
Kapitel 1 Grundlagen

§ 1 Mengen und Zahlen

Inhalt:

Mengen und ihre Elemente, Beschreibung von Mengen

z.B. $\{1, 2, 3\}$, $\{a_1, a_2, a_3, \dots\}$ oder $\{x \in M : E(x)\}$,

die Menge \mathbb{N} der natürlichen Zahlen, Vereinigung, Schnittmenge, Differenz ($A \setminus B$), ein bißchen Aussagen-Logik, Teilmengen, leere Menge, disjunkte Mengen, logische Implikation und Äquivalenz.

Die Menge \mathbb{R} der reellen Zahlen (als Menge unendlicher Dezimalbrüche), die Teilmengen \mathbb{N} , \mathbb{N}_0 , \mathbb{Z} und \mathbb{Q} .

Rechenregeln (Kommutativ- und Assoziativgesetze, Distributivgesetz, Null, Eins, Negatives, Inverses), Begriff des Körpers (Beispiele: \mathbb{R} , \mathbb{Q} , \mathbb{F}_2).

Die Menge \mathbb{R}_+ der positiven reellen Zahlen, die Ordnungsrelationen $<$, \leq , $>$, \geq , offene, abgeschlossene und halboffene Intervalle, $\pm\infty$, erweiterte Zahlengerade, unbeschränkte Intervalle.

Objekte der Mathematik sind

Zahlen, Punkte, Geraden, Ebenen, Vektoren, Felder, Funktionen, ...

Neue Objekte werden gebildet, indem man schon bekannte Objekte zu *Mengen* zusammenfaßt. Zur Abkürzung benutzt man meist ein Symbol oder einen Buchstaben für die Menge; die zusammengefaßten Objekte werden als Variable geschrieben. Man nennt sie auch die *Elemente* der Menge.

Ist x ein Element der Menge M , so schreibt man:

$$x \in M .$$

Liegt x **nicht** in der Menge M , so schreibt man:

$$x \notin M .$$

Zwei Mengen A und B sind genau dann *gleich*, wenn sie die gleichen Elemente enthalten. Man schreibt dann:

$$A = B.$$

Kleine Mengen beschreibt man am besten, indem man ihre Elemente aufzählt, z.B. ist die Menge $M = \{1, 2, 3\}$ die Menge der Zahlen 1, 2 und 3.

Wenn klar ist, was gemeint ist, kann man auch Elemente weglassen und durch Pünktchen ersetzen, z.B.

$$\{1, 2, \dots, 7\} \quad \text{an Stelle von} \quad \{1, 2, 3, 4, 5, 6, 7\}.$$

Mengen können auch unendlich viele Elemente enthalten. Manchmal ist das Bildungsgesetz für eine solche Menge so klar, daß man wieder mit Pünktchen arbeiten kann, etwa im Falle der *Menge der natürlichen Zahlen*

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

Bleiben wir kurz bei der Menge \mathbb{N} . Folgende Eigenschaften setzen wir ohne nähere Begründung als bekannt voraus:

- Es gibt unendlich viele natürliche Zahlen. Was das genau bedeutet, werden wir uns später überlegen.
- Natürliche Zahlen können addiert und multipliziert werden. Die Rechenregeln sind die, die jeder von der Schule her kennt.
- Eine Zahl $n \in \mathbb{N}$ wird *kleiner* als eine Zahl $m \in \mathbb{N}$ genannt (in Zeichen: $n < m$), falls es eine natürliche Zahl k mit $m = n+k$ gibt. Jede aus natürlichen Zahlen gebildete Menge besitzt ein kleinstes Element, insbesondere ist die 1 die kleinste unter allen natürlichen Zahlen.
- Jede natürliche Zahl n besitzt einen unmittelbaren *Nachfolger*, nämlich die Zahl $n + 1$.

Meistens beschreibt man eine Menge mit Hilfe einer Eigenschaft. Steht der Ausdruck $E(x)$ für eine Eigenschaft, die gewissen Elementen x zukommt, so besteht die Menge

$$M = \{x : E(x)\}$$

aus genau denjenigen Objekten x , die diese Eigenschaft $E(x)$ besitzen.

Beispiele.

1. Eine Zahl $n \in \mathbb{N}$ heißt durch eine Zahl $q \in \mathbb{N}$ *teilbar*, falls es eine Zahl $m \in \mathbb{N}$ mit $n = q \cdot m$ gibt. So ist z.B. 28 durch 4 teilbar, weil $28 = 4 \cdot 7$ ist. Die Menge $\{x : x \in \mathbb{N} \text{ und } x \text{ durch } 2 \text{ teilbar}\}$ ist dann die Menge der *geraden natürlichen Zahlen*.

Da die Eigenschaft „ x ist durch 2 teilbar“ aus der Menge aller natürlichen Zahlen die Menge der geraden Zahlen aussondert, schreiben wir letztere auch in der Form

$$\{x \in \mathbb{N} : x \text{ ist durch } 2 \text{ teilbar}\}.$$

In beiden Fällen darf man übrigens den Doppelpunkt auch durch einen senkrechten Strich ersetzen, das ergibt z.B. $\{x \in \mathbb{N} | x \text{ ist durch } 2 \text{ teilbar}\}$.

2. Sind A und B zwei schon gegebene Mengen, so kann man auch die Mengen

$$A \cup B, \quad A \cap B \text{ und } A \setminus B$$

bilden. Und zwar nennt man

$$\begin{aligned} A \cup B &= \{x : x \in A \text{ oder } x \in B\} && \text{die Vereinigungsmenge,} \\ A \cap B &= \{x : x \in A \text{ und } x \in B\} && \text{die Schnittmenge,} \\ \text{und } A \setminus B &= \{x : x \in A \text{ und } x \notin B\} && \text{die Differenz von } A \text{ und } B. \end{aligned}$$

Hier sind *logische Verknüpfungen* von Eigenschaften ins Spiel gekommen. Das sollte ich etwas genauer erläutern.

Daß ein Objekt x eine gewisse Eigenschaft besitzt, ist eine *Aussage*.¹ Eine solche Aussage ist entweder *wahr* oder *falsch*. Andere Werte kann sie in der Mathematik nicht annehmen. In der Elektrotechnik ist es genauso. Entweder ist ein Stromkreis geschlossen (und es fließt Strom) oder der Stromkreis ist unterbrochen (und es fließt kein Strom).

Seien nun $E(x)$ und $F(x)$ zwei Aussagen, die man über das Objekt x machen kann. Dann versteht man unter

$$E(x) \vee F(x) \quad (\text{in Worten: „}E(x) \text{ oder } F(x)\text{“})$$

diejenige Aussage, die genau dann wahr ist, wenn wenigstens eine der beiden Aussagen $E(x)$, $F(x)$ wahr ist. Ist etwa $E(x)$ die Aussage „ x ist ein männlicher Einwohner von Wuppertal“ und $F(x)$ die Aussage „ x wohnt in Wuppertal und ist über 21“, so steht $E(x) \vee F(x)$ für die Aussage „ x ist entweder ein männlicher Wuppertaler oder eine Wuppertalerin über 21“.

Die Aussagenverknüpfung

$$E(x) \wedge F(x) \quad (\text{in Worten: „}E(x) \text{ und } F(x)\text{“})$$

ist genau dann wahr, wenn $E(x)$ und $F(x)$ beide wahr sind. In unserem Beispiel ergibt sich die Aussage „ x ist ein männlicher Einwohner Wuppertals über 21“.

Schließlich ist die Aussage

$$\neg E(x) \quad (\text{in Worten „nicht } E(x)\text{“})$$

¹Die Logiker sprechen auch von einer *Aussageform* oder einem *Prädikat*.

die logische Verneinung der Aussage $E(x)$. Das ist die Aussage, die genau dann wahr ist, wenn $E(x)$ falsch ist. In unserem Beispiel ist $\neg E(x)$ die Aussage „ x ist weibliche Einwohnerin Wuppertals“ (zumindest, wenn es keine Zwitter gibt), und $\neg F(x)$ die Aussage „ x wohnt in Wuppertal und ist nicht über 21“. Ist A eine beliebige Menge, so ist die logische Verneinung der Aussage „ $x \in A$ “ die Aussage „ $x \notin A$ “.

3. Ist

$$A = \{x \in \mathbb{N} : x < 4\} = \{1, 2, 3\}$$

und $B = \{x \in \mathbb{N} : 5 < x\} = \{6, 7, 8, \dots\}$,

so enthält $A \cap B$ kein einziges Element. Auch dafür gibt es eine Bezeichnung. Die *leere Menge* ist die Menge, die kein Element enthält. Sie wird mit dem Symbol \emptyset bezeichnet.

Ist $A \cap B = \emptyset$, so nennt man A und B *disjunkt*.

Definition:

Die Menge A heißt *Teilmenge* der Menge B (in Zeichen: $A \subset B$), falls jedes Element von A auch Element von B ist.

Man nennt dann B auch *Obermenge* von A .

In der Definition steckt eine weitere logische Konstruktion. A ist Teilmenge von B , falls gilt:

Wenn x Element von A ist, **dann** ist x Element von B .

Man nennt so etwas eine logische Folgerung oder *Implikation* und schreibt:

$$x \in A \implies x \in B$$

Sind $E(x)$ und $F(x)$ Aussagen, so ist auch $E(x) \implies F(x)$ eine Aussage. Aber wann ist sie wahr? Ist $E(x)$ wahr und folgt $F(x)$ aus $E(x)$, so muß auch $F(x)$ wahr sein. Um zu sehen, was passiert, wenn $E(x)$ falsch ist, betrachten wir ein Beispiel:

Ist x eine natürliche Zahl, so schreibt man bekanntlich auch x^2 für das Produkt $x \cdot x$. Dann ist offensichtlich

$$\{x \in \mathbb{N} : x^2 = 9\} \subset \{1, 2, 3\},$$

aber auch $\{x \in \mathbb{N} : x^2 = 5\} \subset \{12, 15, 37\}$,

denn da im zweiten Fall kein Element in der linken Menge liegt, muß auch für kein x geprüft werden, ob x in der rechten Menge liegt. Mit anderen Worten: Die leere

Menge ist in jeder anderen Menge enthalten. Und das bedeutet: Aus einer falschen Aussage kann man alles folgern! Die Aussage „Wenn $2 \cdot 2 = 5$ ist, dann bin ich der Papst“ ist also logisch korrekt. Zur Übersicht hier eine „Wahrheitstafel“ für die Implikation:

$E(x)$	$F(x)$	$E(x) \implies F(x)$
w	w	w
w	f	f
f	w	w
f	f	w

Man überzeugt sich leicht davon, daß sich die gleiche Wahrheitswert-Verteilung bei der logischen Verknüpfung $F(x) \vee \neg E(x)$ ergibt.

Gilt sowohl $E(x) \implies F(x)$ als auch $F(x) \implies E(x)$, so schreibt man kurz $E(x) \iff F(x)$ und sagt: Die Aussagen $E(x)$ und $F(x)$ sind äquivalent. Das bedeutet, daß $E(x)$ genau dann wahr ist, wenn $F(x)$ es ist. Eine Anwendung ist der Beweis der Gleichheit von Mengen:

Um zu zeigen, daß die Mengen $A = \{x : E(x)\}$ und $B = \{x : F(x)\}$ gleich sind, reicht es zu zeigen, daß $E(x) \iff F(x)$ für alle x gilt.

Beispiel.

Sei $A = \{n \in \mathbb{N} : n^2 = 9\}$ und $B = \{n \in \mathbb{N} : (n + 1)^2 = 16\}$.

Sei n eine beliebige natürliche Zahl. Ist $n^2 = 9$, so muß $n = 3$ sein. Dann ist aber $n + 1 = 4$ und $(n + 1)^2 = 16$. Damit haben wir gezeigt:

$$n^2 = 9 \implies (n + 1)^2 = 16.$$

Für die Äquivalenz beider Aussagen brauchen wir noch die umgekehrte Implikation. Sei also $(n + 1)^2 = 16$. Dann kann nur $n + 1 = 4$ gelten. Innerhalb der Menge \mathbb{N} muß man zwar mit dem Subtrahieren vorsichtig sein, aber aus der Gleichung $n + 1 = 4$ folgt doch durch Subtraktion von 1 (auf beiden Seiten) die Gleichung $n = 3$. Und dann ist $n^2 = 9$. Das bedeutet:

$$(n + 1)^2 = 16 \implies n^2 = 9.$$

Übrigens wird sehr häufig eine der Beweisrichtungen vergessen!

Bekanntlich läßt sich jede *positive reelle Zahl* als ein unendlicher Dezimalbruch darstellen, z.B. durch $3.1415\dots$ oder $2.000\dots$,² und umgekehrt liefert jeder unendliche Dezimalbruch

$$a_{-n}a_{-(n-1)}\dots a_{-1}a_0.a_1a_2a_3a_4\dots,$$

²Ich verwende in der Vorlesung den Dezimalpunkt an Stelle des sonst in Deutschland üblichen Dezimalkommas.

wo jedes a_i eine der Ziffern $0, 1, 2, \dots, 9$ bedeutet, eine positive reelle Zahl.

Die Darstellung ist nicht ganz eindeutig. Ist z.B. $a_k \neq 9$, so ist

$$0.a_1 \dots a_{k-1} a_k 999 \dots = 0.a_1 \dots a_{k-1} (a_k + 1) 000 \dots$$

Stehen hinter dem Dezimalpunkt nur Nullen, so kann man diese weglassen und man erhält eine natürliche Zahl. Bezeichnen wir die Menge der positiven reellen Zahlen mit \mathbb{R}_+ , so ist $\mathbb{N} \subset \mathbb{R}_+$.

Die Menge \mathbb{R} aller *reellen Zahlen* besteht aus \mathbb{R}_+ , der Null und den negativen reellen Zahlen, d.h. den Zahlen $-x$ mit $x \in \mathbb{R}_+$. Man beachte: Das Minuszeichen vor einer Variablen bedeutet nicht, daß diese Variable für eine negative Zahl steht! Nur wenn x positiv ist (in Zeichen: $x > 0$), dann ist $-x$ negativ. Ist dagegen x negativ, so ist $-x$ positiv.

In \mathbb{R} enthalten ist die Menge $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ und die Menge $\mathbb{Z} = \mathbb{N}_0 \cup \{-x : x \in \mathbb{N}\}$ der ganzen Zahlen.

Auch das Rechnen mit den reellen Zahlen lernt man schon an der Schule. Hier sollen die Regeln noch einmal wiederholt werden, in einer zumindest teilweise etwas ungewohnten Form. Dabei verwenden wir zur Abkürzung die folgenden Symbole:

$$\begin{aligned} \forall x : E(x) & \quad \text{heißt:} \quad \mathbf{Für alle } x \text{ gilt } E(x), \\ \exists x : E(x) & \quad \text{heißt:} \quad \mathbf{Es gibt (wenigstens) ein } x \text{ mit } E(x). \end{aligned}$$

(I) Die Regeln der Addition

1. **Assoziativgesetz:** $\forall x, y, z \in \mathbb{R} : (x + y) + z = x + (y + z).$
2. **Kommutativgesetz:** $\forall x, y \in \mathbb{R} : x + y = y + x.$
3. **Existenz der Null:** \exists eine Zahl $0 \in \mathbb{R}$, so daß $\forall x \in \mathbb{R}$ gilt: $x + 0 = x.$
4. **Negatives:** $\forall x \in \mathbb{R} \exists$ eine Zahl $(-x) \in \mathbb{R}$ mit $x + (-x) = 0.$

(II) Die Regeln der Multiplikation

1. **Assoziativgesetz:** $\forall x, y, z \in \mathbb{R} : (x \cdot y) \cdot z = x \cdot (y \cdot z).$
2. **Kommutativgesetz:** $\forall x, y \in \mathbb{R} : x \cdot y = y \cdot x.$
3. **Existenz der Eins:** \exists eine Zahl $1 \in \mathbb{R}$, so daß $\forall x \in \mathbb{R}$ gilt: $x \cdot 1 = x.$
4. **Inverses:** $\forall x \in \mathbb{R}$ mit $x \neq 0 \exists x^{-1} \in \mathbb{R}$ mit $x \cdot x^{-1} = 1.$

Die Gesetze der Multiplikation sehen fast genauso aus wie bei der Addition, mit einer Ausnahme: Bei der Existenz des Inversen wird $x \neq 0$ vorausgesetzt. „Durch Null zu teilen ist verboten!“ Warum das so sein muß? Ist $x = 0$, so ist $x \cdot y = 0 \cdot y = 0$ für jede Zahl y . Es kann gar kein y mit $0 \cdot y = 1$ geben.

Der Zusammenhang zwischen Addition und Multiplikation wird folgendermaßen hergestellt:

(III) Das Distributivgesetz

Distributivgesetz: $\forall x, y, z \in \mathbb{R} : x \cdot (y + z) = x \cdot y + x \cdot z.$

Hierbei dürfen Addition und Multiplikation nicht vertauscht werden! Bei den Mengenverknüpfungen $A \cup B$ und $A \cap B$ gelten auch die Assoziativ- und Kommutativgesetze, darüber hinaus aber **zwei** Distributivgesetze:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

und

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Definition:

Eine Menge K mit zwei Rechen-Verknüpfungen $+$ und \cdot heißt ein *Körper*, falls alle Regeln der Addition und der Multiplikation, sowie das Distributivgesetz erfüllt sind.

Die Menge der reellen Zahlen bildet einen Körper. Daß wir einen eigenen Namen dafür haben, läßt einen vermuten, daß es noch weitere Beispiele von Körpern gibt. Das ist tatsächlich der Fall:

Beispiele.

1. Alle reellen Zahlen, deren Dezimaldarstellung schließlich periodisch wird, nennt man auch *rationale Zahlen*. Sie können stets in der Form eines Bruches

$$x = \frac{a}{b} \quad \text{mit } a \in \mathbb{Z} \text{ und } b \in \mathbb{N}$$

geschrieben werden.

Ist etwa $x = 0.\underline{142857}142857\dots$, so ist $10^6 \cdot x - x = 142857$, also

$$x = \frac{142857}{999999} = \frac{15873}{111111} = \frac{1443}{10101} = \frac{481}{3367} = \frac{1}{7}.$$

Umgekehrt besitzt jeder Bruch eine periodische Dezimaldarstellung. Diese aus der Schule bekannte Tatsache wollen wir hier nicht beweisen.

Addition und Multiplikation zweier Brüche ergibt wieder einen Bruch. Null und Eins kann man als Brüche schreiben. Das Negative eines Bruches bzw. das Inverse eines Bruches $\neq 0$ ist ebenfalls wieder einen Bruch. Deshalb bildet die Menge \mathbb{Q} aller *rationalen Zahlen* einen Körper.

2. Sei $\mathbb{F}_2 := \{0, 1\}$. Eine Addition und eine Multiplikation auf \mathbb{F}_2 definiert man durch die folgenden Verknüpfungstabellen:

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \qquad \begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

Beide Verknüpfungen sind kommutativ, weil die Tabellen symmetrisch zur Diagonalen sind. Auch sieht man sofort, daß 0 das neutrale Element bzgl. der Addition und 1 das neutrale Element bzgl. der Multiplikation ist. Zu jedem $x \in \mathbb{F}_2$ existiert das Negative (z.B. ist $-1 = 1$) und zu der 1 (dem einzigen Element $\neq 0$) existiert ein multiplikatives Inverses, es ist $1^{-1} = 1$.

Was zunächst reichlich unsinnig anmutet, wird durchaus sinnvoll, wenn man die Dinge etwas anders interpretiert. Es gibt bekanntlich gerade und ungerade ganze Zahlen. Wir schreiben g für „gerade“ und u für „ungerade“. Ersetzen wir nun 0 durch g und 1 durch u , so erhalten wir die folgenden Tabellen:

$$\begin{array}{c|cc} + & g & u \\ \hline g & g & u \\ u & u & g \end{array} \qquad \begin{array}{c|cc} \cdot & g & u \\ \hline g & g & g \\ u & g & u \end{array}$$

Addieren wir etwa eine gerade und eine ungerade Zahl, so erhalten wir als Ergebnis eine ungerade Zahl. Auf diese Weise können wir die Tabellen interpretieren. Daß das so gut klappt, liegt an folgender Tatsache: Teilt man eine gerade Zahl durch 2, so bleibt der Rest 0. Teilt man eine ungerade Zahl durch 2, so bleibt der Rest 1. Addieren wir zwei Zahlen, so addieren sich die Reste. Aber den Rest 2 müssen wir wieder durch den Rest 0 ersetzen.

