Mathematische Grundlagen 1: Logik und Algebra

Dr. Viktoriya Ozornova

20. Januar 2018

Warnung

Dieses Skript enthält unter Umständen Fehler und Ungenauigkeiten jeder Art. Falls Sie solche entdecken, wäre es nett, wenn Sie mir Ihre Anmerkungen per E-Mail an

viktoriya.ozornova@rub.de

schicken könnten. Dieses Skript erhebt keinerlei Anspruch auf Originalität. Insbesondere habe ich bei der Erstellung dieses Skriptes die vorherigen Materialen von Emanuele Delucchi und Martin Dlugosch herangezogen, sowie die Bücher „Mathematik für Informatiker“ von G. Teschl und S. Teschl, „Mathematik für Informatiker“ von B. Kreußler und G. Pfister und „Mathematik für Informatiker“ von P. Hartmann.

Ferner möchte ich den Leuten danken, die bei der Erstellung dieses Skriptes auf die eine oder andere Weise hilfreich waren, insbesondere Roman Bruckner, Emanuele Delucchi, Martin Dlugosch, Eva-Maria Feichtner, Dimitry Feichtner-Kozlov, Tim Haga, Damien Imbs, Jan-Philipp Litza, Lennart Meier, Jan Senge, Ingolf Schäfer, Kirsten Schmitz sowie den vielen Studierenden der Informatik, die mir zu dem Skript Rückmeldung gegeben haben.

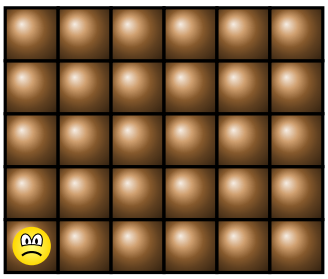
Inhaltsverzeichnis

Teil I
Allgemeine Grundlagen

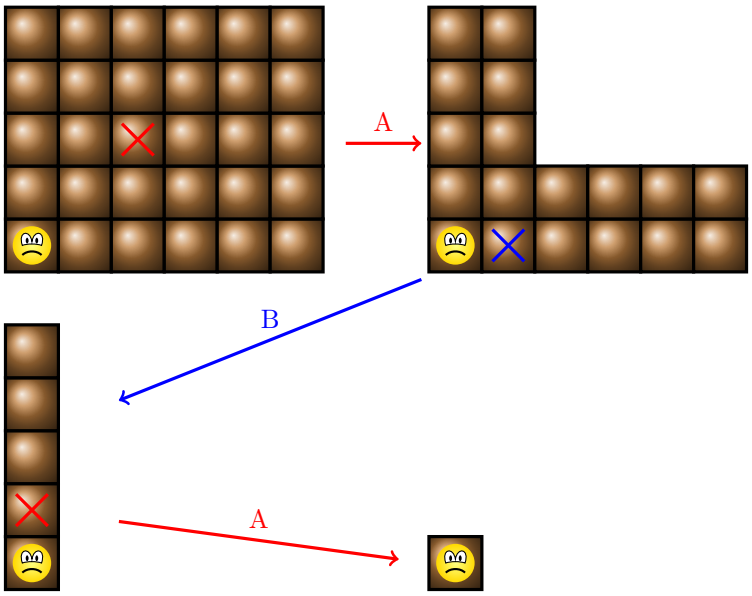
0 Chomp

Nach der Schule assoziieren die meisten mit der Mathematik vor allem Zahlen und Formeln; häufig wird z.B. die Geometrie, die eine andere Art und Weise ist, Mathematik zu betreiben, dabei vergessen. Bei Chomp haben wir ein neues Beispiel dafür, wie Mathematik aussehen kann.

Chomp ist ein Spiel. Es gibt jeweils zwei Spieler, die A und B genannt werden. Gespielt wird auf einem Spielfeld, das aus einer rechteckigen Tafel Schokolade besteht. Diese Tafel ist in $k \times n$ kleine quadratische Stückchen unterteilt, und das untere linke Stückchen ist vergiftet.



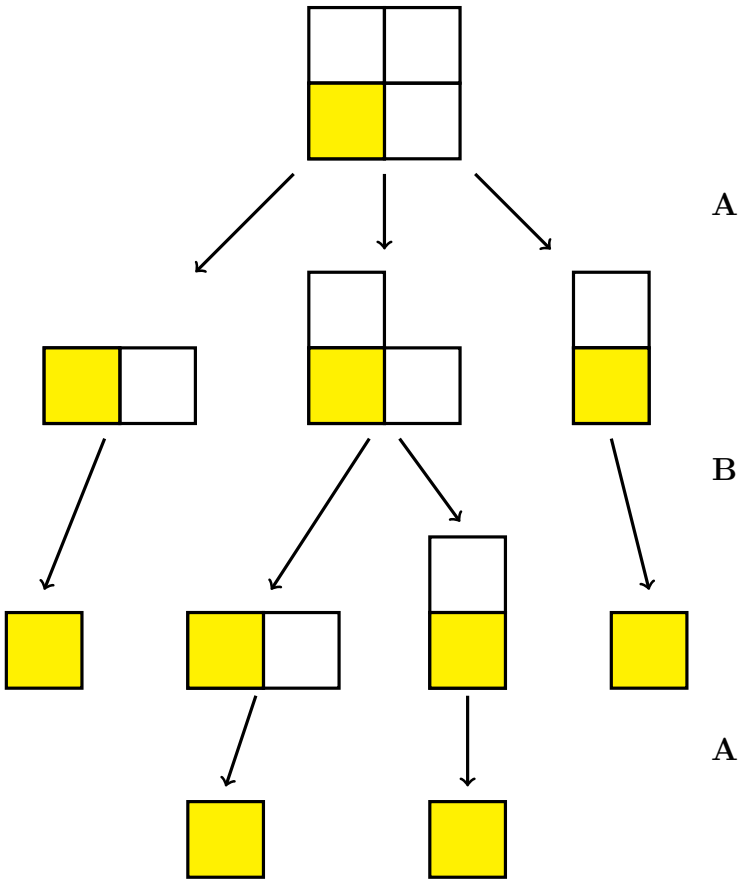
Die Spieler ziehen abwechselnd. Ein Spielzug, also ein Biss, besteht darin, sich eins der quadratischen Stückchen, die noch da sind, auszusuchen und dieses Stückchen sowie alle Stückchen rechts davon und alle oberhalb davon abzubeißen. Beim Abbeißen entsteht ein Geräusch, das ungefähr wie „Chomp“ klingt. Man muss jedes Mal ziehen und kann keinen Zug aussetzen. Verloren hat, wer die vergiftete Ecke isst. Ein Beispielspiel könnte also wie folgt aussehen:



An dieser Stelle bleibt B nichts anderes übrig, als die vergiftete Ecke zu essen; somit hat B also verloren.

Wenn man das Spiel einige Male spielt, so fragt man sich, ob es ein Rezept gibt, der einem erlaubt, immer zu gewinnen, eine sogenannte „Gewinnstrategie“. Etwas Ähnliches kennen die meisten aus Tic-Tac-Toe: da gibt es ein Rezept, wie man stets zumindest ein Unentschieden erreichen kann. An dieser Stelle ist es eine wichtige Beobachtung - beim Chomp kann es kein Unentschieden geben; nach einer gewissen Zeit wird alle Schokolade aufgegessen sein, denn sie wird mit jedem Zug immer weniger, und am Ende von diesem Prozess hat jemand die Giftecke gegessen. Es gibt also einen Verlierer und einen Gewinner.

Für jedes konkrete Chomp können wir die komplette Liste aller möglichen Spielstände und aller Züge aufstellen. Wir machen das am Beispiel vom 2×2-Chomp. Übrigens kann jeder Spieler sich jederzeit für die giftige linke untere Ecke zum Abbeißen entscheiden - allerdings ist es ziemlich dumm, denn dann verliert er sofort. In dem Diagramm, wo die Schokolade ab jetzt etwas vereinfacht dargestellt wird, lassen wir diese „dumme“ Option aus.



Aus dieser Tabelle sehen wir - macht A am Anfang den richtigen Zug (in diesem Fall ist es, die rechte obere Ecke abzubeißen), so kann er sich den Gewinn sichern. Wenn er allerdings dumm spielt, hilft ihm diese theoretische Möglichkeit nicht. In diesem Fall können wir also ganz einfach und sehr konkret eine Liste von Anweisungen erstellen, wie A spielen soll, um zu gewinnen, und für jeden Zug von B einen Gegenzug für A bestimmen, der ihm den Gewinn sichert. Eine solche Beschreibung, was ein Spieler in jeder Situation tun soll, um mit Sicherheit - solange er die Anweisungen befolgt - zu gewinnen, heißt **Gewinnstrategie**.

Wie sieht es mit anderen Chomps aus, hat da ein Spieler stets eine Gewinnstrategie? Wenn man etwas rumprobiert, kommt man auf den Gedanken, dass A eventuell eine Gewinnstrategie haben sollte. Für 2×3- oder 3×3-Chomps ist diese Gewinnstrategie noch ganz leicht hinzuschreiben, aber es wird immer komplizierter, je größer die Chomps werden. Tatsächlich ist keine ganz allgemeine Strategie für beliebige $k \times n$-Chomps bekannt. Allerdings wissen wir trotzdem, dass - außer in dem uninteressanten Fall von einem 1×1-Chomp - der erste Spieler stets eine Gewinnstrategie hat. Das wollen wir an dieser Stelle *beweisen*.

Satz 0.1. *Für jedes $k \times n$-Chomp, das nicht das 1×1-Chomp ist, hat der*

beginnende Spieler eine Gewinnstrategie.

Beweis. Seien k und n fest, und wir nehmen an, dass nicht der 1×1-Chomp vor uns liegt. Wir haben festgestellt, dass es in jeder Chomp-Partie einen Gewinner gibt. Wir stellen uns vor, dass wir, ähnlich wie beim 2×2-Chomp, uns die Liste aller möglichen Spielverläufe anlegen - diese ist unter Umständen sehr lang. Allerdings wissen wir mit Sicherheit, dass man von jeder Spielstellung heraus entweder durch die richtige Zugfolge sich den Gewinn sichern kann, oder notwendigerweise verliert. Das heißt: Einer von den beiden Spielern muss also eine Gewinnstrategie haben. Es können natürlich nicht gleichzeitig A und B eine Gewinnstrategie haben; also ist es entweder A oder B.

Nun stellen wir uns vor, B hätte eine solche Gewinnstrategie. An dieser Stelle kommt der „Strategiediebstahl“ ins Spiel. B hat einen Gegenzug für jeden Eröffnungszug von A; er weiß also insbesondere, welche Zugfolge er wählen muss, wenn A nur die rechte obere Ecke abbeißt. Er muss dann - laut seiner Gewinnstrategie - ein bestimmtes Feld, nennen wir dieses Feld X, für seinen Zug auswählen und auch danach hat er für jeden Zug von A einen Gegenzug. Aber: Dann könnte A ja damit anfangen, dass er das Feld X auswählt, was genau denselben Effekt wie soeben der zweite Zug von B. Jetzt nimmt sich A einfach die Strategie von B und tut einfach genau das, was die Strategie für B vorgesehen hat - er stiehlt die Strategie von B. Dann hätte also A eine Gewinnstrategie. Das ist allerdings im Widerspruch zur Annahme, dass B die Gewinnstrategie hat. Diese Annahme muss also falsch sein! Also hat A eine Gewinnstrategie. Das ist genau das, was wir beweisen wollten. Insbesondere wissen wir jetzt, dass die Aussage des Satzes wahr ist, ohne die konkrete Strategie hingeschrieben zu haben! □

1 Aussagenlogik

1.1 Was ist eine Aussage?

Die Grundbausteine der Mathematik sind Aussagen.

Definition 1.1. Eine **Aussage** ist ein grammatikalischer Satz, der entweder wahr oder falsch ist.

Das ist eine Definition - also in etwa so viel wie eine Vokabel. Die Definitionen an sich sind weder richtig noch falsch, sondern eine Festlegung. Es verhält sich jedoch anders z.B. im Rahmen eines Mathematik-Kurses - hier sorgen einheitliche Definitionen dafür, dass wir über dieselben Objekte sprechen, wenn wir ein bestimmtes Wort benutzen, und wollen genau wie Vokabeln gelernt sein. Insbesondere haben sich diese Definitionen als sinnvoll bewährt, während andere Begriffe, die man vielleicht hätte definieren können, sich nicht durchgesetzt haben.

Damit ein Satz eine Aussage ist, ist es nicht notwendig, dass man entscheiden kann, ob dieser Satz wahr oder falsch ist; es reicht das Wissen aus, dass (genau) eins von beiden zutreffen muss. Hier sind einige Beispiele von Aussagen:

- Alle Schafe sind schwarz. (falsch)
- Die Zahl 7 ist gerade. (falsch)
- Die vorherige Aussage ist falsch. (wahr)
- Die Summe der Winkel in jedem Dreieck beträgt 180°. (wahr)
- Die Zahl $\sqrt{250}$ ist größer als 16. (falsch)

Ob etwas eine Aussage ist, hängt teilweise vom Kontext ab und davon, ob alle darin vorkommende Objekte definiert bzw. festgelegt sind.

- $(y-7)^2 \geq 0$ ist ohne Kontext keine Aussage.
- Die Decke ist blau. - Dies ist keine Aussage, solange nicht klar ist, welche Decke gemeint ist.
- $5+7$. - Dies ist weder wahr noch falsch, also keine Aussage.
- Grün. - Das ist nicht mal ein Satz, also insbesondere keine Aussage.
- $a^2+b^2=c^2$. - Dies ist ohne Kontext keine Aussage.
- Wo ist mein Auto? - Das ist eine Frage, die kann weder wahr noch falsch sein, also ist das keine Aussage.

- Dieser Satz ist falsch. - Das ist ebenfalls keine Aussage, denn dieser Satz kann widerspruchsfrei weder wahr noch falsch sein.

Hingegen ist der Satz des Pythagoras eine Aussage (eine wahre, wie wir in der Schule gelernt haben):

Satz (Pythagoras)**.** *Sind a, b die Kathetenlängen und c die Hypotenusenlänge in einem rechtwinkligen Dreieck, so gilt für diese Zahlen $a^2 + b^2 = c^2$.*

Unter den Aussagen sind solche hervorzuheben, die „für alle Objekte x“ oder „für mindestens ein Objekt x“ gelten. Diese kommen häufig vor und deswegen betrachten wir einige solche Aussagen etwas genauer.

- Für alle reellen Zahlen x gilt: $x^2 \geq 0$. (wahr)
- Für alle reellen Zahlen x gilt $x^2 > 0$. (falsch)
- Es gibt reelle Zahlen x, für die gilt: $x^2 > 0$.(wahr)
- Es gibt reelle Zahlen x, für die gilt: $x^2 < 0$. (falsch)

Solche Aussagen werden uns in Zukunft noch häufig begegnen.

1.2 Verneinung von Aussagen

Man kann Aussagen auf verschiedene Arten und Weisen verknüpfen. Die erste Operation, die wir kennenlernen, ist die Verneinung. Diese entspricht ungefähr unserem intuitiven Verständnis: Die Verneinung ist „das Gegenteil“ der vorgegebenen Aussage. Die Verneinung (man sagt auch: die Negation) einer Aussage ist genau dann wahr, wenn die ursprüngliche Aussage falsch ist, und umgekehrt.

Notation. Für eine Aussage A schreiben wir $\neg A$ für die Verneinung von A.

Wir verdeutlichen das Prinzip der Verneinung an einigen Beispielen:

- Aussage A: Die Zahl 7 ist gerade.
  Verneinung von A: Die Zahl 7 ist ungerade.
- Aussage B: Alle Schafe sind schwarz.
  Verneinung von B: Nich alle Schafe sind schwarz.
  Äquivalente Formulierung: Es gibt Schafe, die nicht schwarz sind.
  Warnung: „Alle Schafe sind weiß“ ist keine gleichwertige Formulierung für die Verneinung. Es stimmt nämlich sowohl, dass nicht alle Schafe weiß sind, als auch, dass nicht alle Schafe schwarz sind!

Wie wir am letzten Beispiel gesehen haben, ist es etwas subtiler, Aussagen zu verneinen, die „für alle Objekte x“ oder „für mindestens ein Objekt x“ gelten. Zunächst schauen wir uns zwei Beispiele für Verneinung von „für alle“-Aussagen.

- Aussage: Für alle reellen Zahlen x gilt: $x^2 \geq 0$.
  Verneinung: Es gibt reelle Zahlen x, für die $x^2 < 0$ gilt.

- Aussage: Für alle ganzen Zahlen k gilt: $k^2 + 7$ ist ungerade.
  Verneinung: Es gibt eine ganze Zahl k, für die $k^2 + 7$ gerade ist.

Generell haben wir die folgende Merkregel: Will man eine Aussage der Form „Für alle Objekte x mit Eigenschaft P gilt $A(x)$.“ verneinen, so kann man das als „Es gibt Objekte x mit Eigenschaft P, für die $\neg A(x)$ gilt.“ formulieren.

Nun zu den „es gibt“-Aussagen (diese werden auch Existenz-Aussagen genannt):

- Aussage: Es gibt eine reelle Zahl y mit $y > 7$.
  Verneinung: Für alle reellen Zahlen y gilt: $y \leq 7$.
  Auch möglich: Es gibt keine reelle Zahl y mit $y > 7$.

- Aussage: Es gibt eine natürliche Zahl k, für die $k^2 + 2$ eine Primzahl ist.
  Verneinung: Für alle natürlichen Zahlen k gilt: k^2+2 ist keine Primzahl.

In diesem Fall ist die Merkregel wie folgt. Will man eine Aussage der Form „Es gibt Objekte x mit Eigenschaft P, für die $A(x)$ gilt.“ verneinen, so kann man das als „Für alle Objekte x mit Eigenschaft P gilt $\neg A(x)$.“ formulieren.

1.3 Und- und Oder-Verknüpfungen

Verneinung ist eine Operation, bei der man aus einer Aussage eine neue Aussage machen kann. Jetzt beschäftigen wir uns mit zwei Arten, jeweils zwei Aussagen zu einer zu verbinden. Die erste ist die „Und“-Verknüpfung. Damit eine Aussage der Form „A und B“ wahr ist, muss sowohl die Aussage A als auch die Aussage B wahr sein. Hier sind einige Beispiele:

- Alle Schafe sind schwarz und 7 ist ungerade. (falsch, da die erste Teilaussage falsch und die zweite wahr ist).

- Die Zahl 7 ist prim und ungerade. (wahr, da beide Teilaussagen wahr sind).

- Im 2×2-Chomp hat der erste Spieler eine Gewinnstrategie und für alle reellen Zahlen x gilt: $x^2 < 0$. (falsch, da die erste Teilaussage wahr und die zweite falsch ist).

- Die Zahl $\frac{1}{2}$ ist natürlich und größer 1. (falsch, da beide Teilaussagen falsch sind).

Notation. Wir bezeichnen die Und-Verknüpfung mit dem Symbol $\wedge$.

Um den Wahrheitswert der Verknüpfung zweier Variablen in Abhängigkeit von den Werten von Variablen festzuhalten, verwendet man *Wahrheitstabellen*. Die Wahrheitstabelle der Und-Verknüpfung sieht wie folgt aus:

A	B	$A \wedge B$
w	w	w
w	f	f
f	w	f
f	f	f

Dabei stehen „w“ und „f“ abkürzend für „wahr“ bzw. „falsch“.

Die zweite Verknüpfung, die wir kennenlernen, ist die „Oder“-Verknüpfung. Das mathematische „Oder“ ist etwas anders festgelegt, als wir es im Alltagsgebrauch handhaben. „A oder B“ ist wahr, wenn A wahr ist, oder B wahr ist, oder beides. Mit anderen Worten: Die „A oder B“-Aussage ist nur falsch, wenn sowohl A als auch B falsch sind.

Notation. Wir bezeichnen die Oder-Verknüpfung mit dem Symbol $\vee$.

In diesem Fall sieht die Wahrheitstabelle wie folgt aus:

A	B	$A \vee B$
w	w	w
w	f	w
f	w	w
f	f	f

In der Programmierung kommt der Typ „boolean“ vor, der sehr ähnlich zu Aussagen behandelt werden kann. In Java steht „!“ für die Verneinung, „&&“ für „Und“ und „||“ für Oder (dabei gibt es einige Feinheiten zu beachten; es soll an dieser Stelle nur als Analogie betrachtet werden).

Auch beim Verknüpfen von All-Aussagen oder Existenz-Aussagen mit „Und“ und „Oder“ ist Vorsicht geboten. So sind beispielsweise die Aussagen „Alle natürlichen Zahlen sind gerade“ und „Alle natürlichen Zahlen sind ungerade“ beide falsch, also ist deren Oder-Verknüpfung „Alle natürlichen Zahlen sind gerade oder alle natürlichen Zahlen sind ungerade“ falsch. Allerdings ist die Aussage „Alle natürlichen Zahlen sind gerade oder ungerade“ wiederum wahr. Die Aussagenteile „Für alle“ oder „Es gibt“ können nicht mit der Oder- und Und-Verknüpfung vertauscht werden.

1.4 Rechnen mit logischen Termen

Versucht man eine Aussage, die sich als Und-Verknüpfung zweier Aussagen zusammensetzt, zu verneinen, so merkt man, dass dabei etwas unintuitives passiert. Tatsächlich ist die Aussage „Es gibt Schafe, die nicht schwarz sind, und 7 ist eine Primzahl“ nicht das Gegenteil der Aussage „Alle Schafe sind schwarz und 7 ist keine Primzahl“, wie man schon daran sieht, dass beide Aussagen falsch sind. Um eine allgemeine Regel anzugeben, wie man Und-Aussagen verneint, führen wir die logischen Terme und logische Äquivalenz von solchen ein.

Definition 1.2 (vorläufig)**.** Ein Ausdruck in Variablen $A, B, C, \ldots$, die durch $\neg, \wedge, \vee$ verknüpft sind, heißt **(logischer) Term**.

Beispiel. $A \wedge B$, $(\neg(A \vee \neg B)) \vee C$, A sind Terme.

Definition 1.3. Zwei Terme heißen **logisch äquivalent**, wenn sie für alle Werte der darin vorkommenden Variablen denselben Wahrheitswert haben.

Wir machen ein einfaches Beispiel, um zu sehen, wie logische Äquivalenz nachgewiesen werden kann.

Proposition 1.4. *A und $\neg(\neg A)$ sind logisch äquivalente Terme.*

Beweis. Wir prüfen, dass die Terme A und $\neg(\neg A)$ für jeden Wert der Variable A denselben Wahrheitswert haben.

A	A	$\neg A$	$\neg(\neg A)$
w	w	f	w
f	f	w	f

Die 2. und 4. Spalte, die zu A bzw. $\neg(\neg A)$ gehören, stimmen überein, also sind A und $\neg(\neg A)$ logisch äquivalent. □

Nun stellen wir allgemeine Regeln zum Verneinen von Und- und Oder-Ausdrücken auf. Außerdem führen wir noch weitere Rechenregeln für Und- und Oder-Verknüpfungen auf. Dabei erinnern einige Regeln an die Rechenregeln, die wir aus der Schule für Addition und Multiplikation der reellen Zahlen kennen; diese Regeln tragen dann die entsprechenden Namen.

Satz 1.5 (De Morganschen Regeln)**.**

1. *Verneinung von Und:* $\neg(A \wedge B) \leftrightarrow (\neg A) \vee (\neg B)$.
2. *Verneinung von Oder:* $\neg(A \vee B) \leftrightarrow (\neg A) \wedge (\neg B)$.
3. *Und-Oder-Distributivgesetz:* $(A \wedge B) \vee C \leftrightarrow (A \vee C) \wedge (B \vee C)$.
4. *Oder-Und-Distributivgesetz:* $(A \vee B) \wedge C \leftrightarrow (A \wedge C) \vee (B \wedge C)$.
5. *Assoziativität von Und:* $(A \wedge B) \wedge C \leftrightarrow A \wedge (B \wedge C)$.

6. *Assoziativität von Oder:* $(A \vee B) \vee C \leftrightarrow A \vee (B \vee C)$.

7. *Kommutativität von Und:* $A \wedge B \leftrightarrow B \wedge A$.

8. *Kommutativität von Oder:* $A \vee B \leftrightarrow B \vee A$.

Beweis. Der Beweis ist in allen Fällen mit Hilfe von Wahrheitstabellen zu erbringen. Man hat in jedem Teilpunkt zu zeigen, dass die Wahrheitswerte der beiden Ausdrücke für alle Werte von A, B und ggf. C dieselben sind. In den letzten vier Punkten wird dies als einfache Übung dem/der LeserIn überlassen. Für die ersten vier Fälle stellen wir die Wahrheitstabellen auf. Die Werte der zu untersuchender Ausdrücke werden jeweils hervorgehoben.

1. Für die logischen Terme $\neg(A \wedge B)$ und $(\neg A) \vee (\neg B)$ erhalten wir:

A	B	$A \wedge B$	$\neg(A \wedge B)$	$\neg A$	$\neg B$	$(\neg A) \vee (\neg B)$
w	w	w	f	f	f	f
w	f	f	w	f	w	w
f	w	f	w	w	f	w
f	f	f	w	w	w	w

   Somit sind die beiden Terme logisch äquivalent.

2. Für die logischen Terme $\neg(A \vee B)$ und $(\neg A) \wedge (\neg B)$ erhalten wir:

A	B	$A \vee B$	$\neg(A \vee B)$	$\neg A$	$\neg B$	$(\neg A) \wedge (\neg B)$
w	w	w	f	f	f	f
w	f	w	f	f	w	f
f	w	w	f	w	f	f
f	f	f	w	w	w	w

   Somit sind die beiden Terme logisch äquivalent.

3. Für die logischen Terme $(A \wedge B) \vee C$ und $(A \vee C) \wedge (B \vee C)$ sieht die Wahrheitstabelle wie folgt aus:

A	B	C	$A \wedge B$	$(A \wedge B) \vee C$	$A \vee C$	$B \vee C$	$(A \vee C) \wedge (B \vee C)$
w	w	w	w	w	w	w	w
w	w	f	w	w	w	w	w
w	f	w	f	w	w	w	w
w	f	f	f	f	w	f	f
f	w	w	f	w	w	w	w
f	w	f	f	f	f	w	f
f	f	w	f	w	w	w	w
f	f	f	f	f	f	f	f

Da die Spalten, die zu $(A \wedge B) \vee C$ bzw. $(A \vee C) \wedge (B \vee C)$ gehören, übereinstimmen, sind die beiden Terme logisch äquivalent.

4. Für die logischen Terme $(A \vee B) \wedge C$ und $(A \wedge C) \vee (B \wedge C)$ sieht die Wahrheitstabelle wie folgt aus:

A	B	C	$A \vee B$	$(A \vee B) \wedge C$	$A \wedge C$	$B \wedge C$	$(A \wedge C) \vee (B \wedge C)$
w	w	w	w	w	w	w	w
w	w	f	w	f	f	f	f
w	f	w	w	w	w	f	w
w	f	f	w	f	f	f	f
f	w	w	w	w	f	w	w
f	w	f	w	f	f	f	f
f	f	w	f	f	f	f	f
f	f	f	f	f	f	f	f

Da die Spalten, die zu $(A \wedge B) \vee C$ bzw. $(A \vee C) \wedge (B \vee C)$ gehören, übereinstimmen, sind die beiden Terme logisch äquivalent.

□

Die Rechenregeln können benutzt werden, um wiederum neue logische Äquivalenzen zu beweisen.

Beispiel. Wir vereinfachen den Term $\neg((X \vee Y) \wedge \neg Z)$ mit Variablen X, Y, Z mit Hilfe der Rechenregeln. Durch die De Morgan-Regel für die Negation von Und erhalten wir:

$$\neg((X \vee Y) \wedge \neg Z) \leftrightarrow (\neg(X \vee Y)) \vee (\neg(\neg Z)).$$

Wir lösen die erste Klammer auf, indem wir De Morgan-Regel für die Negation von Oder anwenden:

$$(\neg(X \vee Y)) \vee (\neg(\neg Z)) \leftrightarrow (\neg X \wedge \neg Y) \vee (\neg(\neg Z)).$$

Schließlich lösen wir die doppelte Negation auf:

$$(\neg X \wedge \neg Y) \vee (\neg(\neg Z)) \leftrightarrow (\neg X \wedge \neg Y) \vee Z.$$

Hier könnte man noch das Distributivgesetz anwenden; allerdings wird der Term dadurch nicht kürzer.

1.5 Implikation

Eine weitere Art, zwei Aussagen zu einer zu verknüpfen, ist die „wenn-dann"-Verknüpfung. Hier sind einige Beispiele:

- Wenn es regnet, ist der Boden nass. (wahr)
- Wenn p eine Primzahl ist, so ist $p+1$ auch eine Primzahl. (falsch)
- Wenn k eine gerade Zahl ist, dann ist 7 ungerade. (wahr)
- Wenn 7 ungerade ist, dann ist $2+3=7$. (falsch)
- Wenn eine ganze Zahl k gerade ist, dann ist k^2 auch gerade. (wahr)
- Wenn x eine positive reelle Zahl ist, dann ist $x^2 > 0$. (wahr)

Dabei ist der Wahrheitswert der Verknüpfung „Aus A folgt B" durch die folgende Wahrheitstabelle vorgegeben:

A	B	$A \Rightarrow B$
w	w	w
w	f	f
f	w	w
f	f	w

Insbesondere folgt aus einer falschen Aussage alles; hingegen kann aus einer wahren Aussage nur eine wahre Aussage folgen.

Um ein besseres Verständnis der Implikation zu erlangen, zeigen, wir, dass die folgende Umformulierung dafür gilt.

Proposition 1.6. *Die Terme $A \Rightarrow B$ und $(\neg A) \vee B$ sind logisch äquivalent.*

Beweis. Wir prüfen anhand der Wahrheitstabelle, dass die Terme $A \Rightarrow B$ und $(\neg A) \vee B$ für jeden Wert der Variablen A, B denselben Wahrheitswert haben.

A	B	$\neg A$	$(\neg A) \vee B$	$A \Rightarrow B$
w	w	f	w	w
w	f	f	f	f
f	w	w	w	w
f	f	w	w	w

Da die letzten beiden Spalten, die zu den Termen $A \Rightarrow B$ bzw. $(\neg A) \vee B$ gehören, übereinstimmen, sind diese Terme logisch äquivalent. □

Beispiel. Die Aussage „Wenn es regnet, ist der Boden nass" ist äquivalent zu „Es regnet nicht, oder der Boden ist nass (oder vielleicht beides)".

Ein weiterer äquivalenter Ausdruck für die Implikation $A \Rightarrow B$ ist $\neg B \Rightarrow \neg A$, wie in der folgenden Proposition gezeigt wird. Dies mag im ersten Moment kontraintuitiv erscheinen.

Proposition 1.7. *Die Terme $A \Rightarrow B$ und $\neg B \Rightarrow \neg A$ sind logisch äquivalent.*

Beweis. Wir zeigen die Behauptung, indem wir die bisher bewiesenen Resultate verwenden, insbesondere die logischen Rechenregeln und die Proposition 1.6. Dabei fangen wir mit dem zweiten Ausdruck an.

$\neg B \Rightarrow \neg A$	$\leftrightarrow$	$\neg(\neg B) \vee \neg A$	Proposition 1.6
	$\leftrightarrow$	$B \vee \neg A$	Doppelte Neg.
	$\leftrightarrow$	$\neg A \vee B$	Kommutativität
	$\leftrightarrow$	$A \Rightarrow B$	Proposition 1.6

Damit haben wir also gezeigt, dass die logischen Terme $A \Rightarrow B$ und $\neg B \Rightarrow \neg A$ logisch äquivalent sind. □

Will man also eine Aussage vom Typ $A \Rightarrow B$ beweisen, so ist es äquivalent, stattdessen die Aussage $\neg B \Rightarrow \neg A$ zu beweisen. Diese Beweistechnik nennt man auch *Kontraposition*.

Beispiel. Die Aussage „Wenn es regnet, ist der Boden nass“ ist äquivalent zu „Wenn der Boden nicht nass ist, regnet es nicht.“.

Als nächstes beschäftigen wir uns damit, wie man eine „Wenn-dann“-Aussage verneint. Hat man beispielsweise die Behauptung „Für alle natürlichen Zahlen k gilt: Ist k durch 7 teilbar, so ist $2k+3$ eine Primzahl“ und will diese widerlegen, so muss man sich klar machen, was das Gegenteil dieser Aussage ist. Dafür haben wir - zunächst für logische Terme - die folgende Proposition.

Proposition 1.8. *Die logischen Terme $\neg(A \Rightarrow B)$ und $A \wedge \neg B$ sind logisch äquivalent.*

Beweis. Wir zeigen die Behauptung erneut durch Anwendung der Rechenregeln und der Proposition 1.6. Dabei fangen wir mit dem ersten Ausdruck an.

$\neg(A \Rightarrow B)$	$\leftrightarrow$	$\neg(\neg A \vee B)$	Proposition 1.6
	$\leftrightarrow$	$\neg(\neg A) \wedge \neg B$	Neg. von Oder
	$\leftrightarrow$	$A \wedge \neg B$	Doppelte Neg.

Daraus folgt, dass die logischen Terme $\neg(A \Rightarrow B)$ und $A \wedge \neg B$ logisch äquivalent sind. □

Um das zu verdeutlichen, betrachten wir erneut einige Beispiele.

Beispiel.

- Die Verneinung von „Wenn es regnet, ist der Boden nass“ ist „Es regnet und der Boden nicht nass“.
- Die Verneinung der Aussage „Für alle natürlichen Zahlen k gilt: Ist k durch 7 teilbar, so ist $2k+3$ eine Primzahl“ lautet „Es gibt eine natürliche Zahl k, die durch 7 teilbar ist und sodass $2k+3$ keine Primzahl ist“. Diese zweite Aussage ist wahr, denn $k = 21$ erfüllt die genannten Bedingungen.
- Wir betrachten die Aussage „Für alle reellen Zahlen x gibt es eine reelle Zahl y, sodass gilt: Wenn $(x-y)^2 \geq 7$, so ist $x \geq 2$.“ Die Verneinung dieser Aussage lautet: „Es gibt eine reelle Zahl x, sodass für alle reellen Zahlen y gilt: $(x-y)^2 \geq 7$ und $x < 2$.“ In diesem Fall ist die ursprüngliche Aussage wahr: Für jede reelle Zahl x können wir $y = x$ wählen; dann ist die Prämisse in der Aussage „Wenn $(x-y)^2 \geq 7$, so ist $x \geq 2$.“ falsch, und somit die Aussage wahr.

1.6 Äquivalenz

Wir lernen eine letzte Verknüpfung von Aussagen kennen: Die Äquivalenz von Aussagen. Sind A und B Aussagen, so lautet die zusammengesetzte Aussage „A ist äquivalent zu B“ oder „A ist genau dann wahr, wenn B wahr ist“, in Zeichen $A \Leftrightarrow B$. Diese Verknüpfung ist durch die folgende Wahrheitstabelle definiert:

A	B	$A \Leftrightarrow B$
w	w	w
w	f	f
f	w	f
f	f	w

Wir suchen wiederum einige Umformulierungen der Äquivalenz, um diese besser zu verstehen:

Proposition 1.9. *Die folgenden logischen Terme sind logisch äquivalent:*

- $A \Leftrightarrow B$,
- $(A \Rightarrow B) \wedge (B \Rightarrow A)$,
- $(A \wedge B) \vee (\neg A \wedge \neg B)$,
- $(A \Rightarrow B) \wedge (\neg A \Rightarrow \neg B)$.

Beweis. Um die logische Äquivalenz der ersten drei logischen Terme zu beweisen, stellen wir die Wahrheitstabelle für diese auf. Abkürzend schreiben wir $C = (A \Rightarrow B) \wedge (B \Rightarrow A)$ und $D = (A \wedge B) \vee (\neg A \wedge \neg B)$.

A	B	$A \Leftrightarrow B$	$A \Rightarrow B$	$B \Rightarrow A$	C	$A \wedge B$	$\neg A \wedge \neg B$	D
w	w	w	w	w	w	w	f	w
w	f	f	f	w	f	f	f	f
f	w	f	w	f	f	f	f	f
f	f	w	w	w	w	f	w	w

Da die Spalten der Wahrheitstabelle, die zu den Ausdrücken $A \Leftrightarrow B$, $C = (A \Rightarrow B) \wedge (B \Rightarrow A)$ und $D = (A \wedge B) \vee (\neg A \wedge \neg B)$ gehören, dieselben Wahrheitswerte aufweisen, sind diese drei logischen Terme logisch äquivalent.

Die logische Äquivalenz vom vierten und zweiten Term ist eine Folgerung der Proposition 1.7. Somit ist die logische Äquivalenz aller vier logischen Terme bewiesen. □

Die obige Proposition ist häufig nützlich, um eine Äquivalenz zweier Aussagen zu beweisen.

Beispiel. Sei k eine natürliche Zahl. Wir betrachten die folgende Aussage: Die Zahl k ist genau dann gerade, wenn ihr Quadrat k^2 gerade ist. Nach der vorherigen Proposition sind die folgenden vier Aussagen äquivalent:

- Die Zahl k ist genau dann gerade, wenn ihr Quadrat k^2 gerade ist.
- Ist k gerade, so auch k^2, und ist k^2 gerade, dann auch k.
- Ist k gerade, so ist k^2 gerade, und ist k ungerade, so ist k^2 ungerade.
- k und k^2 sind beide gerade, oder k und k^2 sind beide ungerade.

1.7 Disjunktive und konjunktive Normalformen

Als letztes in diesem Kapitel wollen wir über zwei standardisierte Formen von Termen sprechen, die disjunktive Normalform und die konjunktive Normalform. Deren Konstruktion beruht auf der folgenden Beobachtung. Hat man die Variablen $X_1, \ldots, X_n$ vorgegeben und wählt sich von jeder Variable X_i entweder die Variable selbst oder deren Negation als Term Y_i, so ist der Term $Y_1 \wedge \ldots \wedge Y_n$ für genau eine Belegung der Variablen $X_1, \ldots, X_n$ wahr und für alle anderen falsch. Verknüpft man nun mehrere Terme, die diese Eigenschaft haben, mit der Oder-Verknüpfung, so erhält man einen neuen Term, der genau dort die Wahr-Einträge in der Wahrheitstabelle hat, wo eines der Und-Terme den Wert „Wahr“ hat. Umgekehrt sieht es aus, wenn man sich nur den „Falsch“-Werten in der Wahrheitstabelle widmet.

Als grobe Idee lässt sich festhalten: Die konjunktive Normalform ist ein besonders einfacher Ausdruck, der als Und-Verknüpfung von Oder-Termen besteht; bei disjunktiver Normalform ist es die Oder-Verknüpfung von Und-Termen. Jetzt können wir diesen Begriff präzisieren.

Definition 1.10. Wir sagen, ein Term Z in den Variablen $X_1, X_2, \ldots, X_n$ ist in **konjunktiver Normalform**, falls er sich schreiben lässt als

$$Z = Z_1 \wedge Z_2 \wedge \ldots \wedge Z_k$$

für ein $k \geq 1$, wobei sich jeder der Terme Z_i für $1 \leq i \leq k$ wiederum schreiben lassen soll als

$$Z_i = Y_1 \vee Y_2 \vee \ldots \vee Y_n,$$

und jedes Y_j entweder die Variable X_j oder deren Negation $\neg X_j$ ist. (Dabei sind natürlich die Y_j für jedes Z_i unterschiedlich. Wir befolgen dabei die Konvention, dass alle Variablen in jedem Z_j vorkommen sollen.)

Sehr ähnlich ist die Definition der disjunktive Normalform:

Definition 1.11. Wir sagen, ein Term Z in den Variablen $X_1, X_2, \ldots, X_n$ ist in **disjunktiver Normalform**, falls er sich schreiben lässt als

$$Z = Z_1 \vee Z_2 \vee \ldots \vee Z_k$$

für ein $k \geq 1$, wobei sich jeder der Terme Z_i für $1 \leq i \leq k$ wiederum schreiben lassen soll als

$$Z_i = Y_1 \wedge Y_2 \wedge \ldots \wedge Y_n,$$

und jedes Y_j entweder die Variable X_j oder deren Negation $\neg X_j$ ist.

Wir betrachten einige Beispiele.

Beispiel.

- Der Term $Z = (A \wedge B \wedge C) \vee (\neg A \wedge B \wedge \neg C) \vee (\neg A \wedge \neg B \wedge \neg C)$ ist in disjunktiver Normalform. Denn dieser lässt sich schreiben als $Z = Z_1 \vee Z_2 \vee Z_3$, wobei $Z_1 = A \wedge B \wedge C$, $Z_2 = \neg A \wedge B \wedge \neg C$, $Z_3 = \neg A \wedge \neg B \wedge \neg C$. Insbesondere ist jeder der Terme Z_1, Z_2, Z_3 eine Und-Verknüpfung, in der jede der Variablen A, B, C (ob verneint oder nicht) einmal vorkommt.

- Der Term $X = \neg(A \wedge B)$ ist weder in konjunktiver noch in disjunktiver Normalform. Es ist zwar eine Und- bzw. eine Oder-Verknüpfung von einem einzelnen Term mit sich selbst, aber dieser ist weder eine Und- noch eine Oder-Verknüpfung von Variablen und ihren Negationen.

- Der Term $Y = (A \vee B \vee C) \wedge (\neg A \vee B \vee \neg C)$ ist in konjunktiver Normalform. Denn dieser lässt sich schreiben als $Y = Y_1 \wedge Y_2$, wobei $Y_1 = A \vee B \vee C$ und $Y_2 = \neg A \vee B \vee \neg C$. Insbesondere ist sowohl Y_1 als auch Y_2 eine Oder-Verknüpfung, in der jede der Variablen A, B, C (ob verneint oder nicht) einmal vorkommt.

- Der Term $T = A \wedge B$ ist in disjunktiver Normalform (in Variablen A, B): Dieser ist die Oder-Verknüpfung $T = T_1$ von einem einzigen Und-Term, in dem jede Variable einmal vorkommt. Nach unserer Definition ist dieser Term allerdings nicht in konjunktiver Normalform, da in den Teiltermen A bzw. B nicht jeweils alle Variablen vorkommen. Es gibt auch Konventionen, nach denen das die konjunktive Normalform ist und unsere Normalform dann „vollständige konjunktive Normalform" heißt.

Dass die Normalformen wirklich zum Standardisieren geeignet sind, sieht man an dem folgenden Satz. Wir werden diesen nicht beweisen, da der Beweis zwar nicht schwierig, doch langwierig ist, allerdings schauen wir uns Beispiele an, durch die die Grundidee des Beweises deutlich wird.

Satz 1.12. *Jeder logische Term in den Variablen $X_1, \ldots, X_n$ ist zu einem logischen Term in konjunktiver Normalform in den Variablen $X_1, \ldots, X_n$ logisch äquivalent, und letzterer ist bis auf die Reihenfolge der Teilterme, die durch Und verknüpft werden, eindeutig.*

Jeder logische Term in den Variablen $X_1, \ldots, X_n$ ist zu einem logischen Term in disjunktiver Normalform in den Variablen $X_1, \ldots, X_n$ logisch äquivalent, und letzterer ist bis auf die Reihenfolge der Teilterme, die durch Oder verknüpft werden, eindeutig.

Beispiel. Manchmal lässt sich die konjunktive bzw. disjunktive Normalform „erraten", oder man kann diese leicht durch logische Umformungen herleiten. Beispielsweise wissen wir für den Term $A \Leftrightarrow B$ in den Variablen A, B nach Proposition 1.9:

$$(A \Leftrightarrow B) \leftrightarrow (A \wedge B) \vee (\neg A \wedge \neg B),$$

und die rechte Seite ist in der disjunktiver Normalform. Für die konjunktive Normalform starten wir mit einer weiteren Beschreibung aus der Proposition 1.9 und wenden dann die Proposition 1.6 an:

$$(A \Leftrightarrow B) \leftrightarrow (A \Rightarrow B) \wedge (B \rightarrow A) \leftrightarrow (\neg A \vee B) \wedge (A \vee \neg B),$$

und letzter Term ist in konjunktiver Normalform.

Beispiel. Wir wollen für den Term $T = \neg((A \vee B \vee C) \wedge (A \Rightarrow C))$ jeweils einen logisch äquivalenten Term in disjunktiver bzw. konjunktiver Normalform. Bevor wir das tun, ist die folgende Beobachtung hilfreich: Ein Term $Y_1 \wedge Y_2 \wedge \ldots \wedge Y_n$ in den Variablen $X_1, \ldots, X_n$, in dem jedes Y_j entweder die Variable X_j oder ihre Negation $\neg X_j$ ist, ist für genau eine Belegung von Variablen $X_1, \ldots X_n$ wahr und für alle andere Belegungen falsch. Umgekehrt gibt es zu jeder Belegung der Variablen $X_1, \ldots X_n$ einen Und-Term der obigen Form, der für diese Belegung wahr ist und für alle anderen Belegungen

falsch. Hat man eine nun eine disjunktive Normalform $Z = Z_1 \vee \ldots \vee Z_k$ vorgegeben, so ist jeder der Terme Z_i für genau eine Belegung der Variablen wahr; nimmt man die Wahr-Belegungen für alle der k-Terme $Z_1, \ldots, Z_k$ zusammen, so erhält man genau die Belegungen, an denen der Term Z wahr ist. Umgekehrt liefert uns das eine Methode, wie wir zu einem vorgegebenen Term die disjunktive Normalform bilden können: Für jede Belegung der Variablen, bei der der vorgegebene Term wahr ist, erhalten wir einen Und-Term, und durch die Oder-Verknüpfung dieser Und-Terme erhalten wir einen zu dem vorgegebenen logisch äquivalenten Term in der disjunktiven Normalform. Bei konjunktiver Normalform muss man umgekehrt verfahren und für jede Falsch-Belegung den passenden Oder-Term finden, die man dann mit dem Und verknüpft. Das demonstrieren wir in diesem Beispiel.

Dafür stellen wir zunächst die Wahrheitstabelle von $T = \neg((A \vee B \vee C) \wedge (A \Rightarrow C))$ auf. Abkürzend schreiben wir $S = (A \vee B \vee C) \wedge (A \Rightarrow C)$. In den letzten beiden Spalten geben wir einen Und-Term an, der genau an dieser Stelle wahr ist, oder einen Oder-Term an, der genau an dieser Stelle falsch ist - abhängig davon, ob T an dieser Stelle wahr oder falsch ist.

A	B	C	$A \vee B \vee C$	$A \Rightarrow C$	S	T	Und-Term	Oder-Term
w	w	w	w	w	w	f		$\neg A \vee \neg B \vee \neg C$
w	w	f	w	f	f	w	$A \wedge B \wedge \neg C$	
w	f	w	w	w	w	f		$\neg A \vee B \vee \neg C$
w	f	f	w	f	f	w	$A \wedge \neg B \wedge \neg C$	
f	w	w	w	w	w	f		$A \vee \neg B \vee \neg C$
f	w	f	w	w	w	f		$A \vee \neg B \vee C$
f	f	w	w	w	w	f		$A \vee B \vee \neg C$
f	f	f	f	w	f	w	$\neg A \wedge \neg B \wedge \neg C$	

Somit erhalten wir den Term $(A \wedge B \wedge \neg C) \vee (A \wedge \neg B \wedge \neg C) \vee (\neg A \wedge \neg B \wedge \neg C)$ in disjunktiver Normalform, der logisch äquivalent zum vorgegebenen Term T ist. Außerdem haben wir den Term

$$(\neg A \vee \neg B \vee \neg C) \wedge (\neg A \vee B \vee \neg C) \wedge (A \vee \neg B \vee \neg C) \wedge (A \vee \neg B \vee C) \wedge (A \vee B \vee \neg C)$$

in konjunktiver Normalform, der zum vorgegebenen Term T logisch äquivalent ist.

2 Elementare Mengenlehre

2.1 Grundlegende Definitionen

Mengen gehören zu den grundlegenden Objekten, die wir in der Mathematik betrachten. Einige davon haben wir bereits in der Schule kennengelernt, wie etwa die Menge der natürlichen Zahlen oder Lösungsmengen von Gleichungen. Die folgende Definition, die auf G. Cantor zurückgeht, legen wir unserem Begriff von Menge zugrunde. Natürlicherweise besteht dabei das Problem, dass wir auf andere Begriffe zurückgreifen müssen, die wir noch nicht definiert haben. Dieses Problem ist eines von grundsätzlicher Natur und soll uns zunächst nicht weiter stören.

Definition 2.1. Eine **Menge** ist eine Zusammenfassung von bestimmten und wohlunterschiedenen Objekten unseres Denkens oder unserer Anschauung zu einem Ganzen. Diese Objekte werden die **Elemente** der Menge genannt.

Beispiel. Es gibt einige Wege, eine Menge zu spezifizieren. Die einfachste ist, alle Elemente der Menge einfach aufzulisten; als Notation verwenden wir geschweifte Klammern, um das Zusammenfassen von Objekten hervorzuheben, z.B. $\{0, 1\}$ für die Menge, die die Objekte 0 und 1 zusammenfasst. Die Objekte der Menge müssen in keiner Verbindung miteinander stehen: Beispielsweise können wir die Menge $\{0, 1, \frac{7}{8}, x, y, \text{Erde}\}$ betrachten, wobei x und y als Symbole zu verstehen sind.

Weiterhin haben wir für gewisse Mengen, die uns in der Mathematik besonders häufig begegnen, feste Notation reserviert: $\mathbb{N}$ für die Menge der natürlichen Zahlen (je nach Kontext mit oder ohne 0; wenn wir betonen wollen, dass die 0 dazugehört, schreiben wir $\mathbb{N}_0$), $\mathbb{Z}$ für die Menge der ganzen Zahlen, $\mathbb{Q}$ für die Menge der rationalen Zahlen und $\mathbb{R}$ für die Menge der reellen Zahlen.

Eine weitere Menge, die besonderer Erwähnung bedarf, ist die **leere Menge** $\varnothing$. Diese ist als die Menge definiert, die *keine* Objekte zusammenfasst.

Für die Aussage „Objekt x ist ein Element der Menge M“ haben wir die Kurzschreibweise $x \in M$, für „x ist kein Element der Menge M“ schreiben wir $x \notin M$. Beispielsweise ist $\sqrt{2} \in \mathbb{R}$, aber $\sqrt{2} \notin \mathbb{Q}$. Für alle Objekte x ist die Aussage „$x \notin \varnothing$“ wahr und „$x \in \varnothing$“ falsch, denn kein Objekt gehört zu den Objekten, die von der leeren Menge zusammengefasst werden.

Eine vorerst letzte Möglichkeit, eine Menge zu beschreiben, besteht darin, Eigenschaften von Objekten zu fordern. Dies geschieht immer in der ungefähren Form „$\{$Objekte $x \mid x$ hat Eigenschaft $P\}$“.

Beispielsweise ist $\{x \in \mathbb{R} \mid x > 3\}$ die Menge aller reellen Zahlen, die größer als 3 sind; die Menge $\{q \in \mathbb{N} \mid 7 \text{ teilt } q\}$ ist die Menge der natürlichen Zahlen, die durch 7 teilbar sind.

Um Mengen zu vergleichen, benötigen wir als erstes die folgende Definition:

Definition 2.2. Zwei Mengen sind **gleich**, wenn sie dieselben Elemente enthalten.

Insbesondere wird die Reihenfolge, in der Elemente aufgelistet werden, nicht berücksichtigt, und die Mehrfachnennungen werden zusammengefasst. So ist beispielsweise $\{0,1\} = \{1,0\}$ und $\{z,z\} = \{z\}$. Das steht im Gegensatz zu geordneten Paaren (kurz auch: Paaren) von Objekten, bei denen es auf die Reihenfolge ankommt und auch Mehrfachnennungen (sinnvollerweise) möglich sind. Solche Paare begegneten uns im Schulunterricht als Koordinaten von Punkten in der Ebene. So ist der Punkt $(0,1)$ ein anderer als der Punkt $(1,0)$, und der Punkt $(3,3)$ ist ebenfalls eine sinnvolle Koordinatenangabe.

Manchmal ist Gleichheit von Mengen auch nicht gänzlich offensichtlich. Hier sind einige Beispiele:

Beispiel.
- $\{x \in \mathbb{R} \mid x < 2 \text{ und } x > 3\} = \varnothing$, da es keine reelle Zahl gibt, die beide Ungleichungen gleichzeitig erfüllt.
- $\{y \in \mathbb{R} \mid 7y + 14 = 0\} = \{-2\}$, denn -2 ist die eindeutige Lösung der Gleichung $7y + 14 = 0$.
- $\{a \in \mathbb{R} \mid a^2 \geq 0\} = \mathbb{R}$, denn die Ungleichung $a^2 \geq 0$ gilt für jede reelle Zahl a.

