Mathematatik I für Studierende der Informatik

(Diskrete Mathematik)

– Vorlesungsskript WiSe 2024/25 –

Stefan Geschke und Mathias Schacht
Fachbereich Mathematik, Universität Hamburg

Vorwort

Dies ist das Skript für die Vorlesung *Mathematik I für Studierende der Informatik (Diskrete Mathematik)* des Wintersemesters 2024/25. Dieses Skript orientiert sich an dem Skript vom Thomas Andreae aus dem Wintersemester 2013/14 zur gleichnamigen Vorlesung und wurde mehrfach von Stefan Geschke und Mathias Schacht überarbeitet. Ziel der Vorlesung ist die Vermittlung allgemeiner mathematischer Grundlagen und Beweistechniken. Die folgenden Themen werden besprochen:

- Grundlagen der Mathematik und Logik,
- die natürlichen Zahlen und vollständige Induktion,
- Elementare Zahlentheorie,
- Elementare Kombinatorik,
- Restklassenringe und das RSA-Verschlüsselungsverfahren,
- Algebraische Strukturen (Gruppen, Ringe und Körper),
- Polynome,
- Vektoren und Matrizen.

Im Unterschied zu früheren Semestern lassen wir die Graphentheorie aus und stellen den Aufbau der Vorlesung etwas um. Im Laufe des Semesters werden sich einige Stellen im Skript noch ändern.

Hamburg, Herbst 2024 Stefan Geschke und Mathias Schacht

Ergänzende Literatur

[1] M. Aigner, *Diskrete Mathematik*, 5th ed., Vieweg Studium: Aufbaukurs Mathematik, Friedr. Vieweg & Sohn, Wiesbaden, 2004.

[2] G. Fischer, *Lineare Algebra*, 18th ed., Grundkurs Mathematik: Eine Einführung für Studienanfänger, Springer, 2014.

[3] J. Matoušek and J. Nešetřil, *Diskrete Mathematik: Eine Entdeckungsreise*, 2nd ed., Springer, 2007.

[4] A. Steger, *Diskrete Strukturen, Band 1: Kombinatorik, Graphentheorie, Algebra*, 2nd ed., Springer, 2007.

[5] G. Teschl and S. Teschl, *Mathematik für Informatiker, Band 1: Diskrete Mathematik und Lineare Algebra*, 4th ed., Springer, 2013.

Inhaltsverzeichnis

KAPITEL 1

Mathematische Grundlagen und Logik

§1.1 Mengen

Definition 1.1. *Eine* Menge *ist eine Zusammenfassung bestimmter, wohlunterschiedener Objekte, die die* Elemente *der Menge genannt werden.*

Bei Mengen kommt es nicht auf die Reihenfolge der Elemente an. Auch können Elemente in einer Menge nicht mehrfach vorkommen. Eine Menge ist durch ihre Elemente eindeutig bestimmt. Daher schreiben wir $A = B$ für zwei Mengen A und B, wenn A und B dieselben Elemente haben.

Definition 1.2. *Ist x ein Element der Menge M, so schreiben wir $x \in M$. $x \notin M$ bedeutet, dass x kein Element von M ist. Sind A und B Mengen, so schreiben wir $A \subseteq B$, wenn A eine* Teilmenge *von B ist, also wenn jedes Element von A auch Element von B ist. Die (eindeutig bestimmte) Menge, die keine Elemente hat, heißt die* leere Menge. *Sie wird als $\{\}$ oder $\varnothing$ notiert.*

Mengen kann man notieren, indem man ihre Elemente in geschweiften Klammern angibt. $\{4, 7, 13\}$ bezeichnet zum Beispiel die Menge, deren Elemente genau die Zahlen 4, 7 und 13 sind. Da es nur auf die Elemente selbst und nicht auf deren Reihenfolge ankommt, bezeichnen $\{3, 4, 5\}$ und $\{4, 5, 3\}$ dieselbe Menge. Wenn ein Element mehrfach genannt wird, so wird das ignoriert, da eine Menge jedes Element nur einmal enthält. Daher bezeichnen $\{1, 2, 1, 1\}$ und $\{1, 2\}$ dieselbe Menge. $\mathbb{Z} = \{\dots, -1, 0, 1, 2, \dots\}$ ist die Menge der ganzen Zahlen. $\mathbb{N}$ ist die Menge $\{1, 2, 3, \dots\}$ der natürlichen Zahlen. Viele Autoren lassen die natürlichen Zahlen bei 0 anfangen. Wir definieren $\mathbb{N}_0$ als die Menge der natürlichen Zahlen zusammen mit der 0, also $\mathbb{N}_0 = \{0, 1, 2, \dots\}$.

$$\{n \colon n \text{ ist eine natürliche Zahl mit } 5 < n < 10\}$$

ist die Menge der natürlichen Zahlen, die echt größer als 5 und echt kleiner als 10 sind, also die Menge $\{6, 7, 8, 9\}$. Auf diese Weise kann man auch unendliche Mengen notieren. So ist

$$\{n \colon n \text{ ist eine durch 2 teilbare natürliche Zahl}\}$$

die Menge der geraden natürlichen Zahlen.

§1.2 Elementare Logik

Definition 1.3. *Eine* Aussage *ist ein Satz, von dem man im Prinzip eindeutig feststellen kann, ob er wahr oder falsch ist. Ob eine Aussage wahr oder falsch ist, ist der* Wahrheitswert *der Aussage. Der Wahrheitswert „wahr“ wird dabei oft mit „w“ oder „1“ abgekürzt, der Wahrheitswert „falsch“ mit „f“ oder „0“.*

Der Satz „Die Straße ist nass“ ist eine Aussage. Ebenso sind „$2 + 5 = 7$“ und „$2 + 5 < 3$“ Aussagen, wobei die erste wahr und die zweite falsch ist. „Guten Abend!“ ist keine Aussage. Ebenso ist „$n^2 = 4$“ keine Aussage, da wir nicht feststellen können, ob diese Formel wahr oder falsch ist, solange wir nicht wissen, was n ist.

Aussagen können mit den logischen Verknüpfungen „und“, „oder“ und „nicht“ verknüpft werden. Allerdings ist die Bedeutung dieser Wörter in der Umgangssprache nicht immer ganz eindeutig. Daher ist es sinnvoll, diese Verknüpfungen für formale Zwecke zu präzisieren.

Definition 1.4. *Ist a eine Aussage, so ist die* Negation *von a die Aussage, die genau dann wahr ist, wenn a falsch ist. Die Negation von a wird $\neg a$ geschrieben und „nicht a“ gelesen. Sind a und b Aussagen, so ist die* Konjunktion *von a und b die Aussage, die genau dann wahr ist, wenn sowohl a als auch b wahr ist. Die Konjunktion von a und b wird $a \wedge b$ geschrieben und „a und b“ gelesen. Die* Disjunktion *von a und b ist die Aussage, die genau dann wahr ist, wenn mindestens eine der Aussagen a und b wahr ist. Die Disjunktion von a und b wird $a \vee b$ geschrieben und „a oder b“ gelesen.*

Den Wahrheitswert einer durch logische Verknüpfungen aus anderen Aussagen gebildeten Aussage in Abhängigkeit der Wahrheitswerte der Ausgangsaussagen kann man in Form einer *Wahrheitstafel* beschreiben:

a	$\neg a$
0	1
1	0

a	b	$a \wedge b$	$a \vee b$
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	1

Definition 1.5. *Weitere wichtige logische Verknüpfungen sind die* Implikation $\rightarrow$, *die* Äquivalenz $\leftrightarrow$ *und das exklusive Oder* xor. *Wir definieren diese Verknüpfungen mit Hilfe einer Wahrheitstafel.*

a	b	$a \rightarrow b$	$a \leftrightarrow b$	*xor*
0	*0*	*1*	*1*	*0*
0	*1*	*1*	*0*	*1*
1	*0*	*0*	*0*	*1*
1	*1*	*1*	*1*	*0*

Die Aussage $a \to b$ ist also immer dann wahr, wenn a falsch ist oder b wahr. Ist $a \to b$ wahr, so sagen wir „b folgt aus a“ oder „a impliziert b“. Die Aussage $a \leftrightarrow b$ ist immer dann wahr, wenn a und b entweder beide falsch oder beide wahr sind. Ist $a \leftrightarrow b$ wahr, so nennen wir a und b äquivalent. Die Zeichen $\to$ und $\leftrightarrow$ werden normalerweise nur in formalen Ausdrücken verwendet, während wir im normalen mathematischen Text $\Rightarrow$ und $\Leftrightarrow$ benutzen. Ein klassisches Beispiel ist die Aussage „wenn es regnet, ist die Straße nass“, die sich mit Hilfe von $\Rightarrow$ so schreiben lässt:

$$\text{Es regnet} \Rightarrow \text{Die Straße ist nass.}$$

(Wir ignorieren in diesem Beispiel das Problem, dass die Wahrheitswerte von „es regnet“ und „die Straße ist nass“ natürlich von Ort und Zeitpunkt abhängen. Wir können uns zum Beispiel vorstellen, dass wir Ort und Zeit schon fest gewählt haben.) Die Aussage $a \operatorname{xor} b$ ist genau dann wahr, wenn die Wahrheitswerte von a und b unterschiedlich sind.

Mit Hilfe von Wahrheitstafeln können wir die Wahrheitswerte komplizierterer Aussagen untersuchen, die durch Verknüpfungen einfacherer Aussagen entstanden sind. Seien zum Beispiel a, b und c Aussagen und e die Aussage $a \wedge (b \vee c)$. Falls die Wahrheitswerte von a, b und c bekannt sind, so können wir zunächst den Wahrheitswert von $b \vee c$ bestimmen und dann den von $a \wedge (b \vee c)$. Auf diese Weise erhält man folgende Wahrheitstafel:

a	b	c	$b \vee c$	$a \wedge (b \vee c)$
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	1	0
1	0	0	0	0
1	0	1	1	1
1	1	0	1	1
1	1	1	1	1

Wenn man eine entsprechende Wahrheitstafel für $(a \wedge b) \vee (a \wedge c)$ aufstellt, sieht man, dass $a \wedge (b \vee c)$ und $(a \wedge b) \vee (a \wedge c)$ äquivalent sind, unabhängig davon, welche Wahrheitswerte die Aussagen a, b und c haben. Auf diese Weise lassen sich Rechenregeln für $\vee$, $\wedge$ und $\neg$ nachweisen. Das ist das *Wahrheitstafelverfahren*. Wir halten zunächst folgenden Satz fest:

Satz 1.6. *Sind a, b und c Aussagen, so ist $a \wedge (b \vee c)$ äquivalent zu $(a \wedge b) \vee (a \wedge c)$.*

Eine weitere wichtige Regel ist die sogenannte *Kontraposition*, die man oft in Beweisen anwenden kann.

Satz 1.7. *Seien a und b Aussagen. Die Aussage $a \to b$ ist äquivalent zu $\neg b \to \neg a$.*

Beweis. Wir schreiben die entsprechende Wahrheitstafel auf.

a	b	$\neg a$	$\neg b$	$a \to b$	$\neg b \to \neg a$
0	0	1	1	1	1
0	1	1	0	1	1
1	0	0	1	0	0
1	1	0	0	1	1

Wie man leicht abliest, sind $a \to b$ und $\neg b \to \neg a$ in der Tat äquivalent. □

Beispiel 1.8. Der Satz „wenn es neblig ist, ist die Sicht schlecht“ ist äquivalent zu „wenn die Sicht nicht schlecht ist, dann ist es nicht neblig“.

Unter dem Stichwort „Boolesche Algebra“ werden wir später noch weitere Rechenregeln für logische Verknüpfungen festhalten.

Definition 1.9. *Eine* Aussageform *ist eine Aussage, in der eine Konstante durch eine Variable ersetzt wurde. So erhält man aus einer Aussage a eine Aussageform $a(x)$.*

„$2 + 5 = 7$“ ist eine Aussage. Daraus lässt sich zum Beispiel die Aussageform „$2 + x = 7$“ ableiten. Sei $a(x)$ diese Aussageform. Ein Wahrheitswert von $a(x)$ lässt sich nicht angeben, da wir nicht wissen, welchen Wert x hat. Wenn wir für x einen Wert einsetzen, dann erhalten wir wieder eine Aussage. So ist $a(5)$, also die ursprüngliche Aussage, wahr, während $a(2)$, also die Aussage „$2 + 2 = 7$“, falsch ist.

Auch Aussageformen können mittels logischer Verknüpfungen verknüpft werden. Ist $a(x)$ die Aussageform „$2+x \leqslant 7$“, so ist $\neg a(x)$ die Aussageform „$2+x \nleqslant 7$“ oder, anders geschrieben, „$2 + x > 7$“. Ist $a(x)$ die Aussageform „$x = 2$“ und $b(x)$ die Aussageform „$x^2 = 4$“, so verstehen wir, was „$a(x) \Rightarrow b(x)$“ bedeutet:

$$\text{Wenn } x = 2 \text{ ist, so ist } x^2 = 4.$$

Setzen wir für x konkrete natürliche Zahlen ein, so erhalten wir immer eine wahre Aussage. Mit anderen Worten, die Aussage

$$\text{Für alle natürlichen Zahlen } x \text{ gilt: } a(x) \Rightarrow b(x)$$

ist wahr. Den Satzteil „für alle natürlichen Zahlen x“ nennen wir einen *Quantor*. Mit Hilfe von Quantoren können wir aus Aussageformen wieder Aussagen machen.

Definition 1.10. *Sei $a(x)$ eine Aussageform und M eine Menge. Dann ist*

$$(\exists x \in M)a(x)$$

die Aussage, die genau dann wahr ist, wenn es mindestens ein Element x der Menge M gibt, so dass $a(x)$ gilt. $(\exists x \in M)a(x)$ wird „es gibt ein x in M mit $a(x)$“ gelesen. Das Zeichen $\exists$ ist der Existenzquantor*.*

$$(\forall x \in M)a(x)$$

ist die Aussage, die genau dann wahr ist, wenn $a(x)$ für alle Elemente x der Menge M gilt. $(\forall x \in M)a(x)$ *wird „für alle x in M gilt $a(x)$" gelesen. Das Zeichen $\forall$ ist der* Allquantor*.*

Im Zusammenhang mit Quantoren, und auch sonst, werden wir Klammern immer so setzen, beziehungsweise weglassen, dass die Lesbarkeit optimal ist.

Ein typisches Beispiel einer *Existenzaussage*, also einer Aussage, die mit einem Existenzquantor beginnt, ist die Aussage $\exists x \in \mathbb{N}(x^2 = 4)$. Ein typisches Beispiel einer *Allaussage*, also einer Aussage, die mit einem Allquantor beginnt, ist die Aussage $\forall x \in \mathbb{N}(x^2 > 0)$.

Oft betrachten wir Aussageformen wie „$(n+1)^2 = n^2+2n+1$". Bei dieser Aussageform ist klar, dass für n eine Zahl eingesetzt werden soll, und nicht anderes. Außerdem steht die Variable n üblicherweise für eine natürliche Zahl. Unsere Erfahrung sagt uns also, dass wir, wenn wir „$(n + 1)^2 = n^2 + 2n + 1$" hinschreiben, oft eigentlich „$\forall n \in \mathbb{N}((n + 1)^2 = n^2 + 2n + 1)$" meinen.

Die Negation $\neg(\forall x \in M)a(x)$ der Allaussage $(\forall x \in M)a(x)$ ist äquivalent zu der Existenzaussage $(\exists x \in M)\neg a(x)$. Das wird an einem Beispiel schnell klar: „Alle Autos in Hamburg sind blau" ist sicher falsch, es gilt vielmehr „nicht alle Autos in Hamburg sind blau", was äquivalent zu der Aussage „es gibt in Hamburg (mindestens) ein Auto, das nicht blau ist" ist. Analog ist $\neg(\exists x \in M)a(x)$ zu $(\forall x \in M)\neg a(x)$ äquivalent.

§1.3 Mengenoperationen

Wir definieren einige Verknüpfungen von Mengen, mit denen sich ganz ähnlich rechnen lässt wie mit den Verknüpfungen $\wedge$, $\vee$ und $\neg$ von Aussagen. Die Rechengesetze, die für die logischen Verknüpfungen (von Aussagen) und für die entsprechenden Verknüpfungen von Mengen gelten, fasst man unter dem Begriff „Boolesche Algebra" zusammen.

Definition 1.11. *Seien A und B Mengen. Dann ist die* Vereinigung *von A und B definiert als*

$$A \cup B := \{x\colon x \in A \vee x \in B\}.$$

(Hier benutzen wir das Zeichen := um auszudrücken, dass es sich um eine Definition handelt.) Der Schnitt *oder* Durchschnitt *von A und B ist die Menge*

$$A \cap B := \{x\colon x \in A \wedge x \in B\}.$$

Zwei Mengen A und B heißen disjunkt, *falls $A \cap B = \varnothing$. Die* mengentheoretische Differenz *von A und B ist die Menge*

$$A \setminus B := \{x \in A\colon x \notin B\}.$$

Schon anhand der Definition von $\cup$ und $\cap$ sieht man, dass $\cup$ etwas mit $\vee$ zu tun hat und $\cap$ mit $\wedge$. Und in der Tat verhalten sich $\cap$ und $\cup$ ähnlich wie $\wedge$ und $\vee$. Eine Operation auf Mengen, die sich analog zur Negation verhält, ist die Komplementbildung.

Definition 1.12. *Für eine Menge M sei*

$$\wp(M) := \{x \colon x \subseteq M\}$$

die Potenzmenge *von M. Wir fixieren M und betrachten nur Teilmengen von M. Für $A \in \wp(M)$ sei*

$$\overline{A} := \{x \in M \colon x \notin A\}$$

das Komplement *von A in M.*

Wir stellen fest, dass $\wp(M)$ unter $\cup$, $\cap$ und Komplementbildung *abgeschlossen* ist. D.h., für alle $A, B \in \wp(M)$ sind $A \cap B$, $A \cup B$ und $\overline{A}$ wieder Elemente von $\wp(M)$.

Rechenregeln für die Mengenoperationen $\cap$, $\cup$ und Komplementbildung können wir wieder mit dem Wahrheitstafelverfahren herleiten. Seien zum Beispiel A, B und C Teilmengen einer Menge M.

Satz 1.13. *Es gilt $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.*

Beweis. Wir wissen schon, dass $A \cap (B \cup C)$ und $(A \cap B) \cup (A \cap C)$ Teilmengen von M sind. Also müssen wir nur zeigen, dass die beiden Mengen genau dieselben Elemente von M enthalten.

Es gilt

$$A \cap (B \cup C) = \{x \in M \colon x \in A \wedge (x \in B \vee x \in C)\}$$

sowie

$$(A \cap B) \cup (A \cap C) = \{x \in M \colon (x \in A \wedge x \in B) \vee (x \in A \wedge x \in C)\}.$$

Wir fixieren nun ein beliebiges Element x von M. Sei a die Aussage $x \in A$, b die Aussage $x \in B$ und c die Aussage $x \in C$. Man beachte, dass wir hier so tun, als wären a, b und c Aussagen, da wir das x vorher fixiert haben und wir es jetzt wie eine Konstante behandeln können.

Nach Satz 1.6 sind $a \wedge (b \vee c)$ und $(a \wedge b) \vee (a \wedge c)$ äquivalent. Damit gilt

$$x \in A \cap (B \cup C) \Leftrightarrow a \wedge (b \vee c) \Leftrightarrow (a \wedge b) \vee (a \wedge c) \Leftrightarrow x \in (A \cap B) \cup (A \cap C)$$

Also haben $A \cap (B \cup C)$ und $(A \cap B) \cup (A \cap C)$ dieselben Elemente und sind damit gleich. □

Wir haben bisher die Frage nach der Gleichheit zweier Mengen auf die Frage zurückgeführt, ob zwei Aussagen äquivalent sind. Die letztere Frage ließ sich mit Hilfe des Wahrheitstafelverfahrens klären. Damit lässt sich das Wahrheitstafelverfahren manchmal einsetzen, um die Gleichheit zweier Mengen nachzuweisen. Im allgemeinen ist es allerdings meistens ratsam, die Gleichheit zweier Mengen A und B nachzurechnen, indem man zunächst $A \subseteq B$ und dann $B \subseteq A$ zeigt.

Beispiel 1.14. Wir beweisen die Gleichung $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ ohne das Wahrheitstafelverfahren. Als erstes zeigen wir $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$. Dazu müssen wir zeigen, dass jedes Element von $A \cap (B \cup C)$ auch ein Element von $(A \cap B) \cup (A \cap C)$ ist.

Sei also $x \in A \cap (B \cup C)$. Dann ist x sowohl in A als auch in $B \cup C$ enthalten. Also ist x in B oder in C enthalten. Ist x in B enthalten, so gilt $x \in A \cap B$. Ist x in C enthalten, so gilt $x \in A \cap C$. Damit ist x in $A \cap B$ oder in $A \cap C$ enthalten. Also gilt $x \in (A \cap B) \cup (A \cap C)$.

Das zeigt $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$. Wir zeigen nun $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$.

Sei $x \in (A \cap B) \cup (A \cap C)$. Dann ist x in $A \cap B$ oder in $A \cap C$ enthalten. Wir nehmen zunächst an, dass $x \in A \cap B$ gilt. Dann ist x in A und in B enthalten. Damit ist x aber auch in $B \cup C$ enthalten. Es folgt $x \in A \cap (B \cup C)$.

Nun nehmen wir an, dass $x \in A \cap C$ gilt. Wie eben sehen wir, dass $x \in A \cap (B \cup C)$ gilt.

Also gilt $x \in A \cap (B \cup C)$ unabhängig davon, ob x ein Element von $A \cap B$ oder $A \cap C$ ist.

Das zeigt $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. Insgesamt folgt nun die Gleichheit $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Definition 1.15. *Sind A und B Mengen, so bezeichnet man mit $A \times B$ die Menge*

$$\{(a, b) \colon a \in A \text{ und } b \in B\}$$

aller geordneten Paare (a, b), deren erste Komponente a ein Element von A ist und deren zweite Komponente b ein Element von B ist. $A \times B$ heißt das kartesische Produkt *der Mengen A und B. Mit A^2 bezeichnet man die Menge $A \times A$.*

*A^3 ist die Menge $\{(a_1, a_2, a_3) \colon a_1, a_2, a_3 \in A\}$ aller Tripel von Elementen von A. Analog ist für jede natürliche Zahl $n \geqslant 1$ A^n die Menge $\{(a_1, \ldots, a_n) \colon a_1, \ldots, a_n \in A\}$ aller n-*Tupel *von Elementen von A.*

Zum Beispiel ist

$$\{1, 2, 3\} \times \{4, 5\} = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}.$$

§1.4 Abbildungen

Definition 1.16. *Eine* Abbildung *von einer Menge A in eine Menge B ist eine Zuordnung, die jedem Element von A ein Element von B zuordnet. Abbildungen werden oft auch* Funktionen *genannt. Ist f eine Abbildung von A nach B, so schreiben wir*

$f\colon A \longrightarrow B$. Dabei wird A der Definitionsbereich *von f genannt und B der* Wertevorrat. *Auch die Begriffe* Vorbereich *für A und* Nachbereich *für B sind sinnvoll.*

Für jedes $a \in A$ bezeichnen wir mit $f(a)$ das Element von B, das die Funktion f dem Element a zuordnet. Falls f einem Element $a \in A$ also $b \in B$ zuordnet, so schreiben wir $f(a) = b$ und sagen „f bildet a auf b ab“. Das Element b heißt der Wert *oder der* Funktionswert *von f an der Stelle a. Man kann anstelle von $f(a) = b$ auch $a \longmapsto b$ schreiben, wenn klar ist, welche Funktion f gemeint ist.*

Das Bild *von f ist die Menge $\{f(x) \colon x \in A\}$.*

Der Name *Wertebereich* wird von manchen Autoren für das Bild einer Funktion verwendet und von anderen für den Wertevorrat. Um Missverständnissen vorzubeugen, verwenden wir diesen Begriff gar nicht.

Beispiel 1.17. (1) Eine Funktion f von der Menge $\mathbb{N}$ der natürlichen Zahlen in die natürlichen Zahlen kann zum Beispiel durch eine Formel gegeben sein: $f(n) = n^2$. Ein Schreibweise, die alle wesentlichen Informationen beinhaltet, wäre dann

$$f\colon \mathbb{N} \longrightarrow \mathbb{N}; n \longmapsto n^2.$$

(2) Der Ausdruck $g\colon \mathbb{N}^2 \longrightarrow \mathbb{N}, (m,n) \longmapsto m+n$ beschreibt eine Funktion von der Menge der Paare natürlicher Zahlen in die Menge der natürlichen Zahlen, die der Gleichung $g((m,n)) = m+n$ genügt. Anstelle von $g((m,n))$ schreiben wir auch $g(m,n)$.

(3) Funktionen mit endlichem Definitionsbereich kann man auch in Form einer Tabelle angeben. Sei zum Beispiel $A = \{1,2,3,4,5\}$ und $B = \{q,w,e,r,t,z\}$. Dann definiert die folgende Tabelle die Funktion $f\colon A \longrightarrow B$:

a	1	2	3	4	5
$f(a)$	w	q	t	w	e

Es gilt nun $f(1) = w$, $f(2) = q$ und so weiter.

Definition 1.18. *Eine Abbildung $f\colon A \longrightarrow B$ heißt*

(1) injektiv, *falls für alle $x, y \in A$ gilt: Ist $x \neq y$, so ist $f(x) \neq f(y)$.*
(2) surjektiv, *falls es für alle $b \in B$ mindestens ein $a \in A$ gibt, so dass $f(a) = b$ gilt.*
(3) bijektiv, *falls sie injektiv und surjektiv ist.*

Beispiel 1.19. (1) Sei $A = \{1,2,3\}$ und $B = \{1,2,3\}$. Die Abbildung $f\colon A \longrightarrow B$ mit $f(1) = 1$, $f(2) = 1$ und $f(3) = 2$ ist weder injektiv noch surjektiv.

(2) Seien A und B wie in (1). Die Funktion $g\colon A \longrightarrow B$ mit $g(1) = 2$, $g(2) = 3$ und $g(3) = 1$ ist sowohl injektiv als auch surjektiv, also bijektiv.

(3) Sei wieder $A = \{1,2,3\}$ aber $B = \{3,7\}$. Die Abbildung $f\colon A \longrightarrow B$ mit $f(1) = 3$, $f(2) = 7$ und $f(3) = 3$ ist surjektiv, aber nicht injektiv.

(4) Sei nun A wie in (1) – (3) und $B = \{1, 2, 3, 4\}$. Die Funktion $f\colon A \longrightarrow B$ mit $f(1) = 2$, $f(2) = 1$, $f(3) = 4$ ist injektiv, aber nicht surjektiv.

(5) Die Abbildung $h\colon \mathbb{N} \longrightarrow \mathbb{N}; n \longmapsto n^2$ ist nicht surjektiv, da es zum Beispiel kein $a \in \mathbb{N}$ gibt, für das $h(a) = 3$ gilt.

Das kann man wie folgt einsehen: Angenommen, es gäbe doch ein $a \in \mathbb{N}$ mit $h(a) = a^2 = 3$. Dann ist a entweder $\sqrt{3}$ oder $-\sqrt{3}$. Beide Zahlen, $\sqrt{3}$ und $-\sqrt{3}$, sind aber keine Elemente von $\mathbb{N}$. Das widerspricht der Annahme $a \in \mathbb{N}$.

Eine andere Möglichkeit zu zeigen, dass 3 nicht im Bild von f liegt ist die folgende: Es gelten $1^2 = 1 < 3$ und $2^2 = 4 > 3$. Für alle $n \geqslant 2$ ist $n^2 \geqslant 2^2$ und damit $n^2 > 3$. Damit gibt es kein $n \in \mathbb{N}$ mit $n^2 = 3$.

Die Abbildung h ist aber injektiv. Seien nämlich $x, y \in \mathbb{N}$ mit $x \neq y$, dann ist entweder $x < y$ oder $y < x$. Wir betrachten nur den ersten Fall, der zweite Fall kann genauso behandelt werden. Wir nehmen also $x < y$ an. (Später werden wir in so einer Situation zum Beispiel schreiben: „Ohne Beschränkung der Allgemeinheit (o.B.d.A.) können wir $x < y$ annehmen“.) Sei $a = y - x$. Dann ist $y = x + a$ und $y^2 = x^2 + 2xa + a^2$. Wegen $x, a > 0$ gilt $2xa + a^2 > 0$ und damit ist $y^2 > x^2$. Insbesondere gilt

$$h(x) = x^2 \neq y^2 = h(y).$$

Das zeigt, dass h injektiv ist.

Definition 1.20. *Für eine natürliche Zahl n versteht man unter einer n-*stelligen Verknüpfung *oder einer n-*stelligen Operation *auf einer Menge M eine Abbildung $f\colon M^n \longrightarrow M$.*

Der wichtigste Spezialfall ist der einer *binären* Verknüpfung $f\colon M^2 \longrightarrow M$. Beispiele binärer Verknüpfungen sind die Addition $+\colon \mathbb{N}^2 \longrightarrow \mathbb{N}; (m, n) \longmapsto m + n$ und die Multiplikation $\cdot\colon \mathbb{N}^2 \longrightarrow \mathbb{N}; (m, n) \longmapsto m \cdot n$.

§1.5 Boolesche Algebra

Wir haben schon gesehen, dass sich die Mengenoperationen $\cap$, $\cup$ und Komplementbildung ganz analog zu den logischen Verknüpfungen $\wedge$, $\vee$ und $\neg$ verhalten. Und in der Tat kann man die Mengenoperationen und die logischen Verknüpfungen mit einem gemeinsamen Begriff beschreiben.

Definition 1.21. *Gegeben sei eine Menge B, die mindestens die zwei verschiedenen Elemente 1 und 0 enthält, zusammen mit der einstelligen Verknüpfung $\neg\colon B \longrightarrow B$ und den zwei zweistelligen Verknüpfungen $\sqcap, \sqcup\colon B^2 \longrightarrow B$. $(B, \sqcap, \sqcup, \neg, 0, 1)$ heißt eine* Boolesche Algebra, *wenn für alle $a, b, c \in B$ die folgenden Gleichungen gelten:*

(A1) Assoziativgesetze:
- $a \sqcap (b \sqcap c) = (a \sqcap b) \sqcap c$

- $a \sqcup (b \sqcup c) = (a \sqcup b) \sqcup c$

(A2) Kommutativgesetze:
- $a \sqcap b = b \sqcap a$
- $a \sqcup b = b \sqcup a$

(A3) Distributivgesetze:
- $a \sqcap (b \sqcup c) = (a \sqcap b) \sqcup (a \sqcap c)$
- $a \sqcup (b \sqcap c) = (a \sqcup b) \sqcap (a \sqcup c)$

(A4) Beschränkheit:
- $a \sqcap 1 = a$
- $a \sqcup 0 = a$

(A5) Komplementierung:
- $a \sqcap \neg a = 0$
- $a \sqcup \neg a = 1$

Die Aussagen (A1)–(A5) in Definition 1.21 sind die *Axiome* für Boolesche Algebren.

Beispiel 1.22. (1) Die *Schaltalgebra* ist die Menge $\{0, 1\}$ der Wahrheitswerte mit den Verknüpfungen $\wedge$, $\vee$ und $\neg$. Die Schaltalgebra ist eine Boolesche Algebra, wie man mit Hilfe des Wahrheitstafelverfahrens leicht nachrechnen kann.

(2) Ist $M \neq \varnothing$ eine Menge, so ist $\wp(M)$ mit den Verknüpfungen $\cap$, $\cup$ und Komplementbildung sowie den Konstanten $1 := M$ und $0 := \varnothing$ eine Boolesche Algebra, die *Potenzmengenalgebra* von M. Dass Potenzmengenalgebren wirklich Boolesche Algebren sind, folgt aus der Tatsache, dass die Schaltalgebra die Axiome einer Booleschen Algebra erfüllt, zusammen mit der Übersetzung von Fragen der Gleichheit von Mengen in Fragen der Äquivalenz von Aussagen, die wir oben schon diskutiert haben.

(3) Wir betrachten noch einen speziellen Fall, nämlich eine Boolesche Algebra, die im wesentlichen genau die Potenzmengenalgebra auf einer achtelementigen Menge ist, die wir aber anders aufschreiben. Es sei $B := \{w, f\}^8$, also die Menge aller 8-Tupel der Wahrheitswerte w und f. Man kann B zum Beispiel als Menge aller möglichen Bytes interpretieren. Weiter sei

$$1 := (w, w, w, w, w, w, w, w) \quad \text{und} \quad 0 := (f, f, f, f, f, f, f, f).$$

Die Operationen definieren wir jetzt wie folgt:

Für $a, b \in B$ mit $a = (a_1, \dots, a_8)$ und $b = (b_1, \dots, b_8)$ sei

$$a \sqcap b := (a_1 \wedge b_1, \dots, a_8 \wedge b_8),$$

$$a \sqcup b := (a_1 \vee b_1, \dots, a_8 \vee b_8)$$

und

$$\neg a := (\neg a_1, \dots, \neg a_8).$$

Dann ist $(B, \sqcap, \sqcup, \neg, 0, 1)$ eine Boolesche Algebra, wie man leicht nachrechnet.

Alle Aussagen, die sich aus (A1)–(A5) ableiten lassen, gelten für alle Booleschen Algebren, inbesondere also für die Schaltalgebra und alle Potenzmengenalgebren. Diese Allgemeinheit ist die Stärke der *axiomatischen Methode*, bei der Sätze aus Axiomen gefolgert werden und nicht nur für bestimmte Strukturen, wie zum Beispiel die natürlichen Zahlen oder eine bestimmte Boolesche Algebra, bewiesen werden.

Wir geben Beispiele für die axiomatische Methode und beweisen ein paar einfache Regeln für Boolesche Algebren. Sei $(B, \sqcap, \sqcup, \neg, 0, 1)$ eine Boolesche Algebra.

Satz 1.23. *Für alle $a \in B$ gilt $a \sqcap a = a$ und $a \sqcup a = a$.*

Beweis. Es gilt

$$a \sqcap a \overset{(A4)}{=} (a \sqcap a) \sqcup 0 \overset{(A5)}{=} (a \sqcap a) \sqcup (a \sqcap \neg a) \overset{(A3)}{=} a \sqcap (a \sqcup \neg a) \overset{(A5)}{=} a \sqcap 1 \overset{(A4)}{=} a.$$

Auf dieselbe Weise rechnen wir $a \sqcup a = a$ nach.

$$a \sqcup a \overset{(A4)}{=} (a \sqcup a) \sqcap 1 \overset{(A5)}{=} (a \sqcup a) \sqcap (a \sqcup \neg a) \overset{(A3)}{=} a \sqcup (a \sqcap \neg a) \overset{(A5)}{=} a \sqcup 0 \overset{(A4)}{=} a.$$

Damit haben wir die beiden Gleichung aus den Axiomen (A1)–(A5) hergeleitet. □

In diesem Beweis fällt auf, dass wir den Beweis der Gleichung $a \sqcap a = a$ in den Beweis der Gleichung $a \sqcup a = a$ übersetzen können, indem wir $\sqcap$ und $\sqcup$ vertauschen und ebenso 0 und 1. Das funktioniert, da die Axiome (A1)–(A5) aus Paaren von Gleichungen bestehen, die jeweils durch diese Vertauschungen auseinander hervorgehen.

Satz 1.24 (Dualitätsprinzip für Boolesche Algebren)**.** *Jede Aussage, die eine Folgerung aus den Axiomen (A1)–(A5) ist, geht in eine gültige Aussage über, wenn man in ihr überall die Zeichen $\sqcap$ und $\sqcup$ sowie die Zeichen 0 und 1 vertauscht.*

Satz 1.25. *Für alle $a \in B$ gilt $a \sqcap 0 = 0$ und $a \sqcup 1 = 1$.*

Beweis. Es gilt

$$a \sqcap 0 = a \sqcap (a \sqcap \neg a) = (a \sqcap a) \sqcap \neg a = a \sqcap \neg a = 0.$$

Die Behauptung $a \sqcap 1 = 1$ folgt aus $a \sqcap 0 = 0$ nach dem Dualitätsprinzip. □

Wir schließen diesen Abschnitt mit zwei wichtigen Regeln für Boolesche Algebren, die aus den Axiomen folgen, deren Beweis wir aber nicht angeben.

Satz 1.26 (De Morgansche Regeln)**.** *Für alle $a, b \in B$ gilt $\neg(a \sqcap b) = \neg a \sqcup \neg b$ und $\neg(a \sqcup b) = \neg a \sqcap \neg b$.*

Der Beweis der De Morganschen Regeln aus den Axiomen (A1)–(A5) ist deutlich aufwendiger als die Beweise der Sätze 1.23 und 1.25. Mit Hilfe des Wahrheitstafelverfahrens lassen sich die De Morganschen Regeln für die Schaltalgebra leicht nachrechen. Man kann zeigen, dass alle Gleichungen, wie zum Beispiel die De Morganschen Regeln, die in der Schaltalgebra gelten, auch in allen anderen Booleschen Algebren gelten. Damit kann das Wahrheitstafelverfahren für Gleichungen, in denen nur die Konstanten 0 und 1 auftreten, in beliebigen Booleschen Algebren eingesetzt werden.

§1.6 Relationen

In Definition 1.15 haben wir das kartesische Produkt $A \times B$ zweier Mengen A und B als die Menge aller Paare (a, b) mit $a \in A$ und $b \in B$ definiert.

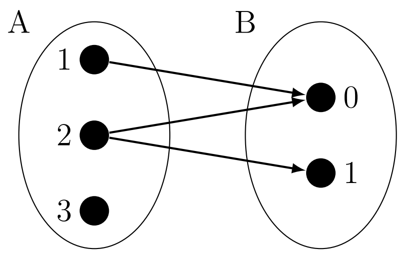
Definition 1.27. *Eine* Relation *von A nach B ist eine Teilmenge R von $A \times B$. Eine Relation auf A ist eine Teilmenge von $A \times A$. Für $(a, b) \in R$ schreiben wir auch aRb.*

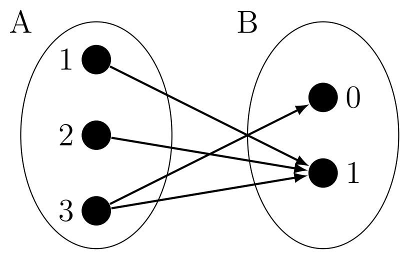
Beispiel 1.28.

(1) Sei $A = \{1, 2, 3\}$ und $B = \{0, 1\}$, dann sind $R_1, \ldots, R_4$ Relationen von A nach B:
  (a) $R_1 = \{(1, 0), (2, 0), (2, 1)\}$.
  (b) $R_2 = \{(1, 1), (2, 1), (3, 0), (3, 1)\}$
  (c) $R_3 = A \times B$
  (d) $R_4 = \varnothing$.

(2) $R = \{(a, b) \colon a, b \in \mathbb{N} \wedge a < b\}$, $S = \{(a, b) \colon a, b \in \mathbb{N} \wedge a \leqslant b\}$ und $T = \{(a, b) \colon a, b \in \mathbb{N} \wedge a = b\}$ sind Relationen auf $\mathbb{N}$. Üblicherweise identifizieren wir $<$ mit R, $\leqslant$ mit S und T mit $=$.

Wir können Relationen ähnlich wie Funktionen mit Hilfe von Pfeildiagrammen notieren. Hier sind zwei Diagramme für die Relationen R_1 und R_2.





Dass diese Pfeildiagramme von Relationen denen von Funktionen ähneln, ist kein Zufall. Wir können eine Funktion $f : A \longrightarrow B$ auch als eine Relation von A nach B auffassen, nämlich als die Relation

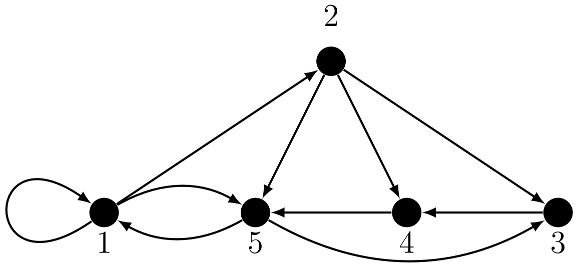
$$\{(a, b) : a \in A \wedge b \in B \wedge f(a) = b\}.$$

Eine Relation R von A nach B ist genau dann eine Funktion von A nach B, wenn es für jedes $a \in A$ genau ein $b \in B$ gibt, so dass $(a, b) \in R$ gilt. Das eindeutig bestimmte b mit $(a, b) \in R$ ist der Funktionswert der Funktion an der Stelle a.

Eine Relation R auf einer Menge A können wir als *gerichteten Graphen* darstellen, wobei für jedes Element von A ein Punkt gezeichnet wird und für jedes Paar $(a, b) \in R$ ein Pfeil von dem Punkt, der a entspricht zu dem, der b entspricht. Sei zum Beispiel $A = \{1, 2, 3, 4, 5\}$ und

$$R = \{(1, 1), (1, 2), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (4, 5), (5, 1), (5, 3)\}.$$

Dann sieht der entsprechende gerichtete Graph wie folgt aus:



Die Punkte 1, 2, 3, 4 und 5 nennt man die *Knoten* des Graphen. Einen Pfeil von einem Knoten zu einem Knoten nennt man auch eine *gerichtete Kante*. Eine Kante von einem Knoten auf sicher selber nennt man auch eine *Schlinge*.

Definition 1.29. *Sei A eine Menge und sei R eine Relation auf A.*

(1) R heißt reflexiv*, falls für alle $a \in A$ das Paar (a, a) in R ist.*
(2) R heißt irreflexiv*, falls R kein Paar der Form (a, a) enthält.*
(3) R heißt symmetrisch*, falls für alle $(a, b) \in R$ auch $(b, a) \in R$ gilt.*
(4) R heißt antisymmetrisch*, falls aus $(a, b) \in R$ und $a \neq b$ stets $(b, a) \notin R$ folgt.*
(5) R heißt transitiv*, falls aus $(a, b) \in R$ und $(b, c) \in R$ stets $(a, c) \in R$ folgt.*

Wir diskutieren die Bedeutung dieser Begriffe anhand der gerichteten Graphen, mit denen wir Relationen veranschaulichen.

Beispiel 1.30. Sei R eine Relation auf der Menge A.

(1) R ist reflexiv, falls jeder Knoten im zugehörigen gerichteten Graphen eine Schlinge hat.
(2) R ist irreflexiv, falls kein Knoten im zugehörigen gerichteten Graphen eine Schlinge hat.
(3) R ist symmetrisch, wenn im gerichteten Graphen für jeden Pfeil von a nach b auch der Pfeil zurück von b nach a vorhanden ist.
(4) R ist antisymmetrisch, wenn für je zwei verschiedene Knoten im gerichteten Graphen höchstens ein Pfeil zwischen den beiden Knoten a und b vorhanden ist.
(5) R ist transitiv, wenn für den gerichteten Graphen folgendes gilt: Immer wenn man entlang der Pfeile (in Pfeilrichtung) von einem Knoten a zu einem Knoten b laufen kann, dann ist bereits ein direkter Pfeil von a nach b vorhanden.

Man beachte, dass irreflexiv nicht dasselbe ist wie nicht reflexiv. Ebenso ist antisymmetrisch nicht dasselbe wie nicht symmetrisch.

Der Vollständigkeit halber sei hier noch erwähnt, dass in der Mathematik und auch in der Informatik oft (*ungerichtete*) *Graphen* betrachtet werden, die keine Schlingen haben und bei denen die Kanten keine Richtung haben. Es gibt also in einem ungerichteten

Graphen genau dann eine Kante von einem Knoten u zu einem Knoten v, wenn es eine Kante von v nach u gibt. Man könnte also einen ungerichteten Graphen als eine irreflexive, symmetrische Relation definieren.

1.6.1. Partitionen und Äquivalenzrelationen.

Definition 1.31. *Eine Relation R auf einer Menge A heißt Äquivalenzrelation, falls R reflexiv, transitiv und symmetrisch ist.*

Ist R eine Äquivalenzrelation auf A, so bezeichnen wir für jedes $a \in A$ mit $[a]_R$ die Menge $\{b \in A\colon (a,b) \in R\}$ und nennen diese Menge die Äquivalenzklasse *von a.*

Beispiel 1.32. Es sei A die Menge der Studierenden, die für diese Vorlesung angemeldet sind. Für zwei Studierende $a, b \in A$ gelte aRb genau dann, wenn a und b am selben Tag Geburtstag haben. Man überprüft schnell, dass R wirklich eine Äquivalenzrelation auf A ist. Die Äquivalenzklassen der Relation R sehen nun wie folgt aus:

Wir fixieren einen Tag, so dass es ein $a \in A$ gibt, so dass a tatsächlich an diesem Tag Geburtstag hat. Die Äquivalenzklasse $[a]_R$ besteht nun aus allen $b \in A$, die an dem einen fixierten Tag Geburtstag haben.

