Mathematik II für Studierende der Informatik

Markus Junker
Albert-Ludwigs-Universität Freiburg

Sommersemester 2013

Vorbemerkung

Dieses Skript entsteht nach und nach und soll im Wesentlichen den Inhalt der im Sommersemester 2013 an der Albert-Ludwigs-Universität in Freiburg gehaltenen Vorlesung „Mathematik II für Studierende der Informatik“ wiedergeben. An einzelnen Stellen wird das Skript über die Vorlesung hinausgehen; umgekehrt sind nicht alle Bemerkungen, Erläuterungen und Beispiele aus der Vorlesung enthalten.

Das Skript verdankt viel einer von Lisa Schüttler angefertigten Mitschrift der gleichen Vorlesung im Sommersemester 2012.

„Plagiats-Disclaimer“

Das Skript ist nach in der Mathematik gängiger Vorgehensweise angefertigt. Dies bedeutet, dass es keinen Anspruch auf eine eigene wissenschaftliche Leistung erhebt und keine eigenen Ergebnisse wiedergibt, sondern die Ergebnisse anderer darstellt. Diese Ergebnisse sind über Jahrhunderte gewachsen; da Mathematik weitgehend ahistorisch betrieben wird, lässt sich in der Regel nicht mehr zurückverfolgen, von wem welche Fragestellungen, Begriffe, Sätze, Beweise oder Beweistechniken stammen. Vereinzelt gibt es überlieferte Zuweisungen von Sätzen oder von Beweisen zu Mathematikern (die aber nicht immer historisch exakt sein müssen).

Die Darstellung des Stoffes orientiert sich an den von mir selbst gehörten Vorlesungen, an Skripten von Kollegen und an Büchern. Diese verschiedenen Einflüsse sind nicht zu trennen und können daher nicht einzeln dargelegt werden. Fehler dagegen sind von mir zu verantworten. Insbesondere bei Formeln empfiehlt sich eine kritische Lektüre, da kleine Tippfehler aufgrund mangelnder Redundanz gleich massive Fehler bewirken.

Inhaltsverzeichnis

Kapitel 1

Lineare Algebra

Im ersten Abschnitt werden etwas kompakt einige grundlegende algebraische Strukturen vorgestellt: Monoïde, Gruppen, Ringe und Körper. Gruppen und Körper braucht man für die Definition von Vektorräumen, welche das zentrale Objekt der Linearen Algebra sind. Gruppen, Ringe und Körper werden dann im Kapitel 2 weiter untersucht werden. Die Monoide werden keine weitere Rolle mehr spielen, stehen aber der Vollständigkeit halber hier, da sie auf dem Weg zu den Gruppen ohne großen Aufwand „mitgenommen“ werden können und das Beispiel des Monoid der Wörter über einem Alphabet in der Informatik häufiger vorkommt.

1.1 Grundlegende algebraische Strukturen

Monoide

Definition 1.1.1 *Ein* Monoid [1] *besteht aus einer nicht-leeren Menge M und einer zweistelligen Verknüpfung $\circ$ auf M, also einer Abbildung $\circ : M \times M \to M$, die*

- assoziativ *ist, d. h.* $(m_1 \circ m_2) \circ m_3 = m_1 \circ (m_2 \circ m_3)$ *für alle* $m_1, m_2, m_3 \in M$ *erfüllt,*
- *und ein* neutrales Element $e \in M$ *besitzt, d. h. es gilt* $e \circ m = m \circ e = m$ *für alle* $m \in M$.

Ein Monoid $(M, \circ)$ *heißt* kommutatives Monoid, *wenn die Verknüpfung* $\circ$ *zusätzlich*

- kommutativ *ist, d. h.* $m_1 \circ m_2 = m_2 \circ m_2$ *für alle* $m_1, m_2 \in M$ *erfüllt.*

Bemerkungen:

Das neutrale Element ist eindeutig bestimmt, denn falls e, e' neutrale Elemente sind, so gilt $e = e \circ e' = e'$.

Wegen der Assoziativität kann man bei iterierten Verknüpfungen Klammern weglassen.

Beispiele:

- Die natürlichen Zahlen $\mathbb{N}$ bilden mit der Addition $+$ ein kommutatives Monoid mit neutralem Element 0.
- Die natürlichen Zahlen $\mathbb{N}$ bilden mit der Multiplikation $\cdot$ ein kommutatives Monoid mit neutralem Element 1.
- Die echt positiven natürlichen Zahlen $\mathbb{N} \setminus \{0\}$ bilden mit der Multiplikation $\cdot$ ein kommutatives Monoid mit neutralem Element 1.

[1] „Monoid“ ist sächlich („das Monoid“), Aussprache: „Mono-id“ mit Betonung auf der letzten Silbe.

- Die Abbildungen $\mathrm{Abb}(A, A)$ einer Menge A in sich selbst bilden unter der Komposition $\circ$, d. h. der Hintereinanderausführung von Funktionen, ein (i. a. nicht kommutatives) Monoid, dessen neutrales Element die identische Abbildung id_A ist.
- Wenn A eine Menge ist (in diesem Kontext auch „Alphabet“ genannt), bildet die Menge A^* der endlichen Folgen von Elementen aus A (auch „Wörter über A“ genannt) mit der „Konkatenation“ (d. h. Hintereinandersetzen) $\widehat{\ }$ ein (i. a. nicht kommutatives) Monoid. Mit $A = \{a, b, c\}$ ist also z. B. $abaac\widehat{\ }ccb = abaacccb$. Das neutrale Element ist das „leere Wort“, also die Folge der Länge 0, das oft mit λ oder ε bezeichnet wird.

Gegenbeispiele:

- Die echt positiven natürlichen Zahlen $\mathbb{N} \setminus \{0\}$ bilden mit der Addition $+$ kein Monoid, da es kein neutrales Element gibt.
- Die natürlichen Zahlen $\mathbb{N}$ bilden mit der Exponentiation kein Monoid, da die Exponentiation nicht assoziativ ist, denn z. B. ist $2^{3^2} = 2^9 = 512$, aber $(2^3)^2 = 8^2 = 64$. Zudem gibt es zwar ein „rechtneutrales Element“ (da $n^1 = n$ für alle $n \in \mathbb{N}$), aber kein „linksneutrales“.

Zur Notation: Wenn die Menge M mit der Verknüpfung $\circ$ und neutralem Element e ein Monoid bildet, schreibt man dafür üblicherweise $(M, \circ, e)$ oder $(M, \circ)$, da e eindeutig festgelegt wird. Wenn man sauber arbeitet, unterscheidet man notationell zwischen der Struktur und der zugrundeliegenden Menge und schreibt dann gerne für die Struktur den entsprechenden Buchstaben in einem anderen Schriftart, also z. B. $\mathcal{M}$ oder $\mathfrak{M}$ für ein Monoid mit Grundmenge M. Oft erlaubt man sich aber die Unsauberkeit, für die Struktur und die Grundmenge das gleiche Symbol (hier z. B. M) zu verwenden.

Bei der Angabe eines Monoids entfällt bisweilen die Angabe der Verknüpfung, wenn aus dem Kontext heraus offensichtlich ist, welche gemeint ist, oder wenn es eine besonders natürliche Verknüpfung gibt. Wenn man z. B. vom Monoid der Wörter über einem Alphabet spricht oder dem Monoid der Abbildungen einer Menge in sich selbst, meint man die oben angegebenen Standardbeispiele. Das doppelte Beispiel der natürlichen Zahlen – einmal mit Addition und einmal mit Multiplikation – zeigt aber, dass man i. a. auf die Angabe der Verknüpfung nicht verzichten kann und selbst eine natürlich wirkende Operation nicht unbedingt einen Alleinstellungsanspruch hat.

Wenn mehrere Monoide betrachtet werden, werden oft die gleichen Notationen für die Verknüpfungen und neutralen Elemente gebraucht. Es kann also vorkommen, dass man Monoid $(M, \circ, e)$ und $(N, \circ, e)$ betrachtet. Zur Verdeutlichung schreibt man dann manchmal $\circ_M$ und e_M für Verknüpfung und neutrales Element von M und analog $\circ_N$ und e_N für die von N.

Analoge Bemerkungen zur Notation gelten für alle im weiteren eingeführten algebraischen Strukturen!

Gruppen

Definition 1.1.2 *Ein* Gruppe *besteht aus einer nicht-leeren Menge G und einer zweistelligen Verknüpfung $\circ$ auf G, die*

- assoziativ *ist,*
- *ein* neutrales Element $e \in M$ *besitzt,*
- *und bezüglich der es* inverse Elemente *gibt, d. h. zu jedem $g \in G$ gibt es ein Element $h \in G$ mit $h \circ g = g \circ h = e$.*

Eine Gruppe $(G, \circ)$ *heißt* kommutative Gruppe[2]*, wenn die Verknüpfung* $\circ$ *zusätzlich* kommutativ *ist.*

Bemerkungen und Notation: Jede (kommutative) Gruppe ist also insbesondere ein (kommutatives) Monoid.

Die inversen Elemente sind eindeutig bestimmt, denn sind h_1, h_2 invers zu g, so gilt

$$h_1 = h_1 \circ e = h_1 \circ (g \circ h_2) = (h_1 \circ g) \circ h_2 = e \circ h_2 = h_2.$$

Das bezüglich $\circ$ zu $g \in G$ inverse Element wird mit g^{-1} bezeichnet.

Es gibt drei gebräuchliche Notationen für Gruppen:

	Verknüpfung	neutrales Element	inverses Element
allgemein:	$\circ$	e	g^{-1}
multiplikativ:	$\cdot$	1	g^{-1}
additiv:	$+$	0	$-g$

Die additive Schreibweise ist im allgemeinen kommutativen Gruppen vorbehalten. Bei der multiplikativen Schreibweise lässt man den Multiplikationspunkt auch gerne weg.

Beispiele:

- $(\mathbb{Z}, +, 0)$ ist kommutative Gruppe.
- $(\mathbb{Q}, +, 0)$, $(\mathbb{Q} \setminus \{0\}, \cdot, 1)$ und $(\mathbb{Q}^{>0}, \cdot, 1)$ mit $\mathbb{Q}^{>0} = \{q \in \mathbb{Q} \mid q > 0\}$ sind kommutative Gruppen.
- Ebenso mit $\mathbb{R}$ statt $\mathbb{Q}$ und – außer für das letzte Beispiel – mit $\mathbb{C}$ statt $\mathbb{Q}$, .
- $(\mathrm{Sym}(A), \circ, \mathrm{id})$ ist eine i. a. nicht kommutative Gruppe, wobei $\mathrm{Sym}(A)$ die Menge der Bijektionen einer Menge A in sich bezeichnet.
- Wichtiges Beispiel einer Gruppe wird die „verallgemeinerte Uhren-Arithmetik“ sein, d. i. die kommutative Gruppe $\mathbb{Z}_m = \big(\{0, \ldots, m-1\}, +_m, 0\big)$, wobei

$$x +_m y = \text{„Rest von } x + y \text{ bei Division durch } m\text{“} = \begin{cases} x + y & \text{falls } x + y < m \\ x + y - m & \text{falls } x + y \geq m \end{cases}$$

  Für $n = 12$ ist dies die Art, wie man mit Uhrzeiten rechnet („*8 Uhr + 5 Stunden = 1 Uhr*“).

Gegenbeispiele: $(\mathbb{Z} \setminus \{0\}, \cdot, 1)$ oder $(\mathbb{Q}, \cdot, 1)$ sind keine Gruppen; im ersten Fall fehlen allen Elementen außer 1 und -1 die Inversen; im zweiten Fall hat 0 kein Inverses.

Bemerkung: Man sieht bei genauerem Anschauen, dass manche dieser Gruppen von einfachen Beispielen für Monoide herstammen. Zum Beispiel ist $(\mathbb{Z}, +, 0)$ eine Gruppe, die aus dem Monoid $(\mathbb{N}, +, 0)$ dadurch entsteht, dass man Inverse (also hier die negativen Elemente) hinzunimmt. Das funktioniert nicht immer, wie man am Beispiel der Null bezüglich der Multiplikation weiß oder am Beispiel der Abbildungen sehen kann: Wenn $h \circ g_1 = h \circ g_2$ für $g_1 \neq g_2$ gilt, kann kein (Links-)Inverses für h gefunden werden und die Abbildung assoziativ bleiben. Bei dem Abbildungsmonoid $(\mathrm{Abb}(A), \circ)$ kommt man nur dadurch offensichtlich zu einer Gruppe, dass man die Elemente herausgreift, die schon ein Inverses im Monoid haben (also die Bijektionen).

[2] oder auch *Abelsche Gruppe*, nach dem norwegischen Mathematiker Niels Henrik Abel (1802–1829)

Ringe

Definition 1.1.3 *Ein* Ring *besteht aus einer nicht-leeren Menge R, zwei zweistelligen Verknüpfungen $+$ und $\cdot$ auf R (Addition und Multiplikation) und Elemente 0 und 1 (Null und Eins), für die gilt:*

- $(R, +, 0)$ *ist eine kommutative Gruppe;*
- $(R, \cdot, 1)$ *ist ein Monoid;*
- *es gelten die* Distributivgesetze*, d. h. für alle $r_1, r_2, s \in R$ gilt:*

$$(r_1 + r_2) \cdot s = (r_1 \cdot s) + (r_2 \cdot s)$$
$$s \cdot (r_1 + r_2) = (s \cdot r_1) + (s \cdot r_2)$$

Ein Ring $(R, +, \cdot)$ heißt kommutativer Ring*, wenn die Multiplikation zusätzlich* kommutativ *ist.*

Bemerkungen und Notation: Genauer handelt es sich hier um „Ringe mit Eins“ oder „unitäre Ringe“. Es gibt auch ein allgemeineres Konzept von Ring, bei dem es kein neutrales Element der Multiplikation zu geben braucht. Bei der Lektüre anderer Skripte oder Bücher muss man also vorsichtig sein, welche Definition jeweils benutzt wird.

In einem kommutativen Ring folgt natürlich jedes der beiden Distributivgesetze aus dem anderen.

Zur Ersparnis von Klammern führt man die üblichen „Vorfahrtsregeln“ ein, also „Punkt vor Strich“; den Multiplikationspunkt lässt man auch gerne weg. Das erste Distributivgesetz kann man also kurz als $(r_1 + r_2)s = r_1 s + r_2 s$ schreiben.

Man rechnet nach, dass $r \cdot 0 = 0 \cdot r = 0$ für alle $r \in R$ ist (denn z. B. $r \cdot 0 = r \cdot (0+0) = r \cdot 0 + r \cdot 0$).

Ähnlich sieht man, dass $(-r) \cdot s = r \cdot (-s) = -(r \cdot s)$ für alle $r, s \in R$. Auch hier kann man daher Klammern einsparen.

Beispiele:

- Die Definition verbietet nicht, dass $0 = 1$ ist. In diesem Fall folgt aber $r = r \cdot 1 = r \cdot 0 = 0$ für alle $r \in R$, und es liegt der sogenannte *triviale Ring* vor, der nur aus einem Element besteht.
- $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ – jeweils mit der normalen Addition und Multiplikation – sind kommutative Ringe.
- Die Gruppe $\mathbb{Z}_m$ kann durch eine analog definierte Multiplikation $\cdot_m$ zu einem kommutativen Ring gemacht werden: $x \cdot_m y$ rechnet man dadurch aus, dass man von dem normalen Produkt in $\mathbb{Z}$ den Rest bei der Division durch m nimmt, also solange m abzieht, bis man im Bereich $\{=, \ldots, m-1\}$ landet.
- Die Polynome mit Koeffizienten in einem Ring R und der Unbekannten X bilden mit der bekannten Polynomaddition und -multiplikation den *Polynomring* $R[X]$, also z. B. $\mathbb{R}[X]$ (Polynome mit einer Unbekannten X und Koeffizienten in $\mathbb{R}$) oder $\mathbb{Z}[X]$ (Polynome mit einer Unbekannten X und Koeffizienten in $\mathbb{Z}$).
  Nimmt man mit einer neuen Unbekannten Y z. B. den Polynomring $\mathbb{R}[X]$ als Koeffizientenbereich, erhält man den Polynomring mit zwei Unbekannten X und Y mit Koeffizienten in $\mathbb{R}$, also $\mathbb{R}[X][Y] = \mathbb{R}[X, Y]$.

Körper

Definition 1.1.4 *Ein* Körper *besteht aus einer nicht-leeren Menge K, zwei zweistelligen Verknüpfungen $+$ und $\cdot$ auf R (Addition und Multiplikation) und Elemente 0 und 1 (Null und Eins), für die gilt:*

- $0 \neq 1$;
- $(K, +, 0)$ *und* $K \setminus \{0\}, \cdot, 1$ *sind kommutative Gruppen;*
- *es gelten die* Distributivgesetze.

Bemerkung: Jeder Körper ist also insbesondere ein kommutativer, nicht-trivialer Ring.[3]

Beispiele:

- $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$ sind Körper.
- Für Primzahlen p ist $\mathbb{Z}_p$ ein Körper und wird dann oft $\mathbb{F}_p$ geschrieben.

Besonders interessant für die Informatik ist der Körper $\mathbb{F}_2$, der aus den beiden Elementen 0 und 1 besteht mit folgenden Verknüpfungen:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Genauer handelt es sich hier um „Ringe mit Eins“ oder „unitäre Ringe“. Es gibt auch ein allgemeineres Konzept von Ring, bei dem es kein neutrales Element der Multiplikation zu geben braucht. Bei der Lektüre anderer Skripte oder Bücher muss man also vorsichtig sein, welche Definition jeweils benutzt wird.

In einem kommutativen Ring folgt natürlich jedes der beiden Distributivgesetze aus dem anderen.

Zur Ersparnis von Klammern führt man die üblichen „Vorfahrtsregeln“ ein, also „Punkt vor Strich“; den Multiplikationspunkt lässt man auch gerne weg. Das erste Distributivgesetz kann man also kurz als $(r_1 + r_2)s = r_1 s + r_2 s$ schreiben.

Man rechnet nach, dass $r \cdot 0 = 0 \cdot r = 0$ für alle $r \in R$ ist (denn z. B. $r \cdot 0 = r \cdot (0+0) = r \cdot 0 + r \cdot 0$).

Ähnlich sieht man, dass $(-r) \cdot s = r \cdot (-s) = -(r \cdot s)$ für alle $r, s \in R$. Auch hier kann man daher Klammern einsparen.

1.2 Vektorräume

Sei K ein Körper, also z. B. $K = \mathbb{R}$ oder $K = \mathbb{F}_2$ (dies werden die hauptsächlichen Beispiele in dieser Vorlesung sein). Zur Verdeutlichung sind die Körperelemente und -operationen vorübergehend mit einem Index K gekennzeichnet, also $+_K, -_K, \cdot_K, 0_K, 1_K$.

Definition 1.2.1 *Ein* K-Vektorraum *V besteht aus einer nicht-leeren Menge V zusammen mit einer zweistelligen „inneren“ Verknüpfung $+ : V \times V \to V$ (der „Addition“) und einer „äußeren“ Verknüpfung $\cdot : K \times V \to V$ (der „Skalarmultiplikation“), für die gilt:*

[3] Es gibt auch ein etwas allgemeineres Konzept eines „Schiefkörper“, bei dem die Multiplikation nicht kommutativ zu sein braucht.

- $(V, +)$ *ist eine kommutative Gruppe mit neutralem Element* 0_V*;*
- *es gelten folgende Regeln für die Skalarmultiplikation:*

$$\begin{aligned} 1_K \cdot v &= v \\ (k_1 +_K k_2) \cdot v &= (k_1 \cdot v) + (k_2 \cdot v) \\ (k_1 \cdot_K k_2) \cdot v &= k_1 \cdot (k_2 \cdot v) \\ k \cdot (v_1 + v_2) &= (k \cdot v_1) + (k \cdot v_2) \end{aligned}$$

für alle $k, k_1, k_2 \in K$ *und* $v, v_1, v_2 \in V$*.*

Bemerkungen und Notation: Falls aus dem Kontext klar ist, um welchen Körper K es geht, spricht man auch kurz von „Vektorraum“ statt von „K-Vektorraum“. Elemente von V heißen *Vektoren*, Elemente von K *Skalare*.

Im Unterschied zu einem Ring kann man Vektoren in einem allgemeinen Vektorraum nicht miteinander multiplizieren[4]; es gelten aber ähnliche Regeln, die man ganz analog beweisen kann. So gilt $k \cdot 0_V = 0_V$ und $0_K \cdot v = 0_V$ und $k \cdot (-v) = (-_K k) \cdot v = -(k \cdot v)$ für alle $k \in K$ und $v \in V$.

Für Vektorräume erlaubt man sich die gleichen notationellen Kurzformen wie bei Ringen (Klammersparregeln und Weglassen des Multiplikationspunktes). Auch werde ich von nun an die Indizes K und V in der Regel weglassen; dadurch bekommen 0, $+$ und $\cdot$ eine doppelte Bedeutung, es sollte aber aus der Situation immer klar werden, welche Null bzw. welche Addition und Multiplikation gemeint sind. Eine Skalarmultiplikation liegt immer dann vor, wenn links ein Körperelement und rechts ein Vektor steht; wenn auf beiden Seiten ein Körperelement steht, handelt es sich um die Multiplikation im Körper. Die Addition kann nur zwischen zwei Vektoren oder zwischen zwei Körperelementen stehen.

Beispiele:

⋮

1.3 Untervektorräume und Erzeugende

In diesem Abschnitt sei stets V ein K-Vektorraum.

Definition 1.3.1 *$U \subseteq V$ heißt K-Untervektorraum von V, falls U unter den eingeschränkten Operationen selbst ein K-Vektorraum ist, d. h. $0 \in U$ und für alle $u, u_1, u_2 \in U$ und $k \in K$ liegen die Elemente $u_1 + u_2$, $-u$ und $k \cdot u$ in U. Man schreibt dafür $U \leqslant V$.*

Bemerkungen: Sämtliche Regeln wie Assoziativität, Kommutativität und Distributivität oder die Neutralität von 0 übertragen sich automatisch auf Teilmengen.

Die Abgeschlossenheit bezüglich Negation folgt automatisch aus den anderen Regeln, da $-u = (-1) \cdot u$. Wenn $U \neq \emptyset$, etwa $u \in U$, folgt $0 = u + (-u) \in U$. Untervektorräume sind also genau die nicht-leeren, bezüglich Addition und Skalarmultiplikation abgeschlossenen Teilmengen.

[4] dessen ungeachtet gibt es in speziellen Fällen gewisse Vektorprodukte

Wenn der Körper K durch den Kontext bekannt ist, sagt man auch kurz „Untervektorraum“ statt „K-Untervektorraum“. Außerdem verkürzt man bisweilen „Untervektorraum“ zu „Unterraum“.

Beispiel: Sei $K = \mathbb{R}$ und $V = \mathbb{R}$. Die $\mathbb{R}$-Untervektorräume von V sind dann:

- der triviale Untervektorraum $\{0\}$;
- alle Teilmengen $\{(x, y) \in \mathbb{R}^2 \mid ax + by = 0\}$ für feste $a, b \in \mathbb{R}$ – dies sind die Geraden durch den Ursprung $(0, 0)$;
- der ganze Vektorraum $\mathbb{R}^2$.

Man überzeugt sich leicht davon, dass ein Schnitt von beliebig vielen K-Untervektorräumen von V wieder ein K-Untervektorraum von V ist.

Definition 1.3.2 *Seien $v_1, \ldots, v_n \in V$. Der von $v_1, \ldots, v_n$ erzeugte Untervektorraum von V, bezeichnet mit $\langle v_1, \ldots, v_n \rangle$ ist äquivalenterweise definiert als*

- *der kleinste Untervektorraum von V, der $v_1, \ldots, v_n$ enthält;*
- *der Schnitt aller Untervektorräumen von V, die $v_1, \ldots, v_n$ enthalten;*
- $\{k_1 v_1 + \cdots + k_n v_n \mid k_1, \ldots, k_n \in K\}$

Zur Äquivalenz der Definitionen: Man sieht leicht, dass jeder K-(Unter-)Vektorraum mit $v_1, \ldots, v_n$ auch jede sogenannte *Linearkombination* von $v_1, \ldots, v_n$, d. i. jeden Ausdruck der Form $k_1 v_1 + \cdots + k_n v_n$ enthalten muss. Umgekehrt sieht man auch leicht, dass die Summen und skalare Vielfache von Linearkombinationen von $v_1, \ldots, v_n$ wieder Linearkombinationen von $v_1, \ldots, v_n$ sind und sich jedes v_i als Linearkombination von $v_1, \ldots, v_n$ schreiben lässt (z. B. $v_1 = 1 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_n$), also bilden die Linearkombinationen von $v_1, \ldots, v_n$ einen $v_1, \ldots, v_n$ enthaltenden Untervektorraum.

Spezialfall: Für $n = 0$ erhält man die leere Menge und es ist $\langle \emptyset \rangle = \{0\}$; damit die Definition auch in diesem Fall stimmt, definiert man 0 als den Wert der „leeren Summe“.