Eine weitere Art, Mengen zueinander in Verbindung zu setzen, ist durch den Begriff der Teilmenge gegeben.

Definition 2.3. Eine Menge A heißt **Teilmenge** einer Menge B, falls jedes Element von A auch ein Element von B ist. Wir schreiben dafür $A \subset B$ oder $A \subseteq B$. In Zeichen sieht die Definition nochmal wie folgt aus:

$$(A \subset B) :\Leftrightarrow (x \in A \Rightarrow x \in B).$$

Eine Menge A heißt **echte** Teilmenge einer Menge B (in Zeichen: $A \subsetneq B$), wenn A eine Teilmenge von B ist und zusätzlich $A \neq B$ gilt.

Beispiel.
- $\{2,7\}$ ist eine Teilmenge der Menge $\{2,7\}$, aber keine echte Teilmenge.
- $\{1,3,5\}$ ist eine echte Teilmenge der Menge $\{1,2,3,4,5\}$.
- $\mathbb{N}$ ist eine echte Teilmenge der Menge $\mathbb{Z}$. Jede natürliche Zahl ist insbesondere eine ganze Zahl, also gilt $\mathbb{N} \subset \mathbb{Z}$. Da es auch ganze Zahlen gibt, die keine natürlichen Zahlen sind, z.B. $-7 \in \mathbb{Z}$, aber $-7 \notin \mathbb{N}$, gilt außerdem $\mathbb{N} \neq \mathbb{Z}$.

- Genauso gilt $\mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}$. Für letzteres wissen wir, dass es reelle Zahlen gibt, wie z.B. $\sqrt{2}$ oder π, die keine rationalen Zahlen sind.
- Für jede Menge A gilt: $A \subset A$ und $\varnothing \subset A$. Um letzteres zu überprüfen, merken wir, dass jedes Element der leeren Menge - also kein Objekt - auch ein Element von A ist.

Manchmal ist es sinnvoll, die Gesamtheit aller Teilmengen einer Menge zu betrachten.

Definition 2.4. Sei A eine Menge. Die **Potenzmenge** von A (in Zeichen: $\mathcal{P}(A)$) ist die Menge aller Teilmengen von A.

Beispiel. In kleinen Beispielen kann man sich die Potenzmenge einer Menge A komplett hinschreiben. Ist $A = \{1, 2\}$, so gilt:

$$\mathcal{P}(\{1, 2\}) = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\}.$$

An dieser Stelle solle man sich nochmal bewusst werden, dass die Zahl 2 und die Menge $\{2\}$ nicht dasselbe sind. So gilt $2 \in A$, aber $2 \notin \mathcal{P}(A)$. Hingegen ist $\{2\} \notin \{1, 2\}$, sondern $\{2\} \subset \{1, 2\}$, denn jedes Element von $\{2\}$ - und das ist einzig die Zahl 2 - ist auch Element der Menge $\{1, 2\}$ ist. Deswegen gilt $\{2\} \in \mathcal{P}(A)$, aber $\{2\} \not\subset \mathcal{P}(A)$, denn nicht alle Elemente von $\{2\}$ - in diesem Fall geht es um das einzige Element 2 - sind Elemente von $\mathcal{P}(A)$. Hingegen gilt $\{\{2\}\} \subset \mathcal{P}(A)$, denn das einzige Element der Menge $\{\{2\}\}$ - nämlich die Menge $\{2\}$ - ist auch ein Element von $\mathcal{P}(A)$.

Die Potenzmengen von größeren werden schnell größer; darauf werden wir später nochmal genauer eingehen. Als Beispiel sei die Potenzmenge von $\{1, 2, 3\}$,

$$\mathcal{P}(\{1, 2, 3\}) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

angeführt, die bereits 8 Elemente hat. Manche Potenzmengen kann man zunächst nicht weiter beschreiben. So ist $\mathcal{P}(\mathbb{N}) = \{A \,|\, A \subset \mathbb{N}\}$ schon eine Beschreibung, mit der man sich zufriedengeben muss. Trotzdem kann man natürlich mit diesem Objekt arbeiten; beispielsweise können wir sagen, dass $\varnothing \in \mathcal{P}(\mathbb{N}), \{1, 2, 3\} \in \mathcal{P}(\mathbb{N}), \mathbb{N} \in \mathcal{P}(\mathbb{N})$; auch können wir sehen, dass

$$\{\{1\}, \{2\}, \{7, 9, 10\}\} \subset \mathcal{P}(\mathbb{N}).$$

Ein sehr intuitiver Begriff ist die **Anzahl** der Elemente einer Menge A. Wir notieren diese mit $|A|$ oder $\#A$.

Beispiel.
- $|\{1, 2, 4\}| = 3$,
- $|\mathcal{P}(\{1, 2\})| = 4$, wie wir im vorherigen Beispiel gesehen haben,
- $|\varnothing| = 0$,

- Die Anzahl der Elemente von $\mathbb{N}$ kann durch keine natürliche Zahl angegeben werden - diese Anzahl ist unendlich. Wir schreiben dafür $|\mathbb{N}| = \infty$. Im Prinzip gibt es auch Methoden, um mit ∞ zu hantieren und auch, um verschiedene Unendlichkeiten zu unterscheiden, doch uns soll vorerst der einfachste Begriff des Unendlichen genügen.

2.2 Mengenoperationen

In diesem Abschnitt lernen wir einige Methoden kennen, um aus vorgegebenen Mengen neue Mengen zu konstruieren.

Definition 2.5. Seien A, B Mengen. Die **Vereinigung** von Mengen A und B (in Zeichen: $A \cup B$) ist die Menge, die genau alle Elemente von A und alle Elemente von B enthält. In Zeichen:

$$(x \in A \cup B) :\Leftrightarrow ((x \in A) \vee (x \in B)).$$

Beispiel.

- $\{1, 2, 3\} \cup \{3, 4, 5\} = \{1, 2, 3, 4, 5\}$, insbesondere werden Mehrfachnennungen zusammengefasst.
- Die Objekte der jeweiligen Mengen müssen nichts miteinander zu tun haben: Sind x, y Symbole, so können wir auch $\{x, y\} \cup \{1, 2\} = \{x, y, 1, 2\}$ bilden.
- Manchmal lässt sich die Vereinigung zweier Mengen nicht in vereinfachter Form hinschreiben: $\mathbb{N} \cup \{\sqrt{2}, \sqrt{3}\}$ ist die einfachste Art, diese Menge hinzuschreiben.
- Ist B eine Teilmenge von einer Menge A, so gilt $B \cup A = A$. Denn jedes Element von B ist bereits in A enthalten, d.h. jedes Element von A oder von B ist insbesondere ein Element von A. Andersrum ist jedes Element von A ein Element von A oder ein Element von B.
- Insbesondere gilt für jede Menge A: $A \cup A = A$ und $A \cup \varnothing = A$.

Definition 2.6. Seien A, B Mengen. Der **Durchschnitt** der Mengen A und B (in Zeichen: $A \cap B$) ist die Menge, die genau alle Elemente enthält, die sowohl in A als auch in B enthalten sind. In Zeichen:

$$(x \in A \cap B) :\Leftrightarrow ((x \in A) \wedge (x \in B)).$$

Beispiel.

- $\{1, 2, 3\} \cap \{3, 4, 5\} = \{3\}$, denn 3 ist das einzige Element, das in beiden Mengen vorkommt.
- Sind x, y Symbole, so gilt $\{x, y\} \cap \{1, 2\} = \varnothing$.
- Für jede Menge A gilt: $A \cap A = A$.

- Wir bezeichnen das **abgeschlossene Intervall** mit eckigen Klammern und das **offene Intervall** mit runden Klammern, z.B.

$$\begin{aligned}[3,5] &= \{x \in \mathbb{R} \mid 3 \leq x \leq 5\} \\ (3,5) &= \{x \in \mathbb{R} \mid 3 < x < 5\}\end{aligned}$$

  Hier haben wir $(3,5) \cap \mathbb{N} = \{4\}$ und $[3,5] \cap \mathbb{N} = \{3,4,5\}$.

Definition 2.7. Seien A, B Mengen. Die **Mengendifferenz** der Mengen A und B (in Zeichen: $A \setminus B$) ist die Menge, die genau alle Elemente enthält, die in A, aber nicht in B enthalten sind. In Zeichen:

$$(x \in A \setminus B) :\Leftrightarrow ((x \in A) \wedge (x \notin B)).$$

Ist B eine Teilmenge von einer Menge A, so wird $A \setminus B$ auch das **Komplement von B in A** genannt.

Beispiel.
- Im Allgemeinen muss keine der beiden Mengen Teilmenge der anderen sein, um Mengendifferenz zu bilden. So ist beispielsweise $\{1,2,3\} \setminus \{3,4,5\} = \{1,2\}$.
- Manchmal kann man eine Mengendifferenz nicht weiter beschreiben, etwa $\mathbb{R} \setminus \{1,2\}$. Solche Mengen trifft man häufig schon in der Schule als Definitionsbereiche von Funktionen.
- $[2,4] \setminus (2,4) = \{2,4\}$.

Definition 2.8. Zwei Mengen A, B heißen **disjunkt**, falls $A \cap B = \varnothing$. Eine Ansammlung $A_1, \ldots, A_n$ von Mengen heißt **paarweise disjunkt**, falls je zwei von diesen Mengen disjunkt sind; mit anderen Worten, die Mengen A_i, A_j sind für $1 \leq i, j \leq n$, $i \neq j$, disjunkt.

Um die verschiedenen Mengenoperationen besser zu verstehen, beweisen wir exemplarisch die folgende Proposition.

Proposition 2.9. *Seien A, B beliebige Mengen. Dann gilt:*

1. $A = (A \setminus B) \cup (A \cap B)$,
2. $(A \setminus B) \cap (A \cap B) = \varnothing$.

Beweis. 1. Zwei Mengen sind nach Definition gleich, wenn sie dieselben Elemente haben. Wir zeigen nun, dass ein Objekt x genau dann in A liegt, wenn es in $(A \setminus B) \cup (A \cap B)$ liegt. Wir fangen mit der letzteren Aussage an. Nach Definition von Vereinigung gilt:

$$x \in (A \setminus B) \cup (A \cap B) \quad \Leftrightarrow \quad ((x \in A \setminus B) \vee (x \in A \cap B)).$$

Wir nutzen nun die Definitionen von Mengendifferenz und vom Durchschnitt:

$$((x \in A \setminus B) \vee (x \in A \cap B)) \quad \Leftrightarrow \quad (x \in A \wedge x \notin B) \vee (x \in A \wedge x \in B).$$

Auf diese Aussagen kann man nun das Distributivgesetz anwenden:

$$(x \in A \wedge x \notin B) \vee (x \in A \wedge x \in B) \Leftrightarrow (x \in A) \wedge (x \notin B \vee x \in B).$$

Da die Aussage $x \notin B \vee x \in B$ stets wahr ist, folgt

$$(x \in A) \wedge (x \notin B \vee x \in B) \Leftrightarrow (x \in A).$$

Nimmt man die Kette der Äquivalenzen zusammen, so haben wir gezeigt, dass $x \in (A \setminus B) \cup (A \cap B)$ genau dann wahr ist, wenn $x \in A$ gilt. Damit haben wir die Gleichheit der Mengen bewiesen.

2. Wir gehen wie im ersten Teil vor: Wir zeigen, dass die Aussage $x \in (A \setminus B) \cap (A \cap B)$ immer falsch ist und die Menge somit dieselben Elemente wie die leere Menge hat, nämlich gar keine. Wir benutzen zunächst die Definition vom Durchschnitt:

$$x \in (A \setminus B) \cap (A \cap B) \Leftrightarrow ((x \in A \setminus B) \wedge (x \in A \cap B)).$$

Im nächsten Schritt setzen wir die Definitionen von Mengendifferenz und vom Durchschnitt wieder ein:

$$((x \in A \setminus B) \wedge (x \in A \cap B)) \Leftrightarrow (x \in A \wedge x \notin B) \wedge (x \in A \wedge x \in B).$$

Nun nutzen wir das Assoziativ- und das Kommutativgesetz für die Und-Verknüpfung und erhalten:

$$(x \in A \wedge x \notin B) \wedge (x \in A \wedge x \in B) \Leftrightarrow (x \notin B \wedge x \in B) \wedge (x \in A \wedge x \in A).$$

Nun ist die Aussage $x \notin B \wedge x \in B$ immer falsch, und somit die gesamte Aussage $(x \notin B \wedge x \in B) \wedge (x \in A \wedge x \in A)$ immer falsch. Somit folgt die Behauptung.

□

Wir führen die vorerst letzte Mengenoperation ein.

Definition 2.10. Seien A, B Mengen. Das **(kartesische) Produkt** $A \times B$ der Mengen A und B ist die Menge aller (geordneter) Paare (a, b) mit $a \in A$ und $b \in B$. In Zeichen:

$$A \times B = \{(a, b) | a \in A, b \in B\}.$$

Man beachte, dass ein Paar stets geordnet ist, wir sagen nur manchmal „geordnete Paare“, um das nochmal zusätzlich zu unterstreichen.

Beispiel.
- Ein Beispiel ist meist aus der Schule bekannt: Die Menge von Paaren (x, y) mit $x \in \mathbb{R}$ und $y \in \mathbb{R}$ wird meist als $\mathbb{R}^2$ geschrieben und wird benutzt, um Koordinaten von Punkten in der Ebene anzugeben.
- $\{1, 2\} \times \{a, b\} = \{(1, a), (1, b), (2, a), (2, b)\}$. Man beachte, dass dies nicht dieselbe Menge ist wie

$$\{a, b\} \times \{1, 2\} = \{(a, 1), (b, 1), (a, 2), (b, 2)\}.$$

Wir wollen auch wissen, wie man die Anzahl der Elemente einer Menge ermittelt, die man aus vorgegebenen Mengen durch Verwendung von Mengenoperationen bekommen hat. Eine erste Aussage in diese Richtung ist die folgende.

Proposition 2.11. *Seien A, B endliche Mengen. Dann gilt:*

1. $|A \cup B| = |A| + |B| - |A \cap B|$,
2. *Sind A und B disjunkt, so ist $|A \cup B| = |A| + |B|$.*

Beweis.
1. Nach Definition enthält $A \cup B$ genau die Elemente, die in A oder in B enthalten sind. Insgesamt ergibt es zunächst $|A| + |B|$ Elemente, allerdings werden manche Elemente hierbei doppelt gezählt, nämlich diejenigen, die sowohl in A als auch in B enthalten sind. Die Menge solcher Elemente ist nach Definition genau $A \cap B$. Berücksichtigt man die Mehrfachnennungen, so erhält man also genau $|A \cup B| = |A| + |B| - |A \cup B|$.
2. Nach Definition ist der Durchschnitt zweier disjunkten Mengen leer und hat somit 0 Elemente. Somit folgt der zweite Teil unmittelbar aus dem ersten.

□

Daraus erhalten wir als Folgerung:

Korollar 2.12. *Seien $A_1, \ldots, A_n$ paarweise disjunkte, endliche Mengen. Dann gilt:*

$$|A_1 \cup A_2 \cup \ldots \cup A_n| = |A_1| + \ldots + |A_n|.$$

3 Vollständige Induktion

Will man eine Aussage für alle natürlichen Zahlen beweisen, so steht man zunächst vor einem Problem, denn es sind unendlich viele Zahlen und eine Aussage, die für $1, 2, 3, 4, 5$ gilt, braucht nicht für 6 zu gelten. Einen Ausweg bietet das Induktionsprinzip. Bevor wir dieses genauer formulieren, beschäftigen wir uns mit einem Beispiel.

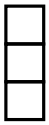
Frage. Sei ein 1024×1024-großes Quadrat vorgegeben, das in 1×1-Kästchen unterteilt ist, wie kariertes Papier. Ist es möglich, dieses Quadrat mit L-Teilchen (s. Bild) zu pflastern, sodass keine zwei sich überlappen, kein Feld frei bleibt und jedes Kästchen von jedem L-Teilchen auf einem Kästchen vom großen Quadrat liegt? (Wir sagen dafür in Zukunft nur noch „pflastern“, diese Bedingungen werden dabei stets angenommen. Das L-Teilchen darf aber gedreht werden.)

Diese erste Frage ist leicht zu beantworten: Wäre das Quadrat mit L-Teilchen gepflastert, so wäre es insbesondere in 3er-Gruppen unterteilt, also müsste die Anzahl der Kästchen in dem Quadrat durch 3 teilbar sein. Auf der anderen Seite ist aber $1024 \cdot 1024 = 2^{20}$ nicht durch 3 teilbar, also kann die Pflasterung nie aufgehen.

Wir modifizieren also die Frage:

Frage. Kann man ein 1024×1024-Quadrat ohne eine Ecke mit L-Teilchen pflastern?

Das Nachrechnen zeigt, dass $2^{20} - 1$ tatsächlich durch 3 teilbar ist. Allerdings kann die Teilbarkeit durch 3 nicht das einzige Kriterium sein, denn das folgende Gebilde kann man auch nicht mit L-Teilchen pflastern:



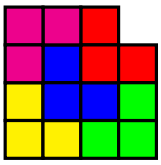
Da 1024 zu groß zum Ausprobieren ist, fragen wir uns, wie es denn für kleinere Potenzen von 2 aussieht.

Frage. Kann man ein 2×2-Quadrat ohne eine Ecke mit L-Teilchen pflastern?

Die Antwort auf diese Frage ist ziemlich klar: Ein 2×2-Quadrat ohne eine Ecke ist genau ein L-Teilchen; die Antwort lautet also „ja“. Wenden wir uns dem nächsten Fall zu:

Frage. Kann man ein 4×4-Quadrat ohne eine Ecke mit L-Teilchen pflastern?

Diese Frage ist schon schwieriger, allerdings auch positiv zu beantworten. Man hat beispielsweise die folgende Pflasterung:



Es ist noch deutlich aufwendiger, den 8×8-Fall zu untersuchen. Macht man sich die Mühe, so stellt man auch dort fest, dass eine Pflasterung möglich ist. Wir wollen also den allgemeinen Satz beweisen:

Satz 3.1. *Man kann ein $2^n \times 2^n$-Quadrat ohne eine Ecke für jede natürliche Zahl $n \geq 1$ mit L-Teilchen pflastern.*

Die Hauptidee wird wie folgt sein: Hat man ein $2^{n+1} \times 2^{n+1}$-Quadrat, so kann man dieses in vier $2^n \times 2^n$-Quadrate zerschneiden, indem man jede Seite halbiert. Dann werden wir unser bereits erlangtes Können - nämlich zupflastern von kleineren Quadraten ohne Ecke - anwenden können, um das große Quadrat zu pflastern. Dabei muss das Vorgehen noch etwas genauer erläutert werden. Dass eine solche Methode überhaupt funktioniert, wird durch das sogenannte *Induktionsaxiom* gesichert. Es ist ein Axiom, also eine Aussage, die wir als wahr annehmen, ohne diese zu beweisen. Ohne Axiome - also grundlegende Aussagen, die nicht bewiesen werden - können wir kaum etwas beweisen, ähnlich wie wir ohne Bedeutung von gewissen Worten als bekannt vorauszusetzen nicht die Bedeutung von anderen Worten erklären können. Man versucht allerdings im Allgemeinen in der Mathematik, so wenige Axiome wie möglich zu verwenden, und diese Annahmen, die nicht bewiesen werden, möglichst einfach und plausibel zu wählen. Wir benutzen die folgenden beiden Formulierungen.

Axiom (Induktionsaxiom I)**.** Sei $X \subseteq \mathbb{N}_0$ eine Teilmenge von natürlichen Zahlen. Wir nehmen an, dass die folgenden zwei Bedingungen für X gelten:

(I1) $0 \in X$ und

(I2) Ist eine natürliche Zahl k ein Element von X, so ist auch ihr Nachfolger $k + 1$ in X.

Dann folgt bereits, dass $X = \mathbb{N}_0$ ist, also X alle natürlichen Zahlen enthält.

Die zweite Version ist nur geringfügig anders:

Axiom (Induktionsaxiom II)**.** Sei $X \subseteq \{x \in \mathbb{N}_0 | x \geq m\}$ eine Teilmenge von natürlichen Zahlen, die größer oder gleich der festen natürlichen Zahl m sind. Wir nehmen an, dass die folgenden zwei Bedingungen für X gelten:

(I1) $m \in X$ und

(I2) Ist eine natürliche Zahl k ein Element von X, so ist auch ihr Nachfolger $k+1$ in X.

Dann folgt bereits, dass $X = \{x \in \mathbb{N}_0 \,|\, x \geq m\}$ ist, also X alle natürlichen Zahlen ab m enthält.

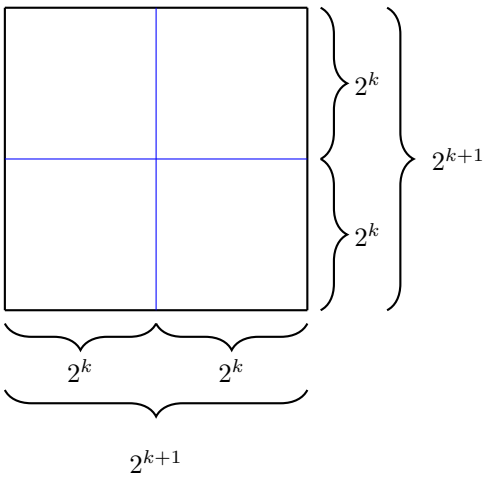
Will man nun eine Behauptung $B(n)$ für alle natürlichen Zahlen ab einer bestimmten natürlichen Zahl m beweisen, so braucht man nur überprüfen, ob $B(m)$ wahr ist und ob die Aussage $B(k) \Rightarrow B(k+1)$ wahr ist. Mit anderen Worten: Wir müssen überprüfen, ob die Aussage für den kleinsten gewünschten Wert m gilt, und dann, ob die Gültigkeit der Aussage für eine natürliche Zahl $k \geq m$ auch die Gültigkeit der Aussage für die nachfolgende natürliche Zahl $k+1$ impliziert. Dann folgt nach dem Induktionsaxiom die Gültigkeit der Aussage $B(n)$ für alle natürlichen Zahlen ab m. (Um darauf das Induktionsaxiom anzuwenden, betrachten wir nämlich die Menge $X = \{x \in \mathbb{N}_0 \,|\, x \geq m \text{ und } B(x) \text{ wahr}\}$.)

Nun können wir diese neue Technik anwenden, um den Satz 3.1 zu beweisen.

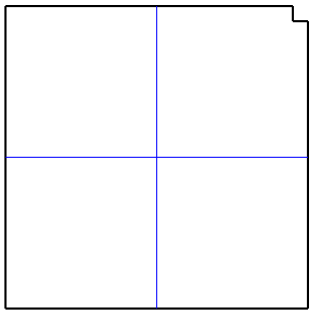
Beweis des Satzes 3.1. Wir beweisen den Satz durch Induktion nach n.

Induktionsanfang: Zunächst müssen wir zeigen, dass die Aussage für den kleinsten relevanten Wert wahr ist. Da wir die Existenz der Pflasterung für alle $n \geq 1$ beweisen wollen, müssen wir für den Induktionsanfang die Aussage für $n = 1$ beweisen. Das haben wir oben bereits getan: Ein 2×2-Quadrat ohne Ecke ist genau das L-Teilchen, und somit lässt sich dieses durch (genau eines) L-Teilchen pflastern. Also ist die Aussage des Satzes wahr für $n = 1$.

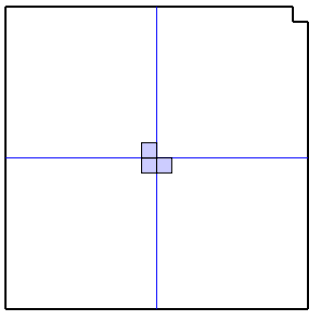
Induktionsschritt: Im Induktionsschritt muss die folgende Aussage gezeigt werden: Gilt die Behauptung des Satzes für eine feste natürliche Zahl $k \geq 1$, so gilt sie auch für Ihren Nachfolger $k+1$. Wir nehmen also an (ohne es bislang zu wissen!), dass die Aussage für $k \geq 1$ wahr ist. Diese Annahme nennen wir **Induktionsvoraussetzung**. Unter dieser Annahme zeigen wir dann, dass die Aussage auch für $k+1$ wahr ist. Konkret heißt es in diesem Fall: Wir nehmen an, wir hätten bereits eine Pflasterung für ein $2^k \times 2^k$-Quadrat ohne Ecke, das ist die Induktionsvoraussetzung. Wir wollen diese nun nutzen, um zu zeigen, dass ein $2^{k+1} \times 2^{k+1}$-Quadrat ohne Ecke ebenfalls mit L-Teilchen gepflastert werden kann. Dafür ist es ganz nützlich, wenn man ein Objekt, das in der Induktionsvoraussetzung vorkommt, in dem größeren Objekt wiederfinden können. Hierfür ist die folgende Beobachtung zentral: Teilt man die Seiten des $2^{k+1} \times 2^{k+1}$-Quadrats jeweils in zwei Hälften auf, so erhält man vier Quadrate der Größe $2^k \times 2^k$:



Macht man dies mit dem $2^{k+1} \times 2^{k+1}$-Quadrat ohne Ecke, so bekommt man ein $2^k \times 2^k$-Quadrat ohne Ecke und drei „unbeschädigte“ $2^k \times 2^k$-Quadrate:



Man weiß *nach Induktionsvoraussetzung* bereits, wie man das kleinere Quadrat ohne Ecke pflastert und muss sich für die restlichen, ‘unbeschädigten“ $2^k \times 2^k$-Quadrate etwas überlegen. Da wir daraus auch Quadrate ohne eine Ecke machen wollen, stellt es sich als sinnvoll heraus, ein L-Teilchen in die Mitte zu legen, sodass es ein Ecke aus jedem dieser drei kleineren Quadrate herausnimmt:



Dadurch entstehen drei weitere $2^k \times 2^k$-Quadrate ohne Ecke, die wir *nach Induktionsvoraussetzung* bereits pflastern können. Damit haben wir aber - unter unserer Annahme, das $2^k \times 2^k$-Quadrat ohne Ecke pflastern zu können

- das $2^{k+1} \times 2^{k+1}$-Quadrat ohne Ecke mit L-Teilchen gepflastert. Das vervollständigt den Induktionsschritt. Somit greift das Induktionsprinzip, und nach dem Induktionsaxiom ist die Aussage bewiesen. □

Wir wollen im Folgenden an einigen Beispielen weiter verdeutlichen, wie Induktionsbeweise funktionieren. Dabei haben die Beweise stets ähnliches Grundmuster, aber im Induktionsschritt (und manchmal schon beim Induktionsanfang) werden stets Methoden herangezogen, die für die Aussage spezifisch sind.

Proposition 3.2. *Für alle natürlichen Zahlen $n \geq 1$ gilt: $2^n > n$.*

Beweis. Wir beweisen die Proposition durch Induktion nach n.

Induktionsanfang: Für $n = 1$ haben wir die Aussage $2^1 > 1$ zu prüfen. Diese ist offensichtlich wahr.

Induktionsschritt: Für den Induktionsschritt müssen wir die folgende Aussage beweisen: Gilt für eine natürliche Zahl $k \geq 1$, dass $2^k > k$, so folgt daraus auch $2^{k+1} > k + 1$. Dafür nutzen wir die Umformungsregeln für Ungleichungen, die wir kennengelernt haben. Dabei starten wir mit der *Induktionsvoraussetzung* und multiplizieren beide Seiten mit 2:

$$\begin{aligned} & 2^k > k && \mid \cdot 2 \\ \Rightarrow \quad & 2^{k+1} > 2k && . \end{aligned}$$

Das ist möglich, da Multiplikation mit einer positiven reellen Zahl die Gültigkeit einer Ungleichung nicht ändert. Nun schreiben wir $2k = k+k$ und addieren zu beiden Seiten der Voraussetzung $k \geq 1$ die reelle Zahl k, was die Gültigkeit der Ungleichung nicht ändert und die neue Ungleichung $k+k \geq k+1$ liefert. Zusammengesetzt erhalten wir

$$2^{k+1} > 2k \geq k + 1,$$

was $2^{k+1} > k+1$ impliziert. Das vervollständigt den Induktionsschritt. Daraus folgt mit dem Induktionsaxiom die Behauptung.

□

Die nächste Proposition ist eine der berühmtesten Anwendungen der Induktion.

Proposition 3.3. *Für alle natürlichen Zahlen $n \geq 1$ gilt:*

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

Dabei ist mit der Summe auf der linken Seite die Summe aller natürlichen Zahlen von 1 bis einschließlich n gemeint. Um das präziser handhaben zu können, führen wir die folgende Notation ein:

Notation. Ist $P(k)$ ein Term in Abhängigkeit von einer natürlichen Zahl k, so schreiben wir die Summe der Werte von P für natürliche Zahlen zwischen m und n als

$$\sum_{k=m}^{n} P(k).$$

Beispiel.

- $\sum_{i=1}^{7} i$ steht für die Summe der natürlichen Zahlen zwischen 1 und 7, also

$$\sum_{i=1}^{7} i = 1+2+3+4+5+6+7 = 28.$$

- $\sum_{i=3}^{4}(i^2+7i) = (3^2+7\cdot 3)+(4^2+7\cdot 4) = 74.$
- $\sum_{i=0}^{3} 2^i = 2^0+2^1+2^2+2^3 = 15.$

Mit dieser Notation lässt sich nun die obige Proposition schreiben als $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Diese Aussage wollen wir nun beweisen.

Beweis der Proposition 3.3. Wir beweisen die Proposition durch Induktion nach n.

Induktionsanfang: Für $n = 1$ haben wir die Aussage $1 = \frac{1\cdot(1+1)}{2}$ zu prüfen. Diese ist wahr, wie man leicht nachrechnet.

Induktionsschritt: Für den Induktionsschritt müssen wir die folgende Aussage beweisen: Gilt für eine natürliche Zahl $m \geq 1$ die Aussage:

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

so folgt daraus auch $\sum_{i=1}^{m+1} i = \frac{(m+1)((m+1)+1)}{2}$.

Diesmal fangen wir mit der linken Seite der Behauptung an und formen diese um. Erste wichtige Beobachtung bereitet die Anwendung der Induktionsvoraussetzung vor.

$$\begin{aligned}\sum_{i=1}^{m+1} i &= 1+2+\ldots+m+(m+1)\\ &= (1+2+\ldots+m)+(m+1)\\ &= \sum_{i=1}^{m} i+(m+1).\end{aligned}$$

Die Summe der natürlichen Zahlen von 1 bis m können wir *nach Induktionsvoraussetzung* bereits durch $\frac{m(m+1)}{2}$ ersetzen. Danach klammern wir den Faktor $m+1$ aus.

$$\sum_{i=1}^{m} i + (m+1) \overset{\text{IV}}{=} \frac{m(m+1)}{2} + (m+1) = (m+1)\left(\frac{m}{2}+1\right).$$

Nun vereinfachen wir den Term in der zweiten Klammer und formen den um, bis wir das gewünschte Ergebnis bekommen haben. .

$$(m+1)\left(\frac{m}{2}+1\right) = (m+1)\cdot\frac{m+2}{2} = \frac{(m+1)((m+1)+1)}{2}.$$

Insgesamt haben wir also durch Termumformungen unter Zuhilfenahme der Induktionsvoraussetzung gezeigt:

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

Damit ist der Induktionsschritt vollständig. Die Behauptung folgt nun nach dem Induktionsaxiom. □

In der nächsten Proposition soll es darum gehen, die Summe der ungeraden natürlichen Zahlen bis zu einer gewissen Zahl zu bestimmen. Bevor wir das tun können, müssen wir uns überlegen, wie man die ungeraden natürlichen Zahlen ausdrücken kann. Die i-te ungerade natürliche Zahl lässt sich als $2i-1$ schreiben, z.B. ist die erste ungerade natürliche Zahl $2\cdot 1-1=1$, die zweite $2\cdot 2-1=3$, die zweiundzwanzigste $2\cdot 22-1=43$. Damit lässt sich nun die folgende Proposition formulieren.

Proposition 3.4. *Die Summe der ersten k ungerader natürlicher Zahlen beträgt k^2. Als Formel:*

$$\sum_{j=1}^{k}(2j-1) = k^2$$

für alle natürlichen Zahlen $k \geq 1$.

Beweis. Wir beweisen die Proposition wieder mittels vollständiger Induktion nach k.

Induktionsanfang: Für $k=1$ haben wir die Aussage $\sum_{j=1}^{1}(2j-1) = 1^2$ zu prüfen. Die Summe lässt sich als $2\cdot 1-1$ schreiben, und dieser Ausdruck hat tatsächlich den Wert 1.

Induktionsschritt: Hier müssen wir zeigen, dass aus der Gültigkeit der Gleichung $\sum_{j=1}^{k}(2j-1) = k^2$ für eine natürliche Zahl k auch die Gleichung

$$\sum_{j=1}^{k+1}(2j-1) = (k+1)^2$$

folgt. Wieder fangen wir mit der linken Seite an und merken, dass die Summe um einen einzigen Summanden im Vergleich zur *Induktionsvoraussetzung* erweitert worden ist. Also können wir wieder die Induktionsvoraussetzung anwenden und erhalten:

$$\sum_{j=1}^{k+1}(2j-1) = \sum_{j=1}^{k}(2j-1) + (2(k+1)-1) \overset{\text{IV}}{=} k^2 + (2(k+1)-1).$$

Nach Auflösen der Klammern sehen wir sofort, dass wir die 1-te binomische Formel anwenden können:

$$k^2 + (2(k+1)-1) = k^2 + 2k + 1 = (k+1)^2.$$

Das zeigt die Behauptung im Induktionsschritt und somit die Behauptung der Proposition. □

Das nächste Beispiel für die Anwendung der Induktion ist von etwas anderer Natur. Um dieses genauer behandeln zu können, wollen wir nochmal den Begriff von Teilbarkeit präziser machen.

Definition 3.5. Wir sagen, die natürliche Zahl $k \neq 0$ **teilt** die natürliche Zahl m, falls $\frac{m}{k}$ eine ganze Zahl ist.

Damit können wir nun die folgende Proposition beweisen.

Proposition 3.6. *Die Zahl $8^m - 3^m$ ist für jede natürliche Zahl $m \geq 1$ durch 5 teilbar.*

Beweis. Wir beweisen die Proposition durch Induktion nach m.

Induktionsanfang: Für $m = 1$ müssen wir nachprüfen, ob $8^1 - 3^1$ durch 5 teilbar ist. Da $8^1 - 3^1 = 5$, ist die Aussage der Proposition für $m = 1$ wahr.

Induktionsschritt: Hier müssen wir beweisen, dass die Zahl $8^{m+1} - 3^{m+1}$ durch 5 teilbar sein muss, falls wir bereits wissen, dass $8^m - 3^m$ durch 5 teilbar ist. Wir benutzen die obige Definition der Teilbarkeit. Wir müssen also prüfen, ob $\frac{8^{m+1}-3^{m+1}}{5}$ eine ganze Zahl ist. Dafür formen wir diesen Ausdruck etwas um. Als erstes fügen wir dabei einen Summanden $3 \cdot 8^m$ im Zähler hinzu und ziehen ihn wieder ab:

$$\frac{8^{m+1} - 3^{m+1}}{5} = \frac{8^{m+1} - 3 \cdot 8^m + 3 \cdot 8^m - 3^{m+1}}{5}$$

Wir benutzen nun, dass $8^{m+1} = 8 \cdot 8^m$ und $3^{m+1} = 3 \cdot 3^m$, und klammern die Faktoren 8^m bzw. 3 jeweils aus:

$$\frac{8^{m+1} - 3 \cdot 8^m + 3 \cdot 8^m - 3^{m+1}}{5} = \frac{8^m \cdot (8-3) + 3 \cdot (8^m - 3^m)}{5}.$$

Nun teilen wir diese Summe auf:

$$\frac{8^m \cdot (8-3) + 3 \cdot (8^m - 3^m)}{5} = 8^m + 3 \cdot \frac{8^m - 3^m}{5}.$$

Nach *Induktionsvoraussetzung* wissen wir, dass $8^m - 3^m$ durch 5 teilbar ist, also folgt aus der Induktionsvoraussetzung, dass $\frac{8^m-3^m}{5}$ eine ganze Zahl ist. Das bleibt genauso, wenn wir diese Zahl mit 3 multipliziert wird, und auch, wenn wir eine weitere ganze Zahl 8^m dazu addieren. Also ist

$$\frac{8^{m+1} - 3^{m+1}}{5} = 8^m + 3 \cdot \frac{8^m - 3^m}{5}$$

eine ganze Zahl, und folglich haben wir aus der Induktionsvoraussetzung hergeleitet, dass $8^{m+1} - 3^{m+1}$ durch 5 teilbar ist. Also ist der Induktionsschritt vollständig und die Proposition damit bewiesen.

□

4 Abzählen I

Wir wollen nun einige Methoden lernen, um die Anzahl von Objekten einer Menge (die beispielsweise durch Mengenoperationen, die wir kennengelernt haben, aus anderen Mengen bekannter Größe entstanden ist). Ein Beispiel hierfür haben wir bereits in der Proposition 2.11 bereits gesehen. Ein weiteres Resultat, das man leicht einsieht, ist das folgende:

Proposition 4.1. *Seien A und B endliche Mengen, so gilt: $|A\times B| = |A|\cdot|B|$.*

Beweis. Listet man die Elemente von $A \times B$ auf, also Paare von Elementen (a, b) mit $a \in A$ und $b \in B$, so erhält man für jedes Element von A genau $|B|$ Elemente, eins für jedes Element von B. Da das für jedes Element von A gilt, erhält man insgesamt $|A| \cdot |B|$. □

Nun wollen wir eine weitere, deutlich schwierigere Aussage von diesem Typ mit Hilfe von vollständiger Induktion beweisen.

Satz 4.2. *Sei A eine n-elementige Menge. Dann hat die Potenzmenge von A, $\mathcal{P}(A)$, genau 2^n Elemente. Mit anderen Worten: Die Menge A hat genau 2^n Teilmengen.*

Beweis. Wir beweisen die Aussage wieder durch Induktion nach n.

Induktionsanfang: Für $n = 0$ haben wir genau eine Menge, die 0 Elemente hat, nämlich die leere Menge $\varnothing$. Die leere Menge hat genau eine Teilmenge, und zwar die leere Menge selbst. Also gilt $\mathcal{P}(\varnothing) = \{\varnothing\}$ und insbesondere $|\mathcal{P}(\varnothing)| = 1$, was genau 2^0 ist. Die Aussage des Satzes stimmt also für $n = 0$.

Induktionsschritt: Nun müssen wir zeigen: Wenn für jede k-elementige Menge A gilt, dass $|\mathcal{P}(A)| = 2^k$ gilt (für ein festes k), so gilt auch für jede $k + 1$-elementige Menge B: $|\mathcal{P}(B)| = 2^{k+1}$. Sei also B eine $k + 1$-elementige Menge. Da k eine natürliche Zahl ist, ist $k \geq 0$ und folglich $k+1 \geq 1$, sodass wir wissen, dass die Menge B nicht die leere Menge ist. Insbesondere können wir uns ein Element $b \in B$ auswählen. Wir betrachten nun die folgenden beiden Teilmengen der Potenzmenge von B:

$$\begin{aligned} X &= \{C \subset B \,|\, b \in C\} \text{ und} \\ Y &= \{C \subset B \,|\, b \notin C\}. \end{aligned}$$

Da jede Teilmenge von B entweder das Element b enthält oder es nicht enthält, gilt

$$\mathcal{P}(B) = X \cup Y.$$

Da ferner nur eine von beiden Möglichkeit jeweils eintritt, sind die Mengen X und Y disjunkt. Daraus folgt nun mit Proposition 2.11:

$$|\mathcal{P}(B)| = |X| + |Y|.$$

Man muss also nur die Anzahlen von Elementen von X und von Y bestimmen, um $|\mathcal{P}(B)|$ zu berechnen. Dazu benötigen wir zwei Schritte. Zunächst bestimmen wir dabei die Anzahl der Elemente von Y, und danach zeigen wir, dass X und Y gleich viele Elemente haben.

Schritt 1: In diesem Schritt bestimmen wir die Anzahl der Elemente von Y. Dafür machen wir die folgende Beobachtung: Ist C eine Teilmenge von B, die b nicht enthält, so ist C auch eine Teilmenge von $B\setminus\{b\}$. Umgekehrt kann jede Teilmenge von $B \setminus \{b\}$ als eine Teilmenge von B gesehen werden, die b nicht enthält. Somit können wir sagen, dass $Y = \mathcal{P}(B \setminus \{b\})$ gilt, da wir uns gerade überlegt haben, dass beide Mengen dieselbe Elemente besitzen. Insbesondere folgt daraus $|Y| = |\mathcal{P}(B \setminus \{b\})|$. Da $b \in B$, hat die Menge $B\setminus\{b\}$ genau k Elemente, da wir ein Element aus der $(k+1)$-elementigen Menge B entfernt haben. Das versetzt uns in die Lage, auf die Menge $B \setminus \{b\}$ die *Induktionsvoraussetzung* anzuwenden. Darin haben wir angenommen, dass wir bereits die Anzahl der Elemente in der Potenzmenge einer k-elementigen Menge bestimmen können. Wenden wir das auf die k-elementige Menge $B \setminus \{b\}$ an, so erhalten wir:

$$|Y| = |\mathcal{P}(B \setminus \{b\})| = 2^k.$$

Die Menge Y hat also 2^k Elemente.

Schritt 2: In diesem zweiten Schritt wollen wir zeigen, dass X und Y gleich viele Elemente haben. Dabei gehen wir wie folgt vor: Wir geben zu jedem Element von X einen „Partner“ in Y an, und zeigen dann, dass jedes Element in X genau einen Partner bekommt, dass kein Element von Y zwei oder mehr Partner bekommen hat und dass auch jedes Element von Y an ein Element von X vergeben wurde. (Man könnte sich vorstellen, man würde eine Gruppe von Personen in Tanzpaare aufteilen, und wir müssen nachprüfen, dass die Aufteilung genau aufgeht.)

Jedem Element $C \in X$ wollen wir also ein Element von Y zuordnen. Wir starten also mit einem Element C von X. Nach Definition ist ein Element von X eine Teilmenge von B, die das Element b enthält. Wir wollen dieser ein Element von Y zuordnen, also eine Teilmenge von B, die das Element b nicht enthält. Dafür tun wir etwas naheliegendes: Wir nehmen die Teilmenge C von B und entfernen daraus das Element b, bilden also die Menge $C \setminus \{b\}$. Das ist wieder eine Teilmenge von B, und außerdem enthält sie das Element b nicht, denn das haben wir herausgenommen. Also ist $C \setminus \{b\} \in Y$. Wir haben also zu jedem Element von X einen Partner in Y gefunden. Jetzt müssen wir zwei Aussagen prüfen: Auf diese Weise haben wir jedem Element von X ein Element von Y zugeordnet. A priori ist es allerdings unklar, ob jedes Element von Y dabei vergeben wurde, und ob nicht manche Elemente von Y doppelt vergeben wurden.

Kann es also sein, dass wir zwei unterschiedliche Teilmengen C_1, C_2 von B genommen haben, die beide b enthalten, und durch das Entfernen des Elementes b die Mengen gleich wurden, also $C_1 \setminus \{b\} = C_2 \setminus \{b\}$ gilt? Das kann nicht vorkommen, denn mindestens eine von den Mengen C_1, C_2 muss ein Element enthalten, das nicht in der anderen liegt, sonst wären die Mengen gleich. Dieses Element kann aber nicht b sein, denn b ist sowohl in C_1 als auch in C_2 enthalten. Also ist dieses Element nach wie vor in genau einer der Mengen $C_1 \setminus \{b\}, C_2 \setminus \{b\}$ enthalten, und diese Mengen wären also unterschiedlich. Es ist also kein Element von Y doppelt vergeben worden.

Könnte es nun möglicherweise sein, dass manche Elemente aus Y gar nicht an Elemente von X vergeben wurden? Auch das kann nicht passieren: Ist $D \in Y$, so ist D nach Definition von Y eine Teilmenge von B, die das Element b nicht enthält. Wir suchen also ein Element von X, dem D zugeordnet wurde, also eine Teilmenge von B, die b enthält und die nach Entfernen von b zu der Menge D wird. Eine solche Menge ist durch $D \cup \{b\}$ gegeben. Diese Menge ist nach Definition ein Element von X, dessen Partner in Y genau D ist.

Somit haben wir nachgewiesen, dass X und Y genau dieselbe Anzahl von Elementen haben, also $|X| = |Y| = 2^k$.

Nimmt man nun die vorher hergeleitete Formel

$$|\mathcal{P}(B)| = |X| + |Y|$$

und setzt $|X| = |Y| = 2^k$ ein, so erhält man

$$|\mathcal{P}(B)| = 2^k + 2^k = 2^{k+1}.$$

Das vervollständigt den Induktionsschritt und zeigt somit die Behauptung. □

Man sollte insbesondere bemerken, dass die Größe der Potenzmenge in Abhängigkeit von der Größe der ursprünglichen Menge sehr schnell wächst, und zwar exponentiell.

In nächsten Schritt will man noch etwas präziser werden und die Anzahl der Teilmengen einer festen Größe in der vorgegebenen Menge ermitteln. Dafür führen wir Binomialkoeffizienten ein. Es werden dieselben sein, die vermutlich aus dem Schulunterricht bekannt sind, allerdings werden wir sie anders definieren. Wir sehen später, dass beide Definitionen äquivalent sind.

Definition 4.3. Die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge ist der Binomialkoeffizient $\binom{n}{k}$ (Sprechweise: „n über k").

Beispiel.
- Eine n-elementige Menge A hat genau eine 0-elementige Teilmenge, nämlich die leere Menge. Deswegen gilt $\binom{n}{0} = 1$ für alle $n \geq 0$.

- Eine n-elementige Menge A hat auch genau eine n-elementige Teilmenge, nämlich sich selbst. Daher gilt stets $\binom{n}{n} = 1$.
- Eine n-elementige Menge A hat für $n \geq 1$ genau n unterschiedliche einelementige Teilmengen, denn man kann jedes der Elemente von A als das einzige Element unserer Teilmenge aussuchen. Deswegen gilt $\binom{n}{1} = n$ für alle $n \geq 1$.
- Konkreter wollen wir noch $\binom{3}{2}$ bestimmen. Dafür betrachten wir die 3-elementige Menge $\{a, b, c\}$ und listen alle möglichen 2-elementigen Teilmengen von dieser auf: Das sind $\{a, b\}$, $\{a, c\}$, $\{b, c\}$. Also gilt $\binom{3}{2} = 3$.
- Die Binomialkoeffizienten können im *Pascalschen Dreieck* angeordnet werden. Die bis jetzt bestimmen Binomialkoeffizienten sowie die Binomialkoeffizienten der Form $\binom{4}{k}$ ergeben:

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

$$\binom{1}{0} = 1 \qquad \binom{1}{1} = 1$$

$$\binom{2}{0} = 1 \qquad \binom{2}{1} = 2 \qquad \binom{2}{2} = 1$$

$$\binom{3}{0} = 1 \qquad \binom{3}{1} = 3 \qquad \binom{3}{2} = 3 \qquad \binom{3}{3} = 1$$

$$\binom{4}{0} = 1 \qquad \binom{4}{1} = 4 \qquad \binom{4}{2} = 6 \qquad \binom{4}{3} = 4 \qquad \binom{4}{4} = 1$$

oder etwas gewohnter:

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

Im Pascalschen Dreieck ergibt die Summe der beiden Binomialkoeffizienten, die rechts und links oberhalb von einer Stelle stehen, genau den Binomialkoeffizienten an dieser Stelle. Das halten wir im folgenden Satz fest.

Satz 4.4. *Für alle natürlichen Zahlen n,k gilt:* $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$.

Da der Beweis dem des Satzes 4.2 ähnelt und recht aufwändig zum präzisen Aufschreiben ist, geben wir hier nur die Beweisidee für diesen Satz. Diese ist wichtig, weil solche und ähnliche Zählmethoden häufiger zum Beweis Identitäten dieser Art herangezogen werden.

Beweisidee. Wir gehen ganz ähnlich vor wie im Beweis des Satzes 4.2.

Die Zahl $\binom{n+1}{k}$ beschreibt die Anzahl der k-elementigen Teilmengen einer $n+1$-elementigen Menge, die wir für diesen Beweis B nennen. Da $n+1 \geq 1$, hat die Menge B mindestens ein Element. Wir betrachten ein festes Element $b \in B$. Jede k-elementige Teilmenge von B enthält b oder enthält b nicht. Zunächst fragen wir uns, wie viele k-elementige Teilmengen von B das Element b nicht enthalten. Diese sind genau die k-elementigen Teilmengen von $B \setminus \{b\}$, also einer n-elementigen Menge. Da gibt es nach Definition der Binomialkoeffizienten genau $\binom{n}{k}$ verschiedene k-elementige Teilmengen.

Als nächstes müssen wir also die k-elementigen Teilmengen zählen, die b enthalten. Nimmt man das Element b aus solcher Teilmenge heraus, so erhält man eine $(k-1)$-elementige Teilmenge von der n-elementigen Menge $B \setminus \{b\}$. Ähnlich wie im Satz 4.2 kann man sich überlegen, dass jeder k-elementigen Teilmenge von B, die b enthält, so genau eine $(k-1)$-elementige Teilmenge von $B \setminus \{b\}$ zugeordnet wird. Umgekehrt kann jeder $(k-1)$-elementigen Teilmenge von $B \setminus \{b\}$ eine k-elementige Teilmenge von B zugeordnet werden, die b enthält, einfach indem wir b als Element hinzufügen. Auf diese Weise kann man nachweisen, dass die Anzahl der k-elementigen Teilmengen von B, die b enthalten, genau $\binom{n}{k-1}$ ist.

Daraus folgt die Behauptung. □

Wie das Verfahren aus dieser Beweisidee etwas konkreter aussieht, soll am folgenden Beispiel veranschaulicht werden.

Beispiel. Die 3-elementigen Teilmengen der Menge $\{a, b, c, d, e\}$ sind:

- Die dreielementigen Teilmengen von $\{a, b, c, d, e\}$, die e nicht als Element enthalten:

  $$\{a, b, c\},\ \{a, b, d\},\ \{a, c, d\},\ \{b, c, d\},$$

  insgesamt $\binom{4}{3} = 4$ Mengen. (Diese Teilmengen sind insbesondere die dreielementigen Teilmengen der Menge $\{a, b, c, d\}$.

- Die dreielementigen Teilmengen von $\{a, b, c, d, e\}$, die das Element e enthalten:

  $$\{a, b, e\},\ \{a, c, e\},\ \{a, d, e\},\ \{b, c, e\},\ \{b, d, e\},\ \{c, d, e\},$$

  insgesamt $\binom{4}{2} = 6$ Mengen. (Diese Teilmengen entstehen aus den 2-elementigen Teilmengen der Menge $\{a, b, c, d\}$ durch Hinzufügen des Elementes e.)

Die Binomialkoeffizienten werden in der folgenden Verallgemeinerung der ersten binomischen Formel benutzt.

Satz 4.5 (Binomischer Lehrsatz)**.** *Für alle $x, y \in \mathbb{R}$ und alle natürlichen Zahlen n gilt:*

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

Bevor wir eine Beweisidee für diesen Satz geben, wollen wir uns vergegenwärtigen, in wie fern dies eine Verallgemeinerung der ersten binomischen Formel ist. Dafür setzen wir $n = 2$ ein und erhalten, indem wir die Summe wieder ausschreiben:

$$(x+y)^2 = \sum_{k=0}^{2} \binom{2}{k} x^k y^{2-k} = \binom{2}{0} x^0 y^{2-0} + \binom{2}{1} x^1 y^{2-1} + \binom{2}{2} x^2 y^{2-2}.$$

Setzt man die vorher bestimmten Werte für Binomialkoeffizienten ein, so erhalten wir

$$\begin{aligned} \binom{2}{0} x^0 y^{2-0} + \binom{2}{1} x^1 y^{2-1} + \binom{2}{2} x^2 y^{2-2} &= 1 \cdot 1 \cdot y^2 + 2 \cdot xy + 1 \cdot x^2 \cdot 1 \\ &= y^2 + 2xy + x^2, \end{aligned}$$

also genau die erste binomische Formel.

