GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Lineare Algebra und Diskrete Mathematik für die Informatik

Prof. Dr. Bastian von Harrach

Goethe-Universität Frankfurt am Main
Institut für Mathematik

Wintersemester 2022/23

`http://numerical.solutions`

Vorläufige unvollständige Version, wird während der Vorlesungszeit laufend ergänzt und korrigiert.
Letzte Aktualisierung: 4. Februar 2023

Inhaltsverzeichnis

Kapitel 1

Grundlegende Begriffe und algebraische Strukturen

Wir wiederholen in diesem Kapitel zunächst ohne Beweise einige mathematische Grundlagen und führen Begriffe für häufig anzutreffende mathematische Strukturen ein.

1.1 Zahlen und Mengen

1.1.1 Der naive Mengen- und Zahlbegriff

Die moderne Mathematik baut auf dem Begriff der *Menge* auf. Eine Menge ist eine „Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unsrer Anschauung oder unseres Denkens (welche die ‚Elemente' von M genannt werden) zu einem Ganzen" (Georg Cantor, Mathematische Annalen 1895). Dieser sogenannte *naive Mengenbegriff* ist nicht komplett widerspruchsfrei[1], für praktische Anwendungen der Mathematik aber völlig ausreichend.

Wir schreiben Mengen mit *geschweiften Klammern*, z.B. bezeichnet $\{1,2,5\}$ die Menge, die als Elemente die Zahlen 1, 2 und 5 enthält. Wir schreiben:

$$1 \in \{1,2,5\}, \quad 2 \in \{1,2,5\}, \quad 5 \in \{1,2,5\}, \quad \text{aber} \quad 4 \notin \{1,2,5\}.$$

Für Mengen ist nur relevant, welche verschiedenen Objekte enthalten sind. Die Reihenfolge der Elemente spielt dabei keine Rolle, z.B. ist

$$\{1,2,5\} = \{1,1,1,1,1,2,2,2,2,2,5\} = \{2,5,5,5,5,2,1\} \neq \{1,1,1,2,2\}$$

[1] Ein bekanntes Paradox ist die Frage, ob sich die Menge aller Mengen, die sich nicht selbst enthalten, selbst enthält.

Die Menge der *natürlichen Zahlen* wird gebildet, indem beginnend mit 1 zu jeder enthalten Zahl auch die nächstgrößere (d.h. um 1 erhöhte) Zahl hinzugenommen wird

$$\mathbb{N} := \{1, 2, 3, 4, 5, \ldots\}$$

$\mathbb{N}$ enthält unendlich viele Elemente. Beginnt man diese Konstruktion mit 0, so erhält man die *natürlichen Zahlen mit Null*

$$\mathbb{N}_0 := \{0, 1, 2, 3, 4, 5, \ldots\}.$$

Wir nehmen dabei die Existenz der natürlichen Zahlen als naturwissenschaftliche Gegebenheit hin (*naiver Zahlbegriff*).

Durch Hinzunahme der negativen Zahlen erhält man die Menge der *ganzen Zahlen*

$$\mathbb{Z} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

Durch Hinzunahme von Brüchen erhält man die Menge der *rationalen Zahlen*

$$\mathbb{Q} := \{\frac{p}{q} : \ p, q \in \mathbb{Z},\ q \neq 0\}.$$

Man kann zeigen, dass Dezimalzahlen mit unendlich vielen und sich dabei nicht periodisch wiederholenden Nachkommastellen keine rationale Zahlen sind (z.B. $\pi = 3,1415\ldots$ oder $\sqrt{2} = 1,4142\ldots$), sondern sich nur als Grenzwerte rationaler Zahlen ergeben. Durch Hinzunahme solcher Grenzwerte erhält man die Menge der *reellen Zahlen* $\mathbb{R}$.

1.1.2 Mengenlehre

Sei M eine Menge. Wir haben bereits die folgende Notation verwendet.

- $m \in M$: m ist *Element* der Menge M (d.h. m ist in M enthalten)
- $m \notin M$: m ist nicht Element der Menge M (d.h. m ist nicht in M enthalten)

Wir führen außerdem ein

- $M = \emptyset$: M ist die *leere Menge* (die gar keine Elemente enthält).
- $|M|$: Anzahl der Elemente von M (für unendliche Mengen: $|M| = \infty$).

Mengen lassen sich über die explizite Angabe ihrer Elemente oder die Angabe einer für die Mengenzugehörigkeit entscheidenden Eigenschaft definieren:

- Umfangsdefinition: $M = \{2,4,6,8\}$
- Inhaltsdefinition: $M = \{m : \ m \text{ ist gerade natürliche Zahl} \leq 8\}$

Für zwei Mengen A, B schreiben wir:

- $A \subseteq B$: A ist *Teilmenge* von B, d.h. alle Elemente von A sind auch in B enthalten. Für alle $a \in A$ gilt auch $a \in B$.
- $A = B$: A ist *gleich* B, d.h. A enthält genau die gleichen Elemente wie B, es gilt $A \subseteq B$ und $B \subseteq A$.
- $A \subset B$: A ist *echte Teilmenge* von B, d.h. es gilt $A \subseteq B$ aber nicht $A = B$.

Wir verwenden auch analoge gespiegelte Notationen, z.B. bedeutet $B \supset A$ das gleiche wie $A \subset B$ und statt $a \in A$ können wir auch $A \ni a$ schreiben.[2]

Für endliche Mengen A und B gilt:

$$\text{Falls} \quad |A| = |B| < \infty \quad \text{und} \quad A \subseteq B, \quad \text{dann ist} \quad A = B.$$

Für unendliche Mengen gilt dies nicht, z.B. ist $|\mathbb{N}| = \infty$, $|\mathbb{Z}| = \infty$, aber $\mathbb{N} \subset \mathbb{Z}$.

Mit *Mengenoperationen* lassen sich aus bekannten Mengen neue kombinieren. Seien A und B zwei Mengen.

- $A \cup B := \{a : \ a \in A \text{ oder } a \in B\}$: *Vereinigung* von A und B
- $A \cap B := \{a : \ a \in A \text{ und } a \in B\}$: *Schnittmenge* von A und B
- $A \setminus B := \{a : \ a \in A \text{ aber } a \notin B\}$: *Differenzmenge* von A und B („A ohne B“)

Für endliche Mengen A, B gilt

$$|A \setminus B| = |A| - |A \cap B| \quad \text{und} \quad |A \cup B| = |A| + |B| - |A \cap B|.$$

Zwei Mengen A, B heißen disjunkt, falls $A \cap B = \emptyset$. Für die Vereinigung disjunkter Mengen schreibt man auch $A \dot{\cup} B$. Ist aus dem Zusammenhang klar, welche Obermenge $M \supseteq A$ gemeint ist, dann schreiben wir auch

- $A^C := M \setminus A$: *Komplement* von A, d.h. alles was nicht in A enthalten ist.

[2] Achtung: In der Literatur wird manchmal auch $\subset$ für Teilmengen verwendet und echte Teilmengen mit $\subsetneq$ gekennzeichnet.

Z.B. ist die Menge der geraden Zahlen $\{2,4,6,\ldots\}$ das Komplement der ungeraden Zahlen $\{1,3,5,\ldots\}$, wenn wir stillschweigend voraussetzen, dass als Obermenge zur Komplementbildung die Menge der natürlichen Zahlen gemeint ist.

Für die eingeführten Mengenoperationen gelten die folgenden Rechenregeln. Es seien A, B und C Mengen.

- Transitivität: Gilt $A \subseteq B$ und $B \subseteq C$, so gilt auch $A \subseteq C$.
- Kommutativität: $A \cup B = B \cup A$, $A \cap B = B \cap A$.
- Assoziativität: $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$.
- Distributivität: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
- De-morgansche Regeln: $(A \cap B)^C = A^C \cup B^C$, $(A \cup B)^C = A^C \cap B^C$..

Die Elemente einer Menge können selbst Mengen sein, z.B. ist

$$M := \{\{1\},\{1,2\}\}$$

die Menge, die als Elemente die Menge $\{1\}$ und die Menge $\{1,2\}$ enthält. Dabei wird streng zwischen der Menge und ihren Elementen unterschieden. So gilt in diesem Beispiel $\{1\} \in M$, jedoch $1 \notin M$ und $\{1\} \not\subseteq M$. Zu einer Menge M definieren wir

- Potenzmenge $\mathscr{P}(A) := \{B : B \subseteq A\}$.

Für $A = \{1,2,3\}$ ergibt sich z.B.

$$\mathscr{P}(A) = \{\emptyset,\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}.$$

Es gilt stets $\emptyset \subseteq A$ und $A \subseteq A$ und deshalb $\emptyset \in \mathscr{P}(A)$ und $A \in \mathscr{P}(A)$. Für eine Menge A erhalten wir eine Teilmenge dadurch, dass wir für jedes Element entscheiden, ob es Element der gebildeten Teilmenge sein soll oder nicht. Für das erste Element gibt es also zwei Möglichkeiten, für das nächste wieder zwei, usw.. Für eine Menge A mit $|A|$ Elementen gibt es daher $2^{|A|}$ Möglichkeiten eine Teilmenge zu bilden und entsprechend gilt

$$|\mathscr{P}(A)| = 2^{|A|}.$$

1.1.3 Tupel und kartesische Produkte

Seien A und B Mengen. Zwei Elemente $a \in A$ und $b \in B$ können wir zu einem geordneten Paar (a,b) zusammenfassen. Diese Paare nennen wir auch *Tupel* und die Menge aller solcher Tupel heißt

- kartesisches Produkt $A \times B := \{(a,b) : \ a \in A, \ b \in B\}$

Im Unterschied zur Mengenbildung ist bei Tupeln die Reihenfolge der Elemente wichtig. Für zwei Tupel (a,b) und (c,d) gilt Gleichheit $(a,b) = (c,d)$ nur dann, wenn $a = c$ und $b = d$ gilt. Es ist also z.B.

$$\{1,2\} = \{2,1\} \subseteq \mathbb{R}, \quad \text{aber} \quad (1,2) \neq (2,1) \in \mathbb{R} \times \mathbb{R}.$$

Analog fassen wir auch mehr als zwei Objekte zusammen. Gegeben n Mengen $A_1,\ldots,A_n$ definieren wir

$$A_1 \times A_2 \times \cdots \times A_n := \{(a_1,a_2,\ldots,a_n) : \ a_i \in A_i \ \text{ für alle } i = 1,\ldots,n\}.$$

Wird n-mal die gleiche Menge verwendet, so schreiben wir auch

$$A^n := A \times A \times \cdots \times A.$$

Tupel aus 3 Elementen nennt man auch *Tripel*. Tupel aus n Elementen nennt man auch n-Tupel.

1.2 Aussagenlogik und Funktionen

1.2.1 Logische Aussagen

Eine mathematische (auch:logische) *Aussage* ist eine präzise formulierte Feststellung, die entweder wahr oder falsch ist. „4 ist durch 2 teilbar (ohne Rest)" ist eine offenbar wahre Aussage. „Die letzte Ziffer von 123^{25} ist eine 3" ist eine Aussage, die entweder wahr oder falsch ist, auch wenn wir es ggf. noch nicht wissen „Gerade Zahlen sind besser als ungerade Zahlen." ist keine mathematische Aussage, solange nicht präzisiert wurde, was genau mit „besser" gemeint ist.

Seien A und B Aussagen. Dann können wir bilden:

- $\neg A = \overline{A}$: *Negation von* A, genau dann wahr, wenn A falsch ist.

- $A \wedge B$: *Konjunktion*, genau dann wahr, wenn A und B wahr sind.
- $A \vee B$: *Disjunktion*, genau dann wahr, wenn A oder B (oder beide) wahr sind.

In der Informatik wird eine Variable, die die Werte „wahr“ und „falsch“ annehmen kann, als boolsche Variable bezeichnet und üblicherweise 1 für „wahr“ und 0 für „falsch“ gewählt. Die oben definierten logischen Operatoren werden auch NOT, AND und OR genannt. Außerdem sind gebräuchlich

- A XOR B: *Exklusives Oder*, genau dann wahr, wenn A oder B wahr ist, aber nicht beide wahr sind (also $(A \wedge \neg B) \vee (\neg B \wedge A)$)
- A NAND B: *Negierte Konjunktion*, d.h. $\neg(A \wedge B)$.

Eine Aussage kann aus einer anderen folgen. Dies schreiben wir wie folgt:

- $A \Longrightarrow B$: *Implikation*, aus A folgt B
- $A \Longleftrightarrow B$: *Äquivalenz*, aus A folgt B und aus B folgt A

Im Falle der Implikation $A \Longrightarrow B$ sagt man auch:

- A ist *hinreichend* für B.
- B ist *notwendig* für A.
- B gilt, *wenn* A gilt.

If Falle der Äquivalenz $A \Longleftrightarrow B$, sagt man auch:

- A ist *notwendig und hinreichend* für B.
- B ist *notwendig und hinreichend* für A.
- A gilt, *genau dann wenn* B gilt.
- B gilt, *genau dann wenn* A gilt.

Statt $A \Longleftrightarrow B$ schreibt man auch kurz $A = B$.

Für Aussagen A, B und C gelten die folgenden Rechenregeln:

- Transitivität: Aus $A \Longrightarrow B$ und $B \Longrightarrow C$ folgt auch $A \Longrightarrow C$.
  Aus $A \Longleftrightarrow B$ und $B \Longleftrightarrow C$ folgt auch $A \Longleftrightarrow C$.

- Kommutativität: $\begin{array}{lll} A \vee B & \iff & B \vee A \\ A \wedge B & \iff & B \wedge A \end{array}$
- Assoziativität: $\begin{array}{lll} (A \vee B) \vee C & \iff & A \vee (B \vee C), \\ (A \wedge B) \wedge C & \iff & A \wedge (B \wedge C). \end{array}$
- Distributivität: $\begin{array}{lll} A \wedge (B \vee C) & \iff & (A \wedge B) \vee (A \wedge C), \\ A \vee (B \wedge C) & \iff & (A \vee B) \wedge (A \vee C). \end{array}$
- Doppelte Negation: $\neg(\neg A) \iff A$
- De-morgansche Regeln: $\begin{array}{lll} \neg(A \vee B) & \iff & (\neg A) \wedge (\neg B) \\ \neg(A \wedge B) & \iff & (\neg A) \vee (\neg B) \end{array}$
- Kontraposition: $(A \implies B) \iff (\neg B \implies \neg A)$

In einem mathematischen Beweis zeigt man, dass aus einer Voraussetzung V eine behauptete Aussage A folgt, also $V \implies A$. Dazu nutzt man die obigen Regeln aus:

- Direkter Beweis: Zeige $V \implies A$.
- Indirekter Beweis (Kontraposition): Zeige $\neg A \implies \neg V$.
- Indirekter Beweis (Widerspruch): Zeige $V \wedge \neg A \implies B$ und B ist falsch.

Dabei können wir verwenden:

- Zwischenschritte: $V \implies Z_1, Z_1 \implies Z_2, \ldots, Z_{n-1} \implies Z_n, Z_n \implies A$.
- Fallunterscheidungen: $V \implies F_1 \vee F_2, \quad F_1 \implies A\,, \quad F_2 \implies A$

(analog beim Beweis durch Kontraposition oder Widerspruch).

1.2.2 Quantoren

Aussagen können von einem Element x einer Menge X abhängen (oder auch von mehreren). Dazu führen wir den *Allquantor* und den *Existenzquantor* ein:

- $\forall x \in X : A(x)$. *Für alle* $x \in X$ gilt die Aussage $A(x)$.
- $\exists x \in X : A(x)$. *Es existiert (mindestens) ein* $x \in X$ für das $A(x)$ gilt.
- $\exists! x \in X : A(x)$. *Es existiert genau ein* $x \in X$ für das $A(x)$ gilt.

Für Quantoren gelten die folgenden Regeln:

- Negation von Quantoren:
$$\begin{array}{lll} \neg(\forall x \in X:\ A(x)) & \iff & \exists x \in X:\ \neg A(x) \\ \neg(\exists x \in X:\ A(x)) & \iff & \forall x \in X:\ \neg A(x) \end{array}$$

- Vertauschung/Kombination von Allquantoren:
$$\begin{aligned} \forall x \in X:\ \forall y \in Y:\ A(x,y) &\iff \forall y \in Y:\ \forall x \in X:\ A(x,y) \\ &\iff \forall (x,y) \in X \times Y:\ A(x,y) \end{aligned}$$

- Vertauschung/Kombination von Existenzquantoren:
$$\begin{aligned} \exists x \in X:\ \exists y \in Y:\ A(x,y) &\iff \exists y \in Y:\ \exists x \in X:\ A(x,y) \\ &\iff \exists (x,y) \in X \times Y:\ A(x,y) \end{aligned}$$

- Vertauschung von Existenz- und Allquantoren:
$$\exists x \in X:\ \forall y \in Y:\ A(x,y) \implies \forall y \in Y:\ \exists x \in X:\ A(x,y).$$

Im letzten Punkt gilt nur „$\implies$", die umgekehrte Vertauschung von Existenz- und Allquantoren gilt nicht! Z.B. ist die Aussage

$$\forall n \in \mathbb{N}:\ \exists m \in \mathbb{N}:\ m \geq n$$

richtig (zu jeder natürlichen Zahl gibt es eine noch größere Zahl), aber die Aussage

$$\exists m \in \mathbb{N}:\ \forall n \in \mathbb{N}:\ m \geq n$$

ist falsch (es gibt keine größte natürliche Zahl).

1.2.3 Funktionen

Seien X und Y Mengen. Eine *Funktion von X nach Y* ordnet jedem Element $x \in X$ genau ein Element $y \in Y$ zu. Wir schreiben:

$$f:\ X \to Y, \quad x \mapsto y = f(x)$$

X heißt *Definitionsmenge*, Y heißt *Zielmenge*. Funktionen werden auch *Abbildungen*, *Operatoren* oder (besonders für $Y = \mathbb{R}$) Funktionale genannt.

Eine Funktion $f:\ X \to Y$ heißt

- injektiv, falls zwei verschiedenen Elemente $x, x' \in X$ nicht das gleiche y zugeordnet wird, d.h. in Formelsprache:

$$\forall x, x' \in X : \; f(x) = f(x') \quad \Longrightarrow \quad x = x'.$$

- surjektiv, falls jedes Element von Y erreicht wird, d.h. in Formelsprache:

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

- bijektiv, falls f injektiv und surjektiv ist, d.h. in Formelsprache:

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

Für bijektive Funktionen können wir die *Umkehrfunktion* f^{-1} definieren durch

$$f^{-1} : \; Y \to X, \quad f^{-1}(y) := x, \quad \text{wobei } x \in X \text{ erfüllt } f(x) = y.$$

Sind $f : \; X \to Y$, $g : \; Y \to Z$, zwei Funktionen zwischen Mengen X, Y, Z, dann definieren wir

- Komposition/Hintereinanderausführung

$$g \circ f : \; X \to Z, \quad x \mapsto (g \circ f)(x) := g(f(x)).$$

Sind f und g beide bijektiv, dann ist $f \circ g$ bijektiv und es gilt

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

Wir wenden Funktionen auch auf Mengen an und erhalten Mengen der erreichten Elemente. Es sei $f : \; X \to Y$ eine Funktion und $X' \subseteq X$, $Y' \subseteq Y$ Teilmengen der Definitions- und Zielmenge. Dann definieren wir

- Bild der Menge X' unter f:

$$f(X') := \{f(x) : x \in X'\} = \{y \in Y : \; \exists x \in X' : \; f(x) = y\} \subseteq Y,$$

- Urbild der Menge Y' unter f:

$$f^{-1}(Y') := \{x \in X : \; f(x) \in Y'\} \subseteq X.$$

Die Schreibweise f^{-1} verwendet man sowohl für die Umkehrfunktion als auch für das Urbild einer Menge. Für ein einzelnes Element $y \in Y$ ist $f^{-1}(y) \in X$ die Anwendung der Umkehrfunktion und nur für bijektive Funktionen definiert. Das Urbild $f^{-1}(Y')$ einer Menge Y' ist aber auch für nicht bijektive Funktionen definiert. Insbesondere sind die Mengen

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

auch für nicht bijektive Funktionen definiert. Für nicht bijektive Funktionen können die Mengen $f^{-1}(\{y\})$ leer sein oder mehrere Elemente enthalten. Für bijektive Funktionen bestehen diese Mengen für jedes $y \in Y$ immer aus genau einem Element, nämlich $f^{-1}(\{y\}) = \{f^{-1}(y)\}$.

Da eine Funktion jedem Element aus X ein Element aus Y zuordnet gilt immer $|f(X)| \leq |X|$ und für injektive Funktionen gilt $|f(X)| = |X|$. Surjektive Funktionen erreichen jedes Element von Y, daher gilt für sie $f(X) = Y$ und damit insbesondere auch $|f(X)| = |Y|$. Im Falle endlicher Mengen gilt sogar:

- Ist $|X| < \infty$, dann gilt

$$f \text{ injektiv} \quad \Longleftrightarrow \quad |f(X)| = |X|.$$

- Ist $|Y| < \infty$, dann gilt

$$f \text{ surjektiv} \quad \Longleftrightarrow \quad |f(X)| = |Y|.$$

- Ist $|X| = |Y| < \infty$, dann gilt

$$f \text{ injektiv} \quad \Longleftrightarrow \quad f \text{ surjektiv} \quad \Longleftrightarrow \quad f \text{ bijektiv.}$$

1.3 Algebraische Strukturen

Zahlen lassen sich addieren oder multiplizieren, Mengen lassen sich vereinigen oder schneiden, Aussagen lassen sich durch Konjunktion oder Disjunktion kombinieren, Drehungen oder Verschiebungen lassen sich hintereinanderführen. In all diesen Fällen werden zwei mathematische Objekte miteinander zu einem neuen Objekt verknüpft. Dabei beobachtet man häufig gewisse Gesetzmäßigkeiten, z.B. Kommutativität oder Assoziativität. Wir abstrahieren und formulieren diese Eigenschaften für beliebige Mengen, geben den so entstehenden Strukturen Namen und nennen wichtige Beispiele.

Sei A eine Menge.

- Eine Verknüpfung ist eine Funktion

$$\circ : A \times A \to A,$$

  also eine Funktion die zwei Elemente aus A auf ein Element in A abbildet (*verknüpft*). Für $a, b \in A$ schreiben wir statt $\circ(a,b)$ auch $a \circ b$.

Auf den natürlichen Zahlen $\mathbb{N}$ bildet die Addition „$+$´´ eine Verknüpfung und diese ist assoziativ, d.h.

$$(l+m)+n = l+(m+n) \quad \text{für alle } l,m,n \in \mathbb{N}.$$

In den natürlichen Zahlen mit Null $\mathbb{N}_0$ ist die Addition ebenfalls eine assoziative Verknüpfung, die außerdem noch ein sogenanntes neutrales Element (nämlich die Null) besitzt, d.h. ein Element durch dessen Addition sich eine Zahl nicht ändert. In den ganzen Zahlen $\mathbb{Z}$ ist die Addition weiterhin eine assoziative Verknüpfung mit neutralem Element und es existiert außerdem zu jedem Element ein inverses Element (nämlich die negierte Zahl) die eine Addition mit dieser Zahl rückgängig macht. Gleiches gilt für $\mathbb{Q}$ und $\mathbb{R}$ bezüglich der Addition, für $\mathbb{Q} \setminus \{0\}$ und $\mathbb{R} \setminus \{0\}$ auch bezüglich der Multiplikation (mit 1 als neutralem Element und dem Kehrwert $1/q = q^{-1}$ als inversem Element zu einer Zahl q), und auch für viele andere wichtige Strukturen, z.B. der Hintereinanderausführung von Drehungen, Verschiebungen, oder allgemeinen invertierbare Abbildungen, oder auch für die in der Kryptographie wichtigen Restklassen. Wir geben solchen Strukturen deshalb einen Namen:

- Eine Gruppe $(G, \circ)$ ist eine (nichtleere) Menge G zusammen mit einer Verknüpfung $\circ : G \times G \to G$, für die gilt:

  (G1) *Assoziativität:*

$$(a \circ b) \circ c = a \circ (b \circ c) \quad \text{für alle } a,b,c \in G.$$

  (G2) *Existenz des neutralen Elements:*

$$\exists e \in G : \quad a \circ e = e \circ a = a \quad \text{für alle } a \in G.$$

  (G3) *Existenz inverser Elemente:*

$$\forall a \in G : \exists a' \in G : \quad a \circ a' = a' \circ a = e \quad \text{für alle } a \in G.$$

(G1)–(G3) heißen auch *Gruppenaxiome*. Ist die Verknüpfung die Addition zweier Zahlen (oder damit verwandt), schreiben wir statt $\circ$ auch $+$, für das neutrale Element auch 0, für das inverse Element $-a$ und statt $b+(-a)$ schreiben wir $b-a$. Ist die Verknüpfung die Multiplikation zweier Zahlen (oder damit verwandt), schreiben wir $*$, nennen das neutrale Element 1 und das inverse Element a^{-1}.

Wir definieren außerdem:

- Eine Gruppe $(G,\circ)$ heißt kommutativ (auch: *abelsch*), wenn zusätzlich zu (G1)–(G3) gilt:

  (G4) *Kommutativität:*

  $$a\circ b = b\circ a \quad \text{für alle } a,b\in G.$$

Ist die Verknüpfung die Multiplikation zweier Zahlen (oder damit verwandt) und kommutativ, so können wir statt $a^{-1}b$ oder ba^{-1} auch $\frac{b}{a}$ schreiben.

$(\mathbb{Z},+)$, $(\mathbb{Q},+)$, $(\mathbb{R},+)$, $(\mathbb{Q}\setminus\{0\},*)$, $(\mathbb{R}\setminus\{0\},*)$ sind kommutative Gruppen. Für eine (nichtleere) Menge X definieren wir

$$\mathscr{F} := \{f:\ X\to X:\ f \text{ ist bijektiv.}\}$$

und bezeichnen mit $\circ$ wieder die Hintereinanderausführung von Funktionen. Dann ist $(\mathscr{F},\circ)$ eine Gruppe mit inversem Element f^{-1} (die Umkehrfunktion zu f) und neutralem Element

$$\mathrm{id}:\ X\to X,\quad x\mapsto x \quad \text{für alle } x\in X.$$

$(\mathscr{F},\circ)$ ist aber im Allgemeinen nicht kommutativ, z.B. sind auf $X=\mathbb{R}$ die Funktionen $f:\ x\mapsto x+1$ und $g:\ x\mapsto 2x$ offensichtlich bijektiv, aber

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

Nützlich ist auch der folgende abgeschwächte Begriff:

- Eine (nichtleere) Menge G mit Verknüpfung $\circ$ heißt Halbgruppe, falls (G1) erfüllt ist (also die Verknüpfung assoziativ ist).
- $(G,\circ)$ heißt kommutative Halbgruppe, falls (G1) und (G4) erfüllt sind (also die Verknüpfung assoziativ und kommutativ ist).

$(\mathbb{N},+)$, $(\mathbb{N},*)$, $(\mathbb{Q},*)$, $(\mathbb{R},*)$ sind kommutative Halbgruppen. Die Menge aller Abbildungen eine Menge in sich ist eine Halbgruppe, die im Allgemeinen nicht kommutativ ist.

Wir führen auch Begriffe für Strukturen ein, in denen (wie in $\mathbb{Q}$ oder $\mathbb{R}$) zwei Verknüpfungen definiert sind:

- Ein Ring $(R,+,*)$ ist eine (nichtleere) Menge R zusammen mit zwei Verknüpfungen $+:\ R\times R\to R$ und $*:\ R\times R\to R$, für die gilt:

  (R1) $(R,+)$ ist eine kommutative Gruppe.

  (R2) $(R,*)$ ist eine Halbgruppe.

  (R3) Es gelten die *Distributivgesetze*, d.h. für alle $a,b,c\in R$ gilt:

$$a*(b+c)=a*b+a*c$$
$$(a+b)*c=a*c+b*c.$$

- Ein Ring $(R,+,*)$ heißt kommutativ, wenn die Verknüpfung „$*$´´ zusätzlich kommutativ ist.

- Ein Ring $(R,+,*)$ heißt Ring mit Eins, wenn es ein neutrales Element bezüglich der Verknüpfung „$*$´´ gibt.

$(\mathbb{Z},+,*)$, $(\mathbb{Q},+,*)$ und $(\mathbb{R},+,*)$ sind kommutative Ringe mit Eins.

Für eine (nichtleere) Menge X und einen Ring $(R,+,*)$ betrachten wir den Funktionenraum

$$\mathscr{F}:=\{f:\ X\to R\}$$

und definieren darauf die *komponentenweise* Addition und Multiplikation

$$+:\mathscr{F}\times\mathscr{F}\to\mathscr{F},\quad (f+g)(x):=f(x)+g(x)\quad \forall x\in X,$$
$$*:\mathscr{F}\times\mathscr{F}\to\mathscr{F},\quad (f*g)(x):=f(x)*g(x)\quad \forall x\in X.$$

Damit ist $(\mathscr{F},+,*)$ ein Ring. Ist $(R,+,*)$ kommutativ, dann ist es auch $(\mathscr{F},+,*)$. Ist $(R,+,*)$ ein Ring mit Eins, so auch $(\mathscr{F},+,*)$.

$\mathbb{Q}$ und $\mathbb{R}$ sind (nach Weglassen der Null) auch bezüglich der Multiplikation nicht nur Halbgruppen sondern kommutative Gruppen. Auch für diese Eigenschaft führen wir einen Begriff ein:

- Ein Körper $(K,+,*)$ ist eine (nichtleere) Menge K zusammen mit zwei Verknüpfungen $+:\ K\times K\to K$ und $*:\ K\times K\to K$, für die gilt:

  (K1) $(K,+)$ ist eine kommutative Gruppe. Das neutrale Element bezüglich „$+$´´ bezeichnen wir mit 0.

  (K2) $(K\setminus\{0\},*)$ ist eine kommutative Gruppe.

  (K3) Es gelten die *Distributivgesetze*, d.h. für alle $a,b,c\in K$ gilt:

$$a*(b+c)=a*b+a*c$$
$$(a+b)*c=a*c+b*c.$$

$(\mathbb{Q}, +, *)$ und $(\mathbb{R}, +, *)$ sind Körper. Später werden wir auch Körper mit nur endlich vielen Elementen kennenlernen, die in Kryptographie verwendet werden.