Verallgemeinerung: Für eventuell unendlich viele Vektoren, also für eine Teilmenge $A \subseteq V$, kann man ebenfalls den von A erzeugten Untervektorraum von V als den kleinsten A enthaltenden Untervektorraum definieren. Es gilt dann

$$\langle A \rangle = \{k_1 a_1 + \cdots + k_n a_n \mid n \in \mathbb{N}, a_1, \ldots a_n \in A, k_1, \ldots, k_n \in K\}.$$

Mengenklammern lässt man in der Kombination mit den spitzen Klammern für das Erzeugnis meist weg, d. h. man schreibt kurz $\langle v_1, v_2, \ldots \rangle$ statt $\langle \{v_1, v_2, \ldots\} \rangle$ oder $\langle v_i \mid i \in I \rangle$ statt $\langle \{v_i \mid i \in I\} \rangle$.

Terminologie: Falls $V = \langle v_i \mid i \in I \rangle$, so sagt man

- die v_i $(i \in I)$ „erzeugen V“ oder
- die v_i $(i \in I)$ „sind Erzeuger (oder Erzeugende) von V“ oder
- $\{v_i \mid i \in I\}$ „ist ein *Erzeugendensystem* von V“

oder Varianten hiervon.

V heißt *endlich erzeugt*, falls es ein endliches Erzeugendensystem gibt.

Beispiele:

⋮

1.4 Lineare Unabhängigkeit, Basis, Dimension

Sei wieder stets V ein K-Vektorraum, und sei $A \subseteq V$ eine Menge von Vektoren.

Definition 1.4.1 *Ein Vektor $v \in V$ ist* linear abhängig *von A, falls $v \in \langle A \rangle$, d. h. falls es $a_1, \ldots, a_n \in A$ und $k_1, \ldots, k_n \in K$ gibt mit $v = k_1 a_1 + \cdots + k_n a_n$.*

A ist linear unabhängig*, falls kein $a \in A$ linear abhängig von $A \setminus \{a\}$ ist.*

Aus der Definition folgt unmittelbar, dass eine Menge von unendlich vielen Vektoren genau dann linear unabhängig ist, wenn jede endliche Teilmenge linear unabhängig ist.

Vorsicht vor den Tücken der Mengenschreibweise bei Doppelnennungen:
Angenommen $v_1 = v_2$ ist linear unabhängig von v_3. Dann ist v_1 linear abhängig von $\{v_2, v_3\}$, aber die Menge $\{v_1, v_2, v_3\}$ ist linear unabhängig, denn $\{v_1, v_2, v_3\} = \{v_1, v_3\}$.
Dieses Problem führt in der Folge zu etwas umständlich wirkenden Formulierungen: Unter eine Menge $\{v_i \mid i \in I\}$ *ohne Doppelnennungen* ist gemeint, dass $v_i \neq v_j$ für $i \neq j$, also dass die Elemente v_i für $i \in I$ paarweise verschieden sind. Anders ausgedrückt: die Abbildung $I \to V$, $i \mapsto v_i$ ist injektiv, oder, noch einmal anders ausgedrückt, $v_{i_0} \notin \{v_i \mid i \in I \setminus \{i_0\}\}$ für alle $i_0 \in I$.

Lemma 1.4.2 *$\{v_1, \ldots, v_n\}$ ist genau dann linear unabhängig und ohne Doppelnennungen, wenn nur die „triviale Linearkombination“ 0 ergibt, d. h. wenn $k_1 v_1 + \cdots + k_n v_n = 0$ nur für $k_1 = 0, \ldots, k_n = 0$ gilt.*

Beweis: Wenn die Menge linear abhängig ist oder Doppelnennungen vorliegen, wenn also z. B. $v_1 = k_2 v_2 + \cdots + k_n v_n$, dann folgt $(-1) \cdot v_1 + k_2 v_2 + \cdots + k_n v_n = 0$. Wenn es umgekehrt eine Darstellung $k_1 v_1 + \cdots + k_n v_n = 0$ gibt, bei der etwa $k_1 \neq 0$, so folgt $v_1 = -\frac{k_2}{k_1} v_2 + \cdots + (-\frac{k_n m}{k_1}) v_n$, also ist die Menge $\{v_1, \ldots, v_n\}$ linear abhängig oder hat Doppelnennungen. □

Daraus folgt sofort die Version für unendlich viele Vektoren: Eine Menge ohne Doppelnennungen $\{v_i \mid i \in I\}$ ist genau dann linear unabhängig, wenn es keine nicht-triviale Linearkombination von endlich vielen Vektoren aus der Menge 0 ergibt.

Definition 1.4.3 *Eine* Basis *von V ist ein linear unabhängiges Erzeugendensystem.*

Satz 1.4.4 *$\{v_i \mid i \in I\}$ ist eine Basis von V*
$\iff$ $\{v_i \mid i \in I\}$ ist eine maximale linear unabhängige Teilmenge von V
$\iff$ $\{v_i \mid i \in I\}$ ist ein minimales Erzeugendensystem von V

(„maximal“ und „minimal“ sind bezüglich der Teilmengenbeziehung)

Beweis:

⋮

□

Folgerung 1.4.5 *Jeder endlich erzeugte Vektorraum besitzt Basen; jedes endliche Erzeugendensystem enthält eine Basis und jede linear unabhängige Teilmenge lässt sich zu einer Basis vergrößern.*

Folgerung 1.4.5 gilt auch für unendlich dimensionale Vektorräume, ist aber langwieriger zu beweisen.

Ohne Beweis benutzen wir:[5]

Satz 1.4.6 *Jeder Vektorraum besitzt Basen, und je zwei Basen eines Vektorraums haben die gleiche Mächtigkeit (d. h. stehe in Bijektion zueinander; im endlichen Fall: haben die gleiche Anzahl von Elementen. Diese Mächtigkeit bzw. Anzahl heißt* Dimension *von V (über K). Man schreibt dafür $\dim_K V$ oder kurz $\dim V$, wenn K im Kontext festgeschrieben ist.*

Beispiele:

⋮

Satz 1.4.7 *Seien $v_1, \dots, v_n$ paarweise verschiedene Elemente. Dann ist $\{v_1, \dots, v_n\}$ genau dann eine Basis von V, wenn es für jedes $v \in V$ eine eindeutige Darstellung $v = k_1v_1 + \cdots + k_nv_n$ gibt.*

Die Koeffizienten $k_1, \dots, k_n$ werden dann häufig die *Koordinaten* von v bezüglich der Basis genannt.

Beweis: Zunächst ist klar, dass genau dann für jedes $v \in V$ solch eine Darstellung existiert, wenn $\{v_1, \dots, v_n\}$ ein Erzeugendensystem ist. Angenommen nun $v = k_1v_1 + \cdots + k_nv_n = k'_1v_1 + \cdots + k'_nv_n$. Dann gilt $0 = (k_1 - k'_1)v_1 + \cdots + (k_n - k'_n)v_n$, d. h. es gibt genau dann zwei verschiedene Darstellungen für einen Vektor, falls es eine nicht-triviale Linearkombination der Null gibt, was nach Lemma 1.4.2 genau dann der Fall ist, wenn $\{v_1, \dots, v_n\}$ nicht linear unabhängig ist. □

Für unendlich viele Vektoren folgt, dass eine Teilmenge von V ohne Doppelnennungen $\{v_i \mid i \in I\}$ genau dann eine Basis von V ist, wenn es für jedes $v \in V$ eine eindeutige Darstellung $v = \sum_{i \in I} k_iv_i$ mit $k_i \in K$ gibt. Hierbei gilt für die Summe die Konvention, dass nur endlich viele Indizes $k_i \neq 0$; die Summe steht also für eine endliche Linearkombination von Vektoren.

1.5 Lineare Abbildungen

Seien V und W K-Vektorräume.

Definition 1.5.1 *Eine Abbildung $\varphi : V \to W$ ist eine* K-lineare Abbildung *oder ein* K-Vektorraum-Homomorphismus*, falls φ mit der Gruppenstruktur und der Skalarmultiplikation verträglich ist, d. h. falls für alle $v, v_1, v_2 \in V$ und $k \in K$ gilt*[6]*:*

- $\varphi(v_1 +_V v_2) = \varphi(v_1) +_W \varphi(v_2)$, $\varphi(0_V) = 0_W$ *und* $\varphi(-_V v) = -_W \varphi(v)$
- $\varphi(k \cdot_V v) = k \cdot_W \varphi(v)$.

Eine Abbildung $\varphi : V \to W$ ist ein K-Vektorraum-Isomorphismus*, falls φ eine bijektive Abbildung ist und sowohl φ als auch die Umkehrabbildung φ^{-1} K-linear sind.*

V und W heißen isomorph *(als K-Vektorräume), falls ein K-Vektorraum-Isomorphismus $\varphi : V \to W$ existiert. Man schreibt dafür $V \cong W$.*

[5] Ein Beweis für endliche-dimensionale Vektorräume folgt später aus dem Gauß-Verfahren; man muss sich aber davon überzeugen, dass der Satz für das Gauß-Verfahren nicht gebraucht wird.

[6] Der Deutlichkeit halber sind wieder vorübergehend bei den Operationen Indizes V und W angebracht, je nachdem, in welcher Struktur gerechnet wird.

Bemerkungen: Man kann zeigen, dass die beiden Bedingungen $\varphi(0) = 0$ und $\varphi(-v) = -\varphi(v)$ aus der Additivität $\varphi(v_1 + v_2) = \varphi(v_1) + \varphi(v_2)$ folgt, da $(V, +)$ eine Gruppe ist.

Ebenso kann man zeigen, dass die Umkehrabbildung einer bijektiven K-linearen Abbildung automatisch K-linear ist.

Falls aus dem Kontext klar ist, um welchen Körper K es sich handelt, spricht man auch kurz von „linearen Abbildungen“ bzw. „Vektorraum-Homo- und -isomorphismen“.

Der Begriff „isomorph“ und die Notation $V \cong W$ werden auch für andere Strukturen verwendet (z. B. Gruppen, Ringe). Wenn sie ohne nähere Spezifikation verwendet werden, setzen sie voraus, dass aus dem Kontext klar ist, welche Art von Strukturen betrachtet werden, hier also K-Vektorräume. In diesem Fall spricht man auch noch kürzer von „Homomorphismen“ und „Isomorphismen“.

Satz 1.5.2 *Sei $\{v_i \mid i \in I\}$ eine Basis von V ohne Doppelnennungen, und seien w_i für $i \in I$ beliebige Elemente von W. Dann gibt es genau eine lineare Abbildung $\varphi : V \to W$ mit $\varphi(v_i) = w_i$ für alle $i \in I$. Außerdem ist φ ist dann ein Isomorphismus, wenn $\{w_i \mid i \in I\}$ eine Basis von W ohne Doppelnennungen ist.*

Beweis: Wenn es überhaupt solch eine lineare Abbildung gibt, muss $\varphi(k_1 v_{i_1} + \cdots + k_n v_{i_n}) = k_1 w_{i_1} + \cdots + k_n w_{i_n}$ gelten. Da nach Satz 1.4.7 jedes v eine eindeutige Darstellung $v = \sum_{j_1}^{n} k_j v_{i_j}$ besitzt mit $n \in \mathbb{N}$, $k_j \in K$ und paarweise verschiedenen $i_j \in I$, kann man durch $\varphi(v) := \sum_{j_1}^{n} k_j w_{i_j}$ auch tatsächlich eine Abbildung $V \to W$ definieren. Man sieht dann auch leicht ein, dass diese Abbildung tatsächlich linear ist.

Das Bild von φ besteht dann aus den Vektoren $\varphi(\sum_{j_1}^{n} k_j v_{i_j}) = \sum_{j_1}^{n} k_j \varphi(v_{i_j}) = \sum_{j_1}^{n} k_j w_{i_j}$, also ist φ genau dann surjektiv, wenn $\{w_i \mid i \in I\}$ ein Erzeugendensystem ist. Wenn $\{w_i \mid i \in I\}$ eine Basis ohne Doppelnennungen ist, folgt aus der Eindeutigkeit der Darstellung (Satz 1.4.7) auch die Injektivität von φ. Wenn umgekehrt φ bijektiv ist, muss zum einen $\{w_i \mid i \in I\}$ eine Menge ohne Doppelnennungen sein (da φ injektiv), und zum andern gilt $w = \sum_{j_1}^{n} k_j w_{i_j}$ genau dann, wenn $\varphi^{-1}(w) = \sum_{j_1}^{n} k_j \varphi^{-1}(w_{i_j}) = \sum_{j_1}^{n} k_j \varphi(v_{i_j})$. Aus der Eindeutigkeit der Darstellung bezüglich der Basis $\{v_i \mid i \in I\}$ folgt damit die Eindeutigkeit der Darstellung bezüglich $\{w_i \mid i \in I\}$, d. h. wieder mit Satz 1.4.7, dass es sich um eine Basis handelt. □

Ein Isomorphismus ist soviel wie eine Umbenennung der Elemente des Vektorraums und überträgt alle aus der Vektorraumsstruktur definierbaren Eigenschaften. Insbesondere bildet er also eine Basis auf eine Basis ab und kann also nur zwischen Vektorräumen gleicher Dimension bestehen!

Folgerung 1.5.3 *Genau dann gibt es einen K-Vektorraum-Isomorphismus $\varphi : V \to W$, wenn $\dim_K V = \dim_K W$.*

Folgerung 1.5.4 *Eine lineare Abbildung $\varphi : V \to W$ ist durch die Bilder einer Basis festgelegt.*

Eine *angeordnete Basis* $(v_1, \ldots, v_n)$ ist eine Basis $\{v_1, \ldots, v_n\}$ ohne Doppelnennungen zusammen mit einer festen Reihenfolge der Elemente (nämlich der Anordnung, in der die Elemente als Komponenten des n-Tupels auftreten).[7]

[7] Wenn man die Basiselemente durch die Variablen $v_1, \ldots, v_n$ bezeichnet, scheint eine Reihenfolge bereits durch die Anordnung der Indizes gegeben zu sein; daher wirkt die Definition auf den ersten Blick vielleicht überflüssig. In einer konkreten Situation steht aber $\{v_1, \ldots, v_n\}$ z. B. für die Menge $\{(1, 1, 2), (-2, 3, 3), (2, -2, 0)\}$, auf der keine vorgegebene Reihenfolge erkennbar ist.

Folgerung 1.5.5 *Sei V ein n-dimensionaler K-Vektorraum. Dann wird durch jede angeordnete Basis $B = (v_1, \ldots, v_n)$ ein Vektorraum-Isomorphismus $i_B : V \to K^n$, $v_i \mapsto e_i$ festgelegt. Dabei wird $v = k_1v_1 + \cdots + k_nv_n$ auf seine Koordinaten $(k_1, \ldots, k_n)$ bezüglich der Basis B abgebildet. Umgekehrt bestimmt jeder Vektorraum-Isomorphismus $i : V \to K^n$ eine angeordnete Basis B von V, nämlich $(i^{-1}(e_1), \ldots, i^{-1}(e_n))$, und es ist $i = i_B$.*

Beispiele: Sei nun stets $K = \mathbb{R}$ (wobei abgesehen von der geometrischen Anschauung die Überlegungen ebenso für jeden anderen Körper K gelten) und $\varphi : V \to W$ eine $\mathbb{R}$-lineare Abbildung zwischen endlich-dimensionalen $\mathbb{R}$-Vektorräumen $V = \mathbb{R}^n$ und $W = \mathbb{R}^m$. Dann ist φ festgelegt durch die Bilder der Standardbasis $\{e_1, \ldots, e_n\}$. Es ist nun üblich und günstig, die Elemente von V und W als Spaltenvektoren zu schreiben. Wir betrachten zunächst drei Spezialfälle:

- Sei $n = m = 1$. Dann ist $e_1 = 1$ und $\varphi(1) = \lambda \in \mathbb{R}$. Es gilt also
$$\varphi(r) = \varphi(r \cdot 1) = r \cdot \varphi(1) = \lambda \cdot r.$$
Die linearen Abbildungen $\mathbb{R} \to \mathbb{R}$ sind also genau die Multiplikationen mit reellen Zahlen.

- Sei n beliebig, $m = 1$ und $\varphi(e_1) = \lambda_1, \ldots, \varphi(e_n) = \lambda_n$. Es gilt dann also:
$$\varphi\left(\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}\right) = \varphi\Big(\sum_{i=1}^{n} r_i e_i\Big) = \sum_{i=1}^{n} r_i \varphi(e_i) = \sum_{i=1}^{n} r_i \lambda_i = \lambda_1 r_1 + \cdots + \lambda_n r_n$$
Die Urbilder der Elemente der Bildraums $\mathbb{R}$ bilden parallele, zu $(\lambda_1, \ldots, \lambda_n)$ senkrechte Hyperebenen im $\mathbb{R}^n$; man kann die Abbildung daher auffassen als die Projektion auf die Gerade durch den Ursprung in Richtung $(\lambda_1, \ldots, \lambda_n)$, skaliert (d. h. gestreckt oder gestaucht) um die Länge von $(\lambda_1, \ldots, \lambda_n)$, also den Faktor $\sqrt{\lambda_1^2 + \cdots + \lambda_n^2}$.

- Sei $n = 1$, m beliebig und $\varphi(1) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}$. Dann gilt
$$\varphi(r) = \varphi(r \cdot 1) = r \cdot \varphi(1) = r \cdot \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = \begin{pmatrix} r\mu_1 \\ \vdots \\ r\mu_m \end{pmatrix}$$
Das Bild von φ ist also die Gerade durch den Punkt $\varphi(1)$; die Abbildung φ bildet $\mathbb{R}$ unter Streckung bzw. Stauchung (Skalierung um die Länge von $\varphi(1)$) auf diese Gerade ab.

- Seien schließlich n und m beliebig und
$$\varphi(e_1) = \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{m1} \end{pmatrix}, \ \varphi(e_2) = \begin{pmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{m2} \end{pmatrix}, \ \ldots, \ \varphi(e_n) = \begin{pmatrix} \mu_{1n} \\ \mu_{2n} \\ \vdots \\ \mu_{mn} \end{pmatrix}$$
Dann ist
$$\varphi\left(\begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}\right) = \varphi\Big(\sum_{i=1}^{n} r_i e_i\Big) = \sum_{i=1}^{n} r_i \varphi(e_i) = \sum_{i=1}^{n} r_i \begin{pmatrix} \mu_{1i} \\ \vdots \\ \mu_{mi} \end{pmatrix} = \begin{pmatrix} \mu_{11} r_1 + \cdots + \mu_{1n} r_n \\ \vdots \\ \mu_{m1} r_1 + \cdots + \mu_{mn} r_n \end{pmatrix}$$

Um diese Abbildungen besser beschreiben zu können, führt man Matrizen ein.

Definition 1.5.6 (a) *Eine $(m \times n)$-Matrix über eine Körper K ist eine rechteckige Anordnung von mn Körperelementen a_{ij} für $i = 1, \dots, m$ („Zeilenindex“) und $j = 1, \dots, m$ („Spaltenindex“) in m Zeilen und n Spalten:*

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

Die Menge aller $(m \times n)$-Matrizen mit Einträgen aus K wird mit $\mathrm{Mat}_{m\times n}(K)$ bezeichnet.

Konvention: Wenn nicht explizit anders angegeben, werden die Einträge einer mit einem Großbuchstaben bezeichneten Matrix durch die entsprechenden Kleinbuchstaben beschrieben. Es hat also z. B. die Matrix C in der Regel Einträge c_{ij}, d. h.. $C = (c_{ij})_{\substack{i=1,\dots,n \\ j=1,\dots,m}}$.

(b) *Man definiert die Multiplikation einer $(m \times n)$-Matrix mit einem Spaltenvektor aus K^n durch die Formel*

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} a_{11}r_1 + a_{12}r_2 + \dots + a_{1n}r_n \\ a_{21}r_1 + a_{22}r_2 + \dots + a_{2n}r_n \\ \vdots \\ a_{m1}r_1 + a_{m2}r_2 + \dots + a_{mn}r_n \end{pmatrix}$$

Wichtige Beobachtung: Durch diese Definition ergibt sich, dass die linearen Abbildungen $K^n \to K^m$ genau die Multiplikationen (von links) mit $(m \times n)$-Matrizen sind, wobei die Spalten der Matrix die Bilder der Standardbasisvektoren sind.

Jeder linearen Abbildung $\varphi : K^n \to K^m$ lässt sich also eindeutig die $(m \times n)$-Matrix

$$\left(\varphi(e_1) \;\middle|\; \dots \;\middle|\; \varphi(e_m) \right)$$

zuordnen, wobei die $\varphi(e_i)$ hier Spaltenvektoren sind (was durch die senkrechten Striche angedeutet sein soll). Man sagt dafür auch, dass die lineare Abbildung durch die Matrix *dargestellt* wird. Oft werde ich auch stillschweigend die $(m \times n)$-Matrix A mit der Abbildung $K^n \to K^m$, $v \mapsto A \cdot v$ identifizieren.

1.6 Matrizenmultiplikation

Seien $\varphi : K^n \to K^m$ und $\psi : K^m \to K^l$ beides K-lineare Abbildungen, wobei φ durch eine $(m \times n)$-Matrix A und ψ durch eine $(l \times m)$-Matrix B dargestellt wird. Man stellt nun zunächst leicht fest, dass $\psi \circ \varphi : K^n \to K^l$ ebenfalls K-linear ist. Also wird $\psi \circ \varphi$ durch eine $(l \times n)$-Matrix C dargestellt. Wie hängt nun C mit A und B zusammen? Wie kann man C aus A und B ausrechnen? Dazu rechnet man $C \cdot v = (B \cdot A) \cdot v$ aus (siehe Formelkasten in Abbildung 1.1) und stellt fest, dass der (i, k)-Eintrag der Matrix C sich berechnet als

$$c_{ik} = \sum_{j=1}^{m} b_{ij}a_{jk} = \begin{pmatrix} b_{i1} & \dots & b_{im} \end{pmatrix} \cdot \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} = \begin{matrix} \\ \\ i\text{-te Zeile} \end{matrix} \begin{pmatrix} \dots & \dots & \dots \\ b_{i1} & \dots & b_{im} \\ \dots & \dots & \dots \end{pmatrix} \cdot \begin{pmatrix} & j\text{-te Spalte} & \\ \vdots & a_{1k} & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & a_{mk} & \vdots \end{pmatrix},$$

$$C \cdot \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = \psi(\varphi(\begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix})) = B \cdot (A \cdot \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix}) =$$

$$= B \cdot \begin{pmatrix} \sum_{i=1}^{n} a_{1i}k_i \\ \vdots \\ \sum_{i=1}^{n} a_{mi}k_i \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{m} b_{1i} \sum_{i=1}^{n} a_{ji}k_i \\ \vdots \\ \sum_{j=1}^{m} b_{li} \sum_{i=1}^{n} a_{ji}k_i \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{m} \sum_{i=1}^{n} b_{1i}a_{ji}k_i \\ \vdots \\ \sum_{j=1}^{m} \sum_{i=1}^{n} b_{li}a_{ji}k_i \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} (\sum_{j=1}^{m} b_{1j}a_{ji})k_i \\ \vdots \\ \sum_{i=1}^{n} (\sum_{j=1}^{m} b_{lj}a_{ji})k_i \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{m} b_{1j}a_{j1} & \dots & \sum_{j=1}^{m} b_{1j}a_{jn} \\ \vdots & & \vdots \\ \sum_{j=1}^{m} b_{lj}a_{j1} & \dots & \sum_{j=1}^{m} b_{lj}a_{jn} \end{pmatrix} \cdot \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix}$$

Abbildung 1.1: Matrizenmultiplikation

wobei hierfür die i-te Zeile von B mit der k-ten Spalte von A so multipliziert wird, wie im letzten Abschnitt definiert (dies heißt auch *Skalarprodukt* des i-ten Zeilenvektors von B mit dem k-ten Spaltenvektor von A, siehe Defintion 1.10.2).

Definition 1.6.1 *Das* Matrixprodukt $B \cdot A$ *einer* $(l \times m)$*-Matrix* B *mit einer* $(m \times n)$*-Matrix* A *ist die* $(l \times n)$*-Matrix* C *mit Einträgen* $c_{ik} = \sum_{j=1}^{m} b_{ij}a_{jk}$.

Das Matrixprodukt $B \cdot A$ ist also dann und nur dann definiert, wenn die Anzahl der Spalten von B gleich der Anzahl der Zeilen von A ist. Als Merkregel für die Dimensionen der Matrizen kann man sich „$(l \times m) \cdot (m \times n) = (l \times n)$“ einprägen; der gemeinsame mittlere Term verschwindet also.

Das Matrixprodukt $B \cdot A$ ist genau so definiert, dass $B \cdot A$ die Verknüpfung der durch B und A beschriebenen linearen Abbildungen beschreibt, d. h. es gilt

$$(A \cdot B) \cdot v = A \cdot (B \cdot v).$$

Bemerkung: In Definition 1.5.6 wurde das Produkt $A \cdot v$ einer $(m \times n)$-Matrix A mit einem Spaltenvektor $v \in K^n$ definiert. Nun ist solch ein Spaltenvektor v nichts anderes als eine $(n \times 1)$-Matrix. Somit ist also das Produkt $A \cdot v$ eigentlich doppelt definiert, aber man kann sich leicht anhand der Formeln davon überzeugen, dass beide Definitionen übereinstimmen.