Da auch die Assoziativgesetze und das Distributivgesetz in \mathbb{F}_2 gelten (wie man durch Testen der endlich vielen Möglichkeiten schnell herausfindet), ist \mathbb{F}_2 tatsächlich ein Körper.

Eine reelle Zahl x ist entweder positiv (d.h. $x \in \mathbb{R}_+$) oder $= 0$ oder negativ (d.h. $-x \in \mathbb{R}_+$). Sind x, y positiv, so ist auch $x + y$ und $x \cdot y$ positiv.

Allgemein heißt ein Körper K *angeordnet*, wenn man $K \setminus \{0\}$ so in die Menge der positiven Elemente, die Null und die Menge der negativen Elemente zerlegen kann, daß Summe und Produkt von positiven Elementen wieder positiv ist. Also ist \mathbb{R} (und genauso \mathbb{Q}) ein angeordneter Körper. Der Körper \mathbb{F}_2 ist nicht angeordnet, denn es ist $-1 = 1$, und das bedeutet, daß 1 gleichzeitig positiv und negativ sein müßte.

Man definiert:

$$\begin{aligned}
a < b & : \iff b - a > 0, \\
a \leq b & : \iff a < b \text{ oder } a = b, \\
a > b & : \iff b < a, \\
a \geq b & : \iff b \leq a.
\end{aligned}$$

Noch eine Anmerkung zum Definieren: Wird ein neues Objekt definiert, so verwendet man das Gleichheitszeichen, mit einem Doppelpunkt auf der Seite, wo das neue Objekt steht, z.B.

$$\mathbb{R}_- := \{x \in \mathbb{R} : x < 0\}.$$

Wird dagegen eine neue Eigenschaft definiert (z.B. $a < b$), so verwendet man das Äquivalenzzeichen, auch wieder mit einem Doppelpunkt auf der Seite, auf der die zu definierende Eigenschaft steht.

Satz. a, b, c seien reelle Zahlen. Ist $a < b$ und $b < c$, so ist auch $a < c$.

BEWEIS: Nach Definition ist $b - a > 0$ und $c - b > 0$. Dann ist aber auch die Summe $c - a = (b - a) + (c - b) > 0$, also $a < c$. ■

Satz. a, b, c seien reelle Zahlen, und $a < b$.

1. Es ist $a + c < b + c$.
2. Ist $c > 0$, so ist auch $a \cdot c < b \cdot c$.
Ist $c < 0$, so ist $a \cdot c > b \cdot c$.

BEWEIS: 1) Aus $b - a > 0$ folgt $(b + c) - (a + c) > 0$.

2) Ist $c > 0$, so ist $b \cdot c - a \cdot c = (b - a) \cdot c > 0$.

Ist $c < 0$, so ist $-c > 0$, also $a \cdot (-c) < b \cdot (-c)$. Daraus folgt: $(b - a) \cdot (-c) > 0$, also $ac - bc > 0$ und damit $ac > bc$. ■

Unter einem *Intervall* in \mathbb{R} versteht man eine der Mengen

$$\begin{aligned}
[a, b] &= \{x \in \mathbb{R} : a \leq x \leq b\}, \\
(a, b) &= \{x \in \mathbb{R} : a < x < b\}, \\
[a, b) &= \{x \in \mathbb{R} : a \leq x < b\} \\
\text{und } (a, b] &= \{x \in \mathbb{R} : a < x \leq b\}.
\end{aligned}$$

Wir führen noch zwei zusätzliche Symbole $-\infty$ und $+\infty$ ein. Sie gehören nicht zur Menge der reellen Zahlen, aber es gelten die Beziehungen $-\infty < +\infty$ und $-\infty < x$ bzw. $x < +\infty$ für $x \in \mathbb{R}$. Die Menge $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ nennt man die *erweiterte Zahlengerade*. Jetzt können wir auch Intervalle der Form

$$\begin{aligned}
(-\infty, b) &= \{x \in \mathbb{R} : x < b\}, \\
(-\infty, b] &= \{x \in \mathbb{R} : x \leq b\}, \\
(a, +\infty) &= \{x \in \mathbb{R} : x > a\} \\
\text{und } [a, +\infty) &= \{x \in \mathbb{R} : x \geq a\}
\end{aligned}$$

betrachten. Insbesondere ist $(-\infty, +\infty) = \mathbb{R}$.

Sind A_1, \dots, A_n endlich viele Mengen, so ist

$$\begin{aligned} A_1 \cup \dots \cup A_n &= \{x : (x \in A_1) \vee \dots \vee (x \in A_n)\} \\ &= \{x : \exists i \text{ mit } 1 \leq i \leq n \text{ und } x \in A_i\} \end{aligned}$$

und

$$\begin{aligned} A_1 \cap \dots \cap A_n &= \{x : (x \in A_1) \wedge \dots \wedge (x \in A_n)\} \\ &= \{x : \forall i \text{ mit } 1 \leq i \leq n \text{ gilt: } x \in A_i\} \end{aligned}$$

Das läßt sich auf Systeme von unendlich vielen Mengen übertragen:

Sind A_1, A_2, A_3, \dots unendlich viele Mengen, so definiert man:

$$\begin{aligned} \bigcup_{i=1}^{\infty} A_i &= \{x : \exists i \in \mathbb{N} \text{ mit } x \in A_i\} \\ \text{und } \bigcap_{i=1}^{\infty} A_i &= \{x : \forall i \in \mathbb{N} \text{ ist } x \in A_i\}. \end{aligned}$$

§ 2 Abbildungen

Inhalt:

Kartesisches Produkt von Mengen, anschauliche Ebene, Geraden, Abstandsbegriff, Kreise.

Begriff der Abbildung (=Funktion) als eindeutige Zuordnung zwischen Mengen, Graph einer Funktion, Beispiele für Abbildungen (z.B. Translation, ggT, Betragsfunktion), Bildmenge, injektive, surjektive und bijektive Abbildungen, Komposition von Abbildungen, Umkehrabbildung, Nachweis der Bijektivität über die Umkehrabbildung.

Ein paar Anmerkungen über endliche, abzählbare und überabzählbare Mengen.

Definition:

Sind M und N zwei Mengen, so bezeichnet man die Menge

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

als *kartesisches (oder direktes) Produkt* von M und N . Die Elemente (x, y) nennt man *(geordnete) Paare*.

Die Mengen $\{2, 3\}$, $\{3, 2\}$ und $\{2, 2, 3, 2, 3\}$ sind alle gleich, weil sie die gleichen Elemente besitzen. Bei einem Paar (x, y) kommt es dagegen auf die Reihenfolge an. Es ist

$$(x, y) = (x', y') \iff x = x' \text{ und } y = y'.$$

Sind Mengen M_1, M_2, \dots, M_n gegeben, so ist ihr *kartesisches Produkt* die Menge

$$M_1 \times M_2 \times \dots \times M_n = \{(x_1, x_2, \dots, x_n) : x_i \in M_i \text{ für } i = 1, \dots, n\}.$$

Die Elemente (x_1, \dots, x_n) nennt man *n -Tupel*.

Ist $M_1 = M_2 = \dots = M_n = M$, so schreibt man auch M^n an Stelle des n -fachen Produktes $M \times \dots \times M$.

Mit Hilfe eines Koordinatensystems kann man die Punkte der Ebene (bzw. des Raumes) durch die Elemente von \mathbb{R}^2 (bzw. $\mathbb{R}^3, \mathbb{R}^4, \dots$) beschreiben. Im Augenblick beschränken wir uns auf die Ebene \mathbb{R}^2 .

Der Punkt $P_0 := (0, 0)$ heißt *Nullpunkt* oder *Ursprung*. Eine Menge der Gestalt

$$L = \{(x, y) \in \mathbb{R}^2 : ax + by = r\}, \text{ mit } a, b, r \in \mathbb{R} \text{ und } (a, b) \neq (0, 0)$$

heißt *Gerade*. Ist $b = 0$, so ist $L = \{(x, y) \in \mathbb{R}^2 : x = r/a\}$ eine sogenannte „vertikale Gerade“. In jedem anderen Fall hat man die Darstellung

$$\{(x, y) \in \mathbb{R}^2 : y = mx + c\}, \text{ mit } m = -\frac{a}{b} \text{ und } c = \frac{r}{b}.$$

Die Zahl m nennt man die *Steigung* der Geraden, und die Zahl c den *Ordinatenabschnitt*.

Sind $P_1 = (x_1, y_1)$ und $P_2 = (x_2, y_2)$ zwei Punkte des \mathbb{R}^2 , so versteht man unter dem *Abstand* von P_1 und P_2 die Zahl

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Diese Zahl ergibt sich aus dem Satz des Pythagoras.

Der *Kreis* mit dem *Mittelpunkt* (x_0, y_0) und dem *Radius* $r > 0$ ist die Menge

$$\{(x, y) \in \mathbb{R}^2 : d((x, y), (x_0, y_0)) = r\} = \{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 = r^2\}.$$

Es seien nun X und Y zwei nicht-leere Mengen. Unter einer *Abbildung* oder *Funktion* f von X nach Y versteht man eine Vorschrift, die jedem $x \in X$ in eindeutiger Weise genau ein $y \in Y$ zuordnet. Man schreibt dann

$$f : X \rightarrow Y.$$

Die Menge X nennt man den *Definitionsbereich* und die Menge Y den *Wertebereich* oder die *Zielmenge* von f . Ordnet f einem Element $x \in X$ das Element $y \in Y$ zu, so schreibt man

$$f : x \mapsto y \quad \text{oder} \quad y = f(x).$$

Man nennt dann y das Bild von x unter f . Es ist eindeutig bestimmt. Umgekehrt heißt x in diesem Fall ein Urbild von y . Dieses ist nicht eindeutig bestimmt, ein y kann mehrere Urbilder oder auch gar kein Urbild besitzen.

Jeder Abbildung $f : X \rightarrow Y$ kann die Menge der Paare (x, y) mit $x \in X$ und $y = f(x)$ zugeordnet werden, der *Graph* von f :

$$G_f := \{(x, f(x)) : x \in X\}.$$

Zu einer Abbildung gehören Definitionsbereich, Zuordnungsvorschrift und Wertebereich. Alles zusammen ist durch den Graphen bestimmt.

Beispiele.

1. Häufig wird eine Abbildung durch eine Funktionsvorschrift definiert, z.B.

$$f : \mathbb{Z} \rightarrow \mathbb{N}_0 \quad \text{mit } f(z) := z^2.$$

2. Zu jeder Menge M gibt es die *identische Abbildung* $\text{id}_M : M \rightarrow M$, mit $\text{id}_M : x \mapsto x$.

3. Sind zwei Zahlen $a, b \in \mathbb{R}$ gegeben, so nennt man die Abbildung

$$T : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \quad \text{mit} \quad T(x, y) := (x + a, y + b)$$

eine *Translation* der Ebene. Sie verschiebt jeden Punkt um a Einheiten nach rechts und b Einheiten nach oben.

4. Manchmal ist die Zuordnungsvorschrift durch einen Algorithmus gegeben. So kann man z.B. jedem Paar $(n, m) \in \mathbb{N} \times \mathbb{N}$ den *größten gemeinsamen Teiler* $\text{ggT}(n, m)$ zuordnen. Zur Berechnung dient der *Euklidische Algorithmus*.

Sei etwa $n > m$. Jeder kennt aus der Schule die *Division mit Rest*: Es gibt Zahlen $q, r \in \mathbb{N}_0$, so daß gilt:

$$(a) \quad n = q \cdot m + r,$$

$$(b) \quad 0 \leq r < m.$$

Dabei nennt man q den *ganzzahligen Quotient* von n und m , und r den *Divisionsrest*. Es gibt sogar Taschenrechner, die q und r auf Knopfdruck liefern.

Der Euklidische Algorithmus verlangt, daß wir wie folgt fortgesetzt ganzzahlig mit Rest dividieren:

$$\begin{aligned} n &= q_0 \cdot m + r_1, & \text{mit } 0 \leq r_1 < m, \\ m &= q_1 \cdot r_1 + r_2, & \text{mit } 0 \leq r_2 < r_1, \\ r_1 &= q_2 \cdot r_2 + r_3, & \text{mit } 0 \leq r_3 < r_2, \\ &\vdots \\ r_{N-2} &= q_{N-1} \cdot r_{N-1} + r_N, & \text{mit } 0 \leq r_N < r_{N-1}, \\ r_{N-1} &= q_N \cdot r_N. \end{aligned}$$

Da $n > m > r_1 > r_2 > \dots \geq 0$ ist, muß die Division irgendwann aufgehen, der Prozeß also abbrechen. Bezeichnet man für eine natürliche Zahl a die Menge aller (ganzzahligen) Teiler von a mit T_a , so ist

$$\text{ggT}(a, b) = \max(T_a \cap T_b).$$

Mit $\max(M)$ bezeichnen wir dabei die größte Zahl einer endlichen Menge M . Ist etwa $a = 6$ und $b = 15$, so ist $T_a = \{1, 2, 3, 6\}$ und $T_b = \{1, 3, 5, 15\}$, $T_a \cap T_b = \{1, 3\}$, also $\text{ggT}(a, b) = 3$.

Im Falle des Euklidischen Algorithmus rollen wir die Sache von hinten auf. Aus der Gleichung $r_{N-1} = q_N \cdot r_N$ ergibt sich, daß $T_{r_N} \subset T_{r_{N-1}}$ ist, und damit sogar $T_{r_N} \cap T_{r_{N-1}} = T_{r_N}$. Also ist $\text{ggT}(r_N, r_{N-1}) = \max(T_{r_N}) = \max\{1, \dots, \text{weitere Teiler } \dots, r_N\} = r_N$.

Aus der Gleichung $r_{N-2} = q_{N-1} \cdot r_{N-1} + r_N$ ergibt sich die Beziehung $T_{r_N} \cap T_{r_{N-1}} = T_{r_{N-1}} \cap T_{r_{N-2}}$ und damit $\text{ggT}(r_N, r_{N-1}) = \text{ggT}(r_{N-1}, r_{N-2})$. So fährt man fort und erhält schließlich:

$$r_N = \text{ggT}(r_{N-1}, r_N) = \text{ggT}(r_{N-2}, r_{N-1}) = \dots = \text{ggT}(n, m).$$

Da (von oben nach unten betrachtet) jeder Schritt eindeutig ist, liefert der Algorithmus eine Abbildung.

Um etwa den ggT von 378 und 550 zu bestimmen, führen wir folgende Divisionen aus:

$$\begin{aligned} 550 &= 1 \cdot 378 + 172, \\ 378 &= 2 \cdot 172 + 34, \\ 172 &= 5 \cdot 34 + 2, \\ 34 &= 17 \cdot 2 + 0. \end{aligned}$$

Also ist $\text{ggT}(378, 550) = 2$.

5. Aus der Schule sind vor allem Abbildungen $f : \mathbb{R} \rightarrow \mathbb{R}$ bekannt, also sogenannte reellwertige Funktionen. Als Beispiel soll hier die *Betragsfunktion* erwähnt werden:

$$\text{Für } x \in \mathbb{R} \text{ sei } |x| := \begin{cases} x & \text{falls } x \geq 0, \\ -x & \text{falls } x < 0. \end{cases}$$

Mit Hilfe der Betragsfunktion kann man sich auch andere Funktionen zu-rechtbasteln, etwa $f : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f(x) := |2x + 1|.$$

Es ist $2x + 1 \geq 0 \iff x \geq -1/2$. Also gilt:

$$f(x) = \begin{cases} 2x + 1 & \text{falls } x \geq -1/2, \\ -2x - 1 & \text{falls } x < -1/2. \end{cases}$$

Definition:

Ist $f : X \rightarrow Y$ eine Abbildung und $M \subset X$, so nennt man

$$f(M) := \{f(x) : x \in M\}$$

die *Bildmenge* von M unter f .

Definition:

Es sei $f : X \rightarrow Y$ eine beliebige Abbildung.

1. f heißt *injektiv*, falls für alle $x_1, x_2 \in X$ gilt:

$$x_1 \neq x_2 \implies f(x_1) \neq f(x_2).$$

2. f heißt *surjektiv*, falls gilt:

$$f(X) = Y, \quad \text{d.h. } \forall y \in Y \exists x \in X \text{ mit } f(x) = y.$$

3. f heißt *bijektiv*, falls f injektiv und surjektiv ist.

f ist also genau dann injektiv, wenn die Gleichung $f(x) = y$ für jedes $y \in f(X)$ eindeutig lösbar ist. Und f ist genau dann surjektiv, wenn die Gleichung $f(x) = y$ für jedes y aus dem Wertebereich Y lösbar ist. Insbesondere ist f genau dann bijektiv, wenn die Gleichung $f(x) = y$ für jedes $y \in Y$ eindeutig lösbar ist.

Bei einer reellwertigen Funktion (und nur bei einer solchen!) kann man Injektivität und Surjektivität wie folgt am Graphen ablesen:

f ist injektiv, wenn jede horizontale Gerade $\{(x, y) : y = \text{const.}\}$ den Graphen höchstens in einem Punkt trifft. f ist surjektiv, wenn jede solche horizontale Gerade (mit y im Wertebereich) den Graphen mindestens einmal trifft.

Beispiele.

1. Die durch $f(x) := x^2$ definierte Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ ist nicht injektiv, denn es ist $2 \neq -2$, aber $f(2) = f(-2)$. Sie ist auch nicht surjektiv, denn zu $y = -1$ gibt es kein Urbild.

Allerdings ist die „Einschränkung“ von f auf $X := \mathbb{R}_+$ injektiv. Ist nämlich $0 < x_1 < x_2$, also $x_2 = x_1 + d$ mit $d > 0$, so ist

$$f(x_2) = (x_1 + d)^2 = x_1^2 + 2x_1d + d^2 > f(x_1).$$

Benutzt man $Y := \mathbb{R}_+$ als Wertebereich, so ist $f : X \rightarrow Y$ zusätzlich surjektiv und damit bijektiv.

2. Die „Projektion“ $p : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit $p(x, y) := x$ ist zwar surjektiv, aber sie ist nicht injektiv, und man kann sie auch auf keine Weise injektiv machen.
3. Die Abbildung $(m, n) \mapsto \text{ggT}(m, n)$ ist surjektiv, aber nicht injektiv.
4. Die Abbildung $f : \mathbb{N} \rightarrow \mathbb{N}$ mit $f(n) := 2 \cdot n$ ist injektiv, aber nicht surjektiv.
5. Jede Translation $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ ist bijektiv.

Definition:

Es seien zwei Abbildungen $f : X \rightarrow Y$ und $g : Y \rightarrow Z$ gegeben. Dann kann man die *Verknüpfung* (*Komposition*, *Verkettung*) $g \circ f : X \rightarrow Z$ bilden. Sie ist definiert durch

$$(g \circ f)(x) := g(f(x)).$$

Achtung! Obwohl man zuerst f und dann g anwendet, schreibt man (aus nahe liegenden Gründen) $g \circ f$. Das führt am Anfang manchmal zu Verwirrungen.

Beispiele.

1. Seien $f : \mathbb{R} \rightarrow \mathbb{R}$ und $g : \mathbb{R} \rightarrow \mathbb{R}$ definiert durch $f(x) := x^2$ und $g(x) := x - \alpha$ (mit einer festen reellen Zahl α). Dann kann man sowohl $g \circ f$ als auch $f \circ g$ bilden, und es ist

$$(g \circ f)(x) = x^2 - \alpha \quad \text{und} \quad (f \circ g)(x) = (x - \alpha)^2 = x^2 - 2\alpha x + \alpha^2.$$

Es kommt also auf die Reihenfolge an!