Ein Körper K ist also eine mathematische Struktur, in der die von den reellen Zahlen bekannten Rechenregeln gelten. Für $x, y, z \in K$ gilt:

- Kommutativgesetz: $\begin{array}{l} x+y = y+x \\ x*y = y*x \end{array}$
- Assoziativgesetz: $\begin{array}{l} (x+y)+z = x+(y+z), \\ (x*y)*z = x*(y*z). \end{array}$
- Distributivgesetz: $x*(y+z) = x*y+x*z.$
- Neutrale Elemente: $\begin{array}{l} x+0 = x, \\ 1*x = x. \end{array}$
- Inverse Elemente: $\begin{array}{l} x+(-x) = 0, \\ x*x^{-1} = 1 \text{ für } x \neq 0. \end{array}$

Für die Multiplikation lassen wir das Symbol „$*$" auch weg und schreiben z.B. xy statt $x*y$. Statt xy^{-1} schreiben wir auch $\frac{x}{y}$.

Kapitel 2

Diskrete Mathematik

Die *diskrete Mathematik* beschäftigt sich mit endlichen Mengen und mit solchen unendlichen Mengen, in denen (anschaulich gesprochen) die Elemente nicht beliebig nah zusammenliegen, wie z.B. $\mathbb{N}$ oder $\mathbb{Z}$.

2.1 Natürliche Zahlen und vollständige Induktion

2.1.1 Das Prinzip der vollständigen Induktion

Wir erinnern daran, dass die Menge der natürlichen Zahlen dadurch gebildet wurde, dass beginnend mit 1 zu jeder natürlichen Zahl auch die um 1 erhöhte dazugenommen wurde:

$$\mathbb{N} = \{1, 2, 3, \ldots\}.$$

Hieraus ergibt sich eine sehr mächtige Beweismethode, um eine Aussage $A(n)$ für alle natürlichen Zahlen $n \in \mathbb{N}$ zu beweisen:

Satz 2.1 (Vollständige Induktion)
*Gilt eine Aussage $A(n)$ für $n = 1$ (*Induktionsanfang*) und außerdem die Implikation (*Induktionsanfang*)*

$$A(n) \quad \Longrightarrow \quad A(n+1),$$

so gilt $A(n)$ für alle $n \in \mathbb{N}$.

Beweis: Die Menge der Zahlen, für die $A(n)$ gilt, enthält 1 und zu jeder enthaltenen Zahl auch die um 1 erhöhte. Dies ist nach Konstruktion ganz $\mathbb{N}$. □

Wir betonen, dass im Induktionsschritt $A(n) \Longrightarrow A(n+1)$ bereits vorausgesetzt wird, dass $A(n)$ für ein $n \in \mathbb{N}$ gilt und *unter dieser Voraussetzung* dann gezeigt wird, dass auch $A(n+1)$ gilt.

Anschaulich können wir uns das wie bei einer Dominobahn vorstellen. $A(n)$ bezeichne dabei, dass der n-te Dominostein umfällt. Wenn sichergestellt ist, dass der erste Stein fällt (also $A(1)$ gilt) und dass jeder umfallende Stein auch den jeweils nächsten umwirft (also $A(n) \Longrightarrow A(n+1)$ gilt), dann ist sicher, dass jeder Stein fallen wird.

Durch Wahl des Induktionsanfangs $n = 0$ können wir mit vollständiger Induktion natürlich genauso auch Aussagen für alle $n \in \mathbb{N}_0$ beweisen oder mit Induktionsanfang $n = k \in \mathbb{Z}$ Aussagen für alle ganzen Zahlen $\geq k$ beweisen.

Beispiel 2.2 (Gaußsche Summenformel)
Für alle $n \in \mathbb{N}$ gilt die Summenformel

$$\sum_{k=1}^{n} k = 1+2+\ldots+n = \frac{n(n+1)}{2}.$$

Beweis: *Induktionsanfang:* Für $n = 1$ ist

$$\sum_{k=1}^{n} k = \sum_{k=1}^{1} k = 1 = \frac{1(1+1)}{2} = \frac{n(n+1)}{2}.$$

Induktionsschluss: Angenommen, die Behauptung gilt für ein $n \in \mathbb{N}$, also

$$\sum_{k=1}^{n} k = 1+2+\ldots+n = \frac{n(n+1)}{2}. \tag{2.1}$$

Wir müssen zeigen, dass sie dann auch für $n+1$ gilt, also dass

$$\sum_{k=1}^{n+1} k = 1+2+\ldots+n+n+1 = \frac{(n+1)(n+2)}{2}. \tag{2.2}$$

Wir betonen noch einmal, dass wir zum Beweis von (2.2) die *Induktionsannahme* (2.1) benutzen dürfen! Damit ist der Beweis einfach:

$$\begin{aligned}\sum_{k=1}^{n+1} k &= 1+2+\ldots+n+(n+1) = (1+2+\ldots+n)+(n+1) \\ &\overset{(2.1)}{=} \frac{n(n+1)}{2}+(n+1) = \frac{n(n+1)+2(n+1)}{2} = \frac{(n+1)(n+2)}{2}.\end{aligned}$$

Die Behauptung ist damit durch vollständige Induktion bewiesen. □

Bemerkung 2.3
Die vollständige Induktion ermöglicht es uns, geratene oder heuristisch gefundene Vermutungen mathematisch rigoros zu beweisen. Sie zeigt uns aber meist nicht, wie wir die richtige Formel finden. Eine anschaulichere Herleitung der obigen Summenformel ergibt sich, wenn wir die komplette Summe noch einmal in umgedrehter Reihenfolge dazuzuaddieren und in folgendem Schema übereinanderstehende Summanden addieren:

$$\begin{array}{cccccccc} & 1+ & 2+ & \ldots+ & (n\text{-}1)+ & n & \\ + & n+ & (n\text{-}1)+ & \ldots+ & 2+ & 1 & \\ \hline = & (n+1)+ & (n+1)+ & \ldots+ & (n+1)+ & (n+1) & = n(n+1). \end{array}$$

Zweimal die Summe $1+2+\ldots+n$ *ergibt also* $n(n+1)$ *und damit ist*

$$1+2+\ldots+n = \frac{n(n+1)}{2}.$$

2.1.2 Kombinatorik und binomischer Lehrsatz

Wir betrachten ein Experiment, in dem $n \in \mathbb{N}$ verschiedene Objekte hintereinander ohne Zurücklegen gezogen werden. Beim ersten Zug gibt es n Möglichkeiten, beim nächsten Zug nur noch $(n-1)$ usw. Insgesamt gibt es daher

$$n! := n*(n-1)*\ldots*1 = \prod_{j=1}^{n} j$$

verschiedene Reihenfolgen, in denen wir die n Objekte ziehen können. $n!$ heißt Fakultät von n. Es vereinfacht die späteren Formeln, wenn wir außerdem $0! = 1$ definieren. Dies entspricht der üblichen Konvention, den Wert des *leeren Produktes* $\prod_{j=1}^{0} j$ auf 1 zu setzen.

Ziehen wir lediglich k der n Objekte (mit $k, n \in \mathbb{N}$, $k \leq n$), dann gibt es dafür dafür

$$\underbrace{n*(n-1)*\ldots*(n-k+1)}_{k \text{ Faktoren}} = \frac{n!}{(n-k)!}$$

Möglichkeiten. Interessiert uns nur, welche k Objekte gezogen wurde (aber nicht die Reihenfolge, in der diese k gezogen wurden), dann müssen wir berücksichtigen, dass wir dieselben k Objekte auf $k!$ verschiedenen Reihenfolgen ziehen konnten. Je $k!$ der $\frac{n!}{(n-k)!}$ Möglichkeiten liefern also dieselben k Objekte (nur in verschiedener Reihenfolge) und es gibt also

$$\binom{n}{k} := \frac{n*(n-1)*\ldots*(n-k+1)}{k*(k-1)*\ldots*1} = \frac{n!}{(n-k)!k!}$$

Möglichkeiten, k Objekte aus n Objekten auszuwählen. $\binom{n}{k}$ heißt n über k oder auch (wegen Satz 2.6) Binomialkoeffizient. Entsprechend der Konvention $0! = 1$ ist für alle $n \in \mathbb{N}$

$$\binom{0}{0} = \binom{n}{0} = \binom{n}{n} = 1.$$

Beispiel 2.4
Die Anzahl der dreielementigen Teilmengen einer fünfelementigen Menge ist

$$\binom{5}{3} = \frac{5*4*3}{3*2*1} = 10.$$

Für $\{1,2,3,4,5\}$ *sind dies gerade*

$$\{1,2,3\},\quad \{1,2,4\},\quad \{1,3,4\},\quad \{2,3,4\},$$
$$\{1,2,5\},\quad \{1,3,5\},\quad \{1,4,5\},\quad \{2,3,5\},\quad \{2,4,5\},\quad \{3,4,5\}$$

In Beispiel 2.4 haben wir alle dreielementigen Teilmengen von $\{1,2,3,4,5\}$ aufgelistet, indem wir zunächst alle dreielementigen Teilmengen hingeschrieben haben, die das letzte Element 5 nicht erhalten (also alle dreielementigen Teilmengen von $\{1,\ldots,4\}$) und danach alle Teilmengen, die die 5 enthalten (also die 5 kombiniert mit allen zweielementigen Teilmengen von $\{1,2,3,4\}$). Mit diesem Argument folgt auch anschaulich, dass die Anzahl der k-elementigen Teilmengen einer $n+1$-elementigen Menge gerade die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge (Anzahl der Teilmengen, die das letze Element nicht enthalten) zuzüglich der Anzahl der $k-1$-elementigen Menge einer n-elementigen Menge ist (Anzahl der Teilmengen, die das letzte Element enthalten). Wir formulieren und beweisen diese nützliche Summenformel noch mathematisch:

Satz 2.5
Für $k,n \in \mathbb{N}$, $n \geq k$ *gilt*

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}.$$

Beweis: Es ist

$$\begin{aligned}\binom{n}{k-1} + \binom{n}{k} &= \frac{n!}{(n-(k-1))!(k-1)!} + \frac{n!}{(n-k)!\,k!} \\ &= \frac{n!k}{(n-k+1)!k!} + \frac{n!(n-k+1)}{(n-k+1)!\,k!} = \frac{n!(k+n-k+1)}{(n+1-k)!k!} \\ &= \frac{n!(n+1)}{(n+1-k)!k!} = \frac{(n+1)!}{(n+1-k)!k!} = \binom{n+1}{k}.\end{aligned}$$

Diese Summenformel lässt sich auch über das Pascalsche Dreieck visualisieren. Wir ordnen die Binomialkoeffizienten in Pyramidenform an:

$$\begin{array}{ccccccccccc}
 & & & & \binom{0}{0} & & & & & & \\
 & & & \binom{1}{0} & & \binom{1}{1} & & & & & \\
 & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & & \\
 & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & & \\
 \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & & \\
 & \binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5}
\end{array}$$

Aufgrund der Summenformel in Satz 2.5 entspricht jeder Eintrag im Schema der Summe der unmittelbar links und rechts darüber stehenden Einträge und die Randwerte ohne solche beidseitigen oberen Nachbarn sind (wegen $\binom{n}{0} = 1 = \binom{n}{n}$ immer eins. Wir erhalten also für die Werte der eingetragenen Binomialkoeffizienten:

$$\begin{array}{ccccccccccc}
 & & & & 1 & & & & & & \\
 & & & 1 & & 1 & & & & & \\
 & & 1 & & 2 & & 1 & & & & \\
 & 1 & & 3 & & 3 & & 1 & & & \\
 1 & & 4 & & 6 & & 4 & & 1 & & \\
 1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

Diese Zahlen treten in allgemeinen binomischen Formeln auf. Für $a, b \in \mathbb{R}$ ist

$$\begin{aligned}
(a+b)^0 &= 1 \\
(a+b)^1 &= 1a + 1b \\
(a+b)^2 &= 1a^2 + 2ab + 1b^2 \\
(a+b)^3 &= 1a^3 + 3a^2b + 3ab^2 + 1b^3 \\
(a+b)^4 &= 1a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + 1b^4 \\
(a+b)^5 &= 1a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + 1b^5
\end{aligned}$$

Dabei ergibt sich jede Zeile durch Multiplikation der obendrüber stehenden Zeile mit $a+b$, wobei sich beim Ausmultiplizieren ein Eintrag durch die Summe des links darüber stehenden Eintrags (multiplizert mit b) und des rechts darüber stehenden Eintrags (multipliziert mit a) ergibt. Die Koeffizienten der allgemeinen binomischen Formeln sind daher gerade die Binomialkoeffizienten.

Satz 2.6
Für $a, b \in \mathbb{R}$ und $n \in \mathbb{N}_0$ gilt

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Beweis: Wir beweisen die Aussage durch vollständige Induktion über n.

Induktionsanfang: Für $n = 0$ ist

$$\sum_{k=0}^{0} \binom{0}{k} a^{0-k} b^k = \binom{0}{0} a^0 b^0 = 1 = (a+b)^0.$$

Induktionsschritt: Angenommen die Behauptung gilt für ein $n \in \mathbb{N}$, also

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Wir müssen zeigen, dass sie dann auch für $n+1$ gilt, d.h.

$$(a+b)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k} b^k. \tag{2.3}$$

Dies erhalten wir aber unter Anwendung der Summenformel in Satz 2.5 aus

$$\begin{aligned}
(a+b)^{n+1} &= (a+b)(a+b)^n = (a+b) \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n+1-k} b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^{k+1} \\
&= \sum_{k=0}^{n} \binom{n}{k} a^{n+1-k} b^k + \sum_{k=1}^{n+1} \binom{n}{k-1} a^{n-(k-1)} b^k \\
&= \binom{n}{0} a^{n+1-0} b^0 + \sum_{k=1}^{n} \left(\binom{n}{k} + \binom{n}{k-1} \right) a^{n+1-k} b^k + \binom{n}{n} a^0 b^{n+1} \\
&= \binom{n+1}{0} a^{n+1} b^0 + \sum_{k=1}^{n} \binom{n+1}{k} a^{n+1-k} b^k + \binom{n+1}{n+1} a^0 b^{n+1} \\
&= \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k} b^k,
\end{aligned}$$

womit die Behauptung durch vollständige Induktion für alle $n \in \mathbb{N}_0$ bewiesen ist. □

2.1.3 Minima, Maxima und Prinzip des kleinsten Verbrechers

Definition 2.7 (Minimum und Maximum)
Sei $M \subseteq \mathbb{R}$.

- *Existiert ein Element $x_- \in M$ mit $x_- \leq x$ für alle $x \in M$, dann nennen wir x_- das Minimum (auch: kleinstes Element) von M und schreiben $x_- = \min X$.*
- *Existiert ein Element $x_+ \in M$ mit $x_+ \geq x$ für alle $x \in M$, dann nennen wir x_+ das Maximum (auch: größtes Element) von M und schreiben $x_+ = \max X$.*

Wenn eine Menge ein Maximum besitzt, dann ist es offenbar eindeutig. Gleiches gilt für das Minimum.

Mengen besitzen aber nicht immer ein größtes oder kleinstes Element. Z.B. enthält die Menge $M := \mathbb{R}$ beliebig große und beliebig kleine Elemente und deshalb weder Maximum noch Minimum. Aber auch die nach oben und unten beschränkte Menge

$$M := \{x \in \mathbb{R} : \ 0 < x < 1\}$$

enthält weder ein Maximum noch ein Minimum. Eine Zahl größer als Null kann nicht das Minimum sein, denn sie wird von Mengenelemente unterschritten. Eine Zahl kleiner gleich Null liegt nicht in der Menge und ist deshalb nicht das Minimum. Die Menge M besitzt daher kein Minimum und analog folgt, dass sie auch kein Maximum besitzt.

In der diskreten Mathematik ist die Existenz von Minima und Maxima meist durch den folgenden Satz und die anschließende Folgerung sichergestellt.

Satz 2.8
(a) Jede endliche Teilmenge $\emptyset \neq M \subseteq \mathbb{R}$ besitzt ein kleinstes Element $\min M$.

(b) Jede Teilmenge $\emptyset \neq M \subseteq \mathbb{N}$ besitzt ein kleinstes Element $\min M$.

Beweis: (a) Zum Beweis werden wir die Aussage

$$A(n): \text{ Jedes } M \subseteq \mathbb{R} \text{ mit } |M| = n \text{ Elementen besitzt ein kleinstes Element.}$$

für alle $n \in \mathbb{N}$ beweisen. Wir verwenden dazu die Methode der vollständigen Induktion.

$A(1)$ gilt, da in einer einelementigen Menge $M = \{x\}$ gilt $\min M = x$.

Zum Beweis von $A(n) \Longrightarrow A(n+1)$ nehmen wir an, dass $A(n)$ gelte und müssen zeigen, dass dann auch jede $n+1$-Elementige Menge ein Minimum besitzt. Sei also $M \subseteq \mathbb{R}$ mit $|M| = n+1$. Wir wählen ein $x \in M$ und definieren $M' := M \setminus \{x\}$. Dann hat M' nur noch n Elemente und besitzt daher (wegen der Induktionsannahme $A(n)$) ein kleinstes Element $\min M'$. Falls $x \geq \min M'$, dann ist $\min M'$ auch das kleinste Element in M, ansonsten ist x das kleinste Element in M. M besitzt also auch ein Minimum.

(b) Sei $\emptyset \neq M \subseteq \mathbb{N}$. Angenommen, M besitzt kein kleinstes Element. Wir werden folgende Aussage für alle $n \in \mathbb{N}$ per vollständiger Induktion zeigen:

$$A(n): \ M \cap \{1,\ldots,n\} = \emptyset.$$

Damit ist dann gezeigt, dass $M \cap \mathbb{N} = \emptyset$ und damit $M = \emptyset$ ist, was $M \neq \emptyset$ widerspricht.

$A(1)$ gilt, da im Falle $1 \in M$ die 1 auch das kleinste Element von M wäre. Angenommen $A(n)$ gilt, also M enthält keine der Zahlen 1,...,n. Dann kann M aber auch nicht die Zahl $n+1$ enthalten, da die sonst das kleinste Element von M wäre. Also gilt auch $A(n+1)$. □

Folgerung 2.9

(a) Jede endliche Teilmenge $M \subseteq \mathbb{R}$ besitzt ein größtes Element $\max M$.

(b) Jede nach unten beschränkte Teilmenge $M \subseteq \mathbb{Z}$ besitzt ein kleinstes Element $\min M$.

(c) Jede nach oben beschränkte Teilmenge $M \subseteq \mathbb{Z}$ besitzt ein größtes Element $\max M$.

Beweis: (a) Sei $M \subseteq \mathbb{R}$ mit $|M| < \infty$. Wir definieren

$$-M := \{-x: \ x \in M\}.$$

Dann hat auch $-M$ nur endlich viele Elemente und deshalb nach Satz 2.8(a) ein kleinstes Element. Die Negation des kleinsten Element von $-M$ ist dann aber das größte Element von M.

(b) Da M nach unten beschränkt ist, existiert ein $a \in \mathbb{N}$, so dass $x \geq -a$ für alle $x \in M$. Für die Menge

$$M + a = \{x + a: \ x \in M\}$$

gilt deshalb, dass $M + a \subseteq \mathbb{N}$ und nach Satz 2.8(b) existiert also ein kleinstes Element von $M + a$. Dieses Element minus a ist dann aber kleinstes Element von M.

(c) folgt aus (b) mit dem gleichen Argument wie in (a). □

Die Existenz eines Minimums lässt sich im folgenden *Prinzip des kleinsten Verbrechers* (auch: *Prinzip des kleinsten Elements*) ausnutzen.

Folgerung 2.10 (Prinzip des kleinsten Verbrechers)
Gilt eine Aussage nicht für alle $n \in \mathbb{N}$, dann existiert ein kleinstes n, für das die Aussage nicht gilt.

Beweis: Die Menge aller $n \in \mathbb{N}$, für die die Aussage nicht gilt, muss gemäß Satz 2.8(b) entweder leer sein oder ein kleinstes Element besitzen. □

2.2 Teilbarkeit ganzer Zahlen

2.2.1 Teilbarkeit und Division mit Rest

Definition 2.11 (Teilbarkeit)
Seien $a, b \in \mathbb{Z}$ und $b \neq 0$. a heißt durch b teilbar, *wenn*

$$\exists q \in \mathbb{Z} : \ a = bq.$$

Wir sagen auch „b teilt a“und schreiben kurz „$b|a$“.

Beachte, dass nach dieser Definition Teiler sowohl positiv als auch negativ sein können und jede Zahl ein Teiler der Null ist (also $a|0$ für alle $a \in \mathbb{Z}$), aber die Null keine Zahl teilt.

Satz 2.12
Seien $a, b, c \in \mathbb{Z} \setminus \{0\}$

(a) Aus $a|b$ folgt $|a| \leq |b|$.

(b) Es gilt immer

$$1|a, \quad -1|a, \quad a|a \text{ und } \quad -a|a.$$

Wir nennen $\{-a, -1, 1, a\}$ die trivialen Teiler *von a.*

(c) Aus $a|b$ und $b|c$ folgt $a|c$.

(d) Aus $a|b$ und $b|a$ folgt $a \in \{-b, b\}$. Für $a, b \in \mathbb{N}$ folgt aus $a|b$ und $b|a$, dass $a = b$.

(e) Aus $a|b$ und $a|c$ folgt $a|(b+c)$.

Beweis: Leichtes Nachrechnen. □

Ist b nicht durch a teilbar, dann ist der Quotient $\frac{b}{a}$ nur noch eine rationale aber keine ganze Zahl. In einigen Anwendungen ist es sinnvoller stattdessen mit der Division mit Rest zu arbeiten. Um dazu $\frac{b}{a}$ nach oben oder unten auf die nächstgelegene ganze Zahl runden zu können, führen wir ein:

Definition 2.13 (Floor- und Ceiling-Funktion)
Für eine reelle Zahl $x \in \mathbb{R}$ bezeichnen wir die nächstkleinere und nächstgrößere ganze Zahlen mit

- $\lfloor x \rfloor := \max\{k \in \mathbb{Z} : \ k \leq x\}$ (Floor-Funktion)

- $\lceil x \rceil := \min\{k \in \mathbb{Z} : \ k \geq x\}$ (Ceiling-Funktion)

*Statt $\lfloor x \rfloor$ schreibt man auch $\mathrm{floor}(x)$ und statt $\lceil x \rceil$ schreibt man auch $\mathrm{ceil}(x)$. In der mathematischen Literatur schreibt man statt $\lfloor x \rfloor$ auch $[x]$ (*Gaußklammer*).*

Für positive Zahlen entsprechen die Ceiling- und Floor-Funktionen gerade dem Auf- und Abrunden. Für negative Zahlen liefert floor ebenfalls die nächstkleinere (also noch negativere) und ceil die nächstgrößere (also weniger negative), z.B.

$$\lfloor 3.5 \rfloor = 3, \quad \lceil 3.5 \rceil = 4, \quad \lfloor -3.5 \rfloor = -4, \quad \lceil -3.5 \rceil = -3.$$

Die Rundung zur Null hin bezeichnet man auch als *Fix-Funktion*:

- $\mathrm{fix}(x) := \begin{cases} \mathrm{floor}(x) & \text{für } x \geq 0 \\ \mathrm{ceil}(x) & \text{für } x < 0 \end{cases}$

Damit gilt dann $\mathrm{fix}(3.5) = 3$, $\mathrm{fix}(-3.5) = -3$.

Wir fassen einige nützliche Eigenschaften der Floor- und Ceiling-Funktionen zusammen:

Satz 2.14
Für $x \in \mathbb{R}$ gilt

(a) $\lfloor x \rfloor \leq x \leq \lceil x \rceil$.

(b) $0 \leq x - \lfloor x \rfloor < 1$

(c) $0 \leq \lceil x \rceil - x < 1$

Beweis: (a) folgt sofort aus den in Definition 2.13 verwendeten Mengen. Damit sind auch die unteren Abschätzung in (b) und (c) gezeigt. Zum Beweis der oberen Abschätzung in (b) nehmen wir an, dass sie nicht gelten würde, also $x - \lfloor x \rfloor \geq 1$ wäre. Dann wäre aber $k := \lfloor x \rfloor + 1 \in \mathbb{Z}$ eine ganze Zahl, die größer wäre als $\lfloor x \rfloor$ und trotzdem immer noch $x \geq k$ erfüllt. Das würde der Maximalität von $\lfloor x \rfloor$ widersprechen. Die obere Abschätzung in (b) muss also doch gelten. Analog folgt die obere Abschätzung in (c) aus der Minimalität von $\lceil x \rceil$. □

Definition und Satz 2.15 (Division mit Rest)
Seien $a \in \mathbb{Z}$ und $b \in \mathbb{N}$. Wir definieren

$$q := \left\lfloor \frac{a}{b} \right\rfloor \quad \textit{und} \quad r := a - b * q.$$

Dann sind $q, r \in \mathbb{Z}$,

$$a = bq + r, \quad \text{und} \quad 0 \leq r \leq b - 1.$$

q heißt Ganzzahlquotient und r heißt Rest der Division. Wir schreiben

$$a : b = q \text{ Rest } r$$

oder auch

$$a \text{ div } b = q, \quad a \text{ mod } b = r.$$

Beweis: Setze $q := \left\lfloor \frac{a}{b} \right\rfloor$ und $r := a - b * q$. Damit ist

$$b * q + r = b * q + (a - b * q) = a.$$

Da die Floor-Funktion nur ganzzahlige Ergebnisse hat sind $q, r \in \mathbb{Z}$. Außerdem ist

$$\frac{r}{b} = \frac{a}{b} - q = \frac{a}{b} - \left\lfloor \frac{a}{b} \right\rfloor$$

und damit wegen Satz 2.14(b)

$$0 \leq \frac{r}{b} < 1,$$

und damit $0 \leq r < b$. □

Bemerkung 2.16
Wir haben in Definition und Satz 2.15 nur positive Divisoren $b \in \mathbb{N}$ betrachtet, aber der Dividend $a \in \mathbb{Z}$ darf positiv oder negativ sein. Der enstehende Rest a mod b ist nach unserer Definition immer nicht-negativ, z.B. ist

$$-9 : 4 = -3 \text{ Rest } 3.$$

Diese Definition heißt auch die mathematische Variante *der Division mit Rest.*

Verwendet man bei der Definition der Division mit Rest statt der Floor-Funktion die Fix-Funktion, so ergibt sich für positive $a, b \in \mathbb{N}$ der gleiche Wert bei der Division mit Rest, aber

$$-9 : 4 = -2 \text{ Rest } -1.$$

Diese Definition heißt (aufgrund der Symmetrie zu $9 : 4 = 2$ Rest 1) auch die symmetrische Variante *der Division mit Rest.*

In Programmiersprachen wird das nicht einheitlich gehandhabt. Beispielsweise verwendet Python die mathematische Variante und der Befehl

`print(-9 % 4)` *ergibt* `3`,

aber C verwendet die symmetrische Variante und

`printf("%d",-9 % 4);` *ergibt* `-1`.

In Matlab heißt die mathematische Variante `mod` *und die symmetrische Variante* `rem`.

Mit den gleichen Formeln lässt sich auch für negative Divisoren eine Division mit Rest definieren. Dabei liefert die mathematische Variante (mit Floor-Funktion) negative Reste und die symmetrische Variante (mit Fix-Funktion) positive Reste. Wir werden in dieser Vorlesung die Division mit Rest immer nur in der mathematischen Variante und nur für positive Divisoren verwenden. Der Rest ist dadurch immer nicht-negativ.

Wir haben die Division mit Rest über explizite Formeln definiert. Die darin vorkommende Floor-Funktion macht Rechnungen mit diesen Formeln aber schwierig. Oft ist es einfacher, stattdessen die folgende äquivalente definierende Eigenschaft zu verwenden:

Satz 2.17
Seien $a \in \mathbb{Z}$ und $b \in \mathbb{N}$. Die beiden Zahlen

$$q := a \text{ div } b = \left\lfloor \frac{a}{b} \right\rfloor$$
$$r := a \text{ mod } b = a - b * \left\lfloor \frac{a}{b} \right\rfloor$$

sind das einzige Zahlenpaar $(q,r) \in \mathbb{Z}^2$ mit der Eigenschaft, dass

$$a = bq + r, \quad \textit{und} \quad 0 \le r \le b-1. \tag{2.4}$$

Beweis: Das $q := a$ div b und $r := a$ mod b die beiden Eigenschaften in (2.4) erfüllen haben wir in Definition und Satz 2.15 gezeigt. Angenommen, es gäbe noch zwei weitere Zahlen $\tilde{q}, \tilde{r} \in \mathbb{Z}$, mit $\tilde{q} \neq q$ oder $\tilde{r} \neq r$, so dass diese auch (2.4) erfüllen, also

$$a = b\tilde{q} + \tilde{r}, \quad \text{und} \quad 0 \le \tilde{r} \le b-1.$$

Dann wäre

$$b(\tilde{q} - q) + \tilde{r} - r = a - a = 0. \tag{2.5}$$

Wir werden zeigen, dass dies in allen möglichen Fällen $\tilde{r} > r$, $\tilde{r} = r$ und $\tilde{r} < r$ zum Widerspruch führt:

(a) $\tilde{r} > r$: Dann ist $b > \tilde{r} - r > 0$. Wir unterscheiden die Fälle $\tilde{q} \ge q$ und $\tilde{q} < q$

(i) $\tilde{q} \geq q$: Dann wäre $b(\tilde{q}-q)+\tilde{r}-r \geq \tilde{r}-r > 0$. Dies widerspricht (2.5).

(ii) $\tilde{q} < q$: Dann wäre $\tilde{q}-q \leq -1$ und damit

$$b(\tilde{q}-q)+\tilde{r}-r \geq \tilde{r}-r \leq -b+\tilde{r}-r < b-b = 0.$$

Dies widerspricht (2.5).

(b) $\tilde{r} = r$: Dann wäre $b(\tilde{q}-q) =$ und damit $\tilde{q} = q$. Dies widerspricht unserer Annahme, dass $\tilde{q} \neq q$ oder $\tilde{r} \neq r$.

(c) $\tilde{r} < r$: Dies führt genau wie in (a) zu einem Widerspruch.

$q = a \text{ div } b$ und $r = a \text{ mod } b$ ist also das einzige Paar ganzer Zahlen, das (2.4) erfüllt. □

Der folgende Satz zeigt, dass bei der Addition und Multiplikation von Zahlen die Restbestimmung sowohl vor als auch nachher durchgeführt werden kann.