Dass dies kein Zufall ist, sieht man folgendermaßen ein: Man kann einen Vektor $v \in K^n$ mit der linearen Abbildung $K^1 \to K^n$, $1 \mapsto v$ identifizieren, deren Matrix gerade der Spaltenvektor v ist (die Abbildung ist also die Multiplikation mit v). Die Verküpfung dieser Abbildung mit der durch A beschriebenen linearen Abbildung ist dann die lineare Abbildung $K^1 \to K^m$, welche 1 auf $A(v)$ abbildet. Die Matrix dieser Abbildung berechnet sich als das Matrixprodukt von A und v, ist aber andererseits der Spaltenvektor $A(v)$.

Beispiele

- $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \cdot \begin{pmatrix} -1 & 0 \\ 0 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 1\cdot(-1)+2\cdot 0+3\cdot 1 & 1\cdot 0+2\cdot 2+3\cdot 3 \\ 4\cdot(-1)+5\cdot 0+6\cdot 1 & 4\cdot 0+5\cdot 2+6\cdot 3 \end{pmatrix} = \begin{pmatrix} 2 & 13 \\ 2 & 28 \end{pmatrix}$
- Die Verküpfung „Spiegelung an der y-Achse ∘ Spiegelung an der x-Achse“ wird beschrieben durch
$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$
ergibt also die Matrix der Punktpsiegelung am Urpsrung.
- Eine Drehung um den Winkel α mit anschließender Drehung um den Winkel β ergibt insgesamt eine Drehung um $\alpha+\beta$. Aus der Berchnung des Matrixprodukts ergeben sich dadurch die Additionstheoreme für Sinus und Cosinus:
$$\begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \cdot \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad \begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix}$$
$$= \begin{pmatrix} \cos\alpha\cos\beta-\sin\alpha\sin\beta & -\sin\alpha\cos\beta-\cos\alpha\sin\beta \\ \cos\alpha\sin\beta+\sin\alpha\cos\beta & -\sin\alpha\sin\beta+\cos\alpha\cos\beta \end{pmatrix} =$$

Definition 1.6.2 *Die zur Identitätsabbildung* $\mathrm{id} : K^n \to K^n$ *gehörige Matrix ist die* Einheitsmatriz *genannte* $(n \times n)$*-Matrix* I_n*, deren Spalten (bzw. Zeilen) gerade die Standardbasisvektoren sind. Bei* Nullmatrizen *sind alle Einträge* 0*. Sie gehören zu den konstanten Nullabbildungen* $K^n \to K^m$*,* $v \mapsto 0$*, und werden oft kurz, aber uneindeutig, ebenfalls mit* 0 *geschrieben.*

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \qquad 0 = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

Kleiner Exkurs: Matrizenmultiplikationen spielen in vielen algorithmischen Anwendungen eine große Rolle; es ist daher interessant und nützlich, möglichst schnelle Verfahren zu finden. Das Verfahren, das der Definition folgt, läuft für $(n \times n)$-Matrizen in $O(n^3)$: pro Eintrag n Multiplikationen und $n-1$ Additionen. Für große Matrizen gibt es aber schnellere Verfahren: Das erste solche stammt 1969 von Volker Strassen[8] und läuft in $O(n^{2{,}807})$. Er wurde nach und nach verbessert; den letzten großen Schritt lieferte 1990 der Coppersmith-Winograd-Algorithmus[9] mit $O(n^{2{,}3737})$. Etwas überraschend kam 2010 nochmals eine Verbesserung durch Andrew Stothers; der derzeit letzte Stand ist ein Algorithmus von Virginia Vassilevska Williams aus dem Jahre 2011 mit einer Laufzeit von $O(n^{2{,}3727})$. Als untere Schranke hat man sicher $O(n^2)$, da n^2 Einträge auszurechnen sind; einige Forscher vermuten, dass diese untere Schranke optimal ist, also dass es Algorithmen in $O(n^2)$ gibt.

(Zu bedenken ist dabei, dass kleinere Exponenten wegen der in der O-Notation versteckten Konstanten evtl. nur für sehr große Matrizen Verbesserungen bringen; außerdem sagt die Laufzeit nichst über die Güte des Algorithmus hinsichtlich Stabilität (Fehleranfälligkeit) aus. Die Verbesserung des Exponenten in der dritten Nachkommastelle scheint zunächst vernachlässigbar, es ist aber bereits $1000^{2{,}3737} - 1000^{2{,}3727} \approx 10^5$; bei vielen Multiplikationen großer Matrizen kann sich also ein spürbarer Effekt ergeben.)

[8] Volker Strassen (∗ 1936), ehemaligere Student der Universität Freiburg, zuletzt Professor in Konstanz.
[9] nach Don Coppersmith (∗ ca. 1950) und Shmuel Winograd (∗ 1936), damals IBM.

Satz 1.6.3 *Die Matrizenmulitplikation ist assoziativ, aber i. a. nicht kommutativ (auch bei $(n \times n)$-Matrizen untereinander. Die Einheitsmatrizen sind neutrale Elemente in dem Sinn, dass $I_m \cdot = A$ und $A \cdot I_n = A$ für jede $(m \times n)$-Matrix A gelten. Nullmatrizen sind* absorbierende Elemente*, d. h. es gilt $0 \cdot A = A$ und $A \cdot 0 = A$ (für die Nullmatrix passender Größe, so dass also die Multiplikationen definiert sind).*

BEWEIS: Die nicht vorhandene Kommutativität sieht man z. B. an

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 0 & 1 \end{pmatrix}.$$

Die anderen Eigenschaften folgen daraus, dass sie auf Seite der zugehörigen Abbildungen gelten. □

Bemerkungen: Eine (1×1)-Matriz (a_{11}) kann man mit der Zahl a_{11} identifizieren. Die Multiplikation von (1×1)-Matrizen ist also kommutativ.

Abgesehen von der fehlenden Kommutativität gibt es noch andere Eigenschaften, welche die Matrizenmultiplikation von der Multiplikation z. B. reeller Zahlen unterscheidet. So gibt es sogenannte „nilpotente“ Elemente, das sind Matrizen $A \neq 0$ mit $A^n = 0$ für ein $n > 0$. Zum Beispiel gilt:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Insbesondere folgt für Matrizen aus $A \cdot B = 0$ nicht $A = 0$ oder $B = 0$!

Abbildungen $\varphi, \psi : K^n \to K^m$ kann man addieren durch $(\varphi + \psi)(v) := \varphi(v) + \psi(v)$ und skalar multiplizieren durch $(k \cdot \varphi)(v) := k \cdot \varphi(v)$; sie bilden dadurch einen K-Vektorraum $\mathrm{Abb}(K^n, K^m)$, mit den linearen Abbildungen als Untervektorraum $\mathrm{Lin}(K^n, K^m)$. Man kann nun die Addition und Skalarmultiplikation mittels der Identifikation von linearen Abbildungen und Matrizen in 1.5.6 (b) auf Matrizen ausdehnen, so dass die Menge $\mathrm{Mat}_{m\times n}(K)$ zu einem zu $\mathrm{Lin}(K^n, K^m)$ isomorphen K-Vektorraum wird. Man kann nun leicht nachrechnen, dass die folgende Definition die *Matrizenaddition* und die *Skalarmultiplikation von Matrizen* beschreibt:

Definition 1.6.4 *Seien A und B $(m \times n)$-Matrizen über K und $k \in K$. Dann ist*

$$A + B = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{pmatrix} := \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{pmatrix}$$

$$k \cdot A = k \cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} := \begin{pmatrix} k \cdot a_{11} & \dots & k \cdot a_{1n} \\ \vdots & & \vdots \\ k \cdot a_{m1} & \dots & k \cdot a_{mn} \end{pmatrix}$$

Es gelten nun die folgenden Eigenschaften:

Satz 1.6.5 (a) *Die $\mathrm{Mat}_{m\times n}(K)$ bilden einen mn-dimensionalen, zu $\mathrm{Lin}(K^n, K^m)$ isomorphen K-Vektorraum; insbesondere sind die $(m \times n)$-Matrizen bezüglich der Addition eine*

kommutative Gruppe, deren neutrales Element die $(m \times n)$-Nullmatrix ist. Die Standardbasis besteht aus Matrizen E_{ij}, deren (i,j)-Eintrag 1 ist und alle anderen Einträge 0. Jede Aufzählung der Standardbasis liefert einen Vektorraum-Isomorphismus $\mathrm{Mat}_{m\times n}(K) \to K^{mn}$.

(b) *Es gelten die Distributivgesetze, d. h immer, wenn die Operationen definiert sind, gilt* $A \cdot (B_1 + B_2) = (A \cdot B_1) + (A \cdot B_2)$ *bzw.* $(B_1 + B_2) \cdot A = (B_1 \cdot A) + (B_2 \cdot A)$.

(c) *Die quadratischen $(n \times n)$-Matrizen* $\mathrm{Mat}_{n\times n}(K)$ *bilden mit Matrizenaddition und -multiplikation einen nicht-kommutativen Ring mit Eins I_n.*

(d) *$k \cdot I_n$ ist die $(n \times n)$-„Diagonalmatrix“ mit Einträgen k auf der Hauptdiagonale von links oben nach rechts unten und Einträgen 0 an allen anderen Stellen. Es gilt dann $k \cdot A = (k \cdot I_n) \cdot A = A \cdot (k \cdot I_n)$. Die Skalarmultiplikation vertauscht mit der Matrizenmultiplikation, d. h. es gilt $k \cdot (A \cdot B) = (k \cdot A) \cdot B = A \cdot (k \cdot B)$ (sofern das Produkt $A \times B$ definiert ist).*

Beweis: Manche Eigenschaften sind offensichtlich, weil sie auf der Seite der Abbildungen gelten; die anderen rechnet man anhand der Formeln nach. □

1.7 Basiswechsel und invertierbare Matrizen

Definition 1.7.1 *Eine $(n \times n)$-Matrix A über K heißt* invertierbar*, wenn die zugehörige lineare Abbildung $\varphi : K^n \to K^n$ invertierbar ist, d. h. wenn eine $(n \times n)$-Matrix A^{-1} existiert (nämlich die Matrix zu φ^{-1}) mit*

$$A \cdot A^{-1} = A^{-1} \cdot A = I_n.$$

Aus Folgerung 1.5.5 ergibt sich, dass A genau dann invertierbar ist, wenn die Spaltenvektoren von A, also $\varphi(e_1), \ldots, \varphi(e_n)$, eine Basis von K^n bilden. Die Umkehrabbildung φ^{-1} ist dann durch $\varphi(e_i) \mapsto e_i$ festgelegt.

Offensichtlich ist A^{-1} selbst wieder invertierbar und es gilt $(A^{-1})^{-1} = A$.

Ziel dieses Abschnitts ist es nun, lineare Abbildungen zwischen beliebigen endlich dimensionalen Vektorräumen durch Matrizen zu beschreiben. Da beliebige Vektorräume i. a. keine ausgezeicheten Standardbasen haben, wird es – abhängig von gewählten Basen – verschiedene darstellenden Matrizen geben, und ein Hauptproblem wird sein zu verstehen, wie diese miteinander zusammenhängen.

Definition 1.7.2 *Sei V ein n-dimensionaler und W ein m-dimensionaler K-Vektorraum und $\varphi : V \to W$ eine K-lineare Abbildung. Sei außerdem $v_1, \ldots, v_n$ eine angeordnete Basis B von V und $w_1, \ldots, w_m$ eine angeordnete Basis B' von W. Nach Folgerung 1.5.5 legen B und B' Isomorphismen $i_B : V \to K^n$ und $i_{B'} : W \to K^m$ fest, so dass sich folgendes Diagramm ergibt:*

$$K^n \xleftarrow{i_B} V \xrightarrow{\varphi} W \xrightarrow{i_{B'}} K^m$$

Die Matrix von φ bezüglich der Basen B und B' *wird mit ${}_{B'}\varphi_B$ bezeichnet und ist die Matrix der Abbildung $i_{B'} \circ \varphi \circ i_B^{-1} : K^n \to K^m$, d. h. die Spaltenvektoren sind die Koordinaten von $\varphi(v_1), \ldots, \varphi(v_n)$ bezüglich der Basis B'.*

Falls $V = W$ und $B = B'$, schreibt man kurz φ_B für ${}_B\varphi_B$.

Satz 1.7.3 *Seien V, W, X K-Vektorräume mit angeordneten Basen B, B', B'' und seien $\varphi : V \to W$ und $\psi : W \to X$ lineare Abbildungen. Dann gilt*

$$ {}_{B''}(\psi \circ \varphi)_B \;=\; ({}_{B''}\psi_{B'}) \cdot ({}_{B'}\varphi_B) $$

BEWEIS: ${}_{B''}(\psi \circ \varphi)_B$ ist die Matrix von $i_{B''} \circ (\psi \circ \varphi) \circ i_B^{-1} = i_{B''} \circ \psi \circ i_{B'}^{-1} \circ i_{B'} \circ \varphi \circ i_B^{-1}$, was gerade das Produkt der Matrix von $i_{B''} \circ \psi \circ i_{B'}^{-1}$ mit der Matrix von $i_{B'} \circ \varphi \circ i_B^{-1}$ ist, also $({}_{B''}\psi_{B'}) \cdot ({}_{B'}\varphi_B)$. □

Folgerung 1.7.4 *Sei wieder $\varphi : V \to W$ lineare, und seien B_1, B_2 angeordnete Basen von V, B_1', B_2' angeordnete Basen von W. Dann gilt*

$$ {}_{B_2'}\varphi_{B_2} = ({}_{B_2'}\mathrm{id}_{W\,B_1'}) \cdot ({}_{B_1'}\varphi_{B_1}) \cdot ({}_{B_1}\mathrm{id}_{V\,B_2}) $$

Spezialfall $V = W$ und $B_i' = B_i$: Dann gilt

$$ \varphi_{B_2} = ({}_{B_2}\mathrm{id}_{V\,B_1}) \cdot \varphi_{B_1} \cdot ({}_{B_1}\mathrm{id}_{V\,B_2}) = ({}_{B_1}\mathrm{id}_{V\,B_2})^{-1} \cdot \varphi_{B_1} \cdot ({}_{B_1}\mathrm{id}_{V\,B_2}). $$

BEWEIS: Der erste Teil folgt direkt aus dem Satz, da $\varphi = \mathrm{id}_W \circ \varphi \circ \mathrm{id}_V$. Wegen $({}_{B_2}\mathrm{id}_{V\,B_1}) \cdot ({}_{B_1}\mathrm{id}_{V\,B_2}) = {}_{B_2}(\mathrm{id}_V \circ \mathrm{id}_V)_{B_2} = {}_{B_2}\mathrm{id}_{V\,B_2} = I_{\dim V}$ folgt auch die rechte Seite der Gleichung im Spezialfall. □

Die Matrizen ${}_{B_2'}\mathrm{id}_{W\,B_1'}$ und ${}_{B_1}\mathrm{id}_{V\,B_2}$ heißen *Basiswechselmatrizen*. Sie sind also stets invertierbar mit $({}_{B_1}\mathrm{id}_{V\,B_2})^{-1} = {}_{B_2}\mathrm{id}_{V\,B_1}$.

Wie rechnet man die Basiswechselmatrizen aus? Ist die Basis $B_1 = (v_1, \ldots, v_n)$ von V gegeben und ist v' der j-te Vektor in B_2, so muss man also die Koeffizienten a_{ij} mit $v' = a_{1j}v_1 + \cdots + a_{nj}v_n$ berechnen; diese stehen als j-te Spalte in der Basiswechselmatrix ${}_{B_1}\mathrm{id}_{V\,B_2}$. Wenn die Basiselemente als Vektoren in K^n gegeben sind (also mit ihren Koordinaten bezüglich der Standardbasis), dann ergibt die Gleichung ein lineares Gleichungssystem, das z. B. nach dem Gauß-Verfahren (siehe folgender Abschnitt 1.8) gelöst werden kann. Auch das Invertieren von Matrizen geschieht am besten mit dem Gauß-Verfahren. Besonders einfach ist es, wenn $V = K^n$ und B_1 die Standardbasis ist: Dann besteht die Basiswechselmatrix ${}_{B_1}\mathrm{id}_{V\,B_2}$ aus den Vektoren von B_2 als Spaltenvektoren.

Beispiele:

- ...
- Ein weiteres Beispiel findet sich bei der Diagonalisierung einer Drehung über den komplexen Zahlen auf Seite 23.
- *Was passiert, wenn der Basiswechsel in einer Umordnung der Basis besteht?*
  Wenn B die Basis $(v_1, \ldots, v_n)$ ist, wird eine Umordnung beschrieben durch eine *Permutation* der Indizes, also eine Bijektion $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$, wobei die neu angeordnete Basis B_σ dann $(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$ ist.[10]
  Die Basiswechselmatrix $M(\sigma) := {}_{B_\sigma}\mathrm{id}_B$ hat nun Einträge 1 an den Stellen $(i, \sigma(i))$ und 0 an allen anderen Stellen. Solche Matrizen heißen auch *Permutationsmatrix*: Sie sind quadratische Matrizen, die in jeder Zeile und in jeder Spalte genau eine 1 haben

[10] Bei dieser Version gibt σ also an, welcher Vektor an die jeweilige Stelle gesetzt wird, d. h. $\sigma(2) = 3$ bedeutet, dass v_3 in der neu angeordneten Basis an zweiter Stelle steht. Alternativ könne man als neu angeordnete Basis $(v_{\sigma^{-1}(1)}, \ldots, v_{\sigma^{-1}(n)})$ nehmen. Dann würde σ angeben, an welche Stelle der jeweilige Vektor geschoben wird, d. h. $\sigma(2) = 3$ würde bedeuten, dass v_2 in der neu angeordneten Basis an dritter Stelle käme.

und sonst überall 0. Die Inverse zu $M(\sigma)$ ist $M(\sigma^{-1})$, also die Permutationsmatrix mit Einträge 1 an den Stellen $(\sigma(i), i)$, also die an der Diagonalen gespiegelte Matrix.

Beispiel: Sei $n = 3$ und $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$. Dann ist

$$M(\sigma) = {}_{B_\sigma}\mathrm{id}_B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \text{ und } M(\sigma^{-1}) = {}_{B}\mathrm{id}_{B_\sigma} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

(Kleiner Vorgriff auf Abschnitt 2.1: Die Permutationen von $\{1, \ldots, n\}$ bilden eine Gruppe $\mathrm{Sym}(n)$, die *symmetrische Gruppen*, und die Abbildunge $\sigma \mapsto M(\sigma)$ ist ein Gruppenhomomorphismus von $\mathrm{Sym}(n)$ in die multiplikative Gruppe der invertierbaren $(n \times n)$-Matrizen.)

Spezialfall: Transpositionen sind spezielle Permutatione, die nur zwei Elemente vertauschen (und damit selbst-invers sind). Die Transposition τ, welche die Elemente i und j vertauscht, schreibt man auch (ij). Der Lesbarkeit halber schreibe ich $M_{(ij)}$ für $M((ij))$. Es gilt dann (alle nicht aufgeführten Einträge sind gleich 0):

$$M_{(ij)} = M_{(ij)}^{-1} = \left(\begin{array}{ccccccc} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ \hline & & & & 1 & & \\ \hline & & & 1 & & & \\ & & & & & \ddots & \\ & & & & & & 1 \\ \hline & & 1 & & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{array}\right) \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ i\text{-te Zeile} \\ \\ \\ \end{array} \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ j\text{-te Zeile} \end{array}$$

Es ist also etwa (zweite und dritte Zeile und Spalte jeweils vertauschen!)

$$M_{(23)} \cdot \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 0 & 1 & 2 \\ 3 & 4 & 5 & 6 \end{pmatrix} \cdot M_{(32)} = \begin{pmatrix} 1 & 3 & 2 & 4 \\ 9 & 1 & 0 & 2 \\ 5 & 7 & 6 & 8 \\ 3 & 5 & 4 & 6 \end{pmatrix}.$$

Ein Ziel der linearen Algebra besteht darin, zu einer gegebenen linearen Abbildung $\varphi : V \to V$ eine Basis B zu finden, so dass die Matrix φ_B möglichst „schön“ ist. Hierzu gibt es eine ganze Reihe von Ergebnissen über sogenannte Normalformen von Matrizen. „Besonders schön“ ist eine Matrix in Diagonalgestalt, also von der Form

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Für die Basisvektoren $v_1, \ldots, v_n$ gilt dann $\varphi(v_i) = \lambda v_i$ und für beliebige Vektoren $\varphi(a_1 v_1 + \cdots + a_n v_n) = \lambda a_1 v_1 + \cdots + \lambda a_n v_n$.

Definition 1.7.5 *Ein Vektor $v \neq 0$ heißt* Eigenvektor *der linearen Abbildung $\varphi : V \to V$ zum* Eigenwert *$\lambda \in K$, falls $\varphi(v) = \lambda v$.*

Der Idealfall besteht also darin, dass man zu einer linearen Abbildung eine Basis aus Eigenvektoren findet. (Wenn man weiß, dass λ ein Eigenwert ist und φ durch die Matrix A

beschrieben ist, kann man die Eigenvektoren durch Lösen des linearen Gleichungssystemes $A \cdot x = \lambda x$ mit Unbekannten Koeffizienten für x finden. Jedes skalare Vielfache eines Eigenvektors ($\neq 0$) ist wieder ein Eigenvektor. Die Eigenwerte wiederum kann man als Nullstellen des sogenannten *charakteristischen polynoms* bestimmen.)

Im allgemeinen findet man aber keine Basis aus Eigenvektoren. Es gibt zwei Hinderungsgründe:

(1) *Drehungen im $\mathbb{R}^2$ haben i. a. keine Eigenvektoren, wie geometrisch sofort ersichtlich ist.* (Nur wenn der Drehwinkel ein ganzzahliges Vielfaches von 180° ist, gibt es Eigenvektoren in $\mathbb{R}^2$.)

Diese Problem lässt sich dadurch beheben, dass man den Körper erweitern, hier zu den komplexen Zahlen $\mathbb{C}$. So hat z. B. die Drehung um 90° bezüglich der Standardbasis die Matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ – ohne Eigenvektoren in $\mathbb{R}^2$ – aber als Matrix über den komplexen Zahlen sind $\left(\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1 \\ -i \end{smallmatrix}\right)$ zwei linear unabhängige Eigenvektoren zu den Eigenwerten $-i$ und i, d. h. bezüglich der aus diesen beiden Vektoren gebildeten Basis ergibt sich die Diagonalform $\left(\begin{smallmatrix} -i & 0 \\ 0 & i \end{smallmatrix}\right)$.

Man kann an diesem Beispiel noch einmal schön den Basiswechsel nachvollziehen: Da eine der basen die Standardbasis ist, besteht eine der beiden Basiswechselmatrizen aus den Vektoren der anderen Basis als Spalten und die andere Basiswechselmatrix ist deren Inverse:

$$\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \text{ und } \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2i} \\ \frac{1}{2} & -\frac{1}{2i} \end{pmatrix};$$

man kann auch nachrechnen, dass diese Matrix tatsächlich die Koeffizienten der Standardbasis bezüglich der neuen Basis enthält, da

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ i \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 \\ -i \end{pmatrix} \text{ und } \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2i}\begin{pmatrix} 1 \\ i \end{pmatrix} - \frac{1}{2i}\begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

Auch den Basiswechsel lässt sich nachrechnen; es gilt:

$$\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}\begin{pmatrix} \frac{1}{2} & \frac{1}{2i} \\ \frac{1}{2} & -\frac{1}{2i} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ und } \begin{pmatrix} \frac{1}{2} & \frac{1}{2i} \\ \frac{1}{2} & -\frac{1}{2i} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}$$

(2) *Scherungen im $\mathbb{R}^2$ haben i. a. (bis auf skalare Vielfache) nur einen Eigenvektor.*

Diese Problem kann nicht durch Vergörßerung des Körpers behoben werden; eine Scherung wie z. B. $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ bildet einen Vektor $v = ae_1 + be_2$ auf $ae_1 + b(e_2 + e_1)$ ab. Man kann leicht nachrechnen, dass nur die skalaren Vielfachen von e_1 Eigenvektoren sind, also wenn $b = 0$.

Man kann zeigen, dass die über $\mathbb{C}$ der einzige Hinderungsgrund ist; man kann durch geeignete Basiswahl die sogenannte Jordan'sche Normalform erreichen, bei der die Matrix aus Teilmatrizen der folgenden Form ausgebaut ist:

$$\begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \dots & \dots & 0 & \lambda \end{pmatrix}$$

1.8 Lineare Gleichungssysteme

Definition 1.8.1 *Sei $\varphi : V \to W$ eine lineare Abbildung. Das* Bild von φ *ist definiert als* $\mathrm{Bild}(\varphi) := \{\varphi(v) \mid v \in V\}$*; der* Kern von φ *als* $\mathrm{Kern}(\varphi) := \{v \in V \mid \varphi(v) = 0\}$.

Das Bild wird ebenso für beliebige Abbildungen definiert. Bild und Kern werden auch (nach dem englischen *image* und *kernel*) als $\mathrm{im}(\varphi)$ und $\ker(\varphi)$ bezeichnet.

