Karlsruher Institut für Technologie

Vorkurs Mathematik

Skriptum

Institut für Analysis
Abteilung für Didaktik

Version vom 1. Oktober 2025

Inhaltsverzeichnis

Vorwort **3**

Einleitung **4**
Grundbegriffe 6

1 Aussagenlogik **8**
1.1 Aussagen und Quantoren 8
1.2 Logische Verknüpfungen 13
1.3 Tautologien und Beweisstrategien 20

2 Mengenlehre **29**
2.1 Mengen 29
2.2 Verknüpfungen von Mengen 38

3 Ungleichungen **49**
3.1 Allgemeingültige Ungleichungen 49
3.2 Lösungsstrategien für (Un-)Gleichungen 60

4 Abbildungen **74**
4.1 Abbildungen als Relationen 74
4.2 Eigenschaften von Abbildungen 89

Anhang **107**
Polynomdivision 107
Partialbruchzerlegung 109

Index **111**

Vorwort

Dieser Vorkurs wiederholt Inhalte aus der Schulmathematik in der Sprache der Universitätsmathematik. Dadurch sollen unterschiedliche Lernvoraussetzungen angeglichen und eine Basis für das Mathematikstudium geschaffen werden. Gleichzeitig werden Inhalte und Methoden vorbereitet und erläutert, die in der Anfangsphase des Studiums häufig Schwierigkeiten bereiten. All dies hat zum Ziel, den Einstieg in das Studium zu erleichtern und so die Freude am mathematischen Arbeiten zu erhalten und zu fördern.

Der Vorkurs beginnt mit den bekannten Zahlenbereichen und Rechenregeln. In Kapitel 1 wird mit Aussagenlogik und Beweisstrategien die Grundlage für das mathematische Arbeiten und Formulieren gelegt. Kapitel 2 behandelt Schreibweisen für Mengen und ihre Verknüpfungen. In Kapitel 3 werden sowohl der Umgang mit Ungleichungen als auch Lösungsstrategien für einige Arten von Gleichungen und Ungleichungen diskutiert. Kapitel 4 führt das wichtige Konzept *Abbildung* anhand von Relationen ein und erläutert dann einige ihrer Verknüpfungen und Eigenschaften.

Der Vorkurs ist als Blockveranstaltung an fünf Terminen konzipiert, die jeweils aus einer Vorlesung und einem Tutorium bestehen. Zur Vorbereitung und für die Tutorien werden Übungsaufgaben mit Lösungen angeboten. Mathematik lernt man, indem man sie betreibt – daher sei den Teilnehmern hier und in ihrem Studium empfohlen, in der Vorlesung mitzudenken, aktiv im Skriptum zu lesen, die Übungsaufgaben selbstständig zu bearbeiten, das Gelernte aufzuschreiben, zu durchdenken, mit anderen zu diskutieren und Fragen zu stellen.

Dieses Skriptum basiert auf einer früheren Version, die ich für den Kurs math4mint am KIT erstellt habe. Dabei flossen einige Ideen aus dem Skriptum des Instituts für Analysis ein, das ursprünglich für diesen Kurs konzipiert wurde. Gedankt sei in chronologischer Reihenfolge den Autoren Johanna Dettweiler (2009), Alexander Ullmann, Andreas Bolleyer, Andreas Hirsch (2016), Michael Ullmann (2019), Sebastian Ohrem (2020), Christopher Bresch (2021) sowie allen, die Anmerkungen und Korrekturen beigetragen haben.

Karlsruhe, den 01. Oktober 2025 Wolf Wechinger

Einleitung

Die Mathematik ist ein Denkgebäude, eine **abstrakte** Welt, die nicht den Regeln der Natur folgt, sondern den Regeln, die wir ihr geben (wenngleich natürlich ihre Begriffe von der Natur inspiriert sind und ihre Regeln gelegentlich die Natur nachbilden, also *modellieren* sollen). Im Gegensatz zum Alltag und zu vielen anderen Wissenschaften ist die Mathematik streng **deduktiv**, erlaubt also nur den logischen Schluss von einer allgemeingültigen Regel auf einen Spezialfall; nicht erlaubt ist hingegen, aus einer Betrachtung von Beispielen ein allgemeines Prinzip abzuleiten (*induktives* Schließen) oder mithilfe von Beobachtungen in der Natur zu argumentieren (*Empirie*).
Wenn aber neues Wissen nur aus bereits Bekanntem gefolgert werden darf, wo fängt man dann an? Der Startpunkt jeder mathematischen Argumentation ist eine Reihe von Aussagen, die wir als wahr annehmen, *ohne dies zu begründen*, den sogenannten **Axiomen**; diese Axiome können sich aus unserer Erfahrung mit der Natur ergeben, aber auch ganz willkürlich festgelegt sein. Aus den Axiomen werden mithilfe bestimmter Schlussregeln *rein deduktiv* neue Aussagen hergeleitet; den Nachweis über die Wahrheit einer behaupteten Aussage nennt man einen **Beweis**. Der Beweis macht aus der Behauptung einen gültigen **Satz**; ab dann kann dieser Satz zum Beweis neuer Behauptungen herangezogen werden. Um sich in all den logischen Beziehungen zurechtzufinden und darüber besser sprechen zu können, versieht man bestimmte Eigenschaften und Objekte mit Namen, legt also **Definitionen** fest. So entsteht ein immer größeres und komplexeres Denkgebäude, das aber ein einfaches Fundament aus einigen Axiomen und logischen Schlussregeln besitzt. Dieses Vorgehen ist charakteristisch für die moderne Mathematik und hat viele Vorteile, denn für einen Mathematiker steckt Erkenntnis nicht nur in neuen Sätzen, sondern auch im logischen Weg von einem Startpunkt bis dorthin.

Unter anderem den Bemühungen von David Hilbert ist es zu verdanken, dass die Mathematik inzwischen (fast) ein einziges riesiges Denkgebäude ist, das auf wenigen *gemeinsamen Axiomen* beruht. Das Ziel des Mathematikstudiums in den ersten drei Semestern ist es, dieses Gebäude „von unten" schrittweise kennenzulernen und einige höhere Etagen zu erkunden, also bei gewissen Axiomen zu beginnen und sich in die wichtigsten Teilbereiche der Mathematik einzuarbeiten: (Lineare) Algebra, Geometrie, Analysis, Numerik und Stochastik. Man beginnt diesen Weg nicht „beim Fundament", also den Grundlagen der Mathematik, da diese schwierig und von der Schulmathematik sehr verschieden sind; sie werden im späteren Studium behandelt. Stattdessen betreten wir das Mathematikgebäude „im Erdgeschoss": Wir starten mit einfachen Begriffen und Zusammenhängen, die aus der Schule bekannt sind, und arbeiten uns mit diesen nach der deduktiven Methode „aufwärts“. Einige der grundlegenden Begriffe werden hier durch Axiome oder Definitionen genauer gefasst; bei anderen findet diese *Präzisierung* „abwärts“ erst im frühen oder sogar späten Teil des Studiums statt. Aus diesem Grund sind die einzelnen Kapitel gerade am Anfang häufig unpräzise und arbeiten erst darauf aufbauend exakt.

Auch wenn dieses bewährte Vorgehen der Verständlichkeit dienen soll, so sorgt es doch gelegentlich dafür, dass man im Denkgebäude der Mathematik die Orientierung verliert, also geltende Sätze, Voraussetzungen und beweisbedürftige Behauptungen durcheinander bringt. Dies ist eine der Herausforderungen, mit denen alle Anfänger im Mathematikstudium zu kämpfen haben und von denen Sie sich nicht verunsichern lassen sollten:

- Mathematik ist schwierig: Sie handelt von *abstrakten* Objekten und Begriffen, sie sucht nach Strukturen und *allgemeinen* Konzepten, dadurch entzieht sie sich oft unserer Anschauung. Gleichzeitig ist Mathematik *präzise* und erwartet ein streng logisches, deduktives Vorgehen. Sie werden viel Mühe, Durchhaltevermögen und Konzentration dabei aufwenden, sich in und durch die Mathematik zu arbeiten.

- Die Universitätsmathematik behandelt nicht nur andere Objekte und Begriffe als die Schulmathematik, sie verfolgt auch eine andere Methode: Der Fokus liegt weniger auf Ergebnissen und Verfahren und dafür mehr auf Begründungen und Verständnis. Manche Dinge werden Sie neu und ganz anders lernen, manche werden Sie vergessen; aber auf jeden Fall werden Sie wissen, *warum* etwas funktioniert und *warum* Sie etwas tun.

- In den Mathematikvorlesungen und -übungen werden Sie in hohem Tempo mit einer großen Fülle an komplexen Inhalten konfrontiert. Sie werden viel Zeit nur dafür brauchen, überhaupt mitzukommen und zu verstehen, worum es geht. Bleiben Sie dran! Viele Dinge werden Sie erst bemerken oder verstehen, wenn sie Ihnen zum fünften Mal in Ihrem Studium begegnen – das ist für unser Lernen völlig normal.

- Die neue Arbeitsweise und Schwierigkeit im Mathematikstudium werden Ihnen auch eine persönliche Herausforderung sein. Mathematikerinnen und Mathematiker werden geschätzt für ihre Gelassenheit und ihr flexibles Denken, für ihre Beharrlichkeit und Frustrationstoleranz, ihren Struktursinn und ihre Korrektheit – und diese Eigenschaften werden auch Sie sich hart erarbeiten (müssen).

Lassen Sie sich von dieser Aufzählung nicht entmutigen! Schon bald werden Sie sich an das Mathematiktreiben an der Universität gewöhnt und Freude daran haben, verschiedenste Situationen zu durchschauen, zu strukturieren und darin Probleme zu lösen. Bleiben Sie neugierig, gönnen Sie sich Erholung und Ausgleich zu Ihrem Studium und suchen Sie sich Freunde und Gleichgesinnte, mit denen Sie auch (aber nicht nur) über Mathematik sprechen und zusammen studieren können. Viel Spaß und Erfolg bei Ihrem Mathematikstudium am KIT!

Grundbegriffe

Von den folgenden Grundlagen sollte das meiste aus der Schule bekannt sein, daher werden sie nur stichpunktartig aufgelistet.

(1) Die Zahlenbereiche $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ und $\mathbb{R}$. Eine Schreibweise wie $x \in \mathbb{R}$ bedeutet, dass x eine reelle Zahl ist, siehe Definition 2.1.

- Die **natürlichen Zahlen** $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$. Beachte: 0 ist keine natürliche Zahl und wir schreiben $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$.[1] Es gibt eine kleinste natürliche Zahl 1 und jede Zahl $n \in \mathbb{N}$ hat einen Nachfolger $n + 1 \in \mathbb{N}$, es gibt also keine größte natürliche Zahl. $\mathbb{N}$ ist bezüglich $+$ und $\cdot$ *abgeschlossen*, d.h. die Summe und das Produkt zweier natürlicher Zahlen liegt wieder in $\mathbb{N}$.

- Die **ganzen Zahlen** $\mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}$. $\mathbb{Z}$ besteht aus den natürlichen Zahlen, Null und den negativen ganzen Zahlen $\{-n : n \in \mathbb{N}\}$. Es gibt keine kleinste oder größte ganze Zahl. $\mathbb{Z}$ ist bezüglich $+, -$ und $\cdot$ abgeschlossen.

- Die **rationalen Zahlen** $\mathbb{Q} = \{\frac{z}{n} : z \in \mathbb{Z}, n \in \mathbb{N}\}$. $\mathbb{Q}$ besteht aus allen (positiven und negativen) Brüchen, also abbrechenden oder periodischen Dezimalzahlen. Zwischen zwei rationalen Zahlen liegt stets eine weitere rationale Zahl. $\mathbb{Q}$ ist bezüglich aller Grundrechenarten abgeschlossen.[2]

- Die **reellen Zahlen** $\mathbb{R}$. $\mathbb{R}$ enthält auch irrationale Zahlen, also Dezimalzahlen, die weder abbrechend noch periodisch sind, zum Beispiel $\sqrt{2}, \pi, e$ oder $\log 2$. Auch $\mathbb{R}$ ist bezüglich aller Grundrechenarten abgeschlossen, hat aber keine „Lücken“ mehr.[3] $\mathbb{R}$ wird in Analysis 1 präzise definiert.

Gebräuchlich sind außerdem die **komplexen Zahlen** $\mathbb{C}$ und die **Primzahlen** $\mathbb{P}$.

(2) Die Regeln des Zahlenrechnens und der Algebra:

- Die Regeln der Grundrechenarten, insbesondere Assoziativ-, Kommutativ- und Distributivgesetz, die binomischen Formeln und Vorzeichenregeln.
- Die Regeln der Bruchrechnung.
- Die Potenzgesetze, insbesondere Wurzeln.
- Die Logarithmengesetze.

(3) Der Umgang mit Termen und Gleichungen:

- Das Einsetzen und zielgerichtete Umformen von Termen. Im Anhang werden Polynomdivision und Partialbruchzerlegung kurz erklärt.
- Das Lösen von (Un-)Gleichungen, insbesondere Äquivalenzumformungen und die Lösungsformel für quadratische Gleichungen; siehe dazu Kapitel 3.

[1] Das ist die übliche Konvention am KIT, manche Mathematiker schreiben stattdessen $\mathbb{N} = \{0, 1, 2, \ldots\}$ und $\mathbb{N}^* = \{1, 2, 3, \ldots\}$.

[2] Natürlich ist die Division durch 0 nicht erlaubt.

[3] Zum Beispiel hat die Gleichung $x^2 = 2$ zwei reelle Lösungen, aber keine rationale Lösung.

Potenzen und Wurzeln

Zu einer **Basis** $a \in \mathbb{R}$ und einem **Exponenten** $n \in \mathbb{N}$ definiert[4] man die **Potenz** $a^n := a \cdot \ldots \cdot a$ mit n Faktoren sowie $a^0 := 1$. Für $a \neq 0$ definiert man außerdem $a^{-n} := \frac{1}{a^n}$. (*ganzzahlige Potenz*)

Für $a \geqslant 0$ hat die Gleichung $x^n = a$ genau eine nichtnegative Lösung $x \geqslant 0$, diese nennt man $\sqrt[n]{a} = a^{\frac{1}{n}}$. Damit ist auch $a^{\frac{z}{n}} = (a^{\frac{1}{n}})^z$ für alle $a > 0$ und $\frac{z}{n} \in \mathbb{Q}$ definiert.[5](*rationale Potenz*)

Die *reelle Potenz* a^x für $a > 0$ und $x \in \mathbb{R}$ wird in Analysis 1 definiert; dort werden auch die Details zu diesen Definitionen geklärt.
Für alle Potenzen gelten die **Potenzgesetze**

$$a^r \cdot a^s = a^{r+s}, \qquad a^r \cdot b^r = (a \cdot b)^r, \qquad (a^r)^s = a^{rs}.$$

Einige Beispiele:[6]

$$64^{\frac{1}{2}} \cdot 64^{-\frac{2}{3}} = 64^{\frac{1}{2}-\frac{2}{3}} = 64^{-\frac{1}{6}} = \frac{1}{\sqrt[6]{64}} = \frac{1}{2}, \qquad \frac{2^{3^2}}{512} = \frac{2^9}{2^9} = 2^{9-9} = 1,$$

$$\sqrt{6} \cdot \sqrt{10} \cdot \sqrt{15} = \sqrt{6 \cdot 10 \cdot 15} = \sqrt{900} = 30, \qquad \sqrt[6]{125} = \sqrt{\sqrt[3]{125}} = \sqrt{5}.$$

Logarithmus

Zu einer **Basis** $b > 0, b \neq 1$ und einem **Numerus** $a > 0$ hat die Gleichung $b^x = a$ genau eine Lösung $x > 0$, diese nennt man den **Logarithmus** $\log_b(a)$. Der Logarithmus ist die Umkehrfunktion der Exponentialfunktion, d.h.

$$\log_b(b^x) = x, \qquad b^{\log_b(x)} = x.$$

Aus den Potenzgesetzen folgen sofort die **Logarithmengesetze**

$$\log_b(x \cdot y) = \log_b(x) + \log_b(y), \qquad \log_b(x^y) = y \cdot \log_b(x)$$

sowie die Umrechnungsformeln $\log_a(x) = \log_a(b) \cdot \log_b(x)$ und $a^x = b^{x \cdot \log_b(a)}$. Daher genügt ein einziger Logarithmus; man wählt den **natürlichen Logarithmus** zur Basis der **Eulerschen Zahl** $e \approx 2{,}718$, bezeichnet mit log oder ln. Einige Beispiele:

$$\log_b(1) = 0, \ \log_b(b) = 1, \ \log_2\left(\frac{1}{16}\right) = -4, \ \log\left(\sqrt{e}\right) = \frac{1}{2},$$

$$\log(18) + \log(81) = \log(2 \cdot 3^2) + \log(3^4) = \log(2) + 6\log(3).$$

[4]Eine Schreibweise wie $A := B$ bedeutet, dass das Objekt A durch das Objekt B definiert wird.

[5]Prinzipiell lassen sich ungerade Wurzeln $\sqrt[3]{x}, \sqrt[5]{x}, \ldots$ auch für negative x definieren, dann würden aber die Potenzgesetze nicht mehr gelten: $\sqrt[3]{x} = x^{\frac{1}{3}} = x^{\frac{2}{6}} = \sqrt[6]{x^2}$ (setze $x = -8$ ein).

[6]Ein Ergebnis wie $\sqrt{5}$ kann nicht weiter vereinfacht werden. Manchmal ist es sinnvoll, die Größenordnung $\sqrt{5} \approx 2{,}236$ anzugeben, aber die Schreibweise $\sqrt{5} = 2{,}236$ ist falsch.

1 Aussagenlogik

1.1 Aussagen und Quantoren

Definition 1.1. Eine (mathematische) **Aussage** ist ein Satz, der eindeutig wahr oder falsch ist. „Wahr“ (w) oder „falsch“ (f) heißt der **Wahrheitswert** der Aussage.

Dass jede Aussage immer entweder wahr oder falsch ist, aber niemals beides zugleich, bezeichnet man als das *Prinzip der Zweiwertigkeit*. Dies ist eines der wichtigsten logischen Grundprinzipien der Mathematik.

Beispiel 1.2. (1) „$2 < 1$.“ ist eine (falsche) Aussage.

(2) „97 ist eine Primzahl.“ ist eine (wahre) Aussage.

(3) „Heute ist das Wetter schön.“ ist keine Aussage, denn der Wahrheitswert ist nicht eindeutig bestimmbar.

(4) Aus dem gleichen Grund ist „$2x + 7 = 9$.“ keine Aussage, denn der Wahrheitswert hängt von x ab. „$2x + 7 = 9$.“ ist eine **Aussageform** $A(x)$, siehe Definition 1.5.

(5) „Dieser Satz ist falsch.“ ist keine Aussage, denn der Wahrheitswert ist nicht eindeutig bestimmbar: Ist der Satz wahr, so ist er falsch und umgekehrt. Definition 1.1 schließt auch widersprüchliche Sätze (*Paradoxa*) aus der Mathematik aus.

(6) „In jedem rechtwinkligen Dreieck ist die Summe der Kathetenquadrate gleich dem Hypotenusenquadrat.“ ist eine (wahre) Aussage, nämlich der Satz des Pythagoras. Ist auch die Aussage „Ist in einem Dreieck die Summe der Kathetenquadrate gleich dem Hypotenusenquadrat, so ist dieses Dreieck rechtwinklig.“ wahr?

(7) „Jede gerade Zahl größer als 2 lässt sich als Summe zweier Primzahlen darstellen.“ ist eine Aussage, nämlich die Goldbachsche Vermutung. Ihr Wahrheitswert wurde bis heute nicht ermittelt![1]

(8) „Mathematik macht Spaß“ ist natürlich eine wahre Aussage. ;)

[1] Ein Satz von Gödel besagt, dass es sogar Aussagen gibt, deren Wahrheitswert *prinzipiell* nicht festgestellt werden kann. Ein Beispiel dafür ist das sogenannte *Auswahlaxiom*: „Ist eine Ansammlung von Mengen gegeben, so lässt sich aus jeder Menge ein Element auswählen.“

Definition 1.3. Die **Negation** einer Aussage A ist wahr, wenn A falsch ist, und falsch, wenn A wahr ist. Die Negation von A wird mit $\neg A$ oder $\overline{A}$ bezeichnet (lies: „nicht A“).

Diese Situation kann man in einer **Wahrheitstafel** wie folgt darstellen:

A	$\overline{A}$
w	f
f	w

In der ersten Spalte stehen die möglichen Wahrheitswerte von A, in der zweiten Spalte die zugehörigen Wahrheitswerte von $\overline{A}$.

Beispiel 1.4.

(1) Die Negation der (wahren) Aussage „8 ist gerade.“ lautet „8 ist ungerade.“ und diese Aussage ist falsch.

(2) Negiert man den Satz des Pythagoras, so entsteht die (falsche) Aussage „Nicht in jedem rechtwinkligen Dreieck ist die Summe der Kathetenquadrate gleich dem Hypotenusenquadrat.“ oder gleichwertig „Es gibt ein rechtwinkliges Dreieck, in dem die Summe der Kathetenquadrate ungleich dem Hypotenusenquadrat ist.“

(3) Die Negation der Goldbachschen Vermutung lautet „Es gibt eine gerade Zahl größer als 2, die sich *nicht* als Summe zweier Primzahlen darstellen lässt.“ Auch von dieser Aussage ist natürlich nicht bekannt, ob sie wahr oder falsch ist.

(4) Bildet man die Negation von $\overline{A}$, so erkennt man anhand einer Wahrheitstafel, dass in den Spalten von A und $\overline{\overline{A}}$ die gleichen Wahrheitswerte stehen:

A	$\overline{A}$	$\overline{\overline{A}}$
w	f	w
f	w	f

Man sagt dann, A und $\overline{\overline{A}}$ sind **tautologisch äquivalent**, und schreibt $A \leftrightarrow \overline{\overline{A}}$. Zum Beispiel sind die Aussagen $A:$ „8 ist gerade.“ und $\overline{\overline{A}}:$ „8 ist nicht ungerade.“ tautologisch äquivalent; die beiden Negationen heben sich gegenseitig auf.[2]

Definition 1.5. Eine **Aussageform** enthält eine (oder mehrere) Variable x, sodass für jede Belegung von x eine Aussage $A(x)$ entsteht.

[2]In der Alltagssprache ist das anders: „Ich widerspreche dir nicht.“ ist *nicht* dasselbe wie „Ich stimme dir zu.“ Aber in der Mathematik gibt es nur zwei Möglichkeiten: wahr oder falsch.

Beispiel 1.6.

(1) $A(t)$: „t ist ein Wochentag.“ ist eine Aussageform. Zum Beispiel für die Belegung $t =$ Montag entsteht die (wahre) Aussage $A(\text{Montag})$: „Montag ist ein Wochentag.“, hingegen ist $A(\text{Tomate})$: „Tomate ist ein Wochentag.“ eine falsche Aussage. Da $A(t)$ für mindestens eine Belegung wahr ist, nennt man $A(t)$ *erfüllbar*.

(2) *Gleichungen* und *Ungleichungen* sind die häufigsten Beispiele für Aussageformen. Die Gleichung $A(x)$: „$2x + 7 = 9$.“ ist eine Aussageform, da für jede Belegung der Variable x eine (wahre oder falsche) Aussage $A(x)$ entsteht. Zum Beispiel ist die Aussage

$$A(1) \text{ wahr, da } 2 \cdot 1 + 7 = 9.$$
$$A(-1) \text{ falsch, da } 2 \cdot (-1) + 7 \neq 9.$$

Die Menge aller Belegungen, für die $A(x)$ eine wahre Aussage ist, heißt die **Lösungsmenge** $\mathcal{L}$ der Gleichung: (zur Schreibweise von Mengen siehe Beispiel 2.2)

$$\mathcal{L} = \{x : A(x) \text{ ist wahr}\} = \{x : \text{„}2x + 7 = 9.\text{“ ist wahr}\} = \{1\}.$$

Die Lösungsmenge der Ungleichung $B(x)$: „$-2x + 8 < -2$.“ ist $\mathcal{L} = \{x : x > 5\}$, das bedeutet: Für jede Belegung von x mit einer Zahl größer als 5 ist $B(x)$ eine wahre Aussage, sonst eine falsche.

(3) Der Term $A(a, b) = (a+b)^2$ ist *keine* Aussageform, da durch Belegen der Variablen a und b keine Aussage, sondern eine Zahl entsteht: $A(1, 2) = (1 + 2)^2 = 9$. Hingegen ist die Gleichung $B(a, b)$: „$(a + b)^2 = a^2 + 2ab + b^2$.“ eine Aussageform, die für *jede* Belegung von a und b eine wahre Aussage liefert. Daher nennt man die erste binomische Formel $B(a, b)$ *allgemeingültig*.

Die Allgemeingültigkeit oder Erfüllbarkeit von Aussageformen kann man mithilfe von Quantoren als Aussage formulieren.

Definition 1.7. Eine Aussageform sei allgemeingültig, also $A(x)$ für *jede* Belegung von x eine wahre Aussage. Dann schreibt man mithilfe des **Allquantors** $\forall$ (lies: „für alle“)

$$\forall x : A(x).$$

Ist die Aussageform erfüllbar, also $A(x)$ für (mindestens) eine Belegung von x eine wahre Aussage, so schreibt man mithilfe des **Existenzquantors** $\exists$ (lies: „es existiert“)

$$\exists x : A(x).$$

Gelegentlich verwendet man auch die Schreibweisen

$$\nexists x : A(x) \text{ für „Es existiert kein } x \text{, sodass } A(x) \text{ wahr ist.“}$$
$$\exists! x : A(x) \text{ für „Es existiert genau ein } x \text{, sodass } A(x) \text{ wahr ist.“}$$

Beispiel 1.8. (1) Die erste binomische Formel $B(a, b)$ aus Beispiel 1.6(3) ist eine allgemeingültige Aussageform und daher

$$\forall a\, \forall b : (a+b)^2 = a^2 + 2ab + b^2$$

eine wahre Aussage. Die Variablen a und b sind hier nur Platzhalter für beliebige Zahlen (sogenannte *gebundene Variable*), genauso richtig ist auch

$$\forall x\, \forall y : (x+y)^2 = x^2 + 2xy + y^2.$$

In der Regel erklärt man, dass a und b Platzhalter für beliebige *reelle Zahlen* sind, indem man schreibt

$$\forall a \in \mathbb{R}\, \forall b \in \mathbb{R} : (a+b)^2 = a^2 + 2ab + b^2 \text{ oder kürzer}$$
$$\forall a, b \in \mathbb{R} : (a+b)^2 = a^2 + 2ab + b^2.$$

(2) Die Gleichung $2x^2 + 7 = 9$ hat die Lösungsmenge $\mathcal{L} = \{-1, 1\}$, ist also *lösbar*. Das bedeutet, dass die Aussageform $2x^2 + 7 = 9$ für die Belegungen $x = -1$ und $x = 1$ wahr und damit auch die folgende Aussage wahr ist:

$$\exists x \in \mathbb{R} : 2x^2 + 7 = 9.$$

Der Existenzquantor behauptet nur, dass die Gleichung *mindestens* eine Lösung hat; die folgenden Aussagen sind falsch:

$$\forall x \in \mathbb{R} : 2x^2 + 7 = 9.$$
$$\exists!\, x \in \mathbb{R} : 2x^2 + 7 = 9.$$

(3) Die Gleichung $x^2 = 2$ hat bekanntlich keine rationale Lösung, d.h. $\sqrt{2}$ ist irrational. Die Aussage

$$\exists x \in \mathbb{Q} : x^2 = 2 \text{ („2 ist das Quadrat einer rationalen Zahl.“)}$$

ist also falsch, die Aussageform $x^2 = 2$ ist *unerfüllbar* über $\mathbb{Q}$.[3] Daher ist die Negation dieser Aussage wahr, sie lautet:

$$\nexists x \in \mathbb{Q} : x^2 = 2, \text{ („2 ist nicht das Quadrat einer rationalen Zahl.“)}$$
$$\text{äquivalent } \forall x \in \mathbb{Q} : x^2 \neq 2. \text{ („Das Quadrat jeder rationalen Zahl ist ungleich 2.“)}$$

Genauso ist die Negation der (wahren) Aussage $\forall x \in \mathbb{R} : x^2 \geqslant 0$ die (falsche) Aussage $\exists x \in \mathbb{R} : x^2 < 0$. **Merkregel:**

$$\neg\, (\forall x : A(x)) \leftrightarrow \exists x : \neg A(x), \qquad \neg\, (\exists x : A(x)) \leftrightarrow \forall x : \neg A(x).$$

[3] Die Aussage $\exists x \in \mathbb{R} : x^2 = 2$ ist natürlich wahr, hier kann man $x = \sqrt{2}$ oder $x = -\sqrt{2}$ wählen.

In komplexen Aussagen können mehrere Quantoren gleichzeitig auftreten; und auch diese Aussagen können negiert werden.

Beispiel 1.9. (1) Wie in Beispiel 1.6(3) ist die Reihenfolge mehrerer Allquantoren unerheblich, die folgenden (wahren) Aussagen sind gleichbedeutend:

$$\forall a \in \mathbb{R}\, \forall b \in \mathbb{R} : (a+b)^2 = a^2 + 2ab + b^2,$$
$$\forall b \in \mathbb{R}\, \forall a \in \mathbb{R} : (a+b)^2 = a^2 + 2ab + b^2 \text{ oder kürzer}$$
$$\forall a, b \in \mathbb{R} : (a+b)^2 = a^2 + 2ab + b^2.$$

Dasselbe gilt bei mehreren Existenzquantoren; folgende (wahre) Aussagen sind äquivalent:

$$\exists k \in \mathbb{Z}\, \exists l \in \mathbb{Z} : 2k + 3l = 1,$$
$$\exists l \in \mathbb{Z}\, \exists k \in \mathbb{Z} : 2k + 3l = 1 \text{ oder kürzer}$$
$$\exists k, l \in \mathbb{Z} : 2k + 3l = 1.$$

Die Negation dieser Aussage ist falsch und lautet $\forall k, l \in \mathbb{Z} : 2k + 3l \neq 1$.

(2) Die Aussage „Zu jeder Geraden gibt es eine parallele Gerade" schreibt sich mithilfe von Quantoren als

$$\forall \text{ Gerade } g\, \exists \text{ Gerade } h : g \parallel h.$$

Diese Aussage ist wahr, die folgende Aussage ist hingegen falsch:

$$\exists \text{ Gerade } h\, \forall \text{ Gerade } g : g \parallel h.$$

Das liegt daran, dass die Gerade h in der oberen Aussage passend zu g gewählt werden kann, also von g abhängen darf; in der unteren Aussage muss hingegen „zuerst" h gewählt werden, also unabhängig von g sein.[4] Die (wahre) Negation der unteren Aussage lautet übrigens

$$\forall \text{ Gerade } h\, \exists \text{ Gerade } g : g \nparallel h.$$

(3) Ein anderes Beispiel für das Phänomen in (2) ist das folgende:

$$\forall x \in \mathbb{Z}\, \exists y \in \mathbb{Z} : x + y = 0,$$
$$\exists y \in \mathbb{Z}\, \forall x \in \mathbb{Z} : x + y = 0.$$

Auch hier ist die obere Aussage wahr, da $y = -x$ abhängig von x gewählt werden kann. Die untere Aussage ist hingegen falsch, da ihre Negation wahr ist:

$$\forall y \in \mathbb{Z}\, \exists x \in \mathbb{Z} : x + y \neq 0.$$

Im Gegensatz zu (1) stellen wir daher fest: **Treten ungleichartige Quantoren gleichzeitig auf, so ist die Reihenfolge wichtig!**

[4] Insbesondere folgt die obere Aussage aus der unteren Aussage, d.h. in Implikationen dürfen Allquantoren „nach links" und Existenzquantoren „nach rechts" verschoben werden.

1.2 Logische Verknüpfungen

Wir wollen nun zwei gegebene Aussagen zu einer neuen Aussage verknüpfen, sodass diese logischen Verknüpfungen unserer intuitiven Vorstellung von Logik entsprechen.[5]

Definition 1.10. Seien A, B zwei Aussagen. Die Aussagen $A \wedge B$ (lies: „A **und** B“) sowie $A \vee B$ (lies: „A **oder** B“) werden durch folgende Wahrheitstafel definiert:

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

Das bedeutet: Sind A und B wahr, so ist auch die Aussage $A \wedge B$ wahr (erste Zeile), andernfalls ist $A \wedge B$ falsch (zweite bis vierte Zeile); sind A und B beide falsch, so ist auch $A \vee B$ falsch (vierte Zeile), andernfalls ist $A \vee B$ wahr (erste bis dritte Zeile).

Beispiel 1.11. (1) Die Aussage „$5 < 1$ und $3 \leqslant 3$“ ist falsch, denn die Teilaussage „$5 < 1$“ ist falsch und die Teilaussage „$3 \leqslant 3$“ ist wahr (dritte Zeile).
Genauso ist die Aussage „Manchmal regnet es oder alle Schüler mögen Mathematik.“ wahr, denn die Teilaussage „Manchmal regnet es.“ ist wahr. (zweite Zeile)

(2) Während das mathematische „und“ im Wesentlichen der Alltagssprache entspricht, wird das mathematische „oder“ immer nicht-ausschließend im Sinne von „und/oder“ verstanden (lat. vel). Zum Beispiel ist die Aussage

„F ist ein Parallelogramm oder eine Raute.“

auch dann wahr, wenn F ein Parallelogramm *und* eine Raute ist, also ein Quadrat. Hingegen bedeutet die Frage

„Möchten Sie Ihren Hamburger mit Pommes oder Salat?“

für gewöhnlich nicht, dass Pommes *und* Salat als Beilage möglich ist. Dem umgangssprachlichen ausschließenden „oder“ (lat. aut) entspricht das mathematische „entweder oder“, zum Beispiel in der Aussage

„Zwei Ebenen sind entweder parallel oder sie schneiden sich.“

[5] Allgemein versteht man unter einer *Verknüpfung* einen Vorgang, bei dem zwei Objekte einer Art zu einem dritten Objekt derselben Art kombiniert werden.

(3) Das Symbol „$\leqslant$“ bedeutet „$<$ oder $=$“; insbesondere ist die Aussage $a \leqslant b$ wahr, wenn $a = b$ ist. Zum Beispiel sind die folgenden Aussagen alle wahr:

$$-1 \leqslant 0, \qquad -1 < 0, \qquad -1 \leqslant -1, \qquad -1 \geqslant -1, \qquad -1 = -1,$$

ihre Negationen sind hingegen alle falsch:

$$-1 > 0, \qquad -1 \geqslant 0, \qquad -1 > -1, \qquad -1 < -1, \qquad -1 \neq -1.$$

(4) Statt der Aussage „Für alle ganzen Zahlen k ist $2k$ eine gerade Zahl und für alle ganzen Zahlen k ist $2k+1$ eine ungerade Zahl.“ kann man einfacher sagen: „Für alle ganzen Zahlen k ist $2k$ eine gerade und $2k+1$ eine ungerade Zahl.“ Symbolisch bedeutet dies, dass die folgenden (wahren) Aussagen gleichbedeutend sind:

$$(\forall k \in \mathbb{Z} : 2k \text{ ist gerade}) \text{ und } (\forall k \in \mathbb{Z} : 2k+1 \text{ ist ungerade})$$
$$\forall k \in \mathbb{Z} : 2k \text{ ist gerade und } 2k+1 \text{ ist ungerade.}$$

Der Allquantor ist also mit der logischen Verknüpfung „und“ *verträglich*; genauso ist der Existenzquantor mit der logischen Verknüpfung „oder“ verträglich: Für eine ganze Zahl x sind zum Beispiel folgende (wahre) Aussagen äquivalent:

$$(\exists k \in \mathbb{Z} : x = 2k) \text{ oder } (\exists k \in \mathbb{Z} : x = 2k+1)$$
$$\exists k \in \mathbb{Z} : x = 2k \text{ oder } x = 2k+1.$$

Der Allquantor ist hingegen mit „oder“ *unverträglich*, der Existenzquantor ist mit „und“ unverträglich. Machen Sie sich hierfür selbst ein Beispiel!

(5) Von den Aussagen A und $\overline{A}$ ist genau eine immer wahr, daher muss die Aussage $A \vee \overline{A}$ immer wahr sein, eine sogenannte **Tautologie**. In der Wahrheitstafel steht in der Spalte einer Tautologie nur „wahr“:

A	$\overline{A}$	$A \vee \overline{A}$
w	f	w
f	w	w

Die Tautologie $A \vee \overline{A}$ heißt das *Gesetz vom ausgeschlossenen Dritten* („tertium non datur“), ihre Negation $A \wedge \overline{A}$ ist immer falsch (**Kontradiktion**) und heißt das *Gesetz vom ausgeschlossenen Widerspruch.*

Wir kommen nun zur wichtigsten logischen Verknüpfung in der Mathematik: Der **Implikation** „Wenn A, dann B“. Viele verschiedene Verbalisierungen dieser Verknüpfung sind üblich, zum Beispiel:

- Wenn A wahr ist, dann ist auch B wahr.
- Aus A folgt B.
- Aus der Gültigkeit von A lässt sich B schlussfolgern.
- A ist hinreichend für B.
- Unter der Voraussetzung A ist B richtig.
- Wenn A gilt, dann muss schon B gelten.
- Damit A wahr sein kann, muss auch B wahr sein.
- Nur wenn B gilt, kann auch A gelten.
- B ist notwendig für A.

All diese Sprechweisen sind gleichbedeutend und werden mit $A \Rightarrow B$ bezeichnet.

Definition 1.12. Seien A, B zwei Aussagen. Die **Implikation** $A \Rightarrow B$ (lies: „A impliziert B“) wird durch folgende Wahrheitstafel definiert:

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

Das bedeutet: Die Implikation $A \Rightarrow B$ ist falsch, wenn A wahr, aber B falsch ist (zweite Zeile), also eine wahre Aussage eine falsche Aussage impliziert; in allen anderen Fällen ist die Implikation $A \Rightarrow B$ wahr.

Beispiel 1.13. (1) Die Aussage „Wenn es regnet, dann ist die Straße nass.“ ist eine Implikation $A \Rightarrow B$ mit den Teilaussagen A : „Es regnet.“ und B : „Die Straße ist nass.“ Diese Implikation ist wahr,

wenn es regnet und die Straße nass ist (erste Zeile),
wenn es nicht regnet und die Straße nass ist (dritte Zeile),
wenn es nicht regnet und die Straße nicht nass ist (vierte Zeile).

Regnet es hingegen, ohne dass die Straße nass ist, so muss die Implikation „Wenn es regnet, dann ist die Straße nass.“ falsch sein (zweite Zeile).

Man erkennt hieran: Die Implikation bewertet nur die Gültigkeit der *Schlussfolgerung*, nicht die Gültigkeit der einzelnen Teilaussagen! Wenn es überhaupt nicht regnet, dann ist die Schlussfolgerung „Wenn es regnet, dann ist die Straße nass.“ korrekt – unabhängig davon, ob die Straße nass ist. Eine falsche Aussage impliziert also jede beliebige wahre oder falsche Aussage. („ex falso quodlibet“)

(2) Die Implikation „Wenn $0 = 1$, dann ist $\sqrt{2}$ rational.“ ist eine wahre Aussage, denn die Teilaussage „$0 = 1$“ ist falsch. Wahlweise sieht man das auch so: Wir nehmen die Gleichung $0 = 1$ und multiplizieren beide Seiten mit $\sqrt{2}$; dann ist $0 = \sqrt{2}$ und 0 ist eine rationale Zahl, also auch $\sqrt{2}$. An diesem (etwas absurden) Beispiel erkennt man nochmal, warum sich die Mathematik nur für wahre Aussagen interessiert: Aus falschen Aussagen lässt sich alles schlussfolgern.

(3) Sei $f : D \to \mathbb{R}$ differenzierbar in einem inneren Punkt x_0 von D. Eine bekannte (wahre) Implikation aus der Analysis ist die Aussage

$$\text{Wenn } f \text{ bei } x_0 \text{ ein lokales Extremum besitzt, dann ist } f'(x_0) = 0. \tag{1.1}$$

In einer Implikation $A \Rightarrow B$ kann man A als **hinreichende** Bedingung für B und B als **notwendige** Bedingung für A auffassen. Das bedeutet hier: (1.1) ist ein

a) hinreichendes Kriterium für die Eigenschaft „$f'(x_0) = 0$“, d.h. aus einem lokalen Extremum bei x_0 lässt sich $f'(x_0) = 0$ schlussfolgern.

b) notwendiges Kriterium für Eigenschaft „f hat bei x_0 ein lokales Extremum“, d.h. für $f'(x_0) \neq 0$ kann f bei x_0 kein lokales Extremum haben.

Beachte: Hat f bei x_0 *kein* lokales Extremum, so kann $f'(x_0) = 0$ oder $f'(x_0) \neq 0$ gelten; gilt $f'(x_0) = 0$, so kann f bei x_0 ein lokales Extremum haben oder auch nicht. Wie sehen Beispiele für all diese Fälle aus?

(4) Implikationen $A \Rightarrow B$ werden in der Mathematik häufig als **Satz** formuliert, man sagt: Aus der **Voraussetzung** A folgt die **Behauptung** B. Jeder Satz erfordert einen **Beweis**, also einen Nachweis darüber, dass die Implikation $A \Rightarrow B$ wahr ist. Für die Aussage (1.1) könnte das so aussehen:

Satz. Sei $f : D \to \mathbb{R}$ differenzierbar in einem inneren Punkt x_0 von D und x_0 ein lokales Extremum von f. Dann ist $f'(x_0) = 0$.
BEWEIS (Dieser Beweis wird in der Analysis 1 geführt.) □[6]

Ist die Implikation $A \Rightarrow B$ wahr und die Aussage A wahr, so ist auch die Aussage B wahr („modus ponens“). Man sagt: Wir verwenden den bewiesenen Satz ($A \Rightarrow B$), indem wir seine Voraussetzung (A) überprüfen. Zum Beispiel hat die Funktion $f(x) = x^2$ ein lokales Minimum bei $x_0 = 0$, mit (1.1) folgt daraus $f'(0) = 0$.

[6] Das Symbol □ steht genau wie ■ oder q.e.d. („quod erat demonstrandum“) am Ende eines Beweises.

(5) Für ganze Zahlen $t, a, b \in \mathbb{Z}$ wollen wir folgende Implikation beweisen:[7]

$$t|a \text{ oder } t|b \Rightarrow t|ab. \tag{1.2}$$

Ist die Aussage „$t|a$ oder $t|b$“ falsch, so ist (1.2) stets wahr. Ist hingegen die Aussage „$t|a$ oder $t|b$“ wahr, so bedeutet das

$$\exists k \in \mathbb{Z} : tk = a \text{ oder } tk = b.$$

Also gilt $ab = tkb$ oder $ab = atk$, d.h. die Aussage „$t|ab$“ ist wahr; damit ist auch in diesem Fall die Implikation (1.2) wahr. An diesem Beispiel erkennt man: Um nachzuweisen, dass eine Implikation $A \Rightarrow B$ wahr ist, braucht man nur zu zeigen:

Wenn A wahr ist, dann ist auch B wahr.

Man nennt dies einen **direkten Beweis**. Die *Umkehrung* von (1.2) ist die Implikation

$$t|ab \Rightarrow t|a \text{ oder } t|b.$$

Sie ist nicht für alle ganzen Zahlen t wahr, sondern nur für Primzahlen $t \in \mathbb{P}$.

(6) Das folgende Schema kann helfen, den Überblick zwischen Bewiesenem und Behauptetem zu behalten.
Voraussetzung: Seien $x, y \in \mathbb{Z}$.
Behauptung: Sind x und y ungerade, so ist $x + y$ gerade.
Beweis Sind x, y ungerade, so existieren ganze Zahlen $k, l \in \mathbb{Z}$ mit $x = 2k + 1$ und $y = 2l + 1$. Daraus folgt

$$x + y = 2k + 1 + 2l + 1 = 2k + 2l + 2 = 2(k + l + 1),$$

d.h. $x + y$ ist von der Form $2m$ mit $m = k + l + 1 \in \mathbb{Z}$, also gerade. □

Bemerkung. Wir haben in Beispiel 1.13 gesehen, wie man die Wahrheitstafeln in den Definitionen 1.10, 1.12 und 1.14 nicht nur „vorwärts“ als logische Verknüpfungen, sondern auch „rückwärts“ als **Beweisstrategien** lesen kann. Dabei interessieren uns vor allem wahre Aussagen: Will man beweisen, dass eine Aussage der Form

- $A \Rightarrow B$ wahr ist, setzt man A als wahr voraus und zeigt dann, dass B wahr ist. (direkter Beweis)
- $A \wedge B$ wahr ist, zeigt man, dass A wahr ist und dass B wahr ist.
- $A \vee B$ wahr ist, zeigt man, dass A wahr ist oder dass B wahr ist.

Man kann die Wahrheitstafeln auch so lesen: Wenn $A \Rightarrow B$ und A beide wahr sind, so muss auch B wahr sein; wenn $A \wedge B$ falsch und A wahr ist, so muss B falsch sein etc. Wir werden im nächsten Abschnitt weitere Beweisstrategien kennenlernen.

[7] Die Schreibweise $t|a$ bedeutet „t teilt a“ oder „a ist ein Vielfaches von t“, formal $\exists k \in \mathbb{Z} : a = tk$.

Definition 1.14. Seien A, B zwei Aussagen. Die **Äquivalenz** $A \Leftrightarrow B$ (lies: „A äquivalent B") wird durch folgende Wahrheitstafel definiert:

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

Die Äquivalenz $A \Leftrightarrow B$ ist also nur dann wahr, wenn A und B denselben Wahrheitswert haben, andernfalls ist die Äquivalenz $A \Leftrightarrow B$ falsch. Das Symbol $\Leftrightarrow$ ist eine Kombination der Symbole $\Rightarrow$ und $\Leftarrow$, denn wir werden in Satz 1.19(6) sehen, dass die Aussage $A \Leftrightarrow B$ zur Aussage $(A \Rightarrow B) \wedge (A \Leftarrow B)$ tautologisch äquivalent ist. Auch für die Äquivalenz $A \Leftrightarrow B$ sind verschiedene Sprechweisen üblich:

- A ist äquivalent zu B.
- A gilt genau dann, wenn B gilt.
- A ist wahr dann und nur dann, wenn B wahr ist.
- A ist hinreichend und notwendig für B.

Beispiel 1.15.

(1) Die Äquivalenz „Der Mars ist ein Planet genau dann, wenn der Atlantik Salzwasser enthält." ist eine wahre Aussage, denn beide Teilaussagen „Der Mars ist ein Planet." und „Der Atlantik enthält Salzwasser." sind wahr (erste Zeile).
Genauso ist die Äquivalenz „Die Sonne ist ein Planet genau dann, wenn der Pazifik Süßwasser enthält." wahr, da beide Teilaussagen falsch sind (vierte Zeile).
Die Aussage „Der Atlantik enthält Salzwasser genau dann, wenn der Pazifik Süßwasser enthält." ist hingegen falsch, da der Wahrheitswert der Teilaussagen unterschiedlich ist (zweite Zeile). Wieder erkennt man: Die Äquivalenz trifft keine Aussage über die Gültigkeit der Teilaussagen oder einen kausalen Zusammenhang, sondern nur über die Korrektheit der *Schlussfolgerung*!

(2) Für alle reellen Zahlen $x, y \in \mathbb{R}$ gilt der **Satz vom Nullprodukt**:

$$xy = 0 \Leftrightarrow x = 0 \text{ oder } y = 0.$$

Das bedeutet ausführlich: Wenn einer der Faktoren x, y Null ist, dann ist das Produkt xy gleich Null („$\Leftarrow$"); ist umgekehrt das Produkt xy gleich Null, so muss schon einer der Faktoren x, y Null sein („$\Rightarrow$").

(3) Gleichungen sind Aussageformen, ihre Lösungsmenge $\mathcal{L}$ enthält alle Belegungen der Variablen, sodass die Gleichung eine wahre Aussage liefert. Zwei äquivalente Gleichungen haben dieselbe Lösungsmenge: Ist zum Beispiel in der Äquivalenz

$$15 - x = 3x - 1 \Leftrightarrow 16 = 4x$$

die linke Gleichung wahr, also x eine Lösung dieser Gleichung, so ist auch die rechte Gleichung wahr; liegt umgekehrt x in der Lösungsmenge der rechten Gleichung, so ist auch die linke Gleichung eine wahre Aussage. Sucht man systematisch äquivalente Gleichungen, so spricht man von **Äquivalenzumformungen**:

$$\begin{aligned} 15 - x = 3x - 1 &\Leftrightarrow \\ 16 = 4x \quad &\Leftrightarrow \\ 4 = x. & \end{aligned}$$

Die Lösungsmenge $\mathcal{L} = \{4\}$ der letzten Gleichung kann man einfach ablesen; da alle Gleichungen äquivalent sind, hat dann auch die ursprüngliche Gleichung $15 - x = 3x - 1$ die Lösungsmenge $\mathcal{L} = \{4\}$.

