

Vorkurs Mathematik

September 2016

Janko Latschev und Immanuel van Santen

Inhaltsverzeichnis

1	Elementares Rechnen	4
1.1	Zahlenbereiche und deren Erweiterungen	4
1.2	Weitere Notationen und nützliche Formeln	15
2	Gleichungen	19
2.1	Lineare Gleichungen	21
2.2	Quadratische Gleichungen	21
2.3	Lösung von Gleichungen durch Substitution	23
2.4	Bruch- und Wurzelgleichungen	23
2.5	Gleichungen mit Beträgen	26
2.6	Schlussbemerkung	27
3	Ungleichungen	28
3.1	Lineare Ungleichungen	29
3.2	Quadratische Ungleichungen	30
3.3	Betragsungleichungen	32
3.4	Zusammenfassung	35
4	Lineare Algebra und Geometrie	37
4.1	Vektoren	37
4.2	Geraden im \mathbb{R}^2	45
4.3	Ebenen im \mathbb{R}^3	50
4.4	Lineare Gleichungssysteme in 2 bzw. 3 Variablen	53
4.5	Relative Lage von Punkten, Geraden und Ebenen	60
5	Die Sprache der Mathematik	65
5.1	Definitionen	65
5.2	Aussagen	66
5.3	Wie man aus Aussagen neue Aussagen gewinnt	67
6	Abbildungen und Funktionen	75
6.1	Etwas mehr über Mengen	75
6.2	Abbildungsbegriff	78
6.3	Elementare Funktionen	82
7	Was ist eigentlich ein Grenzwert?	91
7.1	Folgen	93
7.2	Konvergenz	94

7.3	Reihen: Ein Ausblick	97
8	Was ist eine stetige Funktion?	100
9	Äquivalenzrelationen	101
9.1	Definition und erste Beispiele von Äquivalenzrelationen	101
9.2	Äquivalenzklassen	103
9.3	Die Konstruktion der rationalen Zahlen aus den ganzen Zahlen	105
9.4	Vektoren als Pfeilklassen	107
10	Beweismethoden	109
10.1	Direkter Beweis	110
10.2	Beweis durch Fallunterscheidungen	115
10.3	Beweis durch Widerspruch	118
10.4	Beweis durch Kontraposition	119
10.5	Beweis durch vollständige Induktion	121
10.6	Die Rolle von Beispielen und Gegenbeispielen	126
10.7	Zusammenfassung	129
11	Differentialrechnung	130
11.1	Tangenten und die Definition der Differenzierbarkeit	130
11.2	Ableitungen elementarer Funktionen	133
11.3	Stetigkeit von differenzierbaren Abbildungen	134
11.4	Differentiationsregeln	135
11.5	Anwendungen	140
12	Integralrechnung	145
12.1	Definition des Integrals	145
12.2	Stammfunktionen	149
12.3	Integrationsregeln	151
12.4	Flächenberechnung mittels Integration	158

1 Elementares Rechnen

Die meisten Menschen verbinden mit dem Begriff *Mathematik* in erster Linie diverse mehr oder weniger komplizierte Rechenschemata, welche sie in der Schule gelernt haben. Auch wir beginnen diesen Vorkurs mit einem kurzen Überblick über Zahlbereiche und ihre Eigenschaften, sowie elementare Rechenoperationen.

1.1 Zahlenbereiche und deren Erweiterungen

Wir werden die folgenden aus der Schule wohlbekanntesten Zahlenbereiche wiederholen:

- die Menge \mathbb{N} der natürlichen Zahlen,
- die Menge \mathbb{Z} der ganzen Zahlen,
- die Menge \mathbb{Q} der rationalen Zahlen und
- die Menge \mathbb{R} der reellen Zahlen.

Dabei verwenden wir einen naiven Mengenbegriff.

Definition. Eine *Menge* ist eine Sammlung wohlunterscheidbarer Objekte. Die einzelnen Bestandteile einer Menge nennt man die *Elemente der Menge*.

Diese "Definition" ist nicht sehr präzise, aber zusammen mit einigen Beispielen wird sie uns zunächst genügen. Im Verlaufe des Vorkurses werden wir dann noch etwas mehr über Mengen zu sagen haben. Es gibt verschiedene Möglichkeiten, Mengen symbolisch zu beschreiben, zum Beispiel kann man die Elemente einer Menge aufzählen, wenn es nicht allzu viele sind. So bezeichnet

$$M = \{1, 2, 3, 5, 7\}$$

die Menge M , die aus den fünf Zahlen 1, 2, 3, 5 und 7 besteht. Geschweifte Klammern geben hier den Beginn und das Ende der Aufzählung an. Um auszudrücken, dass ein gewisses Objekt Element einer Menge ist, schreiben wir $m \in M$, also zum Beispiel

$$3 \in \{1, 2, 3, 5, 7\} \quad \text{oder kurz} \quad 3 \in M,$$

da wir ja bereits die Bezeichnung M für unsere Beispielmenge eingeführt haben. Um auszudrücken, dass ein gewisses Objekt nicht Element einer Menge ist, schreiben wir $n \notin M$, also zum Beispiel

$$4 \notin \{1, 2, 3, 5, 7\} \quad \text{oder kurz} \quad 4 \notin M.$$

Eine *Teilmenge* einer gegebenen Menge M ist eine Menge N , so dass jedes Element von N auch ein Element von M ist. Man schreibt dann $N \subseteq M$, also zum Beispiel

$$\{1, 3, 5\} \subseteq M.$$

Eine Menge (oder eine Teilmenge einer Menge) kann auch *leer* sein, d.h. keine Elemente enthalten. Für eine solche leere Menge schreiben wir $\{\}$, oder noch öfter \emptyset . Die leere Menge ist Teilmenge jeder Menge.

Dies soll zunächst genügen.

Natürliche Zahlen

Definition. Als *natürlichen Zahlen* bezeichnen wir die Elemente der Menge

$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}.$$

Außerdem definieren wir die Menge $\mathbb{N}_0 = \{0, 1, 2, 3, 4, \dots\}$, in der neben den natürlichen Zahlen auch die Null mit eingeschlossen ist.

Diese Bezeichnungen sind nicht einheitlich, häufig wird in Lehrbüchern und Vorlesungen auch die 0 als natürliche Zahl bezeichnet.

Wie man bereits in den ersten Schuljahren lernt, können natürliche Zahlen addiert und multipliziert werden. Beide Operationen sind kommutativ, das heißt, sie hängen nicht von der Reihenfolge ab: $a + b = b + a$ bzw. $a \cdot b = b \cdot a$. Beide Operationen sind auch assoziativ, d.h. man kann bei mehrfacher Anwendung derselben Operation die Klammern "verschieben": $(a + b) + c = a + (b + c)$ und $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. Bei der Multiplikation lässt man den Punkt meist weg, was wegen der Assoziativität unproblematisch ist. Außerdem gilt das sogenannte Distributivgesetz $a \cdot (b + c) = a \cdot b + a \cdot c$.

Für kleine Zahlen entsprechen diese Regeln unserer Anschauung, und können durch einfaches Abzählen verifiziert werden. Allgemeiner sind sie die wesentlichen Eigenschaften der natürlichen Zahlen, welche allen Rechnungen zugrunde liegen. Wenn Sie zum Beispiel jemandem erklären wollen, wie und warum schriftliche Addition funktioniert, werden Sie ohne das Distributivgesetz nicht sehr weit kommen.

Während die Addition und Multiplikation für alle Paare natürlicher Zahlen definiert sind und wieder eine natürliche Zahl liefern, sind deren Umkehrungen, die Subtraktion und die Division, nur für gewisse geordnete Paare natürlicher Zahlen definiert. So gibt es auf den natürlichen Zahlen eine Relation $<$ mit deren Hilfe sie angeordnet werden können: $1 < 5$, $13 < 21$ usw. Um eine Zahl $m \in \mathbb{N}$ von einer Zahl $n \in \mathbb{N}$ subtrahieren zu können, müssen die beiden Zahlen die Relation $m < n$ erfüllen. So ist zum Beispiel $8 - 3 = 5 \in \mathbb{N}$, aber $3 - 8$ ist keine natürliche Zahl.

Um auszudrücken, dass eine natürliche Zahl n durch eine andere natürliche Zahl m dividiert werden kann, gibt es den Begriff des Teilers.

Definition. Eine natürliche Zahl m heißt *Teiler* einer natürlichen Zahl n , falls eine weitere natürliche Zahl $k \in \mathbb{N}$ existiert, so dass

$$n = m \cdot k.$$

Zum Beispiel hat 12 die Teiler

1, 2, 3, 4, 6 und 12,

und die Teiler von 27 sind

1, 3, 9 und 27.

Man sieht nun sofort ein, dass die Division von n durch m innerhalb der natürlichen Zahlen genau dann funktioniert, wenn m ein Teiler von n ist. Eine interessante Klasse von natürlichen Zahlen bilden die sogenannten Primzahlen.

Definition. Eine natürliche Zahl ist eine *Primzahl*, wenn sie größer als 1 ist und nur 1 und sich selbst als Teiler besitzt.

Die Definition der Primzahl geht schon auf die alten Griechen zurück. Die Primzahlen sind bis in die heutige Zeit von enormer Bedeutung, sowohl innerhalb der Mathematik als auch im täglichen Leben. So basieren gängige Verschlüsselungsalgorithmen für Daten auf der Tatsache, dass es keine effiziente Methode gibt, sehr große Zahlen (mit mehreren Hundert Dezimalstellen) in Primfaktoren zu zerlegen.

Die ersten Primzahlen lauten

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, . . .

Es gibt außer der Definition noch eine andere nützliche Charakterisierung der Primzahlen.

Lemma 1 (Lemma von Euklid). *Eine natürliche Zahl $p > 1$ ist genau dann eine Primzahl, wenn sie folgende Eigenschaft besitzt: Teilt p ein beliebiges Produkt ab zweier natürlicher Zahlen, dann teilt p mindestens eine der Zahlen a oder b .*

Der Beweis dieses Lemmas ist ohne weitere Vorüberlegungen erstaunlich mühsam. Wir werden ihn darum später führen, wenn wir etwas mehr Theorie zur Verfügung haben. Andererseits ist das Lemma ein wichtiges Hilfsmittel im Beweis der folgenden Aussage.

Satz (Fundamentalsatz der Arithmetik). *Sei $n \geq 2$ eine natürliche Zahl.*

(a) *Die Zahl n kann als Produkt von Primzahlen geschrieben werden, wobei Faktoren mehrfach auftreten dürfen.*

(b) *Die Zerlegung in (a) ist eindeutig bis auf Reihenfolge.*

In Kurzform, sagt man, dass jede natürliche Zahl eine eindeutige Primfaktorzerlegung hat. Produkte mit nur einem Faktor sind zugelassen (und auch nötig: alle Primzahlen haben diese Form). Die Aussage ist erstaunlicher, als es vielleicht auf den ersten Blick aussieht. Sind Sie sich zum Beispiel wirklich sicher, dass

$$37 \cdot 59 \cdot 61 \neq 7 \cdot 19009?$$

Beweis. Falls n eine Primzahl sind, so ist (a) schon trivialerweise erfüllt. Falls n keine Primzahl ist, so gibt es natürliche Zahlen n_1 und n_2 beide größer als 1 und kleiner als n , so dass

$$n = n_1 \cdot n_2.$$

Nun können wir uns wieder die Frage Stellen, ob n_1 eine Primzahl ist. Wenn ja, lassen wir n_1 stehen und wenn nein, können wir n_1 wieder als Produkt von zwei natürlichen Zahlen, welche größer als 1 und kleiner als n_1 sind schreiben. Genau gleich verfahren wir mit n_2 . Nach endlich vielen Schritten haben wir dann n in ein Produkt von Primzahlen zerlegt.

Wir kommen nun zum Beweis von (b). Seien also

$$n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_\ell$$

zwei Zerlegungen von n in Primfaktoren. Die Zahl p_1 teilt n , also das Produkt $q_1 q_2 \cdots q_\ell$. Mit Hilfe des Lemmas von Euklid und der Definition von Primzahlen überlegt man sich, das hieraus folgen muss, dass p_1 eine der Zahlen q_1, \dots, q_ℓ teilen muss. Da diese aber Primzahlen sind, muss also p_1 mit einer der Zahlen q_r übereinstimmen. Nach Umsortieren können wir annehmen, dass $p_1 = q_1$. Dann folgt aber auch

$$p_2 \cdots p_k = q_2 \cdots q_\ell.$$

Wendet man dasselbe Argument nun endlich oft an, so erhalten wir, dass $k = \ell$ gelten muss und die Zahlen p_2, \dots, p_k und q_2, \dots, q_k sich nur bis auf die Reihenfolge unterscheiden. Dies war aber gerade die Behauptung. \square

Wir betrachten noch einige Beispiele für die Primfaktorzerlegung natürlicher Zahlen:

$$12 = 2 \cdot 2 \cdot 3 = 2^2 \cdot 3$$

$$27 = 3 \cdot 3 \cdot 3 = 3^3$$

$$42 = 2 \cdot 3 \cdot 7$$

$$649 = 11 \cdot 59$$

$$7531 = 17 \cdot 443$$

$$19017 = 3 \cdot 3 \cdot 2113 = 3^2 \cdot 2113.$$

Sei a eine natürliche Zahl. Die Teiler von a , welche Primzahlen sind, heißen die *Primteiler* von a . Falls nun p_1, \dots, p_n die Primteiler von a sind, so gibt es wegen der eindeutigen Primfaktorzerlegung natürliche Zahlen r_1, \dots, r_n , so dass

$$a = p_1^{r_1} \cdot p_2^{r_2} \cdots p_n^{r_n}.$$

Satz 2 (Euklid). *Es gibt unendlich viele Primzahlen.*

Beweis. Wir nehmen an, es gäbe nur endlich viele Primzahlen. Wir nennen diese

$$p_1, p_2, \dots, p_n$$

und betrachten die natürliche Zahl

$$a = 1 + p_1 p_2 \cdots p_n.$$

Die Zahl a ist nicht durch p_j teilbar für $j = 1, 2, \dots, n$, sonst gäbe es eine natürliche Zahl b , so dass

$$1 + p_1 p_2 \cdots p_n = a = b p_j$$

also wäre (wir klammern p_j aus dem Produkt $p_1 \cdots p_n$ aus)

$$1 = p_j (b - p_1 \cdots p_{j-1} \cdot p_{j+1} \cdots p_n),$$

ein Widerspruch, da $p_j \geq 2$. Da man aber jede natürliche Zahl ≥ 2 als Produkt von Primzahlen schreiben kann, also insbesondere auch a , muss es noch Primzahlen geben, welche bei der Auflistung

$$p_1, p_2, \dots, p_n$$

nicht vorkommen. Das wiederum steht aber im Widerspruch zur Annahme, dass die Zahlen p_1, p_2, \dots, p_n gerade alle Primzahlen sind. Also war unsere Grundannahme falsch, dass es nur endlich viele Primzahlen gibt, und wir haben somit bewiesen, dass es unendlich viele Primzahlen gibt. \square

Bemerkung. Die Beweismethode, die wir hier angewendet haben, ist bekannt als „Beweis durch Widerspruch“ oder „reductio ad absurdum“. Sie ist ein häufiges Beweisverfahren in der Mathematik, weshalb wir sie auch später noch wiedersehen werden.

Definition. Es seien a und b natürliche Zahlen. Der *größte gemeinsame Teiler* von a und b die größte natürliche Zahl g , welche a teilt und b teilt. Wir schreiben dann

$$g = \text{ggT}(a, b).$$

Das *kleinste gemeinsame Vielfache* von a und b ist die kleinste natürliche Zahl k , welche a und b als Teiler hat. Wir schreiben dann

$$k = \text{kgV}(a, b).$$

Bemerkung. Man kann analog auch den ggT und das kgV für ganze Zahlen a_1, \dots, a_n , welche von 0 verschieden sind, definieren, d.h. $\text{ggT}(a_1, \dots, a_n)$ ist die größte natürliche Zahl, welche alle Zahlen a_1, \dots, a_n teilt und $\text{kgV}(a_1, \dots, a_n)$ bezeichnet die kleinste natürliche Zahl, welche alle Zahlen a_1, \dots, a_n als Teiler besitzt.

Wir betrachten wieder einige Beispiele:

$$\text{ggT}(2, 6) = 2$$

$$\text{kgV}(2, 6) = 6$$

$$\text{ggT}(15, 27) = 3$$

$$\text{kgV}(5, 7) = 35$$

$$\text{ggT}(24, 36) = 12$$

$$\text{kgV}(32, 48) = 96$$

Für Primzahlen p und q gilt $\text{ggT}(p, q) = \text{kgV}(p, q) = p$ falls $p = q$, und $\text{ggT}(p, q) = 1$ sowie $\text{kgV}(p, q) = p \cdot q$ falls $p \neq q$. Allgemeiner betrachten wir zwei natürliche Zahlen a und b mit Primfaktorzerlegungen

$$a = p_1^{r_1} \cdot p_2^{r_2} \cdots p_n^{r_n} \quad \text{und} \quad b = p_1^{s_1} \cdot p_2^{s_2} \cdots p_n^{s_n} .$$

Hier bezeichnen p_1, p_2, \dots, p_n gerade alle Primzahlen, die entweder a oder b teilen, so dass in der obigen Darstellung auch einige der Exponenten s_i und r_i null sein können. Dann gilt für den ggT und das kgV von a und b :

$$\begin{aligned} \text{ggT}(a, b) &= p_1^{\min\{r_1, s_1\}} \cdots p_n^{\min\{r_n, s_n\}} \quad \text{und} \\ \text{kgV}(a, b) &= p_1^{\max\{r_1, s_1\}} \cdots p_n^{\max\{r_n, s_n\}} . \end{aligned} \tag{1.1}$$

Dabei bezeichnet $\min\{r_i, s_i\}$ die kleinere der beiden Zahlen r_i und s_i und $\max\{r_i, s_i\}$ bezeichnet die größere der beiden Zahlen r_i und s_i . Auch dazu ein Beispiel:

Beispiel. Sei $a = 48$ und $b = 64$. Dann gilt für die Primfaktorzerlegung

$$a = 2^4 \cdot 3^1 \quad \text{und} \quad b = 2^6 = 2^6 \cdot 3^0 .$$

Damit gilt für den ggT und das kgV :

$$\text{ggT}(a, b) = 2^{\min\{4, 6\}} \cdot 3^{\min\{1, 0\}} = 2^4 \cdot 3^0 = 16$$

und

$$\text{kgV}(a, b) = 2^{\max\{4, 6\}} \cdot 3^{\max\{1, 0\}} = 2^6 \cdot 3^1 = 64 \cdot 3 = 192 .$$

Die Darstellung von (1.1) liefert auch noch folgenden Zusammenhang des ggT und des kgV von zwei natürlichen Zahlen a und b . Es gilt nämlich

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

denn

$$\begin{aligned} \text{ggT}(a, b) \text{kgV}(a, b) &= \left(p_1^{\min\{r_1, s_1\}} \cdots p_n^{\min\{r_n, s_n\}} \right) \cdot \left(p_1^{\max\{r_1, s_1\}} \cdots p_n^{\max\{r_n, s_n\}} \right) \\ &= p_1^{\min\{r_1, s_1\} + \max\{r_1, s_1\}} \cdots p_n^{\min\{r_n, s_n\} + \max\{r_n, s_n\}} \\ &= p_1^{r_1 + s_1} \cdots p_n^{r_n + s_n} \\ &= (p_1^{r_1} \cdots p_n^{r_n}) \cdot (p_1^{s_1} \cdots p_n^{s_n}) = a \cdot b . \end{aligned}$$

Dabei haben wir verwendet, dass für ganze Zahlen $r, s \geq 0$ stets

$$\min\{r, s\} + \max\{r, s\} = r + s$$

gilt, worüber Sie dann in den Übungen noch einmal nachdenken werden.

Häufig ist es schwierig, für Zahlen a und b eine Primfaktorzerlegung hinzuschreiben, und dann können wir auch nicht die Formeln (1.1) für die Berechnung des ggT und des kgV verwenden. Es gibt aber eine viel effizientere Methode den ggT zu berechnen, welche wieder auf die alten Griechen zurück geht. Bevor wir aber diese Methode vorstellen, diskutieren wir zunächst die Division mit Rest.

Satz 3 (Division mit Rest). Seien a, b natürliche Zahlen. Dann existieren eindeutige Zahlen $q, r \in \mathbb{N}_0$ mit

$$a = qb + r \text{ und } 0 \leq r < b.$$

Beweis. Wir beweisen zuerst die Existenz von solchen Zahlen q und r . Falls $a < b$, so setzen wir $q = 0$ und $r = a$. Falls $b \geq a$, dann subtrahieren wir so oft b von a , bis das Ergebnis das erste Mal kleiner als b ist. Die Anzahl der Subtraktionen bezeichnen wir mit q , und den erhaltenen Rest $a - qb$ mit r . In beiden Fällen erfüllen unsere gewählten q und r die Aussage des Satzes.

Seien nun

$$a = qb + r = q'b + r'$$

zwei Zerlegungen von a mit $0 \leq r < b$ und $0 \leq r' < b$. Wir wollen zeigen, dass dann $q = q'$ und $r = r'$ gelten müssen. Dazu schreiben wir die Gleichung

$$qb + r = q'b + r'$$

um als

$$r - r' = q'b - qb = (q' - q)b.$$

Da die beiden nichtnegativen Zahlen r und r' strikt kleiner als b sind, erfüllt ihre Differenz die Ungleichungen

$$-b < -r' < r - r' < r < b.$$

Sie soll aber gleich der durch b teilbaren Zahl $(q' - q)b$ sein. Dies geht nur, wenn $r - r' = 0$, und hieraus folgt dann auch (wegen $b \neq 0$) dass $q' - q = 0$, d.h. $q = q'$. Also waren unsere Zerlegungen gleich. Da dies für je zwei beliebig vorgegebene Zerlegungen gilt, beweist es die behauptete Eindeutigkeit. \square

Nun kommen wir zur versprochenen effizienteren Methode, um den ggT zu berechnen. Wir beschreiben hier das Verfahren nur anhand eines Beispiels. Vorweg machen wir noch folgende Bemerkung: für natürliche Zahlen c, d und q mit $q \cdot c < d$ gilt

$$\text{ggT}(d - q \cdot c, c) = \text{ggT}(d, c). \quad (1.2)$$

Dies ist klar: Jede Zahl, welche d und c teilt, teilt auch $d - q \cdot c$, und umgekehrt teilt jede Zahl, welche die Differenz und c teilt, auch d .

Euklidischer Algorithmus

Wir wollen $\text{ggT}(a, b)$ für $a = 456$ und $b = 356$ berechnen. Dazu seien q_1 und r_1 die ganzen Zahlen, welche

$$a = q_1b + r_1 \quad \text{und} \quad 0 \leq r_1 < b$$

erfüllen, d.h. in unserem Beispiel $q_1 = 1$ und $r_1 = 100$. Es gilt also

$$\text{ggT}(a, b) = \text{ggT}(1 \cdot 356 + 100, 356) = \text{ggT}(100, 356).$$

Also müssen wir nun $\text{ggT}(100, 356)$ berechnen. Es gilt

$$356 = 3 \cdot 100 + 56,$$

also ist $\text{ggT}(356, 100) = \text{ggT}(56, 100)$. Weiter ist

$$100 = 1 \cdot 56 + 44,$$

also ist $\text{ggT}(100, 56) = \text{ggT}(44, 56)$. Analoges Weiterrechnen ergibt nun

$$\text{ggT}(56, 44) = \text{ggT}(44, 12) = \text{ggT}(12, 8) = \text{ggT}(8, 4) = 4.$$

Der Algorithmus besagt also, dass man mit a und b beginnend in jedem Schritt die größere der beiden aktuellen Zahlen durch den Rest der Division durch die kleinere ersetzt. Das Verfahren bricht in dem Moment ab, wenn die Division keinen Rest hinterlässt. Da wegen der Gleichung (1.2) der ggT jeweils erhalten bleibt, ist die am Schluss gewonnene Zahl tatsächlich der gesuchte ggT von a und b . In unserem Beispiel ist also $\text{ggT}(456, 356) = 4$.

Weil das Produkt zweier natürlicher Zahlen gerade das Produkt von ggT und kgV ist, lässt sich dann auch das kgV effizient berechnen. In unserem Beispiel erhalten wir

$$\text{kgV}(456, 356) = (456 \cdot 356) : 4 = 40584.$$

Soweit zunächst zum Rechnen mit natürlichen Zahlen. Um den “Mangel” zu beheben, dass Subtraktion und Division nicht immer ausführbar sind, erweitert man die betrachtete Zahlenmenge. Die beiden Erweiterungen sind dabei unabhängig und in beiden Reihenfolgen ausführbar.

Ganze Zahlen

Um die uneingeschränkte Ausführbarkeit der Subtraktion zu ermöglichen, führen wir die ganzen Zahlen ein.

Definition. Die Menge der *ganzen Zahlen* ist definiert als

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, 4, -4, \dots\}$$

Die natürlichen Zahlen bilden offenbar eine (echte) Teilmenge aller ganzen Zahlen, $\mathbb{N} \subseteq \mathbb{Z}$. Die Addition und die Multiplikation ist für alle ganzen Zahlen definiert, und erfüllt nach wie vor die wesentlichen Eigenschaften (Kommutativität, Assoziativität und Distributivgesetz). Die Relation $<$ lässt sich ebenfalls auf die ganzen Zahlen fortsetzen.

Neu ist die Subtraktion für alle geordneten Paare ganzer Zahlen definiert. Wir wollen dies ausnutzen und noch einmal einen Blick auf den Euklidischen Algorithmus werfen.

Satz 4. Sind a und b natürliche Zahlen, und ist $g = \text{ggT}(a, b)$, so existieren ganze Zahlen x und y mit

$$g = xa + yb.$$

Beweis. Wir können annehmen, dass $b \leq a$. Nach Satz 3 gibt es nun $q_1 \in \mathbb{N}$ und $r_1 \in \mathbb{N}_0$ mit $0 \leq r_1 < b$, so dass

$$a = q_1 b + r_1.$$

Dies können wir umschreiben als

$$r_1 = a - q_1 b.$$

Wir können also den Rest r_1 in der Form $r_1 = x_1 a + y_1 b$ schreiben. Division mit Rest von b durch $r_1 < b$ liefert uns nun eine neue Gleichung

$$b = q_2 r_1 + r_2$$

mit $q_2 \in \mathbb{N}$ und $0 \leq r_2 < r_1$. Es folgt also

$$r_2 = b - q_2 r_1 = b - q_2(a - q_1 b) = (1 + q_1 q_2)b - q_2 a.$$

Wieder können wir also den Rest r_2 in der Form $r_2 = x_2 a + y_2 b$ schreiben, wobei $x_2 = -q_2$ und $y_2 = 1 + q_1 q_2$ ganze Zahlen sind.

Im nächsten Schritt dividieren wir r_1 durch r_2 , und erhalten wie zuvor eine neue Darstellung des Restes $0 \leq r_3 < r_2$ als

$$r_3 = r_1 - q_3 r_2 = (x_1 a + y_1 b) - q_3(x_2 a + y_2 b).$$

Fassen wir neu zusammen, so folgt hieraus

$$r_3 = x_3 a + y_3 b$$

mit ganzen Zahlen $x_3 = x_1 - q_3 x_2$ und $y_3 = y_1 - q_3 y_2$. Diese Formeln gelten ganz allgemein, d.h. für $k \geq 3$ erhalten wir

$$r_k = x_k a + y_k b$$

mit

$$x_k = x_{k-2} - q_k x_{k-1} \quad \text{und} \quad y_k = y_{k-2} - q_k y_{k-1}.$$

Da die Reste r_k , welche bei unseren Divisionen entstehen, stets nichtnegativ sind aber immer kleiner werden, wird nach endlich vielen Schritten (sagen wir, nach n Schritten) der Fall $r_n = 0$ erreicht. Der größte gemeinsame Teiler von a und b ist dann aber gerade r_{n-1} , und unser Schema liefert eine Darstellung

$$r_{n-1} = x_{n-1} a + y_{n-1} b.$$

□

Beispiel. Wir betrachten noch einmal unser Zahlenbeispiel mit $a = 456$ und $b = 356$. Die bei den einzelnen Divisionen erhaltenen Reste waren

$$\begin{array}{llll} r_1 = 100 & = 1 \cdot 456 - 1 \cdot 356 & & = 1 \cdot a - 1 \cdot b \\ r_2 = 56 & = 1 \cdot 356 - 3 \cdot 100 & = 1 \cdot b - 3(1 \cdot a - 1 \cdot b) & = -3 \cdot a + 4 \cdot b \\ r_3 = 44 & = 1 \cdot 100 - 1 \cdot 56 & = (1 \cdot a - 1 \cdot b) - (-3 \cdot a + 4 \cdot b) & = 4 \cdot a - 5 \cdot b \\ r_4 = 12 & = 1 \cdot 56 - 1 \cdot 44 & = (-3 \cdot a + 4 \cdot b) - (4 \cdot a - 5 \cdot b) & = -7 \cdot a + 9 \cdot b \\ r_5 = 8 & = 1 \cdot 44 - 3 \cdot 12 & = (4 \cdot a - 5 \cdot b) - 3(-7 \cdot a + 9 \cdot b) & = 25 \cdot a - 32 \cdot b \\ r_6 = 4 & = 1 \cdot 12 - 1 \cdot 8 & = (-7 \cdot a + 9 \cdot b) - (25 \cdot a - 32 \cdot b) & = -32 \cdot a + 41 \cdot b. \end{array}$$

Dies ist die im Satz behauptete Darstellung des größten gemeinsamen Teilers.

Als kleine Anwendung kommen wir noch einmal auf die Primzahlzerlegung zurück. Mit Hilfe von Satz 4 ist der Beweis von Lemma 1 nämlich relativ problemlos möglich. Hier zur Erinnerung noch einmal die Aussage:

Lemma. *Eine natürliche Zahl $p > 1$ ist genau dann eine Primzahl, wenn sie folgende Eigenschaft besitzt: Teilt p ein beliebiges Produkt ab zweier natürlicher Zahlen, dann teilt p mindestens eine der Zahlen a oder b .*

Beweis. Wir nehmen zunächst an, dass p eine Primzahl ist, welche das Produkt ab teilt, aber den Faktor a nicht teilt, und werden beweisen, dass dann p den Faktor b teilt.

Da p eine Primzahl ist, also nur die Teiler 1 und p besitzt, und andererseits p die Zahl a nach unserer Annahme nicht teilt, muss

$$\text{ggT}(p, a) = 1$$

gelten. Mit Hilfe von Satz 4 erhalten wir nun ganze Zahlen x und y mit

$$1 = xp + ya.$$

Multiplizieren wir diese Gleichung mit b , so erhalten wir

$$b = xpb + yab.$$

Nach Voraussetzung teilt p das Produkt ab , es gibt also eine natürliche Zahl z mit $ab = pz$. Setzen wir dies in unsere Gleichung ein, erhalten wir

$$b = xpb + yab = xpb + ypz = p(xb + yz),$$

und somit haben wir b als Produkt von p mit einer ganzen Zahl $k = (xb + yz)$ dargestellt. Da sowohl b also auch p natürliche Zahlen sind, muss übrigens auch k eine natürliche Zahl sein. Die Primzahl p teilt also b , was zu beweisen war.

Ist andererseits p keine Primzahl, so können wir p als Produkt $p = ab$ mit Faktoren $1 < a, b < p$ schreiben. Dann ist aber klar, dass p weder a noch b teilen kann (warum?). \square

Bemerkungen. Wir wollen an dieser Stelle noch zwei Bemerkungen machen. Zunächst einmal ist bei einem wie hier auf später verschobenem Beweis stets große Vorsicht geboten. Haben wir wirklich nichts verwendet, was auf der zu beweisenden Aussage basiert, und in der Zwischenzeit entwickelt wurde? Auch im vorliegenden Fall sollten Sie also die gesamte Argumentationskette noch einmal genau prüfen!

Die andere Bemerkung ist struktureller Natur. Lemma 1 ist eine Aussage über natürliche Zahlen, und es gibt auch Beweise, welche ausschließlich in den natürlichen Zahlen argumentieren und nicht auf ganze Zahlen zurückgreifen (für eine mögliche Argumentationskette betrachte man die Übungsaufgabe A 8.12 in Kapitel 8 von Griesers Buch [?]). Es werden Ihnen in der Mathematik aber immer wieder Aussagen begegnen, deren Beweis kürzer und manchmal auch verständlicher wird, wenn man über die in der Aussage selbst vorkommenden Objekte hinaus allgemeinere Objekte oder Theorien benutzt.

Rationale Zahlen

Division von ganzen Zahlen ist manchmal möglich, aber nicht immer. Zum Beispiel ist $12 : 3 = 4 \in \mathbb{Z}$ und $-15 : 5 = -3 \in \mathbb{Z}$, aber $5 : 4$ entspricht keiner ganzen Zahl. Will man Division für alle Zahlenpaare durchführen, so führt dies zu einer anderen Erweiterung des Zahlenbereiches.

Die *rationalen Zahlen* sind diejenigen Zahlen, welche sich als Quotient $\frac{p}{q}$ aus zwei ganzen Zahlen p und q mit $q \neq 0$ ausdrücken lassen. Die Menge der rationalen Zahlen bezeichnen wir mit \mathbb{Q} .

Bemerkung. Jeder vernünftige Mensch wird an dieser Stelle protestieren: woher weiß ich denn, was eine *Zahl* ist? Die korrekte Definition der rationalen Zahlen ist subtil, weil eben nicht jeder Bruch $\frac{p}{q}$ eine eigene Zahl darstellt: die Darstellung einer ganzen Zahl als Bruch nicht eindeutig, denn es gilt zum Beispiel

$$\frac{-2}{3} = \frac{-240004}{360006} = \frac{6}{-9},$$

und man schreibt für diese Zahl meist einfach $-\frac{2}{3}$. Allgemein beschreiben zwei Brüche $\frac{a}{b}$ und $\frac{c}{d}$ dieselbe rationale Zahl, falls $ad = bc$. Eine präzise Definition werden wir später geben, wenn wir Äquivalenzrelationen diskutiert haben.

Die rationalen Zahlen sind eine echte Erweiterung der ganzen Zahlen, d.h. es gilt $\mathbb{Z} \subset \mathbb{Q}$, weil sich jedes $n \in \mathbb{Z}$ zum Beispiel als $n = \frac{n}{1}$ schreiben lässt. Rationale Zahlen können addiert, multipliziert, subtrahiert und dividiert werden. Einzige Ausnahme: Division durch 0 ist nicht erlaubt. Es gelten die Regeln

$$\begin{array}{ll} \frac{a}{b} + \frac{c}{d} = \frac{ad + cb}{bd} & \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd} \\ \frac{a}{b} - \frac{c}{d} = \frac{ad - cb}{bd} & \frac{a}{b} : \frac{c}{d} = \frac{ad}{bc} \end{array}$$

Addition und Multiplikation sind wie schon zuvor kommutativ und assoziativ, Subtraktion und Division nicht; es gilt das Distributivgesetz. Die rationalen Zahlen lassen sich weiterhin bezüglich $<$ anordnen: $\frac{2}{3} < \frac{4}{5}$, $\frac{-3}{2} < \frac{1}{7}$, etc. Es gibt aber, wie wir gleich sehen werden, immer noch Gleichungen, die keine Lösung in den rationalen Zahlen haben.

Satz 5. *Es gibt keine rationale Zahl x für die $x^2 = 2$ gilt.*

Beweis. Angenommen es gibt eine solche Zahl, dann hat sie eine Darstellung $x = \frac{p}{q}$. Wir können p und q außerdem so wählen, dass ihr kleinster gemeinsamer Teiler 1 ist (gekürzter Bruch). Nun wissen wir $x^2 = (\frac{p}{q})^2 = 2$, also

$$p^2 = 2 \cdot q^2.$$

Somit muss p^2 und damit auch p selbst eine gerade Zahl sein. Es gibt also eine ganze Zahl r sodass $p = 2 \cdot r$ und es folgt $2q^2 = p^2 = 4r^2$ beziehungsweise

$$2r^2 = q^2.$$

Das bedeutet, q ist auch gerade.

Somit haben wir gezeigt: falls es eine rationale Lösung der Gleichung $x^2 = 2$ gibt, dann gibt es ein Paar p, q von Zahlen ohne gemeinsamen Teiler, die aber beide gerade sind. Da das unmöglich ist, kann es auch keine solche Lösung geben. \square

Reelle Zahlen

Um eine Lösung von Gleichungen der Form $x^2 = 2$ (und vielen anderen Gleichungen) zu ermöglichen, erweitern wir \mathbb{Q} zur Menge der reellen Zahlen \mathbb{R} . Intuitiv ist diese Erweiterung schwerer zu fassen als die vorhergehenden. Anschaulich kann man sich die reellen Zahlen als die Menge aller (unendlichen) Dezimalzahlen (mit gewissen Identifikationen, wie z.B. $0.999\cdots = 1$) beziehungsweise als Menge aller Punkte auf dem Zahlenstrahl vorstellen, welcher "keine Löcher hat". Eine genauere Untersuchung der grundlegenden Eigenschaften der reellen Zahlen ist ein zentrales Thema der ersten Wochen der Analysis-Vorlesung.

Beispiele. • Die positive Lösung der Gleichung $x^2 = 2$ ist eine reelle Zahl. Sie wird mit $\sqrt{2} \approx 1.414213562$ bezeichnet. Allgemein sind alle Wurzeln aus positiven Zahlen (d.h. positive Lösungen \sqrt{y} von Gleichungen der Form $x^2 = y$) in \mathbb{R} enthalten.

- Die Kreiszahl $\pi \approx 3.1415926$ ist ebenfalls eine reelle Zahl.

Wir betrachten einige Eigenschaften reeller Zahlen:

- $\mathbb{Q} \subset \mathbb{R}$, da jede rationale Zahl auch als Dezimalzahl dargestellt werden kann.
- reelle Zahlen lassen sich addieren, multiplizieren, subtrahieren und dividieren. Außerdem können Wurzeln aus allen nicht-negativen reellen Zahlen gezogen werden.
- reelle Zahlen lassen sich durch $<$ anordnen.
- Da x^2 immer größer oder gleich 0 ist, lässt sich für die Gleichung $x^2 = -1$ in den reellen Zahlen keine Lösung finden.

Man kann die reellen Zahlen nochmals erweitern, zu den sogenannten komplexen Zahlen. In den komplexen Zahlen hat dann die Gleichung $x^2 = -1$ eine Lösung.

1.2 Weitere Notationen und nützliche Formeln

Intervalle

Intervalle sind häufig auftretende Teilmengen von \mathbb{R} . Sie bestehen gerade aus allen reellen Zahlen die zwischen zwei gegebenen reellen Zahlen a und b liegen.

Definition. Seien $a, b \in \mathbb{R}$ mit $a < b$, dann definieren wir

- $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$ (offenes Intervall)
- $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$ (abgeschlossenes Intervall)

- $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$ und $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$ (halboffene Intervalle)

Hier benutzen wir eine neue Schreibweise für Mengen: wir betrachten eine Teilmenge einer bereits bekannten Menge (in diesem Fall \mathbb{R}), welche alle Objekte mit einer gewissen Eigenschaft enthält. diese Eigenschaft wird nach einem Trennstrich $|$ in die Mengenklammern geschrieben. Man liest das Ganze dann zum Beispiel als

Das offene Intervall (a, b) besteht aus allen $x \in \mathbb{R}$, so dass $a < x < b$.

Der einzige Unterschied in den Definitionen von offenen, abgeschlossenen und halboffenen Intervallen besteht darin, ob die Randpunkte a und b in der Menge enthalten sind oder nicht.

Außerdem werden häufig auch Mengen der Form $\{x \in \mathbb{R} \mid x > a\}$, $\{x \in \mathbb{R} \mid x \geq a\}$, $\{x \in \mathbb{R} \mid x < b\}$ oder $\{x \in \mathbb{R} \mid x \leq b\}$ betrachtet. Auch diese Mengen kann man als Intervalle schreiben, indem man ∞ beziehungsweise $-\infty$ als Intervallgrenze zulässt:

- $(a, \infty) = \{x \in \mathbb{R} \mid x > a\}$,
- $[a, \infty) = \{x \in \mathbb{R} \mid x \geq a\}$,
- $(-\infty, b) = \{x \in \mathbb{R} \mid x < b\}$,
- $(-\infty, b] = \{x \in \mathbb{R} \mid x \leq b\}$.

Man beachte, dass die Intervalle bei ∞ und $-\infty$ immer offen sind. Diese Notation macht insofern Sinn, da $-\infty$ und ∞ keine reellen Zahlen sind und somit nicht im Intervall enthalten sein können.

Summen und Produkte

Summen- und Produktzeichen (\sum und \prod) bieten eine kompakte Schreibweise für eben Summen und Produkte. Sie können wie folgt definiert werden.

Definition. Angenommen wir haben reelle Zahlen $a_1, a_2, a_3, \dots, a_n$ und $m < n$ ist eine natürliche Zahl. Dann definieren wir

$$\sum_{i=m}^n a_i = a_m + a_{m+1} + a_{m+2} + \dots + a_n,$$

$$\prod_{i=m}^n a_i = a_m \cdot a_{m+1} \cdot a_{m+2} \cdot \dots \cdot a_n.$$

$\sum_{i=m}^n a_i$ wird gelesen als „die Summe der Zahlen a_i von m bis n “, und $\prod_{i=m}^n a_i$ wird gelesen als „das Produkt der Zahlen a_i von m bis n “.

Falls $m > n$, dann setzt man per Konvention $\sum_{i=m}^n a_i = 0$ und $\prod_{i=m}^n a_i = 1$.

Beispiele. • Angenommen $a_1 = 1, a_2 = 2, a_3 = 3, \dots$, d.h. $a_i = i$. Dann ist

$$\sum_{i=3}^6 a_i = a_3 + a_4 + a_5 + a_6 = 3 + 4 + 5 + 6 = 18$$

$$\prod_{i=1}^4 a_i = a_1 \cdot a_2 \cdot a_3 \cdot a_4 = 1 \cdot 2 \cdot 3 \cdot 4 = 24 \quad \text{und}$$

$$\sum_{i=7}^4 a_i = 0$$

• Falls $a_i = 5$ für alle i , dann ist

$$\sum_{i=1}^3 a_i = a_1 + a_2 + a_3 = 5 + 5 + 5 = 3 \cdot 5 = 15$$

beziehungsweise allgemein

$$\sum_{i=1}^n 5 = n \cdot 5.$$

Klarerweise spielte der genaue Wert der Zahl 5 nicht wirklich eine Rolle und man bekommt auch analoge Ergebnisse wenn alle a_i den gleichen Wert haben, unabhängig davon, was dieser Wert ist.

Man kann nun in Analogie zum letzten Beispiel Produkte benutzen, um eine Potenz (zugegebenermaßen etwas kompliziert) zu beschreiben. Für eine reelle Zahl x und eine natürliche Zahl n gilt nämlich

$$x^n = \prod_{i=1}^n x.$$

Satz 6. *Folgende Rechenregeln gelten für Summen und Produkte:*

1. *Assoziativgesetz:*

$$\sum_{i=1}^{m+n} a_i = \sum_{i=1}^m a_i + \sum_{i=m+1}^{m+n} a_i, \quad \prod_{i=1}^{m+n} a_i = \left(\prod_{i=1}^m a_i \right) \left(\prod_{i=m+1}^{m+n} a_i \right)$$

2. *Kommutativgesetz:*

$$\sum_{i=1}^n (a_i + b_i) = \sum_{i=1}^n a_i + \sum_{i=1}^n b_i, \quad \prod_{i=1}^n (a_i b_i) = \left(\prod_{i=1}^n a_i \right) \left(\prod_{i=1}^n b_i \right)$$

3. *Distributivgesetz:*

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

Der Beweis dieser Rechenregeln kann zum Beispiel über vollständige Induktion geführt werden. Wir werden diese Beweismethode später genauer behandeln.

Der Binomische Lehrsatz

Als Beispiel für die Anwendung der grundlegenden Rechenregeln wollen wir die sogenannten binomischen Formeln herleiten, die häufig nützlich sein können (beispielsweise werden wir sie beim Lösen quadratischer Gleichungen verwenden).

Satz 7 (Binomische Formeln). *Seien $a, b \in \mathbb{R}$, dann gilt*

1. $(a + b)^2 = a^2 + 2ab + b^2$,
2. $(a - b)^2 = a^2 - 2ab + b^2$,
3. $(a + b)(a - b) = a^2 - b^2$.

Beweis. Wir beweisen nur die erste der Formeln, die anderen beiden können auf komplett gleiche Weise hergeleitet werden.

$$\begin{aligned}(a + b)^2 &= (a + b)(a + b) \\ &= (a + b)a + (a + b)b && \text{(Distributivgesetz)} \\ &= a(a + b) + b(a + b) && \text{(Kommutativgesetz der Multiplikation)} \\ &= a^2 + ab + ba + b^2 && \text{(Distributivgesetz)} \\ &= a^2 + ab + ab + b^2 && \text{(Kommutativgesetz der Multiplikation)} \\ &= a^2 + 2ab + b^2.\end{aligned}$$

□

Die binomische Formel $(a + b)^2 = a^2 + 2ab + b^2$ kann auf beliebig viele Faktoren verallgemeinert werden, wobei die erhaltene Formel am einfachsten mit unserer neuen Summen- und Produktschreibweise formuliert werden kann.

Satz 8 (Binomischer Lehrsatz). *Für $a, b \in \mathbb{R}$ und $n \in \mathbb{N}$ gilt*

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

□

Hier bezeichnet $\binom{n}{k}$ den Binomialkoeffizienten “ n über k ” und ist definiert als

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

wobei $n!$ die Fakultät von n bezeichnet und als

$$n! = \prod_{i=1}^n i = 1 \cdot 2 \cdot \dots \cdot n$$

definiert ist. Aus der Schule ist (hoffentlich) bekannt, dass $\binom{n}{k}$ gerade die Anzahl der k -elementigen Teilmengen einer n -elementigen Menge bezeichnet. Zum Beispiel bezeichnet $\binom{50}{4}$ die Anzahl der Möglichkeiten, aus 50 Äpfeln an einem Marktstand 4 auszuwählen.

2 Gleichungen

Eine mathematische Gleichung besteht aus zwei gleichartigen Ausdrücken (auch *Terme* genannt), nämlich der rechten und linken Seite, die durch das Gleichheitszeichen in Verbindung gebracht werden. Üblicherweise enthalten Gleichungen eine oder mehrere Variablen (Unbekannte). Im Folgenden werden wir vor allem Gleichungen in einer oder zwei Variablen betrachten, und diese meist mit x und y bezeichnen.

Um ganze Beispielfamilien behandeln zu können, werden in unseren Gleichungen oft weitere Symbole (meist a, b, c usw.) auftreten, welche die Rolle von "Parametern" haben. Spezifiziert man konkrete Werte für diese, so erhält man ein konkretes Beispiel für eine Gleichung.

Beispiel. Die allgemeine quadratische Gleichung in einer (reellen) Variablen x hat die Form

$$ax^2 + bx + c = 0.$$

Konkrete quadratische Gleichungen erhält man, indem man konkrete Zahlenwerte für a, b und c einsetzt, also mit $a = \sqrt[3]{17}$, $b = \frac{67}{13}$ und $c = -1$ zum Beispiel die Gleichung

$$\sqrt[3]{17}x^2 + \frac{67}{13}x - 1 = 0.$$

Bemerkung. In diesem Kapitel geht es uns um Gleichungen zwischen Termen, in denen die Variablen Platzhalter für Zahlen sind. Es gibt aber auch Gleichungen zwischen Ausdrücken anderer Form: Gleichungen zwischen Matrizen, Gleichungen zwischen Mengen, Gleichungen zwischen Gruppen, usw.

Warnung: *Es ist bei der Benutzung des Gleichheitszeichen **unerlässlich**, dass die Objekte auf beiden Seiten vom selben Typ sind.*

Eine Gleichung der Form $1 = \{1, 2, 3\}$ zwischen einer Zahl und einer Menge ist zum Beispiel einfach nur Unsinn.

Wir werden hier als möglichen Wertevorrat für unsere Variablen die reellen Zahlen annehmen. Allerdings kann es sein, dass die auftretenden Ausdrücke für gewisse Werte der Variablen nicht definiert sind. Der *Definitionsbereich eines Ausdrucks* ist (im Fall einer Variablen) die Teilmenge aller reellen Zahlen, für die der Ausdruck sinnvoll ist. So ist zum Beispiel der Definitionsbereich des Ausdrucks $\sqrt{1-x}$ die Menge aller reellen Zahlen x , so dass $1-x \geq 0$ ist, also gerade alle reellen Zahlen $x \leq 1$.

Der *Definitionsbereich der Gleichung* ist die Teilmenge aller reellen Zahlen, für die beide Seiten sinnvoll sind. So ist zum Beispiel $x = 0$ nicht im Definitionsbereich der Gleichung $\frac{1}{x} = \sin x$ enthalten, d.h. der Definitionsbereich der Gleichung wäre die Teilmenge

$$D = \{x \in \mathbb{R} \mid x \neq 0\} \subset \mathbb{R}.$$

Definition. Eine *Lösung* einer Gleichung ist eine Belegung der Variablen mit Werten aus dem Definitionsbereich der Gleichung, sodass die Gleichung zu einer wahren Aussage wird.

Beispiele. • Die Gleichung $x^2 = 2 - x$ hat zwei Lösungen, $x = 1$ und $x = -2$.

- Die Gleichung $x^2 = 0$ besitzt genau eine Lösung, nämlich $x = 0$.
- Die Gleichung $\sin x = 0$ besitzt unendlich viele Lösungen, jedes Element der Menge $\{k\pi \mid k \in \mathbb{Z}\}$ ist nämlich eine Lösung.
- Der Begriff der Lösung beschränkt sich nicht auf Gleichungen in einer Variablen: die Gleichung $y^2 = x + 1$ hat beispielsweise die Lösung $(x, y) = (0, 1)$, aber auch unendlich viele andere Lösungen.
- Nicht alle Gleichungen sind lösbar, so hat zum Beispiel die Gleichung $\sin x = 7$ keine (reelle!) Lösung, da der Sinus nur Werte im Intervall $[-1, 1]$ annimmt.

Definition. Als *Lösungsmenge* einer Gleichung bezeichnen wir die Menge aller Lösungen der Gleichung. Eine Gleichung zu lösen bedeutet, ihre Lösungsmenge zu bestimmen.

Bemerkung. Je nach Anwendung kann es manchmal auch genügen, nur eine Lösung einer Gleichung zu bestimmen.

In den obigen Beispielen haben wir in allen Fällen außer der Gleichung $y^2 = x + 1$ nicht nur eine Lösung, sondern die gesamte Lösungsmenge angegeben. Wie wir gesehen haben, ist es durchaus möglich, dass die Lösungsmenge leer ist, aber auch, dass sie unendlich viele Elemente enthält.

Eine wichtige Rolle beim Lösen von Gleichungen spielt der Begriff der Äquivalenzumformung.

Definition. Eine *Äquivalenzumformung* einer Gleichung ist eine Umformung, die die Lösungsmenge unverändert lässt.

Einige Beispiele für Äquivalenzumformungen sind:

- Vertauschen von linker und rechter Seite,
- Addition oder Subtraktion des gleichen Ausdrucks auf beiden Seiten,
- Multiplikation beider Seiten mit demselben Faktor $c \neq 0$,
- Division beider Seiten durch dasselbe $c \neq 0$.

Natürlich ist Division durch c nichts anderes als Multiplikation mit $c' = \frac{1}{c}$, so dass die letzten beiden Punkte nicht wirklich verschieden sind.

Eine Warnung ist hier aber noch angebracht: Bei der Multiplikation und Division mit Ausdrücken, die eine Variable enthalten muss man sicherstellen, dass diese Ausdrücke nicht Null sind bzw. den Fall, dass sie gleich Null sind separat betrachten. Zum Beispiel hat die Gleichung $x - 1 = 0$ die Lösung $x = 1$. Dividiert man aber beide Seiten durch $(x - 1)$ erhält man die unlösbare Gleichung $1 = 0$.

2.1 Lineare Gleichungen

Definition. Eine Gleichung für eine Variable x heißt *linear*, wenn sie sich durch Äquivalenzumformungen in die Form $ax + b = 0$ bringen lässt, wobei a und b von x unabhängige Parameter mit $a \neq 0$ sind.

Beispiele für lineare Gleichungen in x sind

$$4x + 7 = 3x - 3 \quad \text{oder} \quad (c^7 + 3c^3)x - \sqrt{c} = 7,$$

wobei wir bei der zweiten Gleichung voraussetzen, dass der Parameter c nur positive Werte annimmt.

Mithilfe der folgenden Äquivalenzumformungen lässt sich eine lineare Gleichung immer lösen:

$$\begin{array}{ll} ax + b = 0 & | -b \text{ (subtrahiere } b \text{ auf beiden Seiten)} \\ ax = -b & | : a \text{ (dividiere beide Seiten durch } a) \\ x = -\frac{b}{a} & \end{array}$$

Somit ist $x = -\frac{b}{a}$ die eindeutige Lösung der linearen Gleichung $ax + b = 0$.

Beispiel. Wir bestimmen die Lösungsmenge der Gleichung $2x + 4 = 3x + 1$:

$$\begin{array}{ll} 2x + 4 = 3x + 1 & | -(3x + 1) \text{ (Auf mögliche Vorzeichenfehler achten!)} \\ -x + 3 = 0 & | -3 \\ -x = -3 & | : (-1) \\ x = 3 & \end{array}$$

2.2 Quadratische Gleichungen

Definition. Wie schon erwähnt heißt eine Gleichung für eine Variable x *quadratisch*, wenn sie sich durch Äquivalenzumformungen in die Form $ax^2 + bx + c = 0$ mit $a \neq 0$ bringen lässt. Eine quadratische Gleichung der Form $x^2 + px + q = 0$ (also mit $a = 1$) heißt quadratische Gleichung *in Normalform*.

Durch die Äquivalenzumformung „Division durch a “ kann jede quadratische Gleichung in Normalform gebracht werden. Außerdem können (ähnlich wie im linearen Fall) Gleichungen der Form $a_1x^2 + b_1x + c_1 = a_2x^2 + b_2x + c_2$ auf die gewünschte Form gebracht werden, indem man die rechte Seite subtrahiert.

p-q-Formel und Diskriminante

Die folgende Formel zur Lösung von quadratischen Gleichungen in Normalform sollte schon aus der Schule bekannt sein.

Satz 9. Sei $x^2 + px + q = 0$ eine quadratische Gleichung in Normalform, und sei $D = \frac{p^2}{4} - q$. Dann gilt:

(a) Ist $D < 0$, so besitzt die Gleichung keine reelle Lösung.

(b) Ist $D \geq 0$, so sind alle Lösungen der Gleichung gegeben durch

$$x_1 = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q} \quad \text{und} \quad x_2 = -\frac{p}{2} - \sqrt{\frac{p^2}{4} - q}$$

Ist insbesondere $D = 0$, so stimmen die Ausdrücke für x_1 und x_2 überein, und $-\frac{p}{2}$ ist die eindeutige Lösung.

Der Ausdruck $D = \frac{p^2}{4} - q$ unter der Wurzel wird als *Diskriminante* bezeichnet.

Quadratisches Ergänzen

Wir wollen die p - q -Formel nicht nur verwenden, sondern auch verstehen, warum sie funktioniert. Wir werden dies anhand eines Beispiels durchführen, alle Schritte funktionieren aber auch für allgemeines p und q (siehe Übung).

Angenommen wir haben die folgende quadratische Gleichung in Normalform gegeben

$$x^2 + 3x + 2 = 0.$$

Die p - q -Formel gibt als Lösungen

$$x_1 = -\frac{3}{2} + \sqrt{\frac{3^2}{4} - 2} = -1 \quad \text{und} \quad x_2 = -\frac{3}{2} - \sqrt{\frac{3^2}{4} - 2} = -2.$$

Um diese Lösung ohne die Formel zu erhalten, erinnern wir uns zuerst an die binomische Formel $(a + b)^2 = a^2 + 2ab + b^2$. Wir würden nun gerne den Teil $x^2 + 3x$ der linken Seite unserer Gleichung so ergänzen, dass er die Form $a^2 + 2ab + b^2$ besitzt:

$$x^2 + 3x = x^2 + 2 \cdot \frac{3}{2}x,$$

also sollte $b = \frac{3}{2}$ bzw. $b^2 = \frac{3^2}{4}$ sein.

