

---

# Kapitel I

## Die Objekte der Mathematik

Objekte der Mathematik sind

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

Neue Objekte werden gebildet, indem man schon bekannte Objekte zu *Mengen* zusammenfaßt.

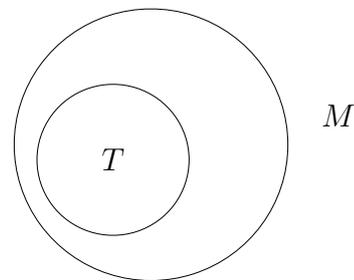
Ist  $x$  ein Element der Menge  $M$ , so schreibt man:

$$x \in M.$$

Liegt  $x$  nicht in  $M$ , so schreibt man:  $x \notin M$ .

Faßt man einen Teil der Elemente von  $M$  zu einer eigenen Menge  $T$  zusammen, so nennt man  $T$  eine *Teilmenge* von  $M$  und schreibt:

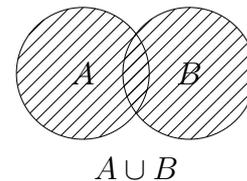
$$T \subset M.$$



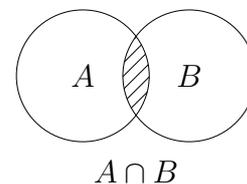
Enthält  $T$  wenigstens ein Element, das nicht zu  $M$  gehört, so ist  $T$  keine Teilmenge von  $M$ , in Zeichen:  $T \not\subset M$ .

Sind schließlich  $A$  und  $B$  irgendwelche Mengen, so kann man aus ihnen neue Mengen bilden:

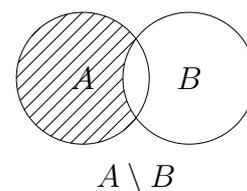
1. Die *Vereinigung*  $A \cup B$ , die alle Elemente von  $A$  und  $B$  zusammen enthält.



2. Den *Durchschnitt*  $A \cap B$ , der nur die Elemente enthält, die zugleich in  $A$  und  $B$  liegen.



3. Die *Differenz*  $A \setminus B$ , die alle diejenigen Elemente von  $A$  enthält, die nicht in  $B$  liegen.



Schließlich erweist es sich als zweckmäßig, ein Symbol für die *leere Menge* einzuführen, die überhaupt kein Element enthält, nämlich  $\emptyset$ .

## §1 Zahlen

Mit  $\mathbb{R}$  bezeichnen wir die Menge der *reellen Zahlen*, die wir uns anschaulich als die Menge der Punkte auf einer Geraden vorstellen können.

$\mathbb{N} := \{1, 2, 3, 4, \dots\}$  ist die Teilmenge der *natürlichen Zahlen*.

(Wir verwenden hier die *aufzählende Schreibweise* für die Darstellung einer Menge. Das Zeichen „:=“ bedeutet: „ist definiert als“)

$\mathbb{N}_0 := \mathbb{N} \cup \{0\} = \{0, 1, 2, 3, 4, \dots\}$  ist die Menge der natürlichen Zahlen einschließlich der Null.

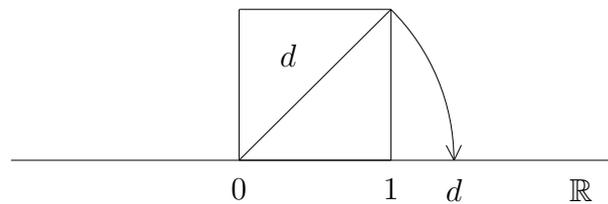
$\mathbb{Z} := \{0, +1, -1, +2, -2, \dots\}$  ist die Menge der *ganzen Zahlen*.

$\mathbb{Q} := \left\{ \frac{a}{b} \mid a \in \mathbb{Z} \text{ und } b \in \mathbb{N} \right\}$  ist die Menge der *rationalen Zahlen*. Man beachte, daß die Nenner hier automatisch immer  $\neq 0$  sind. Wir haben zur Darstellung von  $\mathbb{Q}$  die *beschreibende Schreibweise* benutzt. Zu beachten ist auch noch, daß **jeder** Ausdruck der Gestalt  $\frac{a}{b}$  mit  $a, b \in \mathbb{R}$  und  $b \neq 0$  als *Bruch* bezeichnet wird. *Rational* sind nur solche Brüche, bei denen Zähler und Nenner ganzzahlig sind. Die Bruch-Darstellung ist bekanntlich nicht eindeutig, man kann kürzen und erweitern.

$\mathbb{R} \setminus \mathbb{Q}$  ist die Menge der *irrationalen Zahlen*.

**I.1.1 Satz.** *Es gibt irrationale Zahlen.*

BEWEIS: Wir beginnen mit einer geometrischen Konstruktion:



Für die Diagonale  $d$  eines Quadrates mit der Seitenlänge 1 gilt nach dem Satz des Pythagoras:  $1^2 + 1^2 = d^2$ , also  $d^2 = 2$ . Mit Hilfe eines Zirkels kann man die Länge  $d$  auf der Zahlengeraden abtragen.

Wir wollen zeigen, daß  $d$  irrational ist. Dazu verwenden wir das Prinzip des Widerspruchsbeweises:

Annahme,  $d \in \mathbb{Q}$ . Dann ist  $d = \frac{m}{n}$ , mit  $m, n \in \mathbb{N}$ . Durch Kürzen kann man erreichen, daß  $m$  und  $n$  außer 1 keinen gemeinsamen Teiler haben.

Da  $d^2 = 2$  ist, folgt:  $m^2 = 2 \cdot n^2$ , also ist  $m^2$  eine gerade Zahl. Das ist nur möglich, wenn  $m$  selbst gerade ist, also von der Form  $m = 2k$ . Setzen wir das oben wieder ein, so erhalten wir:  $4k^2 = 2n^2$ . Also ist auch  $n^2$  eine gerade Zahl. Dann ist aber auch  $n$  gerade, und  $n$  und  $m$  haben die 2 als gemeinsamen Teiler. Das ist ein Widerspruch! Also war die Annahme falsch, die Zahl  $d = \sqrt{2}$  ist irrational!  $\square$

Das Rechnen in  $\mathbb{R}$  wird hier als bekannt vorausgesetzt, ich erinnere nur an einige Regeln:

$$\left. \begin{array}{l} a + b = b + a \\ a \cdot b = b \cdot a \end{array} \right\} \text{Kommutativgesetze}$$

$$\left. \begin{array}{l} a + (b + c) = (a + b) + c \\ a \cdot (b \cdot c) = (a \cdot b) \cdot c \end{array} \right\} \text{Assoziativgesetze}$$

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{Distributivgesetz}$$

Im Gegensatz zu solchen allgemeingültigen Gleichungen kennen wir auch sogenannte *Bestimmungsgleichungen*. Sie treten z.B. in der Beschreibung einer Menge auf:

$$M := \{x \in \mathbb{R} \mid a \cdot x + b = c\}.$$

Um  $M$  genauer zu bestimmen, müssen wir die Gleichung  $a \cdot x + b = c$  lösen. Dabei sind  $a, b, c$  als bekannte Konstanten aufzufassen,  $x$  ist variabel. Wir müssen diejenigen Werte für  $x$  finden, für die die Gleichung erfüllt ist. Dabei gibt es mancherlei Fallen. Der Faktor  $a$  könnte z.B.  $= 0$  sein, dann reduziert sich die Bestimmungsgleichung auf  $b = c$ , und das  $x$  kommt nicht mehr vor. Ist tatsächlich  $b = c$ , so ist  $M = \mathbb{R}$ , ist jedoch  $b \neq c$ , so ist  $M = \emptyset$ . Wenn aber  $a \neq 0$  ist, dann ist die Gleichung eindeutig lösbar, und es ist  $M = \{\frac{1}{a}(c - b)\}$  eine Menge mit genau einem Element.

Interessanter ist schon eine quadratische Gleichung. Wir beginnen mit dem einfachsten Fall, nämlich der Gleichung

$$x^2 = a.$$

Hier kommt ins Spiel, daß es positive und negative reelle Zahlen gibt.

Ist  $a > 0$ , so gibt es zwei verschiedene Lösungen, nämlich  $x = +\sqrt{a}$  und  $x = -\sqrt{a}$ .

Ist  $a = 0$ , so gibt es nur die Lösung  $x = 0$ .

Ist  $a < 0$ , so gibt es keine Lösung!

Zur Begründung des letzten Falles müssen wir uns kurz an das Rechnen mit Ungleichungen erinnern:

Sind  $a > 0$  und  $b > 0$ , so ist auch  $a + b > 0$  und  $a \cdot b > 0$ . Außerdem ist  $1 > 0$ . Diese Tatsachen werden einfach so festgelegt. Jede Zahl  $a \neq 0$  ist entweder positiv oder negativ, und wenn  $a < 0$  ist, so ist  $-a > 0$ . Da schließlich  $(-a) \cdot (-a) = a \cdot a = a^2$  ist, ist  $a^2 > 0$  für jedes  $a \neq 0$ .

Weiter definiert man:

$$\begin{aligned} a < b & : \iff b - a > 0, & 1 \\ a \leq b & : \iff a < b \text{ oder } a = b, \\ a > b & : \iff b < a, \\ \text{und } a \geq b & : \iff a > b \text{ oder } a = b. \end{aligned}$$

So ist – im Gegensatz zur Umgangssprache – die Aussage „ $3 \geq 3$ “ mathematisch völlig korrekt.

<sup>1</sup>„:  $\iff$ “ steht für „bedeutet definitionsgemäß“. Eine *Definition* ist die Erklärung eines neu eingeführten Begriffes oder Symbols.

Ist nun etwa  $a < b$  und  $c$  beliebig, so ist  $b - a > 0$  und  $(b + c) - (a + c) = b - a > 0$ , also  $a + c < b + c$ . Auf diese Art kann man die bekannten Regeln herleiten:

$$\begin{aligned} a < b \text{ und } c \text{ beliebig} &\implies a + c < b + c, &^2 \\ a < b \text{ und } c > 0 &\implies a \cdot c < b \cdot c, \\ a < b \text{ und } c < 0 &\implies a \cdot c > b \cdot c, \\ 0 < a < b &\implies 0 < \frac{1}{b} < \frac{1}{a}. \end{aligned}$$

Gewisse Teilmengen von  $\mathbb{R}$  können durch Ungleichungen beschrieben werden:

**Definition.**

$$(a, b) := \{x \in \mathbb{R} \mid a < x < b\}$$

heißt *offenes Intervall* mit den Grenzen  $a$  und  $b$ ,

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

heißt *abgeschlossenes Intervall* mit den Grenzen  $a$  und  $b$ .

Beim abgeschlossenen Intervall gehören die Grenzen dazu, beim offenen nicht!<sup>3</sup>

Man beachte noch:

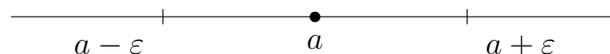
$a < x < b$  bedeutet eigentlich:  $a < x$  **und**  $x < b$ .

Ein Punkt  $x$  liegt also genau dann *nicht* in  $(a, b)$ , wenn entweder  $x \leq a$  **oder**  $x \geq b$  ist.

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

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

die  $\varepsilon$ -Umgebung von  $a$ .



Die Menge  $\mathbb{R}_+ := \{x \in \mathbb{R} \mid x > 0\}$  heißt die *positive Halbgerade* und  $\mathbb{R}_- := \{x \in \mathbb{R} \mid x < 0\}$  die *negative Halbgerade*.

Wir kommen nun zur allgemeinen quadratischen Gleichung

$$ax^2 + bx + c = 0, \text{ mit } a \neq 0.$$

Zur Lösung erinnere man sich an die binomische Formel

$$(x + y)^2 = x^2 + 2xy + y^2.$$

<sup>2</sup>„ $\implies$ “ steht für „daraus folgt“. „ $\iff$ “ steht für „gilt genau dann, wenn“.

<sup>3</sup>Für das offene Intervall  $(a, b)$  sind auch andere Symbole im Gebrauch, wie etwa  $< a, b >$  oder  $]a, b[$ .

Man versucht, die quadratische Gleichung so umzuformen, daß ihre linke Seite die gleiche Form wie die rechte Seite der binomischen Formel annimmt. Das geschieht mit Hilfe der sogenannten „quadratischen Ergänzung“:

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

Der Term  $\Delta := b^2 - 4ac$  heißt *Diskriminante* der quadratischen Gleichung. Wie bei der einfachen quadratischen Gleichung gibt es 3 Möglichkeiten:

1. Ist  $\Delta > 0$ , so gibt es 2 Lösungen:  $x = \frac{1}{2a}(-b \pm \sqrt{\Delta})$ ,
2. Ist  $\Delta = 0$ , so gibt es nur die Lösung  $x = -\frac{b}{2a}$ .
3. Ist  $\Delta < 0$ , so gibt es keine Lösung!

Im ersten Fall kann man schreiben:

$$\{x \in \mathbb{R} \mid ax^2 + bx + c = 0\} = \left\{ \frac{-b + \sqrt{\Delta}}{2a}, \frac{-b - \sqrt{\Delta}}{2a} \right\}.$$

### Definition.

Der *Absolutbetrag* einer reellen Zahl ist gegeben durch

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

So ist z.B.  $|3| = 3$  und  $|-7| = 7$ . Für das Rechnen mit Beträgen haben wir die folgenden Regeln:

1. Den Betrag einer mit Vorzeichen und Dezimalziffern geschriebenen reellen Zahl erhält man, indem man einfach das Vorzeichen wegläßt.
2.  $|a \cdot b| = |a| \cdot |b|$ . Wegen (1) ist das klar.
3.  $|a - b| = |b - a|$  ist der Abstand von  $a$  und  $b$  auf der Zahlengeraden: Ist nämlich etwa  $a < b$  so ist  $|a - b| = b - a$  und  $a + |a - b| = a + (b - a) = b$ .
4. Es ist stets  $-|a| \leq a \leq +|a|$ .
5. Sei  $c > 0$ . Dann ist  
 $|x| < c \iff -c < x < +c$ .

6. Es gilt die *Dreiecks-Ungleichung*:

$$|a + b| \leq |a| + |b|.$$

Zum Beweis: Wegen (4) ist  $-|a| - |b| \leq a + b \leq |a| + |b|$ .  
Wegen (5) folgt daraus die Dreiecksungleichung.

7. Es ist  $|a - b| \geq |a| - |b|$ .

Zum Beweis benutzt man einen beliebigen Trick:  
Es ist  $|a| = |(a - b) + b| \leq |a - b| + |b|$ .

Als Konsequenz erhalten wir z.B.:

$$\begin{aligned} x \in U_\varepsilon(a) &\iff a - \varepsilon < x < a + \varepsilon \\ &\iff -\varepsilon < x - a < +\varepsilon \\ &\iff |x - a| < \varepsilon. \end{aligned}$$

Die  $\varepsilon$ -Umgebung von  $a$  ist also die Menge derjenigen  $x$ , deren Abstand von  $a$  kleiner als  $\varepsilon$  ist.

### Beispiele:

1. Zu bestimmen sind alle reellen Zahlen  $x$  mit

$$|x - 5| < |x + 1|.$$

Wie geht man vor? Die Beträge  $|x - 5|$  und  $|x + 1|$  wechseln ihre Vorzeichen bei  $x = 5$  bzw.  $x = -1$ . Das führt zu folgender Tabelle:

	$x < -1$	$-1 \leq x < 5$	$x \geq 5$
$ x - 5 $	$-x + 5$	$-x + 5$	$x - 5$
$ x + 1 $	$-x - 1$	$x + 1$	$x + 1$
	(I)	(II)	(III)

Im Falle (I) erhält man:

$$-x + 5 < -x - 1 \iff 5 < -1.$$

Das ist immer falsch, also gibt es hier keine Lösung.

Im Falle (II) erhält man:

$$-x + 5 < x + 1 \iff 2x > 4 \iff x > 2.$$

Und im Falle (III) erhält man:

$$x - 5 < x + 1 \iff -5 < 1.$$

Das ist immer richtig, also ist hier jedes  $x$  eine Lösung.

Zusammengefaßt ist die Lösungsmenge gerade  $\{x \in \mathbb{R} \mid x > 2\}$ .

Man kann sich das leicht auch anschaulich überlegen: Gesucht werden diejenigen Zahlen  $x$ , die von 5 weniger weit entfernt sind als von  $-1$ . Da die Mitte zwischen  $-1$  und  $+5$  gerade bei 2 liegt, muß es sich um alle  $x$  mit  $x > 2$  handeln.

2. Es soll eine quadratische Ungleichung gelöst werden:

$$-2x^2 + 9x - 4 < 0.$$

Division durch  $-2$  ändert die Richtung der Ungleichung:

$$x^2 - \frac{9}{2}x + 2 > 0.$$

Zur Lösung benutzen wir wieder die quadratische Ergänzung:

$$x^2 - 2 \cdot \frac{9}{4}x + \left(\frac{9}{4}\right)^2 > \left(\frac{9}{4}\right)^2 - 2,$$

also

$$\left(x - \frac{9}{4}\right)^2 > \frac{49}{16} = \left(\frac{7}{4}\right)^2.$$

Wie ist nun eine Ungleichung vom Typ  $y^2 > c^2$  mit  $c > 0$  zu lösen? Wir erinnern uns an eine weitere binomische Formel:

$$(x - y) \cdot (x + y) = x^2 - y^2.$$

Also ist

$$\begin{aligned} y^2 > c^2 &\iff (y - c) \cdot (y + c) > 0 \\ &\iff y - c, y + c \text{ haben gleiches Vorzeichen.} \end{aligned}$$

Eine reelle Zahl  $x$  ist daher genau dann Lösung von  $(x - \frac{9}{4})^2 > \frac{49}{16}$ , wenn  $(x - 4) \cdot (x - \frac{1}{2}) > 0$  ist, und das ist genau dann der Fall, wenn

(a) entweder  $x > 4$  und  $x > \frac{1}{2}$

(b) oder  $x < 4$  und  $x < \frac{1}{2}$

ist, wenn also entweder  $x > 4$  oder  $x < \frac{1}{2}$  ist.

Die Lösungsmenge ist demnach die Menge

$$\{x \in \mathbb{R} \mid x < \frac{1}{2}\} \cup \{x \in \mathbb{R} \mid x > 4\}.$$

### 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  $\Sigma$  und  $\Pi$  nennt man *Summenzeichen* und *Produktzeichen*.

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

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

Speziell gilt:

1.  $\sum_{i=k}^k a_i = a_k.$

2. Ist  $m < k$ , so wird die „leere Summe“  $\sum_{i=k}^m a_i$  **definitionsgemäß**  $= 0$  gesetzt.

3. Es kann vorkommen, daß  $a_i = a$  ist, für  $i = 1, \dots, n$ . Dann ist

$$\text{offensichtlich } \sum_{i=1}^n a = n \cdot a.$$

Für das Produktzeichen gelten analoge Regeln. Ist  $m < k$ , so wird das „leere Produkt“  $\prod_{i=k}^m a_i = 1$  gesetzt.

Für  $n \in \mathbb{N}_0$  ist  $a^n := \prod_{i=1}^n a := \underbrace{a \cdot a \cdot \dots \cdot a}_{n\text{-mal}}$  die  $n$ -te Potenz von  $a$ . Insbesondere ist  $a^0 = 1$ .

Man beachte, daß  $0^n = 0$  für jedes  $n \in \mathbb{N}$  ist, aber  $0^0 = 1$ .

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

$a^{n+m} = a^n \cdot a^m$
und $a^{n \cdot m} = (a^n)^m.$

Man kann auch negative Exponenten zulassen:

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

Speziell ist  $a^{-1} = \frac{1}{a}$  das *Inverse* zu  $a$ , und allgemein ist

$$a^{-n} = (a^{-1})^n = (a^n)^{-1}.$$

### 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:

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

Wir wollen die Zahl  $n!$  auf andere Weise interpretieren. Dazu brauchen wir ein weiteres Beweisprinzip, und um das vernünftig formulieren zu können, müssen wir erst einmal präzisieren, was eine Aussage ist.

Als *Aussagen* bezeichnen wir vernünftige mathematische Sätze, z.B. „37 ist eine Primzahl“ oder „Zwei Geraden schneiden sich in höchstens einem Punkt“. Man kann von einer Aussage im Prinzip immer entscheiden, ob sie wahr oder falsch ist.

Enthält ein Satz eine Variable  $x$  (z.B. „ $x$  ist eine gerade Zahl“), so kann man den Wahrheitsgehalt nicht mehr ermitteln. Dennoch hat der Satz die *Form* einer Aussage und wird deshalb als *Aussageform* bezeichnet. Ist nun  $\mathcal{A}(x)$  eine solche Aussageform, so gibt es zwei Standardmethoden, daraus eine Aussage zu machen:

1. Die *Existenz-Aussage*  $\exists x : \mathcal{A}(x)$ ,  
in Worten: Es gibt ein  $x$  mit der Eigenschaft  $\mathcal{A}(x)$ .

Z.B.: „Es gibt eine gerade ganze Zahl  $x$ “. Um die Wahrheit festzustellen, braucht man nur ein Beispiel anzugeben. Und es stört keineswegs, daß es in Wirklichkeit sehr viele Beispiele gibt.

2. Die *All-Aussage*  $\forall x : \mathcal{A}(x)$ ,  
in Worten: Für alle  $x$  gilt  $\mathcal{A}(x)$ .

Z.B.: „Alle ganzen Zahlen  $x$  sind gerade“. Dieser Satz ist natürlich falsch, und zum Beweis brauchen wir nur ein Gegenbeispiel anzugeben, etwa  $x = 5$ . Schwieriger wird es bei einer All-Aussage, deren Wahrheit wir zeigen wollen, z.B. „Aus jeder positiven reellen Zahl kann man die Wurzel ziehen“. Eigentlich müßten unendlich viele Aussagen bewiesen werden. Das ist nur zu schaffen, indem man einen allgemein-gültigen Beweis findet.

Wenn sich eine All-Aussage nur auf die Elemente einer bestimmten Menge  $M$  bezieht, dann schreibt man das in der Form

$$\forall x \in M : \mathcal{A}(x).$$

Ist nun zufällig  $M = \mathbb{N}$ , so gibt es ein besonderes Beweisverfahren, den *Beweis durch vollständige Induktion*:

Um den Induktionsbeweis zu verstehen, stelle man 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 alle Steine um.

Um eine Aussage der Form  $\forall n \in \mathbb{N} : \mathcal{A}(n)$  zu beweisen, reicht es, die Aussage  $\mathcal{A}(1)$  zu beweisen und zu zeigen, daß – für beliebiges  $n$  – aus  $\mathcal{A}(n)$  stets  $\mathcal{A}(n + 1)$  folgt. Wohlgemerkt: Man braucht nicht zu zeigen, daß  $\mathcal{A}(n)$  wahr ist! Man muß nur zeigen, daß  $\mathcal{A}(n + 1)$  automatisch wahr wäre, wenn  $\mathcal{A}(n)$  schon wahr wäre.

**Beispiel:**

Wir wollen folgende Aussage beweisen:

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

Eigentlich haben wir nur eine Aussageform  $\mathcal{A}(n)$  hingeschrieben, und wir wollen natürlich die Aussage  $\forall n \in \mathbb{N} : \mathcal{A}(n)$  beweisen. In der Umgangssprache ist klar, daß das gemeint ist.

Nach dem Domino-Prinzip müssen wir zunächst den *Induktionsanfang* erledigen, also die Aussage  $\mathcal{A}(1)$  beweisen: „Es gibt genau eine Möglichkeit, die Zahl 1 anzuordnen“. Nun, das ist so evident, da brauchen wir gar nichts zu zeigen.

Im zweiten Schritt müssen wir den *Induktionsschluß* durchführen, d.h. wir müssen uns davon überzeugen, daß die Domino-Steine richtig stehen. 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.

Jetzt sind wir mit dem Induktionsbeweis fertig! Aus der Aussage  $\mathcal{A}(1)$  folgt wegen des Induktionsschlusses die Aussage  $\mathcal{A}(2)$ , hieraus dann die Aussage  $\mathcal{A}(3)$  usw. Die Dominosteine kippen alle um!

Ein weiteres Beispiel ist die **verallgemeinerte Dreiecks-Ungleichung**:

**I.1.2 Satz.** Sind  $a_1, a_2, \dots, a_n$  reelle Zahlen, so ist

$$|a_1 + a_2 + \dots + a_n| \leq |a_1| + |a_2| + \dots + |a_n|.$$

**BEWEIS:** Wir führen Induktion nach  $n$ .

Der Fall  $n = 1$  ( $|a_1| \leq |a_1|$ ) ist trivial.

Die Aussage sei schon für ein  $n \geq 1$  bewiesen. Dann gilt:

$$\begin{aligned} |a_1 + \dots + a_n + a_{n+1}| &= |(a_1 + \dots + a_n) + a_{n+1}| \\ &\leq |a_1 + \dots + a_n| + |a_{n+1}| \\ &\leq (|a_1| + \dots + |a_n|) + |a_{n+1}| \\ &= |a_1| + \dots + |a_n| + |a_{n+1}|. \end{aligned}$$

Hier wurde das Assoziativgesetz, die gewöhnliche Dreiecksungleichung und die Induktionsvoraussetzung benutzt. □

Ein weiteres Beispiel:

**I.1.3 Satz.** Für alle  $n \in \mathbb{N}$  ist  $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ .

BEWEIS: Wieder führen wir Induktion nach  $n$ :

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

$n \rightarrow n + 1$ : Es ist

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

□

Solche Formeln werden gerne benutzt, um den Induktionsbeweis zu üben. Dabei ist keine Rede davon, wie man die Formel findet, und es wird auch verschwiegen, daß es einen viel einfacheren Beweis gibt:

Es ist nämlich

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

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

Wir werden das Induktionsprinzip künftig nur dort anwenden, wo es nötig ist.

#### I.1.4 Satz. (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:

#### I.1.5 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

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

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 noch 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  $\frac{n!}{k!(n - k)!} = \binom{n}{k}$  Teilmengen von  $\{1, 2, \dots, n\}$  mit  $k$  Elementen.

### Beispiel:

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

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

Bei sovielen Möglichkeiten sollte man die Ausgabe für den nächsten Lottoschein vielleicht doch noch einmal überdenken.

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

### I.1.6 Satz. (Binomische Formel)

Für  $a, b \in \mathbb{R}$  und  $n \in \mathbb{N}$  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  $N(n, k) = \binom{n}{k}$ .  $\square$

Für kleines  $n$  ist diese Formel wohlbekannt:

$$(a + b)^2 = a^2 + 2ab + b^2,$$

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

Zwei Folgerungen wollen wir noch notieren:

**I.1.7 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.

**I.1.8 Folgerung 2.** *Ist  $x \geq 0$ , so ist*

$$(1 + x)^n \geq 1 + nx.$$

BEWEIS: Es ist

$$(1 + x)^n = 1 + nx + \underbrace{\frac{(n-1)n}{2}x^2 + \dots + x^n}_{\geq 0}.$$

□

## §2 Grenzwerte

Die *Folge* der natürlichen Zahlen  $1, 2, 3, 4, \dots$  bricht niemals ab. Wenn wir nun irgend ein unendliches System von reellen Zahlen  $a_1, a_2, a_3, \dots$  vorliegen haben, das mit einer bestimmten Zahl  $a_1$  beginnt und zu dem eine Vorschrift existiert, welche Zahl auf die  $n$ -te Zahl  $a_n$  folgt, so sprechen wir auch von einer

*Folge reeller Zahlen*  $(a_n)$  .

### Definition.

1. Eine Folge  $(a_n)$  heißt *konstant*, falls es ein  $a \in \mathbb{R}$  gibt, so daß  $a_n = a$  für alle  $n \in \mathbb{N}$  ist.
2. Eine Folge  $(a_n)$  heißt

*nach oben beschränkt* (bzw. *nach unten beschränkt*),

falls es ein  $c \in \mathbb{R}$  gibt, so daß für alle  $n \in \mathbb{N}$  gilt:

$$a_n \leq c \quad (\text{bzw. } c \leq a_n).$$

3.  $(a_n)$  heißt *beschränkt*, falls  $(a_n)$  nach oben **und** nach unten beschränkt ist.
4. Eine Folge heißt *unbeschränkt*, falls sie nicht beschränkt ist. (Sie ist dann entweder *nicht* nach oben beschränkt **oder** *nicht* nach unten beschränkt oder in beiden Richtungen nicht beschränkt.)

Wir betrachten einige

### Beispiele :

1. Durch  $a_n := n$  wird die Folge der natürlichen Zahlen  $1, 2, 3, 4, \dots$  gegeben. Unsere Erfahrung sagt uns, daß sie nicht beschränkt ist. Tatsächlich kann man diese Aussage nicht beweisen, die Mathematiker fordern sie als *Axiom*, also als eine unbeweisbare Tatsache:

#### Das Axiom des Archimedes:

$$\forall x \in \mathbb{R} \exists n \in \mathbb{N} \text{ mit } n > x.$$

Die natürlichen Zahlen wachsen über jede Schranke hinaus.

2. Durch  $a_n := (-1)^n$  erhält man die seltsame Folge

$$-1, +1, -1, +1, \dots$$

Sie ist offensichtlich beschränkt, es werden ja nur die beiden Werte  $-1$  und  $+1$  angenommen.

**Hinweis:** Ist  $(a_n)$  eine Folge, so kann man dazu die Menge der Folgenglieder bilden:  $M := \{a_n \mid n \in \mathbb{N}\}$ . Diese Menge ist *etwas ganz anderes* als die Folge selbst.

Die Folge in unserem Beispiel besitzt unendlich viele Glieder, die zugehörige Menge  $M = \{-1, +1\}$  besteht aber nur aus 2 Elementen. Der Grund dafür liegt in der genauen Mengendefinition, die ich am Anfang der Vorlesung unterschlagen habe. Man muß die Elemente einer Menge stets voneinander unterscheiden können. Kommt ein Objekt mehrfach als Element vor, so darf es nur einmal gezählt werden! Bei einer Folge sind die Glieder dagegen schon durch ihre Nummer voneinander zu unterscheiden, die Werte dürfen dann ruhig gleich sein.

3. Durch  $a_n := \frac{1}{n}$  wird die Folge der „Stammbrüche“  $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$  definiert. Diese Folge ist nach unten durch 0 und nach oben durch 1 beschränkt.

Die zuletzt betrachtete Folge hat eine bemerkenswerte Eigenschaft: Die Werte der Folgeglieder kommen der Null immer näher, ja, sogar beliebig nahe! Wie kann man das, was anschaulich klar ist, exakt formulieren? Tatsächlich wird der Wert 0 ja nie angenommen.

Stellen Sie sich vor, wir wären über das Problem in Streit geraten. Sie glauben von ihrer Anschauung her, daß die Folgeglieder der Null beliebig nahe kommen, ich dagegen bin skeptisch: „Wie nahe kommen die  $a_n$  der Null, etwa näher als  $1/1000$ ?“

„Sicher!“, sagen Sie. „Wenn  $n \geq 1001$  ist, dann ist  $a_n < \frac{1}{1001} < 1/1000$ .“

„Ja, aber kommen die  $a_n$  der Null auch näher als  $1/3\,000\,000$ ?“

„Natürlich, sobald  $n \geq 3\,000\,001$  ist.“

Und wenn wir nicht mit dem Streitgespräch aufhören, dann sind wir am Ende des Semesters immer noch dabei. Die Tatsache, daß Sie immer wieder eine passende Antwort finden, muß uns als Begründung reichen.

### Definition.

Eine Folge  $(a_n)$  konvergiert gegen die reelle Zahl  $a$ , falls gilt:

Zu jeder vorgegebenen Genauigkeit  $\varepsilon > 0$  gibt es eine Nummer  $n_0$ , so daß für alle weiteren Nummern  $n \geq n_0$  das  $a_n$  um weniger als  $\varepsilon$  von  $a$  entfernt ist.

In Formeln kann man das etwas kürzer ausdrücken:

$$\forall \varepsilon > 0 \exists n_0, \text{ so daß } \forall n \geq n_0 \text{ gilt: } |a - a_n| < \varepsilon.$$

Man nennt dann  $a$  den *Grenzwert* oder *Limes* der Folge und schreibt:

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

Wenn die Folge nicht konvergiert, heißt sie *divergent*.

Die Verschachtelung von „Quantoren“  $\forall \dots \exists \dots \forall \dots$  bereitet zu Anfang Schwierigkeiten, und mancher lernt's nie. Deshalb kommt hier noch eine weitere Formulierung der Konvergenz:

Wir sagen, *fast alle Folgeglieder*  $a_n$  liegen in einer Menge  $M$ , wenn es höchstens endlich viele Ausnahmen gibt. Geht man die Folgeglieder der Reihe nach durch, so hat man irgendwann alle Ausnahmen hinter sich gebracht und damit ein  $n_0$  erreicht, so daß für  $n \geq n_0$  alle  $a_n$  in  $M$  liegen. Mit dieser Sprechweise können wir sagen:

$(a_n)$  konvergiert genau dann gegen  $a$ , wenn gilt:

In jeder  $\varepsilon$ -Umgebung von  $a$  liegen fast alle  $a_n$ .

### I.2.1 Satz.

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0.$$

BEWEIS: Wir müssen nur feststellen, daß das Streitgespräch um die Genauigkeit der Annäherung beliebig lange weitergeführt werden kann.

Sei ein beliebiges  $\varepsilon > 0$  vorgegeben. Dann ist  $\frac{1}{\varepsilon}$  eine reelle Zahl, und nach dem Axiom des Archimedes gibt es eine natürliche Zahl  $n_0$ , die größer als  $\frac{1}{\varepsilon}$  ist. (Man bedenke, daß man  $\varepsilon$  sehr klein wählen wird, und daß dann  $\frac{1}{\varepsilon}$  eine sehr große Zahl ist).

Ist nun  $n \geq n_0$ , so ist

$$\left| \frac{1}{n} - 0 \right| = \frac{1}{n} \leq \frac{1}{n_0} < \varepsilon.$$

Also ist alles bestens! □

Es kommt manchmal vor, daß eine Folge zwar nicht konvergiert, sich aber dennoch einem bestimmten Wert beliebig annähert.

### Definition.

$a \in \mathbb{R}$  heißt *Häufungspunkt* der Folge  $(a_n)$ , falls gilt:

In jeder  $\varepsilon$ -Umgebung von  $a$  liegen unendlich viele Folgenglieder  $a_n$ .

### Beispiele :

1.  $1, 2, 3, 4, \dots$  hat keinen Häufungspunkt.
2.  $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$  hat die 0 als einzigen Häufungspunkt. Der ist zugleich der Limes.
3.  $a_n := (-1)^n = -1, +1, -1, +1, -1, \dots$  hat zwei Häufungspunkte, nämlich  $+1$  und  $-1$ . Sie konvergiert weder gegen  $+1$  noch gegen  $-1$ , weil sie sich immer wieder von diesen Punkten entfernt.

Man muß also sorgfältig unterscheiden, ob in einer Umgebung „nur“ unendlich viele oder sogar fast alle Folgenglieder liegen.

4. Die Folge

$$a_n := \begin{cases} n & \text{falls } n \text{ gerade} \\ \frac{1}{n} & \text{falls } n \text{ ungerade} \end{cases}$$

hat einen Häufungspunkt, nämlich die 0. Sie konvergiert aber nicht gegen 0, sie ist sogar unbeschränkt!

Ein Häufungspunkt ist also i.a. nicht der Grenzwert. Umgekehrt gilt aber offensichtlich:

*Wenn die Folge  $(a_n)$  gegen  $a \in \mathbb{R}$  konvergiert, dann ist  $a$  auch der einzige Häufungspunkt der Folge.*

Wir haben an früherer Stelle gelernt, daß es eine reelle Zahl  $d$  mit  $d^2 = 2$  gibt. Diese Zahl bezeichnen wir auch mit  $\sqrt{2}$ , und ihre Existenz haben wir geometrisch begründet. Außerdem haben wir gezeigt, daß  $\sqrt{2}$  keine rationale Zahl ist. Aber was ist  $\sqrt{2}$  wirklich?