(4) Das Quadrieren ist *keine* Äquivalenzumformung: Zum Beispiel ist zwar jede Lösung der Gleichung $\sqrt{x} = 2 - x$ auch eine Lösung der Gleichung

$$x = 4 - 4x + x^2,$$

es gehen also keine Lösungen verloren. Aber die zweite Gleichung hat offenbar die Lösungsmenge $\mathcal{L} = \{1, 4\}$ und $x = 4$ ist *keine* Lösung der ersten Gleichung $\sqrt{x} = 2 - x$, durch das Quadrieren kommt also die *Scheinlösung* $x = 4$ hinzu. Da „$\Rightarrow$“ wahr und „$\Leftarrow$“ falsch ist, ist insbesondere die folgende Äquivalenz falsch:

$$\sqrt{x} = 2 - x \Leftrightarrow x = 4 - 4x + x^2.$$

Das liegt letztlich daran, dass auch die folgende Äquivalenz falsch ist:

$$x = \sqrt{a} \Leftrightarrow x^2 = a.$$

Zur Übersicht hier nochmal alle logischen Verknüpfungen in einer Tabelle:

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

1.3 Tautologien und Beweisstrategien

Die Gleichung $(a+b)^2 = a^2 + 2ab + b^2$ ist eine allgemeingültige Aussageform. Mit ihr kann man *innerhalb* eines komplexen Terms eine Vereinfachung vornehmen:

$$(n+1)^2 - 4n \overset{\downarrow}{=} n^2 + 2n + 1 - 4n = n^2 - 2n + 1 \overset{\downarrow}{=} (n-1)^2.$$

In Beispiel 1.4(4) haben wir die tautologische Äquivalenz $A \leftrightarrow \overline{\overline{A}}$ gefunden, in Beispiel 1.11(5) die Tautologie $A \vee \overline{A}$. Wir werden in diesem Abschnitt weitere solche allgemeingültigen Sätze beweisen und sehen, wie sie sich zur Vereinfachung eines komplexen logischen Ausdrucks oder als Beweisstrategie eignen.

Definition 1.16. Seien A, B Aussagen.
A und B heißen **tautologisch äquivalent**, wenn sie in der Wahrheitstafel die gleichen Einträge haben; in diesem Fall schreibt man $A \leftrightarrow B$.
A heißt **Tautologie**, falls jeder Eintrag von A in der Wahrheitstafel „wahr“ ist.
A heißt **Kontradiktion**, falls jeder Eintrag von A in der Wahrheitstafel „falsch“ ist.

Häufig schreibt man $\top$ für die Aussage, die immer wahr ist, und $\bot$ für die Aussage, die immer falsch ist; dann gilt offensichtlich:

$$\begin{aligned} &A \leftrightarrow \top \text{ genau dann, wenn } A \text{ eine Tautologie ist.}\\ &A \leftrightarrow \bot \text{ genau dann, wenn } A \text{ eine Kontradiktion ist.}\end{aligned}$$

Satz 1.17. Seien A, B, C Aussagen. Es gelten die folgenden tautologischen Äquivalenzen:

(1) $A \wedge B \leftrightarrow B \wedge A$. (**Kommutativgesetz** für $\wedge$)

(2) $A \vee B \leftrightarrow B \vee A$. (**Kommutativgesetz** für $\vee$)

(3) $(A \wedge B) \wedge C \leftrightarrow A \wedge (B \wedge C)$. (**Assoziativgesetz** für $\wedge$)

(4) $(A \vee B) \vee C \leftrightarrow A \vee (B \vee C)$. (**Assoziativgesetz** für $\vee$)

(5) $(A \wedge B) \vee C \leftrightarrow (A \vee C) \wedge (B \vee C)$. (Erstes **Distributivgesetz**)

(6) $(A \vee B) \wedge C \leftrightarrow (A \wedge C) \vee (B \wedge C)$. (Zweites **Distributivgesetz**)

(7) $\overline{A \wedge B} \leftrightarrow \overline{A} \vee \overline{B}$. (Erste **Regel von De Morgan**)

(8) $\overline{A \vee B} \leftrightarrow \overline{A} \wedge \overline{B}$. (Zweite **Regel von De Morgan**)

Beweis Alle Aussagen lassen sich beweisen, indem man für alle möglichen Wahrheitswerte von A, B, C anhand einer Wahrheitstafel feststellt, dass jeweils der Wahrheitswert der linken Aussage mit dem Wahrheitswert der rechten Aussage übereinstimmt. Wir tun dies hier exemplarisch für die Aussagen (1) und (3):

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

Aus dem Vergleich der vierten und fünften Spalte ergibt sich Aussage (1), aus dem Vergleich der letzten beiden Spalten die Aussage (3); auf ähnliche Weise beweist man auch die übrigen Aussagen.
Aussage (8) lässt sich auch aus (7) schlussfolgern, indem man zuerst (7) mit $\overline{A}, \overline{B}$ statt A, B verwendet: $\overline{\overline{A} \wedge \overline{B}} \leftrightarrow \overline{\overline{A}} \vee \overline{\overline{B}}$. Dann setzt man die tautologische Äquivalenz $A \leftrightarrow \overline{\overline{A}}$ aus Beispiel 1.4(4) ein und erhält

$$\overline{A \vee B} \leftrightarrow \overline{\overline{\overline{A}} \vee \overline{\overline{B}}} \leftrightarrow \overline{\overline{\overline{A} \wedge \overline{B}}} \leftrightarrow \overline{A} \wedge \overline{B}.$$

An diesem Beispiel erkennt man, wie sich tautologische Äquivalenzen zur Vereinfachung komplexer logischer Verknüpfungen einsetzen lassen; sie sind gewissermaßen Formeln und Gleichungen mit logischen Ausdrücken. □

Beispiel 1.18. (1) Die Regeln (1) bis (6) in Satz 1.17 verwenden wir ganz selbstverständlich, weil sie unserem natürlichen Sprachgebrauch entsprechen: Zum Beispiel gibt es keinen Unterschied zwischen den Aussagen „Es regnet und die Straße ist nass.“ und „Die Straße ist nass und es regnet.“ (Kommutativgesetz für $\wedge$), ebenso unterscheiden wir nicht die Klammerung in Sätzen wie „Die Ampel leuchtet rot, grün oder gelb.“ (Assoziativgesetz für $\vee$). Das Distributivgesetz erkennt man zum Beispiel in der Äquivalenz der Aussagen

„Schokolade ist weiß oder dunkel, aber auf jeden Fall süß.“
„Schokolade ist weiß und süß oder sie ist dunkel und süß.“

(2) Das Assoziativgesetz für $\wedge$ erlaubt es uns, die Schreibweise $A \wedge B \wedge C$ zu verwenden: Hier macht es nach Satz 1.17(3) keinen Unterschied, ob zuerst $A \wedge B$ oder $B \wedge C$ verknüpft wird.[8] Für vielfache Verknüpfungen von $\wedge$ schreibt man entsprechend

$$A_1 \wedge A_2 \wedge \ldots \wedge A_n.$$

[8] Zum Vergleich: Der Ausdruck $a + b + c$ ist sinnvoll, da die Addition $+$ assoziativ ist; der Ausdruck $a : b : c$ ist hingegen nicht sinnvoll, da die Division *nicht* assoziativ ist.

Eine solche Und-Verknüpfung ist genau dann wahr, wenn alle n Teilaussagen $A_1, A_2, \ldots, A_n$ wahr sind. Nach Satz 1.17(4) kann man entsprechend auch $A \vee B \vee C$ schreiben; eine Oder-Verknüpfung der Form $A_1 \vee A_2 \vee \ldots \vee A_n$ ist genau dann wahr, wenn (mindestens) eine der Teilaussagen $A_1, \ldots, A_n$ wahr ist.

(3) Die Regeln von De Morgan sind ein wichtiges Werkzeug, um komplexe Aussagen zu verneinen, sie treten in der Alltagssprache aber nur selten auf. Ein Beispiel: Die Abiturprüfung ist nur dann bestanden, wenn

„in Block I mindestens 200 Punkte
und in Block II mindestens 100 Punkte erreicht wurden."

Die Abiturprüfung nach Satz 1.17(7) also dann *nicht* bestanden, wenn

„in Block I weniger als 200 Punkte
oder in Block II weniger als 100 Punkte erreicht wurden."

(4) Mithilfe der Rechenregeln in Satz 1.17 können wir nun komplexe logische Aussagen systematisch vereinfachen:

$$\begin{aligned}\overline{A} \vee \overline{B} \vee (A \wedge (B \vee C)) &\leftrightarrow \overline{A \wedge B} \vee (A \wedge B) \vee (A \wedge C)\\ &\leftrightarrow \top \vee (A \wedge C) \leftrightarrow \top,\end{aligned}$$

die Aussage ist also eine Tautologie. Hierbei haben wir die dritte der folgenden vier tautologischen Äquivalenzen verwendet:

$$\begin{aligned} A \wedge \top &\leftrightarrow A, & A \wedge \bot &\leftrightarrow \bot,\\ A \vee \top &\leftrightarrow \top, & A \vee \bot &\leftrightarrow A.\end{aligned}$$

Satz 1.19. Es seien A, B, C Aussagen. Es gelten die folgenden tautologischen Äquivalenzen:

(1) $\overline{A} \Rightarrow \bot \leftrightarrow A$. („Reductio ad absurdum")

(2) $A \Rightarrow B \leftrightarrow \overline{A} \vee B$.

(3) $A \Rightarrow B \leftrightarrow A \wedge \overline{B} \Rightarrow \bot$. (**Beweis durch Widerspruch**)

(4) $A \Rightarrow B \leftrightarrow \overline{B} \Rightarrow \overline{A}$. (**Kontraposition**)

(5) $A \Rightarrow B \leftrightarrow (A \wedge C \Rightarrow B) \wedge (A \wedge \overline{C} \Rightarrow B)$. (**Fallunterscheidung**)

(6) $A \Leftrightarrow B \leftrightarrow (A \Rightarrow B) \wedge (B \Rightarrow A)$.

(7) $A \Leftrightarrow B \leftrightarrow \overline{A} \Leftrightarrow \overline{B}$.

(8) $A \Leftrightarrow B \leftrightarrow (A \wedge B) \vee (\overline{A} \wedge \overline{B})$.

(9) $\overline{A \Leftrightarrow B} \leftrightarrow \overline{A} \Leftrightarrow B$.

Beweis Wir beweisen zunächst die Aussagen (1) und (2) wie in Satz 1.17 mithilfe einer Wahrheitstafel:

A	B	$\bot$	$\overline{A}$	$\overline{A} \Rightarrow \bot$	$A \Rightarrow B$	$\overline{A} \vee B$
w	w	f	f	w	w	w
w	f	f	f	w	f	f
f	w	f	w	f	w	w
f	f	f	w	f	w	w

Daraus folgt nun mit der Regel von De Morgan (Satz 1.17(8)) die Aussage (3):

$$A \wedge \overline{B} \Rightarrow \bot \overset{(1)}{\leftrightarrow} \overline{A \wedge \overline{B}} \leftrightarrow \overline{A} \vee B \overset{(2)}{\leftrightarrow} A \Rightarrow B.$$

In ähnlicher Weise führt man (4) mit dem Kommutativgesetz für $\vee$ auf (2) zurück:

$$\overline{B} \Rightarrow \overline{A} \overset{(2)}{\leftrightarrow} \overline{\overline{B}} \vee \overline{A} \leftrightarrow \overline{A} \vee B \overset{(2)}{\leftrightarrow} A \Rightarrow B.$$

Die Aussagen (5), (6) und (8) sollen zur Übung selbst bewiesen werden. Damit führen wir auch die Aussagen (7) und (9) auf bereits bewiesene Aussagen zurück:

$$\begin{aligned} &A \Leftrightarrow B \leftrightarrow (A \Rightarrow B) \wedge (B \Rightarrow A) \leftrightarrow (\overline{B} \Rightarrow \overline{A}) \wedge (\overline{A} \Rightarrow \overline{B}) \leftrightarrow \overline{A} \Leftrightarrow \overline{B}, \\ &\overline{A \Leftrightarrow B} \leftrightarrow \overline{A \Rightarrow B} \vee \overline{B \Rightarrow A} \leftrightarrow (A \wedge \overline{B}) \vee (B \wedge \overline{A}) \leftrightarrow \overline{A} \Leftrightarrow B. \ \square \end{aligned}$$

Beispiel 1.20. Die Aussagen in Satz 1.19 stellen weitere Rechenregeln bereit, um komplexe logische Ausdrücke zu vereinfachen. Ihre wirkliche Bedeutung für die Mathematik liegt aber darin, sie als **Beweisstrategien** zu interpretieren.

(1) Sei $x \in \mathbb{Z}$. Zur Übung beweise man folgende Aussage direkt: $(A \Rightarrow B)$

$$x \text{ ungerade} \Rightarrow x^2 \text{ ungerade.} \tag{1.3}$$

Wir wollen nun die Umkehrung von (1.3) zeigen, also die Aussage $(B \Rightarrow A)$

$$x^2 \text{ ungerade} \Rightarrow x \text{ ungerade.} \tag{1.4}$$

Versucht man wieder einen direkten Beweis, so wird man feststellen, dass sich mit der Aussage $x^2 = 2k+1$ wenig anfangen lässt. Stattdessen nutzen wir Satz 1.19(4): Aussage (1.4) ist tautologisch äquivalent zur **Kontraposition** $(\overline{A} \Rightarrow \overline{B})$

$$x \text{ gerade} \Rightarrow x^2 \text{ gerade.}$$

Diese Aussage ist viel einfacher zu zeigen: Aus $x = 2k$ folgt sofort, dass auch $x^2 = (2k)^2 = 2(2k^2)$ gerade ist. $\square$

Wir wissen nun, dass die beiden Aussagen (1.3) und (1.4) wahr sind; nach Satz 1.19(6) ist dann auch die folgende Äquivalenz wahr: $(A \Leftrightarrow B)$

$$x \text{ ungerade} \Leftrightarrow x^2 \text{ ungerade.}$$

In der gleichen Weise kann man eine Äquivalenz „$\Leftrightarrow$“ in zwei Teilaussagen „$\Rightarrow$“ und „$\Leftarrow$“ zerlegen. Aus Satz 1.19(7) folgt wiederum, dass auch folgende Äquivalenz wahr ist: $(\overline{A} \Leftrightarrow \overline{B})$

$$x \text{ gerade} \Leftrightarrow x^2 \text{ gerade.}$$

(2) Die Regel in Satz 1.19(1) entspricht dem **indirekten Beweis** oder **Beweis durch Widerspruch**. Wir wollen die Aussage A beweisen; dazu treffen wir die *Annahme*, dass A falsch ist, und folgern daraus eine falsche Aussage $(\overline{A} \Rightarrow \perp)$.
Als Beispiel beweisen wir einen Klassiker: $\sqrt{2}$ ist irrational. *Annahme:* $\sqrt{2}$ ist nicht irrational, sondern rational. (Jetzt beginnt der indirekte Beweis und wir müssen eine falsche Aussage folgern.) Wenn $\sqrt{2}$ rational ist, lässt sich $\sqrt{2}$ als Bruch darstellen, d.h.

$$\exists z \in \mathbb{Z}, n \in \mathbb{N} : \sqrt{2} = \frac{z}{n}.$$

Indem wir diesen Bruch so weit wie möglich kürzen, können wir annehmen, dass z und n teilerfremd sind, also keinen gemeinsamen Teiler größer als 1 haben. $\sqrt{2}$ ist die positive Zahl, die im Quadrat 2 ergibt, daher quadrieren wir die obige Gleichung:

$$\sqrt{2} = \frac{z}{n} \Rightarrow 2 = \frac{z^2}{n^2} \Rightarrow 2n^2 = z^2.$$

Also ist die Zahl z^2 gerade; und nach (1) ist dann auch z gerade, das bedeutet $\exists k \in \mathbb{Z} : z = 2k$. Setzen wir dies in die obige Gleichung ein, so erhalten wir

$$2n^2 = (2k)^2 = 4k^2 \Rightarrow n^2 = 2k^2.$$

Mit dem gleichen Argument wie soeben ist n^2 und damit auch n gerade. Aber dann sind z und n beide gerade, haben also den gemeinsamen Teiler 2. Dies ist ein Widerspruch dazu, dass z und n teilerfremd sind. ↯ (Wir haben erfolgreich eine falsche Aussage gefolgert, jetzt ist der indirekte Beweis beendet.) Daher ist unsere Annahme falsch und $\sqrt{2}$ ist irrational. □

(3) Wir machen noch ein zweites Beispiel für einen Widerspruchsbeweis, nun in der Form $A \Rightarrow B$ von Satz 1.19(3). Wir wollen folgende Aussage beweisen:

$$\text{„Quadratische Funktionen haben keine Wendestelle.“} \tag{1.5}$$

Für den indirekten Beweis nehmen wir an, f ist eine quadratische Funktion mit einer Wendestelle $(A \wedge \overline{B})$, und wollen nun eine falsche Aussage folgern.

Eine quadratische Funktion ist eine Funktion der Form $f(x) = ax^2 + bx + c$; dabei ist $a \neq 0$, da f sonst nicht quadratisch, sondern linear wäre. Jede Wendestelle x_0 von f ist ein lokales Extremum der Ableitung f'; und diese Extrema erfüllen notwendig $f''(x_0) = 0$, siehe (1.1). Nun ist aber

$$f'(x) = 2ax + b, \qquad\qquad f''(x) = 2a \neq 0,$$

die Aussage $f''(x_0) = 0$ ist also ein Widerspruch zu $a \neq 0$. ↯ Damit ist (1.5) bewiesen. □

(4) Wir haben bewiesen, dass $\sqrt{2}$ irrational ist, d.h. $\sqrt{2} \notin \mathbb{Q}$. Nun wollen wir folgende Aussage beweisen:

$$\exists a, b \notin \mathbb{Q} : a^b \in \mathbb{Q}. \tag{1.6}$$

Dazu führen wir eine **Fallunterscheidung** (Satz 1.19(5)) durch: Wir dürfen eine Aussage C ganz beliebig wählen und müssen dann in beiden Fällen „C ist wahr“ und „C ist falsch“ zeigen, dass die Behauptung (1.6) wahr ist. Wir entscheiden uns (geschickt) für die Aussage C : „$\sqrt{2}^{\sqrt{2}}$ ist rational.“[9]

<u>Fall 1:</u> $\sqrt{2}^{\sqrt{2}}$ ist rational. Dann wählen wir $a = b = \sqrt{2} \notin \mathbb{Q}$
und die Aussage $a^b = \sqrt{2}^{\sqrt{2}} \in \mathbb{Q}$ ist wahr.

<u>Fall 2:</u> $\sqrt{2}^{\sqrt{2}}$ ist irrational. Dann wählen wir $a = \sqrt{2}^{\sqrt{2}} \notin \mathbb{Q}, b = \sqrt{2} \notin \mathbb{Q}$
und erhalten die Aussage $a^b = (\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^2 = 2 \in \mathbb{Q}$. □

In Fallunterscheidungen können natürlich auch drei oder mehr Fälle unterschieden werden, solange insgesamt alle Fälle abgedeckt werden; auch müssen sich die einzelnen Fälle nicht gegenseitig ausschließen.

(5) Wir möchten eine Aussage mit einem **Allquantor** beweisen, zum Beispiel

$$\forall x, y \geqslant 0 : x \leqslant y \Rightarrow x^2 \leqslant y^2. \tag{1.7}$$

Dazu nimmt man sich zwei $x, y \geqslant 0$ beliebig, quasi „zufällig“ her und zeigt für diese die Aussage „$x \leqslant y \Rightarrow x^2 \leqslant y^2$.“ Die Variablen x, y fungieren als Platzhalter für *beliebige* Zahlen, die Aussage wird quasi simultan für alle möglichen $x, y \geqslant 0$ bewiesen. Um diese Allgemeinheit nicht zu verletzen, darf man keine *speziellen* x, y wählen (zum Beispiel $x = 2, y = 3$), sondern muss mit den Variablen x, y arbeiten.

[9] Beachte: Wir wissen nicht, ob die Aussage C wahr ist; und das ist für die Fallunterscheidung auch unerheblich.

Durch Multiplikation der Ungleichung $x \leqslant y$ mit $x \geqslant 0$ erhält man die Ungleichung $x^2 \leqslant xy$, durch Multiplikation mit $y \geqslant 0$ entsprechend $xy \leqslant y^2$. Daraus folgt zusammen $x^2 \leqslant xy \leqslant y^2$, insbesondere $x^2 \leqslant y^2$. Die Ungleichung (1.7) ist jetzt *für alle* $x, y \geqslant 0$ bewiesen, da $x, y \geqslant 0$ beliebig gewählt wurden. □
Die folgende Aussage ist hingegen falsch:

$$\forall x, y \in \mathbb{R} : x \leqslant y \Rightarrow x^2 \leqslant y^2. \tag{1.8}$$

Um dies zu beweisen, müssen wir ein **Gegenbeispiel** angeben; wir zeigen also, dass die Negation von (1.8) wahr ist:

$$\exists x, y \in \mathbb{R} : x \leqslant y \text{ und } x^2 > y^2.$$

Da (1.7) wahr ist, muss dafür x oder y negativ sein; ein mögliches Gegenbeispiel ist $x = -2, y = 1$, denn

$$-2 \leqslant 1 \text{ und } (-2)^2 = 4 > 1 = 1^2. \ \square$$

Man beachte, dass die sprachliche Formulierung „Es gilt nicht $x \leqslant y \Rightarrow x^2 \leqslant y^2$.“ missverständlich ist, sie kann bedeuten:

$$\neg(\forall x, y \in \mathbb{R} : x \leqslant y \Rightarrow x^2 \leqslant y^2),$$
$$\forall x, y \in \mathbb{R} : \neg(x \leqslant y \Rightarrow x^2 \leqslant y^2).$$

Die erste Aussage haben wir soeben bewiesen, die zweite Aussage ist aber falsch. Um dieses Missverständnis zu vermeiden, verwendet man in der Mathematik die Formulierung „Es gilt *im Allgemeinen* nicht $x \leqslant y \Rightarrow x^2 \leqslant y^2$.“ und meint damit die schwächere Aussage, dass es ein Gegenbeispiel zur Aussage (1.8) gibt.

(6) Beweisen Sie als Übung die folgende tautologische Äquivalenz, den sogenannten **Ringschluss**:[10]

$$(A \Rightarrow B) \wedge (B \Rightarrow C) \wedge (C \Rightarrow A) \leftrightarrow (A \Leftrightarrow B) \wedge (B \Leftrightarrow C) \wedge (C \Leftrightarrow A).$$

Wir können also die Äquivalenz von drei Aussagen zeigen, indem wir einen „Ring“ aus drei Implikationen $A \Rightarrow B, B \Rightarrow C, C \Rightarrow A$ beweisen; dabei darf im zweiten Beweisschritt $B \Rightarrow C$ die Aussage A nicht mehr vorausgesetzt werden etc. Dieses Verfahren funktioniert auch mit mehr als drei Aussagen und komplexeren Ketten von Implikationen, solange man von jeder Aussage jede andere Aussage erreicht. Als Beispiel beweisen wir für eine Zahl $n \in \mathbb{Z}$ die Aussage

$$n \text{ ist ungerade} \Leftrightarrow 3n - 1 \text{ ist gerade} \Leftrightarrow \frac{n+1}{2} \in \mathbb{Z}.$$

[10] Für die rechte Aussage schreibt man kürzer $A \Leftrightarrow B \Leftrightarrow C$. Achtung: Das ist *nicht* gleichbedeutend mit den Aussagen $(A \Leftrightarrow B) \Leftrightarrow C \leftrightarrow A \Leftrightarrow (B \Leftrightarrow C)$.

Beweis „$A \Rightarrow B$": Ist n ungerade, so ist $n = 2k - 1$ für eine Zahl $k \in \mathbb{Z}$; daraus folgt $3n - 1 = 6k - 4 = 2(3k - 2)$, d.h. $3n - 1$ ist gerade.
„$B \Rightarrow C$": Ist $3n - 1 = 2k$ für ein $k \in \mathbb{Z}$, so lässt k beim Teilen durch 3 den Rest 1; daher ist $\frac{n+1}{2} = \frac{k+2}{3}$ eine ganze Zahl.
„$C \Rightarrow A$": Aus $\frac{n+1}{2} = k \in \mathbb{Z}$ folgt sofort $n = 2k - 1$, d.h. n ist ungerade. □

Bemerkung. Wir führen hier noch den Beweis durch **vollständige Induktion** auf, der in den Vorlesungen des ersten Semesters behandelt wird.[11] Um eine Aussage der Form $\forall n \in \mathbb{N} : A(n)$ zu beweisen, kann man wie folgt vorgehen:

(1) **Induktionsanfang:** Beweise die Aussage $A(1)$.

(2) **Induktionsschluss:** Beweise für alle $n \in \mathbb{N}$ die Implikation $A(n) \Rightarrow A(n+1)$.
(**Induktionsvoraussetzung $\Rightarrow$ Induktionsbehauptung**)

Aus $A(1)$ und (2) folgt dann $A(2)$, aus $A(2)$ und (2) folgt $A(3)$ usw., sodass die Aussage $A(n)$ für alle natürliche Zahlen $n \in \mathbb{N}$ bewiesen ist. Dieses Verfahren funktioniert *nicht* für Aussagen der Form $\forall x \in \mathbb{R} : A(x)$. Es gibt viele Varianten der vollständigen Induktion: Der Induktionsanfang muss nicht bei $n = 1$ liegen, als Induktionsvoraussetzung kann man $A(k)$ für alle $k \leqslant n$ wählen, auch Induktionen über $\mathbb{Z}$ mit zwei Induktionsschritten sind möglich. Als Beispiel beweisen wir folgende Aussage:

$$\forall n \in \mathbb{N} : 7 | 2^{n+1} + 3^{2n-1}.$$

Induktionsanfang: Für $n = 1$ ist $2^2 + 3^1 = 7$ durch 7 teilbar.
Induktionsschluss: Für ein $n \in \mathbb{N}$ gelte $7|2^{n+1} + 3^{2n-1}$, es ist $7|2^{n+2} + 3^{2n+1}$ zu zeigen. Wir versuchen, die Induktionsbehauptung auf die Induktionsvoraussetzung zurückzuführen und berechnen daher

$$2^{n+2} + 3^{2n+1} = 2 \cdot 2^{n+1} + 9 \cdot 3^{2n-1} = 2 \cdot (2^{n+1} + 3^{2n-1}) + 7 \cdot 3^{2n-1}.$$

Nach Induktionsvoraussetzung ist der Term in der Klammer durch 7 teilbar, also gleich $7k$ für eine ganze Zahl $k \in \mathbb{Z}$. Daraus folgt

$$2^{n+2} + 3^{2n+1} = 2 \cdot 7k + 7 \cdot 3^{2n-1} = 7(2 + 3^{2n-1}),$$

d.h. auch $2^{n+2} + 3^{2n+1}$ ist durch 7 teilbar. □

Es sei nochmal darauf hingewiesen, dass in einem mathematischen Beweis eines Satzes oder einer Behauptung *ausschließlich* die folgenden Aussagen zur Argumentation herangezogen werden dürfen:

- Als wahr angenommene Aussagen (Axiome),
- bereits bewiesene Sätze,
- die Voraussetzungen der zu beweisenden Aussage.

Diese strenge Unterscheidung und das logische Vorgehen machen das deduktive Prinzip aus, das die Grundlage der gesamten Mathematik ist.

[11] *Vollständig* soll darauf hindeuten, dass es sich dabei um eine deduktive Methode handelt.

Bemerkung. Wir haben gesehen, wie man mithilfe tautologischer Äquivalenzen wie in den Sätzen 1.17 und 1.19 komplexe logische Ausdrücke umformen und vereinfachen kann. Die Regeln von De Morgan erlauben zum Beispiel, jede Disjunktion $\vee$ durch eine Kombination aus Negation $\neg$ und Konjunktion $\wedge$ auszudrücken; mithilfe von Satz 1.19(2) und (8) sind wir auch in der Lage, Implikationen und Äquivalenzen durch $\neg, \wedge, \vee$, also schließlich durch $\neg$ und $\wedge$ auszudrücken. Damit lässt sich *jeder* logische Ausdruck durch eine Kombination von $\neg$ und $\wedge$ darstellen; diese Symbole bilden ein **vollständiges System**. Weitere vollständige Systeme sind zum Beispiel $\neg, \vee$; das „entweder oder" bzw. „XOR" $\dot\vee$; das „NAND" $\bar\wedge$, eine Kombination von NOT und AND ($A \mathbin{\bar\wedge} B \leftrightarrow \overline{A \wedge B}$) oder die analog definierten Verknüpfungen „XAND" und „NOR".
Vollständige Systeme sind besonders in der *Digitaltechnik* wichtig, also bei der technischen Umsetzung logischer Verknüpfungen zum Beispiel in elektronischen Schaltkreisen von Computern. Dort fasst man die Wahrheitswerte „wahr" und „falsch" als Bits 1 und 0 auf; aus der logischen Verknüpfung $\vee$ wird dann

$$1 \vee 1 = 1, \qquad 1 \vee 0 = 1, \qquad 0 \vee 1 = 1, \qquad 0 \vee 0 = 0,$$

die Verknüpfung $\wedge$ ist einfach die Multiplikation; die Rechenregeln in Satz 1.17 sind hier besonders wichtig. Aus einem vollständigen System lassen sich so beliebig komplexe logische Schaltungen realisieren, von digitalen Uhren und Multiplikationen im Dezimalsystem bis hin zu Taschenrechnern und modernen Betriebssystemen.

In der Mathematik verwendet man fast ausschließlich die logische Verknüpfung $\Leftrightarrow$ anstelle der tautologischen Äquivalenz $\leftrightarrow$, weil man nur an wahren Aussagen interessiert ist; dieses Vorgehen wird durch den folgenden Satz 1.21 gerechtfertigt: Zwei Aussagen sind genau dann tautologisch äquivalent, wenn ihre Äquivalenz eine Tautologie ist. Zum Vergleich: Wir können $(a+b)^2$ durch $a^2 + 2ab + b^2$ ersetzen genau dann, wenn $(a+b)^2 = a^2 + 2ab + b^2$ eine allgemeingültige Aussageform ist.

Satz 1.21. Seien A, B Aussagen. Dann gilt:

$$A \leftrightarrow B \text{ genau dann, wenn } A \Leftrightarrow B \text{ eine Tautologie ist.}$$

Beweis Es gelte $A \leftrightarrow B$, d.h. A und B haben in der Wahrheitstafel dieselben Einträge. Ist dann A wahr, so muss auch B wahr sein, d.h. $A \Leftrightarrow B$ ist wahr. Ist hingegen A falsch, so muss auch B falsch sein, also ist wiederum $A \Leftrightarrow B$ wahr. Damit ist $A \Leftrightarrow B$ eine Tautologie:

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

Sei umgekehrt $A \Leftrightarrow B$ eine Tautologie, also immer wahr. Ist dann A wahr, so muss nach Definition 1.14 auch B wahr sein; genauso muss B falsch sein, wenn A falsch ist. In der Wahrheitstafel ist die Situation also dieselbe wie zuvor und A, B sind tautologisch äquivalent. $\square$

2 Mengenlehre

2.1 Mengen

Eine exakte mathematische Definition davon, was eine Menge ist, ist mit hohem Aufwand verbunden und wurde in der Geschichte der Mathematik erst sehr spät gegeben. Wir wollen uns daher mit dem folgenden, „naiven“ Mengenbegriff zufriedengeben; später werden wir aber auch ein Beispiel sehen, für das dieser Begriff zu ungenau ist.

Definition 2.1. Eine **Menge** ist (nach Georg Cantor) „die Zusammenfassung bestimmter, wohlunterschiedener Objekte unserer Anschauung oder unseres Denkens zu einem Ganzen.“ Die Objekte, die zu einer Menge gehören, nennt man **Elemente** dieser Menge.

Die Menge, die keine Elemente enthält, heißt die **leere Menge** und wird mit $\emptyset$ oder $\{\}$ bezeichnet. Ist x ein Element der Menge M, so schreibt man $x \in M$ (lies: „x Element M“ oder „x in M“), andernfalls $x \notin M$.

Die Anzahl der Elemente einer (endlichen) Menge M heißt die **Mächtigkeit** oder Kardinalität von M und wird mit $\#M$ bezeichnet. Haben zwei (endliche) Mengen M, N die gleiche Mächtigkeit, also $\#M = \#N$, so nennt man M und N **gleichmächtig**.

Beispiel 2.2. (1) Es gibt im Wesentlichen zwei Schreibweisen, um eine Menge anzugeben: Fasst man die Objekte $1, 2, 3$ in einer Menge zusammen, so verwendet man die *aufzählende* Schreibweise $\{1, 2, 3\}$; es ist also $1 \in \{1, 2, 3\}$, aber zum Beispiel $4 \notin \{1, 2, 3\}$. Die Menge $\{1\}$ enthält nur das Element 1, aber es ist $1 \neq \{1\}$, denn links steht eine Zahl, rechts eine Menge. Das Wort „wohlunterschieden“ in Definition 2.1 bedeutet, dass gleiche oder mehrfach vorkommende Elemente nicht unterschieden werden; außerdem kommt es nicht auf die Reihenfolge an:

$$\{1, 2, 3\} = \{1, 1, 1, 2, 3\} = \{2, 1, 3\} = \{3, 1, \sqrt{2}^2, 9 - 7\}.$$

Alle diese Mengen haben die Kardinalität $\#\{1, 2, 3\} = 3$, auch wenn in der zweiten Darstellung mehr als drei Objekte stehen. In der Regel meint man daher mit einer Schreibweise wie $\{a, b, c\}$, dass die Elemente a, b, c *verschieden* sind.[1]

[1] Genauer würde man sagen, dass sie *paarweise verschieden* sind, also $a \neq b, b \neq c, c \neq a$.

(2) Mengen können nicht nur Zahlen enthalten, sondern ganz beliebige Objekte. Zum Beispiel ist $\{\Delta, \diamond, \square\}$ eine Menge von geometrischen Figuren mit Kardinalität 3, die Mengen $\{1, 2, 3\}$ und $\{\Delta, \diamond, \square\}$ gleichmächtig. Eine Gerade ist eine Menge von Punkten bzw. Vektoren $\vec{x}$, die von der Form $\vec{x} = \vec{a} + s\vec{v}$ sind; ebenso ist eine Ebene eine Menge von Punkten bzw. Vektoren. Die Menge der Stammfunktionen von $f(x) = 2x$ enthält alle Funktionen $g(x) = x^2 + C$ mit $C \in \mathbb{R}$.
Mengen können wiederum Mengen enthalten, was gelegentlich zu Verwirrung führt. Zum Beispiel enthält die Menge $\{1, \{2\}\}$ das Element 1, aber nicht das Element 2, sondern die Menge $\{2\}$, die selbst das Element 2 enthält:

$$1 \in \{1, \{2\}\}, \qquad 2 \notin \{1, \{2\}\} \qquad \{1\} \notin \{1, \{2\}\}, \qquad \{2\} \in \{1, \{2\}\}.$$

Die leere Menge $\emptyset$ enthält kein Element, daher ist $\#\emptyset = 0$ und es gilt immer $x \notin \emptyset$. Die Menge $\{\emptyset\}$ enthält hingegen genau ein Element, nämlich die leere Menge $\emptyset$, daher ist $\{\emptyset\} \neq \emptyset$ und $\#\{\emptyset\} = 1$.[2] Die Menge $\{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$ enthält alle aus der Schule bekannten Zahlenmengen; obwohl jede einzelne Zahlenmenge unendlich ist, ist diese Menge selbst endlich mit Kardinalität $\#\{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\} = 4$.

(3) Die aufzählende Schreibweise ist für sehr große oder unendliche Mengen etwas unpräzise: Zum Beispiel soll

$$\{1, \ldots, 101\}$$

die Menge der natürlichen Zahlen von 1 bis 101 bezeichnen, sie könnte aber auch für die Menge der ungeraden Zahlen zwischen 1 und 101 stehen. Entsprechend könnte $\{2, 4, 6, \ldots\}$ die (unendliche) Menge der geraden Zahlen bezeichnen – ganz klar ist aber nicht, wie die Folge $2, 4, 6$ weitergehen soll. Aus diesem Grund beschreibt man Mengen in der Regel mithilfe einer *Aussageform* (vgl. Definition 1.5): Die Menge

$$\{n \in \mathbb{N} : n \text{ ist gerade}\} \tag{2.1}$$

bezeichnet genau die Menge der geraden natürlichen Zahlen. Sie ist so zu lesen: Vor dem Doppelpunkt stehen die potentiellen Elemente der Menge, hier natürliche Zahlen n, hinter dem Doppelpunkt steht die zu erfüllende Bedingung („n ist gerade“), damit n tatsächlich Element der Menge ist. Die Elemente n der **Grundmenge** $\mathbb{N}$ werden also durch die Bedingung hinter dem Doppelpunkt eingeschränkt bzw. ausgewählt. Eine natürliche Zahl n ist genau dann Element der Menge (2.1), wenn sie beim Einsetzen in die Aussageform „n ist gerade“ eine wahre Aussage liefert; in diesem Sinne ist

$$\{n \in \mathbb{N} : n \text{ ist gerade}\} = \{2, 4, 6, \ldots\}.$$

Statt dem Doppelpunkt schreibt man gleichbedeutend auch $\{n \in \mathbb{N} \,|\, n \text{ ist gerade}\}$. Die obige Menge der natürlichen Zahlen von 1 bis 101 könnte man zum Beispiel so darstellen:

$$\{n \in \mathbb{N} : 1 \leqslant n \leqslant 101\}.$$

[2]Ein Sack, der einen leeren Sack enthält, ist nicht leer.

(4) Allgemeiner sei $A(x)$ ein Aussageform. Dann enthält die Menge

$$\{x : A(x)\}$$

genau die Elemente y, für die $A(y)$ eine wahre Aussage ist; symbolisch:

$$y \in \{x : A(x)\} \Leftrightarrow A(y) \text{ ist wahr.} \tag{2.2}$$

Überprüft man die Aussageform $A(x)$ nur für Elemente einer **Grundmenge** G, so schreibt man entsprechend

$$\{x \in G : A(x)\}.$$

Jedes Element dieser Menge ist immer auch ein Element der Menge G. Mithilfe von (2.2) lassen sich die Elemente einer Menge (linke Seite) mit einer logischen Aussage (rechte Seite) vergleichen; wir werden im folgenden solche Verknüpfungen nutzen, um die Aussagenlogik aus Kapitel 1 in Mengenschreib- und sprechweise zu übertragen. Da x der Schreibweise $\{x : A(x)\}$ nur ein Platzhalter ist (**gebundene Variable**), kann man die Variable jederzeit umbenennen: $\{x : A(x)\} = \{y : A(y)\}$.

(5) Die Gleichung $2x + 7 = 9$ hat die **Lösungsmenge** $\mathcal{L} = \{x : 2x + 7 = 9\}$, denn nach (2.2) gilt

$$x \in \{x : 2x + 7 = 9\} \Leftrightarrow 2x + 7 = 9 \text{ ist wahr,}$$

vergleiche Beispiel 1.6(2). Sucht man nach Lösungen in der Grundmenge der reellen Zahlen, so schreibt man $\mathcal{L} = \{x \in \mathbb{R} : 2x + 7 = 9\}$. Entsprechend ist zum Beispiel

$$\begin{aligned} &\{x \in \mathbb{R} : x^2 = 2\} = \{-\sqrt{2}, \sqrt{2}\}, \qquad &&\{x \in \mathbb{Z} : x^2 = 1\} = \{-1, 1\}, \\ &\{x \in \mathbb{Q} : x^2 = 2\} = \emptyset, &&\{x \in \mathbb{N} : x^2 = 1\} = \{1\}. \end{aligned}$$

An diesen Beispielen erkennt man, dass Mengen mit einer Aussageform endliche oder die leere Menge beschreiben können; aber auch unendliche (Lösungs-)Mengen sind möglich, die sich nicht aufzählend oder mit „...“ beschreiben lassen:

$$\{x \in \mathbb{R} : (x-3)^2 - x^2 = 9 - 6x\} = \mathbb{R}.$$

(6) Wir geben noch ein paar Beispiele für verschiedene Darstellungen von Mengen.
a) Jede gerade Zahl $n \in \mathbb{N}$ lässt sich in der Form $n = 2k$ für ein $k \in \mathbb{N}$ schreiben, daher kann man die Menge (2.1) auch darstellen als $\{2k : k \in \mathbb{N}\}$ oder gleichbedeutend $\{2n : n \in \mathbb{N}\}$. Damit ist

$$\{n \in \mathbb{N} : n \text{ ist gerade}\} = \{2n : n \in \mathbb{N}\} = \{2, 4, 6, 8, 10, \ldots\},$$

wobei n in der linken und in der mittleren Menge verschiedene Rollen spielt.

b) Eine Gerade im Raum wird in der analytischen Geometrie beschrieben durch die Gleichung $g : \vec{x} = \vec{a} + s\vec{v}$, das bedeutet: g besteht aus allen Punkten bzw. Vektoren $\vec{x}$, die von der Form $\vec{x} = \vec{a} + s\vec{v}$ für ein $s \in \mathbb{R}$ sind:

$$g = \{\vec{x} \in \mathbb{R}^3 : \exists s \in \mathbb{R} : \vec{x} = \vec{a} + s\vec{v}\} = \{\vec{a} + s\vec{v} : s \in \mathbb{R}\}.$$

Entsprechend gilt für eine Ebene E mit Stützvektor $\vec{a}$ und Normalenvektor $\vec{n}$

$$E = \{\vec{x} \in \mathbb{R}^3 : (\vec{x} - \vec{a}) \cdot \vec{n} = 0\}.$$

c) Würfelt man einmal mit einem sechsseitigen Würfel (W6), so nennt man die möglichen Augenzahlen $1, 2, 3, 4, 5, 6$ die *Ergebnisse* dieses Zufallsexperiments und die Menge $\{1, 2, 3, 4, 5, 6\}$ den *Ergebnisraum*. Verschiedene Teilmengen des Ergebnisraums können als *Ereignisse* interpretiert werden, zum Beispiel

$\{2, 4, 6\}$: Es wird eine gerade Zahl gewürfelt.
$\{4, 5, 6\}$: Es wird eine Zahl größer als 3 gewürfelt.
$\{2, 3, 5\}$: Es wird eine Primzahl gewürfelt.

Bemerkung. Definition 2.1 ist nicht präzise genug, damit die Bildung von Mengen wie in (2.2) für jede beliebige Aussageform funktioniert. Betrand Russell fand 1903 als Erster eine Aussageform $A(x)$, für die $\{x : A(x)\}$ keine Menge sein kann, für die (2.2) also zu einer Inkonsistenz in der Mathematik führt (**Russellsches Paradoxon**). Er betrachtete das Objekt

$$M := \{x : x \notin x\},$$

das also alle Mengen x enthält, die sich nicht selbst als Element enthalten. Die Frage ist dann: Enthält M sich selbst, gilt also $M \in M$?
Wenn $M \in M$ gilt, dann hat M die Eigenschaft $M \notin M$, ein Widerspruch. Wenn hingegen $M \notin M$ gilt, dann muss die Bedingung „$x \notin x$“ für $x = M$ falsch sein und es gilt $M \in M$, ebenfalls ein Widerspruch.
In der modernen Mathematik wird dieses Problem gelöst, indem man eine sehr präzise, aber auch sehr aufwendige Definition des Begriffs Menge angibt und Mengenbildungen der Form (2.2) nicht mehr für alle, sondern nur noch für bestimmte Aussageformen zulässt, sodass immer eine Menge entsteht. Diese von Ernst Zermelo und Abraham Fraenkel um 1930 entwickelte axiomatische Mengenlehre (**ZF-System**) bildet heute das Fundament der gesamten Mathematik – für unsere Zwecke und die der allermeisten Mathematiker genügt aber das Wissen, dass Mengenbildungen der Form (2.2) für fast alle Aussageformen widerspruchsfrei funktionieren.

Definition 2.3. Zwei Mengen M, N heißen **gleich**, wenn sie genau dieselben Elemente haben, wenn also jedes Element von N auch ein Element von M ist und umgekehrt. Symbolisch bedeutet das: Es gilt $N = M$ genau dann, wenn

$$\forall x : x \in N \Leftrightarrow x \in M.$$

N heißt **Teilmenge** von M, wenn jedes Element von N auch ein Element von M ist; symbolisch: Es gilt $N \subseteq M$ (lies: „N Teilmenge M") genau dann, wenn

$$\forall x : x \in N \Rightarrow x \in M.$$

In dieser Situation nennt man M eine **Obermenge** von N. Die Menge aller Teilmengen von M heißt die **Potenzmenge** von M:

$$\mathcal{P}(M) := \{N : N \subseteq M\}.$$

Ist $N \subseteq M$ eine Teilmenge, aber nicht gleich M, so schreibt man auch $N \subsetneqq M$ und sagt, N ist eine **echte Teilmenge** von M. Das Symbol $\subset$ wird uneinheitlich sowohl im Sinne von $\subseteq$ als auch im Sinne von $\subsetneqq$ gebraucht.

Beispiel 2.4. (1) Jedes Element $1, 2$ der Menge $\{1, 2\}$ ist auch Element der Menge $\{1, 2, 3\}$, daher ist $\{1, 2\}$ eine Teilmenge von $\{1, 2, 3\}$, symbolisch $\{1, 2\} \subseteq \{1, 2, 3\}$. Hingegen ist $3 \in \{1, 2, 3\}$ kein Element der Menge $\{1, 2\}$, die beiden Mengen sind also *nicht* gleich: In der Aussage

$$\forall x : x \in \{1, 2\} \Rightarrow x \in \{1, 2, 3\}$$

wäre die Richtung „$\Leftarrow$" für $x = 3$ falsch. Daher nennt man $\{1, 2\}$ eine echte Teilmenge von $\{1, 2, 3\}$ und schreibt $\{1, 2\} \subsetneqq \{1, 2, 3\}$. Definition 2.3 klärt auch, weshalb die Mengen $\{1, 2\}$ und $\{2, 1, 1, 2\}$ gleich sind: In der (wahren) Aussage

$$\forall x : x \in \{1, 2\} \Leftrightarrow x \in \{2, 1, 1, 2\}$$

sind die Reihenfolge und die Häufigkeit der Elemente unerheblich; es kommt nur auf die Elemente selbst an (*Prinzip der Extensionalität*). Unsere Vorstellung von einer Menge hängt eng mit dem Begriff der Gleichheit zweier Mengen zusammen.

(2) Jedes Quadrat ist auch ein Rechteck, daher ist

$$\{A : A \text{ ist ein Quadrat}\} \subseteq \{A : A \text{ ist ein Rechteck}\}.$$

Die Quadrate sind eine echte Teilmenge der Rechtecke, da es Rechtecke gibt, die keine Quadrate sind; man könnte hier also das Symbol „$\subseteq$" durch „$\subsetneqq$" ersetzen.

In der gleichen Weise ist die Menge der natürlichen Zahlen eine (echte) Obermenge sowohl der geraden als auch der ungeraden Zahlen:

$$\{n \in \mathbb{N} : n \text{ ist gerade}\} = \{2n : n \in \mathbb{N}\} \subseteq \mathbb{N},$$
$$\{n \in \mathbb{N} : n \text{ ist ungerade}\} = \{2n-1 : n \in \mathbb{N}\} \subseteq \mathbb{N}.$$

Für die erste Inklusion könnte man etwas unpräzise auch schreiben $\{2,4,6,\ldots\} \subseteq \{1,2,3,\ldots\}$. Allgemein gilt $\{x \in G : A(x)\} \subseteq G$ wie in Beispiel 2.2(4).

(3) In Definition 2.3 verwendet die Gleichheit = zweier Mengen die Äquivalenz $\Leftrightarrow$ und die Teilmengenrelation $\subseteq$ die Implikation $\Rightarrow$; damit ist es möglich, Aussagenlogik in die Mengenlehre zu übertragen. Zum Beispiel folgt aus Satz 1.19(6)

$$\begin{aligned} N = M &\Leftrightarrow (\forall x : x \in N \Leftrightarrow x \in M) \\ &\Leftrightarrow (\forall x : x \in N \Rightarrow x \in M \text{ und } x \in M \Rightarrow x \in N) \\ &\Leftrightarrow (\forall x : x \in N \Rightarrow x \in M) \text{ und } (\forall x : x \in M \Rightarrow x \in N) \\ &\Leftrightarrow N \subseteq M \text{ und } M \subseteq N. \end{aligned}$$

Zwei Mengen sind also genau dann gleich, wenn beide Teilmengen voneinander sind; wir können die Gleichheit $M = N$ in zwei Richtungen $M \subseteq N$ und $M \supseteq N$ zerlegen (**Antisymmetrie** von $\subseteq$), genauso wie wir die Äquivalenz „$\Leftrightarrow$" in „$\Rightarrow$" und „$\Leftarrow$" zerlegen können. Als Beispiel für dieses Prinzip beweisen wir die Aussage[3]

$$\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} = \{(\cos\varphi, \sin\varphi) : \varphi \in \mathbb{R}\}. \tag{2.3}$$

Für die Inklusion „$\subseteq$" sei (x,y) ein Punkt aus der linken Menge, also mit der Eigenschaft $x^2 + y^2 = 1$. Diese Gleichung kann nur gelten, wenn der Betrag von x und y nicht zu groß ist, wenn nämlich $-1 \leqslant x, y \leqslant 1$. Daher finden wir ein (sogar zwei) φ, sodass $\cos\varphi = x$, und die Gleichung $x^2 + y^2 = 1$ liefert $y^2 = 1 - x^2 = 1 - \cos^2\varphi = \sin^2\varphi$. Wir wählen jetzt φ mit $\cos\varphi = x$ so, dass auch $\sin\varphi = y$ gilt; dafür betrachtet man am besten die Graphen von sin und cos. Dann ist $(x,y) = (\cos\varphi, \sin\varphi)$, also (x,y) ein Element der rechten Menge und die Inklusion „$\subseteq$" bewiesen.
Für die andere Richtung „$\supseteq$" sei $(\cos\varphi, \sin\varphi)$ ein Punkt aus der rechten Menge. Dann gilt $\cos^2\varphi + \sin^2\varphi = 1$ und daher liegt $(\cos\varphi, \sin\varphi)$ in der linken Menge. Insgesamt ist damit die Gleichheit (2.3) gezeigt. □
Entsprechend beweist man mithilfe von Satz 1.19(2) die allgemeinen Aussagen

$$\begin{aligned} &N \not\subseteq M \Leftrightarrow \exists x : x \in N \text{ und } x \notin M, \\ &N \neq M \Leftrightarrow \exists x : (x \in N \text{ und } x \notin M) \text{ oder } (x \in M \text{ und } x \notin N). \end{aligned}$$

[3] Mit (x,y) bezeichnen wir Punkte in der Ebene $\mathbb{R}^2$ mit den Koordinaten x und y.

