilden, die eine bestimmte Eigenschaft besitzen, zum Beispiel die Menge der geraden natürlichen Zahlen. Das schreibt man so:

$$\big\{n \in \mathbb{N} \mid n \text{ ist gerade}\big\}.$$

Ist allgemein A eine Menge und P eine (sinnvolle) Eigenschaft, dann können wir die Teilmenge der Elemente mit dieser Eigenschaft bilden:[10]

$$\big\{a \in A \mid a \text{ besitzt die Eigenschaft } P\big\}.$$

Ein Beispiel dafür sind die Geraden: Aus der Menge $\mathbb{R}^n$ aller Vektoren fassen wir diejenigen zusammen, die auf der Geraden L mit Fußpunkt $\vec{v}$ und Richtung $\vec{w} \neq \vec{0}$ liegen:

$$L = \big\{\vec{x} \in \mathbb{R}^n \mid \exists t \in \mathbb{R} \colon \vec{x} = \vec{v} + t\vec{w}\big\}.$$

Wir können das auch ohne den Existenzquantor als $L = \{\vec{v}+t\vec{w} \mid t \in \mathbb{R}\}$ schreiben.

Definition Seien A und B zwei Mengen. Das **kartesische Produkt**

$$A \times B$$

von A und B besteht aus allen **geordneten Paaren** (a, b) mit $a \in A$ und $b \in B$.

Dabei gilt $(a, b) = (a', b')$ genau dann, wenn $a = a'$ und $b = b'$ gelten. Insbesondere kommt es auf die Reihenfolge an. Das Paar (a, b) ist nicht dasselbe wie die Menge $\{a, b\}$, bei der die Reihenfolge egal ist.
Genauso bildet man für endlich viele Mengen $A_1, \dots, A_n$ das kartesische Produkt

$$A_1 \times \cdots \times A_n$$

dessen Elemente die **geordneten n-Tupel** (oder *Listen*) $(a_1, \dots, a_n)$ mit $a_i \in A_i$ für $i = 1, \dots, n$ sind. Zu jeder Menge A und $n \in \mathbb{N}$ ist außerdem

$$A^n = \underbrace{A \times \cdots \times A}_{\text{n } \textit{mal}}$$

das n-fache kartesische Produkt der Menge A mit sich selbst, die **n-te kartesische Potenz von** A, die wir für $A = \mathbb{R}$ schon als Menge von Vektoren verwendet haben.

[10] Die Notation dafür ist nicht ganz einheitlich: Statt eines senkrechten Strichs wird oft auch ein Doppelpunkt benutzt.

3 Lösungsräume

Wir können mit dem Gauß-Algorithmus jedes lineare Gleichungssystem lösen. Jetzt wird es darum gehen, die Struktur der Lösungsmenge besser zu verstehen. Außerdem wollen wir die Verbindung mit der Geometrie in $\mathbb{R}^n$ herstellen.

3.1 Das Matrix-Vektor-Produkt

Wir vereinfachen noch einmal die Notation. Ein lineares Gleichungssystem

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n &= b_2 \\ \vdots \qquad\quad \vdots \quad \vdots \qquad\quad \vdots & \\ a_{m,1}x_1 + a_{m,2}x_2 + \ldots + a_{m,n}x_n &= b_m \end{aligned}$$

mit m Gleichungen in n Unbekannten, können wir unter Verwendung von Summenzeichen und Indizes so schreiben:

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \qquad (i = 1, \ldots, m).$$

(Die Summennotation mit dem Großbuchstaben Sigma ist vermutlich inzwischen bekannt: Sie steht abkürzend für $\sum_{i=1}^{n} c_i = c_1 + c_2 + \cdots + c_n$.)

Definition Ist A eine $m \times n$-Matrix (das heißt mit m Zeilen und n Spalten) mit reellen Einträgen a_{ij} und ist $\vec{v} \in \mathbb{R}^n$ ein Vektor, dann heißt der Vektor in $\mathbb{R}^m$ mit den Einträgen

$$\sum_{j=1}^{n} a_{ij}v_j \qquad (i = 1, \ldots, m)$$

das **Matrix-Vektor-Produkt** von A und $\vec{v}$, geschrieben

$$A \cdot \vec{v} \quad \text{oder} \quad A\vec{v}.$$

Damit das Matrix-Vektor-Produkt definiert ist, muss der Spaltenvektor $\vec{v}$ also genauso hoch sein, wie die Matrix A breit ist.

3.1 Beispiel Es gilt

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 \\ 0 \cdot 1 + 1 \cdot 2 + 2 \cdot 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 8 \end{pmatrix} \qquad \diamond$$

Wir schreiben $\mathrm{Mat}_{m\times n}(\mathbb{R})$ für die Menge aller reellen $m\times n$-Matrizen. Für alle $A, B \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ und alle $\vec{v}, \vec{w} \in \mathbb{R}^n$ und $c \in \mathbb{R}$ gelten die Rechenregeln

$$A \cdot (\vec{v} + \vec{w}) = A\vec{v} + A\vec{w}, \quad (A + B)\vec{v} = A\vec{v} + B\vec{v}, \quad A(c\vec{v}) = (cA)\vec{v} = c(A\vec{v}).$$

Das kann man nachrechnen, aber wir werden es später allgemeiner beweisen. Bilden wir nun aus einem linearen Gleichungssystem

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \qquad (i = 1, \ldots, m)$$

die Koeffizientenmatrix $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ und den Vektor $\vec{b} \in \mathbb{R}^m$ der rechten Seite, dann können wir umgekehrt die Unbekannten $x_1, \ldots, x_n$ in einen Vektor $\vec{x}$ schreiben und das lineare Gleichungssystem kurz in der Form

$$A\vec{x} = \vec{b}$$

notieren. Gesucht ist also ein Vektor $\vec{x} \in \mathbb{R}^n$, der sich mit A zum gegebenen Vektor $\vec{b} \in \mathbb{R}^m$ multipliziert.

3.2 Lösungsraum eines Gleichungssystems

Definition Es sei $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ und $\vec{b} \in \mathbb{R}^m$. Der **Lösungsraum** des linearen Gleichungssystems $A\vec{x} = \vec{b}$ ist die Menge

$$L = \left\{ \vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{b} \right\}$$

aller Lösungen des Systems. Falls $L \neq \emptyset$ gilt, dann nennen wir die Anzahl der freien Unbekannten des Systems die **Dimension** des Lösungsraums.

Die Anzahl der freien Unbekannten können wir ablesen, nachdem wir die Matrix A in Zeilenstufenform gebracht haben: Es ist die Anzahl der Spalten, in denen keine Stufe anfängt. Wir geben bald eine allgemeinere und bessere[1] Definition.

[1] Die Zeilenstufenform hängt davon ab, welche Umformungen wir genau vornehmen und von der Reihenfolge der Gleichungen (also der Zeilen von A). Wir haben noch nicht gezeigt, dass die Anzahl der Stufen davon unabhängig ist. Unsere Definition steht also auf etwas wackligen Füßen.

3.2 Beispiel Gegeben sei das lineare Gleichungssystem

$$x_1 + 2x_2 + 3x_3 + 4x_4 = 5$$
$$x_1 + 2x_2 + 4x_3 + 6x_4 = 8.$$

Wir bringen die erweiterte Koeffizientenmatrix in Zeilenstufenform:

$$\left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 6 & 8 \end{array}\right) \begin{array}{l} \\ \cdot(-1) \\ + \end{array} \rightsquigarrow \left(\begin{array}{cccc|c} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 1 & 2 & 3 \end{array}\right)$$

Hier fängt in der zweiten und vierten Spalte keine neue Stufe an. Durch Rückeinsetzen bekommen wir die Lösungen

$$x_1 = -2x_2 + 2x_4 - 4, \qquad x_3 = 3 - 2x_4$$

in den freien Unbekannten x_2 und x_4. Der Lösungsraum hat die Dimension 2. ◇

Besteht das lineare Gleichungssystem $A\vec{x} = \vec{b}$ aus m Gleichungen (die Anzahl der Zeilen von A), dann hat jede einzelne dieser Gleichungen einen Lösungsraum L_i. Der Lösungsraum L des ganzen Systems ist dann der Durchschnitt

$$L = L_1 \cap \cdots \cap L_m$$

der einzelnen Lösungsräume.

Definition Das lineare Gleichungssystem $A\vec{x} = \vec{b}$ heißt **homogen**, wenn $\vec{b}$ der Nullvektor ist. Andernfalls heißt es **inhomogen**.

Ein homogenes lineares Gleichungssystem $A\vec{x} = \vec{0}$ ist immer lösbar, es besitzt nämlich die **triviale Lösung** oder **Nulllösung** $\vec{x} = \vec{0}$. Die ist natürlich reichlich uninteressant. Das Eliminationsverfahren findet aber *alle* Lösungen, ausgedrückt durch die freien Unbekannten.

3.3 Satz (Unterbestimmte Gleichungssysteme)
Es sei $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ mit $m < n$. Dann hat der Lösungsraum des homogenen linearen Gleichungssystems

$$A\vec{x} = \vec{0}$$

mindestens die Dimension $n - m > 0$.

Beweis. Da das System homogen ist, ist der Lösungsraum nicht leer. Nach Satz 1.3 können wir die Koeffizientenmatrix A auf Zeilenstufenform bringen, was den

Lösungsraum nicht ändert. Eine Matrix mit m Zeilen hat höchstens m Stufen. Es gibt also mindestens $n - m$ Spalten, in denen keine Stufe anfängt und in denen damit freie Unbekannte stehen, deren Werte wir beliebig vorgeben können. ■

3.4 Satz (Struktursatz für inhomogene Gleichungssysteme)
Es sei $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$. Ist L der Lösungsraum des homogenen linearen Gleichungssystems $A\vec{x} = \vec{0}$ und L' der Lösungsraum des inhomogenen Systems $A\vec{x} = \vec{b}$ für ein $\vec{b} \in \mathbb{R}^m$, dann gilt entweder $L' = \emptyset$ oder für jede Lösung $\vec{v} \in L'$ gilt[2]

$$L' = \vec{v} + L = \{\vec{v} + \vec{w} \mid \vec{w} \in L\}.$$

Beweis. Ist $\vec{x} \in L'$, dann gilt $A\vec{x} = \vec{b}$. Wegen $A\vec{v} = \vec{b}$ folgt daraus $A(\vec{x} - \vec{v}) = A\vec{x} - A\vec{v} = \vec{0}$, also $\vec{x} - \vec{v} \in L$ und damit $\vec{x} \in \vec{v} + L$. Ist umgekehrt $\vec{w} \in L$, dann also $A\vec{w} = \vec{0}$ und damit $A(\vec{v} + \vec{w}) = A\vec{v} + A\vec{w} = \vec{b} + \vec{0} = \vec{b}$, also $\vec{v} + \vec{w} \in L'$. ■

3.3 Geraden in der Ebene

Der Vektor $\vec{v}$ in Satz 3.4 spielt dieselbe Rolle wie der Fußpunkt einer Geraden. Eine solche Gerade L ist ja gegeben durch den Fußpunkt $\vec{v} \in \mathbb{R}^n$ und einen Richtungsvektor $\vec{w} \in \mathbb{R}^n$, $\vec{w} \neq \vec{0}$ als die Menge $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$. Wir stellen, zunächst für $n = 2$, den Zusammenhang mit linearen Gleichungen her.

3.5 Satz *Für eine Teilmenge $L \subset \mathbb{R}^2$ sind folgende Aussagen äquivalent:*

(1) Die Menge L ist eine Gerade durch den Ursprung (also mit $\vec{0} \in L$).

(2) Die Menge L ist der Lösungsraum einer homogenen linearen Gleichung

$$a_1x_1 + a_2x_2 = 0$$

mit Koeffizienten $a_1, a_2 \in \mathbb{R}$, die nicht beide Null sind.

Beweis. Dieser Satz behauptet wieder eine Äquivalenz und der Beweis zerfällt in zwei Teile: Aus der ersten Aussage folgt die zweite, aus der zweiten folgt die erste.

(2)⇒(1). Wenn $a_1 \neq 0$ ist, dann ist $\vec{x} = (x_1, x_2)$ genau dann eine Lösung von $a_1x_1 + a_2x_2 = 0$, wenn $x_1 = -\frac{a_2}{a_1}x_2$ gilt. Dann ist also

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -\frac{a_2}{a_1}x_2 \\ x_2 \end{pmatrix} = x_2 \cdot \begin{pmatrix} -\frac{a_2}{a_1} \\ 1 \end{pmatrix}.$$

[2] Die Notation $\vec{v} + L$ (formal zunächst sinnlos, was ist die Summe eines Vektors und einer Menge?) ist eine Abkürzung für die Menge $\{\vec{v} + \vec{w} \mid \vec{w} \in L\}$ auf der rechten Seite.

Das heißt gerade, dass $\vec{x}$ auf der Geraden mit Fußpunkt $\vec{0}$ und Richtungsvektor $(-\frac{a_2}{a_1}, 1)$ liegt. Ist andererseits $a_1 = 0$, dann muss $a_2 \neq 0$ sein und die Gleichung lautet $a_2 x_2 = 0$. Das ist äquivalent zu $x_2 = 0$. Der Lösungsraum ist dann die Gerade mit Fußpunkt $\vec{0}$ und Richtungsvektor $(1, 0)$.

(1)⇒(2). Sei L eine Gerade mit $\vec{0} \in L$. Nach Kor. 2.4 können wir den Nullvektor als Fußpunkt nehmen und

$$L = \{t\vec{w} \mid t \in \mathbb{R}\}$$

für $\vec{w} = (w_1, w_2) \in \mathbb{R}^2$ mit $\vec{w} \neq \vec{0}$ schreiben. Wir setzen $a_1 = w_2$ und $a_2 = -w_1$. Das ist so gewählt, dass $a_1 w_1 + a_2 w_2 = 0$ gilt. Betrachte den Lösungsraum

$$M = \{\vec{x} \in \mathbb{R}^2 \mid w_2 x_1 - w_1 x_2 = 0\}.$$

Wie wir in der anderen Beweisrichtung gesehen haben, ist M eine Gerade. Da L und M beide $\vec{0}$ und $\vec{w}$ enthalten, stimmen sie nach Satz 2.5 überein. ■

3.6 Satz *Für eine Teilmenge $L \subset \mathbb{R}^2$ sind folgende Aussagen äquivalent:*

(1) Die Menge L ist eine Gerade.

(2) Die Menge L ist der Lösungsraum einer linearen Gleichung

$$a_1 x_1 + a_2 x_2 = b$$

mit Koeffizienten $a_1, a_2, b \in \mathbb{R}$, wobei a_1, a_2 nicht beide Null sind.

Beweis. (2)⇒(1). Es sei

$$L = \{\vec{x} \in \mathbb{R}^2 \mid a_1 x_2 + a_2 x_2 = b\}.$$

Nach Satz 3.5 ist $L_0 = \{\vec{x} \in \mathbb{R}^2 \mid a_1 x_1 + a_2 x_2 = 0\}$ eine Gerade durch den Ursprung. Es gibt also $\vec{w} \in \mathbb{R}^2$, $\vec{w} \neq \vec{0}$, mit $L_0 = \{t\vec{w} \mid t \in \mathbb{R}\}$. Weil a_1, a_2 nicht beide 0 sind, hat die Gleichung $a_1 x_1 + a_2 x_2 = b$ eine Lösung $\vec{v} \in \mathbb{R}^2$. Nach Satz 3.4 ist die Menge aller Lösungen L gerade die Gerade $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$.

(1)⇒(2). Es sei $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$ und setze $L_0 = \{t\vec{w} \mid t \in \mathbb{R}\}$, also die Gerade durch den Ursprung mit demselben Richtungsvektor wie L. Nach Satz 3.5 ist L_0 der Lösungsraum einer homogenen linearen Gleichung $a_1 x_1 + a_2 x_2 = 0$. Sei $b = a_1 v_1 + a_2 v_2$. Der Lösungsraum der Gleichung $a_1 x_1 + a_2 x_2 = b$ ist nach der anderen Beweisrichtung eine Gerade. Sie enthält $\vec{v}$ (nach Definition von b) und außerdem $\vec{v} + \vec{w}$, denn es gilt $a_1(v_1 + w_1) + a_2(v_2 + w_2) = a_1 v_1 + a_2 v_2 + a_1 w_1 + a_2 w_2 = b + 0 = b$. Nach Satz 2.5 stimmt sie daher mit L überein. ■

Bei einer einzigen Gleichung wie in (2) ist immer eine der beiden Unbekannten eine freie Unbekannte und der Lösungsraum hat die Dimension 1. Die Geraden sind also die **eindimensionalen Lösungsräume**.

Die Gleichung $a_1x_1 + a_2x_2 = b$ entspricht im Prinzip der aus der Schule bekannten Geradengleichung. Da heißen die beiden Unbekannten meistens x und y statt x_1 und x_2. Ist der Koeffizient von y ungleich 0, dann kann man ihn durch Teilen auf 1 normieren und die Gleichung in der Form $y = mx + b$ schreiben. Das sind alle Geraden in $\mathbb{R}^2$ bis auf diejenigen, die parallel zur y-Achse sind.

Aus Satz 3.6 haben wir jetzt zwei verschiedene Beschreibungen einer Geraden:

(1) eine **parametrische** (oder **explizite**) Beschreibung $\{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$;

(2) eine **implizite** Beschreibung als Lösungsraum einer linearen Gleichung $a_1x_1 + a_2x_2 = b$ in zwei Unbekannten.

3.4 Gleichungssysteme in zwei Unbekannten

Eine lineare Gleichung in zwei Unbekannten, deren Koeffizienten nicht beide Null sind, beschreibt nach Satz 3.6 eine Gerade. Ist nun

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

ein System von zwei solchen Gleichungen, dann liefert die erste Gleichung eine Gerade L_1 und die zweite eine Gerade L_2. Der Lösungsraum des Systems ist der Durchschnitt $L_1 \cap L_2$ der beiden Geraden. Wenn wir die erweiterte Koeffizientenmatrix

$$\left(\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{array}\right)$$

in Zeilenstufenform bringen, dann gibt es für die Gestalt dieser Stufenform im Wesentlichen drei Möglichkeiten, nämlich

(3.7)
$$\begin{array}{lll} (1) & \left(\begin{array}{cc|c} 1 & * & * \\ 0 & 1 & * \end{array}\right) & \\ (2) & \left(\begin{array}{cc|c} 1 & * & * \\ 0 & 0 & 0 \end{array}\right) & \text{oder} \quad \left(\begin{array}{cc|c} 0 & 1 & * \\ 0 & 0 & 0 \end{array}\right) \\ (3) & \left(\begin{array}{cc|c} 1 & * & * \\ 0 & 0 & 1 \end{array}\right) & \text{oder} \quad \left(\begin{array}{cc|c} 0 & 1 & * \\ 0 & 0 & 1 \end{array}\right) \end{array}$$

Wie wir wissen, hat der erste Typ eine eindeutige Lösung, der zweite hat unendlich viele Lösungen, und der dritte ist unlösbar. Für die beiden Geraden entspricht das geometrisch den Fällen

(1) $L_1 \neq L_2$ und $L_1 \cap L_2 \neq \emptyset$; die Geraden schneiden sich in einem Punkt.

(2) $L_1 = L_2 = L_1 \cap L_2$; die Geraden sind gleich.

(3) $L_1 \neq L_2$ und $L_1 \cap L_2 = \emptyset$; die Geraden sind parallel.[3]

Schreibt man die Geraden in parametrischer Form $L_1 = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$ und $L_2 = \{\vec{v}' + t\vec{w}' \mid t \in \mathbb{R}\}$ mit $\vec{v}, \vec{v}', \vec{w}, \vec{w}' \in \mathbb{R}^2$, $\vec{w}, \vec{w}' \neq \vec{0}$, dann wird der zweite Fall durch Satz 2.3 charakterisiert. Zusätzlich kann man sich überlegen, dass der dritte Fall genau dann vorliegt, wenn $\vec{w}$ und $\vec{w}'$ kollinear sind wie in Satz 2.3, jedoch $\vec{v} \notin L_2$ (oder, äquivalent, $\vec{v}' \notin L_1$) gilt (Übung).

3.8 Satz *Für den Lösungsraum eines linearen Gleichungssystems in zwei Unbekannten in $\mathbb{R}^2$ gibt es die folgenden Möglichkeiten:*

(1) ein einzelner Punkt;

(2) eine Gerade;

(3) die leere Menge;

(4) die ganze Ebene $\mathbb{R}^2$.

Beweis. Gegeben ein System aus m Gleichungen in zwei Unbekannten, dann ist die erweiterte Koeffizientenmatrix eine Matrix mit m Zeilen und drei Spalten. Das kann die Nullmatrix sein, wenn alle Koeffizienten 0 sind. Dann stellt das Gleichungssystem keine Bedingungen und es tritt Fall (4) ein. Ansonsten hat die Zeilenstufenform höchstens drei Stufen. Sind es genau drei Stufen, dann fängt die letzte Stufe auf der rechten Seite an und das System ist unlösbar (Fall (3)). Sind es nur eine oder zwei Stufen, dann sieht die Zeilenstufenform aus wie oben in (3.7), aber gefolgt von $m - 2$ zusätzlichen Nullzeilen, die keien Rolle spielen. Die drei Typen entsprechen den Fällen (1)–(3). Das haben wir schon diskutiert. ∎

[3] Zwei Geraden sind parallel, wenn ihre Richtungsvektoren kollinear sind. Wenn sich zwei verschiedene Geraden in der Ebene nicht schneiden, dann müssen sie parallel sein (siehe Übungen). Für Geraden im drei- oder höherdimensionalen Raum stimmt das nicht.

4 Lineare Unterräume in $\mathbb{R}^n$

In dieser Vorlesung setzen wir die Untersuchung von Lösungsräumen linearer Gleichungssysteme fort, wobei wir uns nur auf die Struktur konzentrieren und die Gleichungen zunächst vergessen.

4.1 Lineare Unterräume

Gegeben sei ein homogenes lineares Gleichungssystem

$$A\vec{x} = \vec{0}$$

aus m Gleichungen in n Unbekannten. Sind $\vec{u}, \vec{v} \in \mathbb{R}^n$ zwei Lösungen des Systems, dann ist auch $\vec{u} + \vec{v}$ eine Lösung, denn es gilt

$$A(\vec{u} + \vec{v}) = A\vec{u} + A\vec{v} = \vec{0} + \vec{0} = \vec{0}.$$

Außerdem ist für jede Lösung $\vec{u}$ und jedes $c \in \mathbb{R}$ auch $c \cdot \vec{u}$ wieder eine Lösung, denn es gilt

$$A(c\vec{u}) = c \cdot A\vec{u} = c \cdot \vec{0} = \vec{0}.$$

Für diese Eigenschaften des Lösungsraum gibt es einen allgemeinen Namen.

Definition Eine Teilmenge $U \subset \mathbb{R}^n$ heißt ein **linearer Unterraum** (oder kurz *Unterraum*), wenn sie die folgenden Eigenschaften hat:

(NL) Es gilt $U \neq \emptyset$.

(ADD) Für alle $\vec{u}, \vec{v} \in U$ gilt $\vec{u} + \vec{v} \in U$.

(SKM) Für alle $\vec{u} \in U$ und alle $a \in \mathbb{R}$ gilt $a \cdot \vec{u} \in U$.

Die Eigenschaften (ADD) und (SKM) sagen, dass U *abgeschlossen* unter der Vektoraddition und unter Skalarmultiplikation ist. Man kann diese beiden Eigenschaften auch in einer einzigen zusammenfassen:

(LK) Für alle $\vec{u}, \vec{v} \in U$ und alle $a, b \in \mathbb{R}$ gilt $a\vec{u} + b\vec{v} \in U$.

4.1 Beispiele (1) Bei jedem homogenen linearen Gleichungssystem ist der Lösungsraum ein linearer Unterraum. Eigenschaft (NL) ist erfüllt, weil ein homogenes System immer die Nulllösung $\vec{0}$ besitzt. Die Eigenschaften (ADD) und (SKM) haben wir gerade nachgerechnet.

(2) Jede Gerade durch den Ursprung in $\mathbb{R}^n$ ist ein linearer Unterraum. Denn eine solche Gerade hat (nach Kor. 2.4) die Form

$$L = \{t\vec{w} \mid t \in \mathbb{R}\}$$

für ein $\vec{w} \in \mathbb{R}^n$, $\vec{w} \neq \vec{0}$. Eigenschaft (NL) ist klar, denn es gilt $\vec{0} = 0 \cdot \vec{w} \in L$. Sind nun $\vec{u}, \vec{v} \in L$ und $a \in \mathbb{R}$, dann gibt es $t_0, t_1 \in \mathbb{R}$ mit $\vec{u} = t_0\vec{w}$ und $\vec{v} = t_1\vec{w}$ und es folgt (LK):

$$a\vec{u} + b\vec{v} = at_0\vec{w} + bt_1\vec{w} = (at_0 + bt_1)\vec{w} \in L.$$

(3) Die Menge $\{\vec{0}\}$, die nur aus dem Nullvektor besteht, ist ein linearer Unterraum, der **Nullraum**. Die drei Eigenschaften sind offensichtlich erfüllt. ◇

4.2 Lemma[1] *Jeder lineare Unterraum enthält den Nullvektor.*

Beweis. Sei $U \subset \mathbb{R}^n$ ein linearer Unterraum. Dann gilt $U \neq \emptyset$ nach (NL). Es gibt also einen Vektor $\vec{u} \in U$. Nach (SKM) gilt dann auch $\vec{0} = 0 \cdot \vec{u} \in U$. ■

Zum Beispiel ist eine Gerade in $\mathbb{R}^n$, die nicht durch den Nullpunkt geht, *kein* linearer Unterraum. Da jeder Unterraum den Nullvektor enthält, kann man Eigenschaft (NL) in der Definition auch durch die Forderung $\vec{0} \in U$ ersetzen.

4.2 Linearkombinationen

Ist $U \subset \mathbb{R}^n$ ein linearer Unterraum und sind $\vec{v}_1, \ldots, \vec{v}_m \in U$ und $c_1, \ldots, c_m \in \mathbb{R}$, dann gilt auch

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m \in U.$$

Das sieht man durch wiederholtes Anwenden von (LK): Als erstes liegt $c_1\vec{v}_1 + c_2\vec{v}_2$ in U, nach (LK), dann als nächstes auch $(c_1\vec{v}_1 + c_2\vec{v}_2) + c_3\vec{v}_3$, usw.

Definition Es seien $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{R}^n$ Vektoren. Jeder Vektor der Form

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m$$

mit $c_1, \ldots, c_m \in \mathbb{R}$ heißt eine **Linearkombination** von $\vec{v}_1, \ldots, \vec{v}_m$. Die Skalare $c_1, \ldots, c_m$ heißen die *Koeffizienten* der Linearkombination.

[1] Ein Lemma ist ein Hilfssatz, der oft eine wiederkehrende Tatsache ausdrückt. Zwischen Lemma und Satz gibt es keine scharfe Abgrenzung. Das Wort stammt aus dem Altgriechischen. Der Plural ist »Lemmata« (oder, wenn man mag, auch »Lemmas«).

Wir bezeichnen mit $\mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m)$ die Menge dieser Linearkombinationen, also

$$\mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m) = \big\{ c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m \mid c_1, \ldots, c_m \in \mathbb{R} \big\}.$$

4.3 Lemma *Für $\vec{v}_1, \ldots, \vec{v}_m \in \mathbb{R}^n$ ist die Menge $\mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m)$ der kleinste lineare Unterraum von $\mathbb{R}^n$, der die Vektoren $\vec{v}_1, \ldots, \vec{v}_m$ enthält. Das bedeutet:*

(1) $\mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m)$ *ist ein linearer Unterraum mit $\vec{v}_1, \ldots, \vec{v}_m \in \mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m)$.*

(2) Ist $U \subset \mathbb{R}^n$ ein linearer Unterraum und gilt $\vec{v}_1, \ldots, \vec{v}_m \in U$, dann gilt auch $\mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m) \subset U$.

Beweis. (1) Es gilt $\vec{v}_1 \in \mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m)$, denn $\vec{v}_1$ ist die Linearkombination $\vec{v}_1 = 1 \cdot \vec{v}_1 + 0 \cdot \vec{v}_2 + \cdots + 0 \cdot \vec{v}_m$. Genauso folgt $\vec{v}_2, \ldots, \vec{v}_m \in \mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m)$. Damit ist auch (NL) erfüllt. Sind $\vec{x}, \vec{x}' \in \mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m)$, dann haben wir Darstellungen

$$\vec{x} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m \quad \text{und} \quad \vec{x}' = c_1'\vec{v}_1 + \cdots + c_m'\vec{v}_m$$

für Skalare $c_1, \ldots, c_m, c_1', \ldots, c_m' \in \mathbb{R}$ und damit auch

$$\vec{x} + \vec{x}' = (c_1 + c_1')\vec{v}_1 + \cdots + (c_m + c_m')\vec{v}_m \in \mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m).$$

Das zeigt (ADD). Für $c \in \mathbb{R}$ gilt außerdem

$$c\vec{x} = cc_1\vec{v}_1 + \cdots + cc_m\vec{v}_m \in \mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m),$$

also gilt auch (SKM).

(2) Es sei $U \subset \mathbb{R}^n$ ein linearer Unterraum mit $\vec{v}_1, \ldots, \vec{v}_m \in U$. Dann enthält U auch alle Linearkombinationen von $\vec{v}_1, \ldots, \vec{v}_m$, wie wir schon bemerkt haben. Mit anderen Worten, es gilt $\mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m) \subset U$. ∎

Definition Der lineare Unterraum $\mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_m)$ heißt **der von $\vec{v}_1, \ldots, \vec{v}_m$ aufgespannte Unterraum von $\mathbb{R}^n$** oder kurz[2] der **Spann von $\vec{v}_1, \ldots, \vec{v}_m$**.

4.4 Beispiele (1) Für einen einzelnen Vektor $\vec{w} \in \mathbb{R}^n$ mit $\vec{w} \neq \vec{0}$ ist

$$\mathrm{Lin}(\vec{w}) = \{ c \cdot \vec{w} \mid c \in \mathbb{R} \}$$

die Ursprungsgerade mit Richtungsvektor $\vec{w}$.

(2) Sind $\vec{v}, \vec{w} \in \mathbb{R}^n$ zwei Vektoren, beide ungleich $\vec{0}$, dann gibt es für $\mathrm{Lin}(\vec{v}, \vec{w})$ zwei unterschiedliche Fälle:

[2] Üblich sind auch die Bezeichnungen »erzeugter Unterraum« oder »lineare Hülle«.

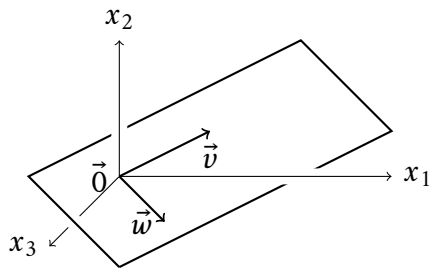


Abb. 4.1: Ebene in $\mathbb{R}^3$

- Sind $\vec{v}$ und $\vec{w}$ kollinear, dann gibt es $c_0 \in \mathbb{R}$ mit $\vec{v} = c_0\vec{w}$. Jedes Element von $\mathrm{Lin}(\vec{v}, \vec{w})$ hat die Form $c_1\vec{v} + c_2\vec{w} = c_1c_0\vec{w} + c_2\vec{w} = (c_0c_1 + c_2)\vec{w}$. Deshalb ist $\mathrm{Lin}(\vec{v}, \vec{w}) = \mathrm{Lin}(c_0\vec{w}, \vec{w}) = \mathrm{Lin}(\vec{w})$ wieder eine Ursprungsgerade.

- Wenn $\vec{v}$ und $\vec{w}$ nicht kollinear sind, dann ist $E = \mathrm{Lin}(\vec{v}, \vec{w})$ eine **Ebene** durch den Ursprung (siehe Abb. 4.1), die die beiden Geraden $\mathrm{Lin}(\vec{v})$ und $\mathrm{Lin}(\vec{w})$ enthält. Obwohl E von den beiden Vektoren $\vec{v}$ und $\vec{w}$ aufgespannt wird, ist E *nicht* die Vereinigungsmenge $\mathrm{Lin}(\vec{v}) \cup \mathrm{Lin}(\vec{w})$. Denn außer den Vielfachen von $\vec{v}$ und $\vec{w}$ enthält sie viele weitere Vektoren, zum Beispiel $\vec{v} + \vec{w}$.

(3) Die Vektoren

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots \ \vec{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

heißen die **Einheitsvektoren** in $\mathbb{R}^n$. Der Vektor $\vec{e}_i$ hat also eine 1 an der i-ten Stelle, sonst nur Nullen. Jeder Vektor $\vec{x} \in \mathbb{R}^n$ ist eine Linearkombination der Einheitsvektoren, nämlich

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_n \end{pmatrix} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n.$$

Es gilt also $\mathrm{Lin}(\vec{e}_1, \ldots, \vec{e}_n) = \mathbb{R}^n$. Für alle k mit $1 \leqslant k < n$ gilt außerdem

$$\begin{aligned}\mathrm{Lin}(\vec{e}_1,\dots,\vec{e}_k) &= \left\{\sum\nolimits_{i=1}^{k} c_i\vec{e}_i \mid c_1,\dots,c_k\in\mathbb{R}\right\}\\ &= \left\{(c_1,\dots,c_k,0,\dots,0) \mid c_1,\dots,c_k\in\mathbb{R}\right\}\\ &= \left\{\vec{x}\in\mathbb{R}^n \mid x_{k+1}=\dots=x_n=0\right\}\subset\mathbb{R}^n. \qquad \diamond\end{aligned}$$

Es sei $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ eine $m\times n$-Matrix und seien $\vec{a}_1,\dots,\vec{a}_m \in \mathbb{R}^n$ die Zeilenvektoren[3] von A und $\vec{b}_1,\dots,\vec{b}_n \in \mathbb{R}^m$ die Spaltenvektoren. Schematisch können wir das so darstellen:

$$A = \begin{pmatrix} — & \vec{a}_1 & — \\ & \vdots & \\ — & \vec{a}_m & — \end{pmatrix} = \begin{pmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_n \\ | & & | \end{pmatrix}$$

Der lineare Unterraum $\mathrm{Lin}(\vec{a}_1,\dots,\vec{a}_m)\subset\mathbb{R}^n$ heißt der **Zeilenraum** von A und $\mathrm{Lin}(\vec{b}_1,\dots,\vec{b}_n)\subset\mathbb{R}^m$ der **Spaltenraum**.

Für jeden Vektor $\vec{x}\in\mathbb{R}^n$ können wir das Matrix-Vektor-Produkt $A\vec{x}$ ausschreiben in der Form

$$A\vec{x} = \begin{pmatrix} | & | & & | \\ \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \\ | & | & & | \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\vec{b}_1 + \dots + x_n\vec{b}_n.$$

In Worten: Der Vektor $A\vec{x}$ ist eine Linearkombination der Spaltenvektoren von A, wobei die Koeffizienten die Einträge von $\vec{x}$ sind.

Wir können diese Erkenntnis auf lineare Gleichungssysteme anwenden: Eine *Lösung* des Systems $A\vec{x}=\vec{b}$ ist dasselbe wie eine Darstellung des Vektors $\vec{b}$ der rechten Seite als Linearkombination der Spaltenvektoren von A.

4.5 Satz *Das lineare Gleichungssystem $A\vec{x}=\vec{b}$ ist genau dann lösbar, wenn*

$$\vec{b}\in\mathrm{Lin}(\vec{b}_1,\dots,\vec{b}_n)$$

gilt, wobei $\vec{b}_1,\dots,\vec{b}_n\in\mathbb{R}^m$ die Spaltenvektoren der Koeffizientenmatrix A sind.

Beweis. Das haben wir gerade schon bewiesen. ■

[3] Eigentlich müssen wir die Zeilenvektoren $\vec{a}_1,\dots,\vec{a}_m$ in Spalten verwandeln, wenn wir darauf bestehen, dass $\mathbb{R}^n$ ein Raum von Spaltenvektoren ist. Im Moment ist das nicht so wichtig.

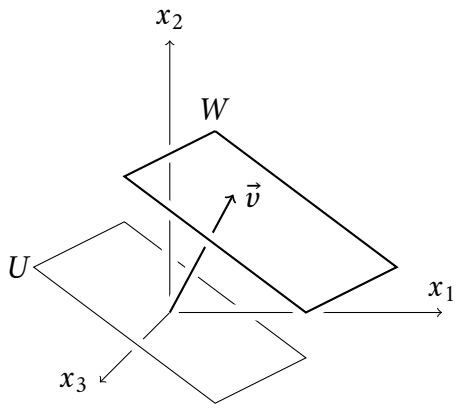


Abb. 4.2: Affiner Unterraum in $\mathbb{R}^3$

4.3 Affine Unterräume

Der Lösungsraum eines homogenen linearen Gleichungssystems ist ein linearer Unterraum. Der Lösungsraum eines *inhomogenen* Systems ist *kein* linearer Unterraum, weil er den Nullvektor nicht enthält. Allerdings haben wir das in Satz 3.4 schon untersucht: Ist L der Lösungsraum des homogenen linearen Gleichungssystems $A\vec{x} = \vec{0}$ und L' der Lösungsraum des inhomogenen Systems $A\vec{x} = \vec{b}$ für ein $\vec{b} \in \mathbb{R}^m$, dann gilt entweder $L' = \emptyset$ oder für jede Lösung $\vec{v} \in L'$ gilt $L' = \vec{v} + L = \{\vec{v} + \vec{w} \mid \vec{w} \in L\}$.

Definition Eine Teilmenge $W \subset \mathbb{R}^n$ heißt ein **affiner Unterraum**[4], wenn es einen linearen Unterraum $U \subset \mathbb{R}^n$ und einen Vektor $\vec{v} \in \mathbb{R}^n$ gibt mit

$$W = \vec{v} + U,$$

oder wenn $W = \emptyset$ gilt (Abb. 4.2).

4.6 Beispiele Jede Gerade in $\mathbb{R}^n$ ist ein affiner Unterraum. Der Lösungsraum eines linearen Gleichungssystems ist ein affiner Unterraum. Jeder lineare Unterraum ist auch ein affiner Unterraum (weil der Verschiebungsvektor $\vec{v}$ auch der Nullvektor sein darf). ◇

4.7 Lemma *Ein affiner Unterraum von $\mathbb{R}^n$ ist genau dann ein linearer Unterraum, wenn er den Nullvektor enthält.*

Beweis. Übung. ■

[4] Üblich ist auch die Bezeichnung »affin-linearer« Unterraum.

5 Abbildungen

In dieser Vorlesung führen wir Abbildungen ein und viele Sprechweisen, die damit zusammenhängen. Wir konzentrieren uns gleich auf Beispiele aus der linearen Algebra, die sich von den Funktionen aus der Analysis I unterscheiden.

5.1 Einführung

Es seien X und Y zwei Mengen. Eine **Abbildung**

$$f\colon X \to Y$$

ordnet jedem Element x in der *Quelle* X ein Element $f(x)$ im *Ziel* Y zu.

Die Begriffe »Abbildung« und »Funktion« bedeuten das Gleiche. Allerdings wird »Funktion« bevorzugt, wenn das Ziel eine Menge von Zahlen ist. In der Schule, und in der Vorlesung Analysis I, werden vor allem Funktionen $\mathbb{R} \to \mathbb{R}$ ausführlich behandelt, zum Beispiel so etwas wie

$$f\colon \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ x & \mapsto & 3x^2 + 7. \end{array} \right.$$

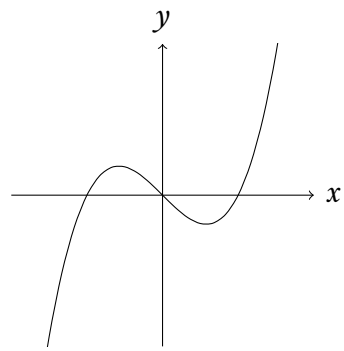
(Man beachte die zwei verschiedenen Stile von Pfeilen $\to$ und $\mapsto$ zwischen den Mengen und den Elementen.) Wir schreiben auch verkürzt $f(x) = 3x^2 + 7$.

Definition Sei $f\colon X \to Y$ eine Abbildung. Die Menge

$$\Gamma_f = \big\{(x, y) \in X \times Y \mid y = f(x)\big\}$$

heißt der **Graph von** f.

Der Graph einer Abbildung $\mathbb{R} \to \mathbb{R}$ ist eine Teilmenge der Ebene $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. In vielen Fällen kann man dann einen Ausschnitt des Graphen zeichnen. Das kennt man aus der Schule.



Graph der Funktion
$f\colon \mathbb{R} \to \mathbb{R},\ x \mapsto x^3 - x$

5.2 Vektorwertige Funktionen

In der Vektorrechnung kommen naheliegenderweise auch Funktionen vor, deren Werte Vektoren sind. Zum Beispiel gehört zu jeder Geraden $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$, gegeben durch den Fußpunkt $\vec{v} \in \mathbb{R}^3$ und die Richtung $\vec{w} \in \mathbb{R}^3$, $\vec{w} \neq \vec{0}$, die vektorwertige Abbildung

$$\varphi\colon \begin{cases} \mathbb{R} & \to & \mathbb{R}^3 \\ t & \mapsto & \vec{v} + t\vec{w} \end{cases}$$

welche die Gerade L *parametrisiert*. Sie weist also jedem Wert des Parameters t den entsprechenden Punkt auf L zu.
Die Gerade L ist die **Bildmenge** oder kurz das **Bild** der Abbildung φ:

$$\varphi(\mathbb{R}) = \big\{\varphi(t) \mid t \in \mathbb{R}\big\} = \big\{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\big\} = L.$$

Im Unterschied zu einer Funktion $\mathbb{R} \to \mathbb{R}$ kann man den Graph einer vektorwertigen Funktion nach $\mathbb{R}^n$ für $n \geqslant 3$ nicht mehr zeichnen.

Genauso können wir für zwei verschiedene Punkte $\vec{x}, \vec{y} \in \mathbb{R}^3$ die Verbindungsgerade $L' = \{\vec{x} + t(\vec{y} - \vec{x}) \mid t \in \mathbb{R}\}$ bilden, die durch die Abbildung

$$\psi\colon \begin{cases} \mathbb{R} & \to & \mathbb{R}^3 \\ t & \mapsto & \vec{x} + t(\vec{y} - \vec{x}) \end{cases}$$

parametrisiert wird. Die Gerade L' ist das Bild von ψ. Wir können die Abbildung ψ auch **einschränken** auf das Intervall $[0, 1]$. Die Bildmenge dieses Intervalls unter der Abbildung ψ ist

$$\psi([0,1]) = \big\{\vec{x} + t(\vec{y} - \vec{x}) \mid t \in \mathbb{R},\ 0 \leqslant t \leqslant 1\big\},$$

die *Verbindungsstrecke* von $\vec{x}$ und $\vec{y}$. Sie endet in den Punkten $\vec{x}$ (für $t = 0$) und $\vec{y}$ (für $t = 1$); die ganze Gerade L' ist die Verlängerung der Strecke über die beiden Endpunkte hinaus.

Als weiteres Beispiel betrachten wir die Abbildung

$$\pi\colon \begin{cases} \mathbb{R}^3 & \to & \mathbb{R}^2 \\ (x_1, x_2, x_3) & \mapsto & (x_1, x_2) \end{cases}$$

welche die letzte Koordinate einfach weglässt, eine **Koordinatenprojektion**.

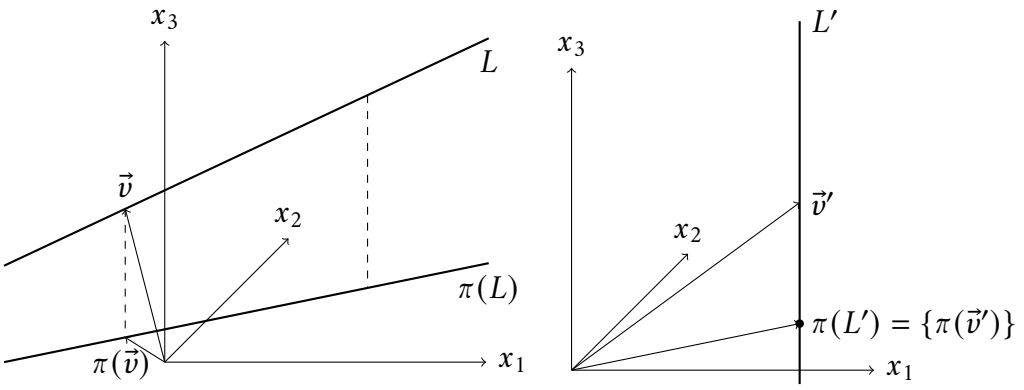


Abb. 5.1: Projektion einer Geraden aus dem Raum in die Ebene

5.3 Komposition

Wir können die Abbildungen π und φ wie oben **komponieren**, also hintereinander ausführen: Für $t \in \mathbb{R}$ ist

$$\pi\big(\varphi(t)\big) = \pi(\vec{v} + t\vec{w}) = \pi\begin{pmatrix} v_1 + tw_1 \\ v_2 + tw_2 \\ v_3 + tw_3 \end{pmatrix} = \begin{pmatrix} v_1 + tw_1 \\ v_2 + tw_2 \end{pmatrix} = \pi(\vec{v}) + t\pi(\vec{w}).$$

Aus φ und π wird so eine neue Abbildung

$$\pi \circ \varphi \colon \mathbb{R} \to \mathbb{R}^2,\ t \mapsto \pi\big(\varphi(t)\big).$$

Das Bild von $\pi \circ \varphi$ ist die Projektion der Geraden $L = \varphi(\mathbb{R})$ in $\mathbb{R}^3$ in die Ebene und damit eine Gerade in $\mathbb{R}^2$, sofern der Vektor $\pi(\vec{w})$ nicht der Nullvektor ist.

5.1 Beispiel Wir betrachten die beiden Geraden $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$ und $L' = \{\vec{v}' + t\vec{w}' \mid t \in \mathbb{R}\}$ mit

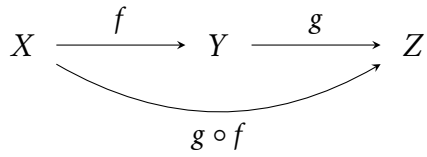
$$\vec{v} = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix},\ \vec{w} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix},\ \vec{v}' = \begin{pmatrix} 3 \\ 2 \\ 2 \end{pmatrix},\ \vec{w}' = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Die Projektion $\pi(L)$ ist die Gerade mit Fußpunkt $\pi(\vec{v})$ und Richtung $\pi(\vec{w})$, während die Projektion $\pi(L')$ nur aus dem Punkt $\pi(\vec{v}')$ besteht (siehe Abb. 5.1). Das liegt daran, dass $\pi(\vec{w}')$ der Nullvektor ist. ◇

Definition Sind $f\colon X \to Y$ und $g\colon Y \to Z$ zwei Abbildungen, bei denen das Ziel der ersten die Quelle der zweiten ist, dann ist die **Komposition von f und g** definiert durch

$$g \circ f\colon \begin{cases} X & \to & Z \\ x & \mapsto & g(f(x)) \end{cases}$$

(gelesen »g nach f«).



Die Abbildung, die zuerst an der Reihe ist, steht dabei rechts, weil das Argument, auf das die Funktion angewendet wird, immer nach rechts geschrieben wird: $(g \circ f)(x) = g(f(x))$ für $x \in X$.

In unserem Beispiel haben wir aus $\varphi\colon \mathbb{R} \to \mathbb{R}^3$ und $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$ die Komposition $\pi \circ \varphi\colon \mathbb{R} \to \mathbb{R}^2$ gebildet. Dagegen ist $\varphi \circ \pi$ überhaupt nicht sinnvoll definiert, weil die Quelle von φ und das Ziel von π nicht zusammenpassen.

5.4 Teilmengen und Potenzmenge

Wenn wir zu zwei Vektoren $\vec{v}, \vec{w} \in \mathbb{R}^n$ mit $\vec{w} \neq \vec{0}$ die Gerade $L = \{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\}$ bilden, dann können wir nicht nur die Parametrisierung φ wie oben betrachten, sondern auch die Zuordnung $(\vec{v}, \vec{w}) \mapsto L$. Das ist ein völlig anderer Typ von Abbildung. Ihre Quelle ist die Menge

$$\mathbb{R}^n \times (\mathbb{R}^n \setminus \{\vec{0}\}) = \left\{(\vec{v}, \vec{w}) \in \mathbb{R}^n \times \mathbb{R}^n \mid \vec{w} \neq \vec{0}\right\}$$

und ihr Ziel ist die Menge $\mathcal{P}(\mathbb{R}^n) = \{T \mid T \subset \mathbb{R}^n\}$ deren *Elemente* die *Teilmengen* von $\mathbb{R}^n$ sind. Wir können also

$$\Phi\colon \begin{cases} \mathbb{R}^n \times (\mathbb{R}^n \setminus \{\vec{0}\}) & \to & \mathcal{P}(\mathbb{R}^n) \\ (\vec{v}, \vec{w}) & \mapsto & \left\{\vec{v} + t\vec{w} \mid t \in \mathbb{R}\right\} \end{cases}.$$

schreiben und bekommen eine Abbildung Φ, die dem Paar $(\vec{v}, \vec{w})$ die Gerade mit Fußpunkt $\vec{v}$ und Richtung $\vec{w}$ zuordnet. Sie ist *mengenwertig*.

Definition Zu jeder Menge A ist die **Potenzmenge** $\mathcal{P}(A)$ eine Menge, deren Elemente die Teilmengen von A sind.

Dass aus den Teilmengen plötzlich Elemente werden, ist im ersten Moment sehr verwirrend. Dazu noch einige Beispiele.

5.2 Beispiele (1) Für jede Menge A sind die Menge A selbst und die leere Menge $\emptyset$ Teilmengen von A. Es gilt also immer

$$\emptyset \in \mathcal{P}(A) \quad \text{und} \quad A \in \mathcal{P}(A).$$

(2) Die Potenzmenge der Menge $A = \{a, b, c\}$ mit drei Elementen ist

$$\mathcal{P}(A) = \big\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\big\}.$$

Dabei ist die einelementige Menge $\{a\}$ nicht dasselbe wie a. Es gilt $a \in A$, aber $\{a\} \subset A$ und deshalb $\{a\} \in \mathcal{P}(A)$.

(3) Die Potenzmenge der leeren Menge ist $\mathcal{P}(\emptyset) = \{\emptyset\}$, besteht also aus einem Element! Die Potenzmenge einer Menge ist niemals leer.

(4) Die Elemente von $\mathcal{P}(\mathbb{R}^n)$ sind alle möglichen Mengen von Vektoren. Darunter befinden sich beispielsweise die Geraden und die linearen Unterräume, aber natürlich auch unzählige weitere Teilmengen. ◇

Es sei $f\colon X \to Y$ eine Abbildung. Für jede Teilmenge $A \subset X$ ist

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

die **Bildmenge von** A **unter** f, die wir schon betrachtet haben. Wie bei der Parametrisierung einer Geraden können wir das auch so schreiben:

$$f(A) = \{f(x) \mid x \in A\} = \{y \in Y \mid \exists x \in A\colon f(x) = y\}.$$

Für $B \subset Y$ ist die **Urbildmenge von** B **unter** f die Teilmenge

$$f^{-1}(B) = \big\{x \in X \mid f(x) \in B\big\}.$$

Für $y \in Y$ heißt jedes Element der Menge[1]

$$f^{-1}(\{y\}) = \{x \in X \mid f(x) = y\}$$

ein **Urbild** von y unter f.

[1]Die Schreibweise $f^{-1}(y)$ statt $f^{-1}(\{y\})$ ist naheliegend aber nicht korrekt. Vor allem darf man den Urbildoperator nicht mit der Umkehrabbildung verwechseln (siehe unten).

Während f jedem Element von X ein Element von Y zuordnet, findet die Zuordnung hier auf der Ebene von Teilmengen von X und Y statt. Wir können den **Bildoperator** und den **Urbildoperator** als Abbildungen

$$f\colon \begin{cases} \mathcal{P}(X) & \to & \mathcal{P}(Y) \\ A & \mapsto & f(A) \end{cases} \quad \text{und} \quad f^{-1}\colon \begin{cases} \mathcal{P}(Y) & \to & \mathcal{P}(X) \\ B & \mapsto & f^{-1}(B) \end{cases}$$

zwischen den Potenzmengen von X und Y auffassen. Das ist allerdings sehr formal ausgedrück. Wichtiger ist, dass man den richtigen Umgang mit Bild- und Urbildmengen an Beispielen übt.

5.3 Beispiele (1) Ist $\varphi\colon \mathbb{R} \to \mathbb{R}^3$ weiterhin eine Parametrisierung $\varphi(t) = \vec{v} + t\vec{w}$ einer Geraden L in $\mathbb{R}^n$ mit Fußpunkt $\vec{v}$ und Richtung $\vec{w} \neq \vec{0}$, dann hat zum Beispiel die Menge $B = \{\vec{v} + \vec{w}, \vec{v} + 2\vec{w}\}$ die Urbildmenge $\varphi^{-1}(B) = \{1, 2\}$. Ist andererseits $B' \subset \mathbb{R}^3$ eine Menge von Vektoren, die allesamt nicht auf L liegen, also mit $B' \cap L = \emptyset$, dann gilt auch $\varphi^{-1}(B') = \emptyset$.

(2) Unter der Projektion $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$, $(x_1, x_2, x_3) \mapsto (x_1, x_2)$ gilt für jeden Vektor $\vec{y} = (y_1, y_2) \in \mathbb{R}^2$ die Gleichheit

$$\pi^{-1}(\{\vec{y}\}) = \big\{(y_1, y_2, t) \mid t \in \mathbb{R}\big\}.$$

Denn auf der rechten Seite stehen genau die Vektoren, die $\vec{y}$ ergeben, wenn man den letzten Eintrag weglässt. ◇

5.5 Injektive, surjektive und bijektive Abbildungen

Wir kommen nun zu zwei wichtigen Eigenschaften von Abbildungen, die in der modernen Mathematik häufig verwendet werden, um ganz verschiedene Sachverhalte auszudrücken.

Definition Eine Abbildung $f\colon X \to Y$ zwischen zwei Mengen heißt

- **injektiv** oder eine **Injektion**, wenn verschiedene Elemente von X auch verschiedene Bilder unter f haben, wenn also gilt:

  $$\forall x, x' \in X\colon \quad \big(x \neq x' \implies f(x) \neq f(x')\big).$$

  Eine logisch äquivalente Formulierung (Kontraposition) ist:

  $$\forall x, x' \in X\colon \quad \big(f(x) = f(x') \implies x = x'\big).$$

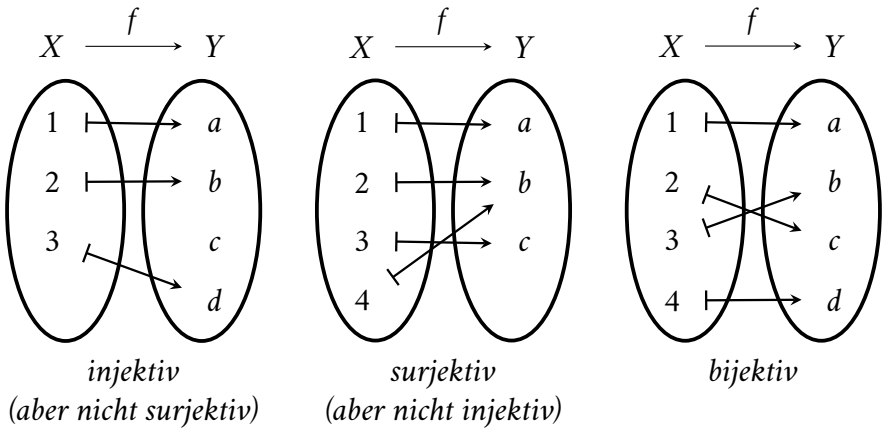


Abb. 5.2: Injektive/surjektive/bijektive Abbildungen

- **surjektiv** oder eine **Surjektion**, wenn jedes Element von Y im Bild von f liegt, wenn also gilt:

$$\forall y \in Y \exists x \in X\colon \ f(x) = y.$$

  Mit anderen Worten, die Bildmenge von f ist ganz Y. Das können wir auch kurz als Mengengleichheit $f(X) = Y$ schreiben.
- **bijektiv** oder eine **Bijektion**, wenn sie injektiv und surjektiv ist.