Wir schreiben nun die linke Seite der Gleichung folgendermaßen um:

$$x^2 + 3x + 2 = \left(x^2 + 3x + \frac{3^2}{4}\right) - \frac{3^2}{4} + 2 = \left(x + \frac{3}{2}\right)^2 - \frac{3^2}{4} + 2$$

Um den Zusammenhang zur p - q -Formel zu verdeutlichen, vereinfachen wir den Ausdruck $-\frac{3^2}{4} + 2$ nicht weiter. Die Äquivalenzumformung „subtrahiere $-\frac{3^2}{4} + 2$ auf beiden Seiten“ führt nun zu folgender Gleichung:

$$\left(x + \frac{3}{2}\right)^2 = \frac{3^2}{4} - 2.$$

Diese Gleichung hat dieselbe Lösungsmenge wie unsere ursprüngliche Gleichung. Es ist bekannt, dass die Gleichung $y^2 = a$ für eine reelle Zahl $a > 0$ gerade die beiden Lösungen $y = \sqrt{a}$ und $y = -\sqrt{a}$ besitzt und für $a = 0$ nur die Lösung $y = 0$ besitzt.

Somit erfüllen die Lösungen der ursprünglichen quadratischen Gleichung eine der folgenden Gleichungen:

$$x_1 + \frac{3}{2} = \sqrt{\frac{3^2}{4} - 2} \quad \text{und} \quad x_2 + \frac{3}{2} = -\sqrt{\frac{3^2}{4} - 2}.$$

Die Äquivalenzumformung „subtrahiere $\frac{3}{2}$ auf beiden Seiten“ führt nun genau zu den in Satz 9 genannten Lösungen, welche die p - q -Formel liefert.

2.3 Lösung von Gleichungen durch Substitution

Substitution ist ein Verfahren, bei dem eine Gleichung in zwei Schritten gelöst wird. Wichtig hierfür ist, dass die Variable x in der Gleichung nur im selben (Teil-)Term vorkommt. Wir ersetzen (substituieren) dann diesen Term durch eine neue Variable und lösen die resultierende Gleichung. Aus den Lösungen dieser Gleichung ermittelt man dann die Lösungen für die ursprüngliche Gleichung.

Beispiel. Wir wollen die Gleichung $x^4 - x^2 - 6 = 0$ lösen. Dazu schreiben wir die Gleichung in der Form $(x^2)^2 - (x^2) - 6 = 0$ und substituieren $x^2 = u$. Das führt zu einer quadratischen Gleichung $u^2 - u - 6 = 0$, die wir mithilfe der p - q -Formel lösen können. Wir erhalten Lösungen $u_1 = -2$ und $u_2 = 3$.

Um die Lösungen der ursprünglichen Gleichungen zu ermitteln müssen wir nun nur noch die Gleichungen $x^2 = u_1$ und $x^2 = u_2$ lösen. Für u_1 gibt es keine Lösung in den reellen Zahlen, da x^2 nicht negativ sein kann. Für u_2 erhalten wir die Lösungen $x_1 = \sqrt{3}$ und $x_2 = -\sqrt{3}$.

Bemerkung. Substitution ist natürlich auch in komplizierteren Ausdrücken sinnvoll. Wenn Sie zum Beispiel die Gleichung

$$(x - 12)^3 - 27 = 0$$

lösen wollen, so ist es viel effizienter, mit der Substitution $y = x - 12$ zu arbeiten, statt erst alles auszumultiplizieren.

2.4 Bruch- und Wurzelgleichungen

Als Bruchgleichungen bezeichnet man Gleichungen, bei denen die Variable im Nenner eines Bruches vorkommt. Hier ist die Bestimmung der Definitionsmenge von zentraler Bedeutung für die Lösungen der Gleichung, da eine naiv ausgerechnete Lösung unter Umständen nicht im Definitionsbereich der Gleichung liegt.

Beispiel. Wir suchen alle Lösungen der Gleichung $\frac{x}{x-1} = \frac{1}{x-1}$. Multipliziert man beide Seiten mit $(x-1)$, so erhält man die Gleichung $x = 1$. Das kann aber keine Lösung unserer ursprünglichen Gleichung sein, da weder $\frac{x}{x-1}$ noch $\frac{1}{x-1}$ an der Stelle $x = 1$ definiert ist.

Das obige Beispiel führt zu folgendem Lösungsverfahren für Bruchgleichungen:

1. Wir bestimmen den Definitionsbereich der Gleichung, d.h. wir schließen alle Werte von x aus, für die der Nenner Null wird.
2. Vor jeder Multiplikation und Division klären wir, für welche Werte von x wir mit 0 multiplizieren bzw. durch 0 dividieren. Für diese Werte müssen wir die Gleichung gegebenenfalls noch einmal explizit überprüfen (außer sie wurden schon im ersten Schritt ausgeschlossen).

Mit dieser Vorgehensweise sehen wir, dass die Gleichung im obigen Beispiel nicht lösbar ist. Wir wollen nun noch ein weiteres Beispiel betrachten.

Beispiel. Wir suchen die Lösungen der Gleichung

$$\frac{2x}{x-3} = \frac{x-1}{x+4}.$$

Der Definitionsbereich der Gleichung ist offenbar $\mathbb{R} \setminus \{3, -4\}$. Für $x \neq 3, \neq -4$ multiplizieren wir beide Seiten mit $(x-3)(x+4)$ und erhalten die Gleichung

$$2x(x+4) = (x-1)(x-3).$$

Ausmultiplizieren liefert

$$2x^2 + 8x = x^2 - 4x + 3$$

und „Subtraktion von $x^2 - 4x + 3$ auf beiden Seiten“ führt uns zur Gleichung

$$x^2 + 12x - 3.$$

Mittels p - q -Formel erhalten wir also die beiden Lösungen

$$-\frac{12}{2} + \sqrt{\frac{(12)^2}{4} + 3} = -6 + \sqrt{39} \quad \text{und} \quad -\frac{12}{2} - \sqrt{\frac{(12)^2}{4} + 3} = -6 - \sqrt{39}.$$

Da sich beide erhaltenen Lösungen von 3 und -4 unterscheiden, sind die beiden Zahlen in (2.4) gerade die Lösungen unserer ursprünglichen Gleichung $\frac{2x}{x-3} = \frac{x-1}{x+4}$.

Ein ähnliches Vorgehen ist immer möglich, wenn die Gleichung nicht überall definiert ist. Nehmen wir beispielsweise Wurzelgleichungen, als Gleichungen in denen die Variable unter einer Wurzel vorkommt. Anstatt Werten, für die der Nenner Null wird, schließen wir in diesem Fall solche Werte aus, für die der Ausdruck unter der Wurzel negativ wird.

Beispiel. Wir suchen alle Lösungen der Gleichung $\sqrt{1+x} = \sqrt{1-x}$. Zuerst stellen wir fest, dass die linke Seite nur für $x \geq -1$ und die rechte Seite nur für $x \leq 1$ definiert ist. Somit sind nur Lösungen aus dem Intervall $[-1, 1]$ zulässig. Da für $x \in [-1, 1]$ beide Seiten der Gleichung nichtnegativ sind, liefert quadrieren die äquivalente Gleichung $1+x = 1-x$. Diese lineare Gleichung hat die eindeutige Lösung $x = 0$. Da die Lösung im Definitionsbereich liegt, ist sie auch die eindeutige Lösung der ursprünglichen Gleichung.

Ein wichtiger Hinweis zum Quadrieren. Beim Lösen von Wurzelgleichungen ist stets zu beachten, dass Quadrieren im Allgemeinen keine Äquivalenzumformung ist (bzw. nur dann eine Äquivalenzumformung ist, wenn für alle Belegungen der Variablen entweder beide Seiten der Gleichung nichtnegativ, oder beide Seiten nichtpositiv sind). Allerdings ist klar, dass $L = R$ stets $L^2 = R^2$ impliziert. Das heißt, wir verlieren durch Quadrieren keine Lösung, aber unter Umständen erhalten wir zusätzliche (Schein-)Lösungen. Diese kann man nachträglich mittels einer Probe eliminieren: man setzt alle erhaltenen Lösungen in die ursprüngliche Gleichung ein und behält nur jene, die tatsächlich eine wahre Aussage ergeben.

Beispiel. Wir suchen alle Lösungen der Gleichung

$$\sqrt{2x + \sqrt{x^2 + 1} + 1} = x + 1.$$

Zuerst quadrieren wir beide Seiten der Gleichung und erhalten

$$2x + \sqrt{x^2 + 1} + 1 = (x + 1)^2.$$

Wir subtrahieren $2x + 1$, um die Wurzel auf einer Seite zu isolieren, und erhalten

$$\sqrt{x^2 + 1} = x^2.$$

Nun quadrieren wir erneut:

$$x^2 + 1 = x^4.$$

Subtrahieren von $x^2 + 1$ und vertauschen der beiden Seiten ergibt

$$x^4 - x^2 - 1 = 0.$$

Durch Substitution von $u = x^2$ in obiger Gleichung, erhalten wir die Lösungen

$$u_1 = \frac{1}{2} + \sqrt{\frac{1}{4} + 1} = \frac{1 + \sqrt{5}}{2} \quad \text{und} \quad u_2 = \frac{1}{2} - \sqrt{\frac{1}{4} + 1} = \frac{1 - \sqrt{5}}{2}$$

und damit sind

$$x_1 = \sqrt{\frac{1 + \sqrt{5}}{2}} \quad \text{und} \quad x_2 = -\sqrt{\frac{1 + \sqrt{5}}{2}}$$

die beiden reellen Lösungen der Gleichung $x^4 - x^2 - 1 = 0$ (weil u_2 negativ ist, entsprechen dieser Lösung der Hilfsgleichung keine Lösungen der ursprünglichen Gleichung, denn $x^2 = u_2$ hat für negative u_2 keine Lösung).

Zu guter Letzt führen wir mit unseren ermittelten Lösungen x_1 und x_2 noch die Probe durch. Setzen wir x_1 in unsere ursprüngliche Gleichung ein, so bekommen wir eine wahre Aussage. Setzen wir hingegen x_2 ein, bekommen wir eine falsche Aussage, weil man sich leicht überzeugt, dass $x_2 + 1$ negativ ist. Somit ist x_1 unsere einzige Lösung der ursprünglichen Gleichung.

2.5 Gleichungen mit Beträgen

Definition. Die *Betragsfunktion* $|\cdot|$ ist eine Funktion von \mathbb{R} nach \mathbb{R} , welche durch

$$\begin{aligned} |x| &= x && \text{wenn } x \geq 0 \\ |x| &= -x && \text{wenn } x < 0 \end{aligned}$$

definiert ist. Alternativ kann man die Betragsfunktion auch durch Quadrieren und Wurzelziehen beschreiben:

$$|x| = \sqrt{x^2} \quad \text{für alle reellen Zahlen } x.$$

Fasst man die reellen Zahlen als Punkte einer Gerade auf, so kann man $|x|$ als Abstand der Zahl x zum Nullpunkt $0 \in \mathbb{R}$ interpretieren. Eine Gleichung in der die Variable im Inneren eines Betrages auftritt heißt Betragsgleichung. Um solche Gleichungen zu lösen bietet es sich an, zuerst den Betrag mittels der obigen Unterscheidung aufzulösen.

Beispiel. Wir suchen alle Lösungen der Gleichung $|x - 1| + 1 = 0$. Wir betrachten zwei Fälle:

1. $x - 1 \geq 0$: In diesem Fall vereinfacht sich die Gleichung zu $x = 0$, allerdings gilt für diese Lösung nicht $x - 1 \geq 0$, somit ist sie nicht zulässig.
2. $x - 1 < 0$: In diesem Fall vereinfacht sich die Gleichung zu $-x + 2 = 0$, die Lösung ist somit $x = 2$, wiederum ist die Lösung nicht zulässig, da sie die Gleichung $x - 1 < 0$ nicht erfüllt.

Die Lösungsmenge obiger Gleichung ist somit leer. Dies ist auch wenig überraschend, denn für alle $x \in \mathbb{R}$ gilt $|x - 1| \geq 0$, und wenn man zu einer nichtnegativen Zahl 1 addiert, kann man nicht 0 erhalten.

Mit einer ähnlichen Fallunterscheidung wie im eben behandelten Beispiel kann man auch allgemeinere Betragsgleichungen lösen:

1. Zerlege \mathbb{R} in Intervalle deren Grenzen dadurch gegeben sind, dass ein Argument der vorkommenden Betragsfunktionen dort das Vorzeichen ändert.
2. Auf jedem der Intervalle können wir die Beträge auflösen und die entsprechende Gleichung lösen. Danach überprüfen wir noch, ob die gefundenen Lösungen im entsprechenden Intervall liegen (falls nicht, verwerfen wir sie).
3. Die Lösungsmenge der Gleichung ist nun die Vereinigung der Lösungsmengen aus den einzelnen Intervallen.

Beispiel. Wir lösen die Gleichung $|x + 4| - |x - 2| = 1$.

1. In der Gleichung kommen zwei Beträge vor, die das Vorzeichen bei -4 bzw. bei 2 ändern. Die Intervalle in die wir die reellen Zahlen zerlegen sind somit $(-\infty, -4)$, $[-4, 2]$ und $(2, \infty)$.

2. Nun betrachten wir die Gleichung auf jedem der drei Intervalle separat:
- Im Intervall $(-\infty, -4)$ gilt $x + 4 < 0$ und $x - 2 < 0$. Somit vereinfacht sich die Gleichung zu $-(x + 4) + (x - 2) = 1$ bzw. zu $-6 = 1$, was eine falsche Aussage ist.
 - Im Intervall $[-4, 2]$ gilt $x + 4 \geq 0$ und $x - 2 \leq 0$. Somit vereinfacht sich die Gleichung zu $(x + 4) + (x - 2) = 1$ bzw. zu $2x + 1 = 0$. Diese Gleichung hat die Lösung $x = -\frac{1}{2}$. Da diese Lösung im betrachteten Intervall $[-4, 2]$ liegt, ist sie tatsächlich eine Lösung der ursprünglichen Gleichung.
 - Im Intervall $(2, \infty)$ gilt $x + 4 \geq 0$ und $x - 2 \geq 0$. Somit vereinfacht sich die Gleichung zu $(x + 4) - (x - 2) = 1$ bzw. $2 = 1$, somit gibt es auch hier keine Lösung.
3. Zuletzt vereinigen wir die Lösungsmengen aus den verschiedenen Teilintervallen und erhalten die Gesamtlösungsmenge, also in unserem Fall $\{-\frac{1}{2}\}$.

2.6 Schlussbemerkung

Wir haben Ihnen in diesem Kapitel einige einfache Gleichungen und Methoden für deren Lösung, d.h. das Auffinden der vollständigen und exakten Lösungsmenge aufgezeigt. In der Schule und auch an der Universität haben die meisten in den Übungsaufgaben auftretenden Gleichungen tatsächlich Lösungen, welche durch mehr oder weniger clevere Äquivalenzumformungen und weitere Tricks gefunden werden können. Allerdings ist dies eine stark eingeschränkte Auswahl aus den *möglichen* Gleichungen, welche einem in der Praxis begegnen können.

In der Tat ist es so, dass *die meisten Gleichungen nicht explizit lösbar sind*, d.h. man kann die Lösungsmengen nicht wie in unseren Beispielen explizit angeben. Stattdessen gibt es viele, oft speziell auf die jeweilige Situation angepasst Verfahren, um *Näherungslösungen* für Gleichungen zu bestimmen, an denen man aus praktischen Gründen interessiert ist. Einige dieser Verfahren werden Sie zum Beispiel in Vorlesungen über numerische Mathematik kennenlernen. Hier treffen Theorie und Praxis aufeinander: einerseits will man möglichst effiziente Algorithmen, andererseits benötigt man auch zuverlässige Aussagen, unter welchen Umständen das jeweilige Verfahren tatsächlich zum gewünschten Ergebnis führt und wie lange man braucht, um eine vorgegebene Genauigkeit für die gewünschte Approximation zu erreichen.

3 Ungleichungen

Ersetzt man das Gleichheitszeichen in einer Gleichung durch ein Ordnungszeichen $>$, $<$, \geq , \leq so erhält man eine Ungleichung. Wie bei Gleichungen nennen wir die Teile links und rechts des Ordnungszeichens linke bzw. rechte Seite der Ungleichung. Eine Lösung einer Ungleichung ist (analog zum Fall von Gleichungen) eine Variablenbelegung, die die Ungleichung zu einer wahren Aussage macht. Die Lösungsmenge einer Ungleichung ist die Menge aller Lösungen. Ebenfalls analog zum Fall der Gleichungen wird der Definitionsbereich einer Ungleichung definiert.

Beim Lösen von Ungleichungen spielen erneut Äquivalenzumformungen eine Rolle, wobei eine Äquivalenzumformung auch hier wieder eine Umformung ist, die die Lösungsmenge unverändert lässt.

Die folgenden Umformungen sind Beispiele für Äquivalenzumformungen:

- Vertauschen von rechter und linker Seite bei gleichzeitigem Umkehren des Ordnungssymbols (aus $<$ wird $>$, aus \leq wird \geq und umgekehrt).
- Addition und Subtraktion desselben Ausdrucks auf beiden Seiten.
- Multiplikation beider Seiten mit demselben Faktor $c > 0$.
- Multiplikation beider Seiten mit demselben $c < 0$ bei gleichzeitigem Umkehren des Ordnungssymbols.
- Division beider Seiten durch dasselbe $c > 0$.
- Division beider Seiten durch dasselbe $c < 0$ bei gleichzeitigem Umkehren des Ordnungssymbols.

Man beachte, dass die beiden letzten Beispiele von Äquivalenzumformungen wieder den beiden vorangehenden Beispielen entsprechen, da Division durch ein $c \neq 0$ gerade der Multiplikation mit $\frac{1}{c}$ entspricht.

Bei Multiplikation und Division mit Ausdrücken ist hier noch mehr Vorsicht geboten als im Fall von Gleichungen. Man muss nicht nur sicherstellen, dass der Ausdruck nicht den Wert 0 besitzt, sondern auch feststellen, ob er größer oder kleiner als 0 ist, um zu wissen, ob das Ordnungssymbol umgedreht werden muss. Ein Beispiel dazu:

Beispiel. Wir starten mit der Ungleichung

$$\frac{1}{x} < 5.$$

Durch Umstellen der Terme erscheint es natürlich, auf die Ungleichung

$$\frac{1}{5} < x$$

zu kommen. **Dieser Schluss ist falsch**, denn wir haben keine Äquivalenzumformung vorgenommen. Wenn man zum Beispiel in der ersten Gleichung $x = -2$ einsetzt, dann erhält man die wahre Aussage $-\frac{1}{2} < 5$. Wenn man aber in der zweiten Gleichung $x = -2$ einsetzt, so erhält man die falsche Aussage $\frac{1}{5} < -2$. Das Problem liegt darin, dass man glaubt, man könne beide Seiten mit x multiplizieren. Das geht auch gut, solange x positiv ist, aber eben nicht für negative x . Das Problem kann aber durch eine Fallunterscheidung behoben werden. Wenn wir mit $\frac{1}{x} < 5$ starten, dann gibt es folgende zwei Fälle.

Fall 1: x ist positiv. Wir erhalten dann $\frac{1}{5} < x$;

Fall 2: x ist negativ. Wir erhalten dann $\frac{1}{5} > x$. Dies gilt für alle negativen x .

Insgesamt hat die Lösungsmenge unserer Ungleichung also die Form

$$\left\{ x \in \mathbb{R} \mid x < 0 \text{ oder } x > \frac{1}{5} \right\}.$$

3.1 Lineare Ungleichungen

Definition. Eine *lineare Ungleichung* für eine Variable x ist eine Ungleichung, welche sich durch Äquivalenzumformungen auf eine der Formen

$$ax + b > 0, \quad ax + b < 0, \quad ax + b \geq 0 \quad \text{oder} \quad ax + b \leq 0,$$

mit gegebenen Parametern $a, b \in \mathbb{R}$ mit $a \neq 0$ bringen lässt.

Um eine lineare Ungleichung zu lösen, kann man genauso vorgehen wie bei linearen Gleichungen. Wir werden dies anhand einiger Beispiele sehen. Man beachte, dass sich das Ordnungssymbol bei manchen Äquivalenzumformungen umkehrt. Kommt in dem Ausdruck, mit dem wir multiplizieren eine Variable vor, so müssen wir gegebenenfalls eine Fallunterscheidung durchführen.

Beispiel. Wir wollen die Lösungsmenge der linearen Ungleichung $2x + 4 > 0$ bestimmen.

$$\begin{array}{ll} 2x + 4 > 0 & | - 4 \\ 2x > -4 & | : 2 \text{ (das Ordnungssymbol bleibt gleich, weil } 2 > 0) \\ x > -2 & \end{array}$$

Die Lösungsmenge der Ungleichung ist somit das Intervall $(-2, \infty)$.

Beispiel. Wir bestimmen die Lösungsmenge der linearen Ungleichung $-3x + 6 > 0$:

$$\begin{array}{ll} -3x + 6 > 0 & | - 6 \\ -3x > -6 & | : (-3) \text{ (das Ordnungssymbol kehrt sich um, weil } -3 < 0) \\ x < 2 & \end{array}$$

Die Lösungsmenge dieser Gleichung ist somit das Intervall $(-\infty, 2)$.

Diese beiden Beispiele kann man nun leicht auf allgemeine a und b verallgemeinern und erhält:

- falls $a > 0$, dann ist die Lösungsmenge der Ungleichung $ax + b > 0$ genau das Intervall $(-\frac{b}{a}, \infty)$.
- falls $a < 0$, dann ist die Lösungsmenge der Ungleichung $ax + b > 0$ genau das Intervall $(-\infty, -\frac{b}{a})$.

Analoge Aussagen gelten auch für die anderen linearen Ungleichungen.

3.2 Quadratische Ungleichungen

Definition. Als *quadratische Ungleichung* für eine Variable x bezeichnet man eine Ungleichung, welche sich durch Äquivalenzumformungen in eine der folgenden Formen bringen lässt, wobei a, b und c reelle Parameter mit $a \neq 0$ sind:

$$\begin{array}{ll} ax^2 + bx + c > 0, & ax^2 + bx + c < 0, \\ ax^2 + bx + c \geq 0, & ax^2 + bx + c \leq 0. \end{array}$$

Genau wie quadratische Gleichungen können auch quadratische Ungleichungen mittels Division durch a auf eine Normalform

$$\begin{array}{ll} x^2 + px + q > 0, & x^2 + px + q < 0, \\ x^2 + px + q \geq 0, & x^2 + px + q \leq 0 \end{array}$$

gebracht werden. Hierbei ist zu beachten, dass sich das Ordnungssymbol je nach Vorzeichen von a umkehren kann.

Um quadratische Ungleichungen in Normalform zu lösen, betrachten wir zunächst die Funktion¹ $f(x) = x^2 + px + q$. Der Graph dieser Funktion ist eine nach oben geöffnete Parabel, somit ist es leicht, festzustellen in welchen Bereichen der Funktionswert größer bzw. kleiner als 0 ist. In der Abbildung 3.1 sind die drei verschiedenen Fälle abgebildet:

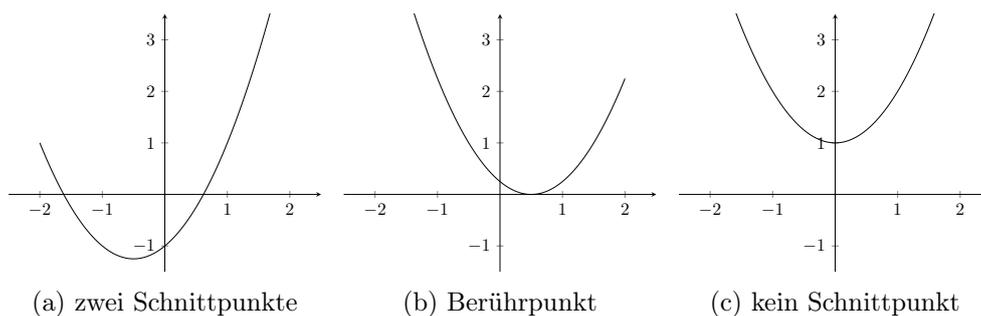


Abbildung 3.1: Lage nach oben geöffneter Parabeln relativ zur x -Achse

¹Funktionen werden wir in der nächsten Woche noch ausführlich diskutieren, hier appellieren wir an Schulwissen.

- Schneidet die Parabel die x -Achse in 2 Punkten x_1 und x_2 , so ist der Funktionswert zwischen den beiden Schnittpunkten kleiner als 0 und außerhalb der Schnittpunkte größer als 0. An den Schnittpunkten ist der Funktionswert gleich 0.
- Berührt die Parabel die x -Achse in einem Punkt, so ist der Funktionswert überall außer in diesem Punkt größer als 0. Am Berührungspunkt ist der Funktionswert gleich 0.
- Verläuft die Parabel komplett oberhalb der x -Achse, so ist der Funktionswert überall größer als 0.

Aufgrund der obigen Eigenschaften der Funktion $f(x) = x^2 + px + q$ kann man nun folgendes Verfahren zur Lösung quadratischer Ungleichungen herleiten.

1. Wir bringen die Ungleichung auf Normalform.
2. Wir bestimmen die Lösungen der quadratischen Gleichung, die aus der Ungleichung entsteht, wenn wir das Ordnungssymbol durch ein Gleichheitszeichen ersetzen. Das können wir beispielsweise mit der p - q -Formel machen. Nenne diese Lösungsmenge L .
3. Wir nutzen die obigen Eigenschaften der Funktion $f(x) = x^2 + px + q$, um die Lösungsmenge unserer ursprünglichen Ungleichung zu bestimmen. Die folgende Tabelle fasst zusammen, wie diese Lösungsmenge aussehen muss. Das ist abhängig vom Ordnungssymbol in der Ungleichung sowie der Lösungsmenge L der zugehörigen Gleichung (im Fall einer zweielementigen Lösungsmenge schreiben wir L als $\{x_1, x_2\}$ mit $x_1 < x_2$).

Ungleichung	$L = \emptyset$	$L = \{x_0\}$	$L = \{x_1, x_2\}$
$x^2 + px + q > 0$	\mathbb{R}	$\mathbb{R} \setminus \{x_0\}$	$(-\infty, x_1) \cup (x_2, \infty)$
$x^2 + px + q < 0$	\emptyset	\emptyset	(x_1, x_2)
$x^2 + px + q \geq 0$	\mathbb{R}	\mathbb{R}	$(-\infty, x_1] \cup [x_2, \infty)$
$x^2 + px + q \leq 0$	\emptyset	$\{x_0\}$	$[x_1, x_2]$

Beispiel. Wir suchen die Lösungsmenge der Ungleichung $x^2 + 2x + 3 < 0$.

1. Im ersten Schritt ist nichts zu tun, da die Ungleichung schon in Normalform vorliegt.
2. Die quadratische Gleichung $x^2 + 2x + 3 = 0$ besitzt keine Lösung, da die Diskriminante $\frac{2^2}{4} - 3 = -2$ kleiner als 0 ist.
3. Somit ist die Lösungsmenge der Ungleichung die leere Menge.

Beispiel. Gesucht ist die Lösungsmenge der Ungleichung $-3x^2 + 3x + 6 \geq 0$.

1. Im ersten Schritt dividieren wir beide Seiten der Ungleichung durch -3 . Da $-3 < 0$ ist, müssen wir das Ordnungssymbol umkehren. Somit erhalten wir folgende Ungleichung in Normalform: $x^2 - x - 2 \leq 0$

- Wir lösen die zugehörige Gleichung mittels der p - q -Formel (oder einer der binomischen Formeln) und erhalten die Lösungen $x_1 = -1$ und $x_2 = 2$.
- Die Lösungsmenge der Gleichung besteht also aus dem Intervall $[-1, 2]$.

3.3 Betragsungleichungen

Als Betragsungleichungen bezeichnen wir Ungleichungen, in denen die Variable x innerhalb einer Betragsfunktion auftritt.

Der Lösungsweg für Betragsungleichungen ist im wesentlichen der gleiche wie für Betragsgleichungen.

- Wir teilen den Definitionsbereich in Intervalle auf, in denen keiner der vorkommenden Beträge das Vorzeichen ändert.
- Dann lösen wir die Ungleichung auf jedem der Intervalle und behalten nur jene Lösungen, die auch tatsächlich im entsprechenden Intervall liegen.
- Die Gesamtlösungsmenge ergibt sich wieder als Vereinigung der Lösungsmengen der einzelnen Intervalle.

Wir werden dieses Verfahren nun anhand von Beispielen veranschaulichen.

Beispiel. Wir suchen alle Lösungen der Ungleichung $|x + 1| - |2x - 4| + 7 < 0$.

- Zuerst teilen wir \mathbb{R} in Intervalle auf. $(x + 1)$ wechselt das Vorzeichen bei $x = -1$, $(2x - 4)$ wechselt das Vorzeichen bei $x = 2$. Somit sind die Intervalle, die wir betrachten müssen $(-\infty, -1)$, $[-1, 2)$ und $[2, \infty)$.
- Wir lösen nun die Ungleichung in jedem der Intervalle.

- Im Intervall $(-\infty, -1)$ ist $x + 1 < 0$ und $2x - 4 < 0$. Somit wird die linke Seite der Ungleichung zu $-(x + 1) + (2x - 4) + 7 = x + 2$. Auf dem Intervall $(-\infty, -1)$ ist also unsere ursprüngliche Ungleichung äquivalent zur Ungleichung $x + 2 < 0$. Die Lösungsmenge dieser Ungleichung ist also $\{x \mid x < -2\} = (-\infty, -2)$.

Wir müssen noch auf das Intervall $(-\infty, -1)$ einschränken und erhalten somit als Lösungsmenge in diesem ersten Fall $(-\infty, -1) \cap (-\infty, -2) = (-\infty, -2)$.

- Im Intervall $[-1, 2)$ ist $x + 1 \geq 0$ und $2x - 4 < 0$. Die linke Seite der Ungleichung wird somit zu $(x + 1) + (2x - 4) + 7 = 3x + 4$. Somit haben wir auf dem Intervall $[-1, 2)$ wieder eine lineare Ungleichung, nämlich $3x + 4 < 0$. Die Lösungsmenge dieser Ungleichung ist $\{x \mid x < -\frac{4}{3}\} = (-\infty, -\frac{4}{3})$.

Wieder schränken wir die Lösungsmenge auf das betrachtete Intervall ein und erhalten $[-1, 2) \cap (-\infty, -\frac{4}{3}) = \emptyset$.

- Im Intervall $[2, \infty)$ ist $x + 1 \geq 0$ und $2x - 4 \geq 0$. Somit erhalten wir als linke Seite $(x + 1) - (2x - 4) + 7 = -x + 12$. Die Ungleichung wird damit zu $-x + 12 < 0$ auf dem Intervall $[2, \infty)$. Die Lösungsmenge dieser Ungleichung

ist $\{x \mid x > 12\} = (12, \infty)$. (Vorsicht: beim Lösen dieser linearen Ungleichung muss das Ordnungssymbol umgekehrt werden!)

Wir schränken erneut die Lösungsmenge auf das betrachtete Intervall ein und erhalten $[2, \infty) \cap (12, \infty) = (12, \infty)$.

3. Insgesamt erhalten wir also als Lösungsmenge $(-\infty, -2) \cup (12, \infty)$.

Beispiel. Wir suchen die Lösungsmenge der Ungleichung $|x + 2| + |x - 2| - |x| - 3 \leq 0$.

1. Die Punkte, an denen die Beträge das Vorzeichen wechseln sind -2 , 0 und 2 . Die Intervalle, die wir betrachten müssen sind somit $(-\infty, -2)$, $[-2, 0)$, $[0, 2)$ und $[2, \infty)$.

2. Wir lösen die Gleichung in jedem der Intervalle:

- In $(-\infty, -2)$ erhalten wir für die linke Seite $-(x+2) - (x-2) + x - 3 = -x - 3$. Die Ungleichung wird zu $-x - 3 \leq 0$, die Lösungsmenge dieser Ungleichung ist $\{x \mid x \geq -3\} = [-3, \infty)$.

Schränken wir diese Menge auf das betrachtete Intervall ein, so erhalten wir $(-\infty, -2) \cap [-3, \infty) = [-3, -2)$.

- In $[-2, 0)$ erhalten wir für die linke Seite $(x+2) - (x-2) + x - 3 = x + 1$. Die Ungleichung wird zu $x + 1 \leq 0$, die Lösungsmenge dieser Ungleichung ist $\{x \mid x \leq -1\} = (-\infty, -1]$.

Die Menge der Lösungen, die im betrachteten Intervall liegen ist also $[-2, 0) \cap (-\infty, -1] = [-2, -1]$.

- In $[0, 2)$ erhalten wir für die linke Seite $(x+2) - (x-2) - x - 3 = -x + 1$. Die Ungleichung wird zu $-x + 1 \leq 0$, die Lösungsmenge dieser Ungleichung ist $\{x \mid x \geq 1\} = [1, \infty)$.

Wir schränken diese Menge auf das Intervall $[0, 2)$ ein und erhalten $[0, 2) \cap [1, \infty) = [1, 2)$.

- In $[2, \infty)$ erhalten wir für die linke Seite $(x+2) + (x-2) - x - 3 = x - 3$. Die Ungleichung wird zu $x - 3 \leq 0$, die Lösungsmenge dieser Ungleichung ist $\{x \mid x \leq 3\} = (-\infty, 3]$.

Schränken wir diese Menge auf das betrachtete Intervall ein, so erhalten wir $[2, \infty) \cap (-\infty, 3] = [2, 3]$.

3. Die Lösungsmenge der ursprünglichen Betragsungleichung ergibt sich nun als Vereinigung der Lösungsmengen aus den einzelnen Fällen:

$$L = [-3, -2) \cup [-2, -1] \cup [1, 2) \cup [2, 3] = [-3, -1] \cup [1, 3].$$

Wir haben gesehen, dass die Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$, welche durch $f(x) = |x + 2| + |x - 2| - |x| - 3$ definiert ist, auf den einzelnen Intervallen

$$(-\infty, -2), \quad [-2, 0), \quad [0, 2), \quad [2, \infty)$$

linear ist. Es gilt

$$f(-2) = -1 \quad f(0) = 1 \quad f(2) = -1.$$

Damit können wir nun den Graphen von f in einem Koordinatensystem zeichnen:

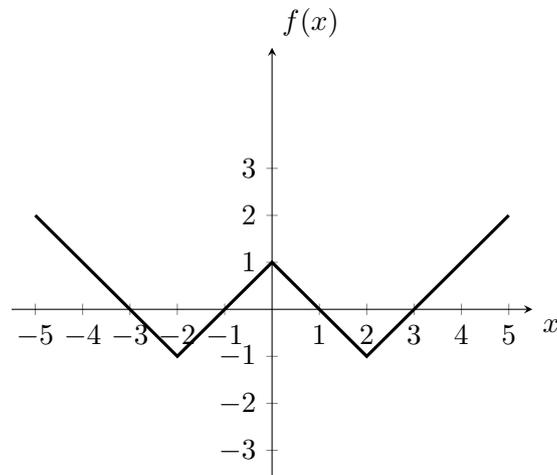


Abbildung 3.2: Graph der Funktion $f(x) = |x + 2| + |x - 2| - |x| - 3$

Zuletzt wollen wir nun auch noch ein Beispiel für eine quadratische Betrags-Ungleichung machen.

Beispiel. Wir wollen die Ungleichung $|x^2 - 1| + |x^2 + 1| > 0$ lösen.²

1. Die Punkte, wo $|x^2 - 1|$ das Vorzeichen wechselt, sind gerade die reellen Nullstellen von $x^2 - 1$, also wir das Vorzeichen in ± 1 gewechselt. Da $x^2 + 1$ immer positiv ist, gilt $|x^2 + 1| = x^2 + 1$ auf ganz \mathbb{R} . Also unterteilen wir \mathbb{R} in die Intervalle $(-\infty, -1)$, $[-1, 1)$ und $[1, \infty)$.
2. Wir lösen nun wieder unsere Ungleichung auf jedem Teilintervall separat.
 - In $(-\infty, -1)$ ist unsere Ungleichung gegeben als $x^2 - 1 + x^2 + 1 = 2x^2 > 0$. Die Lösungsmenge auf diesem Intervall ist also $(-\infty, -1) \cap \mathbb{R} \setminus \{0\} = (-\infty, -1)$.
 - In $[-1, 1)$ ist unsere Ungleichung gegeben als $-(x^2 - 1) + x^2 + 1 = 2 > 0$. Die Lösungsmenge auf diesem Intervall ist also gerade $[-1, 1)$.
 - In $[1, \infty)$ ist unsere Ungleichung gegeben als $x^2 - 1 + x^2 + 1 = 2x^2 > 0$. Die Lösungsmenge auf diesem Intervall ist also $[1, \infty) \cap \mathbb{R} \setminus \{0\} = [1, \infty)$.
3. Insgesamt erhalten wir also für die Lösungsmenge

$$(-\infty, -1) \cup [-1, 1) \cup [1, \infty) = \mathbb{R},$$

d.h. unsere ursprüngliche Ungleichung ist für alle reellen Zahlen erfüllt.

²Bevor Sie weiterlesen: welche Lösungsmenge würden Sie erwarten?

In Anbetracht der Tatsache, dass $|x^2 + 1| \geq |1| = 1$ und $|x^2 - 1| \geq 0$ für alle $x \in \mathbb{R}$ gilt, ist dies auch wenig überraschend. Der gerade beschriebene Lösungsweg funktioniert aber auch, wenn die Lösung weniger “offensichtlich” ist.

Wir kennen die Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = |x^2 - 1| + |x^2 + 1|$, nun auf den einzelnen Teilintervallen $(-\infty, -1)$, $[-1, 1)$, $[1, \infty)$, und zwar gilt

$$\begin{aligned} f(x) &= 2x^2 && \text{für } x \text{ in } (-\infty, -1), \\ f(x) &= 2 && \text{für } x \text{ in } [-1, 1), \\ f(x) &= 2x^2 && \text{für } x \text{ in } [1, \infty). \end{aligned}$$

Also ist der Graph von f gegeben durch

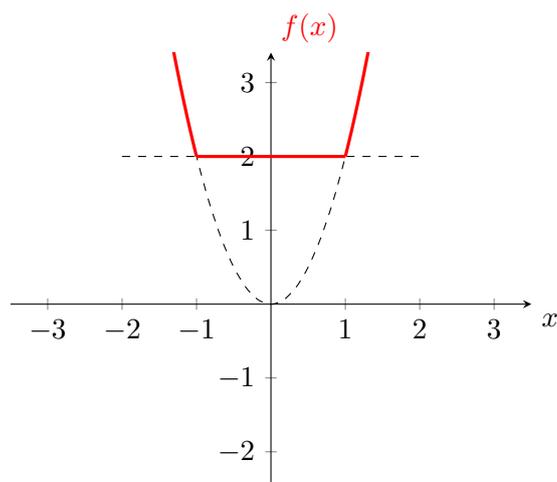


Abbildung 3.3: Graph der Funktion $f(x) = |x^2 - 1| + |x^2 + 1|$

Insgesamt sehen wir (aus den Formeln der Funktion f) auf den einzelnen Teilintervallen, dass wir diese Funktion einfacher schreiben können als

$$|x^2 - 1| + |x^2 + 1| = f(x) = \max\{2x^2, 2\} = 2 \max\{x^2, 1\},$$

wobei $\max\{a, b\}$ gerade wieder die größere der beiden Zahlen a und b ist.

3.4 Zusammenfassung

Abschließend lässt sich zu diesem Kapitel noch Folgendes sagen: Ungleichungen präsentieren im Prinzip keine wesentlichen neuen Schwierigkeiten im Vergleich zu Gleichungen, allerdings muss man besonders sorgfältig arbeiten, um in den nun gehäuft auftretenden Fallunterscheidungen die Übersicht nicht zu verlieren.

Für “gutartige” Ungleichungen einer Variablen (genauer: solche von der Form $f(x) > 0$, $f(x) \geq 0$ usw., wobei f eine *stetige Funktion* der Variablen x ist), ist es oft am praktischsten, zunächst die zugehörige Gleichung $f(x) = 0$ zu lösen. Dies liefert einem mit etwas

Glück endlich viele Punkte, welche die reellen Zahlen in gewisse Teilintervalle zerlegen. Im Inneren jedes dieser Teilintervalle hat die Funktion $f(x)$ ein festes Vorzeichen³, so dass das entsprechende Intervall je nach ursprünglicher Ungleichung entweder vollständig zur Lösungsmenge oder vollständig nicht zur Lösungsmenge gehört. Genau dieses Verfahren hatten wir im Fall der quadratischen Ungleichungen oben bereits vorgestellt.

³Dies folgt aus dem sogenannten *Zwischenwertsatz*, welchen Sie in der Analysis kennenlernen werden.

4 Lineare Algebra und Geometrie

Zweitausend Jahre lang wurde Geometrie eher synthetisch/deskriptiv betrieben, und Euklids *Elemente* waren das ultimative Lehrbuch dazu.

Ein großer Umbruch in der Geometrie geschah durch die Einführung von Koordinaten, welche man zur Erinnerung an René Descartes auch *kartesische Koordinaten* nennt. Dadurch kann man Punkte im Raum, welche geometrische Objekte beschreiben, mittels Systemen von Zahlen beschreiben und untersuchen. Diese Methode zur Behandlung geometrischer Fragen durch (lineare) Algebra nennt man *analytischen Geometrie*.

4.1 Vektoren

In vielen Anwendungen (beispielsweise in der Physik) ist es nicht nur relevant, wie groß eine Entität (zum Beispiel eine Kraft) ist, sondern auch, welche Richtung sie hat. Wir betrachten also Pfeile in unserem Raum, sagen wir zuerst einmal der Einfachheit halber in der Ebene:

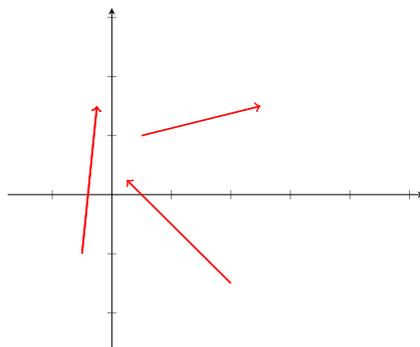


Abbildung 4.1: Pfeile in der Ebene

Wir betrachten nun zwei Pfeile, welche durch eine Verschiebung ineinander überführt werden können, also die gleiche Länge und die gleiche “Richtung” besitzen, als den gleichen *Vektor*. So werden alle Pfeile der Abbildung 4.2 als der gleiche Vektor betrachtet.

Bemerkung. Wie schon bei den rationalen Zahlen ist auch der Übergang von Pfeilen zu Vektoren ein Beispiel für eine Äquivalenzrelation, wie wir sie später diskutieren werden.

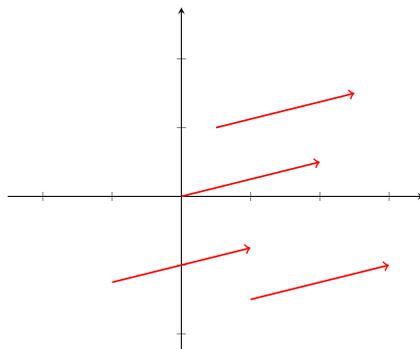


Abbildung 4.2: Pfeile, welche sich nur durch eine Verschiebung unterscheiden beschreiben denselben Vektor.

Haben wir zwei Koordinatenachsen in unserer Ebene gewählt, so können wir jeden Punkt der Ebene durch zwei reelle Koordinaten beschreiben. Den Schnittpunkt der Achsen bezeichnen wir als *Nullpunkt* oder *Ursprung*, er hat die Koordinaten $(0, 0)$.

Nun können wir den Startpunkt eines beliebigen Pfeils in den Ursprung verschieben und ändern dabei nicht den Vektor, welcher durch diesen Pfeil beschrieben wird. Solch ein Pfeil mit Startpunkt im Ursprung ist aber durch seinen Endpunkt eindeutig bestimmt. Also kann ein Vektor in der Ebene auch eindeutig durch ein Paar von reellen Zahlen beschrieben werden, nämlich die Koordinaten dieses Endpunktes. Analog lassen sich Vektoren im dreidimensionalen Raum durch ein Tripel von reellen Zahlen beschreiben. Für ein allgemeines n , kann also ein n -dimensionaler Vektor als ein n -Tupel (x_1, x_2, \dots, x_n) von reellen Zahlen $x_1, \dots, x_n \in \mathbb{R}$ geschrieben werden.

Definition. Die Gesamtheit aller n -dimensionalen Vektoren wird als

$$\mathbb{R}^n = \{(x_1, \dots, x_n) \mid x_1, \dots, x_n \in \mathbb{R}\}$$

notiert.

Bemerkung. Der \mathbb{R}^n beschreibt sowohl die *Punkte* einer Ebene als auch ihre *Vektoren*. Dies führt zu Beginn manchmal zu Verwirrung, ist aber oft auch sehr praktisch. Viele Autoren sind in der Trennung zwischen Punkten und Vektoren auch nicht immer ganz konsequent. Wir werden diese Tradition hier fortführen.

Als Unterscheidung zu den Vektoren, werden die reellen Zahlen in diesem Zusammenhang üblicherweise *Skalare* genannt. Um Vektoren besser von Skalaren unterscheiden zu können, werden Vektoren in diesem Text fett dargestellt. So ist \mathbf{x} ein Vektor, aber x ein Skalar. Andere gebräuchliche Notationen für Vektoren sind \vec{x} und \underline{x} , es wird aber auch häufig darauf verzichtet, den Unterschied zwischen Skalar und Vektor typographisch hervorzuheben.

Im Rahmen dieses Vorkurses werden wir uns hauptsächlich mit Vektoren im \mathbb{R}^2 und \mathbb{R}^3 auseinandersetzen. Dies hat den Vorteil, dass wir die Vektoren (nach der Identifikation mit Punkten) auch zeichnen können.

Für Vektoren kann man folgende Rechenoperationen definieren:

- Addition von Vektoren:

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

- Multiplikation mit Skalaren:

$$c \cdot (x_1, x_2, \dots, x_n) = (cx_1, cx_2, \dots, cx_n)$$

Geometrisch kann die Addition von zwei Vektoren \mathbf{x} und \mathbf{y} wie folgt gedeutet werden: Wähle zunächst Pfeile, welche \mathbf{x} und \mathbf{y} repräsentieren, so dass der Startpunkt des zu \mathbf{y} gehörenden Pfeils mit dem Endpunkt des zu \mathbf{x} gehörenden Pfeils übereinstimmt. Dann entspricht $\mathbf{x} + \mathbf{y}$ dem Pfeil, welcher am Startpunkt des Repräsentanten von \mathbf{x} beginnt und am Endpunkt des Repräsentanten von \mathbf{y} endet. Man erhält dasselbe Resultat, wenn die Rollen von \mathbf{x} und \mathbf{y} vertauscht werden. Multiplikation mit einer reellen Zahl $c \in \mathbb{R}$ bedeutet, den Vektor um den entsprechenden Faktor zu skalieren (darum die Bezeichnung *Skalare*). Falls $|c| > 1$ entspricht dies einer echten Streckung, für $0 \leq c < 1$ wird der Vektor verkürzt. Im Fall $c < 0$ wird außerdem die Richtung des Vektors umgedreht.

Beispiel. Sei $\mathbf{x} = (1, 4)$ und $\mathbf{y} = (-3, 1)$. Dann ist $\mathbf{x} + \mathbf{y} = (1 - 3, 4 + 1) = (-2, 5)$ Sei weiter $\mathbf{v} = (1, 3)$. Dann ist $2 \cdot \mathbf{v} = (2, 6)$ und außerdem ist $-\mathbf{v} = (-1) \cdot \mathbf{v} = (-1, -3)$.

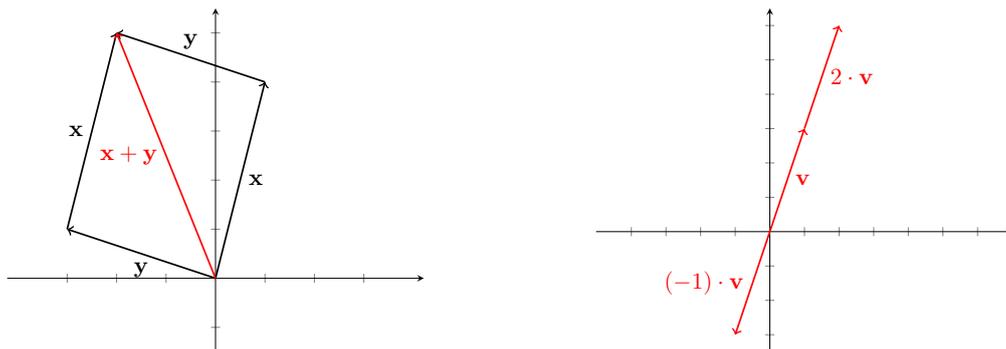


Abbildung 4.3: Addition von Vektoren und Multiplikation von Vektoren mit Skalaren

Der Vektor $\mathbf{y} - \mathbf{x}$ hat auch eine interessante Interpretation: er wird gerade von dem Pfeil repräsentiert, welcher im Endpunkt von \mathbf{x} startet und im Endpunkt von \mathbf{y} endet. Dies folgt direkt aus der Gleichung $\mathbf{x} + (\mathbf{y} - \mathbf{x}) = \mathbf{y}$.

Beispiel. Sei $\mathbf{x} = (-3, 2)$ und $\mathbf{y} = (2, 5)$. Dann ist $\mathbf{y} - \mathbf{x} = (2 - (-3), 5 - 2) = (5, 3)$.

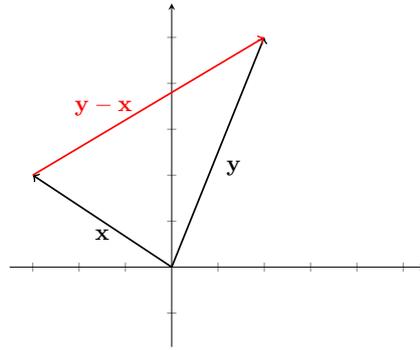


Abbildung 4.4: Differenz von Vektoren

Definition. Der Vektor $(0, 0, \dots, 0)$ heißt *Nullvektor*. Er wird häufig mit dem Symbol $\mathbf{0}$ bezeichnet (Vorsicht: Verwechslungsgefahr mit dem Skalar 0).

Satz 10. Für $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in \mathbb{R}^n und Skalare c, d in \mathbb{R} gelten die folgenden Rechenregeln:

- *Kommutativität der Addition:*

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}.$$

- *Assoziativität der Addition:*

$$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}.$$

- *Assoziativität der Multiplikation mit Skalaren:*

$$c \cdot (d \cdot \mathbf{x}) = (c \cdot d) \cdot \mathbf{x}.$$

- *Distributivität von Skalarmultiplikation und Addition:*

$$c \cdot (\mathbf{x} + \mathbf{y}) = c \cdot \mathbf{x} + c \cdot \mathbf{y} \quad \text{und} \quad (c + d) \cdot \mathbf{x} = c \cdot \mathbf{x} + d \cdot \mathbf{x}$$

Beweis. Wir setzen in die Definition der Addition bzw. Skalarmultiplikation ein und benutzen die entsprechenden Rechenregeln in \mathbb{R} .

Exemplarisch führen wir dies für die Kommutativität der Addition durch, die anderen Regeln lassen sich genau gleich beweisen. Sei $\mathbf{x} = (x_1, \dots, x_n)$ und $\mathbf{y} = (y_1, \dots, y_n)$. Dann ist

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, \dots, x_n + y_n) = (y_1 + x_1, \dots, y_n + x_n) = \mathbf{y} + \mathbf{x}. \quad \square$$

Bemerkung. Die letzten beiden Gesetze sind eventuell etwas verwirrend, da für Addition in \mathbb{R} und in \mathbb{R}^n sowie für Multiplikation in \mathbb{R} und Skalarmultiplikation in \mathbb{R}^n dieselben Symbole benutzt wurden. Manchmal wird aus diesem Grund in der Literatur auch \oplus für Vektoraddition und \odot für Skalarmultiplikation geschrieben. Die Assoziativität der Multiplikation mit Skalaren wird dann zu

$$c \odot (d \odot \mathbf{x}) = (c \cdot d) \odot \mathbf{x}$$

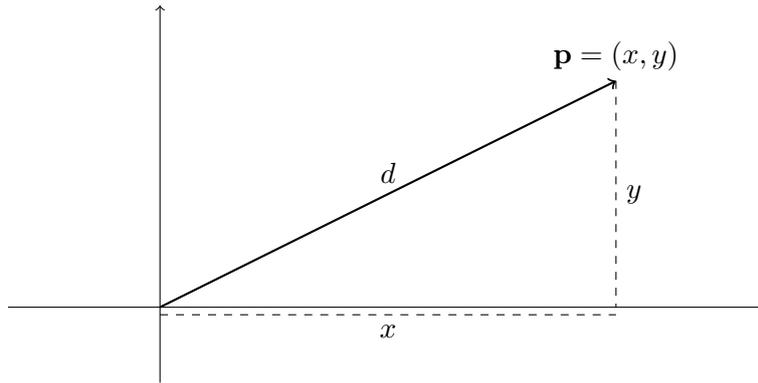


Abbildung 4.5: Abstand eines Punktes zum Ursprung in \mathbb{R}^2

und die Distributivität von Skalarmultiplikation und Addition wir dann zu

$$c \odot (\mathbf{x} \oplus \mathbf{y}) = c \odot \mathbf{x} \oplus c \odot \mathbf{y} \quad \text{und} \quad (c + d) \odot \mathbf{x} = c \odot \mathbf{x} \oplus d \odot \mathbf{x}$$

Man beachte weiter, dass bei Ausdrücken, wo Additionen von Vektoren und Skalare Multiplikationen vorkommen, die Regel „Punkt vor Strich“ gilt.

Abstand zweier Punkte

Man betrachte einen Punkt $\mathbf{p} = (x, y)$ in \mathbb{R}^2 . Gesucht ist der Abstand d von \mathbf{p} zum Ursprung, also die Länge der Verbindungsstrecke von $\mathbf{0}$ zu \mathbf{p} . Die Situation ist in Abbildung 4.5 dargestellt.

Zur Lösung dieses Problems können wir den Satz von Pythagoras benutzen: Wenn a , b und c die Seitenlängen in einem rechtwinkligen Dreieck (wobei c die Länge der Hypotenuse ist) sind, so gilt

$$a^2 + b^2 = c^2.$$

Somit gilt für den Abstand d von \mathbf{p} zum Ursprung

$$d^2 = x^2 + y^2.$$

Also muss d entweder $\sqrt{x^2 + y^2}$ oder $-\sqrt{x^2 + y^2}$ sein. Da Abstände immer positiv sind, kommt nur ersteres in Frage.

Betrachten wir nun dasselbe Problem für einen Punkt $\mathbf{p} = (x, y, z)$ im dreidimensionalen Raum \mathbb{R}^3 (siehe Abbildung 4.6).

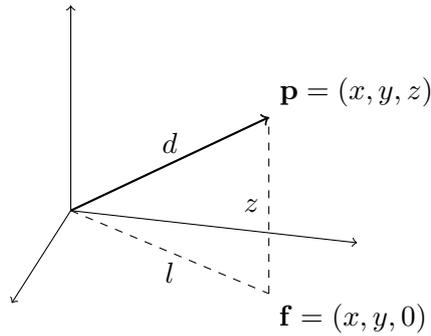


Abbildung 4.6: Abstand eines Punktes zum Ursprung in \mathbb{R}^3

Hier können wir den Satz von Pythagoras nicht direkt anwenden, da wir die Länge l der Strecke von 0 nach $\mathbf{f} = (x, y, 0)$ nicht kennen. Aber die Länge l dieser Strecke kann man ebenfalls mittels des Satzes von Pythagoras bestimmen:

$$l^2 = x^2 + y^2.$$

Somit ist $l = \sqrt{x^2 + y^2}$ und wir erhalten

$$d^2 = l^2 + z^2 = x^2 + y^2 + z^2,$$

beziehungsweise

$$d = \sqrt{x^2 + y^2 + z^2}.$$

Diese Vorgehensweise lässt sich auf \mathbb{R}^n für beliebiges n verallgemeinern: wenn $\mathbf{p} = (x_1, \dots, x_n)$ ein Punkt ist, dann können wir seinen Abstand zum Ursprung definieren als $\sqrt{x_1^2 + \dots + x_n^2}$. Außerdem kann man den Abstand zweier beliebiger Punkte $\mathbf{p} = (x, y, z)$ und $\mathbf{q} = (u, v, w)$ in \mathbb{R}^3 mit derselben Vorgehensweise berechnen. Man erhält dann

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(x - u)^2 + (y - v)^2 + (z - w)^2}.$$

Das passt auch mit der Interpretation von $\mathbf{p} - \mathbf{q}$ als der in den Ursprung verschobene Pfeil von \mathbf{q} nach \mathbf{p} zusammen.