(4) Statt der Aussage $1 \in \mathbb{N}$ kann man auch $\{1\} \subseteq \mathbb{N}$ schreiben; umgekehrt gilt aber $\{1\} \notin \mathbb{N}$ und $1 \nsubseteq \mathbb{N}$. Wenn eine Menge selbst wieder Mengen enthält, können auch alle Aussagen zugleich richtig sein:

$$1 \in \{1, \{1\}\}, \qquad \{1\} \subseteq \{1, \{1\}\}, \qquad \{1\} \in \{1, \{1\}\} \qquad \{\{1\}\} \subseteq \{1, \{1\}\}.$$

Sei $A \subseteq \mathbb{N}$ eine Teilmenge; alle Teilmengen sind echt außer $A = \mathbb{N}$, die Aussage $A \subsetneqq \mathbb{N}$ schließt also nur den Fall $A = \mathbb{N}$ aus. Das Symbol „$\subsetneqq$" ist eine stärkere Aussage als „$\subseteq$":

$$A \subsetneqq B \Leftrightarrow A \subseteq B \text{ und } A \neq B.$$

Umgekehrt ist „$\subseteq$" eine schwächere Aussage als „$\subsetneqq$":

$$A \subseteq B \Leftrightarrow A \subsetneqq B \text{ oder } A = B.$$

Man kann diese Situation mit den Symbolen „$\leqslant$" und „$<$" für Zahlen vergleichen: Die Eigenschaft $a \leqslant b$ kann $a < b$ oder $a = b$ bedeuten, aber $a < b$ bedeutet sowohl $a \leqslant b$ als auch $a \neq b$ (mehr dazu in Kapitel 3). Eine Hierarchie zwischen den verschiedenen Teilmengenbeziehungen sieht so aus:

$$= \Rightarrow \subseteq \Leftarrow \subsetneqq \Rightarrow \neq$$

(5) Es seien A, B zwei (endliche) Mengen. Aus der Gleichheit $A = B$ folgt $\#A = \#B$, d.h. die Mengen A und B sind gleichmächtig. Die Umkehrung dieser Aussage gilt nicht: Die Mengen $\{1, 2\}$ und $\{3, 4\}$ sind zwar gleichmächtig, aber nicht gleich. Entsprechend folgt aus der Inklusion $A \subseteq B$ die Ungleichung $\#A \leqslant \#B$ und man könnte vermuten, dass aus der echten Inklusion $A \subsetneqq B$ auch die echte Ungleichung $\#A < \#B$ folgt. Für endliche Mengen ist das richtig, aber für unendliche Mengen[4] ist die Situation überraschend anders, wie zuerst Georg Cantor mit zwei Diagonalargumenten gezeigt hat: Es gilt zwar

$$\mathbb{N} \subsetneqq \mathbb{Z} \subsetneqq \mathbb{Q} \subsetneqq \mathbb{R},$$

aber $\#\mathbb{N} = \#\mathbb{Z} = \#\mathbb{Q}$. Hingegen ist $\#\mathbb{N} \neq \#\mathbb{R}$, obwohl beide Mengen unendlich sind; es gibt also *verschiedene* unendliche Mächtigkeiten. Dies war einer der Gründe für die ZF-Mengenlehre und warf die sogenannte *Kontinuumshypothese* auf: Gibt es noch andere Mächtigkeiten zwischen $\#\mathbb{N}$ und $\#\mathbb{R}$?

[4] Bei einer unendlichen Menge können wir nicht wie in Definition 2.1 die Elemente zählen, um die Mächtigkeit zu bestimmen; stattdessen nennt man zwei (unendliche) Mengen M, N gleichmächtig, wenn es eine bijektive Abbildung $f : M \to N$ gibt.

(6) Die Potenzmenge $\mathcal{P}(M)$ einer Menge M enthält alle Teilmengen von M, d.h.

$$N \in \mathcal{P}(M) \Leftrightarrow N \subseteq M.$$

$\mathcal{P}(M)$ ist eine Menge von Mengen, ein sog. *Mengensystem*, und kann für große Mengen M schnell sehr unübersichtlich werden; für kleine Mengen wie $M = \{0\}$ oder $M = \{0, 1\}$ lässt sich die Potenzmenge aber noch gut darstellen:

$$\begin{aligned} \mathcal{P}(\{0\}) &= \{\emptyset, \{0\}\}, \\ \mathcal{P}(\{0,1\}) &= \{\emptyset, \{0\}, \{1\}, \{0,1\}\}. \end{aligned}$$

Für jede Menge gilt $M \subseteq M$ (**Reflexivität** von $\subseteq$) und $\emptyset \subseteq M$, daher ist stets $M \in \mathcal{P}(M)$ und $\emptyset \in \mathcal{P}(M)$; im Spezialfall $M = \emptyset$ gibt es nur eine einzige Teilmenge der leeren Menge, nämlich die leere Menge selbst: $\mathcal{P}(\emptyset) = \{\emptyset\}$.
Hat M nur endlich viele Elemente, so gibt es auch nur endlich viele verschiedene Teilmengen von M; durch Abzählen aller Möglichkeiten erhält man

$$\#\mathcal{P}(M) = 2^{\#M}.$$

Für unendliche Mengen ist die Situation komplexer, zum Beispiel $\#\mathcal{P}(\mathbb{N}) = \#\mathbb{R}$.

Definition 2.5. Die Zahlenmengen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ wurden bereits eingeführt. Alle (reellen) Zahlen lassen sich auf der **Zahlengeraden** veranschaulichen und ihrer Größe nach sortieren. Zusammenhängende Abschnitte der Zahlengeraden werden als **Intervalle** bezeichnet; man unterscheidet ($a, b \in \mathbb{R}$)

(1) **Offene** Intervalle[5]
$$\begin{aligned} (a,b) &:= \{x \in \mathbb{R} : a < x < b\}, \\ (a,\infty) &:= \{x \in \mathbb{R} : x > a\}, \\ (-\infty,b) &:= \{x \in \mathbb{R} : x < b\}. \end{aligned}$$

(2) **Abgeschlossene** Intervalle
$$\begin{aligned} [a,b] &:= \{x \in \mathbb{R} : a \leqslant x \leqslant b\}, \\ [a,\infty) &:= \{x \in \mathbb{R} : x \geqslant a\}, \\ (-\infty,b] &:= \{x \in \mathbb{R} : x \leqslant b\}. \end{aligned}$$

(3) **Halboffene** Intervalle
$$\begin{aligned} (a,b] &:= \{x \in \mathbb{R} : a < x \leqslant b\}, \\ [a,b) &:= \{x \in \mathbb{R} : a \leqslant x < b\}. \end{aligned}$$

Beachte, dass bei den Grenzen $\pm\infty$ immer eine runde (ausschließende) Klammer steht.

[5] Manchmal schreibt man $]a, b[$ statt (a, b) etc., um dies nicht mit dem geordneten Paar zu verwechseln.

Für alle Arten von Intervallen sind weitere Sprech- und Schreibweisen üblich:

- Für die **Intervallgrenzen** a, b gilt in der Regel $a < b$, in diesem Fall nennt man das Intervall **echt**. Ist hingegen $a \geqslant b$, so heißt das Intervall I **unecht**; dann ist $I = \emptyset$ außer im Fall $[a, a] = \{a\}$.
- Intervalle der Form $(a, b), [a, b], (a, b], [a, b)$ heißen **beschränkt**, das Intervall $[a, b]$ auch **kompakt**; alle anderen Intervalle heißen **unbeschränkt**. Die **Länge** von echten beschränkten Intervallen ist $|I| := b - a$, von unechten Intervallen $|I| := 0$.
- Ein leeres Intervall $I = \emptyset$ nennt man sowohl offen als auch abgeschlossen; dasselbe gilt für den Sonderfall $(-\infty, \infty) := \mathbb{R}$. Ein Intervall der Form $(a - \varepsilon, a + \varepsilon)$ mit $\varepsilon > 0$ heißt auch eine **ε-Umgebung** von a oder eine **ε-Kugel** um a.

Satz 2.6. Sei $I \subseteq \mathbb{R}$. Die folgende Charakterisierung von Intervallen wird in Analysis 1 bewiesen:

$$I \text{ ist ein Intervall} \Leftrightarrow \forall x, y \in I \ \forall z \in \mathbb{R} : x < z < y \Rightarrow z \in I.$$

(Von der Richtung „$\Rightarrow$“ können Sie sich jetzt schon überzeugen.)

Beispiel 2.7. (1) Genauso wie die Rechenregeln für Zahlen ist aus der Schule bekannt, dass alle Zahlenmengen durch *Zahlbereichserweiterungen* entstehen.
a) Ausgangspunkt sind die natürlichen Zahlen $\mathbb{N} = \{1, 2, 3, \ldots\}$, durch Hinzufügen der Null erhält man $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$. Daher gilt $\mathbb{N} \subseteq \mathbb{N}_0, 0 \in \mathbb{N}_0, 0 \notin \mathbb{N}$ und zum Beispiel $101 \in \mathbb{N}, -1 \notin \mathbb{N}_0$.
b) Erweitert man $\mathbb{N}_0$ um die negativen Zahlen, so erhält man $\mathbb{Z} = \{0, 1, -1, \ldots\}$; dabei sind alle natürlichen Zahlen auch ganze Zahlen: $\mathbb{N} \subseteq \mathbb{N}_0 \subseteq \mathbb{Z}$. Es gilt zum Beispiel $0 \in \mathbb{Z}, 15 \in \mathbb{Z}, -22 \in \mathbb{Z}$, aber $\frac{1}{2} \notin \mathbb{Z}$ und $4{,}2 \notin \mathbb{Z}$.
c) Fügt man den ganzen Zahlen die Brüche bzw. die abbrechenden und periodischen Dezimalzahlen hinzu, so erhält man die rationalen Zahlen $\mathbb{Q}$. Wiederum sind alle ganzen Zahlen auch rationale Zahlen, d.h. $\mathbb{Z} \subseteq \mathbb{Q}$ und zum Beispiel $0 \in \mathbb{Q}, -4 \in \mathbb{Q}, \frac{4}{3} \in \mathbb{Q}$ und $-0{,}1 \in \mathbb{Q}$, aber $\sqrt{5} \notin \mathbb{Q}, \log 3 \notin \mathbb{Q}, e \notin \mathbb{Q}$.
d) Bei der letzten Erweiterung zu den reellen Zahlen $\mathbb{R}$ werden *alle noch fehlenden*[6] Zahlen ergänzt, das sind genau die irrationalen Zahlen bzw. nicht periodischen, nicht abbrechenden Dezimalzahlen wie $\sqrt{2}$ oder π. Wieder sind alle rationalen Zahlen auch reelle Zahlen, es gilt also $\mathbb{Q} \subseteq \mathbb{R}$; zusammenfassend:

$$\mathbb{N} \subseteq \mathbb{N}_0 \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}.$$

Da bei jeder Zahlenbereichserweiterung neue Zahlen hinzukommen, kann man hier jede Inklusion „$\subseteq$“ auch durch „$\subsetneqq$“ ersetzen.

[6] Diese Zahlbereichserweiterung wird im Laufe des Mathematikstudiums präzisiert.

(2) Intervalle enthalten immer reelle Zahlen, sie beschränken sich niemals nur auf rationale, ganze oder natürliche Zahlen. Das Intervall $(0, 1)$ umfasst alle Zahlen zwischen 0 und 1, also einen zusammenhängenden Bereich auf der Zahlengeraden. Dabei bedeutet die runde (offene/ ausschließende) Klammer, dass die Grenzen 0 und 1 *nicht* mehr zum Intervall gehören; eine Schreibweise wie $[0, 1]$ mit eckigen (abgeschlossenen/ einschließenden) Klammern bedeutet hingegen, dass die Grenzen 0 und 1 zum Intervall gehören. Zum Beispiel ist $0 \notin (0, 1)$ und $0 \in [0, 1]$, aber in beiden Fällen $\frac{1}{2} \in (0, 1)$ und $\frac{1}{2} \in [0, 1]$. Offensichtlich gilt auch

$$(a, b) \subseteq (a, b] \subseteq [a, b], \qquad (a, b) \subseteq [a, b) \subseteq [a, b].$$

Nach rechts unbeschränkte Intervalle wie $(1, \infty)$ oder nach links unbeschränkte Intervalle wie $(-\infty, 1]$ werden symbolisch mit den Grenzen ∞ und $-\infty$ geschrieben, auch wenn diese Symbole natürlich *keine Zahlen* sind. Einige Beispiele:

$$\begin{array}{cc} (1,2) \subseteq (1,\infty) \subseteq [1,\infty), & [-1,0) \subseteq (-\infty,0) \subseteq (-\infty,0], \\ [1,0) \subseteq [1,\infty) \subseteq [-1,\infty), & (-\infty,-1] \subseteq (-\infty,0) \subseteq (-\infty,1]. \end{array}$$

Man kann die gegenseitige Lage, Teilmengenbeziehungen und die Länge verschiedener Intervalle gut auf der Zahlengeraden veranschaulichen. Dabei ist eine generelle Konvention, dass offene Grenzen mit einer runden Klammer, einem leeren Kreis oder einer gestrichelten Linie dargestellt werden, abgeschlossene Grenzen mit einer eckigen Klammer, einem ausgefüllten Kreis oder einer durchgezogenen Linie.

(3) Man beachte, dass die Zahlenmengen selbst Mengen sind, auch wenn keine Mengenklammern sichtbar sind; zum Beispiel ist $3 \in \mathbb{R}$, denn die Zahl 3 ist in der Menge $\mathbb{R}$ enthalten. Umgekehrt enthält die Menge $\{\mathbb{R}\}$ nur ein Element, es ist $\mathbb{R} \neq \{\mathbb{R}\}$ und $3 \notin \{\mathbb{R}\}$.
Das Intervall $(0, \infty) = \{x \in \mathbb{R} : x > 0\}$ beschreibt die Menge der positiven Zahlen und wird auch mit $\mathbb{R}_+$ bezeichnet; entsprechend bezeichnet $\mathbb{R}_- = (-\infty, 0) = \{x \in \mathbb{R} : x < 0\}$ die Menge der negativen Zahlen. Der jeweils gegenüberliegende Teil der Zahlengeraden sind die nichtpositiven Zahlen $(-\infty, 0] = \{x \in \mathbb{R} : x \leqslant 0\}$ bzw. die nichtnegativen Zahlen $[0, \infty) = \{x \in \mathbb{R} : x \geqslant 0\}$.

2.2 Verknüpfungen von Mengen

Wir haben die logischen Verknüpfungen $\Rightarrow$ und $\Leftrightarrow$ in die Beziehungen $\subseteq$ und $=$ zwischen Mengen übertragen, nun wollen wir das auch für die Verknüpfungen $\wedge, \vee$ und $\neg$ tun; ferner werden wir Mengen mit dem kartesischen Produkt $\times$ verknüpfen.

Definition 2.8. Seien M, N Mengen. Die **Schnittmenge** $M \cap N$ (lies: „M geschnitten N") enthält alle Elemente, die sowohl in M als auch in N liegen:

$$M \cap N := \{x : x \in M \text{ und } x \in N\}.$$

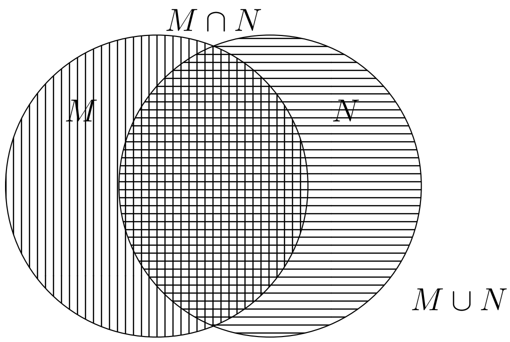
Gibt es keine gemeinsamen Elemente, also $M \cap N = \emptyset$, so nennt man M und N **disjunkt**. Die **Vereinigungsmenge** $M \cup N$ (lies: „M vereinigt N") enthält alle Elemente, die in M oder in N liegen:

$$M \cup N := \{x : x \in M \text{ oder } x \in N\}.$$

Mithilfe von (2.2) wird der Zusammenhang zur Aussagenlogik deutlich:

$$x \in M \cap N \Leftrightarrow x \in M \wedge x \in N,$$
$$x \in M \cup N \Leftrightarrow x \in M \vee x \in N.$$

Man kann Mengenbeziehungen häufig gut in einem *Venn-Diagramm* veranschaulichen: Hier entspricht $M \cap N$ dem Bereich, der sowohl von M als auch von N überdeckt wird; $M \cup N$ entsteht, wenn man die Bereiche von M und N zusammenfasst.



Beispiel 2.9. (1) a) Für den Durchschnitt der Mengen $\{1, 2, 3\}$ und $\{1, 2, 4\}$ gilt $\{1, 2, 3\} \cap \{1, 2, 4\} = \{1, 2\}$, denn (genau) die Elemente 1 und 2 sind in beiden Mengen enthalten. Für die Vereinigung gilt $\{1, 2, 3\} \cup \{1, 2, 4\} = \{1, 2, 3, 4\}$, denn die Elemente $1, 2, 3, 4$ kommen in (mindestens) einer der beiden Mengen vor; die Elemente $1, 2$ werden wie üblich nicht doppelt aufgeführt.

b) Im Alltag treten Schnitt und Vereinigung zum Beispiel bei der Suche in Datenbanken auf: Eine Suche nach „Mathelabor Karlsruhe" ergibt üblicherweise alle Einträge, die „Mathelabor" *und* „Karlsruhe" enthalten, also den Schnitt

$$\{x : x \text{ enthält "Mathelabor"}\} \cap \{x : x \text{ enthält "Karlsruhe"}\}.$$

Eine Suche nach „Mathelabor OR Karlsruhe" ergibt hingegen alle Einträge mit „Mathelabor" und alle Einträge mit „Karlsruhe", also die Vereinigung[7]

$$\{x : x \text{ enthält "Mathelabor"}\} \cup \{x : x \text{ enthält "Karlsruhe"}\}.$$

[7]Moderne Suchmaschinen nutzen natürlich zusätzliche Features wie Priorisierung nach Relevanz oder Sponsoring, automatische Vervollständigung oder Korrektur. Weitere nützliche Befehle sind zum Beispiel die exakte Suche „", der Platzhalter * oder - zum Ausschließen von Suchbegriffen.

c) Ist A ein Rechteck, aber auch eine Raute, so ist A ein Quadrat; mithilfe von Mengen schreibt sich diese Eigenschaft als Durchschnitt,

$$\{A : A \text{ ist Rechteck}\} \cap \{A : A \text{ ist Raute}\} = \{A : A \text{ ist Quadrat}\}.$$

Hingegen gibt es keine Dreiecke, die sowohl rechtwinklig als auch gleichseitig sind:

$$\{\Delta : \Delta \text{ ist rechtwinklig}\} \cap \{\Delta : \Delta \text{ ist gleichseitig}\} = \emptyset,$$

diese beiden Mengen sind disjunkt. Es gilt immer $A \cup \emptyset = A$ und $A \cap \emptyset = \emptyset$, die leere Menge ist also disjunkt zu jeder anderen Menge.[8] Übrigens ist für mehr als zwei Mengen die Formulierung „A, B, C sind disjunkt" unklar, denn man kann Mengen A, B, C finden, sodass

$$A \cap B \cap C = \emptyset, \text{ aber } A \cap B \neq \emptyset, B \cap C \neq \emptyset, C \cap A \neq \emptyset.$$

Finden Sie ein Beispiel hierfür? Mit der Formulierung „A, B, C sind paarweise disjunkt" meint man tatsächlich $A \cap B = \emptyset, B \cap C = \emptyset, C \cap A = \emptyset$.

(2) Die Aussage „$x \in A \cap B$" ist eine stärkere Aussage als „$x \in A$", denn sie stellt an die Elemente von A die zusätzliche Bedingung „$x \in B$". Daher gilt immer $A \cap B \subseteq A$ und genauso $A \cap B \subseteq B$, die Schnittmenge $A \cap B$ ist in der Regel kleiner als A bzw. B. Zum Beispiel gilt

$$\mathbb{P} \cap \{n \in \mathbb{N} : n \text{ ist ungerade}\} = \{n \in \mathbb{P} : n \text{ ist ungerade}\} \subseteq \mathbb{P},$$

die Primzahlen $n \in \mathbb{P}$ werden durch die Bedingung „n ist ungerade" eingeschränkt. In der gleichen Weise ist „$x \in A \cup B$" eine schwächere Aussage als „$x \in A$", denn aus „$x \in A$ oder $x \in B$" folgt nicht zwangsläufig „$x \in A$". Jedes Element von A bzw. B ist auch ein Element von $A \cup B$, daher gilt immer $A \subseteq A \cup B$ und $B \subseteq A \cup B$. In Satz 2.11 werden wird zeigen, dass in diesen beiden Regeln $A \cap B \subseteq A$ und $A \subseteq A \cup B$ genau dann Gleichheit gilt, wenn $A \subseteq B$. Ein paar weitere Beispiele:

$$\begin{aligned} \mathbb{N} \cup \{0\} &= \mathbb{N}_0, & [1,2] \cup (2,3] &= [1,3], & (-1,3) \cap [0,5] &= [0,3), \\ \mathbb{R} \cap \mathbb{Q} &= \mathbb{Q}, & (-\infty,-1) \cup (-2,0] &= (-\infty,0], & (-\infty,1) \cap [-1,\infty) &= [-1,1). \end{aligned}$$

(3) Die Lösungen eines Gleichungssystems wie

$$\begin{aligned} 3x^2 + 5 &= 8, \\ 5 - 2x &= 7, \end{aligned}$$

sind die Zahlen x, die die obere *und* die untere Gleichung simultan erfüllen; die Lösungsmenge ist daher der Durchschnitt $\mathcal{L} = \{x : 3x^2 + 5 = 8\} \cap \{x : 5 - 2x = 7\}$ der Lösungsmengen der beiden einzelnen Gleichungen.

[8] Man vergleiche diese Aussagen mit Beispiel 1.18(4).

Umgekehrt entsteht eine Vereinigung von Lösungsmengen, wenn man den Satz vom Nullprodukt anwendet: Jede Lösung der Gleichung $(x-2)(x+1)=0$ erfüllt $x-2=0$ *oder* $x+1=0$, daher ist

$$\begin{aligned}\mathcal{L} &= \{x : (x-2)(x+1)=0\} = \{x : x-2=0\} \cup \{x : x+1=0\} \\ &= \{2\} \cup \{-1\} = \{2,-1\}.\end{aligned}$$

(4) Sind A, B endliche Mengen, so sind auch $A \cap B$ und $A \cup B$ endlich. Die Inklusionen $A \cap B \subseteq A$ und $A \subseteq A \cup B$ in (2) ergeben die Ungleichungen $\#(A \cap B) \leqslant \#A \leqslant \#(A \cup B)$. Möchte man die Anzahl der Elemente in $A \cap B$ exakt abzählen, so zählt man die Elemente in A und in B separat, hat dann aber die Elemente in $A \cap B$ doppelt gezählt. Daher gilt für die Kardinalitäten

$$\#A + \#B = \#(A \cup B) + \#(A \cap B). \tag{2.4}$$

Diese Formel macht man sich häufig in der Wahrscheinlichkeitsrechnung zunutze: Wird ein W6 einmal gewürfelt, so ist $\{1,2,3,4,5,6\}$ der Ergebnisraum, die Teilmenge $A = \{2,4,6\}$ beschreibt das Ereignis „Gerade Zahl gewürfelt" und $B = \{1,2,3,4\}$ das Ereignis „Zahl kleiner als 5 gewürfelt". Für die Ereignisse

$A \cap B$: "Gerade Zahl kleiner als 5 gewürfelt",

$A \cup B$: "Gerade Zahl oder Zahl kleiner als 5 gewürfelt"

gilt $A \cap B = \{2,4\}$ und nach (2.4) daher

$$\#(A \cup B) = \#A + \#B - \#(A \cap B) = 3 + 4 - 2 = 5.$$

Da das Würfeln mit einem sechsseitigen Würfel ein *Laplace-Experiment* ist, hat also $A \cup B$ die *Wahrscheinlichkeit* $\frac{5}{6}$. Diesen Wert konnten wir mithilfe von (2.4) bestimmen, ohne die Elemente von $A \cup B$ zu kennen.

Satz 2.10. Seien A, B, C Mengen. Es gelten die folgenden Rechenregeln:

(1) $A \cap B = B \cap A$. (**Kommutativgesetz** für $\cap$)

(2) $A \cup B = B \cup A$. (**Kommutativgesetz** für $\cup$)

(3) $(A \cap B) \cap C = A \cap (B \cap C)$. (**Assoziativgesetz** für $\cap$)

(4) $(A \cup B) \cup C = A \cup (B \cup C)$. (**Assoziativgesetz** für $\cup$)

(5) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$. (Erstes **Distributivgesetz**)

(6) $(A \cup B) \cap C = (A \cap C) \cup (B \cup C)$. (Zweites **Distributivgesetz**)

Beweis Die Aussagen entsprechen Satz 1.17(1) bis (6) und können elementweise übersetzt werden, zum Beispiel für (1)

$$x \in A \cap B \Leftrightarrow x \in A \wedge x \in B \Leftrightarrow x \in B \wedge x \in A \Leftrightarrow x \in B \cap A.$$

Daraus folgt $A \cap B = B \cap A$ nach Definition 2.3; genauso geht man auch für die übrigen Aussagen vor. $\square$

Bemerkung. Analog zu den logischen Verknüpfungen $\wedge$ und $\vee$ erlaubt uns das Assoziativgesetz in Satz 2.10(3) und (4), mehrfache Schnitte und Vereinigungen zu definieren und dabei die Klammern wegzulassen; für die Mengen $A_1 \cap \ldots \cup A_n$ und $A_1 \cup \ldots \cup A_n$ gilt also

$$x \in A_1 \cap \ldots \cap A_n \Leftrightarrow x \in A_1 \wedge \ldots \wedge x \in A_n,$$
$$x \in A_1 \cup \ldots \cup A_n \Leftrightarrow x \in A_1 \vee \ldots \vee x \in A_n.$$

Mit dieser Methode kann man auch unendliche Schnitte und Vereinigungen bilden und schreibt dafür analog zum Reihenzeichen $\sum_{n=1}^{\infty}$

$$\bigcap_{n=1}^{\infty} \left(-\frac{1}{n}, \frac{1}{n}\right) = (-1,1) \cap \left(-\frac{1}{2}, \frac{1}{2}\right) \cap \left(-\frac{1}{3}, \frac{1}{3}\right) \cap \ldots = \{0\}, \qquad \bigcap_{b>1} [0,b] = [0,1],$$
$$\bigcup_{n=1}^{\infty} \left[\frac{1}{n}, 1\right] = \{1\} \cup \left[\frac{1}{2}, 1\right] \cup \left[\frac{1}{3}, 1\right] \cup \ldots = (0,1], \qquad \bigcup_{a>0} (-a,a) = \mathbb{R}.$$

Auch für solche Schnitte und Vereinigungen bleiben Kommutativ-, Assoziativ- und Distributivgesetz gültig. Für ein Mengensystem $\mathcal{A}$ schreibt man gelegentlich auch

$$\bigcap \mathcal{A} = \bigcap_{A \in \mathcal{A}} A, \qquad \bigcup \mathcal{A} = \bigcup_{A \in \mathcal{A}} A,$$

in diesem Sinne ist zum Beispiel $\bigcup\{(-a,a) : a > 0\} = \mathbb{R}$.

Satz 2.11. Seien A, B Mengen. Die folgenden Aussagen sind äquivalent:

(1) $A \subseteq B$,

(2) $A \cap B = A$,

(3) $A \cup B = B$.

Beweis Um die Aussage (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3) zu zeigen, machen wir einen Ringschluss und beweisen die Implikationen (1) $\Rightarrow$ (2), (2) $\Rightarrow$ (3) und (3) $\Rightarrow$ (1), vgl. Beispiel 1.20(6).

„(1) $\Rightarrow$ (2)“: Es gelte $A \subseteq B$, zu zeigen ist $A \cap B = A$. Diese Mengengleichheit zerlegen wir in die beiden Inklusionen „$\subseteq$“ und „$\supseteq$“, vgl. Beispiel 2.4(3). $A \cap B \subseteq A$ ist klar. Für „$\supseteq$“ sei $x \in A$, aus der Voraussetzung $A \subseteq B$ folgt dann $x \in B$, also insgesamt $x \in A \cap B$. Damit ist $A \cap B = A$ und die Implikation „(1) $\Rightarrow$ (2)“ bewiesen.
„(2) $\Rightarrow$ (3)“: Es gelte $A \cap B = A$, zu zeigen ist $A \cup B = B$. Wieder ist $A \cup B \supseteq B$ klar. Für „$\subseteq$“ sei $x \in A \cup B$, also $x \in A$ oder $x \in B$. Ist $x \in A$, so folgt aus (2) auch $x \in A \cap B$, insbesondere $x \in B$. Also folgt aus $x \in A \cup B$ sicher $x \in B$ und wir haben insgesamt $A \cup B = B$ bewiesen.
„(3) $\Rightarrow$ (1)“: Es gelte $A \cup B = B$, zu zeigen ist $A \subseteq B$. Sei $x \in A$, dann ist auch $x \in A \cup B$, nach (3) also $x \in B$. Damit ist $A \subseteq B$ gezeigt. $\square$

Wir wollen nun die Negation $\neg$ auf Mengen übertragen. In einer Definition der Form $x \in \overline{N} \Leftrightarrow x \notin N$ ist aber die Aussageform „$x \notin N$“ zu unspezifisch, um eine Menge zu bilden. Daher schränken wir auf eine gegebene Grundmenge M ein; die Differenzmenge $M \setminus N$ ist dann je nach Sichtweise eine Verknüpfung der Mengen M und N oder die Negation von N innerhalb der Grundmenge M.

Definition 2.12. Seien M, N Mengen. Die **Differenzmenge** $M \setminus N$ (lies: „M ohne N“) enthält alle Elemente von M, die nicht in N liegen:

$$M \setminus N := \{x \in M : x \notin N\}.$$

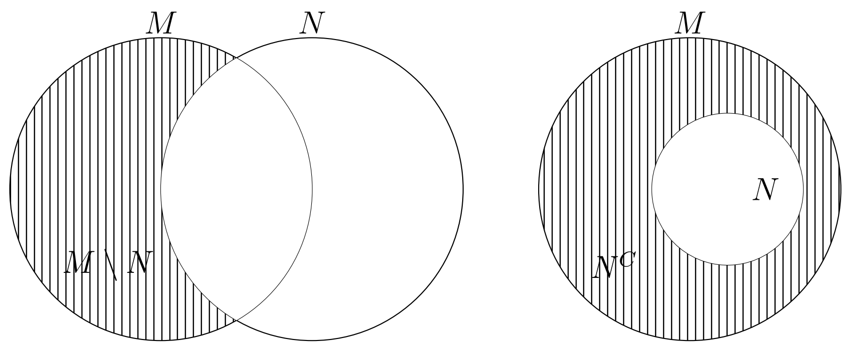
Aus der Menge M werden also alle Elemente entfernt, die zu N gehören; es gilt

$$\forall x : x \in M \setminus N \Leftrightarrow x \in M \text{ und } x \notin N.$$

Gilt zusätzlich $N \subseteq M$, so kann man M als **Grundmenge** interpretieren und nennt $N^C := M \setminus N$ (lies: „N Komplement“) das **Komplement** von N in M. Da alle Elemente von N dann auch zu M gehören, hat man die einfachere Aussage

$$\forall x \in M : x \in N^C \Leftrightarrow x \notin N.$$

In einem Venn-Diagramm umfasst $M \setminus N$ den Bereich von M, der nicht zu N gehört; entsprechend ist N^C der Bereich, mit dem N zur Grundmenge M ergänzt werden kann.



Beispiel 2.13.

(1) a) Die Differenz der Mengen $\{1, 2, 3\}$ und $\{1, 3\}$ ist $\{1, 2, 3\} \setminus \{1, 3\} = \{2\}$; da $\{1, 3\} \subseteq \{1, 2, 3\}$ gilt, kann man auch $M = \{1, 2, 3\}$ als Grundmenge interpretieren und das Komplement $\{1, 3\}^C = \{2\}$ bilden. Die Differenz $\{1, 2, 3\} \setminus \{1, 4\} = \{2, 3\}$ lässt sich hingegen *nicht* als Komplementbildung interpretieren, da $\{1, 4\}$ keine Teilmenge von $\{1, 2, 3\}$ ist.

b) Das Komplement der (positiven) geraden Zahlen $G = \{n \in \mathbb{N} : n \text{ ist gerade}\} = \{2, 4, 6, \ldots\}$ in den natürlichen Zahlen $\mathbb{N}$ ist die Menge der ungeraden Zahlen $U = \{n \in \mathbb{N} : n \text{ ist ungerade}\} = \{1, 3, 5, \ldots\}$:

$$G^C = \mathbb{N} \setminus G = \{n \in \mathbb{N} : n \notin G\} = \{n \in \mathbb{N} : n \text{ ist ungerade}\} = U,$$

denn für eine Zahl $n \in \mathbb{N}$ sind die Aussagen $n \notin G$ und $n \in U$ äquivalent. Bildet man hingegen das Komplement von G in den ganzen Zahlen $\mathbb{Z}$, so kann $n \notin G$ auch $n \in \{0, -1, -2, \ldots\}$ bedeuten; man erhält also ein anderes Komplement:

$$G^C = \mathbb{Z} \setminus G = \{n \in \mathbb{Z} : n \notin G\} = \{-n : n \in \mathbb{N}_0\} \cup U.$$

Man erkennt: Bei der Schreibweise G^C muss immer klar sein, auf welche Grundmenge sich das Komplement bezieht.
c) Das Komplement der Primzahlen $\mathbb{P}$ in der Grundmenge $\mathbb{N}$ besteht aus den zusammengesetzten Zahlen mit der Eins: $\mathbb{P}^C = \mathbb{N} \setminus \mathbb{P} = \{1, 4, 6, 8, 10, 12, 14, 15, \ldots\}$.
d) Das Komplement der positiven Zahlen $\mathbb{R}_+ = (0, \infty)$ in der Grundmenge $\mathbb{R}$ sind gerade die nichtpositiven Zahlen $(-\infty, 0]$; umkehrt sind die nichtnegativen Zahlen $(-\infty, 0)^C = \mathbb{R} \setminus (-\infty, 0) = [0, \infty)$ das Komplement der negativen Zahlen. Das Komplement eines beschränkten Intervalls ist kein Intervall mehr, sondern die Vereinigung von zwei Intervallen; zum Beispiel gilt

$$x \notin [2, 3) \Leftrightarrow \neg(2 \leqslant x < 3) \Leftrightarrow x < 2 \text{ oder } x \geqslant 3 \Leftrightarrow x \in (-\infty, 2) \cup [3, \infty)$$

und daher $[2, 3)^C = (-\infty, 2) \cup [3, \infty)$. Aus der Grundmenge von Gleichungen oder dem Definitionsbereich von Funktionen wird häufig genau eine Zahl ausgeschlossen:

$$\{1\}^C = \mathbb{R} \setminus \{1\} = (-\infty, 1) \cup (1, \infty).$$

e) Weitere Beispiele:

$$\begin{array}{lll} \mathbb{N}_0 \setminus \mathbb{N} = \{0\}, & \mathbb{Q} \setminus \mathbb{R} = \emptyset, & [-1, 1] \setminus \{0\} = [-1, 0) \cup (0, 1], \\ [-1, 0] \setminus \mathbb{Z} = (-1, 0), & [1, \infty) \setminus [1, 2] = (2, \infty), & (-\infty, 2) \setminus (-\infty, 1) = [1, 2). \end{array}$$

(2) An der Differenz $\{1, 2, 3\} \setminus \{1, 4\} = \{2, 3\}$ erkennt man auch, dass für $M \setminus N$ nur die Elemente von N relevant sind, die auch in M liegen. Man kann daher $M \setminus N$ stets als Komplement von $N \cap M$ in M auffassen: In der Grundmenge M ist

$$M \setminus N = M \setminus (N \cap M) = (N \cap M)^C.$$

(3) Die Aussage „$x \in N \cap N^C$“ bedeutet „$x \in N$ und $x \notin N$“, was nach dem Gesetz vom ausgeschlossenen Widerspruch niemals wahr ist. Daher ist $N \cap N^C = \emptyset$, eine Menge also immer disjunkt zu ihrem Komplement. Genauso gilt $N \cup N^C = M$ nach dem Gesetz vom ausgeschlossenen Dritten, die Vereinigung einer Menge mit ihrem Komplement ergibt also die ganze Grundmenge.[9]

[9] Ist M endlich, folgt mit (2.4) daraus übrigens sofort $\#M = \#N + \#N^C$ für jede Menge $N \subseteq M$.

Ganz ähnlich kann man auch andere Tautologien in die Mengenlehre übersetzen, indem man die Negation durch das Komplement ersetzt: Aus Satz 1.19(3) und (5) wird

$$\begin{aligned} &A \subseteq B \Leftrightarrow A^C \cap B = \emptyset, \\ &A \subseteq B \Leftrightarrow A \cap C \subseteq B \text{ und } A \cap C^C \subseteq B. \end{aligned}$$

In Satz 2.14 werden wir weitere allgemeine Regeln beweisen. Aber: Neben den Gesetzen von De Morgan gelten im Allgemeinen keine Distributivgesetze zwischen der Differenz $\setminus$ und dem Schnitt $\cap$ bzw. der Vereinigung $\cup$, stattdessen hat man

$$\begin{aligned} &(A \setminus B) \cap C = (A \cap C) \setminus B = A \cap (C \setminus B), \\ &(A \setminus B) \cup C = (A \cup C) \setminus (B \setminus C). \end{aligned}$$

Wie sehen Venn-Diagramme zu diesen beiden Formeln aus?

Satz 2.14. Es seien A, B Teilmengen einer Grundmenge M. Dann gilt:

(1) $(A^C)^C = A$.

(2) $A \setminus B = A \cap B^C$.

(3) $(A \cap B)^C = A^C \cup B^C$. (Erste **Regel von De Morgan**)

(4) $(A \cup B)^C = A^C \cap B^C$. (Zweite **Regel von De Morgan**)

(5) $A \subseteq B \Leftrightarrow B^C \subseteq A^C$. (**Kontraposition**)

(6) $A = B \Leftrightarrow A^C = B^C$.

Beweis Alle Aussagen sind wieder Übersetzungen aus der Aussagenlogik, zum Beispiel folgt (1) aus der Regel $\overline{\overline{A}} = A$ für die doppelte Negation, die Gesetze von De Morgan (3) und (4) aus den Aussagen (7) und (8) in Satz 1.17. Wir beweisen die übrigen Aussagen: (2) folgt aus der Äquivalenz

$$x \in A \setminus B \Leftrightarrow x \in A \text{ und } x \notin B \Leftrightarrow x \in A \text{ und } x \in B^C \Leftrightarrow x \in A \cap B^C,$$

die für alle $x \in M$ gilt. In (5) gilt nach Satz 1.19(6)

$$\begin{aligned} A \subseteq B &\Leftrightarrow \forall x \in M : (x \in A \Rightarrow x \in B) \Leftrightarrow \forall x \in M : (x \notin B \Rightarrow x \notin A) \\ &\Leftrightarrow \forall x \in M : \left(x \in B^C \Rightarrow x \in A^C\right) \Leftrightarrow B^C \subseteq A^C. \end{aligned}$$

Daraus folgt sofort (6), denn $A = B \Leftrightarrow A \subseteq B$ und $B \subseteq A$. □

Bemerkung. Ähnlich wie die Regeln in Satz 2.10 lassen sich auch die Regeln von De Morgan in Satz 2.14(3) und (4) auf mehrfache Durchschnitte und Vereinigungen verallgemeinern. Sind zum Beispiel $A_1, A_2, \ldots$ Teilmengen einer Grundmenge M, so gilt

$$\left(\bigcap_{n=1}^{\infty} A_n\right)^C = \bigcup_{n=1}^{\infty} A_n^C, \qquad \left(\bigcup_{n=1}^{\infty} A_n\right)^C = \bigcap_{n=1}^{\infty} A_n^C.$$

Definition 2.15. Seien M, N Mengen, $x \in M$ und $y \in N$. Mit dem Symbol (x, y) wird das **geordnete Paar** der Elemente x und y bezeichnet. Wir fassen also (x, y) als *ein* neues Element mit dem *Komponenten* x und y auf, dabei ist die Reihenfolge wichtig. Die Menge aller geordneten Paare heißt das **kartesische Produkt** $M \times N$ (lies: „M Kreuz N") von M und N:

$$M \times N := \{(x, y) : x \in M, y \in N\}.$$

Da es auf die Reihenfolge ankommt, ist im Allgemeinen $(x, y) \neq (y, x)$ und $M \times N \neq N \times M$. Entsprechend kann man **Tripel** (x, y, z), **Quadrupel** (w, x, y, z) oder allgemeiner n-**Tupel** $(x_1, x_2, \ldots, x_n)$ von Elementen aus Mengen $M_1, \ldots, M_n$ betrachten und definiert das kartesische Produkt als

$$M_1 \times \ldots \times M_n := \{(x_1, \ldots, x_n) : x_1 \in M_1, \ldots, x_n \in M_n\}.$$

Für das kartesische Produkt einer Menge mit sich selbst schreibt man $M^2 := M \times M$, $M^3 := M \times M \times M$ und für n Faktoren entsprechend $M^n := M \times \ldots \times M$.

Beispiel 2.16. (1) a) Das kartesische Produkt der Mengen $\{\square, \Delta\}$ und $\{1, 2, 3\}$ ist je nach Reihenfolge eine der Mengen

$$\{\square, \Delta\} \times \{1, 2, 3\} = \{(\square, 1), (\square, 2), (\square, 3), (\Delta, 1), (\Delta, 2), (\Delta, 3)\},$$
$$\{1, 2, 3\} \times \{\square, \Delta\} = \{(1, \square), (1, \Delta), (2, \square), (2, \Delta), (3, \square), (3, \Delta)\}.$$

Die beiden Mengen sind verschieden, enthalten aber beide 6 Elemente, sind also gleichmächtig. Allgemein hat das kartesische Produkt zweier (endlicher) Mengen A, B die Kardinalität $\#(A \times B) = \#A \cdot \#B$.[10]

b) Das bekannteste Beispiel eines kartesischen Produkts ist das Koordinatensystem bzw. die **Ebene**, die aus Punkten (x, y) mit $x, y \in \mathbb{R}$ besteht. Die Menge all dieser Punkte ist das kartesische Produkt

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

Die Zahlen x, y heißen die **Koordinaten** des Punktes (x, y), die x-Koordinate auch **Abszisse**, die y-Koordinate auch **Ordinate**. Häufig stellt man die Ebene in einem **kartesischen Koordinatensystem** dar, in dem sich die zwei **Koordinatenachsen** für x- und y-Koordinate senkrecht im **Ursprung** $(0, 0)$ schneiden und dabei vier **Quadranten** bilden; aber auch andere Darstellungen sind möglich. Ganz analog wird der dreidimensionale **Raum** beschreiben durch das kartesische Produkt

$$\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) : x, y, z \in \mathbb{R}\}$$

mit x-, y- und z-Koordinate, wo sich wieder drei paarweise senkrechte Koordinatenachsen im Ursprung $(0, 0, 0)$ schneiden und dabei acht Oktanten bilden. In der analytischen Geometrie unterschiedet man häufig nicht zwischen einem Punkt $P = (x, y) \in \mathbb{R}^2$ und dem zugehörigen Ortsvektor $\vec{P} = \binom{x}{y}$.

[10] „Die Kardinalität eines kartesischen Produktes ist das Produkt der Kardinalitäten."

c) So wie sich die Punkte $(1, 0)$ und $(0, 1)$ im Koordinatensystem unterschieden, ist bei jedem geordneten Paar die Reihenfolge wichtig. Das Symbol (a, b) ist daher zu unterscheiden von der Menge $\{a, b\} = \{b, a\}$, in der die Reihenfolge keine Rolle spielt. Für zwei Paare (a, b) und (c, d) (und analog für n-Tupel) gilt

$$(a, b) = (c, d) \Leftrightarrow a = c \text{ und } b = d.$$

Für kartesische Produkte gilt entsprechend

$$A \times B = C \times D \Leftrightarrow A = C \text{ und } B = D,$$
$$A \times B \subseteq C \times D \Leftrightarrow A \subseteq C \text{ und } B \subseteq D.$$

d) Die Menge $\mathbb{N}^2$ beschreibt ein *Gitter* von Punkten mit Abstand 1 im ersten Quadranten, die Menge

$$\mathbb{Z}^2 = \{(x, y) : x, y \in \mathbb{Z}\} = \{(0, 0), (1, 0), (-1, 0), (0, 1), (0, -1), (2, 0), \ldots\}$$

ein entsprechendes Gitter über dem gesamten Koordinatensystem. Für einen anderen Gitterabstand $a > 0$ verwendet man die folgende Menge:

$$\{(ax, ay) : x, y \in \mathbb{Z}\}.$$

Häufig schreibt man $a\mathbb{Z} := \{ax : x \in \mathbb{Z}\}$ und daher $(a\mathbb{Z})^2$ für dieses Gitter.

(2) Kartesische Produkte tauchen im Alltag überall dort auf, wo auch Koordinatensysteme oder Diagramme verwendet werden können: Zum Beispiel wird ein Schachbrett beschrieben durch die 64 Elemente des kartesischen Produkts

$$\{a, \ldots, h\} \times \{1, \ldots, 8\} = \{a1, a2, \ldots, h7, h8\}.$$

Werden Daten aufgenommen, so liegen n-Tupel vor: Die Verteilung der Temperatur in Karlsruhe besteht aus 5-Tupeln (t, x, y, z, T), wobei t den Messzeitpunkt, (x, y, z) die Koordinaten des Messortes und T die gemessene Temperatur angibt. Auch beim Modellieren von Zufallsexperimenten nutzt man kartesische Produkte: Während $\{1, \ldots, 6\}$ den Ergebnisraum eines einzelnen Wurfs mit einem W6 beschreibt, ist

$$\{1, \ldots, 6\}^{100} = \{(x_1, \ldots, x_{100}) : x_1, \ldots, x_{100} \in \{1, \ldots, 6\}\}$$

der Ergebnisraum von 100 solchen Würfen. Hierfür schreibt man auch

$\{(x_1, \ldots, x_{100}) : x_i \in \{1, \ldots, 6\} \text{ für } i \in \{1, \ldots, 100\}\}$ oder etwas unsauber
$\{(x_1, \ldots, x_{100}) : x_i \in \{1, \ldots, 6\} \text{ für } i = 1, \ldots, 100\}$.

(3) Das kartesische Produkt von zwei oder drei Intervallen kann man sich als Rechteck in der Ebene bzw. als Quader im Raum vorstellen. Zu Beispiel bezeichnet

$$[0,1]^3 = [0,1] \times [0,1] \times [0,1] = \{(x,y,z) \in \mathbb{R}^3 : 0 \leqslant x,y,z \leqslant 1\} \subseteq \mathbb{R}^3$$

den Einheitswürfel im $\mathbb{R}^3$, also den Würfel mit Kantenlänge 1 und den Eckpunkten

$$\{0,1\}^3 = \{(0,0,0),(0,0,1),(0,1,0),(0,1,1),(1,0,0),(1,0,1),(1,1,0),(1,1,1)\}.$$

Demgegenüber fehlen dem Würfel $[0,1)^3$ einige Seitenflächen, nämlich welche? Der Flächeninhalt eines Rechtecks wie $[0,3] \times [-1,1] \subseteq \mathbb{R}^2$ ist gleich dem Produkt der Seiten- bzw. Intervalllängen $3 \cdot 2 = 6$; Analoges gilt für das Volumen eines Quaders.

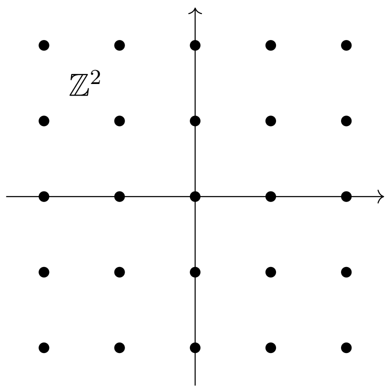
(4) Mithilfe zusätzlicher Bedingungen kann man auch komplexe Teilmengen der Ebene darstellen: Zum Beispiel ist

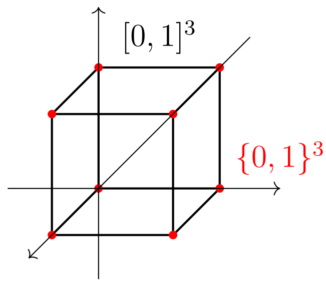
$$\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$

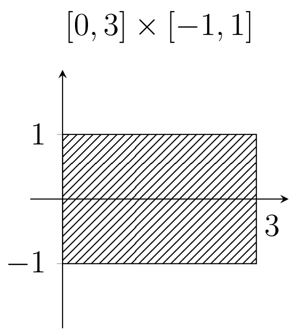
der schon in (2.3) diskutierte Einheitskreis im $\mathbb{R}^2$ mit Radius 1 und Mittelpunkt $(0,0)$. Die Menge

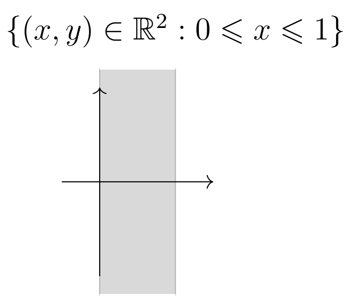
$$\{(x,y) \in \mathbb{R}^2 : y = x^2\}$$

beschreibt die Normalparabel, also den Graphen der Funktion $f(x) = x^2$, die Menge $\{(x,y) \in \mathbb{R}^2 : 0 \leqslant x \leqslant 1\}$ ist ein Streifen parallel zur y-Achse.