In Worten kann man diese Eigenschaften auch so ausdrücken: Eine Abbildung ist

- *injektiv*, wenn jedes Element der Zielmenge *höchstens ein Urbild* besitzt.
- *surjektiv*, wenn jedes Element der Zielmenge *mindestens ein Urbild* besitzt.
- *bijektiv*, wenn jedes Element der Zielmenge *genau ein Urbild* besitzt.

5.4 Beispiele (1) Die Parametrisierung

$$\varphi\colon \mathbb{R} \to \mathbb{R}^3,\ t \mapsto \vec{v} + t\vec{w}$$

der Geraden $L = \varphi(\mathbb{R})$ ist injektiv, was bedeutet, dass verschiedene Werte des Parameters auch verschiedene Punkte auf L ergeben. Die Parametrisierung läuft die Gerade genau einmal entlang. Rechnen wir das zur Übung noch einmal nach: Sind $t, t' \in \mathbb{R}$ zwei Werte mit $\varphi(t) = \varphi(t')$, dann bedeutet das $\vec{v} + t\vec{w} = \vec{v} + t'\vec{w}$. Wenn wir alles auf eine Seite bringen, bekommen wir $(t - t')\vec{w} = \vec{0}$. Da der

Richtungsvektor $\vec{w}$ nicht der Nullvektor ist, folgt daraus $t - t' = 0$, also $t = t'$. Das zeigt, dass φ injektiv ist.

Wir sehen auch: Ist $\vec{w} = \vec{0}$ der Nullvektor, dann ist Abbildung $t \mapsto \vec{v} + t\vec{w} = \vec{v}$ nicht injektiv, denn sie bildet alle Zahlen t auf denselben Punkt $\vec{v}$ ab. Es gilt beispielsweise $\varphi(1) = \varphi(2) = \vec{v}$. Ihr Bild ist keine Gerade, sondern besteht nur aus der einelementigen Menge $\{\vec{v}\}$.

In jedem Fall ist φ nicht surjektiv: Es gibt immer Vektoren in $\mathbb{R}^3$, die nicht auf der Geraden L, also nicht im Bild von φ liegen.

(2) Die Projektion

$$\pi\colon \mathbb{R}^3 \to \mathbb{R}^2,\ (x_1, x_2, x_3) \mapsto (x_1, x_2)$$

ist nicht injektiv, denn es gilt zum Beispiel $\pi(0,0,1) = \pi(0,0,2)$, aber $(0,0,1) \neq (0,0,2)$. Sie ist aber surjektiv, denn für jeden Vektor $\vec{y} = (y_1, y_2) \in \mathbb{R}^2$ gilt $\vec{y} = \pi(y_1, y_2, 0)$.

(3) Die Funktion $f\colon \mathbb{R} \to \mathbb{R}, x \mapsto x^2$ ist nicht injektiv, denn es gilt zum Beispiel $f(1) = f(-1) = 1$. Sie ist auch nicht surjektiv, da ihr Bild keine negativen Zahlen enthält. Dagegen ist $g\colon \mathbb{R} \to \mathbb{R}, x \mapsto x^3$ bijektiv, denn jede reelle Zahl besitzt eine eindeutige dritte Wurzel, wie man in der Analysis-Vorlesung beweist. ◇

Für jede Menge X gibt es die Abbildung

$$\mathrm{id}_X\colon \left\{ \begin{array}{ccc} X & \to & X \\ x & \mapsto & x \end{array} \right.$$

die jedes Element auf sich selbst abbildet, also

$$\forall x \in X\colon\ \mathrm{id}_X(x) = x$$

erfüllt. Sie wird die **Identität auf** X (oder manchmal auch *identische Abbildung*) genannt. Diese Abbildung scheint ziemlich uninteressant zu sein, wird aber verwendet, um verschiedene Sachverhalte über andere Abbildungen auszudrücken:

5.5 Satz *Sind $f\colon X \to Y$ und $g\colon Y \to X$ zwei Abbildungen mit*

$$g \circ f = \mathrm{id}_X,$$

dann ist f injektiv und g ist surjektiv.

Dabei ist $g \circ f = \mathrm{id}_X$ eine *Gleichheit zwischen Abbildungen*: Der Wert ist für alle $x \in X$ derselbe, also $g(f(x)) = (g \circ f)(x) = \mathrm{id}_X(x) = x$ für alle $x \in X$.

Beweis. Seien $x, x' \in X$ mit $f(x) = f(x')$. Dann folgt $x = \mathrm{id}_X(x) = g(f(x)) = g(f(x')) = \mathrm{id}_X(x') = x'$. Also ist f injektiv. Für $x \in X$ gilt außerdem $x = g(y)$ mit $y = f(x)$, was zeigt, dass g surjektiv ist. ∎

5.6 Satz *Es seien $f\colon X \to Y$ und $g\colon Y \to Z$ zwei Abbildungen.*

(1) Sind f und g injektiv bzw. surjektiv bzw. bijektiv, dann hat auch $g \circ f$ die entsprechende Eigenschaft.

(2) Ist $g \circ f$ injektiv, dann ist f injektiv.

(3) Ist $g \circ f$ surjektiv, dann ist g surjektiv.

(4) Ist $g \circ f$ bijektiv, dann ist f injektiv und g surjektiv.

Beweis. Übung. ∎

5.6 Umkehrabbildung

5.7 Satz *Sei $f\colon X \to Y$ eine Abbildung. Genau dann ist f bijektiv, wenn es eine Abbildung $g\colon Y \to X$ gibt mit*

$$g \circ f = \mathrm{id}_X \quad \textit{und} \quad f \circ g = \mathrm{id}_Y .$$

In diesem Fall ist die Abbildung g durch f eindeutig bestimmt.

Beweis. Angenommen f ist bijektiv. Dann definieren wir $g\colon Y \to X$ dadurch, dass wir jedem $y \in Y$ das eindeutige Element $x \in X$ mit $f(x) = y$ zuordnen. Für $y \in Y$ gilt dann also $f(g(y)) = y$ per Definition und damit $f \circ g = \mathrm{id}_Y$. Außerdem gilt für $x \in X$ auch $g(f(x)) = x$ per Definition, also $g \circ f = \mathrm{id}_X$.

Wenn umgekehrt g wie angegeben existiert, dann ist f nach Satz 5.5 injektiv und surjektiv, also bijektiv.

Zusätzlich müssen wir zeigen, dass g durch f eindeutig bestimmt ist. Seien also $g_1\colon Y \to X$ und $g_2\colon Y \to X$ zwei Abbildungen mit

$$g_i \circ f = \mathrm{id}_X \quad \text{und} \quad f \circ g_i = \mathrm{id}_Y \quad \text{für } i = 1, 2.$$

Für $y \in Y$ gilt dann $g_1(y) = g_1(\mathrm{id}_Y(y)) = g_1(f(g_2(y))) = \mathrm{id}_X(g_2(y)) = g_2(y)$, was $g_1 = g_2$ beweist. ∎

Definition Es sei $f\colon X \to Y$ eine Bijektion. Die eindeutig bestimmte Abbildung $g\colon Y \to X$ aus Satz 5.7 heißt die **Umkehrabbildung** oder **Inverse** von f. Sie erfüllt also

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

Satz 5.7 sagt gerade, dass eine Abbildung genau dann bijektiv ist, wenn sie eine Umkehrabbildung besitzt. Die Umkehrabbildung und der Urbildoperator sind nicht dasselbe, auch dann nicht, wenn die Abbildung bijektiv ist[2].

Sind $f\colon X \to Y$ und $g\colon Y \to Z$ zwei Bijektionen, dann dreht sich beim Übergang zur Umkehrabbildung die Reihenfolge der Komposition um:

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

(Am Morgen ziehen wir erst die Socken an und dann die Schuhe, aber am Abend müssen wir erst die Schuhe wieder ausziehen und dann die Socken!)

5.8 Beispiel Wir betrachten die Abbildungen

$$\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix} \quad \text{und} \quad \psi\colon \mathbb{R}^2 \to \mathbb{R}^2,\ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mapsto \frac{1}{2}\begin{pmatrix} y_1 + y_2 \\ y_1 - y_2 \end{pmatrix}.$$

Es gelten

$$\psi\Big(\varphi\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\Big) = \psi\begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} (x_1 + x_2) + (x_1 - x_2) \\ (x_1 + x_2) - (x_1 - x_2) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

und genauso

$$\varphi\Big(\psi\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\Big) = \psi\begin{pmatrix} \frac{1}{2}(y_1 + y_2) \\ \frac{1}{2}(y_1 - y_2) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}(y_1 + y_2) + \frac{1}{2}(y_1 - y_2) \\ \frac{1}{2}(y_1 + y_2) - \frac{1}{2}(y_1 - y_2) \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Es gilt also

$$\varphi \circ \psi = \mathrm{id}_{\mathbb{R}^2} \quad \text{und} \quad \psi \circ \varphi = \mathrm{id}_{\mathbb{R}^2},$$

was zeigt, dass $\psi = \varphi^{-1}$ die Umkehrabbildung von φ ist. Nach Satz 5.7 ist φ damit eine bijektive Abbildung. Prüfen wir die Injektivität und die Surjektivität von φ zur Übung noch einmal direkt nach: Sind $\vec{x}, \vec{x}' \in \mathbb{R}^2$ zwei Vektoren mit $\varphi(\vec{x}) = \varphi(\vec{x}')$, dann bedeutet das

$$\begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix} = \begin{pmatrix} x_1' + x_2' \\ x_1' - x_2' \end{pmatrix}.$$

Wenn wir die erste Gleichung auf die zweite addieren, bekommen wir $2x_1 = 2x_1'$, also $x_1 = x_1'$. Aus der ersten Gleichung folgt dann auch $x_2 = x_2'$. Es gilt also $\vec{x} = \vec{x}'$, was zeigt, dass φ injektiv ist.

[2] Ist $f\colon X \to Y$ eine bijektive Abbildung und $y \in Y$, dann ist $f^{-1}(y)$ ein Element von X, während die Urbildmenge $f^{-1}(\{y\})$ die einelementige Teilmenge $\{f^{-1}(y)\}$ von X ist.

Für die Surjektivität sei $\vec{y} \in \mathbb{R}^2$ beliebig, dann müssen wir zeigen, dass $\vec{x} \in \mathbb{R}^2$ gibt mit $\varphi(\vec{x}) = \vec{y}$. Das bedeutet die Gleichheit

$$\varphi(\vec{x}) = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Das ist ein lineares Gleichungssystem in den Unbekannten $\vec{x}$ mit rechter Seite $\vec{y}$. Wir können dieses System lösen und bekommen $x_2 = \frac{1}{2}(y_1 - y_2)$ und $x_1 = \frac{1}{2}(y_1 + y_2)$. Das ist gerade $\vec{x} = \psi(\vec{y})$, mit anderen Worten, wir haben die Umkehrabbildung auf diese Weise ausgerechnet. ◇

5.7 Ergänzungen

Zum Abschluss noch zwei ergänzende Bemerkungen:

1. Ein Begriff, der im Zusammenhang mit Abbildungen immer herumgeistert, ist die **Wohldefiniertheit**. Damit ist nur gemeint, dass eine Abbildung jedem Element der Quelle *genau ein* Element des Ziels zuordnen muss. Das folgende Beispiel ist typisch für die Probleme, die es dabei geben kann: Angenommen wir möchten eine Abbildung

$$f\colon \mathbb{Q} \to \mathbb{Z},\ \frac{a}{b} \mapsto a$$

definieren, die also jeder rationalen Zahl den Zähler zuordnet. So, wie es da steht, kann das nicht gehen, weil zum Beispiel $\frac{1}{2} = \frac{2}{4} = \frac{3}{6}$ *dieselbe Zahl* ist, aber repräsentiert durch verschiedene Brüche. Aus unserer Definition wird nicht klar, welchen Wert die Abbildung nun annehmen soll, sie ist *nicht wohldefiniert*. Die Sprechweise ist etwas irreführend. Korrekter wäre es zu sagen, dass die angegebene »Abbildung« gar keine ist, sie ist eine Mogelpackung.

2. Wenn man allzu sorglos Mengen von Mengen bildet, kann man in Teufels Küche kommen. Problematisch ist beispielsweise die »Menge aller Mengen«. Ist nämlich A die Menge aller Mengen, dann gilt insbesondere $A \in A$, da A ja selbst eine Menge ist. Die Menge A enthält sich also selbst als Element. Kein Problem? Nun, dann sollte aber auch das Folgende sinnvoll sein: Es sei $B = \{X \in A \mid X \notin X\}$. Das ist also die Menge aller Mengen, die sich nicht selbst als Element enthalten. Die Eine-Million-Euro-Frage lautet: Gilt $B \in B$ oder gilt $B \notin B$? Wenn man darüber nachdenkt, stellt man fest, dass beide Möglichkeiten der Definition von B direkt widersprechen! (Denn ist $B \in B$, dann ist B ja eine Menge, die sich selbst als Element enthält, und damit kein Element von B. Es folgt also $B \notin B$. Aber dann ist B ja eine Menge, die sich nicht selbst als Element enthält. Also doch $B \in B$…)

Solche Widersprüche haben Philosophen, Logikern und Mathematikern eine Zeitlang ziemliches Kopfzerbrechen bereitet und brachten das Ende der sogenannten »naiven Mengenlehre« mit sich (Stichwort: *Grundlagenkrise*). Wenn man die Mengenlehre systematisch korrekt aufbauen möchte, gibt es verschiedene Lösungsansätze. Die heute gebräuchliche Axiomatik der Mengenlehre nach Zermelo-Fraenkel (~1930) schließt unter anderem aus, dass sich eine Menge selbst als Element enthält.

In praktisch dieselbe Falle kann man auch in der Logik laufen: »Der Barbier rasiert die und nur die, die sich nicht selbst rasieren. Wer rasiert den Barbier?«[3] Dieses Phänomen ist in der Logik als *Russelsches Paradoxon* (oder korrekter *Russelsche Antinomie*) bekannt.

Wir verwenden des Öfteren griechische Buchstaben, vor allem bei Abbildungen. Hier eine Übersicht:

Zeichen	Name	Zeichen	Name
A, α	Alpha	N, ν	Ny
B, β	Beta	Ξ, ξ	Xi
Γ, γ	Gamma	O, o	Omikron
Δ, δ	Delta	Π, π	Pi
E, ε	Epsilon	P, ρ	Rho
Z, ζ	Zeta	Σ, σ	Sigma
H, η	Eta	T, τ	Tau
Θ, ϑ	Theta	Y, u	Ypsilon
I, ι	Iota	Φ, φ	Phi
$K, \varkappa$	Kappa	X, χ	Chi
Λ, λ	Lambda	Ψ, ψ	Psi
M, μ	My	Ω, ω	Omega

Die Zeichen, die auch im lateinischen Alphabet vorkommen, kann man natürlich nicht separat verwenden.

[3] »You can define the barber as ›one who shaves all those, and those only, who do not shave themselves.‹ The question is, does the barber shave himself?« aus: Bertrand Russell: The Philosophy of Logical Atomism (1918)

6 Aussagenlogik

Von alltäglichen Sachverhalten haben wir ein intuitives Verständnis, das uns vor vielen Irrtümern bewahrt. »Alle Hunde sind Haustiere, und alle Hunde bellen, aber auch die Katzen sind Haustiere, und folglich bellen sie.«[1] Wir merken sofort, dass das Unsinn ist. Das liegt aber nicht nur an der fehlerhaften Logik: Wir *wissen* eben, dass Katzen nicht bellen. In der Mathematik dagegen sind die Objekte oft so abstrakt, dass sie sich unserer Anschauung entziehen und wir logische Schlüsse nicht mehr so einfach auf ihre Plausibilität überprüfen können. Deshalb müssen wir beim Argumentieren sehr sorgfältig vorgehen.

Teil der Aussagenlogik ist eine logische Formelsprache, um mit Aussagen formal zu operieren, ähnlich wie mit mathematischen Formeln. Größere Zusammenhänge lassen sich aber nur in Worten verständlich ausdrücken, weshalb wir lange Reihen von Zeichen und Formeln möglichst vermeiden sollten. Wir müssen deshalb auch der deutschen Sprache ein sehr hohes Maß an Präzision abverlangen.

Wir vergessen für den Moment den mathematischen Gehalt und konzentrieren uns auf die Logik. Jede Aussage hat einen *Wahrheitswert*, nämlich wahr oder falsch (es sei denn, sie ist zu schwammig formuliert!). Betrachten wir die Aussagen

(P) »Anna ist Bastians Mutter.«

(Q) »Bastian ist Annas Sohn.«

(R) »Anna und Bastian sind verwandt.«

Wir können diese Aussagen auf verschiedene Weisen mit einander kombinieren:

Konjunktion (»und«)	$P \wedge Q$	»Anna ist Bastians Mutter und Bastian ist Annas Sohn.«
Disjunktion (»oder«)	$P \vee R$	»Anna ist Bastians Mutter oder Anna und Bastian sind verwandt.«
Negation (»nicht«)	$\neg P$	»Anna ist nicht Bastians Mutter.«
Implikation (»wenn…dann«)	$P \Longrightarrow R$	»Wenn Anna Bastians Mutter ist, dann sind Anna und Bastian verwandt.«
Äquivalenz (»genau dann…, wenn«)	$P \Leftrightarrow Q$	»Genau dann ist Anna Bastians Mutter, wenn Bastian Annas Sohn ist.«

[1] Aus einem Kneipengespräch in Umberto Ecos Roman *Das Foucaultsche Pendel*, Kap. 3

Das logische »oder« ist hier nicht ausschließend zu verstehen. Das heißt, $X \vee Y$ ist wahr, wenn eine der Aussagen X, Y oder alle beide wahr sind.

Wichtig ist der folgende grundlegende Unterschied zwischen diesen Aussagen: Ob etwa $P \vee Q$ richtig oder falsch ist, können wir nicht mit rein logischen Mitteln entscheiden. Es hängt von realen Sachverhalten ab. Dagegen ist die Implikation $P \Rightarrow R$ hier aus rein logischen Gründen wahr, ganz unabhängig davon, ob P wahr ist und wer Anna und Bastian überhaupt sind. In einer Implikation $X \Rightarrow Y$ heißt X die *Prämisse* und Y die *Konklusion*. Die Implikation ist nur dann falsch, wenn die Prämisse wahr ist, die Konklusion aber falsch. Ist dagegen die Prämisse falsch, so wird gar nichts behauptet und die Implikation ist immer wahr. (Deshalb darf man die Implikation $P \Rightarrow R$ in unserem Beispiel nicht mit »Anna ist Bastians Mutter, also sind Anna und Bastian verwandt« übersetzen.)

Die Richtigkeit der Implikation $P \Rightarrow R$ begründet sich hier also allein aus der Bedeutung der Worte. Wenn Anna nicht Bastians Mutter ist, dann ist P zwar falsch (und R könnte richtig oder falsch sein), aber die Implikation $P \Rightarrow R$ ist in jedem Fall logisch richtig. Genauso ist es mit der Äquivalenz $P \Leftrightarrow Q$. Von dieser Art sind im Prinzip auch alle wahren Aussagen der Mathematik. Sie gelten unabhängig davon, was wir in der Realität vorfinden.

Ein entsprechendes Beispiel aus der Mathematik ist die Aussage »Wenn eine natürliche Zahl n größer als 5 ist, dann ist sie auch größer als 2.« Diese Aussage ist wahr, unabhängig davon, ob n nun tatsächlich größer als 5 ist oder nicht.

Dagegen ist in den obigen Beispielen die Implikation $R \Rightarrow P$ *im Allgemeinen* falsch, denn Anna könnte ja auch mit Bastian verwandt sein, obwohl sie nicht seine Mutter ist.

Die vier logischen Operatoren $\wedge, \vee, \Rightarrow, \neg$ kann man kombinieren, wobei man die Aussagen gegebenenfalls klammern muss, um keine mehrdeutigen Aussagen zu bekommen. Zum Beispiel ist $P \wedge Q \Rightarrow R$ nicht eindeutig lesbar, denn $P \wedge (Q \Rightarrow R)$ ist nicht dasselbe wie $(P \wedge Q) \Rightarrow R$.

Manche zusammengesetzten Aussagen sind *immer wahr*, unabhängig davon, was die Bestandteile überhaupt sind, zum Beispiel $P \vee (\neg P)$. Diese Aussage ist in der Logik als *Satz vom ausgeschlossenen Dritten* (»tertium non datur«) bekannt, weil eben außer P und $\neg P$ keine dritte Möglichkeit besteht. Zusammengesetzte Aussagen, die immer wahr sind, werden **Tautologien** genannt. Eine Tautologie, die in der Mathematik häufig eine Rolle spielt, ist

$$(P \Rightarrow Q) \Longleftrightarrow (\neg Q \Rightarrow \neg P).$$

Zum Beispiel ist die Aussage »Wenn heute Mittwoch ist, dann ist morgen Donnerstag« logisch äquivalent zur Aussage »Wenn morgen nicht Donnerstag ist, dann

ist heute nicht Mittwoch«. Die Implikation $\neg Q \Rightarrow \neg P$ wird als **Kontraposition** der Implikation $P \Rightarrow Q$ bezeichnet. Manchmal lässt sich die Kontraposition einer Aussage leichter verstehen oder beweisen als die ursprüngliche Aussage.

Außer reinen Aussagen brauchen wir auch Ausdrücke, in denen manche Teile variabel vorkommen, sogenannte *Prädikate*. Zum Beispiel:

$P(x)$	»x ist eine ungerade Zahl« (für x aus den natürlichen Zahlen)
$Q(x, y)$	»x ist mit y verheiratet« (für x und y aus der Menge aller Menschen)
$R(x, y)$	»x ist größer als y« (für x und y aus den reellen Zahlen).

Zu einem Prädikat $P(x)$ gehört eine Menge M von *Werten*, die man für x einsetzen darf. Dabei kann man über die Variablen *quantifizieren*. Das haben wir in Beispielen schon gesehen.

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

bedeutet »für alle x in M gilt $P(x)$«. Das Zeichen $\forall$ ist ein umgedrehtes A für »alle«. Ebenso oft ist man nur daran interessiert, ob es wenigstens ein $x \in M$ gibt, für das die Aussage $P(x)$ wahr ist. In diesem Fall schreibt man

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

und sagt »es gibt ein x in M derart, dass $P(x)$ gilt«. Das umgedrehte E steht für »existiert«. Die Zeichen $\forall$ und $\exists$ werden *Allquantor* bzw. *Existenzquantor* genannt, beide sind **Quantoren**. Im Unterschied zu einer Aussage ist ein Prädikat nicht wahr oder falsch. Erst wenn man für die Variablen Elemente einsetzt oder über sie quantifiziert, wird aus dem Prädikat eine Aussage. Deshalb werden Prädikate auch als *Aussagefunktionen* bezeichnet.

Bei mehreren Quantoren kommt es auf die Reihenfolge an!

$$\forall x \in A \exists y \in B\colon P(x, y) \quad \text{und} \quad \exists y \in B \forall x \in A\colon P(x, y)$$

sind im allgemeinen nicht dasselbe. Ist zum Beispiel A die Menge aller Tierarten, B die Menge aller Zoologen und $P(x, y)$ die Aussage »Zoologe y kennt Tierart x«, dann sind das also die beiden Aussagen

$\forall x \in A \exists y \in B\colon P(x, y)$	:	»Zu jeder Tierart gibt es einen Zoologen, der diese Tierart kennt.«
$\exists y \in B \forall x \in A\colon P(x, y)$	:	»Es gibt einen Zoologen, der alle Tierarten kennt«.

Wie man sieht, ist das keineswegs dasselbe. Eigentlich ist das ganz einfach. Es wird aber dadurch erschwert, dass wir in der Normalsprache den Allquantor oft

nach hinten stellen. »Die Aussage $P(x)$ gilt für alle x« klingt meistens flüssiger als »Für alle x gilt die Aussage $P(x)$.« Das ist kein Problem, solange kein weiterer Quantor dazu kommt. Ein entsprechendes Beispiel ist die wahre Aussage

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

(das *archimedische Axiom*). Wenn man das in Worte fassen will, muss man sorgfältig formulieren. Schlecht ist zum Beispiel: »Es gibt eine Zahl $n \in \mathbb{N}$ die größer als x ist, für alle $x \in \mathbb{R}$.« Denn was ist gemeint? »Zu jeder reellen Zahl gibt es eine größere natürliche Zahl« oder »Es gibt eine natürliche Zahl, die größer als jede reelle Zahl ist«? Das erste ist die richtige Aussage oben, das zweite ist die Aussage mit umgedrehten Quantoren und natürlich falsch.

Aufpassen muss man auch bei der Verneinung: Die Aussage »Für alle x gilt $P(x)$« wird falsch, sobald $P(x)$ für ein einziges x nicht gilt. Es gilt deshalb

$$\neg\big(\forall x \in M\colon P(x)\big) \quad \Longleftrightarrow \quad \exists x \in M\colon \neg P(x).$$

Die Verneinung der Aussage »Jede natürliche Zahl ist ein Produkt von Primzahlen« ist also »Es gibt eine natürliche Zahl, die nicht Produkt von Primzahlen ist«.

Diesen ganzen Abschnitt kann man praktisch so zusammenfassen:

Genau lesen und genau formulieren!

6.1 Witz Eine Ingenieurin, eine Physikerin und eine Mathematikerin fahren im Zug durch Schottland und sehen im Vorbeifahren ein schwarzes Schaf. Darauf die Ingenieurin: »Wow, in Schottland sind die Schafe schwarz.« Die Physikerin: »Das kann man so allgemein nicht sagen. Aber jedenfalls gibt es in Schottland schwarze Schafe.« Darauf die Mathematikerin: »Eigentlich können wir doch bloß Folgendes sagen: In Schottland existiert mindestens ein Schaf, das von mindestens einer Seite schwarz ist.«

Etwas ernsthafter: Man soll es mit der Genauigkeit am Anfang ruhig ein bisschen übertreiben. Mit der Zeit bekommt man ein Gefühl dafür, was man ganz genau erklären sollte und was nicht. Ein mathematischer Beweis ist im Prinzip wie ein Dialog zwischen zwei Personen: Eine, die erklärt, und eine, die zuhört, bei Bedarf nachfragt oder widerspricht. Nur muss man beim Aufschreiben die möglichen Nachfragen schon mit bedenken.

»Eine Erklärung dient dazu, ein Mißverständnis zu beseitigen, oder zu verhüten – also eines, das ohne die Erklärung eintreten würde; aber nicht: jedes, welches ich mir vorstellen kann.« (L. Wittgenstein, *Philosophische Untersuchungen*,1953)

II

Lineare Algebra und Geometrie in $\mathbb{R}^n$

The beauty of mathematics only shows itself to more patient followers.

Maryam Mirzakhani (1977–2017)

7 Lineare Unabhängigkeit, Basen und Dimension

Die Einheitsvektoren $\vec{e}_1, \ldots, \vec{e}_n$ in $\mathbb{R}^n$ haben die schöne Eigenschaft, dass jeder andere Vektor $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ als Linearkombination

$$\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$$

dargestellt werden kann. Diese Darstellung ist außerdem eindeutig: Die Koeffizienten $x_1, \ldots, x_n$ sind ja genau die Einträge des Vektors $\vec{x}$. Die Einheitsvektoren bilden damit ein *Koordinatensystem* in $\mathbb{R}^n$.

Wenn wir es nicht mit $\mathbb{R}^n$ zu tun haben, sondern mit einem linearen Unterraum U, dann möchten wir ein Koordinatensystem innerhalb von U finden. Das führt auf den Begriff der Basis. Um zu verstehen, wie das funktioniert, brauchen wir erst einen anderen Begriff, der zu den wichtigsten der linearen Algebra gehört.

7.1 Lineare Unabhängigkeit

7.1 Beispiel Betrachte die drei Vektoren

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \ \vec{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \ \vec{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

in $\mathbb{R}^3$ und sei $U = \mathrm{Lin}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$. Es gilt hier

$$\vec{v}_3 = \tfrac{1}{2}(\vec{v}_1 + \vec{v}_2).$$

Deswegen wird der Unterraum U bereits von den beiden Vektoren $\vec{v}_1$ und $\vec{v}_2$ aufgespannt. Denn für jede Linearkombination $c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 \in U$ gilt

$$c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 = c_1\vec{v}_1 + c_2\vec{v}_2 + \tfrac{c_3}{2}(\vec{v}_1 + \vec{v}_2) = \left(c_1 + \tfrac{c_3}{2}\right)\vec{v}_1 + \left(c_2 + \tfrac{c_3}{2}\right)\vec{v}_2 \in \mathrm{Lin}(\vec{v}_1, \vec{v}_2).$$

Wir können also jede Linearkombination von $\vec{v}_1, \vec{v}_2, \vec{v}_3$ so umschreiben, dass $\vec{v}_3$ nicht mehr vorkommt. Den Vektor $\vec{v}_3$ können wir deshalb bei der Erzeugung von U auch weglassen. Tatsächlich gilt hier

$$\mathrm{Lin}(\vec{v}_1, \vec{v}_2, \vec{v}_3) = \mathrm{Lin}(\vec{v}_1, \vec{v}_2) = \mathrm{Lin}(\vec{v}_1, \vec{v}_3) = \mathrm{Lin}(\vec{v}_2, \vec{v}_3).$$

Von den drei Vektoren $\vec{v}_1, \vec{v}_2, \vec{v}_3$ ist im Hinblick auf den Spann immer einer überflüssig: Sie sind linear abhängig. Das drückt sich in der Gleichheit

$$\frac{1}{2}\vec{v}_1 + \frac{1}{2}\vec{v}_2 - \vec{v}_3 = \vec{0}$$

aus, die zur folgenden allgemeinen Definition passt. ◇

Definition (1) Ein m-Tupel $(\vec{v}_1, \dots, \vec{v}_m)$ aus $\mathbb{R}^n$ nennen wir ein **System** von Vektoren und lassen die Klammern meistens weg. Ein System von Vektoren hat also eine Reihenfolge und die Vektoren dürfen sich auch wiederholen. Die Zahl $m \in \mathbb{N}_0$ ist die **Länge** des Systems. (Das *leere* System hat die Länge 0.)

(2) Es sei $\vec{v}_1, \dots, \vec{v}_m$ ein System von Vektoren in $\mathbb{R}^n$. Jede Darstellung

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}$$

(mit $c_1, \dots, c_m \in \mathbb{R}$) des Nullvektors als Linearkombination von $\vec{v}_1, \dots, \vec{v}_m$ nennen wir eine **lineare Relation** zwischen $\vec{v}_1, \dots, \vec{v}_m$. Die Relation heißt **nicht-trivial**, wenn mindestens einer der Koeffizienten $c_1, \dots, c_m$ ungleich 0 ist.

(3) Das System von Vektoren $\vec{v}_1, \dots, \vec{v}_m$ heißt **linear abhängig**, wenn eine nicht-triviale lineare Relation zwischen $\vec{v}_1, \dots, \vec{v}_m$ existiert. Andernfalls heißt das System **linear unabhängig**.

In der Regel wird die Definition so verwendet: Das System $\vec{v}_1, \dots, \vec{v}_m$ von Vektoren ist linear unabhängig, wenn die Implikation

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0} \quad \Longrightarrow \quad c_1 = \cdots = c_m = 0 \tag{7.2}$$

für alle $c_1, \dots, c_m \in \mathbb{R}$ gilt.

7.3 Beispiel Die drei Vektoren $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^3$ in Beispiel 7.1 sind linear abhängig, denn es besteht zwischen ihnen die nicht-triviale lineare Relation $\frac{1}{2}\vec{v}_1 + \frac{1}{2}\vec{v}_2 - \vec{v}_3 = \vec{0}$. Dagegen sind zum Beispiel $\vec{v}_1$ und $\vec{v}_2$ alleine linear unabhängig. Denn sind $c_1, c_2 \in \mathbb{R}$ mit $c_1\vec{v}_1 + c_2\vec{v}_2 = \vec{0}$, dann gilt also

$$c_1\vec{v}_1 + c_2\vec{v}_2 = \begin{pmatrix} c_1 \\ 2c_1 \\ 3c_1 \end{pmatrix} + \begin{pmatrix} c_2 \\ 0 \\ c_2 \end{pmatrix} = \begin{pmatrix} c_1 + c_2 \\ 2c_1 \\ 3c_1 + c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} .$$

Das entspricht einem homogenen linearen Gleichungssystem mit drei Gleichungen in den Unbekannten c_1 und c_2. Dieses System besitzt nur die Nulllösung. (Aus der zweiten Gleichung folgt $c_1 = 0$ und aus der ersten damit auch $c_2 = 0$). ◇

Die Beobachtung in diesem Beispiel halten wir noch als Aussage fest.

7.4 Satz (Kriterium für lineare Unabhängigkeit in $\mathbb{R}^n$)
Genau dann ist ein System $\vec{a}_1, \dots, \vec{a}_m$ von Vektoren in $\mathbb{R}^n$ linear unabhängig, wenn das homogene lineare Gleichungssystem

$$\sum_{j=1}^{m} a_{ij} x_j = 0, \quad (i = 1, \dots, n)$$

nur die triviale Lösung besitzt.

Beweis. Denn eine Lösung dieses Systems ist genau dasselbe wie eine lineare Relation zwischen den Vektoren $\vec{a}_1, \dots, \vec{a}_m$. ∎

Noch ein paar weitere Beispiele zur linearen Unabhängigkeit:

7.5 Beispiele (1) Die Einheitsvektoren $\vec{e}_1, \dots, \vec{e}_n$ in $\mathbb{R}^n$ sind linear unabhängig. Das zugehörige lineare Gleichungssystem ist in diesem Fall einfach

$$x_1 = 0, \dots, x_n = 0$$

und hat nur die triviale Lösung.

(2) Ein System $\vec{v}_1, \dots, \vec{v}_m$, das den Nullvektor enthält, ist immer linear abhängig. Denn ist zum Beispiel $\vec{v}_1 = \vec{0}$, dann ist

$$1 \cdot \vec{0} + 0 \cdot \vec{v}_2 + \dots + 0 \cdot \vec{v}_m = \vec{0}$$

eine nicht-triviale lineare Relation.

(3) Ebenso ist jedes System, in dem ein Vektor zweimal vorkommt, linear abhängig. Denn ist zum Beispiel $\vec{v}_1 = \vec{v}_2$, dann ist

$$1 \cdot \vec{v}_1 + (-1) \cdot \vec{v}_2 + 0 \cdot \vec{v}_3 + \dots + 0 \cdot \vec{v}_m = \vec{0}$$

eine nicht-triviale lineare Relation.

(4) Ein System $\vec{v}, \vec{w}$ aus zwei Vektoren, mit $\vec{v}, \vec{w} \neq \vec{0}$ ist genau dann linear abhängig, wenn $\vec{v}$ und $\vec{w}$ kollinear sind, das heißt, wenn es $c \in \mathbb{R}$ gibt mit $c \neq 0$ und $\vec{v} = c\vec{w}$. Denn ist $\vec{v} = c\vec{w}$, dann ist $\vec{v} + (-c)\vec{w} = \vec{0}$ eine nicht-triviale lineare Relation zwischen $\vec{v}$ und $\vec{w}$. Sei umgekehrt

$$c\vec{v} + d\vec{w} = \vec{0}$$

eine nicht-triviale lineare Relation. Wäre $c = 0$, dann würde wegen $\vec{w} \neq \vec{0}$ auch $d = 0$ folgen, ein Widerspruch. Also ist $c \neq 0$ und es ist $\vec{v} = -\frac{d}{c}\vec{w}$.

(5) Es sei A eine Matrix mit m Zeilen und n Spalten, die in Zeilenstufenform vorliegt. Seien $\vec{v}_1, \dots, \vec{v}_m \in \mathbb{R}^n$ die Zeilenvektoren von A. Schematisch sieht das also so aus:

$$A = \begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * & * & \cdots & * & * & \cdots & * \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & 1 & \cdots & * & * & \cdots & * \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \cdots & 1 & * & \cdots & * \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} — & \vec{v}_1 & — \\ — & \vec{v}_2 & — \\ & \vdots & \\ — & \vec{v}_m & — \end{pmatrix}$$

Wenn hier die letzte Stufe in der k-ten Zeile steht, dann sind $\vec{v}_1, \dots, \vec{v}_k \neq \vec{0}$ und $\vec{v}_{k+1} = \cdots = \vec{v}_m = \vec{0}$. Die Stufenform impliziert in diesem Fall, dass das System $\vec{v}_1, \dots, \vec{v}_k$ linear unabhängig ist. Um das zu sehen, sei

$$c_1\vec{v}_1 + \cdots + c_k\vec{v}_k = \vec{0}$$

eine lineare Relation. Wenn die erste 1 in der ersten Zeile an der Stelle $(1, \ell)$ steht, dann ist also der ℓ-te Eintrag von $\vec{v}_1$ gleich 1, während der ℓ-te Eintrag in den Vektoren $\vec{v}_2, \dots, \vec{v}_k$ überall 0 ist. Aus der Relation folgt

$$c_1 \cdot \underbrace{v_{1l}}_{=1} + c_2 \cdot \underbrace{v_{2l}}_{=0} + \cdots + c_k \cdot \underbrace{v_{kl}}_{=0} = 0,$$

also $c_1 = 0$. Wir haben damit die Relation zu $c_2\vec{v}_2 + \cdots + c_k\vec{v}_k = \vec{0}$ verkürzt. Mit dem gleichen Argument folgt nun $c_2 = 0$ usw., bis insgesamt alle Koeffizienten 0 sind. ◇

7.6 Satz *Es sei $\vec{v}_1, \dots, \vec{v}_m$ ein System von Vektoren in $\mathbb{R}^n$.*

(1) Genau dann ist $\vec{v}_1, \dots, \vec{v}_m$ linear abhängig, wenn es einen Index k gibt mit $\mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_m) = \mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_{k-1}, \vec{v}_{k+1}, \dots, \vec{v}_m)$.

(2) Sind $\vec{v}_1, \dots, \vec{v}_m$ linear unabhängig und ist $\vec{v}_{m+1} \in \mathbb{R}^n$ ein Vektor, der nicht in $\mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_m)$ enthalten ist, dann sind auch $\vec{v}_1, \dots, \vec{v}_{m+1}$ linear unabhängig.

Beweis. (1) Angenommen $\vec{v}_1, \dots, \vec{v}_m$ sind linear abhängig. Dann gibt es eine nichttriviale Relation $c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}$. Gilt dabei $c_k \neq 0$, dann können wir

$$\vec{v}_k = -\frac{1}{c_k}(c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} + c_{k+1}\vec{v}_{k+1} + \cdots + c_m\vec{v}_m)$$

schreiben. Wir können deshalb $\vec{v}_k$ in jeder Linearkombination von $\vec{v}_1, \dots, \vec{v}_m$ ersetzen. Es gilt also $\mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_m) = \mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_{k-1}, \vec{v}_{k+1}, \dots, \vec{v}_m)$.

Umgekehrt gelte $\mathrm{Lin}(\vec{v}_1,\dots,\vec{v}_m) = \mathrm{Lin}(\vec{v}_1,\dots,\vec{v}_{k-1},\vec{v}_{k+1},\dots,\vec{v}_m)$. Dann hat $\vec{v}_k$ eine Darstellung $\vec{v}_k = c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} + c_{k+1}\vec{v}_{k+1} + \cdots + c_m\vec{v}_m$. Also ist

$$c_1\vec{v}_1 + \cdots + c_{k-1}\vec{v}_{k-1} + (-1)\cdot\vec{v}_k + c_{k+1}\vec{v}_{k+1} + \cdots + c_m\vec{v}_m = \vec{0}$$

eine nicht-triviale lineare Relation zwischen den Vektoren $\vec{v}_1,\dots,\vec{v}_m$.

(2) Es sei $U = \mathrm{Lin}(\vec{v}_1,\dots,\vec{v}_m)$. Falls $U = \mathbb{R}^n$, dann ist $\mathbb{R}^n \setminus U = \emptyset$ und es ist nichts zu zeigen. Andernfalls sei $\vec{v}_{m+1} \in \mathbb{R}^n \setminus U$ und sei

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m + c_{m+1}\vec{v}_{m+1} = \vec{0}$$

eine lineare Relation. Wäre $c_{m+1} \neq 0$, dann könnten wir

$$\vec{v}_{m+1} = -\frac{1}{c_{m+1}}(c_1\vec{v}_1 + \cdots + c_m\vec{v}_m) \in U$$

schreiben, im Widerspruch zu $\vec{v}_{m+1} \notin U$. Also muss $c_{m+1} = 0$ gelten. Dann folgt auch $c_1 = \cdots = c_m = 0$, weil $\vec{v}_1,\dots,\vec{v}_m$ linear unabhängig sind. ■

7.2 Basen

Definition Es sei $U \subset \mathbb{R}^n$ ein linearer Unterraum.

(1) Ein System $\vec{u}_1,\dots,\vec{u}_m$ von Vektoren aus U mit $U = \mathrm{Lin}(\vec{u}_1,\dots,\vec{u}_m)$ heißt ein **Erzeugendensystem** von U.

(2) Eine **Basis** von U ist ein **linear unabängiges Erzeugendensystem** von U.

7.7 Beispiele (1) Die Einheitsvektoren $\vec{e}_1,\dots,\vec{e}_n$ sind eine Basis von $\mathbb{R}^n$, die **Standardbasis**.

(2) Betrachten wir wieder die drei Vektoren

$$\vec{v}_1 = \begin{pmatrix}1\\2\\3\end{pmatrix},\ \vec{v}_2 = \begin{pmatrix}1\\0\\-1\end{pmatrix},\ \vec{v}_3 = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

in $\mathbb{R}^3$ und sei $U = \mathrm{Lin}(\vec{v}_1,\vec{v}_2,\vec{v}_3)$. Das System $\vec{v}_1,\vec{v}_2,\vec{v}_3$ ist linear abhängig (Beispiel 7.1), also keine Basis von U. Aber $\vec{v}_1,\vec{v}_2$ sind linear unabhängig und erfüllen immer noch $U = \mathrm{Lin}(\vec{v}_1,\vec{v}_2)$. Diese beiden Vektoren bilden also eine Basis von U. ◇

7.8 Satz (Basisauswahlsatz) *Es seien $\vec{u}_1, \dots, \vec{u}_m \in \mathbb{R}^n$ und sei $U = \operatorname{Lin}(\vec{u}_1, \dots, \vec{u}_m)$. Dann gibt es unter den Vektoren $\vec{u}_1, \dots, \vec{u}_m$ eine Basis von U.*

Beweis. Wenn $\vec{u}_1, \dots, \vec{u}_m$ linear unabhängig sind, dann bilden sie schon eine Basis und es gibt nichts zu tun. Andernfalls können wir nach Satz 7.6(1) von den Vektoren $\vec{u}_1, \dots, \vec{u}_m$ mindestens einen weglassen und haben immer noch ein Erzeugendensystem von U. Wenn wir diesen Schritt wiederholen, haben wir nach spätestens m Schritten ein linear unabhängiges Erzeugendensystem gefunden. ∎

7.9 Satz *Es sei $U \subset \mathbb{R}^n$ ein linearer Unterraum und seien $\vec{u}_1, \dots, \vec{u}_m \in U$. Die folgenden Aussagen sind äquivalent:*

(1) Das System $\vec{u}_1, \dots, \vec{u}_m$ ist eine Basis von U.

(2) Jeder Vektor $\vec{v} \in U$ hat eine eindeutige Darstellung

$$\vec{v} = x_1\vec{u}_1 + \cdots + x_m\vec{u}_m$$

als Linearkombination von $\vec{u}_1, \dots, \vec{u}_m$.

(3) Das System $\vec{u}_1, \dots, \vec{u}_m$ ist erzeugend und nicht verkürzbar, das heißt, kein echtes Teilsystem von $\vec{u}_1, \dots, \vec{u}_m$ erzeugt U.

(4) Das System $\vec{u}_1, \dots, \vec{u}_m$ ist linear unabhängig und nicht verlängerbar, das heißt, für jedes $\vec{u} \in U$ ist das System $\vec{u}_1, \dots, \vec{u}_m, \vec{u}$ der Länge $m + 1$ linear abhängig.

Beweis. (1) $\Rightarrow$ (2). Angenommen $\vec{u}_1, \dots, \vec{u}_m$ ist eine Basis von U. Dann ist jeder Vektor aus U eine Linearkombination von $\vec{u}_1, \dots, \vec{u}_m$, da jede Basis erzeugend ist. Es bleibt die Eindeutigkeit zu zeigen. Ist $\vec{v} \in U$ ein Vektor mit zwei Darstellungen

$$\vec{v} = x_1\vec{u}_1 + \cdots + x_m\vec{u}_m = y_1\vec{u}_1 + \cdots + y_m\vec{u}_m$$

dann folgt

$$(x_1 - y_1)\vec{u}_1 + \cdots + (x_m - y_m)\vec{u}_m = \vec{0}.$$

Da $\vec{u}_1, \dots, \vec{u}_m$ linear unabhängig sind, muss diese Relation trivial sein, das heißt, es gilt $x_1 = y_1, \dots, x_m = y_m$, was die Eindeutigkeit der Darstellung von $\vec{v}$ beweist.

(2) $\Rightarrow$ (3). Für jeden Index k hat der Vektor $\vec{u}_k$ die Darstellung $\vec{u}_k = 1 \cdot \vec{u}_k$. Da diese Darstellung nach (2) eindeutig ist, kann $\vec{u}_k$ nicht auch Linearkombination von $\vec{u}_1, \dots, \vec{u}_{k-1}, \vec{u}_{k+1}, \dots, \vec{u}_m$ sein. Wir können $\vec{u}_k$ also nicht weglassen.

(3) $\Rightarrow$ (4). Wäre das System $\vec{u}_1, \dots, \vec{u}_m$ linear abhängig, dann könnten wir nach Satz 7.6(1) einen der Vektoren weglassen und hätten immer noch ein Erzeugendensystem von U, im Widerspruch zu (3). Aus dem gleichen Grund ist das Sys-

tem nicht verlängerbar: Denn da $\vec{u}_1, \dots, \vec{u}_m$ bereits U erzeugen, muss jedes längere System nach Satz 7.6(1) linear abhängig sein.

(4) $\Rightarrow$ (1). Setze $W = \operatorname{Lin}(\vec{u}_1, \dots, \vec{u}_m)$. Wäre $W \neq U$, dann wäre für jedes $\vec{u}_{m+1} \in U \setminus W$ auch $\vec{u}_1, \dots, \vec{u}_m, \vec{u}_{m+1}$ linear unabhängig, nach Satz 7.6(2). Da das System nicht verlängerbar ist, gilt $W = U$, und $\vec{u}_1, \dots, \vec{u}_m$ ist eine Basis von U. ∎

7.10 Satz *Es sei $U = \operatorname{Lin}(\vec{w}_1, \dots, \vec{w}_m) \subset \mathbb{R}^n$ ein linearer Unterraum mit einem Erzeugendensystem der Länge m. Dann ist jedes System von Vektoren aus U, dessen Länge größer als m ist, linear abhängig.*

Beweis. Es sei $\vec{u}_1, \dots, \vec{u}_k$ ein System von Vektoren aus U mit $k > m$. Nach Voraussetzung hat jeder der Vektoren $\vec{u}_1, \dots \vec{u}_k$ eine Darstellung

$$\vec{u}_j = \sum_{i=1}^{m} a_{ij} \vec{w}_i \qquad (j = 1, \dots, k).$$

Die Koeffizienten a_{ij} bilden eine $m \times k$-Matrix. Dazu gehört das homogene lineare Gleichungssystem

$$\sum_{j=1}^{k} a_{ij} x_j = 0 \qquad (i = 1, \dots, m)$$

in Unbekannten $x_1, \dots, x_k$. Weil wir $k > m$ vorausgesetzt haben, hat dieses Gleichungssystem nach Satz 3.3 eine nicht-triviale Lösung $c_1, \dots, c_k$. Es folgt

$$\sum_{j=1}^{k} c_j \vec{u}_j = \sum_{j=1}^{k} c_j \left(\sum_{i=1}^{m} a_{ij} \vec{w}_i \right) = \sum_{i=1}^{m} \underbrace{\left(\sum_{j=1}^{k} a_{ij} c_j \right)}_{=0} \vec{w}_i = \vec{0}.$$

Also sind $\vec{u}_1, \dots, \vec{u}_k$ linear abhängig. ∎

7.11 Korollar *Jeder lineare Unterraum von $\mathbb{R}^n$ besitzt eine Basis.*

Beweis. Es sei $U \subset \mathbb{R}^n$ ein linearer Unterraum. Da $\mathbb{R}^n$ die Basis $\vec{e}_1, \dots, \vec{e}_n$ der Länge n hat, hat jedes linear unabhängige System in U nach Satz 7.10 höchstens die Länge n. Unter allen linear unabhängigen Systemen in U können wir also ein System $\vec{u}_1, \dots, \vec{u}_m$ von größtmöglicher Länge auswählen. Ein solches System ist nicht verlängerbar, also eine Basis von U nach Satz 7.9(4)⇒(1). ∎

7.12 Korollar (Invarianz der Basislänge) *Ist $U \subset \mathbb{R}^n$ ein linearer Unterraum und sind $\vec{u}_1, \dots, \vec{u}_k$ und $\vec{w}_1, \dots, \vec{w}_m$ zwei Basen von U, dann gilt $k = m$.*

Beweis. Da $\vec{u}_1, \dots, \vec{u}_k$ ein Erzeugendensystem ist und $\vec{w}_1, \dots, \vec{w}_m$ linear unabhängig, folgt $m \leqslant k$ aus dem vorangehenden Satz. Indem man die Rollen der beiden Systeme vertauscht, folgt genauso $k \leqslant m$, also insgesamt $k = m$. ∎

7.13 Satz (Basisergänzungssatz) *Es seien $U, V \subset \mathbb{R}^n$ zwei lineare Unterräume mit $U \subset V$. Ist $\vec{u}_1, \dots, \vec{u}_m \in U$ eine Basis von U, dann gibt es $k \in \mathbb{N}_0$ und Vektoren $\vec{v}_1, \dots, \vec{v}_k \in V$ derart, dass das System $\vec{u}_1, \dots, \vec{u}_m, \vec{v}_1, \dots, \vec{v}_k$ eine Basis von V ist.*

Beweis. Falls $U = V$ gilt, dann ist $\vec{u}_1, \dots, \vec{u}_m$ bereits eine Basis von V und es ist nichts zu zeigen. Es gelte also $U \subsetneq V$, und wir wählen ein $\vec{v}_1 \in V \setminus U$ beliebig. Nach Satz 7.6(2) ist dann das System $\vec{u}_1, \dots, \vec{u}_m, \vec{v}_1$ linear unabhängig. Betrachte nun den Unterraum $U_1 = \mathrm{Lin}(\vec{u}_1, \dots, \vec{u}_m, \vec{v}_1)$. Es gilt $U \subsetneq U_1$ und $\vec{u}_1, \dots, \vec{u}_m, \vec{v}_1$ ist eine Basis von U_1. Falls $U_1 = V$, dann sind wir fertig. Andernfalls wiederholen wir diese Konstruktion. Nach k Schritten haben wir dann einen Unterraum U_k mit Basis $\vec{u}_1, \dots, \vec{u}_m, \vec{v}_1, \dots, \vec{v}_k$ konstruiert. Ist andererseits etwa $\vec{w}_1, \dots, \vec{w}_l$ eine Basis von V, dann muss $m + k \leqslant l$ gelten, nach Satz 7.10. Nach spätestens $k \leqslant l - m$ Schritten muss in unserer Konstruktion also der Fall $U_k = V$ eintreten. ∎

7.14 Beispiel Unsere beiden Lieblingsvektoren $\vec{v}_1 = (1, 2, 3)$ und $\vec{v}_2 = (1, 0, -1)$ sind linear unabhängig und damit eine Basis des Unterraums $U = \mathrm{Lin}(\vec{v}_1, \vec{v}_2)$. Wir können $\vec{v}_1, \vec{v}_2$ zu einer Basis von $\mathbb{R}^3$ ergänzen, indem wir einen beliebigen Vektor in $\mathbb{R}^3 \setminus U$ dazunehmen, zum Beispiel $\vec{v}_3 = (1, 0, 0)$. ◇

7.3 Dimension

Da wir bewiesen haben, dass alle Basen eines linearen Unterraums dieselbe Länge haben, ist die folgende Definition sinnvoll.

Definition Die **Dimension** eines linearen Unterraums $U \subset \mathbb{R}^n$ ist die Länge einer Basis von U und wird mit $\dim(U)$ bezeichnet.

7.15 Beispiele (1) Der ganze Raum $\mathbb{R}^n$ hat die Basis $\vec{e}_1, \dots, \vec{e}_n$ und damit die Dimension n. (Es wäre wirklich etwas faul, wenn es anders wäre.)

(2) Für den Nullraum $\{\vec{0}\}$ ist das leere System der Länge 0 eine Basis. Er hat also die Dimension 0.

(3) Jede Gerade $\mathrm{Lin}(\vec{w})$ mit $\vec{w} \neq \vec{0}$ in $\mathbb{R}^n$ hat die Dimension 1.

(4) Sind $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$ zwei linear unabhängige Vektoren, dann hat die $U = \mathrm{Lin}(\vec{v}_1, \vec{v}_2)$ die Dimension 2. Es ist eine **Ebene**.

(5) Allgemeiner nennt man jeden linearen Unterraum der Dimension $n - 1$ in $\mathbb{R}^n$ eine **Hyperebene**. ◇

Sind U_1, U_2 zwei lineare Unterräume von $\mathbb{R}^n$, dann ist ihr Durchschnitt $U_1 \cap U_2$ wieder ein linearer Unterraum, der in U_1 und U_2 enthalten ist (siehe Übungen). Wir können andererseits auch die **Summe**

$$U_1 + U_2 \;=\; \left\{\vec{u}_1 + \vec{u}_2 \mid \vec{u}_1 \in U_1,\ \vec{u}_2 \in U_2\right\} \subset \mathbb{R}^n$$

der beiden Unterräume bilden. Das ist ebenfalls ein linearer Unterraum, der nun U_1 und U_2 beide enthält (siehe Übungen). Die wichtige *Dimensionsformel* setzt die Dimensionen all dieser Unterräume in Beziehung:

7.16 Satz (Dimensionsformel für Unterräume) *Es seien U_1 und U_2 zwei lineare Unterräume von $\mathbb{R}^n$.*

(1) *Falls $U_1 \subset U_2$, dann gilt $\dim(U_1) \leqslant \dim(U_2)$. Dabei ist $\dim(U_1) = \dim(U_2)$ genau dann, wenn $U_1 = U_2$ gilt.*

(2) *Es gilt*

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

Beweis. (1) Nach Satz 7.13 können wir jede Basis von U_1 zu einer Basis von U_2 verlängern. Daraus folgt $\dim(U_1) \leqslant \dim(U_2)$. Ist außerdem $U_1 \subsetneq U_2$, dann muss die Basis bei der Ergänzung länger werden und es folgt $\dim(U_1) < \dim(U_2)$.