Als Norm bezeichnet man in der Mathematik eine Funktion, die jedem Vektor eine Größe (bzw. „Länge“) zuordnet, und gewisse Standardeigenschaften besitzt. Der Abstand eines Vektors \mathbf{v} zum Ursprung ist eine solche Größenzuordnung, da dieser Abstand in gewisser Weise der Länge von \mathbf{v} entspricht. Die daraus entstehende Norm wird als euklidische Norm bezeichnet.

Definition. Die (euklidische) Norm eines Vektors $\mathbf{v} = (v_1, \dots, v_n)$ ist definiert als

$$\|\mathbf{v}\| = \sqrt{v_1^2 + \dots + v_n^2}.$$

Beispiel. Mit Hilfe der Norm können wir gewisse geometrische Objekte sehr elegant beschreiben. Die Oberfläche einer Kugel mit Mittelpunkt in $\mathbf{0} = (0, 0, 0)$ und Radius r im \mathbb{R}^3 ist zum Beispiel gerade die Lösungsmenge der Gleichung

$$\|\mathbf{v}\| = r, \quad \mathbf{v} \in \mathbb{R}^3.$$

Alternativ schreibt man diese Gleichung auch oft als

$$v_1^2 + v_2^2 + v_3^2 = r^2.$$

Ein Kreis vom Radius r um $\mathbf{0} = (0, 0)$ im \mathbb{R}^2 mit Koordinaten x und y ist auf ganz analoge Weise die Lösungsmenge der Gleichung

$$x^2 + y^2 = 0.$$

Das Skalarprodukt

In diesem Abschnitt werden wir eine weitere Verknüpfung zwischen Vektoren einführen, deren Ergebnis immer ein Skalar, also eine reelle Zahl ist.

Definition. Seien $\mathbf{x} = (x_1, \dots, x_n)$ und $\mathbf{y} = (y_1, \dots, y_n)$ zwei Vektoren aus dem \mathbb{R}^n . Dann ist das *Skalarprodukt* von \mathbf{x} und \mathbf{y} definiert als

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = \sum_{i=1}^n x_i y_i.$$

Beispiel. Sei $\mathbf{x} = (1, 3, 2)$ und $\mathbf{y} = (-2, 1, 1)$, dann ist das Skalarprodukt

$$\langle \mathbf{x}, \mathbf{y} \rangle = 1 \cdot (-2) + 3 \cdot 1 + 2 \cdot 1 = 3.$$

Bemerkung. Aus der Definition von Norm und Skalarprodukt ist sofort klar, dass

$$\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle.$$

Satz 11 (Eigenschaften des Skalarprodukts). *Seien $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ und $c \in \mathbb{R}$. Dann gilt*

1. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ und Gleichheit gilt genau dann, wenn \mathbf{x} der Nullvektor ist.
2. $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$.
3. $\langle c \cdot \mathbf{x}, \mathbf{y} \rangle = c \cdot \langle \mathbf{x}, \mathbf{y} \rangle$.
4. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.

Beweis. Durch Einsetzen in die Definition des Skalarprodukts. □

Aus diesen Eigenschaften kann man Zusammenhänge zwischen Skalarprodukten zweier Vektoren und dem Winkel, den diese Vektoren einschließen herausarbeiten.

Satz 12. *Seien \mathbf{x} und \mathbf{y} zwei Vektoren aus \mathbb{R}^2 , die einen rechten Winkel einschließen. Dann ist*

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0.$$

Beweis. Wenn wir \mathbf{x} und \mathbf{y} als Katheten eines rechtwinkligen Dreiecks betrachten, dann ist die Länge der Hypotenuse dieses Dreiecks gegeben durch $\|\mathbf{y} - \mathbf{x}\|$. Wir wissen aufgrund des Satzes von Pythagoras, dass

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2.$$

Andererseits können wir die Norm folgendermaßen berechnen:

$$\begin{aligned} \|\mathbf{y} - \mathbf{x}\|^2 &= \langle \mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x} \rangle \\ &= \langle \mathbf{y}, \mathbf{y} - \mathbf{x} \rangle - \langle \mathbf{x}, \mathbf{y} - \mathbf{x} \rangle \\ &= \langle \mathbf{y} - \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{y} - \mathbf{x}, \mathbf{x} \rangle \\ &= \langle \mathbf{y}, \mathbf{y} \rangle - \langle \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{x} \rangle \\ &= \|\mathbf{y}\|^2 - 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{x}\|^2 \end{aligned}$$

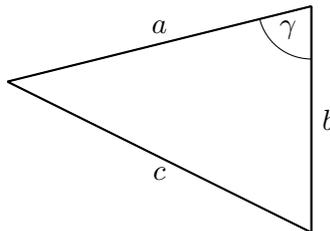
Der Term $-2\langle \mathbf{x}, \mathbf{y} \rangle$ muss also 0 sein, und somit gilt $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. □

Der obige Satz ist ein Spezialfall folgender (allgemeineren) Aussage:

Satz 13. Seien \mathbf{x} und \mathbf{y} zwei Vektoren aus der Ebene \mathbb{R}^2 und sei α der Winkel zwischen \mathbf{x} und \mathbf{y} . Dann gilt

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos \alpha.$$

Beweis. Wir sehen \mathbf{x} und \mathbf{y} wieder als Seiten eines Dreiecks an. Aus der Schule sollte der Cosinussatz bekannt sein: Für beliebige Dreiecke



gilt

$$c^2 = a^2 + b^2 - 2ab \cos \gamma.$$

Also ergibt der Cosinussatz in unserer Situation

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2 \cdot \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos \alpha.$$

Andererseits gibt dieselbe Rechnung wie im obigen Beweis

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{y}\|^2 - 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{x}\|^2$$

Durch Gleichsetzen erhalten wir also

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos \alpha.$$

Somit ist der Satz bewiesen. □

Bemerkung. In der linearen Algebra wird häufig der umgekehrte Weg gewählt: Man definiert Winkel über das Skalarprodukt und zeigt dann, dass sie beispielsweise im \mathbb{R}^2 mit der anschaulichen Definition von Winkeln übereinstimmen.

4.2 Geraden im \mathbb{R}^2

Normalform

Eine Gerade in \mathbb{R}^2 mit den Koordinaten x und y ist die Lösungsmenge einer linearen Gleichung in den Variablen x und y . Die allgemeine Form einer solchen linearen Gleichung ist

$$ax + by + c = 0, \quad (4.1)$$

wobei a, b und c vorgegebene reelle Zahlen sind, und a und b nicht gleichzeitig gleich 0 sind (was passiert in dem Fall?). Da sich die Lösungsmenge der Gleichung (4.1) nicht ändert, wenn man beide Seiten mit einer von Null verschiedenen Konstante multipliziert, sind die Parameter a, b und c durch die Gerade nur bis auf ein gemeinsames Vielfaches aus $\mathbb{R} \setminus \{0\}$ bestimmt: zum Beispiel haben die Gleichungen

$$3x + 4y + 2 = 0 \quad \text{und} \quad \frac{3}{2}x + 2y + 1 = 0$$

dieselbe Lösungsmenge, definieren also dieselbe Gerade

$$g = \{(x, y) \in \mathbb{R}^2 \mid 3x + 4y + 2 = 0\} = \{(x, y) \in \mathbb{R}^2 \mid \frac{3}{2}x + 2y + 1 = 0\}.$$

Die Form der Geradengleichungen in (4.1) wird als *parameterfreie Form* oder *Normalform* bezeichnet.

In der Schule haben Sie Geraden zum Beispiel als Graphen linearer Funktionen kennengelernt. Diese haben dann eine Gleichung der Form

$$y = mx + d,$$

welche natürlich problemlos auf Normalform gebracht werden kann. Umgekehrt kann eine Geradengleichung der Form (4.1) in diese Gestalt gebracht werden, falls $b \neq 0$.

Beispiel. Die Gerade mit der Gleichung $5x + 3y - 3 = 0$ kann durch Äquivalenzumformungen auf die Form $y = -\frac{5}{3}x + 3$ gebracht werden:

$$\begin{array}{rcl} 5x + 3y - 3 = 0 & & | - (5x - 3) \\ 3y = -5x + 3 & & | : 3 \\ y = -\frac{5}{3}x + 1. & & \end{array}$$

Parameterform

Eine weitere Möglichkeit, eine Gerade anzugeben, bietet die sogenannte *Parameterform*. Hierfür definieren wir zwei lineare Funktionen $x(t)$ und $y(t)$, die beide von derselben Variablen t wie folgt abhängen:

$$\begin{aligned}x(t) &= tm_x + d_x \\y(t) &= tm_y + d_y.\end{aligned}$$

Dabei sind m_x, d_x, m_y und d_y vorgegebene Werte und m_x, m_y sind nicht gleich null. Die Gerade ist dann die Menge

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

In anderen Worten erhalten wir für jeden Wert von t einen Punkt $(x(t), y(t))$ auf der Geraden und zu jedem Punkt auf der Geraden einen zugehörigen Wert für t . Die Variable t von der sowohl x als auch y abhängen wird als Parameter bezeichnet.

Bemerkung. Wenn eine Gerade in Parameterform gegeben ist, dann ist $\mathbf{p} = (d_x, d_y)$ immer ein Punkt auf der Geraden (setze $t = 0$). Außerdem ist leicht zu sehen, dass sich zwei beliebige Punkte auf der Geraden nur um ein Vielfaches des Vektors $\mathbf{r} = (m_x, m_y)$ unterscheiden. Dieser Vektor gibt somit die Richtung der Geraden an, und wird deshalb häufig als ein *Richtungsvektor der Geraden* bezeichnet. Man schreibt Geraden in Parameterform auch oft als

$$g = \{\mathbf{p} + t \cdot \mathbf{r} \in \mathbb{R}^2 \mid t \in \mathbb{R}\},$$

wobei \mathbf{p} und \mathbf{r} ein gegebener Punkt und ein gegebener Vektor sind. Folgende Graphik soll die Situation veranschaulichen:

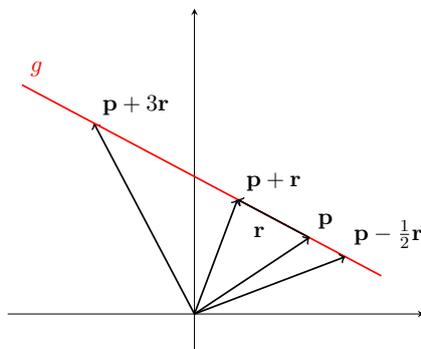


Abbildung 4.7: Parameterform einer Geraden

Die folgenden Beispiele demonstrieren, wie man die parameterfreie Form und die Parameterform ineinander überführen kann.

Beispiel. Wir betrachten die Gerade, die durch $x(t) = 2t - 3$ und $y(t) = t - 2$ definiert wird. Um diese Gerade in parameterfreie Form zu bringen, müssen wir den Parameter t eliminieren. Wir formen also beispielsweise die zweite Gleichung um:

$$\begin{array}{rcl} y = t - 2 & & | + 2 \\ y + 2 = t & & | \text{Seiten vertauschen} \\ t = y + 2. & & \end{array}$$

Somit haben wir t in Abhängigkeit von y dargestellt. Diese Darstellung von t können wir nun in die Gleichung für x einsetzen um den Parameter zu eliminieren:

$$x = 2(y + 2) - 3.$$

Die letzte Gleichung kann durch Äquivalenzumformungen auf die Form $x - 2y - 1 = 0$ gebracht werden.

Beispiel. Wir wollen nun die Gerade mit der Gleichung $2x + y - 3 = 0$ in Parameterform überführen. Zu diesem Zweck wählen wir $x(t)$ beliebig (aber von der richtigen Form), zum Beispiel

$$x(t) = t + 2.$$

Danach setzen wir dieses x in die Geradengleichung ein und Lösen nach y auf:

$$\begin{array}{rcl} 2(t + 2) + y - 3 = 0 & & \\ 2t + y + 1 = 0 & & | - (2t + 1) \\ y = -2t - 1 & & \end{array}$$

Somit haben wir $y(t)$ bestimmt und damit auch die Parameterform der Gerade. Aufmerksame Leser werden bereits bemerkt haben, dass die Wahl $x(t) = t + 2$ unnötig kompliziert war. Üblicherweise wählt man eine Funktion, die so wenig Rechenaufwand wie möglich macht, zum Beispiel $x(t) = t$.

Achtung. In obigem Beispiel konnte man $x(t)$ als eine beliebige Funktion der Form $x(t) = t + d$ wählen. Wenn wir aber mit einer Geradengleichung der Form $x + c = 0$ starten, so funktioniert dieser Ansatz nicht. In diesem Fall kann man aber $y(t) = t$ wählen und erhält dann als Parameterform $g = \{(-c, t) \mid t \in \mathbb{R}\}$.

Bemerkung. Die beiden hier vorgestellten Methoden um die verschiedenen Darstellungen von Geraden in der Ebene ineinander überzuführen sind relativ schnell und einfach durchzuführen. Es gibt aber noch unzählige andere Möglichkeiten die beiden Darstellungsformen in die jeweils andere zu transformieren.

Wir listen auch noch einige weitere Möglichkeiten auf, eine Gerade zu definieren:

- Für zwei verschiedene Punkte $\mathbf{p}, \mathbf{q} \in \mathbb{R}^2$ gibt es genau eine Gerade, die beide Punkte enthält. Diese Gerade wird als Verbindungsgerade von \mathbf{p} und \mathbf{q} bezeichnet. Ein Richtungsvektor ist durch $\mathbf{r} = \mathbf{q} - \mathbf{p}$ gegeben, die Parameterform kann also mittels

$$\mathbf{x}(t) = \mathbf{p} + t \cdot (\mathbf{q} - \mathbf{p})$$

beschrieben werden.

- Für einen Punkt \mathbf{p} im \mathbb{R}^2 und einen Vektor $\mathbf{n} \in \mathbb{R}^2$ gibt es genau eine Gerade, die \mathbf{p} enthält und deren Richtungsvektoren \mathbf{r} mit \mathbf{n} einen rechten Winkel einschließen. Der Vektor \mathbf{n} heißt *Normalenvektor* der Geraden. Die Situation sieht wie folgt aus:

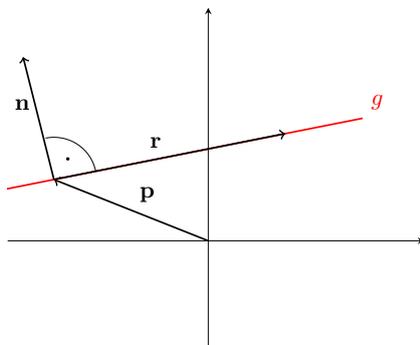


Abbildung 4.8: Parameterform einer Geraden

Zur Bestimmung dieser Geraden stellen wir zuerst fest, dass die Verbindungsstrecke zwischen je zwei Punkten auf der Gerade ein Richtungsvektor ist. Insbesondere steht für jeden Punkt \mathbf{x} auf der Gerade der Vektor $\mathbf{x} - \mathbf{p}$ senkrecht auf \mathbf{n} . Somit wissen wir, dass

$$\langle \mathbf{x} - \mathbf{p}, \mathbf{n} \rangle = 0$$

sein muss. Setzen wir die gegebenen \mathbf{p} und \mathbf{n} in die Definition des Skalarprodukts ein, so erhalten wir die Geradengleichung in Normalform.

Beispiel. Wir wollen die Gerade durch den Punkt $\mathbf{p} = (1, 3)$ mit dem Normalenvektor $\mathbf{n} = (1, 2)$ bestimmen. Dazu setzen wir einen (allgemeinen) Punkt $\mathbf{x} = (x, y)$ der Gerade in die Gleichung $\langle \mathbf{x} - \mathbf{p}, \mathbf{n} \rangle = 0$ ein. Für die linke Seite erhalten wir

$$\begin{aligned} \langle \mathbf{x} - \mathbf{p}, \mathbf{n} \rangle &= \langle (x - 1, y - 3), (1, 2) \rangle \\ &= (x - 1) \cdot 1 + (y - 3) \cdot 2 \\ &= x + 2y - 7. \end{aligned}$$

Die Gleichung der Geraden in Normalform lautet somit $x + 2y - 7 = 0$.

Schnitt zweier Geraden

Die Schnittmenge zweier Geraden g und h ist die Menge aller Punkte, die sowohl auf g als auch auf h liegen. In der Ebene kann diese Menge entweder leer sein (parallele Geraden), einen Punkt enthalten (sich schneidende Geraden) oder unendlich sein (die beiden Geraden stimmen überein). Um die Schnittmenge zu bestimmen, ist es notwendig ein lineares Gleichungssystem zu lösen. Wir werden lineare Gleichungssysteme dann später auch noch eingehender studieren.

Beispiel. Wir bestimmen den Schnittpunkt der Geraden g und h mit den Gleichungen $x + y - 2 = 0$ und $2x - y - 1 = 0$. Wir müssen also die Punkte finden, welche beide Gleichungen gleichzeitig erfüllen. Zuerst lösen wir die Geradengleichung von g nach x auf:

$$\begin{array}{rcl} x + y - 2 = 0 & & | - (y - 2) \\ x = -y + 2. & & \end{array}$$

Dieses x setzen wir in die Geradengleichung von h ein:

$$\begin{array}{rcl} 2(-y + 2) - y - 1 = 0 & & \\ -3y + 3 = 0 & & | - 3 \\ -3y = -3 & & | : (-3) \\ y = 1. & & \end{array}$$

Somit haben g und h genau einen Schnittpunkt y -Koordinate 1 und x -Koordinate $-y + 2 = -1 + 2 = 1$.

Winkel zwischen Geraden

Laut Satz 13 wissen wir, dass für den Winkel α zwischen zwei Vektoren \mathbf{x} und \mathbf{y} gilt, dass

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos \alpha.$$

Außerdem ist der Kosinus, wenn man ihn als Funktion $[0, \pi] \rightarrow [-1, 1]$ betrachtet umkehrbar, d.h. zu jedem Wert $t = [-1, 1]$ gibt es genau einen Winkel $\alpha \in [0, \pi]$ mit $\cos \alpha = t$. Er besitzt somit eine Umkehrfunktion, die mit \arccos (manchmal auch \cos^{-1}) bezeichnet wird. Mit Hilfe dieser Funktion können wir die obige Gleichung nach α auflösen:

$$\begin{array}{rcl} \langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos \alpha & & | : (\|\mathbf{x}\| \cdot \|\mathbf{y}\|) \text{ und Seiten vertauschen} \\ \cos \alpha = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} & & | \arccos \text{ auf beide Seiten anwenden} \\ \alpha = \arccos \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \right) & & \end{array}$$

Die letzte Formel kann man nun verwenden, um den Winkel zwischen zwei Vektoren zu berechnen. Der Winkel zwischen zwei Geraden ist definiert als der Winkel zwischen den jeweiligen Richtungsvektoren (falls die Geraden einander schneiden).

Beispiel. Wir wollen die Schnittmenge und gegebenenfalls den Winkel zwischen den Geraden

$$g = \{(x, y) = t \cdot (1, 2) + (3, 1) \in \mathbb{R}^2 \mid t \in \mathbb{R}\}$$

und

$$h = \{(x, y) = s \cdot (2, -1) + (1, 2) \in \mathbb{R}^2 \mid s \in \mathbb{R}\}$$

bestimmen.

Für die Schnittmenge suchen wir einen Vektor (x, y) der sowohl als $t \cdot (1, 2) + (3, 1)$ als auch als $s \cdot (2, -1) + (1, 2)$ für gewisse zu bestimmende reelle Zahlen s, t dargestellt werden kann. Dass wir in der Gleichung von h den Parameter durch s ersetzt haben hat den Grund, dass der Parameter in den beiden Darstellungen von g und h verschieden sein kann. Wir erhalten somit die Gleichung

$$t \cdot (1, 2) + (3, 1) = s \cdot (2, -1) + (1, 2),$$

welche in zwei Gleichungen aufgespaltet werden kann:

$$\begin{aligned} t + 3 &= 2s + 1 \\ 2t + 1 &= -s + 2 \end{aligned}$$

Löst man diese Gleichungen so erhält man die Lösung $t = 0$ und $s = 1$. Den Schnittpunkt erhalten wir, indem wir in der Geradengleichung von g den Parameter 0 setzen (oder in der Geradengleichung von h den Parameter 1 setzen). Das Ergebnis ist in beiden Fällen $(3, 1)$.

Die beiden Geraden haben somit genau einen Schnittpunkt und wir können den Winkel als Winkel ihrer Richtungsvektoren bestimmen:

$$\begin{aligned} \alpha &= \arccos \left(\frac{\langle (1, 2), (2, -1) \rangle}{\|(1, 2)\| \cdot \|(2, -1)\|} \right) \\ &= \arccos \left(\frac{1 \cdot 2 + 2 \cdot (-1)}{\sqrt{1^2 + 2^2} \cdot \sqrt{2^2 + (-1)^2}} \right) \\ &= \arccos \left(\frac{0}{5} \right) \\ &= \frac{\pi}{2} \end{aligned}$$

Die Geraden g und h schneiden sich also im Punkt $(3, 1)$ in einem rechten Winkel.

Bemerkung. Um herauszufinden, ob sich zwei Geraden in der Ebene in Parameterform in genau einem Punkt schneiden, reicht es zu überprüfen, dass die beiden Richtungsvektoren nicht Vielfache voneinander sind. Für die Berechnung des Winkels zwischen den Geraden ist es dann gar nicht nötig diesen Schnittpunkt konkret auszurechnen.

4.3 Ebenen im \mathbb{R}^3

Normalform und Parameterform

Wie schon in \mathbb{R}^2 kann man auch in \mathbb{R}^3 eine Gerade durch einen Punkt \mathbf{p} und einen Richtungsvektor \mathbf{r} definieren. Man erhält dann die Gerade in Parameterform:

$$g = \{\mathbf{x} = t \cdot \mathbf{r} + \mathbf{p} \in \mathbb{R}^3 \mid t \in \mathbb{R}\}$$

wobei $\mathbf{x} = (x_1, x_2, x_3)$ ein allgemeiner Punkt der Gerade und t ein reellwertiger Parameter ist. Ebenso kann man wie im zweidimensionalen Fall eine Gerade durch zwei Punkte \mathbf{p} und \mathbf{q} bestimmen indem man als Richtungsvektor die Differenz $\mathbf{q} - \mathbf{p}$ von \mathbf{p} nach \mathbf{q} wählt.

Im 3-dimensionalen Raum \mathbb{R}^3 können Geraden nicht mehr in Normalform angegeben werden. Grund dafür ist, dass die Lösungsmengen von Gleichungen der Form

$$ax + by + cz + d = 0$$

im Allgemeinen Ebenen sind. Gibt man eine Ebene durch eine solche Gleichung an spricht man von einer *Ebenengleichung* in Normalform bzw. *parameterfreier Form*.

Bemerkung. Ebenso wie bei Geraden im \mathbb{R}^2 ist der Vektor (a, b, c) hier der Normalenvektor der Ebene im Raum. Man kann also, wenn man einen Punkt \mathbf{p} auf der Ebene und ihren Normalenvektor \mathbf{n} kennt die Ebenengleichung in parameterfreier Form durch

$$\langle \mathbf{x} - \mathbf{p}, \mathbf{n} \rangle = 0$$

berechnen, wobei $\mathbf{x} = (x, y, z)$ ein allgemeiner Punkt der Ebene ist.

Des Weiteren kann man auch Ebenen in Parameterform angeben. Man benötigt dafür einen Punkt \mathbf{p} auf der Ebene und zwei von $\mathbf{0}$ verschiedene Richtungsvektoren \mathbf{r}_1 und \mathbf{r}_2 , wobei \mathbf{r}_1 kein Vielfaches von \mathbf{r}_2 ist. Die Beschreibung der Ebene E zu diesen Daten sieht dann in Parameterform folgendermaßen aus:

$$E = \{ \mathbf{x} = \mathbf{p} + t_1 \cdot \mathbf{r}_1 + t_2 \cdot \mathbf{r}_2 \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R} \}.$$

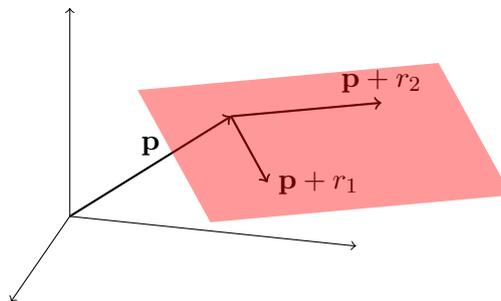


Abbildung 4.9: Parameterform einer Ebene

Bemerkung. Falls \mathbf{r}_1 ein Vielfaches von \mathbf{r}_2 ist beschreibt diese Parameterform eine Gerade.

Die Parameterform erlaubt es uns, auf relativ einfache Weise eine Ebene durch drei Punkte $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ zu bestimmen. Die benötigten Richtungsvektoren berechnen wir als $\mathbf{r}_1 = \mathbf{p}_2 - \mathbf{p}_1$ und $\mathbf{r}_2 = \mathbf{p}_3 - \mathbf{p}_1$. Sofern $\mathbf{p}_1, \mathbf{p}_2$ und \mathbf{p}_3 nicht auf einer gemeinsamen Geraden liegen ist das so berechnete \mathbf{r}_1 kein Vielfaches von \mathbf{r}_2 und wir erhalten tatsächlich eine Ebene.

Beispiel. Wir wollen die Parameter- und Normalform der Ebene E durch die Punkte $\mathbf{p}_1 = (1, -1, 1)$, $\mathbf{p}_2 = (1, 1, -1)$ und $\mathbf{p}_3 = (-1, 1, -1)$ finden.

Für die Parameterform berechnen wir

$$\mathbf{r}_1 = \mathbf{p}_2 - \mathbf{p}_1 = (1, 1, -1) - (1, -1, 1) = (0, 2, -2)$$

und

$$\mathbf{r}_2 = \mathbf{p}_3 - \mathbf{p}_1 = (-1, 1, -1) - (1, -1, 1) = (-2, 2, -2).$$

Da die Vektoren \mathbf{r}_1 und \mathbf{r}_2 nicht Vielfache voneinander sind, kann die Ebene somit in Parameterform als

$$E = \{\mathbf{x} = (1, -1, 1) + t_1 \cdot (0, 2, -2) + t_2 \cdot (-2, 2, -2) \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R}\}$$

geschrieben werden. Um in Normalform umzurechnen betrachten wir die Parameterform in jeder Koordinate:

$$\begin{aligned}x_1 &= 1 + 0t_1 - 2t_2, \\x_2 &= -1 + 2t_1 + 2t_2, \\x_3 &= 1 - 2t_1 - 2t_2.\end{aligned}$$

Aus der ersten Gleichung erhalten wir durch Äquivalenzumformungen

$$t_2 = -\frac{x_1}{2} + \frac{1}{2}.$$

Setzen wir das in die zweite Gleichung ein, so bekommen wir die Gleichung

$$x_2 = -1 + 2t_1 + 2\left(-\frac{x_1}{2} + \frac{1}{2}\right),$$

die wir durch weitere Äquivalenzumformungen auf die Form

$$t_1 = \frac{x_1}{2} + \frac{x_2}{2}$$

bringen können. Nun setzen wir die Werte für t_1 und t_2 in die dritte Gleichung ein und erhalten

$$x_3 = 1 - 2\left(\frac{x_1}{2} + \frac{x_2}{2}\right) - 2\left(-\frac{x_1}{2} + \frac{1}{2}\right).$$

Diese Gleichung kann durch Äquivalenzumformungen auf die Form

$$x_2 + x_3 = 0$$

gebracht werden. Dies ist dann gerade die Normalform der Ebene E .

4.4 Lineare Gleichungssysteme in 2 bzw. 3 Variablen

Viele der Probleme, die wir bisher in diesem Kapitel betrachtet haben, lassen sich auf ein lineares Gleichungssystem zurückführen, d.h. auf eine Menge von Gleichungen, wo die Unbekannten nur linear vorkommen. Wir beginnen mit Gleichungssystemen in 2 Variablen.

Lineare Gleichungssysteme in 2 Variablen

Beispiel. Zwei Geraden g und h seien durch die Gleichungen $x + y - 3 = 0$ und $3x - y + 2 = 0$ beschrieben. Um die Schnittpunkte zu finden, muss also die Lösungsmenge des Gleichungssystems

$$\begin{aligned}x + y - 3 &= 0 \\3x - y + 2 &= 0\end{aligned}$$

bestimmt werden.

Beispiel. Eine Ebene E sei – wie in einem früheren Beispiel – in Parameterform gegeben:

$$E = \{\mathbf{x} = (1, -1, 1) + t_1 \cdot (0, 2, -2) + t_2 \cdot (-2, 2, -2) \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R}\}.$$

Um die Normalform von E zu berechnen, betrachteten wir das linear Gleichungssystem

$$\begin{aligned}x_1 &= 1 + 0t_1 - 2t_2 \\x_2 &= -1 + 2t_1 + 2t_2 \\x_3 &= 1 - 2t_1 - 2t_2.\end{aligned}$$

Die Idee war dann das System, welches aus den ersten beiden Gleichungen besteht

$$\begin{aligned}x_1 &= 1 + 0t_1 - 2t_2 \\x_2 &= -1 + 2t_1 + 2t_2\end{aligned}$$

nach t_1 und t_2 aufzulösen, wobei x_1 und x_2 als Parameter betrachtet werden. Diese beiden Lösungen für t_1 und t_2 (welche von x_1 und x_2 abhängen) wurden dann in die dritte Gleichung

$$x_3 = 1 - 2t_1 - 2t_2$$

eingesetzt und man erhält dadurch die gewünschte Normalform (welche eine lineare Gleichung in x_1 , x_2 und x_3 ist).

Allgemein ist ein *lineares Gleichungssystem in 2 Variablen* eine Menge von linearen Gleichungen in 2 Variablen (hier x und y)

$$\begin{aligned}a_1x + b_1y + c_1 &= 0 \\a_2x + b_2y + c_2 &= 0 \\&\vdots \\a_kx + b_ky + c_k &= 0,\end{aligned}$$

wobei die a_i , b_i und c_i jeweils vorgegebene reelle Zahlen sind. Gesucht sind nun Variablenbelegungen, so dass alle Gleichungen gleichzeitig richtig sind. Jede solche Variablenbelegung heißt *Lösung* des linearen Gleichungssystems. Die Menge aller Lösungen bezeichnen wir (wie schon im Kapitel über Gleichungen) als *Lösungsmenge*. In diesem Abschnitt werden wir uns mit Methoden zur Bestimmung der Lösungsmenge von linearen Gleichungssystemen beschäftigen.

Graphische Interpretation von linearen Gleichungssystemen in 2 Variablen

Jede einzelne Gleichung in einem linearen Gleichungssystem in 2 Variablen ist die Gleichung einer Geraden in Normalform. Wenn wir nun ein solches Gleichungssystem lösen wollen, müssen wir x und y finden, die Lösungen aller Gleichungen sind, d.h. die entsprechenden Punkte (x, y) liegen auf jeder der Geraden. Ein lineares Gleichungssystem in zwei Variablen zu lösen ist somit gleichbedeutend mit der Aufgabe, den Schnitt einer Menge von Geraden im \mathbb{R}^2 zu bestimmen.

Beispiel. Die Geraden g_1 , g_2 und g_3 , gegeben durch die Gleichungen $x + 7y - 9 = 0$, $x + y - 3 = 0$ und $2x - 3y + 3 = 0$ enthalten alle den Punkt $\mathbf{p} = (2, 1)$ (siehe Abbildung 4.10). Also erfüllt $x = 2, y = 1$ alle drei Geradengleichungen und ist somit eine Lösung des linearen Gleichungssystems

$$\begin{aligned}x + 7y - 9 &= 0, \\x + y - 3 &= 0, \\2x - 3y + 3 &= 0.\end{aligned}$$

Aus der Abbildung 4.10 sieht man auch leicht, dass $\mathbf{p} = (2, 1)$ der einzige Punkt ist, welcher auf den Geraden g_1 , g_2 und g_3 liegt.

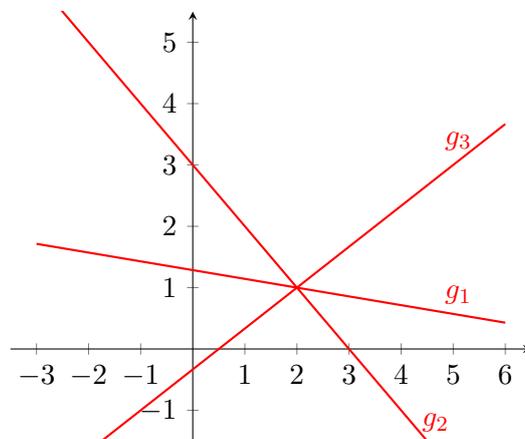


Abbildung 4.10: Interpretation der Lösung eines linearen Gleichungssystems als Schnitt von Geraden

Anhand dieser geometrischen Interpretation kann man auch leicht erkennen, wie die Lösungsmengen von linearen Gleichungssystemen in zwei Variablen aussehen:

- Falls die Geraden alle übereinstimmen, so ist jeder Punkt dieser Gerade eine Lösung des Gleichungssystems. Die Lösungsmenge ist also eine Gerade.
- Falls die Geraden nicht übereinstimmen, können sie einander in höchstens einem Punkt schneiden. Tun sie das, so hat das Gleichungssystem eine eindeutige Lösung.
- Andernfalls haben die Geraden keinen gemeinsamen Punkt (das ist zum Beispiel bei zwei parallelen Geraden und bei drei oder mehr "zufälligen" Geraden der Fall). In diesem Fall ist die Lösungsmenge des Gleichungssystems leer.

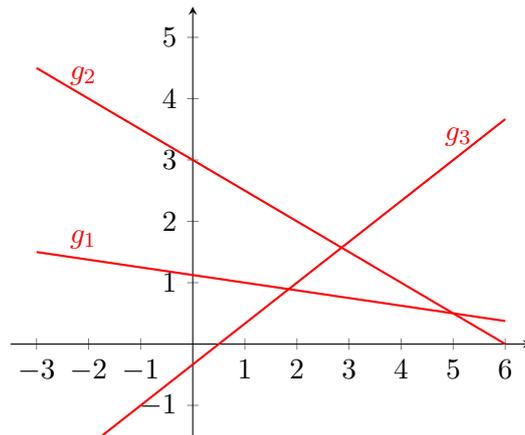


Abbildung 4.11: Typischerweise haben drei (oder mehr) Geraden in der Ebene keinen gemeinsamen Punkt, die Lösungsmenge des zugehörigen linearen Gleichungssystems ist also leer.

Lineare Gleichungssysteme in 3 Variablen

Analog zum 2-Variablen Fall, ist eine lineares Gleichungssystem in 3 Variablen eine Menge von linearen Gleichungen in 3 Variablen.

Beispiel. Um die Schnittpunkte der beiden Ebenen E_1 und E_2 mit den Gleichungen $x + y + z + 1 = 0$ und $x - y + z - 1 = 0$ zu berechnen, müssen wir alle Lösungen des linearen Gleichungssystems

$$\begin{aligned}x + y + z + 1 &= 0 \\x - y + z - 1 &= 0\end{aligned}$$

finden.

Lineare Gleichungen in 3 Variablen entsprechen genau den Ebenengleichungen in Normalform. Lineare Gleichungssysteme in 3 Variablen zu lösen ist somit gleichbedeutend damit, den Schnitt der entsprechenden Ebenen zu finden, ganz analog zu den Gleichungssystemen in 2 Variablen.

Lösung durch Substitution

Gegeben sei ein lineares Gleichungssystem in den Variablen x, y, z . Bisher haben wir solche Systeme mit ad-hoc-Methoden gelöst. Dabei folgten wir aber meistens in groben Zügen folgendem Schema:

1. Löse eine Gleichung durch Äquivalenzumformungen nach einer Variablen auf, sagen wir x .
2. Ersetze in allen anderen Gleichungen die Variable x durch das Resultat des ersten Schrittes. Dadurch erhalten wir ein Gleichungssystem mit einer Variable weniger.
3. Wiederhole die ersten beiden Schritte bis jede Variable nur noch in einer Gleichung auftritt. Löse dann diese Gleichungen. Wenn mehrere Variablen in einer Gleichung auftreten, dürfen wir alle bis auf eine durch sogenannte „freie Parameter“ ersetzen. Jede beliebige Belegung dieser Parameter mit reellen Zahlen führt zu einer Lösung des Gleichungssystems.
4. Durch Einsetzen in die Gleichungen die wir jeweils im ersten Schritt erhalten haben erhalten wir dann alle Lösungen. Auch hier gilt: gibt es Variablen in einer dieser Gleichungen, die noch nicht belegt sind, dürfen wir diese wieder als freie Parameter wählen.

Diese Methode ein lineares Gleichungssystem zu Lösen wird manchmal als Substitutionsmethode bezeichnet. Bei Systemen mit zwei oder drei Variablen ist sie oft relativ einfach anwendbar, bei größeren Systemen wird sie aber schnell unübersichtlich und daher fehleranfällig.

Beispiel. Wir lösen das Gleichungssystem

$$\begin{aligned}x + y - 1 &= 0 \\x - y - 3 &= 0.\end{aligned}$$

Zuerst formen wir die erste Gleichung zu $y = -x + 1$ um und setzen das in die zweite Gleichung ein. Somit erhalten wir $x - (-x + 1) - 3 = 0$ beziehungsweise $2x - 4 = 0$. Folglich muss $x = 2$ und $y = -x + 1 = -1$ sein.

Beispiel. Wir lösen das lineare Gleichungssystem

$$\begin{aligned}x - y + z + 1 &= 0 \\-x + y + z - 2 &= 0\end{aligned}$$

mit der Substitutionsmethode. Zuerst formen wir die erste Gleichung um:

$$\begin{array}{rcl}x - y + z + 1 = 0 & & | -x + y - 1 \\z = -x + y - 1 & & \end{array}$$

Jetzt ersetzen wir z in der zweiten Gleichung durch $-x + y - 1$ und erhalten $-x + y + (-x + y - 1) - 2 = 0$ beziehungsweise

$$-2x + 2y - 3 = 0.$$

Durch Äquivalenzumformungen sehen wir, dass $y = x + \frac{3}{2}$ sein muss. Da wir keine weiteren Gleichungen zur Verfügung haben, kommen alle Paare (x, y) mit der Eigenschaft $y = x + \frac{3}{2}$ als Lösungen in Frage. Wir ersetzen x durch den freien Parameter t , erhalten dadurch $y = \frac{3}{2} + t$ und setzen das in die Gleichung $z = -x + y - 1$ ein. Somit erhalten wir

$$z = -t + \left(\frac{3}{2} + t\right) - 1 = \frac{1}{2}.$$

Die Elemente der Lösungsmenge haben also die Form

$$\begin{aligned}x &= t \\y &= t + \frac{3}{2} \\z &= \frac{1}{2}\end{aligned}$$

Schreibt man diese Lösung als Vektor, so erhält man $(x, y, z) = t \cdot (1, 1, 0) + (0, 1, 0)$, was genau die Parameterform einer Geraden im \mathbb{R}^3 ist. Das passt auch mit unserer Interpretation von Gleichungssystemen in 3 Variablen als Schnitt von Ebenen zusammen.

Das Gauß-Verfahren

Wir betrachten das folgende lineare Gleichungssystem:

$$\begin{aligned}x + 2y - 2z &= 3 \\-y + 2z &= -1 \\z &= 1\end{aligned}$$

Dieses System ist sehr leicht zu lösen: aus der letzten Gleichung sieht man sofort, dass $z = 1$ gelten muss. Setzt man das in die zweite Gleichung ein, stellt man fest, dass $y = 3$ ist. Setzt man dann in der ersten Gleichung $y = 3$ und $z = 1$ ein, so sieht man, dass $x = -1$ ist.

Dieses System ist deshalb so leicht zu lösen, weil es in sogenannter Zeilenstufenform vorliegt: Jede Zeile enthält eine Variable, die in den nachfolgenden Zeilen nicht mehr vorkommt (ordnet man die Variablen richtig an, so ergibt sich tatsächlich ein Stufenmuster wie im obigen Beispiel).

Das Gauß-Verfahren ist eine Methode zum Lösen linearer Gleichungssysteme, die den Umstand nutzt, dass lineare Gleichungssysteme in Zeilenstufenform sehr leicht zu lösen sind. Bei dieser Methode bringen wir ein gegebenes Gleichungssystem zuerst mittels spezieller Äquivalenzumformungen (sogenannte elementare Zeilenumformungen) auf Zeilenstufenform und lösen dann dieses einfachere Gleichungssystem. Da Äquivalenzumformungen die Lösungsmenge nicht verändern, haben wir dadurch das ursprüngliche Gleichungssystem gelöst.

Die folgenden Umformungen werden hierfür benötigt:

- Äquivalenzumformungen innerhalb einer Gleichung (insbesondere Multiplikation beider Seiten mit demselben $c \neq 0$)
- Vertauschen von zwei Zeilen
- Addieren eines beliebigen Vielfachen einer Zeile zu einer anderen Zeile

Bemerkung. Substitution ist ebenfalls eine Äquivalenzumformung. Sie besteht aus zwei Teilen: zuerst werden Äquivalenzumformungen auf eine Gleichung angewandt, dann wird ein Vielfaches dieser Gleichung von jeder anderen Gleichung, die die entsprechende Variable enthält subtrahiert. In gewisser Weise sind Substitution und Gauß-Verfahren somit dasselbe Lösungsverfahren. Wie schon erwähnt ist Substitution bei einem kleinen Gleichungssystem (wenige Gleichungen/Variablen) leicht durchzuführen; je größer das System wird, desto eher lohnt sich die strukturierte Herangehensweise des Gauß-Verfahrens.

Mit Hilfe von Matrizen kann man das Gauß-Verfahren noch strukturierter durchführen. Dazu betrachten wir die sogenannte erweiterte Koeffizientenmatrix eines linearen Gleichungssystems:

Gegeben sei ein lineares Gleichungssystem der Form

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= b_1 \\ a_{21}x + a_{22}y + a_{23}z &= b_2 \\ &\vdots \\ a_{n1}x + a_{n2}y + a_{n3}z &= b_n \end{aligned}$$

Mit wesentlich weniger Schreibaufwand können die wesentlichen Informationen über dieses Gleichungssystem in einer sogenannten Matrix dargestellt werden.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & b_n \end{pmatrix}$$

Die zuvor erwähnten Äquivalenzumformungen lassen sich nun auch an dieser Matrix durchführen:

- Multiplikation aller Werte einer Zeile mit demselben $c \neq 0$.
- Vertauschen von zwei Zeilen.
- Addieren eines beliebigen Vielfachen einer Zeile zu einer anderen Zeile.

Beispiel. Wir betrachten das Gleichungssystem

$$\begin{aligned} x + 2y + 3z &= 1 \\ -x + y - 4z &= 3 \\ 2x - y + 5z &= 2 \end{aligned}$$

Die entsprechende Matrixdarstellung ist dann gegeben durch

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ -1 & 1 & -4 & 3 \\ 2 & -1 & 5 & 2 \end{pmatrix}$$

Falls wir nun z.B. die zweite Gleichung mit 2 multiplizieren, so ändert sich obige Matrix zu

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ -2 & 2 & -8 & 6 \\ 2 & -1 & 5 & 2 \end{pmatrix}$$

Falls wir nun z.B. in diesem neuen Gleichungssystem die zweite Zeile zur ersten addieren, so wird die entsprechende Matrix zu

$$\begin{pmatrix} -1 & 4 & -5 & 7 \\ -2 & 2 & -8 & 6 \\ 2 & -1 & 5 & 2 \end{pmatrix}.$$

Beispiel. Wir lösen das folgende Gleichungssystem mit Hilfe des Gauß-Verfahrens:

$$\begin{aligned} x + 2y - z &= -1 \\ x + 2y + z &= 3 \\ 2x + 3y + z &= -1 \end{aligned}$$

Zuerst ermitteln wir die entsprechende Matrix:

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

Wir betrachten die erste Spalte. Wir wollen, dass hier nur noch der erste Eintrag $\neq 0$ ist. Zu diesem Zweck subtrahieren wir die erste Zeile von der zweiten Zeile, und das doppelte der ersten Zeile von der dritten Zeile. Diese Umformungen schreibt man oft in der Form

$$\begin{pmatrix} 1 & 2 & -1 & -1 \\ 1 & 2 & 1 & 3 \\ 2 & 3 & 1 & -1 \end{pmatrix} \begin{array}{l} -I \\ -2 \cdot I \end{array} \rightsquigarrow \begin{pmatrix} 1 & 2 & -1 & -1 \\ 0 & 0 & 2 & 4 \\ 0 & -1 & 3 & 1 \end{pmatrix}$$

Als nächstes betrachten wir die Spalte mit dem niedrigsten Index, so dass es außerhalb der ersten Zeile noch Einträge $\neq 0$ gibt. Das ist in diesem Fall die zweite Spalte. Wir wollen nun, dass hier nur noch in der ersten und zweiten Zeile ein Eintrag $\neq 0$ steht. Das können wir erreichen, indem wir die zweite und dritte Zeile vertauschen.

$$\begin{pmatrix} 1 & 2 & -1 & -1 \\ 0 & 0 & 2 & 4 \\ 0 & -1 & 3 & 1 \end{pmatrix} \begin{array}{l} \\ \updownarrow \end{array} \rightsquigarrow \begin{pmatrix} 1 & 2 & -1 & -1 \\ 0 & -1 & 3 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix}$$

Das dieser Matrix entsprechende Gleichungssystem ist in Zeilenstufenform:

$$\begin{aligned}x + 2y - z &= -1 \\ -y + 3z &= 1 \\ 2z &= 4\end{aligned}$$

Aus der letzten Gleichung erhalten wir $z = 2$. Setzen wir das in die zweite Gleichung ein, so bekommen wir $y = 5$. Zuletzt setzen wir noch in der ersten Gleichung $y = 5$ und $z = 2$ ein und erhalten $x = -9$.

4.5 Relative Lage von Punkten, Geraden und Ebenen

In diesem Abschnitt werden wir darauf eingehen, wie Punkte, Geraden und Ebenen relativ zueinander im \mathbb{R}^3 liegen können. Wir werden auch jeweils eine Möglichkeit angeben, mit Hilfe eines linearen Gleichungssystems zu überprüfen, welcher Fall vorliegt. Ebenen und Geraden werden wir in Parameterform schreiben.

Punkt-Punkt Der einfachste Fall ist die Lage zweier Punkte $\mathbf{p} = (p_1, p_2, p_3)$ und $\mathbf{q} = (q_1, q_2, q_3)$ zueinander. Die beiden Punkte können entweder übereinstimmen, oder verschieden sein. Sie stimmen überein genau dann, wenn $p_1 = q_1$, $p_2 = q_2$ und $p_3 = q_3$.

Punkt-Gerade Ein Punkt \mathbf{p} kann entweder auf einer Geraden g mit Parameterform $\mathbf{x}(t) = \mathbf{q} + t \cdot \mathbf{r}$ liegen, oder nicht. Er liegt auf der Geraden genau dann, wenn das Gleichungssystem $\mathbf{p} = t \cdot \mathbf{r} + \mathbf{q}$ für die Variable $t \in \mathbb{R}$ eine Lösung besitzt.

Punkt-Ebene Ein Punkt \mathbf{p} kann entweder in einer Ebene $E = \mathbf{q} + t_1 \cdot \mathbf{r}_1 + t_2 \cdot \mathbf{r}_2$ liegen, oder nicht. Er liegt genau dann in der Ebene, wenn das Gleichungssystem $\mathbf{p} = \mathbf{q} + x \cdot \mathbf{r}_1 + y \cdot \mathbf{r}_2$ in den Variablen x und y eine Lösung besitzt.

Gerade-Gerade Zwei Geraden g und h in \mathbb{R}^3 können entweder identisch sein, oder parallel sein, oder einander in einem Punkt schneiden, oder windschief zueinander liegen. Wenn diese Geraden in Parameterform als

$$g = \{t \cdot \mathbf{r} + \mathbf{p} \in \mathbb{R}^3 \mid t \in \mathbb{R}\}$$

und

$$h = \{t \cdot \mathbf{s} + \mathbf{q} \in \mathbb{R}^3 \mid t \in \mathbb{R}\}$$

gegeben sind¹, so können die einzelnen Fälle wie folgt überprüft werden:

¹Die "interne" Variable t in der Beschreibung der Geraden ist in diesen beiden Zeilen unabhängig. Wenn man Schnittprobleme der hier auftretenden Art betrachtet, benennt man besser als erstes eine dieser beiden Variablen um, um unsinnige Fehler zu vermeiden. Dies ändert natürlich nichts an den Geraden als Punktmengen.

- g und h sind identisch, wenn \mathbf{r} ein Vielfaches von \mathbf{s} ist und \mathbf{p} auf h liegt, falls also die Gleichungssysteme

$$\mathbf{r} = c \cdot \mathbf{s} \quad (4.2)$$

und

$$\mathbf{p} = t \cdot \mathbf{s} + \mathbf{q} \quad (4.3)$$

Lösungen $c, t \in \mathbb{R}$ besitzen.

- g und h sind parallel (aber nicht identisch), wenn \mathbf{r} ein Vielfaches von \mathbf{s} ist und \mathbf{p} nicht auf h liegt, d.h. wenn (4.2) eine Lösung $c \in \mathbb{R}$ aber (4.3) keine Lösung $t \in \mathbb{R}$ besitzt.
- g und h schneiden einander in einem Punkt, wenn \mathbf{r} kein Vielfaches von \mathbf{s} ist (d.h. (4.2) hat keine Lösung $c \in \mathbb{R}$), jedoch das Gleichungssystem

$$x \cdot \mathbf{r} + \mathbf{p} = y \cdot \mathbf{s} + \mathbf{q} \quad (4.4)$$

in den beiden Variablen x und y eine Lösung besitzt. Der Punkt $x \cdot \mathbf{r} + \mathbf{p} = y \cdot \mathbf{s} + \mathbf{q}$ heißt in diesem Fall Schnittpunkt von g und h .

- Falls keine der obigen Aussagen zutrifft, d.h. falls weder (4.2) noch (4.4) Lösungen besitzen, sind g und h windschief.

Bemerkung. Das Gleichungssystem (4.4) kann auch benutzt werden, um zu überprüfen, ob die Geraden übereinstimmen. In diesem Fall ist die Lösungsmenge nämlich unendlich. Falls die Lösungsmenge aber leer ist, so können wir daraus nicht ohne zusätzliche Überlegungen erkennen, ob die Geraden parallel oder windschief sind.

Gerade-Ebene Gegeben sind eine Gerade g und eine Ebene E , jeweils in Parameterform

$$g = \{\mathbf{p} + t \cdot \mathbf{r} \in \mathbb{R}^3 \mid t \in \mathbb{R}\}$$

und

$$E = \{\mathbf{q} + t_1 \cdot \mathbf{r}_1 + t_2 \cdot \mathbf{r}_2 \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R}\}.$$

Dann kann g entweder in E liegen, oder g schneidet E in einem Punkt, oder g liegt parallel zu E . Um festzustellen, welcher Fall zutrifft, betrachten wir das lineare Gleichungssystem in drei Variablen $x, y, z \in \mathbb{R}$

$$x \cdot \mathbf{r} + \mathbf{p} = \mathbf{q} + y \cdot \mathbf{r}_1 + z \cdot \mathbf{r}_2. \quad (4.5)$$

Es ergeben sich folgende Alternativen:

- Ist die Lösungsmenge unendlich, dann liegt die Gerade g in der Ebene E .
- Existiert eine eindeutige Lösung, dann schneiden g und E einander in genau einem Punkt.
- Falls keine Lösung existiert, so liegt g parallel zu E .

Ebene-Ebene Zuletzt betrachten wir noch den Fall zweier Ebenen E und F , ebenfalls gegeben in Parameterform

$$E = \{\mathbf{p} + t_1 \cdot \mathbf{r}_1 + t_2 \cdot \mathbf{r}_2 \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R}\}$$

und

$$F = \{\mathbf{q} + t_1 \cdot \mathbf{s}_1 + t_2 \cdot \mathbf{s}_2 \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R}\}.$$

Die Ebenen können entweder identisch sein, oder parallel, oder sie schneiden einander in einer Geraden. Um festzustellen, welcher Fall zutrifft, betrachten wir das Gleichungssystem in 4 Variablen x, y, z und w

$$\mathbf{p} + x \cdot \mathbf{r}_1 + y \cdot \mathbf{r}_2 = \mathbf{q} + z \cdot \mathbf{s}_1 + w \cdot \mathbf{s}_2,$$

wobei sich die folgenden Alternativen ergeben:

- Falls die Lösungsmenge unendlich ist und durch einen reellen Parameter beschrieben werden kann, so schneiden die Ebenen einander in einer Geraden.
- Falls wir zwei Parameter benötigen um die Lösungsmenge zu beschreiben, dann sind die Ebenen identisch.
- Gibt es keine Lösung, so sind die Ebenen parallel.

Sind die Ebenen durch Gleichungen in Normalform beschrieben, d.h.

$$E = \{\mathbf{x} = (x, y, z) \in \mathbb{R}^3 \mid ax + by + cz + d = 0\}$$

und

$$F = \{\mathbf{x} = (x, y, z) \in \mathbb{R}^3 \mid ex + fy + gz + h = 0\},$$

so betrachtet man stattdessen das Gleichungssystem in drei Variablen x, y und z :

$$\begin{aligned} ax + by + cz + d &= 0 \\ ex + fy + gz + h &= 0. \end{aligned}$$

Die Alternativen sind dieselben wie oben beschrieben, mit etwas einfacheren Kriterien:

- Falls die Vektoren (a, b, c) und (e, f, g) nicht Vielfache voneinander sind, so schneiden die Ebenen einander in einer Geraden.
- Falls die Vektoren (a, b, c, d) und (e, f, g, h) Vielfache voneinander sind, dann sind die Ebenen identisch.
- Gelten beide Bedingungen nicht, sind also (a, b, c) und (e, f, g) Vielfache voneinander aber (a, b, c, d) und (e, f, g, h) nicht, so sind die Ebenen parallel.

Beispiel. Wir wollen überprüfen, ob der Punkt $\mathbf{p} = (1, 2, 1)$ in der Ebene E liegt, welche durch die Parameterform

$$E = \{t_1 \cdot (1, 2, 3) + t_2 \cdot (1, 0, 1) + (2, 0, 0) \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R}\}$$

gegeben ist. Wir stellen also das entsprechende Gleichungssystem auf:

$$\begin{aligned} 1 &= x + y + 2 \\ 2 &= 2x \\ 1 &= 3x + y \end{aligned}$$

Wir lösen dieses System durch Substitution: aus der zweiten Gleichung erhalten wir direkt $x = 1$. Setzen wir das in die erste Gleichung ein, so erhalten wir die Gleichung $1 = 1 + y + 2$, also $y = -2$. Um zu überprüfen, ob $x = 1$ und $y = -2$ tatsächlich eine Lösung des Gleichungssystems ist, müssen wir noch in die dritte Gleichung einsetzen. Hier erhalten wir $1 = 3 - 2$, eine wahre Aussage. Das Gleichungssystem besitzt also eine Lösung und somit liegt der Punkt \mathbf{p} in der Ebene E .

