Johannes Kepler University Linz (JKU)
Research Institute for Symbolic Computation (RISC)

Skriptum zum Kurs

MATHEMATIK UND LOGIK

FÜR WIRTSCHAFTSINFORMATIK

Wolfgang Windsteiger[1]

ab Wintersemester 2017/18

[1] Dieses Skriptum basiert im Wesentlichen auf den Unterlagen aus den Vorjahren von Veronika Pillwein.

Inhaltsverzeichnis

KAPITEL 1

LOGIK

Aufgabe der *Logik* ist es, einen formalen Rahmen zu schaffen, in dem mathematische Sachverhalte präzise *beschrieben* und *überprüft* werden können. Die Logik gibt also auf der einen Seite einen sprachlichen Rahmen, auf der anderen Seite beschreibt sie, wie auf Basis dieser Sprache mathematische Argumente aufgebaut werden können.

AUSSAGENLOGIK

Aufgabe der *Aussagenlogik* ist es, vereinfacht und verkürzt gesagt, die Wahrheit von kompliziert ausgebauten Aussagen auf Basis der Wahrheit ihrer elementaren Bestandteile zu beurteilen. Nicht mehr und nicht weniger ...

Eine *Aussage* ist jeder Ausdruck (der Umgangssprache), dem die Eigenschaft *wahr* (w) oder *falsch* (f) zugesprochen werden kann.

Beispiele für Aussagen sind

- Meine Haare sind schwarz.
- $3^3 = 81$
- $-1 < 14$
- Die letzte Stelle der 247. Primzahl ist 3.

Keine Aussagen sind

- Ist π eine ganze Zahl?
- $(a + b)^2$

In jeder Sprache, somit auch in der Sprache der Logik, unterscheidet man zwischen *Syntax* und *Semantik*. Syntax bezieht sich immer auf die *äußere Form* von Ausdrücken, Semantik hingegen beschreibt die *Bedeutung* von Ausdrücken der Sprache.

A	$\neg A$
w	f
f	w

A	B	$A \vee B$	$A \wedge B$	$A \Rightarrow B$	$A \Leftrightarrow B$
w	w	w	w	w	w
w	f	w	f	f	f
f	w	w	f	w	f
f	f	f	f	w	w

Tabelle 1.1: Wahrheitstabellen für aussagenlogische Junktoren.

Syntax: Es stehen 5 Junktoren $\neg$ (Negation), $\vee$ (Disjunktion), $\wedge$ (Konjunktion), $\Rightarrow$ (Implikation) und $\Leftrightarrow$ (Äquivalenz) zur Verfügung. Mit deren Hilfe können aus gegebenen Aussagen neue Aussagen gebildet werden. Seien A und B Aussagen, dann sind

$$\neg A \qquad A \vee B \qquad A \wedge B \qquad A \Rightarrow B \qquad A \Leftrightarrow B$$

ebenfalls zulässige Aussagen.

Semantik: Für jede Aussage gibt es nur 2 mögliche Bedeutungen. Um die Bedeutung von $\neg A$ festzulegen, müssen wir lediglich die Möglichkeiten „A wahr“ bzw. „A falsch“ in Betracht ziehen. Bei den restlichen Junktoren müssen wir alle Kombinationen „A wahr“ bzw. „A falsch“ mit „B wahr“ bzw. „B falsch“ durchspielen, das sind jeweils genau 4 Möglichkeiten, die wir in sogenannten *Wahrheitstabellen* festlegen, siehe Tabelle 1.1.

Aus den Wahrheitstabellen 1.1 ist zu erkennen, dass

- $\neg A$ genau dann wahr ist, wenn A *nicht* wahr ist;
- $A \vee B$ genau dann wahr ist, wenn A wahr ist *oder* B wahr ist (oder beide);
- $A \wedge B$ genau dann wahr ist, wenn A wahr ist *und* B wahr ist;
- $A \Rightarrow B$ genau dann wahr ist, wenn *falls* A wahr ist, *dann auch* B wahr ist;
- $A \Leftrightarrow B$ genau dann wahr ist, wenn A *genau dann* wahr ist, *wenn* B wahr ist;

Die Aussagen $A \vee B$ und $B \vee A$ sind *voneinander verschiedene Aussagen*, sie sind jedenfalls *nicht gleich*. Betrachtet man aber ihre Wahrheitstabellen, so stellt man fest, dass ihre *Bedeutung* in allen Situationen übereinstimmt.

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

Definition 1.1: Gleichwertige Aussagen, logische Konsequenz

Zwei Aussagen A und B nennt man *gleichwertig* (*äquivalent*[a]) genau dann, wenn ihre ihre Bedeutung in allen Situationen übereinstimmt, d.h. wenn ihre Wertetabellen übereinstimmen. Man schreibt dafür $A \equiv B$.

Die Aussage B ist eine *aussagenlogische Konsequenz* von A (man sagt auch: B *folgt aus* A) genau dann, wenn B in allen Situationen wahr ist, in denen auch A wahr ist. Man schreibt dafür $A \models B$.

[a] Es sei erwähnt, dass auch auf Ebene der Syntax oft für die Aussage „$A \Leftrightarrow B$“ gesprochen wird „A äquivalent B“. Dieser Unterschied ist etwas subtil, $\Leftrightarrow$ ist ein syntaktischer Bestandteil der Sprache, und $A \Leftrightarrow B$ ist selbst eine Aussage, wohingegen $\equiv$ sich auf die Semantik bezieht, und $A \equiv B$ *keine Aussage* ist, sondern eine *Beziehung zwischen Aussagen* (das ist eine andere Sprachebene!).

Das einleitende Beispiel hat schon die *Kommutativität* der Disjunktion gezeigt, analog dazu können mittels Wahrheitstafeln weitere Gesetze hergeleitet werden.

$$A \vee B \equiv B \vee A \qquad A \wedge B \equiv B \wedge A \qquad \text{(Kommutativität)}$$
$$(A \vee B) \vee C \equiv A \vee (B \vee C) \qquad (A \wedge B) \wedge C \equiv A \wedge (B \wedge C) \qquad \text{(Assoziativität)}$$

Beispiel 1.2

Wir zeigen $(A \wedge B) \wedge C \equiv A \wedge (B \wedge C)$ und bestimmen dazu die Wahrheitstafeln für beide Seiten.

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

Beispiel 1.3

Aus der Wahrheitstafel in Beispiel 1.2 können wir auch ablesen, dass B eine aussagenlogische Konsequenz von $A \wedge B$ ist. Wir vergleichen dazu die Spalten zu B bzw. $A \wedge B$ und beachten nur jene Zeilen, in denen bei $A \wedge B$ der Wert „w“ steht.

SATZ 1.4

Seien A, B, C Aussagen, dann gelten die Distributivgesetze

$$A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C) \qquad A \wedge (B \vee C) \equiv (A \wedge B) \vee (A \wedge C).$$

Beweis. Wir zeigen diese Aussage (lt. Definition von $\equiv$) über Wahrheitstafeln:

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

□

Die *Gesetze von De Morgan* beschreiben, wie sich „und“ bzw. „oder“ beim Verneinen verhalten.

SATZ 1.5: DE MORGAN

Seien A, B Aussagen. Dann gilt

$$\neg(A \wedge B) \equiv \neg A \vee \neg B \qquad \neg(A \vee B) \equiv \neg A \wedge \neg B \qquad \text{(Gesetze von De Morgan)}$$

Beweis. Wir zeigen Teil 1 mittels Wahrheitstafeln, der Beweis zu Teil 2 geht analog.

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

□

LEMMA 1.6

Seien A, B Aussagen. Dann gelten folgende Beziehungen:

$$A \Leftrightarrow B \equiv (A \Rightarrow B) \wedge (B \Rightarrow A) \tag{1.1}$$
$$A \Rightarrow B \equiv \neg A \vee B \tag{1.2}$$
$$A \Rightarrow B \equiv \neg B \Rightarrow \neg A \tag{1.3}$$
$$\neg\neg A \equiv A \tag{1.4}$$

Alle Beziehungen $X \equiv Y$ können als *Umformungsregeln* betrachtet werden, die eine Ersetzung von X durch Y und umgekehrt erlauben, wie man es von Gleichheit ($=$) gewohnt ist.

BEISPIEL 1.7

Seien A, B Aussagen: Die Aussage $\neg(A \Rightarrow B)$ kann durch Anwenden von Umformungsregeln durch $\neg, \vee, \wedge$ ausgedrückt werden:

$$\neg\,(A \Rightarrow B) \quad \equiv \quad \neg\,(\neg A \vee B) \quad \equiv \quad A \wedge \neg B.$$

Es gilt daher auch $\neg(A \Rightarrow B) \equiv A \wedge \neg B$.

DEFINITION 1.8: TAUTOLOGIE, KONTRADIKTION

Aussagen, die immer wahr sind, heißen *Tautologien*. Aussagen, die immer falsch sind, heißen *Kontradiktionen*.

Ein einfaches Beispiel für eine Tautologie ist $A \vee \neg A$, ein einfaches Beispiel für eine Kontradiktion ist $A \wedge \neg A$. Ist A eine Tautologie, so symbolisieren[1] wir dies durch $A \equiv \top$, ist A eine Kontradiktion, so schreiben wir $A \equiv \bot$. In diesem Sinne ist

$$A \vee \neg A \equiv \top \qquad\qquad A \wedge \neg A \equiv \bot.$$

Auch solche Aussagen lassen sich als Umformungsregeln gebrauchen. Führt eine (eventuell auch lange) Kette von Umformungen einer Aussage X schlussendlich auf $\top$, so heißt das $X \equiv \top$, also ist X eine Tautologie. Analog dazu, falls eine Umformungskette auf $\bot$ führt, so heißt das $X \equiv \bot$, also ist X eine Kontradiktion[2].

Wir fassen diese Aussagen und noch ein paar weitere einfache Beispiele in folgendem Lemma zusammen.

[1] $\top$ und $\bot$ sind neue Ausdrücke auf Ebene der Syntax der Aussagenlogik. Der Junktor Negation ist 1-stellig, hängt also von 1 Grundaussage ab, die anderen Junktoren hängen jeweils von 2 Grundaussagen ab, sie heißen 2-stellig. In diesem Sinn könnte man $\top$ und $\bot$ als 0-stellige Junktoren auffassen, die konstante Aussagen bilden, eben eine, die immer wahr ist und eine, die immer falsch ist.

[2] Vergleiche unsere Ausführungen zum Thema $\equiv$ vs. $\Leftrightarrow$. Tautologie und Kontradiktion sind Konzepte, die semantische Eigenschaften von Aussagen beschreiben, $\top$ und $\bot$ hingegen sind syntaktische Bausteine der Sprache.

LEMMA 1.9

Sei A eine Aussage. Dann gilt:

$$A \wedge A \equiv A \qquad A \wedge \neg A \equiv \bot \qquad A \wedge \top \equiv A \qquad A \vee \top \equiv \top$$
$$A \vee A \equiv A \qquad A \vee \neg A \equiv \top \qquad A \wedge \bot \equiv \bot \qquad A \vee \bot \equiv A$$

BEISPIEL 1.10

Welche der folgenden Aussagen sind Tautologien, Kontradiktionen oder weder noch?

- $A \vee (A \Rightarrow B)$
- $(A \Rightarrow \neg A) \wedge A$
- $(A \Rightarrow \neg B) \wedge (A \Rightarrow B)$
- $(A \wedge (A \Rightarrow B)) \Rightarrow B$

PRÄDIKATENLOGIK

In der Aussagenlogik können wir jede Grundaussage mit einem Aussagensymbol bezeichnen und kompliziertere Aussagen mit Junktoren zusammenbauen. Mit Hilfe von Wahrheitstabellen können wir dann die Bedeutung der Aussagen unter die Lupe nehmen.

BEISPIEL 1.11

Nehmen wir als Grundaussagen nun $x < 10$ (A) und $x < 15$ (B). Haben wir weiters eine Aussage $x \geq 10$, so schreiben wir dafür sinnvollerweise $\neg A$ anstatt eine neue Aussage C einzuführen. Die Aussage „x ist kleiner 10 *oder* x ist größer gleich 10“ ist *immer wahr*, es handelt sich um $A \vee \neg A$, was rein auf Basis der Aussagenlogik als Tautologie erkannt wird (Lemma 1.9). Die Aussage „x ist kleiner 10 *und* x ist größer gleich 10“ ist *immer falsch*, es handelt sich um die bekannte Kontradiktion $A \wedge \neg A$ aus Lemma 1.9.

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

Betrachten wir nun „wenn x kleiner als 10 ist, dann ist x auch kleiner als 15“, in die Aussagenlogik zu übersetzen als $A \Rightarrow B$. Aufgrund der Wahrheitstabelle können wir weder sagen, dass die Aussage immer wahr ist, noch dass sie immer falsch ist.

Die Aussage *ist* aber immer wahr, die Sprache der Aussagenlogik ist lediglich zu schwach, um den inneren Zusammenhang der beiden Aussagen A und B auszudrücken. Es handelt sich bei den beiden ja um keine beliebigen zwei Aussagen, die innere Struktur der beiden Aussagen sorgt nämlich dafür, dass die Kombination A wahr und B falsch in der Realität nicht auftreten

kann. In allen möglichen verbleibenden Fällen ist die Aussage dann wahr. Die *Prädikatenlogik* wird uns die dafür nötigen Werkzeuge liefern.

In der Sprache der Prädikatenlogik unterscheiden wir zwischen *Termen* und *Aussagen*. Terme sollen Objekte bezeichnen, Aussagen stehen – wie in der Aussagenlogik – für Sachverhalte, die wahr oder falsch sein können.

Auf der Ebene der Syntax gibt es:

- *Objektkonstante* (z.B. 1, -12, π);
- *Variable* (z.B. x, y, n);
- *Funktionssymbole* für konkrete Operationen (z.B. $+$, $/$, etc.);
- *Prädikatensymbole* für konkrete Eigenschaften (z.B. $<$, $=$, $\in$, $\subseteq$, etc.);
- die aus der Aussagenlogik bekannten *Junktoren* $\neg$, $\vee$, $\wedge$, $\Rightarrow$ bzw. $\Leftrightarrow$;
- die *Quantoren* $\forall$ und $\exists$.

Die Bedeutung von Ausdrücken (Semantik) wird festgelegt durch:

- einen nicht-leeren Grundbereich G bestehend aus den gerade interessierenden Objekten;
- eine Interpretation, die
  - jeder Konstanten ein konkretes Objekt aus G zuordnet,
  - jedem Funktionssymbol eine „passende Operation“ auf G zuordnet und
  - jedem Prädikatensymbol eine „passende Eigenschaft“ auf G zuordnet;
- eine Belegung der freien Variablen.

Der Zweck von Beweisen ist, ausgehend von Aussagen, die in einer bestimmten Realität gelten, die Gültigkeit einer Aussage in derselben Realität zu erschließen. Dabei muss jeder der endlich vielen Schlussschritte abstrakt, korrekt und kontrollierbar sein.

Ein *Beweis* einer Aussage A basierend auf einem bestimmten „Grundwissen“ Γ (über der betrachteten Realität) ist ein Argument bestehend aus einer endlichen Folge von Aussagen, wobei jede Aussage der Folge entweder aus dem Grundwissen Γ ist oder aus den vorherigen Aussagen des Arguments in einem Schlussschritt entsteht. Die letzte Aussage dieser Folge ist A. Man schreibt dafür $\Gamma \vdash A$. Es ist interessant anzumerken, dass jeder Beweisschritt aus einer rein syntaktischen Umformung von Aussagen besteht.

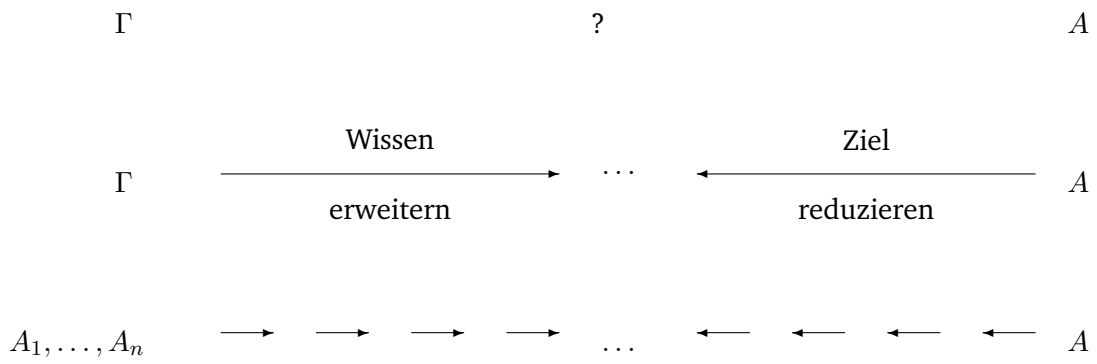


Abbildung 1.1: Praktisches Beweisen: Kombination von Vorwärts- und Rückwärtsschließen.

Das Grundwissen besteht zum Beispiel aus früher bewiesenen Aussagen, aus allgemeingültigen Aussagen, Definitionen, usw. Die Beweisschritte sollen nicht zu groß sein (damit der Beweis nachvollziehbar bleibt) und nicht zu klein (damit der Beweis überschaubar bleibt).

Um eine Aussage zu beweisen (oder zu widerlegen), muss man sich zunächst klar darüber sein, was die *Voraussetzungen* sind, die man als bekannt annimmt ($= \Gamma$), und was zu zeigen ist ($= A$). Eine zu zeigende Aussage A kann aus dem Grundwissen (inklusive der Voraussetzungen) direkt hergeleitet werden oder auf die bekannten Voraussetzungen zurückgeführt werden. Eine Aussage A kann aber auch *indirekt* bewiesen werden, indem ihre Negation ($\neg A$) mit dem Grundwissen auf einen Widerspruch geführt wird. Um eine Aussage A zu widerlegen, beweist man einfach $\neg A$. Es sei an dieser Stelle darauf hingewiesen, dass in der Praxis meist nur gesagt wird, was zu beweisen ist; das zur Verfügung stehende Grundwissen muss meist aus dem Zusammenhang erraten werden[3].

Stimmt die zu beweisende Aussage A mit einer der als wahr bekannten Voraussetzungen in Γ überein, ist natürlich nichts mehr zu tun. Üblicherweise besteht aber zwischen Γ und A eine gewisse „Diskrepanz“, die es mit Hilfe eines Beweises zu überbrücken gilt. In der Praxis kombiniert man dabei meist *Vorwärtsschritte*, in denen man neues Wissen aus schon Bekanntem herleitet, mit *Rückwärtsschritten*, in denen man die zu beweisende Aussage auf eine andere (hoffentlich einfacher zu beweisende) Aussage reduziert, um sich schlussendlich „in der Mitte zu treffen“, siehe Abbildung 1.1.

Die bekannten Beweisregeln der Prädikatenlogik bauen allesamt auf der syntaktischen Struktur der beteiligten Aussagen auf.

[3]Das ist die „Praxis“ beim Erlernen des Beweisens. Im echten Leben ist es so, dass man es mit bisher nicht bekannten Aussagen zu tun hat, die zu beweisen oder widerlegen sind. Als Grundwissen ist alles erlaubt, was irgendwoher als „wahr“ zur Verfügung steht.

Terme und atomare Aussagen

Syntax:

- Eine Konstante ist ein Term.
- Eine Variable ist ein Term.
- Ist f ein Funktionssymbol und $t_1, \ldots, t_n$ Terme, dann ist $f(t_1, \ldots, t_n)$ ebenfalls ein (zusammengesetzter) Term. Funktionssymbole müssen nicht unbedingt mit Klammern angewendet werden, es gibt unzählige mathematische *Notationen* (Infix, Präfix, Postfix, etc.).
- Ist p ein Prädikatensymbol und $t_1, \ldots, t_n$ Terme, dann ist $p(t_1, \ldots, t_n)$ eine atomare Aussage. Prädikatensymbole müssen nicht unbedingt mit Klammern angewendet werden, es gibt unzählige mathematische *Notationen* (Infix, Präfix, Postfix, etc.).

Terme und atomare Aussagen sind jene Ausdrücke, die durch endlich viele Anwendungen dieser Regeln entstehen können.

Semantik:

- Die Bedeutung von Konstanten ist in der Interpretation festgelegt.
- Die Bedeutung von Variablen ist aus der Variablenbelegung abzulesen.
- Die Bedeutung eines zusammengesetzten Terms $f(t_1, \ldots, t_n)$ erhält man, indem man zuerst rekursiv die Bedeutung der $t_1, \ldots, t_n$ ermittelt und dann die von f bezeichnete Operation auf diese Objekte anwendet.
- Die Bedeutung einer atomaren Aussage $p(t_1, \ldots, t_n)$ erhält man, indem man zuerst rekursiv die Bedeutung der $t_1, \ldots, t_n$ ermittelt und dann die von p bezeichnete Eigenschaft dieser Objekte überprüft.

Beweisregeln: Für atomare Aussagen gibt es keine speziellen Regeln. Ist eine Aussage $p(t_1, \ldots, t_n)$ in den Voraussetzungen enthalten oder zu beweisen, so muss man sich darauf verlassen, dass weiteres Wissen über p in den Vorausetzungen verfügbar ist. Dies kann zum Beispiel durch eine *Definition* von p gegeben sein, d.h. eine Aussage der Form[4]

$$p(x_1, \ldots, x_n) :\Leftrightarrow R_{x_1,\ldots,x_n}.$$

Einsetzen einer Definition heißt dann nichts anderes als jedes Auftreten der linken Seite der Definition durch die ensprechende rechte Seite zu ersetzen, mit anderen Worten jedes Auftreten von $p(t_1, \ldots, t_n)$ durch $R_{x_1 \to t_1,\ldots,x_n \to t_n}$ zu ersetzen. Im Normalfall sind in R dann wieder Junktoren und Quantoren enthalten, sodass mit den später gezeigten Regeln für Junktor- bzw. Quantoraussagen weitergemacht werden kann.

Auch zusammengesetzte Terme sind üblicherweise durch Definitionen eingeführt. Analog zu oben sind das Aussagen der Form

$$f(x_1, \ldots, x_n) := T_{x_1,\ldots,x_n},$$

[4]Der Doppelpunkt in einer Definition hat keine logische Bedeutung, er dient nur dazu, eine Definition von anderen Aussagen unterscheiden zu können, und sie auch optisch sofort als solche zu erkennen.

und diese können dazu verwendet werden, um jedes Auftreten der linken Seite der Definition durch die ensprechende rechte Seite zu ersetzen, mit anderen Worten jedes Auftreten von $f(t_1, \ldots, t_n)$ durch $T_{x_1 \to t_1, \ldots, x_n \to t_n}$ zu ersetzen.

Die einzige Ausnahme bildet das 2-stellige Prädikatensymbol „Gleichheit“ ($=$). Die Gleichheit ist in der Mathematik nicht durch eine Definition beschrieben, vielmehr ist die Bedeutung der Gleichheit direkt in die Logik „eingebaut“, indem eine Beweisregel den Umgang mit der Gleichheit regelt.

Regel ($=$-wissen): Wenn $t = s$ wahr ist, dann darf überall t durch s ersetzt werden.

BEISPIEL 1.12: SYNTAX

3 sei eine Konstante und x sei eine Variable. Dann ist $x + 3$ ein Term. In diesem Fall ist $+$ ein Funktionssymbol und wir verwenden eine Infix-Notation $x + 3$ statt $+(x, 3)$. Ebenfalls ein Term ist $3!$, hier ist $!$ ein Funktionssymbol, das in Postfix-Notation $3!$ statt $!(3)$ geschrieben ist.

$x < 3$ ist eine atomare Aussage, $<$ ist ein Prädikatensymbol, das ebenfalls in Infix-Schreibweise verwendet wird. $ist\_prim(3)$ und $ist\_prim(x)$ sind ebenfalls atomare Aussagen, in denen im Unterschied zu den bisherigen Beispielen keine spezielle mathematische Notation zum Einsatz kommt.

BEISPIEL 1.13: SEMANTIK

Es sei eine Interpretation wie folgt festgelegt: Grundbereich seien die natürlichen Zahlen, Interpretation von 3 ist die natürliche Zahl „drei“, Interpretation von $+$ sei die Addition[a], Interpretation von $!$ sei die Fakultätsfunktion, Interpretation von $<$ sei die kleiner-Relation und die Interpretation von ist_prim sei die Eigenschaft einer natürlichen Zahl, eine Primzahl zu sein.

Unter der Variablenbelegung $[x \mapsto$ „eins“$]$ ist

- die Bedeutung von $x + 3$ dann die *Addition* angewendet auf die Bedeutung von x und die Bedeutung von 3, also die Addition angewendet auf „eins“ und „drei“, also „vier“;
- die Bedeutung von $3!$ dann die *Fakultät* angewendet auf die Bedeutung von 3, also die Fakultät angewendet auf „drei“, also „sechs“;
- die Bedeutung von $x < 3$ dann die *kleiner-Relation* angewendet auf die Bedeutung von x und die Bedeutung von 3, also die kleiner-Relation angewendet auf „eins“ und „drei“, also „w“;
- die Bedeutung von $ist\_prim(x)$ dann die *Primzahl-Eigenschaft* angewendet auf die Bedeutung von x, also die Primzahl-Eigenschaft angewendet auf „eins“, also „f“;

Unter der Variablenbelegung $[x \mapsto$ „elf“$]$ ist

- die Bedeutung von $x + 3$ dann die *Addition* angewendet auf die Bedeutung von x und die Bedeutung von 3, also die Addition angewendet auf „elf“ und „drei“, also „vierzehn“;
- die Bedeutung von $3!$ dann die *Fakultät* angewendet auf die Bedeutung von 3, also die Fakultät angewendet auf „drei“, also „sechs“;
- die Bedeutung von $x < 3$ dann die *kleiner-Relation* angewendet auf die Bedeutung von x und die Bedeutung von 3, also die kleiner-Relation angewendet auf „elf“ und „drei“, also „f“;
- die Bedeutung von $ist\_prim(x)$ dann die *Primzahl-Eigenschaft* angewendet auf die Bedeutung von x, also die Primzahl-Eigenschaft angewendet auf „elf“, also „w“;

[a] Funktions- und Prädikatensymbole müssen in dem Sinn „passend“ interpretiert werden, dass die Anzahl der Terme, auf die die jeweiligen Symbole syntaktisch angewendet werden, mit der Anzahl der Parameter der Operation bzw. Eigenschaft übereinstimmt.

Der Unterschied zwischen Konstanten und Variablen besteht genau darin, dass die Bedeutung von Variablen sich ändern kann, indem die Variablenbelegung verändert wird, und Konstante durch Festlegung der Interpretation fixiert sind. Enthält ein Term oder eine atomare Aussage Variable, so kann eine Bedeutung nur angegeben werden, wenn für alle Variablen eine Belegung gegeben ist.

Junktoraussagen

Syntax: Seien A und B beliebige Aussagen, dann sind

$$\neg A \qquad A \vee B \qquad A \wedge B \qquad A \Rightarrow B \qquad A \Leftrightarrow B$$

ebenfalls zulässige Aussagen, sogenannte *Junktoraussagen* (siehe Aussagenlogik in Abschnitt 1.1).

Semantik: Ermittle im Falle von $\neg A$ die Bedeutung von A, ansonsten die Bedeutungen von A und B und lese die Bedeutung der Junktoraussage aus den *Wahrheitstabellen* ab, siehe Aussagenlogik Abschnitt 1.1, Tabelle 1.1.

Beweisregeln: Wir werden für jeden Typ von Junktoraussage Regeln kennenlernen, wie diese Aussagen zu beweisen sind oder als Teil der Wissensbasis zu behandeln sind.

BEISPIEL 1.14: SYNTAX

Es seien die Voraussetzungen wie in Beispiel 1.12. Die Aussagen

$$x < 3 \Rightarrow \neg ist\_prim(x) \qquad x < 3! \wedge ist\_prim(x) \tag{1.5}$$

sind allesamt Junktoraussagen. Der Ausdruck $x + 3 \Rightarrow x < 3$ ist syntaktisch inkorrekt.

BEISPIEL 1.15: SEMANTIK

Es seien die Voraussetzungen wie in den Beispielen 1.12 und 1.13. Ohne eine Variablenbelegung festzulegen, kann den Aussagen in (1.5) keine Bedeutung zugeordnet werden.

Unter der Variablenbelegung $[x \mapsto$ „eins“$]$ ist $x < 3 \Rightarrow \neg ist\_prim(x)$ eine *wahre Aussage*:

- die Bedeutung von $x < 3$ ist „w“;
- die Bedeutung von $ist\_prim(x)$ ist „f“;
- laut Wahrheitstabelle für die Negation ist daher $\neg ist\_prim(x)$ „w“;
- die Wahrheitstabelle für die Implikation w $\Rightarrow$ w ergibt „w“.

Beweisregeln für Junktoraussagen

Regel (Widerspruchsbeweis, indirekter Beweis): Die Grundidee ist es, statt die zu beweisende Aussage A zu zeigen, ihre Negation $\neg A$ zu widerlegen, indem man unter Annahme der Negation $\neg A$ einen Widerspruch herleitet. Die Negation $\neg A$ kann somit nicht wahr sein, also muss A wahr sein (eine andere Möglichkeit gibt es nicht, „*tertium non datur*“, wie man so schön sagt).

Insbesondere dann, wenn eine Aussage der Form $\neg A$ zu beweisen ist, bietet sich ein indirekter Beweis an. In diesem Fall ist dann die Negation $\neg\neg A$, also einfach A.