Satz 1.33. *Sei A eine Menge und R eine Äquivalenzrelation auf A. Dann gilt für alle $a, b \in A$ entweder $[a]_R \cap [b]_R = \varnothing$ oder $[a]_R = [b]_R$. Der zweite Fall tritt genau dann ein, wenn aRb gilt.*

Beweis. Seien $a, b \in A$ mit $[a]_R \cap [b]_R \neq \varnothing$. Sei $c \in [a]_R \cap [b]_R$. Dann gilt aRc und bRc. Wegen Symmetrie und Transitivität von R folgt daraus aRb. Wieder wegen Symmetrie und Transitivität von R ist jedes Element von A, das zu a äquivalent ist, auch zu b äquivalent und umgekehrt. Damit sind $[a]_R$ und $[b]_R$ gleich. $\square$

Für eine Äquivalenzrelation R auf einer Menge A ist $\{[a]_R\colon a \in A\}$ eine *Partition* von A.

Definition 1.34. *Sei A eine Menge, I eine Indexmenge und für alle $i \in I$ sei $K_i \subseteq A$. $P = \{K_i\colon i \in I\}$ ist eine* Partition *von A, falls gilt:*

(1) Für alle $i \in I$ ist $K_i \neq \varnothing$.
(2) Für alle $i, j \in I$ mit $i \neq j$ ist $K_i \cap K_j = \varnothing$.
(3) Es gilt $\bigcup_{i \in I} K_i = A$.

Dabei ist $\bigcup_{i \in I} K_i$ die Menge $\{x\colon \exists i \in I (x \in K_i)\}$.

In Beispiel 1.32 sieht die Partition, die zu der Äquivalenzrelation R gehört wie folgt aus: Für jeden Tag, der als Geburtstag einer Person $a \in A$ auftritt, ist die Menge aller $b \in A$, die an diesem Tag Geburtstag haben, eine der Mengen in der Partition. Die Mengen der Partition sind also mit den möglichen Geburtstagen indiziert. Die Partition unterteilt also A in die Mengen der Studierenden, die an einem festen Tag Geburtstag haben. Falls ein bestimmter Tag nicht als Geburtstag auftritt, so tritt die entsprechende Menge von Studierenden, in diesem Fall die leere Menge, nicht in der Partition auf.

Umgekehrt kann man einer Partition $P = \{K_i \colon i \in I\}$ von A eine Äquivalenzrelation auf A zuordnen, deren Äquivalenzklassen genau die Mengen K_i sind. Sei nämlich $P = \{K_i \colon i \in I\}$ eine Partition von A. Sei

$$R := \{(a,b) \in A \times A \colon \exists i \in I(a, b \in K_i)\}.$$

Wir nennen also zwei Elemente a und b von A äquivalent, wenn sie in derselben Menge K_i liegen.

Wegen $\bigcup_{i\in I} K_i = A$ gibt es für jedes $a \in A$ ein $i \in I$ mit $a \in K_i$. Damit steht jedes $a \in A$ zu sich selbst in Relation. R ist also reflexiv. Gilt $a, b \in K_i$, so gilt auch $b, a \in K_i$. Damit ist R symmetrisch. Seien schließlich $a, b, c \in A$ mit aRb und bRc. Dann gibt es $i, j \in I$ mit $a, b \in K_i$ und $b, c \in K_j$. Nun gilt $b \in K_i \cap K_j$. Da die Mengen in der Partition paarweise disjunkt sind, muss $K_i = K_j$ gelten. Also gilt $a, c \in K_i$. Damit ist aRc. Das zeigt die Transitivität von R.

Korollar 1.35. *Es sei A eine Menge. Für jede Äquivalenzrelation auf A bilden die Äquivalenzklassen eine Partition von A. Umgekehrt gibt es für jede Partition von A eine Äquivalenzrelation, deren Äquivalenzklassen genau die Mengen in der Partition sind.*

Beispiel 1.36. Wir hatten bereits die Menge

$$\mathbb{Z} = \{\dots, -1, 0, 1, 2, \dots\}$$

der ganzen Zahlen eingeführt. Für $a, b \in \mathbb{Z}$ und $m \in \mathbb{N}$ schreiben wir $a \equiv b(\text{mod } m)$, wenn bei der Division von a und b durch m derselbe Rest auftritt. So ist zum Beispiel $7 \equiv 12(\text{mod } 5)$, weil 7 und 12 beide den Rest 2 bei Division durch 5 haben.

Sei $m \in \mathbb{N}$ und $R = \{(a,b) \in \mathbb{Z} \times \mathbb{Z} \colon a \equiv b(\text{mod } m)\}$. Dann ist R eine Äquivalenzrelation auf $\mathbb{Z}$, deren Äquivalenzklassen *Restklassen modulo m* genannt werden.

Die Anzahl der Restklassen modulo m ist genau m. Die verschiedenen Restklassen sind die Mengen

$$\{m \cdot q + 0 \colon q \in \mathbb{Z}\}, \quad \{m \cdot q + 1 \colon q \in \mathbb{Z}\}, \quad \dots, \quad \{m \cdot q + (m-1) \colon q \in \mathbb{Z}\}.$$

Diese Mengen bilden eine Partition von $\mathbb{Z}$.

Die Menge $\mathbb{Q}$ der *rationalen Zahlen* ist die Menge aller Brüche $\frac{m}{n}$ mit $m, n \in \mathbb{Z}$ und $n \neq 0$. Dabei sind zwei Brüche gleich, wenn sie durch erweitern oder kürzen auseinander hervorgehen.

Beispiel 1.37. Wir können die rationalen Zahlen wie folgt konstruieren:

Betrachte zunächst die Menge

$$M = \{(m,n) \in \mathbb{Z} \times \mathbb{Z} : n \neq 0\}.$$

Ein Paar (m, n) steht dabei für den Bruch $\frac{m}{n}$.

Nun identifizieren wir Brüche, die durch Erweitern oder Kürzen auseinander hervorgehen. Dazu sei $(m, n) \sim (m', n')$, falls $mn' = m'n$ gilt. (Wenn man schon mit Brüchen rechnen kann, dann sieht man, dass $mn' = m'n$ genau dann gilt, wenn $\frac{m}{n} = \frac{m'}{n'}$ ist.)

Man rechnet leicht nach, dass $\sim$ eine Äquivalenzrelation auf M ist. Nun definieren wir $\mathbb{Q}$ als die Menge der Äquivalenzklassen der Relation $\sim$ auf M. Es ist also

$$\mathbb{Q} = \{[(m, n)]_\sim : (m, n) \in \mathbb{Z} \times \mathbb{Z} \wedge n \neq 0\}.$$

Für die Äquivalenzklasse von (m, n) schreiben wir $\frac{m}{n}$. Jetzt müsste man eigentlich noch die Operationen $+$, $\cdot$, $-$ und $:$ auf $\mathbb{Q}$ definieren. Das sparen wir uns aber aus Zeitgründen.

Nachdem die rationalen Zahlen einmal korrekt eingeführt wurden, kann man die Details dieser Implementation wieder vergessen. Es kommt nur auf die Rechenregeln an, die wir im Abschnitt über Zahlentheorie besprechen.

Die Menge der Äquivalenzklassen einer Äquivalenzrelation $\sim$ auf einer Menge M nennt man auch die *Faktormenge* oder den *Quotienten* von M nach der Relation $\sim$. Diese Menge der Äquivalenzklassen wird mit $M/\sim$ bezeichnet.

Ist M endlich mit m Elementen und haben alle $\sim$-Äquivalenzklassen gleichviele Elemente, zum Beispiel n Stück, so hat $M/\sim$ genau m/n Elemente.

Beispiel 1.38. Auch die ganzen Zahlen können wir als Faktormenge konstruieren:

Betrachte zunächst die Menge

$$M = \mathbb{N}_0 \times \mathbb{N}_0.$$

Ein Paar (m, n) steht dabei für die Differenz $m - n$.

Nun identifizieren wir Paare $(m, n), (m', n') \in M$, die dieselben Differenzen beschreiben. Dazu sei $(m, n) \sim (m', n')$, falls $m + n' = m' + n$ gilt. (Wenn wir schon mit Differenzen rechnen können, sehen wir, dass $m + n' = m' + n$ genau dann gilt, wenn $m - n = m' - n'$ ist.)

Man rechnet leicht nach, dass $\sim$ eine Äquivalenzrelation auf M ist. Nun definieren wir $\mathbb{Z}$ als die Menge der Äquivalenzklassen der Relation $\sim$ auf M, also als die Faktormenge $M/\sim$.

Die Addition auf $\mathbb{Z}$ wird dann wie folgt definiert:

$$[(m, n)] + [(m', n')] := [(m + m', n + n')]$$

Das entspricht der Rechenregel

$$(m - n) + (m' - n') = (m + m') - (n + n').$$

1.6.2. Ordnungsrelationen.

Definition 1.39. *Sei A eine Menge und R eine Relation auf A. Dann ist R eine* Ordnungsrelation*, falls R reflexiv, antisymmetrisch und transitiv ist. Ordnungsrelationen nennt man auch* Halbordnungen *oder* partielle Ordnungen*. Das Paar (A, R) ist eine* halbgeordnete *oder* partiell geordnete Menge*.*

Ordnungsrelationen werden oft mit $\leqslant$ oder einem ähnlichen Zeichen bezeichnet. Man schreibt dann praktisch immer $a \leqslant b$ anstelle von $(a, b) \in \leqslant$. Man beachte, dass dabei nicht unbedingt die bekannte $\leqslant$-Relation auf den reellen Zahlen gemeint ist.

Beispiel 1.40. Sei $A := \{a, b, c, d\}$ und

$$R := \{(a, a), (b, b), (c, c), (d, d), (a, b), (a, c), (a, d), (b, d), (c, d)\}.$$

Wie man leicht sieht, ist R reflexiv, transitiv und antisymmetrisch.

Beispiel 1.41. Sei $A := \{a, b, c, d\}$ und

$$R := \{(a, a), (b, b), (c, c), (d, d), (a, b), (a, c), (a, d), (b, c), (b, d), (c, d)\}.$$

Wieder sieht man leicht, dass R reflexiv, transitiv und antisymmetrisch ist.

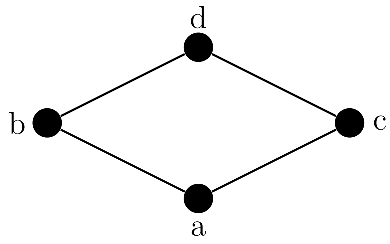
Beispiel 1.42.
(1) Die Relation $\leqslant$ ist eine Ordnungsrelation auf $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ und auf der Menge $\mathbb{R}$ der reellen Zahlen.
(2) Für jede Menge M ist $\subseteq$ eine Ordnungsrelation auf $\wp(M)$.
(3) Die Teilbarkeitsrelation $\mid$ ist eine Ordnungsrelation auf $\mathbb{N}$. Da $a \mid b$ genau dann, wenn es ein $k \in \mathbb{N}$ gibt, so dass $b = ka$ gilt, wenn also b ein Vielfaches von a ist.

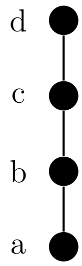
Definition 1.43. *Eine Ordnungsrelation R auf einer Menge R heißt* lineare Ordnung*, falls für alle $a, b \in A$ mit $a \neq b$ entweder aRb oder bRa gilt. Lineare Ordnungen nennt man auch* totale Ordnungen*.*

Beispiel 1.44. Die Relation $\leqslant$ auf $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ und $\mathbb{R}$ ist jeweils eine lineare Ordnung. Die Relation R aus Beispiel 1.41 ist ebenfalls eine lineare Ordnung, während die Relation aus Beispiel 1.40 keine lineare Ordnung ist, da die Elemente b und c nicht *vergleichbar* sind, also da weder (b, c) noch (c, b) in R ist. Ebenso ist $\subseteq$ keine lineare Ordnung auf $\wp(M)$, falls M mindestens zwei Elemente hat.

Wir betrachten noch einmal die Beispiele 1.40 und 1.41. Wenn man von einer Relation R auf einer Menge A schon weiß, dass es sich um eine Ordnungsrelation handelt, dann kann man in dem gerichteten Graphen die Schlingen an den einzelnen Knoten weglassen sowie gerichtete Kanten, deren Existenz aus der Transitivität der Relation folgt. Schließlich können wir noch vereinbaren, dass Kanten immer nach oben zeigen, so dass wir die Pfeilspitzen weglassen können. Diese Darstellung nennt man ein *Hassediagramm* einer geordneten Menge.

Folgende Diagramme sind Hassediagramme der Relationen in Beispiel 1.40 und 1.41.





Insbesondere sieht man, dass die Knoten im Falle der linearen Ordnung auf einer Linie angeordnet sind.

§1.7 Hüllenbildungen

Sei R eine Relation auf einer Menge A. Falls R nicht bereits reflexiv ist, so kann man R zu einer reflexiven Relation R' machen, indem man für jedes $a \in A$ das Paar (a, a) zu R hinzufügt.

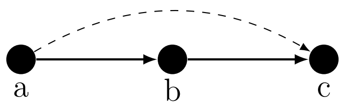
Definition 1.45. *Für eine Relation R auf einer Menge A sei*

$$R' := R \cup \{(a, a) \colon a \in A\}.$$

R' ist die kleinste reflexive Relation, die R umfasst, und wird die reflexive Hülle *von R genannt.*

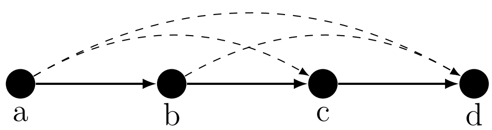
Sei zum Beispiel $<$ die übliche $<$-Relation auf $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ oder $\mathbb{R}$. Dann ist die Relation $\leqslant$ auf derselben Menge die reflexive Hülle von $<$.

Auf ähnliche Weise können wir aus einer Relation R eine transitive Relation machen. Sei $A = \{a, b, c\}$ und $R = \{(a, b), (b, c)\}$.



Damit R transitiv wird, müssen wir das Paar (a, c) zu R hinzufügen.

Wir betrachten noch die folgende, etwas kompliziertere Situation. Sei $A = \{a, b, c, d\}$ und $R = \{(a, b), (b, c), (c, d)\}$.



Hier müssen wir zunächst (a, c) und (b, d) zu R hinzufügen. Aber die Relation $R \cup \{(a, c), (b, d)\}$ ist immer noch nicht transitiv, denn obwohl

$$(a, b), (b, d) \in R \cup \{(a, c), (b, d)\}$$

gilt, ist das Paar (a, d) nicht in der Relation $R \cup \{(a, c), (b, d)\}$ enthalten. Wenn wir jedoch auch noch (a, d) hinzufügen, so erhalten wir eine transitive Relation.

Im Allgemeinen gilt für eine transitive Relation R: Falls

$$(a_1, a_2), \ldots, (a_{n-1}, a_n) \in R$$

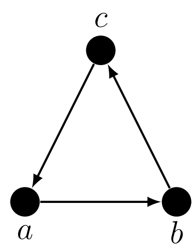
gilt, so ist auch $(a_1, a_n) \in R$. Das erklärt die folgende Definition:

Definition 1.46. *Sei R eine Relation auf einer Menge A. Dann ist*

$$R^+ := \{(a, b) : \ \textit{es gibt } n \geqslant 2 \textit{ und } a_1, \ldots, a_n \in A \textit{ mit}$$
$$a = a_1,\ b = a_n \textit{ und } (a_1, a_2), \ldots, (a_{n-1}, a_n) \in R\}$$

die kleinste transitive Relation mit $R \subseteq R^+$. R^+ ist die transitive Hülle *von R.*

Man sieht schnell, dass R^+ transitiv ist. Man beachte, dass es durchaus vorkommen kann, dass $(a_1, a_2), \ldots, (a_{n-1}, a_n) \in R$ gilt und dabei $a_1 = a_n$ ist. So ist die transitive Hülle der Relation $R = \{(a, b), (b, c), (c, a)\}$ auf der Menge A die Relation $R^+ = A \times A$.

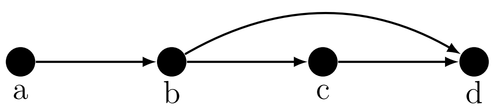


Schließlich kombinieren wir noch die transitive und die reflexive Hülle.

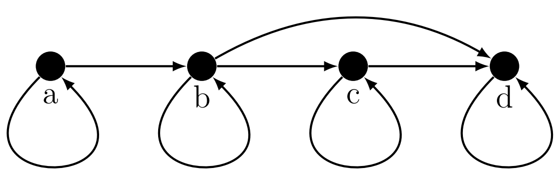
Definition 1.47. *Sei R eine Relation auf einer Menge A. Dann ist $R^* = R^+ \cup R'$ die* reflexive, transitive Hülle *von R. R^* ist die kleinste reflexive, transitive Relation, die R umfasst.*

Beispiel 1.48. Sei $A = \{a, b, c, d\}$ und $R = \{(a, b), (b, c), (c, d), (b, d)\}$. Wir geben die reflexive Hülle, die transitive Hülle und die reflexive, transitive Hülle von R an.

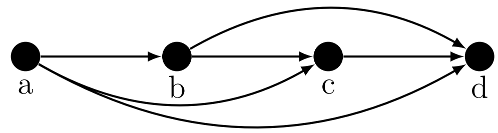
$R = \{(a, b), (b, c), (c, d), (b, d)\}$



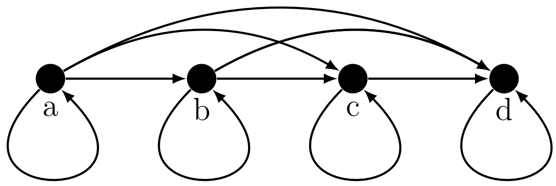
$R' = \{(a, a), (b, b), (c, c), (d, d), (a, b), (b, c), (c, d), (b, d)\}$



$R^+ = \{(a, b), (a, c), (a, d), (b, c), (b, d), (c, d)\}$



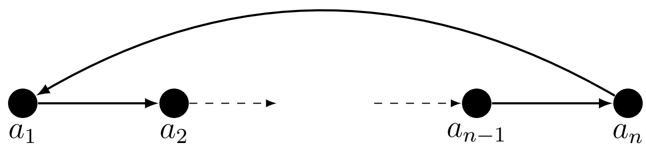
$$R^* = \{(a,a),(b,b),(c,c),(d,d),(a,b),(a,c),(a,d),(b,c),(b,d),(c,d)$$



Die reflexive, transitive Hülle R^* einer Relation R ist immer reflexiv und transitiv. Aber R^* muss natürlich nicht antisymmetrisch sein. Da reflexive, transitive Relationen aber relativ häufig vorkommen, bekommen sie einen eigenen Namen.

Definition 1.49. *Eine reflexive, transitive Relation heißt* Quasiordnung.

Die reflexive, transitive Hülle einer Relation ist also immer eine Quasiordnung, aber nicht unbedingt eine Ordnungsrelation. Es stellt sich heraus, dass R^* genau dann eine Ordnungsrelation ist, wenn es in R keine Kreise der Form



mit $n \geqslant 2$ gibt.

§1.8 Mehrstellige Relationen

In Definition 1.15 hatten wir schon kartesische Produkte der Form A^n betrachtet. Analog können wir auch kartesische Produkte zwischen verschiedenen Mengen definieren.

Definition 1.50. *Sei $n \geqslant 1$ und seien $A_1 \dots, A_n$ Mengen. Dann ist*

$$A_1 \times \ldots \times A_n = \{(a_1, \ldots, a_n) \colon a_1 \in A_1 \wedge \cdots \wedge a_n \in A_n\}$$

das kartesische Produkt *der Mengen $A_1, \ldots, A_n$.*

*Eine n-*stellige Relation über *$A_1, \ldots, A_n$ ist eine Teilmenge R des Produkts $A_1 \times \ldots \times A_n$. Eine n-*stellige Relation auf *einer Menge A ist eine Teilmenge R von A^n.*

Im vorigen Abschnitt haben wir nur *binäre*, also zweistellige Relationen diskutiert. Einstellige Relationen auf einer Menge A sind einfach Teilmengen der Menge A.

Beispiel 1.51. Seien $A = \{1,2,3\}$, $B = \{0,1\}$ und $C = \{2,3\}$. Dann sind $R_1 = \varnothing$, $R_2 = \{(2,0,2)\}$, $R_3 = \{(1,0,2),(1,1,2),(2,1,3)\}$ und $R_4 = A \times B \times C$ Relationen über A, B und C.

§1.9 Mehr über Abbildungen

Definition 1.52. *Seien A und B Mengen und $f\colon A \longrightarrow B$ eine Abbildung. Für $A' \subseteq A$ ist die Menge*

$$f[A'] = \{b \in B\colon \exists a \in A'(f(a) = b)\} = \{f(a)\colon a \in A'\}$$

das Bild von A' unter f. *Anstelle von $f[A']$ schreibt man auch $f(A')$.*

Für $B' \subseteq B$ ist die Menge

$$f^{-1}[B'] = \{a \in A\colon f(a) \in B'\}$$

das Urbild von B' unter f.

Beispiel 1.53. Sei $A = \{1, 2, 3, 4, 5\}$ und $B = \{0, 1, 2\}$. Weiter sei $f\colon A \longrightarrow B$ definiert durch $f(1) = f(2) = 0$, $f(3) = f(5) = 1$ und $f(4) = 2$. Schließlich seien $A' = \{3, 4, 5\}$ und $B' = \{0, 2\}$. Dann gilt $f[A'] = \{1, 2\}$ und $f^{-1}[B'] = \{1, 2, 4\}$.

Satz 1.54. *Es seien A und B Mengen und $f\colon A \longrightarrow B$ eine Funktion. Für alle $A_1, A_2 \subseteq A$ und $B_1, B_2 \subseteq B$ gelten die folgenden Aussagen:*

(1) $f[A_1 \cap A_2] \subseteq f[A_1] \cap f[A_2]$
(2) $f[A_1 \cup A_2] = f[A_1] \cup f[A_2]$
(3) $f^{-1}[B_1 \cap B_2] = f^{-1}[B_1] \cap f^{-1}[B_2]$
(4) $f^{-1}[B_1 \cup B_2] = f^{-1}[B_1] \cup f^{-1}[B_2]$
(5) $f^{-1}[f[A_1]] \supseteq A_1$
(6) $f[f^{-1}[B_1]] \subseteq B_1$

Beweis. Wir zeigen (1), (3) und (5) und lassen (2), (4) und (6) als Übungen.

(1) Sei $b \in f[A_1 \cap A_2]$. Dann existiert $a \in A_1 \cap A_2$ mit $f(a) = b$. Wegen $a \in A_1$ gilt $b = f(a) \in f[A_1]$. Wegen $a \in A_2$ gilt $b = f(a) \in f[A_2]$. Also ist $b \in f[A_1] \cap f[A_2]$. Damit gilt $f[A_1 \cap A_2] \subseteq f[A_1] \cap f[A_2]$.

(3) Sei $a \in f^{-1}[B_1 \cap B_2]$. Dann gilt $f(a) \in B_1 \cap B_2$. Also ist $f(a) \in B_1$ und $f(a) \in B_2$. Damit ist $a \in f^{-1}[B_1]$ und $a \in f^{-1}[B_2]$. Es folgt $a \in f^{-1}[B_1] \cap f^{-1}[B_2]$. Das zeigt $f^{-1}[B_1 \cap B_2] \subseteq f^{-1}[B_1] \cap f^{-1}[B_2]$.

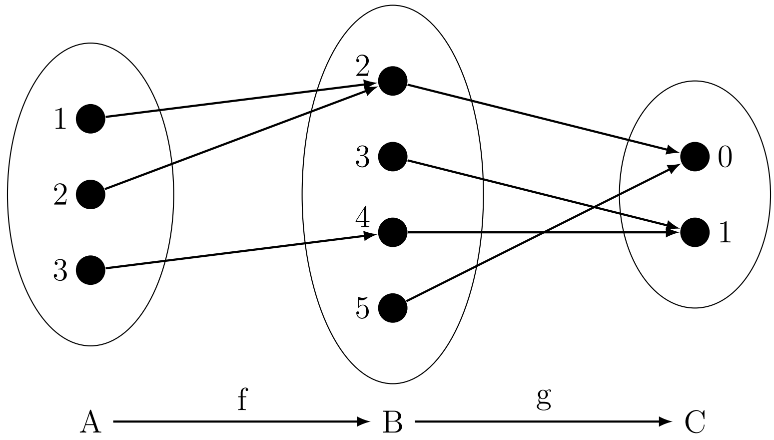
Sei nun $a \in f^{-1}[B_1] \cap f^{-1}[B_2]$. Dann ist $a \in f^{-1}[B_1]$ und $a \in f^{-1}[B_2]$. Also gilt $f(a) \in B_1$ und $f(a) \in B_2$. Damit ist $f(a) \in B_1 \cap B_2$. Es folgt $a \in f^{-1}[B_1 \cap B_2]$. Das zeigt $f^{-1}[B_1] \cap f^{-1}[B_2] \subseteq f^{-1}[B_1 \cap B_2]$.

(5) Sei $a \in A_1$. Dann ist $f(a) \in f[A_1]$. Also gilt $a \in f^{-1}[f[A_1]]$. Das zeigt $A_1 \subseteq f^{-1}[f[A_1]]$. □

Definition 1.55. *Sind $f\colon A \longrightarrow B$ und $g\colon B \longrightarrow C$ Funktionen, so definieren wir die* Komposition *von f und g als die Funktion $g \circ f\colon A \longrightarrow C; a \longmapsto g(f(a))$. Die Komposition $g \circ f$ wird „g nach f“ gelesen.*

Beispiel 1.56. Es seien $A = \{1, 2, 3\}$, $B = \{2, 3, 4, 5\}$ und $C = \{0, 1\}$. Die Funktionen $f : A \longrightarrow B$ und $g : B \longrightarrow C$ seien definiert durch $f(1) = f(2) = 2$, $f(3) = 4$, $g(2) = g(5) = 0$ und $g(3) = g(4) = 1$. Dann gilt $(g \circ f)(1) = (g \circ f)(2) = 0$ sowie $(g \circ f)(3) = 1$.

Die Komposition $g \circ f$ kann man sich leicht vorstellen, wenn man die entsprechenden Pfeildiagramme betrachtet.



Die Komposition von Abbildungen erfüllt das Assoziativgesetz.

Satz 1.57. *Seien $f : A \longrightarrow B$, $g : B \longrightarrow C$ und $h : C \longrightarrow D$ Abbildungen. Dann gilt $h \circ (g \circ f) = (h \circ g) \circ f$.*

Beweis. Wir müssen zeigen, dass für alle $a \in A$ die Gleichung

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

gilt. Sei also $a \in A$. Dann ist

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

und der Satz folgt. □

Definition 1.58. *Sei $f : A \longrightarrow B$ eine Funktion und $A' \subseteq A$. Unter der* Einschränkung *oder* Restriktion *von f auf A' versteht man die Funktion $g : A' \longrightarrow B; a \longmapsto f(a)$. Für die Einschränkung von f auf A' schreibt man $f \restriction A'$ oder $f|_{A'}$.*

Definition 1.59. *Sei $f : A \longrightarrow B$ eine injektive Funktion. Dann kann man eine Funktion $g : f[A] \longrightarrow A$ so definieren, dass für alle $b \in f[A]$ und $a \in A$ die Gleichung $g(b) = a$ genau dann gilt, wenn $f(a) = b$ ist. Die Funktion g ist die* Umkehrfunktion *von f. Für die Umkehrfunktion von f schreibt man f^{-1}.*

Bemerkung 1.60. Sei $f : A \longrightarrow B$ eine Bijektion und sei $B_1 \subseteq B$. Die Schreibweise $f^{-1}[B_1]$ erscheint zunächst mehrdeutig, da entweder das Urbild von B_1 unter f, oder das Bild von B_1 unter der Abbildung f^{-1} gemeint sein könnte. Allerdings sind diese Mengen identisch. Es gilt

$$\{a \in A : f(a) \in B_1\} = \{f^{-1}(b) : b \in B_1\}.$$

Also ist diese Mehrdeutigkeit unproblematisch.

§1.10 Die Russellsche Antinomie und die Mengenlehre als Grundlage der Mathematik

Die Mengenlehre bildet die Grundlage der Mathematik, weil sich alle mathematischen Objekte als Mengen auffassen lassen. Damit ist sämtliche Mathematik eigentlich Mengenlehre. Unser naiver Mengenbegriff führt aber leider zu Widersprüchen. Bertrand Russell entdeckte 1901 die folgende „Menge“:

$$R := \{x : x \notin x\}$$

Warum ist diese Definition problematisch? Wir fragen uns, ob $R \in R$ gilt oder nicht.

1. Fall: $R \in R$. In diesem Fall erfüllt R, eingesetzt für x, nicht die Aussage $x \notin x$. Also ist R nach der Definition von R kein Element von R. Damit gilt $R \notin R$, was unserer Annahme $R \in R$ widerspricht.

2. Fall: $R \notin R$. Jetzt hat R, eingesetzt für x, die Eigenschaft $x \notin x$. Also ist, entsprechend der Definition von R, R selbst ein Element von R. Es gilt also $R \in R$. Das widerspricht unserer Annahme $R \notin R$.

Einer der beiden Fälle muss eintreten, aber beide führen zu einem Widerspruch. Das ist die Russellsche Antinomie.

Wie kann dieses Problem behoben werden? Russell selbst hat das Problem durch die Einführung einer Typentheorie gelöst, bei der eine Mengen nur Elemente echt niedrigeren Typs haben können. Es gibt dann zum Beispiel einen Typ für Zahlen, einen höheren Typ für Mengen von Zahlen, einen noch höheren Typ für Mengen von Mengen von Zahlen und so weiter. Insbesondere ist in der Typentheorie klar, dass eine Menge nicht Element von sich selbst sein kann. Typentheorien werden heutzutage insbesondere im Bereich des automatischen Beweisens betrachtet.

In der allgemeinen Mathematik löst man die Russellsche Antinomie dadurch, dass man nicht einfach jede Menge der Form $\{x : E(x)\}$, wobei $E(x)$ eine Aussageform ist, bilden darf. Stattdessen legt man sich auf ein System von Axiomen fest, welche einem im wesentlichen sagen, wie man aus bekannten Mengen weitere Mengen konstruieren kann. Das populärste Axiomensystem ist ZFC, die Zermelo-Fraenkelsche Mengenlehre zusammen mit dem Auswahlaxiom.

Die ZFC-Axiome besagen zum Beispiel, dass die leere Menge existiert, dass für je zwei Mengen x und y auch die Menge $\{x, y\}$ existiert, oder dass für jede Menge x auch die Potenzmenge von x existiert.

Beispiel 1.61. Der Vollständigkeit halber schreiben wir die Axiome der Zermelo-Fraenkelschen Mengenlehre und das Auswahlaxiom in natürlicher Sprache auf.

- Ext (Extensionalitätsaxiom): Zwei Mengen sind genau dann gleich, wenn sie die gleichen Elemente haben.

- Paar (Paarmengenaxiom): Für je zwei Mengen x und y ist $\{x, y\}$ eine Menge.
- Ver (Vereinigungsaxiom): Für jede Menge x ist die Vereinigung aller Elemente von x wieder eine Menge.
- Pot (Potenzmengenaxiom): Für jede Menge x ist $\wp(x)$ eine Menge.
- Fund (Fundierungsaxiom): Jede nichtleere Menge x hat ein Element y, so dass kein $z \in x$ Element von y ist. Zusammen mit den anderen Axiomen folgt daraus zum Beispiel, dass keine Menge Element von sich selber ist.
- Un (Unendlichkeitsaxiom): Es gibt eine unendliche Menge. Genauer, es gibt eine nichtleere Menge, die unter der Abbildung $y \longmapsto y \cup \{y\}$ abgeschlossen ist.
- Auss (Aussonderungsaxiom): Für jede Eigenschaft $E(x)$ von Mengen und jede Menge y ist $\{x \in y : E(x)\}$ eine Menge.
- Ers (Ersetzungsaxiom): Für jede (definierbare) Abbildung F und jede Menge y ist $\{F(x) : x \in y\}$ eine Menge.
- Aus (Auswahlaxiom): Ist I eine Menge und ist A_i für jedes $i \in I$ eine nichtleere Menge, so existiert eine Funktion f mit Definitionsbereich i, so dass für jedes $i \in I$ gilt: $f(i) \in A_i$

Diese Axiome können auch in der formalen Sprache der Mengenlehre aufgeschrieben werden, die abgesehen von Variablen, Klammern, Quantoren und logischen Verknüpfungen mit dem Relationssymbol $\in$ auskommt. So lautet das Extensionalitätsaxiom formal

$$\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y).$$

Möchte man das Aussonderungsaxiom oder das Ersetzungsaxiom formal aufschreiben, so muss man beachten, dass man in der formalen Sprache der Mengenlehre nur über Mengen quantifizieren darf, nicht aber über Eigenschaften oder Definitionen von Funktionen. Deshalb muss man für jede Eigenschaft $E(x)$ von Mengen ein eigenes Aussonderungsaxiom aufschreiben und für jede Definition einer Funktion F ein eigenes Ersetzungsaxiom. Damit enthält ZFC unendlich viele Axiome, was aber in Wirklichkeit unproblematisch ist.

Die Russellsche „Menge“ R, die wir oben definiert haben, ist in ZFC keine Menge, sondern eine Zusammenfassung im Sinne von Cantor, die wir Klasse nennen. Eine Klasse ist also ein Objekt der Form $\{x : E(x)\}$, wobei $E(x)$ eine Eigenschaft von Mengen ist. Nicht alle Klassen sind Mengen, wie das Beispiel R zeigt. Aber jede Menge y lässt sich mit der Klasse $\{x : x \in y\}$ identifizieren. Die Elemente von Klassen sind Mengen. Klassen, die keine Mengen sind, sogenannte echte Klassen, treten niemals als Elemente von irgendetwas auf.

Wir werden in dieser Vorlesung nicht wieder auf die Zermelo-Fraenkelsche Mengenlehre zurückkommen. In der Tat benutzen die meisten Mathematiker ZFC nie explizit.

Überraschender Weise ist es trotzdem so, dass sich Beweise mathematischer Sätze praktisch immer unbewusst im Rahmen von ZFC abspielen.

KAPITEL 2

Die natürlichen Zahlen und vollständige Induktion

§2.1 Summen- und Produktzeichen

Bevor wir uns eingehend mit den bekannten Zahlenbereichen $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$ befassen, führen wir eine Notation ein, die sich bald als nützlich erweisen wird. Die *reellen Zahlen* $\mathbb{R}$ sind die bekannten Zahlen auf der Zahlengerade wie -1, 0, 2.5, $-\frac{10}{7}$, e und π, für die die üblichen Rechenregeln gelten.

Definition 2.1. *Für reelle Zahlen $a_1, \ldots, a_n$ sei*

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \ldots + a_n.$$

Dabei heißt i der Laufindex, *1 ist die* untere Summationsgrenze *und n die* obere Summationsgrenze.

Der Laufindex muss nicht mit i bezeichnet werden und die untere Summationsgrenze muss nicht 1 sein. So ist zum Beispiel

$$\sum_{j=0}^{4} 2^j = 2^0 + 2^1 + 2^2 + 2^3 + 2^4 = 31.$$

Summen mit wechselnden Vorzeichen, wie zum Beispiel $a_1 - a_2 + a_3 - a_4$, kann man bequem mit Hilfe von Potenzen von -1 schreiben. Dabei muss man aber genau aufpassen, welche Vorzeichen man erzeugt:

$$\sum_{i=1}^{4} (-1)^i a_i = -a_1 + a_2 - a_3 + a_4$$

$$\sum_{i=1}^{4} (-1)^{i+1} a_i = a_1 - a_2 + a_3 - a_4$$

Falls $a_1 = \cdots = a_n = a$ gilt, so ist $\sum_{i=1}^{n} a_i = na$.

Das bekannte Distributivgesetz lautet $a(b + c) = ab + ac$. Das Gesetz gilt auch für mehr als zwei Summanden. Für alle reellen Zahlen $a, b_1, \ldots, b_n$ ist

$$a \sum_{i=1}^{n} b_i = a(b_1 + \ldots + b_n) = ab_1 + \ldots + ab_n = \sum_{i=1}^{n} ab_i.$$

Mit Hilfe des Distributivgesetzes können wir Ausdrücke wie $(a + b)(c + d)$ ausmultiplizieren und erhalten

$$(a + b)(c + d) = ac + ad + bc + bd.$$

Allgemein gilt

$$(a_1 + \ldots + a_m)(b_1 + \ldots + b_n) = a_1b_1 + \ldots + a_1b_n + \ldots + a_mb_1 + \ldots + a_mb_n.$$

Mit dem Summenzeichen geschrieben erhalten wir

$$\left(\sum_{i=1}^{m} a_i\right)\left(\sum_{j=1}^{n} b_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_ib_j.$$

Da wir nach dem Kommutativgesetz für die Addition die Summanden vertauschen können ohne den Wert der Summe zu ändern, ist

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_ib_j = \sum_{j=1}^{n}\sum_{i=1}^{m} a_ib_j.$$

Auf der Änderung der Summationsreihenfolge beruht auch die Gleichung

$$\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i.$$

Oft kann man dieselben Summen unterschiedlich aufschreiben. So ist zum Beispiel

$$\sum_{i=0}^{3} a_{2i+1} = a_1 + a_3 + a_5 + a_7 = \sum_{i=1}^{4} a_{2i-1}.$$

Bemerkung 2.2. Analog zum Summenzeichen kann man auch das Produktzeichen definieren. Sind $a_1, \ldots, a_n$ reelle Zahlen, so setzt man

$$\prod_{i=1}^{n} a_i := a_1 \cdot a_2 \cdot \ \ldots \ \cdot a_n.$$

§2.2 Die natürlichen Zahlen

Auf den natürlichen Zahlen $\mathbb{N} = \{1, 2, 3, \ldots\}$ gelten die bekannten Rechengesetze:

(1) Assoziativgesetze:
- $a + (b + c) = (a + b) + c$
- $a \cdot (b \cdot c) = (a \cdot b) \cdot c$

(2) Kommutativgesetze:
- $a + b = b + a$
- $a \cdot b = b \cdot a$

(3) Distributivgesetz:
- $a \cdot (b + c) = a \cdot b + a \cdot c$

(4) Existenz eines neutralen Elements der Multiplikation:
- $a \cdot 1 = a$

Eine weitere wichtige Eigenschaft von $\mathbb{N}$ ist das Funktionieren der *vollständigen Induktion*.

§2.3 Das Prinzip der vollständigen Induktion

Sei $A(n)$ eine Aussageform. Dann gilt $\forall n \in \mathbb{N}\colon A(n)$ genau dann, wenn folgende zwei Bedingungen erfüllt sind:

(1) *Induktionsanfang*: $A(1)$ ist wahr.
(2) *Induktionsschritt*: Für jedes $n \in \mathbb{N}$ gilt: Falls $A(n)$ wahr ist, so ist auch $A(n+1)$ wahr.

Kompakt geschrieben gilt also für jede Aussageform $A(n)$:

$$(A(1) \wedge \forall n \in \mathbb{N}(A(n) \Rightarrow A(n+1))) \Rightarrow \forall n \in \mathbb{N}\colon A(n)$$

Als Beispiel beweisen wir einen Satz über die Summe der ersten n natürlichen Zahlen.

Satz 2.3. *Für alle $n \in \mathbb{N}$ gilt:*

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

Beweis. Sei $A(n)$ die Aussageform $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Wir wollen zeigen, dass $A(n)$ für alle $n \in \mathbb{N}$ gilt.

Induktionsanfang. A(1) ist wahr.

A(1) ist nämlich die Aussage $\sum_{i=1}^{1} i = \frac{1\cdot(1+1)}{2}$. Es gilt $\sum_{i=1}^{1} i = 1 = \frac{1\cdot(1+1)}{2}$. Das zeigt $A(1)$.

Induktionsschritt. Für alle $n \in \mathbb{N}$ gilt: $A(n) \Rightarrow A(n+1)$

Um das zu zeigen, nehmen wir uns ein beliebiges $n \in \mathbb{N}$ her und zeigen $A(n) \Rightarrow A(n+1)$. Wir müssen also zeigen, dass $A(n+1)$ wahr ist, falls $A(n)$ wahr ist. Wenn $A(n)$ falsch ist, ist nichts zu zeigen.

Wir können also annehmen, dass $A(n)$ wahr ist. Das ist die *Induktionsannahme*. Nun zeigen wir $A(n+1)$ unter dieser Annahme. $A(n+1)$ ist die Aussage

$$\sum_{i=1}^{n+1} i = \frac{(n+1)((n+1)+1)}{2},$$

also

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

Es gilt

$$\sum_{i=1}^{n+1} i = \sum_{i=1}^{n} i + (n+1).$$

Nach der Induktionsannahme ist $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Mit dieser Information erhalten wir

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

Das zeigt $A(n+1)$.

Damit haben wir den Induktionsanfang und den Induktionsschritt bewiesen. Es folgt, dass $A(n)$ für alle $n \in \mathbb{N}$ gilt. $\square$

Wir geben ein weiteres Beispiel. Für ganze Zahlen a und b schreiben wir $a \mid b$, falls a ein Teiler von b ist.

Satz 2.4. *Für alle $n \in \mathbb{N}$ ist $n^3 - n$ durch 3 teilbar.*

Beweis. Sei $A(n)$ die Aussageform „3 teilt $n^3 - n$“. Wir wollen zeigen, dass $A(n)$ für alle $n \in \mathbb{N}$ gilt.

Induktionsanfang. A(1) ist wahr.

A(1) ist nämlich die Aussage $3 \mid 1^3 - 1$, also $3 \mid 0$. Diese Aussage ist wahr.

Induktionsschritt. Für alle $n \in \mathbb{N}$ gilt: $A(n) \Rightarrow A(n+1)$

Sei also $n \in \mathbb{N}$. Wieder nehmen wir an, dass $A(n)$ wahr ist, und zeigen $A(n+1)$. Die Induktionsannahme ist also $3 \mid n^3 - n$.

$A(n+1)$ ist die Aussage $3 \mid (n+1)^3 - (n+1)$. Wir vereinfachen:

$$(n+1)^3 - (n+1) = n^3 + 3n^2 + 3n + 1 - n - 1 = n^3 + 3n^2 + 2n$$

Wir wollen zeigen, dass $n^3 + 3n^2 + 2n$ durch 3 teilbar ist, und dürfen die Induktionsannahme, dass $n^3 - n$ durch 3 teilbar ist, dafür verwenden. Es gilt

$$n^3 + 3n^2 + 2n = (n^3 - n) + 3n^2 + 3n.$$

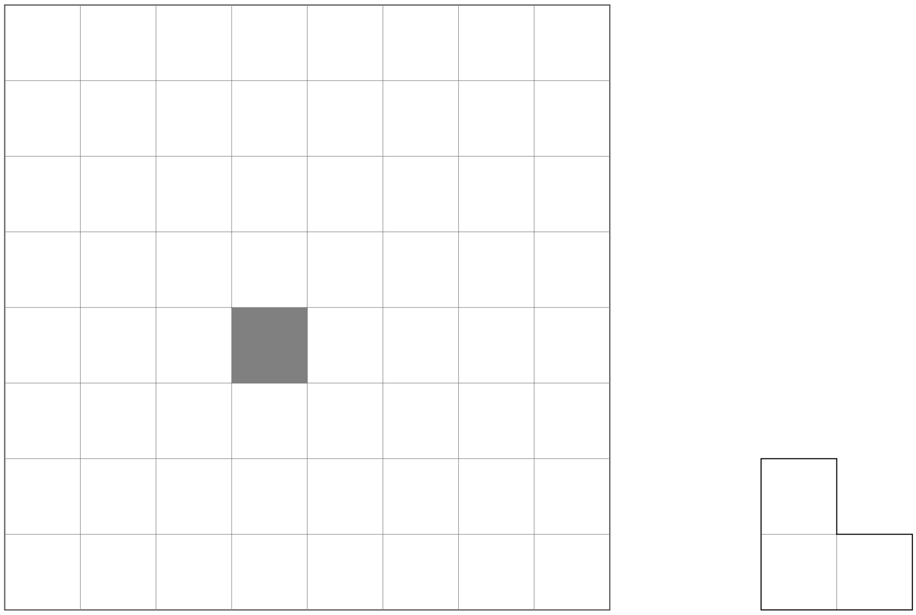
Der erste Summand der rechten Seite dieser Gleichung, $n^3 - n$, ist nach Induktionsannahme durch 3 teilbar. Der Rest, $3n^2 + 3n$, ist offenbar auch durch 3 teilbar. Das zeigt $3 \mid (n+1)^3 - (n+1)$ und damit $A(n+1)$.