Beispiel. Wir bestimmen den Schnitt der Ebenen E und F , gegeben in Parameterform als

$$E = \{(0, 1, 2) + t_1(1, 1, 2) + t_2(0, 1, 3) \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R}\}$$

und

$$F = \{(1, 2, -2) + t_1(2, -1, 1) + t_2(-1, 0, 1) \in \mathbb{R}^3 \mid t_1, t_2 \in \mathbb{R}\}.$$

Wir betrachten also das Gleichungssystem $(0, 1, 2) + x(1, 1, 2) + y(0, 1, 3) = (1, 2, -2) + z(2, -1, 1) + w(-1, 0, 1)$, beziehungsweise

$$\begin{aligned} x &= 1 + 2z - w \\ 1 + x + y &= 2 - z \\ 2 + 2x + 3y &= -2 + z + w \end{aligned}$$

Wir werden dieses Gleichungssystem mit Hilfe des Gauß-Verfahrens lösen. Zuerst bringen wir es durch Äquivalenzumformungen in jeder Zeile auf die Standardform

$$\begin{aligned} x - 2z + w &= 1 \\ x + y + z &= 1 \\ 2x + 3y - z - w &= -4 \end{aligned}$$

Nun stellen wir die entsprechende Matrix auf und formen um:

$$\begin{aligned} \begin{pmatrix} 1 & 0 & -2 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 2 & 3 & -1 & -1 & -4 \end{pmatrix} &\begin{array}{l} -I \\ -2 \cdot I \end{array} \rightsquigarrow \begin{pmatrix} 1 & 0 & -2 & 1 & 1 \\ 0 & 1 & 3 & -1 & 0 \\ 0 & 3 & 3 & -3 & -6 \end{pmatrix} \begin{array}{l} \\ -3 \cdot II \end{array} \rightsquigarrow \\ &\rightsquigarrow \begin{pmatrix} 1 & 0 & -2 & 1 & 1 \\ 0 & 1 & 3 & -1 & 0 \\ 0 & 0 & -6 & 0 & -6 \end{pmatrix} \end{aligned}$$

Das Gleichungssystem, das dieser Matrix entspricht ist in Zeilenstufenform:

$$\begin{aligned}x - 2z + w &= 1 \\y + 3z - w &= 0 \\-6z &= -6\end{aligned}$$

Aus der letzten Gleichung erhalten wir sofort $z = 1$. Setzen wir das in die zweite Gleichung ein, so erhalten wir $y + 3 - w = 0$. Hier sind noch zwei Variablen vorhanden, wir dürfen eine davon als Parameter wählen. Wählen wir beispielsweise $w = t$, dann erhalten wir $y = -3 + t$. Setzen wir das in die erste Gleichung ein, so erhalten wir $x - 2 + t = 1$, was sich zu $x = 3 - t$ umformt. Da die Lösungsmenge des Gleichungssystem nur von einem Parameter abhängt, schneiden sich E und F in einer Geraden.

Wir bestimmen nun noch die Schnittgerade. Jeder Punkt im Schnitt der Ebenen E und F kann durch $(0, 1, 2) + x(1, 1, 2) + y(0, 1, 3)$ beschrieben werden, wobei x und y Werte aus der Lösung (x, y, z, w) des Gleichungssystems annehmen. Das entspricht der Parameterdarstellung

$$(0, 1, 2) + (3 - t)(1, 1, 2) + (-3 + t)(0, 1, 3),$$

beziehungsweise

$$(3, 1, -1) + t(-1, 0, 1).$$

Einsetzen von z und w in die rechte Seite des ursprünglichen Gleichungssystems hätte zum selben Ergebnis geführt.

5 Die Sprache der Mathematik

*Die Mathematiker sind eine Art Franzosen:
redet man zu ihnen, so übersetzen sie es in ihre Sprache,
und dann ist es alsobald ganz etwas anderes.*
Johann Wolfgang von Goethe, *Maximen und Reflexionen*

Um Mathematik zu kommunizieren, bedienen wir uns der natürlichen Sprache. Andererseits gibt es diverse mehr oder weniger strikte interne Regeln, welche zum Teil durchaus vom üblichen Sprachgebrauch abweichen. In diesem Kapitel wollen wir uns mit der Verwendung der Sprache in der Mathematik ausführlicher beschäftigen.

Zwei prinzipielle Bausteine der mathematischen Kommunikation sind die *Definition* und die *Aussage*.

5.1 Definitionen

Definitionen sind typischerweise Begriffsklärungen, d.h. sie führen oft einen neuen Namen für ein Konzept ein, das dann häufiger benutzt werden soll. Hier einige Beispiele, die wir zum Teil bereits gesehen hatten:

Beispiele. • Ein *Teiler* einer ganzen Zahl n ist eine ganze Zahl m , für die eine ganze Zahl k mit $n = m \cdot k$ existiert.

- Eine *Primzahl* ist eine natürliche Zahl größer als 1, welche nur sich selbst und 1 als positive Teiler besitzt.
- Eine ganze Zahl heißt *gerade*, wenn sie das Produkt einer anderen ganzen Zahl mit 2 ist.
- Drei Punkte einer Ebene heißen *kollinear*, falls es eine Gerade dieser Ebene gibt, welche alle drei Punkte enthält.
- Eine Funktion $f : U \rightarrow \mathbb{C}$ heißt *holomorph*, falls sie in jedem Punkt $z \in U$ komplex differenzierbar ist.

Wie man schon an den Beispielen sieht, enthalten Definitionen typischerweise Begriffe (in den obigen Beispielen „ganze Zahl“, „natürliche Zahl“, „Produkt“, „Punkt“, „Ebene“, „Gerade“, „Funktion“, „komplex differenzierbar“) sowie eventuell Symbole (z.B. U und \mathbb{C}), welche bereits als bekannt vorausgesetzt werden. Eine Definition führt dann darauf aufbauend *einen* neuen Begriff ein. Definitionen sind also ein schöpferischer Akt, dem oft ein längerer Denkprozess vorausgeht. Nicht selten wird eine Definition im Laufe der

Zeit mehrfach verändert, bevor sie ihre „endgültige“ Form erhält. Wichtig ist aber, dass in einem Text oder einer Vorlesung eine einmal getroffene Definition *bindend* ist, d.h. der Autor oder die Vorlesende gibt das Versprechen ab, den Begriff nur in diesem Sinne zu verwenden. Oft entstehen Definitionen aus dem Versuch, ein anschauliches Konzept mathematisch präzise zu fassen. Hat man aber die Definition erst einmal formuliert, tritt die zugrunde liegende Anschauung in den Hintergrund und *man argumentiert nur noch mit den in der Definition formulierten Eigenschaften*.

Eine andere Art der Definition führt neue Symbole ein.

Beispiele. • Sei \mathbb{N} die Menge der natürlichen Zahlen.

- Sei p_0 der kleinste Primfaktor der natürlichen Zahl $n > 1$.

Definitionen wie die letzte werden oft in Beweisen verwendet, um das dann folgende Argument effizienter formulieren zu können. Diese sind dann *lokaler Natur*, d.h. das entsprechende Symbol wird spätestens am Ende des Beweises wieder „freigegeben“, verliert also dann die verabredete Bedeutung.

5.2 Aussagen

Als *Aussage* bezeichnen wir in der Mathematik einen Satz oder eine Formel, welche entweder wahr oder falsch sein kann.

Beispiele. • $1 + 1 = 2$.

- Wenn die natürliche Zahl n gerade ist, so ist $n + 1$ ungerade.
- Es gibt unendlich viele Primzahlen.
- $5 - 7 > 0$.
- Alle Primzahlen sind ungerade.
- Jede gerade Zahl größer als 2 lässt sich als Summer zweier Primzahlen schreiben.

Hier sind die ersten zwei Aussagen wahr, die nächsten beiden falsch, und von der letzten Aussage (der sogenannten *Goldbachschen Vermutung*) ist nicht bekannt, ob sie wahr oder falsch ist. Wir müssen also den Wahrheitswert eines Satzes nicht kennen, um zu klären, ob es sich um eine Aussage handelt.

Um den Begriff zu verstehen, müssen wir aber auch abgrenzen können, welche Formulierungen und Ausdrücke nicht Aussagen in unserem Sinne sind. Nehmen wir zum Beispiel die Formel

$$a^2 + b^2 = c^2.$$

Diese weckt zwar sicher Erinnerungen, d.h. man assoziiert damit eine aus der Schule bekannte Aussage. *Die Formel ist aber für sich genommen keine Aussage in unserem Sinne*, denn es ist nicht geklärt, was für Objekte a , b und c sein sollen. Haben wir konkrete Zahlenwerte gegeben, z.B. $a = 3$, $b = 4$ und $c = 5$, so wird die Formel zu einer (im Beispiel wahren) Aussage. Auch falls wir aus dem Kontext wüssten, dass a ,

b und c die Seitenlängen eines rechtwinkligen ebenen Dreiecks sind, wobei $c > a$ und $c > b$ ist, wird die Formel zu einer bekannten *wahren Aussage*. Wird aber behauptet, die Formel gelte für die Seitenlängen a , b und c jedes beliebigen ebenen Dreiecks, oder für die konkreten Zahlen $a = 4$, $b = 7$ und $c = 8$, so wird diese Aussage *falsch*. Streng genommen handelt sich bei der Formel also um eine *Aussageform*, in der noch variable Teile (hier die „Variablen“ a , b und c) vorkommen. Werden diese genauer spezifiziert, so wird aus der Aussageform eine Aussage.

Im Rest dieses Kapitels wollen wir die Symbole A , B , C usw. jeweils abstrakt als Platzhalter für eine Aussage verwenden. Hängt eine Aussageform noch von Variablen a , b und c ab, so schreiben wir entsprechend $A(a, b, c)$.

5.3 Wie man aus Aussagen neue Aussagen gewinnt

Die Negation

Am einfachsten erhält man aus einer Aussage A eine neue Aussage, indem man sie negiert.

Definition. Die *Negation* der Aussage A ist wahr, wenn A falsch ist und falsch, wenn A wahr ist. Sie wird oft mit $\neg A$ bezeichnet.

Beispiele. Die Negation einer einfachen Aussage ist meist sehr leicht zu formulieren.

- $A =$ „3 ist eine Primzahl.“
 $\neg A =$ „3 ist keine Primzahl.“
- $A =$ „ $1 + 1 = 3$ “
 $\neg A =$ „ $1 + 1 \neq 3$ “
- $A =$ „Jedes Dreieck ist rechtwinklig.“
 $\neg A =$ „Nicht jedes Dreieck ist rechtwinklig.“

Verknüpfungen

Eine nützliche Methode, aus Aussagen neue Aussagen zu erzeugen sind die Verknüpfungen mit „und“ und „oder“.

Definition. Die Aussage „ A und B “ ist wahr, wenn sowohl A also auch B wahr sind, und sonst falsch. Sie wird oft mit $A \wedge B$ bezeichnet.

Die Aussage „ A oder B “ ist wahr, wenn mindestens eine der Aussagen A oder B wahr ist, und falsch, wenn beide falsch sind. Sie wird oft mit $A \vee B$ bezeichnet.

Mit diesen Konventionen sind dann zum Beispiel die Aussagen

$$\text{„}\neg(A \text{ und } B)\text{“} \quad \text{und} \quad \text{„}(\neg A) \text{ oder } (\neg B)\text{“} \quad (5.1)$$

gleichwertig, d.h. sie haben immer denselben Wahrheitswert (sind also entweder beide falsch oder beide wahr).

Beispiel. Wir betrachten die Aussagen A „3 ist eine Primzahl“ und B „4 ist eine Primzahl“. Dann lautet die Aussage

- $(A \text{ und } B)$ „3 und 4 sind Primzahlen.“
- $\neg(A \text{ und } B)$ „Es stimmt nicht, dass 3 und 4 Primzahlen sind.“
- $\neg A$ „3 ist keine Primzahl.“
- $\neg B$ „4 ist keine Primzahl.“
- $((\neg A) \text{ oder } (\neg B))$ „Eine der Zahlen 3 und 4 ist keine Primzahl.“

Wir sehen also, dass zumindest in diesem Beispiel die Aussagen C „ $\neg(A \text{ und } B)$ “ und D „ $(\neg A) \text{ oder } (\neg B)$ “ denselben Wahrheitswert haben.

Bevor wir fortfahren, hier noch eine **Warnung**: In der Umgangssprache wird „oder“ manchmal ohne Kennzeichnung exklusiv verwendet. In den Sätzen

„Wir gehen heute ins Kino oder ins Theater.“ oder

„Ich fahre im Urlaub nach Amerika oder nach Australien.“

wird implizit gesagt, dass nur eine der beiden Alternativen zutrifft. *In der Mathematik ist eine mit „oder“ formulierte Aussage immer einschließend gemeint.* Will man das ausschließende „oder“ formulieren, so muss man explizit *entweder A oder B* sagen.

Die Implikation

Eine in der Mathematik sehr oft verwendete Konstruktion setzt zwei gegebene Aussagen A und B zu einer neuen Aussage „Wenn A , dann B “ zusammen. Man schreibt dann formal auch einfach „ $A \implies B$ “.

Definition. Die *Implikation* $A \implies B$ ist falsch, wenn A wahr und B falsch ist, und richtig in allen anderen Fällen.

Beispiele. Wir geben einige Beispiele für Implikationen.

- Wenn 3 die Zahl 9007199254740992 teilt, dann ist die Zahl 9007199254740992 keine Primzahl.
- Wenn $x = \pi$, dann gilt $\sin(x) = 0$.
- Wenn das Dreieck D gleichschenkelig ist, dann ist D gleichseitig.
- Wenn die natürliche Zahl n ungerade ist, dann ist sie ohne Rest durch 3 teilbar.

Streng genommen sind die letzten drei Implikationen Aussageformen, denn sie enthalten noch Variablen: die Zahl x , das Dreieck D oder die natürliche Zahl n . In einem Kontext, in dem diese Objekte bereits eingeführt wurden, d.h. wenn wir schon wissen, von welcher Zahl x oder welchem Dreieck D oder welcher natürlichen Zahl n die Rede ist, dann sind unsere Beispiele tatsächlich gültige Implikationen, welche zwei nachprüfbare Aussagen verknüpfen.

In einer Implikation $A \implies B$ nennt man die Aussage A die *Voraussetzung*, und die Aussage B die *Behauptung*. Über den Wahrheitsgehalt dieser beiden Teile macht die Implikation selbst keine Aussage. Insbesondere *folgt aus einer falschen Aussage jede andere Aussage*. Zum Beispiel ist die Implikation

Wenn $-1 = 1$, dann gilt $5 = 8$.

als Implikation eine wahre Aussage, auch wenn die beiden Teile falsch sind. Wir werden bald sehen, warum dies eine sehr nützliche Konvention ist.

In mathematischen Texten werden nicht immer die Signalworte *Wenn ... , dann ...* verwendet, um Implikationen auszudrücken. Hier sind einige Beispiele - können Sie diese als „wenn ... , dann ...“-Sätze umformulieren?

Beispiele. Die folgenden Aussagen sind wahr. Wieder nehmen wir an, dass die vorkommenden Symbole bereits mit konkreten Werten belegt sind.

- Falls x eine positive reelle Zahl ist, so ist auch x^2 eine positive reelle Zahl.
- Die natürliche Zahl n ist nur dann durch 4 teilbar, wenn sie durch 2 teilbar ist.
- Damit das Viereck $ABCD$ im dreidimensionalen Raum in einer Ebene liegt, müssen sich die beiden Diagonalen AC und BD schneiden.

Auch bei der Implikation gibt es wichtige Unterschiede zur umgangssprachlichen Verwendung. Betrachten wir als Beispiel folgenden Satz, den vielleicht Eltern zu Ihrem Kind sagen (wie sinnvoll solche Ansagen sind, wollen wir hier nicht diskutieren):

Wenn Du jetzt Dein Zimmer aufräumst, dann gehen wir nachher Eis essen.

Typischerweise ist in einer solchen Situation die Aussage „*Wenn Du Dein Zimmer nicht aufräumst, dann gibt es auch kein Eis.*“ gleich mit gemeint. Streng logisch sind diese beiden Aussagen aber erst einmal unabhängig, d.h. aus der einen lässt sich die andere nicht herleiten.

Die Negation der Implikation

Wir haben bereits allgemein erklärt, wie man Aussagen negiert, und wir haben Implikationen betrachtet. Nun ist es nicht schwer, die Negation einer Implikation zu formulieren - dazu müssen wir die beiden bekannten Regeln nur verbinden. Wir erinnern uns: die Implikation „ $A \implies B$ “ ist wahr, wenn entweder A falsch ist (und B beliebig) oder auch wenn A und B beide wahr sind, und ist falsch, wenn A wahr ist und B falsch. Durch Negation erhalten wir also die neue Regel: Die negierte Implikation $\neg(A \implies B)$ ist wahr, wenn A wahr und B falsch ist, und falsch in allen anderen Fällen.

Beispiel. Für die Aussagen $A =$ „3 ist eine Primzahl“ und $B =$ „4 ist eine Primzahl“ haben wir

- $(A \implies B) =$ „Wenn 3 eine Primzahl ist, dann ist auch 4 eine Primzahl.“

- $(\neg(A \implies B)) =$ „Es stimmt nicht, dass wenn 3 eine Primzahl ist, dann auch 4 eine Primzahl sein muss.“ Schöner formuliert man das als: „Aus der Tatsache, dass 3 eine Primzahl ist, folgt nicht, dass 4 eine Primzahl ist.“

Die ursprüngliche Implikation war hier falsch, die negierte Implikation ist wahr.

Man kann sich leicht überzeugen, dass die Negation der Implikation $A \implies B$ gleichbedeutend mit der Aussage „ A und nicht B “ ist. Vergleichen wir dies noch einmal mit dem Beispiel vom Aufräumen und Eis essen, so sehen wir, dass die logische Negation der Aussage „Wenn Du jetzt Dein Zimmer aufräumst, dann gehen wir nachher Eis essen.“ die Aussage „Du räumst jetzt Dein Zimmer auf und wir gehen nachher nicht Eis essen.“ ist, aus der Sicht des Kindes keine besonders attraktive Aussicht.

Spannender (und in der Mathematik sehr wichtig) ist die Kontraposition der Implikation $A \implies B$, d.h. die Aussage $(\neg B) \implies (\neg A)$. Diese ist in der Tat äquivalent zur ursprünglichen Implikation, d.h. sie ist genau dann wahr, wenn die Implikation $A \implies B$ wahr ist, und falsch, wenn diese falsch ist. Diese Beobachtung benutzt man oft in Beweisen: statt direkt die Implikation $A \implies B$ zu zeigen, beweist man stattdessen, dass $(\neg B) \implies (\neg A)$.

Beispiele. Hier zwei Beispiele für (wahre) Implikationen und ihre Kontraposition. Welche Aussagen können Sie leichter beweisen?

- $A \implies B$: Ist eine natürliche Zahl durch 3 teilbar, so ist auch ihre Quersumme durch 3 teilbar.
 $(\neg B) \implies (\neg A)$: Ist die Quersumme einer natürlichen Zahl nicht durch 3 teilbar, so ist auch die Zahl selbst nicht durch 3 teilbar.
- $A \implies B$: Wenn das Dreieck D gleichseitig ist, dann ist es auch gleichschenkelig.
 $(\neg B) \implies (\neg A)$: Wenn das Dreieck D nicht gleichschenkelig ist, dann ist es auch nicht gleichseitig.

Bemerkung. Im Zusammenhang mit Implikationen spricht man manchmal auch von einer *notwendigen Bedingung* oder einer *hinreichenden Bedingung*: Ist die Implikation $A \implies B$ wahr, so ist B eine notwendige Bedingung für A , und A ist eine hinreichende Bedingung für B .

In unserem Alltagsbeispiel ist die Bedingung „Du räumst jetzt Dein Zimmer auf“ hinreichend für „wir gehen nachher Eis essen“ (rein logisch aber nicht notwendig). Hier sind Logik und Alltag also durchaus im Einklang. Die Kontraposition von „Wenn Du jetzt Dein Zimmer aufräumst, dann gehen wir nachher Eis essen.“ die Aussage „Wenn wir nachher nicht Eis essen gehen, dann hast Du jetzt Dein Zimmer nicht aufgeräumt.“. Auch dies ist logisch gleichwertig zur ursprünglichen Aussage.

Äquivalenz von Aussagen

Wir hatten ja bereits gesehen, dass manchmal zwei Aussagen A und B gleichwertig sind. Mit Hilfe der Implikation kann man das als „ $A \implies B$ und $B \implies A$ “ ausdrücken, wie man schnell nachprüft. Man sagt in solchen Fällen auch, die Aussagen A und B seien

äquivalent, und schreibt dafür $A \iff B$, gesprochen „A genau dann, wenn B“. Manchmal sagt man dafür auch „A dann und nur dann, wenn B“.

Beispiel. • Sei n eine natürliche Zahl. Die Zahl n ist genau dann durch 3 teilbar, wenn die Quersumme (die Summe der Ziffern von n im Dezimalsystem) durch 3 teilbar ist.

- Sei p eine natürliche Zahl größer 1. Die Zahl p ist genau dann eine Primzahl, wenn p nur die positiven Teiler 1 und p besitzt.

Unser zweites Beispiel verdeutlicht noch einmal einen wichtigen Aspekt von Definitionen: sie führen einen neuen Begriff ein, indem sie diesen als äquivalent zu einer anderen (meist längeren) Aussage erklären.¹

Unter den Äquivalenzen gibt es eine besondere Gruppe, die sogenannten *Tautologien*. Diese sind zusammengesetzte Aussagen, welche unabhängig vom Wahrheitswert der darin vorkommenden einzelnen Bestandteile *immer wahr sind*. Solche aussagenlogischen Formeln spielen bei logischen Schlussfolgerungen eine wichtige Rolle, wie wir später sehen werden, wenn wir uns mit mathematischen Beweisen auseinandersetzen.

Satz 14. Die folgenden Aussagen sind Beispiele für Tautologien, d.h. für beliebige Aussagen A , B und C immer wahr:

- A oder $\neg A$
- $\neg(A \text{ und } \neg A)$
- $\neg\neg A \iff A$
- $(A \implies B) \iff (\neg B \implies \neg A)$
- $(A \text{ und } (A \implies B)) \implies B$
- $(A \text{ und } (\neg B \implies \neg A)) \implies B$
- $((A \implies B) \text{ und } (B \implies C)) \implies (A \implies C)$
- $(A \text{ und } B) \iff (B \text{ und } A)$
 $(A \text{ oder } B) \iff (B \text{ oder } A)$
- $((A \text{ und } B) \text{ und } C) \iff A \text{ und } (B \text{ und } C)$
 $((A \text{ oder } B) \text{ oder } C) \iff A \text{ oder } (B \text{ oder } C)$
- $A \text{ und } (B \text{ oder } C) \iff (A \text{ und } B) \text{ oder } (A \text{ und } C)$
 $A \text{ oder } (B \text{ und } C) \iff (A \text{ oder } B) \text{ und } (A \text{ oder } C)$

¹In der Formulierung von Definitionen wird das „genau dann“ oft weggelassen. Dies ist aus rein logischer Sicht bedenklich, sprachlich aber für manche Ohren angenehmer, und daher die Ausnahme, die die Regel bestätigt.

- $\neg(A \text{ und } B) \iff (\neg A \text{ oder } \neg B)$
 $\neg(A \text{ oder } B) \iff (\neg A \text{ und } \neg B)$

Um diesen Satz zu beweisen, muss man für jeden einzelnen Punkt überprüfen, dass tatsächlich die angegebenen Verknüpfungen für alle möglichen Wahrheitswerte der Aussagen A , B und C jeweils äquivalent sind. Für die meisten dieser Aussagen ist dies auch nicht schwer, wenn man sie in Alltagssprache übersetzt und in Ruhe darüber nachdenkt. Eine Möglichkeit, einen systematischen Beweis zu geben, ist das Aufstellen von sogenannten *Wahrheitstabellen*. Wir betrachten als Beispiel die fünfte Zeile unserer Liste:

$$(A \text{ und } (A \implies B)) \implies B.$$

Jede der beiden Aussagen A oder B kann wahr oder falsch sein, wir erhalten also 4 mögliche Kombinationen. In allen diesen Fällen können wir nun für die einzelnen Bestandteile der gesamten Aussage den Wahrheitswert bestimmen. Durch Auflösung der Verknüpfungen erhalten wir schließlich den Wahrheitswert unserer Gesamtaussage:

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

In der rechten Spalte sehen wir, dass die zu beweisende Aussage in allen Fällen wahr ist. Dies entspricht genau der Behauptung, dass es sich um eine Tautologie handelt.

Die Quantoren „Für alle“ und „Es gibt“

Wir haben bereits Aussagen des folgenden Typs erwähnt:

- Für jede natürliche Zahl n teilt 3 das Produkt $n(n + 1)(n + 2)$.
- Ist p eine Primzahl, so ist $p + 1$ gerade.
- Es existiert ein ebenes Dreieck, dessen Innenwinkelsumme größer als 180° ist.
- Es existiert eine natürliche Zahl n , die kleiner ist als jede andere natürliche Zahl.

Diesen Aussagen enthalten wieder eine Aussageform $A(x)$, welche noch eine Variable x enthält – in unseren Beispielen die natürliche Zahl n oder die Primzahl p oder das ebene Dreieck. Statt wie vorher implizit anzunehmen, dass diese Symbole aus dem Kontext schon erklärt sind, werden die Aussageformen hier zu einer Aussage gemacht, indem mit Hilfe eines der Quantoren „Für alle“ oder „Es gibt“ die zulässigen Werte der Variablen spezifiziert werden. In logischen Formeln benutzt man die Symbole \forall und \exists für diese beiden Quantoren, also:

$$\forall x \in X \dots \quad \text{bedeutet} \quad \text{„Für alle } x \in X \dots\text{“}$$

und

$\exists x \in X \dots$ bedeutet „Es existiert ein $x \in X \dots$ “,

In unserem Beispiel ist die erste Aussage wahr. Die zweite Aussage ist falsch, wie das Gegenbeispiel $p = 2$ zeigt. Die dritte Aussage ist ebenfalls falsch, denn für jedes ebene Dreieck ist die Innenwinkelsumme bekanntlich gleich 180° . Die vierte Aussage ist wieder wahr, denn 1 ist eine Zahl mit der verlangten Eigenschaft.

Wir betrachten zum Schluss noch ein weiteres Beispiel:

Für jede natürliche Zahl $n < 0$ ist $n + 1$ irrational.

Wir verlangen hier von n zwei Dinge: es soll eine natürliche Zahl sein, und es soll negativ sein. Da diese Bedingungen nicht beide gleichzeitig erfüllbar sind, ist die Menge der zulässigen Werte für n die *leere Menge*.

In der Mathematik sind alle Aussagen der Form $\forall x \in M : A(x)$ wahr, wenn M die leere Menge ist.

Insbesondere ist also auch die gerade erwähnte Aussage wahr. Dies ist sehr nützlich und auch sinnvoll, da es ja nichts gibt, was der Aussage widerspricht. Allerdings weicht diese Konvention vom Gebrauch in der Umgangssprache ab. Wenn Sie zum Beispiel sagen

„Ich habe in meiner Geldbörse nur 1-Euro-Münzen.“,

so werden die meisten Menschen davon ausgehen, dass Sie momentan tatsächlich mehrere dieser Münzen besitzen. Nur mit den Gesetzen der Logik lässt sich dies aus Ihrer Aussage aber nicht schließen, d.h. innerhalb der Mathematik ist dieser Satz auch wahr, falls Sie *gar keine Münzen* oder *nur eine 1-Euro-Münze* in Ihrer Geldbörse haben.

Negation von Aussagen mit Quantoren

Wie negiert man nun Aussagen, welche Quantoren enthalten? Ein kurzer Blick auf die Wahrheitswerte zeigt: die Negation der Aussage „Für alle x gilt die Aussage $A(x)$.“ ist die Aussage „Es gibt ein x , so dass die Aussage $A(x)$ nicht gilt.“. Analog ist die Negation der Aussage „Es existiert ein x , so dass die Aussage $A(x)$ gilt.“ die Aussage „Für alle x gilt die Aussage $A(x)$ nicht.“ In Formeln sieht das so aus

$$\begin{aligned}\neg(\forall x : A(x)) &\iff \exists x : \neg A(x) \\ \neg(\exists x : A(x)) &\iff \forall x : \neg A(x)\end{aligned}$$

Beispiele. Im Folgenden ist jeweils die ursprüngliche Aussage wahr und die Negation falsch.

- Die Negation der Aussage „Jede natürliche Zahl größer als 1 ist durch eine Primzahl teilbar.“ ist die Aussage „Es gibt eine natürliche Zahl größer als 1, die durch keine Primzahl teilbar ist.“
- Die Negation der Aussage „Es gibt reelle Zahlen, welche nicht rational sind.“ ist die Aussage „Jede reelle Zahl ist rational.“

Dies hat unmittelbare praktische Bedeutung: Um eine \forall -Aussage zu widerlegen, genügt es, ein einziges Gegenbeispiel zu finden. Um hingegen eine \exists -Aussage zu widerlegen, muss man etwas über *alle* betrachteten Objekte zeigen.

Man kann Quantoren auch kombinieren, und erhält somit schnell relativ komplexe Aussagen.

Beispiele. Welche der folgenden Aussagen sind wahr?

- Zu jeder natürlichen Zahl n gibt es eine natürliche Zahl m mit $m > n$.
- Es gibt eine natürliche Zahl n , so dass für jede natürliche Zahl m gilt, dass $m > n$.
- Für jede natürliche Zahl n und jede natürliche Zahl m gilt $m > n$.
- Es gibt eine natürliche Zahl n , so dass eine natürliche Zahl m mit $m > n$ existiert.

Bei der Negation solcher Verkettungen von Quantoren geht man relativ mechanisch vor, indem man von „außen“ nach „innen“ die obigen Regeln anwendet:

$$\begin{aligned} \neg(\forall x \exists y \forall z : A(x, y, z)) & \text{ ist äquivalent zu } \exists x : \neg(\exists y \forall z : A(x, y, z)) \\ & \text{ und äquivalent zu } \exists x \forall y : \neg(\forall z : A(x, y, z)) \\ & \text{ und äquivalent zu } \exists x \forall y \exists z : \neg(A(x, y, z)) \end{aligned}$$

Konkret sieht das dann so aus.

Beispiel. Die Negation der Aussage

„Für alle $x \in (a, b)$ und für jedes $\epsilon > 0$ existiert ein $\delta > 0$ so dass für alle $y \in (a, b)$ mit $|x - y| < \delta$ gilt, dass $|f(x) - f(y)| < \epsilon$.“

ist die Aussage

„Es gibt ein $x \in (a, b)$ und ein $\epsilon > 0$, so dass für jedes $\delta > 0$ ein $y \in (a, b)$ mit $|x - y| < \delta$ existiert, für das $|f(x) - f(y)| \geq \epsilon$ gilt.“

Wie immer nehmen wir an, dass die auftretenden Symbole im Kontext bereits hinreichend erklärt wurden. Wie Sie in der Analysis sehen werden, beschreibt die Aussage die Stetigkeit einer Funktion $f : (a, b) \rightarrow \mathbb{R}$ auf dem gesamten Intervall $(a, b) \subseteq \mathbb{R}$. Die Negation formuliert also die Bedingung „ f ist in (a, b) nicht (überall) stetig“.

6 Abbildungen und Funktionen

Der Begriff der Abbildung ist absolut zentral für die Mathematik, und Funktionen bilden eine wichtige Beispielklasse unter den Abbildungen. Intuitiv geht es darum, Zusammenhänge zwischen verschiedenen variablen Objekten oder „Größen“ auszudrücken.

Wir beginnen mit einigen Beispielen von Zuordnungen, deren Gemeinsamkeiten wir danach formalisieren wollen.

Beispiele. (a) Zu einem festen Zeitpunkt können wir jedem Punkt auf der Erdoberfläche den in diesem Moment dort herrschenden Luftdruck zuordnen. Alternativ können wir auch in einem festen Messpunkt in einer gewissen Zeitspanne den Luftdruck aufzeichnen. Schließlich können wir die beiden Alternativen auch kombinieren, und jedem Paar (Zeitpunkt, Ort) den gemessenen Luftdruck zuordnen.

(b) Jedem Auto mit Hamburger Kennzeichen können wir zu jedem Zeitpunkt innerhalb der nächsten Stunde seine Geschwindigkeit zuordnen.

(c) Jedem Dreieck in einer Ebene können wir seinen Schwerpunkt zuordnen.¹

(d) Jedem Zeitpunkt in einer gewissen Zeitspanne können wir die in diesem Moment auf einer vorher ausgewählten (Digital-)Uhr angezeigte Uhrzeit zuordnen.

(e) Einem Paar natürlicher Zahlen a und b kann man ihre Summe $a + b$ zuordnen.

6.1 Etwas mehr über Mengen

Wir hatten bereits am Anfang des Vorkurses kurz etwas über Mengen gesagt. Zur Erinnerung, hier noch einmal die Definition.

Definition. Eine *Menge* ist eine Sammlung wohlunterscheidbarer Objekte. Die einzelnen Bestandteile einer Menge nennt man die *Elemente der Menge*.

Bisher waren unsere Beispiele für Mengen vor allem Mengen von Zahlen. In der Mathematik fasst man aber auch viele andere Objekte zu Mengen zusammen. In den einführenden Beispielen oben tauchten zum Beispiel die folgenden Mengen auf:

- in (a) die Menge der Punkte der Erdoberfläche, die Menge der möglichen Werte für den Luftdruck, und auch die Menge der Zeiten in einer gewissen Zeitspanne. Nach Wahl von Maßeinheiten können die letzten beiden Mengen mit Teilmengen der reellen Zahlen identifiziert werden, die erste hingegen scheint etwas komplizierter.

¹Stellt man sich das Dreieck aus einem homogenen Material gefertigt vor, so ist der Schwerpunkt derjenige Punkt, an dem man es auf einem spitzen Stock balancieren kann.

- in (b) die Menge der Autos mit Hamburger Kennzeichen und die Menge der Zeiten innerhalb der nächsten Stunde sowie die Menge der möglichen Geschwindigkeiten.
- in (c) die Menge der Dreiecke der gegebenen Ebene sowie die Menge der Punkte in dieser Ebene (die ja alle als Schwerpunkt auftreten können).
- in (d) die Menge der Zeiten in einer gewissen Zeitspanne (die wir uns wieder als Intervall in \mathbb{R} vorstellen), und die Menge der von unserer Digitaluhr anzeigbaren Uhrzeiten (diese Menge ist endlich).
- in (e) die Menge der Paare natürlicher Zahlen sowie die Menge der natürlichen Zahlen.

Bei manchen dieser Mengen ist es unpraktisch, alle Elemente aufzuzählen. Oft kann man stattdessen die Elemente der Menge durch ihre Eigenschaften charakterisieren. Wenn wir zum Beispiel die Menge aller Primzahlen, welche kleiner als 10^{30} sind angeben wollen, so können wir dafür

$$\{p \in \mathbb{N} \mid p \text{ ist Primzahl und } p < 10^{30}\}$$

schreiben. Analog schreiben wir

$$\{\text{Symbole der Form } h : m \mid h, m \in \mathbb{N}_0, 0 \leq h \leq 23, 0 \leq m \leq 59\}$$

für die Menge der auf unserer Digitaluhr anzeigbaren Uhrzeiten.² Manchmal benutzt man auch statt des Trennstrichs | einen Doppelpunkt, wenn es die Lesbarkeit erhöht. Zum Beispiel bezeichnet

$$\{x \in \mathbb{R} : |x| < 100\}$$

die Menge der reellen Zahlen, deren Betrag kleiner als 100 ist, also das Intervall $(-100, 100)$.

Man kann aus einer oder mehreren Mengen auf verschiedene Art neue Mengen konstruieren. Sei zum Beispiel F die Menge aller Früchte an einem Verkaufsstand eines Wochenmarktes. Wir definieren

$$A = \{x \in F \mid x \text{ ist gelb}\}$$

als die Teilmenge der gelben Früchte, und

$$B = \{x \in F \mid x \text{ ist ein Apfel}\}$$

als die Teilmenge der Äpfel. Der *Durchschnitt* $A \cap B$ der Mengen A und B ist die Menge aller Objekte, welche in beiden Mengen enthalten sind, also hier

$$A \cap B = \{x \in F \mid x \text{ ist ein gelber Apfel}\}.$$

Zwei Mengen A und B heißen *disjunkt*, wenn ihr Durchschnitt $A \cap B$ die leere Menge ist. Dies könnte in unserem Beispiel der Fall sein, wenn der Marktstand keine gelben Äpfel im Angebot hat.

²Wenn wir genauer sein wollen, könnten wir noch dazuschreiben, dass die Zahlen stets zweistellig, notfalls mit 0 beginnend, dargestellt sein sollen.

Die *Vereinigung* $A \cup B$ der Mengen A und B ist die Menge aller Objekte, welche in mindestens einer der Mengen enthalten sind. In unserem Beispiel ist dies

$$A \cup B = \{x \in F \mid x \text{ ist gelb oder ein Apfel}\}.$$

Die *Differenz* $A \setminus B$ zweier Mengen ist die Menge aller Objekte, welche in A aber nicht in B enthalten sind. In unserem Beispiel ist dies

$$A \setminus B = \{x \in F \mid x \text{ ist gelb, aber kein Apfel}\}.$$

Um Mengen anschaulich darzustellen, verwenden wir manchmal abstrakte Diagramme, in denen die einzelnen Mengen als ebene Figuren dargestellt sind. Überlappende Regionen stellen dabei Durchschnitte von Mengen dar.

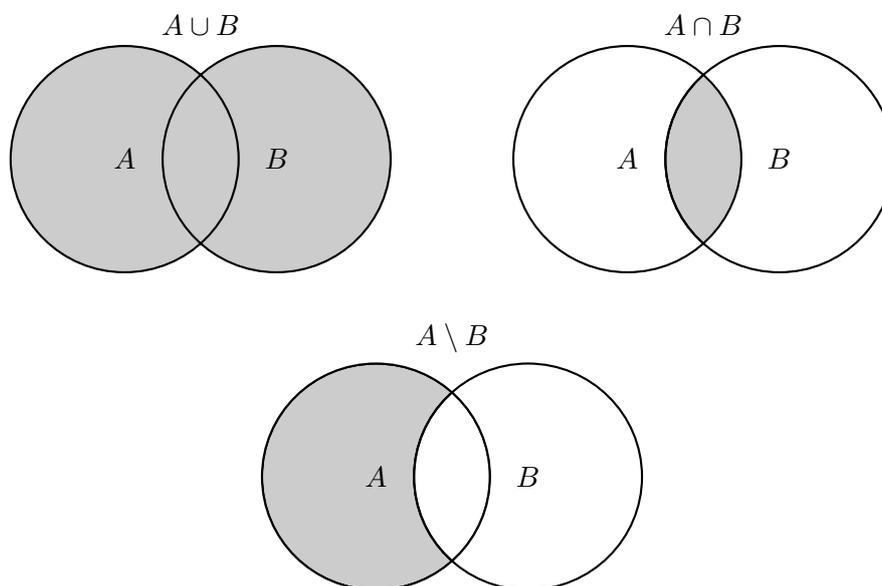


Abbildung 6.1: Vereinigung, Durchschnitt und Differenz von Mengen A und B

Eine andere nützliche Konstruktion einer neuen Menge aus einer bekannten Menge M ist die *Menge aller Teilmengen von M* , welche man oft mit $\mathcal{P}(M)$ (für Potenzmenge) bezeichnet. Wir erwähnen dieses Beispiel hier, um anzumerken, dass die Elemente einer Menge selbst auch Mengen sein können. So gilt zum Beispiel

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

Es ist wichtig, in der Notation zwischen Elementen und Teilmengen immer klar zu unterscheiden. So gilt zum Beispiel $1 \in \{1, 2, 3\}$, *aber nicht* $\{1\} \in \{1, 2, 3\}$. Andererseits gilt $\{1\} \subseteq \{1, 2, 3\}$, *aber nicht* $1 \subseteq \{1, 2, 3\}$. Schließlich gilt $\{1\} \in \mathcal{P}(\{1, 2, 3\})$, *aber nicht* $1 \in \mathcal{P}(\{1, 2, 3\})$.

Bemerkung. Den letzten Absatz sollten Sie so oft lesen und so lange mit anderen Studierenden diskutieren, bis er Ihnen restlos klar ist. Dies wird Ihnen später viel Verwirrung und vielleicht auch verlorene Punkte bei Übungsaufgaben und Klausuren ersparen.

Schließlich ist es immer wieder einmal nützlich, geordnete Paare von Elementen gegebener Mengen zu betrachten. Sind also X und Y zwei Mengen, so definieren wir die *Produktmenge* $X \times Y$ als die Menge aller geordneter Paare (x, y) mit $x \in X$ und $y \in Y$, in Symbolen

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

Beispiel. In den Vorlesungen über Geometrie und lineare Gleichungssysteme hatten wir über die Ebene als \mathbb{R}^2 gesprochen. In unserer neuen Notation von Produktmengen gilt

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}.$$

6.2 Abbildungsbegriff

Die folgende Definition abstrahiert die wesentlichen Eigenschaften dieser Zuordnungen und macht daraus einen mathematischen Begriff.

Definition. Eine *Abbildung* f von einer Menge X in eine Menge Y (wir schreiben dafür $f : X \rightarrow Y$) ist eine Vorschrift, die jedem Element $x \in X$ genau ein Element $y \in Y$ zuordnet. Wir schreiben auch $y = f(x)$ für den *Wert* der Abbildung f im Punkt $x \in X$.

X heißt *Definitionsbereich* der Abbildung f , und Y heißt *Wertebereich* der Abbildung.

Beispiele. Wir betrachten noch einmal die Zuordnungen aus Beispiel 6.

- (a) Im ersten Fall war der Definitionsbereich die Menge der Punkte der Erdoberfläche, und der Wertebereich ist die Menge der *potentiell möglichen Werte* für den Luftdruck. Diese kann man (z.B.) mit den positiven reellen Zahlen identifizieren. Im zweiten Fall war der Definitionsbereich ein gewisses Zeitintervall, und der Wertebereich wieder die Menge der möglichen Werte für den Luftdruck.
- (b) Hier war der Definitionsbereich die (endliche) Menge der Autos mit Hamburger Kennzeichen, und der Wertebereich die Menge der möglichen Geschwindigkeiten. Dies ist wieder ein gewisses Teilintervall der reellen Zahlen, welches als Teil der Definition der Abbildung gewählt wird.
- (c) Hier ist der Definitionsbereich die Menge aller Dreiecke der gegebenen Ebene, und der Wertebereich ist die Menge aller Punkte derselben Ebene.
- (d) Hier ist der Definitionsbereich ein gewisses Zeitintervall, und der Wertebereich sind alle von unserer ausgewählten Uhr anzeigbaren Uhrzeiten (Zum Beispiel die Uhrzeiten von 00 : 00 bis 23 : 59, eine Menge mit $24 \cdot 60 = 1440$ Elementen).
- (e) Bei der Addition natürlicher Zahlen ist der Definitionsbereich $\mathbb{N} \times \mathbb{N}$ und der Wertebereich wird typischerweise auch als \mathbb{N} gewählt, obwohl auch die natürlichen Zahlen ≥ 2 als Wertebereich in Frage kämen, da die Summe zweier natürlicher Zahlen mindestens 2 ist.

Wie wir sehen, haben wir bei der konkreten Definition einer Abbildung einen gewissen Spielraum, was Definitions- und Wertebereich angeht. Um Unklarheiten zu vermeiden gibt man diese also bei der Einführung einer neuen Abbildung normalerweise explizit an. Ist der Definitionsbereich sehr klein, so kann man eine Abbildung auch als Wertetabelle schreiben. Folgende Tabelle beschreibt etwa eine Abbildung $p : X \rightarrow Y$ mit $X = Y = \{1, 2, 3, 4, 5\}$:

x	1	2	3	4	5
$p(x)$	2	3	4	5	1

Man beachte aber, dass die Tabelle zwar den Definitionsbereich der Abbildung, nicht aber den Wertebereich deutlich macht. Dieser muss also zusätzlich mit angegeben werden, wenn er nicht aus dem Kontext klar ist.

Definition. Abbildungen, deren Wertebereich ein Zahlbereich ist (also eine Teilmenge von \mathbb{R} oder sogar \mathbb{C} , den komplexen Zahlen) nennt man auch *Funktion*.

Funktionen werden oft (aber nicht immer) durch Angabe einer Funktionsvorschrift definiert. Um solche Definitionen zu kennzeichnen, benutzt viele Autoren hier *einmal* das Zeichen „:=“. In Beispielen sieht das dann wie folgt aus.

Beispiele. • Wir definieren eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ durch $f(x) := x^{2016} - 2016x$.

- Wir definieren eine Funktion $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ durch $g(x, y, z) := x + y + z - 1$.
- Der *Absolutbetrag* ist die Funktion $|\cdot| : \mathbb{R} \rightarrow \mathbb{R}$, welche durch

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

definiert ist.

Funktionen, deren Definitions- und Wertebereich Teilmengen von \mathbb{R} sind, werden oft durch einen *Funktionsgraphen* veranschaulicht. Sind $X, Y \subseteq \mathbb{R}$ und $f : X \rightarrow Y$, so ist der Funktionsgraph gerade die Teilmenge

$$\text{graph}(f) = \{(x, f(x)) \mid x \in X\} \subseteq X \times Y \subseteq \mathbb{R}^2$$

aller Punkte der Ebene deren erste Koordinate in X liegt und deren zweite Koordinate gerade der Funktionswert der ersten Koordinate ist. Man bemerke, dass die Definition für beliebige Mengen X und Y funktioniert.

Definition. Seien X und Y Mengen und $f : X \rightarrow Y$ eine Abbildung.

- Das *Bild* einer Teilmenge $A \subseteq X$ ist die Menge

$$f(A) := \{f(x) \mid x \in A\}.$$

Insbesondere bezeichnet man die Teilmenge $f(X) \subseteq Y$ als *das Bild von f* .

- Das *Urbild* einer Teilmenge $B \subset Y$ ist die Teilmenge

$$f^{-1}(B) := \{x \in X \mid f(x) \in B\} \subseteq X.$$

Hat die Teilmenge $B \subset Y$ nur ein Element, d.h. $B = \{y\} \subset Y$, so schreibt man zur besseren Lesbarkeit ausnahmsweise nur $f^{-1}(y)$ statt $f^{-1}(\{y\})$. Für das Urbild $f^{-1}(Y)$ gibt es keinen speziellen Namen, da stets $f^{-1}(Y) = X$ gilt (warum?).

Beispiele. Wir betrachten noch einmal einige Funktionen aus den vorherigen Beispielen.

- Für die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ aus Beispiel 6.2 zeigt eine kurze Rechnung, dass das Bild $f(\mathbb{R})$ von f gerade die Teilmenge $\{y \in \mathbb{R} \mid y \geq -2015\}$ ist, das Bild $f([0, 1])$ ist gerade das Intervall $[-2015, 0]$. Das Urbild der Teilmenge $(-3000, -2500)$ ist leer, und das Urbild des Punktes $0 \in \mathbb{R}$ ist die Teilmenge $\{0, \sqrt[2015]{2016}\}$. Können Sie diese Behauptungen beweisen?
- Das Bild der Funktion $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ aus Beispiel 6.2 ist ganz \mathbb{R} . Was ist das Urbild des Punktes $0 \in \mathbb{R}$ unter g ?
- Das Bild des Absolutbetrages ist die Teilmenge $[0, \infty) \subset \mathbb{R}$. Das Urbild eines Intervalles $[a, b]$ mit $0 < a < b$ besteht aus der Vereinigung der beiden Teilintervalle $[-b, -a] \cup [a, b] \subset \mathbb{R}$. Insbesondere hat jedes $x > 0$ das Urbild $\{-x, x\}$, während das Urbild der Null nur aus der Null besteht, und das Urbild von $x < 0$ leer ist.
- Ist $h : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion, so können wir eine neue Abbildung $H : \mathbb{R} \rightarrow \mathbb{R}^2$ durch $H(x) = (x, h(x))$ definieren. Das Bild dieser Abbildung H ist dann gerade der Graph der Funktion h in \mathbb{R}^2 .

Wichtig für den Begriff der Abbildung ist es, dass *jedem* Element aus dem Definitionsbereich *genau ein* Element des Wertebereichs zugeordnet wird. Wie die Beispiele zeigen, kann dabei *verschiedenen* Elementen des Definitionsbereichs durchaus *dasselbe* Element des Wertebereichs zugeordnet werden. Auch beschreibt der Wertebereich nur die festgelegte Menge *möglicher* Werte, d.h. nicht jedes Element des Wertebereichs muss tatsächlich auch als Wert vorkommen.

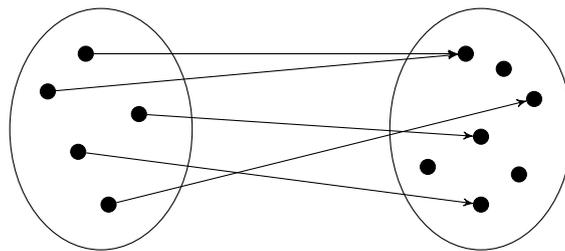


Abbildung 6.2: Schematische Darstellung einer Abbildung. Der Definitionsbereich ist links, der Wertebereich rechts dargestellt.

Manchmal ist es aber durchaus nützlich anzunehmen, dass tatsächlich jedes Element des Wertebereichs einer Abbildung $f : X \rightarrow Y$ ein Bildpunkt von f ist, d.h. $f(X) = Y$

gilt. Abbildungen mit dieser Eigenschaft nennt man *surjektiv*. Entsprechen andererseits verschiedenen Elementen von X stets verschiedene Bildpunkte in Y , so nennt man die Abbildung $f : X \rightarrow Y$ *injektiv*. Eine Abbildung, welche beide diese Bedingungen erfüllt, nennt man *bijektiv*. Für bijektive Abbildungen $f : X \rightarrow Y$ gilt:

- zu jedem $x \in X$ gibt es genau ein $y \in Y$ mit $f(x) = y$ (dies gilt nach Definition für *jede* Abbildung), und
- zu jedem $y \in Y$ gibt es genau ein $x \in X$ mit $f(x) = y$ (dies sind die beiden Bedingungen „ f ist injektiv“ (es gibt höchstens ein $x \dots$) und „ f ist surjektiv“ (es gibt mindestens ein $x \dots$)).

Unter diesen Voraussetzungen ist die Abbildung $f : X \rightarrow Y$ *umkehrbar*: es existiert dann eine Abbildung $g : Y \rightarrow X$ so dass die folgenden Beziehungen gelten:

$$\forall x \in X : g(f(x)) = x \quad \text{und} \quad \forall y \in Y : f(g(y)) = y.$$

In diesem Fall nennt man die Abbildung $g : Y \rightarrow X$ die *Umkehrabbildung* oder *inverse Abbildung* zu $f : X \rightarrow Y$. Diese wird meist mit $f^{-1} : Y \rightarrow X$ bezeichnet.

Beispiel. Wir betrachten noch einmal die Abbildung $p : \{1, 2, 3, 4, 5\} \rightarrow \{1, 2, 3, 4, 5\}$, welche durch die Tabelle

x	1	2	3	4	5
$p(x)$	2	3	4	5	1

definiert wird. Diese Abbildung ist umkehrbar, und die Umkehrabbildung $p^{-1} : \{1, 2, 3, 4, 5\} \rightarrow \{1, 2, 3, 4, 5\}$ ist gegeben durch

y	1	2	3	4	5
$p^{-1}(y)$	5	1	2	3	4

In solchen einfachen Beispielen mit Wertetabelle erhält man also, wenn überhaupt möglich, die Umkehrabbildung, indem man „die Zeilen vertauscht“.

Beispiele. (a) Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, definiert durch $f(x) = x + x^3$ ist streng monoton wachsend (aus $x_1 < x_2$ folgt $f(x_1) < f(x_2)$) und nimmt jeden Wert $y \in \mathbb{R}$ an. Sie besitzt also eine Umkehrfunktion $f^{-1} : \mathbb{R} \rightarrow \mathbb{R}$, für die man allerdings keine elementare Formel angeben kann. Dies ist „im wirklichen Leben“ durchaus typisch.

(b) Die Funktion $q : \mathbb{R} \rightarrow \mathbb{R}$, $q(x) = x^2$, besitzt keine Umkehrfunktion, da negative Zahlen im Wertebereich kein Urbild und die positiven Zahlen zwei Urbilder besitzen. Schränkt man die Funktion aber ein zu einer Funktion $\tilde{q} : [0, \infty) \rightarrow [0, \infty)$, so verschwinden diese beiden Probleme. Die Umkehrfunktion dieser Einschränkung \tilde{q} ist natürlich nichts anderes als die Quadratwurzel $\sqrt{} : [0, \infty) \rightarrow [0, \infty)$.

(c) Unsere Abbildung, welche jedem ebenen Dreieck seinen Schwerpunkt zuordnet, hat als Bild den gesamten Wertebereich, denn jeder Punkt der Ebene kann als Schwerpunkt auftreten. Allerdings ist die Abbildung nicht umkehrbar, denn verschiedene Dreiecke können durchaus denselben Schwerpunkt besitzen.

Bei der Diskussion der Umkehrabbildung trat bereits eine Konstruktion auf, die wir wegen ihrer Wichtigkeit noch einmal explizit herausstellen wollen.

Definition. Seien $f : X \rightarrow Y$ und $g : Y \rightarrow Z$ Abbildungen. Dann definieren wir ihre *Verknüpfung* (oder *Komposition*) als die Abbildung $g \circ f : X \rightarrow Z$ (sprich: „ g nach f “ oder „ g verknüpft mit f “), gegeben durch

$$(g \circ f)(x) = g(f(x)).$$

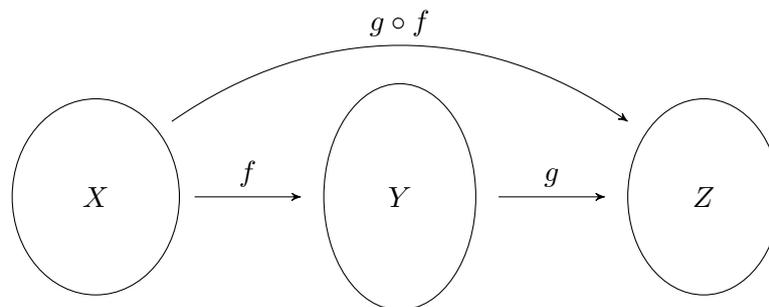


Abbildung 6.3: Schematische Darstellung der Verknüpfung (oder auch *Komposition* zweier Abbildungen $f : X \rightarrow Y$ und $g : Y \rightarrow Z$).

Beispiel. Seien X die Menge der Schüler und Schülerinnen einer Schulklasse, Y die Menge der möglichen Punktzahlen in einer Mathematikprüfung, und Z die Menge der möglichen Zensuren. Nachdem der Korrektur gibt es eine Abbildung $f : X \rightarrow Y$, welche jeder Schülerin und jedem Schüler seine Punktzahl zuordnet. Außerdem hat die Lehrerin vermutlich vor der Prüfung eine Abbildung $g : Y \rightarrow Z$ bekanntgegeben, welche angibt, wie die Punktzahlen in Zensuren umgerechnet werden. Eine Schülerin kann dann ihre Zensur bestimmen, indem sie die Verknüpfung dieser beiden Abbildungen bildet, und bei sich selbst auswertet.

6.3 Elementare Funktionen

In diesem Abschnitt behandeln wir einige Klassen von Funktionen $\mathbb{R} \rightarrow \mathbb{R}$, die aus der Schule schon bekannt sein sollten. Es werden jeweils einige interessante Eigenschaften der maximal mögliche Definitionsbereich D sowie der Bildbereich $f(D)$ angegeben. Wir werden jeweils auch den Graphen der Funktionen skizzieren.

Potenzfunktionen

Definition. Funktionen mit den Abbildungsvorschrift

$$f(x) = x^q$$

wobei q eine rationale Zahl ist heißen *Potenzfunktionen*.

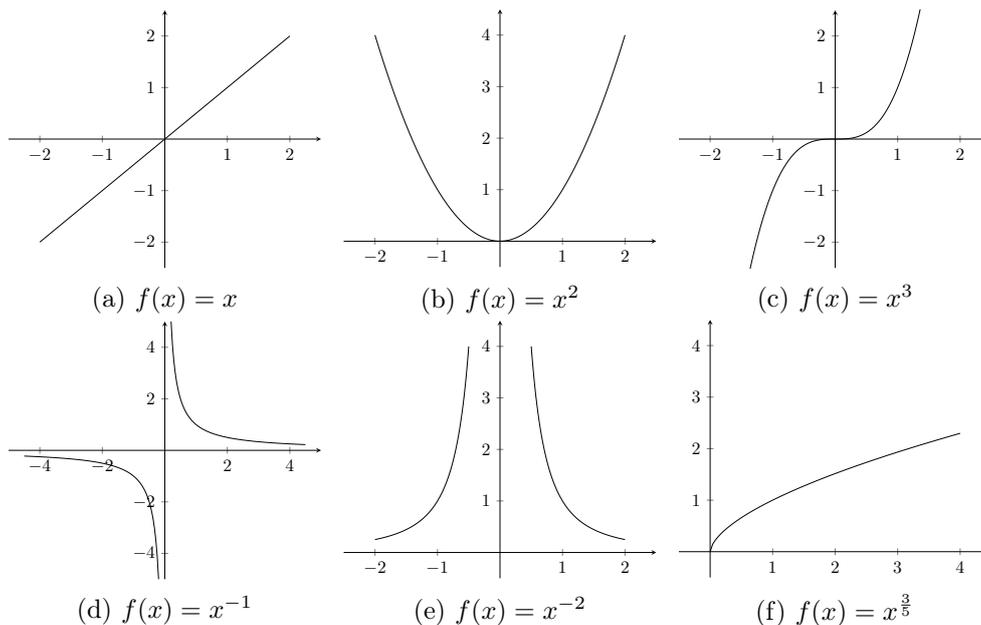


Abbildung 6.4: Funktionsgraphen einiger Potenzfunktionen mit verschiedenen Exponenten

Definitionsbereich und Bildbereich hängen davon ab, ob der Exponent q ganzzahlig ist oder nicht und ob er positiv oder negativ ist. Wir werden nun auf die verschiedenen Fälle im Detail eingehen.