Satz 1.8.2 [11]
(a) *Kern und Bild einer linearen Abbildung $\varphi : V \to W$ sind Unterräume von V bzw. W.*
(b) *Falls $w_0 = \varphi(v_0) \in \mathrm{Bild}(\varphi)$, so ist $\varphi^{-1}[w_0] = v_0 + \mathrm{Kern}(\varphi) := \{v_0 + v \mid v \in \mathrm{Kern}(\varphi)\}$. Mit anderen Worten, es gilt g $\varphi(v) = \varphi(v') \iff v - v' \in \mathrm{Kern}(\varphi)$.*

Beweis: (a) Da $\varphi(0_V) = 0_W$ ist $0_V \in \mathrm{Kern}(\varphi)$ und $0_W \in \mathrm{Bild}(\varphi)$.

Seien $w_1 = \varphi(v_1)$ und $w_2 = \varphi(v_2)$ in $\mathrm{Bild}(\varphi)$. Dann sind $w_1 + w_2 = \varphi(v_1 + v_2)$ und $k \cdot w_1 = \varphi(k \cdot v_1)$ ebenfalls in $\mathrm{Bild}(\varphi)$, also ist $\mathrm{Bild}(\varphi)$ ein Untervektorraum.

Seien $v_1, v_2 \in \mathrm{Kern}(\varphi)$. Dann ist $\varphi(v_1 + v_2) = \varphi(v_1) + \varphi(v_2) = 0 + 0 = 0$ und $\varphi(k \cdot v_1) = k \cdot \varphi(v_1) = k \cdot == 0$. Also ist auch $\mathrm{Kern}(\varphi)$ ein Untervektorraum.

(b) Es ist $\varphi(v) = \varphi(v') \iff \varphi(v - v') = \varphi(v) - \varphi(v') = 0 \iff v - v' \in \mathrm{Kern}(\varphi)$. □

Für ein $w \in W$ gibt es also zwei mögliche Fälle: Entweder $w \notin \mathrm{Bild}(\varphi)$ und $\varphi^{-1}[w] = \emptyset$; oder $w \in \mathrm{Bild}(\varphi)$ und $\varphi^{-1}[w]$ ist eine sogenannte „Nebenklasse“ $v + \mathrm{Kern}(\varphi)$ von $\mathrm{Kern}(\varphi)$ mit $\varphi(v) = w$.

Natürlich ist φ surjektiv, wenn $\mathrm{Bild}(\varphi) = W$ (gilt für beliebige Abbildungen $\varphi : V \to W$).

Folgerung 1.8.3 *φ ist injektiv $\iff \mathrm{Kern}(\varphi) = \{0\} \iff \dim \mathrm{Kern}(\varphi) = 0$.*
Wenn W endlich-dimensional ist, so ist φ surjektiv $\iff \dim \mathrm{Bild}(\varphi) = \dim W$.

Beispiel

⋮

Satz 1.8.4 *Sei $\varphi : V \to W$ linear. Dann gilt $\dim \mathrm{Kern}(\varphi) + \dim \mathrm{Bild}(\varphi) = \dim V$.*

Beweis: [12] Sei $l = \dim \mathrm{Kern}(\varphi)$ und $n = \dim V$ und wähle eine Basis $\{v_1, \ldots, v_l\}$ von $\mathrm{Kern}(\varphi)$. Diese ist eine lineare unabhängige Teilmenge von V, kann also zu einer maximal linear unabhängigen Teilmenge $\{v_1 \ldots, v_l, v_{l+1}, \ldots, v_n\}$ ergänzt werden, d. h. zu einer Basis von V. Zu zeigen ist also $n - l = \dim \mathrm{Bild}(\varphi)$, indem gezeigt wird, dass $\varphi(v_{l+1}), \ldots, \varphi(v_n)$ eine Basis ohne Doppelnennungen von $\mathrm{Bild}(\varphi)$ ist.

Sei $w = \varphi(a_1 v_1 + \cdots + a_n v_n) \in \mathrm{Bild}(\varphi)$. Dann ist $w = a_1\varphi(v_1) + \cdots + a_n\varphi(v_n) = a_{l+1}\varphi(v_{1+1}) + \cdots + a_n\varphi(v_n)$, da $\varphi(v_1) = \cdots = \varphi(v_l) = 0$. Also ist $\varphi(v_{l+1}), \ldots, \varphi(v_n)$ ein Erzeugendensystem von $\mathrm{Bild}(\varphi)$.

Zu zeigen bleibt die lineare Unabhängigkeit, mit Lemma 1.4.2: Sei also $0 = b_{l+1}\varphi(v_{l+1}) + \cdots + b_n\varphi(v_n) = \varphi(b_{l+1}v_{l+1} + \cdots + b_n v_n) \in \mathrm{Kern}(\varphi)$. Da $v_1, \ldots, v_l$ eine Basis von $\mathrm{Kern}(\varphi)$ ist, gibt es $b_1, \ldots, b_l$ mit $b_{l+1}v_{l+1} + \cdots + b_n v_n = b_1 v_1 + \cdots + b_l v_l$, oder $(-b_1)v_1 + \cdots + (-b_l)v_l + b_{l+1}v_{l+1} + \cdots + b_n v_n = 0$. Aus der linearen Unabhängigkeit der Basis $v_1, \ldots, v_n$ folgt nun aber $b_1 = \cdots = b_n = 0$. □

[11] Notation: Ist $f : A \to B$ eine beliebige Abbildung und $b \in B$, so bezeichnet $f^{-1}[b]$ die Menge $\{a \in A \mid f(a) = b\}$. Meist wird dafür $f^{-1}(b)$ geschrieben. Falls f bijektiv ist und die Umkehrfunktion $f^{-1} : B \to A$ existiert, so wird die Schreibweise mit runden Klammern aber zweideutig. Mit der exakteren Schreibweise gilt $f^{-1}[b] = \{f^{-1}(b)\}$.

[12] Für endlich-dimensionales V; der Beweis funktioniert mit den entsprechenden Modifikationen aber auch für unendlich-dimensionale Vektorräume.

Folgerung 1.8.5 *Wenn $\varphi : V \to W$ bijektiv ist, dann gilt $\dim V = \dim W$. Wenn $\dim V = \dim W$ endlich ist, dann ist φ genau dann injektiv, wenn surjektiv (und damit genau dann, wenn bijektiv).*

BEWEIS: Wenn $\varphi : V \to W$ bijektiv ist, so ist $\dim \mathrm{Kern}(\varphi) = 0$, da φ injektiv, und $\dim \mathrm{Bild}(\varphi) = W$, da φ surjektiv, also $\mathrm{Bild}(\varphi) = W$. Es folgt $\dim V = \dim \mathrm{Kern}(\varphi) + \dim \mathrm{Bild}(\varphi) = 0 + \dim W$.

Wenn $\dim V = \dim W$ endlich und φ injektiv, dann ist $\dim W = \dim V = \dim \mathrm{Kern}(\varphi) + \dim \mathrm{Bild}(\varphi) = 0 + \dim \mathrm{Bild}(\varphi)$, also φ surjektiv.

Wenn $\dim V = \dim W$ endlich und φ surjektiv, dann ist $\dim \mathrm{Kern}(\varphi) = \dim V - \dim \mathrm{Bild}(\varphi) = \dim W - \dim \mathrm{Bild}(\varphi) = 0$, also φ injektiv. □

Definition 1.8.6 *Ein lineares Gleichungssystem (über einem Körper K, meist $K = \mathbb{R}$) besteht aus linearen Gleichungen*

$$\begin{array}{ccccccc} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + \cdots + & a_{1n} \cdot x_n & = & b_1 \\ & & & \vdots & & \vdots & \\ a_{m1} \cdot x_1 & + & a_{m2} \cdot x_2 & + \cdots + & a_{mn} \cdot x_n & = & b_m \end{array}$$

mit $a_{ij}, b_i \in K$ und Unbekannten $x_1, \ldots, x_n$. Eine Lösung des Gleichungssystems besteht aus Werten $k_1, \ldots, k_n \in K$, welche gleichzeitig alle m Gleichungen erfüllen.

Das zugehörige homogene (lineare) Gleichungssystem ist

$$\begin{array}{ccccccc} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + \cdots + & a_{1n} \cdot x_n & = & 0 \\ & & & \vdots & & \vdots & \\ a_{m1} \cdot x_1 & + & a_{m2} \cdot x_2 & + \cdots + & a_{mn} \cdot x_n & = & 0 \end{array} .$$

Offenbar kann man das lineare Gleichungssystem in einer Matrix zusammenfassen als

$$A \cdot x = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} = b$$

Eine Lösung des homogenen Gleichungssystems $A \cdot x = 0$ ist dann ein Vektor aus dem Kern von A; die *Lösungsmenge* des homogenen Gleichungssystems (d. h. die Menge aller Lösungen) ist genau $\mathrm{Kern}(A)$.

Für das allgemeine Gleichungssystem $A \cdot x = b$ gibt es die beiden bereits besprochenen Möglichkeiten: Entweder $b \notin \mathrm{Bild}(A)$ und es gibt keine Lösung, oder $b \in \mathrm{Bild}(A)$ und die Lösungsmenge besteht aus einer Nebenklasse $c + \mathrm{Kern}(A)$, wobei c irgendeine Lösung des Gleichungssystems ist. Um die Lösungsmenge des Gleichungssytems zu bestimmen, muss man also eine sogenannte *spezielle Lösung* c finden – sofern sie existiert! – und den Kern von A bestimmen. Ist $v_1, \ldots, v_l$ eine Basis des Kerns, so besteht die Lösungsmenge also aus allen Vektoren der Form $c + k_1 v_1 + \cdots + k_l vl$ mit $k_i \in K$ („die *allgemeine Lösung*“).

Definition 1.8.7 *Der Rang einer $(m \times n)$-Matrix A, $\mathrm{rg}(A)$, ist die Dimension des Bildes von A als linearer Abbildung $K^n \to K^m$, d. h. die Dimension des von den Spalten $A \cdot e_1, \ldots, A \cdot e_n$ von A erzeugten Unterraums.*

Nach Definition ist $\operatorname{rg}(A) \leqslant m$ und $= m$ genau dann, wenn A surjektiv ist. Außerdem gilt $n = \dim \operatorname{Kern}(A) + \operatorname{rg}(A)$ nach Satz 1.8.4.

Folgerung 1.8.8 *Falls A die Matrix eines homogenen linearen Gleichunssystemes mit m Gleichungen und n Unbekannten ist, dann ist die Dimension des Lösungsraums $n - \operatorname{rg}(A)$.*

Das Gauß-Verfahren zum Lösen linearer Gleichungssysteme

Die Idee des Verfahrens besteht darin, das Gleichungssystem bzw. die Matrix durch eine Reihe „elementarer Umformungen“, die die Löungsmenge nicht oder in einer kontroliierten Weise ändern, in eine „schöne Form“ zu bringen, aus der man die Lösungsmenge leicht errechnen kann.

Hier betrachten wir drei Arten von elementaren Umformungen. Jede der Umformungen entsprich der Multiplikation von links mit einer invertierbaren Matrix M. Dann gilt

$$A \cdot v = M^{-1}MA \cdot v = b \iff MA \cdot v = M \cdot b$$

und wegen $M \cdot 0 = 0$ ist insbesondere $\operatorname{Kern}(MA) = \operatorname{Kern}(A)$, d. h. die Lösungsmenge ändert sich nicht, wenn A und c gleichermaßen umgeformt werden.

(1) *Vertauschung der i-ten mit der j-ten Gleichung bzw.*
Vertauschung der i-ten mit der j-ten Zeile der Matrix A und des Vektors b.

Dies entspricht der Multiplikation von links mit der Matrix $M_{(ij)} = M_{(ij)}^{-1}$ (siehe Seite 22).

(2) *Addition des k-fachen der j-ten Gleichung zur i-ten Gleichung bzw.*
Addition des k-fachen der j-ten Zeile der Matrix A und des Vektors b zur i-ten Zeile.

Dies entspricht der Multiplikation von links mit der Matrix $E_{ij}(k) = I_m + k \cdot E_{ij}$. ($E_{ij}$ ist die Standardbasenmatrix aus Satz 1.6.5). Man sieht leicht ein, dass $E_{ij}(k)^{-1} = E_{ij}(-k)$.

(3) *Multiplikation der i-ten Gleichung mit $k \neq 0$ bzw.*
Multiplikation der i-ten Zeile der Matrix A und des Vektors b mit $k \neq 0$.

Dies entspricht der Multiplikation von links mit der Matrix $E_i(k) = I_m + (k-1) \cdot E_{ii}$. Man sieht wiederum leicht ein, dass $E_i(k)^{-1} = E_i(k^{-1})$.

Ergänzend können auch Operationen auf den Spalten der Matrix vorgenommen werden (z. B. Vertauscheungen der Spalten, die dann den entsprechenden Vertauschungen der Unbekannten entsprechen). Diese sind aber nur zur Verbesserung von Algorithmen hinsichtlich Stabilität notwendig.

Definition 1.8.9 *Eine Matrix A ist in* Zeilenstufenform, *falls sie Einträge $a_{1j_1} \neq 0, \ldots, a_{rj_r} \neq 0$ mit $j_1 < \cdots < j_r$ hat, so dass*

$$a_{ij} = 0, \text{ falls} \begin{cases} i > i_k \text{ und } j \leqslant j_k \\ \text{oder } j < j_1 \\ \text{oder } i > i_r \end{cases} .$$

Die Elemente a_{ij_i} heißen Pivot-Elemente, *die Spalten $j_1, \ldots, j_r$* Pivot-Spalten.

Schematisch angedeutet sieht eine Zeilenstufenform (mit $r = 3$) wie folgt aus; $*$ steht für

beliebige Elemente:

$$\begin{pmatrix} 0 & a_{1j_1} & * & * & * & * & * \\ 0 & 0 & a_{2j_2} & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & a_{3j_3} & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Satz 1.8.10 (a) *Jede Matrix kann durch elementare Umformungen der Art (1) und (2) in Zeilenstufenform gebracht werden.*

(b) *Jede invertierbare Matrix kann durch elementare Umformungen der Art (1) und (2) in eine Diagonalmatrix und durch elementare Umformungen der Art (1), (2) und (3) in die Identitätmatrix überführt werden.*

Beweis: Für (a) gibt es den in Abbildung 1.2 dargestellten Algorithmus, das sogenannte *Gauß-Verfahren*. Die Matrix wird spaltenweise von links nach rechts und zeilenweise von oben nach unten so abgearbeitet, dass die gewünschten Nullen auftreten. Betrachtet wird immer nur der Teil unterhalb der aktuellen Stelle: Ein eventuell vorhandener Eintrag $\neq 0$ in der Spalte wird ggf. durch Zeilenvertauschung an die betrachtete Stelle gebracht; durch die Addition eines passenden Vielfachens der Zeile werden unterhalb der betrachteten Stelle Nullen erzeugt. (Ein formaler Korrektheitsbeweis unterbleibt hier).

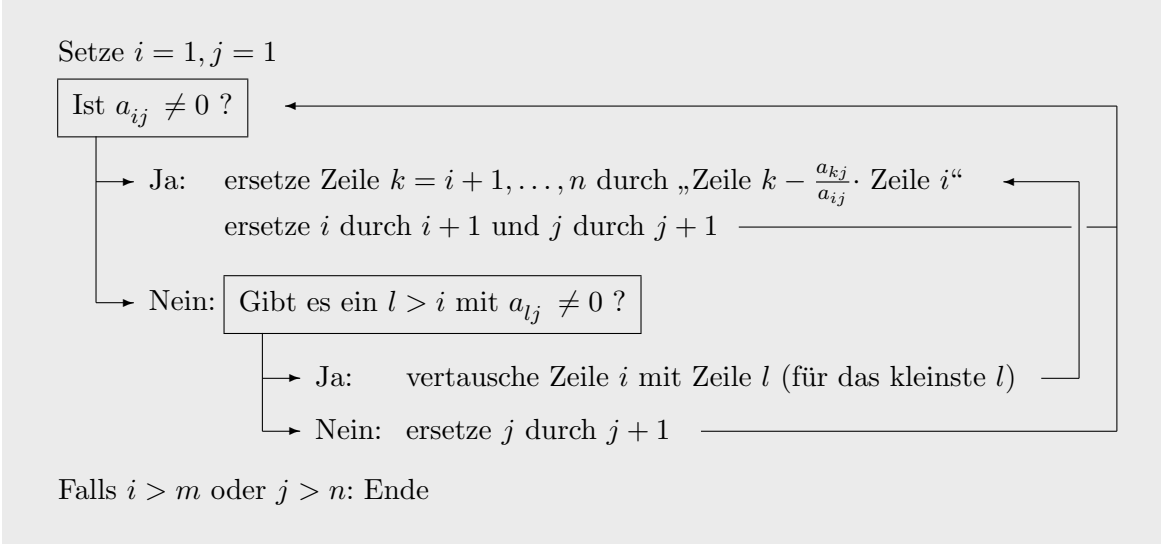


Abbildung 1.2: Gauß-Verfahren

(b) Durch das Gauß-Verfahren bringt man zunächst die Matrix in Zeilenstufenform. Sie ist genau dann invertierbar, wenn die Zeilenstufenform Dreiecksform hat, d. h. wenn die Pivot-Elemente die Diagonalelemente $a_{11}, \ldots, a_{nn}$ sind. Man kann auf diese Matrix nun das Gauß-Verfahren gewissermaßen „punktgespiegelt“, also spaltenweise von rechts nach links und zeilenweise von unten nach oben anwenden, und erhält eine Diagonalmatrix (d. h. $a_{ii} \neq 0$, aber $a_{ij} = 0$ für alle $i \neq j$.) Durch Umformungen der Art (3) kann man schließlich die Diagonaleinträge auf 1 bringen. diese Verfahren heißt manchmal auch *Gauß-Jordan-Verfahren*. □

Beispiel

⋮

Was kann mit dem Gauß-Verfahren berechnet werden? Sei stets E eine invertierbare Matrix, die A in Zeilenstufenform bringt, d. h. E ist eine Matrix $E_k \cdot \ldots \cdot E_1$, wobei $E_1, \ldots, E_k$ Matrizen zu elementaren Umformungen sind, welche nach dem Gauß-Verfahren A in eine Matrix $E_k \cdot \ldots \cdot E_1 \cdot A$ in Zeilenstufenform umformen.

- *Den Rang einer Matrix berechnen:*
  Der Rang der Matrix ist die Anzahl der Pivot-Elemente in der Zeilenstufenform.
- *Testen, ob eine Matrix invertierbar ist:*
  Eine Matrix ist genau dann invertierbar, wenn sie quadratisch ist und der Rang mit der Anzahl der Zeilen/Spalten übereinstimmt.
- *Eine spezielle Lösung eines linearen Gleichungssystems ausrechnen:*
  Man bringt das Gleichungssystem in Zeilenstufenform $EA \cdot x = E \cdot b$ und löst die Gleichungen von unten nach oben auf („Rückwärteinsetzen“). Sind Unbekannte durch eine Gleichung und die vorherigen Festsetzungen nicht eindeutig bestimmt, setzt man einen beliebigen Wert (z. B. 0) ein.
- *Eine Basis des Kerns bestimmen:*
  Man bringt das homogene Gleichungssystem in Zeilenstufenform $EA \cdot x = E \cdot 0 = 0$. Ist der Rang der Matrix gleich n (= Anzahl der Unbekannten), so ist Kern $= \{0\}$, die Basis also die leere Menge. Andernfalls löst man die Gleichungen von unten nach oben durch Rückwärteinsetzen auf. Für jede Unbekannte, die nicht eindeutig festgelegt ist, bekommt man einen Basisvektor des Kerns, indem man diese Unbekannte auf 1 setzt und alle andern dann nicht festgelegten Unbekannten auf 0.
- *Eine Basis des Bilds bestimmen:*
  Spalten $Ae_{i_1}, \ldots, Ae_{i_l}$ von A sind genau dann linear unabhängig, wenn die entsprechenden Spalten $EAe_{i_1}, \ldots, EAe_{i_l}$ der Matrix in Zeilenstufenform linear unabhängig sind. Also bilden die Spalten von A, die Pivot-Spalten von EA sind, eine Basis des Bildes.
- *Eine Menge linear unabhängiger Vektoren zu einer Basis ergänzen:*
  Man fügt die Vektoren als Spalten zu einer $(m \times n)$-Matrix A zusammen und bestimmt eine Basis des Bildes der $(m \times (n+m))$-Matrix $(A \mid I_m)$ wie oben beschrieben.
  alternativ: Man fügt die Vektoren als Zeilen zu der $(n \times m)$-Matrix A^{T} zusammen und bringt sie in Zeilenstufenform. Diejenigen Standardbasisvektoren e_i, für die i keine Pivot-Spalte ist, ergänzen die gegebenen Vektoren zu einer Basis.
- *Das Inverse einer Matrix berechnen:*
  Die elementaren Umformungen, welche A nach dem Gauß-Jordan-Verfahren in die Identitätsmatrix umformen, formen gleichzeitig die Identitätsmatrix in die Inverse von A um: falls $E \cdot A = I_n$, so ist $E \cdot I_n = A^{-1}$.

Mathematische Folgerungen

Definition 1.8.11 *Die* Transponierte A^{T} *einer* $(m \times n)$*-Matrix* $A = (a_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$ *ist die „an der Diagonalen gespiegelte“* $(n \times m)$*-Matrix* $(a_{ji})_{\substack{j=1,\ldots,m \\ i=1,\ldots,n}}$.

Also beispielsweise

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad A^{\mathrm{T}} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Man sieht auch leicht aus der Multiplikationsformel, dass $(A \cdot B)^{\mathrm{T}} = B^{\mathrm{T}} \cdot A^{\mathrm{T}}$. Insbesondere ist die Transponierte einer invertierbaren Matrix selbst invertierbar mit $(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}$.

Satz 1.8.12 $\mathrm{rg}(A) = \mathrm{rg}(A^{\mathrm{T}})$.

Beweis: Man sieht, dass der Satz für Matrizen in Zeilenstufenform gilt. Da Isomorphismen die Dimension bewahren, ändert die Multiplikation von rechts oder links mit einer invertierbaren Matrix nicht den Rang einer Matrix. Sei also $E \cdot A$ in Zeilenstufenform für ein invertierbares E. Dann gilt: $\mathrm{rg}(A) = \mathrm{rg}(E \cdot A) = \mathrm{rg}((E \cdot A)^{\mathrm{T}}) = \mathrm{rg}(A^{\mathrm{T}} \cdot E^{\mathrm{T}}) = \mathrm{rg}(A^{\mathrm{T}})$. □

Aus dem Gauß-Verfahren gewinnt man auch einen Beweis für Satz 1.4.6 im endlich-dimensionalen Fall. Allerdings müsste man sich noch davon überzeugen, dass der Satz für das Gauß-Verfahren nicht gebraucht wurde (und man muss aufpassen, dass man keine Begriffe oder Argumente verwendet, welche bereits auf der Dimension beruhen, wie z. B. den Rang).

Satz 1.8.13 *Wenn ein Vektorraum V eine Basis mit endlich vielen Elementen besitzt, dann haben alls Basen von V die gleiche Anzahl von Elementen.*

Beweis: Angenommen V hat Basen mit m und mit n Elementen, $m < n$. Über die eine Basis ist V isomorph zu K^m; man kann also annehmen, dass $V = K^m$. Nun stellt man die $(m \times n)$-Matrix A auf, deren Spalten die Vektoren der Basis mit n Elementen sind. Diese sind nach Annahme linear unabhängig, also müssen auch die Spalten der in Zeilenstufenform gebrachten Matrix linear unabhängig sein. Ind er Zeilenstufenform sieht man aber, dass maximal m Spalten linear unabhängig sein können: Widerspruch. □

1.9 Determinanten

⋮

1.10 Länge, Winkel, Skalarprodukt

In diesem Abschnitt soll stets $K = \mathbb{R}$ sein; alle betrachteten Vektorräume seien über $\mathbb{R}$. Es sollen nun die geometrisch anschaulichen Begriffe der Länge eines Vektors, des Abstandes zweier Vektoren und des Winkels zwischen zwei Vektoren eingeführt werden. Dabei ist das Vorgehen – ähnlich wie schon bei der Determinante – wie folgt: Man findet eine Formel für die Berechnung, die in den Fällen der Dimension 1, 2 und 3 das Richtige tut und die Eigenschaften besitzt, die man von den Begriffen erwartet. In den höherdimensionalen Fällen, wo eine direkte geometrische Anschauung fehlt, definiert man die Begriffe dann durch diese Formel.

Länge und Abstand

Definition 1.10.1 *Die* Länge *eines Vektors* $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$) *ist*

$$\|v\| := \sqrt{v_1^2 + \cdots + v_n^2}.$$

Der Abstand (*oder die* Distanz) *zweier Vektoren ist*

$$\mathrm{d}(v, w) := \|v - w\|.$$

Es gilt also $\|v\| = d(v, 0)$. Im $\mathbb{R}^1$ ist $\|v\| = |v|$; in $\mathbb{R}^2$ und $\mathbb{R}^3$ sieht man mit dem Satz von Pythagoras, dass die Definition den gewöhnlichen Längenbegriff wiedergibt.