In unserer Not greifen wir zum Taschenrechner, und der zeigt uns auf Knopfdruck:

$$\sqrt{2} = 1.4142136 \dots$$

Ein solches Ergebnis bleibt zwangsläufig immer etwas ungenau, aber es liefert uns auch Informationen:

$\sqrt{2}$  liegt im Intervall  $I_1 := [1, 2]$  (der Länge 1),

die Zahl liegt aber auch in den Intervallen

$I_2 := [1.4, 1.42]$  (der Länge 0.02),

$I_3 := [1.414, 1.415]$  (der Länge 0.001)

usw. Die Intervalle sind ineinander enthalten, und es sieht so aus, als ob man sie beliebig klein bekommen könnte.

**Definition.**

Eine Folge von Intervallen  $I_n := [a_n, b_n]$  mit

$$I_1 \supset I_2 \supset I_3 \supset \dots$$

heißt *Intervallschachtelung*, falls  $\lim_{n \rightarrow \infty} (b_n - a_n) = 0$  ist.

Es ist dann

$$\dots \leq a_n \leq a_{n+1} < b_{n+1} \leq b_n \leq \dots,$$

und daher  $b_n - a_n \geq b_{n+1} - a_{n+1} \geq \dots$ . Wenn diese Intervall-Längen zudem gegen Null konvergieren, dann kann man sich fragen, ob im Inneren aller Intervalle noch eine Zahl übrig bleibt. Z.B. ist  $I_n := [-\frac{1}{n}, +\frac{1}{n}]$  eine Intervallschachtelung, die den Nullpunkt als einzigen gemeinsamen Punkt enthält, und wir haben den Anfang einer Intervallschachtelung konstruiert, bei der alle Intervalle die Zahl  $\sqrt{2}$  enthalten. Daß letzteres auch mit unendlich vielen Schritten geht, besagt das

**Vollständigkeitsaxiom:**

Zu jeder Intervallschachtelung  $I_n := [a_n, b_n]$  gibt es genau eine reelle Zahl  $c$ , die in sämtlichen Intervallen  $I_n$  enthalten ist.

Dieses nicht beweisbare Vollständigkeitsaxiom besagt, daß es – so, wie es uns die geometrische Anschauung sagt – in  $\mathbb{R}$  keine Lücken gibt. Üblicherweise wird man auf diesem Wege versuchen, irrationale Zahlen durch aufeinander zulaufende rationale Folgen zu approximieren. Wir sprechen dann von einer *rationalen Intervallschachtelung*. Das Axiom sichert zu, daß es die approximierte Zahl gibt, und daß sie eindeutig bestimmt ist. Umgekehrt kann man natürlich um jede reelle Zahl  $c$  herum eine Intervallschachtelung konstruieren, die genau diese Zahl enthält, etwa  $[c - \frac{1}{n}, c + \frac{1}{n}]$ . Mit etwas mehr Mühe kann man auch erreichen, daß die Intervallgrenzen *rationale Zahlen* sind: Weil sich jede reelle Zahl  $c$  als unendlicher Dezimalbruch schreiben läßt, finden wir in jeder  $\varepsilon$ -Umgebung von  $c$  genügend rationale Zahlen.

Ein mathematisch sauberer Beweis für die Existenz von  $\sqrt{2}$  wird nun so geführt, daß man eine geeignete Intervallschachtelung angibt (wir haben bis jetzt ja nur den Anfang einer solchen kennengelernt). Allerdings wollen wir das hier nicht ausführen, und ich möchte auch nur erwähnen, daß man mit Hilfe von Intervallschachtelungen folgenden Satz beweisen kann:

**I.2.2 Satz.** Zu jeder **positiven** reellen Zahl  $a$  und jedem  $n \in \mathbb{N}$  gibt es genau eine positive Lösung der Gleichung  $x^n = a$ .

Die Lösung schreibt man in der Form  $x = \sqrt[n]{a} = a^{1/n}$ .

Ist  $q = \frac{m}{n}$  eine positive rationale Zahl und  $a > 0$  reell, so kann man

$$a^q := (a^m)^{1/n} = \sqrt[n]{a^m}$$

setzen und damit die Potenzen erweitern. Die bekannten Rechenregeln für Potenzen gelten nach wie vor, aber man muß die Bedingung  $a > 0$  beachten!

Setzen wir noch  $a^{-q} := \frac{1}{a^q}$ , so haben wir den Ausdruck  $a^x$  für jedes positive  $x \in \mathbb{R}$  und jedes rationale  $x \in \mathbb{Q}$  definiert. Das genügt uns aber immer noch nicht, wir wollen für den Exponenten  $x$  sogar beliebige reelle Zahlen zulassen. Und das geht folgendermaßen:

Wir betrachten zunächst den Fall  $a > 1$ . Ist  $x \in \mathbb{R}$  beliebig, so gibt es eine rationale Intervallschachtelung

$$I_1 \supset I_2 \supset I_3 \supset \dots \text{ mit } x \in I_k \text{ für alle } k.$$

Ist  $I_k = [p_k, q_k]$ , mit  $p_k, q_k \in \mathbb{Q}$ , so setzen wir  $I_k^* := [a^{p_k}, a^{q_k}]$ . Man kann nachrechnen, daß die  $I_k^*$  wieder eine Intervallschachtelung bilden (die i.a. nicht mehr rational ist!), und die dadurch approximierte reelle Zahl bezeichnen wir mit  $a^x$ .

Der Fall  $a = 1$  ist besonders einfach, natürlich wird stets  $1^x = 1$  gesetzt.

Ist  $0 < a < 1$ , so ist  $\frac{1}{a} > 1$  und man kann  $a^x := (\frac{1}{a})^{-x}$  setzen.

Wir fassen zusammen:

Definitionsgemäß ist  $0^0 = 1$ .

Ist  $a \in \mathbb{R}$ ,  $a \neq 0$  und  $n \in \mathbb{Z}$ , so ist  $a^n$  definiert.

Ist  $a \in \mathbb{R}_+$  und  $x \in \mathbb{R}$ , so ist ebenfalls  $a^x$  definiert.

Ist  $a \leq 0$  und  $x \notin \mathbb{Z}$ , so ist  $a^x$  i.a. nicht definiert! Es gibt aber Ausnahmen, wie etwa  $(-27)^{1/3} = -3$ .

Wenn  $a^x$  und  $a^y$  beide definiert sind, dann gilt:

$$\begin{aligned} a^{x+y} &= a^x \cdot a^y, \\ a^{x \cdot y} &= (a^x)^y = (a^y)^x, \\ a > 1 \text{ und } x \leq y &\implies a^x \leq a^y, \\ a < 1 \text{ und } x \leq y &\implies a^x \geq a^y. \end{aligned}$$

Wir kommen jetzt zu einem sehr wichtigen Satz über Folgen:

**I.2.3 Satz von Bolzano–Weierstraß.** *Jede beschränkte Folge  $(a_n)$  besitzt (mindestens) einen Häufungspunkt.*

BEWEIS: Da  $(a_n)$  beschränkt ist, gibt es reelle Zahlen  $\alpha_1, \beta_1$ , so daß

$$\alpha_1 \leq a_n \leq \beta_1 \text{ für alle } n \in \mathbb{N} \text{ ist.}$$

Wir halbieren das Intervall  $I_1 := [\alpha_1, \beta_1]$ . In wenigstens einer der beiden Hälften müssen unendlich viele Folgeglieder liegen. Die Grenzen dieser Intervallhälfte liefern uns reelle Zahlen  $\alpha_2, \beta_2$  mit  $\alpha_1 \leq \alpha_2 \leq \beta_2 \leq \beta_1$  und  $\beta_2 - \alpha_2 = \frac{1}{2} \cdot (\beta_1 - \alpha_1)$ , so daß unendlich viele  $a_n$  zwischen  $\alpha_2$  und  $\beta_2$  liegen.

Jetzt halbieren wir dieses Intervall wieder und finden  $\alpha_3, \beta_3$  mit  $\beta_3 - \alpha_3 = \frac{1}{4} \cdot (\beta_1 - \alpha_1)$ , so daß unendlich viele  $a_n$  dazwischen liegen. So fahren wir fort und erhalten eine Intervallschachtelung, bei der jedes Teilintervall unendlich viele Folgeglieder  $a_n$  enthält. Die reelle Zahl  $c$ , die dadurch approximiert wird, ist offensichtlich ein Häufungspunkt der Folge.  $\square$

**I.2.4 Satz.** *Eine Folge ist genau dann konvergent, wenn sie beschränkt ist und genau einen Häufungspunkt besitzt.*

*Der einzige Häufungspunkt ist auch der Grenzwert.*

BEWEIS: 1) Sei  $(a_n)$  konvergent gegen die reelle Zahl  $a$ ,  $\varepsilon > 0$  eine irgendwie gewählte Zahl. Da in  $U_\varepsilon(a)$  stets fast alle Folgeglieder liegen, ist  $a$  auch der einzige Häufungspunkt. Die verbliebenen endlich vielen Folgeglieder, die vielleicht nicht in  $U_\varepsilon(a)$  liegen, haben einen beschränkten Abstand von  $a$ . Also ist  $(a_n)$  auch beschränkt.

2) Sei umgekehrt  $(a_n)$  beschränkt und  $a$  der einzige Häufungspunkt. Sei  $\varepsilon > 0$  beliebig vorgegeben. Wir müssen zeigen, daß fast alle  $a_n$  in  $U_\varepsilon(a)$  liegen. Wir nehmen an, das sei nicht der Fall. Dann liegen unendlich viele  $a_n$  in  $\mathbb{R} \setminus U_\varepsilon(a)$ . Wenn wir diese unendlich vielen Folgeglieder durchnummerieren, dann erhalten wir eine „Teilfolge“, die natürlich auch beschränkt ist. Nach Bolzano–Weierstraß besitzt die Teilfolge einen Häufungspunkt  $b$ , und der muß um mindestens  $\varepsilon$  von  $a$  entfernt sein. Insbesondere ist  $a \neq b$ . Das ist ein Widerspruch, denn  $b$  ist auch Häufungspunkt der ursprünglichen Folge, und die sollte nur einen besitzen.  $\square$

**Beispiel:**

Sei  $a_n := (-1)^n \frac{n}{n+1}$ . Das ist die Folge

$$-\frac{1}{2}, \quad \frac{2}{3}, \quad -\frac{3}{4}, \quad \frac{4}{5}, \quad -\frac{5}{6}, \quad \dots$$

Für die positiven Terme  $a_{2n} = \frac{2n}{2n+1}$  gilt:

$$|1 - a_{2n}| = \left| \frac{2n+1}{2n+1} - \frac{2n}{2n+1} \right| = \frac{1}{2n+1}.$$

Da  $2n+1$  über alle Grenzen wächst, kommt  $\frac{1}{2n+1}$  der Null beliebig nahe, d.h.:

$$\forall \varepsilon > 0 \text{ liegen fast alle } a_{2n} \text{ in } U_\varepsilon(1).$$

Also ist 1 ein Häufungspunkt der Folge. Genauso zeigt man, daß auch  $-1$  ein Häufungspunkt ist. Aus dem gerade bewiesenen Satz folgt nun, daß  $(a_n)$  nicht konvergiert, also eine divergente Folge ist.

Im Beispiel ist zu sehen, daß es für jeden der beiden Häufungspunkte eine Teilfolge gibt, die gegen diesen Punkt konvergiert. Das ist ein allgemeingültiges Prinzip:

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

Wir haben gesehen, daß es schon bei relativ einfachen Folgen ziemlich mühsam ist, Häufungspunkte oder Grenzwerte zu bestimmen. Zum Glück haben die Mathematiker eine Reihe von Regeln zur Berechnung von Grenzwerten gefunden, die wir hier ohne Beweis übernehmen:

**I.2.5 Satz.** Die Folgen  $(a_n)$  bzw.  $(b_n)$  seien konvergent gegen  $a$  bzw.  $b$ . Dann gilt:

$$\begin{aligned} \lim_{n \rightarrow \infty} a_n \pm b_n &= a \pm b, \\ \text{und} \quad \lim_{n \rightarrow \infty} a_n \cdot b_n &= a \cdot b. \\ \text{Sind } b \text{ und alle } b_n &\neq 0, \text{ so ist} \\ \lim_{n \rightarrow \infty} \frac{a_n}{b_n} &= \frac{a}{b}. \end{aligned}$$

Ist  $(x_n)$  eine weitere Folge und  $a_n \leq x_n \leq b_n$  für alle  $n$ , so gelten folgende „Vergleichsregeln“:

1. Konvergiert  $(x_n)$  gegen  $x$ , so ist  $a \leq x \leq b$ .
2. Ist über  $(x_n)$  nichts bekannt, aber  $a = b$ , so folgt, daß  $(x_n)$  gegen  $a$  konvergiert.

Bisher sind wir mit dem Begriff „unendlich“ etwas zaghaft umgegangen. Wir wollen uns jetzt das Leben erleichtern:

Die *erweiterte Zahlengerade*  $\overline{\mathbb{R}}$  besteht aus der Menge  $\mathbb{R}$  und zwei weiteren Elementen  $-\infty$  und  $+\infty$ , die nicht zu  $\mathbb{R}$  gehören und auch voneinander verschieden sind:

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}.$$

Die Rechenregeln von  $\mathbb{R}$  dürfen nicht auf  $-\infty$  und  $+\infty$  angewandt werden, aber es gilt folgende Ordnungs-Beziehung:

$$\forall x \in \mathbb{R} : \quad -\infty < x < +\infty.$$

Insbesondere soll das auch bedeuten, daß  $-\infty < +\infty$  ist.

Wir definieren jetzt:

$$\begin{aligned} \lim_{n \rightarrow \infty} a_n = +\infty & : \iff \forall c \in \mathbb{R} \text{ gilt: Fast alle } a_n \text{ sind } < c. \\ \lim_{n \rightarrow \infty} a_n = -\infty & : \iff \forall c \in \mathbb{R} \text{ gilt: Fast alle } a_n \text{ sind } < c. \end{aligned}$$

**Beispiele :**

1.  $a_n := \frac{3n}{n+1} = \frac{3}{1+\frac{1}{n}}$  konvergiert gegen  $\frac{3}{1+0} = 3$ .
2.  $a_n := \frac{n(n-2)}{5n^2+3} = \frac{n^2-2n}{5n^2+3} = \frac{1-2\cdot\frac{1}{n}}{5+3\cdot\frac{1}{n}\cdot\frac{1}{n}}$  konvergiert gegen  $\frac{1}{5}$ .

Der Trick ist immer der gleiche: Man kürzt die höchste Potenz von  $n$  heraus. Das geht allerdings nur so lange gut, wie diese höchste Potenz im Nenner auftritt. Enthält der Zähler  $n$  in höherer Potenz als der Nenner, so passiert z.B. folgendes:

$$a_n := \frac{n^3-1}{5+n} = \frac{n^2-\frac{1}{n}}{\frac{5}{n}+1} \text{ konvergiert gegen } +\infty.$$

3. Nun zu etwas komplizierteren Beispielen:

Sei  $b \geq 1$  und  $a_n := \sqrt[n]{b}$ . Da  $a_n \geq 1$  ist, können wir schreiben:

$$a_n = 1 + h_n, \text{ mit } h_n \geq 0.$$

Dann ist  $b = (a_n)^n = (1 + h_n)^n \geq 1 + n \cdot h_n$ , also

$$0 \leq h_n \leq \frac{b-1}{n}.$$

Daraus folgt, daß  $\lim_{n \rightarrow \infty} h_n = 0$  und  $\lim_{n \rightarrow \infty} a_n = 1$  ist.

4. Den Trick von eben sollte man sich merken, denn jetzt wird's noch schwieriger:

Sei  $a_n := \sqrt[n]{n}$ . Eine Vorhersage über das Ergebnis ist nicht so ohne weiteres möglich. Wir gehen wie beim vorigen Beispiel an die Sache heran, benutzen aber noch einen weiteren Trick:

Da  $a_n > 1$  ist, ist  $b_n := \sqrt{a_n} = 1 + h_n$  mit  $h_n > 0$ .

Dabei ist  $b_n = \sqrt[n]{\sqrt{n}}$ , also

$$\sqrt{n} = (b_n)^n = (1 + h_n)^n \geq 1 + n \cdot h_n.$$

Damit ist

$$h_n \leq \frac{\sqrt{n}-1}{n} = \frac{1-1/\sqrt{n}}{\sqrt{n}} \leq \frac{1}{\sqrt{n}},$$

also

$$1 < a_n = (b_n)^2 = (1 + h_n)^2 = 1 + 2h_n + (h_n)^2 \leq 1 + \frac{2}{\sqrt{n}} + \frac{1}{n}.$$

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

5. Wir betrachten nun die Folge  $q^n$  für  $q \in \mathbb{R}_+$ .

- (a) Ist  $q > 1$ , also  $q = 1 + h$  mit  $h > 0$ , so ist  $q^n = (1 + h)^n \geq 1 + n \cdot h$ , und dieser Ausdruck konvergiert gegen  $+\infty$ .

(b) Ist  $q = 1$ , so ist  $q^n = 1$  eine konstante Folge.

(c) Sei  $0 < q < 1$ . Dann ist  $\frac{1}{q} > 1$ , also  $\left(\frac{1}{q}\right)^n \geq 1 + n \cdot h$  nach oben unbeschränkt.

Daraus folgt:

$$0 < q^n \leq \frac{1}{1 + n \cdot h} \text{ für alle } n.$$

Also ist in diesem Falle  $\lim_{n \rightarrow \infty} q^n = 0$ .

Eine ganz andere Klasse von Beispielen erhält man folgendermaßen:

$$\text{Sei } a_n := \sum_{i=0}^n q^i = \frac{q^{n+1} - 1}{q - 1} \quad (\text{für } q \neq 1).$$

Wir betrachten hier nur den Fall  $q > 0$ :

1. Ist  $q > 1$ , so ist  $q^n$  und damit erst recht  $a_n = 1 + q + q^2 + \dots + q^n$  unbeschränkt.

2. Ist  $0 < q < 1$ , so strebt  $q^{n+1} = q \cdot q^n$  gegen Null, also

$$a_n \text{ gegen } \frac{1}{q-1}.$$

Es ist

$$\begin{aligned} a_1 &= 1 + q, \\ a_2 &= 1 + q + q^2, \\ a_3 &= 1 + q + q^2 + q^3, \\ &\vdots \end{aligned}$$

Im Grenzwert werden unendlich viele Terme addiert. Man spricht daher von einer *unendlichen Reihe* und schreibt

$$\lim_{n \rightarrow \infty} a_n = \sum_{i=0}^{\infty} q^i = \frac{1}{1 - q}.$$

Wir werden zu einem späteren Zeitpunkt noch viele andere Reihen kennenlernen, das hier betrachtete Beispiel ist die sogenannte *geometrische Reihe*. Sie kommt in der Mathematik sehr häufig vor.

### Beispiele:

1. Wir betrachten einen unendlichen Dezimalbruch, z.B.

$$0.9999 \dots = 9 \cdot 10^{-1} + 9 \cdot 10^{-2} + 9 \cdot 10^{-3} + \dots$$

Einen Sinn bekommt das erst, wenn wir schreiben:

$$\begin{aligned} 0.9999 \dots &= \lim_{n \rightarrow \infty} \sum_{i=1}^n 9 \cdot \left(\frac{1}{10}\right)^i \\ &= 9 \cdot \lim_{n \rightarrow \infty} \sum_{i=1}^n \left(\frac{1}{10}\right)^i \\ &= 9 \cdot \left(\frac{1}{1 - 1/10} - 1\right) \\ &= 9 \cdot \left(\frac{10}{9} - \frac{9}{9}\right) = 1. \end{aligned}$$

2. Es ist

$$\sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^i = \frac{1}{1 - \frac{1}{2}} = 2,$$

also  $\sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = 1$ . Das liefert die wundersame Formel

$$\boxed{\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 1.}$$

Sie zeigt, warum Achilles bei seinem Wettlauf mit der Schildkröte doch gewonnen hat.

Eine andere wichtige Klasse von Folgen sind die monotonen Folgen:

**Definition.**

Eine Folge  $(a_n)$  heißt *monoton wachsend* (bzw. *monoton fallend*), falls gilt:

$$a_n \leq a_{n+1} \quad (\text{bzw. } a_n \geq a_{n+1}) \quad \text{für alle } n \in \mathbb{N}.$$

Z.B. ist die Folge  $a_n := \frac{1}{n}$  monoton fallend, während die Folge  $a_n := \sum_{i=0}^n q^i$  (mit  $q > 0$ ) monoton wachsend ist.

**I.2.6 Satz von der monotonen Konvergenz.** *Die Folge  $(a_n)$  sei monoton wachsend und nach oben beschränkt. Dann ist sie konvergent.*

**BEWEIS:** Es gibt nach Voraussetzung eine reelle Zahl  $c$ , so daß  $a_1 \leq a_n \leq c$  für alle  $n \in \mathbb{N}$  ist. Also ist  $(a_n)$  beschränkt und besitzt nach dem Satz von Bolzano-Weierstraß einen Häufungspunkt  $a$ . Wir wollen zeigen, daß es der einzige Häufungspunkt und damit schon der Grenzwert ist.

Wenn es ein  $n_0 \in \mathbb{N}$  mit  $a_{n_0} > a$  gäbe, dann wäre wegen der Monotonie sogar  $a_n > a$  für alle  $n \geq n_0$ , also für fast alle  $n$ . Wählt man dann ein  $\varepsilon$  zwischen 0 und  $a_{n_0} - a$ , so können höchstens endlich viele  $a_n$  in  $U_\varepsilon(a)$  liegen. Das widerspricht der Tatsache, daß  $a$  Häufungspunkt ist.

Also ist  $a_n \leq a$  für alle  $n \in \mathbb{N}$ . Eine Zahl  $b > a$  hat dann keine Chance mehr, Häufungspunkt von  $(a_n)$  zu sein. Wenn es aber ein  $b < a$  gäbe, das Häufungspunkt ist, dann müßte – nach der gleichen Argumentation wie eben – schon  $a_n \leq b$  für alle  $n \in \mathbb{N}$  sein, und die Zahl  $a$  könnte kein Häufungspunkt der Folge sein. Also besitzt  $(a_n)$  nur einen einzigen Häufungspunkt und ist damit beschränkt!  $\square$

Der Satz von der monotonen Konvergenz ist ein äußerst nützliches Kriterium, um die Konvergenz einer Folge zu beweisen. Ist die Konvergenz erst einmal gesichert, dann gibt es oftmals Mittel und Wege, auch noch den Grenzwert herauszubekommen.

**Beispiel:**

Es sei  $a_1 := \sqrt{2}$ ,  $a_2 := \sqrt{2 + \sqrt{2}}$ ,  $a_3 := \sqrt{2 + \sqrt{2 + \sqrt{2}}}$  und allgemein  $a_{n+1} := \sqrt{2 + a_n}$ .

Das ist eine *rekursive Definition*. Daß auf diese Weise alle Folgenglieder bestimmt sind, ergibt sich unmittelbar aus dem Prinzip der vollständigen Induktion. Rekursive Formeln lassen sich besonders gut mit dem Computer bearbeiten.

Wir wollen zeigen, daß  $(a_n)$  monoton wachsend und nach oben durch 2 beschränkt ist.

Wegen der rekursiven Definition müssen wir fast zwangsläufig das Induktionsprinzip verwenden: Offensichtlich ist  $a_1 = \sqrt{2} < 2$ , und da  $2 < 2 + \sqrt{2}$  ist, ist  $a_1 = \sqrt{2} < \sqrt{2 + \sqrt{2}} = a_2$ . Damit ist der Induktionsanfang erledigt.

1. Ist nach Induktionsvoraussetzung  $a_n < 2$ , so ist  $a_n + 2 < 4$ , und damit  $a_{n+1} = \sqrt{a_n + 2} < \sqrt{4} = 2$ . Das war ganz einfach.
2. Bei der Monotonie lautet die Induktionsvoraussetzung  $a_n < a_{n+1}$ . Aber dann ist  $2 + a_n < 2 + a_{n+1}$ , und damit

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

Damit ist auch hier der Induktionsschluß vollzogen.

Aus dem Satz von der monotonen Konvergenz folgt jetzt, daß  $(a_n)$  konvergent ist. Mit einem einfachen Trick ermitteln wir den Grenzwert:

Mit  $(a_n)$  ist natürlich auch  $(b_n)$  mit  $b_n := a_{n+1}$  konvergent. Beide Folgen konvergieren gegen den gleichen Grenzwert, eine reelle Zahl  $a$ . Nun ist  $b_n = \sqrt{2 + a_n}$ , also

$$a^2 = \left(\lim_{n \rightarrow \infty} b_n\right)^2 = \lim_{n \rightarrow \infty} (b_n)^2 = \lim_{n \rightarrow \infty} (2 + a_n) = 2 + a.$$

Das bedeutet, daß  $a^2 - a - 2 = 0$  ist, also  $a = \frac{1 \pm \sqrt{1 + 2 \cdot 4}}{2} = \frac{1 \pm 3}{2}$ . Da alle Folgenglieder  $\geq 0$  sind, muß auch  $a \geq 0$  sein, also  $a = 2$ .

Gelegentlich kann man auch neue (irrationale) Zahlen als Grenzwerte von Folgen einführen:

Es sei

$$a_n := \sum_{\nu=0}^n \frac{1}{\nu!} = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \cdots + \frac{1}{n!}.$$

Es handelt sich mal wieder um eine Reihe, diesmal aber keine geometrische Reihe. Offensichtlich ist  $(a_n)$  monoton wachsend. Da die einzelnen Summanden ab einer gewissen Stelle stets schneller schrumpfen als die Summanden der geometrischen Reihe, dürfen wir vermuten, daß  $(a_n)$  konvergiert.

Tatsächlich ist

$$\frac{1}{\nu!} = \frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{\nu} < \frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2} = \frac{1}{2^{\nu-1}},$$

also

$$\begin{aligned}
 a_n &< 1 + 1 + \sum_{\nu=2}^n \left(\frac{1}{2}\right)^{\nu-1} \\
 &= 1 + \sum_{\mu=0}^{n-1} \left(\frac{1}{2}\right)^{\mu} \\
 &= 1 + \frac{1 - \left(\frac{1}{2}\right)^n}{1 - \frac{1}{2}} \\
 &< 1 + 2 = 3.
 \end{aligned}$$

Als beschränkte monotone Folge ist  $(a_n)$  konvergent, und der Grenzwert

$$e := \lim_{n \rightarrow \infty} = \sum_{\nu=0}^{\infty} \frac{1}{\nu!}$$

heißt *Eulersche Zahl*.

Man sieht sofort, daß  $2 < e \leq 3$  ist. Eine genauere Analyse ergibt

$$e = 2.718281\dots,$$

und man kann relativ leicht zeigen, daß  $e$  irrational ist.

Als nächstes untersuchen wir die Folge

$$b_n := \left(1 + \frac{1}{n}\right)^n \quad (= 2, 2.25, 2.37\dots, \dots).$$

Mit der binomischen Formel erhalten wir:

$$\begin{aligned}
 b_n &= \sum_{k=0}^n \binom{n}{k} \frac{1}{n^k} = \sum_{k=0}^n \frac{1}{k!} \cdot \frac{(n-k+1) \cdot \dots \cdot (n-1)n}{n^k} \\
 &= \sum_{k=0}^n \frac{1}{k!} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \dots \cdot \frac{n-k+1}{n} \\
 &= \sum_{k=0}^n \frac{1}{k!} \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right) \\
 &< \sum_{k=0}^n \frac{1}{k!} < e.
 \end{aligned}$$

Also ist  $(b_n)$  nach oben beschränkt durch die Eulersche Zahl. Weiter gilt:

$$\begin{aligned}
 b_{n+1} &= \left(1 + \frac{1}{n+1}\right)^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} \frac{1}{(n+1)^k} \\
 &= \sum_{k=0}^{n+1} \frac{1}{k!} \cdot \frac{(n+1)n(n-1) \cdot \dots \cdot (n-k+2)}{(n+1)^k} \\
 &\geq \sum_{k=0}^n \dots
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^n \frac{1}{k!} \cdot \frac{n(n-1) \cdot \dots \cdot (n-(k-2))}{(n+1)^{k-1}} \\
&= \sum_{k=0}^n \frac{1}{k!} \cdot \frac{(n+1)-1}{n+1} \cdot \frac{(n+1)-2}{n+1} \cdot \dots \cdot \frac{(n+1)-(k-1)}{n+1} \\
&= \sum_{k=0}^n \frac{1}{k!} \cdot \left(1 - \frac{1}{n+1}\right) \left(1 - \frac{2}{n+1}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n+1}\right) \\
&\geq \sum_{k=0}^n \frac{1}{k!} \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right) \\
&= b_n.
\end{aligned}$$

Damit ist  $(b_n)$  monoton wachsend, also konvergent gegen eine Zahl  $e^*$ . Um den Grenzwert näher zu bestimmen, brauchen wir einen Trick! Ist  $m \geq 2$  irgend eine feste natürliche Zahl und  $n > m$ , so gilt:

$$\begin{aligned}
b_n &= \sum_{k=0}^n \frac{1}{k!} \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right) \\
&\geq \sum_{k=0}^m \frac{1}{k!} \cdot \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdot \dots \cdot \left(1 - \frac{k-1}{n}\right).
\end{aligned}$$

Läßt man  $n \rightarrow \infty$  gehen, so erhält man:

$$e^* = \lim_{n \rightarrow \infty} b_n \geq \sum_{k=0}^m \frac{1}{k!}, \text{ für jedes } m \geq 2.$$

Also ist schließlich  $e^* \geq e$ , und damit

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^{\infty} \frac{1}{k!}.$$

So haben wir zwei verschiedene Darstellungen der Eulerschen Zahl gewonnen.

### §3 Paare und Abbildungen

$A$  und  $B$  seien zwei Mengen. Dann kann man je ein Element  $x \in A$  und ein Element  $y \in B$  zu einem (*geordneten*) *Paar*  $(x, y)$  zusammenfassen. Man nennt dann  $x$  die *erste Komponente* und  $y$  die *zweite Komponente* des Paares  $(x, y)$ . Es kommt auf die Reihenfolge an, d.h., es gilt:

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

Die Menge aller möglichen Paare nennt man die *Paarmenge* oder das *kartesische Produkt* von  $A$  und  $B$  und bezeichnet sie mit  $A \times B$ . Es gilt also:

$$A \times B = \{(x, y) \mid x \in A \text{ und } y \in B\}.$$

**Beispiel:**

Es sei

$$A := \{\diamond, \heartsuit, \spadesuit, \clubsuit\}$$

die Menge der Farben beim Skat, und

$$B := \{7, 8, 9, 10, \text{Bube, Dame, König, As}\}$$

die Menge der möglichen Werte. Dann ist  $A \times B$  die Menge der Karten eines Skat-spieles, ein Element ist z.B. das Paar  $(\heartsuit, \text{As})$ , das „Herz-As“.

Das kartesische Produkt von  $A$  mit sich selbst bezeichnet man auch mit  $A^2$ . Hier muß besonders auf die Reihenfolge der Komponenten geachtet werden. Sind  $x, y$  zwei verschiedene Elemente von  $A$ , so ist zwar  $\{x, y\} = \{y, x\}$ , aber  $(x, y) \neq (y, x)$  !

Die Menge  $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$  wird auch als *Euklidische Ebene* bezeichnet. Sie ist ein Modell für die Menge der Punkte der Anschauungsebene.

Ist  $P_0 = (x_0, y_0) \in \mathbb{R}^2$  ein solcher Punkt, so nennt man  $x_0$  die *x-Koordinate* oder *Abzisse* und  $y_0$  die *y-Koordinate* oder *Ordinate* von  $P_0$ . Der Punkt  $O := (0, 0)$  heißt der *Koordinatenursprung*.

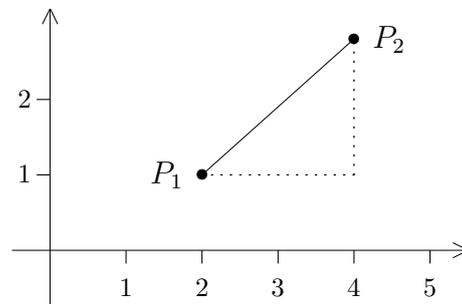
Sind  $P_1 := (x_1, y_1)$  und  $P_2 := (x_2, y_2)$  zwei Punkte in der Ebene, so ist die Verbindungsstrecke von  $P_1$  und  $P_2$  die Hypotenuse eines rechtwinkligen Dreiecks mit den Katheten  $|y_2 - y_1|$  und  $|x_2 - x_1|$ . Sie hat also die Länge

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

Man nennt diese Zahl auch den *Abstand* oder die *Distanz* von  $P_1$  und  $P_2$ .

Ist  $P_0 = (x_0, y_0)$  ein Punkt im  $\mathbb{R}^2$  und  $r > 0$  eine reelle Zahl, so nennt man die Menge

$$\begin{aligned} D_r(P_0) &:= \{P \in \mathbb{R}^2 \mid d(P, P_0) < r\} \\ &= \{(x, y) \in \mathbb{R}^2 \mid (x - x_0)^2 + (y - y_0)^2 < r^2\} \end{aligned}$$



die (*offene*) Kreisscheibe um  $P_0$  mit Radius  $r$ . Ersetzt man „ $<$ “ durch „ $\leq$ “, so spricht man von einer *abgeschlossenen Kreisscheibe*. Die Kreisscheiben spielen im  $\mathbb{R}^2$  die Rolle, die auf der Zahlengeraden die Intervalle innehaben. So konvergiert z.B. eine Folge  $(P_n)$  von Punkten der Ebene genau dann gegen einen Punkt  $P_0$ , wenn in jeder Kreisscheibe  $D_\varepsilon(P_0)$  fast alle Folgeglieder liegen.

Ist  $F(x, y)$  ein Term, der  $x$  und  $y$  als Variable enthält, so nennt man die Menge

$$K = \{(x, y) \in \mathbb{R}^2 \mid F(x, y) = 0\}$$

eine *ebene Kurve*. Dahinter kann sich vielerlei verbergen:

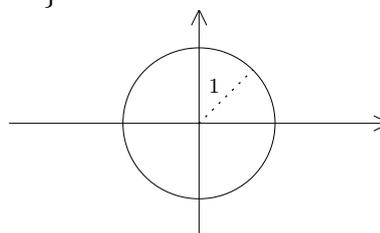
### Beispiele:

1.  $S^1 := \{(x, y) \mid x^2 + y^2 = 1\} = \{P \in \mathbb{R}^2 \mid d(P, O) = 1\}$

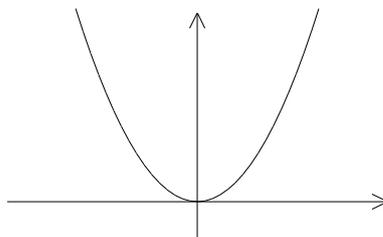
ist der sogenannte *Einheitskreis*. Allgemeiner ist

$$K_r(P_0) = \{P \in \mathbb{R}^2 \mid d(P, P_0) = r\}$$

der Kreis um  $P_0$  mit Radius  $r$ .



2.  $N := \{(x, y) \in \mathbb{R}^2 \mid y = x^2\}$  ist die „Normal-Parabel“.



3. Wenn man Pech hat, kommt gar keine Linie heraus. So ist z.B.

$$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 0\} = \{(0, 0)\}$$
 nur der Nullpunkt.

4. Das Wort „Kurve“ soll keineswegs mit einer gekrümmten Linie gleichgesetzt werden. Unser wichtigstes Beispiel ist die allgemeine *Gerade*

$$L = \{(x, y) \in \mathbb{R}^2 \mid ax + by = s\},$$

mit reellen Zahlen  $a, b, s$  und  $(a, b) \neq (0, 0)$ . Auch das ist ein Beispiel für eine Kurve.