Für $q = 0$ definiert man $x^0 = 1$, das heißt, die Potenzfunktion ist die Funktion, die überall den Wert 1 hat (konstante Funktion, $D = \mathbb{R}$, $f(D) = \{1\}$).

Der nächste Fall ist, wenn q eine natürliche Zahl ist. In diesem Fall ist die Funktion $f(x) = x^q$ für jede reelle Zahl x definiert. Falls q ungerade ist, so ist der Bildbereich ganz \mathbb{R} und die Funktion ist sogar bijektiv. Für gerades q ist der Bildbereich $f(\mathbb{R}) = \{x \in \mathbb{R} \mid x \geq 0\} = [0, \infty)$.

Beispiele. Abbildungen 6.4a, 6.4b und 6.4c zeigen Potenzfunktionen, die in diese Kategorie fallen.

Lassen wir für q auch negative Zahlen zu, so erhalten wir Funktionen der Form $x^{-1} = \frac{1}{x}$, $x^{-2} = \frac{1}{x^2}$, $x^{-3} = \frac{1}{x^3}$, ... Da wir durch jede reelle Zahl mit Ausnahme der 0 dividieren können, sind diese Funktionen für beliebige x mit Ausnahme der 0 definiert, also $D = \mathbb{R} \setminus \{0\}$. Ihr Bildbereich ist $\mathbb{R} \setminus \{0\}$, falls q ungerade ist und $\{x \in \mathbb{R} \mid x > 0\} = (0, \infty)$, falls q gerade ist.

Beispiele. Abbildungen 6.4d und 6.4e zeigen Funktionen, die in diese Klasse fallen.

Lassen wir für q beliebige Brüche zu, so kommen Wurzeln ins Spiel: $x^{\frac{m}{n}}$ ist definiert als $(\sqrt[n]{x})^m$, also zum Beispiel $x^{\frac{1}{2}} = \sqrt{x}$ bzw. $x^{-\frac{3}{4}} = \frac{1}{(\sqrt[4]{x})^3}$. Üblicherweise kürzt man den Bruch im Exponenten soweit wie möglich, d.h. man verwendet gekürzte Brüche.

Der Definitionsbereich von Funktionen dieser Bauart hängt vom Definitionsbereich der

Wurzel ab. Gerade Wurzeln sind nur für Zahlen ≥ 0 definiert, ungerade Wurzeln überall. Außerdem muss man für negatives q wieder die 0 aus dem Definitionsbereich ausschließen.
Beispiel. Abbildung 6.4f zeigt den Graphen einer Funktion dieses allgemeinen Typs.

Polynome

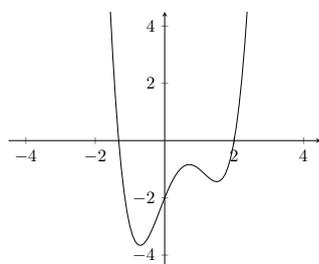
Definition. Funktionen der Form

$$f(x) = \sum_{i=0}^n a_i x^i = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

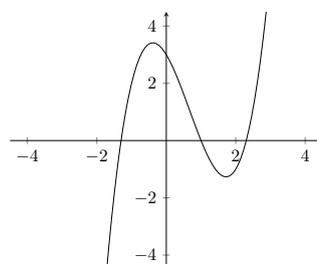
wobei $a_n \neq 0$ heißen Polynome n -ten Grades. a_n heißt Leitkoeffizient des Polynoms.

Da Polynome aus Potenzfunktionen mit Exponenten aus \mathbb{N}_0 zusammengesetzt sind, ist ihr Definitionsbereich ganz \mathbb{R} . Polynome vom Grad 0 sind die konstanten Funktionen, Polynome vom Grad 1 heißen lineare Funktionen (ihre Funktionsgraphen sind Geraden), Polynome vom Grad 2 bezeichnet man als quadratische Funktionen.

Beispiele. Abbildung 6.5 zeigt 2 Graphen von Polynomfunktionen.



(a) $f(x) = x^4 - 2x^3 - x^2 + 3x - 2$



(b) $f(x) = x^3 - 2x^2 - 2x + 3$

Abbildung 6.5: Graphen von Polynomen

Bemerkung. Ist der Index n des Leitkoeffizienten eines Polynoms

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

ungerade, so ist das Bild $f(\mathbb{R})$ stets ganz \mathbb{R} . Für gerades n ist die Sache etwas komplizierter: je nach Vorzeichen von a_n ist das Bild dann ein Intervall der Form $(-\infty, b]$ oder $[a, \infty)$. Dies können Sie nach dem ersten Semester Analysis problemlos beweisen.

Rationale Funktionen

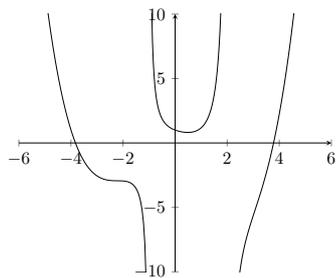
Rationale Funktionen lassen sich auf dieselbe Weise aus den Polynomen gewinnen, wie rationale Zahlen aus den ganzen Zahlen gewonnen werden.

Definition. Als *rationale Funktionen* bezeichnen wir Funktionen der Form

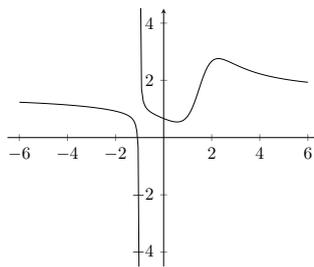
$$f(x) = \frac{p(x)}{q(x)},$$

wobei p und q Polynome sind und q nicht die konstante Nullfunktion ist.

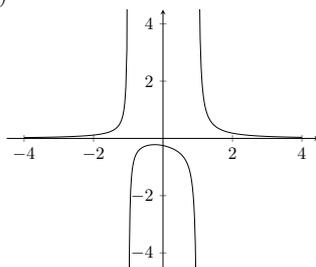
Beispiele. Abbildung 6.6 zeigt 3 Graphen von rationalen Funktionen.



(a) $f(x) = \frac{x^6 - 14x^4 - 4x^2 + 10x - 40}{20(x^2 - x - 2)}$



(b) $g(x) = \frac{3x^3 - 2x^2 - 2x + 4}{2x^3 - 4x^2 + 6}$



(c) $h(x) = \frac{2x^2 + x + 1}{4x^4 - 4}$

Abbildung 6.6: Graphen rationaler Funktionen

Der maximale Definitionsbereich einer solchen rationalen Funktion $f(x) = \frac{p(x)}{q(x)}$ ist die Menge

$$\mathbb{R} \setminus \{x \in \mathbb{R} \mid q(x) = 0\}$$

derjenigen reellen Zahlen, in denen der Nenner von Null verschieden ist. Da Polynome nur endliche viele Nullstellen haben, ist die Ausnahmemenge für jede rationale Funktion endlich. Für die in Abbildung 6.6 gezeigten Funktionen sind die Definitionsbereiche

$$D(f) = \mathbb{R} \setminus \{-2, 1\},$$

$$D(g) = \mathbb{R} \setminus \{-1\},$$

$$D(h) = \mathbb{R} \setminus \{-1, 1\}.$$

Bemerkung. Das Verhalten einer rationalen Funktion $f(x) = \frac{p(x)}{q(x)}$ im Unendlichen hängt von den Graden der Polynome p und q ab. Können Sie eine allgemeine Regel formulieren?

Exponentialfunktionen

Exponentialfunktionen sehen zwar ähnlich aus wie Potenzfunktionen, im Gegensatz zu diesen steht die Variable x aber nun im Exponenten.

Definition. Als *Exponentialfunktionen* bezeichnen wir Funktionen der Form

$$f(x) = a^x$$

wobei $a > 0$ eine reelle Zahl ist.

Der Definitionsbereich dieser Funktionen ist ganz \mathbb{R} . Dass sie für rationales x definiert sind, haben wir gerade eben gesehen. Die Definition für beliebiges $x \in \mathbb{R}$ kann Beispielsweise mithilfe von Grenzwerten erfolgen, einem Konzept, das wir später in diesem Kurs noch kennenlernen werden. Der Bildbereich ist die Menge aller positiven reellen Zahlen.

Das Monotonieverhalten von Exponentialfunktionen mit Parameter $a < 1$ und $a > 1$ ist sehr unterschiedlich, wie Abbildung 6.7 zeigt. Der Grenzfall $a = 1$ gibt die konstante Funktion $f(x) = 1$ und ist daher nur mäßig interessant .

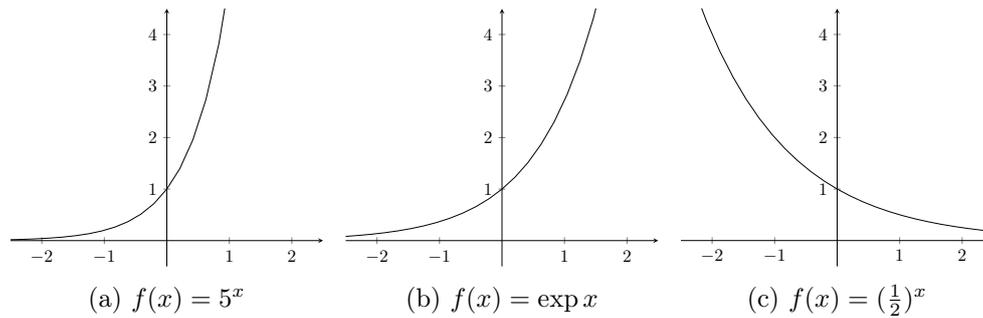


Abbildung 6.7: Graphen von Exponentialfunktionen

Satz 15 (Rechenregeln für Potenzen). *Es gilt:*

1. $a^x \cdot a^y = a^{x+y}$,
2. $a^{cx} = (a^c)^x$.

Definition. Die Exponentialfunktion $\exp(x) = e^x$ mit Basis $e = 2.71828182\dots$ heißt *natürliche Exponentialfunktion*.

Der Grund, aus dem man Exponentialfunktionen gerne zur Basis e darstellt ist, dass die Funktion $f(x) = e^x$ gleich ihrer eigenen Ableitung ist (mehr dazu später). Satz 15 sagt auch, dass wir jede Exponentialfunktion auf eine natürliche Exponentialfunktion zurückführen können indem wir ein b finden, sodass $a = e^b$. Dann ist nämlich $a^x = (e^b)^x = e^{bx} = \exp(bx)$.

Logarithmusfunktion

Die Exponentialfunktionen $f(x) = a^x$ sind für $a \neq 1$ bijektiv als Abbildungen von \mathbb{R} nach $(0, \infty)$. Dies erlaubt es uns, ihre Umkehrfunktionen zu betrachten.

Definition. Zu vorgegebenen Zahlen $a > 0$, $a \neq 1$, und $u > 0$ ist der Logarithmus $\log_a(u)$ von u zur Basis a die eindeutig bestimmte Zahl s , so dass die Gleichung $a^s = u$ gilt.

Beispiele. • $\log_3(9) = 2$, weil $3^2 = 9$.

- $\log_4(2) = \frac{1}{2}$, weil $4^{\frac{1}{2}} = \sqrt{4} = 2$.
- $\log_{10}(10000) = 4$.
- $\log_2(64) = 6$.

Aus der Definition des Logarithmus sieht man sofort, dass $\log_a(a^s) = s$ und dass $a^{\log_a(u)} = u$ sein muss. Der Logarithmus zur Basis 10 wird manchmal mit $\log_{10}(u) = \lg(u)$ bezeichnet, der Logarithmus zur Basis e der natürlichen Exponentialfunktion mit $\log_e(u) = \ln(u)$. Wenn die Basis a aus dem Zusammenhang klar ist, dann lässt man sie auch manchmal weg und schreibt nur $\log(u)$ anstatt $\log_a(u)$.

Satz 15 über Rechenregeln für Potenzen hat direkte Implikationen für das Rechnen mit Logarithmen.

Satz 16 (Rechenregeln für Logarithmen). *Es gilt*

1. $\log_a(uv) = \log_a(u) + \log_a(v)$ und $\log_a\left(\frac{u}{v}\right) = \log_a(u) - \log_a(v)$.
2. $\log_a(u^t) = t \log_a(u)$ und $\log_a(\sqrt[t]{u}) = \frac{1}{t} \log_a(u)$

Außerdem kann man Logarithmen zu einer bestimmten Basis relativ leicht in eine andere Basis umrechnen:

$$\log_a(u) = \log_a(b^{\log_b(u)}) = \log_b(u) \log_a(b).$$

Somit ist $\log_b(u) = \frac{\log_a(u)}{\log_a(b)}$. Diese Regel kann man beispielsweise benutzen um beliebige Logarithmen zu berechnen, wenn der Taschenrechner nur den natürlichen Logarithmus berechnen kann: $\log_a(u) = \frac{\ln(u)}{\ln(a)}$.

Da der Logarithmus zur Basis a für alle positiven Zahlen definiert ist, können wir eine Funktion mit Definitionsbereich $(0, \infty)$ definieren, die jeder Zahl x den Logarithmus $\log_a(x)$ zuordnet.

Definition. Funktionen $g : (0, \infty) \rightarrow \mathbb{R}$ der Form $g(x) = \log_a(x)$ heißen *Logarithmusfunktionen*.

Dies sind natürlich gerade die Umkehrfunktionen der jeweiligen Exponentialfunktion $f(x) = a^x$. In Abbildung 6.8 sind die Graphen einiger Logarithmusfunktionen dargestellt.

Trigonometrische Funktionen

Als trigonometrische Funktionen oder auch Winkelfunktionen bezeichnet man die aus der Schule schon bekannten Funktionen Sinus, Kosinus, Tangens, etc. Wir geben hier nur einige der elementarsten Eigenschaften dieser Funktionen an.

Definition. Sei $x \in \mathbb{R}$ und sei (u, v) der Punkt am Einheitskreis, der mit der positiven u -Achse einen Winkel von x einschließt (Bogenmaß, gegen den Uhrzeigersinn gerechnet). Dann definieren wir den *Kosinus von x* durch $\cos x = u$ und den *Sinus von x* durch $\sin x = v$ (siehe Abbildung 6.9).

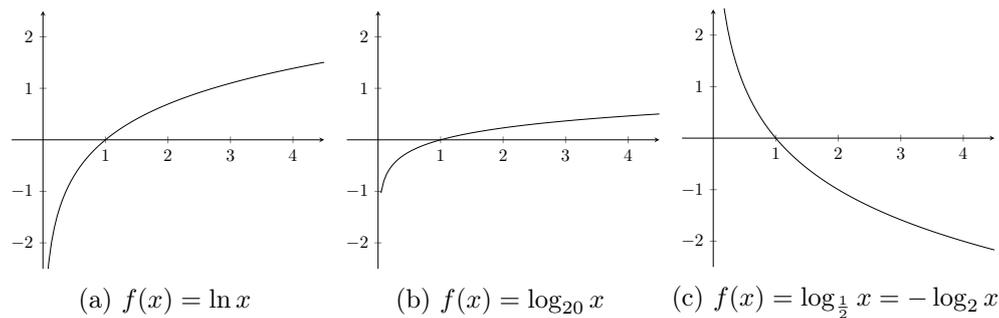


Abbildung 6.8: Graphen von Logarithmusfunktionen

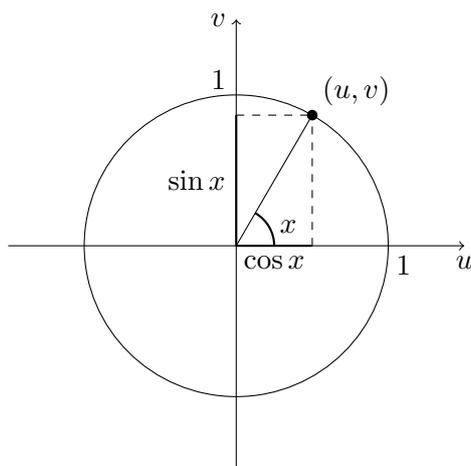


Abbildung 6.9: Definition von Sinus und Kosinus

Aus der Definition sieht man sofort, dass $\sin x$ und $\cos x$ die Längen der beiden Katheten eines rechtwinkligen Dreiecks sind, dessen Hypotenuse Länge 1 besitzt. Somit folgt mittels des Satzes von Pythagoras, dass

$$\forall x \in \mathbb{R} : (\sin x)^2 + (\cos x)^2 = 1.$$

Man schreibt dafür auch $\sin^2 x + \cos^2 x = 1$.

Sinus und Kosinus haben gleichartigen Verlauf. Die beiden Graphen sind zueinander um das Bogenmaß $\frac{\pi}{2}$ verschoben, was einem Viertelkreis entspricht. In Abbildung 6.10 ist dies gut zu sehen. Beide Funktionen sind auf ganz \mathbb{R} definiert und nehmen Werte in $[-1, 1]$ an.

Außerdem sind die beiden Funktionen periodisch mit Periode 2π , also

$$\begin{aligned} \sin(x + 2\pi) &= \sin x \\ \cos(x + 2\pi) &= \cos x \end{aligned}$$

Das sieht man leicht daran, dass der Winkel 2π einem vollen Kreis entspricht, danach wiederholt sich die Bahn des Punktes (u, v) .

Definition. Die Funktionen *Tangens* beziehungsweise *Kotangens* sind definiert durch die Gleichungen

$$\tan x = \frac{\sin x}{\cos x} \quad \text{beziehungsweise} \quad \cot x = \frac{\cos x}{\sin x}.$$

Tangens und Kotangens sind überall definiert wo der jeweilige Nenner nicht gleich Null ist. Im Fall des Tangens ist das $\mathbb{R} \setminus \{x \mid \exists n \in \mathbb{N} : x = (\frac{1}{2} + n)\pi\}$, im Fall des Kotangens $\mathbb{R} \setminus \{x \mid \exists n \in \mathbb{N} : x = n\pi\}$. Der Bildbereich von Tangens und Cotangens ist jeweils ganz \mathbb{R} . Beide Funktionen sind periodisch mit Periode π .

Die folgende Tabelle enthält einige Werte für die in diesem Abschnitt definierten Winkelfunktionen, „n.d.“ bedeutet, dass die Funktion an der entsprechenden Stelle nicht definiert ist.

Winkel	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	π	$\frac{3\pi}{2}$	2π
Winkel in Grad	0°	30°	45°	60°	90°	180°	270°	360°
$\sin x$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	0	-1	0
$\cos x$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	-1	0	1
$\tan x$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	n.d.	0	n.d.	0
$\cot x$	n.d.	$\sqrt{3}$	1	$\frac{\sqrt{3}}{3}$	0	n.d.	0	n.d.

Zusammengesetzte Funktionen

Viele Funktionen, die in praktischen Anwendungen eine Rolle spielen, entsprechen zwar nicht genau den elementaren Funktionen die bisher behandelt wurden, können aber relativ einfach aus ihnen zusammengesetzt werden. Beispielsweise kann man Funktionen addieren, subtrahieren, multiplizieren und dividieren, indem man die entsprechende Operation auf die Funktionswerte in jedem Punkt x anwendet.

Man nennt diese Operation punktweise Addition, Subtraktion, Multiplikation bzw. Division. Sie ist überall definiert, wo beide Funktionen definiert sind. Ausnahme ist die

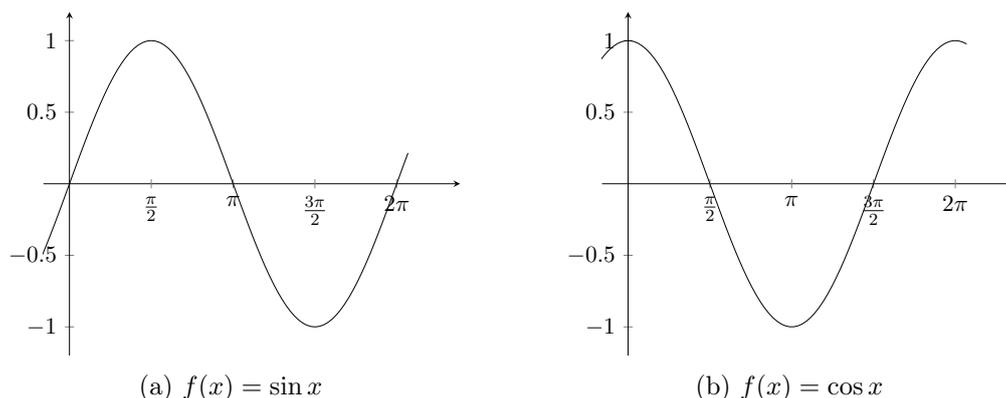


Abbildung 6.10: Die Graphen der Sinus- und Kosinusfunktion

Division, hier muss auch noch sichergestellt werden, dass der Wert des Nenners im jeweiligen Punkt nicht 0 ist.

Definition. Seien f und g zwei Funktionen mit Definitionsbereichen D_f und D_g . Dann kann man folgende Funktionen definieren:

$$\begin{aligned} (f + g): & & D_f \cap D_g & \rightarrow \mathbb{R} & & x \mapsto f(x) + g(x) \\ (f - g): & & D_f \cap D_g & \rightarrow \mathbb{R} & & x \mapsto f(x) - g(x) \\ (f \cdot g): & & D_f \cap D_g & \rightarrow \mathbb{R} & & x \mapsto f(x) \cdot g(x) \\ \left(\frac{f}{g}\right): & & \{x \in D_f \cap D_g \mid g(x) \neq 0\} & \rightarrow \mathbb{R} & & x \mapsto \frac{f(x)}{g(x)} \end{aligned}$$

Beispiele. • $f(x) = 1 + 3x^2$, $g(x) = 2x - x^2$, dann ist $f + g : \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch $(f + g)(x) = 1 + 2x + 2x^2$

- $f(x) = e^x$, $g(x) = e^{3x}$, dann ist $f \cdot g : \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch $(f \cdot g)(x) = e^x \cdot e^{3x} = e^{4x}$
- $f(x) = \sin x$, $g(x) = \cos x$, dann ist $\frac{f}{g}$ gegeben durch $x \mapsto \tan x$ (mit dem entsprechenden Definitionsbereich).

Neben punktweiser Kombination von Funktionen kann man auch neue Funktionen erhalten, indem man den Funktionswert einer Funktion als Argument einer anderen benutzt. Diese Verknüpfung (oder auch Komposition) hatten wir im allgemeinen Teil zu Abbildungen bereits beschrieben.

Wir betrachten abschließend noch einige Beispiele für Verknüpfungen von Funktionen. Der Definitionsbereich einer solchen Verknüpfung $g \circ f$ von Funktionen $f : D_f \rightarrow \mathbb{R}$ und $g : D_g \rightarrow \mathbb{R}$ besteht aus allen x aus dem Definitionsbereich von f , so dass $f(x)$ im Definitionsbereich von g liegt - in Symbolen

$$D_{g \circ f} = \{x \in D_f \mid f(x) \in D_g\}.$$

Beispiele. • Sei $f(x) = 1 + x$ und $g(x) = 1 - x + x^2$, dann ist $(g \circ f)(x) = 1 - f(x) + f(x)^2 = 1 - (1 + x) + (1 + x)^2$ mit Definitionsbereich \mathbb{R} .

- Falls $g(x) = f(x) = \frac{1}{x}$ ist, dann ist $(g \circ f)(x) = \frac{1}{\left(\frac{1}{x}\right)} = x$ mit Definitionsbereich $\mathbb{R} \setminus \{0\}$.
- Falls $f(x) = x^3$ und $g(x) = \sqrt[3]{x}$, so ist $g \circ f(x) = \sqrt[3]{x^3} = x$ mit Definitionsbereich \mathbb{R} .
- Die letzten beiden Beispiele lassen sich allgemein formulieren: Ist $f : X \rightarrow Y$ eine bijektive Abbildung so ist $f^{-1} \circ f$ die Abbildung von X nach X , welche jedes $x \in X$ auf sich selbst abbildet. Analog ist $f \circ f^{-1}$ die Abbildung von Y nach Y , welches jedes $y \in Y$ auf sich selbst abbildet. Als weiteres konkretes Beispiel für zueinander inverse Funktionen hatten wir ja bereits Logarithmen und Exponentialfunktionen betrachtet.

7 Was ist eigentlich ein Grenzwert?

Grenzwerte sind ein zentrales Konzept der Analysis, welches Ihnen im ersten Semester begegnen wird und von da an immer wieder in diversen Kontexten auftritt. Zur Motivation beginnen wir mit einigen Fragen:

- In welchem Sinne gilt die Gleichung $1 = 0.999\dots$ (unendlicher Dezimalbruch)?
- Wie kann man den Umfang oder den Flächeninhalt eines Kreises näherungsweise bestimmen?
- (Zenos Paradox) Achilles und eine Schildkröte sollen um die Wette laufen. Achilles ist zehnmal so schnell wie die Schildkröte, darum erhält die Schildkröte 10 Meter Vorsprung. Nun erklärt der Philosoph Zeno, dass Achilles das Wettrennen gar nicht gewinnen kann: Wenn er an dem Ort ankommt, an dem die Schildkröte gestartet ist, dann ist sie ja schon ein Stück weiter, und kommt er an diesem nächsten Punkt an, dann ist sie schon wieder etwas weiter, und so fort... Hat Zeno recht?

Wie wir sehen werden, haben alle diese Fragen mit Grenzwerten zu tun. Was ist nun aber ein Grenzwert?!? Betrachten wir dazu als Beispiel die Frage nach dem Umfang eines Kreises vom Radius 1. Erste Näherungslösungen erhalten wir, indem wir statt des eigentlichen Kreises regelmäßige Polygone betrachten, so dass entweder die Ecken oder die Kantenmittelpunkte auf dem Kreis liegen. Im folgenden Bild sehen wir einige Beispiele für solche Polygone.

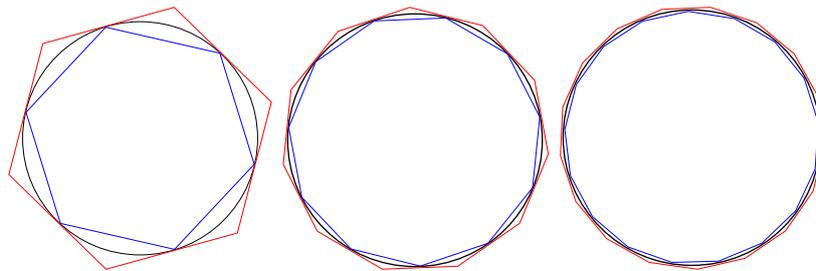


Abbildung 7.1: Approximationen einer Kreislinie durch Polygone mit 6, 11 und 17 Ecken, jeweils von innen (blau) und von außen (rot).

Die Bilder suggerieren mehrere Dinge. Zunächst einmal ist klar, dass die blauen Polygone einen kleineren Umfang haben als die roten, denn um von einem blauen zu einem roten Polygon zu kommen, ersetzen wir jeweils jede Kante durch die beiden anderen Seiten eines Dreiecks mit dieser Basis. Auch scheint die Länge der Kreislinie zwischen den

Längen der blauen und der roten Polygone zu liegen. Um also erste Abschätzungen für die Länge der Kreislinie zu erhalten, betrachten wir der Einfachheit halber einmal die Approximation durch Sechsecke. Da unsere Polygone regelmäßig gewählt waren, ist der

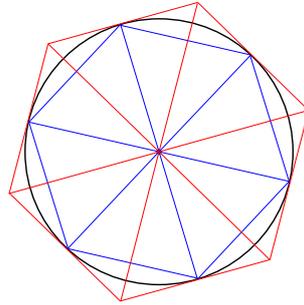


Abbildung 7.2: Die inneren Dreiecke sind gleichseitig und haben als Kantenlänge gerade den Radius des Kreises. Außerdem entspricht diese Kantenlänge gerade der Höhe der äußeren Dreiecke.

Scheitelwinkel der Dreiecke im Zentrum gerade $\frac{360^\circ}{6} = 60^\circ$, so dass es sich um gleichseitige Dreiecke handelt. Die inneren, blauen Dreiecke haben als Kantenlänge gerade den Radius unseres Kreises, den wir als $r = 1$ angenommen hatten, und wir erhalten für den Umfang des blauen Sechsecks $6 \cdot 1 = 6$. Für die äußeren, roten Dreiecke bilden die blauen Kanten gerade die Höhe, so dass man mit einer kurzen Rechnung als Kantenlänge der roten Dreiecke $\frac{2}{\sqrt{3}}$ erhält. Insgesamt erhalten wir für die Länge ℓ unserer Kreislinie die Abschätzungen

$$6 < \ell < 6 \cdot \frac{2}{\sqrt{3}} \approx 6.9282.$$

Betrachten wir nun noch einmal die Abbildung 7.1, so bemerken wir auch, dass „*sich die Polygone der Kreislinie immer mehr annähern, wenn man die Anzahl der Ecken erhöht*“. Diese Aussage zu präzisieren ist nicht ganz einfach, denn es genügt für die folgende Argumentation nicht, festzustellen, dass die Polygone als Punktmenge immer näher an der Kreislinie liegen. In der Tat ist es nicht völlig trivial, die Länge der Kreislinie überhaupt zu definieren. Nehmen wir aber einmal an, dass es eine sinnvolle Definition für diese Länge ℓ gibt, und dass für jedes n der Umfang u_n des inneren n -Ecks und der Umfang U_n des äußeren n -Ecks die Ungleichungen

$$u_n < \ell < U_n$$

erfüllen. Man kann nachrechnen, dass die u_n eine aufsteigende Folge bilden, d.h. es gilt

$$u_3 < u_4 < u_5 < u_6 < \dots$$

Analog bilden die U_n eine absteigende Folge, d.h. es gilt

$$U_3 > U_4 > U_5 > U_6 > \dots$$

Anders ausgedrückt erhalten wir eine Folge von Intervallen

$$[u_3, U_3] \supset [u_4, U_4] \supset [u_5, U_5] \supset [u_6, U_6] \supset \cdots \supset [u_n, U_n] \supset [u_{n+1}, U_{n+1}] \supset \cdots,$$

welche ineinander geschachtelt sind. In diesem konkreten Beispiel besteht der Durchschnitt dieser unendlich vielen Intervalle aus genau einer reellen Zahl x_0 , und da die Länge ℓ unseres Kreisbogens offenbar in jedem der Intervalle und somit auch in diesem Durchschnitt enthalten ist, muss sie mit der reellen Zahl x_0 übereinstimmen. Dies ist der Grenzwert der Umfänge sowohl der einbeschriebenen als auch der umbeschriebenen n -Ecke „wenn man n gegen Unendlich gehen lässt“. In der Tat wissen Sie aus der Schule vermutlich bereits die Antwort: es gilt $\ell = x_0 = 2\pi$.

7.1 Folgen

Um nun das Wesentliche aus dem obigen Beispiel zu extrahieren, formalisieren wir noch einmal einen bereits benutzten Begriff.

Definition. Eine *Folge reeller Zahlen* ist eine Abbildung von \mathbb{N} nach \mathbb{R} . Wir schreiben solche Folgen in der Form $(x_n)_{n \in \mathbb{N}}$ oder manchmal auch nur (x_n) , womit gemeint ist, dass n auf x_n abbildet. Die Einträge x_n heißen Glieder der Folge.

Häufig schreibt man Folgen auch als (x_1, x_2, x_3, \dots) , wobei diese Schreibweise nur dann Sinn macht, wenn klar ist, wie die Folge fortgesetzt werden muss. Zum Beispiel enthält die Folge $(1, 2, 3, 4, \dots)$ alle natürlichen Zahlen in der Reihenfolge, die durch die Ordnung $<$ gegeben ist. Bei der Folge $(8, 8, 15, 2, 1, 37, \dots)$ sind die Punkte aber eher fehl am Platz, da nicht sehr klar ist, wie es weitergehen soll. Hier müsste schon genauer gesagt werden, wie die übrigen Folgenglieder aussehen.

In der Praxis ist es häufig nicht von Bedeutung, dass wir es mit Abbildungen zu tun haben. Wichtiger ist, dass die Elemente einer Folge eine bestimmte Reihenfolge haben, also x_1 vor x_2 vor x_3 etc. Insofern kann man eine Folge auch als unendliche Aufzählung von reellen Zahlen auffassen, deren Reihenfolge eine Rolle spielt. Im Gegensatz dazu ist eine Menge eine (möglicherweise unendliche) Aufzählung von Elementen, deren Reihenfolge keine Rolle spielt. So sind beispielsweise die Mengen $\{1, 2, 3, 4, \dots\}$ und $\{3, 1, 2, 4, 5, 6, 7, \dots\}$ gleich, die Folgen $(1, 2, 3, 4, \dots)$ und $(3, 1, 2, 4, 5, 6, 7, \dots)$ aber nicht.

Beispiele. • Die Folge $(n)_{n \in \mathbb{N}} = (1, 2, 3, \dots)$ ist die Folge der natürlichen Zahlen, die wir schon zuvor als Beispiel verwendet haben.

- $(\frac{1}{n})_{n \in \mathbb{N}}$ definiert die Folge $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$.
- $((-1)^n)_{n \in \mathbb{N}}$ definiert die (periodische) Folge $(-1, 1, -1, 1, -1, \dots)$. Periodisch bedeutet in diesem Fall, dass sich die Folge immer wieder wiederholt.
- Manchmal sind die Folgenglieder nicht mit 1 beginnend nummeriert, sondern beginnend mit einer anderen natürlichen Zahl.¹ So bilden auch die Umfänge $(u_n)_{n \geq 3}$ und

¹Dies ist kein großes Problem, schließlich könnte man, falls unbedingt nötig, die Folgenglieder umnummerieren, ohne die Reihenfolge zu ändern, so dass die Nummerierung doch wieder mit 1 beginnt.

$(U_n)_{n \geq 3}$ der einbeschriebenen bzw. umbeschriebenen n -Ecke unseres Kreisbeispiels Folgen reeller Zahlen.

- Manchmal definiert man die einzelnen Einträge einer Folge nicht explizit, sondern gibt eine Vorschrift an, wie ein Eintrag aus dem vorigen gewonnen werden kann. Das bezeichnet man als rekursiv definierte Folge: Die Folge, die durch die Vorschrift

$$x_1 = 1, \quad x_2 = 1 \quad \text{und} \quad x_n = x_{n-2} + x_{n-1}$$

definiert ist heißt *Fibonacci-Folge*. Ihre ersten Glieder sehen folgendermaßen aus:

$$(1, 1, 2, 3, 5, 8, 13, 21, 34, \dots)$$

- Manche Folgen kann man leicht sowohl explizit als auch rekursiv definieren. So liefern zum Beispiel die Beschreibungen $(2^n)_{n \in \mathbb{N}}$ und $x_1 = 2$ und $x_{n+1} = 2 \cdot x_n$ für $n \geq 1$ äquivalente Beschreibungen derselben Folge, nämlich $(2, 4, 8, 16, 32, \dots)$.

Da wir Folgen als Abbildungen von (Teilmengen von) \mathbb{N} nach \mathbb{R} betrachten können, können wir sie addieren, subtrahieren, multiplizieren und dividieren (sofern der Nenner nicht 0 ist). Außerdem können wir für eine Folge (x_n) und eine Funktion f , so dass jedes Folgenglied im Definitionsbereich der Funktion liegt, durch $(f(x_n)) = (f(x_1), f(x_2), f(x_3), \dots)$ eine neue Folge definieren.

Beispiele. • Sei $(x_n) = (n)_{n \in \mathbb{N}}$ und $f(x) = \frac{1}{x}$, dann ist $(f(x_n)) = (\frac{1}{n})_{n \in \mathbb{N}}$.

- Wenn $x_n = \pi n$ ist und $f = \sin(x)$, dann ist $(f(x_n))_{n \in \mathbb{N}} = (\sin(\pi n))_{n \in \mathbb{N}} = (0)_{n \in \mathbb{N}}$ die konstante Null-Folge, da der Sinus von beliebigen ganzzahligen Vielfachen von π immer 0 ist.

7.2 Konvergenz

Bei der Untersuchung von Folgen geht es oft darum, gewisse Muster zu erkennen und zu verstehen. Wie wir ja schon bemerkt haben, ist zum Beispiel die Folge $((-1)^n)_{n \in \mathbb{N}}$ periodisch mit Periode 2. Andere Folgen, wie zum Beispiel $(\frac{1}{n})_{n \in \mathbb{N}}$ oder unsere Folgen $(u_n)_{n \in \mathbb{N}}$ und $(U_n)_{n \in \mathbb{N}}$ „kommen einer gewissen reellen Zahl immer näher“. Formal ist der Begriff der Konvergenz folgendermaßen definiert:

Definition. Eine Folge reeller Zahlen $(x_n)_{n \in \mathbb{N}}$ heißt *konvergent mit Grenzwert a* wenn es zu jedem $\epsilon > 0$ eine natürlich Zahl n_0 gibt, so dass für jedes $n \geq n_0$ gilt, dass

$$|x_n - a| < \epsilon.$$

Wir sagen in diesem Fall auch „ x_n konvergiert gegen a “ oder „ x_n geht gegen a “ und schreiben

$$\lim_{n \rightarrow \infty} x_n = a \quad \text{beziehungsweise kurz} \quad x_n \rightarrow a.$$

Mit Hilfe von Quantoren kann man die obige Definition auch schreiben als

$$\forall \epsilon > 0 \exists n_0 \in \mathbb{N} \forall n \geq n_0: |x_n - a| < \epsilon.$$

Diese Definition ist hinreichend kompliziert, dass man sie sich einmal in Ruhe ansehen sollte. Es wird verlangt, dass für jede reelle Zahl $\epsilon > 0$ ab einem gewissen Index n_0 der Abstand der Folgenglieder x_n zu unserem Grenzwert a kleiner sein muss als der vorgegebene Wert ϵ . Hat man für ein gewisses $\epsilon_0 > 0$ einmal ein geeignetes $n_0 \in \mathbb{N}$ gefunden, so dass die Bedingung erfüllt ist, so gilt sie offenbar auch für alle größeren $\epsilon > \epsilon_0$. Um also Konvergenz zu überprüfen, genügt es dann, für alle $0 < \epsilon < \epsilon_0$ die Bedingung nachzuweisen.

Bemerkung. Das folgende sprachliche Bild kann hilfreich sein: Eine Folge $(x_n)_{n \in \mathbb{N}}$ konvergiert gegen $a \in \mathbb{R}$, wenn wir zu jeder fest vorgegebenen „Messgenauigkeit“ $\epsilon > 0$ einen Index $n \in \mathbb{N}$ finden, so dass man mit dieser Messgenauigkeit alle späteren Folgenglieder nicht mehr von a unterscheiden kann.

Wie bei allen \forall -Aussagen genügt es, ein einziges $\epsilon > 0$ anzugeben, für das die Bedingung nicht erfüllbar ist, um Konvergenz auszuschließen. Eine manchmal etwas unbequeme Eigenschaft unserer Definition ist es übrigens, dass man einen Kandidaten für den Grenzwert braucht, um Konvergenz zu beweisen oder zu widerlegen. In der Analysis werden Sie alternative Kriterien kennenlernen, die Ihnen Konvergenzbeweise ermöglichen, ohne den Grenzwert bereits zu kennen.

Beispiele. • Die Folge $(\frac{1}{n})$ konvergiert gegen 0: für jedes $\epsilon > 0$ können wir eine natürliche Zahl $n_0 > \frac{1}{\epsilon}$ finden.² Für $n \geq n_0 > \frac{1}{\epsilon}$ gilt dann $0 < \frac{1}{n} < \epsilon$ und somit

$$|x_n - a| = \left| \frac{1}{n} - 0 \right| = \frac{1}{n} < \epsilon.$$

- Die Folge $((-1)^n)$ konvergiert nicht. Dies beweisen wir indirekt, nehmen also an, es gäbe einen Grenzwert $a \in \mathbb{R}$. Nun wählen wir

$$\epsilon_0 = \max(|a - 1|, |a + 1|)$$

als größeren der beiden Abstände von a zu 1 bzw. von a zu -1 . Man kann sich überlegen, dass dieses ϵ_0 mindestens 1 sein muss, egal welches $a \in \mathbb{R}$ wir als Grenzwert annehmen, so dass insbesondere $\epsilon_0 > 0$. Nun müsste es ein n_0 geben, so dass für $n > n_0$ gilt, dass $|x_n - a| < \epsilon_0$. Da die Folgenglieder x_n aber abwechselnd die Werte 1 und -1 annehmen, müsste also gelten, dass $|1 - a| < \epsilon_0$ und $|-1 - a| = |a + 1| < \epsilon_0$. Dies widerspricht unserer Wahl von ϵ_0 .

- Für $q \in (0, \infty)$ definieren wir die *geometrische Folge* durch $x_n = q^n$. Das Konvergenzverhalten dieser Folge hängt vom Wert von q ab.
 - Ist $q > 1$, dann konvergiert die Folge nicht: Jedes $q > 1$ lässt sich als $q = 1 + r$ mit $r > 0$ schreiben. Wir haben dann

$$q^n = (r + 1)^n = \sum_{k=0}^n \binom{n}{k} r^k 1^{n-k} = 1 + nr + \dots \geq 1 + nr,$$

²Dies ist eine Eigenschaft der reellen Zahlen.

da die weggelassenen Summanden positiv sind. Ist $M > 0$ gegeben, so wählen wir ein $n_0 \in \mathbb{N}$ mit $n_0 > \frac{M-1}{r}$ und stellen fest, dass dann für alle $n \geq n_0$ die Ungleichung

$$1 + nr > 1 + n_0 r > M \quad (7.1)$$

gilt. Wir haben also gezeigt: für jede reelle Zahl $M > 0$ existiert ein $n_0 \in \mathbb{N}$, so dass für alle natürlichen Zahlen $n \geq n_0$ die Ungleichung

$$q^n \geq 1 + nr > M.$$

gilt. Diese Tatsache schließt Konvergenz gegen irgendein $a \in \mathbb{R}$ aus (warum?).³

- Ist $q = 1$, dann ist jedes q^n ebenfalls gleich 1. Die Folge konvergiert somit gegen 1.
- Falls $q < 1$, so konvergiert die Folge gegen 0. Wir führen dies auf das Argument im ersten Fall zurück, und betrachten dafür $p = \frac{1}{q} > 1$. Ist nun $\epsilon > 0$ gegeben, so definieren wir $M := \frac{1}{\epsilon}$. Zu diesem M finden wir ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$ die Abschätzung

$$p^n > M$$

gilt, welche aber äquivalent zu

$$|q^n - 0| = q^n = \frac{1}{p^n} < \frac{1}{M} = \epsilon$$

ist. Dies beweist die behauptete Konvergenz.

In der Analysis werden Folgen und deren Grenzwerte sehr viel detaillierter behandelt. Einen der vielen Gründe dafür werden wir in den nächsten Kapiteln sehen, wo wir Grenzwerte benutzen werden, um Eigenschaften von Funktionen herauszuarbeiten. Zuerst formulieren wir aber noch einige Rechenregeln für Grenzwerte, die nicht schwer zu beweisen sind.

Satz 17. *Seien (x_n) und (y_n) konvergente Folgen mit den Grenzwerten a und b . Dann gilt:*

$$(a) \lim_{n \rightarrow \infty} (x_n + y_n) = a + b.$$

$$(b) \lim_{n \rightarrow \infty} (x_n - y_n) = a - b.$$

$$(c) \lim_{n \rightarrow \infty} (x_n \cdot y_n) = a \cdot b.$$

$$(d) \lim_{n \rightarrow \infty} \left(\frac{x_n}{y_n} \right) = \frac{a}{b}, \text{ falls } b \neq 0.$$

Die letzte Aussage des Satzes ist wie folgt zu verstehen: Für hinreichend große $n \in \mathbb{N}$ ist y_n von 0 verschieden, so dass die Quotientenfolge definiert ist, und diese Folge konvergiert gegen den angegebenen Grenzwert.

³Die hier formulierte Bedingung „Für alle $M > 0$ existiert ein $n_0 \in \mathbb{N}$ so dass für alle $n \in \mathbb{N}$ mit $n \geq n_0$ die Ungleichung $x_n > M$ gilt.“ charakterisiert die *bestimmte Divergenz* der Folge $(x_n)_{n \in \mathbb{N}}$ gegen ∞ .

7.3 Reihen: Ein Ausblick

Wir wollen nun noch einmal auf die beiden noch nicht beantworteten Fragen vom Anfang des Kapitels zurückkommen. Um die Frage

In welchem Sinne gilt die Gleichung $1 = 0.999 \dots$ (unendlicher Dezimalbruch)?

zu beantworten, müssen wir uns zunächst überlegen, wofür so ein *unendlicher* Dezimalbruch genau steht. Dies tun wir in Analogie zu *endlichen* Dezimalbrüchen. Es gilt bekanntlich

$$0.z_1z_2\dots z_n = z_1 \cdot \frac{1}{10} + z_2 \cdot \frac{1}{100} + \dots + z_n \cdot \frac{1}{10^n} = \sum_{i=1}^n \frac{z_i}{10^i},$$

wenn $z_1, \dots, z_n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ die Ziffern unserer Dezimalzahl sind. Haben wir nun einen unendlichen Dezimalbruch vor uns, so entspricht dieser offenbar einer unendlichen Summe (auch *Reihe* genannt)

$$\sum_{k=1}^{\infty} \frac{z_k}{10^k}.$$

Man kann unendlichen Summen nicht immer auf sinnvolle Weise einen Wert zuweisen. Im vorliegenden Fall können wir jedoch beschließen, den unendlichen Dezimalbruch „zu runden“, indem wir nach irgendeiner Anzahl n von Ziffern abbrechen. Formal erhalten wir also aus einem unendlichen Dezimalbruch

$$x = 0.z_1z_2z_3\dots$$

eine Folge von (rationalen) Zahlen $(x_n)_{n \in \mathbb{N}}$ mit

$$x_n = 0.z_1z_2\dots z_n = \sum_{k=1}^n \frac{z_k}{10^k}.$$

Diese endlichen Summen nennt man *Partialsommen* der Reihe.

Definition. Eine Reihe $\sum_{k=1}^{\infty} a_k$ heißt *konvergent*, falls der Grenzwert der Folge der Partialsommen $s_n = \sum_{k=1}^n a_k$ existiert. Man nennt dann diesen Grenzwert $\lim_{n \rightarrow \infty} s_n$ der Partialsommen den *Wert der unendlichen Reihe*.

In unserem konkreten Beispiel

$$x = 0.999\dots$$

erhalten wir als Partialsommen

$$x_1 = \frac{9}{10}, \quad x_2 = \frac{99}{100}, \quad x_3 = \frac{999}{1000}, \dots$$

Allgemein gilt

$$x_n = \sum_{i=1}^n \frac{9}{10^i} = 9 \cdot \sum_{i=1}^n \frac{1}{10^i}.$$

Wenn wir den Faktor 9 vorübergehend ignorieren, so erhalten wir eine neue Folge $(y_n)_{n \in \mathbb{N}}$ mit

$$y_n = \sum_{i=1}^n \frac{1}{10^i} = 0.\underbrace{11\dots1}_{n \text{ Einsen}}.$$

Summen dieser Form sind zum Glück relativ einfach auszurechnen, denn man kann zeigen, dass für jede reelle Zahl $q \in \mathbb{R}$ gilt:

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

Andererseits hatten wir bereits gesehen, dass die Folge $(q^n)_{n \in \mathbb{N}}$ für $0 < q < 1$ gegen 0 konvergiert. Dasselbe gilt dann natürlich auch für die Folge $(q^{n+1})_{n \in \mathbb{N}}$ (warum?). Die Folge der *Partialsommen*

$$\sum_{i=1}^n q^i$$

der sogenannten *geometrischen Reihe*

$$\sum_{i=1}^{\infty} q^i$$

konvergiert also für $0 < q < 1$ gegen

$$\lim_{n \rightarrow \infty} \frac{q^{n+1} - q}{q - 1} = \frac{-q}{q - 1} = \frac{q}{1 - q},$$

d.h. die geometrische Reihe konvergiert:

$$\sum_{i=1}^{\infty} q^i = \lim_{n \rightarrow \infty} \sum_{i=1}^n q^i = \frac{q}{1 - q}.$$

In unserem konkreten Fall rechnen wir mit $q = \frac{1}{10}$ und erhalten

$$\sum_{i=1}^{\infty} \frac{1}{10^i} = \sum_{i=1}^{\infty} \left(\frac{1}{10}\right)^i = \frac{\frac{1}{10}}{1 - \frac{1}{10}} = \frac{\frac{1}{10}}{\frac{9}{10}} = \frac{1}{9}.$$

Da unsere Folgenglieder x_n nun gerade jeweils das Neunfache der Partialsommen der geometrischen Reihe waren, erhalten wir

$$x = \lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} 9y_n = 9 \cdot \lim_{n \rightarrow \infty} y_n = 9 \cdot \frac{1}{9} = 1.$$

Zusammenfassend können wir also sagen: unsere Gleichung

$$1 = 0.999\dots$$

gilt, weil wir die rechte Seite als konvergente Reihe interpretieren können (und im Wesentlichen müssen!), deren Grenzwert gerade 1 ist.

Bemerkung. Wenn Ihnen dies jetzt zu schnell ging, so können Sie einfach abwarten, bis dieser Stoff noch einmal Thema in der Analysis sein wird. Es lohnt sich aber, dieses einfache Beispiel auch jetzt schon zu durchdenken. Sie werden dann in der Analysis effiziente Methoden kennenlernen, um zu beweisen, dass *jeder* unendliche Dezimalbruch tatsächlich gegen eine reelle Zahl konvergiert.

Schließlich bleibt noch Zenos Paradox. Haben Sie vielleicht selbst eine Idee, wie man damit umgeht? Es gibt verschiedene Möglichkeiten: Sie können Zenos Argumentation im Wesentlichen folgen und mit Reihen argumentieren, um den Denkfehler aufzudecken, oder aber Sie beschreiben die Situation anders (aber äquivalent!), so dass das Paradox (hoffentlich!) gar nicht erst auftritt...

8 Was ist eine stetige Funktion?

9 Äquivalenzrelationen

Objekte grob zu kategorisieren ist etwas, das Menschen regelmäßig und meist ohne darüber überhaupt nachzudenken tun. Wenn jemand zum Beispiel eine Tisch sieht, so wird diese Person erkennen, dass es sich um einen Tisch handelt, auch wenn er oder sie vielleicht diese konkrete Bauart eines Tisches noch nie gesehen hat. Wir können auf einen Blick die Fruchtsorte einer uns vorgelegten einzelnen Frucht angeben (zumindest bei nicht allzu exotischen Früchten), und wir wissen aus Erfahrung, dass es bei einem Pflaumenkuchen nicht genau darauf ankommt, dass wir bestimmte Früchte nehmen – irgendwelche (reife) Pflaumen genügen.

In der Mathematik formalisiert der Begriff der *Äquivalenzrelation* die konkrete Art und Weise, wie wir mathematische Objekte (genauer: Elemente einer Menge) zu neuen, allgemeineren Objekten (sogenannten *Äquivalenzklassen*) zusammenfassen. Meist geschehen solche „Kategorisierungen“, wie auch in den obigen Beispielen aus dem Alltag, weil für die folgenden Überlegungen oder Konstruktionen nur noch ein Teil der Eigenschaften der Objekte wesentlich sind, und es egal ist, wenn sie sich in anderen Eigenschaften noch unterscheiden.

Wir sind in diesem Vorkurs bereits mehreren Situationen begegnet, in denen wir Objekte zu allgemeineren Objekten zusammengefasst haben:

- verschiedene Brüche beschreiben „dieselbe rationale Zahl“,
- verschiedene Pfeile in der Ebene beschreiben „denselben Vektor“,
- verschiedene Dezimalbrüche beschreiben „dieselbe reelle Zahl“.

Die genaue mathematische Diskussion der Objekte in Anführungszeichen involviert jeweils gewisse Äquivalenzrelationen. Die folgende Diskussion dieses Konzepts folgt im Wesentlichen der Darstellung in [Hou12].

9.1 Definition und erste Beispiele von Äquivalenzrelationen

Bevor wir den Begriff der Äquivalenzrelation einführen, wollen wir aber zunächst den allgemeineren Begriff der Relation betrachten.

Definition. Sei X eine Menge. Eine *Relation auf X* ist eine Teilmenge $R \subseteq X \times X$. Bei der Untersuchung von Relationen schreiben wir $x \sim y$ um auszudrücken, dass das geordnete Paar (x, y) zur Teilmenge $R \subseteq X \times X$ gehört, und $x \not\sim y$, falls $(x, y) \notin R$.

Beispiel. Wir betrachten wieder einmal einen hypothetischen Marktstand, und bezeichnen mit F die Menge der dort angebotenen Früchte (jede einzelne Frucht ist also ein Element von F). Wir führen nun eine Relation auf F ein, bei der zwei Früchte $x, y \in F$ genau

dann zueinander in der Beziehung $x \sim y$ stehen, wenn beide zu derselben *Fruchtsorte* gehören.¹

Beispiele. Hier stellen wir einige Beispiele für mathematische Relationen vor.

- (1) Auf einer beliebigen Menge X betrachten wir die Relation, bei der $x \sim y$ genau dann, wenn $x = y$.
- (2) Auf der Menge $X = \mathbb{R}$ betrachten wir die Relation, bei der $x \sim y$ genau dann, wenn $x < y$.
- (3) Zu einer gegebenen Zahl $n \in \mathbb{N}$ betrachten wir auf der Menge $X = \mathbb{Z}$ die Relation, bei der $x \sim y$ genau dann, wenn x und y bei der Division durch n denselben Rest lassen.
- (4) Auf der Menge $X = \mathbb{Z}$ betrachten wir die Relation, bei der $x \sim y$ genau dann, wenn x ein Teiler von y ist.
- (5) Auf der Menge $X = \mathbb{R}$ betrachten wir die Relation, bei der $x \sim y$ genau dann, wenn $xy = 0$.

Obwohl in der Definition von einer Relation als Teilmenge $R \subseteq X \times X$ die Rede ist, beschreibt man Relationen in der Praxis meist wie in diesen Beispielen. Man spricht daher oft auch von \sim als einer Relation auf X , obwohl dies rein technisch gesehen nicht ganz korrekt ist. Solange aber klar ist, wie man die Teilmenge R aus der Beschreibung von \sim rekonstruieren könnte, ist dies nicht weiter schlimm.

Die folgenden Definition beschreibt einige Eigenschaften, welche Relationen haben können.

Definition. Sei \sim eine Relation auf der Menge X . Die Relation heißt

- *reflexiv*, falls für alle $x \in X$ gilt, dass $x \sim x$.
- *symmetrisch*, falls für alle $x, y \in X$ aus $x \sim y$ auch $y \sim x$ folgt.
- *transitiv*, falls für alle $x, y, z \in X$ aus $x \sim y$ und $y \sim z$ auch $x \sim z$ folgt.

Wir wollen für die oben betrachteten Beispiele prüfen, welche dieser Eigenschaften sie jeweils besitzen.

Beispiel. Unsere Relation auf der Menge F der Früchte eines Marktstandes besitzt alle drei genannten Eigenschaften.

Beispiele. Wir betrachten noch einmal die Beispiele 9.1 für Relationen.

- (1) Die erste oben erwähnte Relation hat alle drei Eigenschaften, wie sofort aus der Definition klar ist.

¹Ob wir hier nun verschiedene Apfelsorten usw. als verschiedene Fruchtsorten betrachten oder nicht, ist für den Rest der Diskussion unwesentlich.

- (2) Die Relation $x < y$ auf \mathbb{R} ist nicht reflexiv und nicht symmetrisch, sie ist jedoch transitiv.
- (3) Sei $n \in \mathbb{N}$ gegeben. Die Relation auf \mathbb{Z} , bei der $x \sim y$, falls beide bei der Division durch n denselben Rest hinterlassen, hat alle drei erwähnten Eigenschaften. Reflexivität und Symmetrie sind direkt aus der Definition klar, und wenn sowohl x und y also auch y und z jeweils denselben Rest hinterlassen, so ist dieser Rest offenbar bei allen drei Divisionen durch n gleich, d.h. insbesondere $x \sim z$.
- (4) Jede ganze Zahl ist ein Teiler von sich selbst, so dass die vierte Relation reflexiv ist. Sie ist nicht symmetrisch, wie das Gegenbeispiel der beiden Zahlen 2 und 6 zeigt. Sie ist jedoch transitiv: teilt x die Zahl y und teilt y die Zahl z , so teilt x offenbar auch z .
- (5) Die letzte erwähnte Relation ist nicht reflexiv, denn z.B. gilt $1 \not\sim 1$. Sie ist symmetrisch, denn $xy = yx$. Sie ist nicht transitiv, denn zwar gilt $1 \sim 0$ und $0 \sim 2$, aber $1 \not\sim 2$.