2. Sei $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ die durch $T(x, y) := (x + a, y + b)$ gegebene Translation und $p : \mathbb{R}^2 \rightarrow \mathbb{R}$ die durch $p(x, y) := x$ gegebene Projektion. Dann kann man nur die Verknüpfung $p \circ T : \mathbb{R}^2 \rightarrow \mathbb{R}$ bilden. Offensichtlich ist $(p \circ T)(x, y) = x + a$.
3. Sei $I \subset \mathbb{R}$ ein Intervall und $f : I \rightarrow \mathbb{R}$ eine Funktion. Wir definieren dann eine Abbildung $F : I \rightarrow \mathbb{R}^2$ durch $F(t) := (t, f(t))$. Dann ist $F(I) = G_f$. Sind $p_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$ und $p_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ die durch $p_1(x, y) := x$ und $p_2(x, y) := y$ gegebenen Projektionen auf die erste bzw. zweite Komponente, so ist $p_1 \circ F = \text{id}_I$ und $p_2 \circ F = f$.
4. Sei $k \in \mathbb{N}$ und $M_k : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$ definiert durch $M_k(n, m) := (kn, km)$. Dann ist $(\text{ggT} \circ M_k)(n, m) = k \cdot \text{ggT}(n, m)$.

Ist $f : X \rightarrow Y$ **bijektiv**, so besitzt jedes Element $y \in Y$ ein eindeutig bestimmtes Urbild x mit $f(x) = y$. Durch $f^{-1}(y) := x$ wird dann eine Abbildung $f^{-1} : Y \rightarrow X$ definiert, die *Umkehrabbildung* von f .

Beispiele.

1. Die Abbildung $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ mit $f(x) := x^2$ ist bijektiv. Die Umkehrabbildung ist wohlbekannt, es ist $f^{-1}(y) = \sqrt{y}$.
2. Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ definiert durch $f(x) := x^3 - x$. Diese Funktion ist sicher nicht injektiv, denn es ist z.B. $f(0) = f(1)$. Ich behaupte aber, daß sie auf dem offenen Intervall $I = (1, \infty)$ injektiv ist.

Um das zu beweisen, benutze ich das logische Prinzip der *Kontraposition*:

$$(\mathcal{A} \implies \mathcal{B}) \iff (\neg \mathcal{B} \implies \neg \mathcal{A})$$

Anstatt die Implikation $(x_1 \neq x_2) \implies (f(x_1) \neq f(x_2))$ zu zeigen, beweisen wir die Implikation $(f(x_1) = f(x_2)) \implies (x_1 = x_2)$.

Seien $x_1, x_2 \in I$ mit $f(x_1) = f(x_2)$. Dann ist $x_1^3 - x_2^3 = x_1 - x_2$. Aus der Beziehung

$$(x_1 - x_2)^3 = x_1^3 - 3x_1^2x_2 + 3x_1x_2^2 - x_2^3$$

folgt dann: $(x_1 - x_2)^3 + 3x_1x_2(x_1 - x_2) = x_1 - x_2$. Ist $x_1 = x_2$, so ist alles gut. Ist aber $x_1 \neq x_2$, so kann man durch $x_1 - x_2$ kürzen und erhält:

$$(x_1 - x_2)^2 + 3x_1x_2 = 1, \quad \text{also } x_1^2 + x_1x_2 + x_2^2 = 1.$$

Das ist unmöglich, denn für $x_1, x_2 \in I$ ist die linke Seite größer als 3.

Leider ist die Einschränkung $f|_I : I \rightarrow \mathbb{R}$ nicht surjektiv. Für $x > 1$ ist $f(x) = x(x^2 - 1) > 0$. Wir werden später sehen, daß jeder Wert aus \mathbb{R}_+ angenommen wird. Also ist $f : I \rightarrow \mathbb{R}_+$ bijektiv, und es muß dazu eine Umkehrabbildung geben. Wir sind allerdings (im Augenblick) nicht in der Lage, sie explizit anzugeben.

3. Die Translation $T : (x, y) \mapsto (x + a, y + b)$ ist bijektiv, ihre Umkehrabbildung ist wieder eine Translation, nämlich $T^{-1} : (x, y) \mapsto (x - a, y - b)$.

Ist $f : X \rightarrow Y$ bijektiv, so gilt:

$$f^{-1} \circ f = \text{id}_X \quad \text{und} \quad f \circ f^{-1} = \text{id}_Y.$$

Die erste Gleichung ist nur eine andere Schreibweise für die Definition der Umkehrabbildung. Aber auch die zweite Gleichung folgt ganz leicht: Ist $y_0 \in Y$, so gibt es wegen der Bijektivität von f genau ein $x_0 \in X$ mit $f(x_0) = y_0$. Nach Definition der Umkehrabbildung ist $f^{-1}(y_0) = x_0$ und daher $f \circ f^{-1}(y_0) = f(x_0) = y_0$.

Interessant ist nun die Umkehrung dieser Beziehung:

Nachweis der Bijektivität über die Umkehrabbildung

Es sei eine Abbildung $f : X \rightarrow Y$ gegeben. Wenn es eine Abbildung $g : Y \rightarrow X$ mit $g \circ f = \text{id}_X$ und $f \circ g = \text{id}_Y$ gibt, dann ist f bijektiv und $g = f^{-1}$.

BEWEIS: 1) Ist $f(x_1) = f(x_2)$, so ist auch $x_1 = g(f(x_1)) = g(f(x_2)) = x_2$. Also ist f injektiv.

2) Ist $y_0 \in Y$, so liegt $x_0 := g(y_0)$ in X , und es ist $f(x_0) = f(g(y_0)) = y_0$. Das zeigt, daß f surjektiv ist.

Weil $g(f(x)) = x$ ist, ist g die Umkehrabbildung. ■

Folgerung. *Die Umkehrabbildung einer bijektiven Abbildung ist wieder bijektiv.*

Beispiele.

1. Sei $f(x) := \frac{2x+3}{x-3}$. Diese Funktion ist nicht auf ganz \mathbb{R} definiert, es muß $x \neq 3$ sein. Wir fragen uns, welche Werte vorkommen können. Sei also $y \in \mathbb{R}$ und $f(x) = y$. Dann ist $y(x-3) = 2x+3$, also $x(y-2) = 3y+3$. Ist $y_0 \neq 2$, so ist $x_0 := \frac{3y_0+3}{y_0-2}$ eine reelle Zahl, und es ist

$$f(x_0) = \frac{2x_0+3}{x_0-3} = \frac{(6y_0+6)/(y_0-2)+3}{(3y_0+3)/(y_0-2)-3} = \frac{9y_0}{9} = y_0.$$

Wäre andererseits $f(x) = 2$, so wäre $2(x-3) = 2x+3$, also $-6 = 3$. Das kann nicht sein, also ist $f : \mathbb{R} \setminus \{3\} \rightarrow \mathbb{R} \setminus \{2\}$ surjektiv. Die Abbildung $g : \mathbb{R} \setminus \{2\} \rightarrow \mathbb{R}$ mit $g(y) := \frac{3y+3}{y-2}$ kann nicht den Wert 3 annehmen, denn dann wäre $3(y-2) = 3y+3$, also wieder $-6 = 3$.

Wir haben nachgerechnet, daß $f(g(y)) = y$ ist, und genauso erhält man, daß $g(f(x)) = x$ ist. Die Abbildung f ist bijektiv und g ihre Umkehrabbildung. Wir haben zwar implizit die Surjektivität von f bewiesen, indem wir die Umkehrabbildung konstruiert haben, aber ein expliziter Beweis der Injektivität war nicht mehr erforderlich.

2. Sei $F : \mathbb{R} \rightarrow \mathbb{R}^2$ definiert durch

$$F(t) := \left(\frac{2t}{t^2+1}, \frac{t^2-1}{t^2+1} \right).$$

Dann gilt für jeden Bildpunkt $(x, y) := F(t)$ die Gleichung

$$x^2 + y^2 = \frac{4t^2 + (t^2 - 1)^2}{(t^2 + 1)^2} = \frac{(t^2 + 1)^2}{(t^2 + 1)^2} = 1.$$

Das bedeutet, daß (x, y) auf dem Einheitskreis

$$K = \{(x, y) \in \mathbb{R}^2 : d((x, y), (0, 0)) = 1\}$$

liegt. Allerdings kommt $(0, 1)$ nicht als Bildpunkt vor, denn wenn $x = 0$ ist, muß $t = 0$ und damit $y = -1$ sein.

Definieren wir nun $H : K \setminus \{(0, 1)\} \rightarrow \mathbb{R}$ durch

$$H(x, y) := \frac{x}{1 - y},$$

so ist

$$H(F(t)) = \frac{(2t)/(t^2 + 1)}{1 - (t^2 - 1)/(t^2 + 1)} = \frac{2t}{(t^2 + 1) - (t^2 - 1)} = t$$

Das bedeutet insbesondere, daß H surjektiv ist. Für $(x, y) \in K \setminus \{(0, 1)\}$ gilt andererseits:

$$\begin{aligned} F(H(x, y)) &= F\left(\frac{x}{1 - y}\right) \\ &= \left(\frac{(2x)/(1 - y)}{x^2/(1 - y)^2 + 1}, \frac{x^2/(1 - y)^2 - 1}{x^2/(1 - y)^2 + 1} \right) \\ &= \left(\frac{2x(1 - y)}{2 - 2y}, \frac{x^2 - (1 - y)^2}{2 - 2y} \right) \\ &= \left(\frac{2x(1 - y)}{2(1 - y)}, \frac{2y(1 - y)}{2(1 - y)} \right) = (x, y). \end{aligned}$$

Damit ist F bijektiv und $H = F^{-1}$. Man nennt F die *rationale Parametrisierung* des Einheitskreises. Für die Abbildung H gibt es eine geometrische Deutung: Die Gerade durch den Punkt $(0, 1)$ und einen Punkt (x, y) des Einheitskreises schneidet die x -Achse genau im Punkt $(H(x, y), 0)$.

Satz. Ist $n < m$, so gibt es keine bijektive Abbildung von $A_n := \{1, 2, \dots, n\}$ auf $A_m := \{1, 2, \dots, m\}$.

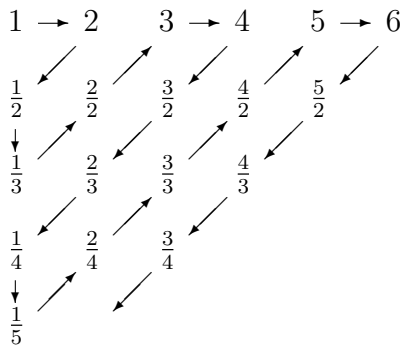
Der BEWEIS (z.B. durch Widerspruch) ist eine kleine Knobelaufgabe und sei dem interessierten Leser überlassen.

Definition:

Eine Menge M heißt *endlich*, falls es eine bijektive Abbildung von einer Menge $A_n = \{1, 2, 3, \dots, n\}$ nach M gibt. Die Zahl n nennt man dann die *Anzahl der Elemente* von M .

Ist M nicht endlich, so heißt M eine *unendliche Menge*. Speziell heißt M *abzählbar*, falls es eine bijektive Abbildung von \mathbb{N} nach M gibt. Ist M weder endlich noch abzählbar, so nennt man M *überabzählbar*.

Die Mengen \mathbb{N} , \mathbb{N}_0 und \mathbb{Z} sind natürlich abzählbar. Aber auch \mathbb{Q} ist abzählbar, wie das *Cantorsche Diagonalverfahren* zeigt:



Satz. Die Menge \mathbb{R} der reellen Zahlen ist nicht abzählbar.

BEWEIS: Wir beschränken uns auf reelle Zahlen zwischen 0 und 1 und führen den Beweis durch Widerspruch. Wäre die Menge der reellen Zahlen zwischen 0 und 1 abzählbar, so könnte man alle diese Zahlen in einer unendlichen Kolonne hintereinander aufschreiben:

$$\begin{aligned}
 x_1 &= 0 . a_{11} a_{12} a_{13} \dots , \\
 x_2 &= 0 . a_{21} a_{22} a_{23} \dots , \\
 x_3 &= 0 . a_{31} a_{32} a_{33} \dots , \\
 &\vdots
 \end{aligned}$$

Die Ziffern a_{ij} nehmen dabei wie üblich Werte zwischen 0 und 9 an.

Nun konstruieren wir eine reelle Zahl $y = 0 . c_1 c_2 c_3 \dots$ wie folgt:

$$\text{Es sei } c_i := \begin{cases} 5 & \text{falls } a_{ii} \neq 5 \\ 4 & \text{falls } a_{ii} = 5 \end{cases}$$

Offensichtlich liegt y zwischen 0 und 1 und muß unter den Folgengliedern x_1, x_2, x_3, \dots vorkommen. Es gibt also ein $n \in \mathbb{N}$, so daß $y = x_n$ ist. Dann ist $c_n = a_{nn}$, im Widerspruch zur Definition der c_i . ■

§ 3 Elementare Kombinatorik

Inhalt:

Vollständige Induktion, Summen- und Produktzeichen, Potenzen und Fakultäten, geometrische Summenformel, Binomialkoeffizienten, binomische Formel, Bernoullische Ungleichung.

Wir betrachten jetzt ein Beweisverfahren, das sich auf die Struktur der Menge \mathbb{N} stützt.

Sei $E(n)$ eine Aussage über natürliche Zahlen und n_0 ein festes Element aus \mathbb{N} . Unser Ziel ist es, die Aussage $E(n)$ für alle $n \geq n_0$ zu beweisen. Ist etwa $E(n)$ die Aussage „ $2^n > 2$ “, so können wir das sicher für $n \geq 2$ beweisen. Wir fangen einfach mit $n = 2$ an:

$$2^2 = 4 > 2 \text{ (weiß ja jeder).}$$

Bei $n = 3$ können wir jetzt so schließen:

$$2^3 = 2 \cdot 2^2 = 4 + 4 > 2 + 2 = 4 > 2.$$

Dann geht's weiter:

$$2^4 = 2 \cdot 2^3 = 2^3 + 2^3 > 2 + 2 = 4 > 2,$$

$$2^5 = 2 \cdot 2^4 = 2^4 + 2^4 > 2 + 2 = 4 > 2, \text{ usw.}$$

Bei jedem neuen Schritt verwenden wir das vorhergehende Ergebnis. So gewinnen wir eine unendliche Kette von kleinen Schlüssen, an deren „Ende“ die gewünschte Aussage bewiesen ist.

Allgemein funktioniert dieses Prinzip so:

1. Zunächst zeigen wir, daß $E(n_0)$ wahr ist. Das ist der sogenannte **Induktionsanfang**.
2. Sei $n \geq n_0$ beliebig. Unter der Voraussetzung, daß $E(n)$ wahr ist, beweisen wir, daß auch $E(n + 1)$ wahr ist. Das ist der **Induktionsschluß** (oder **Induktionsschritt**).

Man stelle sich die natürlichen Zahlen als eine unendliche Reihe von Domino-Steinen vor. Kippt man den ersten Stein um, und sind alle Steine so aufgestellt, daß der n -te Stein, wenn er kippt, automatisch auch den $(n + 1)$ -ten Stein zum Kippen bringt, dann kippen sie alle.

Wichtig ist, daß man folgendes versteht: Es geht nicht darum, zu zeigen, daß $E(n)$ wahr ist! Vielmehr soll man aus der vorangegangenen Aussage $E(n)$ die Aussage $E(n+1)$ folgern.

Hier ist ein einfaches Beispiel:

Gaußsche Summenformel

$$\text{Für alle } n \in \mathbb{N} \text{ ist } 1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

BEWEIS: Wir führen Induktion nach n :

Induktionsanfang: Die linke Seite ist $= 1$, und die rechte $= \frac{1 \cdot (1+1)}{2} = 1$.

Induktionsschluß: Wir setzen voraus, daß die Formel für ein $n \in \mathbb{N}$ gilt. Dann ist

$$\begin{aligned} 1 + 2 + \cdots + n + (n+1) &= (1 + 2 + \cdots + n) + (n+1) = \frac{n(n+1)}{2} + (n+1) \\ &= \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}. \end{aligned}$$

Also gilt die Formel auch für $n+1$. ■

Solche Formeln werden gerne benutzt, um den Induktionsbeweis zu üben. In diesem Falle gibt es aber einen einfacheren Beweis, der zudem die Formel liefert:

Wir schreiben die Summe zweimal auf, dabei einmal in verkehrter Reihenfolge:

$$\begin{aligned} 2 \cdot (1 + 2 + \cdots + n) &= (1 + 2 + 3 + \cdots + n) \\ &\quad + (n + (n-1) + (n-2) + \cdots + 1) \\ &= n \cdot (n+1). \end{aligned}$$

In ähnlicher Form hat der kleine Gauß schon in der Schule das Problem gelöst, die Zahlen von 1 bis 100 zu addieren.

Als nächstes führen wir Abkürzungen für Summen und Produkte ein.

Definition:

Sind $a_1, a_2, a_3, \dots, a_n$ reelle Zahlen, so setzt man

$$\sum_{i=1}^n a_i := a_1 + a_2 + \dots + a_n$$

und

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

Die Symbole \sum und \prod nennt man *Summenzeichen* und *Produktzeichen*.

Der Summationsindex i darf auch mit einem anderen Buchstaben bezeichnet werden, und die Untergrenze muß nicht die 1 sein. So ist z.B.

$$a_k + a_{k+1} + \dots + a_{k+m} = \sum_{i=k}^{k+m} a_i = \sum_{j=0}^m a_{k+j}.$$

Ist $m < k$, so nennt man $\sum_{i=k}^m a_i$ eine *leere Summe*. Sie wird **definitionsgemäß** gleich Null gesetzt. Für das Produktzeichen gelten analoge Regeln. Das *leere Produkt* wird gleich Eins gesetzt.

Für $n \in \mathbb{N}$ wird die n -te *Potenz von a* definiert durch

$$a^n := \prod_{i=1}^n a = \underbrace{a \cdot a \cdot \dots \cdot a}_{n\text{-mal}}.$$

Wegen der Regelung für das leere Produkt ist $a^0 = 1$. Man beachte, daß dann $0^n = 0$ für jedes $n \in \mathbb{N}$ ist, aber $0^0 = 1$.

Aus der Definition ergeben sich sofort die **Rechenregeln für Potenzen**:

$$\begin{array}{l} a^{n+m} = a^n \cdot a^m \\ \text{und } a^{n \cdot m} = (a^n)^m. \end{array} \quad (0)$$

Man kann auch negative Exponenten zulassen:

$$a^{-n} := \left(\frac{1}{a}\right)^n, \text{ für } n \in \mathbb{N}, a \neq 0.$$

Definition:

Mit n -Fakultät bezeichnet man die Zahl

$$n! := \prod_{i=1}^n i = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n.$$

Dann ist $0! = 1! = 1$, und allgemein gilt die Rekursionsformel:

$$(n + 1)! = n! \cdot (n + 1).$$

Weitere Werte:

$$2! = 2, \quad 3! = 2 \cdot 3 = 6, \quad 4! = 2 \cdot 3 \cdot 4 = 24, \quad \dots, \quad 10! = 3\,628\,800.$$

Man kann die Zahl $n!$ auch auf andere Weise interpretieren:

Satz. *Es gibt genau $n!$ Möglichkeiten, die Zahlen $1, 2, 3, \dots, n$ anzuordnen.*

BEWEIS:

Induktionsanfang: Es gibt genau eine Möglichkeit, die Zahl 1 anzuordnen. Das ist so evident, da brauchen wir gar nichts zu zeigen.

Induktionsschluß: Wenn man die Zahlen $1, 2, 3, \dots, n$ auf genau $n!$ verschiedene Weisen anordnen kann, dann sieht die Situation bei den Zahlen $1, 2, 3, \dots, n, n + 1$ folgendermaßen aus:

Es gibt $n!$ mögliche Anordnungen, bei denen die Zahl $n + 1$ an letzter Stelle steht. Es kann aber auch jede andere Zahl zwischen 1 und $n + 1$ an letzter Stelle stehen, und für die verbliebenen n Zahlen stehen jedesmal wieder $n!$ mögliche Anordnungen zur Verfügung. Zusammen sind das $(n + 1) \cdot n! = (n + 1)!$ Möglichkeiten. ■

Geometrische Summenformel

Ist $q \neq 1$ eine reelle Zahl, so ist

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

BEWEIS: Man kann einfach ausmultiplizieren:

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

Division durch $q - 1$ ergibt das Ergebnis. ■

Der Trick mit den beiden Summen, wo sich die Summanden – bis auf das erste Glied der einen und das letzte Glied der anderen Summe – gegenseitig wegheben, kann häufig benutzt werden. Man spricht auch von „Teleskop-Summen“. Hier ist noch ein weiteres wichtiges Beispiel:

Satz. Ist $x \neq y$, so ist

$$\frac{x^n - y^n}{x - y} = \sum_{i=0}^{n-1} x^i y^{n-i-1}.$$

BEWEIS: Es ist

$$\begin{aligned} (x - y) \cdot \sum_{i=0}^{n-1} x^i y^{n-i-1} &= \\ &= \sum_{i=0}^{n-1} x^{i+1} y^{n-i-1} - \sum_{i=0}^{n-1} x^i y^{n-i} \\ &= (xy^{n-1} + \dots + x^{n-1}y + x^n) - (y^n + xy^{n-1} + \dots + x^{n-1}y) \\ &= x^n - y^n. \end{aligned}$$

■

Beispiele.