Damit ist für alle $n \in \mathbb{N}$ die Implikation $A(n) \Rightarrow A(n+1)$ bewiesen. Zusammen mit dem Induktionsanfang folgt $3 \mid n^3 - n$ für alle $n \in \mathbb{N}$. □

Als nächstes diskutieren wir ein Beispiel, das zeigt, dass der Erfolg einer Induktion von der geschickten Wahl des Induktionsanfangs abhängen kann. Außerdem liefert der folgende Beweis einen Algorithmus, also ein Verfahren, zur Lösung des vorgelegten Problems.

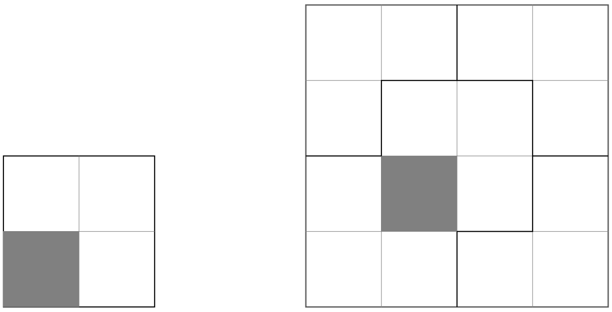
Problem 2.5. Ein quadratischer Hof mit der Seitenlänge 2^n soll mit L-förmigen Fliesen gefliest werden. Dabei soll ein Quadrat mit der Seitenlänge 1 in der Mitte des Hofes frei bleiben, weil da eine Statue aufgestellt werden soll. Die Fliesen haben die Form von drei aneinander gesetzten Quadraten mit Seitenlänge eins, so wie in der Skizze. Ist es möglich, den Hof bis auf das Quadrat in der Mitte vollständig mit den Fliesen zu überdecken, ohne dass die Fliesen sich überlappen und ohne Fliesen zu zerschneiden?

Im Folgenden betrachten wir nur Quadrate, deren Seitenlängen ganzzahlig sind. Auch stellen wir uns immer vor, dass die Quadrate in der Ebene liegen, wobei die Koordinaten der Ecken der Quadrate alle ganzzahlig sind.



Hof Fliese

Wir betrachten zunächst die Fälle $n = 1$ und $n = 2$ und sehen, dass wir den Hof wie gewünscht fliesen können. Schon der Fall $n = 1$ genügt für den Induktionsanfang.



$n = 1$ $n = 2$

Eine naheliegende Induktionsannahme wäre die Aussageform $A(n)$: „Jeder quadratische Hof mit der Kantenlänge 2^n kann bis auf ein fehlendes Quadrat der Kantenlänge 1 in der Mitte vollständig mit L-förmigen Fliesen gefliest werden.“

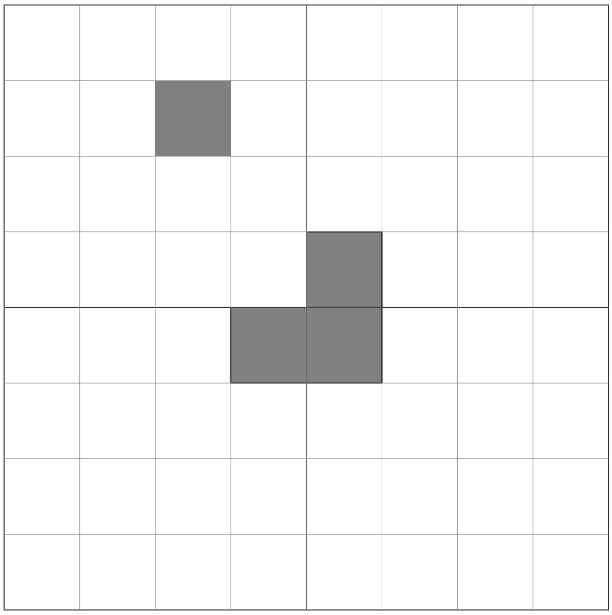
Es stellt sich heraus, dass wir Schwierigkeiten haben, die gewünschte Induktion mit dieser Induktionsannahme durchzuführen. Einen Hof der Kantenlänge 2^{n+1} können wir in vier quadratische Teile mit der Kantenlänge 2^n zerlegen, aber das fehlende Quadrat in der Mitte des Quadrats mit Kantenlänge 2^{n+1} liegt nun am Rand eines der Quadrate mit Kantenlänge 2^n. Bei den anderen drei Quadraten mit Kantenlänge fehlt kein Quadrat.

Eine Verstärkung von $A(n)$ führt schließlich zum Erfolg. $B(n)$ sei die Aussageform: „Jeder quadratische Hof mit der Kantenlänge 2^n kann bis auf ein beliebig vorgegebenes fehlendes Quadrat der Kantenlänge 1 vollständig mit L-förmigen Fliesen gefliest werden“.

Wir zeigen, dass $B(n)$ für alle $n \in \mathbb{N}$ gilt. Der Induktionsanfang ist einfach: $B(1)$ gilt, da von einem Quadrat der Kantenlänge 2 nach Entfernen eines Quadrates der Kantenlänge 1 eine L-förmige Fliese übrig bleibt.

Induktionsschritt: Wir zeigen, dass für alle $n \in \mathbb{N}$ die Implikation $B(n) \Rightarrow B(n+1)$ gilt. Sei also $n \in \mathbb{N}$. Wir nehmen an, dass $B(n)$ gilt. Sei nun ein Quadrat mit Kantenlänge 2^{n+1} vorgegeben, in dem ein Quadrat der Kantenlänge 1 markiert ist, welches beim Überdecken ausgelassen werden soll.

Wir zerlegen dieses Quadrat in vier Quadrate der Kantenlänge 2^n. Das markierte Quadrat der Kantenlänge 1 liegt in einem dieser vier Quadrate. Nun legen wir eine der L-förmigen Fliesen so in die Mitte des Quadrats mit Kantenlänge 2^{n+1}, dass die drei Quadrate der Fliese alle in je einem der vier Quadrate der Kantenlänge 2^n zum liegen kommen, wobei dasjenige der vier Quadrate, das das markierte Quadrat enthält, nicht getroffen wird.



Zerlegung des Quadrats der Kantenlänge 2^{n+1} und Lage der ersten Fliese

Nun genügt es, jedes der vier Quadrate mit Kantenlänge 2^n mit L-förmigen Fliesen zu überdecken, wobei jeweils ein Quadrat der Kantenlänge 1 ausgelassen werden muss. Das ist aber nach der Induktionsannahme $B(n)$ möglich. Das zeigt die Implikation $B(n) \Rightarrow B(n+1)$. Also gilt $B(n)$ für alle $n \in \mathbb{N}$. Das löst Problem 2.5.

Wir bemerken noch, dass diese Lösung des Problems auch ein Verfahren liefert, den Hof wie gewünscht zu fliesen:

- Wenn der Hof die Kantenlänge 2 hat, so bleibt neben dem markierten Quadrat genau Platz für eine L-förmige Fliese.
- Wenn der Hof für ein $n > 1$ die Kantenlänge 2^n hat, so unterteile den Hof in vier Quadrate der Kantenlänge 2^{n-1} und lege eine Fliese so in die Mitte des Hofes, dass sie genau die drei Quadrate der Kantenlänge 2^{n-1} trifft, die nicht das markierte Quadrat enthalten.

- Führe den Algorithmus für die vier Quadrate der Kantenlänge 2^{n-1} durch, wobei das ursprünglich markierte Quadrat und die drei Quadrate, die von der ersten Fliese überdeckt werden, markiert werden.

Wir betrachten zwei weitere Varianten der vollständigen Induktion. So muss man zum Beispiel den Induktionsanfang nicht unbedingt bei $n = 1$ machen. Ein Induktionsanfang bei $n = 0$ kommt recht häufig vor, andere Startwerte sind aber auch möglich.

2.3.1. Vollständige Induktion mit beliebigem Startwert. Es sei n_0 eine ganze Zahl und $A(n)$ eine Aussageform. Dann gilt $A(n)$ genau dann für alle ganzen Zahlen $n \geqslant n_0$, wenn $A(n_0)$ wahr ist und die Implikation $A(n) \Rightarrow A(n+1)$ für alle $n \geqslant n_0$ gilt.

Als Beispiel beweisen wir eine einfache Ungleichung.

Satz 2.6. *Für alle natürlichen Zahlen $n \geqslant 3$ gilt $2n + 1 < 2^n$.*

Beweis. $A(n)$ sei die Aussageform $2n + 1 < 2^n$.

Induktionsanfang. $A(3)$ gilt.

Um das zu sehen, setzen wir 3 für n ein. Es ist $2 \cdot 3 + 1 = 7 < 8 = 2^3$.

Induktionsschritt. Für alle $n \geqslant 3$ gilt: $A(n) \to A(n+1)$

Wir nehmen an, dass $A(n)$ für ein gewisses $n \geqslant 3$ gilt, und haben $A(n+1)$ nachzuweisen. Es ist

$$2(n+1) + 1 = 2n + 3 = 2n + 1 + 2 \overset{\text{I.A.}}{<} 2^n + 2 \overset{n\geqslant 2}{<} 2^n + 2^n = 2^{n+1}.$$

Das zeigt $A(n+1)$.

Es folgt, dass $A(n)$ für alle $n \geqslant 3$ gilt. □

Wir beweisen noch eine Formel, die sich in der Analysis als nützlich erweisen wird. Sei q eine reelle Zahl $\neq 1$ und $n \in \mathbb{N}_0$. Wir wollen einen einfachen Ausdruck für die Summe $\sum_{i=0}^{n} q^i = 1 + q + \ldots + q^n$ herleiten. Dazu formen wir die Summe um:

$$\begin{aligned}\sum_{i=0}^{n} q^i &= 1 + \sum_{i=1}^{n} q^i = 1 + q\sum_{i=1}^{n} q^{i-1} = 1 + q\sum_{i=0}^{n-1} q^i = 1 + q\sum_{i=0}^{n-1} q^i + q^{n+1} - q^{n+1}\\ &= 1 + q\left(\sum_{i=0}^{n-1} q^i + q^n\right) - q^{n+1} = 1 + q\sum_{i=0}^{n} q^i - q^{n+1}\end{aligned}$$

Wenn man den Term $q\sum_{i=0}^{n} q^i$ auf die linke Seite dieser Gleichung bringt, erhält man

$$(1-q)\sum_{i=0}^{n} q^i = 1 - q^{n+1}.$$

Da $q \neq 1$ ist, können wir auf beiden Seiten durch $1 - q$ teilen und erhalten so die *geometrische Summenformel*:

Satz 2.7 (Geometrische Summenformel)**.** *Sei q eine reelle Zahl $\neq 1$ und $n \in \mathbb{N}_0$. Dann gilt*

$$\sum_{i=0}^{n} q^i = \frac{1-q^{n+1}}{1-q}.$$

Beweis. Wir haben die geometrische Summenformel zwar korrekt hergeleitet, geben aber trotzdem noch einen Beweis mittels vollständiger Induktion an.
Induktionsanfang. Für $n = 0$ stimmt die geometrische Summenformel, denn es gilt

$$\sum_{i=0}^{0} q^i = 1 = \frac{1-q^1}{1-q}.$$

Induktionsschritt. Die Induktionsannahme besagt, dass die geometrische Summenformeln für ein gewisses $n \geqslant 0$ gilt und wir zeigen, dass die Formel auch für $n+1$ gilt:

$$\sum_{i=0}^{n+1} q^i = \sum_{i=0}^{n} q^i + q^{n+1} \overset{\text{I.A.}}{=} \frac{1-q^{n+1}}{1-q} + q^{n+1} = \frac{1-q^{n+1}}{1-q} + \frac{q^{n+1}(1-q)}{1-q}$$
$$= \frac{1-q^{n+1}+q^{n+1}-q^{n+2}}{1-q} = \frac{1-q^{n+2}}{1-q}$$

Damit ist die geometrische Summenformel für alle $n \in \mathbb{N}_0$ bewiesen. □

2.3.2. Vollständige Induktion mit mehreren Vorgängern. Wieder sei $A(n)$ eine Aussageform. Dann gilt $A(n)$ genau dann für alle natürlichen Zahlen n, wenn $A(1)$ wahr ist und für alle $n \in \mathbb{N}$ die folgende Implikation gilt: $A(1) \wedge \cdots \wedge A(n) \Rightarrow A(n+1)$.

Bei dieser Variante ist die Induktionsannahme die Annahme, dass $A(1), \ldots, A(n)$ wahr sind.

Eng mit der vollständigen Induktion verwandt sind *rekursive Definitionen.*

Beispiel 2.8. Wir definieren einen Folge natürlicher Zahlen a_n:

(1) $a_1 = 1$ und
(2) $a_{n+1} = 2a_n + 1$.

Dadurch ist a_n für jede natürliche Zahl n eindeutig bestimmt. Nach (1) gilt $a_1 = 1$. Wenden wir (2) auf den Fall $n = 1$ an, so erhalten wir $a_2 = 2 \cdot 1 + 1 = 3$. Wenden wir (2) auf den Fall $n = 2$ an, so ergibt sich $a_3 = 2 \cdot 3 + 1 = 7$.

Ein weiteres Beispiel für eine rekursive Definition sind die bekannten Fibonacci-Zahlen.

Definition 2.9. *Es sei $f_0 = 0$ und $f_1 = 1$. Für alle $n \geqslant 1$ sei $f_{n+1} = f_{n-1} + f_n$.*

Die Zahlen $f_0, f_1, f_2, \ldots$ heißen Fibonacci-Zahlen. *Die ersten 10 Glieder der Folge $f_0, f_1, f_2, \ldots$ lauten* $0, 1, 1, 2, 3, 5, 8, 13, 21, 34$.

Man kann für die n-te Fibonacci-Zahl f_n eine geschlossene Formel angeben, also einen Ausdruck, der keine Rekursion benutzt.

Satz 2.10. *Für alle $n \in \mathbb{N}_0$ gilt*

$$f_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

Beweis. Wir beweisen den Satz durch vollständige Induktion, wobei wir Induktion mit mehreren Vorgängern anwenden. Das liegt daran, dass in der rekursiven Definition von f_{n+1} auch auf mehrere Vorgänger zurückgegriffen wird.

Um die Rechnung übersichtlicher zu gestalten, führen wir zwei Abkürzungen ein. Es seien $\varphi := \frac{1+\sqrt{5}}{2}$ und $\psi := \frac{1-\sqrt{5}}{2}$. Sei $A(n)$ die Aussageform

$$f_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}.$$

Wir wollen also zeigen, dass $A(n)$ für alle $n \in \mathbb{N}_0$ gilt.

Als Induktionsannahme wählen wir $A(n-1) \wedge A(n)$. Das können wir natürlich nur annehmen, falls n mindestens 1 ist, da f_{-1} ja nicht definiert ist und wir nicht wissen, was $A(-1)$ bedeutet. Im Induktionsschritt zeigen wir dann für alle $n \geqslant 1$, dass aus $A(n-1)$ und $A(n)$ zusammen $A(n+1)$ folgt.

Wenn wir für den Induktionsanfang nur $A(0)$ zeigen, dann wissen wir nicht, ob $A(1)$ überhaupt gilt, da im Induktionsschritt $A(n-1) \wedge A(n) \Rightarrow A(n+1)$ nur für $n \geqslant 1$ vorausgesetzt wird. Daher müssen wir beim Induktionsanfang auch noch $A(1)$ explizit zeigen.

Induktionsanfang. Es gilt

$$\frac{\varphi^0 - \psi^0}{\sqrt{5}} = \frac{1-1}{\sqrt{5}} = 0 = f_0$$

sowie

$$\frac{\varphi^1 - \psi^1}{\sqrt{5}} = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2}\right) = \frac{1}{\sqrt{5}} \cdot \frac{2\sqrt{5}}{2} = 1 = f_1.$$

Induktionsschritt. Wir zeigen $A(n-1) \wedge A(n) \Rightarrow A(n+1)$ für alle $n \geqslant 1$. Dazu nehmen wir an, dass für ein gewisses $n \geqslant 1$ die Aussage $A(n-1) \wedge A(n)$ gilt. Dann ist

$$f_{n+1} = f_{n-1} + f_n = \frac{\varphi^{n-1} - \psi^{n-1} + \varphi^n - \psi^n}{\sqrt{5}} = \frac{\varphi^n\left(1 + \frac{1}{\varphi}\right) - \psi^n\left(1 + \frac{1}{\psi}\right)}{\sqrt{5}}.$$

Es gilt

$$\begin{aligned} 1 + \frac{1}{\varphi} = 1 + \frac{2}{1+\sqrt{5}} &= \frac{1+\sqrt{5}+2}{1+\sqrt{5}} \\ &= \frac{(3+\sqrt{5})(1-\sqrt{5})}{(1+\sqrt{5})(1-\sqrt{5})} = \frac{-2-2\sqrt{5}}{1-5} = \frac{1+\sqrt{5}}{2} = \varphi \end{aligned}$$

und analog $1 + \frac{1}{\psi} = \psi$. Damit ergibt sich

$$f_{n+1} = \frac{\varphi^{n+1} - \psi^{n+1}}{\sqrt{5}},$$

also $A(n+1)$.

Insgesamt gilt $A(n)$ für alle $n \in \mathbb{N}_0$. □

§2.4 Die Peano-Axiome

Wir haben bisher noch nicht diskutiert, warum die vollständige Induktion überhaupt funktioniert. Unsere intuitive Vorstellung von den natürlichen Zahlen ist wie folgt: Wenn wir bei 1 anfangen zu zählen und dann in Einerschritten immer weiter zählen, so erreichen wir schließlich jede natürliche Zahl. Oder anders gesagt, die natürlichen Zahlen sind genau die Zahlen, die wir erreichen können, wenn wir bei 1 zu zählen anfangen und dann in Einerschritten immer weiter zählen.

Ist $A(n)$ eine Aussageform und gelten $A(1)$ und $\forall n \in \mathbb{N}(A(n) \Rightarrow A(n+1))$, so können wir die Menge $S = \{n \in \mathbb{N} \colon A(n) \text{ ist wahr}\}$ betrachten und stellen Folgendes fest:

(1) $1 \in S$
(2) $n \in S \Rightarrow n+1 \in S$

Eine Menge mit den Eigenschaften (1) und (2) nennen wir *induktiv*. Wir können also anfangen bei 1, in Einerschritten zu zählen, ohne jemals die Menge S zu verlassen. Unserer Vorstellung von den natürlichen Zahlen folgend, erreichen wir dadurch alle natürlichen Zahlen. Also gilt $\mathbb{N} \subseteq S$. Andererseits ist $S \subseteq \mathbb{N}$. Es folgt $S = \mathbb{N}$. Also gilt $A(n)$ für alle $n \in \mathbb{N}$.

Die folgende Axiome präzisieren unsere Vorstellung von den natürlichen Zahlen. Hierbei steht n' für den Nachfolger von n in den natürlichen Zahlen, also für $n+1$.

Definition 2.11. *Die folgenden Axiome sind die* Peano-Axiome *für die natürlichen Zahlen.*

(*P1*) $1 \in \mathbb{N}$
(*P2*) $n \in \mathbb{N} \Rightarrow n' \in \mathbb{N}$
(*P3*) $n \in \mathbb{N} \Rightarrow n' \neq 1$
(*P4*) $m, n \in \mathbb{N} \Rightarrow (m' = n' \Rightarrow m = n)$
(*P5*) $(1 \in S \wedge \forall n \in \mathbb{N}(n \in S \Rightarrow n' \in S)) \Rightarrow \mathbb{N} \subseteq S$

Das Axiom (5) ist das *Induktionsaxiom*, welches garantiert, dass wir Sätze mittels vollständiger Induktion beweisen können. Normalsprachlich lauten die Axiome wie folgt:

(*P1*) 1 ist eine natürliche Zahl.
(*P2*) Der Nachfolger einer natürlichen Zahl ist wieder eine natürliche Zahl.
(*P3*) 1 ist nicht Nachfolger einer natürlichen Zahl.
(*P4*) Die Nachfolgerfunktion $n \longmapsto n'$ ist injektiv.
(*P5*) Jede induktive Menge enthält alle natürlichen Zahlen.

Auf Basis dieser Axiome kann man nun die bekannten Operationen $+$ und $\cdot$, sowie die Relation $\leqslant$ auf $\mathbb{N}$ rekursiv definieren, was wir aber als Übung für den interessierten

Leser lassen. Vollständige Induktion liefert uns interessante Informationen über die Menge der natürlichen Zahlen.

Satz 2.12. *Jede nichtleere Menge natürlicher Zahlen hat ein kleinstes Element.*

BEWEIS. Sei A eine nichtleere Menge natürlicher Zahlen, also $A \subseteq \mathbb{N}$ und $A \neq \varnothing$. Falls A kein kleinstes Element hat, so betrachte $B = \mathbb{N} \setminus A$. Wir zeigen mittels vollständiger Induktion, dass B alle natürlichen Zahlen enthält und A damit leer ist, im Widerspruch zur Annahme.

Sei $P(n)$ die Aussageform $n \in B$. 1 ist das kleinste Element von $\mathbb{N}$. Also gilt $1 \notin A$, da sonst 1 das kleinste Element von A wäre. Damit ist $1 \in B$. Das zeigt $P(1)$. Das ist der Induktionsanfang.

Nun nehmen wir an, dass die Zahlen $1, \ldots, n$ Elemente von B sind, dass also $P(1), \ldots, P(n)$ gelten. Die Zahl n' kann nicht in A liegen, da n' dann das kleinste Element von A wäre. Also liegt n' in B. Das zeigt $P(n')$. Das ist der Induktionsschritt.

Damit gilt $\mathbb{N} \subseteq B$. Also ist $A = \varnothing$, im Widerspruch zu $A \neq \varnothing$. Damit hat A ein kleinstes Element. □

Wir haben hier die Induktion mit mehreren Vorgängern durchgeführt. Um zu sehen, dass das wirklich dasselbe ist, wie die Standardform der Induktion, kann man zum Beispiel anstelle der Aussageform $P(n)$ die folgende Aussageform $Q(n)$ betrachten: $\forall k \in \mathbb{N}(k \leqslant n \Rightarrow k \in B)$

Dann kann man an Stelle der Induktionsannahme $P(1) \wedge \cdots \wedge P(n)$ einfach $Q(n)$ schreiben. Man beweist dann im Induktionsschritt nicht $(P(1) \wedge \cdots \wedge P(n)) \Rightarrow P(n')$, sondern $Q(n) \Rightarrow Q(n')$. Der Beweis selbst bleibt aber eigentlich derselbe.

Wir haben dann gezeigt, dass $Q(n)$ für alle $n \in \mathbb{N}$ gilt, und zwar mit der Standardform der Induktion. Aber $(\forall n \in \mathbb{N})Q(n)$ ist natürlich äquivalent zu $(\forall n \in \mathbb{N})P(n)$.

KAPITEL 3

Elementare Zahlentheorie

§3.1 Ganze und rationale Zahlen

Da wir jede ganzen Zahl m mit dem Bruch $\frac{m}{1}$ identifizieren können, fassen wir $\mathbb{Z}$ als eine Teilmenge von $\mathbb{Q}$ auf. Wir erinnern uns kurz daran, wie man Brüche addiert und multipliziert:

$$\frac{m}{n} + \frac{m'}{n'} = \frac{m \cdot n' + m'n}{n \cdot n'}$$
$$\frac{m}{n} \cdot \frac{m'}{n'} = \frac{m \cdot m'}{n \cdot n'}$$

Die folgenden Rechenregeln für rationale Zahlen a, b, c setzen wir als bekannt voraus:

(K1) Assoziativgesetze
- $a + (b + c) = (a + b) + c$
- $a \cdot (b \cdot c) = (a \cdot b) \cdot c$

(K2) Kommutativgesetze
- $a + b = b + a$
- $a \cdot b = b \cdot a$

(K3) Distributivgesetz
- $a \cdot (b + c) = a \cdot b + a \cdot c$

(K4) Existenz neutraler Elemente bezüglich der Addition und der Multiplikation
- $a + 0 = a$
- $1 \cdot a = a$

(K5) Existenz inverser Elemente bezüglich der Addition und der Multiplikation
- Es gibt ein Element $-a$ mit $a + (-a) = 0$.
- Falls $a \neq 0$ ist, so gibt es ein Element a^{-1} mit $a \cdot a^{-1} = 1$.

Da diese Rechengesetze so wichtig sind, bekommen Strukturen, in denen diese Gesetze erfüllt sind, einen eigenen Namen.

Definition 3.1. *Sei K eine Menge, 0 und 1 zwei verschiedene Elemente von K und $+\colon K \times K \longrightarrow K$ und $\cdot\colon K \times K \longrightarrow K$ Abbildungen. Dann heißt K zusammen mit 0, 1, $+$ und $\cdot$ ein* Körper*, falls die Axiome (K1)–(K5) erfüllt sind.*

Wie oben schon bemerkt, erfüllt $\mathbb{Q}$ mit der üblichen Addition und Multiplikation und mit den bekannten Konstanten 0 und 1 die Körperaxiome (K1)–(K5). Die ganzen Zahlen $\mathbb{Z}$ mit den üblichen Rechenoperationen erfüllen zwar (K1)–(K4), aber sie bilden keinen Körper, da zum Beispiel 2 in $\mathbb{Z}$ kein multiplikatives Inverses besitzt: Es gibt keine ganze Zahl n mit $2 \cdot n = 1$.

Neben der Struktur eines Körpers haben die rationalen Zahlen noch eine weitere wichtige Eigenschaft. Sie werden durch die Kleiner-Beziehung $<$ *angeordnet*. Für je zwei verschiedene rationale Zahlen a und b gilt entweder $a < b$ („a kleiner b") oder $a > b$ („a größer b"). Es gelten folgende Regeln:

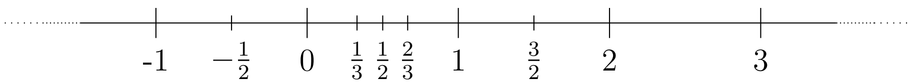
(1) $a < b \wedge b < c \Rightarrow a < c$
(2) $a < b \Rightarrow a + c < b + c$
(3) $a < b \Rightarrow a \cdot c < b \cdot c$, falls $c > 0$.
(4) $a < b \Rightarrow a \cdot c > b \cdot c$, falls $c < 0$.

Wir schreiben $a \leqslant b$ für $(a < b \vee a = b)$ und lesen $\leqslant$ als „kleiner-gleich" und $\geqslant$ als „größer-gleich".

Für $\leqslant$ gelten ähnliche Regeln wie für $<$:

(1) $a \leqslant b \wedge b \leqslant c \Rightarrow a \leqslant c$,
(2) $a \leqslant b \Rightarrow a + c \leqslant b + c$,
(3) $a \leqslant b \Rightarrow a \cdot c \leqslant b \cdot c$, falls $c \geqslant 0$,
(4) $a \leqslant b \Rightarrow a \cdot c \geqslant b \cdot c$, falls $c \leqslant 0$.

Die ganzen und die rationalen Zahlen lassen sich gut auf dem Zahlenstrahl veranschaulichen. Wir stellen uns vor, dass die Gerade horizontal von links nach rechts verläuft. Nun markieren wir einen Punkt auf der Geraden und nennen ihn 0. Rechts von der 0 markieren wir einen weiteren Punkt und nennen ihn 1. Ist nun n eine natürlich Zahl, so entspricht n dem Punkt auf der Geraden, den man erreicht, wenn man von der 0 ausgehend n-mal die Strecke von der 0 zur 1 abträgt. Sind m und n natürliche Zahlen, so erhält man den Punkt auf der Geraden, der $\frac{m}{n}$ entspricht, in dem man die Strecke von 0 nach m in n gleiche Teile unterteilt. Damit finden wir alle rationalen Zahlen > 0 auf der Zahlengeraden. Für natürliche Zahlen m und n finden wir den Punkt auf der Geraden, der $-\frac{m}{n}$ entspricht, indem man von 0 ausgehend nach links die Länge der Strecke von 0 bis $\frac{m}{n}$ abträgt.



Offenbar kann man zum Beispiel $\frac{3}{2}$ auch erreichen, indem man zuerst die Strecke von 0 nach 1 halbiert, um $\frac{1}{2}$ zu erhalten, und dann dreimal von 0 ausgehend nach rechts die Länge der Strecke von 0 bis $\frac{1}{2}$ abträgt.

Die rationalen Zahlen liegen *dicht* auf der Zahlengeraden. D.h., zwischen je zwei verschiedenen Punkten auf der Geraden liegt eine rationale Zahl. Wir werden jedoch gleich sehen, dass es Punkte auf der Geraden gibt, die keiner rationalen Zahl entsprechen, dass die rationalen Zahlen also Lücken haben.

§3.2 Die reellen Zahlen

Mit $\sqrt{2}$ bezeichnen wir die positive Lösung der Gleichung $x^2 = 2$. Es stellt sich heraus, dass $\sqrt{2}$ keine rationale Zahl ist.

Bevor wir das beweisen können, stellen wir Folgendes fest.

Lemma 3.2. *Sei m eine ganze Zahl. Falls m^2 gerade ist, so ist auch m selbst gerade.*

Beweis. Wir beweisen die Kontraposition dieser Aussage: Wenn m ungerade ist, so ist auch m^2 ungerade.

Sei m ungerade. Dann ist $m-1$ gerade. Also gibt es eine ganze Zahl k mit $2k = m-1$. Es gilt also $m = 2k+1$. Nun ist $m^2 = (2k+1)^2 = 4k^2+4k+1$. Da $4k^2+4k$ gerade ist, ist $4k^2+4k+1$ ungerade. Also ist m^2 ungerade. □

Satz 3.3. *Es gibt keine rationale Zahl a mit $a^2 = 2$.*

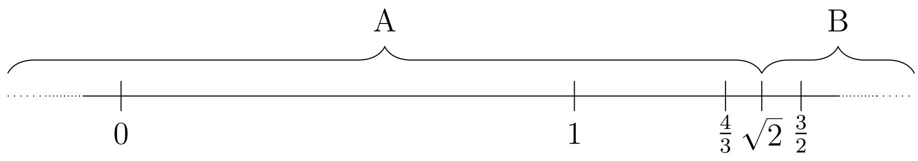
Beweis. Der Beweis dieses Satzes ist ein sogenannter *Widerspruchsbeweis*. Wir nehmen dazu an, dass es eine rationale Zahl a mit $a^2 = 2$ gibt und folgern daraus eine offensichtlich falsche Aussage. Sei A die Aussage „es gibt eine rationale Zahl a mit $a^2 = 2$“ und B eine falsche Aussage. Wenn wir $A \Rightarrow B$ zeigen können und B falsch ist, so muss A falsch sein, was wir leicht der Wahrheitstafel für $\to$ entnehmen können. Wir haben also $\neg A$ bewiesen.

Zum eigentlichen Beweis. Wie eben schon angekündigt, nehmen wir an, dass es eine rationale Zahl a mit $a^2 = 2$ gibt. Die Zahl a lässt sich als Bruch $\frac{m}{n}$ schreiben, wobei m und n ganze Zahlen sind und $n \neq 0$ gilt. Gilt $a^2 = 2$, so gilt auch $(-a)^2 = 2$. Daher können wir annehmen, dass a positiv ist und dass m und n natürliche Zahlen sind. Schließlich können wir noch annehmen, dass der Bruch $\frac{m}{n}$ gekürzt ist, dass also m und n keine gemeinsame Teiler > 1 haben. Es gilt $a^2 = \frac{m^2}{n^2} = 2$. Multiplikation mit n^2 liefert $m^2 = 2n^2$. Also ist m^2 durch 2 teilbar. Nach Lemma 3.2 ist damit auch m durch 2 teilbar.

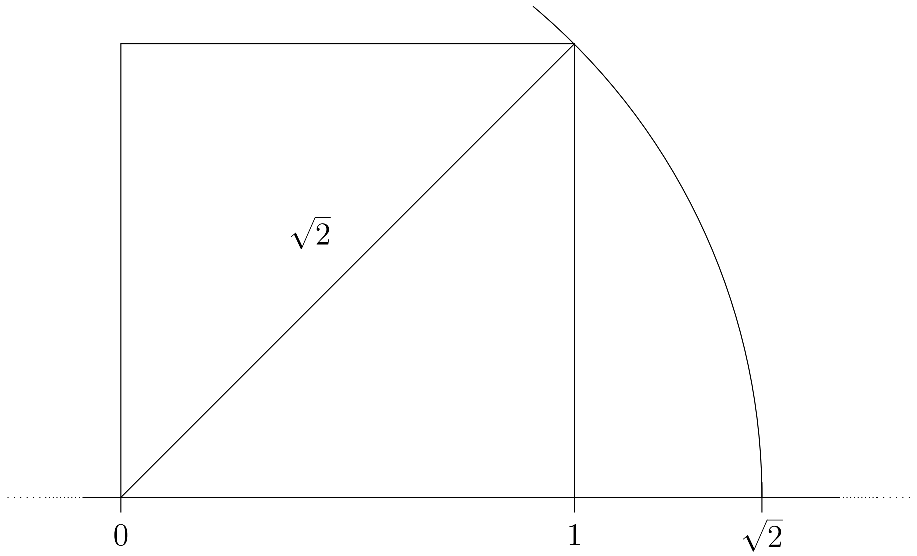
Wenn aber m von 2 geteilt wird, so wird m^2 von 4 geteilt. Wegen $m^2 = 2n^2$ wird dann aber auch n^2 von 2 geteilt. Wie oben für m ergibt sich, dass n gerade ist. Das heißt aber, dass man den Bruch $\frac{m}{n}$ durch 2 kürzen kann, ein Widerspruch zur Annahme, dass der Bruch bereits gekürzt ist.

Die Aussage „der Bruch $\frac{m}{n}$ ist gekürzt und der Bruch $\frac{m}{n}$ lässt sich kürzen“, ist offenbar falsch. Also haben wir aus der Aussage „es gibt eine rationale Zahl a mit $a^2 = 2$“ eine falsche Aussage abgeleitet. Damit ist diese Aussage selbst falsch und es gilt stattdessen, was wir zeigen wollten: Es gibt keine rationale Zahl a mit $a^2 = 2$. □

Trotzdem finden wir einen Punkt auf der Zahlengeraden, der der Zahl $\sqrt{2}$ entspricht, nämlich den eindeutig bestimmten Punkt, der rechts von allen Zahlen in der Menge $A := \{x \in \mathbb{Q} \colon x < 0 \vee x^2 < 2\}$ und links von allen Zahlen in der Menge $B := \{x \in \mathbb{Q} \colon x > 0 \wedge x^2 > 2\}$ liegt.



Die Existenz eines Punktes auf der Zahlengeraden, dessen Abstand von 0 genau $\sqrt{2}$ ist, sieht man wie folgt: Auf der Strecke von 0 nach 1 errichte man ein Quadrat mit der Kantenlänge 1. Die Diagonale dieses Quadrats hat nach dem Satz von Pythagoras die Länge $\sqrt{2}$. Wenn wir von 0 ausgehend nach rechts die Länge der Diagonalen des Quadrats auf der Zahlengeraden abtragen, so erreichen wir den Punkt, der $\sqrt{2}$ entspricht.



Es gibt viele Punkte auf der Zahlengeraden, denen keine rationale Zahl entspricht. Wir können $\mathbb{Q}$ aber so zur Menge $\mathbb{R}$ der *reellen Zahlen* erweitern, dass jedem Punkt auf der Zahlengeraden eine reelle Zahl entspricht und umgekehrt jede reelle Zahl einem Punkt auf der Zahlengeraden. Wir können reelle Zahlen addieren und multiplizieren, wobei wir bei Einschränkung dieser Operationen auf $\mathbb{Q}$ genau die bekannten Operationen auf den rationalen Zahlen erhalten. Mit diesen Operationen bilden die reellen Zahlen einen Körper, wie die rationalen Zahlen auch.

Die Kleiner-Beziehung $<$ zwischen reellen Zahlen ist so erklärt, dass für reelle Zahlen a und b die Beziehung $a < b$ genau dann gilt, wenn der Punkt auf der Zahlengeraden, der a entspricht, links von dem Punkt liegt, der b entspricht. Es gelten dieselben Rechenregeln für $<$ auf $\mathbb{R}$ wie auf $\mathbb{Q}$.

Es gibt verschiedene Möglichkeiten, die reellen Zahlen ausgehend von den rationalen Zahlen zu konstruieren. Wir werden allerdings nicht näher auf die Konstruktion eingehen. Alle reellen Zahlen lassen sich als (eventuell unendliche) Dezimalbrüche darstellen. Die rationalen Zahlen entsprechen den Dezimalbrüchen, die entweder nach endlich vielen Nachkommastellen abbrechen oder periodisch werden.

Die reellen Zahlen, die nicht rational sind, heißen *irrational.* Beispiele für irrationale Zahlen sind $\sqrt{2}$, $\sqrt{3}$, e, π und $\sqrt[3]{5}$.

§3.3 Die Abzählbarkeit von $\mathbb{Q}$ und die Überabzählbarkeit von $\mathbb{R}$

Wir haben schon gesehen, dass es reelle Zahlen gibt, die nicht rational sind, wie zum Beispiel $\sqrt{2}$. In diesem Abschnitt werden wir sehen, dass es sogar viel mehr reelle als rationale Zahlen gibt.

Definition 3.4. *Zwei Mengen A und B heißen* gleichmächtig, *wenn es eine Bijektion $f\colon A \longrightarrow B$ gibt.*

Diese Definition ist auch für unendliche Mengen sinnvoll. So ist

$$f\colon \mathbb{Z} \longrightarrow \{a \in \mathbb{Z}\colon a \text{ ist gerade}\}; a \longmapsto 2a$$

eine Bijektion zwischen den ganzen Zahlen und den (positiven sowie negativen) geraden Zahlen. $\mathbb{Z}$ und die Menge aller geraden Zahlen sind also gleichmächtig.

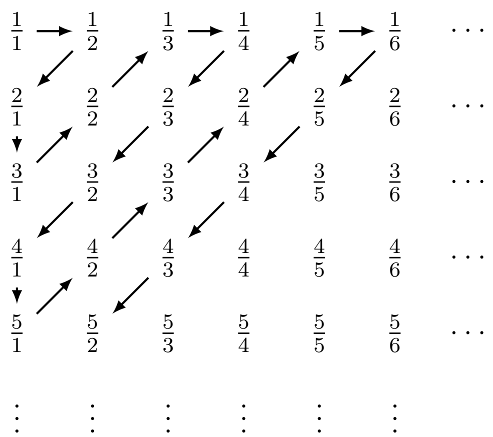
Definition 3.5. *Eine Menge M heißt abzählbar, wenn M entweder endlich ist oder es eine Bijektion $f\colon \mathbb{N} \longrightarrow M$ gibt. Eine Menge, die nicht abzählbar ist, heißt überabzählbar.*

Man kann leicht zeigen, dass eine Menge M genau dann abzählbar ist, wenn sie entweder leer ist oder es eine surjektive Abbildung $f\colon \mathbb{N} \longrightarrow M$ gibt. Eine Surjektion $f\colon \mathbb{N} \longrightarrow M$ nennt man eine *Aufzählung* von M. Eine Aufzählung f von M kann man einfach in der Form $f(1), f(2), \dots$ notieren.

So ist zum Beispiel $0, 1, -1, 2, -1, \dots$ eine Aufzählung von $\mathbb{Z}$. Die Menge der ganzen Zahlen ist also abzählbar. Etwas verblüffender ist folgender Satz, der von Cantor bewiesen wurde.

Satz 3.6. *Die Menge $\mathbb{Q}$ der rationalen Zahlen ist abzählbar.*

Beweis. Wir geben zunächst eine Aufzählung $q_1, q_2, \dots$ der Menge der rationalen Zahlen > 0 an. Man erhält die Aufzählung, indem man im folgenden Bild bei dem Bruch $\frac{1}{1}$ beginnt und den Pfeilen folgt.



Die Aufzählung lautet also

$$q_1 = \frac{1}{1}, \quad q_2 = \frac{1}{2}, \quad q_3 = \frac{2}{1}, \quad q_4 = \frac{3}{1}, \quad q_5 = \frac{2}{2}, \quad \dots$$

Die Tatsache, dass viele rationale Zahlen hierbei doppelt auftreten, zum Beispiel 1 als $\frac{1}{1}$ und $\frac{2}{2}$, spielt keine Rolle, da eine Aufzählung nicht injektiv sein muss. Es ist aber klar, das jede rationale Zahl > 0 in dieser Aufzählung irgendwann einmal auftritt.

Mit dieser Aufzählung der rationalen Zahlen > 0 können wir nun aber leicht eine Aufzählung aller rationalen Zahlen angeben:

$$0, q_1, -q_1, q_2, -q_2, \dots$$

leistet das Gewünschte. □

Satz 3.7. *Die Menge* $\mathbb{R}$ *der reellen Zahlen ist überabzählbar.*

Beweis. Wir zeigen, dass schon die Menge der reellen Zahlen, die echt größer als 0 und echt kleiner als 1 sind, überabzählbar ist. Wir führen einen Widerspruchsbeweis.

Angenommen, es gibt eine Aufzählung $s_1, s_2, s_3, \dots$ der reellen Zahlen s mit $0 < s < 1$. Die Zahlen s_n, $n \in \mathbb{N}$ lassen sich als Dezimalzahlen ohne Vorzeichen mit einer 0 vor dem Dezimalpunkt schreiben. Für alle $i, j \in \mathbb{N}$ sei s_{ij} die Ziffer, die in der j-ten Nachkommastelle der Dezimaldarstellung von s_i steht. Dann können wir die Aufzählung $s_1, s_2, \dots$ wie folgt notieren:

$$\begin{array}{ccl} s_1 & = & 0.s_{11}s_{12}s_{13}\dots \\ s_2 & = & 0.s_{21}s_{22}s_{23}\dots \\ s_3 & = & 0.s_{31}s_{32}s_{33}\dots \\ \vdots & & \quad\vdots \end{array}$$

Nun definieren wir eine weitere reelle Zahl a, die echt zwischen 0 und 1 liegt, die in der Aufzählung aber nicht auftritt. Das widerspricht der Annahme, dass $s_1, s_2, s_3, \dots$ eine Aufzählung der reellen Zahlen ist, die echt zwischen 0 und 1 liegen.

Wir geben die Nachkommastellen $a_1a_2a_3\dots$ der Zahl a an. Für $i \in \mathbb{N}$ sei

$$a_i := \begin{cases} 4, & \text{falls } s_{ii} \neq 4 \text{ ist und} \\ 5, & \text{sonst.} \end{cases}$$

Es ist klar, dass $a = 0.a_1a_2a_3\dots$ echt zwischen 0 und 1 liegt. a ist so gewählt, dass es sich an der i-ten Nachkommastelle von s_i unterscheidet. Da die Nachkommastellen von a nicht irgendwann konstant 0 oder konstant 9 werden, ist a damit von allen s_i, $i \in \mathbb{N}$, verschieden. □

§3.4 Teilbarkeit und Primzahlen

Wir haben bereits Teilbarkeit durch 2 betrachtet. Dennoch wiederholen wir die formale Definition von Teilbarkeit.

Definition 3.8. *Eine ganze Zahl a ist ein* Teiler *einer ganzen Zahl b, falls eine ganze Zahl c mit $b = a \cdot c$ existiert. Wenn a ein Teiler von b ist, so nennt man b ein* Vielfaches *von a. Ist a ein Teiler von b, so schreiben wir $a \mid b$. Ist a kein Teiler von b, so schreiben wir $a \nmid b$.*

Man beachte, dass jede ganze Zahl a die 0 teilt. Es ist nämlich $0 = 0 \cdot a$. Umgekehrt teilt 0 nur sich selber und keine andere ganze Zahl. Ebenso beachte man, dass für alle ganzen Zahlen a und b Folgendes gilt:

$$a \mid b \Leftrightarrow -a \mid b \Leftrightarrow -a \mid -b \Leftrightarrow a \mid -b$$

Damit kann man die Teilbarkeitsbeziehung zwischen ganzen Zahlen immer auf die Teilbarkeitsbeziehung zwischen natürlichen Zahlen zurückführen.

Satz 3.9. *Die Teilbarkeitsbeziehung $\mid$ hat folgende Eigenschaften:*

(1) Gilt $a \mid b$ und $b \mid c$, so gilt auch $a \mid c$.
(2) Aus $a_1 \mid b_1$ und $a_2 \mid b_2$ folgt $a_1 \cdot a_2 \mid b_1 \cdot b_2$.
(3) Aus $a \cdot b \mid a \cdot c$ und $a \neq 0$ folgt $b \mid c$.
(4) Aus $a \mid b_1$ und $a \mid b_2$ folgt für alle $c_1, c_2 \in \mathbb{Z}$ die Beziehung $a \mid b_1 \cdot c_1 + b_2 \cdot c_2$.