(2) Auch das zeigen wir durch Basisergänzung. Es sei $d = \dim(U_1 \cap U_2)$ und sei $\vec{u}_1, \dots, \vec{u}_d$ eine Basis von $U_1 \cap U_2$. Nach Satz 7.13 können wir diese Basis zu einer Basis $\vec{u}_1, \dots, \vec{u}_d, \vec{v}_1, \dots, \vec{v}_k$ von U_1 (mit $\dim(U_1) = d + k$) und zu einer Basis $\vec{u}_1, \dots, \vec{u}_d, \vec{w}_1, \dots, \vec{w}_l$ von U_2 (mit $\dim(U_2) = d + l$) ergänzen. Wir behaupten, dass das System

$$\vec{u}_1, \dots, \vec{u}_d, \vec{v}_1, \dots, \vec{v}_k, \vec{w}_1, \dots, \vec{w}_l$$

eine Basis von $U_1 + U_2$ ist. Daraus folgt dann $\dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2) = (d + k) + (d + l) - d = d + k + l = \dim(U_1 + U_2)$. Außerdem ist klar, dass dieses System $U_1 + U_2$ erzeugt. Wir müssen nur seine lineare Unabhängigkeit beweisen. Es sei

$$\underbrace{a_1\vec{u}_1 + \cdots + a_d\vec{u}_d + b_1\vec{v}_1 + \cdots + b_k\vec{v}_k}_{\in U_1} + \underbrace{c_1\vec{w}_1 + \cdots + c_l\vec{w}_l}_{\in U_2} = \vec{0}$$

eine lineare Relation. Dann ist

$$c_1\vec{w}_1 + \cdots + c_l\vec{w}_l = -(a_1\vec{u}_1 + \cdots + a_d\vec{u}_d + b_1\vec{v}_1 + \cdots + b_k\vec{v}_k) \in U_1 \cap U_2.$$

Es gibt also Koeffizienten $\alpha_1, \dots, \alpha_d \in K$ mit

$$c_1\vec{w}_1 + \cdots + c_l\vec{w}_l = \alpha_1\vec{u}_1 + \cdots + \alpha_d\vec{u}_d.$$

Da das System $\vec{u}_1, \dots, \vec{u}_d, \vec{w}_1, \dots, \vec{w}_l$ eine Basis von U_2 ist, folgt daraus $c_1 = \cdots = c_l = 0$. In der gleichen Weise können wir auch $b_1 = \cdots = b_k = 0$ schließen. Von der ursprünglichen Relation bleibt dann nur noch $a_1\vec{u}_1 + \cdots + a_d\vec{u}_d = \vec{0}$ übrig, und da $\vec{u}_1, \dots, \vec{u}_d$ auch linear unabhängig sind, folgt $a_1 = \cdots = a_d = 0$. ∎

7.4 Lösungsräume

Nachdem wir jetzt die grundlegenden Eigenschaften von Basen und Dimension gezeigt haben, stellt sich die Frage, wie man eine Basis eines Unterraums findet. Der wichtigste Fall ist, dass der Unterraum der Lösungsraum eines homogenen linearen Gleichungssystems $A\vec{x} = \vec{0}$ ist. Wenn die Koeffizientenmatrix A in Zeilenstufenform vorliegt, dann kann in jeder Spalte, in der keine Stufe anfängt, die zugehörige Unbekannte bekanntlich frei gewählt werden. Das können wir jetzt genauer so ausdrücken:

7.17 Satz (Basis des Lösungsraums) *Gegeben sei ein homogenes lineares Gleichungssystem $A\vec{x} = \vec{0}$ mit A in Zeilenstufenform. Man erhält eine Basis des Lösungsraums, indem man jeweils eine der freien Unbekannten gleich* 1 *setzt und die anderen gleich* 0*. Wenn es keine freien Unbekannten gibt, ist der Lösungsraum der Nullraum.*

Beweis. Wenn es k freie Unbekannte gibt, dann erhalten wir in dieser Weise k verschiedene Lösungsvektoren $\vec{u}_1, \dots, \vec{u}_k$. Diese sind linear unabhängig, denn in jedem von ihnen gibt es einen Eintrag, der gleich 1 ist, in allen anderen aber 0 (wie in Beispiel 7.5(5)). Jede Lösung entsteht, indem man den freien Unbekannten Werte zuweist. Sind diese Werte $a_1, \dots, a_k$, dann ist die Linearkombination

$$a_1\vec{u}_1 + \cdots + a_k\vec{u}_k$$

die zugehörige Lösung. Also sind $\vec{u}_1, \dots, \vec{u}_k$ eine Basis des Lösungsraums. ∎

7.18 Beispiel Gegeben sei das Gleichungssystem mit Koeffizientenmatrix

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

in Unbekannten $x_1, \dots, x_5$. Hier gilt also

$$x_3 = 2x_4 - x_5 \quad \text{und} \quad x_1 = -2x_3 - x_4 - x_5$$

und x_2, x_4 und x_5 sind die freien Unbekannten. Der Lösungsraum hat die Basis

$$\begin{matrix} x_2 = 1 \\ x_4 = 0 \\ x_5 = 0 \end{matrix} \rightsquigarrow \vec{u}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{matrix} x_2 = 0 \\ x_4 = 1 \\ x_5 = 0 \end{matrix} \rightsquigarrow \vec{u}_2 = \begin{pmatrix} -5 \\ 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{matrix} x_2 = 0 \\ x_4 = 0 \\ x_5 = 1 \end{matrix} \rightsquigarrow \vec{u}_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}.$$

Für jede Wahl $x_2 = a_1$, $x_4 = a_2$ und $x_5 = a_3$ erhalten wir den Lösungsvektor $a_1\vec{u}_1 + a_2\vec{u}_2 + a_3\vec{u}_3$ als Linearkombination der $\vec{u}_1, \vec{u}_2, \vec{u}_3$. Außerdem sehen wir wie im Beweis, dass das System $\vec{u}_1, \vec{u}_2, \vec{u}_3$ linear unabhängig ist. Denn ist

$$c_1\vec{u}_1 + c_2\vec{u}_2 + c_3\vec{u}_3 = \begin{pmatrix} -5c_2 + c_3 \\ c_1 \\ 2c_2 - c_3 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

eine lineare Relation, dann liest man direkt $c_1 = c_2 = c_3 = 0$ ab. ◇

7.19 Korollar *Die Dimension des Lösungsraums eines homogenen linearen Gleichungssystems stimmt mit der Anzahl der freien Unbekannten überein.*

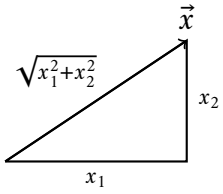
So hatten wir die Dimension eines Lösungsraums zunächst *ad hoc* definiert.

Beweis. Der vorangehende Satz sagt gerade, dass wir für jede freie Unbekannte einen Basisvektor erhalten, so dass die Behauptung folgt. ■

8 Norm und Skalarprodukt

8.1 Länge und Norm

In der Ebene $\mathbb{R}^2$ ist die Länge eines Vektors $\vec{x} = \binom{x_1}{x_2}$ die Zahl $\sqrt{x_1^2 + x_2^2}$.



Das ist, wenn man so will, der Satz des Pythagoras. Entsprechend definieren wir:

Definition Für einen Vektor $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ heißt

$$\|\vec{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

die **Länge** oder die **Norm** von $\vec{x}$.

Die Quadratwurzel ist dabei wie üblich die positive Quadratwurzel einer positiven reellen Zahl. Insbesondere gilt $\sqrt{x^2} = |x|$ für alle $x \in \mathbb{R}$.

8.1 Lemma *(1) Für alle $\vec{x} \in \mathbb{R}^n$ gelten $\|\vec{x}\| \geqslant 0$ und*

$$\|\vec{x}\| = 0 \quad \Longleftrightarrow \quad \vec{x} = \vec{0}.$$

(2) Für alle $\vec{x} \in \mathbb{R}^n$ und $c \in \mathbb{R}$ gilt $\|c\vec{x}\| = |c| \cdot \|\vec{x}\|$.

Beweis. Übung. ∎

Ein Vektor $\vec{x} \in \mathbb{R}^n$ heißt **normiert**, wenn er die Länge 1 hat. Jeden Vektor $\vec{x} \in \mathbb{R}^n$, der nicht der Nullvektor ist, können wir **normieren**, das heißt den normierten Vektor

$$\frac{1}{\|\vec{x}\|} \cdot \vec{x}$$

bilden.

8.2 Das Skalarprodukt

Der Norm liegt noch ein anderer Begriff zu Grunde.

Definition Für $\vec{x}, \vec{y} \in \mathbb{R}^n$ heißt

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{n} x_i y_i$$

das **Skalarprodukt** von $\vec{x}$ und $\vec{y}$.

Es gilt offenbar

$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} \quad \text{und damit} \quad \langle \vec{x}, \vec{x} \rangle = \|\vec{x}\|^2.$$

8.2 Beispiel Für jeden Vektor $\vec{x} \in \mathbb{R}^n$ gilt

$$\langle \vec{x}, \vec{e}_i \rangle = x_i. \qquad \diamond$$

8.3 Lemma (Rechenregeln für das Skalarprodukt) *Das Skalarprodukt ist **bilinear**, **symmetrisch** und **positiv definit**: Für alle $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$ und $c \in \mathbb{R}$ gelten:*

(1) $\langle \vec{x} + \vec{y}, \vec{z} \rangle = \langle \vec{x}, \vec{z} \rangle + \langle \vec{y}, \vec{z} \rangle$ *und* $\langle \vec{x}, \vec{y} + \vec{z} \rangle = \langle \vec{x}, \vec{y} \rangle + \langle \vec{x}, \vec{z} \rangle$

(2) $\langle c\vec{x}, \vec{y} \rangle = \langle \vec{x}, c\vec{y} \rangle = c\langle \vec{x}, \vec{y} \rangle$

(3) $\langle \vec{x}, \vec{0} \rangle = \langle \vec{0}, \vec{x} \rangle = 0$

(4) $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$

(5) $\langle \vec{x}, \vec{x} \rangle \geqslant 0$, *wobei* $\langle \vec{x}, \vec{x} \rangle = 0 \Leftrightarrow \vec{x} = \vec{0}$

Beweis. Übung. ∎

Das Skalarprodukt kann man umgekehrt wieder aus der Norm berechnen:

8.4 Lemma (Polarisationsformel) *Für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ gilt*

$$\langle \vec{x}, \vec{y} \rangle = \frac{1}{2}\big(\|\vec{x} + \vec{y}\|^2 - \|\vec{x}\|^2 - \|\vec{y}\|^2\big).$$

Beweis. Es gilt

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 - \|\vec{x}\|^2 - \|\vec{y}\|^2 &= \langle \vec{x} + \vec{y}, \vec{x} + \vec{y} \rangle - \langle \vec{x}, \vec{x} \rangle - \langle \vec{y}, \vec{y} \rangle \\ &= \langle \vec{x}, \vec{x} \rangle + \langle \vec{x}, \vec{y} \rangle + \langle \vec{y}, \vec{x} \rangle + \langle \vec{y}, \vec{y} \rangle - \langle \vec{x}, \vec{x} \rangle - \langle \vec{y}, \vec{y} \rangle \\ &= 2\langle \vec{x}, \vec{y} \rangle \end{aligned}$$

und damit die Behauptung. ∎

8.3 Cauchy-Schwarz- und Dreiecksungleichung

8.5 Satz (Cauchy-Schwarz-Ungleichung[1]) *Für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ gilt*

$$|\langle \vec{x}, \vec{y} \rangle| \leqslant \|\vec{x}\| \cdot \|\vec{y}\|.$$

Beweis. Für alle $a, b \in \mathbb{R}$ gilt

$$\begin{aligned} 0 \leqslant \|a\vec{x} - b\vec{y}\|^2 &= \langle a\vec{x} - b\vec{y}, a\vec{x} - b\vec{y} \rangle = \langle a\vec{x}, a\vec{x} \rangle - \langle a\vec{x}, b\vec{y} \rangle - \langle b\vec{y}, a\vec{x} \rangle + \langle b\vec{y}, b\vec{y} \rangle \\ &= a^2\|\vec{x}\|^2 - 2ab\langle \vec{x}, \vec{y} \rangle + b^2\|\vec{y}\|^2 \end{aligned}$$

und damit

$$2ab\langle \vec{x}, \vec{y} \rangle \leqslant a^2\|\vec{x}\|^2 + b^2\|\vec{y}\|^2.$$

Wir setzen nun $a = \|\vec{y}\|$ und $b = \|\vec{x}\|$ ein und erhalten

$$2\|\vec{x}\| \cdot \|\vec{y}\| \cdot \langle \vec{x}, \vec{y} \rangle \leqslant 2\|\vec{x}\|^2 \cdot \|\vec{y}\|^2.$$

Falls $\|\vec{x}\| = 0$ oder $\|\vec{y}\| = 0$, dann ist die Aussage des Satzes offensichtlich richtig. Wir können hier also $\|\vec{x}\| \cdot \|\vec{y}\| \neq 0$ annehmen und kürzen. Dann steht da

$$\langle \vec{x}, \vec{y} \rangle \leqslant \|\vec{x}\| \cdot \|\vec{y}\|$$

wie gewünscht. Schließlich gilt noch

$$-\langle \vec{x}, \vec{y} \rangle = \langle -\vec{x}, \vec{y} \rangle \leqslant \| -\vec{x}\| \cdot \|\vec{y}\| = \|\vec{x}\| \cdot \|\vec{y}\|$$

also auch $|\langle \vec{x}, \vec{y} \rangle| \leqslant \|\vec{x}\| \cdot \|\vec{y}\|$. ∎

Aus der Cauchy-Schwarz-Ungleichung folgt die wichtige Dreiecksungleichung.

8.6 Satz (Dreiecksungleichung) *Für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ gilt*

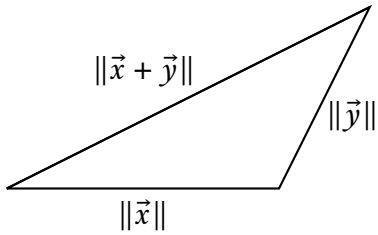
$$\|\vec{x} + \vec{y}\| \leqslant \|\vec{x}\| + \|\vec{y}\|.$$

Beweis. Nach der Cauchy-Schwarz-Ungleichung gilt

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 &= \langle \vec{x} + \vec{y}, \vec{x} + \vec{y} \rangle = \|\vec{x}\|^2 + 2\langle \vec{x}, \vec{y} \rangle + \|\vec{y}\|^2 \\ &\leqslant \|\vec{x}\|^2 + 2\|\vec{x}\| \cdot \|\vec{y}\| + \|\vec{y}\|^2 = \big(\|\vec{x}\| + \|\vec{y}\|\big)^2. \end{aligned}$$

Also folgt die Behauptung durch Wurzelziehen auf beiden Seiten. ∎

[1] nach Augustin-Louis Cauchy (1789–1857) und Hermann Schwarz (1843–1921)



Die Dreiecksungleichung ist deshalb wichtig, weil sie etwas über die kürzeste Strecke zwischen zwei Punkten aussagt.

Definition Für $\vec{x}, \vec{y} \in \mathbb{R}^n$ heißt $\|\vec{x}-\vec{y}\|$ der euklidische **Abstand** von $\vec{x}$ und $\vec{y}$.

Dieser Abstand hat die Eigenschaften einer **Metrik** auf $\mathbb{R}^n$, das heißt, es gelten

(1) $\|\vec{x}-\vec{y}\| \geqslant 0$

(2) $\|\vec{x}-\vec{y}\| = 0 \;\Leftrightarrow\; \vec{x} = \vec{y}$

(3) $\|\vec{x}-\vec{y}\| = \|\vec{y}-\vec{x}\|$ (*Symmetrie*)

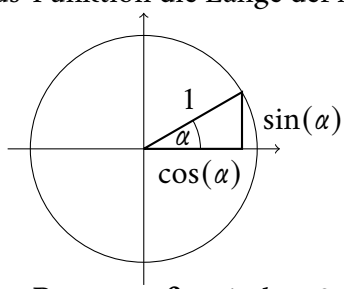
(4) $\|\vec{x}-\vec{z}\| \leqslant \|\vec{x}-\vec{y}\| + \|\vec{y}-\vec{z}\|$ (*Dreiecksungleichung*)

für alle $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$. Die Variante der Dreiecksungleichung in (4) folgt, indem man die normale Dreiecksungleichung auf $\|\vec{x}-\vec{z}\| = \|(\vec{x}-\vec{y})+(\vec{y}-\vec{z})\|$ anwendet. Sie sagt anschaulich, dass die Entfernung zwischen $\vec{x}$ und $\vec{z}$ nicht kürzer werden kann, wenn man den Umweg über $\vec{y}$ nimmt.

8.4 Winkelmessung

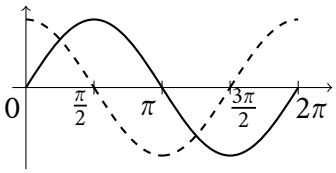
Während die Norm den Abstand zwischen Punkten angibt, kann man mit dem Skalarprodukt auch Winkel zwischen Vektoren messen. Dazu brauchen wir etwas Trigonometrie, die eigentlich nicht zur linearen Algebra gehört. Wir nehmen ein paar Anleihen bei der Schulmathematik bzw. der Analysis-Vorlesung.

Die Sinus-Funktion ordnet einem Winkel α in der Ebene die (signierte) Länge der *Gegenkathete* des entsprechenden rechtwinkligen Dreiecks mit Hypotenusenlänge 1 zu, und die Cosinus-Funktion die Länge der *Ankathete*:



Der Winkel α läuft dabei im **Bogenmaß** zwischen 0 und 2π einmal im Kreis.

Die Graphen von Sinus und Cosinus (gestrichelt) als Funktionen des Winkels zeigen die bekannten Wellenfunktionen:



Sinus und Cosinus erfüllen nach dem Satz des Pythagoras die Identität

$$\sin(\alpha)^2 + \cos(\alpha)^2 = 1$$

und parametrisieren den **Einheitskreis**

$$\{\vec{x} \in \mathbb{R}^2 \mid \|x\| = 1\} = \{(\cos(\alpha), \sin(\alpha)) \in \mathbb{R}^2 \mid \alpha \in [0, 2\pi)\}.$$

Einige Male werden wir die beiden *Additionstheoreme*

$$\begin{aligned}\sin(\alpha \pm \beta) &= \sin(\alpha)\cos(\beta) \pm \cos(\alpha)\sin(\beta)\\ \cos(\alpha \pm \beta) &= \cos(\alpha)\cos(\beta) \mp \sin(\alpha)\sin(\beta)\end{aligned}$$

brauchen, die später in der Analysis-Vorlesung bewiesen werden[2].

Für zwei normierte Vektoren in $\mathbb{R}^2$, also auf dem Einheitskreis, ist der eingeschlossene Winkel durch das Skalarprodukt gegeben: Sind

$$\vec{x} = \begin{pmatrix}\cos(\alpha)\\ \sin(\alpha)\end{pmatrix} \quad \text{und} \quad \vec{y} = \begin{pmatrix}\cos(\beta)\\ \sin(\beta)\end{pmatrix}$$

dann gilt nach dem Additionstheorem für den Cosinus

$$\langle \vec{x}, \vec{y} \rangle = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta) = \cos(\alpha - \beta).$$

Entsprechend können wir nun Winkel in $\mathbb{R}^n$ definieren: Die Cosinus-Funktion bildet das Intervall $[0, \pi]$ bijektiv auf das Intervall $[-1, 1]$ ab. Zu jedem Wert $-1 \leqslant \beta \leqslant 1$ gibt es also genau eine Zahl $\alpha \in [0, \pi]$ mit $\cos(\alpha) = \beta$. Sind $\vec{x}, \vec{y} \in \mathbb{R}^n$ zwei Vektoren, beide ungleich $\vec{0}$, dann gilt nach der Cauchy-Schwarz-Ungleichung

$$-1 \leqslant \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\|\|\vec{y}\|} \leqslant 1.$$

[2]Die Notation mit $\pm/\mp$ bedeutet, dass in der ersten Gleichung links und rechts dasselbe Vorzeichen steht, in der zweiten das umgekehrte.

Definition Für $\vec{x}, \vec{y} \in \mathbb{R}^n$ mit $\vec{x}, \vec{y} \neq \vec{0}$ definieren wir den **Winkel** zwischen $\vec{x}$ und $\vec{y}$ als die eindeutig bestimmte Zahl $\alpha \in [0, \pi]$ mit

$$\cos(\alpha) = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\| \|\vec{y}\|}$$

Dabei hat der so definierte Winkel kein Vorzeichen, das heißt, er hängt nicht von der Reihenfolge von $\vec{x}$ und $\vec{y}$ ab.

8.5 Orthogonale Vektoren

Für $\alpha \in [0, 2\pi]$ gilt $\cos(\alpha) = 0$ genau für $\alpha = \frac{\pi}{2} = 90°$. Deshalb definieren wir:

Definition Zwei Vektoren $\vec{x}, \vec{y} \in \mathbb{R}^n$ sind **orthogonal** oder stehen **senkrecht** aufeinander, wenn gilt:

$$\langle \vec{x}, \vec{y} \rangle = 0.$$

8.7 Beispiele (1) Ein Vektor $\vec{x} \in \mathbb{R}^n$ steht genau dann senkrecht auf dem i-ten Einheitsvektor $\vec{e}_i$, wenn $x_i = 0$ gilt. Insbesondere stehen die Einheitsvektoren $\vec{e}_1, \ldots, \vec{e}_n$ paarweise aufeinander senkrecht, das heißt es gilt

$$\langle \vec{e}_i, \vec{e}_j \rangle = 0 \quad \text{für } i \neq j.$$

Für $n = 2$ und $n = 3$ passt dies dazu, dass wir die Einheitsvektoren senkrecht aufeinander zeichnen.

(2) Zwei Vektoren $\vec{x}, \vec{y} \in \mathbb{R}^2$ sind genau dann orthogonal, wenn sie

$$x_1 y_1 + x_2 y_2 = 0$$

erfüllen. Insbesondere kann man zu gegebenem $\vec{x} = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ sehr leicht einen orthogonalen Vektor angeben, nämlich $\vec{y} = \left(\begin{smallmatrix} x_2 \\ -x_1 \end{smallmatrix}\right)$ oder $\vec{y} = \left(\begin{smallmatrix} -x_2 \\ x_1 \end{smallmatrix}\right)$.

(3) Sind $\vec{x}, \vec{y} \in \mathbb{R}^3$, dann ist

$$\vec{x} \times \vec{y} = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}$$

ein Vektor, der senkrecht auf $\vec{x}$ und auf $\vec{y}$ steht. Der Vektor $\vec{x} \times \vec{y}$ heißt das **Kreuzprodukt** von $\vec{x}$ und $\vec{y}$ (auch *Vektorprodukt*, weil das Ergebnis ein Vektor ist). Man kann das nachrechnen und auch sonst noch allerhand Eigenschaften beweisen. Wir werden das Kreuzprodukt aber nicht benutzen. ◇

8.6 Orthonormalbasen

Die Standardbasis $\vec{e}_1, \dots, \vec{e}_n$ von $\mathbb{R}^n$ hat die Eigenschaft, dass die Basisvektoren (und damit die Koordinatenachsen) senkrecht aufeinander stehen. Die Einträge eines Vektors $\vec{x}$ sind außerdem gerade die Skalarprodukte mit den Basisvektoren, das heißt, es gilt immer $\langle \vec{e}_i, \vec{x} \rangle = x_i$ und damit

$$\vec{x} = \langle \vec{e}_1, \vec{x} \rangle \cdot \vec{e}_1 + \cdots + \langle \vec{e}_n, \vec{x} \rangle \cdot \vec{e}_n.$$

Außer der Standardbasis gibt es noch andere Basen mit diesen Eigenschaften.

Definition Ein System von Vektoren $\vec{v}_1, \dots, \vec{v}_m$ in $\mathbb{R}^n$ heißt ein **Orthonormalsystem**, wenn gilt:

$$\langle \vec{v}_i, \vec{v}_i \rangle = 1 \text{ für alle } i = 1, \dots, m \qquad \text{und} \qquad \langle \vec{v}_i, \vec{v}_j \rangle = 0 \text{ für alle } i \neq j.$$

Die erste Eigenschaft sagt also, dass die Vektoren in einem Orthonormalsystem die Länge 1 haben (denn $\langle \vec{v}, \vec{v} \rangle = 1$ ist gleichbedeutend mit $\|\vec{v}\| = 1$). Die zweite Eigenschaft sagt, dass die Basisvektoren paarweise aufeinander senkrecht stehen. Als Abkürzung für die Fallunterscheidung in den Indizes verwendet man gern die Notation

$$\langle \vec{v}_i, \vec{v}_j \rangle = \delta_{ij} \quad \text{mit} \quad \delta_{ij} = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{falls } i \neq 0 \end{cases}$$

genannt das **Kronecker-Delta**.

8.8 Beispiel Für jedes $k \leqslant n$ bilden die ersten k Einheitsvektoren $\vec{e}_1, \dots, \vec{e}_k$ in $\mathbb{R}^n$ ein Orthonormalsystem. ◇

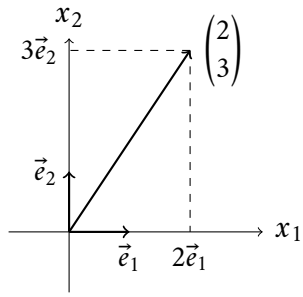
8.9 Satz *Es sei $\vec{v}_1, \dots \vec{v}_m$ ein Orthonormalsystem in $\mathbb{R}^n$.*

(1) $\vec{v}_1, \dots, \vec{v}_m$ sind linear unabhängig.

(2) Für jedes $\vec{u} \in \operatorname{Lin}(\vec{v}_1, \dots, \vec{v}_m)$ gilt

$$\vec{u} = \sum_{i=1}^{m} \langle \vec{v}_i, \vec{u} \rangle \cdot \vec{v}_i.$$

Das Orthonormalsystem ist also eine **Orthonormalbasis** des aufgespannten Unterraums. Nach (2) sind die Koordinaten eines Vektors in einer Orthonormalbasis die Skalarprodukte mit den Basisvektoren, genau wie in der Standardbasis.



Beweis. (1) Es sei $c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}$ eine lineare Relation, $c_1, \ldots, c_m \in \mathbb{R}$. Wir müssen zeigen, dass alle Koeffizienten $c_1, \ldots, c_m$ Null sind. Für jeden Index j berechnen wir

$$0 = \langle \vec{0}, \vec{v}_j \rangle = \langle c_1\vec{v}_1 + \cdots + c_m\vec{v}_m, \vec{v}_j \rangle = \langle c_j\vec{v}_j, \vec{v}_j \rangle + \sum_{i \neq j} \langle c_i\vec{v}_i, \vec{v}_j \rangle$$

$$= c_j \underbrace{\langle \vec{v}_j, \vec{v}_j \rangle}_{=1} + \sum_{i \neq j} c_i \underbrace{\langle \vec{v}_i, \vec{v}_j \rangle}_{=0} = c_j.$$

(2) Das rechnen wir einfach aus: Sei $\vec{u} \in U$. Da $\vec{v}_1, \ldots, \vec{v}_m$ eine Basis von U ist, besitzt $\vec{u}$ jedenfalls irgendeine Darstellung als Linearkombination von $\vec{v}_1, \ldots, \vec{v}_m$, etwa $\vec{u} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m$. Für jedes $i = 1, \ldots, m$ gilt dann

$$\langle \vec{v}_i, \vec{u} \rangle = \langle \vec{v}_i, \sum\nolimits_{j=1}^{m} c_j\vec{v}_j \rangle = c_i \underbrace{\langle \vec{v}_i, \vec{v}_i \rangle}_{=1} + \sum_{j \neq i} c_j \underbrace{\langle \vec{v}_i, \vec{v}_j \rangle}_{=0} = c_i. \qquad \blacksquare$$

Eine Basis, deren Basisvektoren nicht aufeinander senkrecht stehen, lässt sich schrittweise in eine Orthonormalbasis verwandeln, mit Hilfe des Gram-Schmidt-Verfahrens, das wir nun beschreiben.

8.10 Satz (Gram-Schmidt-Verfahren[3]) *Sei $\vec{v}_1, \ldots, \vec{v}_m$ ein linear unabhängiges System von Vektoren in $\mathbb{R}^n$. Dann gibt es ein Orthonormalsystem $\vec{u}_1, \ldots, \vec{u}_m$ mit*

$$\mathrm{Lin}(\vec{v}_1, \ldots, \vec{v}_k) = \mathrm{Lin}(\vec{u}_1, \ldots, \vec{u}_k)$$

für alle $k = 1, \ldots, m$. Insbesondere besitzt jeder lineare Unterraum von $\mathbb{R}^n$ eine Orthonormalbasis.

[3] nach Jørgen Pedersen Gram (1850–1916) und Erhard Schmidt (1876–1959)

Beweis. Da $\vec{v}_1, \dots, \vec{v}_m$ linear unabhängig sind, ist $\vec{v}_1 \neq \vec{0}$. Als erstes normieren wir $\vec{v}_1$, setzen also

$$\vec{u}_1 = \frac{1}{\|\vec{v}_1\|}\vec{v}_1.$$

Es ist dann $\mathrm{Lin}(\vec{u}_1) = \mathrm{Lin}(\vec{v}_1)$. Als nächstes ersetzen wir $\vec{v}_2$ durch

$$\vec{u}_2' = \vec{v}_2 - \langle \vec{u}_1, \vec{v}_2 \rangle \vec{u}_1.$$

Wegen $\vec{v}_2 \notin \mathrm{Lin}(\vec{v}_1) = \mathrm{Lin}(\vec{u}_1)$ gilt $\vec{u}_2' \neq \vec{0}$ und außerdem

$$\langle \vec{u}_1, \vec{u}_2' \rangle = \langle \vec{u}_1, \vec{v}_2 \rangle - \langle \vec{u}_1, \vec{v}_2 \rangle \langle \vec{u}_1, \vec{u}_1 \rangle = 0.$$

Jetzt normieren wir $\vec{u}_2'$, setzen also

$$\vec{u}_2 = \frac{1}{\|\vec{u}_2'\|}\vec{u}_2'.$$

In dieser Weise fahren wir fort: Haben wir für ein k mit $1 \leqslant k < m$ ein Orthonormalsystem $\vec{u}_1, \dots, \vec{u}_k$ mit $\mathrm{Lin}(\vec{u}_1, \dots, \vec{u}_k) = \mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_k)$ konstruiert, dann setzen wir

$$\vec{u}_{k+1}' = \vec{v}_{k+1} - \sum\nolimits_{i=1}^{k} \langle \vec{u}_i, \vec{v}_{k+1} \rangle \vec{u}_i.$$

Es gilt dann $\vec{u}_{k+1}' \in \mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_{k+1})$, und $\vec{u}_{k+1}' \neq \vec{0}$ wegen $\vec{v}_{k+1} \notin \mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_k) = \mathrm{Lin}(\vec{u}_1, \dots, \vec{u}_k)$. Wir können also

$$\vec{u}_{k+1} = \frac{1}{\|\vec{u}_{k+1}'\|}\vec{u}_{k+1}'$$

definieren und haben $\mathrm{Lin}(\vec{u}_1, \dots, \vec{u}_{k+1}) = \mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_{k+1})$. Für $j = 1, \dots, k$ gilt

$$\langle \vec{u}_j, \vec{u}_{k+1}', \rangle = \langle \vec{u}_j, \vec{v}_{k+1} \rangle - \sum\nolimits_{i=1}^{k} \langle \vec{u}_i, \vec{v}_{k+1} \rangle \langle \vec{u}_j, \vec{u}_i \rangle = \langle \vec{u}_j, \vec{v}_{k+1} \rangle - \langle \vec{u}_j, \vec{v}_{k+1} \rangle = 0$$

und damit auch $\langle \vec{u}_j, \vec{u}_{k+1} \rangle = 0$. Nach m Schritten ist $\vec{u}_1, \dots, \vec{u}_m$ ein Orthonormalsystem mit $\mathrm{Lin}(\vec{u}_1, \dots, \vec{u}_m) = \mathrm{Lin}(\vec{v}_1, \dots, \vec{v}_m)$.

Ist insbesondere U ein linearer Unterraum von $\mathbb{R}^n$, dann können wir eine Basis von U wählen und anschließend das gerade beschriebene Verfahren anwenden. Wir bekommen dann eine Orthonormalbasis von U. ∎

Im Prinzip funktioniert das Gram-Schmidt-Verfahren immer. Beim Rechnen per Hand entstehen allerdings oft unangenehme Wurzelausdrücke.[4] Wir rechnen dazu nur ein ganz kleines Beispiel, welches das Problem hinreichend illustriert.

8.11 Beispiel Gegeben seien die Vektoren

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \quad \text{und} \quad \vec{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} .$$

Wir berechnen mit Hilfe des Gram-Schmidt-Verfahrens eine Orthonormalbasis $\vec{u}_1, \vec{u}_2, \vec{u}_3$ von $\mathbb{R}^3$ mit $\mathrm{Lin}(\vec{v}_1, \vec{v}_2) = \mathrm{Lin}(\vec{u}_1, \vec{u}_2)$. Als erstes ergänzen wir $\vec{v}_1, \vec{v}_2$ zu einer Basis, etwa indem wir $\vec{v}_3 = \vec{e}_3$ hinzufügen. Als nächstes setzen wir

$$\vec{u}_1 = \frac{1}{\|\vec{v}_1\|}\vec{v}_1 = \frac{1}{\sqrt{14}}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

und dann

$$\vec{u}_2' = \vec{v}_2 - \langle \vec{u}_1, \vec{v}_2 \rangle \vec{u}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \frac{1}{7}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{1}{7}\begin{pmatrix} 8 \\ 2 \\ -4 \end{pmatrix} .$$

Es gilt $\|\vec{u}_2'\| = \sqrt{\frac{84}{49}} = \sqrt{\frac{12}{7}}$ und wir setzen

$$\vec{u}_2 = \frac{1}{\|\vec{u}_2'\|}\vec{u}_2' = \frac{1}{\sqrt{21}}\begin{pmatrix} 4 \\ 1 \\ -2 \end{pmatrix} .$$

Dann auch noch

$$\vec{u}_3' = \vec{v}_3 - \langle \vec{u}_1, \vec{v}_3 \rangle \vec{u}_1 - \langle \vec{u}_2, \vec{v}_3 \rangle \vec{u}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \frac{3}{14}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \frac{2}{21}\begin{pmatrix} 4 \\ 1 \\ -2 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

und zu guter Letzt $\|\vec{u}_3'\| = \frac{1}{\sqrt{6}}$ und damit

$$\vec{u}_3 = \frac{1}{\|\vec{u}_3'\|}\vec{u}_3' = \sqrt{6} \cdot \vec{u}_3' = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

Damit ist die gesuchte Orthonormalbasis gefunden. ◇

[4] Auch beim numerischen Rechnen macht das Gram-Schmidt-Verfahren oft Probleme, wenn die Orthogonalität durch Rundungsfehler gestört wird. Deshalb wird in der Praxis meistens ein leicht modifiziertes Verfahren verwendet.

9 Lineare Abbildungen

Mit Abbildungen haben wir uns allgemein schon in Kap. 5 beschäftigt. Für die lineare Algebra mit Abstand am wichtigsten sind Abbildungen $\mathbb{R}^n \to \mathbb{R}^m$, die also Vektoren auf Vektoren abbilden, und außerdem *linear* sind.

9.1 Linearität

Jede $m \times n$-Matrix können wir mit einem Vektor aus $\mathbb{R}^n$ multiplizieren und erhalten einen Vektor aus $\mathbb{R}^m$. Mit anderen Worten, jedes $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ definiert die Abbildung

$$\varphi_A \colon \mathbb{R}^n \to \mathbb{R}^m,\ \vec{x} \mapsto A\vec{x}.$$

Im Prinzip haben wir diese Abbildung schon betrachtet: Ist $A\vec{x} = \vec{b}$ mit $\vec{b} \in \mathbb{R}^m$ ein lineares Gleichungssystem mit Koeffizientenmatrix A, dann ist eine Lösung des Systems ein Vektor in $\mathbb{R}^n$, der von φ_A auf $\vec{b}$ abgebildet wird. Mit anderen Worten, die Lösungen sind die *Urbilder* von $\vec{b}$ unter der Abbildung φ_A.

Nach den Rechenregeln für das Matrix-Vektor-Produkt gilt

$$\varphi_A(\vec{x} + \vec{y}) = A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} = \varphi_A(\vec{x}) + \varphi_A(\vec{y})$$

und außerdem

$$\varphi_A(c \cdot \vec{x}) = A(c \cdot \vec{x}) = c \cdot A\vec{x} = c \cdot \varphi_A(\vec{x})$$

für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ und alle $c \in \mathbb{R}$. Das führt zu folgender Definition.

Definition Eine Abbildung $\varphi \colon \mathbb{R}^n \to \mathbb{R}^m$ heißt **linear**, wenn sie die folgenden beiden Eigenschaften besitzt:

(1) Für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ gilt $\varphi(\vec{x} + \vec{y}) = \varphi(\vec{x}) + \varphi(\vec{y})$. (*Additivität*)

(2) Für alle $\vec{x} \in \mathbb{R}^n$ und alle $c \in \mathbb{R}$ gilt $\varphi(c \cdot \vec{x}) = c \cdot \varphi(\vec{x})$. (*Homogenität*)

Ist $\varphi \colon \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung und $c_1\vec{v}_1 + \cdots + c_k\vec{v}_k$ eine Linearkombination von Vektoren $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$, dann können wir die Linearität von φ mehrfach anwenden und bekommen

$$\varphi(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k) = c_1\varphi(\vec{v}_1) + \cdots + c_k\varphi(\vec{v}_k)$$

Man kann die lineare Abbildung φ also aus der Linearkombination *herausziehen* und in sie *hineinziehen*.

9.1 Lemma *Jede lineare Abbildung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ erfüllt $\varphi(\vec{0}) = \vec{0}$.*

Beweis. Denn aus der Linearität von φ folgt $\varphi(\vec{0}) = \varphi(0 \cdot \vec{0}) = 0 \cdot \varphi(\vec{0}) = \vec{0}$. ∎

9.2 Lemma *(1) Ist $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ mit Spaltenvektoren $\vec{b}_1, \dots, \vec{b}_n \in \mathbb{R}^m$, dann gilt*

$$\varphi_A(\vec{e}_i) = \vec{b}_i \quad \text{für } i = 1, \dots, n.$$

Die Spalten von A sind also die Bilder der Einheitsvektoren. Für jedes $\vec{x} \in \mathbb{R}^n$ gilt dann

$$\varphi_A(\vec{x}) = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n.$$

(2) Jede lineare Abbildung kommt von einer Matrix, das heißt, für jede lineare Abbildung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ existiert eine Matrix $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ mit $\varphi = \varphi_A$.

Beweis. (1) Aus der Definition des Matrix-Vektor-Produkts sehen wir direkt

$$A \cdot \vec{e}_1 = \begin{pmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_n \\ | & & | \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} | \\ \vec{b}_1 \\ | \end{pmatrix} \quad \dots \quad A \cdot \vec{e}_n = \begin{pmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_n \\ | & & | \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} | \\ \vec{b}_n \\ | \end{pmatrix}$$

und das ist die Behauptung. Für jedes $\vec{x} \in \mathbb{R}^n$ können wir $\vec{x} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$ schreiben und bekommen mit der Linearität von φ_A die behauptete Gleichheit $\varphi_A(\vec{x}) = \varphi_A(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n) = x_1\varphi_A(\vec{e}_1) + \cdots + x_n\varphi_A(\vec{e}_n) = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n$.

(2) Es sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ linear, dann setzen wir $\vec{b}_i = \varphi(\vec{e}_i) \in \mathbb{R}^m$ für $i = 1, \dots, n$ und bilden die $m \times n$-Matrix A mit Spalten $\vec{b}_1, \dots, \vec{b}_n$. Nach (1) erfüllt die lineare Abbildung φ_A dann für jedes $\vec{x} \in \mathbb{R}^n$ die Gleichheit

$$\varphi_A(\vec{x}) = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n = x_1\varphi(\vec{e}_1) + \cdots + x_n\varphi(\vec{e}_n) = \varphi(\vec{x}).$$

Also stimmen φ und φ_A überein. ∎

9.3 Korollar (Lineare Ausdehnung) *Zu jeder Wahl von n Vektoren $\vec{v}_1, \dots, \vec{v}_n \in \mathbb{R}^m$ gibt es genau eine lineare Abbildung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ mit*

$$\varphi(\vec{e}_1) = \vec{v}_1, \ \dots, \ \varphi(\vec{e}_n) = \vec{v}_n.$$

Beweis. Denn wir können die Matrix A mit Spalten $\vec{v}_1, \dots, \vec{v}_n$ bilden, dann hat φ_A die gewünschte Eigenschaft. Da jede lineare Abbildung von einer Matrix kommt, kann es auch keine andere lineare Abbildung mit dieser Eigenschaft geben. ∎

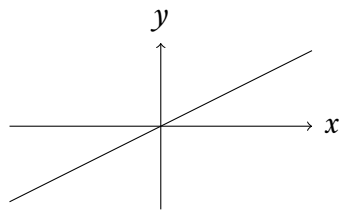
9.2 Beispiele

Damit haben wir abstrakt völlig verstanden, welche linearen Abbildungen es gibt. Wir betrachten diese Abbildungen nun etwas näher und wollen vor allem auch die Geometrie dahinter verstehen.

9.4 Beispiel Die linearen Abbildungen $\mathbb{R} \to \mathbb{R}$ entsprechen den 1×1-Matrizen und sind damit von der Form

$$\varphi_a \colon x \mapsto a \cdot x$$

für ein $a \in \mathbb{R}$. Das sind genau die Funktionen $\varphi \colon \mathbb{R} \to \mathbb{R}$, deren Graph $\Gamma_\varphi = \{(x, y) \in \mathbb{R}^2 \mid y = \varphi(x)\}$ eine Gerade durch den Ursprung ist.



Graph der Funktion $\varphi \colon \mathbb{R} \to \mathbb{R},\ x \mapsto \frac{1}{2}x$

Das ist nicht gerade eine Fülle an interessanten Funktionen. Dasselbe geht immerhin in jeder Dimension: Für jedes $a \in \mathbb{R}$ ist die Abbildung

$$\varphi \colon \mathbb{R}^n \to \mathbb{R}^n,\ \vec{x} \mapsto a \cdot \vec{x}$$

linear, die **Streckung** um den Faktor a. Die zugehörige Matrix hat die Gestalt

$$\begin{pmatrix} a & 0 & \cdots & 0 \\ 0 & a & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & a \end{pmatrix}$$

◇

Definition Die $n \times n$-Matrix

$$\mathbb{1}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{pmatrix}$$

heißt die **Einheitsmatrix** der Größe n.

Zur Einheitsmatrix gehört die Streckung um den Faktor 1. Also ist $\varphi_{\mathbb{1}_n}$ die **Identität** $\mathrm{id}_{\mathbb{R}^n} \colon \vec{x} \mapsto \vec{x}$, die jeden Vektor auf sich selbst abbildet.

9.5 Beispiele In Dimension 2 gibt es schon viel mehr Möglichkeiten als nur Streckungen. Wir betrachten lineare Abbildungen $\mathbb{R}^2 \to \mathbb{R}^2$ der reellen Ebene in sich. Solche Abbildungen entsprechen nach Lemma 9.2 den 2×2-Matrizen. Ist

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} | & | \\ \vec{b}_1 & \vec{b}_2 \\ | & | \end{pmatrix}$$

dann ist φ_A explizit die Abbildung

$$\varphi_A \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = A\vec{x} = x_1\vec{b}_1 + x_2\vec{b}_2 = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}.$$

Sehen wir uns diese Abbildung für verschiedene Wahlen der beiden Vektoren $\vec{b}_1$ und $\vec{b}_2$ genauer an.

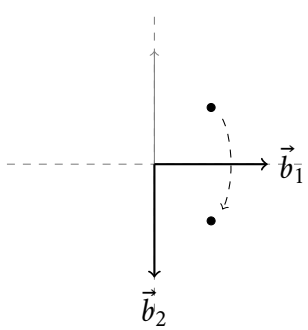
(1) Für $\vec{b}_1 = \vec{e}_1$ und $\vec{b}_2 = -\vec{e}_2$ bekommen wir die Abbildung $\varphi_A \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} x_1 \\ -x_2 \end{smallmatrix}\right)$. Das ist die **Spiegelung** an der waagerechten Achse $\mathrm{Lin}(\vec{e}_1)$.

(2) Für $\vec{b}_1 = \vec{e}_2$ und $\vec{b}_2 = \vec{e}_1$ erhalten wir die Abbildung $\varphi_A \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} x_2 \\ x_1 \end{smallmatrix}\right)$, die die beiden Einträge vertauscht. Auch das ist eine Spiegelung, nämlich an der Diagonalen $\mathrm{Lin}(\vec{e}_1 + \vec{e}_2)$.

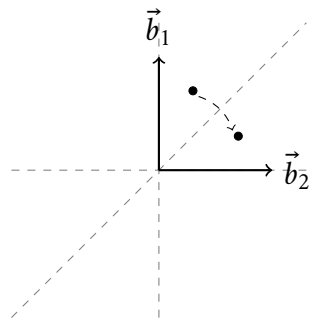
(3) Für $\vec{b}_1 = \vec{e}_2$ und $\vec{b}_2 = -\vec{e}_1$ bekommen wir $\varphi_A \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} -x_2 \\ x_1 \end{smallmatrix}\right)$. Das ist eine Drehung der Ebene um 90 Grad gegen den Uhrzeigersinn.

(4) Für einen Parameter $c \in \mathbb{R}$, $c \neq 0$, setzen wir $\vec{b}_1 = \vec{e}_1$ und $\vec{b}_2 = \vec{e}_2 + c\vec{e}_1$. Die zugehörige lineare Abbildung ist $\varphi_A \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = x_1\vec{e}_1 + x_2(\vec{e}_2 + c\vec{e}_1) = \left(\begin{smallmatrix} x_1 + cx_2 \\ x_2 \end{smallmatrix}\right)$ und heißt eine **Scherung** zum Parameter c.

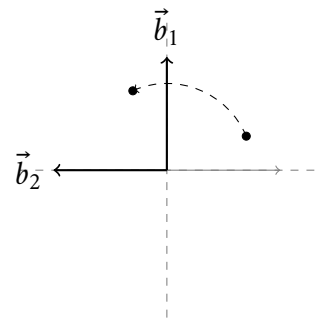
(5) Für $\vec{b}_1 = \vec{e}_1$ und $\vec{b}_2 = \vec{0}$ bekommen wir $\varphi_A \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} x_1 \\ 0 \end{smallmatrix}\right)$. Das ist die Projektion der Ebene auf die waagerechte Achse $\mathrm{Lin}(\vec{e}_1)$. ◇



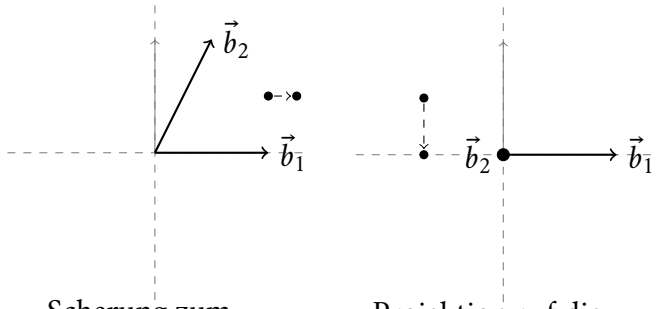
Spiegelung an der waagerechten Achse



Spiegelung an der Diagonalen



Drehung um 90 Grad



Scherung zum Parameter $c = \frac{1}{2}$ | Projektion auf die waagerechte Achse

Diese Bilder zeigen nicht die Graphen der Abbildungen $\mathbb{R}^2 \to \mathbb{R}^2$. (Die sind ja Teilmengen von $\mathbb{R}^4$!) Stattdessen können wir lediglich einige Vektoren zusammen mit ihren Bildvektoren in derselben Zeichnung darstellen.

9.3 Orthogonale Abbildungen

Unter den Beispielen für lineare Abbildungen haben wir Drehungen und Spiegelungen gesehen. Das wollen wir noch etwas allgemeiner untersuchen.

Definition (1) Eine Abbildung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ heißt eine **Isometrie**, wenn sie den Abstand erhält, also

$$\|\varphi(\vec{x}) - \varphi(\vec{y})\| = \|\vec{x} - \vec{y}\|$$

für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ erfüllt.

(2) Eine **orthogonale Abbildung** ist eine Isometrie $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$, die zusätzlich $\varphi(\vec{0}) = \vec{0}$ erfüllt, also den Ursprung fixiert.

9.6 Beispiele (1) Für jedes $\vec{v} \in \mathbb{R}^n$ ist die **Translation** (Verschiebung)

$$\tau_{\vec{v}}\colon \mathbb{R}^n \to \mathbb{R}^n,\ \vec{x} \mapsto \vec{x} + \vec{v}$$

eine Isometrie, denn es gilt $\|\tau_{\vec{v}}(\vec{x}) - \tau_{\vec{v}}(\vec{y})\| = \|\vec{x} + \vec{v} - (\vec{y} + \vec{v})\| = \|\vec{x} - \vec{y}\|$. Für $\vec{v} \neq \vec{0}$ ist $\tau_{\vec{v}}$ aber keine orthogonale Abbildung, denn es gilt $\tau_{\vec{v}}(\vec{0}) = \vec{v}$.

(2) Für $a \in \mathbb{R}$ ist die Streckung $\mathbb{R}^n \to \mathbb{R}^n,\ \vec{x} \mapsto a\vec{x}$ um den Faktor a nur für $a = \pm 1$ eine Isometrie, und dann auch eine orthogonale Abbildung.

(3) Die beiden Spiegelungen in Beispiel 9.5(1),(2) und die Drehung 9.5(3) sind orthogonal. Denn sie bilden $\vec{0}$ auf $\vec{0}$ ab (wie jede lineare Abbildung) und ändern sonst nur Vorzeichen, die sich bei der Berechnung der Abstände wieder herausheben (Übung). ◇

9.7 Lemma

(1) Jede Isometrie ist injektiv.[1]

*(2) Jede orthogonale Abbildung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ ist **längentreu**, das heißt, sie erfüllt*

$$\|\varphi(\vec{x})\| = \|\vec{x}\|$$

für alle $\vec{x} \in \mathbb{R}^n$.

Beweis. (1) Zwei Punkte in $\mathbb{R}^n$ sind genau dann gleich, wenn ihr Abstand 0 ist. Sind also $\vec{x}, \vec{y} \in \mathbb{R}^n$ mit $\vec{x} \neq \vec{y}$, dann gilt $\|\varphi(\vec{x}) - \varphi(\vec{y})\| = \|\vec{x} - \vec{y}\| \neq 0$, und damit auch $\varphi(\vec{x}) \neq \varphi(\vec{y})$. Das zeigt die Injektivität von φ.

(2) Die Länge eines Vektors ist nichts anderes, als der Abstand zum Nullpunkt. Also folgt $\|\varphi(\vec{x})\| = \|\varphi(\vec{x}) - \vec{0}\| = \|\varphi(\vec{x}) - \varphi(\vec{0})\| = \|\vec{x} - \vec{0}\| = \|\vec{x}\|$. ∎

Wir beweisen nun einige grundlegende Eigenschaften orthogonaler Abbildungen, unter anderem, dass jede solche Abbildung linear ist.

9.8 Lemma *Ist $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ eine orthogonale Abbildung, dann gilt*

$$\langle \varphi(\vec{x}), \varphi(\vec{y}) \rangle = \langle \vec{x}, \vec{y} \rangle$$

für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$.

Da wir mit dem Skalarprodukt die Winkel zwischen den Vektoren messen, sagt das Lemma gerade, dass orthogonale Abbildungen nicht nur längentreu, sondern auch **winkeltreu** sind.

Beweis. Für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ gilt nach der Polarisationsformel 8.4 die Gleichheit

$$\langle \vec{x}, \vec{y} \rangle = -\frac{1}{2}\big(\|\vec{x} - \vec{y}\|^2 - \|\vec{x}\|^2 - \|\vec{y}\|^2\big).$$

Da φ orthogonal ist, folgt damit

$$\begin{aligned}\langle \varphi(\vec{x}), \varphi(\vec{y}) \rangle &= -\frac{1}{2}\Big(\|\varphi(\vec{x}) - \varphi(\vec{y})\|^2 - \|\varphi(\vec{x})\|^2 - \|\varphi(\vec{y})\|^2\Big) \\ &= -\frac{1}{2}\Big(\|\vec{x} - \vec{y}\|^2 - \|\vec{x}\|^2 - \|\vec{y}\|^2\Big) = \langle \vec{x}, \vec{y} \rangle.\end{aligned}$$ ∎

[1] Wir werden später sehen, dass Isometrien auch surjektiv, also bijektiv sind.

9.9 Satz *Eine Abbildung $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ ist genau dann orthogonal, wenn sie die folgenden beiden Eigenschaften besitzt:*

(1) Das System $\varphi(\vec{e}_1), \dots, \varphi(\vec{e}_n)$ ist eine Orthonormalbasis von $\mathbb{R}^n$.

(2) Die Abbildung φ ist linear.

Beweis. Es sei φ orthogonal, dann müssen wir die Eigenschaften nachprüfen:

(1) Setze $\vec{v}_i = \varphi(\vec{e}_i)$ für $i = 1, \dots, n$. Nach dem vorangehenden Lemma gilt $\langle \vec{v}_i, \vec{v}_j \rangle = \langle \vec{e}_i, \vec{e}_j \rangle = \delta_{i,j}$ (Kronecker-Delta). Das heißt, $\vec{v}_1, \dots, \vec{v}_n$ ist ein Orthonormalsystem in $\mathbb{R}^n$ und damit linear unabhängig nach Satz 8.9(1). Also ist $U = \operatorname{Lin}(\vec{v}_1, \dots, \vec{v}_n)$ ein n-dimensionaler linearer Unterraum von $\mathbb{R}^n$ und damit $U = \mathbb{R}^n$ (Satz 7.16(1)). Also ist $\vec{v}_1, \dots, \vec{v}_n$ eine Orthonormalbasis von $\mathbb{R}^n$.

(2) Es sei $\vec{v}_1, \dots, \vec{v}_n$ die Orthonormalbasis wie oben. Für jedes $\vec{x} \in \mathbb{R}^n$ gilt

$$\varphi(\vec{x}) = \sum\nolimits_{i=1}^{n} \langle \vec{v}_i, \varphi(\vec{x}) \rangle \vec{v}_i = \sum\nolimits_{i=1}^{n} \langle \vec{e}_i, \vec{x} \rangle \vec{v}_i = \sum\nolimits_{i=1}^{n} x_i \vec{v}_i$$

nach Satz 8.9(2). Also ist φ dasselbe wie die lineare Abbildung φ_A für die Matrix A mit Spalten $\vec{v}_1, \dots, \vec{v}_n$.