Wir wollen die Geraden in der Ebene etwas näher untersuchen:

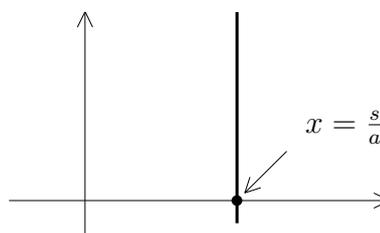
Dazu unterscheiden wir zwei Fälle:

1. Ist  $b = 0$  und  $a \neq 0$ ,

so ist

$$L = \{(x, y) \in \mathbb{R}^2 \mid x = \frac{s}{a}\}$$

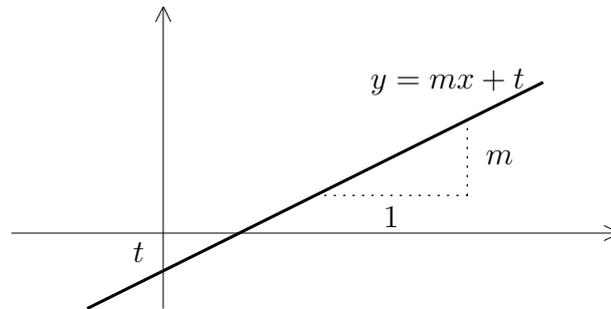
eine „vertikale Gerade“.



2. Ist  $b \neq 0$ , so kann man die Geraden-Gleichung  $ax + by = s$  nach  $y$  auflösen. Wir setzen  $m := -\frac{a}{b}$  und  $t := \frac{s}{b}$ . Dann ist

$$L = \{(x, y) \in \mathbb{R}^2 \mid y = mx + t\}$$

eine „schräge Gerade“:



Wenn sich die Abszisse von  $x$  um eine Einheit auf  $x + 1$  erhöht, dann erhöht sich die Ordinate von  $y = mx + t$  auf  $m(x + 1) + t = y + m$ , also um  $m$  Einheiten. Daher nennt man  $m$  die „Steigung“ der Geraden.

- (a) Ist  $m > 0$ , so liegt eine *aufsteigende Gerade* vor.
- (b) Ist  $m < 0$ , so liegt eine *fallende Gerade* vor.
- (c) Ist  $m = 0$ , so verläuft die Gerade horizontal. Wir fassen das dennoch als Spezialfall einer schrägen Geraden auf.

Meistens werden Geraden durch zwei Punkte definiert:

Zunächst betrachten wir die schräge Gerade mit der Steigung  $m$  durch einen Punkt  $P_1 = (x_1, y_1)$ :

$$\begin{aligned} \text{Allgemein ist } y &= mx + t, \\ \text{und speziell } y_1 &= mx_1 + t, \\ \text{insgesamt also } y - y_1 &= m(x - x_1). \end{aligned}$$

Das ergibt die Geradengleichung

$$\boxed{y = y_1 + m(x - x_1)}.$$

Soll die Gerade zusätzlich durch  $P_2 = (x_2, y_2)$  gehen, so muß  $y_2 = y_1 + m(x_2 - x_1)$  sein, also

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

Einsetzen ergibt die Geradengleichung

$$\boxed{y = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)}.$$

Man kann das noch in eine symmetrischere Form bringen:

$$\frac{y - y_1}{x - x_1} = \frac{y_2 - y_1}{x_2 - x_1}.$$

Man spricht dann von der SZwei-Punkte-Form“.

Als nächstes suchen wir nach dem Schnittpunkt zweier Geraden

$$\begin{aligned} L_1 &= \{ax + by = s\} \\ \text{und } L_2 &= \{cx + dy = r\}. \end{aligned}$$

Wir müssen mehrere Fälle unterscheiden:

1.  $L_1 = \{(x, y) \mid x = p\}$  und  $L_2 = \{(x, y) \mid x = q\}$  sind beide vertikal. Dann ist

$$L_1 \cap L_2 = \begin{cases} L_1 & \text{falls } p = q \\ \emptyset & \text{falls } p \neq q. \end{cases}$$

2.  $L_1 = \{(x, y) \mid x = p\}$  ist vertikal und  $L_2 = \{(x, y) \mid y = mx + t\}$  ist schräg. Dann ist offensichtlich

$$L_1 \cap L_2 = \{(p, mp + t)\}$$

ein einziger Punkt.

Ein analoges Ergebnis erhält man, wenn  $L_1$  schräg und  $L_2$  vertikal ist.

3. Der interessanteste Fall liegt vor, wenn beide Geraden schräg sind. Sei  $L_1 = \{(x, y) \mid y = mx + t\}$  und  $L_2 = \{(x, y) \mid y = nx + u\}$ . Ist  $m = n$ , so haben die Geraden die gleiche Richtung, sind also *parallel*. Dann muß auch noch  $t = u$  sein, damit die Geraden zusammenfallen, andernfalls sind sie *disjunkt* (d.h. sie haben einen leeren Durchschnitt).

Ist  $m \neq n$ , so kann man die Gleichung  $(m - n)x + (t - u) = 0$  nach  $x$  auflösen und erhält genau einen Schnittpunkt, nämlich

$$(x, y) = \left(-\frac{t - u}{m - n}, \frac{mu - nt}{m - n}\right).$$

Man kann den Schnittpunkt auch bestimmen, wenn die Geraden nicht in der Steigungsform gegeben sind. Das führt auf ein Gleichungssystem der folgenden Art:

$$\begin{aligned} (I) \quad ax + by &= s, \\ \text{und } (II) \quad cx + dy &= r. \end{aligned}$$

So etwas nennt man ein „Lineares Gleichungssystem“.

Subtrahiert man das  $b$ -fache der zweiten Gleichung vom  $d$ -fachen der ersten Gleichung, so erhält man:

$$(ad - bc)x = ds - br.$$

Subtrahiert man das  $c$ -fache der ersten Gleichung vom  $a$ -fachen der zweiten Gleichung, so erhält man:

$$(ad - bc)y = ar - sc.$$

Die Zahl  $d := ad - bc$  heißt die *Determinante* des Gleichungssystems. Sie entscheidet über die Lösbarkeit:

1. Ist  $d \neq 0$ , so gibt es genau eine Lösung, nämlich

$$x = \frac{ds - br}{d},$$

und

$$y = \frac{ar - sc}{d}.$$

2. Ist dagegen  $d = 0$ , so ist  $ad = bc$ . Ist außerdem  $ds = br$  und  $ar = cs$ , so gibt es unendlich viele Lösungen, d.h. die beiden Geraden fallen zusammen. Ist etwa  $b \neq 0$ , so ergibt sich die zweite Gleichung aus der ersten durch Multiplikation mit  $\frac{d}{b}$ :

$$c = a \cdot \frac{d}{b}, \quad d = b \cdot \frac{d}{b} \quad \text{und} \quad r = s \cdot \frac{d}{b}.$$

Ist  $b = 0$ , so muß  $a \neq 0$  sein, und die zweite Gleichung ergibt sich wieder aus der ersten, diesmal durch Multiplikation mit  $\frac{c}{a}$ :

$$c = a \cdot \frac{c}{a}, \quad d = b \cdot \frac{c}{a} \quad \text{und} \quad r = s \cdot \frac{c}{a}.$$

Eine andere Situation liegt vor, wenn  $ds \neq br$  oder  $ar \neq cs$  ist. Dann wird das Gleichungssystem von keinem  $(x, y)$  erfüllt, d.h. die Geraden sind disjunkt und zueinander parallel. In diesem Fall ist zwar auch die linke Seite der zweiten Gleichung ein Vielfaches der linken Seite der ersten Gleichung, aber die rechten Seiten passen nicht zueinander.

Soweit die Theorie der Geraden in der Ebene. In Kapitel II werden wir Lineare Gleichungssysteme sehr viel allgemeiner behandeln.

### Definition.

$A$  und  $B$  seien zwei beliebige Mengen. Unter einer *Abbildung von  $A$  nach  $B$*  versteht man eine Vorschrift, durch die **jedem Element  $x \in A$  genau ein Element  $y \in B$**  zugeordnet wird.

Die Zuordnungsvorschrift sei mit dem Buchstaben  $f$  bezeichnet.

Dann schreibt man:

$$\boxed{f : A \rightarrow B}$$

Wird dem Element  $x \in A$  das Element  $x \in B$  zugeordnet, so drückt man das durch  $f(x) = y$  oder  $f : x \mapsto y$  aus.

Die Menge  $A$  nennt man den *Definitionsbereich* von  $f$ . Sie ist wesentlicher Bestandteil der Abbildung  $f$  und wird daher auch gerne mit  $D_f$  bezeichnet. Die Menge  $B$  nennt man *Wertebereich*, *Wertevorrat* oder *Zielmenge*.

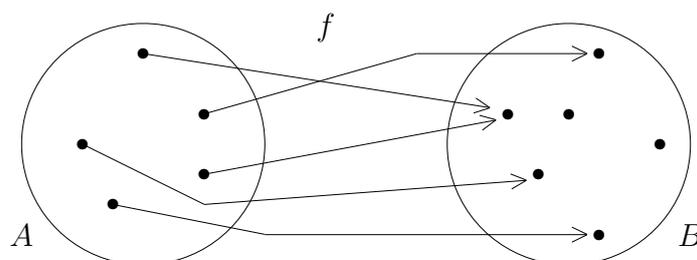
**Beispiel:**

Es sei  $A$  die Menge aller Postsendungen, die ein Briefträger am Morgen zum Austragen erhält, und  $B$  die Menge aller Anwohner seines Zustellbezirkes. Die Zuordnung  $f: A \rightarrow B$  wird durch die Empfänger-Daten auf den Postsendungen definiert. Wir nehmen einmal an, daß diese Daten alle leserlich und korrekt sind, und daß keine Einschreiben oder Nachnahmen dabei sind.

Am Ende seiner Tour hat der Briefträger keine Sendung mehr in der Tasche, es wurde also jeder Sendung ein Empfänger zugeordnet. Und natürlich wurde kein Brief zerrissen und an zwei Leute verteilt. Es liegt also tatsächlich eine „Abbildung“ von  $A$  nach  $B$  vor.

Es ist dabei möglich, daß ein Empfänger mehrere Briefe bekommt, und genauso, daß ein anderer Empfänger leer ausgeht.

In diesem Anfangsstadium können wir eine Abbildung durch ein Pfeildiagramm veranschaulichen:



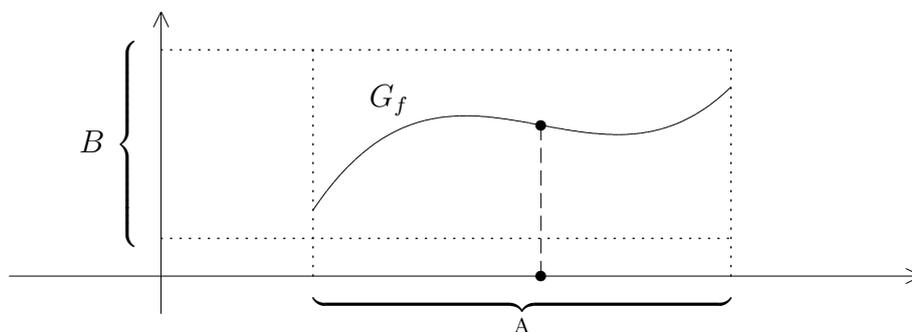
Sobald es um große oder gar unendliche Mengen geht, versagt diese Methode der Veranschaulichung, wir müssen uns nach anderen Hilfsmitteln umsehen.

**Definition.**

Ist  $f: A \rightarrow B$  eine Abbildung, so versteht man unter dem *Graphen* von  $f$  die Menge

$$G_f := \{(x, y) \in A \times B \mid y = f(x)\} = \{(x, f(x)) \mid x \in A\}.$$

Ist etwa  $A \subset \mathbb{R}$  und  $B \subset \mathbb{R}$ , so kann man den Graphen sehr einfach aufzeichnen:



Die typischen Eigenschaften einer Abbildung sind leicht zu erkennen:

1. Da jedem  $x \in A$  etwas zugeordnet werden soll, muß jede vertikale Gerade, die  $A$  trifft, auch den Graphen treffen.

2. Da jedem  $x$  nur ein  $y$  zugeordnet werden kann, darf eine vertikale Gerade den Graphen auch stets nur einmal treffen.

Ist der Wertebereich einer Abbildung eine Menge von Zahlen, so sagt man an Stelle von „Abbildung“ meistens „Funktion“. Das ist allerdings nur eine Tradition, keine feste Vorschrift. Und der Begriff „Zahl“ kann dabei recht weit ausgelegt werden.

Der Graph einer reellwertigen Funktion  $f$  mit  $D_f \subset \mathbb{R}$  ist eine Teilmenge von  $\mathbb{R}^2$  der Gestalt  $\{(x, y) \mid y - f(x) = 0\}$ , also eine Kurve. Umgekehrt ist jedoch noch lange nicht jede Kurve im  $\mathbb{R}^2$  der Graph einer Funktion, wie etwa das Beispiel eines Kreises zeigt.

### Definition.

Sei  $f : A \rightarrow B$  eine Abbildung. Ist  $M \subset A$  eine Teilmenge, so nennt man die Menge

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

das *Bild* (oder die *Bildmenge*) von  $M$  unter  $f$ .

Im Briefträger-Beispiel sei z.B.  $M$  die Menge aller derjenigen Sendungen, auf denen als Empfänger „Franz Xaver Hintermoser“ angegeben ist. Die Bildmenge  $f(M)$  wird dann voraussichtlich nur aus einem einzigen Element bestehen. Anders sieht die Situation aus, wenn man für  $M$  die Menge aller Postkarten nimmt. Schließlich ist es auch möglich,  $M = A$  zu wählen. In unserem Beispiel ist  $f(A)$  die Menge der Bewohner, die Post bekommen. Das ist i.a. etwas anderes als die Zielmenge  $B$ !

### Beispiele:

1.  $f : \mathbb{R} \rightarrow \mathbb{R}$  sei definiert durch  $f(x) := mx + t$ , mit  $m \neq 0$ . Das ist eine „Lineare Funktion“, und der Graph ist eine schräge Gerade. Eine vertikale Gerade kann nicht der Graph einer Funktion sein. Warum nicht?
2.  $f : \mathbb{R} \rightarrow \mathbb{R}$  sei definiert durch  $f(x) := ax^2 + bx + c$ , mit  $a \neq 0$ . Das ist eine „Quadratische Funktion“, und ihr Graph ist eine sogenannte „Parabel“. Was kann man über die Gestalt dieser Parabel sagen?
  - (a) Die Lösungen der quadratischen Gleichung  $f(x) = 0$  sind die „Nullstellen“ von  $f$ , also die Schnittpunkte von  $G_f$  mit der  $x$ -Achse.
  - (b) Nach dem Prinzip der quadratischen Ergänzung kann man schreiben:

$$f(x) = a((x - \alpha)^2 - \beta),$$

mit  $\alpha := -\frac{b}{2a}$  und  $\beta := \frac{\Delta}{4a^2}$ . Dann ergibt sich sofort:

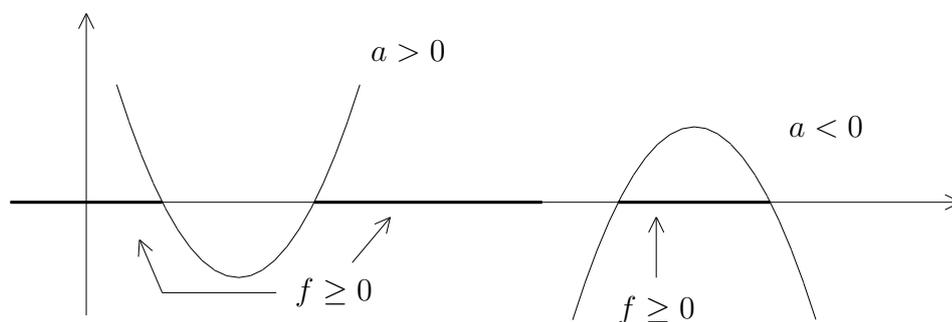
$$f(\alpha - x) = f(\alpha + x),$$

d.h.,  $G_f$  liegt symmetrisch zur vertikalen Geraden  $\{(x, y) \mid x = \alpha\}$ . Die Parabel ist entweder nach oben oder nach unten geöffnet.

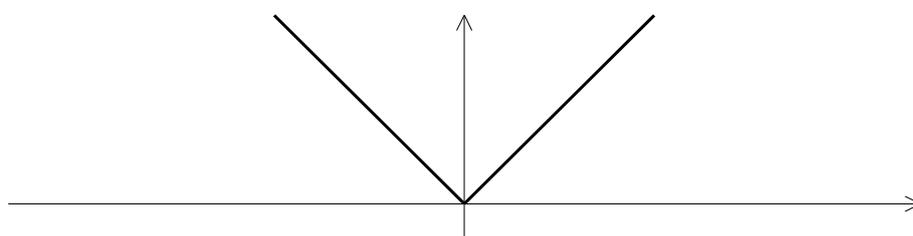
- (c) Ist  $a > 0$ , so ist  $f(x) \geq -a \cdot \beta = -\frac{\Delta}{4a}$ . Dann besitzt  $f$  einen Scheitelpunkt bei  $x = \alpha$ , und  $G_f$  ist nach oben geöffnet.

Ist  $a < 0$ , so ist  $f(x) \leq -a \cdot \beta$  und  $G_f$  nach unten geöffnet.

- (d) Man kann jetzt die Lösungsmenge einer quadratischen Ungleichung aus der Gestalt des Graphen entnehmen.

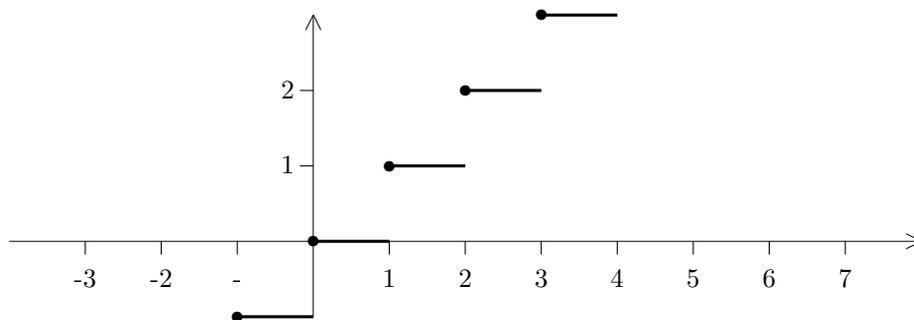


3. Die Funktion  $f(x) := |x|$  hat folgenden Graphen:

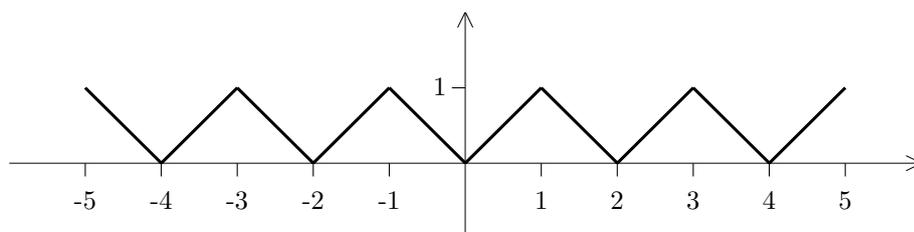


4. Für eine reelle Zahl  $x$  sei  $[x]$  die größte ganze Zahl  $n \in \mathbb{Z}$  mit  $n \leq x$ . Man spricht auch von der „Gauß-Klammer“. Z.B. ist  $[3.5] = 3$ ,  $[37] = 37$  und  $[-\frac{20}{7}] = -3$ .

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



5. Die Gauß-Klammer ist recht nützlich. Wir wollen die Definition der Funktion finden, deren Graph die folgende Zackenkurve ist:



Ist  $2n \leq x < 2n + 1$ , so soll  $f(x) = x - 2n$  sein, in dem betrachteten Intervall ist aber  $2n = [x]$ .

Ist  $2n + 1 \leq x < 2n + 2$ , so soll  $f(x) = -(x - (2n + 2))$  sein. Hier ist  $[x] = 2n + 1$ , also  $f(x) = 1 - (x - [x])$ . Zusammen ergibt das:

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

Bis jetzt haben wir nur reellwertige Funktionen betrachtet. Ist  $I = [a, b]$  ein Intervall und  $\varphi : I \rightarrow \mathbb{R}^2$  eine Abbildung, so nennt man  $\varphi$  eine *parametrisierte Kurve*. Der Fall  $I = \mathbb{R}$  ist auch zugelassen. Das Bild  $|\varphi| := \varphi(I)$  nennt man auch die *Spur* von  $\varphi$ . In vielen Fällen (aber nicht immer) kann man die Spur einer parametrisierten Kurve auch durch eine Gleichung  $F(x, y) = 0$  beschreiben.

### Beispiele :

1. Sind  $P_1 = (x_1, y_1)$ ,  $P_2 = (x_2, y_2)$  zwei verschiedene Punkte im  $\mathbb{R}^2$ , so wird durch  $\varphi(t) := (x_1 + t(x_2 - x_1), y_1 + t(y_2 - y_1))$  eine Abbildung  $\varphi : \mathbb{R} \rightarrow \mathbb{R}^2$  definiert, mit  $\varphi(0) = P_1$  und  $\varphi(1) = P_2$ . Ein Punkt  $(x, y)$  liegt genau dann auf  $|\varphi|$ , wenn es ein  $t \in \mathbb{R}$  gibt, so daß gilt:

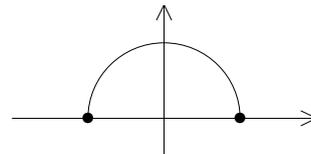
$$\begin{aligned} x &= x_1 + t(x_2 - x_1) \quad \text{und} \quad y = y_1 + t(y_2 - y_1), \\ \text{also} \quad \frac{x - x_1}{x_2 - x_1} &= \frac{y - y_1}{y_2 - y_1}. \end{aligned}$$

Das ist die Zwei-Punkte-Form einer Geraden. Wir haben mit  $\varphi$  also die Parametrisierung einer Geraden durch  $P_1$  und  $P_2$  gefunden. Ein Punkt  $\varphi(t)$  der Geraden liegt genau dann zwischen  $P_1$  und  $P_2$ , wenn  $0 < t < 1$  ist. Beschränkt man also  $\varphi$  auf den Definitionsbereich  $[0, 1]$ , so erhält man die Parametrisierung der *Verbindungsstrecke* von  $P_1$  und  $P_2$ .

2. Einen Kreis können wir momentan noch nicht so schön parametrisieren, das holen wir später nach. Aber eine Parametrisierung des halben Einheitskreises können wir immerhin angeben:

$$\varphi(t) := (t, \sqrt{1 - t^2}),$$

für  $-1 \leq t \leq +1$ .



3. Das vorige Beispiel ist ein Spezialfall der folgenden Situation: Ist  $f : I \rightarrow \mathbb{R}$  eine Funktion, so parametrisiert  $\varphi(t) := (t, f(t))$  den Graphen von  $f$ .

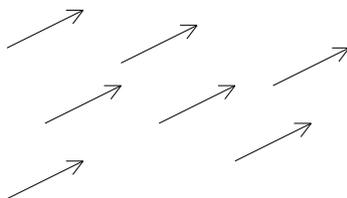
Während man Funktionen am besten durch ihren Graphen darstellt, veranschaulicht man Kurven durch ihre Spur. Schwierig wird es, wenn eine Funktion oder Abbildung von mehreren Variablen abhängt.

Sei etwa  $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  definiert durch

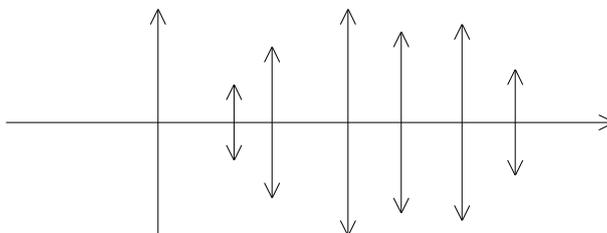
$$\Phi(x, y) := (x + a, y + b).$$

Diese Abbildung verschiebt jeden Punkt der Ebene um  $a$  Einheiten nach rechts und  $b$  Einheiten nach oben. Man nennt sie daher auch eine *Verschiebung* oder *Translation*. Der

Graph wäre eine Menge in  $\mathbb{R}^2 \times \mathbb{R}^2$ , und das kann man sich beim besten Willen nicht mehr vorstellen. Das Bild des Definitionsbereiches  $\mathbb{R}^2$  ist der ganze Zielbereich  $\mathbb{R}^2$ , also kann man auch nicht von einer Spur sprechen. Aber man kann z.B. jeden Punkt  $X$  der Ebene mit seinem Bildpunkt  $\Phi(X)$  durch einen Pfeil verbinden. Das ergibt eine unendliche Schar von Pfeilen gleicher Richtung und gleicher Länge. Dieses Bild sollte man sich für später merken.



Die Abbildung  $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  mit  $\Psi(x, y) := (x, -y)$  stellt die Spiegelung an der x-Achse dar. Verbindet man hier Punkt mit Bildpunkt, so erhält man eine unendliche Schar von vertikalen Doppelpfeilen, die alle von der x-Achse halbiert werden.



Schließlich wollen wir noch ein Beispiel ganz anderer Art betrachten:

Wir wählen eine Anordnung  $i_1, i_2, \dots, i_n$  der Zahlen  $1, 2, \dots, n$  und schreiben die beiden Zahlenreihen untereinander:

$$\begin{bmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{bmatrix}$$

Ordnet man jetzt der Zahl  $\nu$  jeweils die darunter stehende Zahl  $i_\nu$  zu, so ergibt das eine Abbildung  $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ . Dabei kommt unter den Bildelementen  $\sigma(i)$  jede der Zahlen  $1, 2, \dots, n$  wieder genau einmal vor. Man spricht auch von einer *Permutation* (der Zahlen  $1, 2, \dots, n$ ). Die Menge aller Permutationen bezeichnen wir mit dem Symbol  $\mathcal{S}_n$ . Da jedes Element von  $\mathcal{S}_n$  genau einer der möglichen Anordnungen der Zahlen  $1, 2, \dots, n$  entspricht, enthält  $\mathcal{S}_n$  genau  $n!$  Elemente.

Z.B. enthält  $\mathcal{S}_3$  die Elemente

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}.$$

### Definition.

Es sei  $f : A \rightarrow B$  eine beliebige Abbildung.

1.  $f$  heißt *injektiv* (oder *eindeutig*), falls gilt:

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

2.  $f$  heißt *surjektiv* (oder eine *Abbildung auf  $B$* ), falls gilt:

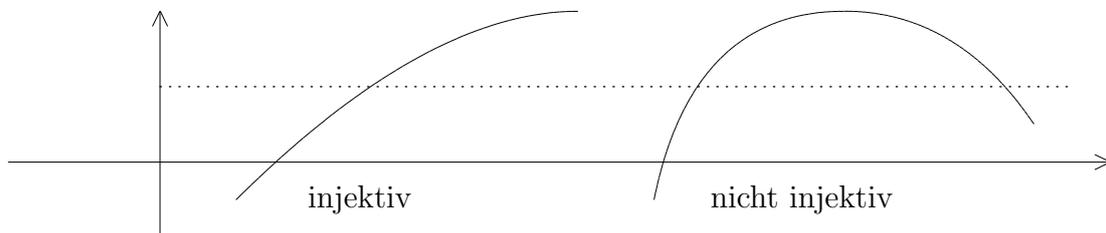
$$\forall y \in B \exists x \in A \text{ mit } f(x) = y.$$

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

$f$  ist also genau dann injektiv, wenn die Gleichung  $f(x) = y$  bei gegebenem  $y$  – wenn überhaupt – eindeutig lösbar ist. Und  $f$  ist genau dann surjektiv, wenn die Gleichung  $f(x) = y$  bei gegebenem  $y$  stets lösbar ist (aber nicht unbedingt eindeutig).

An Hand des Graphen  $G_f$  kann man sich diese Eigenschaften veranschaulichen:

$f$  ist genau dann injektiv, wenn der Graph von jeder horizontalen Geraden höchstens einmal getroffen wird.



Man weist Injektivität meist nach folgendem Schema nach:  $f(x_1) = f(x_2) \implies x_1 = x_2$ .

*Statt  $(\mathcal{A} \implies \mathcal{B})$  zu zeigen, kann man auch  $(\text{nicht } \mathcal{B} \implies \text{nicht } \mathcal{A})$  beweisen. Das nennt man „Beweis durch Kontraposition“. Das Verfahren ist dem Widerspruchsbeweis sehr ähnlich.*

$f$  ist genau dann surjektiv, wenn jede horizontale Gerade, die  $B$  trifft, auch den Graphen  $G_f$  trifft.

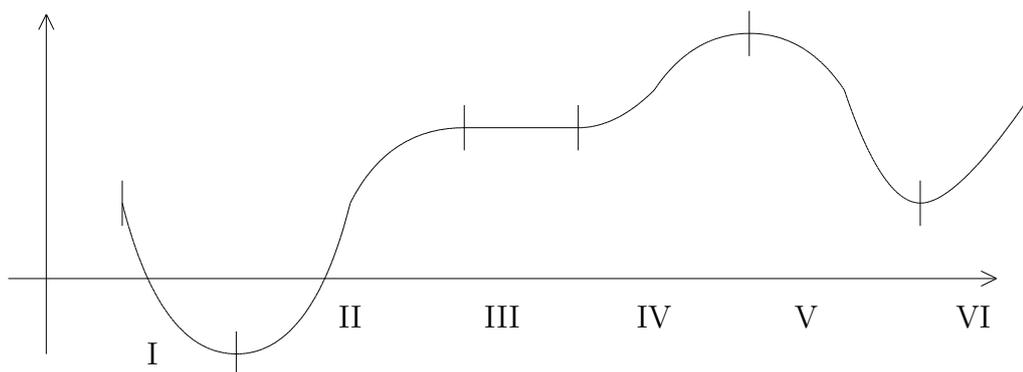
### Beispiele:

1.  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit  $f(x) := mx + t$  und  $m \neq 0$  ist injektiv und surjektiv, also bijektiv.
2.  $f : \mathbb{R} \rightarrow \mathbb{R}$  mit  $f(x) := x^2$  ist weder injektiv (denn  $f(-x) = f(x)$ ) noch surjektiv (denn negative Werte kommen nicht vor).
3. Surjektivität kann man immer durch Verkleinern der Zielmenge erreichen. So ist z.B.  $f : \mathbb{R} \rightarrow \mathbb{R}_+ \cup \{0\}$  mit  $f(x) := x^2$  surjektiv.
4. Eine Funktion  $f : I \rightarrow \mathbb{R}$  heißt *monoton wachsend* (bzw. *streng monoton wachsend*), falls für  $x_1, x_2 \in I$  mit  $x_1 < x_2$  gilt:

$$f(x_1) \leq f(x_2) \quad (\text{bzw. } f(x_1) < f(x_2)).$$

Analog definiert man „(streng) monoton fallend“.

Normalerweise zeigt eine Funktion auf verschiedenen Abschnitten unterschiedliches Monotonieverhalten:



In den Bereichen (I) und (V) fällt  $f$  streng monoton, in den Bereichen (II), (IV) und (VI) wächst  $f$  streng monoton, und in (III) ist  $f$  konstant, also zugleich monoton wachsend und fallend, aber weder streng monoton wachsend, noch streng monoton fallend.

Offensichtlich gilt: Ist  $f$  streng monoton (wachsend oder fallend), so ist  $f$  injektiv.

Betrachten wir z.B.  $f(x) := x^3$  und  $x_1 < x_2$ . Dann ist

$$f(x_2) - f(x_1) = (x_2)^3 - (x_1)^3 = (x_2 - x_1) \cdot (x_1^2 + x_1x_2 + x_2^2).$$

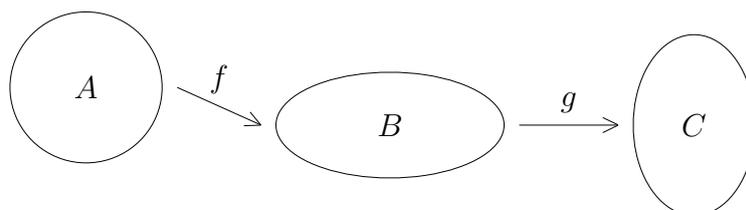
Haben  $x_1, x_2$  das gleiche Vorzeichen, so ist der Ausdruck in der rechten Klammer  $> 0$ , und da auch  $x_2 - x_1 > 0$  ist, muß  $f(x_1) < f(x_2)$  sein. Ist jedoch  $x_1 < 0 < x_2$ , so ist auch  $f(x_1) < 0 < f(x_2)$ . Also ist  $x \mapsto x^3$  streng monoton wachsend und damit erst recht injektiv.

Im übrigen ist  $x \mapsto x^3$  auch surjektiv. Zu jeder positiven reellen Zahl  $a$  gibt es eine positive Zahl  $x$  mit  $x^3 = a$ . Aber dann ist auch  $-a = (-x)^3$ .

### Definition.

$A, B$  und  $C$  seien irgendwelche Mengen. Sind Abbildungen  $f : A \rightarrow B$  und  $g : B \rightarrow C$  gegeben, so kann man diese Zuordnungen hintereinander ausführen und so zu einer neuen Abbildung  $g \circ f : A \rightarrow C$  verknüpfen, mit

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



**Beispiele :**

1. Die Funktionen  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  seien definiert durch  $f(x) := x^2 + 2x$  und  $g(y) := y + 1$ . Dann gilt:

$$\begin{aligned} g \circ f(x) &= x^2 + 2x + 1, \\ f \circ g(y) &= (y + 1)^2 + 2(y + 1) = y^2 + 4y + 3. \end{aligned}$$

Im allgemeinen ist also  $f \circ g \neq g \circ f$ .

2. Sind  $\Phi : (x, y) \mapsto (x + a, y + b)$  und  $\Psi : (x, y) \mapsto (x + c, y + d)$  zwei Translationen der Ebene, so ist  $\Psi \circ \Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  gegeben durch

$$(\Psi \circ \Phi)(x, y) = \Psi(x + a, y + b) = (x + (a + c), y + (b + d)),$$

also wieder eine Translation.

3. Ist  $S : (x, y) \mapsto (x, -y)$  die Spiegelung an der x-Achse, so ist  $S \circ S(x, y) = (x, y)$ .

Ist  $A$  eine beliebige Menge, so wird die Abbildung  $\text{id}_A : A \rightarrow A$  mit  $\text{id}_A(x) := x$  als *identische Abbildung* von  $A$  bezeichnet. Wir haben gerade gesehen, daß  $S \circ S = \text{id}_{\mathbb{R}^2}$  ist.

Sind  $\sigma_1 := \begin{bmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{bmatrix}$  und  $\sigma_2 := \begin{bmatrix} 1 & 2 & \dots & n \\ k_1 & k_2 & \dots & k_n \end{bmatrix}$  zwei Permutationen, so kann man sie zu einer neuen Permutation  $\sigma_2 \circ \sigma_1$  miteinander verknüpfen. Dabei wird z.B. die 1 zunächst auf  $\sigma_1(1) = i_1$  und  $i_1$  dann auf  $\sigma_2(i_1) = k_{i_1}$  abgebildet.

So ist etwa

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} \circ \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}.$$

Eine Permutation heißt *Transposition* oder *Vertauschung*, wenn nur zwei Zahlen miteinander vertauscht werden und alle anderen fest bleiben, z.B. wird bei

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

nur die 1 mit der 4 vertauscht. Eine beliebige Permutation entsteht nun, indem man genügend viele Vertauschungen hintereinander ausführt. Leider ist die Zerlegung einer Permutation in Transpositionen nicht eindeutig. Allerdings gilt:

**I.3.1 Satz.** *Wird eine Permutation auf zweierlei Weisen durch eine Folge von Transpositionen erzeugt, so ist die Anzahl der benötigten Vertauschungen in beiden Fällen entweder gerade oder ungerade.*

Der Beweis ist nicht ganz einfach und wird deshalb weggelassen. Die Idee besteht darin, beide Anzahlen zu addieren und zu zeigen, daß dabei eine gerade Zahl herauskommt, weil man nur durch eine gerade Anzahl von Vertauschungen einen früheren Zustand wieder herstellen kann.