BEISPIEL 1.16

Die Quadratwurzel aus 2 lässt sich nicht als gekürzte Bruchzahl darstellen ($\sqrt{2} \neq \frac{p}{q}$ mit p, q so, dass p und q außer 1 keine gemeinsamen Faktoren enthalten).

Beweis: Angenommen $\sqrt{2} = \frac{p}{q}$ mit p, q so, dass p und q außer 1 keine gemeinsamen Faktoren enthalten. Das bedeutet aber, dass $2 = \frac{p^2}{q^2}$ ist, und damit $p^2 = 2q^2$. Das hat zur Folge, dass p^2 gerade ist, ...

Durch weitere Argumentation muss nun versucht werden, einen Widerspruch herzuleiten. Das wäre etwa dann der Fall, wenn wir erreichen würden, dass p^2 nicht gerade ist. In diesem Fall wäre der Beweis erfolgreich zu Ende gebracht und damit bewiesen, dass sich $\sqrt{2}$ sich nicht als gekürzte Bruchzahl darstellen lässt.

Die Details dazu folgen später, wir benötigen zur Herleitung des Widerspruchs noch mehr Beweisregeln.

BEISPIEL 1.17

Sei x eine positive reelle Zahl. Dann gilt $\frac{x+1}{x+2} < \frac{x+3}{x+4}$.

Beweis. Angenommen es wäre

$$\frac{x+1}{x+2} \geq \frac{x+3}{x+4}. \tag{1.6}$$

Da $x > 0$ ist, gilt auch $x + 2 > 0$ und $x + 4 > 0$. Damit können wir die Ungleichung in (1.6) mit $(x + 2)(x + 4)$ multiplizieren und erhalten

$$x^2 + 5x + 4 \geq x^2 + 5x + 6,$$

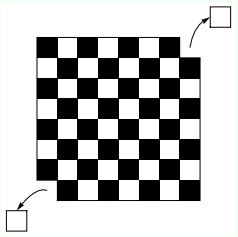
was gleichbedeutend ist mit

$$4 \geq 6.$$

Widerspruch! □

BEISPIEL 1.18

Wir betrachten ein Schachbrett, bei dem zwei (weiße) Felder an gegenüberliegenden Ecken entfernt wurden. Für dieses Brett gilt: Es gibt keine überlappungsfreie Überdeckung des Bretts mit 2×1-Dominosteinen (d.h. von der Größe von zwei Feldern).



Beweis. Wir zeigen diese Aussage mittels Widerspruch: Angenommen, es gibt eine überlappungsfreie Überdeckung. Jeder Stein in dieser Überdeckung deckt ein weißes und ein schwarzes Feld ab. Von den ursprünglich 64 Feldern sind nach Wegnahme der zwei weißen Felder noch 62 (32 schwarze und 30 weiße) übrig, für eine überlappungsfreie Überdeckung brauchen wir daher genau 31 Steine. Egal, wie die Überdeckung gelegt wird, mit 31 Steinen werden 31 weiße und 31 schwarze Felder abgedeckt, Widerspruch! □

Ein alternativer Weg zum Umgang mit Negationen ist, mit Hilfe der Regeln von de Morgan die Negation in die Aussage A hineinzuziehen, und dann mit anderen Regeln weiterzumachen.

Regel ($\wedge$-beweisen): Wenn $A \wedge B$ zu beweisen ist, müssen beide Aussagen A und B bewiesen werden.

Regel ($\wedge$-wissen): Wenn $A \wedge B$ als wahr bekannt ist, dann dürfen sowohl A als auch B als wahr angenommen werden.

Die beiden Regeln für Konjunktion sind derart trivial, dass sie üblicherweise in Beweisen nicht extra erwähnt werden. Die Richtigkeit der Regeln erklärt sich aus der Wahrheitstafel von „$\wedge$“.

Regel ($\vee$-beweisen): Wenn $A \vee B$ zu beweisen ist, reicht es aus, eine der Aussagen A oder B zu beweisen.

Regel ($\vee$-wissen, Fallunterscheidung): Wenn $A \vee B$ als wahr bekannt ist, kann eine Fallunterscheidung gemacht werden:

- Fall 1: Annahme, dass A wahr ist: ...
- Fall 2: Annahme, dass B wahr ist: ...

Wichtig ist, dass in *beiden Fällen* die zu beweisende Aussage bewiesen werden muss.

Ist $A_1 \vee \ldots \vee A_n$ als wahr bekannt, so müssen analog n Fälle unterschieden werden.

Regel ($\Rightarrow$-beweisen): Um eine Aussage $A \Rightarrow B$ zu beweisen, darf A als wahr angenommen werden. Unter dieser zusätzlichen Annahme ist dann B zu beweisen.

Regel ($\Rightarrow$-wissen, Modus Ponens): Wenn sowohl A als auch $A \Rightarrow B$ wahr sind, dann muss auch B wahr sein.

Die Richtigkeit der Regeln erklärt sich wieder aus der Wahrheitstabelle von „$\Rightarrow$“.

Regel ($\Leftrightarrow$-beweisen): Um eine Aussage $A \Leftrightarrow B$ zu beweisen, beweist man zuerst $A \Rightarrow B$ und dann $B \Rightarrow A$.

Regel ($\Leftrightarrow$-wissen): Wenn eine Aussage $A \Leftrightarrow B$ als wahr bekannt ist, dann darf überall A durch B ersetzt werden.

Die Regel ($\Leftrightarrow$-wissen) gewinnt insbesondere daher an Bedeutung, da für gleichbedeutende Aussagen A und B (d.h. $A \equiv B$, egal ob aussagenlogisch oder prädikatenlogisch) die Aussage $A \Leftrightarrow B$ *immer wahr* ist. Somit können alle bekannten Umformungsregeln für Aussagen (de Morgan, etc.) als Beweisschritte verwendet werden. In diesem Licht ist ($\Leftrightarrow$-beweisen) gar keine eigene Regel, weil es nur die Regel ($\Leftrightarrow$-wissen) für die gleichbedeutenden Aussagen $A \Leftrightarrow B$ und $(A \Rightarrow B) \wedge (B \Rightarrow A)$. Die Regel ($\Leftrightarrow$-wissen) erlaubt uns daher auch, $A \vee B$ zu beweisen, indem wir B unter der Annahme $\neg A$ beweisen (warum?). Die Regel ($\Leftrightarrow$-wissen) ist es auch, die das Expandieren von Definitionen zum Ersetzen von atomaren Aussagen rechtfertigt.

Wir betrachten ein weiteres einfaches Beispiel und führen dazu zwei neue Definitionen ein.

DEFINITION 1.19

Eine ganze Zahl a ist *gerade*, genau dann, wenn sie mit einer ganzen Zahl b als $a = 2b$ geschrieben werden kann.

DEFINITION 1.20

Eine ganze Zahl a ist *ungerade*, genau dann, wenn sie mit einer ganzen Zahl b als $a = 2b + 1$ geschrieben werden kann.

Und hier der Satz, den wir beweisen wollen:

SATZ 1.21

Wenn die Zahl n ungerade ist, dann ist auch die Zahl n^2 ungerade.

Beweis. Wir nehmen an, n sei ungerade. Nach Definition gilt dann mit einer ganzen Zahl k, dass $n = 2k + 1$.

Zu zeigen ist, dass n^2 ungerade ist, also mit einer ganzen Zahl m als $n^2 = 2m + 1$ geschrieben werden kann. Nun gilt

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2\underbrace{(2k^2 + 2k)}_{=:m} + 1.$$

Da k eine ganze Zahl ist, ist auch $m = 2k^2+2k$ eine ganze Zahl, und somit ist n^2 ungerade. □

Quantoraussagen

Syntax: Ist x eine Variable und A eine beliebige Aussage, dann sind

$$\forall x : A \qquad\qquad \exists x : A$$

ebenfalls zulässige Aussagen, sogenannte *Quantoraussagen*. Jedes Auftreten der Variable x in $\forall x : A$ bzw. $\exists x : A$ heißt *gebunden*, alle anderen Variablen heißen *frei*.

Semantik:

- Die Aussage $\forall x : A$ ist „w“ genau dann, wenn für alle Belegungen der Variable x die Bedeutung von A gleich „w“ ist.
- Die Aussage $\exists x : A$ ist w genau dann, wenn es eine Belegung der Variable x gibt, sodass die Bedeutung von A gleich „w“ ist.

Beweisregeln: Wir werden für jeden Typ von Quantoraussage Regeln kennenlernen, wie diese Aussagen zu beweisen sind oder als Teil der Wissensbasis zu behandeln sind.

BEISPIEL 1.22: SYNTAX

Es seien die Voraussetzungen wie in Beispiel 1.12. Die Aussagen

$$\forall x : x < y \Rightarrow \neg ist\_prim(x) \qquad\qquad \exists x : x < 3 \wedge ist\_prim(x) \tag{1.7}$$

sind allesamt Quantoraussagen. Der Ausdruck $\forall x : x + 3$ ist syntaktisch inkorrekt. Der

erste Ausdruck enthält eine gebundene Variable x und eine freie Variable y. Der zweite Ausdruck enthält eine gebundene Variable x und keine freien Variablen.

Beispiel 1.23: Semantik

Um die Bedeutung von $\forall x : x < y \Rightarrow \neg ist\_prim(x)$ unter den Belegung $[y \mapsto 2]$ bzw. $[y \mapsto 4]$ zu bestimmen, muss die Bedeutung von $x < y \Rightarrow \neg ist\_prim(x)$ *für alle Belegungen* der Variable x studiert werden.

	$x \mapsto 0$	$x \mapsto 1$	$x \mapsto 2$	$x \mapsto 3$	$x \mapsto 4$	$\ldots$
$y \mapsto 2$	w ⇒ w ≡ w	w ⇒ w ≡ w	f ⇒ f ≡ w	f ⇒ f ≡ w	f ⇒ w ≡ w	f ⇒ $\ldots$ ≡ w
$y \mapsto 4$	w ⇒ w ≡ w	w ⇒ w ≡ w	w ⇒ f ≡ f	w ⇒ f ≡ f	f ⇒ w ≡ w	f ⇒ $\ldots$ ≡ w

Unter der Belegung $[y \mapsto 2]$ ist die Aussage also wahr, unter $[y \mapsto 4]$ ist sie falsch. Betrachten wir nun $\exists x : x < 3 \wedge ist\_prim(x)$. Da diese Aussage keine freien Variablen enthält, ist ihre Bedeutung unabhängig von jeglichen Variablenbelegungen. Tatsächlich ist sie wahr, weil es eine Belegung für x gibt, nämlich $[x \mapsto 2]$, sodass $x < 3 \wedge ist\_prim(x)$ wahr wird.

	$x \mapsto 0$	$x \mapsto 1$	$x \mapsto 2$	$x \mapsto 3$	$\ldots$
$x < 3 \wedge ist\_prim(x)$	w ∧ f ≡ f	w ∧ f ≡ f	w ∧ w ≡ w	f ∧ w ≡ f	f ∧ $\ldots$ ≡ f

Beispiel 1.24

Was sind die gebundenen bzw. die freien Variablen der folgenden Aussagen:

- $x > 3 \wedge y < 0 \Rightarrow -y^2 \leq x$
- $\forall x : x > 3 \wedge y > 5 \Rightarrow x + y \geq 10$
- $\forall x : \text{reell}(x) \wedge -1 \leq x \leq 1 \Rightarrow x^2 \geq 1$

Die Schreibweisen $A \equiv B$ bzw. $A \models B$ werden aus der Aussagenlogik sinngemäß übernommen. $A \equiv B$ bedeutet, dass A und B in allen Situationen die gleiche Bedeutung haben, $A \models B$ heißt, dass B in allen Situationen wahr ist, in denen auch A wahr ist. $\{A_1, \ldots, A_n\} \models B$ bedeutet, dass B in allen Situationen wahr ist, in denen auch alle A_1 bis A_n wahr sind. „In allen Situationen“ heißt in der Prädikatenlogik: bezüglich aller denkbaren Grundbereiche, in allen Interpretationen und unter allen möglichen Variablenbelegungen, siehe Semantik auf Seite 8. Im Gegensatz zur Aussagenlogik kann $A \equiv B$ und $A \models B$ für beliebige Aussagen A und B in der Prädikatenlogik nicht mehr auf Basis von Wahrheitstabellen begründet werden, da „in allen Situationen“ nun typischerweise unendlich viele Situationen sind.

In Beispiel 1.23 haben wir gesehen, dass für die Beurteilung der Bedeutung einer Allaussage der Form $\forall x : C \Rightarrow A$ nur diejenigen x relevant sind, für die C erfüllt ist, für alle anderen Belegungen von x wird $C \Rightarrow A$ trivialerweise wahr. Analog dazu verhält es sich in Existenzaussagen der Form $\exists x : C \wedge A$. Diejenigen x, für die C erfüllt ist, sind entscheidend, für

alle anderen Belegungen von x wird $C \wedge A$ trivialerweise falsch. Die Rolle von C ist also in beiden Fällen die einer *einschränkenden Bedingung* an die Laufvariable des Quantors. Da dies sehr häufig gebraucht wird, sind dafür eigene Schreibweisen vereinbart, nämlich

$$\forall C_x : A_x \quad \text{für } \forall x : C_x \Rightarrow A_x \qquad\qquad \exists C_x : A_x \quad \text{für } \exists x : C_x \wedge A_x$$

Wie werden Quantoraussagen negiert? Die Aussage $\neg\forall x : A$ ist genau dann wahr, wenn $\forall x : A$ falsch ist. Das bedeutet eben, dass *nicht für alle* x die Aussage A gilt, somit für *mindestens ein* x die Aussage A *nicht* gilt. Die Aussage $\neg\exists x : A$ ist genau dann wahr, wenn $\exists x : A$ falsch ist. Das bedeutet dann, dass es *kein* x gibt, für das die Aussage A gilt, somit muss *für alle* x die Aussage A *nicht* gelten.

LEMMA 1.25: NEGATION VON QUANTORAUSSAGEN

$$\neg\forall\, x : A \equiv \exists\, x \colon \neg A \qquad\qquad \neg\exists\, x : A \equiv \forall\, x \colon \neg A$$

Bei der Negation von quantifizierten Aussagen mit einschränkender Bedingung ist zu beachten, dass die Bedingung *nicht* verändert wird, also

$$\neg\forall\, C : A \equiv \exists\, C : \neg A \qquad\qquad \neg\exists\, C : A \equiv \forall\, C : \neg A.$$

BEISPIEL 1.26

Sei A die Aussage $\forall\ \text{reell}(x) : x^2 < 0$. Wir bestimmen die Negation dieser Aussage:

$$\begin{aligned}
\neg A &\equiv \neg\left(\forall\, x, \text{reell}(x) \colon x^2 < 0\right) \\
&\equiv \neg\left(\forall\, x \colon \text{reell}(x) \Rightarrow x^2 < 0\right) \\
&\equiv \exists\, x \colon \neg\left(\text{reell}(x) \Rightarrow x^2 < 0\right) \\
&\equiv \exists\, x \colon \neg\left(\neg(\text{reell}(x)) \vee (x^2 < 0)\right) \\
&\equiv \exists\, x \colon \text{reell}(x) \wedge \neg(x^2 < 0) \\
&\equiv \exists\, x \colon \text{reell}(x) \wedge x^2 \geq 0 \\
&\equiv \exists\ \text{reell}(x) \colon x^2 \geq 0.
\end{aligned}$$

Der quantorfreie Teil dieser Aussage ist zum Beispiel für $[x \mapsto 0]$ erfüllt, d.h., $\neg A$ ist wahr, damit ist die ursprüngliche Aussage falsch.

Quantoren gleichen Typs (Allquantor, Existenzquantor) können in ihrer Reihenfolge vertauscht werden, der Quantor wird oft nur einmal hingeschrieben, d.h.

$$\begin{aligned}
&\forall x : \forall y : A \equiv \forall y : \forall x : A && \text{kurz: } \forall x, y : A \\
&\exists x : \exists y : A \equiv \exists y : \exists x : A && \text{kurz: } \exists x, y : A
\end{aligned}$$

Die Reihenfolge von *Quantoren verschiedenen Typs* darf *nicht vertauscht* werden.

BEISPIEL 1.27

Betrachten wir die Aussage, dass es zu jeder natürlichen Zahl eine größere natürliche Zahl gibt, d.h. mit einem Prädikatensymbol „nat“

$$\forall\ \text{nat}(x)\ \exists\ \text{nat}(y)\colon x < y.$$

Die Aussage ist *wahr*, weil es *für jede Belegung* von x eine Belegung von y gibt, sodass $\text{nat}(y) \wedge x < y$ wahr wird. Wir müssen nur solche Belegungen $[x \mapsto \bar{x}]$ von x in Betracht ziehen, bei denen $\text{nat}(\bar{x})$ wahr ist, andernfalls ist die Aussage trivialerweise wahr (warum?). Als Belegung von y können wir etwa $[y \mapsto \bar{x} + 1]$ nehmen, unter dieser Belegung wird aus $\text{nat}(y) \wedge x < y$ die *wahre Aussage* $\text{nat}(\bar{x} + 1) \wedge \bar{x} < \bar{x} + 1$.

Jetzt vertauschen wir die Reihenfolge der Quantoren und betrachten die Aussage

$$\exists\ \text{nat}(y)\ \forall\ \text{nat}(x)\colon x < y.$$

In Worten: Es gibt eine natürliche Zahl, die größer ist als jede natürliche Zahl ist. Anders gesagt: es gibt eine größte natürliche Zahl. Das ist nicht nur eine andere Aussage als oben, es klingt auch falsch. Um zu zeigen, dass diese Aussage tatsächlich falsch ist, zeigen wir, dass die Negation wahr ist. Wir negieren die Aussage schrittweise:

$$\begin{aligned}\neg\,(\exists\ \text{nat}(y)\ \forall\ \text{nat}(x)\colon x < y) &\equiv \forall\ \text{nat}(y)\colon \neg\,(\forall\ \text{nat}(x)\colon x < y)\\ &\equiv \forall\ \text{nat}(y)\colon \exists\ \text{nat}(x)\colon \neg\,(x < y)\\ &\equiv \forall\ \text{nat}(y)\colon \exists\ \text{nat}(x)\colon x \geq y.\end{aligned}$$

Der Beweis dieser Aussage kann analog zum obigen Beweis geführt werden.

Um zwischen „es existiert (mindestens) ein“ und „es existiert *genau* ein“ zu unterscheiden wird der Quantor $\exists!$ verwendet, z.B.,

$$\exists!\ \text{reell}(x)\colon x = -x^2.$$

Wir benötigen keine eigenen Regeln für diesen „neuen Quantor“, da $\exists! x : A_x$ lediglich eine Kurzschreibweise ist für

$$\exists x : (A_x \wedge \forall y : A_{x \to y} \Rightarrow (y = x)) \qquad \text{oder auch}$$
$$\exists x : A_x \wedge \forall x, y : A_x \wedge A_{x \to y} \Rightarrow (y = x)$$

konkret in obigem Beispiel

$$\begin{aligned}\exists!\ \text{reell}(x)\colon x = -x^2 &\equiv \exists\ \text{reell}(x)\colon (x = -x^2 \wedge \forall \text{reell}(y) : ((y = -y^2) \Rightarrow (y = x)))\\ &\equiv \exists\ \text{reell}(x)\colon x = -x^2 \wedge \forall \text{reell}(x, y) : ((x = -x^2) \wedge (y = -y^2) \Rightarrow (y = x))\end{aligned}$$

BEISPIEL 1.28

Formulieren Sie die folgenden Aussagen mit Quantoren:

(a) Es gibt eine ganze Zahl, deren Quadrat 16 ist.

(b) Alle Katzen sind grau.

(c) Reelle Zahlen sind genau dann größer als eins, wenn ihre Kuben größer als eins sind.

(d) Es gibt natürliche Zahlen, die gerade sind.

Antwort zu (d): Wir führen Prädikatensymbole „nat“ und „gerade“ ein. Dann können wir schreiben

$$\exists\, x\colon \mathrm{nat}(x) \wedge \mathrm{gerade}(x)$$

bzw. in Kurzschreibweise

$$\exists \mathrm{nat}(x)\colon \mathrm{gerade}(x) \qquad \text{oder auch} \qquad \exists \mathrm{gerade}(x)\colon \mathrm{nat}(x).$$

BEISPIEL 1.29

Gegeben ist die folgenden Aussage: „Jede natürliche Zahlen ist gerade oder ungerade, aber nicht beides“. (a) Geben Sie die Aussage mit Quantoren an. (b) Negieren Sie die quantifizierte Aussage. (c) Formulieren Sie die Negation als Satz in der Umgangssprache.

Man beachte, dass Quantoren oft versteckt vorkommen, wie zum Beispiel in der Aussage: „Sei n eine natürliche Zahl. Dann kann n als Produkt zweier natürlicher Zahlen geschrieben werden“. Gemeint ist, dass *jede natürliche Zahl* als Produkt zweier natürlicher Zahlen geschrieben werden kann. In Quantorenschreibweise:

$$\forall\, \mathrm{nat}(n)\ \exists \mathrm{nat}(a,b) : n = a \cdot b.$$

Oft muss der Laufbereich der gebundenen Variablen (i.e. die Angabe des Grundbereichs, siehe Seite 8) aus dem Kontext abgelesen werden: Für natürliche Zahlen gilt

$$\forall\, n\ \exists a, b : n = a \cdot b.$$

Wir betrachten ein konkretes Beispiel und definieren dazu zuerst den Begriff der *Teilbarkeit*.

DEFINITION 1.30

Seien a, b ganze Zahlen. Wir sagen, dass a *teilt* b genau dann, wenn eine ganze Zahl q existiert, sodass $b = aq$. In Zeichen $a \mid b$.

SATZ 1.31

Für ganze Zahlen x, y, z gilt: wenn z sowohl x als auch y teilt, dann teilt z auch die Summe $x + y$ und das Produkt xy.

Mit Quantoren angeschrieben lautet die Aussage:

$$\forall\, x, y, z\colon (z \mid x \ \wedge z \mid y) \Rightarrow (z \mid x + y \ \wedge \ z \mid xy).$$

Beweis. Seien x, y, z beliebige aber fixe ganze Zahlen. Angenommen $z \mid x$ und $z \mid y$. Zu zeigen ist, dass z dann auch $x + y$ und $x \cdot y$ teilt, laut Definition müssen also ganze Zahlen $\hat{q}$ und $\tilde{q}$ gefunden werden mit

$$x + y = z\hat{q} \qquad\qquad xy = z\tilde{q}. \tag{1.8}$$

Aus $z \mid x$ folgt nach Definition, dass eine ganze Zahl q existiert mit $x = zq$. Wir können also eine ganzzahlige Konstante $\bar{q}$ neu wählen und nehmen an $x = z\bar{q}$. Analog folgt aus $z \mid y$, dass eine neue ganzzahlige Konstante q' gewählt werden kann mit $y = zq'$. Dann gilt

$$x + y = z\bar{q} + zq' = z\underbrace{(\bar{q} + q')}_{=:\hat{q}} \qquad\qquad xy = z\bar{q} \cdot zq' = z\underbrace{(\bar{q}zq')}_{=:\tilde{q}}.$$

Weil $\bar{q}$ und q' ganze Zahlen sind, sind auch $\bar{q} + q'$ und $\bar{q}zq'$ ganzzahlig, die Wahlen $\hat{q} := \bar{q} + q'$ und $\tilde{q} := \bar{q}zq'$ erfüllen daher alle Forderungen in (1.8). □

Übung: Beweisen Sie, dass für alle ganzen Zahlen n gilt:

1. n ist genau dann gerade, wenn n^2 gerade ist.
2. n ist genau dann ungerade, wenn n^2 ungerade ist.

Wir können nun unseren Widerspruchsbeweis aus Beispiel 1.16, dass sich $\sqrt{2}$ nicht als gekürzte Bruchzahl darstellen lässt, zu Ende bringen.

BEISPIEL 1.32

$\neg\exists p, q\colon \sqrt{2} = \frac{p}{q} \wedge p$ und q enthalten außer 1 keine gemeinsamen Faktoren.

Beweis. Angenommen es existieren solche p und q. Seien nun p und q solche ganze Zahlen mit $\sqrt{2} = \frac{p}{q}$ so, dass p und q außer 1 keine gemeinsamen Faktoren enthalten. Das bedeutet aber, dass $2 = \frac{p^2}{q^2}$ ist, und damit $p^2 = 2q^2$. Das hat zur Folge, dass p^2 gerade ist, was nur dann der Fall ist, wenn auch p gerade ist, also als $p = 2m$ geschrieben werden kann. Damit ist

$$2 = \frac{p^2}{q^2} = \frac{4m^2}{q^2},$$

woraus sich $q^2 = 2m^2$ ergibt. Dies bedeutet aber, dass q^2, und damit auch q selbst,

gerade Zahlen sind. Somit haben p und q aber mindestens einen gemeinsamen Faktor 2 im Widerspruch zu obiger Annahme! □

Beweis durch Fallunterscheidung

Die Technik eines Beweises mittels Fallunterscheidung haben wir schon kennengelernt im Fall, dass ein Disjunktion $A \vee B$ in der Wissensbasis ist. Eine andere Situation, in der eine Fallunterscheidung gemacht werden kann, ist, wenn Begriffe durch Fallunterscheidung definiert sind. Wir betrachten ein Beispiel und führen dazu zuerst den Absolutbetrag ein.

DEFINITION 1.33

Sei x eine reelle Zahl. Dann definieren wir den Absolutbetrag $|x|$ als

$$|x| := \begin{cases} x, & x \geq 0, \\ -x, & x < 0. \end{cases}$$

SATZ 1.34: DREIECKSUNGLEICHUNG

Für alle reellen Zahlen x, y gilt:

$$|x + y| \leq |x| + |y|.$$

Beweis. Die Aussage ist für alle reellen Zahlen x, y zu zeigen, es seien daher x, y beliebige aber fixe reelle Zahlen. Nun wissen wir aber

$$(x \geq 0 \vee x < 0) \wedge (y \geq 0 \vee y < 0),$$

was aufgrund des Distributivgesetzes aber gleichbedeutend ist mit

$$(x \geq 0 \wedge y \geq 0) \vee (x \geq 0 \wedge y < 0) \vee (x < 0 \wedge y \geq 0) \vee (x < 0 \wedge y < 0).$$

Wir unterscheiden daher 4 Fälle:

1. Angenommen $x \geq 0$ und $y \geq 0$. Dann gilt $|x| = x$ und $|y| = y$. Außerdem gilt $x + y \geq 0$ und damit gilt $|x+y| = x+y$ nach Definition des Absolutbetrags. Zusammengefasst gilt daher
$$|x + y| \leq |x + y| = x + y = |x| + |y|.$$
Damit ist die Dreiecksungleichung erfüllt.

2. Angenommen $x \geq 0$ und $y < 0$. Damit gilt $|x| = x$ und $|y| = -y$. Um für $|x + y|$ ebenfalls die Definition einsetzen zu können, erinnern wir uns, dass natürlich auch $x + y \geq 0 \vee x + y < 0$ gilt, wir müssen eine weitere Fallunterscheidung treffen.

(a) Angenommen $x + y \geq 0$, damit ist $|x + y| = x + y$. Für die Dreiecksungleichung ist zu zeigen, dass $|x + y| \leq |x| + |y|$, also $x + y \leq x - y$, also $y \leq -y$, was wegen $y < 0$ und $-y > 0$ erfüllt ist.

(b) Angenommen $x + y < 0$ und damit $|x + y| = -(x + y) = -x - y$. Zu zeigen ist wieder $|x + y| \leq |x| + |y|$, also $-x - y \leq x - y$, also $-x \leq x$, was wegen $x \geq 0$ und $-x \leq 0$ erfüllt ist.

3. Angenommen $x < 0 \wedge y \geq 0$: analog (Übung).

4. Angenommen $x < 0$ und $y < 0$. Dann gilt $|x| = -x$ und $|y| = -y$. Außerdem gilt $x + y < 0$ und damit $|x + y| = -(x + y) = -x - y$. Zusammengefasst folgt damit

$$|x + y| \leq |x + y| = -x - y = (-x) + (-y) = |x| + |y|.$$

Damit ist die Dreiecksungleichung erfüllt.

□

Übung: Nehmen Sie obigen Beweis von $\sqrt{2} \notin \mathbb{Q}$ aus Beispiel 1.32 und begründen Sie jeden Beweisschritt mit einer der zur Verfügung stehenden Beweisregeln. Es kann auch sein, dass kleine Schritte ergänzt werden müssen bzw. einzelne Schritte durch eine Aneinanderreihung mehrerer Regeln begründbar sind.

Spezielle Schreibweisen

Wir führen nun die *Summen- und Produktschreibweise* ein.

DEFINITION 1.35: SUMME, PRODUKT

Es sei a_k ein beliebiger Term mit freier Variable k und es seien l, n ganze Zahlen, dann ist:

$$\sum_{k=l}^{n} a_k := \begin{cases} 0, & n < l \\ \left(\sum_{k=0}^{n-1} a_k\right) + a_{k\to n}, & n \geq l \end{cases}$$

$$\prod_{k=l}^{n} a_k := \begin{cases} 1, & n < l \\ \left(\prod_{k=0}^{n-1} a_k\right) \cdot a_{k\to n}, & n \geq l. \end{cases}$$

Vom Standpunkt der Logik gesehen, handelt es sich bei $\sum$ und $\prod$ um spezielle Quantoren, die die Variable k binden.