1. Der Fall $n = 2$ liefert die Formel $x^2 - y^2 = (x + y)(x - y)$.
2. Ist $y = 1$, so erhält man $\frac{x^n - 1}{x - 1} = 1 + x + x^2 + \dots + x^{n-1}$.

Definition:

Sei $0 \leq k \leq n$. Die Zahlen $\binom{n}{k} := \frac{n!}{k!(n-k)!}$ (in Worten: „ n über k “) nennt man *Binomialkoeffizienten*.

Hier sind einige Eigenschaften:

1. Es ist $\binom{n}{k} = \frac{(n-k+1) \cdot (n-k+2) \cdot \dots \cdot n}{1 \cdot 2 \cdot \dots \cdot k}$.
2. $\binom{n}{0} = \binom{n}{n} = 1$.

$$3. \binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}.$$

Überprüfen wir die letzte Aussage: Es ist

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

Die gerade bewiesene Formel erlaubt eine besonders einfache Berechnung der Binomialkoeffizienten – ohne Multiplikation! Allerdings müssen die Koeffizienten der Reihenfolge nach berechnet werden. Man kann sie dann in Form des sogenannten Pascalschen Dreiecks anordnen:

$$\begin{array}{ccccccc} n=0 & & & & & & 1 \\ n=1 & & & & 1 & & 1 \\ n=2 & & & 1 & 2 & 1 & \\ n=3 & & 1 & 3 & 3 & 1 & \\ & & & & \dots & & \end{array}$$

Es gibt auch eine kombinatorische Interpretation der Binomialkoeffizienten:

Wieviele Möglichkeiten gibt es, aus der Menge $\{1, 2, 3, \dots, n\}$ eine Teilmenge mit k Elementen auszuwählen? Dazu kann man folgendermaßen vorgehen:

Wir wählen eine beliebige Anordnung aller Zahlen von 1 bis n (davon gibt es bekanntlich $n!$). Anschließend nehmen wir die ersten k Zahlen. Auf diese Weise erhalten wir sicherlich alle k -elementigen Teilmengen, aber sie treten mehrfach auf. Wenn wir die ersten k Zahlen untereinander vertauschen (und dafür gibt es $k!$ Möglichkeiten), oder wenn wir die hinteren $n-k$ Zahlen untereinander vertauschen (dafür gibt es $(n-k)!$ Möglichkeiten), dann ändert sich nichts an unserer Auswahl.

Also gibt es $\frac{n!}{k!(n-k)!} = \binom{n}{k}$ k -elementige Teilmengen von $\{1, 2, \dots, n\}$.

Beispiel.

Beim Lotto müssen 6 Zahlen aus 49 ausgewählt werden. Nun ist

$$\binom{49}{6} = \frac{49!}{6!43!} = \frac{44 \cdot 45 \cdot 46 \cdot 47 \cdot 48 \cdot 49}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 13\,983\,816.$$

Wir kommen nun zu dem Satz, der den Binomialkoeffizienten ihren Namen gab:

Die Binomische Formel

$$\text{Für } a, b \in \mathbb{R} \text{ und } n \in \mathbb{N} \text{ ist } (a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}.$$

BEWEIS: Es ist $(a + b)^n = (a + b) \cdot (a + b) \cdot \dots \cdot (a + b)$, mit n Faktoren. Beim Ausmultiplizieren erhält man Terme der Gestalt

$$\dots + N(n, k) \cdot a^k \cdot b^{n-k} + \dots,$$

wobei $N(n, k)$ die Anzahl der Möglichkeiten bedeutet, aus k Klammern ein a und aus $n - k$ Klammern ein b zu holen. Aber dann ist $N(n, k)$ die Anzahl der k -elementigen Teilmengen der Menge aller n Klammern, also $N(n, k) = \binom{n}{k}$. ■

Für kleines n ist diese Formel wohlbekannt:

$$(a + b)^2 = a^2 + 2ab + b^2 \quad \text{und} \quad (a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

Zwei Folgerungen wollen wir noch notieren:

Folgerung 1. Es ist $\sum_{k=0}^n \binom{n}{k} = 2^n$.

Zum Beweis setze man einfach $a = b = 1$ in der binomischen Formel.

Ist M eine beliebige Menge, so nennt man die Menge $P(M)$ **aller** Teilmengen von M die *Potenzmenge* von M . Ist z.B. $M = \{1, 2, 3\}$, so ist

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

Bezeichnen wir allgemein mit $|M|$ die Anzahl der Elemente von M , so ist in diesem Fall $|M| = 3$ und $|P(M)| = 8 = 2^3$. Die Formel aus der Folgerung 1 zeigt: Ist $|M| = n$, so ist $|P(M)| = 2^n$. Das funktioniert übrigens auch in Extremfällen: Es ist $|\emptyset| = 0$ und $P(\emptyset) = \{\emptyset\}$, also $|P(\emptyset)| = 1 = 2^0$.

Folgerung 2. Ist $x \geq 0$, so ist $(1 + x)^n \geq 1 + nx$.

BEWEIS: Es ist $(1 + x)^n = 1 + nx + \frac{(n-1)n}{2} x^2 + \dots + x^n \geq 1 + nx$. ■

Diese Aussage ist ein Spezialfall der sogenannten **Bernoullischen Ungleichung**:

$$\text{Ist } x \geq -1, x \neq 0 \text{ und } n > 1, \text{ so ist } (x + 1)^n > 1 + nx.$$

Der Beweis wird mit Induktion geführt ($x > -1, n \geq 2$).

§ 4 Konvergenz und Vollständigkeit

Inhalt:

Dreiecksungleichung und ε -Umgebungen, Konvergenz von Folgen, Archimedes-Axiom und Vollständigkeitsaxiom, Umgang mit Intervallschachtelungen.

Beschränkte Mengen, Supremum und Infimum, monotone Konvergenz, Grenzwertsätze, Häufungspunkte, Satz von Bolzano-Weierstraß, Divergenz-Kriterium.

Sind a, b zwei reelle Zahlen, so ist $|a - b| = |b - a|$ der Abstand von a und b auf der Zahlengeraden. Speziell ist $|a|$ der Abstand der Zahl a vom Nullpunkt.

Eigenschaften der Betragsfunktion

1. $|a \cdot b| = |a| \cdot |b|$.
2. Es ist stets $-|a| \leq a \leq +|a|$.
3. Sei $c > 0$. Dann ist $|x| < c \iff -c < x < +c$.
4. Es gilt die **Dreiecks-Ungleichung**: $|a + b| \leq |a| + |b|$.
5. Es ist $|a - b| \geq |a| - |b|$.

Zum BEWEIS: (1) und (2) erhält man durch Fallunterscheidung.

3) Ist $|x| < c$, so ist $-|x| > -c$ und daher

$$-c < -|x| \leq x \leq |x| < c.$$

Ist umgekehrt $-c < x < +c$, so unterscheiden wir wieder 2 Fälle: Ist $x \geq 0$, so ist $|x| = x < c$. Ist $x < 0$, so ist ebenfalls $|x| = -x < -(-c) = c$ (wegen $x > -c$).

4) Wegen (2) ist

$$-(|a| + |b|) = -|a| - |b| \leq a + b \leq |a| + |b|.$$

Wegen (3) folgt daraus die Dreiecksungleichung.

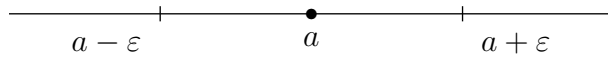
Zum Beweis von (5) benutzt man einen beliebigen Trick:

Es ist $|a| = |(a - b) + b| \leq |a - b| + |b|$. ■

Für beliebiges $a \in \mathbb{R}$ und $\varepsilon > 0$ nennt man die Menge

$$U_\varepsilon(a) := (a - \varepsilon, a + \varepsilon) = \{x \in \mathbb{R} \mid a - \varepsilon < x < a + \varepsilon\} = \{x \in \mathbb{R} : |x - a| < \varepsilon\}$$

die ε -Umgebung von a . Sie besteht aus allen Punkten x auf der Zahlengeraden, deren Abstand von a kleiner als ε ist.



Satz. Für zwei beliebige reelle Zahlen a, b gilt:

$$a = b \iff \forall \varepsilon > 0 \text{ gilt: } |a - b| < \varepsilon.$$

BEWEIS: Es reicht zu zeigen: $x = 0 \iff \forall \varepsilon : |x| < \varepsilon$.

Ist $x = 0$ und $\varepsilon > 0$, so ist selbstverständlich $|x| = 0 < \varepsilon$.

Ist umgekehrt das Kriterium erfüllt, so müssen wir zeigen, daß $x = 0$ ist. Das geht nur durch Widerspruch. Wir nehmen an, es sei $x \neq 0$. Dann ist $|x| > 0$. Setzen wir $\varepsilon := |x|/2$, so ist $0 < \varepsilon < |x|$. Das ist ein Widerspruch. ■

Bemerkung. Im Computer-Zeitalter benutzt man oft ein anderes Prinzip. Durch die Rechengenauigkeit eines Computers ist eine Zahl ε_0 bestimmt, so daß man sagt:

$$x = 0 \iff |x| < \varepsilon_0.$$

In der Mathematik reicht das nicht.

Ist für jedes $n \in \mathbb{N}$ eine reelle Zahl a_n gegeben, so sprechen wir von einer (unendlichen) *Zahlenfolge*. Der obige Satz motiviert die folgende Sprechweise: Die Folge (a_n) kommt der Null beliebig nahe, wenn $|a_n| < \varepsilon$ für jedes $\varepsilon > 0$ und alle $n \in \mathbb{N}$ (mit höchstens endlich vielen Ausnahmen) gilt. Wir nennen (a_n) dann eine *Nullfolge*.

Um das noch etwas präziser ausdrücken zu können, sagen wir: Eine Eigenschaft $E(n)$ gilt für *fast alle* $n \in \mathbb{N}$, falls es ein n_0 gibt, so daß $E(n)$ für alle $n \geq n_0$ (also bis auf endlich viele Ausnahmen) erfüllt ist.

Definition:

Die Folge (a_n) heißt eine *Nullfolge*, falls für jedes $\varepsilon > 0$ fast alle a_n in $U_\varepsilon(0)$ liegen. Man schreibt dann auch:

$$\lim_{n \rightarrow \infty} a_n = 0.$$

Will man tatsächlich zeigen, daß eine Folge gegen Null konvergiert, so gibt es ein Problem. Betrachten wir etwa die Folge $a_n := 1/n$. Würden alle Folgenglieder a_n

oberhalb einer Schranke $\varepsilon > 0$ liegen, so wäre stets $n \cdot \varepsilon \leq 1$. Das kommt einem spanisch vor, aber warum kann es nicht sein? Tatsächlich brauchen wir an dieser Stelle ein zusätzliches Axiom:

Axiom des Archimedes

Zu zwei reellen Zahlen $a, b > 0$ gibt es stets eine natürliche Zahl n mit $n \cdot a > b$.

Das bedeutet insbesondere (im Falle $a = 1$), daß die natürlichen Zahlen beliebig groß werden, und das ist ja durchaus plausibel.

Nun können wir zeigen, daß $1/n$ eine Nullfolge ist.

Satz. *Es ist $\lim_{n \rightarrow \infty} \frac{1}{n} = 0$.*

BEWEIS: Sei $\varepsilon > 0$ beliebig vorgegeben. Nach Archimedes gibt es ein $n_0 \in \mathbb{N}$ mit $n_0 \cdot \varepsilon > 1$. Für $n \geq n_0$ ist dann auch $n \cdot \varepsilon > 1$, also $1/n < \varepsilon$. Da dies für alle $\varepsilon > 0$ funktioniert, ist $1/n$ eine Nullfolge. ■

Beispiel.

Wir wollen zeigen:

Ist $|q| < 1$ und $a \in \mathbb{R}$ beliebig, so ist $\lim_{n \rightarrow \infty} a \cdot q^n = 0$.

Da wir sowieso $|a \cdot q^n|$ betrachten müssen, können wir annehmen, daß $0 < q < 1$ und $a > 0$ ist.

Sei $\varepsilon > 0$ beliebig vorgegeben. Um das passende n_0 zu finden, machen wir erst mal einen **Ansatz**: $a \cdot q^n < \varepsilon$. Daraus ergibt sich: $q^n < \varepsilon/a$. Dabei haben wir a auf die andere Seite der Ungleichung gebracht, weil wir n isolieren wollen. Der Übergang zum Kehrwert liefert $(1/q)^n > a/\varepsilon$.

Jetzt wissen wir nicht mehr weiter. Allerdings haben wir die Voraussetzung $0 < q < 1$ noch gar nicht richtig ausgenutzt. Sie hat zur Folge, daß $1/q > 1$ ist. Nun wenden wir einen Trick an. Wir setzen $1/q = 1 + h$, mit $h > 0$. Dann können wir nämlich die Bernoullische Ungleichung benutzen:

$$\left(\frac{1}{q}\right)^n = (1 + h)^n \geq 1 + nh > nh.$$

Manch einer hat inzwischen vielleicht den Faden verloren. Wir sind aber fast am Ziel, denn jetzt können wir alles systematisch von hinten aufrollen. Nach Archimedes gibt es ein $n_0 \in \mathbb{N}$, so daß $nh > a/\varepsilon$ für $n \geq n_0$ ist. Für solche n ist dann erst recht $(1/q)^n > a/\varepsilon$, also $q^n < \varepsilon/a$, also $a \cdot q^n < \varepsilon$. Und das war zu zeigen.

Wie sieht es nun aus, wenn eine Folge (a_n) einer beliebigen reellen Zahl a beliebig nahe kommt?

Definition:

Eine Folge (a_n) *konvergiert* gegen eine reelle Zahl a , falls in jeder ε -Umgebung von a fast alle a_n liegen. Man schreibt dann: $\lim_{n \rightarrow \infty} a_n = a$.

Es gilt:

$$a_n \in U_\varepsilon(a) \iff |a_n - a| < \varepsilon.$$

Und daß eine Aussage für fast alle n richtig ist, bedeutet:

$$\exists n_0 \in \mathbb{N} \text{ s.d. } \forall n \geq n_0 \text{ gilt: } \dots$$

Im Zusammenhang ergibt sich folgende logische Struktur:

$$\lim_{n \rightarrow \infty} a_n = a : \iff \forall \varepsilon > 0 \exists n_0 \in \mathbb{N} \text{ s.d. } \forall n \geq n_0 \text{ gilt: } |a_n - a| < \varepsilon.$$

Satz. *Es seien (a_n) , (b_n) und (c_n) drei Folgen.*

1. *Ist $a_n \leq b_n$, $\lim_{n \rightarrow \infty} a_n = a$ und $\lim_{n \rightarrow \infty} b_n = b$, so ist auch $a \leq b$.*
2. *Ist $a_n \leq c_n \leq b_n$ und $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n$, so konvergiert auch (c_n) gegen den gleichen Grenzwert.*

BEWEIS: 1) Wir nehmen an, es sei $a > b$, und versuchen, einen Widerspruch herbeizuführen.

Sei $\varepsilon := (a - b)/2$. Weil (a_n) gegen a und (b_n) gegen b konvergiert, gibt es ein n_0 , so daß für $n \geq n_0$ gilt:

$$|a_n - a| < \varepsilon \text{ und } |b_n - b| < \varepsilon.$$

Daraus folgt $a_n > a - \varepsilon$ und $b_n < b + \varepsilon$, also

$$a_n - b_n > (a - \varepsilon) - (b + \varepsilon) = a - b - 2\varepsilon = 0.$$

Demnach wäre $a_n > b_n$ für genügend großes n . Das ist ein Widerspruch.

2) Es seien a bzw. b die Grenzwerte von (a_n) bzw. (b_n) , und es sei ein $\varepsilon > 0$ vorgegeben. Dann ist $a = b$, und es gibt ein n_0 , so daß für $n \geq n_0$ gilt:

$$|a_n - a| < \varepsilon \quad \text{und} \quad |b_n - a| < \varepsilon.$$

Dann ist

$$a - \varepsilon < a_n \leq c_n \leq b_n < a + \varepsilon,$$

also auch $|c_n - a| < \varepsilon$. ■

Als Anwendung ergibt sich: Ist (a_n) eine Nullfolge und $|b_n| < |a_n|$, so ist auch (b_n) eine Nullfolge.

Bemerkung. Der Grenzwert einer Folge ist – wenn er überhaupt existiert – eindeutig bestimmt.

Wir spielen noch etwas mit dem Archimedes-Axiom.

Satz. Zu jeder reellen Zahl x gibt es genau ein $n \in \mathbb{Z}$ mit $n \leq x < n + 1$.

BEWEIS: Wir beschränken uns auf den Fall $x > 0$, denn für $x = 0$ ist die Aussage trivial und den negativen Fall kann man auf den positiven relativ leicht zurückführen.

Die Menge $M := \{k \in \mathbb{N} : k > x\}$ ist nach Archimedes nicht leer. Aber jede Teilmenge von \mathbb{N} besitzt ein kleinstes Element. Im Falle der Menge M sei k_0 dieses kleinste Element. Dann können wir $n := k_0 - 1$ setzen. ■

Definition:

Ist $x \in \mathbb{R}$ und $n \in \mathbb{N}$ die eindeutig bestimmte Zahl mit $n \leq x < n + 1$, so nennt man $[x] := n$ die *Gaußfunktion* von x .

Es ist z.B. $[\frac{18}{7}] = 2$ und $[-3.7] = -4$.

Wir setzen die Untersuchungen konvergenter Folgen fort.

Regeln für die Berechnung von Grenzwerten

Die Folgen (a_n) bzw. (b_n) seien konvergent gegen a bzw. b . Dann gilt:

$$\lim_{n \rightarrow \infty} a_n \pm b_n = a \pm b \quad \text{und} \quad \lim_{n \rightarrow \infty} a_n \cdot b_n = a \cdot b.$$

Sind b und fast alle $b_n \neq 0$, so ist $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{a}{b}$.

Den (relativ langweiligen) BEWEIS überlassen wir den Mathematikern. Wir werden diese Eigenschaften künftig als „Grenzwertsätze“ zitieren.

Beispiele.

1. Typisches Anwendungsbeispiel ist etwa eine Folge der Gestalt

$$a_n := \frac{n(n-2)}{5n^2+3}.$$

Kürzen durch n^2 ergibt $a_n = \frac{1-2/n}{5+3/n^2}$. Weil $1/n$ eine Nullfolge ist, konvergiert der Zähler gegen 1 und der Nenner gegen 5, also (a_n) gegen $1/5$.

2. Allzu sorgloses Umgehen mit den Grenzwertsätzen kann aber zu Fehlern führen. Eine gemeine Falle stellt etwa die Folge $a_n := \frac{1+2+\dots+n}{n^2}$ dar. Es ist $a_n = 1/n^2 + 2/n^2 + \dots + (n-1)/n^2 + 1/n$. Jeder Summand konvergiert gegen 0, also scheint auch (a_n) gegen 0 zu konvergieren. Das ist jedoch falsch, denn die Anzahl der Summanden erhöht sich bei jedem Schritt. Man kann deshalb keinen der Grenzwertsätze anwenden. In Wirklichkeit ist

$$a_n = \frac{1}{n^2} \cdot \frac{n(n+1)}{2} = \frac{1}{2} \cdot \left(1 + \frac{1}{n}\right) \quad (\text{Gaußsche Summenformel}).$$

Das bedeutet, daß (a_n) gegen $1/2$ konvergiert.

Die Folge

$$a_n := \left(1 + \frac{1}{n}\right)^n$$

konvergiert **nicht** gegen 1, wie man vermuten könnte (weil $1/n$ gegen Null, also $1 + 1/n$ gegen 1 konvergiert und 1^n stets $= 1$ ist). In Wirklichkeit kann man keinen der Grenzwertsätze anwenden. Konvergiert die Folge überhaupt, und wenn ja, wogegen? Um dieses Problem zu lösen, müssen wir etwas weiter ausholen.