**Definition.**

Eine Permutation  $\sigma \in \mathcal{S}_n$  heißt *gerade*, wenn sie sich aus einer geraden Anzahl von Vertauschungen zusammensetzt. Andernfalls heißt sie *ungerade*. Wir setzen

$$\operatorname{sgn}(\sigma) := \begin{cases} +1 & \text{falls } \sigma \text{ gerade ist,} \\ -1 & \text{falls } \sigma \text{ ungerade ist,} \end{cases}$$

und nennen diese Zahl das *Signum* der Permutation.

Zur Abkürzung setzen wir  $(i_1, i_2, \dots, i_n) := \begin{bmatrix} 1 & 2 & \dots & n \\ i_1 & i_2 & \dots & i_n \end{bmatrix}$ . Dann ist

$$\begin{aligned} \operatorname{sgn}(1, 2, 4, 3, 5) &= -1, \\ \operatorname{sgn}(3, 4, 1, 2, 5) &= +1 \\ \text{und } \operatorname{sgn}(3, 5, 4, 1, 2) &= -1. \end{aligned}$$

Wir werden dies später in der Determinanten-Theorie brauchen.

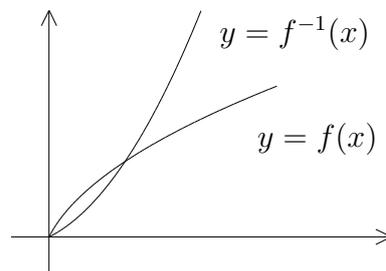
Sei  $f : A \rightarrow B$  eine **bijektive** Abbildung. Dann gibt es wegen der Surjektivität zu jedem  $y \in B$  ein  $x \in A$  mit  $f(x) = y$ , und wegen der Injektivität ist dieses  $x$  eindeutig bestimmt. Also ist die Zuordnung  $y = f(x) \mapsto x$  auch eine Abbildung. Man nennt sie die *Umkehrabbildung* zu  $f$  und bezeichnet sie mit  $f^{-1}$ . Es gilt dann:

$$\begin{aligned} f^{-1} \circ f &= \operatorname{id}_A \\ \text{und } f \circ f^{-1} &= \operatorname{id}_B. \end{aligned}$$

**Beispiele:**

1. Sei  $f : \mathbb{R} \rightarrow \mathbb{R}$  definiert durch  $f(x) := mx + t$ , mit  $m \neq 0$ . Dann ist  $f^{-1}(y) = \frac{1}{m}(y - t)$ .
2. Sei  $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  definiert durch  $f(x) := x^2$ . Dann ist  $f^{-1}(y) = \sqrt{y}$ .

Den Graphen der Umkehrfunktion  $f^{-1}$  erhält man übrigens, indem man den Graphen von  $f$  an der Winkelhalbierenden spiegelt.



**I.3.2 Satz.** Eine Abbildung  $f : A \rightarrow B$  ist genau dann bijektiv, wenn es eine weitere Abbildung  $g : B \rightarrow A$  gibt, so daß  $g \circ f = \operatorname{id}_A$  und  $f \circ g = \operatorname{id}_B$  ist.

**BEWEIS:** Ist  $f$  bijektiv, so wählt man für  $g$  einfach die Umkehrabbildung  $f^{-1}$ . Ist umgekehrt ein  $g$  mit den angegebenen Eigenschaften vorhanden, so muß man zeigen, daß  $f$  bijektiv ist:

- 1) Ist  $y \in B$ , so ist  $f(g(y)) = y$ , also  $f$  surjektiv.

2) Ist  $f(x_1) = f(x_2)$ , so ist  $x_1 = g(f(x_1)) = g(f(x_2)) = x_2$ , also  $f$  injektiv.  $\square$

Dieses Kriterium ist sehr nützlich. Z.B. ist die Translation  $\Phi : (x, y) \mapsto (x + a, y + b)$  bijektiv, weil man sofort eine Umkehrabbildung angeben kann:

$$\Phi^{-1} : (x, y) \mapsto (x - a, y - b).$$

Wir fassen noch einige Eigenschaften der Translationen zusammen:

1.  $\text{id} = \text{id}_{\mathbb{R}^2}$  ist eine Translation, und es ist  $\Phi \circ \text{id} = \text{id} \circ \Phi = \Phi$  für jede Translation  $\Phi$ .
2. Mit  $\Phi$  und  $\Psi$  ist auch deren Verknüpfung  $\Phi \circ \Psi$  eine Translation.
3. Es ist  $\Phi \circ (\Psi \circ \Lambda) = (\Phi \circ \Psi) \circ \Lambda$ .
4. Jede Translation ist bijektiv, mit  $\Phi$  ist auch  $\Phi^{-1}$  eine Translation, und es ist  $\Phi \circ \Phi^{-1} = \Phi^{-1} \circ \Phi = \text{id}$ .

Eine ähnliche Struktur trifft man in der Mathematik sehr oft an. Man spricht dann von einer *Gruppe*.

### Definition.

Eine *Gruppe* ist eine Menge  $G$ , in der je zwei Elemente  $g$  und  $h$  zu einem Element  $gh \in G$  verknüpft werden können, so daß gilt:

1.  $g(hk) = (gh)k$  für alle  $g, h, k \in G$ .
2. Es gibt ein Element  $e \in G$ , so daß  $ge = eg = g$  für alle  $g \in G$  ist.
3. Zu jedem  $g \in G$  gibt es ein  $g^{-1} \in G$ , so daß  $gg^{-1} = g^{-1}g = e$  ist.

Das Element  $e$  ist eindeutig bestimmt und wird als *neutrales Element* bezeichnet. Das Element  $g^{-1}$  heißt *das Inverse* zu  $g$ .

Ist  $gh = hg$  für alle  $g, h \in G$ , so nennt man die Gruppe *kommutativ*.

### Beispiele :

1. Die Permutationen  $\sigma \in S_n$  bilden mit ihrer Verknüpfung auch eine Gruppe. Da i.a.  $\sigma_1 \circ \sigma_2 \neq \sigma_2 \circ \sigma_1$  ist, nennt man die Gruppe *nicht kommutativ*. Die Gruppe der Translationen ist dagegen kommutativ, wie man leicht sieht.
2. Die reellen Zahlen mit der Verknüpfung  $+$ , dem „Inversen“  $-x$  und dem „neutralen Element“  $0$  bilden eine kommutative Gruppe. Und genauso die Menge  $\mathbb{R}^* := \mathbb{R} \setminus \{0\}$  mit der Verknüpfung  $\cdot$ , dem Inversen  $x^{-1}$  und dem neutralen Element  $1$ .

## §4 Elementare Funktionen

### A) 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  $a_n \neq 0$  heißt *Polynom vom Grad  $n$* . Die Zahlen  $a_i$  nennt man die *Koeffizienten* des Polynoms.

Eine konstante Funktion  $f(x) \equiv c$  mit  $c \neq 0$  ist ein Polynom vom Grad 0. Das „Nullpolynom“  $f(x) \equiv 0$  bekommt definitionsgemäß den Grad  $-1$ .

**Bemerkung:**  $f(x) \equiv c$  bedeutet, daß  $f(x) = c$  für **jedes**  $x \in \mathbb{R}$  ist. Man darf dann auch kürzer schreiben:  $f = c$ . Und man kann sagen, daß  $f$  durch  $f(x) := c$  definiert wird.

Dagegen bedeutet die Gleichung  $f(x) = c$  (ohne den Doppelpunkt), daß  $f$  an einer Stelle  $x$  den Wert  $c$  annimmt, in  $x$  also eine  $c$ -Stelle besitzt. Besonders interessieren einen die Nullstellen, also die Lösungen der Gleichung  $f(x) = 0$ , während die Funktion  $f(x) \equiv 0$  (das Nullpolynom) höchst langweilig ist.

Das Rechnen mit Polynomen dürfte bekannt sein, hier soll nur an die Grad-Formel

$$\deg(f \cdot g) = \deg(f) + \deg(g)$$

erinnert werden.

Eins der wichtigsten Hilfsmittel in der Theorie der Polynome ist die *Division mit Rest*:

**I.4.1 Satz (Division mit Rest für Polynome).** *Es seien  $f, g$  Polynome mit  $\deg(g) \leq \deg(f)$ . Dann gibt es Polynome  $q$  und  $r$ , so daß gilt:*

$$1. \quad f = q \cdot g + r \quad \left( \text{also } \frac{f}{g} = q + \frac{r}{g} \right).$$

2. *Entweder ist  $r = 0$  (d.h., die Division geht auf),  
oder es ist  $0 \leq \deg(r) < \deg(g)$ .*

**BEWEIS:** Wir kennen die Division mit Rest von den positiven ganzen Zahlen: Sind  $a, b \in \mathbb{N}$  mit  $a \geq b$ , so dividiert man  $a$  durch  $b$  mit Rest, indem man  $b$  so oft von  $a$  subtrahiert, bis eine Zahl  $r$  mit  $0 \leq r < b$  übrigbleibt. Ist  $q$  die Anzahl der Subtraktionen, so ist  $a = q \cdot b + r$ .

Bei den Polynomen geht das nicht so einfach. Es sei

$$f(x) = a_n x^n + \cdots + a_1 x + a_0 \quad \text{und} \quad g(x) = b_m x^m + \cdots + b_1 x + b_0, \quad \text{mit } m \leq n.$$

Wir wollen  $g(x)$  „so oft“ von  $f(x)$  subtrahieren, daß ein Polynom  $r(x)$  mit einem Grad  $< m$  übrigbleibt, wir bilden also Ausdrücke der Gestalt  $f - q \cdot g$ . Ist der Faktor  $q$  nur eine Zahl

und etwa  $n > m$ , so können wir auf diese Weise die höchsten Terme von  $f$  nie zum Verschwinden bringen. So einfach geht es also nicht, wir müssen für  $q$  auch Polynome zulassen. Setzen wir z.B.

$$q(x) := \frac{a_n}{b_m} x^{n-m},$$

so ist  $q(x) \cdot g(x) = a_n x^n +$  Terme kleineren Grades, und  $f(x) - q(x) \cdot g(x)$  enthält keinen Term vom Grad  $n$  mehr. Beginnt der verbliebene Rest mit einem Term  $c_k x^k$  mit  $k \geq m$ , so subtrahieren wir davon als nächstes das Polynom

$$\frac{c_k}{b_m} x^{k-m} \cdot g(x)$$

und verringern erneut den Grad des Restpolynoms. Das verbesserte  $q$  hat jetzt die Gestalt

$$q(x) = \frac{a_n}{b_m} x^{n-m} + \frac{c_k}{b_m} x^{k-m}.$$

Dieses Verfahren können wir solange fortsetzen, bis  $r(x) = f(x) - q(x) \cdot g(x)$  einen Grad  $< m$  aufweist oder gar das Nullpolynom ist. Dann sind wir fertig.  $\square$

### Rechenbeispiel:

$$\begin{array}{r} (x^5 \qquad \qquad \qquad +3x^2 \quad -2x \quad +7) : (x^3 \quad +4x^2) = x^2 \quad -4x \quad +16 \\ \underline{x^5 \quad +4x^4} \qquad \qquad \qquad +3x^2 \quad -2x \quad +7 \\ \quad -4x^4 \quad -16x^3 \qquad \qquad \qquad +3x^2 \quad -2x \quad +7 \\ \quad \quad \underline{16x^3} \quad +3x^2 \quad -2x \quad +7 \\ \quad \quad \quad 16x^3 \quad +64x^2 \\ \quad \quad \quad \quad \underline{-61x^2} \quad -2x \quad +7 \\ \quad \quad \quad \quad \quad \uparrow \text{ Dies hat einen Grad } < 3 \end{array}$$

Also ist  $f(x) = (x^2 - 4x + 16) \cdot g(x) + (-61x^2 - 2x + 7)$ .

**I.4.2 Satz.** 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:

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

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

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

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

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

$\square$

**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, daß

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

ist. Zur Probe kann man ausmultiplizieren.

**Definition.**

Sei  $p(x)$  ein Polynom vom Grad  $n \geq 0$ . Eine Zahl  $c$  heißt *k-fache Nullstelle* von  $p$ , falls es ein Polynom  $q(x)$  vom Grad  $n - k$  gibt, so daß gilt:

1.  $p(x) = (x - c)^k \cdot q(x)$ .
2.  $q(c) \neq 0$ .

**Bemerkung:** Da  $q(c) \neq 0$  sein soll, kann  $q$  nicht das Nullpolynom sein. Also muß sein Grad  $n - k \geq 0$  sein, d.h.  $k \leq n$ .

**I.4.3 Satz.** Sei  $p(x)$  ein Polynom vom Grad  $n \geq 0$ .

1.  $p$  besitzt höchstens  $n$  Nullstellen (mit Vielfachheit gezählt).
2. Sind  $c_1, \dots, c_m$  die verschiedenen Nullstellen von  $p$ ,  $m \leq n$ , mit Vielfachheiten  $k_1, \dots, k_m$ , so ist

$$p(x) = (x - c_1)^{k_1} \cdot \dots \cdot (x - c_m)^{k_m} \cdot q(x),$$

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

**BEWEIS:** Ist  $\deg(p) = 0$ , so ist  $p$  konstant und die Aussage trivial.

Sei also  $\deg(p) \geq 1$ . Es sei weiter  $m \geq 0$  und  $c_1, \dots, c_m$  ein System von (paarweise verschiedenen) Nullstellen von  $p$ , mit den Vielfachheiten  $k_1, \dots, k_m$ . Die Möglichkeit  $m = 0$  lassen wir zu, dann gibt es eben keine Nullstellen.

Durch wiederholte Polynomdivision kann man ein Polynom  $q(x)$  finden, so daß die Gleichung

$$p(x) = (x - c_1)^{k_1} \cdot \dots \cdot (x - c_m)^{k_m} \cdot q(x)$$

erfüllt ist und  $c_1, \dots, c_m$  keine Nullstellen von  $q$  mehr sind.  $q$  kann nicht das Nullpolynom sein.

Da die Polynome auf der linken und der rechten Seite der Gleichung gleich sind, haben sie auch den gleichen Grad, und es muß

$$k_1 + \dots + k_m + \deg(q) \leq n$$

sein. Da nach Konstruktion  $\deg(q) \geq 0$  ist und alle  $k_i \geq 1$  sind, muß  $k_1 + \dots + k_m \leq n$  und insbesondere  $m \leq n$  sein, es kann also – selbst mit Vielfachheit gezählt – höchstens  $n$  Nullstellen geben. Wenn man dafür sorgt, daß die Menge  $\{c_1, \dots, c_m\}$  bereits alle Nullstellen enthält, dann kann  $q(x)$  keine Nullstelle mehr haben.  $\square$

**Beispiel:**

Wir betrachten ein quadratisches Polynom

$$p(x) = ax^2 + bx + c = a(x^2 + \beta x + \gamma), \text{ mit } a \neq 0, \beta := \frac{b}{a} \text{ und } \gamma := \frac{c}{a}.$$

Wir benutzen die Diskriminante  $\Delta := \beta^2 - 4\gamma$ . Ist  $\Delta \geq 0$ , so gilt:

$$\begin{aligned} p(x) = 0 &\iff x^2 + \beta x + \gamma = 0 \\ &\iff x = x_{1,2} = \frac{-\beta \pm \sqrt{\Delta}}{2}. \end{aligned}$$

Dann sieht man sofort:

$$\begin{aligned} x_1 + x_2 &= -\beta \\ \text{und } x_1 \cdot x_2 &= \gamma. \end{aligned}$$

Diese Beziehung wird auch als „Vietascher Wurzelsatz“ bezeichnet. Nun ergibt sich:

$$p(x) = a(x - x_1)(x - x_2).$$

Ist  $\Delta = 0$  und daher  $x_1 = x_2 =: c$ , so ist  $p(x) = a(x - c)^2$ , und  $p$  hat in  $x = c$  eine Nullstelle 2. Ordnung. Ist  $\Delta < 0$ , so hat  $p$  überhaupt keine Nullstelle. Das ist z.B. bei  $p(x) = 1 + x^2$  der Fall.

**B) 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  $q$  ein Polynom vom Grad  $\geq 0$  ist, besitzt  $q$  höchstens endlich viele Nullstellen. Die muß man aus  $\mathbb{R}$  herausnehmen, um  $D_f$  zu bekommen.

Was passiert in einem Punkt  $x_0$ , wo  $q(x_0) = 0$  ist?

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  $\tilde{q}(x_0) \neq 0$ . Wenn schon  $p(x_0) \neq 0$  ist, muß man  $k = 0$  setzen.

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 Wirklichkeit 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 wird der erste Faktor bei Annäherung an  $x_0$  beliebig groß, während der zweite Faktor beschränkt bleibt. Also kann man  $f$  in  $x_0$  keinen vernünftigen Wert zuordnen. Man sagt in dieser Situation:  $f$  besitzt in  $x_0$  eine *Polstelle* der Ordnung  $l - k$ .

### C) Exkurs über Grenzwerte und Stetigkeit von Funktionen

Gerade war von dem Verhalten einer Funktion  $f(x)$  bei Annäherung an ein  $x$  die Rede. Das müssen wir noch etwas klarer formulieren:

#### Definition.

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

1. Wir sagen,  $f$  *strebt bei Annäherung an  $x_0$  gegen den Wert  $c$* , 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:  $\lim_{x \rightarrow x_0} f(x) = c$ .

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

2. Jetzt sei  $I = [a, \infty) = \{x \in \mathbb{R} \mid x \geq a\}$  und  $f$  auf ganz  $I$  definiert. Wir sagen,  $f$  *strebt für  $x$  gegen Unendlich (asymptotisch) gegen  $c$* , falls gilt:

Für **jede** Folge  $(x_n)$  in  $I$  mit  $\lim_{n \rightarrow \infty} x_n = +\infty$  ist  $\lim_{n \rightarrow \infty} f(x_n) = c$ .

Man schreibt dann:  $\lim_{x \rightarrow +\infty} f(x) = c$ .

Analog definiert man  $\lim_{x \rightarrow -\infty} f(x) = c$ .

#### Beispiele:

1. Die Geschichte mit der hebbaren Unbestimmtheitsstelle wird nun klarer:

$$\lim_{x \rightarrow -1} \frac{x^2 - 1}{x + 1} = \lim_{x \rightarrow -1} (x - 1) = -2.$$

2. Ist  $f(x) := x$ , so ist  $\lim_{x \rightarrow x_0} f(x) = x_0 = f(x_0)$  für jedes  $x_0$  aus dem Definitionsbereich. Ein solches Verhalten werden wir noch genauer studieren.

3. Es ist offensichtlich  $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$ .

Problematischer wird es, wenn man  $\lim_{x \rightarrow 0} \frac{1}{x}$  berechnen möchte. Läßt man nur positive  $x$  zu, so ist der Grenzwert  $+\infty$ , läßt man dagegen nur negative  $x$  zu, so ist der Grenzwert  $-\infty$ .

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

**Definition.**

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

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

Man schreibt dann:

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

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

Ist  $-\infty < b \leq +\infty$  und  $a < b$ , so kann man analog den linksseitigen Grenzwert  $\lim_{x \rightarrow b-} f(x) = f(b-)$  erklären.

**Beispiel:**

Es ist

$$\lim_{x \rightarrow 0-} \frac{1}{x} = -\infty \text{ und } \lim_{x \rightarrow 0+} \frac{1}{x} = +\infty.$$

Aus den Grenzwert-Sätzen folgt für Funktionen:

**I.4.4 Satz.** Es seien  $f$  und  $g$  zwei Funktionen auf  $I \setminus \{a\}$ , wobei  $I$  ein Intervall und  $a$  ein Punkt aus dem Inneren oder vom Rand des Intervalls ist. Wenn die Grenzwerte  $\lim_{x \rightarrow a} f(x) = c$  und  $\lim_{x \rightarrow a} g(x) = d$  existieren, dann gilt:

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  $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{c}{d}$ .

Sei  $a > 0$ ,  $I := [a, \infty)$  und  $f : I \rightarrow \mathbb{R}$  eine Funktion. Dann ist  $f(\frac{1}{x})$  für alle  $x \in (0, \frac{1}{a}]$  definiert. Offensichtlich gilt:

Wenn  $\lim_{x \rightarrow 0+} f(\frac{1}{x})$  existiert, dann existiert auch  $\lim_{x \rightarrow +\infty} f(x)$ , und beide Grenzwerte sind gleich.

Entsprechende Überlegungen gelten für die Annäherung an  $-\infty$ .

Daraus folgt z.B.:

**I.4.5 Satz.** Sind  $p$  und  $q$  zwei Polynome mit  $\deg(p) < \deg(q)$ , so ist

$$\lim_{x \rightarrow +\infty} \frac{p(x)}{q(x)} = 0.$$

BEWEIS: Sei

$$p(x) = \sum_{i=0}^m a_i x^i \quad \text{und} \quad q(x) = \sum_{k=0}^n b_k x^k,$$

mit  $m < n$ ,  $a_m \neq 0$  und  $b_n \neq 0$ . Dann gilt für  $f(x) := \frac{p(x)}{q(x)}$ :

$$\begin{aligned} f\left(\frac{1}{x}\right) &= \frac{a_m x^{-m} + \dots + a_0}{b_n x^{-n} + \dots + b_0} \\ &= \frac{a_m x^{n-m} + \dots + a_0 x^n}{b_n + b_{n-1}x + \dots + b_0 x^n}. \end{aligned}$$

Also ist  $\lim_{x \rightarrow 0+} f\left(\frac{1}{x}\right) = 0$ . □

Als Folgerung ergibt sich für Polynome  $p$  und  $q$  mit  $\deg(p) > \deg(q)$ :

Ist  $p(x) = g(x) \cdot q(x) + r(x)$  mit  $\deg(r) < \deg(q)$  (Division mit Rest), so ist

$$\lim_{x \rightarrow +\infty} \frac{p(x)}{q(x)} = \lim_{x \rightarrow +\infty} g(x).$$

### Definition.

Sei  $f$  eine Funktion.

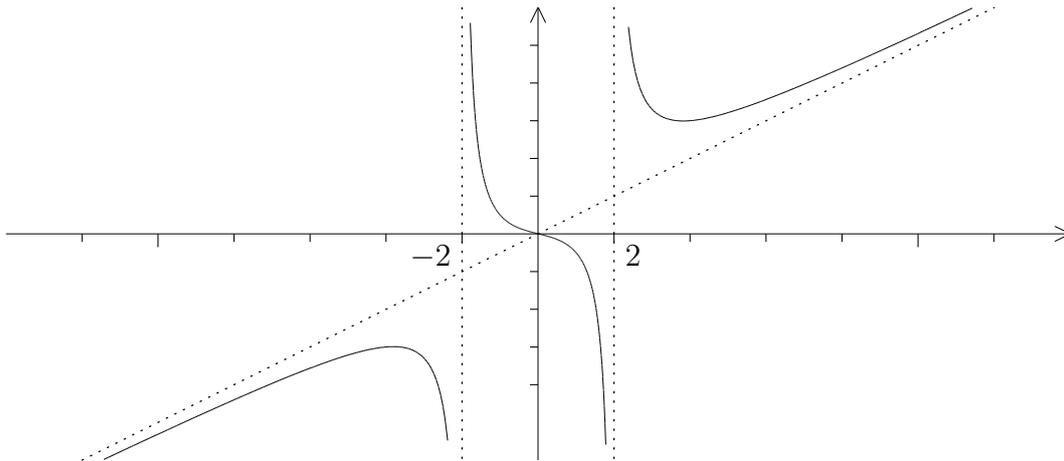
1. Ist  $x_0 \in \mathbb{R}$  und  $\lim_{x \rightarrow x_0} f(x) = \pm\infty$ , so heißt die Gerade  $\{(x, y) \mid x = x_0\}$  eine *vertikale Asymptote* von  $f$ .
2. Ist  $g$  eine weitere Funktion und  $\lim_{x \rightarrow \pm\infty} (f(x) - g(x)) = 0$ , so heißt der Graph von  $g$  eine *asymptotische Kurve* für  $f$ . Ist  $g$  sogar eine lineare Funktion vom Typ  $g(x) = mx + t$ , so spricht man auch hier von einer *Asymptoten*.

### Beispiel:

Sei  $f(x) := \frac{x^3 + 2x}{2x^2 - 8}$  für  $x \neq \pm 2$ . Polynomdivision ergibt:

$$f(x) = \frac{x}{2} + \frac{3x}{(x-2)(x+2)}.$$

Die Geraden  $\{x = -2\}$  und  $\{x = +2\}$  sind offensichtlich vertikale Asymptoten. Die Gerade  $\{y = \frac{1}{2}x\}$  ist eine weitere Asymptote,  $x = 0$  die einzige Nullstelle. Es ergibt sich dann folgendes Bild:



Im Falle der Funktion  $f(x) := x$  war stets  $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ . Dieses besonders schöne Verhalten bekommt einen eigenen Namen:

**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.

Nun ist unmittelbar klar:

**I.4.6 Satz.** Sind  $f$  und  $g$  beide in  $x_0$  stetig, so sind auch  $f \pm g$  und  $f \cdot g$  in  $x_0$  stetig.

Da  $f(x) := x$  und natürlich auch alle konstanten Funktionen stetig sind, ist jedes Polynom auf ganz  $\mathbb{R}$  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:

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

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

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

ist dann stetig (zur Stetigkeit von rationalen Funktionen kommen wir weiter unten), und man sagt,  $f$  ist *nach 1 stetig fortsetzbar*.

Betrachten wir schließlich das Beispiel  $f(x) := [x]$ .

Im Innern der Intervalle  $I_n := [n, n + 1]$  ist  $f$  konstant und daher stetig. An den Endpunkten  $x_n := n$  existieren der linksseitige und der rechtsseitige Grenzwert:

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

Also ist  $f$  in  $x_n$  nicht stetig.

Allgemein gilt:

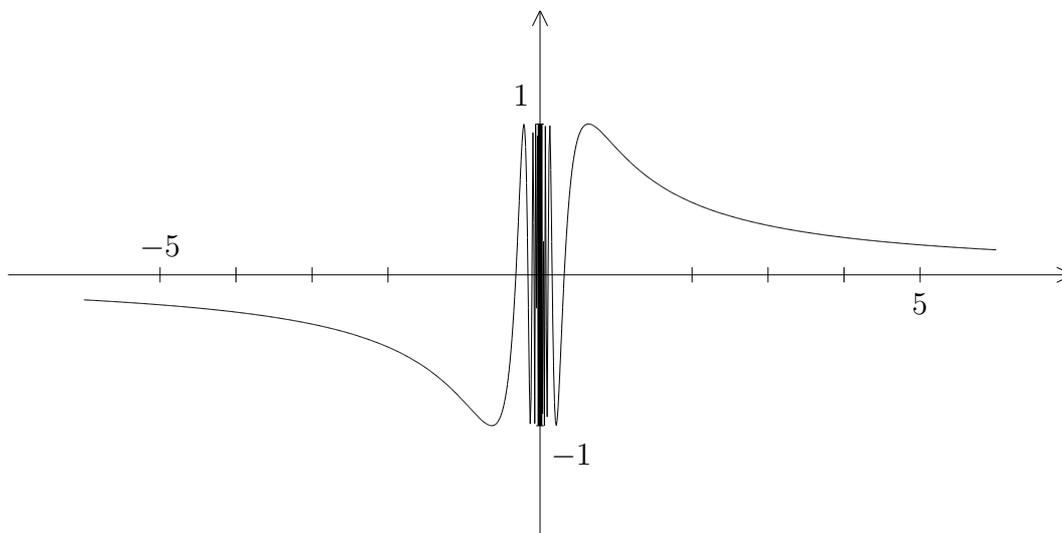
**I.4.7 Satz.** Eine Funktion  $f$  ist genau dann in  $x_0$  stetig, wenn  $f(x_0+)$  und  $f(x_0-)$  beide existieren und gleich sind.

**Definition.**

Die Funktion  $f$  sei in der Nähe von  $x_0$  definiert. Wenn  $f(x_0+)$  und  $f(x_0-)$  existieren und nicht gleich sind, nennt man  $x_0$  eine *Sprungstelle* von  $f$ .

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

Wenn einer der einseitigen Grenzwerte (oder gar beide) nicht existieren, dann liegt in  $x_0$  eine kompliziertere Unstetigkeitsstelle vor. Das ist z.B. bei der Funktion  $f(x) := \sin(\frac{1}{x})$  der Fall.



Nun folgen weitere wichtige Sätze zur Stetigkeit:

**I.4.8 Satz.** Die Funktion  $f : I \rightarrow \mathbb{R}$  sei in  $x_0 \in I$  stetig. Ist  $f(x_0) > 0$  (bzw.  $f(x_0) < 0$ ), so gibt es eine  $\varepsilon$ -Umgebung  $U_\varepsilon(x_0)$ , so daß  $f(x) > 0$  (bzw.  $f(x) < 0$ ) für alle  $x \in I \cap U_\varepsilon(x_0)$  ist.

Wir begnügen uns mit einer Plausibilitätsbetrachtung: Da  $f$  in  $x_0$  stetig ist, ist

$$\lim_{n \rightarrow \infty} f(x_n) = f(x_0) \text{ für jede Folge } (x_n) \text{ in } I \text{ mit } \lim_{n \rightarrow \infty} x_n = x_0.$$

Also kommen die Werte von  $f$  in der Nähe von  $x_0$  dem Wert  $f(x_0)$  beliebig nahe.

**I.4.9 Folgerung.** Ist  $f$  in  $x_0$  stetig und  $f(x_0) \neq 0$ , so ist auch  $\frac{1}{f}$  in  $x_0$  stetig.

Also sind rationale Funktionen überall dort stetig, wo sie definiert sind.

**I.4.10 Satz.** Ist  $f$  stetig in  $x_0$  und  $g$  stetig in  $y_0 := f(x_0)$ , so ist die zusammengesetzte Funktion  $g \circ f$  stetig in  $x_0$ .

Der Beweis ist trivial, er ergibt sich aus der Definition der Stetigkeit.

Der folgende Satz ist besonders wichtig:

**I.4.11 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). Außerdem sei zur Vereinfachung  $c = 0$ .

Der Mittelpunkt des Intervalls  $I_1 := [a, b]$  ist  $x_1 := \frac{a+b}{2}$ . Ist  $f(x_1) = 0$ , so ist man fertig. Ist  $f(x_1) < 0$ , so wähle man  $I_2 := [x_1, b]$ , ist  $f(x_1) > 0$ , so wähle man  $I_2 := [a, x_1]$ . Auf jeden Fall erfüllt  $f$  auf  $I_2$  die Voraussetzungen des Satzes. So fährt man fort. Entweder ist der gesuchte Punkt  $\xi$  nach endlich vielen Schritten gefunden, oder es ergibt sich eine Intervallschachtelung  $(I_k)$  mit der Eigenschaft, daß  $f$  an den Intervallgrenzen stets verschiedenes Vorzeichen hat. In dem Punkt  $\xi$ , der in allen  $I_k$  liegt, muß  $f$  dann eine Nullstelle haben.  $\square$

Nicht nur der Satz, sondern auch sein Beweis ist für die Anwendung interessant: Es sei etwa  $f(x) := x^2 - 2$  auf  $I_1 := [0, 2]$ . Dann gilt:

$$\begin{aligned} f(0) = -2 < 0 < +2 = f(2), \\ f(1) = -1 < 0 < +2 = f(2), \text{ also } I_2 = [1, 2], \\ f(1) = -1 < 0 < +\frac{1}{4} = f\left(\frac{3}{2}\right), \text{ also } I_3 = [1, 1.5], \\ f\left(\frac{5}{4}\right) < 0 < f\left(\frac{6}{4}\right), \text{ also } I_4 = [1.25, 1.5], \\ f\left(\frac{11}{8}\right) < 0 < f\left(\frac{12}{8}\right), \text{ also } I_5 = [1.375, 1.5], \\ f\left(\frac{22}{16}\right) < 0 < f\left(\frac{23}{16}\right), \text{ also } I_6 = [1.375, 1.4375]. \end{aligned}$$

Durch dieses *Bisektions-Verfahren* gewinnt man eine Intervallschachtelung für  $\sqrt{2}$ .

Schneller konvergiert die sogenannte *regula falsi*, die auf einer ähnlichen Idee beruht, aber komplizierter ist. Wir wollen hier nicht darauf eingehen.

Wichtig ist auch der folgende Satz vom Minimum und Maximum:

**I.4.12 Satz von Weierstraß.** Die Funktion  $f$  sei auf dem abgeschlossenen Intervall  $[a, b]$  stetig. Dann nimmt  $f$  in gewissen Punkten des Intervalls seinen kleinsten und seinen größten Wert an.

Daß diese Aussage nicht selbstverständlich ist, zeigt folgendes einfache Beispiel: die Funktion  $\frac{1}{x}$  ist auf dem offenen Intervall  $I = (0, 1)$  stetig, nimmt dort aber keinen größten Wert an.

Den Beweis des Satzes von Weierstraß lassen wir aus. Er benutzt keine wirklich neuen Ideen, ist aber technisch etwas komplizierter.

Als Anwendung untersuchen wir das Verhalten von Polynomen für  $x \rightarrow \pm\infty$ . Zunächst ist klar:

$$\lim_{x \rightarrow +\infty} x^m = +\infty \quad (\text{für alle } m \in \mathbb{N}),$$

$$\lim_{x \rightarrow -\infty} x^m = \begin{cases} +\infty & \text{falls } m \text{ gerade} \\ -\infty & \text{falls } m \text{ ungerade} \end{cases}.$$

**I.4.13 Satz.** Sei  $f(x) = a_m x^m + \dots + a_1 x + a_0$  ein Polynom vom Grad  $m$ . Ist  $a_m > 0$ , so verhält sich  $f$  für  $x \rightarrow \pm\infty$  wie  $x^m$ . Ist  $a_m < 0$ , so verhält sich  $f$  wie  $-x^m$ .

BEWEIS: Da  $a_m \neq 0$  ist, gibt es ein Polynom  $g$  vom Grad  $m - 1$ , so daß gilt:

$$f(x) = a_m x^m \cdot \left(1 + \frac{1}{x} \cdot g\left(\frac{1}{x}\right)\right).$$

Für  $x \rightarrow \pm\infty$  strebt  $\frac{1}{x}$  gegen Null. Da  $g$  auf ganz  $\mathbb{R}$  stetig ist, strebt dann  $1 + \frac{1}{x} \cdot g\left(\frac{1}{x}\right)$  gegen 1, also  $f(x)$  gegen den gleichen Wert wie  $a_m x^m$ . Daraus ergibt sich die Behauptung.  $\square$

### Beispiel:

Wir suchen nach Nullstellen eines Polynoms dritten Grades. Dabei können wir voraussetzen, daß das Polynom *normiert* ist, daß also der höchste Koeffizient = 1 ist:

$$\text{Sei } f(x) := x^3 + ax^2 + bx + c.$$

Für  $x \rightarrow \pm\infty$  verhält sich  $f(x)$  wie  $x^3$ . Der Grad ist ungerade, also gilt:

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

Beschränkt man die stetige Funktion  $f$  auf ein Intervall  $[a, b]$ , wo  $a \ll 0$  und  $b \gg 0$  ist, so kann man den Zwischenwertsatz benutzen und erhält ein  $\xi \in \mathbb{R}$  mit  $f(\xi) = 0$ . Mit Hilfe des Bisektions-Verfahrens läßt sich  $\xi$  noch näher bestimmen.

Mit Polynomdivision erhält man schließlich ein Polynom  $g(x)$  vom Grad 2, so daß  $f(x) = (x - \xi) \cdot g(x)$  ist.