Satz 2.18
Seien $a_1, a_2 \in \mathbb{Z}$, $b \in \mathbb{N}$. Dann gilt

$$\begin{aligned}(a_1 \text{ mod } b) \text{ mod } b &= a_1 \text{ mod } b,\\ a_1 + a_2 \text{ mod } b &= ((a_1 \text{ mod } b) + (a_2 \text{ mod } b)) \text{ mod } b,\\ a_1 a_2 \text{ mod } b &= ((a_1 \text{ mod } b) * (a_2 \text{ mod } b)) \text{ mod } b.\end{aligned}$$

Beweis: Es seien

$$q_1 := a_1 \text{ div } b, \qquad r_1 := a_1 \text{ mod } b,$$

Dann ist $r_1 = 0 * r_1 + r_1$ mit $0 \in \mathbb{Z}$, $r_1 \in \mathbb{Z}$ und $0 \leq r_1 \leq b-1$. Da diese Eigenschaften nach Satz 2.17 nur das Zahlenpaar $r_1 \text{ div } b$ und $r_1 \text{ mod } b$ besitzt, folgt also dass

$$r_1 \text{ div } b = 0 \quad \text{und} \quad r_1 \text{ mod } b = r_1$$

und damit ist die erste Aussage bewiesen.

Um die zweite Aussage zu zeigen definieren wir noch

$$\begin{aligned} q_2 &:= a_2 \text{ div } b, & r_2 &:= a_2 \text{ mod } b,\\ q_+ &:= r_1 + r_2 \text{ div } b, & r_+ &:= r_1 + r_2 \text{ mod } b.\end{aligned}$$

und argumentieren wieder mitttels Satz 2.17. Es ist

$$\begin{aligned} a_1 + a_2 &= bq_1 + r_1 + bq_2 + r_2 = b(q_1+q_2) + r_1 + r_2\\ &= b(q_1+q_2) + bq_+ + r_+ = b(q_1+q_2+q_+) + r_+\end{aligned}$$

und da $q_1+q_2+q_+ \in \mathbb{Z}$, $r_+ \in \mathbb{Z}$ und $0 \leq r_+ \leq b-1$ folgt aus Satz 2.17

$$\begin{aligned} a_1+a_2 \text{ div } b &= q_1+q_2+r_+, \\ a_1+a_2 \text{ mod } b = r_+ &= r_1+r_2 \text{ mod } b = ((a_1 \text{ mod } b)+(a_2 \text{ mod } b)) \text{ mod } b. \end{aligned}$$

Die dritte Behauptung zeigen wir analog der zweiten. Wir setzen

$$q_* := r_1 r_2 \text{ div } b, \qquad r_* := r_1 r_2 \text{ mod } b.$$

Damit ergibt sich

$$\begin{aligned} a_1 a_2 &= (bq_1+r_1)(bq_2+r_2) = bq_1bq_2+bq_1r_2+r_1bq_2+r_1r_2 \\ &= b(q_1q_2b+q_1r_2+r_1q_2)+r_1r_2 = b(q_1q_2b+q_1r_2+r_1q_2)+q_*b+r_* \\ &= b(q_1q_2b+q_1r_2+r_1q_2+q_*)+r_* \end{aligned}$$

und da $q_1q_2b+q_1r_2+r_1q_2+q_* \in \mathbb{Z}$, $r_* \in \mathbb{Z}$ und $0 \leq r_* \leq b-1$ folgt aus Satz 2.17

$$\begin{aligned} a_1a_2 \text{ div } b &= q_1q_2b+q_1r_2+r_1q_2+q_*, \\ a_1a_2 \text{ mod } b = r_* &= r_1r_2 \text{ mod } b = ((r_1 \text{ mod } b)*(r_2 \text{ mod } b)) \text{ mod } b. \end{aligned}$$

Damit sind beide Behauptungen bewiesen. □

Ausdrücke der Form a^n für gegebene $a,n \in \mathbb{N}$ lassen sich effizient berechnen, indem man ausnutzt, dass für $n \geq 2$ gilt

$$a^n = \begin{cases} a^{n/2} * a^{n/2} & \text{für } n \text{ gerade,} \\ a * a^{(n-1)/2} * a^{(n-1)/2} & \text{für } n \text{ ungerade.} \end{cases}$$

Wir müssen also zur Berechnung von a^n also nur rekursiv $a^{n/2}$ und eine weitere Multiplikation bzw. rekursiv $a^{(n-1)/2}$ und zwei weitere Multiplikationen berechnen. Diese Vorgehen nennt man *binäre Exponentiation* oder auch *Square-and-Multiply*.

Beispiel 2.19
Wir berechnen 2^{20}. *Es ist*

$$\begin{aligned} 2^{20} &= 2^{10} * 2^{10} \\ 2^{10} &= 2^5 * 2^5 \\ 2^5 &= 2 * 2^2 * 2^2 \\ 2^2 &= 2 * 2 \end{aligned}$$

Einsetzen von unten nach oben liefert deshalb

$$\begin{aligned} 2^2 &= 2*2 = 4 \\ 2^5 &= 2*4*4 = 32 \\ 2^{10} &= 32*32 = 1024 \\ 2^{20} &= 1024*1024 = 1048576 \end{aligned}$$

und wir haben somit 2^{20} *mit lediglich* 5 *Multiplikationen berechnet.*

Ausdrücke der Form $a^n \mod m$ für gegebene $a,n,m \in \mathbb{N}$ lassen sich damit noch effizienter berechnen, da bei jeder Multiplikation die Restbestimmung schon auf den Faktoren durchgeführt werden kann (wodurch immer nur mit Zahlen $< m$ gerechnet werden muss).

Beispiel 2.20
Wir berechnen die letzte Dezimalziffer von 123^{25}. *Zunächst nutzen wir aus, dass*

$$123^{25} \mod 10 = (123 \mod 10)^{25} \mod 10 = 3^{25} \mod 10.$$

Dieses berechnen wir jetzt mit (modularer) binärer Exponentiation:

$$\begin{aligned} 3^{25} \mod 10 &= 3*(3^{12} \mod 10)*(3^{12} \mod 10) \mod 10 \\ 3^{12} \mod 10 &= (3^6 \mod 10)*(3^6 \mod 10) \mod 10 \\ 3^6 \mod 10 &= (3^3 \mod 10)*(3^3 \mod 10) \mod 10 \\ 3^3 \mod 10 &= 3*(3 \mod 10)*(3 \mod 10) \mod 10 \end{aligned}$$

Einsetzen von unten nach oben liefert deshalb

$$\begin{aligned} 3^3 \mod 10 &= 3*3*3 \mod 10 = 27 \mod 10 = 7 \\ 3^6 \mod 10 &= 7*7 \mod 10 = 49 \mod 10 = 9 \\ 3^{12} \mod 10 &= 9*9 \mod 10 = 81 \mod 10 = 1 \\ 3^{25} \mod 10 &= 3*1*1 \mod 10 = 3. \end{aligned}$$

Die letzte Dezimalziffer von 123^{25} *ist also eine* 3.

2.2.2 Primfaktorzerlegung

Mit dem Begriff der Teilbarkeit können wir Primzahlen definieren.

Definition 2.21 (Primzahl)
Eine Primzahl ist eine Zahl $a \in \mathbb{N} \setminus \{1\}$, die nur die trivialen Teiler $\{-a,-1,1,a\}$ besitzt. Insbesondere ist die Zahl 1 keine Primzahl.

Satz 2.22 (Primfaktorzerlegung)
Für jede Zahl $a \in \mathbb{N}$, $a \neq 1$, existieren endlich viele Primzahlen $p_1,\ldots,p_m \in \mathbb{N}$, $m \in \mathbb{N}$, so dass

$$a = p_1 * p_2 * \ldots * p_m.$$

Beweis: Wir führen den Beweis mit dem Prinzip des kleinsten Verbrechers aus Folgerung 2.10. Wenn es natürliche Zahl außer Eins gibt, die keine Primfaktorzerlegung besitzen, dann muss es eine kleinste solche Zahl geben. Wir werden dies zum Widerspruch führen, womit dann gezeigt ist, dass es gar keine solchen Zahlen geben kann.

Sei also a diese kleinste natürliche Zahl außer Eins, die keine Primfaktorzerlegung besitzt. Dann kann a selbst schon keine Primzahl sein, denn sonst würde die Behauptung einfach mit $p_1 = a$, $n = 1$ und $e_1 = 1$ gelten. Da a also keine Primzahl ist, existiert ein nicht-trivialer Teiler $b \in \mathbb{N}$. Es gibt also ein $q \in \mathbb{N}$ mit $a = bq$. Da b kein trivialer Teiler ist, gilt $1 < b < a$ und damit auch $1 < q < a$ (q ist also auch ein nicht-trivialer Teier von a). Da b und q beide kleiner sind als a (und ungleich 1), müssen sie wegen der kleinstmöglichen Wahl von a beide eine Primfaktorzerlegung besitzen. Es existieren also Primzahlen $r_1,\ldots,r_l$ ($l \in \mathbb{N}$) und $\tilde{r}_1,\ldots,\tilde{r}_k$ ($k \in \mathbb{N}$) mit

$$b = r_1 * \ldots * r_l \quad \text{und} \quad q = \tilde{r}_1 * \ldots * \tilde{r}_k.$$

Damit ist aber auch

$$a = bq = r_1 * \ldots * r_l * \tilde{r}_1 * \ldots * \tilde{r}_k$$

ein Produkt von Primzahlen. Dies widerspricht der Tatsache, dass a keine solche Zerlegung besitzt. Es gibt also gar keine natürliche Zahl außer Eins, die keine Primfaktorzerlegung besitzen, und damit ist die Behauptung bewiesen. □

Bemerkung 2.23
Man kann zeigen, dass für eine Primzahl $p \in \mathbb{N}$ und zwei Zahlen $a,b \in \mathbb{N}$ stets gilt:

$$p|ab \quad \Longrightarrow \quad p|a \ \vee \ p|b. \tag{2.6}$$

Daraus folgt dann auch, dass die Primfaktorzerlegung bis auf die Reihenfolge der Faktoren eindeutig ist. Gilt also für eine Zahl $a \in \mathbb{N}$, $a \neq 1$,

$$a = p_1 p_2 \cdots p_n \quad \textit{und auch} \quad a = q_1 q_2 \cdots q_m,$$

mit (der Größe nach sortierten) Primzahlen $p_1 \leq p_2 \leq \ldots \leq p_n$ und (ebenfalls sortierten) Primzahlen $q_1 \leq q_2 \leq \ldots \leq q_m$, $n,m \in \mathbb{N}$, dann ist $m = n$ und $p_j = q_j$ für alle $j = 1,\ldots,n$.

Üblicherweise fasst man in der Primfaktorzerlegung gleiche Primzahlen zusammen und schreibt

$$a = p_1^{e_1} p_2^{e_2} \cdots p_n^{e_n}$$

mit verschiedenen Primzahlen $p_1 < p_2 < \ldots < p_n$ und Exponenten $e_1,\ldots,e_n \in \mathbb{N}$.

Damit folgt dann, dass die positiven Teiler von a genau die Zahlen der Form

$$b = p_1^{d_1} p_2^{d_2} \cdots p_n^{d_n}$$

mit $0 \leq d_j \leq e_j$ (für alle $j \in \{1,\ldots,n\}$) sind.

2.2.3 Der ggT und der Euklidische Algorithmus

Definition 2.24
Seien $a,b,c \in \mathbb{N}$. Wenn $c|a$ und $c|b$, dann heißt c gemeinsamer Teiler von a und b. Die größte solche Zahl heißt größter gemeinsamer Teiler $\mathrm{ggT}(a,b)$.

Falls $a|c$ und $b|c$, dann heißt c gemeinsames Vielfaches. Die kleinste solche Zahl heißt kleinstes gemeinsames Vielfaches $\mathrm{kgV}(a,b)$.

Entsprechend unserer Konvention, dass jede Zahl die Null teilt, aber die Null selbst keine Zahl teilt, definieren wir für $a,b \in \mathbb{N}$:

$$\mathrm{ggT}(a,0) := a \quad \text{und} \quad \mathrm{ggT}(0,b) := b. \tag{2.7}$$

Für einen später (in (2.9)) gezeigten Zusammenhang zwischen ggT und kgV ist es nützlich, außerdem folgendes zu definieren:

$$\mathrm{kgV}(a,0) := 0, \quad \mathrm{kgV}(0,b) := 0, \quad \mathrm{ggT}(0,0) := 0, \quad \mathrm{kgV}(0,0) := 0. \tag{2.8}$$

Da die Zahl 1 jede andere Zahl teilt, gilt immer

$$\mathrm{ggT}(a,b) \geq 1 \quad \text{für alle } a,b \in \mathbb{N}.$$

Zwei Zahlen $a,b \in \mathbb{N}$ mit $\mathrm{ggT}(a,b) = 1$ heißen auch *teilerfremd*.

$\mathrm{ggT}(a,b)$ und $\mathrm{kgV}(a,b)$ lassen sich leicht aus der Primfaktorzerlegung von $a \in \mathbb{N}$ und $b \in \mathbb{N}$ ablesen. Wir sortieren die Primfaktoren von a und b, so dass die ersten $k \in \mathbb{N}_0$ Faktoren übereinstimmen, also

$$\begin{aligned} a &= p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k}\ p_{k+1}^{e_{k+1}} \cdots p_n^{e_n} \\ b &= p_1^{d_1} p_2^{d_2} \cdots p_k^{d_k}\ q_{k+1}^{d_{k+1}} \cdots q_m^{e_m}. \end{aligned}$$

Dann gilt offenbar

$$\begin{aligned}\mathrm{ggT}(a,b) &= p_1^{\min(e_1,d_1)} p_2^{\min(e_2,d_2)} \cdots p_k^{\min(e_k,d_k)} \\ \mathrm{kgV}(a,b) &= p_1^{\max(e_1,d_1)} p_2^{\max(e_2,d_2)} \cdots p_k^{\max(e_k,d_k)}\ p_{k+1}^{e_{k+1}} \cdots p_n^{e_n}\ q_{k+1}^{d_{k+1}} \cdots q_m^{e_m}.\end{aligned}$$

Wegen $\min(d_j,e_j)\max(d_j,e_j) = e_j d_j$ folgt daraus auch

$$\mathrm{ggT}(a,b) * \mathrm{kgV}(a,b) = ab. \tag{2.9}$$

Wegen (2.7), (2.8) gilt dies auch für $a = 0$ oder $b = 0$.

Die Berechnung der Primfaktorzerlegung einer großen Zahl ist aber sehr aufwändig. Effizienter lässt sich der ggT zweier Zahlen (und damit wegen (2.9) auch das kgV) mit dem *Euklidischen Algorithmus* bestimmen. Der Algorithmus beruht auf der folgenden Beobachtung.

Satz 2.25
Seien $a,b \in \mathbb{N}$.

(a) Ist $a \geq b$, *dann ist* $\mathrm{ggT}(a,b) = \mathrm{ggT}(a-b,b)$.

(b) Ist $b \geq a$, *dann ist* $\mathrm{ggT}(a,b) = \mathrm{ggT}(a,b-a)$.

Beweis: Wir müssen nur (a) beweisen. (b) folgt dann aus $\mathrm{ggT}(a,b) = \mathrm{ggT}(b,a)$. Im Falle $a = b$ stimmt die Behauptung, da

$$\mathrm{ggT}(a,b) = a = b = \mathrm{ggT}(0,b) = \mathrm{ggT}(a-b,b).$$

Sei also $a > b$. Da $\mathrm{ggT}(a,b)$ sowohl a als auch b teilt, teilt $\mathrm{ggT}(a,b)$ auch $a-b$ (siehe Satz 2.12(d)). $\mathrm{ggT}(a,b)$ ist also ein gemeinsamer Teiler von $a-b$ und b und da $\mathrm{ggT}(a-b,b)$ der größte solche ist, folgt $\mathrm{ggT}(a,b) \leq \mathrm{ggT}(a-b,b)$.

Umgekehrt ist aber $\mathrm{ggT}(a-b,b)$ ein Teiler von $a-b$ und von b und damit auch ein Teiler von $a = a-b+b$ (siehe wieder Satz 2.12(d)). $\mathrm{ggT}(a-b,b)$ ist also ein gemeinsamer Teiler von $a-b$ und b und da $\mathrm{ggT}(a,b)$ der größte solche ist, folgt $\mathrm{ggT}(a-b,b) \leq \mathrm{ggT}(a,b)$. Insgesamt gilt also $\mathrm{ggT}(a-b,b) = \mathrm{ggT}(a,b)$. □

Mit Satz 2.25 lässt sich die Berechnung des ggT zweier Zahlen auf die ggT Berechnung kleinerer Zahlen reduzieren. Dabei verringert sich immer eine der beiden Zahlen, aber ohne das sie negativ wird. Da dies nur endlich oft möglich ist, wird eine der Zahlen irgendwann Null und der ggT kann (in der anderen Zahl) abgelesen werden. Algorithmus 1 formuliert das Vorgehen in Pseudocode in einer rekursiven Variante. Wir sparen uns dabei den unnötigen letzten Schritt, in dem eine Zahl gleich Null wird, und fangen direkt den davor auftretenden Fall $a = b$ ab.

Algorithm 1 Euklidischer Algorithmus (klassisch und rekursiv)

```
Gegeben a, b ∈ ℕ
function c = ggT(a, b)
    if a > b then
        c := ggT(a − b, b)
    else if a < b then
        c := ggT(a, b − a)
    else
        c := a
    end if
    return c.
end function
```

Beispiel 2.26
Der klassische Euklidische Algorithmus berechnet den $\mathrm{ggT}(143,117)$ *mit den folgenden Schritten:*

$$\begin{aligned}
&\mathrm{ggT}(143,117) = \mathrm{ggT}(143-117,117)\\
&= \mathrm{ggT}(26,117) = \mathrm{ggT}(26,117-26)\\
&= \mathrm{ggT}(26,91) = \mathrm{ggT}(26,91-26)\\
&= \mathrm{ggT}(26,65) = \mathrm{ggT}(26,65-26)\\
&= \mathrm{ggT}(26,39) = \mathrm{ggT}(26,39-26)\\
&= \mathrm{ggT}(26,13) = \mathrm{ggT}(26-13,13)\\
&= \mathrm{ggT}(13,13) = 13.
\end{aligned}$$

Ist $a > b$ dann kann durch wiederholte Anwendung von Satz 2.25 a so oft um b verringert werden, bis ein Rest verbleibt, der kleiner ist als b. Dies entpricht gerade der Division mit Rest, und wir können die Reduktion auch gleich damit formulieren.

Satz 2.27
Seien $a,b \in \mathbb{N}$.

(a) Ist $a \geq b$*, dann ist* $\mathrm{ggT}(a,b) = \mathrm{ggT}(a \bmod b, b)$.

(b) Ist $b \geq a$*, dann ist* $\mathrm{ggT}(a,b) = \mathrm{ggT}(a, b \bmod a)$.

Beweis: Wieder müssen wir nur (a) beweisen und (b) folgt dann aus der Symmetrie $\mathrm{ggT}(a,b) = \mathrm{ggT}(b,a)$. Im Falle $a = b$ ist $a \bmod b = 0$ und die Behauptung stimmt wegen

$$\mathrm{ggT}(a,b) = a = b = \mathrm{ggT}(0,b) = \mathrm{ggT}(a \bmod b, b).$$

Sei also $a > b$. Nach Definition und Satz 2.15 ist $a \mod b = a - bq \in \mathbb{N}_0$ mit $q := \lfloor \frac{a}{b} \rfloor \in \mathbb{N}_0$. Durch mehrmalige Verwendung von Satz 2.25 erhalten wir deshalb

$$\begin{aligned} \mathrm{ggT}(a,b) &= \mathrm{ggT}(a-b,b) = \mathrm{ggT}(a-2*b,b) = \ldots = \mathrm{ggT}(a-(q-1)*b,b) \\ &= \mathrm{ggT}(a-q*b,b) = \mathrm{ggT}(a \mod b,b). \end{aligned}$$

Damit ist (a) bewiesen. □

Algorithm 2 Euklidischer Algorithmus (modern, rekursiv)

```
Gegeben a, b ∈ ℕ_0
function c = ggT(a, b)
    if a = 0 then
        c := b
    else if b = 0 then
        c := a
    else if a > b then
        c := ggT(a mod b, b)
    else
        c := ggT(a, b mod a)
    end if
    return c.
end function
```

Beispiel 2.28
Der moderne Euklidische Algorithmus berechnet den $\mathrm{ggT}(143,117)$ *mit den folgenden Schritten:*

$$\begin{aligned} \mathrm{ggT}(143,117) &= \mathrm{ggT}(143 \mod 117,117) \\ &= \mathrm{ggT}(26,117) = \mathrm{ggT}(26,117 \mod 26) \\ &= \mathrm{ggT}(26,13) = \mathrm{ggT}(26 \mod 13),13) \\ &= \mathrm{ggT}(0,13) = 13. \end{aligned}$$

2.2.4 Lemma von Bézout und erweiterter Euklid

Wir werden noch eine Erweiterung des Euklidischen Algorithmus benötigen. Dazu formulieren wir zunächst das Lemma von Bézout.

Satz 2.29 (Lemma von Bézout)
Für alle $a,b \in \mathbb{N}_0$ *existieren* $s,t \in \mathbb{Z}$*, so dass*

$$\mathrm{ggT}(a,b) = sa + tb.$$

Beweis: Wir entwickeln im Rest des Abschnitts einen Algorithmus zur Konstruktion von s und t. Damit ist dann auch das Lemma gezeigt. □

Den gewünschten Algorithmus zur Berechnung von s und t enthalten wir aus dem Euklidischen Algorithmus. Wir betrachten dazu nochmal Beispiel 2.28:

Beispiel 2.30
Der moderne Euklidische Algorithmus berechnet den $\mathrm{ggT}(143,117)$ *mit den folgenden Schritten:*

$$\begin{aligned}\mathrm{ggT}(143,117) &= \mathrm{ggT}(143 \bmod 117, 117)\\ &= \mathrm{ggT}(26,117) = \mathrm{ggT}(26, 117 \bmod 26)\\ &= \mathrm{ggT}(26,13) = \mathrm{ggT}(26 \bmod 13), 13)\\ &= \mathrm{ggT}(0,13) = 13.\end{aligned}$$

Die dabei durchgeführten Divisionen mit Rest lauten ausgeschrieben:

$$\begin{aligned}143 &= 1*117+26\\ 117 &= 4*26+13\\ 26 &= 2*13+0.\end{aligned}$$

Löst man die vorletzte Zeile nach dem ggT *13 auf und setzt das jeweils in die vorherigen Zeilen ein, dann erhält man*

$$13 = 117-4*26 = 117-4*(143-1*117) = 5*117-4*143.$$

Dies ist gerade die gewünschte Darstellung.

Wir formulieren diese Idee in einem rekursiven Algorithmus, der zu gegebenen Zahlen $a,b \in \mathbb{N}_0$ nicht nur $c = \mathrm{ggT}(a,b)$ sondern auch zwei Zahlen $s,t \in \mathbb{Z}$ mit $c = sa+tb$ berechnet. Im Rekursionsschritt verwenden wir dabei für $a > b$ wie im Euklidischen Algorithmus, dass

$$\mathrm{ggT}(a,b) = \mathrm{ggT}(a \bmod b, b)$$

gilt. Zusätzlich verwenden wir, dass sich aus einer Bézout-Darstellung der Form

$$\mathrm{ggT}(a \bmod b, b) = s'*(a \bmod b)+t'*b$$

wegen $a \bmod b = a-b*(a \operatorname{div} b)$ auch eine Bézout-Darstellung von $\mathrm{ggT}(a,b)$ ergibt, nämlich

$$\begin{aligned}\mathrm{ggT}(a,b) = \mathrm{ggT}(a \bmod b, b) &= s'*(a \bmod b)+t'*b\\ &= s'*(a-b*(a \operatorname{div} b))+t'*b\\ &= s'*a+(t'-s'*(a \operatorname{div} b))*b = sa+tb,\end{aligned}$$

mit $s := s'$ und $t := t'-s'*(a \operatorname{div} b)$. Analoges gilt für $a < b$. Als Rekursionsanfang verwenden wir wieder $\mathrm{ggT}(a,0) = a$ und $a = 1*a+0*b$ sowie die analogen Formeln für $\mathrm{ggT}(0,b)$.

Algorithm 3 Erweiterter Euklidischer Algorithmus

Gegeben $a, b \in \mathbb{N}_0$
function $[c, s, t] = \text{erweiterter\_ggT}(a, b)$
  if $a = 0$ **then**
    $c := b$, $s := 0$, $t := 1$
  else if $b = 0$ **then**
    $c := a$, $s := 1$, $t := 0$
  else if $a > b$ **then**
    $[c, s', t'] := \text{erweiterter\_ggT}(a \bmod b, b)$
    $s := s'$, $t := t' - s' * (a \text{ div } b)$
  else
    $[c, s', t'] :=:= \text{erweiterter\_ggT}(a, b \bmod a)$
    $t := t'$, $s := s' - t' * (b \text{ div } a)$
  end if
  return $[c, s, t]$.
end function

2.2.5 Chinesischer Restsatz

Zu $m \in \mathbb{N}$ und $a \in \mathbb{Z}$ besitzt die Gleichung

$$x \bmod m = a \bmod m$$

für x die Lösungen a, $a \pm m$, $a \pm 2m$, ...

Schwieriger ist es, eine Zahl $x \in \mathbb{Z}$ zu finden, die mehrere solcher Gleichungen erfüllt. Wir betrachten zunächst den Fall zweier solcher Gleichungen.

Satz 2.31 (Chinesischer Restsatz für zwei Gleichungen)
Seien $m_1, m_2 \in \mathbb{N}$, $a_1, a_2 \in \mathbb{Z}$. Ist $\text{ggT}(m_1, m_2) = 1$, dann existiert eine Lösung $x \in \mathbb{Z}$ von

$$x \bmod m_1 = a_1 \bmod m_1, \tag{2.10}$$
$$x \bmod m_2 = a_2 \bmod m_2. \tag{2.11}$$

Mit den Zahlen $s, t \in \mathbb{Z}$ aus der Bézout-Darstellung[1]

$$1 = \text{ggT}(m_1, m_2) = sm_1 + tm_2$$

ist eine Lösung gegeben durch

$$x = a_1 t m_2 + a_2 s m_1$$

und die Menge aller Lösungen ist $\{x, x \pm M, x \pm 2M, \ldots\}$ mit $M := m_1 m_2$.

[1] also mit Algorithmus 3: $[1, s, t] = \text{erweiterter\_ggT}(m_1, m_2)$

Beweis: $x := a_1 t m_2 + a_2 s m_1$ erfüllt

$$\begin{aligned} x \bmod m_1 &= a_1 t m_2 \bmod m_1 = a_1(1 - s m_1) \bmod m_1 \\ &= a_1 - a_1 s m_1 \bmod m_1 = a_1 \bmod m_1. \end{aligned}$$

und genauso folgt

$$x \bmod m_2 = a_2 \bmod m_2.$$

x löst also die beiden Gleichungen (2.10) und (2.11).

Sei $\tilde{x} \in \mathbb{Z}$ eine weitere Lösung der Gleichungen (2.10) und (2.11). Dann gilt

$$\begin{aligned} x - \tilde{x} \bmod m_1 &= a_1 - a_1 \bmod m_1 = 0, \\ x - \tilde{x} \bmod m_2 &= a_2 - a_2 \bmod m_2 = 0, \end{aligned}$$

also sowohl $m_1 | (x - \tilde{x})$ als auch $m_2 | (x - \tilde{x})$. Da m_1 und m_2 teilerfremd sind folgt dass auch $M := m_1 m_2$ die Differenz $x - \tilde{x}$ teilt, und daher ist $\tilde{x} = x + kM$ mit einem $k \in \mathbb{Z}$. Umgekehrt ist offensichtlich auch jedes solche $\tilde{x} = x + kM$ eine Lösung Lösung der Gleichungen (2.10) und (2.11). □

Die Aussage von Satz 2.31 bedeutet, dass die Lösungsmenge der beiden Gleichungen (2.10) und (2.11) übereinstimmt mit der Lösungsmenge der Gleichung

$$x \bmod (m_1 m_2) = a_1 t m_2 + a_2 s m_1 \bmod (m_1 m_2). \tag{2.12}$$

Es gilt also

$$(2.10) \wedge (2.11) \quad \Longleftrightarrow \quad (2.12).$$

Damit können wir den Chinesischen Restsatz leicht auf mehr als zwei Gleichungen erweitern. Wir kombinieren dazu die ersten zwei Gleichungen zu einer Gleichung, dann die enstandendene Gleichung mit der dritten Gleichung und die dann enstandendene mit der vierten und fahren so fort, bis wir alle Gleichungen kombiniert haben. Die Kombination setzt dabei jeweils voraus, dass die betrachteten modulo Gleichungen zu teilerfremden Divisoren stattfinden. Dazu müssen die Divisoren der ursprünglich gegebenen Gleichungen *paarweise* teilerfremd sein.

Für nicht teilerfremde Divisoren ist die Lösbarkeit nicht immer gesichert. Zum Beispiel existiert keine Lösung $x \in \mathbb{Z}$ von

$$x \bmod 2 = 1 \quad \text{und} \quad x \bmod 4 = 0,$$

da jede durch 4 teilbare Zahl auch durch 2 teilbar ist. Man kann aber zeigen, dass die Gleichungen

$$x \bmod m_1 = a_1 \bmod m_1 \quad \text{und} \quad x \bmod m_2 = a_2 \bmod m_2$$

genau dann lösbar sind, wenn

$$a_1 \bmod \mathrm{ggT}(m_1,m_2) = a_2 \bmod \mathrm{ggT}(m_1,m_2)$$

und auch hierfür eine explizite Lösungsformel angeben (allgemeiner Chinesischer Restsatz) und dies mit der oben skizzierten Ideen auch auf mehr als zwei Gleichungen anwenden.