Beweis. (1)–(4) lassen sich leicht nachrechnen. Zum Beispiel kann man (4) wie folgt nachrechnen:

Wegen $a \mid b_1$ und $a \mid b_2$ existieren $d_1, d_2 \in \mathbb{Z}$ mit $b_1 = a \cdot d_1$ und $b_2 = a \cdot d_2$. Für alle $c_1, c_2 \in \mathbb{Z}$ gilt nun

$$b_1 \cdot c_1 + b_2 \cdot c_2 = a \cdot d_1 \cdot c_1 + a \cdot d_2 \cdot c_2 = a \cdot (d_1 \cdot c_1 + d_2 \cdot c_2).$$

Das zeigt $a \mid b_1 \cdot c_1 + b_2 \cdot c_2$. □

Definition 3.10. *Eine natürliche Zahl $n \geqslant 2$ heißt* Primzahl, *wenn n nur durch -1, 1, n und $-n$ teilbar ist. Die Zahlen ± 1 und $\pm n$ nennt man die* trivialen *Teiler von n.*

Satz 3.11 (Euklid). *Es gibt unendlich viele Primzahlen.*

Beweis. Wir führen wieder einen Widerspruchsbeweis. Angenommen, es gibt nur endlich viele Primzahlen $p_1, \dots, p_n$. Betrachte das Produkt $a = p_1 \cdot \ldots \cdot p_n$.

Sei p die kleinste natürliche Zahl $\geqslant 2$, die $a + 1$ teilt. Dann ist p eine Primzahl. Hat nämlich p einen Teiler q, der von -1, 1, p und $-p$ verschieden ist, so ist q oder $-q$ eine natürliche Zahl $\geqslant 2$, die $a + 1$ teilt und kleiner als p ist. Das widerspricht aber der Wahl von p als kleinstem Teiler von $a + 1$ mit $p \geqslant 2$.

Da p eine Primzahl ist, existiert ein $i \in \{1, \ldots, n\}$ mit $p = p_i$. Damit teilt p sowohl a als auch $a + 1$. Also teilt p auch $1 = (a + 1) - a$. Das widerspricht aber der Wahl von p als einer ganzen Zahl $\geqslant 2$. □

Ohne Beweis geben wir einen wichtigen Satz über die Darstellung natürlicher Zahlen als Produkte von Primzahlen an.

Satz 3.12. *Jede natürliche Zahl $n \geqslant 2$ ist ein Produkt der Form $p_1^{\alpha_1} \cdot \ldots \cdot p_k^{\alpha_k}$, wobei k eine natürliche Zahl $\geqslant 1$ ist, $p_1, \ldots, p_k$ paarweise verschiedene Primzahlen sind und $\alpha_1, \ldots, \alpha_k$ natürliche Zahlen sind. Dabei ist die Produktdarstellung $n = p_1^{\alpha_1} \cdot \ldots \cdot p_k^{\alpha_k}$ bis auf die Reihenfolge der Faktoren eindeutig.*

Zum Beispiel ist $12 = 2^2 \cdot 3$ und $500 = 2^2 \cdot 5^3$.

Eine wichtige Folgerung aus diesem Satz ist die Folgende:

Korollar 3.13. *Teilt eine Primzahl p ein Produkt $a \cdot b$ natürlicher Zahlen, so teilt p eine der beiden Zahlen a und b.*

Beweis. Wir schreiben a und b als Produkte von Primzahlen, $a = p_1^{\alpha_1} \cdot \ldots \cdot p_n^{\alpha_n}$ und $b = q_1^{\beta_1} \cdot \ldots \cdot q_m^{\beta_m}$. Dann ist

$$a \cdot b = p_1^{\alpha_1} \cdot \ldots \cdot p_n^{\alpha_n} \cdot q_1^{\beta_1} \cdot \ldots \cdot q_m^{\beta_m}.$$

Gilt $p \mid a \cdot b$, so existiert eine natürliche Zahl c mit $a \cdot b = p \cdot c$. Schreibt man nun c als Produkt von Primzahlen, so erhält man eine Darstellung von $a \cdot b$ als Produkt von Primzahlen, in dem der Faktor p auftritt. Wegen der Eindeutigkeit der Darstellung von $a \cdot b$ als Produkt von Primzahlen ist der Faktor p ein Element der Menge $\{p_1, \ldots, p_n, q_1, \ldots, q_m\}$. Damit teilt p die Zahl a oder die Zahl b. □

Die Aussage dieses Korollars wird falsch, wenn man die Bedingung weglässt, dass p eine Primzahl ist. Zum Beispiel teilt 6 das Produkt $4 \cdot 9$, während 6 weder 4 noch 9 teilt.

§3.5 Größter gemeinsamer Teiler, kleinstes gemeinsames Vielfaches und der euklidische Algorithmus

Definition 3.14. *Seien a und b natürliche Zahlen. Der* größte gemeinsame Teiler *von a und b ist die größte natürliche Zahl c, die sowohl a als auch b teilt. Der größte gemeinsame Teiler von a und b wird mit $\mathrm{ggT}(a, b)$ bezeichnet. Das* kleinste gemeinsame Vielfache *von a und b ist die kleinste natürliche Zahl, die sowohl von a als auch von b geteilt wird. Das kleinste gemeinsame Vielfache von a und b wird mit $\mathrm{kgV}(a, b)$ bezeichnet.*

Für ganze Zahlen a und $b \in \mathbb{Z} \setminus \{0\}$ definiert man $\mathrm{ggT}(a, b) = \mathrm{ggT}(|a|, |b|)$ und $\mathrm{kgV}(a, b) = \mathrm{kgV}(|a|, |b|)$. Des Weiteren setzt man für alle $a \in \mathbb{Z}$

$$\mathrm{ggT}(0, a) = \mathrm{ggT}(a, 0) = |a| \qquad \textit{und} \qquad \mathrm{kgV}(0, a) = \mathrm{kgV}(a, 0) = 0\,.$$

Der größte gemeinsame Teiler zweier natürlicher Zahlen a und b existiert, da es einerseits nur endliche viele gemeinsame Teiler von a und b gibt und andererseits 1 ein gemeinsamer Teiler von a und b ist. Das kleinste gemeinsame Vielfache von a und b existiert, da es mindestens ein gemeinsames Vielfaches gibt, nämlich $a \cdot b$, und jede nichtleere Menge natürlicher Zahlen ein kleinstes Element hat.

Ist die Zerlegung von a und b in Primfaktoren gegeben, so können wir $\mathrm{ggT}(a, b)$ und $\mathrm{kgV}(a, b)$ leicht berechnen. Sei p eine Primzahl, c ein gemeinsamer Teiler von a und b und $\alpha \in \mathbb{N}$, so dass $p^\alpha \mid c$ gilt. Dann gilt auch $p^\alpha \mid a$ und $p^\alpha \mid b$. Damit können wir den größten gemeinsamen Teiler von a und b wie folgt bestimmen:

In der Primfaktorzerlegung des größten gemeinsamen Teilers von a und b treten für jede Primzahl p die höchsten Potenzen p^α auf, die sowohl a als auch b teilen. Genauer: Sei $\{p_1, \dots, p_n\}$ die Menge der Primzahlen, die sowohl a als auch b teilen. Für jedes $i \in \{1, \dots, n\}$ sei α_i die größte natürliche Zahl, so dass $p_i^{\alpha_i}$ sowohl a als auch b teilt. Dann ist $p_1^{\alpha_1} \cdot \ldots \cdot p_n^{\alpha_n}$ der größte gemeinsame Teiler von a und b.

Das kleinste gemeinsame Vielfache von a und b lässt sich auf ähnliche Weise finden. Ist nämlich c ein Vielfaches von a und von b, so gilt für jede Primzahl p und jede natürliche Zahl α: Wenn p^α die Zahl a oder die Zahl b teilt, so teilt p^α auch c. Sei nun $\{p_1, \dots, p_n\}$ die Menge der Primzahlen, die a oder b teilen. Für jedes $i \in \{1, \dots, n\}$ sei $\alpha_i \in \mathbb{N}$ die größte natürliche Zahl, sodass $p_i^{\alpha_i} \mid a$ oder $p_i^{\alpha_1} \mid b$ gilt. Dann ist $p_1^{\alpha_1} \cdot \ldots \cdot p_n^{\alpha_n}$ das kleinste gemeinsame Vielfache von a und b.

Man beachte, dass man $\mathrm{ggT}(a, b)$ aus $\mathrm{kgV}(a, b)$ berechnen kann und umgekehrt. Es gilt nämlich die Beziehung

$$\mathrm{ggT}(a, b) \cdot \mathrm{kgV}(a, b) = a \cdot b.$$

Beispiel 3.15. (1) Sei $a = 60$ und $b = 70$. Dann ist $a = 2^2 \cdot 3 \cdot 5$ and $b = 2 \cdot 5 \cdot 7$. Es gilt $\mathrm{ggT}(a, b) = 2 \cdot 5 = 10$ und $\mathrm{kgV}(a, b) = 2^2 \cdot 3 \cdot 5 \cdot 7 = 420$.

(2) Sei

$$a = 2^4 \cdot 3 \cdot 5^2 \cdot 7 \cdot 13^4$$

und

$$b = 2^2 \cdot 5 \cdot 7^2 \cdot 13^3 \cdot 17 \cdot 23.$$

Dann ist

$$\mathrm{ggT}(a, b) = 2^2 \cdot 5 \cdot 7 \cdot 13^3$$

und

$$\mathrm{kgV}(a, b) = 2^4 \cdot 3 \cdot 5^2 \cdot 7^2 \cdot 13^4 \cdot 17 \cdot 23.$$

Die Zerlegung ganzer Zahlen in ihre Primfaktoren dauert bei Zahlen mit sehr großen Primfaktoren unter Umständen sehr lange. Diese Tatsache ist zum Beispiel wichtig für das weit verbreitete Verschlüsselungsverfahren RSA.

Es gibt aber einen schnellen Algorithmus, mit dem man den größten gemeinsamen Teiler zweier natürlicher Zahlen bestimmen kann, der auf Euklid zurückgeht und damit seit über 2000 Jahren bekannt ist. Der Algorithmus benutzt die Division mit Rest.

Satz 3.16. *Für alle $m \in \mathbb{Z}$ und alle $n \in \mathbb{N}$ gibt es eindeutig bestimmte Zahlen q und r mit $0 \leqslant r < n$ und $m = q \cdot n + r$.*

In der Darstellung $m = q \cdot n + r$ nennt man q den *Quotienten* von m und n und r den *Rest*. Die Funktion, die m und n den Quotienten q zuordnet, wird mit div bezeichnet. Die Funktion, die m und n den Rest r zuordnet, heißt mod. Es gilt also für alle $m \in \mathbb{Z}$ und alle $n \in \mathbb{N}$ die Gleichung

$$m = (m \operatorname{div} n) \cdot n + (m \bmod n).$$

Beispiel 3.17. (1) Sei $m = 27$ und $n = 12$. Dann ist $27 = 2 \cdot 12 + 3$. Der Quotient ist also 2 und der Rest 3.

(2) Sei $m = -10$ und $n = 3$. Dann ist $-10 = -4 \cdot 3 + 2$. Wir haben also $q = -4$ und $r = 2$. Es gilt zwar auch $-10 = -3 \cdot 3 - 1$, aber die Zahlen q und r werden bei der Division mit Rest immer so gewählt, dass $0 \leqslant r < n$ gilt.

Wir stellen Folgendes fest: Ist a ein gemeinsamer Teiler von m und n und gilt $m = q \cdot n + r$, so ist a auch ein Teiler von $r = m - q \cdot n$. Umgekehrt ist jeder gemeinsame Teiler von n und r auch ein Teiler von m. Es folgt, dass die beiden Zahlen m und n dieselben gemeinsamen Teiler haben wie die beiden Zahlen n und r. Für jede natürliche Zahl n ist $\operatorname{ggT}(n, 0) = n$. Das erklärt, warum der folgende Algorithmus zur Berechnung des größten gemeinsamen Teilers zweier natürlicher Zahlen funktioniert.
Der euklidische Algorithmus. Seien $m, n \in \mathbb{N}_0$ mit $m > n$.

(1) Falls $n = 0$ ist, dann wird m als größter gemeinsamer Teiler ausgegeben.
(2) Falls $n \neq 0$ ist, dann bestimme ganze Zahlen q und r mit $0 \leqslant r < n$ und $m = q \cdot n + r$.
(3) Setze $m := n$ und $n := r$ gehe zurück zu (1).

Nach unserer Vorbemerkung haben m und n in jedem Durchlauf der Schleife in diesem Algorithmus denselben größten gemeinsamen Teiler. Auf der anderen Seite wird n in jedem Durchlauf der Schleife echt kleiner. Also ist nach endlich vielen Schritten $n = 0$ und der Algorithmus terminiert.

Beispiel 3.18. (1) Wir berechnen wieder den größten gemeinsamen Teiler von 70 und 60, aber diesmal mit dem euklidischen Algorithmus. Setze zunächst $m = 70$ und $n = 60$. Wegen $n \neq 0$, führen wir eine Division mit Rest durch. Es gilt $70 = 1 \cdot 60 + 10$. Wir setzen $m := 60$ und $n := 10$. Immer noch gilt $n \neq 0$. Division mit Rest liefert $60 = 6 \cdot 10 + 0$. Wir setzen $m := 10$ und $n := 0$. Nun ist $n = 0$ und der größte gemeinsame Teiler von 10 und 0 ist 10. Die ursprünglichen Zahlen 70 und 60 haben denselben größten gemeinsamen Teiler und daher gilt $\operatorname{ggT}(70, 60) = 10$.

(2) Sei $m = 816$ und $n = 294$. Die Rechnung lautet nun wie folgt:

$$\begin{aligned} 816 &= 2 \cdot 294 + 228 \\ 294 &= 1 \cdot 228 + 66 \\ 228 &= 3 \cdot 66 + 30 \\ 66 &= 2 \cdot 30 + 6 \\ 30 &= 5 \cdot 6 + 0 \end{aligned}$$

Damit ergibt sich $\mathrm{ggT}(816, 294) = 6$.

§3.6 Modulare Arithmetik

Definition 3.19. *Es sei m eine natürliche Zahl. Zwei ganze Zahlen a und b sind* kongruent modulo m, *falls a und b denselben Rest bei Division durch m haben. Ist a kongruent zu b modulo m, so schreiben wir $a \equiv b \pmod{m}$.*

Wir stellen kurz fest, dass $a \equiv b \pmod{m}$ genau dann gilt, wenn $a - b$ durch m teilbar ist. Ist $a \equiv b \pmod{m}$, so existieren ganze Zahlen q_a, q_b und r mit $a = q_a \cdot m + r$, $b = q_b \cdot m + r$ und $0 \leqslant r < m$. Es gilt $a - b = (q_a \cdot m + r) - (q_b \cdot m + r) = (q_a - q_b) \cdot m$. Also ist $a - b$ durch m teilbar.

Sei umgekehrt $a - b$ durch m teilbar. Es gibt ganze Zahlen q_a, q_b, r_a und r_b mit $a = q_a \cdot m + r_a$, $b = q_b \cdot m + r_b$, $0 \leqslant r_a < m$ und $0 \leqslant r_b < m$. Es gilt

$$a - b = (q_a \cdot m + r_a) - (q_b \cdot m + r_b) = (q_a - q_b) \cdot m + (r_a - r_b).$$

Da $a - b$ durch m teilbar ist, ist auch $r_a - r_b$ durch m teilbar. Wegen $0 \leqslant r_a, r_b < m$ gilt $-m < r_a - r_b < m$. Wenn aber eine ganze Zahl, die echt größer als $-m$ und echt kleiner als m ist, durch m teilbar ist, so kann diese Zahl nur 0 sein. Damit ist $r_a - r_b = 0$. Also gilt $a \equiv b \pmod{m}$.

Beispiel 3.20. (1) $23 \equiv 8 \pmod{5}$, da $23 - 8 = 15$ durch 5 teilbar ist. Außerdem ist $23 = 4 \cdot 5 + 3$ und $8 = 1 \cdot 5 + 3$, also $23 \bmod 5 = 3 = 8 \bmod 5$.

(2) $-7 \equiv 2 \pmod{3}$, da $-7 = -3 \cdot 3 + 2$ und $2 = 0 \cdot 3 + 2$, also $-7 \bmod 3 = 2 = 2 \bmod 3$.

(3) $8227 \not\equiv 11 \pmod{3}$, da $8227 - 11 = 8216$ nicht durch 3 teilbar ist.

Wir betrachten die Menge aller ganzen Zahlen, die modulo m kongruent zu einer festen Zahl sind.

Beispiel 3.21. Sei $m = 3$. Die Menge der Zahlen, deren Rest bei Division durch 3 genau 0 ist, ist die Menge

$$K_0 = \{\dots, -6, -3, 0, 3, 6, 9, \dots\}.$$

Die Menge der Zahlen, bei denen der Rest genau 1 ist, ist

$$K_1 = \{\dots, -5, -2, 1, 4, 7, 10, \dots\}.$$

Für den Rest 2 erhalten wir die Menge

$$K_2 = \{\dots, -4, -1, 2, 5, 8, 11, \dots\}.$$

Definition 3.22. *Für jede natürliche Zahl m und jede ganze Zahl a heißt die Menge $[a]_m := \{b \in \mathbb{Z} \colon b \bmod m = a \bmod m\}$ die* Restklasse von a modulo m*.*

Wir stellen fest, dass es für jede natürliche Zahl m genau m verschiedene Restklassen modulo m gibt, nämlich $[0]_m, \dots, [m-1]_m$. Diese Restklassen sind paarweise disjunkt und es gilt $\mathbb{Z} = [0]_m \cup \dots \cup [m-1]_m$.

Folgender Satz sammelt die wichtigsten Regeln für das Rechnen mit Kongruenzen.

Satz 3.23. *Für alle $m \in \mathbb{N}$ und alle $a, b, c, d \in \mathbb{Z}$ gilt:*

(1) $a \equiv a \pmod m$
(2) $a \equiv b \pmod m \Rightarrow b \equiv a \pmod m$
(3) $a \equiv b \pmod m \wedge b \equiv c \pmod m \Rightarrow a \equiv c \pmod m$
(4) $a \equiv b \pmod m \Rightarrow -a \equiv -b \pmod m$
(5) $a \equiv c \pmod m \wedge b \equiv d \pmod m \Rightarrow a + b \equiv c + d \pmod m$
(6) *Gilt* $\mathrm{ggT}(c, m) = 1$*, so folgt aus* $c \cdot a \equiv c \cdot b \pmod m$ *die Kongruenz* $a \equiv b \pmod m$*.*

Diese Rechenregeln kann man direkt mit Hilfe der Definition von $a \equiv b \pmod m$ nachrechnen

Beispiel 3.24. In Satz 3.23 (6) muss man wirklich $\mathrm{ggT}(c, m) = 1$ voraussetzen. Zum Beispiel gilt $8 \cdot 3 \equiv 8 \cdot 6 \pmod 6$ aber nicht $3 \equiv 6 \pmod 6$.

Nützliche Operationen auf den reellen Zahlen, mit deren Hilfe man zum Beispiel auch die Funktionen div und mod berechnen kann, sind das Auf- und Abrunden.

Definition 3.25. *Für eine reelle Zahl r ist $\lceil r \rceil$ die kleinste ganze Zahl $\geqslant r$. Analog ist $\lfloor r \rfloor$ die größte ganze Zahl $\leqslant r$. Man nennt $\lceil \ \rceil$ die* obere Gaußklammer *und $\lfloor \ \rfloor$ die* untere Gaußklammer*.*

Beispiel 3.26. Es gilt

$$\begin{array}{rclrcl} \lceil 3.14 \rceil & = & 4, & \lfloor 3.14 \rfloor & = & 3, \\ \lceil \sqrt{2} \rceil & = & 2, & \lfloor \sqrt{2} \rfloor & = & 1, \\ \lceil 5 \rceil & = & 5, & \lfloor 5 \rfloor & = & 5, \\ \lceil -1.2 \rceil & = & \text{-1}, & \lfloor -1.2 \rfloor & = & \text{-2}. \end{array}$$

Für alle $m \in \mathbb{Z}$ und $n \in \mathbb{N}$ gilt $m \operatorname{div} n = \lfloor \frac{m}{n} \rfloor$ sowie $m \bmod n = m - n \cdot \lfloor \frac{m}{n} \rfloor$.

KAPITEL 4

Elementare Kombinatorik

§4.1 Fakultät, fallende Faktorielle und Binomialkoeffizienten

Definition 4.1. *Für eine endliche Menge M sei $|M|$ die Anzahl der Elemente von M.*

Satz 4.2. *(1) (Additionsregel) M sei eine endliche Menge und $M_1, \dots, M_n$ seien disjunkte Teilmengen von M mit $M = M_1 \cup \dots \cup M_n$. Dann gilt*

$$|M| = \sum_{i=1}^{n} |M_i|.$$

(2) (Multiplikationsregel) Seien $A_1, \dots, A_n$ endliche Mengen. Dann gilt

$$|A_1 \times \dots \times A_n| = |A_1| \cdot \ \dots \ \cdot |A_n| = \prod_{i=1}^{n} |A_i|.$$

(3) (Gleichheitsregel) Seien A und B zwei endliche Mengen. Dann gilt $|A| = |B|$ genau dann, wenn es eine Bijektion $f\colon A \longrightarrow B$ gibt.

Eine typische Anwendung der Multiplikationsregel ist die folgende:
Für ein $n \in \mathbb{N}$ betrachten wir n Kästchen $K_1, \dots, K_n$.

$$\begin{array}{cccc} \square & \square & \cdots & \square \\ K_1 & K_2 & \cdots & K_n \end{array}$$

In das erste Kästchen K_1 legen wir ein Objekt a_1, in das zweite Kästchen K_2 ein Objekt a_2 und so weiter. Wenn wir k_1 Möglichkeiten haben, das erste Kästchen K_1 zu belegen, k_2 Möglichkeiten, das zweite Kästchen K_2 zu belegen und so weiter, dann gibt es insgesamt $k_1 \cdot k_2 \cdot \ \dots \ \cdot k_n$ Möglichkeiten, die n Kästchen zu belegen.

Beispiel 4.3. (1) Eine Kennziffer bestehe aus drei Buchstaben und vier darauffolgenden Ziffern, wie $FAB3447$ oder $ARR5510$. Wieviele derartige Kennziffern gibt es?

Nach der Multiplikationsregel gibt es

$$26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 26^3 \cdot 10^4 = 175760000$$

Kennziffern.

(2) Wieviele Kennziffern wie in (1) gibt es, in denen kein Buchstabe und keine Ziffer doppelt vorkommen?

Nach der Multiplikationsregeln ergibt sich

$$26 \cdot 25 \cdot 24 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 78624000.$$

(3) Gegeben seien 15 unterschiedliche Bücher, von denen 8 auf Englisch, 3 auf Deutsch und 4 auf Russisch sind. Auf wie viele Arten kann man zwei Bücher in verschiedenen Sprachen auswählen?

Nach Additions- und Multiplikationsregel ergibt sich

$$8 \cdot 3 + 8 \cdot 4 + 3 \cdot 4 = 68.$$

Wir diskutieren im Folgenden fünf grundlegende Fragestellungen, die wir Grundaufgaben nennen.

Vorher definieren wir noch Tupel der Länge 0.

Definition 4.4. *Für eine beliebige Menge M sei $\varnothing$ das eindeutig bestimmte 0-Tupel von Elementen von M. Mit anderen Worten, $M^0 = \{\varnothing\}$.*

Grundaufgabe 1. Es seien $n, k \in \mathbb{N}_0$. Wie viele k-Tupel von Elementen einer n-elementigen Menge gibt es?

Antwort: n^k

Diese Antwort ergibt sich sofort mit Hilfe der Multiplikationsregel. □

Beispiel 4.5. (1) Sei $M = \{a, b\}$. Dann gibt es $2^3 = 8$ 3-Tupel von Elementen von M. Es gilt

$$M^3 = \{(a, a, a), (a, a, b), (a, b, a), (a, b, b), (b, a, a), (b, a, b), (b, b, a), (b, b, b)\}.$$

(2) Sei $M = \{a, b, c, d, e, f, g\}$. Dann gibt es $7^3 = 343$ 3-Tupel von Elementen von M.

Grundaufgabe 2. Es seien $n, k \in \mathbb{N}_0$. Wieviele k-Tupel von Elementen einer n-elementigen Menge gibt es, in denen kein Element doppelt vorkommt?

Antwort: Falls $k \geqslant 1$ ist, so gibt es nach der Multiplikationsregel $n \cdot (n-1) \cdot \ldots \cdot (n-(k-1))$ k-Tupel von Elementen einer n-elementigen Mengen, in denen kein Element doppelt vorkommt. Ist $k = 0$, so gibt es genau ein k-Tupel. □

Diese Antwort legt folgende Definition nahe:

Definition 4.6. *Für $n, k \in \mathbb{N}_0$ sei*

$$n^{\underline{k}} := \begin{cases} n \cdot (n-1) \cdot \ldots \cdot (n-k+1), & \text{falls } k \geqslant 1 \text{ und} \\ 1, & \text{sonst.} \end{cases}$$

Beispiel 4.7. (1) $7^{\underline{0}} = 1$
(2) $7^{\underline{1}} = 7$
(3) $7^{\underline{2}} = 7 \cdot 6 = 42$
(4) $7^{\underline{3}} = 7 \cdot 6 \cdot 5 = 210$

Beispiel 4.8. Sei $M = \{a, b, c, d, e, f, g\}$. Dann gibt es $7^{\underline{3}} = 210$ 3-Tupel von Elementen von M, in denen kein Element doppelt vorkommt.

Definition 4.9. *Sei M eine Menge. Eine* Permutation *von M ist eine Bijektion $\pi\colon M \longrightarrow M$.*

Beispiel 4.10. Sei $M = \{1, 2, 3\}$. Wir definieren $\pi\colon M \longrightarrow M$ durch $\pi(1) = 3$, $\pi(2) = 1$ und $\pi(3) = 2$. Dann ist π eine Permutation auf M.

Ist M eine endliche Menge $\{m_1, \dots, m_n\}$, wobei wir annehmen, dass die m_i paarweise verschieden sind, so kann man eine Permutation $\pi\colon M \longrightarrow M$ in der Form

$$\begin{pmatrix} m_1 & m_2 & \dots & m_n \\ \pi(m_1) & \pi(m_2) & \dots & \pi(m_n) \end{pmatrix}$$

darstellen. In dieser Schreibweise lautet die Permutation aus Beispiel 4.10

$$\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

Aus der Grundaufgabe 2 ergibt sich, dass die Anzahl der Permutationen einer n-elementigen Menge genau $n^{\underline{n}} = n \cdot (n-1) \cdot \ \dots \ \cdot 1$ ist. Anstelle von $n^{\underline{n}}$ schreibt man üblicherweise $n!$ (gelesen „n Fakultät“).

Beispiel 4.11. $0! = 0^{\underline{0}} = 1$, $1! = 1^{\underline{1}} = 1$, $2! = 2^{\underline{2}} = 2 \cdot 1 = 2$, $10! = 10^{\underline{10}} = 10 \cdot 9 \cdot \ldots \cdot 2 \cdot 1 = 3628800$.

Beispiel 4.12. (1) Sei $M = \{1, 2, 3\}$. Dann gibt es genau $3! = 3 \cdot 2 \cdot 1 = 6$ Permutationen von M:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

(2) Sei $M = \{a, b, c, d, e, f, g\}$. Dann gibt es $7! = 5040$ Permutationen von M.

Grundaufgabe 3. Es sei $n \geqslant k \geqslant 0$. Wieviele k-elementige Teilmengen einer n-elementigen Menge gibt es?

Antwort: Es gibt $\frac{n^{\underline{k}}}{k!}$ k-elementige Teilmengen einer n-elementigen Menge.

Das kann man wie folgt sehen: Nach Grundaufgabe 2 wissen wir schon, dass es für eine n-elementige Menge M genau $n^{\underline{k}}$ k-Tupel von Elementen von M gibt, in denen kein Element doppelt vorkommt. Für jedes k-Tupel $(m_1, \dots, m_k)$ von Elementen von M können wir nun die k-elementige Menge $\{m_1, \dots, m_k\}$ betrachten. Jede k-elementige Teilmenge von M entsteht auf diese Weise. Für jede k-elementige Teilmenge $\{m_1, \dots, m_k\}$ von M gibt es genau $k!$ k-Tupel, deren Komponenten genau die Elemente $m_1, \dots, m_k$ sind. Das liegt daran, dass jedes solche k-Tupel einer Permutation der Menge

$\{m_1, \ldots, m_k\}$ entspricht. Da also je $k!$ k-Tupel dieselbe k-elementige Teilmenge von M liefern, gibt es insgesamt $\frac{n^{\underline{k}}}{k!}$ k-elementige Teilmengen von M. $\square$

Für die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge schreibt man auch $\binom{n}{k}$. Es gilt

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!} = \frac{n^{\underline{k}} \cdot (n-k)!}{k! \cdot (n-k)!} = \frac{n!}{k! \cdot (n-k)!}.$$

Ist $k \geqslant 1$, so können wir auch

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k \cdot (k-1) \cdot \ldots \cdot 1}$$

schreiben.

Definition 4.13. *Für $n, k \in \mathbb{N}_0$ mit $n \geqslant k \geqslant 0$ nennt man die Zahl $\binom{n}{k} = \frac{n^{\underline{k}}}{k!}$ einen* Binomialkoeffizienten*.*

Beispiel 4.14. Sei $M = \{a, b, c, d, e, f, g\}$. Dann hat M genau

$$\binom{7}{3} = \frac{7 \cdot 6 \cdot 5}{3 \cdot 2 \cdot 1} = 35$$

3-elementige Teilmengen.

Satz 4.15 (Rekursive Berechnung der Binomialkoeffizienten)**.** *Für alle $n, k \in \mathbb{N}$ mit $n \geqslant 2$ und $1 \leqslant k \leqslant n-1$ gilt*

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

Beweis. Es gilt

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

$\square$

Wir ordnen die Binomialkoeffizienten wie folgt im *Pascalschen Dreieck* an:

$$\begin{array}{ccccccccc} & & & & \binom{0}{0} & & & & \\ & & & \binom{1}{0} & & \binom{1}{1} & & & \\ & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\ & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\ \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} \\ \iddots & & & & \vdots & & & & \ddots \end{array}$$

Dabei ist jeder Binomialkoeffizient im Innern des (unendlichen) Dreiecks nach Satz 4.15 die Summe der beiden Binomialkoeffizienten, die sich rechts und links darüber befinden. Auf diese Weise lassen sich leicht die Werte der Binomialkoeffizienten berechnen:

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
\iddots & & & & & & \vdots & & & & & & \ddots
\end{array}
$$

Die Binomialkoeffizienten verdanken ihren Namen dem folgenden Satz:

Satz 4.16 (Binomischer Lehrsatz)**.** *Seien $a, b \in \mathbb{R}$. Dann gilt für alle $n \in \mathbb{N}_0$*

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

Man beachte, dass der Ausdruck $\sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i$ auch für $n = 0$ definiert ist, während $a^n + \binom{n}{1} a^{n-1} b + \ldots + \binom{n}{n-1} a b^{n-1} + b^n$ nur für $n \geqslant 3$ sinnvoll ist. Das zeigt den Vorteil der Schreibweise mit dem Summenzeichen gegenüber der unexakten Pünktchen-Schreibweise.

Beweis. Wir beweisen den Satz durch Induktion über n.

Induktionsanfang. Für $n = 0$ gilt

$$(a+b)^n = (a+b)^0 = 1 = a^0 b^0.$$

Induktionsschritt. Wir nehmen an, dass

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

für ein gewisses $n \in \mathbb{N}_0$ gilt (Induktionsannahme).

Dann gilt

$$\begin{aligned}(a+b)^{n+1} = (a+b)^n \cdot (a+b) &\overset{\text{I.A.}}{=} \left(\sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i\right) \cdot (a+b) \\ &= \left(\sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i\right) \cdot a + \left(\sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i\right) \cdot b \\ &= \sum_{i=0}^{n} \binom{n}{i} a^{n+1-i} b^i + \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^{i+1} \\ &= \sum_{i=0}^{n} \binom{n}{i} a^{n+1-i} b^i + \sum_{i=1}^{n+1} \binom{n}{i-1} a^{n+1-i} b^i \\ &= a^{n+1} + \sum_{i=1}^{n} \left(\binom{n}{i} + \binom{n}{i-1}\right) a^{n+1-i} b^i + b^{n+1} = \sum_{i=0}^{n+1} \binom{n+1}{i} a^{n+1-i} b^i,\end{aligned}$$

wobei sich das letzte Gleichungszeichen aus Satz 4.15 ergibt. □

Beispiel 4.17.

- Für $n = 2$ ist Satz 4.16 die bekannte binomische Formel $(a+b)^2 = a^2 + 2ab + b^2$.
- Für $n = 3$ gilt $(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$.
- Für $n = 4$ gilt $(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$.

Wir bemerken noch zwei wichtige Regeln für Binomialkoeffizienten.

Korollar 4.18.

(1) *Für alle $n \in \mathbb{N}_0$ gilt $2^n = \sum_{i=0}^{n} \binom{n}{i}$.*
(2) *Für alle $n, k \in \mathbb{N}_0$ mit $n \geqslant k$ gilt $\binom{n}{k} = \binom{n}{n-k}$.*

Beweis. (1) Nach Satz 4.16 gilt

$$2^n = (1+1)^n = \sum_{i=0}^{n} \binom{n}{i} 1^{n-i} 1^i = \sum_{i=0}^{n} \binom{n}{i}.$$

(2) Es gilt

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{n!}{(n-k)! \cdot (n-(n-k))!} = \binom{n}{n-k}.$$

Wir geben noch ein weiteres Argument für diese Gleichung an. Sei M eine n-elementige Menge. Die Komplementbildung ist eine Bijektion zwischen der Menge der k-elementigen Teilmengen von M und der Menge der $(n-k)$-elementigen Teilmengen von M. Damit gibt es genausoviele k-elementige Teilmengen von M wie $(n-k)$-elementige. Nach Grundaufgabe 3 und der Gleichheitsregel gilt also $\binom{n}{k} = \binom{n}{n-k}$. □

Korollar 4.19. *Sei $n \in \mathbb{N}_0$ und sei M eine n-elementige Menge. Dann hat $\wp(M)$ genau 2^n Elemente.*

Wir geben zwei Beweise dieser wichtigen Tatsache an.

Erster Beweis. Für $k \in \mathbb{N}_0$ mit $0 \leqslant k \leqslant n$ sei P_k die Menge der k-elementigen Teilmengen von M. Nach Grundaufgabe 3 wissen wir, dass $|P_k| = \binom{n}{k}$ gilt. Außerdem sind die P_k disjunkt und es gilt $\wp(M) = P_0 \cup \cdots \cup P_n$. Nach der Additionsregel und nach Korollar 4.18 ist damit $|\wp(M)| = \sum_{k=0}^{n} \binom{n}{k} = 2^n$. □

Zweiter Beweis. Sei

$$P := \{f\colon f \text{ ist eine Funktion von } M \text{ nach } \{0,1\}\}.$$

Da M genau n Elemente hat, können wir M als $\{m_1, \ldots, m_n\}$ schreiben. Jeder Funktion $f \in P$ ordnen wir nun das n-Tupel $(f(m_1), f(m_2), \ldots, f(m_n))$ zu. Das liefert eine Bijektion zwischen der Menge P und der Menge $\{0,1\}^n$. Nach der Gleichheitsregel ist also $|P| = |\{0,1\}^n|$. Nach Grundaufgabe 1 ist $|\{0,1\}^n| = 2^n$. Damit ist $|P| = 2^n$.

Für jede Menge $A \subseteq M$ betrachte die *charakteristische Funktion* $\chi_A\colon M \longrightarrow \{0,1\}$ von A, die wie folgt definiert ist: Für jedes $x \in M$ sei

$$\chi_A(x) = \begin{cases} 0, & \text{falls } x \notin A \text{ und} \\ 1, & \text{falls } x \in A. \end{cases}$$

Die Abbildung $A \longmapsto \chi_A$ ist eine Bijektion von $\wp(M)$ nach P. Wieder nach der Gleichheitsregel folgt daraus $|\wp(M)| = |P| = 2^n$. □

Grundaufgabe 4. Sei $n \in \mathbb{N}$ und $k \in \mathbb{N}_0$. Es seien n Gefäße $K_1, \ldots, K_n$ gegeben, auf die k ununterscheidbare Kugeln verteilt werden sollen. Wieviele Möglichkeiten gibt es, die Kugeln zu verteilen?

Antwort. Es gibt $\binom{n+k-1}{k}$ Möglichkeiten, die Kugeln zu verteilen.

Das sehen wir wie folgt ein: Wir beschreiben die Verteilung der Kugeln durch eine Folge von Nullen und Einsen. Wir beginnen mit so vielen Nullen, wie Kugeln in P_1 liegen. Dann schreiben wir eine Eins. Es folgen so viele Nullen, wie in P_2 liegen. Darauf schreiben wir wieder eine Eins und so weiter.

Sei zum Beispiel $n = 4$ und $k = 5$. Angenommen, in P_1 liegen 2 Kugeln, in P_2 eine, in P_3 keine und in P_4 die restlichen zwei. Das liefert die Folge 00101100.

Bei n Gefäßen und k Kugeln erhalten wir eine Folge mit k Nullen und $n-1$ Einsen. Umgekehrt ist klar, dass wir aus jeder Folge mit k Nullen und $n-1$ Einsen eindeutig eine Belegung der n Gefäße mit k Kugeln ablesen können.

Mit anderen Worten, es gibt eine Bijektion zwischen der Menge der Belegungen der n Gefäße mit k Kugeln und den Folgen der Länge $n+k-1$ mit $n-1$ Einsen und k Nullen. Die Folgen der Länge $n+k-1$ mit $n-1$ Einsen und k Nullen können wir als charakterische Funktionen von $(n-1)$-elementigen Teilmengen einer $n+k-1$-elementigen Menge interpretieren. Damit gibt es genau $\binom{n+k-1}{n-1} = \binom{n+k-1}{k}$ mögliche Belegungen der n Gefäße mit k Kugeln. □

Beispiel 4.20. Angenommen, k Abgeordnete wählen je einen von n Kandidaten. Keiner der Abgeordneten enthält sich. Dann gibt es $\binom{n+k-1}{k}$ mögliche Verteilungen der k Stimmen auf die n Kandidaten.

Grundaufgabe 5. Gegeben seien r verschiedene Zeichen $Z_1, \dots, Z_r$. Wie viele verschiedene Zeichenfolgen der Länge n kann man aus den Zeichen $Z_1, \dots, Z_r$ bilden, wenn man verlangt, dass das Zeichen Z_1 genau n_1-mal auftritt, das Zeichen Z_2 genau n_2-mal und so weiter.

Beispiel 4.21. Wie viele Wörter können aus den Buchstaben des Wortes ANAGRAMM gebilden werden, wobei alle Buchstaben verwendet werden sollen?

Die Zeichen, die in diesem Beispiel auftreten, sind Z_1 =A, Z_2 =G, Z_3 =M, Z_4 =N und Z_5 =R. Hier kommt das A dreimal vor. Es darf also auch dreimal verwendet werden und $n_1 = 3$. Analog sind $n_2 = 1$, $n_3 = 2$, $n_4 = 1$ und $n_5 = 1$.

Eine Zeichenkette, die aus den Buchstaben in ANAGRAMM gebildet ist, wie zum Beispiel AMMAGRAN, ändert sich nicht, wenn wir die A's untereinander vertauschen oder wenn wir die M's vertauschen. Die drei A's können wir auf 3!=6 Arten permutieren und die M's auf 2!=2 Arten. Insgesamt gibt es also $3! \cdot 2! = 12$ Permutationen der Zeichen in AMMAGRAN, die genau dieselbe Zeichenfolge liefern.

Das gleich Argument zeigt für jede Zeichenfolge aus den Buchstaben von ANAGRAMM, dass es genau 12 Permutationen der Zeichen gibt, die dieselbe Zeichenfolge liefern.

Ingesamt gibt es $8! = 40320$ Permutationen der acht Zeichen in dem Wort ANAGRAMM, von denen wir aber jeweils Klassen von 12 Permutationen nicht unterscheiden können. Damit gibt es $\frac{8!}{3! \cdot 2!} = \frac{40320}{12} = 3360$ mögliche Zeichenfolgen aus den Buchstaben des Wortes ANAGRAMM.

Antwort zu Grundaufgabe 5. Es gibt genau

$$\frac{(n_1 + \ldots + n_r)!}{n_1! \cdot \ldots \cdot n_r!}$$

Zeichenfolgen aus den Zeichen $Z_1, \dots, Z_r$, in denen für jedes $i \in \{1, \dots, r\}$ das Zeichen Z_i genau n_i-mal vorkommt.

Das sieht man genauso, wie in Beispiel 4.21. Wir betrachten die Zeichenfolge W, in der zunächst n_1-mal das Zeichen Z_1 auftritt, dann n_2-mal das Zeichen Z_2 und so weiter. Die Wörter aus den Zeichen $Z_1, \dots, Z_r$, die in der Grundaufgabe 5 gebildet werden dürfen, entstehen durch Permutation der Zeichen in W. W hat die Länge $n_1 + \ldots + n_r$. Also gibt es $(n_1 + \ldots + n_r)!$ solcher Permutationen.

Die Menge dieser Permutationen zerfällt wieder in Klassen disjunkter Mengen, die ununterscheidbare Zeichenfolgen liefern. Die Größe einer jeden solchen Klasse ist $n_1! \cdot \ldots \cdot n_r!$, nämlich die Anzahl der Permutationen der Zeichen Z_1 in einem Wort, multipliziert mit der Anzahl der Permutationen der Zeichen Z_2 in einem Wort und so weiter.

Insgesamt erhalten wir $\frac{(n_1+\ldots+n_r)!}{n_1!\cdot\ldots\cdot n_r!}$ Zeichenfolgen.

Definition 4.22. *Seien $n_1, \ldots, n_r \in \mathbb{N}_0$ und $n = \sum_{i=1}^r n_i$. Dann nennt man*

$$\binom{n}{n_1, \ldots, n_r} = \frac{n!}{n_1! \cdot \ldots \cdot n_r!}$$

einen Multinomialkoeffizienten.

Wegen $0! = 1$ sind die Multinomialkoeffizienten auch definiert, wenn für ein oder mehrere $i \in \{1, \ldots, r\}$ die Gleichung $n_i = 0$ gilt. Auch die Lösung der Grundaufgabe 5 stimmt in dieser Situation. Extrem ist der Fall $n = n_1 + \ldots + n_r = 0$. Aber auch hier geht alles glatt. Es gibt genau eine Zeichenfolge der Länge 0, die leere Zeichenfolge.

Im Spezialfall $r = 2$ sind die Multinomialkoeffizienten genau die schon betrachteten Binomialkoeffizienten. Sei nämlich $n = n_1 + n_2$. Dann gilt

$$\binom{n}{n_1, n_2} = \frac{n!}{n_1! \cdot n_2!} = \frac{n!}{n_1! \cdot (n-n_1)!} = \binom{n}{n_1} = \frac{n!}{n_2! \cdot (n-n_2)!} = \binom{n}{n_2}.$$

§4.2 Ziehen von Elementen einer Menge

Die ersten vier Grundaufgaben gehen alle auf dieselbe grundlegende Frage zurück: Wieviele Möglichkeiten gibt es, k Elemente aus einer n-elementigen Menge zu ziehen?

Dabei wird auf unterschiedliche Weisen gezogen, und die Ergebnisse werden auf unterschiedliche Arten gezählt. Es gibt folgende Möglichkeiten:

(1) Ziehen mit Zurücklegen, wobei die Reihenfolge, in der die Elemente gezogen werden, berücksichtigt wird.
(2) Ziehen ohne Zurücklegen, mit Berücksichtigung der Reihenfolge.
(3) Ziehen ohne Zurücklegen, ohne Berücksichtigung der Reihenfolge.
(4) Ziehen mit Zurücklegen, ohne Berücksichtigung der Reihenfolge.

Satz 4.23. *Seien $n, k \in \mathbb{N}_0$. Dann gibt es genau n^k Möglichkeiten, k Elemente mit Zurücklegen aus einer n-elementigen Menge zu ziehen, wobei die Reihenfolge, in der die Elemente gezogen werden, berücksichtigt wird.*

Beweis. Die Möglichkeiten, die k Elemente zu ziehen, entsprechen genau den k-Tupeln von Elementen der n-elementigen Menge. Gemäß der Lösung von Grundaufgabe 1 gibt es also genau n^k Möglichkeiten. □

Satz 4.24. *Seien $n, k \in \mathbb{N}_0$ mit $k \leqslant n$. Dann gibt es genau $n^{\underline{k}}$ Möglichkeiten, k Elemente ohne Zurücklegen aus einer n-elementigen Menge zu ziehen, wobei die Reihenfolge, in der die Elemente gezogen werden, berücksichtigt wird.*

Beweis. Die Möglichkeiten, die k Elemente zu ziehen, entsprechen genau den k-Tupeln von Elementen der n-elementigen Menge, in denen kein Element doppelt vorkommt. Gemäß der Lösung von Grundaufgabe 2 gibt es also genau $n^{\underline{k}}$ Möglichkeiten. □

Satz 4.25. *Seien $n, k \in \mathbb{N}_0$ mit $k \leqslant n$. Dann gibt es genau $\binom{n}{k}$ Möglichkeiten, k Elemente ohne Zurücklegen aus einer n-elementigen Menge zu ziehen, wobei die Reihenfolge, in der die Elemente gezogen werden, nicht berücksichtigt wird.*

Beweis. Die Möglichkeiten, die k Elemente zu ziehen, entsprechen genau den k-elementigen Teilmengen der n-elementigen Menge. Gemäß der Lösung von Grundaufgabe 3 gibt es also genau $\binom{n}{k}$ Möglichkeiten. □

Satz 4.26. *Seien $n, k \in \mathbb{N}_0$. Dann gibt es genau $\binom{n+k-1}{k}$ Möglichkeiten, k Elemente mit Zurücklegen aus einer n-elementigen Menge zu ziehen, wobei die Reihenfolge, in der die Elemente gezogen werden, nicht berücksichtigt wird.*

Beweis. Wir führen den Satz auf die Lösung der Grundaufgabe 4 zurück. Wenn die Reihenfolge, in der die Elemente gezogen werden, keine Rolle spielt, so müssen wir nur zählen, wie oft jedes Element der n-elementigen Menge gezogen wurde.