Satz (Existenz von Intervallschachtelungen)

Ist $x \in \mathbb{R}$, so gibt es rationale Zahlen $p_n < q_n$, so daß gilt:

1. *x liegt in jedem der Intervalle $I_n := [p_n, q_n]$.*
2. *Es ist stets $I_{n+1} \subset I_n$, und $q_n - p_n$ konvergiert gegen Null.*
3. *x ist der einzige Punkt, der in allen Intervallen I_n gleichzeitig liegt.*

BEWEIS: Sei $p_0 := [x]$ und $q_0 := p_0 + 1$. Dann liegt $x \in [p_0, q_0]$. Es gibt zwei Möglichkeiten:

Ist $x \leq p_0 + \frac{1}{2}$, so setzen wir $p_1 := p_0$ und $q_1 := p_0 + \frac{1}{2}$. Andernfalls setzen wir $p_1 := p_0 + \frac{1}{2}$ und $q_1 := q_0$. Damit ist $x \in [p_1, q_1]$ und $q_1 - p_1 < \frac{1}{2}$.

So fährt man (per Induktion) fort. Bei jedem Schritt wird das Intervall wieder halbiert und dann die passende Hälfte ausgesucht. Das liefert die gewünschte Folge von Intervallen I_n . Offensichtlich ist stets $I_{n+1} \subset I_n$, und $q_n - p_n < 2^{-n}$ konvergiert gegen Null.

Wenn es zwei verschiedene Punkte x und y im Durchschnitt aller Intervalle I_n gäbe, dann wäre $d := |x - y| > 0$. Nach Archimedes gibt es ein $n \in \mathbb{N}$, so daß $2n > 1/d$ ist. Dann ist erst recht $1 + 2n > 1/d$, und weil $2^n > 1 + 2n$ für $n \geq 3$ ist, ist $d > 2^{-n}$, und x, y können nicht beide in I_n liegen. Widerspruch! ■

Allgemein wird eine Folge (I_n) von abgeschlossenen Intervallen als *Intervallschachtelung* bezeichnet, falls gilt:

1. $I_{n+1} \subset I_n$.
2. Die Folge der Längen der Intervalle bildet eine Nullfolge.

Wenn alle Intervallgrenzen rationale Zahlen sind, spricht man von einer *rationalen Intervallschachtelung*.

Wie oben kann man zeigen:

Ist (I_n) eine Intervallschachtelung, so kann es höchstens eine Zahl geben, die gleichzeitig in allen I_n liegt.

Wie wir gesehen haben, gibt es zu jeder reellen Zahl x eine rationale Intervallschachtelung (I_n) , so daß x in allen I_n enthalten ist. Man sagt dann, daß x die durch (I_n) bestimmte Zahl ist. Es bleibt die Frage, ob **jede** Intervallschachtelung eine Zahl bestimmt. Tatsächlich ist das nicht klar, und in \mathbb{Q} wäre es auch nicht wahr. Jetzt sind wir bei dem Punkt angekommen, wo sich \mathbb{Q} und \mathbb{R} voneinander unterscheiden. Es gibt noch ein weiteres Axiom, das dafür sorgt, daß die Menge \mathbb{R} der reellen Zahlen eindeutig festgelegt ist.

Vollständigkeitsaxiom

In \mathbb{R} bestimmt jede Intervallschachtelung (eindeutig) eine Zahl.

Beispiele.

1. Sei $a > 1$ eine reelle Zahl, $x_0 := 1$ und $y_0 := a$. Wir konstruieren rekursiv eine Intervallschachtelung $I_n = [x_n, y_n]$. Für $n \geq 1$ setzen wir

$$y_n := \frac{1}{2} \left(y_{n-1} + \frac{a}{y_{n-1}} \right)$$

und $x_n := \frac{a}{y_n}$.

Wir zeigen zunächst, daß stets $x_n^2 < a < y_n^2$ ist. Und zwar ist

$$\begin{aligned} y_n^2 - a &= \frac{1}{4} \left(y_{n-1}^2 + 2a + \frac{a^2}{y_{n-1}^2} - 4a \right) \\ &= \frac{1}{4} \left(y_{n-1} - \frac{a}{y_{n-1}} \right)^2 \geq 0, \end{aligned}$$

also $a \leq y_n^2$ und $x_n^2 = \frac{a^2}{y_n^2} \leq a$.

Wäre $a = y_n^2$ für ein n , so wäre auch $a = y_{n-1}^2$ u.s.w., also $a = y_0^2 = a^2$. Das kann nicht sein. Deshalb sind die Ungleichungen strikt, und dann ist auch $x_n < y_n$ für jedes n .

Als nächstes zeigen wir, daß $I_{n+1} \subset I_n$ ist. Es ist

$$\begin{aligned} y_n - y_{n+1} &= y_n - \frac{1}{2}\left(y_n + \frac{a}{y_n}\right) \\ &= \frac{y_n^2 - a}{2y_n} \geq 0 \end{aligned}$$

und dann natürlich auch $x_{n+1} \geq x_n$.

Schließlich brauchen wir noch, daß $y_n - x_n$ eine Nullfolge ist. Nach Konstruktion ist $y_n = \frac{1}{2}(y_{n-1} + x_{n-1})$, also

$$y_n - x_n \leq y_n - x_{n-1} \leq \frac{1}{2}(y_{n-1} - x_{n-1}).$$

Das bedeutet, daß schließlich $y_n - x_n \leq \frac{1}{2^n}(a-1)$ ist, und das ist eine Nullfolge.

Wir wissen, daß es genau eine reelle Zahl c mit $x_n \leq c \leq y_n$ für $n \in \mathbb{N}$ gibt. Es ist aber auch $x_n^2 < a < y_n^2$ für alle n . Das ist nur möglich, wenn $c^2 = a$ ist. Also $c = \sqrt{a}$ die durch die Intervallschachtelung bestimmte Zahl.

Schon in der Schule lernt man, daß $\sqrt{2}$ keine rationale Zahl ist. Das zeigt noch einmal, daß das Vollständigkeitsaxiom in \mathbb{Q} nicht gelten kann.

2. Wir können natürlich auch bisher unbekannte reelle Zahlen mit Hilfe von Intervallschachtelungen aufspüren. Als Beispiel betrachten wir die durch

$$x_n := \left(1 + \frac{1}{n}\right)^n \quad \text{und} \quad y_n := \left(1 + \frac{1}{n}\right)^{n+1}$$

gegebenen Intervalle $I_n = [x_n, y_n]$.

Offensichtlich ist stets $x_n < y_n$, und aus den Ungleichungen

$$(a + \varepsilon)^k - a^k = \varepsilon \cdot \sum_{i=0}^{k-1} (a + \varepsilon)^i a^{k-i-1} < \varepsilon \cdot k \cdot (a + \varepsilon)^{k-1}$$

$$\text{und} \quad (a + \varepsilon)^k - a^k > \varepsilon \cdot k \cdot a^{k-1}$$

folgt:

$$\begin{aligned}
\left(1 + \frac{1}{n}\right) \cdot x_n &= \left(1 + \frac{1}{n}\right)^{n+1} \\
&< \left(1 + \frac{1}{n+1}\right)^{n+1} + \left(\frac{1}{n} - \frac{1}{n+1}\right) \cdot (n+1) \cdot \left(1 + \frac{1}{n}\right)^n \\
&= x_{n+1} + \frac{1}{n} \cdot x_n,
\end{aligned}$$

und

$$\begin{aligned}
\left(1 + \frac{1}{n}\right) \cdot y_n &= \left(1 + \frac{1}{n}\right)^{n+2} \\
&> \left(1 + \frac{1}{n+1}\right)^{n+2} + \left(\frac{1}{n} - \frac{1}{n+1}\right) \cdot (n+2) \cdot \left(1 + \frac{1}{n+1}\right)^{n+1} \\
&= y_{n+1} + \frac{n+2}{n(n+1)} \cdot y_{n+1} \cdot \left(\frac{n+2}{n+1}\right)^{-1} \\
&= \left(1 + \frac{1}{n}\right) \cdot y_{n+1}.
\end{aligned}$$

also $x_n < x_{n+1}$ und $y_{n+1} < y_n$.

Wegen

$$\left(1 + \frac{1}{2n}\right)^n < 1 + n \cdot \frac{1}{2n} \cdot \left(1 + \frac{1}{2n}\right)^{n-1} = 1 + \frac{1}{2} \left(1 + \frac{1}{2n}\right)^n.$$

ist $\left(1 + \frac{1}{2n}\right)^n < 2$, also $x_{2n} < 4$. Weil stets $x_n < x_{n+1}$ ist, muß dann sogar $x_n < 4$ für alle n gelten. Daher ist

$$y_n - x_n = \left(1 + \frac{1}{n}\right)^n \cdot \left(\left(1 + \frac{1}{n}\right) - 1\right) = x_n \cdot \frac{1}{n} \leq \frac{4}{n}.$$

Es liegt tatsächlich eine Intervallschachtelung vor.

Die dadurch eindeutig bestimmte reelle Zahl e nennt man die *Eulersche Zahl*.

Wir rechnen die ersten Intervalle aus und bekommen:

$$\begin{array}{rclclcl}
x_1 & = & 2 & \leq e \leq & 4 & = & y_1, \\
x_2 & = & 2.25 & \leq e \leq & 3.375 & = & y_2, \\
x_3 & \approx & 2.37 & \leq e \leq & 3.16 & \approx & y_3, \\
x_4 & \approx & 2.44 & \leq e \leq & 3.05 & \approx & y_4, \\
x_5 & \approx & 2.488 & \leq e \leq & 2.986 & \approx & y_5, \\
x_6 & \approx & 2.52 & \leq e \leq & 2.94 & \approx & y_6, \\
x_7 & \approx & 2.546 & \leq e \leq & 2.91 & \approx & y_7, \\
& & & & \vdots & & \\
x_{100} & \approx & 2.705 & \leq e \leq & 2.732 & \approx & y_{100}, \\
& & & & \vdots & & \\
x_{1000} & \approx & 2.717 & \leq e \leq & 2.72 & \approx & y_{1000}.
\end{array}$$

Tatsächlich ist $e = 2.718281828459\dots$, und man sieht, daß hier das Intervallschachtelungsverfahren sehr schlecht konvergiert.

Das muß nicht immer so sein. Die Intervallschachtelung für $\sqrt{2}$ aus dem vorigen Beispiel ergibt z.B.:

$$\begin{aligned} x_1 &\approx 1.333 \leq \sqrt{2} \leq 1.5 &&= y_1, \\ x_2 &\approx 1.41176 \leq \sqrt{2} \leq 1.416667 &&\approx y_2, \\ x_3 &\approx 1.414212 \leq \sqrt{2} \leq 1.414216 &&\approx y_3, \\ x_4 &\approx 1.414213 \leq \sqrt{2} \leq 1.414214 &&\approx y_4, \\ x_5 &\approx 1.4142136 \leq \sqrt{2} \leq 1.4142136 &&\approx y_5. \end{aligned}$$

Daß das Verfahren hier so schnell konvergiert, liegt an der Abschätzung

$$y_n - x_n \leq 2^{-n}.$$

Definition:

Eine Menge $M \subset \mathbb{R}$ heißt *nach oben (bzw. nach unten) beschränkt*, falls es eine reelle Zahl c gibt, so daß $x \leq c$ (bzw. $x \geq c$) für $x \in M$ ist. Die Zahl c nennt man dann eine *obere (bzw. untere) Schranke* für M .

M heißt *beschränkt*, wenn M nach oben und nach unten beschränkt ist.

Beispiel.

Jedes abgeschlossene Intervall ist beschränkt, aber auch jedes offene Intervall. Die Menge \mathbb{N} der natürlichen Zahlen ist nach unten durch 1, aber nicht nach oben beschränkt (wegen Archimedes).

Definition:

Sei $M \subset \mathbb{R}$ eine nach oben beschränkte Menge. Wenn die Menge aller oberen Schranken von M ein kleinstes Element a besitzt, so nennt man diese kleinste obere Schranke das *Supremum* von M (in Zeichen: $a = \sup(M)$).

Ist M nach unten beschränkt, so nennt man die größte untere Schranke das *Infimum* von M (in Zeichen: $\inf(M)$).

Satz. Sei $M \subset \mathbb{R}$ eine nicht-leere Menge.

1. Ist M nach oben beschränkt, so besitzt M ein Supremum.
2. Ist M nach unten beschränkt, so besitzt M ein Infimum.

BEWEIS: Wir betrachten nur den Fall des Supremums.

Sei b_1 eine obere Schranke von M . Weiter sei $a_1 < b_1$ eine Zahl, die keine obere Schranke von M ist (z.B. $a_1 = x - 1$ für ein Element $x \in M$).

Ist $I_n = [a_n, b_n]$ konstruiert, so setzen wir $d_n := \frac{1}{2}(a_n + b_n)$ (Mittelpunkt des Intervalls). Nun gibt es zwei Möglichkeiten:

1. Fall: d_n ist obere Schranke von M . Dann sei $a_{n+1} := a_n$ und $b_{n+1} := d_n$.
2. Fall: d_n ist keine obere Schranke von M . Dann sei $a_{n+1} := d_n$ und $b_{n+1} := b_n$.

Induktiv erhalten wir so eine Intervallschachtelung (I_n) , die eindeutig eine Zahl s bestimmt. Dabei ist die obere Grenze b_n von I_n jedesmal eine obere Schranke von M , und die untere Grenze a_n ist niemals eine obere Schranke von M . Wir wollen zeigen, daß $s = \sup(M)$ ist.

s ist obere Schranke, denn sonst gäbe es ein $x \in M$ mit $x > s$. Weil stets $a_n \leq s \leq b_n$ ist, wäre dann $a_n < x$ und $b_n - a_n \geq x - s > 0$. Das ist ein Widerspruch, denn $b_n - a_n$ konvergiert gegen Null.

s ist aber auch die kleinste obere Schranke, denn sonst gäbe es eine obere Schranke c von M mit $c < s$. Dann wäre stets $b_n - a_n \geq s - c > 0$. Da das nicht sein kann, ist s das Supremum. ■

Beispiel.

Es ist $\sup(a, b) = \sup[a, b] = b$. Das Supremum einer Menge kann zu der Menge gehören, muß es aber nicht.

Eine wichtige Anwendung ist der Satz von der monotonen Konvergenz.

Definition:

Eine Folge (a_n) heißt *monoton wachsend* (bzw. *monoton fallend*), falls stets $a_n \leq a_{n+1}$ (bzw. $a_n \geq a_{n+1}$) ist.

Ist $I_n = [a_n, b_n]$ eine Intervallschachtelung, so ist (a_n) monoton wachsend und (b_n) monoton fallend.

Wir nennen eine Folge (a_n) *nach oben beschränkt* (bzw. *nach unten beschränkt* oder überhaupt *beschränkt*), wenn das auf die Menge $\{a_n : n \in \mathbb{N}\}$ zutrifft.

Satz von der monotonen Konvergenz

Ist (a_n) monoton wachsend und nach oben beschränkt, so ist (a_n) konvergent.

BEWEIS: Da die Menge $M := \{a_n : n \in \mathbb{N}\}$ nicht leer und nach oben beschränkt ist, besitzt sie ein Supremum $a \in \mathbb{R}$. Dieses a ist unser Kandidat.

Da a eine obere Schranke von M ist, ist stets $a - a_n \geq 0$. Ist $\varepsilon > 0$, so gibt es ein $n_0 \in \mathbb{N}$, so daß $(a - \varepsilon) < a_{n_0}$ ist, sonst wäre a nicht die kleinste obere Schranke. Wegen der Monotonie der Folge gilt dann für $n \geq n_0$:

$$a - \varepsilon - a_n \leq a - \varepsilon - a_{n_0} < 0, \quad \text{also } a - a_n < \varepsilon.$$

Das bedeutet, daß (a_n) gegen a konvergiert. ■

Beispiel.

Die Folge (a_n) sei rekursiv definiert durch

$$a_1 := \sqrt{2} \quad \text{und} \quad a_{n+1} := \sqrt{2 + a_n}.$$

Dann ist $a_2 = \sqrt{2 + \sqrt{2}}$, $a_3 = \sqrt{2 + \sqrt{2 + \sqrt{2}}}$ u.s.w.

Wir zeigen, daß diese Folge monoton wachsend und nach oben beschränkt ist. Als Hilfsmittel benutzen wir das Induktionsprinzip.

a) Offensichtlich ist $a_1 < 2$. Ist allgemein $a_n < 2$, so ist auch $a_{n+1} = \sqrt{2 + a_n} < \sqrt{2 + 2} = 2$.

b) Da $2 < 2 + \sqrt{2}$ ist, ist $a_1 < a_2$. Und wenn man schon weiß, daß $a_n < a_{n+1}$ ist, dann ist auch $2 + a_n < 2 + a_{n+1}$, also

$$a_{n+1} = \sqrt{2 + a_n} < \sqrt{2 + a_{n+1}} = a_{n+2}.$$

Damit ist alles gezeigt, und nach dem Satz von der monotonen Konvergenz strebt die Folge (a_n) gegen einen Grenzwert a . Der Satz sagt nichts darüber aus, wie man den Grenzwert findet. Dafür gibt es einen netten kleinen Trick. Wenn (a_n) gegen a konvergiert, dann konvergiert auch $b_n := a_{n+1}$ gegen a . Es ist aber $b_n^2 = 2 + a_n$. Die rechte Seite strebt offensichtlich gegen $2 + a$, die linke Seite gegen a^2 . Wegen der Eindeutigkeit des Grenzwertes muß $a^2 = 2 + a$ sein. Diese quadratische Gleichung hat nur eine positive Lösung, nämlich $a = 2$. Das ist der gesuchte Grenzwert.

Betrachten wir jetzt die Folge $a_n := (-1)^n$. Da $(-1) \cdot (-1) = 1$ ist, folgt:

$$a_n = \begin{cases} 1 & \text{falls } n \text{ gerade,} \\ -1 & \text{sonst.} \end{cases}$$

Diese Folge kann nicht konvergieren. Es ist bloß schwierig, das auf mathematisch saubere Weise auszudrücken. Deshalb suchen wir uns jetzt ein paar Kriterien, die einwandfrei belegen, daß eine Folge **nicht** konvergiert.

Satz. *Jede konvergente Folge ist beschränkt.*

BEWEIS: Sei (a_n) eine konvergente Folge mit Grenzwert a . Es gibt ein n_0 , so daß für alle $n \geq n_0$ alle a_n in $U_1(a)$ liegen. Ist nun $R := \max\{|a - a_n| : n = 1, 2, \dots, n_0\}$, so liegen alle a_n in $U_{R+1}(a)$. Das bedeutet, daß die Folge beschränkt ist. ■

Im Umkehrschluß kann eine unbeschränkte Folge niemals konvergent sein. Leider hilft uns das bei unserem Beispiel nicht weiter. Die Folge $a_n = (-1)^n$ ist beschränkt.

Definition:

Eine Zahl $a \in \mathbb{R}$ heißt *Häufungspunkt* der Folge (a_n) , falls in jeder ε -Umgebung von a unendlich viele a_n liegen.

Jede konvergente Folge hat einen Häufungspunkt, nämlich ihren Grenzwert. Die Folge $a_n = (-1)^n$ besitzt **zwei** Häufungspunkte, nämlich $+1$ und -1 . Das scheint der Knackpunkt zu sein! Zunächst stellen wir folgende erstaunliche Tatsache fest:

Satz von Bolzano-Weierstraß

Jede beschränkte Folge besitzt mindestens einen Häufungspunkt.