Beispiel 2.32
Wir betrachten das Problem eine Zahl zu finden, die durch 2, 3, 4, 5 *und* 6 *jeweils mit Rest* 1 *teilbar ist und durch* 7 *ohne Rest teilbar ist.*[2] *Wir suchen also* $x \in \mathbb{Z}$ *mit*

$$\begin{array}{lll} x \bmod 2 = 1, & x \bmod 4 = 1, & x \bmod 6 = 1, \\ x \bmod 3 = 1, & x \bmod 5 = 1, & x \bmod 7 = 0. \end{array}$$

Die ersten fünf Gleichungen sind äquivalent dazu, dass

$$2|(x-1) \wedge 3|(x-1) \wedge 4|(x-1) \wedge 5|(x-1) \wedge 6|(x-1).$$

Da jede Zahl, die durch 2 *und* 3 *teilbar ist, auch durch* $2*3=6$ *teilbar ist, und jede durch* 4 *teilbare Zahl auch durch* 2 *teilbar ist, ist dies äquivalent zu*

$$3|(x-1) \wedge 4|(x-1) \wedge 5|(x-1)$$

und dies ist wegen der Teilerfremdheit von 3, 4 *und* 5 *genau dann erfüllt, wenn* $x-1$ *durch* $3*4*5=60$ *teilbar ist. Die ersten fünf Gleichungen sind also äquivalent zu*

$$x \bmod 60 = 1$$

und wir suchen daher eine Lösung von

$$x \bmod 60 = 1 \quad \textit{und} \quad x \bmod 7 = 0 \tag{2.13}$$

und können dazu (wegen $\mathrm{ggT}(60,7)=1$*) Satz 2.31 anwenden.*

Dazu benötigen wir zunächst eine Bézout-Darstellung der Form

$$1 = s*60 + t*7.$$

Mit dem erweiterten Euklidischen Algorithmus (oder geschicktem Ausprobieren) erhalten wir

$$1 = 2*60 - 17*7, \quad \textit{also} \quad s=2,\ t=-17.$$

[2] `https://de.wikipedia.org/wiki/Eieraufgabe_des_Brahmagupta`

Mit Satz 2.31 folgt dass die Lösungen von (2.13) *gegeben sind durch*

$$x,\ x \pm M, x \pm 2M, \ldots$$

mit $M = 60 * 7 := 420$ *und* $x := 1 * (-17) * 7 + 0 * 2 * 60 = -119$. *Die kleinste natürliche Lösung von* (2.13) *ist daher* $-119 + 420 = 301$ *und dies erfüllt damit auch unsere sechs Gleichungen*

$$\begin{array}{lll} 301 \bmod 2 = 1, & 301 \bmod 4 = 1, & 301 \bmod 6 = 1, \\ 301 \bmod 3 = 1, & 301 \bmod 5 = 1, & 301 \bmod 7 = 0. \end{array}$$

2.3 Restklassenarithmetik

2.3.1 Restklassenring und Primzahlkörper

Sei $m \in \mathbb{N}$. In Satz 2.18 haben wir gezeigt, dass wir bei der Berechnung der Summe und des Produktes zweier Zahlen modulo m, auch schon vorher die Summanden bzw. Faktoren modulo m berechnen können, um so mit kleineren Zahlen zu rechnen. Bisher haben wir das explizit ausgeschrieben:

$$\begin{aligned} (a+b) \bmod m &= ((a \bmod m) + (b \bmod m)) \bmod m, \\ ab \bmod m &= ((a \bmod m) * (b \bmod m)) \bmod m. \end{aligned}$$

Der Vorteil dieser Schreibweise ist, dass alle Rechnungen in $\mathbb{Z}$ stattfinden (wo die bekannten Rechenregeln für ganze Zahlen gelten) und die Divisionen mit Rest explizit genannt werden. Gerade für längere Rechnungen, bei denen nur das Endergebnis modulo m interessiert, wäre es zur besseren Lesbarkeit wünschenswert, solche langen Ausdrücke mit häufigen Wiederholungen von mod m zu vermeiden. Deshalb sind in der Literatur noch andere Schreibweisen für die Modulo-Rechnung gebräuchlich.

(a) *Integration ins Gleichheitszeichen:* Statt $a \bmod m = b \bmod m$ schreiben wir

$$a \equiv b \pmod{m}$$

Wenn aus dem Kontext klar ist, bezüglich welcher Zahl der Rest gebildet wird, schreiben wir auch noch kürzer: $a \equiv b$. In der Literatur wird auch manchmal $a \equiv_m b$ geschrieben. $\equiv$ heißt *Äquivalenzrelation*, da es die folgenden drei vom Gleichheitszeichen bekannten Eigenschaften besitzt: Für alle $a, b, c \in \mathbb{Z}$ gilt:

$$\begin{aligned} a &\equiv a && && && \text{(Reflexivität)} \\ a &\equiv b && \iff && b \equiv a && \text{(Symmetrie)} \\ a \equiv b \ \wedge\ b &\equiv c && \implies && a \equiv c && \text{(Transitivität).} \end{aligned}$$

Aufgrund von Satz 2.18 gilt außerdem für alle $a, b, c \in \mathbb{Z}$

$$\begin{aligned} a \equiv b \quad &\Longrightarrow \quad a + c \equiv b + c \\ a \equiv b \quad &\Longrightarrow \quad ac \equiv bc. \end{aligned}$$

Wir können also mit dem Symbol $\equiv$ statt dem Gleichheitszeichen und den aus $\mathbb{Z}$ gewohnten Rechenregeln zur Multiplikation und Addition rechnen, wenn uns das Ergebnis nur modulo m interessiert.

(b) *Integration ins Verknüpfungssymbol:* Die Reste einer Division durch m liegen immer in der endlichen Menge

$$\mathbb{Z}_m := \{0, 1, \ldots, m-1\}.$$

Auf diesen Zahlen können wir auch gleich die Addition und Multiplikation so definieren, dass die anschließende Restbildung durchgeführt wird. Wir definieren also die Verknüpfungen $\oplus : \mathbb{Z}_m \times \mathbb{Z}_m \to \mathbb{Z}_m$ und $\otimes : \mathbb{Z}_m \times \mathbb{Z}_m \to \mathbb{Z}_m$ durch

$$\begin{aligned} a \oplus b &= a + b \mod m \\ a \otimes b &= a * b \mod m. \end{aligned}$$

Dies entspricht der normalen Addition und Multiplikation in $\mathbb{Z}_m$ mit der Maßgabe, dass bei einem Überlauf über $m-1$ hinaus, es wieder mit 0 losgeht. Z.B. führen Drehungen um ganze Gradzahlen in natürlicher Weise auf die Menge

$$\mathbb{Z}_{360} = \{0, 1, \ldots, 359\}.$$

Eine Drehung um 359 Grad gefolgt von einer um 1 Grad entspricht einer effektiven Drehung um 0 Grad, und 37 Drehungen um jeweils 10 Grad entsprechen einer Drehung um 10 Grad.

(c) *Integration in die Zahl:* Die dritte Möglichkeit, besteht darin, die Zahl $a \in \mathbb{Z}$ zu ersetzen durch die Menge aller Zahlen, die den gleichen Rest modulo m besitzen. Wir bilden

$$[a] := \{\ldots, a-2m, a-m, a, a+m, a+2m, \ldots\}.$$

$[a]$ heißt auch *Äquivalenzklasse* oder *Restklasse*. a heißt *Repräsentant* der Restklasse. Es gilt

$$[a] = [b] \quad \Longleftrightarrow \quad a \equiv b \pmod{m} \quad \Longleftrightarrow \quad a - b \in m\mathbb{Z},$$

wobei $m\mathbb{Z} = \{mz : z \in \mathbb{Z}\}$. Die Menge aller solcher Restklassen bezeichnen wir auch mit $\mathbb{Z}/m\mathbb{Z}$. Auf den Restklassen definiert man die Addition und Multiplikation durch

$$[a] + [b] := [a + b], \quad [a][b] := [ab].$$

Im Folgenden verwenden die Schreibweise aus (b) und zeigen, dass die endliche Menge $\mathbb{Z}_m$ mit der modulo m gebildeten Addition $\oplus$ und Multiplikation $\otimes$ für allgemeine m ein Ring und für Primzahlen m sogar ein Körper ist. Alle folgenden Ergebnisse gelten genauso auch für $\mathbb{Z}/m\mathbb{Z}$.

Satz 2.33
$(\mathbb{Z}_m, \oplus)$ *ist eine kommutative Gruppe, d.h es gilt*

(G1) Assoziativität: *Für alle* $a, b, c \in \mathbb{Z}_m$ *gilt*

$$(a \oplus b) \oplus c = a \oplus (b \oplus c) = a + b + c \mod m.$$

(G2) Existenz des neutralen Elements:

$$a \oplus 0 = 0 \oplus a = a \quad \textit{für alle } a \in \mathbb{Z}_m.$$

(G3) Existenz inverser Elemente: *Für alle* $a \in \mathbb{Z}_m \setminus \{0\}$ *ist*

$$\ominus a = -a \mod m = m - a \in \mathbb{Z}_m$$

das additiv inverse Element zu a, d.h. es gilt $a \ominus a = 0$. $0 \in \mathbb{Z}_m$ *ist zu sich selbst inverses Element, denn* $0 \oplus 0 = 0$.

(G4) Kommutativität:

$$a \oplus b = b \oplus a \quad \textit{für alle } a, b \in \mathbb{Z}_m.$$

Beweis: Durch wiederholtes Anwenden von Satz 2.18 erhalten wir

$$\begin{aligned}(a \oplus b) \oplus c &= ((a + b \mod m) + c) \mod m \\ &= (a + b \mod m) + (c \mod m)) \mod m \\ &= ((a + b) + c \mod m)) \mod m \\ &= a + b + c \mod m.\end{aligned}$$

Genauso folgt

$$a \oplus (b \oplus c) = a + b + c \mod m,$$

so dass (G1) gezeigt ist.

(G2) und (G3) folgen durch einfaches Nachrechnen, und wegen

$$a \oplus b = a + b \mod m = b + a \mod m = b \oplus a$$

folgt auch (G4). □

Satz 2.34
$(\mathbb{Z}_m, \oplus, \otimes)$ *ist ein kommutativer Ring mit Eins, d.h.*

(R1) $(\mathbb{Z}_m, \oplus)$ *ist eine kommutative Gruppe.*

(R2) $(\mathbb{Z}_m, \otimes)$ *ist eine Halbgruppe.*

(R3) Distributivität: Für alle $a, b, c \in \mathbb{Z}_m$ *gilt:*

$$a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$$
$$(a \oplus b) \otimes c = (a \otimes c) \oplus (b \otimes c),$$

$(\mathbb{Z}_m, \otimes)$ *ist kommutativ und* $1 \in \mathbb{Z}_m$ *ist neutrales Element bezüglich* $\otimes$.

Beweis: Leichtes Nachrechnen. □

In $\mathbb{Z}_m$ lassen sich Additionen und Multiplikationen leicht durchführen, da sie einfach in $\mathbb{Z}$ durchgeführt werden können und danach der Rest modulo m berechnet wird. Die Rechnung ist üblichweise sogar einfacher als in $\mathbb{Z}$, da schon Zwischenergebnisse durch Restbildung verkleinert werden können, wodurch mit kleineren Zahlen gerechnet werden kann. Die Subtraktion (also das Inverse zur Addition $\oplus a$) kann ebenso einfach entweder durch Subtraktion in $\mathbb{Z}$ und anschließende Restbildung durchgeführt werden.

Um in $\mathbb{Z}_m$ auch dividieren zu können, müssten Elemente in $\mathbb{Z}_m$ auch eine multiplikative Inverse besitzen, d.h. $\mathbb{Z}_m$ müsste sogar ein Körper sein. Dies ist aber nicht immer der Fall. Beispielsweise ist in $\mathbb{Z}_4 = \{0, 1, 2, 3\}$

$$0 \otimes 2 = 0, \quad 1 \otimes 2 = 2, \quad 2 \otimes 2 = 0, \quad \text{und} \quad 3 \otimes 2 = 2,$$
$$0 \otimes 3 = 0, \quad 1 \otimes 3 = 3, \quad 2 \otimes 3 = 2, \quad \text{und} \quad 3 \otimes 3 = 1.$$

2 besitzt daher in $\mathbb{Z}_4$ kein multiplikatives Inverses (durch 2 können wir also in $\mathbb{Z}_4$ nicht teilen), aber 3 besitzt in $\mathbb{Z}_4$ das multiplikative Inverse 3 (durch 3 zu teilen ist in $\mathbb{Z}_4$ das gleiche, wie mit 3 zu multiplizieren).

Das erscheint erstmal nicht überraschend, da ja auch $\mathbb{Z}$ keine multiplikatives Inversen enthält (also nur ein kommutativer Ring mit Eins aber kein Körper ist). Erst die Menge der rationalen Zahlen $\mathbb{Q}$ enthält z.B. das multiplikativ Inverse $2^{-1} = \frac{1}{2}$ zur ganzen Zahl 2.

Der folgende Satz gibt ein Kriterium, wann eine Zahl in $\mathbb{Z}_m$ doch ein multiplikatives Inverses besitzt.

Satz 2.35 (Diskretes multiplikatives Inverses)
Sei $m \in \mathbb{N}$. Eine Zahl $a \in \mathbb{Z}_m$ besitzt genau dann ein multiplikatives Inverses in $\mathbb{Z}_m$, wenn

$$\text{ggT}(a,m) = 1. \tag{2.14}$$

Gilt (2.14) *und ist*

$$1 = \text{ggT}(a,m) = s * a + t * m$$

mit $s,t \in \mathbb{Z}$ die zugehörige Bézout-Darstellung, dann ist

$$s \mod m \in \mathbb{Z}_m$$

die multiplikative Inverse von a in $\mathbb{Z}_m$.

Beweis: Sei $m \in \mathbb{N}$. $a \in \mathbb{Z}_m$ besitze ein multiplikatives Inverses $b \in \mathbb{Z}_m$, dann gilt also

$$1 = a \otimes b = ab \mod m.$$

Es existiert also ein $q \in \mathbb{Z}$ mit

$$ab = qm + 1 \quad \text{und damit} \quad 1 = ab - qm.$$

Jeder gemeinsame Teiler von a und m muss daher auch ein Teiler von 1 sein und der einzige positive gemeinsame Teiler ist daher $\text{ggT}(a,m) = 1$.

Gilt umgekehrt $\text{ggT}(a,m) = 1$, dann existieren nach dem Lemma von Bézout (Satz 2.29) Zahlen $s,t \in \mathbb{Z}$ mit

$$1 = \text{ggT}(a,m) = s * a + t * m.$$

Es ist also

$$\begin{aligned}(s \mod m) \otimes a &= ((s \mod m) * a) \mod m \\ &= s * a \mod m = (1 - t * m) \mod m = 1.\end{aligned}$$

a besitzt also ein multiplikatives Inverses und dieses ist $s \mod m$. □

Folgerung 2.36
In $\mathbb{Z}_m$ besitzt genau dann jedes Element ein multiplikatives Inverses, falls m eine Primzahl ist. Für jede Primzahl $p \in \mathbb{N}$ ist $(\mathbb{Z}_p^\times, \otimes)$ mit $\mathbb{Z}_p^\times := \mathbb{Z}_p \setminus \{0\}$ eine kommutative Gruppe und daher $(\mathbb{Z}_p, \oplus, \otimes)$ ein Körper (der sogenannte Primzahlkörper*). Addition, additive Inverse und Multiplikation können wie in $\mathbb{Z}$ berechnet werden (gefolgt von der Restbildung modulo m). Die multiplikative Inverse kann gemäß Satz 2.35 über den erweiterten Euklidischen Algorithmus bestimmt werden.*

Beispiel 2.37
Gemäß Folgerung 2.36 ist $\mathbb{Z}_2 = \{0,1\}$ *ein Körper. Dabei ist*

$$0 \oplus 0 = 0, \quad 0 \oplus 1 = 1, \quad 1 \oplus 0 = 1, \quad 1 \oplus 1 = 0,$$
$$0 \otimes 0 = 0, \quad 0 \otimes 1 = 0, \quad 1 \otimes 0 = 0, \quad 1 \otimes 1 = 1.$$

2 Werden 0 *und* 1 *als Wahrheitswerte interpretiert, dann entspricht die Addition* $\oplus$ *in* $\mathbb{Z}_2$ *also gerade der* XOR*-Verknüpfung und die Multiplikation der* AND*-Verknüpfung.* 0 *und* 1 *sind in* $\mathbb{Z}_2$ *jeweils zu sich selbst additiv invers, und* 1 *ist zu sich selbst multiplikativ invers.*

2.3.2 Diskrete Wurzel und diskreter Logarithmus

Sei $p \in \mathbb{N}$ eine Primzahl und $a, b \in \mathbb{Z}_p$. Können wir ein $x \in \mathbb{Z}_p$ finden, dass die Gleichung

$$x^a \bmod p = b \quad \text{oder} \quad a^x \bmod p = b$$

erfüllt, d.h. können wir in $\mathbb{Z}_p$ die a-te *diskrete Wurzel* ziehen?

Die diskrete Wurzel existiert nicht immer, z.B. gilt in $\mathbb{Z}_3$

$$0 \otimes 0 = 0, \quad 1 \otimes 1 = 1, \quad 2 \otimes 2 = 1.$$

Die Gleichung $x^2 \bmod 3 = b$ besitzt also für $b = 0$ eine eindeutige Lösung, für $b = 1$ zwei Lösungen und für $b = 2$ keine Lösung. Diskrete Wurzeln können daher nur unter zusätzlichen Annahmen gezogen werden.

Genauso können wir versuchen, ein $x \in \mathbb{Z}_p$ zu finden, dass

$$a^x \bmod p = b$$

erfüllt, d.h. den *diskreten Logarithmus* zur Basis a zu bilden. Auch außerhalb der Trivialfälle $a = 0$, $a = 1$ oder $b = 0$ ist dies aber im Allgemeinen nicht eindeutig lösbar. In $\mathbb{Z}_5$ gilt beispielsweise

$$\begin{aligned}
2^1 \bmod 5 &= 2, & 3^1 \bmod 5 &= 3, & 4^1 \bmod 5 &= 4,\\
2^2 \bmod 5 &= 4, & 3^2 \bmod 5 &= 4, & 4^2 \bmod 5 &= 1,\\
2^3 \bmod 5 &= 3, & 3^3 \bmod 5 &= 2, & 4^3 \bmod 5 &= 4,\\
2^4 \bmod 5 &= 1, & 3^4 \bmod 5 &= 1, & 4^4 \bmod 5 &= 1.
\end{aligned}$$

Die Gleichung $2^x \bmod 5 = b$ sowie die Gleichung $3^x \bmod 5 = b$ besitzt also für alle $b \in \mathbb{Z}_5 \setminus \{0\}$ eine eindeutige Lösung $x \in \mathbb{Z}_5 \setminus \{0\}$. Die Gleichung $4^x \bmod 5 = b$ besitzt hingegen für $b = 1$ und $b = 4$ jeweils zwei Lösungen und für $b = 2$ oder $b = 3$ keine Lösung in $\mathbb{Z}_5 \setminus \{0\}$. Auch der diskrete Logarithmus kann also nur unter Zusatzannahmen eindeutig gebildet werden.

2.3.3 Eulersche phi-Funktion und Satz von Euler-Fermat

Zur Einführung der RSA-Verschlüsselung benötigen wir noch den Satz von Euler-Fermat. Wir zeigen zunächst eine einfachere Version.

Satz 2.38 (Kleiner Satz von Fermat)
Sei $p \in \mathbb{N}$ eine Primzahl und $a \in \mathbb{Z}$ nicht durch p Teilbar, dann gilt

$$a^{p-1} \mod p = 1.$$

Beweis: Sei $\overline{a} := a \mod p \in \mathbb{Z}_p$. Dann ist $\overline{a} \neq 0$, da a nicht durch p teilbar ist. $\overline{a}$ hat deshalb ein multiplikatives Inverses $\overline{a}^{-1} \in \mathbb{Z}_p \setminus \{0\}$, d.h. $\overline{a}^{-1} \otimes \overline{a} = 1 = \overline{a} \otimes \overline{a}^{-1}$.

Wir definieren die Funktion

$$f : \mathbb{Z}_p \setminus \{0\} \to \mathbb{Z}_p \setminus \{0\}, \quad x \mapsto \overline{a} \otimes x.$$

Aus $f(x) = f(y)$ folgt $\overline{a} \otimes x = \overline{a} \otimes y$ und damit auch

$$\begin{aligned} x &= 1 \otimes x = (\overline{a}^{-1} \otimes \overline{a}) \otimes x = \overline{a}^{-1} \otimes (\overline{a} \otimes x) = \overline{a}^{-1} \otimes f(x) \\ &= \overline{a}^{-1} \otimes f(y) = \overline{a}^{-1} \otimes (\overline{a} \otimes y) = (\overline{a}^{-1} \otimes \overline{a}) \otimes y = 1 \otimes y = y. \end{aligned}$$

Die Funktion f ist also injektiv, und da $\mathbb{Z}_p \setminus \{0\}$ nur endlich viele Elemente enthält, ist f sogar bijektiv. Die $p-1$ Werte $f(1), \ldots, f(p-1)$ sind also genau die $p-1$ Zahlen $1, \ldots, p-1$ (nur eben in anderer Reihenfolge). Deshalb gilt

$$\begin{aligned} 1 \otimes 2 \otimes \ldots \otimes p-1 &= f(1) \otimes f(2) \otimes \ldots \otimes f(p-1) \\ &= (\overline{a} \otimes 1) \otimes (\overline{a} \otimes 2) \otimes \ldots \otimes (\overline{a} \otimes (p-1)) \\ &= \underbrace{(\overline{a} \otimes \ldots \otimes \overline{a})}_{p-1 \text{ mal}} \otimes 1 \otimes 2 \otimes \ldots \otimes p-1. \end{aligned}$$

Da $1, 2, \ldots, p-1$ alle multiplikative Inverse in $\mathbb{Z}_p$ besitzen können wir die Gleichung von rechts mit der Inversen von $p-1$, dann dem von $p-2$, usw., multiplizieren und erhalten schließlich

$$1 = \underbrace{(\overline{a} \otimes \ldots \otimes \overline{a})}_{p-1 \text{ mal}} = \overline{a}^{p-1} \mod p = a^{p-1} \mod p.$$

Damit ist die Behauptung bewiesen. □

Wir skizzieren nun, wie sich der kleine Satz von Fermat nun auf den Fall der Restbildung modulo m erweitern lässt, wobei m keine Primzahl sein muss. Im Beweis von Satz 2.38 war die Primzahleigenschaft von p entscheidend dafür, dass sowohl

$\overline{a} \in \mathbb{Z}_p$ als auch alle daran multiplizierte Zahlen $1, \ldots, p-1$ jeweils multiplikative Inverse besitzen. Zur Existenz der Inversen von $\overline{a}$ in $\mathbb{Z}_m$ genügt die Forderung $\mathrm{ggT}(a,m) = 1$. Die im Beweis definierte Funktion $f: x \mapsto \overline{a} \otimes x$ wenden wir dann nur auf diejenigen natürlichen Zahlen $1 \leq x \leq m$ an, die ebenfalls multiplikative Inverse besitzen, also auf die Menge

$$Z := \{x \in \{1, \ldots, m\} : \mathrm{ggT}(x,m) = 1\}.$$

Für jedes $x \in Z$ besitzt auch $f(x) = \overline{a} \otimes x$ ein multiplikatives Inverses (nämlich $x^{-1} \otimes \overline{a}^{-1}$. Es gilt also $f: Z \to Z$ und wie zuvor folgt, dass f weiterhin eine injektive und damit auch bijektive Abbildung ist, und damit die Zahlen aus Z lediglich in eine andere Reihenfolge bringt. Genau wie im Beweis von von Satz 2.38 folgt deshalb

$$\overline{a} \otimes \ldots \otimes \overline{a} = 1,$$

allerdings ist die Anzahl der Faktoren auf der linken Seite dieser Gleichung jetzt nicht mehr $p-1$, sondern nur noch die Anzahl $|Z|$ der Zahlen mit $1 \leq x \leq m$ mit $\mathrm{ggT}(x,m) = 1$ (also die Anzahl der zu m teilerfremden Zahlen $\leq m$). Wir geben dieser Anzahl einen Namen.

Definition 2.39
Die Eulersche Phi-Funktion $\varphi: \mathbb{N} \to \mathbb{N}$ *ist definiert durch*

$$\varphi(m) := |\{x \in \{1, \ldots, m\} : \mathrm{ggT}(x,m) = 1\}|.$$

Mit der obigen Beweisskizze zur Erweiterung des kleinen Satzes von Fermat erhalten wir den folgenden Satz.

Satz 2.40 (Satz von Euler-Fermat)
Sind $a, m \in \mathbb{N}$ *und* $\mathrm{ggT}(a,m) = 1$*, dann gilt*

$$a^{\varphi(m)} \bmod m = 1.$$

Beweis: Dies folgt mit dem oben skizzierten Anpassungen wie im Beweis von Satz 2.38. □

Wir haben am Ende von Abschnitt 2.2.1 schon gesehen, wie wir durch binäre modulare Exponentiation Ausdrücke der Form $a^n \bmod m$ mit $a, m, n \in \mathbb{N}$ effizient berechnen können. Dabei haben wir ausgenutzt, dass wir in einem Produkt schon vor der Multiplikation die Restbildung durchführen konnten (und damit kleinere Zahlen multiplizieren mussten). Der kleine Satz von Fermat und der allgemeinere Satz von Euler-Fermat erlauben es unter gewissen Bedingungen auch den Exponenten selbst schon durch Restbildung zu reduzieren:

Satz 2.41
Ist $p \in \mathbb{N}$ eine Primzahl, und $a \in \mathbb{N}$ eine Zahl, die nicht von p geteilt wird, dann gilt für alle $n \in \mathbb{N}$

$$a^n \bmod p = a^{n \bmod (p-1)} \bmod p$$

Allgemeiner gilt: Sind $a, m \in \mathbb{N}$ teilerfremd, dann gilt für alle $n \in \mathbb{N}$.

$$a^n \bmod m = a^{n \bmod \varphi(m)} \bmod m.$$

Beweis: Mit $q := n \text{ div } (p-1)$ und $r := n \bmod (p-1)$ ist $n = (p-1)q + r$ und damit

$$a^n = a^{(p-1)q} a^r = (a^{p-1})^q a^r.$$

Aufgrund des kleinen Satz von Fermat (Satz 2.38) ist

$$a^{p-1} \bmod p = 1$$

und damit

$$\begin{aligned} a^n \bmod p &= ((a^{p-1})^q \bmod p)(a^r \bmod p) \\ &= ((a^{p-1} \bmod p)^q \bmod p)(a^r \bmod p) = (1^q \bmod p)(a^r \bmod p) \\ &= a^r \bmod p \end{aligned}$$

und dies ist die erste behauptete Aussage.

Genauso folgt mit $\varphi(m)$ statt $p-1$ die allgemeinere zweite Aussage aus dem Satz von Euler-Fermat (Satz 2.40). □

Zur praktischen Anwendung fehlt uns noch eine Möglichkeit, die Eulersche Phi-Funktion effizient zu berechnen. Für unsere Zwecke genügt dazu das folgende Resultat:[3]

Satz 2.42
Für alle Primzahlen $p, q \in \mathbb{N}$, $p \neq q$, gilt.

(a) $\varphi(p) = p - 1$

(b) $\varphi(pq) = (p-1)(q-1) = \varphi(p)\varphi(q)$.

[3]Mit einem schwierigeren Beweis kann man zeigen, dass $\varphi(ab) = \varphi(a)\varphi(b)$ sogar für alle teilerfremden $a, b \in \mathbb{N}$ gilt.

Beweis: Zu einer Primzahl $p \in \mathbb{N}$ sind alle Zahlen $1, 2, \ldots, p-1$ teilerfremd. Daher gilt (a).

Für jedes $1 \leq z \leq pq$ gilt entweder $\mathrm{ggT}(z, pq) = 1$, $\mathrm{ggT}(z, pq) = p$, $\mathrm{ggT}(z, pq) = q$ oder $\mathrm{ggT}(z, pq) = pq$. Es gilt also

$$\begin{aligned}\{1, \ldots, pq\} = \{z \in \{1, \ldots, pq\} :\ \mathrm{ggT}(z, pq) = 1\} \cup \{z = kp,\ 1 \leq k \leq q-1\} \\ \cup \{z = kq,\ 1 \leq k \leq p-1\} \cup \{pq\}\end{aligned}$$

und die drei Mengen sind paarweise disjunkt. Daher ist

$$\begin{aligned}\varphi(pq) &= |\{z \in \{1, \ldots, pq\} :\ \mathrm{ggT}(z, pq) = 1\}| \\ &= |\{1, \ldots, pq\}| - |\{z = kp,\ 1 \leq k \leq q-1\}| \\ &\quad - |\{z = kq,\ 1 \leq k \leq p-1\}| - |\{pq\}| \\ &= pq - (q-1) - (p-1) - 1 = pq - q - p + 1 = (1-p)(1-q)\end{aligned}$$

und damit ist auch (b) gezeigt. □

2.4 Kryptographie

2.4.1 Das RSA-Verfahren

Der Empfänger wählt zwei (große und ungleiche) Primzahlen p und q und berechnet

$$N = pq.$$

Wegen Satz 2.42 gilt dann $\varphi(N) = (p-1)(q-1)$. Diese Zahl berechnet der Empfänger ebenfalls und wählt dann eine zu $\varphi(N)$ teilerfremde Zahl $e \in \mathbb{N}$ mit $1 < e < \varphi(N)$.

Zu e berechnet er das multiplikativ Inverse in $\mathbb{Z}_{\varphi(N)}$, also eine Zahl d mit

$$ed \mod \varphi(N) = 1.$$

Das Tupel (e, N) gibt er als öffentlichen Schlüssel bekannt. Das Tupel (d, N) bildet den privaten Schlüssel. d wird dazu geheim gehalten werden (und ebenfalls die dafür verwendeten Zwischengrößen p, q, $\varphi(N)$).

Ein Sender kann nun eine Nachricht $M < N$ mit dem öffentlichen Schlüssel verschlüsseln, indem das *Chiffrat*

$$C := M^e \mod N$$

berechnet und C an den Empfänger verschickt wird.

Der Empfänger kann aus dem empfangenen Chiffrat C die ursprüngliche Nachricht berechnen mit der Formel

$$M = C^d \bmod N.$$

Wir beweisen die Korrektheit dieser Entschlüsselungsformel in dem folgenden Satz.

Satz 2.43
Es gilt $M = C^d \bmod N$.

Beweis: Da $N = pq$ Produkt zweier Primzahlen ist und $M < N$ gilt entweder $\mathrm{ggT}(M,N) = 1$, $\mathrm{ggT}(M,N) = p$ oder $\mathrm{ggT}(M,N) = q$.