Diese Situation können wir wie folgt kodieren: Sei $M = \{a_1, \ldots, a_n\}$ eine n-elementige Menge. Für jedes Element a_i der n-elementigen Menge M betrachten wir ein Gefäß K_i. Nun ziehen wir die k Elemente der n-elementigen Menge mit Zurücklegen. Immer wenn wir ein Element a_i ziehen, tun wir eine Kugel in das Gefäß K_i.

Jede Verteilung von k Kugeln auf die Gefäße $K_1, \ldots, K_n$ entspricht genau einer Ziehung von k Elementen der n-elementigen Menge und umgekehrt. Nach der Lösung von Grundaufgabe 4 gibt es $\binom{n+k-1}{k}$ mögliche Verteilungen von k Kugeln auf die n Gefäße. Also gibt es auch $\binom{n+k-1}{k}$ Möglichkeiten, k Elemente ohne Zurücklegen aus einer n-elementigen Menge zu ziehen, wenn man die Reihenfolge, in der die Elemente gezogen werden, nicht berücksichtigt. □

§4.3 Der Multinomialsatz

Satz 4.27 (Multinomialsatz)**.** *Seien $r, n \in \mathbb{N}_0$ mit $r \geqslant 1$. Für alle $x_1, \ldots, x_r \in \mathbb{R}$ gilt*

$$(x_1 + \ldots + x_r)^n = \sum_{n_1+\ldots+n_r=n} \binom{n}{n_1, \ldots, n_r} x_1^{n_1} \cdot \ldots \cdot x_r^{n_r}.$$

Diese Summe läuft über alle r-Tupel $(n_1, \ldots, n_r) \in \mathbb{N}_0^r$ mit $n_1 + \ldots + n_r = n$.

Man beachte, dass man für $r = 2$ aus dem Multinomialsatz genau den Binomialsatz erhält.

Beweis. Den Binomialsatz hatten wir mittels vollständiger Induktion bewiesen. Für den Multinomialsatz geben wir einen kombinatorischen Beweis an, der nur die

Lösung von Grundaufgabe 5 benutzt. Wir können

$$(x_1 + \ldots + x_r)^n = \underbrace{(x_1 + \ldots + x_r) \cdot \ldots \cdot (x_1 + \ldots + x_r)}_{n \text{ Faktoren}}$$

durch Ausmultiplizieren berechnen. Für $n_1, \ldots, n_r \in \mathbb{N}_0$ mit $n_1 + \ldots + n_r = n$ zählen wir, wie oft das Produkt $x_1^{n_1} \cdot \ldots \cdot x_r^{n_r}$ beim Ausmultiplizieren auftritt. Beim Ausmultiplizieren wählen wir aus jedem der n Faktoren $(x_1 + \ldots + x_r)$ eine Variable aus. Wir wählen also ein Wort der Länge n aus den Zeichen $x_1, \ldots, x_r$. Um das Produkt $x_1^{n_1} \cdot \ldots \cdot x_r^{n_r}$ zu erhalten, muss in dem Wort, das wir auswählen, die Variable x_1 genau n_1-mal auftreten, die Variable x_2 n_2-mal und so weiter. Nach der Lösung von Grundaufgabe 5 gibt es genau $\binom{n}{n_1,\ldots,n_r}$ Wörter der Länge n, in denen für alle $i \in \{1, \ldots, r\}$ das Zeichen x_i genau n_i-mal auftritt. Damit ist der Koeffizient vor dem Produkt $x_1^{n_1} \cdot \ldots \cdot x_r^{n_r}$, der sich beim Ausmultiplizieren von $(x_1 + \ldots + x_r)^n$ ergibt, die Zahl $\binom{n}{n_1,\ldots,n_r}$. Das zeigt den Multinomialsatz. $\square$

Beispiel 4.28. Nach Ausmultiplizieren von $(x + y + z)^{10}$ ist der Koeffizient vor dem Produkt $x^5 y^3 z^2$ die Zahl

$$\binom{10}{5,3,2} = \frac{10!}{5! \cdot 3! \cdot 2!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{3! \cdot 2!} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{2} = 7 \cdot 4 \cdot 9 \cdot 10 = 2520.$$

§4.4 Das Schubfachprinzip (pigeonhole principle)

Satz 4.29 (Schubfachprinzip)**.** *Seien $m, n \in \mathbb{N}$ mit $m > n$. Wenn m Objekte auf n Fächer verteilt werden, so gibt es mindestens ein Fach mit mindestens zwei Objekten.*

Eine andere Formulierung dieses Satzes ist die folgende: Sind m und n natürliche Zahlen mit $m > n$, so gibt es keine injektive Abbildung $f\colon \{1, \ldots, m\} \longrightarrow \{1, \ldots, n\}$.

Beispiel 4.30. In einer Menge von 13 Menschen gibt es mindestens zwei, die im gleichen Monat Geburtstag haben. In einer Menge von 367 Menschen gibt es mindestens zwei, die am gleichen Tag Geburtstag haben. (Der 29. Februar ist ein möglicher Geburtstag.)

Wir beweisen eine Verstärkung von Satz 4.29.

Satz 4.31. *Seien $m, n \in \mathbb{N}$. Wenn m Objekte auf n Fächer verteilt werden, so gibt es mindestens ein Fach mit mindestens $\lceil \frac{m}{n} \rceil$ Objekte.*

Beweis. Angenommen, das ist nicht der Fall. Dann enthält jedes Fach höchstens $\lceil \frac{m}{n} \rceil - 1$ Objekte.

Damit enthalten die Fächer insgesamt nicht mehr als $n \cdot \left(\lceil \frac{m}{n} \rceil - 1\right)$ Objekte. Es gilt also

$$m \leqslant n \cdot \left(\left\lceil \frac{m}{n} \right\rceil - 1\right).$$

Umformen liefert

$$1 \leqslant \left\lceil \frac{m}{n} \right\rceil - \frac{m}{n}.$$

Das ist aber unmöglich, da für jede reelle Zahl a der Abstand zwischen $\lceil a \rceil$ und a echt kleiner als 1 ist. $\square$

Es gibt auch Versionen des Schubfachprinzips für unendliche Mengen.

Satz 4.32. *Sei M eine unendliche Menge und $n \in \mathbb{N}$. Sind $M_1, \dots, M_n$ Teilmengen von M mit $M = M_1 \cup \dots \cup M_n$, so ist eine der Mengen $M_1, \dots, M_n$ unendlich.*

Beweis. Sind die Mengen $M_1, \dots, M_n$ alle endlich, so sei m maximale Mächtigkeit einer der Mengen $M_1, \dots, M_n$. Dann hat $M_1 \cup \dots \cup M_n$ höchstens die Mächtigkeit $m \cdot n$ und ist damit endlich. Das widerspricht aber unserer Annahme, dass $M = M_1 \cup \dots \cup M_n$ unendlich ist. $\square$

Aus diesem Satz folgt sofort, dass für jede Funktion f von einer unendlichen Menge A in eine endliche Menge B ein $b \in B$ existiert, so dass die Menge

$$\{a \in A \colon f(a) = b\}$$

unendlich ist.

§4.5 Das Prinzip der Inklusion und Exklusion (Siebformel)

Seien $A_1, \dots, A_n$ endliche Mengen. Wir suchen eine Formel für die Mächtigkeit der Vereinigung der Mengen A_i, $i \in \{1, \dots, n\}$, also für die Mächtigkeit $|A_1 \cup \dots \cup A_n|$ der Menge $A_1 \cup \dots \cup A_n$.

Wir betrachten zunächst den Fall zweier Mengen, A_1 und A_2. Eine naheliegende Vermutung ist, dass $|A_1 \cup A_2|$ einfach die Summe von $|A_1|$ und $|A_2|$ ist. Das stimmt aber nur, wenn A_1 und A_2 disjunkt sind.

Ist $A_1 = \{1, 2, 3\}$ und $A_2 = \{2, 3, 4\}$, so ist $|A_1 \cup A_2| = 4$, $|A_1| = 3$, $|A_2| = 3$ und damit $|A_1| + |A_2| = 6$. Das Problem ist, dass die Elemente des Durchschnitts $A_1 \cap A_2 = \{2, 3\}$ in der Rechnung $|A_1| + |A_2|$ doppelt gezählt werden. Um die korrekte Mächtigkeit von $A_1 \cup A_2$ zu berechnen, können wir $|A_1|$ und $|A_2|$ addieren und dann die Mächtigkeit $|A_1 \cap A_2|$ des Durchschnitts, der doppelt gezählt wurde, abziehen:

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| \tag{4.1}$$

In unserem Beispiel erhalten wir $|A_1 \cup A_2| = 4$ und

$$|A_1| + |A_2| - |A_1 \cap A_2| = 3 + 3 - 2 = 4.$$

Nun betrachten wir drei Mengen A_1, A_2 und A_3. Wie wir schon gesehen haben, gilt für zwei endliche Mengen B und C die Formel $|B \cup C| = |B| + |C| - |B \cap C|$. Setzt man $B := A_1 \cup A_2$ und $C = A_3$, so ergibt sich

$$|A_1 \cup A_2 \cup A_3| = |A_1 \cup A_2| + |A_3| - |(A_1 \cup A_2) \cap A_3|. \tag{4.2}$$

Nun ist $(A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3)$. Also gilt

$$|(A_1 \cup A_2) \cap A_3)| = |(A_1 \cap A_3) \cup (A_2 \cap A_3)| = |A_1 \cap A_3| + |A_2 \cap A_3| - |A_1 \cap A_2 \cap A_3|. \quad (4.3)$$

Einsetzen von (1) und (3) in (2) liefert

$$\begin{aligned} &|A_1 \cup A_2 \cup A_3| \\ &\quad = |A_1| + |A_2| - |A_1 \cap A_2| + |A_3| - (|A_1 \cap A_3| + |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|) \\ &\qquad = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|. \end{aligned}$$

An dieser Gleichung sehen wir schon das allgemeine Prinzip der Inklusion und Exklusion.

Satz 4.33 (Prinzip der Inklusion und Exklusion, Siebformel)**.** *Sei $n \in \mathbb{N}$ und seien $A_1, \ldots, A_n$ endliche Mengen. Dann gilt*

$$|A_1 \cup \cdots \cup A_n| = \sum_{k=1}^{n} \left((-1)^{k-1} \cdot \sum_{1 \leqslant n_1 < \cdots < n_k \leqslant n} |A_{n_1} \cap \cdots \cap A_{n_k}| \right).$$

Die innere Summe auf der rechten Seite der Gleichung läuft dabei über alle k-Tupel $(n_1, \ldots, n_k)$ natürlicher Zahlen mit $1 \leqslant n_1 < \cdots < n_k \leqslant n$.

Für den Beweis dieses Satzes benutzen wir folgendes Lemma:

Lemma 4.34. *Jede nichtleere endliche Menge M hat genauso viele Teilmengen mit gerader Mächtigkeit wie mit ungerader Mächtigkeit.*

Beweis. Sei n die Mächtigkeit von M. Wir nehmen zunächst an, dass n ungerade ist. Dann ist die Abbildung $a \longmapsto M \setminus a$ eine Bijektion zwischen der Menge der Teilmengen von M, die eine gerade Mächtigkeit haben, und der Menge der Teilmengen von M, deren Mächtigkeit ungerade ist. Also hat M genauso viele Teilmengen mit gerader Mächtigkeit wie mit ungerader Mächtigkeit.

Sei nun n gerade. Dann hat M genau

$$\binom{n}{1} + \cdots + \binom{n}{n-1} = \sum_{k=0}^{\frac{n}{2}-1} \binom{n}{2k+1}$$

Teilmengen mit ungerader Mächtigkeit. Nach Satz 4.15 gilt $\binom{n}{2k+1} = \binom{n-1}{2k} + \binom{n-1}{2k+1}$. Also ist

$$\sum_{k=0}^{\frac{n}{2}-1} \binom{n}{2k+1} = \sum_{k=0}^{\frac{n}{2}-1} \left(\binom{n-1}{2k} + \binom{n-1}{2k+1} \right) = \sum_{i=0}^{n-1} \binom{n-1}{i} = 2^{n-1}.$$

Da M insgesamt 2^n Teilmengen hat, hat genau die Hälfte aller Teilmengen eine gerade Mächtigkeit. □

Beweis von Satz 4.33. Sei $a \in A_1 \cup \cdots \cup A_n$. Auf der linken Seite der Gleichung wird a genau einmal gezählt. Wir zeigen, dass a auch auf der rechten Seite der Gleichung insgesamt genau 1 beiträgt. Sei $B := \{i\colon 1 \leqslant i \leqslant n \wedge a \in A_i\}$ und $\ell := |B|$. Die Zahl ℓ gibt also an, in wie vielen der Mengen A_i das Element a vorkommt.

Die Summanden auf der rechten Seite der Siebformel haben alle die Form $(-1)^{k-1} \cdot |A_{n_1} \cap \cdots \cap A_{n_k}|$, wobei k mindestens 1 ist und $1 \leqslant n_1 < \cdots < n_k \leqslant n$ gilt. Das Element a trägt nur dann etwas zu einem solchen Summanden bei, wenn $a \in A_{n_1} \cap \cdots \cap A_{n_k}$ gilt, wenn also $n_1, \ldots, n_k$ Elemente von B sind. Das heißt, a trägt genau dann zu einem Summanden $(-1)^{k-1} \cdot |A_{n_1} \cap \cdots \cap A_{n_k}|$ bei, wenn $\{n_1, \ldots, n_k\} \subseteq B$ gilt. Wir wissen für jedes $k \leqslant \ell$, dass B genau $\binom{\ell}{k}$ Teilmengen der Mächtigkeit k hat.

Damit kann man den Beitrag von a zu den Summanden auf der rechten Seite der Siebformel als

$$\sum_{k=1}^{\ell} (-1)^{k-1} \binom{\ell}{k}$$

schreiben. Nach Lemma 4.34 hat jede ℓ-elementige Menge genauso viele Teilmengen mit gerader Mächtigkeit wie mit ungerader Mächtigkeit. Es gilt also

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

Damit ist

$$\sum_{k=1}^{\ell} (-1)^{k-1} \binom{\ell}{k} = \binom{\ell}{0} = 1.$$

Damit ist der Beitrag von a zur rechten Seite der Siebformel ebenfalls genau 1.

Da dieses Argument für jedes $a \in A_1 \cup \cdots \cup A_n$ stimmt, sind die beiden Seiten der Siebformel tatsächlich gleich. $\square$

KAPITEL 5

Restklassenringe und das RSA-Verschlüsselungsverfahren

Dieses Kapitel ist eine Fortführung der Themen aus der Elementaren Zahlentheorie und es schließt direkt an Kapitel 3.6 an. Das Rechnen mit Kongruenzen führt uns zu den Restklassenringen, die wir im nächsten Abschnitt einführen. Als eine wichtige Anwendung werden wir dann dass RSA-Verschlüsselungsverfahren kennenlernen.

§5.1 Restklassenringe

Sei $m \in \mathbb{N}$. Wir erinnern uns an die Definition der Kongruenz modulo m. Zwei Zahlen $a, b \in \mathbb{Z}$ sind *kongruent modulo* m,

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

falls a und b bei Division durch m denselben Rest haben. Die Kongruenz $a \equiv b \pmod{m}$ gilt genau dann, wenn $a - b$ durch m teilbar ist.

Die folgenden drei Eigenschaften aus Satz 3.23 zeigen, dass die Kongruenz modulo m eine Äquivalenzrelation ist:

(1) $a \equiv a \pmod{m}$ (Reflexivität)
(2) $a \equiv b \pmod{m} \Rightarrow b \equiv a \pmod{m}$ (Symmetrie)
(3) $a \equiv b \pmod{m} \wedge b \equiv c \pmod{m} \Rightarrow a \equiv c \pmod{m}$ (Transitivität)

Die Äquivalenzklassen dieser Äquivalenzrelation haben wir *Restklassen* genannt und die Restklasse einer Zahl a mit $[a]_m$ bezeichnet. Es ist also

$$[a]_m = \{b \in \mathbb{Z} \colon a \equiv b \pmod{m}\} = \{\dots, a-m, a, a+m, a+2m, \dots\}.$$

Es gibt genau m verschiedene Restklassen modulo m, nämlich

$$[0]_m, [1]_m, \dots, [m-1]_m.$$

Definition 5.1. *Es sei*

$$\mathbb{Z}_m := \mathbb{Z}/m\mathbb{Z} := \{[0]_m, [1]_m, \dots, [m-1]_m\}$$

die Menge der Restklassen modulo m.

Für eine gegebene Restklasse K modulo m nennen wir ein Element $a \in K$ einen *Repräsentanten* oder *Vertreter* der Restklasse K. Ist a ein Repräsentant von K, so gilt $K = [a]_m$. Wählen wir aus jeder Restklasse genau einen Repräsentanten, so spricht man von einem *Repräsentanten-* oder *Vertretersystem*. Das *Standardrepräsentantensystem* für die Restklassen in $\mathbb{Z}/m\mathbb{Z}$ sind die Zahlen $0, 1, \dots, m-1$.

Wir definieren Rechenoperationen $\oplus$ und $\odot$ zwischen Restklassen modulo m.

Definition 5.2. *Für $a, b \in \mathbb{Z}$ sei*

$$[a]_m \oplus [b]_m := [a+b]_m$$

und

$$[a]_m \odot [b]_m := [a \cdot b]_m.$$

Man beachte, dass diese Definition nur dann sinnvoll ist, wenn die Definition unabhängig von der Wahl der Repräsentanten a und b der Restklassen $[a]_m$ und $[b]_m$ ist, wenn also für alle $c, d \in \mathbb{Z}$ mit $[a]_m = [c]_m$ und $[b]_m = [d]_m$ gilt:

$$[a+b]_m = [c+d]_m \text{ und } [a \cdot b]_m = [c \cdot d]_m$$

An dieser Stelle erinnern wir uns wieder an Satz 3.23. Es gilt:

(5) $a \equiv c \pmod{m} \wedge b \equiv d \pmod{m} \Rightarrow a + b \equiv c + d \pmod{m}$

Mit anderen Worten, wenn $[a]_m = [c]_m$ und $[b]_m = [d]_m$ gilt, dann gilt auch

$$[a+b]_m = [c+d]_m.$$

Das heißt, dass unsere Definition von $[a]_m \oplus [b]_m$ tatsächlich nur von den Restklassen $[a]_m$ und $[b]_m$ abhängt, und nicht von der Wahl der Repräsentanten a und b. Man sagt, dass $\oplus$ *wohldefiniert* ist.

Beispiel 5.3. Sei $m = 7$, $a = 5$ und $b = 8$. Dann ist

$$[a]_m \oplus [b]_m = [5]_7 \oplus [8]_7 = [5+8]_7 = [13]_7 = [6]_7.$$

Wählt man nun $c = -2$ und $d = 1$, so gilt $a - c = 7$ und $b - d = 7$. Es gilt also $a \equiv c \pmod{m}$ und $c \equiv d \pmod{m}$ und damit $[a]_m = [c]_m$ und $[b]_m = [d]_m$. Nun ist

$$[c]_m \oplus [d]_m = [-2]_7 \oplus [1]_7 = [-2+1]_7 = [-1]_7 = [6]_7.$$

Also ist $[a+b]_m = [c+d]_m$, wie erwartet.

Wir müssen noch zeigen, dass auch $\odot$ wohldefiniert ist. Seien $a, b, c, d \in \mathbb{Z}$ mit $a \equiv c \pmod{m}$ und $b \equiv d \pmod{m}$. Dann existieren $r_1, r_2, q_a, q_b, q_c, q_d \in \mathbb{Z}$ mit $a = q_a \cdot m + r_1$, $b = q_b \cdot m + r_2$, $c = q_c \cdot m + r_1$, $d = q_d \cdot m + r_2$ sowie $0 \leqslant r_1, r_2 < m$.

Wir betrachten $a \cdot b$ und $c \cdot d$. Es gilt

$$a \cdot b = (q_a \cdot m + r_1) \cdot (q_b \cdot m + r_2) = q_a \cdot q_b \cdot m^2 + r_1 \cdot q_b \cdot m + r_2 \cdot q_a \cdot m + r_1 \cdot r_2$$

und

$$c \cdot d = (q_c \cdot m + r_1) \cdot (q_d \cdot m + r_2) = q_c \cdot q_d \cdot m^2 + r_1 \cdot q_d \cdot m + r_2 \cdot q_c \cdot m + r_1 \cdot r_2.$$

Also ist $a \cdot b \equiv c \cdot d \pmod{m}$.

Das zeigt, dass $[a \cdot b]_m$ unabhängig von der Wahl der Repräsentanten a und b der Restklassen $[a]_m$ und $[b]_m$ ist. Damit ist auch $\odot$ wohldefiniert.

Satz 5.4. *Für alle $a, b, c \in \mathbb{Z}$ gilt:*

(1) Kommutativgesetz:
- $[a]_m \oplus [b]_m = [b]_m \oplus [a]_m$
- $[a]_m \odot [b]_m = [b]_m \odot [a]_m$

(2) Assoziativgesetz:
- $([a]_m \oplus [b]_m) \oplus [c]_m = [b]_m \oplus ([a]_m \oplus [c]_m)$
- $([a]_m \odot [b]_m) \odot [c]_m = [b]_m \odot ([a]_m \odot [c]_m)$

(3) Existenz neutraler Elemente:
- $[a]_m \oplus [0]_m = [a]_m$
- $[a]_m \odot [1]_m = [a]_m$

(4) Distributivgesetz:
- $[a]_m \odot ([b]_m \oplus [c]_m) = ([a]_m \odot [b]_m) \oplus ([a]_m \odot [c]_m)$

(5) Existenz additiver Inverser.
- $[a]_m \oplus [-a]_m = [0]_m$

Beweis. Alle diese Eigenschaften folgen leicht aus den entsprechenden Eigenschaften von $\mathbb{Z}$. Als Beispiel rechnen wir (4) nach. Es gilt

$$\begin{aligned}
[a]_m \odot ([b]_m \oplus [c]_m) &= [a]_m \odot [b+c]_m = [a \cdot (b+c)]_m \\
&= [a \cdot b + a \cdot c]_m = [a \cdot b]_m \oplus [a \cdot c]_m = ([a]_m \odot [b]_m) \oplus ([a]_m \odot [c]_m).
\end{aligned}$$

Das zeigt (4). □

Wir geben für $m = 2, 3, 4, 5$ Additionstabellen und Multiplikationstabellen an, wobei wir anstelle von $[r]_m$ zur Abkürzung r schreiben.

$m = 2$:

$\oplus$	0	1
0	0	1
1	1	0

$\odot$	0	1
0	0	0
1	0	1

$m = 3$:

$\oplus$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\odot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

$$m = 4: \quad \begin{array}{c|cccc} \oplus & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1 & 2 & 3 \\ 1 & 1 & 2 & 3 & 0 \\ 2 & 2 & 3 & 0 & 1 \\ 3 & 3 & 0 & 1 & 2 \end{array} \qquad \begin{array}{c|cccc} \odot & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 & 3 \\ 2 & 0 & 2 & 0 & 2 \\ 3 & 0 & 3 & 2 & 1 \end{array}$$

$$m = 5: \quad \begin{array}{c|ccccc} \oplus & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 2 & 3 & 4 & 0 \\ 2 & 2 & 3 & 4 & 0 & 1 \\ 3 & 3 & 4 & 0 & 1 & 2 \\ 4 & 4 & 0 & 1 & 2 & 3 \end{array} \qquad \begin{array}{c|ccccc} \odot & 0 & 1 & 2 & 3 & 4 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 2 & 3 & 4 \\ 2 & 0 & 2 & 4 & 1 & 3 \\ 3 & 0 & 3 & 1 & 4 & 2 \\ 4 & 0 & 4 & 3 & 2 & 1 \end{array}$$

Wir schreiben von nun an einfach $+$ und $\cdot$ für $\oplus$ und $\odot$ und stellen fest, dass sich nicht jede Rechenregel von $\mathbb{Z}$ auf $\mathbb{Z}/m\mathbb{Z}$ überträgt. Die Kürzungsregel, dass also für $a \neq 0$ aus $ab = ac$ immer $b = c$ folgt, gilt zum Beispiel im Allgemeinen nicht in $\mathbb{Z}/m\mathbb{Z}$. Zum Beispiel gilt $[2]_4 \cdot [1]_4 = [2]_4 = [6]_4 = [2]_4 \cdot [3]_4$ und $[2]_4 \neq [0]_4$, aber $[1]_4 \neq [3]_4$. Dieses Beispiel hängt damit zusammen, dass $[2]_4 \cdot [2]_4 = [4]_4 = [0]_4$ gilt, dass es also in $\mathbb{Z}/4\mathbb{Z}$ von 0 verschiedene Elemente gibt, deren Produkt 0 ist.

Definition 5.5. *Sei $[a]_m \in \mathbb{Z}/m\mathbb{Z}$. Ein Element $[b]_m \in \mathbb{Z}/m\mathbb{Z}$ heißt* multiplikatives Inverses *von $[a]_m$, falls*

$$[a]_m \cdot [b]_m = [1]_m$$

gilt. Besitzt $[a]_m$ ein multiplikatives Inverses, so nennt man $[a]_m$ invertierbar.

Beispiel 5.6. $[3]_4$ ist invertierbar. Es gilt nämlich $[3]_4 \cdot [3]_4 = [9]_4 = [1]_4$.

$[2]_4$ ist nicht invertierbar, da in $\mathbb{Z}/4\mathbb{Z}$ kein Element $[b]_4$ existiert, so dass $[2]_4 \cdot [b]_4 = [1]_4$ gilt. Das liest man an der entsprechenden Multiplikationstabelle ab.

$[2]_5$ ist invertierbar. Es gilt $[2]_5 \cdot [3]_5 = [6]_5 = [1]_5$.

Satz 5.7. *Ein Element von $\mathbb{Z}/m\mathbb{Z}$ hat höchstens ein multiplikatives Inverses.*

Beweis. Angenommen, $[b]_m$ und $[c]_m$ sind beide multiplikative Inverse von $[a]_m$. Dann gilt

$$[b]_m = [b]_m \cdot [1]_m = [b]_m \cdot ([a]_m \cdot [c]_m) = ([b]_m \cdot [a]_m) \cdot [c]_m = [1]_m \cdot [c]_m = [c]_m.$$

Also gibt es keine zwei verschiedenen multiplikativen Invsersen von $[a]_m$. $\square$

Satz 5.8. *Ein Element $[a]_m \in \mathbb{Z}/m\mathbb{Z}$ ist genau dann invertierbar, wenn a und m teilerfremd sind. Insbesondere ist jedes Element $[a]_p \in \mathbb{Z}/p\mathbb{Z} \setminus \{[0]_p\}$ invertierbar, wenn p eine Primzahl ist.*

Beweis. Sei zunächst $[a]_m \in \mathbb{Z}/m\mathbb{Z}$ invertierbar. Dann existiert $[b]_m \in \mathbb{Z}/m\mathbb{Z}$ mit

$$[a]_m \cdot [b]_m = [1]_m.$$

Es gilt also $ab \equiv 1 \pmod{m}$. Damit existiert ein $k \in \mathbb{Z}$ mit $ab - 1 = km$. Es folgt $ab - km = 1$. Ist $g \in \mathbb{Z}$ ein Teiler von a und m, so teilt g auch $ab - km = 1$. Damit ist g entweder 1 oder -1. Also sind a und m teilerfremd.

Nun nehmen wir an, dass a und m teilerfremd sind. Wir betrachten die Restklassen

$$[0 \cdot a]_m, [1 \cdot a]_m, \ldots, [(m-1) \cdot a]_m$$

und zeigen zunächst, dass sie paarweise verschieden sind.

Seien nämlich $r, s \in \mathbb{Z}$. Angenommen $[ra]_m = [sa]_m$. Dann ist $ra - sa = (r-s)a$ durch m teilbar. Da a und m teilerfremd sind, folgt daraus, dass $r - s$ durch m teilbar ist. Also gilt $[r]_m = [s]_m$. Es folgt, dass für $r, s \in Z$ mit $r \neq s$ und $0 \leqslant r, s < m$ die beiden Restklassen $[ra]_m$ und $[sa]_m$ verschieden sind.

Da die m Restklassen

$$[0 \cdot a]_m, [1 \cdot a]_m, \ldots, [(m-1) \cdot a]_m$$

paarweise verschieden sind, muss die Restklasse $[1]_m$ unter ihnen sein. Also gibt es ein $b \in \mathbb{Z}$ mit $0 \leqslant b < m$ und $[b \cdot a]_m = [1]_m$. Es gilt also $[b]_m \cdot [a]_m = [b \cdot a]_m = [1]_m$ und damit ist $[a]_m$ invertierbar. □

Aus den Sätzen 5.4 und 5.8 folgt sofort das nächste Korollar.

Korollar 5.9. *Ist p eine Primzahl, so ist $\mathbb{Z}/p\mathbb{Z}$ ein Körper.*

Der Beweis des nächsten Satzes zeigt, wie man multiplikative Inverse von invertierbaren Elementen von $\mathbb{Z}/m\mathbb{Z}$ berechnen kann.

Satz 5.10. *Seien $a, b \in \mathbb{N}$ und $d = \mathrm{ggT}(a, b)$. Dann gibt es $\lambda, \mu \in \mathbb{Z}$ mit $d = \lambda a + \mu b$.*

Beweis. Wir können annehmen, dass $a \leqslant b$ gilt und beweisen den Satz durch vollständige Induktion über die Anzahl der Schritte, die im euklidischen Algorithmus durchgeführt werden, um $\mathrm{ggT}(a, b)$ zu berechnen.

Induktionsanfang: Wenn der euklidische Algorithmus bereits nach dem ersten Schritt terminiert, so ist a ein Teiler von b. In diesem Falle ist $\mathrm{ggT}(a, b) = a$ und es gilt $a = 1 \cdot a + 0 \cdot b$.

Induktionsschritt: Sei $n \in \mathbb{N}$ so gewählt, dass der euklidische Algorithmus zur Berechnung von $\mathrm{ggT}(a, b)$ nach n Schritten terminiert und gelte $n > 1$. Angenommen der Satz gilt für alle $a', b' \in \mathbb{N}$, bei denen der euklidische Algorithmus nach weniger als n Schritten terminiert.

Wir führen den ersten Schritt des euklidischen Algorithmus für a und b durch und wählen $r, q \in \mathbb{Z}$ mit $b = q \cdot a + r$ und $0 \leqslant r < a$. Es gilt $d = \mathrm{ggT}(a, b) = \mathrm{ggT}(r, a)$. Nun lässt sich $\mathrm{ggT}(r, a)$ in weniger als n Schritten berechnen und nach Induktionsannahme existieren $\lambda', \mu' \in \mathbb{Z}$ mit $d = \lambda' r + \mu' a$. Es gilt $r = b - qa$ und damit

$$d = \lambda'(b - qa) + \mu' a = \lambda' b + (\mu' - \lambda' q)a.$$

Setzt man also $\mu := \lambda'$ und $\lambda := \mu' - \lambda' q$, so ergibt sich $d = \lambda a + \mu b$. □

Man beachte, dass für teilerfremde $a, m \in \mathbb{N}$ aus Satz 5.10 folgt, dass es $b, k \in \mathbb{Z}$ gibt, so dass $1 = ab + km$ gilt. Es folgt auf etwas andere Weise als im Satz 5.8, dass $[a]_m$ invertierbar ist, nämlich mit dem multiplikativen Inversen $[b]_m$. Man kann den euklidischen Algorithmus also auch einsetzen, um Elemente von $\mathbb{Z}/m\mathbb{Z}$ zu invertieren.

Beispiel 5.11. a) Sei $a = 228$ und $b = 294$. Wir berechnen den größten gemeinsamen Teiler von a und b mit dem euklidischen Algorithmus. Es gilt:

$$\begin{aligned} 294 &= 1 \cdot 228 + 66 \\ 228 &= 3 \cdot 66 + 30 \\ 66 &= 2 \cdot 30 + 6 \\ 30 &= 5 \cdot 6 + 0 \end{aligned}$$

Der größte gemeinsame Teiler von 228 und 66 ist also 6. Aus der vorletzten Gleichung erhalten wir $6 = 66 - 2 \cdot 30$. Aus der zweiten Gleichung ergibt sich $30 = 228 - 3 \cdot 66$. Einsetzen liefert $6 = 66 - 2 \cdot (228 - 3 \cdot 66) = 7 \cdot 66 - 2 \cdot 228$. Die erste Gleichung liefert $66 = 294 - 1 \cdot 228$. Durch Einsetzen in $6 = 7 \cdot 66 - 2 \cdot 228$ folgt

$$6 = 7 \cdot (294 - 1 \cdot 228) - 2 \cdot 228 = 7 \cdot 294 - 9 \cdot 228.$$

b) Sei $a = 15$ und $m = 28$. Wir wollen $[a]_m$ invertieren. Der euklidische Algorithmus liefert

$$\begin{aligned} 28 &= 1 \cdot 15 + 13 \\ 15 &= 1 \cdot 13 + 2 \\ 13 &= 6 \cdot 2 + 1 \\ 2 &= 2 \cdot 1 + 0. \end{aligned}$$

Der größte gemeinsame Teiler von 15 und 28 ist also 1. Auflösen der Gleichung in diesem Durchlauf des euklidischen Algorithmus und Rückwärtseinsetzen liefert

$$\begin{aligned} 1 = 13 - 6 \cdot 2 = 13 - 6 \cdot (15 - 1 \cdot 13) &= 7 \cdot 13 - 6 \cdot 15 \\ &= 7 \cdot (28 - 1 \cdot 15) - 6 \cdot 15 = 7 \cdot 28 - 13 \cdot 15 \end{aligned}$$

Es gilt also

$$1 \equiv -13 \cdot 15 \pmod{28}.$$

Damit ist $[-13]_{28} = [15]_{28}$ das multiplikative Inverse von $[15]_{28}$ in $\mathbb{Z}/28\mathbb{Z}$.

Auf ähnliche Weise wie Satz 5.8 können wir auch den folgenden Satz beweisen, der wichtige Anwendungen in der Kryptographie hat.

Definition 5.12 (Eulersche φ-Funktion)**.** *Für $n \in \mathbb{N}$ sei $\varphi(n)$ die Anzahl der zu n teilerfremden natürlichen Zahlen $\leqslant n$.*

Beispiel 5.13. (*a*) Es gilt $\varphi(1) = 1$, da $\operatorname{ggT}(1,1) = 1$ gilt und damit 1 und 1 teilerfremd sind.

(*b*) Für eine Primzahl p ist $\varphi(p) = p - 1$, da alle kleineren natürlichen Zahlen zu p teilerfremd sind.

(*c*) Die Zahlen $1, 5, 7, 11$ sind zu 12 teilerfremd, während $2, 3, 4, 6, 8, 9, 10$ nichttriviale gemeinsame Teiler mit 12 haben. Also ist $\varphi(12) = 4$.

(*d*) Sind p und q verschiedene Primzahlen, so gilt

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

Eine Zahl $a \leqslant p \cdot q$ hat nämlich genau dann einen nichttrivialen gemeinsamen Teiler mit $p \cdot q$, wenn a ein Vielfaches von p oder q ist. Das kleinste gemeinsame Vielfache von p und q ist $p \cdot q$. Es gibt also p Vielfache von q und q Vielfache von p, die nicht größer als $p \cdot q$ sind. Dabei wird das gemeinsame Vielfache $p \cdot q$ doppelt gezählt. Insgesamt gibt es also $p + q - 1$ natürliche Zahlen $\leqslant p \cdot q$, die nicht zu $p \cdot q$ teilerfremd sind. Das zeigt $\varphi(p \cdot q) = (p-1) \cdot (q-1)$.

Satz 5.14 (Der Satz von Fermat-Euler)**.** *Seien $m, n \in \mathbb{N}$ teilerfremd. Dann gilt*

$$n^{\varphi(m)} \equiv 1 \pmod{m}.$$

Beweis. Seien $r_1, \ldots, r_{\varphi(m)}$ die natürlichen Zahlen $\leqslant m$, die zu m teilerfremd sind. Wie im Beweis von Satz 5.8 sind die Restklassen

$$[r_1 \cdot n]_m, [r_2 \cdot n]_m, \ldots, [r_{\varphi(m)} \cdot n]_m$$

paarweise verschieden. Für jedes $i \in \{1, \ldots, \varphi(m)\}$ sind r_i und n beide zu m teilerfremd. Es folgt, dass auch $r_i \cdot n$ zu m teilerfremd ist. Also gilt

$$\{[r_1 \cdot n]_m, [r_2 \cdot n]_m, \ldots, [r_{\varphi(m)} \cdot n]_m\} = \{[r_1]_m, [r_2]_m, \ldots, [r_{\varphi(m)}]_m\}$$

und damit auch

$$[r_1 \cdot n]_m \cdot [r_2 \cdot n]_m \cdot \ldots \cdot [r_{\varphi(m)} \cdot n]_m = [r_1]_m \cdot [r_2]_m \cdot \ldots \cdot [r_{\varphi(m)}]_m.$$

Daher gilt für $v = r_1 \cdot r_2 \cdot \ldots \cdot r_{\varphi(m)}$ die Kongruenz

$$v \equiv (r_1 \cdot n) \cdot (r_2 \cdot n) \cdot \ldots \cdot (r_{\varphi(m)} \cdot n) \equiv v \cdot n^{\varphi(m)} \pmod{m}.$$

Da v ein Produkt von zu m teilerfremden Zahlen ist, ist auch v selbst zu m teilerfremd. Also ist $[v]_m$ nach Satz 5.8 invertierbar und es existiert $[b]_m \in \mathbb{Z}/m\mathbb{Z}$ mit $[b]_m \cdot [v]_m = [1]_m$. Multiplikation der Gleichung $[v]_m = [v \cdot n^{\varphi(m)}]_m$ mit $[b]_m$ liefert $[1]_m = [n^{\varphi(m)}]_m$, also $n^{\varphi(m)} \equiv 1 \pmod{m}$. $\square$

Korollar 5.15 (Der kleine Satz von Fermat)**.** *Sei $n \in \mathbb{N}$ und p eine Primzahl, die n nicht teilt. Dann gilt*

$$n^{p-1} \equiv 1 \pmod{p}.$$

§5.2 Das RSA-Verschlüsselungsverfahren

Die *RSA-Verschlüsselung* wurde 1977 von den Mathematikern Rivest, Shamir und Adleman entwickelt und ist immer noch wichtiger Bestandteil heute gängiger Verschlüsselungsmethoden. Dabei wird ein Nachrichtentext vom Sender zunächst auf irgendeine sinnvolle Weise als natürliche Zahl m kodiert, so dass sich die Nachricht vom Empfänger aus m leicht wieder dekodieren lässt. Uns interessiert nur, wie wir nun die Zahl m verschlüsseln und an den Empfänger versenden können, ohne dass Dritte die Nachricht entschlüsseln können.

Es gibt beim RSA-Verfahren zwei Schlüssel, einen *öffentlichen Schlüssel* (*public key*) und einen *privaten Schlüssel* (*private key*). Die beiden Schlüssel werden vom Empfänger der Nachricht erzeugt. Nur der öffentliche Schlüssel wird an den Sender weitergeleitet. Der private Schlüssel ist nur dem Empfänger bekannt. Es ist dabei unwichtig, ob der öffentliche Schlüssel Dritten bekannt wird.

Der öffentliche Schlüssel ist ein Zahlenpaar (e, N) und der private Schlüssel ein Zahlenpaar (d, N), wobei N in beiden Fällen dieselbe Zahl ist. Man nennt N den *RSA-Modul*, e den *Verschlüsselungsexponenten* und d den *Entschlüsselungsexponenten.* Die Schlüssel werde wie folgt erzeugt:

(1) Wähle zufällig zwei verschiedene Primzahlen p und q.
(2) Berechne den RSA-Modul $N = p \cdot q$.
(3) Berechne $\varphi(N) = (p-1) \cdot (q-1)$.
(4) Wähle eine zu $\varphi(N)$ teilerfremde Zahl e mit $1 < e < \varphi(N)$.
(5) Berechne das multiplikative Inverse $[d]_{\varphi(N)}$ von $[e]_{\varphi(N)}$.

Die Zahlen p, q und $\varphi(N)$ werden nun nicht mehr benötigt und können gelöscht werden. Die Zahl m, die verschlüsselt werden soll, muss kleiner als das RSA-Modul N sein.

Verschlüsselt wird nun wie folgt: Der Sender benutzt den öffentlichen Schlüssel (e, N) und berechnet $[m^e]_N$. Die Restklasse $[m^e]_N$ wird dann in Form eines Repräsentanten zwischen 0 und N angegeben und an den Empfänger übermittelt. Ohne Kenntnis des privaten Schlüssels (d, N) lässt sich m nicht in sinnvoller Zeit aus $[m^e]_N$ rekonstruieren, obwohl man ja eigentlich nur in $\mathbb{Z}/N\mathbb{Z}$ die e-te Wurzel aus $[m^e]_N$ ziehen muss. Aber das geht eben nicht innerhalb eines sinnvollen Zeitrahmens.

Der Empfänger benutzt den privaten Schlüssel (d, N) und berechnet $[(m^e)^d]_N$. Das geht wiederum schnell, da Potenzieren auch in $\mathbb{Z}/N\mathbb{Z}$ einfach ist. Wegen

$$e \cdot d \equiv 1 \pmod{\varphi(N)}$$

existiert ein $r \in \mathbb{Z}$ mit $e \cdot d = r \cdot \varphi(N) + 1$. Wenn wir annehmen, dass die Nachricht m teilerfremd von N ist, dann gilt nach Satz 5.14

$$(m^e)^d \equiv m^{e \cdot d} \equiv m^{r \cdot \varphi(N)+1} \equiv (m^{\varphi(N)})^r \cdot m \equiv 1^r \cdot m \equiv m \pmod N$$

und damit $[(m^e)^d]_N = [m]_N$. Damit ist die Nachricht entschlüsselt. Für den allgemeinen Fall argumentiert man wie folgt. Zuerst zeigt man

$$m^{ed} \equiv m \pmod p \quad \text{und} \quad m^{ed} \equiv m \pmod q \tag{5.1}$$

und in einem zweiten Schritt (siehe Übung) schließt man aus (5.1), dass mit $N = pq$ dann auch

$$m^{ed} \equiv m \pmod N$$

gilt. Wir zeigen die erste Kongruenz in (5.1) (die zweite folgt analog).

Falls m und p teilerfremd sind, dann folgt aus Korollar 5.15

$$m^{p-1} \equiv 1 \pmod p\,. \tag{5.2}$$

Wegen $e \cdot d = r \cdot \varphi(N) + 1$ und $\varphi(N) = \varphi(pq) = (p-1)(q-1)$ ist $ed = r(p-1)(q-1)+1$ und somit folgt (5.1) durch

$$m^{ed} = m^{r(p-1)(q-1)+1} = \left(m^{p-1}\right)^{r(q-1)} \cdot m \overset{(5.2)}{\equiv} 1^{r(q-1)} \cdot m \equiv m \pmod p\,.$$

Falls m und p nicht teilerfremd sind, dann ist m ein Vielfaches von p und somit gilt $m \equiv 0 \pmod p$ und trivialerweise auch

$$m^{ed} \equiv 0 \equiv m \pmod p\,.$$

In beiden Fällen gilt also die erste Kongruenz aus (5.1).

In der Praxis werden noch diverse weitere Forderungen an p, q und e gestellt, damit das Verfahren effizient und sicher durchgeführt werden kann. Man beachte, dass man den privaten Schlüssel (d, N) aus (e, N) berechnen kann, indem man N in seine Primfaktoren p und q zerlegt. Das dauert aber zu lange, wenn p und q ausreichend groß sind. Im September 2009 wurde eine 232-stellige Zahl (768 Bits) mit einem Rechenaufwand von mehreren Jahren auf hunderten von Rechnern in ihre Primfaktoren zerlegt. Eine gängige Größe für das RSA-Modul ist 2048 Bit, also etwa 600 Dezimalstellen.