3 Ungleichungen

3.1 Allgemeingültige Ungleichungen

Definition 3.1. Alle reellen Zahlen lassen sich ihrer Größe nach vergleichen. Dafür nutzt man die folgenden Ungleichheitszeichen:

$$\begin{aligned}
a \leqslant b &: a \text{ ist kleiner oder gleich } b.\\
a \geqslant b &: a \text{ ist größer oder gleich } b.\\
a < b &: a \text{ ist (echt) kleiner } b.\\
a > b &: a \text{ ist (echt) größer } b.
\end{aligned}$$

Mithilfe des Gleichheitszeichens = lassen sich Aussagen wie $1 = 1$ oder Aussageformen (**Gleichungen**) wie $3x - 4 = 2$ formulieren, genauso entstehen mithilfe der Ungleichheitszeichen Aussagen wie $1 < 2$ oder Aussageformen (**Ungleichungen**) wie $3x-4 \leqslant 2$. Der Umgang mit Gleichungen und Ungleichungen ist ähnlich, aber nicht gleich, denn die grundlegenden Eigenschaften von = und $\leqslant$ unterscheiden sich: Für alle $a, b, c \in \mathbb{R}$ gilt

$a = b \Rightarrow b = a$. (**Symmetrie**)	$a \leqslant b$ und $b \leqslant a \Rightarrow a = b$. (**Antisymmetrie**)
$a = b$ und $b = c \Rightarrow a = c$.	$a \leqslant b$ und $b \leqslant c \Rightarrow a \leqslant c$. (**Transitivität**)
Stets ist $a = a$.	Stets ist $a \leqslant a$. (**Reflexivität**)
	Stets ist $a \leqslant b$ oder $b \leqslant a$. (**Totalität**)
$a = b \Rightarrow a + c = b + c$.	$a \leqslant b \Rightarrow a + c \leqslant b + c$. (**Verträglichkeit** mit +)
$a = b \Rightarrow ac = bc$.	$a \leqslant b$ und $0 \leqslant c \Rightarrow ac \leqslant bc$. (**Verträglichkeit** mit $\cdot$)

Sowohl = als auch $\leqslant$ sind reflexiv und transitiv, aber Symmetrie und Antisymmetrie sind sehr verschiedene Eigenschaften; Totalität für = ist nicht sinnvoll. Auch die Verträglichkeit mit den Grundrechenarten + und $\cdot$ (damit auch : und $-$) ist ähnlich, bei $\leqslant$ ist aber nur eine Multiplikation mit Zahlen $c \geqslant 0$ erlaubt.
Aus den Eigenschaften von = auf der linken Seite lassen sich alle Regeln im Umgang mit Gleichungen herleiten, genauso folgen aus den Eigenschaften von $\leqslant$ auf der rechten Seite alle weiteren Regeln für Ungleichungen, vgl. Satz 3.3. Wir wollen diese **Anordnungsaxiome** nicht beweisen[1], sondern als Ausgangspunkt der weiteren Argumentation verwenden.

[1]So wie wir auch Assoziativ-, Distributiv- und Kommutativgesetze für reelle Zahlen nicht beweisen.

Beispiel 3.2. (1) Auf der Zahlengeraden bedeutet $a \leqslant b$ anschaulich, dass a links von b bzw. b rechts von a liegt. Dabei ist auch $a = b$ möglich; dieser Fall wird mit der Ungleichung $a < b$ ausgeschlossen. Möchte man $a \neq b$ betonen, so liest man das Symbol $<$ als „echt kleiner“, sonst als „kleiner“, das Symbol $\leqslant$ liest man als „kleiner-gleich“; entsprechendes gilt für $>$ und $\geqslant$.
Auch die Anordnungsaxiome lassen sich an der Zahlengeraden interpretieren: Liegt zum Beispiel a links von b und b links von c, so muss auch a links von c liegen (Transitivität). Diese anschauliche Argumentation und Vorstellung von großen und kleinen Zahlen muss in Beweisen sauber getrennt werden von einer deduktiven Argumentation anhand der Anordnungsaxiome.
Zahlen $a > 0$ liegen (echt) rechts der Null und heißen **positiv**, Zahlen $a < 0$ liegen (echt) links der Null und heißen **negativ**. Die Negation dessen sind die **nichtpositiven** $a \leqslant 0$ und die **nichtnegativen** $a \geqslant 0$ Zahlen. Beachte: *Alle* negativen Zahlen sind kleiner als *alle* positiven Zahlen, zum Beispiel ist $-2 < 1$ und $-2 < -\frac{1}{2}$, auch wenn man aufgrund des Betrages (der „Größe“) dieser Zahlen etwas anderes vermuten würde.

(2) Das Symbol $\leqslant$ ist eine Kombination der Symbole $<$ und $=$, denn $a \leqslant b$ bedeutet „$a < b$ oder $a = b$“; entsprechend bedeutet $a \geqslant b$ gerade „$a > b$ oder $a = b$“. Die Bedingung $a \leqslant b$ ist also eine *schwächere* Bedingung als $a < b$, aber auch eine schwächere Bedingung als $a = b$.
Umgekehrt hat $a < b$ die Bedeutung „$a \leqslant b$ und $a \neq b$“, und $a > b$ bedeutet „$a \geqslant b$ und $a \neq b$“. Die Bedingung $a < b$ ist also eine *stärkere* Bedingung als $a \leqslant b$, aber auch eine stärkere Bedingung als $a \neq b$. Ein Diagramm für die Hierarchie zwischen den Symbolen $=, \neq, \leqslant, \geqslant, <, >$ sieht so aus:

$$\begin{array}{ccccc} \leqslant & \Leftarrow & = & \Rightarrow & \geqslant \\ \Uparrow & & & & \Uparrow \\ < & \Rightarrow & \neq & \Leftarrow & > \end{array}$$

Bei der Negation der Aussage $a \leqslant b$ entsteht die Aussage $a > b$, entsprechend ist $a \geqslant b$ die Negation der Aussage $a < b$. In (1.7) und den Übungen haben wir gesehen, dass für Zahlen $x, y \geqslant 0$ die Äquivalenz $x \leqslant y \Leftrightarrow x^2 \leqslant y^2$ richtig ist. Durch Kontraposition (Satz 1.19(7)) erhält man daraus die Aussage

$$\forall x, y \geqslant 0 : x < y \Leftrightarrow x^2 < y^2.$$

Vollziehen Sie die folgenden Beispiele aufmerksam nach:

- Aus $1 < 2$ folgt $1 \leqslant 2$.
- $1 > -1$ ist eine stärkere Aussage als $1 \geqslant -1$.
- Die Aussage $-1 > 0$ ist falsch, daher ist $-1 \leqslant 0$ richtig.
- Aus $-3 \leqslant 3$ und $-3 \neq 3$ folgt $-3 < 3$.
- Es gilt $5 = 5$ und daher sowohl $5 \leqslant 5$ als auch $5 \geqslant 5$.
- Die Aussage $-1 \leqslant -2$ ist falsch, denn es gilt weder $-1 < -2$ noch $-1 = -2$.

- Die Aussage $1 < 1$ ist falsch, aber die Aussage $1 \leqslant 1$ ist richtig.
- Aus $0 \leqslant 2$ und $0 < 2$ folgt $0 < 2$.
- Die Aussage $0 \leqslant 0$ ist richtig, denn $0 = 0$.
- Aus $0 < 4$ folgt $0 \neq 4$.
- Die Aussage $7 \leqslant 5$ ist falsch, daher ist $7 > 5$ richtig.
- Aus $-6 \leqslant -6$ und $-6 \geqslant -6$ folgt $-6 = -6$.

(3) Bei Termumformungen sind wir gewohnt, Gleichheitszeichen aneinanderzuhängen:

$$\frac{2-\sqrt{2}}{\sqrt{2}+1} = \frac{(2-\sqrt{2})(\sqrt{2}-1)}{(\sqrt{2}+1)(\sqrt{2}-1)} = \frac{3\sqrt{2}-2-2}{2-1} = 3\sqrt{2}-4,$$

mit der **Transitivität** folgt daraus $\frac{2-\sqrt{2}}{\sqrt{2}+1} = 3\sqrt{2}-4$. Aufgrund der **Symmetrie** können wir diese Gleichung auch von rechts nach links lesen, also die Differenz als Quotienten darstellen. Genauso kann man Ungleichungen aneinanderhängen: Aus der Ungleichungskette[2]

$$3\sqrt{2}-4 \geqslant 3 \cdot \frac{7}{5} - 4 = \frac{1}{5} \geqslant 0$$

folgt aufgrund der Transitivität $3\sqrt{2}-4 \geqslant 0$, die Zahl $3\sqrt{2}-4$ ist also nichtnegativ. Da $a \leqslant b$ aus $a = b$ folgt, können auch Gleichungen mit Ungleichungen kombiniert werden:

$$\frac{2-\sqrt{2}}{\sqrt{2}+1} = 3\sqrt{2}-4 \leqslant 3 \cdot 2 - 4 = 2,$$

also $\frac{2-\sqrt{2}}{\sqrt{2}+1} \leqslant 2$. Da hier die Ungleichung $2 \leqslant \frac{2-\sqrt{2}}{1+\sqrt{2}}$ eine falsche Aussage ist (Warum?), gilt sogar $\frac{2-\sqrt{2}}{\sqrt{2}+1} < 2$.

(4) Die **Antisymmetrie** von $\leqslant$ drückt aus, dass im Gegensatz zu $a = b$ die Seiten einer Ungleichung nur dann vertauscht werden dürfen, wenn eigentlich eine Gleichung vorliegt; $=$ ist also ein ungerichtetes und $\leqslant$ ein gerichtetes Symbol. Gilt für zwei Zahlen

$$a \leqslant b \leqslant a,$$

so folgt aus der Antisymmetrie $a = b$. Auf diese Weise kann man die Gleichheit $a = b$ in die zwei Ungleichungen $a \leqslant b$ und $b \leqslant a$ zerlegen, vgl. dieses Prinzip mit Satz 1.19(6) und Beispiel 2.4(3).

(5) Auch an der Verträglichkeit mit der Multiplikation erkennt man, dass $\leqslant$ eine ausgezeichnete Richtung besitzt: Während alle Gleichungen mit allen Zahlen multipliziert werden dürfen, dürfen Ungleichungen nur mit nichtnegativen Zahlen $c \geqslant 0$ multipliziert werden. Beispiel: Aus $1 = 1$ folgt durch Multiplikation mit -1 die Gleichung $-1 = -1$, aber

$$1 \leqslant 2 \text{ ist wahr,} \qquad -1 \leqslant -2 \text{ ist falsch.}$$

[2] Die Ungleichung $\sqrt{2} \geqslant \frac{7}{5}$ folgt aus der Ungleichung $(5\sqrt{2})^2 = 50 \geqslant 49 = 7^2$.

(6) Reelle Zahlen weisen die Besonderheit auf, dass sie sich immer vergleichen lassen (**Totalität**), also eine der Ungleichungen $a \leqslant b$ oder $b \leqslant a$ immer wahr ist.[3] Man kann daher die Ungleichung $a \leqslant b$ beweisen, indem man die Ungleichung $b \leqslant a$ zum Widerspruch führt. Dazu ein kleines Beispiel: Für eine Zahl $a \geqslant 0$ gelte

$$\forall \varepsilon > 0 : a \leqslant \varepsilon.$$

Behauptung: Dann ist $a = 0$. Beweis Angenommen, es wäre $a > 0$. Dann gilt $\frac{a}{2} > 0$ und $2a > a$, also $a > \frac{a}{2}$. Setzt man $\varepsilon = \frac{a}{2}$ in der Voraussetzung, so erhält man aber die Aussage $a \leqslant \frac{a}{2}$. $\lightning$ Widerspruch! Also ist $a \leqslant 0$ (Totalität), zusammen mit $a \geqslant 0$ folgt daraus $a = 0$ (Antisymmetrie). $\square$

(7) Die sechs Anordnungsaxiome gelten sinngemäß, wenn man alle $\leqslant$ durch $\geqslant$ ersetzt, außer bei der Verträglichkeit mit der Multiplikation: Hier muss weiterhin $c \geqslant 0$ sein, damit die Ungleichung erhalten bleibt. Die Anordnungsaxiome gelten *nicht* für die Symbole $<$ und $>$, denn hier sind Antisymmetrie, Reflexivität und Totalität verletzt; die übrigen Eigenschaften werden in Satz 3.3 bewiesen.

Satz 3.3. Seien $a, b, c \in \mathbb{R}$. Es gelten die folgenden Rechenregeln:

(1) Stets ist entweder $a < b$ oder $a = b$ oder $a > b$. (**Totalität**)

(2) $a < b$ und $b \leqslant c \Rightarrow a < c$. Analog $a \leqslant b$ und $b < c \Rightarrow a < c$. (**Transitivität**)

(3) $a < b \Rightarrow a+c < b+c$. Entsprechend $a < b$ und $0 < c \Rightarrow ac < bc$. (**Verträglichkeit**)

(4) $a \leqslant b$ und $c \leqslant 0 \Rightarrow ac \geqslant bc$. Analog $a < b$ und $c < 0 \Rightarrow ac > bc$.

(5) $0 < a \leqslant b \Rightarrow 0 < \frac{1}{b} \leqslant \frac{1}{a}$. Analog $a \leqslant b < 0 \Rightarrow \frac{1}{b} \leqslant \frac{1}{a} < 0$.

(6) Sind $a, b \geqslant 0$, so gilt $a \leqslant b \Leftrightarrow a^2 \leqslant b^2 \Leftrightarrow \sqrt{a} \leqslant \sqrt{b}$.[4]

Merkregel für (4) und (5): Beim Multiplizieren mit einer negativen Zahl oder beim Bilden des Kehrwertes dreht sich das Ungleichheitszeichen um.
Beweis Der Beweis dieser Rechenregeln ist ein gutes Beispiel für deduktives Schließen: Unsere Intuition der Zahlen ist *kein* gültiges Argument, sondern wir verwenden nur die Anordnungsaxiome und daraus hergeleitete Aussagen.

(1) Nach der Totalität ist die Aussage „$a \leqslant b$ oder $b \leqslant a$“ stets wahr, nach Definition 1.10 gibt es dafür genau drei Möglichkeiten:

Fall 1: $a \leqslant b$ ist wahr und $a \geqslant b$ ist wahr.
Fall 2: $a \leqslant b$ ist wahr und $a \geqslant b$ ist falsch.
Fall 3: $a \leqslant b$ ist falsch und $a \geqslant b$ ist wahr.

Im ersten Fall folgt $a = b$ nach der Antisymmetrie; im zweiten Fall folgt $a \leqslant b$ und $a < b$, also $a < b$; im dritten Fall folgt $a > b$ und $a \geqslant b$, also $a > b$. Diese drei Fälle schließen sich gegenseitig aus, da keine der Teilaussagen zugleich wahr und falsch sein kann.

[3] Die Inklusion $\subseteq$ verhält sich ähnlich wie $\leqslant$, aber es gilt weder $\{1\} \subseteq \{2\}$ noch $\{2\} \subseteq \{1\}$.
[4] Mit etwas mehr Aufwand kann man für alle reellen Zahlen $x > 0$ zeigen, dass $a \leqslant b \Leftrightarrow a^x \leqslant b^x$.

(2) Aus $a < b$ folgt $a \leqslant b$, zusammen mit $b \leqslant c$ und der Transitivität also $a \leqslant c$. Wäre nun $a = c$, so wäre $a < b$ und $b \leqslant a$, was nach (1) unmöglich ist. Also ist $a < c$; genauso zeigt man auch die zweite Aussage.
(3) Aus der Verträglichkeit von $\leqslant$ folgt $a+c \leqslant b+c$ bzw. $ac \leqslant bc$. Wäre nun $a+c = b+c$, so folgte $a = b$ im Widerspruch zu $a < b$, also ist $a+c < b+c$. Genauso folgt aus $ac = bc$ auch $a = b$ (beachte $c \neq 0$) im Widerspruch zu $a < b$, also ist $ac < bc$.
(4) Addiert man auf beiden Seiten der Ungleichung $c \leqslant 0$ die Zahl $-c$, so erhält man $0 \leqslant -c$. Mit der Verträglichkeit folgt daraus $a(-c) \leqslant b(-c)$, also $-ac \leqslant -bc$. Addiert man wieder $ac+bc$ auf beiden Seiten, so folgt wie behauptet $bc \leqslant ac$. Die zweite Aussage ist eine Übungsaufgabe.
(5) Aus $a > 0$ folgt $\frac{1}{a} > 0$, denn sonst wäre $\frac{1}{a} \leqslant 0$ und die Multiplikation mit $a > 0$ ergäbe den Widerspruch $1 \leqslant 0$. Also ist auch $\frac{1}{b} > 0$ und $\frac{1}{ab} > 0$ und wir können die Ungleichung $a \leqslant b$ mit $\frac{1}{ab}$ multiplizieren, woraus wir $\frac{1}{b} = \frac{a}{ab} \leqslant \frac{b}{ab} = \frac{1}{a}$ erhalten. Die zweite Aussage ist wieder eine Übungsaufgabe.
(6) siehe Beispiel 3.2(2). □

Beispiel 3.4. (1) Nach Satz 3.3(2) genügt in einer Ungleichungskette ein einziges $<$ oder $>$, um die Gleichheit auszuschließen: Im obigen Beispiel

$$3\sqrt{2} - 4 \geqslant 3 \cdot \frac{7}{5} - 4 = \frac{1}{5} > 0$$

ist also $3\sqrt{2} - 4 > 0$ positiv, insbesondere $3\sqrt{2} - 4 \neq 0$. Oft will man bei solchen **Abschätzungen** die Größe eines Terms mit einem anderen vergleichen; mit $\leqslant$ macht man den Term größer und spricht von einer Abschätzung *nach oben*, entsprechend heißt $\geqslant$ eine Abschätzung *nach unten*. Als Beispiel beweisen wir die Abschätzung $4\sqrt{3} - 7 < 0$: Wir schätzen nach oben ab, etwa durch Einsetzen von $\sqrt{3} \leqslant 2$:

$$4\sqrt{3} - 7 \leqslant 4 \cdot 2 - 7 = -1.$$

Hoppla! Unsere Abschätzung ist „nicht scharf genug“, statt der gewünschten Aussage haben wir nur die schwächere Aussage $4\sqrt{3} - 7 \leqslant -1$ bewiesen. Wir können nun nicht (wie bei $=$) unsere Rechenschritte rückgängig machen: Aus -1 lässt sich der Term $4\sqrt{3} - 7$ nicht zurückgewinnen.[5] Wir starten also einen neuen Versuch von vorne: Aus der Ungleichung $(4\sqrt{3})^2 = 48 < 49 = 7^2$ folgt $\sqrt{3} < \frac{7}{4}$ (siehe Satz 3.3(6) und (3)) und daraus wie gewünscht

$$4\sqrt{3} - 7 < 4 \cdot \frac{7}{4} - 7 = 0. \ \square$$

Natürlich ist in der Abschätzung $\sqrt{3} < \frac{7}{4}$ der Bruch $\frac{7}{4}$ so gewählt, dass sich auf der rechten Seite 0 ergibt; dieses „Rückwärtsdenken“ ist bei vielen Abschätzungen eine hilfreiche Strategie.

[5] Abschätzungen sind also nur Implikationen, während Termumformungen Äquivalenzen sind.

(2) Die Verträglichkeit mit Addition und Multiplikation erlaubt es, beim Abschätzen Gleichungen und Ungleichungen ineinander einzusetzen; dabei sind aber negative Zahlen, Brüche und Quadrate eine häufige Fehlerquelle. Es gelte zum Beispiel $a \leqslant b$ und $c \leqslant d$ für gewisse Zahlen; ist dann die Ungleichung $\frac{a}{d} \leqslant \frac{b}{c}$ korrekt? Im Allgemeinen nicht, die Vorzeichen aller beteiligten Variablen sind entscheidend:

$$\begin{array}{ccc} \dfrac{2}{7} \leqslant \dfrac{3}{6}, & \dfrac{-2}{7} \leqslant \dfrac{3}{6}, & \dfrac{-3}{7} \leqslant \dfrac{-2}{6}, \\ \dfrac{2}{7} \geqslant \dfrac{3}{-6}, & \dfrac{-2}{7} \geqslant \dfrac{3}{-6}, & \dfrac{-3}{7} \leqslant \dfrac{-2}{-6}, \\ \dfrac{2}{-6} \geqslant \dfrac{3}{-7}, & \dfrac{-2}{-6} \geqslant \dfrac{3}{-7}, & \dfrac{-3}{-6} \geqslant \dfrac{-2}{-7}. \end{array}$$

Weitere Beispiele mit Brüchen und Quadraten:

$$\begin{array}{ll} \text{Aus } 6 \geqslant 4 \text{ folgt } -\dfrac{6}{7} \leqslant -\dfrac{4}{7}, & \text{es gilt} -2 \leqslant 1, \text{ aber } (-2)^2 \geqslant 1^2, \\ \text{aus } \dfrac{1}{2} \leqslant 3 \text{ folgt } 2 \geqslant \dfrac{1}{3}, & \text{aus } -\dfrac{1}{2} \leqslant 3 \text{ folgt } -2 \leqslant \dfrac{1}{3}. \end{array}$$

Daher sollte man beim Abschätzen in komplexen Termen die Verschachtelung sorgfältig von innen nach außen aufbauen: Um zum Beispiel die Ungleichung $x \leqslant -1$ im Term $\frac{1}{1-x}$ zu verwenden, denkt man

$$x \leqslant -1 \Rightarrow -x \geqslant 1 \Rightarrow 1 - x \geqslant 2 \Rightarrow \frac{1}{1-x} \leqslant \frac{1}{2}.$$

(3) Es sei noch angemerkt, dass die Kombination verschieden gerichteter Ungleichungen nicht zum Ziel führt: Zum Beispiel lässt sich aus der Ungleichung $a \leqslant 1 \geqslant b$ nichts über das Verhältnis von a und b folgern, es kann sowohl $-1 \leqslant 1 \geqslant 0$ als auch $0 \leqslant 1 \geqslant -1$ sein.

Wir wollen nun die Bedeutung und Verwendung der Regeln in Satz 3.3 an einigen einfachen allgemeingültigen Ungleichungen illustrieren.

Satz 3.5. Seien $a, b, c, d \in \mathbb{R}$. Es gelten die folgenden elementaren Ungleichungen:

(1) Ist $a < b$, so liegt dazwischen das *arithmetische Mittel* $a < \frac{a+b}{2} < b$.

(2) Stets ist $a^2 \geqslant 0$. Dabei ist $a^2 = 0 \Leftrightarrow a = 0$.

(3) Stets ist $2ab \leqslant a^2 + b^2$ und daher $4ab \leqslant (a+b)^2 \leqslant 2(a^2 + b^2)$.

(4) Es gilt die **Cauchy-Schwarz-Ungleichung** $(ab + cd)^2 \leqslant (a^2 + c^2)(b^2 + d^2)$. Allgemeiner gilt für Zahlen $a_1, \ldots, a_n, b_1, \ldots b_n \in \mathbb{R}$:

$$(a_1 b_1 + \ldots + a_n b_n)^2 \leqslant (a_1^2 + \ldots + a_n^2)(b_1^2 + \ldots + b_n^2).$$

Beweis (1) Wir addieren a auf beiden Seiten von $a < b$ und erhalten $2a < a+b$, genauso addieren wir b und erhalten $a + b < 2b$ (Satz 3.3(3)); aus der Transitivität (Satz 3.3(2)) folgt $2a < a + b < 2b$. Multiplikation mit $\frac{1}{2} > 0$ (Satz 3.3(3)) ergibt $a < \frac{a+b}{2} < b$.
(2) Nach der Totalität ist $a \geqslant 0$ oder $0 \geqslant a$. Im ersten Fall multiplizieren wir die Ungleichung mit $a \geqslant 0$ und erhalten $a^2 \geqslant 0$; im zweiten Fall multiplizieren wir die Ungleichung mit $0 \geqslant a$ und erhalten $0 \leqslant a^2$ nach Satz 3.3(4). Damit ist die erste Aussage gezeigt; in der zweiten Aussage ist „$\Leftarrow$“ klar. Für die Richtung „$\Rightarrow$“ sei $a^2 = 0$, wir wollen $a = 0$ zeigen. Dazu betrachten wir die Aussage

$$a > 0 \text{ oder } a < 0 \Rightarrow a^2 > 0.$$

Diese Aussage beweist man wie die Aussage $a^2 \geqslant 0$ mithilfe von Satz 3.3(3) und (4). Damit ist auch ihre Kontraposition richtig:

$$a^2 \leqslant 0 \Rightarrow \neg(a > 0) \text{ und } \neg(a < 0).$$

Da $a^2 = 0$, ist die linke Aussage erfüllt und es folgt die rechte Aussage. Nach Satz 3.3(1) muss dann $a = 0$ sein.
(3) Nach (2) ist $0 \leqslant (a - b)^2$, Ausmultiplizieren und Addition von $2ab$ auf beiden Seiten ergibt $2ab \leqslant a^2 + b^2$. Addiert man nochmal $2ab$, so erhält man $4ab \leqslant a^2 + b^2 + 2ab = (a + b)^2$; nun setzt man $2ab \leqslant a^2 + b^2$ hier ein und erhält $(a + b)^2 \leqslant 2a^2 + 2b^2$.
(4) In der ersten Ungleichung von (3) ersetzen wir a durch ad und b durch bc, es ergibt sich $2abcd \leqslant a^2d^2 + b^2c^2$. Für die Cauchy-Schwarz-Ungleichung addieren wir auf beiden Seiten den fehlenden Term $a^2b^2 + c^2d^2$:

$$(ab + cd)^2 = 2abcd + a^2b^2 + c^2d^2 \leqslant a^2d^2 + b^2c^2 + a^2b^2 + c^2d^2 = (a^2 + c^2)(b^2 + d^2).$$

Im allgemeinen Fall verläuft der Beweis analog, man verwendet für die Mischterme auf der linken Seite die Ungleichung $2a_ib_ia_jb_j \leqslant a_i^2b_j^2 + b_i^2a_j^2$. □

Beispiel 3.6. (1) Satz 3.5(1) besagt, dass zwischen zwei reellen Zahlen stets eine weitere liegt, zwischen zwei rationalen Zahlen sogar eine weitere rationale Zahl; man kann also beliebig nah an die Zahlengerade heranzoomen. Tatsächlich gilt auch

$$\forall a, b \in \mathbb{R} : (a < b \Rightarrow \exists r \in \mathbb{Q} : a < r < b)\,,$$

man sagt dazu: Die rationalen Zahlen $\mathbb{Q}$ liegen **dicht** in den reellen Zahlen $\mathbb{R}$.

(2) Mithilfe von Satz 3.5(2) wird häufig so argumentiert: Gilt für gewisse Zahlen $a, b \in \mathbb{R}$ die Gleichung

$$a^2 + b^2 = 0,$$

so muss $a = b = 0$ sein. Denn: Eine Summe von Zahlen $\geqslant 0$ kann nur dann Null sein, wenn jede einzelne Zahl Null ist; und aus $a^2 = b^2 = 0$ folgt $a = b = 0$.

(3) Mit Satz 3.5(3) können Produkte durch Summen abgeschätzt werden, dabei kann das Produkt ab auch negativ sein. Ein Beispiel: $(7 \cdot 81 = 567)$

$$7 \cdot 81 \leqslant \frac{1}{2}\left(7^2 + 81^2\right) = 3305, \qquad 7 \cdot 81 \leqslant \frac{1}{4}(7+81)^2 = 1936.$$

Für die Cauchy-Schwarz-Ungleichung in Satz 3.5(4) kann man ähnliche Zahlenbeispiele machen, man kann sie aber auch geometrisch verstehen: Nach Wurzelziehen interpretiert man $ab + cd$ als das Skalarprodukt $\vec{x} \cdot \vec{y}$ zweier Vektoren $\vec{x} = \binom{a}{c}$ und $\vec{y} = \binom{b}{d}$. Auf der rechten Seite steht dann das Produkt der Längen $\|\vec{x}\| = \sqrt{a^2 + c^2}$ und $\|\vec{y}\| = \sqrt{b^2 + d^2}$, die Cauchy-Schwarz-Ungleichung hat also die Form

$$|\vec{x} \cdot \vec{y}| \leqslant \|\vec{x}\| \cdot \|\vec{y}\|.$$

Das bedeutet für den Cosinus des Winkels zwischen $\vec{x}$ und $\vec{y}$ gerade

$$|\cos(\sphericalangle(\vec{x}, \vec{y}))| = \frac{|\vec{x} \cdot \vec{y}|}{\|\vec{x}\| \cdot \|\vec{y}\|} \leqslant 1.$$

Übrigens gilt Gleichheit in der Cauchy-Schwarz-Ungleichung genau dann, wenn die Vektoren $\vec{x}$ und $\vec{y}$ linear abhängig sind, denn dann ist $|\cos(\sphericalangle(\vec{x}, \vec{y}))| = 1$ und der Winkel $0°$ oder $180°$ groß.

Definition 3.7. Sei $a \in \mathbb{R}$. Der **Betrag** von a ist wie folgt definiert:

$$|a| := \begin{cases} a, & \text{falls } a \geqslant 0, \\ -a, & \text{falls } a < 0. \end{cases}$$

Der Betrag lässt positive Zahlen unverändert und macht negative Zahlen positiv; man beachte, dass $-a$ im zweiten Fall positiv ist. Mit der Vorzeichenfunktion **Signum**

$$\operatorname{sign}(a) := \begin{cases} 1, & \text{falls } a > 0, \\ 0, & \text{falls } a = 0, \\ -1, & \text{falls } a < 0, \end{cases}$$

kann man auch schreiben $|a| = a \operatorname{sign}(a)$; der Betrag „verdoppelt“ also das Vorzeichen, sodass negative Vorzeichen verschwinden.

Beispiel 3.8. (1) Es ist $|2| = 2$ und $|0| = 0$ nach der ersten Zeile in Definition 3.7, $|-1| = -(-1) = 1$ und $|-\frac{1}{2}| = \frac{1}{2}$ nach der zweiten Zeile. Genauso gilt

$$\operatorname{sign}(\pi) = 1, \qquad \operatorname{sign}(0) = 0, \qquad \operatorname{sign}(-2) = -1$$

und zum Beispiel $|-2| = -2 \cdot \operatorname{sign}(-2) = -2 \cdot (-1) = 2$.

Man erkennt: Der Betrag ist stets nichtnegativ, also $|a| \geqslant 0$; daraus folgt sofort $||a|| = |a|$. Die Zahlen a und $-a$ unterscheiden sich nur um ein Vorzeichen, haben also denselben Betrag $|-a| = |a|$; die Schreibweise $|a| = \pm a$ kann das andeuten. Beachte: Es gilt $-2 \leqslant 1$, aber $|-2| = 2 > 1 = |1|$, der Betrag „erhält“ Ungleichungen also *nicht*; aus $a \leqslant b$ kann sowohl $|a| \leqslant |b|$ als auch $|a| \geqslant |b|$ folgen.[6]

(2) Auf der Zahlengeraden ist rechts der Null einfach $|a| = a$, Zahlen $a < 0$ links der Null werden auf die positive Seite $|a| = -a$ gespiegelt. Daher ist stets $a \leqslant |a|$ und $-a \leqslant |a|$. Der Betrag $|a|$ gibt den (gemessenen) Abstand der Zahl a vom Ursprung 0 an; entsprechend ist $|a-b|$ der Abstand der Zahlen a und b auf der Zahlengeraden. Für $a \geqslant b$ ist dieser Abstand einfach die Differenz $a - b$, für $a \leqslant b$ die Differenz $-(a-b) = b-a$ in der gespiegelten Situation. Wegen $|a-b| = |-(b-a)| = |b-a|$ ist natürlich der Abstand von a zu b derselbe Abstand wie von b zu a. Beispiel:

$$|3-1| = |2| = 2, \qquad |-3-(-1)| = |1-3| = |-2| = 2$$

ist sowohl der Abstand der Zahlen 3 und 1 als auch der Zahlen -3 und -1;

$$|-3-1| = |-4| = 4, \qquad |3-(-1)| = |4| = 4$$

ist sowohl der Abstand der Zahlen -3 und 1 als auch der Zahlen 3 und -1.

(3) Mit dieser Anschauung lassen sich einige Gleichungen mit Beträgen sehr einfach lösen: Zum Beispiel hat $|x| = 2$ die Lösungsmenge $\mathcal{L} = \{-2, 2\}$, denn genau diese beiden Zahlen haben den Abstand 2 vom Ursprung. Die Lösungen der Gleichungen

$$|x-3| = 2, \qquad |x-3| = -2,$$

haben den Abstand 2 bzw. -2 von der Zahl 3; im ersten Fall ist $\mathcal{L} = \{1, 5\}$, im zweiten Fall $\mathcal{L} = \emptyset$, da Abstände immer nichtnegativ sind. Die Gleichung

$$|x-1| = |x+3|$$

sucht nach Zahlen, die den gleichen Abstand von 1 und -3 haben; das erfüllt nur der Mittelwert $x = (1-3)/2 = -1$. Im nächsten Abschnitt werden wir weitere systematische Lösungsmethoden kennenlernen, die aber aufwändiger sind.

(4) Es gibt einen wichtigen Zusammenhang zwischen Betrag, Wurzel und Quadraten: Sei $a \geqslant 0$. Die Zahl $\sqrt{a}$ ist (nach Definition) die nichtnegative Lösung der Gleichung $x^2 = a$; aus den beiden Lösungen $2, -2$ von $x^2 = 4$ wird also $\sqrt{4} = 2$ ausgewählt.[7] Nach Definition gilt $\sqrt{a}^2 = a$.

[6] Die Betragsfunktion $x \mapsto |x|$ ist nicht monoton.

[7] Diese Konvention für das Symbol $\sqrt{a}$ macht man in den komplexen Zahlen $\mathbb{C}$ nicht; dort steht $\sqrt{4}$ für *irgendeine* der Lösungen $2, -2$.

Nun sei $a \in \mathbb{R}$ beliebig. Die Gleichung $x^2 = a^2$ hat die Lösungen $a, -a$, also ist $\sqrt{a^2}$ die positive dieser beiden Zahlen – welche ist das? Für positives a ist $\sqrt{a^2} = a$, für negatives a ist $\sqrt{a^2} = -a$, kurz:

$$\sqrt{a^2} = |a|.$$

Wegen $(\sqrt{a^2})^2 = a^2$ und $a^2 \geqslant 0$ gilt insbesondere $|a|^2 = a^2 = |a^2|$. Daraus kann man auch etwas über den Umgang mit Gleichungen lernen: Falls beide Seiten nichtnegativ sind, ist das Wurzelziehen erlaubt und eine Äquivalenzumformung. Zweitens kann man *jede* Gleichung quadrieren; dies ist aber nur dann eine Äquivalenzumformung, wenn beide Seiten nichtnegativ sind.[8] Die obige Gleichung hätten wir also auch so lösen können:

$$\begin{aligned} |x-1| &= |x+3| && \Leftrightarrow \\ |x-1|^2 &= |x+3|^2 && \Leftrightarrow \\ (x-1)^2 &= (x+3)^2 && \Leftrightarrow \\ x^2 - 2x + 1 &= x^2 + 6x + 9 && \Leftrightarrow \\ x &= -1. \end{aligned}$$

(5) Wir bezeichnen das **Maximum** zweier Zahlen $a, b \in \mathbb{R}$ mit $\max\{a, b\}$ und das **Minimum** mit $\min\{a, b\}$, das bedeutet

$$\max\{a,b\} := \begin{cases} a, & \text{falls } a \geqslant b, \\ b, & \text{falls } b \geqslant a, \end{cases} \qquad \min\{a,b\} := \begin{cases} a, & \text{falls } a \leqslant b, \\ b, & \text{falls } b \leqslant a. \end{cases}$$

Offensichtlich ist immer $\min\{a, b\} \leqslant a, b \leqslant \max\{a, b\}$ und gerade für das Minimum bzw. Maximum gilt „=“. Beispiel:

$$\begin{aligned} \max\{1, 2, 3\sqrt{2}, 4\} &= 3\sqrt{2}, & \min\{1, 2, 3\sqrt{2}, 4\} &= 1, \\ \max\{-2, 0, 1\} &= 1, & \min\{-2, 0, 1\} &= -2, \\ \max\{\sqrt{2}, \sqrt[3]{3}, \sqrt[4]{4}\} &= \sqrt[3]{3}, & \min\{\sqrt{2}, \sqrt[3]{3}, \sqrt[4]{4}\} &= \sqrt{2} = \sqrt[4]{4}. \end{aligned}$$

Dabei haben wir $3\sqrt{2} - 4 > 0$ und in der dritten Zeile $3^4 = 81 \geqslant 64 = 4^3$ verwendet. Zwischen Betrag, Minimum und Maximum bestehen folgende Zusammenhänge:

$$\begin{aligned} |a| &= \max\{a, -a\} = \max\{a, 0\} - \min\{a, 0\}, \\ \max\{a, b\} &= \frac{1}{2}\left(a + b + |a - b|\right), \\ \min\{a, b\} &= \frac{1}{2}\left(a + b - |a - b|\right). \end{aligned}$$

[8] Ungleichungen dürfen nur quadriert werden, wenn beide Seiten nichtnegativ sind, siehe Satz 3.3(6).

Satz 3.9. Seien $a, b \in \mathbb{R}$. Dann gilt:

(1) Stets ist $|a| \geqslant 0$. Dabei ist $|a| = 0 \Leftrightarrow a = 0$. (**positiv definit**)

(2) $|a| \leqslant b \Leftrightarrow -b \leqslant a \leqslant b$. Entsprechend ist $|a| \geqslant b \Leftrightarrow a \geqslant b$ oder $-b \geqslant a$.

(3) $|ab| = |a||b|$. (**Homogenität**)

(4) $|a + b| \leqslant |a| + |b|$. (**Dreiecksungleichung**)

(5) $\big||a| - |b|\big| \leqslant |a - b|$. (**umgekehrte Dreiecksungleichung**)

(6) *Merkhilfe:* $|a| - |b| \leqslant |a \pm b| \leqslant |a| + |b|$.

Beweis (1) Nach der Totalität ist $a \geqslant 0$ oder $a \leqslant 0$. Im ersten Fall ist $|a| = a \geqslant 0$, im zweiten Fall ebenso $|a| = -a \geqslant 0$. Von der zweiten Aussage ist „$\Leftarrow$“ klar, und „$\Rightarrow$“ auch: Aus $|a| = 0$ folgt $a = 0$ oder $-a = 0$ je nach Vorzeichen von a, jedenfalls $a = 0$.
(2) Ist die linke Seite der ersten Aussage wahr, so gilt $b \geqslant 0$ nach (1). Im Fall $a \geqslant 0$ gilt dann $a \leqslant b$ und $a \geqslant 0 \geqslant -b$, also die rechte Seite. Im Fall $a \leqslant 0$ gilt analog $-a \leqslant b$, also $-b \leqslant a$, und $a \leqslant 0 \leqslant b$. Nun gelte umgekehrt $-b \leqslant a \leqslant b$, dann ist $a \leqslant b$ und $-a \leqslant b$, also sicherlich $|a| \leqslant b$. Für die zweite Aussage bildet man die Kontraposition der ersten und beachtet $|a| = b \Leftrightarrow a = b$ oder $a = -b$.
(3) $|ab|$ und ab unterscheiden sich nur um ein Vorzeichen, ebenso $|a|, a$ und $|b|, b$; dafür schreiben wir kurz $|ab| = \pm ab = \pm|a||b|$. Nun ist $|ab| \geqslant 0$ und $|a||b| \geqslant 0$ nach (1), daher muss das letzte Vorzeichen „+“ sein.
(4) Es gilt $-|a| \leqslant a \leqslant |a|$ und dasselbe für b, daraus folgt $-|a| - |b| \leqslant a + b \leqslant |a| + |b|$. Nach (2) bedeutet das $|a + b| \leqslant |a| + |b|$.
(5) Mithilfe von (4) erhalten wir $|a| = |(a - b) + b| \leqslant |a - b| + |b|$, also $|a| - |b| \leqslant |a - b|$.[9] Dasselbe tun wir für b und erhalten $|b| - |a| \leqslant |b - a| = |a - b|$, also $|a| - |b| \geqslant -|a - b|$. Insgesamt bedeutet das $-|a - b| \leqslant |a| - |b| \leqslant |a - b|$, nach (2) also $\big||a| - |b|\big| \leqslant |a - b|$.
(6) ist eine Kombination der Ungleichungen in (4) und (5). Wir ersetzen b durch $-b$ und erhalten wegen $|-b| = |b|$ dann $|a - b| \leqslant |a| + |b|$ und $\big||a| - |b|\big| \leqslant |a + b|$. Beachtet man noch $|a| - |b| \leqslant \big||a| - |b|\big|$, so kann man alle Ungleichungen zusammensetzen. $\square$

Beispiel 3.10. (1) Die Aussagen in Satz 3.9 sind nicht schwierig, aber der Umgang mit Beträgen in Ungleichungen erfordert etwas Übung. Nach (2) bedeutet $|a| \leqslant b$, dass a in einer Umgebung des Ursprungs von b bis $-b$ liegt, und entsprechend $|a| \geqslant b$, dass a außerhalb dieser Umgebung liegt.
Zum Beispiel sind die Lösungen der Ungleichung $|x| \leqslant 4$ alle Zahlen $-4 \leqslant x \leqslant 4$ in einer Umgebung des Ursprung mit „Radius“ 4, also $\mathcal{L} = [-4, 4]$. Die Ungleichung

$$|x + 2| \leqslant 3$$

ist äquivalent zu $-3 \leqslant x + 2 \leqslant 3$ oder $-5 \leqslant x \leqslant 1$, also $\mathcal{L} = [-5, 1]$. Und alle Lösungen von $|x - 1| \geqslant 2$ liegen außerhalb einer Umgebung mit Radius 2 um die Zahl 1, das ist gerade der Bereich $\mathcal{L} = (-\infty, -1] \cup [3, \infty)$.

[9] Hier wurde die Null „$b - b$“ geschickt eingefügt – das ist besonders bei Beträgen ein nützlicher Trick.

(2) Die Dreiecksungleichung in Satz 3.9(4) ist ein wichtiges Hilfsmittel im Umgang mit Beträgen. Sie bedeutet anschaulich, dass der Weg vom Ursprung nach $a+b$ nicht länger ist als die Wege nach a und b zusammen; für gleiche Vorzeichen von a, b ist das klar, aber für verschiedene Vorzeichen liegt eben $a+b$ näher am Ursprung als $a-b$ und $b-a$. Gleichheit tritt in der Dreiecksungleichung daher genau dann auf, wenn $ab \geqslant 0$, also a, b dasselbe Vorzeichen haben, andernfalls „$<$“:

$$|2+(-1)| = 1 < 3 = |2| + |-1|.$$

Die Dreiecksungleichung gilt allgemeiner für n Zahlen $a_1, \ldots, a_n \in \mathbb{R}$ in der Form

$$|a_1 + \ldots + a_n| \leqslant |a_1| + \ldots + |a_n|.$$

Nach der umgekehrten Dreiecksungleichung in Satz 3.9(5) ist der Abstand zweier Zahlen a, b mindestens so groß wie der Abstand ihrer Beträge. Man mache einige Zahlenbeispiele, um die verschiedenen Bedeutungen des Betrages zu verstehen.

(3) Wir schließen den Abschnitt mit einem etwas komplexeren Beispiel und lösen die Ungleichung

$$|x-3||x+3| < 16.$$

Mithilfe der Homogenität (Satz 3.9(3)) schreiben wir zunächst $|x-3||x+3| = |(x-3)(x+3)| = |x^2-9|$. Aus Satz 3.9(2) folgt dann

$$|x^2-9| < 16 \Leftrightarrow -16 < x^2 - 9 < 16 \Leftrightarrow -7 < x^2 < 25.$$

Da die erste Ungleichung $-7 < x^2$ immer erfüllt ist, ist diese Aussage wiederum mit Satz 3.3(6) und Satz 3.9(2) äquivalent zu

$$x^2 < 25 \Leftrightarrow |x| < 5 \Leftrightarrow -5 < x < 5.$$

Die gesuchte Lösungsmenge ist also $\mathcal{L} = \{x \in \mathbb{R} : -5 < x < 5\} = (-5, 5)$.

3.2 Lösungsstrategien für (Un-)Gleichungen

In diesem Abschnitt werden wir verschiedene Strategien und Techniken vorstellen, um die Lösungsmenge einer Gleichung oder Ungleichung rechnerisch zu bestimmen.[10] Auch wenn wir den Lösungsweg anhand der bisherigen Sätze begründen, gerät dadurch das deduktive Argumentieren etwas in den Hintergrund. Wir lösen vor allem Ungleichungen und gehen systematisch von einfachen zu schwierigen Gleichungen. Dabei werden wir allerdings nicht ganze Klassen von Gleichungen lösen oder Algorithmen formulieren, sondern verschiedene Aspekte, Ideen und Methoden anhand von Beispielen zeigen. Dies tun wir auch und gerade deshalb, weil es allgemeine Verfahren für komplexe Gleichungen häufig nicht gibt; das Lösen von Gleichungen profitiert daher neben Kreativität und vielfältigen Werkzeugen vor allem von Übung und Erfahrung.

[10] Neben diesem *algebraischen* Lösen gibt es natürlich auch andere, etwa graphische oder numerische Methoden, um (hinreichend genaue) Lösungen zu finden.

Gleichungen und Ungleichungen sind Aussageformen, ergeben also durch Belegen der darin vorkommenden Variablen wahre oder falsche Aussagen. Beim Lösen einer Gleichung suchen wir alle Belegungen, für die eine *wahre* Aussage entsteht, und fassen diese **Lösungen** in der **Lösungsmenge** $\mathcal{L}$ zusammen. Das Ziel ist es in der Regel, die gesamte Lösungsmenge in einfacher Form anzugeben, also aufzählend oder als Vereinigung von Intervallen; gelegentlich interessiert auch nur die Lösbarkeit $\mathcal{L} \neq \emptyset$, eine einzelne Lösung $x \in \mathcal{L}$, alle Lösungen mit einer bestimmten Eigenschaft $N \subseteq \mathcal{L}$ oder allgemeine Eigenschaften aller Lösungen $M \supseteq \mathcal{L}$.[11]

In Gleichungen mit einer Variablen x ist gewöhnlich $\mathbb{R}$ die **Grundmenge**, in Gleichungen mit zwei Variablen $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ usw. Manchmal werden nur rationale, ganzzahlige oder Lösungen in einem bestimmten Intervall gesucht; und häufig kann man die Variable nicht sinnvoll mit *jeder* reellen Zahl belegen. In diesen Fällen wählt man für jede Variable eine Teilmenge von $\mathbb{R}$ als Grundmenge der Gleichung.[12] Zum Beispiel ist bei $\frac{1}{x-1} = 2$ die Belegung $x = 1$ nicht sinnvoll und daher $\mathbb{R} \setminus \{1\} = (-\infty, 1) \cup (1,\infty)$ die größtmögliche Grundmenge. In der Gleichung

$$(-2)^x = 8$$

sollte $\mathbb{Z}$ als Grundmenge gewählt werden, da der Ausdruck $(-2)^x$ für Brüche und irrationale x nicht definiert ist. Wie in Beispiel 2.2(5) ist die Angabe der Grundmenge entscheidend für die Lösungsmenge, da Elemente außerhalb der Grundmenge niemals Lösungen sein können. Ist keine Grundmenge angegeben, so wählt man immer die *größtmögliche* Grundmenge, in der alle Variablen sinnvoll belegt werden können.

Verschiedene Lösungsmengen können sehr vielfältig sein: Ist die Gleichung **unlösbar**, so ist $\mathcal{L} = \emptyset$, andernfalls ist sie **lösbar**. Die lineare Gleichung $2x + 5 = 7$ hat *genau eine* Lösung $\mathcal{L} = \{1\}$, die quadratische Gleichung $2x^2 + 5 = 7$ hat *mehr als eine* Lösung $\mathcal{L} = \{-1, 1\}$. Man kann grob unterscheiden zwischen Gleichungen mit *endlich* vielen Lösungen $\mathcal{L} = \{x_1, \ldots, x_n\}$, mit *(abzählbar) unendlich* vielen Lösungen $\mathcal{L} = \{x_1, x_2, \ldots\}$ und mit *(überabzählbar) unendlich* vielen Lösungen, die sich nicht mit „...“ schreiben lassen, zum Beispiel $\mathcal{L} = \mathbb{R}$ oder $\mathcal{L} = [0, 1)$. Üblicherweise (aber nicht immer) haben Gleichungen eine endliche Lösungsmenge und Ungleichungen Lösungsmengen aus der Vereinigung von Intervallen: Zum Beispiel hat $x(x^2 - 1) = 0$ die Lösungsmenge $\mathcal{L} = \{-1, 0, 1\}$ und für die Ungleichung $x(x^2 - 1) > 0$ ist $\mathcal{L} = (-1, 0) \cup (1,\infty)$.