Wir kommen nun zur zentralen Definition dieses Kapitels.

Definition. Sei X eine Menge. Eine Relation \sim auf X heißt *Äquivalenzrelation*, wenn sie reflexiv, symmetrisch und transitiv ist.

Wie wir in den Beispielen 9.1 gesehen haben, sind die erste und dritte unserer eingangs betrachteten mathematischen Relationen Äquivalenzrelationen, während die anderen es nicht sind. Die zweite und vierte Relation sind Beispiele für sogenannte *Ordnungsrelationen*, auf die wir hier aber nicht näher eingehen werden. Die letzte Relation hat keinen speziellen Namen und wurde vor allem zur Illustration der Eigenschaften betrachtet. Auch die Relation auf der Menge F der Früchte unseres Marktstandes ist eine Äquivalenzrelation im Sinne unserer Definition.

Bemerkung. Wenn ein Begriff wie hier durch eine Reihe von Eigenschaften charakterisiert wird, so lohnt es sich zu überprüfen, dass die angegebenen Bedingungen tatsächlich unabhängig voneinander sind, d.h. in unserem Fall, dass keine der drei Bedingungen bereits aus den anderen folgt. Dies ist in der Tat so, und es ist instruktiv, Beispiele zu suchen, die diese Behauptung belegen.

9.2 Äquivalenzklassen

Wie eingangs erwähnt interessieren wir uns für Äquivalenzrelationen als eine Möglichkeit, Elemente von Mengen zu neuen Objekten zusammenzufassen. Es ist also nicht verwunderlich, wenn diese neuen Objekte einen eigenen Namen bekommen.

Definition. Sei X eine Menge und \sim eine Äquivalenzrelation auf X . Die Äquivalenzklasse eines Elements $x \in X$ bezüglich der Äquivalenzrelation \sim ist die Teilmenge

$$[x] := \{y \in X \mid x \sim y\} \subseteq X.$$

Beispiele. • Für die Relation auf der Menge der Früchte F unseres Marktstandes sind die Äquivalenzklassen gerade die Fruchtsorten.

- Für die Relation auf einer Menge X mit $x \sim y$ genau dann, wenn $x = y$, bestehen alle Äquivalenzklassen jeweils nur aus einem Element, d.h. wir haben $[x] = \{x\} \subseteq X$ für alle $x \in X$.
- Für jedes $n \in \mathbb{N}$ haben wir oben durch „ $x \sim y$ genau dann, wenn x und y denselben Rest bei Division durch n lassen“, eine Äquivalenzrelation auf \mathbb{Z} definiert. Aufgrund ihrer Bedeutung in der Arithmetik haben diese Äquivalenzrelationen ihre eigene Notation: man schreibt statt $x \sim y$ hier

$$x \equiv y \pmod{n}.$$

So wird die Wichtigkeit von $n \in \mathbb{N}$ angemessen unterstrichen. Wir haben folgende Äquivalenzklassen:

$$\begin{aligned} [0] &= \{0, n, -n, 2n, -2n, 3n, -3n, \dots\} \subseteq \mathbb{Z} \\ [1] &= \{1, 1+n, 1-n, 1+2n, 1-2n, 1+3n, 1-3n, \dots\} \subseteq \mathbb{Z} \\ [2] &= \{2, 2+n, 2-n, 2+2n, 2-2n, 2+3n, 2-3n, \dots\} \subseteq \mathbb{Z} \\ &\dots \\ [n-1] &= \{n-1, 2n-1, -1, 3n-1, -1-n, 4n-1, -1-2n, \dots\} \subseteq \mathbb{Z} \\ [n] &= \{0, n, -n, 2n, -2n, 3n, -3n, \dots\} = [0] \subseteq \mathbb{Z} \end{aligned}$$

- Ist X ein Satz Spielkarten, so können wir darauf zwei verschiedene Äquivalenzrelationen betrachten:
 - die erste Äquivalenzrelation ist definiert durch $x \sim y$ genau dann, wenn x und y dieselbe Farbe haben. Die Äquivalenzklassen sind dann Kreuz, Pik, Herz und Karo.
 - die zweite Äquivalenzrelation ist definiert durch $x \sim y$ genau dann, wenn x und y denselben Kartenwert haben. Beispiele für Äquivalenzklassen sind dann die Teilmenge der Siebenen, die Teilmenge der Könige oder die Teilmenge der Asse.

Der folgende Satz enthält eine wichtige Charakterisierung der Äquivalenzrelationen.

Satz 18. Sei \sim eine Äquivalenzrelation auf der Menge X .

(a) $\bigcup_{x \in X} [x] = X.$

(b) $x \sim y$ gilt genau dann, wenn $[x] = [y].$

(c) $x \not\sim y$ gilt genau dann, wenn $[x] \cap [y] = \emptyset.$

In Worten besagt der Satz, dass die Äquivalenzrelation die Menge X in paarweise disjunkte Teilmengen, nämlich gerade die Äquivalenzklassen, zerlegt.

Beweis.

zu (a): Für jedes $x \in X$ gilt, weil \sim reflexiv ist, dass $x \in [x]$. Also folgt auch $x \in \bigcup_{y \in X} [y]$, insgesamt also $X \subseteq \bigcup_{y \in X} [y]$.

Die umgekehrte Inklusion ist offensichtlich, da $\bigcup_{y \in X} [y]$ als Vereinigung von Teilmengen von X selbst Teilmenge von X ist. Dies zeigt die Behauptung.

zu (b): Da eine Äquivalenz zweier Aussagen behauptet wird, beweisen wir zwei Implikationen.

Behauptung 1: Wenn $[x] = [y]$, dann gilt $x \sim y$.

Dies gilt, denn wegen $y \in [y]$ und $[x] = [y]$ haben wir $y \in [x]$, d.h. $x \sim y$ nach Definition unserer Äquivalenzklassen.

Behauptung 2: Wenn $x \sim y$, dann gilt $[x] = [y]$.

Auch diese Behauptung zerlegen wir in zwei Teilaussagen, und zeigen zunächst $[x] \subseteq [y]$. Sei also $z \in X$ ein Element aus $[x]$, d.h. er gilt $x \sim z$. Wegen der Symmetrie gilt dann auch $z \sim x$. Aus der Transitivität folgt dann mit $x \sim y$ auch $z \sim y$. Wieder mit der Symmetrie folgt nun $y \sim z$, d.h. $z \in [y]$. Dies beweist $[x] \subseteq [y]$.

Umgekehrt zeigen wir schließlich $[y] \subseteq [x]$. Ist $z \in [y]$, so gilt $y \sim z$. Mit $x \sim y$ und der Transitivität folgt dann $z \in [x]$. Dies zeigt $[y] \subseteq [x]$, und somit ist Behauptung 2 bewiesen.

zu (c): Wie in (b) zeigen wir die beiden Implikationen getrennt.

Behauptung 1: Wenn $x \not\sim y$, dann folgt $[x] \cap [y] = \emptyset$.

Diese Behauptung beweisen wir durch Kontraposition, d.h. wir nehmen $[x] \cap [y] \neq \emptyset$ an und folgern daraus $x \sim y$. Ist nämlich $z \in [x] \cap [y]$, so gilt $x \sim z$ und $y \sim z$. Symmetrie impliziert $z \sim y$, und aus der Transitivität folgt nun $x \sim y$ wie behauptet.

Behauptung 2: Wenn $[x] \cap [y] = \emptyset$, dann folgt $x \not\sim y$.

Wieder argumentieren wir unter Verwendung der Kontraposition, d.h. wir nehmen an, dass $x \sim y$. Dann gilt nach Definition $y \in [x]$. Da aber auch $y \in [y]$, folgt hieraus $y \in [x] \cap [y]$, d.h. dieser Durchschnitt ist nicht leer. \square

Es gilt übrigens auch die folgende Umkehrung des Satzes, deren Beweis wir den Leserinnen und Lesern als Übungsaufgabe überlassen.

Satz 19. *Ist X eine Menge, und ist $\{X_i \mid i \in I\} \subseteq \mathcal{P}(X)$ eine Familie von Teilmengen von X , so dass*

$$\text{aus } i \neq j \text{ folgt } X_i \cap X_j = \emptyset,$$

dann wird durch

$$x \sim y \text{ genau dann, wenn } \exists i \in I : x \in X_i \text{ und } y \in X_i$$

eine Äquivalenzrelation auf X definiert. \square

9.3 Die Konstruktion der rationalen Zahlen aus den ganzen Zahlen

Wir wollen nun mit Hilfe des Begriffs der Äquivalenzrelation die Konstruktion der rationalen Zahlen aus den ganzen Zahlen beschreiben. Dazu betrachten wir die Brüche als

Elemente der Menge

$$\mathcal{B} = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid b \neq 0\} \subset \mathbb{Z} \times \mathbb{Z}.$$

Lemma 20. Die Relation \sim auf \mathcal{B} , definiert durch

$$(a, b) \sim (c, d) \iff ad = bc$$

ist eine Äquivalenzrelation.

Beweis. Reflexivität und Symmetrie folgen direkt aus der Definition. Um Transitivität zu prüfen, betrachten wir drei Elemente (a, b) , (c, d) und (e, f) aus \mathcal{B} , und nehmen an, dass $(a, b) \sim (c, d)$ und $(c, d) \sim (e, f)$. Dies impliziert

$$ad = bc \quad \text{und} \quad cf = de.$$

Multiplikation der ersten Gleichung mit f und der zweiten Gleichung mit b sind Äquivalenzumformungen, da nach Voraussetzung b und f von 0 verschieden sind. Wir erhalten dann

$$adf = bcf \quad \text{und} \quad bcf = bde,$$

woraus $adf = bde$ folgt. Da $d \neq 0$, folgt hieraus auch $af = be$, und dies bedeutet $(a, b) \sim (e, f)$. \square

Wir können nun eine korrekte Definition der rationalen Zahlen geben.

Definition. Wir definieren die Menge \mathbb{Q} der *rationalen Zahlen* als die Menge der Äquivalenzklassen der Relation \sim auf \mathcal{B} .

Eine rationale Zahl ist nun also eine Äquivalenzklasse von Brüchen. Der entscheidende Vorteil dieser Definition gegenüber unserer früheren „Definition“ ist, dass kein unerklärter Begriff „Zahl“ mehr vorkommt. Der Nachteil der fehlenden Wiedererkennung kann leicht dadurch behoben werden, dass man statt für die Elemente aus \mathcal{B} statt (a, b) wieder wie üblich $\frac{a}{b}$ schreibt.

Wir haben eine injektive Abbildung

$$I : \mathbb{Z} \rightarrow \mathbb{Q} \\ I(n) = \left[\frac{n}{1} \right].$$

Nun möchten wir Rechenregeln für rationale Zahlen festlegen, welche die Regeln für das Rechnen mit ganzen Zahlen verallgemeinern. Die Idee ist einfach: wir wählen aus gegebenen Äquivalenzklassen jeweils einen Repräsentanten, definieren deren Summe bzw. Produkt als Bruch, und gehen dann wieder zur Äquivalenzklasse über. Damit dies tatsächlich eine sinnvolle Definition der Rechenregeln für die *Äquivalenzklassen* ist, müssen wir zeigen, dass das Resultat unabhängig von den getroffenen Wahlen ist. Mit anderen Worten: wenn wir aus *denselben* gegebenen Äquivalenzklassen *andere Repräsentanten* wählen, und dann dieselbe Rechenregel anwenden, so erhalten wir als Ergebnis *dieselbe Äquivalenzklasse* wie vorher.

Bemerkung. Diese Unabhängigkeit der Definition von den zwischendurch getroffenen willkürlichen Wahlen nennt man *Wohldefiniertheit* des erhaltenen Begriffes.

Man kann nun durch geduldiges Nachrechnen folgenden Satz zeigen:

Satz 21. *Die folgenden Rechenoperationen auf der Menge \mathbb{Q} der rationalen Zahlen sind wohldefiniert:*

$$\left[\frac{a}{b}\right] + \left[\frac{c}{d}\right] := \left[\frac{ad + bc}{bd}\right] \quad , \quad \left[\frac{a}{b}\right] \cdot \left[\frac{c}{d}\right] := \left[\frac{ac}{bd}\right].$$

Sie sind assoziativ, kommutativ und erfüllen das Distributivgesetz. Außerdem respektiert die injektive Abbildung $I : \mathbb{Z} \rightarrow \mathbb{Q}$ die Rechenoperationen, d.h. für $m, n \in \mathbb{Z}$ gilt

$$I(m + n) = I(m) + I(n) \quad \text{und} \quad I(m \cdot n) = I(m) \cdot I(n). \quad \square$$

Bemerkung. Da die rationalen Zahlen sehr häufig benutzt werden, macht man bei ihrer Verwendung eine Ausnahme von der Regel, Äquivalenzklassen mit eckigen Klammern zu kennzeichnen. Wir schreiben also wie üblich $\frac{a}{b}$, wenn wir die rationale Zahl $\left[\frac{a}{b}\right]$ meinen.

9.4 Vektoren als Pfeilklassen

Hier wollen wir noch ganz kurz unsere Diskussion von Vektoren im Kapitel 4 in einen historischen Kontext stellen.

Eine der großen mathematischen Überlieferungen der griechischen Antike ist die axiomatische Beschreibung der Geometrie einer Ebene E in Euklids *Elementen*. Aus den dort eingeführten Grundbegriffen Punkt und Gerade sowie den ihre Beziehungen beschreibenden Axiomen ergibt sich sowohl das Konzept der *parallelen Geraden* als auch das Konzept der *Spiegelung an einer gegebenen Geraden* (als Abbildung $S : E \rightarrow E$ der Ebene). Man kann dann eine *Verschiebung* $V : E \rightarrow E$ als die Verknüpfung von zwei Spiegelungen an parallelen Geraden *definieren*. All dies kommt völlig ohne Koordinaten aus.

Nun definiert man einen Pfeil einfach als ein geordnetes Paar von Punkten, wobei wir den ersten als Anfangspunkt und den zweiten als Endpunkt interpretieren wollen. Die Menge der Pfeile entspricht also gerade

$$\mathcal{P} = E \times E,$$

denn jeder Punkt kann als Anfangs- oder Endpunkt auftreten. Auf dieser Menge definieren wir nun eine Äquivalenzrelation, bei der $(x_1, y_1) \sim (x_2, y_2)$ genau dann, wenn es eine Verschiebung $V : E \rightarrow E$ mit $V(x_1) = x_2$ und $V(y_1) = y_2$ gibt. Die Menge $\mathcal{V}(E)$ der Vektoren der Ebene definieren wir nun wieder als die Menge der Äquivalenzklassen von Pfeilen. Aus dem geometrischen Konzept der *zentrischen Streckung* erhält man die Definition einer *Skalarmultiplikation* als Abbildung

$$\mathbb{R} \times \mathcal{V}(E) \rightarrow \mathcal{V}(E),$$

und außerdem kann man die Addition von Vektoren

$$+ : \mathcal{V}(E) \times \mathcal{V}(E) \rightarrow \mathcal{V}(E)$$

geometrisch mit Hilfe der Parallelogrammregel. Für beide muss man (ähnlich wie bei den Rechenregeln für rationale Zahlen) die *Wohldefiniertheit* beweisen, da sie ja jeweils über gewisse Operationen auf Repräsentanten definiert werden. Insgesamt erhält man so auf $\mathcal{V}(E)$ die (aus den Axiomen abgeleitete) Struktur eines Vektorraumes (die Definition dieses Begriffes werden Sie bald in der linearen Algebra kennenlernen), und zwar immer noch ganz ohne Koordinaten. Die heute meist bevorzugte Darstellung, bei der man von Anfang an mit Koordinaten arbeitet, kann man in diesem Zugang bei Bedarf nachträglich einführen.

10 Beweismethoden

Beweise sind das Rückgrat der Mathematik. Mit Hilfe von Beweisen überzeugen wir uns und andere, dass unsere Behauptungen „wahr“ und unsere Lösungen für Probleme richtig und vollständig sind. Es ist übrigens wichtig, zu bemerken, dass diese „Wahrheit“ etwas weniger absolut ist, als in der Schule typischerweise thematisiert wird. Jede mathematische Theorie ist auf gewisse Grundannahmen, meist *Axiome* genannt, angewiesen, die als Voraussetzungen für alle folgenden Aussagen dienen, aus denen die Theorie also entwickelt wird, und die selbst nicht bewiesen werden. Für den üblichen Umgang mit Zahlen und daraus entwickelten Konzepten ist das recht allgemein akzeptierte Axiomensystem das sogenannte „ZFC“ – die Axiome der Mengenlehre nach Zermelo und Frenkel einschließlich des sogenannten *Auswahlaxioms* („*axiom of choice*“). Es war zunächst ein Schock, als Gödel 1931 bewiesen hat, dass man ausgehend von jedem halbwegs interessanten Axiomensystem Aussagen finden kann, die *mit diesen Grundannahmen weder beweisbar noch widerlegbar sind*. Auch kann man für die meisten Axiomensysteme die *Widerspruchsfreiheit der Axiome* innerhalb des Systems selbst nicht beweisen. Beide Aussagen treffen auf „ZFC“ zu, trotzdem hatten sie für die praktische Entwicklung der Mathematik erstaunlich geringe Bedeutung. Die meisten Mathematiker und Mathematikerinnen gehen davon aus, dass das Axiomensystem „ZFC“ widerspruchsfrei ist, oder arbeiten zumindest so, als ob es das wäre.

Ein Beweis einer mathematischen Aussage (eines Satzes) beginnt also bei als wahr bekannten Aussagen (Axiomen oder bereits aus ihnen abgeleiteten Feststellungen) und formuliert eine Folge von korrekten Schlussfolgerungen, an deren Ende die Wahrheit der neu behaupteten Aussage steht. Die Aussagen, die man beweisen will sind meist in der Form „wenn A , dann B “ gegeben. Allerdings ist A manchmal nicht explizit formuliert, sondern aus dem Kontext zu erschließen. Häufig findet man auch Aussagen der Form „ A genau dann, wenn B “, welche man dann oft (aber nicht immer!) in die beiden Teilaussagen „aus A folgt B “ und „aus B folgt A “ zerlegt.

Ein mathematischer Satz sollte grundsätzlich erst dann als wahr angenommen werden, wenn er bewiesen ist¹. Es ist besser, man gewöhnt sich diese „Berufsethik“ schon bei einfachsten Aussagen an, denn bei komplizierteren Aussagen ist schnell tatsächlich nicht mehr „klar“, ob sie nun wahr oder falsch sind, und nur ein Beweis einer der beiden Alternativen kann hier Gewissheit bringen. Erfahrungsgemäß tut man sich als Anfänger relativ schwer, präsentierte Beweise zu verstehen oder selbstständig Beweise zu führen. Dies mag auch daran liegen, dass es meist *nicht die eine richtige Lösung* gibt, wenn es

¹In diesem Vorkurs sind viele Beweise nur angedeutet oder werden ganz ausgelassen. Gründe dafür sind sowohl Zeitmangel als auch manchmal fehlende Diskussion der Grundlagen, welche für einen echten Beweis notwendig wären, wie zum Beispiel eine fundierte Einführung der reellen Zahlen. Im Studium wird dies anders sein.

um Beweise geht.

Beispiel. Wir werden zur Illustration drei Beweise der Aussage „Wenn n ungerade ist, dann ist auch n^2 ungerade.“ geben.

Beweis 1. Ist n ungerade, so gibt es eine natürliche Zahl k , so dass $n = 2k - 1$. Dann gilt jedoch

$$n^2 = (2k - 1)^2 = 4k^2 - 4k + 1 = 2(2k^2 - 2k) + 1,$$

so dass auch n^2 ungerade ist. □

Beweis 2. Wir beweisen: *Ist n^2 gerade, so ist auch n gerade.*

Ist n^2 gerade, so ist 2 ein Primfaktor von $n^2 = n \cdot n$. Da die Primfaktorzerlegung eindeutig ist, tritt jeder Primfaktor von n^2 bereits in der Primfaktorzerlegung von n auf, so dass 2 auch ein Primfaktor von n ist. Also ist n gerade. □

Beweis 3. Ist n ungerade, so ist 2 kein Primfaktor von n . Ist n^2 gerade, so ist 2 aber ein Primfaktor von n^2 , also auch ein Primfaktor von n . Dieser Widerspruch zeigt, dass nicht gleichzeitig n ungerade und n^2 gerade sein können, und somit ist die Aussage bewiesen. □

Einen Beweis zu finden ist ein kreativer Prozess, dessen Erfolg stark vom individuellen Vorwissen abhängt. Je mehr wahre Aussagen, je mehr Möglichkeiten der Umformulierung der gegebenen Voraussetzungen und Behauptungen und je mehr prinzipielle Beweisstrategien wir kennen, um so eher werden wir ans Ziel gelangen. In diesem Kapitel geht es in erster Linie darum, einige wichtige Beweistechniken vorzustellen, die häufig Anwendung in der Mathematik finden. Das soll dazu dienen, ein besseres Verständnis für die „Anatomie“ von Beweisen zu entwickeln, damit man Beweise besser lesen kann, aber auch besser selbst Beweise führen kann.

Zur Formulierung von Beweisen gilt: je einfacher ein Beweis ist und je leichter er zu verstehen ist, desto besser. Wenn ein Sachverhalt leichter mit einem deutschen Satz als mit einer Formel ausgedrückt werden kann, dann sollte man das auch tun. Manchmal kann es auch sinnvoll sein, eine Formel zusätzlich noch in Worten auszudrücken.

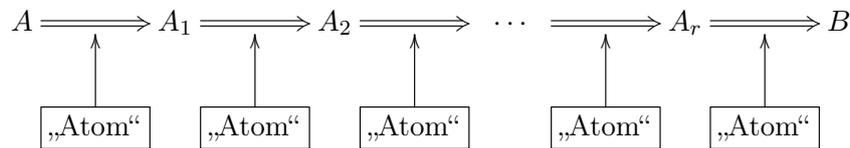
Die Auflistung der Beweismethoden in diesem Kapitel ist bei weitem nicht vollständig. Alle existierenden Beweismethoden aufzulisten ist praktisch nicht möglich. Außerdem sei angemerkt, dass mathematische Beweise häufig eine bunte Mischung verschiedener Methoden enthalten. Oft ist es so, dass man den Beweis einer Aussage in mehrere Teilschritte zerlegt, und es ist völlig normal, dass für die verschiedenen Teilschritte unterschiedliche Methoden verwendet werden.

Dieses Kapitel hat viele Anregungen und Beispiele aus Houstons Buch [Hou12] übernommen.

10.1 Direkter Beweis

Der direkte Beweis ist die konzeptuell einfachste Beweisform. Um die Aussage „aus A folgt B “ zu zeigen, geht man hier von der Aussage A aus und leitet durch logische

Schlussfolgerungen (d.h. mit Hilfe von als wahr bekannten Implikationen) die Aussage B her. Um also direkt zu beweisen, dass B aus A folgt, zeigt man, dass A_1 aus A folgt, dass A_2 aus A_1 folgt, dass A_3 aus A_2 folgt und so weiter, bis schliesslich aus einem gewissen A_r dann B folgt. Die einzelnen Teilschritte sollten dann offensichtlich sein, es sollten also keine großen, schwer nachvollziehbaren Gedankensprünge notwendig sein. Diese Teilschritte, kann man dann sozusagen als „Atome“ auffassen.



Wir haben bereits einige direkte Beweise gesehen - auch der erste Beweis aus dem Beispiel 10 war ein direkter Beweis. Wir wollen nun noch einige weitere Beispiele von Aussagen, die direkt bewiesen werden können, betrachten.

Beispiele für direkte Beweise

Satz. $\sum_{i=n}^{n+4} i$ ist durch 5 teilbar.

Dieser Satz ist zunächst nicht in der Form $A \implies B$ formuliert. Nachdem die Summe aber nur für $n \in \mathbb{N}$ definiert ist, könnte man ihn zu „Wenn n eine natürliche Zahl ist, dann ...“ umformulieren, die Aussage A steckt somit implizit in der Formulierung der Aussage B . Präziser könnte man also diesen Satz wie folgt formulieren, und damit ist dann offensichtlich, was die Aussage A ist und was die Aussage B ist.

Satz. Wenn n eine natürliche Zahl ist, dann ist $\sum_{i=n}^{n+4} i$ durch 5 teilbar.

Beweis. Wir schreiben die Summe aus und formen um:

$$\sum_{i=n}^{n+4} i = n + (n+1) + (n+2) + (n+3) + (n+4) = 5n + 10 = 5(n+2).$$

Da $\sum_{i=n}^{n+4} i$ eine Darstellung als $5k$ für ein $k \in \mathbb{N}$ hat, muss es durch 5 teilbar sein. \square

Satz. Seien $k, n \in \mathbb{N}$ mit $k > 1$. Falls n durch k teilbar ist, dann ist $n+1$ nicht durch k teilbar.

Die Aussage A besteht in diesem Satz aus mehreren Teilaussagen, und zwar ist A die durch „und“ verknüpfte Aussage der Aussagen „ $k \in \mathbb{N}$ “, „ $n \in \mathbb{N}$ “, „ $k > 1$ “ und „ k teilt n “. Die Aussage B ist „ $n+1$ ist nicht durch k teilbar.“

Beweis. Da k ein Teiler von n ist, gibt es eine ganze Zahl q mit $n = qk$. Somit ist aber $n+1 = qk+1$, und da $0 \leq 1 < k$ gilt, ist dies das eindeutige Ergebnis der Division mit Rest von $n+1$ durch k . Insbesondere ist der Rest $r = 1$ verschieden von 0, so dass $n+1$ nicht durch k teilbar ist. \square

Im nächsten Beispiel wollen wir uns nicht nur den Beweis anschauen, sondern auch die (möglichen) Denkvorgänge, welche zu einem Beweis führen können.

Satz. *Es sei $p \in \mathbb{Q}$ mit $p^2 \in \mathbb{Z}$. Dann ist auch $p \in \mathbb{Z}$.*

Was sind die Voraussetzungen, was ist also die Aussage A ? Die Aussage A ist „ $p \in \mathbb{Q}$ und $p^2 \in \mathbb{Z}$ “. Die Schlussfolgerung, also die Aussage B , ist „ $p \in \mathbb{Z}$ “. Bis jetzt haben wir noch nicht so viele Sätze über die rationalen Zahlen kennengelernt. Eigentlich haben wir bis jetzt nur die Definition kennengelernt, welche besagt, dass $p = \frac{a}{b}$ für ganze Zahlen a und b . Wir können annehmen, dass dieser Bruch so weit wie möglich gekürzt ist. Die zweite Voraussetzung besagt, dass $p^2 \in \mathbb{Z}$. Demnach ist also $(\frac{a}{b})^2 \in \mathbb{Z}$ und damit $\frac{a^2}{b^2} \in \mathbb{Z}$. Dieser Bruch ist ebenfalls gekürzt, und dies ist nur möglich, wenn $b^2 = 1$. Man erhält $b = \pm 1$. Also sieht man, dass $p = \frac{a}{\pm 1} = \pm a \in \mathbb{Z}$ sein muss.

Das sind nun die Gedanken, mit welchen wir aus der Aussage A die Aussage B gefolgert haben. Wenn man die eigenen Gedanken aufgeschrieben hat, muss man diese noch einmal durchgehen, damit man dann zu einem sauberen Beweis kommt.

Beweis. Nach Voraussetzung ist $p = \frac{a}{b}$ mit ganzen Zahlen a und b , wobei dieser Bruch gekürzt sei. Es folgt $p^2 = (\frac{a}{b})^2 = \frac{a^2}{b^2}$. Da $p^2 \in \mathbb{Z}$ und dieser Bruch ebenfalls gekürzt ist, muss $b^2 = 1$ sein. Demnach gilt $b = \pm 1$, womit wir $p = \frac{a}{\pm 1} = \pm a \in \mathbb{Z}$ erhalten. \square

Bemerkung. Ein häufiger Anfängerfehler in direkten Beweisen ist es, mit der Behauptung zu beginnen und so lange umzuformen, bis man auf eine wahre Aussage kommt. *Dies beweist leider gar nichts!* Wie wir im Kapitel über Logik gesehen haben, kann man aus einer falschen Aussage *jede* Aussage folgern, also auch jede wahre Aussage. Nur aus der Tatsache, dass aus der Behauptung etwas Wahres folgt, lässt sich also gar nichts über diese Behauptung selbst ableiten.

Beweis einer Gleichung

Es gibt ganz verschiedene Methoden eine Gleichung zu beweisen. Ein guter Ansatz ist, mit der schwierigeren Seite zu beginnen und solange umzuformen, bis man zur einfacheren Seite gelangt.

Beispiel. Wir wollen (für $x \notin \{k\pi \mid k \in \mathbb{Z}\}$) die Gleichung

$$\frac{1}{\tan^2(x) + 1} = \cos^2(x)$$

beweisen. Die kompliziertere Seite ist offenbar die linke Seite:

$$\begin{aligned}
 \frac{1}{\tan^2(x) + 1} &= \frac{1}{\frac{\sin^2(x)}{\cos^2(x)} + 1} && \text{nach Definition des Tangens} \\
 &= \frac{1}{\frac{\sin^2(x) + \cos^2(x)}{\cos^2(x)}} \\
 &= \frac{1}{\frac{1}{\cos^2(x)}} && \text{weil } \sin^2(x) + \cos^2(x) = 1 \\
 &= \cos^2(x) && \text{solange } \cos^2(x) \neq 0, \text{ d.h. } \cos(x) \neq 0.
 \end{aligned}$$

Es gibt auch andere Methoden, eine Gleichung zu beweisen. Je nach Situation nutzt man zum Beispiel manchmal eine der folgenden Äquivalenzen:

- $x = y \iff x - y = 0$.
- $x = y \iff (x \leq y \text{ und } x \geq y)$.
- $x = y \iff (x = z \text{ und } z = y)$.

Die letzten beiden Methoden haben den Vorteil gegenüber einem direkten Beweis, dass man die Aussage in zwei getrennte Teilaussagen zerlegt hat, welche eventuell für sich genommen jeweils einfacher zu beweisen sind.

Beweis einer Äquivalenz

Wie schon in der Einleitung erwähnt, kann man eine Aussage der Form „A genau dann, wenn B“ beweisen, indem man die Implikationen „aus A folgt B“ und „aus B folgt A“ beweist. Wir betrachten ein Beispiel.

Satz. Sei n eine natürliche Zahl größer 1. Die Zahl n ist eine Primzahl genau dann, wenn für alle $a, b \in \mathbb{N}$ mit $n = ab$ gilt, dass $a = n$ oder $b = n$.

Proof. Beweis Die Aussage hat die Form $A \iff B$, wobei A = „ n ist eine Primzahl“ und B = „für jede Darstellung $n = ab$ von n als Produkt von zwei natürlichen Zahlen a und b gilt $a = n$ oder $b = n$ “. Wir wollen nun einen Beweis finden, indem wir jede Richtung einzeln betrachten. Wir beginnen mit dem „genau dann“-Teil, also der Richtung $A \implies B$. Wir müssen also zu natürlichen Zahlen a und b mit $n = ab$ zeigen, dass $a = n$ oder $b = n$ gilt. Da aber n eine Primzahl ist und a, b Teiler von n sind, muss $a = n$ und $b = 1$ oder $a = 1$ und $b = n$ gelten. Dies beweist die Richtung $A \implies B$.

Jetzt wollen wir den „wenn“-Teil, also die Richtung $A \impliedby B$ beweisen. Wir betrachten einen beliebigen Teiler a von n , was bedeutet, dass es eine natürliche Zahl b gibt, so dass $n = ab$. Nach Voraussetzung gilt dann entweder $a = n$ (und $b = 1$) oder $b = n$ (und $a = 1$). Wir sehen also, dass unser „beliebiger“ Teiler a nur die Werte n oder 1 annehmen kann. Zusammen mit der Bedingung $n > 1$, welche wir sowieso voraussetzen, ist dies gerade die Definition dafür, dass n eine Primzahl ist. Wir haben also die Richtung $A \impliedby B$ bewiesen. \square

Zum Abschluss halten wir hier noch einige Bemerkungen fest.

- Man kann eine Aussage der Form „ A genau dann, wenn B “ auch direkt beweisen, indem man die Äquivalenz $A \iff B$ in kleine Teilschritte

$$A \iff A_1 \iff A_2 \iff \dots \iff A_r \iff B$$

zerlegt. Wie bei den Gleichungen empfiehlt es sich auch hier, mit der komplizierteren der beiden Aussagen A oder B zu beginnen und dann zur einfacheren zu gelangen.

- Eine weitere Strategie für Äquivalenzbeweise ist, statt $A \iff B$ die beiden Teilaussagen $A \implies B$ und $(\neg A) \implies (\neg B)$ zu beweisen. Dies werden wir bei der Diskussion des Beweises durch Kontraposition noch einmal näher erläutern.
- Ist die Äquivalenz mehrerer Aussagen zu zeigen, kann es effizient sein, diese „in einem Ring“ anzuordnen. Um zum Beispiel die Äquivalenz von drei Aussagen, also

$$A \iff B \iff C$$

zu beweisen, kann man versuchen, die drei Implikationen

$$A \implies B \implies C \implies A$$

zu zeigen, denn die Verknüpfung zweier dieser Implikationen ergibt jeweils die Umkehrung der dritten. Je mehr Aussagen in der Liste vorkommen, um so effizienter wird dieses Verfahren, da man für die Äquivalenz von $n + 1$ Aussagen bestenfalls statt $2n$ Implikationen (für n Äquivalenzen) nur noch $n + 1$ Implikationen beweisen muss.

Beweise von Aussagen mit Mengen

Häufig wird man mit dem Problem konfrontiert von zwei Mengen zu beweisen, dass eine der Menge in der anderen enthalten ist, oder dass beide Mengen gleich sind.

Es seien A und B zwei Mengen. Wenn wir zeigen wollen, dass $A \subseteq B$ gilt, dann müssen wir beweisen, dass für *jedes* $a \in A$ auch $a \in B$ gilt. Betrachten wir dazu ein Beispiel.

Satz. *Es sei $f: X \rightarrow Y$ eine Abbildung und A, B seien Teilmengen von X . Dann gilt $f(A \cup B) \subseteq f(A) \cup f(B)$.*

Beweis. Wir starten also mit einem Element $y \in f(A \cup B)$. Nach Definition gibt es ein $x \in A \cup B$, so dass $f(x) = y$. Da $x \in A \cup B$ folgt $x \in A$ oder $x \in B$. Wir können ohne Beschränkung der Allgemeinheit annehmen, dass $x \in A$ gilt. Also folgt $y = f(x) \in f(A)$ und damit haben wir $y \in f(A) \cup f(B)$ gezeigt. \square

Um zu zeigen, dass zwei Mengen A und B gleich sind, empfiehlt es sich zu zeigen, dass $A \subseteq B$ und $B \subseteq A$ gilt. Beide dieser Teilaussagen können dann mit obiger Methode einzeln bewiesen werden.

Ohne Beschränkung der Allgemeinheit (o.B.d.A.)

Im Beweis unseres letzten Satzes haben wir „ohne Beschränkung der Allgemeinheit angenommen“, dass x in A liegt. Oft wird „ohne Beschränkung der Allgemeinheit“ mit o.B.d.A. abgekürzt. Was bedeutet dieses „ohne Beschränkung der Allgemeinheit“ hier und in anderen Beweisen?

Diese Floskel wird verwendet, wenn eine Voraussetzung eingeführt wird, aufgrund derer die neue Aussage wie eine Spezialisierung *aussieht*, es aber tatsächlich nicht ist. Es wird also der bisherige Grad an Allgemeinheit erhalten. Man muss sich aber bei jedem Auftauchen eines o.B.d.A. überlegen, ob zusätzliche Annahme wirklich so harmlos war, d.h. ob die allgemeine Aussage tatsächlich aus dem bewiesenen Spezialfall folgt.

In unserem Beweis, wussten wir, dass $x \in A$ oder $x \in B$ gilt. Wir haben dann o.B.d.A. $x \in A$ angenommen. Wenn aber $x \in B$ gewesen wäre, so hätte man einfach die Mengen A und B vertauschen können, und hätte dasselbe beweisen können. Die Aussage in unserem Satz bleibt nämlich gleich, wenn die Mengen A und B vertauscht werden.

In diesem konkreten Beispiel wäre es nicht wesentlich länger, für die Leser aber einfacher gewesen, wenn man statt „o.B.d.A. ist $x \in A$ “ zu verwenden die beiden Fälle $x \in A$ und $x \in B$ nacheinander betrachtet hätte. Wenn man aber den folgenden allgemeineren Satz beweisen will, so wird man lieber „o.B.d.A.“ eine Annahme machen als alle Fälle aufzulisten.

Satz. *Es sei $f: X \rightarrow Y$ eine Abbildung und $\{A_i\}_{i \in I}$ eine Familie von Teilmengen von X . Dann gilt $f(\bigcup_{i \in I} A_i) \subseteq \bigcup_{i \in I} f(A_i)$. \square*

10.2 Beweis durch Fallunterscheidungen

Beim Beweis durch Fallunterscheidung, zerlegt man die Aussage $A \implies B$ in mehrere Teilaussagen. Zuerst wird die Aussage A in endlich viele Fälle zerlegt („Wenn A gilt, dann gilt entweder A_1 , oder A_2 , oder... oder A_r “). Danach wird in jedem dieser Fälle die Gültigkeit von B bewiesen (also $A_1 \implies B$, $A_2 \implies B$, ..., $A_r \implies B$). Insgesamt haben wir also bewiesen, dass, wenn A gilt, auch mindestens eine der Aussagen A_1 , A_2 , ..., A_r und somit auch B gelten muss. In Formeln ausgedrückt, verläuft ein Beweis durch Fallunterscheidungen wie folgt:

$$A \implies A_1 \vee A_2 \vee A_3 \vee \dots \vee A_r \implies B$$

Betrachten wir dazu ein Beispiel.

Satz. *Für jede natürliche Zahl n ist 3 ein Teiler von $n(n^2 - 1)$.*

Beweis. Wenn n eine natürliche Zahl ist, dann lässt n bei Division durch 3 entweder 0, oder 1, oder 2 Rest, also entweder $n = 3k$, oder $n = 3k + 1$, oder $n = 3k + 2$ für geeignetes $k \in \mathbb{N}_0$. Wir werden in jedem dieser Fälle zeigen, dass 3 ein Teiler von $n(n^2 - 1)$ ist.

- Falls $n = 3k$ ist, dann ist $n(n^2 - 1) = 3k(9k^2 - 1) = 3r$ für $r = k(9k^2 - 1)$. Somit ist 3 ein Teiler.

- Falls $n = 3k + 1$ ist, dann ist

$$\begin{aligned}
 n(n^2 - 1) &= (3k + 1)((3k + 1)^2 - 1) \\
 &= (3k + 1)(9k^2 + 6k + 1 - 1) \\
 &= (3k + 1)(9k^2 + 6k) \\
 &= 3(3k + 1)(3k^2 + 2k),
 \end{aligned}$$

und wiederum ist $n(n^2 - 1)$ durch 3 teilbar.

- Falls $n = 3k + 2$ ist, dann ist

$$\begin{aligned}
 n(n^2 - 1) &= (3k + 2)((3k + 2)^2 - 1) \\
 &= (3k + 2)(9k^2 + 12k + 4 - 1) \\
 &= (3k + 2)(9k^2 + 12k + 3) \\
 &= 3(3k + 2)(3k^2 + 4k + 1).
 \end{aligned}$$

Auch in diesem Fall ist $n(n^2 - 1)$ durch 3 teilbar.

Somit haben wir für alle auftretenden Fälle gezeigt, dass $n(n^2 - 1)$ durch 3 teilbar ist. Das beweist den Satz. \square

Ein Beweis durch Fallunterscheidung ist vergleichsweise fehleranfällig, da man geneigt ist, einen unangenehmen Fall zu übersehen. Es lohnt sich also, selbst wenn man einen richtigen Beweis durch Fallunterscheidung gefunden hat, noch einmal zu überlegen, ob man nicht einen etwas konzeptionelleren Beweis ohne Fallunterscheidung findet. Im obigen Fall würde folgende Alternative wohl von den meisten Mathematikerinnen und Mathematikern als „besser“ angesehen werden, weil sie kürzer ist und den „Grund“ für die Richtigkeit der Aussage besser erklärt.

Alternativer Beweis. Für $n \in \mathbb{N}$ ist $n(n^2 - 1) = (n - 1) \cdot n \cdot (n + 1)$ das Produkt dreier aufeinanderfolgender ganzer Zahlen. Da aber zwischen je zwei aufeinanderfolgenden Vielfachen von 3 nur zwei andere ganze Zahlen liegen, muss einer der drei Faktoren durch 3 teilbar sein. Also ist auch das Produkt durch 3 teilbar. \square

Wir wollen noch einen zweiten Beweis geben, bei dem eine Fallunterscheidung nützlich ist.

Satz. *Für jedes ebene Dreieck ist der Flächeninhalt gleich der Hälfte des Produktes einer Seitenlänge mit der dazugehörenden Höhe.*

Beweis. Die Behauptung lautet für ein Dreieck, dass wie in Abbildung 10.1 aussieht, dass der Flächeninhalt sich als $\frac{ch}{2}$ berechnen lässt. Typischerweise führt man die Aussage nun auf die einfachere Aussage für rechtwinklige Dreiecke zurück.

Behauptung: *Die Aussage des Satzes gilt für den Fall, dass wir eine der beiden kürzeren Seiten in einem rechtwinkligen Dreieck betrachten.*

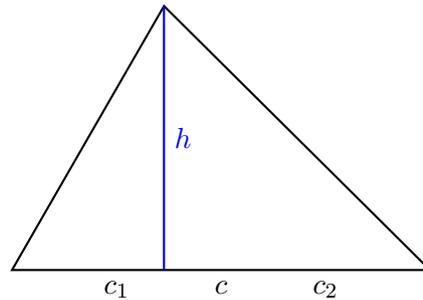
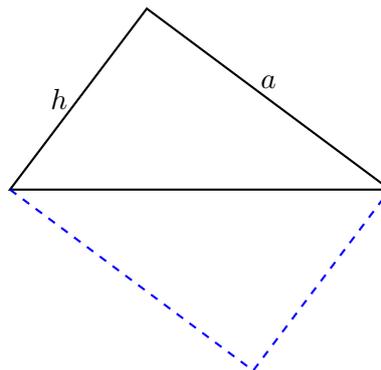


Abbildung 10.1: Ein Dreieck mit eingezeichneter Höhe.



Um diesen Spezialfall zu beweisen, betrachten wir die folgende Skizze eines rechtwinkligen Dreiecks.

Aus der (als bekannt angenommenen) Tatsache, dass der Flächeninhalt eines Rechtecks das Produkt der Seitenlängen ist, sowie der Tatsache, dass die Diagonale ein Rechteck in zwei kongruente rechtwinklige Dreiecke zerlegt, erkennen wir sofort die Wahrheit der Behauptung.

Wir können nun diesen Spezialfall auf die beiden rechtwinkligen Teildreiecke unseres ursprünglichen Dreiecks anwenden, und erhalten als Flächeninhalte $\frac{c_1 \cdot h}{2}$ und $\frac{c_2 \cdot h}{2}$. Da die Summe dieser beiden Flächeninhalte der Gesamtflächeninhalt ist, und andererseits die Summe der Seitenlängen c_1 und c_2 gerade die Seitenlänge c unseres Anfangsdreiecks, folgt hieraus die Behauptung des Satzes für Dreiecke wie das in der Abbildung 10.1 gezeichnete.

Ist der Satz nun bereits bewiesen? Nun, die Frage lautet: „sieht jedes Dreieck so aus wie das in Abbildung 10.1?“, und die Antwort ist „nein“:

In einem stumpfwinkligen Dreieck zerlegt die Höhe unser Dreieck nicht unbedingt in zwei rechtwinklige Dreiecke, so dass die Argumentation wie oben nicht zulässig ist. Allerdings kann man diesen Fall leicht retten, denn hier können wir unser gegebenes Dreieck als *Differenz* zweier rechtwinkliger Dreiecke auffassen, wobei das kleinere die Kathetenlängen h und c_1 und das größere die Kathetenlängen h und $c_1 + c$ besitzt (siehe

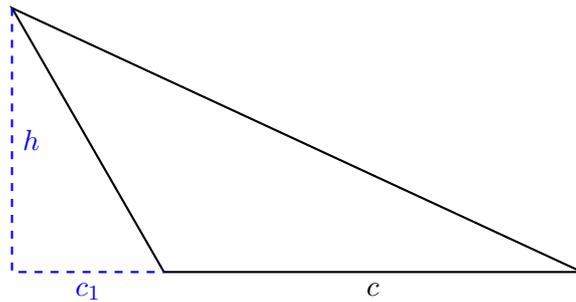


Abbildung 10.2: Ein stumpfwinkliges Dreieck

Abbildung 10.2). Nun erhalten wir für unseren Flächeninhalt

$$A = \frac{(c + c_1)h}{2} - \frac{c_1 h}{2} = \frac{ch}{2},$$

d.h. die Aussage folgt auch in diesem Fall. □

Bemerkung. Oft treten in Beweisen Fallunterscheidungen auf, bei denen es einen „Standardfall“ sowie gewisse „Spezialfälle“ gibt, welche separat behandelt werden. Ein typisches Beispiel hatten wir schon bei Äquivalenzumformungen von Gleichungen gesehen: wenn mit Termen multipliziert wird, in denen noch Variablen vorkommen, muss separat der Fall betrachtet werden, in dem der verwendete Faktor gleich 0 ist.

10.3 Beweis durch Widerspruch

Der Beweis durch Widerspruch ist eine der interessantesten Beweistechniken in der Mathematik. Um die Aussage „aus A folgt B “ zu beweisen gehen wir hierbei von der Negation der Aussage, beziehungsweise von der zu dieser Negation äquivalenten Aussage „ A und nicht B “ aus und zeigen, dass daraus eine falsche Aussage (ein Widerspruch) folgt. Da aber aus einer wahren Aussage keine falsche Aussage folgen kann, muss die Negation von „aus A folgt B “ falsch sein, also ist „aus A folgt B “ selbst wahr. Man geht also bei einem Widerspruchsbeweis wie folgt vor:

$$A \wedge \neg B \implies C_1 \implies C_2 \implies \dots \implies C_r \implies \text{!}$$

Wir haben einen solchen Beweis auch schon angetroffen, als wir gezeigt haben, dass es keine rationale Zahl gibt, deren Quadrat gleich 2 ist. Auch der dritte Beweis in Beispiel 10 war ein Widerspruchsbeweis. Wir geben nun noch ein weiteres Beispiel.

Satz. Die Gleichung $2x^2 + 2x - 1 = 2y^2$ hat keine ganzzahligen Lösungen.

Beweis. Wir nehmen umgekehrt an, dass ganze Zahlen x und y existieren, so dass $2x^2 + 2x - 1 = 2y^2$. In diesem Fall ist aber die rechte Seite der Gleichung eine gerade Zahl, während die linke Seite eine ungerade Zahl ist. Dieser Widerspruch zeigt, dass unsere Annahme falsch war, d.h. ganzzahlige Lösungen der Gleichung existieren nicht. □

Wie wir im letzten Beispiel gesehen haben, konnten wir die Nichtexistenz einer gewissen Lösung mittels eines Widerspruchsbeweises zeigen. Auch die Irrationalität von $\sqrt{2}$ wurde durch einen Widerspruchsbeweis gezeigt. Genau genommen haben wir hier die Nichtexistenz einer rationalen Lösung der Gleichung $x^2 = 2$ gezeigt. Ganz allgemein ist es empfehlenswert einen Widerspruchsbeweis zu versuchen, wenn man zeigen muss, dass etwas nicht existiert. Es ist oft einfacher von etwas auszugehen, das existiert (wie zum Beispiel einer ganzzahligen Lösung von $2x^2 + 2x - 1 = 2y^2$ oder einem Bruch $\frac{a}{b}$, so dass $(\frac{a}{b})^2 = 2$) und dies dann auf einen Widerspruch zu führen, als direkt die Nichtexistenz zu zeigen.

Noch einen Hinweis, wie man einen Widerspruchsbeweis sauber aufschreibt:

- (i) Erklären Sie, dass Sie die Negation der Behauptung annehmen. Erfahrene Leserinnen und Leser erkennen dann, dass ein Beweis durch Widerspruch folgen soll.
- (ii) Formulieren Sie diese Negation explizit.
- (iii) Ziehen Sie Folgerungen daraus, bis Sie zu einem Widerspruch kommen.
- (iv) Erklären Sie, dass ein Widerspruch aufgetreten ist, und worin genau dieser besteht.

10.4 Beweis durch Kontraposition

Die Kontraposition einer Implikation „aus A folgt B “ haben wir schon kennengelernt: es ist die Aussage „wenn B nicht gilt, so gilt auch A nicht“. Wir haben auch gesehen, dass die Äquivalenz dieser beiden Aussagen eine Tautologie ist, d.h. sie gilt unabhängig vom Wahrheitsgehalt der Aussagen A und B immer. Um also „aus A folgt B “ zu beweisen, können wir stattdessen „aus $(\neg B)$ folgt $(\neg A)$ “ beweisen. Unser zweiter Beweis in Beispiel 10 war von dieser Form. Wir wollen nun ein weiteres Beispiel zu dieser Beweismethode betrachten.

Satz. Sind a und b positive reelle Zahlen und ist $a < b$, so gilt $\sqrt[3]{b} - \sqrt[3]{a} < \sqrt[3]{b-a}$.

Wir wollen also zeigen, dass die Aussage „ a und b sind positive reelle Zahlen mit $a < b$ “ die Aussage „ $\sqrt[3]{b} - \sqrt[3]{a} < \sqrt[3]{b-a}$ “ impliziert. Die erste Aussage scheint weniger kompliziert als die zweite Aussage zu sein, (was auch auf die Negationen der Aussagen zutrifft). Wir haben schon gesehen, dass es beim Beweis einer Gleichung günstig ist mit der komplizierteren Seite der Gleichung zu beginnen, um dann auf die einfachere Seite zu kommen. Analog ist es auch bei einem Beweis oft einfacher, von einer komplizierteren Aussage durch logische Schlüsse zu einer einfacheren zu kommen. Aus diesem Grund bietet sich für den Beweis unseres Satzes ein Beweis durch Kontraposition an, da wir dann von der Negation der zweiten Aussage (kompliziert) zur Negation der ersten Aussage (weniger kompliziert) kommen müssen.

Beweis. Es seien a und b positive reelle Zahlen. Wir nehmen an, dass $\sqrt[3]{b} - \sqrt[3]{a} \geq \sqrt[3]{b-a}$,

und können dann folgende Kette äquivalenter Ungleichungen finden:

$$\begin{aligned}
 \sqrt[3]{b} - \sqrt[3]{a} &\geq \sqrt[3]{b-a} && \text{(zu dritten Potenzen übergehen } (x \mapsto x^3 \text{ ist monoton))} \\
 b - 3\sqrt[3]{b^2a} + 3\sqrt[3]{ba^2} - a &\geq b - a && (+ (a - b)) \\
 -3\sqrt[3]{b^2a} + 3\sqrt[3]{ba^2} &\geq 0 && (+ 3\sqrt[3]{b^2a}) \\
 3\sqrt[3]{ba^2} &\geq 3\sqrt[3]{b^2a} && \text{(durch 3 dividieren)} \\
 \sqrt[3]{ba^2} &\geq \sqrt[3]{b^2a} && \text{(zu dritten Potenzen übergehen)} \\
 ba^2 &\geq ab^2 && \text{(durch } ab \text{ dividieren (möglich, da } a, b > 0)) \\
 a &\geq b
 \end{aligned}$$

Damit haben wir bewiesen, dass aus $a < b$ die Ungleichung $\sqrt[3]{b} - \sqrt[3]{a} < \sqrt[3]{b-a}$ folgt. \square

Wenn wir nun mit der Aussage „ $a < b$ “ gestartet hätten und dies direkt zu $\sqrt[3]{b} - \sqrt[3]{a} < \sqrt[3]{b-a}$ hätten umformen wollen, wäre uns dies wohl schwerer gefallen. Wie hätte man zum Beispiel zu $\sqrt[3]{b-a}$ kommen können? Beim gewählten Weg im Beweis ist es wohl ziemlich einleuchtend, dass man durch Übergang zu dritten Potenzen versucht, die Wurzeln aufzulösen, und dann weiter umformt.

Der Beweis durch Widerspruch ist dem Beweis durch Kontraposition sehr ähnlich. Tatsächlich kann jeder Beweis durch Kontraposition auch als Beweis durch Widerspruch geschrieben werden:

Anstatt $(\neg B) \implies (\neg A)$ zu zeigen, kann man $A \wedge \neg B$ annehmen, dann aus $\neg B$ die Aussage $\neg A$ folgern und somit die widersprüchliche Aussage $A \wedge \neg A$ erhalten.

Dadurch enthält der Beweis allerdings unnötige Beweisschritte, was ihn unter Umständen schwerer verständlich macht. Deshalb ist es bei einem Widerspruchsbeweis häufig sinnvoll zu überprüfen, ob man ihn nicht besser als Beweis durch Kontraposition formulieren könnte.

Beweis einer Äquivalenz

Um eine Äquivalenz „ A genau dann, wenn B “ zu beweisen, kann es manchmal hilfreich sein, statt „aus A folgt B “ und „aus B folgt A “ die beiden Aussagen „aus A folgt B “ und „aus $(\neg A)$ folgt $(\neg B)$ “ zu zeigen. Da die zweite Aussage die Kontraposition von „aus B folgt A “ ist, ist damit die Äquivalenz ebenfalls bewiesen. Konkret betrachten wir noch einmal die Aussage aus Beispiel 10, und präzisieren diese zu einer Äquivalenz.

Satz. *Eine natürliche Zahl n ist genau dann ungerade, wenn n^2 ungerade ist.*

Beweis. Wir hatten bereits bewiesen, dass n^2 ungerade ist, falls n ungerade ist. Für die Umkehrung zeigen wir die Kontraposition: *Ist n gerade, so ist auch n^2 gerade.* Dies ist in der Tat nicht schwer: Ist n gerade, so ist 2 ein Primfaktor von n , und somit ist 2 auch ein Primfaktor von $n^2 = n \cdot n$, d.h. n^2 ist ebenfalls gerade. \square

Bemerkung. Vergleichen Sie diesen Beweis noch einmal mit dem zweiten Beweis in Beispiel 10. Dort hatten wir die Kontraposition benutzt, um die ursprüngliche Aussage zu zeigen.

10.5 Beweis durch vollständige Induktion

Vollständige Induktion ist eine Beweistechnik mit deren Hilfe man Aussagen über alle natürliche Zahlen (oder allgemeiner Mengen von Objekten, die durch die natürlichen Zahlen indiziert werden) beweisen kann. Sie basiert auf den fundamentalen Eigenschaften der natürlichen Zahlen, wie sie in *Peanos Axiomen* formalisiert werden:

Definition (Peano-Axiome für die natürlichen Zahlen). Die *natürlichen Zahlen* sind eine Menge \mathbb{N} zusammen mit einer Abbildung $s : \mathbb{N} \rightarrow \mathbb{N}$ und einem ausgezeichnetem Element $1 \in \mathbb{N}$, die folgende Eigenschaften haben:

- (P1) Für alle $x, y \in \mathbb{N}$ folgt aus $s(x) = s(y)$ auch $x = y$.
- (P2) Es existiert kein $x \in \mathbb{N}$ mit $s(x) = 1$.
- (P3) Ist $A \subseteq \mathbb{N}$ eine Teilmenge mit den beiden Eigenschaften
 - (a) $1 \in A$, und
 - (b) für alle $a \in A$ ist auch $s(a) \in A$,dann gilt $A = \mathbb{N}$.

Alle anderen Eigenschaften der natürlichen Zahlen lassen sich aus diesen Axiomen herleiten. Man kann insbesondere Addition und Multiplikation definieren und aus der Definition und den Axiomen deren übliche Eigenschaften nachweisen.

Die Abbildung s formalisiert die Bildung des Nachfolgers (Englisch *successor*). Wir schreiben daher für $s(x)$ auch $x+1$, und benutzen auch weiter die üblichen Bezeichnungen $2 = s(1)$, $3 = s(2)$ etc.