Beispiel 5.16. Wir wählen die zwei Primzahlen $p = 11$ und $q = 13$. Das liefert den RSA-Modul $N = 143$. Es gilt $\varphi(N) = (p-1) \cdot (q-1) = 10 \cdot 12 = 120$. Die Zahl $e = 23$ ist zu 120 teilerfremd. Wir wählen $(23, 143)$ als den öffentlichen Schlüssel. Mit dem euklidischen Algorithmus bestimmen wir das multiplikative Inverse von $[23]_{120}$ in $\mathbb{Z}/120\mathbb{Z}$. Es gilt $\mathrm{ggT}(23, 120) = 1 = 23 \cdot 47 - 9 \cdot 120$. Damit ist $23 \cdot 47 \equiv 1 \pmod{120}$ und wir setzen $d = 47$. Der private Schlüssel ist also $(47, 143)$.

Angenommen, die Zahl 7 soll verschlüsselt werden. Es gilt

$$7^{23} \mod 143 = 27368747340080916343 \mod 143 = 2.$$

Die verschlüsselte Nachricht ist also 2.

Zum Entschlüsseln müssen wir mit $d = 47$ potenzieren. Es gilt

$$2^{47} \mod 143 = 140737488355328 \mod 143 = 7.$$

KAPITEL 6

Algebraische Strukturen

§6.1 Einfache Strukturen

Definition 6.1. *Eine algebraische Struktur ist eine Menge M zusammen mit endlich vielen endlichstelligen Operationen $f_1, \ldots, f_k$ auf M. Dabei sind ausdrücklich auch 0-stellige Funktionen erlaubt, die wir als Konstanten interpretieren. Formal schreibt man für die algebraische Struktur $\mathcal{M} = (M, f_1, \ldots, f_k)$. Dabei heißt M die $\mathcal{M}$* unterliegende Menge. *Oft wird jedoch nicht zwischen einer algebraischen Struktur und ihrer unterliegenden Menge unterschieden. So bezeichnet $\mathbb{R}$ sowohl die Menge der reellen Zahlen als auch die algebraische Struktur* $(\mathbb{R}, +, \cdot, 0, 1)$.

Beispiel 6.2. Wir haben schon einiger Beispiele algebraischer Strukturen kennengelernt.

(*a*) Ein Körper ist eine Menge K zusammen mit zwei zweistelligen Operationen $+$ und $\cdot$ sowie Konstanten 0 und 1, sodass die Axiome (K1)–(K5) erfüllt sind. Damit sind Körper algebraische Strukturen. Das gilt insbesondere für $(\mathbb{Q}, +, \cdot, 0, 1)$ und $(\mathbb{R}, +, \cdot, 0, 1)$.

(*b*) $(\mathbb{Z}, +, \cdot)$ und $(\mathbb{N}, +, \cdot)$ sind ebenfalls algebraische Strukturen.

(*c*) Boolesche Algebren sind algebraische Strukturen mit zwei zweistelligen Operationen $\sqcup$ und $\sqcap$ sowie einer einstelligen Operation $\neg$ und zwei Konstanten 0 und 1.

(*d*) Für eine Menge A sei $F(A)$ die Menge der Funktionen von A nach A. Dann ist $(F(A), \circ)$ eine algebraische Struktur. Ist $\mathcal{S}(A)$ die Menge der Bijektionen von A nach A, so ist $(\mathcal{S}(A), \circ)$ eine algebraische Struktur. Man beachte, dass die Komposition $\circ$ von Abbildungen tatsächlich eine zweistellige Operation auf $\mathcal{S}(A)$ ist, da die Komposition zweier Bijektionen wieder eine Bijektion ist.

Definition 6.3. *Sei $(M, *)$ eine algebraische Struktur mit einem zweistelligen Operator $*$. Ein Element $e \in M$ wird* neutrales Element *(bezüglich $*$) genannt, falls für alle $a \in M$ gilt:*

$$e * a = a * e = a$$

Beispiel 6.4. (*a*) Die 0 ist ein neutrales Element bezüglich $+$ in $\mathbb{R}$, $\mathbb{Q}$ und $\mathbb{Z}$. In denselben Strukturen ist 1 ein neutrales Element bezüglich $\cdot$.

(*b*) In einer Booleschen Algebra ist 1 neutral bezüglich $\sqcap$ und 0 ist neutral bezüglich $\sqcup$.

(*c*) In $F(A)$ und $\mathcal{S}(A)$ ist die identische Abbildung

$$\mathrm{id}_A\colon A \longrightarrow A; x \longmapsto x$$

ein neutrales Element bezüglich $\circ$.

(*d*) Es gibt nicht in jeder algebraischen Struktur mit einer zweistelligen Operation ein neutrales Element. Ein Beispiel ist $(\mathbb{N}, +)$.

Lemma 6.5. *Ist $*$ eine zweistellige Operation auf M, so gibt es höchstens ein neutrales Element bezüglich $*$.*

Beweis. Seien c und d neutrale Elemente bezüglich $*$. Dann gilt $c = c * d = d$. □

Definition 6.6. *Sei $*$ eine zweistellige Operation auf M mit einem neutralen Element e. Für $a \in M$ heißt $b \in M$* invers *zu a (bezüglich $*$), falls $a * b = b * a = e$ gilt. Falls für $a \in M$ ein $b \in M$ existiert, das zu a invers ist, so heißt a* invertierbar.

Beispiel 6.7. (*a*) Für jedes a in $\mathbb{Z}$, $\mathbb{Q}$ oder $\mathbb{R}$ ist $-a$ das zu a inverse Element bezüglich $+$.

(*b*) Für jedes a in $\mathbb{Q}$ oder $\mathbb{R}$ mit $a \neq 0$ ist a^{-1} das zu a inverse Element bezüglich $\cdot$.

(*c*) In $(\mathbb{N}, \cdot)$ ist 1 das neutrale Element und auch invertierbar mit dem Inversen 1, jedoch ist keine natürliche Zahl > 1 invertierbar.

(*d*) In keinem Körper besitzt 0 ein Inverses bezüglich der Multiplikation.

(*e*) Wie wir bereits gesehen haben, hat das Element $[2]_4$ in $\mathbb{Z}/4\mathbb{Z}$ kein Inverses bezüglich der Multiplikation. Andererseits ist $[3]_4$ in $\mathbb{Z}/4\mathbb{Z}$ invertierbar bezüglich $\cdot$ und zu sich selbst invers.

(*f*) Bezüglich $+$ sind alle Elemente $[a]_m$ von $\mathbb{Z}/m\mathbb{Z}$ invertierbar, wobei $[-a]_m$ zu $[a]_m$ invers ist.

Definition 6.8. *Es sei $(M, *)$ eine algebraische Struktur mit einer zweistelligen Verknüpfung $*$. Gilt für alle $a, b, c \in M$ das Assoziativgesetz*

$$a * (b * c) = (a * b) * c,$$

so ist $(M, *)$ eine Halbgruppe.

Hat $(M, *)$ außerdem ein neutrales Element, so nennt man $(M, *)$ ein Monoid.

Beispiel 6.9. (*a*) Die Strukturen $(\mathbb{N}, \cdot)$, $(\mathbb{R}, +)$, $(\mathbb{R}, \cdot)$ und $(F(A), \circ)$ sind Monoide. $(\mathbb{N}, +)$ ist jedoch kein Monoid, da es in $\mathbb{N}$ bezüglich $+$ kein neutrales Element gibt.

(*b*) Für eine Menge A, die wir in diesem Zusammenhang *Alphabet* nennen. sei A^* die Menge aller endlichen Folgen von Zeichen aus A. Die Elemente von A^* nennen wir *Wörter* über A. Für zwei Wörter $v = a_1 \dots a_n$ und $w = b_1 \dots b_m$ definieren wir die *Verkettung* $v^\frown w$ von v und w als das Wort $a_1 \dots a_n b_1 \dots b_m$. Dann ist $(A^*, {}^\frown)$ ein Monoid. Dabei ist das leere Wort das neutrale Element.

(*c*) Ist $(K, +, \cdot)$ ein Körper, so ist sowohl $(K \setminus \{0\}, \cdot)$ als auch $(K, \cdot)$ ein Monoid.

(*d*) Für $m \geqslant 2$ ist $(\mathbb{Z}/m\mathbb{Z}, \cdot)$ ein Monoid. Nach (*c*) ist $(\mathbb{Z}/m\mathbb{Z} \setminus \{[0]_m\}, \cdot)$ ein Monoid, falls m eine Primzahl ist. Ist m keine Primzahl, so ist $(\mathbb{Z}/m\mathbb{Z}\setminus\{[0]_m\}, \cdot)$ nicht einmal eine algebraische Struktur. Seien nämlich $k, \ell \in \mathbb{N}$ mit $m = k \cdot \ell$ und $k, \ell \neq 1$. Dann gilt $[k]_m \cdot [\ell]_m = [k \cdot \ell]_m = [m]_m = [0]_m$. Damit sind $[k]_m$ und $[\ell]_m$ in $\mathbb{Z}/m\mathbb{Z}\setminus\{[0]_m\}$, während $[k]_m \cdot [\ell]_m$ kein Element von $\mathbb{Z}/m\mathbb{Z}\setminus\{[0]_m\}$ ist. In diesem Falle ist $\cdot$ also gar keine Operation auf $\mathbb{Z}/m\mathbb{Z} \setminus \{[0]_m\}$.

Satz 6.10. *Ist $(M, *)$ ein Monoid, so besitzt jedes Element a von M höchstens ein Inverses.*

Beweis. Der Beweis ist eine allgemeine Fassung des Beweises von Satz 5.7. Seien $b, c \in M$ Inverse von $a \in M$. Dann gilt $b = b * e = b * (a * c) = (b * a) * c = e * c = c$. □

§6.2 Gruppentheorie

Definition 6.11. *Eine* Gruppe *ist ein Monoid, in dem jedes Element invertierbar ist. Der Übersichtlichkeit halber geben wir die Axiome für Gruppen noch einmal gesammelt an.*

Sei $(G, *)$ eine algebraische Struktur mit einer zweistelligen Verknüpfung $*$. Dann heißt $(G, *)$ eine Gruppe, *falls gilt:*

(G1) Für alle $a, b, c \in G$ gilt: $a * (b * c) = (a * b) * c$ (Assoziativgesetz)

(G2) Es gibt ein Element $e \in G$, sodass für alle $a \in G$ gilt: $a * e = e * a = a$ (Existenz eines neutralen Elements)

(G3) Für alle $a \in G$ existiert ein $b \in G$, sodass für das eindeutig bestimmte neutrale Element $e \in G$ gilt: $a * b = b * a = e$ (Existenz inverser Elemente)

Nachdem wir die entsprechenden Tatsachen für Monoide bewiesen haben, wissen wir, dass das neutrale Element einer Gruppe eindeutig bestimmt ist. Ebenso ist für jedes Element einer Gruppe das Inverse eindeutig bestimmt.

Beispiel 6.12.

(*a*) Wir haben schon zahlreiche Beispiele für Gruppen gesehen. So sind $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ und $(\mathbb{R}, +)$ Gruppen. Ebenso ist für jedes $m \geqslant 2$ die Struktur $(\mathbb{Z}/m\mathbb{Z}, +)$ eine Gruppe.

(*b*) Auch $(\mathbb{Q} \setminus \{0\}, \cdot)$ und $(\mathbb{R} \setminus \{0\}, \cdot)$ sind Gruppen. Ist m eine Primzahl, so ist $(\mathbb{Z}/m\mathbb{Z} \setminus \{[0]_m\}, \cdot)$ eine Gruppe.

(*c*) Sei A eine Menge und sei $\mathcal{S}(A)$ wieder die Menge der Bijektionen von A nach A. Dann ist $(\mathcal{S}(A), \circ)$ eine Gruppe. Für jede Funktion $f \in \mathcal{S}(A)$ ist die Umkehrfunktion f^{-1} das zu f inverse Element. Die Gruppe $(\mathcal{S}(A), \circ)$ heißt die *symmetrische Gruppe* auf A. Besonders wichtig sind die Gruppen $\mathcal{S}_n = (\mathcal{S}(\{1, \ldots, n\}), \circ)$ für $n \in \mathbb{N}$. Im Gegensatz zu den Gruppen, die wir bisher diskutiert haben, erfüllt $(\mathcal{S}(A), \circ)$ nicht das Kommutativgesetz, falls A mindestens drei Elemente hat.

Seien nämlich $a, b, c \in A$ verschieden und seien $f, g\colon A \longrightarrow A$ Permutationen, die alle $x \in A \setminus \{a, b, c\}$ wieder auf x abbilden. Weiter sei $f(a) = b$, $f(b) = a$, $f(c) = c$, $g(a) = b$, $g(b) = c$ und $g(c) = a$. Dann gilt

$$(f \circ g)(a) = f(g(a)) = f(b) = a$$

und

$$(g \circ f)(a) = g(f(a)) = g(b) = c\,.$$

Also ist $f \circ g \neq g \circ f$.

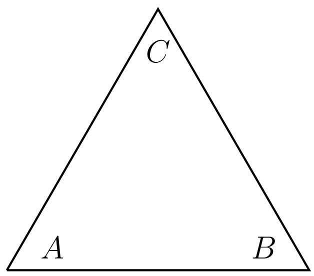
(d) Sei $m \geqslant 2$ und $(\mathbb{Z}/m\mathbb{Z})^\times = \{[a]_m\colon a \text{ und } m \text{ sind teilerfremd}\}$. $(\mathbb{Z}/m\mathbb{Z})^\times$ ist also genau die Menge der invertierbaren Elemente von $\mathbb{Z}/m\mathbb{Z}$. Dann ist $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot)$ eine Gruppe, die *Einheitengruppe* von $\mathbb{Z}/m\mathbb{Z}$. Die Elemente von $(\mathbb{Z}/m\mathbb{Z})^\times$ nennt man *Einheiten* von $\mathbb{Z}/m\mathbb{Z}$. Anstelle von $(\mathbb{Z}/m\mathbb{Z})^\times$ schreibt man auch $(\mathbb{Z}/m\mathbb{Z})^*$.

Dass die Einheiten eine Gruppe bilden, sieht man wie folgt: Zunächst müssen wir zeigen, dass $\cdot$ überhaupt eine Operation auf $(\mathbb{Z}/m\mathbb{Z})^\times$ ist, d.h., dass das Produkt zweier Einheiten wieder eine Einheit ist.

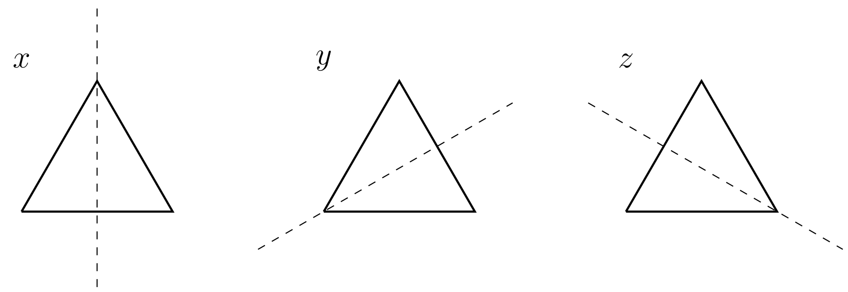
Seien $a, b \in \mathbb{Z}$ teilerfremd zu m. Dann gibt es $c, d \in \mathbb{Z}$, sodass $[c]_m$ und $[d]_m$ zu $[a]_m$ und $[b]_m$ invers sind. Damit ist aber $[c]_m \cdot [d]_m$ zu $[a]_m \cdot [b]_m$ invers. Also ist $[a]_m \cdot [b]_m \in (\mathbb{Z}/m\mathbb{Z})^\times$.

Dass $\cdot$ das Assoziativgesetz erfüllt, wissen wir schon. $[1]_m$ ist das neutrale Element von $(\mathbb{Z}/m\mathbb{Z})^\times$. Auch wissen wir, das alle Elemente von $(\mathbb{Z}/m\mathbb{Z})^\times$ in $\mathbb{Z}/m\mathbb{Z}$ invertierbar sind. Wir müssen noch zeigen, dass das Inverse einer Einheit auch wieder eine Einheit ist. Das ist aber klar: Ist $[b]_m$ zu $[a]_m$ invers, so ist $[a]_m$ zu $[b]_m$ zu invers. Also ist für jedes Element von $(\mathbb{Z}/m\mathbb{Z})^\times$ auch sein Inverses eine Einheit.

(e) Wir betrachten nun noch ein geometrisches Beispiel, die Gruppe $G_\triangle$ der Symmetrien eines gleichseitigen Dreiecks, also der Transformationen der Ebene, die das Dreieck auf das Dreieck abbilden. Die zweistellige Operation auf der Menge dieser Symmetrien ist die Komposition von Abbildungen. Diese Gruppe nennen wir kurz die Dreiecksgruppe.



Diese Transformationen sind zunächst die Identität, die jeden Punkt der Ebene wieder auf sich selbst abbildet. Die Identität bezeichnen wir mit i. Weiter sei r die Drehung um $120°$ entgegen dem Uhrzeigersinn, also im mathematisch positiven Drehsinn. Es sei s die Drehung um $240°$ entgegen dem Uhrzeigersinn. Schließlich seien x, y und z die Spiegelungen entlang der in der Zeichnung angegebenen Achsen.



Diese Symmetrien sind jeweils eindeutig dadurch bestimmt, auf welche Ecken die Ecken des Dreiecks abgebildet werden. Damit entspricht jede Symmetrie einer Permutation der Menge $\{A, B, C\}$.

Wir listen die Entsprechungen auf.

i	r	s
$\begin{pmatrix} A & B & C \\ A & B & C \end{pmatrix}$	$\begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}$	$\begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}$

x	y	z
$\begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix}$	$\begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix}$	$\begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix}$

Wir wissen, dass die Komposition von Abbildungen das Assoziativgesetz erfüllt. Auch wissen wir, dass die Identität ein neutrales Element bezüglich der Komposition ist. Um zu zeigen, dass die Menge $G = \{i, r, s, x, y, z\}$ mit der Komposition von Abbildungen tatsächlich eine Gruppe ist, müssen wir noch zeigen, dass die Komposition je zwei der Abbildungen in G wieder in G ist und dass jede Abbildung in G eine Umkehrfunktion in G hat. Dazu berechnen wir alle Kompositionen von Elementen von G und stellen das Ergebnis in einer Multiplikationstabelle dar. Multiplikationstabellen werden in diesem Zusammenhang auch *Gruppentafeln* genannt. In der Zeile rechts neben dem Element a und der Spalte unter dem Element b steht das Produkt $a \circ b$.

$\circ$	i	r	s	x	y	z
i	i	r	s	x	y	z
r	r	s	i	z	x	y
s	s	i	r	y	z	x
x	x	y	z	i	r	s
y	y	z	x	s	i	r
z	z	x	y	r	s	i

Dieser Gruppentafel entnehmen wir, dass für je zwei Elemente $a, b \in G$ die Komposition $a \circ b$ wieder in G liegt und dass jedes Element von G invertierbar ist. So sind i, x, y und z zu sich selbst invers, während r zu s invers ist.

Wir stellen fest, dass in der Gruppentafel in Beispiel 6.12 (e) in jeder Zeile und Spalte jedes Element genau einmal auftaucht. Das folgende Lemma zeigt, dass das kein Zufall ist. Im folgenden schreiben wir für $a * b$ kurz ab. Außerdem schreiben wir e für das neutrale Element einer Gruppe und a^{-1} für das Inverse eines Elements a.

Lemma 6.13. *Sei G eine Gruppe.*

(i) *Seien $a, b, c \in G$. Gilt $ab = ac$, so ist $b = c$. Genauso folgt aus $ba = ca$, dass $b = c$ gilt.*

(ii) *Die Gleichungen $ax = b$ und $xa = b$, wobei x eine Unbekannte ist, sind eindeutig lösbar.*

Beweis. (i) Es gelte $ab = ac$. Wir multiplizieren diese Gleichung von links mit a^{-1} und erhalten $a^{-1}ab = a^{-1}ac$, also $eb = ec$ und damit $b = c$, wie behauptet. Man beachte, dass wir aufpassen müssen, von welcher Seite wir mit a^{-1} multiplizieren, da in G nicht unbedingt das Kommutativgesetz gilt. Es könnte also sein, dass $b = a^{-1}ab$ und aba^{-1} verschieden sind.

Falls $ba = ca$ gilt, so multiplizieren wir diese Gleichung von rechts mit a^{-1} und erhalten $b = c$.

(ii) Ist die Gleichung $ax = b$ gegeben, so multiplizieren wir wieder von links mit a^{-1}. Das liefert $x = a^{-1}b$. Die Gleichung wird also von dem Gruppenelement $a^{-1}b$ gelöst. Mit Hilfe einer Multiplikation von rechts sehen wir, dass $xa = b$ die Lösung $x = ba^{-1}$ hat. $\square$

Teil (i) dieses Lemmas zeigt, dass in einer Gruppentafel in jeder Zeile und Spalte jedes Element höchstens einmal auftritt. Teil (ii) zeigt, dass in jeder Zeile und in jeder Spalte einer Gruppentafel jedes Element mindestens einmal auftritt.

Beispiel 6.14. Wir betrachten wieder die Dreiecksgruppe $G_\triangle$. Wir benutzen X als Unbekannte, um die Unbekannte von dem Gruppenelement x zu unterscheiden. Angenommen, wir wollen die Gleichung $Xs = y$ lösen. Multiplikation von rechts mit s^{-1}

liefert $X = ys^{-1}$. In der Gruppentafel von $G_\triangle$ lesen wir ab, dass $s^{-1} = r$ gilt und dass $yr = z$ ist. Damit löst $X = z$ die Gleichung $Xs = y$.

6.2.1. Die Ordnung eines Gruppenelements. Gegeben sei eine Gruppe $(G, *)$. Dann definiert man die Potenzen a^n eines Gruppenelements a wie folgt: Es sei $a^0 := e$. Für $n \in \mathbb{N}_0$ sei $a^{n+1} := a^n * a$. Potenzen mit negativen Exponenten definiert man durch $a^{-n} := (a^{-1})^n$

Wie für Potenzen reeller Zahlen rechnet man schnell für alle $a \in G$ und alle $m, n \in \mathbb{Z}$ die folgenden Rechenregeln nach:

$$a^m a^n = a^{m+n} \quad \text{und} \quad (a^m)^n = a^{mn}.$$

Definition 6.15. *Sei G eine Gruppe mit dem neutralen Element e und sei $a \in G$. Falls ein $m \in \mathbb{N}$ existiert, sodass $a^m = e$ gilt, so definiert man die* Ordnung *von a als das kleinste $m \in \mathbb{N}$ mit $a^m = e$. Falls kein solches m exisitiert, so sagen wir, dass a die Ordnung ∞ hat.*

Die Ordnung *einer Gruppe G ist einfach ihre Mächtigkeit.*

Den Zusammenhang zwischen der Ordnung einer Gruppe und der Ordnung eines Gruppenelements werden wir später noch näher betrachten.

Satz 6.16. *In einer endlichen Gruppe hat jedes Element eine endliche Ordnung.*

Beweis. In einer endlichen Gruppe G gibt es nur endlich viele Möglichkeiten für die Potenzen eines Elements. Ist also $a \in G$ und G endlich, so gibt es $m, n \in \mathbb{N}$ mit $m < n$ und $a^m = a^n$. Nun gilt $a^{n-m}a^m = a^n = a^m = ea^m$. Da man in Gruppen kürzen kann, folgt $a^{n-m} = e$. Damit existiert eine natürliche Zahl k mit $a^k = e$. Also hat a eine endliche Ordnung. □

Beispiel 6.17. (*a*) Zunächst beachte man, dass mit unserer Schreibweise das neutrale Element e von $(\mathbb{Z}, +)$ einfach 0 ist. Auch steht unsere allgemeine Schreibweise a^n im Fall von $(\mathbb{Z}, +)$ für die Zahl $n \cdot a$. Die ganze Zahl 1 hat in $(\mathbb{Z}, +)$ unendliche Ordnung.

(*b*) In $G_\triangle$ haben r und s die Ordnung 3, x, y und z die Ordnung 2 und i die Ordnung 1.

(*c*) In $(\mathbb{Z}/15\mathbb{Z}, +)$ hat $[3]_{15}$ die Ordnung 5. Das Element $[4]_{15}$ hat die Ordnung 15.

(*d*) Wir betrachten die Gruppe $((\mathbb{Z}/10\mathbb{Z})^\times, \cdot)$. Die Zahl 7 ist zu 10 teilerfremd, und damit gilt $[7]_{10} \in (\mathbb{Z}/10\mathbb{Z})^\times$. Wir berechnen die Potenzen von $[7]_{10}$ in $\mathbb{Z}/10\mathbb{Z}$. Es gilt

$$7^2 \equiv 49 \equiv 9 \pmod{10},$$

$$7^3 \equiv 9 \cdot 7 \equiv 63 \equiv 3 \pmod{10}$$

und

$$7^4 \equiv 49 \cdot 49 \equiv 9 \cdot 9 \equiv 81 \equiv 1 \pmod{10}.$$

Also ist 4 die kleinste natürliche Zahl m mit $[7]_{10}^m = [1]_{10}$. Damit ist 4 die Ordnung von $[7]_{10}$ in $(\mathbb{Z}/10\mathbb{Z})^\times$.

(e) Die Permutation $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 1 & 5 \end{pmatrix}$ hat in $\mathcal{S}_n$ die Ordnung 4.

Satz 6.18. *Sei G eine Gruppe und sei $a \in G$ ein Element von endlicher Ordnung m. Dann gilt für alle $n \in \mathbb{Z}$ genau dann $a^n = e$, wenn m ein Teiler von n ist.*

Beweis. Sei zunächst m ein Teiler von n. Dann existiert $q \in \mathbb{Z}$ mit $n = qm$. Nun ist $a^n = a^{qm} = (a^m)^q = e^q = e$.

Sei umgekehrt $a^n = e$. Wähle $q, r \in \mathbb{Z}$ mit $0 \leqslant r < m$ und $n = qm + r$. Dann gilt

$$e = a^n = a^{qm+r} = (a^m)^q a^r = e^q a^r = ea^r = a^r.$$

Da nun m die kleinste natürliche Zahl mit $a^m = e$ ist und da $r < m$ ist, muss $r = 0$ gelten. Damit ist $n = qm$ und $m|n$. □

6.2.2. Isomorphie von Gruppen.

Definition 6.19. *Seien $(G, *_G)$ und $(H, *_H)$ zwei Gruppen. Eine Bijektion*

$$f: G \longrightarrow H$$

heißt ein Isomorphismus *von Gruppen (oder* Gruppenisomorphismus*), falls für alle $a, b \in G$ gilt:*

$$f(a *_G b) = f(a) *_H f(b)$$

Falls ein Isomorphismus zwischen zwei Gruppen G und H existiert, so nennt man die Gruppen isomorph *und schreibt $G \cong H$.*

Wir haben die Operationen $*_G$ und $*_H$ nur der Deutlichkeit halber unterschieden. In unserer normalen Schreibweise lautet die Gleichung $f(a *_G b) = f(a) *_H f(b)$ einfach $f(ab) = f(a)f(b)$.

Lemma 6.20. (i) *Ist $f: G \longrightarrow H$ ein Isomorphismus von Gruppen, so auch*

$$f^{-1}: H \longrightarrow G.$$

(ii) *Sind $f: F \longrightarrow G$ und $g: G \longrightarrow H$ Gruppenisomorphismen, so ist auch*

$$g \circ f: F \longrightarrow H$$

ein Isomorphismus.

(iii) *Ist $f: G \longrightarrow H$ ein Gruppenisomorphismus und sind e_G und e_H die neutralen Elemente von G bzw. H, so gilt $f(e_G) = e_H$. Für jedes $a \in G$ gilt*

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

Beweis. *(i)* Es ist klar, dass f^{-1} eine Bijektion ist. Seien $x, y \in H$. Dann existieren $a, b \in G$ mit $f(a) = x$ und $f(b) = y$. Es gilt $f^{-1}(x) = a$ und $f^{-1}(y) = b$. Da f ein Isomorphismus ist, gilt $f(ab) = f(a)f(b) = xy$. Also ist

$$f^{-1}(xy) = ab = f^{-1}(x)f^{-1}(y).$$

Damit ist f^{-1} ein Isomorphismus.

(ii) Wir wissen schon, dass die Komposition von Bijektionen wieder eine Bijektion ist. Seien $a, b \in F$. Dann gilt

$$(g \circ f)(ab) = g(f(ab)) = g(f(a)f(b)) = g(f(a))g(f(b)) = (g \circ f)(a)(g \circ f)(b)$$

damit ist $g \circ f$ ein Isomorphismus.

(iii) Wir erinnern uns zunächst daran, dass neutrale und inverse Elemente in Gruppen eindeutig bestimmt sind.

Sei $x \in H$. Dann existiert ein $a \in A$ mit $f(a) = x$. Es gilt

$$f(a) = f(e_G a) = f(e_G)f(a) = f(e_G)x.$$

Genauso sieht man, dass $xf(e_G) = x$ gilt. Das zeigt $f(e_G) = e_H$.

Für die Inversen sei wieder $x \in H$ und $a \in G$ mit $f(a) = x$. Dann gilt

$$xf(a^{-1}) = f(a)f(a^{-1}) = f(aa^{-1}) = f(e_G) = e_H.$$

Genauso sieht man $f(a^{-1})x = e_H$. Das zeigt $f(a^{-1}) = x^{-1} = (f(a))^{-1}$. □

Dieses Lemma zeigt unter anderem, dass die Relation $\cong$ zwischen Gruppen symmetrisch und transitiv ist. Da für jede Gruppe G die identische Abbildung

$$\mathrm{id}_G \colon G \longrightarrow G \quad \text{mit} \quad a \longmapsto a$$

ein Isomorphismus ist, ist $\cong$ auch reflexiv.

Beispiel 6.21. Die Gruppen $G_\triangle$ und $\mathcal{S}_3$ sind isomorph.

In Beispiel 6.12 *(e)* hatten wir bereits jeder Transformation in $G_\triangle$ eine Permutation der Menge $\{A, B, C\}$ zugeordnet. Man rechnet leicht nach, dass es sich bei dieser Zuordnung um einen Isomorphismus handelt. Es ist klar, dass die Gruppen $\mathcal{S}_3$ und $\mathcal{S}(\{A, B, C\})$ isomorph sind.

6.2.3. Zyklische Gruppen.

Definition 6.22. *Eine Gruppe G heißt* zyklisch, *wenn es ein Element $a \in G$ mit*

$$G = \{a^n : n \in \mathbb{Z}\}$$

gibt, wenn G also aus den Potenzen eines einzigen Elements besteht. Gilt

$$G = \{a^n : n \in \mathbb{Z}\},$$

so sagt man, dass G von a erzeugt *wird.*

Beispiel 6.23. (*a*) Die Gruppe $(\mathbb{Z}, +)$ ist zyklisch. Alle ganzen Zahlen sind Vielfache von 1. Das Element $a = 1$ erzeugt also die Gruppe $\mathbb{Z}$. Man erinnere sich daran, dass aus dem Vielfachen $n \cdot 1$ in der multiplikativen Schreibweise, die wir für allgemeine Gruppen benutzen, die Potenz a^n wird. Das Element -1 erzeugt ebenfalls die Gruppe $\mathbb{Z}$.

(*b*) Für alle $m \in \mathbb{N}$ ist die Gruppe $(\mathbb{Z}/m\mathbb{Z}, +)$ zyklisch. Diese Gruppe wird von $[1]_m$ erzeugt.

(*c*) Die Gruppe $G_\triangle$ ist nicht zyklisch. Wir weisen diese Behauptung nach, indem wir zeigen, dass kein Element von $G_\triangle$ die ganze Gruppe erzeugt. Für $a = x, y, z$ gilt $a^2 = i$, $a^3 = a$, $a^4 = i$ und so weiter. Mittels vollständiger Induktion weist man leicht nach, dass für alle geraden $n \in \mathbb{Z}$ $a^n = i$ gilt, während für alle ungeraden n $a^n = a$ ist. Also sind nur zwei verschiedene Element von G Potenzen von a.

Für $a = i$ ist jede Potenz von a das Element i. Also erzeugt auch i nicht die ganze Gruppe. Für $a = r, s$ gilt $a^0 = i$, $a^1 = a$, $a^2 = a^{-1}$ und $a^3 = i$. Mittels vollständiger Induktion rechnet man schnell nach, dass $a^n = a^{n \bmod 3}$ gilt. Damit sind nur drei verschiedene Gruppenelemente Potenzen von a.

Wir haben also gesehen, dass es kein $a \in G_\triangle$ gibt, das sechs verschiedene Potenzen hat. Also ist $G_\triangle$ nicht zyklisch.

Satz 6.24. *Eine zyklische Gruppe G ist entweder zu $(\mathbb{Z}, +)$ isomorph oder es gibt ein $m \in \mathbb{N}$ mit $G \cong (\mathbb{Z}/m\mathbb{Z}, +)$.*

Beweis. Da G zyklisch ist, existiert ein $a \in G$ mit

$$G = \{a^n : n \in \mathbb{Z}\}.$$

Sei $f : \mathbb{Z} \longrightarrow G$ definiert durch $f(n) = a^n$.

Ist a von unendlicher Ordnung, so ist f injektiv:

Sonst gäbe es nämlich $m, n \in \mathbb{Z}$ mit $m \neq n$ und $a^m = a^n$. Wir können annehmen, dass $m < n$ gilt. Es ist $a^{n-m} = a^n a^{-m} = a^n (a^m)^{-1} = a^n a^{-n} = e$. Also hat a doch eine endliche Ordnung. Ein Widerspruch.

Da G von a erzeugt wird, ist f auch surjektiv. Nun zeigen wir, dass f ein Isomorphismus ist. Das ist aber einfach. Für alle $m, n \in \mathbb{Z}$ gilt nämlich

$$f(m+n) = a^{m+n} = a^m a^n = f(m)f(n).$$

Damit sind G und $\mathbb{Z}$ isomorph.

Sei nun a von der endlichen Ordnung m. Seien $n, n' \in \mathbb{Z}$, sodass $f(n) = f(n')$ gilt. Dann ist $a^n = a^{n'}$. Damit gilt $a^{n-n'} = e$. Nach Satz 6.18 folgt daraus, dass $n - n'$ ein Vielfaches von m ist. Es gilt also $n \equiv n' \pmod{m}$.

Ist umgekehrt $n \equiv n' \pmod{m}$, so ist $a^{n-n'} = e$, also $a^n = a^{n'}$ und damit $f(n) = f(n')$. Das zeigt, dass die Abbildung $g : \mathbb{Z}/m\mathbb{Z} \longrightarrow G; [n]_m \longmapsto a^n$ wohldefiniert und injektiv ist. Da a die Gruppe G erzeugt, ist g auch surjektiv.

Für alle $n, n' \in \mathbb{Z}$ gilt außerdem

$$g([n]_m + [n']_m) = g([n + n']_m) = a^{n+n'} = a^n a^{n'} = g([n]_m)g([n']_m).$$

Damit ist g ein Isomorphismus. $\square$

Wir haben schon festgestellt, dass die Gruppen $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{Q} \setminus \{0\}, \cdot)$ und $(\mathbb{R} \setminus \{0\}, \cdot)$ das Kommutativgesetz erfüllen, während zum Beispiel $G_\triangle$ nicht das Kommutativgesetz erfüllt.

Definition 6.25. *Eine Gruppe G heißt* kommutativ *oder* abelsch*, wenn für je zwei Elemente $a, b \in G$ gilt: $ab = ba$*

Korollar 6.26. *Alle zyklischen Gruppen sind abelsch.*

Beweis. Ist G zyklisch, so ist G isomorph zu $(\mathbb{Z}, +)$ oder zu einer der Gruppen $(\mathbb{Z}/m\mathbb{Z}, +)$ für ein $m \in \mathbb{N}$. In jedem Falle ist G zu einer abelschen Gruppen isomorph. Damit ist G auch selbst abelsch. $\square$

Die Umkehrung dieses Korollars stimmt nicht. So ist $(\mathbb{Q}, +)$ abelsch, aber nicht zyklisch. Ist nämlich $a \in \mathbb{Q}$ und $a \neq 0$, so ist $\frac{a}{2} \in \mathbb{Q}$, aber $\frac{a}{2}$ ist kein Vielfaches von a.

6.2.4. Untergruppen und Nebenklassen.

Definition 6.27. *Sei $(G, *)$ eine Gruppe. Dann heißt $U \subseteq G$ eine* Untergruppe*, von G, falls U zusammen mit der Einschränkung der Operation $*$ auf $U \times U$ wieder eine Gruppe ist.*

Beispiel 6.28. (*a*) Für $m \in \mathbb{N}$ sei $m\mathbb{Z} = \{m \cdot a \colon a \in \mathbb{Z}\}$ die Menge aller Vielfachen von m. Dann ist $m\mathbb{Z}$ eine Untergruppe von $(\mathbb{Z}, +)$. Um das nachzuweisen, müssen wir zunächst zeigen, dass $+$ überhaupt eine zweistellige Operation auf $m\mathbb{Z}$ ist.

Seien also $a, b \in m\mathbb{Z}$. Dann existieren $c, d \in \mathbb{Z}$ mit $a = mc$ und $b = md$. Wegen

$$a + b = mc + md = m(c + d)$$

ist $a + b$ wieder ein Element von $m\mathbb{Z}$. Damit ist die Einschränkung von $+$ auf $m\mathbb{Z} \times m\mathbb{Z}$ tatsächlich eine Operation auf $m\mathbb{Z}$. Wegen $0 \in m\mathbb{Z}$ hat $m\mathbb{Z}$ ein neutrales Element. Für jedes $ma \in m\mathbb{Z}$ ist $-ma = m(-a) \in m\mathbb{Z}$. Damit existiert in $m\mathbb{Z}$ zu jedem Element ein Inverses. Also ist $m\mathbb{Z}$ eine Untergruppe von $\mathbb{Z}$.

(*b*) Für jede Gruppe G sind $\{e\}$ und G selbst Untergruppen von G.

(*c*) Wir betrachten Untergruppen von $G_\triangle$. Die kleinste Untergruppe ist $\{i\}$, die grösste ist $G_\triangle$ selbst. Weiter sind $\{i, x\}$, $\{i, y\}$ und $\{i, z\}$ Untergruppen, da die Transformationen x, y und z jeweils zu sich selbst invers sind. Schließlich $\{i, r, s\}$ eine Untergruppe von $G_\triangle$.

Das sind alle Untergruppen von $G_\triangle$, wie wir demnächst sehen werden.

Satz 6.29. *Sei G eine Gruppe und $U \subseteq G$.*

(i) U ist genau dann eine Untergruppe von G, wenn für alle $a, b \in U$ gilt:

$$e, a^{-1}, ab \in U$$

(ii) U ist genau dann eine Untergruppe von G, wenn U nicht leer ist und für alle $a, b \in U$ gilt:

$$ab^{-1} \in U$$

(iii) Ist U endlich, so ist U bereits dann eine Untergruppe von G, wenn U nicht leer ist und für alle $a, b \in U$ gilt:

$$ab \in U$$

Beweis. *(i)* Sei U eine Untergruppe von G. Da die Operation von G auf U eingeschränkt immer noch eine zweistellige Operation auf U ist, gilt für alle $a, b \in U$ auch $ab \in U$.

Sei e_U das neutrale Element der Gruppe U. Dann gilt in U die Gleichung $e_U e_U = e_U$. Damit gilt in G die Gleichung $e_U e_U = e_U e$, wobei e das neutrale Element von G ist. Nach Lemma 6.13 *(i)* folgt aus der Gleichung $e_U e_U = e_U e$, dass $e_U = e$ gilt. Also ist $e \in U$ und die neutralen Elemente von U und G stimmen überein.

Für $a \in U$ existiert $b \in U$ mit $ab = e$. Bezeichne a^{-1} das Inverse von a in G. Dann ist $ab = aa^{-1}$. Aus Lemma 6.13 *(i)* folgt $a^{-1} = b$. Insbesondere gilt $a^{-1} \in U$.

Gelte umgekehrt für alle $a, b \in U$

$$e, a^{-1}, ab \in U.$$

Dann ist die Operation von G eingeschränkt auf U eine zweistellige Operation auf U. Außerdem enthält U das neutrale Element von G, welches auch ein neutrales Element von U ist. Für jedes $a \in U$ ethält U auch das Inverse a^{-1}. Da $aa^{-1} = e$ in G gilt, gilt die Gleichung auch in U. Also ist a^{-1} auch in U zu a invers. Das zeigt, dass U eine Untergruppe von G ist.

(ii) Ist U eine Untergruppe von G und sind a und b in U, so gilt nach *(i)* $b^{-1} \in U$. Ebenfalls nach *(i)* gilt: $ab^{-1} \in U$

Gelte nun für alle $a, b \in U$ auch $ab^{-1} \in U$ und sei $U \neq \varnothing$. Sei $a \in U$. Dann gilt $e = aa^{-1} \in U$. Also gilt für alle $a \in U$ auch $a^{-1} = ea^{-1} \in U$. Seien nun $a, b \in U$. Dann ist $b^{-1} \in U$. Es folgt $ab = a(b^{-1})^{-1} \in U$. Damit ist U eine Untergruppe von G.

(iii) Sei $a \in U$. Nach Lemma 6.13 sind die Elemente ab, $b \in U$, paarweise verschieden. Da sie auch Elemente von U sind, muss es ein $b \in U$ mit $ab = a$ geben. Wieder nach Lemma 6.13 gilt $b = e$. Damit ist $e \in U$. Also gibt es ein $b \in U$ mit $ab = e$. Es gilt $b = a^{-1}$. Nach *(i)* ist U eine Untergruppe von G. □

Definition 6.30. *Sei G eine Gruppe und $U \subseteq G$ eine Untergruppe. Für $a \in G$ schreiben wir aU für die Menge $\{ag\colon g \in U\}$ sowie Ua für die Menge $\{ga\colon g \in U\}$. Wir nennen die Mengen der Form aU* Linksnebenklassen *von U und die Mengen der Form Ua* Rechtsnebenklassen.

Beispiel 6.31. (*a*) Sei $G = (\mathbb{Z}, +)$, und $U = 6\mathbb{Z}$. Dann ist die Rechtsnebenklasse von 4 von U die Menge $6\mathbb{Z} + 4 = \{\dots, -2, 4, 10, \dots\} = [4]_6$. Hierbei beachte man, dass die Operation die Gruppe G die Addition ist, auch wenn wir die Operation auf einer Gruppe im Allgemeinen multiplikativ schreiben. Die Linksnebenklasse von 4 von U ist die Menge $4 + 6\mathbb{Z}$, die aber mit $6\mathbb{Z} + 4$ übereinstimmt, da $+$ das Kommutativgesetz erfüllt.

(*b*) Wir betrachten die Gruppe $G_\triangle$ und die Untergruppe $U = \{i, y\}$. Dann gilt $iU = \{i, y\}$, $xU = \{x, r\}$, $yU = \{y, i\}$, $zU = \{z, s\}$, $rU = \{r, x\}$ und $sU = \{s, z\}$, wie man leicht an der Gruppentafel von $G_\triangle$ abliest. Die verschiedenen Linksnebenklassen von U in $G_\triangle$ sind also die Mengen $iU = yU = U = \{i, y\}$, $xU = rU = \{r, x\}$ und $zU = sU = \{z, s\}$.

Die entsprechende Rechnung liefert die Rechtsnebenklassen $Ui = Uy = U = \{i, y\}$, $Ux = Us = \{x, s\}$ und $Uz = Ur = \{z, r\}$.

Satz 6.32. *Sei G eine Gruppe und $U \subseteq G$ eine Untergruppe.*

(*i*) *Für jedes $a \in G$ ist $a \in aU$ und $a \in Ua$.*

(*ii*) *Für alle $c \in U$ ist $cU = U = Uc$.*

(*iii*) *Für $a, b \in G$ mit $b \in aU$ gilt $aU = bU$. Für $a, b \in G$ mit $b \in Ua$ gilt $Ua = Ub$.*

(*iv*) *Für $a, b \in G$ sind die Linksnebenklassen aU und bU entweder disjunkt oder gleich. Auch die Rechtsnebenklassen Ua und Ub sind entweder disjunkt oder gleich.*

(*v*) *Für alle $a \in G$ sind aU, U und Ua gleichmächtig.*

Beweis. (*i*) Wegen $e \in U$ gilt $a = ae \in aU$ und $a = ea \in Ua$.

(*ii*) Es ist klar, dass $cU, Uc \subseteq U$ gilt. Sei nun $d \in U$. Dann ist $c^{-1}d \in U$. Also ist $d = cc^{-1}d \in U$. Das zeigt $U \subseteq cU$. Auf ähnliche Weise sieht man $U \subseteq Uc$.