Gleichungen und Ungleichungen löst man mithilfe zielgerichteter **Äquivalenzumformungen**; das sind Umformungen, die *die Lösungsmenge nicht ändern.* Kann die Lösungsmenge der neuen Gleichung abgelesen oder (einfacher) bestimmt werden, so ist auch die ursprüngliche Gleichung gelöst. Die Umformung einer Gleichung $A(x) = B(x)$ zu $C(x) = D(x)$ innerhalb der Grundmenge G ist eine Äquivalenzumformung, wenn

$$\forall x \in G : A(x) = B(x) \Leftrightarrow C(x) = D(x).$$

[11] Nicht nur Zahlen können Lösungen sein, z.B. in *Differentialgleichungen* werden Funktionen gesucht.
[12] Diese Grundmengen können voneinander abhängen, betrachte das Beispiel $\frac{1}{x-y} = 1$.

Fehlt eine der Richtungen $\Rightarrow$ oder $\Leftarrow$, so erhält man nur eine Teil- oder Obermenge von $\mathcal{L}$. Ist f eine *injektive* Funktion (siehe Kapitel 4) und liegen $A(x), B(x)$ für alle $x \in G$ im Definitionsbereich von f, so ist das Anwenden von f auf beiden Seiten der Gleichung eine Äquivalenzumformung. Beispiele dafür sind:

- Die Addition einer beliebigen Zahl.
- Die Multiplikation mit einer Zahl ungleich Null.
- Das Quadrieren, wenn beide Seiten nichtnegativ sind.
- Das Wurzelziehen, wenn beide Seiten nichtnegativ sind.

Für *Ungleichungen* gilt im Wesentlichen dasselbe, nur müssen wir zusätzlich die Richtung von $\leqslant$ berücksichtigen; daher muss f sogar streng monoton wachsend sein (siehe Kapitel 4). Beispiele für Äquivalenzumformungen bei Ungleichungen sind:

- Die Addition einer beliebigen Zahl.
- Die Multiplikation mit einer *positiven* Zahl.
- Das Quadrieren, wenn beide Seiten nichtnegativ sind.
- Das Wurzelziehen, wenn beide Seiten nichtnegativ sind.

Auch Umformungen, die keine Äquivalenzumformungen sind, können zum Ziel führen: In der Grundmenge $[-2, \infty)$ ist das Quadrieren

$$\sqrt{x+2} = x \Rightarrow x + 2 = x^2$$

erlaubt, aber keine Äquivalenzumformung.[13] Die Lösungsmenge der zweiten Gleichung $\{-1, 2\}$ ist größer als $\mathcal{L}$, es können *Scheinlösungen* hinzugekommen sein; aber in $\{-1, 2\}$ liegen sicher alle Lösungen von $\sqrt{x+2} = x$. Diese identifiziert man mithilfe einer *Probe*: Beim Einsetzen in $\sqrt{x+2} = x$ zeigt sich, dass nur $x = 2$ eine Lösung ist; damit ist $\mathcal{L} = \{2\}$ bewiesen. Entsprechend gehen bei der Umformung

$$x(x-1)(x+2) = 0 \Leftarrow x - 1 = 0$$

Lösungen verloren, die Lösungsmenge $\{1\}$ der zweiten Gleichung ist kleiner als $\mathcal{L} = \{-2, 0, 1\}$. Solche Umformungen sind sinnvoll, wenn man eine einzelne Lösung finden will oder bereits Informationen über $\mathcal{L}$ hat. Zum Beispiel hat die kubische Gleichung

$$x^3 - 2x^2 = 3x$$

höchstens drei Lösungen, also $\mathcal{L}$ höchstens drei Elemente; findet man (etwa durch Einsetzen) die drei Lösungen $0, -1$ und 3, so hat man schon $\mathcal{L} = \{-1, 0, 3\}$ bewiesen.

Merkhilfe: Durch das Lösen mit notwendigen Bedingungen „$\Rightarrow$“ kommen Lösungen *hinzu*. Durch das Lösen mit hinreichenden Bedingungen „$\Leftarrow$“ gehen Lösungen *verloren*.

[13] Da die rechte Seite nicht zwangsläufig nichtnegativ ist, folgt aus $x + 2 = x^2$ nur wieder $\sqrt{x+2} = |x|$.

Beispiel 3.11 (Polynome). (1) Die einfachsten Gleichungen sind *lineare Gleichungen*, sie können stets in die Form $ax = b$ gebracht werden. Ihre Grundmenge ist $\mathbb{R}$, $\mathcal{L}$ enthält entweder genau ein Element, ist die leere Menge $\emptyset$ oder ganz $\mathbb{R}$. *Lineare Ungleichungen* löst man wie lineare Gleichungen, zum Beispiel

$$-3(x-2) \leqslant x-6 \Leftrightarrow -3x+6 \leqslant x-6 \Leftrightarrow 12 \leqslant 4x \Leftrightarrow 3 \leqslant x,$$

also $\mathcal{L} = \{x \in \mathbb{R} : -3(x-2) \leqslant x-6\} = \{x \in \mathbb{R} : 3 \leqslant x\} = [3, \infty)$. Auch hier können die Sonderfälle $\mathcal{L} = \mathbb{R}$ oder $\mathcal{L} = \emptyset$ auftreten:

$$x-2 \leqslant x-6 \Leftrightarrow -2 \leqslant -6, \text{ also } \mathcal{L} = \emptyset,$$
$$x-2 \leqslant x+6 \Leftrightarrow -2 \leqslant 6, \text{ also } \mathcal{L} = \mathbb{R}.$$

(2) Gleichungen, in denen nicht nur die Variable x auftritt, sondern auch Brüche, Potenzen usw. mit x (zum Beispiel $x^2, \frac{1}{x}, \sqrt{x}, e^x, \log x$), heißen *nichtlinear*. Nichtlineare Gleichungen können komplexe Verkettungen enthalten und sind daher viel schwieriger als lineare Gleichungen. Die einfachsten nichtlinearen Gleichungen sind *quadratische Gleichungen* wie

$$2x^2 - 10 = 8x.$$

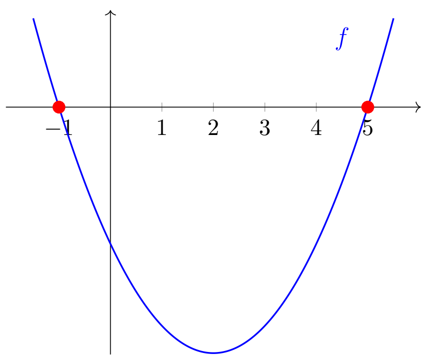
Hier gibt es viele Lösungsverfahren, die schon aus der Schule bekannt sind: Lösungsformeln wie *pq*-Formel oder *abc*-Formel, quadratische Ergänzung oder geschicktes Raten mit dem Satz von Vieta. Gleichungen ohne absoluten Term -10 oder linearen Term $8x$ können einfacher mit dem Satz vom Nullprodukt bzw. Wurzelziehen gelöst werden. Die Lösungsmenge $\mathcal{L} = \{-1, 5\}$ ist letztlich dasselbe wie die Nullstellen des quadratischen Polynoms $f(x) = 2x^2 - 8x - 10$; damit kann f in ein Produkt zweier linearer Polynome zerlegt werden (**Linearfaktorisierung**):

$$2x^2 - 8x - 10 = 2(x+1)(x-5).$$

Aus dieser Darstellung können wir das Vorzeichen von f zwischen den Nullstellen bestimmen und damit auch *quadratische Ungleichungen* lösen:

$$\begin{aligned}
&\text{Für } x < -1 \text{ ist } x+1 < 0 \text{ und } x-5 < 0, \text{ also } 2(x+1)(x-5) > 0,\\
&\text{Für } -1 < x < 5 \text{ ist } x+1 > 0 \text{ und } x-5 < 0, \text{ also } 2(x+1)(x-5) < 0,\\
&\text{Für } 5 < x \text{ ist } x+1 > 0 \text{ und } x-5 > 0, \text{ also } 2(x+1)(x-5) > 0.
\end{aligned}$$

Nach der zweiten Zeile hat die Ungleichung $2x^2 - 10 \leqslant 8x$ die Lösungsmenge $\mathcal{L} = [-1, 5]$, dort liegt f unterhalb der x-Achse; nach den anderen Zeilen hat $2x^2 - 10 > 8x$ die Lösungsmenge $\mathcal{L} = (-\infty, -1) \cup (5, \infty)$, dort liegt f echt oberhalb der x-Achse.



Diese allgemeine Strategie kann man variieren: An der Ungleichung

$$x^2 - 2x + 2 < 0 \Leftrightarrow (x-1)^2 < -1$$

erkennt man sofort $\mathcal{L} = \emptyset$, da Quadrate niemals negativ sind. Die Lösungsmenge von $x^2 - 2 < 2x + 1$ ist der Bereich, wo die Parabel $x^2 - 2$ echt unterhalb der Geraden $2x + 1$ liegt, also das Intervall $\mathcal{L} = (-1, 3)$.

(3) Allgemeine *Polynomgleichungen* können auf folgende Form gebracht werden:

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0 = 0.$$

Hier heißen die Zahlen $a_0, \ldots, a_n \in \mathbb{R}$ **Koeffizienten** und für $a_n \neq 0$ die Potenz n der **Grad** des **Polynoms** $a_n x^n + \ldots + a_0$; lineare Gleichungen haben den Grad 1, quadratische Grad 2. Die Lösungsmenge der Gleichung besteht aus den Nullstellen des Polynoms, die Lösungsmenge entsprechender Ungleichungen erhält man aus dem Vorzeichen des Polynoms zwischen diesen Nullstellen.
Alle Nullstellen eines Polynoms zu bestimmen, ist im Allgemeinen ein schwieriges Problem und wird in der Algebra studiert. Lösungsformeln wie die *pq*-Formel gibt es noch für Grad 3 (**Cardanische Formeln**) und Grad 4, für Grade $n \geqslant 5$ können sie nicht existieren.[14] In der Praxis bestimmt man einzelne Nullstellen näherungsweise oder errät sie, dabei helfen notwendige Kriterien wie das Lemma von Gauß[15]; dann reduziert man den Grad mithilfe einer **Polynomdivision** (siehe Anhang). Ein Beispiel: Das Polynom

$$x^6 + 5x^5 + 6x^4 - 4x^3 - 8x^2 = x^2(x^4 + 5x^3 + 6x^2 - 4x - 8)$$

hat offenbar die Nullstelle $x = 0$, im Restpolynom vom Grad 4 errät man die Nullstelle $x = 1$. Daher dividiert man durch den Linearfaktor $(x - 1)$:

$$x^4 + 5x^3 + 6x^2 - 4x - 8 = (x-1)(x^3 + 6x^2 + 12x + 8).$$

Das Restpolynom vom Grad 3 ist gerade $(x+2)^3$, wir haben also die Linearfaktorisierung

$$x^6 + 5x^5 + 6x^4 - 4x^3 - 8x^2 = x^2(x-1)(x+2)^3$$

und die Nullstellen $0, 1, -2$ bestimmt. Für die Vorzeichen zwischen den Nullstellen berechnet man

$$\begin{aligned} &> 0 \text{ für } x < -2, && < 0 \text{ für } -2 < x < 0, \\ &< 0 \text{ für } 0 < x < 1, && > 0 \text{ für } 1 < x, \end{aligned}$$

daher hat zum Beispiel die Ungleichung $x^6 + 5x^5 + 6x^4 \leqslant 4x^3 + 8x^2$ die Lösungsmenge $\mathcal{L} = [-2, 1]$.

[14] Dieser berühmte Satz wurde 1824 von Niels Henrik Abel bewiesen.
[15] Hat ein Polynom ganzzahlige Koeffizienten, so gilt für jede rationale Nullstelle $\frac{p}{q}$: $p|a_0$ und $q|a_n$.

Nullstellen von Polynomen

(1) Jedes Polynom vom Grad n hat höchstens n verschiedene Nullstellen. Es können durchaus weniger oder keine (reellen) Nullstellen auftreten; tatsächlich hat jedes Polynom eine Darstellung

$$a_n(x-x_1)\dots(x-x_k)(x^2+p_1x+q_1)\dots(x^2+p_lx+q_l),$$

aus linearen und quadratischen Faktoren, wobei $x_1,\dots,x_k \in \mathbb{R}$ genau die (reellen) Nullstellen sind, die quadratischen Faktoren keine (reellen) Nullstellen haben und die Summe der Grade aller Faktoren gleich n ist.

(2) Man verwendet nun die **komplexen Zahlen** $\mathbb{C}$ aus folgendem Grund: Über $\mathbb{C}$ ist *jede* quadratische Gleichung lösbar, jedes quadratische Polynom hat also eine Linearfaktorisierung $(x-z)(x-w)$, wo die Nullstellen $z, w \in \mathbb{C}$ komplexe Zahlen sind. Hierin steckt der **Fundamentalsatz der Algebra**: Jedes Polynom vom Grad $n \geqslant 1$ hat über $\mathbb{C}$ mindestens eine Nullstelle. Damit lässt sich jedes Polynom vom Grad n als Produkt von genau n Linearfaktoren schreiben:

$$a_n(x-z_1)(x-z_2)\dots(x-z_n).$$

Die reellen z_i entsprechen den reellen Nullstellen x_i, die komplexen z_i entsprechen den Nullstellen der quadratischen Faktoren $x^2+p_ix+q_i$.[16]

(3) Die n Linearfaktoren in (2) müssen nicht verschieden sein. Man kann sie in der Form

$$a_n(x-z_1)^{r_1}(x-z_2)^{r_2}\dots(x-z_m)^{r_m}$$

zusammenfassen, dann heißen die Zahlen $r_1,\dots,r_m \in \mathbb{N}$ die **Vielfachheiten** der m verschiedenen (komplexen) Nullstellen $z_1,\dots,z_m \in \mathbb{C}$; es ist $r_1+\dots+r_m = n$. Das Polynom aus obigem Beispiel

$$x^6+5x^5+6x^4-4x^3-8x^2 = x^2(x-1)(x+2)^3$$

hat Grad 6 und genau drei (reelle) Nullstellen: Eine zweifache bei 0, eine einfache bei 1 und eine dreifache bei -2. Im Graphen entspricht das einem Berührpunkt mit der x-Achse bei $x=0$, einem einfachen Nulldurchgang bei $x=1$ und einem Sattelpunkt auf der x-Achse bei $x=-2$.

[16] Man kann zeigen: Komplexe Nullstellen z treten immer paarweise mit ihrem Konjugierten $\overline{z}$ auf.

Beispiel 3.12 (Betrag)**.** Im vorherigen Abschnitt haben wir Betragsgleichungen und -ungleichungen mithilfe einer *Abstandsvorstellung* an der Zahlengeraden gelöst: Die Ungleichung $|3x+6| < 1 \Leftrightarrow |x+2| < \frac{1}{3}$ hat die Lösungsmenge $\mathcal{L} = (-\frac{7}{3}, -\frac{5}{3})$ und für die Gleichung $|x-7| = |x+2|+1$ ist $\mathcal{L} = \{2\}$. In der Ungleichung $|x-1|+|x+3| \leqslant 4$ sollen die Abstände von x zu 1 und zu -3 zusammen nicht größer als 4 sein; das erfüllen gerade die Zahlen im Intervall $\mathcal{L} = [-3, 1]$. Da für alle $x, y \in \mathbb{R}$ die Äquivalenz $x^2 \leqslant y^2 \Leftrightarrow |x| \leqslant |y|$ gilt, kann *Quadrieren* erfolgreich sein: Die Ungleichung

$$|1-2x| \geqslant |x-1| \Leftrightarrow (1-2x)^2 \geqslant (x-1)^2 \Leftrightarrow x(3x-2) \geqslant 0$$

hat die Lösungsmenge $\mathcal{L} = (-\infty, 0] \cup [\frac{2}{3}, \infty)$. Umgekehrt können auch Ungleichungen mit Quadraten auf Betragsungleichungen zurückgeführt werden; zum Beispiel ist

$$20 + x^2 \leqslant 12x \Leftrightarrow (6-x)^2 \leqslant 16 \Leftrightarrow |6-x| \leqslant 4,$$

und daher $\mathcal{L} = [2, 10]$. Eine allgemeine Lösungsstrategie ist es, alle Beträge durch *Fallunterscheidungen* aufzulösen; der Aufwand verringert sich dabei erheblich, wenn man die Fälle geschickt wählt und mit anderen Strategien kombiniert. In der Ungleichung

$$|2x-3| < x-1$$

ist das Quadrieren keine Äquivalenzumformung (die rechte Seite kann negativ sein), daher unterscheiden wir nach dem Vorzeichen des Terms $2x-3$:

$$\text{Fall 1: } x \geqslant \frac{3}{2}. \text{ Dann ist } |2x-3| < x-1 \Leftrightarrow 2x-3 < x-1 \Leftrightarrow x < 2.$$
$$\text{Fall 2: } x < \frac{3}{2}. \text{ Dann ist } |2x-3| < x-1 \Leftrightarrow 3-2x < x-1 \Leftrightarrow \frac{4}{3} < x.$$

Eine Lösung muss im ersten Fall $x \geqslant \frac{3}{2}$ *und* $x < 2$ erfüllen, dieser Fall hat also den Durchschnitt $[\frac{3}{2}, \infty) \cap (-\infty, 2) = [\frac{3}{2}, 2)$ als Lösungsmenge; der zweite Fall entsprechend $(-\infty, \frac{3}{2}) \cap (\frac{4}{3}, \infty) = (\frac{4}{3}, \frac{3}{2})$. Da alle Lösungen von $|2x-3| < x-1$ entweder in Fall 1 *oder* in Fall 2 liegen, ist die gesuchte Lösungsmenge die Vereinigung $\mathcal{L} = [\frac{3}{2}, 2) \cup (\frac{4}{3}, \frac{3}{2}) = (\frac{4}{3}, 2)$. In der Ungleichung

$$||x+1| - |x+3|| < 1$$

wird man sinnvollerweise die drei Fälle $-1 \leqslant x, -3 \leqslant x < -1, x < -3$ unterscheiden, da $-1 \leqslant x$ und $x < -3$ nicht zugleich eintreten können. Im ersten und letzten Fall ist die Differenz der Abstände zu -1 und -3 immer gleich 2, sie tragen also nichts zur Lösungsmenge bei. Im mittleren Fall $-3 \leqslant x < -1$ berechnen wir:

$$||x+1| - |x+3|| < 1 \Leftrightarrow |-x-1-(x+3)| < 1 \Leftrightarrow |2x+4| < 1 \Leftrightarrow |x+2| < \frac{1}{2}.$$

Damit ist die gesuchte Lösungsmenge $\mathcal{L} = (-\frac{5}{2}, -\frac{3}{2})$.

Beispiel 3.13 (Wurzeln)**.** Bei Wurzelgleichungen und -ungleichungen ist die Grundmenge häufig nicht mehr ganz $\mathbb{R}$, da $\sqrt{x}$ nur für $x \geqslant 0$ definiert ist. Zum Beispiel muss in der Ungleichung $\sqrt{x+2} - \sqrt{|x-1|-1} \geqslant 2$ sowohl $x \geqslant -2$ als auch $|x-1| \geqslant 1$ gelten, daher ist $[-2,0] \cup [2,\infty)$ größtmögliche Grundmenge.

Zur Lösung von Wurzelungleichungen wird oft *systematisch quadriert*. Das ist nur dann eine Äquivalenzumformung, wenn beide Seiten nichtnegativ sind; andernfalls ist es bei Ungleichungen nicht erlaubt, bei Gleichungen entstehen *Scheinlösungen*. Argumente wie $x^2 \geqslant 0, |x| \geqslant 0, \sqrt{x} \geqslant 0$, eine Verkleinerung der Grundmenge oder notfalls Fallunterscheidungen können helfen, das Quadrieren zu rechtfertigen. Einige Beispiele:
a) Beide Seiten der Ungleichung $\sqrt{x+1} > \sqrt{x^2-1}$ sind nichtnegativ und daher

$$\sqrt{x+1} > \sqrt{x^2-1} \Leftrightarrow x+1 > x^2-1 \Leftrightarrow x^2-x-2<0 \Leftrightarrow (x-2)(x+1)<0;$$

auf der Grundmenge $\{-1\} \cup [1,\infty)$ erhalten wir die Lösungsmenge $\mathcal{L} = [1,2)$.
b) Die Gleichung $\sqrt{2x^2-2}+x=0$ kann nur für $x \leqslant 0$ gelten, wir können also die Grundmenge $(-\infty,-1] \cup [1,\infty)$ auf $(-\infty,-1]$ reduzieren. Für $x \leqslant -1$ ist dann das Quadrieren eine Äquivalenzumformung:

$$\sqrt{2x^2-2}+x=0 \Leftrightarrow \sqrt{2x^2-2}=-x \Leftrightarrow 2x^2-2=x^2 \Leftrightarrow x^2=2$$

und die Gleichung hat die Lösungsmenge $\mathcal{L} = \{-\sqrt{2}\}$.
c) Der Ungleichung $\sqrt{x-1}+x^2<0$ sehen wir die Unlösbarkeit $\mathcal{L} = \emptyset$ sofort an: Die Summe zweier nichtnegativer Terme ist nie negativ.
d) In der Gleichung $\sqrt{x+1} = 5-x$ mit der Grundmenge $[-1,\infty)$ kann die rechte Seite negativ sein; wir quadrieren dennoch:

$$\sqrt{x+1}=5-x \Rightarrow x+1=(5-x)^2 \Leftrightarrow x^2-11x+24=0 \Leftrightarrow (x-3)(x-8)=0.$$

Damit ist die Obermenge von Lösungen $\{3,8\} \supseteq \mathcal{L}$ gefunden, mit einer Probe schließen wir 8 aus und erhalten $\mathcal{L} = \{3\}$.
e) In der analogen Ungleichung $\sqrt{x+1} \leqslant 5-x$ dürfen wir *nicht* quadrieren; stattdessen unterscheiden wir die Fälle $-1 \leqslant x \leqslant 5, 5 < x$. Im ersten Fall ist $5-x \geqslant 0$ und wir berechnen wie soeben

$$\sqrt{x+1} \leqslant 5-x \Leftrightarrow x+1 \leqslant (5-x)^2 \Leftrightarrow 0 \leqslant (x-3)(x-8),$$

was für $x \in (-\infty,3] \cup [8,\infty)$ erfüllt ist; dieser Fall trägt also den Durchschnitt

$$((-\infty,3] \cup [8,\infty)) \cap [-1,5] = ((-\infty,3] \cap [-1,5]) \cup ([8,\infty) \cap [-1,5]) = [-1,3]$$

zur Lösungsmenge bei. Im zweiten Fall ist hingegen $5-x<0$ und daher die Gleichung $\sqrt{x+1} \leqslant 5-x$ unlösbar.[17] Damit ist $\mathcal{L} = [-1,3]$ die gesuchte Lösungsmenge.

[17] Allgemeine Strategie: Bei den Vorzeichen $+ \leqslant +$ kann man quadrieren, bei $- \leqslant -$ nach Vertauschen der Seiten ebenfalls; Ungleichungen der Form $+ \leqslant -$ sind unlösbar, Ungleichungen der Form $- \leqslant +$ immer erfüllt. Ggf. macht man Fallunterscheidungen nach dem Vorzeichen.

Gelegentlich reicht einfaches Quadrieren nicht aus: Die Gleichung $\sqrt{x}+\sqrt{x+3} = \sqrt{x+8}$ hat die Grundmenge $[0,\infty)$, man berechnet dort

$$\sqrt{x} + \sqrt{x+3} = \sqrt{x+8} \Leftrightarrow x + 2\sqrt{x(x+3)} + x + 3 = x + 8 \Leftrightarrow 2\sqrt{x(x+3)} = 5 - x.$$

Das zweite Quadrieren ist keine Äquivalenzumformung mehr:

$$\begin{aligned} 2\sqrt{x(x+3)} = 5 - x \Rightarrow 4x(x+3) = (5-x)^2 \\ \Leftrightarrow 3x^2 + 22x - 25 = 0 \Leftrightarrow 3(x-1)\left(x + \frac{25}{3}\right) = 0. \end{aligned}$$

$-\frac{25}{3}$ liegt nicht in der Grundmenge $[0,\infty)$ und nach einer Probe ist 1 tatsächlich eine Lösung von $2\sqrt{x(x+3)} = 5 - x$; damit ist die gesuchte Lösungsmenge $\mathcal{L} = \{1\}$.

Wie bei biquadratischen Gleichungen ($x^4 + 3x^2 + 1 = 0$) ist *Substitution* oft erfolgreich: In der Ungleichung $\sqrt{x+3} \geqslant x + 1$ kann man mit einer Fallunterscheidung quadrieren oder $t = \sqrt{x+3}$ substituieren:

$$\sqrt{x+3} \geqslant x + 1 \Leftrightarrow t \geqslant t^2 - 2 \Leftrightarrow 0 \geqslant (t-2)(t+1);$$

dies gilt genau auf dem Intervall $[-1,2]$. Bei der Rücksubstitution $t = \sqrt{x+3}$ muss $t \geqslant 0$ gelten; für $t \in [0,2]$ nimmt $x = t^2 - 3$ Werte in $[-3,1]$ an; und dies liegt in der Grundmenge $[-3,\infty)$, daher folgt $\mathcal{L} = [-3,1]$. Analog ergeben sich bei einer Substitution $t = x^2$ aus jeder Lösung $t \geqslant 0$ zwei Lösungen $|x| = \sqrt{t}$.

Der Umgang mit allgemeinen Wurzeln $\sqrt[n]{x}$ ist trickreich: Aus der Gleichung $\sqrt[3]{x+3} + \sqrt[6]{x+3} = 2$ wird mit der Substitution $t = \sqrt[6]{x+3}$

$$\sqrt[3]{x+3} + \sqrt[6]{x+3} = 2 \Leftrightarrow t^2 + t = 2 \Leftrightarrow (t+2)(t-1) = 0.$$

Rücksubstitution ergibt nur für $t = 1$ eine Lösung und $x = -2$ liegt auch in der Grundmenge $[-3,\infty)$, also ist $\mathcal{L} = \{-2\}$. Gelegentlich hilft auch *gemeinsames Potenzieren*:

$$\sqrt{x+1} = \sqrt[6]{7x+1} \Leftrightarrow (x+1)^3 = 7x + 1 \Leftrightarrow x^3 + 3x^2 - 4x = 0 \Leftrightarrow x(x-1)(x+4) = 0;$$

in der Grundmenge $[-\frac{1}{7},\infty)$ ist daher $\mathcal{L} = \{0,1\}$. In diesem Zusammenhang sind folgende Rechenregeln nützlich:

$$\begin{array}{llll} \sqrt{x^2} = |x|, & \sqrt{x^3} = |x|\sqrt{x}, & \sqrt{x^4} = x^2, & \sqrt{x^5} = x^2\sqrt{x}, \ldots \\ \sqrt{x^2} = |x|, & \sqrt[3]{x^3} = x, & \sqrt[4]{x^4} = |x|, & \sqrt[5]{x^5} = x, \ldots \end{array}$$

Für alle Basen $x, y \geqslant 0$ und Exponenten $a > 0$ gilt

$$x \leqslant y \Leftrightarrow x^a \leqslant y^a, \tag{3.1}$$

daher sind das Potenzieren ($a = 2, 3, \ldots$) und das Wurzelziehen ($a = \frac{1}{2}, \frac{1}{3}, \ldots$) Äquivalenzumformungen, wenn beide Seiten nichtnegativ sind.[18]

[18] Bei ungeraden Exponenten $a \in \mathbb{Z}$ gilt (3.1) sogar für alle $x, y \in \mathbb{R}$. *Frage:* Gilt (3.1) auch für $a = 0$? Gilt (3.1) auch mit „<“ statt „$\leqslant$“?

Beispiel 3.14 (Brüche)**.** Ähnlich wie bei Wurzeln ist bei Bruchgleichungen und -ungleichungen die Grundmenge häufig nicht ganz $\mathbb{R}$, da $\frac{1}{x}$ nur für $x \neq 0$ definiert ist. Zum Beispiel hat das Polynom $x^2 - 5x + 6$ die Nullstellen $2, 3$ und daher hat

$$\frac{1}{x-1} = \frac{1}{x^2 - 5x + 6}$$

die maximale Grundmenge $\mathbb{R} \setminus \{1, 2, 3\} = (-\infty, 1) \cup (1,2) \cup (2,3) \cup (3, \infty)$. Beim Lösen von Bruchgleichungen ist es häufig erfolgreich, durch Multiplikation mit dem *Hauptnenner* alle Nenner zu beseitigen: Im Beispiel

$$\frac{x+2}{2x+3} = \frac{5x+2}{6x+9} \Leftrightarrow (x+2)(6x+9) = (5x+2)(2x+3) \Leftrightarrow 4x^2 - 2x - 12 = 0$$

ist $(2x+3)(6x+9)$ der Hauptnenner; von den Lösungen 2 und $-\frac{3}{2}$ der Polynomgleichung liegt $-\frac{3}{2}$ nicht in der Grundmenge, daher ist $\mathcal{L} = \{2\}$. Bei *Bruchungleichungen* muss sehr sorgfältig auf das Vorzeichen der Nenner geachtet werden, meist sind Fallunterscheidungen notwendig. Zum Beispiel unterscheidet man bei der Ungleichung

$$\frac{1}{x-2} + \frac{1}{2x+2} \leqslant \frac{1}{2x-4}$$

mit der Grundmenge $\mathbb{R} \setminus \{-1, 2\}$ die Fälle $x < -1, -1 < x < 2$ und $2 < x$, sodass der Hauptnenner $(x-2)(2x+2)(2x-4)$ negatives/ positives/ positives Vorzeichen besitzt. Hier ist es geschickt, zuerst die passenden Bruchterme zusammenzufassen:

$$\frac{1}{x-2} + \frac{1}{2x+2} \leqslant \frac{1}{2x-4} \Leftrightarrow \frac{1}{2(x-2)} + \frac{1}{2(x+1)} \leqslant 0 \Leftrightarrow \frac{2x-1}{2(x-2)(x+1)} \leqslant 0.$$

In den Fällen $x < -1$ und $2 < x$ ist der Nenner $(x-2)(x+1)$ positiv und daher

$$\frac{2x-1}{2(x-2)(x+1)} \leqslant 0 \Leftrightarrow 2x - 1 \leqslant 0 \Leftrightarrow x \leqslant \frac{1}{2},$$

im Fall $-1 < x < 2$ ist der Nenner negativ und man erhält dasselbe mit $\geqslant$ statt $\leqslant$. Damit besteht die Lösungsmenge aus der Vereinigung beider Fälle

$$\mathcal{L} = \Big(((-\infty, -1) \cup (2, \infty)) \cap (-\infty, \tfrac{1}{2}] \Big) \cup ((-1, 2) \cap [\tfrac{1}{2}, \infty)) = (-\infty, -1) \cup [\tfrac{1}{2}, 2).$$

Auch negative Exponenten stellt man als Brüche dar, denn im Gegensatz zu (3.1) ist das Potenzieren mit negativen Exponenten keine Äquivalenzumformung: Es gilt $-2 \leqslant 1 \leqslant 3$, aber

$$-\frac{1}{2} \leqslant 1 \text{ ist richtig,} \qquad 1 \leqslant \frac{1}{3} \text{ ist falsch.}$$

Beispiel: Die Ungleichung $x + x^{-1} \geqslant 4$ hat die Grundmenge $\mathbb{R} \setminus \{0\}$, für positive x gilt

$$x + x^{-1} \geqslant 4 \Leftrightarrow x + \frac{1}{x} \geqslant 4 \Leftrightarrow x^2 + 1 \geqslant 4x \Leftrightarrow (x - 2 - \sqrt{3})(x - 2 + \sqrt{3}) \geqslant 0,$$

für negative x dasselbe mit $\leqslant$ statt $\geqslant$. Die Lösungsmenge ist daher

$$\begin{aligned}\mathcal{L} &= \Big((-\infty,0)\cap[2-\sqrt{3},2+\sqrt{3}]\Big)\cup\Big((0,\infty)\cap\Big((-\infty,2-\sqrt{3}]\cup[2+\sqrt{3},\infty)\Big)\Big)\\ &= (0,2-\sqrt{3}]\cup[2+\sqrt{3},\infty).\end{aligned}$$

In komplexen Ungleichungen können Beträge, Wurzeln und Brüche zugleich auftreten, was eine Kombination aller bisherigen Strategien erforderlich macht: Für die Ungleichung $\frac{|x-1|}{2x+2}\geqslant 1$ mit der Grundmenge $\mathbb{R}\setminus\{-1\}$ unterscheidet man die Fälle $x<-1$, $-1<x<1$ und $1\leqslant x$; man erhält die drei Ungleichungen

$$1-x\leqslant 2x+2,\qquad 1-x\geqslant 2x+2,\qquad x-1\geqslant 2x+2.$$

Nach kurzer Rechnung ergibt sich die Lösungsmenge

$$\begin{aligned}\mathcal{L} &= \left((-\infty,-1)\cap\left[-\frac{1}{3},\infty\right)\right)\cup\left((-1,1)\cap\left(-\infty,-\frac{1}{3}\right]\right)\cup\Big([1,\infty)\cap(-\infty,-3]\Big)\\ &= \left(-1,-\frac{1}{3}\right].\end{aligned}$$

Beispiel 3.15 (Exponentialfunktion, Logarithmus und Trigonometrie)**.** Jede beliebige Gleichung lässt sich in der Form $f(x)=0$ schreiben; mithilfe einer *Kurvendiskussion* der Funktion f kann man daher die Lösungsmenge aus den Nullstellen von f bestimmen, genauso bei Ungleichungen wie $f(x)\geqslant 0$ aus dem Vorzeichen zwischen diesen Nullstellen. Komplexe Gleichungen, die neben Polynomen, Beträgen, Wurzeln und Brüchen zum Beispiel noch Variable im Exponenten, Logarithmen oder trigonometrische Funktionen beinhalten, lassen sich oft nicht algebraisch lösen.[19] Die Kurvendiskussion von f wird daher numerisch mit dem Computer durchgeführt, Nullstellen können zum Beispiel mit dem *Newton-Verfahren* näherungsweise bestimmt werden.

Manche komplexen Gleichungen haben auch eine algebraische Lösung, in diesen Spezialfällen wird oft Wissen über die Funktion f genutzt. Zum Beispiel sind die Gleichungen

$$3\cdot 2^{x^2}+5=0,\qquad e^{2-x}\leqslant -1,\qquad \sin(x+1)=2$$

allesamt unlösbar ($\mathcal{L}=\emptyset$), da für die Exponentialfunktion $e^x>0$ gilt und der Sinus nur Werte im Intervall $[-1,1]$ annimmt.

Logarithmen verkleinern die Grundmenge, da $\log x$ nur für $x>0$ definiert ist; zum Beispiel hat die Ungleichung $\log_2(3x-1)+\log_2(x+5)<6$ die Grundmenge $(\frac{1}{3},\infty)$. Die Anwendung der natürlichen Exponentialfunktion und des (natürlichen) Logarithmus' ist eine Äquivalenzumformung, denn für alle Basen $a>1$ und Exponenten $x,y\in\mathbb{R}$ gilt

$$x\leqslant y\Leftrightarrow a^x\leqslant a^y, \tag{3.2}$$

[19] Man versuche es mit der Gleichung $e^x+x=0$.

man vergleiche dies mit (3.1). Für obige Ungleichung ist daher

$$\begin{aligned}\log_2(3x-1)+\log_2(x+5)<6 &\Leftrightarrow (3x-1)(x+5)<2^6\\ &\Leftrightarrow 3x^2+14x-69<0 \Leftrightarrow 3(x-3)\left(x+\frac{23}{3}\right)<0\end{aligned}$$

und innerhalb der Grundmenge $(\frac{1}{3},\infty)$ ist die Lösungsmenge $\mathcal{L}=(\frac{1}{3},3)$. Für Basen $0<a<1$ gilt im Gegensatz zu (3.2) $x\leqslant y\Leftrightarrow a^x\geqslant a^y$; dieser Fehler tritt häufig beim Logarithmieren von Ungleichungen auf. Wendet man $\log_{\frac{1}{2}}$ auf die Ungleichung $(\frac{1}{2})^x\geqslant 8$ an, so ergibt sich

$$\left(\frac{1}{2}\right)^x\geqslant 8\Leftrightarrow x\leqslant\log_{\frac{1}{2}}(8)=-4,$$

also $\mathcal{L}=(-\infty,-4]$. Es ist daher empfehlenswert, nur den (natürlichen) Logarithmus log zu verwenden und dabei $\log x<0$ für $0<x<1$ zu beachten:

$$\left(\frac{1}{2}\right)^x\geqslant 8\Leftrightarrow x\log\left(\frac{1}{2}\right)\geqslant\log 8\Leftrightarrow x\leqslant\frac{\log 8}{\log(\frac{1}{2})}=\frac{\log_{\frac{1}{2}}8}{\log_{\frac{1}{2}}(\frac{1}{2})}=-4.$$

Auch Substitution ist bei komplexen Gleichungen oft erfolgreich: Bei $2^{4x}+1024=2^{2x+6}$ substituiert man $y=2^x$ und erhält

$$2^{4x}+1024=2^{2x+6}\Leftrightarrow y^4+2^{10}=2^6\cdot y^2\Leftrightarrow(y^2-2^5)^2=0.$$

Aus $y^2=2^5\Leftrightarrow|y|=2^{\frac{5}{2}}$ ergibt sich nur für $y=2^{\frac{5}{2}}$ eine Lösung, nämlich $\mathcal{L}=\{\frac{5}{2}\}$.

Beispiel 3.16 (Systeme von Gleichungen)**.** Die Lösungsmenge eines Systems von Gleichungen ist die Schnittmenge der einzelnen Lösungsmengen: Das System

$$\begin{cases}3x^2+5=8\\ 5-2x\ \ =7\end{cases}\tag{3.3}$$

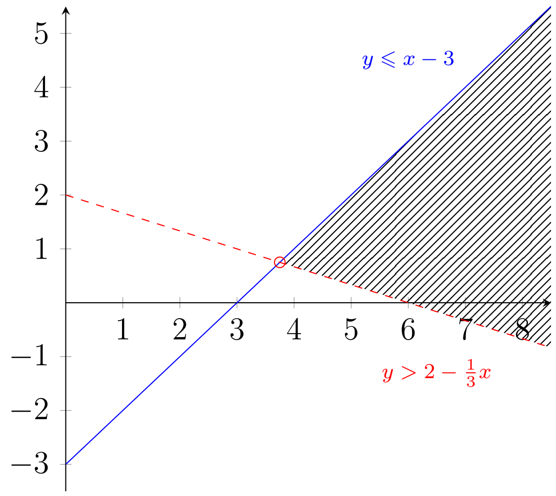
hat die Grundmenge $\mathbb{R}$ und die Lösungsmenge $\mathcal{L}=\{-1,1\}\cap\{-1\}=\{-1\}$.

Wie in einzelnen Gleichungen können auch in Gleichungssystemen mehrere Variable zugleich auftreten. Sind alle Gleichungen des Systems linear, so heißt es ein **lineares Gleichungssystem** (LGS); das System (3.3) ist nichtlinear. Für LGS kennt man aus der Schule das Einsetzungs- und das Additionsverfahren; dies wird in der Linearen Algebra 1 zum *Gauß-Algorithmus* verallgemeinert und ausführlich untersucht.

Für ein LGS mit zwei Variablen und zwei Gleichungen kann man $\mathcal{L}$ durch die Schnittmenge zweier Geraden in der Ebene veranschaulichen: Bei identischen Geraden ist $\mathcal{L} = \mathbb{R}^2$ unendlich, bei sich schneidenden Geraden enthält $\mathcal{L}$ nur den Schnittpunkt $S(x, y)$ und bei parallelen Geraden ist $\mathcal{L} = \emptyset$. Genauso kann man die Lösungsmenge eines Systems linearer Ungleichungen als Schnittmenge zweier Halbebenen darstellen: Im System

$$\begin{cases} 3x - 3y \geqslant 9 \\ 2x + 6y > 12 \end{cases} \Leftrightarrow \begin{cases} y \leqslant x - 3 \\ y > 2 - \frac{1}{3}x \end{cases} \tag{3.4}$$

entspricht die erste Gleichung der Halbebene unterhalb der Geraden $y = x - 3$, die zweite der Halbebene echt oberhalb von $y = 2 - \frac{1}{3}x$, die Schnittmenge ist $\mathcal{L}$. Analog kann man sich Lösungsmengen von linearen Systemen mit drei Variablen im Raum vorstellen: Die Ungleichung $x + 2y - 3z \geqslant 0$ ist ein abgeschlossener *Halbraum* im $\mathbb{R}^3$; mit der Schnittmenge einiger Halbräume lassen sich zum Beispiel *Polyeder* (Quader, Tetraeder, Oktaeder usw.) beschreiben.



Für nichtlineare Systeme gibt es keine allgemeine Lösungsstrategie, vor allem bei mehreren Variablen in gekoppelten Gleichungen. Ungleichungssysteme treten häufig als *Ungleichungskette* auf: Die Bedingung $x + 1 \leqslant 2|x| \leqslant x + 2$ ist einfach das System

$$\begin{cases} x + 1 \leqslant 2|x|, \\ 2|x| \leqslant x + 2. \end{cases}$$

Hier erhält man mit einer Fallunterscheidung

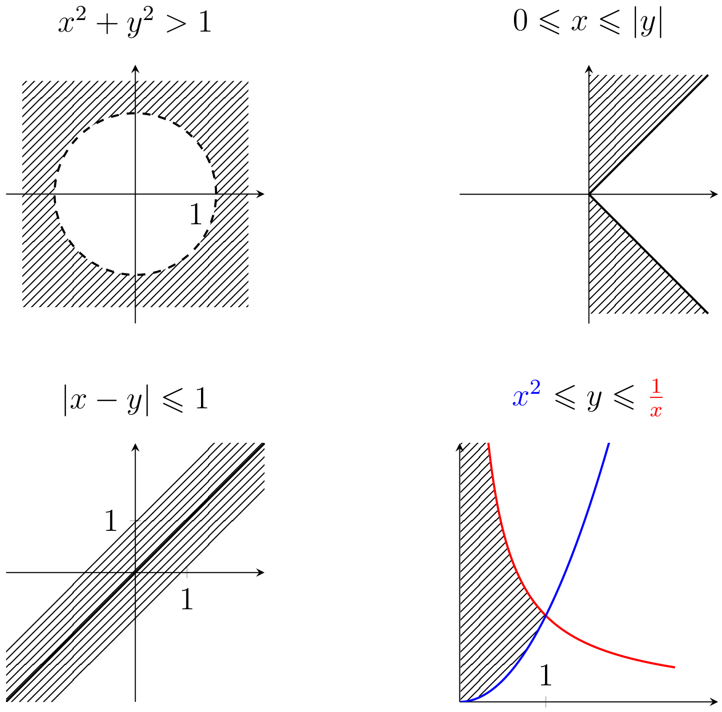
Fall 1: $x \geqslant 0$. Dann $x + 1 \leqslant 2|x| \leqslant x + 2 \Leftrightarrow x + 1 \leqslant 2x \leqslant x + 2 \Leftrightarrow 1 \leqslant x \leqslant 2$.

Fall 2: $x < 0$. Dann $x + 1 \leqslant 2|x| \leqslant x + 2 \Leftrightarrow x + 1 \leqslant -2x \leqslant x + 2 \Leftrightarrow -\dfrac{1}{3} \geqslant x \geqslant -\dfrac{2}{3}$.

Die Lösungsmenge ist also $\mathcal{L} = [-\frac{2}{3}, -\frac{1}{3}] \cup [1, 2]$.

Auch die Lösungsmengen nichtlinearer Systeme mit zwei Variablen lassen sich als Teilmengen des $\mathbb{R}^2$ interpretieren:

- Die Menge $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 > 1\}$ beschreibt die Fläche (echt) außerhalb des Einheitskreises.
- $0 \leqslant x \leqslant |y|$ ist die Vereinigung der Flächen oberhalb der ersten und unterhalb der zweiten Winkelhalbierenden im ersten bzw. zweiten Quadranten.
- $|x - y| \leqslant 1$ ist ein diagonaler Streifen der Breite $\sqrt{2}$ um die Gerade $y = x$.
- $x^2 \leqslant y \leqslant \frac{1}{x}$ ist die Fläche zwischen den Funktionen $f(x) = x^2$ und $g(x) = \frac{1}{x}$.



4 Abbildungen

4.1 Abbildungen als Relationen

Wann immer die Objekte zweier Mengen einander zugeordnet werden, spricht die Mathematik von einer **Relation**. Unter diesem allgemeinen Konzept werden wir die bisherigen Kapitel zusammenfassen und den Begriff **Abbildung** neu einführen.

Definition 4.1. Seien A, B Mengen. Eine Teilmenge $R \subseteq A \times B$ heißt eine **Relation** zwischen A und B. Ist $(x, y) \in R$, so steht $x \in A$ in der Relation R zu $y \in B$ und man schreibt auch xRy. Ist $A = B$, so nennt man $R \subseteq A^2$ eine Relation *auf* A.

Beispiel 4.2. (1) Gehören Ihnen diese Magazine? Die Mathematik antwortet: Sei A die Menge der Menschen, B die Menge der Magazine und die Relation $R \subseteq A \times B$:

$$R = \{(x, y) \in A \times B : \text{Person } x \text{ gehört das Magazin } y\}.$$

$(x, y) \in R$ bedeutet dann, dass Mensch $x \in A$ das Magazin $y \in B$ gehört. Paul gehört das Magazin „Times“, also $(\text{Paul}, \text{Times}) \in R$, aber nicht Clara: $(\text{Clara}, \text{Times}) \notin R$. Sie besitzt auch kein anderes Magazin: $(\text{Clara}, y) \in R$ für kein $y \in B$, aber alle teilen sich „Nature“: $(x, \text{Nature}) \in R$ für jedes $x \in A$. Als zweites betrachten wir auf A die Relation

$$S = \{(x, y) \in A^2 : x \text{ ist die Mutter von } y\}.$$

$(\text{Esther}, \text{Clara}) \in S$ oder kürzer Esther S Clara bedeutet, dass Esther die Mutter von Clara ist; zusammen mit Ruth S Esther ist Ruth eine Großmutter von Clara. Paul Sy ist für keinen Menschen $y \in A$ erfüllt, ebenso ySy; hingegen gibt es genau ein $x \in A$ mit xS Paul – von Regenbogenfamilien mal abgesehen. Als drittes sei

$$T = \{(x, y) \in A^2 : x \text{ ist vor } y \text{ geboren}\}.$$

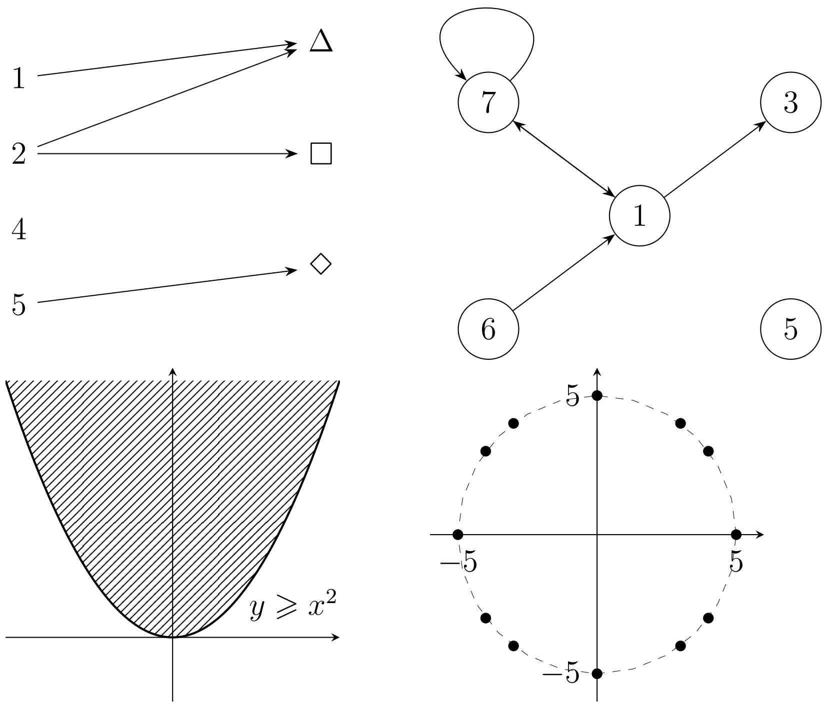
Dann gilt Esther T Clara und Ruth T Esther, also auch Ruth T Clara. Für je zwei $x, y \in A$ ist xTy oder yTx immer richtig; sind beide wahr, so wurden x, y zum gleichen Zeitpunkt geboren. Mithilfe der Relation T lassen sich alle Menschen auf einem Zeitstrahl *anordnen* und vergleichen.

(2) Relationen lassen sich als **Pfeildiagramme** darstellen: Jedem Paar $(x, y) \in R \subseteq A \times B$ entspricht ein Pfeil von $x \in A$ nach $y \in B$. Im Fall $A = B$ erhält man einen *gerichteten Graphen*, dessen *Knoten* die Elemente $x \in A$ und dessen *Kanten* die Relationen $(x, y) \in R \subseteq A^2$ sind. Die Abbildung zeigt die beiden Relationen

$$\begin{aligned} R_1 &= \{(1,\Delta),(2,\Delta),(2,\Box),(5,\Diamond)\} &&\text{auf } \{1,2,4,5\} \times \{\Delta,\Box,\Diamond\},\\ R_2 &= \{(1,3),(1,7),(6,1),(7,1),(7,7)\} &&\text{auf } \{1,3,5,6,7\}^2. \end{aligned}$$

Ist zum Beispiel $A = \mathbb{R}$ eine *Zahlenmenge*, so kann man eine Relation $R \subseteq A^2 = \mathbb{R}^2$ auch im Koordinatensystem darstellen: Jeder Relation xRy entspricht ein Punkt $(x, y) \in R$ in der Ebene. In der Abbildung sehen wir die Relationen

$$\begin{aligned} R_3 &= \{(x,y) \in \mathbb{R}^2 : x^2 \leqslant y\},\\ R_4 &= \{(x,y) \in \mathbb{Z}^2 : x^2 + y^2 = 25\}. \end{aligned}$$



(3) Wir haben in diesem Kurs schon viele Relationen kennengelernt:

$$\begin{array}{lllll} a \leqslant b, & a < b, & a = b, & a \neq b & \text{für Zahlen } a, b \in \mathbb{R},\\ A \subseteq B, & A \subsetneq B, & A = B, & \#A = \#B & \text{für Mengen } A, B. \end{array}$$

In allen Fällen schreiben wir ganz natürlich xRy statt $(x, y) \in R$, zum Beispiel $1 < 3$ statt $(1, 3) \in <$ oder $\mathbb{Z} \subseteq \mathbb{Q}$ statt $(\mathbb{Z}, \mathbb{Q}) \in \subseteq$.