Eigenschaften von Länge und Abstand Für alle $u, v, w \in \mathbb{R}^n$ und $k \in \mathbb{R}$ gilt:

Positivität:	$\|v\| \geq 0$	$\mathrm{d}(v, w) \geq 0$
	$\|v\| = 0 \Leftrightarrow v = 0$	$\mathrm{d}(v, w) = 0 \Leftrightarrow v = w$
Symmetrie:	$\|v\| = \|-v\|$	$\mathrm{d}(v, w) = \mathrm{d}(w, v)$
Dreiecksungleichung:	$\|v + w\| \leq \|v\| + \|w\|$	$\mathrm{d}(v, w) \leq \mathrm{d}(v, u) + \mathrm{d}(u, w)$
Skalierung:	$\|r \cdot v\| = \|r\| \cdot \|v\|$	$\mathrm{d}(r \cdot v, r \cdot w) = \|r\| \cdot \mathrm{d}(v, w)$

Neben diesem gewöhnlichen Längenbegriff (der auch „euklidische Norm“ oder „2-Norm“ $\|v\|_2$ genannt wird), gibt es im Mehrdimensionalen auch weitere Längenbegriffe, etwa die „1-Norm“ $\|v\|_1 = |v_1| + \cdots + |v_n|$ oder die „Maximumsnorm“ $\|v\|_\infty := \max\{|v_1|, \ldots, |v_n|\}$, die ebenfalls alle oben aufgeführten Eigenschaften aufweisen.

Skalarprodukt, Winkel, Orthogonalität

Der Winkel zwischen zwei Vektoren wird üblicherweise über das Skalarprodukt ausgerechnet. Das Skalarprodukt selbst misst keine ganz elementare geometrische Größe wie Länge oder Winkel, sondern beides in Kombination. Im $\mathbb{R}^2$ wird das Skalarprodukt von $v = (v_1, v_2)$ und $w = (w_1, w_2)$ durch die Formel $\langle v, w\rangle = v_1 w_1 + v_2 w_2$ berechnet; die geometrische Interpretation dieser Größe ist: „$\|v\|$ mal $\|w\|$ mal Cosinus des Winkels zwischen v und w“.

Da in diesem Fall die geometrische Interpretation der Formel viel weniger ersichtlich ist als bei der Länge von Vektoren, soll sie auf zwei Arten erklärt werden (die zwar im wesentlichen übereinstimmen, aber Verschiedenes voraussetzen).

Erste Methode Da es bei dem Winkel nicht auf die Längen der Vektoren ankommt, kann man o.E. annehmen, dass v und w Länge 1 haben (indem man sie durch $\frac{v}{\|v\|}$ bzw. $\frac{w}{\|w\|}$ ersetzt; den Fall $v = 0$ oder $w = 0$ kann man außer Acht lassen, da $\langle 0, w\rangle = \langle v, 0\rangle = 0$). Falls $v = e_1 = (1, 0)$, so ist w_1 gerade der Cosinus des eingeschlossenen Winkels zwischen e_1 und w (und w_2 ist die (orientierte) Höhe der von e_2 und w aufgespannten Raute, also im wesentlichen deren Flächeninhalt, da die Grundseite e_1 Länge 1 hat). Winkel sollten unter Drehungen invariant sein; man kann daher den allgemeinen Fall auf diesen speziellen Fall durch die Drehung von v auf e_1 zurückführen. Also ist der Cosinus des Winkels zwischen v und w die erste Koordinate von

$$\begin{pmatrix} v_1 & v_2 \\ -v_2 & v_1 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} v_1 w_1 + v_2 w_2 \\ v_1 w_2 - v_2 w_1 \end{pmatrix}.$$

Man sieht auch, dass die zweite Komponente die orientierte Höhe der Fläche der von v und w aufgespannten Raute ist, also die Volumenveränderung der Abbildung $\begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}$ angibt. Also stimmt neben der Formel für das Skalarprodukt auch die Determinantenformel im $\mathbb{R}^2$.

Zweite Methode Geht man von der gewünschten geometrischen Interpretation des Skalarprodukts aus, so ist klar, dass $\langle v, k \cdot v\rangle = k \cdot \|v\|^2$ sein muss und dass $\langle v, v\rangle = 0$ gelten muss, wenn v und w senkrecht aufeinander stehen. Sicher steht $v = (v_1, v_2)$ senkrecht auf $(-v_2, v_1)$ und allen seinen Vielfachen. Nun schreibt man (nachrechnen durch Ausmultiplizieren!)

$$(w_1, w_2) = \frac{v_1 w_1 + v_2 w_2}{v_1^2 + v_2^2} \cdot (v_1, v_2) + \frac{v_1 w_2 - v_2 w_1}{v_1^2 + v_2^2} \cdot (-v_2, v_1).$$

Der linke Summand gibt dann gerade die orthogonale Projektion von w auf v an; die Länge dieses Vektors mal die Länge von v ist dann gerade $v_1 w_1 + v_2 w_2$.

Ähnliche Überlegungen kann man für den $\mathbb{R}^3$ anstellen (oder man führt, indem man die beiden Vektoren zunächst in die $\{e_1, e_2\}$-Ebene dreht, den dreidimensionalen auf den zweidimensionalen Fall zurück).

Definition 1.10.2 *Das* (Standard-)Skalarprodukt *im Vektorraum* $\mathbb{R}^n$ *ist die folgende Abbildung* $\langle \cdot, \cdot \rangle$*:* $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$*:*

$$\langle v, w\rangle := (v_1, \ldots, v_n) \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_2 \end{pmatrix} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i.$$

Eigenschaften des Skalarprodukts Für alle $v, v', w, w' \in \mathbb{R}^n$ und $r \in \mathbb{R}$ gilt:

Positivität:	$\langle v, v\rangle = \|v\|^2 =\geq 0$	
	$\langle v, v\rangle = 0 \Leftrightarrow v = 0$	
Symmetrie:	$\langle v, w\rangle = \langle w, v\rangle$	
Bilinearität:	$\langle v + v', w\rangle = \langle v, w\rangle + \langle v', w\rangle$	$\langle v, w + w'\rangle = \langle v, w\rangle + \langle v, w'\rangle$
	$\langle r \cdot v, w\rangle = r \cdot \langle v, w\rangle$	$\langle v, r \cdot w\rangle = r \cdot \langle v, w\rangle$

d. h. das Skalarprodukt ist sowohl im ersten als auch im zweiten Argument eine lineare Abbildung.

Satz 1.10.3 (Cauchy-Schwarz[13]) *Seien $v, w \in \mathbb{R}^n$, dann gilt*

$$|\langle v, w\rangle| \quad \leq \quad \|v\| \cdot \|w\|$$

oder (quadriert)

$$\left(\sum_{i=1} v_i w_i\right)^2 \leq \sum_{i=1} v_1^2 \cdot \sum_{i=1} w_1^2.$$

Folgerung 1.10.4 *Für $v \neq 0$ und $w \neq 0$ gilt*

$$-1 \ \leq \ \frac{\langle v, w\rangle}{\|v\| \cdot \|w\|} \ = \ \left\langle \frac{v}{\|v\|}, \frac{w}{\|w\|} \right\rangle \ \leq \ 1;$$

somit findet man einen eindeutigen Winkel $\alpha \in [0, 2\pi]$ mit

$$\langle v, w\rangle = \|v\| \cdot \|w\| \cdot \cos(\alpha).$$

Per Definition nennt man α den zwischen v und w eingeschlossenen Winkel $\sphericalangle(v, w)$.[14]

Beweis: Der Fall $w = 0$ ist klar (beide Seiten ergeben 0); sei also $w \neq 0$. Dann ist

$$\begin{aligned} 0 \ &\leq \ \left\langle v - \frac{\langle v, w\rangle}{\|w\|^2} \cdot w, v - \frac{\langle v, w\rangle}{\|w\|^2} \cdot w \right\rangle \\ &= \ \langle v, v\rangle - 2 \cdot \frac{\langle v, w\rangle}{\|w\|^2} \langle v, w\rangle + \frac{\langle v, w\rangle^2}{\|w\|^4} \langle w, w\rangle \\ &= \ \frac{1}{\|w\|^2} \Big(\langle v, v\rangle \cdot \langle w, w\rangle - 2 \cdot \langle v, w\rangle^2 + \langle v, w\rangle^2 \Big) = \ \frac{1}{\|w\|^2} \Big(\langle v, v\rangle \cdot \langle w, w\rangle - \langle v, w\rangle^2 \Big) \end{aligned}$$

Daraus folgt also $0 \leq \langle v, v\rangle \cdot \langle w, w\rangle - \langle v, w\rangle^2$ bzw. $\langle v, w\rangle^2 \leq \langle v, v\rangle \cdot \langle w, w\rangle = \|v\|^2 \cdot \|w\|^2$, also nach Wurzelziehen das gewünschte Ergebnis. □

Insbesondere gilt also:

$$\begin{aligned} \langle v, w\rangle = 0 \ &\Longleftrightarrow \ \cos \sphericalangle(v, w) \text{ ist } \frac{\pi}{2} \text{ oder } \frac{3}{2}\pi \quad (\text{d. h. } 90^\circ \text{ oder } 270^\circ) \\ &\Longleftrightarrow \ v \text{ und } w \text{ stehen } \textit{senkrecht} \text{ aufeinander.} \end{aligned}$$

In den Dimensionen 1, 2 und 3 stimmt dies also mit dem anschaulichen geometrischen Begriff überein; in den höheren Dimensionen ist es eine sinnvolle Verallgemeinerung. In abstrakten n-dimensionalen Räumen gibt es dagegen kein Standard-Skalarprodukt, also auch keinen natürlichen Winkelbegriff. Durch die Wahl einer Basis kann man aber das Skalarprodukt des $\mathbb{R}^n$ übertragen. Das Standard-Skalarprodukt geht axiomatisch davon aus, dass die Standardbasis eine sogenannte *Orthonormalbasis* ist, also die Basisvektoren Länge 1 haben und paarweise aufeinander senkrecht stehen. Darauf beruhen alle weiteren Längen- und Winkelbestimmungen. Die Übertragung des Standard-Skalarprodukts des $\mathbb{R}^n$ auf einen abstrakten n-dimensionalen Vektorraum durch Wahl einer Basis bedeutet, dass man diese Basis zur Orthonormalbasis erklärt.

[13] Augustin Louis Cauchy (1789-1857), Hermann Amandus Schwarz (1843-1921)

[14] Dieser Begriff ist nicht orientiert, d. h. der Winkel zwischen v und w ist gleich dem Winkel zwischen w und v. Dem entspricht, dass der Cosinus eine gerade Funktion ist, also $\cos(\alpha) = \cos(-\alpha)$ ist.

Satz 1.10.5 *Sei $v \neq 0$, dann ist die* orthogonale Projektion *von w auf v gleich*

$$w_v = \frac{\langle w, v\rangle}{\|v\|} \cdot \frac{v}{\|v\|} = \frac{\langle w, v\rangle}{\langle v, v\rangle} \cdot v.$$

Wenn $v_1, \ldots, v_n$ eine Orthonormalbasis ist, dann gilt

$$w = \sum_{i=1}^{n} \langle w, v_i\rangle \cdot v_i.$$

BEWEIS: Wenn man $\left\langle \sum_{i=1}^{n} \langle w, v_i\rangle \cdot v_i, v_j \right\rangle$ mit Hilfe der Bilinearität des Skalarprodukts ausrechnet, ergibt sich sofort der zweite Teil. Der erste Teil folgt aus der geometrischen Interpretation des Skalarprodukts (bzw. durch Skalieren und Ergänzen von v zu einer Orthonormalbasis). □

Satz 1.10.6 (Verallgemeinerter Satz des Pythagoras[15]; Cosinussatz)
Für $v, w \in \mathbb{R}^n$ gilt:

$$\|v + w\|^2 = \|v\|^2 + 2 \cdot \langle v, w\rangle + \|w\|^2.$$

Insbesondere gilt $\|v + w\|^2 = \|v\|^2 + \|w\|^2$ genau dann, wenn $\langle v, w\rangle = 0$, also wenn v und w senkrecht aufeinander stehen.

BEWEIS: Einfach ausrechnen:

$$\|v + w\| = \sum_{i=1}^{n} (v_i + w_i)(v_i + w_i) = \sum_{i=1}^{n} v_i^2 + \sum_{i=1}^{n} w_i^2 + 2 \cdot \sum_{i=1}^{n} v_i w_i = \|v\|^2 + 2\langle v, w\rangle + \|w\|^2$$

□

Orthogonale Abbildungen

Definition 1.10.7 *Eine lineare Abbildung $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ heißt* orthogonal, *wenn φ das Skalarprodukt erhält, wenn also $\langle \varphi(v), \varphi(w)\rangle = \langle v, w\rangle$ für alle $v, w \in \mathbb{R}^n$ gilt.*[16]

Eine $(n \times n)$-Matrix A heißt orthogonal, wenn die zugehörige lineare Abbildung orthogonal ist.

Definition 1.10.8 *Eine Basis $v_1, \ldots, v_n$ des $\mathbb{R}$ ist eine* Orthonormalbasis, *wenn*

$$\langle v_i, v_j\rangle := \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{falls } i \neq j. \end{cases}$$

Bemerkung: Man rechnet leicht nach, dass

$$\langle Av, w\rangle = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ji} v_i \right) \cdot w_j = \sum_{i=1}^{n} v_i \cdot \left(\sum_{j=1}^{n} a_{ji} w_j \right) = \langle v, A^{\mathrm{T}} w\rangle.$$

[15] Pythagoras (ca. 570 bis ca. 510 v. Chr.)
[16] Achtung: Die orthogonale Projektion aus Satz 1.10.5 ist keine orthogonale Abbildung.

Satz 1.10.9 *Die folgenden Aussagen sind äquivalent für eine $(n \times n)$-Matrix über $\mathbb{R}^n$:*
(a) *A ist orthogonal;*
(b) *A ist invertierbar und $A^{-1} = A^{\mathrm{T}}$;*
(c) *$Ae_1, \ldots, Ae_n$ ist eine Orthonormalbasis.*

BEWEIS: Klar ist, dass eine orthogonale Abbildung eine Orthonormalbasis auf eine Orthonormalbasis abbilden muss, also gilt (a)⇒(c). Der (i,j)-Eintrag von $A^{\mathrm{T}} \cdot A$ ist genau $\langle Ae_i, Ae_j \rangle$, also ist $Ae_1, \ldots, Ae_n$ genau dann eine Orthonormalbasis, wenn $A^{\mathrm{T}} \cdot A = \mathrm{Id}$, also wenn (b) gilt. Schließlich folgt aus (b), dass $\langle Av, Aw \rangle = \langle v, A^{\mathrm{T}} Aw \rangle = \langle v, w \rangle$, also dass A orthogonal ist. □

Beispiel: Drehungen im $\mathbb{R}^2$ sind orthogonal: $\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}^{-1} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$. Ebenso sind Spiegelungen orthogonal. Drehungen und Spiegelungen sind die einzigen orthogonalen Abbildungen der Ebene (dabei sind die Drehungen orientierungserhaltend, die Spiegelungen nicht).

Bemerkung: Orthogonale Abbildungen sind *längentreu*, d. h. $\|Av\| = \|v\|$ für alle v, und *winkeltreu*, d. h. $\sphericalangle(Av, Aw) = \sphericalangle(v, w)$ für alle v, w. Aus $A^{-1} = A^{\mathrm{T}}$ folgt $\det(A)^{-1} = \det(A^{\mathrm{T}}) = \det(A)$ und somit $\det A = \pm 1$. Orthogonale Abbildungen sind also zudem *volumentreu*, allerdings nur im unorientierten Sinn; die Orientierung kann sich ändern (wie man am Beispiel der Spiegelungen sieht).

Scherungen sind Beispiele von volumenerhaltenden Abbildungen, die weder längen- noch winkeltreu sind; Streckungen (aller Vektoren um den gleichen Faktor) sind Beispiele von winkeltreuen Abbildungen, die weder längen- noch volumentreu sind. Man kann aber zeigen, dass längentreue Abbildungen bereits orthogonal sind (dies folgt unmittelbar aus dem verallgemeinerten Satz von Pythagoras). Ebenso sind Abbildungen, die winkel- und volumentreu sind, schon orthogonal.[17]

1.11 Lineare Codes

Einführung, Beispiele, Definitionen

In der Codierungstheorie geht es um folgende Situation bzw. Problematik: Informationen werden in Folgen von Symbolen aufgeschrieben bzw. festgehalten. Man sagt dazu auch, dass die Informationen „codiert“ werden, z. B. durch Morse-Zeichen oder durch Zahlenfolgen im ASCII-Code. Bei der Übermittlung von Nachrichten (z. B. Übertragung durch Funk oder Kabel oder Speicherung der Information über längere Zeiträume) können Übertragungsfehler passieren oder Teile der Information verloren gehen.

Kann man nun die Codierung so wählen, dass Übertragungsfehler erkannt und teilweise korrigiert werden können, und die Informationsübermittlung dennoch möglichst effizient geschieht? Es geht also darum, in die Codierung eine Redundanz einzubauen. Die einfachste Art der Redundanz besteht darin, die Nachricht mehrfach zu wiederholen. Stimmen die empfangenen Informationen nicht überein, so weiß man, dass Übertragungsfehler eingetreten sein. Indem man gegebenenfalls die am häufigsten empfangene Version als die richtige

[17] Winkelerhaltend heißt, dass $Ae_1, \ldots, Ae_n$ eine Orthogonalbasis ist, also $A^{\mathrm{T}} \cdot A$ eine Diagonalbasis ist. Betrachtet man den Winkel zwischen e_i und $e_i + e_j$, sieht man schnell, dass alle Diagonaleinträge gleich sein müssen. Da die Abbildung Determinante ± 1 hat, müssen die Diagonaleinträge $= 1$ sein.

ansieht, kann man u. U.. auch Übertragungsfehler ausgleichen. Die Codierung durch Wiederholung ist aber insofern ineffizient, als sich die Länge der übermittelten Nachricht (und damit Zeit und Kosten) vervielfacht. Die Anforderung der Effizienz bezieht sich aber auch auf die Durchführung von Codierung, Decodierung und die eventuelle Fehlerkorrektur: hierfür sollen schnelle Algorithmen vorliegen.

Konkret betrachtet man folgende Situation: Man verfügt über ein endliches Alphabet (eine Symbolmenge) A mit q Elementen und nimmt Wörter der festen Länge n über A, d. h. Elemente $(a_1, \ldots, a_n)$ von A^n, um Nachrichten zu codieren. Die Menge dieser n-Tupel wird auch der *Hamming-Raum*[18] $H(n, A)$ genannt, bzw. $H(n, q)$, wenn es nur auf die Anzahl der Elemente von A ankommt.

Oft nimmt man als Alphabet eine endliche Gruppe oder einen endlichen Körper, etwa $\mathbb{F}_q$, da die algebraische Struktur beim Ver- und Entschlüsseln helfen kann und geschickte Codierungen ermöglicht. $H(n, \mathbb{F}_q) = \mathbb{F}_q^n$ ist dann ein n-dimensionaler Vektorraum über $\mathbb{F}_q$. Besonders häufig ist der Fall $q = 2$ mit $\mathbb{F}_2 = \{0, 1\}$. Der Hamming-Raum $H(8, \mathbb{F}_2)$ ist zum Beispiel die Menge der möglichen Bytes. Den Hamming-Raum $H(4, \mathbb{F}_2)$ kann man mit den hexadezimalen Ziffern identifizieren.

Beispiele für Codes mit Redundanzen

- Im ursprünglichen ASCII-Code wurden Zeichen durch ein Byte $(a_1, \ldots, a_8)$, also ein 8-Tupel über $\mathbb{F}_2$, codiert. Dabei bildeten die ersten sieben Ziffern $a_1, \ldots, a_7$ die eigentliche Information: als Binärzahl gelesen geben sie die Stelle des codierten Zeichens (Buchstabe, Ziffer, Satz- oder Steuerungszeichen) in der Liste der ASCII-Zeichen an. Die letzte Ziffer a_8 war eine Kontrollziffer, welche den sogenannten *parity check* durchführte: a_8 war so gewählt, dass $a_1 + \cdots + a_8 = 0$ in $\mathbb{F}_2$ gilt. Der Code erkennt, wenn an einer Stelle ein Übertragungsfehler passiert, da dann die Prüfrechnung nicht mehr stimmt. Geht bei der Übertragung eine Stelle verloren, kann man sie errechnen.

- Der alte ISBN-Code bestand aus einer neunstelligen Dezimalzahl, die man als 9-Tupel $(b_1, \ldots, b_9)$ über $\mathbb{F}_{11}$ aufgefasst und um eine Prüfziffer $b_{10} \in \mathbb{F}_{11}$ so ergänzt hat, dass $\sum_{i=0}^{10} i \cdot b_i = 0$ in $\mathbb{F}_{11}$ gilt. (Das Element 10 in $\mathbb{F}_{11}$ wurde übrigens X geschrieben.)
  Dieser Code erkennt eine falsche Ziffer und auch Vertauschungen von zwei Ziffern. („Erkennen“ heißt dabei, dass die Prüfrechnung nicht mehr stimmt.)

- Der aktuelle ISBN-Code ist ein 13-Tupel über $\mathbb{Z}/10\mathbb{Z}$, wobei wieder die letzte Ziffer eine Prüfziffer ist, die so gewählt wird, dass $b_1 + 3b_2 + b_3 + 3b_4 + \cdots + b_{13} = 0$ in $\mathbb{Z}/10\mathbb{Z}$ gilt. Dieser Code erkennt wieder eine falsche Ziffer, aber nur noch gewisse Vertauschungen.

Fehler und die Hamming-Metrik

Anschaulich gesprochen ist ein Code gut, wenn er besonders viele Fehler erkennt oder sogar deren Korrektur zulässt. Um dies zu präzisieren, muss man festlegen, was Fehler sind und wie man ihre Anzahl misst. Im üblichen Setting legt man dazu fest, dass es nur um die Anzahl der Stellen geht, die nicht übereinstimmen. Eine Vertauschung von zwei (verschiedenen) Ziffern zählt also als zwei Fehler, da anschließend zwei Stellen nicht mehr stimmen. Insbesondere werden alle Stellen als gleichwertig gezählt (während man z. B. bei Dezimalzahlen Fehler in den höheren Stellen als gewichtiger ansehen würde als in den niederen Stellen) und alle

[18] Richard Hamming (1915-1998)

Elemente des Alphabets werden ebenfalls untereinander als gleichwertig gezählt (d. h. es ist gleichermaßen ein einziger Fehler, ob z. B. 2 statt 1 empfangen wird oder 9 statt 1).

Mathematischer wird dies in der Hamming-Metrik präzisiert:

Definition 1.11.1 *Für $v = (v_1, \ldots, v_n)$ und $w = (w_1, \ldots, w_n)$ in $H(n, A)$ definiert man den* Hamming-Abstand *(die* Hamming-Metrik*) als* $\mathrm{d}(v, w) := \big|\{i \mid v_i \neq w_i\}\big|$.

Satz 1.11.2 *Die Hamming-Metrik ist eine Metrik auf $H(n, A)$, d. h. es gilt:*

- *Positivität:* $\mathrm{d}(v, w) \geqslant 0$ *und* $\mathrm{d}(v, w) = 0 \iff v = w$
- *Symmetrie:* $\mathrm{d}(v, w) = \mathrm{d}(w, v)$
- *Dreiecksungleichung:* $\mathrm{d}(u, v) \leqslant \mathrm{d}(u, v) + \mathrm{d}(v, w)$.

Falls $(A, +)$ eine (kommutative) Gruppe ist, dann gilt zusätzlich:

- *Translationsinvarianz:* $\mathrm{d}(v, w) = \mathrm{d}(v + u, w + u)$,
  insbesondere $\mathrm{d}(v, w) = \mathrm{d}(v - w, 0) = \mathrm{d}(-w, -v)$

Falls A ein K-Vektorraum ist, dann gilt außerdem:

- *Invarianz unter Skalarmultiplikation:* $\mathrm{d}(v, w) = \mathrm{d}(kv, kw)$ *für* $k \in K \setminus \{0\}$

Beweis: Die ersten beiden Eigenschaften folgen unmittelbar aus der Definition. Die Dreiecksungleichung sieht man aus der Transitivität der Gleichheit: Wenn $u_i \neq w_i$, dann gilt $u_i \neq v_i$ oder $v_i \neq w_i$. Offensichtlich gilt $\mathrm{d}(v, w) \geqslant \mathrm{d}(f(v), f(w))$ für eine beliebige Abbildung $f : H(n, A) \to H(n, A)$, also $\mathrm{d}(v, w) \geqslant \mathrm{d}(f(v), f(w)) \geqslant \mathrm{d}(f^{-1}(f(v)), f^{-1}(f(w))) = \mathrm{d}(v, w)$ für bijektive f. Damit folgt die Invarianz unter Translationen und unter Skalarmultiplikation, da die Abbildungen $v \mapsto v + u$ und $v \mapsto k \cdot v$ für $k \neq 0$ bijektiv sind (Umkehrabbildungen sind $v \mapsto v - u$ und $v \mapsto k^{-1} \cdot v$). □

Während die übliche euklidische Metrik $\|v - w\|$ im $\mathbb{R}^n$ ebenfalls translationsinvariant ist, gilt dort $\|rv - rw\| = |r| \cdot \|v - w\|$. Die Invarianz der Hamming-Metrik unter Skalarmultiplikation ist also eine „ungeometrische“ Eigenschaft.