Seien umgekehrt (1) und (2) erfüllt. Nach Lemma 9.1 gilt dann $\varphi(\vec{0}) = \vec{0}$. Wir müssen zeigen, dass φ auch eine Isometrie ist. Sind $\vec{x}, \vec{y} \in \mathbb{R}^n$ beliebig, dann gilt $\|\varphi(\vec{x}) - \varphi(\vec{y})\| = \|\varphi(\vec{x} - \vec{y})\|$, da φ linear ist. Es genügt also zu beweisen, dass φ längentreu ist, denn dann folgt $\|\varphi(\vec{x} - \vec{y})\| = \|\vec{x} - \vec{y}\|$.

Um $\|\varphi(\vec{x})\| = \|\vec{x}\|$ zu zeigen, schreiben wir wieder $\vec{v}_i = \varphi(\vec{e}_i)$ und berechnen

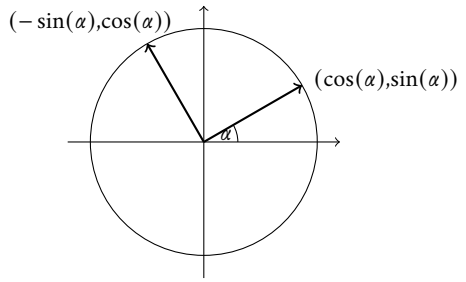
$$\begin{aligned} \|\varphi(\vec{x})\|^2 &= \|\varphi(x_1\vec{e}_1 + \cdots + x_n\vec{e}_n)\|^2 = \|x_1\varphi(\vec{e}_1) + \cdots + x_n\varphi(\vec{e}_n)\|^2 \\ &= \|x_1\vec{v}_1 + \cdots + x_n\vec{v}_n\|^2 = \left\langle \sum\nolimits_{i=1}^{n} x_i\vec{v}_i, \sum\nolimits_{j=1}^{n} x_j\vec{v}_j \right\rangle \\ &= \sum_{i,j=1}^{n} x_i x_j \langle \vec{v}_i, \vec{v}_j \rangle = \sum_{i=1}^{n} x_i^2 = \|x\|^2. \end{aligned}$$

Wurzelziehen auf beiden Seiten ergibt $\|\varphi(\vec{x})\| = \|\vec{x}\|$, und wir sind fertig. ∎

9.10 Beispiel Rotieren wir die beiden Einheitsvektoren $\vec{e}_1$ und $\vec{e}_2$ in $\mathbb{R}^2$ um den Winkel $\alpha \in [0, 2\pi)$, dann erhalten wir die Vektoren

$$\vec{b}_1 = \begin{pmatrix} \cos(\alpha) \\ \sin(\alpha) \end{pmatrix} \quad \text{und} \quad \vec{b}_2 = \begin{pmatrix} -\sin(\alpha) \\ \cos(\alpha) \end{pmatrix}$$

auf dem Einheitskreis.



Es gibt genau eine lineare Abbildung $\rho_\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$ mit $\rho_\alpha(\vec{e}_1) = \vec{b}_1$ und $\rho_\alpha(\vec{e}_e) = \vec{b}_2$, nämlich gegeben durch die Matrix

$$R_\alpha = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}, \qquad \rho_\alpha \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \cos(\alpha) - x_2 \sin(\alpha) \\ x_1 \sin(\alpha) + x_2 \cos(\alpha) \end{pmatrix}.$$

Nach Satz 9.9 ist diese Abbildung auch orthogonal, da $\vec{b}_1, \vec{b}_2$ eine Orthonormalbasis von $\mathbb{R}^2$ bilden. ◇

Zum Abschluss noch ein paar Bemerkungen zur geometrischen Vorstellung: Wenn wir uns auf einen Drehstuhl setzen, dann dreht sich der Stuhl (und wir mit ihm), aber der umgebende Raum dreht sich nicht. Drehungen *als Abbildungen* in der Mathematik funktionieren nicht so: Die Abbildung ρ_α im vorigen Beispiel ist auf der ganzen Ebene $\mathbb{R}^2$ definiert und bewegt alles auf einmal. Sie wird nicht selektiv auf ein bestimmtes Objekt angewendet.

Außerdem sind Abbildungen in der Mathematik keine *Aktionen*, es zählt nur das Ergebnis. Eine Drehung um den Winkel 2π (also 360 Grad) ist mathematisch genau dasselbe wie die Identität, die alle Punkte festlässt. Für die Drehmatrix oben gilt $R_{2k\pi} = \mathbb{1}_n$ (Einheitsmatrix) für alle ganzen Zahlen k. Wenn wir uns dagegen auf dem Drehstuhl fünfzigmal schnell um die eigene Achse drehen, haben wir anschließend wahrscheinlich nicht das Gefühl, dass *überhaupt nichts* passiert ist.

9.4 Komposition und Matrizenprodukt

Ganz allgemein können wir mehrere Abbildungen hintereinander ausführen, was wir in §5.3 als *Komposition* bezeichnet haben. Ist $\varphi_B\colon \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung, gegeben durch eine Matrix $B \in \mathrm{Mat}_{m\times n}(\mathbb{R})$, und ist $\varphi_A\colon \mathbb{R}^m \to \mathbb{R}^l$ eine weitere lineare Abbildung, gegeben durch $A \in \mathrm{Mat}_{l\times m}(\mathbb{R})$, dann ist ihre Komposition die Abbildung

$$\varphi_A \circ \varphi_B\colon \mathbb{R}^n \to \mathbb{R}^l, \vec{x} \mapsto A \cdot B\vec{x}.$$

Dabei ist das Matrix-Vektor-Produkt $B\vec{x}$ ein Vektor in $\mathbb{R}^m$, den wir mit der Matrix A multiplizieren können. Wir können aber auch zuerst die beiden Matrizen A und B multiplizieren. Wenn man das ausrechnet, kommt man auf Folgendes:

Definition Gegeben seien Matrizen

$$A \in \mathrm{Mat}_{l\times m}(\mathbb{R}) \quad \text{und} \quad B \in \mathrm{Mat}_{m\times n}(\mathbb{R}).$$

Dann heißt die $l \times n$-Matrix mit den Einträgen

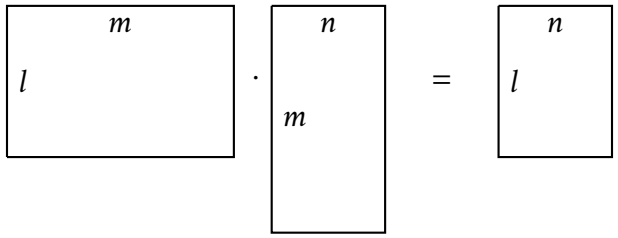
$$\sum_{k=1}^{m} a_{ik}b_{kj}, \quad (i = 1,\dots,l,\ j = 1,\dots,n)$$

das **Matrizenprodukt** (oder kurz **Produkt**) von A und B, geschrieben

$$A \cdot B \quad \text{oder} \quad AB.$$

Die Zuordnung $(A, B) \mapsto AB$ wird als **Matrizenmultiplikation** bezeichnet.

Damit man zwei Matrizen überhaupt multiplizieren kann, muss also die erste Matrix so viele *Spalten* haben wie die zweite *Zeilen* hat.



9.11 Beispiele (1) Das Produkt von 2×2-Matrizen ist gegeben durch:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{pmatrix}.$$

(2) Das Produkt aus einer $m \times n$-Matrix A und einer $n \times 1$-Matrix $\vec{x}$ (Spaltenvektor) ist dasselbe wie das Matrix-Vektor-Produkt $A\vec{x}$.

(3) Das Produkt aus einem Zeilenvektor und einem Spaltenvektor (derselben

Länge) ist eine 1×1-Matrix, also ein Skalar:

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Das ist dasselbe wie das Skalarprodukt der beiden Vektoren, wenn man $\vec{x}$ ebenfalls als Spaltenvektor schreibt.

(4) Das Produkt aus einem Spaltenvektor der Länge m und einem Zeilenvektor der Länge n ist dagegen eine $m \times n$-Matrix:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix} = \begin{pmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_m y_1 & x_m y_2 & \cdots & x_m y_n \end{pmatrix} \qquad \diamond$$

Die Matrizenmultiplikation genügt den folgenden **Rechenregeln**:
Für Matrizen A, B, C gelten

(1) $(AB)C = A(BC)$ (*Assoziativität*)

(2) $A\mathbb{1} = \mathbb{1}A = A$ (*Einheitsmatrix*)

(3) $(A + B)C = AC + BC$ und $A(B + C) = AB + AC$ (*Distributivität*)

Dabei müssen die Matrizen A, B, C und $\mathbb{1}$ in jeder dieser Aussagen zueinander passende Formate haben (welche genau?), damit die Summen und Produkte alle definiert sind.

Es ist nicht schwierig aber etwas mühsam, diese Regeln direkt nachzurechnen. Wir verzichten an dieser Stelle darauf und erledigen dies später durch ein abstraktes Argument.

Wir wollen aber noch überprüfen, dass die Matrizenmultiplikation tatsächlich der Komposition der linearen Abbildungen entspricht.

9.12 Satz *Es seien $A \in \mathrm{Mat}_{l\times m}(\mathbb{R})$ und $B \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ und seien $\varphi_A \colon \mathbb{R}^m \to \mathbb{R}^l$ und $\varphi_B \colon \mathbb{R}^n \to \mathbb{R}^m$ die zugehörigen linearen Abbildungen. Dann ist die Komposition $\varphi_A \circ \varphi_B \colon \mathbb{R}^n \to \mathbb{R}^l$ ebenfalls linear und es gilt*

$$\varphi_A \circ \varphi_B = \varphi_{AB}.$$

Beweis. Dazu rechnen wir $\varphi_A\big(\varphi_B(\vec{x})\big)$ für $\vec{x} \in \mathbb{R}^n$ aus: Der Vektor $\varphi_B(\vec{x}) = B\vec{x} \in \mathbb{R}^m$ hat den k-ten Eintrag $\sum_{j=1}^{n} b_{kj}x_j$. Deshalb hat $A \cdot B\vec{x}$ den i-ten Eintrag

$$\sum_{k=1}^{m} a_{ik}\left(\sum_{j=1}^{n} b_{kj}x_j\right) = \sum_{j=1}^{n}\left(\sum_{k=1}^{m} a_{ik}b_{kj}\right)x_j.$$

Das ist dasselbe wie der i-te Eintrag von $(AB)\vec{x}$. Die Behauptung ist bewiesen. ■

9.13 Beispiel Betrachten wir in Beispiel 9.5 die Spiegelung an der waagerechten Achse

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad , \quad \varphi_A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}$$

und die Spiegelung an der Diagonalen

$$B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \varphi_B \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}.$$

Die Komposition der beiden Abbildungen ist

$$(\varphi_A \circ \varphi_B)(\vec{x}) = \varphi_A\big(\varphi_B(\vec{x})\big) = \varphi_A \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix}$$

beziehungsweise

$$(\varphi_B \circ \varphi_A)(\vec{x}) = \varphi_B\big(\varphi_A(\vec{x})\big) = \varphi_B \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix} = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} \quad .$$

Dazu gehören die beiden Matrizenprodukte

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{und} \quad BA = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} .$$

Wir sehen auch, dass $AB \neq BA$ gilt, weil auch $\varphi_A \circ \varphi_B$ und $\varphi_B \circ \varphi_A$ nicht dieselben Abbildungen sind. (Welche Abbildungen sind das geometrisch betrachtet?) ◇

9.5 Dimensionsformel für lineare Abbildungen

In diesem Abschnitt geht es darum, wie lineare Abbildugen mit linearen Unterräumen zusammenspielen. Ziel ist die wichtige Dimensionsformel für lineare Abbildungen.

9.14 Lemma *Seien $U \subset \mathbb{R}^n$ und $V \subset \mathbb{R}^m$ lineare Unterräume, sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung. Dann sind die Bildmenge $\varphi(U) \subset \mathbb{R}^m$ und die Urbildmenge $\varphi^{-1}(V) \subset \mathbb{R}^n$ ebenfalls lineare Unterräume.*

Beweis. Für das Bild: Es gilt $\vec{0} = \varphi(\vec{0}) \in \varphi(U)$, also $\varphi(U) \neq \emptyset$. Sind $\vec{w}_1, \vec{w}_2 \in \varphi(U)$ und $c_1, c_2 \in \mathbb{R}$, dann gibt es $\vec{u}_1, \vec{u}_2 \in U$ mit $\varphi(\vec{u}_1) = \vec{w}_1$ und $\varphi(\vec{u}_2) = \vec{w}_2$. Es folgt

$$c_1\vec{w}_1 + c_2\vec{w}_2 = c_1\varphi(\vec{u}_1) + c_2\varphi(\vec{u}_2) = \varphi(c_1\vec{u}_1 + c_2\vec{u}_2) \in \varphi(U).$$

Damit ist gezeigt, dass $\varphi(U)$ ein linearer Unterraum ist.

Für das Urbild: Wegen $\varphi(\vec{0}) = \vec{0} \in V$ gilt $\varphi^{-1}(V) \neq \emptyset$. Sind $\vec{v}_1, \vec{v}_2 \in \varphi^{-1}(V)$ und $c_1, c_2 \in \mathbb{R}$, dann gilt also $\varphi(\vec{v}_1), \varphi(\vec{v}_2) \in V$ und damit auch

$$\varphi(c_1\vec{v}_1 + c_2\vec{v}_2) = c_1\varphi(\vec{v}_1) + c_2\varphi(\vec{v}_2) \in V.$$

Es folgt $c_1\vec{v}_1 + c_2\vec{v}_2 \in \varphi^{-1}(V)$, so dass $\varphi^{-1}(V)$ ein linearer Unterraum ist. ∎

Definition Es sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung.

(1) Der **Kern** von φ ist der lineare Unterraum

$$\mathrm{Kern}(\varphi) = \{\vec{v} \in \mathbb{R}^n \mid \varphi(\vec{v}) = \vec{0}\}.$$

(2) Das **Bild** von φ ist der lineare Unterraum

$$\mathrm{Bild}(\varphi) = \varphi(\mathbb{R}^n).$$

Der Kern ist die Urbildmenge $\mathrm{Kern}(\varphi) = \varphi^{-1}(\{\vec{0}\})$ (und damit nach Lemma 9.14 ein linearer Unterraum). Den Kern betrachtet man im Unterschied zum Bild in aller Regel nur bei linearen Abbildungen.

9.15 Lemma *Es sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung.*

(1) Genau dann ist φ injektiv, wenn $\mathrm{Kern}(\varphi) = \{\vec{0}\}$ gilt.

(2) Genau dann ist φ surjektiv, wenn $\mathrm{Bild}(\varphi) = \mathbb{R}^m$ gilt.

Die zweite Aussage ist nichts Besonderes, denn sie gilt für jede Abbildung. Das ist einfach die Definition von Surjektivität. Die erste Aussage dagegen gilt in der Regel nur für lineare Abbildungen. Sie macht es oft leichter zu überprüfen, ob eine lineare Abbildung injektiv ist.

Beweis. (2) ist trivial.

(1) Es sei φ injektiv und sei $\vec{v} \in \operatorname{Kern}(\varphi)$. Dann folgt $\varphi(\vec{v}) = \vec{0} = \varphi(\vec{0})$ und damit $\vec{v} = \vec{0}$, weil φ injektiv ist. Also gilt $\operatorname{Kern}(\varphi) = \{\vec{0}\}$.

Es gelte umgekehrt $\operatorname{Kern}(\varphi) = \{\vec{0}\}$. Es seien $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$ mit $\varphi(\vec{v}_1) = \varphi(\vec{v}_2)$ gegeben. Dann folgt

$$\varphi(\vec{v}_1 - \vec{v}_2) = \varphi(\vec{v}_1) - \varphi(\vec{v}_2) = \vec{0}$$

also $\vec{v}_1 - \vec{v}_2 \in \operatorname{Kern}(\varphi)$ und somit $\vec{v}_1 - \vec{v}_2 = \vec{0}$, was $\vec{v}_1 = \vec{v}_2$ bedeutet. Damit ist gezeigt, dass φ injektiv ist. ■

9.16 Beispiele (1) Die Drehungen und Spiegelungen der Ebene in Beispiel 9.5 sind sowohl injektiv als auch surjektiv. Allgemeiner ist jede orthogonale Abbildung injektiv (Lemma 9.7). Der Kern all dieser Abbildungen ist also der Nullraum.

(2) Wir betrachten die Projektion

$$\pi\colon \mathbb{R}^n \to \mathbb{R}^{n-1},\ (x_1, \dots, x_n) \mapsto (x_1, \dots, x_{n-1})$$

welche die letzte Koordinate vergisst. Sie ist surjektiv, denn für jeden Vektor $\vec{y} \in \mathbb{R}^{n-1}$ gilt $\vec{y} = \pi(y_1, \dots, y_{n-1}, 0)$. Sie ist aber nicht injektiv. Ihr Kern ist

$$\operatorname{Kern}(\pi) = \{\vec{x} \in \mathbb{R}^n \mid x_1 = \cdots = x_{n-1} = 0\} = \left\{ \begin{pmatrix} 0 \\ \vdots \\ 0 \\ t \end{pmatrix} \;\middle|\; t \in \mathbb{R} \right\}.$$

Das ist eine Gerade in $\mathbb{R}^n$. ◇

9.17 Satz (Dimensionsformel für lineare Abbildungen) *Es sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ eine lineare Abbildung. Dann gilt*

$$\dim\big(\operatorname{Kern}(\varphi)\big) + \dim\big(\operatorname{Bild}(\varphi)\big) = n.$$

Beweis. Es sei $r = \dim(\operatorname{Kern}(\varphi))$ und sei $\vec{v}_1, \dots, \vec{v}_r$ eine Basis von $\operatorname{Kern}(\varphi)$. Nach dem Basisergänzungssatz 7.13 gibt es Vektoren $\vec{v}_{r+1}, \dots, \vec{v}_n$ derart, dass $\vec{v}_1, \dots, \vec{v}_n$ insgesamt eine Basis von $\mathbb{R}^n$ ist. Wir behaupten, dass das System

$$\varphi(\vec{v}_{r+1}), \dots, \varphi(\vec{v}_n)$$

eine Basis des linearen Unterraums $\mathrm{Bild}(\varphi)$ ist. Wenn das gezeigt ist, dann folgt $\dim(\mathrm{Bild}(\varphi)) = n - r = n - \dim(\mathrm{Kern}(\varphi))$ und der Satz ist bewiesen.

Wir zeigen als erstes, dass $\mathrm{Bild}(\varphi) = \mathrm{Lin}(\varphi(\vec{v}_{r+1}), \dots, \varphi(\vec{v}_n))$ gilt: Ist $\vec{w} \in \mathrm{Bild}(\varphi)$, dann gibt es $\vec{v} \in \mathbb{R}^n$ mit $\vec{w} = \varphi(\vec{v})$. Wir stellen $\vec{v}$ in der Basis $\vec{v}_1, \dots, \vec{v}_n$ da, etwa $\vec{v} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$, dann folgt

$$\begin{aligned}\vec{w} = \varphi(\vec{v}) &= \varphi(x_1\vec{v}_1 + \cdots + x_n\vec{v}_n)\\ &= \underbrace{\varphi(x_1\vec{v}_1 + \cdots + x_r\vec{v}_r)}_{=\vec{0}} + \varphi(x_{r+1}\vec{v}_{r+1} + \cdots + x_n\vec{v}_r)\\ &= x_{r+1}\varphi(\vec{v}_{r+1}) + \cdots + x_n\varphi(\vec{v}_n) \in \mathrm{Lin}(\varphi(\vec{v}_{r+1}, \dots, \varphi(\vec{v}_n))).\end{aligned}$$

Zweitens zeigen wir die lineare Unabhängigkeit von $\varphi(\vec{v}_{r+1}), \dots, \varphi(\vec{v}_n)$. Es sei

$$c_{r+1}\varphi(\vec{v}_{r+1}) + \cdots + c_n\varphi(\vec{v}_n) = \vec{0}$$

eine lineare Relation, dann müssen wir $c_{r+1} = \cdots = c_n = 0$ zeigen. Setze

$$\vec{u} = c_{r+1}\vec{v}_{r+1} + \cdots + c_n\vec{v}_n.$$

Da φ linear ist, folgt

$$\varphi(\vec{u}) = c_{r+1}\varphi(\vec{v}_{r+1}) + \cdots + c_n\varphi(\vec{v}_n) = \vec{0}$$

und damit $\vec{u} \in \mathrm{Kern}(\varphi)$. Der Vektor $\vec{u}$ besitzt deshalb eine Darstellung $\vec{u} = y_1\vec{v}_1 + \cdots + y_r\vec{v}_r$. Das bedeutet aber

$$y_1\vec{v}_1 + \cdots + y_r\vec{v}_r - c_{r+1}\vec{v}_{r+1} + \cdots - c_n\vec{v}_n = \vec{u} - \vec{u} = \vec{0}.$$

Da $\vec{v}_1, \dots, \vec{v}_n$ linear unabhängig sind, folgt daraus $y_1 = \cdots = y_r = c_{r+1} = \cdots = c_n = 0$, wie gewünscht. ∎

Dieser Beweis war ein bisschen mühsam aufzuschreiben. Wir werden später noch einen abstrakteren und in der Notation wesentlich kürzeren Beweis dieser Dimensionsformel sehen.

9.18 Korollar *Ist $\varphi\colon \mathbb{R}^n \to \mathbb{R}^m$ eine bijektive lineare Abbildung, dann gilt $m = n$.*

Beweis. Es gilt $\mathrm{Kern}(\varphi) = \{\vec{0}\}$ und $\mathrm{Bild}(\varphi) = \mathbb{R}^m$ nach Lemma 9.15. Die Dimensionsformel sagt also $n = \dim(\mathrm{Kern}(\varphi)) + \dim(\mathrm{Bild}(\varphi)) = 0 + m = m$. ∎

9.19 Korollar *Es sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ eine lineare Abbildung. (Jetzt also $m = n$).*

(1) Wenn φ injektiv ist, dann ist φ auch surjektiv.

(2) Wenn φ surjektiv ist, dann ist φ auch injektiv.

Beweis. (1) Aus φ injektiv folgt $\dim(\mathrm{Kern}(\varphi)) = 0$, also $\dim(\mathrm{Bild}(\varphi)) = n$ nach der Dimensionsformel. Damit ist $\mathrm{Bild}(\varphi)$ ein n-dimensionaler Unterraum von $\mathbb{R}^n$. Deshalb muss $\mathrm{Bild}(\varphi) = \mathbb{R}^n$ gelten.

(2) Ist φ surjektiv, dann gilt $\dim(\mathrm{Bild}(\varphi)) = n$ und aus der Dimensionsformel folgt $\dim(\mathrm{Kern}(\varphi)) = 0$. Also ist $\mathrm{Kern}(\varphi) = \{0\}$ und damit φ injektiv. ■

9.20 Korollar *Jede orthogonale Abbildung $\mathbb{R}^n \to \mathbb{R}^n$ ist bijektiv.*

Beweis. Alle orthogonalen Abbildungen sind linear und injektiv (Lemma 9.7 und Satz 9.9). Nach dem vorangehenden Korollar sind sie dann auch surjektiv. ■

10 Rang und invertierbare Matrizen

10.1 Rang einer Matrix

Definition Es sei $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ eine $m \times n$-Matrix und seien $\vec{a}_1, \ldots, \vec{a}_m \in \mathbb{R}^n$ die Zeilenvektoren von A und $\vec{b}_1, \ldots, \vec{b}_n \in \mathbb{R}^m$ die Spaltenvektoren:

$$A = \begin{pmatrix} — & \vec{a}_1 & — \\ — & \vec{a}_2 & — \\ & \vdots & \\ — & \vec{a}_m & — \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \\ | & | & & | \end{pmatrix}.$$

Der lineare Unterraum $\mathrm{Lin}(\vec{a}_1, \ldots, \vec{a}_m) \subset \mathbb{R}^n$ heißt der **Zeilenraum** von A und $\mathrm{Lin}(\vec{b}_1, \ldots, \vec{b}_n) \subset \mathbb{R}^m$ der **Spaltenraum**. Die Dimension des Zeilenraums heißt der **Zeilenrang** und die des Spaltenraums der **Spaltenrang** von A.

10.1 Lemma *(1) Der Zeilenrang einer Matrix in Zeilenstufenform ist die Anzahl der Stufen.*

(2) Der Zeilenraum und damit der Zeilenrang einer Matrix bleiben bei elementaren Zeilenumformungen unverändert.

Beweis. (1) Wir haben in Beispiel 7.5(8) überprüft, dass die Zeilenvektoren, in denen eine neue Stufe anfängt, linear unabhängig sind. Die übrigen Zeilen sind $\vec{0}$, tragen also zum Zeilenraum nichts bei.

(2) Es sei A eine $m \times n$-Matrix mit Zeilenvektoren $\vec{a}_1, \ldots, \vec{a}_m \in \mathbb{R}^n$. Ist A' eine $m \times n$-Matrix mit Zeilenvektoren $\vec{a}'_1, \ldots, \vec{a}'_m$, die aus A durch elementare Zeilenumformungen hervorgegangen ist, dann gilt $\mathrm{Lin}(\vec{a}_1, \ldots, \vec{a}_m) = \mathrm{Lin}(\vec{a}'_1, \ldots, \vec{a}'_m)$. Es reicht, das für eine einzige Zeilenumformung zu überprüfen: Bei Typ I (Vertauschen von zwei Zeilen) und Typ II (Multiplikation einer Zeile mit einem Skalar $\neq 0$) ist es offensichtlich. Und wenn bei Typ III etwa das c-fache der k-ten Zeile zur l-ten Zeile addiert wird ($c \neq 0$ und $k \neq l$), dann gilt also $\vec{a}'_l = \vec{a}_l + c\vec{a}_k$ und $\vec{a}'_i = \vec{a}_i$ für $i \neq l$. Es ist dann $\vec{a}'_l \in \mathrm{Lin}(\vec{a}_1, \ldots, \vec{a}_m)$ und umgekehrt $\vec{a}_l = \vec{a}'_l - c\vec{a}'_k \in \mathrm{Lin}(\vec{a}'_1, \ldots, \vec{a}'_m)$. Daraus folgt die Behauptung. ∎

10.2 Beispiel Wir betrachten die Matrix

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

über $\mathbb{R}$. Die Spaltenvektoren hatten wir schon öfter. Wir wissen, dass sie linear abhängig sind und einen zweidimensionalen Unterraum von $\mathbb{R}^3$ aufspannen. Der Spaltenrang von A ist also 2. Um den Zeilenrang zu bestimmen, bringen wir die Matrix in Zeilenstufenform.

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & 1 \\ 3 & -1 & 1 \end{pmatrix} \overset{\leftarrow +,\ \leftarrow +}{\rightsquigarrow} \begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & -1 \\ 0 & -4 & -2 \end{pmatrix} \overset{|\cdot(-\frac{1}{2})}{\rightsquigarrow} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & -4 & -2 \end{pmatrix} \overset{\cdot 4,\ \leftarrow +}{\rightsquigarrow} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

Der Zeilenrang ist damit ebenfalls 2. Wir beweisen gleich, dass Zeilen- und Spaltenrang immer übereinstimmen. ◇

10.3 Satz *Es sei A eine $m \times n$-Matrix und*

$$A\vec{x} = \vec{0}$$

das zugehörige homogene lineare Gleichungssystem mit Koeffizientenmatrix A. Ist d die Dimension des Lösungsraums und z der Zeilenrang von A, dann gilt

$$d + z = n.$$

Beweis. Nach Lemma 10.1(2) ändert sich der Zeilenrang nicht, wenn wir die Matrix A zuerst auf Zeilenstufenform bringen. Der Lösungsraum ändert sich dabei auch nicht. Wir können also ohne Einschränkung annehmen, dass die Matrix A bereits in Zeilenstufenform vorliegt. Ihr Zeilenrang ist dann nach 10.1(1) die Anzahl der Stufen. Die Dimension des Lösungsraums ist nach Satz 7.17 dagegen die Anzahl der Spalten, in denen keine Stufe anfängt und damit $d = n - z$. ■

10.4 Satz *Für jede Matrix stimmen Zeilen- und Spaltenrang überein.*

Beweis. Es sei A eine $m \times n$-Matrix und $\varphi_A \colon \mathbb{R}^n \to \mathbb{R}^m$ die lineare Abbildung $\vec{x} \mapsto A\vec{x}$. Sei z der Zeilenrang und s der Spaltenrang von A. Das Bild von φ_A ist der Spaltenraum von A, denn sind $\vec{b}_1, \ldots, \vec{b}_n$ die Spaltenvektoren von A, dann ist $A \cdot \vec{x}$ die Linearkombination $x_1\vec{b}_1 + \cdots + x_n\vec{b}_n$. Der Kern von φ_A ist der Lösungsraum des linearen Gleichungssystems $A\vec{x} = \vec{0}$. Ist $d = \dim(\mathrm{Kern}(\varphi_A))$, dann gilt nach der Dimensionsformel für lineare Abbildungen (Satz 9.17) $d + s = n$. Andererseits haben wir gerade auch $d + z = n$ bewiesen. Also gilt $s = z$. ■

10.5 Beispiel Wenn man die Aussage mal nur für sich betrachtet, ist sie überhaupt nicht offensichtlich. Nehmen wir die Matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{pmatrix}.$$

Die ersten beiden Zeilen sind linear unabhängig, aber die dritte ist ihre Summe. Der Zeilenrang ist deshalb 2. Nach Satz 10.4 ist auch der Spaltenrang 2. Die zehn Vektoren in $\mathbb{R}^3$ können also nicht den ganzen Raum aufspannen, sondern nur eine Ebene. Daran ändert sich auch nichts, wenn wir die Matrix nach demselben Muster immer größer machen. ◇

Definition Der **Rang** einer Matrix A ist ihr Zeilenrang, geschrieben

$$\operatorname{Rang}(A).$$

Da der Rang einer $n \times m$-Matrix A auch ihr Spaltenrang ist, stimmt er mit der Dimension des Bildes der linearen Abbildung $\varphi_A \colon \mathbb{R}^n \to \mathbb{R}^m$ überein.

10.2 Umkehrabbildung und Invertierung

Es sei $\varphi \colon \mathbb{R}^n \to \mathbb{R}^n$ eine bijektive lineare Abbildung. Die Umkehrabbildung (§5.6) $\varphi^{-1} \colon \mathbb{R}^n \to \mathbb{R}^n$ ist dadurch charakterisiert, dass sie den Effekt der Abbildung φ rückgängig macht, das heißt, es gelten

$$\varphi^{-1} \circ \varphi = \mathrm{id}_{\mathbb{R}^n} \quad \text{und} \quad \varphi \circ \varphi^{-1} = \mathrm{id}_{\mathbb{R}^n}$$

was explizit

$$\varphi^{-1}\big(\varphi(\vec{x})\big) = \vec{x} \quad \text{und} \quad \varphi\big(\varphi^{-1}(\vec{x})\big) = \vec{x}$$

für alle $\vec{x} \in \mathbb{R}^n$ bedeutet.

10.6 Lemma *Ist $\varphi \colon \mathbb{R}^n \to \mathbb{R}^n$ eine bijektive lineare Abbildung, dann ist die Umkehrabbildung φ^{-1} ebenfalls linear.*

Beweis. Es seien $\vec{w}_1, \vec{w}_2 \in \mathbb{R}^n$ und $c_1, c_2 \in \mathbb{R}$. Setze $\vec{v}_1 = \varphi^{-1}(\vec{w}_1)$ und $\vec{v}_2 = \varphi^{-1}(\vec{w}_2)$. Dann gilt

$$\begin{aligned} \varphi^{-1}(c_1\vec{w}_1 + c_2\vec{w}_2) &= \varphi^{-1}\big(c_1\varphi(\vec{v}_1) + c_2\varphi(\vec{v}_2)\big) = \varphi^{-1}\big(\varphi(c_1\vec{v}_1 + c_2\vec{v}_2)\big) \\ &= c_1\vec{v}_1 + c_2\vec{v}_2 = c_1\varphi^{-1}(\vec{w}_1) + c_2\varphi^{-1}(\vec{w}_2) \end{aligned}$$

was zeigt, dass φ^{-1} linear ist. ■

Ist nun A eine $n \times n$-Matrix, für welche die Abbildung φ_A bijektiv ist, dann gibt es nach Lemma 9.2 eine $n \times n$-Matrix B mit $(\varphi_A)^{-1} = \varphi_B$. Die Matrix B erfüllt dann nach Satz 9.12 die Gleichheit

$$AB = BA = \mathbb{1}_n.$$

Definition Eine Matrix $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ heißt **invertierbar**, wenn es eine Matrix $B \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ gibt mit $AB = BA = \mathbb{1}_n$. Wir schreiben in diesem Fall A^{-1} für B und nennen A^{-1} die **Inverse** von A.

Aufgrund der Beziehung zwischen A und φ_A folgt unmittelbar:

10.7 Satz *Genau dann ist eine $n \times n$-Matrix A invertierbar, wenn die lineare Abbildung φ_A bijektiv ist.* ∎

10.8 Lemma *Es seien $A, B \in \mathrm{Mat}_{n\times n}$ zwei invertierbare Matrizen. Dann ist auch AB invertierbar und es gilt*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Beweis. Sind φ_A und φ_B die zugehörigen bijektiven linearen Abbildungen, dann sind auch $\varphi_A \circ \varphi_B$ und $\varphi_B \circ \varphi_A$ bijektiv und linear und es gilt $(\varphi_A \circ \varphi_B)^{-1} = \varphi_B^{-1} \circ \varphi_A^{-1}$. Daraus folgt die Behauptung. ∎

Für jede invertierbare Matrix A und jedes $c \in \mathbb{R} \setminus \{0\}$ gilt außerdem

$$(cA)^{-1} = \frac{1}{c}A^{-1}$$

wie man leicht nachrechnet. Dagegen ist die Summe $A + B$ zweier invertierbarer Matrizen A und B im allgemeinen nicht invertierbar, und selbst wenn sie es ist, gibt es keine allgemeine Vereinfachung für den Ausdruck $(A + B)^{-1}$.

Wir werden mit der Zeit verschiedene Kriterien entwickeln, um festzustellen, ob eine gegebene Matrix invertierbar ist. Als erstes Zwischenergebnis in diese Richtung halten wir folgendes fest.

10.9 Lemma *Eine $n \times n$-Matrix ist genau dann invertierbar, wenn ihr Rang n ist.*

Beweis. Sei $\varphi_A \colon \mathbb{R}^n \to \mathbb{R}^n$, $\vec{x} \mapsto A\vec{x}$ die zugehörige lineare Abbildung. Das Bild von φ_A ist der Spaltenraum. Ist A invertierbar, dann ist φ_A bijektiv und damit $\mathrm{Bild}(\varphi_A) = \mathbb{R}^n$. Also hat A den Rang n.

Hat umgekehrt A den Rang n, dann der Spaltenraum gleich $\mathbb{R}^n$, also φ_A surjektiv. Nach Kor. 9.19 ist φ_A dann auch injektiv, also bijektiv. ∎

10.3 Berechnung der Inversen

Um das Inverse einer Matrix zu berechnen, kann man die Gleichung $A \cdot X = \mathbb{1}_n$ als lineares Gleichungssystem in den Spaltenvektoren von X auffassen und dieses System mit dem Gauß-Algorithmus lösen.

10.10 Beispiel Wir möchten die 2×2-Matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ invertieren. Das tun wir nach dem gerade beschriebenen Schema:

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \mid \cdot \tfrac{1}{a} \rightsquigarrow \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ c & d & 0 & 1 \end{array}\right) \begin{array}{l} \\ \longleftarrow + \end{array} \begin{array}{l} \\ \rule{0.4em}{0.4pt}\!\rule{0.4pt}{0.8em} \; -c \end{array} \rightsquigarrow \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & d - c\frac{b}{a} & -\frac{c}{a} & 1 \end{array}\right) \mid \cdot \tfrac{a}{ad-bc}$$

$$\rightsquigarrow \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right) \begin{array}{l} \\ \rule{0.4em}{0.4pt}\!\rule{0.4pt}{0.8em} \; -\frac{b}{a} \end{array} \rightsquigarrow \left(\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right) .$$

Man kann das Ergebnis einfach kontrollieren, indem man das Produkt von A und der Matrix auf der rechten Seite berechnet. Dabei haben wir $a \neq 0$ und $ad-bc \neq 0$ benutzt. Wie sieht es aus, wenn diese Bedingungen nicht erfüllt sind? Ist $a \neq 0$, aber $ad - bc = 0$, so folgt $d = \frac{bc}{a}$ und

$$\begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} b \\ \frac{bc}{a} \end{pmatrix} = \frac{b}{a} \cdot \begin{pmatrix} a \\ c \end{pmatrix}.$$

Die beiden Spaltenvektoren sind also kollinear und der Spaltenrang von A ist 1. Also ist A nicht invertierbar, nach dem vorangehenden Lemma. Ist $a = 0$, aber $ad-bc \neq 0$, dann ist also $c \neq 0$ und man kann nach Vertauschen der beiden Zeilen genauso vorgehen wie oben und bekommt die gleiche Formel für die Inverse. Wir haben also bewiesen, dass die Matrix A genau dann invertierbar ist, wenn gilt:

$$ad - bc \neq 0. \qquad \diamond$$

Das Rechenverfahren zum Invertieren einer quadratischen Matrix A ist also:

(1) Bringe A auf Zeilenstufenform.

(2) Forme die Zeilenstufenform durch weitere elementare Zeilenumformungen von unten nach oben zur Einheitsmatrix um. (Wenn das nicht geht, weil Nullen auf der Diagonalen stehen, dann ist die Matrix nicht invertierbar.)

(3) Führe alle diese elementaren Zeilenumformungen gleichzeitig an der Einheitsmatrix aus (rechte Seite). Das Ergebnis ist A^{-1}.

Wir untersuchen den Zusammenhang zwischen invertierbaren Matrizen und dem Gauß-Algorithmus genauer. Wir können elementare Zeilenumformungen

auch durch Matrizenmultiplikationen ausdrücken, zum Beispiel

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c + ta & d + tb \end{pmatrix}.$$

In dieser Weise können wir zu jedem Typ von elementarer Zeilenumformung eine entsprechende Matrix hinschreiben.

Definition Wir definieren drei Typen von **Elementarmatrizen**.

(I) Für ein Paar i, j mit $1 \leqslant i, j \leqslant n$ und $i \neq j$ sei

$$T_{ij} = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & \cdots & \cdots & \cdots & 1 & \\ & & & \vdots & 1 & & & \vdots & \\ & & & \vdots & & \ddots & & \vdots & \\ & & & \vdots & & & 1 & \vdots & \\ & & & 1 & \cdots & \cdots & \cdots & 0 & \\ & & & & & & & & 1 \\ & & & & & & & & & \ddots \\ & & & & & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{matrix}$$

(II) Für ein i mit $1 \leqslant i \leqslant n$ und $c \in \mathbb{R} \setminus \{0\}$ sei

$$M_i(c) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & c & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \\ i \\ \\ \\ \\ \end{matrix}$$

(III) Für ein Paar i, j mit $1 \leqslant i, j \leqslant n$, $i \neq j$ und ein $c \in \mathbb{R}$ sei

$$S_{ij}(c) = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & \cdots & c & \\ & & & \ddots & \vdots & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \\ i \\ \\ j \\ \\ \end{matrix}$$

Die Elementarmatrizen entstehen also gerade dadurch, dass man die entsprechende elementare Zeilenumformung auf die Einheitsmatrix $\mathbb{1}_n$ anwendet. Ist dann A eine beliebige $m \times n$-Matrix, dann ergibt die Multiplikation mit einer $m \times m$-Elementarmatrix von links genau die jeweilige Zeilenumformung:

Matrix	Zeilenumformung
$T_{ij} \cdot A$	Vertauschen der i-ten und der j-ten Zeile in A
$M_i(c) \cdot A$	Multiplikation der i-ten Zeile von A mit c
$S_{ij}(c) \cdot A$	Addition des c-fachen der j-ten Zeile zur i-ten Zeile von A

Die Elementarmatrizen sind allesamt invertierbar, weil wir die entsprechende Zeilenumformung ja auch wieder rückgängig machen können. Es gelten demnach

$$T_{ij}^{-1} = T_{ij}, \qquad M_i(c)^{-1} = M_i(c^{-1}), \qquad S_{ij}(c)^{-1} = S_{ij}(-c).$$

Dass wir jede Matrix in eine Zeilenstufenform bringen können (mit dem Gauß-Algorithmus), lässt sich damit auch so formulieren:

10.11 Satz *Zu jeder $m \times n$-Matrix A gibt es eine invertierbare $m \times m$-Matrix S, die ein Produkt von Elementarmatrizen ist, derart, dass die Matrix*

$$S \cdot A$$

Zeilenstufenform besitzt. ■

Dieses Verfahren wenden wir nun auf invertierbare Matrizen an.

10.12 Satz *Eine $n \times n$-Matrix ist genau dann invertierbar, wenn sie sich durch elementare Zeilenumformungen in die Einheitsmatrix $\mathbb{1}_n$ überführen lässt.*

Beweis. Es sei A eine $n \times n$-Matrix. Angenommen A ist invertierbar. Als erstes finden wir ein Produkt S von Elementarmatrizen derart, dass SA Zeilenstufenform hat, wie gerade bemerkt. Da A und S invertierbar sind, ist auch $B = SA$ invertierbar und hat deshalb nach Lemma 10.9 den Rang n. Da der Rang auch die Anzahl der Stufen ist, hat die Matrix B die Gestalt

$$B = \begin{pmatrix} 1 & * & \cdots & * \\ & \ddots & & \vdots \\ & & \ddots & * \\ & & & 1 \end{pmatrix}$$

Nun können wir weitere elementare Zeilenumformungen anwenden und alle Einträge oberhalb der Diagonalen zu 0 machen (wie im Gauß-Algorithmus, nur von unten nach oben). Wir erhalten so ein Produkt T von Elementarmatrizen mit $TB = TSA = \mathbb{1}_n$.

Da Elementarmatrizen invertierbar sind, folgt aus der Existenz einer solchen Darstellung umgekehrt, dass A invertierbar ist. ■

10.13 Korollar *Jede invertierbare Matrix ist ein Produkt von Elementarmatrizen.*

Beweis. Denn ist A eine invertierbare $n \times n$-Matrix, dann gibt es nach dem gerade bewiesenen Satz ein Produkt S von Elementarmatrizen mit $SA = \mathbb{1}_n$. Daraus folgt $A = S^{-1}$. Da die Inverse einer Elementarmatrix wieder eine Elementarmatrix ist, folgt daraus die Behauptung. ∎

10.4 Anwendung auf lineare Gleichungssysteme

In diesem kurzen Abschnitt wenden wir einige vorige Ergebnisse auf lineare Gleichungssysteme an.

10.14 Satz *Es sei $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ und sei $\vec{b} \in \mathbb{R}^m$. Dann sind äquivalent:*

(1) Das lineare Gleichungssystem $A\vec{x} = \vec{b}$ ist lösbar.

(2) Der Vektor $\vec{b}$ liegt im Bild der linearen Abbildung $\varphi_A \colon \mathbb{R}^n \to \mathbb{R}^m,\ \vec{x} \mapsto A\vec{x}$.

(3) Der Vektor $\vec{b}$ liegt im Spaltenraum der Matrix A.

(4) Für die erweiterte Koeffizientenmatrix $(A|\vec{b})$ gilt $\mathrm{Rang}(A|\vec{b}) = \mathrm{Rang}(A)$.

Beweis. (1)⇔(2): Die Lösungen von $A\vec{x} = \vec{b}$ sind die Urbilder von $\vec{b}$ unter φ_A.

(2)⇔(3): Der Spaltenraum ist genau das Bild von φ_A.

(3)⇒(4): Wenn $\vec{b}$ im Spaltenraum von A liegt, dann haben A und $(A|\vec{b})$ also denselben Spaltenraum und damit denselben (Spalten)rang.

(4)⇒(3): Aus $\mathrm{Rang}(A|\vec{b}) = \mathrm{Rang}(A)$ folgt, dass der Spaltenraum von A und der von $A|\vec{b}$ übereinstimmen. Also muss $\vec{b}$ im Spaltenraum von A liegen. ∎

Die Äquivalenz zwischen (1) und (4) beschreibt genau die Bedingung, die wir überprüfen, wenn wir in der Zeilenstufenform die letzte Stufe betrachten, um die Lösbarkeit des Gleichungssystems zu entscheiden.

Für quadratische Matrizen ($m = n$) können wir auch noch den Zusammenhang zur Invertierbarkeit der Koeffizientenmatrix herstellen.

10.15 Satz *Es sei $A \in \mathrm{Mat}_{n\times n}(\mathbb{R})$. Dann sind äquivalent:*

(1) Für jedes $\vec{b} \in \mathbb{R}^n$ ist das lineare Gleichungssystem $A\vec{x} = \vec{b}$ lösbar.

(2) Für jedes $\vec{b} \in \mathbb{R}^n$ ist das lineare Gleichungssystem $A\vec{x} = \vec{b}$ eindeutig lösbar.

(3) Das homogene lineare Gleichungssystem $A\vec{x} = \vec{0}$ hat nur die Lösung $\vec{x} = \vec{0}$.

(4) Die Matrix A hat Rang n, ist also invertierbar.

Beweis. (1)⇒ (2): Die Aussage (1) bedeutet, dass die lineare Abbildung $\varphi_A\colon \vec{x} \mapsto A\vec{x}$ surjektiv ist (siehe auch Satz 10.14). Wegen $m = n$ ist sie nach Kor. 9.19 dann aber auch injektiv, was die Eindeutigkeit der Lösung zeigt.

(2)⇒ (3): Dazu muss man nur (2) auf $\vec{b} = \vec{0}$ anwenden.

(3)⇒ (4): Aus (3) folgt $\operatorname{Kern}(\varphi_A) = \{\vec{0}\}$. Also ist die lineare Abbildung φ_A injektiv und damit, wieder nach Kor. 9.19, auch surjektiv. Sie ist also bijektiv und A damit invertierbar. Nach Lemma 10.9 ist das äquivalent zu $\operatorname{Rang}(A) = n$.

(4)⇒(1): Ist A invertierbar, dann ist $\vec{x} = A^{-1}\vec{b}$ eine Lösung. ∎

Ein homogenes lineares Gleichungssystem bestimmt einen linearen Unterraum, den Lösungsraum. Der Gauß-Algorithmus läuft darauf hinaus, eine Basis des Lösungsraums zu bestimmen (Satz 7.17). Wir können auch den umgekehrten Weg gehen und aus einem linearen Unterraum mit Basis ein dazu passendes lineares Gleichungssystem herstellen.

10.16 Satz *Jeder lineare Unterraum der Dimension m von $\mathbb{R}^n$ ist die Lösungsmenge eines homogenen linearen Gleichungssystem mit $n-m$ Gleichungen in n Unbekannten.*

Beweis. Es sei $U \subset \mathbb{R}^n$ ein linearer Unterraum der Dimension m. Dann gibt es also $\vec{u}_1, \dots, \vec{u}_m \in \mathbb{R}^n$ mit $U = \operatorname{Lin}(\vec{u}_1, \dots, \vec{u}_m)$. Sei A die Matrix mit Zeilenvektoren $\vec{u}_1, \dots, \vec{u}_m$. Diese $m \times n$-Matrix hat dann also den Zeilenrang m. Nach Satz 10.3 hat der Lösungsraum des zugehörigen homogenen linearen Gleichungssystem mit Koeffizientenmatrix A die Dimension $n - m$. Es sei $\vec{v}_1, \dots, \vec{v}_{n-m}$ eine Basis dieses Lösungsraums. Für alle $i = 1, \dots, m$ und $k = 1, \dots, n - m$ gilt also

$$\sum_{j=1}^{n} u_{ij} v_{kj} = 0.$$

In diesem Gleichungssystem vertauschen wir nun die Rollen der Koeffizienten und der Lösungen: Die Vektoren $\vec{u}_1, \dots, \vec{u}_m$ sind Lösungsvektoren des homogenen linearen Gleichungssystems

$$\sum_{j=1}^{n} v_{kj} x_j = 0 \qquad (k = 1, \dots, n - m)$$

in den Unbekannten x_j. Die Koeffizientenmatrix hat Zeilenvektoren $\vec{v}_1, \dots, \vec{v}_{n-m}$ und die Vektoren $\vec{u}_1, \dots, \vec{u}_m$ sind Lösungen. Ist V der Lösungsraum dieses neuen Gleichungssystems, dann gilt also $U \subset V$. Wiederum nach Satz 10.3 gilt nun

$$\dim(V) = n - (n - m) = m = \dim(U).$$

Wegen $U \subset V$ folgt daraus $U = V$. Der Unterraum U ist also der Lösungsraum des homogenen linearen Gleichungssystems mit den Zeilenvektoren $\vec{v}_1, \dots, \vec{v}_{n-m}$. ∎

Jeder lineare Unterraum von $\mathbb{R}^n$ hat also zwei verschiedene Beschreibungen: Eine **implizite Beschreibung**, als die Lösungsmenge eines homogenen linearen Gleichungssystems und eine **explizite Beschreibung** (oder **Parametrisierung**) durch eine Basis, oder wenigstens ein Erzeugendensystem. Ein lineares Gleichungssystem zu lösen bedeutet, von einer impliziten Beschreibung des Lösungsraums zu einer expliziten überzugehen. Manchmal ist aber auch die implizite Beschreibung einfacher als die explizite:

10.17 Beispiel Sei $U \subset \mathbb{R}^n$ ein linearer Unterraum. Angenommen, wir möchten wissen, ob U einen gegebenen Vektor $\vec{v}$ enthält. Ist U implizit als Lösungsraum eines linearen Gleichungssystem gegeben, dann müssen wir nur einsetzen, um zu testen, ob $\vec{v}$ diese Gleichungen löst oder nicht. Ist dagegen $U = \operatorname{Lin}(\vec{v}_1, \dots, \vec{v}_k)$ explizit gegeben, dann müssen wir das lineare Gleichungssystem

$$\sum_{i=1}^{k} x_i \vec{v}_i = \vec{v}$$

in k Unbekannten $x_1, \dots, x_k$ lösen. ◇

Das alles gilt genauso für inhomogene lineare Gleichungssysteme und affine Unterräume.

10.18 Korollar *Jeder affine Unterraum der Dimension m von $\mathbb{R}^n$ ist die Lösungsmenge eines inhomogenen linearen Gleichungssystems mit $n - m$ Gleichungen in n Unbekannten.*

Die Dimension eines affinen Unterraums $\vec{v} + U$ ist dabei die Dimension des linearen Unterraums U.

Beweis. Es sei $A = \vec{v} + U$ mit $\vec{v} \in \mathbb{R}^n$ und $U \subset \mathbb{R}^n$ ein linearer Unterraum. Nach Satz 10.16 ist U der Lösungsraum eines homogenen linearen Gleichungssystems

$$\sum_{j=1}^{n} a_{ij} x_j = 0, \quad i = 1, \dots, n - m.$$

Setze $b_i = \sum_{j=1}^{n} a_{ij} v_j$. Dann ist $\vec{v} + U$ nach Satz 3.4 der affine Lösungsraum des inhomogenen linearen Gleichungssystems

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, \dots, n - m.$$ ■

III

Zahlen und algebraische Strukturen

The science of operations, as derived from mathematics more especially, is a science of itself, and has its own abstract truth and value.

Ada Lovelace (1815–1852)

11 Gruppen

Bisher haben wir die lineare Algebra in $\mathbb{R}^n$ studiert. Dabei haben wir die reellen Zahlen mit ihren Eigenschaften und Rechenregeln als gegeben vorausgesetzt. Viele dieser Regeln sind allen Zahlbereichen gemeinsam. Außerdem haben wir Rechenregeln für Vektoraddition, Matrizenprodukt usw. gesehen. Das betrachten wir nun sehr viel abstrakter und allgemeiner, indem wir Strukturen mit bestimmten Rechenregeln systematisch untersuchen.

11.1 Verknüpfungen

Definition Eine **Verknüpfung** auf einer Menge M ist eine Abbildung

$$M \times M \to M.$$

Eine Verknüpfung ordnet also jedem Paar von Elementen aus M ein neues Element aus M zu.

11.1 Beispiele (1) Addition und Multiplikation sind Verknüpfungen auf den Zahlmengen $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$.

(2) Das Matrizenprodukt ist eine Verknüpfung auf der Menge $\mathrm{Mat}_{n\times n}(\mathbb{R})$, die einem Paar (A, B) von $n \times n$-Matrizen ihr Produkt $A \cdot B$ zuordnet. Die Matrizenmultiplikation ist allgemeiner eine Abbildung $\mathrm{Mat}_{l\times m}(\mathbb{R}) \times \mathrm{Mat}_{m\times n}(\mathbb{R}) \to \mathrm{Mat}_{l\times n}(\mathbb{R})$, aber keine Verknüpfung, da die drei Mengen nicht identisch sind.

(3) Aus dem gleichen Grund sind die Skalarmultiplikation $\mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ und das Skalarprodukt $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ von Vektoren keine Verknüpfungen.

(3) Die Subtraktion ist keine Verknüpfung auf $\mathbb{N}$, denn für $a, b \in \mathbb{N}$ mit $b > a$ ist $a - b$ keine natürliche Zahl. Im Begriff der Abbildung ist enthalten, dass die Verknüpfung für *alle* Paare von Elementen der zugrundeliegenden Menge definiert sein muss. Die Subtraktion ist eine Verknüpfung auf der Menge $\mathbb{Z}$.

(4) Die Division ist keine Verknüpfung auf $\mathbb{R}$, denn $(x, y) \mapsto \frac{x}{y}$ ist nicht definiert für $y = 0$. Sie ist aber eine Verknüpfung auf $\mathbb{R} \setminus \{0\}$. ◇

11.2 Gruppenaxiome

Obwohl die Zahlbereiche immer zwei Verknüpfungen tragen, Addition und Multiplikation, betrachten wir als erstes Strukturen mit nur einer Verknüpfung. Die wichtigste solche Struktur ist die einer Gruppe.

Definition Eine **Gruppe** ist ein Paar $(G, *)$ bestehend aus einer nicht-leeren Menge G und einer Verknüpfung

$$\begin{cases} G \times G & \to & G \\ (a, b) & \mapsto & a * b \end{cases}$$

derart, dass die folgenden Eigenschaften erfüllt sind:

(A) Für alle $a, b, c \in G$ gilt $(a * b) * c = a * (b * c)$.
(*Assoziativität*)

(N) Es gibt ein Element $e \in G$ mit $e * a = a$ für alle $a \in G$.
(*Existenz des neutralen Elements*)

(I) Zu jedem $a \in G$ gibt es ein Element $a' \in G$ mit $a' * a = e$.
(*Existenz von Inversen*)

Die Gruppe G heißt **kommutativ** oder **abelsch**[1], wenn zusätzlich gilt:

(K) Für alle $a, b \in G$ gilt $a * b = b * a$. (*Kommutativität*)

11.2 Beispiele (1) Die ganzen Zahlen $\mathbb{Z}$ bilden eine abelsche Gruppe mit der Addition + als Verknüpfung. Das neutrale Element ist 0 und das additive Inverse einer Zahl $a \in \mathbb{Z}$ ist $a' = -a$.