BEISPIEL 1.36

Einfache Beispiele sind:

$$\sum_{k=0}^{5} k = \Big(\sum_{k=0}^{4} k\Big) + 5 = \ldots = 0+1+2+3+4+5.$$

$$\sum_{k=1}^{7} k^2 = \Big(\sum_{k=1}^{6} k^2\Big) + 7^2 = \ldots = 1^2+2^2+3^2+4^2+5^2+6^2+7^2.$$

$$\sum_{k=0}^{n} (2k+1) = \Big(\sum_{k=0}^{n-1} (2k+1)\Big) + (2n+1) = \ldots = 1+3+\cdots+(2n+1).$$

$$\prod_{k=0}^{5} k = \Big(\prod_{k=0}^{4} k\Big) \cdot 5 = \ldots = 0 \cdot 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 0.$$

$$\prod_{k=1}^{7} k^2 = \Big(\prod_{k=1}^{6} k^2\Big) \cdot 7^2 = \ldots = 1^2 \cdot 2^2 \cdot 3^2 \cdot 4^2 \cdot 5^2 \cdot 6^2 \cdot 7^2.$$

$$\prod_{k=1}^{n} (2k+1) = \Big(\prod_{k=1}^{n-1} (2k+1)\Big) \cdot (2n+1) = \ldots = 1 \cdot 3 \cdot \cdots \cdot (2n+1).$$

Was ist $\sum_{k=1}^{n} 1$? Was $\sum_{k=0}^{n} 1$? Was ist $\prod_{k=1}^{n} a$?

KAPITEL 2

MENGEN

MENGENBILDUNG

Mengen sind die grundlegenden Bausteine zum Aufbau der Mathematik. Eine Menge ist eine *Zusammenfassung von Objekten*, wobei klar sein muss, ob ein Objekt zur Menge gehört oder nicht. Beispiele für Mengen sind

$$\{1, 3, 7, 12, 47\} \qquad \{\text{rot}, \text{schwarz}, \text{blau}\} \qquad \{\Delta, \bigcirc, \spadesuit, \square, \diamondsuit\}$$

Diesen Beispielen ist gemeinsam, dass die Mengen *aufzählend* – d.h. durch Aufzählung aller Objekte, die zur Menge gehören – angegeben sind. Mengen können auch *beschreibend* angegeben werden. Zum Beispiel können wir die Menge der geraden natürlichen Zahlen beschreibend angeben als

$$\{m \mid \text{nat}(m) \wedge m \text{ ist gerade}\}.$$

Des weiteren wird es sich als praktisch herausstellen, Mengen durch ein *Erzeugungsprinzip* festlegen zu können, z.B.

$$\{3t - 1 \mid \text{nat}(t) \wedge t < 5\}$$

für „die Menge aller Elemente x der Form $x = 3t - 1$, wobei t eine natürliche Zahl ist mit $t < 5$“.

Die Objekte einer Menge M heißen *Elemente* von M. Wir schreiben $m \in M$ für „m ist ein Element der Menge M“ und $m \notin M$ für „m ist *kein* Element der Menge M“.

BEISPIEL 2.1: RUSSELL'SCHES PARADOXON

Sei $R := \{x \mid x \notin x\}$. Gilt $R \in R$ oder gilt $R \notin R$?

- Angenommen, es wäre $R \in R$. Dann muss R die charakterisierende Eigenschaft von R erfüllen, d.h. $R \notin R$. ↯
- Angenommen, es wäre $R \notin R$. Dann erfüllt R die charakterisierende Eigenschaft von R, also $R \in R$. ↯

Es gilt also gleichermaßen $R \in R$ wie auch das logische Gegenteil $R \notin R$, ein logischer Widerspruch in der Theorie der Mengenlehre!

Ein logischer Widerspruch in der Theorie der Mengenlehre? Das kann nicht sein[1]. Das Beispiel von Russell zeigt lediglich, dass man es sich nicht so leicht machen kann. Die naive Auffassung nämlich, dass man

1. zu jeder Eigenschaft P_x die Menge $\{x \mid P_x\}$ bilden kann, und dass
2. $y \in \{x \mid P_x\}$ genau dann gilt, wenn $P_{x \to y}$ erfüllt ist[2],

führt zu den oben geschilderten Problemen. Um diesen aus dem Weg zu gehen, müssen in Mengen der Form $\{x \mid P_x\}$ bestimmte Einschränkungen an die Gestalt von P_x in Kauf genommen werden[3]. Wir gehen diesen theoretischen Überlegungen nicht weiter auf den Grund, vielmehr wollen wir ab nun die gängigen Mengen und Mengenbildungsprozesse vorstellen und verlassen uns darauf, dass diese in der zugrundeliegenden Mengenlehre auch tatsächlich abgesichert sind. Bei allen zukünftig vorgestellten Mengen der Form $\{x \mid P_x\}$ gilt

$$y \in \{x \mid P_x\} \quad \text{genau dann, wenn} \quad P_{x \to y}.$$

DEFINITION 2.2: AUFZÄHLUNG, AUSSONDERUNG, ERSETZUNG

Endliche Mengen können durch *Aufzählung* der Elemente angegeben werden, d.h.

$$\{e_1, \ldots, e_n\} := \{x \mid x = e_1 \vee \ldots \vee x = e_n\}.$$

Mengen können durch *Aussonderung* aus einer schon bestehenden Menge gebildet werden, siehe Abbildung 2.1, d.h.

$$\{x \in A \mid P_x\} := \{x \mid x \in A \wedge P_x\}.$$

Mengen können auch durch *funktionale Ersetzung* aus bestehenden Mengen gebildet werden, d.h.

$$\{t_x \mid x \in A \wedge P_x\} := \{y \mid \exists x \colon x \in A \wedge P_x \wedge y = t_x\}.$$

[1] Fußnoten unterbrechen den Lesefluss, das wissen wir! Deshalb setzen wir sie auch so selten ein wie möglich. Wir verwenden Fußnoten für weiterführende Erklärungen, die der mathematisch interessierten Leserin zur Erbauung gereichen, dem unmittelbaren Verständnis des Stoffes aber nicht notwendigerweise dienen. In diesem Fall geht es um *Widersprüche in mathematischen Theorien*. Was ist das Schlimme daran? Ein Widerspruch ist eine falsche Aussage. Die Regeln der Logik – wir werden darauf in Kapitel 1 zurückkommen – besagen, dass aus einer falschen Aussage *jede beliebige Aussage* hergeleitet werden kann. Eine solche Theorie, in der alles wahr ist, ist aber wertlos.

[2] Der Index x in P_x soll andeuten, dass in P_x die Variable x vorkommt. $P_{x \to y}$ steht für P mit jedem x ersetzt durch y.

[3] Die genaue Untersuchung, welche Mengenbildungen erlaubt sind und welche nicht, ist Aufgabe der (axiomatischen) Mengenlehre. Dort wird durch Axiome festgelegt, von welchen Mengen man als existent ausgeht, und durch welche Mengenbildungsprozesse neue Mengen gebildet werden dürfen. Ein solches Axiomensystem bildet etwa die *Zermelo-Fraenkel Mengenlehre*.

Der Ausdruck „funktionale Ersetzung“ kommt daher, dass durch einen Term t_x eine Funktion $f : x \mapsto t_x$ beschrieben ist. Die Menge $\{t_x \mid x \in A \wedge P_x\}$ entspricht dann genau der Menge aller Funktionswerte $f(x)$, es werden also alle x durch die Funktionswerte $f(x)$ ersetzt, siehe Abbildung 2.1.

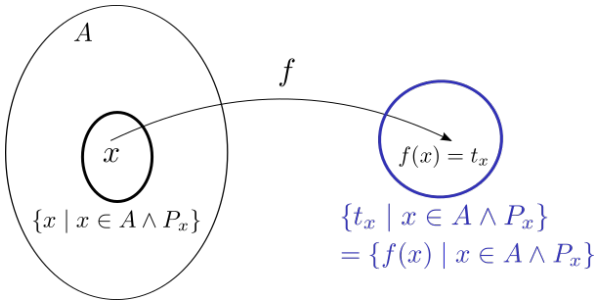


Abbildung 2.1: Mengenbildung durch Aussonderung und funktionale Ersetzung

BEISPIEL 2.3

Die Menge $\{\text{rot}, \text{schwarz}, \text{blau}\}$ ist eine durch Aufzählung gebildete endliche Menge. Laut Definition ist

$$\text{rot} \in \{\text{rot}, \text{schwarz}, \text{blau}\}, \text{ weil } \underline{\text{rot=rot}} \text{ (oder rot=schwarz oder rot=blau).}$$
$$\text{gelb} \notin \{\text{rot}, \text{schwarz}, \text{blau}\}, \text{ weil weder gelb=rot noch gelb=schwarz noch gelb=blau.}$$

BEISPIEL 2.4: NATÜRLICHE ZAHLEN

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

Die natürlichen Zahlen sind eine *unendliche Menge*, die wir ausnahmsweise durch „Aufzählen der Elemente“ symbolisieren[a]. Die Menge der natürlichen Zahlen $\mathbb{N}$ beginnt also mit 1, benötigen wir die Menge der natürlichen Zahlen beginnend mit 0, so verwenden wir

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

[a] Aufzählen geht nur bei endlichen Mengen. Wir gehen allerdings davon aus, dass wir eine genügend gute intuitive Vorstellung der natürlichen Zahlen haben, ohne diese durch eine exakte mathematische Definition einzuführen. In der „echten Mengenlehre“ ist die Existenz einer unendlichen Menge entsprechend den natürlichen Zahlen durch ein nicht-triviales Axiom gesichert. Wir ersparen uns die Details. Es sollte uns aber schon bewußt sein, dass „...“ jeden mathematischen Ausdruck ungenau macht und deswegen wann immer möglich vermieden werden sollte!

BEISPIEL 2.5

In einigen Beispielen haben wir bisher ein Prädikatensymbol „nat“ verwendet, um mit $\text{nat}(n)$ auszudrücken, dass n eine natürliche Zahl ist. Dieses Sachverhalt können wir ab nun auch durch $n \in \mathbb{N}$ beschreiben.

Beispiel 2.6: Singleton, Leere Menge

Eine einelementige Menge $\{e_1\}$ wird oft *Singleton* genannt. Diese ist als Spezialfall mit $n = 1$ durch Definition 2.2 bereits abgedeckt, d.h.

$$\{e_1\} = \{x \mid x = e_1\}.$$

Beachte aber, dass der Fall $n = 0$ in Definition 2.2 *nicht* umfasst ist.

$$\{\} := \{x \in \mathbb{N} \mid x \neq x\}.$$

Ganz egal, welches y wir betrachten, die charakterisierende Eigenschaft ist für y nie erfüllt, da $y \neq y$ immer falsch ist! Man schreibt anstelle von $\{\}$ oft $\emptyset$. Die Menge $\emptyset$ enthält also keine Elemente, man nennt $\emptyset$ deswegen die *leere Menge*. Die leere Menge ist ein Beispiel für eine nach dem Prinzip der Aussonderung gebildete Menge.

Beispiel 2.7

$$\{-n \mid n \in \mathbb{N}\}$$

ist eine durch funktionale Ersetzung gebildete Menge. Es gilt $-1 \in \{-n \mid n \in \mathbb{N}\}$, weil es ein $n \in \mathbb{N}$ gibt mit $-1 = -n$, nämlich $n = 1$. Analog kann argumentiert werden, dass $-2 \in \{-n \mid n \in \mathbb{N}\}$, $-3 \in \{-n \mid n \in \mathbb{N}\}$, etc. Wir können uns diese Menge also „vorstellen“ als die Menge $\{-1, -2, -3, \ldots\}$.

Beispiel 2.8

$$8 \in \{3t - 1 \mid t \in \mathbb{N}_0 \wedge t < 5\},$$

weil es ein $t \in \mathbb{N}_0$ gibt mit $t \in \mathbb{N}_0$ und $t < 5$ und $8 = 3t - 1$, nämlich $t = 3$. Man sieht leicht, dass die Menge $\{3t - 1 \mid t \in \mathbb{N}_0 \wedge t < 5\}$ auch geschrieben werden kann als $\{-1, 2, 5, 8, 11\}$.

Funktionale Ersetzung bedeutet also, dass die Menge $\{t_x \mid x \in A \wedge P_x\}$ entsteht, indem man „die Menge $\{x \mid x \in A \wedge P_x\}$ durchläuft“ und jeweils anstelle von x den Wert t_x in die Menge aufnimmt.

Aber was bedeutet eigentlich „=“ bei Mengen? Es stellen sich Fragen wie „$\{1, 2, 3\} = \{2, 3, 1\}$?“ oder „$\{1, 2, 2\} = \{2, 1\}$?“ Die Antwort ist in beiden Fällen „ja“, d.h. die Reihenfolge der Elemente in einer Menge soll keine Rolle spielen, und jedes Element ist einer Menge „nur einmal enthalten“.

Definition 2.9: Teilmenge, Gleichheit

1. Eine Menge A heißt *Teilmenge* der Menge B (wir schreiben $A \subseteq B$) genau dann, wenn jedes Element der Menge A auch ein Element der Menge B ist, d.h.

$$\forall x \colon x \in A \Rightarrow x \in B.$$

2. Zwei Mengen A, B sind *gleich* (wir schreiben $A = B$) genau dann, wenn $x \in A$ genau dann gilt, wenn $x \in B$ ist, d.h.

$$\forall x \colon x \in A \Leftrightarrow x \in B.$$

Gleiche Mengen enthalten die gleichen Elemente.

Ist $A = B$, dann gilt sowohl $A \subseteq B$ als auch $B \subseteq A$.

BEISPIEL 2.10

$$\{n \in \mathbb{N} \mid n < 5\} \subseteq \mathbb{N} \qquad \{3t - 1 \mid t \in \mathbb{N}_0 \wedge t < 5\} = \{-1, 2, 5, 8, 11\}.$$

Für jede Menge A gilt $A \subseteq A$ und $\emptyset \subseteq A$. Außerdem: falls $A \subseteq B$ und $B \subseteq C$, dann ist $A \subseteq C$.

BEISPIEL 2.11: LEERE MENGE TEILMENGE JEDER MENGE

Die Tatsache $\emptyset \subseteq A$ verdient genauere Betrachtung. Laut Definition bedeutet dies ja

$$\forall x \in \emptyset : x \in A,$$

was wiederum nach Konvention bedeutet

$$\forall x : \underbrace{\underbrace{x \in \emptyset}_{\text{f}} \Rightarrow x \in A}_{\text{w}}.$$

Die Elemente einer Menge können durchaus auch selbst wieder Mengen sein, Mengen können also *verschachtelt sein*, z.B. $\{1, \{2\}, \{\{\{\emptyset\}\}\}\}$. Das *Regularitätsaxiom* sorgt dafür, dass bei solchen Verschachtelungen alles mit rechten Dingen zugeht. Das Axiom selbst ist relativ technisch, für uns wichtig sind zwei Folgerungen aus der Regularität:

1. Keine Menge kann in sich selbst enthalten sein, d.h. $A \notin A$.
2. Es gibt keine Kette $A_1 \in A_2 \in \ldots \in A_n \in A_1$.

Folgerung 1 ist dabei lediglich ein Spezialfall von Folgerung 2 für $n = 1$.

MENGENOPERATIONEN

Wir können aus bestehenden Mengen neue Mengen bilden.

DEFINITION 2.12: VEREINIGUNG, DURCHSCHNITT, DIFFERENZ

Seien A, B Mengen. Dann definieren wir

- die *Vereinigung* von A und B als die Menge, die genau jene Elemente enthält, die in A oder in B liegen in Zeichen: $A \cup B$ („A vereinigt B“);

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

- den *Durchschnitt* von A und B als die Menge, die genau jene Elemente enthält, die sowohl in A als auch in B liegen, in Zeichen $A \cap B$ („A geschnitten B“);

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

- die *Differenzmenge* von A und B als die Menge, die genau jene Elemente enthält, die in A aber nicht in B liegen, in Zeichen $A \backslash B$ („A ohne B“);

$$A \backslash B := \{x \mid x \in A \ \wedge \ x \notin B\}.$$

DEFINITION 2.13: DISJUNKT

Zwei Mengen heißen *disjunkt* genau dann, wenn $A \cap B = \emptyset$.

BEISPIEL 2.14

Seien $A = \{1, 2, 3, 4, 5, 6\}$ und $B = \{4, 5, 6, 7, 8\}$, dann:

$$A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8\} \qquad A \cap B = \{4, 5, 6\}$$
$$A \backslash B = \{1, 2, 3\} \qquad B \backslash A = \{7, 8\}.$$

Die Mengen A und B sind nicht disjunkt.

Die Mengenoperationen können mittels *Venn-Diagrammen* veranschaulicht werden. Auch Eigenschaften von Mengen können manchmal mittels Venn Diagrammen dargestellt werden, siehe Abbildung 2.2.

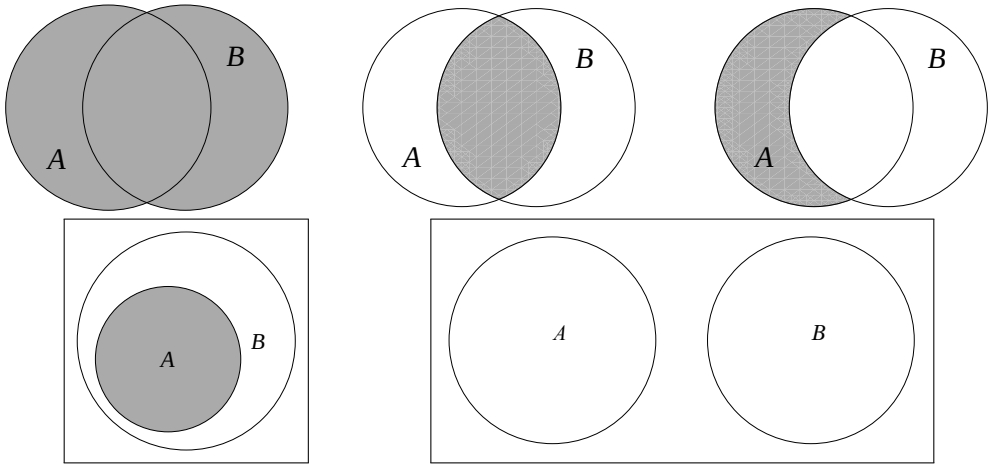


Abbildung 2.2: Oben (von links nach rechts): $A \cup B$, $A \cap B$ und $A \backslash B$.
Unten $A \subseteq B$ (links) bzw. A und B disjunkt (rechts).

Wir sind nun in der Lage, die neben den natürlichen Zahlen „bekannten Zahlenmengen“ einzuführen.

DEFINITION 2.15: $\mathbb{Z}$ UND $\mathbb{Q}$

$$\mathbb{Z} := \mathbb{N}_0 \cup \{-n \mid n \in \mathbb{N}\} \qquad \text{(ganze Zahlen)}$$

$$\mathbb{Q} := \{\frac{p}{q} \mid p \in \mathbb{Z} \wedge q \in \mathbb{N} \wedge p, q \text{ haben keine gemeinsamen Teiler außer } 1\}$$

(rationale Zahlen)

Die ganzen Zahlen stellen wir uns als $\{\dots, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, \dots\}$ vor, die rationalen Zahlen sind genau die Zahlen, die sich als Bruch darstellen lassen[4].

BEISPIEL 2.16

Seien $A = \mathbb{N}$ und $B = \mathbb{Z}$, dann:

$$A \cup B = \mathbb{Z} \qquad A \cap B = \mathbb{N} \qquad A \backslash B = \emptyset \qquad B \backslash A = \{\dots, -3, -2, -1, 0\}.$$

Wir ersparen uns eine exakte Definition der *reellen Zahlen*, wir stellen sie uns als „lückenlose Zahlengerade“ vor und bezeichnen sie mit $\mathbb{R}$. Weiters verwenden wir

$$\mathbb{R}^+ := \{x \in \mathbb{R} \mid x > 0\} \qquad \mathbb{R}^- = \{x \in \mathbb{R} \mid x < 0\}$$
$$\mathbb{R}_0^+ := \{x \in \mathbb{R} \mid x \geq 0\} \qquad \mathbb{R}_0^- := \{x \in \mathbb{R} \mid x \leq 0\}.$$

Analoge Schreibweisen können für $\mathbb{Z}$ und $\mathbb{Q}$ verwendet werden, wir könnten in Beispiel 2.16 auch schreiben $B \backslash A = \mathbb{Z}_0^-$.

[4] Der interessierten Leserin ist natürlich aufgefallen, dass wir nicht definiert haben, was $\frac{p}{q}$ eigentlich bedeutet. Wir verlassen uns hier auf Ihre Vorbildung.

DEFINITION 2.17: INTERVALLE

Das *geschlossene Intervall* $[a, b]$ für reelle Zahlen a, b bezeichnet die Menge aller reellen Zahlen, die größer oder gleich a und kleiner oder gleich b sind, kurz:

$$[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}.$$

Das *offene Intervall* $]a, b[$ für reelle Zahlen a, b bezeichnet die Menge aller reellen Zahlen, die größer als a und kleiner als b sind, kurz:

$$]a, b[:= \{x \in \mathbb{R} \mid a < x < b\}.$$

Analog werden die *halboffenen Intervalle* $]a, b]$ und $[a, b[$ wie folgt definiert

$$]a, b] := \{x \in \mathbb{R} \mid a < x \leq b\} \qquad [a, b[:= \{x \in \mathbb{R} \mid a \leq x < b\}.$$

Offene bzw. halboffene Intervalle werden manchmal auch als (a, b), $(a, b]$ bzw. $[a, b)$ geschrieben.

BEISPIEL 2.18

Seien $I_1 = [0, 5]$, $I_2 = [2.4, 4]$, $I_3 = [3, 7]$. Bestimme (a) $I_2 \cap \mathbb{N}$ (b) $I_1 \cup I_3$ (c) $I_2 \cap I_3$ (d) $I_1 \backslash I_2$.

BEISPIEL 2.19

Seien A, B zwei Mengen, für die gilt: $A \subseteq B$. Was ist (a) $A \cup B$ (b) $A \cap B$ (c) $A \backslash B$ (d) $B \backslash A$?

Für Mengenoperationen gelten ganz ähnliche Gesetze wie für die logischen Junktoren, vergleiche etwa Satz 1.4. Einige Rechenregeln für die Mengenoperationen aus Defintion 2.12 sind im folgenden Satz zusammengefasst. Aus diesen können andere Rechenregeln abgeleitet werden.

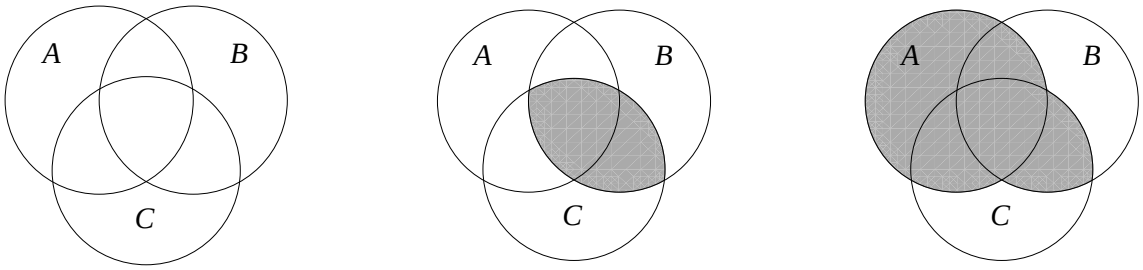
SATZ 2.20

Seien A, B, C Mengen. Dann gilt:

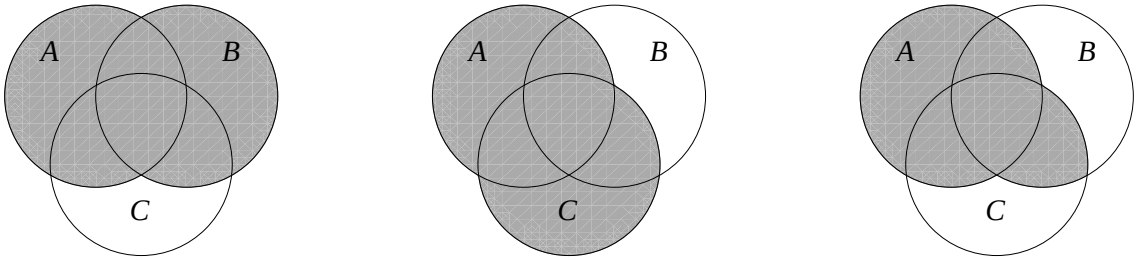
$$\begin{aligned}
A \cup B &= B \cup A & A \cap B &= B \cap A && \text{(Kommutativgesetz)}\\
(A \cup B) \cup C &= A \cup (B \cup C) & (A \cap B) \cap C &= A \cap (B \cap C) && \text{(Assoziativgesetz)}\\
A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) && && \text{(Distributivgesetz)}\\
A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) && && \text{(Distributivgesetz)}\\
A \backslash (B \cap C) &= (A \backslash B) \cup (A \backslash C) & A \backslash (B \cup C) &= (A \backslash B) \cap (A \backslash C) && (2.1)
\end{aligned}$$

Beweis. über Venn-Diagramme: die Venn-Diagramme für die rechte und die linke Seite der Identität müssen übereinstimmen. Wir betrachten den Beweis für das erste Distributivgesetz:

Für die linke Seite beginnen wir mit drei Mengen und zeichnen zunächst $B \cap C$ ein und im zweiten Schritt bilden wir die Vereinigung davon mit A:



Für die rechte Seite bilden wir zunächst die beiden geklammerten Ausdrücke $A \cup B$ und $A \cup C$ und bilden dann den Durchschnitt:



□

BEISPIEL 2.21

Stellen Sie mittels Venn-Diagrammen fest, welche der folgenden Aussagen gelten:

- $(A \cap B) \backslash C = (A \backslash C) \cap (B \backslash C)$
- $(A \backslash B) \backslash C = A \backslash (B \backslash C)$

Oft kann man von einem „fixen Universum“ ausgehen, in dem gerechnet wird. In diesem Fall, wenn also unmissverständlich klar ist, über welcher *fixen Grundmenge* Ω (Universum) gearbeitet wird, dann kann man einfach A^c für $\Omega \backslash A$ schreiben. Die Menge A^c nennt man das *Komplement* von A bzw. die *Komplementärmenge* von A, sie enthält genau diejenigen Elemente (von Ω), die *nicht in* A sind. In dem Sinn, mit $A = \Omega$, können die Identitäten (2.1) aus Satz 2.20 auch so geschrieben werden:

$$(B \cap C)^c = B^c \cup C^c \qquad (B \cup C)^c = B^c \cap C^c. \tag{2.2}$$

DEFINITION 2.22: POTENZMENGE

Sei A eine gegebene Menge. Die *Potenzmenge* von A (in Zeichen: $P(A)$) ist die Menge aller Teilmengen von A:

$$P(A) := \{B \mid B \subseteq A\}.$$

BEISPIEL 2.23

Sei $A = \{1, 2, 3\}$, dann ist

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

Es gilt immer $\emptyset \in P(A)$ und $A \in P(A)$.

Als Gegenstück zu endlichen Mengen $\{e_1, \ldots, e_n\}$ gibt es n-Tupel $(e_1, \ldots, e_n)$. Diese unterscheiden sich von Mengen im Wesentlichen dadurch, dass die Reihenfolge der Elemente entscheidend ist, und dass Elemente auch mehrfach vorkommen können. Dies spiegelt sich in der Definition der Gleichheit von Tupel wider.

$(a_1, \ldots, a_n) = (b_1, \ldots, b_m)$ genau dann, wenn $n = m$ und für alle $1 \leq i \leq n$ gilt $a_i = b_i$.

Anstelle von 2-Tupel sagt man *(geordnetes) Paar*.

DEFINITION 2.24: KARTESISCHES PRODUKT

Sei $n \in \mathbb{N}$ und $n \geq 2$. Die Menge

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

nennt man das *kartesische Produkt* (oder auch die *Produktmenge*) von $A_1, \ldots, A_n$.

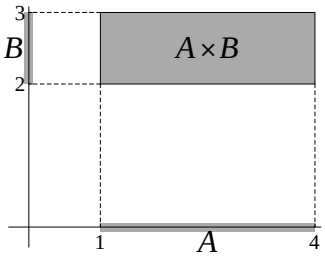
Wenn $A_1 = \cdots = A_n = A$ gilt, dann wird das kartesische Produkt oft kurz mit A^n bezeichnet. Das kartesische Produkt nur einer Menge ist in obiger Definition nicht umfasst, da ja $n \geq 2$ gefordert ist. Sollte man es einmal brauchen, bieten sich folgende Erweiterungen an:

$$A^1 := A \qquad\qquad A^0 := \emptyset.$$

BEISPIEL 2.25

Sei $A = \{a, b, c\}$ und $B = \{1, 2\}$. Dann ist

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



BEISPIEL 2.26

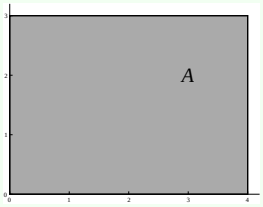
Sei $A = [1, 4]$ und $B = [2, 3]$. Dann ist

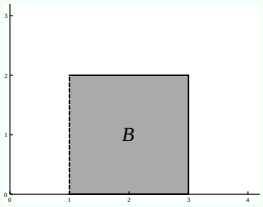
$$A \times B = \{(x, y) \mid 1 \leq x \leq 4 \ \wedge \ 2 \leq y \leq 3\}.$$

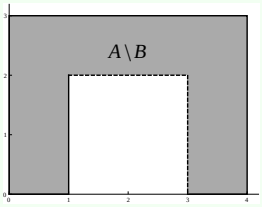
BEISPIEL 2.27

Seien $A = [0,4] \times [0,3]$, $B = (1,3] \times [0,2]$, und $C = [-1,2] \times [0,3]$. Bestimmen Sie die Mengen $A \cap B$, $A \cap C$, $A \cup C$, und $A \setminus B$. Für die Differenzmenge gilt

$$A \setminus B = [0,1] \times [0,3] \cup (3,4] \times [0,3] \cup (1,3] \times (2,3].$$







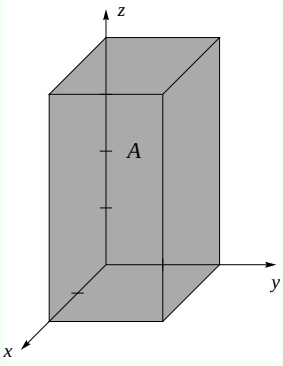
Im Dreidimensionalen lassen sich die Produktmengen auch noch graphisch veranschaulichen.