(4) Seien $n, t \in \mathbb{Z}$. Auch die *Teilbarkeit* $t|n$ ist eine Relation:

$$| = \{(t, n) \in \mathbb{Z}^2 : \exists k \in \mathbb{Z} : n = kt\} \subseteq \mathbb{Z}^2.$$

Ist t kein Teiler von n, so entsteht beim Teilen ein **Rest** r: Wir können n schreiben in der Form $n = kt + r$, wobei $k \in \mathbb{Z}$ und $r \in \{0, \dots, t-1\}$ gewählt wird. Für zwei Zahlen $n, m \in \mathbb{Z}$ definieren wir die **Modulo**-Relation

$$n \equiv m \mod t :\Leftrightarrow n \text{ und } m \text{ lassen beim Teilen durch } t \text{ den gleichen Rest,}$$

dann gelten die Äquivalenzen

$$n \equiv m \mod t \Leftrightarrow n - m \equiv 0 \mod t \Leftrightarrow t|n - m. \tag{4.1}$$

Einige Beispiele:

$$\begin{array}{lll} 14 = 2 \cdot 5 + 4, & 15 = 3 \cdot 5, & 16 = 3 \cdot 5 + 1, \\ 11 \equiv 16 \mod 5, & 7 \equiv 1 \mod 2, & 6 \equiv 0 \mod 3. \end{array}$$

Definition 4.3. Es sei A eine Menge und $R \subseteq A^2$ eine Relation auf A.

(1) Die Relation R heißt **reflexiv**, falls stets xRx gilt.
(2) Die Relation R heißt **symmetrisch**, falls

$$\forall x, y \in A : xRy \Rightarrow yRx.$$

(3) Die Relation R heißt **antisymmetrisch**, falls

$$\forall x, y \in A : xRy \text{ und } yRx \Rightarrow x = y.$$

(4) Die Relation R heißt **transitiv**, falls

$$\forall x, y, z \in A : xRy \text{ und } yRz \Rightarrow xRz.$$

(5) Die Relation R heißt **total**, falls stets xRy oder yRx gilt.

Beispiel 4.4. (1) Sei A die Menge der Reisenden am Hbf Karlsruhe und

$$R = \{(x, y) \in A^2 : x \text{ fährt mit demselben Zug wie } y\}.$$

Die Relation R ist reflexiv (Ich fahre mit dem Zug, mit dem ich fahre.), symmetrisch (Fährst du mit meinem Zug, fahre ich mit deinem Zug.) und transitiv (Fährst du mit meinem Zug und ich mit seinem Zug, fährst auch du mit seinem Zug.), aber weder antisymmetrisch noch total.

(2) Symmetrische Relationen sind ungerichtet, xRy und yRx sind äquivalent. Beispiele dafür sind die Gleichheit zweier Zahlen $a = b$ oder zweier Mengen $A = B$, die Relation „x unterscheidet sich von y um ein Vielfaches von 2π“ oder „E und F haben denselben Flächeninhalt“. Die Relationen $a \leqslant b$ und $A \subseteq B$ sind *nicht* symmetrisch, sie haben eine Richtung; beide sind antisymmetrisch (s. Beispiel 2.4(3)):

$$a \leqslant b \text{ und } b \leqslant a \Rightarrow a = b,$$
$$A \subseteq B \text{ und } B \subseteq A \Rightarrow A = B.$$

Während eine der Aussagen $a \leqslant b$ und $b \leqslant a$ immer wahr ist, können beide Aussagen $A \subseteq B$ und $B \subseteq A$ falsch sein (Beispiel?); $\leqslant$ ist also total, $\subseteq$ ist *nicht* total. Aus Definition 3.1 kennen wir die Reflexivität und Transitivität der Relationen $=$ und $\leqslant$:

$$a = b \text{ und } b = c \Rightarrow a = c,$$
$$a \leqslant b \text{ und } b \leqslant c \Rightarrow a \leqslant c.$$

Auch die Teilmengenrelation $\subseteq$ ist reflexiv und transitiv. Die „echten" Varianten $a < b$ und $A \subsetneqq B$ sind ebenfalls transitiv, aber weder reflexiv noch antisymmetrisch noch total.[1] Die gleichen Aussagen erhält man natürlich für $\geqslant, \supseteq, >$ und $\supsetneqq$.

(3) Viele Relationen sind nicht transitiv, zum Beispiel „x ist verheiratet mit y" oder das Spiel Schere-Stein-Papier. Auch „A und B sind disjunkt" ist eine nicht reflexive, nicht transitive Relation: Für die Mengen $A = \{1\}, B = \{2\}$ und $C = \{1\}$ gilt

$$A \cap B = \emptyset, B \cap C = \emptyset, \text{ aber } \qquad A \cap C = \{1\} \neq \emptyset.$$

(4) Die Teilerrelation $t|n$ ist *reflexiv*, denn für alle $n \in \mathbb{Z}$ gilt $n = 1 \cdot n$, also $n|n$. Sie ist auch *transitiv*: Sind $n_1, n_2, n_3 \in \mathbb{Z}$ mit $n_1|n_2$ und $n_2|n_3$ gegeben, so gibt es $k_1, k_2 \in \mathbb{Z}$, sodass $n_1 = k_1 n_2, n_2 = k_2 n_3$. Daraus folgt $n_1 = k_1 n_2 = k_1 k_2 n_3$, also $n_1|n_3$. Da die Teilbarkeit offenbar weder symmetrisch noch total ist, bleibt die Antisymmetrie zu untersuchen: Für zwei $n, m \in \mathbb{Z}$ gelte $n|m$ und $m|n$, dann gibt es wieder $k_1, k_2 \in \mathbb{Z}$ mit der Eigenschaft $n = k_1 m, m = k_2 n$.

Folgt daraus $n = m$? Auf den ganzen Zahlen $\mathbb{Z}$ im Allgemeinen nicht, wie man am Beispiel $1 = (-1) \cdot (-1)$ und $-1 = (-1) \cdot 1$ erkennt. Sind aber $n, m \in \mathbb{N}$ *natürliche* Zahlen, so sind auch $k_1, k_2 \in \mathbb{N}$ und es folgt (bei $n = 0$ ist $m = 0$)

$$n = k_1 m = k_1 k_2 n \Rightarrow k_1 k_2 = 1 \Rightarrow k_1 = k_2 = 1 \Rightarrow n = m.$$

Die Eigenschaften einer Relation $R \subseteq A^2$ hängen also von der Menge A ab: Die Teilbarkeit ist *antisymmetrisch* auf $\mathbb{N}$, aber nicht auf $\mathbb{Z}$.

Definition 4.5. Es sei A eine Menge und $R \subseteq A^2$ eine Relation auf A.

(1) Ist die Relation R reflexiv, symmetrisch und transitiv, so heißt R eine **Äquivalenzrelation** auf A; in diesem Fall nennt man zwei Elemente xRy *äquivalent*.

(2) Ist die Relation R reflexiv, antisymmetrisch und transitiv, so nennt man R eine (partielle) **Ordnungsrelation** auf A und die Menge A eine (partiell) **geordnete Menge**. Ist R zusätzlich total auf A, so nennt man R eine **totale Ordnungsrelation** und A eine **total geordnete Menge**.

[1] Man erkennt: Die Eigenschaften in Definition 4.3 sind unabhängig voneinander; und es kann sein, dass eine Relation weder symmetrisch noch antisymmetrisch ist.

Beispiel 4.6. Wir fassen hier unsere bisherigen Ergebnisse zusammen, diese Begriffe werden im Mathematikstudium weiter untersucht.

(1) Die Gleichheit zweier Zahlen $a = b$ oder zweier Mengen $A = B$ ist eine Äquivalenzrelation. Die Kleiner-Gleich-Relation $a \leqslant b$ ist eine (totale) Ordnungsrelation und die Teilmengenrelation $A \subseteq B$ eine (partielle) Ordnungsrelation, entsprechend auch die Relationen $a \geqslant b$ und $A \supseteq B$. Die Relationen $a < b, a > b$ sowie $A \subsetneqq B, A \supsetneqq B$ sind weder Äquivalenz- noch Ordnungsrelationen. Der Begriff „Ordnungsrelation" drückt aus, dass wir damit Objekte der Größe nach sortieren können:

$$\emptyset \subseteq \{2\} \subseteq \{1,2\} \subseteq \{2,1\} \subseteq \{1,2,3,4\} \subseteq \dots$$

Unsere Vorstellung von der Sortierung aller Zahlen auf der Zahlengeraden haben wir in Definition 3.1 mithilfe der **Anordnungsaxiome** ausgedrückt, diese Axiome besagen nun: Die Relation $\leqslant$ ist eine totale Ordnungsrelation auf $\mathbb{R}$, die mit Addition und Multiplikation verträglich ist.

(2) Die Teilbarkeit $t|n$ ist eine (partielle) Ordnungsrelation auf $\mathbb{N}$, aber nicht auf $\mathbb{Z}$. Wir können also natürliche Zahlen nicht nur mit $\leqslant$ ordnen, sondern auch in der Form

$$1 \,|\, 2 \,|\, 6 \,|\, 30 \,|\, 210 \,|\, 630 \,|\, \dots$$

Die Modulo-Relation $n \equiv m \mod t$ aus (4.1) ist für jedes $t \in \mathbb{N}$ eine Äquivalenzrelation auf $\mathbb{Z}$, beweisen Sie das zur Übung. Äquivalente Zahlen modulo t unterscheiden sich nur um Vielfache von t: $-5 \equiv 1 \equiv 7 \equiv \dots \mod 6$. Möchte man nur über Reste bei der Division durch 6 sprechen, so ist jede dieser Zahlen ein *Repräsentant* der **Restklasse**

$$[1]_6 = \{6k+1 : k \in \mathbb{Z}\} = \{\dots, -11, -5, 1, 7, 13, \dots\}.$$

Jede ganze Zahl liegt in genau einer der sechs Restklassen $[0]_6, [1]_6, [2]_6, [3]_6, [4]_6, [5]_6$; die Menge dieser Restklassen wird mit $\mathbb{Z}/6\mathbb{Z}$ oder $\mathbb{Z}_6$ bezeichnet. Allgemein wird jede Menge A durch eine Äquivalenzrelation $\sim \subseteq A^2$ in *Äquivalenzklassen* $[x]_\sim$ zerlegt.

(3) Die Gleichmächtigkeit $\#M = \#N$ von (endlichen) Mengen M, N ist offenbar reflexiv, symmetrisch und transitiv, also eine Äquivalenzrelation. Das gilt auch für unendliche Mengen, siehe Abschnitt 4.2.

(4) Die tautologische Äquivalenz ist eine Äquivalenzrelation auf der Menge aller Aussagen, siehe Satz 1.21: $A \Leftrightarrow A$ ist eine Tautologie (Reflexivität), aus $A \leftrightarrow B$ folgt $B \leftrightarrow A$ (Symmetrie) und aus $A \leftrightarrow B$ und $B \leftrightarrow C$ folgt auch $A \leftrightarrow C$ (Transitivität). Alle Aussagen mit denselben Einträgen in der Wahrheitstafel fallen in dieselbe Äquivalenzklasse, werden also auf ihren Wahrheitswert reduziert. Genauso kann man eine Ordnungsrelation $A \to B$ definieren durch

$$A \to B \text{ genau dann, wenn } A \Rightarrow B \text{ eine Tautologie ist.}$$

Von Relationen $R \subseteq A^2$ auf einer Menge A kommen wir nun zurück zu Relationen $R \subseteq A \times B$ zwischen zwei Mengen A, B.

Definition 4.7. Seien A, B Mengen und $R \subseteq A \times B$ eine Relation zwischen A und B.

(1) Die Relation R heißt **linksvollständig**, falls jedes $x \in A$ zu mindestens einem $y \in B$ in der Relation R steht, d.h.

$$\forall x \in A\ \exists y \in B : xRy.$$

Entsprechend heißt die Relation R **rechtsvollständig**, falls jedes $y \in B$ zu mindestens einem $x \in A$ in der Relation R steht, d.h.

$$\forall y \in B\ \exists x \in A : xRy.$$

(2) Die Relation R heißt **linkseindeutig**, falls jedes $y \in B$ zu höchstens einem $x \in A$ in Relation steht, d.h.

$$\forall y \in B\ \forall x, \tilde{x} \in A : xRy \text{ und } \tilde{x}Ry \Rightarrow x = \tilde{x}.$$

Entsprechend heißt die Relation R **rechtseindeutig**, falls jedes $x \in A$ zu höchstens einem $y \in B$ in Relation steht, d.h.

$$\forall x \in A\ \forall y, \tilde{y} \in B : xRy \text{ und } xR\tilde{y} \Rightarrow y = \tilde{y}.$$

Beispiel 4.8. (1) a) Auf der Menge A aller Autoren und der Menge B aller Bücher betrachten wir

$$xRy \Leftrightarrow x \text{ ist Autor des Buches } y.$$

R ist linksvollständig (Jeder Autor hat ein Buch verfasst) und rechtsvollständig (Jedes Buch hat einen Autor), aber weder links- noch rechtseindeutig (Ein Buch kann mehrere Autoren haben, ein Autor mehrere Bücher verfassen). Ist M die Menge aller Menschen, so ist die Relation $R \subseteq M \times B$ nicht mehr linksvollständig.
b) Ein Geschäft bietet Waren $x \in W$ aus einem Sortiment W zu verschiedenen Preisen $y > 0$ an, diese Preisbildung definiert eine Relation $R \subseteq W \times (0, \infty)$:

$$xRy \Leftrightarrow \text{Ware } x \text{ hat den Preis } y.$$

Man stellt Preise in einer Tabelle dar, denn der Preis jeder Ware ist *eindeutig*: Jede Ware hat mindestens einen Preis (R ist linksvollständig) und höchstens einen Preis (R ist rechtseindeutig). Umgekehrt kann ein Preis gar nicht oder mehrfach vergeben sein, R ist also weder rechtsvollständig noch linkseindeutig.
c) Sei N die Menge aller möglichen Matrikelnummern am KIT und $R \subseteq M \times N$ die Relation

$$xRy \Leftrightarrow \text{ Person } x \text{ hat die Matrikelnummer } y.$$

R ist weder weder links- noch rechtsvollständig, aber links- und rechtseindeutig: Jeder Student hat eine eindeutige Matrikelnummer und kann anhand der Matrikelnummer identifiziert werden, die Zuordnung R : Student $\mapsto$ Matrikelnummer ist sogar *eineindeutig*.

(2) Im Pfeildiagramm ist die Relation „x ist das Quadrat von y“ auf der Menge $\{0, 1, 2, 3, 4\} \times \{-3, -2, -1, 0, 1\}$ dargestellt. Sie ist weder linksvollständig (von $2, 3$ geht kein Pfeil aus) noch rechtsvollständig (bei -3 kommt kein Pfeil an); die Einschränkung auf $\{0, 1, 2, 3, 4\} \times \{-2, -1, 0, 1\}$ wäre rechtsvollständig. Außerdem ist die Relation linkseindeutig (zum Pfeilende ist der Startpunkt eindeutig), aber nicht rechtseindeutig (von 1 gehen zwei Pfeile aus).

(3) Eine Relation $R \subseteq A \times B$ lässt sich als **Zuordnung** von Elementen $x \in A$ auf Elemente $y \in B$ verstehen, xRy bedeutet „x wird y durch R zugeordnet“. Jedem $x \in A$ wird durch eine linksvollständige Relation *mindestens* ein $y \in B$ zugeordnet, durch eine rechtseindeutige Relation *höchstens* ein $y \in B$, bei beiden Eigenschaften also *genau ein* $y \in B$. Genauso wird jedes $y \in B$ durch eine rechtsvollständige Relation *mindestens* einem $x \in A$ zugeordnet, durch eine linkseindeutige Relation *höchstens* einem $x \in A$, bei beiden Eigenschaften *genau einem* $x \in A$.

(4) Die Begriffe aus Definition 4.3 sind für Relationen $R \subseteq A \times B$ nicht sinnvoll; aber die Begriffe aus Definition 4.7 kann man auf Relationen $R \subseteq A^2$ anwenden:
a) Der diagonale „Spalt“ $\{(x, y) \in \mathbb{R}^2 : |x - y| > 1\}$ ist eine symmetrische Relation auf $\mathbb{R}$; er ist auch links- und rechtsvollständig, da jedem $x \in \mathbb{R}$ (unter anderem) das Element $x + 2 \in \mathbb{R}$ zugeordnet wird, jedem $y \in \mathbb{R}$ das Element $y - 2 \in \mathbb{R}$. Man sieht die Links-(Rechts-)vollständigkeit daran, dass jede Parallele zur y-(zur x-)Achse den Spalt mindestens einmal schneidet.

b) Auch die *Normalparabel* $y = x^2$ und ihre Spiegelung an der ersten Winkelhalbierenden $x = y^2$ sind Relationen auf $\mathbb{R}$:

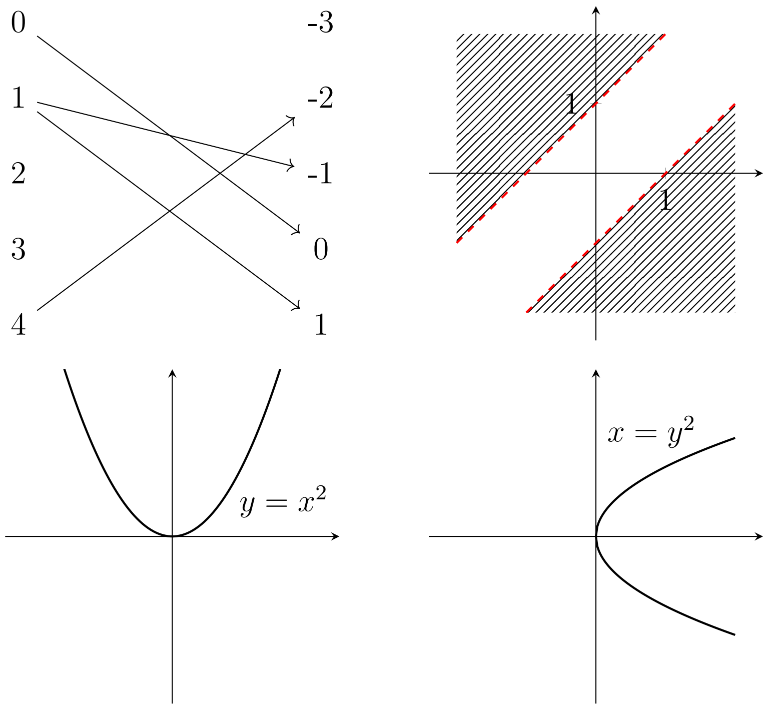
$$R_x = \{(x, y) \in \mathbb{R}^2 : y = x^2\} = \{(x, x^2) : x \in \mathbb{R}\} \subseteq \mathbb{R}^2,$$
$$R_y = \{(x, y) \in \mathbb{R}^2 : y^2 = x\} = \{(y^2, y) : y \in \mathbb{R}\} \subseteq \mathbb{R}^2.$$

Jede Parallele zur y-Achse schneidet die Parabel $y = x^2$ genau einmal, daher ist R_x linksvollständig und rechtseindeutig; R_y hat keine dieser beiden Eigenschaften. Umgekehrt schneidet jede Parallele zur x-Achse $y^2 = x$ genau einmal, daher ist R_y rechtsvollständig und linkseindeutig; R_x hat keine dieser beiden Eigenschaften.
Bei R_x kann man sinnvoll sagen, dass jedem x ein y eindeutig zugeordnet wird, bei R_y geschieht diese Zuordnung nicht immer und nicht eindeutig: Zum Beispiel ist $(-1, y) \notin R_y$ für alle $y \in \mathbb{R}$, aber sowohl $(1, 1) \in R_y$ als auch $(1, -1) \in R_y$. Relationen wie R_x sind von besonderer Bedeutung und heißen **Abbildungen**; in dieser Sprechweise hat jedes **Urbild** x ein (eindeutig bestimmtes) **Bild** y unter der Abbildung R_x, bezeichnet mit $R_x(x)$:

$$R_x = \{(x, R_x(x)) \in \mathbb{R}^2 : R_x(x) = x^2\}.$$

Bei R_y wäre die Schreibweise $R_y(x)$ nicht sinnvoll, da es zum Beispiel $R_y(-1)$ nicht gibt und das Symbol $R_y(1)$ sowohl -1 als auch 1 bezeichnen würde.[2]

[2] Die Asymmetrie zwischen R_x und R_y beruht auf der Konvention, dass Abbildungen jedem $x \in A$ ein $y \in B$ zuordnen. Ordnet man stattdessen jedem $y \in B$ ein $x \in A$ zu, so vertauschen sich die Rollen von R_x und R_y; daher bezeichnet man R_y auch als **Umkehrrelation** von R_x.



Definition 4.9. Seien A, B Mengen und $f \subseteq A \times B$ eine Relation zwischen A und B. Ist f linksvollständig und rechtseindeutig, so heißt f eine **Abbildung** von A nach B und man schreibt $f : A \to B$. In diesem Fall wird durch die Abbildung f jedem $x \in A$ *genau ein* $y \in B$ zugeordnet; das zu $x \in A$ gehörige $y \in B$ bezeichnet man mit dem Symbol $f(x)$ (lies: „f von x“), statt $(x, y) \in f$ oder xfy schreibt man

$$x \mapsto f(x) \text{ (\textit{Abbildungsvorschrift})} \qquad \text{oder} \qquad y = f(x) \text{ (\textit{Funktionsgleichung})}.$$

A heißt **Definitionsmenge** oder Definitionsbereich, B **Wertemenge** oder Wertebereich von f; in der Zuordnung $x \mapsto f(x)$ heißt $x \in A$ **Stelle**, **Argument** oder **Urbild** von $f(x) \in B$ und $f(x) \in B$ heißt **Wert** oder **Bild** von $x \in A$ unter der Abbildung f.
Sind A, B Zahlenmengen, zum Beispiel $f : \mathbb{R} \to \mathbb{R}$, so heißt f auch eine **Funktion** und $f(x) \in B$ ein **Funktionswert** von f an der Stelle $x \in A$.
Die Menge aller Abbildungen $f : A \to B$ wird mit $\mathrm{Abb}(A, B)$ oder B^A bezeichnet.

Beispiel 4.10. (1) Die Normalparabel $y = x^2$ ist eine linksvollständige und rechtseindeutige Relation $R_x \subseteq \mathbb{R}^2$, also eine Abbildung $R_x : \mathbb{R} \to \mathbb{R}$. Wir schreiben f statt R_x; f ordnet jeder Stelle x aus der Definitionsmenge $\mathbb{R}$ den (eindeutig bestimmten) Wert x^2 aus der Wertemenge $\mathbb{R}$ zu, die Abbildungsvorschrift ist also $x \mapsto x^2$, die Funktionsgleichung $f(x) = x^2$.

f ist eine Menge von Paaren $f = \{(x, y) \in \mathbb{R} \times \mathbb{R} : y = x^2\}$; genauso lässt sich jede Abbildung $f : A \to B$ schreiben:

$$f = \{(x, y) \in A \times B : y = f(x)\} \subseteq A \times B. \tag{4.2}$$

Stellt man diese Menge im Koordinatensystem dar, so entsteht der Graph von f; in diesem Sinne sind die Abbildung f und ihr Graph dasselbe, nämlich die Menge (4.2). Häufig sieht man eine Abbildung stattdessen als dreiteiliges Objekt an, bestehend aus Definitionsmenge A, Wertemenge B und Abbildungsvorschrift $x \mapsto f(x)$:

$$f : \mathbb{R} \to \mathbb{R}, f(x) = x^2.$$

f bezeichnet dann die gesamte Abbildung – im Unterschied zum **Graphen** von f, der aus allen Paaren $(x, f(x))$ besteht:

$$\mathrm{gr}(f) = \{(x, f(x)) : x \in A\} \subseteq A \times B.$$

Zwischen den Auffassungen einer Abbildung als Menge (4.2) und als Tripel $(A, B, x \mapsto f(x))$ wechselt man zweckmäßig in der Mathematik.

(2) Die Abbildung $g : \mathbb{R} \to [0, \infty), g(x) = x^2$ hat dieselbe Funktionsgleichung und denselben Graphen wie f, ist im Gegensatz zu f aber rechtsvollständig; f und g sind *nicht* dieselbe Abbildung. Den Unterschied zwischen f und der Abbildung $h : [0, \infty) \to \mathbb{R}, h(x) = x^2$ sieht man auch am Graphen: h ist nur der rechte Zweig der Normalparabel. Man nennt h auch die **Einschränkung** von f auf $[0, \infty)$; f ist eine mögliche **Fortsetzung** von h auf $\mathbb{R}$, eine andere ist die *abschnittweise definierte* Funktion

$$\tilde{h} : \mathbb{R} \to \mathbb{R}, \tilde{h}(x) := \begin{cases} 0, & \text{falls } x < 0, \\ x^2, & \text{falls } x \geqslant 0. \end{cases}$$

Allgemein sind zwei Abbildungen erst dann **gleich**, wenn Definitionsmenge, Wertemenge und Abbildungsvorschrift übereinstimmen; daher auch die Auffassung $(A, B, x \mapsto f(x))$ in (1). So wie gleiche Mengen unterschiedliche Darstellungen haben können, geht das auch für Abbildungen: $u : \{0, 1\} \to \{0, 1\}, u(x) = x^2$ und $v : \{0, 1\} \to \{0, 1\}, v(x) = x$ sind dieselbe Abbildung, da sie auf gleicher Definitions- und Wertemenge dieselben Zuordnungen $0 \mapsto 0$ und $1 \mapsto 1$ umfassen.[3] Auch die Funktion

$$l : (0, \infty) \to \mathbb{R}, y \mapsto x,$$
$$\text{wobei } x \text{ die eindeutige Lösung der Gleichung } e^x = y \text{ ist,}$$

ist dieselbe wie der Logarithmus $\log : (0, \infty) \to \mathbb{R}$.

(3) Bei Änderungen von Definitions- oder Wertemenge, abschnittsweisen oder impliziten Abbildungsvorschriften etc. besteht die Gefahr, dass die definierte Relation keine Abbildung, also nicht linksvollständig oder nicht rechtseindeutig ist; dann wäre auch die Schreibweise $f : A \to B, x \mapsto f(x)$ nicht definiert. Um diesen Umstand zu umgehen, ist die Sprechweise „die Abbildung f ist **wohldefiniert**" üblich; das bedeutet einfach, dass die Relation f tatsächlich eine Abbildung ist.

[3] Im Gegensatz dazu sind die Polynome x^2 und x natürlich *nicht* gleich.

Zum Beispiel ist

$$g : \mathbb{R} \to (0, \infty), g(x) = x^2$$

keine wohldefinierte Abbildung, da $g(0) = 0$ nicht in der Wertemenge $(0, \infty)$ liegt, g ist also nicht linksvollständig. Genauso ist

$$h : [0, 2] \to \mathbb{R}, h(x) = \begin{cases} x^2 + 1, & \text{falls } x \in [0, 1], \\ 1 - x, & \text{falls } x \in [1, 2], \end{cases}$$

keine wohldefinierte Funktion: Es ist sowohl $h(1) = 2$ nach der ersten Zeile als auch $h(1) = 0$ nach der zweiten Zeile, h also nicht rechtseindeutig.[4] Die Funktion

$$k : \mathbb{R} \to \mathbb{R}, k(x) = \frac{1}{x}$$

hat an der Stelle $x = 0$ eine *Definitionslücke*; sie ist nicht wohldefiniert, also eigentlich gar keine Funktion. Wohldefiniert ist zum Beispiel die Einschränkung von k auf $\mathbb{R} \setminus \{0\}$ oder

$$\tilde{k} : \mathbb{R} \to \mathbb{R}, \tilde{k}(x) = \begin{cases} \frac{1}{x}, & \text{falls } x \neq 0, \\ 0, & \text{falls } x = 0. \end{cases}$$

(4) Nicht jede Teilmenge des $\mathbb{R}^2$ lässt sich als Graph einer Funktion auffassen: Die Kreislinie

$$\{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 = r^2\}$$

mit Radius $r > 0$ und Mittelpunkt (x_0, y_0) ist *kein* Graph, da manche Parallelen zur y-Achse mehr als einen Schnittpunkt mit ihr haben, manche keinen. Wir beschränken uns daher auf $x \in [x_0 - r, x_0 + r]$ und zerlegen die Kreislinie in

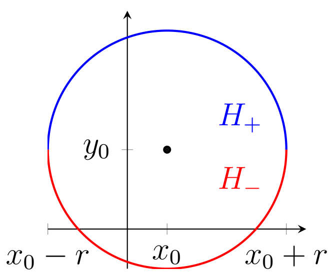
$$H_+ := \{(x, y) \in [x_0 - r, x_0 + r] \times [y_0, \infty) : (x - x_0)^2 + (y - y_0)^2 = r^2\},$$
$$H_- := \{(x, y) \in [x_0 - r, x_0 + r] \times (-\infty, y_0] : (x - x_0)^2 + (y - y_0)^2 = r^2\}.$$

Im Halbkreis H_+ kann man die Kreisgleichung auflösen zu $y = y_0 + \sqrt{r^2 - (x - x_0)^2}$, d.h. H_+ ist der Graph der Funktion

$$f_+ : [x_0 - r, x_0 + r] \to [y_0, \infty), f_+(x) = y_0 + \sqrt{r^2 - (x - x_0)^2}.$$

Entsprechend ist H_- der Graph von

$$f_- : [x_0 - r, x_0 + r] \to (-\infty, y_0], f_-(x) = y_0 - \sqrt{r^2 - (x - x_0)^2}.$$



[4] Relationen wie g heißen gelegentlich eine *partielle Funktion*, h eine *Multifunktion*, auch wenn beide *keine* Funktionen sind.

(5) Die Untersuchung von Funktionen ist ein wesentlicher Gegenstand der *Analysis*; hier einige Beispiele interessanter Eigenschaften einer Funktion $f : \mathbb{R} \to \mathbb{R}$:

a) f heißt **periodisch** mit *Periode* $p > 0$, falls $f(x + p) = f(x)$ für alle $x \in \mathbb{R}$ gilt. Beispiele dafür sind die trigonometrischen Funktionen.

b) f heißt **gerade**, falls stets $f(-x) = f(x)$ gilt, also der Graph symmetrisch zur y-Achse ist. Beispiele dafür sind der Cosinus, die Betragsfunktion oder Polynomfunktionen, in denen nur gerade Exponenten auftreten.

c) f heißt **ungerade**, falls stets $f(-x) = -f(x)$ gilt, also der Graph symmetrisch zum Ursprung ist. Beispiele dafür sind der Sinus, die Signumfunktion oder Polynomfunktionen, in denen nur ungerade Exponenten auftreten.

d) Falls aus $x < y$ stets

- $f(x) \leqslant f(y)$ folgt, so heißt f **monoton wachsend**.
- $f(x) \geqslant f(y)$ folgt, so heißt f **monoton fallend**.
- $f(x) < f(y)$ folgt, so heißt f **streng monoton wachsend**.
- $f(x) > f(y)$ folgt, so heißt f **streng monoton fallend**.

Zum Beispiel ist die Funktion $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ weder monoton wachsend noch fallend ($f(-1) > f(0)$ und $f(0) < f(1)$); aber die Einschränkung $f : [0, \infty) \to \mathbb{R}, g(x) = x^2$ ist streng monoton wachsend: Für $x, y > 0$ gilt

$$x < y \Rightarrow f(x) - f(y) = x^2 - y^2 = (x + y)(x - y) < 0 \Rightarrow f(x) < f(y).$$

Die Exponentialfunktion $f(x) = a^x$ ist für $a \geqslant 1$ monoton wachsend, für $0 < a \leqslant 1$ monoton fallend; dasselbe gilt für den Logarithmus. Der Sinus ist weder monoton wachsend noch fallend, konstante Funktionen sind beides.

e) Gibt es eine Zahl $C \geqslant 0$, sodass $-C \leqslant f(x) \leqslant C$ für alle $x \in \mathbb{R}$ gilt, so heißt f **beschränkt**, andernfalls **unbeschränkt**. Zum Beispiel ist der Sinus beschränkt, Potenz- und Exponentialfunktionen unbeschränkt. Jede monoton wachsende Funktion $f : [a, b] \to \mathbb{R}$ ist beschränkt: Es gilt $f(a) \leqslant f(x) \leqslant f(b)$ für alle $x \in [a, b]$ und man kann $C = \max\{-f(a), f(b)\} \geqslant 0$ wählen.

(6) Abbildungen sind ein zentrales Konzept der gesamten Mathematik, die folgenden Beispiele sollen das illustrieren und werden im Mathematikstudium vertieft.

a) Alle Arten von Rechenoperationen und **Verknüpfungen** sind Abbildungen $* : M \times M \to M$, zum Beispiel Addition und Multiplikation

$$+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (x, y) \mapsto x + y,$$
$$\cdot : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (x, y) \mapsto xy.$$

$+$ und $\cdot$ sind *zweistellige* Abbildungen, weil sie ein Paar von Argumenten (x, y) entgegennehmen; genauso gibt es auch *mehrstellige* Abbildungen $f(x_1, \ldots, x_n)$. Eine Teilmenge $N \subseteq M$ heißt *abgeschlossen* unter der Verknüpfung $*$, falls $* : N \times N \to N$ wohldefiniert ist.

b) Geometrische Operationen wie **Spiegelungen**, **Drehungen**, **Verschiebungen** oder **zentrische Streckungen** sind Abbildungen der Ebene $\mathbb{R}^2$ in sich. Ist zum Beispiel σ die Spiegelung an einer Achse g und Δ ein Dreieck, so ist $\sigma(\Delta) = \Delta'$ das Bilddreieck und $\sigma(\sigma(\Delta)) = \Delta$. Für Drehungen D_α um den Winkel α um ein gemeinsames Drehzentrum Z gilt $D_\alpha(D_\beta(\Delta)) = D_{\alpha+\beta}(\Delta)$.

c) Abbildungen der Form $a : \mathbb{N} \to \mathbb{R}$ heißen eine (reelle) **Folge**, man schreibt a_n statt $a(n)$ und (a_n) statt a. Die Folge (a_n) kann durch ein *Bildungsgesetz* in expliziter ($a_n = \sqrt{n+1} - \sqrt{n}$) oder impliziter Form ($a_{n+2} = a_{n+1} + a_n$, $a_0 = a_1 = 1$) gegeben sein; mit Folgen beschreibt man Wachstumsprozesse, aber auch Iterationen und Algorithmen.

d) $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ ist der Ergebnisraum beim zweifachen Würfelwurf, die Abbildung $S : \Omega \to \mathbb{R}, S(x, y) = x + y$ die Augensumme bei diesem *Zufallsexperiment*. In komplexeren Situationen kann eine **Zufallsvariable** S den Gewinn bei einem Spiel, den Verlauf eines Aktienkurses oder die Ausschüttungshöhe einer Rentenversicherung beschreiben.

e) Sei $D(\mathbb{R}, \mathbb{R})$ die Menge aller differenzierbaren Funktionen $f : \mathbb{R} \to \mathbb{R}$, dann ist die **Ableitung** $' : D(\mathbb{R}, \mathbb{R}) \to \mathrm{Abb}(\mathbb{R}, \mathbb{R}), f \mapsto f'$ eine Abbildung. Entsprechend sind das (bestimmte) **Integral** und die Integralfunktion Abbildungen auf der Menge $I(\mathbb{R}, \mathbb{R})$ aller integrierbaren Funktionen:

$$\begin{aligned} \int_a^b &: I(\mathbb{R}, \mathbb{R}) \to \mathbb{R}, f \mapsto \int_a^b f \,\mathrm{d}x, \\ \int_a &: I(\mathbb{R}, \mathbb{R}) \to \mathrm{Abb}(\mathbb{R}, \mathbb{R}), f \mapsto I_f, \\ &\quad \text{definiert durch } I_f : \mathbb{R} \to \mathbb{R}, I_f(x) = \int_a^x f \,\mathrm{d}t. \end{aligned}$$

Abbildungen wie $\int_a^b$, die Funktionen auf Zahlen abbilden, heißen **Funktionale**; Abbildungen wie die Ableitung oder $\int_a$, die Funktionen auf Funktionen abbilden, heißen **Operatoren**.

Beispiel 4.11 (Reelle Funktionen)**.** Wir stellen nun einige Klassen *reeller* Funktionen $f : \mathbb{R} \to \mathbb{R}$ dar, die zumeist aus der Schule bekannt sind.
Man versuche, sich den häufig unsauberen Umgang mit Funktionen bewusst zu machen und zu vermeiden: Bei der „Funktion $f(x) = x^2$“ fehlen Definitions- und Wertemenge, man wählt dann gewöhnlich die größtmöglichen Mengen $f : \mathbb{R} \to \mathbb{R}$; auch bei der „Funktion $f(x) = 2x - 1$ für $x \in \mathbb{R}$“ fehlt die Wertemenge. Man unterscheide genau zwischen der Funktion f und einem Element des Wertebereichs $f(x)$: „Die Funktion $f(x)$“ (zum Beispiel „die Funktion x^2“) ist eine sehr unsaubere Sprechweise.

(1) Die Funktionen $f : \mathbb{R} \to \mathbb{R}, f(x) = ax + b$ mit $a, b \in \mathbb{R}$ heißen **affin-lineare Funktionen**, ihre Graphen sind *Geraden*. Im Fall $a = 0$ erhalten wir die **konstante**

Funktion $\mathbb{R} \to \mathbb{R}, x \mapsto b$, deren Graph parallel zur x-Achse verläuft; Geraden der Form $x = b$ parallel zur y-Achse sind keine Graphen, da sie als Relation nicht rechtseindeutig sind. Im Fall $b = 0$ ist der Graph eine *Ursprungsgerade* und f eine **lineare Funktion**, erfüllt also die Eigenschaften[5]

$$\forall x, y, \lambda \in \mathbb{R} : f(x+y) = f(x) + f(y), \quad f(\lambda x) = \lambda f(x).$$

Eine besondere lineare Funktion ist die **Identität**

$$\mathrm{id} : \mathbb{R} \to \mathbb{R}, x \mapsto x,$$

deren Graph die erste Winkelhalbierende $y = x$ ist.

(2) Die Funktionen $f : \mathbb{R} \to \mathbb{R}$ der Form $f(x) = ax^2 + bx + c$ heißen *quadratisch*, der Form $f(x) = ax^3 + bx^2 + cx + d$ *kubisch*, allgemein der Form

$$f(x) = a_n x^n + \ldots + a_1 x + a_0$$

eine **Polynomfunktion** oder *ganzrationale Funktion*. Im Fall $a_n \neq 0$ heißt $n \in \mathbb{N}_0$ auch der **Grad** von f; die Eigenschaften von f hängen eng mit den Nullstellen und Vorzeichenwechseln des Polynoms $a_n x^n + \ldots + a_0$ zusammen, siehe Abschnitt 3.2. Die einfachsten Polynomfunktionen $p_n : \mathbb{R} \to \mathbb{R}, p_n(x) = x^n$ heißen **Monomfunktionen**: $p_0(x) = 1$ ist konstant, $p_1(x) = x$ die Identität, $p_2(x) = x^2$ usw.

(3) Die Funktion $f : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}, f(x) = \frac{1}{x}$ heißt **Inversion**, ihr Graph ist eine *Hyperbel*. f hat eine *Definitionslücke* bei $x = 0$, kann dort also nicht durch die Funktionsgleichung $f(x) = \frac{1}{x}$ (wohl-)definiert werden. Durch Inversion von Polynomfunktionen entstehen **(gebrochen-)rationale Funktionen**

$$f : D \to \mathbb{R}, f(x) = \frac{a_n x^n + \ldots + a_1 x + a_0}{b_m x^m + \ldots + b_1 x + b_0},$$

die Definitionsmenge $D = \{x \in \mathbb{R} : b_m x^m + \ldots + b_1 x + b_0 \neq 0\}$ hat Lücken bei den Nullstellen des Nennerpolynoms. Je nach Verhältnis der Grade n, m können diese Definitionslücken von f *hebbar* oder nicht hebbare *Pole* sein; für $x \to \pm\infty$ können sich *Asymptoten* an andere Polynomfunktionen ergeben.[6]

(4) Die **Potenzfunktion** $f(x) = x^a$ hat eine unterschiedliche Definitions- und Wertemenge je nach dem Zahlenbereich, aus dem der Exponent a kommt:

$$\begin{aligned} \text{Für } a \in \mathbb{N}_0 \text{ ist } f \text{ eine Abbildung } & f : \mathbb{R} \to \mathbb{R}, \\ \text{Für } a \in \mathbb{Z} \text{ ist } f \text{ eine Abbildung } & f : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}, \\ \text{Für \textit{nichtnegative} } a \in \mathbb{Q} \text{ oder } a \in \mathbb{R} \text{ ist } f \text{ eine Abbildung } & f : [0, \infty) \to [0, \infty), \\ \text{Für beliebige } a \in \mathbb{Q} \text{ oder } a \in \mathbb{R} \text{ ist } f \text{ eine Abbildung } & f : (0, \infty) \to (0, \infty). \end{aligned}$$

$a \in \mathbb{N}_0$ entspricht gerade den Monomfunktionen p_n; ihre Umkehrfunktionen heißen **Wurzelfunktionen** $w_n : [0, \infty) \to [0, \infty), w_n(x) = \sqrt[n]{x}$, siehe Beispiel 4.23(1).

[5] Im Fall $a \neq 0$ sind affin-lineare Funktionen *nicht* linear.

[6] Dabei sind Polynomdivision und Partialbruchzerlegung wichtige Verfahren, siehe Anhang.

(5) Analog hängen auch bei der **Exponentialfunktion** $f(x) = a^x$ Definitions- und Wertemenge vom Zahlenbereich ab, aus dem die Basis a kommt:

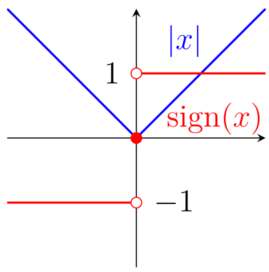
$$\begin{aligned} \text{Für positive } a > 0 \text{ ist } f \text{ eine Abbildung } f &: \mathbb{R} \to \mathbb{R}, \\ \text{Für nichtnegative } a \geqslant 0 \text{ ist } f \text{ eine Abbildung } f &: [0, \infty) \to [0, \infty), \\ \text{Für } a \neq 0 \text{ ist } f \text{ eine Abbildung } f &: \mathbb{Z} \to \mathbb{Z}, \\ \text{Für beliebige } a \in \mathbb{R} \text{ ist } f \text{ eine Abbildung } f &: \mathbb{N}_0 \to \mathbb{N}_0. \end{aligned}$$

Für $a > 0$ ist die Umkehrfunktion von f die **Logarithmusfunktion** (zur Basis a) $\log_a : (0, \infty) \to \mathbb{R}$. Zur Basis der **Eulerschen Zahl** $e \approx 2{,}718$ heißen $f(x) = e^x$ und $\log(x)$ die *natürliche* Exponential- und Logarithmusfunktion.

(6) Aus Abschnitt 3.1 kennen wir **Betrag-** und **Signumfunktion**:[7]

$$|\cdot| : \mathbb{R} \to [0, \infty), \qquad |x| = \begin{cases} x, & \text{falls } x \geqslant 0, \\ -x, & \text{falls } x < 0, \end{cases}$$

$$\operatorname{sign} : \mathbb{R} \to \{-1, 0, 1\}, \quad \operatorname{sign}(x) = \begin{cases} 1, & \text{falls } x > 0, \\ 0, & \text{falls } x = 0, \\ -1, & \text{falls } x < 0. \end{cases}$$



sign hat eine *Sprungstelle* bei $x = 0$ und ist daher *unstetig*; die Betragsfunktion hat einen Knick bei $x = 0$ und ist ist daher *stetig*, aber *nicht differenzierbar*. Außerhalb von $x = 0$ ist $\operatorname{sign}(x) = \frac{x}{|x|} = \pm 1$ die Ableitung der Betragsfunktion.

(7) Vom Dreieck kennt man die **trigonometrischen Funktionen Sinus** und **Cosinus**, sie können am Einheitskreis auf $[0, 2\pi]$ fortgesetzt werden.[8] Diese setzt man wiederum zu 2π-periodischen Funktionen $\sin : \mathbb{R} \to \mathbb{R}$ und $\cos : \mathbb{R} \to \mathbb{R}$ fort.

- Es gilt der „*trigonometrische Pythagoras*“ $\sin^2(x) + \cos^2(x) = 1$ und die Beschränktheit
$$-1 \leqslant \sin(x) \leqslant 1, \quad -1 \leqslant \cos(x) \leqslant 1.$$

- Der Sinus $\sin(-x) = -\sin(x)$ ist ungerade, der Kosinus $\cos(-x) = \cos(x)$ gerade und es gelten die Zusammenhänge

$$\begin{aligned} \sin\left(x + \frac{\pi}{2}\right) &= \cos(x), & \cos\left(x + \frac{\pi}{2}\right) &= -\sin(x), \\ \sin(x + \pi) &= -\sin(x), & \cos(x + \pi) &= -\cos(x). \end{aligned}$$

[7] Man schreibt $|\cdot|$ für die *Funktion*, die an jeder Stelle x den *Funktionswert* $|x|$ hat.

[8] Dabei misst man Winkel zweckmäßig im Bogenmaß $b \in [0, 2\pi]$ statt im Gradmaß $\alpha \in [0°, 360°]$; es gilt $b = 2\pi \cdot \frac{\alpha}{180°}$.

- Es gelten die *Additionstheoreme*

$$\sin(x \pm y) = \sin(x)\cos(y) \pm \cos(x)\sin(y),$$
$$\cos(x \pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y)$$

  und daher die nützlichen Formeln

$$\sin(2x) = 2\sin(x)\cos(x), \qquad \cos(2x) = \cos^2(x) - \sin^2(x),$$
$$\sin^2(x) = \frac{1}{2}(1-\cos(2x)), \qquad \cos^2(x) = \frac{1}{2}(1+\cos(2x)).$$

- *Wichtige Funktionswerte* zeigt die folgende Tabelle:

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	π
$\sin(x)$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	0
$\cos(x)$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	-1

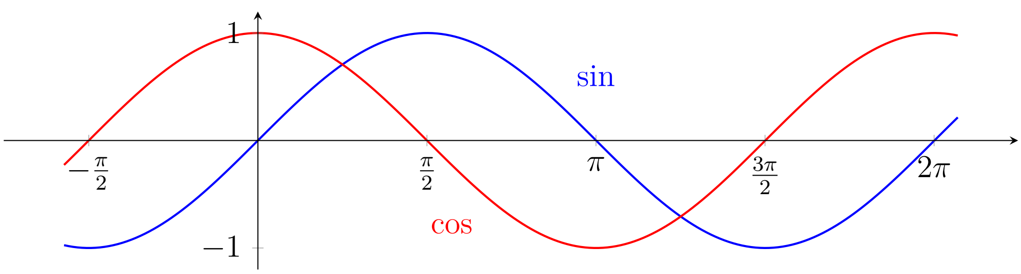
  Daraus kann man weitere Funktionswerte bestimmen: Zum Beispiel für $x := \sin(\frac{\pi}{12}) > 0$ gilt $\frac{1}{4} = \sin^2(2 \cdot \frac{\pi}{12}) = 4\sin^2(\frac{\pi}{12})\cos^2(\frac{\pi}{12}) = 4x^2(1-x^2)$ und diese Gleichung hat die positive Lösung $x = \frac{1}{2}\sqrt{2-\sqrt{3}}$.

Die Nullstellenmengen von $\sin, \cos$ sind $N_s = \{x \in \mathbb{R} : \sin(x) = 0\} = \{k\pi : k \in \mathbb{Z}\}$ und $N_c = \{x \in \mathbb{R} : \cos(x) = 0\} = \{\frac{\pi}{2} + k\pi : k \in \mathbb{Z}\}$. Damit definiert man **Tangens** und **Cotangens**:

$$\tan : \mathbb{R} \setminus N_c \to \mathbb{R}, \tan(x) = \frac{\sin(x)}{\cos(x)}, \qquad \cot : \mathbb{R} \setminus N_s \to \mathbb{R}, \cot(x) = \frac{\cos(x)}{\sin(x)}.$$

Beide Funktionen sind π-periodisch und ungerade. Geeignete Einschränkungen der trigonometrischen Funktionen haben Umkehrfunktionen, die **Arkusfunktionen**:

$$\arcsin : [-1,1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \qquad \arccos : [-1,1] \to [0,\pi],$$
$$\arctan : \mathbb{R} \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \qquad \operatorname{arccot} : \mathbb{R} \to (0,\pi).$$



(8) An der Hyperbel $x^2 - y^2 = 1$ (statt dem Einheitskreis), mithilfe der Eulerschen Formel $e^{iz} = \cos(z) + i\sin(z)$ $(z \in \mathbb{C})$ oder der Exponentialfunktion definiert man analog zu (7) **hyperbolische Funktionen**, etwa *Sinus und Cosinus hyperbolicus*:

$$\sinh : \mathbb{R} \to \mathbb{R}, \sinh(x) = \frac{1}{2}\left(e^x - e^{-x}\right), \quad \cosh : \mathbb{R} \to \mathbb{R}, \cosh(x) = \frac{1}{2}\left(e^x + e^{-x}\right).$$

Es gelten analoge Eigenschaften wie $\cosh(x)^2 - \sinh(x)^2 = 1$, Additionstheoreme oder die Ableitungsregeln $\sinh'(x) = \cosh(x), \cosh'(x) = \sinh(x)$. Auch Umkehrfunktionen existieren, die **Areafunktionen**:

$$\operatorname{arsinh} : \mathbb{R} \to \mathbb{R}, \operatorname{arsinh}(x) = \log\left(x + \sqrt{x^2+1}\right),$$
$$\operatorname{arcosh} : [1,\infty) \to [0,\infty), \operatorname{arcosh}(x) = \log\left(x + \sqrt{x^2-1}\right).$$

4.2 Eigenschaften von Abbildungen

Definition 4.12. Seien A, B Mengen, $f : A \to B$ eine Abbildung und $X \subseteq A, Y \subseteq B$.