(*iii*) Ist $b \in aU$, so existiert $c \in U$ mit $b = ac$. Es gilt $bU = acU = aU$. Auf ähnliche Weise sieht man $U = Ub$, falls $b \in Ua$ gilt.

(*iv*) Falls $aU \cap bU$ nicht leer ist, so existiert $c \in aU \cap bU$. Nach (*iii*) gilt $aU = cU = bU$. Auf ähnliche Weise sieht man, dass Ua und Ub entweder gleich oder disjunkt sind.

(*v*) Wir zeigen nur, dass U und aU gleichmächtig sind, indem wir eine Bijektion zwischen beiden Mengen angeben. Die Gleichmächtigkeit von U und Ua kann auch ähnliche Weise nachgerechnet werden.

Sei $f\colon U \longrightarrow aU$ mit $b \longmapsto ab$. Aus der Definition von aU folgt sofort, dass f surjektiv ist. Seien nun $b, c \in U$ mit $ab = f(b) = f(c) = ac$. Nach Lemma 6.13 (*i*) folgt daraus $b = c$. Damit ist f injektiv. Also sind U und aU in der Tat gleichmächtig. □

Beispiel 6.33. Sei G eine Gruppe und $a \in G$. Dann ist $\langle a \rangle := \{a^n : n \in \mathbb{Z}\}$ eine Untergruppe von G, die von a *erzeugte Untergruppe* von G. Die Ordnung von U ist genau die Ordnung von a.

Korollar 6.34 (Satz von Lagrange)**.** *Ist G eine endliche Gruppe und U eine Untergruppe von G, so ist die Ordnung von U ein Teiler der Ordnung von G. Insbesondere ist die Ordnung von jedem Element von G ein Teiler von $|G|$.*

Beweis. Nach Satz 6.32 bilden die Rechtsnebenklassen von U eine Partition von G in Klassen der Mächtigkeit $|U|$. Ist m die Anzahl der verschiedenen Rechtsnebenklassen, so gilt $|G| = m \cdot |U|$. Die Ordnung eines Elements a von G ist die Ordnung der von a erzeugten Untergruppe und damit ein Teiler der Ordnung von G. $\square$

Definition 6.35. *Sei G eine Gruppe und U eine Untergruppe von G. Die Zahl der Rechtsnebenklassen von U in G (die identisch ist mit der Zahl der Linksnebenklassen) nennt man den* Index *von U in G. Man schreibt $[G : U]$ für den Index von U in G.*

Der Beweis des Satzes von Lagrange zeigt also für jede endliche Gruppe G und jede Untergruppe U die Gleichung

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

was auch die Notation $[G : U]$ erklärt.

Beispiel 6.36. Wir betrachten wieder die Dreiecksgruppe $G_\triangle$. Die Gruppe hat 6 Elemente. Also sind die möglichen Ordnungen von Untergruppen von G die Zahlen 1, 2, 3 und 6. Die einzige Untergruppe der Ordnung 1 ist $\{i\}$. Diese Untergruppe hat den Index 6.

Ist $U \subseteq G_\triangle$ eine Untergruppe der Ordnung 2, so enthält U das Element i und ein weiteres Element, dass die Ordnung 2 haben muss. Damit sind die Untergruppen der Ordnung 2 genau $\{i, x\}$, $\{i, y\}$ und $\{i, z\}$. Diese Untergruppen haben den Index 3.

Sei nun U eine Untergruppe von G der Ordnung 3. Nach Korollar 6.34 hat jedes Element von U eine Ordnung, die die Zahl 3 teilt. Also hat U nur Elemente der Ordnung 1 und 3. Damit ist $U = \{i, r, s\}$. Diese Untergruppe hat den Index 2.

Die einzige Untergruppe von $G_\triangle$ mit 6 Elementen ist $G_\triangle$ selbst. Diese Untergruppe hat den Index 1.

Wir bestimmen die Nebenklassen der Untergruppen von $G_\triangle$. Für jede Untergruppe U ist $U = iU = Ui$ sowohl eine Rechts- als auch Linksnebenklasse. $U = G_\triangle$ hat nur die Nebenklasse U, und hierbei ist es egal, ob wir Rechts- oder Linksnebenklassen betrachten.

$U = \{i, r, s\}$ hat die Rechts und Linksnebenklasse U. Da die Nebenklassen alle dieselbe Mächtigkeit haben wie U und eine Partition von $G_\triangle$ bilden, gibt es genau eine weitere Nebenklasse, nämlich $\{x, y, z\}$. Diese Menge ist wieder sowohl Rechts- als auch Linksnebenklasse.

Nun betrachten wir eine Untergruppe der Ordnung 2, zum Beispiel $U = \{i, x\}$. Es gibt insgesamt 3 Rechts- und 3 Linksnebenklassen. Eine Nebenklasse, die sowohl Rechts- als auch Linksnebenklasse ist, ist U selbst. Es gilt $yU = \{y, s\}$, wie wir der Gruppentafel von $G_\triangle$ entnehmen. $\{y, s\}$ ist also eine Linksnebenklasse von U. Da die Linksnebenklassen von U eine Partition von $G_\triangle$ bilden und alle dieselbe Mächtigkeit haben, hat U noch eine dritte Linksnebenklasse, nämlich $\{z, r\}$.

Auf dieselbe Weise rechnet man nach, dass die Rechtsnebenklassen von U genau die Mengen U, $Uy = \{y, r\}$ und $\{z, s\}$ sind. Insbesondere sind die Linksnebenklassen von U in $G_\triangle$ nicht identisch mit den Rechtsnebenklassen.

Die Nebenklassen von $U = \{i\}$ sind die Einermengen $U = \{i\}$, $\{x\}$, $\{y\}$, $\{z\}$, $\{r\}$ und $\{s\}$. Hierbei stimmen wieder die Links- und Rechtsnebenklassen überein, auch wenn $G_\triangle$ nicht abelsch ist.

Beispiel 6.37. Auch wenn die Gruppe G und ihre Untergruppe U unendlich sind, kann es sein, dass der Index von U in G endlich ist. Für jedes $m \in \mathbb{N}$ ist $m\mathbb{Z}$ eine Untergruppe von $\mathbb{Z}$ und es gilt

$$[\mathbb{Z} : m\mathbb{Z}] = m,$$

da die Mengen $[0]_m, \dots, [m-1]_m$ genau die verschiedenen Nebenklassen von $m\mathbb{Z}$ in $\mathbb{Z}$ sind. In $\mathbb{Z}$ ist es nicht nötig, zwischen Links- und Rechtsnebenklassen zu unterscheiden, da die Gruppe abelsch ist.

Beispiel 6.38. Aus dem Satz von Lagrange (Korollar 6.34) können wir sehr einfach den Satz von Fermat und Euler (Satz 5.14) folgern. Sei $m \geqslant 2$ und $n \in \mathbb{Z}$ zu m teilerfremd. Dann ist $[n]_m \in (\mathbb{Z}/m\mathbb{Z})^\times$ und $(\mathbb{Z}/m\mathbb{Z})^\times$ hat die Ordnung $\varphi(m)$. Nach dem Satz von Lagrange ist die Ordnung von $[n]_m$ in $(\mathbb{Z}/m\mathbb{Z})^\times$ ein Teiler der Ordnung $\varphi(m)$ von $(\mathbb{Z}/m\mathbb{Z})^\times$. Damit gilt aber $([n]_m)^{\varphi(m)} = [1]_m$, also $n^{\varphi(m)} \equiv 1 \pmod m$.

Satz 6.39. *Sei G eine zyklische Gruppe. Ist U eine Untergruppe von G, so ist auch U zyklisch.*

Beweis. Sei a das erzeugende Element von G, also $G = \{a^n : n \in \mathbb{Z}\}$. Ist $U = \{e\}$, so ist U zyklisch. Wir können also annehmen, dass U ein von e verschiedenes Element enthält. Also gibt es ein $n \in \mathbb{Z}$ mit $n \neq 0$ und $a^n \in U$. Mit a^n ist auch $a^{-n} = (a^n)^{-1}$ in U. Damit existiert ein $n > 0$ mit $a^n \in U$.

Sei nun m die kleinste natürliche Zahl mit $a^m \in U$. Wir zeigen, dass alle Elemente von U Potenzen von a^m sind. Sei $a^n \in U$. Wir zeigen, dass n ein Vielfaches von m ist. Wieder können wir annehmen, dass $n > 0$ ist.

Seien $q, r \in \mathbb{Z}$ mit $n = qm + r$ und $0 \leqslant r < m$. Dann gilt $a^n a^{-qm} = a^r \in U$. Aus $r < m$ und der Wahl von m als kleinste natürliche Zahl mit $a^m \in U$ folgt $r = 0$. Damit ist $n = qm$ und $a^n = (a^m)^q$. Das zeigt, dass U zyklisch ist. □

Beispiel 6.40. Wir betrachten die Untergruppen der Gruppe $\mathbb{Z}/12\mathbb{Z}$. Die möglichen Ordnungen sind 1, 2, 3, 4, 6 und 12 und alle Untergruppen sind zyklisch.

Für alle $m \in \{1, \ldots, 11\}$ die zu 12 teilerfremd sind, erzeugt $[m]_{12}$ die ganze Gruppe $\mathbb{Z}/12\mathbb{Z}$. $[2]_{12}$ und $[10]_{12}$ erzeugen jeweils die Untergruppe

$$\{[0]_{12}, [2]_{12}, [4]_{12}, [6]_{12}, [8]_{12}, [10]_{12}\}.$$

$[3]_{12}$ und $[9]_{12}$ erzeugen jeweils die Untergruppe

$$\{[0]_{12}, [3]_{12}, [6]_{12}, [9]_{12}\}.$$

$[4]_{12}$ und $[8]_{12}$ erzeugen jeweils die Untergruppe

$$\{[0]_{12}, [4]_{12}, [8]_{12}\}.$$

$[6]_{12}$ erzeugt die Untergruppe

$$\{[0]_{12}, [6]_{12}\}.$$

$[0]_{12}$ erzeugt schließlich die Untergruppe $\{[0]_{12}\}$. Das sind alle Untergruppen von $\mathbb{Z}/12\mathbb{Z}$.

Satz 6.41. *Ist G eine Gruppe, deren Ordnung eine Primzahl p ist. Dann ist G zyklisch und die einzigen Untergruppen von G sind G und $\{e\}$.*

Beweis. Sei $a \in G$. Nach dem Satz von Lagrange ist die Ordnung von a ein Teiler von p. Damit hat a entweder die Ordnung 1 oder p. Im ersten Fall gilt $a = e$. Im zweiten Fall ist $G = \{a^n : n \in \mathbb{Z}\}$. □

§6.3 Permutationen

Man kann zeigen, dass jede Gruppe zu einer Menge von Permutationen isomorph ist. Daher ist das Studium von Permutationen in der Gruppentheorie von besonderem Interesse.

Zur Erinnerung: Eine Permutation einer Menge A ist eine Bijektion von A nach A. Die Komposition $g \circ f$ zweier Permutationen einer Menge A ist wieder eine Permutation von A. Die Menge aller Permutationen einer Menge A zusammen mit der Komposition $\circ$ ist eine Gruppe $\mathcal{S}(A)$. Das neutrale Element ist die Identität

$$\mathrm{id}_A : A \longrightarrow A, \quad x \longmapsto x.$$

Für jede Permutation $\pi \in \mathcal{S}(A)$ ist die Umkehrfunktion π^{-1} das zu π inverse Element von $\mathcal{S}(A)$.

Ist A endlich, also zum Beispiel $A = \{a_1, \ldots, a_n\}$, so können wir eine Permutation $\pi : A \longrightarrow A$ als

$$\begin{pmatrix} a_1 & \ldots & a_n \\ \pi(a_1) & \ldots & \pi(a_n) \end{pmatrix}$$

aufschreiben.

Beispiel 6.42. Es gilt

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 1 & 4 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 5 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}.$$

Die Permutation auf der rechten Seite der Gleichung ist $\mathrm{id}_{\{1,2,3,4,5\}}$. Damit sind die beiden Permutationen auf der linken Seite der Gleichung in $\mathcal{S}_5 = \mathcal{S}(\{1,2,3,4,5\})$ invers zueinander.

Wir betrachten die Permutation $\pi := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 5 & 1 & 4 \end{pmatrix}$ etwas eingehender. Es gilt $\pi(2) = 2$. Die 2 wird also durch π auf sich selbst abgebildet. Die 1 wird durch π auf 3 abgebildet, die 3 auf die 5, die 5 auf die 4 und die 4 wieder auf die 1. Iteriert man also die Anwendung von π auf 1 so landet man zunächst bei 3, dann bei 5, bei 4 und schließlich wieder bei 1.

Lemma 6.43. *Ist A eine endliche Menge und $\pi \in \mathcal{S}(A)$, so existiert für jedes $a \in A$ ein $n \in \mathbb{N}$ mit $\pi^n(a) = a$.*

Beweis. Da A endlich ist, gibt es $k, \ell \in \mathbb{N}$ mit $k < \ell$ und $\pi^k(a) = \pi^\ell(a)$. Nun gilt $a = (\pi^{-k} \circ \pi^k)(a) = (\pi^{-k} \circ \pi^\ell(a) = \pi^{\ell-k}(a)$. Setzt man $n := \ell - k$, so ergibt sich $\pi^n(a) = a$. □

Definition 6.44. *Sei A eine Menge, $n \geqslant 2$ und $a_1, \ldots, a_n$ paarweise verschiedene Elemente von A. Dann bezeichen wir mit $(a_1 a_2 \ldots a_n)$ die Permutation π von A, die wie folgt definiert ist:*

$$\pi(a) = \begin{cases} a, \text{ falls } a \in A \setminus \{a_1, \ldots, a_n\}, \\ a_{i+1}, \text{ falls } a = a_i \text{ für ein } i \in \{1, \ldots, n-1\} \text{ und} \\ a_1, \text{ falls } a = a_n. \end{cases}$$

Die Permutation $(a_1 a_2 \ldots a_n)$ nennen wir einen Zyklus *der Länge n.*

Zwei Zyklen $(a_1 \ldots a_n)$ und $(b_1 \ldots b_m)$ heißen disjunkt, falls die Mengen

$$\{a_1, \ldots, a_n\} \text{ und } \{b_1, \ldots, b_m\}$$

disjunkt sind. Zyklen der Länge 2 heißen Transpositionen.

Satz 6.45. *Sei A eine endliche Menge.*

(*i*) *Jede Permutation π von A ist ein Produkt von paarweise disjunkten Zyklen. Eine Darstellung von π als Produkt disjunkter Zyklen heißt* Zyklenzerlegung *von π. Die Zyklenzerlegung von π ist bis auf die Reihenfolge eindeutig.*

(*ii*) *Jeder Zyklus ist ein Produkt von Transpositionen.*

(*iii*) *Jede Permutation von A ist ein Produkt von Transpositionen.*

Beweis. (*i*) Für $a, b \in A$ schreiben wir $a \sim b$, falls es ein $n \in \mathbb{Z}$ mit $\pi^n(a) = b$ gibt. Die Relation $\sim$ ist eine Äquivalenzrelation auf A. Sei nun $a \in A$. Nach Lemma

6.43 existiert ein $m \in \mathbb{N}$ mit $\pi^m(a) = a$. Sei nun $b \sim a$. Dann existiert ein $n \in \mathbb{Z}$ mit $\pi^n(a) = b$. Wähle $q, r \in \mathbb{Z}$ mit $n = q \cdot m + r$ und $0 \leqslant r < m$. Dann gilt

$$b = \pi^n(a) = \pi^{q \cdot m + r}(a) = \pi^r((\pi^m)^q(a)) = \pi^r(a).$$

Das zeigt, dass die Äquivalenzklasse von a genau die Menge $\{\pi^0(a), \ldots, \pi^{m-1}(a)\}$ ist.

Ist $m = 1$, so besteht diese Äquivalenzklasse nur aus dem Element a und a wird von π nicht bewegt. Ist $m > 1$, so ist π auf der Äquivalenzklasse von a genau der Zyklus $(\pi^0(a), \ldots, \pi^{m-1}(a))$.

Für jede Äquivalenzklasse mit mindestens zwei Elementen erhalten wir also einen Zyklus, dessen Einträge genau die Elemente dieser Äquivalenzklasse sind. Da die Äquivalenzklassen paarweise disjunkt sind, sind diese Zyklen disjunkt. Die Permutation π ist das Produkt dieser Zyklen.

(ii) Es gilt $(a_1, \ldots, a_n) = (a_1 a_2) \circ \cdots \circ (a_{n-1} a_n)$.

(iii) Die Behauptung folgt sofort aus *(ii)* und *(iii)*. □

Beispiel 6.46. Sei $A = \{1, 2, 3, 4, 5, 6\}$ und

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 1 & 3 & 6 & 2 \end{pmatrix}.$$

Dann gilt

$$\pi = (143) \circ (256).$$

Weiter gilt

$$(143) = (14) \circ (43)$$

und

$$(256) = (25) \circ (56).$$

Damit ist

$$\pi = (14) \circ (43) \circ (25) \circ (56).$$

Satz 6.47. *Sei π eine Permutation einer endlichen Menge A. Ist π ein Produkt von gerade vielen Transpositionen, so hat jede Darstellung von π als Produkt von Transpositionen eine gerade Anzahl von Faktoren. In diesem Falle nennen wir π eine* gerade *Permutation. Permutationen, die nicht gerade sind, nennen wir* ungerade.

Beweis. Es ist hinreichend zu zeigen, dass sich die Identität id_A nur durch eine gerade Anzahl von Transpositionen darstellen lässt (Wieso?). Sei also

$$\mathrm{id}_A = \tau_1 \circ \cdots \circ \tau_k \tag{6.1}$$

für ein $k \in \mathbb{N}_0$. Wir zeigen mit Induktion nach k, dass k gerade sein muss. Da eine einzelne Transposition nicht die Identität darstellen kann, können wir annehmen, dass

$k \geqslant 3$ ist. Ausgehend von (6.1) werden wir zeigen, dass es k Transpositionen $\tau_1', \dots, \tau_k'$ gibt, sodass für ein $i \in [k-1]$ gilt

$$\tau_i' = \tau_{i+1}'\,. \tag{6.2}$$

Da Transpositionen selbstinvers sind, folgt damit

$$\mathrm{id}_A = \tau_1' \circ \cdots \circ \tau_{i-1}' \circ \tau_{i+2}' \circ \cdots \circ \tau_k'$$

und wegen der Induktionsvoraussetzung ist somit $k-2$ und deswegen auch k gerade.

Sei $\tau_k = (x\,y)$ für zwei verschiedene Elemente x, $y \in A$. Wir betrachten vier Fälle in Abhängigkeit von τ_{k-1}. Falls $\tau_{k-1} = \tau_k$, dann haben wir (6.2) gezeigt und sind fertig.

Falls τ_{k-1} disjunkt von τ_k ist, d. h. falls $\tau_{k-1} = (a\,b)$ mit $\{a, b\} \cap \{x, y\} = \varnothing$, dann können wir

$$\tau_{k-1} \circ \tau_k = (a\,b) \circ (x\,y) \qquad \text{durch} \qquad \tau_{k-1}' \circ \tau_k' = (x\,y) \circ (a\,b)$$

ersetzen. Durch die Ersetzung erreichen wir, dass das Element x nicht mehr in der letzten Transposition vorkommt. In den restlichen beiden Fällen

$$\tau_{k-1} = (x\,b) = (b\,x) \qquad \text{b. z. w.} \qquad \tau_{k-1} = (a\,y) = (y\,a)$$

werden wir ähnlich argumentieren

Tatsächlich können wir sowohl

$$\tau_{k-1} \circ \tau_k = (x\,b) \circ (x\,y) \qquad \text{durch} \qquad \tau_{k-1}' \circ \tau_k' = (x\,y) \circ (y\,b)$$

und

$$\tau_{k-1} \circ \tau_k = (a\,y) \circ (x\,y) \qquad \text{durch} \qquad \tau_{k-1}' \circ \tau_k' = (x\,a) \circ (a\,y)$$

ersetzen. In jedem Fall ist das Element x nicht mehr Teil der letzten Transposition τ_k'.

Nun wiederholen wir das gleiche Argument und betrachten die die vorletzte Transposition τ_{k-1}' die als erstes Element x enthält und die vorvorletzte Transposition mit Index $k-2$. Durch geeignete Ersetzungen dieser beiden Transpositionen (genau wie zuvor) erreichen wir, dass x in keiner der beiden letzten Transpositionen vorkommt.

Wenn wir dieses Argument $(k-1)$-Mal wiederholen könnten, ohne dass jemals die Situation (6.2) eintritt, dann würde das Element x nur noch in der ersten Transposition (mit Index 1) auftauchen. Somit würde aber x nicht auf sich selbst abgebildet werden, im Widerspruch dazu, dass die Transpositionen die Identität id_A darstellen. D. h. irgendwann muss Situation (6.2) eintreten, was den Beweis abschließt. ☐

Korollar 6.48. *Sei A eine endliche Menge. Die geraden Permutationen bilden eine Untergruppe der Gruppe aller Permutationen von A vom Index* 2.

Beweis. Es ist klar, dass das Produkt zweier gerader Permutationen wieder gerade ist. Man sieht auch schnell, dass das Inverse einer geraden Permutation wieder gerade ist.

Die Untergruppe U von $\mathcal{S}(A)$ der geraden Permutationen hat genau zwei Nebenklassen, nämlich U selbst und die Menge der ungeraden Permutationen. □

Beispiel 6.49. Die Gruppe $\mathcal{S}_3$ hat $3! = 6$ Elemente. Damit gibt es 3 gerade Permutationen und 3 ungerade Permutationen. Die die geraden Permutationen sind die Identität, $(123) = (12)(23)$ und $(321) = (32)(21)$. Die ungeraden Permutationen sind (12), (13) und (23). Man beachte, dass die Darstellungen von Permutationen als Produkt von Transpositionen nicht eindeutig ist. Es gilt zum Beispiel

$$(123) = (12)(23) = (231) = (23)(31) = (312) = (31)(12).$$

Auch die Anzahl der Transpositionen ist nicht eindeutig:

$$(321) = (32)(21) = (123)^2 = (12)(23)(31)(12)$$

Was aber nach Satz 6.47 eindeutig ist, ist die Anzahl der Transpositionen modulo 2.

§6.4 Ringe und Körper

Definition 6.50. *Eine Menge R zusammen mit zwei binären Operationen $+$ und $\cdot$ und zwei verschiedenen Konstanten* 0 *und* 1 *heißt ein* Ring *(mit* 1*), falls für alle $a, b, c \in R$ die folgenden Axiome gelten:*

(*R1*) *Assoziativgesetze*
- $a + (b + c) = (a + b) + c$
- $a \cdot (b \cdot c) = (a \cdot b) \cdot c$

(*R2*) *Kommutativgesetz der Addition:*
- $a + b = b + a$

(*R3*) *Distributivgesetze*
- $a \cdot (b + c) = a \cdot b + a \cdot c$
- $(b + c) \cdot a = b \cdot a + c \cdot a$

(*R4*) *Existenz neutraler Elemente bezüglich der Addition und der Multiplikation*
- $a + 0 = a$
- $1 \cdot a = a$

(*R5*) *Existenz inverser Elemente bezüglich der Addition*
- *Es gibt ein Element $-a$ mit $a + (-a) = 0$.*

Man beachte, dass der offizielle Name für hier definierten Strukturen „Ring mit 1“ lautet. Wir werden aber keine Ringe ohne 1 betrachten und sagen daher abkürzend einfach „Ring“, obwohl wir eigentlich „Ring mit 1“ meinen. Unter Verwendung der Begriffe Gruppe und Monoid können wir Ringe auch in der folgenden kompakten Form definieren.

Definition 6.51. *Eine Menge R mit zwei binären Operationen $+$ und $\cdot$ ist ein* Ring *(mit 1) falls gilt:*

(RI) $(R,+)$ *ist eine kommutative Gruppe.*
(RII) $(R,\cdot)$ *ist ein Monoid.*
(RIII) *Es gelten die Distributivgesetze, d.h., für alle* $a,b,c \in R$ *gilt:*
- $a \cdot (b+c) = a \cdot b + a \cdot c$
- $(b+c) \cdot a = b \cdot a + c \cdot a$

Bei dieser Definition definieren wir 0 *als das neutrale Element der Addition und* 1 *als das neutrale Element der Multiplikation.*

Wie üblich schreiben wir $-a$ für das additive Inverse eines Ringelements a und a^{-1} für das multiplikative Inverse, falls es denn existiert.

Beispiel 6.52. (a) Jeder Körper ist ein Ring. Umgekehrt ist ein Ring $(R,+,\cdot)$ ein Körper, wenn das Kommutativgesetz für $\cdot$ gilt und jedes von 0 verschiedene Element ein multiplikatives Inverses besitzt.
(b) Die ganzen Zahlen mit Addition, Multiplikation und den üblichen Konstanten 0 und 1 bilden einen Ring, aber bekanntlich keinen Körper.
(c) Für jedes $m \geqslant 2$ ist $(\mathbb{Z}/m\mathbb{Z},+,\cdot)$ ein Ring.

Definition 6.53. *Sei* $(R,+,\cdot)$ *ein Ring. Die* Einheitengruppe $R^\times$ *von* R *ist die Menge derjenigen Elemente von* R*, die ein mutliplikatives Inverses besitzen, zusammen mit der Multiplikation.*

Wir hatten schon gesehen, dass die Einheitengruppe eines Ringes der Form $\mathbb{Z}/m\mathbb{Z}$, $m \geqslant 2$, tatsächlich eine Gruppe ist. Das gleiche Argument liefert die entsprechende Aussage für beliebige Ringe:

Satz 6.54. *Für jeden Ring* R *ist* $R^\times$ *eine Gruppe.*

Beweis. Zunächst müssen wir zeigen, dass $\cdot$ überhaupt eine Operation auf $R^\times$ ist, dass also das Produkt zweier invertierbarer Elemente von R wieder invertierbar ist. Seien also $a,b \in R^\times$. Dann ist

$$(ab)(b^{-1}a^{-1}) = aa^{-1} = 1 = b^{-1}b = (b^{-1}a^{-1})(ab).$$

Also ist ab invertierbar und es gilt $(ab)^{-1} = b^{-1}a^{-1}$.

Die 1 ist zu sich selbst invers und damit gilt $1 \in R^\times$. Es ist auch klar, dass mit $a \in R$ auch a^{-1} invertierbar ist. Das Inverse von a^{-1} ist nämlich einfach a. Damit ist $R^\times$ tatsächlich eine Gruppe. □

Beispiel 6.55. (a) Für jeden Körper K ist $K^\times = K \setminus \{0\}$. Insbesondere ist $\mathbb{R}^\times = \mathbb{R} \setminus \{0\}$, $\mathbb{Q}^\times = \mathbb{Q} \setminus \{0\}$ und $(\mathbb{Z}/p\mathbb{Z})^\times = \mathbb{Z}/p\mathbb{Z} \setminus \{[0]_p\}$ für jede Primzahl p.
(b) Es gilt $\mathbb{Z}^\times = \{-1,1\}$.
(c) Es gilt

$$(\mathbb{Z}/8\mathbb{Z})^\times = \{[1]_8,[3]_8,[5]_8,[7]_8\}$$

und
$$(\mathbb{Z}/12\mathbb{Z})^\times = \{[1]_{12}, [5]_{12}, [7]_{12}, [11]_{12}\}\,.$$

Im Allgemeinen folgt aus Satz 5.8
$$(\mathbb{Z}/n\mathbb{Z})^\times = \big\{[a]_n\colon \operatorname{ggT}(a,n) = 1\big\}\,.$$

Wir werden später noch zwei wichtige Beispiele für Ringe kennenlernen, nämlich Polynomringe und Matrizenringe. Matrizenringe werden unser erstes Beispiel für Ringe sein, bei denen die Multiplikation im Allgemeinen nicht das Kommutativgesetz erfüllt. Bevor zir zu den Polynomringen kommen, führen wir noch einen weiteren wichtigen Körper ein.

§6.5 Die komplexen Zahlen

Einer der Gründe, warum man anstelle der rationalen Zahlen mit den reellen Zahlen arbeitet ist der, dass sich gewisse Gleichungen in $\mathbb{Q}$ nicht lösen lassen, während in $\mathbb{R}$ Lösungen existieren. Ein Beispiel ist die Gleichung $x^2 = 2$, die die irrationalen Lösungen $\pm\sqrt{2}$ hat. Da das Quadrat jeder reellen Zahl $\geqslant 0$ ist, lässt sich aber zum Beispiel die Gleichung $x^2 = -1$ in $\mathbb{R}$ nicht lösen. Dieses Problem lösen wir, indem wir ein letztes Mal den Zahlenbereich erweitern und von den reellen Zahlen zu den *komplexen Zahlen* übergehen. Die komplexen Zahlen werden in vielen Anwendungen der Mathematik benötigt, etwa in der Physik oder in der Elektrotechnik.

Definition 6.56. *Wir setzen*
$$\mathbb{C} = \mathbb{R}^2 = \{(a,b) : a, b \in \mathbb{R}\}.$$
Auf $\mathbb{C}$ definieren wir eine Addition und eine Multiplikation. Für alle $a, b, c, d \in \mathbb{R}$ sei
$$(a,b) + (c,d) := (a+c, b+d)$$
und
$$(a,b) \cdot (c,d) := (ac - bd, ad + bc).$$

Satz 6.57. *Die Menge $\mathbb{C}$ bildet bezüglich der in Definition 6.56 definierten Addition und Multiplikation einen Körper, den wir auch mit $\mathbb{C}$ bezeichnen.*

Beweis. Die Assoziativ- und Kommutativgesetze sowie das Distributivgesetz rechnet man schnell nach. Das neutrale Element bezüglich der Addition ist $(0,0)$. Wie üblich bezeichnen wir dieses neutrale Element ebenfalls mit 0. Wir wissen auch schon, dass $(\mathbb{C}, +)$ eine abelsche Gruppe ist, nämlich dieselbe abelsche Gruppe wie die Gruppe $(\mathbb{R}^2, +)$. Das zu (a,b) inverse Element bezüglich $+$ ist also $-(a,b) = (-a,-b)$.

Das neutrale Element bezüglich $\cdot$ ist $(1,0)$. Für alle $(a,b) \in \mathbb{C}$ gilt nämlich
$$(a,b)\cdot(1,0) := (a\cdot 1 - b\cdot 0, a\cdot 0 + b\cdot 1) = (a,b).$$

Ist $(a,b) \neq 0$ so ist $\left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2}\right)$ das zu (a,b) inverse Element bezüglich der Multiplikation:

$$
\begin{aligned}
(a,b) \cdot & \left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2}\right) \\
& = \left(a \cdot \frac{a}{a^2+b^2} - b \cdot \frac{-b}{a^2+b^2}, a \cdot \frac{-b}{a^2+b^2} + b \cdot \frac{a}{a^2+b^2}\right) = (1,0)
\end{aligned}
$$

□

Um nun den Körper $\mathbb{C}$ der *komplexen Zahlen* als Erweiterung von $\mathbb{R}$ auffassen zu können, müssen wir $\mathbb{R}$ mit einer geeigneten Teilmenge von $\mathbb{C}$ identifizieren. Diese Teilmenge ist einfach die x-Achse in $\mathbb{R}^2$, also die Menge

$$R = \{(a,0) : a \in \mathbb{R}\}.$$

In der Tat rechnet man schnell nach, dass für alle $a, b \in \mathbb{R}$ gilt:

$$(a,0) + (b,0) = (a+b,0) \text{ und } (a,0) \cdot (b,0) = (ab - 0 \cdot 0, a \cdot 0 + 0 \cdot b) = (ab,0)$$

Das zeigt, dass die Abbildung $a \longmapsto (a,0)$ ein Isomorphismus von Körpern zwischen dem Körper $\mathbb{R}$ der reellen Zahlen und dem Unterkörper R von $\mathbb{C}$ ist.

Wir können also tatsächlich so tun, als ob die Menge $\mathbb{R}$ eine Teilmenge von $\mathbb{C}$ ist. Wir vereinfachen nun unsere Notation wie folgt:

Definition 6.58. *Die komplexe Zahl* $(0,1)$ *nennen wir die* imaginäre Einheit *und bezeichnen sie mit i. Anstelle von (a,b) mit $a, b \in \mathbb{R}$ schreiben wir $a + ib$. Dabei nennen wir a den Realteil der komplexen Zahl $a + ib$ und b den Imaginärteil.*

Es gilt

$$i^2 = (0,1) \cdot (0,1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) = -1.$$

Praktisch kann man nun mit komplexen Zahlen in der Form $a + ib$ wie folgt rechnen:

- Es dürfen die bekannten Rechenregeln in Körpern angewendet werden.
- Immer wenn wir einem Term i^2 begegnen, können wir diesen durch -1 ersetzen.
- Am Ende der Rechnung ordnen wir die Terme so, dass wir wieder einen Ausdruck der Form $c + id$ erhalten.

Beispiel 6.59. a) Es gilt $(a+ib) + (c+id) = a + c + i(b+d)$. Das entspricht genau unserer Definition von $(a,b) + (c,d)$.

b) Es gilt

$$(a+ib) \cdot (c+id) = ac + aid + ibc + i^2bd = ac - bd + i(ad + bc).$$

Das entspricht genau unserer Definition von $(a,b) \cdot (c,d)$.

c) Wir dividieren zwei komplexe Zahlen. Sei $c + id \neq 0$, d.h., wir nehmen an, dass c und d nicht beide 0 sind. Dann gilt

$$\frac{a+ib}{c+id} = \frac{(a+ib)\cdot(c-id)}{(c+id)\cdot(c-id)} = \frac{ac+bd+i(bc-ad)}{c^2+d^2}.$$

Insbesondere ist

$$\frac{1}{c+id} = \frac{c-id}{c^2+d^2}.$$

Definition 6.60. *Für eine komplexe Zahl $z = a + ib$ ist $\overline{z} := a - ib$ die zu z* konjugiert komplexe Zahl. *Der* Betrag *von z ist die Zahl $|z| := \sqrt{a^2+b^2}$.*

Satz 6.61. *a) Für jede komplexe Zahl z ist $|z|^2 = z \cdot \overline{z}$. Der Realteil von z ist die Zahl $\frac{1}{2}(z+\overline{z})$. Der Imaginärteil ist die Zahl $\frac{1}{2}(z-\overline{z})$.*

b) Für zwei komplexe Zahlen z_1, z_2 gilt

$$\overline{z_1+z_2} = \overline{z_1} + \overline{z_2}$$

sowie

$$\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$$

Beweis. a) Es gilt

$$(a+ib)\cdot(a-ib) = a^2 - i^2b^2 = a^2 + b^2 = |z|^2.$$

Die Aussage über den Real- und Imaginärteil von z ist noch einfacher nachzurechnen.

b) rechnet man ebenfalls leicht nach. □

Der Grund, warum die komplexen Zahlen eine so wichtige Rolle spielen, ist die Tatsache, dass wir in den komplexen Zahlen beliebige Wurzeln auch aus negativen Zahlen ziehen können. Warum das so ist, werden wir im nächsten Semester klären, wenn wir die trigonometrischen Funktionen sin und cos zur Verfügung haben.

Allerdings können die komplexen Zahlen die reellen Zahlen nicht ersetzen. In $\mathbb{C}$ ist nämlich $i^2 = -1$, während in einem angeordneten Körper Quadrate immer positiv sind. Die komplexen Zahlen tragen also keine Ordnung, die sie zu einem angeordneten Körper machen. In der Analysis ist es aber wichtig, dass wir mit einem angeordneten Körper wie zum Beispiel $\mathbb{R}$ arbeiten. Damit behalten die reellen Zahlen ihre Bedeutung, auch wenn wir in anderen Zusammenhängen lieber mit komplexen Zahlen rechnen.

KAPITEL 7

Polynome

§7.1 Polynomringe

Definition 7.1. *Ist K ein Körper, so bezeichnen wir einen Ausdruck der Form*

$$a_0X^0 + a_1X^1 + a_2X^2 + \cdots + a_nX^n\,,$$

wobei die Koeffizienten $a_0, \ldots, a_n$ aus K stammen und X eine Unbekannte ist, als Polynom *(in der Unbestimmten X) über K. Die Menge aller Polynome über K bezeichnen wir mit $K[X]$. Polynome der Form a_0 nennen wir* konstant. *Die Elemente von K identifizieren wir mit den konstanten Polynomen und fassen so K als Teilmenge von $K[X]$ auf.*

Bemerkung 7.2. In unserer Definition von Polynomen haben wir die verschiedenen Potenzen von X in aufsteigender Reihenfolge angegeben. Meistens werden die Potenzen jedoch in absteigender Reihenfolge angegeben. Statt

$$a_0X^0 + a_1X^1 + a_2X^2 + \cdots + a_nX^n$$

schreibt man also

$$a_nX^n + a_{n-1}X^{n-1} + \cdots + a_0X^0.$$

Die Potenz X^0 hat für alle X den Wert 1. Deshalb lässt man den Term X^0 normalerweise weg. Anstelle von X^1 schreibt man einfach X. Mit diesen Konventionen lautet das Polynom also

$$a_nX^n + \cdots + a_1X + a_0.$$

Ist für ein i der Koeffizient a_i gleich 0, so lässt man den Term a_iX^i weg. Bei negativen Koeffizienten zieht man das Minuszeichen mit dem vorhergehenden Pluszeichen zu einem Minuszeichen zusammen. Koeffizienten, die den Wert 1 haben, lässt man weg, falls es sich nicht um den Koeffizienten vor X^0 handelt. Anstelle von

$$1X^0 + (-5)X^1 + 0X^2 + 1X^3$$

schreibt man also

$$X^3 - 5X + 1.$$

Beispiel 7.3. (*a*) Aus der Schule sind Polynome mit reellen oder rationalen Koeffizienten bekannt, also Polynome über $\mathbb{R}$ oder $\mathbb{Q}$, wie das oben genannte Beispiel $X^3 - 5X + 1$. Streng genommen sind die Koeffizienten dieses Polynoms sogar ganzzahlig, sodass man von einem Polynom über $\mathbb{Z}$ sprechen könnte. Wir werden jedoch nur Polynome über Körpern betrachten.

(*b*) Wir kennen auch schon weitere Körper außer $\mathbb{Q}$, $\mathbb{R}$ und $\mathbb{C}$, nämlich die endlichen Körper $\mathbb{Z}/p\mathbb{Z}$ für Primzahlen p. So ist zum Beispiel $X^2 - X + 1$ ein Polynom über $\mathbb{Z}/2\mathbb{Z}$, wobei wir 1 für das neutrale Element der Multiplikation schreiben. Wir könnten dieses Polynom auch $X^2 - X + [1]_2$ oder $[1]_2X^2 + [-1]_2X^1 + [1]_2$ schreiben. Man beachte, dass für alle $a \in \mathbb{Z}/2\mathbb{Z}$ die Gleichung $a = -a$ gilt. Damit ist dieses Polynom identisch mit $X^2 + X + 1$. Man sieht, dass es in diesem Falle wichtig ist, festzulegen, über welchem Körper man das Polynom betrachtet.

(*c*) Wenn man Polynome über $\mathbb{Z}/p\mathbb{Z}$ betrachtet, wird es schnell lästig, die Koeffizienten in der Form $[n]_p$ zu schreiben. Deshalb schreiben wir in diesem Zusammenhang anstelle der Restklassen einfach die Standardrepräsentanten der Restklassen. Für das Polynom $X^3 + [2]_3X^2 + [-2]_3X + [1]_3$ über $\mathbb{Z}/3\mathbb{Z}$ schreiben wir also einfach X^3+2X^2+X+1. Die Schreibweise X^3+2X^2-2X+1 ist aber auch akzeptabel.

(*d*) Spezielle Polynome sind die sogenannten *Monome* X^n, $n \in \mathbb{N}_0$.

Wir haben schon intuitiv zwei Polynome gleich genannt, wenn sie dieselben Koeffizienten haben. An dieser Stelle müssen wir jedoch vorsichtig sein. Was ist zu Beispiel mit den Polynomen $0X^2 + X - 1$ und $X - 1$?

Definition 7.4. *Sei $p = a_0X^0 + \cdots + a_nX^n$ ein Polynom über einem Körper K. Der* Grad $\operatorname{grad}(p)$ *von p ist das größte $i \in \{0, \ldots, n\}$ mit $a_i \neq 0$, falls solch ein i existiert. Existiert kein i mit $a_i \neq 0$, so nennt man p das Nullpolynom und setzt $\operatorname{grad}(p) := -\infty$. Polynome vom Grad $\leqslant 0$ nennen wir* konstant.

Ist $\operatorname{grad}(p) \geqslant 0$, so nennt man den Koeffizienten $a_{\operatorname{grad}(p)}$ den Leitkoeffizienten *von p. Das Polynom p heißt normiert, falls der Leitkoeffizient 1 ist.*

Wir nennen zwei Polynome $p = a_0X^0 + \cdots + a_nX^n$ und $q = b_0X^0 + \cdots + b_mX^m$ über demselben Körper K gleich, wenn sie denselben Grad k haben und für alle $i \in \{0, \ldots, k\}$ die Koeffizienten a_i und b_i gleich sind.

Insbesondere sind also die Polynome $0X^2 + X - 1$ und $X - 1$ gleich. Beide Polynome haben den Grad 1 und die Koeffizienten vor X^1 und X^0 sind jeweils dieselben. Man beachte, dass es in diesem Beispiel egal ist, über welchem Körper man die Polynome betrachtet, solange es für beide Polynome derselbe Körper ist.

Als nächstes definieren wir Summen und Produkte von Polynomen.

Definition 7.5. *Seien $p = a_0X^0 + \cdots + a_nX^n$ und $q = b_0X^0 + \cdots + b_mX^m$ Polynome über demselben Körper K. Sei $k = \max(m, n)$. Für alle $i \in \mathbb{Z}$ mit $n < i \leqslant k$ sei $a_i := 0$. Für alle $j \in \mathbb{Z}$ mit $m < j \leqslant k$ sei $b_j := 0$. Dann gilt $p = a_0X^0 + \cdots + a_kX^k$ und*

$q = b_0X^0 + \cdots + b_kX^k$. Nun sei $p + q := (a_0 + b_0) + \cdots + (a_k + b_k)X^k$. Wir definieren die Summe zweier Polynome also „koeffizientenweise“.

Das Produkt von p und q definieren wir durch Ausmultiplizieren. Das Produkt $p \cdot q$ sei das Polynom $c_0 + \cdots + c_{n+m}X^{n+m}$ mit $c_i = a_0b_i + a_1b_{i-1} + \cdots + a_ib_0$.

Beispiel 7.6. Addition und Multiplikation von Polynomen über $\mathbb{Q}$ und $\mathbb{R}$ setzen wir als bekannt voraus.

(*a*) Wir betrachten Polynome über $\mathbb{Z}/5\mathbb{Z}$. Sei $p = X^3+3X^2+2$ und $q = 2X^2-X+4$. Dann ist

$$p + q = X^3 + (3 + 2)X^2 - X + (2 + 4) = X^3 + 4X + 1$$

und

$$\begin{aligned} p \cdot q &= (X^3 + 3X^2 + 2) \cdot (2X^2 - X + 4) \\ &= 2X^5 + (-1 + 3 \cdot 2)X^4 + (4 - 3)X^3 + (3 \cdot 4 + 2 \cdot 2)X^2 - 2X + 2 \cdot 4 \\ &= 2X^5 + X^3 + X^2 + 3X + 3. \end{aligned}$$

Insbesondere ist

$$\operatorname{grad}(p \cdot q) = \operatorname{grad}(p) + \operatorname{grad}(q).$$

Wie man leicht nachrechnet, gilt diese Gleichung für je zwei Polynome über demselben Körper.

(*b*) Wir betrachten wieder Polynome über $\mathbb{Z}/5\mathbb{Z}$. Sei $p = X^3 + 3X^2 + 2$ wie oben und $q = -X^3 + X^2 - 3$. Dann gilt

$$p + q = (1 - 1)X^3 + (3 + 1)X^2 + (2 - 3) = 4X^2 - 1 = 4X^2 + 4.$$

Insbesondere ist

$$\operatorname{grad}(p + q) < \operatorname{grad}(p), \operatorname{grad}(q).$$

Das ist aber ein Spezialfall. Sind p und q Polynome von verschiedenem Grad, so ist

$$\operatorname{grad}(p + q) = \max(\operatorname{grad}(p), \operatorname{grad}(q)).$$

Sind p und q Polynome vom selben Grad und ist der Leitkoeffizient von p nicht genau das additive Inverse des Leitkoeffizienten von q, so ist

$$\operatorname{grad}(p + q) = \operatorname{grad}(p) = \operatorname{grad}(q).$$

Satz 7.7. *Die Menge $K[X]$ zusammen mit den eben definierten Operationen $+$ und $\cdot$ für Polynome bildet einen Ring, in dem das Kommutativgesetz für $\cdot$ gilt. (Damit ist $K[X]$ ein* kommutativer Ring*.) Diesen Ring nennt man den* Polynomring *(in der Unbestimmten X) über K.*

Beweis. Die Axiome für Ringe und das Kommutativgesetz der Multiplikation rechnet man leicht nach. □

§7.2 Polynomdivision

Für Polynome können wir die Teilbarkeitsrelation wie für ganze Zahlen definieren.