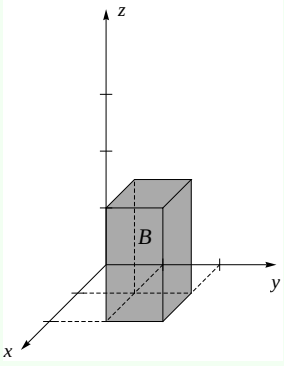
BEISPIEL 2.28

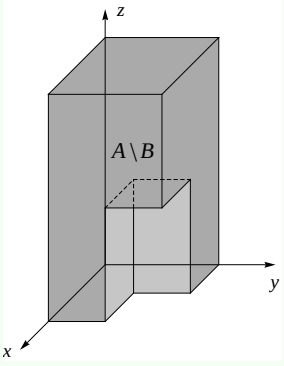
Gegeben sind die Mengen $A = [0,2] \times [0,2] \times [0,4]$ und $B = [1,2] \times [1,2] \times [0,2]$ wie im Bild unten links skizziert. Die Menge $A \setminus B$ ist zum Beispiel gegeben durch

$$A \setminus B = [0,2] \times [0,1) \times [0,4] \cup [0,1) \times [1,2] \times [0,4] \cup [1,2] \times [1,2] \times (2,4],$$

siehe die Skizze ganz rechts. Die Flächen $[1,2] \times \{1\} \times [0,2]$, $\{1\} \times [1,2] \times [0,2]$ und $[1,2] \times [1,2] \times \{2\}$ sind *nicht* in der Menge $A \setminus B$ enthalten.







INDUKTIONSBEWEIS

„Induktion“ ist keine allgemeingültige Beweistechnik der Logik, sondern eine Spezialtechnik, die zum Beweis von Allaussagen über der Menge der natürlichen Zahlen angewendet werden kann[5]. Der Schlüssel dazu ist der spezielle Aufbau der natürlichen Zahlen:

[5] Es gibt auch andere *induktiv aufgebaute Bereiche*, in denen Induktionsbeweise (in etwas modifizierter Form) möglich sind. Wir gehen darauf aber nicht näher ein.

- $0 \in \mathbb{N}_0$ **und**
- für jedes n gilt: $n \in \mathbb{N}_0 \Rightarrow n+1 \in \mathbb{N}_0$.

Damit sind die natürlichen Zahlen eindeutig bestimmt (und können aufzählend generiert werden).

Regel (Induktion in $\mathbb{N}_0$): Um eine Aussage $\forall n \in \mathbb{N}_0 : A$ zu beweisen, reicht es demnach

1. $A_{n\to 0}$ zu beweisen (*Induktionsanfang*) **und**
2. $\forall m \in \mathbb{N}_0 : A_{n\to m} \Rightarrow A_{n\to m+1}$ zu beweisen. Dazu
   (a) wählt man $\bar{n} \in \mathbb{N}_0$ beliebig aber fix,
   (b) *nimmt an*, dass $A_{n\to\bar{n}}$ gilt (*Induktionsannahme*),
   (c) *und zeigt*, dass $A_{n\to\bar{n}+1}$ gilt (*Induktionsschritt*).

SATZ 2.29

Für alle $n \in \mathbb{N}_0$ gilt

$$\sum_{k=0}^{n} 2k = n(n+1).$$

Beweis. Wir führen einen Induktionsbeweis durch:

1. *Induktionsanfang*: Wir zeigen, dass die Aussage für $n = 0$ stimmt.
$$\sum_{k=0}^{0} 2k = 2 \cdot 0 = 0 = 0(0+1).$$
2. *Induktionsannahme*: Wir nehmen an, dass die Aussage für eine beliebige natürliche Zahl $\bar{n}$ gilt, d.h.
$$\sum_{k=0}^{\bar{n}} 2k = \bar{n}(\bar{n}+1). \tag{2.3}$$
3. *Induktionsschritt*: Wir zeigen, dass die Aussage auch gilt, wenn n durch $\bar{n}+1$ ersetzt wird, d.h. wir zeigen:
$$\sum_{k=0}^{\bar{n}+1} 2k = (\bar{n}+1)(\bar{n}+2).$$
Wir beginnen mit der linken Seite und formen schrittweise um. Zuerst kann der letzt Summand aus der Summe herausgezogen werden:
$$\sum_{k=0}^{\bar{n}+1} 2k = \sum_{k=0}^{\bar{n}} 2k + 2(\bar{n}+1) \overset{(2.3)}{=} \bar{n}(\bar{n}+1) + 2(\bar{n}+1) = (\bar{n}+1)(\bar{n}+2).$$

□

BEISPIEL 2.30

Es ist zu beweisen, dass für alle $n \in \mathbb{N}_0$ gilt:

$$\sum_{k=0}^{n}(2k+1) = (n+1)^2. \tag{2.4}$$

1. *Induktionsanfang*:

$$\sum_{k=0}^{0}(2k+1) = 2 \cdot 0 + 1 = 1 = (0+1)^2.$$

2. *Induktionsannahme*: es gelte

$$\sum_{k=0}^{n}(2k+1) = (n+1)^2. \tag{2.5}$$

3. *Induktionsschritt*: Zu zeigen ist:

$$\sum_{k=0}^{n+1}(2k+1) = (n+2)^2.$$

Dazu betrachten wir

$$\begin{aligned}\sum_{k=0}^{n+1}(2k+1) &= \sum_{k=0}^{n}(2k+1) + (2(n+1)+1) \overset{(2.5)}{=} \\ &= (n+1)^2 + (2n+3) = n^2 + 2n + 1 + 2n + 3 = n^2 + 4n + 4 = \\ &= (n+2)^2.\end{aligned}$$

Man beachte, dass wir in Beispiel 2.30 die allgemein übliche Angewohnheit übernommen haben, für die neu gewählte Konstante statt wie bisher $\bar{n}$ nur kurz n zu verwenden, wir haben mittlerweile ja Routine im Beweisen! Das führt dazu, dass die Induktionshypothese (2.5) auf den ersten Blick identisch mit der ursprünglich zu beweisenden Aussage (2.4) ist, d.h. es sieht aus, als nähme man als wahr an, was eigentlich zu beweisen ist, das kann nicht sein! Aufgrund dieses Missverständnisses sind Beginner oft Induktion gegenüber skeptisch. Andererseits führt es auch dazu, dass bei der Induktionsannahme oft geschrieben/gesagt wird: „Jetzt nehmen wir (2.4) an." Da diese Annahme ja die zu beweisende Aussage ist, kann das vom Leser/Hörer natürlich nicht verstanden werden! Aus diesem Grund bevorzugen wir eigentlich die Schreibweise mit $\bar{n}$ anstelle des kürzeren n, Verständnis ist wichtiger als Einsparung von Schreibarbeit!

Der Induktionsanfang muss nicht bei 0 gewählt werden, man kann auch bei jeder anderen

Zahl $l \in \mathbb{N}$ beginnen. Die Aussage ist dann aber nur für alle natürlichen Zahl *ab dem Index* l nachgewiesen.

Regel (Eingeschränkte Induktion in $\mathbb{N}$): Um eine Aussage $\forall n \in \mathbb{N}, n \geq l : A$ zu beweisen, reicht es,

1. $A_{n \to l}$ zu beweisen (*Induktionsanfang*) **und**
2. $\forall m \in \mathbb{N}, m \geq l : A_{n \to m} \Rightarrow A_{n \to m+1}$ zu beweisen. Dazu
   (a) wählt man $\bar{n} \in \mathbb{N}$ mit $\bar{n} \geq l$ beliebig aber fix,
   (b) *nimmt an*, dass $A_{n \to \bar{n}}$ gilt (*Induktionsannahme*),
   (c) *und zeigt*, dass $A_{n \to \bar{n}+1}$ gilt (*Induktionsschritt*).

BEISPIEL 2.31

Es ist zu beweisen, dass für alle $n \in \mathbb{N}$ mit $n \geq 6$ gilt: $7n < 2^n$. Wir verwenden Induktion.

1. *Induktionsanfang*: Für $n = 6$ haben wir $7 \cdot 6 = 42 < 64 = 2^6$.
2. *Induktionsannahme*: Für $\bar{n} \in \mathbb{N}$ mit $\bar{n} \geq 6$ gelte
$$7\bar{n} < 2^{\bar{n}}. \tag{2.6}$$
3. *Induktionsschritt*: Zu zeigen ist $7(\bar{n}+1) < 2^{\bar{n}+1}$. Diese Aussage ist wahr wegen
$$7(\bar{n}+1) = 7\bar{n} + 7 \overset{\bar{n}>1}{<} 7\bar{n} + 7\bar{n} \overset{(2.6)}{<} 2^{\bar{n}} + 2^{\bar{n}} = 2 \cdot 2^{\bar{n}} = 2^{\bar{n}+1}.$$

KAPITEL 3

FUNKTIONEN

DER FUNKTIONSBEGRIFF IN DER MATHEMATIK

Es seien X, Y beliebige Mengen. Unter einer Funktion f von X nach Y wollen wir uns eine „Zuordnung“ vorstellen, die jedem $x \in X$ genau ein $y \in Y$ zuordnet. Das einem $x \in X$ durch die Funktion f zugeordnete Objekt bezeichnet man mit $f(x)$.

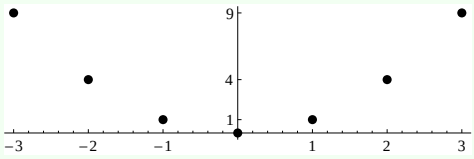
Es gibt verschiedene Möglichkeiten, solch eine „Zuordnung“ zu definieren.

BEISPIEL 3.1

Eine Funktion kann durch eine *Wertetabelle* definiert sein. Es ist dabei nicht von Belang, ob aus der Tabelle ein gesetzmäßiger Zusammenhang zwischen x und y erkennbar ist oder nicht. (Die Tabelle *ist* das Gesetz.)

Wertetabelle:

x	y
-3	9
-2	4
-1	1
0	0
1	1
2	4
3	9



Für die durch diese Tabelle charakterisierte Funktion f gilt:

$$f(-2) = 4, \qquad f(0) = 0, \qquad f(1) = 1, \qquad \text{etc.}$$

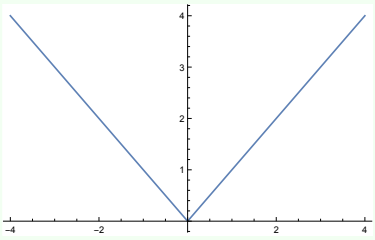
Es handelt sich bei f um eine Funktion von $\{-3, -2, -1, 0, 1, 2, 3\}$ nach $\{0, 1, 4, 9\}$.

BEISPIEL 3.2

Eine Funktion kann auch durch eine Abbildungsvorschrift definiert werden. Unter Zuhilfenahme des Absolutbetrags (siehe Definition 1.33) können wir eine Funktion f definieren, die jedem $x \in \mathbb{R}$ den Wert $y = f(x) = |x| \in \mathbb{R}$ zuordnet. Man schreibt dafür üblicherweise etwas in der Art

$$f\colon \mathbb{R} \to \mathbb{R}, \quad x \mapsto |x|$$

Eine Wertetabelle hat bei einer derartigen Funktion wenig Sinn, da sie ja unendlich viele Werte enthalten müsste, um die Funktion vollständig zu charakterisieren. Ein „Funktionsgraph“ in $\mathbb{R}^2$ ist da schon eher passend, siehe nebenstehendes Bild.



Eine Funktion f von X nach Y ist durch alle Paare (x, y) mit $x \in X$, $y \in Y$ und $y = f(x)$ charakterisiert, in Beispiel 3.1 ist dies genau die Menge

$$f = \{(-3, 9), (-2, 4), (-1, 1), (0, 0), (1, 1), (2, 4), (3, 9)\} \subseteq \{-3, -2, -1, 0, 1, 2, 3\} \times \{0, 1, 4, 9\},$$

in Beispiel 3.2 ist es die Menge

$$f = \{(x, |x|) \mid x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}.$$

DEFINITION 3.3

Seien X, Y Mengen. Eine Menge $f \subseteq X \times Y$ *ist eine Funktion* von X nach Y genau dann, wenn gilt

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

Man schreibt dafür $f\colon X \to Y$. Die Menge X heißt der *Definitionsbereich* von f und die Menge Y der *Zielbereich* von f.

Die Definition einer Funktion durch eine Wertetabelle besagt nun nichts anderes als die Festlegung der Funktion als endliche Paarmenge, eine Definition durch Zuordnungsvorschrift ist nichts anderes als die Festlegung der Funktion als (üblicherweise unendliche) Paarmenge durch funktionale Ersetzung (d.h. Bildungsvorschrift, siehe Definition 2.2), siehe Abbildung 3.1. Abbildung 3.1 zeigt überdies, dass in einer Funktionsdefinition der Art

$$f\colon X \to Y, \quad x \mapsto t_x \tag{3.1}$$

die Angabe der Zielmenge Y eigentlich überflüssig ist. Genau genommen sagt man mit (3.1) aus

$$f = \{(x, t_x) \mid x \in X\} \wedge f : X \to Y,$$

man definiert die Zuordnung und *behauptet* gleichzeitig, dass es sich dabei um eine Funktion von X nach Y handelt[1].

[1] ... was natürlich auch eine falsche Behauptung sein kann, man könnte ja „definieren“ $f\colon \mathbb{R} \to \mathbb{N}, x \mapsto |x|$.

Funktionsdefinition		Bedeutung
$\begin{array}{c\|c} x & y \\ \hline x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{array}$	$\rightsquigarrow$	$f = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$
$f\colon X \to Y, \quad x \mapsto t_x$	$\rightsquigarrow$	$f = \{(x, t_x) \mid x \in X\}$

Abbildung 3.1: Bedeutung verschiedener Funktionsdefintionen

BEISPIEL 3.4

Seien $X = \{a, b, c, d\}$, $Y = \{1, 2, 3\}$ und seien $f_1 = \{(a, 1), (b, 3), (c, 2), (d, 3)\}$ und $f_2 = \{(a, 1), (b, 2), (b, 3), (c, 1), (d, 3)\}$ und $f_3 = \{(a, 1), (b, 2), (c, 3)\}$ gegeben. Welche dieser Teilmengen von $X \times Y$ sind Funktionen?

BEISPIEL 3.5

Welche der folgenden Teilmengen von $X \times Y$ definiert eine Funktion von X nach Y? Skizzieren Sie die angegebenen Punktmengen.

(a) $f = \{(x, 5) \mid x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$

(b) $f = \{(5, x) \mid x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$

(c) $f = \{(x, y) \mid x^2 + y^2 = 1\} \subseteq [-1, 1] \times \mathbb{R}$

(d) $f = \{(x, x^2) \mid x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$

DEFINITION 3.6

Sei $f\colon X \to Y$. Die Menge

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

heißt das *Bild* (oder der *Bildbereich* oder der *Wertebereich*) von X unter f. Ist $y \in f(X)$, dann wird ein $x \in X$ mit $f(x) = y$ als *ein Urbild* von y bezeichnet. Sei $Z \subseteq Y$. Die Menge

$$f^{-1}(Z) := \{x \in X \mid \exists\, y \in Z\colon y = f(x)\} \subseteq X$$

heißt *Urbildbereich* von Z unter f.

BEISPIEL 3.7

Wir betrachten wieder die Funktion aus Beispiel 3.1. Was ist $f(X)$? Was ist $f^{-1}(\{4,9\})$? Was ist $f^{-1}(\{1\})$? Was ist $f^{-1}(\{5\})$?

BEISPIEL 3.8

Sei $X = [-5,5]$ und $Y = \mathbb{R}$ und $f\colon X \to Y$ die Abbildung $x \mapsto |x|$.

(a) Was ist $f(X)$?

(b) Sei $Z = [1,3]$. Was ist $f^{-1}(Z)$?

BEISPIEL 3.9

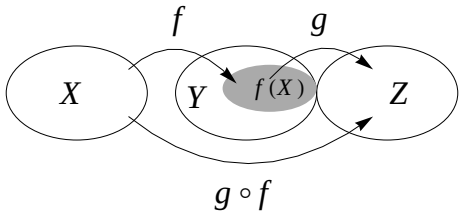
Gegeben sind die Mengen $X_1 = [-3,3]$, $X_2 = \mathbb{R}_0^+$ $X_3 = [-1,1]$ und $Y_1 = [-20,16]$, $Y_2 = \mathbb{R}$, $Y_3 = [-4,0]$. Wir betrachten die Abbildungen $f_k\colon X_k \to Y_k, x \mapsto (x+1)^2(x-2)$ für $k = 1,2,3$. (a) Wie sieht der Funktionsgraph aus? (b) Was ist jeweils der Bildbereich von X_k ($k = 1,2,3$) unter f_k?

DEFINITION 3.10: KOMPOSITION

Seien X, Y, Z Mengen und seien $f\colon X \to Y$ und $g\colon f(X) \to Z$. Dann nennt man $g \circ f$ die *Komposition* (*Hintereinanderausführung*) von f und g; sie ist definiert als

$$g \circ f\colon X \to Z, \quad x \mapsto g(f(x)),$$

d.h. also $(g \circ f)(x) := g(f(x))$. Für $g \circ f$ spricht man „g nach f“.



SATZ 3.11

Die Komposition zweier Funktionen ist wieder eine Funktion.

BEISPIEL 3.12

Seien $X = [0, \pi], Y = [0, 2\pi], Z = [-1, 1]$ und seien f und g gegeben durch

$$f\colon X \to Y, \quad x \mapsto 2x \qquad\qquad g\colon Y \to Z, \quad x \mapsto \sin(x).$$

Die Hintereinanderausführung $g \circ f$ ist gegeben durch

$$g \circ f\colon [0, \pi] \to [-1, 1], \quad x \mapsto \sin(2x).$$

BEISPIEL 3.13

Die Funktionen f, g seien gegeben durch

$$f\colon \mathbb{R} \to \mathbb{R}, \quad x \mapsto x^3 \qquad\qquad g\colon \mathbb{R} \to \mathbb{R}, \quad x \mapsto x + 1.$$

Bestimmen Sie $g \circ f$ und $f \circ g$.

DEFINITION 3.14: IDENTISCHE FUNKTION

Sei X eine Menge. Die Funktion von X nach X, die jedes Element $x \in X$ auf sich selbst abbildet, heißt die *identische Abbildung* auf X und wird mit id_X bezeichnet.

BEMERKUNG 3.15

Sei $f\colon X \to Y$ eine Funktion. Dann gilt für alle $x \in X$:

$$(id_Y \circ f)(x) = id_Y(f(x)) = f(x) \qquad (f \circ id_X)(x) = f(id_X(x)) = f(x).$$

EIGENSCHAFTEN VON FUNKTIONEN

DEFINITION 3.16

Sei $f\colon X \to Y$. f heißt *invertierbar* genau dann, wenn es ein $g\colon Y \to X$ gibt, sodass

$$g \circ f = id_X \qquad\qquad f \circ g = id_Y.$$

In diesem Fall schreibt man für g auch f^{-1} und nennt f^{-1} die *inverse Funktion* (oder *Umkehrfunktion*) zu f.

Die Umkehrfunktion kann auch mittels Produktmengen definiert werden: Sei $f \subseteq X \times Y$. Wenn

$$\{(y, x) \mid y \in Y, x \in X \text{ und } (x, y) \in f\} \subseteq Y \times X \tag{3.2}$$

eine Funktion ist, dann ist f invertierbar und f^{-1} ist durch (3.2) bestimmt.

Wie berechnet man die Umkehrfunktion? Bringt man die Funktionsgleichung $f(x) = y$ in die Form $x = g(y)$, so gilt für dieses g:

$$\begin{aligned} \text{für alle } x \in X\text{:} \quad & x = g(y) = g(f(x)) = (g \circ f)(x) \quad \text{d.h. } g \circ f = id_X \\ \text{für alle } y \in Y\text{:} \quad & y = f(x) = f(g(y)) = (f \circ g)(y) \quad \text{d.h. } f \circ g = id_Y \end{aligned}$$

Ist $g : Y \to X$, so ist $g = f^{-1}$.

BEISPIEL 3.17

Sei $f\colon [0,4] \to [-1,1], x \mapsto \frac{x}{2} - 1$. Um die Umkehrfunktion $f^{-1}\colon [-1,1] \to [0,4]$ zu bestimmen, formen wir $y = f(x)$ um auf $x = f^{-1}(y)$:

$$y = \frac{x}{2} - 1 \quad \rightsquigarrow \quad y + 1 = \frac{x}{2} \quad \rightsquigarrow \quad x = 2y + 2.$$

Man sieht leicht, dass $\{(y, 2y+2) \mid y \in [-1,1]\} \subseteq [-1,1] \times [0,4]$ eine Funktion von $[-1,1]$ nach $[0,4]$ ist. Damit ist die Umkehrfunktion f^{-1} gegeben durch

$$f^{-1}\colon [-1,1] \to [0,4], \quad y \mapsto 2y + 2.$$

Zur Probe berechnen wir $f \circ f^{-1}$ und $f^{-1} \circ f$ (ob tatsächlich die Identität herauskommt):

$$\begin{aligned} f \circ f^{-1}(x) &= f(f^{-1}(x)) = f(2x+2) = \frac{2x+2}{2} - 1 = x \\ f^{-1} \circ f(x) &= f^{-1}(f(x)) = f^{-1}\left(\tfrac{x}{2} - 1\right) = 2\left(\tfrac{x}{2} - 1\right) + 2 = x \end{aligned}$$

BEISPIEL 3.18

Welche der folgenden Funktionen ist invertierbar? Wie lautet die Umkehrfunktion der invertierbaren Funktionen?

- $f = \{(x, 5) \mid x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$
- $g\colon]1, \infty[\to]1, \infty[, x \mapsto \frac{x}{x-1}$
- $h\colon [-2,2] \to [0,4], x \mapsto 4 - x^2$

Wann ist eine Funktion nun invertierbar und wann nicht?

DEFINITION 3.19: INJEKTIV, SURJEKTIV, BIJEKTIV

Sei $f\colon X \to Y$.

- Man nennt f *injektiv* (von X nach Y) genau dann, wenn

$$\forall x_1, x_2 \in X\colon f(x_1) = f(x_2) \Rightarrow x_1 = x_2.$$

- Man nennt f *surjektiv* (von X nach Y) genau dann, wenn

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

- f heißt *bijektiv* (von X nach Y) genau dann, wenn f sowohl injektiv (von X nach Y) als auch surjektiv (von X nach Y) ist.

BEMERKUNG 3.20

- f ist injektiv (von X nach Y), wenn jedes Element im Bild von X *genau ein* Urbild in X besitzt, wenn also jedes $y \in Y$ höchstens ein Urbild in X hat.
- f ist surjektiv (von X nach Y), wenn jedes $y \in Y$ im Bild von X liegt, wenn also jedes $y \in Y$ mindestens ein Urbild in X hat, d.h. $Y \subseteq f(X)$. Da $f(X) \subseteq Y$ per Definition gilt, bedeutet das insgesamt $Y = f(X)$.

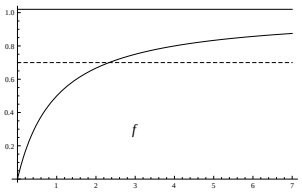
BEISPIEL 3.21

Für welche Mengen X_k, Y_k ist $f_k \subseteq X_k \times Y_k$ aus Beispiel 3.9 (a) injektiv (b) surjektiv (c) bijektiv?

Die Begriffe injektiv und surjektiv hängen mit der Lösbarkeit der Gleichung $f(x) = y$ bei gegebener Funktion $f \colon X \to Y$ und gegebenen $y \in Y$ zusammen:

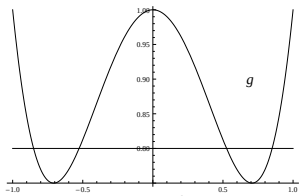
- Wenn f *injektiv* (von X nach Y) ist, dann besitzt die Gleichung $f(x) = y$ *höchstens* eine Lösung $x \in X$.
- Wenn f *surjektiv* (von X nach Y) ist, dann besitzt die Gleichung $f(x) = y$ *mindestens* eine Lösung $x \in X$.
- Wenn f *bijektiv* (von X nach Y) ist, dann besitzt die Gleichung $f(x) = y$ *genau* eine Lösung $x \in X$.

Die Lösung (falls existent) kann gefunden werden, indem der Funktionsgraph mit der horizontalen Gerade durch y geschnitten wird.



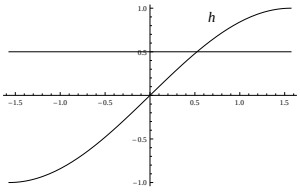
$f \colon \mathbb{R}_0^+ \to \mathbb{R}_0^+,$

$x \mapsto 1 - \dfrac{1}{x+1}$



$g \colon [-1, 1] \to [\frac{3}{4}, 1],$

$x \mapsto x^4 - x^2 + 1$



$h \colon [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$

$x \mapsto \sin(x)$

Für jede nicht-leere Menge X ist id_X ein sehr einfaches Beispiel für eine bijektive Funktion von X nach X.

Eine Funktion ist also invertierbar genau dann, wenn jedem $y \in Y$ ein eindeutig bestimmtes $x \in X$ zugeordnet werden kann, für das $f(x) = y$ gilt.

SATZ 3.22

Sei $f : X \to Y$. Dann gilt:

$$f \text{ ist invertierbar} \Leftrightarrow f \text{ ist bijektiv}.$$

DEFINITION 3.23

Sei $f\colon X \to Y$ und $A \subseteq X$.

- f heißt *streng monoton wachsend* auf A genau dann, wenn für alle $x_1, x_2 \in A$ gilt:
$$x_1 < x_2 \Rightarrow f(x_1) < f(x_2).$$
- f heißt *monoton wachsend* auf A genau dann, wenn für alle $x_1, x_2 \in A$ gilt:
$$x_1 < x_2 \Rightarrow f(x_1) \leq f(x_2).$$
- f heißt *streng monoton fallend* auf A genau dann, wenn für alle $x_1, x_2 \in A$ gilt:
$$x_1 < x_2 \Rightarrow f(x_1) > f(x_2).$$
- f heißt *monoton fallend* auf A genau dann, wenn für alle $x_1, x_2 \in A$ gilt:
$$x_1 < x_2 \Rightarrow f(x_1) \geq f(x_2).$$

Wird A nicht extra erwähnt, so meint man $A = X$.

BEISPIEL 3.24

- Die Betragsfunktion $f\colon [-5,5] \to [0,5], x \mapsto |x|$ ist im Intervall $[-5,0]$ streng monoton fallend und im Intervall $[0,5]$ streng monoton wachsend.
- Die Vorzeichenfunktion
$$\operatorname{sign}\colon [-5,5] \to [-1,1], x \mapsto \begin{cases} 1, & x \geq 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$
ist monoton wachsend.

SATZ 3.25

Die Hintereinanderausführung zweier Funktionen f und g ist monoton wachsend, wenn f und g beide monoton wachsend oder beide monoton fallend sind.

Die Hintereinanderausführung zweier Funktionen f und g ist monoton fallend, wenn f monoton wachsend und g monoton fallend ist, oder, wenn f monoton fallend und g monoton wachsend ist.

SPEZIELLE TYPEN VON FUNKTIONEN

Für $d \in \mathbb{N}_0$ heißt ein Ausdruck der Form $a_0 + a_1 x + \cdots + a_d x^d$ mit $a_0, \ldots, a_d \in \mathbb{R}$ und $a_d \neq 0$ ein *reelles Polynom (von Grad d)*.

DEFINITION 3.26: POLYNOMFUNKTION, RATIONALE FUNKTION

Man nennt p eine *(reelle) Polynomfunktion* genau dann, wenn es $X, Y \subseteq \mathbb{R}$ gibt mit

- $p : X \to Y$ und
- es ein $d \in \mathbb{N}_0$ und reelle Zahlen $a_0, \ldots, a_d$ mit $a_d \neq 0$ gibt mit

$$p(x) = a_0 + a_1 x + \cdots + a_d x^d \quad \text{für alle } x \in X.$$

Wie bei Polynomen nennt man auch bei Polynomfunktionen d den *Grad* von p und schreibt dafür $\deg(p)$.