(1) Die Menge $f(X) := \{f(x) : x \in X\} \subseteq B$ heißt das **Bild** von X unter f. Die Menge $f(A) \subseteq B$ heißt das *Bild von* f.

(2) Die Menge $f^{-1}(Y) := \{x \in A : f(x) \in Y\} \subseteq A$ heißt das **Urbild** von Y unter f. Für jede Abbildung gilt $f^{-1}(B) = A$.

Beispiel 4.13. (1) a) Das Bild der Funktion $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ ist die Menge

$$f(\mathbb{R}) = \{f(x) : x \in \mathbb{R}\} = \{x^2 : x \in \mathbb{R}\} = [0,\infty),$$

denn („$\subseteq$") jeder Funktionswert x^2 liegt in $[0,\infty)$ und („$\supseteq$") jede Zahl $y \in [0,\infty)$ ist ein Funktionswert $y = f(\sqrt{y})$; das Bild $[0,\infty)$ ist etwas anderes als die Wertemenge $\mathbb{R}$. Auch die Funktionen $g : \mathbb{R} \to [0,\infty), g(x) = x^2$ und $h : [0,\infty) \to \mathbb{R}, h(x) = x^2$ haben das Bild $g(\mathbb{R}) = h([0,\infty)) = [0,\infty)$.
b) Das Bild einer Funktion hängt nicht von der Wertemenge, aber von der Definitionsmenge ab: Die Inversion $f : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}, f(x) = \frac{1}{x}$ hat das Bild $f(\mathbb{R} \setminus \{0\}) = \mathbb{R} \setminus \{0\}$, ihre Einschränkung $g : (0,\infty) \to \mathbb{R} \setminus \{0\}, g(x) = \frac{1}{x}$ das Bild $g((0,\infty)) = (0,\infty)$. Das Bild der Einschränkung von f auf $[1,2]$ ist dasselbe wie das Bild von $[1,2]$ unter f, nämlich

$$f([1,2]) = \left\{\frac{1}{x} : x \in [1,2]\right\} = \left[\frac{1}{2}, 1\right].$$

Genauso ist zum Beispiel $f([-1,0) \cup \{3\}) = (-\infty,-1] \cup \{\frac{1}{3}\}$.
c) Für das Bild der Funktion $f : \mathbb{Z} \to \mathbb{Z}, f(x) = x^2$ gilt $f(\mathbb{Z}) \subseteq \mathbb{N}_0$, aber nicht „$\supseteq$", denn zum Beispiel $3 \in \mathbb{N}_0$ ist keine Quadratzahl. Entsprechend ist $f(\mathbb{Z}) \supseteq \{4, 9, 16\}$, aber die Einschränkung $g : \mathbb{Z} \to \{4, 9, 16\}, g(x) = x^2$ nicht wohldefiniert.

d) Im Koordinatensystem findet man das Bild $f(A)$ einer Abbildung $f : A \to B$ auf der y-Achse: Es enthält alle $y \in B$, durch die die Parallele zur x-Achse den Graphen von f (mindestens einmal) schneidet, die also von einem $x \in A$ unter f „getroffen“ werden. Für das Bild $f(X)$ von Teilmengen $X \subseteq A$ bestimmt man entsprechend, welche $y \in B$ von Elementen $x \in X$ „getroffen“ werden. Erinnerung: Schneidet *jede* Parallele zur x-Achse durch ein $y \in B$ den Graphen mindestens einmal, so ist f rechtsvollständig; schneidet jede Parallele höchstens einmal, so ist f linkseindeutig.
e) Der Sinus $\sin : \mathbb{R} \to \mathbb{R}$ hat das Bild $\sin(\mathbb{R}) = [-1, 1]$, ebenso der Cosinus; es gilt $\sin([0, 2\pi]) = [-1, 1]$ und $\sin([0, \pi]) = [0, 1]$. Für die Exponentialfunktion $f : \mathbb{R} \to \mathbb{R}, f(x) = e^x$ ist $f(\mathbb{R}) = (0, \infty)$ und $f((0, \infty)) = (1, \infty)$.

(2) a) Nach dem Urbild $f^{-1}(Y)$ einer Teilmenge $Y \subseteq B$ fragt man: Für welche $x \in A$ liegt $f(x)$ in Y? Für die Funktion $f : \mathbb{R} \to \mathbb{R}, f(x) = 2x$ ist zum Beispiel

$$f^{-1}([-2, 2]) = \{x \in \mathbb{R} : f(x) \in [-2, 2]\} = \{x \in \mathbb{R} : 2x \in [-2, 2]\} = [-1, 1],$$

denn aus $2x \in [-2, 2]$ folgt $x \in [-1, 1]$ und umgekehrt. Genauso bestimmt man $f^{-1}(\{-4\} \cup (0, 2)) = \{-2\} \cup (0, 1)$ und $f^{-1}(\mathbb{R} \setminus \{0\}) = \mathbb{R} \setminus \{0\}$. Es gilt auch $f^{-1}(\mathbb{R}) = \mathbb{R}$, allgemein immer $f^{-1}(B) = A$, denn zur Linksvollständigkeit gehört, dass für jedes $x \in A$ das Bild $f(x)$ in B liegt.
b) Das Urbild einer Menge $Y \subseteq B$ hängt nicht von der Wertemenge, aber von der Definitionsmenge ab: Für $f : \mathbb{R} \to \mathbb{R}, g : \mathbb{R} \to [0, \infty), h : [0, \infty) \to \mathbb{R}$, $f(x) = g(x) = h(x) = x^2$ gilt zum Beispiel

$$\begin{aligned} f^{-1}((0, 4]) &= \{x \in \mathbb{R} : x^2 \in (0, 4]\} = [-2, 2] \setminus \{0\} = [-2, 0) \cup (0, 2], \\ g^{-1}((0, 4]) &= \{x \in \mathbb{R} : x^2 \in (0, 4]\} = [-2, 2] \setminus \{0\} = [-2, 0) \cup (0, 2], \\ h^{-1}((0, 4]) &= \{x \in [0, \infty) : x^2 \in (0, 4]\} = (0, 2]. \end{aligned}$$

Es ist auch $f^{-1}([-1, 4]) = [-2, 2]$, aber $g^{-1}([-1, 4])$ ist nicht definiert, da $[-1, 4]$ keine Teilmenge der Wertemenge $[0, \infty)$ ist.
c) Für die Funktion $f : \mathbb{R} \setminus \{1\} \to \mathbb{R}, f(x) = \frac{1}{1-x}$ gilt $f^{-1}([0, \infty)) \supseteq \{-1, 0, \frac{1}{2}\}$, da die Zahlen $f(-1), f(0)$ und $f(\frac{1}{2})$ nichtnegativ sind. Umgekehrt ist $f^{-1}([0, \infty)) \subseteq (-\infty, 1]$, da für $x \neq 1$ aus der Ungleichung $\frac{1}{1-x} \geqslant 0$ sofort $1 \geqslant x$ folgt.
d) Im Koordinatensystem findet man Urbilder von Teilmengen $Y \subseteq B$ auf der x-Achse: $x \in A$ liegt genau dann in $f^{-1}(Y)$, wenn das zugehörige $f(x)$ in Y liegt. Die Parallele zur y-Achse durch x schneidet den Graphen der Abbildung genau einmal; ein $y \in Y$ kann aber durchaus kein oder mehrere Urbilder haben.
e) Für den Cosinus $\cos : \mathbb{R} \to \mathbb{R}$ ist zum Beispiel[9] $\cos^{-1}(\{2\}) = \emptyset, \cos^{-1}(\{1\}) = \{2k\pi : k \in \mathbb{Z}\}$ und

$$\cos^{-1}([0, 1]) = \ldots \cup \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \cup \left[\frac{3\pi}{2}, \frac{5\pi}{2}\right] \cup \ldots = \bigcup_{k \in \mathbb{Z}} \left[2k\pi - \frac{\pi}{2}, 2k\pi + \frac{\pi}{2}\right].$$

[9]Nicht die Schreibweise $\cos^{-1}$ für das Urbild mit der Umkehrfunktion arccos verwechseln!

(3) Die Lösungen $x \in A$ einer Gleichung $f(x) = y$ kann man als die Urbilder von $y \in B$ unter f auffassen: Die Lösungsmenge ist

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

Hat dieses Urbild kein, ein oder mehrere Elemente, so hat die Gleichung $f(x) = y$ keine, eine oder mehrere Lösungen $x \in A$. Zum Beispiel sind die Urbilder von $y = 1$ unter $f : \mathbb{R} \to \mathbb{R}, f(x) = (x-2)^2$ gerade die Lösungen von $(x-2)^2 = 1$, also

$$f^{-1}(\{1\}) = \{x \in \mathbb{R} : (x-2)^2 = 1\} = \{1, 3\}.$$

(4) Eine Abbildung ordnet jedem x genau ein y zu, aber verschiedene Urbilder x können dasselbe Bild $f(x)$ haben; das Bild $f(X)$ ist also eher „kleiner" als $X \subseteq A$. Umgekehrt kann jedes y kein oder auch mehrere Urbilder haben, das Urbild $Y \subseteq B$ also sowohl „größer" als auch „kleiner" als $f^{-1}(Y)$ sein. Beispiel: Das Bild der Nullfunktion $f : \mathbb{R} \to \mathbb{R}, f(x) = 0$ ist $f(\mathbb{R}) = \{0\}$; andererseits ist $f^{-1}(\{0\}) = \mathbb{R}$ größer und $f^{-1}(\{1\}) = \emptyset$ kleiner als eine einelementige Menge.

(5) Für das Bild $f(A)$ einer Abbildung $f : A \to B$ sind auch die Schreibweisen $\mathrm{Bild}(f)$, $R(f)$ für „range" oder $\mathrm{im}(f)$ für „image" gebräuchlich.
Gelegentlich enthält die Definitionsmenge einer Abbildung auch Mengen, dann kann die Schreibweise $f(X)$ für das Bild von $X \subseteq A$ uneindeutig sein und man schreibt besser $f[X]$: Bei der (etwas künstlichen) Abbildung

$$f : \{1, 2, \{2\}\} \to \mathbb{R}, \qquad 1 \mapsto 2, 2 \mapsto 2, \{2\} \mapsto 1.$$

gilt $f(1) = f(2) = 2$ und $f(\{2\}) = 1$; man könnte $f(\{2\})$ hier (miss-)verstehen als das Bild $f[\{2\}] = \{f(x) : x \in \{2\}\} = \{2\}$.

(6) Wir werden später zu bijektiven Abbildungen $f : A \to B$ eine Umkehrfunktion $f^{-1} : B \to A$ definieren; die Urbilder von f sind dasselbe wie die Bilder von f^{-1}. Falls aber f nicht bijektiv ist, so ist f^{-1} *keine* Funktion und die Schreibweise $f^{-1}(Y)$ steht rein symbolisch für das Urbild. Daher schreibt man gelegentlich auch $f^{-1}(y)$ für die Menge

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

Ist f bijektiv, so enthält $f^{-1}(\{y\})$ nur ein Element, nämlich den Funktionswert der Umkehrfunktion $f^{-1}(y)$; andernfalls kann $f^{-1}(\{y\})$ auch kein oder mehrere Elemente enthalten. Die Schreibweise f^{-1} ist außerdem nicht mit dem Kehrwert $\frac{1}{f}$ aus Definition 4.17 zu verwechseln, eine völlig andere Funktion.

Satz 4.14. Seien A, B Mengen, $f : A \to B$ eine Abbildung, $X, \tilde{X} \subseteq A$ und $Y, \tilde{Y} \subseteq B$.

(1) Das Urbild ist mit Vereinigung, Durchschnitt, Differenz und der Mengeninklusion verträglich, d.h.

$$f^{-1}(Y \cup \tilde{Y}) = f^{-1}(Y) \cup f^{-1}(\tilde{Y}), \qquad f^{-1}(Y \cap \tilde{Y}) = f^{-1}(Y) \cap f^{-1}(\tilde{Y}),$$
$$f^{-1}(Y \setminus \tilde{Y}) = f^{-1}(Y) \setminus f^{-1}(\tilde{Y}), \qquad Y \subseteq \tilde{Y} \Rightarrow f^{-1}(Y) \subseteq f^{-1}(\tilde{Y}).$$

Für das Komplement gilt $f^{-1}(B \setminus Y) = A \setminus f^{-1}(Y)$, also $f^{-1}(Y^C) = f^{-1}(Y)^C$.

(2) Für das Bild gelten lediglich die Eigenschaften

$$f(X \cup \tilde{X}) = f(X) \cup f(\tilde{X}), \qquad f(X \cap \tilde{X}) \subseteq f(X) \cap f(\tilde{X}),$$
$$f(X \setminus \tilde{X}) \supseteq f(X) \setminus f(\tilde{X}), \qquad X \subseteq \tilde{X} \Rightarrow f(X) \subseteq f(\tilde{X}).$$

(3) In der Kombination von Bild und Urbild gilt

$$f(X) \subseteq Y \Leftrightarrow X \subseteq f^{-1}(Y),$$

insbesondere $X \subseteq f^{-1}(f(X))$ und $f(f^{-1}(Y)) \subseteq Y$.

Beweis (1) Für alle $x \in A$ gilt

$$x \in f^{-1}(Y \cup \tilde{Y}) \Leftrightarrow f(x) \in Y \cup \tilde{Y} \Leftrightarrow f(x) \in Y \vee f(x) \in \tilde{Y} \Leftrightarrow x \in f^{-1}(Y) \cup f^{-1}(\tilde{Y}),$$
$$x \in f^{-1}(Y \cap \tilde{Y}) \Leftrightarrow f(x) \in Y \cap \tilde{Y} \Leftrightarrow f(x) \in Y \wedge f(x) \in \tilde{Y} \Leftrightarrow x \in f^{-1}(Y) \cap f^{-1}(\tilde{Y}),$$
$$x \in f^{-1}(Y \setminus \tilde{Y}) \Leftrightarrow f(x) \in Y \setminus \tilde{Y} \Leftrightarrow f(x) \in Y \wedge f(x) \notin \tilde{Y} \Leftrightarrow x \in f^{-1}(Y) \setminus f^{-1}(\tilde{Y}),$$

das zeigt die ersten drei Gleichheiten. Die letzte Implikation folgt aus $x \in f^{-1}(Y) \Rightarrow f(x) \in Y \subseteq \tilde{Y} \Rightarrow x \in f^{-1}(\tilde{Y})$; für das Komplement beachte $f^{-1}(B) = A$.
(2) Aufgrund der Verträglichkeit des Existenzquantors mit $\vee$ gilt zunächst für alle $y \in B$

$$\begin{aligned} y \in f(X \cup \tilde{X}) &\Leftrightarrow \exists x \in X \cup \tilde{X} : f(x) = y \\ &\Leftrightarrow \Big(\exists x \in X : f(x) = y\Big) \vee \Big(\exists \tilde{x} \in \tilde{X} : f(\tilde{x}) = y\Big) \Leftrightarrow y \in f(X) \cup f(\tilde{X}), \end{aligned}$$

also die erste Gleichheit, bei der zweiten Aussage hingegen nur

$$\begin{aligned} y \in f(X \cap \tilde{X}) &\Leftrightarrow \exists x \in X \cap \tilde{X} : f(x) = y \\ &\Rightarrow \Big(\exists x \in X : f(x) = y\Big) \wedge \Big(\exists \tilde{x} \in \tilde{X} : f(\tilde{x}) = y\Big) \Leftrightarrow y \in f(X) \cap f(\tilde{X}). \end{aligned}$$

Die dritte Aussage beweisen wir genauso:

$$\begin{aligned} y \in f(X) \setminus f(\tilde{X}) &\Leftrightarrow \Big(\exists x \in X : f(x) = y\Big) \wedge \Big(\forall \tilde{x} \in \tilde{X} : f(\tilde{x}) \neq y\Big) \\ &\Rightarrow \exists x \in X \setminus \tilde{X} : f(x) = y \Leftrightarrow y \in f(X \setminus \tilde{X}). \end{aligned}$$

Die letzte Implikation ist einfach: $y \in f(X) \Rightarrow \exists x \in X \subseteq \tilde{X} : f(x) = y \Rightarrow y \in f(\tilde{X})$.
(3) Für „$\Rightarrow$“ sei $x \in X$, dann ist $f(x) \in f(X) \subseteq Y$, also $x \in f^{-1}(Y)$. Für „$\Leftarrow$“ sei umgekehrt $y \in f(X)$, dann ist $y = f(x)$ für ein $x \in X \subseteq f^{-1}(Y)$, also $y = f(x) \in Y$. Mit der Wahl $Y = f(X)$ bzw. $X = f^{-1}(Y)$ folgt daraus auch die letzte Behauptung. $\square$

Definition 4.15. Seien A, B, C Mengen und $f : A \to B, g : B \to C$ Abbildungen. Die **Verkettung** von f und g ist eine Abbildung $g \circ f : A \to C$ (lies: „g nach f“), definiert durch die Abbildungsvorschrift

$$x \mapsto g(f(x)) \qquad \text{bzw.} \qquad (g \circ f)(x) = g(f(x)).$$

Sei D eine weitere Menge und $h : C \to D$ eine weitere Abbildung, sind auch die Verkettungen $h \circ g : B \to D$ sowie $h \circ (g \circ f) : A \to D$ und $(h \circ g) \circ f : A \to D$ definiert.
Behauptung: Die Verkettung ist *assoziativ*, d.h. es gilt $h \circ (g \circ f) = (h \circ g) \circ f$ und man kann einfacher $h \circ g \circ f : A \to D$ schreiben.
Beweis Wir wollen zeigen, dass $(h \circ (g \circ f))(x) = ((h \circ g) \circ f)(x)$ für alle $x \in A$ gilt; und tatsächlich ist nach Definition der Verkettung

$$(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x))) = (h \circ g)(f(x)) = ((h \circ g) \circ f)(x). \square$$

Zu dieser Aussage sagt man auch, das folgende Diagramm *kommutiert*; damit ist gemeint, dass verschiedene Wege durch das Diagramm zum gleichem Ergebnis führen.

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[rr, bend right, "g \circ f"'] & B \arrow[r, "g"] \arrow[rr, bend left, "h \circ g"] & C \arrow[r, "h"] & D
\end{tikzcd}$$

Beispiel 4.16. Die Verkettung kann schwierig sein, weil im Term $g(f(x))$ das Element $f(x)$ sowohl das „y“ der Funktion f als auch das „x“ der Funktion g ist.

(1) a) Der Wertebereich von $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ ist gleich dem Definitionsbereich von $g : \mathbb{R} \to \mathbb{R}, g(x) = 2x - 1$, daher ist die Verkettung $g \circ f : \mathbb{R} \to \mathbb{R}$ definiert:

$$(g \circ f)(x) = g(f(x)) = g(x^2) = 2x^2 - 1.$$

Dasselbe gilt umgekehrt, die Verkettung $f \circ g : \mathbb{R} \to \mathbb{R}$ hat die Vorschrift

$$(f \circ g)(x) = f(g(x)) = f(2x - 1) = (2x - 1)^2 = 4x^2 - 4x + 1.$$

Es ist $g \circ f \neq f \circ g$, die Verkettung also im Allgemeinen *nicht kommutativ.*
b) Wir können f auch mit der Funktion $h : \mathbb{R} \setminus \{0\} \to \mathbb{R}, h(x) = \frac{1}{x}$ verketten und erhalten $f \circ h : \mathbb{R} \setminus \{0\} \to \mathbb{R}, (f \circ h)(x) = f(\frac{1}{x}) = \frac{1}{x^2}$. Die Verkettung $h \circ f$ ist nicht definiert, da h an der Stelle 0 des Wertebereichs von f nicht definiert ist.

Wir beheben dieses Problem: $k : \mathbb{R} \to \mathbb{R} \setminus \{0\}, k(x) = x^2$ ist nicht wohldefiniert, aber die Einschränkung $\tilde{f} : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}, \tilde{f}(x) = x^2$ von f ist es; und nun ist $h \circ \tilde{f} : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ definiert:

$$(h \circ \tilde{f})(x) = h(\tilde{f}(x)) = h(x^2) = \frac{1}{x^2}.$$

Hier ist zufällig $f \circ h = h \circ \tilde{f}$.

c) Verketten wir $f \circ h$ mit g, so ergibt sich $g \circ f \circ h : \mathbb{R} \setminus \{0\} \to \mathbb{R}$,

$$(g \circ f \circ h)(x) = g((f \circ h)(x)) = g\left(\frac{1}{x^2}\right) = \frac{2}{x^2} - 1.$$

Dasselbe erhalten wir natürlich für die Verkettung von $g \circ f$ mit h:

$$(g \circ f)(h(x)) = (g \circ f)\left(\frac{1}{x}\right) = 2\left(\frac{1}{x}\right)^2 - 1 = \frac{2}{x^2} - 1.$$

(2) Durch Verkettung lassen sich aus „einfachen“ Funktionen komplexere Funktionen erzeugen. Verkettet man eine Funktion $f : \mathbb{R} \to \mathbb{R}$ mit der affin-linearen Funktion $g : \mathbb{R} \to \mathbb{R}, g(x) = ax + b$, so entstehen $f \circ g : \mathbb{R} \to \mathbb{R}, g \circ f : \mathbb{R} \to \mathbb{R}$,

$$(f \circ g)(x) = f(ax + b), \qquad (g \circ f)(x) = af(x) + b.$$

Hier kann man den Einfluss der Parameter $a, b \in \mathbb{R}$ gut am Graphen von f interpretieren (Übung). Dasselbe gilt für die Verkettung mit der Betragsfunktion $\mathbb{R} \to \mathbb{R}, x \mapsto |x|$: Der Graph von $x \mapsto f(|x|)$ entsteht durch Spiegelung des Teilgraphen von f rechts der y-Achse an der y-Achse („nach links umklappen“), $x \mapsto |f(x)|$ durch Spiegelung des Teils unterhalb der x-Achse an der x-Achse („nach oben umklappen“). Zum Beispiel entsteht der Graph der Funktion $f : \mathbb{R} \to \mathbb{R}, f(x) = -2|(x-1)^2 - 3|$ aus der Normalparabel $y = x^2$ durch

- Verschiebung um 1 nach rechts und drei nach unten,
- Umklappen unterhalb der x-Achse nach oben,
- Streckung mit Faktor 2 und Spiegeln an der x-Achse.

(3) Die Assoziativität der Verkettung erlaubt, bei der Funktion $h \circ g \circ f = h \circ (g \circ f) = (h \circ g) \circ f$ die Klammern wegzulassen; für eine Funktion $f : A \to A$ ist daher auch die mehrfache Verkettung $f^n = f \circ \ldots \circ f$ mit $n \in \mathbb{N}$ Faktoren definiert. Ein Beispiel: Für $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ ist

$$\begin{aligned} f(x) &= x^2, & f^2(x) &= f(f(x)) = f(x^2) = x^4, \\ f^3(x) &= x^6, & f^n(x) &= (f \circ \ldots \circ f)(x) = x^{2n}. \end{aligned}$$

Achtung: Bei „benannten“ Funktionen wie dem Sinus meint man mit $\sin^2(x)$ häufig nicht $\sin(\sin(x))$, sondern $\sin(x)^2$. Die Abbildung $\mathrm{id}_A : A \to A, x \mapsto x$ heißt die **Identität** auf der Menge A; sie erfüllt für jede Funktion $f : A \to A$

$$(f \circ \mathrm{id}_A)(x) = f(\mathrm{id}_A(x)) = f(x), \qquad (\mathrm{id}_A \circ f)(x) = \mathrm{id}_A(f(x)) = f(x),$$

also $f \circ \mathrm{id}_A = \mathrm{id}_A \circ f = f$. Man sagt dazu, id_A ist das **neutrale Element** bezüglich der Verkettung und definiert $f^0 := \mathrm{id}_A$. Für $A \subseteq B$ nennt man übrigens $\mathrm{id} : A \to B, x \mapsto x$ die **Einbettung** von A in B.

(4) *Für die Physiker:* In der Physik kann eine Gleichung nach jeder vorkommenden Größe aufgelöst werden und so implizit eine Funktion definieren. Verkettung bedeutet oft nur einen Wechsel der Abhängigkeit: Die Energie $E = E(h)$ hängt von der Höhe ab, die Höhe $h(t)$ von der Zeit, für die Verkettung $E(h(t))$ schreibt der Physiker $E(h = h(t))$ oder einfach $E(t)$. Auch Koordinatenwechsel sind Verkettungen: Ein Zentralpotential $V(\vec{r}) = V(r)$ ist eigentlich die Verkettung $V = \tilde{V} \circ |\cdot|$ des Abstandes $|\cdot|$ mit einem Skalarpotential $\tilde{V}$.

Neben der Verkettung sind punktweise Verknüpfungen eine wichtige Methode, um aus bekannten Funktionen neue zu erzeugen.

Definition 4.17. Sei A eine Menge, $\lambda \in \mathbb{R}$ und $f : A \to \mathbb{R}, g : A \to \mathbb{R}$ Funktionen.

(1) Die Funktionen $\lambda f : A \to \mathbb{R}, |f| : A \to \mathbb{R}, f + g : A \to \mathbb{R}$ und $fg : A \to \mathbb{R}$ werden **punktweise** definiert durch die Funktionsgleichungen

$$(\lambda f)(x) := \lambda f(x), \qquad |f|(x) := |f(x)|,$$
$$(f+g)(x) := f(x) + g(x), \qquad (fg)(x) := f(x)g(x).$$

Auf der Menge $B := \{x \in A : g(x) \neq 0\}$ wird außerdem die Funktion $\frac{f}{g} : B \to \mathbb{R}$ definiert durch

$$\left(\frac{f}{g}\right)(x) := \frac{f(x)}{g(x)}.$$

Analog kann man auch andere Funktionen *punktweise* definieren, zum Beispiel $\max\{f, g\} : A \to \mathbb{R}$ und $\min\{f, g\} : A \to \mathbb{R}$ durch die Vorschriften

$$(\max\{f, g\})(x) := \max\{f(x), g(x)\}, \qquad (\min\{f, g\})(x) := \min\{f(x), g(x)\}.$$

(2) Auf $\mathrm{Abb}(A, \mathbb{R})$ werden die Relationen $f = g, f \leqslant g, f < g$ und analog auch $f \geqslant g, f > g$ **punktweise** definiert durch

$$f = g \Leftrightarrow \forall x \in A : f(x) = g(x),$$
$$f \leqslant g \Leftrightarrow \forall x \in A : f(x) \leqslant g(x),$$
$$f < g \Leftrightarrow \forall x \in A : f(x) < g(x).$$

Beispiel 4.18. (1) a) Mithilfe der Monomfunktionen $p_n : \mathbb{R} \to \mathbb{R}, p_n(x) = x^n$ hat $f : \mathbb{R} \to \mathbb{R}, f(x) = 2x - 1$ die Form

$$f(x) = 2x - 1 = 2p_1(x) - p_0(x), \text{ also } \qquad f = 2p_1 - p_0;$$

für die Funktion $g : \mathbb{R} \to \mathbb{R}, g(x) = (x-1)^2$ gilt

$$(f+g)(x) = f(x) + g(x) = 2x - 1 + (x-1)^2 = x^2 = p_2(x),$$

also $f + g = p_2$. Genauso lassen sich alle Polynomfunktionen als *Linearkombination* der p_n schreiben.

Weiter ist $|g| = g$ und $fg = 2p_3 - 5p_2 + 4p_1 - p_0$:

$$|g|(x) = |g(x)| = |(x-1)^2| = (x-1)^2 = g(x),$$
$$(fg)(x) = f(x)g(x) = (2x-1)(x-1)^2 = 2x^3 - 5x^2 + 4x - 1.$$

Für $x \neq 1$ ist $g(x) \neq 0$ und daher $\frac{f}{g} : \mathbb{R} \setminus \{1\} \to \mathbb{R}, \left(\frac{f}{g}\right)(x) = \frac{2x-1}{(x-1)^2}$ wohldefiniert. Zu all diesen Verknüpfungen vergleiche man auch die Graphen mit f, g.
b) Das Produkt $f \cdot g$ ist nicht mit der Verkettung $f \circ g$ zu verwechseln: Zum Beispiel ist $p_2 \cdot p_3 = p_5$, denn $x^2 \cdot x^3 = x^5$, aber $p_2 \circ p_3 = p_6$, denn $(x^3)^2 = x^6$.

(2) a) Für $f : \mathbb{R} \to \mathbb{R}, f(x) = 2x - 1$ und $g : \mathbb{R} \to \mathbb{R}, g(x) = 2x + 1$ gilt

$$\forall x \in \mathbb{R} : f(x) = 2x - 1 < 2x + 1 = g(x),$$

also $f < g$; natürlich ist auch $f \leqslant g$. Der Graph von f liegt an jeder Stelle echt unterhalb des Graphen von g, für $h : \mathbb{R} \to \mathbb{R}, h(x) = 2 - x$ ist hingegen

$$f(0) = -1 < 2 = h(0), \qquad f(2) = 3 > 0 = h(2),$$

also weder $f \leqslant h$ noch $h \leqslant f$. Die Relationen $\leqslant, <$ treffen immer eine Aussage über *alle* Stellen im Definitionsbereich.
b) Für $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ und $g : \mathbb{R} \to \mathbb{R}, g(x) = -x^2$ gilt offenbar $g \leqslant f$; wegen $f(0) = 0 = g(0)$ ist aber $g < f$ *nicht* richtig.
c) Die Relation $f = g$ ist dieselbe wie in Beispiel 4.10(2), insbesondere stimmen Definitions- und Wertemenge überein. Mit der Schreibweise $f \neq g$ meint man *nicht* $\forall x \in A : f(x) \neq g(x)$, sondern die Negation von $f = g$:

$$\exists x \in A : f(x) \neq g(x).$$

Zum Beispiel bedeutet $f \neq 0$, dass f nicht die Nullfunktion ist, also mindestens einen Funktionswert $f(x) \neq 0$ hat; der Sinus hat unendlich viele Nullstellen, aber $\sin \neq 0$. Keine Nullstellen fordert man mit der Aussage

$$\forall x \in A : f(x) \neq 0.$$

d) In a) gilt für die Einschränkungen von f und h auf $(-\infty, 1)$ die Relation $f < h$, dazu sagt man „$f < h$ auf $(-\infty, 1)$“. Genauso gilt $f \leqslant h$ auf $(-\infty, 1]$ und $f > h$ auf $(1, \infty)$, aber auf $[0, 2]$ stehen f und h in keiner Relation zueinander.

(3) Für Analysis ist interessant, ob sich Eigenschaften von Funktionen auf Verknüpfungen übertragen: Zum Beispiel ist $f_a : \mathbb{R} \to \mathbb{R}, f_a(x) = \sin(ax)$ für $a \neq 0$ periodisch mit (kleinster) Periode $\frac{2\pi}{a}$, aber

$$f_1 + f_2 \text{ ist periodisch,} \qquad f_1 + f_{\sqrt{2}} \text{ ist nicht periodisch.}$$

Die Summe $p_3 + p_5$ der ungeraden Funktionen p_3 und p_5 ist weiterhin eine ungerade Funktion, aber das Produkt $p_3 \cdot p_5 = p_8$ ist nicht ungerade, sondern gerade.

Die Funktionen $f : \mathbb{R} \setminus \{1\} \to \mathbb{R}, f(x) = \frac{1}{1-x}$ und $g : \mathbb{R} \setminus \{1\} \to \mathbb{R}, g(x) = \frac{x}{x-1}$ haben jeweils einen Pol an der Stelle $x = 1$, ihre Summe $f + g : \mathbb{R} \setminus \{1\} \to \mathbb{R}$,

$$(f+g)(x) = f(x) + g(x) = \frac{1}{1-x} + \frac{x}{x-1} = \frac{1-x}{1-x} = 1$$

ist hingegen stetig fortsetzbar an der Stelle $x = 1$.

(4) Wie bei Zahlen ist die Aussage $f < g$ stärker als $f \leqslant g$ und $f \neq g$, die Aussage $f \leqslant g$ schwächer als $f < g$ und $f = g$. Die Relation $\leqslant$ ist reflexiv, antisymmetrisch und transitiv, aber nicht total, also eine (partielle) Ordnungsrelation auf $\mathrm{Abb}(A, \mathbb{R})$; die Relation $<$ ist nur transitiv.

(5) Die punktweisen Verknüpfungen erzeugen eine Struktur[10] auf der Menge $\mathrm{Abb}(A, \mathbb{R})$ aller Funktionen $f : A \to \mathbb{R}$, die den reellen Zahlen $\mathbb{R}$ ähnelt: Es gelten zum Beispiel das Assoziativgesetz $(f + g) + h = f + (g + h)$, das Kommutativgesetz $fg = gf$ oder die Distributivgesetze

$$(\lambda + \mu)f = \lambda f + \mu f, \qquad \lambda(f + g) = \lambda f + \lambda g.$$

Die Nullfunktion ist das neutrale Element bezüglich dieser Addition, denn $f + 0 = 0 + f = f$; $p_0 = 1$ ist das neutrale Element bezüglich der Multiplikation, denn $f \cdot 1 = 1 \cdot f = f$ für jedes $f : A \to \mathbb{R}$. Es gibt aber auch Unterschiede zu $\mathbb{R}$: Die Relation $\leqslant$ zwischen Funktionen ist nicht total, nicht durch jede Funktion $f \neq 0$ kann man dividieren und auf $\mathrm{Abb}(\mathbb{R}, \mathbb{R})$ hat man mit der Verkettung eine zweite Multiplikation: $\circ$ ist nicht kommutativ, aber assoziativ und es gelten die Distributivgesetze

$$(f + g) \circ h = (f \circ h) + (g \circ h), \qquad f \circ (g + h) = f \circ g + f \circ h.$$

Ein Beweis all dieser Aussagen verläuft schematisch: Man berechnet zum Beispiel

$$\begin{aligned} ((f+g) \circ h)(x) &= (f+g)(h(x)) = f(h(x)) + g(h(x)) = (f \circ h)(x) + (g \circ h)(x) \\ &= ((f \circ h) + (g \circ h))(x) \end{aligned}$$

für alle $x \in A$, das bedeutet $(f + g) \circ h = (f \circ h) + (g \circ h)$; genauso beweist man auch die anderen Aussagen. □

(6) Von einer Abbildung $f : A \to \mathbb{R}$ definieren wir den *Positivteil* $f_+ : A \to \mathbb{R}$ und den *Negativteil* $f_- : A \to \mathbb{R}$ durch

$$\begin{aligned} f_+(x) &:= \max\{f(x), 0\}, \text{ also } f_+ = \max\{f, 0\}, \\ f_-(x) &:= -\min\{f(x), 0\}, \text{ also } f_- = -\min\{f, 0\}, \end{aligned}$$

wobei 0 auch für die Nullfunktion $0 : \mathbb{R} \to \mathbb{R}, x \mapsto 0$ steht.

[10] $\mathrm{Abb}(A, \mathbb{R})$ ist ein $\mathbb{R}$-*Vektorraum*; das wird in der Linearen Algebra behandelt.

Der Positivteil f_+ nimmt nur den Graphen von f oberhalb der x-Achse, der Negativteil f_- klappt den Teil unterhalb der x-Achse nach oben. Mit Beispiel 3.8(5) beweist man die Eigenschaften $f_+, f_- \geqslant 0$, $f = f_+ - f_-$, $|f| = f_+ + f_-$ sowie

$$f \geqslant 0 \Leftrightarrow f = f_+, \qquad\qquad f \leqslant 0 \Leftrightarrow f = -f_-.$$

Definition 4.19. Seien A, B Mengen und $f : A \to B$ eine Abbildung.

(1) f heißt **surjektiv**, falls f rechtsvollständig ist. Äquivalent dazu ist jeweils:

a) $\forall y \in B\, \exists x \in A : f(x) = y$.
b) Jedes $y \in B$ hat mindestens ein Urbild, d.h. $f^{-1}(\{y\}) \neq \emptyset$.
c) Jedes $y \in B$ liegt im Bild von f, d.h. $f(A) = B$.

(2) f heißt **injektiv**, falls f linkseindeutig ist. Äquivalent dazu ist jeweils:

a) $\forall x, \tilde{x} \in A : f(x) = f(\tilde{x}) \Rightarrow x = \tilde{x}$.
b) Für je zwei verschiedene $x, \tilde{x} \in A$ ist auch $f(x) \neq f(\tilde{x})$.
c) Zu jedem $y \in B$ existiert höchstens ein $x \in A$, sodass $f(x) = y$.
d) Jedes $y \in B$ hat höchstens ein Urbild, d.h. $f^{-1}(\{y\})$ höchstens ein Element.

(3) f heißt **bijektiv**, falls f injektiv und surjektiv ist. Äquivalent dazu ist jeweils:

a) Zu jedem $y \in B$ existiert genau ein $x \in A$, sodass $f(x) = y$.
b) Jedes $y \in B$ hat genau ein Urbild, d.h. $f^{-1}(\{y\}) = \{x\}$.

Beispiel 4.20. (1) a) $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ ist nicht surjektiv, da es zum Wert $-1 \in \mathbb{R}$ kein Urbild $x \in \mathbb{R}$ gibt, also die Gleichung $f(x) = -1$ unlösbar ist. f ist auch nicht injektiv, da an den Stellen 2 und -2 derselbe Funktionswert $f(2) = f(-2) = 4$ vorliegt.
b) Ändert man die Wertemenge auf $f(\mathbb{R}) = [0, \infty)$, so ist $g : \mathbb{R} \to [0, \infty), g(x) = x^2$ surjektiv: Jedes $y \geqslant 0$ hat die Urbilder $x = \sqrt{y}$ und $x = -\sqrt{y}$. Allgemein ist für jede Abbildung $f : A \to B$ die Einschränkung $f : A \to f(A)$ surjektiv.
c) Ändert man die Definitionsmenge von f auf $[0, \infty)$, so ist $h : [0, \infty) \to \mathbb{R}$, $h(x) = x^2$ injektiv: Für alle $x, \tilde{x} \geqslant 0$ gilt

$$h(x) = h(\tilde{x}) \Rightarrow x^2 = \tilde{x}^2 \Rightarrow |x| = |\tilde{x}| \Rightarrow x = \tilde{x}.$$

Die Surjektivität kann durch Verkleinerung der Wertemenge „erzwungen“ werden, die Injektivität durch Verkleinerung der Definitionsmenge; daher sagt man auch, eine Funktion ist injektiv auf $\tilde{A} \subseteq A$ oder surjektiv auf $\tilde{B} \subseteq B$.[11]
d) Die Funktion $k : [0, \infty) \to [0, \infty), k(x) = x^2$ ist injektiv und surjektiv, also bijektiv: Jedes $y \geqslant 0$ hat genau ein Urbild $x = \sqrt{y} \geqslant 0$. Allgemein ist für jede injektive Abbildung $f : A \to B$ die Einschränkung $f : A \to f(A)$ bijektiv.

[11] Zum Beispiel für konstante Funktionen muss $\tilde{A}$ sehr klein sein.

(2) Für ein $y \in B$ ist die Lösungsmenge der Gleichung $f(x) = y$ gerade das Urbild $f^{-1}(\{y\}) = \{x \in A : f(x) = y\}$. Ist nun für *jedes* $y \in B$

a) die Gleichung $f(x) = y$ lösbar, so ist f surjektiv.
b) die Lösung der Gleichung $f(x) = y$ eindeutig, so ist f injektiv.
c) die Gleichung $f(x) = y$ eindeutig lösbar, so ist f bijektiv.

Umgekehrt ist zum Beispiel $f : \mathbb{R} \to \mathbb{R}, f(x) = x^3$ bijektiv und daher die Gleichung $x^3 = -4$ eindeutig lösbar.

(3) Im Koordinatensystem bedeutet die Surjektivität (Rechtsvollständigkeit), dass für jedes $y \in B$ die Parallele zur x-Achse durch y den Graphen von f *mindestens* einmal schneidet; bei Injektivität (Linkseindeutigkeit) schneiden diese Parallelen entsprechend *höchstens* einmal, bei Bijektivität *genau* einmal.
Eine bijektive Abbildung $f : A \to B$ ist nach Vertauschen der x- und y-Achse wieder eine Abbildung: Jedem y wird genau ein x zugeordnet. Die Abbildung $f(x) \mapsto x$ heißt die **Umkehrfunktion** $f^{-1} : B \to A$ von f, siehe Definition 4.22; auch f^{-1} ist bijektiv. Ein Beispiel: Sei $a \neq 0$, dann hat die Gleichung $ax + b = y$ für jedes $y \in \mathbb{R}$ genau eine Lösung $x = \frac{y-b}{a} = \frac{1}{a}y - \frac{b}{a} \in \mathbb{R}$. Daher ist die Funktion $f : \mathbb{R} \to \mathbb{R}, f(x) = ax + b$ bijektiv mit der Umkehrfunktion

$$f^{-1} : \mathbb{R} \to \mathbb{R}, f^{-1}(y) = \frac{1}{a}y - \frac{b}{a}.$$

Wegen $\frac{1}{a} \neq 0$ ist f^{-1} ebenfalls bijektiv mit der Umkehrfunktion $(f^{-1})^{-1} = f$.

Satz 4.21. Seien A, B Mengen und $f : A \to B$ eine Abbildung.

(1) Die folgenden Aussagen sind äquivalent:

a) f ist surjektiv.
b) Für alle Mengen $Y \subseteq B$ gilt $f(f^{-1}(Y)) = Y$.
c) Für alle Mengen $Y, \tilde{Y} \subseteq B$ gilt $Y \subseteq \tilde{Y} \Leftrightarrow f^{-1}(Y) \subseteq f^{-1}(\tilde{Y})$.

(2) Die folgenden Aussagen sind äquivalent:

a) f ist injektiv.
b) Für alle Mengen $X \subseteq A$ gilt $X = f^{-1}(f(X))$.
c) Für alle Mengen $X, \tilde{X} \subseteq A$ gilt $X \subseteq \tilde{X} \Leftrightarrow f(X) \subseteq f(\tilde{X})$.
d) Für alle Mengen $X, \tilde{X} \subseteq A$ gilt $f(X \cap \tilde{X}) = f(X) \cap f(\tilde{X})$.
e) Für alle Mengen $X, \tilde{X} \subseteq A$ gilt $f(X) \setminus f(\tilde{X}) = f(X \setminus \tilde{X})$.

Man beachte, dass in allen Aussagen die Inklusion „$\subseteq$“ bzw. die Implikation „$\Rightarrow$“ immer wahr ist, siehe Satz 4.14.
Beweis Wir beweisen die Äquivalenz der Aussagen jeweils mit einem Ringschluss.

(1) „a) $\Rightarrow$ b)“: Sei f surjektiv, es ist nur „$\supseteq$“ in b) zu zeigen. Sei $y \in Y$, dann gibt es ein $x \in A$ mit $f(x) = y$; das bedeutet $x \in f^{-1}(Y)$ und daher $y \in f(f^{-1}(Y))$.
„b) $\Rightarrow$ c)“: Wieder ist nur „$\Leftarrow$“ in c) zu zeigen. Es gelte $f^{-1}(Y) \subseteq f^{-1}(\tilde{Y})$, aus Satz 4.14(2) folgt dann $f(f^{-1}(Y)) \subseteq f(f^{-1}(\tilde{Y}))$ und aus b) $Y \subseteq \tilde{Y}$.
„c) $\Rightarrow$ a)“: Stets gilt

$$f^{-1}(f(A)) = \{x \in A : f(x) \in f(A)\} = A = f^{-1}(B).$$

Mit $Y = B$ und $\tilde{Y} = f(A) \subseteq B$ in c) folgt daraus $f(A) = B$, d.h. f ist surjektiv.

(2) „a) $\Rightarrow$ b)“: Sei $x \in f^{-1}(f(X))$, dann gilt $f(x) \in f(X)$ und daher existiert ein $\tilde{x} \in X$ mit $f(\tilde{x}) = f(x)$. Aus der Injektivität von f folgt $x = \tilde{x} \in X$.
„b) $\Rightarrow$ c)“ folgt wie in (1) aus Satz 4.14(1).
„c) $\Rightarrow$ d)“: Sei $y \in f(X) \cap f(\tilde{X})$ gegeben, dann gibt es $x \in X$ mit $y = f(x) \in f(\tilde{X})$. Das bedeutet gerade $f(\{x\}) = \{f(x)\} \subseteq f(\tilde{X})$ und aus c) folgt $\{x\} \subseteq \tilde{X}$, also $x \in \tilde{X}$. Damit ist $x \in X \cap \tilde{X}$ und $y = f(x) \in f(X \cap \tilde{X})$.
„d) $\Rightarrow$ e)“: Aus $X \setminus \tilde{X} \subseteq X$ folgt mit Satz 4.14(2) $f(X \setminus \tilde{X}) \subseteq f(X)$ und daher $f(X \setminus \tilde{X}) \setminus f(\tilde{X}) \subseteq f(X) \setminus f(\tilde{X})$. Nach d) ist aber[12]

$$f(X \setminus \tilde{X}) \cap f(\tilde{X}) = f((X \setminus \tilde{X}) \cap \tilde{X}) = f(\emptyset) = \emptyset$$

und daher $f(X \setminus \tilde{X}) = f(X \setminus \tilde{X}) \setminus f(\tilde{X}) \subseteq f(X) \setminus f(\tilde{X})$.
„e) $\Rightarrow$ a)“: Seien $x, \tilde{x} \in A$ mit $f(x) = f(\tilde{x})$ gegeben, dann ist nach e)

$$f(\{x\} \setminus \{\tilde{x}\}) = f(\{x\}) \setminus f(\{\tilde{x}\}) = \{f(x)\} \setminus \{f(\tilde{x})\} = \emptyset,$$

also $\{x\} \setminus \{\tilde{x}\} = \emptyset$. Daraus folgt $\{x\} \subseteq \{\tilde{x}\}$, also $x = \tilde{x}$ und f ist injektiv. $\square$

Definition 4.22. Seien A, B Mengen und $f : A \to B$ bijektiv. Dann gilt

$$\begin{aligned} &\forall x \in A\ \exists! y \in B : f(x) = y, \text{ da } f \text{ linksvollständig und rechtseindeutig ist,} \\ \text{und } &\forall y \in B\ \exists! x \in A : f(x) = y, \text{ da } f \text{ rechtsvollständig und linkseindeutig ist.} \end{aligned}$$

Nach der zweiten Zeile ist die Umkehrrelation

$$f^{-1} := \{(y, x) \in B \times A : f(x) = y\} \subseteq B \times A$$

linksvollständig und rechtseindeutig, also eine Abbildung $f^{-1} : B \to A$; sie heißt die **Umkehrfunktion** oder **Inverse** von f. f^{-1} ordnet jedem $y \in B$ das eindeutige $x \in A$ zu, sodass $f(x) = y$ gilt; dieses zu y gehörige x wird mit $f^{-1}(y)$ bezeichnet.
Daher gilt $f(f^{-1}(y)) = y$ für alle $y \in B$ und genauso $f^{-1}(f(x)) = x$ für alle $x \in A$, kurz

$$f \circ f^{-1} = \mathrm{id}_B, \qquad f^{-1} \circ f = \mathrm{id}_A, \tag{4.3}$$

für die Umkehrung dieser Aussage siehe Satz 4.24(3)c). Die Umkehrrelation von f^{-1}

$$(f^{-1})^{-1} = \{(x, y) \in A \times B : f^{-1}(y) = x\} = \{(x, y) \in A \times B : f(x) = y\}$$

ist f, also eine Abbildung; daher ist auch f^{-1} bijektiv und $(f^{-1})^{-1} = f$.

[12]Beachte hier $A \cap B = \emptyset \Leftrightarrow A = A \setminus B$ und in „e) $\Rightarrow$ a)“ $A \setminus B = \emptyset \Leftrightarrow A \subseteq B$.