Definition 7.8. *Seien p und q Polynome über einem Körper K. Wir sagen, dass p das Polynom q teilt, wenn es ein Polynom r über K gibt, sodass $q = p \cdot r$ gilt. In diesem Falle heißt q ein* Vielfaches *von p und wir schreiben $p \mid q$.*

Ein Polynom r ist ein gemeinsamer Teiler *von p und q, wenn r sowohl p als auch q teilt. Das Polynom r ist ein* größter gemeinsamer Teiler *von p und q, wenn r ein gemeinsamer Teiler von p und q von maximalem Grad ist.*

Beispiel 7.9. (*a*) Wir rechnen wieder über $\mathbb{Z}/5\mathbb{Z}$. Die Gleichung

$$(X^3 + 3X^2 + 2) \cdot (2X^2 - X + 4) = 2X^5 + X^3 + X^2 + 3X + 3,$$

zeigt, dass $X^3 + 3X^2 + 2$ und $2X^2 - X + 4$ Teiler von $2X^5 + X^3 + X^2 + 3X + 3$ sind.

(*b*) Wir rechnen über $\mathbb{R}$. Die Zahlen 2.5 und π, aufgefasst als konstante Polynome, werden beide von allen reellen Zahlen $\neq 0$ geteilt. Für jedes $a \in \mathbb{R} \setminus \{0\}$ gilt nämlich $2.5 = a \cdot \frac{2.5}{a}$ und $\pi = a \cdot \frac{\pi}{a}$. Für jedes Polynom $p \in \mathbb{R}[X]$ vom Grad $\geqslant 1$ und jedes $r \in \mathbb{R}[X]$ mit $r \neq 0$ ist $\operatorname{grad}(p \cdot r) \geqslant 1$ und damit $p \cdot r \neq 2.5$. Die Zahl 2.5 wird also nur von konstanten Polynomen geteilt, aber von allen von 0 verschiedenen konstanten Polynomen. Dasselbe gilt für π. Damit sind genau die konstanten Polynome $\neq 0$ größte gemeinsame Teiler von 2.5 und π. Insbesondere sind größte gemeinsame Teiler in Polynomringen im Allgemeinen nicht eindeutig bestimmt.

Wie im Falle von $\mathbb{Z}$ lassen sich größte gemeinsame Teiler in $K[X]$ mit dem euklidischen Algorithmus bestimmen. Dazu müssen wir zunächst die Division mit Rest von Polynomen einführen, die sogenannte *Polynomdivision*.

Satz 7.10. *Seien p und m Polynome über einem Körper K. Ist $m \neq 0$, so existieren Polynome q und r über K mit $p = q \cdot m + r$ und $\operatorname{grad}(r) < \operatorname{grad}(m)$.*

Beweis. Ist m konstant, also zum Beispiel $m = b_0 \in K$, so setzen wir

$$q := \frac{a_n}{b_0} X^n + \cdots + \frac{a_0}{b_0}$$

und $r := 0$. Dann gilt $p = q \cdot m + r$ und die Gradbedingung ist erfüllt.

Ist $\operatorname{grad}(m) \geqslant 1$, so beweisen wir den Satz durch vollständige Induktion über den Grad von p.

Induktionsanfang: Ist $\operatorname{grad}(p) < \operatorname{grad}(m)$, so setzen wir $q := 0$ und $r := p$. Dann gilt $p = q \cdot m + r$, wobei r die gewünschte Gradbedingung erfüllt.

Induktionsschritt: Sei nun der Grad von p mindestens so hoch wie der Grad von m.

Wir nehmen an, dass für alle Polynome p' mit $\mathrm{grad}(p') < \mathrm{grad}(p)$ Polynome q' und r' mit $p' = q' \cdot m + r'$ und $\mathrm{grad}(r') < \mathrm{grad}(m)$ existieren (Induktionsannahme).

Wir suchen Polynome q und r mit $p = q \cdot m + r$ und $\mathrm{grad}(r) < \mathrm{grad}(m)$.

Sei $n = \mathrm{grad}(p)$, $k = \mathrm{grad}(m)$, $p = a_n X^n + \cdots + a_0$ und $m = b_k X^k + \cdots + b_0$. Wir setzen

$$p' := p - \frac{a_n}{b_k} \cdot X^{n-k} \cdot m$$

und berechnen den Koeffizienten c_n von X^n in p'. $X^{n-k} \cdot m$ ist ein Polynom vom Grad $n - k + k = n$ mit dem Leitkoeffizienten b_k. Damit ist $c_n = a_n - \frac{a_n}{b_k} \cdot b_k = 0$. Also ist p' ein Polynom mit $\mathrm{grad}(p') < n = \mathrm{grad}(p)$.

Nach Induktionsannahme existieren Polynome q' und r' mit $p' = q' \cdot m + r'$ und $\mathrm{grad}(r') < \mathrm{grad}(m)$. Nach Wahl von p' gilt

$$p = \frac{a_n}{b_k} \cdot X^{n-k} \cdot m + p'.$$

Setzt man nun für p' den Ausdruck $q' \cdot m + r'$ ein, so ergibt sich

$$p = \frac{a_n}{b_k} \cdot X^{n-k} \cdot m + q' \cdot m + r' = \left(\frac{a_n}{b_k} \cdot X^{n-k} + q' \right) \cdot m + r'.$$

Wir setzen $r := r'$ und $q := \left(\frac{a_n}{b_k} \cdot X^{n-k} + q' \right)$. Nun gilt $p = q \cdot m + r$, wobei die Gradbedingung $\mathrm{grad}(r) < \mathrm{grad}(m)$ erfüllt ist. Das beendet den Induktionsschritt. □

Der Beweis von Satz 7.10 liefert ein rekursives Verfahren, mit dem sich der Quotient q und damit auch der Rest r bei Division von p durch m berechnen lässt. Wesentlicher Punkt dieser *Polynomdivision* ist die folgende Bemerkung.

Bemerkung 7.11. Sei $\mathrm{grad}(p) \geqslant \mathrm{grad}(m) \geqslant 1$. Im Beweis von Satz 7.10 haben wir gesehen, dass es Polynome q und r mit $\mathrm{grad}(r) < \mathrm{grad}(m)$ und $p = q \cdot m + r$ gibt, wobei q die Form $\frac{a_n}{b_k} \cdot X^{n-k} + q'$ hat. Dabei gilt $p' = q' \cdot m + r'$ für ein Polynom p' mit $\mathrm{grad}(p') < \mathrm{grad}(p)$. Also ist der Grad von q' kleiner als $n - k$, wobei n der Grad von p und k der Grad von m ist. Damit ist $\frac{a_n}{b_k}$ der Leitkoeffizient von q.

Außerdem ist der Rest r bei der Division von p durch m einfach das Polynom r', also der Rest bei der Division von p' durch m.

Wir beschreiben den Algorithmus zur Division von Polynomen, der sich aus dem Beweis von Satz 7.10 ergibt.

Polynomdivision. Seien zwei Polynome

$$p = a_n X^n + \cdots + a_0$$

und

$$m = b_k X^k + \cdots + b_0$$

über einem festen Körper K gegeben. Das Polynom m habe den Grad $k \geqslant 0$. Wir wollen Polynome q und r wie in Satz 7.10 bestimmen.

Ist $k = 0$, so ist p durch m teilbar und man erhält den Quotienten q, indem man jeden Koeffizienten von p durch $m \in K$ teilt. Der Rest ist in diesem Fall $r = 0$.

Nun nehmen wir an, dass $k \geqslant 1$ gilt. Wir halten p und m im Laufe der Berechnung fest und verändern die Variablen $\bar{p}$ und $\bar{n}$. Dabei seien $\bar{a}_{\bar{n}}, \dots, \bar{a}_0$ immer die Koeffizienten des Polynoms $\bar{p}$. Die Koeffizienten $c_{n-k}, \dots, c_0$ des Quotienten q werden nach und nach berechnet, falls $n \geqslant k$ ist.

(1) Setze $\bar{n} := n$ und $\bar{p} := p$.

(2) Ist $\bar{n} < k$, so ist $r = \bar{p}$ der Rest bei der Division von p durch m. Ist $n \geqslant k$, so ist $q = c_{n-k}X^{n-k} + \cdots + c_0$ der Quotient bei der Division von p durch m. Ist $n < k$, so lautet der Quotient $q = 0$ und es wurden auch keine c_i berechnet. Die Berechnung endet hier.

(3) Ist $\bar{n} \geqslant k$, so speichere den Koeffizienten
$$c_{\bar{n}-k} := \frac{\bar{a}_{\bar{n}}}{b_k}$$
und setze
$$\bar{p} := \bar{p} - c_{\bar{n}-k} \cdot X^{\bar{n}-k} \cdot m.$$

(4) Ist $\bar{p}$ das Nullpolynom, so setze $\bar{n} := -\infty$ und fahre mit Schritt (2) fort.

(5) Ist $\bar{p} \neq 0$, so setze $\bar{n} := \bar{n} - 1$ und fahre mit Schritt (2) fort.

Bemerkung 7.12. Seien p und m wie im Algorithmus zur Polynomdivision. Wir nehmen an, dass $n \geqslant k \geqslant 1$ ist. Dann kann man die Berechnung des Algorithmus wie folgt aufschreiben: Wir starten mit der Zeile
$$(a_nX^n + \cdots + a_0) = (b_kX^k + \cdots + b_0) \cdot \Big(\dots \Big)$$
Zunächst berechnen wir den Koeffizienten $c_{n-k} = \frac{a_n}{b_k}$ und tragen ihn zusammen mit der passenden Potenz X^{n-k} auf der rechten Seite in der Klammer ein.

Das liefert
$$(a_nX^n + \cdots + a_0) = (b_kX^k + \cdots + b_0) \cdot \left(\frac{a_n}{b_k}X^{n-k} + \dots \right)$$
Als nächstes multiplizieren wir m mit $\frac{a_n}{b_k}X^{n-k}$. Das liefert ein Polynom vom Grad n, das wir unter das Polynom p schreiben. Als nächstes ziehen wir $\frac{a_n}{b_k}X^{n-k} \cdot m$ von p ab und schreiben das Ergebnis ebenfalls darunter. Die dritte Zeile lautet nun
$$0 + \left(a_{n-1} - b_{k-1}\frac{a_n}{b_k} \right) X^{n-1} + \dots$$
Wir setzen dann die Polynomdivision mit dem Polynom in der dritten Zeile fort, und zwar solange, bis der Grad der letzten Differenz kleiner als der Grad von m geworden ist. Dabei schreiben wir die neu berechneten Terme c_iX^i von q oben rechts hinter den Ausdruck $\frac{a_n}{b_k}X^{n-k}$. Am Schluss steht das gesamte Polynom q auf der rechten Seite der Gleichung zwischen den Klammern und die Differenz in der letzten Zeile ist der Rest

bei der Division von p durch m. Damit das Gleichheitszeichen gerechtfertigt ist, tragen wir am Schluss den Rest r noch als zusätzlichen Summanden in die oberste Zeile.

Es ist übrigens nicht nötig, die Differenzen immer vollständig aufzuschreiben, da alle bis auf die ersten $k-1$ Summanden mit den entsprechenden Summanden von p übereinstimmen.

Beispiel 7.13. Wir rechnen über $\mathbb{Q}$.

(a) Sei $p = X^3 - 2X^2 + 4X + 7$ und $m = X + 1$. Die Polynomdivision sieht dann wie folgt aus:

$$\begin{array}{l}
X^3 - 2X^2 + 4X + 7 = \big(X+1\big)\big(X^2 - 3X + 7\big) + 0 \\
\underline{-X^3 \;\; - X^2} \\
\qquad\; -3X^2 + 4X \\
\qquad\quad \underline{3X^2 + 3X} \\
\qquad\qquad\quad 7X + 7 \\
\qquad\qquad\; \underline{-7X - 7} \\
\qquad\qquad\qquad\quad 0
\end{array}$$

In diesem Fall ergibt sich der Rest 0. Insbesondere ist p durch m teilbar.

(b) Sei $p = X^3 - 2X^2 + 5X + 6$ und $m = X^2 - X + 1$. Die Polynomdivision sieht dann wie folgt aus:

$$\begin{array}{l}
X^3 - 2X^2 + 5X + 6 = \big(X^2 - X + 1\big)\big(X - 1\big) + 3X + 7 \\
\underline{-X^3 \;\; + X^2 \;\; - X} \\
\qquad\; -X^2 + 4X + 6 \\
\qquad\quad \underline{X^2 \;\; - X + 1} \\
\qquad\qquad\quad 3X + 7
\end{array}$$

Hier ist der Quotient $X - 1$ und der Rest $3X + 7$.

Wie bei ganzen Zahlen kann man größte gemeinsame Teiler von Polynomen mit Hilfe des euklidischen Algorithmus berechnen. Dabei spielt der Grad die Rolle des Betrages bei den ganzen Zahlen. Ein Unterschied zur Situation bei den ganzen Zahlen besteht darin, dass es durchaus passieren kann, dass zwei Polynomen denselben Grad haben, ohne dass die beiden Polynomen einander teilen. In diesem Falle ist es egal, ob man zunächst das eine Polynom durch das andere teilt oder umgekehrt.

Beispiel 7.14. Wir wollen einen größten gemeinsamen Teiler der Polynome

$$p = X^3 - 3X^2 + 5X - 3$$

und

$$q = X^3 - 1$$

bestimmen. Eigentlich müssten wir beim euklidischen Algorithmus zunächst das Polynom vom höheren Grad durch das vom niedrigeren Grad teilen. Die beiden Grade sind aber gleich. Deshalb ist es egal, ob wir zunächst p durch q teilen oder umgekehrt. Wir starten mit der Division von p durch q.

$$\begin{array}{lll}
X^3 - 3X^2 + 5X - 3 & = & \big(X^3 - 1\big)1 - 3X^2 + 5X - 2 \\
\underline{-X^3 \qquad\qquad\qquad\quad + 1} & & \\
\qquad -3X^2 + 5X - 2 & &
\end{array}$$

Der Rest ist also $-3X^2 + 5X - 2$. Also dividieren wir im nächsten Schritt q durch $-3X^2 + 5X - 2$.

$$\begin{array}{lll}
X^3 \qquad\qquad\qquad\qquad - 1 & = & \big(-3X^2 + 5X - 2\big)\big(-\frac{1}{3}X - \frac{5}{9}\big) + \frac{19}{9}X - \frac{19}{9} \\
\underline{-X^3 + \frac{5}{3}X^2 - \frac{2}{3}X} & & \\
\qquad \frac{5}{3}X^2 - \frac{2}{3}X - 1 & & \\
\qquad \underline{-\frac{5}{3}X^2 + \frac{25}{9}X - \frac{10}{9}} & & \\
\qquad\qquad \frac{19}{9}X - \frac{19}{9} & &
\end{array}$$

Das liefert den Rest $\frac{19}{9}(X - 1)$. Man beachte, dass das Polynom $\frac{19}{9}(X - 1)$ genau dieselben Teiler wie $X - 1$ hat und auch genau dieselben Polynome teilt. Damit können wir im nächsten Schritt der Einfachheit halber durch $X - 1$ anstelle von $\frac{19}{9}(X - 1)$ teilen.

$$\begin{array}{lll}
-3X^2 + 5X - 2 & = & \big(X - 1\big)\big(-3X + 2\big) + 0 \\
\underline{\;3X^2 - 3X\;} & & \\
\qquad 2X - 2 & & \\
\qquad \underline{-2X + 2} & & \\
\qquad\qquad 0 & &
\end{array}$$

Der Rest ist dabei 0. Also ist $X - 1$ ein größter gemeinsamer Teiler von p und q.

§7.3 Polynomfunktionen und Nullstellen von Polynomen

Definition 7.15. *Sei K ein Körper und $p = a_0 + \cdots + a_nX^n \in K[X]$. Dann ist die Funktion*

$$f_p\colon K \longrightarrow K \quad \textit{mit} \quad x \longmapsto a_0 + \cdots + a_nx^n$$

die zu p gehörige Polynomfunktion.

Man berechnet also f_p, indem man ein gegebenes Körperelement x (nicht zu verwechseln mit der Unbestimmten X) für X in das Polynom einsetzt.

Beispiel 7.16. (*a*) Sei $p = 2X^2 - 3X + 7 \in \mathbb{Q}[X]$. Dann ist

$$f_p(3) = 2 \cdot 3^2 - 3 \cdot 3 + 7 = 18 - 9 + 7 = 16.$$

(*b*) Sei $p = X^3 - 2X + 1 \in (\mathbb{Z}/3\mathbb{Z})[X]$. Dann ist

$$f_p(2) = 2^3 - 2 \cdot 2 + 1 = 2 - 1 + 1 = 2.$$

Wir schreiben wieder Standardvertreter anstelle von Restklassen und rechnen modulo 3.

Der Grund, weshalb wir zwischen Polynomen und den zugerörigen Polynomfunktionen unterscheiden, ist, dass es über einem endlichen Körper K zwar unendlich viele Polynome gibt, aber nur endlich viele Polynomfunktionen. Es gibt also verschiedene Polynome p und q über K, deren Polynomfunktionen übereinstimmen.

Beispiel 7.17. Sei $p = X^4 + X + 2$ und $q = X^3 + X^2 + 2$, wobei wir p und q als Polynome über $\mathbb{Z}/3\mathbb{Z}$ auffassen. Dann ist $p \neq q$, und zwar schon deshalb, weil p und q unterschiedlichen Grad haben. In $\mathbb{Z}/3\mathbb{Z}$ gilt aber

$$f_p(0) = 0 + 0 + 2 = 2\,, \quad f_p(1) = 1 + 1 + 2 = 1\,, \quad f_p(2) = 1 + 2 + 2 = 2$$

und

$$f_q(0) = 0 + 0 + 2 = 2\,, \quad f_q(1) = 1 + 1 + 2 = 1\,, \quad f_q(2) = 2 + 1 + 2 = 2\,.$$

Damit sind die Polynomfunktionen f_p und f_q gleich.

Ist $p \in K[X]$ und $x \in K$, so schreibt man in der Praxis anstelle von $f_p(x)$ eher $p(x)$. Für ein Körperelement x steht $p(x)$ also für das Körperelement, das man erhält, wenn man für die Unbestimmte X das Körperelement x in das Polynom einsetzt.

Definition 7.18. *Sei K ein Körper und $p \in K[X]$. Dann heißt $a \in K$ eine* Nullstelle *von p, falls $p(a) = 0$ ist.*

Satz 7.19. *Ein Körperelement $a \in K$ ist genau dann eine Nullstelle von $p \in K[X]$, wenn $X - a$ ein Teiler von p ist.*

Beweis. Angenommen, $X - a$ teilt p. Dann existiert $q \in K[X]$ mit $p = q \cdot (X - a)$. Es gilt

$$p(a) = q(a) \cdot (a - a) = q(a) \cdot 0 = 0.$$

Also ist $X - a$ eine Nulsstelle von p.

Sei umgekehrt $p(a) = 0$. Nach Satz 7.10 existieren Polynome $q, r \in K[X]$ mit $p = q \cdot (X - a) + r$ und $\operatorname{grad}(r) < \operatorname{grad}(X - a) = 1$. Das Polynom r ist also konstant. Es gilt

$$0 = p(a) = q(a) \cdot (a - a) + r = r$$

und damit $p = q \cdot (X - a)$. Damit teilt $(X - a)$ das Polynom p. □

Korollar 7.20. *Ein Polynom $p \in K[X]$ vom Grad $n > 0$ hat höchstens n verschiedene Nullstellen.*

Beweis. Wir zeigen das Korollar durch Induktion über n.
Induktionsanfang: Sei $n = 1$. Dann ist p von der Form $a_1X + a_0$ mit $a_0, a_1 \in K$ und $a_1 \neq 0$. Sei $x \in K$ mit $p(x) = 0$. Dann gilt $a_1x + a_0 = 0$ und damit $x = -a_0 \cdot a_1^{-1}$. Insbesondere hat p genau eine Nullstelle, nämlich $-a_0a_1^{-1}$.
Induktionsschritt: Sei $n \in \mathbb{N}$. Angenommen, jedes Polynom vom Grad n hat höchstens n verschiedene Nullstellen. Sei $p \in K[X]$ ein Polynom vom Grad $n + 1$ und $a \in K$ eine Nullstelle von p. Nach Satz 7.19 existiert $q \in K[X]$ mit $p = q \cdot (X - a)$. Sei $b \in K$ eine weitere, also von a verschiedene Nullstelle von p.

Dann gilt $0 = p(b) = q(b) \cdot (b - a)$. Wegen $b \neq a$ ist $b - a \neq 0$. Also ist $q(b) = 0$. Jede von a verschiedene Nullstelle von p ist also eine Nullstelle von q. Das Polynom q hat den Grad n. Nach Induktionsannahme hat q aber höchstens n verschiedene Nullstellen. Damit hat p höchstens n verschiedene Nullstellen, die von a verschieden sind. Also hat p höchstens $n + 1$ verschiedene Nullstellen. □

Der Beweis dieses Korollars liefert ein rekursives Verfahren, alle Nullstellen eines Polynoms zu bestimmen, wenn man in der Lage ist, einzelne Nullstellen zu bestimmen:

Sei $p \in K[X]$ ein Polynom vom Grad $n > 0$. Bestimme eine Nullstelle a_1 von p und teile p durch $(X - a_1)$. Wiederhole das Verfahren mit $p/(X - a_1)$. Iteriere das Verfahren solange, wie der Grad des Polynoms > 0 ist.

Um Nullstellen von Polynomen zweiten Grades über $\mathbb{R}$ zu bestimmen, gibt es die bekannte p-q-Formel: Das Polynom $X^2 + pX + q$ hat die Nullstellen

$$x_1 = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}$$

und

$$x_2 = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q},$$

falls die *Diskriminante* $\frac{p^2}{4} - q$ nicht negativ ist. Ist $\frac{p^2}{4} - q < 0$, so hat $X^2 + pX + q$ keine reellen Nullstellen.
Herleitung der p-q-Formel: Gegeben sei eine quadratische Gleichung der Form

$$X^2 + pX + q = 0.$$

Diese Gleichung lässt sich nicht einfach nach X auflösen. Eine Gleichung der Form

$$(X + a)^2 = b$$

lässt sich allerdings einfach nach X auflösen:

Aus $(X + a)^2 = b$ folgt $b \geqslant 0$ und $X + a = \pm\sqrt{b}$. Ist $(X + a)^2 = b$ genau dann lösbar, wenn $b \geqslant 0$ gilt, und die Lösungen sind die Zahlen $x_{1,2} = -a \pm \sqrt{b}$.

Die Gleichung $X^2 + pX + q = 0$ können wir aber auf die Form $(X + a)^2 = b$ bringen:

$$\begin{aligned} X^2 + pX + q &= 0 \\ X^2 + 2\frac{p}{2}X + \left(\frac{p}{2}\right)^2 - \left(\frac{p}{2}\right)^2 + q &= 0 \\ X^2 + 2\frac{p}{2}X + \left(\frac{p}{2}\right)^2 &= \left(\frac{p}{2}\right)^2 - q \\ \left(X + \frac{p}{2}\right)^2 &= \left(\frac{p}{2}\right)^2 - q \end{aligned}$$

Setzt man also $a := \frac{p}{2}$ und $b = \left(\frac{p}{2}\right)^2 - q = \frac{p^2}{4} - q$, so hat man die Gleichung $X^2 + pX + q = 0$ in die Form $(X + a)^2 = b$ überführt.

Damit ist $X^2 + pX + q = 0$ genau dann lösbar, wenn $\left(\frac{p}{2}\right)^2 - q \geqslant 0$ gilt. In diesem Falle lauten die Lösungen

$$x_{1,2} = -a \pm \sqrt{b} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}.$$

Das erklärt die Gültigkeit der p-q-Formel.

Indem man ein von 0 verschiedenes Polynom durch seinen Leitkoeffizienten teilt, kann man es normieren, ohne die Nullstellen zu verändern. Damit löst die p-q-Formel das Problem des Findens von Nullstellen von Polynomen vom Grad 2 über $\mathbb{R}$. Nullstellen von Polynomen vom Grad 1 lassen sich direkt durch Auflösen einer Gleichung mittels Äquivalenzumformungen finden. Für Polynome 3. und 4. Grades über $\mathbb{R}$ gibt es auch Formeln, die aber zu umfangreich sind, um sie hier zu präsentieren. Man kann beweisen, dass es zur Berechnung von Nullstellen von Polynomen 5. Grades über $\mathbb{R}$ keine allgemeinen Formeln mehr gibt. Allerdings kann man mit Hilfe numerischer Verfahren immer noch Näherungslösungen für Gleichungen der Form $p(x) = 0$ finden.

Hilfreich ist allerdings folgender Satz, der auf Gauß zurückgeht.

Satz 7.21. *Sei $p = X^n + a_{n-1}X^{n-1} + \cdots + a_0$ ein normiertes Polynom vom Grad $n > 0$ mit ganzzahligen Koeffizienten. Dann ist jede Nullstelle $b \in \mathbb{Q}$ von p eine ganze Zahl, die a_0 teilt.*

Beweis. Sei $b \in \mathbb{Q} \setminus \{0\}$ eine Nullstelle von p und $b = \frac{y}{z}$ für teilerfremde ganze Zahlen y und z mit $y \neq 0$ and $z \geqslant 1$. Wir zeigen zuerst $z = 1$.

Da $b = y/z$ eine Nullstelle von p ist, gilt

$$0 = p(y/z) = \left(\frac{y}{z}\right)^n + a_{n-1} \cdot \left(\frac{y}{z}\right)^{n-1} + \cdots + a_1 \cdot \left(\frac{y}{z}\right) + a_0\,. \tag{7.1}$$

Wir multiplizieren die Gleichung mit z^n und stellen nach y^n um und erhalten

$$y^n = z \cdot \big(- a_{n-1}y^{n-1} - \cdots - a_1yz^{n-2} - a_0z^{n-1}\big)\,.$$

Da alle Koeffizienten $a_0, \ldots, a_{n-1}$ sowie y und z ganzzahlig sind, ist die rechte Seite ein ganzzahliges Vielfaches von z. Somit muss y^n ein ganzzahliges Vielfaches von z sein.

Da $y \neq 0$ und $z \geqslant 1$ teilerfremd sind, kann z nur 1 sein. Insbesondere ist $b = y$ also ganzzahlig.

Es bleibt zu zeigen, dass $b = y$ ein ganzzahliger Teiler von a_0 ist. Ausgangspunkt ist wieder (7.1). Da wir aber bereits wissen, dass $z = 1$ ist und somit $b = y \neq 0$ ist, erhalten wir

$$0 = b^n + a_{n-1}b^{n-1} + \cdots + a_1 b + a_0\,.$$

Diesmal stellen wir nach a_0 um und Klammern b aus. Somit gilt

$$a_0 = b\big(- b^{n-1} - a_{n-1}b^{n-2} - \cdots - a_2 b - a_1\big)\,.$$

Nun folgt aus der Ganzzahligkeit von $b = y$ und $a_{n-1}, \ldots, a_0$, dass die rechte Seite ein ganzahliges Vielfaches von b ist und somit folgt auch dass a_0 ein ganzzahliges Vielfaches von b ist. $\square$

Der Satz zeigt, dass die rationalen Nullstellen eines normierten Polynoms mit ganzzahligen Koeffizienten ganzzahlig sind und man diese einfach durch Ausprobieren der Teiler des konstanten Summanden des Polynoms finden kann. Es ist aber zu beachten, dass der Satz reelle Nullstellen nicht ausschließt. Dies wäre auch falsch, wie das normierter Polynom $X^2 - 2$ mit den reellen Nullstellen $\pm\sqrt{2}$ zeigt.

Beispiel 7.22. Sei $p = X^3 - 6X^2 + 11X - 6 \in \mathbb{Q}[X]$. Wir wollen die rationalen Nullstellen von p finden. Nach Satz 7.21 sind die rationalen Nullstellen in Wirklichkeit ganze Zahlen, die -6 teilen. Die Kandidaten sind also $-6, -3, -2, -1, 1, 2, 3, 6$. Als erstes probieren wir 1 aus, weil in diesem Fall die Rechnung am einfachsten ist.

Es gilt $p(1) = 1 - 6 + 11 - 6 = 0$. Damit haben wir die erste Nullstelle $a_1 = 1$ von p gefunden. Nun teilen wir p durch $X - 1$.

$$\begin{array}{l}
X^3 - 6X^2 + 11X - 6 = \big(X-1\big)\big(X^2 - 5X + 6\big) + 0 \\
\underline{-X^3 \;\; + X^2} \\
\qquad -5X^2 + 11X \\
\qquad \underline{\;\;\; 5X^2 \;\; - 5X} \\
\qquad\qquad\qquad 6X - 6 \\
\qquad\qquad\quad \underline{-6X + 6} \\
\qquad\qquad\qquad\qquad 0
\end{array}$$

Die weiteren Nullstellen von p sind Nullstellen des Quotienten $q = X^2 - 5X + 6$. Da q ein Polynom zweiten Grades ist, können wir die p-q-Formel benutzen, um die Nullstellen zu finden. Die Diskriminante ist in diesem Falle

$$D = \frac{25}{4} - 6 = \frac{25}{4} - \frac{24}{4} = \frac{1}{4} = \left(\frac{1}{2}\right)^2.$$

Es gilt

$$a_2 = -\frac{-5}{2} + \sqrt{D} = \frac{5}{2} + \frac{1}{2} = 3$$

und

$$a_3 = -\frac{-5}{2} - \sqrt{D} = \frac{5}{2} - \frac{1}{2} = 2.$$

Damit haben wir alle Nullstellen von p gefunden.

KAPITEL 8

Vektoren und Matrizen

§8.1 Vektorrechnung

Definition 8.1. *Sei K ein Körper. Wir nennen die Elemente von K^n* Vektoren. *Die Summe zweier Vektoren $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in K^n$ definieren wir komponentenweise. Es sei*

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) := (a_1 + b_1, \ldots, a_n + b_n).$$

Außerdem definieren wir die Multiplikation von Vektoren mit Elementen des Körpers K. Sei $\alpha \in K$ und $v = (a_1, \ldots, a_n) \in K^n$. Dann sei

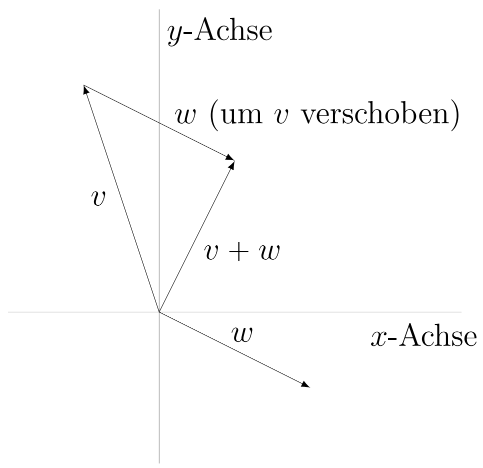
$$\alpha v := (\alpha a_1, \ldots, \alpha a_n).$$

In diesem Zusammenhang nennt man α einen Skalar *mit dem der Vektor v* skaliert *wird.*

Beispiel 8.2. Wir stellen uns Vektoren in $\mathbb{R}^2$ als Punkte in der Anschauungsebene oder als Pfeile vom Nullpunkt zu einem Punkt in der Ebene vor. Die Summe von Vektoren lässt sich dann geometrisch als Aneinanderreihung von Pfeilen interpretieren.

Entsprechendes gilt in $\mathbb{R}^3$ oder ganz allgemein in $\mathbb{R}^n$, wobei unsere Anschauung im Falle $n > 3$ natürlich sehr herausgefordert wird.

Sei $v := (-1, 3)$ und $w := (2, -1)$.



Sei $\alpha = 2.5$ und $v := (-1, 3)$. Dann ist $\alpha v = (-2.5, 7.5)$.

Die Multiplikation mit dem Skalar α entspricht einer Streckung um den Faktor α.

Satz 8.3. *Sei K ein Körper und $n \in \mathbb{N}$.*

a) $(K^n, +)$ *ist eine abelsche Gruppe. Das neutrale Element der Addition ist der Vektor* $(0, \dots, 0)$*, den wir den* Nullpunkt *nennen.*

b) Für alle $v, w \in K^n$ *und alle* $\alpha, \beta \in K$ *gilt:*

(1) $\alpha(v + w) = \alpha v + \alpha w$

(2) $(\alpha + \beta)v = \alpha v + \beta v$

(3) $(\alpha \cdot \beta)v = \alpha(\beta v)$

(4) $1v = v$.

Eine Struktur der Form K^n mit der Operation $+$ und der Multiplikation mit Skalaren ist ein *Vektorraum*. Wir werden Vektorräume im zweiten Teil der Vorlesung studieren.

Beispiel 8.4. Wir betrachten wieder $\mathbb{R}^2$ und $v = (-1, 3)$. Sei

$$U = \{\alpha v : \alpha \in \mathbb{R}\}.$$

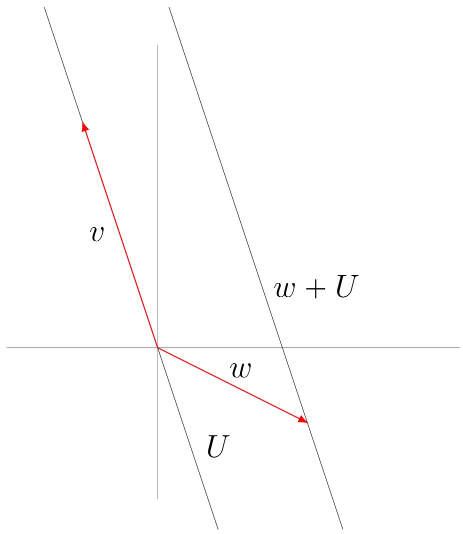
Dann ist U eine Untergruppe von $(\mathbb{R}^2, +)$.

Für $\alpha, \beta \in \mathbb{R}$ gilt nämlich nach Satz 8.3 $\alpha v - \beta v = (\alpha + (-\beta))v \in U$.

Nach unserem Kriterium für Untergruppen folgt nun, dass U tatsächlich eine Untergruppe von $(\mathbb{R}^2, +)$ ist.

Die Menge U der skalaren Vielfachen von v ist einfach die Gerade durch den 0-Punkt, die den Vektor v enthält. Die Nebenklassen von U in $\mathbb{R}^2$ sind die Geraden in $\mathbb{R}^2$, die zu der Geraden U parallel sind.

Sei $v := (-1, 3)$ und $w := (2, -1)$ und $U := \{\alpha v : \alpha \in \mathbb{R}\}$.



Wir definieren noch eine weitere Operation zwischen Vektoren in $\mathbb{R}^n$.

Definition 8.5. *Seien* $(a_1, \dots, a_n), (b_1, \dots, b_n) \in \mathbb{R}^n$*. Dann ist das (*Standard-*)* Skalarprodukt *von* $v = (a_1, \dots, a_n)$ *und* $w = (b_1, \dots, b_n)$ *das Körperelement*

$$\langle v, w \rangle := a_1 b_1 + \cdots + a_n b_n.$$

Die Bezeichnungen „Skalarprodukt“ und „Multiplikation mit einem Skalar“ geben leicht Anlass zur Verwirrung. Es handelt sich um die Standardbezeichnungen und man muss aufpassen, dass man sich immer genau klarmacht, worum es geht.

Beispiel 8.6. Wir rechnen wieder über dem Körper $\mathbb{R}$. Sei $v = (1, 2, 3)$ und $w = (-1, 2, 1)$. Dann gilt

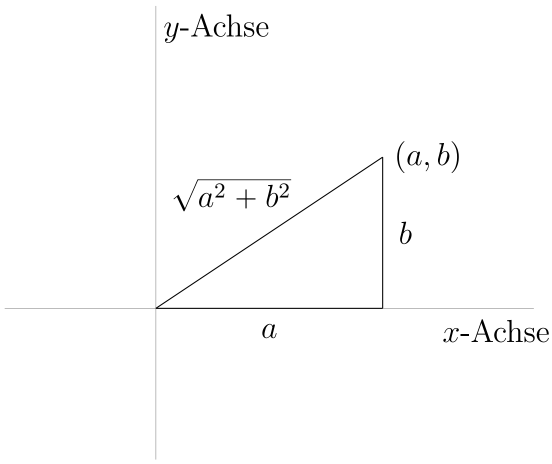
$$\langle v, w\rangle = 1 \cdot (-1) + 2 \cdot 2 + 3 \cdot 1 = -1 + 4 + 3 = 6.$$

Man erinnere sich an den Satz von Pythagoras:

In einem rechtwinkligen Dreieck, in dem die Längen der Katheten, also der Seiten, die am rechten Winkel anliegen, a und b sind, gilt für die Länge c der Hypothenuse, also der Seite, die dem rechten Winkel genüber liegt, die Gleichung

$$a^2 + b^2 = c^2.$$

Insbesondere ist der Abstand des Punktes $(a, b) \in \mathbb{R}^2$ vom Nullpunkt genau $\sqrt{a^2 + b^2} = \sqrt{\langle(a, b), (a, b)\rangle}$.



In höheren Dimensionen gilt das Entsprechende. Daher nennen wir für einen Vektor $v \in \mathbb{R}^n$ die Zahl $|v| = \sqrt{\langle v, v\rangle}$ den *Betrag* von v. Der Betrag von v ist nichts anderes als der Abstand von v vom 0-Punkt. Der Betrag $|\lambda|$ einer reellen Zahl λ ist der Wert den man erhält, wenn man das Vorzeichen von λ weglässt. So ist $|-5| = 5$, $|2.5| = 2.5$ und $|0| = 0$.

Der folgende Satz fasst die Eigenschaften des Standardsskalarprodukts und des Betrages zusammen.

Satz 8.7. *a) Sei $n \in \mathbb{N}$. Dann gelten folgende Aussagen für alle $\alpha \in \mathbb{R}$ und alle $u, v, w \in \mathbb{R}^n$:*

(1) $\langle v, w\rangle = \langle w, v\rangle$
(2) $\langle \alpha v, w\rangle = \alpha\langle v, w\rangle$
(3) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$

b) Für alle $n \in \mathbb{N}$, alle $v, w \in \mathbb{R}^n$ und alle $\lambda \in \mathbb{R}$ gelten die folgenden Aussagen:

(1) $|v| \geqslant 0$

(2) $|v| = 0 \Leftrightarrow v = (0, \dots, 0)$

(3) $|\lambda v| = |\lambda| \cdot |v|$

(4) $|v + w| \leqslant |v| + |w|$ *(Dreiecksungleichung)*

§8.2 Matrizenringe

Definition 8.8. *Seien $m, n \in \mathbb{N}$ und sei K ein Körper. Eine $m \times n$-*Matrix *über K ist ein rechteckiges Zahlenschema der Form*

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

wobei die a_{ij} Elemente von K sind.

Wir schreiben eine solche Matrix kürzer $(a_{ij})_{1 \leqslant i \leqslant m \wedge 1 \leqslant j \leqslant n}$ oder auch einfach (a_{ij}), wenn die Dimension $m \times n$ der Matrix klar ist.

In einer Matrix $(a_{ij})_{1 \leqslant i \leqslant m \wedge 1 \leqslant j \leqslant n}$ nennen wir $(a_{i1}, \dots, a_{in})$ die i-te Zeile und

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

die j-te Spalte.

Die Menge der $m \times n$-Matrizen über dem Körper K bezeichnen wir mit $K^{m \times n}$.

Definition 8.9. *Für zwei $m \times n$-Matrizen $A = (a_{ij})$ und $B = (b_{ij})$ über einem Körper K sei $A + B$ die Matrix $(a_{ij} + b_{ij})$. Der Eintrag in der i-ten Zeile und der j-ten Spalte der Matrix $A + B$ lautet also $a_{ij} + b_{ij}$. Die $m \times n$-Matrix, deren Einträge alle 0 sind, nennen wir eine Nullmatrix.*

Für $\alpha \in K$ und $A = (a_{ij}) \in K^{m \times n}$ sei

$$\alpha A := (\alpha a_{ij}).$$

Wie im Falle von K^n sieht man schnell, dass $(K^{m \times n}, +)$ eine abelsche Gruppe ist. Neben der Addition von Matrizen und der Multiplikation von Matrizen gibt es eine weitere Verknüpfung von Matrizen, die fast noch wichtiger ist als die beiden schon genannten Operationen, nämlich die Matrizenmultiplikation.

Definition 8.10. *Sei K ein Körper und seien $\ell, m, n \in \mathbb{N}$. Weiter sei $A = (a_{ij}) \in K^{\ell \times m}$ und $B = (b_{jk}) \in K^{m \times n}$. Dann ist $AB = A \cdot B$ die $\ell \times n$-Matrix $C = (c_{ik})$, deren Eintrag c_{ik} das Körperelement*

$$a_{i1} b_{1k} + \dots + a_{im} b_{mk},$$

also das Skalarprodukt der i-ten Zeile von A mit der k-ten Spalte von B, ist. Es gilt also

$$AB = \left(\sum_{j=1}^{m} a_{ij}b_{jk}\right)_{1\leqslant i\leqslant \ell \wedge 1\leqslant k\leqslant n}.$$

Eine wichtige, nichttriviale Eigenschaft der Matrizenmultiplikation ist die Assoziativität.

Satz 8.11. *Sei K ein Körper und seien $k, \ell, m, n \in \mathbb{N}$. Sind $A \in K^{k\times \ell}$, $B \in K^{\ell\times m}$ und $C \in K^{m\times n}$, so gilt*

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

Betrachtet man $n \times n$-Matrizen für ein festes n, so kann man die Matrizen in beliebiger Reihenfolge multiplizieren.

Satz 8.12. *Sei $n \in \mathbb{N}$ und sei K ein Körper. Dann ist $(K^{n\times n}, +, \cdot)$ ein Ring, der* Ring der $n \times n$-Matrizen über K.

Das neutrale Element bezüglich der Multiplikation in $K^{n\times n}$ ist die *Einheitsmatrix*

$$E_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

bei der auf der Diagonalen Einsen stehen und sonst nur Nullen.

Die Einheitengruppe des Matrizenringes $K^{n\times n}$ besteht aus den *invertierbaren Matrizen*. Der Matrizenring $K^{n\times n}$ ist für $n > 1$ nicht kommutativ.

Matrizen und ihre Multiplikation spielen eine wesentliche Rolle beim Lösen linearer Gleichungssystem und damit zum Beispiel auch in der Optimierung oder für bildgebende Verfahren. In der Graphentheorie kann man Matrizenmultiplikation zum berechnen von Wegen in Graphen einsetzen. Matrizen spielen auch eine wesentliche Rolle im Page-Rank-Algorithmus mit dem zum Beispiel Google die Reihenfolge der Suchergebnisse festlegt.

Notation

$\wp(M)$: Potenzmenge $\{A\colon A \subseteq M\}$ von M

$|M|$: Anzahl der Elemente einer endlichen Menge M, aber auch als $|a|$ die Ordnung eines Gruppenelementes a

$\mathbb{N}$: natürliche Zahlen $\{1, 2, 3, \dots\}$

$\mathbb{N}_0$: natürliche Zahlen mit Null $\{0, 1, 2, 3, \dots\}$

$[n]$: ersten n natürliche Zahlen $\{1, 2, 3, \dots, n\}$

$\mathbb{Z}$: ganze Zahlen $\{\dots, -2, -1, 0, 1, 2, \dots\}$

$\mathbb{Z}/m\mathbb{Z}$: Restklassenring $\big\{[0]_m, [1]_m, \dots, [m-1]_m\big\}$ der Kongruenzen modulo m, wird auch als $\mathbb{Z}_m$ geschrieben

$\mathbb{Q}$: rationale Zahlen $\big\{\frac{a}{b}\colon a \in \mathbb{Z} \text{ und } b \in \mathbb{Z} \setminus \{0\}\big\}$

$\mathbb{R}$: reelle Zahlen

$|\xi|$: Absolutbetrag einer reellen Zahl ξ

$\mathbb{F}_q$: endlicher Körper mit $q = p^k$ Elementen für eine Primzahl p und $k \in \mathbb{N}$

$R^\times$: Einheitengruppe eines Rings R mit 1

$K[X]$: Polynomring über den Körper K

$\mathrm{grad}(p)$: Grad des Polynoms p

$x \mid y$: x ist ein Teiler von y bzw. y ist ein Vielfaches von x

$x \nmid y$: x ist kein Teiler von y

$\mathrm{ggT}(x, y)$: größter gemeinsamer Teiler ganzer Zahlen x und y

$\mathrm{kgV}(x, y)$: kleinstes gemeinsames Vielfaches ganzer Zahlen x und y

$\langle v, w\rangle$: Skalarprodukt der zwei Vektoren v und w