(a) $\mathrm{ggT}(M,N) = 1$: Dann gilt mit unserer Folgerung aus dem Satz von Euler-Fermat in Satz 2.41:

$$\begin{aligned} C^d \bmod N &= (M^e \bmod N)^d \bmod N = M^{ed} \bmod N \\ &= M^{ed \bmod \varphi(N)} \bmod N = M^1 \bmod N = M, \end{aligned}$$

wobei das letzte Gleichheitszeichen wegen $M < N$ gilt.

(b) $\mathrm{ggT}(M,N) = p$: Dann ist M ein Vielfaches von p, es gilt aber (wegen $M < N$) immer noch $\mathrm{ggT}(M,q) = 1$. Wir können Satz 2.41 deshalb zwar nicht auf Rechnungen modulo N anwenden, aber auf Rechnungen modulo q.

Mit $k := M^e \operatorname{div} N$ gilt

$$M^e = kN + (M^e \bmod N) = kpq + C \tag{2.15}$$

und daher ist $M^e \bmod q = C \bmod q$. Wir erhalten nun mit Satz 2.41

$$\begin{aligned} C^d \bmod q &= (C \bmod q)^d = (M^e \bmod q)^d = M^{ed} \bmod q \\ &= M^{ed \bmod \varphi(N)} \bmod q = M \bmod q. \end{aligned}$$

Dies zeigt $q|(C^d - M)$, also $C^d - M = lq$ mit einem $l \in \mathbb{Z}$. Aus $p|M$ folgt mit (2.15) auch $p|C$ und daher muss auch l ein Vielfaches von p sein, also $l = jp$ mit einem $j \in \mathbb{Z}$. Damit ist aber dann $C^d - M = jpq = jN$ und daher

$$C^d \bmod N = M \bmod N = M,$$

wobei das letzte Gleichheitszeichen wieder wegen $M < N$ gilt.

(c) $\mathrm{ggT}(M,N) = q$: folgt aus (b) durch Vertauschung von p und q. □

2.4.2 Ausblick: Aufwand und Sicherheit des RSA-Verfahren

Wir geben noch einen oberflächlichen Ausblick zum Aufwand und zur Sicherheit des RSA-Verfahrens und verweisen für tiefere Einblicke auf die Spezialvorlesungen zur Komplexitätstheorie und Kryptologie.

Zur Implementierung des RSA-Verfahrens müssen zunächst zwei große und ungleiche Primzahlen p und q gefunden werden sowie eine zur $(p-1)(q-1)$ teilerfremde Zahl e. Es sind sehr effiziente Algorithmen zur Erzeugung einer zufälligen Primzahl bekannt (z.B. basierend auf dem sogenannten Miller-Rabin-Test). Daher wird typischerweise zunächst e gewählt (in vielen Implementierungen ist $e := 65537 = 2^{16}+1$), und dann solange große zufällige Primzahlen generiert, bis die Teilerfremdheitsbedingung erfüllt ist. Die Zahl d kann als multiplikativ Inverses von e effizient mit dem erweiterten Euklidschen Algorithmus berechnet werden. Die Verschlüsselung (mit dem öffentlichen Schlüssel) und die Entschlüsselung (mit dem privaten Schlüssel) benötigen jeweils eine Potenzbildung $M^e \mod N$ bzw $C^d \mod N$. Diese kann sehr effizient wie am Ende von Abschnitt 2.2.1 berechnet werden.

Mögliche Ansätze zur Entschlüsselung ohne Kenntnis von d wären:

- Berechnung von M aus $C = M^e \mod N$ durch Bildung der e-ten diskreten Wurzel modulo N. Für die diskrete Wurzelbildung ist jedoch kein effizientes Verfahren bekannt.
- Verschlüsselung einer selbstgewählten Nachricht M und Berechnung von d aus $M = C^d \mod N$ durch Bildung des diskreten Logarithmus (zur Basis C und modulo N). Für die Bildung des diskreten Logarithmus ist jedoch kein effizientes Verfahren bekannt.
- Berechnung von $\varphi(N)$ aus N. Hierfür ist jedoch kein effizientes Verfahren bekannt.
- Berechnung der Primfaktorzerlegung von $N = pq$. Hierfür ist jedoch kein effizientes Verfahren bekannt.

Alle diese Angriffe lassen sich prinzipiell durch einfaches Durchprobieren aller natürlichen Zahlen bis N (für die Primfaktorzerlegung reicht sogar Durchprobieren bis $\sqrt{N}$) durchführen. Für eine k-Bit-Zahl sind dazu 2^k (bzw. $\sqrt{2^k} = 2^{k/2}$) Möglichkeiten zu prüfen. Diese Anzahl wächst also *exponentiell* mit der Bitlänge der eingegeben Zahl und wir haben dabei noch nicht einmal berücksichtigt, wie aufwändig die Prüfung dieser Zahl ist. Für die angegebenen vier Angriffsmöglichkeiten ist kein Algorithmus bekannt, der dies mit bezüglich der Bitlänge von

N polynomiellem Aufwand berechnen könnte. Die Sicherheit des RSA-Verfahren beruht daher auf der Tatsache, dass für hinreichend große verwendete Schlüssel die Schlüsselgenerierung, Ver- und Entschlüsselung (mit privatem Schlüssel) sehr effizient mit wenig Rechenleistung möglich ist, während die Entschlüsselung ohne privaten Schlüssel aufgrund des immensen Rechenaufwands nicht durchführbar ist.[4]

Die hier eingeführte einfache Version des RSA-Algorithmus (auch: Textbook-RSA genannt) hat allerdings den Nachteil, dass aus der gleichen Nachricht M immer das gleiche Chiffrat C generiert wird. Ein Angriffer kann sich die verschlüsselten Version vieler naheliegender Nachrichten (wie „ja", „nein", etc.) berechnen und mit dem empfangenen Chiffrat vergleichen. In der Praxis wird die RSA-Verschlüsselung deshalb mit sogenannten *padding*-Verfahren noch zusätzlich *randomisiert*.

[4]Für die Diskussion der möglicherweisen Lösung des Problems in polynomieller Zeit auf zukünftigen Quantencomputern mittels des Shor-Algorithmus verweisen wir auf Vorlesungen zur Quanteninformatik.

Kapitel 3

Lineare Algebra

3.1 Vektorrechnung

3.1.1 Vektorräume

Viele Anwendungen führen auf eine geordnete Zusammenstellung von Zahlen in Tupeln (vgl. Abschnitt 1.1.3)

$$\mathbb{R}^n := \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R} = \{(v_1, v_2, \ldots, v_n) : \ v_j \in \mathbb{R} \ \forall j = 1, \ldots, n\}.$$

Solche Tupel können auf natürliche Weise miteinander addiert werden (indem alle jeweiligen Einträge addiert werden) und sie können mit einzelnen Zahlen multipliziert werden (indem jeder Eintrag mit der einzelnen Zahl multipliziert wird).

Beispiel 3.1

(a) Deutschland holte 2016 in Rio de Janeiro $(17, 10, 15)$ *Gold-, Silber- und Bronzemedaillen und 2020 in Tokio* $(10, 11, 16)$ *Gold-, Silber- und Bronzemedaillen. Dies sind zusammen*

$$(10, 11, 16) + (17, 10, 15) = (10 + 17, 11 + 10, 16 + 15) = (27, 21, 31).$$

Würden wir anstreben, die Medaillenzahl aus Tokio zu verdoppeln, dann wären dafür nötig

$$2 * (10, 11, 16) = (2 * 10, 2 * 11, 2 * 16) = (20, 22, 32).$$

(b) Die Koeffizienten des quadratischen Polynoms $p(x) := 3x^2 + 2x + 1$ *lauten* $(3, 2, 1)$. *Für* $p(x) = 3x^2 + 2x + 1$ *und* $q(x) = 5x^2 + 3$ *ist*

$$p(x) + q(x) = 3x^2 + 2x + 1 + 5x^2 + 3 = 8x^2 + 2x + 4$$

und es ist

$$5p(x) = 5 * (3x^2 + 2x + 1) = 15x^2 + 10x + f.$$

Dies könnten wir auch direkt aus den Koeffizienten berechnen:

$$(3,2,1) + (5,0,3) = (8,2,4) \quad \textit{und} \quad 5 * (3,2,1) = (15,10,5).$$

(c) In der zweidimensionalen Ebene bezeichne das Tupel (a,b) einen Schritt um a nach rechts und b nach oben. Der Schritt $(1,2)$ gefolgt vom Schritt $(3,3)$ entspricht insgesamt dem Schritt

$$(1,2) + (3,3) = (4,5)$$

und dreimal den Schritt $(1,2)$ auszuführen, entspricht insgesamt dem Schritt

$$3 * (1,2) = (3,6).$$

Bezüglich der Addition gelten dabei die üblichen Rechenregeln einer kommutativen Gruppe (also Assoziativität, Existenz des neutralen und des inversen Elements, sowie Kommutatitivität). Die Multiplikation ist aber nicht wie in einem Ring für zwei Tupel, sondern für eine Tupel und einer einzelnen Zahl aus dem Körper $\mathbb{R}$ (im Folgenden auch Skalar genannt) definiert. Aber auch für die Multiplikation gelten natürliche Regeln wie Assoziativität, Distributivität und zur Multiplikation mit 1. Wir geben solchen Strukturen einen Namen:

Definition 3.2
Es sei K ein Körper. Ein Vektorraum $(V,+,*)$ besteht aus einer kommutativen Gruppe $(V,+)$ und einer skalaren Multiplikation[1]

$$* : K \times V \to V, \quad (a,v) = a * v \quad \textit{für alle } a \in K,\ v \in V,$$

so dass für das Einselement des Körpers $1 \in K$ und alle $a,b \in K$ und $v,w \in V$ gilt:

(V1) Assoziativität *a*(b*v)=(a*b)*v,*

(V2) Einselement *1*v=v,*

(V3) Distributivität I *a*(v+w)=(a*v)+(a*w),*

[1] Ein *Produkt* bezeichnet das Ergebnis einer *Multiplikation*. In Vektorräumen bezeichnen jedoch die Begriffe *skalare Multiplikation* und *Skalarprodukt* etwas Verschiedenes. Die hier eingeführte *skalare Multiplikation* verknüpft ein Skalar mit einem Vektor zu einem Vektor (das Skalar wird multipliziert). Das später (in Abschnitt 3.1.3) eingeführte *Skalarprodukt* verknüpft zwei Vektoren zu einem Skalar (d.h. das Produkt ist ein Skalar).

(V4) Distributivität II *(a+b)*v=(a*v)+(b*v).*

Wir definieren auch die Rechtsmultiplikation $v * a = a * v$ *und schreiben statt* $a * v$ *auch* $a \cdot v$ *oder* av *(und entsprechend* va *oder* $v \cdot a$*). Insbesondere gilt damit auch* $a * v * b = (ab)v$*. Entsprechend der üblichen Konvention „Punkt- vor Strichrechnung" lassen wir die Klammern auf der rechten Seite von (V3) und (V4) auch weg, also* $a(v+w) = av + aw$ *und* $(a+b)v = av + bv$*.*

Beispiel 3.3

(a) Sei K ein Körper. Wir definieren auf der Menge der n-Tupel K^n *die Addition zweier Tupel und die Multiplikation mit einem Skalar aus* $\mathbb{R}$ *wie folgt: Für alle* $(v_1, \ldots, v_n) \in K^n$*,* $(w_1, \ldots, w_n) \in K^n$ *und* $a \in \mathbb{R}$ *ist*

$$(v_1, \ldots, v_n) + (w_1, \ldots, w_n) := (v_1 + w_1, \ldots, v_n + w_n) \in K^n,$$
$$a * (v_1, \ldots, v_n) := (a * v_1, \ldots, a * v_n) \in K^n.$$

Man rechnet leicht nach, dass K^n *mit diesen Operationen ein Vektorraum ist. Das neutrale Element (Nullelement) der Addition ist dabei* $(0, \ldots, 0) \in K^n$ *und das additiv Inverse zu* $(v_1, \ldots, v_n) \in K^n$ *ist* $(-v_1, \ldots, -v_n) \in K^n$*.*

(b) Die Menge der quadratischen Polynome

$$\Pi_2 := \{p(x) = ax^2 + bx + c : \ a, b, c \in \mathbb{R}\}$$

bildet einen Vektorraum über dem Körper $\mathbb{R}$ *bezüglich der* punktweisen *Addition und skalaren Multiplikation*

$$(p+q)(x) := p(x) + q(x), \quad (a * p)(x) = a * (p(x)), \quad p, q \in \Pi_2,\ a \in \mathbb{R}.$$

Gleiches gilt für die Menge aller Polynome vom Grad $n \in \mathbb{N}$

$$\Pi_n := \{p(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 : \ a_1, \ldots, a_n \in \mathbb{R}\},$$

für die Menge aller Polynome beliebigen Grades Π *(auch:* $\mathbb{R}[x]$*) und sogar für die Menge aller Funktionen*

$$\mathscr{F} := \{f : \ \mathbb{R} \to \mathbb{R}\}.$$

(c) Eine Zeichenkette aus insgesamt $n \in \mathbb{N}$ *Nullen und Einsen können wir als Element des Vektorraums* $\mathbb{Z}_2^n$ *auffassen. Für die Addition gilt dann (vgl. Beispiel 2.37)*

$$(v_1, \ldots, v_n) + (w_1, \ldots, w_n) = (v_1 \text{ XOR } w_1, \ldots, v_n \text{ XOR } w_n)$$

und für die skalare Multiplikation gilt

$$a * (v_1, \ldots, v_n) = (a \text{ AND } v_1, \ldots, a \text{ AND } v_n),$$

also

$$0 * (v_1, \ldots, v_n) = (0, \ldots, 0) \quad \textit{und} \quad 1 * (v_1, \ldots, v_n) = (v_1, \ldots, v_n).$$

Die Elemente eines Vektorraums nennen wir Vektoren. Der Begriff beruht auf der geometrischen Vorstellung, ein Tupel $(v_1, v_2) \in \mathbb{R}^2$ (oder analog im $\mathbb{R}^3$) wie in Beispiel 3.1 als Schritt von einem Punkt der Ebene um v_1 nach rechts und v_2 nach oben aufzufassen. Dieser Schritt wird als Pfeil gemalt. Die Addition zweier Vektoren entspricht dann geometrisch der Hintereinanderausführung zweier solcher Schritte bzw. das Hintereinanderhängen zweier Pfeile und die skalare Multiplikation die entsprechende Verlängerung (mit Umorientierung bei Multiplikation mit negativen Elementen), vgl. die in der Vorlesung gemalten Skizzen. Der Begriff des Vektorraums wurde erst im 19. und 20. Jahrhundert in die Mathematik eingeführt, um geometrische Erkenntnisse und Vorstellungen auf andere Strukturen zu erweitern.

In den Anwendungen ist es üblich, Vektoren durch eine darüberliegenden Pfeil oder durch Fettdruck zu kennzeichnen, z.B. schreibt man

$$\vec{v} = (v_1, \ldots, v_n) \quad \text{oder} \quad \mathbf{v} = (v_1, \ldots, v_n).$$

Entsprechend schreibt man für das Nullelement im $\mathbb{R}^n$

$$\vec{0} = (0, \ldots, 0) \quad \text{oder} \quad \mathbf{0} = (0, \ldots, 0).$$

In der Mathematik ist so eine Kennzeichnung nicht üblich. Wir schreiben ohne weitere Kennzeichnung $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ und 0 bezeichnet sowohl das Tupel $0 = (0, \ldots, 0) \in \mathbb{R}^n$ als auch das Skalar $0 \in \mathbb{R}$. Mehr noch: Mit $v_1, v_2, \ldots, v_n \in K$ bezeichnen wir die Komponenten eines Vektors $v \in K^n$, wir verwenden die gleiche Notation aber auch zur Kennzeichnung mehrerer Vektoren, also z.B.

$$v_1 = (1,2,3) \in \mathbb{R}^3, \quad v_2 = (2,3,5) \in \mathbb{R}^3, \quad \ldots$$

Wenn wir sowohl mehrere Vektoren als auch ihre Komponenten bezeichnen wollen, dann stellen wir den Index auch nach oben und setzen ihn zur Unterscheidung von einer Potenz in Klammern. Wir würden also beispielsweise drei Vektoren im $\mathbb{R}^2$ bezeichnen mit $v^{(1)}, v^{(2)}, v^{(3)} \in \mathbb{R}^2$ und sie komponentenweise schreiben als

$$v^{(1)} = (v_1^{(1)}, v_2^{(1)}) \in \mathbb{R}^2, \quad v^{(2)} = (v_1^{(2)}, v_2^{(2)}) \in \mathbb{R}^2, \quad v^{(3)} = (v_1^{(3)}, v_2^{(3)}) \in \mathbb{R}^2.$$

Aus (V1)–(V4) ergeben sich weitere naheliegende Rechenregeln in Vektorräumen:

Satz 3.4
Es sei V ein Vektorraum über einem Körper K. Dann gilt für alle $v \in V$ und $a \in K$

$$a * 0 = 0$$
$$0 * v = 0$$
$$(-1) * v = -v$$

Beweis: Einfaches Nachrechnen. □

3.1.2 Basis und Dimension

Wir betrachten im Folgenden nur noch Vektorräume über dem Körper $\mathbb{R}$ (sogenannte *reelle* Vektorräume). Alle Ergebnisse gelten aber analog auch über dem in den späteren Vorlesungen eingeführten wichtigen Körper der komplexen Zahlen $\mathbb{C}$. Außerdem schreiben wir die Einträge von Tupeln $v = (v_j)_{j=1}^n \in \mathbb{R}^n$ ab sofort untereinander, also

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n \quad \text{statt} \quad v = (v_1, \ldots, v_n) \in \mathbb{R}^n.$$

Diese Schreibweise ist sperriger, ermöglicht aber jetzt schon die einfachere schematische Durchführung der Addition und skalaren Multiplikation

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix}, \qquad a \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} av_1 \\ av_2 \\ \vdots \\ av_n \end{pmatrix}$$

und ist noch vorteilhafter im Zusammenhang mit der in Abschnitt 3.2 eingeführten Matrizenrechnung.

Definition 3.5
Sei V ein reeller Vektorraum. Eine Teilmenge $U \subseteq V$ heißt Unterraum (auch: Untervektorraum), falls U bezüglich der Addition und skalaren Multiplikation aus V selbst ein Vektorraum ist.

Da alle Elemente aus U auch Elemente von V sind, ist die Addition und skalare Multiplikation auch auf den Elementen von U definiert und erfüllt auch die gleichen Regelnregeln wie in V. Trotzdem ist nicht jede Teilmenge $U \subseteq V$ auch ein

Unterraum, denn die Addition und skalare Multiplikation von Elementen aus U könnten ein Ergebnis außerhalb U liefern oder U könnte nicht das neutrale Element der Addition $0 \in V$ oder das zu einem Vektor $u \in U$ additiv Inverse $-u$ enthalten.

Satz 3.6
Sei V ein reeller Vektorraum und $\emptyset \neq U \subseteq V$. U ist genau dann ein Unterraum von V, wenn für alle $u_1, u_2 \in U$ und $a \in \mathbb{R}$ gilt

$$u_1 + u_2 \in U \quad \text{und} \quad au_1 \in U. \tag{3.1}$$

Beweis: Ist U ein Unterraum, dann bildet die Addition und die skalare Multiplikation nach U ab, es muss also (3.1) gelten.

Gelte nun umgekehrt (3.1). Bezüglich der Addition gilt dann die Assoziativität (G1) und die Kommutativität (G4), da sie in V gelten. Für jedes $u \in U$ ist wegen (3.1) auch $0 = 0u \in U$ und $-u = (-1)u \in U$, also enthält U auch die neutralen Elemente und die inversen Elemente der Addition, so dass auch (G2) und (G3) erfüllt sind. $(U, +)$ ist also eine kommutative Gruppe. Die zusätzlichen Vektorraumaxiome (V1)–(V4) aus Definition 3.2 gelten wiederum, da sie in V gelten. □

Da wir nach Satz 3.4 das neutrale Element durch Multiplikation mit $0 \in \mathbb{R}$ und das Inverse durch Multiplikation mit $-1 \in \mathbb{R}$ erhalten können, ist $U \subseteq V$ genau dann ein Unterraum, wenn für alle $u_1, u_2 \in U$, $a \in \mathbb{R}$ gilt

Beispiel 3.7
(a) Betrachte die Mengen

$$U_1 := \left\{ \begin{pmatrix} v_1 \\ v_2 \\ 0 \end{pmatrix} : v_1, v_2 \in \mathbb{R} \right\} \subseteq \mathbb{R}^3 \text{ und } U_2 := \left\{ \begin{pmatrix} v_1 \\ 2 \\ 0 \end{pmatrix} : v_1 \in \mathbb{R} \right\} \subseteq \mathbb{R}^3.$$

U_1 bildet einen Unterraum von $\mathbb{R}^3$. U_2 bildet keinen Unterraum von $\mathbb{R}^3$.

(b) Die quadratischen Polynome Π_2 bilden einen Unterraum der kubischen Polynome Π_3, diese bilden für $n \geq 3$ einen Unterraum von Π_n, dieser einen Unterraum von Π und dieser einen Unterraum des Vektorraums aller reller Funktionen $\mathscr{F}$.

Wir können eine Menge von Vektoren um ihre Summen und skalare Vielfachen ergänzen, um so einen Unterraum zu erhalten.

Definition 3.8
Sei V ein reeller Vektorraum und $v_1, \ldots, v_n \in V$. Ein Vektor der Form

$$v = \sum_{j=1}^{n} a_j v_j = a_1 v_1 + \ldots + a_n v_n \quad \text{mit } a_1, \ldots, a_n \in \mathbb{R}$$

heißt Linearkombination der Vektoren $v_1, \ldots, v_n$. Die Menge aller solcher Linearkombinationen

$$\operatorname{span}\{v_1, \ldots, v_n\} := \{v = a_1 v_1 + \ldots + a_n v_n : \ a_1, \ldots, a_n \in \mathbb{R}\}$$

heißt Erzeugnis (engl.: span*) der Vektoren $v_1, \ldots, v_n$. Anstelle von* $\operatorname{span}\{v_1, \ldots, v_n\}$ *schreibt man auch $\langle v_1, \ldots, v_n \rangle$.*

Man überzeugt sich leicht, dass $U := \operatorname{span}\{v_1, \ldots, v_n\} \subseteq V$ ein Unterraum von V ist und zwar der kleinste, der alle Vektoren $v_1, \ldots, v_n$ enthält. Die Vektoren $\{v_1, \ldots, v_n\}$ heißen auch Erzeugendensystem dieses Unterraums. Die Linearkombination $0 = 0v_1 + \ldots + 0v_n$ nennen wir auch *triviale Linearkombination*.

Beispiel 3.9
(a) Jeden Vektor $v \in \mathbb{R}^3$ können wir schreiben als

$$v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Daher ist

$$\operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\} = \mathbb{R}^3$$

(b) Wir können jeden Vektor $v \in \mathbb{R}^3$ auch schreiben als

$$v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = v_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (v_2 - v_1) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + (v_3 - v_1) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 5 \\ 7 \\ 2 \end{pmatrix}.$$

Daher ist auch

$$\operatorname{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 7 \\ 2 \end{pmatrix} \right\} = \mathbb{R}^3$$

(c) *Für U_1 aus Beispiel 3.7 gilt*

$$U_1 = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\} \subseteq \mathbb{R}^3.$$

(d) *Jedes quadratische Polynom $p(x) = ax^2 + bx + c \in \Pi_2$ ist eine Linearkombination der* Monome $1 \in \Pi_2$, $x \in \Pi_2$ *und* $x^2 \in \Pi_2$. *Also ist* $\Pi_2 = \operatorname{span}\{1, x, x^2\}$.

In Beispiel 3.9(b) enthielt das Erzeugendensystem einen unnötigen vierten Vektor, der auch schon in der Linearkombination der anderen enthalten war. Wir geben dieser Eigenschaft einen Namen:

Definition 3.10
Sei V ein reeller Vektorraum. Eine Menge von Vektoren $v_1, v_2, \ldots, v_n \in V$ heißt linear unabhängig, falls nur die triviale Linearkombination den Nullvektor erzeugt, d.h. für alle $a_1, a_2, \ldots, a_n \in \mathbb{R}$ gilt

$$a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0 \quad \Longrightarrow \quad a_1 = a_2 = \ldots = a_n = 0$$

Ansonsten heißen die Vektoren linear abhängig.

Beachte, dass für linear abhängige Vektoren $v_1, \ldots, v_n \in V$ eine Linearkombination existiert mit

$$a_1 v_1 + a_2 v_2 + \ldots + a_n v_n = 0,$$

wobei aber mindestens einer der Koeffizienten a_1, a_2,..., a_n nicht Null ist. Sei dies der j-te, also $a_j \neq 0$, dann ist

$$v_j = \frac{1}{a_j} \sum_{\substack{k=1 \\ k \neq j}}^{n} a_k v_k.$$

Der Vektor v_j kann also als Linearkombination der anderen $n-1$ Vektoren dargestellt werden und es gilt

$$\operatorname{span}\{v_1, \ldots, v_n\} = \operatorname{span}\left(\{v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots v_n\} \setminus \{v_j\}\right).$$

Umgekehrt ist eine Menge von Vektoren, bei der ein Vektor Linearkombination der anderen ist, offenbar linear abhängig. Lineare unabhängige Vektoren bilden daher ein *minimales Erzeugendensystem*, bei dem kein Vektor weggelassen werden kann ohne den erzeugten Unterraum zu verändern. Mehr noch: Jeder Vektor

im Erzeugnis linear unabhängiger Vektoren kann nur auf genau eine Art und Weise als Lienarkombination der linear unabhängigen Vektoren dargestellt werden, denn aus

$$v = a_1 v_1 + \ldots + a_n v_n = b_1 v_1 + \ldots b_n v_n$$

folgt

$$(a_1 - b_1)v_1 + \ldots + (a_n - b_n)v_n = 0$$

und damit $a_1 = b_1, \ldots a_n = b_n$.

Man zeigt leicht, dass die Erzeugendensysteme in Beispiel 3.9(a),(b) und (d) linear unabhängig sind.

Definition 3.11
Sei V ein reeller Vektorraum und $v_1, \ldots, v_n \in V$. Ist $\{v_1, \ldots, v_n\}$ linear unabhängig und $\operatorname{span}\{v_1, \ldots, v_n\} = V$*, dann heißt $\{v_1, \ldots, v_n\}$ eine (endliche) Basis von V.*

Oft ist auch die Reihenfolge der Basiselemente wichtig, in dem Fall bezeichnet man eine Basis statt mit $\{v_1, \ldots, v_n\} \subseteq V$ auch mit $(v_1, \ldots, v_n) \in V \times V \times \cdot \times V$.

Für die Menge der Polynome vom Grad $n \in \mathbb{N}$ ist die Menge der Monome

$$\{1, x, x^2, \ldots, x^n\} \subseteq \Pi_n$$

offenbar ein Erzeugendensystem. Man kann zeigen, dass dieses auch linear unabhängig ist und deshalb eine Basis von Π_n ist. Der Vektorraum Π der Polynome beliebigen Grades besitzt keine endliche Basis. Man kann die Begriffe Erzeugnis und lineare Unabhängigkeit auch auf unendliche Mengen erweitern, indem das Erzeugnis alle endlichen Linearkombinationen umfasst und lineare Unabhängigkeit bedeutet, dass der Nullvektor nicht als nicht-triviale endliche Linearkombination geschrieben werden kann. Damit lässt sich dann zeigen, dass jeder Vektorraum eine (endliche oder unendliche) Basis besitzt, im Falle der Polynome belieben Grades wäre das beispielsweise die unendliche Menge aller Monome beliebig hohen Grades

$$\{x^n : \ n \in \mathbb{N}_0\} \subseteq \Pi.$$

Diese Form der Behandlung unendlicher Mengen als Basis (mit endlichen Linearkombination) nennt man auch *Hamelbasis*. Daneben gibt es auch Definitionen, die unendliche Linearkombination zulassen (z.B. die sogenannte *Schauderbasis*).

Wir werden in dieser Vorlesung nur Vektorräume mit endlicher Basis betrachten. Ein wichtiger Satz zeigt, dass es zwar mehrere mögliche Basen gibt, die Anzahl der Basiselemente aber immer gleich ist.

Definition und Satz 3.12
Sei $\mathscr{B} := \{v_1, \ldots, v_n\} \subseteq V$, $n \in \mathbb{N}$, eine Basis eines reellen Vektorraums V. Dann hat auch jede andere Basis von V genau n Elemente und jede linear unabhängige Menge von n Vektoren ist eine Basis. n heißt die Dimension des Vektorraums V. Wir schreiben $\dim V = n$ und nennen Vektorräume mit endlicher Basis auch endlich-dimensional (kurz: $\dim V < \infty$).

Beweis: Wir skizieren nur die grundlegende Idee. Gegeben zwei Basen von V kann man zeigen (Austauschlemma von Steinitz), dass ein Basiselement der ersten Basis gegen eines der zweiten ausgetauscht werden kann ohne dass die Basiseigenschaft verlorengeht. Würde eine Basis weniger Elemente als die andere enthalten, so könnten wir in der ersten Basis nacheinander alle Elemente gegen solche aus der zweiten Basis austauschen und hätten damit eine Basis die aus einer Teilmenge der zweiten Basis bestünde. Alle verbleibenden Elemente der zweiten Basis müssten sich dann aber als Linearkombination dieser Teilmenge schreiben lassen, was der linearen Unabhängigkeit widerspricht. Die beiden Basen müssen daher die gleiche Anzahl von Elementen haben. □

Bemerkung 3.13
Sei V ein endlich-dimensionaler reeller Vektorraum und $n = \dim V$. Mit der gleichen Beweisenideen wie in Satz 3.12 kann man auch zeigen:

(a) Jedes linear unabhängige Teilmenge von V besitzt höchstens n Elemente. Jede linear unabhängige Teilmenge mit genau n Elementen ist auch ein Erzeugendensystem und damit eine Basis von V.

(b) Jedes Erzeugendensystem von V besitzt mindestens n Elemente. Jedes Erzeugendensystem mit genau n Elementen ist auch linear unabhängig und damit eine Basis von V.