BEWEIS: Sei (a_n) eine beschränkte Folge. Dann ist $c_n := \inf\{a_n, a_{n+1}, \dots\}$ eine monoton wachsende Folge, die wieder beschränkt ist. Nach dem Satz von der monotonen Konvergenz strebt (c_n) gegen eine reelle Zahl c . Die c_n kommen c beliebig nahe, und die Zahlen a_k mit $k \geq n$ kommen dem c_n beliebig nahe. Daraus folgt, daß c ein Häufungspunkt von (a_n) ist. ■

Satz. *Eine konvergente Folge hat nur einen Häufungspunkt.*

BEWEIS: Sei (a_n) eine konvergente Folge mit Grenzwert a . Dann ist a auch ein Häufungspunkt. Annahme, es gibt einen weiteren Häufungspunkt $b \neq a$. Dann wählen wir ein ε mit $0 < \varepsilon < \frac{1}{2}|b - a|$. Dann ist $U_\varepsilon(a) \cap U_\varepsilon(b) = \emptyset$, und fast alle a_n

liegen in $U_\varepsilon(a)$. Da dann für $U_\varepsilon(b)$ nur noch höchstens endlich viele a_n übrigbleiben, ist das ein Widerspruch. ■

Zusammengefaßt erhalten wir:

Divergenz-Kriterium

Eine Folge (a_n) ist genau dann divergent (d.h. nicht konvergent), wenn eine der beiden folgenden Bedingungen erfüllt ist:

1. (a_n) ist unbeschränkt.
2. (a_n) ist beschränkt und hat mindestens zwei verschiedene Häufungspunkte.

BEWEIS: Wenn das Kriterium erfüllt ist, kann (a_n) nicht konvergieren. Umgekehrt, wenn (a_n) divergent und beschränkt ist, so muß sie mindestens zwei Häufungspunkte besitzen. ■

Ist (a_n) eine Folge reeller Zahlen und (n_ν) eine *streng monoton wachsende* Folge natürlicher Zahlen (d.h. $n_{\nu+1} > n_\nu$), so kann man die Folge (a_{n_ν}) bilden. Diese reelle Zahlenfolge nennt man eine *Teilfolge* der ursprünglichen Folge (a_n) . Ist (a_n) konvergent, so konvergiert jede Teilfolge gegen den gleichen Grenzwert.

Satz. *Ist a ein Häufungspunkt der Folge (a_n) , so gibt es eine Teilfolge, die gegen a konvergiert.*

BEWEIS: Für ε wählen wir sukzessive die Zahlen $1, 1/2, 1/3, \dots$. Weil a ein Häufungspunkt ist, finden wir in jeder Umgebung $U_{1/k}(a)$ ein Folgenglied a_{n_k} . Da wir jedesmal sogar unendlich viele zur Auswahl haben, können wir erreichen, daß (n_k) streng monoton wächst. So bekommen wir eine Teilfolge, die offensichtlich gegen a konvergiert. ■

§ 5 Elementare Funktionen

Inhalt:

Affin-lineare und quadratische Funktionen, Polynome, Nullstellen und Linearfaktoren, Division mit Rest für Polynome, Vielfachheiten von Nullstellen, rationale Funktionen, Polstellen und hebbare Unbestimmtheitsstellen, Partialbruchzerlegung.

Grenzwerte von Funktionen, Stetigkeit, Sprungstellen, Monotonie.

Exponentialfunktion, Logarithmus, Bogenmaß, Winkelfunktionen, Additionstheoreme, Arcus-Funktionen, Hyperbelfunktionen.

Dieser Abschnitt behandelt zu einem erheblichen Teil Schulstoff. Soweit dies der Fall ist, dient er nur der Wiederholung, und auf Beweise wird dann verzichtet, auch wenn davon auszugehen ist, daß die Ergebnisse auch an der Schule ohne Beweis präsentiert worden sind.

5.1 Affin-lineare Funktionen

$f : \mathbb{R} \rightarrow \mathbb{R}$ sei definiert durch

$$f(x) := mx + c, \text{ mit } m \neq 0.$$

Dann nennt man f eine *affin-lineare Funktion*, und der Graph ist die Gerade $\{(x, y) \in \mathbb{R}^2 \mid mx - y = -c\}$.

Der Graph geht genau dann durch den Nullpunkt, wenn $c = 0$ ist. Dann spricht man von einer *linearen Funktion* oder einer *proportionalen Zuordnung* (mit dem *Proportionalitätsfaktor* m).

Definition:

Sei $M \subset \mathbb{R}$. Eine Funktion $f : M \rightarrow \mathbb{R}$ heißt *monoton wachsend* (bzw. *monoton fallend*), wenn für $x_1, x_2 \in M$ mit $x_1 < x_2$ stets $f(x_1) \leq f(x_2)$ ist. Gilt sogar stets die strikte Ungleichung, so nennt man f *streng monoton wachsend* (bzw. *fallend*).

Die affin-lineare Funktion $f(x) = mx + c$ ist streng monoton wachsend (bzw. fallend), wenn $m > 0$ (bzw. < 0) ist. Ist $m = 0$, also $f(x) = c$ eine *konstante Funktion*, so ist f gleichzeitig monoton wachsend und fallend (im schwachen Sinne).

Eine streng monotone Funktion $f : M \rightarrow \mathbb{R}$ ist automatisch injektiv. Ist $f(M) = N$, so ist dann $f : M \rightarrow N$ bijektiv und besitzt eine Umkehrabbildung

Satz. Die Umkehrfunktion einer streng monoton wachsenden Funktion ist ebenfalls streng monoton wachsend.

BEWEIS: Seien $y_1, y_2 \in f(M)$ mit $y_1 < y_2$. Es gibt Elemente $x_1, x_2 \in M$ mit $f(x_i) = y_i$ für $i = 1, 2$. Sicher ist dann $x_1 \neq x_2$, denn sonst wäre $f(x_1) = f(x_2)$. Wäre aber $x_1 > x_2$, so müßte wegen der strengen Monotonie von f auch $y_1 > y_2$ sein. Das ist nicht der Fall, also ist $x_1 < x_2$ und damit f^{-1} streng monoton wachsend. ■

5.2 Quadratische Funktionen

$f : \mathbb{R} \rightarrow \mathbb{R}$ sei definiert durch

$$f(x) := ax^2 + bx + c, \text{ mit } a \neq 0.$$

Das ist eine *quadratische Funktion*, und ihr Graph ist eine *Parabel*. Die Gestalt dieser Parabel kann man folgendermaßen bestimmen.

1. Die Lösungen der quadratischen Gleichung $f(x) = 0$ sind die „Nullstellen“ von f , also die Schnittpunkte von G_f mit der x-Achse. Das Prinzip der quadratischen Ergänzung liefert:

$$\begin{aligned} ax^2 + bx + c = 0 &\iff a \cdot \left(x^2 + \frac{b}{a}x\right) = -c \\ &\iff a \cdot \left(\left(x + \frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2\right) = -c \\ &\iff \left(x + \frac{b}{2a}\right)^2 - \left(\frac{b}{2a}\right)^2 = -\frac{c}{a} \\ &\iff \left(x + \frac{b}{2a}\right)^2 = \frac{1}{4a^2}(b^2 - 4ac). \end{aligned}$$

Die Größe $\Delta := b^2 - 4ac$ nennt man die *Diskriminante*. Ist $\Delta \geq 0$, so erhält man die Lösungen

$$x_{1,2} = \frac{1}{2a} \cdot (-b \pm \sqrt{b^2 - 4ac}).$$

Ist $\Delta = 0$, so fallen die beiden Lösungen zusammen. Ist $\Delta < 0$, so gibt es keine (reelle) Lösung.

2. Aus der Darstellung

$$f(x) = a \cdot \left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a^2} \right] = a((x - \alpha)^2 - \beta)$$

mit $\alpha := -\frac{b}{2a}$ und $\beta := \frac{\Delta}{4a^2}$ folgt sofort:

$$f(\alpha - x) = f(\alpha + x),$$

d.h., G_f liegt symmetrisch zur vertikalen Geraden $\{(x, y) \mid x = \alpha\}$.

3. Ist $a > 0$, so ist $f(x) \geq -a \cdot \beta$, und die Parabel ist nach oben geöffnet. Ist $a < 0$, so ist sie nach unten geöffnet.

Ist $a > 0$, so ist f rechts von α streng monoton wachsend und links von α streng monoton fallend. Ist $a < 0$, so verhält es sich umgekehrt.

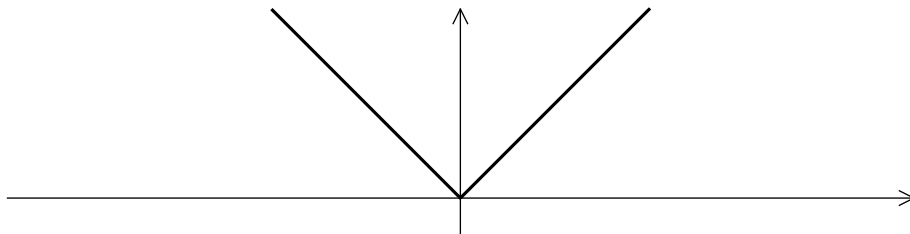
Definition:

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt *gerade* (bzw. *ungerade*), wenn $f(-x) = f(x)$ (bzw. $f(-x) = -f(x)$) für alle x ist.

Die quadratische Funktion $f(x) = ax^2$ ist ein typisches Beispiel für eine gerade Funktion.

5.3 Die Betragsfunktion

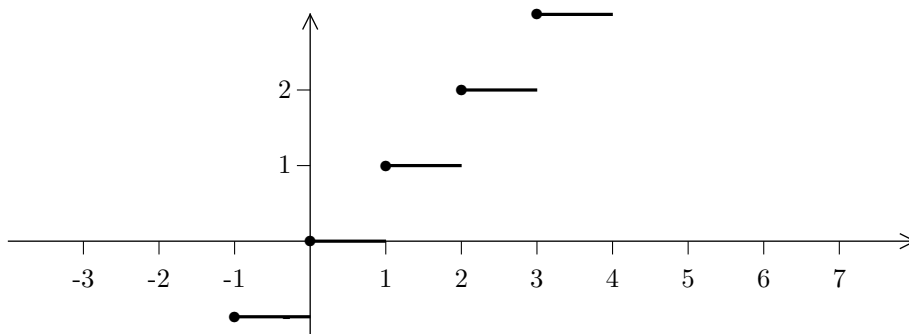
Die Funktion $f(x) := |x|$ hat folgenden Graphen:



Die Betragsfunktion ist gerade. Rechts von 0 ist sie streng monoton wachsend, links davon streng monoton fallend.

5.4 Die Gauß-Klammer

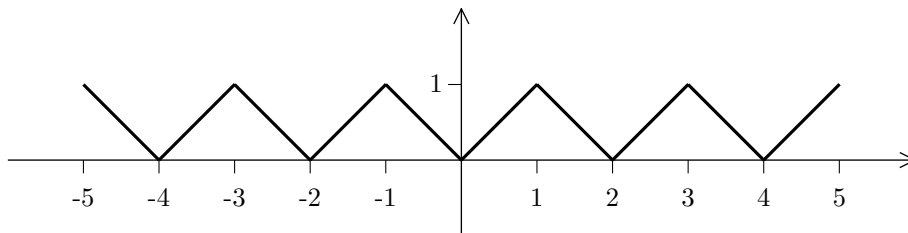
Die Funktion $f(x) := [x]$ hat folgende Gestalt:



Die Gaußfunktion ist monoton wachsend. Sie ist weder gerade noch ungerade.

5.5 Die Zickzack-Funktion:

Wir wollen die Definition der Funktion finden, deren Graph die folgende Zackenkurve ist:



Der Graph von f besteht aus Geradenstücken. Ist $2n \leq x < 2n + 1$, so ist $f(x) = x - 2n$. Ist $2n + 1 \leq x < 2n + 2$, so ist $f(x) = -x + (2n + 2)$. Zusammen ergibt das:

$$f(x) = \begin{cases} x - 2n & \text{für } 2n \leq x < 2n + 1 \\ -x + (2n + 2) & \text{für } 2n + 1 \leq x < 2n + 2. \end{cases}$$

Dies ist ein typisches Beispiel einer „zusammengesetzten Funktion“. Sie ist gerade und abschnittsweise streng monoton wachsend bzw. fallend. Natürlich ist sie nicht global injektiv, wohl aber auf jedem Intervall, auf dem sie monoton ist.

Definition:

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt *periodisch* (mit Periode T), falls $f(x + T) = f(x)$ für alle $x \in \mathbb{R}$ ist.

Ist f nicht konstant, so besitzt f eine kleinste positive Periode. Wenn nichts anderes gesagt wird, verstehen wir immer unter dieser Zahl die Periode von f .

Die Zickzack-Funktion ist periodisch mit Periode 2.

5.6 Polynome

Eine Funktion der Gestalt

$$p(x) := \sum_{i=0}^n a_i x^i = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

mit *Koeffizienten* $a_i \in \mathbb{R}$ heißt *reelles Polynom*. Die Nummer n des höchsten nicht verschwindenden Koeffizienten nennt man den *Grad des Polynoms*.

So ist etwa $\text{grad}(x - x^4) = 4$.

Das „Nullpolynom“ $f(x) \equiv 0$ hat eigentlich keinen Grad, wird aber manchmal definitionsgemäß mit dem Grad $-\infty$ versehen. Eine konstante Funktion $f(x) \equiv c$ mit $c \neq 0$ ist ein Polynom vom Grad 0.

Sind $f, g : M \rightarrow \mathbb{R}$ zwei Funktionen mit gleichem Definitionsbereich, so kann man ihre Summe und ihr Produkt bilden. $(f + g) : M \rightarrow \mathbb{R}$ und $(f \cdot g) : M \rightarrow \mathbb{R}$ werden definiert durch

$$(f + g)(x) := f(x) + g(x)$$

und

$$(f \cdot g)(x) := f(x) \cdot g(x).$$

Es gelten die üblichen Rechengesetze, allerdings kann man i.a. nicht dividieren. Ist $0 \in M$, so ist die Funktion $1/x$ auf M nicht definiert, man kann also auf M z.B. nicht durch x dividieren.

Wenn Polynome addiert und multipliziert werden, so ist das Ergebnis jeweils wieder ein Polynom:

$$\text{Sei } f(x) := \sum_{i=0}^n a_i x^i \quad \text{und} \quad g(x) := \sum_{j=0}^m b_j x^j.$$

Ist $n = m$, so ist

$$(f + g)(x) = \sum_{i=0}^n (a_i + b_i) x^i.$$

Ist $n \neq m$, etwa $n > m$, so ergänzt man in g noch die Terme $b_{m+1}x^{m+1}, \dots, b_n x^n$, mit $b_{m+1} = \dots = b_n = 0$, und verfährt dann wie oben.

Das Produkt $f \cdot g$ ist gegeben durch

$$(f \cdot g)(x) := \sum_{k=0}^{n+m} \left(\sum_{i+j=k} a_i b_j \right) x^k$$

$$= a_0 b_0 + (a_1 b_0 + a_0 b_1) x + (a_2 b_0 + a_1 b_1 + a_0 b_2) x^2 + \dots + a_n b_m x^{n+m}.$$

Hieraus ergeben sich die Grad-Formeln:

1. $\text{grad}(f \cdot g) = \text{grad}(f) + \text{grad}(g)$.
2. $\text{grad}(f + g) \leq \max(\text{grad}(f), \text{grad}(g))$.

Soll ein Polynom $f(x) = a_n x^n + \dots + a_1 x + a_0$ ausgewertet werden, so bedient man sich meist des sogenannten *Horner-Schemas*:

$$f(x) = a_0 + x(a_1 + x(a_2 + \dots + x(a_{n-1} + x a_n) \dots)).$$

Auf diese Weise kann man die Anzahl der Multiplikationen halbieren.

Ist f nicht die Nullfunktion, so interessiert man sich für die Nullstellen von f , also die Lösungen der Gleichung $f(x) = 0$.

Abspaltung von Linearfaktoren

Sei $p(x)$ ein Polynom vom Grad n und c eine Nullstelle von p . Dann gibt es ein Polynom $q(x)$ vom Grad $n - 1$, so daß gilt:

$$p(x) = (x - c) \cdot q(x).$$

Man kann also einen „Linearfaktor“ $(x - c)$ abspalten.

BEWEIS: Es ist $p(c) = 0$, also

$$\begin{aligned} p(x) = p(x) - p(c) &= (a_n x^n + \dots + a_0) - (a_n c^n + \dots + a_0) \\ &= a_n(x^n - c^n) + a_{n-1}(x^{n-1} - c^{n-1}) + \dots + a_1(x - c), \end{aligned}$$

und dieses Polynom ist durch $(x - c)$ teilbar, nach der Formel

$$x^k - c^k = (x - c) \cdot \sum_{i=0}^{k-1} x^i c^{k-i-1}.$$

■

Hat man eine Nullstelle gefunden (das ist i.a. ein schwieriges und oft sogar ein unlösbares Problem), so kann man $q(x)$ durch Polynomdivision ermitteln.

Beispiel.

Sei $p(x) := x^5 - 2x^4 - x + 2$. Durch Probieren stellt man schnell fest, daß p bei $x = 1$ eine Nullstelle besitzt. Durch Polynomdivision erhält man:

$$p(x) = (x - 1) \cdot (x^4 - x^3 - x^2 - x - 2).$$

Im Detail geht das so:

1. $(x^5 - 2x^4 - x + 2) - x^4(x - 1) = -x^4 - x + 2,$
2. $(-x^4 - x + 2) - (-x^3)(x - 1) = -x^3 - x + 2,$
3. $(-x^3 - x + 2) - (-x^2)(x - 1) = -x^2 - x + 2,$
4. $(-x^2 - x + 2) - (-x)(x - 1) = -2x + 2,$
5. $(-2x + 2) - (-2)(x - 1) = 0.$

Zur Probe kann man ausmultiplizieren.

Wenn man nicht durch eine Nullstelle dividiert, dann braucht die Division natürlich nicht aufzugehen. Es gilt aber:

mit $k_1 + \dots + k_m \leq n$ und einem Polynom q ohne Nullstellen.

Zum BEWEIS dividiert man sukzessive alle Nullstellen heraus, bis nur noch ein Polynom ohne Nullstellen übrigbleibt. Dessen Grad ist $= n - (k_1 + \dots + k_m) \geq 0$.

Beispiel.

Sei $p(x) = x^7 - 2x^6 + 3x^5 - 4x^4 + 2x^3$.

Offensichtlich ist $p(x) = x^3 \cdot q_1(x)$, mit $q_1(x) := x^4 - 2x^3 + 3x^2 - 4x + 2$.

Da $q_1(0) \neq 0$ ist, hat p in $x = 0$ eine Nullstelle der Vielfachheit 3.

Probieren zeigt, daß $q_1(1) = 0$ ist. Polynomdivision ergibt:

$$q_1(x) = (x - 1) \cdot q_2(x), \text{ mit } q_2(x) := x^3 - x^2 + 2x - 2.$$

Da auch $q_2(1) = 0$ ist, kann man noch einmal durch $(x - 1)$ dividieren und erhält:

$$q_1(x) = (x - 1)^2 \cdot q_3(x), \text{ mit } q_3(x) := x^2 + 2.$$

$q_3(x)$ besitzt keine Nullstelle mehr. Also ist

$$p(x) = x^3 \cdot (x - 1)^2 \cdot (x^2 + 2)$$

die bestmögliche Zerlegung von $p(x)$.

Im allgemeinen ist es nicht möglich, Nullstellen von Polynomen höheren Grades zu bestimmen (oder jedenfalls nur näherungsweise, mit numerischen Methoden). In einem Spezialfall gibt es aber ein nettes Hilfsmittel.

Satz (Nullstellen von Polynomen mit ganzen Koeffizienten)

Sei $p(x) = a_n x^n + \dots + a_0$ ein Polynom vom Grad n mit ganzzahligen Koeffizienten a_i . Sind $a, b \in \mathbb{N}$ teilerfremd und ist $\alpha := a/b$ eine Nullstelle von $p(x)$, so muß gelten:

1. a ist Teiler von a_0 .
2. b ist Teiler von a_n .

BEWEIS: Ist $p(\alpha) = 0$, so ist also

$$a_n \left(\frac{a}{b}\right)^n + \dots + a_1 \frac{a}{b} + a_0 = 0.$$

Das ist genau dann der Fall, wenn gilt:

$$a_0b^n = a(-a_1b^{n-1} - a_2ab^{n-2} - \dots - a_n a^{n-1})$$

und

$$a_n a^n = b(-a_0b^{n-1} - a_1ab^{n-2} - \dots - a_{n-1}a^{n-1}).$$

Dann folgt: a teilt a_0b^n und daher a_0 (weil a und b teilerfremd sind). Und b teilt $a_n a^n$, also auch a_n . ■

Folgerung. Ist $p(x) = a_n x^n + \dots + a_0$ ein Polynom mit ganzzahligen Koeffizienten und α eine ganzzahlige Nullstelle von $p(x)$, so teilt α den Koeffizienten a_0 .