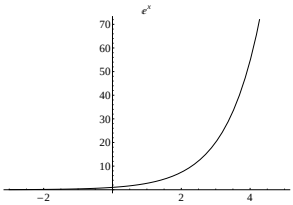
Man nennt r eine *rationale Funktion* genau dann, wenn es $X, Y \subseteq \mathbb{R}$ gibt mit

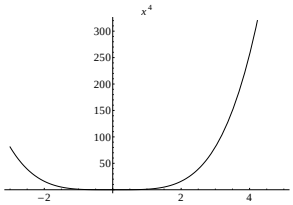
- $r : X \to Y$ und
- es Polynomfunktionen p, q gibt mit $r(x) = p(x)/q(x)$ für alle $x \in X$.

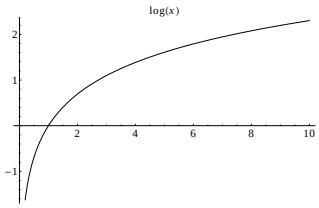
BEMERKUNG 3.27

Die Hintereinanderausführung zweier Polynomfunktionen ist wieder eine Polynomfunktion. Wenn p_1 eine Polynomfunktion vom Grad d_1 ist und p_2 eine Polynomfunktion vom Grad d_2, dann ist $p_1 \circ p_2$ eine Polynomfunktion vom Grad $d_1 d_2$.

Ebenso ist die Hintereinanderausführung zweier rationaler Funktionen wieder eine rationale Funktion.







Sei $a \in \mathbb{R}^+ \backslash \{1\}$. Jede Funktion der Form $f\colon X \to Y, x \mapsto a^x$ (z.B. von $\mathbb{R}$ nach $\mathbb{R}^+$) heißt *eine Exponentialfunktion* zur Basis a. Ein Spezialfall ist die Funktion $\exp\colon \mathbb{R} \to \mathbb{R}^+, x \mapsto e^x$, wobei $e \approx 2.7182818$ die *Eulersche Zahl* ist[2]. Die Exponentialfunktion ist bijektiv von $\mathbb{R}$ nach $\mathbb{R}^+$. Die daher existierende inverse Funktion zur Exponentialfunktion $\exp$ wird mit $\ln$ bezeichnet und *natürliche Logarithmusfunktion* genannt. Klarerweise gilt

$$\exp \circ \ln = id_{\mathbb{R}^+} \qquad\qquad \ln \circ \exp = id_{\mathbb{R}}.$$

Mit anderen Worten:

$$\forall x \in \mathbb{R}^+ : e^{\ln(x)} = x \qquad\qquad \forall x \in \mathbb{R} : \ln(e^x) = x \tag{3.3}$$

SATZ 3.28

Für die Exponentialfunktion gelten die gewohnten Rechenregeln für Potenzen, d.h.

$$\exp(x+y) = e^{x+y} = e^x \cdot e^y = \exp(x) \cdot \exp(y) \quad \exp(x \cdot y) = e^{x \cdot y} = (e^x)^y = \exp(x)^y.$$

Wegen (3.3) gilt natürlich auch $a = e^{\ln(a)}$, daher ist

$$a^x = e^{x \cdot \ln(a)} = \exp(x \cdot \ln(a)),$$

jede Exponentialfunktion ist also durch $\exp$ darstellbar[3]. Für die inverse Funktion einer Exponentialfunktion zur Basis a – wir sprechen von einer *Logarithmusfunktion zur Basis* a – bedeutet das:

$$y = a^x = \exp(x \cdot \ln(a)) \quad \rightsquigarrow \quad x = \frac{\ln(y)}{\ln(a)} =: \log_a(y),$$

es ist also auch jede Logarithmusfunktion durch $\ln$ darstellbar. Damit unterscheiden sich die Logarithmen bezüglich verschiedener Basen nur um ein (konstantes) Vielfaches, siehe Abbildung 3.2.

[2] Ist von *der Exponentialfunktion* die Rede, so meint man immer $\exp$.

[3] Wenn Sie sich jetzt fragen, wie man $\exp(x)$ für $x \in \mathbb{R}$ ausrechnen kann bzw. wie $\exp(x)$ eigentlich definiert ist: $\exp(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \ldots$. Das ist eine unendliche Summe, eine sogenannte *Reihe*, die als Grenzwert der entsprechenden endlichen Teilsummen festgelegt ist. Dass so ein Grenzwert für jedes $x \in \mathbb{R}$ tatsächlich existiert, ist keineswegs selbstverständlich. Die Eulersche Zahl e ist übrigens genau definiert als dieser Grenzwert im Fall $x = 1$. Alles weitere würde hier den Rahmen sprengen.

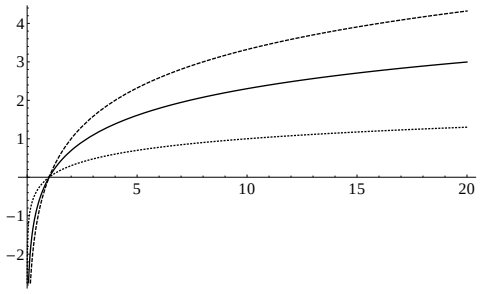


Abbildung 3.2: $\ln(x)$ (—), $\log_{10}(x)$ ($\cdots$) und $\log_2(x)$ (- - -).

SATZ 3.29

Für Logarithmen gelten die folgenden Rechenregeln. Für $x, y > 0$ bzw. $d \in \mathbb{R}$

$$\log_a(xy) = \log_a(x) + \log_a(y) \qquad \log_a(x^d) = d \log_a(x).$$

Beweis. Wir betrachten die erste dieser Identitäten: Sei $u = \log_a(x)$ und $v = \log_a(y)$, d.h. nach Definition des Logarithmus gilt $x = a^u$ und $y = a^v$. Damit erhalten wir:

$$xy = a^u a^v = a^{u+v}, \quad \text{und somit} \quad \log_a(xy) = \log_a\left(a^{u+v}\right) = u + v = \log_a(x) + \log_a(y). \quad \square$$

DER BEGRIFF DER ANZAHL VON ELEMENTEN IN EINER MENGE

Im Fall endlicher Mengen hat man eine recht anschauliche Vorstellung des Begriffs der „Anzahl von Elementen in einer Menge“. Aber selbst hier muss man vorsichtig sein, man kann nicht allgemein sagen, dass die Menge $\{a_1, a_2, \ldots, a_n\}$ aus n Elementen besteht. Warum? Nehmen Sie $\{1, 2, 3, 2, 1\}$. Diese Menge besteht nicht aus 5 Elementen, sondern nur aus 3. Wir brauchen also eine ordendliche Definition, am besten eine, die sich auch auf den Fall unendlicher Mengen wie $\mathbb{N}$, $\mathbb{Q}$ oder $\mathbb{R}$ übertragen lässt. Spätestens hier nämlich lässt uns unsere Intuition im Stich, hier müssen wir uns strikt auf Definitionen stützen.

DEFINITION 3.30: GLEICHMÄCHTIG

Zwei Mengen X, Y heißen *gleichmächtig* (besitzen die gleiche Anzahl von Elementen, haben die gleiche Kardinalität) genau dann, wenn eine bijektive Funktion $f\colon X \to Y$ existiert.

BEISPIEL 3.31

Es gibt gleich viele gerade wie ungerade natürliche Zahlen. Wir bezeichnen die geraden natürlichen Zahlen mit $\mathbb{N}_g$ und die ungeraden natürlichen Zahlen mit $\mathbb{N}_u$. Eine bijektive Funktion von $\mathbb{N}_g$ nach $\mathbb{N}_u$ ist beispielsweise durch $f\colon \mathbb{N}_g \to \mathbb{N}_u, x \mapsto x + 1$ gegeben.

BEISPIEL 3.32

Die Abbildung

$$f\colon \mathbb{N} \to \mathbb{N}_g, \quad x \mapsto 2x$$

ist bijektiv. Folglich haben $\mathbb{N}$ und $\mathbb{N}_g$ die gleiche Anzahl von Elementen[a].

[a] Intuitiv würde man doch meinen, dass $\mathbb{N}$ doppelt so viele Elemente hat wie $\mathbb{N}_g$, und „doppelt so viele" ist doch „mehr". Aber im Unendlichen reicht das nicht, durch Verdoppeln bleiben es gleich viele.

BEISPIEL 3.33

Die Abbildung

$$f\colon [0,4] \to [-1,1], \quad x \mapsto \frac{x}{2} - 1$$

ist bijektiv. Folglich haben $[0,4]$ und $[-1,1]$ die gleiche Anzahl von Elementen.

Im folgenden verwenden wir für $m, n \in \mathbb{N}_0$ die Abkürzung

$$\mathbb{N}_{m,n} := \{i \in \mathbb{N} \mid m \leq i \leq n\}.$$

DEFINITION 3.34: ENDLICHE MENGEN, UNENDLICHE MENGEN, KARDINALITÄT

Man nennt eine Menge A *endlich* genau dann, wenn ein $n \in \mathbb{N}_0$ und eine bijektive Funktion $f : \mathbb{N}_{1,n} \to A$ existieren. Andernfalls nennt man A eine *unendliche* Menge. Die *Kardinalität* (*Mächtigkeit*) einer endlichen Menge A ist definiert als

$$|A| := \text{dasjenige } n \in \mathbb{N}_0, \text{ für das eine bijektive Funktion } f : \mathbb{N}_{1,n} \to A \text{ existiert.}$$

Für unendliche Mengen A lassen wir die Kardinalität $|A|$ *undefiniert*. Es hat sich aber für unendliche Mengen A die Schreibweise $|A| = \infty$ eingebürgert, nur um den Umstand auszudrücken, dass A eine unendliche Menge ist[4].

BEISPIEL 3.35

$$|\{1,3,7,12\}| = 4 \qquad |\{2,2\}| = 1 \qquad |\mathbb{N}| = \infty \qquad |[0,1]| = \infty \qquad |\mathbb{R}| = \infty$$

[4] „$|A| = \infty$" ist einfach eine *symbolische Schreibweise*, weder ist „∞" eine Zahl, noch ist darin „=" das übliche mathematische Gleichheitszeichen. Wäre das nämlich so, so kämen wir sofort in Teufels Küche: es sind ja sowohl $\mathbb{N}$ als auch $\mathbb{R}$ unendliche Mengen, also $|\mathbb{N}| = \infty$ und $|\mathbb{R}| = \infty$, was mit den Regeln für Gleichheit sofort auf $|\mathbb{N}| = |\mathbb{R}|$ führt. Das ist aber *falsch*, wie man überall nachlesen kann! Am besten kommen Sie durchs Leben, wenn Sie einsehen, dass ∞ keine Zahl ist (also insbesondere $\infty \notin \mathbb{N}$ und $\infty \notin \mathbb{R}$), und Sie bei jedem Auftauchen von ∞ (z.B. im Zusammenhang mit Grenzwerten und $\lim$-Schreibweisen) immer davon ausgehen, dass es sich um rein symbolische Schreibweisen handelt.

BEMERKUNG 3.36

Intuitiv soll die Anzahl der Elemente in der leeren Menge natürlich 0 sein, also $|\emptyset| = 0$. Es ist ein gute Übung in abstraktem Denken, sich davon zu überzeugen, dass mit obigen Definitionen *tatsächlich* $|\emptyset| = 0$ gilt.

Für Mengen A und B steht $|A| = |B|$ für die Tatsache, dass A und B gleichmächtig sind, siehe Definition 3.30. Im Falle endlicher Mengen ist $|.|$ genau die Kardinalität aus Definition 3.34, im Falle unendlicher Mengen ist es wieder eine rein symbolische Schreibweise. Die Aussagen aus den Beispielen 3.31 bis 3.33 lassen sich also wie folgt zusammenfassen:

$$|\mathbb{N}| = |\mathbb{N}_g| = |\mathbb{N}_u| \qquad\qquad |[0,4]| = |[-1,1]|.$$

Ohne Beweis wollen wir Folgendes festhalten:

BEMERKUNG 3.37

$$|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}| \qquad\qquad |\mathbb{N}| \neq |\mathbb{R}| \qquad\qquad |[0,1]| = |\mathbb{R}|.$$

SATZ 3.38

Es seien A, B, C endliche Mengen mit $A \cap B = \emptyset$ und $C \subseteq A$. Dann gilt:

$$|A \cup B| = |A| + |B| \qquad\qquad |A \backslash C| = |A| - |C|.$$

Beweis. Direkt basierend auf Definition 3.34. Man muss die benötigten bijektiven Funktionen konstruieren bzw. kann im zweiten Teil eventuell Teil 1 schon verwenden. Gute Übung! □

SATZ 3.39

Für beliebige endliche Mengen A, B gilt:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Beweis. Sei $C := A \cap B$. Man überzeugt sich rasch davon, dass

$$A \cup B = (A \backslash C) \cup B \quad \text{und} \quad (A \backslash C) \cap B = \emptyset,$$

somit wegen Satz 3.38

$$|A \cup B| = |A \backslash C| + |B|.$$

Da $C \subseteq A$, gilt wieder wegen Satz 3.38 $|A \backslash C| = |A| - |C|$ und daher insgesamt

$$|A \cup B| = |A \backslash C| + |B| = |A| - |C| + |B| = |A| + |B| - |A \cap B|.$$ □

Die Anzahl der Elemente in einer Produktmenge ist gleich dem Produkt der Kardinalitäten der jeweiligen (endlichen) Mengen.

SATZ 3.40: KARDINALITÄT VON PRODUKTMENGEN

Für endliche Mengen A, B gilt: $|A \times B| = |A| \cdot |B|$.

SATZ 3.41: KARDINALITÄT VON POTENZMENGEN

Sei A eine endliche Menge. Dann gilt $|P(A)| = 2^{|A|}$.

Beweis. Wir verwenden Induktion über endliche Mengen[5]. Induktionsanfang: Sei $A = \emptyset$. Dann ist die Potenzmenge $P(A) = \{\emptyset\}$ und es gilt $|P(A)| = 1 = 2^0$.

Induktionsannahme: Sei $n \in \mathbb{N}_0$, wir nehmen an, für alle X mit $|X| = n$ gilt $|P(X)| = 2^n$.

Induktionsschritt: Wir betrachten jetzt eine beliebige $(n+1)$-elementige Menge A, zu zeigen ist $P(A) = 2^{n+1}$.

$$A = \{a_1, a_2, \ldots, a_n, a_{n+1}\} = \{a_1, a_2, \ldots, a_{n-1}, a_n\} \cup \{a_{n+1}\}.$$

Sei $A_1 = \{a_1, a_2, \ldots, a_{n-1}, a_n\}$ (d.h., $A = A_1 \cup \{a_{n+1}\}$). Wegen $|A_1| = n$ gilt laut Induktionsvoraussetzung $|P(A_1)| = 2^n$.

Wir konstruieren die Potenzmenge von A jetzt folgendermaßen: Für jedes $B \in P(A_1)$ gilt $B \in P(A)$. Weiters gilt für jedes $B \in P(A_1)$, dass $C = B \cup \{a_{n+1}\} \in P(A)$. Die Potenzmenge von A enthält keine weiteren Elemente mehr, d.h.

$$P(A) = P(A_1) \cup \{B \cup \{a_{n+1}\} \mid B \in P(A_1)\}.$$

Aufgrund von Satz 3.38 ist daher

$$|P(A)| = |P(A_1)| + |\{B \cup \{a_{n+1}\} \mid B \in P(A_1)\}| = |P(A_1)| + |P(A_1)| = 2 \cdot 2^n = 2^{n+1}. \quad \square$$

[5]Um eine Aussage für alle endlichen Mengen zu beweisen, geht man wie folgt vor: 1) Induktionsanfang: man beweist die Aussage für die leere Menge. 2) Induktionsannahme: man nimmt an, dass die Aussage *für alle Mengen* mit einer fixen Anzahl von n Elementen gilt. 3) Induktionsschritt: man beweist die Aussage *für alle Mengen* mit $n + 1$ Elementen.

KAPITEL 4

RELATIONEN

DER RELATIONSBEGRIFF IN DER MATHEMATIK

Relationen sollen dazu dienen, *Beziehungen zwischen Objekten* zu modellieren. Dazu betrachtet man üblicherweise zwei Mengen, zwischen deren Elementen diese Beziehungen bestehen. Hat man nur eine Menge, so kann man Beziehungen zwischen den Elementen untereinander untersuchen.

BEISPIEL 4.1

Es sei A eine Menge von Personen und G eine Menge von Gütern. Dann kann man die Beziehung „Peter besitzt ein Smartphone“ durch das Paar $(\text{Peter}, \text{Smartphone})$. Die gesamte „Besitzrelation“ B kann durch Zusammenfassen all dieser Paare zu einer Menge dargestellt werden, die „Besitzrelation“ besteht also aus allen Paaren, die miteinander in dieser Beziehung stehen, z.B.

$$B := \{(\text{Peter}, \text{Smartphone}), (\text{Peter}, \text{Rad}), (\text{Sue}, \text{Rad})\} \subseteq A \times G.$$

Die Beziehung „x besitzt y“ besteht dann genau dann, wenn $(x, y) \in B$.

Man sieht nun natürlich sofort, dass auf diese Weise jede Menge $R \subseteq X \times Y$ eine Beziehung zwischen den Elementen von X und Y darstellt.

DEFINITION 4.2

Seien X, Y Mengen. Man nennt R eine *(binäre) Relation zwischen* X und Y genau dann, wenn $R \subseteq X \times Y$. Ist $X = Y$, so spricht man von einer *Relation auf* X. Für $x \in X$ und $y \in Y$ stehen die Schreibweisen

$$(x, y) \in R \quad \text{oder} \quad xRy$$

für „x steht in Relation R zu y“.

Zur Angabe einer Relation können wir auf alle Möglichkeiten zurückgreifen, wie Teilmengen von $X \times Y$ definiert werden können. Endliche Relationen können durch Aufzählung aller Paare angegeben werden, dies ist vergleichbar mit der Definition einer Funktion durch eine

Wertetabelle. Auch hier muss nicht unbedingt eine Gesetzmäßigkeit hinter der Paarbildung stehen. Unendliche Relationen in der allgemeinsten Form sind Mengen

$$R = \{(x, y) \in X \times Y \mid A_{x,y}\},$$

wobei $A_{x,y}$ irgendeine Aussage mit freien Variablen x, y ist. Da für solche Mengen $(x, y) \in R$ gleichbedeutend ist mit $A_{x,y}$, legt man derartige Relationen üblicherweise fest, indem man nur $A_{x,y}$ definiert.

BEISPIEL 4.3

Seien $X = \{0, 1, 2\}$, $Y = \{a, b, c\}$ dann definieren

$$R_1 := \{(0, a), (0, b), (2, c)\} \subseteq X \times Y \qquad R_2 := \{(0, c), (1, b), (2, a)\} \subseteq X \times Y$$

Relationen zwischen X und Y.

BEISPIEL 4.4

Sei $X = \{1, 2, 3, 4, 6, 12\}$ dann wird durch die Bedingung „x teilt y“ eine Relation auf X definiert. In Zeichen:

$$xR_3y :\Leftrightarrow x \mid y,$$

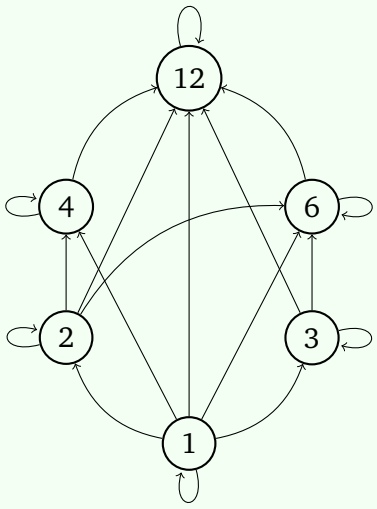
womit man nichts anderes sagt als

$$R_3 := \{(x, y) \in X^2 \mid x \text{ teilt } y\}.$$

Wir können R_3 auch explizit anschreiben als

$$\begin{aligned} R_3 = &\{(1,1), (1,2), (1,3), (1,4), (1,6), (1,12), (2,2), (2,4), (2,6), (2,12), (3,3), (3,6), (3,12), \\ &(4,4), (4,12), (6,6), (6,12), (12,12)\} \subseteq X \times X \end{aligned}$$

oder graphisch darstellen, wobei Pfeile $x \to y$ für xR_3y stehen:



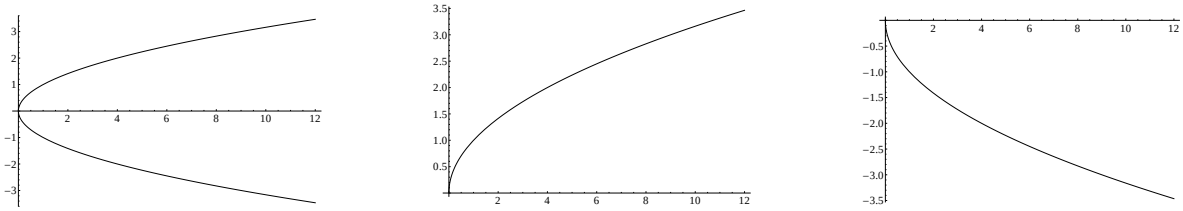


Abbildung 4.1: Relation W (links), Funktionen $\sqrt{\ }$ (mitte) und W_- (rechts).

BEISPIEL 4.5

Sei X eine Menge von Geraden im $\mathbb{R}^2$ und zu $g, h \in X$ definieren wir gRh genau dann, wenn g und h parallel sind.

BEISPIEL 4.6

Sei $A = \{1, 2\}$ und $X = P(A)$ (die Potenzmenge von A). Wir definieren die Relation $R \subseteq X \times X$ durch aRb genau dann, wenn $a \subseteq b$.

Jede Funktion ist also auch eine Relation, aber natürlich nicht umgekehrt.

BEISPIEL 4.7

Wir betrachten die Relation

$$W := \{(x, y) \in \mathbb{R}_0^+ \times \mathbb{R} \mid x = y^2\},$$

siehe Abbildung 4.1 (links). Jedes $x > 0$ steht mit mehreren $y \in \mathbb{R}$ in Relation, z.B. $(4, 2) \in W$ aber auch $(4, -2) \in W$. W ist also *keine* Funktion von R_0^+ nach $\mathbb{R}$. Wenn der Zielbereich (sinnvoll) eingeschränkt wird, dann kann man tatsächlich von einer Funktion sprechen. Die übliche „Wurzelfunktion“ ist dabei jene, in der der Wertebereich auf $\mathbb{R}_0^+$ eingeschränkt wird, d.h.

$$\sqrt{\ } := \{(x, y) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+ \mid x = y^2\}$$

ist eine Funktion. Für $x \in \mathbb{R}_0^+$ bezeichnet $\sqrt{x}$ also *dasjenige* $y \in \mathbb{R}_0^+$ mit $x = y^2$. Mit dieser Schreibweise können wir aus W auch eine andere Funktion ableiten, nämlich

$$W_- \colon \mathbb{R}_0^+ \to \mathbb{R}_0^-, \quad x \mapsto -\sqrt{x},$$

siehe Abbildung 4.1.

DEFINITION 4.8: VORBEREICH, NACHBEREICH

Sei R eine Relation auf den Mengen X, Y. Dann ist der *Vorbereich* von R definiert als

$$\mathrm{dom}(R) := \{x \in X \mid \exists\, y \in Y : xRy\}.$$

Der *Nachbereich* von R ist definiert als

$$\mathrm{cod}(R) := \{y \in Y \mid \exists\, x \in X : xRy\}.$$

DEFINITION 4.9: KOMPLEMENT, INVERSE

Das *Komplement* von R ist $\overline{R} := (X \times Y) \backslash R$.

Die *Inverse* von R ist jene Relation $R^{-1} \subseteq Y \times X$, die durch

$(y, x) \in R^{-1} :\Leftrightarrow (x, y) \in R$ mit anderen Worten: $R^{-1} := \{(y, x) \in Y \times X \mid (x, y) \in R\}$

definiert ist.

DEFINITION 4.10: KOMPOSITION

Es seien X, Y, Z Mengen und $R \subseteq X \times Y$ und $S \subseteq Y \times Z$ Relationen. Die *relationale Komposition* (bzw. das *Relationenprodukt*) $S \circ R \subseteq X \times Z$ von R und S ist definiert als

$$S \circ R := \{(x, z) \mid \exists\, y \in Y : xRy \wedge ySz\}.$$

Definition 4.10 besagt, dass $(x, z) \in S \circ R$ genau dann, wenn es ein $y \in Y$ gibt mit $(x, y) \in R \wedge (y, z) \in S$. Für Relationen, die in einem Pfeildiagramm veranschaulicht werden können, bedeutet das anschaulich, dass es einen „Weg“ (eine Verbindung) von x nach z gibt, in dem man den ersten Schritt in R (nach y) und den zweiten Schritt in S (von y weiter zu z) geht. Man beachte, dass im Fall von Funktionen R und S (jede Funktion ist eine Relation!) das Relationenprodukt $S \circ R$ genau mit der Komposition von Funktionen (Definition 3.10) übereinstimmt.

BEISPIEL 4.11

Seien $X = \{0, 1, 2\}, Y = \{a, b, c\}$ und $Z = \{x, y, z\}$ und die beiden Relationen $R_1 \subseteq X \times Y$ und $R_2 \subseteq Y \times Z$ gegeben durch

$$R_1 = \{(0, a), (0, b), (1, c), (2, a)\} \qquad R_2 = \{(a, z), (b, y), (c, z)\}.$$

Dann gilt

$$R_2 \circ R_1 = \{(0, y), (0, z), (1, z), (2, z)\} \subseteq X \times Z.$$

DEFINITION 4.12: POTENZEN EINER RELATION

Sei X eine Menge und $R \subseteq X \times X$ eine Relation. Mit $id_X \subseteq X \times X$ bezeichnen wir die *Identitätsrelation* definiert durch

$$id_X := \{(x, x) \mid x \in X\}.$$

Die *Potenzen* von R werden dann rekursiv definiert durch

$$R^0 := id_X \qquad\qquad R^n := R \circ R^{n-1}, \quad \text{für } n \geq 1.$$

Eigenschaften von Relationen

Definition 4.13: eindeutig, total

$R \subseteq X \times Y$ heißt *rechtseindeutig* genau dann, wenn gilt

$$\forall\, x \in X\, \forall\, y, z \in Y : (xRy \wedge xRz) \Rightarrow y = z,$$

$R \subseteq X \times Y$ heißt *rechtstotal* genau dann, wenn $\mathrm{cod}(R) = Y$. $R \subseteq X \times Y$ heißt *linkseindeutig* genau dann, wenn R^{-1} rechtseindeutig ist, bzw. *linkstotal*, wenn R^{-1} rechtstotal ist.

Noch einmal zurück zu Funktionen: Seien X, Y Mengen. Eine Relation $f \subseteq X \times Y$ heißt *partielle Funktion*, wenn f rechtseindeutig ist. f heißt *Funktion* (oder auch *totale Funktion*, wenn f außerdem linkstotal ist, was unserer Definition einer Funktion aus dem letzten Kapitel entspricht.

Eine (totale) Funktion heißt *injektiv*, wenn jeder Wert im Bildbereich höchstens einmal erreicht wird. In der Sprache der Relationen bedeutet das, dass $f \subseteq X \times Y$ *injektiv* ist genau dann, wenn f *linkseindeutig* ist.

Eine (totale) Funktion heißt *surjektiv*, wenn jeder Wert im Bildbereich mindestens einmal erreicht wird. In der Sprache der Relationen bedeutet das, dass $f \subseteq X \times Y$ *surjektiv* ist genau dann, wenn f *rechtstotal* ist.

Definition 4.14

Eine Relation R auf einer Menge X heißt

- *reflexiv* (auf X) genau dann, wenn $\forall\, x \in X : xRx$
- *symmetrisch* (auf X) genau dann, wenn $\forall\, x, y \in X : xRy \Rightarrow yRx$
- *antisymmetrisch* (auf X) genau dann, wenn $\forall\, x, y \in X : xRy \wedge yRx \Rightarrow x = y$
- *transitiv* (auf X) genau dann, wenn $\forall\, x, y, z \in X : xRy \wedge yRz \Rightarrow xRz$

Beispiel 4.15

Wir betrachten wieder die Teilbarkeitsrelation R_3 aus Beispiel 4.4. Wir untersuchen diese Relation auf die Eigenschaften aus Definition 4.14:

- Reflexivität: Jede Zahl teilt sich selbst, daher gilt für jedes $x \in X = \{1, 2, 3, 4, 6, 12\}$, dass xR_3x. Somit ist R_3 reflexiv.
- Symmetrie: Die Relation ist nicht symmetrisch, was durch *ein* Gegenbeispiel gezeigt werden kann (Symmetrie müsste für alle Elemente der Relation gelten): es gilt $2 \mid 4$ aber $4 \nmid 2$.
- Antisymmetrie: Seien $x, y \in X$ mit $x \mid y$ und $y \mid x$. Aus $x \mid y$ folgt, dass es ein $p \in \mathbb{N}_0$ gibt sodass $y = px$. Aus $y \mid x$ folgt, dass es ein $q \in \mathbb{N}_0$ gibt mit $x = qy$. Zusammengefasst heißt das, dass $y = pq \cdot y$ für natürliche Zahlen p, q. Das kann nur gelten, wenn $p = q = 1$ ist, d.h. $x = y$. Die Relation ist also antisymmetrisch.
- Transitivität: Seien $x, y, z \in X$ mit $x \mid y$ und $y \mid z$. Analog zum letzten Punkt bedeutet das, dass $p, q \in \mathbb{N}_0$ existieren mit $y = px$ und $z = qy$. Folglich gilt $z = qp \cdot x$. Da $qp \in \mathbb{N}_0$ ist, folgt daraus, dass $x \mid z$. Die Relation ist also transitiv.