*(c) Jeder Unterraum U von V ist endlich-dimensional mit $\dim U \leq \dim V$. Jede Basis $\{u_1, \ldots, u_k\} \subseteq U$ des Unterraums U, $k := \dim U$, kann durch $n-k$ Elemente $v_{k+1}, \ldots, v_n \in V$ zu einer Basis $\{u_1, \ldots, u_k, v_{k+1}, \ldots, v_n\}$ des Vektorraums V ergänzt werden (*Basisergänzungssatz*). Falls $\dim U = \dim V$, so ist $U = V$.*

Beispiel 3.14
(a) Offenbar besitzt der Raum $\mathbb{R}^n$ die Dimension n und eine Basis des $\mathbb{R}^n$ bilden die Einheitsvektoren

$$e_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 := \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, e_n := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}. \tag{3.2}$$

(b) Der Raum Π_n besitzt Dimension $n+1$ und die Monome $\{1, x, x^2, \ldots, x^n\}$ bilden eine Basis des Π_n.

(c) Für einen Vektorraum V betrachten wir die einelementige Menge $\{0\}$, die nur den Nullvektor $0 \in V$ enthält. Diese Menge erfüllt ebenfalls die Vektorraumaxiome, sie bildet also einen Unterraum von V. Wir nennen $\{0\}$ den Nullvektorraum, auch hierfür ist die Abkürzung $0 = \{0\}$ üblich.

Beispiel 3.15 (Quadratische Interpolation)
Wir betrachten das Problem zu drei vorgegebenen Punkten

$$(x_1, y_1), (x_2, y_2), (x_3, y_3) \in \mathbb{R}^2$$

mit paarweisen verschiedenen $x_1, x_2, x_3 \in \mathbb{R}$ ein quadratisches Polynom (Parabel) zu finden, das durch diese drei Punkte verläuft (man sagt auch: diese Punkte interpoliert*).*

*Wir definieren die folgenden drei Polynome (*Lagrange-Grundpolynome*):*

$$l_1(x) := \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} \in \Pi_2,$$
$$l_2(x) := \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} \in \Pi_2,$$
$$l_3(x) := \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} \in \Pi_2.$$

Diese erfüllen offenbar

$$\begin{aligned} l_1(x_1) &= 1, \quad l_1(x_2) = 0, \quad l_1(x_3) = 0, \\ l_2(x_1) &= 0, \quad l_2(x_2) = 1, \quad l_2(x_3) = 0, \\ l_3(x_1) &= 0, \quad l_3(x_2) = 0, \quad l_1(x_3) = 1. \end{aligned}$$

Das Polynom

$$p(x) = y_1 l_1(x) + y_2 l_2(x) + y_3 l_3(x)$$

besitzt daher die gewünschte Eigenschaft

$$\begin{aligned} p(x_1) &= y_1 * 1 + y_2 * 0 + y_3 * 0 = y_1, \\ p(x_2) &= y_1 * 0 + y_2 * 1 + y_3 * 0 = y_2, \\ p(x_3) &= y_1 * 0 + y_2 * 0 + y_3 * 1 = y_3. \end{aligned}$$

Mehr noch: $\{l_1, l_2, l_3\}$ ist eine Basis des Π_2 und jedes Polynom $p \in \Pi_2$ lässt sich schreiben als Linearkombination

$$p(x) = a_1 l_1(x) + a_2 l_2(x) + a_3 l_3(x) \quad \text{mit } a_1 = p(x_1),\ a_2 = p(x_2),\ a_3 = p(x_3).$$

3.1.3 Normen und Skalarprodukte

Für einen Vektor $v \in \mathbb{R}^n$ heißt

$$\|v\|_2 := \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2} \in \mathbb{R}$$

die euklidische Norm des Vektors. Im $\mathbb{R}^2$ und $\mathbb{R}^3$ entspricht dies nach dem Satz des Pythagoras gerade der Länge der zugehörige Strecke.

Die euklidische Norm erfüllt die folgenden Eigenschaften:

- Definitheit: $\|v\|_2 \geq 0$ für alle $v \in \mathbb{R}^n$ und
$$\|v\|_2 = 0 \quad \Longleftrightarrow \quad v = 0.$$
- Homogenität: Für alle $v \in \mathbb{R}^n$ und $r \in \mathbb{R}$ gilt
$$\|rv\|_2 = |r| \, \|v\|_2 .$$
- Dreiecksungleichung: Für alle $v, w \in \mathbb{R}^n$ gilt
$$\|v + w\|_2 \leq \|v\|_2 + \|w\|_2 .$$

Allgemein nennen wir jede Abbildung $\|\cdot\| : V \to \mathbb{R}$ auf einem reellen Vektorraum V, die diese drei Eigenschaften erfüllt eine Norm. Wichtige weitere Normen auf dem $\mathbb{R}^n$ sind:

- Die Maximumsnorm: $\|v\|_\infty := \max_{i=1,\ldots,n} |v_i|$.
- Die Summennorm: $\|v\|_1 := |v_1| + |v_2| + \ldots + |v_n|$.
- Die p-Norm: $\|v\|_p := \sqrt[p]{v_1^p + v_2^p + \ldots + v_n^p}$ mit $1 \leq p < \infty$.

Das Euklidische Skalarprodukt (auch: Standardskalarprodukt) zweier Vektoren $v, w \in \mathbb{R}^n$ ist definiert durch

$$\langle v, w \rangle := v_1 w_1 + v_2 w_2 + \ldots + v_n w_n .$$

Statt $\langle v, w \rangle$ schreibt man auch $v \cdot w$ oder $v^T w$ (siehe Abschnitt 3.2). Geometrisch gilt

$$\langle v, w \rangle = \|v\|_2 \, \|w\|_2 \cos \alpha$$

wobei α der Winkel zwischen den Vektoren v und w ist.

Das euklidische Skalarprodukt erfüllt die Eigenschaften:

- Linearität in jeder Komponente: Für alle $u, v, w \in \mathbb{R}^n$, $r \in \mathbb{R}$ gilt

$$\begin{aligned} \langle u+v, w\rangle &= \langle u, w\rangle + \langle v, w\rangle, \qquad & \langle ru, v\rangle &= r\langle u, v\rangle, \\ \langle u, v+w\rangle &= \langle u, v\rangle + \langle u, w\rangle, \qquad & \langle u, rv\rangle &= r\langle u, v\rangle. \end{aligned}$$

- Symmetrie:

$$\langle u, v\rangle = \langle v, u\rangle \quad \text{für alle } u, v \in \mathbb{R}^n.$$

- Positive Definitheit: Für alle $v \in \mathbb{R}^n$ ist $\langle v, v\rangle \geq 0$ und

$$\langle v, v\rangle = 0 \quad \Longleftrightarrow \quad v = 0.$$

Allgemein nennen wir jede Abbildung $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ auf einem reellen Vektorraum V, die diese drei Eigenschaften erfüllt ein Skalarprodukt.

Offenbar gilt

$$\|v\|_2 = \sqrt{\langle v, v\rangle} \quad \text{für alle } v \in \mathbb{R}^n$$

und genauso definiert auch im allgemeinen Fall ein Skalarprodukt auf einem reellen Vektorraum immer eine Norm (induzierte Norm). Es ist aber nicht jede Norm von einem Skalarprodukt induziert, z.B. existiert zu $\|\cdot\|_p$ für $p \neq 2$ kein zugehöriges Skalarprodukt.

Zwei Vektoren $v, w \in V$ mit $\langle v, w\rangle = 0$ heißen orthogonal. Die Koeffizienten einer Linearkombination orthogonaler Vektoren lassen sich durch Orthogonalprojektion besonders einfach bestimmen. Sind nämlich $v_1, \ldots, v_k \in V$ paarweise orthogonale Vektoren in einem reellen Vektorraum mit Skalarprodukt $\langle \cdot, \cdot \rangle$ und davon induzierter Norm $\|\cdot\|$ und ist $w \in V$ eine Linearkombination dieser Vektoren

$$w = a_1 v_1 + a_2 v_2 + \ldots + a_k v_k \in V,$$

dann gilt

$$\begin{aligned} \langle v_j, w\rangle &= \langle v_j, a_1 v_1 + a_2 v_2 + \ldots + a_k v_k\rangle \\ &= a_1\langle v_j, v_1\rangle + a_2\langle v_j, v_2\rangle + \ldots + a_k\langle v_j, v_k\rangle = a_j\langle v_j, v_j\rangle = a_j \left\|v_j\right\|^2. \end{aligned}$$

Das zeigt insbesondere auch, dass paarweise orthogonale Vektoren immer linear unabhängig sind. Eine Basis $v_1, \ldots, v_n$ von V mit paarweise orthogonalen und normierten Vektoren, d.h.

$$\langle v_j, v_k\rangle = \begin{cases} 0 & \text{für } j \neq k, \\ 1 & \text{für } j = k, \end{cases}$$

heißt Orthonormalbasis (ONB) von V.

3.2 Lineare Abbildungen und Matrizen

3.2.1 Der Vektorraum der Matrizen

Eine Matrix $A \in \mathbb{R}^{m\times n}$ ist eine tabellarische Anordnung von insgesamt mn reellen Zahlen in m Zeilen und n Spalten ($m,n \in \mathbb{N}$). Wir bezeichnen den Eintrag in der i-ten Zeile und j-ten Spalte mit $a_{i,j} \in \mathbb{R}$ und schreiben

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} = (a_{i,j})_{\substack{i=1,\dots,m,\\ j=1,\dots,n.}}$$

Oft lassen wir das Komma im Index auch weg und schreiben a_{ij} statt $a_{i,j}$ und z.B. a_{12} statt $a_{1,2}$. Statt

$$A = (a_{i,j})_{\substack{i=1,\dots,m,\\ j=1,\dots,n}} \in \mathbb{R}^{m\times n}$$

schreiben wir auch kurz

$$A = (a_{ij}) \in \mathbb{R}^{m\times n}.$$

Für $m = n$ nennen wir die Matrix quadratisch.

Definition 3.16
Für $A \in \mathbb{R}^{m\times n}$ definieren wir die transponierte Matrix $A^T \in \mathbb{R}^{n\times m}$ durch Vertauschung der Zeilen und Spalten, d.h. der Eintrag in der i-ten Zeile und j-Spalte von A^T ($i = 1,\dots,n,\ j = 1,\dots,m$) ist also a_{ji} und es ist

$$A^T = \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{m,n} \end{pmatrix}.$$

Definition 3.17 (Matrix-Vektor-Produkt)
Sei $x = (x_j)_{j=1}^n \in \mathbb{R}^n$ und $A = (a_{ij})_{\substack{i=1,\dots,m\\ j=1,\dots,n}} \in \mathbb{R}^{m\times n}$ (die Anzahl der Spalten von A stimmt also mit der Anzahl der Einträge in x überein). Dann definieren wir das Matrix-Vektor-Produkt

$$Ax = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix}$$

Der i-te Eintrag von $Ax \in \mathbb{R}^m$ ist also gegeben durch $\sum_{j=1}^n a_{ij}x_j$.

Vektoren können wir als Matrizen mit einer Spalte auffassen. Entsprechend gilt für $x, y \in \mathbb{R}^n$

$$x^T y = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = \langle x, y \rangle.$$

und $\|x\|_2^2 = x^T x$.

Wegen des in Abschnitt 3.2.3 diskutierten Zusammenhangs zwischen Matrizen und linearen Abbildungen nennt man das Matrix-Vektor-Produkt auch *Anwendung* einer Matrix auf einen Vektor.

Wir betonen nochmal, dass wir dabei vorausgesetzt haben, dass die Anzahl der Spalten von A mit der Anzahl der Einträge in x übereinstimmt. Für $A \in \mathbb{R}^{m,r}$ und $x \in \mathbb{R}^n$ mit $r \neq n$ ist das Matrix-Vektor-Produkt Ax nicht definiert!

Beispiel 3.18

(a) Wir betrachten ein lineares Gleichungssystem (LGS) mit drei Unbekannten. Suche $x_1, x_2, x_3 \in \mathbb{R}$, so dass

$$\begin{aligned} x_1 + 2x_2 + 5x_3 &= 8, \\ 3x_1 - x_2 + 2x_3 &= 1, \\ x_2 - x_3 &= 0. \end{aligned}$$

In Matrix-Vektor-Schreibweise lautet das LGS: Finde $x = (x_j)_{j=1}^3 \in \mathbb{R}^3$, so dass

$$\begin{pmatrix} 1 & 2 & 5 \\ 3 & -1 & 2 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 1 \\ 0 \end{pmatrix}.$$

(b) Die Spiegelung eines Punktes $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ in der zweidimensionalen Ebene an der x_1-Ebene ergibt den Punkt

$$\begin{pmatrix} x_1 \\ -x_2 \end{pmatrix} = \begin{pmatrix} 1x_1 + 0x_2 \\ 0x_1 + (-1)x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Die folgende Eigenschaft nennen wir auch *Linearität des Matrixanwendung*. Wir werden in Abschnitt 3.2.3 sehen, dass wir jede lineare Abbildung durch eine Matrix beschreiben können.

Satz 3.19 (Linearität der Matrixanwendung)
Für alle $\lambda \in \mathbb{R}$, $x, y \in \mathbb{R}^n$ und $A \in \mathbb{R}^{m\times n}$ gilt:

(a) Additivität*:* $A(x+y) = Ax + Ay$,

(b) Homogenität*:* $\lambda(Ax) = A(\lambda x)$.

Beweis: Leichtes Nachrechnen. □

Bemerkung 3.20
Für die Einheitsvektoren $e_1, \ldots, e_n \in \mathbb{R}^n$ aus (3.2) *gilt*

$$Ae_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad Ae_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \quad Ae_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix},$$

Ae_j ergibt also gerade die j-te Zeile der Matrix und A ist durch Kenntnis von Ae_1, $Ae_2, \ldots, Ae_n$ bereits eindeutig bestimmt.

Definition 3.21
Für zwei Matrizen gleicher Dimension $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, $B = (a_{ij}) \in \mathbb{R}^{m\times n}$ definieren wir die Summe durch

$$A + B := \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{pmatrix} \in \mathbb{R}^{m\times n}.$$

Für eine Matrix $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ und eine reelle Zahl $\lambda \in \mathbb{R}$ definieren wir außerdem die skalare Multiplikation

$$\lambda A := \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{pmatrix} \in \mathbb{R}^{m\times n}.$$

Satz 3.22 (Vektorraum der Matrizen)
$\mathbb{R}^{m\times n}$ bildet mit dieser Addition und skalare Multiplikation einen Vektorraum. $(\mathbb{R}^{m\times n}, +)$ ist also eine kommutative Gruppe:

(G1) Assoziativität:

$$A + (B + C) = (A + B) + C \quad \text{für alle } A, B, C \in \mathbb{R}^{m\times n}.$$

(G2) Existenz des neutralen Elements:

$$\textit{Mit} \quad 0 := \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \quad \textit{gilt} \quad A+0=0+A=A \quad \forall A \in \mathbb{R}^{m\times n}.$$

(G3) Existenz inverser Elemente: Sei $A \in \mathbb{R}^{m\times n}$.

$$-A := \begin{pmatrix} -a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{m1} & \cdots & -a_{mn} \end{pmatrix} \quad \textit{erfüllt} \quad (-A)+A=A+(-A)=0.$$

(G4) Kommutativität:

$$A+B=B+A \quad \textit{für alle } A,B \in \mathbb{R}^{m\times n},$$

un die sklare Multiplikation erfüllt

(V1) Assoziativität:

$$\lambda(\mu A)=(\lambda\mu)A \quad \textit{für alle } \lambda,\mu \in \mathbb{R},\ A \in \mathbb{R}^{m\times n},$$

(V2) Einselement: Für die skalare Eins $1 \in \mathbb{R}$ *gilt*

$$1*A=A \quad \textit{für alle } A \in \mathbb{R}^{m\times n},$$

(V3) Distributivität I:

$$\lambda(A+B)=(\lambda A)+(\lambda B) \quad \textit{für alle } \lambda \in \mathbb{R},\ A,B \in \mathbb{R}^{m\times n},$$

(V4) Distributivität II:

$$(\lambda+\mu)A=(\lambda A)+(\mu A) \quad \textit{für alle } \lambda,\mu \in \mathbb{R},\ A \in \mathbb{R}^{m\times n}.$$

Beweis: Einfaches Nachrechnen. □

Wie üblich definieren wir auch die skalare Multiplikation von rechts durch

$$A\lambda := \lambda A \quad \text{für alle } \lambda \in \mathbb{R},\ A \in \mathbb{R}^{m\times n},$$

schreiben die Addition des additiv Inversen kurz als

$$A-B := A+(-B) \quad \text{für alle } A,B \in \mathbb{R}^{m\times n}$$

und folgen der Konvention „Punkt- vor Strichrechnung", also

$$(\lambda A)+(\lambda B)=\lambda A+\lambda B \quad \text{und} \quad (\lambda A)+(\mu A)=\lambda A+\mu A$$

(für alle $\lambda,\mu\in\mathbb{R}$, $A,B\in\mathbb{R}^{m\times n}$.

Zusätzlich zu (G1)–(G4), (V1)–(V4) gelten die folgenden Verträglichkeitseigenschaften für das Matrix-Vektor-Produkt.

Satz 3.23
Für alle $\lambda\in\mathbb{R}$, $v,w\in\mathbb{R}^n$ und $A,B\in\mathbb{R}^{m\times n}$ gilt:

(a) $(A+B)v=Av+Bv$,

(b) $A(v+w)=Av+Aw$,

(c) $(\lambda A)v=A(\lambda v)$.

Beweis: Einfaches Nachrechnen. □

3.2.2 Die Multiplikation von Matrizen

Bemerkung 3.24 (Motivation der Matrixmultiplikation)
Wir betrachten nun den Fall, dass wir eine Matrix A auf einen Vektor $y=Bx$ anwenden, der wiederum aus einem Matrix-Vektorprodukt entstanden ist, berechnen also

$$Ay=A(Bx).$$

Wir werden sehen, dass wir dies auch in eine Matrix C zusammenfasen können, also eine Matrix C finden, so dass

$$Cx=A(Bx).$$

Diese Matrix werden wir das Produkt von A und B nennen, $C=AB$. Mit der so definierten Matrixmultiplikation gilt dann also

$$(AB)x=A(Bx).$$

Das geht nur wenn die Dimensionen zueinander passen. Die Matrix $A\in\mathbb{R}^{m\times n}$ können wir nur auf einen Vektor $y\in\mathbb{R}^n$ anwenden. $y=Bx$ besitzt soviele Einträge, wie B Zeilen besitzt, und x muss soviele Einträge besitze, wie B Spalten besitzt. $A(Bx)$ ist also nur definiert für

$$A\in\mathbb{R}^{m\times n},\quad B\in\mathbb{R}^{n\times r},\quad x\in\mathbb{R}^r$$

und entsprechend können wir das Matrix-Produkt AB nur bilden für $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times r}$ (die Anzahl der Spalten der linken Matrix muss also gleich der Anzahl der Zeilen der rechten Matrix sein).

Um C so zu definieren, dass tatsächlich $Cx = A(Bx)$ gilt, nutzen wir aus, dass gemäß Bemerkung 3.20 die k-te Spalte von C durch Ce_k gegeben ist. Der (i,k)-te Eintrag von C muss also der i-te Eintrag von $Ce_k = A(Be_k)$ sein. Wiederum nach Bemerkung 3.20 ist

$$y := Be_k = \begin{pmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{pmatrix}$$

und daher muss gelten

$$c_{ik} = \sum_{j=1}^{n} a_{ij} y_j = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

Definition 3.25 (Matrixmultiplikation)
Für

$$A = (a_{ij}) \in \mathbb{R}^{m\times n} \quad \text{und} \quad B = (b_{kl}) \in \mathbb{R}^{n\times r}$$

definieren wir das Produkt *AB durch $AB := (c_{ik}) \in \mathbb{R}^{m\times r}$ mit*

$$c_{ik} := \sum_{j=1}^{n} a_{ij} b_{jk}.$$

Es ist also

$$AB = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11}+a_{12}b_{21}+\ldots+a_{1n}b_{n1} & a_{11}b_{12}+a_{12}b_{22}+\ldots+a_{12}b_{n2} & \cdots & a_{11}b_{1n}+a_{12}b_{2n}+\ldots+a_{1n}b_{nr} \\ a_{21}b_{11}+a_{22}b_{21}+\ldots+a_{2n}b_{n1} & a_{21}b_{12}+a_{22}b_{22}+\ldots+a_{2n}b_{n2} & \cdots & a_{21}b_{1n}+a_{22}b_{2n}+\ldots+a_{2n}b_{nr} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11}+a_{m2}b_{21}+\ldots+a_{mn}b_{n1} & a_{m1}b_{12}+a_{m2}b_{22}+\ldots+a_{mn}b_{n2} & \cdots & a_{m1}b_{1r}+a_{m2}b_{2r}+\ldots+a_{mn}b_{nr} \end{pmatrix}.$$

Wir betonen nochmal, dass für Definition 3.25 die Anzahl der Spalten in der linken Matrix A mit der Anzahl der Zeilen in der rechten Matrix B übereinstimmen muss. Das Ergebnis ist eine Matrix mit der Zeilenanzahl der linken Matrix A und der Spaltenanzahl der rechten Matrix B. Für andere Kombinationen von Matrizen ist die Matrixmultiplikation nicht definiert.

Vektoren im $\mathbb{R}^n$ können wir auch als Matrizen mit n Zeilen und 1 Spalte auffassen, also als Elemente von $\mathbb{R}^{n\times 1}$. Das in Definition 3.17 definierte Matrix-Vektor-Produkt ist damit ein Spezialfall des hier definierten Matrix-Matrix-Produktes.

Ein weiterer wichtiger Spezialfall ist die Multiplikation einer $1\times n$ Matrix (*liegender* Vektors) mit einer $n\times 1$-Matrix (*stehender* Vektor):

$$\begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = a_1x_1 + \ldots + a_nx_n.$$

Der (i,k)-te Eintrag des Matrix-Produkts ist

$$c_{ik} = \sum_{j=1}^{n} a_{ij}b_{jk} = \underbrace{\begin{pmatrix} a_{i1} & \cdots & a_{in} \end{pmatrix}}_{=i\text{-te Zeile von } A} \underbrace{\begin{pmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{pmatrix}}_{=j\text{-te Spalte von } B}$$

Die Zuordnung lässt sich mit dem folgenden Schema leicht merken:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1r} \\ c_{21} & c_{22} & \cdots & c_{2r} \\ c_{31} & c_{32} & \cdots & c_{3r} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mr} \end{pmatrix}.$$

Eine weiter nützliche Beobachtung ist, dass sich die j-te Spalte von $C = AB$ gerade durch Anwendung von A auf die j-te Spalte von B ergibt, also

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix} = \begin{pmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{mj} \end{pmatrix}.$$

Satz 3.26 (Rechenregeln für das Matrixprodukt)
(a) Für $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{p\times r}$ gilt das Assoziativgesetz

$$(AB)C = A(BC).$$

(b) Für $A, B \in \mathbb{R}^{m\times n}$ und $C, D \in \mathbb{R}^{n\times p}$ gelten die Distributivgesetze

$$(A+B)C = AC+BC \quad \text{und} \quad A(C+D) = AC+AD.$$

(c) Für die $(n\times n)$-Einheitsmatrix

$$I := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}$$

gilt $AI = A$ für jede Matrix $A \in \mathbb{R}^{m\times n}$ und $IB = B$ für jede Matrix $B \in \mathbb{R}^{n\times p}$.

(d) Für $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$ gilt

$$(AB)^T = B^T A^T \in \mathbb{R}^{p\times m}.$$

Beweis: Seien

$$A = (a_{ij}) \in \mathbb{R}^{m\times n}, \quad B = (b_{jk}) \in \mathbb{R}^{n\times p} \quad \text{und} \quad C = (c_{kl}) \in \mathbb{R}^{p\times r}.$$

AB ist nach Definition 3.17 eine $(m\times p)$-Matrix und ihr (i,k)-ter Eintrag lautet

$$\sum_{j=1}^{n} a_{ij}b_{jk} \quad \text{für alle } i = 1,\dots,m,\ k = 1,\dots p.$$

$(AB)C$ ist daher nach Definition 3.17 eine $(m\times r)$-Matrix und ihr (i,l)-ter Eintrag lautet

$$\sum_{k=1}^{p}\left(\sum_{j=1}^{n} a_{ij}b_{jk}\right)c_{kl} \quad \text{für alle } i = 1,\dots,m,\ l = 1,\dots r.$$

BC ist nach Definition 3.17 eine $(n\times r)$-Matrix und ihr (j,l)-ter Eintrag lautet

$$\sum_{k=1}^{p} b_{jk}c_{kl} \quad \text{für alle } j = 1,\dots,n,\ l = 1,\dots r.$$

$A(BC)$ ist daher nach Definition 3.17 eine $(m\times r)$-Matrix und ihr (i,l)-ter Eintrag lautet

$$\sum_{j=1}^{n} a_{ij}\left(\sum_{k=1}^{p} b_{jk}c_{kl}\right) \quad \text{für alle } i = 1,\dots,m,\ l = 1,\dots r.$$

Da für alle $i = 1,\dots,m$ und $l = 1,\dots r$

$$\sum_{k=1}^{p}\left(\sum_{j=1}^{n} a_{ij}b_{jk}\right)c_{kl} = \sum_{k=1}^{p}\sum_{j=1}^{n} a_{ij}b_{jk}c_{kl} = \sum_{j=1}^{n}\sum_{k=1}^{p} a_{ij}b_{jk}c_{kl}$$
$$= \sum_{j=1}^{n} a_{ij}\left(\sum_{k=1}^{p} b_{jk}c_{kl}\right),$$

stimmen die (i,l)-ten Einträge von $(AB)C$ und $A(BC)$ jeweils überein, also ist $(AB)C = A(BC)$. Damit ist (a) gezeigt.

(b), (c) und (d) folgen ähnlich (aber einfacher) aus Definition 3.17. □

Bemerkung 3.27

(a) Das Matrixprodukt ist nach Satz 3.26 assoziativ, aber dies gilt nur wenn die Dimensionen passen. *Seien beispielsweise*

$$A := \begin{pmatrix}1 & 2\end{pmatrix} \in \mathbb{R}^{1\times 2}, \quad B := \begin{pmatrix}3\\4\end{pmatrix} \in \mathbb{R}^{2\times 1}, \quad C := \begin{pmatrix}5\\6\end{pmatrix} \in \mathbb{R}^{2\times 1}.$$

Dann ist

$$(AB)C = \left(\begin{pmatrix}1 & 2\end{pmatrix}\begin{pmatrix}3\\4\end{pmatrix}\right)\begin{pmatrix}5\\6\end{pmatrix} = 11\begin{pmatrix}5\\6\end{pmatrix} = \begin{pmatrix}55\\66\end{pmatrix} \in \mathbb{R}^{2\times 1},$$

aber

$$A(BC) = \begin{pmatrix}1 & 2\end{pmatrix}\left(\begin{pmatrix}3\\4\end{pmatrix}\begin{pmatrix}5\\6\end{pmatrix}\right) \quad \textit{ist nicht definiert,}$$

da die Anzahl der Spalten von $\begin{pmatrix}3\\4\end{pmatrix}$ *nicht mit der Anzahl der Zeilen von* $\begin{pmatrix}5\\6\end{pmatrix}$ *übereinstimmt.*

(b) Für die Multiplikation von Matrizen gilt im Allgemeinen nicht das Kommutativgesetz $AB = BA$. Seien beispielsweise

$$A := \begin{pmatrix}1 & 2\\3 & 4\end{pmatrix}, \quad B := \begin{pmatrix}5\\6\end{pmatrix},$$

dann ist $AB = \begin{pmatrix}17\\39\end{pmatrix}$*, aber BA ist nicht definiert.*

Für das Beispiel

$$A := \begin{pmatrix}1 & 2\end{pmatrix}, \quad B := \begin{pmatrix}5\\6\end{pmatrix},$$

gilt

$$AB = \begin{pmatrix}1 & 2\end{pmatrix}\begin{pmatrix}5\\6\end{pmatrix} = 17 \in \mathbb{R}^{1\times 1}, \quad BA = \begin{pmatrix}5\\6\end{pmatrix}\begin{pmatrix}1 & 2\end{pmatrix} = \begin{pmatrix}5 & 10\\6 & 12\end{pmatrix} \in \mathbb{R}^{2\times 2}$$

sind AB und BA beide definiert aber nicht mal ihre Dimensionen stimmen überein.

Für das Beispiel mit quadratischen Matrizen

$$A := \begin{pmatrix}1 & 1\\0 & 0\end{pmatrix} \in \mathbb{R}^{2\times 2}, \quad B := \begin{pmatrix}0 & 1\\0 & 1\end{pmatrix} \in \mathbb{R}^{2\times 2}$$

gilt

$$AB = \begin{pmatrix}1 & 1\\0 & 0\end{pmatrix}\begin{pmatrix}0 & 1\\0 & 1\end{pmatrix} = \begin{pmatrix}0 & 2\\0 & 0\end{pmatrix}, \quad BA = \begin{pmatrix}0 & 1\\0 & 1\end{pmatrix}\begin{pmatrix}1 & 1\\0 & 0\end{pmatrix} = \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}.$$

AB und BA sind also von gleicher Dimension, aber $AB \neq BA$.

(c) *Das letzte Beispiel aus (b) zeigt auch das die Matrixmultiplikation nicht* nullteilerfrei *ist, d.h. für $A \neq 0$ und $B \neq 0$ kann trotzdem $AB = 0$ sein.*

3.2.3 Lineare Abbildungen

Definition 3.28
Sei $f : V \to W$ eine Abbildung zwischen zwei reellen Vektorräumen V und W. f heißt lineare Abbildung (kurz: f ist linear*), falls gilt*

(a) Additivität*: $f(u+v) = f(u) + f(v)$ für alle $u, v \in V$.*

(b) Homogenität*: $f(av) = af(v)$ für alle $a \in \mathbb{R}$, $v \in V$.*

Für eine lineare Abbildung $f : V \to W$ definieren wir

(a) *Kern von f (engl.: Nullspace):*

$$\mathcal{N}(f) := \{v \in V : f(v) = 0\} \subseteq V.$$

(b) *Bild von f (engl.: Range):*

$$\mathcal{R}(f) := f(V) = \{f(v) : v \in V\} \subseteq W.$$

Satz 3.29
Sei $f : V \to W$ eine lineare Abbildung zwischen reellen Vektorräumen V und W.