(2) Die natürlichen Zahlen $\mathbb{N}_0$ bilden mit der Addition keine Gruppe, da sie zwar das neutrale Element 0 besitzen, es zu $n \in \mathbb{N}_0$ mit $n \neq 0$ aber kein additives Inverses gibt.

(3) Die Vektoren in $\mathbb{R}^n$ bilden unter der Addition die **Vektorgruppe**. Das neutrale Element ist der Nullvektor $\vec{0}$.

(4) Es gibt auch Gruppen mit nur endlich vielen Elementen. Auf einer einelementigen Menge $G = \{e\}$ gibt es für die Verknüpfung nur eine Möglichkeit, nämlich $e * e = e$. Die Gruppenaxiome sind dann alle erfüllt, so dass $(G, *)$ eine Gruppe (mit neutralem Element e und Inversem $e' = e$) ist. Diese Gruppe heißt die **triviale Gruppe**.

[1] Niels Henrik Abel (1802–1829) war ein norwegischer Mathematiker. Nach ihm ist heute auch der höchst renommierte Abel-Preis benannt.

(5) Bei endlichen Gruppe können wir die Verknüpfung in einer Tabelle festhalten, der sogenannten **Gruppentafel**. Zum Beispiel bildet die Menge $G = \{e, a\}$ mit zwei verschiedenen Elementen e und a eine Gruppe mit der Verknüpfung $e * a = a * e = a$ und $e * e = a * a = e$. Die Gruppentafel dazu ist also

G	e	a
e	e	a
a	a	e

Es ist leicht zu überprüfen, dass die Gruppenaxiome alle erfüllt sind.

(6) Auch auf einer Menge mit drei Elementen $G = \{e, a, b\}$ gibt es eine prima Gruppenstruktur, und zwar

G	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

Hier wird es aber schon ein wenig mühsam, dass Assoziativgesetz nur anhand der Gruppentafel nachzuprüfen, während man $a' = b$ und $b' = a$ leicht abliest. Außerdem sehen wir, dass auch diese Gruppe abelsch ist. ◇

11.3 Die allgemeine lineare Gruppe

Die interessanteste Gruppe in der linearen Algebra ist die Gruppe der invertierbaren Matrizen:

11.3 Satz *Für jedes $n \in \mathbb{N}$ bildet die Menge der invertierbaren $n \times n$-Matrizen*

$$\mathrm{GL}_n(\mathbb{R}) = \big\{A \in \mathrm{Mat}_{n\times n}(\mathbb{R}) \mid A \text{ ist invertierbar}\big\}$$

eine Gruppe mit der Matrizenmultiplikation als Verknüpfung. Das neutrale Element ist die Einheitsmatrix $\mathbb{1}_n$. Für $n \geqslant 2$ ist $\mathrm{GL}_n(\mathbb{R})$ nicht abelsch.

Die Gruppe $\mathrm{GL}_n(\mathbb{R})$ heißt die **allgemeine lineare Gruppe**.

Beweis. Zunächst halten wir fest, dass das Produkt zweier invertierbarer Matrizen wieder invertierbar ist, nach Lemma 10.8. Deshalb ist die Matrizenmultiplikation auch wirklich eine Verknüpfung auf der Menge $\mathrm{GL}_n(\mathbb{R})$. Wir wissen auch schon, dass die Einheitsmatrix $\mathbb{1}_n A = A$ für alle $n \times n$-Matrizen A erfüllt. Sie ist also ein neutrales Element. Die Existenz von Inversen ist auch klar nach Definition, nämlich $A' = A^{-1}$. Es bleibt die Assoziativität der Matrizenmultiplikation

zu zeigen. Wir könnten das, etwas mühsam, direkt nachrechnen. Einfacher ist der Umweg über lineare Abbildungen: Seien $A, B, C \in \mathrm{GL}_n(\mathbb{R})$ und $\varphi_A, \varphi_B, \varphi_C$ die zugehörigen bijektiven linearen Abbildungen $\mathbb{R}^n \to \mathbb{R}^n$. Die Assoziativität der Komposition von Abbildungen ist offensichtlich: Es gilt $(\varphi_A \circ \varphi_B) \circ \varphi_C = (\varphi_A \circ (\varphi_B \circ \varphi_C)$, denn beide Abbildungen geben angewendet auf jedes $\vec{x} \in \mathbb{R}^n$ einfach $\varphi_A(\varphi_B(\varphi_C(\vec{x})))$. Mit Satz 9.12 folgt nun

$$\varphi_{(AB)C} = \varphi_{AB} \circ \varphi_C = (\varphi_A \circ \varphi_B) \circ \varphi_C = \varphi_A \circ (\varphi_B \circ \varphi_C) = \varphi_{A(BC)}.$$

Da sich Matrizen und lineare Abbildungen gegenseitig eindeutig bestimmen, muss $(AB)C = A(BC)$ gelten. Also ist $\mathrm{GL}_n(\mathbb{R})$ eine Gruppe. Die Rechnung

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{aber} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

zeigt außerdem, dass $\mathrm{GL}_2(\mathbb{R})$ nicht abelsch ist. Also ist $\mathrm{GL}_n(\mathbb{R})$ für $n \geqslant 3$ erst recht nicht abelsch (Übung). ■

11.4 Folgerungen aus den Gruppenaxiomen

Wir sehen, dass bisher schon ziemlich viele verschiedene Gruppen vorgekommen sind, die alle ihre Besonderheiten haben. Wenn wir aber generelle Folgerungen aus den Gruppenaxiomen selbst ziehen, dann gelten sie in jeder Gruppe, ohne dass man das je wieder überprüfen muss.

11.4 Lemma *Es sei $(G, *)$ eine Gruppe.*

(1) *Das neutrale Element $e \in G$ (mit der Eigenschaft $e * a = a$ für alle $a \in G$) erfüllt auch $a * e = a$ für alle $a \in G$.*

(2) *Für alle $a \in G$ erfüllt das inverse Element $a' \in G$ (mit der Eigenschaft $a' * a = e$) auch $a * a' = e$.*

(3) *Das neutrale Element ist eindeutig bestimmt.*

(4) *Das inverse Element a' ist durch a eindeutig bestimmt.*

(5) *Es gilt $e' = e$, das heißt, das neutrale Element ist zu sich selbst invers.*

(6) *Für alle $a \in G$ gilt $(a')' = a$.*

(7) *Für alle $a, b \in G$ gilt*

$$(a * b)' = b' * a'.$$

Beweis. Solche Beweise sind wie ein Puzzlespiel, in dem ausschließlich die drei Gruppenaxiome (A), (N) und (I) verwendet werden dürfen, um daraus die Behauptungen eine nach der anderen zusammen zu setzen. Die Kunst besteht unter anderem darin, die einfachste Reihenfolge zu finden.

Wir beginnen mit (2). Es sei $a \in G$ und sei $a' \in G$ mit $a' * a = e$. Zu a' gibt es wieder ein Inverses a'' mit $a'' * a' = e$. Daraus folgt

$$\begin{aligned} a * a' &= e * (a * a') = (a'' * a') * (a * a') \\ &= a'' * \big((a' * a) * a'\big) = a'' * (e * a') = a'' * a' = e. \end{aligned}$$

(1) Unter Verwendung von (2) gilt für jedes $a \in G$ nun

$$a * e = a * (a' * a) = (a * a') * a = e * a = a.$$

(3) Die Eindeutigkeit bedeutet hier Folgendes: Es sei $f \in G$ ein Element mit $f * a = a$ für alle $a \in G$. Dann müssen wir $f = e$ beweisen. Nach (1) wissen wir aber schon $a * e = a$ für alle a, und für $a = f$ damit

$$f = f * e = e.$$

(4) Sei $b \in G$ mit $b * a = e$. Mit (1) folgt dann

$$a' = e * a' = (b * a) * a' = b * (a * a') = b * e = b.$$

(5) Es gilt $e * e = e$ per Definition, also $e' = e$.

(6) Ebenso gilt $a * a' = e$ nach (2) und damit $a = (a')'$.

(7) Es gilt $(b' * a') * (a * b) = b' * (a' * a) * b = b' * e * b = b' * b = e$. Nach (4) folgt daraus $b' * a' = (a * b)'$. ∎

12 Ringe und ganze Zahlen

12.1 Ringaxiome

Nach den Gruppen mit nur einer Verknüpfung kommen wir zu den Ringen mit zwei Verknüpfungen, die näher an den bekannten Zahlbereichen sind.

Definition Ein **Ring** ist eine Menge R mit zwei Verknüpfungen

$$\begin{cases} R \times R & \to & R \\ (x,y) & \mapsto & x+y \end{cases} \quad \text{und} \quad \begin{cases} R \times R & \to & R \\ (x,y) & \mapsto & x \cdot y \end{cases}$$

(Addition und Multiplikation) mit den folgenden Eigenschaften:

(AA) Für alle $x, y, z \in R$ gilt $(x+y)+z = x+(y+z)$.
(*Assoziativität der Addition*)

(AK) Für alle $x, y \in R$ gilt $x+y = y+x$.
(*Kommutativität der Addition*)

(AN) Es gibt ein Element $0_R \in R$ mit $x + 0_R = x$ für alle $x \in R$.
(*Existenz der Null*)

(AI) Für jedes $x \in R$ gibt es ein Element $-x \in R$ mit $x + (-x) = 0$.
(*Existenz von additiven Inversen*)

(MA) Für alle $x, y, z \in R$ gilt $(xy)z = x(yz)$.
(*Assoziativität der Multiplikation*)

(MN) Es gibt ein Element $1_R \in R$ mit $x \cdot 1_R = 1_R \cdot x = x$ für alle $x \in R$.
(*Existenz der Eins*)

(D) Für alle $x, y, z \in R$ gelten $(x+y)z = xz + yz$ und $x(y+z) = xy + xz$.
(*Distributivität*)

Die Eigenschaften (AA), (AK), (AN) und (AI) sagen gerade, dass $(R, +)$ eine abelsche Gruppe ist. Dagegen ist $(R, \cdot)$ in aller Regel keine Gruppe. Das Distributivgesetz (D) ist das einzige, das beide Operationen zusammen bringt.

Den Index R an der 0 und der 1 lässt man in aller Regel weg, aber für den Moment bleiben wir dabei, um die Null und die Eins in einem abstrakten Ring von den Zahlen 0 und 1 zu unterscheiden.

Definition Ein Ring R heißt **kommutativ**, wenn zusätzlich Folgendes gilt:

(MK) Für alle $x, y \in R$ gilt $xy = yx$. (*Kommutativität der Multiplikation*)

12.1 Beispiele (1) Die Mengen $\mathbb{Z}$, $\mathbb{Q}$ und $\mathbb{R}$ mit der üblichen Addition und Multiplikation sind kommutative Ringe. Dagegen ist $\mathbb{N}$ kein Ring, da die Ringgesetze (AN) und (AI) in $\mathbb{N}$ nicht erfüllt sind.

(2) Die $n \times n$-Matrizen bilden mit der eintragsweisen Addition

$$A + B = (a_{ij} + b_{ij})_{i,j=1,\dots,n}$$

und dem Matrizenprodukt AB den **Matrizenring** $\mathrm{Mat}_{n\times n}(\mathbb{R})$. Die additiven Eigenschaften (AA), (AK), (AN), (AI) übertragen sich sofort von $(\mathbb{R}, +)$ auf $(\mathrm{Mat}_{n\times n}(\mathbb{R}), +)$. Die Assoziativität der Multiplikation kann man genauso beweisen, wie wir es schon für invertierbare Matrizen in Satz 11.3 getan haben. Die Eins ist die Einheitsmatrix $\mathbb{1}_n$. Wir müssten noch das Distributivgesetz nachrechnen. Das lassen wir an dieser Stelle aus und holen es später nach. Der Matrizenring ist für $n \geqslant 2$ nicht kommutativ, da die Matrizenmultiplikation schon in $\mathrm{GL}_n(\mathbb{R})$ nicht kommutativ ist.

(3) Es sei $R = \{a\}$ die Menge mit nur einem Element a. Darauf gibt es die Addition und Multiplikation $a + a = a$ und $a \cdot a = a$. Alle Ringgesetze sind trivialerweise erfüllt. Es gilt allerdings $0_R = 1_R = a$. Dieser Ring heißt der **Nullring**. Er ist natürlich reichlich langweilig, aber es gibt systematische Gründe, ihn nicht auszuschließen. ◇

In den Ringen $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ von Zahlen gelten noch weitere Rechengesetze außer den Ringgesetzen. Aber welche davon lassen sich allein aus den Ringgesetzen folgern und gelten damit in jedem Ring?

12.2 Satz *Es sei R ein Ring.*

(1) Für alle $x, y, z \in R$ gilt die Implikation $x + z = y + z \implies x = y$.

(2) Für alle $x \in R$ gilt $0_R \cdot x = x \cdot 0_R = 0_R$.

(3) Für alle $x \in R$ gilt $-1_R \cdot x = -x$.

(4) Es gilt $(-1_R)(-1_R) = 1_R$.

Beweis. (1) Es gelte $x + z = y + z$. Nach den Ringgesetzen folgt dann[1]

$$\begin{aligned} x &= x + 0_R = x + (z + (-z)) = (x + z) + (-z) = (y + z) + (-z) \\ &= y + (z + (-z)) = y + 0_R = y. \end{aligned}$$

(2) Es gilt

$$0_R \cdot x = (0_R + 0_R) \cdot x = 0_R \cdot x + 0_R \cdot x.$$

Es gilt also $0_R + 0_R \cdot x = 0_R \cdot x + 0_R \cdot x$ und damit $0_R \cdot x = 0_R$ nach (1). Der Beweis für $x \cdot 0_R$ geht genauso.

(3) Nach (2) gilt $0_R = 0_R \cdot x = (1_R + (-1_R))x = x + (-1_R)x$. Addition von $-x$ auf beiden Seiten liefert also $(-1_R)x = -x$ wie behauptet.

(4) Nach (3) gilt $(-1_R) \cdot (-1_R) = -(-1_R) = 1_R$. ∎

12.3 Lemma *In einem Ring sind die Eins, die Null und die additiven Inversen eindeutig bestimmt.*

Beweis. Sei R ein Ring. Da $(R, +)$ eine Gruppe ist, haben wir die Eindeutigkeit der Null und der additiven Inversen schon mit Lemma 11.4 erledigt. Auch der Beweis für die Eindeutigkeit der Eins geht im Prinzip genauso: Seien $e, e' \in R$ zwei Elemente mit der Eigenschaft der Eins, also

$$\forall x \in R\colon\ x \cdot e = e \cdot x = x \quad \text{und} \quad \forall x \in R\colon\ x \cdot e' = e' \cdot x = x.$$

Dann können wir die linke Eigenschaft für $x = e'$ anwenden und die rechte für $x = e$ und erhalten $e' = e' \cdot e = e$. Es gibt also in R nur ein einziges solches Element, das wir mit 1_R bezeichnen. ∎

In jedem Ring R können wir die **Subtraktion** als Verknüpfung $R \times R \to R$ durch $x - y = x + (-y)$ definieren[2]. Ist ferner $n \in \mathbb{N}$ eine natürliche Zahl und $a \in R$, dann schreiben wir

$$n \cdot x = \underbrace{x + \cdots + x}_{n \text{ mal}}$$

sowie $0 \cdot x = 0_R$. Es gilt dann $(m + n)x = mx + nx$ für alle $m, n \in \mathbb{N}$.
Entsprechend definieren wir[3]

$$x^n = \underbrace{x \cdots x}_{n \text{ mal}}$$

[1] Jeder Schritt benutzt genau ein Ringgesetz. Welches ist es jeweils?

[2] Das Minuszeichen hat damit wie üblich zwei verschiedene Bedeutungen.

[3] Es ist wichtig, dass der Exponent eine natürliche Zahl ist. Für $x, y \in R$ hat die Schreibweise x^y in der Regel keine Bedeutung.

und es gilt dann

$$x^{m+n} = x^m \cdot x^n.$$

Außerdem definieren wir noch $a^0 = 1$.

12.4 Beispiel Sei R ein Ring und seien $x, y \in R$ mit $xy = yx$. Für jedes $n \in \mathbb{N}_0$ gilt dann die **binomische Formel**

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Dabei ist $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ der Binomialkoeffizient. Das kann man genauso beweisen wie für reelle Zahlen (siehe Analysis-Vorlesung). Die Voraussetzung $xy = yx$ ist wesentlich: Sind zum Beispiel $A, B \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ zwei nicht-kommutierende Matrizen, dann ist

$$(A+B)^2 = (A+B)(A+B) = A(A+B) + B(A+B) = A^2 + AB + BA + B^2$$

und die Summe $AB + BA$ lässt sich wegen $AB \neq BA$ nicht zusammenfassen. ◇

12.2 Ganze Zahlen

Die natürlichen Zahlen $\mathbb{N}$ bilden keinen Ring, sind aber im Ring $\mathbb{Z}$ enthalten. Eine wichtige Eigenschaft, die die natürlichen Zahlen von den anderen bekannten Zahlbereichen $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ unterscheidet, ist die folgende:

Jede nicht-leere Menge natürlicher Zahlen besitzt ein kleinstes Element.

(Diese Tatsache begründet unter anderem das Beweisprinzip der vollständigen Induktion.) Wir nehmen das als gegeben an und versuchen nicht, es aus Axiomen oder Rechenregeln für natürliche Zahlen herzuleiten.

Bei den ganzen Zahlen $\mathbb{Z}$ kommen zu den natürlichen Zahlen die Null und die negativen Zahlen hinzu. Viele interessante Fragen über ganze und natürliche Zahlen betreffen die **Teilbarkeit**. Sind a und b ganze Zahlen, dann sagen wir b **teilt** a oder b **ist ein Teiler von** a, wenn es eine ganze Zahl q gibt mit

$$a = bq$$

und schreiben in diesem Fall kurz

$$b|a.$$

Außerdem schreiben wir $q = a/b$ oder $q = \frac{a}{b}$, falls $b \neq 0$. Wenn b kein Teiler von a ist, dann schreiben wir $b \nmid a$.

12.5 Satz (Eigenschaften der Teilbarkeit) *Es seien $a, b, c, d \in \mathbb{Z}$.*

(1) *Es gilt $a|a$.* *(Reflexivität)*

(2) *Für $a \neq 0$ gilt $a|0$ aber $0 \nmid a$.*

(3) *Aus $a \neq 0$ und $b|a$ folgt $|b| \leqslant |a|$.*

(4) *Falls $b|a$ und $a|b$, dann gilt $a = \pm b$.*

(5) *Aus $c|b$ und $b|a$ folgt $c|a$.* *(Transitivität)*

(6) *Aus $c|a$ und $d|b$ folgt $cd|ab$.* *(Verträglichkeit mit der Multiplikation)*

(7) *Aus $c|a$ und $c|b$ folgt $c|(a+b)$.* *(Verträglichkeit mit der Addition)*

(8) *Ist $c \neq 0$ und gilt $bc|ac$, so folgt $b|a$.*

Beweis. (1) Es gilt $a = a \cdot 1$.

(2) Für jedes $a \in \mathbb{Z}$ gilt $0 = a \cdot 0$, also $a|0$. Andererseits bedeutet $0|a$, dass es $q \in \mathbb{Z}$ mit $a = q \cdot 0$ gibt. Es folgt $a = 0$.

(3) Es sei $a = bq$. Dann gilt auch $|a| = |b||q|$. Wegen $a \neq 0$ gilt auch $q \neq 0$ und damit $|q| \geqslant 1$. Es folgt $|a| = |b||q| \geqslant |b|$.

(4) Es sei $a = bq$ und $b = aq'$, dann folgt $a = aqq'$. Falls $a = 0$, dann ist auch $b = 0$ und die Behauptung richtig. Andernfalls folgt $qq' = 1$ nach der Kürzungsregel und damit $q = q' = \pm 1$.

(5) Es sei $b = cq$ und $a = bq'$, dann folgt $a = bq' = cqq'$, also $c|a$.

(6) Es sei $a = cq$ und $b = dq'$, also $ab = cqdq' = (cd)(qq')$ und damit $cd|ab$.

(7) Es sei $a = cq$ und $b = cq'$, dann folgt $a + b = c(q + q')$, also $c|(a + b)$.

(8) Es gelte $bc|ac$, also $ac = bcq$ für ein $q \in \mathbb{Z}$. Es folgt $c(a - bq) = ca - cbq = 0$. Wegen $c \neq 0$ folgt mit der Kürzungsregel $a - bq = 0$, also $a = bq$ und $b|a$. ∎

Nicht jede Zahl teilt jede andere, aber bekanntlich kann man immer mit Rest teilen: Wir zeigen nun, wie man das aus den Rechenregeln herleiten kann.

12.6 Satz (Division mit Rest) *Gegeben zwei ganze Zahlen $a, b \in \mathbb{Z}$ mit $b > 0$, dann gibt es $q, r \in \mathbb{Z}$ mit*

$$a = bq + r \quad \textit{und} \quad 0 \leqslant r < b.$$

Dabei sind q und r durch a und b eindeutig festgelegt. Insbesondere gilt $r = 0$ genau dann, wenn b ein Teiler von a ist.

Die Zahl q heißt der **Quotient** von a und b und r der **Rest** von a **modulo** b. Wie üblich schreibt man auch

$$a \,:\, b = q \quad \text{Rest } r.$$

Beweis. Existenz. Wir betrachten die Menge

$$S = \big\{x \in \mathbb{N}_0 \mid \exists n \in \mathbb{Z}\colon x = a - bn\big\}.$$

Sie ist nicht leer (für $n = -a^2$ ist $a - bn = a + a^2 b \geqslant 0$), enthält also ein kleinstes Element r, das dann die Form $r = a - bq$ für ein $q \in \mathbb{Z}$ hat. Per Definition gilt $r \geqslant 0$. Wir müssen nur $r < b$ zeigen. Angenommen es wäre $r \geqslant b$, dann würde $r - b = a - b(q+1) \in S$ folgen. Wegen $b > 0$ ist aber $r - b < r$. Das ist ein Widerspruch zur Minimalität von r. Also muss $r < b$ gelten, wie behauptet.

Eindeutigkeit. Gegeben seien zwei Darstellungen

$$a = bq + r \quad \text{und} \quad a = bq' + r'$$

mit $0 \leqslant r < b$ und $0 \leqslant r' < b$ und etwa $r' \geqslant r$. Dann folgt $bq + r = bq' + r'$, also $b(q - q') = r' - r$. Wäre $q - q' \neq 0$, dann würde $r' - r \geqslant b$ folgen. Wegen $r \geqslant 0$ und $r' \geqslant r$ gilt aber $0 \leqslant r' - r \leqslant r' < b$. Also muss die linke Seite 0 sein und damit auch die rechte. Es folgen $q = q'$ und $r = r'$. ∎

12.3 Der euklidische Algorithmus

12.7 Satz *Es seien a und b zwei ganze Zahlen, nicht beide Null. Dann gibt es eine eindeutig bestimmte natürliche Zahl d mit den folgenden beiden Eigenschaften:*

(1) Die Zahl d ist ein gemeinsamer Teiler von a und b, das heißt, es gilt

$$d|a \quad \textit{und} \quad d|b.$$

(2) Jeder gemeinsame Teiler von a und b ist ein Teiler von d, das heißt, es gilt:

$$\forall c \in \mathbb{Z}\colon \big((c|a \;\wedge\; c|b) \Longrightarrow c|d\big).$$

Definition Die Zahl d in Satz 12.7 ist der **größte gemeinsame Teiler** von a und b und wird mit $\mathrm{ggT}(a,b)$ bezeichnet. Die ganzen Zahlen a und b heißen **teilerfremd**, wenn $\mathrm{ggT}(a,b) = 1$ gilt.

12.8 Beispiele Es gelten zum Beispiel $\mathrm{ggT}(6,4) = 2$ und $\mathrm{ggT}(24,18) = 6$. Wir werden gleich sehen, wie man den ggT allgemein bestimmen kann. ◇

Wir verschieben den Beweis von Satz 12.7 für den Moment und diskutieren zuerst noch einige Eigenschaften des größten gemeinsamen Teilers. Per Definition gilt $\mathrm{ggT}(a,b) = \mathrm{ggT}(b,a)$ für alle a, b und, weil der ggT eine natürliche Zahl ist, auch $\mathrm{ggT}(a,b) = \mathrm{ggT}(|a|,|b|)$.

Da 0 von jeder Zahl geteilt wird, gilt für alle $a \in \mathbb{Z}$, $a \neq 0$, immer

$$\mathrm{ggT}(a,0) = \mathrm{ggT}(0,a) = |a|.$$

(Hingegen ist $\mathrm{ggT}(0,0)$ undefiniert). Für alle $a, b, c \in \mathbb{Z}$ (mit $b, c \neq 0$) gilt

$$\mathrm{ggT}(ca,cb) = |c| \cdot \mathrm{ggT}(a,b).$$

Der größte gemeinsame Teiler kann mit dem **euklidischen Algorithmus** berechnet werden, der auf der folgenden Beobachtung beruht.

12.9 Lemma *Es seien $a, b \in \mathbb{Z}$ mit $b > 0$ und es gelte $a : b = q$ Rest r, also*

$$a = bq + r \quad \textit{und} \quad 0 \leqslant r < b.$$

Dann gilt $\mathrm{ggT}(a,b) = \mathrm{ggT}(b,r)$.

Beweis. Wegen $a = bq + r$ ist jeder gemeinsame Teiler von b und r auch ein Teiler von a (nach Satz 12.5(7)). Umgekehrt ist wegen $r = a - bq$ jeder gemeinsame Teiler von a und b auch ein Teiler von r. Beide Zahlenpaare haben also die gleichen gemeinsamen Teiler und damit auch den gleichen ggT. ∎

Für die Berechnung des größten gemeinsamen Teilers zweier natürlicher Zahlen $a, b \in \mathbb{N}$ mit $a \geqslant b$ macht man sich das folgendermaßen zu Nutze.

(1) Teile a durch b und erhalte den Rest r.

(2) Wenn $r = 0$ ist, dann gilt $\mathrm{ggT}(a,b) = b$ und wir sind fertig. Wenn $r \neq 0$ ist, dann ersetze a durch b und b durch r und wiederhole den ersten Schritt.

(3) Erhalte so eine Folge von Resten, bis irgendwann der Rest 0 bleibt. Der letzte von 0 verschiedene Rest ist der größte gemeinsame Teiler von a und b.

12.10 Beispiele 1. Bestimme den größten gemeinsamen Teiler von 1309 und 102.

$$\begin{aligned} 1309 : 102 &= 12 \quad \text{Rest } 85 \\ 102 : 85 &= 1 \quad \text{Rest } 17 \\ 85 : 17 &= 5 \quad \text{Rest } 0. \end{aligned}$$

Dabei ist 17 der letzte von 0 verschiedene Rest. Also gilt $\mathrm{ggT}(1309, 102) = 17$.

2. Bestimme den größten gemeinsamen Teiler von 228 und 87.

$$\begin{aligned} 228 : 87 &= 2 \quad \text{Rest } 54 \\ 87 : 54 &= 1 \quad \text{Rest } 33 \\ 54 : 33 &= 1 \quad \text{Rest } 21 \\ 33 : 21 &= 1 \quad \text{Rest } 12 \\ 21 : 12 &= 1 \quad \text{Rest } 9 \\ 12 : 9 &= 1 \quad \text{Rest } 3 \\ 9 : 3 &= 3 \quad \text{Rest } 0. \end{aligned}$$

Dabei ist 3 der letzte von 0 verschiedene Rest. Also gilt $\operatorname{ggT}(228, 87) = 3$.

12.11 Satz (Euklidischer Algorithmus) *Der größte gemeinsame Teiler zweier natürlicher Zahlen $a, b \in \mathbb{N}$ kann folgendermaßen bestimmt werden: Setze $r_0 = a$ und $r_1 = b$ und definiere für $k \geqslant 1$ die Zahl r_{k+1} als Rest der Division $r_{k-1} : r_k$, also durch die Gleichung*

$$r_{k-1} = r_k q_k + r_{k+1}, \qquad 0 \leqslant r_{k+1} < r_k$$

so lange, bis $r_{k+1} = 0$ gilt. Die ganze Folge der Divisionen mit Rest sieht also so aus:

$$\begin{aligned} a &= bq_1 + r_2 \\ b &= r_2 q_2 + r_3 \\ r_2 &= r_3 q_3 + r_4 \\ &\vdots \\ r_{k-2} &= r_{k-1} q_{k-1} + r_k \\ r_{k-1} &= r_k q_k + \underbrace{r_{k+1}}_{=0}. \end{aligned} \tag{EA}$$

Dann ist r_k der größte gemeinsame Teiler von a und b.

Beweis. Wir bemerken zunächst, dass die Abbruchbedingung $r_{k+1} = 0$ immer eintritt, denn es gilt $b > r_2 > r_3 > \cdots \geqslant 0$ usw. Da dies eine Folge natürlicher Zahlen ist, kann sie nicht unendlich absteigen. Nach endlich vielen Schritten muss also in der Division der Rest 0 bleiben.

Nach Lemma 12.9 gilt nun für die größten gemeinsamen Teiler

$$\begin{aligned} \operatorname{ggT}(a, b) &= \operatorname{ggT}(r_0, r_1) = \cdots = \operatorname{ggT}(r_{k-1}, r_k) = \operatorname{ggT}(r_k, r_{k+1}) \\ &= \operatorname{ggT}(r_k, 0) = r_k. \end{aligned}$$

■

Aus der Berechnung des größten gemeinsamen Teilers mit dem euklidischen Algorithmus folgt nebenbei auch seine Existenz:

Beweis von Satz 12.7. Es seien $a, b \in \mathbb{Z}$, nicht beide Null. Wenn eine der Zahlen Null ist, ist die andere der größte gemeinsame Teiler. Andernfalls können wir, weil Vorzeichen keine Rolle spielen, $a, b \in \mathbb{N}$ annehmen. In dieser Situation ist der euklidische Algorithmus anwendbar und beweist die Existenz des größten gemeinsamen Teilers.

Um die Eindeutigkeit zu zeigen, seien d und d' zwei natürliche Zahlen mit den Eigenschaften (1) und (2) in Satz 12.7. Dann folgt aus (2) sowohl $d|d'$ als auch $d'|d$ und damit $d = \pm d'$. Wegen $d, d' \in \mathbb{N}$ muss $d = d'$ gelten. ∎

12.4 Das Lemma von Bézout

Mit Hilfe des euklidischen Algorithmus können wir noch eine andere wichtige Eigenschaft des größten gemeinsamen Teilers zeigen:

12.12 Lemma (Bézout) *Für $a, b \in \mathbb{Z}$, nicht beide Null, existieren $u, v \in \mathbb{Z}$ mit*

$$\mathrm{ggT}(a, b) = ua + vb.$$

Jede solche Darstellung von $\mathrm{ggT}(a, b)$ heißt eine **Bézout-Identität**.

Beweis. Indem wir die Vorzeichen von a und u bzw. b und v ändern, können wir annehmen, dass a und b natürliche Zahlen sind. Wir wenden den euklidischen Algorithmus an und erhalten eine Folge (EA) von Divisionen mit Rest wie oben. Wir haben dabei

$$r_2 = a - q_1 b \quad \text{und} \quad r_3 = b - r_2 q_2 = b - (a - bq_1)q_2 = (-q_2)a + (1 + q_1 q_2)b$$

Haben wir nun Darstellungen $r_i = u_i a + v_i b$ und $r_{i+1} = u_{i+1} a + v_{i+1} b$ für ein i gefunden, dann bekommen wir daraus eine Darstellung

$$r_{i+2} = r_i - r_{i+1} q_{i+1} = (u_i - u_{i+1} q_{i+1})a + (v_i - v_{i+1} q_{i+1})b$$

von r_{i+2}. Insbesondere bekommen wir nach $k - 1$ Schritten eine Darstellung von $r_k = \mathrm{ggT}(a, b)$. ∎

12.13 Beispiele (1) Man kann eine Bézout-Identität mit Hilfe des Euklidischen Algorithmus berechnen, so wie gerade im Beweis. Für unser voriges Beispiel $17 = \mathrm{ggT}(1309, 102)$ geht das also so:

$$\begin{aligned}
1309 : 102 &= 12 \quad \text{Rest } 85 & 85 &= 1309 - 12 \cdot 102 \\
102 : 85 &= 1 \quad \text{Rest } 17 & 17 &= 102 - 85 = 102 - (1309 - 12 \cdot 102) \\
& & &= -1309 + 13 \cdot 102 \\
85 : 17 &= 5 \quad \text{Rest } 0.
\end{aligned}$$

Eine Bézout-Identität ist also $17 = (-1) \cdot 1309 + 13 \cdot 102$. (Dieses Rechenverfahren lässt sich noch etwas optimieren, worauf wir hier aber verzichten.)

(2) Für $\mathrm{ggT}(228, 87) = 3$ hat man die Bézout-Identität

$$3 = (-8) \cdot 228 + 21 \cdot 87. \qquad \diamond$$

12.14 Korollar *Genau dann sind $a, b \in \mathbb{Z} \setminus \{0\}$ teilerfremd, wenn es $u, v \in \mathbb{Z}$ gibt derart, dass gilt:*

$$ua + vb = 1.$$

Beweis. Wenn a und b teilerfremd sind, dann gilt $\mathrm{ggT}(a, b) = 1$ und die Behauptung folgt aus dem Lemma von Bézout. Gilt umgekehrt $ua + vb = 1$ für $u, v \in \mathbb{Z}$, dann ist jeder gemeinsame Teiler von a und b auch ein Teiler von 1 und damit gleich ± 1. Also sind a und b teilerfremd. ■

12.15 Beispiel Zwei benachbarte natürliche Zahlen n und $n + 1$ sind immer teilerfremd, denn es gilt $1 = 1 \cdot (n + 1) + (-1) \cdot n$. $\diamond$

12.16 Lemma *Es seien $a_1, \ldots, a_k, b \in \mathbb{Z}$. Wenn a_i und b für jedes $i = 1, \ldots, k$ teilerfremd sind, dann sind auch das Produkt $a_1 \cdots a_k$ und b teilerfremd.*

Beweis. Nach dem Lemma von Bézout gibt es $u_1, \ldots, u_k, v_1, \ldots, v_k \in \mathbb{Z}$ mit $1 = u_i a_i + v_i b$ für $i = 1, \ldots, k$. Es folgt

$$u_1 \cdots u_k \cdot a_1 \cdots a_k = (1 - v_1 b) \cdot \cdots \cdot (1 - v_k b).$$

Die rechte Seite hat nach Ausmultiplizieren die Form $1 - vb$ für $v \in \mathbb{Z}$, und es folgt $1 = u_1 \cdots u_k a_1 \cdots a_k + vb$. Also sind $a_1 \cdots a_k$ und b teilerfremd nach Kor. 12.14. ■

12.5 Primzahlen

Definition Eine **Primzahl** ist eine natürliche Zahl $p > 1$, die außer 1 und p keine positiven Teiler besitzt.

12.17 Satz *Es sei $p \in \mathbb{N}$ eine Primzahl und seien $a_1, \dots, a_k \in \mathbb{Z}$. Ist p ein Teiler des Produkts $a_1 \cdots a_k$, dann teilt p eine der Zahlen $a_1, \dots, a_k$.*

Beweis. Das beweisen wir durch Kontraposition. Wir zeigen also: Wenn p keine der Zahlen $a_1, \dots, a_k$ teilt, dann teilt p auch das Produkt $a_1, \dots, a_k$ nicht. Aus $p \nmid a_i$ folgt aber bereits, dass p und a_i teilerfremd sind, weil p eine Primzahl ist. Nach Lemma 12.16 sind dann auch p und $a_1 \cdots a_k$ teilerfremd. Insbesondere ist p kein Teiler von $a_1 \cdots a_k$. ■

12.18 Beispiel Für Zahlen, die nicht prim sind, ist das natürlich völlig falsch. Zum Beispiel sind die Zahlen 6 und 10 beide nicht durch 4 teilbar, aber ihr Produkt $6 \cdot 10 = 60 = 4 \cdot 15$ ist durch 4 teilbar. ◇

12.19 Satz (Primfaktorzerlegung) *Zu jeder natürlichen Zahl $n \in \mathbb{N}$ existieren Primzahlen $p_1, \dots, p_k$ mit*

$$n = p_1 \cdots p_k.$$

Diese Darstellung ist eindeutig bis auf die Reihenfolge: Sind $q_1, \dots, q_l$ Primzahlen mit $n = q_1 \cdots q_l$, dann gilt $k = l$ und nach Umnummerieren $p_1 = q_1, \dots, p_k = q_k$.

Beweis. Wir beweisen als erstes die Existenz einer solchen Darstellung. Dazu zeigen wir durch Induktion nach n, dass jede Zahl kleiner oder gleich n als Produkt von Primzahlen darstellbar ist. Für $n = 1$ nehmen wir $k = 0$, das leere Produkt, das per Definition gleich 1 ist. Angenommen, die Behauptung gilt für $n \in \mathbb{N}$. Falls $n+1$ selbst eine Primzahl ist, dann können wir $k = 1$ und $p_1 = n+1$ nehmen und haben die Behauptung auch für $n+1$ bewiesen. Andernfalls besitzt $n+1$ einen Teiler $m \in \mathbb{N}$ mit $1 < m < n+1$, etwa $n+1 = mq$. Es folgt auch $1 < q < n+1$. Nach Induktionsvoraussetzung gibt es also Primzahlen $p_1, \dots, p_k$ und $p'_1, \dots, p'_l$ mit $m = p_1 \cdots p_k$ und $q = p'_1 \cdots p'_l$. Also ist auch $n+1 = mq = p_1 \cdots p_k p'_1 \cdots p'_l$ ein Produkt von Primzahlen, und die Behauptung ist bewiesen.

Für die Eindeutigkeit zeigen wir die folgende Behauptung: Ist $k \in \mathbb{N}_0$ und sind $p_1, \dots, p_k, q_1, \dots, q_l$ Primzahlen mit $p_1 \cdots p_k = q_1 \cdots q_l$ dann folgen $l = k$ und $p_i = q_i$ für $i = 1, \dots, k$ nach Umnummerieren. Das zeigen wir durch Induktion nach k. Für den Induktionsanfang müssen wir bei $k = 0$ starten. In diesem Fall folgt aus $1 = q_1 \cdots q_l$ mit $q_i \geqslant 2$ auch $l = 0$ und die Behauptung ist richtig. Sei $k \in \mathbb{N}$ beliebig und die Behauptung richtig für $k-1$. Aus $p_k | n$ folgt $p_k | q_1 \cdots q_l$

und damit $p_k | q_i$ für ein $i \in \{1, \dots, l\}$ nach Satz 12.17. Weil q_i prim ist und $p_k > 1$ muss also $p_k = q_i$ gelten. Nach Umnummerieren können wir ohne Einschränkung $i = l$ annehmen. In der Gleichheit $p_1 \dots p_k = q_1 \cdots q_l$ können wir also p_k auf beiden Seiten kürzen und bekommen $p_1 \cdots p_{k-1} = q_1 \cdots q_{l-1}$. Nach Induktionsvoraussetzung gilt nun $k - 1 = l - 1$ und $p_j = q_j$ für alle $j = 1, \dots, k - 1$, nach Umnummerieren. Damit ist die Behauptung bewiesen. ∎

12.20 Satz (Euklid) *Es gibt unendlich viele Primzahlen.*

Beweis. Es seien $p_1, \dots, p_k$ endlich viele Primzahlen. Betrachte die Zahl

$$n = p_1 \cdots p_k.$$

Die Zahlen n und $n + 1$ sind teilerfremd (siehe Beispiel 12.15). Also ist $n + 1$ durch keine der Zahlen $p_1, \dots, p_k$ teilbar. Andererseits ist $n+1$ nach Satz 12.19 als Produkt von Primzahlen darstellbar. Es muss also außer $p_1, \dots, p_k$ noch weitere Primzahlen geben. Da das für jede endliche Menge von Primzahlen gilt, kann die Menge aller Primzahlen insgesamt nicht endlich sein. ∎

12.21 Bemerkung Primzahlen üben seit jeher eine große Faszination aus und ihre Erforschung ist eines der zentralen Themen der *Zahlentheorie*. Sie sind Gegenstand vieler großer Sätze und Vermutungen. Einige Beispiele:

- Die Anzahl der Primzahlen im Intervall $[1, N]$ für $N \in \mathbb{N}$ ist asymptotisch gegeben durch $\frac{N}{\log(N)}$ (*Primzahlsatz*). Niemand weiß genau, wie stark die tatsächliche Anzahl von dieser Näherung abweichen kann (Stichwort: *Riemannsche Vermutung*).
- Es ist nicht bekannt, ob jede gerade Zahl größer als 2 als Summe zweier Primzahlen darstellbar ist (*Goldbachsche Vermutung*).
- Es ist nicht bekannt, ob es unendlich viele sogenannte Primzahlzwillinge gibt. Das sind zwei Primzahlen mit Abstand 2, zum Beispiel $(3, 5)$, $(5, 7)$, $(17, 19)$, $(641, 643)$.
- Die Primzahlen enthalten beliebig lange *arithmetische Progressionen*: Für jedes $k \in \mathbb{N}$ gibt es eine Primzahl p und einen Abstand m derart, dass die Zahlen $p, p + m, p + 2m, \dots, p + km$ allesamt prim sind (*Satz von Green-Tao*, 2004).
- Die größte derzeit bekannte Primzahl ist $2^{82\,589\,933} - 1$ (eine sogenannte *Mersenne-Zahl*), deren Dezimaldarstellung über 24 Millionen Stellen hat (*GIMPS-Projekt* 2018). Zum Vergleich: Nach Schätzungen der Physik ist die Anzahl der Atome im Universum eine Zahl mit deutlich weniger als hundert Stellen.

12.6 Kongruenzrechnung

Als Kongruenzrechnung oder modulare Arithmetik wird das Rechnen mit Resten bei der Division ganzer Zahlen bezeichnet. Vertraut ist das vom Rechnen mit Uhrzeiten her: 6 Stunden nach 21 Uhr ist es nicht 27 Uhr, sondern wieder 3 Uhr — dabei ist 3 der Rest von 27 bei Division durch 24. Es wird *modulo* 24 gerechnet. Entsprechend wird bei den Minuten und Sekunden modulo 60 gerechnet. Es zeigt sich, dass die modulare Arithmetik noch für viele andere Dinge von Nutzen ist, sowohl in der Mathematik, als auch in Anwendungen.

Wir fixieren eine Zahl $n \in \mathbb{N}$, den *Modulus*. Zwei ganze Zahlen $a, b \in \mathbb{Z}$ heißen **kongruent modulo** n, wenn sie bei Division durch n denselben Rest ergeben. Wir schreiben in diesem Fall

$$a \equiv b \pmod{n}.$$

12.22 Beispiel Es gelten zum Beispiel

$$9 \equiv 5 \pmod 4, \quad 14 \equiv 0 \pmod 7, \quad 6 \equiv 72 \equiv 39 \pmod{11}. \qquad \diamond$$

12.23 Lemma *Für $a, b \in \mathbb{Z}$ und $n \in \mathbb{N}$ gilt*

$$a \equiv b \pmod n \quad \Longleftrightarrow \quad n|(a-b).$$

In Worten: Genau dann sind zwei Zahlen kongruent modulo n, wenn ihre Differenz durch n teilbar ist. Insbesondere gilt $a \equiv 0 \pmod n$ genau dann, wenn a durch n teilbar ist.

Beweis. Es gelte $a \equiv b \pmod n$. Per Definition bedeutet das, dass a und b bei Division durch n denselben Rest liefern. Es gibt also $0 \leqslant r < n$ und $q, q' \in \mathbb{Z}$ mit

$$a = qn + r \quad \text{und} \quad b = q'n + r.$$

Daraus folgt $a - b = (q - q')n$, also $n|(a-b)$.

Es gelte umgekehrt $n|(a-b)$. Dann gibt es also $c \in \mathbb{Z}$ mit $a - b = cn$. Ist nun r der Rest von b bei Division durch n, etwa $b = qn + r$, dann folgt $a = b + cn = (q + c)n + r$. Also ist r auch der Rest von a bei Division durch n. ∎

12.24 Satz (Eigenschaften der Kongruenz) *Seien $k, n \in \mathbb{N}$ und $a, a', b, b', c \in \mathbb{Z}$.*

(1) Es gilt $a \equiv a \pmod n$. *(Reflexivität)*

(2) Es gilt $a \equiv b \pmod n$ genau dann, wenn $b \equiv a \pmod n$ gilt. *(Symmetrie)*

(3) Aus $a \equiv b \pmod n$ und $b \equiv c \pmod n$ folgt $a \equiv c \pmod n$. *(Transitivität)*

(4) *Falls* $a \equiv a' \text{ und } b \equiv b' \pmod{n}$, *so* $a + b \equiv a' + b' \pmod{n}$.

(5) *Falls* $a \equiv a' \text{ und } b \equiv b' \pmod{n}$, *so* $ab \equiv a'b' \pmod{n}$.

(6) *Sind* c *und* n *teilerfremd und gilt* $ca \equiv cb \pmod{n}$, *so folgt* $a \equiv b \pmod{n}$.

(7) *Falls* $a \equiv b \pmod{kn}$, *dann auch* $a \equiv b \pmod{n}$.

Beweis. (1) und (2) sind offensichtlich.

(3) Aus $n|(a-b)$ und $n|(b-c)$, etwa $a - b = qn$ und $b - c = q'n$, folgt $a - c = (a - b) + (b - c) = qq'n$.

(4) Aus $a - a' = qn$ und $b - b' = q'n$ folgt $(a + b) - (a' + b') = (q + q')n$, also $a + b \equiv a' + b' \pmod{n}$.

(5) Entsprechend folgt aus der gleichen Voraussetzung

$$ab - a'b' = a\underbrace{(q'n + b')}_{=b} - \underbrace{(a - qn)}_{=a'}b' = aq'n + b'qn + ab' - ab' = (aq' + b'q)n,$$

also $ab \equiv a'b' \pmod{n}$.

(6) Aus $ca - cb = qn$ folgt $n|c(a-b)$. Weil n und c teilerfremd sind, folgt daraus $n|(a-b)$ (denn dies gilt für jeden Primteiler von n). Also gilt $a \equiv b \pmod{n}$.

(7) Aus $kn|(a-b)$ folgt insbesondere $n|(a-b)$. ∎

Definition Es sei $a \in \mathbb{Z}$. Wir schreiben

$$[a]_n = \{b \in \mathbb{Z} \mid a \equiv b \pmod{n}\}.$$

für die Menge aller Zahlen, die zu a kongruent sind, die **Kongruenzklasse von** a **modulo** n.

Wenn zwei Zahlen kongruent sind, dann sind die Kongruenzklassen *gleich*:

$$a \equiv b \pmod{n} \quad \Longleftrightarrow \quad [a]_n = [b]_n.$$

12.25 Beispiel (1) Es sei $n = 4$. Beim Teilen durch 4 können vier verschiedene Reste bleiben, nämlich 0, 1, 2 oder 3. Da der Rest einer Zahl $a \in \mathbb{Z}$ die Kongruenzklasse $[a]_4$ bestimmt, gibt es also genau vier Kongruenzklassen:

$$\begin{aligned}
[0]_4 &= \{\ldots, -16, -12, -8, -4, \mathbf{0}, 4, 8, 12, 16, \ldots\} = 4\mathbb{Z} \\
[1]_4 &= \{\ldots, -15, -11, -7, -3, \mathbf{1}, 5, 9, 13, 17, \ldots\} = 4\mathbb{Z} + 1 \\
[2]_4 &= \{\ldots, -14, -10, -6, -2, \mathbf{2}, 6, 10, 14, 18, \ldots\} = 4\mathbb{Z} + 2 \\
[3]_4 &= \{\ldots, -13, -9, -5, -1, \mathbf{3}, 7, 11, 15, 19, \ldots\} = 4\mathbb{Z} + 3.
\end{aligned}$$

Jede Zahl liegt in genau einer dieser vier Klassen:

$$\mathbb{Z} = \{\dots, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, \dots\}.$$

Außerdem bestimmen kongruente Zahlen dieselbe Kongruenzklasse:

$$[8]_4 = [4]_4 = [0]_4,\ [9]_4 = [5]_4 = [1]_4,\ [11]_4 = [7]_4 = [3]_4 \text{ usw.}$$

(2) Für $n = 2$ bestehen die beiden Kongruenzklassen $[0]_2$ und $[1]_2$ aus allen geraden bzw. allen ungeraden Zahlen. ◇

Das gilt entsprechend für jeden Modulus.

12.26 Lemma *Es sei $n \in \mathbb{N}$. Die Kongruenzklassen modulo n sind*

$$[0]_n, [1]_n, [2]_n, \dots, [n-1]_n$$

und $\mathbb{Z}$ ist die disjunkte Vereinigung dieser Kongruenzklassen.

Beweis. Jede Zahl liegt in genau einer Kongruenzklasse, deshalb ist $\mathbb{Z}$ die disjunkte Vereinigung der Kongruenzklassen (siehe auch Lemma 12.30). Da der Rest beim Teilen durch n immer in $\{0, \dots, n-1\}$ liegt, kann es außer $[0]_n, [1]_n, \dots, [n-1]_n$ keine weiteren Kongruenzklassen geben. Diese sind auch alle verschieden, denn für $a, b \in \{0, 1, \dots, n-1\}$ gilt immer $|a-b| < n$ und damit $n|(a-b) \iff a = b$. ■

Wir schreiben

$$\mathbb{Z}/n = \big\{[0]_n, [1]_n, \dots, [n-1]_n\big\}$$

für die Menge der n verschiedenen Kongruenzklassen modulo n.

12.27 Satz *Die Menge $\mathbb{Z}/n$ wird mit der Addition bzw. Multiplikation*

$$[a]_n + [b]_n = [a+b]_n \quad \text{bzw.} \quad [a]_n \cdot [b]_n = [ab]_n$$

zu einem kommutativen Ring. Die Null ist $[0]_n$ und die Eins ist $[1]_n$.

Beweis. Alle Ringgesetze übertragen sich sofort von $\mathbb{Z}$ auf $\mathbb{Z}/n$. Zum Beispiel ist die Addition kommutativ, denn für alle $a, b \in \mathbb{Z}$ gilt

$$[a]_n + [b]_n = [a+b]_n = [b+a]_n = [b]_n + [a]_n.$$

Es gibt aber noch einen Haken: Wir müssen zeigen, dass die Addition und die Multiplikation überhaupt wohldefinierte Verknüpfungen sind. Das kommt auf Folgendes raus: Gegeben $a, b, a', b' \in \mathbb{Z}$ mit $[a]_n = [a']_n$ und $[b]_n = [b']_n$, dann

haben wir $[a]_n + [b]_n$ einerseits durch $[a+b]_n$ definiert, andererseits aber auch durch $[a'+b']_n$. Es sollte also besser

$$[a+b]_n = [a'+b']_n$$

gelten, sonst ist unsere Definition eine Mogelpackung. Nun bedeutet $[a]_n = [a']_n$ gerade $a \equiv a' \pmod{n}$ und entsprechend $b \equiv b' \pmod{n}$. Nach Satz 12.24(4) gilt dann auch $a + b \equiv a' + b' \pmod{n}$, was gerade $[a+b]_n = [a'+b']_n$ bedeutet. Analog folgt die Wohldefiniertheit der Multiplikation aus Satz 12.24(5). ■

Für jede natürliche Zahl n haben wir damit einen Ring mit n Elementen konstruiert, dessen Elemente die Kongruenzklassen $[0]_n, [1]_n, \dots, [n-1]_n$ sind. Wenn klar ist, dass wir in $\mathbb{Z}/n$ arbeiten, lassen wir die Klammern und den Index n weg.

12.28 Beispiele (1) Es sei $n = 2$. Der Ring $\mathbb{Z}/2$ besteht aus den beiden Elementen $0 = [0]_2$ und $1 = [1]_2$, also nur aus seiner Null und seiner Eins.

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

(2) Für $n = 4$ sehen die Additions- und Multiplikationstabellen so aus:

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

·	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

(3) Die Multiplikationstabelle von $\mathbb{Z}/7$ (ohne die Null) ist

+	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

◇

12.7 Äquivalenzrelationen

Die Kongruenz von ganzen Zahlen ist ein Beispiel für eine Äquivalenzrelation, was wir kurz allgemein diskutieren. Eine Relation ist eine Beziehung, die zwischen verschiedenen Objekten besteht. Man schreibt dann etwa

$$x \sim y$$

wenn die Objekte x und y in der Relation $\sim$ stehen. Das Konzept ist so allgemein, dass man viele alltägliche Beispiele dafür geben kann. Auf der Menge aller Menschen sind etwa Verwandtschaft oder Freundschaft Relationen. Formal definiert ist eine Relation einfach irgendeine Menge von Paaren, also eine Teilmenge des kartesischen Produkts.

Definition Es sei X eine Menge. Eine **Äquivalenzrelation auf** X ist eine Relation $\sim$, die *reflexiv*, *symmetrisch* und *transitiv* ist. Das bedeutet Folgendes:

(1) Für alle $x \in X$ gilt $x \sim x$. (*Reflexivität*)

(2) Für alle $x, y \in X$ gilt: $x \sim y \iff y \sim x$. (*Symmetrie*)

(3) Für alle $x, y, z \in X$ gilt: $\big(x \sim y \wedge y \sim z\big) \implies x \sim z$. (*Transitivität*)

Ist $\sim$ eine Äquivalenzrelation auf der Menge X und $x \in X$ ein Element, dann heißt

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

die **Äquivalenzklasse von** x bezüglich $\sim$.

12.29 Beispiele (1) Sei X die Menge der Spielfiguren in »Mensch ärgere dich nicht«. Es gibt vier Farben, rot, blau, grün und gelb, und in jeder Farbe vier Spielfiguren. Die Menge X hat also 16 Elemente. Die Relation

$$x \sim y \quad \Leftrightarrow \quad x \text{ und } y \text{ haben dieselbe Farbe}$$

ist eine Äquivalenzrelation auf X. Die Äquivalenzklasse $[x]$ einer Spielfigur x besteht aus vier Elementen, nämlich den Spielfiguren derselben Farbe.

(2) Für jede natürliche Zahl n ist die Kongruenz modulo n eine Äquivalenzrelation auf $\mathbb{Z}$. Die Reflexivität, Symmetrie und Transitivität haben wir in Satz 12.24 nachgeprüft. Die Äquivalenzklassen sind die Kongruenzklassen. ◇

In beiden Beispielen sieht man das Typische einer Äquivalenzrelation: Äquivalente Elemente müssen nicht gleich sein, sie sind nur gleich hinsichtlich einer

bestimmten Eigenschaft: im ersten Beispiel die Farbe, im zweiten der Rest bei Division durch n. Eine Äquivalenzrelation ist sozusagen eine vergröberte Gleichheit, die nur auf gewisse Eigenschaften konzentriert ist. Sie unterteilt die gegebene Menge dabei in die Äquivalenzklassen. Genauer gilt Folgendes:

12.30 Satz *Eine Menge X mit einer Äquivalenzrelation $\sim$ ist die disjunkte Vereinigung aller Äquivalenzklassen, das heißt:*

(1) Es gilt $X = \bigcup_{x \in X} [x]$.

(2) Für alle $x, y \in X$ gilt: Falls $[x] \neq [y]$, so $[x] \cap [y] = \emptyset$.