Damit ist gezeigt: Ein Polynom dritten Grades besitzt immer eine Nullstelle. Darüber hinaus kann es noch zwei weitere Nullstellen oder eine weitere Nullstelle zweiter Ordnung geben, muß es aber nicht.

## D) Exponentialfunktion und Logarithmus

Es sei  $a > 0$  eine feste reelle Zahl. Dann wird durch

$$f : x \mapsto a^x$$

eine Funktion auf ganz  $\mathbb{R}$  definiert, die *Exponentialfunktion zur Basis  $a$* . Aus den Rechenregeln für Potenzen folgt:

1. Ist  $a = 1$ , so ist  $f(x) \equiv 1$ .

2. Ist  $a > 0$  beliebig, so ist

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

Man nennt das die *Funktionalgleichung* der Exponentialfunktion.

3. Es ist  $f(0) = 1$  und  $f(1) = a$ .

4. Ist  $a > 1$ , so ist  $f$  streng monoton wachsend.

Ist  $a < 1$ , so ist  $f$  streng monoton fallend.

In beiden Fällen ist  $f$  injektiv.

Einige Bemerkungen zum BEWEIS: Wir können uns auf den Fall  $a > 1$  beschränken, also  $a = 1 + h$  mit  $h > 0$ . Da  $(1 + h)^n \geq 1 + n \cdot h$  ist, ist dann  $a^n > 1$  für alle  $n \in \mathbb{N}$ . Daraus folgt sehr leicht, daß auch  $a^{1/n} = \sqrt[n]{a} > 1$  für alle  $n$  und schließlich  $a^q > 1$  für alle positiven rationalen Zahlen  $q$  ist. Mit Hilfe von Intervallschachtelungen überträgt sich das auf Exponenten  $x \in \mathbb{R}_+$ .

Ist nun  $x_1 < x_2$ , so ist  $x := x_2 - x_1 > 0$  und  $a^{x_2} = a^{x_1+x} = a^{x_1} \cdot a^x > a^{x_1}$ .

5.  $f : x \mapsto a^x$  bildet (für  $a \neq 1$ )  $\mathbb{R}$  surjektiv auf  $\mathbb{R}_+$  ab.

Der Beweis ist etwas technisch. Man geht vor wie beim Zwischenwertsatz, an Stelle der Stetigkeit benutzt man die strenge Monotonie. Daß keine Sprünge auftauchen können, liegt daran, daß die Werte  $a^q$  mit rationalem  $q$  genügend dicht liegen.

Das folgende Kriterium ist recht nützlich:

**I.4.14 Satz.** 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.

Ist sogar  $f(I) = J$ , so ist  $f$  auf ganz  $I$  stetig.

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.  $\square$

Hieraus folgt:

Die Funktion  $f : x \mapsto a^x$  ist überall stetig.

**Definition.**

Sei  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  die Exponentialfunktion  $x \mapsto a^x$ , mit  $a \neq 1$ .

Die Umkehrfunktion  $f^{-1} : \mathbb{R}_+ \rightarrow \mathbb{R}$  heißt *Logarithmus zur Basis  $a$* . Man schreibt:

$$f^{-1}(y) = \log_a(y).$$

Aus der Definition des Logarithmus folgt mit Hilfe der Eigenschaften der Exponentialfunktion:

**I.4.15 Satz.**

1.  $\log_a(a^x) = x$  und  $a^{\log_a(y)} = y$ .
2.  $\log_a(1) = 0$  und  $\log_a(a) = 1$ .
3.  $\log_a(u \cdot v) = \log_a(u) + \log_a(v)$  für  $u, v > 0$ .
4.  $\log_a\left(\frac{u}{v}\right) = \log_a(u) - \log_a(v)$  für  $u, v > 0$ .
5.  $\log_a(y^x) = x \cdot \log_a(y)$  für  $y > 0$ .

Zum BEWEIS:  $x := \log_a(y)$  ist die *Hochzahl*, mit der man  $a$  potenzieren muß, um  $y$  zu erhalten.

Z.B. wollen wir  $a^? = y^x$  lösen. Es ist aber  $a^{\log_a(y)} = y$ , also

$$y^x = (a^{\log_a(y)})^x = a^{x \cdot \log_a(y)}.$$

**I.4.16 Folgerung.** Sind  $a, b \in \mathbb{R}_+$ , beide  $\neq 1$ , so kann man den Logarithmus zur Basis  $a$  in den zur Basis  $b$  umrechnen: Für  $x > 0$  ist

$$\log_b(x) = \frac{1}{\log_a(b)} \cdot \log_a(x).$$

Insbesondere gilt:

$$\log_a(b) \cdot \log_b(a) = 1.$$

BEWEIS: Es ist

$$\log_b(x) \cdot \log_a(b) = \log_a(b^{\log_b(x)}) = \log_a(x).$$

Die zweite Behauptung ergibt sich, wenn man  $x := a$  setzt. □

Zum Beispiel ist

$$2^{-2} = 0.25, 2^{-1} = 0.5, 2^0 = 1, 2^1 = 2, 2^2 = 4, 2^3 = 8, \dots,$$

also

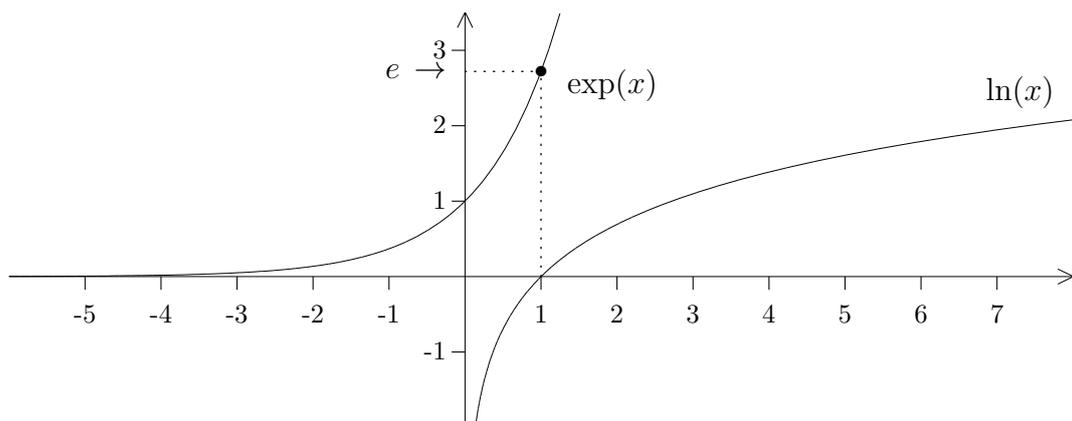
$$\log_2(0.25) = -2, \log_2(0.5) = -1, \log_2(1) = 0, \log_2(2) = 1, \log_2(4) = 2, \dots$$

Das ist ein einfaches Beispiel einer sogenannten „Logarithmentafel“. 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. Mit  $\lg(x)$  bezeichnet man den *Zehner-Logarithmus* oder *Briggschen Logarithmus*  $\log_{10}(x)$ .

In mathematisch-naturwissenschaftlichen Anwendungen interessiert eher der Fall der Basis  $e$ :

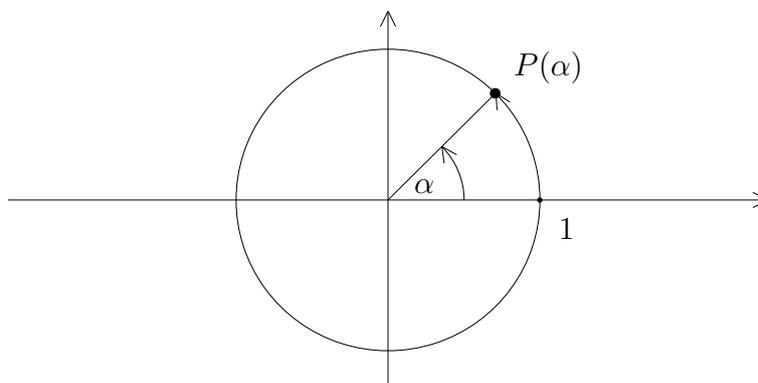
$$\begin{aligned} \exp(x) &:= e^x && \text{heißt } \textit{Exponentialfunktion}, \\ \ln(x) &:= \log_e(x) && \text{heißt } \textit{Natürlicher Logarithmus}. \end{aligned}$$

Die Graphen dieser Funktionen haben folgende Gestalt:



### E) Winkelfunktionen

Wir betrachten einen Punkt  $P$  im  $\mathbb{R}^2$ , der sich gegen den Uhrzeigersinn auf dem Rand des Einheitskreises bewegt:



Vom Ursprung aus sieht man  $P$  unter einem gewissen Winkel  $\alpha$ , gemessen von der positiven  $x$ -Achse aus. Der Winkel bewegt sich zwischen  $0^\circ$  und  $360^\circ$ . Mit  $arc(\alpha)$  (*Arkus von  $\alpha$* ) bezeichnet man die Länge des zurückgelegten Bogens. Sie ist proportional zum Winkel.

#### Definition.

Mit  $\pi$  bezeichnet man die Länge des halben Einheitskreisbogens.

Offensichtlich ist dann  $arc(\alpha) = \frac{\pi}{180} \cdot \alpha$ .

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

$$\pi = 3.141592653\dots$$

Die Ziffern 1, 1, 3, 3, 5, 5 liefern bei geeigneter Anordnung eine erstaunlich gute rationale Näherung von  $\pi$ :

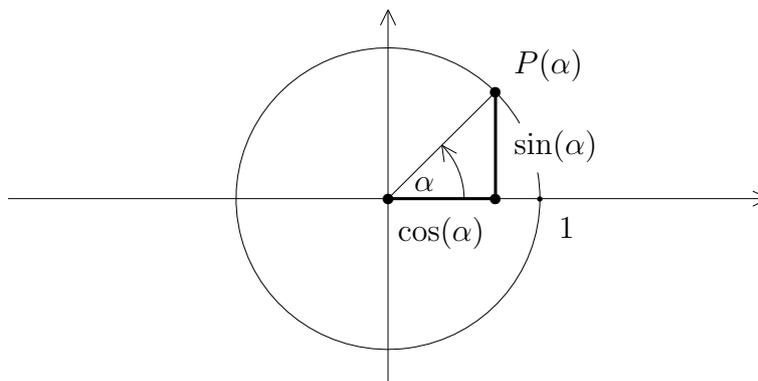
$$\frac{355}{113} = 3.1415929$$

Der Umfang eines Kreises vom Radius  $r$  ist die Zahl  $U = 2r\pi$ , seine Fläche die Zahl  $F = r^2\pi$ .

Hier ist noch eine Umrechnungstabelle zwischen Grad und Bogenmaß:

$\alpha$	$0^\circ$	$30^\circ$	$45^\circ$	$60^\circ$	$90^\circ$	$180^\circ$	$270^\circ$	$360^\circ$
$\text{arc}(\alpha)$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\pi$	$\frac{3\pi}{2}$	$2\pi$

Noch einmal betrachten wir  $P = P(\alpha) = (x, y)$  auf dem Einheitskreis.



### Definition.

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

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

Der Winkel darf dabei wahlweise im Grad- oder im Bogenmaß angegeben werden, d.h. es ist  $\cos(45^\circ) = \cos(\frac{\pi}{4})$ . Für mathematische Zwecke ist es aber günstiger, mit dem dimensionslosen Bogenmaß zu arbeiten. Dann sind Sinus und Cosinus auf dem Intervall  $[0, 2\pi]$  definiert. Läßt man jedoch auch mehrere Umdrehungen und Winkel in Uhrzeigerichtung zu, so kann man die Winkelfunktionen auf ganz  $\mathbb{R}$  definieren.

### Definition.

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

Offensichtlich sind die erweiterten Funktionen Sinus und Cosinus periodisch mit der Periode  $2\pi$ . Das gleiche gilt für davon abgeleitete Funktionen wie

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

### Definition.

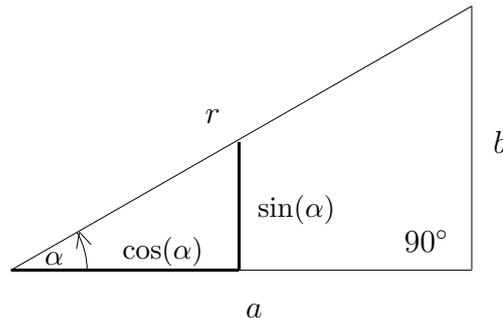
Eine Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}$  heißt *gerade*, falls  $f(-x) = f(x)$  für alle  $x \in \mathbb{R}$  ist. Sie heißt *ungerade*, falls  $f(-x) = -f(x)$  für alle  $x \in \mathbb{R}$  ist.

**Beispiel :**

$f(x) := x^2$  ist offensichtlich gerade, während  $f(x) := x^3$  ungerade ist.

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.

Um einige Werte der Winkelfunktionen zu berechnen, erinnern wir an eine andere Charakterisierung dieser Funktionen:



Wir betrachten ein rechtwinkliges Dreieck mit den Katheten  $a$  und  $b$  und der Hypotenuse  $r$ . Der Winkel zwischen  $a$  und  $r$  sei mit  $\alpha$  bezeichnet. Dann nennt man  $a$  auch die *Ankathete* zu  $\alpha$  und  $b$  die *Gegenkathete* zu  $\alpha$ .

Aus dem Strahlensatz folgt:

$$\frac{b}{\sin(\alpha)} = \frac{a}{\cos(\alpha)} = \frac{r}{1},$$

also

$$a = r \cdot \cos(\alpha) \text{ und } b = r \cdot \sin(\alpha).$$

Das bedeutet z.B., daß durch

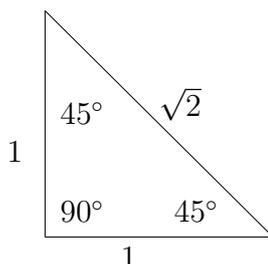
$$\varphi(t) := (x_0 + r \cdot \cos(t), y_0 + r \cdot \sin(t))$$

eine Parametrisierung  $\varphi : [0, 2\pi] \rightarrow \mathbb{R}^2$  des Kreises mit Radius  $r$  um  $(x_0, y_0)$  gegeben ist.

Außerdem lesen wir ab, daß – unabhängig von der Größe des Dreiecks – folgende Beziehung gilt:

$$\begin{aligned} \frac{a}{r} &= \frac{\text{Ankathete}}{\text{Hypotenuse}} = \cos(\alpha), \\ \frac{b}{r} &= \frac{\text{Gegenkathete}}{\text{Hypotenuse}} = \sin(\alpha). \end{aligned}$$

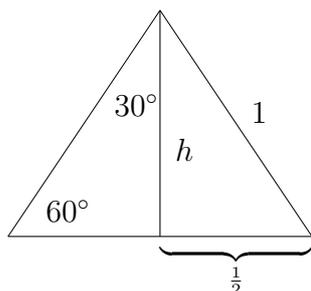
An einem gleichschenkelig-rechtwinkligen Dreieck mit Kathetenlänge 1 kann man jetzt ablesen:



Es ist

$$\begin{aligned} \sin\left(\frac{\pi}{4}\right) &= \cos\left(\frac{\pi}{4}\right) \\ &= \frac{1}{\sqrt{2}} = \frac{1}{2}\sqrt{2}. \end{aligned}$$

In einem gleichseitigen Dreieck mit Seitenlänge 1 und Höhe  $h$  folgt:



Es ist  $h^2 = 1^2 - (\frac{1}{2})^2 = \frac{3}{4}$ ,  
also  $h = \frac{1}{2}\sqrt{3}$ . Daraus folgt:

$$\sin\left(\frac{\pi}{6}\right) = \frac{1}{2} \text{ und } \cos\left(\frac{\pi}{6}\right) = \frac{1}{2}\sqrt{3}.$$

Bei  $\frac{\pi}{3}$  geht es umgekehrt.

Fassen wir die Ergebnisse in einer Tabelle zusammen:

$x$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\pi$
$\cos(x)$	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

Um weitere Werte zu ermitteln, braucht man die

**I.4.17 Additionstheoreme.** Für beliebige Winkel  $\alpha, \beta$  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.

**I.4.18 Folgerung.** Für beliebige Winkel  $\alpha$  gilt:

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

Aus der Definition von Sinus und Cosinus am Einheitskreis folgt außerdem:

**I.4.19 Satz.** Für alle  $x \in \mathbb{R}$  ist  $\sin(x)^2 + \cos(x)^2 = 1$ .

**I.4.20 Folgerung.**

1.  $\sin(\alpha + \frac{\pi}{2}) = \cos(\alpha)$  und  $\cos(\alpha + \frac{\pi}{2}) = -\sin(\alpha)$ .
2.  $\sin(\alpha + \pi) = -\sin(\alpha)$  und  $\cos(\alpha + \pi) = -\cos(\alpha)$ .
3. Für beliebiges  $\alpha$  ist  $|\cos(\alpha)| \leq 1$  und  $|\sin(\alpha)| \leq 1$ .

Es folgt insbesondere, daß die Graphen von Sinus und Cosinus gleich aussehen, sie sind nur um  $\frac{\pi}{2}$  gegeneinander verschoben. Und es folgt, daß Tangens und Cotangens sogar periodisch mit der Periode  $\pi$  sind.

**I.4.21 Satz.** Die Funktion  $\sin(x)$  ist auf  $[0, \frac{\pi}{2}]$  streng monoton wachsend.

Wir verzichten auf einen exakten BEWEIS, die Behauptung läßt sich anschaulich am Einheitskreis leicht überprüfen.

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.

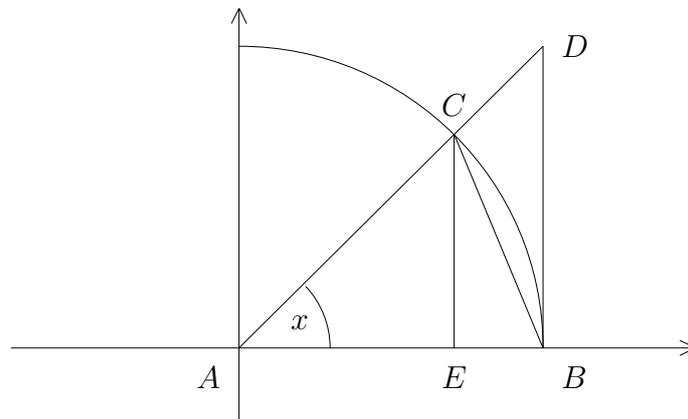
**I.4.22 Satz.** *Es ist*

$$\lim_{x \rightarrow 0} \sin(x) = 0, \quad \lim_{x \rightarrow 0} \cos(x) = 1, \quad \lim_{x \rightarrow 0} \frac{\sin(x)}{x} = 1$$

$$\text{und} \quad \lim_{x \rightarrow 0} \frac{1 - \cos(x)}{x} = 0.$$

BEWEIS: Die ersten beiden Aussagen sind offensichtlich richtig.

Da  $\frac{\sin(-x)}{-x} = \frac{\sin(x)}{x}$  ist, genügt es bei der dritten Aussage, den rechtsseitigen Grenzwert zu betrachten, und da es nur auf kleine positive Werte von  $x$  ankommt, können wir folgende Skizze benutzen:



Es sei  $AB = AC = 1$  und  $x := \angle(BAC)$ . Dann ist  $EC = \sin(x)$  und  $BD = \tan(x)$ .

Die Fläche des Dreiecks  $\triangle ABC$  beträgt  $\frac{1}{2} \cdot AB \cdot EC = \frac{\sin(x)}{2}$ ,

die Fläche des Kreissektors  $ABC$  beträgt  $\frac{x}{2\pi} \cdot \{\text{Fläche des Einheitskreises}\} = \frac{x}{2}$

und die Fläche des Dreiecks  $\triangle ABD$  beträgt  $\frac{1}{2} \cdot AB \cdot BD = \frac{\tan(x)}{2}$ .

Also erhält man folgende Ungleichungskette:

$$\frac{\sin(x)}{2} < \frac{x}{2} < \frac{\tan(x)}{2},$$

$$\text{und daher} \quad 0 < \sin(x) < x < \tan(x) = \frac{\sin(x)}{\cos(x)}.$$

Ist  $0 < x < \frac{\pi}{2}$ , so kann man durch  $\sin(x)$  dividieren und erhält:

$$1 < \frac{x}{\sin(x)} < \frac{1}{\cos(x)}, \text{ also } \cos(x) < \frac{\sin(x)}{x} < 1.$$

Da  $\lim_{x \rightarrow 0} \cos(x) = 1$  ist, folgt die 3. Aussage.

Für  $0 < x < \frac{\pi}{2}$  ist außerdem

$$\begin{aligned} 0 < 1 - \cos(x) &= (1 - \cos(x)) \cdot \frac{1 + \cos(x)}{1 + \cos(x)} \\ &= \frac{1 - \cos^2(x)}{1 + \cos(x)} = \frac{\sin^2(x)}{1 + \cos(x)} \\ &< \sin^2(x) < x^2, \end{aligned}$$

also

$$0 < \frac{1 - \cos(x)}{x} < x.$$

Das ergibt die 4. Grenzwertbeziehung. □

**I.4.23 Folgerung 1.** *Sinus und Cosinus sind auf ganz  $\mathbb{R}$  stetig.*

BEWEIS: Auf Grund des gerade bewiesenen Satzes sind  $\sin(x)$  und  $\cos(x)$  in  $x = 0$  stetig. Wir betrachten jetzt nur den Sinus, beim Cosinus geht's analog.

Sei  $x_0 \in \mathbb{R}$  beliebig,  $h := x - x_0$ . Dann gilt:

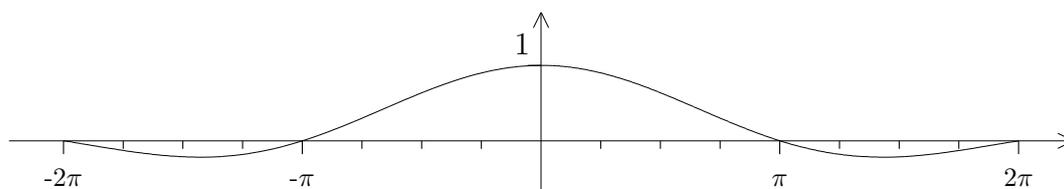
$$\sin(x) = \sin(x_0 + h) = \sin(x_0) \cos(h) + \cos(x_0) \sin(h).$$

Wenn  $(x_n)$  gegen  $x_0$  konvergiert, dann konvergiert  $h_n := x_n - x_0$  gegen 0, und wegen der Stetigkeit von  $\sin(x)$  und  $\cos(x)$  im Nullpunkt konvergiert  $\sin(x_n)$  gegen  $\sin(x_0)$ . Also ist der Sinus in  $x_0$  stetig. □

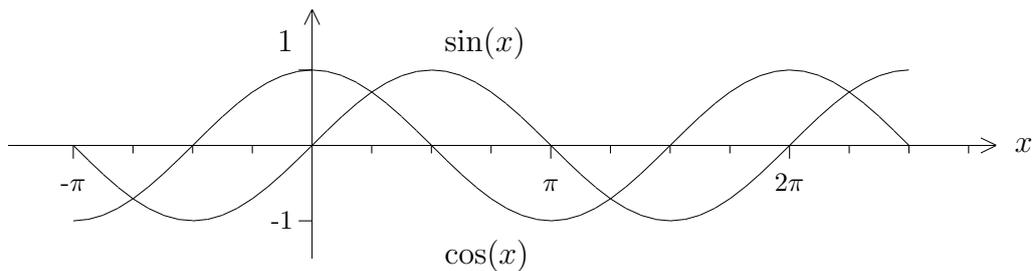
**I.4.24 Folgerung 2.**

Die Funktion  $f(x) := \begin{cases} \frac{\sin(x)}{x} & \text{falls } x \neq 0 \\ 1 & \text{falls } x = 0 \end{cases}$  ist überall stetig.

Der BEWEIS ist klar.

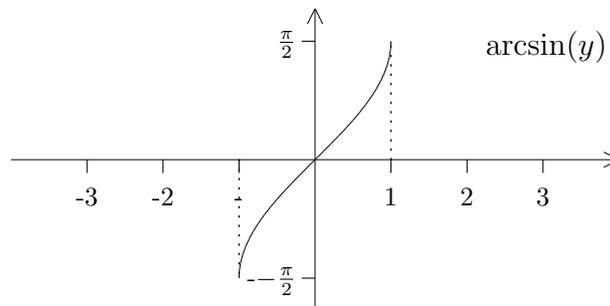


Für die Funktionsgraphen von Sinus und Cosinus ergibt sich nun folgendes Bild:



Als streng monoton wachsende Funktion ist  $\sin : [-\frac{\pi}{2}, +\frac{\pi}{2}] \rightarrow [-1, +1]$  bijektiv.

Die Umkehrabbildung  $\sin^{-1} : [-1, +1] \rightarrow [-\frac{\pi}{2}, +\frac{\pi}{2}]$  wird als (*Hauptwert des*) *Arcussinus* bezeichnet, kurz:  $\sin^{-1}(y) = \arcsin(y)$ . 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 ist der *Arcuscosinus*, die Funktion  $\arccos : [-1, 1] \rightarrow [0, \pi]$ .

Zwischen 0 und  $2\pi$  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:

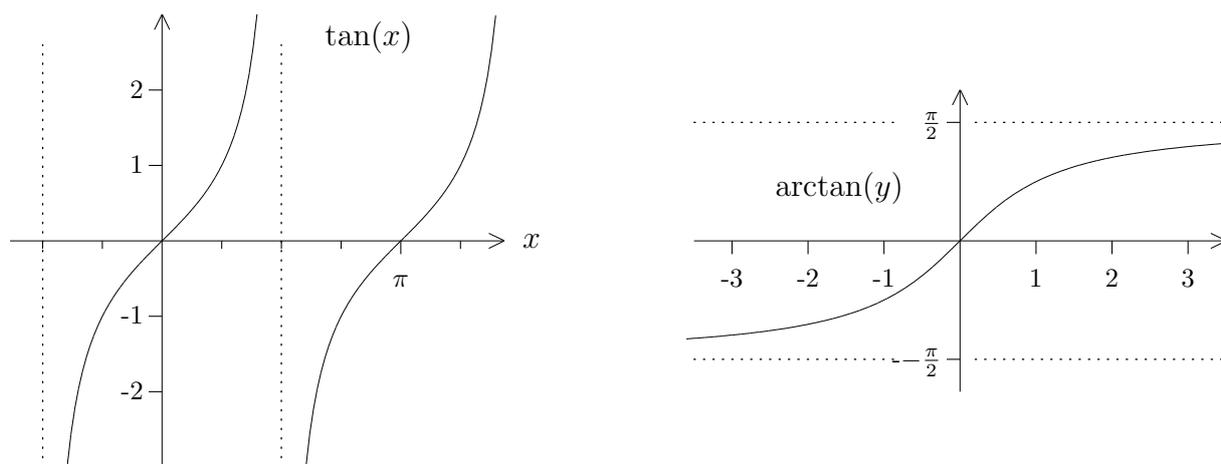
$$\lim_{x \rightarrow -\pi/2+} \tan(x) = -\infty \text{ und } \lim_{x \rightarrow +\pi/2-} \tan(x) = +\infty.$$

Man kann sich auch überlegen, daß

$$\tan : \left(-\frac{\pi}{2}, +\frac{\pi}{2}\right) \rightarrow \mathbb{R}$$

streng monoton wachsend und stetig ist, also bijektiv.

Die Umkehrfunktion  $\tan^{-1} : \mathbb{R} \rightarrow \left(-\frac{\pi}{2}, +\frac{\pi}{2}\right)$  ist der *Arcustangens*, kurz  $\arctan(x)$ . Die Graphen von Tangens und Arcustangens sehen folgendermaßen aus:



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

Wir betrachten einige Anwendungen der Winkelfunktionen:

1. Sei  $P = (x, y)$  ein Punkt der Ebene, aber nicht der Ursprung. Es sei

$$r(P) := d(P, O) = \sqrt{x^2 + y^2} > 0$$

und

$$u := \frac{x}{r} \text{ und } v := \frac{y}{r}.$$

Dann ist  $u^2 + v^2 = 1$ , der Punkt  $(u, v)$  liegt auf dem Einheitskreis und es gibt einen eindeutig bestimmten Winkel  $\alpha \in [0, 2\pi)$ , so daß  $u = \cos(\alpha)$  und  $v = \sin(\alpha)$  ist. Wir setzen

$$\arg(P) := \alpha \quad (\text{Argument von } \alpha).$$

Dann gilt:

$$P = (r(P) \cdot \cos(\arg(P)), r(P) \cdot \sin(\arg(P))).$$

Wir haben eine bijektive Beziehung zwischen den Punkten der Ebene (mit Ausnahme des Ursprungs) und den Paaren  $(r, \alpha)$  mit  $r > 0$  und  $\alpha \in [0, 2\pi)$ . Man nennt  $(r, \alpha)$  die *Polarkoordinaten* des Punktes  $P$ .

Die Zuordnung  $(r, \alpha) \mapsto P := (r \cdot \cos(\alpha), r \cdot \sin(\alpha))$  ist sehr einfach, aber wie sieht explizit die Umkehrabbildung aus?

$P = (x, y) \mapsto (r, \alpha)$  ist gegeben durch

$$r := \sqrt{x^2 + y^2}$$

und

$$\alpha := \begin{cases} \arctan\left(\frac{y}{x}\right) & \text{falls } x > 0 \text{ und } y \geq 0 \\ \frac{\pi}{2} & \text{falls } x = 0 \text{ und } y > 0 \\ \arctan\left(\frac{y}{x}\right) + \pi & \text{falls } x < 0 \\ \frac{3\pi}{2} & \text{falls } x = 0 \text{ und } y < 0 \\ \arctan\left(\frac{y}{x}\right) + 2\pi & \text{falls } x > 0 \text{ und } y < 0. \end{cases}$$

2. Gegeben sei eine schräge Gerade  $L = \{(x, y) \mid y = mx + b\}$ . Sind  $x_0, x \in \mathbb{R}$  mit  $x_0 < x$ , und  $y_0, y$  gegeben durch

$$y_0 := mx_0 + b \text{ und } y := mx + b,$$

so erhält man für die Steigung der Geraden:

$$m = \frac{y - y_0}{x - x_0}.$$

Aber  $x - x_0$  und  $y - y_0$  sind die Katheten eines rechtwinkligen Dreiecks, dessen Hypotenuse auf der Geraden  $L$  liegt. Ist  $\alpha$  der Winkel, den die Gerade mit der  $x$ -Achse einschließt, so ist offensichtlich  $m = \tan(\alpha)$ .

Ist also z.B. nach der Geraden gefragt, die mit der  $x$ -Achse einen Winkel von  $30^\circ$  einschließt, und die außerdem durch den Punkt  $(3, 1)$  geht, so hat man folgende Informationen:

Die Gerade ist durch eine Gleichung der Form  $y = mx + b$  gegeben, mit  $m = \frac{1}{3}\sqrt{3}$ , und es ist  $3m + b = 1$ , also  $b = 1 - 3m = 1 - \sqrt{3}$ . Zusammen ergibt das die Geradengleichung

$$y = \frac{1}{3}\sqrt{3} \cdot x + (1 - \sqrt{3}) = 1 + \sqrt{3}\left(\frac{x}{3} - 1\right).$$

3. Es seien  $A, \omega, \alpha$  feste reelle Zahlen. Dann wird die Funktion

$$f(t) := A \cdot \sin(\omega t + \alpha)$$

als eine *harmonische Sinus-Schwingung* bezeichnet. Dabei heißt  $\alpha$  die *Anfangsphase* (auch *Phasenkonstante* oder *Nullphasenwinkel*),  $A$  die *Amplitude* und  $\omega$  die *Kreisfrequenz*.

Es gilt nun: Eine Funktion  $f(t)$  ist genau dann eine harmonische Sinusschwingung mit der Kreisfrequenz  $\omega$ , wenn es reelle Konstanten  $a$  und  $b$  gibt, so daß gilt:

$$f(t) = a \cdot \sin(\omega t) + b \cdot \cos(\omega t).$$

Zum BEWEIS braucht man in der einen Richtung nur das Additionstheorem anzuwenden:

$$\begin{aligned} f(t) &= A \cdot [\sin(\omega t) \cos(\alpha) + \cos(\omega t) \sin(\alpha)] \\ &= a \cdot \sin(\omega t) + b \cdot \cos(\omega t), \end{aligned}$$

mit  $a := A \cdot \cos(\alpha)$  und  $b := A \cdot \sin(\alpha)$ .

Ist umgekehrt  $f$  von dieser Gestalt, so setze man  $A := \sqrt{a^2 + b^2}$ . Dann ist

$$\left(\frac{a}{A}\right)^2 + \left(\frac{b}{A}\right)^2 = \frac{a^2 + b^2}{A^2} = 1,$$

und es gibt ein  $\alpha$  mit  $\cos(\alpha) = \frac{a}{A}$  und  $\sin(\alpha) = \frac{b}{A}$ . Aber das bedeutet, daß  $f$  eine harmonische Sinusschwingung mit der Amplitude  $A$ , der Kreisfrequenz  $\omega$  und der Anfangsphase  $\alpha$  ist.

Als nächstes betrachten wir *zwei* harmonische Sinusschwingungen mit der *gleichen* Kreisfrequenz  $\omega$ :

$$f(t) := A_1 \sin(\omega t + \alpha_1) \text{ und } g(t) := A_2 \sin(\omega t + \alpha_2).$$

$h(t) := f(t) + g(t)$  ist die Schwingung, die durch *Überlagerung* (*Superposition*) von  $f$  und  $g$  entsteht. Nun kann man aber schreiben:

$$\begin{aligned} f(t) &= a_1 \sin(\omega t) + b_1 \cos(\omega t) \\ \text{und } g(t) &= a_2 \sin(\omega t) + b_2 \cos(\omega t), \end{aligned}$$

mit geeigneten Konstanten  $a_1, b_1, a_2, b_2$ . Und dann ist

$$f(t) + g(t) = (a_1 + a_2) \sin(\omega t) + (b_1 + b_2) \cos(\omega t),$$

und das ist wieder eine harmonische Sinusschwingung mit der Kreisfrequenz  $\omega$ .

Sinusschwingungen mit *verschiedenen* Kreisfrequenzen überlagern sich übrigens nicht mehr zu einer harmonischen Sinusschwingung!

## F) Hyperbelfunktionen

### Definition.

Die auf ganz  $\mathbb{R}$  definierten Funktionen

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

heißen *Sinus hyperbolicus* und *Cosinus hyperbolicus*.

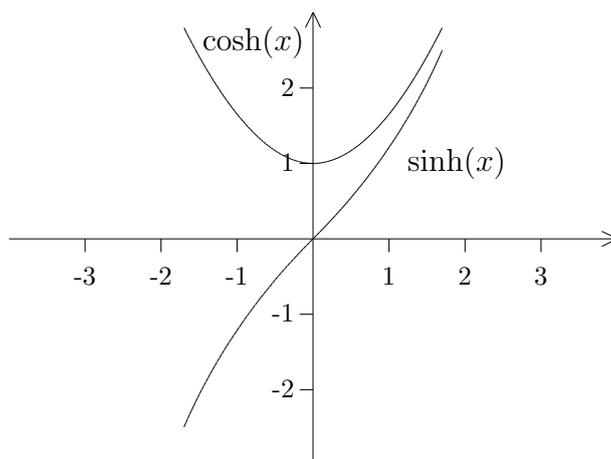
Offensichtlich ist  $\cosh(x)$  eine gerade und  $\sinh(x)$  eine ungerade Funktion. Weiter gilt:

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

sowie

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

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. Daher haben die Hyperbelfunktionen ihren Namen.