Bemerkung: Falls $(A, +)$ eine kommutative Gruppe ist und p eine Primzahl, dann ist A genau dann ein $\mathbb{F}_p$-Vektorraum, wenn $\underbrace{a + \cdots + a}_{p \text{ mal}} = 0$ für alle $a \in A$ gilt. Die Skalarmultiplikation ist dann durch $m \cdot a = \underbrace{a + \cdots + a}_{m \text{ mal}}$ gegeben und es gilt $-a = (p - 1) \cdot a$.

Insbesondere folgt daraus: Wenn $C \subseteq \mathbb{F}_p^n$ unter Addition abgeschlossen ist, dann ist C bereits ein Untervektorraum!

Definition 1.11.3

(a) *Ein* Code *ist eine Teilmenge von $H(n, A)$ bzw. $H(n, q)$. Man spricht von einem „*Code der Länge n über A*“ bzw. einem „*q-ären Code der Länge n*“. Der* Minimalabstand *des Codes ist* $\min\big\{\mathrm{d}(v, w) \;\big|\; v, w \in C, v \neq w\big\}$.

(b) *Ein* linearer Code *ist ein Untervektorraum von $\mathbb{F}_q^n$. Das* Gewicht *von $v \in C$ ist* $\mathrm{d}(v, 0)$ *und das* Minimalgewicht *des Codes ist* $\min\big\{\mathrm{d}(v, 0) \;\big|\; v \in C, v \neq 0\big\}$.

(c) *Ein Code C* erkennt *(mindestens) e* Fehler*, falls der Minimalabstand größer als e ist.*

(d) *Ein Code C* korrigiert *(mindestens) e* Fehler*, falls es zu jedem $v \in H(n, A)$ höchstens ein $c \in C$ gibt mit* $\mathrm{d}(v, c) \leqslant e$.

Lineare Codes werden meist durch zwei oder drei Parameter beschrieben: als „(q-äre) $[n,k]$-*Codes*“ oder „(q-äre) $[n,k,d]$-*Codes*“. Dann ist stets n die Länge der Wörter, $k = \dim C$ und d das Minimalgewicht. Es gilt $|C| = q^k$ bzw. $k = \log_q |C|$.[19] Wegen $\mathrm{d}(v,w) = \mathrm{d}(v-w,0)$ ist das Minimalgewicht eines linearen Codes gleich seinem Minimalabstand. Unmittelbar aus der Definition sieht man auch:

Satz 1.11.4
(a) *Ein Code mit Minimalabstand d erkennt $d-1$ Fehler und korrigiert $\left\lfloor \frac{d-1}{2} \right\rfloor$ Fehler.*
(b) *Ein Code, der e Fehler korrigiert, erkennt $2e$ Fehler und hat Minimalabstand mindestens $2e+1$.*

Beispiele:

- Der alte ISBN-Code ist ein 11-ärer Code der Länge 10, der einen Fehler erkennt und keinen korrigiert.
- Der ursprüngliche ASCII-Code ist ein binärer linearer $[8,7,2]$-Code, der also einen Fehler erkennt und keinen korrigiert.
- Der Wiederholungscode $\{(x,x,x) \mid x \in \mathbb{F}_q\} \subseteq H(3,q)$ ist ein q-ärer linearer $[3,1,3]$-Code, der zwei Fehler erkennt und einen korrigiert.

Definition 1.11.5 *Der Ball vom Radius e um v ist*[20]

$$B_e(v) := \big\{ w \in H(n,A) \;\big|\; \mathrm{d}(v,w) \leqslant e \big\}.$$

Die Anzahl der Elemente eines solchen Balles kann man ausrechnen durch

$$\big|B_e(v)\big| = \sum_{i=0}^{e} \binom{n}{i} (q-1)^i;$$

hierbei durchläuft i die möglichen Abstände zu v; der Binomialkoeffizient gibt die Anzahl der Möglichkeiten für die i Stellen, an denen die Abweichungen auftreten; $q-1$ ist für jede Stelle die Anzahl der alternativen Elemente des Alphabets.

Für $q = 2$ gilt insbesondere

$$\big|B_e(c)\big| = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{e}.$$

Es gilt nun offensichtlich:

- C erkennt genau dann e Fehler, wenn $c' \notin B_e(c)$ für $c, c' \in C$, $c \neq c'$.
- C korrigiert genau dann e Fehler, wenn die Bälle $B_e(c)$ für $c \in C$ paarweise disjunkt sind.

Gütekriterien und Schranken für Codes

Zwei Beispiele für einen 1-fehlerkorrigierenden Code Ausgangslage: Man hat als eigentliche Information Wörter der Länge 4 über $\mathbb{F}_2$ (also etwa die Binärdarstellung von hexadezimalen Zeichen). Man möchte den Code nun z. B. durch Anhängen von Prüfziffern so verändern, dass er einen Fehler korrigiert.

[19] Manche Autoren bevorzugen, statt der Dimension eines Codes C an zweiter Stelle die Anzahl der Elemente von C anzugeben.

[20] In der Analysis sind Bälle üblicherweise als offene Bälle definiert, d. h. man fordert „$< e$“ statt „$\leqslant e$“. Dies ist in der diskreten Situation hier nicht besonders sinnvoll.

Code 1 Die „naive“ Methode besteht darin, das Ausgangswort dreifach zu senden. Wörter aus $H(4, \mathbb{F}_2)$ werden also codiert als Wörter in $H(12, \mathbb{F}_2)$, nämlich $v = (v_1, v_2, v_3, v_4)$ als $v^\frown v^\frown v := (v_1, v_2, v_3, v_4, v_1, v_2, v_3, v_4, v_1, v_2, v_3, v_4)$.

$C_1 = \left\{ v^\frown v^\frown v \;\middle|\; v \in H(4, \mathbb{F}_2) \right\}$ ist dann ein binärer $[12, 4, 3]$-Code: Die Wortlänge ist 12, die Dimension 4 (da $\dim H(4, \mathbb{F}_2) = 4$) und das Minimalgewicht 3, d. h. der Code erkennt zwei Fehler und korrigiert einen.

Dieser Code ist aber nicht besonders effizient: die Raumgröße ist $|H(12, \mathbb{F}_2)| = 2^{12} = 4.096$. Es gibt 16 Codewörter, die mit Ihren „Korrekturbereichen“ einen Platz von $16 \cdot |B_1(c)| = 16 \cdot 13 = 208$ einnehmen. Es gibt also einen „verschwendeten Platz“ von $4.096 - 208 = 3.888$ Wörtern.

Code 2 C_2 besteht aus folgenden Wörtern in $H(7, \mathbb{F}_2)$:

$(0,0,0,0,0,0,0)$	$(0,1,0,0,1,0,1)$	$(1,0,0,0,0,1,1)$	$(1,1,0,0,1,1,0)$
$(0,0,0,1,1,1,1)$	$(0,1,0,1,0,1,0)$	$(1,0,0,1,1,0,0)$	$(1,1,0,1,0,0,1)$
$(0,0,1,0,1,1,0)$	$(0,1,1,0,0,1,1)$	$(1,0,1,0,1,0,1)$	$(1,1,1,0,0,0,0)$
$(0,0,1,1,0,0,1)$	$(0,1,1,1,1,0,0)$	$(1,0,1,1,0,1,0)$	$(1,1,1,1,1,1,1)$

Ein Wort v aus $H(4, \mathbb{F}_2)$ wird codiert durch dasjenige Wort aus $H(7, \mathbb{F}_2)$ in der Liste, dessen Anfangsstück gerade v ist. Man kann nun überprüfen, dass C_2 ein binärer $[7, 4, 3]$-Code ist. Der Code erkennt also ebenfalls zwei Fehler und korrigiert einen, bei gleicher Anzahl von Codewörtern (d. h. gleicher Dimension 4).

Die Raumgröße ist hier aber $|H(7, \mathbb{F}_2)| = 2^7 = 128$. Die 16 Codewörter nehmen mit Ihren „Korrekturbereichen“ einen Platz von $16 \cdot |B_1(c)| = 16 \cdot 8 = 128$ ein, d. h. es gibt keinen verschwendeten Platz. Solche Codes heißen *perfekte Codes*.

C_2 ist übrigens ein Beispiel für einen *Hamming-Code*. Im folgenden wird erklärt werden, wie man C_2 systematisch konstruieren kann und wie Codierung und Decodierung funktionieren. Denn C_1 hat gegenüber C_2 zunächst den Vorteil, dass die Codierungs- und Decodierungsschritte offensichtlich sind, während man bei C_2 in der Tabelle nachschauen muss.

Ein guter Code sollte also

- möglichst viele Fehler erkennen und korrigieren, d. h. großen Minimalabstand haben;
- möglichst viele Codewörter im Verhältnis zur Wortlänge n haben;
- und dabei eine effiziente Codierung (Verschlüsselung), Decodierung (Entschlüsselung) und ggf. Fehlerkorrektur gestatten.

Für Ver- und Entschlüsselung gibt es immer die Möglichkeit, eine Verschlüsselungstafel aufzustellen. Für die Entschlüsselung eines fehlerhaft übertragenen Worts muss dann nach dem bzgl. der Hamming-Metrik nächstgelegenen Wort in der Tafel suchen. Bei einem großen Hamming-Raum ist dies aber ein eher langwieriger Algorithmen. Schnelle Algorithmen setzen voraus, dass der Code eine interne Struktur besitzt. Daher sind lineare Codes interessant.

Die ersten beiden Anforderungen laufen einander zuwider: Redundanzen (Prüfziffern) erhöhen die Wortlänge. Es gibt daher Schranken für das Verhältnis von Codegröße und Minimalabstand bei gegebenen Hamming-Raum.

Satz 1.11.6 (Die Hamming-Schranke) *Die Anzahl der Codewörter eines q-ären Codes der Länge n mit Mindestabstand $\geqslant d$ ist höchstens*

$$q^n \Big/ \sum_{i=0}^{\lfloor \frac{d-1}{2} \rfloor} \binom{n}{i} \cdot (q-1)^i.$$

BEWEIS: Die Schranke folgt sofort aus der Formel für die Anzahl der Elemente von $|B_e(c)|$ und der Größe des Hamming-Raumes $|H(n,q)| = q^n$. □

Definition 1.11.7 *Ein Code heißt* perfekt, *wenn er die Hamming-Schranke erreicht.*

Beispiele:

- $q = 2, n = 7, d = 3$: Hier ergibt die Hamming-Schranke $2^7/(1+7) = 16$. Der Hamming-Code C_2 im Beispiel oben erreicht als perfekter Code diese Schranke.
- $q = 2, n = 6$: Die Folge der Binomialkoeffizienten $\binom{6}{i}$ ist 1, 6, 15, 20, 15, 6, 1. Keine der Summen $\binom{6}{0} + \ldots \binom{6}{e}$ ist ein Teiler von $2^6 = 64$ außer für $e = 0$ und $e = 6$. Diese entsprechen den sogenannten *trivialen Codes*: es sind alle Wörter Codewörter (bei Minimalabstand 1) oder es gibt überhaupt nur ein Codewort (bei Minimalabstand ∞). Beide Codes sind perfekt, aber aus Sicht der Codierungstheorie vollkommen uninteressant. Darüberhinaus gibt es also keinen perfekten binären Code der Länge 6.

Satz 1.11.8 (Die Gilbert-Schranke[21]) *Gegebenen q, n, d, so gibt es einen q-ären Code der Länge n und vom Minimalabstand mindestens d mit mindestens*

$$q^n \Big/ \sum_{i=0}^{d-1} \binom{n}{i} \cdot (q-1)^i$$

Codewörtern. Ist q eine Primzahlpotenz, so kann man den Code linear wählen.

BEWEIS: Sei C ein Code vom Minimalabstand $\geqslant d$, so dass $|C|$ kleiner als die Gilbert-Schranke ist. Dann gibt es ein $x \in H(n,q)$, welches zu allen $c \in C$ mindestens Abstand d hat, denn nach Annahme gilt $|C| \cdot |B_{d-1}(c)| < q^n = |H(n,q)|$. (Die Größe der Bälle $B_{d-1}(c)$ hängt nicht von c ab!) Dann ist $C \cup \{x\}$ ein größerer Code vom Minimalabstand $\geqslant d$. Durch sukzessives Vergrößern erhält man also einen Code, der die Gilbert-Schranke erfüllt.

Für den linearen Fall nimmt man an, dass $C \subseteq H(n, \mathbb{F}_q)$ bereits ein linearer Code ist (z. B. der triviale Code $\{0\}$) und wählt x wie oben. Statt $C \cup \{x\}$ betrachtet man nun den erzeugten linearen Code $\langle x, C \rangle$, muss aber noch zeigen, dass dieser weiterhin Mindestgewicht $\geqslant d$ hat.

Ein typisches Element darin hat die Form $\alpha x + \beta c$ mit $\alpha, \beta \in \mathbb{F}_q$ und $c \in C$.
– Falls $\alpha = 0$, so ist $\mathrm{d}(\alpha x + \beta c, 0) = \mathrm{d}(\beta c, 0) \geqslant d$ nach Annahme an C.
– Falls $\alpha \neq 0$, so ist $\mathrm{d}(\alpha x + \beta c, 0) = \mathrm{d}(x, -\frac{\beta}{\alpha} c) \geqslant d$ nach Wahl von x, da $\frac{\beta}{\alpha} c \in C$. □

Beispiel: Für $q = 2, n = 7, d = 3$ ergibt die Gilbert-Schranke $2^7/(1+7+21) \approx 4,41$. Die Gilbert-Schranke stellt also die Existenz einen Codes C vom Minimalabstand 3 mit mindestens 5 Codewörtern sicher. Im linearen Fall weiß man, dass die Anzahl der Element von C als Untervektorraum von $\mathbb{F}_2^7$ eine Zweierpotenz sein muss, also erhält man $|C| \geqslant 8$. Aus dem obigen Beispiel wissen wir aber, dass es sogar den Hamming-Code mit 16 Wörtern gibt.

Lineare Codes, Erzeuger- und Prüfmatrizen

Sei C nun ein q-ärer $[n,k]$-Code, also ein k-dimensionaler Unterraum von $H(n,q) = \mathbb{F}_q^n$.

[21] Edgar Gilbert (*1923)

Definition 1.11.9 *Eine* Erzeugermatrix *G für einen linearen $[n,k]$-Code C ist eine $(k \times n)$-Matrix, deren Zeilen eine Basis von C bilden.*

Eine Erzeugermatrix eines Codes ist nicht eindeutig bestimmt. Man kann sie aber durch elementare Umformungen auf die Form

$$\left(\begin{array}{c|c} \mathrm{Id}_k & A \end{array} \right)$$

bringen, wobei A eine $(k \times (n-k))$-Matrix ist. Im allgemeinen wird der Code durch solche Umformungen verändert und durch einen *äquivalenten* Code ersetzt. Äquivalente Codes haben zwar u. U.. andere Codewörter, aber dieselben Parameter wie z. B. Anzahl der Codewörter, Minimalabstand, und werden in der Regel identifiziert. Wir werden auch sehen, dass diese spezielle Form der Erzeugermatrix einer Codierung durch Anhängen von Prüfziffern entspricht, also einer sehr üblichen Art von Codes.

Beispiel: Der $[7,4,3]$-Hamming-Codes hat mit

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

eine Erzeugermatrix in der Form $(\mathrm{Id}_k \mid A)$.

Bemerkung: Man sieht hier leicht, dass die Basisvektoren ein Gewicht und paarweise einen Abstand von mindestens 3 haben. Dies reicht aber nicht um zu folgern, dass der erzeugte Code Minimalabstand mindestens 3 hat. Zum Beispiel haben die Vektoren $(1,1,1,0,0,0)$, $(0,0,0,1,1,1)$ und $(1,1,0,1,1,0)$ Gewicht $\geqslant 3$ und paarweisen Abstand $\geqslant 3$, der von ihnen erzeugte Code hat aber nur Minimalgewicht 2. Man kann nur, wie im Beweis der Gilbert-Schranke, von einem Vektor, der zu einem Untervektorraum einen Minimalabstand hat, auf den Minimalabstand des von beiden erzeugten Untervektorraums schließen.

Codierung mit Hilfe der Erzeugermatrix: Die Codierung eines (Zeilen-)Vektors $v \in H(k,q)$ erfolgt nun[22] durch $v \cdot G = (G^{\mathrm{T}} \cdot v^{\mathrm{T}})^{\mathrm{T}} \in H(n,q)$. Hat G die besondere Form $(\mathrm{Id}_k \mid A)$, so entsteht der Codevektor also durch Anhängen von $n-k$ Prüfziffern, da $v \cdot G$ die Form $v^\frown w$ für einen Vektor w der Länge $n-k$ hat. Die Prüfziffern erhält man als Linearkombination der Prüfziffern der Basiselement.

Im Beispiel wird etwa der Vektor $(1,0,1,1)$, den man als Darstellung der Binärzahl 1011 bzw. der Hexadezimalzahl B auffassen kann, durch $(1,0,1,1) \cdot G = (1,0,1,1,0,1,0)$ codiert.

Satz 1.11.10 *Die Codierungsabbildung ist injektiv.*

Beweis: Im Falle des Anhängens von Prüfziffern ist dies trivialerweise gegeben; da jeder Code zu einem solchen äquivalent ist, also durch Isomorphie dazu übergeht, gilt es auch allgemein.

Alternativ: Die Zeilen von G sind linear unabhängig, also gilt

$$\mathrm{rg}(G) = \mathrm{rg}(G^{\mathrm{T}}) = \dim \mathrm{Bild}(G^{\mathrm{T}}) = k$$

und somit $\dim \mathrm{Kern}(G^{\mathrm{T}}) = 0$. □

[22] Vektoren v sind hier stets als Zeilenvektoren gedacht; v^{T} ist dann der entsprechende Spaltenvektor.

Definition 1.11.11 *Eine* Prüfmatrix *H für einen $[n,k]$-Code C ist eine $((n-k)\times n)$-Matrix, für die $C = \mathrm{Kern}(H)$ gilt.*

Satz 1.11.12 *Die folgenden Aussagen sind äquivalent für einen linearen $[n,k]$-Code C und eine $((n-k)\times n)$-Matrix H:*
(a) *H ist eine Prüfmatrix von C.*
(b) *Es gilt $H\cdot c^{\mathrm{T}} = 0$ für alle $c\in C$ und die Zeilen von H sind linear unabhängig.*
(c) *Es gilt $G\cdot H^{\mathrm{T}} = 0$ und die Zeilen von H sind linear unabhängig.*

Beweis: $G\cdot H^{\mathrm{T}} = 0$ ist äquivalent mit $H\cdot G^{\mathrm{T}} = 0$ und impliziert $H\cdot c^{\mathrm{T}} = 0$ für alle $c\in C$, da jedes $c\in C$ Linearkombination von Zeilen von G ist. Dies bedeutet, dass C im Kern von H liegt. Wegen $\mathrm{rg}(H) = \dim \mathrm{Bild}(H) = n - \dim \mathrm{Kern}(H)$ und $n-k = n-\dim(C)$ ist die lineare Unabhängigkeit der Zeilen von H mit $\dim \mathrm{Kern}(H) = k$ äquivalent und mit der ersten überlegung also zu $\mathrm{Kern}(H) = C$. □

Satz 1.11.13 *Genau dann sind G und H Erzeuger- und Prüfmatrix eines $[n,k]$-Codes, wenn G eine $(k\times n)$-Matrix vom Rang k und H eine $((n-k)\times n)$-Matrix vom Rang $(n-k)$ ist, für die $G\cdot H^{\mathrm{T}} = 0$ ist.*

Beweis: Erzeuger- und Prüfmatrix haben nach Definition und Satz 1.11.12 diese Eigenschaften. Umgekehrt kann es eine $(k\times n)$-Matrix G vom Rang k nur für $k\leqslant n$ geben; solch eine Matrix ist dann per Definition Erzeugermatrix des Codes $\mathrm{Bild}(G^{\mathrm{T}})$. H ist dann wieder nach Satz 1.11.12 eine zugehörige Prüfmatrix. □

Folgerung 1.11.14 *Wenn eine Erzeugermatrix G die Form $(\mathrm{Id}_k \mid A)$ hat, so ist*

$$H = \left(\begin{array}{c|c} -A^{\mathrm{T}} & \mathrm{Id}_{n-k} \end{array} \right)$$

eine zugehörige Prüfmatrix.

Wenn also

$$G = \begin{pmatrix} 1 & 0 & \dots & 0 & a_{11} & \dots & a_{1\,n-k} \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & & \vdots \\ \vdots & & \ddots & \ddots & 0 & \vdots & \vdots \\ 0 & \dots & 0 & 1 & a_{k1} & \dots & a_{k\,n-k} \end{pmatrix},$$

dann ist

$$H = \begin{pmatrix} -a_{11} & \dots & -a_{k1} & 1 & 0 & \dots & \dots & 0 \\ \vdots & & \vdots & 0 & \ddots & \ddots & & \vdots \\ \vdots & & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \vdots & \vdots & & \ddots & \ddots & 0 \\ -a_{1\,n-k} & \dots & -a_{k\,n-k} & 0 & \dots & \dots & 0 & 1 \end{pmatrix},$$

denn man sieht leicht, dass dann $G\cdot H^{\mathrm{T}} = 0$.

Der Code C ist jeweils durch G und durch H festgelegt; umgekehrt sind G und H aber nicht eindeutig durch C bestimmt. In der speziellen Form sind sie zwar durch C festgelegt, für äquivalente Codes sind sie aber im allgemeinen verschieden.

Beispiel: Im Falle des $[7,4,3]$-Hamming-Codes und der Erzeugermatrix G von oben ist

$$H = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

die passende Prüfmatrix.

Decodierung mit Hilfe der Prüfmatrix: Angenommen $w \in H(n,q)$ wird empfangen. Als Decodierung wird zunächst dasjenige $c \in C$ gesucht, welches minimalen Hamming-Abstand zu w hat und dann das Urbild von c unter der Codierung (die ja injektiv ist) bestimmt. Um die Existenz von c sicherzustellen, nehmen wir an, dass entweder ein perfekter, e-fehlerkorrigierender Code vorliegt, oder dass höchstens e Übertragungsfehler vorgekommen sind, wobei $2e+1 \leqslant d$ für das Minimalgewicht d des Codes ist.

Es gilt aufgrund dieser Annahmen dann $w = c + f$ wobei $c \in C$ und $\mathrm{d}(f,0) \leqslant e$. Man berechnet nun zunächst das sogenannte *Syndrom* von w, das ist $H \cdot w^{\mathrm{T}}$. Es gilt dafür

$$H \cdot w^{\mathrm{T}} = H \cdot (c+f)^{\mathrm{T}} = H \cdot c^{\mathrm{T}} + H \cdot f^{\mathrm{T}} = 0 + H \cdot f^{\mathrm{T}} = H \cdot f^{\mathrm{T}},$$

d. h. das Syndrom von w ist gleich dem Syndrom des Fehlers f. Für zwei mögliche Fehler f, f' gilt zudem

$$\mathrm{d}(f-f',0) \leqslant \mathrm{d}(f,0) + \mathrm{d}(-f',0) \leqslant e + e < d,$$

also ist $f - f' \notin C$ und somit $H \cdot f^{\mathrm{T}} - H \cdot f'^{\mathrm{T}} = H \cdot (f-f')^{\mathrm{T}} \neq 0$. Verschiedene Fehler haben also verschiedene Syndrome.

Man kann nun eine Liste der Syndrome der möglichen Fehler aufstellen. Dies sind $|B_e(0)|$ viele; eine im Vergleich mit $|H(n,q)|$ deutlich kleinere Zahl. Die Decodierung geht dann folgendermaßen: Man berechnet das Syndrom $H \cdot w^{\mathrm{T}}$, schaut in der Tabelle nach, welchem Fehler f es entspricht, korrigiert den Fehler, und erhält so das zugehörige Codewort $c \in C$. Zu diesem kann man nun die Umkehrung der Codierungsabbildung bestimmen; hat G die besondere Form $(\mathrm{Id}_k \mid A)$, dann besteht diese Umkehrabbildung einfach im Weglassen der Prüfziffern.

An der Prüfmatrix kann man das Minimalgewicht des Codes ablesen:

Satz 1.11.15 *Sei C ein linearer Code. Dann gilt: C hat genau dann Minimalgewicht mindestens d, wenn je $d-1$ Spalten der Prüfmatrix linear unabhängig sind.*

Beweis: Eine Linearkombination $0 = \sum_{i=0}^{n} a_i S_i$ der Spalten S_i von H entspricht gerade einem Vektor $a = (a_1, \ldots, a_n)$, für welchen $H \cdot a = 0$ gilt, also einem $a \in \mathrm{Kern}(H) = C$. Die Anzahl der Komponenten $a_i \neq 0$ in a, also der tatsächlich vorkommenden Spalten in der Linearkombination, ist gerade das Gewicht von a. □

Insbesondere hat also ein Code Minimalgewicht mindestens 3, wenn je zwei Spalten der Prüfmatrix linear unabhängig sind, d. h. wenn keine null ist und keine das skalare Vielfache einer anderen ist.