Beispiel.

Sei $p(x) = x^3 - 6x^2 - 9x + 14$. Hier besitzt $a_0 = 14$ nur die Teiler 1, 2 und 7. Tatsächlich sind $x_1 = 1$, $x_2 = -2$ und $x_3 = 7$ Nullstellen.

Im Falle des Polynoms $q(x) = x^4 + 4x^3 - 6x^2 + 12x - 2$ ist zwar 2 ein Teiler von $a_0 = 2$, aber trotzdem ist weder $x = 2$ noch $x = -2$ eine Nullstelle. Das Kriterium hilft also nur manchmal weiter.

5.7 Rationale Funktionen:

Definition:

Sind p, q zwei Polynome, von denen q nicht das Nullpolynom ist, so nennt man

$$f(x) := \frac{p(x)}{q(x)}$$

eine *rationale Funktion*.

Der Definitionsbereich D_f einer solchen rationalen Funktion f ist die Menge $D_f := \{x \in \mathbb{R} \mid q(x) \neq 0\}$.

Da das Nennerpolynom q einen Grad ≥ 0 haben soll, besitzt es höchstens endlich viele Nullstellen. Eine rationale Funktion ist also fast überall definiert.

Was passiert in einem Punkt x_0 mit $q(x_0) = 0$?

Wir können den Faktor $(x - x_0)$ in der höchstmöglichen Potenz aus $p(x)$ und $q(x)$ herausziehen: Es gibt Zahlen $k \geq 0$ und $l > 0$, sowie Polynome \tilde{p} und \tilde{q} , so daß gilt:

$$p(x) = (x - x_0)^k \cdot \tilde{p}(x) \text{ und } q(x) = (x - x_0)^l \cdot \tilde{q}(x),$$

mit $\tilde{p}(x_0) \neq 0$ und $\tilde{q}(x_0) \neq 0$. Dabei ist auch $k = 0$ möglich.

Nun sind zwei Fälle zu unterscheiden:

1. Ist $k \geq l$, so ist

$$f(x) = (x - x_0)^{k-l} \cdot \frac{\tilde{p}(x)}{\tilde{q}(x)},$$

und die rechte Seite der Gleichung ist auch in $x = x_0$ definiert. Man nennt x_0 dann eine *hebbare Unbestimmtheitsstelle* von f . Der „unbestimmte Wert“ $f(x_0) = \frac{0}{0}$ kann in diesem Fall durch eine bestimmte reelle Zahl ersetzt werden.

2. Ist $k < l$, so ist

$$f(x) = \frac{1}{(x - x_0)^{l-k}} \cdot \frac{\tilde{p}(x)}{\tilde{q}(x)}.$$

Hier ist der erste Faktor bei x_0 nach wie vor nicht definiert, während der zweite Faktor einen bestimmten Wert annimmt. Man sagt in dieser Situation: f besitzt in x_0 eine *Polstelle* der Ordnung $l - k$.

Ist $f(x) = p(x)/q(x)$ eine beliebige rationale Funktion und $\text{grad}(p) \geq \text{grad}(q)$, so liefert der Satz von der Division mit Rest eine Zerlegung $p(x) = g(x) \cdot q(x) + r(x)$ mit $\text{grad}(r) < \text{grad}(q)$ und damit $f(x) = g(x) + r(x)/q(x)$.

Ist $\text{grad}(p) < \text{grad}(q)$ und $q(x) = (x - c_1)^{m_1} \cdot \dots \cdot (x - c_k)^{m_k}$, so gibt es eine sogenannte *Partialbruchzerlegung*:

$$f(x) = \sum_{j=1}^k \sum_{i=1}^{m_j} \frac{a_{ji}}{(x - c_j)^i}, \text{ mit } a_{ji} \in \mathbb{R}.$$

Wenn $q(x)$ nicht in Linearfaktoren zerfällt, dann wird die Formel noch etwas komplizierter, darauf kommen wir zu einem späteren Zeitpunkt zurück. Es gibt viele Situationen, in denen man eine Partialbruchzerlegung braucht. Deshalb soll hier schon einmal das Verfahren in der einfacheren Situation beschrieben werden. Folgende Schritte sind auszuführen:

1. Zerlegung des Nenners in Linearfaktoren. Meistens scheitert man schon an dieser Stelle.
2. Ansatz mit „unbestimmten Koeffizienten“, so wie in der Formel vorgegeben.
3. Berechnung der Koeffizienten. Das führt auf ein mehr oder weniger kompliziertes Gleichungssystem.

Beispiel.

$$\text{Sei } f(x) := \frac{x^2 + 5x + 2}{(x - 1)(x + 1)^2}.$$

$$\text{Ansatz: } \frac{x^2 + 5x + 2}{(x-1)(x+1)^2} = \frac{a_{11}}{x-1} + \frac{a_{21}}{x+1} + \frac{a_{22}}{(x+1)^2}.$$

Multiplikation mit dem Nenner führt zu

$$\begin{aligned} x^2 + 5x + 2 &= a_{11}(x+1)^2 + a_{21}(x^2 - 1) + a_{22}(x-1) \\ &= (a_{11} + a_{21})x^2 + (2a_{11} + a_{22})x + (a_{11} - a_{21} - a_{22}). \end{aligned}$$

Koeffizientenvergleich liefert die Bestimmungsgleichungen

$$a_{11} + a_{21} = 1, \quad 2a_{11} + a_{22} = 5 \quad \text{und} \quad a_{11} - a_{21} - a_{22} = 2.$$

Als Lösung erhält man: $a_{11} = 2$, $a_{21} = -1$ und $a_{22} = 1$.

5.8 Stetigkeit

Vorbemerkung: Die Folgen $a_n := n$ und $b_n := (-1)^n n$ sind beide unbeschränkt und damit divergent. Trotzdem gibt es einen Unterschied zwischen ihnen. Von der ersten würden wir gefühlsmäßig sagen, daß sie gegen $+\infty$ strebt, von der zweiten nicht. Deshalb sagen wir:

$$\lim_{n \rightarrow \infty} a_n = +\infty : \iff \forall N \in \mathbb{N} \exists n_0, \text{ s.d. } a_n \geq N \text{ für } n \geq n_0,$$

und entsprechend

$$\lim_{n \rightarrow \infty} a_n = -\infty : \iff \forall N \in \mathbb{N} \exists n_0, \text{ s.d. } a_n \leq -N \text{ für } n \geq n_0.$$

Außerdem sei noch einmal daran erinnert, daß $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$ ist.

Definition:

Sei $I \subset \mathbb{R}$ ein Intervall, $x_0 \in I$ ein Punkt und f eine auf $I \setminus \{x_0\}$ definierte Funktion.

Wir sagen, f strebt bei Annäherung an x_0 gegen den Wert $c \in \overline{\mathbb{R}}$, falls gilt: Für **jede** Folge (x_n) in $I \setminus \{x_0\}$ mit $\lim_{n \rightarrow \infty} x_n = x_0$ ist $\lim_{n \rightarrow \infty} f(x_n) = c$. Man schreibt dann

$$\lim_{x \rightarrow x_0} f(x) = c$$

und nennt c den Grenzwert von f für $x \rightarrow x_0$.

Dabei ist auch $x_0 = \pm\infty$ zugelassen.

Beispiele.

1. Sei $f(x) := \frac{x^2 - 1}{x + 1}$. Die Funktion ist auf $\mathbb{R} \setminus \{-1\}$ nicht definiert. Für $x \neq -1$ ist $f(x) = x - 1$. Ist (x_n) eine beliebige Folge, die gegen $x_0 := -1$ konvergiert, so konvergiert $f(x_n) = x_n - 1$ gegen -2 . Also ist $\lim_{x \rightarrow -1} f(x) = -2$.
2. Sei nun $f(x) := \frac{x^2 + 1}{x + 1}$. Diese Funktion ist ebenfalls nicht in $x = -1$ definiert, aber sie läßt sich nicht so leicht vereinfachen.

Strebt x_n gegen -1 , so strebt der Zähler $(x_n)^2 + 1$ gegen 2 , aber der Nenner $x_n + 1$ gegen 0 . Also ist $f(x_n)$ unbeschränkt, und $\lim_{x \rightarrow -1} f(x)$ existiert nicht!

Wir können noch nicht einmal $\lim_{x \rightarrow -1} f(x) = \infty$ sagen, denn für $x_n < -1$ ist $f(x_n) < 0$ und für $x_n > -1$, so ist $f(x_n) > 0$. Bei Annäherung „von links“ erhält man daher den Grenzwert $-\infty$ und bei Annäherung „von rechts“ den Grenzwert $+\infty$.

Wir müssen also auch die einseitige Annäherung an eine Stelle untersuchen:

Definition:

Es sei $-\infty \leq a < b \leq +\infty$ und f auf (a, b) definiert. Wir sagen, f besitzt in a die Zahl c als rechtsseitigen Grenzwert, falls gilt:

Für jede Folge (x_n) mit $x_n > a$ und $\lim_{n \rightarrow \infty} x_n = a$ ist $\lim_{n \rightarrow \infty} f(x_n) = c$.

Wir schreiben dann:

$$c = \lim_{x \rightarrow a+} f(x) = f(a+).$$

Dabei ist auch $c = \pm\infty$ zugelassen.

Analog wird der linksseitige Grenzwert $\lim_{x \rightarrow b-} f(x) = f(b-)$ mit Hilfe von Folgen (x_n) mit $x_n < b$ und $\lim_{n \rightarrow \infty} x_n = b$ erklärt.

Beispiele.

1. Sei $f(x) := [x]$ (Gaußklammer). Dann ist $\lim_{x \rightarrow 1-} f(x) = 0$ und $\lim_{x \rightarrow 1+} f(x) = 1$.
2. Es ist $\lim_{x \rightarrow 0-} \frac{1}{x} = -\infty$ und $\lim_{x \rightarrow 0+} \frac{1}{x} = +\infty$.

Satz. Der Limes von $f(x)$ für $x \rightarrow x_0$ existiert genau dann, wenn der linksseitige und der rechtsseitige Limes von $f(x)$ in x_0 existiert und beide gleich sind.

BEWEIS: Die eine Richtung ist trivial. Setzen wir also voraus, daß die einseitigen Grenzwerte existieren und gleich einer Zahl c sind! Sei (x_n) eine beliebige Folge, die gegen x_0 konvergiert. Befinden sich fast alle x_n auf einer Seite von x_0 , so existiert natürlich der Grenzwert von $f(x_n)$ für $n \rightarrow \infty$. Liegen auf beiden Seiten von x_0 unendlich viele x_n , so kann man sie auf zwei Teilfolgen verteilen, eine links und eine rechts von x_0 . Da beide Bildfolgen gegen c konvergieren, hat $(f(x_n))$ nur einen einzigen Häufungspunkt und ist konvergent. ■

Definition:

Sei $I \subset \mathbb{R}$ ein Intervall, $f : I \rightarrow \mathbb{R}$ eine Funktion und $x_0 \in I$.

f heißt *stetig* in x_0 , falls gilt:

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

f heißt *stetig auf I* , falls f in jedem Punkt $x \in I$ stetig ist.

Satz. Sind f und g beide in x_0 stetig, so sind auch $f \pm g$ und $f \cdot g$ in x_0 stetig. Ist $g(x_0) \neq 0$, so ist f/g in x_0 stetig.

BEWEIS: Aus den Grenzwert-Sätzen folgt für Funktionen f und g auf $I \setminus \{x_0\}$: Wenn die Grenzwerte $\lim_{x \rightarrow a} f(x) = c$ und $\lim_{x \rightarrow a} g(x) = d$ existieren, dann ist

1. $\lim_{x \rightarrow a} (f(x) \pm g(x)) = c \pm d$.
2. $\lim_{x \rightarrow a} (f(x) \cdot g(x)) = c \cdot d$.
3. Ist $d \neq 0$, so ist $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{c}{d}$.

■

Beispiele.

1. Da $f(x) := x$ und natürlich auch alle konstanten Funktionen stetig sind, ist jedes Polynom auf ganz \mathbb{R} stetig. Dann sind aber auch rationale Funktionen auf ihrem Definitionsbereich stetig.

Die rationale Funktion $f(x) := \frac{x^n - 1}{x^m - 1}$ ist in $x = 1$ nicht definiert. Kürzt man durch $x - 1$, so erhält man:

$$f(x) = \frac{1 + x + \cdots + x^{n-1}}{1 + x + \cdots + x^{m-1}} \quad (\text{für } x \neq 1).$$

Also ist $\lim_{x \rightarrow 1} f(x) = \frac{n}{m}$. Deshalb ist die Funktion

$$F(x) := \begin{cases} f(x) & \text{falls } x \neq 1 \\ n/m & \text{falls } x = 1 \end{cases}$$

in $x = 1$ stetig, und man sagt, f ist in $x = 1$ *stetig fortsetzbar*.

- Die Funktion $f(x) := [x]$ ist auf den offenen Intervallen $I_n := (n, n + 1)$ konstant und daher stetig. An den Endpunkten $x_n := n$ existiert der linksseitige und der rechtsseitige Grenzwert, aber es ist

$$\lim_{x \rightarrow n+} f(x) = n = f(n) \quad \text{und} \quad \lim_{x \rightarrow n-} f(x) = n - 1.$$

Also ist f in x_n nicht stetig.

Die Funktion f sei in der Nähe von x_0 definiert. Ist f in x_0 nicht stetig, so nennt man x_0 eine *Unstetigkeitsstelle* von f . Es gibt dann zwei Möglichkeiten.

- Zumindest einer der beiden einseitigen Grenzwerte existiert nicht. Dann nennen wir x_0 eine *wesentliche Unstetigkeit* von f .
- Wenn $f(x_0+)$ und $f(x_0-)$ beide existieren, aber **nicht** gleich sind, so nennt man x_0 eine *Sprungstelle* von f .

Den Wert $|f(x_0+) - f(x_0-)|$ nennt man dann die *Sprunghöhe*.

Satz. Die Funktion $f : I \rightarrow \mathbb{R}$ sei in $x_0 \in I$ stetig. Ist $f(x_0) > 0$, so gibt es eine ε -Umgebung $U_\varepsilon(x_0)$, so daß $f(x) > 0$ für alle $x \in I \cap U_\varepsilon(x_0)$ ist.

Eine analoge Aussage gilt auch, wenn $f(x_0) < 0$ ist.

BEWEIS: Wir verwenden das Widerspruchsprinzip. Wäre die Aussage des Satzes falsch, so gäbe es eine Folge (x_n) , die gegen x_0 konvergiert, so daß $f(x_n) \leq 0$ für alle n ist. Weil f stetig ist, müßte dann auch

$$f(x_0) = \lim_{n \rightarrow \infty} f(x_n) \leq 0$$

sein. ■

Zwischenwertsatz

Sei f auf dem abgeschlossenen Intervall $[a, b]$ stetig, $f(a) \neq f(b)$ und c ein Wert zwischen $f(a)$ und $f(b)$.

Dann gibt es ein $\xi \in [a, b]$ mit $f(\xi) = c$.

BEWEIS: Ist $c = f(a)$ oder $c = f(b)$, so ist nichts zu zeigen.

Es sei also $f(a) < c < f(b)$ (der Fall $f(a) > c > f(b)$ wird ganz analog behandelt). Der Einfachheit halber betrachten wir nur den Fall $c = 0$.

Die Menge $\{x \in [a, b] \mid f(x) < 0\}$ ist nicht leer, denn sie enthält auf jeden Fall die Zahl a . Also existiert $\xi := \sup(M)$, und man kann in $[a, \xi)$ eine monoton wachsende Folge (x_ν) finden, die gegen ξ konvergiert. Wegen der Stetigkeit von f ist $f(\xi) \leq 0$. Wäre sogar $f(\xi) < 0$, so wäre $f(x) < 0$ für alle x in einer ε -Umgebung von ξ . Insbesondere wäre dann z.B. $f(\xi + \varepsilon/2) < 0$, im Widerspruch zur Definition von ξ . Also muß $f(\xi) = 0$ sein. ■

Folgerung. Sei $f(x)$ ein Polynom vom Grad n . Ist n ungerade, so besitzt f wenigstens eine Nullstelle.

BEWEIS: O.B.d.A. sei f normiert, d.h.

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = x^n(1 + a_{n-1}x^{-1} + \cdots + a_1x^{-n+1} + a_0x^{-n}).$$

Für $x \rightarrow \pm\infty$ strebt x^{-1} gegen Null. Deshalb verhält sich $f(x)$ für $x \rightarrow \infty$ wie x^n . Da n ungerade ist, gilt:

$$\lim_{x \rightarrow -\infty} x^n = -\infty \quad \text{und} \quad \lim_{x \rightarrow +\infty} x^n = +\infty.$$

Beschränkt man die stetige Funktion f auf ein Intervall $[a, b]$ mit $a \ll 0$ und $b \gg 0$ ist, so ist $f(a) < 0$ und $f(b) > 0$, und nach dem Zwischenwertsatz erhält man ein $\xi \in \mathbb{R}$ mit $f(\xi) = 0$. ■

Das folgende Kriterium ist recht nützlich:

Stetigkeit und Unstetigkeit monotoner Funktionen

Sei $f : I := [a, b] \rightarrow \mathbb{R}$ streng monoton wachsend, $J := [f(a), f(b)]$. Dann ist $f(I) \subset J$, und f besitzt höchstens Sprungstellen als Unstetigkeitsstellen.

f ist genau dann auf ganz I stetig, wenn $f(I) = J$ ist.

BEWEIS: Mit Hilfe des Satzes von der monotonen Konvergenz kann man folgern, daß überall der linksseitige und der rechtsseitige Grenzwert existiert. Ist $f : I \rightarrow J$ monoton und surjektiv, so kann es aber keine Sprungstellen geben. Umgekehrt folgt aus Monotonie und Stetigkeit die Surjektivität, wegen des Zwischenwertsatzes. ■

5.9 Exponentialfunktion und Logarithmus

Ist $a > 1$ eine reelle Zahl und $q = m/n \in \mathbb{Q}$, so ist $a^q = \sqrt[n]{a^m}$. Man kann zeigen: Konvergiert eine Folge (q_ν) in \mathbb{Q} gegen q , so konvergiert auch a^{q_ν} gegen a^q .

Ist nun x eine beliebige reelle Zahl und $I_n = [p_n, q_n]$ eine rationale Intervallschachtelung, die x definiert, so ist auch $[a^{p_n}, a^{q_n}]$ eine Intervallschachtelung. Die dadurch eindeutig definierte Zahl bezeichnen wir mit a^x .

Die durch

$$\boxed{\exp_a(x) := a^x}$$

definierte Funktion heißt *Exponentialfunktion zur Basis a* . Sie hat folgende Eigenschaften:

1. $\exp_a : \mathbb{R} \rightarrow \mathbb{R}$ ist auf ganz \mathbb{R} stetig,
2. es ist $\exp_a(0) = 1$, $\exp_a(h) > 1$ für $h > 0$ und speziell $\exp_a(1) = a$,
3. und es gilt die „Funktionalgleichung“

$$\exp_a(x_1 + x_2) = \exp_a(x_1) \cdot \exp_a(x_2).$$

Daraus folgt, daß $\exp_a(-x) = (\exp_a(x))^{-1}$ ist. Nun ist auch klar, daß stets $\exp_a(x) > 0$ ist. Außerdem ist \exp_a streng monoton wachsend, denn es ist $\exp_a(x + h) = \exp_a(x) \cdot \exp_a(h) > \exp_a(x)$ für $h > 0$. Es folgt, daß

$$\exp_a : \mathbb{R} \rightarrow \mathbb{R}_+$$

surjektiv und damit bijektiv ist.