BEISPIEL 4.16

Die Relation aus Beispiel 4.5 ist reflexiv, symmetrisch und transitiv, aber nicht antisymmetrisch.

DEFINITION 4.17: HÜLLEN

Sei R eine beliebige Relation auf einer Menge X. Die *transitive Hülle* R^+ ist die kleinste Relation, die R einschließt und die Eigenschaft der Transitivität erfüllt. Die *reflexiv transitive Hülle* R^* ist die kleinste Relation, die R^+ einschließt und zusätzlich die Eigenschaften der Reflexivität erfüllt.

Es gelten die folgenden Beziehungen:

$$R^+ = \bigcup_{n \in \mathbb{N}} R^n, \quad \text{und} \quad R^* = \bigcup_{n \in \mathbb{N}_0} R^n.$$

BEISPIEL 4.18

Sei $X = \{0, 1, 2\}$ und $R = \{(0, 1), (0, 2), (1, 0), (2, 2)\} \subseteq X \times X$. Wir bestimmen die reflexiv transitive Hülle von R:

$$\begin{aligned}
R^0 = id_X &= \{(0, 0), (1, 1), (2, 2)\} \\
R^1 = R &= \{(0, 1), (0, 2), (1, 0), (2, 2)\} \\
R^2 = R \circ R &= \{(0, 0), (0, 2), (1, 1), (1, 2), (2, 2)\} \\
R^3 = R \circ R^2 &= \{(0, 1), (0, 2), (1, 0), (1, 2), (2, 2)\} \\
R^4 = R \circ R^3 &= \{(0, 0), (0, 2), (1, 1), (1, 2), (2, 2)\} = R^2 \\
R^5 = R \circ R^4 &= R \circ R^2 = R^3
\end{aligned}$$

Damit gilt $R^{2n} = R^2$ und $R^{2n+1} = R^3$ für alle $n \geq 2$. Die reflexiv transitive Hülle ergibt

sich also als Vereinigung von R^0, R, R^2 und R^3:

$$R^* = \{(0,0),(0,1),(0,2),(1,0),(1,1),(1,2),(2,2)\}.$$

ÄQUIVALENZRELATIONEN

DEFINITION 4.19: ÄQUIVALENZRELATION

Eine Relation R auf einer Menge X heißt *Äquivalenzrelation* auf X genau dann, wenn R reflexiv, symmetrisch und transitiv auf X ist. Für Äquivalenzrelationen wird häufig die Notation $\sim$ statt R verwendet.

Ein typisches Beispiel für eine Äquivalenzrelation ist die Kongruenz modulo einer natürlichen Zahl:

DEFINITION 4.20: KONGRUENZ MODULO

Seien $n \in \mathbb{N}$ und $a, b \in \mathbb{Z}$. Dann sind a und b *kongruent modulo* n genau dann, wenn $n \mid a - b$. Man schreibt $a \equiv_n b$ (oder $a \equiv b \pmod{n}$ oder $a = b \pmod{n}$ oder $a \sim_n b$).

Kongruenz modulo n ist eng verwandt mit der ganzzahligen Division durch n bzw. der Restbildung bei einer solchen Division. Dazu führen wir zuerst formal die Begriffe Quotient und Rest ein.

SATZ 4.21

Seien $x, y \in \mathbb{N}_0$ mit $y \neq 0$. Dann gibt es eindeutig bestimmte Zahlen $q, r \in \mathbb{N}_0$ mit

$$x = qy + r \quad \textit{und} \quad r < y.$$

DEFINITION 4.22

Für $x, y \in \mathbb{N}_0$ mit $y \neq 0$ nennt man die laut Satz 4.21 eindeutig existierenden $q, r \in \mathbb{N}_0$ den *Quotient*, bzw. den *Rest* bei Division von x durch y. Für den Rest bei Division von x durch y schreibt man oft $\mathrm{mod}(x, y)$ oder $x \bmod y$ (in Programmiersprachen oft $x\%y$)[a].

[a] Achtung, aufpassen mit „mod“: in $a \equiv b \pmod{n}$ ist „mod“ Teil der Notation, es ist nur eine Schreibweise „für das Auge“ im Kontext des *Prädikatensymbols* „kongruent“. In $\mathrm{mod}(x, y)$ und $x \bmod y$ hingegen handelt es sich um ein 2-stelliges *Funktionssymbol*.

Seien $a \equiv_n b$, d.h. $a = b + nq$ für ein $q \in \mathbb{N}_0$, und sei $r = \mathrm{mod}(b, n)$, also $b = ln + r$ für ein $l \in \mathbb{N}_0$. Dann gilt

$$a = b + nq = ln + r + nq = n(l + q) + r.$$

Das bedeutet aber, dass $r = \mathrm{mod}(a, n)$. Man überzeugt sich leicht davon, dass umgekehrt zwei Zahlen, die beide den gleichen Rest bei Division durch n haben, kongruent modulo n sind.

Mit anderen Worten: Kongruente Zahlen (modulo n) sind genau die, die den selben Rest bei Division (durch n) haben.

BEISPIEL 4.23

Einige Beispiele:

$$\begin{aligned} 7 &\equiv_2 5, \quad \text{weil } 2 \mid (7-5), \quad \textit{oder} \quad 7 \bmod 2 = 1 = 5 \bmod 2. \\ 4 &\equiv_2 10, \quad \text{weil } 2 \mid (4-10), \quad \textit{oder} \quad 4 \bmod 2 = 0 = 10 \bmod 2. \\ 27 &\equiv_5 12, \quad \text{weil} \quad 5 \mid (27-12), \quad \textit{oder} \quad 27 \bmod 5 = 2 = 12 \bmod 5. \end{aligned}$$

SATZ 4.24

$\equiv_n$ ist eine Äquivalenzrelation auf $\mathbb{Z}$.

Beweis. Sei $n \in \mathbb{N}$. Zu zeigen ist, dass die Relation $\equiv_n$ reflexiv, symmetrisch und transitiv ist.

- Reflexivität: Sei $a \in \mathbb{Z}$. Dann gilt

  $$a \equiv_n a, \text{ weil } n \mid \underbrace{(a-a)}_{=0}.$$

  Somit ist die Relation reflexiv.

- Symmetrie: Seien $a, b \in \mathbb{Z}$ und es gelte $a \equiv_n b$. Nach Definition von $\equiv_n$ folgt, dass $n \mid a - b$, d.h. $a - b = qn$ für ein $q \in \mathbb{Z}$. Damit gilt $b - a = (-q)n$ und auch $-q \in \mathbb{Z}$. Daraus folgt, dass $n \mid b - a$, was wiederum $b \equiv_n a$ zur Folge hat. Somit ist die Relation symmetrisch.

- Transitivität: Seien $a, b, c \in \mathbb{Z}$ mit $a \equiv_n b$ und $b \equiv_n c$, d.h.

  $$a \equiv_n b \qquad\qquad b \equiv_n c.$$

  Nach Definition bedeutet das, dass für bestimmte $q_1, q_2 \in \mathbb{Z}$ gilt:

  $$a - b = q_1 n \qquad\qquad b - c = q_2 n.$$

  Zu zeigen ist, dass $a \equiv_n c$, d.h. zu zeigen ist, dass ein $q \in \mathbb{Z}$ existiert, sodass $a - c = qn$. Nun gilt:

  $$a - c = (a-b) + (b-c) = q_1 n + q_2 n = \underbrace{(q_1 + q_2)}_{=:q} n.$$

  Mit der Wahl $q := q_1 + q_2 \in \mathbb{Z}$ gilt also $a - c = qn$, damit ist $a \equiv_n c$ bewiesen.

□

BEISPIEL 4.25

Sei $f\colon X \to Y$. Dann ist $\sim_f \subseteq X \times X$ definiert durch

$$a \sim_f b :\Leftrightarrow f(a) = f(b)$$

eine Äquivalenzrelation auf X.

Eine Äquivalenzrelation auf einer Menge X teilt X in Teilmengen auf, die Äquivalenzklassen genannt werden.

DEFINITION 4.26: ÄQUIVALENZKLASSE

Für eine Äquivalenzrelation $\sim$ auf einer Menge X und ein Element $a \in X$ ist die *Äquivalenzklasse* von a bzgl. $\sim$ definiert als

$$[a]_\sim := \{b \in X \mid a \sim b\}.$$

Jedes Element der Äquivalenzklasse $[a]_\sim$ nennt man einen *Repräsentanten* der Klasse.

Wenn aus dem Kontext klar ersichtlich ist, von welcher Äquivalenzrelation die Rede ist, wird oft einfach $[a]$ für die Äquivalenzklasse von a geschrieben. Für die Äquivalenzklasse von a bzgl. $\equiv_n$ schreibt man meist kurz $[a]_n$ (statt $[a]_{\equiv_n}$) oder wieder nur $[a]$.

BEISPIEL 4.27

Auf $\mathbb{Z}$ betrachten wir die Äquivalenzrelation $\equiv_5$. Die Äquivalenzklassen von 1 bzw. 3 lauten:

$$\begin{aligned}
[1]_5 &= \{b \in \mathbb{Z} \mid 1 \equiv_5 b\} = \{b \in \mathbb{Z} \mid \exists q \in \mathbb{Z} : 5q = b - 1\} = \{5q + 1 \mid q \in \mathbb{Z}\} \\
&= \{\dots, -9, -4, 1, 6, 11, \dots\} \\
[3]_5 &= \{b \in \mathbb{Z} \mid 3 \equiv_5 b\} = \{b \in \mathbb{Z} \mid \exists q \in \mathbb{Z} : 5q = b - 3\} = \{5q + 3 \mid q \in \mathbb{Z}\} \\
&= \{\dots, -7, -2, 3, 8, 13, \dots\}
\end{aligned}$$

Aus jeder Klasse $[a]_5$ kann man natürlich a selbst als Repräsentanten wählen, aber natürlich auch jedes andere $b \in [a]_5$, z.B.

$$[1]_5 = [6]_5 = [106]_5 = [-44]_5 = \dots$$

Äquivalenzklassen liefern disjunkte Zerlegungen der gegebenen Menge.

DEFINITION 4.28

Eine Menge (von Mengen) P ist eine *Partition* von M genau dann, wenn

1. $\forall A \in P :\ A \neq \emptyset$,

2. $\forall m : m \in M \Leftrightarrow \exists A \in P : \ m \in A$, [a]

3. $\forall A, B \in P, A \neq B : A \cap B = \emptyset$.

[a] Definiert man die *Vereinigung* von Mengen $\bigcup X := \{x \mid \exists Y \in X : x \in Y\}$, so kann man Bedingung 2 auch schreiben als $\bigcup P = M$. Man beachte: $\bigcup X$ erlaubt die Vereinigung von unendlich vielen Mengen. Im Falle einer endlichen Menge $X = \{X_1, \ldots, X_n\}$ gilt allerdings:

$$\bigcup X = \{x \mid \exists Y \in \{X_1, \ldots, X_n\} : x \in Y\} = \{x \mid x \in X_1 \vee \ldots \vee x \in X_n\} = X_1 \cup \ldots \cup X_n.$$

SATZ 4.29

Für jede Äquivalenzrelation $\sim$ auf einer Menge X ist die Menge der Äquivalenzklassen $\{[a]_\sim \mid a \in X\}$ eine Partition von X.

BEISPIEL 4.30

Auf $\mathbb{Z}$ betrachten wir die Äquivalenzrelation $\equiv_5$. Die Äquivalenzklassen bzgl. $\equiv_5$ lauten:

$$\begin{aligned}
[0] &= \{5q \mid q \in \mathbb{Z}\} = \{\ldots, -10, -5, 0, 5, 10, \ldots\} \\
[1] &= \{5q+1 \mid q \in \mathbb{Z}\} = \{\ldots, -9, -4, 1, 6, 11, \ldots\} \\
[2] &= \{5q+2 \mid q \in \mathbb{Z}\} = \{\ldots, -8, -3, 2, 7, 12, \ldots\} \\
[3] &= \{5q+3 \mid q \in \mathbb{Z}\} = \{\ldots, -7, -2, 3, 8, 13, \ldots\} \\
[4] &= \{5q+4 \mid q \in \mathbb{Z}\} = \{\ldots, -6, -1, 4, 9, 14, \ldots\}
\end{aligned}$$

1. Jede Klasse ist nicht-leer, 2. die Vereinigung aller Klassen ergibt $\mathbb{Z}$, und 3. je zwei Klassen haben leeren Durchschnitt. Somit bildet die Menge

$$\mathbb{Z}_5 := \{[0], [1], [2], [3], [4]\}$$

eine Partition von $\mathbb{Z}$. Jede ganze Zahl liegt in *genau einer* Äquivalenzklasse. Alternativ kann $\mathbb{Z}_5$ auch geschrieben werden als

$$\{[-2], [-1], [0], [1], [2]\}.$$

ORDNUNGSRELATIONEN

DEFINITION 4.31: ORDNUNGSRELATION, TOTALORDNUNG

Eine Relation R auf einer Menge X heißt eine *Ordnungsrelation* auf X genau dann, wenn R reflexiv, antisymmetrisch und transitiv auf X ist. In diesem Fall heißt (X, R) eine *geordnete Menge*. Gilt außerdem

$$\forall x, y \in X : \ xRy \vee yRx,$$

so nennt man R eine *Totalordnung*.

Eine Ordnungsrelation ist eine Totalordnung, wenn je zwei beliebige Elemente der Menge X vergleichbar sind. Eine Ordnungsrelation, die keine Totalordnung ist wird auch *Halbordnung* genannt.

BEISPIEL 4.32

Die Teilbarkeitsrelation aus Beispiel 4.4 ist eine Ordnungsrelation, aber keine Totalordnung. Die Eigenschaften einer Ordnungsrelation haben wir bereits gezeigt. Die Relation ist keine Totalordnung, da nicht alle Elemente miteinander vergleichbar sind. Zum Beispiel gilt weder $3 \mid 4$ noch gilt $4 \mid 3$.

Ordnungsrelationen werden üblicherweise durch sogenannte *Hasse-Diagramme* dargestellt. Dabei wird auf die Kanten verzichtet, die sich durch Reflexivität oder Transitivität ergeben, und die Orientierung der Beziehung wird durch die Anordnung der Knoten veranschaulicht. Abbildung 4.2 zeigt das Hasse-Diagramm der Teilbarkeitsrelation aus Beispiel 4.4.

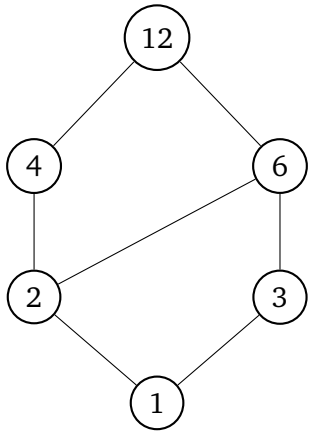


Abbildung 4.2: Hasse-Diagramm der Teilbarkeitsrelation aus Beispiel 4.4.

BEISPIEL 4.33

Die Teilmengenrelation $\subseteq$ aus Beispiel 4.6 auf $X = P(\{1, 2\})$ ist eine Ordnungsrelation, aber keine Totalordnung. Wir überprüfen die Eigenschaften einer Ordnungsrelation:

- Reflexivität: Für jede Menge gilt, dass sie Teilmenge von sich selbst ist, d.h. $\forall a \in X : a \subseteq a$ und somit ist R reflexiv.
- Antisymmetrie: Seien $a, b \in X$ mit $a \subseteq b$ und $b \subseteq a$. Das ist genau die Definition der Gleichheit von Mengen (siehe Kapitel 2, somit gilt $a = b$ und die Relation ist folglich antisymmetrisch.
- Transitivität: Seien $a, b, c \in X$ mit $a \subseteq b$ und $b \subseteq c$. Zu zeigen ist $a \subseteq c$, d.h. $\forall x \in a : x \in c$. Sei dazu x beliebig aber fix mit $x \in a$, zu zeigen ist $x \in c$. Da $a \subseteq b$ ist, gilt $x \in b$. Da $b \subseteq c$ ist, gilt $x \in c$. Die Relation ist somit transitiv.

Die Relation ist keine Totalordnung, da nicht alle Mengen miteinander vergleichbar sind. Zum Beispiel ist $\{1\}$ weder eine Teilmenge von $\{2\}$ noch ist $\{2\}$ eine Teilmenge von $\{1\}$.

BEISPIEL 4.34

Die Relation $\trianglelefteq \subseteq \mathbb{N}^2 \times \mathbb{N}^2$ definiert durch

$$(a,b) \trianglelefteq (c,d) :\Leftrightarrow (a < c) \vee (a = c \wedge b \leq d)$$

ist eine Totalordnung. Zuerst überprüfen wir die Eigenschaften einer Ordnungsrelation:

- Reflexivität: Seien $a, b \in \mathbb{N}$, dann gilt $a = a$ und $b \leq b$ und somit $(a,b) \trianglelefteq (a,b)$
- Antisymmetrie: Seien $a, b, c, d \in \mathbb{N}$ und angenommen es gilt sowohl $(a,b) \trianglelefteq (c,d)$ als auch $(c,d) \trianglelefteq (a,b)$. Da nicht $a < c$ und $c < a$ gelten kann, muss $a = c$ und $b \leq d \wedge d \leq b$ gelten. Daraus folgt $b = d$ und damit $(a,b) = (c,d)$.
- Transitivität: Seien $a, b, c, d, e, f \in \mathbb{N}$ mit $(a,b) \trianglelefteq (c,d)$ und $(c,d) \trianglelefteq (e,f)$, d.h.,

  $$\big(\underbrace{a < c}_{=A_1} \vee \underbrace{(a = c \wedge b \leq d)}_{=A_2}\big) \wedge \big(\underbrace{c < e}_{=B_1} \vee \underbrace{(c = e \wedge d \leq f)}_{=B_2}\big).$$

  Fallunterscheidung:

  - Fall 1: $A_1 \wedge B_1 \colon a < c \wedge c < e$ daher $a < e$ und somit $(a,b) \trianglelefteq (e,f)$.
  - Fall 2: $A_1 \wedge B_2 \colon a < c \wedge c = e \wedge d \leq f$ daher $a < e$ und somit $(a,b) \trianglelefteq (e,f)$.
  - Fall 3: $A_2 \wedge B_1 \colon a = c \wedge b \leq d \wedge c < e$ daher $a < e$ und somit $(a,b) \trianglelefteq (e,f)$.
  - Fall 4: $A_2 \wedge B_2 \colon a = c \wedge b \leq d \wedge c = e \wedge d \leq f$ daher $a = e \wedge b \leq f$ und somit $(a,b) \trianglelefteq (e,f)$.

$\trianglelefteq$ ist eine Totalordnung: Seien $(a,b), (c,d) \in \mathbb{N}^2$. Für a, c gilt entweder $a < c$, $a > c$ oder $a = c$. Im ersten Fall ist $(a,b) \trianglelefteq (c,d)$, im zweiten Fall $(c,d) \trianglelefteq (a,b)$. Im dritten Fall betrachten wir b, d. Auch hier gilt $b < d$ oder $b > d$ oder $b = d$. Hier gilt im ersten Fall $(a,b) \trianglelefteq (c,d)$, im zweiten $(c,d) \trianglelefteq (a,b)$ und im dritten gilt beides. Zwei beliebige Elemente sind also immer vergleichbar.

DEFINITION 4.35: MINIMAL,KLEINSTES

Sei $\preceq$ eine Ordnungsrelation auf X:

- Ein Element $m \in X$ heißt *minimales Element* (in X bzgl. $\preceq$) genau dann, wenn für alle $x \in X$ mit $x \preceq m$ gilt $x = m$.

- Ein Element $k \in X$ heißt *kleinstes Element* (in X bzgl. $\preceq$) genau dann, wenn für alle $x \in X$ gilt $k \preceq x$.

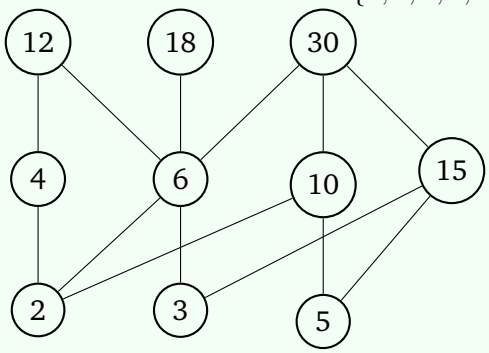
DEFINITION 4.36: MAXIMAL,GRÖSSTES

- Ein Element $m \in X$ heißt *maximales Element* (in X bzgl. $\preceq$) genau dann, wenn für alle $x \in X$ mit $m \preceq x$ gilt $x = m$.
- Ein Element $g \in X$ heißt *größtes Element* (in X bzgl. $\preceq$) genau dann, wenn für alle $x \in X$ gilt $x \preceq g$.

BEISPIEL 4.37

Für die Teilbarkeitsrelation aus Beispiel 4.4 auf der Menge $X = \{1, 2, 3, 4, 6, 12\}$ gilt: 1 ist das minimale und das kleinste Element und 12 ist das maximale und das größte Element. In diesem Beispiel gibt es nur ein minimales Element, das mit dem kleinsten Element übereinstimmt, da 1 mit allen anderen Zahlen in X vergleichbar ist. Da Teilbarkeit keine Totalordnung ist, muss das Gleiche aber nicht für Teilbarkeit als Relation auf beliebigen Mengen gelten.

Betrachten wir nun die Teilbarkeitsrelation auf $Y = \{2, 3, 5, 4, 6, 10, 12, 15, 18, 30\}$.



Die minimalen Elemente in Y bezüglich der Relation $|$ sind $\{2, 3, 5\}$, die maximalen Elemente sind $\{12, 18, 30\}$. Es gibt weder ein kleinstes noch ein größtes Element.

BEISPIEL 4.38

Wir betrachten die Relation aus Beispiel 4.34 auf $X = \{0, 1, 2\}^2$. Auf dieser Teilmenge gilt:

$$(0,0) \trianglelefteq (0,1) \trianglelefteq (0,2) \trianglelefteq (1,0) \trianglelefteq (1,1) \trianglelefteq (1,2) \trianglelefteq (2,0) \trianglelefteq (2,1) \trianglelefteq (2,2).$$

In X ist das kleinste Element bezüglich $\trianglelefteq$ also $(0, 0)$ und das größte Element $(2, 2)$.

KAPITEL 5

ELEMENTARE BEGRIFFE DER ZAHLENTHEORIE

EUKLIDISCHER ALGORITHMUS UND DIOPHANTISCHE GLEICHUNGEN

DEFINITION 5.1

Seien $a, b \in \mathbb{Z}$ mit $a \neq 0$ oder $b \neq 0$, dann ist der *größte gemeinsame Teiler* $\mathrm{ggT}(a, b)$ die (eindeutig bestimmte) größte ganze Zahl d mit $d \mid a$ und $d \mid b$.

Für jede ganze Zahl q mit $q \mid a$ und $q \mid b$ muss gelten, dass $q \leq \mathrm{ggT}(a, b)$. Der größte gemeinsame Teiler ist immer positiv, also eine natürliche Zahl (wäre $-q = \mathrm{ggT}(a, b) < 0$, dann ist auch q ein gemeinsamer Teiler von a und b, aber $q > -q$, Widerspruch!). Außerdem gilt $\mathrm{ggT}(a, b) = \mathrm{ggT}(|a|, |b|) = \mathrm{ggT}(b, a)$, d.h. der größte gemeinsame Teiler hängt nicht vom Vorzeichen der gegebenen Zahlen ab und ist in den Argumenten symmetrisch. Eine erste Methode, die man in der Schule zur Berechnung des ggT kennenlernt, ist über die Primfaktorenzerlegung.

DEFINITION 5.2

Eine Zahl $p \in \mathbb{N}$ ist genau dann eine *Primzahl*, wenn die folgenden zwei Bedingungen erfüllt sind:

1. $p > 1$.
2. Für alle $a, b \in \mathbb{N}$ mit $p = a \cdot b$ gilt $a = 1$ oder $b = 1$.

Die Primzahlen unter 100 sind:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.$$

SATZ 5.3

*Jede natürliche Zahl $n > 1$ lässt sich als Produkt von Primzahlen schreiben (*Primfaktoren-

zerlegung*), d.h. es existieren* $k \in \mathbb{N}$*, Primzahlen* $p_1, \ldots, p_k$ *und* $e_1, \ldots, e_k \in \mathbb{N}$ *mit*

$$n = \prod_{j=1}^{k} p_j^{e_j}.$$

Bis auf die Reihenfolge sind die Faktoren eindeutig bestimmt und heißen Primfaktoren.

BEISPIEL 5.4

Die Primfaktoren können z.B. durch sukzessives Dividieren von Primzahlen beginnend mit dem kleinsten Teiler gewonnen werden:

$$\begin{aligned} 28 &= 2 \cdot 14 = 2 \cdot 2 \cdot 7 = 2^2 \cdot 7, \\ 1235 &= \cdots = 5 \cdot 13 \cdot 19, \\ 102432 &= \cdots = 2^5 \cdot 3 \cdot 11 \cdot 97. \end{aligned}$$

Um $ggT(a, b)$ zu berechnen, kann man zuerst die Primfaktorenzerlegung von a und b bestimmen und anschließend alle Faktoren, die in beiden Zerlegungen vorkommen, aufmultiplizieren, also zum Beispiel $\mathrm{ggT}(100, 60) = 20$, da

$$100 = 2^2 \cdot 5^2 = \underline{2} \cdot \underline{2} \cdot \underline{5} \cdot 5 \quad \text{und} \quad 60 = 2^2 \cdot 3 \cdot 5 = \underline{2} \cdot \underline{2} \cdot 3 \cdot \underline{5} \quad \text{also} \quad \mathrm{ggT}(100, 60) = 2 \cdot 2 \cdot 5 = 20.$$

Es ist derzeit kein effizientes Verfahren bekannt, um die Primfaktorenzerlegung einer beliebigen ganzen Zahl zu berechnen, und es ist eine der großen offenen Fragen der Informatik, ob Primfaktorenzerlegungen in polynomieller Zeit berechnet werden können. Auf dieser Tatsache basieren Verschlüsselungsverfahren wie RSA, das wir später noch betrachten werden.

DEFINITION 5.5

Zwei Zahlen $a, b \in \mathbb{Z}$ heißen *relativ prim* (*teilerfremd*) genau dann, wenn $\mathrm{ggT}(a, b) = 1$.

SATZ 5.6

Seien $a, b \in \mathbb{Z}$ *und* $z \in \mathbb{Z}$*. Dann gilt* $\mathrm{ggT}(a, b) = \mathrm{ggT}(a - zb, b)$.

Beweis. Sei $g = \mathrm{ggT}(a - zb, b)$, wir wissen also

$$g \mid a - zb \tag{5.1}$$
$$g \mid b \tag{5.2}$$
$$\forall c : c \mid a - zb \wedge c \mid b \Rightarrow c \leq g. \tag{5.3}$$

Wir zeigen, dass g die charakteristische Eigenschaft des $\mathrm{ggT}(a, b)$ erfüllt:

1. Wegen (5.2) gilt auch $g \mid zb$. Wegen (5.1) gilt auch $g \mid a - zb + zb$ also $g \mid a$.

2. $g \mid b$ wissen wir schon wegen (5.2), somit ist g ein gemeinsamer Teiler von a und b.
3. Sei nun d ebenfalls ein gemeinsamer Teiler von a und b, dann gilt auch $d \mid a - zb$ und wegen (5.3) ist daher $d \leq g$.

Damit ist g der größte gemeinsame Teiler von a und b. □

Im Speziellen gilt daher auch:

SATZ 5.7

Seien $a, b \in \mathbb{Z}$. *Dann gilt* $\mathrm{ggT}(a, b) = \mathrm{ggT}(a + b, b) = \mathrm{ggT}(a - b, b)$.

Aus Satz 5.6 folgt ein Algorithmus zur Bestimmung des ggT, der erstmals von Euklid (ca. 360 v. Chr. bis ca. 280 v. Chr.) schriftlich erwähnt wurde und daher nach ihm benannt ist.