Schließlich definiert man noch:

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

## §5 Vektoren

Wir haben schon Paare als Elemente eines kartesischen Produktes kennengelernt. Was mit zwei Mengen möglich ist, geht auch mit drei, vier oder noch mehr Mengen:

Sind  $A_1, A_2, \dots, A_n$  irgendwelche Mengen, so kann man aus jeder dieser Mengen  $A_i$  jeweils ein Element  $x_i$  herausnehmen und dann diese  $n$  Elemente  $x_1, x_2, \dots, x_n$  zu einem neuen Gebilde

$$\mathbf{x} = (x_1, \dots, x_n)$$

zusammenfassen, einem sogenannten „ $n$ -Tupel“. Im Falle  $n = 2$  spricht man auch von einem *Paar*, im Falle  $n = 3$  von einem *Tripel* und im Falle  $n = 4$  von einem *Quadrupel*.

Die Menge aller  $n$ -Tupel  $\mathbf{x} = (x_1, \dots, x_n)$  mit  $x_i \in A_i$  bezeichnet man als *kartesisches Produkt* aller  $A_i$  (in Zeichen:  $A_1 \times A_2 \times \dots \times A_n$ ).

Ist  $A_1 = A_2 = \dots = A_n = A$ , so schreibt man:  $A^n := A_1 \times \dots \times A_n$ . Wir kennen das schon bei der Euklidischen Ebene  $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ , und entsprechend ist der  $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$  ein Modell für den Euklidischen Raum. Dadurch, daß jedem Punkt des Raumes ein Tripel  $(x_1, x_2, x_3)$  zugeordnet wird, führt man *kartesische Koordinaten* im Raum ein. Mit Hilfe dieser Koordinaten mißt man Länge, Breite und Höhe von räumlichen Gebilden.

Die Physiker arbeiten im Zusammenhang mit der Relativitätstheorie gerne im  $\mathbb{R}^4$ , die 4. Koordinate mißt dann die Zeit. Science-Fiction-Autoren sprechen manchmal über fünf- oder höherdimensionale Räume, und sie haben dann Schwierigkeiten mit der Interpretation der neu hinzugekommenen Koordinaten. Der Mathematiker braucht nicht zu interpretieren, und er hat deshalb auch keine Probleme, im  $n$ -dimensionalen Raum  $\mathbb{R}^n$  zu arbeiten. Anwenden läßt sich eine solche Theorie dann auch auf Gegebenheiten, wo man mit vielen voneinander unabhängigen Größen rechnen muß, die sich nicht mehr geometrisch deuten lassen.

Vorerst kehren wir zu  $\mathbb{R}^2$  und  $\mathbb{R}^3$  zurück und deuten die Komponenten der zugehörigen Elemente im herkömmlichen Sinne als ebene oder räumliche Koordinaten.

Neben den „statischen“ oder „skalaren“ Größen, die mit Hilfe reeller Zahlen gemessen werden, treten in der Physik auch „dynamische“ Größen auf, die durch eine positive Maßzahl **und** ihre Richtung charakterisiert werden. Sie werden *Vektoren* genannt. Typische Beispiele sind Kräfte oder Geschwindigkeiten. Um nun in der Mathematik ein adäquates Modell zu finden, versuchen wir, die Eigenschaften der physikalischen Vektoren möglichst genau zu analysieren:

1. Eine vektorielle Größe hat eine wohlbestimmte *Richtung*, in die sie wirkt.
2. Sie hat einen meßbaren Wert, die Maßzahl ist stets eine nicht-negative reelle Zahl.
3. Wenn man ein kartesisches Koordinatensystem zugrundelegt und dieses parallel verschiebt, dann ändern sich die Meßwerte nicht.

Ein *Pfeil* oder eine *gerichtete Strecke* haben Größe (Länge) und Richtung. Zur Festlegung genügen zwei Punkte, nämlich Anfangs- und Endpunkt. In der Ebene bräuchte man dazu

$2 \times 2 = 4$  Koordinatenwerte, und im Raum  $2 \times 3 = 6$ . Diese Werte ändern sich, wenn man das Koordinatensystem parallel verschiebt. Invariant bleiben jedoch die Koordinatendifferenzen. Man kann also den Ursprung des Koordinatensystems auf den Anfangspunkt des Pfeils legen. Zur Festlegung des Pfeils genügt dann die Angabe der zwei (bzw. drei) Koordinaten seines Endpunktes.

Es bieten sich nun verschiedene mathematische Modelle für den Vektorbegriff an:

1. Ein *Vektor* ist eine unendliche Schar von gleichlangen und gleichgerichteten Pfeilen (oder gerichteten Strecken). Jeder Punkt der Ebene (bzw. des Raumes) kommt als Anfangspunkt eines dieser Pfeile vor.
2. Ein *Vektor* ist eine Translation  $T$  der Ebene (bzw. des Raumes). Für jeden Punkt  $X$  gibt  $T(X)$  den Endpunkt des in  $X$  angreifenden Vektors an.
3. Ein *Vektor* ist eine aus zwei (bzw. drei) Komponenten bestehende Größe

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad \text{oder} \quad \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}.$$

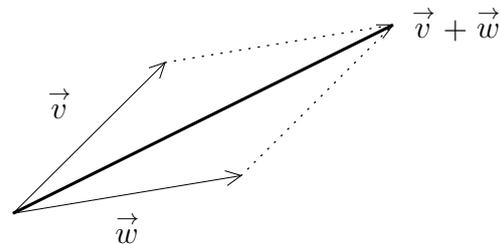
Das Paar  $(v_1, v_2)$  (bzw. das Tripel  $(v_1, v_2, v_3)$ ) bezeichnet jeweils den Endpunkt des am Ursprung angreifenden Vektors.

Welches dieser Modelle zeigt nun den „wahren“ Vektor? Tatsache ist, daß alle drei Modelle Inkarnationen ein und desselben Begriffes sind, und wir können uns von Fall zu Fall jeweils die passendste Möglichkeit aussuchen. Zum Rechnen wird man meistens die dritte wählen, den „Spaltenvektor“. Die anderen Darstellungsarten helfen aber dabei, Eigenschaften von Vektoren zu ermitteln.

Wir arbeiten jetzt mit Vektoren im Raum, es läßt sich aber alles auch auf ebene Vektoren übertragen:

Für  $\mathbf{v} := (v_1, v_2, v_3) \in \mathbb{R}^3$  sei  $\vec{v}$  der zugehörige Spaltenvektor und  $T_{\mathbf{v}}$  die durch  $T_{\mathbf{v}}(x_1, x_2, x_3) = (x_1 + v_1, x_2 + v_2, x_3 + v_3)$  gegebene Translation.

Bei der Beschreibung durch Translationen werden die Vektoreigenschaften am deutlichsten. So wie sich zwei Translationen durch Hintereinanderausführung miteinander verknüpfen lassen, so lassen sich zwei Vektoren  $\vec{v}$  und  $\vec{w}$  zu einem resultierenden Vektor  $\vec{v} + \vec{w}$  verknüpfen. Anschaulich erfolgt diese *Addition von Vektoren* nach dem Prinzip vom Parallelogramm der Kräfte.



Das bedeutet für Spaltenvektoren:

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} := \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ v_3 + w_3 \end{pmatrix},$$

und für Translationen:

$$T_{\mathbf{v}} + T_{\mathbf{w}} := T_{\mathbf{w}} \circ T_{\mathbf{v}}.$$

Da für die Verknüpfung von Abbildungen stets das Assoziativgesetz gilt, ist auch die Vektoraddition assoziativ. Weiter gibt es ein neutrales Element, den *Nullvektor*

$$\vec{0} := \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \text{ bzw. } T_{\mathbf{o}} := \text{id}.$$

Offensichtlich ist  $\vec{v} + \vec{0} = \vec{0} + \vec{v} = \vec{v}$  für jeden Vektor  $\vec{v}$ .

Ein Inverses zu  $\vec{v}$  ist der *negative* oder *entgegengesetzte* Vektor

$$-\vec{v} := \begin{pmatrix} -v_1 \\ -v_2 \\ -v_3 \end{pmatrix}, \text{ bzw. } -T_{\mathbf{v}} := (T_{\mathbf{v}})^{-1}.$$

Insgesamt haben wir nachgewiesen:

*Die Menge  $\mathcal{V}_3$  der räumlichen Vektoren (und genauso die Menge  $\mathcal{V}_2$  der ebenen Vektoren) bildet eine Gruppe.*

An Stelle von Spaltenvektoren hätte man genauso auch „Zeilenvektoren“ (also gewöhnliche Paare oder Tripel) benutzen können:

$$\begin{aligned} (v_1, v_2, v_3) + (w_1, w_2, w_3) &:= (v_1 + w_1, v_2 + w_2, v_3 + w_3), \\ -(v_1, v_2, v_3) &:= (-v_1, -v_2, -v_3). \end{aligned}$$

Dann gilt:

$$T_{\mathbf{v}} + T_{\mathbf{w}} = T_{\mathbf{v}+\mathbf{w}} \text{ und } -T_{\mathbf{v}} = T_{-\mathbf{v}}.$$

Wir haben an Beispielen gesehen, daß in Gruppen das kommutative Gesetz nicht zu gelten braucht. Da aber bei den Vektoren die Verknüpfung auf die Addition von Zahlen zurückgeführt wird, gilt:

*Die (additive) Gruppe der Vektoren ist kommutativ.*

Die Menge der Vektoren ist nicht nur eine Gruppe, sie hat eine noch viel reichhaltigere Struktur. Es ist z.B. anschaulich klar, daß man einen Vektor um einen gewissen reellen Betrag strecken kann:

$$\text{Ist } \alpha \in \mathbb{R} \text{ und } \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \text{ so setzt man } \alpha \cdot \vec{v} := \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \\ \alpha v_3 \end{pmatrix}.$$

Definiert man noch  $\alpha \cdot (v_1, v_2, v_3) := (\alpha v_1, \alpha v_2, \alpha v_3)$ , so ist  $\alpha \cdot T_{\mathbf{v}} = T_{\alpha \cdot \mathbf{v}}$ .

Man kann leicht nachrechnen, daß diese Multiplikation zwischen Skalaren und Vektoren folgende Eigenschaften hat:

1.  $1 \cdot \vec{v} = \vec{v}$ .

2.  $\alpha \cdot (\beta \cdot \vec{v}) = (\alpha\beta) \cdot \vec{v}$ .
3.  $(\alpha + \beta) \cdot \vec{v} = \alpha \cdot \vec{v} + \beta \cdot \vec{v}$ .
4.  $\alpha \cdot (\vec{v} + \vec{w}) = \alpha \cdot \vec{v} + \alpha \cdot \vec{w}$ .

Hieraus folgt z.B. (ohne Rückgriff auf die Definition mit Hilfe von Spaltenvektoren):

$0 \cdot \vec{v} = (0 + 0) \cdot \vec{v} = 0 \cdot \vec{v} + 0 \cdot \vec{v}$ , also  $0 \cdot \vec{v} = \vec{0}$ . Man beachte, daß hier zweierlei Nullen auftreten!

Die Multiplikation mit einem Skalar  $\alpha < 0$  hat übrigens folgende anschauliche Bedeutung: Der Vektor wird um den Faktor  $|\alpha|$  gestreckt (bzw. gestaucht), und dann wird zusätzlich seine Richtung umgedreht. Zur Abkürzung schreiben wir noch  $\vec{v} - \vec{w} := \vec{v} + (-\vec{w})$ .

Man beachte aber, daß man Vektoren nicht miteinander multiplizieren und erst recht sie nicht durcheinander dividieren kann!

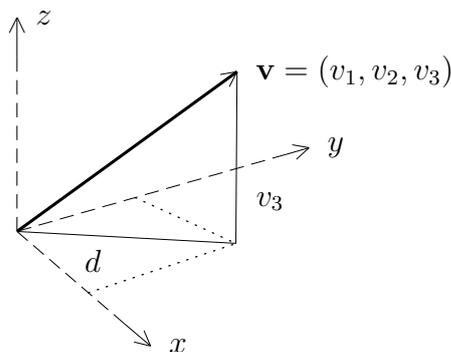
Zu einem Vektor gehört neben seiner Richtung auch seine Größe. Bis jetzt haben wir aber noch keinen Größenbegriff für Vektoren eingeführt. Dazu nehmen wir wieder die geometrische Anschauung zu Hilfe:

Wir beginnen mit dem einfachsten Fall, den ebenen Vektoren:

Der Vektor  $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$  wird anschaulich durch einen Pfeil vom Ursprung  $(0, 0)$  zum Punkt  $\mathbf{v} = (v_1, v_2)$  dargestellt. Die Länge dieses Pfeils wird nach dem Satz des Pythagoras durch die Zahl

$$\|\vec{v}\| := \sqrt{(v_1)^2 + (v_2)^2}$$

angegeben. Im  $\mathbb{R}^3$  ist die Situation ähnlich:



Wird der Vektor  $\vec{v}$  durch einen Pfeil von  $(0, 0, 0)$  nach  $(v_1, v_2, v_3)$  repräsentiert und ist  $d$  die Länge der Projektion von  $\vec{v}$  auf die x-y-Ebene, so gilt:  $d = \sqrt{(v_1)^2 + (v_2)^2}$ .

Aber die Länge  $\|\vec{v}\|$  des Vektors  $\vec{v}$  tritt als Hypotenuse eines rechtwinkligen Dreiecks mit den Katheten  $d$  und  $v_3$  auf, es ist also

$$\|\vec{v}\| = \sqrt{d^2 + (v_3)^2} = \sqrt{(v_1)^2 + (v_2)^2 + (v_3)^2}.$$

Die Zahl  $\|\vec{v}\|$  nennt man die *Länge* oder *Norm* des (räumlichen) Vektors  $\vec{v}$ . Man überlegt sich leicht:

1.  $\|\vec{v}\| \geq 0$  für alle Vektoren  $\vec{v}$ .

$$2. \quad \|\vec{v}\| = 0 \iff \vec{v} = \vec{0}.$$

$$3. \quad \|\alpha \cdot \vec{v}\| = |\alpha| \cdot \|\vec{v}\|.$$

Außerdem gilt

### I.5.1 Die Dreiecksungleichung.

$$\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|.$$

Zum BEWEIS brauchen wir einen Hilfssatz:

### I.5.2 Schwarzsche Ungleichung. $a_1, \dots, a_n$ und $b_1, \dots, b_n$ seien reelle Zahlen.

$$\text{Dann ist} \quad \left( \sum_{i=1}^n a_i b_i \right)^2 \leq \left( \sum_{i=1}^n a_i^2 \right) \cdot \left( \sum_{i=1}^n b_i^2 \right).$$

BEWEIS: Sei

$$A := \sum_{i=1}^n a_i^2, \quad B := \sum_{i=1}^n b_i^2 \quad \text{und} \quad C := \sum_{i=1}^n a_i b_i.$$

Ist  $B = 0$ , so ist  $b_1 = \dots = b_n = 0$ , und die Ungleichung ist trivialerweise erfüllt. Wir nehmen also an, daß  $B > 0$  ist. Dann gilt:

$$\begin{aligned} 0 &\leq \sum_{i=1}^n (Ba_i - Cb_i)^2 \\ &= B^2 \cdot \sum_{i=1}^n a_i^2 + C^2 \cdot \sum_{i=1}^n b_i^2 - 2BC \sum_{i=1}^n a_i b_i \\ &= B^2 A + C^2 B - 2BC^2 \\ &= B^2 A - C^2 B = B(AB - C^2). \end{aligned}$$

Also ist  $C^2 \leq AB$ , und das war zu zeigen. □

Die **Dreiecksungleichung** folgt nun leicht:

$$\begin{aligned} \|\vec{v} + \vec{w}\|^2 &= \sum_{i=1}^3 (v_i + w_i)^2 \\ &= \sum_{i=1}^3 v_i^2 + \sum_{i=1}^3 w_i^2 + 2 \sum_{i=1}^3 v_i w_i \\ &\leq \sum_{i=1}^3 v_i^2 + \sum_{i=1}^3 w_i^2 + 2\sqrt{(\sum_{i=1}^3 v_i^2)(\sum_{i=1}^3 w_i^2)} \\ &= \left( \sqrt{\sum_{i=1}^3 v_i^2} + \sqrt{\sum_{i=1}^3 w_i^2} \right)^2 \\ &= (\|\vec{v}\| + \|\vec{w}\|)^2. \end{aligned}$$

Ein Vektor der Länge 1 wird auch als *Einheitsvektor* bezeichnet. Durch *Normierung* kann man aus jedem Vektor  $\vec{v} \neq \vec{0}$  einen Einheitsvektor machen, nämlich

$$\vec{v}^0 := \frac{1}{\|\vec{v}\|} \cdot \vec{v}.$$

Hier sollte eine **Warnung** ausgesprochen werden:

Die Norm eines Vektors ist eine reelle Zahl und kann daher auch im Nenner auftauchen. Der Vektor selbst darf jedoch nie im Nenner stehen!!!

Im Folgenden bezeichne  $V$  eine der Mengen  $\mathcal{V}_2$  oder  $\mathcal{V}_3$ . Wir nennen eine solche Menge von Vektoren auch einen *Vektorraum*.

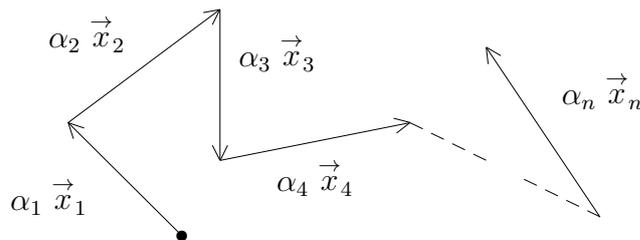
**Definition.**

Sind  $\vec{x}_1, \dots, \vec{x}_n$  Elemente von  $V$  und  $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ , so nennt man

$$\sum_{i=1}^n \alpha_i \cdot \vec{x}_i := \alpha_1 \cdot \vec{x}_1 + \dots + \alpha_n \cdot \vec{x}_n$$

eine *Linearkombination* der Vektoren  $\vec{x}_1, \dots, \vec{x}_n$ .

Eine Linearkombination von endlich vielen Vektoren aus  $V$  liegt wieder in  $V$ . Man spricht auch von „Vektorketten“ oder „Vektorzügen“. Die jeweils um die Faktoren  $\alpha_i$  gestreckten Vektoren  $\vec{x}_i$  werden sukzessive aneinandergelagert und der Pfeil vom Anfang dieser Vektorkette bis zu ihrer Spitze repräsentiert das Ergebnis.



**Beispiel:**

Seien  $\vec{x}_1 := \begin{pmatrix} 7 \\ -2 \end{pmatrix}$ ,  $\vec{x}_2 := \begin{pmatrix} -1 \\ -1 \end{pmatrix}$  und  $\vec{x}_3 := \begin{pmatrix} -4 \\ 1 \end{pmatrix}$  Vektoren aus  $\mathcal{V}_2$ . Dann ist

$$3 \cdot \vec{x}_1 - 5 \cdot \vec{x}_2 + 8 \cdot \vec{x}_3 = \begin{pmatrix} 21 \\ -6 \end{pmatrix} + \begin{pmatrix} 5 \\ 5 \end{pmatrix} + \begin{pmatrix} -32 \\ 8 \end{pmatrix} = \begin{pmatrix} -6 \\ 7 \end{pmatrix}.$$

**Definition.**

Ist  $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ , so nennt man  $\mathbf{x}^\top := \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$  den zugehörigen *Ortsvektor*.

Da die Zuordnung  $\mathbf{x} \mapsto \mathbf{x}^\top$  eine bijektive Abbildung  $\mathbb{R}^3 \rightarrow \mathcal{V}_3$  induziert, ist es möglich, die Geometrie des Raumes mit vektoriellen Mitteln weiterzutreiben. Und entsprechendes gilt natürlich auch für die Ebene!

Sei etwa  $L \subset \mathbb{R}^2$  die Gerade durch die beiden verschiedenen Punkte  $\mathbf{a} = (a_1, a_2)$  und  $\mathbf{b} = (b_1, b_2)$ . Wir haben schon die Parameterdarstellung kennengelernt:

$$L = \{(x_1, x_2) \in \mathbb{R}^2 \mid \exists t \in \mathbb{R} \text{ mit } x_1 = a_1 + t(b_1 - a_1) \text{ und } x_2 = a_2 + t(b_2 - a_2)\}.$$

Vektoriell geschrieben ist das die Menge

$$L = \{\mathbf{x} \in \mathbb{R}^2 \mid \exists t \in \mathbb{R} \text{ mit } \mathbf{x} = \mathbf{a} + t(\mathbf{b} - \mathbf{a})\}$$

oder

$$L = \{\vec{x} \in \mathcal{V}_2 \mid \exists t \in \mathbb{R} \text{ mit } \vec{x} = \vec{a} + t(\vec{b} - \vec{a})\}.$$

Man spricht dann von der vektoriellen Zwei-Punkte-Form der Geraden. Sie läßt sich unmittelbar auf den räumlichen Fall übertragen.

Den Vektor  $\vec{v} := (b_1 - a_1, b_2 - a_2)^\top$  bezeichnen wir als *Richtungsvektor* von  $L$ . Nach Voraussetzung ist er immer  $\neq \vec{0}$ . Mit den Ortsvektoren  $\vec{a} = (a_1, a_2)^\top$  und  $\vec{x} = (x_1, x_2)^\top$  kann man schreiben:

$$L = \{\vec{x} \in \mathcal{V}_2 \mid \exists t \in \mathbb{R} \text{ mit } \vec{x} = \vec{a} + t \cdot \vec{v}\}.$$

Das nennt man die Punkt-Richtungs-Form der Geraden. Auch bei ihr kann  $\mathcal{V}_2$  ohne weiteres durch  $\mathcal{V}_3$  ersetzt werden.

**Definition.**

1. Zwei Vektoren  $\vec{x} = \mathbf{x}^\top$  und  $\vec{y} = \mathbf{y}^\top$  der Ebene oder des Raumes heißen *kollinear*, wenn  $\mathbf{x}$  und  $\mathbf{y}$  auf der gleichen Geraden durch den Ursprung liegen. (Sie unterscheiden sich dann höchstens um einen reellen Faktor)
2. Zwei Geraden heißen *parallel*, falls ihre Richtungsvektoren kollinear sind. Sind sie disjunkt, aber nicht parallel, so nennt man sie *windschief*.

Windschiefe Geraden gibt es erst ab der Dimension 3.

**Definition.**

1. Sei  $\vec{a} \in \mathcal{V}_3$ . Weiter seien  $\vec{v}, \vec{w} \in \mathcal{V}_3$  *nicht* kollinear. Dann heißt

$$E := \{\vec{x} \in \mathcal{V}_3 \mid \exists s, t \in \mathbb{R} \text{ mit } \vec{x} = \vec{a} + s \cdot \vec{v} + t \cdot \vec{w}\}$$

eine *Ebene* in  $\mathcal{V}_3$ .

2. Drei (oder mehr) Vektoren in  $\mathcal{V}_3$  heißen *komplanar*, falls sie in der gleichen Ebene durch  $\vec{0}$  liegen.

Sind  $E_1$  (bzw.  $E_2$ ) zwei Ebenen mit den „Richtungsvektoren“  $\vec{v}_1, \vec{w}_1$  (bzw.  $\vec{v}_2, \vec{w}_2$ ), so gibt es zwei Möglichkeiten:

Sind  $\vec{v}_1, \vec{w}_1, \vec{v}_2$  und  $\vec{w}_2$  allesamt komplanar, so sind die Ebenen „parallel“. Andernfalls schneiden sich  $E_1$  und  $E_2$  längs einer Geraden.

Wir haben vorhin Vektorketten betrachtet. Wir wollen nun untersuchen, was passiert, wenn die Spitze einer Vektorkette wieder genau auf ihren Anfang trifft. Man spricht dann von einer „geschlossenen Vektorkette“.

Es seien  $\vec{x}_1, \dots, \vec{x}_n$  Vektoren  $\neq \vec{0}$ , und es gebe reelle Zahlen  $\alpha_1, \dots, \alpha_n$ , so daß gilt:

$$\sum_{i=1}^n \alpha_i \vec{x}_i = \vec{0}.$$

Zwei Fälle sind denkbar:

1) Es ist  $\alpha_1 = \dots = \alpha_n = 0$ . Das ist natürlich höchst langweilig. Wir sprechen dann vom „trivialen Fall“.

2) Nicht alle Koeffizienten  $\alpha_i$  sind  $= 0$ .

Wann kann dieser Fall überhaupt eintreten? Wir untersuchen das für verschiedene  $n \in \mathbb{N}$ .

$n = 1$ : Ist  $\alpha \cdot \vec{x} = \vec{0}$  und  $\vec{x} \neq \vec{0}$ , so gibt es eine Komponente  $x_i \neq 0$  von  $\vec{x}$  mit  $\alpha \cdot x_i = 0$ . Dann muß  $\alpha = 0$  sein. Der nicht-triviale Fall kann hier garnicht eintreten.

$n = 2$ : Sei  $\alpha \cdot \vec{x} + \beta \cdot \vec{y} = \vec{0}$ . Ist etwa  $\beta \neq 0$ , so ist  $\vec{y} = -\frac{\alpha}{\beta} \cdot \vec{x}$  ein Vielfaches von  $\vec{x}$ . Ist dagegen  $\alpha \neq 0$ , so kann man  $\vec{x}$  als Vielfaches von  $\vec{y}$  schreiben. In beiden Fällen ist einer der Vektoren eine (zugegebenermaßen simple) Linearkombination des anderen. Man sagt dann, die Vektoren sind *linear abhängig*.

*Zwei Vektoren sind also genau dann linear abhängig, wenn sie kollinear sind.*

$n = 3$ : Es sei  $\alpha_1 \cdot \vec{x}_1 + \alpha_2 \cdot \vec{x}_2 + \alpha_3 \cdot \vec{x}_3 = \vec{0}$ .

Ist  $\alpha_3 = 0$ , so liegt in Wirklichkeit der Fall  $n = 2$  vor. Deshalb können wir voraussetzen, daß  $\alpha_3 \neq 0$  ist. Aber dann kann die Gleichung nach  $\vec{x}_3$  aufgelöst werden:

$$\vec{x}_3 = \left(-\frac{\alpha_1}{\alpha_3}\right) \cdot \vec{x}_1 + \left(-\frac{\alpha_2}{\alpha_3}\right) \cdot \vec{x}_2 .$$

Man kann  $\vec{x}_3$  linear aus  $\vec{x}_1$  und  $\vec{x}_2$  kombinieren. Wenn eine solche Abhängigkeit zwischen den drei Vektoren besteht, nennt man sie auch hier wieder *linear abhängig*. Geometrisch gesehen liegen die drei Vektoren dann auf der gleichen Ebene durch den Ursprung. Also gilt:

*Drei Vektoren sind genau dann linear abhängig, wenn sie komplanar sind.*

Unabhängig voneinander sind drei Vektoren also nur dann, wenn sich der Nullvektor  $\vec{0}$  aus ihnen nur auf triviale Weise linear kombinieren läßt.

Diese Fallstudie motiviert die folgende Definition:

**Definition.**

Ein System von Vektoren  $\vec{x}_1, \dots, \vec{x}_n$  heißt *linear abhängig*, falls es reelle Zahlen  $\alpha_1, \dots, \alpha_n$  gibt, die nicht alle  $= 0$  sind, so daß gilt:

$$\sum_{i=1}^n \alpha_i \cdot \vec{x}_i = \vec{0} .$$

Andernfalls nennt man die Vektoren *linear unabhängig*.

**Beispiele :**

1. Ein einzelner Vektor  $\neq \vec{0}$  ist immer linear unabhängig. Kommen mehrere Vektoren zusammen, so können Sie – als System betrachtet – dennoch linear abhängig sein.
2. Die sogenannten *Einheitsvektoren*

$$\vec{e}_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad \vec{e}_3 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

bilden ein linear unabhängiges System in  $\mathcal{V}_3$ :

Ist nämlich

$$\vec{0} = \sum_{i=1}^3 \alpha_i \cdot \vec{e}_i = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix},$$

so muß  $\alpha_1 = \alpha_2 = \alpha_3 = 0$  sein.

3. Wir behaupten:

$$\vec{x}_1 := \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{und} \quad \vec{x}_2 := \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{sind linear unabhängig im } \mathbb{R}^2.$$

Zum Beweis nehmen wir wieder an, es gäbe reelle Zahlen  $\alpha_1$  und  $\alpha_2$ , so daß

$$\alpha_1 \cdot \vec{x}_1 + \alpha_2 \cdot \vec{x}_2 = \vec{0}$$

ist. Dann gilt:

$$\begin{pmatrix} \alpha_1 + \alpha_2 \\ \alpha_1 - \alpha_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{also } \alpha_1 + \alpha_2 = 0 \quad \text{und} \quad \alpha_1 - \alpha_2 = 0.$$

Daraus folgt sofort, daß  $\alpha_1 = \alpha_2 = 0$  sein muß.

4. Die Vektoren  $\vec{x}_1 := \begin{pmatrix} 6 \\ 3 \end{pmatrix}$  und  $\vec{x}_2 := \begin{pmatrix} -10 \\ -5 \end{pmatrix}$  sind linear abhängig, denn es ist  $\vec{x}_2 = -\frac{5}{3} \cdot \vec{x}_1$ .

Die Einheitsvektoren  $\vec{e}_1, \vec{e}_2, \vec{e}_3 \in \mathcal{V}_3$  sind nicht nur linear unabhängig, sie haben noch eine weitere bemerkenswerte Eigenschaft: Ist nämlich  $\vec{x} \in \mathcal{V}_3$  ein beliebiger Vektor mit den Komponenten  $x_1, x_2, x_3$ , so kann man auch schreiben:

$$\vec{x} = x_1 \cdot \vec{e}_1 + x_2 \cdot \vec{e}_2 + x_3 \cdot \vec{e}_3.$$

**Jeder** Vektor in  $\mathcal{V}_3$  läßt sich aus den Einheitsvektoren linear kombinieren!

Natürlich gelten sinngemäße Aussagen auch in  $\mathcal{V}_2$ .

Für die Ebene wollen wir eine noch allgemeinere Aussage beweisen:

**I.5.3 Satz.** Die Vektoren  $\vec{x}_1, \vec{x}_2 \in \mathcal{V}_2$  seien linear unabhängig. Dann läßt sich jeder andere Vektor  $\vec{x} \in \mathcal{V}_2$  aus  $\vec{x}_1$  und  $\vec{x}_2$  linear kombinieren.

BEWEIS: Sei  $\vec{x}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$  und  $\vec{x}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ , und  $\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{V}_2$  sei beliebig vorgegeben. Da  $\vec{x}_1$  und  $\vec{x}_2$  linear unabhängig sind, müssen sie insbesondere beide  $\neq \vec{0}$  sein. Wir suchen reelle Zahlen  $\alpha, \beta$ , so daß  $\vec{x} = \alpha \cdot \vec{x}_1 + \beta \cdot \vec{x}_2$  ist. Das bedeutet, daß wir das folgende lineare Gleichungssystem (mit den Unbekannten  $\alpha$  und  $\beta$ ) lösen müssen:

$$\begin{aligned} \text{(I)} \quad x_1\alpha + x_2\beta &= x, \\ \text{(II)} \quad y_1\alpha + y_2\beta &= y. \end{aligned}$$

Wäre die Determinante  $d := x_1y_2 - x_2y_1 = 0$ , so gäbe es ein  $t \in \mathbb{R}$ , so daß  $(y_1, y_2) = t \cdot (x_1, x_2)$  ist. Aber dann wäre  $\vec{x}_1 = x_1 \cdot \begin{pmatrix} 1 \\ t \end{pmatrix}$  und  $\vec{x}_2 = x_2 \cdot \begin{pmatrix} 1 \\ t \end{pmatrix}$ , d.h.  $\vec{x}_1$  und  $\vec{x}_2$  wären linear abhängig. Das widerspricht der Voraussetzung, wir können also annehmen, daß  $d \neq 0$  ist. Aber dann besitzt das Gleichungssystem genau eine Lösung  $(\alpha, \beta)$ .  $\square$

Mehr als 2 linear unabhängige Vektoren kann man in  $\mathcal{V}_2$  nicht finden. Es gilt nämlich:

**I.5.4 Satz.** *Je drei Vektoren  $\vec{x}_1, \vec{x}_2, \vec{x}_3 \in \mathcal{V}_2$  sind linear abhängig.*

BEWEIS: Sind  $\vec{x}_1$  und  $\vec{x}_2$  schon linear abhängig, so sind erst recht alle drei Vektoren linear abhängig. Sind aber  $\vec{x}_1$  und  $\vec{x}_2$  linear unabhängig, so kann man  $\vec{x}_3$  aus ihnen linear kombinieren:

$$\vec{x}_3 = \alpha \cdot \vec{x}_1 + \beta \cdot \vec{x}_2.$$

Doch dann ist  $\alpha \cdot \vec{x}_1 + \beta \cdot \vec{x}_2 + (-1) \cdot \vec{x}_3 = \vec{0}$  und  $(\alpha, \beta, -1) \neq (0, 0, 0)$ . Also sind die drei Vektoren linear abhängig.  $\square$

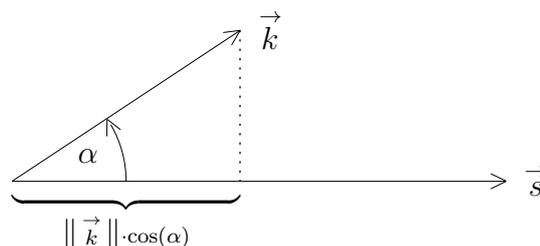
In  $\mathcal{V}_3$  gelten entsprechende Aussagen: Aus drei linear unabhängigen Vektoren läßt sich jeder weitere Vektor linear kombinieren, und mehr als drei linear unabhängige Vektoren kann es in  $\mathcal{V}_3$  nicht geben. Das werden wir später in allgemeinerem Zusammenhang beweisen.

Wir betrachten nun eine Situation aus der Physik:

Eine Kraft – dargestellt durch den Vektor  $\vec{k}$  – wirke auf einen Massenpunkt. Dieser Punkt sei nur in einer bestimmten Richtung bewegbar. Wenn er in dieser Richtung einen gewissen Weg zurücklegt, so kann man das auch mit einem Vektor  $\vec{s}$  beschreiben. Das Produkt

$$A := \text{Kraft} \cdot \text{Weg}$$

ist die dabei verrichtete *Arbeit*. Sie ist eine skalare Größe. Um sie richtig zu berechnen, darf man nur den Anteil der Kraftwirkung berücksichtigen, der in Richtung des Vektors  $\vec{s}$  wirkt:



In der von  $\vec{s}$  und  $\vec{k}$  aufgespannten Ebene schließen die beiden Vektoren einen Winkel  $\alpha$  ein. Dieser Winkel  $\alpha = \angle(\vec{s}, \vec{k})$  ist gerichtet, aber seine Richtung hat keinen Einfluß auf den Wert von  $A$ . Es gilt offensichtlich:

$$A = \|\vec{s}\| \cdot \|\vec{k}\| \cdot \cos(\alpha).$$

Da eine solche Konstruktion häufig vorkommt, erhält sie einen eigenen Namen:

**Definition.**

Sind  $\vec{v}$  und  $\vec{w}$  zwei Vektoren, so heißt

$$\vec{v} \bullet \vec{w} := \|\vec{v}\| \cdot \|\vec{w}\| \cdot \cos(\angle(\vec{v}, \vec{w}))$$

das *innere Produkt* oder *Skalarprodukt* von  $\vec{v}$  und  $\vec{w}$ .

**I.5.5 Satz.** Seien  $\vec{x}$  und  $\vec{y}$  Vektoren. Dann gilt:

1. Es ist  $\vec{x} \bullet \vec{x} = \|\vec{x}\|^2$ .
2. Ist  $\vec{x} = \vec{0}$  oder  $\vec{y} = \vec{0}$ , so ist  $\vec{x} \bullet \vec{y} = 0$ .
3. Es ist  $\vec{x} \bullet \vec{y} = \vec{y} \bullet \vec{x}$ .
4. Ist  $\vec{x} \neq \vec{0}$  und  $\vec{y} \neq \vec{0}$ , so gilt  $\vec{x} \bullet \vec{y} = 0$  genau dann, wenn  $\vec{x}$  und  $\vec{y}$  aufeinander senkrecht stehen.
5. Ist  $\lambda \in \mathbb{R}$ , so ist  $(\lambda \cdot \vec{x}) \bullet \vec{y} = \lambda \cdot (\vec{x} \bullet \vec{y}) = \vec{x} \bullet (\lambda \cdot \vec{y})$ .
6. Es gilt die „Schwarzsche Ungleichung“:

$$(\vec{x} \bullet \vec{y})^2 \leq \|\vec{x}\|^2 \cdot \|\vec{y}\|^2.$$

Dabei bedeutet die Gleichheit, daß  $\vec{x}$  und  $\vec{y}$  linear abhängig sind.