(a) $\mathcal{N}(f)$ *ist Untervektorraum von* V *und* $\mathcal{R}(f)$ *ist Untervektorraum von* W.

(b) f *ist surjektiv, genau dann wenn* $\mathcal{R}(f) = W$.

(c) f *ist injektiv, genau dann wenn* $\mathcal{N}(f) = 0$ *(d.h. wenn* $\mathcal{N}(f)$ *nur den Nullvektor* $0 \in V$ *enthält).*

Beweis: (a) rechnet man mit Satz 3.6 leicht nach. (b) folgt sofort aus der Definition von Surjektivität in Abschnitt 1.2.3.

Zum Beweis von (c) bezeichnen wir abweichend von der bisherigen Notation den Nullvektor in V und W mit $\vec{0} \in V$ und $\vec{0} \in W$ und die reelle Null mit $0 \in \mathbb{R}$. Da f linear ist, gilt

$$f(\vec{0}) = f(0 * \vec{0}) = 0 * f(\vec{0}) = \vec{0}.$$

Für jedes $u \in \mathcal{N}(f)$ gilt daher $f(u) = \vec{0} = f(\vec{0})$. Falls f injektiv ist, dann folgt daraus $u = \vec{0}$, also $\mathcal{N}(f) = \{\vec{0}\}$.

Sei nun umgekehrt $\mathcal{N}(f) = \{\vec{0}\}$. Um zu zeigen, dass f injektiv ist, müssen wir zeigen, dass für $u, v \in V$ gilt

$$f(v) = f(u) \quad \Longrightarrow \quad v = u.$$

Seien also $u, v \in V$ mit $f(v) = f(u)$. Dann ist

$$\begin{aligned} f(u-v) = f(u+(-1)v) = f(u) + f((-1)v) = f(u) + (-1)f(v) \\ = f(u) - f(v) = \vec{0}, \end{aligned}$$

also $u - v \in \mathcal{N}(f) = \{\vec{0}\}$ und damit $u - v = \vec{0}$, womit $u = v$ gezeigt ist. □

Satz 3.30 (Dimensionsformel)
Sei $f : V \to W$ *eine lineare Abbildung zwischen endlich-dimensionalen reellen Vektorräumen* V *und* W. *Dann gilt*

$$\dim V = \dim \mathcal{N}(f) + \dim \mathcal{R}(f).$$

Beweis: Wir skizzieren nur die grundlegende Idee. Es bezeichne $n := \dim V$ und $k := \dim \mathcal{N}(f)$. Außerdem sei $\{u_1, \ldots, u_k\}$ eine Basis von $\mathcal{N}(f)$. Da $\mathcal{N}(f)$ ein Unterraum von $\dim V$ ist, gilt wegen Bemerkung 3.13(c) $k \leq n$ und wir können Vektoren $v_{k+1}, \ldots, v_n \in V$ finden, die die Basis von $\mathcal{N}(f)$ zu einer Basis

$$\{u_1, \ldots, u_k, v_{k+1}, \ldots, v_n\} \subseteq V$$

von V ergänzen. Man kann zeigen, dass dann die Vektoren

$$\{f(v_{k+1}), \ldots, f(v_n)\} \subseteq \mathcal{R}(f) \subseteq W$$

eine Basis von $\mathscr{R}(f)$ bilden. Damit folgt dann

$$\dim \mathscr{R}(f) = n - k = \dim V - \dim \mathscr{N}(f).$$

und damit ist die Dimensionsformel bewiesen. □

Definition 3.31
Eine Matrix $A \in \mathbb{R}^{m\times n}$ definiert eine Abbildung

$$\mathbb{R}^n \to \mathbb{R}^m, \quad x \mapsto Ax.$$

und nach den Rechenregeln für die Matrix-Vektormultiplikation in Satz 3.19 ist diese Abbildung linear.

Wir verwenden die Begriffe aus Definition 3.28 für die Matrix A, wenn sie für diese zugehörige lineare Abbildung $x \mapsto Ax$ gelten, also

(a) Kern von A (engl.: Nullspace):

$$\mathscr{N}(A) := \{x \in \mathbb{R}^n : \ Ax = 0\} \subseteq \mathbb{R}^n.$$

(b) Bild von A (engl.: Range):

$$\mathscr{R}(A) := \{Ax : \ x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

Entsprechend nennen wir eine Matrix A injektiv bzw. surjektiv, wenn die Abbildung $x \mapsto Ax$ injektiv bzw. surjektiv ist.

Folgerung 3.32
Sei $A \in \mathbb{R}^{m\times n}$.

(a) Satz 3.29 zeigt:

$$\begin{aligned} A \text{ ist injektiv} &\iff \mathscr{N}(A) = 0, \\ A \text{ ist surjektiv} &\iff \mathscr{R}(A) = \mathbb{R}^m. \end{aligned}$$

(b) Aus der Dimensionsformel in Satz 3.30 folgt

$$\dim \mathscr{N}(A) + \dim \mathscr{R}(A) = n.$$

(c) Eine Matrix mit mehr Zeilen als Spalten (also $m > n$) ist nicht surjektiv.

(d) Eine Matrix mit mehr Spalten als Zeilen (also $m < n$) ist nicht injektiv.

(e) Für eine quadratische Matrix $A \in \mathbb{R}^{n\times n}$ (also $m = n$) gilt

$$A \text{ injektiv} \iff A \text{ surjektiv} \iff A \text{ bijektiv}.$$

Man nennt $\operatorname{rang}(A) := \dim \mathscr{R}(A)$ *auch den* Rang *der Matrix A. Die Dimensionsformel heißt deshalb auch* Rangformel.

Beweis: (a) folgt aus Satz 3.29.

(b) folgt aus Satz 3.30.

(c) Aus (b) folgt auch dass $\dim \mathscr{R}(A) \leq n$. Für $m > n$ kann also nicht $\mathscr{R}(A) = \mathbb{R}^m$ gelten und daher A nicht surjektiv sein.

(d) Da $\mathscr{R}(A) \subseteq \mathbb{R}^m$ gilt $\dim \mathscr{R}(A) \leq m$. Für $m < n$ folgt aus (b) deshalb

$$\dim \mathscr{N}(A) = n - \dim \mathscr{R}(A) \geq n - m > 0,$$

und damit $\mathscr{N}(A) \neq 0$, also (nach (a)) A nicht injektiv.

(e) Ist A injektiv, dann gilt nach (b)

$$\dim \mathscr{R}(A) = n - \dim \mathscr{N}(A) = n - 0 = n = \dim \mathbb{R}^n$$

und daher $\mathscr{R}(A) = \mathbb{R}^n$, also ist A auch surjektiv und damit auch bijektiv.

Ist A surjektiv, dann gilt

$$\dim \mathscr{N}(A) = n - \dim \mathscr{R}(A) = n - n = 0,$$

also $\mathscr{N}(A) = 0$ und damit ist A auch injektiv und daher auch bijektiv. □

Bemerkung 3.33
Eine Matrix $A \in \mathbb{R}^{m\times n}$ definiert eine lineare Abbildung zwischen den Vektorräumen $\mathbb{R}^n$ und $\mathbb{R}^m$. Umgekehrt lässt sich aber auch jede lineare Abbildung $f : V \to W$ zwischen zwei endlich-dimensionalen reellen Vektorräumen V und W mit $\dim V = n$ und $\dim V = m$ als Matrix $A \in \mathbb{R}^{m\times n}$ schreiben. Wir demonstrieren dies am Beispiel der quadratischen Polynome

$$\Pi_2 := \{p(x) = ax^2 + bx + c : \ a, b, c \in \mathbb{R}\}.$$

Bezüglich der Basis der Monome $p_1(x) = x^2$, $p_2(x) = x$, $p_3(x) = 1$ können wir jedes Polynom $p(x) = ax^2 + bx + c \in \Pi_2$ identifizieren mit dem Vektor $\begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{R}^3$.

Entsprechend wird die Ableitung $p'(x) = 2ax + b \in \Pi_2$ *identifiziert mit dem Vektor* $\begin{pmatrix} 0 \\ 2a \\ b \end{pmatrix} \in \mathbb{R}^3$ *und der Ableitungsoperator*

$$\Pi_2 \to \Pi_2, \quad p(x) \mapsto p'(x).$$

entspricht bezüglich dieser Identifikation der Matrixanwendung

$$\begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 2a \\ b \end{pmatrix}$$

Die Spalten dieser Matrix ergeben sich dabei gerade aus den Vektoren mit denen die Ableitung der verwendeten Basiselemente identifiziert werden $p_1'(x) = 2x$ *entspricht* $\begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$, $p_2'(x) = 1$ *entspricht* $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ *und* $p_3'(x) = 0$ *entspricht* $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

Man beachte beim Übergang von allgemeinen linearen Abbildungen zu Matrizen aber, dass die Identifikation eines allgemeinen Vektors $v \in V$ *mit seinen Entwicklungskoeffizienten und daher auch das Aufstellen der Matrix von der Wahl der Basen der Vektorräume* V *und* W *und auch von der Reihenfolge der Basiselemente abhängt.*

3.2.4 Lineare Gleichungssysteme und die inverse Matrix

Wir haben in Beispiel 3.18 schon gesehen, dass sich lineare Gleichungssysteme (LGS) in Matrix-Vektor-Schreibweise schreiben lassen. In einem LGS mit m Gleichungen und n Unbekannten suchen wir $x_1, x_2, \ldots, x_n \in \mathbb{R}$, so dass

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= y_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= y_m, \end{aligned}$$

wobie die Koeffizienten $a_{ij} \in \mathbb{R}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ und die rechten Seiten $y_1, \ldots, y_m \in \mathbb{R}$ gegeben seien. Dies entspricht der Suche nach einem Vektor $x = (x_i) \in \mathbb{R}^n$, der die Matrix-Vektor-Gleichung

$$Ax = y$$

mit gegebener Matrix $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ und rechter Seite $y = (y_i) \in \mathbb{R}^m$ löst.

Satz 3.34
Seien $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ und $y = (y_i) \in \mathbb{R}^m$.

(a) Das LGS $Ax = y$ besitzt genau dann (mindestens) eine Lösung, wenn $y \in \mathscr{R}(A)$ gilt.

(b) Das LGS $Ax = y$ ist genau dann eindeutig lösbar (also lösbar mit genau einer Lösung), wenn $y \in \mathscr{R}(A)$ und $\mathscr{N}(A) = 0$ gilt.

(c) Ist $\hat{x} \in \mathbb{R}^n$ eine *Lösung des LGS $Ax = y$, dann ist die Menge* aller *Lösungen gegeben durch*

$$\{x \in \mathbb{R}^n : Ax = y\} = \{\hat{x} + x^{(0)} : x^{(0)} \in \mathscr{N}(A)\}.$$

Beweis: (a) folgt aus der Definition von $\mathscr{R}(A)$. (b) folgt, da wegen Folgerung 3.32 $\mathscr{N}(A) = 0$ äquivalent zur Injektivität von A ist.

Zum Beweis von (c) sei $A\hat{x} = y$ und $Ax = y$, dann gilt $A(\hat{x} - x) = A\hat{x} - Ax = 0$, also $\hat{x} - x \in \mathscr{N}(A)$. Umgekehrt gilt für $A\hat{x} = y$ und $x^{(0)} \in \mathscr{N}(A)$ auch

$$A(\hat{x} + x^{(0)}) = A\hat{x} + Ax^{(0)} = y + 0 = y.$$

Aufgrund der Bedeutung von $\mathscr{N}(A)$ für die Eindeutigkeit der Lösung definiert man folgenden Begriff:

Definition 3.35
Für $0 \neq y$ heißt $Ax = y$ auch inhomogenes LGS.

Die Vektoren im $\mathscr{N}(A)$, also diejenigen $x^{(0)} = (x_j^{(0)}) \in \mathbb{R}^n$ mit $Ax^{(0)} = 0$, d.h.

$$\begin{aligned} a_{11}x_1^{(0)} + a_{12}x_2^{(0)} + \ldots + a_{1n}x_n^{(0)} &= 0 \\ a_{21}x_1^{(0)} + a_{22}x_2^{(0)} + \ldots + a_{2n}x_n^{(0)} &= 0 \\ &\vdots \\ a_{m1}x_1^{(0)} + a_{m2}x_2^{(0)} + \ldots + a_{mn}x_n^{(0)} &= 0, \end{aligned}$$

heißen auch Lösungen des zugehörigen homogenen LGS.

Zwei Lösungen eines inhomogenen LGS unterscheiden sich also gerade um eine Lösung des zugehörigen homogenen LGS. Das homogene LGS $Ax = 0$ besitzt natürlich immer die Lösung $x = 0$, diese nennt man auch die *triviale Lösung des homogenen LGS*.

Mit den Ergebnissen des letzten Abschnittes erhalten wir die folgende Charakterisierung der Lösungen eines LGS.

Satz 3.36

(a) Besitzt ein LGS mehr Gleichungen als Unbekannte, dann existieren rechte Seiten, für die keine Lösung existiert.

(b) Besitzt ein LGS mehr Unbekannte als Gleichungen, dann ist die Lösung (falls sie denn existiert) nicht eindeutig.

(c) Für ein LGS mit gleich vielen Unbekannten wie Gleichungen sind äquivalent:

(i) Das zugehörige homogene LGS besitzt nur die triviale Lösung.

(ii) Das LGS besitzt für jede rechte Seite eine Lösung.

(iii) Das LGS besitzt für jede rechte Seite eine Lösung und diese ist eindeutig.

Beweis: (a) Besitzt ein LGS mehr Gleichungen als Unbekannte, dann besitzt die zugehörige Matrix mehr Zeilen als Spalten, kann also nach Folgerung 3.32 nicht surjektiv sein, und damit gibt es rechte Seiten, die nicht im Bild der Matrix liegen, also für die (nach Satz 3.34) das LGS nicht lösbar ist.

(b) Besitzt ein LGS mehr Unbekannte als Gleichungen, dann besitzt die zugehörige Matrix mehr Spalten als Zeilen, kann also nach Folgerung 3.32 nicht injektiv sein, und daher ist (nach Satz 3.34) die Lösung nicht eindeutig.

(c) Besitzt ein LGS gleich viele Unbekannte wie Gleichungen, dann ist die zugehörige Matrix quadratisch und nach 3.32 ist sie daher injektiv genau dann wenn sie surjektiv ist (und damit ist sie dann auch bijektiv). Da Injektivität äquivalent zu (i), Surjektivität äquivalent zu (ii) und Bijektivität äquivalent zu (iii) ist, folgt (c). □

Systematische und auch für große LGS effizient auf dem Computer implementierbare Lösungsalgorithmen für lineare Gleichungssysteme lernen Sie in den Vorlesung zur Numerik kennen. Wir geben aber schon ein einfaches Beispiel für die Bedeutung von Satz 3.36.

Beispiel 3.37

(a) Betrachte eine lineare Gleichung mit drei Unbekannten

$$x_1 + x_2 + x_3 = 3. \tag{3.3}$$

Die Lösung $x_1 = x_2 = x_3 = 1$ fällt sofort ins Auge. Diese kann aber nicht die einzige Lösung sein, da das LGS mehr Unbekannte als Gleichungen enthält.

Das zugehörige homogene LGS lautet

$$x_1^{(0)} + x_2^{(0)} + x_3^{(0)} = 0.$$

Die Menge aller Lösungen des inhomogenen LGS (3.3) *lautet daher*

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{pmatrix} \in \mathbb{R}^3 : \; x_1^{(0)} + x_2^{(0)} + x_3^{(0)} = 0 \right\}.$$

(b) Betrachte das inhomogene LGS

$$\begin{aligned} x_1 - x_2 - x_3 &= 3, \\ x_2 - x_3 &= 7, \\ x_1 - 2x_2 - x_3 &= 12. \end{aligned}$$

Das zugehörige homogene LGS lautet

$$\begin{aligned} x_1^{(0)} - x_2^{(0)} - x_3^{(0)} &= 0, \\ x_2^{(0)} - x_3^{(0)} &= 0, \\ x_1^{(0)} - 2x_2^{(0)} - x_3^{(0)} &= 0. \end{aligned}$$

Ziehen wir die dritte Gleichung von der ersten ab, so erhalten wir $x_2^{(0)} = 0$. *Daraus folgt dann mit der zweiten Gleichung auch* $x_3^{(0)} = 0$ *und dann mit der ersten auch* $x_1^{(0)} = 0$. *Das homogene LGS besitzt also nur die triviale Lösung. Das inhomogene LGS ist deshalb eindeutig lösbar.*

In der folgenden Definition bezeichnet I wieder die Einheitsmatrix in $\mathbb{R}^{n\times n}$.

Definition und Satz 3.38

(a) Sei $A \in \mathbb{R}^{n\times n}$ *bijektiv. Dann exisitiert eine Matrix* $A^{-1} \in \mathbb{R}^{n\times n}$ *mit der Eigenschaft, dass*

$$AA^{-1} = I \quad \text{und} \quad A^{-1}A = I.$$

A^{-1} *heißt die zu* A *inverse Matrix. Eine bijektive Matrix* A *nennt man deshalb auch invertierbar.*

Auch A^{-1} *ist wiederum bijektiv und es gilt* $(A^{-1})^{-1} = A$.

(b) Seien $A, M \in \mathbb{R}^{n\times n}$. *Gilt*

$$AM = I \quad \text{oder} \quad MA = I,$$

dann sind A *und* M *bijektiv und es gilt* $A^{-1} = M$ *und* $M^{-1} = A$. *Insbesondere ist die inverse Matrix eindeutig bestimmt.*

(c) *Sind $A, B \in \mathbb{R}^{n\times n}$ bijektiv, dann ist auch $AB \in \mathbb{R}^{n\times n}$ bijektiv und es gilt*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

(d) *Für $A \in \mathbb{R}^{n\times n}$ gilt genau dann $A^T = A^{-1}$, wenn die Spalten von A eine ONB des $\mathbb{R}^n$ bilden. In diesem Fall heißt A unitär und es gilt $\|Ax\|_2 = \|x\|_2$ für alle $x \in \mathbb{R}^n$.*

Beweis: (a) Wegen der Bijektivität von A ist das LGS $Ax = y$ für jede rechte Seite $y \in \mathbb{R}^n$ eindeutig lösbar.

Bezeichne

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

die Basis aus Einheitsvektoren des $\mathbb{R}^n$. Für $j = 1, \ldots, n$ sei $v_j \in \mathbb{R}^n$ die Lösung von $Av_j = e_j$ und bilde damit spaltenweise die Matrix

$$A^{-1} := \begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Dann ist nach den Regeln der Matrixmultiplikation die j-te Spalte von AA^{-1} gegeben durch $Av_j = e_j$ und damit

$$AA^{-1} = \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} = I \in \mathbb{R}^{n\times n}.$$

A^{-1} ist auch injektiv, denn aus $A^{-1}x = 0$ folgt $x = AA^{-1}x = 0$. Da A^{-1} quadratisch ist, ist es damit auch bijektiv. Mit dem bereits bewiesenen existiert deshalb eine Matrix $(A^{-1})^{-1} \in \mathbb{R}^{n\times n}$, so dass $A^{-1}(A^{-1})^{-1} = I$. Damit folgt dann aber auch, dass

$$A = AI = AA^{-1}(A^{-1})^{-1} = (A^{-1})^{-1}$$

und damit $A^{-1}A = A^{-1}(A^{-1})^{-1} = I$.

(b) Ist $M \in \mathbb{R}^{n\times n}$ eine Matrix mit $AM = I$, dann gilt für jedes $x \in \mathbb{R}^n$

$$x = Ix = AMx = A(Mx) \in \mathscr{R}(A).$$

A ist also surjektiv und damit ist A auch bijektiv. Mit (a) folgt außerdem, dass $M = A^{-1}AM = A^{-1}$.

Ist $M \in \mathbb{R}^{n\times n}$ eine Matrix mit $MA = I$, dann gilt für jedes $x \in \mathbb{R}^n$

$$Ax = 0 \quad \Longrightarrow \quad x = MAx = M0 = 0.$$

A ist also injektiv und damit ist A auch bijektiv. Mit (a) folgt außerdem, dass $M = MAA^{-1} = A^{-1}$.

(c) Mit $M := B^{-1}A^{-1}$ gilt

$$(AB)M = ABB^{-1}A^{-1} = I.$$

Wegen (b) ist daher AB bijektiv und es gilt $(AB)^{-1} = M = B^{-1}A^{-1}$.

(d) Bezeichne $v_1, \ldots, v_n \in \mathbb{R}^n$ die Spalten von A. Dann sind $v_1^T, \ldots, v_n^T \in \mathbb{R}^{1,n}$ die Zeilenvektoren von A^T und der (j,k)-te Eintrag von A^TA ist $v_j^T v_k \in \mathbb{R}$. Daher ist A^TA genau dann die Einheitsmatrix I, wenn

$$v_j^T v_k = \begin{cases} 1 & \text{für } j = k, \\ 0 & \text{für } j \neq k. \end{cases}$$

Ist A unitär, dann gilt

$$\|Ax\|_2^2 = (Ax)^T(Ax) = x^TA^TAx = x^TIx = x^Tx = \|x\|_2^2$$

für alle $x \in \mathbb{R}^n$. □

Folgerung 3.39
Die Lösung eines LGS $Ax = b$ mit invertierbarer Matrix A lautet $x = A^{-1}b$.

Zur effizienten Lösung eines LGS nutzt man aus, dass sich die Lösung nicht ändert, wenn wir

(a) zwei Gleichungen vertauschen, oder

(b) eine Gleichung mit einer Zahl $0 \neq r \in \mathbb{R}$ multiplizieren, oder

(c) eine Gleichung zu einer anderen addieren, oder

(d) ein (von Null verschiedenes) Vielfaches einer Gleichung zu einer anderen addieren.

Durch Vertauschung lässt sich sicherstellen, dass die erste Gleichung die erste Unbekannte enthält. Durch Addition geeigneter Vielfache der ersten Gleichung auf die anderen Gleichungen lässt sich die erste Unbekannte aus der zweiten bis letzten Gleichung eliminieren. Dieses Verfahren wiederholt man mit der zweiten Gleichung und eliminiert so die zweite Unbekannte aus der dritten bis letzten Gleichung. So fährt man fort bis das resultierende LGS Dreiecksgestalt besitzt. Dann kann der Wert der letzten Unbekannten aus der letzten Gleichung abgelesen werden, damit dann der Wert der vorletzten Unbekannten aus der vorletzten Gleichung berechnet werden und damit der vorvorletzte Wert usw.

Dieses Vorgehen heißt Gauß-Verfahren und damit lässt sich eine Matrix A auch auf Invertierbarkeit überprüfen und A^{-1} praktisch berechnen. Dieses und weitere Methoden zur effizienten Lösung linearer Gleichungssysteme sind Gegenstand der Numerik-Vorlesungen.

3.3 Die Determinante

Wir werden in diesem Abschnitt eine Funktion kennenleren, die (für beliebige Dimension $n \in \mathbb{N}$) jeder quadratischen Matrix jeweils eine reelle Zahl zuordnet

$$\det : \mathbb{R}^{n\times n} \to \mathbb{R}.$$

Für $A \in \mathbb{R}^{n\times n}$ heißt $\det A$ die Determinante der Matrix A. Die Determinante wird die folgenden Eigenschaften besitzen:

(a) $\det A \neq 0$ gilt genau dann, wenn A invertierbar ist.

(b) $\det A$ entspricht dem Volumen des n-dimensionalen Körpers, den die n Spaltenvektoren von A aufspannen.

Für eine 1×1-Matrix (also eine einzelne Zahl) $A \in \mathbb{R}^{1\times 1}$ ist $\det A$ diese Zahl. Für $A \in \mathbb{R}^{2\times 2}$ ist $\det A$ der Flächeninhalt des von den beiden Spalten aufgespannten Parallelograms, für $A \in \mathbb{R}^{3\times 3}$ ist $\det A$ das Volumen des von den drei Spalten von A aufgespannten Spats.

Die wichtigsten Methoden zur Definition und Berechnung von $\det A$ stellen wir in den folgenden Abschnitten vor.

3.3.1 Die Leibniz-Formel

Wir definieren die Determinante mittels der sogenannten Leibniz-Formel und beschreiben diese zunächst mit Worten. Sei $A \in \mathbb{R}^{n\times n}$. Wir wählen in jeder Zeile von A genau ein Element aus und wählen dabei aber jedesmal eine andere Spalte (also wenn wir in der 1. Zeile das 2. Element ausgewählt haben, dann nehmen wir in den anderen Zeilen jeweils nicht das 2. Element, usw.).

Für eine $n \times n$-Matix wird in jeder der n-Zeilen eine der Spalten $1,\ldots,n$ ausgewählt und keine Spalte mehrfach gewählt. Dies entspricht dem Ziehen ohne Zurücklegen der n Zahlen $1,\ldots,n$, oder äquivalent der Vertauschung dieser Zahlen. Die Auswahlmöglichkeiten heißen deshalb auch Permutationen und für eine $n \times n$-Matrix gibt es nach Abschnitt 2.1.2 $n!$ solche Permutationen.

Für $n = 2$ gibt es $2! = 2$ Permutationen:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Für $n = 3$ gibt es $3! = 6$ Permutationen:

$$\begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & \boxed{a_{22}} & a_{23} \\ a_{31} & a_{32} & \boxed{a_{33}} \end{pmatrix} \quad \begin{pmatrix} a_{11} & \boxed{a_{12}} & a_{13} \\ \boxed{a_{21}} & a_{22} & a_{23} \\ a_{31} & a_{32} & \boxed{a_{33}} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} & \boxed{a_{13}} \\ \boxed{a_{21}} & a_{22} & a_{23} \\ a_{31} & \boxed{a_{32}} & a_{33} \end{pmatrix}$$

$$\begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & a_{22} & \boxed{a_{23}} \\ a_{31} & \boxed{a_{32}} & a_{33} \end{pmatrix} \quad \begin{pmatrix} a_{11} & \boxed{a_{12}} & a_{13} \\ a_{21} & a_{22} & \boxed{a_{23}} \\ \boxed{a_{31}} & a_{32} & a_{33} \end{pmatrix} \quad \begin{pmatrix} a_{11} & a_{12} & \boxed{a_{13}} \\ a_{21} & \boxed{a_{22}} & a_{23} \\ \boxed{a_{31}} & a_{32} & a_{33} \end{pmatrix}$$

Für jede dieser Permutationen bilden wir das Produkt der ausgewählten Elemente. Außerdem zählen wir die sogenannten Fehlstände jeder Permutation, d.h. wie oft ein ausgewähltes Element weiter links steht als ein zuvor ausgewähltes Element. Dabei berücksichtigen wir die Vielfachheit, d.h. ein Element das weiter links steht als k zuvor ausgewählte Elemente zählt k-fach. Ist diese Anzahl ungerade, so schreiben wir ein Minus vor das zugehörige Produkt, ansonsten ein Plus. Die Determinante der Matrix ist dann die Summe aller so erhalten Terme.

Für $n = 2$ ergibt sich:

$$\begin{array}{cc} \begin{pmatrix} \boxed{a_{11}} & a_{12} \\ a_{21} & \boxed{a_{22}} \end{pmatrix} & \begin{pmatrix} a_{11} & \boxed{a_{12}} \\ \boxed{a_{21}} & a_{22} \end{pmatrix} \\ \text{0 Fehlstände} & \text{1 Fehlstand} \\ +a_{11}a_{22} & -a_{12}a_{21} \end{array}$$

und damit

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Für $n = 3$ ergibt sich:

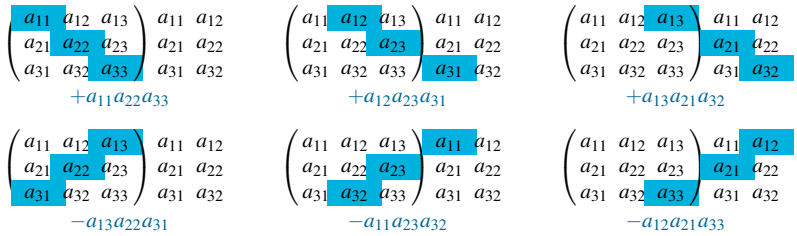
$$\begin{array}{ccc} \begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & \boxed{a_{22}} & a_{23} \\ a_{31} & a_{32} & \boxed{a_{33}} \end{pmatrix} & \begin{pmatrix} a_{11} & \boxed{a_{12}} & a_{13} \\ \boxed{a_{21}} & a_{22} & a_{23} \\ a_{31} & a_{32} & \boxed{a_{33}} \end{pmatrix} & \begin{pmatrix} a_{11} & a_{12} & \boxed{a_{13}} \\ \boxed{a_{21}} & a_{22} & a_{23} \\ a_{31} & \boxed{a_{32}} & a_{33} \end{pmatrix} \\ \text{0 Fehlstände} & \text{1 Fehlstand} & \text{2 Fehlstände} \\ +a_{11}a_{22}a_{33} & -a_{12}a_{21}a_{33} & +a_{13}a_{21}a_{32} \end{array}$$

$$\begin{array}{ccc} \begin{pmatrix} \boxed{a_{11}} & a_{12} & a_{13} \\ a_{21} & a_{22} & \boxed{a_{23}} \\ a_{31} & \boxed{a_{32}} & a_{33} \end{pmatrix} & \begin{pmatrix} a_{11} & \boxed{a_{12}} & a_{13} \\ a_{21} & a_{22} & \boxed{a_{23}} \\ \boxed{a_{31}} & a_{32} & a_{33} \end{pmatrix} & \begin{pmatrix} a_{11} & a_{12} & \boxed{a_{13}} \\ a_{21} & \boxed{a_{22}} & a_{23} \\ \boxed{a_{31}} & a_{32} & a_{33} \end{pmatrix} \\ \text{1 Fehlstand} & \text{2 Fehlstände} & \text{3 Fehlstände} \\ -a_{11}a_{23}a_{32} & +a_{12}a_{23}a_{31} & -a_{13}a_{22}a_{31}, \end{array}$$

und damit (die sogenannte Regel von Sarrus)

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = +a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ -a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

Die Regel von Sarrus kann man sich gut merken, indem man die ersten beiden Spalten der Matrix nocheinmal rechts neben die Matrix schreibt und dann jeweils das Produkt der Elemente der drei nach unten rechts verlaufenden Diagonalen mit Plus und der drei nach unten links verlaufenden Diagonalen mit Minus versehen aufaddiert:



Werden die Matrixeinträge ausgeschrieben, dann schreibt man die Determinante auch kurz als

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = \det\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

Diese Schreibweise ist allerdings nur für $n \geq 2$ sinnvoll, da für negative Zahlen $a_{11} < 0$

$$\det(a_{11}) = a_{11} \neq |a_{11}|.$$

Um die Leibniz-Formel für allgemeines $A \in \mathbb{R}^{n\times n}$ in Formelsprache auszudrücken definieren wir die verwendeten Begriffe mathematisch präzise.