Wir wollen, um die Funktionsweise der Formel besser zu verstehen, uns auch den Fall $n = 3$ ansehen. Genauso wie zuvor erhalten wir

$$\begin{aligned} (x+y)^3 &= \sum_{k=0}^{3} \binom{3}{k} x^k y^{3-k} \\ &= \binom{3}{0} x^0 y^{3-0} + \binom{3}{1} x^1 y^{3-1} + \binom{3}{2} x^2 y^{3-2} + \binom{3}{3} x^3 y^{3-3} \\ &= 1 \cdot 1 \cdot y^3 + 3 \cdot xy^2 + 3 \cdot x^2 y + 1 \cdot x^3 \cdot 1 \\ &= y^3 + 3xy^2 + 3x^2 y + x^3. \end{aligned}$$

Allgemein können wir den binomischen Lehrsatz etwas informeller in der Form

$$(x+y)^n = \binom{n}{0} x^0 y^n + \binom{n}{1} x^1 y^{n-1} + \ldots + \binom{n}{n-1} x^{n-1} y^1 + \binom{n}{n} x^n y^0$$

schreiben.

Um uns auf die Beweisidee vorzubereiten, überlegen wir uns, wie man Klammern simultan ausmultipliziert. Die grundsätzliche Methode ist dieselbe, wie beim üblichen Ausmultiplizieren von Klammern: Für jeden neuen Summanden vom Ergebnis nimmt man jeweils einen Summanden aus jeder Klammer, und jede Kombination von Summanden betrachtet man genau einmal.

So erhalten wir durch ausmultiplizieren von $(x+y)^3$:

$$(x+y)(x+y)(x+y) = x^3 + yx^2 + xyx + x^2y + xy^2 + yxy + y^2x + y^3.$$

Wir überlegen uns am Beispiel von $(x+y)^4$, dass die Teilmengen von $\{1,2,3,4\}$ dazu genutzt werden können, die Summanden im ausmultiplizierten Term zu nummerieren. (Das ist kein allgemeiner Beweis, sondern eben nur ein Beispiel. Hier soll auf einen Beweis verzichtet werden. Es ist häufig beim Lösen von mathematischen Problemen so, dass man sich die grundsätzliche Beweismethode zunächst an Beispielen klarmacht, und im nächsten Schritt dann diese zu einem strikten Beweis formalisiert. Beide Schritte sind wichtig zu lernen; hier wollen wir den Schwerpunkt auf dem ersten Teil setzen.)

Wir lassen uns von jeder Teilmenge von $\{1,2,3,4\}$ also anzeigen, aus welchen Klammern wir den Summanden x auswählen; aus den übrigen Klammern wählen wir den Summanden y aus. Ist beispielsweise die Menge $\{1,3\}$ vorgegeben, so entspricht das dem Produkt $xyxy$, also nehmen wir x in der ersten, y in der zweiten, x in der dritten, y in der vierten Klammer. Die Menge $\{1,2,3,4\}$ würde dem Produkt x^4 entsprechen, also der Wahl von x in jeder Klammer. Die leere Menge hingegen entspricht dem Produkt y^4, denn in keiner Klammer wurde das x ausgewählt - also wurde in jeder Klammer y ausgewählt. Insgesamt erhalten wir vor dem Zusammenfassen 2^4 Summanden, für jede Teilmenge von $\{1,2,3,4\}$ genau einen, und jeder Summand kann so beschrieben werden. Beim Zusammenfassen sehen wir, dass jede k-elementige Teilmenge von $\{1,2,3,4\}$ den Summanden $x^k y^{4-k}$ liefert, denn eine k-elementige Teilmenge lässt uns k-mal das x und $(4-k)$-mal das y auswählen. Das ist auch das Herzstück des Beweises, den wir jetzt etwas allgemeiner, wenn auch nicht ganz formal, festhalten.

Beweisidee zu dem Satz 4.5. Wie in der Vorüberlegung, multiplizieren wir die Klammern in dem Produkt $(x+y)^n$ simultan aus. Dabei entsprechen die Summanden vom Ergebnis vor dem Zusammenfassen den Teilmengen von $\{1,2,\ldots,n\}$. Das sieht man, indem man jeder Teilmenge den Summanden zuordnet, bei dem x genau in den Klammern ausgewählt wurde, die von dieser Teilmenge angegeben sind, und y in den übrigen Klammern. Auf diese Weise entstehen vor dem Zusammenfassen 2^n Summanden, und die k-elementigen Teilmengen entsprechen genau den Summanden $x^k y^{n-k}$. Nun haben wir $\binom{n}{k}$ als die Anzahl der k-elementigen Teilmengen von einer n-elementigen Menge definiert, also wird es nach dem Ausmultiplizieren genau $\binom{n}{k}$ Summanden $x^k y^{n-k}$ geben. Nach dem Zusammenfassen erhalten wir also genau $\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$, wie behauptet. □

Aus dem binomischen Lehrsatz lässt sich nun das folgende Korollar ableiten:

Korollar 4.6. *Für alle natürlichen Zahlen n gilt:* $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Bevor wir das Korollar beweisen, veranschaulichen wir uns die Situation im Pascalschen Dreieck: Die Summe auf der linken Seite ist jeweils die Summe aller Binomialkoeffizienten in der n-ten Zeile, die wir hier rechts neben jeder Zeile notieren:

			1				1
		1		1			2
	1		2		1		4
1		3		3		1	8
1		4		6		4	1	16
1	5		10		10		5	1	32

Beweis des Korollars 4.6. Wir fügen bei den Summanden den Faktor $1 = 1^k \cdot 1^{n-k}$ hinzu, was die Summe nicht verändert.

$$\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} \cdot 1^k \cdot 1^{n-k}.$$

Darauf lässt sich nun der binomische Lehrsatz 4.5 anwenden. Wir erhalten also

$$\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} \cdot 1^k \cdot 1^{n-k} = (1+1)^n = 2^n,$$

was zu beweisen war. □

Alternativer Beweis würde wie folgt aussehen: Jede Teilmenge einer n-elementigen Menge hat eine gewisse Anzahl von Elementen, und diese Anzahl ist eine natürliche Zahl zwischen 0 und n (einschließlich). Wenn wir also die Summe $\sum_{k=0}^{n} \binom{n}{k}$ bilden, erhalten wir die Gesamtzahl aller möglichen Teilmengen einer n-elementigen Menge, die wir nach Größe sortiert aufgeschrieben haben. Auf der anderen Seite wissen wir nach Satz 4.2, dass die Anzahl der Teilmenge einer n-elementigen Menge 2^n beträgt. Das zeigt die Behauptung.

Nun wollen wir auf die Beschreibung der Binomialkoeffizienten zurückkommen, die in der Schule üblicherweise benutzt wird. Dafür müssen wir den Begriff der Fakultät einer natürlichen Zahl wiederholen.

Definition 4.7. Für jede natürliche Zahl $n \geq 1$ ist $n!$ („n **Fakultät**") das Produkt der natürlichen Zahlen von 1 bis n:

$$n! = 1 \cdot 2 \cdot \ldots \cdot n.$$

Für 0 definieren wir $0! = 1$.

Beispiel. Wir bestimmen die Fakultäten für kleine natürliche Zahlen:

$$\begin{aligned} 1! &= 1 \\ 2! &= 1 \cdot 2 = 2 \\ 3! &= 1 \cdot 2 \cdot 3 = 6 \\ 4! &= 1 \cdot 2 \cdot 3 \cdot 4 = 24 \\ 5! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120 \\ 6! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 = 720 \\ 7! &= 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040 \end{aligned}$$

Es ist eine wichtige Beobachtung, dass Fakultäten sehr schnell wachsen.

Die Fakultäten spielen eine wichtige Rolle beim Lösen von Abzählproblemen. Hier ist ein Beispiel dafür.

Satz 4.8. *Es gibt $n!$ unterschiedliche Anordnungen von n Objekten.*

Beispiel. Die Objekte a, b, c können auf $3! = 6$ Arten angeordnet werden:

$$abc, acb, bac, bca, cab, cba$$

Der Beweis dieses Satzes würde durch Induktion erfolgen. Da wir bereits sehr viele formale Induktionsbeweise gesehen haben, beschränken wir uns auch hier auf eine Beweisidee.

Beweisidee. Die Aussage ist wahr für $n = 1$.

Bei einer Anordnung von den n Objekten hat man n Möglichkeiten, das erste Objekt auszuwählen. Für jedes Anfangsobjekt und die restlichen $(n-1)$ Objekte gibt es $(n-1)!$-Möglichkeiten, die restlichen Objekte anzuordnen. Das liefert insgesamt $n \cdot (n-1)! = n!$ Anordnungen. □

Nun können wir die Binomialkoeffizienten mit Hilfe von Fakultäten ausdrücken.

Satz 4.9. *Für alle natürlichen Zahlen $n \geq k$ gilt:* $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Wieder erläutern wir zunächst die Beweisidee an einem Beispiel.

Beispiel. Wir suchen die 2-elementigen Teilmengen von $\{a, b, c, d\}$. Um diese zu erhalten, betrachten wir alle möglichen Anordnungen der Objekte a, b, c, d, und wählen in jeder solchen Anordnung die ersten zwei Elemente zu den Elementen einer zweielementigen Menge aus. Dabei erhalten wir jede 2-elementige Teilmenge von $\{a, b, c, d\}$, allerdings häufiger zweimal. Das

wollen wir konkret durchführen (wir verzichten hier zwecks besseren Übersichtlichkeit auf Klammern):

$abcd$	$acbd$	$bacd$	$bcad$	$cabd$	$cbad$
$abdc$	$adbc$	$badc$	$bdac$	$dabc$	$dbac$
$acdb$	$adcb$	$cadb$	$cdab$	$dacb$	$dcab$
$bcda$	$bdca$	$cbda$	$cdba$	$dbca$	$dcba$

Dabei merken wir, dass wir jede 2-elementige Teilmenge 4 Mal gezählt haben. Dabei erzeugen die möglichen Anordnungen der zweielementigen Menge eine Verdopplung in der Zählung, und die möglichen Anordnungen der übriggebliebenen zwei Elementen erzeugen eine weitere Verdopplung. Daher erhält man 24 2-elementige Teilmengen der Menge $\{a, b, c, d\}$, wobei jede Teilmenge 4 Mal gezählt wurde, also $\frac{24}{4} = 6$ zweielementige Teilmengen. Das Prinzip verallgemeinern wir nun.

Beweisidee. Betrachte eine n-elementige Menge A. Betrachte alle $n!$ möglichen Anordnungen der Elemente der Menge A. Bei jeder Anordnung fassen wir die ersten k Elemente zu einer Teilmenge auf. Auf diese Weise erhalten wir jede k-elementige Teilmenge, jedoch werden diese im Allgemeinen mehrfach vorkommen. Bislang haben wir $n!$ Teilmengen, von denen manche gleich sind, markiert.

Als nächstes ermitteln wir, wie oft jede Teilmenge markiert wurde. Die Teilmenge und die restlichen $(n - k)$ Elemente können beliebig angeordnet werden und liefert dabei immer dieselbe Teilmenge, da es bei Teilmengen nur auf die Elemente und nicht auf deren Reihenfolge ankommt. Somit wurde jede Teilmenge $k! \cdot (n - k)!$ Mal markiert, da wir die k-elementige Menge auf $k!$ verschiedene Arten anordnen können und die restlichen Element auf $(n-k)!$ verschiedene Arten, und diese Anordnungen sind voneinander unabhängig. Daher gibt es $\frac{n!}{k!(n-k)!}$ k-elementige Teilmengen von A, also folgt wie behauptet

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

□

5 Abbildungen

In diesem Abschnitt beschäftigen wir uns mit Abbildungen. Abbildungen geben uns eine Möglichkeit, unterschiedliche Mengen miteinander in Beziehung zu setzen. Abbildungen verallgemeinern den Begriff einer Funktion, der aus der Schule bekannt ist.

Definition 5.1. Seien A,B Mengen. Eine **Abbildung** $f\colon A \to B$ ist eine Zuordnung, die jedem Element von A genau ein Element von B zuordnet. Die Menge A heißt **Quelle** und die Menge B das **Ziel** der Abbildung f.

Man bemerke, dass man von einer Abbildung erst sprechen kann, wenn man Quelle und Ziel festgelegt hat. Ferner ist es wichtig, dass jedes Element von A einen Wert im Ziel B zugeordnet bekommt, und dieses Element sollte eindeutig sein. Letzteres entspricht der aus dem Schulunterricht bekannter Tatsache, dass ein Funktionsgraph niemals zwei unterschiedliche y-Werte für denselben x-Wert haben darf.

Beispiel.

- Funktionen aus dem Schulunterricht liefert Beispiel für Abbildungen, z.B. $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. Alternativ schreibt man auch $x \mapsto x^2$, um die Funktionsvorschrift zu beschreiben.

- In manchen Fällen kann man die Abbildung durch die vollständige Liste der Zuordnungen beschreiben. $g\colon \{1,2,3\} \to \{u,v,w\}$, $g(1) = u$, $g(2) = v$, $g(3) = u$ ist eine Abbildung. Alternativ schreibt man auch:

$$\begin{aligned} 1 &\mapsto u, \\ 2 &\mapsto v, \\ 3 &\mapsto u. \end{aligned}$$

  Man beachte, dass hier zwei Elementen der Quelle, nämlich den Elementen 1 und 3, dasselbe Element zugeordnet wird. Das ist bei Abbildungen durchaus zulässig; ausgeschlossen wird nur, dass zwei oder mehr Elemente des Bildes die Werte für ein Element der Quelle als Werte ausgewählt werden.

- Die Zuordnung {KlausurteilnehmerInnen} $\to$ {Gültige Noten}, die jedem Prüfling seine Klausurnote zuordnet, ist eine Abbildung. Jede(r) KlausurteilnehmerIn erhält nämlich eine Klausurnote, und zwar genau eine.

- Die Zuordnung $g\colon \mathbb{R} \to \mathbb{R}$,

$$a \mapsto \text{ Lösung der Gleichung } x^2 = a,$$

  ist keine Abbildung. Dabei treten gleich zwei Probleme auf: Den negativen Zahlen wird durch diese Vorschrift keine reelle Zahl zugeordnet,

denn die Gleichung $x^2 = a$ hat keine Lösung in den reellen Zahlen für die negativen Werte von a. Für die positiven Werte von a tritt allerdings auch ein Problem auf, denn die Gleichung $x^2 = a$ hat zwei verschiedene Lösungen, und somit ist es nicht eindeutig, welche der beiden Lösungen einer positiven Zahl a zugeordnet werden soll.

- Die Vorschrift $x \mapsto x^2$ ist (ohne Kontext) keine Abbildung, da Ziel und Quelle nicht spezifiziert wurden.
- Die Zuordnung $\varphi\colon [1,2] \to [1,2], x \mapsto x^2$, ist ebenfalls keine Abbildung. Jedem Element der Quelle soll durch die Vorschrift ein Element des Ziels zugeordnet werden, allerdings würde die Vorschrift z.B. dem Element 2 der Quelle die Zahl 4 zuordnen, die kein Element des Ziels ist.

Um Abbildungen zu vergleichen, geben wir noch die folgende Definition. Diese ist weitestgehend intuitiv, allerdings müssen Quelle und Ziel der Abbildungen ebenfalls übereinstimmen.

Definition 5.2. Zwei Abbildungen heißen **gleich**, falls sie dieselbe Quelle und dasselbe Ziel haben und beide Abbildungen jedem Element der Quelle jeweils dasselbe Element im Ziel zuordnen.

Im Folgenden wollen wir die Eigenschaften von Abbildungen studieren. Dabei fangen wir mit Injektivität und Surjektivität an.

Definition 5.3. Eine Abbildung $f\colon X \to Y$ heißt **injektiv**, wenn für alle $x_1, x_2 \in X$ aus $f(x_1) = f(x_2)$ folgt: $x_1 = x_2$.

Nach Proposition 1.7 können wir das unmittelbar zu der folgenden Aussage umformulieren: Eine Abbildung $f\colon X \to Y$ ist genau dann injektiv, wenn für alle $x_1, x_2 \in X$ mit $x_1 \neq x_2$ gilt: $f(x_1) \neq f(x_2)$. Unterschiedlichen Elementen der Quelle werden also unterschiedliche Elemente des Ziels zugeordnet. Das sollte nicht mit der Definition der Abbildung verwechselt werden: Da geht es darum, dass jedes einzelne Element der Quelle etwas Eindeutiges zugeordnet bekommt, während für die Injektivität verglichen wird, wie die Werte für *unterschiedliche* Elemente aussehen.

Beispiel.

- Die Abbildung $\varphi\colon \mathbb{R} \to \mathbb{R}, \varphi(y) = 7y+3$, ist injektiv. Um das zu prüfen, nehmen wir an, dass $y_1, y_2 \in \mathbb{R}$ liegen und die Eigenschaft $\varphi(y_1) = \varphi(y_2)$ haben. Wir müssen unter dieser Annahme zeigen, dass $y_1 = y_2$ folgt. Schreibt man die Annahme expliziter hin, so erhält man

$$7y_1 + 3 = 7y_2 + 3.$$

Diese Gleichung formen wir nun äquivalent um:

$$\begin{aligned} 7y_1 + 3 &= 7y_2 + 3 \,|\, -3 \\ \Leftrightarrow 7y_1 &= 7y_2 \,|\, :7 \\ \Leftrightarrow y_1 &= y_2 \end{aligned}$$

- $f\colon \{1,2,3\} \to \{u,v,w\}$, $f(1) = u$, $f(2) = v$, $f(3) = u$ ist keine injektive Abbildung. Negiert man die Definition der Injektivität, so sieht man sofort: Eine Abbildung $g\colon A \to B$ ist genau dann nicht injektiv, wenn es Elemente $a_1, a_2 \in A$ gibt, sodass $a_1 \neq a_2$, aber $g(a_1) = g(a_2)$ gilt. Wendet man das auf die Abbildung f an, so hat man hier unterschiedliche Elemente $1, 3$ der Quelle auf dasselbe Element u des Ziels abgebildet.

Ergänzend zur Injektivität gibt es auch die Surjektivität, die im Allgemeinen von der Injektivität aber unabhängig ist.

Definition 5.4. Seien X, Y zwei Mengen. Eine Abbildung $f\colon X \to Y$ heißt **surjektiv**, wenn jedes Element $y \in Y$ ein **Urbild** $x \in X$ hat, d.h. zu jedem $y \in Y$ gibt es ein $x \in X$ mit $f(x) = y$.

Dabei kann ein $y \in Y$ auch durchaus mehrere Urbilder in X haben, es ist nur die Existenz wichtig. Die Surjektivität besagt also, informeller ausgedrückt, dass jedes Element vom Ziel mindestens einmal als Bild von etwas in der Quelle fungiert.

Beispiel.

- Die Abbildung $\alpha\colon \{1,2,3\} \to \{a,b\}$, gegeben durch $1 \mapsto b, 2 \mapsto b, 3 \mapsto a$ ist surjektiv. Wir müssen nachprüfen, dass jedes Element in der Zielmenge, also jedes Element von $\{a,b\}$ ein Urbild hat. Das Element a hat ein Urbild, da $\alpha(3) = a$, und das Element b hat ein Urbild, da $\alpha(1) = b$, also ist die Abbildung surjektiv. Dass das Element b noch ein weiteres Urbild hat, ist hierbei nicht relevant.

- Die Abbildung $\{a,b\} \to \mathbb{N}$, $a \mapsto 2$, $b \mapsto 7$, ist nicht surjektiv: Beispielsweise hat das Element $1 \in \mathbb{N}$ kein Urbild unter dieser Abbildung.

Das nächste Beispiel verdeutlicht, dass für Injektivität und Surjektivität die Wahl der Quelle und des Ziels eine entscheidende Rolle spielen.

Beispiel.

- Die Abbildung $\varphi\colon \mathbb{R} \to \mathbb{R}$, $\varphi(x) = x^2$, ist weder injektiv noch surjektiv. Um zu zeigen, dass die Abbildung φ nicht injektiv ist, bemerken wir, dass beispielsweise die unterschiedlichen Elemente -2 und 2 der Quelle auf dasselbe Element des Ziels abgebildet werden, nämlich auf das Element 4. Das zeigt, dass die Abbildung φ nicht injektiv ist. Ferner ist die Abbildung nicht surjektiv, da beispielsweise das Element -1 kein Urbild unter φ besitzt.

- Wir ändern nun die Quelle der Abbildung und betrachten die Abbildung $\psi\colon [0,\infty) \to \mathbb{R}$, $\psi(y) = y^2$. Diese Abbildung ist nun injektiv, aber nicht surjektiv, wie wir gleich zeigen werden. Hat man zwei nichtnegative reelle Zahlen y_1, y_2, für die gilt: $y_1^2 = y_2^2$, so folgt daraus $y_1 = y_2$. Die Abbildung ist nicht surjektiv, da das Element -1 wie auch bei der vorherigen Abbildung kein Urbild hat.

- Die Abbildung $\beta\colon [0,\infty) \to [0,\infty)$, $\beta(x) = x^2$, ist nun injektiv und surjektiv. Für die Injektivität benutzen wir dasselbe Argument wie bei der vorherigen Abbildung. Zur Surjektivität: Jedes Element $y \in [0,\infty)$ des Ziels hat ein Urbild unter β, nämlich $\sqrt{y}$, denn $\beta(\sqrt{y}) = (\sqrt{y})^2 = y$.

Wir geben ein weiteres Beispiel einer Abbildung.

Beispiel. Für jede Menge A gibt es eine *Identitätsabbildung von* A, $\mathrm{id}_A\colon A \to A$, die jedem Element a von A dasselbe Element a von A zuordnet. Diese Abbildung ist sowohl injektiv als auch surjektiv.

Definition 5.5. Eine Abbildung, die sowohl injektiv als auch surjektiv ist, heißt **bijektiv**.

Abbildungen können miteinander in Verbindung gesetzt werden, indem man sie *verkettet*. Wir werden Verkettung insbesondere brauchen, um bijektive Abbildungen besser zu beschreiben; es ist aber auch an sich ein wichtiger Begriff.

Definition 5.6. Seien $f\colon A \to B$ and $g\colon B \to C$ zwei Abbildungen. Die **Komposition** dieser Abbildungen ist eine Abbildung $g \circ f\colon A \to C$, die durch die Vorschrift

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

für alle $a \in A$ definiert ist.

Hierzu sollten einige Dinge angemerkt werden.

Bemerkung.

- Wir sollten uns davon überzeugen, dass der Ausdruck in der Definition tatsächlich Sinn ergibt. Um eine Abbildung von der Menge A in die Menge C zu definieren, müssen wir jedem Element von A genau ein Element von C zuordnen. Dabei ordnen wir dem Element $a \in A$ als Zwischenschritt das Element $f(a) \in B$ zu. Auf dieses Element können wir nun die Abbildung g anwenden, um das Element $g(f(a))$ in C zu erhalten. Somit ist $g \circ f$ in der obigen Definition tatsächlich eine Abbildung.

- Wie wir im letzten Punkt gesehen haben, würde die Definition einer Komposition $\beta \circ \alpha$ nicht funktionieren, wenn die Zielmenge von α und die Quelle von β nicht übereinstimmen würden. Es ist also eine wichtige Bedingung.

- Die Komposition wird synonym auch *Verkettung*, *Verknüpfung*, *Hintereinanderschaltung* bzw. *Hintereinanderausführung* genannt.

- Die Schreibweise $\beta\circ\alpha$ ist potentiell verwirrend, da man auf ein Element der Quelle von α zunächst α anwendet und erst dann β; man liest in diesem Fall gewissermaßen von rechts nach links. Eine Sprechweise für „$\beta \circ \alpha$“ ist allerdings „β nach α“.

Beispiel. Wir betrachten die Abbildungen $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = 2x + 1$, und $g\colon \mathbb{R} \to \mathbb{R}, g(y) = y^3$.

Zuerst wollen wir nachprüfen, dass f sowohl injektiv als auch surjektiv ist. (Das hat nicht wirklich mit Komposition zu tun, sondern soll nochmal die Begriffe „injektiv“ und „surjektiv“ verdeutlichen.) Die Injektivität wurde ähnlich bereits nachgewiesen. Wir betrachten zwei Elemente $x_1, x_2 \in \mathbb{R}$ mit gleichem Bild $f(x_1) = f(x_2)$. Das schreiben wir aus und formen es äquivalent um:

$$\begin{aligned} 2x_1 + 1 &= 2x_2 + 1 \,|\, -1 \\ \Leftrightarrow 2x_1 &= 2x_2 \,|\, :2 \\ \Leftrightarrow x_1 &= x_2. \end{aligned}$$

Also impliziert $f(x_1) = f(x_2)$ bereits $x_1 = x_2$, folglich ist f injektiv.

Die Abbildung f ist surjektiv. Um das zu beweisen, müssen wir zu jedem Element y des Ziels $\mathbb{R}$ ein Urbild angeben. Das Urbild von $y \in \mathbb{R}$ ist durch $\frac{y-1}{2}$ gegeben, denn es gilt:

$$f\left(\frac{y-1}{2}\right) = 2 \cdot \frac{y-1}{2} + 1 = y - 1 + 1 = y.$$

Nun machen wir mit Komposition weiter. Wir bestimmen $f \circ g$, indem wir die Definition der Komposition benutzen: Für alle $z \in \mathbb{R}$ erhalten wir

$$(f \circ g)(z) = f(g(z)) = f(z^3) = 2 \cdot z^3 + 1.$$

Auf der anderen Seite wollen wir die Komposition $g \circ f$ expliziter beschreiben. Dabei benutzen wir den binomischen Lehrsatz 4.5:

$$(g \circ f)(z) = g(f(z)) = g(2z + 1) = (2z + 1)^3 = 8z^3 + 12z^2 + 6z + 1.$$

Wir sehen, dass die Abbildungen $g \circ f$ und $f \circ g$ im Allgemeinen nicht gleich sind. Um das nachzuweisen, setzen wir z.B. das Element 1 ein:

$$\begin{aligned} (f \circ g)(1) &= 2 \cdot 1^3 + 1 = 3, \\ (g \circ f)(1) &= (2 \cdot 1 + 1)^3 = 27. \end{aligned}$$

Dem Element 1 ordnen also die Abbildungen $f \circ g$ und $g \circ f$ unterschiedliche Elemente des Ziels zu, folglich sind diese Abbildungen verschieden. Im Gegensatz zu solchen Operationen wie Und-Verknüpfung von Aussagen oder Addition oder Multiplikation von reellen Zahlen ist die Komposition von Abbildungen nicht kommutativ, d.h. die Reihenfolge der Abbildungen spielt eine Rolle.

Im Folgenden betrachten wir ein weiteres Beispiel, das etwas anders geartet ist.

Beispiel. Sei $A = \{a, b, c, d\}$ und $B = \{x, y, w\}$. Wir definieren die Abbildung $\alpha\colon A \to B$ durch

$$\begin{aligned} a &\mapsto y \\ b &\mapsto w \\ c &\mapsto y \\ d &\mapsto x \end{aligned}$$

Wir bemerken, dass die Abbildung α surjektiv ist, da jedes Element der Zielmenge B als Bild von einem Element aus A auftaucht. Die Abbildung α ist nicht injektiv, da es zwei unterschiedliche Elemente der Quelle gibt, nämlich a und c, die auf dasselbe Element des Ziels abgebildet werden.

Ferner definieren wir die Abbildung $\beta\colon B \to \mathbb{N}$ durch $\beta(x) = 2$, $\beta(y) = 5$, $\beta(w) = 7$. Diese Abbildung ist nicht surjektiv, da beispielsweise das Element 1 des Ziels kein Urbild besitzt. Diese Abbildung ist injektiv, da unterschiedliche Elemente von B auf unterschiedliche Elemente von $\mathbb{N}$ abgebildet werden.

In diesem Beispiel ist die Komposition $\alpha \circ \beta$ nicht definiert, da das Ziel von β (die Menge der natürlichen Zahlen) und die Quelle von α (die Menge A) nicht übereinstimmen. Würden wir versuchen, ein $\beta(b)$ in α einzusetzen, so würden wir das gar nicht auswerten können, da α auf der Menge A und nicht auf den natürlichen Zahlen definiert ist.

Hingegen ist die Komposition $\beta \circ \alpha\colon A \to \mathbb{N}$ definiert. Wir bestimmen die Werte dieser Abbildung auf den Elementen von A:

$$\begin{aligned} (\beta \circ \alpha)(a) &= \beta(\alpha(a)) = \beta(y) = 5, \\ (\beta \circ \alpha)(b) &= \beta(\alpha(b)) = \beta(w) = 7, \\ (\beta \circ \alpha)(c) &= \beta(\alpha(c)) = \beta(y) = 5, \\ (\beta \circ \alpha)(d) &= \beta(\alpha(d)) = \beta(x) = 2. \end{aligned}$$

Insbesondere bemerken wir, dass die Abbildung $\beta \circ \alpha$ weder injektiv noch surjektiv ist. Sie ist nicht surjektiv, da beispielsweise das Element $1 \in \mathbb{N}$ kein Urbild unter dieser Abbildung besitzt. Sie ist nicht injektiv, da es zwei unterschiedliche Elemente a und c in der Quelle gibt, die beide auf dasselbe Element 5 im Ziel abgebildet werden.

Wir setzen unsere Suche nach einem Kriterium für Bijektivität fort. Damit eng verbunden ist der Begriff einer Umkehrabbildung (oder inverser Abbildung)

Definition 5.7. Sei $f\colon A \to B$ eine Abbildung. Eine Abbildung $g\colon B \to A$ heißt **inverse Abbildung zu** f, falls $g \circ f = \mathrm{id}_A$ und $f \circ g = \mathrm{id}_B$.

Schreibt man die Bedingungen in der Definition nochmal aus, so heißt es: $g(f(a)) = \mathrm{id}_A(a) = a$ für alle $a \in A$ und $f(g(b)) = \mathrm{id}_B(b) = b$ für alle $b \in B$. Um sich diesen Begriff zu veranschaulichen, schauen wir uns wieder Beispiele an:

Beispiel. • Zunächst schauen wir uns ein aus der Schule bekanntes Beispiel an. Die Abbildung $\sqrt{}\colon [0,\infty) \to [0,\infty)$, $x \mapsto \sqrt{x}$ ist die inverse Abbildung zu der Abbildung $q\colon [0,\infty) \to [0,\infty), x \mapsto x^2$. Um das nachzuprüfen, müssen wir für jede nicht-negative reelle Zahl x nachprüfen, dass $\sqrt{q(x)} = x$ und $q(\sqrt{x}) = x$ gilt. Setzt man die Definitionen ein, so sind diese Aussagen zu den Aussagen $\sqrt{x^2} = x$ und $(\sqrt{x})^2 = x$ äquivalent, die für jede nicht-negative reelle Zahl x gelten. Somit ist $\sqrt{}$ die inverse Abbildung zu der Abbildung q.

- Die Abbildung $\psi\colon \mathbb{R} \to \mathbb{R}$, $\psi(y) = 2y + 1$, hat die inverse Abbildung $\varphi\colon \mathbb{R} \to \mathbb{R}$, $\varphi(z) = \frac{1}{2}z - \frac{1}{2}$. Um das zu überprüfen, müssen wir die Kompositionen $\psi \circ \varphi$ und $\varphi \circ \psi$ bestimmen. Für jedes $x \in \mathbb{R}$ gilt nun nach Definitionen von Komposition und von den Abbildungen φ, ψ:

$$\begin{aligned}(\psi \circ \varphi)(x) &= \psi(\varphi(x)) = \psi\left(\frac{1}{2}x - \frac{1}{2}\right) = 2 \cdot \left(\frac{1}{2}x - \frac{1}{2}\right) + 1\\ &= x - 1 + 1 = x,\\ (\varphi \circ \psi)(x) &= \varphi(\psi(x)) = \varphi(2x+1) = \frac{1}{2} \cdot (2x+1) - \frac{1}{2}\\ &= x + \frac{1}{2} - \frac{1}{2} = x.\end{aligned}$$

  Das zeigt, dass φ und ψ zueinander inverse Abbildungen sind.

- Die Abbildung $\alpha\colon \{a,b,c\} \to \{1,2,3\}$, die durch die Vorschrift

$$\begin{aligned}a &\mapsto 1,\\ b &\mapsto 2,\\ c &\mapsto 3\end{aligned}$$

  gegeben ist, hat als inverse Abbildung die Abbildung $\beta\colon \{1,2,3\} \to \{a,b,c\}$, die durch die Vorschrift

$$\begin{aligned}1 &\mapsto a,\\ 2 &\mapsto b,\\ 3 &\mapsto c\end{aligned}$$

  gegeben ist. Um das nachzuprüfen, müssen wir die Kompositionen $\beta \circ \alpha$ und $\alpha \circ \beta$ bestimmen. Dabei ergibt sich:

$$\begin{aligned}(\beta \circ \alpha)(a) &= \beta(\alpha(a)) = \beta(1) = a,\\ (\beta \circ \alpha)(b) &= \beta(\alpha(b)) = \beta(2) = b,\\ (\beta \circ \alpha)(c) &= \beta(\alpha(c)) = \beta(3) = c,\end{aligned}$$

und

$$\begin{aligned}(\alpha \circ \beta)(1) &= \alpha(\beta(1)) = \alpha(a) = 1,\\ (\alpha \circ \beta)(2) &= \alpha(\beta(2)) = \alpha(b) = 2,\\ (\alpha \circ \beta)(3) &= \alpha(\beta(3)) = \alpha(c) = 3.\end{aligned}$$

Also gilt $\beta \circ \alpha = \mathrm{id}_{\{a,b,c\}}$ und $\alpha \circ \beta = \mathrm{id}_{\{1,2,3\}}$ und somit sind die Abbildungen α und β zueinander invers.

- Nicht jede Abbildung besitzt eine inverse Abbildung. Betrachten wir beispielsweise die Abbildung $f\colon \mathbb{N}_0 \to \mathbb{N}_0$, gegeben durch $f(x) = x^2$. Wir wollen zeigen, dass diese Abbildung keine inverse Abbildung besitzt. Angenommen, $g\colon \mathbb{N}_0 \to \mathbb{N}_0$ wäre die inverse Abbildung zu f. Wir betrachten die natürliche Zahl $g(2)$, die die Abbildung g der Zahl 2 zuordnen würde. Nach Definition der inversen Abbildung müsste dafür $f(g(2)) = 2$ gelten. Setzt man hier die Definition von f ein, so erhalten wir $(g(2))^2 = 2$. Allerdings gibt es, wie wir wissen, keine natürliche Zahl, dessen Quadrat genau 2 ergibt. Das liefert einen Widerspruch, also war unsere Annahme, dass f eine inverse Abbildung g hat, falsch. Damit ist gezeigt, dass die Abbildung f keine inverse Abbildung haben kann.

Nun können wir unser Bijektivitätskriterium formulieren.

Satz 5.8. *Eine Abbildung $f\colon A \to B$ ist genau dann bijektiv, wenn sie eine inverse Abbildung besitzt.*

Beweis. Die Aussage, die wir beweisen wollen, ist eine Äquivalenz zweier Aussagen (der Aussagen „Die Abbildung f ist bijektiv.“ und „Die Abbildung f besitzt eine inverse Abbildung.”). Um eine solche Aussage zu beweisen, ist es ein übliches Vorgehen, zu zeigen, dass jede der Aussagen die jeweils andere impliziert. Konkreter müssen hier dann zwei Aussagen bewiesen werden:

1) Ist eine Abbildung $f\colon A \to B$ bijektiv, so hat diese Abbildung eine inverse Abbildung.

2) Hat die Abbildung $f\colon A \to B$ eine inverse Abbildung, so ist diese Abbildung f bijektiv.

Wir fangen mit dem zweiten Punkt an, da dieser einfacher ist.

ad 2) In diesem Teil des Beweises nehmen wir an, dass die Abbildung $f\colon A \to B$ eine inverse Abbildung $g\colon B \to A$ besitzt. Wir wollen unter dieser Voraussetzung zeigen, dass f bijektiv ist. Nach Definition der Bijektivität müssen wir also zeigen, dass f injektiv und surjektiv ist.

Wir fangen mit der Injektivität an. Seien also a_1, a_2 Elemente der Quelle, die von f auf dasselbe Element $f(a_1) = f(a_2)$ abgebildet werden. Wendet man nun g auf dieses Element von B an, so erhält man $g(f(a_1)) = g(f(a_2))$. Erinnert man sich nun an die Definition der inversen Abbildung, so sieht man $g(f(a_1)) = (g \circ f)(a_1) = \mathrm{id}_A(a_1) = a_1$ und genauso $g(f(a_2)) = (g \circ f)(a_2) = \mathrm{id}_A(a_2) = a_2$. Also können wir aus der Annahme, dass $f(a_1) = f(a_2)$, folgern, dass $a_1 = a_2$ gilt. Somit ist die Abbildung f injektiv.

Nun bleibt es zu zeigen, dass die Abbildung f surjektiv ist. Sei also $b \in B$ ein beliebiges Element; wir müssen zeigen, dass es ein Urbild unter f besitzt. Betrachte das Element $g(b) \in A$. Wir behaupten, dass dieses Element ein Urbild von b unter f ist. Bestimmt man nämlich $f(g(b))$, so erhält man aus der Definition der inversen Abbildung:

$$f(g(b)) = (f \circ g)(b) = \mathrm{id}_B(b) = b.$$

Also liefert $g(b)$ zu jedem $b \in B$ ein Urbild unter f. Somit ist die Abbildung f surjektiv.

In diesem Beweisteil haben wir also gezeigt, dass die Abbildung f, die eine inverse Abbildung besitzt, notwendigerweise injektiv und surjektiv sein muss.

ad 1) Nun wollen wir den ersten Teil der Aussage beweisen, nämlich dass jede bijektive Abbildung $f\colon A \to B$ eine inverse Abbildung besitzt. Dieser Beweisteil ist insofern schwieriger, dass wir nun eine Abbildung $g\colon B \to A$ angeben müssen, und dann prüfen, dass die beiden Bedingungen aus der Definition der inversen Abbildung erfüllt sind, also dass $f \circ g = \mathrm{id}_B$ und $g \circ f = \mathrm{id}_A$ gilt.

Um die Abbildung g zu definieren, nutzen wir die Surjektivität von f aus. Zu jedem $b \in B$ hat man nach Definition der Surjektivität mindestens ein Urbild $a \in A$. Wir suchen uns also zu jedem $b \in B$ ein solches Urbild $a \in A$ aus und setzen $g(b) = a$. Das definiert zu jeder surjektiven Abbildung f eine Abbildung $g\colon B \to A$ mit der Eigenschaft $f(g(b)) = b$ für alle $b \in B$, allerdings brauchen wir noch die Injektivität von f, um zu zeigen, dass diese Abbildung g auch die Umkehrabbildung von f ist. Dass es ohne Injektivität von f nicht funktioniert, wird in dem nachfolgenden Beispiel demonstriert.

Wir müssen noch zeigen, dass $g \circ f = \mathrm{id}_A$ gilt, mit anderen Worten, dass für alle $a \in A$ gilt: $g(f(a)) = a$. Um das einzusehen, merken wir, dass $g(f(a))$ als ein Element definiert wurde, dessen Bild $f(g(f(a)))$ genau $f(a)$ ist. Ohne Weiteres würde das noch nicht sicherstellen, dass

$g(f(a)) = a$ gilt. Allerdings wissen wir, dass f injektiv ist, und dass $g(f(a))$ und a zwei Elemente von a sind, dessen Bilder unter f übereinstimmen. Nach Definition der Injektivität folgt nun, dass $g(f(a)) = a$ ist.

Damit haben wir also gezeigt, dass jede bijektive Abbildung eine inverse Abbildung besitzt. Das vervollständigt den Beweis des Satzes.

□

Beispiel. Wir betrachten die Abbildung $\alpha\colon \{a, b, c\} \to \{1, 2\}$, die durch $\alpha(a) = 1$, $\alpha(b) = 2$, $\alpha(c) = 2$ gegeben ist. Wir bemerken, dass diese Abbildung surjektiv, aber nicht injektiv ist. Indem wir das Verfahren aus dem Beweis des vorigen Satzes anwenden, können wir eine Abbildung $\beta\colon \{1, 2\} \to \{a, b, c\}$ definieren, die die Eigenschaft $\alpha \circ \beta = \mathrm{id}_{\{1,2\}}$ besitzt. Wir wählen nämlich zu jedem der Elemente $1, 2$ des Ziels ein Urbild unter α und erhalten beispielsweise $\beta(1) = a$ und $\beta(2) = c$. (Hier haben wir eine Entscheidung getroffen.) Es gilt nun:

$$\begin{aligned} \alpha(\beta(1)) &= \alpha(a) = 1, \\ \alpha(\beta(2)) &= \alpha(c) = 2. \end{aligned}$$

Aber die andere Komposition, $\beta \circ \alpha$, ist nicht die Identitätsabbildung der Menge $\{a, b, c\}$: Beispielsweise gilt

$$\begin{aligned} \beta(\alpha(b)) &= \beta(2) = c, \text{ aber} \\ \mathrm{id}_B(b) &= b. \end{aligned}$$

Also ist β keine inverse Abbildung von α. Nach dem Satz, den wir soeben bewiesen haben, kann α keine inverse Abbildung haben, da α nicht injektiv ist.

6 Abzählen II

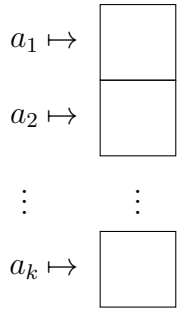
Mit den neuen Begriffen können wir uns wieder den Fragestellungen zuwenden, die wir schon im Kapitel 4 aufgeworfen haben. Zunächst fragen wir uns, wie viele Abbildungen es zwischen zwei endlichen Mengen gibt.

Proposition 6.1. *Sei A eine k-elementige und B eine m-elementige Menge. Dann gibt es m^k unterschiedliche Abbildungen von A nach B, d.h., die Menge*

$$\{f\colon A \to B \mid f \text{ Abbildung}\}$$

hat m^k Elemente.

Beweisidee. Für jeder Element von A wählen wir, um eine Abbildung $f\colon A \to B$ zu bestimmen, ein einziges Element von B. Die Wahl findet für jedes Element von A unabhängig statt. Wenn wir $A = \{a_1, a_2, \ldots, a_k\}$ schreiben, so ergibt sich das folgende Bild.



Wir haben also k Positionen, jede im Bild durch ein Kästchen dargestellt, und auf jede dieser Positionen wird ein Element von B gesetzt. Das ergibt m unabhängige Möglichkeiten für jede der Positionen, also insgesamt

$$\underbrace{m \cdot m \cdot \ldots \cdot m}_{k \text{ Mal}} = m^k$$

Möglichkeiten. □

Im Folgenden geht es darum, unseren Begriff für die Anzahl der Elemente einer Menge zu präzisieren. Dabei erhalten wir auch einen Eindruck davon, wie unterschiedliche Begriffe von „Unendlichkeit“ festgemacht werden können.

Definition 6.2. Zwei Mengen heißen **gleichmächtig**, wenn es eine Bijektion zwischen diesen Mengen gibt.

Eine Menge A hat die **Mächtigkeit** n, wenn es eine Bijektion zwischen dieser Menge A und der Menge $\{1, 2, \ldots, n\}$ gibt.

Anschaulich ist eine Bijektion zwischen der Menge $\{1, 2, \ldots, n\}$ genau das Durchnummerieren der Elemente von A mit den Zahlen $1, 2, \ldots, n$; eine Menge der Mächtigkeit n ist also nur eine sauberere Formulierung des intuitiven Begriffes einer n-elementigen Menge.

Während beim Arbeiten mit endlichen Mengen der intuitive Begriff vollkommen ausreichend ist, trügt uns unsere Intuition häufig, sobald wir über unendliche Mengen sprechen. Intuitiv würden wir sagen, dass es mehr ganze als natürliche Zahlen gibt, also es keine Bijektion zwischen den Mengen $\mathbb{N}_0$ und $\mathbb{Z}$ geben sollte. Das stellt sich allerdings als falsch heraus. Einige Größenvergleiche für unendliche Mengen werden im folgenden Satz zusammengefasst.

Satz 6.3. *1. Die Mengen $\mathbb{N}_0, \mathbb{Z}, \mathbb{Q}$ sind gleichmächtig.*

2. Die Mengen $\mathbb{N}$ und $\mathbb{R}$ sind nicht gleichmächtig.

3. Das offene Intervall $(0,1)$ und die Menge der reellen Zahlen $\mathbb{R}$ sind gleichmächtig.

Der Beweis von diesem Satz ist insgesamt nicht ganz einfach, und wir wollen nur eine der Behauptungen beweisen. Im folgenden zeigen wir, dass $\mathbb{N}_0$ und $\mathbb{Z}$ gleichmächtig sind.

Beweis der Gleichmächtigkeit von $\mathbb{N}_0$ und $\mathbb{Z}$. Um zu zeigen, dass $\mathbb{N}_0$ und $\mathbb{Z}$ gleichmächtig sind, müssen wir eine Bijektion $\alpha\colon \mathbb{N}_0 \to \mathbb{Z}$ konstruieren. Wir geben eine solche Abbildung α an. Um zu zeigen, dass sie bijektiv ist, benutzen wir dann den Satz 5.8, und konstruieren eine Abbildung $\beta\colon \mathbb{Z} \to \mathbb{N}_0$, die invers zu α ist.

Als kleine Vorüberlegung erinnern wir uns, dass jede natürliche Zahl entweder gerade oder ungerade ist. Jede gerade natürliche Zahl x kann in der Form $x = 2k$ für eine natürliche Zahl k geschrieben werden, wie aus der Definition der Teilbarkeit folgt. Jede ungerade natürliche Zahl y kann in der Form $y = 2k - 1$ für eine natürliche Zahl k geschrieben werden, wie wir uns bereits im Kapitel 3 veranschaulicht haben.

Nun definieren wir $\alpha\colon \mathbb{N}_0 \to \mathbb{Z}$ folgendermaßen:

$$x \mapsto \begin{cases} k, \text{ falls } x = 2k \text{ für eine natürliche Zahl } k, \\ -k, \text{ falls } x = 2k-1 \text{ für eine natürliche Zahl } k \geq 1. \end{cases}$$

Informell gesprochen: Jede gerade Zahl wird durch 2 geteilt; jede ungerade Zahl wird durch 2 geteilt, aufgerundet und danach mit einem Minuszeichen versehen. Um die Abbildung besser zu verstehen, schreiben wir einige Werte dieser Abbildung auf.

$$\begin{aligned} \alpha(0) &= 0 \\ \alpha(1) &= -1 \\ \alpha(2) &= 1 \\ \alpha(3) &= -2 \\ \alpha(4) &= 2 \\ \alpha(5) &= -3 \\ \alpha(6) &= 3 \end{aligned}$$

Intuitiv sehen wir bereits, dass es „so weitergeht“ und jede ganze Zahl in dieser Liste der Werte genau einmal auftauchen wird. Um das zu präzisieren, konstruieren wir nun die Umkehrabbildung zu α.

Die Abbildung $\beta\colon \mathbb{Z} \to \mathbb{N}_0$ sei gegeben durch:

$$m \mapsto \begin{cases} 2m, \text{ falls } m \geq 0, \\ 2 \cdot (-m) - 1, \text{ falls } m < 0. \end{cases}$$

Dadurch ordnen wir jeder Zahl m eine eindeutige natürliche Zahl zu. Nun bestimmen wir zuerst $\beta \circ \alpha$. Dafür müssen wir zwei Fälle unterscheiden, nämlich, ob in $\beta \circ \alpha$ eine gerade oder eine ungerade natürliche Zahl eingesetzt wird. Dabei nutzen wir wieder, dass jede gerade natürliche Zahl in der Form $2k$ für eine natürliche Zahl k geschrieben werden kann, und jede ungerade natürliche Zahl als $2k-1$ mit $k \in \mathbb{N}$ und $k \geq 1$. In diesen beiden Fällen erhalten wir nun

$$\begin{aligned} (\beta \circ \alpha)(2k) &= \beta(\alpha(2k)) = \beta(k) \overset{k \geq 0}{=} 2k, \\ (\beta \circ \alpha)(2k-1) &= \beta(\alpha(2k-1)) = \beta(-k) \overset{-k < 0}{=} 2 \cdot (-(-k)) - 1 = 2k - 1. \end{aligned}$$

Kombiniert man die beiden Fälle, so sehen wir, dass $\beta \circ \alpha = \mathrm{id}_{\mathbb{N}_0}$.

Nun müssen wir nachprüfen, dass die Abbildung $\alpha \circ \beta$ genau die Identitätsabbildung der Menge $\mathbb{Z}$ ist. Dafür unterscheiden wir wieder zwei Fälle, nämlich ob die ganze Zahl, die wir einsetzen, negativ oder nicht-negativ ist.

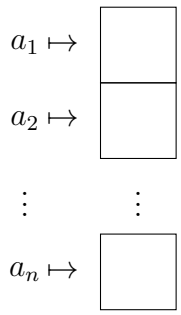
$$\begin{aligned} \text{Für } m \geq 0 : (\alpha \circ \beta)(m) &= \alpha(\beta(m)) = \alpha(2m) = m, \\ \text{Für } m < 0 : (\alpha \circ \beta)(m) &= \alpha(\beta(m)) = \alpha(2(-m) - 1) \overset{-m \in \mathbb{N}}{=} -(-m) = m. \end{aligned}$$

Folglich gilt auch $\alpha \circ \beta = \mathrm{id}_{\mathbb{Z}}$. Somit ist β die inverse Abbildung zu α. Nach Satz 5.8 folgt nun, dass α eine Bijektion zwischen $\mathbb{N}_0$ und $\mathbb{Z}$ ist. Also sind $\mathbb{N}_0$ und $\mathbb{Z}$ nach Definition gleichmächtig. □

Wir bemerken, dass es zwischen zwei endlichen Mengen genau dann eine Bijektion gibt, wenn sie gleich viele Elemente haben. Als nächstes wollen wir bestimmen, wie viele Bijektionen es zwischen zwei endlichen Mengen mit gleicher Anzahl von Elementen geben kann.

Proposition 6.4. *Seien A, B endliche Mengen mit $|A| = |B| = n$. Dann gibt es $n!$ bijektive Abbildungen von A nach B.*

Beweisidee. Seien $A = \{a_1, a_2, \ldots, a_n\}$ und $B = \{b_1, b_2, \ldots, b_n\}$. Wenn wir uns wieder das Bild aus dem Beweis der Proposition 6.1 vor Augen führen:



so kommt erneut in jedes Kästchen ein Element von B, allerdings jetzt mit der Einschränkung, dass jedes Element von B genau einmal vorkommen muss. Das ist allerdings genau dasselbe wie eine Anordnung von B, und da haben wir uns bereits überlegt, dass eine n-elementige Menge auf $n!$ unterschiedliche Arten angeordnet werden kann. □

Wenn die Anzahlen der Elemente zweier endlicher Mengen nicht übereinstimmen, gibt es dazwischen keine Bijektion, aber man kann sich fragen, wie viele injektive oder surjektive Abbildungen es dazwischen gibt. Für die surjektiven Abbildungen ist die Antwort im Allgemeinen kompliziert und soll hier nicht betrachtet werden. Für injektive Abbildungen sieht es allerdings viel besser aus.

Proposition 6.5. *Sei A eine k-elementige Menge und B eine m-elementige Menge. Dann gibt es keine injektive Abbildung von A nach B, wenn $k > m$ ist. Wenn $k \leq m$ ist, dann gibt es $\frac{m!}{(m-k)!}$ unterschiedliche injektive Abbildungen von A nach B.*

Bevor wir die Proposition beweisen, führen wir den folgenden hilfreichen Begriff ein.

Definition 6.6. Sei $\varphi\colon X \to Y$ eine Abbildung. Das **Bild** von φ ist eine Teilmenge von Y, die durch

$$\mathrm{Bild}(\varphi) = \{\varphi(x) \,|\, x \in X\}$$

gegeben ist.

Beispiel.
- Die Abbildung $f\colon [0,\infty) \to [0,\infty)$, $f(x) = 2x + 1$, hat das Bild $\mathrm{Bild}(f) = \{2x + 1 \,|\, x \in [0,\infty)\} = [1,\infty)$, denn $x \geq 0$ ist für reelle Zahlen äquivalent zu $2x + 1 \geq 1$.
- Das Bild der Abbildung $\alpha\colon \{1,2,3\} \to \{a,b,c\}$, $\alpha(1) = a$, $\alpha(2) = b$, $\alpha(3) = a$ ist gegeben durch $\mathrm{Bild}(\alpha) = \{a,b\}$.

Bemerkung. Eine Abbildung $\varphi\colon X \to Y$ ist genau dann surjektiv, wenn $\mathrm{Bild}(\varphi) = Y$ ist, denn das heißt genau, dass jedes Element von Y als Wert der Abbildung φ auftaucht, oder nochmal anders gesprochen, dass jedes Element von Y ein Urbild unter der Abbildung φ hat.

Nun können wir die Beweisidee für die Proposition geben.

Beweisidee der Proposition 6.5. Wir fangen an mit einer kleinen Vorüberlegung. Angenommen, $f\colon A \to B$ ist eine injektive Abbildung. Da die Bilder aller Elemente von A paarweise verschieden sind, können wir schließen, dass $\mathrm{Bild}(f)$ genau k Elemente besitzt.