Definition 1.11.16 *Ein Hamming-Code ist ein linearer Code C vom Minimalgewicht 3, dessen Prüfmatrix zu gegebener Zeilenanzahl die maximale Anzahl von Spalten hat.*

Ist m die Anzahl der Zeilen der Prüfmatrix H, so bilden die Spalten von H ein Repräsentantensystem der Vektoren $\mathbb{F}_q \setminus \{0\}$ bezüglich skalarer Multiplikation. d. h. keine Spalte von H ist die Nullspalte und für jeden Vektor $v \in \mathbb{F}_q^m, v \neq 0$ gibt es genau einen Spaltenvektor von H, der ein skalares Vielfaches von v ist. Mit anderen Worten: die Spaltenvektoren sind Erzeuger eindimensionaler Unterräume von $\mathbb{F}_q^m$ und jeder solche Unterraum kommt genau einmal vor.

Man kann sich leicht überlegen, dass die Anzahl n der Spalten dann genau $\frac{q^m-1}{q-1}$ ist, denn $q^m - 1$ ist die Anzahl der Vektoren $\neq 0$ und $q - 1$ die Anzahl der Skalarfaktoren $\neq 0$). Die Dimension des zugehörigen Hamming-Codes ist dann $k = \frac{q^m-1}{q-1} - m$.

Satz 1.11.17 *Hamming-Codes sind perfekt.*

Beweis: Es ist also $n = \frac{q^m-1}{q-1}$ und $k = \frac{q^m-1}{q-1} - m$ und man rechnet nach, dass

$$|C| \cdot |B_1(c)| = q^{\frac{q^m-1}{q-1}-m} \cdot \left(1 + \tfrac{q^m-1}{q-1} \cdot (q-1)\right) = q^{\frac{q^m-1}{q-1}} = |H(\tfrac{q^m-1}{q-1}, \mathbb{F}_q)|.$$

□

Spezialfall $q = 2$: Hier ist die Situation besonders einfach: Die Spaltenvektoren von H sind sämtliche Vektoren in $\mathbb{F}_q^m$ ohne den Nullvektor. Ihre Anzahl n ist $2^m - 1$, die Dimension des Codes ist $2^m - (m+1)$. Übrigens sind alle binären Hamming-Codes fester Länge aquivalent.

Auch die Decodierung ist im Falle $q = 2$ besonders einfach: Die möglichen Fehler sind gerade die Standardbasisvektoren $e_1, \ldots, e_n$ in $\mathbb{F}_2^n$. Das Syndrom $H \cdot e_i^{\mathrm{T}}$ von e_i ist dann gerade die i-te Spalte von H. Zur Decodierung von w berechnet man also das Syndrom $H \cdot w^{\mathrm{T}}$. Ist das Syndrom 0, so ist kein Fehler aufgetreten. Andernfalls schaut man nach, in welcher Spalte i von H das Syndrom auftritt, ändert das i-te Bit von w und lässt die Prüfziffern, d. h. die letzten $n - k$ Stellen von w, weg.

Liste der perfekten Codes

Ist q eine Primzahlpotenz, so gibt es die folgenden perfekten q-ären Codes (wobei e die Anzahl der korrigierbaren Fehler bezeichnet):

- triviale Codes, die nur aus einem Wort bestehen (nimmt man dafür den Nullvektor, so ist es ein $[n, 0, \infty]$-Code mit $e = n$);
- den triviale $[n, n, 1]$-Code mit $e = 0$ (alle Wörter sind im Code);
- die q-ären Hamming-Codes: $[\frac{q^m-1}{q-1}, \frac{q^m-1}{q-1} - m, 3]$-Codes mit $e = 1$; sowie einige nichtlineare Codes mit gleichen Parametern wie Hamming-Codes;
- die binären Wiederholungscodes ungerader Länge:
  zu jedem $e \in \mathbb{N}$ einen $[2e + 1, 1, 2e + 1]$-Code, der nur die beiden Wörter $(0, 0, \ldots, 0)$ und $(1, 1, \ldots, 1)$ enthält;
- den binären Golay-Code[23]: ein $[23, 12, 7]$-Code mit $e = 3$;
- den ternären Golay-Code: ein $[11, 6, 5]$-Code mit $e = 2$.

[23] Marcel Golay (1902-1989)

Der binäre Wiederholungscode stimmt für $e = 0$ mit dem trivialen $[1, 1, 1]$-Code und für $e = 1$ mit dem $[3, 1, 31]$-Hamming-Code überein.

Falls q keine Primzahlpotenz ist, so weiß man nicht, ob es perfekte nicht-triviale q-äre Codes gibt, und nur für einige wenige Werte weiß man, dass keine perfekten q-ären Codes existieren außer den trivialen. Im allgemeinen weiß man auch wenig darüber, welches die „besten“ Codes sind (hinsichtlich der „Packungsdichte“

Kapitel 2

Algebra

2.1 Gruppen

Zur Erinnerung:

Definition 2.1.1 *Eine* Gruppe *besteht aus einer nicht-leeren Menge G und einer zweistelligen Operation $\circ : G \times G \to G$ auf G, die assoziativ ist, ein neutrales Element e hat, und in der jedes Element $g \in G$ ein inverses Element g^{-1} besitzt.*

G heißt kommutative Gruppe*, wenn $\circ$ zusätzlich kommutativ ist.*

Neutrale Elemente und die jeweiligen Inversen sind eindeutig bestimmt; insbesondere gilt $(g^{-1})^{-1} = g$ und $(g \circ h)^{-1} = h^{-1} \circ g^{-1}$. Es reicht also, wenn man eine Gruppe beschreiben möchte, die Grundmenge und die Operation anzugeben. Bei Standardbeispielen lässt man die Operation auch weg (so ist z. B. klar, dass mit $\mathbb{Z}$ die Gruppe $(\mathbb{Z}, +)$ gemeint ist, und nicht eine etwaige andere Operation.)

Es gibt drei gebräuchliche Schreibweisen:

	Operation	neutrales Element	inverses Element
allgemeine Schreibweise	$\circ$	e	g^{-1}
multiplikative Schreibweise	$\cdot$	1	g^{-1}
additive Schreibweise	$+$	0	$-g$

Multiplikationspunkte (und manchmal auch das Zeichen $\circ$) werden gerne weggelassen; die additive Schreibweise ist üblicherweise kommutativen Gruppen vorbehalten. Bei der Angabe einer Gruppe

Erinnerung auch an wichtige **Beispiele**:

- die kommutative Gruppe $(\mathbb{Z}, +, 0)$;
- die kommutative Gruppe $\mathbb{Z}_m = \big(\{0, \dots, m-1\}, +_m, 0\big)$, wobei

$$x +_m y = \text{„Rest von } x + y \text{ bei Division durch } m\text{“} = \begin{cases} x + y & \text{falls } x + y < m \\ x + y - m & \text{falls } x + y \geq m \end{cases};$$

- die Gruppen $(\mathrm{Sym}(M), \circ, \mathrm{id})$ der Permutationen von M, d. h. der Bijektionen $M \to M$;

- die Gruppe der Vektorraum-Isomorphismen $V \to V$ mit der Hintereinanderausführung von Abbildungen als Operation;
- die Gruppe $\mathrm{GL}(n, K)$ der invertierbaren $(n \times n)$-Matrizen über einem Körper K, d. h. der $(n \times n)$-Matrizen mit Determinante $\neq 0$.

Für $n = 0, 1$ ist $\mathrm{GL}(n, \mathbb{R})$ kommutativ, ebenso $\mathrm{Sym}(M)$ für ein- oder zweielementige Mengen M. In allen anderen Fällen sind diese Gruppen nicht kommutativ.

Definition 2.1.2 *Eine Abbildung $\varphi : G \to H$ zwischen zwei Gruppen $(G, \circ_G, e_G)$ und $(H, \circ_H, e_H)$ heißt* Gruppenhomomorphismus, *falls*

- $\varphi(g_1 \circ_G g_2) = \varphi(g_1) \circ_H \varphi(g_2)$ *für alle* $g_1, g_2 \in G$;
- $\varphi(e_G) = e_H$;
- $\varphi(g^{-1}) = \varphi(g)^{-1}$ *für alle* $g \in G$.

Man kann zeigen, dass es ausreicht, die erste Bedingung zu prüfen, da die zweite und dritte dann automatisch erfüllt sind. Wenn klar ist, dass es um Gruppen geht, spricht man auch kurz von „Homomorphismus“.

Beispiele für Gruppenhomomorphismen:

- Die „Rest-Abbildung“ $\mathbb{Z} \to \mathbb{Z}_m$, die den Rest bei der Division durch m angibt, also eine Zahl $n = qm + r$ mit $0 \leqslant r < m$ auf r abbildet.
- Die Determinante $\det : \mathrm{GL}(n, \mathbb{R}) \to (\mathbb{R} \setminus \{0\}, \cdot)$.
- Die Abbildung $M : S_n \to \mathrm{GL}(n, \mathbb{R})$, die einer Permutation σ die zugehörige Permutationsmatrix $M(\sigma)$ zuordnet.
- Das *Signum* (oder Vorzeichen) $\mathrm{sgn} : S_n \to (\{\pm 1\}, \cdot)$ mit $\mathrm{sgn} = \det \circ M$.

Definition 2.1.3 *Ein Gruppenhomomorphismus $\varphi : G \to H$ heißt* (Gruppen-)Isomorphismus, *falls φ bijektiv ist und die Umkehrabbildung φ^{-1} ebenfalls ein Gruppenhomomorphismus ist.*

Zwei Gruppen G und H heißen isomorph zueinander, *$G \cong H$, falls es einen Gruppenisomorphismus $G \to H$ gibt.*

Man kann zeigen, dass ein bijektiver Gruppenhomomorphismus stets ein Isomorphismus ist, also dass die Umkehrabbildung, falls sie existiert, automatisch ein Homomorphismus ist.

Beispiele:

- Die Gruppe der Vektorraum-Isomorphismen $\mathbb{R}^n \to \mathbb{R}^n$ ist isomorph zu der Matrixgruppe $\mathrm{GL}(n, \mathbb{R})$: jede Basiswahl liefert einen Isomorphismus.
- Sind M und N zwei gleichmächtige Mengen, so liefert jede Bijektion $\beta : M \to N$ einen Gruppenisomorphismus $\hat{\beta} : \mathrm{Sym}(M) \to \mathrm{Sym}(N)$, $\sigma \mapsto \beta \circ \sigma \circ \beta^{-1}$.
  Bis auf Isomorphie ist $\mathrm{Sym}(M)$ also durch die Anzahl der Elemente von M festgelegt; für $|M| = n$ schreibt man dann auch S_n für eine zu $\mathrm{Sym}(M)$ isomorphe Gruppe.
- Man kann zeigen, dass die beiden Gruppen S_3 und $\mathrm{GL}(2, \mathbb{F}_2)$ isomorph zueinander sind.

Definition 2.1.4 *Eine* Untergruppe U *von* G, $U \leqslant G$, *ist eine Teilmenge* $U \subseteq G$, *die abgeschlossen bzgl. der Gruppenoperation ist, das neutrale Element enthält und zu jedem seiner Element auch dessen Inverses.*

Eine *Untergruppe* ist also eine Teilmenge U, die bzgl. der auf U eingeschränkten Operationen selbst wieder eine Gruppe ist.

Beispiele:

- Die Gruppe G selbst und die triviale Gruppe $\{e\}$ sind stets Untergruppen von G.
- Untergruppen von Untergruppen sind Untergruppen: Falls $U \leqslant V$ und $V \leqslant G$, so ist $U \leqslant G$.
- Jeder Untervektorraum eines Vektorraums ist insbesondere auch eine Untergruppe. (Untervektorräume sind die unter Skalarmultiplikation abgeschlossenen Untergruppen.)
- Falls G eine Gruppe ist, so bildet das *Zentrum* $\mathrm{Z}(G) := \{g \in G \mid g \circ h = h \circ g \text{ für alle } h \in G\}$ eine Untergruppe von G.
  Es ist z. B. $\mathrm{Z}(S_3) = \{e\}$ und $\mathrm{Z}\big(\mathrm{GL}(n,\mathbb{R})\big) = \left\{ \begin{pmatrix} r & & 0 \\ & \ddots & \\ 0 & & r \end{pmatrix} \;\middle|\; r \in \mathbb{R}\backslash\{0\} \right\}$.
- Die Gruppe der Vektorraum-Isomorphismen $V \to V$ ist eine Untergruppe von Sym_V.

Definition 2.1.5 *Wenn* $\varphi : G \to H$ *ein Gruppenhomomorphismus ist, so ist der* Kern *definiert als* $\mathrm{Kern}(\varphi) := \{g \in G \mid \varphi(g) = e\}$.

Satz 2.1.6 *Wenn* $\varphi : G \to H$ *ein Gruppenhomomorphismus ist, so ist der Kern von* φ *eine Untergruppe von* G *und das Bild von* φ *eine Untergruppe von* H.

Beweis: Es gilt nach Definition $\varphi(e_G) = e_H$, also ist $e_G \in \mathrm{Kern}(\varphi)$ und $e_H \in \mathrm{Bild}(\varphi)$. Wenn $g_1, g_2 \in \mathrm{Kern}(\varphi)$, so ist $\varphi(g_1 \circ g_2) = \varphi(g_1) \circ \varphi(g_2) = e \circ e = e$, und $\varphi(g_1^{-1}) = \varphi(g_1)^{-1} = e^{-1} = e$, also ist $\mathrm{Kern}(\varphi)$ eine Untergruppe. Wenn $h_1 = \varphi(g_1), h_2 = \varphi(g_2)$, so sind $h_1 \circ h_2 = \varphi(g_1) \circ \varphi(g_2) = \varphi(g_1 \circ g_2)$ und $h_1^{-1} = \varphi(g_1)^{-1} = \varphi(g_1^{-1})$, also ist $\mathrm{Bild}(\varphi)$ eine Untergruppe. □

Der Schnitt einer Menge von Untergruppen von G ist, wie man sich schnell überlegt, wieder eine Untergruppe von G. Also existiert zu jeder Teilmenge $A \subseteq G$ die „kleinste Untergruppe von G, die A enthält“. Diese Untergruppe wird die *von A erzeugte Untergruppe* genannt wird und mit $\langle A \rangle$ bezeichnet wird. Es gilt nun

$$\langle A \rangle = \left\{ {a_1}^{\pm 1} \circ \ldots \circ {a_n}^{\pm 1} \;\middle|\; a_i \in A, n \in \mathbb{N} \right\}$$

mit der Konvention, dass $a^{+1} = a$: Zum einen muss jede A enthaltende Untergruppe auch jedes der Produkte ${a_1}^{\pm 1} \circ \ldots \circ {a_n}^{\pm 1}$ enthalten. Zum andern bildet die Menge dieser Produkte eine A enthaltende Untergruppe: für $n = 0$ enthält man das neutrale Element („leeres Produkt“); mit $n = 0$ enthält man jedes Element von a; die Menge ist offensichtlich unter der Gruppenoperation abgeschlossen, und ebenso unter Inversen, da $({a_1}^{\pm 1} \circ \ldots \circ {a_n}^{\pm 1})^{-1} = {a_n}^{\mp 1} \circ \ldots \circ {a_1}^{\mp 1}$.

Notation: Statt $\langle \{g_1, \ldots, g_k\} \rangle$ schreibt man kurz $\langle g_1, \ldots, g_k \rangle$.

2.2 Zyklische Gruppen

Definition 2.2.1 *Sei G eine Gruppe und $g \in G$. Man definiert für $n \in \mathbb{Z}$:*

$$g^n := \begin{cases} \underbrace{g \circ \cdots \circ g}_{n \text{ mal}} & n > 1 \\ e & n = 0 \\ \underbrace{g^{-1} \circ \cdots \circ g^{-1}}_{|n| \text{ mal}} & n < 1 \end{cases}$$

Insbesondere gilt $g^1 = g$, und g^{-1} stimmt mit dem bisher schon damit bezeichneten Inversen von g überein. Man überzeugt sich leicht, dass für $n < 0$ gilt: $g^n = (g^{-1})^{|n|} = (g^{|n|})^{-1}$. Man kann die Definition auch in induktiver Form angeben:

$$\begin{aligned} g^0 &:= e \\ g^{n+1} &:= g \circ g^n \text{ für } n \in \mathbb{N} \\ g^n &:= (g^{-n})^{-1} \text{ für } n \in \mathbb{Z} \setminus \mathbb{N} \end{aligned}$$

Ist die Gruppe $(G, +)$ additiv geschrieben, so schreibt man ng statt g^n.

Satz 2.2.2 *Es gelten die Potenzgesetze, wie man sie vom Spezialfall der Gruppe $(\mathbb{R} \setminus \{0\}, \cdot)$ her kennt, d. h.:*

$$g^{n+m} = g^n \circ g^m \qquad \text{und} \qquad (g^n)^m = g^{nm}$$

für alle $g \in G$ und $n, m \in \mathbb{Z}$.

Beweis: Für die erste Regel sollte man Fallunterscheidungen nach den Vorzeichen von n und m machen; jeder einzelne Fall ist aber nach Definition klar. Die zweite Regel ist ebenfalls unmittelbar einsichtig für $n, m > 0$. Falls $n = 0$ oder $m = 0$ kommt nach Definition von g^0 auf beiden Seiten e heraus. Ist mindestens einer der beiden Exponenten negativ, so kann man sich durch $g^{-n} = (g^n)^{-1}$ auf den positiven Fall zurückziehen. □

Die Regel $(g^{-1})^{-1} = g$ ist nun übrigens ein Spezialfall des zweiten Potenzgesetzes, ebenso $(g^{-1})^n = (g^n)^{-1}$.

Das erste Potenzgesetz besagt, dass es für jedes $g \in G$ einen Homomorphismus gibt:

$$g^{\smile} : (\mathbb{Z}, +) \to (G, \circ), \; n \mapsto g^n.$$

Das Bild dieses Homomorphismus ist die von g erzeugte Untergruppe $\langle g \rangle$.

Definition 2.2.3 **(a)** *Eine Gruppe heißt* zyklisch, *wenn sie von einem Element erzeugt ist.*

(b) *Die* Ordnung einer Gruppe *G ist die Anzahl der Elemente von G (eine natürliche Zahl oder ∞).*

(c) *Die* Ordnung eines Gruppenelements *$g \in G$,* $\mathrm{ord}(g)$, *ist die Ordnung der von g erzeugten Untergruppe $\langle g \rangle$.*

Beispiele:

- In jeder Gruppe ist e das einzige Element mit Ordnung 1.

- Die Gruppe $(\mathbb{Z}, +)$ ist zyklisch; sie hat zwei Erzeuger: 1 und -1. Die Ordnung der Gruppe ist ∞; alle Elemente außer 0 haben Ordnung ∞.
- Die Gruppe $\mathbb{Z}_{12}$ ist zyklisch: 1, 5, 7 und 11 haben jeweils Ordnung 12 und sind daher Erzeuger. Die Ordnung aller Element sieht man in der folgenden Tabelle:

Element	0	1	2	3	4	5	6	7	8	9	10	11
Ordnung	1	12	6	4	3	12	2	12	3	4	6	12

- Die Gruppe S_n hat Ordnung $n!$ und ist für $n > 2$ nicht zyklisch, da nicht kommutativ.

Bemerkung: Zyklische Gruppen sind kommutativ, denn es gilt

$$g^m \circ g^n = g^{m+n} = g^{n+m} = g^n \circ g^m.$$

Allgemeiner gilt, dass homomorphe Bilder kommutativer Gruppen wieder kommutativ sind, d. h. falls $\varphi : G \to H$ ein surjektiver Gruppenhomomorphismus ist und G kommutativ, dann ist H kommutativ, da $h_1 \circ h_2 = \varphi(g_1) \circ \varphi(g_2) = \varphi(g_1 \circ g_2) = \varphi(g_2 \circ g_1) = \varphi(g_2) \circ \varphi(g_1) = h_2 \circ h_1$.

Satz 2.2.4 *Untergruppen und homomorphe Bilder zyklischer Gruppen sind wieder zyklisch.*

Beweis: Sei $\varphi : G = \langle g \rangle \to H$ ein surjektiver Gruppenhomomorphismus. Dann ist $H = \text{Bild}(\varphi) = \{\varphi(g^n) \mid n \in \mathbb{Z}\} = \{\varphi(g)^n \mid n \in \mathbb{Z}\} = \langle \varphi(g) \rangle$.

Allgemein gilt bei einem Homomorphismus $\varphi : G \to H$ mit $X \subseteq G$, dass $\varphi[\langle X \rangle] = \langle \varphi[X] \rangle$, d. h. die Bilder von Erzeugern sind Erzeuger des Bilds.

Sei nun $U \leqslant \langle g \rangle$. Falls $U = \{e\}$, ist U natürlich zyklisch. Andernfalls gibt es ein $e \neq g^n \in U$, und dann ist auch $g^{-n} = (g^n)^{-1} \in U$. Wähle nun $m > 0$ minimal mit der Eigenschaft, dass $g^m \in U$. Dann ist offensichtlich $\langle g^m \rangle \subseteq U$. Falls $g^n \in U$, so schreibt man $n = qm + r$ mit $r \in \{0, \ldots, m-1\}$ (Division mit Rest) und sieht mit den Potenzgesetzen: $g^r = g^{n-qm} = g^n \circ (g^m)^{-q} \in U$. Aus der Minimalität von m folgt also $r = 0$, und somit $g^n \in \langle g^m \rangle = U$. □

Satz 2.2.5 *Eine zyklische Gruppe ist entweder von unendlicher Ordnung und isomorph zu $(\mathbb{Z}, +)$ oder von endlicher Ordnung m und isomorph zu $(\mathbb{Z}_m, +_m)$.*

Beweis: Sei $G = \langle g \rangle$ zyklisch, und betrachte den surjektiven Homomorphismus $g^{\llcorner} : \mathbb{Z} \to G, n \mapsto g^n$. Falls $g^{\llcorner}$ injektiv ist, so ist es ein Isomorphismus und somit $G \cong \mathbb{Z}$. Andernfalls gibt es $k \neq l$ mit $g^k = g^l$. Dann ist $g^{k-l} = g^k \circ (g^l)^{-1} = e$, d. h. $\text{Kern}(g^{\llcorner}) \neq \{0\}$. Wähle nun $m > 0$ minimal mit $m \in \text{Kern}(g^{\llcorner})$, d. h. mit $g^m = e$. Nach dem Beweis von Satz 2.2.4 ist dann $\text{Kern}(g^{\llcorner}) = \langle m \rangle$; es gilt also $g^k = g^l \iff k - l \in \langle m \rangle$ d. h. wenn m ein Teiler von $k - l$ ist. Somit ist $G = \{g^0, g^1, \ldots, g^{m-1}\}$ und die Abbildung $r \mapsto g^r$ ist ein Isomorphismus $\mathbb{Z}_m \to G$. □

Im nächsten Abschnitt werden wir sehen, dass im zweiten Fall aus dem Homomorphiesatz unmittelbar $G \cong \mathbb{Z}/m\mathbb{Z}$ folgt, wobei $\mathbb{Z}/m\mathbb{Z}$ eine abstrakt gewonnenen, zu $\mathbb{Z}_m$ isomorphe Gruppe ist.

Folgerung 2.2.6 *Die Ordnung von $g \in G$ ist das kleinste $m > 0$ mit $g^m = e$, sofern es existiert; und ∞ sonst. Es gilt genau dann $g^k = e$, wenn m ein Teiler von k ist.*

Bemerkung Wenn $G = \langle g \rangle \cong \mathbb{Z}$ eine unendliche zyklische Gruppe ist und H eine beliebige Gruppe, dann existiert zu jedem $h \in H$ ein eindeutig bestimmter Gruppenhomomorphismus $G \to H$ mit $g \mapsto h$, nämlich die Abbildung $g^n \mapsto h^n$. In diesem Aspekt ähnelt also der Erzeuger einer unendlichen zyklischen Gruppe der Basis eines Vektorraums.

Wenn $G = \langle g \rangle \cong \mathbb{Z}_m$ eine endliche zyklische Gruppe ist und H eine beliebige Gruppe, dann existiert nicht unbedingt ein Gruppenhomomorphismus $G \to H$ mit $g \mapsto h$, denn es gilt $g^m = e$, also muss auch $h^m = \varphi(g)^m = \varphi(g^m) = \varphi(e) = e$ gelten. Dies ist aber das einzige Hindernis, d. h. man kann zeigen, dass ein (eindeutig bestimmter) Gruppenhomomorphismus $G \to H$ mit $g \mapsto h$ genau dann existiert, wenn $h^m = e$, also wenn $\mathrm{ord}(h) \mid \mathrm{ord}(g)$.