Sehr häufig werden Wachstumsvorgänge durch eine Funktion der Gestalt $x \mapsto f(x) = a^x$ beschrieben (wobei x für die Zeit und $f(x)$ für die anwachsende Größe steht). Ganz besonders interessiert man sich für die (*natürliche*) *Exponentialfunktion*

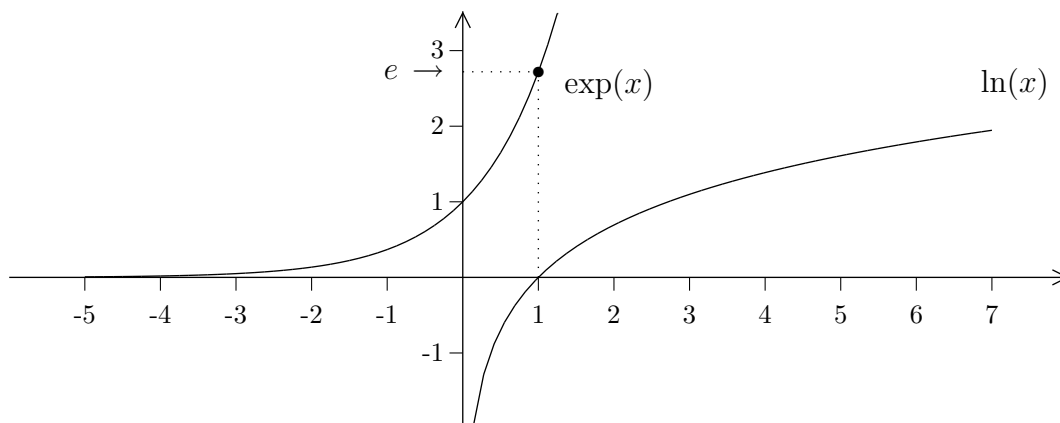
$$\exp(x) = e^x.$$

Definition:

Die Umkehrfunktion $\log_a(x)$ zur Funktion \exp_a bezeichnet man als *Logarithmus zur Basis a* .

Mit $\lg(x)$ bezeichnet man den *Briggschen Logarithmus* $\log_{10}(x)$, mit $\ln(x)$ den *natürlichen Logarithmus* $\log_e(x)$.

Die Graphen von Exponentialfunktion und Logarithmus haben folgende Gestalt:



Aus der Gleichung $\exp(\ln(x)) = x$ (für $x > 0$) folgt die Funktionalgleichung für den Logarithmus:

$$\ln(x \cdot y) = \ln(x) + \ln(y).$$

Entsprechend ist auch $\ln(x/y) = \ln(x) - \ln(y)$ und $\ln(x^\alpha) = \alpha \cdot \ln(x)$ für $\alpha \in \mathbb{Q}$. Da das Logarithmieren Punkt-Rechnungen in Strich-Rechnungen verwandelt, hat man vor der Einführung des Taschenrechners oder Computers viel mit Logarithmentafeln gearbeitet. Wegen unseres Dezimalsystems benutzten diese Tafeln üblicherweise die Basis 10.

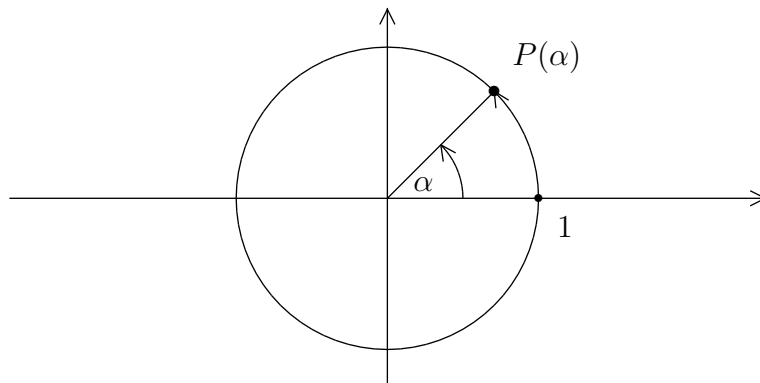
Die allgemeine Exponentialfunktion kann nun in folgender Form geschrieben werden:

$$a^x = \exp(x \cdot \ln(a)).$$

Und die *Potenzfunktion* x^α ist gegeben durch $x^\alpha = \exp(\alpha \cdot \ln(x))$.

5.10 Winkelfunktionen

Wir betrachten einen Punkt P auf dem Rand des Einheitskreises:



Vom Ursprung aus sieht man P unter einem gewissen Winkel α , gemessen von der positiven x-Achse aus. Der Winkel bewegt sich zwischen 0° und 360° . Die Länge des Kreisbogens über dem Winkel ist proportional zu α . Bezeichnet π die Länge des halben Einheitskreisbogens, so ist die Länge des Kreisbogens über α die Zahl

$$\text{Arcus}(\alpha) = \frac{\pi}{180} \cdot \alpha.$$

Man spricht auch vom *Bogenmaß* des Winkels α .

Die Zahl π ist irrational, ja sogar *transzendent*, d.h. sie kann nicht mit Hilfe von Wurzeln aus rationalen Zahlen dargestellt werden. Der Anfang der Dezimalbruchentwicklung von π ist

$$\pi = 3.141592653\dots$$

Die Ziffern 1, 1, 3, 3, 5, 5 liefern übrigens bei geeigneter Anordnung eine erstaunlich gute rationale Näherung von π :

$$\frac{355}{113} = 3.1415929$$

Jeder Punkt $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$ kann eindeutig durch den Abstand $r = \sqrt{x^2 + y^2}$ vom Nullpunkt und den Winkel α (gemessen im Bogenmaß) zwischen der x -Achse und der Geraden durch $(0, 0)$ und (x, y) beschrieben werden. Wir nennen r und α die *Polarkoordinaten* von (x, y) .

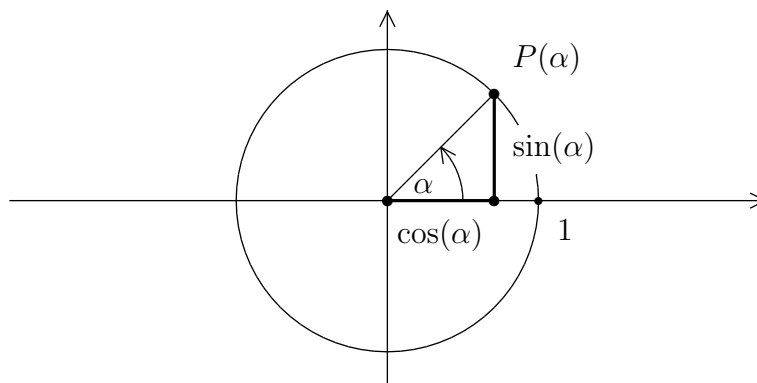
Jede Zahl t aus dem Intervall $[0, 2\pi)$ bestimmt auf eindeutige Weise einen Punkt $P(t) = (x(t), y(t))$ mit den Polarkoordinaten 1 und t . Läßt man zu, daß der Kreis mehrfach durchlaufen wird, und auch in umgekehrter Richtung, so kann t beliebig in \mathbb{R} gewählt werden, und $x(t), y(t)$ werden zu periodischen Funktionen (mit Periode 2π) auf \mathbb{R} .

Definition:

Die Winkelfunktionen *Sinus* und *Cosinus* werden definiert durch

$$P(t) = (\cos(t), \sin(t)).$$

Wir schreiben allerdings auch dann $\cos(\alpha)$ und $\sin(\alpha)$, wenn mit α der Winkel im Gradmaß gemeint ist, d.h. es ist z.B. $\cos(45^\circ) = \cos(\pi/4)$.



Von Sinus und Cosinus werden weitere Funktionen abgeleitet, wie

$$\begin{aligned}\tan(x) &:= \frac{\sin(x)}{\cos(x)} \quad (\text{Tangens}) \\ \text{und } \cot(x) &:= \frac{\cos(x)}{\sin(x)} \quad (\text{Cotangens}).\end{aligned}$$

Da die Punkte $P(\alpha)$ und $P(-\alpha)$ spiegelbildlich zur x-Achse liegen, ist Cosinus eine gerade und Sinus eine ungerade Funktion. Daraus folgt, daß Tangens und Cotangens beides ungerade Funktionen sind.

Mit dem Satz des Pythagoras und etwas Elementargeometrie kann man sehr schnell einige Werte der Winkelfunktionen zu berechnen:

x	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	π
$\cos(x)$	1	$\frac{1}{2}\sqrt{3}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}$	0	-1
$\sin(x)$	0	$\frac{1}{2}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{3}$	1	0
$\tan(x)$	0	$\frac{1}{3}\sqrt{3}$	1	$\sqrt{3}$	∞	0

Um weitere Werte zu ermitteln, braucht man die

Die Additionstheoreme für Sinus und Cosinus

Für beliebige Winkel α, β gilt:

$$\begin{aligned}\sin(\alpha + \beta) &= \sin(\alpha) \cos(\beta) + \cos(\alpha) \sin(\beta), \\ \cos(\alpha + \beta) &= \cos(\alpha) \cos(\beta) - \sin(\alpha) \sin(\beta).\end{aligned}$$

Der Beweis kann – zumindest für spitze Winkel – elementargeometrisch geführt werden.

Folgerung. Für beliebige Winkel α gilt:

$$\begin{aligned}\sin(2\alpha) &= 2 \sin(\alpha) \cos(\alpha) \\ \cos(2\alpha) &= \cos^2(\alpha) - \sin^2(\alpha).\end{aligned}$$

Außerdem ist

$$\sin\left(\alpha + \frac{\pi}{2}\right) = \cos(\alpha) \quad \text{und} \quad \cos\left(\alpha + \frac{\pi}{2}\right) = -\sin(\alpha).$$

Es folgt insbesondere, daß die Graphen von Sinus und Cosinus gleich aussehen, sie sind nur um $\frac{\pi}{2}$ gegeneinander verschoben.

Da $P(t) = (\cos(t), \sin(t))$ stets auf dem Einheitskreis liegt, gilt außerdem:

Satz. Für alle $x \in \mathbb{R}$ ist $\sin(x)^2 + \cos(x)^2 = 1$.

Aus der geometrischen Konstruktion der Winkelfunktionen folgt: $\sin(x) > 0$ für $0 < x < \pi$ und $\cos(x) > 0$ für $0 < x < \pi/2$.

Satz. Die Funktion $\sin(x)$ ist auf $[0, \frac{\pi}{2}]$ streng monoton wachsend.

BEWEIS: Sei $0 \leq x < y \leq \pi/2$. Dann ist

$$\sin(y) - \sin(x) = 2 \cos \frac{y+x}{2} \sin \frac{y-x}{2} > 0.$$

Die benutzte Gleichung folgt aus dem Additionstheorem, unter Verwendung von $y = \frac{y+x}{2} + \frac{y-x}{2}$ und $x = \frac{y+x}{2} - \frac{y-x}{2}$. ■

Da der Sinus eine ungerade Funktion ist, ist $\sin(x)$ sogar auf $[-\frac{\pi}{2}, +\frac{\pi}{2}]$ streng monoton wachsend. Und da

$$\sin\left(\frac{\pi}{2} + x\right) = \sin\left(\frac{\pi}{2} - x\right)$$

gilt, ist $\sin(x)$ auf $[\frac{\pi}{2}, \frac{3\pi}{2}]$ streng monoton fallend.

Satz. Sinus und Cosinus sind auf ganz \mathbb{R} stetig.

BEWEIS: Da die Sinusfunktion das Intervall $[-\pi/2, \pi/2]$ surjektiv auf $[-1, 1]$ abbildet (geometrische Deutung des Sinus) und dabei streng monoton wachsend ist, muß sie dort stetig sein. Speziell ist

$$\lim_{x \rightarrow 0} \sin(x) = 0 = \sin(0),$$

und weil $\cos(x) = \sqrt{1 - \sin^2(x)}$ ist, folgt:

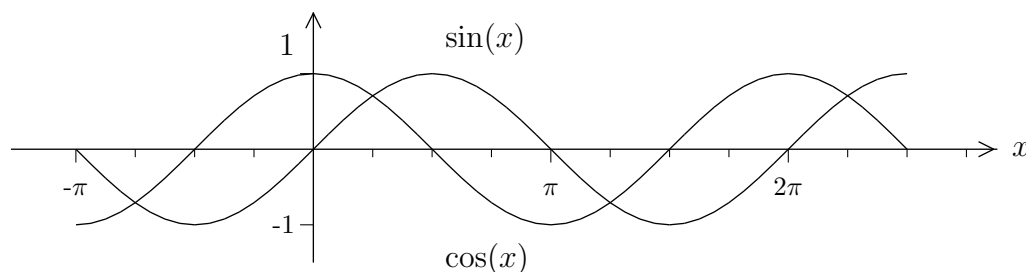
$$\lim_{x \rightarrow 0} \cos(x) = 1 = \cos(0).$$

Sei $x_0 \in \mathbb{R}$ beliebig, (x_n) eine Folge, die gegen x_0 konvergiert, $h_n := x_n - x_0$. Dann gilt:

$$\sin(x_n) = \sin(x_0 + h_n) = \sin(x_0) \cos(h_n) + \cos(x_0) \sin(h_n),$$

und dieser Ausdruck strebt gegen $\sin(x_0)$. Also ist der Sinus (und damit auch der Cosinus) in x_0 stetig. ■

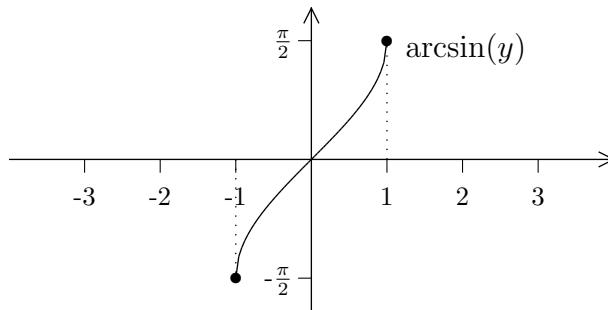
Für die Funktionsgraphen von Sinus und Cosinus ergibt sich nun folgendes Bild:



Da $\sin : [-\frac{\pi}{2}, +\frac{\pi}{2}] \rightarrow [-1, +1]$ bijektiv ist, gibt es eine Umkehrabbildung

$$\arcsin := \sin^{-1} : [-1, +1] \rightarrow [-\frac{\pi}{2}, +\frac{\pi}{2}].$$

Sie wird als (*Hauptwert des*) *Arcussinus* bezeichnet. Vom „Hauptwert“ spricht man, weil man den Sinus natürlich auf jedem Intervall umkehren kann, wo er streng monoton ist. Das ergibt andere „Zweige“ des Arcussinus.



Der Cosinus ist auf $[0, \pi]$ streng monoton fallend, bildet dieses Intervall also bijektiv auf $[-1, 1]$ ab. Die Umkehrfunktion $\arccos : [-1, 1] \rightarrow [0, \pi]$ ist der Hauptwert des *Arcuscosinus*.

Zwischen 0 und 2π hat der Sinus die Nullstellen $0, \pi, 2\pi$ (wie man sofort am Einheitskreis sieht). Also ist

$$\{x \in \mathbb{R} \mid \sin(x) = 0\} = \{x \in \mathbb{R} \mid \exists n \in \mathbb{Z} \text{ mit } x = n\pi\}$$

die Menge aller Nullstellen des Sinus, und entsprechend ist

$$\{x \in \mathbb{R} \mid \cos(x) = 0\} = \{x \in \mathbb{R} \mid \exists n \in \mathbb{Z} \text{ mit } x = \frac{\pi}{2} + n\pi\}$$

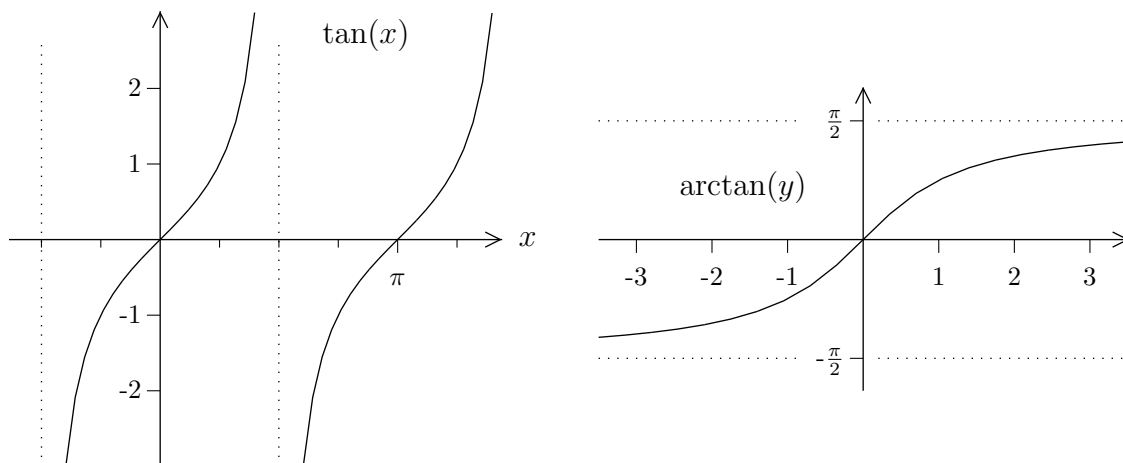
die Menge aller Nullstellen des Cosinus. Außerhalb der letzteren Menge ist der Tangens definiert, und offensichtlich gilt:

$$\lim_{x \rightarrow -\pi/2+} \tan(x) = -\infty \text{ und } \lim_{x \rightarrow +\pi/2-} \tan(x) = +\infty.$$

Man kann sich auch überlegen, daß $\tan : (-\frac{\pi}{2}, +\frac{\pi}{2}) \rightarrow \mathbb{R}$ streng monoton wachsend und stetig ist, also bijektiv. Die Umkehrfunktion ist der *Arcustangens*.

$$\arctan := \tan^{-1} : \mathbb{R} \rightarrow (-\frac{\pi}{2}, +\frac{\pi}{2})$$

Die Graphen von Tangens und Arcustangens sehen folgendermaßen aus:



Schließlich kann man auch den Cotangens auf gewissen Intervallen umkehren und erhält so den Arcuscotangens.

5.11 Hyperbelfunktionen

Definition:

Die auf ganz \mathbb{R} definierten Funktionen

$$\sinh(x) := \frac{1}{2}(e^x - e^{-x}) \quad \text{und} \quad \cosh(x) := \frac{1}{2}(e^x + e^{-x})$$

heißen *Sinus hyperbolicus* und *Cosinus hyperbolicus*.

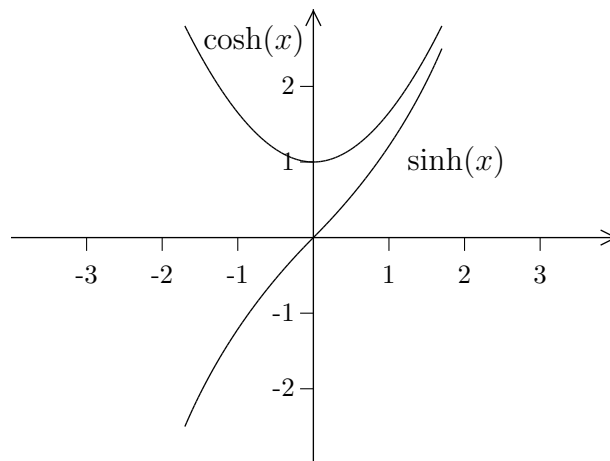
Offensichtlich ist $\cosh(x)$ eine gerade und $\sinh(x)$ eine ungerade Funktion. Weiter gilt:

$$\cosh(0) = 1 \quad \text{und} \quad \lim_{x \rightarrow -\infty} \cosh(x) = \lim_{x \rightarrow +\infty} \cosh(x) = +\infty,$$

sowie

$$\sinh(0) = 0, \quad \lim_{x \rightarrow -\infty} \sinh(x) = -\infty \quad \text{und} \quad \lim_{x \rightarrow +\infty} \sinh(x) = +\infty.$$

Außerdem ist $\sinh(x) < \cosh(x)$ für alle $x \in \mathbb{R}$. Die Graphen sehen also ungefähr folgendermaßen aus:



Es ist

$$\cosh^2(x) - \sinh^2(x) = 1 \quad \text{für alle } x.$$

Deshalb wird durch

$$\varphi(t) := (\cosh(t), \sinh(t))$$

die *Hyperbel* $H := \{(x, y) \in \mathbb{R}^2 \mid x^2 - y^2 = 1\}$ parametrisiert.

Schließlich definiert man noch:

$$\begin{aligned} \tanh(x) &:= \frac{\sinh(x)}{\cosh(x)} \\ \text{und} \quad \coth(x) &:= \frac{\cosh(x)}{\sinh(x)}. \end{aligned}$$

Über Umkehrfunktionen der Hyperbelfunktionen werden wir an späterer Stelle sprechen.