Beweis. (1) Das ist klar, denn jedes Element $x \in X$ ist in einer Äquivalenzklasse enthalten, nämlich in $[x]$. (2) Das zeigen wir durch Kontraposition: Seien $x, y \in X$ mit $[x] \cap [y] \neq \emptyset$, dann müssen wir $[x] = [y]$ zeigen. Nach Voraussetzung gibt es also ein Element $z \in [x] \cap [y]$. Per Definition gelten dann $x \sim z$ und $y \sim z$ und damit wegen Symmetrie auch $z \sim y$. Aus der Transitivität folgt dann $x \sim y$. Sei nun $w \in [x]$ beliebig, das heißt es gelte $x \sim w$. Wegen Transitivität und Symmetrie können wir daraus $y \sim w$ folgern, also $w \in [y]$. Damit haben wir $[x] \subset [y]$ bewiesen. Mit genau dem gleichen Argument, mit x und y vertauscht, können wir auch $[y] \subset [x]$ zeigen. Es gilt also $[x] = [y]$. ∎

12.31 Bemerkung Eine Äquivalenzrelation auf einer Menge X ist grundsätzlich nichts anderes als eine Unterteilung von X in disjunkte Teilmengen. Denn ist $(Y_i)_{i \in I}$ eine Familie von Teilmengen mit $X = \bigcup_{i \in I} Y_i$ und $Y_i \cap Y_j = \emptyset$ für $i \neq j$, dann ist also jedes $x \in X$ in genau einer der Mengen Y_i enthalten. Wir können dann eine Äquivalenzrelation auf X definieren, in der $x \sim y$ genau dann gilt, wenn x und y in derselben Teilmenge Y_i enthalten sind. Die Äquivalenzklassen dieser Relation sind genau die Y_i.

13 Körper

13.1 Körperaxiome

Definition Ein **Körper** ist ein kommutativer Ring K, nicht der Nullring, mit der folgenden zusätzlichen Eigenschaft:

(MI) Für jedes $x \in K$ mit $x \neq 0$ gibt es ein Element $x^{-1} \in K$ mit $x \cdot x^{-1} = 1$.
(*Existenz von multiplikativen Inversen*)

Eine äquivalente Definition ist: Ein Körper ist eine nicht-leere Menge K mit zwei Verknüpfungen + und · mit folgenden Eigenschaften:

(1) $(K, +)$ ist eine abelsche Gruppe mit neutralem Element 0_K.

(2) $(K \setminus \{0_K\}, \cdot)$ ist eine abelsche Gruppe mit neutralem Element 1_K.

(3) Für alle $x, y, z \in K$ gelten die Distributivgesetze

$$(x + y)z = xz + yz \quad \text{und} \quad x(y + z) = xy + xz.$$

13.1 Beispiele Die kommutativen Ringe $\mathbb{Q}$ und $\mathbb{R}$ sind Körper. Dagegen ist $\mathbb{Z}$ zwar ein kommutativer Ring aber kein Körper, weil (MI) nicht erfüllt ist. Weitere Beispiele folgen in den nächsten beiden Abschnitten. ◇

In Körpern gibt es keine Nullteiler außer 0, das heißt, ist ein Produkt 0, dann ist mindestens einer der Faktoren 0: In jedem Körper K gilt also die Implikation

$$\forall a, b \in K\colon \big(ab = 0 \implies (a = 0 \vee b = 0)\big).$$

Denn ist etwa $a \neq 0$, dann können wir die Gleichung $ab = 0$ mit a^{-1} multiplizieren und $b = 0$ folgern. Außerdem darf man in Körpern immer **kürzen**:

$$\forall a, b, c \in K\colon \big((ac = bc \wedge c \neq 0) \implies a = b\big).$$

Noch eine Bemerkung zur Notation: Die Schreibweise als Bruch verwendet man nicht nur für rationale Zahlen, sondern auch ganz allgemein für das multiplikative Inverse in einem Körper. Ist $x \in K \setminus \{0\}$, dann gilt immer

$$x^{-1} = \frac{1}{x},$$

es gibt zwischen den beiden Schreibweisen keinen Unterschied. Man darf sie auch auch mischen oder Brüche von Brüchen bilden, zum Beispiel

$$\frac{1}{\frac{1}{x}} = \frac{1}{x^{-1}} = x, \qquad \frac{\frac{x}{y}}{\frac{z}{w}} = \frac{x}{y} \cdot \frac{w}{z} = \frac{xw}{yz}, \qquad \frac{\frac{x}{y}}{\frac{z}{y}} = \frac{x}{z}, \qquad \text{usw.}$$

13.2 Rationale und reelle Zahlen

Eine rationale Zahl ist repräsentiert durch einen Bruch $\frac{a}{b}$ ganzer Zahlen mit $b \neq 0$. Dabei können verschiedene Brüchen dieselbe Zahl repräsentieren: Es gilt

$$\frac{a}{b} = \frac{a'}{b'} \quad \Leftrightarrow \quad ab' = a'b.$$

Das ist ein Beispiel für eine Äquivalenzrelation auf der Menge der Paare $\mathbb{Z}\times\mathbb{Z}\setminus\{0\}$ gegeben durch $(a, b) \sim (a', b)' \Leftrightarrow ab' = a'b$. Die rationalen Zahlen entsprechen den Äquivalenzklassen (Übung).

13.2 Satz *Jede rationale Zahl $x \in \mathbb{Q}$ besitzt eine eindeutige Darstellung*

$$x = \frac{a}{b}$$

mit $a \in \mathbb{Z}$, $b \in \mathbb{N}$ und $\mathrm{ggT}(a, b) = 1$.

Beweis. Dies folgt aus der eindeutigen Primfaktorzerlegung für Zähler und Nenner (Satz 12.19); Übung. ∎

Ebenfalls aus der Schule bekannt sind die **reellen Zahlen** $\mathbb{R}$, mit denen wir schon die ganze Zeit arbeiten. Es ist aber nicht ganz einfach, mathematisch völlig exakt zu sagen, was eine reelle Zahl überhaupt ist.[1] Die spezifischen Eigenschaften der reellen Zahlen sind von überragender Bedeutung für die Analysis und werden in der Analysis-Vorlesung ausführlich besprochen.

Jede rationale Zahl ist auch eine reelle Zahl. Hinzu kommen aber die *irrationalen Zahlen*, die nicht durch Brüche dargestellt werden können. Die Notwendigkeit für solche Zahlen ergibt sich zum Beispiel aus der folgenden berühmten Aussage:

13.3 Satz (Irrationalität von Quadratwurzeln) *Ist p eine Primzahl, dann gibt es keine rationale Zahl x mit $x^2 = p$.*

[1] Es ist fast, aber nicht ganz korrekt, zu sagen, dass eine reelle Zahl ein unendlicher Dezimalbruch ist. Die beiden Dezimalbrüche $1, \overline{0}$ und $0, \overline{9}$ (Periode) repräsentieren zum Beispiel dieselbe reelle Zahl. Deshalb entsprechen auch reelle Zahlen Äquivalenzklassen von Dezimalbrüchen.

Beweis. Angenommen, es gibt doch eine rationale Zahl x mit $x^2 = p$. Dann gibt es also ganze Zahlen a, b mit $b \neq 0$ und $x = \frac{a}{b}$. Wir können annehmen, dass a und b teilerfremd sind, nach Satz 13.2. Insbesondere sind a und b dann nicht beide durch p teilbar. Aus der Gleichung

$$x^2 = \frac{a^2}{b^2} = p$$

folgt nun $a^2 = pb^2$. Also ist a^2 durch p teilbar und damit auch a selbst: Denn p ist eine Primzahl und teilt das Produkt $a^2 = a \cdot a$. Nach Satz 12.17 teilt p damit auch a. Es gibt also eine ganze Zahl a' mit $a = pa'$. Daraus folgt $pb^2 = a^2 = (pa')^2 = p^2a'^2$ und somit $b^2 = pa'^2$. Es folgt, dass auch b durch p teilbar ist, im Widerspruch zur Annahme, dass a und b teilerfremd sind. Dieser Widerspruch zeigt, dass die ursprüngliche Annahme falsch gewesen sein muss. Es gibt also keine solche rationale Zahl x. ∎

Im Unterschied zu den rationalen Zahlen, kann man in den reellen Zahlen Quadratwurzeln aus jeder nicht-negativen Zahl ziehen.

13.4 Satz *Zu jeder reellen Zahl $a \geqslant 0$ existiert eine reelle Zahl $\sqrt{a} \geqslant 0$ mit $\sqrt{a}^2 = a$.*

Der Beweis dieser Aussage gehört in die Analysis-Vorlesung.[2] Von Satz 13.4 gilt bekanntlich auch die Umkehrung: Das Quadrat einer reellen Zahl ist niemals negativ (denn es gilt $(-1)^2 = 1$).

Das Lösen allgemeiner quadratischer Gleichungen lässt sich mit der bekannten Lösungsformel auf das Quadratwurzelziehen reduzieren.

13.5 Satz (Quadratische Lösungsformel) *Es seien $a, b, c \in \mathbb{R}$ mit $a \neq 0$ und sei $\Delta = b^2 - 4ac$. Ist $\Delta \geqslant 0$, dann hat die quadratische Gleichung $ax^2 + bx + c = 0$ in der Unbekannten x die Lösungen $\frac{-b \pm \sqrt{\Delta}}{2a}$. Ist $\Delta < 0$, dann gibt es keine reelle Lösung.*

Beweis. Wir setzen $p = \frac{b}{2a}$ und $q = \frac{c}{a}$ und teilen die Gleichung $ax^2 + bx + c = 0$ durch a. Sie wird damit zu $x^2 + 2px + q = 0$, was wir durch quadratische Ergänzung zu

$$x^2 + 2px + q = (x + p)^2 - p^2 + q = 0$$

umformen. Die Gleichung $(x+p)^2 = p^2 - q$ hat die beiden Lösungen $-p \pm \sqrt{p^2 - q}$. Einsetzen ergibt die obigen Ausdrücke in a, b, c. ∎

[2] Das schriftliche Wurzelziehen, also die näherungsweise Berechnung einer Dezimalbruchdarstellung, war früher eine wichtige Rechentechnik, die aber schon seit Längerem dem Computer oder Taschenrechner überlassen wird und deshalb auch nicht mehr im Schulunterricht vorkommt. Das allgemeine Prinzip dahinter ist das *Newton-Verfahren*, ein allgemeines Näherungsverfahren zur Bestimmung von Nullstellen, das ebenfalls in der Analysis-Vorlesung behandelt wird.

13.3 Komplexe Zahlen

Der Körper der komplexen Zahlen entsteht dadurch, dass den reellen Zahlen ein zusätzliches Element i hinzugefügt wird, die *imaginäre Einheit*, die der Gleichung

$$i^2 = -1$$

genügt. Es ist klar, dass i keine reelle Zahl sein kann, denn das Quadrat einer reellen Zahl ist immer größer oder gleich 0. Die Konstruktion der komplexen Zahlen führt deshalb aus den reellen Zahlen hinaus.

Wir konstruieren die komplexen Zahlen folgendermaßen: In der reellen Ebene $\mathbb{R}^2 = \{(a, b) \mid a, b \in \mathbb{R}\}$ haben wir die Vektoraddition

$$(a, b) + (a', b') = (a + a', b + b')$$

für $a, a', b, b' \in \mathbb{R}$. Zusätzlich definieren wir eine Multiplikation, allerdings nicht komponentenweise, sondern durch

$$(a, b) \cdot (a', b') = (aa' - bb', ab' + a'b).$$

Diese Definition hat die folgenden erfreulichen Eigenschaften:

(1) Wenn wir uns auf Paare der Form $(a, 0)$ beschränken, dann gelten einfach

$$(a, 0) + (a', 0) = (a + a', 0) \quad \text{und} \quad (a, 0) \cdot (a', 0) = (aa', 0).$$

Solange der zweite Eintrag 0 ist, finden wir also die reellen Zahlen mit der normalen Addition und Multiplikation im ersten Eintrag wieder.

(2) Das Element

$$i = (0, 1)$$

hat die gewünschte Eigenschaft

$$i^2 = (0, 1) \cdot (0, 1) = (-1, 0),$$

denn das Paar $(-1, 0)$ entspricht ja der reellen Zahl -1.

Jetzt müssen wir nur die sperrige Notation mit den Paaren wieder loswerden. Wir schreiben einfach weiter a statt $(a, 0)$ und haben damit

$$a + bi = (a, 0) + (b, 0)(0, 1) = (a, 0) + (0, b) = (a, b)$$

was die übliche Schreibweise für komplexe Zahlen darstellt. Per Definition gilt für die Multiplikation

$$(a+bi)(a'+b'i) = (aa'-bb') + (ab'+a'b)i.$$

Definition Für die Menge $\mathbb{R}^2$ mit der angegebenen Addition und Multiplikation schreiben wir $\mathbb{C}$ und nennen ihre Elemente **komplexe Zahlen**. Für eine komplexe Zahl $z = a + bi$ heißt a der **Realteil** von z und b der **Imaginärteil**, kurz

$$a = \mathrm{Re}(z) \quad \text{und} \quad b = \mathrm{Im}(z).$$

Außerdem heißt

$$\overline{z} = a - bi$$

die **komplex-konjugierte** Zahl zu z.

13.6 Satz *Die komplexen Zahlen bilden mit der gerade definierten Addition und Multiplikation einen Körper.*

Beweis. Seien im folgenden

$$z = a + bi, \;\; z' = a' + b'i, \;\; z'' = a'' + b''i, \;\; w = c + di$$

komplexe Zahlen. Da bei der Addition die erste und die zweite Komponente völlig getrennt bleiben, übertragen sich die Rechengesetze hier einfach von $\mathbb{R}$. Es gelten also die Assoziativität, außerdem $-z = -(a+bi) = -a - bi$ und $z + 0 = 0 + z = z$.

Die Multiplikation ist kommutativ:

$$\begin{aligned} zz' &= (a+bi)(a'+b'i) = (aa'-bb') + (a'b+ab')i \\ &= (a'a-b'b) + (ab'+a'b)i = (a'+b'i)(a+bi) = z'z. \end{aligned}$$

Es gilt außerdem

$$1 \cdot (a+bi) = (1+0i)(a+bi) = a+bi = (a+bi)\cdot 1.$$

Für das Distributivgesetz berechnen wir

$$\begin{aligned} (z+z')w &= \big((a+bi)+(a'+b'i)\big)(c+di) \\ &= \big(a+a'+(b+b')i\big)(c+di) \\ &= (a+a')c-(b+b')d+\big((a+a')d+(b+b')c\big)i \\ &= ac+a'c-bd-b'd+(ad+a'd+bc+b'c)i \end{aligned}$$

$$\begin{aligned}
&= ac - bd + (ad + bc)i + a'c - b'd + (a'd + b'c)i \\
&= (a + bi)(c + di) + (a' + b'i)(c + di) \\
&= zw + z'w.
\end{aligned}$$

Die Assoziativität der Multiplikation ergibt sich aus

$$\begin{aligned}
(zz')z'' &= \big((a + bi)(a' + b'i)\big)(a'' + b''i) = (aa' - bb' + (ab' + a'b)i)(a'' + b''i) \\
&= (aa' - bb')a'' - (ab' + a'b)b'' + \big((aa' - bb')b'' + (ab' + a'b)a''\big)i \\
&= aa'a'' - bb'a'' - ab'b'' - a'bb'' + (aa'b'' - bb'b'' + aa''b' + a'a''b)i \\
&= (a(a'a'' - b'b'') - b(a'b'' + a''b')) + \big(a(a''b') + b(a'a'' - b'b'')\big)i \\
&= (a + bi)(a'a'' - b'b'' + (a'b'' + a''b')i) \\
&= z(z'z'').
\end{aligned}$$

Ist schließlich $z \neq 0$, dann ist

$$z\overline{z} = (a + bi)(a - bi) = a^2 - b^2i^2 = a^2 + b^2 > 0$$

eine reelle Zahl ungleich 0. Daraus folgt

$$z \cdot \frac{\overline{z}}{a^2 + b^2} = 1.$$

Also haben wir das multiplikative Inverse von z gefunden, nämlich

$$z^{-1} = \frac{\overline{z}}{a^2 + b^2} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i.$$

Damit ist alles bewiesen. ■

13.7 Beispiel An den Umgang mit den Formeln

$$(a + bi)(a' + b'i) = (aa' - bb') + (ab' + a'b)i$$
$$\frac{1}{a + bi} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i$$

muss man sich etwas gewöhnen. Es gelten zum Beispiel

$$\frac{1}{i} = -i \quad \text{und} \quad \frac{1}{1 + i} = \frac{1}{2} - \frac{1}{2}i = \frac{1}{2}(1 - i).$$ ◇

In den komplexen Zahlen hat jede Gleichung

$$x^2 + px + q = 0$$

mit $p, q \in \mathbb{R}$ die Lösungen

$$-\frac{p}{2} \pm \frac{\sqrt{\Delta}}{2} \quad \text{wobei } \Delta = p^2 - 4q.$$

Ist $\Delta > 0$, dann ist $\sqrt{\Delta}$ eine positive reelle Zahl und es gibt zwei verschiedene reelle Lösungen der Gleichung, entsprechend den beiden Vorzeichen vor der Wurzel. Für $\Delta = 0$ gibt es nur die eine Lösung $-\frac{p}{2}$. Ist aber $\Delta < 0$, dann hat Δ keine reelle Quadratwurzel. Dafür gilt dann $-\Delta > 0$ und es gilt $(\sqrt{-\Delta} \cdot i)^2 = -\Delta \cdot (-1) = \Delta$. Wir erhalten so die beiden komplexen Lösungen

$$-\frac{p}{2} \pm \frac{\sqrt{-\Delta}}{2} i.$$

Man beachte, dass die beiden Lösungen zu einander komplex-konjugiert sind.

Allgemeiner haben in den komplexen Zahlen auch quadratische Gleichungen mit komplexen Koeffizienten und sogar algebraische Gleichungen von beliebigem positivem Grad eine Lösung (nach dem *Fundamentalsatz der Algebra*). Darauf kommen wir später zurück.

13.8 Bemerkung Die komplexen Zahlen wurden etwa ab dem 16. Jahrhundert aus bestimmten Rechenproblemen heraus entwickelt. Es hat aber lange gedauert, bis sie zu einem festen Bestandteil der Mathematik wurden. Man war der Meinung, dass die imaginäre Einheit i nicht »wirklich existiert«. (Schon der Name deutet darauf hin.) Für die moderne Mathematik sind solche rein formalen Konstruktionen typisch, und komplexe Zahlen kommen in der Mathematik und ihren Anwendungen überall vor.

13.4 Endliche Körper

Für die lineare Algebra ist die Kongruenzrechnung modulo einer Primzahl besonders wichtig:

13.9 Satz *Für jede Primzahl p ist $\mathbb{Z}/p$ ein Körper.*

Für diesen Körper schreibt man statt $\mathbb{Z}/p$ auch $\mathbb{F}_p$.

Erster Beweis. Wir müssen zeigen, dass jedes Element ungleich Null in $\mathbb{F}_p$ ein multiplikatives Inverses besitzt. Ein solches Element ist eine Kongruenzklasse $[a]_p \neq$

$[0]_p$, also gegeben durch eine Zahl $a \in \mathbb{Z}$, die nicht durch p teilbar ist. Da p eine Primzahl ist, sind a und p dann teilerfremd. Nach dem Lemma von Bézout (Kor. 12.14) gibt es deshalb $u, v \in \mathbb{Z}$ mit

$$ua + vp = 1.$$

Nun ist vp durch p teilbar, also $vp \equiv 0 \pmod{p}$, und es folgt

$$ua \equiv ua + vp \equiv 1 \pmod{p},$$

also $[u]_p \cdot [a]_p = [1]_p$ in $\mathbb{F}_p$. Damit ist $[u]_p = [a]_p^{-1}$ das multiplikative Inverse von $[a]_p$ in $\mathbb{F}_p$. ∎

Zweiter Beweis. Man kann die Aussage auch ohne den euklidischen Algorithmus beweisen. Es sei wieder $a \in \mathbb{Z}$ mit $[a]_p \neq [0]_p$ in $\mathbb{F}_p$, also $p \nmid a$. Wir betrachten die Abbildung

$$\varphi_a \colon \mathbb{F}_p \to \mathbb{F}_p,\ [x]_p \mapsto [a_p] \cdot [x]_p = [ax]_p.$$

Die Abbildung φ_a ist injektiv: Denn sind $x, y \in \mathbb{Z}$ mit $[ax]_p = [ay_p]$, dann ist $a(x-y)$ durch p teilbar. Da p eine Primzahl mit $p \nmid a$ ist, muss $p|(x-y)$ gelten und damit $[x]_p = [y]_p$. Da φ_a eine injektive Abbildung der endlichen Menge $\mathbb{F}_p$ in sich ist, ist φ_a dann auch surjektiv. Insbesondere gibt es $[b]_p \in \mathbb{F}_p$ mit $\varphi_a([b_p]) = [1]_p$, was gerade $[a_p] \cdot [b_p] = [1]_p$ und damit $[b]_p = [a]_p^{-1}$ bedeutet. ∎

Wir haben damit bewiesen, dass für jede Primzahl p ein **endlicher Körper** mit p Elementen existiert. Das Rechnen in den Körpern $\mathbb{F}_p$ ist etwas ungewohnt. Das additive Inverse einer Kongruenzklasse $[a]_p \in \mathbb{F}_p$ ist $-[a]_p = [-a]_p = [n-a]_p$ und damit leicht zu bestimmen. Das multiplikative Inverse $[a]_p^{-1}$ lässt sich durch eine Bézout-Identität und damit durch den euklidischen Algorithmus ermitteln. Da es nur endlich viele Elemente gibt, kann man die Lösung aber auch durch Ausprobieren finden. Im Körper $\mathbb{F}_{11}$ gilt zum Beispiel:

a	0	1	2	3	4	5	6	7	8	9	10
$-a$	0	10	9	8	7	6	5	4	3	2	1
a^{-1}	–	1	6	4	3	9	2	8	7	5	10

13.10 Bemerkungen (1) Es gilt auch die Umkehrung von Satz 13.9: Wenn $n \in \mathbb{N}$ keine Primzahl ist, dann ist $\mathbb{Z}/n$ kein Körper (siehe Übungen).

(2) Später (Algebra I) wird bewiesen, dass zu jeder Primzahlpotenz p^k genau ein Körper mit p^k Elementen existiert und es sonst keine weiteren endlichen Körper gibt.

13.5 Lineare Algebra über Körpern

Ist K ein Körper, dann betrachten wir K^n als den Raum der Spaltenvektoren mit Einträgen aus K und können darin genauso rechnen wie in $\mathbb{R}^n$.

Der Bezug zur Geometrie geht dabei leider verloren. Dafür gelten aber die meisten Aussagen, die wir in den Kapiteln I und II über den reellen Zahlen bewiesen haben, auch weiterhin: Die Vektor- und Matrizenrechnung benutzt nichts weiter als die Körperaxiome. Deshalb gelten die Aussagen wortwörtlich weiter, wenn wir überall $\mathbb{R}$ durch K ersetzen, mit identischen Beweisen.

Wenn wir alles noch einmal durchgehen, stoßen wir aber auf eine wesentliche Ausnahme: Die Norm eines Vektors $\|x\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}$ ist in $\mathbb{R}^n$ nur deshalb sinnvoll definiert, weil $\langle \vec{x}, \vec{x} \rangle \geqslant 0$ für alle $\vec{x} \in \mathbb{R}^n$ gilt und jede positive reelle Zahl eine Quadratwurzel besitzt. Beides ist in allgemeinen Körpern nicht richtig.[3]

Zeit für eine Zwischenbilanz. Hier ist eine Übersicht über die Konzepte der linearen Algebra, die wir bisher eingeführt haben.

		Bemerkungen
1.	Lineare Gleichunggsysteme	
	Zeilenstufenform	
	Eliminationsverfahren	
2.	Vektoren (Addition und Skalarmultiplikation)	
	Geraden in K^n	
3.	Lösungsräume linearer Gleichungssysteme	
4.	Lineare Unterräume in K^n	
	Affine Unterräume	
7.	Lineare Unabhängigkeit	
	Basen und Dimension linearer Unterräume	
8.	Norm und Skalarprodukt	nur für $K = \mathbb{R}$
	Orthonormalbasen	nur für $K = \mathbb{R}$
9.	Lineare Abbildungen	
	Matrizenprodukt	
	Dimensionsformel für lineare Abbildungen	
	Orthogonale Abbildungen	nur für $K = \mathbb{R}$
10.	Rang einer Matrix	
	Zeilenrang = Spaltenrang	
	Invertierbare Matrizen	

[3] Über den komplexen Zahlen wäre die Norm sinnvoll definiert, hat aber nicht die gewünschten Eigenschaften. Darauf gehen wir später noch ein.

Lineare Gleichungssysteme lassen sich wie über $\mathbb{R}$ mit dem Gauß-Algorithmus lösen. Dazu noch ein paar Beispiele.

13.11 Beispiele (1) Gegeben sei das lineare Gleichungssystem

$$\begin{aligned}(1+i)x_1 + 2x_2 - (2-10i)x_3 &= 1+3i\\ \frac{1}{2}x_1 + x_2 + (2+4i)x_3 &= 2+\frac{3}{2}i\\ x_1 + x_2 + (4+7i)x_3 &= 2+4i\end{aligned}$$

über den komplexen Zahlen. Wir bringen die erweiterte Koeffizientenmatrix in Zeilenstufenform:

$$\left(\begin{array}{ccc|c} 1+i & 2 & -2+10i & 1+3i \\ \frac{1}{2} & 1 & 2+4i & 2+\frac{3}{2}i \\ 1 & 1 & 4+7i & 2+4i \end{array}\right) \mid \cdot \tfrac{1}{1+i} \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1-i & 4+6i & 2+i \\ \frac{1}{2} & 1 & 2+4i & 2+\frac{3}{2}i \\ 1 & 1 & 4+7i & 2+4i \end{array}\right) \begin{array}{l}\cdot(-\frac{1}{2})\\ \longleftarrow +\end{array}$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1-i & 4+6i & 2+i \\ 0 & \frac{1}{2}+\frac{1}{2}i & i & 1+i \\ 1 & 1 & 4+7i & 2+4i \end{array}\right) \begin{array}{l}\cdot(-1)\\ \longleftarrow +\end{array} \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1-i & 4+6i & 2+i \\ 0 & \frac{1}{2}+\frac{1}{2}i & i & 1+i \\ 0 & i & i & 3i \end{array}\right) \mid \cdot \tfrac{1}{\frac{1}{2}+\frac{1}{2}i}$$

$$\rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1-i & 4+6i & 2+i \\ 0 & 1 & 1+i & 2 \\ 0 & i & i & 3i \end{array}\right) \begin{array}{l}\cdot(-i)\\ \longleftarrow +\end{array} \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1-i & 4+6i & 2+i \\ 0 & 1 & 1+i & 2 \\ 0 & 0 & 1 & i \end{array}\right)$$

(Bei den beiden Divisionen haben wir $(1-i)^{-1} = 2\cdot(1+i)$ benutzt.) Daraus erhalten wir durch Rückeinsetzen die Lösungen

$$x_1 = 6+i,\quad x_2 = 3-i,\quad x_3 = i.$$

(2) Besonders einfach sind Gleichungssysteme über $\mathbb{F}_2$. Weil es nur 1 und 0 gibt, entfällt das Multiplizieren von Zeilen mit Skalaren.

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{array}\right) \longleftarrow + \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right) \longleftarrow + \rightsquigarrow \left(\begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

Hier ist das zugehörige System also unlösbar.

(3) Wir betrachten das homogene lineare Gleichungssystem

$$\begin{aligned}x_1 + 3x_2 + 3x_3 &= 0\\ 2x_1 + x_2 + 4x_3 &= 0\\ 3x_1 + 5x_2 + 6x_3 &= 0\end{aligned}$$

über dem Körper $\mathbb{F}_7$ mit sieben Elementen.

$$\begin{pmatrix}1&3&3\\2&1&4\\3&5&6\end{pmatrix} \overset{\cdot(-2)}{\underset{+}{\longleftarrow}} \rightsquigarrow \begin{pmatrix}1&3&3\\0&2&5\\3&5&6\end{pmatrix} \overset{\cdot(-3)}{\underset{+}{\longleftarrow}} \rightsquigarrow \begin{pmatrix}1&3&3\\0&2&5\\0&3&4\end{pmatrix} \mid \cdot(2^{-1}=4)$$
$$\rightsquigarrow \begin{pmatrix}1&3&3\\0&1&6\\0&3&4\end{pmatrix} \mid \cdot(3^{-1}=5) \rightsquigarrow \begin{pmatrix}1&3&3\\0&1&6\\0&1&6\end{pmatrix} \overset{\cdot(-1)}{\underset{+}{\longleftarrow}} \rightsquigarrow \begin{pmatrix}1&3&3\\0&1&6\\0&0&0\end{pmatrix}$$

In den Lösungen kann also x_3 frei gewählt werden. Rückeinsetzen ergibt $x_2 = -6x_3 = x_3$ und $x_1 = -3x_2 - 3x_2 = -6x_3 = x_3$. Jede Wahl von x_3 in $\mathbb{F}_7$ ergibt eine Lösung des Gleichungssystems, das also insgesamt 7 Lösungen hat, nämlich

$$(x_1, x_2, x_3) \in \{(0,0,0), (1,1,1), (2,2,2), (3,3,3), (4,4,4), (5,5,5), (6,6,6)\}. \quad \diamond$$

14 Ordnungsrelationen

Die Körper $\mathbb{Q}$ und $\mathbb{R}$ besitzen eine Anordnung, das heißt, die Zahlen lassen sich der Größe nach vergleichen. Das ist für allgemeine Körper und etwa für komplexe Zahlen nicht der Fall. In diesem Kontext führen wir noch einige allgemeine Begriffe ein, die in der Mathematik immer wieder auftreten.

14.1 Partielle und lineare Ordnungen

Neben den Äquivalenzrelationen kommen Ordnungsrelationen sehr häufig vor, in denen die Elemente einer Menge in irgendeiner Weise sortiert werden.

Definition Es sei X eine Menge. Eine **partielle Ordnung auf** X ist eine Relation $\leqslant$ auf X, die *reflexiv*, *antisymmetrisch* und *transitiv* ist. Das bedeutet im Einzelnen folgendes:

(1) Für alle $x \in X$ gilt $x \leqslant x$. (*Reflexivität*)

(2) Für alle $x, y \in X$ gilt: $\big(x \leqslant y \ \wedge \ y \leqslant x\big) \implies x = y$. (*Antisymmetrie*)

(3) Für alle $x, y, z \in X$ gilt: $\big(x \leqslant y \wedge y \leqslant z\big) \implies x \leqslant z$. (*Transitivität*)

Eine partielle Ordnung $\leqslant$ auf einer Menge X wird **Ordnung** genannt, wenn sie zusätzlich die folgende Eigenschaft besitzt:

(4) Für alle $x, y \in X$ gilt: $x \leqslant y$ oder $y \leqslant x$. (*Totale Vergleichbarkeit*)

Zur besseren Unterscheidbarkeit von partiellen Ordnungen nennt man die Ordnungen auch **totale** oder **lineare** Ordnungen (weil man sich die Elemente wie auf einer Geraden aufgereiht denken kann).

Wie üblich schreibt man auch

$$x < y \quad \text{für} \quad (x \leqslant y \ \wedge \ x \neq y)$$

und außerdem $x \geqslant y$ für $y \leqslant x$, bzw. $x > y$ für $y < x$. Aufgrund der Antisymmetrie kann für $x, y \in X$ von den drei Möglichkeiten $x < y$, $x > y$, $x = y$ höchstens eine zutreffen, bei einer totalen Ordnung genau eine.

14.1 Beispiele (1) Die übliche Ordnung der natürlichen Zahlen nach ihrer Größe ist eine lineare Ordnung. Eine formale Definition ist

$$m \leqslant n \quad \Longleftrightarrow \quad \exists k \in \mathbb{N}_0 \colon n = m + k \qquad (m, n \in \mathbb{N}). \qquad \diamond$$

(2) Auf der Menge $\mathbb{N} \times \mathbb{N}$ der Paare natürlicher Zahlen ist durch

$$(m_1, m_2) \leqslant_p (n_1, n_2) \quad \Longleftrightarrow \quad m_1 \leqslant n_1 \text{ und } m_2 \leqslant n_2$$

eine partielle Ordnung definiert, die beide Einträge mit einander vergleicht. Diese Ordnung ist nicht linear, denn es gilt beispielsweise weder $(1, 2) \leqslant_p (2, 1)$ noch $(2, 1) \leqslant_p (1, 2)$. Die beiden Paare $(1, 2)$ und $(2, 1)$ sind hier nicht vergleichbar.

(3) Eine lineare Ordnung auf $\mathbb{N} \times \mathbb{N}$ ist die **lexikographische Ordnung**

$$(m_1, m_2) \leqslant_{lex} (n_1, n_2) \quad \Longleftrightarrow \quad m_1 < n_1 \text{ oder } (m_1 = n_1 \text{ und } m_2 \leqslant n_2)$$

die zuerst den ersten Eintrag vergleicht und den zweiten nur betrachtet, wenn der erste übereinstimmt. Der Name kommt daher, dass dies der Methode entspricht, nach der ein Wörterbuch sortiert ist.

(4) Für zwei Vektoren in $\mathbb{R}^n$, $n \geqslant 2$, können wir ihre Norm vergleichen. Allerdings können verschiedene Vektoren dieselbe Norm haben, zum Beispiel $\|\vec{e}_1\| = \|\vec{e}_2\|$, das heißt, die Antisymmetrie ist nicht gegeben. Mit anderen Worten, durch $\vec{x} \leqslant \vec{y} \Leftrightarrow \|\vec{x}\| \leqslant \|\vec{y}\|$ ist *keine* partielle Ordnung auf $\mathbb{R}^n$ definiert. $\diamond$

14.2 Mengeninklusion als partielle Ordnung

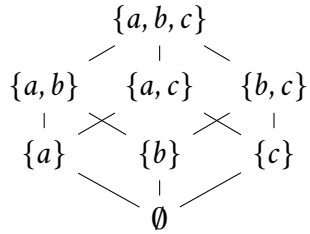
Ein wichtiges Beispiel für eine partielle Ordnung ist die Inklusion von Teilmengen. Ist C irgendeine Menge und $\mathcal{P}(C)$ ihre Potenzmenge, also die Menge aller Teilmengen von C, dann ist $\mathcal{P}(C)$ partiell geordnet durch die Definition[1]

$$A \leqslant B \quad \Longleftrightarrow \quad A \subset B.$$

für Teilmengen $A, B \subset C$. Sobald C mehr als zwei Elemente hat, ist die partielle Ordnung von $\mathcal{P}(C)$ durch Inklusion nicht mehr linear. Denn sind $a, b \in C$ mit $a \neq b$, dann sind die einelementigen Mengen $\{a\}$ und $\{b\}$ nicht vergleichbar.

[1]Unsere Notationen für Ordnungen und Inklusion sind nicht ganz analog, weil wir $\subset$ und nicht $\subseteq$ für $\leqslant$, sowie $\subsetneq$ für $<$ schreiben. Viele Autorinnen und Autoren machen das anders. Es ist aber jedenfalls nicht üblich, mit $\subset$ die strikte Inklusion zu bezeichnen.

14.2 Beispiel Ist $C = \{a, b, c\}$ eine Menge mit drei Elementen, dann hat $\mathcal{P}(C)$ acht Elemente, nämlich $\mathcal{P}(C) = \big\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\big\}$. Die partielle Ordnung von $\mathcal{P}(C)$ lässt sich gut in einem Diagramm visualisieren:



Dabei stehen die größeren Teilmengen oben und die Kanten zeigen die Inklusionen zwischen den Teilmengen von C an. ◇

Definition Es sei $\leqslant$ eine partielle Ordnung auf einer Menge X. Ein Element $x \in X$ heißt ein

- **größtes Element**, wenn $x \geqslant y$ für alle $y \in X$ gilt.
- **kleinstes Element**, wenn $x \leqslant y$ für alle $y \in X$ gilt.
- **maximales Element**, wenn es kein Element $y \in X$ mit $y > x$ gibt.
- **minimales Element**, wenn es kein Element $y \in X$ mit $y < x$ gibt.

In Worten: Ein größtes Element ist größer als jedes andere. Ein maximales Element ist eines, das von keinem anderen übertroffen wird; entsprechend für kleinste und minimale Elemente. Den Unterschied sieht man nur bei partiellen Ordnungen, die nicht total sind.

14.3 Beispiele (1) In den Zahlbereichen $\mathbb{Z}$ und $\mathbb{Q}$ gibt es kein kleinstes und kein größtes Element. In $\mathbb{N}$ ist 1 das kleinste Element und es gibt kein größtes. Die Frage, welche Teilmengen von $\mathbb{Q}$ und $\mathbb{R}$ ein größtes Element besitzen, wird in der Analysis-Vorlesung ausführlich diskutiert.

(2) In der Potenzmenge $\mathcal{P}(C)$ einer Menge C, partiell geordnet durch Inklusion, gibt es immer ein größtes Element, nämlich die ganze Menge C, und ein kleinstes Element, nämlich die leere Menge $\emptyset$. Dagegen gibt es in der Teilmenge $\mathcal{P}^*(C)$ aller nichtleeren Teilmengen von C in aller Regel kein kleinstes Element, dafür aber minimale Elemente, nämlich alle einelementigen Teilmengen von C. ◇

14.3 Angeordnete Körper

Bei den Zahlbereichen $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ ist es wichtig, dass die übliche Ordnungsrelation auch mit Addition und Multiplikation verträglich ist.

Definition Ein **angeordneter Körper** ist ein Körper K zusammen mit einer linearen Ordnung $\leqslant$, die folgende Eigenschaften besitzt:

(A) Für alle $a, b, c \in K$ gilt $a \leqslant b \Rightarrow a + c \leqslant b + c$.

(M) Für alle $a, b, c \in K$ mit $c > 0$ gilt $a \leqslant b \Rightarrow ac \leqslant bc$.

Aus Eigenschaft (A) folgt direkt, dass

$$a \geqslant b \quad \Longleftrightarrow \quad a - b \geqslant 0$$

für alle $a, b \in K$ gilt. In einem angeordneten Körper reicht es also zu wissen, welche Elemente positiv sind.

14.4 Beispiele Die üblichen Ordnungen von $\mathbb{Q}$ und $\mathbb{R}$ machen diese zu angeordneten Körpern und können etwa wie folgt definiert weden:

- Für $x \in \mathbb{Z}$: $x \geqslant 0 \Leftrightarrow x \in \mathbb{N}_0$.
- Für $\frac{a}{b} \in \mathbb{Q}$: $\frac{a}{b} \geqslant 0 \Leftrightarrow ab \geqslant 0$ in $\mathbb{Z}$.
- Für $x \in \mathbb{R}$: $x \geqslant 0 \Leftrightarrow \exists z \in \mathbb{R} \colon x = z^2$.

Dabei ist $\mathbb{Z}$ natürlich kein Körper (aber eine Teilmenge des angeordneten Körpers $\mathbb{Q}$, ein *angeordneter Ring*). ◇

14.5 Lemma *Es sei $(K, \leqslant)$ ein angeordneter Körper.*

(1) Es gilt $a^2 \geqslant 0$ für alle $a \in K$. Insbesondere gilt $1 > 0$.

(2) Es gilt $-1 < 0$.

(3) Für alle $c \in K$ gilt $c > 0 \Leftrightarrow c^{-1} > 0$.

(4) Für alle $a, b, c \in K$ mit $c < 0$ gilt $a \leqslant b \Leftrightarrow ac \geqslant bc$.

(5) Für alle $a, b \in K$ gilt $a \leqslant b \Leftrightarrow -a \geqslant -b$.

Beweis. (1) Falls $a \geqslant 0$, dann folgt $a^2 \geqslant 0$ sofort aus Eigenschaft (M) (mit $b = 0$ und $c = a$). Ist $a < 0$, dann folgt $-a > 0$ aus (A) und damit $a^2 = (-a)^2 > 0$.

(2) folgt aus (1) unter Verwendung von (A).

(3) Sei $c > 0$. Wäre $c^{-1} < 0$, dann würde wegen (M) also $1 < 0$ folgen, im Widerspruch zu (1). Die Umkehrung folgt genauso.

(4) Wegen $a \leqslant b$ und $-c > 0$ folgt $-ac \leqslant -bc$, also $bc \leqslant ac$ durch Addition von $ac + bc$ auf beiden Seiten. Die Umkehrung folgt wegen (3).

(5) folgt direkt aus (2) und (4). ■

14.6 Korollar *Der Körper $\mathbb{C}$ der komplexen Zahlen und die endlichen Körper $\mathbb{F}_p$, für jede Primzahl p, besitzen keine Anordnungen.*

Beweis. Das bedeutet genauer: Ist K einer dieser Körper, dann gibt es keine lineare Ordnung auf der Menge K, durch die K zu einem angeordneten Körper würde.

Denn für $K = \mathbb{C}$ ist $-1 = i^2$ ein Quadrat und müsste damit nach dem vorangehenden Lemma sowohl positiv als auch negativ sein, was nicht möglich ist.

In $K = \mathbb{F}_p$ gilt $-1 = p - 1 = (1 + \cdots + 1)$. Aus $1 > 0$ nach dem Lemma würde dann mit Eigenschaft (A) wieder $-1 > 0$ folgen. ■

IV

Vektorräume und lineare Abbildungen

The power of mathematics is often to change one thing into another, to change geometry into language.

Marcus du Sautoy

15 Abstrakte Vektorräume

In diesem Kapitel wird der wichtigste Begriff der abstrakten linearen Algebra eingeführt, nämlich der des Vektorraums. Er verallgemeinert den Raum der Spaltenvektoren und konzentriert sich nur auf seine formalen Eigenschaften.

Im ganzen Kapitel bezeichnet K einen beliebigen Körper.

15.1 Vektorräume

Definition Ein **Vektorraum über** K (kurz K-Vektorraum) ist eine abelsche Gruppe $(V, +)$ zusammen mit einer Abbildung

$$K \times V \to V,\ (a, \mathbf{v}) \mapsto a \cdot \mathbf{v}$$

genannt **Skalarmultiplikation** mit den folgenden Eigenschaften:

(VA) Für alle $a, b \in K$ und alle $\mathbf{v} \in V$ gilt

$$(ab)\mathbf{v} = a(b\mathbf{v}) \qquad \textit{(Assoziativität).}$$

(VD) Für alle $a, b \in K$ und alle $\mathbf{v}, \mathbf{w} \in V$ gelten

$$(a+b)\mathbf{v} = a\mathbf{v} + b\mathbf{v} \quad \text{und} \quad a(\mathbf{v}+\mathbf{w}) = a\mathbf{v} + a\mathbf{w} \qquad \textit{(Distributivität).}$$

(VE) Für alle $\mathbf{v} \in V$ gilt $\qquad 1 \cdot \mathbf{v} = \mathbf{v}.$

Elemente von V heißen **Vektoren** und die Verknüpfung + die **Vektoradditon**. Das neutrale Element der Vektoraddition ist der **Nullvektor 0**.

15.1 Beispiele (1) Für jedes $n \in \mathbb{N}$ ist K^n mit der üblichen Vektoraddition und Skalarmultiplikation

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}, \quad a \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a \cdot x_1 \\ \vdots \\ a \cdot x_n \end{pmatrix} \qquad \text{für } \vec{x}, \vec{y} \in K^n \text{ und } a \in K$$

ein K-Vektorraum. Die Körperaxiome stellen dabei sicher, dass die Vektorraumaxiome alle erfüllt sind. Es ist $K^1 = K$ und $K^0 = \{\mathbf{0}\}$ der **Nullraum**. Zur besse-

ren Unterscheidung behalten die Spaltenvektoren aus K^n ihr Pfeilchen obendrüber, während Vektoren in allgemeinen Vektorräumen fettgedruckt sind.

(2) Statt nur endlich viele können wir auch unendlich viele Einträge erlauben und von Vektoren zu Folgen übergehen: Die Menge

$$K^{\mathbb{N}} = \{(a_1, a_2, a_3, \dots) \mid a_i \in K \text{ für alle } i \in \mathbb{N}\}$$

aller Folgen in K ist ein Vektorraum: Wir können solche Folgen genau wie Vektoren komponentenweise addieren und mit Skalaren aus K multiplizieren.

(3) Wie zuvor schreiben wir $\mathrm{Mat}_{m\times n}(K)$ für die Menge aller $m \times n$-Matrizen mit Einträgen aus K. Die Menge $\mathrm{Mat}_{m\times n}(K)$ ist ein Vektorraum über K: Für $A, B \in \mathrm{Mat}_{m\times n}(K)$ ist $A + B$ die Matrix mit den Einträgen $a_{ij} + b_{ij}$ und für $c \in K$ ist $c \cdot A$ die Matrix mit den Einträgen $c \cdot a_{ij}$. Mit anderen Worten, das ist einfach der Vektorraum K^{mn} mit der normalen Vektoraddition und Skalarmultiplikation, nur mit doppelter statt mit einfacher Indizierung. ◇

15.2 Lemma *In jedem Vektorraum V über K gelten die folgenden Rechenregeln:*

(1) Sind $\mathbf{v}, \mathbf{w} \in V$ mit $\mathbf{v} + \mathbf{w} = \mathbf{v}$, dann folgt $\mathbf{w} = \mathbf{0}$.

(2) Sind $\mathbf{v}, \mathbf{w} \in V$ mit $\mathbf{v} + \mathbf{w} = \mathbf{0}$, dann folgt $\mathbf{w} = -\mathbf{v}$.

(3) Für alle $\mathbf{v} \in V$ und $a \in K$ gelten $0 \cdot \mathbf{v} = \mathbf{0}$ und $a \cdot \mathbf{0} = \mathbf{0}$.

(4) Ist $\mathbf{v} \in V$ und $a \in K$ mit $a\mathbf{v} = \mathbf{0}$, dann folgt $a = 0$ oder $\mathbf{v} = \mathbf{0}$.

(5) Für alle $\mathbf{v} \in V$ gilt $-1 \cdot \mathbf{v} = -\mathbf{v}$.

Beweis. (1) und (2) Addiere $(-\mathbf{v})$ auf beiden Seiten und vereinfache.

(3) Es gilt $0 \cdot \mathbf{v} = (0 + 0) \cdot \mathbf{v} = 0 \cdot \mathbf{v} + 0 \cdot \mathbf{v}$. Wir addieren $-(0 \cdot \mathbf{v})$ auf beiden Seiten und erhalten $\mathbf{0} = 0 \cdot \mathbf{v}$. Entsprechend folgt aus $a \cdot \mathbf{0} = a \cdot (\mathbf{0}+\mathbf{0}) = a \cdot \mathbf{0} + a \cdot \mathbf{0}$ die zweite Behauptung.

(4) Ist $a\mathbf{v} = \mathbf{0}$ mit $a \neq 0$, so folgt $\mathbf{v} = 1\mathbf{v} = (a^{-1}a)\mathbf{v} = a^{-1}(a\mathbf{v}) = a^{-1}\mathbf{0} = \mathbf{0}$.

(5) Es gilt $\mathbf{v} + (-1)\mathbf{v} = (1 - 1)\mathbf{v} = 0 \cdot \mathbf{v} = \mathbf{0}$. Wegen (2) folgt $(-1) \cdot \mathbf{v} = -\mathbf{v}$. ■

Wie in $\mathbb{R}^n$ gibt es den Begriff des linearen Unterraums.

Definition Es sei V ein K-Vektorraum. Eine Teilmenge $U \subset V$ heißt **linearer Unterraum** (oder *Untervektorraum* oder kurz *Unterraum*), wenn folgendes gilt:

(NL) Es gilt $U \neq \emptyset$.

(ADD) Für alle $\mathbf{u}, \mathbf{v} \in U$ gilt $\mathbf{u} + \mathbf{v} \in U$.

(SKM) Für alle $\mathbf{u} \in U$ und $a \in K$ gilt $a \cdot \mathbf{u} \in U$.

Die beiden Eigenschaften (ADD) und (SKM) kann man wieder zusammenfassen:

(LK) Für alle $\mathbf{u}, \mathbf{v} \in U$ und alle $a, b \in K$ gilt $a\mathbf{u} + b\mathbf{v} \in U$.

15.3 Lemma *Jeder lineare Unterraum enthält den Nullvektor.*

Beweis. Genauso wie für $V = \mathbb{R}^n$ (Lemma 4.2). ∎

15.4 Beispiele (1) Die Lösungsmenge eines homogenen linearen Gleichungssystems mit Koeffizienten in K ein linearer Unterraum von K^n, der **Lösungsraum**.

(2) In jedem Vektorraum V ist der **Nullraum** $\{\mathbf{0}\}$ ein linearer Unterraum. ◇

15.5 Lemma *Es sei V ein K-Vektorraum und U ein linearer Unterraum von V. Dann ist U, mit der Vektoraddition und Skalarmultiplikation aus V, selbst ein K-Vektorraum.*

Beweis. Der Punkt ist folgender: Die Eigenschaften (ADD) und (SKM) sorgen dafür, dass die Einschränkungen der Vektoraddition und Skalarmultiplikation auf U wieder in U landen und damit tatsächlich Abbildungen

$$\begin{cases} U \times U & \to & U \\ (\mathbf{v}, \mathbf{w}) & \mapsto & \mathbf{v} + \mathbf{w} \end{cases} \quad \text{und} \quad \begin{cases} K \times U & \to & U \\ (a, \mathbf{v}) & \mapsto & a \cdot \mathbf{v} \end{cases}$$

definieren. Die Vektorraumaxiome übertragen sich alle sofort von V auf U. ∎

Dieses Lemma zeigt schon einen kleinen Vorteil des abstrakten Vektorraumbegriffs: Wir brauchen Vektorräume und lineare Unterräume in allgemeinen Aussagen nicht zu unterscheiden.

15.6 Lemma *Es sei V ein K-Vektorraum und seien $U_1, U_2 \subset V$ lineare Unterräume.*

(1) Der Durchschnitt $U_1 \cap U_2$ ist wieder ein linearer Unterraum von V.

(2) Die Summe

$$U_1 + U_2 \;=\; \big\{\mathbf{u}_1 + \mathbf{u}_2 \mid \mathbf{u}_1 \in U_1,\ \mathbf{u}_2 \in U_2\big\}$$

ist wieder ein linearer Unterraum von V.

Beweis. (1) Übung. (2) Genauso wie im Fall $U_1, U_2 \subset \mathbb{R}^n$ (siehe Übungen). ∎

Die Aussage über den Durchschnitt gilt allgemeiner für jede Familie $(U_i)_{i \in I}$ von Unterräumen von V: Auch $\bigcap_{i \in I} U_i$ ist ein Unterraum. Die Aussage über die Summe verallgemeinert sich auf endlich viele Summanden $U_1 + \cdots + U_k$.

Wie in $\mathbb{R}^n$ können wir den von gegebenen Vektoren aufgespannten linearen Unterraum bilden. Das definieren wir gleich ein wenig allgemeiner:

Definition Für eine Teilmenge $\mathcal{A} \neq \emptyset$ von V bezeichnen wir mit $\mathrm{Lin}(\mathcal{A})$ den **von $\mathcal{A}$ aufgespannten Unterraum** oder den **Spann** von $\mathcal{A}$ in V, das heißt

$$\mathrm{Lin}(\mathcal{A}) = \left\{ \sum_{i=1}^{m} c_i \mathbf{v}_i \mid c_1, \ldots, c_m \in K,\ \mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathcal{A},\ m \geqslant 1 \right\}.$$

Außerdem setzen wir $\mathrm{Lin}(\emptyset) = \{\mathbf{0}\}$.

Der Spann ist also wieder die Menge der (endlichen) **Linearkombinationen** von Vektoren aus der Menge $\mathcal{A}$. Wieder schreiben wir bei einer endlichen Menge $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ von Vektoren auch kurz $\mathrm{Lin}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$.

15.7 Lemma *Für eine Menge $\mathcal{A} \subset V$ von Vektoren ist $\mathrm{Lin}(\mathcal{A})$ der kleinste lineare Unterraum von V, der die Menge $\mathcal{A}$ enthält. Das heißt es gelten:*

(1) $\mathrm{Lin}(\mathcal{A})$ *ist ein linearer Unterraum von V, der $\mathcal{A}$ enthält.*

(2) Ist U irgendein Unterraum von V mit $\mathcal{A} \subset U$, dann gilt $\mathrm{Lin}(\mathcal{A}) \subset U$.

Beweis. Völlig analog zu Lemma 4.3. ∎

15.2 Funktionenräume

Der Übergang von Spaltenvektoren zu Vektorräumen ist keine abstrakte Spielerei. Die Idee, dass alle möglichen Objekte eine Art »Raum« bilden, hat sich in der modernen Mathematik als sehr nützlich erwiesen.

Ist X eine nicht-leere Menge und K ein Körper, dann ist die Menge

$$\mathrm{Abb}(X, K) = \{f \colon X \to K\}$$

aller Abbildungen von X nach K ein Vektorraum. Denn zu zwei Funktionen $f, g \in \mathrm{Abb}(X, K)$ können wir die Summe $f + g$ durch die Gleichung

$$(f + g)(x) = f(x) + g(x)$$

für $x \in X$ definieren, und für $f \in \mathrm{Abb}(X, K)$ und $a \in K$ ist $a \cdot f$ durch

$$(af)(x) = a \cdot f(x)$$

für $x \in X$ definiert. Der Nullvektor in diesem Raum ist die Nullabbildung, die alle Elemente von X auf 0 abbildet. Dass man Funktionen auch als Vektoren auffassen kann, ist von grundsätzlicher Bedeutung für die moderne Analysis (Stichwort: Funktionalanalysis). In der linearen Algebra wird das nur in Beispielen gelegentlich eine Rolle spielen.

15.8 Beispiele (1) Man kann die Analogie zwischen Spaltenvektoren und Funktionen auch etwas konkreter verstehen als nur über die Vektorraumaxiome: Ein Vektor $\vec{x} \in K^n$ ist »dasselbe« wie die Abbildung $i \mapsto x_i$, die jedem Index den entsprechenden Eintrag zuweist. Mit anderen Worten, K^n entspricht dem Spezialfall $\mathrm{Abb}(X, K)$ für $X = \{1, \dots, n\}$.