Definition 3.40
(a) Eine (n-stellige) Permutation ist eine bijektive Abbildung

$$\sigma : \{1,\dots,n\} \to \{1,\dots,n\}.$$

Die Menge der n-stelligen Permutationen bezeichnen wir mit

$$S_n := \{\sigma : \{1,\dots,n\} \to \{1,\dots,n\},\ \sigma \text{ bijektiv}\}.$$

Nach Abschnitt 2.1.2 gilt $|S_n| = n!$.

(b) Für eine Permutation σ heißt ein Tupel $(i,j) \in \{1,\dots,n\}^2$ Fehlstand (auch: Inversion) von σ falls

$$i < j \quad \text{und} \quad \sigma(i) > \sigma(j).$$

Die Menge aller Fehlstände einer Permutation σ bezeichnen wir mit

$$\mathrm{inv}(\sigma) := \{(i,j) \in \{1,\dots,n\}^2 : \ i < j \ \wedge \sigma(i) > \sigma(j)\}.$$

Die Anzahl der Fehlstände einer Permutation beträgt entsprechend $|\mathrm{inv}(\sigma)|$.

(c) Für eine Permutation σ heißt

$$\mathrm{sgn}(\sigma) = (-1)^{|\mathrm{inv}(\sigma)|} = \begin{cases} +1 & \text{falls Anzahl der Fehlstände gerade,} \\ -1 & \text{falls Anzahl der Fehlstände ungerade,} \end{cases}$$

das Signum (auch: Vorzeichen oder Parität) der Permutation σ.

In Formelsprache lautet die Leibniz-Formel dann für allgemeines $A \in \mathbb{R}^{n\times n}$:

$$\det A := \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}. \tag{3.4}$$

Offenbar hätten wir die gleichen Permutationen auch dadurch erhalten können, dass wir nicht in jeder Zeile sondern in jeder Spalte ein Element auswählen. Man kann außerdem zeigen, dass es auch für das Signum der Permutation keine Rolle spielt, ob die Fehlstände spalten- oder zeilenweise gezählt werden. Daher ändert Vertauschung der Spalten und Zeilen die Determinante nicht und es gilt

$$\det A = \det A^T.$$

Die Leibniz-Formel besitzt den Vorteil einer expliziten Formel. Zur praktischen Berechnung der Determinante lässt sich die Leibniz-Formel aber nur für sehr kleine n verwenden, da die Formeln sehr schnell sehr lang werden. Schon für $n = 4$ entsteht eine Summe aus $4! = 24$ Summanden (die jeweils Produkte von vier Faktoren sind) und es ist dafür keine einfache Merkregel mehr bekannt. Für $n = 10$ entsteht eine Summe aus $10!$, d.h. über 3.5 Millionen Summanden.

3.3.2 Der Laplacesche Entwicklungssatz

Für (etwas) größere Matrizen lässt sich die Determinante mit der folgenden rekursiven Formel berechnen. Für $A \in \mathbb{R}^{n\times n}$ und $i,j \in \{1,\dots,n\}$ bezeichnen wir mit $A_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ diejenige $(n-1)\times(n-1)$-Matrix, die sich aus A durch Streichung (d.h. Löschung) der i-ten Zeile und der j-ten Spalte ergibt. Für

$$A = \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & a_{1j} & a_{1,j+1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j} & a_{i-1,j+1} & \dots & a_{i-1,n} \\ a_{i,1} & \dots & a_{i,j-1} & a_{i,j} & a_{i,j+1} & \dots & a_{i,n} \\ a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j} & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,j-1} & a_{n,j} & a_{n,j+1} & \dots & a_{n,n} \end{pmatrix} \in \mathbb{R}^{n\times n}$$

ist also

$$A_{ij} = \begin{pmatrix} a_{11} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j+1} & \dots & a_{i-1,n} \\ a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{n,n} \end{pmatrix} \in \mathbb{R}^{(n-1)\times(n-1)}.$$

Satz 3.41 (Entwicklungssatz von Laplace)
Sei $A \in \mathbb{R}^{n\times n}$.

(a) Für alle $i = 1,\dots,n$ gilt

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij},$$

Diese Formel heißt auch Entwicklung nach der i-ten Zeile.

(b) Für alle $j = 1,\dots,n$ gilt

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}.$$

Diese Formel heißt auch Entwicklung nach der j-ten Spalte.

Beweis: Mit vollständiger Induktion kann man beweisen, dass die Formel aus (a) das gleiche Ergebnis liefert wie die zur Definition der Determinante verwendete Leibniz-Formel. (b) folgt dann aus $\det A = \det A^T$. □

Beispiel 3.42
Wir berechnen die folgende 4×4*-Determinante durch Entwicklung nach der 2. Zeile*

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 0 & 7 & 0 \\ 9 & 8 & 7 & 4 \\ 3 & 2 & 3 & 3 \end{vmatrix} = \underbrace{(-1)^{2+1} * 5 * \begin{vmatrix} 2 & 3 & 4 \\ 8 & 7 & 4 \\ 2 & 3 & 3 \end{vmatrix}}_{\text{2. Zeile, 1. Spalte gestrichen}} + \underbrace{(-1)^{2+2} * 0 * \begin{vmatrix} 1 & 3 & 4 \\ 9 & 7 & 4 \\ 3 & 3 & 3 \end{vmatrix}}_{\text{2. Zeile, 2. Spalte gestrichen}}$$

$$+ \underbrace{(-1)^{2+3} * 7 * \begin{vmatrix} 1 & 2 & 4 \\ 9 & 8 & 4 \\ 3 & 2 & 3 \end{vmatrix}}_{\text{2. Zeile, 3. Spalte gestrichen}} + \underbrace{(-1)^{2+4} * 0 * \begin{vmatrix} 1 & 2 & 3 \\ 9 & 8 & 7 \\ 3 & 2 & 3 \end{vmatrix}}_{\text{2. Zeile, 4. Spalte gestrichen}}.$$

Die beiden 3×3*-Determinanten, die mit Null multipliziert werden, müssen wir erst gar nicht berechnen. Für die anderen beiden* 3×3*-Determinanten erhalten wir jeweils mit der Regel von Sarrus*

$$\begin{vmatrix} 2 & 3 & 4 \\ 8 & 7 & 4 \\ 2 & 3 & 3 \end{vmatrix} = 2*7*3+3*4*2+4*8*3-4*7*2-2*4*3-3*8*3 = 10,$$

$$\begin{vmatrix} 1 & 2 & 4 \\ 9 & 8 & 4 \\ 3 & 2 & 3 \end{vmatrix} = 1*8*3+2*4*3+4*9*2-4*8*3-1*4*2-2*9*3 = -38.$$

Insgesamt ist also

$$\begin{vmatrix} 1 & 2 & 3 & 4 \\ 5 & 0 & 7 & 0 \\ 9 & 8 & 7 & 4 \\ 3 & 2 & 3 & 3 \end{vmatrix} = -5*10-7*(-38) = 216.$$

Bei der Anwendung des Entwicklungssatzes von Laplace ist es vorteilhaft eine Zeile oder Spalte auszuwählen, die möglichst viele Nullen enthält, damit möglichst wenige kleinere Determinanten ausgerechnet werden müssen. In dem Fall lässt sich damit auch die Determinante (etwas) größerer Matrizen berechnen. Im Allgemeinen wächst aber auch mit dem Entwicklungssatz von Laplace der zu leistende Rechenaufwand sehr schnell an (wiederum ca. wie $n!$) und die Berechnung der Determinante größerer Matrizen ist auch damit nicht praktikabel.

3.3.3 Rechenregeln und Interpretation als Volumenfunktion

Die Determinante erfüllt die folgenden Rechenregeln.

Satz 3.43
Für $A, B \in \mathbb{R}^{n\times n}$ gilt:

(a) Für eine rechte obere Dreiecksmatrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n-1} & a_{1,n} \\ 0 & a_{2,2} & \dots & a_{2,n-1} & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \dots & 0 & a_{n,n} \end{pmatrix}$$

ist die Determinante das Produkt der Hauptdiagonalelemente, d.h.

$$\det A = a_{1,1} * a_{2,2} * \ldots * a_{n,n}.$$

Gleiches gilt für linke untere Dreiecksmatrizen und insbesondere gilt auch

$$\det I = 1.$$

(b) $\det(AB) = \det(A) * \det(B)$

(c) Ist A invertierbar, dann gilt $\det A \neq 0$ *und* $\det(A^{-1}) = \frac{1}{\det A}$.

Beweis: (a) Die in der Leibniz-Formel gebildeten Produkte ergeben Null, wenn in einer Zeile ein Element links der Hauptdiagonale (also eine Null) gewählt wurde. Sie ergeben aber auch Null, wenn in einer Zeile ein Element rechts der Hauptdiagonale gewählt wurde. Geschieht dies nämlich in der i-ten Zeile erstmalig, dann muss ein Element in der i-ten Spalte in einer späteren Zeile gewählt werden und dort steht es dann links der Hauptdiagonale, ist also Null. Alle Produkte, bei denen nicht das Element auf der Hauptdiagonalen gewählt wurden ergeben also Null. Die in der Leibniz-Formel gebildeten Summe aller Produkte enthält nur das Produkt der Hauptdiagonalelemente als von Null verschiedenen Summanden (und zwar ohne Vorzeichenwechsel, da diese Auswahl keine Fehlstände besitzt).

(b) kann durch Einsetzen der Formel für das Matrixprodukt in die Leibniz-Formel hergeleitet werden.

(c) Ist A invertierbar, dann folgt aus (a) und (b)

$$1 = \det I = \det A^{-1}A = \det(A^{-1}) * \det(A).$$

Daher muss $\det(A) \neq 0$ sein und es folgt $\det(A^{-1}) = \frac{1}{\det A}$. □

Eine ähnlich einfache Formel für $\det(A+B)$ existiert nur für den Spezialfall, dass in A und B alle bis auf eine Zeile (oder alle bis auf eine Spalte) übereinstimmen.

Satz 3.44
Seien $(a_{ij}) \in \mathbb{R}^{n\times n}$, $(v_j) \in \mathbb{R}^n$, und $r \in \mathbb{R}$. Dann gilt:

(a) Die Determinante ist multilinear in den Zeilen, d.h. bezüglich jeder einzelnen Zeile linear. Für $i = 1,\dots,n$ ist

$$\begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & a_{i-1,2} & \dots & a_{i-1,n} \\ a_{i,1} & a_{i,2} & \dots & a_{i,n} \\ a_{i+1,1} & a_{i+1,2} & \dots & a_{i+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{vmatrix} + \begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & a_{i-1,2} & \dots & a_{i-1,n} \\ v_1 & v_2 & \dots & v_n \\ a_{i+1,1} & a_{i+1,2} & \dots & a_{i+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{vmatrix} = \begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & a_{i-1,2} & \dots & a_{i-1,n} \\ a_{i,1}+v_1 & a_{i,2}+v_2 & \dots & a_{i,n}+v_n \\ a_{i+1,1} & a_{i+1,2} & \dots & a_{i+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{vmatrix}$$

und

$$\begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & a_{i-1,2} & \dots & a_{i-1,n} \\ ra_{i,1} & ra_{i,2} & \dots & ra_{i,n} \\ a_{i+1,1} & a_{i+1,2} & \dots & a_{i+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{vmatrix} = r \begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & a_{i-1,2} & \dots & a_{i-1,n} \\ a_{i,1} & a_{i,2} & \dots & a_{i,n} \\ a_{i+1,1} & a_{i+1,2} & \dots & a_{i+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{vmatrix}.$$

(b) Die Determinante ist alternierend in den Zeilen, d.h. sie ist Null, falls zwei Zeilen identisch sind. Für $i, j = 1,\dots,n$ ist

$$\begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i-1,1} & a_{i-1,2} & \dots & a_{i-1,n} \\ v_1 & v_2 & \dots & v_n \\ a_{i+1,1} & a_{i+1,2} & \dots & a_{i+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{j-1,1} & a_{j-1,2} & \dots & a_{j-1,n} \\ v_1 & v_2 & \dots & v_n \\ a_{j+1,1} & a_{j+1,2} & \dots & a_{j+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{vmatrix} = 0.$$

(c) Die Determinante ist auch mulilinear und alternierend in den Spalten, d.h. analoge Formeln gelten spaltenweise.

Beweis: Die Aussagen lassen sich mit der Leibniz-Formel zeigen. Für (a) nutzt man dabei aus, dass das in der Leibniz-Formel gebildete Produkt über in jeder Zeile ausgewählte Elemente bezüglich eines einzelnen Faktors linear ist.

Für (b) nutzt man aus, dass die Auswahl des k-ten Elementes in der oberen der beiden identischen Zeilen und des l-ten Elements in der unteren dieser Zeilen das gleiche Produkt ergibt, wie die umgekehrter Auswahl (also des l-ten Elements in oberen und des k-ten in der unteren der identischen Zeilen). Man überlegt sich leicht, dass sich die Anzahl der Fehlstände zwischen diesen Auswahlmöglichkeiten um eine ungerade Zahl unterscheidet (nämlich gerade um eins plus zweimal die Anzahl der spalten- und zeilenweise zwischen diesen Elementen liegenden ausgewählten Matrixeinträgen). In der Summe der Leibniz-Formel kommt deshalb jedes Produkt zweimal vor, jedoch mit unterschiedlichem Vorzeichen, was sich gegenseitig ausgleicht und insgesamt Null ergibt.

(c) folgt wieder aus $\det A = \det A^T$. □

Beim Produkt einer Matrix $A \in \mathbb{R}^{n\times n}$ mit einer Zahl $r \in \mathbb{R}$ wird jeder Eintrag von A (und damit jede der n-Zeilen) mit r multipliziert. Es gilt also

$$\det(rA) = r^n \det(A).$$

Bemerkung 3.45
Der gerichtete Flächeninhalt des Parallelograms, das von zwei Vektoren aufgespannt wird ist ebenfalls multilinear und alternierend bezüglich dieser zwei Vektoren, vgl. die in der Vorlesung gemalten Skizzen. Analoges gilt für das (gerichtete) Volumen eines dreidimensionale Volumen eines Spats, der von drei Vektoren des $\mathbb{R}^3$ aufgespannt wird und es erscheint natürlich, diese Eigenschaft auch von einer Verallgemeinerung des Volumensbegriffs im n-dimensionalen zu fordern. Eine multilineare und alternierende Abbildung

$$\underbrace{\mathbb{R}^n \times \ldots \times \mathbb{R}^n}_{n\text{-mal}} \to \mathbb{R}$$

heißt deshalb auch abstrakte Volumenfunktion.

Schreiben wir n-Vektoren aus dem $\mathbb{R}^n$ spalten- oder zeilenweise in eine Matrix, und wenden darauf die Determinante an, so ist dies wegen Satz 3.44 eine abstrakte Volumenfunktion. Man kann zeigen, dass die Determinante die einzige abstrakte Volumenfunktion ist, die dem von den n-Einheitsvektoren aufgespannten Körper das Volumen 1 zuordnet (letzteres folgt aus Satz 3.43). Jede andere abstrakte Volumenfunktion unterscheidet sich von der Determinante nur um diesen Normierungsfaktor.

Insbesondere ist die Determinante einer 2×2-Matrix der Flächeninhalt des von den beiden Spaltenvektoren aufgespannten Parallelograms und die Determinante einer 3×3-Matrix ist das Volumen des von den drei Spaltenvektoren aufgespannten Spats. Gleiches gilt bezüglich der Zeilenvektoren.

Bemerkung 3.46
Mit Hilfe von Satz 3.44 kann man zeigen:

(a) Die Addition eines Vielfachens einer Zeile zu einer anderen Zeile ändert den Wert der Determinante nicht.

(b) Die Vertauschung zweier Zeilen ändert das Vorzeichen der Determinante.

(c) Beides gilt auch analog für die Spalten der Matrix.

Durch Ausnutzung von (a) und (b) lässt sich durch Vertauschungen und zeilenweisen Additionen wie beim Gauß-Algorithmus zur LGS-Lösung die Berechnung einer Determinante auf die Determinate einer Dreiecksmatrix zurückführen, und dort kann sie dann einfach als Produkt der Hauptdiagonaleinträge abgelesen werden. Man kann zeigen, dass so die Determinantenberechnung einer $n \times n$-Matrix nur ca. n^3 Rechenschritte erfordert.

In Satz 3.43(c) hatten wir bereits gezeigt, dass invertierbare Matrizen eine von Null verschiedene Determinante besitzen. Mit den skizzierten Argumenten zur Zeilenumformung lässt sich auch die Rückrichtung zeigen. Es gilt Es gilt

$$A \text{ invertierbar} \quad \Longleftrightarrow \quad \det(A) \neq 0.$$

Die Determinante von A zeigt, ob ein LGS mit Koeffizientenmatrix A eindeutig lösbar ist oder nicht.

3.4 Eigenwerte und Singulärwerte

3.4.1 Eigenwerte und charakteristisches Polynom

Definition 3.47
Sei $A \in \mathbb{R}^{n\times n}$ eine quadratische Matrix. Eine Zahl $\lambda \in \mathbb{R}$ heißt Eigenwert der Matrix A, falls ein $0 \neq v \in \mathbb{R}^n$ existiert mit

$$Av = \lambda v.$$

v heißt ein zum Eigenwert λ zugehöriger Eigenvektor.

Beispiel 3.48
(a) Für die Einheitsmatrix $I \in \mathbb{R}^{n\times n}$ gilt

$$Iv = 1v \quad \text{für alle } v \in \mathbb{R}^n.$$

I besitzt also den Eigenwert 1 und alle Vektoren $0 \neq v \in \mathbb{R}^n$ sind zugehörige Eigenvektoren.

(b) Für eine Diagonalmatrix

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \ddots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_n \end{pmatrix}$$

gilt $De_j = d_j e_j$. Die Diagonaleinträge $d_1, \ldots, d_n$ sind also Eigenwerte von D und die Einheitsvektoren $e_1, \ldots, e_n \in \mathbb{R}^n$ sind zugehörige Eigenvektoren. Offenbar sind auch ae_j für $0 \neq a \in \mathbb{R}$ ein Eigenvektor zu d_j und im Falle zweier übereinstimmender Diagonaleinträge $d_j = d_k$ ist auch $e_j + e_k$ (sowie jede andere nicht-triviale Linearkombination) ein Eigenvektor zu d_j.

(c) Für die 2×2-Matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

gilt

$$A\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{und} \quad A\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

A besitzt also den Eigenwert 2 und $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ist ein zugehöriger Eigenvektor, und A besitzt außerdem den Eigenwert 0 und $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ ist ein dazugehöriger Eigenvektor. In beiden Fällen sind auch wieder alle nicht-trivialen Vielfachen von $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ bzw. $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ zugehörige Eigenvektoren.

Definition und Satz 3.49
$\lambda \in \mathbb{R}$ ist genau dann Eigenwert der Matrix $A \in \mathbb{R}^{n \times n}$, wenn

$$\det(\lambda I - A) = 0.$$

Die Eigenwerte von A sind also genau die Nullstellen der Abbildung

$$p_A : \mathbb{R} \to \mathbb{R}, \quad x \mapsto p(x) := \det(xI - A).$$

p_A ist ein Polynom vom Grad n und heißt charakteristisches Polynom.

Beweis: Ist $\lambda \in \mathbb{R}$ ein Eigenwert von A, dann existiert $0 \neq v \in \mathbb{R}^n$, so dass $Av = \lambda v$. Es ist also $(\lambda I - A)v = \lambda v - Av = 0$ und damit $v \in \mathcal{N}(\lambda I - A)$. Dies

zeigt $\mathcal{N}(\lambda I - A) \neq 0$, nach Folgerung 3.32 ist also die Matrix $\lambda I - A$ nicht injektiv und deshalb (nach Bemerkung 3.46) $\det(\lambda I - A) = 0$.

Ist umgekehrt $\det(\lambda I - A) = 0$ dann ist (wiederum nach nach Bemerkung 3.46) die Matrix $\lambda I - A$ nicht bijektiv. Da $\lambda I - A$ eine quadratische Matrix ist, folgt aus Folgerung 3.32, dass sie nicht injektiv ist und daher $\mathcal{N}(\lambda I - A) \neq 0$ gilt. Es gibt also ein $0 \neq v \in \mathcal{N}(\lambda I - A)$ und dieses erfüllt dann $(\lambda I - A)v = 0$, also

$$Av = \lambda v.$$

Die Einträge von $xI - A$ sind Polynome vom Grad 1. Die Leibniz-Formel (3.4) zur Berechnung von $p_A(x) = \det(xI - A)$ enthält daher Summanden, die jeweils Produkte aus n Polynomen vom Grad 1 sind. Insgesamt ist deshalb $p_A \in \Pi_n$. □

Beispiel 3.50
(a) Für

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

ist

$$p_A(x) = \det(xI - A) = \begin{vmatrix} x-1 & -1 \\ -1 & x-1 \end{vmatrix} = (x-1)^2 - 1.$$

Es ist

$$0 = p_A(x) = (x-1)^2 - 1 \iff 1 = (x-1)^2 \iff x - 1 = \pm 1 \iff x \in \{0, 2\}.$$

Die Eigenwerte von A sind also 0 *und* 2.

(b) Für

$$B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

erhalten wir analog $p_B(x) = (x-1)^2 + 1$ *und dieses besitzt keine reellen Nullstellen. Mit einer in der Folgevorlesung eingeführten Erweiterung der reellen Zahlen zu den sogenannten komplexen Zahlen, lassen sich für B jedoch komplexe Eigenwerte angeben.*

3.4.2 Eigenwert- und Singulärwertzerlegung

Wir stellen noch eine in der Informatik sehr wichtige Anwendung von Eigenwerten vor.

Die Eigenwertzerlegung symmetrischer Matrizen. Eine quadratische Matrix $A \in \mathbb{R}^{n\times n}$ mit $A = A^T$ heißt symmetrisch. Man kann zeigen, dass für eine symmetrische Matrix n reelle (möglicherweise gleiche) Eigenwerte

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

und zugehörige Eigenvektoren

$$v_1, \ldots, v_n \in \mathbb{R}^n, \quad \text{mit} \quad Av_j = \lambda_j v_j,$$

existieren, so dass die Vektoren $v_1, \ldots, v_n$ eine ONB des $\mathbb{R}^n$ bilden, d.h.

$$v_j^T v_k = \langle v_j, v_k \rangle = \begin{cases} 0 & \text{für } j \neq k, \\ 1 & \text{für } j = k, \end{cases}$$

Jedes $x \in \mathbb{R}^n$ lässt sich in diese ONB entwickeln, also

$$x = x_1 v_1 + x_2 v_2 + \ldots + x_n v_n \quad \text{mit} \quad x_1, \ldots, x_n \in \mathbb{R}.$$

Aus der ONB-Eigenschaft folgt

$$v_j^T x = v_j^T (x_1 v_1 + x_2 v_2 + \ldots + x_n v_n) = x_j.$$

Daher ist

$$\sum_{j=1}^{n} \lambda_j v_j v_j^T x = \sum_{j=1}^{n} \lambda_j x_j v_j = \sum_{j=1}^{n} A(x_j v_j) = A\left(\sum_{j=1}^{n} x_j v_j\right) = Ax.$$

Dies gilt für alle $x \in \mathbb{R}^n$, so dass

$$A = \sum_{j=1}^{n} \lambda_j v_j v_j^T \in \mathbb{R}^{n\times n}.$$

Man nennt diese Darstellung die Eigenwertzerlegung oder Diagonalisierung der symmetrischen Matrix A und schreibt sie auch in der Form

$$A = VDV^T \quad \text{wobei} \quad D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \in \mathbb{R}^{n\times n},$$

und $V = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \in \mathbb{R}^{n\times n}$ spaltenweise die Eigenvektoren $v_1, \ldots, v_n \in \mathbb{R}^n$ enthält. Nach Definition und Satz 3.38(d) ist V unitär und $V^T = V^{-1}$. Insbesondere

ist A invertierbar genau dann, wenn alle Eigenwerte von Null verschieden sind und in dem Fall gilt

$$A^{-1} = VD^{-1}V^T \quad \text{mit} \quad D^{-1} = \begin{pmatrix} \frac{1}{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_n} \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Auch für manche nicht-symmetrische Matrizen lässt sich so eine Eigenwertzerlegung angeben. Dies ist (selbst nach Einführung komplexer Zahlen) jedoch nicht immer möglich. Beispielsweise besitzt

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

nur den Eigenwert 0, deshalb wäre in obiger Darstellung die Matrix D und damit auch die Matrix VDV^T immer die Nullmatrix.

Die Singulärwertzerlegung allgemeiner Matrizen. Auch für den allgemeinen Fall möglicherweiser nicht-symmetrischer oder nicht-quadratischer Matrizen lässt sich eine weitgehend analoge Zerlegung herleiten. Diese beruht darauf, dass für jedes $A \in \mathbb{R}^{m\times n}$ die Matrix $A^TA \in \mathbb{R}^{n\times n}$ quadratisch ist und wegen Satz 3.26(d)

$$(A^TA)^T = A^T(A^T)^T = A^TA,$$

die Matrix $A^TA \in \mathbb{R}^{n\times n}$ auch symmetrisch ist. Es gibt daher n Eigenwerte

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$$

von A^TA und eine zugehörige ONB aus Eigenvektoren des $\mathbb{R}^n$

$$v_1, \ldots, v_n \in \mathbb{R}^n, \quad \text{mit} \quad A^TAv_j = \lambda_j v_j.$$

Für alle $j = 1, \ldots, n$ gilt

$$0 \leq \|Av_j\|_2^2 = (Av_j)^T(Av_j) = v_j^TA^TAv_j = v_j^T(\lambda_j v_j) = \lambda_j v_j^T v_j = \lambda_j,$$

die Eigenwerte λ_j sind also alle nicht-negativ, aber möglicherweise Null. Bezeichne p die Anzahl der von Null verschiedenen λ_j und definiere für diese $j = 1, \ldots, p$ Eigenwerte

$$\sigma_j := \sqrt{\lambda_j} > 0 \quad \text{und} \quad u_j := \frac{1}{\sigma_j} Av_j \in \mathbb{R}^m.$$

Die u_j sind orthonormiert, da

$$u_k^T u_j = \frac{1}{\sigma_j \sigma_k} v_k^T A^T A v_j = \frac{\lambda_j}{\sigma_j \sigma_k} v_k^T v_j = \begin{cases} 0 & \text{für } j \neq k, \\ 1 & \text{für } j = k. \end{cases}$$

Außerdem gilt

$$Av_j = \sigma_j u_j \quad \text{und} \quad A^T u_j = \sigma_j v_j \quad \text{für alle } j = 1, \ldots, p.$$

Für $k > p$ folgt aus $\lambda_k = 0$ auch

$$\|Av_k\|_2^2 = v_k A^T A v_k = \lambda_k v_k^T v_k = 0 \quad \text{und damit} \quad Av_k = 0.$$

Wie bei der Eigenwertzerlegung erhalten wir deshalb

$$A = \sum_{j=1}^{p} \sigma_j u_j v_j^T \in \mathbb{R}^{m \times n}$$

und dies können wir auch in Matrixform schreiben als

$$A = UDV^T \quad \text{wobei} \quad D = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_p \end{pmatrix} \in \mathbb{R}^{p \times p},$$

und $U \in \mathbb{R}^{m \times p}$ und $V \in \mathbb{R}^{n \times p}$ spaltenweise die Vektoren $u_1, \ldots, u_p \in \mathbb{R}^m$ bzw. $v_1, \ldots, v_p \in \mathbb{R}^n$ enthalten. Dies heißt auch Singulärwertzerlegung (SWZ)[2]

Eine unmittelbare Anwendung der SWZ besteht in der Beschleunigung der Matrixmultiplikation. Die Multiplikation einer Matrix $A \in \mathbb{R}^{m \times n}$ mit einem Vektor $x \in \mathbb{R}^n$ benötigt mn Multiplikationen (der Einfachheit halber zählen wir nur diese). Besitzt A nur wenige von Null verschiedene Singulärwerte $\sigma_1 \geq \ldots \geq \sigma_p > 0$ (oder lassen sich im Rahmen der benötigten Genauigkeit alle anderen auf Null setzen), so benötigen die Multiplikationen

$$x \mapsto V^T x \mapsto DV^T x \mapsto UDV^T x$$

nur $np + p + mp$ Multiplikationen (und sogar nur $(n+m)p$, falls die Multiplikation UD vorberechnet wird) und liefern das gleiche Ergebnis $Ax = UDV^T x$ bzw. eine gute Approximation falls kleine Singulärwerte auf Null gesetzt wurden.

[2] Dies ist die sogenannte *ökonomische Variante* der SWZ. Für manche Anwendungen ist es zweckmäßig V mit den Spalten $v_{p+1}, \ldots, v_n$ zu einer $\mathbb{R}^{n \times n}$-Matrix zu ergänzen, D mit Nullen zu einer $\mathbb{R}^{m \times n}$-Matrix aufzufüllen, und die u_j zu einer ONB des ganzen $\mathbb{R}^m$ zu ergänzen und so U zu einer $\mathbb{R}^{m \times m}$-Matrix zu erweitern. U und V sind dann jeweils quadratisch und unitär.

Die SWZ liefert außerdem ein einfaches Verfahren zur Bildkomprimierung. Enthält $A \in \mathbb{R}^{m\times n}$ die Werte eines Graustufenbildes mit $m \times n$-Pixeln (oder eines Farbkanal eines Farbbildes), und werden alle bis auf p Singulärwerte von A durch 0 ersetzt, so müssen zur Speicherung von U, V und D nur $mp+np+p$ Zahlen gespeichert werden. In der folgenden Abbildung komprimieren wir so ein Logo der Größe 600×326 an und verwenden nur die größten 50 Singulärwerte. Die Anzahl der (bei naiver pixelweiser Abspeicherung) zu speichernden Zahlen reduziert sich dabei auf ca. 24% des ursprünglichen Wertes.

Original

$p = 50$ Singulärwerte