Seien $a, b \in \mathbb{N}_0$ mit $a > b$ gegeben. Die Division (siehe Definition 4.22) dieser beiden Zahlen liefert $q, r \in \mathbb{N}_0$ mit $r < b$ sodass

$$a = qb + r \quad \longrightarrow r = a - qb.$$

Laut Satz 5.6 gilt

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

Der $\mathrm{ggT}(a, b)$ ist somit auf den ggT der *kleineren Zahlen* b, r zurückgeführt worden. Dieser Schritt kann rekursiv wiederholt werden: Wir berechnen den $\mathrm{ggT}(34, 28)$:

1. Division liefert $34 = 28 \cdot 1 + 6$, d.h. $q = 1$ und $r = 6$. Dann gilt
$$\mathrm{ggT}(34, 28) = \mathrm{ggT}(34 - 28 \cdot 1, 28) = \mathrm{ggT}(6, 28) = \mathrm{ggT}(28, 6)$$
2. Division liefert $28 = 6 \cdot 4 + 4$, d.h. $q = 4$ und $r = 4$. Dann gilt
$$\mathrm{ggT}(28, 6) = \mathrm{ggT}(28 - 6 \cdot 4, 6) = \mathrm{ggT}(4, 6) = \mathrm{ggT}(6, 4)$$
3. Division liefert $6 = 4 \cdot 1 + 2$, d.h. $q = 1$ und $r = 2$. Dann gilt
$$\mathrm{ggT}(6, 4) = \mathrm{ggT}(6 - 4 \cdot 1, 4) = \mathrm{ggT}(2, 4) = \mathrm{ggT}(4, 2)$$
4. Division liefert $4 = 2 \cdot 2 + 0$, d.h. $q = 2$ und $r = 0$. Dann gilt
$$\mathrm{ggT}(4, 2) = \mathrm{ggT}(4 - 2 \cdot 2, 2) = \mathrm{ggT}(0, 2) = \mathrm{ggT}(2, 0) = 2.$$

In jedem Zwischenschritt sind die Einträge x, y im $\mathrm{ggT}(x, y)$ Linearkombinationen der Eingangszahlen $a = 34$ und $b = 28$, z.B. im ersten Schritt

$$\mathrm{ggT}(a, b) = \mathrm{ggT}(34, 28) = \mathrm{ggT}(34 - 28, 28) = \mathrm{ggT}(28, \underbrace{34 - 28}_{=6}) = \mathrm{ggT}(b, 1 \cdot a + (-1) \cdot b).$$

Im zweiten Schritt erhalten wir:

$$\begin{aligned}\mathrm{ggT}(a, b) = \mathrm{ggT}(b, r) = \mathrm{ggT}(28, \underbrace{34 - 28}_{=6}) &= \mathrm{ggT}(28 - (\underbrace{34 - 28}_{=6}) \cdot 4, \underbrace{34 - 28}_{=6}) \\ &= \mathrm{ggT}(\underbrace{-4 \cdot 34 + 5 \cdot 28}_{=4}, \underbrace{34 - 28}_{=6}) \\ &= \mathrm{ggT}(\underbrace{34 - 28}_{=6}, \underbrace{5 \cdot 28 - 4 \cdot 34}_{=4}) \\ &= \mathrm{ggT}(1 \cdot a + (-1) \cdot b, (-4) \cdot a + 5 \cdot b).\end{aligned}$$

In einer Tabelle können die Koeffizienten der Linearkombinationen mitberechnet werden:

I	34	1	0
II	28	0	1
$III = I - II$	6	1	-1
$IV = II - 4III$	4	-4	5
$V = III - IV$	2	5	-6
$VI = IV - 2V$	0	-14	17

Aus dieser Tabelle können die Koeffizienten 5 und -6 abgelesen werden, d.h.

$$\mathrm{ggT}(34, 28) = 2 = 5 \cdot 34 - 6 \cdot 28.$$

In der letzten Zeile können die Koeffizienten abgelesen werden, für die die Linearkombination aus a, b genau 0 ergibt:

$$0 = -14 \cdot 34 + 17 \cdot 28.$$

Diese Koeffizienten sind *immer* genau $-\frac{b}{\mathrm{ggT}(a,b)}$ und $\frac{a}{\mathrm{ggT}(a,b)}$.

SATZ 5.8

Seien $a, b \in \mathbb{N}$. Dann existieren ganze Zahlen $s, t \in \mathbb{Z}$ mit

$$\mathrm{ggT}(a, b) = s \cdot a + t \cdot b.$$

Der *Erweiterte Euklidische Algorithmus (EEA)* folgt dem Vorgehen aus dem Beispiel und berechnet den größten gemeinsamen Teiler und die Koeffizienten, mit denen der ggT als Linearkombination der Eingabewerte dargestellt werden kann: Gegeben $a, b \in \mathbb{N}$ mit $a > b$:

1. Initialisierung: $r_0 := a, r_1 := b, s_0 := 1, s_1 := 0, t_0 := 0, t_1 := 1$ und $k := 1$
2. While $r_k \neq 0$ Do

(i) $q_k := \text{quot}(r_{k-1}, r_k)$ (Quotient der Division von r_{k-1} durch r_k, d.h. $r_{k-1} = q_k r_k + R$)
(ii) $r_{k+1} := r_{k-1} - q_k r_k$ (d.h. $r_{k+1} = R =$ Rest der Division von r_{k-1} durch r_k)
(iii) $s_{k+1} := s_{k-1} - q_k s_k$ und $t_{k+1} := t_{k-1} - q_k t_k$ (Bestimmung der Linearkoeffizienten)
(iv) $k := k + 1$

3. Return $\text{ggT}(a, b) = r_{k-1}$, (s_{k-1}, t_{k-1})

In jedem Schritt gilt $r_k = s_k a + t_k b$.

BEISPIEL 5.9

Wir bestimmen den $\text{ggT}(126, 81)$ mit dem EEA. Dazu notieren wir in der Tabelle zusätzlich den Schleifenindex k und den Quotienten q:

	k	r	s	t	q
I	0	126	1	0	
II	1	81	0	1	1
$III = I - II$	2	45	1	-1	1
$IV = II - II$	3	36	-1	2	1
$V = III - IV$	4	9	2	-3	4
$VI = IV - 4V$	5	0	-9	14	

Das heißt,

$$\text{ggT}(126, 81) = 9 = 2 \cdot 126 + (-3) \cdot 81 \quad \text{und} \quad 0 = -9 \cdot 126 + 14 \cdot 81.$$

Der erweiterte Euklidische Algorithmus kann verwendet werden, um *ganzzahlige* Lösungen von linearen Gleichungen mit *ganzzahligen* Koeffizienten zu bestimmen. Solche Gleichungen werden *diophantische Gleichungen* genannt.

BEISPIEL 5.10

Ein Bauer hat 500 Taler gespart, um neue Hühner und Kühe zu kaufen. Auf dem Markt stellt er fest, dass eine Kuh 17 Taler kostet und ein Huhn 5 Taler. Wie viele Hühner und Kühe kann er kaufen, wenn er die ganzen 500 Taler ausgeben möchte?

Schritt 1: Wir berechnen mit EEA den ggT von 17 und 5.

$$\text{ggT}(17, 5) = 1 = -2 \cdot 17 + 7 \cdot 5 \tag{5.4}$$

und

$$0 = 5 \cdot 17 - 17 \cdot 5. \tag{5.5}$$

Wenn wir die Gleichung 5.4 mit 500 multiplizieren erhalten wir

$$500 = -1000 \cdot 17 + 3500 \cdot 5, \tag{5.6}$$

wenn der Bauer also -1000 Kühe und 3500 Hühner kauft, gibt er genau 500 Taler aus. Dieser Handel dürfte allerdings schwierig werden. Mithilfe von (5.5) können wir alter-

native Lösungen berechnen. Wenn wir (5.5) mit einer beliebigen Zahl x multiplizieren erhalten wir

$$0 = 5x \cdot 17 - 17x \cdot 5.$$

Diese Gleichung können wir zu (5.6) addieren und erhalten so

$$500 = (-1000 + 5x) \cdot 17 + (3500 - 17x) \cdot 5.$$

Jetzt bleibt festzustellen, ob es ganzzahlige Werte für x gibt, sodass sowohl $-1000+5x \geq 0$ als auch $3500 - 17x \geq 0$.

$$-1000 + 5x \geq 0 \quad \longrightarrow \quad x \geq 200 \quad \text{und} \quad 3500 - 17x = 0 \quad \longrightarrow \quad x \leq \tfrac{3500}{17} \simeq 205.882.$$

Für $200 \leq x \leq 205$ sind demnach beide Koeffizienten nichtnegativ. Das sind nur endliche viele Kombinationen, die wir ausprobieren können:

$$\begin{aligned}
x = 200 \quad &\longrightarrow \quad 500 = 0 \cdot 17 + 100 \cdot 5 \\
x = 201 \quad &\longrightarrow \quad 500 = 5 \cdot 17 + 83 \cdot 5 \\
x = 202 \quad &\longrightarrow \quad 500 = 10 \cdot 17 + 66 \cdot 5 \\
x = 203 \quad &\longrightarrow \quad 500 = 15 \cdot 17 + 49 \cdot 5 \\
x = 204 \quad &\longrightarrow \quad 500 = 20 \cdot 17 + 32 \cdot 5 \\
x = 205 \quad &\longrightarrow \quad 500 = 25 \cdot 17 + 15 \cdot 5
\end{aligned}$$

SATZ 5.11

Die diophantische Gleichung $ax + by = c$ *hat genau dann (mindestens) eine ganzzahlige Lösung, wenn c ein Vielfaches von* $\mathrm{ggT}(a, b)$ *ist, d.h.* $c = q\,\mathrm{ggT}(a, b)$ *für ein* $q \in \mathbb{Z}$.

Betrachten wir die Gleichung $ax + by = c$ mit gegebenen $a, b, c \in \mathbb{Z}$. Sei $d = \mathrm{ggT}(a, b)$, dann können mit dem Erweiterten Euklidischen Algorithmus $d = \mathrm{ggT}(a, b)$ und Koeffizienten x_0, y_0 bzw. x_1, y_1 berechnet werden, sodass

$$ax_0 + by_0 = d \tag{5.7}$$
$$ax_1 + by_1 = 0. \tag{5.8}$$

Wenn c ein Vielfaches des größten gemeinsamen Teilers von a, b ist, wenn also gilt $c = qd$, dann kann (5.7) auf beiden Seiten mit q multipliziert werden, und es gilt

$$a(qx_0) + b(qy_0) = qd,$$

d.h. (qx_0, qy_0) ist eine ganzzahlige Lösung der gegebenen Gleichung. Außerdem kann (5.8) auf beiden Seiten mit einer ganzen Zahl k multipliziert werden, und wir erhalten

$$a(kx_1) + b(ky_1) = 0.$$

Wenn die beiden letzten Gleichungen addiert werden, ergibt das

$$a(qx_0 + kx_1) + b(qy_0 + ky_1) = c,$$

d.h. jedes Tupel der Form $(qx_0 + kx_1, qy_0 + ky_1)$ für beliebiges $k \in \mathbb{Z}$ ist eine *ganzzahlige* Lösung der gegebenen diophantischen Gleichung.

Wir erinnern daran, dass $x_1 = -\frac{b}{\mathrm{ggT}(a,b)}$ und $y_1 = \frac{a}{\mathrm{ggT}(a,b)}$.

Zusammengefasst ergibt das folgenden Satz.

SATZ 5.12: FORTSETZUNG VON SATZ 5.11

Ist (x_0, y_0) eine (mit dem erweiterten Euklidischen Algorithmus berechnete) ganzzahlige Lösung von $ax_0 + by_0 = \mathrm{ggT}(a, b)$ und sei $c = q\,\mathrm{ggT}(a, b)$. Dann ist (qx_0, qy_0) eine Lösung von $ax + by = c$. Seien weiters $x_1 = -\frac{b}{\mathrm{ggT}(a,b)}$ bzw. $y_1 = \frac{a}{\mathrm{ggT}(a,b)}$, dann sind alle ganzzahligen Lösungen der Gleichung sind gegeben durch

$$\{(qx_0 + kx_1, qy_0 + ky_1) \mid k \in \mathbb{Z}\}.$$

BEISPIEL 5.13

- Gegeben ist die diophantische Gleichung $251x + 127y = 16$. Wir wenden den EEA für $a = 251$ und $b = 127$ an und erhalten

  $$251 \cdot 42 + 127 \cdot (-83) = 1,$$

  d.h. $x_0 = 42$, $y_0 = -83$ und $q = 16$. Da der ggT dieser beiden Zahlen 1 ist, ist die diophantische Gleichung sogar für jede (ganzzahlige) rechte Seite lösbar (da jede ganze Zahl durch 1 teilbar ist) und eine Lösung ist

  $$(16 \cdot 42, 16 \cdot (-83)) = (672, -1328).$$

  Laut Formel ist $x_1 = -\frac{b}{\mathrm{ggT}(a,b)} = -127$ und $y_1 = \frac{a}{\mathrm{ggT}(a,b)} = 251$, als Lösungsmenge erhält man damit

  $$L = \{(672 - 127k, -1328 + 251k) \mid k \in \mathbb{Z}\}.$$

- Gegeben ist die diophantische Gleichung $60x + 128y = c$ für ein $c \in \mathbb{Z}$. Wir berechnen mit dem EEA

  $$60 \cdot 15 + 128 \cdot (-7) = 4.$$

  Damit gilt, dass die diophantische Gleichung genau dann ganzzahlige Lösungen besitzt, wenn c ein Vielfaches von 4 ist. Zum Beispiel gilt für $c = -12$, dass

  $$L = \{(-45 - 32k, 21 + 15k) \mid k \in \mathbb{Z}\}.$$

  Wenn c kein Vielfaches von 4 ist, also z.B. $c = 1, 17, -121, \ldots$, dann ist die Lösungsmenge die leere Menge.

Modulare Arithmetik

Wir definieren zu $n \in \mathbb{N}$ die Menge

$$\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}.$$

Diese Menge kann als Menge von Repräsentanten der Äquivalenzrelation *Kongruenz modulo* n (siehe Satz 4.24) betrachtet werden (also auch als die Menge der Zahlen, die als Rest bei einer Division durch n auftreten können). Die herkömmliche Addition und Multiplikation funktionieren auf $\mathbb{Z}_n$ *nicht* mehr, da die Operationen nicht abgeschlossen sind, d.h. die Summe bzw. das Produkt zweier Zahlen in $\mathbb{Z}_n$ liegen nicht mehr notwendigerweise ebenfalls in $\mathbb{Z}_n$. Wir können auf $\mathbb{Z}_n$ aber eine *neue Additionsoperation*, die sog. *modulare Addition* $\oplus_n$, und eine *neue Multiplikationsoperation*, die sog. *modulare Multiplikation* $\odot_n$ so definieren, dass jeweils das Ergebnis modulo n genommen wird, d.h. jeweils der Rest bei Division durch n.

Definition 5.14

Sei $n \in \mathbb{N}$ und seien $a, b \in \mathbb{Z}_n$. Wir definieren auf $\mathbb{Z}_n$

$$a \oplus_n b := (a + b) \bmod n$$
$$a \odot_n b := (a \cdot b) \bmod n$$

Beispiel 5.15

In $\mathbb{Z}_7 = \{0, 1, 2, 3, 4, 5, 6\}$ gilt:

$$3 \oplus_7 4 = 0 \qquad 3 \odot_7 4 = 5 \qquad 4 \odot_7 5 = 6$$

Da $\mathbb{Z}_n$ eine endliche Menge ist, können die Ergebnisse der Addition und Multiplikation von allen Kombinationen von Zahlen in $\mathbb{Z}_n$ berechnet und in Tabellen angegeben werden.

Additions- und Multiplikationstabellen für $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ bzw. $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$:

$\oplus_4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\odot_4$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

$\oplus_5$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\odot_5$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Es ist in der Mathematik üblich, für die modulare Addition und Multiplikation einfach wieder „$+$“ und „$\cdot$“ so wie für die „übliche“ Addition und Multiplikation zu schreiben. In irgendeiner Weise muss durch geeignete Notation angezeigt werden, modulo welchem n gerechnet wird. Dies geschieht häufig durch ein nachgestelltes „$(\mathrm{mod}\ n)$“.

BEISPIEL 5.16

In $\mathbb{Z}_7 = \{0, 1, 2, 3, 4, 5, 6\}$ gilt:

$$3 + 4 = 0 \pmod 7 \qquad 3 \cdot 4 = 5 \pmod 7 \qquad 4 \cdot 5 = 6 \pmod 7$$

DEFINITION 5.17

Sei $n \in \mathbb{N}$. Das neutrale Element bezüglich (modularer) Addition in $\mathbb{Z}_n$ ist 0, d.h. es gilt

$$a + 0 = 0 + a = a \pmod n \quad \text{für jedes } a \in \mathbb{Z}_n.$$

Zu gegebenem $a \in \mathbb{Z}_n$ ist die *additive Inverse* definiert als jenes $b \in \mathbb{Z}_n$, für das gilt

$$a + b = 0 \pmod n.$$

Oft wird für die additive Inverse b auch $-a$ geschrieben, sodass $a + (-a) = 0 \pmod n$.

Für $a \in \mathbb{Z}_n \setminus \{0\}$ gilt $-a \pmod n = n - a$ und $-0 \pmod n = 0$.

SATZ 5.18

Sei $n \in \mathbb{N}$. Jede Zahl $a \in \mathbb{Z}_n$ besitzt eine eindeutig bestimmte additive Inverse.

BEISPIEL 5.19

Die *Caesar-Verschlüsselung* nutzt die Eindeutigkeit der additiven Inversen in $\mathbb{Z}_{26}$ aus. Die Grundidee ist, dass die Buchstaben des Alphabets verschoben werden, zum Beispiel:

A	B	C	D	E	F	G	$\ldots$	U	V	W	X	Y	Z
F	G	H	I	J	K	L	$\ldots$	Z	A	B	C	D	E

Buchstaben können als Zahlen in $\mathbb{Z}_{26}$ codiert werden, d.h.

$$A = 0 \qquad B = 1 \qquad C = 2 \qquad \ldots \qquad X = 23 \qquad Y = 24 \qquad Z = 25.$$

Das Verschieben der Buchstaben entspricht dann einer modularen Addition, das heißt das Zeichen x wird mit dem Schlüssel k zu einem neuen Zeichen y durch

$$y = x + k \pmod{26}.$$

Eine Verschiebung um fünf Zeichen wie oben entspricht also $k = 5$ und

A	B	C	D	E	F	G	$\ldots$	U	V	W	X	Y	Z
0	1	2	3	4	5	6	$\ldots$	20	21	22	23	24	25
5	6	7	8	9	10	11	$\ldots$	25	0	1	2	3	4
F	G	H	I	J	K	L	$\ldots$	Z	A	B	C	D	E

Zum Decodieren muss $y = x + k \pmod{26}$ nach x aufgelöst werden, d.h. $x = y + (-k) \pmod{26}$ berechnet werden. (Der Schlüssel muss also bekannt sein.)

Das verschlüsselte Wort „RFYMJRFYNP“ (in Zahlen: $(17, 5, 24, 12, 9, 17, 5, 24, 13, 15)$) mit dem Schlüssel $k = 5$ wird mit

$$\begin{aligned} 17 + (-5) &= 12 \pmod{26} && = \text{„M“} \\ 5 + (-5) &= 0 \pmod{26} && = \text{„A“} \\ 24 + (-5) &= 19 \pmod{26} && = \text{„T“} \\ &\vdots && = \vdots \\ 13 + (-5) &= 8 \pmod{26} && = \text{„I“} \\ 15 + (-5) &= 10 \pmod{26} && = \text{„K“} \end{aligned}$$

im Klartext zu „MATHEMATIK“.

Es gilt also, dass *modulare Gleichungen* der Form $a + x = b \pmod{n}$ für jedes $n \in \mathbb{N}$ eine eindeutig bestimmte Lösung $x \in \mathbb{Z}_n$ besitzen, da die additive Inverse eindeutig bestimmt ist.

DEFINITION 5.20

Sei $n \in \mathbb{N}$. Das neutrale Element bezüglich (modularer) Multiplikation in $\mathbb{Z}_n$ ist 1, d.h. es gilt

$$a \cdot 1 = 1 \cdot a = a \pmod{n} \quad \text{für alle } a \in \mathbb{Z}_n.$$

Die Zahl a in $\mathbb{Z}_n$ heißt *invertierbar* (mod n), falls ein $b \in \mathbb{Z}_n$ existiert mit

$$a \cdot b = 1 \pmod{n}.$$

Man nennt b die *multiplikative Inverse* zu a. Oft wird als Bezeichnung für die multiplikative Inverse b auch $\frac{1}{a}$ oder a^{-1} geschrieben.

Eine multiplikative Inverse muss nicht zu jeder Zahl in $\mathbb{Z}_n$ existieren. In $\mathbb{Z}_4$ ist zum Beispiel $3^{-1} = 3 \pmod{4}$ (da $3 \cdot 3 = 1 \pmod{4}$), aber $a = 2$ besitzt keine multiplikative Inverse. In $\mathbb{Z}_5$, besitzt jede Zahl außer 0 eine multiplikative Inverse:

$$1^{-1} = 1 \pmod{5} \qquad 2^{-1} = 3 \pmod{5} \qquad 3^{-1} = 2 \pmod{5} \qquad 4^{-1} = 4 \pmod{5}.$$

Etwas überraschend kann der *erweiterte Euklidische Algorithmus* verwendet werden, um die multiplikative Inverse in $\mathbb{Z}_n$ zu bestimmen, sofern sie existiert.

BEISPIEL 5.21

Sei $a = 11$ und wir suchen $a^{-1} \in \mathbb{Z}_{26}$, d.h. gesucht ist $b \in \mathbb{Z}_{26}$ mit $11b = 1 \pmod{26}$. Diese Gleichung ist erfüllt, wenn es ein $q \in \mathbb{Z}$ gibt mit

$$11b + 26q = 1.$$

Das ist eine diophantische Gleichung, die mit dem EEA gelöst werden kann. Zuerst berechnen wir den $\mathrm{ggT}(11, 26) = 1 = (-7) \cdot 11 + 3 \cdot 26$. Nach Satz 5.11 existiert eine Lösung der Gleichung (weil der größte gemeinsame Teiler die rechte Seite der Gleichung teilt). Die Lösungen für b sind gegeben durch

$$\{-7 - 26k \mid k \in \mathbb{Z}\}$$

Für die modulare Inverse müssen wir nur ein $b \in \mathbb{Z}_{26}$ aus dieser Lösungsmenge bestimmen, und das ergibt sich durch die Wahl $k = -1$. Damit erhalten wir $a^{-1} = 19$.

Um die multiplikative Inverse b einer gegebenen Zahl $a \in \mathbb{Z}_n$ zu bestimmen, muss also eine Gleichung der Form

$$a \cdot b + n \cdot q = 1$$

gelöst werden. Nach Satz 5.11 besitzt die Gleichung genau dann eine Lösung, wenn $\mathrm{ggT}(a, n) \mid 1$, d.h. genau dann, wenn $\mathrm{ggT}(a, n) = 1$, also wenn a und n relativ prim sind.

Zu beliebigem $n \in \mathbb{N}$ wird die Menge der invertierbaren Zahlen in $\mathbb{Z}_n$ mit $\mathbb{Z}_n^*$ bezeichnet, d.h.

$$\mathbb{Z}_n^* := \{a \in \mathbb{Z}_n \mid \exists\, b \in \mathbb{Z}_n \colon ab = 1\},$$

und mit dem oben Besprochenen folgt

$$\mathbb{Z}_n^* = \{a \in \mathbb{Z}_n \mid \mathrm{ggT}(a, n) = 1\}.$$

Zum Beispiel gilt

$$\mathbb{Z}_{26}^* = \{1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25\},$$

und für p eine Primzahl gilt $\mathbb{Z}_p^* = \mathbb{Z}_p \backslash \{0\}$.

SATZ 5.22

Sei $n \in \mathbb{N}$. Jede Zahl $a \in \mathbb{Z}_n$ mit $\mathrm{ggT}(a, n) = 1$ besitzt eine eindeutig bestimmte multiplikative Inverse.

Für modulare Gleichungen der Form $a \cdot x = b \pmod{n}$ gilt

- Falls a und n teilerfremd sind, dann besitzt $a \cdot x = b \pmod{n}$ die eindeutig bestimmte Lösung $x := (a^{-1} \cdot b) \mod n$.
- Falls a und n einen (nichttrivialen) gemeinsamen Teiler haben, dann kann die Gleichung $a \cdot x = b \pmod{n}$ *keine* oder *mehr als eine* Lösung besitzen.

BEISPIEL 5.23

Die Gleichung $4x = 2 \pmod{n}$ besitzt für $n = 17$ eine eindeutig bestimmte Lösung. Die multiplikative Inverse von $a = 4$ in $\mathbb{Z}_{17}$ ist $a^{-1} = 13$. Multiplikation mit a^{-1} auf beiden Seiten liefert

$$x = (2 \cdot 13) \bmod 17 = 26 \bmod 17 = 9 \quad .$$

Probe: $4 \cdot 9 = 36 = 2 \pmod{17}$.

Für $n = 6$ besitzt $4x = 2 \pmod{n}$ die beiden Lösungen $x = 2$ und $x = 5$.

BEMERKUNG 5.24

Wenn das Ergebnis von Summe oder Produkt zweier ganzen Zahlen a, b modulo $n \in \mathbb{N}$ gesucht ist, dann spielt es keine Rolle, ob das Ergebnis $a + b$ (oder $a \cdot b$) modulo n genommen wird, oder ob die Summe oder das Produkt von a modulo n mit b modulo n berechnet wird und dann das Ergebnis modulo n genommen wird.

Seien $a, b \in \mathbb{Z}$ und $n \in \mathbb{N}$. Wenn $\alpha, \beta \in \mathbb{Z}_n$ so sind, dass

$$a = \alpha \pmod{n} \quad \text{und} \quad b = \beta \pmod{n},$$

dann gibt es $q, p \in \mathbb{Z}$ mit

$$a = qn + \alpha \quad \text{und} \quad b = pn + \beta.$$

Damit gilt

$$ab = (qn + \alpha)(pn + \beta) = qp\, n^2 + (\alpha p + \beta q)n + \alpha\beta = \alpha\beta \pmod{n}.$$

Also haben wir insgesamt, dass

$$a \cdot b = \alpha \cdot \beta \pmod{n}.$$

SATZ VON FERMAT UND RSA

SATZ 5.25: KLEINER SATZ VON FERMAT

Sei p eine Primzahl. Dann gilt für jedes $x \in \mathbb{Z}$ mit $\mathrm{ggT}(x, p) = 1$:

$$x^{p-1} = 1 \pmod{p}.$$

Beweis. Sei $x \in \mathbb{Z}$ mit $\mathrm{ggT}(x, p) = 1$. Wie in Abschnitt 5.2 sei $\mathbb{Z}_p^* = \{1, 2, \ldots, p-1\}$. Wir definieren (zu x) die Funktion

$$f : \mathbb{Z}_p^* \to \mathbb{Z}_p^*, \quad a \mapsto x \cdot a \bmod p.$$

Diese Funktion ist invertierbar, da

$$f(a) = b \quad \Leftrightarrow \quad b = x \cdot a \pmod{p} \quad \Leftrightarrow \quad a = x^{-1} \cdot b \pmod{p}.$$

Damit ist f auch bijektiv (von $\mathbb{Z}_p^*$ nach $\mathbb{Z}_p^*$), d.h. jedem $a \in \mathbb{Z}_p^*$ wird genau ein $b \in \mathbb{Z}_p^*$ zugeordnet. Damit wird die Menge $\mathbb{Z}_p^*$ bijektiv auf die Menge $f(\mathbb{Z}_p^*) = \{a \cdot x \bmod p \mid a \in \mathbb{Z}_p^*\}$ abgebildet, und das Produkt (modulo p) über alle Zahlen aus $\mathbb{Z}_p^*$ muss mit dem Produkt über alle Zahlen aus $f(\mathbb{Z}_p^*)$ übereinstimmen (die einzelnen Faktoren sind ja durch die bijektive Abbildung f nur umgeordnet!), d.h.

$$\begin{aligned}
&\prod_{b \in f(\mathbb{Z}_p^*)} b = \prod_{a \in \mathbb{Z}_p^*} a \\
\Leftrightarrow \quad &\prod_{a \in \mathbb{Z}_p^*} a \cdot x = \prod_{a \in \mathbb{Z}_p^*} a \pmod{p} \\
\Leftrightarrow \quad &x^{p-1} \prod_{a \in \mathbb{Z}_p^*} a = \prod_{a \in \mathbb{Z}_p^*} a \pmod{p} \\
\Leftrightarrow \quad &x^{p-1} = 1 \pmod{p}.
\end{aligned}$$

Im letzten Schritt wird auf beiden Seiten mit den multiplikativen Inversen von a (die alle existieren, da p eine Primzahl ist) multipliziert. □

Als Anwendung betrachten wir das Grundkonzept der *Datenverschlüsselung mit dem RSA-Algorithmus* (nach Rivest-Shamir-Adleman). Bei der RSA-Verschlüsselung handelt es sich um ein *public-key Verschlüsselungsverfahren*, d.h. es gibt einen öffentlichen Schlüssel (public-key) und einen privaten Schlüssel (private-key).

Wenn Alice eine Nachricht von Bob empfangen will, stellt sie einen öffentlichen Schlüssel e (wie *encrypt*) zur Verfügung, mit dem Bob seine Nachricht verschlüsselt und an Alice schickt. Alice verfügt über ihren privaten Schlüssel d (wie *decrypt*), mit dem sie die Nachricht wieder in Klartext übersetzt.

Die Schlüssel werden wie folgt erzeugt:

1. Alice wählt zwei (große) unterschiedliche Primzahlen p, q und berechnet $n = p \cdot q$ und $m = (p-1)(q-1)$.
2. Für den public-key wählt Alice eine Zahl e mit $\mathrm{ggT}(e, m) = 1$.
3. Für den private-key berechnet sie d mit $ed = 1 \pmod{m}$, also $d = e^{-1} \pmod{m}$. So ein d muss existieren, da e und m teilerfremd sind.
4. Alice gibt als public-key das Paar (n, e) öffentlich bekannt und behält als private-key (n, d).
5. Jeder, der an Alice eine verschlüsselte nachricht schicken will, braucht Zugang zu Alices public-key (n, e). Nur Alice darf ihren private-key (n, d) kennen!