(2) Der Vektorraum $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ aller Funktionen $\mathbb{R} \to \mathbb{R}$ hat eine Reihe von wichtigen linearen Unterräumen, zum Beispiel:

$$\begin{aligned} C^0(\mathbb{R}) &= \{f \colon \mathbb{R} \to \mathbb{R} \mid f \text{ ist stetig}\} \\ C^k(\mathbb{R}) &= \{f \colon \mathbb{R} \to \mathbb{R} \mid f \text{ ist } k\text{-mal stetig differenzierbar}\} \\ C^\infty(\mathbb{R}) &= \{f \colon \mathbb{R} \to \mathbb{R} \mid f \text{ ist beliebig oft differenzierbar}\} \end{aligned}$$

Die Unterraumeigenschaften (ADD) und (SKM) entsprechen dabei aus der Analysis bekannten Eigenschaften von stetigen bzw. differenzierbaren Funktionen. ◇

15.9 Beispiel Die Menge der *quadratsummierbaren Folgen*

$$\ell^2(\mathbb{R}) = \left\{ (a_n)_{n\in\mathbb{N}} \in \mathbb{R}^\infty \mid \sum_{n=1}^{\infty} a_n^2 \text{ ist konvergent} \right\}$$

(gelesen »Klein-Ell-Zwei«) ist ein linearer Unterraum des Folgenraums $\mathbb{R}^\infty$ und eines der beiden wichtigsten Beispiele für einen **Hilbertraum**. Offenbar liegt die Nullfolge in $\ell^2(\mathbb{R})$. Weil die Summanden a_n^2 alle größer gleich 0 sind, genügt es für die Konvergenz, die Beschränktheit der Folge der Partialsummen nachzuprüfen. Also liegt $(a_n)_{n\in\mathbb{N}}$ genau dann in $\ell^2(\mathbb{R})$, wenn es $M \in \mathbb{R}$ gibt derart, dass für alle $N \in \mathbb{N}$ gilt: $\sum_{n=1}^{N} a_n < M$. In diesem Fall folgt für alle $c \in \mathbb{R}$ auch $\sum_{n=1}^{N} (ca_n)^2 < c^2 M$. Also ist (SKM) erfüllt. Ist auch $(a'_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{R})$ und etwa $\sum_{n=1}^{N} a'_n < M'$ für $M' \in \mathbb{R}$, dann folgt außerdem

$$\sum_{n=1}^{N} a_n a'_n \leqslant \sqrt{\sum_{n=1}^{N} a_n^2} \cdot \sqrt{\sum_{n=1}^{N} (a'_n)^2} \leqslant \sqrt{MM'}$$

nach der Cauchy-Schwarz-Ungleichung (Satz 8.5). Daraus folgt dann

$$\sum_{n=1}^{N} (a_n + a'_n)^2 = \sum_{n=1}^{N} a_n^2 + 2 \sum_{n=1}^{N} a_n a'_n + \sum_{n=1}^{N} (a'_n)^2 \leqslant M + 2\sqrt{MM'} + M'$$

für alle $N \in \mathbb{N}$ und damit $(a_n + a'_n)_{n\in\mathbb{N}} \in \ell^2(\mathbb{R})$. ◇

15.10 Beispiel (Polynome) Für $k \in \mathbb{N}_0$ schreiben wir

$$x^k \colon \left\{ \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ a & \mapsto & a^k \end{array} \right. .$$

Das ist ein Element des Vektorraums $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ aller Funktionen $\mathbb{R} \to \mathbb{R}$. Darin eingeschlossen ist die konstante Funktion $x^0 = 1$. Wir schreiben

$$\mathbb{R}[x] = \mathrm{Lin}\big(\{x^k \mid k \in \mathbb{N}_0\}\big) = \big\{a_0 + a_1 x + \cdots a_n x^n \mid a_0, \ldots, a_n \in \mathbb{R},\ n \in \mathbb{N}\big\}.$$

Die Funktionen in $\mathbb{R}[x]$ sind die vertrauten **Polynomfunktionen** (oder kurz Polynome[1]) in einer Variablen x. Sie sind also Linearkombinationen der Potenzfunktionen (Monome) x^k. Ein Polynom $a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d$ mit $a_d \neq 0$ hat den **Grad** d. Für $d \in \mathbb{N}_0$ schreiben wir außerdem

$$\mathbb{R}[x]_{\leqslant d} = \mathrm{Lin}(1, x, \ldots, x^d) = \big\{a_0 + a_1 x + \cdots + a_d x^d \mid a_0, \ldots, a_d \in \mathbb{R}\big\}$$

für den linearen Unterraum aller Polynome vom Grad höchstens d. In der linearen Algebra werden Polynome im zweiten Teil eine größere Rolle spielen. ◇

15.3 Lineare Unabhängigkeit, Basen und Dimension

Wir übertragen nun die Begriffe der linearen Unabhängigkeit, der Basis und der Dimension auf allgemeine Vektorräume.

Definition Es sei V ein K-Vektorraum. Ein System $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ von Vektoren aus V heißt **linear unabhängig**, wenn die Implikation

$$c_1 \mathbf{v}_1 + \cdots + c_m \mathbf{v}_m = \mathbf{0} \quad \Rightarrow \quad c_1 = \cdots = c_m = 0$$

für alle Skalare $c_1, \ldots, c_m \in K$ gilt. Andernfalls heißt das System **linear abhängig**. Allgemeiner definieren wir ein beliebig indiziertes System $(\mathbf{v}_i)_{i \in I}$ von Vektoren aus V als linear unabhängig, wenn das endliche System $\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_m}$ für jede Wahl von endlich vielen verschiedenen Indizes $i_1, \ldots, i_m \in I$ linear unabhängig ist.

15.11 Beispiel Diese Definition ist genauso wie in $\mathbb{R}^n$: Lineare Unabhängigkeit bedeutet, dass keine **nicht-triviale lineare Relation** zwischen den Vektoren des

[1] Die Definition ist im Prinzip für jeden Körper K sinnvoll, nicht nur für $\mathbb{R}$. Es gibt aber einen Unterschied zwischen Polynomen und Polynomfunktionen, der vor allem für endliche Körper relevant ist. Das diskutieren wir später an geeigneter Stelle.

Systems besteht. Ein Beispiel für ein unendliches linear unabhängiges System ist das System

$$1, x, x^2, x^3, \cdots$$

aller Monome im Raum $\mathbb{R}[x]$ der Polynome, also $(\mathbf{v}_i)_{i\in I}$ mit $\mathbf{v}_i = x^i$, $I = \mathbb{N}_0$. Die lineare Unabhängigkeit gilt in diesem Fall, weil ein Polynom

$$a_0 + a_1 x + a_2 x^2 \cdots + a_n x^n$$

nur dann der Nullvektor in $\mathbb{R}[x]$ ist, also die Nullfunktion, wenn die Koeffizienten $a_0, \ldots, a_n$ alle gleich 0 sind (Übung). Das gilt für Polynome von jedem Grad und damit für jede Auswahl einer Linearkombination von endlich vielen verschiedenen Monomen. ◇

15.12 Beispiel Im Raum $C^\infty(\mathbb{R})$ aller beliebig oft differenzierbaren Funktionen $\mathbb{R} \to \mathbb{R}$ betrachten wir das System von Funktionen

$$\sin(x), \sin(2x), \cos(x).$$

Es ist linear unabhängig: Seien $a, b, c, \in \mathbb{R}$ mit

$$a\sin(x) + b\sin(2x) + c\cos(x) = \mathbf{0}$$

(Nullfunktion!), dann muss $a = b = c = 0$ gelten. Um das zu zeigen, können wir Werte für x einsetzen (so wie wir bei Spaltenvektoren einzelne Einträge betrachten können). Wir wissen zum Beispiel, dass $\cos(\frac{\pi}{2}) = 0$, $\sin(\pi) = 0$ und $\sin(\frac{\pi}{2}) = 1$ gelten. Wenn wir also $x = \frac{\pi}{2}$ in die Relation einsetzen, dann bekommen wir

$$0 = a\sin(\tfrac{\pi}{2}) + b\sin(\pi) + c\cos(\tfrac{\pi}{2}) = a$$

also gilt $a = 0$. Von der Relation bleibt nur noch $b\sin(2x) + c\cos(x) = \mathbf{0}$ übrig. Weiter wissen wir $\sin(0) = 0$ und $\cos(0) = 1$, also bekommen wir

$$0 = b\sin(0) + c\cos(0) = c$$

und damit $c = 0$. Aus $b\sin(2x) = \mathbf{0}$ folgt dann noch $b = 0$, da $\sin(2x) \neq \mathbf{0}$ gilt. ◇

15.13 Satz *Es sei $(\mathbf{v}_i)_{i\in I}$ ein System von Vektoren in V.*

(1) Genau dann ist $(\mathbf{v}_i)_{i\in I}$ linear abhängig, wenn es einen Index $k \in I$ gibt mit

$$\mathrm{Lin}(\mathbf{v}_i \mid i \in I) = \mathrm{Lin}(\mathbf{v}_i \mid i \in I,\ i \neq k).$$

(2) *Ist das System $(\mathbf{v}_i)_{i\in I}$ linear unabhängig und ist $\mathbf{v} \in V$ ein Vektor, der nicht in $\mathrm{Lin}(\mathbf{v}_i \mid i \in I)$ enthalten ist, dann ist auch das verlängerte System $(\mathbf{v}_i)_{i\in I\cup\{j\}}$, in dem $\mathbf{v}_j = \mathbf{v}$ für einen neuen Index $j \notin I$ angehängt wird, linear unabhängig.*

Beweis. Analog zu Satz 7.6. ∎

Definition Ein System $(\mathbf{v}_i)_{i\in I}$ in einem Vektorraum V heißt **erzeugend**, wenn $V = \mathrm{Lin}(\mathbf{v}_i \mid i \in I)$ gilt. Ein Vektorraum V heißt **endlichdimensional** (oder *endlich erzeugt*), wenn er von einem endlichen System von Vektoren erzeugt wird.

Definition Ein System von Vektoren $(\mathbf{v}_i)_{i\in I}$ aus V heißt eine **Basis** von V, wenn es linear unabhängig und erzeugend ist.

In endlichdimensionalen Vektorräumen gelten dieselben Aussagen, die wir für lineare Unterräume von $\mathbb{R}^n$ bewiesen haben:

15.14 Satz *Es sei V ein endlichdimensionaler Vektorraum.*

(1) Basisauswahl: *Jedes System $\mathbf{v}_1, \dots, \mathbf{v}_m$ von Vektoren mit $V = \mathrm{Lin}(\mathbf{v}_1, \dots, \mathbf{v}_m)$ enthält eine Basis von V. Insbesondere besitzt V eine endliche Basis.*

(2) *Alle Basen von V haben dieselbe Länge.*

(3) *Für ein System $\mathbf{v}_1, \dots, \mathbf{v}_n$ von Vektoren in V sind äquivalent:*

   (i) *Das System $\mathbf{v}_1, \dots, \mathbf{v}_n$ ist eine Basis von V.*

   (ii) *Jeder Vektor $\mathbf{v} \in V$ hat eine eindeutige Darstellung*

$$\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$$

   als Linearkombination von $\mathbf{v}_1, \dots, \mathbf{v}_n$.

   (iii) *Das System $\mathbf{v}_1, \dots, \mathbf{v}_n$ ist erzeugend und nicht verkürzbar, d.h. kein echtes Teilsystem von $\mathbf{v}_1, \dots, \mathbf{v}_n$ ist erzeugend.*

   (iv) *Das System $\mathbf{v}_1, \dots, \mathbf{v}_n$ ist linear unabhängig und nicht verlängerbar, d.h. jedes längere System, in dem $\mathbf{v}_1, \dots, \mathbf{v}_n$ vorkommen, ist linear abhängig.*

(4) *Jeder lineare Unterraum von V ist endlichdimensional.*

(5) Basisergänzung: *Ist $U \subset V$ ein linearer Unterraum und $\mathbf{u}_1, \dots, \mathbf{u}_m \in U$ eine Basis von U, dann gibt es $k \in \mathbb{N}_0$ und Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ derart, dass das System $\mathbf{u}_1, \dots, \mathbf{u}_m, \mathbf{v}_1, \dots, \mathbf{v}_k$ eine Basis von V ist.*

Beweis. Die Beweise gehen wörtlich genauso wie für die entsprechenden Aussagen für lineare Unterräume von $\mathbb{R}^n$ in Kapitel 7. ∎

Für unendlichdimensionale Vektorräume gelten diese Aussagen entsprechend, wobei man für den Beweis allerdings ein abstraktes Prinzip der Mengenlehre, das *Zornsche Lemma*, heranzieht. Darauf kommen wir im nächsten Semester zurück.

15.15 Beispiel Es gibt trotzdem unendlichdimensionale Vektorräume, für die wir direkt eine Basis angeben können: Das System $1, x, x^2, \dots$ im Vektorraum $\mathbb{R}[x]$ ist linear unabhängig und per Definition erzeugend, also eine Basis. ◇

Definition Die **Dimension** eines endlichdimensionalen Vektorraums V ist die Länge einer Basis von V und wird mit $\dim(V)$ bezeichnet.

15.16 Satz (Dimensionsformel für Unterräume) *Es sei V ein endlichdimensionaler Vektorraum und seien U_1 und U_2 zwei lineare Unterräume von V.*

(1) Falls $U_1 \subset U_2$, dann gilt $\dim(U_1) \leqslant \dim(U_2)$. Dabei ist $\dim(U_1) = \dim(U_2)$ genau dann, wenn $U_1 = U_2$ gilt.

(2) Es gilt

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

Beweis. Genauso wie Satz 7.16. ∎

15.17 Korollar *Ist V ein endlichdimensionaler Vektorraum der Dimension n und ist $\mathbf{v}_1, \dots, \mathbf{v}_n$ ein System aus n Vektoren in V, dann sind äquivalent:*

(1) $\mathbf{v}_1, \dots, \mathbf{v}_n$ sind eine Basis von V.

(2) $\mathbf{v}_1, \dots, \mathbf{v}_n$ erzeugen V.

(3) $\mathbf{v}_1, \dots, \mathbf{v}_n$ sind linear unabhängig.

Beweis. (1) impliziert per Definition (2) und (3). Es gelte (2). Nach Basisauswahl gibt es unter den $\mathbf{v}_1, \dots, \mathbf{v}_n$ dann eine Basis. Deren Länge kann aber nicht kleiner als $n = \dim(V)$ sein. Also folgt (1). Es gelte (3). Dann ist $U = \mathrm{Lin}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ ein linearer Unterraum der Dimension n von V und damit $U = V$ nach dem vorangehenden Satz. ∎

15.18 Beispiel Die drei Vektoren

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \vec{e}_1 + \vec{e}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \vec{e}_1 + \vec{e}_2 + \vec{e}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

in K^3 sind linear unabhängig. Nach dem vorangehenden Korollar sind sie damit eine Basis von K^3. Mit anderen Worten, es muss $\mathrm{Lin}(\vec{e}_1, \vec{e}_1 + \vec{e}_2, \vec{e}_1 + \vec{e}_2 + \vec{e}_3) = K^3$ gelten, ohne dass wir das noch extra nachprüfen müssen. ◇

Sind V und W zwei K-Vektorräume, dann ist das kartesische Produkt

$$V \times W = \{(\mathbf{v}, \mathbf{w}) \mid \mathbf{v} \in V \text{ und } \mathbf{w} \in W\}$$

wieder ein K-Vektorraum, wenn man Addition und Skalarmultiplikation komponentenweise definiert, also durch

$$(\mathbf{v}, \mathbf{w}) + (\mathbf{v}', \mathbf{w}') = (\mathbf{v} + \mathbf{v}', \mathbf{w} + \mathbf{w}') \quad \text{und} \quad a(\mathbf{v}, \mathbf{w}) = (a\mathbf{v}, a\mathbf{w})$$

für $\mathbf{v}, \mathbf{v}' \in V$, $\mathbf{w}, \mathbf{w}' \in W$ und $a \in K$.

Definition Der Vektorraum $V \times W$ heißt das **direkte Produkt** von V und W.

Für $V = K^m$ und $W = K^n$ besteht $K^m \times K^n$ aus allen Paaren $\left(\vec{x}, \vec{y}\right)$. Natürlich kann man die Einträge von $\vec{x}$ und $\vec{y}$ auch in einen Vektor der Länge $m + n$ untereinander schreiben. Meistens unterscheidet man deshalb nicht nicht zwischen $K^m \times K^n$ und K^{m+n}, obwohl es streng genommen nicht genau dasselbe ist.

Entsprechend bildet man das direkte Produkt

$$V_1 \times \cdots \times V_n$$

endlich vieler K-Vektorräume $V_1, \ldots, V_n$.

15.19 Satz *Das direkte Produkt zweier endlichdimensionaler Vektorräume V und W hat die Dimension* $\dim(V \times W) = \dim(V) + \dim(W)$.

Beweis. Es sei $m = \dim(V)$ und $n = \dim(W)$ und seien $\mathbf{v}_1, \ldots, \mathbf{v}_m$ und $\mathbf{w}_1, \ldots, \mathbf{w}_n$ Basen von V bzw. W. Wir behaupten, dass die $m + n$ Paare

$$(\mathbf{v}_1, \mathbf{0}), \ldots, (\mathbf{v}_m, \mathbf{0}), (\mathbf{0}, \mathbf{w}_1), \ldots, (\mathbf{0}, \mathbf{w}_n)$$

eine Basis von $V \times W$ bilden. Denn ist $(\mathbf{v}, \mathbf{w}) \in V \times W$, dann gibt es Darstellungen $\mathbf{v} = \sum_{i=1}^{m} x_i \mathbf{v}_i$ und $\mathbf{w} = \sum_{j=1}^{n} y_j \mathbf{w}_j$ und damit

$$\begin{aligned}(\mathbf{v}, \mathbf{w}) &= (\mathbf{v}, \mathbf{0}) + (\mathbf{0}, \mathbf{w}) = \Big(\sum\nolimits_{i=1}^{m} x_i \mathbf{v}_i, \mathbf{0}\Big) + \Big(\mathbf{0}, \sum\nolimits_{j=1}^{n} y_j \mathbf{w}_j\Big) \\ &= \sum_{i=1}^{m} x_i (\mathbf{v}_i, \mathbf{0}) + \sum_{j=1}^{n} y_j (\mathbf{0}, \mathbf{w}_j).\end{aligned}$$

Für die lineare Unabhängigkeit betrachten wir eine lineare Relation

$$\sum_{i=1}^{m} c_i(\mathbf{v}_i, \mathbf{0}) + \sum_{j=1}^{n} d_j(\mathbf{0}, \mathbf{w}_j) = (\mathbf{0}, \mathbf{0}).$$

Daraus folgen $\sum_{i=1}^{m} c_i\mathbf{v}_i = \mathbf{0}$ und $\sum_{j=1}^{n} d_j\mathbf{w}_j = \mathbf{0}$, also $c_1 = \cdots = c_m = d_1 = \cdots = d_n = 0$, da $\mathbf{v}_1, \ldots, \mathbf{v}_m$ und $\mathbf{w}_1, \ldots, \mathbf{v}_n$ jeweils linear unabhängig sind. ■

Analog oder per Induktion bekommt man die entsprechende Aussage für mehr als zwei Faktoren, das heißt für endlichdimensionale Vektorräume $V_1, \ldots, V_k$ gilt

$$\dim(V_1 \times \cdots \times V_k) = \dim(V_1) + \cdots + \dim(V_k).$$

15.4 Körpererweiterungen

Den Körper $\mathbb{C}$ der komplexen Zahlen können wir auch als Vektorraum über den reellen Zahlen auffassen ($V = \mathbb{C}$, $K = \mathbb{R}$). Tatsächlich haben wir bei der Konstruktion von $\mathbb{C}$ ja den Vektorraum $\mathbb{R}^2$ zu Grunde gelegt. Die Vektoren sind dann komplexe Zahlen $a + bi$ (mit $a, b \in \mathbb{R}$) und die Skalarmultiplikation ist für $\alpha \in \mathbb{R}$ durch $\alpha \cdot (a + bi) = \alpha a + \alpha bi$ gegeben.

In ähnlicher Weise können wir auch die reellen Zahlen als Vektorraum über den rationalen Zahlen auffassen ($V = \mathbb{R}$, $K = \mathbb{Q}$). Die Vektoren sind dann also reelle Zahlen (mit der normalen Addition reeller Zahlen als Vektoraddition) und die Skalare rationale Zahlen (mit der normalen Multiplikation als Skalarmultiplikation). Es ist leicht zu überprüfen, dass die Vektorraumaxiome alle erfüllt sind. Tatsächlich geht das immer, sobald ein Körper in einem anderen enthalten ist (in einer *Körpererweiterung*). Diese Methode, mit Zahlen wie mit Vektoren umzugehen, spielt in der Algebra eine wichtige Rolle.

15.20 Beispiel Wir betrachten $\mathbb{R}$ als $\mathbb{Q}$-Vektorraum. Die beiden Vektoren $1, \sqrt{2}$ in $\mathbb{R}$ sind über $\mathbb{Q}$ linear unabhängig. Denn eine lineare Relation zwischen ihnen hat die Form

$$c_1 \cdot 1 + c_2\sqrt{2} = 0$$

mit $c_1, c_2 \in \mathbb{Q}$. Wäre hier $c_2 \neq 0$, dann würde also $\sqrt{2} = -\frac{c_1}{c_2} \in \mathbb{Q}$ folgen. Wir haben aber bewiesen, dass $\sqrt{2}$ irrational ist (Satz 13.3). Es muss also $c_2 = 0$ gelten und damit auch $c_1 = 0$. ◇

16 Lineare Abbildungen

Lineare Abbildungen zwischen Vektorräumen können wir genauso definieren wie in $\mathbb{R}^n$ in Kap. 9. Darüber hinaus definieren wir weitere Strukturen auf diveresen Mengen von linearen Abbildungen.

16.1 Grundlagen

Definition Es seien V und W zwei Vektorräume über demselben Körper K. Eine Abbildung $\varphi\colon V \to W$ heißt **linear**, wenn sie die folgenden beiden Eigenschaften besitzt:

(ADD) Für alle $\mathbf{v}_1, \mathbf{v}_2 \in V$ gilt $\varphi(\mathbf{v}_1 + \mathbf{v}_2) = \varphi(\mathbf{v}_1) + \varphi(\mathbf{v}_2)$. (*Additivität*)

(H) Für alle $\mathbf{v} \in V$ und $c \in K$ gilt $\varphi(c\mathbf{v}) = c\varphi(\mathbf{v})$. (*Homogenität*)

Man kann auch wieder beide Eigenschaften zu einer zusammenfassen. Genau dann ist eine Abbildung $\varphi\colon V \to W$ linear, wenn folgendes gilt:

(LIN) Für alle $\mathbf{v}_1, \mathbf{v}_2 \in V$ und alle $c_1, c_2 \in K$ gilt

$$\varphi(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\varphi(\mathbf{v}_1) + c_2\varphi(\mathbf{v}_2) \qquad \textit{(Linearität)}$$

Durch wiederholten Anwenden von (LIN) sieht man, dass eine lineare Abbildung φ immer

$$\varphi\big(\sum\nolimits_{i=1}^{k} c_i\mathbf{v}_i\big) = \sum_{i=1}^{k} c_i\varphi(\mathbf{v}_i)$$

für alle $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$ und $c_1, \ldots, c_k \in K$ erfüllt. Daraus folgt insbesondere für alle $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$ die Gleichheit

$$\varphi\big(\mathrm{Lin}(\mathbf{v}_1, \ldots, \mathbf{v}_k)\big) = \mathrm{Lin}\big(\varphi(\mathbf{v}_1), \ldots, \varphi(\mathbf{v}_k)\big).$$

16.1 Beispiele (1) Für $V = \mathbb{R}^n$ und $W = \mathbb{R}^m$ bestimmt jede Matrix $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ die lineare Abbildung $\varphi_A(\vec{x}) = A\vec{x}$, die wir in Kapitel 9 untersucht haben. Dasselbe ist über jedem Körper K sinnvoll.

(2) Für zwei K-Vektorräume V und W ist die **Nullabbildung**, $\mathbf{v} \mapsto \mathbf{0}$, die jeden Vektor in V auf den Nullvektor in W abbildet, eine lineare Abbildung.

(3) Ist V ein Vektorraum und $a \in K$, dann ist

$$\varphi\colon \begin{cases} V & \to & V \\ \mathbf{v} & \mapsto & a\mathbf{v} \end{cases}$$

eine lineare Abbildung. Für $a = 1$ ist das die **Identität** id_V.

(4) Für zwei K-Vektorräume V und W ist die **Nullabbildung**, $\mathbf{v} \mapsto \mathbf{0}$, die jeden Vektor in V auf den Nullvektor in W abbildet, eine lineare Abbildung.

(5) Für zwei K-Vektorräume V und W ist die **Projektion**

$$\pi\colon \begin{cases} V \times W & \to & V \\ (\mathbf{v}, \mathbf{w}) & \mapsto & \mathbf{v} \end{cases}$$

vom direkten Produkt auf den ersten Faktor linear, ebenso die entsprechende Projektion auf den zweiten Faktor.

(6) Es sei $V = C^\infty(\mathbb{R})$ der Raum aller beliebig oft differenzierbaren Funktionen $\mathbb{R} \to \mathbb{R}$. Der **Differentialoperator**

$$\delta\colon \begin{cases} V & \to & V \\ f & \mapsto & f' \end{cases}$$

der einer Funktion ihre Ableitung zuordnet, ist linear. Bekanntlich gelten nämlich die Ableitungsregeln $(f+g)' = f' + g'$ und $(cf)' = cf'$ für $f, g \in V$ und $c \in \mathbb{R}$, und das ist genau die Linearität von δ. ◇

16.2 Kern und Bild

16.2 Lemma *Jede lineare Abbildung bildet Nullvektor auf Nullvektor ab.*

Beweis. Denn ist $\varphi\colon V \to W$ linear, dann gilt

$$\varphi(\mathbf{0}) = \varphi(0 \cdot \mathbf{0}) = 0 \cdot \varphi(\mathbf{0}) = \mathbf{0}.$$ ■

16.3 Lemma *Es seien V und W zwei K-Vektorräume, $U \subset V$ und $X \subset W$ lineare Unterräume, und sei $\varphi\colon V \to W$ eine lineare Abbildung. Dann sind die Bildmenge $\varphi(U) \subset W$ und die Urbildmenge $\varphi^{-1}(X) \subset V$ ebenfalls lineare Unterräume.*

Beweis. Genauso wie Lemma 9.14. ■

Definition Es seien V und W zwei K-Vektorräume und sei $\varphi\colon V \to W$ eine lineare Abbildung.

(1) Der **Kern** von φ ist der lineare Unterraum

$$\operatorname{Kern}(\varphi) = \{\mathbf{v} \in V \mid \varphi(\mathbf{v}) = \mathbf{0}\} \subset V.$$

(2) Das **Bild** von φ ist der lineare Unterraum

$$\operatorname{Bild}(\varphi) = \varphi(V) \subset W.$$

Der Kern ist die Urbildmenge $\operatorname{Kern}(\varphi) = \varphi^{-1}(\{\mathbf{0}\})$ (und damit nach Lemma 16.3 ein linearer Unterraum). Den Kern betrachtet man im Unterschied zum Bild in aller Regel nur bei linearen Abbildungen.

16.4 Lemma *Es seien V und W zwei K-Vektorräume und sei $\varphi\colon V \to W$ eine lineare Abbildung.*

(1) Genau dann ist φ injektiv, wenn $\operatorname{Kern}(\varphi) = \{\mathbf{0}\}$ gilt.

(2) Genau dann ist φ surjektiv, wenn $\operatorname{Bild}(\varphi) = W$ gilt.

Beweis. (2) ist trivial, denn das ist die Definition der Surjektivität.

(1) geht genauso wie in Lemma 9.15, aber es schadet auch nichts, diesen Beweis noch einmal zu wiederholen: Es sei φ injektiv und sei $\mathbf{v} \in \operatorname{Kern}(\varphi)$. Dann folgt $\varphi(\mathbf{v}) = \mathbf{0} = \varphi(\mathbf{0})$ und damit $\mathbf{v} = \mathbf{0}$, weil φ injektiv ist. Also gilt $\operatorname{Kern}(\varphi) = \{\mathbf{0}\}$.

Es gelte umgekehrt $\operatorname{Kern}(\varphi) = \{\mathbf{0}\}$. Es seien $\mathbf{v}_1, \mathbf{v}_2 \in V$ mit $\varphi(\mathbf{v}_1) = \varphi(\mathbf{v}_2)$ gegeben. Dann folgt $\varphi(\mathbf{v}_1 - \mathbf{v}_2) = \varphi(\mathbf{v}_1) - \varphi(\mathbf{v}_2) = \mathbf{0}$, also $\mathbf{v}_1 - \mathbf{v}_2 \in \operatorname{Kern}(\varphi)$ und somit $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$, was $\mathbf{v}_1 = \mathbf{v}_2$ bedeutet. Damit ist gezeigt, dass φ injektiv ist. ■

16.5 Beispiel Der Kern des Differentialoperators

$$\delta\colon \begin{cases} C^\infty(\mathbb{R}) & \to & C^\infty(\mathbb{R}) \\ f & \mapsto & f' \end{cases}$$

besteht aus allen Funktionen, deren Ableitung die Nullfunktion ist. Bekanntlich sind das gerade die konstanten Funktionen. Der Differentialoperator ist also nicht injektiv. Das entspricht der Tatsache, dass die Funktionen $f + a$ und $f + b$ für $f \in C^\infty(\mathbb{R})$ und $a, b \in \mathbb{R}$ dieselbe Ableitung haben. (Der Differentialoperator ist surjektiv nach dem Hauptsatz der Differential- und Integralrechnung; für jedes $f \in C^\infty(\mathbb{R})$ gilt $f = g'$ mit $g(x) = \int_a^x f(t)dt$, mit $a \in \mathbb{R}$ beliebig.) ◇

Definition Die Dimension von Bild(φ) heißt der **Rang**[1] von φ.

16.6 Satz (Dimensionsformel für lineare Abbildungen) *Es sei $\varphi\colon V \to W$ eine lineare Abbildung zwischen endlichdimensionalen Vektorräumen. Dann gilt*

$$\dim(V) = \operatorname{Rang}(\varphi) + \dim\big(\operatorname{Kern}(\varphi)\big).$$

Insbesondere gilt $\dim(\operatorname{Bild}(\varphi)) \leqslant \dim(V)$.

Beweis. Das kann man völlig analog zu Satz 9.17 beweisen. Wir geben später auch noch einen anderen Beweis. ■

16.3 Isomorphismen

16.7 Lemma *Es sei $\varphi\colon V \to W$ eine bijektive lineare Abbildung. Dann ist auch die Umkehrabbildung $\varphi^{-1}\colon W \to V$ linear.*

Beweis. Genauso wie Lemma 10.6. ■

Definition Es seien V und W zwei K-Vektorräume. Eine bijektive lineare Abbildung von V nach W heißt ein **Isomorphismus** zwischen V und W. Die Vektorräume V und W heißen **isomorph**, wenn ein Isomorphismus $V \to W$ existiert.

16.8 Beispiele (1) Ist $\vec{w} \in \mathbb{R}^n$, $\vec{w} \neq \vec{0}$ und $L = \operatorname{Lin}(\vec{w})$ die Ursprungsgerade mit Richtungsvektor $\vec{w}$, dann ist $\varphi\colon \mathbb{R} \to L$, $c \mapsto c\vec{w}$, ein Isomorphismus. Zwei Geraden können verschieden sein, sie können sogar in verschiedenen Räumen $\mathbb{R}^m$ und $\mathbb{R}^n$ leben. Sie sind aber trotzdem als Vektorräume isomorph. Dasselbe gilt über jedem Körper K.

(2) Die Vektorräume $\mathbb{R}[x]_{\leqslant d}$ und $\mathbb{R}^{d+1}$ sind isomorph, denn die Abbildung

$$\varphi\colon \begin{cases} \mathbb{R}[x]_{\leqslant d} & \to & \mathbb{R}^{d+1} \\ a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0 & \mapsto & (a_0, \ldots, a_d) \end{cases}$$

die einem Polynom vom Grad höchstens d den Vektor seiner Koeffizienten zuordnet, ist ein Isomorphismus. ◇

Isomorphismen erhalten die ganze Struktur eines Vektorraums. Alles was sich rein in der Sprache der Vektorräume sagen lässt, ist für isomorphe Vektorräume gleich. Ein Beispiel dafür ist die folgende Aussage.

[1] Die Dimension von Kern(φ) wird manchmal »Defekt« von φ genannt. Wir werden diese Terminologie aber nicht verwenden.

16.9 Lemma *Sei $\varphi\colon V \to W$ ein Isomorphismus zwischen Vektorräumen. Ist V endlichdimensional und $\mathbf{v}_1, \dots, \mathbf{v}_n$ eine Basis von V, dann ist $\varphi(\mathbf{v}_1), \dots, \varphi(\mathbf{v}_n)$ eine Basis von W. Insbesondere haben isomorphe Vektorräume dieselbe Dimension.*

Beweis. Denn ist $\mathbf{w} \in W$, dann gibt es $x_1, \dots, x_n$ mit $\varphi^{-1}(\mathbf{w}) = \sum_{i=1}^n x_i \mathbf{v}_i$, also

$$\mathbf{w} = \varphi(\varphi^{-1}(\mathbf{w})) = \varphi\Big(\sum\nolimits_{i=1}^n x_i \mathbf{v}_i\Big) = \sum\nolimits_{i=1}^n x_i \varphi(\mathbf{v}_i) \in \operatorname{Lin}\big(\varphi(\mathbf{v}_1), \dots, \varphi(\mathbf{v}_n)\big).$$

Ist $\sum_{i=1}^n c_i \varphi(\mathbf{v}_i) = \mathbf{0}$ eine lineare Relation, dann folgt auch

$$\mathbf{0} = \varphi^{-1}(\mathbf{0}) = \varphi^{-1}\Big(\sum\nolimits_{i=1}^n c_i \varphi(\mathbf{v}_i)\Big) = \sum\nolimits_{i=1}^n c_i \mathbf{v}_i,$$

also $c_1 = \dots = c_n = 0$, weil $\mathbf{v}_1, \dots, \mathbf{v}_n$ linear unabhängig sind. Damit ist gezeigt, dass das System $\varphi(\mathbf{v}_1), \dots, \varphi(\mathbf{v}_n)$ eine Basis von W ist. ∎

16.4 Das Prinzip der linearen Ausdehnung

16.10 Satz (Lineare Ausdehnung) *Es seien V und W zwei K-Vektorräume. Ist V endlichdimensional und $\mathbf{v}_1, \dots, \mathbf{v}_n$ eine Basis von V, dann gibt es zu jeder Wahl von Vektoren $\mathbf{w}_1, \dots, \mathbf{w}_n \in W$ genau eine lineare Abbildung $\varphi\colon V \to W$ mit*

$$\varphi(\mathbf{v}_i) = \mathbf{w}_i \quad \textit{für } i = 1, \dots, n.$$

Beweis. Existenz. Da $\mathbf{v}_1, \dots, \mathbf{v}_n$ eine Basis von V bilden, hat jeder Vektor $\mathbf{v} \in V$ genau eine Darstellung $\mathbf{v} = \sum_{i=1}^n x_i \mathbf{v}_i$. Deshalb definieren wir φ einfach durch

$$\varphi(\mathbf{v}) = \sum_{i=1}^n x_i \mathbf{w}_i.$$

Offenbar hat φ die gewünschte Eigenschaft $\varphi(\mathbf{v}_i) = \mathbf{w}_i$. Wir müssen aber noch nachrechnen, dass φ linear ist: Für $\mathbf{v} = \sum_{i=1}^n x_i \mathbf{v}_i$, $\mathbf{v}' = \sum_{i=1}^n x_i' \mathbf{v}_i$ und $c, c' \in K$ gilt die Gleichheit

$$\begin{aligned}\varphi(c\mathbf{v} + c'\mathbf{v}') &= \varphi\Big(\sum\nolimits_{i=1}^n (cx_i + c'x_i')\mathbf{v}_i\Big) = \sum_{i=1}^n (cx_i + c'x_i')\mathbf{w}_i \\ &= c \sum_{i=1}^n x_i \mathbf{w}_i + c' \sum_{i=1}^n x_i' \mathbf{w}_i = c\varphi(\mathbf{v}) + c'\varphi(\mathbf{v}').\end{aligned}$$

Eindeutigkeit. Sind φ und ψ zwei lineare Abbildungen $V \to W$ mit der Eigenschaft

$\varphi(\mathbf{v}_i) = \mathbf{w}_i$ für $i = 1, \dots, n$, dann gilt für jedes $\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i$

$$\varphi(\mathbf{v}) = \varphi\left(\sum\nolimits_{i=1}^{n} x_i \mathbf{v}_i\right) = \sum_{i=1}^{n} x_i \varphi(\mathbf{v}_i) = \sum_{i=1}^{n} x_i \mathbf{w}_i$$
$$= \sum_{i=1}^{n} x_i \psi(\mathbf{v}_i) = \psi\left(\sum\nolimits_{i=1}^{n} x_i \mathbf{v}_i\right) = \psi(\mathbf{v}).$$

Also gilt $\varphi = \psi$. ∎

Beispiele für lineare Ausdehnung haben wir schon in Kapitel 9 in $\mathbb{R}^2$ gesehen. Wir werden das im nächsten Kapitel verwenden, um lineare Abbildungen wieder mit Matrizen in Verbindung zu bringen.

16.5 Räume linearer Abbildungen

Als letztes untersuchen wir noch die Struktur der Menge aller linearen Abbildungen zwischen gegebenen Vektorräumen.

16.11 Satz *Es seien V, W, X drei K-Vektorräume und seien*

$$\varphi\colon V \to W, \quad \varphi'\colon V \to W, \quad \psi\colon W \to X, \quad \psi'\colon W \to X$$

lineare Abbildungen. Sei $a \in K$. Folgende Abbildungen sind ebenfalls linear:

(1) $(\varphi + \varphi')\colon V \to W$ *definiert durch* $(\varphi + \varphi')(\mathbf{v}) = \varphi(\mathbf{v}) + \varphi'(\mathbf{v})$ *für* $\mathbf{v} \in V$.

(2) $(a\varphi)\colon V \to W$ *definiert durch* $(a\varphi)(\mathbf{v}) = a \cdot \varphi(\mathbf{v})$ *für* $\mathbf{v} \in V$.

(3) $(\psi \circ \varphi)\colon V \to X$ *definiert durch* $(\psi \circ \varphi)(\mathbf{v}) = \psi(\varphi(\mathbf{v}))$ *für* $\mathbf{v} \in V$.

Außerdem gelten $\psi \circ (\varphi + \varphi') = \psi \circ \varphi + \psi \circ \varphi'$ *und* $(\psi + \psi') \circ \varphi = \psi \circ \varphi + \psi' \circ \varphi$.

Beweis. Es hilft nichts, wir müssen die Linearität aller dieser Abbildungen nachrechnen: Es seien $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v} \in V$ und $c_1, c_2 \in K$.

(1) Es gilt $(\varphi+\varphi')(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = \varphi(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) + \varphi'(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\varphi(\mathbf{v}_1) + c_2\varphi(\mathbf{v}_2) + c_1\varphi'(\mathbf{v}_1) + c_2\varphi'(\mathbf{v}_2) = c_1(\varphi + \varphi')(\mathbf{v}_1) + c_2(\varphi + \varphi')(\mathbf{v}_2)$.

(2) Es gilt $(a\varphi)(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = a \cdot (\varphi(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)) = a \cdot (c_1\varphi(\mathbf{v}_1) + c_2\varphi(\mathbf{v}_2)) = ac_1\varphi(\mathbf{v}_1) + ac_2\varphi(\mathbf{v}_2) = c_1(a\varphi)(\mathbf{v}_1) + c_2(a\varphi)(\mathbf{v}_2)$.

(3) Es gilt $(\psi\circ\varphi)(c_1\mathbf{v}_1+c_2\mathbf{v}_2) = \psi(\varphi((c_1\mathbf{v}_1+c_2\mathbf{v}_2))) = \psi(c_1\varphi(\mathbf{v}_1)+c_2\varphi(\mathbf{v}_2)) = c_1\psi(\varphi(\mathbf{v}_1)) + c_2\psi(\varphi(\mathbf{v}_2)) = c_1(\psi \circ \varphi)(\mathbf{v}_1) + c_2(\psi \circ \varphi)(\mathbf{v}_2)$.

Außerdem $(\psi \circ (\varphi+\varphi'))(\mathbf{v}) = \psi(\varphi(\mathbf{v}) + \varphi'(\mathbf{v})) = \psi(\varphi(\mathbf{v})) + \psi(\varphi'(\mathbf{v})) = (\psi \circ \varphi+\psi\circ\varphi')(\mathbf{v})$ und $((\psi+\psi')\circ\varphi)(\mathbf{v}) = \psi(\varphi(\mathbf{v}))+\psi'(\varphi(\mathbf{v})) = (\psi\circ\varphi+\psi'\circ\varphi)(\mathbf{v})$. ∎

Für die Menge aller linearen Abbildungen zwischen zwei Vektorräumen V und W schreiben wir

$$\mathcal{L}(V,W) = \{\varphi\colon V \to W \mid \varphi \text{ ist linear}\}.$$

16.12 Korollar *Es seien V und W zwei Vektorräume.*

(1) Die Menge $\mathcal{L}(V,W)$ bildet, mit Addition und Skalarmultiplikation wie in Satz 16.11, einen Vektorraum über K.

(2) Die Menge $\mathcal{L}(V,V)$ bildet, mit der Addition wie in Satz 16.11 und der Verknüpfung $\circ$ als Multiplikation, einen Ring. Die Eins in $\mathcal{L}(V,V)$ ist id_V.

Üblich sind auch die Notationen $\mathrm{Hom}(V,W) = \mathcal{L}(V,W)$ (Homomorphismen) und $\mathrm{End}(V) = \mathcal{L}(V,V)$ (Endomorphismen).

Beweis. Die nötigen Eigenschaften sind in Satz 16.11 nachgewiesen. ∎

Definition Eine lineare Abbildung $V \to V$ eines Vektorraums in sich heißt auch *Endomorphismus* von V. Der Ring $\mathcal{L}(V,V)$ ist der **Endomorphismenring**.

16.13 Korollar *Zwei endlichdimensionale Vektorräume über K sind genau dann isomorph, wenn sie dieselbe Dimension haben.*

Beweis. Dass isomorphe Vektorräume gleiche Dimension haben, haben wir schon bewiesen (Lemma 16.9). Seien umgekehrt V und W zwei Vektorräume derselben Dimension n mit Basen $\mathbf{v}_1,\dots,\mathbf{v}_n$ bzw. $\mathbf{w}_1,\dots,\mathbf{w}_n$. Nach dem Prinzip der linearen Ausdehnung gibt es also eine lineare Abbildung $\varphi\colon V \to W$ mit $\varphi(\mathbf{v}_i) = \mathbf{w}_i$ und eine lineare Abbildung $\psi\colon W \to V$ mit $\psi(\mathbf{w}_i) = \mathbf{v}_i$ (für $i = 1,\dots,n$.) Es folgt $\psi(\varphi(\mathbf{v}_i)) = \psi(\mathbf{w}_i) = \mathbf{v}_i$ für $i = 1,\dots,n$ und $\psi \circ \varphi$ ist linear nach Satz 16.11. Nach der Eindeutigkeit der linearen Ausdehnung ist aber die Identität id_V die einzige lineare Abbildung $V \to V$ mit $\mathbf{v}_i \mapsto \mathbf{v}_i$ für $i = 1,\dots,n$. Es folgt $\psi \circ \varphi = \mathrm{id}_V$ und, mit dem gleichen Argument, auch $\varphi \circ \psi = \mathrm{id}_W$. Also ist φ ein Isomorphismus. ∎

16.14 Korollar *Sei V ein Vektorraum über K und seien $\mathbf{v}_1,\dots,\mathbf{v}_n \in V$. Betrachte die lineare Abbildung*

$$\varphi\colon \left\{ \begin{array}{ccc} K^n & \to & V \\ (x_1,\dots,x_n) & \mapsto & \sum_{i=1}^n x_i\mathbf{v}_i \end{array} \right. .$$

Das Bild von φ ist der lineare Unterraum $\mathrm{Lin}(\mathbf{v}_1,\dots,\mathbf{v}_n)$. Ferner gelten:

(1) Genau dann ist φ injektiv, wenn $\mathbf{v}_1,\dots,\mathbf{v}_n$ linear unabhängig sind.

(2) Genau dann ist φ surjektiv, wenn $\mathbf{v}_1,\dots,\mathbf{v}_n$ ein Erzeugendensystem sind.

(3) Genau dann ist φ ein Isomorphismus, wenn $\mathbf{v}_1,\dots,\mathbf{v}_n$ eine Basis sind.

Beweis. Nach dem Prinzip der linearen Ausdehnung gibt es eine eindeutig bestimmte Abbildung $K^n \to V$ mit $\vec{e}_i \mapsto \mathbf{v}_i$ für $i = 1, \dots, n$. Das ist gerade die angegebene Abbildung φ. Sie ist also linear. Das Bild von φ ist offenbar $\mathrm{Lin}(\mathbf{v}_1, \dots, \mathbf{v}_n)$, was Aussage (2) impliziert. Der Kern von φ besteht aus allen $\mathbf{x} \in K^n$ mit $\sum_{i=1}^n x_i \mathbf{v}_i = \mathbf{0}$. Das zeigt Aussage (1). Aussage (3) folgt aus (1) und (2). ∎

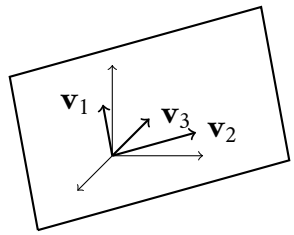
16.15 Beispiel Wir betrachten die drei Vektoren

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \ , \ \vec{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \ , \ \vec{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

in $\mathbb{R}^3$. Nach dem Prinzip der linearen Ausdehnung gibt es eine lineare Abbildung mit $\vec{e}_i \mapsto \vec{v}_i$ für $i = 1, 2, 3$, nämlich die Abbildung

$$\varphi \colon \mathbb{R}^3 \to \mathbb{R}^3, \ \vec{x} \mapsto \sum_{i=1}^{3} x_i \vec{v}_i = \begin{pmatrix} x_1 + x_2 + x_3 \\ 2x_1 + x_3 \\ 3x_1 - x_2 + x_3 \end{pmatrix}.$$

Das Bild von φ ist der lineare Unterraum $E = \mathrm{Lin}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ von $\mathbb{R}^3$. Hier sind $\vec{v}_1, \vec{v}_2, \vec{v}_3$ linear abhängig, nämlich $\vec{v}_1 + \vec{v}_2 - 2\vec{v}_3 = \vec{0}$, und $\vec{v}_1, \vec{v}_2$ bilden bereits eine Basis der Ebene E.



Der Kern besteht aus allen linearen Relationen zwischen $\vec{v}_1, \vec{v}_2, \vec{v}_3$. Das sind die Lösungen des homogenen linearen Gleichungssystems mit Koeffizientenvektoren $\vec{v}_1, \vec{v}_2, \vec{v}_3$. Da das Bild von φ zweidimensional ist, muss dieser Kern nach der Dimensionsformel eindimensional sein. Die lineare Abhängigkeit $\vec{v}_1 + \vec{v}_2 - 2\vec{v}_3 = \vec{0}$ sagt gerade, dass der Vektor $(1, 1, -2)$ im Kern von φ liegt. Da der Kern eindimensional ist, wird er also von diesem Vektor aufgespannt. ◇

16.16 Korollar *Jeder K-Vektorraum der Dimension n ist isomorph zu K^n.*

Beweis. Das folgt wegen $\dim(K^n) = n$ sofort aus Kor. 16.13. Expliziter steht es in Kor. 16.14(3): Jede Wahl einer Basis entspricht einem Isomorphismus mit K^n. ∎

17 Koordinaten und darstellende Matrizen

Die linearen Abbildungen $K^n \to K^m$ entsprechen den $m \times n$-Matrizen: Jede solche lineare Abbildung ist von der Form $\varphi_A \colon \vec{x} \mapsto A\vec{x}$ für $A \in \mathrm{Mat}_{m\times n}(K)$, wie wir in Kapitel 9 gesehen haben. Wenn man dasselbe für beliebige endlichdimensionale Vektorräume machen will, muss man vorher Basen wählen.

17.1 Koordinatenvektoren

Es sei V ein Vektorraum der Dimension n über einem Körper K und sei $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ eine Basis von V. Dann hat jeder Vektor $\mathbf{v} \in V$ eine Darstellung

$$\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$$

und der Spaltenvektor

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in K^n$$

ist der **Koordinatenvektor** von $\mathbf{v}$ bezüglich der Basis $\mathcal{B}$. Wenn wir dafür eine Notation brauchen, schreiben wir

$$\varkappa_{\mathcal{B}}(\mathbf{v}).$$

für den Koordinatenvektor $\vec{x}$ von $\mathbf{v}$. Die Koordinatenvektoren der Basisvektoren $\vec{v}_1, \ldots, \vec{v}_n$ selbst sind die Einheitsvektoren $\vec{e}_1, \ldots, \vec{e}_n$.

17.1 Korollar *Für jede Basis $\mathcal{B}$ des endlichdimensionalen Vektorraums V ist die Abbildung $\varkappa_{\mathcal{B}} \colon V \to K^n$ ein Isomorphismus.*

Beweis. Sei $n = \dim(V)$ und $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Nach Kor. 16.14 ist die lineare Abbildung $\varphi \colon K^n \to V$, $\vec{x} \mapsto \sum_{i=1}^n x_i\mathbf{v}_i$ ein Isomorphismus. Die Abbildung $\varkappa_{\mathcal{B}}$ ist per Definition gerade die Umkehrabbildung von φ. ∎

Nachdem wir eine Basis $\mathcal{B}$ von V gewählt haben, können wir nun statt mit den abstrakten Vektoren in V mit den Spaltenvektoren $\varkappa_{\mathcal{B}}(\mathbf{v})$ für $\mathbf{v} \in V$ rechnen.

17.2 Darstellende Matrizen

Definition Seien V und W zwei endlichdimensionale Vektorräume der Dimension n bzw. m. Gegeben seien eine Basis $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ von V und eine Basis $C = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ von W. Für jede lineare Abbildung $\varphi\colon V \to W$ gibt es dann eindeutige Darstellungen

$$\varphi(\mathbf{v}_j) = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i, \quad (j = 1, \ldots, n).$$

Die $m \times n$-Matrix mit Einträgen a_{ij} heißt die **darstellende Matrix** von φ bezüglich der gewählten Basen $\mathcal{B}$ und C. Wir schreiben dafür

$$M_C^{\mathcal{B}}(\varphi).$$

Die Spalten der darstellenden Matrix entsprechen wieder den Bildern der Basisvektoren. Die darstellende Matrix einer linearen Abbildung $\varphi\colon V \to W$ bezüglich gegebener Basen wie oben bestimmt man also praktisch wie folgt:

Für jeden Index $j = 1, \ldots, n$, finde die Darstellung des Bildvektors $\varphi(\mathbf{v}_j)$ in der Basis $C = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$. Das gibt die j-te Spalte von $M_C^{\mathcal{B}}(\varphi)$.

17.2 Beispiele (1) Sei V ein Vektorraum der Dimension n. Für $a \in K$ hat die Streckung um den Faktor a

$$\varphi\colon \begin{cases} V & \to & V \\ \mathbf{v} & \mapsto & a\mathbf{v} \end{cases}$$

bezüglich jeder Basis $\mathcal{A} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ von V die darstellende Matrix

$$M_{\mathcal{A}}^{\mathcal{A}}(\varphi) = \begin{pmatrix} a & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a \end{pmatrix} = a \cdot \mathbb{1}_n.$$

Insbesondere wird die Identität id_V durch die Einheitsmatrix $\mathbb{1}_n$ dargestellt; ebenso die Nullabbildung durch die Nullmatrix.

(2) Die darstellende Matrix hängt im allgemeinen aber von der Wahl der Basis ab. Betrachten wir die lineare Abbildung

$$\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} .$$

Das ist die Spiegelung an der Diagonalen $\mathrm{Lin}(\vec{e}_1 + \vec{e}_2)$ (Beispiel 9.5). Sie hat in der Standardbasis $\mathcal{E} = (\vec{e}_1, \vec{e}_2)$ die darstellende Matrix

$$M_{\mathcal{E}}^{\mathcal{E}}(\varphi) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Wir können aber auch die Basis $\mathcal{B} = (\vec{v}_1, \vec{v}_2)$ mit

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \ , \ \vec{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

benutzen. Dann gelten $\varphi(\vec{v}_1) = \vec{v}_1$ und $\varphi(\vec{v}_2) = -\vec{v}_2$. Die darstellende Matrix in dieser Basis ist deshalb

$$M_{\mathcal{B}}^{\mathcal{B}}(\varphi) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ .$$

In der Standardbasis $\mathcal{E}$ beschreibt diese Matrix die Spiegelung an der waagerechten Achse $\mathrm{Lin}(\vec{e}_1)$. Die neue Basis $\mathcal{B}$ ist so gewählt, dass der erste Basisvektor die Spiegelungsgerade aufspannt und der zweite dazu senkrecht steht. Wir haben das Koordinatensystem rotiert (und gestreckt), ein Beispiel für einen *Basiswechsel*. Wie sich die darstellende Matrix dabei allgemein verändert, untersuchen wir in Kürze.

(3) Wir betrachten den Differentialoperator

$$\delta \colon \mathbb{R}[x]_{\leqslant d} \to \mathbb{R}[x]_{\leqslant d},\ f \mapsto f'$$

bezüglich der Monombasis

$$\mathcal{B} = (\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d) = (1, x, x^2, \ldots, x^d).$$

Nach den bekannten Ableitungsregeln gilt $\delta(x^i) = ix^{i-1}$ und $\delta(1) = 0$. In der Basis $\mathcal{B}$ bedeutet das also $\delta(\mathbf{v}_0) = \mathbf{0}$ und $\delta(\mathbf{v}_i) = i\mathbf{v}_{i-1}$ für $i = 1, \ldots, d$. Die darstellende Matrix ist deshalb

$$M_{\mathcal{B}}^{\mathcal{B}}(\delta) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & d-1 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & d \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 \end{pmatrix}. \qquad \diamond$$

Wir können darstellende Matrizen etwas abstrakter auch so verstehen: Sind V und W zwei Vektrräume der Dimensionen $n = \dim(V)$ und $m = \dim(W)$ mit Basen $\mathcal{B}$ bzw. $\mathcal{C}$, dann bekommen wir nach Kor. 17.1 Isomorphismen $\varkappa_{\mathcal{B}} \colon V \to K^n$ und $\varkappa_C \colon W \to K^m$. Ist nun $\varphi \colon V \to W$ eine lineare Abbildung, dann bestimmt die darstellende Matrix $A = M_C^{\mathcal{B}}(\varphi)$ die lineare Abbildung $\varphi_A \colon K^n \to K^m$. Per Definition stimmt diese Abbildung mit der Abbildung φ *in den gewählten Koordinaten* überein. Das kann man in einem **kommutierenden Diagramm** bildlich darstellen:

$$\begin{array}{ccc} V & \xrightarrow{\varphi} & W \\ {\scriptstyle \varkappa_{\mathcal{B}}}\downarrow & & \downarrow{\scriptstyle \varkappa_C} \\ K^n & \xrightarrow[\varphi_A]{} & K^m \end{array}$$

Kommutierend bedeutet dabei $\varphi_A \circ \varkappa_{\mathcal{B}} = \varkappa_C \circ \varphi$ oder, äquivalent, $\varphi = \varkappa_C^{-1} \circ \varphi_A \circ \varkappa_{\mathcal{B}}$.

17.3 Eigenschaften darstellender Matrizen

17.3 Satz *Es seien V und W zwei endlichdimensionale Vektorräume. Es sei $n = \dim(V)$ und $m = \dim(W)$. Für jede Wahl von Basen $\mathcal{B}$ von V und $\mathcal{C}$ von W wie oben ist die Zuordnung*

$$M_C^{\mathcal{B}} \colon \left\{ \begin{array}{ccc} \mathcal{L}(V,W) & \to & \mathrm{Mat}_{m\times n}(K) \\ \varphi & \mapsto & M_C^{\mathcal{B}}(\varphi) \end{array} \right.$$

ein Isomorphismus von Vektorräumen.

Das heißt also insbesondere: Zu jeder linearen Abbildung $\varphi \colon V \to W$ gehört, bei fest gewählten Basen, genau eine darstellende Matrix, und umgekehrt ist jede $m \times n$-Matrix die darstellende Matrix genau einer linearen Abbildung.