Betrachten wir nun zunächst den Fall, dass $k > m$ ist. In diesem Fall kann B gar keine k-elementige Teilmenge haben, also kann es keine injektive Abbildung von A nach B geben. Anschaulich gesprochen ist A „zu groß“, sodass nicht jedes Element von A ein eigenes Element von B „nur für sich“ erhalten kann, sondern müssen sich manche Elemente von A dasselbe Bildelement in B teilen.

Nun kommen wir zu dem Fall, dass $k \leq m$ ist, und überlegen hier, wie viele injektive Abbildungen $f\colon A \to B$ es geben kann. Dafür gibt es mehrere möglich Vorgehensweisen, wir entscheiden uns für die folgende. Wir bemerken zunächst wieder, dass $\mathrm{Bild}(f)$ eine k-elementige Teilmenge von f ist. Insbesondere gibt es $\binom{m}{k}$ Möglichkeiten, zunächst das $\mathrm{Bild}(f)$ als Teilmenge festzulegen, bevor man die Abbildung f ganz bestimmt. Um das zu tun, muss man dann die k Elemente des Bildes so auf die k Elemente der Quelle verteilen, dass jedes Element der Quelle genau eins abbekommt und alle Elemente des Bildes genau einmal vergeben werden. Das entspricht genau den Anordnungen von k Objekten, also gibt es bei einem festen Bild $k!$ Möglichkeiten, die injektive Abbildung festzulegen. Nimmt man das für alle möglichen Bildmengen zusammen, so erhalten wir (unter Ausnutzung des Satzes 4.9):

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

injektive Abbildungen von A nach B. Damit ist die Behauptung der Proposition bewiesen. □

Das Zählen von surjektiven Abbildungen ist im Allgemeinen deutlich schwieriger. Allerdings ist der erste Teil der vorherigen Proposition leicht auf den Fall von surjektiven Abbildungen auszuweiten.

Proposition 6.7. *Sei A eine k-elementige Menge und B eine m-elementige Menge, und sei $k < m$. Dann gibt es keine surjektiven Abbildungen von A nach B.*

Die Intuition dahinter ist sehr einfach: In dem beschriebenen Fall hat A „nicht genug Elemente“, um alle Elemente der Zielmenge B zu „treffen“.

Im Allgemeinen haben Injektivität und Surjektivität einer Abbildung nichts miteinander zu tun. Allerdings gibt es ganz spezielle Situationen, in denen doch eine Verbindung existiert. Diese wollen wir als nächstes untersuchen.

Satz 6.8. *Sei A eine* endliche *Menge und $f\colon A \to A$ eine Abbildung. Dann ist f genau dann injektiv, wenn f surjektiv ist.*

Bemerkung. Der Satz gilt auch für Abbildungen $g\colon A \to B$ zwischen zwei *endlichen* Mengen mit *gleicher* Anzahl von Elementen.

Bevor wir uns die Grundidee des Beweises anschauen, wollen wir uns davon überzeugen, dass die Voraussetzungen in dem Satz wirklich nötig sind.

Beispiel. In diesem Beispiel beschreiben wir Abbildungen $\alpha\colon \mathbb{N}_0 \to \mathbb{N}_0$ und $\gamma\colon \mathbb{N}_0 \to \mathbb{N}_0$. Dabei wird α injektiv, aber nicht surjektiv sein, die Abbildung γ hingegen ist surjektiv, aber nicht injektiv. Das veranschaulicht, dass die Endlichkeitsvoraussetzung im obigen Satz von entscheidender Wichtigkeit ist.

Sei die Abbildung $\alpha\colon \mathbb{N}_0 \to \mathbb{N}_0$ durch $\alpha(k) = k + 1$ gegeben. Dann folgt für alle $k_1, k_2 \in \mathbb{N}_0$ aus $\alpha(k_1) = \alpha(k_2)$ bereits $k_1 = k_2$, sodass die Abbildung injektiv ist. Allerdings ist sie nicht surjektiv, da das Element 0 kein Urbild unter α hat.

Die Abbildung $\beta\colon \mathbb{N}_0 \to \mathbb{N}_0$ sei nun durch die folgende Vorschrift definiert.

$$\beta(x) = \begin{cases} x - 1, \text{ falls } x \geq 1, \\ 0, \text{ falls } x = 0. \end{cases}$$

Diese Abbildung ist nicht injektiv, da $\beta(1) = \beta(0) = 0$ ist, aber sie ist surjektiv, da jedes Element $k \in \mathbb{N}_0$ das Element $k + 1$ als Urbild hat.

Beispiel. In diesem Beispiel zeigen wir, dass Endlichkeit alleine nicht ausreicht - hat man endliche Mengen unterschiedlicher Größe, so sind Injektivität und Surjektivität wieder nicht gekoppelt.

Dafür schauen wir uns zunächst die Abbildung $\psi\colon \{1, 2, 3\} \to \{a\}$, die durch $\psi(1) = \psi(2) = \psi(3) = a$ gegeben ist. (Das ist die einzig mögliche Abbildung von $\{1, 2, 3\}$ in die Menge $\{a\}$.) Diese Abbildung ist nicht injektiv, aber surjektiv.

Umgekehrt betrachten wir eine Abbildung $\varphi\colon \{a\} \to \{1, 2, 3\}$, die a auf 2 abbildet. Diese Abbildung ist nun zwar injektiv, aber nicht surjektiv.

Beispiel. Als letztes wollen wir uns ein Beispiel von einer Abbildung anschauen, die den Bedingungen des Satzes genügt. Beispielsweise hat man hier $h\colon \{1, 2, 3, 4\} \to \{1, 2, 3, 4\}$, durch die Vorschrift

$$\begin{aligned} 1 &\mapsto 2 \\ 2 &\mapsto 1 \\ 3 &\mapsto 4 \\ 4 &\mapsto 3 \end{aligned}$$

gegeben ist. Diese Abbildung ist sowohl injektiv als auch surjektiv. Nach der Aussage des Satzes reicht es, eine der beiden Bedingungen zu prüfen, dann ergibt sich die andere in diesem Fall automatisch.

Nun wollen wir einige Überlegungen zu dem Beweis des Satzes festhalten.

Beweisidee des Satzes 6.8. Da wir hier eine Äquivalenz beweisen wollen, können wir diese in zwei Teilaussagen unterteilen, die wir separat beweisen, wie wir bereits für den Satz 5.8 gemacht haben. Sei A also eine endliche Menge und $f\colon A \to A$ eine Abbildung.

Zunächst zeigen wird: Ist f eine injektive Abbildung, so ist f auch surjektiv. Wir haben uns bereits überlegt, dass bei einer injektiven Abbildung mit einer endlichen Quelle gilt: $\mathrm{Bild}(f)$ hat genauso viele Elemente wie die Quelle, also in diesem Fall $|\mathrm{Bild}(f)| = |A|$. Nun ist aber $\mathrm{Bild}(f)$ eine Teilmenge von A, die genauso viele Elemente hat wie A selbst, und deswegen muss $\mathrm{Bild}(f) = A$ gelten. Das ist aber, wie wir bereits bemerkt haben, äquivalent zur Surjektivität von f. Wir haben also gezeigt: Ist $f\colon A \to A$ für eine endliche Menge A injektiv, so ist f auch surjektiv.

Nun kommen wir zu der zweiten Teilaussage. Wir setzen nun voraus, dass f surjektiv ist, und wollen zeigen, dass f injektiv ist. Das machen wir mit einem Widerspruchsbeweis. Angenommen, f wäre surjektiv und nicht injektiv. Das hieße insbesondere, dass es zwei unterschiedliche Elemente a_1, a_2 der Quelle geben müsste, die beide auf dasselbe Element im Ziel abgebildet werden. Da die Bilder dieser beiden Elemente übereinstimmen, können insgesamt höchstens $|A| - 1$ unterschiedliche Bildelemente auftreten, also $|\mathrm{Bild}(f)| \leq |A| - 1$. Das widerspricht aber der Voraussetzung, dass f surjektiv ist, denn diese Voraussetzung ist gleichbedeutend mit $\mathrm{Bild}(f) = A$. Also muss unsere Annahme falsch sein. Daher ist unter unseren Voraussetzungen f notwendigerweise injektiv, wenn f surjektiv ist. □

7 Relationen

Mit Abbildungen können wir Elemente unterschiedlicher Mengen in Beziehung setzen. Bei unserem neuen Thema geht es darum, die Elemente ein und derselben Menge zu vergleichen oder sonstwie in Beziehung miteinander zu setzen. Zunächst sammeln wir einige Beispiele, bevor wir eine abstraktere Formulierung von dem Begriff „Relation“ geben.

Beispiel.
- Die reellen Zahlen vergleichen wir zum Beispiel mit der „$\leq$“-Relation.
- Die natürlichen Zahlen können wir ebenfalls mit „$\leq$“ in Beziehung miteinander setzen. Aber wir können uns auch bei zwei natürlichen Zahlen a, b auch fragen, ob die Zahl a die Zahl b teilt.
- Bei den TeilnehmerInnen X, Y dieser Veranstaltung kann man fragen, ob X mit Y zusammen Übungszettel abgibt.
- Bei den Menschen im Allgemeinen kann bei je zwei Personen fragen, ob diese miteinander verheiratet sind.

Ein formaler Rahmen für den Vergleich von Elementen eine Menge bietet der Begriff einer Relation.

Definition 7.1. Sei A eine Menge. Eine **Relation** auf A ist eine Teilmenge $R \subset A \times A$ vom kartesischen Produkt von A mit sich selbst. Liegt ein Paar (a, b) in R, so schreiben wir dafür aRb oder manchmal auch $a \sim_R b$.

Beispiel.
- Die Menge
$$\{(x, y) \in \mathbb{R}^2 \mid x \leq y\}$$
ist die Menge, die zur $\leq$-Relation auf den reellen Zahlen gehört. Ein Paar (a, b) reeller Zahlen liegt nach Definition genau dann in dieser Menge, wenn $a \leq b$ ist.
- Manchmal kann man eine Relation durch explizite Auflistung aller darin enthaltenen Paare angeben. Ist beispielsweise $A = \{a, b, c\}$, so können wir darauf eine Relation
$$R = \{(a, a), (b, c), (c, b)\} \subset A \times A.$$
Hier gilt beispielsweise aRa, aber nicht bRb. Auch ist bRc wahr, aber nicht aRc.

Relationen können sehr allgemein sein. Um diejenigen Relationen auszuzeichnen, die besondere Eigenschaften haben, führen wir einige Vokabeln ein.

Definition 7.2. Sei A eine Menge und R eine Relation auf A.

- Wir nennen R **reflexiv**, falls für alle Elemente $a \in A$ gilt: aRa, d.h. jedes Element steht in Relation R mit sich selbst.
- Wir nennen R **symmetrisch**, falls für alle Elemente $a, b \in A$ aus aRb folgt, dass auch bRa gilt.
- Wir nennen R **antisymmetrisch**, falls für alle Elemente $a, b \in A$ gilt: Ist aRb und bRa, so folgt daraus sofort $a = b$.
- Wir nennen R **transitiv**, falls für alle Elemente $a, b, c \in A$ aus aRb und bRc folgt, dass auch aRc gilt.

Dabei ist anzumerken, dass „antisymmetrisch“ nicht dasselbe wie „nicht symmetrisch“ bedeutet; es gibt Relationen, die weder antisymmetrisch noch symmetrisch sind, wofür wir später ein Beispiel sehen werden. Andererseits gibt es auch Relationen, die sowohl symmetrisch als auch antisymmetrisch sind. Allerdings sind intuitiv „symmetrisch“ und „antisymmetrisch“ entgegengesetzte Extreme.

Da diese Definitionen recht abstrakt sind, widmen wir uns im Großteil dieses Abschnittes den Beispielen.

Beispiel. Wir betrachten die Menge A aller (lebender) Menschen und die Relation „verheiratet sein“ darauf, d.h. die Person A steht in Relation mit Person B, falls A mit B verheiratet ist. Dabei wollen wir uns mit dem „einfachsten“ Modell der Ehe zufriedengeben.

- Diese Relation ist nicht reflexiv. Um das zu zeigen, würde es reichen, eine einzige Person anzugeben, die nicht mit sich selbst verheiratet ist. Da keine Person mit sich selbst verheiratet ist, ist es insbesondere ausreichend.
- Diese Relation ist symmetrisch: Ist eine Person A mit einer Person B verheiratet, so ist die Person B auch mit der Person A verheiratet.
- Diese Relation ist nicht antisymmetrisch: Nimmt man zwei miteinander verheiratete Personen C, D, so ist C in Relation mit D und D in Relation mit C, allerdings sind C und D dadurch nicht dieselbe Person.
- Diese Relation ist auch nicht transitiv: Nimmt man wieder zwei miteinander verheiratete Personen C, D, so ist C in Relation mit D und D in Relation mit C, also müsste, wenn die Relation transitiv wäre, auch C in Relation zu C stehen, was, wie wir gleich zu Anfang festgestellt haben, nie der Fall ist.

Beispiel. Wir betrachten nun die Relation $\leq$ auf der Menge $\mathbb{R}$ der reellen Zahlen, und untersuchen sie auf Reflexivität, Symmetrie, Antisymmetrie und Transitivität.

- Die Relation $\leq$ ist reflexiv, denn für jede reelle Zahl $x \in \mathbb{R}$ gilt: $x \leq x$.
- Die Relation $\leq$ ist nicht symmetrisch. Um das zu beweisen, reicht es, ein einziges Paar reeller Zahlen (a, b) anzugeben, für die $a \leq b$, aber nicht $b \leq a$ gilt. Das trifft beispielsweise auf $1, 2 \in \mathbb{R}$ zu: Es gilt $1 \leq 2$, aber $2 \not\leq 1$.
- Die Relation $\leq$ ist antisymmetrisch, denn für alle reellen Zahlen $x, y \in \mathbb{R}$ folgt aus $x \leq y$ und $y \leq x$, dass $y = x$ gelten muss.
- Schließlich sehen wir wie folgt, dass $\leq$ eine transitive Relation ist. Für alle Zahlen $x, y, z \in \mathbb{R}$ gilt nämlich: Ist $x \leq y$ und $y \leq z$, so gilt auch $x \leq z$.

Relationen, die sich ähnlich wie $\leq$ verhalten, kommen häufiger in der Mathematik vor und bekommen daher einen eigenen Namen.

Definition 7.3. Sei B eine Menge und R eine Relation auf B. Wir nennen R eine **(partielle) Ordnung** (oder auch eine **Ordnungsrelation**), falls R reflexiv, antisymmetrisch und transitiv ist.

Die Intuition dahinter ist, dass eine solche Relation es erlaubt, Objekte der Menge B zu vergleichen. Es sei an dieser Stelle angemerkt, dass einige Quellen eine weitere Eigenschaft verlangen, um eine Relation „Ordnungsrelation“ zu nennen.

Beispiel. Zum Vergleich betrachten wir die Relation $<$ auf den reellen Zahlen $\mathbb{R}$.

- Diese Relation ist im Gegensatz zu $\leq$ nicht reflexiv. Um das zu zeigen, muss man eine reelle Zahl finden, für die $x \not< x$ gilt. Da dies für alle reelle Zahlen x zutrifft, ist $<$ nicht reflexiv.
- Die Relation $<$ ist, wie auch $\leq$, nicht symmetrisch. Hier können wir dasselbe Beispiel wie oben benutzen: Für $1, 2 \in \mathbb{R}$ gilt $1 < 2$, aber $2 \not< 1$.
- Die Relation $<$ ist antisymmetrisch. Das ist jedoch etwas schwieriger zu sehen als bei $<$. Wir müssen zeigen, dass für alle reellen Zahlen $x, y \in \mathbb{R}$ aus $x < y$ und $y < x$ folgt, dass $y = x$ gelten muss. Allerdings gibt es keine reelle Zahlen $x, y \in \mathbb{R}$, für die sowohl $x < y$ als auch $y < x$ gilt. Da die Prämisse nie erfüllt ist, ist also die gesamte Implikation wahr. Deswegen ist die Relation $<$ antisymmetrisch.
- Der Beweis, dass $<$ eine transitive Relation ist, ist komplett zu $\leq$ analog. Für alle Zahlen $x, y, z \in \mathbb{R}$ gilt nämlich auch: Ist $x < y$ und $y < z$, so gilt auch $x < z$.

Insbesondere sehen wir, dass $<$ keine Ordnungsrelation ist, da diese nicht reflexiv ist. Das ist eine Konventionsfrage; man hätte sicherlich auch andere Begriffe intuitiv zutreffen als „Ordnungsrelationen“ bezeichnen können, hat sich aber z.B. für die obige Version entschieden.

Beispiel. In diesem Beispiel sei M eine feste Menge. Wir betrachten die Relation $\subseteq$ auf der Potenzmenge $\mathcal{P}(M)$ von M, d.h. die Relation, die durch die Menge

$$\{(A, B) \in \mathcal{P}(M) \times \mathcal{P}(M) \mid A \subseteq B\} \subseteq \mathcal{P}(M) \times \mathcal{P}(M)$$

beschrieben wird.

- Diese Relation ist reflexiv, denn für jede Teilmenge A von M gilt: $A \subseteq A$.
- Diese Relation ist antisymmetrisch: Seien A, B Teilmengen von M. Wir müssen zeigen, dass aus $A \subseteq B$ und $B \subseteq A$ folgt, dass $A = B$ ist. Nach Definition von Teilmenge ist $A \subseteq B$ äquivalent dazu, dass jedes Element von A ein Element von B ist, und $B \subseteq A$ bedeutet, dass jedes Element von B auch ein Element von A ist. Insgesamt folgt also, dass die Mengen A und B in diesem Fall dieselben Elemente haben, also $A = B$.
- Diese Relation ist transitiv: Seien A, B, C Teilmengen von M. Gilt $A \subseteq B$ und $B \subseteq C$, so folgt daraus $A \subseteq C$, also ist die Relation $\subseteq$ nach Definition transitiv.
- Ob diese Relation symmetrisch ist, untersuchen wir nur in dem Fall, dass die Menge M nicht die leere Menge ist. In diesem Fall gilt $\varnothing, M \in \mathcal{P}(M)$, ferner $\varnothing \subseteq M$, aber $M \not\subseteq \varnothing$. Also ist die Relation $\subseteq$ nicht symmetrisch, wenn $M \neq \varnothing$ gilt.

Für jede Menge M sehen wir nun, dass die Relation $\subseteq$ auf $\mathcal{P}(M)$ eine Ordnungsrelation ist. Es sei darauf hingewiesen, dass diese Relation einen wesentlichen Unterschied zu der $\leq$-Relation auf den reellen Zahlen aufweist. Hat man zwei reelle Zahlen a, b vorgegeben, so gilt in jedem Fall $a \leq b$ oder $b \leq a$ (oder beides, wenn wir den Fall $a = b$ vorliegen haben). Bei zwei Teilmengen A, B einer Menge M kann es im Allgemeinen passieren, dass weder $A \subseteq B$ noch $B \subseteq A$ gilt. Das motiviert das Adjektiv „partiell“, das manchmal für solche Ordnungen benutzt wird: Die Elemente von $\mathcal{P}(M)$ können nur teilweise verglichen werden.

Beispiel. Bislang haben wir Relationen untersucht, die durch bereits bekannte Arten, Objekte zu vergleichen, hervorgegangen waren. Das ist allerdings im Allgemeinen nicht notwendig. Wir hatten bereits gesehen, dass man

eine Relation auch durch die explizite Angabe der miteinander in Relation stehender Paare definieren kann. Hier betrachten wir die Menge $A = \{a, b, c\}$ und darauf die Relation

$$R = \{(a, a), (a, c), (b, b), (b, c), (c, b)\} \subseteq A \times A.$$

Wir untersuchen wieder die Eigenschaften dieser Relation.

- Diese Relation ist nicht reflexiv, da $(c, c) \notin R$.
- Diese Relation ist nicht symmetrisch: Es gilt aRc, aber $(c, a) \notin R$, also steht c nicht in Relation mit a.
- Diese Relation ist nicht antisymmetrisch: Es gilt nämlich sowohl bRc als auch cRb, aber $b \neq c$.
- Diese Relation ist schließlich auch nicht transitiv: Es gilt nämlich aRc und cRb, aber $(a, b) \notin R$.

Das zeigt, dass es durchaus Relationen gibt, die keine der obigen Eigenschaften haben. Insbesondere ist diese Relation weder antisymmetrisch noch symmetrisch; „antisymmetrisch“ ist also nicht dasselbe wie „nicht symmetrisch“.

Beispiel. Betrachten wir auf der Menge T der TeilnehmerInnen unserer Veranstaltung die Relation „im selben Fachsemester sein“. Die TeilnehmerIn X soll also mit Y in Relation stehen, wenn X im selben Fachsemester ist wie Y.

- Diese Relation ist reflexiv, denn jede(r) Studierende ist im selben Fachsemester wie er/sie selbst.
- Diese Relation ist symmetrisch: Ist X im selben Fachsemester wie Y, so ist auch Y im selben Fachsemester wie X.
- Diese Relation ist nicht antisymmetrisch: Es gibt sicherlich mehrere TeilnehmerInnen im ersten Fachsemester, die allerdings nicht alle ein und dieselbe Person sind.
- Diese Relation ist transitiv: Ist X im selben Fachsemester wie Y und Y im selben Fachsemester wie Z, so sind ja alle drei im selben Fachsemester, also ist insbesondere X im selben Fachsemester wie Z.

Auch Relationen von diesem Typ werden häufiger in der Mathematik behandelt und bekommen wiederum einen Namen.

Definition 7.4. Eine Relation R auf einer Menge A heißt **Äquivalenzrelation**, wenn R reflexiv, symmetrisch und transitiv ist.

Intuitiv wird bei jeder Äquivalenzrelation nur auf die Gleichheit einer bestimmter Eigenschaft der Elemente geachtet, wie der Eigenschaft „Fachsemester“ im obigen Beispiel. Genauso, wie es für manche Zwecke sinnvoll ist, alle Studierende im selben Fachsemester einheitlich zu behandeln, ist es auch manchmal sinnvoll, alle Elemente einer Menge „gleich“ zu behandeln, die miteinander in Relation R stehen, falls R eine Äquivalenzrelation ist. Wir werden das in Kürze noch etwas präzisieren.

Beispiel. Wir betrachten die Menge $\mathbb{R}$ der reellen Zahlen und definieren darauf die folgende Relation:

$$R = \{(x, y) \in \mathbb{R}^2 \mid |x| = |y|\}.$$

Wir wollen zeigen, dass diese Relation eine Äquivalenzrelation ist.

- Diese Relation ist reflexiv, denn $|x| = |x|$ für alle reellen Zahlen x.
- Diese Relation ist symmetrisch, denn aus $|x| = |y|$ folgt auch $|y| = |x|$.
- Diese Relation ist transitiv: Ist für reelle Zahlen x, y, z wahr, dass $|x| = |y|$ und $|y| = |z|$ gilt, so gilt auch $|x| = |z|$.

Ob diese Relation antisymmetrisch ist, ist unerheblich bei der Entscheidung, ob dies eine Äquivalenzrelation ist.

Beispiel. Auf der Menge aller Autos haben wir die Relation „dieselbe Marke haben“, die eine Äquivalenzrelation ist. Tatsächlich ist der Nachweis ganz ähnlich wie im letzten Beispiel:

- Da jedes Auto dieselbe Marke wie es selbst hat, ist die Relation reflexiv.
- Hat das Auto X dieselbe Marke wie das Auto Y, so hat natürlich auch das Auto Y dieselbe Marke wie X. Also ist die Relation symmetrisch.
- Die Relation ist auch transitiv: Ist X von derselben Marke wie Y und Y wie Z, so haben alle drei Autos dieselbe Marke, also insbesondere auch X und Z.

Um die Elemente mit derselben Eigenschaft zusammenzufassen, definiert man Äquivalenzklassen.

Definition 7.5. Sei A eine Menge und R eine Äquivalenzrelation auf A. Sei ferner a ein Element von A. Die **Äquivalenzklasse von** a (bezüglich R) ist die Teilmenge von A, die durch

$$[a] \overset{\text{Def}}{=} \{b \in A \mid bRa\}$$

gegeben ist. Die Menge $[a]$ enthält also genau die Elemente von A, die mit a in Relation R stehen.

Beispiel.
- Im Beispiel 7 sind die Äquivalenzklassen die Menge aller Studierenden im 1. Fachsemester, im 2. Fachsemester, im 3. Fachsemester usw. Insbesondere ist jede(r) TeilnehmerIn in genau einer dieser Äquivalenzklassen.
- Im Beispiel 7 bilden alle Autos einer bestimmten Marke M eine Äquivalenzklasse.
- Im Beispiel 7 gilt beispielsweise $[\sqrt{2}] = \{\sqrt{2}, -\sqrt{2}\}$. Allgemein gilt für jedes $x \in \mathbb{R}$, dass $[x] = \{x, -x\}$ ist. Die Äquivalenzklasse jeder reellen Zahl außer 0 enthält also genau zwei Elemente, während die Äquivalenzklasse von 0 nur das Element 0 enthält.

Die folgende Proposition über Äquivalenzklassen ist eine wichtige Aussage, die wir allerdings nicht beweisen werden.

Proposition 7.6. *Sei A eine Menge, R eine Äquivalenzrelation auf A. Dann gilt:*

a) Für jedes Element $a \in A$ ist a ein Element von $[a]$.

b) Ist $a \in A$ und steht $b \in A$ in Relation R mit a (also $(b, a) \in R$), dann gilt $[a] = [b]$.

c) Die Menge A kann als Vereinigung disjunkter Äquivalenzklassen (von R) geschrieben werden.

Wir veranschaulichen uns diese Proposition anhand der bereits erwähnten Beispiele:

Beispiel.
- Im Beispiel 7 kann man alle TeilnehmerInnen in Gruppen nach Fachsemester einteilen, und jede(r) TeilnehmerIn wird genau einer solchen Gruppe angehören.
- Genauso bei den Autos: Wir können die Autos ihrer Marke entsprechend gruppieren, und jedes Auto gehört dann genau einer solchen Gruppe an.

Nun wollen wir weitere Äquivalenzrelationen untersuchen, die nicht ganz so einfach sind.

Beispiel. Wir betrachten die Relation $\sim$ auf der Menge $\mathbb{R}^2$, also auf der Menge der Paare reeller Zahlen, die durch

$$(x, y) \sim (z, w) \overset{\text{Def}}{\Leftrightarrow} x^2 + y^2 = z^2 + w^2$$

gegeben ist. Dabei wollen wir uns, wie aus dem Schulunterricht bekannt, die Elemente von $\mathbb{R}^2$ als Punkte in der Koordinatenebene veranschaulichen. Wieder wollen wir zeigen, dass dies eine Äquivalenzrelation ist.

- Diese Relation ist reflexiv: Ist $(x, y) \in \mathbb{R}^2$, so gilt $x^2 + y^2 = x^2 + y^2$, also ist nach Definition von $\sim$ die Aussage $(x, y) \sim (x, y)$ wahr.
- Diese Relation ist symmetrisch: Ist $(x, y) \sim (z, w)$, so ist das äquivalent zu $x^2 + y^2 = z^2 + w^2$ und das wiederum zu $z^2 + w^2 = x^2 + y^2$. Letzteres bedeutet nach Definition von $\sim$, dass $(z, w) \sim (x, y)$ gilt.
- Diese Relation ist transitiv: Falls für $(x, y), (z, w), (r, s) \in \mathbb{R}^2$ die Aussagen $(x, y) \sim (z, w)$ und $(z, w) \sim (r, s)$ wahr, so können wir das nach Definition von $\sim$ zu

  $$x^2 + y^2 = z^2 + w^2 \text{ und } z^2 + w^2 = r^2 + s^2$$

  umschreiben. Das impliziert wiederum $x^2 + y^2 = r^2 + s^2$, also gilt auch $(x, y) \sim (r, s)$.

Nun geht es uns darum, die Äquivalenzklassen dieser Relation zu beschreiben. Zunächst betrachten wir den Sonderfall $[(0, 0)]$. Nach Definition von Äquivalenzklassen ist dies

$$\begin{aligned} [(0,0)] &= \{(x,y) \in \mathbb{R} \,|\, (x,y) \sim (0,0)\} \\ &= \{(x,y) \in \mathbb{R} \,|\, x^2 + y^2 = 0^2 + 0^2\} \\ &= \{(x,y) \in \mathbb{R} \,|\, x^2 + y^2 = 0\}. \end{aligned}$$

Nun ist $x^2 \geq 0$ und $y^2 \geq 0$, also kann ihre Summe nur 0 sein, wenn $x^2 = y^2 = 0$ ist, also ist $[(0, 0)] = \{(0, 0)\}$.

Als nächstes setzen wir uns mit dem Spezialfall $[(1, 0)]$ auseinander. Zunächst einige Beispiele. So ist etwa $(-1, 0) \sim (1, 0)$, da $(-1)^2 + 0^2 = 1^2 + 0^2$. Auch ist $(0, 1) \sim (1, 0)$, da $0^2 + 1^2 = 1^2 + 0^2$ gilt. Ferner ist auch $(\frac{3}{5}, \frac{4}{5}) \sim (1, 0)$, da

$$\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2 = \frac{9 + 16}{25} = 1 = 1^2 + 0^2.$$

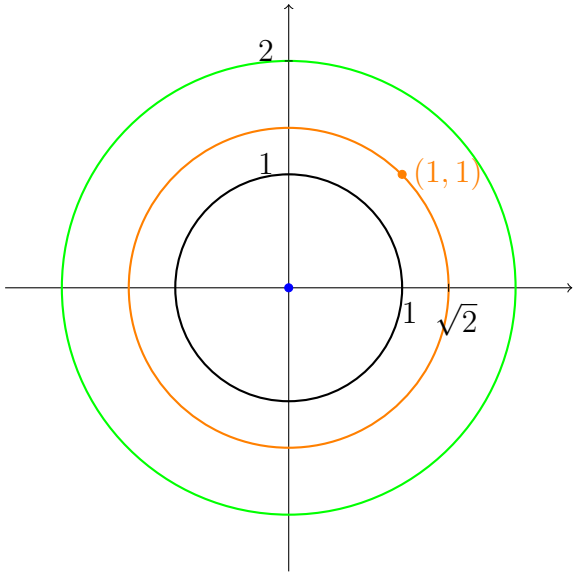
Die Äquivalenzklasse $[(1, 0)]$ ist nach Definition gegeben durch

$$\begin{aligned} [(1,0)] &= \{(x,y) \in \mathbb{R} \,|\, (x,y) \sim (1,0)\} \\ &= \{(x,y) \in \mathbb{R} \,|\, x^2 + y^2 = 1^2 + 0^2\} \\ &= \{(x,y) \in \mathbb{R} \,|\, x^2 + y^2 = 1\}. \end{aligned}$$

Aus dem Schulunterricht ist nun bekannt, dass die Menge der Punkte, die durch die Gleichung $x^2 + y^2 = 1$ beschrieben wird, der Kreis mit Radius 1 mit dem Ursprung als Mittelpunkt ist. Also bilden die Punkte auf diesem Kreis genau die Restklasse von $(1, 0)$.

Ist nun $(a, b) \neq (0, 0)$ ein beliebiges, festes Element von $\mathbb{R}^2$, so kann $[(a, b)]$ als ein Kreis versanschaulicht werden, dessen Mittelpunkt der Ursprung und der den Radius $\sqrt{a^2 + b^2}$ hat.

In dem Bild sind als Beispiel 4 Äquivalenzklassen hervorgehoben: $[(0, 0)]$ in blau, $[(1, 0)]$ in schwarz, $[(1, 1)]$ in orange und $[(0, 2)]$ in grün.



Im nächsten Beispiel betrachten wir eine Äquivalenzrelation einer ganz anderen Natur.

Beispiel. Wir betrachten Elemente der Menge $\mathbb{Z} \times \mathbb{N}$, wobei in diesem Beispiel $\mathbb{N}$ die natürlichen Zahlen *ohne* 0 bezeichne. Darauf definieren wir nun die Relation $\approx$ wie folgt:

$$(a, b) \approx (c, d) \overset{\text{Def}}{\Leftrightarrow} ad = bc.$$

Zum besseren Verständnis schreiben wir uns einige Paare auf, die miteinander in Relation $\approx$ stehen.

- $(2, 1) \approx (6, 3)$, da $2 \cdot 3 = 1 \cdot 6$.
- $(-1, 3) \approx (-5, 15)$, da $(-1) \cdot 15 = 3 \cdot (-5)$.
- $(1, 1) \approx (2, 2)$, $(1, 1) \approx (3, 3)$, allgemeiner, $(1, 1) \approx (a, a)$ für jedes $a \in \mathbb{N}$, da $1 \cdot a = 1 \cdot a$.
- $(0, 1) \approx (0, 2)$, da $0 \cdot 2 = 1 \cdot 0$.

Nun wollen wir zeigen, dass $\approx$ eine Äquivalenzrelation auf $\mathbb{Z} \times \mathbb{N}$ ist.

- Diese Relation ist reflexiv, für jedes Paar $(a, b) \in \mathbb{Z} \times \mathbb{N}$ gilt: $a \cdot b = b \cdot a$, also auch $(a, b) \approx (a, b)$.
- Diese Relation ist symmetrisch: Sei $(a, b) \approx (c, d)$, also nach Definition $ad = bc$. Wir haben zu zeigen: $(c, d) \approx (a, b)$, also müssen wir nach Definition von $\approx$ die Gleichheit $cb = da$ zeigen. Diese folgt unmittelbar aus der Voraussetzung.

- Beim Nachweis der Transitivität müssen wir zum ersten Mal ein etwas längeres Argument geben. Sei also $(a,b) \approx (c,d)$ und $(c,d) \approx (e,f)$. Wir müssen zeigen, dass dann auch $(a,b) \approx (e,f)$ ist, also dass $af = be$ gilt. Nach Definition von $\approx$ ist die Voraussetzung äquivalent zu $ad = bc$ und $cf = de$. Wir nehmen die erste Gleichung und multiplizieren diese mit $f \neq 0$:

$$\begin{aligned} ad &= bc \,|\cdot f \neq 0 \\ \Leftrightarrow adf &= bcf. \end{aligned}$$

  Nun verwenden wir die zweite Gleichung, um auf der rechten Seite cf durch de zu ersetzen, und erhalten $adf = bde$. Da auch $d \neq 0$ ist, können wir beide Seiten der Gleichung durch d teilen und erhalten

$$\begin{aligned} adf &= bde \,|\, : d \neq 0 \\ \Leftrightarrow af &= be, \end{aligned}$$

  was genau die gewünschte Aussage ist. Insgesamt sehen wir also, dass $\approx$ eine Äquivalenzrelation ist.

Als letztes versuchen wir, die Äquivalenzklassen dieser Äquivalenzrelation etwas besser zu beschreiben. Da für $(a,b),(c,d) \in \mathbb{Z} \times \mathbb{N}$ sowohl $b \neq 0$ als auch $d \neq 0$ gilt, ist die Bedingung $ad = bc$ äquivalent zu

$$\frac{a}{b} = \frac{c}{d}.$$

Zwei Paare sind also genau äquivalent, wenn das jeweilige Verhältnis von der ersten zu der zweiten Zahl gleich ist. In der Äquivalenzklasse von $(a,b) \in \mathbb{Z} \times \mathbb{N}$ befinden sich also genau die Elemente $(x,y) \in \mathbb{Z} \times \mathbb{N}$, für die der Bruch $\frac{x}{y}$ dieselbe rationale Zahl darstellt wie $\frac{a}{b}$. All solche Zahlen werden aus $\frac{a}{b}$ durch Kürzen und Erweitern entstehen. Das liefert also eine Beschreibung der Äquivalenzklassen von $\approx$.

Teil II
Zahlentheorie

In der Zahlentheorie beschäftigt man sich, wie der Name schon sagt, mit Zahlen. Im Rahmen dieser Vorlesung werden dies vor allem natürliche und ganze Zahlen sein, die wir in der Zahlentheorie untersuchen. Die Faszination von Zahlen ist sehr alt und gehört zu den Grundsteinen der Mathematik. Uns sind da vor allem die Erkenntnisse der alten Griechen geläufig, die sich von der Welt der Zahlen verzaubern ließen: Euklid, Pythagoras, Eratosthenes, Diophantos sind nur einige wenige aus der langen Liste derer, dessen Erkenntnisse bis heute bewundert werden. Auch schon im alten Ägypten und Babylonien wurde Zahlentheorie betrieben. Viele Jahrtausende lang war Zahlentheorie vor allem eine Art Gedankenspiel, eine Herausforderung für das Denken, die wenig bis keine praktische Verwendung fand. Das ändert sich massiv im 20. Jahrhundert, als - unter anderem durch den Einsatz von Computern - neue Verschlüsselungsmethoden sowohl nötig als auch möglich wurden. Heute spielt die Zahlentheorie eine entscheidende Rolle für die Kryptographie, die Lehre der Verschlüsselung. Als einige Beispiele wären da das sehr bekannte RSA-Verschlüsselungsverfahren zu nennen, aber auch Diffie-Hellman-Schlüsselaustausch und Secure Hash Algorithms (SHA); es gibt auch viele weitere. Wir stehen allerdings ganz am Anfang und wollen zunächst die ersten Grundlagen der Zahlentheorie erlernen.

8 Teilbarkeit

Wir haben bereits über den Begriff der Teilbarkeit gesprochen (Definition 3.5). Wir können die Definition auf ganze Zahlen ausweiten:

Definition 8.1. Wir sagen, die ganze Zahl $k \neq 0$ **teilt** die ganze Zahl m, falls $\frac{m}{k}$ eine ganze Zahl ist.

Der Begriff der Teilbarkeit gehört zu den Besonderheiten der ganzen und natürlichen Zahlen. Hat man etwa zwei rationale Zahlen $x, y \in \mathbb{Q}$, $y \neq 0$, so ist die Zahl $\frac{x}{y}$ stets eine rationale Zahl. Würden wir also jedes Auftreten des Worts „ganz“ durch „rational“ ersetzen, so würde jede rationale Zahl durch jede rationale Zahl $\neq 0$ teilbar sein, was ja bei ganzen Zahlen überhaupt nicht der Fall ist! Genauso ist es mit reellen Zahlen: Sind $u, v \in \mathbb{R}$ mit $v \neq 0$, so ist auch $\frac{u}{v}$ eine reelle Zahl, Teilbarkeit von reellen Zahlen wäre also ganz schön langweilig. Auf den ganzen und natürlichen Zahlen gibt es bei der Teilbarkeit viele spannende Probleme. Wir wollen im Folgenden die Teilbarkeitsrelation genauer untersuchen. Unser erstes Resultat ist wie folgt.

Proposition 8.2. *Die Teilbarkeitsrelation auf* $\mathbb{N} \setminus \{0\}$*, also die Relation*

$$\{(a, b) \in \mathbb{N} \setminus \{0\} \times \mathbb{N} \setminus \{0\} \mid a \text{ teilt } b\},$$

ist eine Ordnungsrelation.

Beweis. Nach Definition der Ordnungsrelation müssen wir nachprüfen, dass die Teilbarkeitsrelation reflexiv, antisymmetrisch und transitiv ist.

- Zur Reflexivität: Sei a eine beliebige positive natürliche Zahl. Dann ist a durch a teilbar, da $\frac{a}{a} \in \mathbb{Z}$ ist. Also ist die Teilbarkeitsrelation reflexiv.
- Zur Antisymmetrie: Seien a, b positive natürliche Zahlen und sei sowohl a ein Teiler von b als auch b ein Teiler von a. Wir wollen zeigen, dass dann $a = b$ gelten muss. Nach Definition von Teilbarkeit ist die Voraussetzung äquivalent dazu, dass $\frac{a}{b}$ und $\frac{b}{a}$ beide ganze Zahlen sind. Da sowohl a als auch b positiv sind, sind auch $\frac{a}{b}$ und $\frac{b}{a}$ positive ganze Zahlen, also beide mindestens 1. Daraus ergibt sich also

  $$\frac{a}{b} \geq 1 \textbf{ und } \frac{b}{a} \geq 1.$$

  Nun nutzen wir erneut aus, dass sowohl a als auch b positiv sind, und multiplizieren beide Seiten der ersten Ungleichung mit b und der zweiten Ungleichung mit a. Das sind jeweils Äquivalenzumformungen, also erhalten wir

  $$\begin{aligned} \frac{a}{b} \geq 1 \,|\cdot b > 0 \quad &\textbf{und} \quad \frac{b}{a} \geq 1 \,|\cdot a > 0 \\ \Leftrightarrow a \geq b \quad &\textbf{und} \quad b \geq a. \end{aligned}$$

  Das wiederum impliziert, dass $a = b$ gilt. Damit haben wir bewiesen, dass die Teilbarkeitsrelation auf $\mathbb{N} \setminus \{0\}$ antisymmetrisch ist.
- Zur Transitivität: Seien also a, b, c positive natürliche Zahlen, sodass die Zahl a die Zahl b teilt und die Zahl b die Zahl c teilt. Wir müssen zeigen, dass dann auch a die Zahl c teilt. Nach Voraussetzung wissen wir, dass sowohl $\frac{b}{a}$ als auch $\frac{c}{b}$ ganze Zahlen sind. Somit ist auch ihr Produkt

  $$\frac{b}{a} \cdot \frac{c}{b} = \frac{c}{a}$$

  auch eine ganze Zahl. Nach Definition von Teilbarkeit bedeutet das aber gerade, dass c durch a teilbar ist. Also ist die Teilbarkeitsrelation auch transitiv. Insgesamt haben wir also gezeigt, dass die Teilbarkeitsrelation auf $\mathbb{N} \setminus \{0\}$ eine partielle Ordnung ist.

□

Da wir häufiger mit der Teilbarkeitsrelation arbeiten werden, führen wir dafür die folgende abkürzende Notation ein.

Notation. Seien a, b ganze Zahlen und $a \neq 0$. Wir schreiben $a|b$ abkürzend für „a teilt b“.

Bemerkung.

- Die Teilbarkeitsrelation auf $\mathbb{N} \setminus \{0\}$ ist nicht symmetrisch: Beispielsweise ist 2 ein Teiler von 4, aber 4 ist kein Teiler von 2.
- Die Teilbarkeitsrelation auf $\mathbb{Z} \setminus \{0\}$ ist keine Ordnungsrelation, da sie zwar reflexiv und transitiv ist, aber nicht antisymmetrisch. Im Beweis der Reflexivität und Transitivität bei der vorangehenden Proposition haben wir die Positivität der Zahlen nicht benutzt; bei der Antisymmetrie spielte sie allerdings eine entscheidende Rolle. Um zu zeigen, dass die Relation nicht antisymmetrisch ist, brauchen wir ein Beispiel von ganzen Zahlen $a, b \in \mathbb{Z} \setminus \{0\}$, die zwar verschieden sind, für die aber sowohl $a|b$ als auch $b|a$ gilt. Das ist beispielsweise bei Zahlen -2 und 2 der Fall: Es ist

$$\frac{2}{-2} = \frac{-2}{2} = -1,$$

also beides eine ganze Zahl. Somit ist 2 durch -2 und -2 durch 2 teilbar.

Weitere Eigenschaften von Teilbarkeit fassen wir in der folgenden Proposition zusammen.

Proposition 8.3. *Seien k, l, m, n ganze Zahlen.*

(1) Sei $k \neq 0$. Ist sowohl l als auch m durch k teilbar, so ist sowohl die Summe $l + m$ als auch die Differenz $l - m$ durch k teilbar. Kurz:

$$\begin{aligned} k|l \wedge k|m &\Rightarrow k|(l+m), \\ k|l \wedge k|m &\Rightarrow k|(l-m). \end{aligned}$$

(2) Ist hingegen l durch $k \neq 0$ und n durch $m \neq 0$ teilbar, so muss im Allgemeinen $(l + n)$ nicht durch $(k + m)$ teilbar sein. Kurz:

$$k|l \wedge m|n \not\Rightarrow (k+m)|(l+n)$$

(3) Ist $k \neq 0$ ein Teiler von l, so ist k auch ein Teiler von jedem Vielfachen $l \cdot n$ von l. Kurz:

$$k|l \Rightarrow k|ln$$

(4) Ist l durch $k \neq 0$ und n durch $m \neq 0$ teilbar, so ist auch ln durch $km \neq 0$ teilbar. Kurz:

$$k|l \wedge m|n \Rightarrow km|ln$$

(5) *Sind $k \neq 0$ und $l \neq 0$ Teiler von m, so muss $kl \neq 0$ im Allgemeinen kein Teiler von m sein. Kurz:*

$$k|m \wedge l|m \quad \not\Rightarrow \quad kl|m$$

Bemerkung. Damit wir von Teilbarkeit sprechen können, sollte die Zahl, durch die wir teilen, nicht 0 sein. Daher wird es manchmal nötig, vorauszusetzen, dass einige der Zahlen nicht 0 sind.

Beweis. Seien k, l, m, n ganze Zahlen.

zu (1): Sei also $k \neq 0$. Ist sowohl l als auch m durch k teilbar, so gilt nach Definition von Teilbarkeit, dass sowohl $\frac{l}{k}$ als auch $\frac{m}{k}$ ganze Zahlen sind. Somit sind sowohl ihre Summe als auch ihre Differenz ganze Zahlen, also $\frac{l+m}{k}$ und $\frac{l-m}{k}$. Das heißt nach Definition von Teilbarkeit aber gerade, dass $l + m$ und $l - m$ durch k teilbar sind.

zu (2): Um zu zeigen, dass dies nicht für alle ganze Zahlen l, k, m, n mit $k \neq 0$ und $m \neq 0$ gilt, reicht es ein Beispiel anzugeben, wo $k|l$ und $m|n$ und $(k+m) \not| (l+n)$ gilt. Beispielsweise gilt $2|4$ und $3|9$, aber $4+9 = 13$ ist nicht durch $2 + 3 = 5$ teilbar.

zu (3): Sei $k \neq 0$ ein Teiler von l. Wir wollen zeigen, dass $l \cdot n$ ebenfalls durch k teilbar ist. Nach Definition von Teilbarkeit ist $\frac{l}{k}$ eine ganze Zahl. Also ist deren Produkt mit der ganzen Zahl n wiederum eine ganze Zahl. Diese können wir als $\frac{l}{k} \cdot n = \frac{ln}{k}$ schreiben. Dass dies eine ganze Zahl ist, bedeutet aber gerade, dass ln durch k teilbar ist.

zu (4): Sei l durch $k \neq 0$ und n durch $m \neq 0$ teilbar. Wir wollen zeigen, dass dann auch ln durch km teilbar. Zunächst bemerken wir, dass aus $k \neq 0$ und $m \neq 0$ folgt, dass $km \neq 0$ ist. Nun wissen wir aus der Voraussetzung, dass $\frac{l}{k}$ und $\frac{n}{m}$ beides ganze Zahlen sind. Folglich ist auch deren Produkt $\frac{l}{k} \cdot \frac{n}{m} = \frac{ln}{km}$ eine ganze Zahl. Das bedeutet aber gerade, dass ln durch km teilbar ist.

zu (5): Hier reicht es wiederum, ein Gegenbeispiel anzugeben. So ist etwa die Zahl 4 sowohl durch 2 als auch durch 4 teilbar, aber nicht durch $2 \cdot 4 = 8$.

□

Bemerkung.

- Obwohl zwei der Resultate aus der letzten Proposition von der Form sind „Es gilt *nicht für alle* ganze Zahlen k, l, m, n, dass...“, gibt es durchaus ganze Zahlen k, l, m, n, für die die entsprechend modifizierte Aussagen wahr sind: Beispielsweise ist 2 ein Teiler von 4 und 3 ein Teiler von 6 und $2 + 3 = 5$ ist auch ein Teiler von $4 + 6 = 10$. So etwas kann also vorkommen, ist aber nicht für alle ganzen Zahlen der Fall.

Ebenfalls gilt: 2 ist ein Teiler von 12 und 3 ist ein Teiler von 12 und auch das Produkt $2 \cdot 3 = 6$ ist ein Teiler von 12. In diesem Fall werden wir später eine recht allgemeine Bedingung angeben, die garantiert, dass eine zu (5) analoge Aussage wahr ist. In der Allgemeinheit der obigen Proposition wird das natürlich nicht funktionieren.

- Im Allgemeinen vertragen sich Teilbarkeit und Division schlecht. Hat man etwa Zahlen l und m, die beide durch eine Zahl $k \neq 0$ teilbar sind, so könnte etwa der Quotient $\frac{l}{m}$ nicht definiert sein (wenn $m = 0$) oder eine rationale, aber keine ganze Zahl sein (etwa bei $l = 1$ und $m = 2$). Selbst wenn $\frac{l}{m}$ eine ganze Zahl ist, braucht sie nicht wieder durch k teilbar sein. So ist beispielsweise $2|6$ und $2|2$, aber 2 ist kein Teiler von $\frac{6}{2} = 3$.

Sind zwei ganze Zahlen durch eine Zahl $k \neq 0$ teilbar, so auch ihre Summe, wie wir in der letzten Proposition bewiesen haben. Als nächstes wollen wir untersuchen, wann die Summe zweier *nicht* durch k teilbarer Zahlen durch k teilbar ist.

Beispiel. Zunächst schauen wir uns für $k = 3$ an, wann Summe zweier nicht durch 3 teilbarer Zahlen selbst durch 3 teilbar ist. Wir stellen fest:

$1 + 1 = 2$	ist nicht durch 3 teilbar
$1 + 2 = 3$	ist durch 3 teilbar
$2 + 1 = 3$	ist durch 3 teilbar
$2 + 2 = 4$	ist nicht durch 3 teilbar
$1 + 4 = 5$	ist nicht durch 3 teilbar
$2 + 4 = 6$	ist durch 3 teilbar

und so weiter. Man kommt durch weiteres Ausprobieren zu der Vermutung, dass man alle natürliche Zahlen in drei Spalten schreiben kann,

1	2	3
4	5	6
7	8	9
10	11	12
...	...	...

und die Summe einer Zahl aus der ersten und einer aus der zweiten Spalte immer durch 3 teilbar ist, während die Summe zweier Zahlen aus der ersten oder die Summe zweier Zahlen aus der zweiten Spalte nie durch 3 teilbar ist. (Dabei sprechen wir nicht über die dritte Spalte, da alle Zahlen darin bereits durch 3 teilbar sind.) Wir bemerken noch, dass die Differenz zweier aufeinanderfolgenden Zahlen in jeder Spalte 3 ist, und die Differenz je zweier Zahl in einer Spalte durch 3 teilbar ist. Um nun dieses und ähnliche Phänomene zu untersuchen, wollen wir uns mit modularer Arithmetik beschäftigen.

9 Modulare Arithmetik

Wir definieren eine Relation auf $\mathbb{Z}$, die genauere Informationen zur Teilbarkeit durch eine feste Zahl m erfasst. Wir definieren nun diese Relation und zeigen als erstes, dass dies eine Äquivalenzrelation ist.

Definition 9.1. Sei m eine feste natürliche Zahl, $m \neq 0$. Wir definieren die Relation „kongruent modulo m“ auf $\mathbb{Z}$ durch

$$\mathrm{Mod}_m = \{(a,b) \in \mathbb{Z} \times \mathbb{Z} \,|\, m \text{ teilt } a-b\}.$$

Ist $(a,b) \in \mathrm{Mod}_m$, so sagen wir, dass a **kongruent zu** b **modulo** m ist und schreiben $a \equiv b \mod m$.

Satz 9.2. *Sei m eine feste natürliche Zahl, $m \neq 0$. Dann ist Mod_m eine Äquivalenzrelation auf $\mathbb{Z}$.*

Beweis. Wir müssen zeigen, dass die Relation Mod_m reflexiv, symmetrisch und transitiv ist.

- Reflexivität: Für jede ganze Zahl a gilt: $a - a = 0$ und somit durch m teilbar. Also gilt $a \equiv a \mod m$ für alle $a \in \mathbb{Z}$ und somit ist die Relation reflexiv.

- Symmetrie: Für alle ganzen Zahlen $a, b \in \mathbb{Z}$ wollen wir zeigen, dass aus $a \equiv b \mod m$ auch schon $b \equiv a \mod m$ folgt. Sei also $a \equiv b \mod m$. Nach Definition der Kongruenz ist das äquivalent dazu, dass m die Zahl $a-b$ teilt, also nach Definition von Teilbarkeit äquivalent dazu, dass $\frac{a-b}{m}$ eine ganze Zahl ist. Dann ist aber auch das Produkt dieser Zahl mit (-1) eine ganze Zahl, also die Zahl

  $$(-1) \cdot \frac{a-b}{m} = \frac{b-a}{m}.$$

  Das bedeutet aber genau, dass $b-a$ durch m teilbar ist, also dass $(b,a) \in \mathrm{Mod}_m$ gilt. Also ist die Relation Mod_m symmetrisch.