Axiom (P1) sagt aus, dass verschiedene natürliche Zahlen auch verschiedene Nachfolger haben, und Axiom (P2) sagt, dass 1 nicht Nachfolger einer natürlichen Zahl ist. Das letzte Axiom nennt man *Induktionsaxiom*. Es besagt, dass man durch Zählen (d.h. man beginnt mit 1 und betrachtet in jedem Schritt den Nachfolger) *jede* natürlichen Zahl erreicht.

Direkt aus diesem Axiom kann man folgende Beweismethode ableiten.

Satz (Prinzip der vollständigen Induktion). *Für jedes $k \in \mathbb{N}$ sei $A(k)$ eine Aussage. Es gelte*

- (i) $A(1)$ ist wahr und
- (ii) Für alle $k \in \mathbb{N}$ folgt aus der Gültigkeit von $A(i)$ für $1 \leq i \leq k$ die Gültigkeit von $A(k+1)$.

Dann ist $A(k)$ für alle $k \in \mathbb{N}$ wahr.

Beweis. Sei $M \subseteq \mathbb{N}$ die Teilmenge aller natürlichen Zahlen $n \in \mathbb{N}$, so dass die Aussagen $A(1), A(2), \dots, A(n-1)$ und $A(n)$ wahr ist. Nach Voraussetzung (i) enthält M das Element $1 \in \mathbb{N}$. Sei nun k ein Element von M . Dann folgt aus Voraussetzung (ii), dass auch $k+1 = s(k)$ in M ist. Nun folgt aus dem Induktionsaxiom (P3), dass $M = \mathbb{N}$ gelten muss, d.h. insbesondere gilt $A(n)$ für alle $n \in \mathbb{N}$. \square

In einem Beweis einer Aussage vom Typ „ $\forall n \in \mathbb{N} : A(n)$ “ nach diesem Prinzip geht man wie folgt vor:

- Man beweist die Richtigkeit von $A(1)$ (dies nennt man *Induktionsanfang* oder *Induktionsverankerung*)
- Für jedes $k \in \mathbb{N}$ beweist man unter Verwendung der Aussage $A(k)$ oder mehrerer Aussagen $A(j)$ mit $1 \leq j \leq k$ die Aussage $A(k+1)$. Hier nennt man die verwendeten Aussagen die *Induktionsvoraussetzung* und die neue Aussage $A(k+1)$ die *Induktionsbehauptung*. Diesen Teil des Beweises nennt man den *Induktionsschritt*.
- Aus dem Prinzip der vollständigen Induktion folgt nun die Behauptung

$$\forall n \in \mathbb{N} : A(n).$$

Hier sind drei Beispiele von Aussagen, die wir mittels vollständiger Induktion beweisen werden:

- Für alle $k \in \mathbb{N}$ ist der Ausdruck $5^k + 7$ durch 4 teilbar.
- Es gilt $\sum_{i=1}^k i = \frac{k(k+1)}{2}$ für alle $k \in \mathbb{N}$.
- Die Ungleichung $2^{k-1} \leq k!$ gilt für alle $k \in \mathbb{N}$.

Wenn man eine mathematische Aussage sieht, sollte man als Mathematikerin oder Mathematiker den Reflex haben, sofort einmal zu prüfen, ob die Aussage plausibel ist. Hier bietet es sich an, die Aussage für einige natürliche Zahlen zu testen. Wir wollen dies beispielhaft für die erste Behauptung tun:

$$\begin{array}{ll} k = 1 : & 5^1 + 7 = 12, & \text{ist teilbar durch 4.} \\ k = 2 : & 5^2 + 7 = 32, & \text{ist teilbar durch 4.} \\ k = 3 : & 5^3 + 7 = 132, & \text{ist teilbar durch 4.} \\ k = 10 : & 5^{10} + 7 = 9765625 + 7 = 9765600 + 32, & \text{ist teilbar durch 4.} \end{array}$$

Dies gibt also eine gewisse Plausibilität, dass die Aussage wahr ist (aber noch lange keinen Beweis; selbst wenn wir dies für *sehr viele* Beispiele testen würden).

Wie auch eingangs gemacht, werden wir für die Aussage, die wir beweisen wollen $A(k)$ schreiben. In diesem Beispiel ist also $A(k)$ die Aussage „ $5^k + 7$ ist durch 4 teilbar“.

Kommen wir nun zu den Beweisen unserer Beispielaussagen.

Satz. Sei $n = 5^k + 7$ für ein $k \in \mathbb{N}$. Dann ist n durch 4 teilbar.

Beweis. Wir wollen für alle $k \in \mathbb{N}$ die Aussage

$$A(k) = \text{„}5^k + 7 \text{ ist durch 4 teilbar“}$$

mittels vollständiger Induktion beweisen.

1. Der Induktionsanfang ist einfach. Wie wir schon gesehen haben ist für $k = 1$ die Zahl $n = 5^1 + 7 = 12$ durch 4 teilbar, also ist $A(1)$ wahr.
2. Für den Induktionsschritt fixieren wir $k \in \mathbb{N}$ und nehmen an, dass die Aussage $A(k)$ wahr ist, dass also $5^k + 7 = 4r$ für ein geeignetes $r \in \mathbb{N}$ gilt. Damit erhalten wir

$$\begin{aligned} 5^{k+1} + 7 &= 5 \cdot 5^k + 7 \\ &= (4 + 1) \cdot 5^k + 7 \\ &= 4 \cdot 5^k + 5^k + 7 \end{aligned}$$

Wir wissen nach Induktionsvoraussetzung, dass $5^k + 7 = 4r$. Nun können wir weiterumformen

$$\begin{aligned} 4 \cdot 5^k + 5^k + 7 &= 4 \cdot 5^k + 4r \\ &= 4(5^k + r). \end{aligned}$$

Diese Zahl ist offensichtlich durch 4 teilbar.

Somit haben wir gezeigt, dass $A(1)$ wahr ist und, dass auch $A(k) \implies A(k+1)$ wahr ist für alle $k \in \mathbb{N}$. Mit dem Prinzip der vollständigen Induktion haben wir also $A(k)$ für alle $k \in \mathbb{N}$ bewiesen. \square

Bemerkung. Die Idee des Beweises war, den Ausdruck $5^{k+1} + 7$ als Summe zu schreiben, wo einer der Summanden $5^k + 7$ ist. Auf diesen Summanden konnten wir dann die Induktionsvoraussetzung anwenden.

Satz. Für alle $k \in \mathbb{N}$ gilt

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

Beweis. Wir beweisen diese Aussage durch vollständige Induktion nach k . Unsere Aussage $A(k)$ ist also die Gleichung

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

1. Für den Induktionsanfang betrachten wir $k = 1$. Die Aussage $A(1)$ lautet

$$\sum_{i=1}^1 i = 1 = \frac{1 \cdot 2}{2},$$

was eine wahre Aussage ist. Also ist der Induktionsanfang bewiesen.

2. Für den Induktionsschritt fixieren wir $k \in \mathbb{N}$ und nehmen an, dass die Aussage $A(k)$ stimmt, dass also $\sum_{i=1}^k i = \frac{k(k+1)}{2}$ gilt. Damit erhalten wir

$$\begin{aligned} \sum_{i=1}^{k+1} i &= \left(\sum_{i=1}^k i \right) + (k+1) && \text{(Definition der Summe)} \\ &= \frac{k(k+1)}{2} + (k+1) && \text{(Induktionsannahme)} \\ &= \frac{k(k+1) + 2(k+1)}{2} \\ &= \frac{(k+2)(k+1)}{2} \\ &= \frac{(k+1)((k+1)+1)}{2} \end{aligned}$$

Also habe wir gezeigt, dass für alle natürlichen $k \in \mathbb{N}$ aus $A(k)$ auch $A(k+1)$ folgt. Zusammen mit dem Induktionsanfang beweist das nach dem Prinzip der vollständigen Induktion die Aussage für alle $k \in \mathbb{N}$ \square

Bemerkung. Auch in diesem Beweis war die Idee den Ausdruck $\sum_{i=1}^{k+1} i$ in zwei Summanden zu schreiben, wovon der eine gerade $\sum_{i=1}^k i$ ist, und dann auf diesen die Induktionsvoraussetzung anzuwenden. Hier noch eine Warnung zu diesem Vorgehen: Wir setzten *nicht* voraus, dass der Fall $A(k+1)$ ebenfalls gilt (das also $\sum_{i=1}^{k+1} i = \frac{(k+1)(k+2)}{2}$ in unserem Beispiel), sondern wir *beweisen*, dass $A(k+1)$ wahr ist, *falls* $A(k)$ gilt. Dieser Fehler, davon auszugehen, dass $A(k+1)$ wahr ist, wird oft von Anfängern begangen.

Kommen wir noch zum Beweis unserer letzten Beispielaussage.

Satz. Für alle $k \in \mathbb{N}$ gilt die Ungleichung $2^{k-1} \leq k!$.

Beweis. Wir wollen für alle $k \in \mathbb{N}$ die Ungleichung

$$A(k) = „2^{k-1} \leq k!“$$

mittels vollständiger Induktion beweisen.

1. Für den Induktionsanfang nehmen wir $k = 1$ an. Damit erhalten wir

$$2^{1-1} = 2^0 = 1 \leq 1!$$

was eine wahre Aussage ist, also ist $A(1)$ bewiesen.

2. Für den Induktionsschritt fixieren wir $k \in \mathbb{N}$ und nehmen an, dass die Aussage

$A(k)$ stimmt, dass also $2^{k-1} \leq k!$ gilt. Damit erhalten wir

$$\begin{aligned}
 2^{(k+1)-1} &= 2^k \\
 &= 2(2^{k-1}) \\
 &\leq 2(k!) && \text{(Induktionsannahme)} \\
 &\leq (k+1)k! && \text{(da } 2 \leq k+1) \\
 &= (k+1)!
 \end{aligned}$$

Damit haben wir für alle k die Implikation $A(k) \implies A(k+1)$ gezeigt. Zusammen mit der Wahrheit von $A(1)$ beweist dies nach dem Prinzip der vollständigen Induktion die Ungleichung $A(k)$ für alle $k \in \mathbb{N}$ \square

Wie schon eingangs erwähnt, empfiehlt es sich oft, mittels vollständiger Induktion zu argumentieren, wenn man eine Aussage für alle $k \in \mathbb{N}$ beweisen muss, wenn also in der zu beweisenden Aussage etwas der Form „Für jede natürliche Zahl...“ oder „...für alle $k \in \mathbb{N}$ “ steht. Manchmal geschieht dies auch versteckt. Wenn man zum Beispiel beweisen soll:

„Die Zahl $n^2 - 1$ ist durch 8 teilbar, falls n eine ungerade natürliche Zahl ist.“,

so sieht dies auf den ersten Blick nicht nach einer Aussage aus, welche durch die natürlichen Zahlen \mathbb{N} indiziert ist, weil ja die geraden Zahlen ausgeschlossen werden. Jedoch lassen sich die ungeraden und die natürlichen Zahlen einander zuordnen: 1 ist die *erste* ungerade Zahl, 3 ist die *zweite* ungerade Zahl, usw. Man kann die Aussage also umformulieren zu einer Aussage, welche durch die natürlichen Zahlen indiziert wird:

„Die Zahl $(2k - 1)^2 - 1$ ist durch 8 teilbar ist, falls k eine natürliche Zahl ist.“

Allerdings gibt es auch Aussagen über natürliche Zahlen (so wie zum Beispiel die gerade erwähnte Behauptung), für die es direkte Beweise gibt, welche kürzer als ein Induktionsbeweis sind.

Wir haben bisher nur die einfachste Variante des Induktionsbeweises gesehen. Tatsächlich ist diese Beweismethode sehr vielseitig. Man kann damit zum Beispiel auch Aussagen des folgenden Typs zeigen:

Satz. *Für jedes $n \geq 3$ lässt sich jedes ebene n -Eck in $n - 2$ Dreiecke zerlegen.*

Beweis. Wenn man nur an *konvexe* n -Ecke denkt, so hat man schnell einen relativ einfachen Beweis vor Augen.

Man verbindet eine Ecke mit allen anderen Eckpunkten, und erhält so die gewünschte Zerlegung. Allerdings wird die Sache schwieriger, wenn das n -Eck nicht konvex ist. Hier gibt es nicht unbedingt eine Ecke, von der aus alle Diagonalen innerhalb des n -Ecks liegen. Allerdings behaupten wir, dass es stets mindestens *eine* solche Diagonale gibt²

Diese Aussage benutzen wir nun für einen Induktionsbeweis. Für $n = 3$ ist die Behauptung des Satzes trivial. Wir nehmen nun als Induktionsvoraussetzung an, dass die

²Ein exakter Beweis dieser Aussage ist nicht ganz trivial – probieren Sie es einmal!

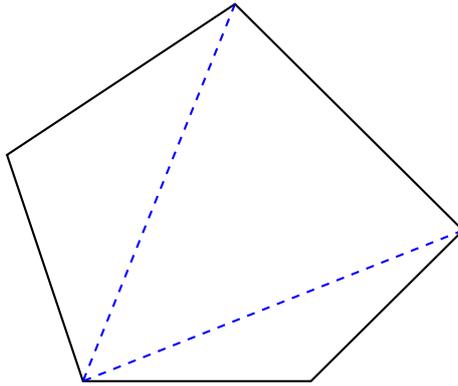


Abbildung 10.3: Ein konvexes Fünfeck mit einer Zerlegung in drei Dreiecke.

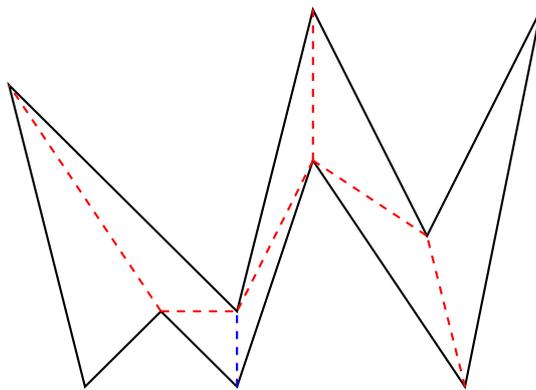


Abbildung 10.4: Ein Zehneck, welches nicht alle seine Diagonalen enthält. Die blaue Diagonale zerlegt es aber beispielsweise in ein Fünfeck und ein Siebeneck.

Aussage für alle $3 \leq j \leq n - 1$ gilt, und betrachten ein beliebiges n -Eck. Darin wählen wir eine Diagonale, die vollständig darin enthalten ist. Diese zerlegt das gegebene n -Eck in zwei kleinere Figuren mit $n_1 \geq 3$ und $n_2 \geq 3$ Ecken, wobei $n_1 + n_2 = n + 2$, denn die Endpunkte der Diagonale werden in beiden Figuren benutzt. Insbesondere gilt $n_1 < n$ und $n_2 < n$. Nach Induktionsvoraussetzung kann man nun diese beiden Figuren auf die gewünschte Art in $n_1 - 2$ bzw. $n_2 - 2$ Dreiecke zerlegen. Zusammengenommen geben diese Dreiecke unsere gewünschte Zerlegung des ursprünglichen n -Ecks in

$$n_1 - 2 + n_2 - 2 = n + 2 - 2 - 2 = n - 2$$

Dreiecke. □

10.6 Die Rolle von Beispielen und Gegenbeispielen

Beispiele und Gegenbeispiele stellen nur in Ausnahmefällen einen Beweis dar.

Beispiele

Die Angabe *eines Beispiels* ist nur dann ein gültiger Beweis, wenn es um eine Existenzaussage geht. In allen anderen Fällen können Beispiele zwar keinen Beweis liefern, sie können aber dabei helfen, einen wirklichen Beweis *zu finden*.

Um eine Idee für einen Beweis einer Aussage zu erhalten (oder auch nur um ein Gefühl für die Aussage zu bekommen), ist es meist hilfreich, sich Beispiele zu überlegen. Mit etwas Glück kann man dann mit Hilfe eines oder mehrerer Beispiele eine Beweisidee erraten.

Deshalb ist es wichtig, dass man für mathematische Begriffe oder Aussagen immer ein Beispiel parat hat. Nachfolgend eine kleine Auswahl von Beispielen zu mathematischen Begriffen und Aussagen aus diesem Vorkurs:

- Beispiele für unendliche Mengen: die natürlichen Zahlen \mathbb{N} , die ganzen Zahlen \mathbb{Z} , die reellen Zahlen \mathbb{R} oder auch \mathbb{P} , die Menge der Primzahlen.
- Beispiel für eine Primzahlzerlegung: $12 = 3 \cdot 2^2$.
- Beispiele für Funktionen: $f: \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch $f(x) = x$ oder $f(x) = x^2$.

Wir wollen nun —als Beispiel— folgende konkrete Aussage beweisen:

„Falls A eine unendliche Menge ist, dann gibt es eine Injektion von den natürlichen Zahlen \mathbb{N} nach A .“

Um eine Idee zu bekommen, betrachten wir also zunächst Beispiele für unendliche Mengen. Nehmen wir also zum Beispiel $A = \mathbb{N}$. Dann können wir leicht durch $f: \mathbb{N} \rightarrow \mathbb{N}$, $f(x) = x$ eine Injektion angeben. Auch wenn wir als Beispiel $A = \mathbb{Z}$ wählen, können wir wieder leicht eine Injektion angeben: $f: \mathbb{N} \rightarrow \mathbb{Z}$, $f(x) = x$. Wie steht es nun aber, wenn $A = \mathbb{P}$? Wir können mit 2 anfangen, dann die nächste Primzahl nehmen, also 3 und so weiter, das hört nie auf, weil es ja unendlich viele Primzahlen gibt. Auf diese Art und Weise kann man eine Injektion $\mathbb{N} \rightarrow \mathbb{P}$ konstruieren.

Funktioniert dieser Beweis vielleicht ganz allgemein? Nun, falls A eine beliebige unendliche Menge ist, dann gibt es sicher ein Element in A – nennen wir es a_1 . Weil A unendlich ist, finden wir auch ein Element in $A \setminus \{a_1\}$, nennen wir es a_2 . Induktiv können wir nun annehmen, dass wir schon paarweise verschiedene Elemente

$$a_1, a_2, \dots, a_n$$

aus A gefunden haben. Weil A unendlich ist, gibt es ein Element in $A \setminus \{a_1, \dots, a_n\}$, nennen wir es a_{n+1} . Induktiv, haben wir nun eine Folge von paarweise verschiedenen Elementen

$$a_1, a_2, \dots$$

gefunden und können nun durch $f: \mathbb{N} \rightarrow A$, $f(n) = a_n$ eine Injektion angeben.

Auch im nächsten Beispiel, wollen wir anhand eines Spezialfalles einen allgemeinen Beweis finden. Wie man im Beispiel sieht, kann das manchmal ganz leicht gehen.

Beispiel. Um zu beweisen, dass für alle Primzahlen p die Gleichung $x^2 = p$ keine rationale Lösung hat, verallgemeinern wir den Beweis im Spezialfall $p = 2$.

Es sei also p eine Primzahl. Wie im Falle $p = 2$ wollen wir einen Widerspruchsbeweis durchführen. Wir nehmen also an, dass ein $x = \frac{a}{b}$ (mit teilerfremden ganze Zahlen a und b) existiert, so dass $x^2 = p$. Es folgt dann

$$a^2 = p \cdot b^2.$$

Also muss a^2 durch p teilbar sein und somit auch a durch p teilbar sein, weil p eine Primzahl ist. Es gibt also eine ganze Zahl r , so dass $a = p \cdot r$ und es folgt $p \cdot b^2 = a^2 = p^2 \cdot r^2$ beziehungsweise

$$b^2 = p \cdot r^2.$$

Das bedeutet aber —analog zu eben— dass b durch p teilbar ist. Dies widerspricht nun aber der Annahme, dass a und b teilerfremd sind.

Bemerkung. Wir betonen noch einmal: wenn man eine Aussage beweisen will, dann reicht es (außer bei Existenzbehauptungen) nicht aus, eines oder mehrere Beispiele anzugeben, in denen die Aussage wahr ist. Die Beispiele können nur dabei helfen, einen Beweis zu finden, ersetzen diesen aber nicht!

Gegenbeispiele

Ein Gegenbeispiel in der Mathematik ist ein Beispiel, das zeigt, dass eine bestimmte Aussage falsch ist. Es heisst *Gegenbeispiel*, weil es „gegen“ eine Aussage gerichtet ist; es zeigt, dass deren *Gegenteil* wahr ist.

Ein einziges Gegenbeispiel kann also dazu benutzt werden, die Falschheit einer Aussage zu zeigen. Dies ist insbesondere bei \forall -Aussagen der Fall. Findet man keinen Beweis für eine solche Aussage, dann kann es also sinnvoll sein, nach einem Gegenbeispiel zu suchen. Falls eines existiert, dann ist die Aussage nämlich falsch und alle Versuche, sie zu beweisen sind zum Scheitern verurteilt.

Beispiele.

- Um die Aussage „Alle Schafe sind weiß“ zu widerlegen, genügt es, ein einziges Schaf zu finden, das nicht weiß ist. So ein Schaf wäre ein Gegenbeispiel zu der Aussage.
- Der Satz „Je zwei Geraden in der euklidischen Ebene schneiden sich in einem Punkt.“ ist falsch, wie man anhand des Gegenbeispiels zweier paralleler Geraden sehen kann.
- Die Aussage „Alle Primzahlen sind ungerade“ ist falsch, weil 2 ein Gegenbeispiel ist.
- Betrachten wir die Aussage „Es seien p und q reelle Zahlen. Falls $p/q \in \mathbb{Q}$, dann gilt auch $p \in \mathbb{Q}$ und $q \in \mathbb{Q}$.“ Diese Aussage könnte auf den ersten Blick vernünftig aussehen. Wenn wir jedoch $p = \sqrt{2}/3$ und $q = \sqrt{2}/2$ setzen, dann erhalten wir $p/q = 2/3 \in \mathbb{Q}$, aber p und q sind reelle Zahlen, die nicht in \mathbb{Q} liegen, weil $\sqrt{2}$

nicht rational ist. Also bilden $p = \sqrt{2}/3$ und $q = \sqrt{2}/2$ ein Gegenbeispiel zu obiger Aussage.

Vorsicht: nur weil wir kein Gegenbeispiel finden, bedeutet das nicht, dass eine Aussage wahr sein muss.

Bemerkung. Auch um eine als wahr bekannte Aussage besser zu verstehen, kann es hilfreich sein zu versuchen, ein Gegenbeispiel zu konstruieren. Man wird dann kein solches Gegenbeispiel finden können, bekommt aber oft ein besseres Gefühl dafür, was die Aussage genau bedeutet, weil man eventuell bei solchen Versuchen versteht, warum das vermeintliche Gegenbeispiel nicht konstruierbar ist. Bestenfalls erhält man so sogar eine Idee für einen Beweis der Aussage.

10.7 Zusammenfassung

Wir haben in diesem Kapitel einige wichtige Beweismethoden vorgestellt. Da man bei der Suche nach einem Beweis nie ganz sicher sein kann, welche Methode erfolgreich sein wird, ist es oft hilfreich, zu Beginn verschiedene Umformulierungen der zu beweisenden Aussage parat zu haben. Für eine Implikation „ $A \implies B$ “ kann man stattdessen zunächst „ $\neg A \vee B$ “ oder die Kontraposition „ $\neg B \implies \neg A$ “ betrachten. Da beide Aussagen äquivalent zur ursprünglichen Aussage sind, können wir, wenn es einfacher erscheint, auch eine dieser beiden Aussagen beweisen.

Bei der Suche nach einem Beweis ist prinzipiell alles erlaubt. Oft lohnt es sich, sowohl die Voraussetzung als auch die Behauptung mehrfach umzuformen, bis man die erhaltenen Aussagen miteinander in Beziehung setzen kann. *Beim Aufschreiben eines Beweises ist es dann aber wichtig, in der richtigen Richtung zu argumentieren.* Hier sollte bei jeder Implikation wirklich von einer Voraussetzung auf eine *daraus folgende* Behauptung geschlossen werden. Auch darf man *nie von der Behauptung ausgehen*, wenn man einen Beweis aufschreibt.

11 Differentialrechnung

11.1 Tangenten und die Definition der Differenzierbarkeit

Für eine „hinreichend glatte“ Kurve kann man in jedem Punkt eine Tangente zeichnen.

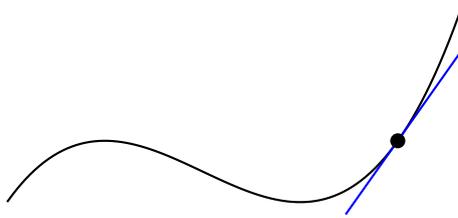


Abbildung 11.1: Tangente an eine Kurve in einem Punkt

Damit ist oft die Vorstellung verbunden, dass diese Tangente zumindest lokal die gegebene Kurve nur in einem Punkt berührt. Wie man schon bei der Betrachtung von Geraden sieht, ist diese Vorstellung etwas zu naiv, denn für diese ist die Tangente in jedem Punkt die Gerade selbst. Dies mag man noch als unwesentlichen Sonderfall abtun, aber was macht man aus folgendem Beispiel?

Beispiel. Wir betrachten den Graphen der Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, definiert als

$$f(x) = \begin{cases} x^2 \sin(\frac{1}{x}) & \text{falls } x \neq 0, \\ 0 & \text{falls } x = 0. \end{cases}$$

Die noch zu formulierende Definition der Tangente an den Graphen von f liefert für $x = 0$ die Gerade $y = 0$, d.h. die x -Achse, welche den Graphen in jedem noch so kleinen Intervall $(-\epsilon, \epsilon)$ unendlich oft schneidet.

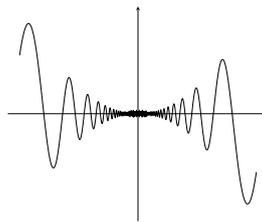


Abbildung 11.2: Die Tangente an den Graphen von f im Nullpunkt ist die x -Achse.

Wie kann man nun also Tangenten formalisieren? Tatsächlich ist die übliche Heuristik für Tangenten recht gut, wenn man sie nur zu Ende denkt und das Ergebnis, so ungewohnt es aus der Schulperspektive in manchen Fällen auch sein mag, akzeptiert. Um

eine Tangente zu approximieren betrachten wir üblicherweise Sekanten. Ist die betrachtete Kurve der Graph einer Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, so erhalten wir die Parameterform der Geradengleichung der Sekante durch die Punkte $(x_0, f(x_0))$ und $(x_1, f(x_1))$ als

$$\begin{cases} x(t) = x_0 + t(x_1 - x_0) \\ y(t) = f(x_0) + t(f(x_1) - f(x_0)) \end{cases}$$

Eliminieren wir den Parameter t , so erhalten wir hieraus die parameterfreie Geradengleichung

$$y = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0} \cdot (x - x_0).$$

Die Steigung ist also gerade der *Differenzenquotient*

$$\Delta f(x_0, x_1) := \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

Nun *definiert* man einfach die Eigenschaft „hinreichend glatt“ für eine Kurve dadurch, dass man die Existenz des Grenzwertes der Differenzenquotienten verlangt.

Definition. Es sei $(a, b) \subset \mathbb{R}$ ein offenes Intervall. Eine Funktion $f : (a, b) \rightarrow \mathbb{R}$ heißt im Punkt $x_0 \in (a, b)$ *differenzierbar*, falls der Grenzwert

$$\lim_{x \rightarrow x_0} \Delta f(x_0, x) = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

existiert. Man bezeichnet diesen Grenzwert mit $f'(x_0)$ und nennt ihn die *Ableitung* von f in x_0 .

Die Gerade durch $(x_0, f(x_0))$ mit der Steigung $f'(x_0)$ nennt man die *Tangente* an den Graphen $\text{graph}(f)$ der Funktion f im Punkt $(x_0, f(x_0))$.

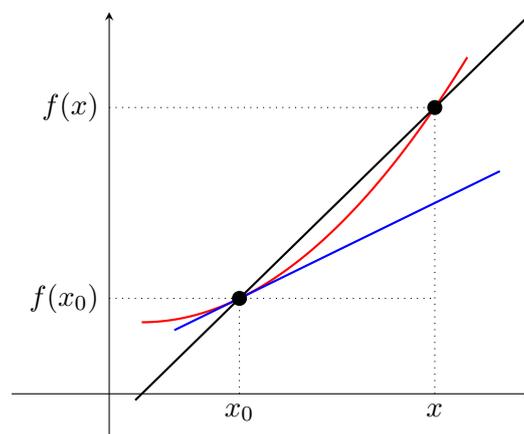


Abbildung 11.3: Beim Grenzübergang $x \rightarrow x_0$ wird die Sekante zur Tangente

Bemerkung. Ist die Funktion $f : (a, b) \rightarrow \mathbb{R}$ im Punkt $x_0 \in (a, b)$ differenzierbar, so ist die Funktion $L_{x_0}(x) = f(x_0) + f'(x_0)(x - x_0)$ gerade diejenige *lineare Funktion*, welche f in der Nähe von x_0 am besten approximiert. Es gilt

$$\lim_{x \rightarrow x_0} \frac{f(x) - L_{x_0}(x)}{x - x_0} = 0, \quad (11.1)$$

und man kann zeigen, dass es höchstens eine lineare Funktion geben kann, für welche (11.1) gilt. In der Tat ist die Existenz einer solchen Funktion gerade äquivalent zur Differenzierbarkeit von f im Punkt x_0 .

Die Ableitung der Funktion f an der Stelle x_0 wird oft auch als

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

geschrieben. Das ist mittels der Substitution $h = x - x_0$ äquivalent zur obigen Definition: $x - x_0$ konvergiert nämlich genau dann gegen 0, wenn x gegen x_0 konvergiert.

Man kann die Definition benutzen, um die Ableitungen einiger elementarer Funktionen zu berechnen.

Beispiel. Ist die Funktion $f : (a, b) \rightarrow \mathbb{R}$ konstant, d.h. $f(x) = c$ für alle $x \in (a, b)$ und ein $c \in \mathbb{R}$, so sind alle Differenzenquotienten

$$\Delta f(x_0, x) = \frac{f(x) - f(x_0)}{x - x_0} = 0,$$

so dass die Ableitung in jedem Punkt $x_0 \in (a, b)$ existiert und gleich 0 ist.

Beispiel. Wir betrachten die Funktion $f(x) = x^2 + x - 1$ an der Stelle $x_0 = 1$. Die folgende Tabelle enthält die Werte einiger Differenzenquotienten:

x	2	1.5	1.25	1.125	1.0625
$\Delta f(x_0, x)$	4	3.5	3.25	3.125	3.0625

Es liegt also die Vermutung nahe, dass die Ableitung an der Stelle $x_0 = 1$ den Wert 3 annimmt. Tatsächlich können wir den Wert der Ableitung auch berechnen:

$$f'(1) = \lim_{x \rightarrow 1} \frac{f(x) - f(1)}{x - 1} = \lim_{x \rightarrow 1} \frac{x^2 + x - 1 - 1}{x - 1} = \lim_{x \rightarrow 1} \frac{(x - 1)(x + 2)}{x - 1} = \lim_{x \rightarrow 1} (x + 2) = 3$$

Im letzten Schritt haben wir benutzt, dass die Funktion $g(x) = x + 2$ stetig ist und somit $\lim_{x \rightarrow 1} g(x) = g(1)$ ist. Somit ist $f(x)$ in $x = 1$ differenzierbar mit $f'(1) = 3$

Bemerkung. Bei der Berechnung der Ableitung von f an der Stelle 1, sind wir davon ausgegangen, dass der Grenzwert $\lim_{x \rightarrow 1} \frac{f(x) - f(1)}{x - 1}$ existiert und haben dann Umformungen durchgeführt, bis wir bei einer konkreten Zahl gelandet sind. Streng genommen ist diese Schlussweise nicht korrekt. Man sieht aber leicht: wenn man beim Resultat, nämlich der 3 gestartet wäre, und dann sukzessive umgeformt hätte, dann wäre man korrekterweise beim Grenzwert $\lim_{x \rightarrow 1} \frac{f(x) - f(1)}{x - 1}$ gelandet und hätte damit dessen Existenz mitbewiesen. Da wir aber in unserer Rechnung auch zeigen wollten, wie man auf das Resultat kommt,

haben wir uns für obige (nicht ganz saubere) Variante entschieden. Wenn wir im Folgenden weitere Ableitungen berechnen werden, werden wir immer wie im obigen Beweis vorgehen und überlassen es den Leserinnen und Lesern, die Existenz der Grenzwerte mit zu überprüfen.

Beispiel. Die Funktion $f(x) = |x|$ ist in $x = 0$ nicht differenzierbar. Falls f in 0 differenzierbar wäre, so müsste für jede gegen 0 konvergente Folge x_n die Folge $\frac{f(x_n)-f(0)}{x_n-0}$ gegen die Ableitung in 0 konvergieren. Um zu zeigen, dass $f(x)$ nicht differenzierbar ist genügt es, ein Gegenbeispiel zu finden, also eine gegen 0 konvergente Folge, sodass die zugehörige Folge von Differenzenquotienten nicht konvergiert.

Wir betrachten die Folge $x_n = \frac{(-1)^n}{n}$, die offensichtlich gegen 0 konvergiert.

- Für gerades n ist $(-1)^n = 1$, somit ist x_n positiv, somit ist $f(x_n) = |x_n| = x_n$ und $\frac{f(x_n)-f(0)}{x_n-0} = \frac{x_n}{x_n} = 1$.
- Falls n ungerade ist, dann ist x_n negativ. Also ist $f(x_n) = |x_n| = -x_n$ und $\frac{f(x_n)-f(0)}{x_n-0} = \frac{-x_n}{x_n} = -1$.

Die Folge $\frac{f(x_n)-f(0)}{x_n-0}$ ist also die Folge $((-1)^n)_{n \in \mathbb{N}}$ gegeben und diese Folge konvergiert nicht.

11.2 Ableitungen elementarer Funktionen

Beispiel. Wir betrachten die Funktion $f(x) = x^2$ an einer beliebigen Stelle x_0 und versuchen, die Ableitung zu bestimmen.

$$\begin{aligned} f'(x_0) &= \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \frac{x^2 - x_0^2}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \frac{(x + x_0)(x - x_0)}{x - x_0} \\ &= \lim_{x \rightarrow x_0} (x + x_0) \\ &= 2x_0 \end{aligned}$$

Im letzten Schritt haben wir benutzt, dass die Funktion $g(x) = x + x_0$ stetig ist und somit $\lim_{x \rightarrow x_0} g(x) = g(x_0) = 2x_0$ gilt.

Obiges Beispiel ist ein Spezialfall des folgenden Satzes:

Satz 22. Wenn $f(x) = x^n$ für eine natürliche Zahl $n \geq 1$, dann ist $f'(x) = nx^{n-1}$.

Man kann diesen Satz problemlos mit einer Verallgemeinerung des gerade präsentierten Beweises für $n = 2$ zeigen. Mit den später diskutierten allgemeinen Rechenregeln erhält man auch alternative Beweise, so dass wir hier auf den Beweis verzichten.

Die folgende Tabelle enthält die Ableitungen einiger weiterer elementarer Funktionen. Um diese Ableitungen direkt aus der Definition zu bestimmen, fehlen uns einige technische Hilfsmittel die zu weit auf die Analysis vorgeifen würden. Daher begnügen wir uns an dieser Stelle damit, die Ableitungen ohne Herleitung anzugeben. Im nächsten Abschnitt werden wir dann sehen, wie wir aus den hier zusammengefassten Ableitungen die Ableitungen anderer Funktionen (z.B. Polynome, allgemeinere Exponentialfunktionen und Logarithmen, Tangens und Kotangens) bestimmen können.

Funktion	Ableitung	Bemerkung
e^x	e^x	Exponentialfunktionen mit anderer Basis folgen im nächsten Abschnitt
$\ln x $	$\frac{1}{x}$	Gilt für $x \neq 0$. Logarithmen mit anderer Basis im nächsten Abschnitt
$\sin x$	$\cos x$	
$\cos x$	$-\sin x$	

Bemerkung. Da wir anstatt $\ln x$ die Funktion $\ln|x|$ betrachten, können wir den Definitionsbereich auf negative Zahlen ausweiten. Nur im Punkt $x = 0$ ist diese Funktion nicht definiert.

11.3 Stetigkeit von differenzierbaren Abbildungen

Bevor wir zu Differentiationsregeln und Anwendungen kommen, wollen wir zunächst noch beweisen, dass die Differenzierbarkeit einer Funktion in einem bestimmten Punkt die Stetigkeit in diesem Punkt impliziert. Da wir momentan nicht viel mehr als die Definitionen haben, werden wir mit diesen argumentieren.

Satz 23. *Wenn eine Funktion $f : (a, b) \rightarrow \mathbb{R}$ in einem Punkt $x_0 \in (a, b)$ differenzierbar ist, dann ist sie in x_0 auch stetig.*

Beweis. Da nach Voraussetzung der Grenzwert $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$ existiert und nach Definition gerade gleich $f'(x_0)$ ist, können wir mit unseren Grenzwertsätzen Folgendes schließen:

$$\begin{aligned}
 0 &= f'(x_0) \cdot 0 \\
 &= \left(\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \right) \cdot \left(\lim_{x \rightarrow x_0} (x - x_0) \right) \\
 &= \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} (x - x_0) \\
 &= \lim_{x \rightarrow x_0} (f(x) - f(x_0)) \\
 &= \left(\lim_{x \rightarrow x_0} f(x) \right) - f(x_0)
 \end{aligned}$$

Dies beweist die Stetigkeit von f in x_0 . □

11.4 Differentiationsregeln

In diesem Abschnitt beschäftigen wir uns mit Rechenregeln für Ableitungen. Da die Ableitung durch einen Grenzwert definiert ist, ist es sinnvoll, sich zuerst nochmals die Rechenregeln für Grenzwerte (Satz 17) ins Gedächtnis zu rufen.

Faktor- und Summenregel

In diesem Abschnitt beschäftigen wir uns mit Ableitungen von Funktionen der Form $f+g$ bzw. $c \cdot f$, wobei f, g differenzierbare Funktionen sind und c eine reelle Konstante ist.

Satz 24 (Faktor- und Summenregel). *Seien $f : (a, b) \rightarrow \mathbb{R}$ und $g : (a, b) \rightarrow \mathbb{R}$ differenzierbare Funktionen und sei $c \in \mathbb{R}$. Dann gilt*

1. $c \cdot f : (a, b) \rightarrow \mathbb{R}$ ist differenzierbar mit Ableitung $c \cdot f'$,
2. $f + g : (a, b) \rightarrow \mathbb{R}$ ist differenzierbar mit Ableitung $f' + g'$.

Beweis. Aus den Rechenregeln für Grenzwerte erhalten wir direkt:

$$\begin{aligned}(c \cdot f)'(x_0) &= \lim_{x \rightarrow x_0} \frac{c \cdot f(x) - c \cdot f(x_0)}{x - x_0} \\ &= \lim_{x \rightarrow x_0} c \cdot \left(\frac{f(x) - f(x_0)}{x - x_0} \right) \\ &= c \cdot \left(\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} \right) \\ &= c \cdot (f'(x_0)),\end{aligned}$$

sowie

$$\begin{aligned}(f + g)'(x_0) &= \lim_{x \rightarrow x_0} \frac{(f(x) + g(x)) - (f(x_0) + g(x_0))}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \left(\frac{f(x) - f(x_0)}{x - x_0} + \frac{g(x) - g(x_0)}{x - x_0} \right) \\ &= \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} + \lim_{x \rightarrow x_0} \frac{g(x) - g(x_0)}{x - x_0} \\ &= (f' + g')(x_0).\end{aligned}$$

□

Bemerkung. Die Summenregel gilt auch für eine beliebige (endliche) Anzahl an Summanden. Das lässt sich leicht mit der Assoziativität der Addition begründen.

Beispiel. Sei $f(x) = x^3 + 2x + 1$. Dann ist

$$\begin{aligned}f' &= (x^3 + 2x + 1)' \\ &= (x^3)' + (2x)' + (1)' && \text{(Summenregel)} \\ &= 3x^2 + 2(x)' + 0 && \text{(Faktorregel für den 2. Summanden)} \\ &= 3x^2 + 2\end{aligned}$$

Dieses Beispiel zeigt schon, wie man ganz allgemein die Ableitungen von Polynomen bestimmen kann. Ein Polynom vom Grad d ist eine Funktion der folgenden Form (wobei die a_i reelle Konstanten sind):

$$p: \mathbb{R} \rightarrow \mathbb{R}, \quad p(x) = \sum_{i=0}^d a_i x^i.$$

Um die Ableitung zu bestimmen, wenden wir zuerst die Summenregel und dann auf jeden Summanden die Faktorregel an und erhalten

$$\begin{aligned} p'(x) &= \left(\sum_{i=0}^d a_i x^i \right)' \\ &= \sum_{i=0}^d (a_i x^i)' && \text{(Summenregel)} \\ &= \sum_{i=0}^d a_i (x^i)' && \text{(Faktorregel für jeden Summanden)} \\ &= \sum_{i=0}^d a_i i x^{i-1} \end{aligned}$$

Beispiel. Die Ableitung des Polynoms

$$f(x) = 2x^4 + 3x^3 - 2x^2 - x + 2$$

ist

$$f'(x) = 2 \cdot 4x^3 + 3 \cdot 3x^2 - 2 \cdot 2x - 1 + 2 \cdot 0 = 8x^3 + 9x^2 - 4x - 1.$$

Neben Polynomfunktionen kann man auch Logarithmen zu einer beliebigen Basis $a \neq 1$ mithilfe der Faktorregel ableiten. Wir erinnern uns daran, dass Logarithmen bezüglich verschiedener Basen mittels der Formel

$$\log_a(x) = \frac{\log_b(x)}{\log_b(a)}$$

ineinander umgerechnet werden können. Da wir wissen, dass $\frac{1}{x}$ die Ableitung von $\ln(x)$ ist, können wir für $f(x) = \log_a(x)$ die Ableitung berechnen als

$$f'(x) = \log'_a(x) = \left(\frac{1}{\ln(a)} \ln(x) \right)' = \frac{1}{\ln(a)} (\ln(x))' = \frac{1}{x \ln(a)}.$$

Produktregel

Satz 25 (Produktregel). *Sind $f: (a, b) \rightarrow \mathbb{R}$ und $g: (a, b) \rightarrow \mathbb{R}$ differenzierbare Funktionen, dann ist auch $f \cdot g: (a, b) \rightarrow \mathbb{R}$ differenzierbar mit Ableitung $f'g + g'f$,*

Beweis. Wir setzen wieder direkt in die Definition der Ableitung ein und erhalten

$$(f \cdot g)'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x)g(x) - f(x_0)g(x_0)}{x - x_0}.$$

Wir addieren $0 = f(x)g(x_0) - f(x)g(x_0)$ im Zähler:

$$(f \cdot g)'(x_0) = \lim_{x \rightarrow x_0} \frac{f(x)g(x) - f(x_0)g(x_0) + f(x)g(x_0) - f(x)g(x_0)}{x - x_0}.$$

Geeignetes Zusammenfassen der Terme und Aufspalten des Grenzwertes in eine Summe von Grenzwerten ergibt

$$\begin{aligned} (f \cdot g)'(x_0) &= \lim_{x \rightarrow x_0} \frac{f(x)(g(x) - g(x_0)) + g(x_0)(f(x) - f(x_0))}{x - x_0} \\ &= \lim_{x \rightarrow x_0} \frac{f(x)(g(x) - g(x_0))}{x - x_0} + \frac{g(x_0)(f(x) - f(x_0))}{x - x_0}. \\ &= \lim_{x \rightarrow x_0} f(x) \frac{(g(x) - g(x_0))}{x - x_0} + \lim_{x \rightarrow x_0} g(x_0) \frac{(f(x) - f(x_0))}{x - x_0} \end{aligned}$$

Wir spalten beide Grenzwerte weiter auf. Beim ersten Grenzwert müssen wir beachten, dass $f(x)$ keine Konstante ist, deshalb müssen wir den Grenzwert des Produktes als Produkt von Grenzwerten schreiben. Die Stetigkeit von f und die Definition der Ableitung gibt schließlich das gewünschte Ergebnis.

$$\begin{aligned} (f \cdot g)'(x_0) &= \lim_{x \rightarrow x_0} f(x) \lim_{x \rightarrow x_0} \frac{(g(x) - g(x_0))}{x - x_0} + g(x_0) \lim_{x \rightarrow x_0} \frac{(f(x) - f(x_0))}{x - x_0} \\ &= f(x_0)g'(x_0) + g(x_0)f'(x_0). \end{aligned}$$

□

Bemerkung. Die Faktorregel ist ein Spezialfall der Produktregel. Da wir aus dem vorigen Abschnitt wissen, dass die Ableitung einer konstanten Funktion $g(x) = c$ die Nullfunktion ist, erhalten wir aus der Produktregel

$$(f \cdot g)' = f' \cdot g + g' \cdot f = c \cdot f' + 0 \cdot f = c \cdot f'.$$

Beispiel. Wir berechnen die Ableitung der Funktion $f(x) = x^2 \cdot x^3$:

$$\begin{aligned} f' &= (x^2)' \cdot (x^3) + (x^3)' \cdot (x^2) \\ &= 2x \cdot x^3 + 3x^2 \cdot x^2 \\ &= 2x^4 + 3x^4 \\ &= 5x^4. \end{aligned}$$

Das stimmt auch damit überein, dass $x^2 \cdot x^3 = x^5$ und nach der Regel für Ableitungen von Potenzfunktionen $f'(x) = 5x^4$ ist.

Beispiel. Gesucht ist die Ableitung von $f(x) = \sin(x) \cdot \ln(x)$. Mit Hilfe der Produktregel berechnen wir diese Ableitung als

$$f'(x) = \cos(x) \ln(x) + \frac{1}{x} \sin(x).$$

Kettenregel

Die Kettenregel beschreibt die Ableitung einer Verknüpfung von differenzierbaren Funktionen.

Satz 26 (Kettenregel). *Seien $f : (a, b) \rightarrow \mathbb{R}$ und $g : (c, d) \rightarrow (a, b)$ differenzierbare Funktionen. Dann ist auch $f \circ g : (c, d) \rightarrow \mathbb{R}$, gegeben als $(f \circ g)(x) = f(g(x))$ differenzierbar. Die Ableitung ist $(f \circ g)' = (f' \circ g) \cdot g'$, an einer Stelle x_0 ist der Wert der Ableitung also gegeben durch $(f \circ g)'(x_0) = f'(g(x_0)) \cdot g'(x_0)$.*

Der Beweis erfolgt wie schon bei der Produktregel durch geschicktes Umformen des Differentialquotienten.

Im Rest dieses Abschnitts betrachten wir einige Anwendungen der Kettenregel.

Ableitungen von Exponentialfunktionen

Eine einfache Anwendung der Kettenregel erlaubt es uns, Ableitungen beliebiger Exponentialfunktionen zu finden. Wir erinnern uns daran, dass $a^x = e^{\ln(a)x}$ für $a > 0$ und beliebige $x \in \mathbb{R}$ gilt. Wir wollen nun die Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$, $h(x) = a^x$ differenzieren (wobei $a > 0$ eine reelle Konstante ist). Wir schreiben zuerst

$$a^x = e^{\ln(a)x} = f(g(x))$$

für $f(x) = e^x$ und $g(x) = \ln(a)x$. Wir bestimmen also

$$f'(g(x)) = e^{g(x)} = e^{\ln(a)x} = a^x$$

und

$$g'(x) = \ln(a).$$

Nach der Kettenregel ist also

$$h'(x) = (f \circ g)'(x) = \ln(a) \cdot a^x.$$

Ableitungen von Potenzfunktionen

Mit Hilfe der Kettenregel können wir die Ableitung einer beliebigen Potenzfunktion berechnen. Insbesondere wollen wir die Ableitung der Wurzelfunktion $x^{\frac{1}{2}}$ berechnen. Wir beschränken daher den Definitionsbereich auf die positiven reellen Zahlen, d.h. auf das Intervall $(0, \infty)$. Wir bestimmen also die Ableitung der Funktion $h_s : (0, \infty) \rightarrow \mathbb{R}$, definiert als $h_s(x) = x^s$, wobei s eine reelle Zahl ungleich 0 ist. Wie im letzten Abschnitt können wir

$$h_s(x) = x^s = e^{\ln(x)s}$$

schreiben. Also gilt $h_s(x) = f(g_s(x))$, wobei $f: \mathbb{R} \rightarrow \mathbb{R}$ die Exponentialfunktion $f(x) = e^x$ ist und $g_s: (0, \infty) \rightarrow \mathbb{R}$ die Funktion $g_s(x) = s \ln(x)$. Aus $f' = f$ erhalten wir

$$f'(g_s(x)) = e^{s \ln(x)} = h_s(x) = x^s,$$

und andererseits gilt

$$g'_s(x) = s \cdot \frac{1}{x}$$

unter Verwendung der Faktorregel und der Ableitung $\frac{1}{x}$ von $\ln(x)$. Mit der Kettenregel erhalten wir also für die Ableitung von h_s in einem Punkt $x \in (0, \infty)$

$$h'_s(x) = f'(g_s(x)) \cdot g'_s(x) = x^s \cdot s \cdot \frac{1}{x} = s x^{s-1}.$$

Wie wir feststellen, sieht das Ergebnis formal genau so aus wie für natürliche Exponenten s .

Die Quotientenregel

Die Quotientenregel ist eine Differentiationsmethode, die häufig unabhängig von der Kettenregel eingeführt wird. Tatsächlich kann man sie aber direkt aus der Kettenregel herleiten.

Satz 27 (Quotientenregel). *Seien f und g differenzierbare Funktionen mit $g(x_0) \neq 0$. Dann gilt*

1. $\frac{1}{g}$ ist differenzierbar in x_0 mit Ableitung $-\frac{g'(x_0)}{g(x_0)^2}$,
2. $\frac{f}{g}$ ist differenzierbar in x_0 mit Ableitung $\frac{f'(x_0)g(x_0) - g'(x_0)f(x_0)}{g(x_0)^2}$.

Beweis. 1. Wir betrachten die Funktion $h(x) = \frac{1}{x}$. dann ist $\frac{1}{g} = (h \circ g)$. Wegen $h'(x) = -\frac{1}{x^2}$ gilt $h'(g(x)) = -\frac{1}{g(x)^2}$ und eine Anwendung der Kettenregel führt zum gewünschten Ergebnis.

2. Wir wenden die Produktregel auf $f \cdot \frac{1}{g}$ an und erhalten

$$\left(\frac{f}{g}\right)' = f' \cdot \frac{1}{g} + \left(\frac{1}{g}\right)' \cdot f = \frac{f'}{g} - \frac{g'f}{g^2} = \frac{f'g - g'f}{g^2}. \quad \square$$

Beispiel. Wir betrachten die Funktion $f(x) = \tan(x) = \frac{\sin(x)}{\cos(x)}$. Ihre Ableitung kann mit Hilfe der Quotientenregel bestimmt werden:

$$f' = \frac{\cos(x) \cdot \cos(x) - (-\sin(x)) \sin(x)}{\cos(x)^2} = \frac{\cos(x)^2 + \sin(x)^2}{\cos(x)^2} = \frac{1}{\cos(x)^2},$$

da $\sin(x)^2 + \cos(x)^2$ bekanntermaßen 1 ist.

Die Ableitung der Funktion $\cot(x) = \frac{1}{\tan(x)}$ können wir entweder auf die gleiche Art und Weise berechnen, oder wir benutzen

$$\cot'(x) = \left(\frac{1}{\tan}\right)'(x) = -\frac{\tan'(x)}{\tan(x)^2} = -\frac{1}{\sin(x)^2}.$$

11.5 Anwendungen

Zahlreiche Eigenschaften einer Funktion können unter Zuhilfenahme der Differentialrechnung untersucht werden. Wir wollen uns in diesem Abschnitt auf drei Eigenschaften konzentrieren.

Monotonieverhalten

Wir erinnern uns: eine Funktion $f: I \rightarrow \mathbb{R}$, welche auf einem reellen Intervall I definiert ist, heißt *monoton wachsend*, wenn $x < y \implies f(x) \leq f(y)$ für alle x, y in I gilt. Für den Differenzenquotienten bedeutet das, dass

$$\frac{f(x) - f(x_0)}{x - x_0} \geq 0$$

und zwar unabhängig von der Wahl von x und x_0 (insbesondere spielt es keine Rolle, ob $x < x_0$ oder $x_0 > x$). Falls die Funktion f nun differenzierbar ist, so können wir folgern, dass für jedes beliebige x in I die Ableitung $f'(x)$ größer oder gleich null sein muss.

Ganz analog wird „monoton fallend“ definiert und man kann argumentieren, dass für eine monoton fallende, differenzierbare Funktion die Ableitung immer ≤ 0 sein muss.

Umgekehrt gilt auch: wenn die Ableitung ≥ 0 ist, dann ist f monoton wachsend, wenn die Ableitung ≤ 0 ist, dann ist f monoton fallend. Intuitiv sollte das klar sein, da die Ableitung die Steigung des Funktionsgraphen widerspiegelt. Dies ist nur ein Plausibilitätsargument. Für einen rigorosen Beweis sei an dieser Stelle auf die Analysis-Vorlesung verwiesen.

Für den folgenden Satz sei noch daran erinnert, dass eine Funktion $f: I \rightarrow \mathbb{R}$ *streng monoton wachsend* heißt, wenn $x < y \implies f(x) < f(y)$ gilt. Analog wird *streng monoton fallend* definiert.

Satz 28. Seien $a, b \in \mathbb{R}$ mit $a < b$ und sei f eine Funktion die in jedem Punkt des Intervalls (a, b) differenzierbar ist. Dann gilt

1. f ist monoton wachsend auf dem Intervall $(a, b) \iff$ für jedes $x \in (a, b)$ gilt $f'(x) \geq 0$,
2. f ist monoton fallend auf $(a, b) \iff$ für jedes $x \in (a, b)$ gilt $f'(x) \leq 0$,
3. falls für jedes $x \in (a, b)$ gilt dass $f'(x) > 0$ (bzw. $f'(x) < 0$), dann ist f streng monoton wachsend (bzw. streng monoton fallend).

Zuerst eine kleine Bemerkung. Ist Ihnen aufgefallen, dass nichts über die Umkehrung der dritten Aussage behauptet wird? Tatsächlich ist sie falsch. Ein Gegenbeispiel ist die Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^3$. Diese Funktion ist zwar streng monoton wachsend auf \mathbb{R} , aber die Ableitung verschwindet in $x_0 = 0$.

Diesen Satz können wir nun verwenden, um differenzierbare Funktionen auf Monotonie zu untersuchen. Um Bereiche zu finden, in denen f monoton wachsend bzw. fallend ist, müssen wir nur noch die Lösungsmengen der Ungleichungen $f'(x) \geq 0$ bzw. $f'(x) \leq 0$

bestimmen. An dieser Stelle sei noch erwähnt, dass eine Funktion, welche stetig auf einem abgeschlossenen oder halboffenen Intervall I ist und im Innern dieses Intervalls monoton wachsend (bzw. fallend) ist, diese Eigenschaft dann auch auf dem gesamten Intervall I besitzt.

Beispiel. Wir zeigen, dass $f(x) = e^x$ streng monoton wächst. Dazu betrachten wir die Ableitung $f'(x) = e^x$. Da die Exponentialfunktion nur Werte größer als 0 annimmt, haben wir $f'(x) > 0$ für alle $x \in \mathbb{R}$ gezeigt. Somit ist f streng monoton wachsend.

Beispiel. Wir wollen die Bereiche bestimmen, auf denen die Funktion $f(x) = 3x^3 - 3x^2 - 3x + 1$ monoton wächst bzw. fällt.

Zu diesem Zweck bestimmen wir zuerst die Ableitung der Funktion:

$$f'(x) = 9x^2 - 6x - 3.$$

Um den Bereich zu bestimmen in dem f monoton wachsend ist, müssen wir nun die Ungleichung $f'(x) \geq 0$ beziehungsweise

$$x^2 - \frac{2}{3}x - \frac{1}{3} \geq 0$$

lösen. Das ist eine quadratische Ungleichung. Wir lösen also zuerst die zugehörige quadratische Gleichung:

$$x_1 = \frac{1}{3} - \sqrt{\frac{1}{9} + \frac{1}{3}} = -\frac{1}{3} \qquad x_2 = \frac{1}{3} + \sqrt{\frac{1}{9} + \frac{1}{3}} = 1$$

Da der Funktionsgraph von $x^2 - \frac{2}{3}x - \frac{1}{3}$ eine nach oben offene Parabel ist, wissen wir somit, dass $f'(x) \geq 0$ für alle x in den Intervallen $(-\infty, -\frac{1}{3}]$ und $[1, \infty)$ gilt. Somit ist f in diesen Intervallen monoton wachsend.