Anders als bei Vektorräumen kann man bei Gruppen also nicht Homomorphismen beliebig auf minimalen Erzeugendensystemen vorschreiben, Das liegt daran, dass in Vektorräumen Basen „frei“ sind, also keine besonderen Abhängigkeiten vorweisen können, während in Gruppen zusätzliche Gleichungen gelten können.

Definition 2.2.7 *Ein* Automorphismus *einer Gruppe G ist ein Isomorphismus $G \to G$. Die Menge der Automorphismen von G bildet unter der Hintereinanderausführung von Abbildungen die* Automorphismengruppe *von G,* $\mathrm{Aut}(G)$.

Satz 2.2.8 *Sei $G = \langle g \rangle$ zyklisch. Dann sind die Automorphismen von G genau die Homomorphismen $\varphi : G \to G$, für die $\varphi(g)$ ein Erzeuger von G ist.*

Beweis: $\mathrm{Bild}(\varphi) = \langle \varphi(g) \rangle$, ist ein Homomorphismen $\varphi : G \to G$ genau dann surjektiv, wenn $\varphi(g)$ ein Erzeuger ist. Im endlichen Fall sind surjektive Abbildungen zwischen gleichmächtigen Mengen automatisch injektiv. Im Fall $G = \mathbb{Z}$ gibt es die beiden Erzeuger 1 und -1; die zugehörigen Abbildungen id und $n \mapsto -n$ sind ebenfalls beide injektiv. □

Beispiel: $\mathbb{Z}_{12}$ hat also vier Automorphismen: die Identität, die „Spiegelung“ $n \mapsto -n(+12)$ (die auf der Uhr der Spiegelung an der Mittelsenkrechten entspricht) und zwei durch $1 \mapsto 5$ und $1 \mapsto 7$ bestimmte Automorphismen, mit folgenden Wertetabellen:

	0	1	2	3	4	5	6	7	8	9	10	11
$1 \mapsto 11$	0	11	10	9	8	7	6	5	4	3	2	1
$1 \mapsto 5$	0	5	10	3	8	1	6	11	4	9	2	7
$1 \mapsto 7$	0	7	2	9	4	11	6	1	8	3	10	5

Dies sind, neben der Identität, also die einzigen mit der Addition $+_{12}$ verträglichen Permutationen von $\{0, \ldots, 11\}$.

Zusammenfassende Ergebnisse über zyklische Gruppen

1. Fall: unendliche Ordnung

- $\mathbb{Z}$ ist bis auf Isomorphie die einzige zyklische Gruppe unendlicher Ordnung.
- Die Ordnung von 0 ist 1, die Ordnung aller anderer Elemente ist ∞.
- 1 und -1 sind Erzeuger.
- Die Untergruppen von $\mathbb{Z}$ sind von der Form $\langle n \rangle = n\mathbb{Z} := \{k \cdot n \mid k \in \mathbb{Z}\}$ für $n \in \mathbb{N}$. Es ist dabei $0\mathbb{Z} = \{0\}$ die triviale Untergruppe und $1\mathbb{Z} = \mathbb{Z}$.
- Die homomorphen Bilder von $\mathbb{Z}$ sind $\mathbb{Z}$ selbst und alle $\mathbb{Z}_m$.
- $\mathrm{Aut}(\mathbb{Z}) = (\{\mathrm{id}, n \mapsto -n\}, \circ) \cong \mathbb{Z}_2$.

2. Fall endliche Ordnung

- $\mathbb{Z}_m$ ist bis auf Isomorphie die einzige zyklische Gruppe der Ordnung m
- Die Ordnung von k in $\mathbb{Z}_m$ ist $\frac{m}{\mathrm{ggT}(k,m)}$, denn dies ist die kleinste Zahl l, so dass m ein Teiler von $l \cdot k$ ist.
- Die Erzeuger sind also die zu m teilerfremden Zahlen k.
- Die Untergruppen von $\mathbb{Z}_m$ sind von der Form

$$\langle n \rangle = n\mathbb{Z}_m := \{0, n, 2n, \ldots, (\tfrac{m}{n} - 1)n\} \cong \mathbb{Z}_{\frac{m}{n}}$$

  für alle Teiler n von m, wobei $1\mathbb{Z}_m = \mathbb{Z}_m$ und $m\mathbb{Z}_m = \{0\}$. Die von einem beliebigen Element k erzeugte Untergruppe ist von $\mathrm{ggT}(k,m)$ erzeugt und zu $Z_{\frac{m}{\mathrm{ggT}(k,m)}}$ isomorph.
- Die homomorphen Bilder von $\mathbb{Z}_m$ sind die $\mathbb{Z}_k$ für Teiler k von m.
- $\mathrm{Aut}(\mathbb{Z}_m)$ wird im Abschnitt 2.5 näher bestimmt.

2.3 Nebenklassen und Faktorgruppen

Sei $(G, \cdot, e)$ eine multiplikativ geschriebene Gruppe. Es ist zunächst nützlich, die Gruppenoperation notationell auf Teilmengen von G auszudehnen, indem man für $X, Y \subseteq G$ definiert:

$$\begin{aligned} X \cdot Y &:= \{x \cdot y \mid x \in X, y \in Y\} \\ X^{-1} &:= \{x^{-1} \mid x \in X\} \end{aligned}$$

Außerdem schreibt man $x_0 \cdot Y$ für $\{x_0\} \cdot Y$ und analog $X \cdot y_0$ für $X \cdot \{y_0\}$ und lässt die Multiplikationspunkte auch weg. Man kann sich leicht davon überzeugen, dass gewisse Rechneregeln auch für die Multiplikation von Teilmengen gelten; beispielsweise ist sie assoziativ und es gilt $g^{-1}gX = eX = X$. Es ist aber Vorsicht geboten; zum Beispiel ist $X \cdot X^{-1} \cdot Y$ im allgemeinen verschieden von Y!

Für additiv geschriebene Gruppen definiert man analog $X + Y$ und $-X$.

Sei nun $U \subseteq G$. Man definiert eine zweistellige Relation ${}_U\!\sim$ durch

$$g_1 \, {}_U\!\sim \, g_2 : \Longleftrightarrow \; g_1^{-1} \cdot g_2 \in U.$$

Satz 2.3.1 *${}_U\!\sim$ ist eine genau dann eine Äquivalenzrelation, wenn U eine Untergruppe von G ist.*

Beweis: Wegen $g \, {}_U\!\sim \, g \Leftrightarrow e = g^{-1} \cdot g \in U$ ist die Relation genau dann reflexiv, wenn $e \in U$.

Weiterhin gilt $g \, {}_U\!\sim \, h \Leftrightarrow g^{-1} \cdot h \in U \Leftrightarrow h^{-1} \cdot g = (g^{-1} \cdot h)^{-1} \in U^{-1} \Leftrightarrow h \, {}_{U^{-1}}\!\sim \, g$. Umgekehrt hat man: $u \in U \Leftrightarrow e^{-1}u \in U \Leftrightarrow e \, {}_U\!\sim \, u$ und $u^{-1} \in U \Leftrightarrow u^{-1}e \in U \Leftrightarrow u \, {}_U\!\sim \, e$. Die Relation ist also genau dann symmetrisch, wenn $U = U^{-1}$.

Schließlich: Falls $g \, {}_U\!\sim \, h$ und $h \, {}_U\!\sim \, i$, so hat man $g^{-1}h \in U$ und $h^{-1}i \in U$, also $g^{-1}i = g^{-1}hh^{-1} \in U \cdot U$, d. h. $g \, {}_{UU}\!\sim \, i$. Umgekehrt: sind $g, h \in U$, so ist $g^{-1} \, {}_U\!\sim \, e$ und $e \, {}_U\!\sim \, h$, und außerdem gilt $gh \in U \Leftrightarrow g^{-1} \, {}_U\!\sim \, h$. Die Relation ist also genau dann transitiv, wenn $U = U \cdot U$. □

Definition 2.3.2 *Wenn $U \leqslant G$, so heißen die Äquivalenzklassen von ${}_U\!\sim$ Linksnebenklassen von U in G. Die Äquivalenzklasse von $g \in G$ ist dabei von der Form $gU := \{gu \mid u \in U\}$. Die Menge der Linksnebenklassen von U in G wird mit G/U bezeichnet.*

Weiter definiert man $\sim_U$ durch

$$g_1 \sim_U g_2 : \iff g_2 \cdot g_1^{-1} \in U$$

und sieht analog, dass auch $\sim_u$ genau dann eine Äquivalenzrelation ist, wenn U eine Untergruppe von G ist.

Definition 2.3.3 *Wenn $U \leqslant G$, so heißen die Äquivalenzklassen von $\sim_U$ Rechtsnebenklassen von U in G. Die Äquivalenzklasse von $g \in G$ ist dabei von der Form $Ug := \{ug \mid u \in U\}$. Die Menge der Rechtsnebenklassen von U in G wird mit $U\backslash G$ bezeichnet.*

Bemerkungen

- Es gilt also

$$gU = hU \iff g \,{}_U\!\sim h \iff g^{-1}h \in U \iff g^{-1}hU = U$$

  und

$$Ug = Uh \iff g \sim_U h \iff hg^{-1} \in U \iff Uhg^{-1} = U$$

- Die Rechts- wie Linksnebenklasse von einem $u \in U$, insbesondere also von e, ist $uU = Uu = U$.
- In kommutativen Gruppen sind stets Rechtsnebenklassen gleich Linksnebenklassen, d. h. es gilt $gU = Ug$. Im allgemeinen sind sie verschieden (z. B. in S_3 die Nebenklassen einer Untergruppe der Ordnung 2).

Im additiven Fall schreibt man natürlich $g + U$ für die Links- und $U + g$ für die Rechtsnebenklasse, wobei man die additive Schreibweise üblicherweise für kommutative Gruppen reserviert, in denen beide übereinstimmen.

Satz 2.3.4 *Sei $U \leqslant G$.*
(a) *Alle Nebenklassen haben die gleiche Anzahl von Elementen wie U.*
(b) *Es gibt ebenso viele Rechts- wie Linksnebenklassen.*

Beweis: (a) $U \to Ug, u \mapsto ug$ ist offenbar eine Bijektion mit, da $x \mapsto xg^{-1}$ die Umkehrabbildung ist. Ebenso ist $U \to gU, u \mapsto gu$ eine Bijektion.

(b) $G/U \to U\backslash G, gU \mapsto Ug^{-1}$ ist eine Bijektion: Es gilt

$$gU = hU \iff U = g^{-1}h \iff g^{-1}h \in U \iff Ug^{-1}h = U \iff Ug^{-1} = Uh^{-1}$$

also ist die Abbildung wohldefiniert und injektiv, und $Ug \mapsto g^{-1}U$ ist die ebenso wohldefinierte Umkehrabbildung. □

Folgerung 2.3.5 (Satz von Lagrange) *Wenn G endlich ist und $U \leqslant G$, dann gilt*

$$|G| = |U| \cdot |G : U|,$$

wobei $|G : U| := |G/U| = |U\backslash G|$ der Index von U in G ist. Die Ordnung einer Untergruppe teilt also die Gruppenordnung; insbesondere teilt die Ordnung eines Gruppenelementes die Gruppenordnung.

Anwendungsbeispiel: Wenn G eine Gruppe ist, deren Ordnung $|G| = p$ eine Primzahl ist, und $g \in G \setminus \{e\}$, dann ist die Ordnung von g ungleich 1 und teilt p, ist also gleich p. Also ist G zyklisch $\cong \mathbb{Z}_p$ und jedes Element $\neq e$ ist ein Erzeuger von G.

Bis auf Isomorphie gibt es also nur eine Gruppe von jeder Primzahlordnung. Für zusammengesetzte Zahlen gibt es i. a. mehrere nicht isomorphe Gruppen, z. B. gibt es die kommutative Gruppe $\mathbb{Z}_6$ und die nicht-kommutative Gruppe S_3 der Ordnung 6. Auch für Ordnung 4 gibt es zwei nicht-isomorphe Gruppen; für Ordnung 8 bereits fünf. (Allerdings gibt es bis auf Isomorphie auch nur eine Gruppe der Ordnung 15.)

Sei nun $\varphi : G \to H$ ein Gruppenhomomorphismus. Dann induziert φ, wie jede Abbildung, eine Äquivalenzrelation auf G, nämlich

$$g_1 \sim_\varphi g_2 \quad :\Longleftrightarrow \quad \varphi(g_1) = \varphi(g_2).$$

Es gilt nun

$$\begin{aligned} g_1 \sim_\varphi g_2 &\iff e = \varphi(g_1)^{-2} \cdot \varphi(g_2) = \varphi(g_1^{-1} \cdot g_2) \iff g_1 \ {}_{\mathrm{Kern}(\varphi)}\!\sim g_1 \\ &\iff e = \varphi(g_2) \cdot \varphi(g_1)^{-2} = \varphi(g_2 \cdot g_1^{-1}) \iff g_1 \sim_{\mathrm{Kern}(\varphi)} g_2 \end{aligned}$$

Die Äquivalenzklassen von $\sim_\varphi$ sind also die Rechts- wie Linksnebenklassen des Kerns von φ.

Definition 2.3.6 *Eine Untergruppe U von G heißt* normale Untergruppe *(oder* Normalteiler*), falls ${}_U\!\sim$ und $\sim_U$ die gleiche Relation sind, also falls Linksnebenklassen und Rechtsnebenklassen übereinstimmen. Man schreibt dafür $U \trianglelefteq G$.*

Kerne von Homomorphismen sind also normale Untergruppen. In kommutativen Gruppen sind offenbar alle Untergruppen normal. Außerdem sind $\{e\}$ und G stets normale Untergruppen einer Gruppe G.

Eine Untergruppe U vom Index 2 ist normal, denn da U eine Rechts- wie Linksnebenklasse ist, ist $G \setminus U$ die jeweils andere Nebenklasse, also auch eine Rechts- und Linksnebenklasse.

Die von einer Transposition erzeugten Untergruppen der S_3 vom Index 2 sind die kleinsten Beispiele von nicht normalen Untergruppen.

Definition 2.3.7 *Eine Äquivalenzrelation $\sim$ auf einer Gruppe G heißt* Kongruenzrelation*, falls die Menge $G/_\sim$ der Äquivalenzklassen so zu einer Gruppe gemacht werden kann, dass die natürliche Abbildung $G \to G/_\sim, g \mapsto g/_\sim$ ein Homomorphismus ist.*

r Gruppe ist genau dann eine Kongruenzrelation, wenn die Festlegung

$$(g/_\sim) \cdot (h/_\sim) := (g \cdot h)/_\sim$$

wohldefiniert ist, wenn also die Äquivalenzklasse des Produktes unabhängig von der Wahl der Repräsentanten ist. (Die Existenz des neutrales Elements und der Inversen ergibt sich dann automatisch als $e/_\sim$ bzw. als $(g/_\sim)^{-1} = (g^{-1}/_\sim.)$

Satz 2.3.8 *Die Kongruenzrelationen für Gruppen sind genau die Nebenklassenrelationen normaler Untergruppen.*

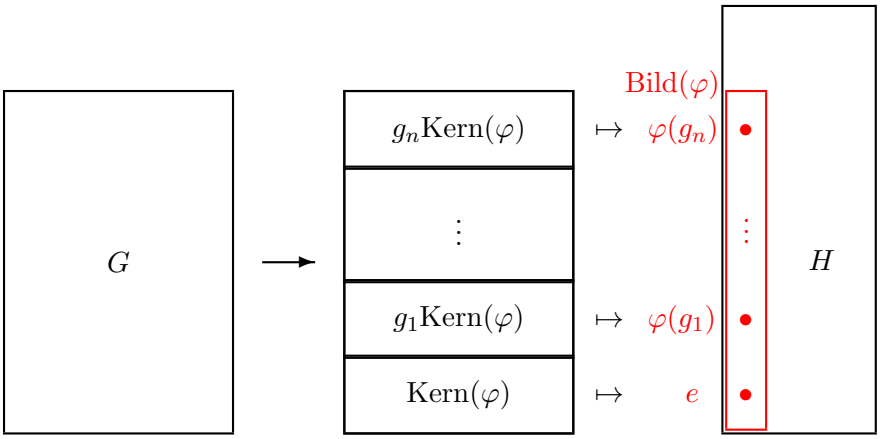
BEWEIS: $\sim$ ist genau dann eine Kongruenzrelation, wenn die natürliche Abbildung $G \to G/_\sim$ ein Homomorphismus ist; also ist sie nach den Überlegungen oben die Nebenklassenrelation des Kerns.

Sei umgekehrt N eine normale Untergruppe und $g_1N = g_2N, h_1N = h_2N$, also $g_2 = g_1n$ und $h_2 = h_1n'$ mit $n, n' \in N$. Wegen $h_1N = Nh_1$ ist $nh_2 = h_1n''$ für ein $n'' \in N$. Es folgt $g_2h_2 = g_1nh_1n' = g_1h_1n''n' \in g_1h_1N$ und somit $g_2h_2N = g_1h_1N$. □

Definition 2.3.9 *Ist N eine normale Untergruppe von G, so heißt die Gruppenstruktur auf der Menge G/N der Nebenklassen von N in G die* Faktorgruppe *oder* Quotientengruppe *von G nach N.*

Folgerung 2.3.10 (Homomorphiesatz) *Ist $\varphi : G \to H$ ein Gruppenhomomorphismus, so kann man φ zusammensetzen als Komposition der folgenden Homomorphismen:*

$$\begin{array}{ccccccc} G & \longrightarrow & G/\mathrm{Kern}(\varphi) & \xrightarrow{\cong} & \mathrm{Bild}(\varphi) & \longrightarrow & H \\ g & \mapsto & g\,\mathrm{Kern}(\varphi) & \mapsto & \varphi(g) & \mapsto & \varphi(g) \\ & & \text{surjektiv} & & \text{bijektiv} & & \text{injektiv} \end{array}$$



Beispiele

- Die Abbildung $\mathbb{Z} \to \mathbb{Z}_m$, die den Rest bei der Division durch m angibt, ist ein surjektiver Homomorphismus mit Kern $m\mathbb{Z}$. Die Faktorgruppe $\mathbb{Z}/m\mathbb{Z}$ ist somit isomorph zu $\mathbb{Z}_m$. Der Unterschied zwischen den beiden Gruppen besteht darin, dass die Gruppenelemente von $\mathbb{Z}/m\mathbb{Z}$ Mengen von ganzen Zahlen sind, wobei die Addition die gewöhnliche Addition von $\mathbb{Z}$ auf den Repräsentanten ist; die Elemente von $\mathbb{Z}_m$ dagegen sind die ganzen Zahlen $0, \ldots, m-1$, die Addition darauf ist aber die „Addition modulo m“.

- Das Signum $\mathrm{sgn} : S_n \to (\{\pm 1\}, \cdot) \cong \mathbb{Z}_2$ ist ein Homomorphismus, dessen Kern die *Alternierende Gruppe A_n* ist. Es gilt also $S_n/A_n \cong \mathbb{Z}_2$.

  Die Drehgruppe D des Würfels permutiert die vier Raumdiagonalen des Würfels; dadurch erhält man einen Homomorphismus $D \to S_4$, von dem man sich überzeugen kann, dass er injektiv ist. Da beide Gruppen die gleiche Anzahl von Elementen haben, ist also $D \cong S_4$. Außerdem permutiert D die drei Mittelsenkrechten des Würfels; dadurch erhält man einen Homomorphismus $S_4 \cong D \to S_3$, der surjektiv ist und dessen Kern die sogenannte *Kleinsche Vierergruppe* ist.

Für $n \geqslant 5$ hat die S_n keine anderen normalen Untergruppen außer $\{e\}$, A_n und S_n. Damit hängt zusammen, dass es für Polynomgleichungen vom Grad mindestens 5 keine allgemeine Lösungsformel mit Wurzelausdrücken gibt.

- $\det : \mathrm{GL}(n, \mathbb{R}) \to (\mathbb{R} \setminus \{0\}, \cdot)$ ist ein surjektiver Gruppenhomomorphimus, dessen Kern die Gruppe $\mathrm{SL}(n, \mathbb{R})$ der Matrizen der Determinante 1 ist. Die Determinante liefert also auch einen Homomorphismus der linearen Abbildungen $\mathbb{R}^n \to \mathbb{R}^n$ in die multiplikative Gruppe von $\mathbb{R}$, dessen Kern die orientierungs- und volumenerhaltenden Abbildungen sind.

- Normale Untergruppen der S_3 sind $\{e\}$, A_3 und S_3; die drei zu $\mathbb{Z}_2$ isomorphen „Spiegelungsgruppen“ sind nicht normal.

Untergruppen und homomorphe Bilder bieten die Möglichkeit, aus einer Gruppe „kleinere Teile“ zu gewinnen und u. U. die Struktur der Gruppe zu verstehen, indem man z. B. die Struktur eines Normalteilers und der zugehörigen Faktorgruppe analysiert. Die einfachste Möglichkeit, wie eine Gruppe aus zwei oder mehreren Bausteinen zusammengesetzt werden kann, ist durch die folgende Konstruktion des sogenannte direkten Produkts beschrieben:

Definition 2.3.11 *Das* direkte Produkt $G_1 \times \cdots \times G_n$ *von Gruppen* $(G_1, \cdot), \ldots, (G_n, \cdot)$ *ist das kartesische Produkt der zugrundeliegenden Mengen mit der komponentenweise Operation*

$$(g_1, \ldots, g_n) \cdot (h_1, \ldots, h_n) := (g_1 \cdot h_1, \ldots, g_n \cdot h_n).$$

Man sieht leicht, dass das direkte Produkt wieder eine Gruppe ist mit neutralem Element $(e_{G_1}, \ldots, e_{G_n})$ und Inversem $(g_1, \ldots, g_n)^{-1} = (g_1^{-1}, \ldots, g_n^{-1})$. Sind alle Gruppen kommutativ, so ist auch das direkte Produkt kommutativ.

(Anmerkung: Es gibt „natürliche“ Isomorphismen zwischen $(G_1 \times G_2) \times G_3$, $G_1 \times (G_2 \times G_3)$ und $G_1 \times G_2 \times G_3$, die notationell im Verschieben bzw. Weglassen der Klammern bestehen. Üblicherweise unterscheidet man in der Mathematik daher nicht zwischen diesen drei Objekten (und analog für größere n), d. h. man begeht hier oft eine Art „Typenfehler“.)

Beispiele

- Die Gruppentafel von $\mathbb{Z}_2 \times \mathbb{Z}_2$ ist:

$+$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$(0,0)$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$(1,0)$	$(1,0)$	$(0,0)$	$(1,1)$	$(0,1)$
$(0,1)$	$(0,1)$	$(1,1)$	$(0,0)$	$(1,0)$
$(1,1)$	$(1,1)$	$(0,1)$	$(1,0)$	$(0,0)$

  Man sieht, dass alle Elemente $\neq (0,0)$ die Ordnung 2 haben. Diese Gruppe ist also nicht isomorph zu $\mathbb{Z}_4$. (Man kann zeigen, dass es keine weiteren Gruppen der Ordnung 4 gibt: jede ist entweder zu $\mathbb{Z}_4$ oder zu $\mathbb{Z}_2 \times \mathbb{Z}_2$ isomorph.)

- In $\mathbb{Z}_2 \times \mathbb{Z}_3$ dagegen sieht man leicht, dass das Element $(1,1)$ die Ordnung 6 hat, dass also $\mathbb{Z}_2 \times \mathbb{Z}_3$ zyklisch ist und damit zu $\mathbb{Z}_6$ isomorph.
  Allgemein gilt, dass die Ordnung eines Elementes $(g_1, \ldots, g_n)$ in $G_1 \times \cdots \times G_n$ das kleinste gemeinsame Vielfache der Ordnungen der g_i ist.

- Die Symmetriegruppe D_4 eines Quadrats hat eine normale, zu $\mathbb{Z}_4$ isomorphe Untergruppe, nämlich die Drehungen des Quadrats, und zu $\mathbb{Z}_2$ isomorphe Untergruppen, die von jeweils einer Spiegelung erzeugt werden. D_4 ist also von einer Drehung (um 90° bzw. um 270°) und von einer (beliebigen) Spiegelung erzeugt (denn die davon erzeugte Untergruppe enthält mit den Drehungen und einer Spiegelung mindestens 5 Elemente, und es gibt keinen echten Teiler von $8 = |D_4|$, der mindestens 5 ist). D_4 ist aber nicht isomorph zu $\mathbb{Z}_4 \times \mathbb{Z}_2$, da D_4 nicht kommutativ ist. Dies ist also ein Beispiel einer Gruppe, die auf kompliziertere Weise aus zwei Bausteinen aufgebaut ist. (Analog gilt dies auch schon für S_3, die Symmetriegruppe eines gleichseitigen Dreiecks).

2.4 Ringe

⋮

2.5 Die endlichen Ringe $\mathbb{Z}/m\mathbb{Z}$

⋮

Kapitel 3

Analysis mehrerer Veränderlicher

3.1 Funktionen mehrerer Veränderlicher

⋮

3.2 Topologie des $\mathbb{R}^n$

⋮

3.3 Differenzierbarkeit

⋮

3.4 Höhere Ableitungen

⋮

3.5 Höherdimensionale Integration

⋮