BEWEIS: Sei  $\alpha := \angle(\vec{x}, \vec{y})$ .

1) Ist  $\vec{x} = \vec{y}$ , so ist  $\alpha = 0$ , also  $\cos(\alpha) = 1$ .

2) Klar!

3) Vertauscht man  $\vec{x}$  und  $\vec{y}$ , so ändert sich lediglich das Vorzeichen von  $\alpha$ . Der Cosinus bleibt dann aber gleich!

4) Unter den angegebenen Voraussetzungen ist  $\|\vec{x}\| > 0$  und  $\|\vec{y}\| > 0$ . Ist nun  $\vec{x} \bullet \vec{y} = 0$ , so muß  $\cos(\alpha) = 0$  sein, und zwischen 0 und  $\pi$  passiert das nur bei  $\alpha = \frac{\pi}{2}$ .

5) Für positive  $\lambda$  folgt die Formel offensichtlich. Da  $\angle(-\vec{x}, \vec{y}) = 180^\circ - \angle(\vec{x}, \vec{y})$  ist, folgt sie auch für negative  $\lambda$ .

6) Die Ungleichung folgt, weil  $\cos^2(\alpha) \leq 1$  ist. Gleichheit gilt genau dann, wenn  $\cos(\alpha) = \pm 1$  ist, also  $\alpha = 0^\circ$  oder  $\alpha = 180^\circ$ . Das ist genau dann der Fall, wenn  $\vec{x}$  und  $\vec{y}$  linear abhängig sind.  $\square$

Das Skalarprodukt mißt, wie weit  $\vec{x}$  und  $\vec{y}$  davon entfernt sind, senkrecht aufeinander zu stehen. Und es unterscheidet sich sehr stark von den Produkten, die wir bisher kennengelernt haben: das Skalarprodukt von zwei *Vektoren* ist nicht mehr länger ein Vektor, es ist ein *Skalar*!

Etwas schwieriger ist folgende Aussage zu beweisen:

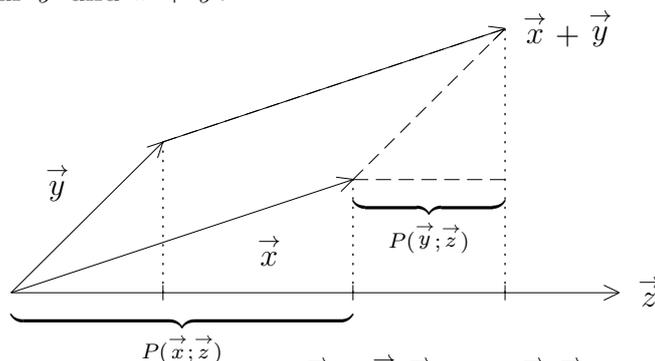
**I.5.6 Satz.** Sind  $\vec{x}$ ,  $\vec{y}$  und  $\vec{z}$  Vektoren, so gilt:

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

BEWEIS: Die (mit Vorzeichen versehene) orthogonale Projektion von  $\vec{x}$  auf  $\vec{z}$  ist die Zahl

$$P(\vec{x}; \vec{z}) := \|\vec{x}\| \cdot \cos(\angle(\vec{z}, \vec{x})) = \frac{\vec{x} \bullet \vec{z}}{\|\vec{z}\|}.$$

Entsprechendes gilt für  $\vec{y}$  und  $\vec{x} + \vec{y}$ .



Wie man der Skizze entnehmen kann, ist  $P(\vec{x} + \vec{y}; \vec{z}) = P(\vec{x}; \vec{z}) + P(\vec{y}; \vec{z})$ . Daraus folgt sofort die Behauptung.  $\square$

Als Anwendung kann das Skalarprodukt in Kartesischen Koordinaten besonders einfach berechnet werden:

**I.5.7 Satz.** Sind  $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$  und  $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$  zwei Vektoren in  $\mathcal{V}_3$ , so ist

$$\vec{x} \bullet \vec{y} = x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3.$$

BEWEIS:  $\vec{e}_1, \vec{e}_2, \vec{e}_3$  seien die Einheitsvektoren in Richtung der Koordinatenachsen. Dann gilt:

$$\begin{aligned} \vec{x} &= x_1 \cdot \vec{e}_1 + x_2 \cdot \vec{e}_2 + x_3 \cdot \vec{e}_3 \\ \text{und } \vec{y} &= y_1 \cdot \vec{e}_1 + y_2 \cdot \vec{e}_2 + y_3 \cdot \vec{e}_3, \end{aligned}$$

sowie

$$\vec{e}_i \bullet \vec{e}_i = 1 \quad \text{für alle } i \quad \text{und} \quad \vec{e}_i \bullet \vec{e}_j = 0 \quad \text{für } i \neq j.$$

Mit dem gerade bewiesenen Distributivgesetz folgt daraus die Behauptung.  $\square$

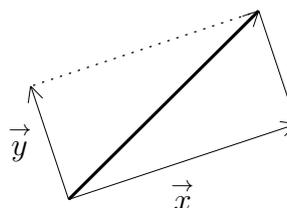
Weitere Folgerungen:

**I.5.8 Satz des Pythagoras.** Sind  $\vec{x}$  und  $\vec{y}$  Vektoren mit  $\vec{x} \bullet \vec{y} = 0$ , so ist

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

BEWEIS: Es ist

$$\begin{aligned} \|\vec{x} + \vec{y}\|^2 &= (\vec{x} + \vec{y}) \bullet (\vec{x} + \vec{y}) \\ &= \vec{x} \bullet \vec{x} + 2 \cdot \vec{x} \bullet \vec{y} + \vec{y} \bullet \vec{y} \\ &= \|\vec{x}\|^2 + \|\vec{y}\|^2. \end{aligned}$$



**I.5.9 Satz.** Sind  $\vec{x}, \vec{y}$  Vektoren  $\neq \vec{0}$ , mit  $\vec{x} \bullet \vec{y} = 0$ , so sind sie linear unabhängig.

BEWEIS: Es seien  $\lambda, \mu$  reelle Zahlen, mit  $\lambda \cdot \vec{x} + \mu \cdot \vec{y} = \vec{0}$ . Multiplikation mit  $\vec{x}$  ergibt:

$$0 = \vec{0} \bullet \vec{x} = \lambda \cdot \|\vec{x}\|^2 + \mu \cdot \vec{y} \bullet \vec{x} = \lambda \cdot \|\vec{x}\|^2.$$

Da  $\vec{x} \neq \vec{0}$  ist, muß  $\lambda = 0$  sein. Aber dann ist auch  $\mu = 0$ .  $\square$

Als nächstes betrachten wir Geraden in der Ebene:

Wir gehen diesmal von der allgemeinen Form in kartesischen Koordinaten aus:

$$L = \{(x_1, x_2) \in \mathbb{R}^2 \mid ax_1 + bx_2 = r\}.$$

Setzt man  $\vec{n} := (a, b)^\top$ , so ergibt sich folgende vektorielle Schreibweise:

$$L = \{\vec{x} \in \mathcal{V}_2 \mid \vec{n} \bullet \vec{x} = r\}.$$

Man kann die Situation noch normieren:

$$\vec{n}^0 := \frac{\vec{n}}{\|\vec{n}\|} \quad \text{und} \quad p := \frac{r}{\|\vec{n}\|}.$$

Dann ist  $\|\vec{n}^0\| = 1$  und

$$L = \{\vec{x} \in \mathcal{V}_2 \mid \vec{n}^0 \bullet \vec{x} - p = 0\}.$$

Man sagt dann,  $L$  ist in der *Hesseschen Normalform* gegeben. Den Vektor  $\vec{n}$  bezeichnet man als *Normalenvektor* und  $\vec{n}^0$  als *Einheitsnormalenvektor*.

Wir vergleichen dies mit den anderen Darstellungsformen von  $L$ :

- a) Ist  $L = \{\vec{x} \mid \vec{x} \bullet \vec{n} = p\}$  in der Hesseschen Normalform gegeben, mit einem normierten Vektor  $\vec{n} = (n_1, n_2)^\top$ , so kann man die Geradengleichung sofort ablesen:

$$n_1 x_1 + n_2 x_2 = p.$$

Wie man umgekehrt von einer Geradengleichung auf die Hessesche Normalform kommt, haben wir gerade oben gesehen.

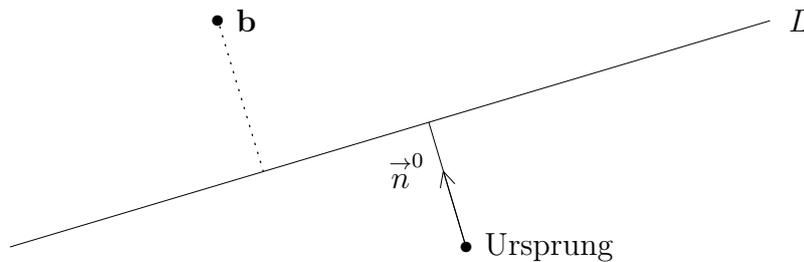
- b) Ist  $L = \{\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2 \mid ax_1 + bx_2 = r\}$ , so steht  $\mathbf{v} := (-b, a)$  auf dem Normalenvektor  $\mathbf{n} = (a, b)$  senkrecht, kann also als Richtungsvektor benutzt werden. Um außerdem einen Ortsvektor auf  $L$  zu bekommen, muß man die Gleichung  $ax_1 + bx_2 = r$  lösen. Dabei ist ein Parameter frei, und wenn etwa  $b \neq 0$  ist, dann ist  $\mathbf{x}_0 := (0, \frac{r}{b})$  ein geeigneter Ortsvektor. Das ergibt die Darstellung

$$L = \{\mathbf{x} \mid \exists t \in \mathbb{R} \text{ mit } \mathbf{x} = \mathbf{x}_0 + t\mathbf{v}\}.$$

Ist umgekehrt  $L$  in einer solchen Punkt-Richtungs-Form gegeben und  $\mathbf{n}$  ein zu  $\mathbf{v}$  orthogonaler Einheitsvektor, so hat die Hessesche Form die Gestalt

$$L = \{\mathbf{x} \mid (\mathbf{x} - \mathbf{x}_0) \bullet \mathbf{n} = 0\}.$$

Die Hessesche Normalform ist besonders gut dazu geeignet, den Abstand eines Punktes  $\mathbf{b} = (b_1, b_2)$  von der Geraden  $L$  zu bestimmen:



Es geht darum, eine reelle Zahl  $d$  zu bestimmen, so daß  $\vec{b} + d \cdot \vec{n}^0 \in L$  ist. Dann ist  $|d|$  der gesuchte Abstand.

Ist  $d$  gefunden und  $\vec{x}_1 := \vec{b} + d \cdot \vec{n}^0$ , so liegt auch  $\vec{x}_2 := \vec{x}_1 + \vec{v}$  auf  $L$ , und es gilt:

$$0 = \vec{v} \bullet \vec{n}^0 = (\vec{x}_2 - \vec{x}_1) \bullet \vec{n}^0 = p - (\vec{b} + d \cdot \vec{n}^0) \bullet \vec{n}^0 = p - \vec{b} \bullet \vec{n}^0 - d.$$

Damit folgt:

Der gesuchte Abstand ist  $|d| = |p - \vec{b} \bullet \vec{n}^0|$ .

**Beispiel:**

Sei  $L = \{(x, y) \mid 3x + 2y = 5\}$ ,  $\vec{b} = (4, 3)^\top$ . Dann ist  $\vec{n} = (3, 2)^\top$  ein Normalenvektor mit Norm  $\|\vec{n}\| = \sqrt{13}$ , also  $\vec{n}^0 = \frac{1}{\sqrt{13}} \cdot \begin{pmatrix} 3 \\ 2 \end{pmatrix}$  und  $p = \frac{5}{\sqrt{13}}$ .

Also ist der gesuchte Abstand

$$= |p - \vec{b} \bullet \vec{n}^0| = \left| \frac{5}{\sqrt{13}} - \frac{1}{\sqrt{13}} \cdot (4 \cdot 3 + 3 \cdot 2) \right| = \frac{13}{\sqrt{13}} = \sqrt{13}.$$

Für Ebenen im Raum kann man ebenfalls eine Hessesche Normalform aufstellen. In Punkt-Richtungs-Form hat eine solche Ebene die Gestalt

$$E = \{\mathbf{x} \in \mathbb{R}^3 \mid \exists s, t \in \mathbb{R} \text{ mit } \mathbf{x} = \mathbf{x}_0 + s\mathbf{a} + t\mathbf{b}\},$$

mit zwei linear unabhängigen Vektoren  $\mathbf{a}$  und  $\mathbf{b}$ . Ist  $\mathbf{n}$  ein Einheitsvektor, der gleichzeitig auf  $\mathbf{a}$  und  $\mathbf{b}$  senkrecht steht, so ergibt sich die Hessesche Normalform

$$E = \{\mathbf{x} \in \mathbb{R}^3 \mid (\mathbf{x} - \mathbf{x}_0) \bullet \mathbf{n} = 0\}.$$

Hieraus kann man sofort die Ebenengleichung

$$(x_1 - x_1^0)n_1 + (x_2 - x_2^0)n_2 + (x_3 - x_3^0)n_3 = 0$$

ablesen.

### Definition.

Ein rechteckiges Zahlenschema

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

wird als (reellwertige) *Matrix* mit  $n$  Zeilen und  $m$  Spalten bezeichnet. Die Einträge  $a_{11}, a_{12}, \dots$  nennt man die *Koeffizienten* der Matrix.

Die Menge aller solchen Matrizen wird mit  $M_{n,m}(\mathbb{R})$  bezeichnet.

### Beispiele :

1. Die Koeffizienten eines Linearen Gleichungssystems

$$\begin{aligned} ax + by &= r \\ cx + dy &= s \end{aligned}$$

kann man zu der Matrix  $A := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2,2}(\mathbb{R})$  zusammenfassen.

2. Die Komponenten  $x_1, x_2, x_3$  eines Punktes im  $\mathbb{R}^3$  kann man zu einer Zeile  $\mathbf{x} = (x_1, x_2, x_3) \in M_{1,3}(\mathbb{R})$  zusammenfassen.

3. Ein Vektor  $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$  bildet eine Spalte aus  $M_{3,1}(\mathbb{R})$ .

4. Sind  $\vec{x}, \vec{y}$  und  $\vec{z}$  (Spalten-)Vektoren in  $\mathcal{V}_3$ , so ist  $A := (\vec{x}, \vec{y}, \vec{z})$  eine Matrix aus  $M_{3,3}(\mathbb{R})$ .

In der  $i$ -ten Zeile und  $j$ -ten Spalte der Matrix  $A$  steht das Element  $a_{ij}$ . Vertauscht man die Komponenten so, daß in der  $i$ -ten Zeile und  $j$ -ten Spalte das Element  $a_{ji}$  steht, so erhält man die sogenannte *transponierte Matrix*  $A^\top$ .

Z.B. ist  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^\top = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ ,  $\mathbf{x}^\top = \vec{x}$  und  $\vec{x}^\top = \mathbf{x}$ .

Ganz allgemein ist  $(A^\top)^\top = A$ .

### Definition.

Ist  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_{2,2}(\mathbb{R})$ , so heißt die Zahl  $\det(A) := ad - bc$  die *Determinante* der Matrix  $A$ .

Da eine Matrix  $A \in M_{2,2}(\mathbb{R})$  aus zwei Spaltenvektoren besteht, kann man die Determinante auch als Abbildung  $\det : \mathcal{V}_2 \times \mathcal{V}_2 \rightarrow \mathbb{R}$  auffassen. Dann hat sie die folgenden Eigenschaften:

**I.5.10 Satz.** Für  $\vec{x}, \vec{y} \in \mathcal{V}_2$  und  $\alpha, \beta \in \mathbb{R}$  gilt:

1.  $\det(\vec{y}, \vec{x}) = -\det(\vec{x}, \vec{y})$ .
2.  $\det(\vec{x}_1 + \vec{x}_2, \vec{y}) = \det(\vec{x}_1, \vec{y}) + \det(\vec{x}_2, \vec{y})$ .
3.  $\det(\alpha \cdot \vec{x}, \beta \cdot \vec{y}) = \alpha\beta \cdot \det(\vec{x}, \vec{y})$ .
4.  $\det(\vec{e}_1, \vec{e}_2) = 1$ .

**BEWEIS:** Sei  $\vec{x} = \begin{pmatrix} a \\ c \end{pmatrix}$ ,  $\vec{y} = \begin{pmatrix} b \\ d \end{pmatrix}$  und  $\vec{x}_i = \begin{pmatrix} a_i \\ c_i \end{pmatrix}$ . Dann gilt:

$$1) \det(\vec{y}, \vec{x}) = \det \begin{pmatrix} b & a \\ d & c \end{pmatrix} = bc - ad = -\det(\vec{x}, \vec{y}).$$

$$2) \det(\vec{x}_1 + \vec{x}_2, \vec{y}) = \det \begin{pmatrix} a_1 + a_2 & b \\ c_1 + c_2 & d \end{pmatrix} = (a_1 + a_2)d - b(c_1 + c_2) = (a_1d - bc_1) + (a_2d - bc_2) = \det(\vec{x}_1, \vec{y}) + \det(\vec{x}_2, \vec{y}).$$

$$3) \det(\alpha \cdot \vec{x}, \beta \cdot \vec{y}) = \det \begin{pmatrix} \alpha a & \beta b \\ \alpha c & \beta d \end{pmatrix} = \alpha\beta \cdot (ad - bc) = \alpha\beta \cdot \det(\vec{x}, \vec{y}).$$

$$4) \det(\vec{e}_1, \vec{e}_2) = \det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1. \quad \square$$

Sind  $\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$  und  $\vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$  zwei Spaltenvektoren aus  $\mathcal{V}_3$ , so kann man aus ihnen die Matrix

$$A = (\vec{a}, \vec{b}) = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix} \in M_{3,2}(\mathbb{R})$$

bilden. Streicht man in  $A$  die  $i$ -te Zeile, so erhält man eine quadratische Matrix  $S_i(A) \in M_{2,2}(\mathbb{R})$ .

**Definition.**

Sind  $\vec{a}, \vec{b} \in \mathcal{V}_3$ , so definiert man das *Vektorprodukt*  $\vec{a} \times \vec{b} \in \mathcal{V}_3$  durch

$$\vec{a} \times \vec{b} := \sum_{i=1}^3 (-1)^{i+1} \det(S_i(A)) \cdot \vec{e}_i.$$

Dann gilt:

$$\begin{aligned}\vec{a} \times \vec{b} &= \det \begin{pmatrix} a_2 & b_2 \\ a_3 & b_3 \end{pmatrix} \cdot \vec{e}_1 - \det \begin{pmatrix} a_1 & b_1 \\ a_3 & b_3 \end{pmatrix} \cdot \vec{e}_2 + \det \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \cdot \vec{e}_3 \\ &= \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}.\end{aligned}$$

Hier haben wir nun glücklich das dritte Produkt im Bereich der Vektorrechnung kennengelernt. Wir kennen inzwischen:

- Die skalare Multiplikation  $\alpha \cdot \vec{x}$  (für  $\alpha \in \mathbb{R}$  und Vektoren  $\vec{x}$ ),
- das Skalarprodukt  $\vec{x} \bullet \vec{y}$  (für Vektoren  $\vec{x}$  und  $\vec{y}$ ),
- das Vektorprodukt  $\vec{x} \times \vec{y}$  (ebenfalls für Vektoren  $\vec{x}$  und  $\vec{y}$ ).

Das Vektorprodukt sieht so aus, wie man es von einem anständigen Produkt erwartet: Zwei Vektoren werden wieder zu einem Vektor verknüpft. Leider ist diese Konstruktion nur in  $\mathcal{V}_3$  möglich! Wir wollen untersuchen, was man sich anschaulich unter dem Vektorprodukt vorzustellen hat:

**I.5.11 Satz.** Für  $\vec{a}, \vec{a}_1, \vec{a}_2, \vec{b} \in \mathcal{V}_3$  und  $\alpha \in \mathbb{R}$  gilt:

1.  $\vec{b} \times \vec{a} = -\vec{a} \times \vec{b}$ , insbesondere  $\vec{a} \times \vec{a} = \vec{0}$ .
2.  $(\vec{a}_1 + \vec{a}_2) \times \vec{b} = \vec{a}_1 \times \vec{b} + \vec{a}_2 \times \vec{b}$ .
3.  $(\alpha \cdot \vec{a}) \times \vec{b} = \alpha \cdot (\vec{a} \times \vec{b})$ .
4.  $\vec{a} \bullet (\vec{a} \times \vec{b}) = \vec{b} \bullet (\vec{a} \times \vec{b}) = 0$ .

**BEWEIS:** Die Eigenschaften (1), (2) und (3) ergeben sich sehr einfach aus der Definition mit Hilfe von Determinanten.

(4): Man muß mit Hilfe der Koordinaten ausrechnen, daß  $\vec{a} \bullet (\vec{a} \times \vec{b}) = 0$  ist. Dann ist auch  $\vec{b} \bullet (\vec{a} \times \vec{b}) = -\vec{b} \bullet (\vec{b} \times \vec{a}) = 0$ . □

Sind die Vektoren  $\vec{a}$  und  $\vec{b}$  linear abhängig, so ist der eine von ihnen ein Vielfaches des anderen, und aus (1) folgt, daß  $\vec{a} \times \vec{b} = \vec{0}$  ist. Sind die Vektoren dagegen linear unabhängig, so spannen sie eine Ebene  $E \subset \mathcal{V}_3$  auf, die den Nullpunkt enthält, und das Vektorprodukt  $\vec{a} \times \vec{b}$  steht auf dieser Ebene senkrecht. Es kann in diesem Falle nicht der Nullvektor sein.<sup>4</sup> Aber in welche der beiden Normalenrichtungen weist  $\vec{a} \times \vec{b}$ ? Und wie lang ist der Vektor?

Betrachten wir zunächst einen Spezialfall: Sind  $\vec{e}_1, \vec{e}_2$  und  $\vec{e}_3$  die Standard-Einheitsvektoren, so gilt:

$$\vec{e}_1 \times \vec{e}_2 = \vec{e}_3, \quad \vec{e}_1 \times \vec{e}_3 = -\vec{e}_2 \quad \text{und} \quad \vec{e}_2 \times \vec{e}_3 = \vec{e}_1.$$

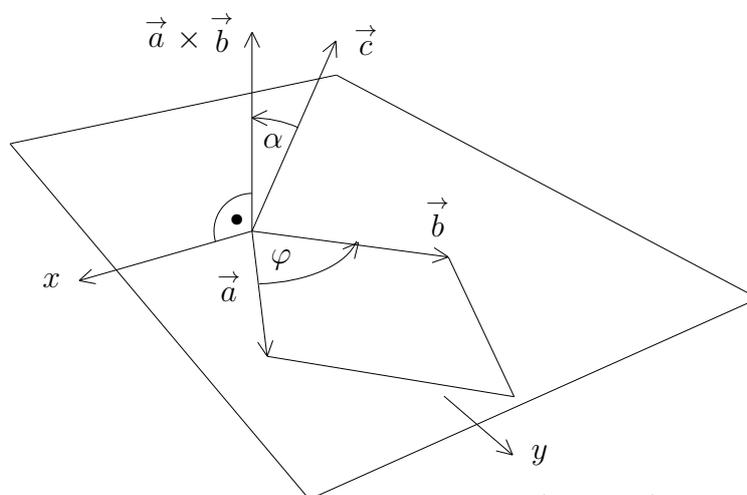
<sup>4</sup>Der Beweis wird sich aus der Determinantentheorie in Kapitel II ergeben.

Insbesondere bilden die Vektoren  $\vec{e}_1$ ,  $\vec{e}_2$  und  $\vec{e}_1 \times \vec{e}_2$  ein sogenanntes „Rechtssystem“. Spreizt man Daumen, Zeigefinger und Mittelfinger der rechten Hand, so bilden die Richtungen, in die die Finger weisen, in der genannten Reihenfolge ein Rechtssystem. Das ist natürlich keine mathematische Definition, aber wir können das Vektorprodukt für eine exakte Definition benutzen:

**Definition.**

Ein System  $\{\vec{a}, \vec{b}, \vec{c}\}$  von linear unabhängigen Vektoren in  $\mathcal{V}_3$  bezeichnet man als *Basis* oder *Dreibein*. Legt man auch noch die Reihenfolge der Vektoren fest, so spricht man von einer *orientierten Basis*. Die Basis heißt *positiv orientiert* oder ein *Rechtssystem*, falls  $\alpha := \angle(\vec{c}, \vec{a} \times \vec{b}) < 90^\circ$  ist. Andernfalls heißt sie negativ orientiert.

Die von  $\vec{a}$  und  $\vec{b}$  aufgespannte Ebene



Sind also  $\vec{a}$  und  $\vec{b}$  zwei linear unabhängige Vektoren, so bilden  $\vec{a}, \vec{b}, \vec{a} \times \vec{b}$  definitionsgemäß eine positiv orientierte Basis. Die oben genannte Rechte-Hand-Regel zeigt, wie die positive Orientierung des Raumes *anschaulich* zu verstehen ist.

**I.5.12 Satz.** *Es ist*

$$\|\vec{a} \times \vec{b}\|^2 = \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 - (\vec{a} \bullet \vec{b})^2.$$

Zum BEWEIS muß man nur beide Seiten der Gleichung ausrechnen.

**I.5.13 Folgerung.** *Es ist*

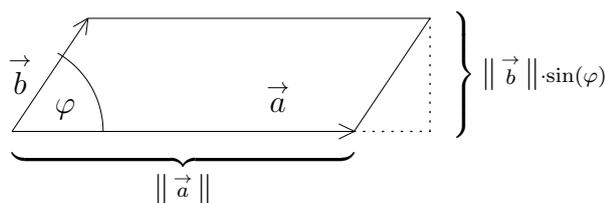
$$\|\vec{a} \times \vec{b}\| = \|\vec{a}\| \cdot \|\vec{b}\| \cdot \sin(\angle(\vec{a}, \vec{b})).$$

BEWEIS: Sei  $\varphi := \angle(\vec{a}, \vec{b})$ . Dann gilt:

$$\begin{aligned} \|\vec{a} \times \vec{b}\|^2 &= \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 - (\vec{a} \bullet \vec{b})^2 \\ &= \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 \cdot (1 - \cos^2(\varphi)) \\ &= \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 \cdot \sin^2(\varphi). \end{aligned}$$

Da  $0^\circ \leq \varphi \leq 180^\circ$  ist, ist  $\sin(\varphi) \geq 0$ . □

Das gewonnene Ergebnis hat anschauliche Bedeutung:



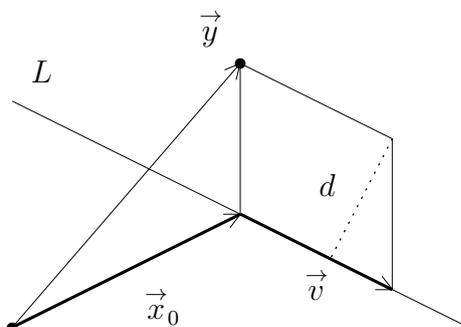
Die Länge von  $\vec{a} \times \vec{b}$  ist genau der Flächeninhalt des von  $\vec{a}$  und  $\vec{b}$  aufgespannten Parallelogramms!

Wir wollen zwei Anwendungen des Vektorproduktes betrachten:

- a) Es sei eine Gerade in Punkt-Richtungs-Form gegeben,

$$L = \{\vec{x} \mid \exists t \in \mathbb{R} \text{ mit } \vec{x} = \vec{x}_0 + t \vec{v}\},$$

sowie ein weiterer Punkt  $\vec{y}$ . Dann spannen die Vektoren  $\vec{v}$  und  $\vec{y} - \vec{x}_0$  ein Parallelogramm auf, dessen Höhe  $d$  den Abstand von  $\vec{y}$  und  $L$  beschreibt:



Die Fläche des Parallelogramms beträgt dann gerade  $\|\vec{v}\| \cdot d$ , und wegen der Formel für den Betrag des Vektorproduktes folgt daraus:

$$d = \frac{\|\vec{v} \times (\vec{y} - \vec{x}_0)\|}{\|\vec{v}\|}.$$

- b) Es soll die Schnittgerade zweier Ebenen bestimmt werden, die beide in der Hesseschen Normalform vorliegen:

$$\begin{aligned} E_1 &= \{\vec{x} \mid (\vec{x} - \vec{x}_1) \bullet \vec{n}_1 = 0\} \\ \text{und } E_2 &= \{\vec{x} \mid (\vec{x} - \vec{x}_2) \bullet \vec{n}_2 = 0\}. \end{aligned}$$

Eine Lösung  $\vec{x}_0$  des Gleichungssystems

$$(\vec{x} - \vec{x}_i) \bullet \vec{n}_i, \quad i = 1, 2,$$

liegt auf jeden Fall auf der Schnittgeraden.

Der Vektor  $\vec{v} := \vec{n}_1 \times \vec{n}_2$  steht gleichzeitig auf  $\vec{n}_1$  und auf  $\vec{n}_2$  senkrecht, liegt also gleichzeitig in  $E_1$  und in  $E_2$ , und damit auf der Schnittgeraden

$$L := \{\vec{x} \mid \exists t \in \mathbb{R} \text{ mit } \vec{x} = \vec{x}_0 + t \vec{v}\}.$$

**Definition.**

Sind  $\vec{a}, \vec{b}, \vec{c} \in \mathcal{V}_3$ , so nennt man

$$[\vec{a}, \vec{b}, \vec{c}] := \vec{a} \bullet (\vec{b} \times \vec{c})$$

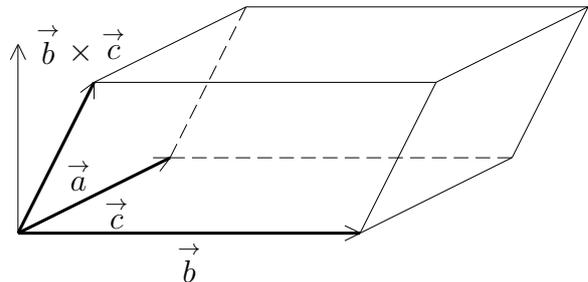
das *Spatprodukt* von  $\vec{a}, \vec{b}$  und  $\vec{c}$ .

Noch ein Produkt!! Diesmal sogar mit 3 Faktoren! Nun, lassen Sie sich nicht irritieren, so schlimm ist es gar nicht. Was haben wir hier vor uns? Das Spatprodukt ist eine reelle Zahl, die wir noch etwas genauer hinschreiben können:

$$[\vec{a}, \vec{b}, \vec{c}] = \|\vec{a}\| \cdot \|\vec{b} \times \vec{c}\| \cdot \cos(\angle(\vec{a}, \vec{b} \times \vec{c})).$$

Das Vorzeichen hängt also nur von dem Winkel  $\beta := \angle(\vec{a}, \vec{b} \times \vec{c})$  ab. Ist  $\beta \leq 90^\circ$ , so ist das Spatprodukt positiv oder Null, andernfalls ist es negativ. Und  $\beta$  ist genau dann kleiner als  $90^\circ$ , wenn  $\{\vec{b}, \vec{c}, \vec{a}\}$  (in der angegebenen Reihenfolge) eine positiv orientierte Basis ist. Da eine zweimalige Vertauschung der Basisvektoren die Orientierung erhält, kann man sagen:

$$[\vec{a}, \vec{b}, \vec{c}] > 0 \iff \{\vec{a}, \vec{b}, \vec{c}\} \text{ positiv orientierte Basis.}$$



Und was ist  $|[\vec{a}, \vec{b}, \vec{c}]|$ ?

Die drei Vektoren spannen ein räumliches Gebilde auf, das man einen *Spat* nennt. Man kann dieses Gebilde als 3-dimensionale Verallgemeinerung des Parallelogramms auffassen. Die Grundfläche  $F$  ist durch die Zahl  $\|\vec{b} \times \vec{c}\|$  gegeben, die Höhe  $h$  durch die Gleichung  $\frac{h}{\|\vec{a}\|} = \cos(\beta)$ . Das Volumen  $V$  des Spates errechnet sich wie bei einem Quader als

$$\begin{aligned} & \text{Grundfläche} \cdot \text{Höhe} \\ &= F \cdot h = \|\vec{b} \times \vec{c}\| \cdot \|\vec{a}\| \cdot \cos(\beta) = |[\vec{a}, \vec{b}, \vec{c}]|. \end{aligned}$$

Damit ergibt sich:

**I.5.14 Satz.**

1. Folgende Aussagen über Vektoren  $\vec{a}, \vec{b}$  und  $\vec{c}$  sind äquivalent:

- (a)  $\vec{a}, \vec{b}$  und  $\vec{c}$  sind linear abhängig.
- (b)  $\vec{a}, \vec{b}$  und  $\vec{c}$  sind komplanar.
- (c)  $[\vec{a}, \vec{b}, \vec{c}] = 0$ .

2. Vertauscht man in  $[\vec{a}, \vec{b}, \vec{c}]$  zwei aufeinander folgende Argumente, so ändert sich das Vorzeichen.

3. Es ist

$$[\vec{a}_1 + \vec{a}_2, \vec{b}, \vec{c}] = [\vec{a}_1, \vec{b}, \vec{c}] + [\vec{a}_2, \vec{b}, \vec{c}]$$

und

$$[\alpha \cdot \vec{a}, \vec{b}, \vec{c}] = \alpha \cdot [\vec{a}, \vec{b}, \vec{c}].$$

4. Es ist  $[\vec{e}_1, \vec{e}_2, \vec{e}_3] = 1$ .

### Definition.

Sei  $A = (\vec{a}_1, \vec{a}_2, \vec{a}_3) \in M_{3,3}(\mathbb{R})$  eine 3-reihige quadratische Matrix. Dann heißt

$$\det(A) := [\vec{a}_1, \vec{a}_2, \vec{a}_3]$$

die Determinante von  $A$ .

Man beachte, daß die 3-reihige Determinante ganz analoge Eigenschaften zur 2-reihigen Determinante hat.

Natürlich kann man die Determinante auch explizit berechnen. Es ist

$$\begin{aligned} \det(\vec{a}, \vec{b}, \vec{c}) &= \vec{a} \bullet (\vec{b} \times \vec{c}) = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \bullet \begin{pmatrix} b_2c_3 - b_3c_2 \\ b_3c_1 - b_1c_3 \\ b_1c_2 - b_2c_1 \end{pmatrix} \\ &= a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1 \\ &= \sum_{\sigma \in S_3} \operatorname{sgn}(\sigma) a_{\sigma(1)} b_{\sigma(2)} c_{\sigma(3)}. \end{aligned}$$

Setzt man die Originaldefinition des Vektorproduktes ein, so erhält man ferner:

### I.5.15 Laplace'scher Entwicklungssatz.

$$\det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = a_1 \cdot \det \begin{pmatrix} b_2 & c_2 \\ b_3 & c_3 \end{pmatrix} - a_2 \cdot \det \begin{pmatrix} b_1 & c_1 \\ b_3 & c_3 \end{pmatrix} + a_3 \cdot \det \begin{pmatrix} b_1 & c_1 \\ b_2 & c_2 \end{pmatrix}.$$

BEWEIS: Es ist

$$\vec{a} \bullet (\vec{b} \times \vec{c}) = \sum_{i=1}^3 (-1)^{i+1} a_i \cdot \det(S_i(\vec{b}, \vec{c})).$$

□

Wir werden später für jedes  $n$  eine Determinantenfunktion

$$\det : M_{n,n} \rightarrow \mathbb{R}$$

definieren, die die gleichen Eigenschaften aufweist, wie wir sie im Falle  $n = 2$  und  $n = 3$  schon nachgewiesen haben, und wir werden zeigen, daß sie durch diese Eigenschaften sogar eindeutig bestimmt ist. In der Theorie der Linearen Gleichungssysteme mit  $n$  Unbestimmten spielen die Determinanten eine wichtige Rolle. Mit Hilfe eines verallgemeinerten Laplaceschen Entwicklungssatzes kann man solche Determinanten induktiv berechnen. Zum Schluß sollen noch zwei „Eselsbrücken“ angegeben werden:

1. Das Vektorprodukt kann formal als Determinante berechnet werden:

$$\vec{a} \times \vec{b} = \det \begin{pmatrix} \vec{e}_1 & a_1 & b_1 \\ \vec{e}_2 & a_2 & b_2 \\ \vec{e}_3 & a_3 & b_3 \end{pmatrix}.$$

Natürlich darf man so etwas nicht wirklich hinschreiben! Die Einträge in der Matrix dürfen eigentlich nur Skalare sein, keine Vektoren.