Die Aussage soll auch sinnvoll sein, wenn V oder W der Nullraum ist. In diesem Fall ist also $m = 0$ oder $n = 0$, und wir definieren $\mathrm{Mat}_{m\times n}(K)$ als den Nullraum (der also eine »Nullmatrix«, aber mit 0 Zeilen oder 0 Spalten enthält).

Beweis. Als erstes müssen wir zeigen, dass $M_C^{\mathcal{B}}$ linear ist. Das heißt also für alle $\varphi_1, \varphi_2 \in \mathcal{L}(V,W)$ und $a \in K$ müssen

$$M_C^{\mathcal{B}}(\varphi_1 + \varphi_2) = M_C^{\mathcal{B}}(\varphi_1) + M_C^{\mathcal{B}}(\varphi_2) \quad \text{und} \quad M_C^{\mathcal{B}}(a\varphi_1) = a \cdot M_C^{\mathcal{B}}(\varphi_1)$$

gelten. Beides rechnet man leicht nach.

Die Surjektivität ergibt sich aus dem Prinzip der linearen Ausdehnung. Denn sei $\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$, $\mathcal{C} = (\mathbf{w}_1, \dots, \mathbf{w}_m)$ und sei eine $m \times n$-Matrix A gegeben. Dann setzen wir

$$\mathbf{w}'_j = \sum_{i=1}^{m} a_{ij}\mathbf{w}_i, \quad (j = 1, \dots, n)$$

und wenden Satz 16.10 auf die Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ und die Vektoren $\mathbf{w}'_1, \dots, \mathbf{w}'_n$ an. Wir erhalten so eine lineare Abbildung $\varphi\colon V \to W$ mit $\varphi(\mathbf{v}_i) = \mathbf{w}'_i$ für $i = 1, \dots, n$. Es gilt dann $M_{\mathcal{C}}^{\mathcal{B}}(\varphi) = A$. Nach Satz 16.10 ist φ außerdem eindeutig, was die Injektivität von $M_{\mathcal{C}}^{\mathcal{B}}$ beweist. ∎

17.4 Korollar *Sind V und W endlichdimensionale Vektorräume, dann ist auch der Vektorraum $\mathcal{L}(V, W)$ endlichdimensional, und es gilt*

$$\dim(\mathcal{L}(V, W)) = \dim(V) \cdot \dim(W).$$

Beweis. Ist $n = \dim(V)$ und $m = \dim(W)$, dann ist $\mathcal{L}(V, W)$ nach Satz 17.3 isomorph zum Raum $\mathrm{Mat}_{m\times n}(K)$. Dieser hat die Dimension mn. ∎

17.5 Satz *Es seien U, V und W drei endlichdimensionale Vektorräume, jeweils mit Basen $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$. Seien $\psi\colon U \to V$ und $\varphi\colon V \to W$ zwei lineare Abbildungen. Dann gilt*

$$M_{\mathcal{C}}^{\mathcal{A}}(\varphi \circ \psi) = M_{\mathcal{C}}^{\mathcal{B}}(\varphi) \cdot M_{\mathcal{B}}^{\mathcal{A}}(\psi).$$

In Worten: Der Komposition von linearen Abbildungen entspricht das Produkt der darstellenden Matrizen.

Erster Beweis. Ist $A = M_{\mathcal{C}}^{\mathcal{B}}(\varphi)$ und $B = M_{\mathcal{B}}^{\mathcal{A}}(\psi)$, dann haben wir in Satz 9.12 bereits die Gleichheit $\varphi_{AB} = \varphi_A \circ \varphi_B$ nachgerechnet. Nun gelten $\psi = \kappa_{\mathcal{B}}^{-1} \circ \varphi_B \circ \kappa_{\mathcal{A}}$ und $\varphi = \kappa_{\mathcal{C}}^{-1} \circ \varphi_A \circ \kappa_{\mathcal{B}}$ und damit

$$\varphi \circ \psi = (\kappa_{\mathcal{C}}^{-1} \circ \varphi_A \circ \kappa_{\mathcal{B}}) \circ (\kappa_{\mathcal{B}}^{-1} \circ \varphi_B \circ \kappa_{\mathcal{A}}) = \kappa_{\mathcal{C}}^{-1} \circ (\varphi_A \circ \varphi_B) \circ \kappa_{\mathcal{A}} = \kappa_{\mathcal{C}}^{-1} \circ \varphi_{AB} \circ \kappa_{\mathcal{A}}.$$

Das ist die Behauptung. ∎

Zweiter Beweis. Man kann das auch noch einmal vollständig nachrechnen: Seien $\mathcal{A} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$, $\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_m)$ und $\mathcal{C} = (\mathbf{w}_1, \dots, \mathbf{w}_l)$. Setze $A = M_{\mathcal{C}}^{\mathcal{B}}(\varphi) \in \mathrm{Mat}_{l\times m}(K)$, $B = M_{\mathcal{B}}^{\mathcal{A}}(\psi) \in \mathrm{Mat}_{m\times n}(K)$, $C = M_{\mathcal{C}}^{\mathcal{A}}(\varphi \circ \psi) \in \mathrm{Mat}_{l\times n}(K)$. Es gelten dann also

$$\psi(\mathbf{u}_j) = \sum_{k=1}^{m} b_{kj}\mathbf{v}_k \quad \text{und} \quad \varphi(\mathbf{v}_k) = \sum_{i=1}^{l} a_{ik}\mathbf{w}_i.$$

Einsetzen des rechten Ausdrucks in den linken ergibt

$$\begin{aligned}\varphi(\psi(\mathbf{u}_j)) &= \varphi\Big(\sum\nolimits_{k=1}^{m} b_{kj}\mathbf{v}_k\Big) = \sum\nolimits_{k=1}^{m} b_{kj}\varphi(\mathbf{v}_k) = \sum\nolimits_{k=1}^{m} b_{kj}\Big(\sum\nolimits_{i=1}^{l} a_{ik}\mathbf{w}_i\Big)\\ &= \sum\nolimits_{i=1}^{l}\Big(\sum\nolimits_{k=1}^{m} a_{ik}b_{kj}\Big)\mathbf{w}_i.\end{aligned}$$

Vergleich mit

$$(\varphi\circ\psi)(\mathbf{u}_j) = \sum_{i=1}^{l} c_{ij}\mathbf{w}_i$$

zeigt $C = AB$, wie behauptet. ∎

17.6 Korollar (Rechenregeln für die Matrizenmultiplikation)
Für Matrizen A, B, C gelten

(1) $(AB)C = A(BC)$ *(Assoziativität)*

(2) $A\mathbb{1} = \mathbb{1}A = A$ *(Einheitsmatrix)*

(3) $(A+B)C = AC + BC$ *und* $A(B+C) = AB + AC$ *(Distributivität)*

Dabei müssen die Matrizen A, B, C und $\mathbb{1}$ in jeder dieser Aussagen zueinander passende Formate haben, damit die Summen und Produkte definiert sind.

Aus (3) ergeben sich auch die beiden Distributivgesetze für das Matrix-Vektor-Produkt, die wir bereits formuliert haben.

Beweis. Es ist nicht schwierig, aber etwas mühsam, diese Regeln direkt nachzurechnen. Es genügt aber auch zu bemerken, dass die entsprechenden Regeln für lineare Abbildungen erfüllt sind (Satz 16.11). Da jede $m\times n$-Matrix A in eindeutiger Weise der linearen Abbildung $\varphi_A\colon K^n \to K^m$ entspricht, folgen daraus die entsprechenden Eigenschaften für Matrizen mit Hilfe von Satz 17.3 und Satz 17.5. ∎

17.7 Korollar *Für jedes $n \in \mathbb{N}$ bilden die $n\times n$-Matrizen mit der Matrizenaddition und Matrizenmultiplikation einen Ring.*

Beweis. Wenn wir nur quadratische Matizen derselben Größe betrachten, dann sind beliebige Summen und Produkte von Matrizen definiert. Addition und Multiplikation sind also Verknüpfungen auf $\mathrm{Mat}_{n\times n}(K)$. Die Ringgesetze ergeben sich alle aus den nun bekannten Rechenregeln. ∎

17.4 Invertierbarkeit und Rang

Zwischen der Invertierbarkeit von linearen Abbildungen und ihren darstellenden Matrizen besteht der erwartete Zusammenhang:

17.8 Korollar *Es seien V und W zwei Vektorrääume derselben endlichen Dimension und sei $\varphi\colon V \to W$ eine lineare Abbildung.*

(1) Ist φ ein Isomorphismus, dann ist die darstellende Matrix $M_C^{\mathcal{B}}(\varphi)$ für jede Wahl von Basen $\mathcal{B}$ und C invertierbar, und es gilt

$$(M_C^{\mathcal{B}}(\varphi))^{-1} = M_{\mathcal{B}}^{C}(\varphi^{-1}).$$

(2) Ist $M_C^{\mathcal{B}}(\varphi)$ für irgendeine Wahl von Basen $\mathcal{B}$ und C invertierbar, dann ist φ ein Isomorphismus.

Beweis. Sei $n = \dim(V) = \dim(W)$. (1) Es gilt $\varphi^{-1} \circ \varphi = \mathrm{id}_V$. Für jede Wahl von Basen $\mathcal{B}$ bzw. C gilt dann

$$M_{\mathcal{B}}^{C}(\varphi^{-1}) \cdot M_C^{\mathcal{B}}(\varphi) = M_{\mathcal{B}}^{\mathcal{B}}(\mathrm{id}_V) = \mathbb{1}_n$$

nach Satz 17.5. Also ist $M_C^{\mathcal{B}}(\varphi)$ invertierbar mit der Inversen $M_{\mathcal{B}}^{C}(\varphi^{-1})$.

(2) Es seien $\mathcal{B}$ bzw. C Basen, für welche $A = M_C^{\mathcal{B}}(\varphi)$ invertierbar ist. Nach Satz 17.3 gibt es dann eine lineare Abbildung $\psi\colon W \to V$ mit $M_{\mathcal{B}}^{C}(\psi) = A^{-1}$. Mit Satz 17.5 folgt dann umgekehrt

$$\mathbb{1}_n = A^{-1} \cdot A = M_{\mathcal{B}}^{C}(\psi) \cdot M_C^{\mathcal{B}}(\varphi) = M_{\mathcal{B}}^{\mathcal{B}}(\psi \circ \varphi).$$

Aus $M_{\mathcal{B}}^{\mathcal{B}}(\psi \circ \varphi) = \mathbb{1}_n$ folgt $\psi \circ \varphi = \mathrm{id}_V$. Genauso zeigt man $\varphi \circ \psi = \mathrm{id}_W$. Also ist φ ein Isomorphismus (mit $\varphi^{-1} = \psi$). ∎

Der Rang einer linearen Abbildung ist per Definition die Dimension ihres Bildes. Für eine Matrix $A \in \mathrm{Mat}_{m\times n}(K)$ ist das Bild der linearen Abbildung φ_A dasselbe wie der Spaltenraum von A, der Rang von φ_A also der (Spalten-)Rang von A. Derselbe Zusammenhang besteht immer.

17.9 Korollar *Ist $\varphi\colon V \to W$ eine lineare Abbildung zwischen endlichdimensionalen Vektorräumen mit Basen $\mathcal{B}$ bzw. C, dann gilt*

$$\mathrm{Rang}(\varphi) = \mathrm{Rang}(M_C^{\mathcal{B}}(\varphi)).$$

Beweis. Setze $A = M_C^{\mathcal{B}}(\varphi)$, dann gilt $\kappa_C \circ \varphi = \varphi_A \circ \kappa_{\mathcal{B}}$. Da $\kappa_{\mathcal{B}}$ und κ_C Isomorphismen sind, haben $\kappa_C \circ \varphi$ und φ genauso wie $\varphi_A \circ \kappa_{\mathcal{B}}$ und φ_A isomorphe Bilder. Also sind auch die Bilder von φ und φ_A isomorph und haben dieselbe Dimension. ∎

18 Koordinatentransformationen

18.1 Basiswechsel und Übergangsmatrix

Es sei V ein Vektorraum der Dimension n über einem Körper K und sei $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ eine Basis von V. Dann hat jeder Vektor $\mathbf{v} \in V$ eine Darstellung

$$\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$$

und der Spaltenvektor $\vec{x} \in K^n$ ist der Koordinatenvektor von $\mathbf{v}$ bezüglich der Basis $\mathcal{B}$.

Es stellt sich die Frage, wie sich die Koordinatenvektoren ändern, wenn wir eine neue Basis $C = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$ von V wählen. Wir wollen dann von der Basis $\mathcal{B}$ zur Basis C übergehen. Als erstes drücken wir die neue Basis in der alten aus, das heißt, wir schreiben

$$\mathbf{w}_j = \sum_{i=1}^{n} t_{ij}\mathbf{v}_i.$$

für eine $n \times n$-Matrix T. Diese Matrix stellt in der Basis $\mathcal{B}$ eine lineare Abbildung dar, nämlich genau den Basiswechsel

$$\mathbf{v}_1 \mapsto \mathbf{w}_1, \ldots, \mathbf{v}_n \mapsto \mathbf{w}_n.$$

Da die $\mathbf{w}_1, \ldots, \mathbf{w}_n$ eine Basis bilden, ist diese Abbildung ein Isomorphismus (nach Kor. 16.14), also ist die Matrix T invertierbar.

Definition Wir nennen die Matrix T die **Übergangsmatrix** von $\mathcal{B}$ nach C.

Ein Vektor $\mathbf{v} \in V$ hat nun zwei Darstellungen

$$\begin{array}{ccccc} \mathbf{v} & = & \sum_{i=1}^{n} x_i\mathbf{v}_i & = & \sum_{i=1}^{n} y_i\mathbf{w}_i \\ & & \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} & & \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ & & \text{alte Koordinaten} & & \text{neue Koordinaten} \end{array}$$

die durch die Übergangsmatrix T zusammenhängen. Es gilt nämlich

$$\mathbf{v} = \sum_{j=1}^{n} y_j \mathbf{w}_j = \sum_{j=1}^{n} y_j \sum_{i=1}^{n} t_{ij} \mathbf{v}_i = \sum_{i=1}^{n} \Big(\sum\nolimits_{j=1}^{n} t_{ij} y_j \Big) \mathbf{v}_i = \sum_{i=1}^{n} x_i \mathbf{v}_i.$$

Das bedeutet gerade $T\vec{y} = \vec{x}$. Die Übergangsmatrix T drückt also die alten Koordinaten durch die neuen aus. In der Regel ist man aber an der anderen Richtung interessiert. Es gilt also

$$\vec{y} = T^{-1}\vec{x}.$$

Wir halten fest:

neue Koordinaten = Inverse der Übergangsmatrix · alte Koordinaten.

18.1 Notation Wie zuvor schreiben wir $\kappa_{\mathcal{B}}(\mathbf{v})$ für den Koordinatenvektor von $\mathbf{v}$ bezüglich der Basis $\mathcal{B}$. Außerdem schreiben wir $T_{\mathcal{B}}^{\mathcal{C}}$ für die Übergangsmatrix von der Basis $\mathcal{B}$ zur Basis $\mathcal{C}$. (Man beachte die Reihenfolge von $\mathcal{B}$ und $\mathcal{C}$!)

Wir haben bewiesen:

18.2 Satz *Sind $\mathcal{B}$ und $\mathcal{C}$ zwei Basen des endlichdimensionalen Vektorraums V, dann gilt für jeden Vektor $\mathbf{v} \in V$*

$$\kappa_{\mathcal{C}}(\mathbf{v}) = \left(T_{\mathcal{B}}^{\mathcal{C}}\right)^{-1} \cdot \kappa_{\mathcal{B}}(\mathbf{v}). \qquad \blacksquare$$

Per Definition gilt $(T_{\mathcal{B}}^{\mathcal{C}})^{-1} = T_{\mathcal{C}}^{\mathcal{B}}$, so dass man die Formel auch ohne Inverse hinschreiben kann. In der Regel ist aber $T_{\mathcal{B}}^{\mathcal{C}}$ die Matrix, die sich direkt bestimmen lässt und anschließend invertiert werden muss.

Zu jedem Basiswechsel gehört also eine invertierbare Matrix. Wir bemerken dazu, dass auch das Umgekehrte gilt.

18.3 Lemma *Es sei V ein Vektorraum der Dimension n mit Basis $\mathcal{B}$ und sei $T \in \mathrm{GL}_n(K)$ eine invertierbare Matrix. Dann gibt es eine Basis $\mathcal{C}$ mit $T = T_{\mathcal{B}}^{\mathcal{C}}$.*

Beweis. Die Matrix T stellt in der Basis $\mathcal{B}$ einen Isomorphismus $\tau\colon V \to V$ mit $T = M_{\mathcal{B}}^{\mathcal{B}}(\tau)$ dar. Ist $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$, dann ist $\mathcal{C} = (\tau(\mathbf{v}_1), \ldots, \tau(\mathbf{v}_n))$ die gesuchte neue Basis von V mit Übergangsmatrix T von $\mathcal{B}$ nach $\mathcal{C}$. $\blacksquare$

Bevor wir Beispiele dazu diskutieren, ein paar allgemeine Bemerkungen:

(1) In K^n rechnet man zunächst immer bezüglich der Standardbasis $\vec{e}_1, \dots, \vec{e}_n$. Es gibt keine Koordinatenvektoren und keine Basiswechsel. Lineare Abbildungen entsprechen einfach Matrizen. Alle sind glücklich und zufrieden.

(2) Ist V ein abstrakter Vektorraum, dann muss man erst eine Basis von V wählen, bevor man überhaupt in Koordinaten rechnen kann. Es ist dann klar, dass alle diese Rechnungen von der Basis abhängen und sich beim Wechsel zu einer anderen Basis ändern. Das ist die Theorie, die wir hier entwickeln.

(3) Etwas verwirrend kann es werden, wenn die beiden Fälle (1) und (2) zusammentreffen. Man rechnet zwar in $\mathbb{R}^n$ oder K^n, hat aber außer der Standardbasis noch andere Basen, zum Beispiel von Lösungsräumen, und muss zwischen diesen Basen hin- und herwechseln. Dabei muss man höllisch aufpassen, die Vektoren selbst (Spaltenvektoren in K^n) und ihre Koordinatenvektoren bezüglich einer gewählten Basis (ebenfalls Spaltenvektoren in K^n) nicht zu verwechseln, sonst kommt man völlig durcheinander.

18.4 Beispiel Wir betrachten das homogene lineare Gleichungssystem

$$A\vec{x} = \vec{0} \quad \text{mit} \quad A = \begin{pmatrix} 1 & 1 & 4 & 2 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

in den Unbekannten x_1, x_2, x_3, x_4. Hier sind x_2 und x_4 freie Unbekannte, und der Lösungsraum U hat die Basisvektoren

$$\vec{u}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{und} \quad \vec{u}_2 = \begin{pmatrix} 2 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

Wir können diese beiden Vektoren zu einer Basis von $\mathbb{R}^4$ ergänzen, indem wir zwei weitere linear unabhängige Vektoren hinzufügen, zum Beispiel $\vec{u}_3 = \vec{e}_3$ und $\vec{u}_4 = \vec{e}_4$. Die Übergangsmatrix von der Standardbasis $\mathcal{B} = (\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4)$ in die neue Basis $C = (\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4)$ und ihre Inverse sind dann

$$T = T_{\mathcal{B}}^{C} = \begin{pmatrix} -1 & 2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \quad \text{und} \quad T^{-1} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 0 & 1 \end{pmatrix}.$$

Für einen Vektor $\vec{v} \in \mathbb{R}^4$ ist der Koordinatenvektor $\varkappa_{\mathcal{B}}(\vec{v})$ bezüglich $\mathcal{B}$ einfach der Vektor $\vec{v}$ selber. Der neue Koordinatenvektor ist dagegen

$$\varkappa_C(\vec{v}) = T^{-1} \cdot \vec{v}$$

Weil $\vec{u}_1$ bzw. $\vec{u}_2$ die ersten beiden Basisvektoren sind, gilt $\varkappa_C(\vec{u}_1) = \vec{e}_1$ und $\varkappa_C(\vec{u}_2) = \vec{e}_2$. Der Lösungsraum U ist in den neuen Koordinaten der Unterraum $\mathrm{Lin}(\vec{e}_1, \vec{e}_2)$. Es gilt beispielsweise

$$\varkappa_C \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \\ -1 \end{pmatrix} . \qquad \diamond$$

18.5 Beispiel Hier ist ein inhaltlich interessanteres Beispiel. Der Raum $\mathbb{R}[x]_{\leqslant d}$ aller Polynome vom Grad höchstens d hat die Basis $\mathcal{M} = (1, x, x^2, \dots, x^d)$. Der Koordinatenvektor eines Polynoms $f = a_d x^d + \cdots + a_1 x + a_0$ ist einfach der Spaltenvektor

$$\varkappa_{\mathcal{M}}(f) = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{pmatrix} \in \mathbb{R}^{d+1}$$

seiner Koeffizienten. Im ersten Moment sieht man keinen Grund, jemals eine andere Basis zu betrachten, aber tatsächlich kommt das in der Approximationstheorie bzw. der numerischen Analysis aus bestimmten praktischen Gründen sehr häufig vor. Auf den Hintergrund gehen wir hier nicht näher ein. Ein Beispiel ist die **Bernstein-Basis**[1]. Sie besteht aus den Polynomen

$$b_{i,d}(x) = \binom{d}{i} x^i (1-x)^{d-i} \qquad \text{für } i = 0, \dots, d.$$

Man kann sich überlegen, dass auch $\mathcal{B} = \big(b_{0,d}(x), \dots, b_{d,d}(x)\big)$ eine Basis von $\mathbb{R}[x]_{\leqslant d}$ bildet. Jedes Polynom $f \in \mathbb{R}[x]_{\leqslant d}$ hat also eine eindeutige Darstellung

$$f(x) = \sum_{i=0}^{d} \beta_i b_{i,d}(x)$$

[1] Sergei Natanowitsch Bernstein (1880–1968), russischer/sowjetischer Mathematiker.

Die Koeffizienten $\beta_0, \dots, \beta_d \in \mathbb{R}$ heißen die Bernstein-Koeffizienten von f. Sie bilden den Koordinatenvektor

$$\varkappa_{\mathcal{B}}(f) = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_d \end{pmatrix} \in \mathbb{R}^{d+1}.$$

Betrachten wir den Fall $d = 3$. Die vier Bernstein-Polynome vom Grad 3 sind

$$\begin{aligned} b_{0,3}(x) &= (1-x)^3 = -x^3 + 3x^2 - 3x + 1 \\ b_{1,3}(x) &= 3x(1-x)^2 = 3x^3 - 6x^2 + 3x \\ b_{2,3}(x) &= 3x^2(1-x) = -3x^3 + 3x^2 \\ b_{3,3}(x) &= x^3 \end{aligned}$$

Wir wollen den Basiswechsel von $\mathcal{M}$ nach $\mathcal{B}$ ausrechenen, gegeben durch

$$1 \mapsto b_{0,3}(x),\ x \mapsto b_{1,3}(x),\ x^2 \mapsto b_{2,3}(x),\ x^3 \mapsto b_{3,3}(x).$$

Die Koordinatenvektoren $\varkappa_{\mathcal{M}}(b_{i,3}(x))$, also die Koeffizienten, bilden die Spalten der Übergangsmatrix $T = T^{\mathcal{B}}_{\mathcal{M}}$. Wir erhalten so die Matrix T und rechnen gleich noch ihre Inverse aus:

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3 & 3 & 0 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{pmatrix} \quad \text{und} \quad T^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & \frac{1}{3} & 0 & 0 \\ 1 & \frac{2}{3} & \frac{1}{3} & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Wir können nun die Bernstein-Koeffizienten eines kubischen Polynoms $f \in \mathbb{R}[x]_{\leqslant 3}$ berechnen, nämlich den Vektor $\varkappa_{\mathcal{B}}(f) = T^{-1}\varkappa_{\mathcal{M}}(f)$. Für zum Beispiel $f = 12x^2 - 9x + 3$ ist dann $\varkappa_{\mathcal{M}} = (3, -9, 12, 0)^t$ und damit

$$\varkappa_{\mathcal{B}} = T^{-1}\varkappa_{\mathcal{M}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & \frac{1}{3} & 0 & 0 \\ 1 & \frac{2}{3} & \frac{1}{3} & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ -9 \\ 12 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 1 \\ 6 \end{pmatrix}$$

was der Gleichung $f(x) = 3b_{0,3}(x) + b_{2,3}(x) + 6b_{3,3}$ entspricht. Wenn wir die Übergangsmatrix einmal bestimmt haben, dann können wir in dieser Weise die Bernstein-Koeffizienten für beliebige kubische Polynome schnell ausrechnen. ◇

18.2 Der Transformationssatz für lineare Abbildungen

Als nächstes müssen wir uns überlegen, was beim Basiswechsel mit den darstellenden Matrizen linearer Abbildungen passiert. Es sei $\varphi\colon V \to W$ eine lineare Abbildung zwischen zwei endlichdimensionalen Vektorräumen der Dimension n bzw. m. Wir wählen Basen $\mathcal{B}$ von V und C von W und können dann φ durch die darstellende Matrix $M_C^{\mathcal{B}}(\varphi) \in \mathrm{Mat}_{m\times n}(K)$ beschreiben. Wir müssen uns nun mit zwei möglichen Basiswechseln befassen, nämlich auf V und auf W.

18.6 Satz (Transformationssatz für lineare Abbildungen) *Es seien V und W zwei Vektorräume der Dimension n bzw. m. Gegeben seien Basen $\mathcal{B}$ und $\mathcal{B}'$ von V sowie C und C' von W mit Übergangsmatrizen*

$$S = T_{\mathcal{B}}^{\mathcal{B}'} \in \mathrm{GL}_n(K) \quad \textit{und} \quad T = T_C^{C'} \in \mathrm{GL}_m(K)$$

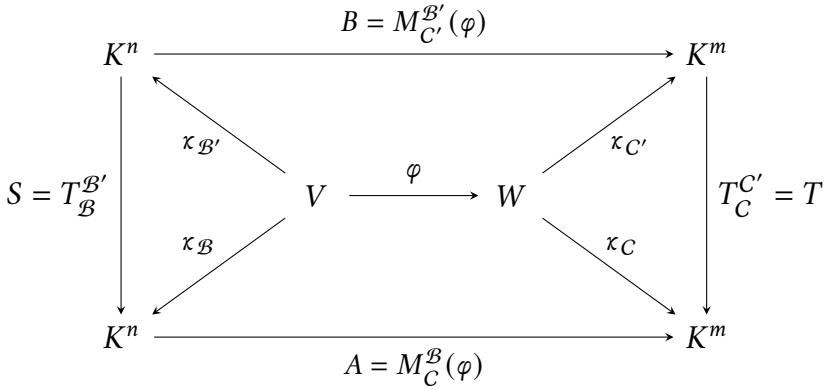
Es sei $\varphi\colon V \to W$ eine lineare Abbildung und seien

$$A = M_C^{\mathcal{B}}(\varphi) \quad \textit{und} \quad B = M_{C'}^{\mathcal{B}'}(\varphi)$$

die darstellenden Matrizen. Dann gilt die Beziehung

$$B \;=\; T^{-1}AS.$$

Erster Beweis. Wir betrachten das folgende wunderbare Diagramm:



Die mit den Matrizen A, B, S, T bezeichneten Pfeile entsprechen genauer den linearen Abbildungen $\varphi_A, \varphi_B, \varphi_S, \varphi_T$. Man überzeugt sich, dass dieses Diagramm kommutiert, was jeweils nur den Definitionen der beteiligten Matrizen entspricht, und kann die Aussage einfach ablesen, durch den Vergleich zwischen dem direkten Weg von links nach rechts oben und dem Umweg über die untere Kante. ■

Zweiter Beweis. Wenn einem dieses Argument zu wolkig vorkommt, kann man die Aussage auch einfach ausrechnen, eine Übung im Verwalten von Indizes: Es seien

$$\begin{aligned} \mathcal{B} &= (\mathbf{v}_1, \dots, \mathbf{v}_n) \quad \text{und} \quad \mathcal{B}' = (\mathbf{v}'_1, \dots, \mathbf{v}'_n) \\ C &= (\mathbf{w}_1, \dots, \mathbf{w}_m) \quad \text{und} \quad C' = (\mathbf{w}'_1, \dots, \mathbf{w}'_m) \end{aligned}$$

Wir können nun auf beiden Seiten die Basen wechseln und bekommen einerseits

$$\varphi(\mathbf{v}'_j) = \varphi\Big(\sum_{i=1}^{n} s_{ij}\mathbf{v}_i\Big) = \sum_{i=1}^{n} s_{ij}\varphi(\mathbf{v}_i) = \sum_{i=1}^{n} s_{ij} \sum_{k=1}^{m} a_{ki}\mathbf{w}_k = \sum_{k=1}^{m}\Big(\sum_{i=1}^{n} a_{ki}s_{ij}\Big)\mathbf{w}_k$$

und andererseits

$$\varphi(\mathbf{v}'_j) = \sum_{i=1}^{m} b_{ij}\mathbf{w}'_i = \sum_{i=1}^{m} b_{ij} \sum_{k=1}^{m} t_{ki}\mathbf{w}_k = \sum_{k=1}^{m}\Big(\sum_{i=1}^{m} t_{ki}b_{ij}\Big)\mathbf{w}_k.$$

Vergleich der Koeffizienten in der Basis C ergibt

$$\sum_{i=1}^{n} a_{ki}s_{ij} = \sum_{i=1}^{m} t_{ki}b_{ij}$$

was gerade $AS = TB$ bedeutet, und damit $B = T^{-1}AS$ wie behauptet. ∎

Wenn man unbedingt will, kann man die Regel in Satz 18.6 auch völlig in Notation verpacken:

$$M_{C'}^{\mathcal{B}'}(\varphi) = (T_C^{C'})^{-1} \cdot M_C^{\mathcal{B}}(\varphi) \cdot T_{\mathcal{B}}^{\mathcal{B}'}$$

18.3 Normalform für darstellende Matrizen

Der folgende Satz zeigt, wie Basiswechsel den Umgang mit linearen Abbildungen vereinfachen können. Im Prinzip lässt sich die darstellende Matrix immer auf dieselbe einfache Gestalt bringen.

18.7 Satz *Es seien V und W Vektorräume der Dimension n bzw. m. Sei $\varphi\colon V \to W$ eine lineare Abbildung vom Rang r. Dann gibt es Basen $\mathcal{B}$ von V und C von W mit*

$$M_C^{\mathcal{B}}(\varphi) = \underbrace{\begin{pmatrix} \mathbb{1}_r & 0 \\ 0 & 0 \end{pmatrix}}_{n}\Bigg\}m \ .$$

Beweis. Nach der Dimensionsformel für lineare Abildungen gilt

$$\dim(\mathrm{Kern}(\varphi)) = \dim(V) - \mathrm{Rang}(\varphi) = n - r.$$

Wir wählen eine Basis $\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ von $\mathrm{Kern}(\varphi)$ und ergänzen sie mit dem Basisergänzungssatz zu einer Basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ von V, das heißt, wir fügen die Vektoren $\mathbf{v}_1, \dots, \mathbf{v}_r$ vorne hinzu. Setze $\mathbf{w}_i = \varphi(\mathbf{v}_i)$ für $i = 1, \dots, r$, dann sind $\mathbf{w}_1, \dots, \mathbf{w}_r$ weiterhin linear unabhängig. Ist $c_1\mathbf{w}_1 + \cdots + c_r\mathbf{w}_r = \mathbf{0}$ eine lineare Relation, dann also

$$\mathbf{0} = c_1\mathbf{w}_1 + \cdots + c_r\mathbf{w}_r = c_1\varphi(\mathbf{v}_1) + \cdots + c_r\varphi(\mathbf{v}_r) = \varphi(c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r)$$

und damit $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r \in \mathrm{Kern}(\varphi)$. Es ist aber $\mathrm{Kern}(\varphi) = \mathrm{Lin}(\mathbf{v}_{r+1}, \dots, \mathbf{v}_n)$ und $\mathbf{v}_1, \dots, \mathbf{v}_n$ sind linear unabhängig. Also folgt $c_1 = \cdots = c_r = 0$. (Dasselbe haben wir ausführlicher im Beweis der Dimensionsformel überprüft.) Wir können die Vektoren $\mathbf{w}_1, \dots, \mathbf{w}_r$ (die eine Basis von $\mathrm{Bild}(\varphi)$ bilden) zu einer Basis $\mathbf{w}_1, \dots, \mathbf{w}_m$ von W ergänzen. In den Basen $\mathcal{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ und $C = (\mathbf{w}_1, \dots, \mathbf{w}_m)$ ist φ dann per Definition bestimmt durch die Zuordnung

$$\mathbf{v}_1 \mapsto \mathbf{w}_1, \dots, \mathbf{v}_r \mapsto \mathbf{w}_r,\ \mathbf{v}_{r+1} \mapsto \mathbf{0}, \dots, \mathbf{v}_n \mapsto \mathbf{0}.$$

Also hat $M_C^{\mathcal{B}}(\varphi)$ die gewünschte Gestalt. ∎

18.8 Korollar *Es sei $A \in \mathrm{Mat}_{m\times n}(K)$ eine Matrix vom Rang r. Dann gibt es invertierbare Matrizen $T \in \mathrm{GL}_m(K)$ und $S \in \mathrm{GL}_n(K)$ derart, dass*

$$T^{-1}AS = \begin{pmatrix} \mathbb{1}_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Beweis. Wende zuerst Satz 18.7 auf die lineare Abbildung $\varphi_A \colon K^n \to K^m, \vec{x} \mapsto A\vec{x}$ an und erhalte Basen $\mathcal{B}$ und C mit $M_C^{\mathcal{B}}(\varphi_A) = B$, wobei B von der gewünschten Form ist. Nach dem Transformationssatz 18.6 gilt dann $B = T^{-1}AS$ für die Übergangsmatrizen S bzw. T von den Standardbasen auf die Basen $\mathcal{B}$ bzw. C. ∎

18.4 Transponierte Matrix

Als Vorbereitung auf den nächsten Abschnitt und für die Zukunft führen wir kurz das Transponieren von Matrizen ein.

Definition Für jede $m \times n$-Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

ist die **Transponierte** von A die $n \times m$-Matrix

$$A^t = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

Üblich sind für die transponierte Matrix auch A^T oder gelegentlich A'. Das Buch von Fischer verwendet die (sonst weniger gebräuchliche) Notation ${}^tA$.

18.9 Beispiele (1) In der Transponierten sind also Zeilen- und Spaltenindizes einfach vertauscht, zum Beispiel

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

(2) Die Transponierte einer $n \times 1$-Matrix (also eines Spaltenvektors) ist eine $1 \times n$-Matrix (also ein Zeilenvektor), und umgekehrt.

(3) Für zwei Spaltenvektoren $\vec{x}, \vec{y} \in K^n$ ergibt das Matrix-Vektor-Produkt $\vec{x}^t \cdot \vec{y}$ den Skalar

$$\vec{x}^t \cdot \vec{y} = \sum\nolimits_{i=1}^{n} x_i y_i$$

was für $K = \mathbb{R}$ das Skalarprodukt auf $\mathbb{R}^n$ ist. ◇

18.10 Lemma *Für das Transponieren gelten die Rechenregeln*

$$(A^t)^t = A, \qquad (AB)^t = B^t A^t, \qquad (C^{-1})^t = (C^t)^{-1}$$

für alle $A \in \mathrm{Mat}_{l\times m}(K)$, $B \in \mathrm{Mat}_{m\times n}(K)$ *und invertierbares* $C \in \mathrm{Mat}_{n\times n}(K)$.

Beweis. Die Gleichheit $(A^t)^t = A$ ist klar. Weiter gilt

$$(AB)_{ij} = \sum\nolimits_{k=1}^{m} a_{ik} b_{kj},$$

also

$$(AB)^t_{ij} = (AB)_{ji} = \sum_{k=1} a_{jk} b_{ki} = \sum_{k=1} b_{ki} a_{jk} = (B^t A^t)_{ij}.$$

Die letzte Aussage folgt nun wegen $(C^{-1})^t C^t = (CC^{-1})^t = \mathbb{1}_n^t = \mathbb{1}_n$, also $(C^{-1})^t = (C^t)^{-1}$, wie behauptet. ∎

18.5 Orthogonale Matrizen

Als nächstes bestimmen wir die darstellenden Matrizen von orthogonalen Abbildungen auf $\mathbb{R}^n$.

18.11 Satz *Es sei A eine reelle $n \times n$-Matrix. Genau dann ist die lineare Abbildung φ_A orthogonal, wenn*

$$A^t A = AA^t = \mathbb{1}_n$$

gilt, mit anderen Worten, wenn A invertierbar ist mit $A^{-1} = A^t$.

Definition Eine reelle $n \times n$-Matrix A mit der Eigenschaft $A^t A = AA^t = \mathbb{1}_n$ wird **orthogonale Matrix** genannt.

Beweis von Satz 18.11. Die Spaltenvektoren $\vec{b}_1, \dots, \vec{b}_n$ von A sind gerade die Bilder $\varphi_A(\vec{e}_1), \dots, \varphi_A(\vec{e}_n)$ der Einheitsvektoren. Nach Satz 9.9 ist φ_A also genau dann orthogonal, wenn $\vec{b}_1, \dots, \vec{b}_n$ eine Orthonormalbasis von $\mathbb{R}^n$ bilden. Die Bedingung $\langle \vec{b}_i, \vec{b}_j \rangle = \delta_{ij}$ ist aber genau dasselbe wie $A^t A = \mathbb{1}_n$. Wir bemerken außerdem, dass aus $A^t A = \mathbb{1}_n$ auch $AA^t = (A^t A)^t = E_n^t = E_n$ folgt, und umgekehrt. ∎

18.12 Lemma *Sind $A, B \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ orthogonal, dann sind es auch AB und A^{-1}.*

Beweis. Es gilt $(AB)^t = B^t A^t = B^{-1} A^{-1} = (AB)^{-1}$, also ist AB orthogonal. Ebenso gilt $(A^{-1})^t = (A^t)^t = A = (A^{-1})^{-1}$, also ist auch A^{-1} orthogonal. ∎

18.13 Korollar *Es sei $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ eine orthogonale Abbildung. Dann wird φ in jeder Orthonormalbasis von $\mathbb{R}^n$ durch eine orthogonale Matrix dargestellt.*

Beweis. Es sei $\mathcal{E}$ die Standardbasis und $\mathcal{B}$ eine Orthonormalbasis von $\mathbb{R}^n$. Nach Satz 18.11 ist $A = M_{\mathcal{E}}^{\mathcal{E}}(\varphi)$ orthogonal. Außerdem ist die Übergangsmatrix $T = T_{\mathcal{E}}^{\mathcal{B}}$ orthogonal, denn ihre Spalten sind die Vektoren in $\mathcal{B}$. Nach dem Transformationssatz gilt $M_{\mathcal{B}}^{\mathcal{B}} = T^{-1} A T$. Nach dem Lemma ist diese Matrix orthogonal. ∎

18.14 Beispiel In der Standardbasis beschreibt die Matrix

$$R_\alpha = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

eine Drehung von $\mathbb{R}^2$ um den Nullpunkt mit Winkel $\alpha \in [0, 2\pi)$ gegen den Uhrzeigersinn. (Dass diese Matrix tatsächlich orthogonal ist, liegt an der Identität $\sin(\alpha)^2 + \cos(\alpha)^2 = 1$.) Außerdem beschreibt die Matrix

$$S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

die Spiegelung an der waagerechten Achse. Wenn wir an einer anderen Ursprungsgeraden spiegeln wollen, etwa an der Geraden L_α mit Winkel α, dann können wir L_α in die Waagerechte drehen, dann die Spiegelung ausführen und dann wieder zurückdrehen. Mit anderen Worten, wir verwenden $R_\alpha^{-1} = R_\alpha^t$ als Koordinatenwechsel und erhalten (unter Benutzung der Additionstheoreme für Sinus und Cosinus) als darstellende Matrix der Spiegelung an L_α die Matrix

$$R_\alpha S R_\alpha^t = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{pmatrix} \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix} = \begin{pmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{pmatrix}. \quad \diamond$$

19 Gruppen linearer Abbildungen

Neben der Gruppe $\mathrm{GL}_n(K)$ der invertierbaren Matrizen gibt es weitere Gruppen von linearen Abbildungen, die wir nun einführen. Bei dieser Gelegenheit gehen wir außerdem auf einige weitere Begriffe der Gruppentheorie ein.

19.1 Lineare Gruppen

Ist V ein Vektorraum über einem Körper K, dann bilden die Isomorphismen, also der bijektiven linearen Abbildungen, $V \to V$ eine Gruppe

$$\mathrm{GL}(V) = \big\{\varphi\colon V \to V \mid \varphi \text{ ist bijektiv}\big\}$$

mit der Komposition als Verknüpfung: Das neutrale Element ist die Identität id_V und das Inverse ist die Umkehrabbildung $\varphi' = \varphi^{-1}$.

Für $V = K^n$ wird jeder Isomorphismus durch eine invertierbare $n \times n$-Matrix dargestellt, zum Beispiel in der Standardbasis. Das gibt die **allgemeine lineare Gruppe**

$$\mathrm{GL}_n(K) = \big\{A \in \mathrm{Mat}_{n\times n}(K) \mid A \text{ ist invertierbar}\big\}$$

aller invertierbaren Matrizen, die wir (für $K = \mathbb{R}$) schon betrachtet haben.

19.1 Beispiel Von den 16 Matrizen in $\mathrm{Mat}_{2\times 2}(\mathbb{F}_2)$ sind 6 invertierbar, nämlich

$$\mathbb{1}_2 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, A = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, B = \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, C = \begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}, D = \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}, E = \begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}.$$

Die Gruppe $\mathrm{GL}_2(\mathbb{F}_2)$ hat also 6 Elemente. Sie ist nicht abelsch, denn es gilt zum Beispiel $AB = E \neq C = BA$. ◇

19.2 Untergruppen

Definition Sei $(G, *)$ eine Gruppe. Eine Teilmenge $H \subset G$ heißt eine **Untergruppe**, wenn gelten

(1) $H \neq \emptyset$;

(2) Für alle $h_1, h_2 \in H$ gilt $h_1 * h_2 \in H$;

(3) Für alle $h \in H$ gilt $h' \in H$.

19.2 Beispiele (1) In der Gruppe $(\mathbb{R}^*, \cdot)$ bilden die positiven reellen Zahlen $\mathbb{R}_{>0}$ eine Untergruppe $(\mathbb{R}_{>0}, \cdot)$.

(2) In der Gruppe $GL_2(K)$ ist

$$H = \left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \mid a \in K \right\}$$

eine Untergruppe, denn für alle $a, b \in K$ gelten

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a+b \\ 0 & 1 \end{pmatrix} \in H \quad \text{und} \quad \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -a \\ 0 & 1 \end{pmatrix} \in H.$$

(3) Allgemeiner bilden die **oberen Dreiecksmatrizen**

$$H = \left\{ A \in \mathrm{GL}_n(K) \mid a_{ij} = 0 \text{ für alle } i < j \right\}$$

und die oberen Dreiecksmatrizen mit Einsen auf der Diagonalen (die *unipotenten* oberen Dreiecksmatrizen)

$$U = \left\{ A \in H \mid a_{ii} = 1 \text{ für alle } i \right\}$$

Untergruppen von $\mathrm{GL}_n(K)$ (Übung).

(4) Nach Lemma 18.12 bildet die Menge der orthogonalen $n \times n$-Matrizen

$$\mathrm{O}_n(\mathbb{R}) = \left\{ A \in \mathrm{GL}_n(\mathbb{R}) \mid A^{-1} = A^t \right\}$$

eine Untergruppe von $\mathrm{GL}_n(\mathbb{R})$, die **orthogonale Gruppe**. ◇

19.3 Homomorphismen und Isomorphismen

Im zweiten Beispiel oben ist H eine Gruppe von 2×2-Matrizen, aber bei der Multiplikation werden die Einträge rechts oben einfach addiert. Mit anderen Worten, die Untergruppe H ist letztlich *dasselbe* wie die additive Gruppe $(K, +)$. Strukturell gleiche Gruppen können in verschiedener Weise in Erscheinung treten. Dazu führen wir den folgenden Begriff ein.

Definition Seien $(G_1, *_1)$ und $(G_2, *_2)$ Gruppen. Eine Abbildung $\varphi\colon G_1 \to G_2$ heißt ein **Homomorphismus** (oder genauer Gruppenhomomorphismus), wenn

$$\varphi(g *_1 h) = \varphi(g) *_2 \varphi(h)$$

für alle $g, h \in G_1$ gilt.

19.3 Beispiele (1) Die Abbildung

$$\varphi\colon \mathbb{Z} \to \mathbb{Z},\ a \mapsto 2a$$

ist ein Homomorphismus $(\mathbb{Z},+) \to (\mathbb{Z},+)$. Denn es gilt $\varphi(a+b) = 2(a+b) = 2a+2b = \varphi(a)+\varphi(b)$ für alle $a,b \in \mathbb{Z}$.

(2) Die Abbildung $\varphi\colon \mathbb{Z} \to \mathbb{Z},\ a \mapsto a+1$ ist kein Homomorphismus $(\mathbb{Z},+) \to (\mathbb{Z},+)$. Denn es gilt zum Beispiel $\varphi(1+1) = \varphi(2) = 3 \neq 4 = 2+2 = \varphi(1)+\varphi(1)$.

(3) Die Exponentialfunktion

$$\varphi\colon \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto e^x$$

ist ein Homomorphismus zwischen der additiven Gruppe $(\mathbb{R},+)$ und der multiplikativen Gruppe $(\mathbb{R}_{>0},\cdot)$ der positiven reellen Zahlen. Das entspricht der Rechenregel

$$\varphi(x+y) = e^{x+y} = e^x \cdot e^y = \varphi(x)\cdot\varphi(y)$$

für die Exponentialfunktion, die man aus der Schule oder der Analysis I kennt.

(4) In ähnlicher Weise ist die Abbildung

$$\varphi\colon K \to \mathrm{GL}_2(K),\ a \mapsto \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

aus Beispiel 19.2(2) ein Homomorphismus $(K,+) \to \big(\mathrm{GL}_2(K),\cdot\big)$. ◇

Im folgenden schreiben wir die Gruppenoperation immer mit $*$, das neutrale Element immer mit e und die Inversen immer mit g', selbst wenn es um zwei verschiedene Gruppen geht.

19.4 Lemma *Es sei $\varphi\colon G_1 \to G_2$ ein Homomorphismus.*

(1) *Es gilt $\varphi(e) = e$.*

(2) *Für alle $g \in G_1$ gilt $\varphi(g') = \varphi(g)'$.*

Beweis. (1) Es gilt $\varphi(e) = \varphi(e*e) = \varphi(e)*\varphi(e)$. Multiplikation mit $\varphi(e)'$ auf beiden Seiten zeigt $\varphi(e) = e$.

(2) Es gilt $\varphi(g)*\varphi(g') = \varphi(g*g') = \varphi(e) = e$ nach (1). Wegen der Eindeutigkeit der Inversen folgt daraus $\varphi(g') = \varphi(g)'$. ■

19.5 Lemma *Ist $\varphi\colon G_1 \to G_2$ ein bijektiver Homomorphismus, dann ist die Umkehrabbildung $\varphi^{-1}\colon G_2 \to G_1$ ebenfalls ein Homomorphismus.*

Beweis. Für $g, h \in G_2$ gilt

$$\begin{aligned}\varphi^{-1}(g * h) &= \varphi^{-1}\big(\varphi(\varphi^{-1}(g)) * \varphi(\varphi^{-1}(h))\big) = \varphi^{-1}\big(\varphi(\varphi^{-1}(g) * \varphi^{-1}(h))\big)\\ &= \varphi^{-1}(g) * \varphi^{-1}(h)\end{aligned}$$ ∎

Definition Ein bijektiver Homomorphimus heißt ein **Isomorphismus**. Zwei Gruppen G_1 und G_2 heißen **isomorph**, wenn es einen Isomorphismus zwischen ihnen gibt. Wir schreiben in diesem Fall $G_1 \cong G_2$.

19.6 Beispiele (1) Die Abbildung

$$\varphi\colon K \to \mathrm{GL}_2(K),\ a \mapsto \begin{pmatrix}1 & a\\ 0 & 1\end{pmatrix}$$

ist ein Isomorphismus zwischen der Gruppe $(K, +)$ und der Untergruppe

$$H = \left\{ \begin{pmatrix}1 & a\\ 0 & 1\end{pmatrix} \;\middle|\; a \in K \right\}$$

von $\big(\mathrm{GL}_2(K), \cdot\big)$ aus Beispiel 19.2.

(2) Die Exponentialfunktion $x \mapsto e^x$ ist ein Isomorphismus $(\mathbb{R}, +) \to (\mathbb{R}_{>0}, \cdot)$.

(3) Für eine Primzahl p gibt es im Körper $\mathbb{F}_p$ genau $p - 1$ Elemente außer 0. Die multiplikative Gruppe $(\mathbb{F}_p \setminus \{0\}, \cdot)$ ist also eine Gruppe mit $p - 1$ Elementen. Wir kennen noch eine andere Gruppe mit $p - 1$ Elementen, nämlich die *additive* Gruppe $(\mathbb{Z}/(p-1), +)$. Tatsächlich kann man zeigen, dass diese beiden Gruppen isomorph sind. (Das beweist man in der Vorlesung *Algebra*). Für den Beweis fehlen uns im Moment die Mittel. Wir können uns aber ein Beispiel anschauen, indem wir die Gruppentafeln von $(\mathbb{Z}/4, +)$ und $(\mathbb{F}_5 \setminus \{0\}, \cdot)$ vergleichen.

$(\mathbb{Z}/4, +)$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{0}$	$\overline{1}$	$\overline{2}$

$(\mathbb{F}_5 \setminus \{0\}, \cdot)$	1	2	4	3
1	1	2	4	3
2	2	4	3	1
4	4	3	1	2
3	3	1	2	4

Dabei habe ich die Elemente von $\mathbb{F}_5 \setminus \{0\}$ in einer anderen Reihenfolge hingeschrieben, damit man die Isomorphie besser sieht. Sie ist durch die Abbildung

$$\overline{0} \mapsto 1, \quad \overline{1} \mapsto 2, \quad \overline{2} \mapsto 4, \quad \overline{3} \mapsto 3$$

gegeben. Dieses Beispiel wird im Weiteren nicht verwendet. ◇

19.4 Isometrische und orthogonale Gruppe

In Kapitel 9 hatten wir Isometrien auf dem Vektorraum $\mathbb{R}^n$ betrachtet. Das sind Abbildungen, die alle Abstände zwischen Punkten erhalten. Die orthogonalen Abbildungen sind die Isometrien, welche den Ursprung fixieren und damit auch die Länge von Vektoren erhalten. Ist $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ eine orthogonale Abbildung, dann hatten wir bewiesen, dass φ auch das Skalarprodukt erhält, also

$$\langle \varphi(\vec{x}), \varphi(\vec{y}) \rangle = \langle \vec{x}, \vec{y} \rangle$$

für alle $\vec{x}, \vec{y} \in \mathbb{R}^n$ gilt. (Die orthogonalen Abbildungen sind damit genau die linearen Abbildungen mit dieser Eigenschaft.)

19.7 Satz *Die Isometrien bilden mit der Komposition als Verknüpfung eine Gruppe und die orthogonalen Abbildungen darin eine Untergruppe.*

Beweis. Das wurde bereits in den Übungen gezeigt. ∎

Ein Beispiel für eine Isometrie, die keine orthogonale Abbildung ist, ist eine Translation $\tau_{\vec{v}}\colon \vec{x} \to \vec{x}+\vec{v}$ um einen Vektor $\vec{v} \neq \vec{0}$. In den Übungen wurde gezeigt, dass jede Isometrie aus einer orthogonalen Abbildung und einer Translation zusammengesetzt ist, das heißt, für jedes $\varphi \in \mathrm{Iso}(\mathbb{R}^n)$ gibt es einen Vektor $\vec{v} \in \mathbb{R}^n$ und eine orthogonale Abbildung $\psi \in \mathrm{O}(\mathbb{R}^n)$ mit

$$\varphi = \tau_{\vec{v}} \circ \psi.$$

Definition Die Gruppe

$$\mathrm{Iso}(\mathbb{R}^n) = \left\{\varphi\colon \mathbb{R}^n \to \mathbb{R}^n \mid \forall \vec{x}, \vec{y} \in \mathbb{R}^n\colon \|\varphi(\vec{x}) - \varphi(\vec{y})\| = \|\vec{x} - \vec{y}\|\right\}$$

aller Isometrien $\mathbb{R}^n \to \mathbb{R}^n$ heißt die **isometrische Gruppe** (oder *Bewegungsgruppe*) in Dimension n. Die Untergruppe

$$\mathrm{O}(\mathbb{R}^n) = \left\{\varphi \in \mathrm{Iso}(\mathbb{R}^n) \mid \varphi(\vec{0}) = \vec{0}\right\}$$

der orthogonalen Abbildungen ist die (abstrakte) **orthogonale Gruppe**.

Da orthogonale Abbildungen linear sind, ist $\mathrm{O}(\mathbb{R}^n)$ auch eine Untergruppe der Gruppe $\mathrm{GL}(\mathbb{R}^n)$ der bijektiven linearen Abbildungen. Wenn wir wollen, können wir auch alle diese Gruppen als Untergruppen der Gruppe $\mathrm{Bij}(\mathbb{R}^n)$ aller bijektiven Abbildungen $\mathbb{R}^n \to \mathbb{R}^n$ interpretieren.

19.5 Darstellende Matrizen in Gruppen

Wir betrachten noch einmal die Übersetzung von linearen Abbildungen in Matrizen im Kontext von Gruppen.

19.8 Satz *Es sei V ein Vektorraum der endlichen Dimension n über K. Für jede Wahl einer Basis $\mathcal{B}$ von V ist*

$$M_{\mathcal{B}}^{\mathcal{B}}\colon \mathrm{GL}(V) \to \mathrm{GL}_n(K), \varphi \mapsto M_{\mathcal{B}}^{\mathcal{B}}(\varphi)$$

ein Isomorphismus von Gruppen.

Beweis. Das haben wir alles schon bewiesen: Die Abbildung $M_{\mathcal{B}}^{\mathcal{B}}$ ist wohldefiniert und surjektiv nach Kor. 17.8, außerdem injektiv nach Satz 17.3. Zusätzlich gilt $M_{\mathcal{B}}^{\mathcal{B}}(\varphi \circ \psi) = M_{\mathcal{B}}^{\mathcal{B}}(\varphi) \cdot M_{\mathcal{B}}^{\mathcal{B}}(\psi)$ für alle $\varphi, \psi \in \mathrm{GL}(V)$ nach Satz 17.5. Also ist $M_{\mathcal{B}}^{\mathcal{B}}$ auch ein Homomorphismus von Gruppen. ∎

Für orthogonale Abbildungen auf $\mathbb{R}^n$ bekommen wir orthogonale Matrizen. Mit Kor. 18.13 folgt wie oben:

19.9 Satz *Für jede Wahl einer Orthonormalbasis $\mathcal{B}$ von $\mathbb{R}^n$ ist*

$$\mathrm{O}(\mathbb{R}^n) \to \mathrm{O}_n(\mathbb{R}),\ \varphi \mapsto M_{\mathcal{B}}^{\mathcal{B}}(\varphi)$$

ein Isomorphismus von Gruppen. ∎