- Transitivität: Seien a, b, c ganze Zahlen. Wir müssen zeigen, dass aus $a \equiv b \mod m$ und $b \equiv c \mod m$ folgt, dass $a \equiv c \mod m$ gilt. Nach Voraussetzung sind $\frac{a-b}{m}$ und $\frac{b-c}{m}$ ganze Zahlen, also ist auch deren Summe $\frac{a-b}{m} + \frac{b-c}{m} = \frac{a-c}{m}$ eine ganze Zahl. Das heißt aber wiederum, dass m die Zahl $a-c$ teilt, also $a \equiv c \mod m$ ist. Somit haben wir die Transitivität von Mod_m nachgewiesen.

Insgesamt folgt also, dass Mod_m eine Äquivalenzrelation ist. □

Zu jeder Äquivalenzrelation, also insbesondere zu Mod_m, können wir die zugehörigen Äquivalenzklassen betrachten. Wir schauen uns das an Beispielen an.

Beispiel. Wir betrachten die Relation Mod_3 auf $\mathbb{Z}$. In der Äquivalenzklasse von 4 bezüglich der Relation Mod_3, die wir mit $[4]_3$ (oder manchmal nur mit $[4]$) notieren, liegen beispielsweise $7, 10, 13, 1, -2, -5$ (und unendlich viele weitere ganze Zahlen), aber nicht etwa 8 oder 12. Die positiven Elemente daraus stimmen genau mit der ersten Spalte im Beispiel, das wir zum Schluss des letzten Kapitels betrachtet haben, überein. Allgemein kann man diese wie folgt beschreiben:

$$[4] = \{3r + 4 \,|\, r \in \mathbb{Z}\}.$$

Um das zu beweisen, müssen wir uns vergewissern, dass jedes Element von $[4]$ als $3r+4$ mit irgendeinem $r \in \mathbb{Z}$ geschrieben werden kann, und dass jedes $3r + 4$ mit $r \in \mathbb{Z}$ tatsächlich in $[4]$ liegt. Letzteres ist die einfachere Hälfte: Um zu zeigen, dass $3r + 4 \equiv 4 \mod 3$ ist, müssen wir zeigen, dass $\frac{(3r+4)-4}{3}$ eine ganze Zahl ist. Da Vereinfachen $\frac{(3r+4)-4}{3} = r$ ergibt, folgt $3r + 4 \in [4]$ für alle $r \in \mathbb{Z}$. Sei nun $a \in [4]$ beliebig, d.h. a ist eine ganze Zahl und $a \equiv 4 \mod 3$. Das bedeutet nach Definitionen von Kongruenz und Teilbarkeit, dass $\frac{a-4}{3}$ eine ganze Zahl ist. Setze $s = \frac{a-4}{3}$, dann gilt: $s \in \mathbb{Z}$ und

$$3s + 4 = 3 \cdot \frac{a-4}{3} + 4 = (a-4) + 4 = a,$$

also ist $a = 3s + 4$ mit $s \in \mathbb{Z}$. Das zeigt also, dass die Mengen $[4]$ und $\{3r + 4 \,|\, r \in \mathbb{Z}\}$ dieselben Elemente haben, also gleich sind.

Wir wollen mit Zahlen „modulo m“ rechnen können. Schon der Titel „modulare Arithmetik“ weist darauf hin, dass es in diesem Abschnitt um Grundrechenarten gehen wird. In dem nächsten Satz zeigen wir, dass Kongruenzen sich gut mit Summe, Produkt und Differenz verhalten. Anschließend sehen wir ein Beispiel dafür, dass Kongruenzen und Quotienten hingegen im Allgemeinen nicht miteinander verträglich sind.

Proposition 9.3. *Sei m eine feste natürliche Zahl, $m \neq 0$.Seien a_1, a_2, b_1, b_2 ganze Zahlen mit $a_1 \equiv a_2 \mod m$ und $b_1 \equiv b_2 \mod m$. Dann gilt auch:*

(1) $a_1 + b_1 \equiv a_2 + b_2 \mod m$

(2) $a_1 - b_1 \equiv a_2 - b_2 \mod m$

(3) $a_1 \cdot b_1 \equiv a_2 \cdot b_2 \mod m$

Beweis. Seien m und a_1, a_2, b_1, b_2 wie in der Proposition. Die Voraussetzungen an a_1, a_2, b_1, b_2 sind äquivalent dazu, dass $\frac{a_1-a_2}{m}$ und $\frac{b_1-b_2}{m}$ beides ganze Zahlen sind.

zu (1): Wir müssen zeigen, dass $a_1 + b_1 \equiv a_2 + b_2 \mod m$ gilt. Das ist äquivalent dazu, dass $\frac{(a_1+b_1)-(a_2+b_2)}{m}$ eine ganze Zahl ist. Wir formen diesen Ausdruck um und erhalten:

$$\begin{aligned}\frac{(a_1+b_1)-(a_2+b_2)}{m} &= \frac{a_1+b_1-a_2-b_2}{m} \\ &= \frac{a_1-a_2+b_1-b_2}{m} = \frac{a_1-a_2}{m} + \frac{b_1-b_2}{m},\end{aligned}$$

also ist diese Zahl Summe zweier Zahlen, die nach Voraussetzung ganze Zahlen sind, und somit selbst eine ganze Zahl. Damit ist die erste Behauptung bewiesen.

zu (2): Diesmal müssen wir zeigen, dass $a_1 - b_1 \equiv a_2 - b_2 \mod m$ gilt. Der Beweis ist dem ersten sehr ähnlich. Die Behauptung ist äquivalent dazu, dass $\frac{(a_1-b_1)-(a_2-b_2)}{m}$ eine ganze Zahl ist. Wir formen diesen Ausdruck auch um und erhalten:

$$\begin{aligned}\frac{(a_1-b_1)-(a_2-b_2)}{m} &= \frac{a_1-b_1-a_2+b_2}{m} \\ &= \frac{a_1-a_2-b_1+b_2}{m} = \frac{a_1-a_2}{m} - \frac{b_1-b_2}{m},\end{aligned}$$

also ist diese Zahl diesmal Differenz zweier Zahlen, die nach Voraussetzung ganze Zahlen sind, und somit selbst eine ganze Zahl. Damit ist die zweite Behauptung bewiesen.

zu (3): Wir müssen zuletzt zeigen, dass $a_1 b_1 \equiv a_2 b_2 \mod m$ gilt. Dies ist etwas schwieriger als in den ersten beiden Fällen. Das ist äquivalent dazu, dass $\frac{a_1 b_1 - a_2 b_2}{m}$ eine ganze Zahl ist. Auch hier formen wir diesen Ausdruck um; hierbei ist allerdings in kleiner „Trick“ notwendig. Wir subtrahieren im Zähler die Zahl $a_2 b_1$ und addieren diese gleich wieder. Dies ändert natürlich nicht den Wert des Zählers und mag zuerst überflüssig anmuten. Allerdings erkennen wir dann, dass wir aus den ersten beiden Summanden b_1 und aus den letzten beiden Summanden a_2 ausklammern können und dabei das Ergebnis wieder die Zahlen aus der Voraussetzung enthält:

$$\begin{aligned}\frac{a_1 b_1 - a_2 b_2}{m} &= \frac{a_1 b_1 - a_2 b_1 + a_2 b_1 - a_2 b_2}{m} \\ &= \frac{(a_1-a_2)b_1 + a_2(b_1-b_2)}{m} = \frac{a_1-a_2}{m} \cdot b_1 + a_2 \cdot \frac{b_1-b_2}{m}.\end{aligned}$$

Diese Zahl ist nun Summe zweier Produkte von ganzen Zahlen, also wieder eine ganze Zahl. Damit ist auch die letzte Behauptung gezeigt.

□

Wir können also, um eine Rechnung zu machen, die nur modulo m korrekt sein muss, und nur Addition, Subtraktion und/oder Multiplikation enthält, die Zahlen nach Belieben durch kongruente Zahlen ersetzen, ohne die Richtigkeit des Ergebnisses zu ändern. Mit Division funktioniert das allerdings nicht, wie im nächsten Beispiel verdeutlicht wird.

Beispiel. Es gilt: $2 \equiv 2 \mod 2$ und $2 \equiv 4 \mod 2$. Aber es ergibt keinen Sinn, danach zu fragen, ob $\frac{2}{2} \overset{?}{\equiv} \frac{2}{4} \mod 2$ gilt, da $\frac{2}{4} = \frac{1}{2}$ keine ganze Zahl ist, und Kongruenzen nur für ganze Zahlen überhaupt definiert sind. Aber auch, wenn wir die umgekehrten Quotienten bilden, also uns fragen, ob $\frac{2}{2} \overset{?}{\equiv} \frac{4}{2} \mod 2$ gilt, so ist das zwar eine Aussage, diese ist allerdings falsch, da $1 \not\equiv 2 \mod 2$ gilt.

Wir wollen uns einige erste Beispiele dafür anschauen, wie modulare Arithmetik funktionieren kann.

Beispiel. Wir wollen eine kleine Zahl bestimmen, die kongruent zu 4^{30} modulo 3 ist. Ein wichtiger Aspekt hiervon ist, dass wir die Zahl 4^{30} gar nicht explizit als Dezimalzahl zu kennen brauchen, denn diese anzugeben, ist schon nicht ganz einfach (diese Zahl ist z.B. größer als die maximale darstellbare Zahl im Java-Datentyp „int“). Allerdings ist 4^{30} ein Produkt, das 30 mal den Faktor 4 hat, und unsere Rechenregeln besagen, dass wir im Produkt die Zahl durch eine kongruente Zahl modulo 3 ersetzen können, ohne das Ergebnis modulo 3 zu verändern. Also nutzen wir $4 \equiv 1 \mod 3$ und erhalten

$$4^{30} \equiv 1^{30} \equiv 1 \mod 3,$$

wobei 1^{30} zu bestimmen jetzt eine leichte Rechnung war.

Auch für 2^{30} können wir eine kleine Zahl finden, die kongruent zu 2^{30} modulo 3 ist. Hier ist es von Vorteil, obwohl unsere Aufgabe nur positive Zahlen benutzt, eine negative Zahl für die Zwischenrechnung zu nutzen. Wir bemerken, dass $2 \equiv (-1) \mod 3$ und die Potenzen von (-1) sind wiederum einfach bestimmt: $(-1)^n$ ist 1, falls n gerade ist, und -1, falls n ungerade ist. Insbesondere ist also

$$2^{30} \equiv (-1)^{30} \equiv 1 \mod 3.$$

Dass wir immer eine „kleine“ Zahl finden können, die zu unserer vorgegebener Zahl kongruent ist modulo einer festen Zahl m, ist eine Konsequenz dessen, dass wir mit Rest teilen können. Wir werden nun die dazugehörigen Resultat präziser formulieren. Dahinter steckt allerdings dieselbe Division mit Rest, mit der man schon seit der Grundschule vertraut ist.

Satz 9.4 (Division mit Rest)**.**

Version 1: Seien a, b natürliche Zahlen und $b \neq 0$. Dann gibt es eindeutig bestimmte natürliche Zahlen $q, r \in \mathbb{N}$ mit $0 \leq r < b$ und

$$a = q \cdot b + r.$$

Version 2: Seien a, b ganze Zahlen und $b \neq 0$. Dann gibt es eindeutig bestimmte ganze Zahlen $q, r \in \mathbb{Z}$ mit $0 \leq r < |b|$ und

$$a = q \cdot b + r.$$

Beispiel.

- Für natürliche Zahlen ist es die altbekannte Division mit Rest, die wir durchführen müssen. Sind etwa $a = 17$ und $b = 4$, so ist $17 = 4 \cdot 4 + 1$, also $q = 4$ und $r = 1$.
- Für negative Zahlen ist es ein wenig anders, weil wir auch hier eine nicht-negative Zahl als Rest erhalten wollen. So würde die Darstellung $-17 = (-4) \cdot 4 - 1$ nicht den Anforderungen für $a = -17$ und $b = 4$ genügen, da -1 negativ ist. Hingegen sind $q = -5$ und $r = 3$ die Zahlen, die sich im obigen Satz für diesen Fall ergeben: Es ist

  $$-17 = (-5) \cdot 4 + 3$$

  und $0 \leq 3 < 4$.
- Führt man Division mit Rest für a geteilt durch b aus und hat man den Fall vorliegen, dass a durch b teilbar ist, so ist der Rest in diesem Fall 0. Etwa für $a = 20$ und $b = 4$ ist $q = 5$ und $r = 0$, also

  $$20 = 5 \cdot 4 + 0.$$

Als Konsequenz dessen, dass Division mit Rest möglich ist, erhalten wir das folgende Korollar.

Korollar 9.5. *Sei m eine feste natürliche Zahl, $m \neq 0$. Dann gibt es zu jeder ganzen Zahl $a \in \mathbb{Z}$ eine natürliche Zahl $r \in \mathbb{N}_0$ mit $0 \leq r \leq m - 1$, sodass*

$$a \equiv r \mod m$$

Mit anderen Worten ist jede ganze Zahl kongruent modulo m zu einer, die zwischen 0 und $m - 1$ (einschließlich) ist, also zu einer der Zahlen $0, 1, 2, \ldots, m - 1$.

Bemerkung. Durch Division mit Rest sehen wir auch, dass die im Korollar erwähnte Zahl r für jedes $a \in \mathbb{Z}$ eindeutig ist.

Bereits diese zahlentheoretische Grundlagen werden in der Praxis verwendet. Ein Beispiel dafür ist ISBN.

Beispiel. Die ISBN-13 ist eine Zahl, die jedes (neuere) Buch eindeutig identifiziert. (Diese hat die frühere Version, ISBN-10, abgelöst.) Die letzte Ziffer der ISBN ist eine Prüfziffer: Diese lässt sich aus den anderen Ziffern bestimmen und trägt keine eigene Information, sondern dient der Überprüfung der Nummer. Stimmt die als Rechenergebnis bestimmte Prüfziffer nicht mit

der Prüfziffer der Nummer überein, so ist die ISBN nicht korrekt. (Andersrum darf allerdings nicht geschlossen werden: Zwar „merkt“ die Prüfziffer der ISBN in jedem Fall, wenn genau eine Ziffer ersetzt wurde, und auch manche Zahlendreher, aber die Prüfziffer bleibt gleich, wenn etwa die Ziffern 5 und 0 vertauscht werden.)

Die Prüfziffer wird wie folgt berechnet:

- Zunächst werden die ersten 12 Ziffern der ISBN abwechselnd mit 1 und 3 multipliziert und die so entstandenen Produkte summiert.
- Im nächsten Schritt bestimmt man den Rest r dieser Summe bei der Division durch 10.
- Ist $r \neq 0$, so nimmt man $10 - r$ als Prüfziffer. Im Fall $r = 0$ ist 0 die errechnete Prüfziffer.

Wir betrachten ein Buch mit der ISBN $978 - 0 - 387 - 95385 - 4$. Wir prüfen, ob die Prüfziffer 4 tatsächlich das Ergebnis der oben beschriebenen Rechnung ist. Hierbei machen wir uns die Rechenregeln der modularen Arithmetik zunutze - will man den Rest bei Division 10 berechnen, so kann man jede in der Rechnung (die nur Multiplikation und Addition enthält) vorkommende Zahl durch eine dazu modulo 10 kongruente Zahl ersetzen, ohne das Ergebnis (ebenfalls $\mod 10$ betrachtet) zu ändern. Dabei ist die Beobachtung wichtig, dass jede natürliche Ziffer kongruent modulo 10 zu ihrer letzten Ziffer ist. (Dies gilt nur in der Dezimaldarstellung, über die wir später noch ausführlicher sprechen werden.) Wir gruppieren also die Summanden (unter Ausnutzung der Assoziativität und der Kommutativität der Addition) so, wie es uns am einfachsten zu rechnen erscheint, und ersetzen diese stets durch ihre letzte Ziffer. Dabei nutzen wir erneut, dass jede natürliche Zahl kongruent modulo 10 zu ihrer letzten Ziffer in der Dezimaldarstellung ist. Wir führen nun die Rechnung durch:

$$\begin{aligned}
&1 \cdot 9 + 3 \cdot 7 + 1 \cdot 8 + 3 \cdot 0 + 1 \cdot 3 + 3 \cdot 8 + 1 \cdot 7 \\
&+3 \cdot 9 + 1 \cdot 5 + 3 \cdot 3 + 1 \cdot 8 + 3 \cdot 5 \\
&= 30 + 32 + 10 + 30 + 35 + 20 + 9 \\
&\equiv 0 + 2 + 0 + 0 + 5 + 0 + 9 \mod 10 \\
&\equiv 16 \equiv 6 \mod 10
\end{aligned}$$

Also ist die errechnete Prüfziffer gegeben durch $10 - 6 = 4$, und stimmt somit mit der angegebenen Prüfziffer überein.

Als nächstes wollen wir ein Beispiel dafür sehen, wie durch modulare Arithmetik Aussagen über Teilbarkeit von natürlichen (oder ganzen) Zahlen bewiesen werden können. Es sei angemerkt, dass diese Proposition auch anders, etwa durch vollständige Induktion, bewiesen werden kann.

Proposition 9.6. *Sei n eine natürliche Zahl. Dann ist die Zahl $n^3 - n$ durch 3 teilbar.*

Beweis. Nach Korollar 9.5 gibt es zu der natürlichen Zahl n eine eindeutige natürliche Zahl r mit $0 \leq r \leq 2$ und $n \equiv r \mod 3$. Dank der Rechenregeln der modularen Arithmetik gilt auch $n^3 - n \equiv r^3 - r \mod 3$. Nun kann die Zahl r nur die Werte $0, 1, 2$ annehmen, da dies alle natürliche Zahlen im Intervall $[0, 2]$ sind. Bei diesen Zahlen können wir jeweils überprüfen, ob $r^3 - r$ durch 3 teilbar ist. Es ergibt sich:

$$n^3 - n \equiv r^3 - r \equiv \begin{cases} 0^3 - 0 \equiv 0 \mod 3, \text{ falls } r = 0, \\ 1^3 - 1 \equiv 0 \mod 3, \text{ falls } r = 1, \\ 2^3 - 2 \equiv 6 \equiv 0 \mod 3, \text{ falls } r = 2. \end{cases}$$

Folglich ist die Zahl $n^3 - n$ für jeden möglichen Wert von n kongruent zu 0 modulo 3, also ist sie stets durch 3 teilbar. □

Im Folgenden wollen wir ein weiteres Beispiel für den Einsatz der modularen Arithmetik betrachten.

Beispiel. Wir wollen untersuchen, ob die Zahl

$$1\underbrace{00\ldots00}_{2014}1$$

das Quadrat einer natürlichen Zahl ist. (Wir können diese Zahl auch als $10^{2015} + 1$ schreiben.) Man nennt eine natürliche Zahl, die das Quadrat einer natürlichen Zahl ist, eine *Quadratzahl*.

Es gibt unterschiedliche Möglichkeiten, wie man an dieses Problem herangehen kann. Wir beschreiben einen Ansatz, der modulare Arithmetik benutzt. Zunächst bemerken wir, dass all die Zahlen $101, 1001, 10001, 100001$ keine Quadratzahlen sind. Das führt uns zu der Vermutung (die wir noch zu beweisen haben), dass die Zahl $10^k + 1$ für keine natürliche Zahl $k \geq 2$ eine Quadratzahl sein kann. Insbesondere würde es zu einer negativen Antwort auf die eingangs gestellte Frage führen. Um diese Aussage zu beweisen, werden wir ein Hilfsresultat benötigen.

Lemma 9.7. *Eine Quadratzahl ist kongruent zu 0 oder 1 modulo 3.*

Formuliert man die Aussage des Lemmas um, so heißt das, dass eine natürliche Zahl, die kongruent zu 2 modulo 3 ist, niemals eine Quadratzahl sein kann. Wir wollen nun zunächst das Lemma beweisen.

Beweis. Wir verfahren genau wie im Beweis der letzten Proposition. Wir betrachten die Quadratzahl n^2, also das Quadrat der natürlichen Zahl n. Auch hier gibt es nach Korollar 9.5 zu der natürlichen Zahl n eine eindeutige

natürliche Zahl r mit $0 \leq r \leq 2$ und $n \equiv r \mod 3$. Wir erhalten erneut $n^2 \equiv r^2 \mod 3$ und $r \in \{0, 1, 2\}$ und müssen also nur diese drei Fälle untersuchen:

$$n^2 \equiv r^2 \equiv \begin{cases} 0^2 \equiv 0 \mod 3, \text{ falls } r = 0, \\ 1^2 \equiv 1 \mod 3, \text{ falls } r = 1, \\ 2^2 \equiv 4 \equiv 1 \mod 3, \text{ falls } r = 2. \end{cases}$$

Das zeigt nun, dass Quadrate natürlicher Zahlen nie den Rest 2 bei der Division durch 3 haben. □

Nun können wir zu dem obigen Beispiel zurückkehren.

Beispiel (Fortsetzung)**.** Wir wollen zeigen, dass $10^k + 1$ für alle natürlichen Zahlen $k \geq 2$ keine Quadratzahl ist. Da $10 \equiv 1 \mod 3$, folgt $10^k + 1 \equiv 1^k + 1 \equiv 2 \mod 3$. Nun folgt nach Lemma 9.7 unmittelbar, dass $10^k + 1$ keine Quadratzahl sein kann.

Als nächstes wollen wir die modulare Arithmetik benutzen, um zu erklären, warum die Teilbarkeitskriterien für Teilbarkeit durch 3 und durch 9 funktionieren. Zunächst formulieren wir eine formale Version der Dezimaldarstellung natürlicher Zahlen.

Satz 9.8. *Jede natürliche Zahl lässt sich eindeutig in der Form* $\sum_{i=0}^{m} 10^i \cdot b_i$ *schreiben, wobei* $b_m \neq 0$*, falls* $m \neq 0$ *ist, alle* b_i *natürliche Zahlen sind und für alle* $0 \leq i \leq m$ *gilt:* $0 \leq b_i \leq 9$.

Die Intuition hinter diesem Satz ist genau die Funktionsweise des Stellenwertssystems, die wir bereits aus der Grundschule kennen: Schreibt man eine Zahl, etwa 987123, so steht diese Notation dafür, dass in der „Hunderttausenderstelle“ die Ziffer 9 steht, man also *neun*mal hunderttausend gezählt hat, und dazu *acht*mal zehntausend, dann *sieben*mal tausend, *ein*mal hundert, *zwei*mal zehn und schießlich noch drei Objekte übriggeblieben sind. Es ist also

$$\begin{aligned} 987123 &= 9 \cdot 100000 + 8 \cdot 10000 + 7 \cdot 1000 + 1 \cdot 100 + 2 \cdot 10 + 3 \cdot 1 \\ &= 9 \cdot 10^5 + 8 \cdot 10^4 + 7 \cdot 10^3 + 1 \cdot 10^2 + 2 \cdot 10^1 + 3 \cdot 10^0. \end{aligned}$$

Das ist genau die Darstellung, die in dem obigen Satz erwähnt ist, und wird Dezimaldarstellung genannt. Die Zahlen b_i aus dem Satz sind genau die Ziffern der Dezimaldarstellung; in diesem Fall ist $m = 6$ (denn die Zahl 987123 ist 6-stellig) und $b_6 = 9$, $b_5 = 8$, $b_4 = 7$, $b_3 = 1$, $b_2 = 2$ und $b_3 = 6$. Die Forderung $0 \leq b_i \leq 9$ bedeutet gerade, dass die Zahlen b_i einstellig sind, also jeweils eine Ziffer.

Bereits in der Schule lernt man andere Stellenwertsysteme kennen. Besonders wichtig in der Praxis sind die Binärdarstellung, die Oktaldarstellung und die Hexadezimaldarstellung einer Zahl. Die mathematische Grundlage dafür wird durch den folgenden Satz geliefert, der dem vorherigen sehr ähnelt.

Satz 9.9. *Sei $k \geq 2$ eine natürliche Zahl. Jede natürliche Zahl lässt sich eindeutig in der Form $\sum_{i=0}^{m} k^i \cdot b_i$ schreiben, wobei $b_m \neq 0$, falls $m \neq 0$ ist, alle b_i natürliche Zahlen sind und für alle $0 \leq i \leq m$ gilt: $0 \leq b_i \leq k-1$.*

In diesem Fall gibt es also k unterschiedliche Ziffern, wie auch in erwähnten Darstellungen bekannt: In der Binärdarstellung werden nur zwei unterschiedliche Ziffern, nämlich 0 und 1 verwendet; bei Oktalzahlen 0,1,2,3,4,5,6,7; und bei Hexadezimalzahlen hat man 16 Ziffern zur Verfügung. Da man in unserem Dezimalsystem üblicherweise nur 10 Ziffern zur Verfügung hat, verwendet man für die letzten 6 stattdessen als Symbole die Großbuchstaben A, B, C, D, E, F. Die Zahl FC etwa entspricht $15 \cdot 16 + 12 = 252$ im Dezimalsystem.

Bevor wir nun die Teilbarkeitskriterien beweisen, beobachten wir noch, dass diese nicht (unbedingt) in anderen Stellenwertsystemen gelten. So ist etwa 1101_2 die Binärdarstellung der Zahl 13, die nicht durch 3 teilbar ist, allerdings ist die Binärquersumme $1 + 1 + 0 + 1 = 3$, also durch 3 teilbar.

Wir erinnern nochmal an die Definition der Quersumme.

Definition 9.10. Ist $\sum_{i=0}^{m} 10^i \cdot b_i$ die Dezimaldarstellung einer natürlichen Zahl, so ist die **Quersumme** dieser Zahl gegeben durch $\sum_{i=0}^{m} b_i$. Mit anderen Worten ist die Quersumme einer natürlichen Zahl die Summe ihrer Ziffern (in der Dezimaldarstellung).

Nun beweisen wir das aus der Schule bekannte Teilbarkeitskriterium. Dabei zeigen wir sogar eine stärkere Aussage.

Proposition 9.11. *1. Jede natürliche Zahl ist kongruent modulo 3 zu ihrer Quersumme.*

2. Jede natürliche Zahl ist kongruent modulo 9 zu ihrer Quersumme.

Beweis. Wir führen den Beweis nur für den ersten Teil; der Beweis des zweiten Teils funktioniert analog.

Sei $\sum_{i=0}^{m} 10^i \cdot b_i$ die Dezimaldarstellung einer natürlichen Zahl. Wir bemerken, dass $10 \equiv 1 \mod 3$ gilt. Folglich gilt auch

$$\sum_{i=0}^{m} 10^i \cdot b_i \equiv \sum_{i=0}^{m} 1^i \cdot b_i \equiv \sum_{i=0}^{m} b_i \mod 3.$$

Letzter Term ist aber gerade die Quersumme der vorgegebenen Zahl. Damit ist die Aussage bewiesen. □

Als nächstes sehen wir ein weiteres Beispiel für die Funktionsweise modularen Arithmetik.

Beispiel. In diesem Fall fragen wir uns, was die letzte Ziffer der Zahl 2009^{2009} ist. Wir erinnern uns, dass die letzte Ziffer einer natürlichen Zahl genau der Rest dieser natürlichen Zahl bei der Division durch 10 ist, also insbesondere eine natürliche Zahl zwischen 0 und 9 (einschließlich), die kongruent zu der ursprünglichen modulo 10 ist.

Wir müssen also eine Zahl $0 \leq r \leq 9$ finden, sodass $r \equiv 2009^{2009} \mod 10$ gilt. Zunächst nutzen wir, dass $2009 \equiv 9 \mod 10$ gilt und somit auch

$$2009^{2009} \equiv 9^{2009} \mod 10.$$

Diese Zahl ist zwar deutlich kleiner als die vorherige, jedoch immer noch zu groß für direkte Berechnungen. Stattdessen betrachten wir die letzte Ziffer der Potenzen 9^k für natürliche Zahlen $k \geq 1$:

$$\begin{aligned} 9^1 &= 9 \equiv 9 \mod 10 \\ 9^2 &= 81 \equiv 1 \mod 10 \\ 9^3 &= 729 \equiv 9 \mod 10 \\ 9^4 &\equiv 9^3 \cdot 9 \equiv 9 \cdot 9 \equiv 1 \mod 10 \end{aligned}$$

Man kommt zu der (korrekten) Vermutung, dass die Potenzen von 9 abwechselnd auf 1 und 9 enden. Das formulieren wir nochmal in dem folgenden Lemma.

Lemma 9.12. *Sei $k \geq 1$ eine natürliche Zahl, dann gilt:*

$$9^k \equiv \begin{cases} 1 \mod 10, & \textit{falls } k \textit{ gerade,} \\ 9 \mod 10, & \textit{falls } k \textit{ ungerade.} \end{cases}$$

Beweis. Da $9 \equiv -1 \mod 10$ ist, folgt die Aussage aus der Tatsache, dass

$$(-1)^k = \begin{cases} 1, \text{ falls } k \text{ gerade,} \\ -1, \text{ falls } k \text{ ungerade.} \end{cases}$$

□

Beispiel (Fortsetzung)**.** Nutzt man nun das obige Lemma und die Tatsache, dass 2009 ungerade ist, so folgt $9^{2009} \equiv 9 \mod 10$. Also endet die Zahl 2009^{2009} auf die Ziffer 9.

Um die letzte Ziffer der Zahl 2014^{2014} zu bestimmen, brauchen wir das folgende Lemma:

Lemma 9.13. *Sei $k \geq 1$ eine natürliche Zahl, dann gilt:*

$$4^k \equiv \begin{cases} 6 \mod 10, & \textit{falls } k \textit{ gerade,} \\ 4 \mod 10, & \textit{falls } k \textit{ ungerade.} \end{cases}$$

Der Beweis des Lemmas (z.B. mittels vollständiger Induktion) ist nicht schwer und wir verzichten an dieser Stelle darauf, ihn aufzuschreiben.

Beispiel. Die letzte Ziffer der Zahl 2014^{2014} ist nach dem vorherigen Lemma 6, da

$$2014^{2014} \equiv 4^{2014} \equiv 6 \mod 10$$

ist.

10 Euklidischer Algorithmus I

Unser Ziel in diesem Kapitel ist es, den euklidischen Algorithmus kennenzulernen. Dieser kann für unterschiedliche Problemstellungen verwendet werden. Wir beginnen mit drei motivierenden Beispielen.

Beispiel. Wir suchen alle ganzen Zahlen k, die Lösungen der modularen Gleichung $2k \equiv 4 \mod 6$ sind. Hätten wir die Gleichung $2x = 4$ für reelle Zahlen zu lösen, so würden wir beide Seiten durch 2 teilen und würden die einzige Lösung $x = 2$ erhalten. Bei den modularen Gleichungen ist es allerdings anders; insbesondere ist im Allgemeinen, wie wir auch bereits gesehen haben, Division schlecht mit Kongruenzen verträglich. Wir bemerken, dass auch die Zahl 5 eine Lösung der modularen Gleichung ist, da $2 \cdot 5 = 10 \equiv 4 \mod 6$. Probiert man weiter, so stellt man auch fest, dass auch 8 und 11 Lösungen der modularen Gleichung sind und kommt zu der Vermutung, dass „jede dritte Zahl“ eine Lösung dieser Gleichung ist. Wir wollen dies beweisen. Zunächst benutzen wir erneut das Korollar 9.5, wonach es zu jeder ganzen Zahl k eine eindeutige natürliche Zahl $0 \leq r \leq 5$ gibt, sodass $k \equiv r$ gilt. Es kommt nur auf diese Zahl r an bei der Frage, ob k der Gleichung genügt. Für die Zahlen 0, 1, 2, 3, 4, 5 ist die Frage schnell geklärt:

$$\begin{aligned} 2 \cdot 0 &\equiv 0 \not\equiv 4 \mod 6 \\ 2 \cdot 1 &\equiv 2 \not\equiv 4 \mod 6 \\ 2 \cdot 2 &\equiv 4 \mod 6 \\ 2 \cdot 3 &\equiv 0 \not\equiv 4 \mod 6 \\ 2 \cdot 4 &\equiv 2 \not\equiv 4 \mod 6 \\ 2 \cdot 5 &\equiv 4 \mod 6 \end{aligned}$$

Wir wissen, dass Mod_6 eine Äquivalenzrelation auf $\mathbb{Z}$ ist. Ferner haben wir gesehen, dass dann $\mathbb{Z}$ sich als disjunkte Vereinigung mancher Äquivalenzklassen schreiben lässt. In diesem Fall gilt

$$\mathbb{Z} = [0]_6 \cup [1]_6 \cup [2]_6 \cup [3]_6 \cup [4]_6 \cup [5]_6.$$

Die Lösungsmenge $\mathbb{L}$ der modularen Gleichung besteht aus allen ganzen Zahlen, die kongruent zu 2 oder zu 5 modulo 6 ist. Schreiben wir das nochmal etwas expliziter auf, so erhalten wir:

$$\begin{aligned} \mathbb{L} &\overset{\text{Def}}{=} \{k \in \mathbb{Z} \mid 2k \equiv 4 \mod 6\} \\ &= [2]_6 \cup [5]_6 \\ &= \{6s + 2 \mid s \in \mathbb{Z}\} \cup \{6s + 5 \mid s \in \mathbb{Z}\} \\ &= \{3l + 2 \mid l \in \mathbb{Z}\}. \end{aligned}$$

Insbesondere bemerken wir, dass, ganz im Gegensatz zu der ähnlichen Gleichung für reelle Zahlen, in diesem Fall die Lösungsmenge unendlich ist.

Beispiel. Das zweite Beispiel ist ganz ähnlich zu dem ersten und wird etwas knapper behandelt. Diesmal suchen wir nach allen Lösungen der modularen Gleichung $4x \equiv 1 \mod 11$. Würde man die Gleichung $4x = 1$ über den reellen Zahlen lösen, so würde man durch 4 teilen, was für Gleichungen in reellen Zahlen eine Äquivalenzumformung ist, und erhalten, dass $x = \frac{1}{4}$ die eindeutige Lösung der Gleichung über den reellen Zahlen ist. Da dies keine ganze Zahl ist, wäre es verlockend zu schließen, dass die modulare Gleichung keine Lösungen hat. Allerdings stimmt das nicht, und schon nach kurzem Ausprobieren sieht man, dass $x = 3$ eine Lösung der Gleichung ist, denn $4 \cdot 3 = 12 \equiv 1 \mod 11$. Wie im letzten Beispiel auch ist x genau dann eine Lösung der modularen Gleichung, wenn der Rest von Division von x durch 11 eine Lösung der Gleichung ist. Also müssen wir wieder nur die Zahlen zwischen 0 und 10 (einschließlich) überprüfen und sehen:

$$\begin{aligned}
4 \cdot 0 &\equiv 0 \mod 11 \\
4 \cdot 1 &\equiv 4 \mod 11 \\
4 \cdot 2 &\equiv 8 \mod 11 \\
4 \cdot 3 &\equiv 1 \mod 11 \\
4 \cdot 4 &\equiv 5 \mod 11 \\
4 \cdot 5 &\equiv 9 \mod 11 \\
4 \cdot 6 &\equiv 2 \mod 11 \\
4 \cdot 7 &\equiv 6 \mod 11 \\
4 \cdot 8 &\equiv 10 \mod 11 \\
4 \cdot 9 &\equiv 3 \mod 11 \\
4 \cdot 10 &\equiv 7 \mod 11.
\end{aligned}$$

Also sehen wir, dass in diesem Fall die Lösungsmenge wie folgt beschrieben wird: $\{11k + 3 \mid k \in \mathbb{Z}\}$.

Die Methode, die wir hier angewandt haben, ist doch recht aufwändig und wird bei größeren Zahlen schnell unpraktikabel. Der euklidische Algorithmus kann dazu eingesetzt werden, solche Gleichungen viel effektiver zu lösen.

Beispiel. Dieses Beispiel ist recht anders in seiner Natur als die beiden vorherigen.

Wir nehmen an, dass wir einen großen Tank Wasser haben, in dem 300 Liter Wasser aufbewahrt werden. Wir haben einen zweiten Tank, der auch 300 Liter fassen kann, allerdings leer ist. Wir würden gerne genau einen Liter in diesen Tank umfüllen. Im ersten Fall haben wir allerdings nur einen 3-Liter-Behälter und einen 6-Liter-Behälter, die jeweils nur ganz gefüllt und umgefüllt werden können. Man sieht schnell ein, dass man so niemals auf einen Liter kommen kann, denn die Literanzahlen, die man so abmessen kann, sind stets durch 3 teilbar. 3 ist nämlich ein gemeinsamer Teiler von

300, 3 und 6, und alle Zahlen, die man durch Addition und Subtraktion dieser Zahlen erhalten kann, werden stets durch 3 teilbar sein.

Hätte man nun einen 3-Liter-Behälter und einen 5-Liter-Behälter vorgegeben, so hätte man keine Schwierigkeiten, genau einen Liter abzumessen: Man würde erstmal $2 \cdot 5$ Liter in den leeren Tank füllen, und dann $3 \cdot 3$ Liter wieder darausnehmen, um $2 \cdot 5 - 3 \cdot 3 = 1$ Liter zu erhalten.

Der entscheidende Faktor hier ist der größte gemeinsame Teiler der Behältergrößen. Wir wiederholen die Definition des größten gemeinsamen Teilers zweier natürlichen Zahlen, die aus der Schule bekannt sein dürfte.

Definition 10.1. Seien $a, b \in \mathbb{N} \setminus \{0\}$ natürliche Zahlen. Der **größte gemeinsame Teiler** von a und b ist die größte natürliche Zahl d, die sowohl a als auch b teilt. Wir schreiben in diesem Fall $d = \mathrm{ggT}(a, b)$.

Bemerkung. Die natürliche Zahl 1 teilt jede natürliche Zahl. Daher ist 1 ein gemeinsamer Teiler von jedem Paar natürlicher Zahlen und somit ist $\mathrm{ggT}(a, b) \geq 1$ für beliebige natürliche Zahlen a, b.

Definition 10.2. Seien $a, b \in \mathbb{N} \setminus \{0\}$ natürliche Zahlen. Die Zahlen a, b heißen **teilerfremd**, falls $\mathrm{ggT}(a, b) = 1$ gilt.

Manchmal will man aus unterschiedlichen Gründen den größten gemeinsamen Teiler zweier natürlichen Zahlen bestimmen. Bei kleinen Zahlen geht es ganz gut durch ausprobieren. Häufig lernt man in der Schule die folgende Methode: Man zerlegt die beiden Zahlen in Primfaktoren, und fasst alle gemeinsamen Primfaktoren zu dem größten gemeinsamen Teiler zusammen. Diese Methode hat einige Vorteile. Für uns hat sie allerdings zwei Nachteile. Der erste Nachteil ist theoretischer Natur: Warum lässt sich jede Zahl in Primfaktoren zerlegen? Wie beweist man die Existenz und Eindeutigkeit einer solchen Zerlegung? Der zweite Nachteil ist näher an die praktischen Schwierigkeiten: Die bekannten Verfahren zum Bilden von Primfaktorzerlegungen sind sehr langsam. Der euklidische Algorithmus ist hingegen ein schnelles Verfahren, um den größten gemeinsamen Teiler zweier natürlicher Zahlen zu bestimmen.

Wir beschreiben zunächst das Verfahren und machen ein Beispiel dazu. Danach sehen wir einige Überlegungen dazu, warum der Algorithmus funktioniert und wie man ihn anwendet. Am einfachsten ist es, das Verfahren rekursiv zu beschreiben.

Euklidischer Algorithmus. Seien natürliche Zahlen $a \geq b \geq 1$ vorgegeben. Um den größten gemeinsamen Teiler von a, b zu bestimmen, gehe man wie folgt vor:

1. Man teile a durch b mit Rest und erhalte $a = qb + r$ mit $q, r \in \mathbb{N}$ und $0 \leq r \leq b$.

2. Ist $r = 0$, so ist $b = \mathrm{ggT}(a, b)$.

3. Ist $r > 0$, so ist $\mathrm{ggT}(a, b) = \mathrm{ggT}(b, r)$. Wende nun den euklidischen Algorithmus an, um letzteren zu bestimmen.

Es dürfte nicht unmittelbar klar sein, warum das funktioniert - also insbesondere wie die Gleichheit im letzten Schritt zustandekommt oder auch, warum das Verfahren terminiert. Wir schauen uns ein Beispiel an, bevor wir uns diesen Fragen widmen.

Beispiel. Wir wollen $\mathrm{ggT}(144, 159)$ bestimmen. $159 > 144$, also müssen wir als erstes 159 durch 144 mit Rest teilen:

$$159 = 1 \cdot 144 + 15.$$

Da der Rest 15 nicht 0 ist, wenden wir wiederum den euklidischen Algorithmus an, um $\mathrm{ggT}(144, 15)$ zu bestimmen. Die nächste Division mit Rest liefert:

$$144 = 9 \cdot 15 + 9.$$

Als nächstes haben wir $\mathrm{ggT}(15, 9)$ zu bestimmen. (Das wäre jetzt klein genug, um durch Ausprobieren zum Ergebnis zu kommen, aber wir wollen den euklidischen Algorithmus so weit wie möglich ausführen, um die Funktionsweise besser zu verstehen.)

$$15 = 1 \cdot 9 + 6.$$

Weiter mit $\mathrm{ggT}(9, 6)$:

$$9 = 1 \cdot 6 + 3.$$

Will man nun $\mathrm{ggT}(6, 3)$ bestimmen, so stellt man fest, dass 6 ohne Rest durch 3 teilbar, und wie im zweiten Schritt des Algorithmus bereits behauptet wurde, ist $\mathrm{ggT}(6, 3) = 3$. Der euklidische Algorithmus liefert also $\mathrm{ggT}(144, 159) = 3$.

11 Zahlentheorie in der Kryptographie

In diesem Kapitel geben wir einige Beispiele für die Anwendungen der Zahlentheorie in der Kryptographie. Will eine Person, nennen wir sie Alice, einer anderen Person - Bob - eine geheime Nachricht übermittelt, so steht sie im Allgemeinen vor einem Problem. Sie kann versuchen, die Botschaft etwa zu verbergen, eine sichere Leitung zu benutzen o.ä. Das ist allerdings nicht der Inhalt unserer Betrachtung. Wir beschäftigen uns mit dem Fall, dass die Botschaft über einen unsicheren Kanal verschickt wird, und stattdesssen der Inhalt der Botschaft verborgen ist. Alice ersetzt also die eigentliche Botschaft durch eine andere, die für einen zufälligen Beobachter nicht verständlich ist, aus der aber Bob die ursprüngliche Botschaft rekonstruieren kann - sie *verschlüsselt* die Botschaft. Die Lehre von den Verschlüsselungen wird **Kryptographie** genannt. Gleichzeitig ist es auch wichtig, sich Gedanken darüber zu machen, wie verschlüsselte Botschaften geknackt werden können - auch wenn man gar keine fremde Botschaften abfangen will, sind solche Überlegungen, auch **Kryptoanalyse** genannt, wichtig, um die Sicherheit von einem Verschlüsselungsverfahren zu analysieren.

Wir fangen mit einem sehr alten Verschlüsselungsverfahren an, das heute nicht mehr benutzt wird. Trotzdem ist es interessant, da es sowohl besonders einfach zu verstehen ist als auch wichtige Erkenntnisse über Verschlüsselungsverfahren im Allgemeinen liefert.

Caesar-Verschlüsselung

In diesem Fall besteht die Botschaft, die wir übermitteln wollen, aus einem Text. Bei Caesar-Verschlüsselung wird jeder Buchstabe nach einem festen Muster durch einen anderen festen Buchstaben ersetzt. Eine Verschlüsselung, bei der jeder Buchstabe durch einen festen Buchstaben ersetzt wird, heißt *monoalphabetisch*. Die Idee bei Caesar-Verschlüsselung, die, soweit man weiß, tatsächlich zur Zeiten Caesars benutzt wurde, ist die Verschiebung von allen Buchstaben um eine feste Größe. Genauer funktioniert das wie folgt. Es wird eine Verschiebung festgelegt, ein sogenannter *Schlüssel*, den vorher Alice und Bob auf irgendeine Art und Weise geheim geteilt haben müssen. In diesem Fall ist es ganz hilfreich, Buchstaben mit Zahlen zu identifizieren, denn damit können wir häufig einfacher arbeiten.

- Man fasst jeden Buchstaben als eine Zahl zwischen 0 und 25 auf.
- Man addiere zu jedem Buchstabenwert den Schlüsselwert mod 26.
- Man übersetze das Ergebnis zurück in Buchstaben.

Hier ist ein Beispiel für diese Prozedur:

Beispiel. Wir möchten den Satz „Wenn es neblig ist, ist die Sicht schlecht.“ verschlüsseln. Der Schlüssel, auf den sich Alice und Bob geeinigt haben, ist Q. Dies ist der 17-te Buchstabe im Alphabet, also entspricht der Zahl 16 (da wir bei 0 zu zählen anfangen). Wir stellen das Verfahren für die ersten Buchstaben vor:

W	E	N	N	E	S	N	E	B	L	I	G
22	4	13	13	4	18	13	4	1	11	8	6
12	20	3	3	20	8	3	20	17	1	24	22
M	U	D	D	U	I	D	U	R	B	Y	W

Insgesamt ist die verschlüsselte Botschaft nun „Mudd ui durbyw yij, yij tyu Iysxj isxbusxj.“. Erhält Bob nun diese Botschaft, so führt er zum Entschlüsseln genau dieselben Schritte in umgekehrten Reihenfolge durch, und hat wieder den ursprünglichen Text.

Auf den ersten Blick erscheint die verschlüsselte Botschaft recht sicher - keiner kann den Inhalt der Botschaft in dieser Form verstehen, also könnte man auf die Idee kommen, dass die Botschaft sicher ist. Allerdings hat das Verfahren mehrere Schwächen, die es ermöglichen, den Text mit einem gewissen Aufwand auch ohne den Schlüssel zu entschlüsseln. Eine naive Möglichkeit ist es, einfach alle möglichen Schlüssel - davon gibt es ja nur 26 - durchzuprobieren. Diese Möglichkeit besteht im Prinzip, allerdings ist sie langwierig. Die Caesar-Verschlüsselung hat eine viel gravierendere Schwäche.

Ein allgemeines Verfahren, das es ermöglicht, monoalphabetische Verschlüsselungen zu knacken, ist die *Häufigkeitsanalyse*. Sie beruht auf der Tatsache, dass in deutscher Sprache (wie auch in jeder anderen Sprache) die Buchstaben nicht gleich häufig vorkommen. Dieses Phänomen gibt es natürlich auch in anderen Sprachen, allerdings sind die Häufigkeitsverteilungen der unterschiedlichen Sprachen unterschiedlich. In einem längeren Text der deutscher Sprache ist etwa jeder sechste Buchstabe ein E. Danach kommt mit deutlichem Abstand ein N, und dann ungefähr in gleichen Anteilen I, S, R, A. Das sind natürlich nur statistische Verteilungen - das heißt in diesem Fall, dass sie zwar im allgemeinen stimmen, aber es durchaus Ausnahmen geben kann. Vor allem in kurzen Texten kann die Verteilung variieren; man kann eine verzerrte Verteilung natürlich auch künstlich herbeiführen, indem man etwa einen Satz wie „Annas Ananas ist rosa und grau. “ konstruiert.

Im Allgemeinen kann man jedoch bei einem längeren Text der deutschen Sprache davon ausgehen, dass E der häufigste und N der zweithäufigste Buchstabe ist. Hat man also einen Geheimtext vorliegen, von dem man vermutet, dass dieser monoalphabetisch verschlüsselt wurde, so kann man die Häufigkeit der vorkommenden Buchstaben bestimmen und annehmen, dass der häufigste Buchstabe E oder vielleicht N im Klartext ist. Das ist bei Caesar besonders gravierend, weil man, sobald man weiß, mit welchem Buchstaben E verschlüsselt wurde, auch den Schlüssel kennt und die gesamte Botschaft

entschlüsseln kann. Aber auch im Allgemeinen erhält man durch eine feinere Häufigkeitsanalyse begründete Vermutungen darüber, welche Buchstaben sich entsprechen, und insgesamt eine Möglichkeit, den Text zu entschlüsseln.

Beispiel. Hat man etwa den mit Caesar verschlüsselten Geheimtext „ydz hvoczhvodfzm ndiy zdiz vmo amviujnzi: mzyzo hvi up dcizi, nj pzwzmnzouzi ndz zn di dcmz nkmvxcz, piy yvii dno zn vgnwvgy zorvn viyzmzn.“ vorliegen, so stellt man eine Tabelle für die Häufigkeit der häufigsten 6 Buchstaben auf (rechts sind die Entsprechungen im Klartext angegeben):

Geheimtextbuchstabe	Anzahl	Klartextbuchstabe
Z	21	E
I	13	N
V	11	A
N	11	S
D	9	I
M	8	R

Man sieht, dass selbst bei diesem recht kurzen Text die Verteilung ähnlich zu der allgemeinen Verteilung ist. Insgesamt kann man den Text nun leicht entschlüsseln und erhält: „Die Mathematiker sind eine Art Franzosen: Redet man zu ihnen, so übersetzen sie es in ihre Sprache, und dann ist es alsbald etwas anderes.“ (Dies ist ein Zitat von J.W. von Goethe.)

Lange Zeit (vom 16. bis zur Mitte des 19. Jahrhunderts) galt eine verbesserte Variante der Caesar-Verschlüsselung, die Vigenère-Verschlüsselung, als unknackbar. Die Idee hierbei ist, dass jetzt ein einziger Buchstabe durch unterschiedliche Buchstaben repräsentiert werden kann, und diese unterschiedliche Buchstaben werden durch ein festes Muster vorgegeben, das durch eine ganze Schlüsselphrase bestimmt wird. Aber im 19. Jahrhundert wurden Verfahren entwickelt, die unter Einsatz von Rechenmaschinen auch diese Verschlüsselung zu knacken erlaubten.

Diffie-Hellman-Schlüsselübergabe

Wie wir schon bei Caesar-Verschlüsselung gesehen haben, stellt sich bei jeder *symmetrischen* Verschlüsselung, also bei einer, wo Alice und Bob denselben Schlüssel brauchen, das Problem, den Schlüssel sicher zu übergeben. Bis ins 20. Jahrhundert scheint die Kryptographie an dieser Stelle kaum eingesetzt worden zu sein. Hier wollen wir ein Verfahren vorstellen, das zur Schlüsselübergabe über einen unsicheren Kanal benutzt werden kann. Das gewissermaßen erstaunliche dabei ist, dass sowohl Alice als auch Bob am Ende denselben Schlüssel besitzen, obwohl der Schlüssel selbst nie über die Leitung übermittelt worden ist. Verfeinerungen von diesem Verfahren werden auch heute eingesetzt.

Eine wichtige Idee dabei ist, sogenannte Einweg-Funktionen zu verwenden, also bijektive Abbildungen, die „in eine Richtung“ einfach zu berechnen sind, während die Urbildbestimmung sehr kompliziert ist. Das kann man sich wie ein herkömmliches Telefonbuch vorstellen: Kennt man den Namen einer Person, so ist es leicht, die Telefonnummer dieser Person in einem Telefonbuch zu finden. Umgekehrt ist es aber fast unmöglich (oder jedenfalls sehr aufwändig), in dem Telefonbuch den Namen der Person zu finden, die zu einer Telefonnummer gehört. Für die Diffie-Hellman-Schlüsselübergabe liefert die Zahlentheorie ein Beispiel solcher Funktionen.

In unserem Fall beruht die Einweg-Funktion auf der folgenden Beobachtung. Sei eine Primzahl p und eine natürliche Zahl x vorgegeben. Es ist einfach, für ein a die Zahl $x^a \mod p$ zu bestimmen. (Insbesondere gibt es dafür schnelle Algorithmen.) Allerdings ist es schwierig, aus $x^a \mod p$ die Zahl a zu bestimmen; dafür sind keine schnellen Algorithmen bekannt. (Dieses schwierige Problem trägt den Namen „Diskreter-Logarithmus-Problem“.)

Darauf basierend, kann man nun den Schlüssel, der jetzt eine natürliche Zahl (mit gewissen Einschränkungen) sein wird, wie folgt übergeben:

- Alice und Bob legen eine Primzahl p und eine (spezielle) Zahl $2 \leq x \leq p-2$ fest und übermitteln diese.
- Alice hat eine geheime Zahl a, Bob hat eine geheime Zahl b.
- Alice sendet $x^a \mod p$ an Bob, Bob sendet $x^b \mod p$ an Alice.
- Alice berechnet $(x^b)^a = x^{ab} \mod p$. Das ist jetzt der gemeinsame Schlüssel.
- Bob berechnet $(x^a)^b = x^{ab} \mod p$.

Die Zahl $x^{ab} \mod p$ selbst wird dabei nie übermittelt. Mit dieser Notation meinen wir hier die natürliche Zahl zwischen 0 und $p-1$ (einschließlich), die kongruent zu x^{ab} modulo p ist.