2. Für die Berechnung der 3-reihigen Determinante (und nur für diese!) gibt es die *Sarrus'sche Regel*:

Schreibe die Matrix-Elemente in der folgenden Form auf:

$$\begin{array}{ccc|cc} a_1 & b_1 & c_1 & a_1 & b_1 \\ a_2 & b_2 & c_2 & a_2 & b_2 \\ a_3 & b_3 & c_3 & a_3 & b_3 \end{array}$$

Die Produkte der Elemente in den „Hauptdiagonalen“ werden mit „+“ versehen, die Produkte der Elemente in den „Nebendiagonalen“ werden mit „-“ versehen.

Schließlich werden alle Produkte aufsummiert:

$$\begin{array}{ccccccc} & & & & - & - & - \\ & & & \nearrow & \nearrow & \nearrow & \\ a_1 & b_1 & c_1 & a_1 & b_1 & & \\ & \searrow & \times & \times & \nearrow & & \\ a_2 & b_1 & c_1 & a_1 & b_1 & & \\ & \nearrow & \times & \times & \searrow & & \\ a_3 & b_1 & c_1 & a_1 & b_1 & & \\ & & & \searrow & \searrow & \searrow & \\ & & & + & + & + & \end{array}$$

## §6 Komplexe Zahlen

Ist  $x$  eine beliebige reelle Zahl, so ist  $x^2 \geq 0$ . Deshalb ist in  $\mathbb{R}$  nicht jede quadratische Gleichung lösbar. Es kann z.B. keine reelle Zahl  $x$  mit  $x^2 + 1 = 0$  geben.

Wir haben jedoch bewiesen, daß eine Gleichung dritten Grades immer eine Lösung besitzt. Wir wollen sehen, wie man eine solche Lösung finden kann:

Es geht also um eine allgemeine kubische Gleichung der Form

$$x^3 + ax^2 + bx + c = 0.$$

Man kann den Anfang der linken Seite,  $x^3 + ax^2$ , als Teil eines Binoms vom Grad 3 auffassen:

$$\left(x + \frac{a}{3}\right)^3 = x^3 + ax^2 + 3 \cdot \left(\frac{a}{3}\right)^2 x + \left(\frac{a}{3}\right)^3.$$

Also ist

$$x^3 + ax^2 + bx + c = \left(x + \frac{a}{3}\right)^3 + \left(b - \frac{a^2}{3}\right)x + \left(c - \left(\frac{a}{3}\right)^3\right).$$

Das bedeutet, daß man nur die folgende *reduzierte Gleichung* lösen muß:

$$\boxed{y^3 + py = q},$$

mit  $p = b - \frac{a^2}{3}$ ,  $q = \frac{ab}{3} - c - \frac{2a^3}{27}$  und  $y = x + \frac{a}{3}$ .

Um weiter zu kommen, braucht man einen Trick, der um 1500 in Italien gefunden wurde:

Es sei zunächst  $p > 0$  und  $q > 0$ . Der Trick besteht in dem Ansatz  $y = u + v$ . Setzt man das ein, so erhält man:

$$u^3 + v^3 + 3uv(u + v) + p(u + v) = q.$$

Wenn man nun  $u$  und  $v$  so wählen kann, daß  $u^3 + v^3 = q$  und  $uv = -\frac{p}{3}$  ist, dann hat man die Gleichung gelöst. Und das ist tatsächlich möglich:

Ist eine quadratische Gleichung  $z^2 + \beta z + \gamma = 0$  mit der Diskriminante  $\Delta := \beta^2 - 4\gamma$  gegeben, und  $\Delta > 0$ , so erhält man die Lösungen

$$z_1 = \frac{-\beta + \sqrt{\Delta}}{2} \quad \text{und} \quad z_2 = \frac{-\beta - \sqrt{\Delta}}{2}.$$

Wir haben in §4 schon die „Gleichungen von Vieta“ kennengelernt:

$$\boxed{\begin{array}{l} z_1 + z_2 = -\beta \\ z_1 \cdot z_2 = \gamma \end{array}}$$

Die Gleichungen  $u^3 + v^3 = q$  und  $u^3 \cdot v^3 = -\frac{p^3}{27}$  sind also sicher dann erfüllt, wenn  $u^3$  und  $v^3$  Lösungen der quadratischen Gleichung  $z^2 - qz - \frac{p^3}{27} = 0$  sind, wenn also gilt:

$$\begin{aligned} u^3 &= \frac{q + \sqrt{\Delta}}{2} \\ \text{und} \quad v^3 &= \frac{q - \sqrt{\Delta}}{2}, \end{aligned}$$

mit  $\Delta = q^2 + \frac{4p^3}{27}$ . Da  $p, q > 0$  vorausgesetzt wurde, ist tatsächlich  $\Delta > 0$ . Also gilt:

$$y = u + v = \sqrt[3]{\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}.$$

Gefunden hat diese Gleichung der Italiener Scipione dal Ferro. Allerdings hielt er das Ergebnis zunächst geheim und verriet es nur seinem Schwiegersohn und einem befreundeten Rechenmeister namens Antonio Maria Fiore. Letzterer forderte voller Stolz den bekannten Rechenmeister Tartaglia, den „Stotterer“, zu einem öffentlichen Wettstreit heraus. Weil Fiore wohl nicht besonders intelligent war und Tartaglia zudem auf dunklen Wegen auch in den Besitz des Geheimnisses gekommen war, gewann der Stotterer den Wettstreit auf glänzende Weise. Er wurde daraufhin von dem Mathematiker Girolamo Cardano bedrängt, ihm doch das Geheimnis zu verraten. Nach langem Zögern gab er nach, ließ sich jedoch unter Eid Verschwiegenheit zusichern. Als Cardano die Formeln studierte, kam er zu der Ansicht, sie seien von Tartaglia nicht auf redliche Weise erworben worden. Er fühlte sich nicht mehr an seinen Eid gebunden und veröffentlichte das Ergebnis innerhalb eines umfangreichen Algebra-Buches. Ein langer Rechtsstreit und wüste öffentliche Beschimpfungen waren die Folge, aber in die Geschichte gingen die Lösungen als „Cardanische Formeln“ ein.

### Beispiel:

Wir betrachten die reduzierte Gleichung  $y^3 + 3y = 4$ .

Dann ist

$$u = \sqrt[3]{2 + \sqrt{5}} \text{ und } v = \sqrt[3]{2 - \sqrt{5}}.$$

Leider erhält man die Lösung so in einer sehr komplizierten Form. Wenn man weiß, daß

$$(1 \pm \sqrt{5})^3 = 1 \pm 3\sqrt{5} + 15 \pm 5\sqrt{5} = 16 \pm 8\sqrt{5} = 8 \cdot (2 \pm \sqrt{5})$$

ist, dann erhält man:

$$y = u + v = \frac{1}{2} \cdot (1 + \sqrt{5}) + \frac{1}{2} \cdot (1 - \sqrt{5}) = 1.$$

Dieses Ergebnis hätte man natürlich durch Probieren schneller gefunden, aber wenn es keine so glatte Lösung gibt, dann ist die „Cardano’sche Formel“ sehr nützlich.

Es gibt Fälle, wo die Formel nicht angewandt werden kann:

Wir betrachten die Gleichung  $y^3 - 6y + 4 = 0$ , die auch vom obigen Typ ist. Hier ist  $p = -6$  und  $q = -4$ , also  $\Delta = 16 - 4 \cdot 2^3 = -16 < 0$ .

Wenn man alle Bedenken über Bord wirft und mit  $\sqrt{\Delta} = 4\sqrt{-1}$  formal weiterrechnet, so erhält man:

$$y = \sqrt[3]{-2 + 2\sqrt{-1}} + \sqrt[3]{-2 - 2\sqrt{-1}}.$$

Das ist ein ziemlich sinnloser Ausdruck, aber man kann sich überlegen, daß

$$(1 \pm \sqrt{-1})^3 = 1 \pm 3\sqrt{-1} - 3 \pm (-\sqrt{-1}) = -2 \pm 2\sqrt{-1}$$

ist, also

$$y = (1 + \sqrt{-1}) + (1 - \sqrt{-1}) = 2.$$

Nun ist  $y^3 - 6y + 4 = (y - 2) \cdot (y^2 + 2y - 2) = (y - 2)(y + 1 - \sqrt{3})(y + 1 + \sqrt{3})$ . Durch die Rechnung mit der „imaginären Größe“  $\sqrt{-1}$  hat man schließlich drei reelle Lösungen gefunden.<sup>5</sup>

Derartige Entdeckungen waren Anlaß genug, auch mit Wurzeln aus negativen Zahlen zu rechnen. Der berühmte Mathematiker Euler führte 1777 die Abkürzung

$$i := \sqrt{-1}$$

ein, und Gauß übernahm sie und erklärte außerdem, was man sich anschaulich unter einer „komplexen Zahl“  $a + ib$  (mit  $a, b \in \mathbb{R}$ ) vorzustellen hat. Die Bezeichnungen haben sich bis heute erhalten, allerdings wird in der Elektrotechnik lieber  $\mathbf{j}$  statt  $i$  geschrieben.

Wenn man mit  $\mathbf{j}$  genauso wie mit anderen Zahlen rechnen kann, dann ergibt sich:

$$\begin{aligned} (a + b\mathbf{j}) + (c + d\mathbf{j}) &= (a + b) + (c + d)\mathbf{j}, \\ \text{und } (a + b\mathbf{j}) \cdot (c + d\mathbf{j}) &= (ac - bd) + (ad + bc)\mathbf{j}. \end{aligned}$$

Das motiviert die folgende

**Definition.**

Unter der Menge der *komplexen Zahlen* versteht man die Ebene  $\mathbb{R}^2$ , auf der neben der vektoriellen Addition zusätzlich eine Multiplikation gegeben ist:

$$(a, b) \cdot (c, d) := (ac - bd, ad + bc).$$

Das Element  $(1, 0)$  wird mit  $\mathbf{1}$  bezeichnet, das Element  $(0, 1)$  mit  $\mathbf{j}$ .

Die Menge der komplexen Zahlen wird auch mit dem Symbol  $\mathbb{C}$  bezeichnet. Jedes Element  $z \in \mathbb{C}$  kann auf eindeutige Weise in der Form  $z = a + b\mathbf{j}$  geschrieben werden, mit  $a, b \in \mathbb{R}$ . Man nennt dann  $\operatorname{Re}(z) := a$  den *Realteil* und  $\operatorname{Im}(z) := b$  den *Imaginärteil* von  $z$ .

Die reellen Zahlen bilden eine Teilmenge der komplexen Zahlen (nämlich genau diejenigen komplexen Zahlen, deren Imaginärteil = 0 ist). Die Erweiterung von  $\mathbb{R}$  nach  $\mathbb{C}$  gibt Anlaß zu der Frage, was denn Zahlen überhaupt sind.

Vielleicht bekommen wir eine Antwort, wenn wir versuchen, die algebraische Struktur zu analysieren, die  $\mathbb{R}$  und  $\mathbb{C}$  gemeinsam haben:

**Definition.**

Ein *Körper* ist eine Menge  $K$  mit zwei Verknüpfungen  $+$  und  $\cdot$ , sowie zwei (voneinander verschiedenen) ausgezeichneten Elementen  $0$  und  $1$ , so daß gilt:

1.  $(K, +, 0)$  ist eine kommutative Gruppe.  
Das „Inverse“  $-x$  wird *das Negative zu  $x$*  genannt.

<sup>5</sup>Schon Cardanos Schüler Rafael Bombelli entdeckte um 1570 diese Methode.

2. Sei  $K^* := K \setminus \{0\}$ . Dann ist  
 $(K^*, \cdot, 1)$  ebenfalls eine kommutative Gruppe.  
 Das „Inverse“  $x^{-1}$  wird *das Reziproke zu  $x$*  genannt.
3. Es gilt das Distributivgesetz  $a \cdot (b + c) = a \cdot b + a \cdot c$ .

Dies ist nach der Gruppenstruktur eine weitere wichtige algebraische Struktur. Prominentestes Beispiel ist natürlich  $\mathbb{R}$ , der Körper der reellen Zahlen. Aber auch die Menge  $\mathbb{Q}$  der rationalen Zahlen bildet einen Körper, es wird ja keine Vollständigkeit gefordert. Und man sollte auch beachten, daß es in einem Körper keine Anordnung zu geben braucht. Die komplexen Zahlen, die man sich ja als Punkte der Ebene vorstellen kann, besitzen keine (mit den algebraischen Strukturen verträgliche) Anordnung.

Ein besonders pathologisches Beispiel stellt die Menge  $K_2 := \{0, 1\}$  dar. Sie bildet einen Körper, wenn man definiert:

$$0 + 0 := 0, 0 + 1 = 1 + 0 := 1 \text{ und } 1 + 1 := 0,$$

$$\text{sowie } 0 \cdot 0 := 0, 1 \cdot 0 = 0 \cdot 1 := 0 \text{ und } 1 \cdot 1 := 1.$$

Dann ist z.B.  $-1 = 1$  (in  $K_2$ !). Was hier wie ein rein mathematischer Irrsinn aussieht, dient nicht nur dazu, daß Sie zu Hause Ihrer Oma erzählen können, Sie hätten gelernt, daß  $1 + 1 = 0$  ist. Im Zeitalter von Elektronik und Binär-Arithmetik haben solche Strukturen durchaus auch Anwendungen außerhalb der Mathematik, etwa in der Codierungstheorie.

Wie verrückt die Beispiele auch aussehen mögen, gewisse Regeln bleiben immer erhalten: Ist etwa  $K$  ein beliebiger Körper und  $x \in K$ , so ist  $0 \cdot x = 0$ .

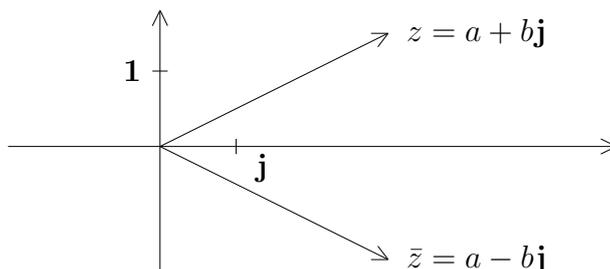
BEWEIS: Es ist  $0 \cdot x = (0 + 0) \cdot x = 0 \cdot x + 0 \cdot x$ .

Da auch  $0 \cdot x = 0 \cdot x + 0$  ist, muß  $0 \cdot x = 0$  sein.  $\square$

Wir wollen nun nachweisen, daß die komplexen Zahlen tatsächlich einen Körper bilden. Was haben wir davon? Wir wissen dann, daß wir mit komplexen Zahlen genauso wie mit reellen Zahlen rechnen können (solange wir nichts benutzen, was mit der Anordnung zu tun hat), und gleichzeitig lernen wir den Umgang mit komplexen Zahlen.

Da wir die Zeilenvektoren von  $\mathbb{R}^2$  mit den Spaltenvektoren von  $\mathcal{V}_2$  identifizieren können, wissen wir aus der Vektorrechnung schon, daß  $(\mathbb{C}, +, 0)$  eine kommutative Gruppe ist. Die 1 ist sicher ein neutrales Element für die Multiplikation. Assoziativität und Kommutativität der Multiplikation kann man – mit genügend Geduld – nachrechnen, genauso das Distributivgesetz. Nun müssen wir noch das Reziproke zu einer komplexen Zahl  $z \neq 0$  finden. Eine direkte Herleitung ist etwas mühsam, deshalb verwenden wir einen einfachen Trick:

Ist  $z = a + b\mathbf{j} \in \mathbb{C}$ , so bezeichnet man  $\bar{z} := a - b\mathbf{j}$  als die zu  $z$  *konjugierte komplexe Zahl*.



Es gilt:

$$1. \overline{z+w} = \bar{z} + \bar{w}.$$

$$\begin{aligned} \overline{(a+b\mathbf{j}) + (c+d\mathbf{j})} &= \overline{(a+c) + (b+d)\mathbf{j}} = (a+c) - (b+d)\mathbf{j} \\ &= (a-b\mathbf{j}) + (c-d\mathbf{j}) = \bar{a} + \bar{b\mathbf{j}} + \bar{c} + \bar{d\mathbf{j}}. \end{aligned}$$

$$2. \overline{z \cdot w} = \bar{z} \cdot \bar{w}.$$

$$\begin{aligned} \overline{(a+b\mathbf{j}) \cdot (c+d\mathbf{j})} &= \overline{(ac-bd) + (ad+bc)\mathbf{j}} = (ac-bd) - (ad+bc)\mathbf{j} \\ \text{und} \\ \overline{(a+b\mathbf{j})} \cdot \overline{(c+d\mathbf{j})} &= (a-b\mathbf{j}) \cdot (c-d\mathbf{j}) = (ac-bd) - (ad+bc)\mathbf{j}. \end{aligned}$$

$$3. \bar{\bar{z}} = z.$$

$$4. \text{ Ist } z = a + b\mathbf{j}, \text{ so ist } z \cdot \bar{z} = (a + b\mathbf{j}) \cdot (a - b\mathbf{j}) = a^2 + b^2 \geq 0. \\ \text{Ist } z \neq 0, \text{ so ist sogar } z \cdot \bar{z} > 0.$$

5. Realteil und Imaginärteil einer komplexen Zahl sind gegeben durch

$$\operatorname{Re}(z) = \frac{1}{2}(z + \bar{z}) \text{ und } \operatorname{Im}(z) = \frac{1}{2\mathbf{j}}(z - \bar{z}).$$

Die reelle Zahl  $|z| := +\sqrt{z\bar{z}} \geq 0$  nennt man den *Betrag* der komplexen Zahl  $z$ . Ist  $z = a + b\mathbf{j}$ , so ist

$$|z| = \sqrt{a^2 + b^2} = \|(a, b)\|$$

die euklidische Norm und besitzt deshalb auch die gleichen Eigenschaften.

Ist nun  $z \neq 0$ , so ist  $z\bar{z} = |z|^2 > 0$ , und es gilt:

$$1 = \frac{z\bar{z}}{z\bar{z}} = z \cdot \frac{\bar{z}}{|z|^2}.$$

Also ist

$$\boxed{z^{-1} = \frac{\bar{z}}{|z|^2}, \text{ für } z \neq 0.}$$

### Definition.

Sei  $I \subset \mathbb{R}$  ein Intervall. Eine Abbildung  $f : I \rightarrow \mathbb{C}$  wird auch als *komplexwertige Funktion* bezeichnet.

Eine komplexwertige Funktion kann immer in der Form  $f = g + \mathbf{j}h$  in Realteil und Imaginärteil zerlegt werden. Da  $g$  und  $h$  gewöhnliche reellwertige Funktionen sind, können wir unsere Erkenntnisse über reelle Funktionen problemlos auf komplexwertige Funktionen ausdehnen:  $f$  ist z.B. stetig, wenn  $g$  und  $h$  es sind.

Schwieriger zu behandeln sind Funktionen, die auch als Argumente komplexe Zahlen zulassen, das werden wir zu einem späteren Zeitpunkt tun. Eine Ausnahme bilden die komplexen Polynome

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0,$$

mit denen wir uns auch schon in diesem Paragraphen beschäftigen werden.

Ist  $z = a + bj$  eine komplexe Zahl mit  $|z| = 1$ , so ist  $a^2 + b^2 = 1$  und es gibt eine eindeutig bestimmte Zahl  $t \in [0, 2\pi)$  mit

$$\cos(t) = a \text{ und } \sin(t) = b.$$

Für beliebige komplexe Zahlen  $z \neq 0$  erhält man dementsprechend eine Darstellung

$$z = |z| \cdot (\cos t + j \sin t).$$

Das ist die sogenannte „Polarkoordinaten-Darstellung“ von  $z$ . Die Zahl  $\arg(z) := t$  nennt man das *Argument* von  $z$ . Erlaubt man, daß  $t$  beliebig in  $\mathbb{R}$  gewählt wird, so ist das Argument nur bis auf  $2\pi$  eindeutig bestimmt. Für  $z = 0$  kann man überhaupt keinen Winkel ermitteln, deswegen haben wir diesen Fall ausgeschlossen.

Nun folgt eine auf den ersten Blick befremdliche Definition:

**Definition.**

Für  $t \in \mathbb{R}$  sei  $e^{jt} := \cos(t) + j \sin(t)$ .

Die Rechtfertigung liefert der folgende Satz:

**I.6.1 Satz.**

1.  $e^{j \cdot 0} = 1$ .
2.  $e^{j(t+s)} = e^{jt} \cdot e^{js}$ .

BEWEIS: 1)  $\cos(0) = 1$  und  $\sin(0) = 0$ .

2) Mit dem Additionstheorem erhält man:

$$\begin{aligned} \cos(t+s) + j \sin(t+s) &= \\ &= \cos(t)\cos(s) - \sin(t)\sin(s) + j[\sin(t)\cos(s) + \cos(t)\sin(s)] \\ &= [\cos(t) + j \sin(t)] \cdot [\cos(s) + j \sin(s)]. \end{aligned}$$

□

**Bemerkung:** Die Beziehung

$$e^{jt} = \cos(t) + j \sin(t)$$

wird als *Eulersche Formel* bezeichnet.

Das Rechnen mit Winkelfunktionen wird auf diese Weise stark vereinfacht. Z.B. gilt:

**I.6.2 Moivre'sche Formel.**

$$(\cos t + j \sin t)^n = \cos(nt) + j \sin(nt).$$

BEWEIS:

$$(e^{\mathbf{j}t})^n = e^{\mathbf{j}nt}.$$

□

Interessant ist auch die folgende „Weltformel“:

$$e^{\mathbf{j}\pi} + 1 = 0.$$

Sie verbindet die wichtigsten Konstanten der Mathematik miteinander: 0, 1,  $\mathbf{j}$ ,  $\pi$  und  $e$ . Der Beweis ist simpel: Es ist  $\cos(\pi) = -1$  und  $\sin(\pi) = 0$ .

**I.6.3 Satz.** Die Gleichung  $z^n = 1$  hat in  $\mathbb{C}$  genau  $n$  Lösungen, nämlich

$$\zeta_{k,n} := e^{\frac{k}{n} \cdot 2\pi \mathbf{j}}, \quad k = 0, 1, \dots, n-1.$$

BEWEIS:

Die Punkte  $\zeta_{k,n} = \cos(k \cdot \frac{2\pi}{n}) + \mathbf{j} \sin(k \cdot \frac{2\pi}{n})$ ,  $k = 0, 1, \dots, n-1$ , liegen auf den Ecken eines (dem Einheitskreis einbeschriebenen) regelmäßigen  $n$ -Ecks. Insbesondere sind sie alle verschieden.

Da  $e^{k \cdot 2\pi \mathbf{j}} = \cos(k \cdot 2\pi) + \mathbf{j} \sin(k \cdot 2\pi) = 1$  ist, ist

$$(\zeta_{k,n})^n = e^{k \cdot 2\pi \mathbf{j}} = 1 \text{ für } k = 0, 1, \dots, n-1.$$

Sei umgekehrt  $w \in \mathbb{C}$  irgendeine Lösung der Gleichung  $z^n = 1$ . Dann ist  $|w|^n = |w^n| = 1$ , also  $|w| = 1$ , also  $w = e^{\mathbf{j}t}$  für ein geeignetes  $t \in [0, 2\pi)$ . Da außerdem  $e^{\mathbf{j}tn} = w^n = 1$  ist, muß gelten:

$$\cos(tn) = 1 \text{ und } \sin(tn) = 0.$$

Das ist nur möglich, wenn  $tn \in \{2\pi k \mid k \in \mathbb{Z}\}$  ist. Da  $t \in [0, 2\pi)$  liegt, kommen für  $tn$  nur die Werte  $0, 2\pi, 4\pi, \dots, (n-1)2\pi$  in Frage. Also muß  $t$  von der Form  $t = \frac{k}{n} \cdot 2\pi$  sein. □

**Definition.**

Die Zahlen  $\zeta_{k,n} := e^{\frac{k}{n} \cdot 2\pi \mathbf{j}}$ ,  $k = 0, 1, \dots, n-1$ , heißen die  $n$ -ten Einheitswurzeln.

Man braucht übrigens zu jedem  $n$  jeweils nur die erste Einheitswurzel zu kennen, denn es ist ja  $\zeta_{k,n} = (\zeta_{1,n})^k$ .

**Beispiel:**

Sei  $n = 3$ . Es ist  $\frac{2\pi}{3} = \text{arc}(120^\circ)$ ,

$$\cos(120^\circ) = -\sin(30^\circ) = -\frac{1}{2} \quad \text{und} \quad \sin(120^\circ) = \sin(60^\circ) = \frac{1}{2}\sqrt{3}.$$

Daraus folgt:

$$\zeta_{0,3} = 1, \quad \zeta_{1,3} = -\frac{1}{2} + \frac{\mathbf{j}}{2}\sqrt{3} \quad \text{und} \quad \zeta_{2,3} = (\zeta_{1,3})^2 = -\frac{1}{2} - \frac{\mathbf{j}}{2}\sqrt{3}.$$

**I.6.4 Satz.** In  $\mathbb{C}$  besitzt jede Zahl  $z \neq 0$  genau  $n$   $n$ -te Wurzeln.

BEWEIS: Sei  $z = re^{jt}$ , mit  $r = |z|$  und einem geeigneten  $t \in [0, 2\pi)$ . Dann setzen wir

$$z_k := \sqrt[n]{r} \cdot e^{\frac{jt}{n}} \cdot \zeta_{k,n}, \quad k = 0, 1, \dots, n-1.$$

Offensichtlich sind dies  $n$  verschiedene komplexe Zahlen  $z_k$  mit  $z_k^n = z$ .

Ist andererseits  $w$  irgendeine Lösung der Gleichung  $w^n = z$ , so ist  $w^n = z_0^n$ , also

$$(wz_0^{-1})^n = 1.$$

Das bedeutet, daß es eine  $n$ -te Einheitswurzel  $\zeta_{k,n}$  gibt, so daß  $w = z_0 \cdot \zeta_{k,n}$  ist. □

Man kann also in  $\mathbb{C}$  nie von *der*  $n$ -ten Wurzel einer Zahl  $z$  sprechen, es gibt stets  $n$  verschiedene. Das gilt auch im Falle  $n = 2$ ! Das Symbol  $\sqrt{z}$  ist also zweideutig. In  $\mathbb{R}$  haben wir dagegen die positive Lösung der Gleichung  $x^2 = a$  als *die* Wurzel aus  $a$  definiert.

Der obige Satz enthält die Aussage:

*Das Polynom  $z^n - w$  besitzt (für jedes  $w \in \mathbb{C}$ ) eine Nullstelle.*

Im Reellen ist das falsch, z.B. hat  $x^2 + 1$  keine Nullstelle in  $\mathbb{R}$ . Das zeigt die Bedeutung des folgenden Satzes:

### I.6.5 Fundamentalsatz der Algebra.

*Jedes nicht konstante komplexe Polynom hat in  $\mathbb{C}$  wenigstens eine Nullstelle.*

Wir verzichten hier auf den Beweis, der eher für Mathematiker interessant ist.

Durch fortgesetzte Polynomdivision erhält man nun:

**I.6.6 Satz.** *Jedes nicht konstante komplexe Polynom läßt sich in Linearfaktoren zerlegen.*

Es gibt eine interessante Anwendung auf reelle Polynome:

### I.6.7 Satz.

*Es sei  $p(z) = \sum_{i=0}^n a_i z^i$  ein Polynom mit **reellen** Koeffizienten. Dann gilt:*

1. *Ist  $\alpha \in \mathbb{C}$  und  $p(\alpha) = 0$ , so ist auch  $p(\bar{\alpha}) = 0$ .*
2. *Die Anzahl der nicht-reellen Nullstellen von  $p$  ist gerade.*
3. *Ist  $n$  ungerade, so besitzt  $p$  mindestens eine reelle Nullstelle.*
4. *Besitzt  $p$  keine reelle Nullstelle, so ist  $p(z)$  Produkt von Potenzen von quadratischen reellen Polynomen.*

BEWEIS: Ist  $\alpha \in \mathbb{C}$ , so ist  $\alpha = \bar{\alpha} \iff \alpha \in \mathbb{R}$ .

1) Da alle  $a_i$  reell sind, ist

$$p(\bar{\alpha}) = \overline{p(\alpha)} = 0.$$

2) Ist  $\alpha \notin \mathbb{R}$ , so ist  $\bar{\alpha} \neq \alpha$ . Also kann man die nicht-reellen Nullstellen in Paaren zusammenfassen.

3) Es gibt  $n = 2m + 1$  komplexe Nullstellen. Wegen (2) muß mindestens eine davon reell sein.

4) Sei  $\alpha = a + b\mathbf{j}$  eine nicht-reelle Nullstelle. Dann enthält  $p(z)$  den Faktor

$$(z - \alpha)(z - \bar{\alpha}) = (z - a - b\mathbf{j})(z - a + b\mathbf{j}) = (z - a)^2 - (b\mathbf{j})^2 = z^2 - 2az + (a^2 + b^2),$$

und der ist ein quadratisches Polynom mit reellen Koeffizienten. Fortgesetzte Polynomdivision ergibt die Behauptung.  $\square$

Zum Schluß dieses Abschnittes soll angedeutet werden, wie komplexe Zahlen in der Wechselstromrechnung angewandt werden:

Ein Punkt  $P$  durchlaufe mit konstanter Geschwindigkeit einen Kreis mit Radius  $A$ , in mathematisch positiver Richtung, also gegen den Uhrzeigersinn.  $T$  sei die Zeit, die er für einen vollen Umlauf benötigt. Der in der Zeit  $t$  überstrichene Winkel beträgt  $\frac{t}{T} \cdot 2\pi$ ,

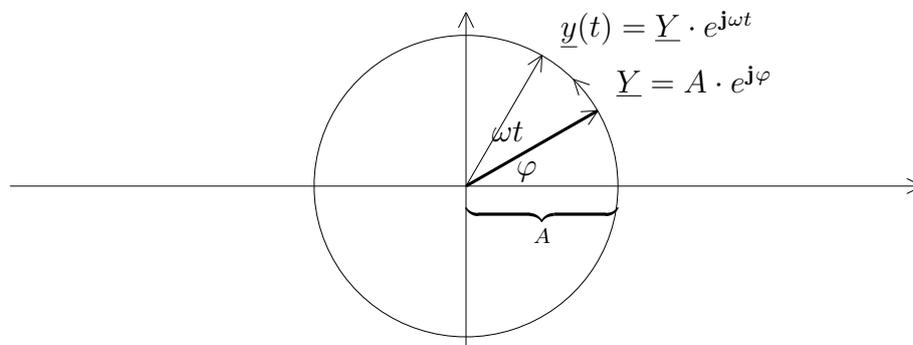
$$\text{die Größe } \omega := \frac{2\pi}{T}$$

ist die Winkelgeschwindigkeit. Nach der Zeitspanne  $t$  hat  $P$  demnach den Winkel  $\omega t$  zurückgelegt.

In komplexer Schreibweise hat  $P$  eine Ausgangsposition  $P_0 = A \cdot e^{j\varphi}$ , den Winkel  $\varphi$  nennt man die *Phase*. Zur Zeit  $t$  hat  $P$  die Position  $P(t) = A \cdot e^{j(\omega t + \varphi)}$ . Die Projektion

$$y(t) := \text{Im } P(t) = A \cdot \sin(\omega t + \varphi)$$

auf die  $y$ -Achse ist eine harmonische Sinusschwingung.



In der Elektrotechnik bezeichnet man die konstante Größe  $\underline{Y} := A \cdot e^{j\varphi}$  als *Zeiger* der Bewegung. Man stelle sich das wie einen Uhrzeiger vor. Die Drehbewegung selber wird durch  $y(t) := \underline{Y} \cdot e^{j\omega t}$  beschrieben, also durch die Drehung des Zeigers (gegen den Uhrzeigersinn) um den Ursprung.

In der Wechselstromtechnik treten Spannung und Stromstärke in Form von harmonischen Schwingungen auf:

$$u(t) := \hat{u} \sin(\omega t + \varphi_u) \quad \text{und} \quad i(t) := \hat{i} \sin(\omega t + \varphi_i).$$

Dabei beschreibt  $\varphi := \varphi_u - \varphi_i$  die Phasenverschiebung zwischen Spannung und Stromstärke. Nun bildet man die komplexen Zeiger

$$\underline{U} := \hat{u} \cdot e^{j\varphi_u} \quad \text{und} \quad \underline{I} := \hat{i} \cdot e^{j\varphi_i},$$

die auch als *komplexe Scheitelwerte* von Spannung und Strom bezeichnet werden.

Die Funktionen

$$\underline{u}(t) := \underline{U} \cdot e^{j\omega t} \quad \text{und} \quad \underline{i}(t) := \underline{I} \cdot e^{j\omega t}$$

sind dann Wechselspannung und Wechselstrom in komplexer Darstellung, und es ist

$$\text{Im } \underline{u}(t) = u(t) \quad \text{und} \quad \text{Im } \underline{i}(t) = i(t).$$

Für jedes  $t$  bezeichnet man das Verhältnis  $\underline{z}(t) := \frac{\underline{u}(t)}{\underline{i}(t)}$  der komplexen Spannung zur komplexen Stromstärke als *komplexen Widerstand*. Es gilt:

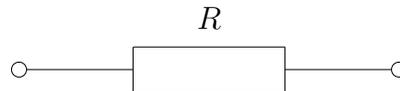
$$\underline{z}(t) := \frac{\underline{U}e^{j\omega t}}{\underline{I}e^{j\omega t}} = \frac{\underline{U}}{\underline{I}} = \frac{\hat{u}}{\hat{i}} e^{j\varphi}.$$

Offensichtlich ist dieser Ausdruck zeitunabhängig, also ein komplexer Zeiger  $\underline{Z}$ . Man nennt  $\underline{Z}$  auch *Widerstandsoperator*, *Scheinwiderstand* oder *Impedanz*.  $R := \text{Re}(\underline{Z})$  heißt *Wirkwiderstand*,  $X := \text{Im}(\underline{Z})$  heißt *Blindwiderstand*. Der Vorteil der komplexen Schreibweise besteht unter anderem darin, daß in Form der Gleichung

$$\underline{U} = \underline{Z} \cdot \underline{I}$$

das Ohmsche Gesetz auch in der Wechselstromtechnik gilt.

Befindet sich in einem Wechselstromkreis ein Ohmscher Widerstand,



so gilt:

$\underline{u}(t) = R \cdot \underline{i}(t)$  für alle  $t$ , also  $\underline{U} = R \cdot \underline{I}$ . Der Ohmsche Widerstand wird daher durch den **reellen** Widerstandsoperator  $\underline{Z} = R$  dargestellt. Insbesondere muß  $\varphi = 0$ , also  $\varphi_u = \varphi_i$  sein, Spannung und Strom sind in Phase.