Die verschlüsselte Nachricht wird wie folgt erstellt:

1. Bob stellt seine Nachricht (nach einem Alice bekannten System) als eine Zahl $x < n$ dar.
2. Bob konstruiert aus x mit dem public-key (von Alice) die verschlüsselte Nachricht y durch
$$y = x^e \bmod n.$$

Alice empfängt y und verwendet dann ihren private-key, um die Originalnachricht (oder zumindest deren Zahlencode) zu rekonstruieren, indem sie
$$y^d \bmod n$$
berechnet. Warum funktioniert das? Zu zeigen ist also, dass
$$y^d = x \pmod n.$$
Aus Bemerkung 5.24 folgt, dass $y^d = (x^e)^d \bmod n$. Wegen $ed = 1 \pmod m$, ist $ed = km + 1$ für ein $k \in \mathbb{N}_0$, d.h.
$$x^{ed} = x^{km+1} = x \cdot x^{k(p-1)(q-1)} = x \cdot \left(x^{p-1}\right)^{k(q-1)}.$$

Wir unterscheiden jetzt zwei Fälle:

- $\mathrm{ggT}(x, p) \neq 1$: dann muss x ein Vielfaches der Primzahl p sein, und es gilt $x = 0 \pmod p$ und damit $x \cdot \left(x^{p-1}\right)^{k(q-1)} = 0 \pmod p$. Damit gilt
$$x^{ed} = x \cdot \left(x^{p-1}\right)^{k(q-1)} = 0 = x \pmod p.$$
- $\mathrm{ggT}(x, p) = 1$: Dann gilt nach dem kleinen Satz von Fermat
$$x^{ed} = x \cdot \left(x^{p-1}\right)^{k(q-1)} = x \cdot 1^{k(q-1)} = x \pmod p$$

In beiden Fällen gilt also $x^{ed} = x \pmod p$. Analog kann man zeigen $x^{ed} = x \pmod q$. Aus diesen beiden Identitäten folgt, dass sowohl p als auch q Teiler von $x^{ed} - x$ sind. Da p, q Primzahlen sind, muss auch das Produkt pq ein Teiler von $x^{ed} - x$ sein. Nun ist aber $n = pq$ und $y^d = x^{ed}$, und n muss $y^d - x$ teilen, d.h.
$$y^d = x \pmod n.$$

BEISPIEL 5.26

Bob will das Wort „MATHEMATIK“ an Alice schicken. Dazu codiert er die einzelnen Buchstaben als Zahlen in $\mathbb{Z}_{26}$ und erhält das Zahlentupel $(12, 0, 19, 7, 4, 12, 0, 19, 8, 10)$.

Inzwischen konstruiert Alice ihre Schlüssel: sie wählt $p = 73$ und $q = 59$, damit sind $n = pq = 4307$ und $m = (p-1)(q-1) = 4176$. Für den public-key wählt sie $e = 107$ und berechnet die multiplikative Inverse in $\mathbb{Z}_m$, also $d = 2771$. Den public-key $(4307, 107)$

stellt sie Bob zur Verfügung. Bob rechnet

$$\begin{aligned} 12^{107} \bmod 4307 &= 29675177\ldots143178743808 \bmod 4307 = 3352 \\ 0^{107} \bmod 4307 &= 0 \\ 19^{107} \bmod 4307 &= 123 \\ &\vdots \\ 10^{107} \bmod 4307 &= 2679 \end{aligned}$$

und erhält insgesamt das verschlüsselte Tupel

$$(3352, 0, 123, 2680, 2318, 3352, 0, 123, 940, 2679).$$

BEISPIEL 5.27

Alice empfängt das verschlüsselte Tupel $(3352, 0, 123, 2680, 2318, 3352, 0, 123, 940, 2679)$ und rechnet

$$\begin{aligned} 3352^{2771} \bmod 4307 &= 12 \\ 0^{2771} \bmod 4307 &= 0 \\ 123^{2771} \bmod 4307 &= 19 \\ &\vdots \\ 2679^{2771} \bmod 4307 &= 10 \end{aligned}$$

und erhält so das entschlüsselte Tupel

$$(12, 0, 19, 7, 4, 12, 0, 19, 8, 10).$$

In ganz ähnlicher Weise kann auch eine *digitale Signatur* basierend auf RSA realisiert werden. Wir bezeichnen mit (n_S, e_S) und (n_S, d_S) die public-/private-keys der Senderin und mit (n_E, e_E) und (n_E, d_E) die public-/private-keys der Empfängerin. Die Senderin will die Nachricht „x_S“ versenden. Dazu *signiert* sie die Nachricht mit *ihrem eigenen private-key* und berechnet $y_S = x_S^{d_S} \bmod n_S$ und verschlüsselt dann wie gewohnt mit dem *public-key der Empfängerin* und versendet $y_S^{e_E} \bmod n_E$. Die Empfängerin erhält $y_E = y_S^{e_E} \bmod n_E$ und entschlüsselt in einem ersten Schritt wie gewohnt mit *ihrem eigenen private-key* zu

$$x_E = y_E^{d_E} \bmod n_E = y_S^{e_E d_E} \bmod n_E = y_S \pmod{n_E}.$$

Die Empfängerin kennt auch den *public-key der Senderin* und berechnet nun die Originalnachricht

$$y_S^{e_S} \bmod n_S = x_S^{d_S e_S} = x_S \pmod{n_S}.$$

Die Nachricht muss also von jemandem stammen, die den private-key der Senderin kennt, und das sollte im Normalfall nur die Senderin selbst sein. Wäre mit einem „falschen“ key d_σ signiert, so würde der letzte Schritt $x_S^{d_\sigma e_S} \neq x_S \pmod{n_S}$ ergeben. Man erhält nicht den Klartext und kann daraus schließen, dass die Signatur gefälscht ist.

Die Sicherheit eines jeden public-key Verschlüsselungsverfahren beruht darauf, dass aus der Kenntnis des public-key der private-key nicht (einfach) ermittelt werden kann. Im Falle des RSA-Verfahrens ist der private-key d ja $e^{-1} \pmod{(p-1)(q-1)}$, was nur mit Kenntnis von p und q berechnet werden kann. Das wiederum bedeutet, dass der public-key $n = pq$ in seine Primfaktoren p und q zerlegt werden müsste. Stand 2015 gilt gesichert, dass bei einer Schlüssellänge für n von 2048 Bit die Primfaktorenzerlegung von n auch unter Ausnützung aller auf der Welt verfügbarer Rechenkapazität nicht schaffbar ist. Wichtig ist auch, dass das Ver- und Entschlüsseln schnell geht, was für das Potenzieren modulo n jedenfalls gegeben ist. In der Praxis wählt man p und q annähernd gleich groß, da dann das Faktorisieren am schwierigsten ist. Würde man p extrem groß, aber $q = 2$ wählen, dann wäre das ungeschickt! In diesem Fall wäre n dann gerade, und jeder Angreifer würde sofort n faktorisieren können. Ebenso schützt die theoretische Sicherheit des Verfahrens natürlich nicht mehr, wenn ein Angreifer auf anderem Weg an den private-key herankommt, z.B. weil man den key auf einem Zettel in der Schreibtischlade aufgeschrieben hat.

KAPITEL 6

ALGEBREN

ALGEBRAISCHE STRUKTUREN

DEFINITION 6.1: BINÄRE OPERATION

Eine *binäre Operation* $\circ$ auf einer Menge A ist eine Funktion von $A \times A$ nach A.

Die Eigenschaft, dass das Bild der Operation auf A eine Teilmenge von A ist, nennt man *Abgeschlossenheit* ($\circ$ ist abgeschlossen auf A).

BEISPIEL 6.2

1. $+ \colon \mathbb{N}_0 \times \mathbb{N}_0, (a, b) \mapsto a + b$ ist eine Operation
2. $- \colon \mathbb{N}_0 \times \mathbb{N}_0, (a, b) \mapsto a - b$ ist *keine* Operation, da zum Beispiel $(2, 4)$ auf $-2 \notin \mathbb{N}_0$ abgebildet wird
3. Sei A eine Menge, dann ist $\cap \colon P(A) \times P(A) \to P(A)$ eine Operation: seien $B, C \in P(A)$, d.h., $B \subseteq A$ und $C \subseteq A$. Dann ist auch der Durchschnitt $B \cap C$ eine Teilmenge von A (eventuell die leere Menge) und damit $B \cap C \in P(A)$.
4. Sei $X = \{W, F\}$, dann ist $\wedge \colon X \times X \to X$ eine Operation.
5. $\mathrm{ggT} \colon \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{N}_0, (a, b) \mapsto \mathrm{ggT}(a, b)$ eine Operation auf $\mathbb{N}_0$.

Eine Algebra wird bestimmt durch

- *Objekte:* Elemente einer nichtleeren Menge (Trägermenge)
- *Grundoperationen:* im einfachsten Fall eine binäre (zweistellige) Operation
- *Gesetze:* beschreiben die Eigenschaften von Objekten, die Beziehungen zwischen Operationen, ...

DEFINITION 6.3: ALGEBRA

Sei A eine Menge und $\circ$ eine (binäre) Operation auf A. Dann nennt man das Paar $(A, \circ)$ eine *Algebra*.

BEISPIEL 6.4

Die Paare $(\mathbb{N}_0, +)$, $(P(X), \cap)$ (wenn X eine Menge ist) sind Algebren.

DEFINITION 6.5: KOMMUTATIV, ASSOZIATIV, NEUTRALES ELEMENT

Sei $\circ$ eine Operation auf einer Menge A. Dann nennt man $\circ$

- *kommutativ* wenn $\forall\, a, b \in A\colon a \circ b = b \circ a$.
- *assoziativ* wenn $\forall\, a, b, c \in A\colon (a \circ b) \circ c = a \circ (b \circ c)$.

Die Operation besitzt ein *neutrales Element*, wenn ein $e \in A$ existiert, sodass für alle $a \in A$ gilt

$$a \circ e = e \circ a = a.$$

Algebren können in verschiedene Klassen eingeteilt werden, je nachdem welche Eigenschaften sie besitzen.

DEFINITION 6.6: HALBGRUPPE, MONOID

Eine Algebra $(A, \circ)$ heißt *Halbgruppe*, falls die Operation $\circ$ *assoziativ* ist. Eine Halbgruppe $(A, \circ)$ heißt ein *Monoid*, falls sie ein *neutrales Element* besitzt. Wenn $\circ$ außerdem kommutativ ist, spricht man von einer kommutativen Halbgruppe (einem kommutativen Monoid).

BEISPIEL 6.7

- $(\mathbb{N}_0, +)$ ist eine Halbgruppe, da für alle $a, b, c \in \mathbb{N}_0$ gilt $(a + b) + c = a + (b + c)$. $(\mathbb{N}_0, +)$ ist außerdem ein Monoid, mit $e = 0$ als neutralem Element.
- $(\mathbb{N}, +)$ ist eine Halbgruppe, aber kein Monoid.
- Sei A eine Menge. Dann ist $(P(A), \cup)$ eine Halbgruppe, da Vereinigung assoziativ ist (siehe Satz 2.20(2)). $(P(A), \cup)$ ist außerdem ein Monoid mit neutralem Element $\emptyset$: für jede Teilmenge B von A (d.h. für jedes Element $B \in P(A)$) gilt $B \cup \emptyset = B$.
- Sei A eine Menge. Dann ist $(P(A), \cap)$ eine Halbgruppe, da Durchschnittsbildung assoziativ ist (siehe Satz 2.20(2)). $(P(A), \cap)$ ist außerdem ein Monoid mit neutralem Element A: für jede Teilmenge B von A (d.h. für jedes Element $B \in P(A)$) gilt $B \cap A = B$.

- $(\mathbb{Q}, \cdot)$ ist ein Monoid (mit neutralem Element 1).
- $(\mathbb{Z}, -)$ ist eine Algebra, da die ganzen Zahlen unter Differenzbildung abgeschlossen sind, aber es ist *keine* Halbgruppe, weil die Operation nicht assoziativ ist:

$$(3-4)-7=-8 \quad \text{aber} \quad 3-(4-7)=6.$$

- $(\mathbb{N}_0, \mathrm{ggT})$ ist eine Halbgruppe, da der größte gemeinsame Teiler assoziativ ist, also für $a, b, c \in \mathbb{N}_0$ gilt

$$\mathrm{ggT}\big(\mathrm{ggT}(a,b),c\big) = \mathrm{ggT}\big(a,\mathrm{ggT}(b,c)\big).$$

  Auf beiden Seiten wird hier die größte natürliche Zahl bestimmt, die $a, b,$ und c teilt. $(\mathbb{N}_0, \mathrm{ggT})$ ist außerdem ein Monoid mit neutralem Element 0, da $\mathrm{ggT}(a,0) = \mathrm{ggT}(0,a) = a$ für alle $a \in \mathbb{N}_0$.

DEFINITION 6.8: INVERTIERBAR

Sei $(A, \circ)$ ein Monoid mit neutralem Element e. Dann heißt $a \in A$ *invertierbar*, wenn

$$\exists\, b \in A\colon a \circ b = b \circ a = e.$$

In dem Fall heißt b das *zu a inverse Element*.

DEFINITION 6.9: GRUPPE

Ein Monoid $(G, \circ)$ heißt *Gruppe*, wenn jedes $g \in G$ in G invertierbar ist. Wenn die Operation $\circ$ außerdem *kommutativ* ist, dann nennt man $(G, \circ)$ eine *Abelsche Gruppe* (oder kommutative Gruppe).

Das neutrale Element einer Gruppe wird auch *Einselement* genannt. Bezeichnet man das Operationssymbol in einer kommutativen Gruppe mit $+$, so nennt man die Algebra oft auch *Modul*. Das neutrale Element nennt man dann *Nullelement*.

In einer Gruppe $(G, \cdot)$ bezeichnet man das inverse Element zu $g \in G$ mit g^{-1}, in einem Modul $(G, +)$ mit $-g$.

BEISPIEL 6.10

- $(\mathbb{Q}^*, \cdot)$ ist eine Abelsche Gruppe.
- $(\mathbb{N}_0, \cdot)$ ist ein Monoid (mit neutralem Element 1), aber keine Gruppe (nur 1 ist invertierbar).
- $(\mathbb{Z}_n, +)$ ist eine kommutative Gruppe (ein kommutatives Modul) für jedes $n \in \mathbb{N}$.
- $(\mathbb{Z}_p^*, \cdot)$ ist nur dann eine Abelsche Gruppe, wenn p eine Primzahl ist, da nur dann

jedes Element in $\mathbb{Z}_p$ eine multiplikative Inverse besitzt.

Die Gruppe der *Permutationen*, die wir als nächstes betrachten, war historisch die erste Gruppe, die untersucht wurde.

DEFINITION 6.11: PERMUTATION

Sei X eine nichtleere Menge. Eine bijektive Funktion $\pi\colon X \to X$ wird eine *Permutation* von X genannt. Die Menge aller Permutationen der Menge X bezeichnen wir mit $\mathrm{Sym}(X)$.

BEISPIEL 6.12

- Die Funktion $p\colon \mathbb{Z}_{26} \to \mathbb{Z}_{26}, x \mapsto x + 5 \pmod{26}$ ist eine Permutation von $\mathbb{Z}_{26}$.
- Sei $X = [0, 4]$, dann sind zum Beispiel

$$\pi_1\colon X \to X, x \mapsto 4 - x \qquad\qquad \pi_2\colon X \to X, x \mapsto \frac{x^3}{16}$$

  Permutationen der Menge X, also $\pi_1, \pi_2 \in \mathrm{Sym}(X)$.

Eine Permutation ist also eine Umordnung der Elemente einer gegebenen Menge. Wir betrachten die Algebra $(\mathrm{Sym}(X), \circ)$, wobei $\circ$ nun die Komposition von Funktionen bezeichnet.

LEMMA 6.13

Sei $X \neq \emptyset$. Dann ist $(\mathrm{Sym}(X), \circ)$ eine Gruppe.

Man nennt $(\mathrm{Sym}(X), \circ)$ die *symmetrische Gruppe* auf X.

Beweis.

- $(\mathrm{Sym}(X), \circ)$ ist eine Algebra, d.h., die symmetrische Gruppe ist abgeschlossen unter Hintereinanderausführung: Dazu ist zu zeigen, dass die Komposition bijektiver Funktionen wieder bijektiv ist. Seien $\pi, \sigma \in \mathrm{Sym}(X)$, dann gilt, dass $\pi(X) = X$ und $\sigma(X) = X$ und damit $\pi \circ \sigma(X) = X$, also ist $\pi \circ \sigma$ surjektiv. Sei jetzt $z \in X$, dann existiert ein eindeutig bestimmtes Urblid $\pi^{-1}(z) = y$, da π bijektiv ist, und, da σ bijektiv, ein eindeutig bestimmtes Urbild $\sigma^{-1}(y) = x$. Also ist die Kompostion injektiv und somit bijektiv.
- $(\mathrm{Sym}(X), \circ)$ ist eine Halbgruppe, d.h., die Hintereinanderausführung ist assoziativ auf $\mathrm{Sym}(X)$: Seien $\pi_1, \pi_2, \pi_3 \in \mathrm{Sym}(X)$ und $x \in X$, dann gilt

$$((\pi_1 \circ \pi_2) \circ \pi_3)(x) = (\pi_1 \circ \pi_2)(\pi_3(x)) = \pi_1(\pi_2(\pi_3(x))),$$

  und

$$(\pi_1 \circ (\pi_2 \circ \pi_3))(x) = \pi_1(\pi_2 \circ \pi_3(x)) = \pi_1(\pi_2(\pi_3(x))).$$

- $(\mathrm{Sym}(X), \circ)$ ist ein Monoid, d.h., es existiert ein neutrales Element in $\mathrm{Sym}(X)$: Das neutrale Element ist durch $id_X \in \mathrm{Sym}(X)$ gegeben.
- $(\mathrm{Sym}(X), \circ)$ ist eine Gruppe, d.h., jedes Element in $\mathrm{Sym}(X)$ ist bezüglich $\circ$ invertierbar: jede bijektive Funktion ist invertierbar und die Inverse ist wieder bijektiv, also ein Element der symmetrischen Gruppe.

□

Wenn wir die Permutation einer endlichen Menge X mit $|X| = n$ betrachten, kann man sich ohne Beschränkung der Allgemeinheit auf die Mengen $\{1, 2, 3, \ldots, n\}$ beschränken. Wenn eine Menge $X = \{x_1, x_2, x_3, \ldots, x_n\}$ gegeben ist, dann entspricht die Anwendung einer Permutation $\sigma \in \mathrm{Sym}(X)$ mit $\sigma(x_i) = x_j$ einer Permutation der Indizes durch eine Permutation $\pi \in \mathrm{Sym}(\{1, 2, 3, \ldots, n\})$ mit $\pi(i) = j$. Wenn die gegebene Menge $X = \{1, 2, 3, \ldots, n\}$ ist, dann schreiben wir für die symmetrische Gruppe $\mathrm{Sym}(X)$ einfach S_n. Eine übliche Schreibweise für Permutationen in S_n ist wie folgt

$$\pi = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ \pi(1) & \pi(2) & \pi(3) & \ldots & \pi(n) \end{pmatrix},$$

d.h., in der zweiten Reihe stehen die Bilder unter π von Elementen aus der ersten Reihe. Mitunter wird auf die erste Reihe verzichtet und nur die zweite Reihe angegeben, also die folgende Schreibweise verwendet:

$$\pi = \big(\pi(1)\ \pi(2)\ \pi(3)\ \ldots\ \pi(n)\big).$$

BEISPIEL 6.14

Wir betrachten die folgenden Element von $\pi, \sigma \in S_6$,

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 2 & 5 & 3 & 4 \end{pmatrix}, \quad \text{und} \quad \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 4 & 3 & 2 & 5 \end{pmatrix}.$$

Dann gilt

$$\pi \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 6 & 5 & 2 & 1 & 3 \end{pmatrix}, \quad \text{und} \quad \sigma \circ \pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 1 & 2 & 4 & 3 \end{pmatrix}.$$

An diesem Beispiel sieht man, dass die symmetrische Gruppe *keine* Abelsche Gruppe ist (die Verknüpfung ist im allgemeinen nicht kommutativ). Außerdem gilt

$$\pi^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 5 & 6 & 4 & 1 \end{pmatrix}$$

BEISPIEL 6.15

Die Elemente der symmetrischen Gruppe S_4 sind gegeben durch (wobei wir die letztge-

nannte Schreibweise verwenden):

$$(1\ 2\ 3\ 4), (1\ 2\ 4\ 3), (1\ 3\ 2\ 4), (1\ 3\ 4\ 2), (1\ 4\ 2\ 3), (1\ 4\ 3\ 2),$$
$$(2\ 1\ 3\ 4), (2\ 1\ 4\ 3), (2\ 3\ 1\ 4), (2\ 3\ 4\ 1), (2\ 4\ 1\ 3), (2\ 4\ 3\ 1),$$
$$(3\ 1\ 2\ 4), (3\ 1\ 4\ 2), (3\ 2\ 1\ 4), (3\ 2\ 4\ 1), (3\ 4\ 1\ 2), (3\ 4\ 2\ 1),$$
$$(4\ 1\ 2\ 3), (4\ 1\ 3\ 2), (4\ 2\ 1\ 3), (4\ 2\ 3\ 1), (4\ 3\ 1\ 2), (4\ 3\ 2\ 1).$$

Für die Anzahl der Element der symmetrischen Gruppe gilt also $|S_4| = 24$.

LEMMA 6.16

Es gilt $|\operatorname{Sym}(X)| = |X|!$ *und als Spezialfall* $|S_n| = n!$.

Beweis. Wenn wir die Arten betrachten, wie die zweite Reihe einer Permutation

$$\pi = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \pi(1) & \pi(2) & \pi(3) & \dots & \pi(n) \end{pmatrix}$$

gebildet werden kann, dann gibt es n mögliche Werte für $\pi(1)$, aber nur $n-1$ mögliche Werte für $\pi(2)$, da $\pi(1)$ nicht wieder verwendet werden darf. Für $\pi(3)$ gibt es dann nur mehr $n-2$ mögliche Werte, usw. Damit gibt es insgesamt $n(n-1)(n-2)\cdots 2 \cdot 1 = n!$ verschiedene Permutationen der Menge $X = \{1, 2, \dots, n\}$. □

BEISPIEL 6.17: VERKNÜPFUNGSTABELLE FÜR S_3

$\circ$	(1 2 3)	(1 3 2)	(2 1 3)	(2 3 1)	(3 1 2)	(3 2 1)
(1 2 3)	(1 2 3)	(1 3 2)	(2 1 3)	(2 3 1)	(3 1 2)	(3 2 1)
(1 3 2)	(1 3 2)	(1 2 3)	(3 1 2)	(3 2 1)	(2 1 3)	(2 3 1)
(2 1 3)	(2 1 3)	(2 3 1)	(1 2 3)	(1 3 2)	(3 2 1)	(3 1 2)
(2 3 1)	(2 3 1)	(2 1 3)	(3 2 1)	(3 1 2)	(1 2 3)	(1 3 2)
(3 1 2)	(3 1 2)	(3 2 1)	(1 3 2)	(1 2 3)	(2 3 1)	(2 1 3)
(3 2 1)	(3 2 1)	(3 1 2)	(2 3 1)	(2 1 3)	(1 3 2)	(1 2 3)

ABBILDUNGEN ZWISCHEN ALGEBRAISCHEN STRUKTUREN

DEFINITION 6.18: HOMOMORPHISMUS

Seien $(A_1, \circ_1)$ und $(A_2, \circ_2)$ zwei algebraische Strukturen. Dann nennt man ein Abbildung $\varphi\colon A_1 \to A_2$ einen *Homomorphismus* zwischen A_1 und A_2, wenn für alle $x, y \in A_1$ gilt:

$$\varphi(x \circ_1 y) = \varphi(x) \circ_2 \varphi(y).$$

Sind $(A_1, \circ_1)$ und $(A_2, \circ_2)$ Gruppen, dann nennt man φ einen *Gruppenhomomorphismus*.

Ein Homomorphismus ist *strukturerhaltende* Abbildung.

BEISPIEL 6.19

Die Abbildung $\varphi\colon (\mathbb{Z},+) \to (\mathbb{Z}_5,\oplus_5)$ mit $x \mapsto x \bmod 5$ ist ein Homomorphismus. Seien $x, y \in \mathbb{Z}$, dann ist zu zeigen, dass

$$\varphi(x+y) = \varphi(x) \oplus_5 \varphi(y) \quad \text{d.h.} \quad (x+y) \bmod 5 = (x \bmod 5) \oplus_5 (y \bmod 5).$$

Seien $a, b \in \mathbb{Z}_5$ so, dass $x \bmod 5 = a$ und $y \bmod 5 = b$, d.h., es gibt $p, q \in \mathbb{Z}$, sodass $x = 5p + a$ und $y = 5q + b$. Dann gilt

$$a + b = 5s + (a \oplus_5 b) \text{ und somit}$$

$$x + y = 5p + a + 5q + b = 5(p+q) + (a+b) = 5(p+q+s) + (a \oplus_5 b).$$

Das bedeutet nun aber genau $(x+y) \bmod 5 = (x \bmod 5) \oplus_5 (y \bmod 5)$.

Durch einen Gruppenhomomorphismus φ zwischen $(G_1, \circ_1)$ und $(G_2, \circ_2)$ wird das Einselement von G_1 auf das Einselement von G_2 abgebildet: sei $g \in G_1$, und sei e_1 das Einselement von G_1 und e_2 das Einselement von G_2, dann gilt

$$\varphi(g) = \varphi(g \circ_1 e_1) = \varphi(g) \circ_2 \varphi(e_1),$$

also muss $\varphi(e_1) = e_2$ gelten. Außerdem werden die Inversen von Elementen aus G_1 auf die Inversen der Abbildung unter G_2 abgebildet, d.h., für $g \in G_1$ gilt

$$\begin{aligned}\varphi(g^{-1}) &= \varphi(g^{-1}) \circ_2 \varphi(g) \circ_2 \varphi(g)^{-1} = \varphi(g^{-1} \circ_1 g) \circ_2 \varphi(g)^{-1} = \varphi(e_1) \circ_2 \varphi(g)^{-1} = \\ &= e_2 \circ_2 \varphi(g)^{-1} = \varphi(g)^{-1}.\end{aligned}$$

BEISPIEL 6.20

Für den Homomorphismus aus Beispiel 6.19 gilt: $\varphi(0) = 0 \bmod 5 = 0$. Sei $x \in \mathbb{Z}$ so, dass $x = 5q + a$ (also $a = \varphi(x) = x \bmod 5 \in \mathbb{Z}_5$), dann gilt

$$\varphi(-x) = \varphi(-5q - a) = -5q - a \bmod 5 = -a \bmod 5 = 5 - a = \ominus_5 \varphi(x),$$

wobei $\ominus_5$ anzeigen soll, dass wir die additive Inverse in $\mathbb{Z}_5$ betrachten.

DEFINITION 6.21: ISOMORPHISMUS

Ein Homomorphismus φ zwischen zwei algebraischen Strukturen $(A_1, \circ_1)$ und $(A_2, \circ_2)$ heißt ein *Isomorphismus*, wenn die Abbildung außerdem *bijektiv* ist. Wenn $(A_1, \circ_1)$ und $(A_2, \circ_2)$ beide Gruppen sind, dann nennt man φ einen *Gruppenisomorphismus*. Wenn es einen Isomorphismus zwischen $(A_1, \circ_1)$ und $(A_2, \circ_2)$ gibt, dann nennt man die beiden Algebren *isomorph* und schreibt $A_1 \simeq A_2$.

Isomorph bedeutet

- gleich bis auf die Bezeichnung der Objekte,
- nicht unterscheidbar aus der Sicht der Algebra,
- die Verknüpfungstabellen sind identisch (bis auf Umbenennung der Objekte).

BEISPIEL 6.22

Die Abbildung $\varphi\colon (\mathbb{R},+) \to (\mathbb{R}^+,\cdot), x \mapsto e^x$ ist ein Gruppenisomorphismus: $(\mathbb{R},+)$ ist eine Abelsche Gruppe mit neutralem Element 0, $(\mathbb{R},\cdot)$ ist eine Abelsche Gruppe mit neutralem Element 1. Es gilt für $x, y \in \mathbb{R}$, dass

$$\varphi(x+y) = e^{x+y} = e^x e^y = \varphi(x)\cdot\varphi(y), \quad \text{und} \quad \varphi(0) = e^0 = 1.$$

Die Funktion ist außerdem bijektiv ($\varphi(\mathbb{R}) = \mathbb{R}^+$ und die Funktion streng monoton steigend).

BEISPIEL 6.23

Sei $X = \{x_1, x_2, \ldots, x_n\}$ eine n-elementige Menge, und sei $\varphi\colon (\mathrm{Sym}(X),\circ) \to (S_n,\circ)$ die Funktion, die einer Permutation $\pi \in \mathrm{Sym}(X)$ die Permutation $\sigma \in S_n$ so zuordnet, dass

$$\pi(x_i) = x_j \quad \Leftrightarrow \quad \sigma(i) = j, \quad \text{d.h.} \quad \pi(x_i) = x_{\sigma(i)}.$$

Dann ist φ ein Gruppenisomorphismus. Seien $\pi_1, \pi_2 \in \mathrm{Sym}(X)$ mit Bildern $\varphi(\pi_k) = \sigma_k$ ($k = 1, 2$), dann gilt

$$(\pi_1 \circ \pi_2)(x_i) = \pi_1(\pi_2(x_i)) = \pi_1(x_{\sigma_2(i)}) = x_{\sigma_1(\sigma_2(i))} = x_{(\sigma_1\circ\sigma_2)(i)},$$

also

$$\varphi(\pi_1 \circ \pi_2) = \varphi(\pi_1) \circ \varphi(\pi_2).$$