Beispiel 4.23. (1) a) Die Monomfunktionen $p_n : \mathbb{R} \to \mathbb{R}, p_n(x) = x^n$ sind für ungerade $n \in \mathbb{N}$ bijektiv, aber für gerade n nicht injektiv. Für alle $n \in \mathbb{N}$ sind hingegen die Einschränkungen $p_n : [0, \infty) \to [0, \infty)$ wohldefiniert und bijektiv, daher existieren die Umkehrfunktionen $w_n = p_n^{-1} : [0, \infty) \to [0, \infty)$, die **Wurzelfunktionen**. Für $y \geqslant 0$ ist $w_n(y)$ die eindeutige Lösung $x \geqslant 0$ der Gleichung $x^n = y$, also die n**-te Wurzel** $w_n(y) = p_n^{-1}(y) = \sqrt[n]{y} = y^{\frac{1}{n}}$. Auch w_n ist bijektiv mit der Umkehrfunktion $w_n^{-1} = p_n$: Es gilt $p_n \circ w_n = \mathrm{id}_{[0,\infty)} = w_n \circ p_n$, zum Beispiel

$$p_2(w_2(x)) = (\sqrt{x})^2 = x, \qquad w_2(p_2(x)) = \sqrt{x^2} = x$$

für $x \geqslant 0$. $p_n(x)$ gibt es auch für negative Argumente, nicht aber $w_n(x)$; daher sind zum Beispiel $p_2(w_2(-4)), w_3(p_3(-2))$ nicht definiert und

$$w_2(p_2(-2)) = \sqrt{(-2)^2} = 2 = |-2|.$$

Analog ist für alle (reellen) Exponenten $a \neq 0$ die Potenzfunktion $f_a : (0, \infty) \to (0, \infty), f_a(x) = x^a$ bijektiv mit der Umkehrfunktion $f_a^{-1} = f_{\frac{1}{a}}$; auf $(0, \infty)$ ist also (f_a) eine *Schar* bijektiver Funktionen.
b) Sei $a > 0$. Für die Exponentialfunktion $g_a : \mathbb{R} \to \mathbb{R}, g_a(x) = a^x$ gilt ähnlich wie in a): g_a ist nicht surjektiv, aber die Einschränkung $\exp_a : \mathbb{R} \to (0, \infty)$, $\exp_a(x) = a^x$ ist bijektiv. Daher existiert die Umkehrfunktion $\exp_a^{-1} = l_a : (0, \infty) \to \mathbb{R}$, für jedes $y > 0$ ist $l_a(y)$ die eindeutige Lösung $x \in \mathbb{R}$ der Gleichung $a^x = y$, also der **Logarithmus** $l_a(y) = \log_a(y)$. Auch l_a ist bijektiv mit der Umkehrfunktion $l_a^{-1} = \exp_a$:

$$l_a(\exp_a(x)) = \log_a(a^x) = x, \qquad \exp_a(l_a(y)) = a^{\log_a(y)} = y$$

für $x \in \mathbb{R}$ und $y > 0$. Die Exponentialfunktion $\exp_a$ erfüllt $a^{x+\tilde{x}} = a^x + a^{\tilde{x}}$ für alle $x, \tilde{x} \in \mathbb{R}$;[13] daraus erhält man mit $y = a^{\log_a(y)} = a^x$ das Logarithmengesetz

$$\log_a(y \cdot \tilde{y}) = \log_a(a^x \cdot a^{\tilde{x}}) = \log(a^{x+\tilde{x}}) = x + \tilde{x} = \log_a(y) + \log_a(\tilde{y})$$

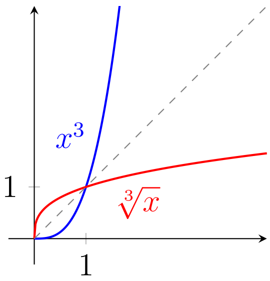
für $y, \tilde{y} > 0$.
c) Die trigonometrischen Funktionen $\sin : \mathbb{R} \to \mathbb{R}$ und $\cos : \mathbb{R} \to \mathbb{R}$ sind nicht injektiv, aber geeignete Einschränkungen auf ein Intervall der Länge π sind bijektiv, zum Beispiel $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$ und $\cos : [0, \pi] \to [-1, 1]$. Die Umkehrfunktionen dieser Einschränkungen heißen **Arkusfunktionen**:

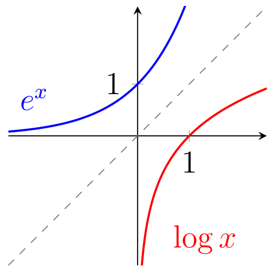
$$\arcsin : [-1, 1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \qquad \arccos : [-1, 1] \to [0, \pi].$$

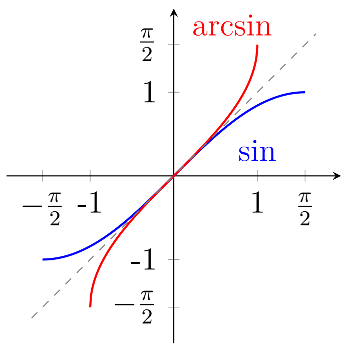
Genauso verfährt man für die trigonometrischen Funktionen $\tan, \cot$ oder die hyperbolischen Funktionen $\sinh, \cosh$: Hier sind die Einschränkungen $\sinh : \mathbb{R} \to \mathbb{R}$ und $\cosh : [0, \infty) \to [1, \infty)$ bijektiv mit den Umkehrfunktionen

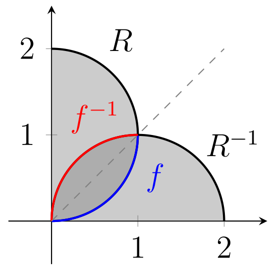
$$\operatorname{arsinh} : \mathbb{R} \to \mathbb{R}, \qquad \operatorname{arcosh} : [1, \infty) \to [0, \infty).$$

[13] Diese wichtige *Funktionalgleichung* wird in der Analysis 1 bewiesen.









(2) Die Relationen f, f^{-1} enthalten zunächst dieselben Paare $(x, y), (y, x)$ mit vertauschten Komponenten, also mit x- bzw. y-Achse als erster Achse. Werden beide in einem gemeinsamen Koordinatensystem dargestellt, so soll für beide die x-Achse erste Achse sein; dazu muss f^{-1} an der Geraden $y = x$ gespiegelt werden. f und f^{-1} liegen also symmetrisch zur ersten Winkelhalbierenden, die Abbildung zeigt das Beispiel

$$R = \{(x, y) \in \mathbb{R}^2 : x^2 + (y-1)^2 \leqslant 1, x \geqslant 0\},$$
$$R^{-1} = \{(y, x) \in \mathbb{R}^2 : (x, y) \in R\} = \{(x, y) \in \mathbb{R}^2 : (x-1)^2 + y^2 \leqslant 1, y \geqslant 0\}.$$

Für Funktionen entspricht dies dem Vertauschen von x und y: Aus dem Paar $(f(x), x) = (y, f^{-1}(y))$ wird $(x, f^{-1}(x))$. Zum Beispiel für den Viertelkreis $f : [0,1] \to [0,1], f(x) = 1 - \sqrt{1-x^2}$ lässt sich $y = 1 - \sqrt{1-x^2}$ auflösen zu $x = \sqrt{1-(y-1)^2}$, die Umkehrfunktion von f ist also

$$f^{-1} : [0,1] \to [0,1], f^{-1}(x) = \sqrt{1-(x-1)^2}.$$

(3) Ist $f : \mathbb{R} \to \mathbb{R}$ streng monoton wachsend, so ist f injektiv:

$$x \neq y \Rightarrow x < y \text{ oder } y < x \Rightarrow f(x) < f(y) \text{ oder } f(y) < f(x) \Rightarrow f(x) \neq f(y). \ \square$$

Dann ist $f : \mathbb{R} \to f(\mathbb{R})$ bijektiv, hat also eine Umkehrfunktion $f^{-1} : f(\mathbb{R}) \to \mathbb{R}$, und f^{-1} ist ebenfalls streng monoton wachsend (Beweisen Sie das zur Übung); dasselbe gilt natürlich für streng monoton fallende Funktionen. Zum Beispiel sind alle p_n streng monoton wachsend auf $[0, \infty)$, ebenso wie ihre Umkehrfunktionen w_n; dasselbe gilt für Exponentialfunktion und Logarithmus.

(4) Ist $f : A \to B$ bijektiv, so erfüllt die Umkehrfunktion $f^{-1} : B \to A$ die Gleichungen (4.3), ist also sowohl *Rechtsinverse* als auch *Linksinverse* von f. In Satz 4.24(1)c) zeigen wir umgekehrt, dass die Inverse f^{-1} als einzige Funktion (4.3) erfüllt. Während keine Funktion mehrere Rechts- *und* mehrere Linksinverse besitzt[14], können Funktionen durchaus mehrere Rechts- *oder* mehrere Linksinverse besitzen: Wie in (1)a) ist $f : \mathbb{R} \to [0, \infty), f(x) = x^2$ surjektiv und mit der Rechtsinversen $g : [0, \infty) \to \mathbb{R}, g(x) = \sqrt{x}$:

$$\forall x \geqslant 0 : (f \circ g)(x) = f(g(x)) = \left(\sqrt{x}\right)^2 = x.$$

Aber auch $h : [0, \infty) \to \mathbb{R}, h(x) = -\sqrt{x}$ ist eine Rechtsinverse von f, daher ist f nicht injektiv und die Inverse f^{-1} existiert nicht. *Merke:* Nur bijektive Funktionen haben eine Umkehrfunktion. Aus einer der beiden Gleichungen $f \circ g = \mathrm{id}_B$ oder $h \circ f = \mathrm{id}_A$ folgt noch nicht, dass g bzw. h die Umkehrfunktion von f ist.

(5) a) Für die mehrfache Verkettung einer Abbildung $f : A \to A$ schreibt man $f^n = f \circ \ldots \circ f$ und $f^0 = \mathrm{id}_A$. Ist f bijektiv, so schreibt man genauso

$$f^{-n} := f^{-1} \circ f^{-1} \circ \ldots \circ f^{-1}$$

für die mehrfache Verkettung der Umkehrfunktion $f^{-1} : A \to A$. Wegen (4.3) gilt das Potenzgesetz

$$f^m \circ f^{-n} = f^{m-n} = f^{-n} \circ f^m$$

für alle $n, m \in \mathbb{N}_0$; insbesondere sind Verkettungen von f und f^{-1} kommutativ, obwohl $\circ$ im Allgemeinen nicht kommutativ ist.
b) Sei $f : A \to B$. Wie in Beispiel 4.13(6) hat das Urbild $f^{-1}(Y) \subseteq A$ einer Menge $Y \subseteq B$ im Allgemeinen nichts mit der Umkehrfunktion $f^{-1} : B \to A$ zu tun. Ist allerdings f bijektiv, so ist das Bild von $Y \subseteq B$ unter f^{-1}

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

gerade das Urbild von Y unter f, denn $f(x) = y \Leftrightarrow f^{-1}(y) = x$. Genauso sind Bilder von f die Urbilder von f^{-1}, die beiden Schreibweisen sind kompatibel.

(6) Sei $n \in \mathbb{N}$. Für eine endliche Menge $M = \{1, \ldots, n\}$ mit n Elementen heißt eine bijektive Abbildung $\pi : M \to M$ eine *Permutation* von M. Zum Beispiel für

$$\pi : \{1, 2, 3\} \to \{1, 2, 3\}, 1 \mapsto 3, 2 \mapsto 2, 3 \mapsto 1$$

ist $\pi \circ \pi = \mathrm{id}_{\{1,2,3\}}$, also $\pi^{-1} = \pi$. Es gibt genau $n! = 1 \cdot \ldots \cdot n$ verschiedene Permutationen von M; sie bilden mit der Verkettung die *symmetrische Gruppe* $(S_n, \circ)$.

[14] Eine solche Funktion wäre nach Satz 4.24(1) bijektiv.

Satz 4.24. Seien A, B Mengen, $A \neq \emptyset$ und $f : A \to B$ eine Abbildung.

(1) Die Abbildung f ist ...

a) ... surjektiv genau dann, wenn es eine Abbildung $g : B \to A$ gibt mit der Eigenschaft $f \circ g = \mathrm{id}_B$. (**Rechtsinverse**)

b) ... injektiv genau dann, wenn es eine Abbildung $h : B \to A$ gibt mit der Eigenschaft $h \circ f = \mathrm{id}_A$. (**Linksinverse**)

c) ... bijektiv genau dann, wenn es Abbildungen $g : B \to A$ und $h : B \to A$ gibt mit den Eigenschaften $f \circ g = \mathrm{id}_B$ und $h \circ f = \mathrm{id}_A$. In diesem Fall ist $g = h = f^{-1}$ die Umkehrfunktion von f.

(2) Sei C eine weitere Menge und $g : B \to C$ eine weitere Abbildung, dann ist die Verkettung $g \circ f : A \to C$ definiert.

a) Sind f und g surjektiv, so auch $g \circ f$.
b) Ist $g \circ f$ surjektiv, so auch g.
c) Sind f und g injektiv, so auch $g \circ f$.
d) Ist $g \circ f$ injektiv, so auch f.
e) Sind f und g bijektiv, so auch $g \circ f$. In diesem Fall gilt $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Beweis (1)a) Es sei $g : B \to A$ eine Rechtsinverse von f und $y \in B$. Dann ist $g(y) \in A$ und $f(g(y)) = (f \circ g)(y) = y$, d.h. $g(y)$ ist ein Urbild von y und f ist surjektiv.
Nun sei f surjektiv und $y \in B$, dann gibt es (mindestens) ein $x \in A$ mit $f(x) = y$. Wir definieren $g(y) := x$, dann ist g rechtseindeutig und wegen der Surjektivität von f auch linksvollständig, also eine Abbildung $g : B \to A$. Ferner gilt $y = f(x) = f(g(y)) = (f \circ g)(y)$ für alle $y \in B$, also $f \circ g = \mathrm{id}_B$.

b) beweist man ganz ähnlich: Ist $h : B \to A$ eine Linksinverse von f und $f(x) = f(\tilde{x})$, so folgt $x = h(f(x)) = h(f(\tilde{x})) = \tilde{x}$, d.h. f ist injektiv.
Umgekehrt sei f injektiv, $x_0 \in A$ ein beliebiges Element und $y \in B$, dann definieren wir $h : B \to A$ wie folgt:

$$h(y) := \begin{cases} x, & \text{falls } y = f(x) \text{ für ein } x \in A, \\ x_0, & \text{sonst.} \end{cases}$$

h ist linksvollständig und wegen der Injektivität von f ist x im ersten Fall eindeutig, d.h. ist h rechtseindeutig und eine Abbildung $h : B \to A$. Nach Definition von h gilt $(h \circ f)(x) = h(f(x)) = x$ für alle $x \in A$, also $h \circ f = \mathrm{id}_A$.

c) Die Äquivalenz folgt aus a) und b), wir beweisen noch den Zusatz: Da f bijektiv ist, existiert die Umkehrfunktion f^{-1} und erfüllt (4.3); daraus folgt

$$\begin{aligned} h(y) &= h(f(f^{-1}(y))) = (h \circ f)(f^{-1}(y)) = f^{-1}(y), \\ g(y) &= f^{-1}(f(g(y))) = f^{-1}((f \circ g)(y)) = f^{-1}(y) \end{aligned}$$

für alle $y \in B$, also $h = f^{-1}$ und $g = f^{-1}$.

(2)a) Sei $z \in C$ und $y \in B$ ein Urbild von z unter g, d.h. $g(y) = z$. Dann gibt es ein Urbild $x \in A$ von y unter f und es folgt $(g \circ f)(x) = g(f(x)) = g(y) = z$, d.h. x ist auch ein Urbild von z unter $g \circ f$ und $g \circ f$ ist surjektiv.
b) Sei $z \in C$ und $x \in A$ ein Urbild von z unter $g \circ f$, dann ist $z = (g \circ f)(x) = g(f(x))$, d.h. $f(x) \in B$ ist ein Urbild von z unter g und g ist surjektiv.
c) Seien $x, \tilde{x} \in A$ mit $(g \circ f)(x) = (g \circ f)(\tilde{x})$, also $g(f(x)) = g(f(\tilde{x}))$. Aus der Injektivität von g folgt $f(x) = f(\tilde{x})$, aus der Injektivität von f dann $x = \tilde{x}$, also ist $g \circ f$ injektiv.
d) Seien $x, \tilde{x} \in A$ mit $f(x) = f(\tilde{x})$, dann ist $g(f(x)) = g(f(\tilde{x}))$, also $(g \circ f)(x) = (g \circ f)(\tilde{x})$. Wegen der Injektivität von $g \circ f$ folgt $x = \tilde{x}$ und f ist injektiv.
e) Die erste Aussage folgt aus a) und c). Für alle $z \in C$ gilt nun

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

und da die Umkehrfunktion $(g \circ f)^{-1}$ existiert, bedeutet das nach Definition $(g \circ f)^{-1}(z) = (f^{-1} \circ g^{-1})(z)$, also $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. □
Alternativer Beweis für (2)e): Es gilt auch $(f^{-1} \circ g^{-1}) \circ (g \circ f) = \mathrm{id}_A$, nach (1)c) ist also $g \circ f$ bijektiv und $f^{-1} \circ g^{-1}$ die Umkehrfunktion.

Man beweise selbst folgende Ergänzung zu (1): f ist surjektiv genau dann, wenn es höchstens eine Linksinverse gibt, und injektiv genau dann, wenn es höchstens eine Rechtsinverse gibt; f ist also bijektiv genau dann, wenn es genau eine Linksinverse oder genau eine Rechtsinverse gibt.

Bemerkung. Mit injektiven, surjektiven und bijektiven Funktionen kann man über unendliche Mengen und deren Mächtigkeit sprechen: Eine Menge A heißt **endlich**, wenn es eine Zahl $n \in \mathbb{N}$ und eine

$$\begin{array}{ll} \text{injektive Abbildung } f : A \to \{1, \dots, n\}, & \text{bijektive Abbildung } f : A \to \{1, \dots, n\} \\ \text{surjektive Abbildung } f : \{1, \dots, n\} \to A & \text{oder bijektive Abbildung } f : \{1, \dots, n\} \to A \end{array}$$

gibt – diese vier Aussagen sind äquivalent. In den beiden Fällen rechts ist $A = \{a_1, \dots, a_n\}$ mit der Mächtigkeit $\#A = n$. Eine nicht endliche Menge heißt **unendlich**.
Entsprechend heißt A **abzählbar**, falls es eine

$$\text{injektive Abbildung } f : A \to \mathbb{N} \text{ oder} \qquad \text{surjektive Abbildung } f : \mathbb{N} \to A \text{ gibt.}$$

Gibt es sogar eine bijektive Abbildung $f : \mathbb{N} \to A$ oder $f : A \to \mathbb{N}$, so ist $A = \{a_1, a_2 \dots\}$ gleichmächtig zu $\mathbb{N}$ und daher *abzählbar unendlich*. Eine nicht abzählbare Menge heißt **überabzählbar**.

Satz 4.25. Seien $k, j \in \mathbb{N}$ und $f : \{1, \dots, k\} \to \{1, \dots, j\}$ eine Abbildung.

(1) Ist f injektiv, so gilt $k \leqslant j$. Ist f surjektiv, so gilt $k \geqslant j$. Ist f bijektiv, so gilt $k = j$, d.h. die Mengen $\{1, \dots, k\}$ und $\{1, \dots, j\}$ sind gleichmächtig.

(2) Ist $k = j$, so gilt: f ist bijektiv $\Leftrightarrow$ f ist injektiv $\Leftrightarrow$ f ist surjektiv.

Gleichmächtigkeit von Mengen

Sei $A = \{a_1, \dots, a_n\}$ eine endliche Menge mit Mächtigkeit $\#A = n$, dann ist $\varphi_A : \{1, \dots, n\} \to A, i \mapsto a_i$ eine bijektive Abbildung. Ist B gleichmächtig zu A, so ist nach Satz 4.24(2)e) auch $\varphi_B \circ \varphi_A^{-1} : A \to B$ bijektiv.
Umgekehrt seien A, B endliche Mengen und $f : A \to B$ eine bijektive Abbildung. Wieder nach Satz 4.24(2)e) ist $\varphi_B^{-1} \circ f \circ \varphi_A : \{1, \dots, \#A\} \to \{1, \dots, \#B\}$ bijektiv, aus Satz 4.25(1) folgt $\#A = \#B$. Wir haben für endliche Mengen A, B bewiesen:

$$\#A = \#B \Leftrightarrow \text{ Es ex. eine bijektive Abbildung } f : A \to B. \tag{4.4}$$

Diese Aussage ist ein *Satz* für endliche Mengen; wir verwenden ihn nun als *Definition* für die Gleichmächtigkeit unendlicher Mengen. Man kann zeigen:

(1) $\#\mathbb{N} = \#\mathbb{Z}$, es gibt „genauso viele" natürliche wie ganze Zahlen.

(2) $\#(0, 1) = \#\mathbb{R}$, es gibt „genauso viele"reelle Zahlen wie im Intervall $(0, 1)$

Zum Beweis muss man nur bijektive Funktionen $f : \mathbb{Z} \to \mathbb{N}, g : (0, 1) \to \mathbb{R}$ finden (Tun Sie das!). Es gibt *unterschiedliche* unendliche Mächtigkeiten:

(3) $\#\mathbb{Q} = \#\mathbb{N}$. (*Erstes Diagonalargument* von Cantor)

(4) $\#\mathbb{R} \neq \#\mathbb{N}$. (*Zweites Diagonalargument* von Cantor)

Der Beweis wird in Analysis 1 geführt. Mengen, die zu $\mathbb{N}$ gleichmächtig sind (zum Beispiel $\mathbb{Z}, \mathbb{Q}$), nennt man **abzählbar unendlich**. $\#\mathbb{N}$ ist die kleinste unendliche Mächtigkeit, denn

(5) Ist $A \subseteq \mathbb{N}$ unendlich, so ist $\#A = \#\mathbb{N}$.

„Größere" Mengen als $\mathbb{N}$ (zum Beispiel $\mathbb{R}$) nennt man **überabzählbar**. Tatsächlich gibt es unendlich viele überabzählbare Mächtigkeiten, die *Kardinalzahlen*; sie sind gerade die Äquivalenzklassen der Äquivalenzrelation (4.4).
Es wurde lange vermutet, dass es keine Mächtigkeiten zwischen $\#\mathbb{N}$ und $\#\mathbb{R}$ gibt, die sog. *Kontinuumshypothese*. Der Logiker Kurt Gödel bewies 1938 die überraschende Antwort: Diese Frage ist unentscheidbar, also sowohl ihre Bejahung als auch ihre Verneinung widerspruchsfrei mit der restlichen Mathematik.

Anhang

Polynomdivision

Seien $P, Q \in \mathbb{Z}$. Bei der Division (mit Rest) $P : Q$ bestimmt man Zahlen $S \in \mathbb{Z}$ und $R \in \{0, \dots, Q-1\}$, sodass $P = S \cdot Q + R$ gilt; im Fall $R = 0$ ist P durch Q teilbar. Dasselbe gilt für Polynome $P(x), Q(x)$: Bei der **Polynomdivision** $P(x) : Q(x)$ sucht man Polynome $S(x)$ und $R(x)$, sodass $R(x)$ einen kleineren Grad als $Q(x)$ hat und $P(x) = S(x) \cdot Q(x) + R(x)$ gilt; ist $R(x)$ das Nullpolynom, so ist $P(x)$ durch $Q(x)$ teilbar und man hat $P(x)$ *faktorisiert*.

Beispiel. (1) Wir wollen $P(x) = x^3 + x^2 - 9x + 4$ durch $Q(x) = x - 1$ dividieren. Dazu dividieren wir den führenden Term x^3 von $P(x)$ durch den führenden Term x von $Q(x)$ und erhalten x^2. Dann berechnen wir $x^2 \cdot Q(x) = x^3 - x^2$ und subtrahieren das Ergebnis von $P(x)$:

$$P(x) - x^2 \cdot Q(x) = (x^3 + x^2 - 9x + 4) - (x^3 - x^2) = 2x^2 - 9x + 4.$$

Wir haben nun einen Teil von $P(x)$ (nämlich $x^2 \cdot Q(x)$) durch $Q(x)$ dividiert, es bleibt ein Restterm $2x^2 - 9x + 4$. Mit diesem tun wir dasselbe wie mit $P(x)$: Wir teilen $2x^2$ durch x, multiplizieren das Ergebnis $2x$ mit $Q(x)$ und subtrahieren

$$P(x) - x^2 \cdot Q(x) - 2x \cdot Q(x) = (2x^2 - 9x + 4) - (2x^2 - 2x) = -7x + 4.$$

Führen wir noch einen Schritt mit dem Restterm $-7x + 4$ aus, so ergibt sich

$$P(x) - x^2 \cdot Q(x) - 2x \cdot Q(x) + 7 \cdot Q(x) = -7x + 4 + (7x - 7) = -3.$$

Der Grad des Restpolynoms $R(x) = -3$ (nämlich 0) ist kleiner als der Grad von $Q(x)$ (nämlich 1), daher ist die Polynomdivision beendet; das Ergebnis ist

$$P(x) = (x^2 + 2x - 7) \cdot Q(x) - 3 = S(x) \cdot Q(x) + R(x).$$

Zur Probe kann man die rechte Seite ausmultiplizieren und mit $P(x)$ vergleichen.

Dieses Verfahren verläuft völlig identisch zur schriftlichen Division, daher wird die Polynomdivision häufig so aufgeschrieben:

$$
\begin{array}{rrrrl}
(x^3 & +x^2 & -9x & +4) & = (x-1)\cdot(x^2 \; +2x \; -7). \\
-(x^3 & -x^2) & & & \\
\cline{1-2}
 & 2x^2 & -9x & & \\
 & -(2x^2 & -2x) & & \\
\cline{2-3}
 & & -7x & +4 & \\
 & & -(-7x & +7) & \\
\cline{3-4}
 & & & -3 &
\end{array}
$$

(2) Um $P(x) = 6x^4 + 2x^3 - 5x + 2$ durch $Q(x) = x^2 - 3$ zu dividieren, berechnen wir:

$$
\begin{array}{rrrrrl}
(6x^4 & +2x^3 & & -5x & +2) & = (x^2-3)\cdot(6x^2 \; +2x \; +18). \\
-(6x^4 & & -18x^2) & & & \\
\cline{1-3}
 & 2x^3 & +18x^2 & -5x & & \\
 & -(2x^3 & & -6x) & & \\
\cline{2-4}
 & & 18x^2 & +x & +2 & \\
 & & -(18x^2 & & -54) & \\
\cline{3-5}
 & & & x & +56 &
\end{array}
$$

Damit ist $P(x) = (6x^2 + 2x + 18)Q(x) + x + 56$.

(3) Vergleiche mit Beispiel 3.11: Hat ein Polynom $P(x)$ die Nullstelle x_0, d.h. $P(x_0) = 0$, so ist $P(x)$ durch den Linearfaktor $(x - x_0)$ teilbar, man kann also die Polynomdivision durch $Q(x) = x - x_0$ ohne Rest durchführen und erhält die Faktorisierung $P(x) = S(x)(x - x_0)$. Kennt man weitere Nullstellen von $S(x)$ (also von $P(x)$), so kann man $S(x)$ weiter faktorisieren; mit jeder Nullstelle reduziert sich der Grad des Restpolynoms um 1. Zum Beispiel hat $P(x) = x^3 - 6x^2 - x + 6$ die Nullstelle $P(1) = 0$, wir können also ohne Rest durch $Q(x) = x - 1$ dividieren:

$$
\begin{array}{rrrrl}
(x^3 & -6x^2 & -x & +6) & = (x-1)\cdot(x^2 \; -5x \; -6). \\
-(x^3 & -x^2) & & & \\
\cline{1-2}
 & -5x^2 & -x & & \\
 & -(-5x^2 & +5x) & & \\
\cline{2-3}
 & & -6x & +6 & \\
 & & -(-6x & +6) & \\
\cline{3-4}
 & & & 0 &
\end{array}
$$

Die übrigen Nullstellen $-1, 6$ des quadratischen Faktors findet man leicht; damit ist $P(x) = (x - 1)(x^2 - 5x - 6) = (x - 1)(x + 1)(x - 6)$.

Partialbruchzerlegung

Eine (gebrochen-)rationale Funktion ist von der Form $f(x) = \frac{P(x)}{Q(x)}$ mit Polynomen $P(x), Q(x)$, siehe Beispiel 4.11(3). Nach der Polynomdivision $P(x) = S(x)Q(x) + R(x)$ hat f die Darstellung $f(x) = S(x) + \frac{R(x)}{Q(x)}$ und $R(x)$ einen kleineren Grad als $Q(x)$.
In dieser Situation möchte man den Bruch $\frac{R(x)}{Q(x)}$ häufig weiter in *Partialbrüche* zerlegen, sodass die Nullstellen von $Q(x)$ isoliert auftreten. Mittels Polynomdivision findet man zunächst eine Linearfaktorisierung von $Q(x)$:

$$Q(x) = (x - x_1)^{r_1} \ldots (x - x_k)^{r_k}.$$

Dabei sind die Nullstellen $x_1, \ldots, x_k$ paarweise verschieden und $r_1, \ldots, r_k \in \mathbb{N}$ deren Vielfachheiten; es ist $r_1 + \ldots + r_k$ der Grad von $Q(x)$. Gesucht ist nun die Partialbruchzerlegung von $\frac{R(x)}{Q(x)}$, also eine Darstellung der Form

$$\frac{R(x)}{Q(x)} = \frac{a_{1,1}}{x - x_1} + \ldots + \frac{a_{1,r_1}}{(x - x_1)^{r_1}} + \ldots + \frac{a_{k,1}}{x - x_k} + \ldots + \frac{a_{k,r_k}}{(x - x_k)^{r_k}}$$

mit (zu bestimmenden) Zahlen $a_{1,1}, \ldots, a_{k,r_k} \in \mathbb{R}$.

Beispiel. (1) Wir wollen den Quotienten $\frac{x-11}{x^2+5x-14}$ in Partialbrüche zerlegen. Dazu faktorisieren wir den Nenner $x^2 + 5x - 14 = (x-2)(x+7)$ und machen den Ansatz

$$\frac{x - 11}{x^2 + 5x - 14} = \frac{a}{x - 2} + \frac{b}{x + 7}.$$

Aus dieser Bruchgleichung (mit der Grundmenge $\mathbb{R} \setminus \{2, -7\}$) sind die Zahlen $a, b \in \mathbb{R}$ zu bestimmen. Multiplikation mit dem Hauptnenner führt auf

$$x - 11 = (a + b)x + (7a - 2b).$$

Wir machen nun einen *Koeffizientenvergleich*: Das rechte und das linke Polynom können auf $\mathbb{R} \setminus \{2, -7\}$ nur dann übereinstimmen, wenn sie identisch sind, also dieselben Koeffizienten haben. Dies führt auf das lineare Gleichungssystem

$$1 = a + b, \qquad -11 = 7a - 2b.$$

Dessen Lösung ist $a = -1, b = 2$ und daher die gesuchte Partialbruchzerlegung

$$\frac{x - 11}{x^2 + 5x - 14} = \frac{-1}{x - 2} + \frac{2}{x + 7}.$$

(2) Für die Partialbruchzerlegung von $\frac{1}{x^3-x^2}$ faktorisieren wir $x^3 - x^2 = x^2(x-1)$, machen den Ansatz

$$\frac{1}{x^3 - x^2} = \frac{a}{x} + \frac{b}{x^2} + \frac{c}{x-1}$$

und erhalten nach Multiplikation mit $x^3 - x^2$ auf der Grundmenge $\mathbb{R} \setminus \{0, 1\}$ die Gleichung $1 = (a + c)x^2 + (-a + b)x - b$; wir könnten nun mit einem Koeffizientenvergleich wieder ein lineares Gleichungssystem ablesen und lösen. Alternativ argumentiert man so: Wenn die Gleichung auf $\mathbb{R} \setminus \{0, 1\}$ gilt, dann auch *für alle* $x \in \mathbb{R}$ und wir können einsetzen:

$$x = 0 \Rightarrow b = -1, \qquad x = 1 \Rightarrow c = 1, \qquad x = -1 \Rightarrow a = -1.$$

Wir haben also die Partialbruchzerlegung

$$\frac{1}{x^3 - x^2} = \frac{-1}{x} + \frac{-1}{x^2} + \frac{1}{x-1}.$$

(3) Im Polynom $x^3 + x^2 + x + 1 = (x + 1)(x^2 + 1)$ hat der Faktor $(x^2 + 1)$ keine reellen Nullstelen, daher machen wir in der folgenden Partialbruchzerlegung den Ansatz

$$\frac{2x^2}{(x+1)(x^2+1)} = \frac{a}{x+1} + \frac{bx+c}{x^2+1}.$$

Nach Multiplikation dieser Gleichung mit $(x+1)(x^2+1)$ erhält man die Gleichung $2x^2 = (a + b)x^2 + (b + c)x + a + c$ und mit den Methoden aus (1) oder (2) die Lösung $a = b = 1, c = -1$, also

$$\frac{2x^2}{(x+1)(x^2+1)} = \frac{1}{x+1} + \frac{x-1}{x^2+1}.$$

Index

$\mathbb{N}, \mathbb{N}_0$, *siehe* natürliche Zahlen
$\mathbb{Z}$, *siehe* ganze Zahlen
$\mathbb{Z}/_{n\mathbb{Z}}, \mathbb{Z}_n$, *siehe* Restklasse
$\mathbb{P}$, *siehe* Primzahlen
$\mathbb{Q}$, *siehe* rationale Zahlen
$\mathbb{R}$, *siehe* reelle Zahlen
$(a, b), (a, b], [a, b), [a, b]$, *siehe* Intervall
$\mathbb{R}^2$, *siehe* Ebene
$\mathbb{R}^3$, *siehe* Raum
$\mathbb{C}$, *siehe* komplexe Zahlen
$A(x)$, *siehe* Aussageform
$\forall, \exists, \nexists, \exists$, *siehe* Quantoren
$\neg A, \overline{A}$, *siehe* Negation
$\wedge$, *siehe* Und
$\vee$, *siehe* Oder
$\Rightarrow$, *siehe* Implikation
$\Leftrightarrow$, *siehe* Äquivalenz
$\leftrightarrow$, *siehe* tautologische Äquivalenz
$\top$, *siehe* Tautologie
$\perp$, *siehe* Kontradiktion
$\in, \notin$, *siehe* Element
$\subseteq, \subsetneqq, \subset$, *siehe* Teilmenge
$\cap$, *siehe* Schnittmenge
$\cup$, *siehe* Vereinigungsmenge
$\setminus$, *siehe* Differenzmenge
M^C, *siehe* Komplement
$\times$, *siehe* Kartesisches Produkt
(x, y), *siehe* Paar
$\#M$, *siehe* Mächtigkeit
$\#M = \#N$, *siehe* Gleichmächtigkeit
$\{x_1, \ldots, x_n\}$, *siehe* endlich
$\{x_1, x_2, \ldots\}$, *siehe* abzählbar
$\emptyset$, *siehe* leere Menge
$\mathcal{P}(M)$, *siehe* Potenzmenge
G, *siehe* Grundmenge
$\mathcal{L}$, *siehe* Lösungsmenge
Ω, *siehe* Ergebnisraum
$\leqslant, <, \geqslant, >$, *siehe* Ungleichheitszeichen
$|x|$, *siehe* Betrag
$|x - y|$, *siehe* Abstand
$\operatorname{sign}(x)$, *siehe* Signum
$\max, \min$, *siehe* Maximum, Minimum
x^a, *siehe* Potenzfunktion
$\sqrt[n]{x}$, *siehe* Wurzelfunktion
a^x, *siehe* Exponentialfunktion
$\log x$, *siehe* Logarithmus
$\sin, \cos$, *siehe* Sinus, Cosinus
$\tan, \cot$, *siehe* Tangens, Cotangens
$\arcsin, \arccos, \arctan, \operatorname{arccot}$, *siehe* Arkusf.
$\sinh, \cosh$, *siehe* hyperbolische F.
$\operatorname{arsinh}, \operatorname{arcosh}$, *siehe* Areaf.
xRy, *siehe* Relation
$f : A \to B$, *siehe* Abbildung
$\operatorname{Abb}(A, B)$, *siehe* Abbildung
B^A, *siehe* Abbildung
(a_n), *siehe* Folge
$x \mapsto f(x)$, *siehe* Abbildungsvorschrift
$f(x)$, *siehe* Wert
$f(X)$, *siehe* Bild
$f^{-1}(Y)$, *siehe* Urbild
$\operatorname{gr}(f)$, *siehe* Graph
id, *siehe* Identität
$\circ$, *siehe* Verkettung
f^n, *siehe* mehrfache Verkettung
f^{-1}, *siehe* Umkehrfunktion
f', *siehe* Ableitung

$*$, *siehe* Verknüpfung
$t|n$, *siehe* Teilbarkeit
$n \equiv m \mod t$, *siehe* Modulo
$y = x^2$, *siehe* Normalparabel
$x^2 = 2$, *siehe* $\sqrt{2}$ ist irrational
e, *siehe* Eulersche Zahl
$\sin^2(x)$, 94
$:=$, 7
$\approx$, 7
$f \neq g$, 96

Abbildung, 80–85
  Gleichheit, 82, 96
  mehrstellige, 84
  wohldefinierte, 82–84
Abbildungsvorschrift, 81–85
abgeschlossen
  bzgl. $*$, 6, 84
  Intervall, 36–38
Ableitung, 85, 87, 89
Abschätzung, 53–56
Abstand, 57, 66
Abszisse, 46
abzählbar, 61, 105, 106
Additionstheorem, 88, 89
Algebra, 6–9, 56, 64, 65, 70–72, 84, 94, 97, 103, 107–109
Analysis, 6, 7, 16, 24, 30, 52, 55, 58, 70, 84, 85, 96, 97, 101, 102, 105, 106
Anordnungsaxiome, 49–52, 78
Äquivalenz, 18, 19, 22–24, 26, 28, 33, 77
  tautologische, 9, 20, 22, 28, 78
Äquivalenzklasse, 78, 106
Äquivalenzumformung, 6, 19, 58, 61, 62, 66–70
Argument, *siehe* Stelle
Arithmetisches Mittel, 54
Assoziativgesetz, 6, 20, 41, 49, 93, 97
Asymptote, 86
Aussage, 8, 9
Aussageform, 8–10, 31, 61
Auswahlaxiom, 8, 106
Axiom, 4

Basis, 7, 68, 70, 87
Behauptung, 16
Betrag, 56–60, 84, 87, 94–98
Beweis, 4, 16
  Beweisstrategien, 17, 23
  direkter, 17, 23
  durch Fallunterscheidung, *siehe* Fallunterscheidung
  Gegenbeispiel, 26, 77
  indirekter, 22, 24
  Kontraposition, 22, 23, 45
  mit Allquantor, 25
  Ringschluss, 26, 42, 99
  vollständige Induktion, 27
  Zerlegung, 24, 28, 34, 42, 51, 92, 104
bijektiv, 91, 98–106
Bild, 80, 81, 89–92, 98–100, 103
Bildungsgesetz, 85
Binomische Formel, 6, 10, 11
Bogenmaß, 87

Cauchy-Schwarz-Ungleichung, 54, 56

Deduktive Methode, 4, 27, 50–52
Definition, 4
Definitionslücke, 44, 83, 86, 109
  hebbar, Pol, 86, 97
Definitionsmenge, -bereich, 81–83, 85–87
dicht, 6, 55
Differentialgleichung, 61
differenzierbar, 85, 87
Differenzmenge, 43–45
disjunkt, 39, 40, 77
Distributivgesetz, 6, 20, 41, 49, 97
Drehung, 85
Dreiecksungleichung, 59, 60
  umgekehrte, 59, 60
Durchschnitt, *siehe* Schnittmenge

Ebene, 34, 46
Einbettung, 94

Einheitskreis, 34, 48, 73, 75, 83, 87, 102
Einheitswürfel, 48
Einschränkung, 82, 88, 96, 98, 101
endlich, 29, 31, 35, 41, 61, 78, 103, 105, 106
Entweder oder, 13
Ereignis, 32, 41
Ergebnisraum, 32, 41, 47, 85
Eulersche Zahl, 7, 87
ex falso quodlibet, 16
Exponent, 7, 68, 70, 86, 101

Fallunterscheidung, 22, 25, 52, 66–69, 72
Flächeninhalt, 48
Folge, 85
Fortsetzung, 82
Funktion, 81–85
  abschnittweise definierte, 82, 87
  (affin-)lineare, 85, 94, 99
  Funktionsgleichung, *siehe* Abbildungsvorschrift
  Funktionswert, *siehe* Wert
  konstante, 84, 86, 97
  monotone, 57, 62, 84, 102
  partielle, Multif., 83
  periodische, 84, 96
  (un)beschränkte, 84
  (un)gerade, 84, 96
Funktional, 85
Funktionalgleichung, 101
Funktionenschar, 101

Gauß-Algorithmus, 71
Gebundene Variable, 11, 31
Geometrie
  analytische, 30, 32, 46, 56, 72, 73
  euklidische, 8, 9, 12, 13, 33, 40, 48, 72, 73, 83, 85
Gesetz vom
  ausgeschlossenen Dritten, tertium non datur, 14, 44
  ausgeschlossenen Widerspruch, 14, 44
Gitter, 47
Gleichmächtigkeit, 29, 35, 75, 78, 106
Goldbachsche Vermutung, 8, 9
Grad, 64, 65, 86, 107, 109
Graph
  aus Knoten und Kanten, 75
  einer Abbildung, 82–84, 94, 98, 102
Grundmenge, 31, 43–45, 61, 67–70, 109

Halbebene, Halbraum, 72
Hauptnenner, 69, 109
hinreichend, 16, 18
Homogenität, 59

Identität, 86, 94
im Allgemeinen, 26
Implikation, 15–17, 22–27, 33, 78
injektiv, 62, 98–100, 104, 105
Integral, 85
Intervall, 36–38, 44, 48, 61
  beschränkt, unbeschränkt, 37, 38
  echt, unecht, Länge, 37
  kompakt, 37
  offen, abgeschlossen, halboffen, 36, 38
Inverse, *siehe* Umkehrfunktion
Inversion, 86

Kardinalität, *siehe* Mächtigkeit
Kardinalzahl, 106
Kartesisches Produkt, 46–48, 74
Koeffizient, 64
Koeffizientenvergleich, 109
Kommutatives Diagramm, 93
Kommutativgesetz, 6, 20, 41, 49, 93, 97
Komplement, 43–45
Kontinuumshypothese, 35, 106
Kontradiktion, 14, 20, 22, 24
Koordinatensystem, 46–48, 63, 72–75, 80–83, 85, 87, 88, 90, 94, 99, 102
Kurvendiskussion, 70

Linearfaktorisierung, 63–65, 108, 109
Linearkombination, 95

Logarithmengesetze, 7, 101
lösbar, 11, 61, 99
Lösungsformel, 6, 63, 64
Lösungsmenge, 10, 11, 19, 31, 40, 61, 62, 71–73, 91, 99

Maximum, Minimum, 58, 97
Menge, 29–35
  Element, 29–33
  Gleichheit, 33–35, 75–78
  leere, 29, 30, 36, 37, 40
  Schreibweise, 29–31
Mengensystem, 36, 42
Modulo, 76, 78
modus ponens, 16
monoton wachsend, fallend, *siehe* monotone Funktion
Mächtigkeit, 29–31, 35, 36, 41, 44, 46, 61, 105, 106

Nachfolger, 6
Negation, 9, 20–23, 26, 43
  bei Quantoren, 11, 12
  doppelte, 9, 21, 45
Neutrales Element, 94, 97
Newton-Verfahren, 70
Normalparabel, 16, 24–26, 30, 48, 50, 73, 75, 80–85, 89, 90, 93, 94, 98, 103
notwendig, 16, 18
Numerus, 7

Obermenge, *siehe* Teilmenge
Oder, 13, 14, 17, 20–23, 39
Oktand, 46
Operator, 85
Ordinate, 46

Paar, 46–48
Paradoxon, 8
  Russellsches, 32
Partialbruchzerlegung, 6, 86, 109
(Partiell) Geordnete Menge, *siehe* Ordnungsrelation
Periode, 84, 87, 88, 96
Permutation, 103
Pfeildiagramm, 75, 80
Physik, 47, 95
Polyeder, 72
Polynom, 63–65, 82, 86, 107–109
Polynomdivision, 6, 64, 86, 107–109
positiv definit, 59
Positivteil, Negativteil, 97
Potenz, *siehe* Potenzfunktion
Potenzgesetze, 7, 103
Potenzmenge, 33, 36
Primzahl, 6, 8, 9, 17, 40, 44
Prinzip
  der Extensionalität, 33
  der Zweiwertigkeit, 8
Probe, 62, 67

Quader, 48
Quadrant, 46
Quadratische Ergänzung, 63
Quantor, 10–12
  Allquantor, 10, 11, 25–27
  Existenzquantor, 10, 11, 26
  Reihenfolge bei, 12
  Verträglichkeit, 14, 92

Raum, 46
Rechenregeln
  für Aussagen, 9, 11, 14, 20, 22
  für Bild und Urbild, 92, 99
  für Mengen, 31, 35, 40–45, 76–78
  für Ungleichungen, 25, 49–54, 58–60, 68, 70, 76–78
  für Verkettungen, 93–97, 100–105
  für Zahlen, 6, 7, 18, 49, 58–60, 68, 76–78, 87–89, 101
Rechteck, 48
Rechtsinverse, Linksinverse, 103–105
Reelle Funktion, 85
  Areaf., 89, 101
  Arkusf., 88, 101

Betragf., *siehe* Betrag
Exponentialf., 7, 70, 84, 87, 101, 102
ganzrationale, *siehe* Polynomf.
gebrochen-rationale, *siehe* rationale
hyperbolische, 89, 101
Logarithmus, 7, 70, 82, 87, 101, 102
Polynomf., 84, 86, 95
Potenzf., 7, 84, 86, 95, 101, 102
rationale, 86, 109
Signumf., *siehe* Signum
Sinus, Cosinus, 84, 87, 88, 90, 101
Tangens, Cotangens, 88
trigonometrische, *siehe* Sinus, Cosinus, Tangens, Cotangens
Wurzelf., 7, 57, 86, 101, 102
Relation, 74–81
antisymmetrische, 34, 49, 51, 76–78
Äquivalenzr., 77, 78, 106
linkseindeutige, 79–82, 98
linksvollständige, 79–83, 90
Ordnungsr., 77, 78, 97
punktweise, 95–97
rechtseindeutige, 79–83, 86
rechtsvollständige, 79–82, 98
reflexive, 36, 49, 76–78
symmetrische, 49, 51, 76–78, 80
totale, 49, 52, 76–78, 97
transitive, 49–52, 76–78
Repräsentant, 78
Rest, 76, 78, 107
Restklasse, 78

Satz, 4, 16
des Pythagoras, 8, 9
Fundamentalsatz der Algebra, 65
Lemma von Gauß, 64
vom Nullprodukt, 18, 41, 63
von de Morgan, 20, 22, 45
von Gödel, 8, 106
von Vieta, 63
Scheinlösung, 19, 62, 67
Schnittmenge, 39–45, 66, 71, 72
Signum, 56, 84, 87
Spiegelung, 85
Sprungstelle, 87
Stelle, Wert, 81–83, 85
stetig, 87
Stochastik, 32, 41, 47, 85
streng monoton wachsend, fallend, *siehe* monotone Funktion
Substitution, 68, 71
Suchmaschine, 39
surjektiv, 98–100, 104, 105
Symmetrische Gruppe, 103

Tautologie, 14, 20, 22, 28, 78
Teilbarkeit, 17, 27, 76–78
Teilmenge, 33–38, 40–45, 75–78
echte, 33, 35, 37, 77, 78
Total geordnete Menge, *siehe* Ordnungsrelation
Tupel, 46–48

überabzählbar, 61, 105, 106
Übungsaufgabe, 8, 14, 16, 21, 23, 26, 34, 36, 40, 45, 53, 55, 59, 68, 76–78, 94, 96–98, 102, 105, 106
Umgebung, 37, 59
Umkehrfunktion, 86–89, 91, 99–105
Umkehrrelation, 80, 100, 102
Umkehrung, 17, 23
Und, 13, 14, 17, 20–23, 39
unendlich, 31, 35, 36, 61, 78, 105, 106
Ungleichheitszeichen, 10, 14, 25, 35, 44, 49–56, 58–60, 72, 75–78
(Un-)Gleichung, 6, 10–12, 19, 31, 41, 49, 60–62, 70, 91, 99, 101
biquadratische, 68
komplexe, 70–73
lineare, 63
mit Beträgen, 57–60, 66
mit Brüchen, 69
mit Polynomen, 63, 64
mit Wurzeln, 67, 68

quadratische, 6, 63
quadrieren, 19, 25, 50–54, 58, 62, 66–68, 101
(Un-)Gleichungssystem, 40, 71–73
lineares (LGS), 71, 72, 109
Ungleichungskette, 72
Urbild, 80, 81, 89–92, 98–100, 103
Ursprung, 46

Vektorraum, 97
Venn-Diagramm, 39, 43, 45
Vereinigungsmenge, 39–45, 61, 66
Verkettung, 93–97, 100–106
mehrfache, 94, 103
Verknüpfung, 13, 84
logische, 13, 19
punktweise, 95
von Mengen, 38
Verschiebung, 85
Verträglichkeit mit $+, \cdot$, 49–54, 62, 78
Vielfachheit, 65, 109
Vollständiges System, 28
Volumen, 48
Voraussetzung, 16

Wahrheitstafel, 9, 13–15, 18–20, 23, 28
Wahrheitswert, 8, 78
Wahrscheinlichkeit, 41
Wertemenge, -bereich, 81–83, 85–87
Wurzel, *siehe* Wurzelfunktion
$\sqrt{2}$ ist irrational, 6, 11, 16, 24, 25, 31, 37

Zahlen
ganze, natürliche, 6, 12, 27, 31, 35–37, 44, 47, 64, 76–78, 86, 88
komplexe, 6, 57, 65, 89
negative, nichtpositive, 38, 44, 50–55
positive, nichtnegative, 38, 44, 49–62
Primzahlen, *siehe* Primzahl
rationale, 6, 11, 25, 31, 35–37, 64, 86
reelle, 6, 11, 31, 35–37, 61, 86
(un)gerade, 7–9, 14, 17, 23–26, 30–34, 40, 44, 84, 101
Zahlengerade, 36–38, 50, 55–57, 66, 78
Zahlenmengen, 6, 30, 35–38, 75, 81
Zentrische Streckung, 85
ZF-System, 32
Zufallsvariable, 85
Zuordnung, 79–82, 91