Als Beobachter ist es nun selbst mit der Kenntnis von $p, x, x^a \mod p, x^b \mod p$ für große Primzahlen p sehr schwierig, den Schlüssel zu berechnen. Allerdings ist dieses Schlüsselaustauschverfahren anfällig für den sogenannten „Man-in-the-middle“-Angriff, also für jemanden, der Nachrichten zwischen Alice und Bob nicht nur abfangen kann, sondern auch Nachrichten verschicken in ihrem Namen verschicken kann. Dieser kann sich seine eigene Zahl m ausdenken und $x^m \mod p$ an Bob und Alice anstelle von den jeweiligen Botschaften verschicken. Damit haben Alice und Angreifer den gemeinsamen Schlüssel $x^{am} \mod p$ und Bob und der Angreifer den gemeinsamen Schlüssel $x^{bm} \mod p$. Weder Alice noch Bob merken, dass alle Kommunikation vom Angreifer ent- und verschlüsselt (und dazwischen möglicherweise verändert) wird.

Wir zeigen an einem Beispiel, wie das grundlegende Verfahren funktioniert. Man bemerke, dass es hier durch Ausprobieren leicht möglich ist, die diskreten Logarithmen zu berechnen; allerdings sind die Primzahlen, die in ähnlichen Verfahren verwendet werden, etwa hundertstellig und liegen somit in ganz anderen Größenordnungen.

Beispiel. Alice und Bob legen die Primzahl $p = 19$ und $x = 2$ fest. Alice hat $a = 5$ gewählt, Bob setzt $b = 13$. Alice übermittelt $2^5 \equiv 13 \mod 19$ an Bob. Bob übermittelt an Alice die Zahl

$$2^{13} \equiv (2^4)^3 \cdot 2 \equiv (-3)^3 \cdot 2 \equiv (-8) \cdot 2 \equiv 3 \mod 19.$$

Alice errechnet den Schlüssel $3^5 \equiv 27 \cdot 9 \equiv 8 \cdot 9 \equiv 72 - 57 \equiv 15 \mod 19$. Bob errechnet den Schlüssel

$$13^{13} \equiv (-6)^{13} \equiv 36^6 \cdot (-6) \equiv (-2)^6 \cdot (-6) \equiv 15 \mod 19.$$

In diesem Fall haben also nochmal gesehen, dass tatsächlich dieselbe Zahl als Schlüssel errechnet wird.

RSA-Verschlüsselung

Die Grundidee des Diffie-Hellman-Algorithmus war die Anwendung einer Einweg-Funktion. Dieser Gedanke erlaubte auch etwas, was auf den ersten Blick unmöglich erscheint: *Asymmetrische* Verschlüsselungsverfahren, also solche, wo der Schlüssel zum Verschlüsseln ein anderer als der Schlüssel zum Entschlüsseln ist. Das ist die Grundidee der RSA-Verschlüsselung, deren Variationen heutzutage weit verbreitet sind. Hier hat jeder Zugriff auf den Schlüssel zum Verschlüsseln, den sogenannten *Public Key*. Schickt Alice eine mit Bobs Public Key verschlüsselte Botschaft an Bob, so kann Bob diese mit seinem privaten und geheimen Private Key entschlüsseln. Diesen Private Key teilt er niemanden mit. Nur mit dem Public Key ist die Entschlüsselung der Botschaft (fast) unmöglich.

In diesem Fall ist die Einweg-Funktion, die benutzt wird, Multiplikation von Primzahlen. Für Multiplikation von großen Zahlen gibt es sehr schnelle Algorithmen. Es ist also einfach, zwei sehr große Primzahlen zu multiplizieren, während es sehr (zeit-)aufwändig ist, aus dem Produkt die beiden Primzahlen abzulesen.

Wir erläutern nun das grundlegende Verfahren und demonstrieren die Funktionsweise an einem Beispiel. Wir zeigen außerdem einige mathematische Aussagen auf, die in die Funktionsweise dieses Verfahrens eingebaut sind. In diesem Fall muss die Botschaft eine Zahl sein. Hat man eine Botschaft, die aus einem Text besteht, so muss man sich vorher eine sinnvolle Möglichkeit überlegen, diesen in eine oder mehrere Zahlen umzuwandeln. Darauf wollen wir hier jedoch nicht näher eingehen. Nun zum eigentlichen Verschlüsselungsverfahren.

Zunächst legt man zwei Primzahlen p und q fest. In der Praxis sind das sehr große Zahlen, an die gewisse Voraussetzungen gestellt werden, damit das Verfahren sicher(er) ist. Für die Demonstration der grundsätzlichen Funktionsweise sollen uns aber $p = 37$ und $q = 47$ genügen. Man bestimmt das Produkt $n = pq$; in unserem Fall ist $n = 1739$. Ferner bestimmt man die Zahl $m = (p-1)(q-1)$, bei uns ist $m = 1656$. Die Zahl n wird öffentlich verfügbar sein; da man aber p und q geheimhält, ist die Zahl m sehr schwer aus n zu bestimmen. Die Zahl m wird ebenfalls geheimgehalten. Nun legt man einen öffentlichen Schlüssel (Public Key) e fest, der teilerfremd zu m sein muss. Wir setzen $e = 29$. Die Zahlen n und e sind also öffentlich verfügbar. Nun berechnet man den privaten Schlüssel d als Lösung der modularen Gleichung $ed \equiv 1 \mod m$ im Bereich 0 bis $m-1$. Das geht schnell mit dem *euklidischen Algorithmus*, allerdings nur, wenn m bekannt ist. Will also jemand aus n und e die Zahl d bestimmen, so steht er vor einer sehr schwierigen Aufgabe. In unserem Fall ist die Lösung der Gleichung $29d \equiv 1 \mod 1656$ im angegebenen Bereich gegeben durch $d = 1085$.

Will man nun eine Botschaft $x = 1345$ übermitteln, so erhält man den Geheimtext mit dem Public Key, indem man $y = x^e \mod n$ berechnet (also diesmal eine Zahl im Bereich 0 bis $n-1$). In unserem Beispiel kann man durch geeignetes Anwenden der Rechenregeln der modularen Arithmetik nachrechnen, dass $1345^{29} \equiv 1206 \mod 1739$ ist. 1206 ist also unsere verschlüsselte Botschaft.

Will nun der Empfänger die Botschaft mit seinem privaten Schlüssel entschlüsseln, so berechnet er $y^d \mod n$. Ist die ursprüngliche Botschaft x im Bereich zwischen 0 und $n-1$, so erhält man als Ergebnis dieser Rechnung genau die ursprüngliche Botschaft. Die zu verschickende Botschaften müssen also unbedingt in diesem Bereich liegen. In diesem Fall kann man nachprüfen, dass $1206^{1085} \equiv 1345 \mod 1739$ gilt.

Wir wollen kurz darauf eingehen, warum man aus der mit dem Public Key nach der beschriebenen Methode verschlüsselten Botschaft y mit dem private Key die ursprüngliche Botschaft wiedergewinnen kann. Dem liegen zwei Sätze zugrunde, die wir soweit noch nicht behandelt haben.

Satz (Kleiner Satz von Fermat)**.** *Sei p eine Primzahl und a eine natürliche Zahl, die zu p teilerfremd ist. Dann ist $a^{p-1} \equiv 1 \mod p$.*

Satz (Chinesischer Restsatz: Spezialfall)**.** *Seien p, q zwei verschiedene Primzahlen und $z \equiv 1 \mod p$ und $z \equiv 1 \mod q$, dann gilt $z \equiv 1 \mod pq$.*

Wir wollen also mit Hilfe dieser Sätze nachprüfen, dass $(x^e)^d = x^{ed} \mod n$ genau die ursprüngliche Nachricht x ergibt. Wir haben aber d gerade so gewählt, dass $ed \equiv 1 \mod m$ gilt, also so, dass es eine natürliche Zahl k gibt, sodass

$$ed - 1 = km = k(p-1)(q-1).$$

Folglich ist

$$x^{ed-1} = x^{k(p-1)(q-1)} = \left(x^{p-1}\right)^{k(q-1)}.$$

Betrachtet man diese Zahl modulo p, so kann man nun den kleinen Satz von Fermat anwenden und erhält

$$\left(x^{p-1}\right)^{k(q-1)} \equiv 1^{k(q-1)} \equiv 1 \mod p.$$

Genauso sieht man, dass $x^{ed-1} \equiv 1 \mod q$. Nun kann man den Spezialfall des Chinesischen Restsatzes anwenden, der oben angeführt ist, und erhält $x^{ed-1} \equiv 1 \mod pq = n$. Also ist

$$x^{ed} = x^{ed-1} \cdot x \equiv 1 \cdot x \equiv x \mod n$$

tatsächlich die ursprüngliche Botschaft.

12 Euklidischer Algorithmus II

Wir wollen nun verstehen, warum der euklidische Algorithmus funktioniert.

Zunächst bemerken wir: Sind $a \geq b \geq 1$ natürliche Zahlen und ist a durch b teilbar, so ist $\mathrm{ggT}(a,b) = b$. Denn wir haben ja vorausgesetzt, dass a durch b teilbar ist, und b ist stets durch b teilbar, also ist b auf jeden Fall ein gemeinsamer Teiler von a und b. Es kann allerdings auch keinen größeren gemeinsamen Teiler von a und b geben, denn b ist der größte Teiler von b: Teilt man b durch eine größere natürliche Zahl, so ist der Quotient eine rationale Zahl strikt zwischen 0 und 1, also keine natürliche Zahl. Das erklärt, warum der Schritt 2 im euklidischen Algorithmus gerechtfertigt ist.

Als nächstes beweisen wir die folgende Proposition, die den Schritt 3 erklärt.

Proposition 12.1. *Seien $a \geq b \geq 1$ natürliche Zahlen. Ist $a = qb + r$ für natürliche Zahlen $q, r \in \mathbb{N}$, $r \neq 0$, so ist* $\mathrm{ggT}(a,b) = \mathrm{ggT}(b,r)$.

Bemerkung.

1. In der Proposition verlangen wir nicht, dass $a = qb + r$ das Ergebnis der Division mit Rest ist. Dafür funktioniert die Proposition insbesondere, aber die Größeneinschränkung für r sind für diese Proposition unnötig.

2. Ist r das Ergebnis der Division mit Rest $a = qb + r$, so kann r auch mit „$a \mod b$“ bezeichnet werden.

Beweis. Der Beweis wird in drei Schritten geführt.

1. Im ersten Schritt zeigen wir, dass jeder - also nicht nur der größte - gemeinsame Teiler d von a und b auch ein Teiler von r ist. Um das zu sehen, stellen wir die Gleichung $a = qb + r$ um und erhalten die äquivalente Gleichung $r = a - qb$. Da b durch d teilbar ist, ist auch qb durch d teilbar. (Zur Erinnerung: Die Teilbarkeitsrechenregeln hatten wir in der Proposition 8.3 zusammengefasst.) Da nun sowohl a als auch qb durch d teilbar ist, ist auch deren Differenz $a - qb = r$ durch d teilbar. Ist also d ein gemeinsamer Teiler von a und b, so ist d notwendigerweise auch ein gemeinsamer Teiler von b und r.

2. Dieser zweite Schritt ist dem ersten sehr ähnlich. Hier zeigen wir, dass jeder gemeinsame Teiler e von b und r auch ein Teiler von a ist. Diesmal folgern wir erneut, dass e auch ein Teiler von qb ist und somit auch von der Summe $qb + r = a$. Ist e also ein gemeinsamer Teiler von b und r, so ist e auch ein gemeinsamer Teiler von a und b.

3. In diesem letzten Schritt führen wir die Aussagen der ersten beiden Schritte zusammen. Daraus sehen wir, dass die folgende Mengengleichheit gilt:

$$\{d \in \mathbb{N} \mid d \text{ teilt } a \text{ und } d \text{ teilt } b\} = \{e \in \mathbb{N} \mid e \text{ teilt } b \text{ und } e \text{ teilt } r\}.$$

Folglich ist auch das größte Element der linken Menge, das nach Definition gerade der größte gemeinsame Teiler von a und b ist, also $\mathrm{ggT}(a,b)$, gleich dem größten Element in der rechten Menge, das nach Definition gerade der größte gemeinsame Teiler von b und r ist, also $\mathrm{ggT}(b,r)$. Somit gilt $\mathrm{ggT}(a,b) = \mathrm{ggT}(b,r)$, was auch zu beweisen war.

□

Diese Proposition klärt also, warum der euklidische Algorithmus stets die richtige Zahl als Ergebnis liefert, sofert er überhaupt eine Zahl ausgibt. Wir wollen nur skizzieren, warum er stets terminiert. Wir behaupten, dass die Summe der beiden in den Algorithmus eingesetzten natürlichen Zahlen in jedem Schritt strikt kleiner wird. Tatsächlich: Startet man mit natürlichen Zahlen a und b und ist man nicht sofort nach dem ersten Schritt fertig, also ist bei der Division mit Rest $a = qb + r$ mit $0 < r < b$, so ist die Summe der beiden als nächstes in den Algorithmus einzusetzenden Zahlen, also von b und r, strikt kleiner als die von a und b, denn wir haben $a \geq b$ angenommen und bei Division mit Rest verlangen wir $b > r$, also erhalten wir insgesamt insgesamt $a+b > b+r$. Ähnlich zeigt man, dass es für jeden Schritt gilt. Würde also der Algorithmus nicht enden, so wäre die Summe der beiden zuletzt eingesetzten natürlichen Zahlen nach spätestens $a + b + 1$ Schritten negativ, denn wir haben mit der Summe $a + b$ angefangen und diese in jedem Schritt um mindestens 1 verkleinert. Da die Summe zweier natürlichen Zahlen nicht negativ sein kann, erhalten wir also aus der Annahme, dass der Algorithmus unendlich weitergeführt wird, einen Widerspruch und somit folgt, dass der Algorithmus stets nach endlich vielen Schritten eine Zahl ausgeben muss; wir haben uns bereits vorher überlegt, dass diese Zahl gerade $\mathrm{ggT}(a,b)$ sein muss.

Wir kehren nun zu einem der motivierenden Beispiele für den euklidischen Algorithmus (in einer leichten Variation) zurück:

Beispiel. Wir nehmen an, dass wir einen extrem großen Tank mit Wasser haben. Insbesondere wird es nicht möglich sein, diesen Tank mit den Behältern, die wir gleich zur Verfügung haben, auszuschöpfen. (Wenn Sie möchten, können Sie sich auch ein Meer anstelle von einem Tank vorstellen.) Wir haben einen zweiten Tank, der auch beliebig viel Wasser fassen kann, allerdings leer ist. Wir würden gerne genau einen Liter (und später: genau d Liter) in diesen leeren Tank umfüllen. Dafür haben wir einen Behälter, der genau a Liter fasst, und einen, der genau b Liter fasst, wobei a und b natürliche Zahlen sind. Wir können einen Behälter jeweils nur ganz füllen. Wir fragen uns, unter welchen Bedingungen an a und b es möglich ist, genau einen Liter abzumessen, und falls es möglich ist, wie wir das zu machen haben.

Wir hatten bereits eingesehen, dass für $a = 3$ und $b = 6$ alle Literanzahlen, die wir abmessen können, durch 3 teilbar sein werden.

Für $a = 3$ und $b = 5$ haben wir bereits geklärt, wie genau ein Liter abgemessen werden kann.

Es ist recht klar, dass wenn $d = \mathrm{ggT}(a, b) > 1$ ist, die Zahlen a und b also durch eine feste Zahl $d > 1$ teilbar sind, auch jede Anzahl, die wir durch Subtraktion und Addition von Vielfachen von a und b erhalten können, durch d teilbar sein muss; für $d > 1$ ist es also unmöglich, genau einen Liter abzumessen. Ist es aber möglich, genau d Liter abzumessen? Die folgende Proposition beantwortet diese Frage positiv.

Proposition 12.2. *Seien $a \geq b \geq 1$ natürliche Zahlen und sei $d = \mathrm{ggT}(a, b)$. Dann gibt es* ganze *Zahlen $x, y \in \mathbb{Z}$, für die*

$$ax + by = d$$

gilt.

Bemerkung.

1. In der obigen Proposition können wir nicht erwarten, dass x und y beide positiv (oder nicht-negativ) sind. Da $d \leq a$ und $d \leq b$, wäre die Summe zweier nicht-negativer Vielfachen von a bzw. b entweder 0 oder bereits mindestens d, und größer, falls nicht $a = b$ gilt. Wir würden also (fast) nur größere Zahlen als d erhalten können, falls beide x und y nicht-negativ sein müssten. Lässt man hingegen ganze Zahlen zu, so stimmt die Aussage der Proposition.

2. Die Proposition sagt nichts über die Eindeutigkeit der Zahlen $x, y \in \mathbb{Z}$, die der obigen Gleichung genügen. Tatsächlich gibt es unendlich viele solche Paare; darauf gehen wir später nochmal ein.

Wir werden keinen vollständigen Beweis dieser Proposition führen, geben aber einen Algorithmus an, um solche Zahlen x, y zu konstruieren (auch den Algorithmus werden wir nicht in voller Allgemeinheit abstrakt spezifizieren). Dieser Algorithmus wird *erweiterter euklidischer Algorithmus* genannt. Wir erläutern die Funktionsweise zuerst an Beispielen.

Beispiel. Seien die Zahlen $a = 217$ und $b = 63$ vorgegeben. Zunächst bestimmen wir den größten gemeinsamen Teiler dieser beiden Zahlen mit Hilfe des euklidischen Algorithmus: Als erstes teilen wir 217 durch 63 mit Rest:

$$217 = 3 \cdot 63 + 28.$$

Da der Rest nicht 0 ist, sagt uns der euklidische Algorithmus, dass $\mathrm{ggT}(217, 63) = \mathrm{ggT}(63, 28)$ ist und wir nun 63 mit Rest durch 28 teilen müssen:

$$63 = 2 \cdot 28 + 7.$$

Da der Rest erneut nicht 0 ist, haben wir $\mathrm{ggT}(63, 28) = \mathrm{ggT}(28, 7)$. (An dieser Stelle - oder vielleicht auch schon früher, wenn man etwas Erfahrung damit

hat - kann man schon „sehen“, also durch Ausprobieren der Teiler erschließen, was der größte gemeinsame Teiler ist. Allerdings fahren wir weiter mit dem euklidischen Algorithmus fort, sowohl, um darin Übung zu bekommen, als auch, weil es gleich für den erweiterten euklidischen Algorithmus nötig sein wird.) Wir teilen also 28 durch 7 und erhalten:

$$28 = 4 \cdot 7 + 0,$$

also ist der Rest 0 und es folgt

$$\mathrm{ggT}(217, 63) = \mathrm{ggT}(63, 28) = \mathrm{ggT}(28, 7) = 7.$$

Die obige Proposition besagt also, dass es ganze Zahlen x, y geben muss, sodass $217x + 63y = 7$ gilt.

Der erweiterte euklidische Algorithmus besteht nun darin, die Gleichungen, die man bei den Divisionen mit Rest erhalten hat, „aufzuwickeln“ und dadurch die gewünschten Zahlen x und y zu erreichen. Wäre man bei der zweiten Division mit Rest gleich fertig gewesen - etwa, wenn man $\mathrm{ggT}(63, 28)$ berechnet hat - so erhält man 7 als Summe bzw. Differenz von Vielfachen von 63 und 28 einfach durch Umstellung der Division mit Rest:

$$7 = 1 \cdot 63 - 2 \cdot 28.$$

Allerdings wollen wir ja 28 gar nicht verwenden, sondern nur Vielfache von 63 und 217 haben. Allerdings können wir 28 durch ebensolche Vielfache ausdrücken, da 28 ja gerade als Rest bei der Division von 217 durch 63 entstanden ist. Also ist

$$28 = 217 - 3 \cdot 63.$$

Nun können wir das anstelle von 28 in die Gleichung $7 = 1 \cdot 63 - 2 \cdot 28$ einsetzen und die Klammern auflösen:

$$7 = 1 \cdot 63 - 2 \cdot (217 - 3 \cdot 63) = 1 \cdot 63 - 2 \cdot 217 + 6 \cdot 63 = 7 \cdot 63 - 2 \cdot 217.$$

Also ist $x = -2$ und $y = 7$ eine ganzzahlige Lösung der Gleichung $217x + 63y = 7$.

Im Allgemeinen werden wir sukzessive den Rest aus der jeweils vorherigen Division mit Rest durch Dividenden und Divisor ausdrücken müssen und in die Gleichung einsetzen, die wir bis dahin erhalten haben, bis wir schließlich bei dem ersten Dividenden a und ersten Divisor b angekommen sind.

Wir demonstrieren das Verfahren an einem weiteren Beispiel.

Beispiel. In diesem Beispiel suchen wir ein Paar ganzer Zahlen (x, y), für die $13x + 21y = 1$ gilt. Die Lösung wird wieder nicht eindeutig sein; wir

interessieren uns zunächst für irgendeine Lösung. Wir führen zunächst den euklidischen Algorithmus durch:

$$\begin{aligned} 21 &= 1 \cdot 13 + 8 \\ 13 &= 1 \cdot 8 + 5 \\ 8 &= 1 \cdot 5 + 3 \\ 5 &= 1 \cdot 3 + 2 \\ 3 &= 1 \cdot 2 + 1 \\ 2 &= 2 \cdot 1 + 0. \end{aligned}$$

Der euklidische Algorithmus zeigt also nochmal, dass $\mathrm{ggT}(21, 13) = 1$ ist. (Kleine Nebenbemerkung: Die Zahlen auf den linken Seiten der Gleichungen sind Fibonacci-Zahlen, die man häufig auch in anderen Zusammenhängen kennenlernt.) Nun stellen wir die vorletzte Division mit Rest so um, dass wir die 1 durch 2 und 3 ausdrücken:

$$1 = 3 - 1 \cdot 2.$$

Die 2 erschien als Rest in der vorherigen Division mit Rest und kann daher durch 3 und 5 ausgedrückt werden:

$$2 = 5 - 1 \cdot 3.$$

Wir setzen das nun in die vorherige Gleichung ein und erhalten

$$\begin{aligned} 1 &= 3 - 1 \cdot (5 - 1 \cdot 3) \\ &= 2 \cdot 3 - 1 \cdot 5. \end{aligned}$$

Die 3 kann nun wiederum, als Rest der Division von 8 durch 5, durch 8 und 5 ausgedrückt werden:

$$3 = 8 - 1 \cdot 5.$$

Das setzen wir in die vorherige Gleichung ein und erhalten:

$$\begin{aligned} 1 &= 2 \cdot (8 - 1 \cdot 5) - 1 \cdot 5 \\ &= 2 \cdot 8 - 3 \cdot 5. \end{aligned}$$

Genauso ist 5 der Rest bei der Division von 13 durch 8 und wir erhalten:

$$5 = 13 - 1 \cdot 8.$$

Einsetzen in die vorherige Gleichung liefert erneut:

$$\begin{aligned} 1 &= 2 \cdot 8 - 3 \cdot (13 - 1 \cdot 8) \\ &= 5 \cdot 8 - 3 \cdot 13. \end{aligned}$$

Als letztes nutzen wir noch

$$8 = 21 - 1 \cdot 13$$

und erhalten

$$\begin{aligned} 1 &= 5 \cdot (21 - 1 \cdot 13) - 3 \cdot 13 \\ &= 5 \cdot 21 - 8 \cdot 13. \end{aligned}$$

Also liefert das Paar $(-8, 5)$ eine Lösung der Gleichung $13x + 21y = 1$.

Wir erläutern den erweiterten euklidischen Algorithmus nun etwas allgemeiner. Seien die natürlichen Zahlen $a \geq b \geq 1$ vorgegeben. Zunächst führt man die ursprüngliche Version des euklidischen Algorithmus durch. Dabei teilen wir zunächst a mit Rest durch b. Ist das Verfahren an dieser Stelle noch nicht zuende, so teilen wir im jeden nächsten Schritt den Divisor der letzten Division mit Rest durch den Rest dieser letzten Division mit Rest, und erhalten erneut einen Quotienten und Rest. Dies wird fortgesetzt, bis man bei einer Division Rest 0 bekommt. Wir schreiben das nochmal in Formeln auf:

$$\begin{aligned} &a = q_1 b + r_1, \quad 0 < r_1 < b; \\ &b = q_2 r_1 + r_2, \quad 0 < r_2 < r_1; \\ &r_1 = q_3 r_2 + r_3, \quad 0 < r_3 < r_2; \\ &\dots \\ &r_{m-4} = q_{m-2} r_{m-3} + r_{m-2}, \quad 0 < r_{m-2} < r_{m-3}; \\ &r_{m-3} = q_{m-1} r_{m-2} + r_{m-1}, \quad 0 < r_{m-1} < r_{m-2}; \\ &r_{m-2} = q_m r_{m-1}. \end{aligned}$$

Wie wir bereits diskutiert haben, ist nun $r_{m-1} = \mathrm{ggT}(a, b)$, und wir suchen ganze Zahlen $x, y \in \mathbb{Z}$, sodass $r_{m-1} = ax + by$ gilt. Zunächst stellen wir die vorletzte Gleichung nach r_{m-1} um und erhalten

$$r_{m-1} = r_{m-3} - q_{m-1} r_{m-2}.$$

Nun können wir r_{m-2} eliminieren, indem wir die Gleichung benutzen, in der r_{m-2} als Rest vorkommt. Das liefert:

$$\begin{aligned} r_{m-1} &= r_{m-3} - q_{m-1} r_{m-2} \\ &= r_{m-3} - q_{m-1}(r_{m-4} - q_{m-2} r_{m-3}) \\ &= (1 + q_{m-1} q_{m-2}) r_{m-3} - q_{m-1} r_{m-4}. \end{aligned}$$

Nun kann man die Gleichung darüber, in der r_{m-3} als Rest vorkommt, wiederum nach r_{m-3} auflösen und in die bis jetzt erreichte Darstellung von r_{m-1} einsetzen (und dadurch r_{m-3} eliminieren). So eliminiert man Schritt für Schritt

Reste aus dieser Gleichung, bis man einen Ausdruck für r_{m-1} in Termen von a und b hat. Das liefert eine Lösung für die obige Gleichung. Im Allgemeinen hat die Gleichung allerdings unendlich viele Lösungen. Wir gehen in Kürze darauf ein.

Wir machen ein weiteres Beispiel zum erweiterten euklidischen Algorithmus.

Beispiel. Wir suchen eine ganzzahlige Lösung der Gleichung $369x + 108y = \mathrm{ggT}(369, 108)$. Um wieder das Beispiel mit den Wassertanks aufzugreifen, wollen wir mit den 369-Liter- und 108-Liter-Behältern genau $\mathrm{ggT}(369, 108)$-Liter abmessen. Zunächst führen wir den euklidischen Algorithmus zur Bestimmung des größten gemeinsamen Teilers durch:

$$\begin{aligned} 369 &= 3 \cdot 108 + 45, \\ 108 &= 2 \cdot 45 + 18, \\ 45 &= 2 \cdot 18 + 9, \\ 18 &= 2 \cdot 9. \end{aligned}$$

Also ist $\mathrm{ggT}(369, 108) = 9$. Nun starten wir mit der vorletzten Gleichung und schreiben diese wie folgt um:

$$9 = 45 - 2 \cdot 18.$$

In Termen von Wasserbehältern hieße das, dass man 9 Liter mit einem 18-Liter-Behälter und einem 45-Liter-Behälter abmessen könnte. Das sind allerdings nicht die Behältergrößen, die uns zur Verfügung stehen. Hätten wir allerdings einen 45-Liter-Behälter, so ließe sich der 18-Liter-Behälter unter Zuhilfenahme des 108-Liter-Behälters simulieren, wie uns die zweite Gleichung verrät:

$$18 = 108 - 2 \cdot 45.$$

Also könnte man mit einem 45-Liter-Behälter und einem 108-Liter-Behälter genau 9 Liter abmessen, und zwar wie folgt:

$$\begin{aligned} 9 &= 45 - 2 \cdot 18 \\ &= 45 - 2 \cdot (108 - 2 \cdot 45) \\ &= 5 \cdot 45 - 2 \cdot 108. \end{aligned}$$

Nun können wir aber den 45-Liter-Behälter wiederum mit den vorgegebenen simulieren und erhalten:

$$\begin{aligned} 9 &= 5 \cdot 45 - 2 \cdot 108 \\ &= 5 \cdot (369 - 3 \cdot 108) - 2 \cdot 108 \\ &= 5 \cdot 369 - 17 \cdot 108. \end{aligned}$$

Somit liefert $x = 5$, $y = -17$ eine ganzzahlige Lösung unserer Gleichung.

Nun wollen wir darauf eingehen, wie viele ganzzahlige Lösungen eine Gleichung dieser Art hat. Um kleinere Zahlen zu haben, kehren wir zu einem früheren Beispiel zurück, in dem wir die Gleichung $13x + 21y = 1$ untersucht haben. Wir haben die Lösung $(-8, 5)$ dieser Gleichung gefunden, da wir mit dem erweiterten euklidischen Algorithmus die Identität

$$-8 \cdot 13 + 5 \cdot 21 = 1$$

gefunden haben. Man rechnet allerdings auch nach, dass das Paar $(13, -8)$ ebenfalls eine Lösung dieser Gleichung ist, da

$$13 \cdot 13 - 8 \cdot 21 = 1$$

gilt. Wir bemerken, dass $-8 + 21 = 13$ und $5 - 13 = -8$ gilt. Das ist allgemeiner wie folgt zu erklären: Hat man eine Lösung (x, y) der Gleichung $ax + by = \mathrm{ggT}(a, b)$ vorgegeben, so ist auch $(x - b, y + a)$ eine Lösung unserer Gleichung, denn wir können diese wie folgt äquivalent umformen:

$$\begin{array}{rrcl} & ax + by & = & \mathrm{ggT}(a, b) \\ \Leftrightarrow & ax - ab + ab + by & = & \mathrm{ggT}(a, b) \\ \Leftrightarrow & a(x - b) + b(a + y) & = & \mathrm{ggT}(a, b) \end{array}$$

Dass man dies beliebig häufig wiederholen könnte, und ähnlich auch die umgekehrten Vorzeichen erreichen könnte, ist die Beweisidee für die erste Aussage der folgenden Proposition.

Proposition 12.3. *Seien $a \geq b \geq 1$ natürliche Zahlen und sei $d = \mathrm{ggT}(a, b)$. Sei ferner $(x_0, y_0) \in \mathbb{Z} \times \mathbb{Z}$ eine Lösung der Gleichung $ax + by = d$. Dann ist für jede ganze Zahl c auch das Paar*

$$(x_0 - cb, y_0 + ca)$$

eine Lösung der obigen Gleichung. Falls $d = 1$ ist, ist zusätzlich jede Lösung der Gleichung $ax + by = 1$ von dieser Form.

Wir wollen jedoch nicht weiter auf den Beweis der Proposition eingehen. Als nächstes wollen wir sehen, wie der erweiterte euklidische Algorithmus zur Lösung der modularen Gleichungen der Form $ax \equiv 1 \mod b$ benutzt werden kann. (Die Lösung solcher Gleichungen war einer der Bausteine im RSA-Verschlüsselungsverfahren.)

Beispiel. Wir suchen alle ganzen Zahlen x, für die $100x \equiv 1 \mod 229$ gilt. Das ist nach Definition äquivalent zu $\frac{100x-1}{229} \in \mathbb{Z}$. Schreiben wir $y = \frac{100x-1}{229}$, so kann man diese Gleichheit äquivalent zu

$$100x - 229y = 1$$

umformen. Wir erhalten also alle Lösungen der obigen modularen Gleichung, wenn wir alle Paare ganzer Zahlen $(x, -y)$ finden, die der Gleichung $100x + 229 \cdot (-y) = 1$ genügen. Wir haben gesehen, dass der erweiterte euklidische Algorithmus für Lösung dieses Problems geeinigt ist. Wir führen diesen durch. Zunächst müssen wir den größten gemeinsamen Teiler von 229 und 100 bestimmen.

$$\begin{aligned} 229 &= 2 \cdot 100 + 29, \\ 100 &= 3 \cdot 29 + 13, \\ 29 &= 2 \cdot 13 + 3, \\ 13 &= 4 \cdot 3 + 1, \\ 3 &= 3 \cdot 1. \end{aligned}$$

Also ist $\mathrm{ggT}(229, 100) = 1$. Nun können wir das Rückeinsetzen durchführen:

$$\begin{aligned} 1 &= 13 - 4 \cdot 3 \\ &= 13 - 4 \cdot (29 - 2 \cdot 13) \\ &= 9 \cdot 13 - 4 \cdot 29 \\ &= 9 \cdot (100 - 3 \cdot 29) - 4 \cdot 29 \\ &= 9 \cdot 100 - 31 \cdot 29 \\ &= 9 \cdot 100 - 31 \cdot (229 - 2 \cdot 100) \\ &= 71 \cdot 100 - 31 \cdot 229. \end{aligned}$$

Also ist $(71, -31)$ eine Lösung der Gleichung $100x + 229z = 1$. Nun besagt die Proposition 12.3, dass alle weiteren Lösungen von der Form $(71 + 229c, -31 - 100c)$ sind. Da wir uns nur für die erste Variable interessieren, erhalten wir:

$$\{x \in \mathbb{Z} \,|\, 100x \equiv 1 \mod 229\} = \{71 + 229c \,|\, c \in \mathbb{Z}\}.$$

Also ist jede ganze Zahl x, für die $100x \equiv 1 \mod 229$ gilt, von der Form $71 + 229c$ für eine ganze Zahl c (und jede solche Zahl ist eine Lösung der modularen Gleichung).

13 Primzahlen

In diesem Kapitel wollen wir die Primzahlen, denen wir schon mehrfach im Laufe dieser Vorlesung begegnet sind, eingehender studieren. Zur Erinnerung nochmal die Definition:

Definition 13.1. Eine natürliche Zahl p heißt **Primzahl**, falls sie genau zwei unterschiedliche Teiler hat.

Insbesondere ist die Zahl 1 keine Primzahl. Das ist in erster Linie eine Konvention; wir sehen später, warum diese Konvention sinnvoll ist.

Die ersten Primzahlen sind $2, 3, 5, 7, 11, 13, \ldots$, die ersten nicht-Primzahlen > 1 (man sagt auch: zusammengesetzte Zahlen) sind 4, 6, 8, 9 ,10, 12, 14, 15, 16, ... Die Primzahlen scheinen immer seltener zu werden; man kann sich also fragen, ob sie irgendwann ganz aufhören. Dieser Frage werden wir nachgehen.

Die Primzahlen sind aus mehreren Gründen interessant. Im gewissen Sinne, den wir bald präzisieren, sind Primzahlen die kleinsten Bausteine, aus denen alle anderen Zahlen zusammengesetzt sind.

Primzahlen sind auch von praktischer Bedeutung. Sie spielen etwa im RSA-Verschlüsselungsverfahren eine entscheidende Rolle. Während es einfach war, die ersten Primzahlen zu finden, ist Suche nach großen Primzahlen im Allgemeinen ein schwieriges Problem. Auch schwierig ist es, eine vorgegebene große Zahl in Primfaktoren zu zerlegen (warum das überhaupt gehen sollte, sehen wir auch bald).

Wir fangen aber mit einer anderen Problemstellung an, in der erste Eigenschaften von Primzahlen deutlich werden.

Beispiel. Seien $a, b, n \in \mathbb{N} \setminus \{0\}$ natürliche Zahlen. Wir nehmen an, dass n das Produkt $a \cdot b$ teilt. Zunächst fragen wir uns, ob dann notwenderweise n auch a teilt. Diese Vermutung ist schnell widerlegt: Beispielsweise teilt 2 das Produkt $1 \cdot 2$, aber 2 teilt nicht 1. Allerdings teilt 2 den anderen Faktor, 2.

Man kann sich also als nächstes Fragen, ob n notwenderweise a oder b (oder gar beide) teilt. Das ist allerdings auch schnell widerlegt: Die Zahl 6 teilt beispielsweise das Produkt $2 \cdot 3$, aber weder 2 noch 3 ist durch 6 teilbar.

Allerdings kommen nun die Primzahlen ins Spiel. Die obige Aussage, die für allgemeine natürliche Zahlen nicht gilt, stellt sich für Primzahlen als wahr heraus. Wir wollen dies nun formulieren und beweisen.

Proposition 13.2. *Seien $a, b \in \mathbb{N}$ natürliche Zahlen und p eine Primzahl. Ist das Produkt ab durch p teilbar, so ist auch a oder b (oder beide) durch p teilbar.*

Bevor wir diese Proposition beweisen, brauchen wir das folgende Lemma:

Lemma 13.3. *Sei p eine Primzahl und a eine natürliche Zahl. Dann gilt:*

$$\mathrm{ggT}(p,a) = \begin{cases} p, & \text{falls } a \text{ durch } p \text{ teilbar ist,} \\ 1, & \text{falls } a \text{ nicht durch } p \text{ teilbar ist.} \end{cases}$$

Beweis. Der Beweis ist in diesem Fall nicht schwer. Da p nur zwei Teiler hat, nämlich 1 und p, muss der größte gemeinsame Teiler von a und p, der ja insbesondere ein Teiler von p ist, entweder 1 oder p sein. Ist nun a durch p teilbar, so ist p der größte gemeinsame Teiler, und andernfalls ist 1 der größte gemeinsame Teiler von a und p. □

Wir beweisen nun die Proposition.

Beweis der Proposition 13.2. Seien $a, b \in \mathbb{N}$ natürliche Zahlen und p eine Primzahl. Sei ab durch p teilbar. Wir führen einen Widerspruchsbeweis. Dafür nehmen wir an, dass weder a noch b durch p teilbar ist, und führen diese Annahme zum Widerspruch.

Ist also weder a noch b durch p teilbar, so folgt nach dem Lemma, das wir soeben bewiesen haben, dass $\mathrm{ggT}(a,p) = 1$ und $\mathrm{ggT}(b,p) = 1$ gilt. Nun haben wir gesehen, dass es dann ganze Zahlen $x, y, u, v \in \mathbb{Z}$ geben muss, sodass

$$\begin{aligned} px + ay &= 1 \text{ und} \\ pu + bv &= 1 \end{aligned}$$

gilt. (Diese Zahlen lassen sich jeweils mit dem erweiterten euklidischen Algorithmus für a und p bzw. für b und p finden.) Wir multiplizieren die erste Gleichung mit bv und erhalten:

$$\begin{aligned} px + ay &= 1 \,|\cdot bv \\ \Rightarrow p \cdot bvx + ab \cdot yv &= bv. \end{aligned}$$

Nun setzen wir diesen Ausdruck für bv in die zweite Gleichung ein und erhalten:

$$1 = pu + bv = p \cdot u + p \cdot bvx + ab \cdot yv.$$

Da nach Voraussetzung ab durch p teilbar ist, ist $\frac{ab}{p}$ eine natürliche Zahl und somit können wir auf der rechten Seite p ausklammern und erhalten:

$$\begin{aligned} 1 &= p \cdot u + p \cdot bvx + ab \cdot yv \\ &= p \cdot (u + bvx + \frac{ab}{p} \cdot yv), \end{aligned}$$

und die Zahl $u + bvx + \frac{ab}{p} \cdot yv$ ist eine ganze Zahl. Das hieße aber, dass $\frac{1}{p}$ eine ganze Zahl ist. Da das für die Primzahl $p \geq 2$ nicht möglich ist, erhalten wir einen Widerspruch. Somit muss unsere Annahme, dass weder a noch b durch p teilbar sind, falsch gewesen sein. Das liefert die Behauptung. □

Nun kommen wir zu einer zentralen Aussage der elementaren Zahlentheorie. Sie rechtfertigt insbesondere die Beschreibung der Primzahlen als kleinste Bausteine der natürlichen Zahlen.

Satz 13.4 (Primfaktorzerlegung)**.** *Jede natürliche Zahl $n > 1$ hat genau eine Zerlegung in der Form*

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k},$$

wobei $p_1 < p_2 < \ldots < p_k$ Primzahlen sind und die Exponenten $\alpha_1, \alpha_2, \ldots, \alpha_k$ natürliche Zahlen ≥ 1 sind.

Wir betrachten zunächst einige Beispiele und skizzieren danach die Beweisidee von diesem Satz.

Beispiel.

- Die Primfaktorzerlegung der Zahl 400 ist gegeben durch $2^4 \cdot 5^2$. Wir haben festgelegt, dass wir die Primzahlen aufsteigend listen, sonst wäre die Darstellung $5^2 \cdot 2^4$ ebenfalls zugelassen. Außerdem fordern wir, dass die Faktoren zusammenfasst werden und wir nicht

  $$2 \cdot 2 \cdot 2 \cdot 2 \cdot 5 \cdot 5$$

  schreiben. (Die Darstellungen unterscheiden sich nur minimal und es ist Konventionsfrage, ob man die Zusatzbedingungen stellt, um etwas eindeutiges zu erhalten, oder eine leichte Ambiguität zulässt.) Die Bedingung, dass die Exponenten nicht 0 sein dürfen, bedeutet, dass wir keine Primzahlen in der Primfaktorzerlegung einer Zahl n haben, durch diese Zahl n gar nicht teilbar ist. Wäre Null als Exponen zulässig, so wäre etwa $400 = 2^4 \cdot 3^0 \cdot 5^2$ eine Primfaktorzerlegung von 400, in der 3 vorkommt, obwohl 400 durch 3 teilbar ist. Schließlich sehen wir bei dieser Gelegenheit einen Grund dafür, dass 1 via Konvention keine Primzahl ist: Wäre 1 prim, so wäre auch $1^{17} \cdot 2^4 \cdot 5^2$ und $1^{2015} \cdot 2^4 \cdot 5^2$ (und unendlich viele weitere Produkte) ebenfalls Primfaktorzerlegungen von 400, was also wiederum die Eindeutigkeit zunichte machen würde.

- Die Primfaktorzerlegung von 323 ist $17 \cdot 19$. Im Allgemeinen ist es ein schwieriges Problem, die Primfaktorzerlegung einer vorgegebener Zahl zu bestimmen (jedenfalls für große Zahlen). Es gibt dafür unterschiedliche Verfahren, und es wird an weiteren Verfahren geforscht. Die naivste Methode ist es, alle Primzahlen $\leq n$ als mögliche Teiler von n auszuprobieren. Allerdings bemerkt man schnell, dass es reicht, bis $\sqrt{n}$ zu gehen: Hat man nämlich einen Teiler $d > \sqrt{n}$, so gilt für den Quotienten $\frac{n}{d} < \sqrt{n}$; zu jedem Teiler über $\sqrt{n}$ muss also ein Teiler $< \sqrt{n}$ gehören, und wir müssen nur Teiler $\leq \sqrt{n}$ überprüfen. In diesem Fall müssen wir allerdings tatsächlich zu der größten Primzahl im Bereich $\leq \sqrt{323}$ gehen, da $\sqrt{323} \approx 17,97$.

- Die Primfaktorzerlegung von 1024 ist 2^{10}. Insbesondere kann es also vorkommen, dass die Primfaktorzerlegung nur eine Primzahl enthält.
- Die Primfaktorzerlegung jeder Primzahl ist durch die Primzahl selbst gegeben.
- Aus den Primfaktorzerlegungen von 400 und 1024 können wir die Primfaktorzerlegung von $400 \cdot 1024$ gewinnen (auch ohne dieses Produkt explizit zu bestimmen): wir sortieren die Primzahlen in dem Produkt $2^4 \cdot 5^2 \cdot 2^{10}$ wieder nach Größe und fassen die Potenzen von denselben Primzahlen (in diesem Fall sind es nur Potenzen von 2) zusammen und erhalten $2^{14} \cdot 5^2$ als Primfaktorzerlegung von $400 \cdot 1024$.

Wir geben nun eine Idee davon, wie der Beweis vom Satz 13.4 funktioniert.

Beweisskizze. Der Beweis hat zwei Teile: Im ersten Teil müssen wir zeigen, dass jede natürliche Zahl > 1 eine Primfaktorzerlegung besitzt. Im zweiten Teil werden wir die Argumentation für die Eindeutigkeit der Zerlegung skizzieren.

In beiden Teilen wird eine Variante von Induktion benutzt: Anstatt das Problem, wie im Falle der bisher betrachteten Induktionsargumente, auf die unmittelbar vorangehende Zahl zurückzuführen, werden wir die Aussage für *alle* kleineren Zahlen als eine Art Induktionsvoraussetzung benutzen. Die Schlussweise ist in etwa wie folgt: Weiß man, dass die Aussage für die Zahl 2 gilt, so auch, dass sie für 3 gilt; um diese für 4 zu beweisen, muss man (unter Umständen) sowohl die Gültigkeit der Aussage für 2 als auch für 3 benutzen; um die Aussage für 5 zu zeigen, braucht man (unter Umständen) die Aussagen für $2, 3, 4$, und so weiter.

Nun fangen wir mit der Existenz einer Primfaktorzerlegung an. Ist n eine Primzahl, so haben wir bereits eingesehen, dass sie dann ihre eigene Primfaktorzerlegung liefert. Ist $n > 1$ nun keine Primzahl, so hat sie mindestens einen Teiler, der weder 1 noch n ist, also muss für diesen Teiler d gelten: $1 < d < n$. Also gilt auch für $e = \frac{n}{d}$: $1 < e < n$. Nun sind $e, d > 1$ zwei Zahlen, die kleiner als n sind, also nach unserer „Induktionsvoraussetzung“ bereits eine Primfaktorzerlegung. Nun können wir, wie in dem Beispiel zuvor, die Faktoren wieder nach Größe sortieren und die Potenzen von gleichen Primzahlen zusammenfassen, und erhalten aus den Primfaktorzerlegungen von d und e auf diese Weise eine Primfaktorzerlegung der Zahl n.

Nun skizzieren wir den Beweis der Eindeutigkeit der Primfaktorzerlegung. Sei also $n > 1$ eine natürliche Zahl und

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k} = q_1^{\beta_1} q_2^{\beta_2} \dots q_l^{\beta_l}$$

zwei Primfaktorzerlegungen von n. Wir müssen nun zeigen, dass beide Primfaktorzerlegung übereinstimmen, also in Wirklichkeit ein und dieselbe Zerlegung sind. Da p_1 in der Primfaktorzerlegung von n auftaucht, ist p_1 insbesondere ein Teiler von n, also auch von dem Produkt $q_1^{\beta_1} q_2^{\beta_2} \dots q_l^{\beta_l}$. Durch

sukzessive Anwendung der Proposition 13.2 erhalten wir zunächst, dass p_1 einen der Faktoren $q_1^{\beta_1}$, $q_2^{\beta_2}$, ..., $q_l^{\beta_l}$ teilt, und durch nochmalige Anwendung der Proposition 13.2 sehen wir, dass p_1 eine der Zahlen $q_1, q_2, \ldots, q_l$ teilt. Da dies alles Primzahlen sind, folgt $p_1 = q_i$ für ein $1 \leq i \leq l$. Somit kann man den Faktor auf beiden Seiten kürzen. Nun erhält man zwei Primfaktorzerlegungen für eine kleinere Zahl, nämlich $\frac{n}{p_1}$, und für diese wissen wir nach „Induktionsvoraussetzung", dass die Primfaktorzerlegung eindeutig ist, die beiden Produkte ohne p_1 also übereinstimmen. Daher müssen Sie auch übereinstimmen, wenn wir beide mit dem Faktor p_1 multiplizieren und die beiden Darstellungen für n erhalten. Das Argument liefert also die Eindeutigkeit der Primfaktorzerlegung. □

Jetzt wissen wir, dass jede natürliche Zahl aus Primzahlen zusammengesetzt ist. Nun wenden wir uns der Frage zu, ob die Primzahlen irgendwann „aufhören" und danach nur noch Zahlen kommen, die aus den bisherigen endlich vielen Primzahlen zusammengesetzt sind. Wir zeigen nun, dass dies nicht vorkommen kann, und es unendlich viele Primzahlen gibt.

Satz 13.5 (Euklid)**.** *Es gibt unendlich viele Primzahlen.*

Beweis. Wir führen erneut einen Widerspruchsbeweis. Dafür nehmen wir an, dass es nur endlich viele Primzahlen gibt, und führen diese Annahme gleich zum Widerspruch. Wenn es also nur endlich viele Primzahlen gäbe, dann könnten wir diese durchnummerieren und eine endliche Auflistung davon anfertigen: $p_1 = 2, p_2 = 3, \ldots, p_k$. (Wir wissen nicht, wie groß k ist, aber laut unserer Annahme gibt es ein solches letztes p_k, das die größte Primzahl ist.)

Wir betrachten nun die Zahl

$$p_1 \cdot p_2 \cdot \ldots \cdot p_k + 1.$$

Da die Zahl $p_1 \cdot p_2 \cdot \ldots \cdot p_k$ durch jede Primzahl teilbar ist, die es laut unserer Annahme gibt, muss also für jedes $1 \leq i \leq k$ gelten:

$$p_1 \cdot p_2 \cdot \ldots \cdot p_k + 1 \equiv 1 \mod p_i.$$

Da nun $0 \not\equiv 1 \mod p_i$ (da die positive Zahl 1 kleiner als p_i ist und somit nicht durch p_i teilbar ist), ist also

$$p_1 \cdot p_2 \cdot \ldots \cdot p_k + 1$$

durch keine der Zahlen $p_1, p_2, \ldots, p_k$ teilbar. Ferner ist $p_1 \cdot \ldots p_k$ als Produkt positiver Zahlen selbst positiv, und somit ist

$$p_1 \cdot p_2 \cdot \ldots \cdot p_k + 1 > 1.$$

Diese Zahl ist also durch keine Primzahl teilbar, muss aber nach dem eben bewiesenen Satz 13.4 eine Primfaktorzerlegung besitzen. Damit haben wir unsere Annahme zum Widerspruch gebracht, und das zeigt wiederum, dass es unendlich viele Primzahlen geben muss. □

Wir wollen nochmal anmerken, dass wir nun zwar wissen, dass nach jeder Primzahl irgendwann eine neue Primzahl kommt, allerdings ist es im Allgemeinen schwer, diese nächste Primzahl zu finden.

Nun wollen wir ein weiteres wichtiges Resultat über die Primzahlen beweisen: Den kleinen Satz von Fermat.

Satz 13.6 (kleiner Satz von Fermat)**.** *Sei p eine Primzahl und a eine Zahl, die nicht durch p teilbar ist. Dann gilt:*

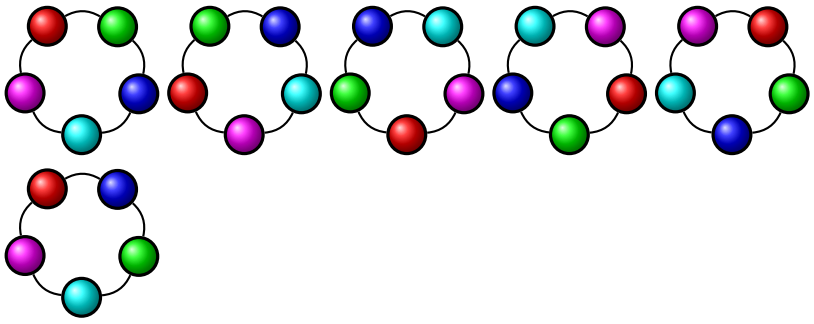
$$a^{p-1} \equiv 1 \mod p.$$

Ist ferner b eine beliebige natürliche Zahl, so gilt $b^p \equiv b \mod p$.

Wir geben zunächst eine Beweisskizze an, die etwas schwerer zu präzisieren ist, allerdings leichter zu erfassen. Danach führen wir einen anderen Beweis, der nicht ganz so anschaulich ist, allerdings leichter aufzuschreiben.

Beweisskizze. Wir skizzieren den Beweis der zweiten Aussage, nämlich dass $a^p \equiv a \mod p$ für jede natürliche Zahl a gilt. Wir stellen uns vor, Perlenketten mit jeweils genau p Perlen zu bilden. Dabei haben wir Perlen in a verschiedenen Farben zur Verfügung. Insgesamt erhalten wir a^p verschiedene Perlenketten. Die einfarbigen Perlenketten sind langweilig, also entfernen wir die a unterschiedlichen einfarbigen Perlenketten aus unserer Sammlung.

Nun legen wir die Perlenketten auf eine Ebene und schließen diese. Dabei binden wir die Fadenenden so gut zusammen, dass man nachher nicht sehen kann, wo die Ketten verbunden worden sind. Dabei kann man nun manche Ketten so drehen, dass sie genau wie bereits vorhandene Ketten aussehen (obwohl sie vor dem Schließen unterschiedlich waren), wie im Beispiel angedeutet.