Analog sieht man, dass $f'(x) \leq 0$ für alle $x \in [-\frac{1}{3}, 1]$ gilt. Somit ist f in diesem Intervall monoton fallend.

Extrema

In Anwendungen versucht man häufig, den Wert einer Funktion zu optimieren, das heißt, so groß (bzw. so klein) wie möglich zu machen. Die Stellen an denen f den maximalen bzw. minimalen Wert annimmt bezeichnet man als Extremstellen. Grundsätzlich unterscheidet man zwischen globalen und lokalen Extremstellen.

Definition. Sei $D \subseteq \mathbb{R}$ eine Teilmenge und $f : D \rightarrow \mathbb{R}$ eine Funktion. Ein Punkt $x_0 \in D$ heißt *globale Maximalstelle* (bzw. *globale Minimalstelle*), falls für alle $x \in D$ die Ungleichung $f(x) \leq f(x_0)$ (bzw. $f(x) \geq f(x_0)$) erfüllt ist. Der Funktionswert $f(x_0)$ wird in diesem Fall als globales Maximum (bzw. globales Minimum) bezeichnet.

Beispiele. • Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x$ hat weder ein globales Maximum noch ein globales Minimum, da es zu jeder reellen Zahl eine größere bzw. kleinere gibt.

- Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^2$ hat das globale Minimum 0 an der Stelle $x = 0$. Die Minimalstelle ist in diesem Fall eindeutig. Es gibt kein globales Maximum, da zu jeder reellen Zahl a eine Zahl b gefunden werden kann, sodass $b^2 > a$.
- Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = \cos(x)$ besitzt das globale Maximum 1 und das globale Minimum -1 . Es gibt unendlich viele globale Maximalstellen (jedes geradzahlige Vielfache von π ist globale Maximalstelle) und unendlich viele globale Minimalstellen (jedes ungeradzahlige Vielfache von π ist globale Minimalstelle).
- Eine stetige Funktion, welche auf einem abgeschlossenen Intervall definiert ist besitzt ein globales Maximum und ein globales Minimum, siehe dazu Satz ?? im Kapitel ?? über die Stetigkeit.

Im Gegensatz zu einer globalen Extremstelle vergleichen wir den Funktionswert an einer lokalen Extremstelle nicht mit den Funktionswerten an allen anderen Stellen, sondern nur an solchen, die „nahe bei x_0 “ liegen. Formal kann man das folgendermaßen ausdrücken.

Definition. Sei $f : (a, b) \rightarrow \mathbb{R}$ eine Funktion. Ein Punkt $x_0 \in (a, b)$ heißt *lokale Maximalstelle* (bzw. *lokale Minimalstelle*), falls es ein $\epsilon > 0$ gibt, sodass für alle $x \in (a, b) \cap (x_0 - \epsilon, x_0 + \epsilon)$ die Ungleichung $f(x) \leq f(x_0)$ (bzw. $f(x) \geq f(x_0)$) erfüllt ist. Der Funktionswert $f(x_0)$ wird in diesem Fall als lokales Maximum (bzw. lokales Minimum) bezeichnet.

Offenbar ist jedes globale Extremum auch ein lokales Extremum: Wenn der Funktionswert größer als alle anderen Funktionswerte ist, dann ist er insbesondere auch größer als alle Funktionswerte in einer kleinen Umgebung. Die Umkehrung gilt jedoch nicht. Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$, gegeben durch

$$f(x) = x^3 - 3x^2 - 20x + 30,$$

hat zwar ein lokales Maximum und ein lokales Minimum, aber keine globalen Extrema (siehe Abbildung 11.4).

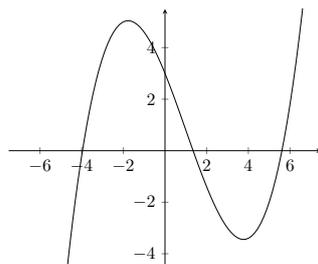


Abbildung 11.4: Eine Funktion ohne globale Extrema

Um lokale Extrema zu finden können wir wieder die Differentialrechnung benutzen. Es ist klar, dass bei einer Funktion f ein lokales Maximum in x_0 vorliegt, wenn f links von x_0 monoton wachsend und rechts von x_0 monoton fallend ist. Präzise ausgedrückt,

bedeutet dies folgendes: Existiert ein $\varepsilon > 0$, so dass f auf dem Intervall $(x_0 - \varepsilon, x_0 + \varepsilon)$ definiert ist, auf $(x_0 - \varepsilon, x_0]$ monoton wachsend und auf $[x_0, x_0 + \varepsilon)$ monoton fallend ist, dann hat f in x_0 ein lokales Maximum. Analoges gilt für ein lokales Minimum. Da wir das Monotonieverhalten einer Funktion durch die Ableitung ausdrücken können, erhalten wir folgenden Satz.

Satz 29. *Seien $x_0 \in \mathbb{R}$ und $\varepsilon > 0$ gegeben und sei f eine auf $(x_0 - \varepsilon, x_0 + \varepsilon)$ differenzierbare Funktion. Falls*

- $f'(x) \geq 0$ für jedes $x \in (x_0 - \varepsilon, x_0]$ und
- $f'(x) \leq 0$ für jedes $x \in [x_0, x_0 + \varepsilon)$,

dann liegt in x_0 ein lokales Maximum vor. Die analoge Bedingung für das Vorliegen eines lokalen Minimums gilt ebenso.

Wenn die Bedingungen des obigen Satzes erfüllt sind, dann ist klar, dass $f'(x_0) = 0$ sein muss. Tatsächlich kann man sogar zeigen, dass wenn f an einer lokalen Extremstelle differenzierbar ist, die Ableitung dort den Wert 0 haben muss. Dies werden Sie in der Analysis-Vorlesung beweisen.

Ist die Funktion zweimal differenzierbar so kann das Verhalten der zweiten Ableitung Aufschluss darüber geben, ob ein Minimum oder ein Maximum vorliegt.

Satz 30. *Sei f eine in x_0 zweimal differenzierbare Funktion.*

1. *Falls $f'(x_0) = 0$ und $f''(x_0) < 0$, dann liegt in x_0 ein lokales Maximum vor.*
2. *Falls $f'(x_0) = 0$ und $f''(x_0) > 0$, dann liegt in x_0 ein lokales Minimum vor.*

Beweis. Wir werden nur die erste Aussage beweisen, da die zweite Aussage ganz analog bewiesen werden kann. Es sei also

$$f''(x_0) = \lim_{x \rightarrow x_0} \frac{f'(x) - f'(x_0)}{x - x_0} < 0.$$

Nach Definition des Grenzwertes gibt es dann ein $\delta > 0$, so dass

$$\frac{f'(x) - f'(x_0)}{x - x_0} < \frac{1}{2} f''(x_0) < 0 \quad \text{für alle } x \text{ mit } 0 < |x - x_0| < \delta.$$

Da $f'(x_0) = 0$, folgt daraus

- $f'(x) < 0$ falls $x_0 < x < x_0 + \delta$ und
- $f'(x) > 0$ falls $x_0 - \delta < x < x_0$.

Also können wir mit Satz 29 schließen, dass ein lokales Maximum in x_0 vorliegt. □

Beispiel. Wir betrachten die Funktion $f(x) = x^3 - 3x^2$. Da diese Funktion überall differenzierbar ist, muss an jedem Extremum $f'(x) = 0$ gelten. Die Ableitung ist $f'(x) = 3x^2 - 6x$. Die Gleichung $f'(x) = 0$ hat somit die Lösungen $x = 0$ und $x = 2$, und das sind unsere Kandidaten für lokale Extremalstellen.

Nun berechnen wir die zweite Ableitung $f''(x) = 6x - 6$. In $x = 0$ gilt $f''(x) = -6 < 0$, somit liegt dort ein lokales Maximum vor. In $x = 2$ gilt $f''(x) = 6 > 0$, also liegt dort ein lokales Minimum vor.

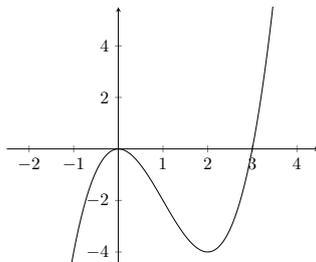


Abbildung 11.5: Graph der Funktion $f(x) = x^3 - 3x^2$

Beispiel. Wir betrachten die Funktion $f(x) = x^4$. Die Ableitung $f'(x) = 4x^3$ nimmt nur an der Stelle $x = 0$ den Wert 0 an. Berechnen wir die zweite Ableitung so erhalten wir $f''(x) = 12x^2$, also ist $f''(0) = 0$ und wir können mit Satz 30 keine Aussage treffen.

Betrachten wir aber die erste Ableitung, so stellen wir fest, dass für $x < 0$ auch $f'(x) = 4x^3 < 0$ gilt. Entsprechend gilt für $x > 0$ auch $f'(x) = 4x^3 > 0$. Somit liegt laut Satz 29 ein lokales Minimum vor.

12 Integralrechnung

12.1 Definition des Integrals

Teile dieses ersten Abschnitts sind von der Behandlung des Themas in [Fri07] inspiriert.

Mittels einfacher Schulmathematik können wir den Flächeninhalt von einfachen Figuren, zum Beispiel denen der folgenden Abbildung, berechnen.

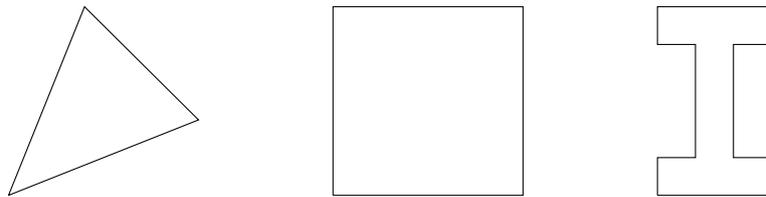


Abbildung 12.1: Verschiedene einfache Figuren

Wie steht es aber mit komplizierteren Figuren, wie zum Beispiel den nachfolgenden „runden Versionen“ der oberen Figuren?

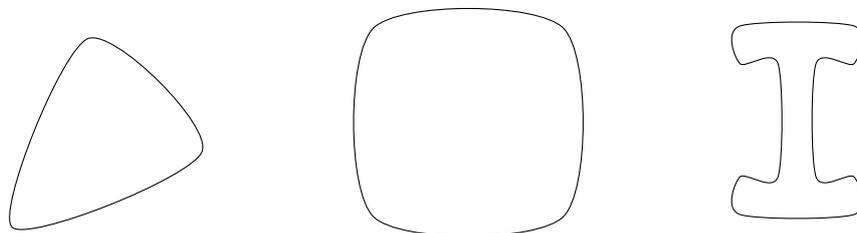


Abbildung 12.2: Verschiedene komplizierte Figuren

Eine Möglichkeit besteht darin, diese Figuren in (hoffentlich einfachere) Teile zu zerlegen. Die Versionen in Abbildung 12.2 können wir zum Beispiel als Varianten der Figuren aus Abbildung 12.1 auffassen, bei denen noch gewisse abgerundete Teile hinzugefügt wurden. Diese kann man (zumindest lokal) als „Fläche unter einem Graphen“ beschreiben.

Genauer formulieren wir also folgendes Modell-Problem. Sei $[a, b]$ ein abgeschlossenes Intervall und $f: [a, b] \rightarrow \mathbb{R}$ eine Funktion. Wir nehmen erst einmal an, dass $f(x) \geq 0$ für alle $x \in [a, b]$. Unser Ziel ist, die Fläche zu berechnen, welche unten durch die x -Achse, oben durch den Graphen von f , von links durch die Gerade $\{x = a\}$ und von rechts durch die Gerade $\{x = b\}$ begrenzt wird.

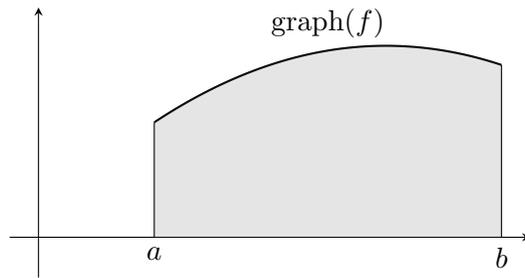


Abbildung 12.3: Unser Modell-Problem: Berechnung der grauen Fläche

Intuitiv ist die Lösung dieses Problems ganz einfach. Man approximiert die gesuchte Fläche durch Flächen, die man tatsächlich berechnen kann. Aber wie soll das gehen? Wenn wir – so wie wir es im Kapitel 7 über Grenzwerte mit dem Kreis getan haben – den Graphen durch Polygonzüge approximieren, so haben wir im Allgemeinen keine Kontrolle darüber, ob die graue Fläche unter dem Polygonzug größer oder kleiner als die gesuchte Fläche ist.

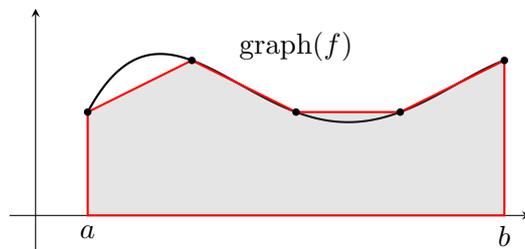


Abbildung 12.4: Approximation durch einen Polygonzug

Wir müssen also etwas geschickter vorgehen. Wir teilen zunächst das Intervall $I = [a, b]$ in n kleinere Intervalle auf:

$$a = x_0 < x_1 < \dots < x_n = b.$$

Dabei brauchen die Teilintervalle $[x_{k-1}, x_k]$ nicht alle gleichlang zu sein. Die Menge der Punkte $Z = \{x_0, x_1, \dots, x_n\}$ heißt *Zerlegung* des Intervalles I . Eine Zerlegung Z' heißt feiner als eine andere Zerlegung Z , falls Z eine echte Teilmenge von Z' ist.

Wir nehmen nun an, dass die Funktion $f : [a, b] \rightarrow \mathbb{R}$, deren Graphen wir betrachten, stetig ist. Auf jedem Teilintervall $[x_{k-1}, x_k]$ nimmt f dann sein Minimum u_k und sein Maximum o_k an, und es gilt natürlich

$$u_k \leq f(x) \leq o_k \quad \text{für alle } x \in [x_{k-1}, x_k].$$

Die Untersumme $U(f, Z)$ und die Obersumme $O(f, Z)$ wird durch

$$U(f, Z) := \sum_{k=1}^n u_k \cdot (x_k - x_{k-1})$$

und

$$O(f, Z) := \sum_{k=1}^n o_k \cdot (x_k - x_{k-1})$$

definiert. Diese Summen beschreiben gerade die Summen der Flächeninhalte der Rechtecke, die in Abbildung 12.5 links bzw. rechts eingezeichnet sind.

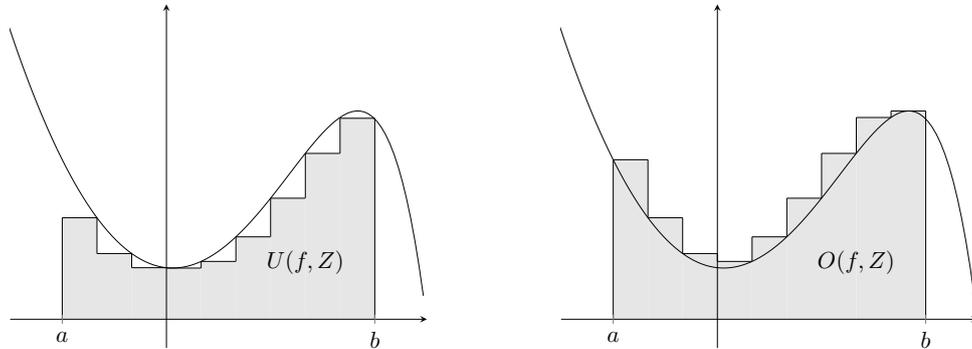


Abbildung 12.5: Unter- und Obersumme einer Funktion bezüglich einer Zerlegung.

Wenn eine sinnvolle Definition des Flächeninhalts A der Fläche unter dem Graphen existiert, so gilt für jede Zerlegung

$$U(f, Z) \leq A \leq O(f, Z),$$

denn die Vereinigung der Rechtecke, deren Fläche in $U(f, Z)$ berechnet wird, liegt vollständig unter dem Graphen, und die Vereinigung der Rechtecke, deren Fläche in $O(f, Z)$ berechnet wird, enthält die gesamte Fläche unter dem Graphen. Außerdem überzeugt man sich schnell, dass die Approximation der Fläche A durch Unter- und Obersummen „besser“ wird, wenn man von einer Zerlegung Z zu einer feineren Zerlegung Z' übergeht, d.h. es gilt dann

$$U(f, Z) \leq U(f, Z') \leq A \leq O(f, Z') \leq O(f, Z).$$

Stetige Funktionen haben nun noch eine andere nützliche Eigenschaft:

Für jedes $\epsilon > 0$ existiert eine Zerlegung $Z = \{a = x_0 < x_1 < \dots < x_n = b\}$ des Intervalls $[a, b]$ (wobei die Anzahl n der Punkte natürlich von ϵ und f abhängt), so dass die Minima und Maxima jedes Teilintervalls $[x_{k-1}, x_k]$ sich um weniger als ϵ unterscheiden, d.h.

$$\forall k \in \mathbb{N}, 1 \leq k \leq n : o_k - u_k < \epsilon.$$

Benutzt man dies und die Definition von Unter- und Obersumme, so sieht man, dass für Z (und jede feinere Zerlegung) die Ungleichung

$$O(f, Z) - U(f, Z) < \epsilon \cdot (b - a)$$

gilt. Hieraus folgt unter Verwendung allgemeiner Eigenschaften der reellen Zahlen, dass der Durchschnitt der Intervalle

$$\bigcap_{Z \text{ ist eine Zerlegung von } [a,b]} [U(f, Z), O(f, Z)] \quad (12.1)$$

über *alle* Zerlegungen aus genau einer reellen Zahl besteht, welche wir nun einfach zur Definition der Fläche unter dem Graphen benutzen.

Definition. Für eine stetige Funktion $f : [a, b] \rightarrow \mathbb{R}$ bezeichnen wir die eben beschriebene eindeutige Zahl

$$A \in \bigcap_Z [U(f, Z), O(f, Z)]$$

als das *Integral von f* im Intervall $[a, b]$ und schreiben dafür

$$A = \int_a^b f(x) dx.$$

Als Konvention setzen wir $\int_b^a f(x) dx$ gleich $-\int_a^b f(x) dx$.

Tatsächlich müssen wir bei obiger Definition nicht annehmen, dass die betrachtete Funktion nur nichtnegative Werte annimmt. Die Flächen, welche unter der x -Achse liegen, leisten einfach einen negativen Beitrag zum Integral.

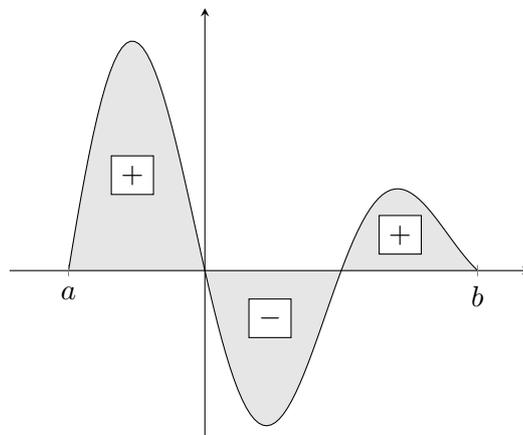


Abbildung 12.6: Flächen oberhalb der x -Achse leisten einen positiven Beitrag, Flächen unterhalb der x -Achse einen Negativen

Um das Integral einer Funktion f zu berechnen genügt es auch, eine Folge von Zerlegungen $(Z_n)_{n \in \mathbb{N}}$ zu finden, sodass die Folgen der Obersummen $(O(f, Z_n))_{n \in \mathbb{N}}$ und der Untersummen $(U(f, Z_n))_{n \in \mathbb{N}}$ den selben Grenzwert I haben. Mithilfe dieser Folgen können wir die Differenz beliebig klein machen, der Grenzwert I ist dann das Integral.

Beispiel. Wir berechnen $\int_0^2 x dx$. Zu diesem Zweck betrachten wir die Zerlegungen von $[0, 2]$ in gleich große Intervalle der Länge $\frac{1}{n}$. Die Obersumme wird dann zu

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

wobei wir auf die Summenberechnung aus dem Kapitel über Beweismethoden zurückgreifen.

Die Untersumme wird entsprechend zu

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

Da die Folge $x_n = \frac{1}{n}$ gegen 0 konvergiert, wissen wir, dass sowohl die Folge der Obersummen als auch die Folge der Untersummen gegen 2 konvergiert. Somit ist

$$\int_0^2 x dx = 2.$$

Das stimmt auch mit der Interpretation des Integrals als Fläche unter dem Funktionsgraphen überein: Für die Funktion $f(x) = x$ ist diese Fläche nämlich ein rechtwinkliges Dreieck in dem beide Katheten Länge 2 haben, Der Flächeninhalt dieses Dreiecks ist $\frac{2 \cdot 2}{2} = 2$.

Bemerkung. Den hier beschriebenen Zugang zur Integration nennt man das *Riemannsches Integral*. Tatsächlich ist es für seine Existenz nicht notwendig, dass die Funktion $f : [a, b] \rightarrow \mathbb{R}$ stetig ist. Sobald man Unter- und Obersummen definieren kann (was für jede beschränkte Funktion f möglich ist), nennt man die Funktion f *Riemann-integrierbar*, wenn der in (12.1) beschriebene Durchschnitt aus genau einer Zahl besteht. Diese Zahl heißt dann *Riemann-Integral* der Funktion f .

12.2 Stammfunktionen

Wie wir im letzten Beispiel gesehen haben, kann man zwar in Einzelfällen Integrale mit Hilfe der Definition zu berechnen, doch diese Methode wird schnell relativ mühsam. Der in diesem Abschnitt formulierte Hauptsatz der Differential- und Integralrechnung stellt einen Zusammenhang zwischen Integration und Differentiation her, welcher deutlich bessere Integrationsmethoden liefert.

Definition. Sei $f : [a, b] \rightarrow \mathbb{R}$ eine stetige Funktion. Eine stetige Funktion $F : [a, b] \rightarrow \mathbb{R}$ heißt *Stammfunktion von f* , wenn F auf dem offenen Intervall (a, b) differenzierbar ist und dort $F'f$ gilt.

Beispiel. $F(x) = \frac{1}{3}x^3$ ist eine Stammfunktion von $f(x) = x^2$, da $F'(x) = \frac{1}{3} \cdot 3x^2 = x^2$ ist. Allgemein ist $F(x) = \frac{1}{k+1}x^{k+1}$ eine Stammfunktion von $f(x) = x^k$ für alle ganzen Zahlen $k \geq 0$.

Beispiel. Wir zeigen, dass $F(x) = x \ln|x| - x$ eine Stammfunktion der Funktion $f(x) = \ln|x|$ ist.¹ Zu diesem Zweck leiten wir F ab und erhalten

$$F'(x) = 1 \cdot \ln|x| + x \cdot \frac{1}{x} - 1 = \ln|x|.$$

Beispiel. $F(x) = \sin x$ und $\tilde{F}(x) = \sin x + \sqrt[4]{3}$ sind beides Stammfunktionen von $f(x) = \cos x$, da die additive Konstante $\sqrt[4]{3}$ beim Ableiten „verschwindet“.

Das obige Beispiel zeigt, dass Stammfunktionen nicht eindeutig definiert sind. Wenn $F(x)$ eine Stammfunktion von $f(x)$ ist, dann ist auch die Funktion $F(x) + c$ für jede beliebige reelle Konstante c eine Stammfunktion von $f(x)$. Da die einzigen Funktionen, deren Ableitung *überall* gleich 0 ist, die konstanten Funktionen sind, gilt auch die Umkehrung: sind F und \tilde{F} zwei Stammfunktionen einer Funktion $f : [a, b] \rightarrow \mathbb{R}$, so gilt $(F - \tilde{F})' = f - f = 0$. Es gibt also eine Konstante $c \in \mathbb{R}$ mit $\tilde{F} - F = c$.

Die Bedeutung von Stammfunktionen erklärt der folgende Satz, demzufolge Integration in gewisser Weise als Umkehrung zur Differentiation betrachtet werden kann. Für seinen Beweis verweisen wir auf die Analysis-Vorlesung.

Satz 31 (Hauptsatz der Differential- und Integralrechnung). *Sei $f : [a, b] \rightarrow \mathbb{R}$ eine stetige Funktion und sei $x_0 \in [a, b]$ ein fester Punkt. Dann gilt:*

(i) *Die Funktion $F : [a, b] \rightarrow \mathbb{R}$, definiert als*

$$F(x) := \int_{x_0}^x f(t) dt,$$

ist eine Stammfunktion von f .

(ii) *Für jede Stammfunktion $F : [a, b] \rightarrow \mathbb{R}$ von f gilt*

$$\int_a^b f(x) dx = F(b) - F(a).$$

Bemerkung. Anstatt $F(b) - F(a)$ schreibt man in diesem Kontext häufig auch $F(x)|_a^b$ oder $[F(x)]_a^b$.

Bemerkung. Wegen ihrer Bedeutung für die Integration schreibt man für die Stammfunktion einer Funktion f oft auch

$$\int f(x) dx.$$

In dieser Schreibweise nennt man Stammfunktionen auch *unbestimmtes Integral* von f , in Abgrenzung zum *bestimmten Integral*

$$\int_a^b f(x) dx,$$

bei dem die Integrationsgrenzen angegeben sind.

¹Diese Behauptung zu überprüfen ist einfach. Wie man darauf kommt, werden wir in Beispiel 12.3 sehen.

Beispiel. Wir berechnen

$$\int_1^e \ln x \, dx.$$

Wir haben vorhin schon überprüft, dass $F(x) = x \ln |x| - x$ eine Stammfunktion von $\ln |x|$ ist. Da alle Punkte im Integrationsintervall größer als 0 sind, gilt dort $|x| = x$. Somit ist

$$\int_1^e \ln x \, dx = F(e) - F(1) = e \ln e - e - (1 \ln 1 - 1) = 1.$$

Beispiel. Wir wollen

$$\int_0^\pi \sin x \, dx$$

berechnen. Wir wissen bereits, dass die Ableitung der Funktion $\cos x$ durch $-\sin x$ gegeben ist. Somit ist $\sin x$ die Ableitung von $-\cos x$ und $-\cos x$ folglich eine Stammfunktion von $\sin x$. Also können wir das Integral berechnen als

$$\int_0^\pi \sin x = -\cos(\pi) - (-\cos(0)) = 1 + 1 = 2.$$

Vorsicht: der obige Satz ist nur dann anwendbar, wenn die Funktion $f : [a, b] \rightarrow \mathbb{R}$ wirklich im ganzen Intervall $[a, b]$ stetig ist. Dafür muss sie insbesondere definiert sein. Das Integral $\int_{-1}^1 \frac{1}{x} \, dx$ können wir so beispielsweise nicht berechnen, weil die Funktion bei $x = 0$ nicht definiert und somit insbesondere nicht stetig ist.

Ein wichtiger Aspekt von Satz 31 besteht darin, die Integration auf die (vergleichsweise einfachere) Differentiation zurückzuführen. Das hilft einerseits, weil wir die Stammfunktionen aller Ableitungen, die wir berechnet haben theoretisch kennen (das haben wir auch eben in den Beispielen benutzt), andererseits ergibt sich dadurch auch die Möglichkeit aus Differentiationsregeln Regeln für die Integration herzuleiten, wie wir im nächsten Abschnitt sehen werden.

12.3 Integrationsregeln

Mittels des Satz 31 haben wir die Integration einer Funktion auf das Auffinden einer Stammfunktion zurückgeführt. Einige einfache Stammfunktionen kennen wir schon, außerdem enthalten viele mathematische Formelsammlungen auch umfangreiche Integraltabellen, also Tabellen, die Stammfunktionen von Funktionen enthalten. Da solche Tabellen aber niemals alle Funktionen enthalten werden, beschäftigen wir uns in diesem Kapitel damit, wie wir Integrale komplizierter Funktionen auf Integrale einfacherer Funktionen zurückführen können.

Summen- und Faktorregel

Die folgenden zwei Regeln ergeben sich direkt aus der Definition des Integrals.

Satz 32. Sind $f : [a, b] \rightarrow \mathbb{R}$ und $g : [a, b] \rightarrow \mathbb{R}$ stetige Funktionen, dann gilt:

1.

$$\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx.$$

2. Für jedes $c \in \mathbb{R}$ gilt

$$\int_a^b c \cdot f(x) dx = c \cdot \int_a^b f(x) dx.$$

Beispiel. Wir berechnen das Integral der Funktion $f(x) = 3 \sin x + 5e^x - \frac{2}{x}$ auf dem abgeschlossenen Intervall $[1, b]$, wobei $b > 1$:

$$\begin{aligned} \int_1^b f(x) dx &= \int_1^b \left(3 \sin x + 5e^x - \frac{2}{x} \right) dx \\ &= \int_1^b 3 \sin x dx + \int_1^b 5e^x dx + \int_1^b \left(-\frac{2}{x} \right) dx \\ &= 3 \int_1^b \sin x dx + 5 \int_1^b e^x dx - 2 \int_1^b \frac{1}{x} dx \\ &= 3(-\cos b + \cos 1) + 5(e^b - e) - 2 \ln b. \end{aligned}$$

Beispiel. Mit Summen- und Faktorregel können wir beliebige Polynome integrieren. Ist ein Polynom durch $p(x) = \sum_{i=0}^k a_i x^i$ gegeben, so können wir das Integral folgendermaßen berechnen:

$$\begin{aligned} \int_0^x p(t) dt &= \int_0^x \left(\sum_{i=0}^k a_i t^i \right) dt \\ &= \sum_{i=0}^k \left(\int_0^x a_i t^i dt \right) && \text{(Summenregel)} \\ &= \sum_{i=0}^k a_i \left(\int_0^x t^i dt \right) && \text{(Faktorregel)} \\ &= \sum_{i=0}^k a_i \left(\frac{1}{i+1} x^{i+1} \right) \\ &= \sum_{i=0}^k \frac{a_i}{i+1} x^{i+1} \end{aligned}$$

Also ist durch die Abbildung $\mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \sum_{i=0}^k \frac{a_i}{i+1} x^{i+1}$ eine Stammfunktion von p gegeben.

Beispiel. Schließlich können wir die Faktorregel benutzen, um auch die Stammfunktionen von Logarithmen bezüglich beliebiger Basen zu bestimmen. Sei also $a > 0$ eine reelle Zahl. Dann gilt für alle $x > 0$:

$$\int_1^x \log_a(t) dt = \int_1^x \frac{1}{\ln a} \ln(t) dt = \frac{1}{\ln a} \int_1^x \ln(t) dt = \frac{1}{\ln a} (x \ln(x) - x + 1).$$

Da wir zu einer Stammfunktion noch eine beliebige Konstante addieren dürfen, ist also die Funktion $F_a : (0, \infty) \rightarrow \mathbb{R}$, gegeben als $F_a(x) = \frac{1}{\ln a}(x \ln(x) - x)$, eine Stammfunktion von $f_a : (0, \infty) \rightarrow \mathbb{R}$, $f_a(x) = \log_a(x)$.

Partielle Integration

Die Faktorregel kann nur angewendet werden, wenn wir es mit einem konstanten Faktor zu tun haben. Wollen wir hingegen eine Funktion der Form $f(x) \cdot g(x)$ integrieren, benötigen wir die sogenannte partielle Integration. Für eine Funktion f mit Stammfunktion F und eine differenzierbare Funktion g gilt laut Produktregel

$$(F \cdot g)'(x) = f(x) \cdot g(x) + F(x) \cdot g'(x),$$

beziehungsweise

$$f(x) \cdot g(x) = (F \cdot g)'(x) - F(x) \cdot g'(x).$$

Diese Überlegungen führen mit Satz 31 zum folgenden Resultat:

Satz 33 (Partielle Integration). *Sei $F : [a, b] \rightarrow \mathbb{R}$ die Stammfunktion einer stetigen Funktion $f : [a, b] \rightarrow \mathbb{R}$ und sei $g : [a, b] \rightarrow \mathbb{R}$ eine stetige Funktion, die auf (a, b) differenzierbar mit stetiger Ableitung g' ist. Dann gilt*

$$\int_a^b f(x) \cdot g(x) dx = F(x) \cdot g(x) \Big|_a^b - \int_a^b F(x) \cdot g'(x) dx.$$

Für unbestimmte Integrale vereinfacht sich dies zu

$$\int f(x) \cdot g(x) dx = F(x) \cdot g(x) - \int F(x) \cdot g'(x) dx.$$

Bemerkung. Partielle Integration ist vor allem dann sinnvoll, wenn man die Stammfunktion F von f bereits kennt oder sehr leicht berechnen kann. Außerdem sollte das Integral auf der rechten Seite nach Möglichkeit leichter (oder zumindest nicht schwerer) zu lösen sein, als das auf der linken Seite.

Beispiel. Wir bestimmen die Stammfunktion von $x \cdot \ln(x)$ für $x \in (0, \infty)$ durch partielle Integration. Dafür wählen wir $f(x) = x$ und $g(x) = \ln(x)$ und berechnen $F(x) = \frac{1}{2}x^2$ und $g'(x) = \frac{1}{x}$. Setzen wir nun in die Formel von Satz 33 ein, so erhalten wir als Stammfunktion

$$\begin{aligned} \int x \cdot \ln(x) dx &= \frac{1}{2}x^2 \ln(x) - \int \frac{1}{2}x^2 \cdot \frac{1}{x} dx = \frac{1}{2}x^2 \ln(x) - \frac{1}{2} \int x dx = \frac{1}{2}x^2 \ln(x) - \frac{1}{4}x^2. \end{aligned}$$

Iterative Anwendung

In manchen Fällen ist es sinnvoll, auf das Integral auf der rechten Seite nochmals partielle Integration anzuwenden. Das ist zum Beispiel der Fall, wenn $g(x)$ ein Polynom vom Grad mindestens 2 ist und $f(x)$ eine Exponentialfunktion oder Sinus bzw. Kosinus ist. In diesen Fällen reduziert sich der Grad des Polynoms g durch Ableiten, während die Funktion F gleich schwer zu integrieren ist wie die Funktion f . Nun kann man so lange partiell integrieren, bis g Grad 0 hat. Das verbleibende Integral ist dann (manchmal) leicht zu lösen.

Beispiel. Wir suchen eine Stammfunktion der Funktion $(x^2 - 1) \cdot e^x$. Für die partielle Integration wählen wir $f(x) = e^x$ und $g(x) = x^2 - 1$. Dann ist $F(x) = e^x$ eine Stammfunktion von $f(x)$ und $g'(x) = 2x$. Wir erhalten also

$$\int (x^2 - 1) \cdot e^x dx = e^x \cdot (x^2 - 1) - \int e^x \cdot 2x dx = e^x \cdot (x^2 - 1) - 2 \int e^x \cdot x dx.$$

Das Integral auf der rechten Seite berechnen wir wieder mittels partieller Integration. Wir wählen $f(x) = e^x$ und $g(x) = x$, bestimmen $F(x) = e^x$ und $g'(x) = 1$ und setzen ein:

$$\int e^x \cdot x dx = e^x \cdot x - \int e^x \cdot 1 dx = (e^x \cdot x - e^x = e^x(x - 1)).$$

Somit erhalten wir

$$\int (x^2 - 1) \cdot e^x dx = e^x \cdot (x^2 - 1) - 2e^x(x - 1) = e^x \cdot (x^2 - 2x + 1).$$

Rekursive Gleichungen

Manchmal kann es passieren, dass nach der partiellen Integration das Integral auf der rechten Seite das gleiche wie auf der linken Seite ist. Diese Möglichkeit besteht insbesondere, wenn wir eine Funktion integrieren, die aus Winkelfunktionen und Exponentialfunktionen zusammengesetzt ist, da diese sich beim Integrieren „nicht stark ändern“. Wenn dieser Fall eintritt, können wir die Stammfunktion bestimmen, indem wir eine lineare Gleichung lösen.

Beispiel. Wir wollen eine Stammfunktion von $\cos x \cdot \sin x$ bestimmen. Wir wählen $f(x) = \cos x$ und $g(x) = \sin x$, und erhalten als Stammfunktion von f die Funktion $F(x) = \sin x$ und $g'(x) = \cos x$ und setzen ein:

$$\int \cos x \cdot \sin x dx = \sin x \cdot \sin x - \int \sin x \cdot \cos x dx$$

Wir stellen fest, dass die Integrale auf der linken und rechten Seite identisch sind. Also können wir diese Integrale auf eine Seite bringen und erhalten

$$2 \int \cos x \cdot \sin x dx = (\sin x)^2,$$

d.h.

$$\int \cos x \cdot \sin x dx = \frac{1}{2}(\sin x)^2.$$

Partielle Integration mit $f \equiv 1$

Ein besonderer Kniff bei der partiellen Integration besteht manchmal darin, eine Funktion zu sehen, wo keine ist. Anstatt der Funktion $g(x)$ betrachten wir die Funktion $1 \cdot g(x)$, d.h. wir setzen $f(x) = 1$, und wenden partielle Integration an.

Beispiel. Wir wollen mit Hilfe dieses Tricks erklären, wie man die Stammfunktion von $\ln(x)$ finden kann, ohne dass sie wie in Beispiel 12.2 vom Himmel fällt. Es sei also $g(x) = \ln x$ und $f(x) = 1$, so dass $F(x) = x$ eine Stammfunktion von f ist und $g'(x) = \frac{1}{x}$. Mittels partieller Integration erhalten wir nun

$$\int \ln(x) dx = x \cdot \ln(x) - \int x \cdot \frac{1}{x} dx = x \cdot \ln(x) - \int 1 dx = x \cdot \ln(x) - x.$$

Dies erklärt die Behauptung aus Beispiel 12.2.

Beispiel. Wir berechnen die Stammfunktion von $g(x) = (\ln x)^2$ für $x > 0$. Dazu wählen wir $f(x) = 1$, also $F(x) = x$ und berechnen $g'(x) = 2 \ln(x) \frac{1}{x}$. Nun setzen wir ein und erhalten

$$\int (\ln x)^2 dx = x \cdot (\ln x)^2 - \int x \cdot 2 \ln(x) \frac{1}{x} dx = x \cdot (\ln x)^2 - 2 \int \ln(x) dx.$$

Die Stammfunktion des Logarithmus haben wir gerade noch einmal bestimmt. Insgesamt erhalten wir

$$G : (0, \infty) \rightarrow \mathbb{R}, \quad G(x) = x ((\ln x)^2 - 2 \ln(x) + 1)$$

als Stammfunktion von $g : (0, \infty) \rightarrow \mathbb{R}$, $g(x) = (\ln x)^2$.

Substitution

Die Substitutionsregel dient der Berechnung von Integralen von Funktionen, welche sich als Verknüpfung $f \circ \varphi$ von zwei Funktionen f und φ schreiben lassen. Sie kann als „Umkehrung“ der Kettenregel angesehen werden.

Falls $F(x)$ eine Stammfunktion von $f(x)$ ist, so besagt die Kettenregel (der Differentialrechnung), dass $F(\varphi(x))$ eine Stammfunktion von $f(\varphi(x)) \cdot \varphi'(x)$ ist. Mit dem Hauptsatz der Differential- und Integralrechnung ergibt sich das folgende Resultat.

Satz 34 (Substitution). *Sei $f : [a, b] \rightarrow \mathbb{R}$ eine stetige Funktion und sei $\varphi : [c, d] \rightarrow [a, b]$ eine stetige Funktion, die auf (c, d) differenzierbar ist und eine stetige Ableitung hat. Dann gilt*

$$\int_{\varphi(c)}^{\varphi(d)} f(t) dt = \int_c^d f(\varphi(x)) \cdot \varphi'(x) dx. \quad (12.2)$$

Die Funktion φ im obigen Satz wird in Anwendungen oft bijektiv gewählt, so dass es eine Umkehrfunktion φ^{-1} gibt. Damit erhält man

$$\int_a^b f(t) dt = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(b)} f(\varphi(x)) \cdot \varphi'(x) dx.$$

Außerdem können wir nun eine Stammfunktion von f bestimmen:

$$F(t) = \int_a^t f(s) ds = \int_{\varphi^{-1}(a)}^{\varphi^{-1}(t)} f(\varphi(x)) \cdot \varphi'(x) dx.$$

Wir erhalten also eine Stammfunktion $F(t)$ von $f(t)$, indem wir $x = \varphi^{-1}(t)$ in eine Stammfunktion $H(x)$ von $f(\varphi(x)) \cdot \varphi'(x)$ einsetzen.

Bemerkung. In den Beispielen in diesem Kurs wird es meistens relativ offensichtlich sein, welche Funktion φ sich für die Substitution anbietet. Bei einer Funktion der Bauart $\sin(h(x))$, $\ln(h(x))$ etc. ist der erste Versuch meistens $t = h(x)$ zu setzen und diese Gleichung nach x aufzulösen. Außerdem gibt es noch einige „Standardsubstitutionen“, die bei Funktionen einer bestimmten Bauart häufig funktionieren. Im Allgemeinen ist es aber eine gewisse Kunst, eine geeignete Substitution zu finden. Momentan sind Menschen mit etwas Übung dabei in komplizierteren Fällen immer noch deutlich besser als Computerprogramme.

In der Praxis ist es oft einfacher, ein gegebenes Integral mit der rechten Seite der Gleichung (12.2) zu identifizieren, wenn man Substitutionen verwendet. Dies werden wir auch in den meisten Beispielen tun.

Lineare Substitution

Der einfachste Fall einer Substitution tritt auf, wenn die zu integrierende Funktion die Form $f(mx + n)$ hat. In diesem Fall betrachtet man die Substitution $\varphi(x) = (mx + n)$, so dass $\varphi'(x) = m$. Die rechte Seite von (12.2) hat dann die Form

$$\int_c^d f(mx + n) \cdot m dx,$$

so dass wir für unser gesuchtes Integral

$$\int_c^d f(mx + n) dx = \frac{1}{m} \int_{\varphi(c)}^{\varphi(d)} f(t) dt \quad (12.3)$$

erhalten.

Beispiel. Wir bestimmen das Integral der Funktion e^{-3x} im Intervall $[0, 2]$. Dazu setzen wir $\varphi(x) = -3x$ und erhalten $\varphi'(x) = -3$, sowie $f(t) = e^t$. Daraus folgt mit (12.3)

$$\int_0^2 e^{-3x} dx = \int_0^2 f(\varphi(x)) dx = -\frac{1}{3} \int_0^{-6} f(t) dt = \frac{1}{3} \cdot \int_{-6}^0 e^t dt = \frac{e^t}{3} \Big|_{-6}^0 = \frac{1}{3} - \frac{1}{3e^6}.$$

Wollen wir eine Stammfunktion bestimmen, so führt die gleiche Substitution zum Ziel. Da $\varphi(x) = -3x$ eine bijektive Funktion auf ganz \mathbb{R} ist, erhalten wir

$$\int_0^x e^{-3y} dy = -\frac{1}{3} \int_0^{-3x} e^t dt = -\frac{e^t}{3} \Big|_0^{-3x} = -\frac{e^{-3x}}{3} + \frac{1}{3}.$$

Also ist $-\frac{1}{3}e^{-3x}$ eine Stammfunktion von e^{-3x} .

Beispiel. Wir wollen die Funktion $\sin(\pi(x - 3))$ im Intervall $[-1, 2]$ integrieren. Eine geeignete Substitution sollte die Funktion vereinfachen, beispielsweise wäre $t = \varphi(x) = \pi(x - 3)$ praktisch. Führen wir diese Substitution durch, erhalten wir $\varphi'(x) = \pi$ und somit wieder unter Verwendung von (12.3)

$$\int_{-1}^2 \sin(\pi(x - 3)) dx = \frac{1}{\pi} \int_{-4\pi}^{-\pi} \sin t dt = \frac{1}{\pi} (-\cos t) \Big|_{-4\pi}^{-\pi} = \frac{1}{\pi} (1 - (-1)) = \frac{2}{\pi}$$

Umkehrung der Kettenregel

Bei der linearen Substitution haben wir ausgenutzt, dass die Ableitung $\varphi'(t)$ eine Konstante war und somit die Berechnung des Integrals nicht erschwert hat. In gewissen anderen Fällen, wenn ein Integrand (fast) in der Form $f(\varphi(x)) \cdot \varphi'(x)$ gegeben ist, kann man Satz 34 ebenfalls direkt anwenden. Man beachte, dass die Funktion φ nicht bijektiv sein muss, damit der Satz anwendbar ist. Leider ist es allerdings nicht immer einfach zu erkennen, dass wir es mit einer Funktion in dieser Form zu tun haben.

Beispiel. Wir wollen das Integral der Funktion $x^3 \cdot e^{(x^4)}$ im Intervall $[-1, 2]$ bestimmen. Da uns der Exponent am meisten stört, ist es verlockend, die Substitution $\varphi(x) = x^4$ auszuprobieren. Es gilt dann $\varphi'(x) = 4x^3$, was bis auf die Konstante 4 dem anderen Faktor des Integranden entspricht. Dies stört nicht sehr, denn wir erhalten aus (12.2)

$$\int_{-1}^2 x^3 \cdot e^{(x^4)} dx = \frac{1}{4} \int_{-1}^2 4x^3 \cdot e^{(x^4)} dx = \frac{1}{4} \int_{(-1)^4}^{2^4} e^t dt = \frac{1}{4} e^t \Big|_1^{16} = \frac{e^{16} - e}{4}.$$

Beispiel. Wir wollen eine Stammfunktion des Tangens, $\tan : (-\frac{\pi}{2}, \frac{\pi}{2}) \rightarrow \mathbb{R}$, bestimmen. Zu diesem Zweck schreiben wir die Funktion als

$$\tan x = \frac{\sin x}{\cos x} = -\frac{1}{\cos x} (-\sin x) = f(\varphi(x)) \cdot \varphi'(x)$$

für $f(t) = -\frac{1}{t}$ und $\varphi(x) = \cos x$. Wir erhalten aus (12.2)

$$\int_0^x \tan(s) ds = \int_1^{\cos x} -\frac{1}{t} dt = -\ln(t) \Big|_1^{\cos x} = -\ln(\cos x).$$

Also ist die auf $(-\frac{\pi}{2}, \frac{\pi}{2})$ definierte Funktion $-\ln(\cos x)$ eine Stammfunktion des Tangens.

Beispiel. Wir geben noch ein weiteres Beispiel für diese Art der Substitution, und betrachten das Integral

$$\int_a^b x^5 (1 - x^3)^{313} dx.$$

Der Integrand ist ein Polynom, welches nach Ausmultiplizieren elementar integrierbar ist. Allerdings ist dieses Vorgehen für praktische Zwecke eher unbrauchbar. Stattdessen versuchen wir es mit der Substitution $\varphi(x) = 1 - x^3$, so dass $\varphi'(x) = -3x^2$. Der Integrand hat also die Form $(\varphi(x))^{313} \cdot \varphi'(x) \cdot (-\frac{1}{3}x^3)$, und den letzten Faktor können wir nun glücklicherweise ebenfalls als Funktion von $\varphi(x)$ ausdrücken, nämlich als $-\frac{1}{3}x^3 = -\frac{1}{3}(1 -$

$\varphi(x)$). Wenn wir also die Funktion f als $f(t) = -\frac{1}{3}(1-t)t^{313}$ wählen, so hat unser Integral die gewünschte Form

$$\int f(\varphi(x)) \cdot \varphi'(x) dx,$$

und wir erhalten mit (12.2)

$$\begin{aligned} & \int_a^b x^5(1-x^3)^{313} dx \\ &= -\frac{1}{3} \int_{1-a^3}^{1-b^3} (1-t)t^{313} dt = \frac{1}{3} \int_{1-b^3}^{1-a^3} (1-t)t^{313} dt = \frac{1}{3} \left(\frac{1}{314}t^{314} - \frac{1}{315}t^{315} \right) \Big|_{1-b^3}^{1-a^3}. \end{aligned}$$

Dieses Ergebnis ist in der Praxis deutlich handhabbarer als die eingangs erwähnte Methode des Ausmultiplizierens.

12.4 Flächenberechnung mittels Integration

Wie bereits am Anfang dieses Kapitels erwähnt, kann das bestimmte Integral als disignierte Fläche zwischen dem Funktionsgraphen und der x -Achse interpretiert werden, wobei Teile der Fläche, die unterhalb der x -Achse liegen, negativ zum Ergebnis beitragen. Diese Interpretation kann benutzt werden, um die Fläche zwischen zwei Funktionen $f(x)$ und $g(x)$ auf einem Intervall $[a, b]$ zu bestimmen (siehe Abbildung 12.7).

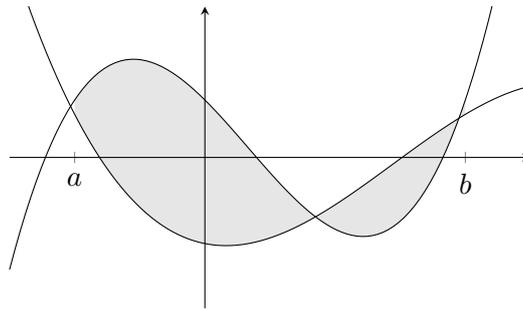


Abbildung 12.7: Fläche Zwischen zwei Funktionen

Gilt für jedes $x \in [a, b]$, dass $f(x) \geq g(x)$, dann können wir die Fläche zwischen den beiden Funktionen als Differenz der Integrale berechnen:

$$A = \int_a^b f(x) dx - \int_a^b g(x) dx = \int_a^b (f(x) - g(x)) dx$$

Für zwei beliebige Funktionen f und g werden wir mit dieser Rechnung nur den signierten Flächeninhalt erhalten, bei dem die Teile mit $f(x) < g(x)$ negativ zum Ergebnis beitragen. Wollen wir die tatsächliche Fläche zwischen zwei beliebigen Funktionsgraphen stetiger Funktionen $f : [a, b] \rightarrow \mathbb{R}$ und $g : [a, b] \rightarrow \mathbb{R}$ bestimmen, so können wir folgendermaßen vorgehen:

1. Wir bestimmen die Nullstellen der Funktion $f - g$ gilt. Das sind genau die x -Koordinaten der Punkte, in denen sich die Funktionsgraphen von f und g schneiden. Wir nehmen an, dass es nur endlich viele solche Stellen gibt, und bezeichnen sie in aufsteigender Reihenfolge mit $x_1 < x_2 < \dots < x_n$. Weiter führen wir die Bezeichnungen $x_0 = a$ und $x_{n+1} = b$ ein.
2. In jedem der Intervalle $[x_k, x_{k+1}]$ gilt entweder $f(x) \geq g(x)$ oder $g(x) \geq f(x)$. Somit können wir die Fläche zwischen den Funktionen in jedem dieser Intervalle entweder durch $\int_{x_k}^{x_{k+1}} (f(x) - g(x)) dx$ oder durch $\int_{x_k}^{x_{k+1}} (g(x) - f(x)) dx$ bestimmen. Da sich die beiden Integrale nur um ein Vorzeichen unterscheiden und wir schon wissen, dass die Fläche immer positiv sein muss, können wir aber auch einfach $\int_{x_k}^{x_{k+1}} (f(x) - g(x)) dx$ berechnen und dann den Betrag des Ergebnisses nehmen.
3. Wir addieren nun die Flächen über den einzelnen Intervallen $[x_k, x_{k+1}]$, um die Gesamtfläche zu erhalten.

Beispiel. Wir berechnen die Fläche, zwischen den Graphen der Funktionen $f(x) = \sin x$ und $g(x) = (\sin x)^2$ im Intervall $[0, 2\pi]$.

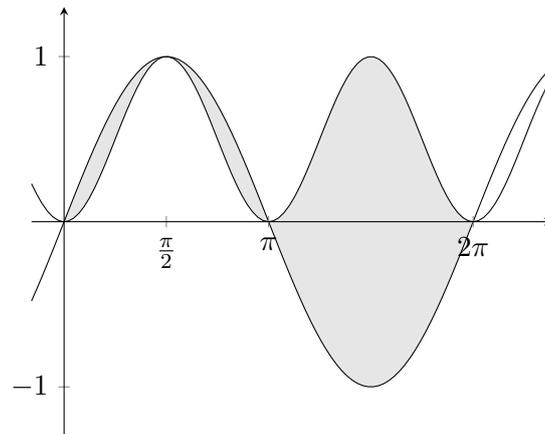


Abbildung 12.8: Fläche zwischen $f(x) = \sin x$ und $g(x) = (\sin x)^2$

1. Zuerst bestimmen wir die Werte von x für die $\sin x = (\sin x)^2$ gilt. Um diese Gleichung zu lösen substituieren wir $r = \sin x$ und erhalten die quadratische Gleichung $r = r^2$ beziehungsweise $r^2 - r = 0$. Die Lösungen dieser Gleichung sind gegeben durch 0 und 1.

Für eine Lösung der ursprünglichen Gleichung $\sin x = (\sin x)^2$ muss also $\sin x = 0$ oder $\sin x = 1$ gelten. Ersteres geschieht im Intervall $[0, 2\pi]$ an den Stellen 0, π und 2π , letzteres ist nur bei $x = \frac{\pi}{2}$ der Fall.

Die zu betrachtenden Teilintervalle für die Integration sind somit $[0, \frac{\pi}{2}]$, $[\frac{\pi}{2}, \pi]$ und $[\pi, 2\pi]$.

2. Um die Funktion $h(x) = f(x) - g(x)$ in den einzelnen Intervallen zu integrieren bestimmen wir zunächst eine Stammfunktion:

$$\int (\sin x - (\sin x)^2) dx = \int (\sin x) dx - \int (\sin x)^2 dx$$

Das erste Integral lässt sich als $-\cos x + c$ schreiben. Für das zweite Integral benutzen wir partielle Integration. Wir wählen $f(x) = \sin x$, $g(x) = \sin x$ und bestimmen $F(x) = -\cos x$ und $g'(x) = \cos x$. Somit erhalten wir

$$\begin{aligned} \int (\sin x)^2 dx &= -\cos x \cdot \sin x - \int (-\cos x) \cos x dx \\ &= -\cos x \cdot \sin x + \int (\cos x)^2 dx. \end{aligned}$$

Wir formen das Integral auf der rechten Seite dieser Gleichung um

$$\int (\cos x)^2 dx = \int (1 - (\sin x)^2) dx = \int 1 dx - \int (\sin x)^2 dx = x - \int (\sin x)^2 dx.$$

Das verbleibende Integral ist gerade das Ausgangsintegral unserer Nebenrechnung. Einsetzen ergibt nun

$$\int (\sin x)^2 dx = -\cos x \cdot \sin x + x - \int (\sin x)^2 dx,$$

und durch einfache Umstellung wird hieraus

$$\int (\sin x)^2 dx = \frac{1}{2} (-\cos x \cdot \sin x + x).$$

Insgesamt erhalten wir also die Stammfunktion

$$H(x) = \int \sin x - (\sin x)^2 dx = -\cos x - \frac{1}{2} (x - \cos x \cdot \sin x)$$

für $h = f - g$. Mit Hilfe dieser Stammfunktion können wir nun den Flächeninhalt auf den einzelnen Teilintervallen bestimmen:

$$\begin{aligned} \left[0, \frac{\pi}{2}\right]: & \quad A_1 := \left| H\left(\frac{\pi}{2}\right) - H(0) \right| = \left| -\frac{\pi}{4} - (-1) \right| = 1 - \frac{\pi}{4} \\ \left[\frac{\pi}{2}, \pi\right]: & \quad A_2 := \left| F(\pi) - F\left(\frac{\pi}{2}\right) \right| = \left| 1 - \frac{\pi}{2} - \left(-\frac{\pi}{4}\right) \right| = 1 - \frac{\pi}{4} \\ [\pi, 2\pi]: & \quad A_3 := |F(2\pi) - F(\pi)| = \left| -1 - \pi - \left(1 - \frac{\pi}{2}\right) \right| = 2 + \frac{\pi}{2} \end{aligned}$$

3. Die Gesamtfläche ergibt sich als Summe der Einzelflächen:

$$A = A_1 + A_2 + A_3 = 1 - \frac{\pi}{4} + 1 - \frac{\pi}{4} + 2 + \frac{\pi}{2} = 4.$$

Literaturverzeichnis

- [Fri07] Klaus Fritzsche, *Mathematik für Einsteiger*, 4. Edition, Spektrum Akademischer Verlag, Heidelberg, 2007.
- [Hou12] Kevin Houston, *Wie man mathematisch denkt*, Springer Spektrum, Heidelberg, 2012.