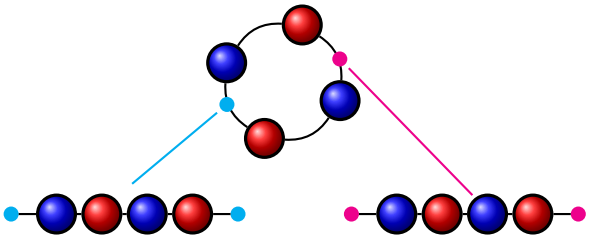


(Die oberen 5 Perlenketten entstehen durch Drehungen aus der ersten; die Perlenkette darunter ist jedoch wirklich anders.)

Bei genauerer Betrachtung merkt man, dass es jede geschlossene Perlenkette nun genau p mal gibt (wir erlauben nur Drehungen, um die Perlenketten als gleich anzusehen): Man kann die geschlossene Perlenkette an jedem der p Zwischenräume aufschneiden und kriegt eine der vorherigen Perlenketten.

Also gibt es $\frac{a^p-a}{p}$ Perlenketten, die jeweils nicht durch Drehungen ineinander überführt werden können. Das muss allerdings eine natürliche Zahl sein, da sie die Anzahl von gewissen Gegenständen beschreibt. Nach Definition von Teilbarkeit und von Kongruenz heißt es allerdings genau, dass $a^p \equiv a \mod p$ ist.

Nun ist Vorsicht geboten. An welcher Stelle haben wir ausgenutzt, dass p eine Primzahl ist? Es scheint, als würde dieser Beweis für beliebige natürliche Zahlen p funktionieren. Allerdings ist z.B. 2 nicht durch 4 teilbar und es gilt trotzdem: $2^4 \equiv 0 \not\equiv 2 \mod 4$. Also muss die Tatsache, dass p eine Primzahl ist, doch irgendwo benutzt worden sein. Tatsächlich ist das der Fall: Ist p keine Primzahl, so bekommt man durch Aufschneiden einer geschlossenen Kette an p verschiedenen Stellen zwar auch p Perlenketten, allerdings sind diese im Allgemeinen nicht alle unterschiedlich, wie wir im folgenden Beispiel sehen:



Da es nun etwas schwieriger ist, zu formalisieren, warum das für Primzahlen nicht passieren kann, führen wir stattdessen einen anderen Beweis. □

Beweis. Der Beweis ist in drei Schritte unterteilt. Erst im letzten Schritt wird der Zusammenhang zwischen den ersten beiden Schritten und der Aussage des Satzes klar.

Schritt 1: Hier zeigen wir zunächst, dass in der Folge der Zahlen

$$1 \cdot a, 2 \cdot a, \ldots, (p-1) \cdot a$$

kein Paar kongruent modulo p ist. Etwas präziser: Wir zeigen, dass für jede natürliche Zahl i und (größere) natürliche Zahl j zwischen 1 und $p-1$, also mit $1 \leq i < j \leq p-1$, gilt: $i \cdot a \not\equiv j \cdot a \mod p$.

Wir führen einen Widerspruchsbeweis. Angenommen also, es gäbe ein $i \in \mathbb{N}$ und ein $j \in \mathbb{N}$ mit $1 \leq i < j \leq p-1$ und $i \cdot a \equiv j \cdot a \mod p$. Wir wollen diese Annahme zum Widerspruch führen. Nach Definition von Kongruenz und von Teilbarkeit ist $i \cdot a \equiv j \cdot a \mod p$ äquivalent dazu, dass $\frac{i \cdot a - j \cdot a}{p}$ eine ganze Zahl ist. Das ist wiederum genau dann der Fall, wenn das Produkt dieser Zahl mit -1 eine ganze Zahl ist, also die Zahl

$$-\frac{i \cdot a - j \cdot a}{p} = \frac{j \cdot a - i \cdot a}{p} = \frac{(j-i)a}{p}.$$

Das Produkt $(j-i)a$ ist also durch die Primzahl p teilbar. Nach Proposition 13.2 muss also mindestens einer der Faktoren $j-i$ und a durch p teilbar sein. Da wir vorausgesetzt haben, dass a nicht durch p teilbar ist, muss also $j-i$ durch p teilbar sein. Andererseits ist $j-i$ eine Zahl, die mindestens 1 ist, da $j > i$ natürliche Zahlen sind, und höchstens $p-2$, da $j \leq p-1$ und $i \geq 1$ ist. Somit liegt der Quotient $\frac{j-i}{p}$ strikt zwischen 0 und 1 und kann somit keine natürliche Zahl sein. Also kann $j-i$ auch nicht durch p teilbar sein. Das ist ein Widerspruch, also muss unsere Annahme falsch gewesen sein und es gilt für alle $i, j \in \mathbb{N}$ mit $1 \leq i < j \leq p-1$:

$$i \cdot a \not\equiv j \cdot a \mod p.$$

Damit ist die Behauptung des ersten Beweisabschnitts gezeigt.

Schritt 2: Als nächstes zeigen wir, dass die Folge

$$1 \cdot a, 2 \cdot a, \ldots, (p-1) \cdot a$$

genau dieselben Reste modulo p hat wie die Folge

$$1, 2, \ldots, (p-1),$$

allerdings im Allgemeinen in einer anderen Reihenfolge geschrieben.

Wir wissen, dass alle möglichen Reste modulo p durch $0, 1, 2, \ldots, p-1$ gegeben sind. Da wir nach Schritt 1 wissen, dass die $p-1$ Zahlen der Folge $1 \cdot a, 2 \cdot a, \ldots, (p-1) \cdot a$ lauter unterschiedliche Reste modulo p haben, reicht es zu zeigen, dass 0 in dieser Folge nicht vorkommt - dann müssen nämlich die $p-1$ verschiedene Reste, die übrigbleiben, alle vorkommen, wenn auch nicht notwenderweise in derselben Reihenfolge.

Wir führen wieder einen Widerspruchsbeweis und nehmen an, für eine natürliche Zahl i mit $1 \leq i \leq p-1$ würde $i \cdot a \equiv 0 \mod p$ sein. Nun läuft das Argument ganz ähnlich wie im Schritt 1: Wir wissen erneut, dass p das Produkt $i \cdot a$ teilen müsste, aber weder die kleinere Zahl $i < p$ noch die nach Voraussetzung nicht durch p teilbare Zahl a teilen kann. Dadurch erhalten wir auch in diesem Fall einen Widerspruch. Also wissen wir, dass in der obigen Folge der Vielfachen von a jeder Rest zwischen 1 und $p-1$ einschließlich genau einmal auftritt.

Schritt 3: Nun führen wir die bisherigen Ergebnisse zusammen. Da bei der Multiplikation von ganzen Zahlen das Ergebnis nicht von der Reihenfolge der Faktoren abhängt, und da wir festgestellt haben, dass die Zahlen $1 \cdot a, 2 \cdot a, \ldots, (p-1) \cdot a$ kongruent sind zu irgendwie umgestellten Zahlen $1, 2, \ldots, (p-1)$, haben also die Produkte beider Folgen denselben Rest bei der Division durch p, also nochmal in Formeln:

$$1 \cdot a \cdot 2 \cdot a \cdot \ldots \cdot (p-1) \cdot a \equiv 1 \cdot 2 \cdot \ldots \cdot (p-1) \mod p.$$

Wir sortieren die Faktoren auf der linken Seite um, sodass wir die $(p-1)$ Faktoren a zu a^{p-1} zusammenfassen. Ferner erinnern wir uns, dass das Produkt aller natürlicher Zahlen zwischen 1 und $p-1$ einschließlich auch mit $(p-1)!$ bezeichnet wird und erhalten:

$$(p-1)! \cdot a^{p-1} \equiv (p-1)! \mod p.$$

Würde es sich um Gleichung in ganzen Zahlen handeln, so könnten wir beide Seiten einfach durch $(p-1)!$ teilen und würden so die Aussage erhalten. Da die Division bei modularen Gleichung im Allgemeinen nicht möglich ist, müssen wir uns diesen Ausdruck wieder genauer anschauen.

Nach Proposition 13.2 wissen wir, dass das Produkt $(p-1)! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (p-1)$ nicht durch p teilbar ist, da p sonst einen der Faktoren teilen müssten, wir aber bereits im ersten Schritt gesehen haben, dass keine Zahl zwischen 1 und $p-1$ einschließlich durch p teilbar ist.

Wir können nun die Identität

$$(p-1)! \cdot a^{p-1} \equiv (p-1)! \mod p.$$

nach Definition der Kongruenz umformulieren zu „p teilt die Differenz $(p-1)! \cdot a^{p-1} - (p-1)! = (p-1)! \cdot (a^{p-1} - 1)$“. Da die Primzahl p dieses Produkt teilt, aber nach der vorangehenden Überlegen nicht die Zahl $(p-1)!$ teilen kann, muss also $a^{p-1} - 1$ durch p teilbar sein.

Das wiederum ist äquivalent zu

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

Das ist gerade die erste Behauptung des Satzes.

Die zweite Behauptung folgt nun unschwer mit einer Fallunterscheidung. Ist b nicht durch p teilbar, so wissen wir bereits $b^{p-1} \equiv 1 \mod p$. Nun multiplizieren wir beide Seiten mit b und erhalten $b^p \equiv b \mod p$. In dem zweiten Fall, nämlich wenn b durch p teilbar ist, gilt $b \equiv 0 \mod p$ und somit auch

$$b^p \equiv 0^p \equiv 0 \equiv b \mod p.$$

In beiden Fällen erhalten wir nun $b^p \equiv b \mod p$. Das vervollständigt den Beweis des Satzes. □

Somit ist nun der kleine Satz von Fermat bewiesen, der, wie wir bereits gesehen haben, im RSA-Verschlüsselungsverfahren eine wesentliche Rolle spielt.

Wir verdeutlichen die Aussage aus dem Schritt 2 des Beweises nochmal in einem Beispiel:

Beispiel. Betrachten wir $p = 11$ und $a = 2$. Dann haben die Vielfachen von 2 die folgenden Reste modulo 11:

$$\begin{array}{rcll}
2 \cdot 1 & \equiv & 2 & \mod 11 \\
2 \cdot 2 & \equiv & 4 & \mod 11 \\
2 \cdot 3 & \equiv & 6 & \mod 11 \\
2 \cdot 4 & \equiv & 8 & \mod 11 \\
2 \cdot 5 & \equiv & 10 & \mod 11 \\
2 \cdot 6 & \equiv & 12 \equiv 1 & \mod 11 \\
2 \cdot 7 & \equiv & 14 \equiv 3 & \mod 11 \\
2 \cdot 8 & \equiv & 16 \equiv 5 & \mod 11 \\
2 \cdot 9 & \equiv & 18 \equiv 7 & \mod 11 \\
2 \cdot 10 & \equiv & 20 \equiv 9 & \mod 11
\end{array}$$

Wir sehen als auch hier konkret, dass jeder Rest zwischen 1 und 10 einschließlich auf der rechten Seide vorkommt, wie bereits im Schritt 2 des obigen Beweises gezeigt wurde.

Es sei noch angemerkt, dass der kleine Satz von Fermat die Grundlage einiger Primzahltests bildet. Die einfachste (und in der Form ungebräuchliche) Version dieses Tests ist in etwa wie folgt: Will man bei einer Zahl $k > 2$ prüfen, ob dies eine Primzahl ist, so kann man zunächst $2^{k-1} \mod k$ ausrechnen und bestimmen, ob der Rest bei der Division von 2^{k-1} durch k genau 1 sein wird. Ist k eine Primzahl, so muss das ja nach dem kleinen Satz von Fermat der Fall sein. Gilt also $2^{k-1} \not\equiv 1 \mod k$ für ein $k > 2$, so wissen wir, dass k keine Primzahl ist. Gilt die Kongruenz $2^{k-1} \equiv 1 \mod k$, so wissen wir nicht, ob k eine Primzahl ist oder nicht. Allerdings können wir anstelle von 2 nun auch weitere Zahlen a einsetzen, die nicht durch k teilbar sind, und hoffen, dass wir irgendwann $a^{k-1} \not\equiv 1 \mod k$ erhalten, sodass wir wissen, dass k keine Primzahl sein kann. Allerdings gibt es Zahlen, die sogenannten Carmichael-Zahlen, die diesen Test für alle natürlichen Zahlen a, die nicht durch k teilbar sind, bestehen, und trotzdem keine Primzahlen sind, diese entziehen sich den Möglichkeiten des beschriebenen Tests. Tatsächlich werden (deutlich verbesserte und dadurch etwas kompliziertere) Tests auf der Grundlage des kleinen Satzes von Fermat in der Praxis genutzt, um große Primzahlen zu finden.

14 Chinesischer Restsatz

In diesem Abschnitt wollen wir die letzte Zutat behandeln, die wir in der mathematischen Behandlung der Grundversion von der RSA-Verschlüsselung brauchten: den chinesischen Restsatz. Wir haben bereits gesehen, dass wir mit dem erweiterten euklidischen Algorithmus modulare Gleichungen lösen können.

Nun geht es darum, Systeme von modularen Gleichungen zu lösen. Wir fangen mit einem einfachen Beispiel an.

Beispiel. Wir suchen alle ganze Zahlen $x \in \mathbb{Z}$, die gleichzeitig beiden modularen Gleichungen

$$\begin{cases} x \equiv 2 \mod 4 \\ x \equiv 0 \mod 3 \end{cases}$$

genügen. In diesem Fall findet man durch Ausprobieren recht einfach die Lösung 6, nach einigen weiteren Versuchen findet man ferner die Zahlen $-6, 18$ und 30, die der modularen Gleichung ebenfalls genügen. Man sieht schnell ein, dass, da $12 \equiv 0 \mod 4$ und $12 \equiv 0 \mod 3$, die Addition oder Subtraktion von 12 zu einer Lösung wieder eine Lösung produzieren muss. Tatsächlich kann man zeigen, dass für alle Lösungen dieses modularen Gleichungssystem genau die Elemente der Menge $\{6 + 12k \mid k \in \mathbb{Z}\}$ sind.

In diesem Beispiel kamen wir durch Ausprobieren zu unserer Lösung. Man kann sich jedoch vorstellen, dass es im Allgemeinen, gerade bei größeren Zahlen, viel schwieriger ist, eine Lösung zu raten. Außerdem ist häufig auch ein Lösungsalgorithmus hilfreich, den man implementieren kann. Eine recht allgemeine Methode zur Behandlung der modularen Gleichungssysteme liefert der chinesische Restsatz, den wir allerdings nicht beweisen werden.

Satz 14.1 (Chinesischer Restsatz). *1. Seien $n_1, n_2, \ldots, n_k \geq 2$ paarweise teilerfremde natürliche Zahlen (paarweise teilerfremd bedeutet: für jedes i, j mit $1 \leq i \neq j \leq n$ gilt: $\mathrm{ggT}(n_i, n_j) = 1$). Dann hat das modulare Gleichungssystem*

$$\begin{cases} x \equiv a_1 \mod n_1 \\ x \equiv a_2 \mod n_2 \\ \ldots \\ x \equiv a_k \mod n_k \end{cases}$$

genau eine Lösung im Bereich $0 \leq x \leq n_1 \cdot n_2 \cdot \ldots \cdot n_k - 1$.

2. Seien $n_1 \geq 2$ und $n_2 \geq 2$ teilerfremde natürliche Zahlen. Nach Proposition 12.2 gibt es ganze Zahlen $u, v \in \mathbb{Z}$, sodass $n_1 u + n_2 v = 1$ gilt.

Ist eine beliebige Darstellung dieser Art vorgegeben, so lassen sich alle Lösungen des modularen Gleichungssystems

$$\begin{cases} x \equiv a_1 \mod n_1 \\ x \equiv a_2 \mod n_2 \end{cases}$$

schreiben als

$$n_1 u a_2 + n_2 v a_1 + n_1 n_2 k,$$

für irgendeine Zahl $k \in \mathbb{Z}$ (und jede Zahl dieser Form ist eine Lösung des modularen Gleichungssystems).

Wir schauen uns zwei Beispiele zu diesem Satz an.

Beispiel. Zunächst wollen wir uns davon überzeugen, dass das modulare Gleichungssystem, mit dem wir angefangen haben, auch mit Hilfe des chinesischen Restsatzes lösbar ist. Eine Darstellung der Form $4u + 3v = 1$ ist durch $u = 1$, $v = -1$ gegeben (es gibt natürlich auch weitere, aber dies ist die einfachste). Nach dem chinesischen Restsatz ist also die Menge der Lösungen des modularen Gleichungssystems

$$\begin{cases} x \equiv 2 \mod 4 \\ x \equiv 0 \mod 3 \end{cases}$$

gegeben durch

$$\{4 \cdot 1 \cdot 0 + 3 \cdot (-1) \cdot 2 + 12k \,|\, k \in \mathbb{Z}\} \;=\; \{-6 + 12k \,|\, k \in \mathbb{Z}\}.$$

Diese Menge ist gleich der, die wir auch durch Ausprobieren bekommen haben, da $-6+12k = -6+12+12(k-1) = 6+12(k-1)$. Da k alle ganzzahlige Werte annimmt, stimmt die Lösungsmenge, die der chinesische Restsatz liefert, mit der vorher ermittelten Lösungsmenge überein.

Beispiel. Nun zu einem etwas anschaulicherem Beispiel: Ein Obsthändler ordnet Orangen an seinem Marktstand zunächst in 7er-Reihen an und stellt dabei fest, dass dabei 3 Orangen über bleiben. Nun ordnet er diese in 10er-Reihen an, und stellt fest, dass dabei noch 2 Orangen übrig bleiben. Er weiß mit Sicherheit, dass er weniger als 100 Orangen hat. Lässt sich aus diesen Daten die genaue Anzahl der Orangen ermitteln?

Die ersten beiden Angaben lassen sich in das folgende modulare Gleichungssystem übersetzen:

$$\begin{cases} x \equiv 3 \mod 7 \\ x \equiv 2 \mod 10. \end{cases}$$

Wir brauchen also ganze Zahlen u, v mit $7u + 10v = 1$. Solche lassen sich mit dem erweiterten euklidischen Algorithmus ermittelt, oder einfach raten, da die Zahlen doch relativ klein sind. Wir betrachten in diesem Beispiel zwei unterschiedliche Paaren solcher Zahlen und vergewissern uns, dass in diesem Fall die Lösungsmenge, die wir aus den beiden Darstellungen erhalten, dieselbe ist. Wir haben also

$$7 \cdot 3 + 10 \cdot (-2) = 1 \text{ und } 7 \cdot (-7) + 10 \cdot 5 = 1.$$

Aus der ersten Darstellung erhalten wir als Lösungsmenge

$$\{7 \cdot 3 \cdot 2 + 10 \cdot (-2) \cdot 3 + 70k \,|\, k \in \mathbb{Z}\} = \{-18 + 70k \,|\, k \in \mathbb{Z}\}.$$

Aus der zweiten Darstellung erhalten wir die Lösungsmenge

$$7 \cdot (-7) \cdot 2 + 10 \cdot 5 \cdot 3 + 70l \,|\, l \in \mathbb{Z}\} = \{52 + 70l \,|\, l \in \mathbb{Z}\}.$$

Da $-18 + 70 = 52$ gilt, sind beide Mengen gleich.

Nun zu der eigentlichen Frage: Da $52 + 70 > 100$ ist und $-18 < 0$ keine mögliche Anzahl von Orangen ist, ist 52 die eindeutige Lösung der ursprünglichen Aufgabe.

Es sei noch angemerkt, dass der chinesische Restsatz benutzt werden kann, um Rechnungen mit sehr großen natürlichen Zahlen durchzuführen - man führt die Rechnung modulo $n_1, \ldots, n_k$ für geeignete Zahlen n_i durch, wobei man dann maximal mit Zahlen der Größe $n_1, n_2, \ldots, n_k$ zu rechnen hat, und setzt das Ergebnis dann mit Hilfe des chinesischen Restsatzes wieder zusammen.

Teil III
Graphentheorie

Zum Schluss wollen wir einige Grundbegriffe der Graphentheorie behandeln. Die Geburtsstunde der Graphentheorie lag im 18. Jahrhundert, und das erste Problem der Graphentheorie wollen wir in einigem Detail behandeln.

Heutzutage dienen Graphen zur Visualisierung vieler Strukturen, etwa von Nahverkehrsnetzen oder allgemeiner von Verkehrsnetzen, insbesondere im Kontext der Routenoptimierung. Die Arbeitsweise vieler Navigationssysteme basiert etwa auf dem Dijkstra-Algorithmus für Graphen, der allerdings üblicherweise in der praktischen Informatik behandelt wird. Graphen werden aber auch eingesetzt, um kompliziert vernetzte Strukturen wie z.B. das Internet zu untersuchen, oder um Datenstrukturen oder Suchalgorithmen darzustellen und zu untersuchen. Insbesondere sind sogenannte Bäume, die wir etwas eingehender betrachten werden, ein Spezialfall von Graphen und diese werden etwa in Form von binären Bäumen in der Informatik verwendet. Auch Stammbäume können als Beispiele solcher Bäume betrachtet werden.

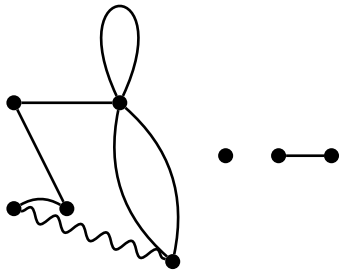
Die Namensgleichheit mit dem Graphen einer Funktion, den man aus der Schule kennt, ist übrigens „zufällig“: Zwar geht die Bezeichnung auf dasselbe griechische Wort zurück, aber es handelt sich um erst einmal unverwandte mathematische Objekte.

15 Grundbegriffe der Graphentheorie

Es sei angemerkt, dass es viele verschiedene Arten und Weisen gibt, Graphen zu behandeln, und dass Graphen in der Literatur in vielen unterschiedlichen Varianten vorkommen. Das ist insbesondere deshalb sinnvoll, weil unterschiedliche Varianten für unterschiedliche Zwecke am besten geeignet sind, insbesondere bei Modellierung von unterschiedlichen realen Sachverhalten.

Die Anschauung hinter einem Graphen ist in etwa wie folgt: Man hat eine Ansammlung von *Knoten*, die etwa die U-Bahn-Stationen, oder einzelne Internet-Nutzer oder Webseiten, oder Paketzentren symbolisieren, und dazwischen hat man Verbindungen, meist *Kanten* genannt, die z.B. die U-Bahn-Linien, Straßen, Autobahnen oder Zugriffe von einer Webseite auf die nächste veranschaulichen. Meist zeichnet man einen solchen Graphen in die Ebene, in dem man Knoten als besondere Punkte in der Ebene hervorhebt, und Kanten als Verbindungslinien einzeichnet, etwa so:

oder so:



Es ist eine Konventionsfrage, ob man *Mehrfachkanten* oder *Schleifen* erlaubt, also Verbindungen wie im folgenden Bild:

Wir wollen zuerst in unseren Betrachtungen sowohl Mehrfachkanten, also mehrere Verbindungen von einem Knoten zum anderen (etwa verschiedene U-Bahn-Linien zwischen gleichen U-Bahn-Stationen) als auch Schlefen (etwa Operationen modellierend, die den Zustand nicht ändern) erlauben.

Ein Graph besteht also aus:

- einer Ansammlung von Punkten (Knoten, manchmal sagt man dazu auch synonym „Ecken“, engl. vertex/vertices),
- einer Ansammlung von Verbindungslinien zwischen zwei Ecken bzw. von einer Ecke zu sich selbst (Kanten, engl. edges).

Wir wollen das mathematisch etwas präziser fassen und geben die folgende Definition.

Definition 15.1. Ein **Graph** Γ besteht aus einer Menge $V(\Gamma)$ von Ecken, einer Menge $E(\Gamma)$ von Kanten und einer Abbildung

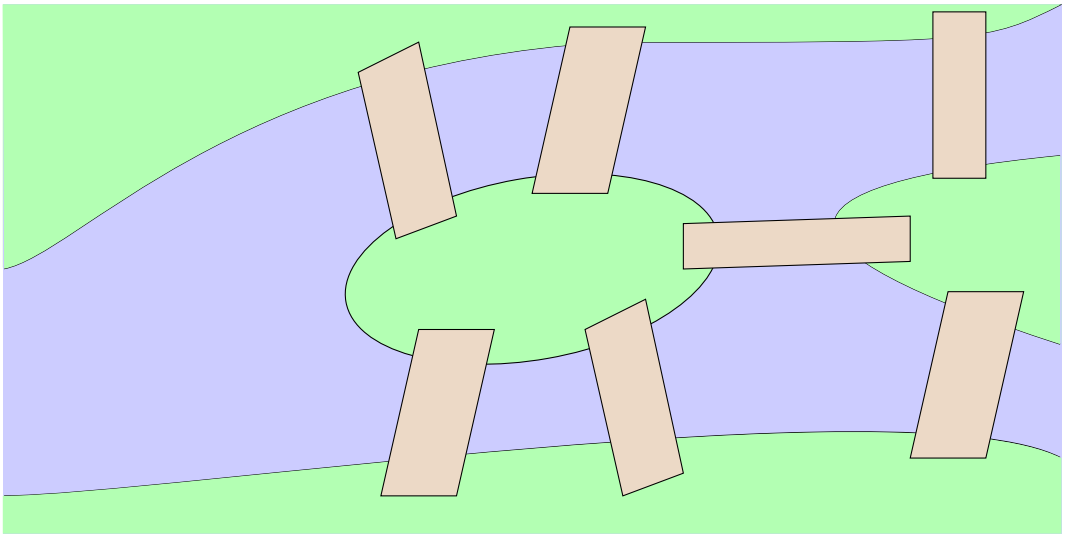
$$d_\Gamma \colon E(\Gamma) \to \{X \subseteq V(\Gamma) \mid 1 \leq |X| \leq 2\}.$$

Bemerkung.

1. Wir werden stets annehmen, dass sowohl die Menge $V(\Gamma)$ von Knoten eines Graphen Γ als auch die Menge $E(\Gamma)$ von Kanten endlich sind.
2. Die Benennung der Mengen leitet sich von den englischen Vokabeln für Knoten und Kanten ab.
3. Ist es aus dem Zusammenhang klar, um welchen Graphen es sich handelt, so werden wir abkürzend V anstatt $V(\Gamma)$ und E anstatt von $E(\Gamma)$ schreiben.

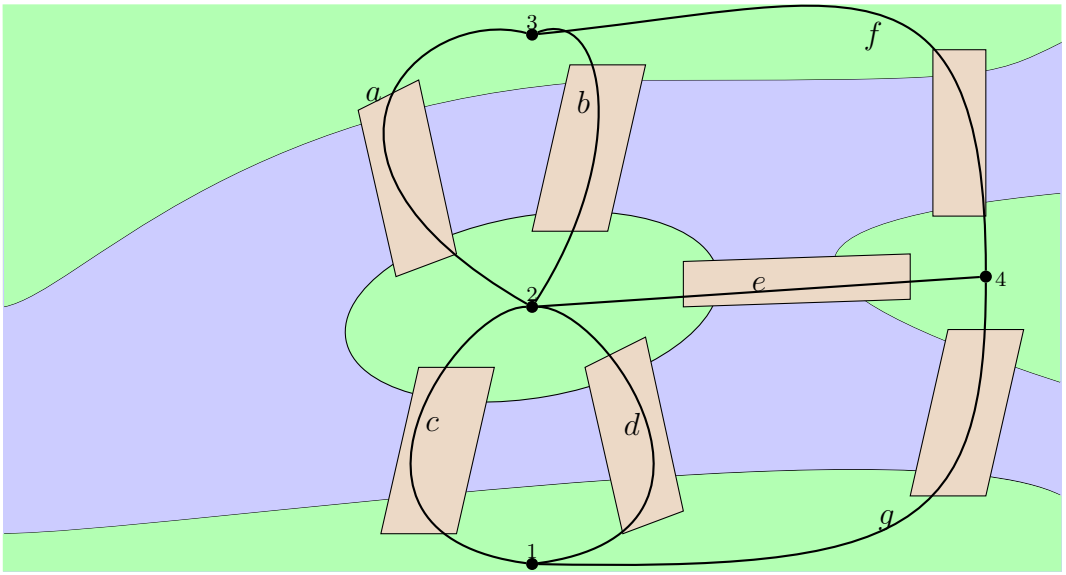
4. Die Abbildung d_Γ mag zuerst kompliziert erscheinen, ordnet aber einfach jeder Kanten die Knoten zu, die diese Kante verbindet. Da wir Schleifen erlauben, können es unserem Fall 1 oder 2 Knoten sein, die von einer Kante verbunden werden, daher die Bedingung $1 \leq |X| \leq 2$ in der Definition. Ferner haben unsere Kanten keine Richtung (d.h. sie sind gewissermaßen keine „Einbahnstraßen“), deswegen können wir keinen Startpunkt und keinen Zielpunkt einer Kante ausmachen, also haben wir nur ein Menge und kein geordnetes Paar von Eckpunkten.

Wir fangen nun mit dem sogenannten *Königsberger Brückenproblem* an. Anfang des 18. Jahrhunderts wurde einem berühmten Mathematiker dieser Zeit, Leonhard Euler, die Frage gestellt, ob es in der Stadt Königsberg (heute Kaliningrad) einen Spaziergang gibt, der über jede der damaligen 7 Brücken über den Fluss Pregel genau einmal führt. Dafür ist die gegenseitige Lage der Brücken sicherlich relevant, und diese kann schematisch wie folgt dargestellt werden:

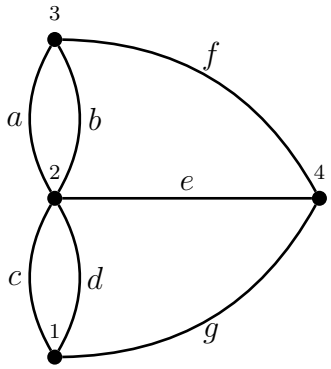


(Dabei sind die grünen Bereiche die unterschiedlichen Stadtteile und die hellbraunen die Brücken.)

Um das Problem mathematisch zu untersuchen, übersetzen wir dieses in die Sprache der Graphen. Dafür betrachten wir jeden Stadtteil als einen Knoten und jede Brücke als eine Kante. Um diese besser wiederfinden zu können, markieren wir die Stadtteile mit Zahlen $1, 2, 3, 4$ und Brücken mit Buchstaben a, b, c, d, e, f, g:



Vergessen wir dabei die Landschaft, so ergibt sich der folgende Graph:



Um die formale Definition eines Graphen besser zu verstehen, machen wir uns in diesem Beispiel klar, wie dieses Gebilde formal zu einem Graphen Γ wird:

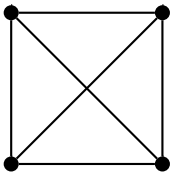
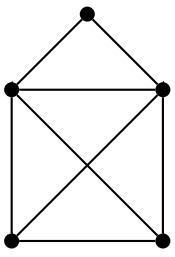
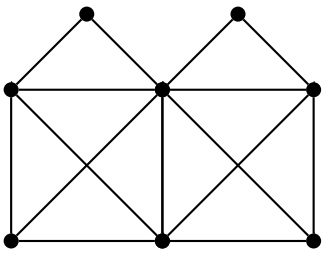
In diesem Fall haben wir $V(\Gamma) = \{1, 2, 3, 4\}$ und $E(\Gamma) = \{a, b, c, d, e, f, g\}$, und die Abbildung d_Γ, die jeder Kante die Menge ihrer Endpunkte zuordnet (also in diesem Fall die Abbildung, die jeder Brücke die Stadtteile zuordnet,

die diese Brücke verbindet), sieht in diesem Fall wie folgt aus:

$$\begin{aligned} d_\Gamma \colon E(\Gamma) &\to V(\Gamma) \\ a &\mapsto \{2,3\}, \\ b &\mapsto \{2,3\}, \\ c &\mapsto \{1,2\}, \\ d &\mapsto \{1,2\}, \\ e &\mapsto \{2,4\}, \\ f &\mapsto \{3,4\}, \\ g &\mapsto \{1,4\}. \end{aligned}$$

Bemerkung. Es stellt sich heraus, dass dieser Graph alle relevanten Daten von diesem Problem erfasst: Es ist nämlich für die Fragestellung nicht wichtig, ob die Stadtteile weit voneinander entfernt sind, ob die Brücken hübsch ist, ob die Brücken breit sind und so weiter. Das einzig Relevante für uns ist, welche Brücken welche Stadtteile verbinden, und diese Information ist in dem Graphen erfasst.

Das Königsberger Brückenproblem ist, vielleicht etwas überraschend, verwandt mit dem folgenden Problem: Welche der folgenden Figuren kann auf dem Papier gezeichnet werden, ohne den Stift vom Papier abzusetzen und ohne eine Kante doppelt zu zeichnen?

a)  b)  c) 

Durch etwas Ausprobieren kommt man schnell zu dem Vermutung, dass es im ersten Fall nicht geht, und findet auch eine Zeichenmöglichkeit für das Haus vom Nikolaus, aber bereits im dritten Fall ist die Antwort nicht so klar. Wir werden nach einem allgemeinen Kriterium suchen, wie man eine solche Frage beantworten kann. Dabei werden wir einige Begriffe, die in Graphentheorie relevant sind, kennenlernen.

16 Wege in Graphen

Zunächst wollen wir so was wie Spaziergänge, Wege und Rundwege in einem Graphen formalisieren.

Definition 16.1. Ein **Weg** in einem Graphen Γ ist eine Folge abwechselnd von Ecken und Kanten von Γ,

$$v_0, e_1, v_1, e_2, v_2, \ldots, v_{k-1}, e_k, v_k,$$

die mit einer Ecke anfängt und mit einer Ecke aufhört und $d_\Gamma(e_j) = \{v_{j-1}, v_j\}$ für alle $1 \leq j \leq k$ erfüllt. Fängt die Folge bei v_0 an und hört mit v_k auf, so sagt man auch, der Weg führt *von* v_0 *nach* v_k.

Ein Weg heißt **geschlossen**, falls $v_0 = v_k$ gilt.

Bemerkung.

1. Man sollte sich die Folge in etwa so vorstellen: Man startet im Knoten v_0 und geht über die Kante e_1 zu dem Knoten v_1; dann geht man vom Knoten v_1 über die Kante e_2 zum Knoten v_2, und so notiert man weiter seinen Weg durch den Graphen, bis man am Ziel, nämlich im Knoten v_k angekommen ist.

2. Es gibt in der Literatur unterschiedliche Begriffe von Wegen in Graphen. Wir erlauben zunächst, dass sowohl Kanten als auch Knoten mehrfach durchlaufen werden.

Da der formale Begriff eines Weges zunächst etwas abstrakt erscheinen mag, verdeutlichen wir ihn durch Beispiele.

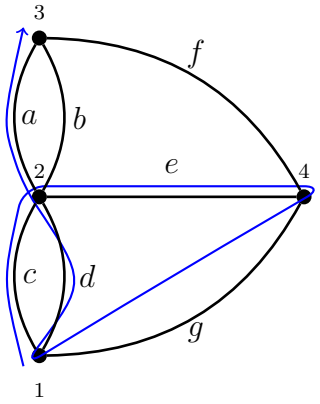
Beispiel.

1. Ein Spaziergang durch Königsberg kann etwa wie folgt aussehen: Man startet im Stadtteil 1, geht über die Brücke c auf die Insel 2, dann über die Brücke e in den östlichen Stadtteil 4, dann kehrt man über die Brücke g in den Stadtteil 1, geht diesmal über die Brücke d auf die Insel 2 und von da über die Brücke a in den nördlichen Stadtteil 3.

   Formal würden wir diesen Weg also als

   $$1, c, 2, e, 4, g, 1, d, 2, a, 3$$

   aufschreiben. In dem Graphen sieht der Weg wie folgt aus:



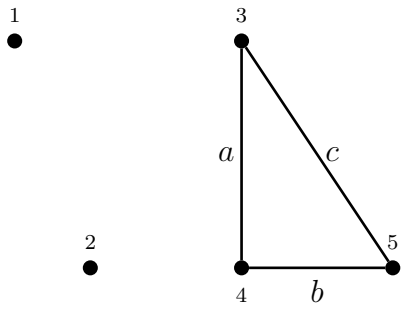
Das ist ein Weg von 1 nach 3, also insbesondere kein geschlossener Weg. (Bei einem Spaziergang würde es etwa dem entsprechen, dass man von Zuhause losgeht, aber am Ende nicht zurückgeht, sondern beispielsweise in ein Café.)

2. Ein anderer Weg in diesem Graphen ist etwa durch $1, c, 2, c, 1$ gegeben. (Man geht über die Brücke c, merkt danach, dass man etwas zuhause vergessen hat, und geht gleich auf demselben Wege wieder zurück - ein recht langweiliger Spaziergang.) Dieser Weg ist geschlossen, da der Anfang und das Ende übereinstimmen.

Wir wollen nun eine Eigenschaft von Graphen einführen, die sich mit Existenz von Wegen beschäftigt.

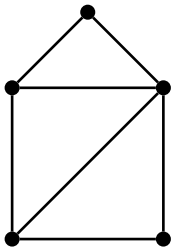
Definition 16.2. Ein Graph heißt **zusammenhängend**, falls je zwei unterschiedliche Ecken in diesem Graphen durch einen Weg verbunden werden können.

Beispiel. 1. Der folgende Graph ist nicht zusammenhängend:



Es gibt keine Möglichkeit, von dem Knoten 1 zu irgendeinem anderen Knoten über einen Weg zu kommen, da keine Kante ein Ende in 1 hat. Genausowenig kann man zum Knoten 2 kommen, oder diesen verlassen.

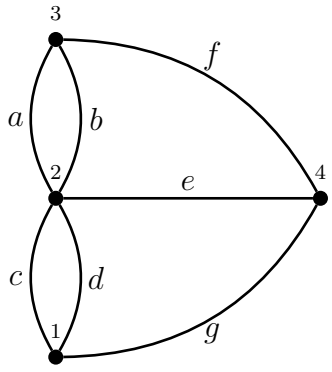
2. Der folgende Graph ist zusammenhängend:



Man kann nachprüfen, dass man je zwei Ecken durch einen Weg verbinden kann. Wir wollen ein solches Argument im Detail im nächsten Beispiel ausführen.

3. Der Graph, der zu dem Königsberger Brückenproblem gehört, ist zusammenhängend. Hier könn man sich einen Teil der Arbeit bei der Überprüfung dieser Eigenschaft ersparen, wenn man z.B. merkt, dass ein Graph, in dem eine bestimmte Ecke X mit jeder anderen Ecke durch einen Weg verbunden werden kann: Will man in diesem Graphen von A nach B gehen, so sucht man sich einen Weg von A nach X und einen von X nach B aus, die beide nach Voraussetzung existieren. Dabei haben wir eine weitere wichtige Beobachtung über Wege in Graphen bereits verwendet: Gibt es in einem Graphen einen Weg von C nach D, so kann man diesen rückwärts gehen und erhält einen Weg von D nach C.

   Zur Erinnerung nochmal der Graph:



   Wir wollen hier einmal zur Veranschaulichung tatsächlich einen Weg zwischen je zwei unterschiedlichen Ecken angeben. Es sei noch angemerkt, dass es im Allgemeinen auch mehrere Wege zwischen den Ecken geben kann und wir uns für einen beliebigen entscheiden können: Es kommt uns nur auf die Existenz mindestens eines Weges an.

   - Weg von 1 nach 2: $1, c, 2$.
   - Weg von 1 nach 3: $1, c, 2, a, 3$.

- Weg von 1 nach 4: $1, c, 2, b, 3, f, 4$. (Hier hätte es auch den kürzeren Weg $1, g, 4$ gegeben, aber in diesem Beispiel suchen wir nicht nach kürzesten Wegen.)
- Weg von 2 nach 1: $2, d, 1$.
- Weg von 2 nach 3: $2, b, 3$.
- Weg von 2 nach 4: $2, e, 4$.
- Weg von 3 nach 1: $3, b, 2, d, 1$.
- Weg von 3 nach 2: $3, a, 2$.
- Weg von 3 nach 4: $3, a, 2, e, 4$.
- Weg von 4 nach 1: $4, g, 1$.
- Weg von 4 nach 2: $4, g, 1, g, 4, e, 2$.
- Weg von 4 nach 3: $4, f, 3$.

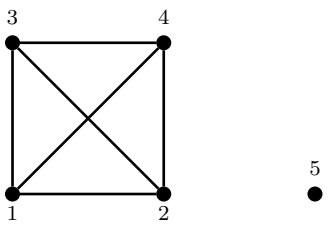
Wir wollen nun untersuchen, wann ein Graph zusammenhängend ist. Dafür werden wir ein *notwendiges* Kriterium angeben, also eine Eigenschaft, die jeder zusammenhängende Graph hat - ohne dass umgekehrt jeder Graph mit dieser Eigenschaft zusammenhängend sein muss.

Satz 16.3. *Ein zusammenhängender Graph mit n Knoten muss mindestens $n-1$ Kanten haben.*

Bevor wir den Beweis des Satzes behandeln, betrachten wir einige Beispiele.

Beispiel. 1. Der folgende Graph hat 5 Knoten und ist zusammenhängend. Nach dem Satz muss er also mindestens 4 Kanten haben, und wir sehen, dass in diesem Beispiel tatsächlich die Minimalanzahl realisiert wird: Dieser Graph hat genau 4 Kanten.

2. Dieser Satz stellt eine *notwenidge*, aber keine *hinreichende* Bedingung dar: Das heißt, dass die Bedingung „der Graph mit n Ecken hat mindestens $n-1$ Kanten“ erfüllt sein muss, damit der Graph zusammenhängend sein kann, aber nicht jeder Graph mit n Ecken und mindestens $n-1$ Kanten ist zusammenhängend. Der folgende Graph hat beispielsweise 5 Knoten und sogar $6 > 4$ Kanten, aber ist trotzdem nicht zusammenhängend.



Wir führen erneut keinen exakten Beweis von diesem Satz, sondern befassen uns nur mit der Beweisidee, die auf einem Algorithmus auf Graphen, der sogenannten *Breitensuche*, basiert.

Beweisidee. Der Breitensuche-Algorithmus (engl. breadth-first-search) liefert in einem beliebigen Graphen zu der vorgegebenen Startecke s alle Ecken, die von s aus durch einen Weg in diesem Graphen erreichbar sind.

Dabei geht man wie folgt vor:

1. Markiere die Startecke s.
2. Markiere alle Knoten, die benachbart sind (d.h. durch eine Kante verbunden) zu den Knoten, die in der vorherigen Ausführung markiert wurden.
3. Sobald in einem Schritt keine Knoten mehr markiert wurden: STOPP.

Alle Knoten, die wir erreichen, sind offenbar mit dem Startknoten s durch einen Weg verbunden. Ferner erreichen wir im 2-ten Schritt alle Ecken, die durch höchstens eine Kante mit s verbunden sind, im dritten Schritt alle Ecken, die durch Wege aus höchstens 2 Kanten mit s verbunden sind, im 4-ten Schritt sind es alle Knoten, die mit s durch einen Weg aus höchstens 3 Kanten verbunden sind, und so weiter. Da jede andere Ecke in einem zusammenhängenden Graphen durch einen Weg bestimmter Länge, etwa der Länge k, mit s verbunden ist, wird diese Ecke spätestens im $(k+1)$-ten Schritt erreicht.

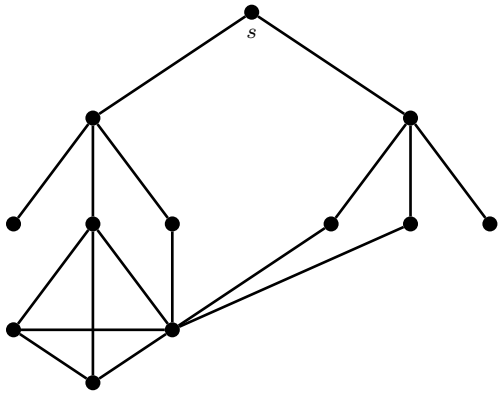
Nun wird für jeden neuen Knoten, den man markiert, eine neue Kante „benutzt“, da man nie zurückgeht. Da wir uns gerade überlegt haben, dass man von s aus jeden der anderen $n-1$ Knoten in dem Graphen erreichen wird, werden dabei auch $n-1$ Kanten „verbraucht“. Also muss der Graph mindestens $n-1$ Kanten haben.

Natürlich muss man, um den Beweis zu vervollständigen, noch Einiges präzisieren, allerdings bildet das die Grundidee für den Beweis dieses Satzes. □

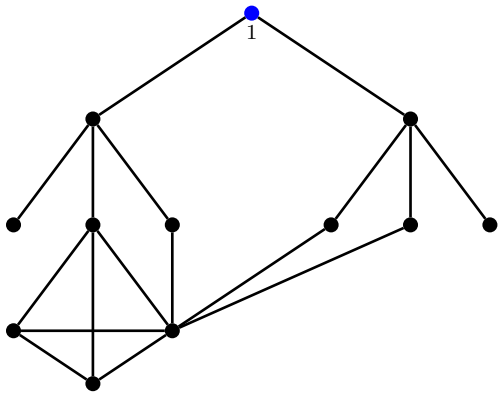
Übrigens heißt dieser Algorithmus Breitensuche, da man gleich „die ganze Breite des Graphen“ auf einmal durchsucht, also in alle möglichen Richtungen auf einmal. (Dem steht eine Tiefensuche entgegen, bei der man zunächst eine Richtung so weit wie möglich verfolgt.)

Wir veranschaulichen die Breitensuche nochmal an einem Beispiel.

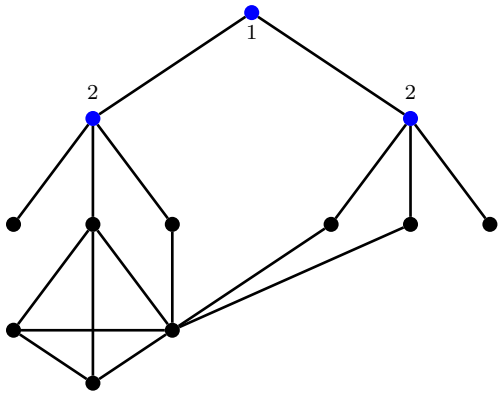
Beispiel. Wir betrachten den Graphen



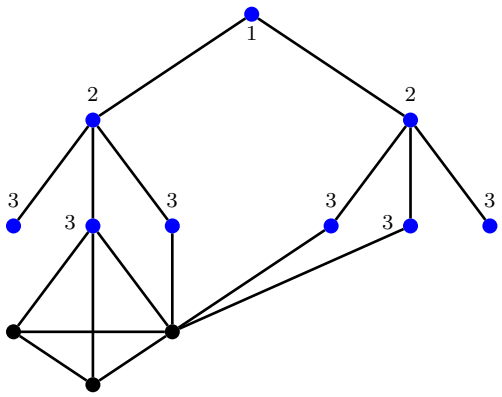
Im ersten Schritt ist nur der Startknoten markiert:



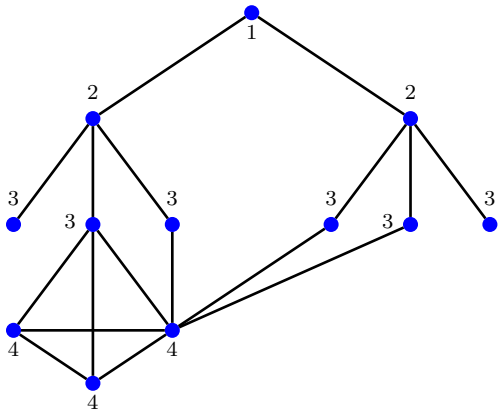
Im zweiten Schritt werden alle Knoten markiert, die mit s durch eine Kante verbunden sind. In diesem Fall sind es zwei weitere Knoten.



Im dritten Schritt kommen nun alle Knoten hinzu, die mit einem der beiden Knoten aus dem letzten Schritt durch eine Kante verbunden sind:



Schließlich markiert man im vierten und vorletzten Schritt alle übrigen Knoten (auch wenn die Ebenendarstellung etwas anderes suggeriert, ist jedoch auch der unterste Knoten mit einem der bereits markierten Knoten verbunden).



Im fünften Schritt werden nun keine weitere Knoten markiert, danach ist die Breitensuche zuende.

17 Bäume

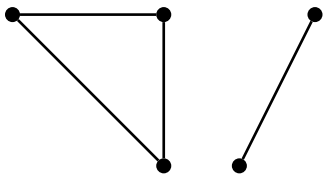
Als nächstes untersuchen wir eine spezielle Form von Graphen, die sogenannten Bäume. Wie schon erwähnt, werden Bäume in unterschiedlichen Formen in der Informatik genutzt. Wir fangen mit einer Definition an, die zunächst die Allgemeinheit der Graphen, die wir betrachten, etwas einschränkt.

Definition 17.1. Ein Graph Γ heißt **einfach**, wenn er weder Mehrfachkanten noch Schleifen besitzt. Formaler heißt das: Die Abbildung

$$d_\Gamma \colon E(\Gamma) \to \{X \subset V(\Gamma) \,|\, 1 \leq |X| \leq 2\}$$

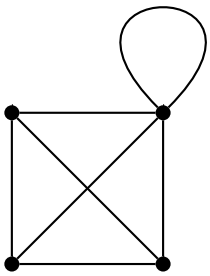
ist injektiv (das entspricht dem Fehlen der Mehrfachkanten) und es gilt $|d_\Gamma(e)| = 2$ für alle Kanten $e \in E(\Gamma)$ (diese Bedingung schließt Schleifen aus).

Beispiel. 1. Der folgende Graph ist einfach, da er weder Mehrfachkanten noch Schleifen besitzt:



2. Der folgende Graph ist nicht einfach, da er Mehrfachkanten besitzt:

3. Der folgende Graph ist nicht einfach, da er eine Schleife besitzt:



Um einen Baum zu definieren, brauchen wir den Begriff eines Kreises in einem Graphen. Dabei handelt es sich um Rundwege, wobei wir die langweiligen Rundwege, bei denen wir auf demselben Weg hin- und zurücklaufen, ausschließen wollen.

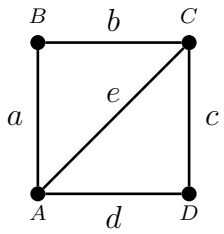
Definition 17.2. Ein **Kreis** in einem Graphen Γ ist ein geschlossener Weg

$$v_0, e_1, v_1, e_2, \ldots, e_k, v_k, e_{k+1}, v_{k+1} = v_0,$$

in dem keine Kante zweimal vorkommt.

Wir wollen uns einige Beispiele dazu anschauen.

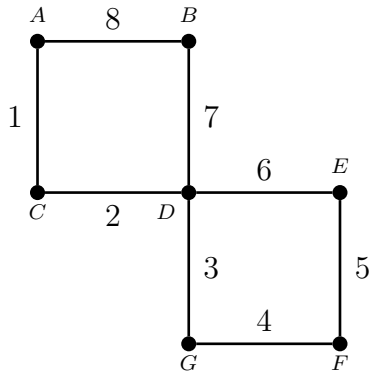
Beispiel. 1. Im Graphen



ist beispielsweise der Weg A, a, B, b, C, e, A ein Kreis, also ein geschlossener Weg, in dem keine Kante zweimal vorkommt. Ein weiterer Kreis in diesem Graphen ist etwa A, d, D, c, C, e, A, oder auch

$$D, d, A, a, B, b, C, c, D.$$

2. In der Literatur wird manchmal auch verlangt, dass die Startecke von einem Kreis die einzige ist, die zweimal in diesem Kreis vorkommt; wir werden diese Einschränkung nicht brauchen, und zwar aus einem Grund, der in dem folgenden Beispiel deutlich werden sollte. Wir werden uns vor allem für die Existenz der Kreise interessieren. Hat man einen Kreis in unserem Sinne in einem einfachen Graphen, so hat man auch einen Kreis, in dem keine Ecke zweimal vorkommt. Wir wollen das zwar nicht beweisen, aber am folgenden Beispiel veranschaulichen.

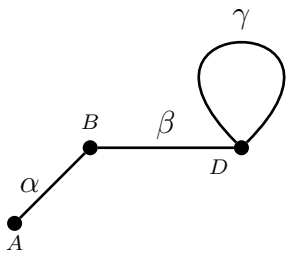


Wir betrachten den Weg

$$A, 1, C, 2, D, 3, G, 4, F, 5, E, 6, D, 7, B, 8, A,$$

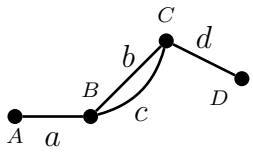
der nach unserer Definition ein Kreis ist, aber die Ecke D zweimal durchläuft. Allerdings haben wir darin einen kürzeren Weg, der bereits ein Kreis ist und keine Ecke doppelt durchläuft (außer dem Startpunkt), zum Beispiel den Weg $D, 3, G, 4, F, 5, E, 6, D$. Hat man allgemeiner einen Kreis, in dem eine Ecke, die nicht Startecke ist, mehrfach vorkommt, so kann man wie hier einen Teil von diesem Kreis zu einem kleineren Kreis machen; deswegen ist es für die Existenz von Kreisen unerheblich, welche Definition man verwendet.

3. In dem Graphen



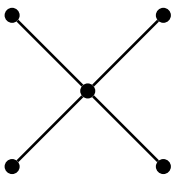
ist der Weg C, γ, C ein Kreis. Das ist der einzige Kreis in diesem Graphen.

4. In dem Graphen



ist B, b, C, c, B ein Kreis. Es ist wieder Konventionsfrage, ob dieser Kreis von dem Kreis B, c, C, b, B unterschieden wird. Dieser Kreis (bzw. diese Kreise) ist der einzige Kreis in diesem Graphen.

5. In dem Graphen



gibt es keine Kreise.

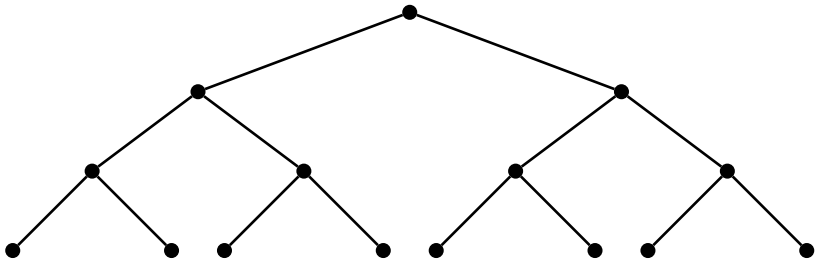
Um die Verallgemeinerung des letzten Beispiels geht es in der folgenden Definition.

Definition 17.3. Ein **Baum** ist ein zusammenhängender Graph Γ, in dem es keine Kreise gibt.

Wie schon angedeutet, werden Bäume in der Informatik etwa für gewisse Datenstrukturen oder für einige Suchalgorithmen genutzt.

Wir wollen nun einige weitere Beispiele von Bäumen geben.

Beispiel. 1. Der Graph

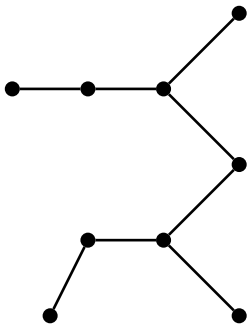


ist ein Baum. Ein solcher Baum wird in der Informatik „ausbalanciert“ genannt.

2. Der Graph

ist ein Baum.

3. Der Graph



ist ebenfalls ein Baum.

Bemerkung. Ein Baum ist immer ein einfacher Graph, d.h. es gibt in einem Graphen weder Mehrfachkanten noch Schleifen. Natürlich ist nicht jeder einfache Graph ein Baum.

Die Definition eines Baums ist zwar nicht schwer zu verstehen, allerdings etwas schwerer nachzuprüfen und für manche Kontexte nicht perfekt geeignet. Deswegen werden wir nun den Beweis des folgenden Satzes skizzieren, der uns drei weitere, äquivalente Bedingungen dafür liefert, wann ein zusammenhängender Graph ein Baum ist.

Satz 17.4. *Sei Γ ein zusammenhängender Graph mit genau n Knoten. Dann sind die folgenden Aussagen äquivalent:*

A: Γ *ist ein Baum.*

B: Γ *hat genau $n-1$ Kanten.*

C: Entfernt man eine beliebige Kante aus Γ, so ist der übrige Graph nicht mehr zusammenhängend.

D: Zwischen je zwei Knoten von Γ gibt es genau einen Weg, der keine Kante doppelt benutzt.

Wir wollen den Beweis dieses Satzes skizzieren.

Beweisskizze. Wir müssen die Äquivalenz der vier vorgegebenen Aussagen zeigen. Allerdings reicht es, anstatt für je zwei dieser Aussagen zu zeigen, dass diese äquivalent sind, wenn wir nur die folgenden vier Implikationen beweisen, denn alle andere Implikationen folgen dann daraus:

1) $A \Rightarrow B$,

2) $B \Rightarrow C$,

3) $C \Rightarrow D$,

4) $D \Rightarrow A$.

Kombiniert man 2) und 3), so sieht man zum Beispiel, dass die Aussage B die Aussage D impliziert. Kombiniert man 4) und 1), so sieht man, dass umkehrt auch B aus D folgt.

Nun fangen wir mit dem Beweis der Implikation 1) an. Sei also Γ ein Baum. Wir wollen zeigen, dass Γ genau $n-1$ Kanten hat. Da Γ zusammenhängend ist, wissen wir, dass Γ mindestens $n-1$ Kanten haben muss. Um zu sehen, dass es genau $n-1$ Kanten sein müssen, erinnern wir uns erneut an die Breitensuche, die wir im Beweis des Satzes 16.3 benutzt haben. Nun markieren wir zusätzlich für jede Ecke außer der Startecke, die wir erreichen, genau eine Kante, über die wir diese Ecke erreichen. Insgesamt haben wir also $n-1$ Kanten markiert, um zu jeder der $n-1$ Ecken, die nicht die Startecke ist, zu kommen. Insbesondere ist es mit den markierten Kanten möglich, von der Startecke zu jeder anderen Ecke zu Kommen. Wäre nun eine weitere Kante außer dieser $n-1$ Kanten von der Ecke A zur Ecke B vorhanden, so könnte

man diese Kante laufen, dann entlang der in der Breitensuche markierten Kanten von B zur Startecke, dann wieder entlang der markierten Kanten zu A und hätte damit einen Kreis in dem Graphen Γ. Das ist unser Widerspruch, da Γ ein Baum ist. Somit ist das wesentliche Argument für 1) erbracht.

Nun kommen wir zum Beweis von 2). Jetzt nehmen wir an, dass unser zusammenhängender Graph Γ genau n Ecken und genau $n-1$ Kanten hat. Wir müssen zeigen: Entfernt man eine beliebige Kante aus Γ, so ist der übrige Graph nicht mehr zusammenhängend. Hat man eine Kante entfernt, so hat man einen Graphen mit n Knoten und $n-2$ Kanten, und dieser kann nach Satz 16.3 unmöglich zusammenhängend sein. Folglich ist 2) bewiesen.

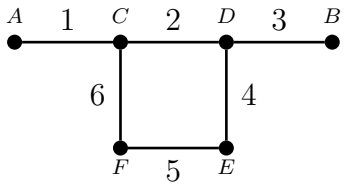
Als nächstes beweisen wir die Implikation 3). Wir nehmen also an, dass das Entfernen einer beliebigen Kante aus unserem zusammenhängenden Graphen Γ einen unzusammenhängenden Graphen macht. Wir müssen zeigen, dass es zwischen je zwei Knoten in diesem Graphen genau einen Weg gibt, der sie verbindet und dabei keine Kante doppelt benutzt. Dass es mindestens einen solchen Weg geben muss, wissen wir, da der Graph zusammenhängend ist. Das garantiert zunächst nur die Existenz eines beliebigen Weges, aber ähnlich wie in dem Beispiel zu der Definition eines Kreises kann man die Mehrfachnutzungen der Kanten eliminieren. Wird nämlich eine Kante doppelt durchlaufen, so kann man alle Schritte zwischen dem ersten und dem letzten Durchlaufen dieser Kante überspringen. (Das wird im Beispiel nach dem Beweis nochmal verdeutlicht.) Wir müssen also nur noch zeigen, dass ein solcher Weg eindeutig ist. Gäbe es nun zwei Wege zwischen zwei Knoten v_0 und v_1, von denen jeder keine Kante doppelt durchläuft, und die unterschiedlich sind, so müsste es ja eine Kante geben, die nur im ersten und nicht in dem zweiten Weg enthalten ist. Entfernt man diese kannte, so kann man sich überlegen, dass der Graph nach wie vor zusammenhängend sein müsste. Es im Detail zu zeigen, ist etwas aufwändig, aber wir sehen sofort, dass v_0 und v_1 nach wie vor durch einen Weg verbunden sind. Für jedes andere Paar von Ecken sucht man eine Verbindung zu v_0 bzw. v_1 und erhält dadurch einen Weg zwischen diesen Knoten. (An dieser Stelle müssten einige Einzelheiten überprüft werden, um den Beweis zu vervollständigen.) Das würde den Widerspruch liefern, da wir angenommen haben, dass das Entfernen einer beliebigen Kante den Graphen unzusammenhängend macht. Das liefert die Grundidee zum Beweis der Implikation 3).

Als letztes beschäftigen wir uns mit dem Beweis der Implikation 4). Wir nehmen also an, dass in dem Graphen Γ zwischen je zwei Knoten genau ein Weg existiert, der keine Kante doppelt nutzt, und müssen nun beweisen, dass Γ ein Baum ist. Nach Definition eines Baums haben wir also zu zeigen, dass Γ keine Kreise besitzt, da Γ bereits nach Voraussetzung zusammenhängend ist. Wir führen erneut einen Widerspruchsbeweis. Angenommen also, wir hätten einen Kreis, in dem eine Kante e zwischen v_0 und v_1 vorkommt. Dann liefert v_0, e, v_1 einen Weg von v_0 nach v_1 in Γ. Läuft man allerdings den

Rest des Kreises von v_0 nach v_1, so entsteht ein anderer Weg von v_0 nach v_1, bei dem weder e vorkommt noch eine Kante doppelt vorkommt, nach unserer Definition eines Kreises. Das widerspricht aber gerade unserer Annahme, dass nur ein solcher Weg existieren darf. Also ist die Implikation 4) bewiesen.

□

Beispiel. Wir wollen nochmal erläutern, wie ein Mehrfachnutzungen der Kanten aus einem Weg in einem Graphen eliminiert werden können, sodass man einen neuen Weg zwischen denselben Endpunkten bekommt, der jetzt weniger Kanten mehrfach nutzt; in unserem Beispiel ist man gleich nach der ersten solchen Abkürzung fertig und erhält einen Weg ohne Mehrfachnutzung der Kanten bei gleichbleibenden Endpunkten.



In diesem Graphen betrachten wir den Weg

$$A, 1, C, 2, D, 4, E, 5, F, 6, C, 2, D, 3, B.$$

In diesem Weg kommt die Kante 2 zweimal vor. Nun haben wir im Beweis zuvor gesagt, dass wir in diesem Fall alle Schritte zwischen dem ersten und dem letzten Durchlaufen der Kante weglaseen können, in diesem Fall ist der Abschnitt

$$D, 4, E, 5, F, 6, C, 2, D$$

ein unnötiger Umweg. Kürzt man diesen ab, so enthält man den Weg

$$A, 1, C, 2, D, 3, B,$$

der nun keine Kante doppelt benutzt.

18 Grad einer Ecke in einem Graphen

In diesem Abschnitt wollen wir weitere Eigenschaften von Graphen untersuchen. Es stellt sich als hilfreich heraus, die Anzahl der Kanten zu untersuchen, die von einem Knoten weggehen. Das ist die Idee vom Grad einer Ecke. Wir müssen jedoch noch eine Kleinigkeit beachten, nämlich müssen wir gesondert berücksichtigen, ob es an der betrachteten Ecke Schleifen gibt. Wir fassen das zunächst etwas präziser in Worte und geben dann die formale Definition.

Definition 18.1. Der **Grad** einer Ecke v im Graphen Γ ist:

- Die Anzahl der Kanten, die ein Ende in v haben, falls es keine Schleifen an v gibt,
- Anzahl der Kanten, die ein Ende in v haben und keine Schleifen sind, plus zweimal die Anzahl der Schleifen an v.

Wir schreiben $\mathrm{grad}_\Gamma(v)$ oder nur $\mathrm{grad}(v)$ für Grad von v. (In der Literatur ist auch die Abkürzung $\deg(v)$ gebräuchlich, vom englischen „degree“.)

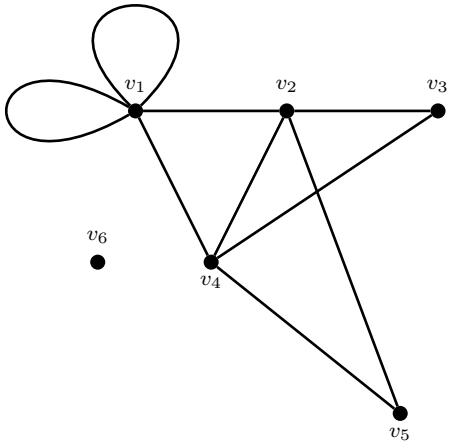
Formaler ausgedrückt wird der Grad von v wie folgt definiert:

$$\begin{aligned}\mathrm{grad}(v) &= |\{e \in E(\Gamma) \mid v \in d_\Gamma(e) \text{ und } |d_\Gamma(e)| = 2\}| \\ &+ \ 2 \cdot |\{e \in E(\Gamma) \mid d_\Gamma(e) = \{v\}\}|.\end{aligned}$$

Die Idee dabei ist, dass eine Schleife zwei Möglichkeiten liefert, von der Ecke, an der sie anliegt, wegzugehen, denn man kann die Schleife in zwei verschiedenen Richtungen durchlaufen.

Wir veranschaulichen uns die Definition an einem Beispiel.

Beispiel. Wir betrachten den folgenden Graphen.



In diesem Graphen gilt: $\mathrm{grad}(v_1) = 6$, da an v_1 zwei Schleifen dranliegen sowie die Kanten nach v_2 und v_4, also erhalten wir insgesamt $2 \cdot 2 + 2 =$

6 Möglichkeiten, v_1 zu verlassen. An allen anderen Ecken haben wir keine Schleifen, und wir brauchen somit nur die Kanten zu zählen, die ein Ende in der jeweiligen Ecke liegen. Die Ecke v_2 ist mit jeder anderen Ecke außer v_6 durch eine Kante verbunden, also ist $\mathrm{grad}(v_2) = 4$. Die Ecken v_3 und v_5 sind jeweils mit v_2 und v_4 verbunden, und somit gilt

$$\mathrm{grad}(v_3) = \mathrm{grad}(v_5) = 2.$$

Auch die Ecke v_4 ist mit jeder anderen Ecke außer v_6 verbunden und es gilt wieder $\mathrm{grad}(v_2) = 4$. Schließlich gibt es keine Kanten, die ein Ende in v_6 haben, also gilt $\mathrm{grad}(v_6) = 0$ - es gibt keinen Weg in unserem Graphen, der die Ecke v_6 verlässt.

Übrigens müssen wir nicht jeden Schnittpunkt zweier Kanten zu einer Ecke machen: Wir können, um einen Graphen zu definieren, einfach Knoten aussuchen und Verbindungslinien dazwischen als Kanten festlegen; hierbei spielt es in vielen Kontexten keine Rolle, ob sie sich schneiden.

Wir wollen nun eine Gradformel beweisen, die die Anzahl der Kanten eines Graphen mit den Graden seiner Ecken in Beziehung setzt.

Satz 18.2. *In jedem Graphen Γ gilt:*

$$\sum_{v \in V(\Gamma)} \mathrm{grad}(v) = 2 \cdot |E(\Gamma)|.$$

Hierbei bedeutet die linke Seite, dass man die Grade aller Knoten im Graphen Γ aufsummiert.

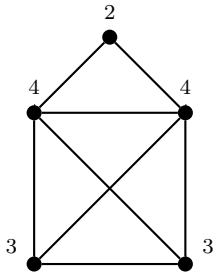
Beispiel.

- Im letzten Beispiel hatten 6 Knoten, und die Summe auf der linken Seite der obigen Aussage wird zu:

$$\begin{aligned} & \sum_{v \in V(\Gamma)} \mathrm{grad}(v) \\ = \ & \mathrm{grad}(v_1) + \mathrm{grad}(v_2) + \mathrm{grad}(v_3) + \mathrm{grad}(v_4) + \mathrm{grad}(v_5) + \mathrm{grad}(v_6) \\ = \ & 6 + 4 + 2 + 4 + 2 + 0 \\ = \ & 18. \end{aligned}$$

  Das passt genau damit zusammen, dass der betrachtete Graph 9 Kanten hat und $2 \cdot 9 = 18$ ist.

- Wir markieren die Knotengrade im Haus vom Nikolaus:



Die Summe der Knotengrade ist in diesem Fall $3+3+4+4+2=16$, und wir haben gerade $\frac{16}{2}=8$ Kanten.

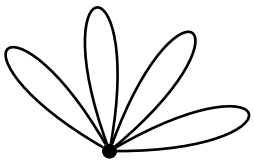
Nun geben wir die wesentliche Idee zum Beweis des obigen Satzes an.

Beweisidee. Jede Kante, die keine Schleife ist, hat genau 2 Enden, und trägt in der Summe auf der linken Seite jeweils eine 1 bei den Graden der beiden ihrer Enden bei. Jede Schleife wird dank unserer Definition des Grades in der Summe auf der linken Seite ebenfalls genau zweimal gezählt. Insgesamt wird für jede Kante im Graphen eine 2 auf der linken Seite addiert, und alle Summanden auf der linken Seite entstehen so. Das impliziert die Behauptung. □

Dieser Satz liefert gewisse Einschränkungen an die möglichen Grade der Ecken in einem Graphen. Um das besser zu verstehen, betrachten wir zunächst das folgende Beispiel.

Beispiel. Wir fragen uns, für welche natürliche Zahlen $n \geq 1$ es einen Graphen gibt, der genau n Knoten hat, jeder von welchen genau den Grad 3 hat.

Im Fall $\mathbf{n=1}$ haben wir eine einzige Ecke, und alle Kanten müssen somit Schleifen an dieser Ecke sein. Der Graph muss also in etwa wie folgt aussehen:

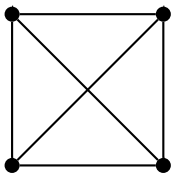


Da jede Schleife zum Grad des einzigen Knoten einen Summanden 2 beiträgt, ist der Grad dieses Knoten gerade die verdoppelte Anzahl der Schleifen und insbesondere stets eine gerade Zahl. Somit können also in einem Graph mit genau einem Knoten nicht alle Knoten den Grad 3 haben.

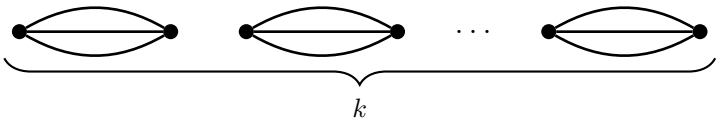
Im Fall $\mathbf{n=2}$ gibt es hingegen Beispiele von Graphen mit genau 2 Ecken, die beide Grad 3 haben:

Für $\mathbf{n=3}$ ist die Frage bereits nicht so klar, während man für $\mathbf{n=4}$ einfach zweimal den Graphen nehmen können, den wir schon eben hatten:

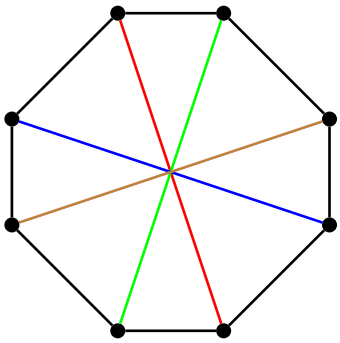
Will man zusätzlich fordern, dass der Graph zusammenhängend sein soll, so hätte man für $\mathbf{n = 4}$ ebenfalls einen Graphen mit den gewünschten Eigenschaften, nämlich



Allgemeiner sieht man nun, dass für eine gerade Anzahl der Ecken, also für $n = 2k$ mit $k \in \mathbb{N}$, ein Graph mit n Ecken jeweils vom Grad 3 erhalten werden kann, indem wir das erste Beispiel k mal wiederholen:

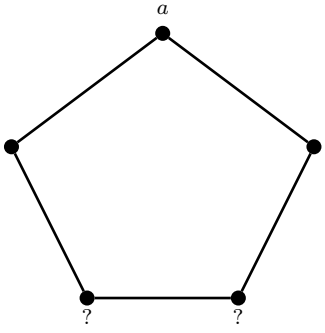


Will man erneut zusätzlich fordern, dass der Graph zusammenhängend ist, so muss man etwas geschickter vorgehen. In diesem Fall betrachten wir ein $2k$-Eck. Da wir eine gerade Anzahl von Ecken haben, besitzt jede Ecke eine eindeutige, ihr gegenüberliegende Ecke; und diese beiden verbinden wir jeweils durch eine Diagonale, etwa wie in dem folgenden Beispiel.



Man kann zeigen, dass man dadurch für jedes $k > 1$ einen Graph mit $2k$ Ecken vom Grad 3 jeweils erhalten kann.

Diese Methode ist aber etwa beim Fünfeck nicht anwendbar, weil es keine eindeutige Ecke gegenüber der vorgegebenen Ecke gibt:



Tatsächlich kommt man nach weiteren Versuchen zu der Vermutung, dass für ein ungerade Gesamtzahl von Knoten es allgemein unmöglich ist, dass jede Ecke genau Grad 3 hat. Das ist der Gegenstand des nächsten Korollars.

Korollar 18.3. *Die Anzahl der Ecken vom ungeraden Grad in einem Graphen ist stets gerade.*

Beweis. Wir erinnern uns daran, dass die *Parität* (also Gerade/Ungerade-Sein) einer Summe zweier Zahlen in der folgenden Art und Weise von der Parität der Summanden abhängt. Sind beide Summanden derselben Parität, also beide gerade oder beide ungerade, so ist die Summe gerade. Sind die Summanden hingegen von unterschiedlicher Parität, also einer gerade und einer ungerade, so wird das Ergebnis gerade sein. Addiert man insbesondere eine beliebige Anzahl von geraden Summanden auf, so ist das Ergebnis wieder eine gerade Zahl. Bei einer Summe von lauter ungeraden Zahlen kommt es hingegen auf die Anzahl der Summanden an: Hat man eine gerade Anzahl ungerader Summanden, so ist die Summe gerade; bei einer ungeraden Anzahl ungerader Summanden muss die Summe ebenfalls ungerade sein.

Nach dieser Vorüberlegung erinnern wir uns außerdem an die Formel

$$\sum_{v \in V(\Gamma)} \operatorname{grad}(v) = 2 \cdot |E(\Gamma)|.$$

für jeden Graphen Γ aus dem Satz 18.2, mit deren Hilfe wir gleich die Behauptung zeigen werden.

Da wir auf der rechten Seite der Formel das doppelte einer natürlichen Zahl haben, haben wir auf der rechten (und somit auch auf der linken Seite) eine gerade Zahl als Ergebnis. Die Summe aller Grade in Γ, von denen manche gerade und manche ungerade sein werden, muss also gerade sein. Die Summe der geraden Grade ist nach unserer Vorüberlegung ebenfalls gerade, also muss der verbleibende Teil der Summe - also die Summe der Grade der Ecken, die ungeraden Grad haben - ebenfalls gerade sein, damit die Summe insgesamt gerade ist. Nach unserer Vorüberlegung kann das allerdings nur der Fall sein, wenn die Anzahl der Ecken ungeraden Grades in unserem Graphen gerade ist. Also erhalten wir gerade die Behauptung des Korollars. □

Beispiel. Es kann also keinen Graphen mit einer ungeraden Anzahl von Knoten geben, in dem alle Knoten genau Grad 3 haben, da dies dem Korollar widersprechen würde.

19 Eulertouren und Eulerkreise in Graphen

Wir gehen nun zu der motivierenden Frage zurück, ob ein Spaziergang durch Königsberg möglich ist, der jede Brücke genau einmal benutzt. Um diese Fragestellung mathematisch besser erfassen zu können, führen wir die Begriffe einer Eulertour und eines Eulerkreises ein.

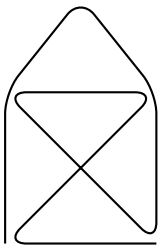
Definition 19.1. Ein Weg in einem Graphen Γ heißt **Eulertour**, falls er jede Kante von Γ genau einmal benutzt.

Ein Weg, der geschlossen ist und gleichzeitig eine Eulertour ist, wird **Eulerkreis** genannt.

Sowohl das Königsberger Brückenproblem als auch die Frage, ob eine Figur gezeichnet werden kann, ohne den Stift abzusetzen, können nun als die Frage nach der Existenz einer Eulertour in dem vorgegebenen Graphen angesehen werden.

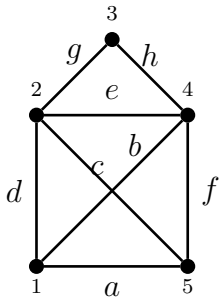
Beispiel.

- In dem Haus von Nikolaus existiert eine Eulertour, zum Beispiel können wir diesen wie folgt zeichnen, ohne den Stift abzusetzen und ohne eine Kante doppelt zu malen:



(Dieses Bild ist aus der dem schönen pgf-manual http://www.texample.net/media/pgf/builds/pgfmanualCVS2012-11-04.pdf übernommen.)

Will man das etwas formaler aufschreiben, so benennt man die Ecken und Kanten des Graphen zum Beispiel wie folgt:



Dann kann man den obigen Weg aufschreiben als

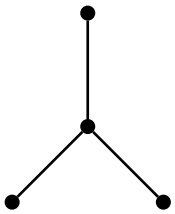
$$1, d, 2, g, 3, h, 4, f, 5, c, 2, e, 4, b, 1, a, 5.$$

Dies ist eine Eulertour in dem vorgegebenen Graphen, allerdings ist das kein Eulerkreis, da wir in einer anderen Ecke angefangen haben, als wir am Ende angekommen sind.

- Im Graphen

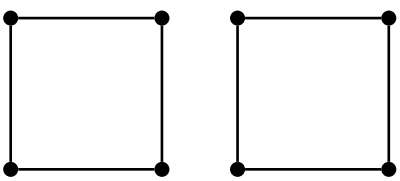
gibt es offenbar eine Eulertour, also einen Weg, der die einzige Kante genau einmal benutzt: Man geht zum Beispiel von der linken zu der rechten Ecke entlang der einzigen Kante. Allerdings ist auch unmittelbar klar, dass dieser Graph keinen Eulerkreis enthalten kann: Hat man die einzige Kante verbraucht, so kann man nicht mehr zurückkommen, um die ursprüngliche Ecke wieder zu erreichen.

- In dem Graphen



gibt es weder eine Eulertour noch einen Eulerkreis. Fängt man in einer der äußeren Ecken an, so muss man im ersten Schritt notwenderweise in die Mitte gehen und im zweiten Schritt zu einer anderen äußeren Ecke, die man allerdings nicht verlassen kann, da man die einzige Kante, die zu dieser Ecke führt, verbraucht hat. Würde man in der Mitte anfangen, so würde dieses Problem sogar gleich nach dem ersten Schritt auftreten. Folglich gibt es in diesem Graphen keine Eulertour.

- Auch in dem folgenden Graphen ist eine Eulertour nicht möglich:



Zwar ist in jedem der Teile sogar ein Eulerkreis möglich (man geht einfach einmal im Kreis), allerdings gibt es eine Möglichkeit, von der

rechten Hälfte des Graphen in die linke zu kommen, oder umgekehrt. Somit kann es keinen Weg in diesem Graphen geben, der jede Kante benutzt. Es veranschaulicht auch, warum wir uns in Zukunft beim Studium von Eulertouren auf zusammenhängende Graphen weitgehend einschränken werden.

Unser Ziel ist es nun, das folgende einfache Kriterium dafür zu verstehen, dass ein Graph eine Eulertour oder einen Eulerkreis besitzt.

Satz 19.2. *Sei* Γ *ein zusammenhängender Graph.*

1. *Es gibt genau dann einen Eulerkreis in* Γ*, wenn jede Ecke in* Γ *einen geraden Grad besitzt.*

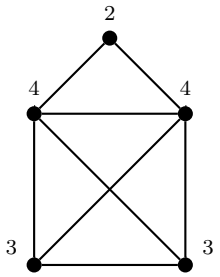
2. *Es gibt genau dann eine Eulertour in* Γ*, wenn es höchstens* 2 *Ecken vom ungeraden Grad in* Γ *gibt.*

Wir haben bereits festgestellt, dass ein Graph unmöglich genau eine Ecke vom ungeraden Grad besitzen kann. Somit kann man die letzte Bedingung wie folgt umformulieren: Es gibt genau dann eine Eulertour im zusammenhängenden Graphen Γ, wenn es in Γ entweder keine oder genau 2 Ecken vom ungeraden Grad gibt.

Bevor wir einige Beweisideen zu diesem Satz durchgehen, betrachten wir Beispiele zu diesem Satz.

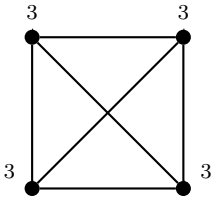
Beispiel.

- Wir betrachten zunächst wieder das Haus vom Nikolaus und markieren wieder die Knotengrade:



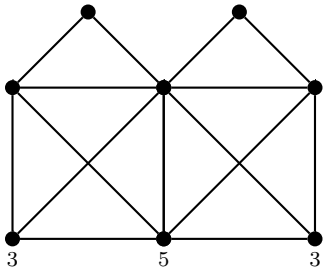
  Wir haben genau zwei Ecken vom ungeraden Grad, nämlich die beiden unteren Ecken. Unser Satz besagt, dass dieser Graph eine Eulertour besitzt (was wir allerdings schon wussten), aber keinen Eulerkreis. Zeichnet man also das Haus des Nikolaus ohne den Stift abzusetzen und ohne eine Kante doppelt zu zeichnen, so kann man dabei niemals in derselben Ecke aufhören, in der man angefangen hat.

- Betrachten wir den Graphen



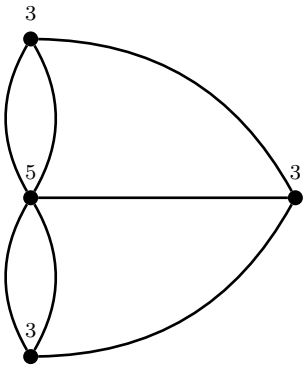
(hier sind wieder die Grade der Ecken notiert.) Dieser Graph hat 4 Ecken ungeraden Grades, und besitzt demzufolge keine Eulertour (und somit erst recht keinen Eulerkreis). Der Satz liefert uns also einen Beweis dafür, dass diese Figur nicht gezeichnet werden kann, ohne den Stift abzusetzen und ohne eine Kante doppelt zu zeichnen.

- Nun können wir auch entscheiden, ob das doppelte Haus vom Nikolaus in einem Zug gezeichnet werden kann. Wir notieren wieder einige Knotengrade:



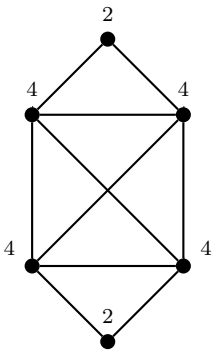
Hier sehen wir schon, dass wir mindestens drei Ecken vom ungeraden Grad in diesem Graphen haben, und somit keine Eulertour durch diesen Graphen existieren kann. Wir müssen also die restlichen Knotengrade nicht mehr ermitteln.

- In dem Graphen, der das Königsberger Brückenproblem beschreibt, notieren wir ebenfalls die Knotengrade:



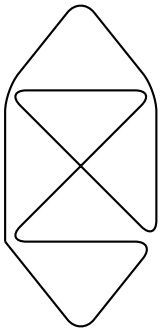
Da in diesem Graphen alle 4 Knoten ungeraden Grad haben, existiert laut dem Satz darin keine Eulertour. Also kann es keinen Spaziergang durch Königsberg geben, der jede Brück genau einmal benutzt - wie bereits Leonhard Euler im 18. Jahrhundert herausgefunden hat.

- Wir fügen zu dem Haus des Nikolaus einen Keller hinzu und ermitteln im neuen Graphen ebenfalls die Knotengrade:



In diesem Graphen sind alle Knotengrade gerade. Somit besagt der Satz, dass es darin einen Eulerkreis geben muss. Der Satz sagt allerdings nichts darüber aus, wie wir einen solchen Eulerkreis finden (teilweise wird das allerdings aus dem Beweis deutlich), aber er garantiert, dass es möglich ist.

Es ist übrigens nicht schwer, einen solchen Eulerkreis zu finden, etwa den im folgenden Bild angedeuteten:



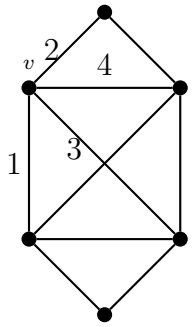
Wir geben nun einige der Beweisideen zum Satz 19.2 an.

Beweisideen zum Satz 19.2. Jeder der beiden Teile der Satzes ist eine „genau dann, wenn“-Aussage. Beide kann man also jeweils in zwei Teile aufteilen.

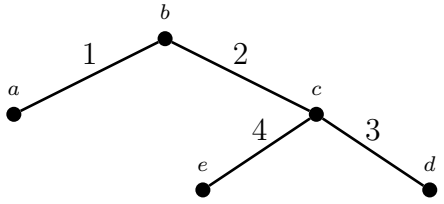
Wir beginnen mit einem Teil der ersten Aussage, nämlich damit, dass in einem Graphen, in dem ein Eulerkreis existiert, alle Knotengrade gerade sein müssen. Schauen wir uns eine Ecke v auf dem Eulerkreis an. Zunächst wollen wir annehmen, dass es nicht die Startecke ist, und auch, dass an

dieser Ecke keine Schleifen dran sind. Die Beweisidee beruht nun auf der folgenden, einleuchtenden Überlegung: Bei jedem Besuch in v „verbrauchen“ wir genau zwei Kanten, nämlich eine, um zu dieser Ecke zu kommen, und ein, um von dieser Ecke wieder wegzugehen. Das gilt bei jedem Besuch in dieser Ecke. Wir bezeichnen die Anzahl dieser Besuche mit b. Dann sieht man, dass man $2 \cdot b$ Kanten benutzt hat, die an der Ecke v anliegen. Da man jede Kante, die an dieser Ecke anliegt (genauso wie jede andere Kante) genau einmal auf unserem Rundweg genutzt hat, ist der Grad der Ecke v also $2b$ und somit gerade. Als nächstes bemerken wir, dass die Schleifen, die wir bis jetzt nicht berücksichtigt haben, immer eine gerade Zahl zu dem Grad einer Ecke beitragen und somit unsere Überlegung nicht stören (geht man nämlich entlang einer Kante von v weg, so kommt man sofort wieder an v an.) Schließlich bemerken wir, dass man in der Startecke zwar am Anfang schon steht, also am Anfang nur eine Kante verbraucht, um diese zu verlassen - allerdings bildet diese Kante am Ende ein Paar mit der Kante, mit der wir am Schluss an dieser Ecke wieder ankommen.

Um sich das nochmal zu veranschaulichen, betrachten wir das letzte Beispiel nochmal. Dabei verwenden wir wieder den Weg, den wir eben angegeben haben. Die Ecke v ist im Bild markiert, sowie die Reihenfolge, in der die an v anliegenden Kanten verwendet werden.



Um etwas besser zu verstehen, warum das funktioniert, betrachen wir zum Vergleich den folgenden Graphen, in dem, wie man leicht sieht, keine Eulertour existiert, und in dem wir stattdessen einen Weg wählen wollen, der jede Kante mindestens einmal benutzt:

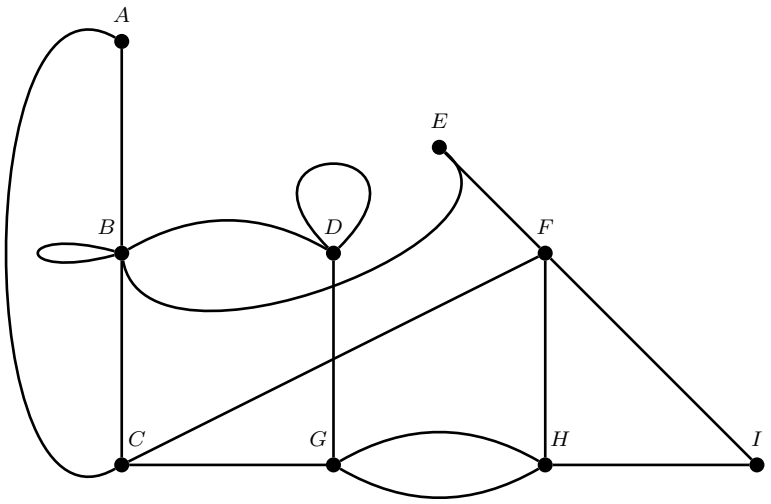


Wir haben also in diesem Graphen den Weg $a, 1, b, 2, c, 3, d, 3, c, 4, e$, der jede Kante mindestens einmal benutzt. Man kann sich also fragen, warum dieselbe Argumentation wie oben nicht funktioniert: Etwa in der Ecke c kommen wir zweimal vorbei auf unserem Weg, müssten da nicht 4 Kanten an c dranliegen?

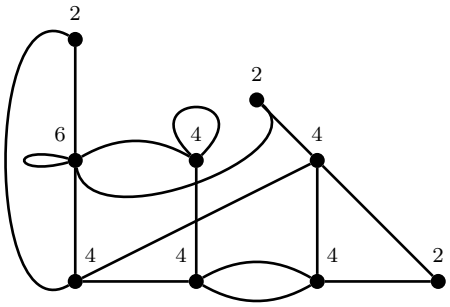
Betrachtet man den Weg genauer, so merkt man, dass es bei Eulertouren in entscheidender Art und Weise ausgenutzt worden war, dass jede Kante genau einmal benutzt wird. Nicht so in diesem Beispiel: Man braucht tatsächlich zwei Kanten für den Hinweg nach c und zwei, um c zu verlassen, allerdings nutzt man dabei die Kante 3 zweimal, und macht damit das Paritätsargument der vorherigen Überlegung zunichte.

Als nächstes wollen wir uns mit der Aussage beschäftigen, dass jeder zusammenhängende Graph mit nur geraden Knotengraden einen Eulerkreis besitzt. Diese Aussage ist schon in ihrer Natur schwieriger zu beweisen: Man muss einen Weg (mit gewissen Eigenschaften) in einem Graphen angeben, über den man nur sehr wenig Kenntnis besitzt. Wir werden, vor allem an einem Beispiel, eine Methode vorstellen, die einen Eulerkreis in unserem Graphen liefert.

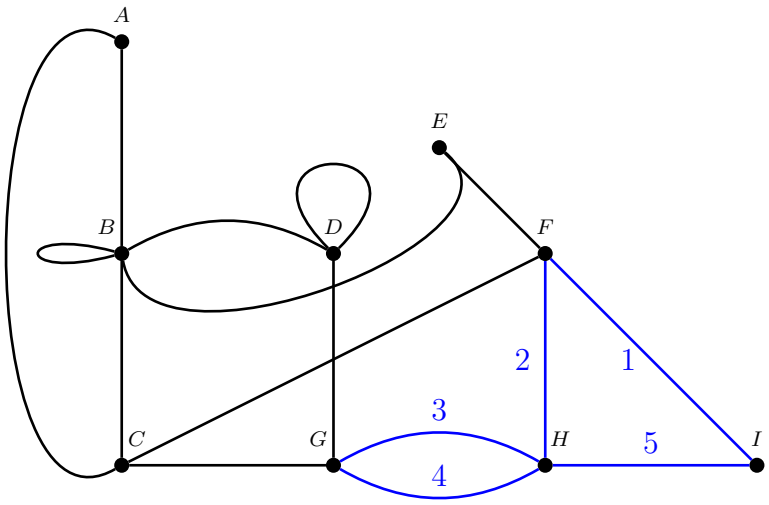
Wir gehen im Wesentlichen wie folgt vor: Wir starten in einer beliebigen Ecke und gehen einen beliebigen Weg, ohne jedoch eine Kante doppelt zu benutzen, bis wir in der Startecke wieder angekommen sind. Hat man noch nicht alle Kanten an der Startecke verbraucht, so wiederholt man den Vorgang, bis man wieder an der Startecke angekommen ist und keine Kanten mehr zur Verfügung hat. Nun hat man einen geschlossenen Weg, der keine Kante doppelt benutzt, der allerdings möglicherweise nicht alle Kanten im Graphen benutzt. Wir wollen danach an diesen Weg weitere Rundwege an Zwischenstationen „dranhängen“. Um das zu veranschaulichen, betrachten wir ein Beispiel.



In diesem Beispiel haben wir die folgenden Knotengrade:

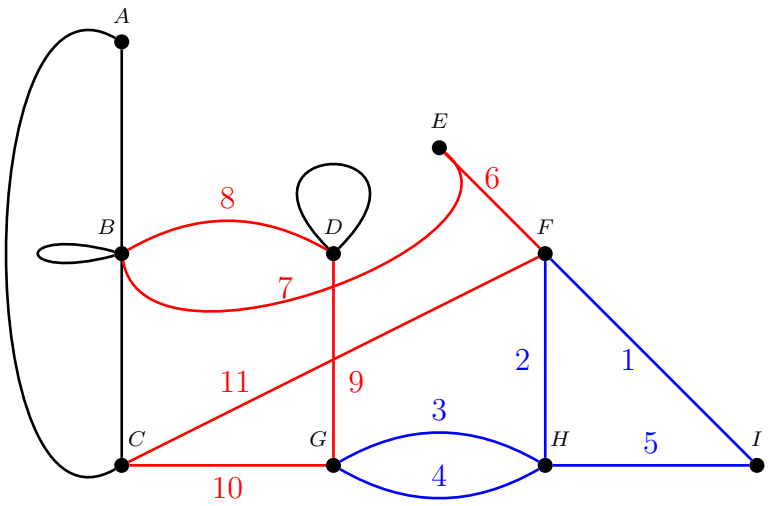


Insbesondere sind alle Knotengrade gerade, es ist also ein Graph, der die Prämisse unserer Behauptung erfüllt. Nun laufen wir in der Ecke I los und nummerieren die Kanten, die wir durchlaufen haben:



Unser erster Weg ist also $I, 1, F, 2, H, 3, G, 4, H, 5, I$.

Nun erweitern wir, wie angekündigt, diesen Weg, indem wir dasselbe Verfahren jetzt mit dem Startpunkt F anwenden, wobei wir die blauen Kanten nicht benutzen dürfen. Dabei erhalten wir den folgenden geschlossenen Weg an F:

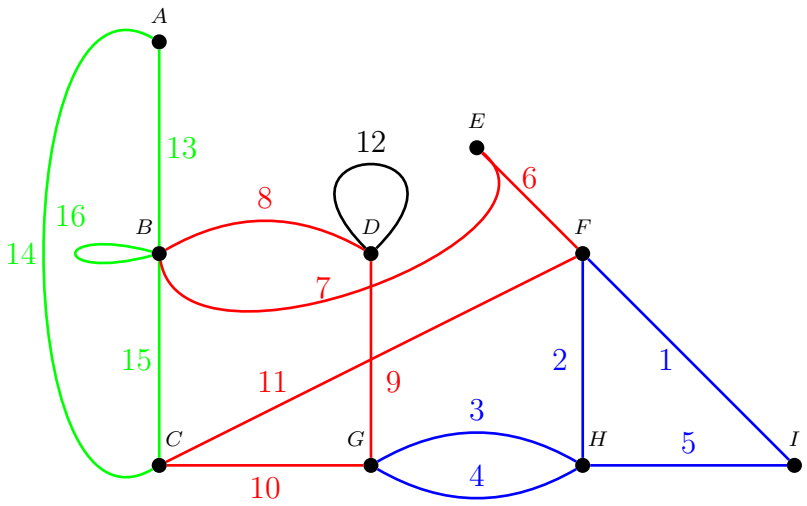


Der rote Weg ist beschrieben durch $F, 6, E, 7, B, 8, D, 9, G, 10, C, 11, F$. Diesen Weg können wir mit unserem ersten Weg kombinieren, indem wir diesen neuen Weg als ein Umweg in F einbauen und erhalten:

$$I, 1, F, 6, E, 7, B, 8, D, 9, G, 10, C, 11, F, 2, H, 3, G, 4, H, 5, I.$$

Damit erhalten wir einen Weg, der jede rote und jede blaue Kante genau einmal benutzt. Nun merkt man allerdings, dass der Rest (also die schwarzen Kanten) keinen zusammenhängenden Graphen mehr bilden. Allerdings hat jedes der übrigen Teile nach wie vor die Eigenschaft, dass jeder Knoten darin mit den verbleibenden Kanten geraden Grad hat. Das folgt aus derselben Überlegung wie im ersten Teil des Beweises - sollten wir eine Ecke besucht haben, so haben wir bei jedem Besuch genau 2 Kanten verbraucht, und den Knotengrad somit bei jedem Besuch um 2 gesenkt. Da der Knotengrad am Anfang gerade war, wird er nach dem Entfernen eines geschlossenen Rundgangs, der jede Kante höchstens einmal benutzt, ebenfalls gerade sein.

Wir hängen also an den bereits gefunden Weg nun zwei weitere Rundwege, jeweils in den übrigen Teilen des Graphen. In einem ist der komplette Rundgang sehr einfach, da er nur aus einer Schleife besteht. Im anderen Teil ist ein Rundgang etwa im folgenden Bild grün markiert:



Der Weg im grünen Teil ist nun gegeben durch

$$B, 13, A, 14, C, 15, B, 16, B.$$

Hier sollte man bemerken, dass wir gerade in B und nicht in A anfangen (C wäre allerdings auch möglich), weil der alte Weg durch A nicht durchgeht und wir sie somit nicht „zusammensetzen“ könnten.

Nun nehmen wir alle ermittelten Wege zusammen und erhalten dadurch einen Eulerkreis im ursprünglichen Graphen:

$$I, 1, F, 6, E, 7, B, 13, A, 14, C, 15, B, 16, B, 8, D, 12, D,$$
$$9, G, 10, C, 11, F, 2, H, 3, G, 4, H, 5, I.$$

Will man es mathematisch formalisieren, so würde man eine Induktion, beispielsweise nach der Anzahl der Ecken im Graphen machen. Dann weiß man nach dem Entfernen des ersten Rundwegs, der jede Kante höchstens einmal benutzt, mitsamt Ecken, die ohne diesen Rundweg mit dem Rest nicht mehr verbunden sind, dass man zumindestens die Startecke entfernt hat und somit weniger Ecken hat. Zwar ist der Rest nicht unbedingt zusammenhängend, wie wir gesehen haben, aber in jedem zusammenhängenden Teil können wir die Induktionsvoraussetzung anwenden und einen Eulerkreis finden, da, wie wir uns bereits überlegt haben, in den restlichen Teilen die Knotengrade nach wie vor gerade sind.. Als letztes muss man sich überlegen, dass die nach Induktionsvoraussetzung gefundenen Eulerkreise an den ersten Rundweg drangehängt werden können, also mindestens eine Ecke mit diesem gemeinsam haben. Damit zeigt man auch, dass der Induktionsschritt funktioniert, und kann dann den ersten Teil des Satzes beweisen.

Wir wollen noch kurz auf den zweiten Teil des Satzes eingehen. Wenn man nun eine Eulertour hat, die kein Eulerkreis ist, so verlässt man die Startecke

einmal häufiger, als man zurückkommt, und umgekehrt in der Zielecke: Da kommt man am Ende über eine Kante an, verbraucht aber keine, um die Ecke wieder zu verlassen. Somit müssen Start- und Zielecke einer Eulertour, sofern sie verschieden sind, unbedingt ungeraden Grad haben.

Hat man also umgekehrt in einem Graphen genau zwei Ecken vom ungeraden Grad vorgegeben, so startet man die Eulertour in einer der beiden besonderen Ecken und beendet Sie in der anderen. Will man also beispielsweise das Haus von Nikolaus in einem Zug zeichnen, so muss man in einer der unteren Ecken anfangen und in der anderen unteren Ecke aufhören, sonst kann es nicht funktionieren.

Ansonsten verfolgt man für den Beweis wieder die Idee, dass man zunächst einen Weg von einer Ecke ungeraden Grades zur anderen nimmt, der keine Kante doppelt nutzt, und dann Umwege daran anfügt, bis man eine Eulertour durch den gesamten Graphen erhalten hat.

Damit wollen wir die Behandlung des Satzes über Eulertouren und Eulerkreise abschließen. □

20 Hamiltonkreise in Graphen

Nach der Frage über Eulertouren, die so erschöpfend mit dem Satz 19.2 geklärt werden konnte, wollen wir uns einer ähnlichen Frage zuwenden, die sich als ungleich viel schwieriger herausstellt. Es geht nämlich darum, nun einen Weg in unserem Graphen zu finden, der nun nicht alle Kanten, sondern alle Ecken genau einmal benutzt (außer der Startecke, in die wir zurückkehren wollen). Das ist der Gegenstand der folgenden Definition.

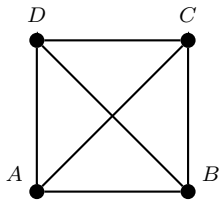
Definition 20.1. Ein **Hamiltonkreis** in einem Graphen ist ein geschlossener Weg, der jeden Knoten dieses Graphen außer dem Start- und Zielknoten genau einmal benutzt.

Ein Kontext, in dem eine solche Route wichtig sein kann, ist die Route eines Handelsreisenden, der aus der Filialniederlassung startet, in diese am Ende zurück will und jeden Kunden genau einmal benutzen will (mehr dazu können Sie unter dem Stichwort „travelling salesman problem“ in der Literatur finden). Ähnliche Probleme der optimalen Routenplanung werden vielfach in der Praxis gebraucht, oft durch die Forderung möglichst geringer Kosten oder möglichst schneller Auftragserfüllung verkompliziert.

Wir veranschaulichen den Begriff eines Hamiltonkreises an Beispielen.

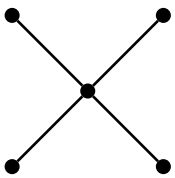
Beispiel.

- Zunächst bemerken wir, dass die Existenz eines Eulerkreises nicht die Existenz eines Hamiltonkreises beeinflusst.



  Wir haben gesehen, dass es in diesem Graphen zwar keinen Eulerkreis geben kann, aber man leicht einen Hamiltonkreis findet: Man geht etwa von A nach B, von da nach C und von C nach D, und dann wieder zurück zu A.

- Nicht alle Graphen besitzen einen Hamiltonkreis:



  In diesem Graphen wissen wir, dass es überhaupt keine Kreise gibt, also jeder geschlossener Weg eine Kante und damit mindestens zwei

Ecken doppelt nutzen muss. (Alternativ sehen wir, dass wir, um von einer beliebigen äußeren Ecke zu einer anderen solchen zu gelangen, stets durch den Mittelpunkt gehen müssen. Da wir diesen nur einmal benutzen können, können wir höchstens zwei äußere Ecken besuchen.)

Das Problem, ob ein einfacher zusammenhängender Graph einen Eulerkreis besitzt, kann in quadratischer Zeit (in Abhängigkeit von der Anzahl der Ecken) gelöst werden: Man muss sich nur zu jeder Ecke jede andere Ecke anschauen, um festzustellen, ob dazwischen eine Kante ist, und damit den Grad jeder Ecke bestimmen. Kennt man den Grad jeder Ecke, so sagt uns Satz 19.2, ob der Graph einen Eulerkreis besitzt.

Hingegen ist die frage, ob ein einfacher zusammenhängender Graph einen Hamiltonkreis besitzt algorithmisch sehr kompliziert, es ist eines der sogenannten NP-vollständigen Problemen. Insbesondere ist kein Algorithmus bekannt, der in polynomiellen Zeit (in Abhängigkeit von der Anzahl der Ecken) entscheidet, ob ein Graph einen Hamiltonkreis besitzt. Würde man einen solchen polynomiellen Algorithmus finden, so würde man damit das berühmte „$P = NP$“-Problem lösen, für das ein Preisgeld von einer Million Dollar ausgeschrieben ist.

Während die Frage nach Hamiltonkreisen im Allgemeinen sehr schwierig ist, kann man doch gewisse Aussagen darüber treffen. Im Folgenden wollen wir noch au ein hinreichendes Kriterium für Existenz der Hamiltonkreise liefern.

Im Satz 19.2 über die Existenz der Eulerkreise kam es auf die Parität der Knotengrade an. In diesem Satz werden die Knotengrade auch eine Rolle spielen, allerdings müssen Sie diesmal eine gewisse Mindestgröße besitzen, um die Existenz eines Hamiltonkreises garantieren zu können.

Satz 20.2. *Jeder einfache Graph mit $n \geq 3$ Knoten, in dem jeder Knoten den Grad $\geq \frac{n}{2}$ hat, besitzt einen Hamiltonkreis.*

Dieser Satz ist ein *hinreichendes* Kriterium für die Existenz eines Hamiltonkreises. Hinreichende Kriterien sind Gegenstücke zu den notwendigen Kriterien, die wir am Beispiel des Satzes 16.3 kennengelernt haben. Ist das Kriterium erfüllt, so garantiert es die Existenz eines Hamiltonkreises, allerdings gibt es Beispiele von Graphen, in denen die Knotengrade auch kleiner als $\frac{n}{2}$ sind, die allerdings trotzdem einen Hamiltonkreis besitzen, wie etwa der folgende Graph mit 8 Kanten, in dem jede Ecke nur den Grad 2 hat und in dem einfach „im Kreis gehen“ einen Hamiltonkreis liefert: