

# **Mathematik für Naturwissenschaftler**

Wolfgang Schuster

Skriptum zur Vorlesung

Albert-Ludwigs-Universität  
Freiburg i. Br.



Da sie (die Mathematik) ... nur allein Entwicklung der Gesetze des menschlichen Geistes ist, so ist sie nicht allein die gewisseste und zuverlässigste aller menschlichen Wissenschaften sondern auch gewiss die *leichteste*.

Georg Ch. Lichtenberg (1742–1799), Sudelbücher I, Heft 3 [103]

## Vorwort

Das vorliegende Skriptum, immer wieder verbessert und erweitert, begleitet die Vorlesung „Mathematik für Naturwissenschaftler“, die ich seit einigen Jahren regelmäßig an der Universität Freiburg halte. Die Kapitel 1–7 bilden jeweils den Stoff des Wintersemesters, die Kapitel 8–10 den des Sommersemesters. Die Vorlesung wendet sich an Biologen, Geologen, Hydrologen und, seit dem WS 2009/10, auch an Studenten des neu eingerichteten Faches Umwelt(natur) wissenschaft. Sie wendet sich also an Nicht-Mathematiker. Dieser Umstand prägt wesentlich Konzeption und Stil von Vorlesung und Skriptum. Einige Aspekte dieser Konzeption sollen im folgenden angedeutet werden. Die im einzelnen behandelten Themen sind dem Inhaltsverzeichnis zu entnehmen.

Eine formalistische Darstellung des Stoffes wird vermieden zugunsten eines möglichst anschaulichen Vorgehens. Dieses wird unterstützt durch die Behandlung von Anwendungen und mathematischen Modellen, überwiegend aus dem Bereich der Biologie. So zeigt sich z.B. der Nutzen der doch recht abstrakten Aussagenlogik als geeignetes Mittel zur Beschreibung der Funktionsweise einer Nervenzelle (Kap. 1, Abschnitt 1.2).

Ausführlich behandelt werden die komplexen Zahlen als Abschluss der sukzessiven Erweiterungen beim Aufbau unseres Zahlensystems (Kap. 2). Dies ist die systematische Begründung. Zum anderen vereinfacht die Kenntnis der komplexen Zahlen und des Rechnens mit ihnen die Herleitung von Beziehungen an anderen Stellen des Stoffes, wie z.B. bei den trigonometrischen Funktionen. Auch die Einführung des Skalarproduktes von Vektoren gelingt mühelos mit Hilfe komplexer Zahlen. Gleichwohl bereitet das Verständnis der komplexen Zahlen, da im allgemeinen kein Schulstoff mehr, gelegentlich Schwierigkeiten. Eine Passage aus Robert Musils Roman „Die Verwirrungen des Zöglings Törleß“ (abgedruckt am Ende des Kap. 2), zeichnet davon ein genaues Bild und mag zum Troste Betroffener dienen.

Anmerkungen zur Geschichte der Mathematik, zu bedeutenden Mathematikern, erweitern gelegentlich die Darstellung des mathematischen Stoffes. Im Zusammenhang mit den Fibonacci-Zahlen und dem Goldenen Schnitt (Abschnitt 4.3) zeigt sich die tiefliegende philosophische Problematik des Zusammenhangs zwischen der Wahrnehmung

des Schönen in Kunst und Natur und seinen mathematischen Grundlagen. – Dies vielleicht ein kleiner Beitrag zugunsten des durch die Bologna–Reform verdrängten Bildungsaspekts des Studiums.

Für die sorgfältige Erstellung des Textes und für geduldige Überarbeitungen und Erweiterungen des Skriptums danke ich Frau Monika Gilg. Den Herren Ingo Jansen und Ephraim Keinath danke ich für die Erstellung der Abbildungen.

Freiburg, den 12. Oktober 2010

Wolfgang Schuster

# Inhaltsverzeichnis

<b>Vorwort</b>	<b>3</b>
<b>1 Grundbegriffe</b>	<b>7</b>
1.1 Aussagenlogik . . . . .	7
1.2 Das logische Neuron . . . . .	12
1.3 Mengen . . . . .	21
1.4 Mengenalgebra . . . . .	25
1.5 Abbildungen . . . . .	28
<b>2 Komplexe Zahlen</b>	<b>33</b>
2.1 Der Aufbau des Zahlensystems . . . . .	34
2.2 Die Zahl $i$ . . . . .	36
2.3 Die komplexe Ebene . . . . .	38
2.4 Die Euler-Formel . . . . .	41
2.5 Die Gruppe der $n$ -ten Einheitswurzeln . . . . .	46
2.6 Der Fundamentalsatz der Algebra . . . . .	48
<b>3 Kombinatorik</b>	<b>53</b>
3.1 Anzahl der Wörter aus einem Alphabet . . . . .	53
3.2 Anzahl der Permutationen eines Wortes . . . . .	54
3.3 Anzahl der $k$ -elementigen Teilmengen einer $n$ -elementigen Menge . . . . .	56
3.4 Der binomische Lehrsatz . . . . .	58
<b>4 Folgen und Reihen</b>	<b>61</b>
4.1 Beispiele für Folgen . . . . .	61
4.2 Logistisches Wachstum . . . . .	64
4.3 Fibonacci-Wachstum . . . . .	70
4.4 Beispiele für Reihen . . . . .	75
4.5 Konvergenz . . . . .	79
<b>5 Elementare Funktionen</b>	<b>83</b>
5.1 Polynome . . . . .	83
5.2 Rationale Funktionen . . . . .	89

5.3	Die Exponentialfunktion . . . . .	92
5.4	Logarithmus-, allgemeine Exponentialfunktion ... . . . .	94
5.5	Die trigonometrischen Funktionen . . . . .	99
<b>6</b>	<b>Differentialrechnung</b>	<b>111</b>
6.1	Die Ableitung einer Funktion . . . . .	111
6.2	Ableitungsregeln . . . . .	117
6.3	Die Ableitung der elementaren Funktionen . . . . .	121
6.4	Anwendungen der Differentialrechnung . . . . .	128
<b>7</b>	<b>Integralrechnung</b>	<b>149</b>
7.1	Eine Flächenberechnung . . . . .	149
7.2	Der Hauptsatz der Differential- und Integralrechnung . . . . .	153
7.3	Integrationsregeln . . . . .	159
7.4	Integration der elementaren Funktionen . . . . .	164
7.5	Taylorreihen . . . . .	172
<b>8</b>	<b>Lineare Algebra</b>	<b>179</b>
8.1	Vektoren . . . . .	179
8.2	Geraden und Ebenen . . . . .	188
8.3	Matrizen . . . . .	196
8.4	Eigenvektoren und Eigenwerte . . . . .	208
8.5	Das Leslie-Modell . . . . .	211
8.6	Lineare Gleichungssysteme . . . . .	217
<b>9</b>	<b>Differentialgleichungen</b>	<b>233</b>
9.1	Exponentielles Wachstum und verwandte Phänomene . . . . .	234
9.2	Logistisches Wachstum; explosives Wachstum . . . . .	241
9.3	Die lineare Differentialgleichung erster Ordnung . . . . .	247
9.4	Schwingungen . . . . .	252
9.5	Räuber-Beute-Systeme . . . . .	259
<b>10</b>	<b>Stochastik</b>	<b>269</b>
10.1	Vorüberlegungen und Beispiele . . . . .	270
10.2	Grundbegriffe und Grundregeln . . . . .	274
10.3	Die Binomialverteilung . . . . .	280
10.4	Die Poisson-Verteilung . . . . .	285
10.5	Die Normalverteilung . . . . .	290
10.6	Testen . . . . .	299
10.7	Schätzen . . . . .	303
10.8	Die Ausgleichsgerade . . . . .	307
10.9	Der Korrelationskoeffizient . . . . .	312
	<b>Literatur</b>	<b>317</b>

# Kapitel 1

## Grundbegriffe

Die Grundbegriffe, die in diesem Kapitel behandelt werden, sind die Begriffe *Aussage*, *Menge* und *Abbildung*. Damit sind die mathematischen Gebiete Aussagenlogik und Mengenlehre angesprochen. Es kann sich aber im vorliegenden Zusammenhang lediglich um die Vermittlung von Anfangsgründen dieser mathematischen Gebiete handeln mit dem bescheidenen Ziel, in eine nützliche Terminologie und Notation einzuführen. Der Behandlung der Anfangsgründe der Aussagenlogik liegt freilich ein zusätzliches Motiv zugrunde: Sie erlaubt im sogenannten McCulloch-Pitts-Neuron die Formulierung eines Modells, das in vereinfachter Form die Funktionsweise einer Nervenzelle und eines Verbunds von Nervenzellen beschreibt. Damit wird ein erstes Beispiel mathematischer Modellbildung in der Biologie gewonnen.

### 1.1 Aussagenlogik

Unter einer Aussage versteht man einen Satz, dem sinnvollerweise eines der Prädikate „wahr“ oder „falsch“ zukommt. Keine Aussage in diesem Sinne ist also der Satz: „Die Sieben ist betrunken“. Beispiele von Aussagen sind hingegen:

- Der Mond ist rund,
- Der Feminismus ist eine Wissenschaft,
- Geld macht glücklich,
- Der Frosch ist ein Lebewesen,
- 4 ist eine Primzahl.

Aussagen werden gewöhnlich durch kleine lateinische Buchstaben  $p, q, r, s, t, \dots$  symbolisch bezeichnet. Jede Aussage  $p$  besitzt einen *Wahrheitswert*. Wenn  $p$  wahr ist, ordnet man  $p$  die Zahl 1 als ihren Wahrheitswert zu und wenn  $p$  falsch ist, die Zahl 0. (Häufig werden entsprechend auch die Buchstaben  $w, f$  verwendet.) Die Umgangssprache besitzt die Möglichkeit, aus gegebenen Aussagen neue Aussagen zu bilden.

Das geschieht einmal durch Bildung der *Verneinung* einer Aussage  $p$  und zum anderen durch die *Verknüpfung* zweier Aussagen  $p, q$  durch die sogenannten *Junktoren* „und“, „oder“, „wenn, dann“. Man erhält so die Aussagen:

- nicht  $p$ ,
- $p$  und  $q$ ,
- $p$  oder  $q$ ,
- wenn  $p$ , dann  $q$ .

Mit Hilfe dieser Grundverknüpfungen lassen sich dann Aussagen beliebiger Komplexität erzeugen.

Zur Vereinfachung der Schreibweise führt man für die Junktoren *Symbole* ein. Für *nicht* steht  $\neg$ , für *und*  $\wedge$ , für *oder*  $\vee$  und für die *wenn, dann*-Verknüpfung steht  $\rightarrow$ .

Die formale Logik sieht vom *Inhalt* der Aussagen ab und betrachtet sie lediglich unter dem Aspekt ihrer *Wahrheit* bzw. *Falschheit*, das heißt, sie interessiert sich für deren *Wahrheitswert* nur unter einem rein formalen Gesichtspunkt. Bei einer komplexen Aussage bedeutet dies zu untersuchen, wie deren Wahrheitswert von den Wahrheitswerten der *Elementarsätze*  $p, q, r, s, t$ , aus den sie aufgebaut ist, abhängt. Daher heißen Aussagen auch *Wahrheitsfunktionen*.

Von besonderem Interesse sind in diesem Zusammenhang die sogenannten *logischen Wahrheiten*, das sind komplexe Aussagen, die immer wahr sind, unabhängig vom Wahrheitswert ihrer Elementarsätze, wie zum Beispiel die Aussage  $p \vee (\neg p)$ . Eine logische Wahrheit heißt auch *Tautologie*.

Zur exakten Beschreibung der Bedeutung der einzelnen Junktoren dienen sogenannte *Wahrheitstabellen*. Die Aufstellung einer Wahrheitstafel erfolgt in Anlehnung an den umgangssprachlichen Gebrauch des jeweiligen Junktors. Im Falle der oder-Verbindung und der wenn-dann-Verbindung, deren umgangssprachliche Bedeutung ambivalent ist, entscheidet sich die Logik für eine der Bedeutungsvarianten.

## Wahrheitstabellen

Durch Wahrheitstabellen wird die Bedeutung der Junktoren  $\neg, \wedge, \vee, \rightarrow$  also präzise festgelegt. Das geschieht dadurch, dass der Wahrheitswert der durch die Junktoren gebildeten komplexen Aussagen in seiner Abhängigkeit von den möglichen Belegungen der Teilaussagen (Elementarsätze) mit Wahrheitswerten definitorisch ausgedrückt wird. Man sagt daher auch, dass eine komplexe Aussage eine *Wahrheitsfunktion* ihrer Elementarsätze ist. Die Bedeutung der Verneinung wird durch die Wahrheitstabelle 1.1 wiedergegeben, die der übrigen Junktoren definiert die Tabelle 1.2. In Übereinstimmung mit dem umgangssprachlichen Gebrauch ist eine durch „und“ verbundene

Tabelle 1.1: Wahrheitstafel für Verneinung

$p$	$\neg p$
1	0
0	1

Tabelle 1.2: Wahrheitstafel für restliche Junktoren

$p$	$q$	$p \wedge q$	$p \vee q$	$p \rightarrow q$
1	1	1	1	1
1	0	0	1	0
0	1	0	1	1
0	0	0	0	1

Aussage nur dann wahr, wenn beide Teilaussagen wahr sind. Eine „Oder–Aussage“ ist auch dann wahr, wenn beide Teilaussagen wahr sind. Es handelt sich also um das nicht ausschließende „oder“, das dem lateinischen „vel“ entspricht, im Gegensatz zum „entweder, oder“ dem lateinischen „aut“.

Man mache sich die Verknüpfung der Wahrheitstafel für die „wenn, dann“ -Beziehung an folgendem Beispiel klar: Jemand behauptet: „Wenn schönes Wetter ist, dann gehe ich spazieren.“ - Formalisiert:  $p =$  Es ist schönes Wetter,  $q =$  Ich gehe spazieren, also:  $p \rightarrow q$ . Nur im Falle der Wahrheitswertkombination  $(1, 0)$  der zweiten Zeile der Tabelle zum Junktor  $\rightarrow$  kann die obige Behauptung als eindeutig widerlegt betrachtet werden. In allen anderen Fällen muss das Verhalten des Aussagenden entweder als Bestätigung seiner Behauptung  $-(1, 1)$  erste Zeile - aufgefasst werden oder es lässt keinen zuverlässigen Schluss zu. In den Fällen  $(0, 1)$ ,  $(0, 0)$  entscheidet sich der Logiker zugunsten der Wahrheitsliebe des Spaziergängers. Seine Behauptung wird durch sein Verhalten wenigstens nicht widerlegt. Aufgrund der obigen Definitionen für die Junktoren  $\neg, \wedge, \vee, \rightarrow$  lässt sich dann die Wahrheitstafel einer beliebig gebildeten zusammengesetzten Aussage bestimmen.

**Beispiel:** Zur komplexen Aussage  $(p \vee q) \rightarrow (p \wedge (\neg q))$  gehört die Wahrheitstafel 1.3. Die Aussage  $(p \vee q) \rightarrow (p \wedge (\neg q))$  ist genau dann wahr, wenn  $p$  wahr und  $q$  falsch ist oder wenn beide Teilaussagen falsch sind. In den verbleibenden Fällen ist sie falsch. Denkt man sich eine primitive Welt, die durch die wahr–falsch Alternativen bei zwei Aussagen  $p$  und  $q$  vollständig beschrieben wird, dann liefert die komplexe Aussage eine Information über diese Welt. Ist die Aussage wahr, dann befindet sich diese

Tabelle 1.3: Wahrheitstafel

$p$	$q$	$p \vee q$	$\neg q$	$p \wedge (\neg q)$	$(p \vee q) \rightarrow (p \wedge (\neg q))$
1	1	1	0	0	0
1	0	1	1	1	1
0	1	1	0	0	0
0	0	0	1	0	1

Welt im Zustand  $(1, 0)$  oder  $(0, 0)$ , ist sie falsch, im Zustand  $(1, 1)$  oder  $(0, 1)$ . Keine Information über den Zustand einer  $p, q, r$ -Welt vermittelt hingegen die Aussage  $(p \wedge (p \rightarrow q) \wedge (q \rightarrow r)) \rightarrow r$ . Es handelt sich bei dieser Aussage daher um eine logische Wahrheit, auch *Tautologie* genannt, die in allen denkbaren  $p, q, r$ -Welten gültig ist (siehe Tabelle 1.4). ■

Tabelle 1.4: Wahrheitstafel einer komplexen Aussage

$p$	$q$	$r$	$p \rightarrow q$	$q \rightarrow r$	$(p \rightarrow q) \wedge (q \rightarrow r)$	$p \wedge (\dots)$	$(\dots) \rightarrow r$
1	1	1	1	1	1	1	1
1	0	1	0	1	0	0	1
0	1	1	1	1	1	0	1
0	0	1	1	1	1	0	1
1	1	0	1	0	0	0	1
1	0	0	0	1	0	0	1
0	1	0	1	0	0	0	1
0	0	0	1	1	1	0	1

## Logische Folgerung

Zwei Aussagen  $p$  und  $q$  können in der logischen Beziehung der *Folgerung* zueinander stehen. Folgt die Aussage  $q$  aus der Aussage  $p$ , dann schreibt man „ $p \Rightarrow q$ “. Den Beweis, dass die Folgerung  $p \Rightarrow q$  zutreffend ist, führt man mit Hilfe einer Wahrheitstafel indem man zeigt, dass die Aussage  $p \rightarrow q$  eine Tautologie ist. Der Doppelpfeil  $p \Leftrightarrow q$  zwischen zwei Aussagen  $p$  und  $q$  bedeutet: Es gilt  $p \Rightarrow q$  und zugleich  $q \Rightarrow p$ . Das heißt, die Aussagen  $p$  und  $q$  folgen wechselseitig auseinander, oder:  $p$  ist genau dann wahr, wenn auch  $q$  wahr ist, das heißt, die Aussage  $p$  und  $q$  sind *logisch äquivalent*. Das aber bedeutet, wenn  $p$  wahr ist, dann ist auch  $q$  wahr und wenn  $p$  falsch

ist, dann ist auch  $q$  falsch. Das heißt, beide Aussagen haben die gleiche Wahrheitstafel.

So kann man zum Beispiel folgern:  $(p \wedge q) \implies p$ , denn die Aussage  $(p \wedge q) \rightarrow p$  ist eine Tautologie, wie die nachstehende Wahrheitstafel zeigt

Tabelle 1.5: Beweis  $(p \wedge q) \implies p$ 

$p$	$q$	$p \wedge q$	$p$	$(p \wedge q) \rightarrow p$
1	1	1	1	1
1	0	0	1	1
0	1	0	0	1
0	0	0	0	1

Die Aussagen  $\neg(p \wedge q)$  und  $(\neg p) \vee (\neg q)$  sind logisch äquivalent:  $\neg(p \wedge q) \Leftrightarrow (\neg p) \vee (\neg q)$ . Dies zeigt die folgende Wahrheitstafel. Da die 4. und 7. Spalte der Tabelle überein-

Tabelle 1.6: Beweis  $\neg(p \wedge q) \Leftrightarrow (\neg p) \vee (\neg q)$ 

$p$	$q$	$p \wedge q$	$\neg(p \wedge q)$	$\neg p$	$\neg q$	$(\neg p) \vee (\neg q)$
1	1	1	0	0	0	0
1	0	0	1	0	1	1
0	1	0	1	1	0	1
0	0	0	1	1	1	1

stimmen, ist die behauptete Äquivalenz gezeigt.

## Algebraisierung

An Stelle der Wahrheitstafeln für die Junktoren  $\neg$ ,  $\wedge$ ,  $\vee$ ,  $\rightarrow$  lassen sich auch algebraische Ausdrücke (Formeln) angeben, die auf rein rechnerischem Wege die jeweils richtigen Zuordnungen zwischen den Wahrheitswerten der Teilaussagen  $p$ ,  $q$  und dem Wahrheitswert der Gesamtaussage liefern, wenn die entsprechenden Wahrheitswerte in die Formeln eingesetzt werden. Dies leisten die Formeln

- $\neg p = 1 - p$ ,
- $p \wedge q = pq$ ,
- $p \vee q = p + q - pq$ ,
- $p \rightarrow q = 1 - p + pq$ ,

wie man leicht nachrechnet. Mit Hilfe dieser Formeln lässt sich dann durch „Ausrechnen“ die Wahrheitstafel einer Aussage erstellen. Bei der Ausführung einer solchen Rechnung hat man lediglich zu beachten, dass immer gilt:  $p^2 = p$ , denn es ist  $0^2 = 0$  und  $1^2 = 1$ .

**Beispiel:** Wir betrachten die Aussage  $((p \rightarrow q) \wedge p) \rightarrow q$  und ersetzen sukzessive die durch die Junktoren  $\wedge$  und  $\rightarrow$  verbundenen Teilaussagen durch die obigen Formeln. Man erhält

$$\begin{aligned}(p \rightarrow q) \wedge p &= (1 - p + pq) \cdot p \\ &= p - p^2 + p^2q \\ &= p - p + pq \\ &= pq,\end{aligned}$$

also gilt:

$$\begin{aligned}(p \rightarrow q) \wedge p) \rightarrow q &= pq \rightarrow q \\ &= 1 - pq + pq \cdot q \\ &= 1 - pq + pq = 1.\end{aligned}$$

Es liegt somit wie bei der Aussage von Tabelle 1.4 eine logische Wahrheit vor, und zwar handelt es sich um die aus der klassischen Logik bekannte Schlussregel des „modus ponens“. Sie erlaubt zum Beispiel aus den Prämissen „Alle Menschen sind sterblich“ und „Sokrates ist ein Mensch“ den Schluss zu ziehen: „Sokrates ist sterblich“.

## 1.2 Das logische Neuron

Das logische Neuron oder McCulloch-Pitts-Neuron (Waren S. McCulloch, 1898-1972, und Walter Pitts, geboren 1924, amerikanische Neurophysiologen, entwickelten 1943 das nach ihnen benannte Modell einer Nervenzelle) stellt ein stark idealisiertes Modell des Baues und der Funktionsweise von Nervenzellen dar, wie sie zum Beispiel im menschlichen Gehirn und Rückenmark zu finden sind. Eine Nervenzelle ist mit benachbarten Zellen durch Nervenfasern verbunden, über die sie elektrische Signale (Inputs) empfängt, verarbeitet und weitergibt (Output). Über die Inputs erhält das Neuron Informationen in Form von elektrischen Impulsen, die zum Beispiel durch Reizung von Sinnesorganen ausgelöst werden. Die Inputs zerfallen dabei in zwei Typen: erregende und hemmende. Durch die eingehenden Impulse wird in der Nervenzelle ein elektrisches Potential aufgebaut. Hierzu leisten aber nur die erregenden Inputs einen positiven Beitrag, während die hemmenden Inputs das Potential vermindern. Erreicht oder überschreitet das Potential einen kritischen Wert (Schwellenwert), dann gibt das Neuron über den Output einen elektrischen Impuls ab, der dann an ein benachbartes Neuron weitergeleitet wird oder direkt eine bestimmte physische Reaktion auslöst.

## Die Funktionsweise des logischen Neurons

In das Neuron münden also zwei Typen von Nervenfasern, erregende ( $\rightarrow$ ) und hemmende ( $\dashrightarrow$ ) ein. Die einlaufenden Impulse bauen im Inneren des Neurons ein elektrisches Potential auf. Zur Bildung dieses Potential liefern aber, wie gesagt, nur die erregenden Fasern einen positiven Beitrag, während ein Impuls (Input) über eine hemmende Faser zur Verminderung des Potentials führt. Erreicht oder überschreitet der Wert  $P$  des Potentials einen Schwellenwert  $\Theta$ , dann sendet die Nervenzelle über die ausgehende Faser ein Signal (Output) an eine benachbarte Zelle aus. Man nehme zu-

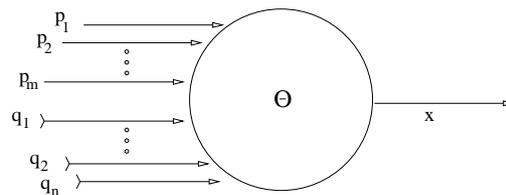


Abbildung 1.1: Schema des logischen Neurons.

sätzlich an, dass die Informationsübertragung auf das Neuron „taktweise“, dass heißt, zu diskreten Zeitpunkten erfolgt. Dabei wird normalerweise nur ein Teil der Inputs aktiv sein (ein Signal übertragen) der Rest bleibt inaktiv. Wir verabreden, einem aktiven Input die Zahl 1, einem inaktiven die Zahl 0 zuzuordnen:

$$\begin{aligned} p &= 1 : && \text{Input ist aktiv} \\ p &= 0 : && \text{Input ist inaktiv} \end{aligned}$$

Münden in die Nervenzellen die erregenden Inputs  $p_1, p_2, \dots, p_m$  und die hemmenden Inputs  $q_1, q_2, \dots, q_n$  ein, dann wird das in der Nervenzelle aufgebaute Potential durch den Ausdruck

$$P = p_1 + p_2 + \dots + p_m - q_1 - q_2 - \dots - q_n$$

repräsentiert, wobei man sich für  $p$  und  $q$  die Werte 0 oder 1 eingesetzt denkt. Erreicht oder überschreitet  $P$  den Wert  $\Theta$ , dann „feuert“ der Output  $x$ , sonst nicht. Der Schwellenwert  $\Theta$  kann die Werte  $0, 1, 2, \dots$  annehmen. Das Verhalten des Neurons lässt sich also charakterisieren durch die Funktion

$$\chi = p_1 + p_2 + \dots + p_m - q_1 - q_2 - \dots - q_n - \Theta,$$

die *charakteristische Funktion* des Neurons. Es gilt die *Regel*:

$$\begin{aligned} \chi &\geq 0 : && \text{Output feuert, dass heißt, } x = 1 \\ \chi &< 0 : && \text{Output feuert nicht, dass heißt, } x = 0. \end{aligned}$$

Im nächsten Abschnitt wird gezeigt, wie einerseits Wahrheitsfunktionen durch Neuronschaltungen realisiert werden können und andererseits einem Neuron eine Wahrheitsfunktion zugeordnet werden kann.

## Die elementaren Wahrheitsfunktionen als Neuronen

Die elementaren Wahrheitsfunktionen  $\neg p$ ,  $p \wedge q$ ,  $p \vee q$ ,  $p \rightarrow q$  lassen sich leicht als Neuronen realisieren. Das Neuron mit einem hemmenden Input  $p$  und dem Schwellenwert  $\Theta = 0$  realisiert offenbar die Verneinung.

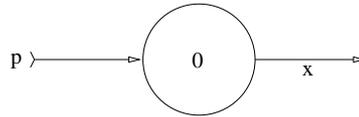


Abbildung 1.2: Nicht-Neuron.

Tabelle 1.7: Funktionsschema  $\neg$ -Neuron

$p$	$\chi = -p$	$x = \neg p$
1	-1	0
0	0	1

Der Vergleich von erster und letzter Spalte der Tabelle 1.7 zeigt, dass die Reaktion des Outputs  $x$  der Wahrheitsfunktion  $\neg p$  entspricht. Das Neuron mit zwei erregenden Inputs  $p$  und  $q$  mit Schwellenwert  $\Theta = 2$  verhält sich wie die „und“-Verbindung.

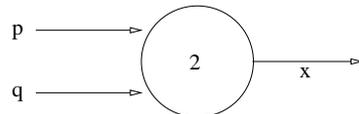


Abbildung 1.3: und-Neuron.

Tabelle 1.8: Funktionsschema  $\wedge$ -Neuron

$p$	$q$	$\chi = p + q - 2$	$x = p \wedge q$
1	1	0	1
1	0	-1	0
0	1	-1	0
0	0	-2	0

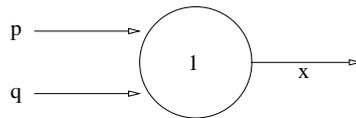


Abbildung 1.4: oder-Neuron.

Tabelle 1.9: Funktionsschema  $\vee$ -Neuron

$p$	$q$	$\chi = p + q - 1$	$x = p \vee q$
1	1	1	1
1	0	0	1
0	1	0	1
0	0	-1	0

Das „oder“-Neuron entsteht aus dem „und“-Neuron, indem man den Schwellenwert auf  $\Theta = 1$  absenkt. Die Struktur des Nicht-Neurons, des und- sowie des oder-Neurons lässt sich so gleichsam erraten. Man kann allerdings auch systematisch vorgehen und die jeweilige Struktur ausrechnen. Wie man dies bewerkstelligt, soll am Fall des Wenn-dann-Neurons gezeigt werden.

Die allgemeine Form der charakteristischen Funktion eines Neurons mit zwei Eingängen  $p$  und  $q$  ist

$$(*) \quad \chi = \alpha p + \beta q - \Theta$$

mit Faktoren  $\alpha, \beta$ , die jeweils die Werte  $\pm 1$  annehmen können. Ist  $\alpha = +1$ , dann ist der Input  $p$  erregend, bei  $\alpha = -1$  ist er hemmend. Dasselbe gilt für den Input  $q$ . Der Schwellenwert  $\Theta$  kann im Prinzip die ganzzahligen Werte  $0, 1, 2, \dots$  annehmen. Da aber der Ausdruck  $\alpha p + \beta q$  höchstens den Wert 2 erreichen kann, wäre ein Neuron mit zwei Inputs und einem Schwellenwert  $\Theta \geq 3$  funktionslos, da in diesem Fall immer  $\chi \leq -1$  gelten würde.

Aufgrund der Wahrheitstabellen für die  $\rightarrow$ -Aussage (vgl. Tabelle 1.1) ergeben sich für die charakteristische Funktion (\*) die folgenden Ungleichungen, wenn man die vier 0,1-Alternativen für die Inputs  $p$  und  $q$  einsetzt:

- (1)  $\alpha + \beta - \Theta \geq 0,$
- (2)  $\alpha - \Theta < 0,$
- (3)  $\beta - \Theta \geq 0,$
- (4)  $- \Theta \geq 0.$

Von den möglichen Werten  $\Theta = 0, 1, 2$  erfüllt nur der Wert  $\Theta = 0$  die Ungleichung (4). Dann folgt aus (3)  $\beta \geq 0$ . Da  $\beta$  nur die Werte  $\pm 1$  annehmen kann, muss  $\beta = 1$  sein. Ungleichung (2) besagt  $\alpha < 0$ , also ist  $\alpha = -1$ . Mit den Werten  $\alpha = -1$ ,  $\beta = 1$  und  $\Theta = 0$  ist aber auch die Ungleichung (1) erfüllt. Also hat das Wenn–dann–Neuron den hemmenden Input  $p$ , den erregenden Input  $q$  und den Schwellenwert  $\Theta = 0$ . Seine charakteristische Funktion hat daher die Form  $\chi = q - p$ .

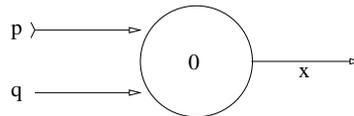


Abbildung 1.5: wenn, dann–Neuron.

Tabelle 1.10: Funktionsschema  $\rightarrow$ -Neuron

$p$	$q$	$\chi = q - p$	$x = p \rightarrow q$
1	1	0	1
1	0	-1	0
0	1	1	1
0	0	0	1

### Bestimmung der einem Neuron zugeordneten Wahrheitsfunktion

Wie man die zu einem gegebenen Neuron gehörige Wahrheitsfunktion findet, wird an einem Beispiel veranschaulicht. Es sei  $X$  das Neuron mit zwei erregenden Inputs  $p, q$ , einem hemmenden Input  $r$  und dem Schwellenwert  $\Theta = 1$ .

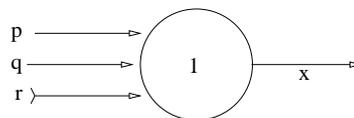


Abbildung 1.6: Bestimmung der Wahrheitsfunktion.

Seine *charakteristische Funktion* ist also  $\chi = p + q - r - 1$ . Wir stellen die zugehörige Input–Output–Tabelle auf:

Die Bestimmung der entsprechenden Wahrheitsfunktion lässt sich nun in drei Schritte zerlegen:

Tabelle 1.11: Input–Output–Tabelle

	$p$	$q$	$r$	$\chi = p + q - r - 1$	$x$
*	1	1	1	0	1
	1	0	1	-1	0
	0	1	1	-1	0
	0	0	1	-2	0
*	1	1	0	1	1
*	1	0	0	0	1
*	0	1	0	0	1
	0	0	0	-1	0

*Schritt 1:* Man markiere diejenigen Zeilen (\*) der Tabelle 1.11, zu denen unter dem Output  $x$  eine 1 steht.

*Schritt 2:* Jeder \*-Zeile wird unter Verwendung der Junktoren  $\neg$  und  $\wedge$  eine Wahrheitsfunktion zugeordnet, die ausschließlich für die in der \*-Zeile links stehenden Kombinationen von Wahrheitswerten wahr ist. Im Beispiel also (von oben nach unten):

$$\begin{aligned}
 & p \wedge q \wedge r \\
 & p \wedge q \wedge (\neg r) \\
 & p \wedge (\neg q) \wedge (\neg r) \\
 & (\neg p) \wedge q \wedge (\neg r)
 \end{aligned}$$

*Schritt 3:* Die so gewonnenen Wahrheitsfunktionen werden durch den  $\vee$ -Junktor verknüpft. Man erhält so im Beispiel die Wahrheitsfunktion

$$x = [p \wedge q \wedge r] \vee [p \wedge q \wedge (\neg r)] \vee [p \wedge (\neg q) \wedge (\neg r)] \vee [(\neg p) \wedge q \wedge (\neg r)].$$

Setzt man in  $x$  eine \*-Kombination von Wahrheitswerten ein, dann wird eine der Teilaussagen in den eckigen Klammern wahr und damit die gesamte  $\vee$ -Verbindung  $x$ . Setzt man eine andere Kombination ein, so sind nach Konstruktion sämtliche dieser Teilaussagen falsch, also auch  $x$ . Damit ist gezeigt, dass das Verhalten der Wahrheitsfunktion  $x$  die Reaktion des Neurons auf die Inputs genau widerspiegelt.

**Bemerkung:** Man mache sich klar, dass eine reine iterierte  $\wedge$ -Verbindung  $p \wedge q \wedge \dots \wedge r$  genau dann wahr ist, wenn sämtliche Elementarsätze wahr sind, und eine reine iterierte  $\vee$ -Verbindung  $p \vee q \vee \dots \vee r$  genau dann falsch ist, wenn sämtliche Elementarsätze falsch sind und dass Klammern in diesen Ausdrücken weggelassen werden können, da

$p \wedge (q \wedge r) = (p \wedge q) \wedge r$  und  $p \vee (q \vee r) = (p \vee q) \vee r$  gilt. Das Gleichheitszeichen zwischen zwei Wahrheitsfunktionen bedeutet, dass ihre Wahrheitstabellen identisch sind. ▲

## Neuronennetze

Jede Wahrheitsfunktion kann durch ein Neuronennetz simuliert werden. Dies zeigen wir an einem Beispiel: Gegeben sei die Wahrheitsfunktion

$$x = (p \vee q) \rightarrow ((\neg q) \wedge r).$$

Mit Hilfe der bereits beschriebenen elementaren Neuronen, welche die Junktoren  $\neg$ ,  $\wedge$ ,  $\vee$ ,  $\rightarrow$  repräsentieren, kann  $x$  durch ein Netzwerk simuliert werden. Damit die Impulse, die über die Inputs  $p, q, r$  eintreffen, im Gleichtakt bei den Neuronen ankommen, müssen Neuronen (\*) der logischen Form  $p$  mit einem erregenden Input  $p$  und dem Schwellenwert 1 zwischengeschaltet werden. Diese leiten einen ankommenden Impuls einfach durch. Das Netzwerk baut man folgendermaßen auf: In ein  $\rightarrow$ -Neuron mit Output  $x$  münden die Inputs  $p \wedge q$  und  $(\neg q) \vee r$ . Diese werden wiederum durch die entsprechenden Elementarneurone dargestellt. Dabei muss auf Gleichtakt beim Durchlauf der Signale geachtet werden. Auf diese Weise entsteht das Netzwerk in Abbildung 1.7. Es reagiert auf eine Impulskombination bei den Eingängen  $p, q, r$  analog zur Wahrheitsfunktion  $x$ .

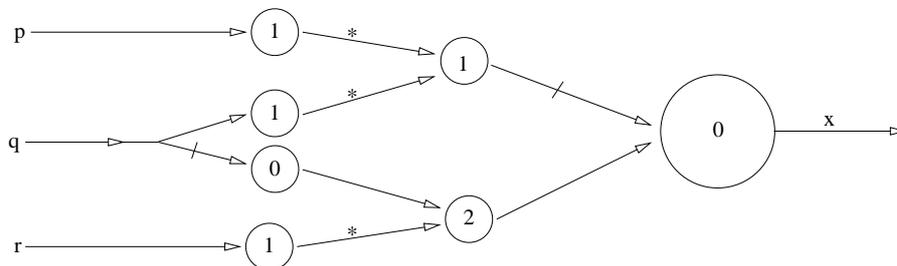


Abbildung 1.7: Wahrheitsfunktion als Neuronennetz.

**Bemerkung:** Es ist nicht immer möglich, eine Wahrheitsfunktion durch ein *einziges* Neuron zu simulieren. Die Anzahl der Wahrheitsfunktionen, die mit  $n$  Elementarsätzen gebildet werden können, ist nämlich größer, als die Zahl der Neuronen mit  $n$  Eingängen. Die Wahrheitstabelle, die zu einer Aussage mit  $n$  Elementarsätzen gehört, hat  $2^n$  Zeilen, denn  $2^n$  ist die Anzahl der Möglichkeiten,  $n$  Positionen mit 0 oder 1 zu besetzen. Die Anzahl der 0, 1-Kombinationen wiederum, die man den  $2^n$  Zeilen zuordnen kann und damit die Anzahl der Wahrheitsfunktionen mit  $n$  Elementarsätzen ist daher gleich  $2^{2^n}$  (vgl. Kapitel 3). Damit die charakteristische Funktion  $\chi$  eines logischen Neurons nicht nur negative Werte annimmt, darf der Schwellenwert  $\Theta$  bei  $n$  Inputs

den Wert  $n$  nicht überschreiten, kann also nur die  $n + 1$  Werte  $0, 1, 2, \dots, n$  annehmen. Da es für die Besetzung mit den Alternativen „erregend“ und „hemmend“ bei  $n$  Eingängen  $2^n$  Möglichkeiten gibt, kann es höchstens  $2^n(n + 1)$  sinnvolle Neuronen mit  $n$  Eingängen geben. Für  $n \geq 2$  ist aber  $2^n(n + 1) < 2^{2^n}$ , so dass die Wahrheitsfunktionen mit zwei oder mehr Elementarsätzen nicht sämtlich durch ein einziges logisches Neuron simuliert werden können. ▲

Das Neuronennetz in Abbildung 1.8 stellt ein einfaches Modell für die Reaktion eines Organismus auf äußere Reize dar. Ein einziger, bei  $p$  eintretender Impuls  $p_1$  durchläuft

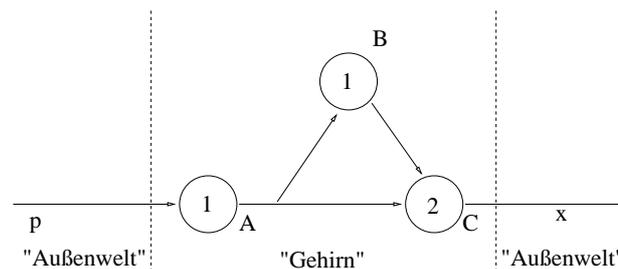


Abbildung 1.8: Modell für eine Reizschwelle.

Neuron  $A$  und wird dann in zwei Inputs aufgespalten, die bei  $B$  und  $C$  eintreffen. Da  $B$  für den ankommenden Impuls durchlässig ist, treffen bei  $C$  zwei Impulse nacheinander ein, ohne dass wegen des Schwellenwertes 2 bei  $C$  der Output  $x$  feuert. Erst zwei nacheinander eintretende Impulse  $p_1$  und  $p_2$  lösen eine Reaktion bei  $x$  aus. Es muss also eine „Reizschwelle“ überschritten werden, damit eine Reaktion des Systems erfolgt.

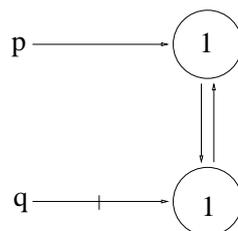


Abbildung 1.9: Primitiv-Modell eines Gedächtnisses.

Das Neuronennetz in Abbildung 1.9 stellt das Primitiv-Modell eines *Gedächtnisses* dar. Ein bei  $p$  eingehender Input (Information) durchläuft beständig das Netzwerk („wird im Gedächtnis behalten“), sofern bei  $q$  kein Input eingeht. Eine einmalige Störung durch einen Input bei  $q$  hingegen unterbricht den Kreislauf, der die Information speichert. Das bedeutet, der Gedächtnisinhalt wird ausgelöscht.

Ein Nervensystem bestehe aus drei Neuronen  $A$ ,  $B$ ,  $C$  mit gewissen Verbindungen (siehe Abbildung 1.10). Ein „Gemütszustand“ des Systems wird bestimmt durch die

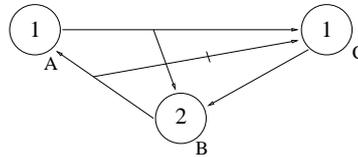


Abbildung 1.10: Nervensystem.

Angabe, welches der Neuronen  $A$ ,  $B$ ,  $C$  gerade feuert oder nicht feuert. Es gibt demnach  $2 \cdot 2 \cdot 2 = 8$  mögliche Zustände des Systems. Jeder Zustand bestimmt eindeutig den nachfolgenden Zustand. Die möglichen Zustände ① bis ⑧ werden in der Tabelle 1.12 zusammen mit ihren Folgezuständen dargestellt. (Durch welchen Einfluß das System in den jeweiligen Anfangszustand versetzt wird, kann hier offengelassen werden.)

Tabelle 1.12: Mögliche Zustände.

	$A$	$B$	$C$	$\rightarrow$	$A$	$B$	$C$	
①	1	1	1	$\rightarrow$	1	1	0	⑤
②	1	0	1	$\rightarrow$	0	1	1	③
③	0	1	1	$\rightarrow$	1	0	0	⑥
④	0	0	1	$\rightarrow$	0	0	0	⑧
⑤	1	1	0	$\rightarrow$	1	0	0	⑥
⑥	1	0	0	$\rightarrow$	0	0	1	④
⑦	0	1	0	$\rightarrow$	1	0	0	⑥
⑧	0	0	0	$\rightarrow$	0	0	0	⑧

Eine Übersicht über das „Gemütsleben“ des Systems, das heißt über die Aufeinanderfolge seiner Seelenzustände, gibt das Diagramm in Abbildung 1.11: wieder. Fassen wir den Zustand ⑧ =  $(0, 0, 0)$  als „Depression“ auf, so können wir folgern: Gleichgültig welcher Anfangszustand vorliegt, das System verfällt schließlich, wenn eine Stimulation von außen ausbleibt, immer in eine Depression.

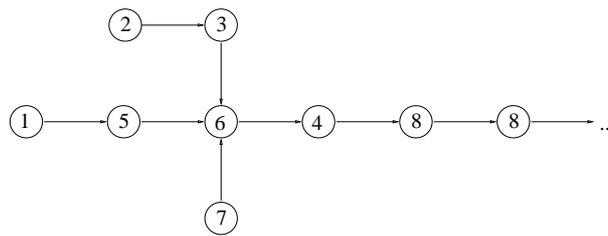


Abbildung 1.11: „Depressives“ Gemütsleben.

## 1.3 Mengen

Die Mengenlehre bildet ein Teilgebiet der Mathematik und ihre Begriffsbildungen sind zu ihrem systematischen Aufbau unentbehrlich. Sie dienen uns im vorliegenden Kontext aber nur zur bequemen und präzisen Klassifizierung von Gegenstandsbereichen und deren logischen Beziehungsverhältnissen. Insbesondere bei der Formulierung der Grundbegriffe der Wahrscheinlichkeitsrechnung (vgl. Kapitel ??) kann auf die Mengenlehre nur schwer verzichtet werden.

Der Begriff der *Menge* ist ein undefinierter Grundbegriff. Das ist aber kein Mangel, denn die Kette der definitorischen Erklärungen kann nicht unendlich fortgesetzt werden, sondern muss irgendwo mit formal undefinierten Grundbegriffen beginnen. Mit Georg Cantor (1845-1918), dem Begründer der Mengenlehre, lässt sich der Mengenbegriff folgendermaßen umschreiben:

Unter einer „Menge“ verstehen wir jede Zusammenfassung  $M$  von bestimmten wohlunterschiedenen Objekten  $m$  unserer Anschauung oder unseres Denkens (welche die „Elemente“ von  $M$  genannt werden) zu einem Ganzen.

Man unterscheidet also zwischen der Menge  $M$  und den Objekten  $m$ , oder den Elementen  $x$ , die zu  $M$  gehören. Folgende *Bezeichnungen* sind üblich:

$M, A, B, C, \dots$	: Mengensymbole
$x, a, b, c, \dots$	: Elementsymbole
$x \in M$	: $x$ ist ein Element von $M$
$x \notin M$	: $x$ ist nicht Element von $M$

Das Symbol  $\in$  ist ein stilisiertes griechisches  $\varepsilon$  (Epsilon). Die Abgrenzung einer Menge  $M$  aus einer größeren Gesamtheit von Objekten geschieht durch Angabe einer *Eigenschaft*  $E$ , die genau den Elementen von  $M$  zukommt. Man schreibt *symbolisch*:

$$M = \{x \mid E(x)\}.$$

Das bedeutet:  $M$  ist die Menge aller derjenigen Objekte  $x$ , denen die Eigenschaft  $E$  zukommt. Hat  $M$  nur endliche viele Elemente, so kann  $M$  auch durch Aufzählung seiner Elemente bestimmt werden

**Beispiel:** Wenn  $W$  die Eigenschaft bezeichnet, einen Wirbel zu besitzen, dann ist  $W = \{x \mid W(x)\}$  die Menge der Wirbeltiere. ■

Wir gehen im Folgenden davon aus, dass sämtliche vorkommenden Mengen in einer Universalmenge  $U$ , einem Mengen–Universum, enthalten sind (siehe Abbildung 1.12).



Abbildung 1.12: Die Universalmenge  $U$ .

### Teilmengen, Gleichheit von Mengen

$A$  ist eine Teilmenge der Menge  $B$ , symbolisch  $A \subset B$  wenn gilt: jedes Element  $x$  von  $A$  ist auch Element von  $B$ . Diese Beziehung kann man formal auch so ausdrücken:  $x \in A \Rightarrow x \in B$  (siehe Abbildung 1.13). Zwei Mengen  $A$  und  $B$  sind gleich, wenn die Beziehung  $A \subset B$  und  $B \subset A$  gleichzeitig erfüllt sind.

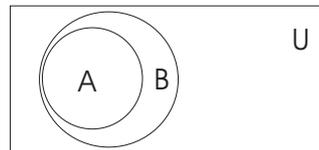


Abbildung 1.13:  $A$  ist Teilmenge von  $B$ .

Aus gegebenen Mengen  $A, B \subset U$  lassen sich nach mehreren Regeln neue Mengen bilden.

### Die leere Menge

Aus systematischen Gründen ist es zweckmäßig, eine Menge  $\emptyset$  zu definieren, die keine Elemente enthält, die *leere Menge*. So handelt es sich zum Beispiel bei der Menge aller ganzen Zahlen, die zugleich gerade und ungerade sind, um die leere Menge. Oder: Bei zwei Mengen  $A$  und  $B$ , die keine Elemente gemeinsam haben, schreibt man symbolisch:

$$A \cap B = \emptyset.$$

Man sagt dann auch:  $A$  und  $B$  sind *disjunkte* Mengen. Die leere Menge ist Teilmen-

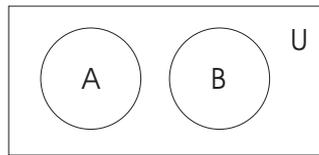


Abbildung 1.14:  $A \cap B = \emptyset$ .

ge einer jeden Menge  $M$ :  $\emptyset \subset M$ . Dies ergibt sich rein formal aus dem Umstand, dass die wenn-dann-Aussage  $x \in \emptyset \rightarrow x \in M$  immer wahr ist, da die Aussage  $x \in \emptyset$  für alle  $x$  falsch ist. Daher ist die Folgerung  $x \in \emptyset \Rightarrow x \in M$  richtig und es gilt  $\emptyset \subset M$ .

### Komplement

Unter dem *Komplement*  $\bar{A}$  einer Menge  $A$  versteht man die Gesamtheit der Elemente  $x$  aus der Universalmenge  $U$ , die nicht in  $A$  liegen:  $\bar{A} = \{x \mid x \in U \wedge x \notin A\}$

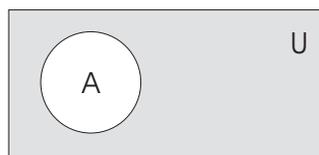


Abbildung 1.15: Das Komplement einer Mengen.

### Vereinigungsmenge

Die *Vereinigungsmenge*  $A \cup B$  zweier Mengen  $A, B$  besteht aus denjenigen Elementen  $x$ , die in der Menge  $A$  oder in der Menge  $B$  liegen:  $A \cup B = \{x \mid x \in A \vee x \in B\}$ .

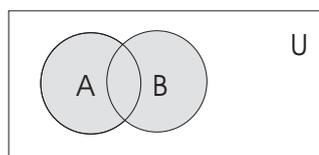


Abbildung 1.16: Die Vereinigungsmenge zweier Mengen.

### Durchschnittsmenge

Die *Durchschnittsmenge*  $A \cap B$  zweier Mengen  $A, B$  besteht aus allen Elementen  $x$ , die zugleich in  $A$  und in  $B$  enthalten sind:  $A \cap B = \{x \mid x \in A \wedge x \in B\}$ . Haben die Mengen  $A$  und  $B$  kein Element gemeinsam, dann ist, wie gesagt, ihr Durchschnitt

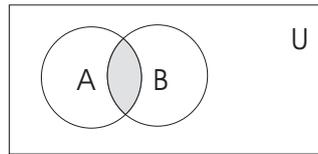


Abbildung 1.17: Der Durchschnitt zweier Mengen.

gleich der *leeren Menge*  $\emptyset$ :  $A \cap B = \emptyset$ .

### Cartesisches Produkt

Einen neuen Typ von Mengen erhält man durch Bildung des *Cartesischen Produktes*  $A \times B$  zwei Mengen  $A, B$ :  $A \times B = \{(x, y) \mid x \in A, y \in B\}$ .  $A \times B$  ist also die Menge aller *geordneten Paare*  $(x, y)$ , wobei die linke Komponente  $x$  ein Element aus  $A$  und die rechte Komponente  $y$  ein Element aus  $B$  ist. Geordnet bedeutet, dass es bei der Paarbildung auf die Reihenfolge der Elemente ankommt, das heißt, es ist  $(x, y) \neq (y, x)$ , sofern  $x \neq y$  ist.

**Beispiel:** Sei  $G = \{A, a\}$  die Menge der zu einem „Locus“ gehörigen allelen Gene. Die Menge der zugehörigen Genotypen lässt sich beschreiben durch die Menge

$$\begin{aligned} G \times G &= \{(x, y) \mid x \in G, y \in G\} \\ &= \{(A, A); (A, a); (a, A); (a, a)\} \end{aligned}$$

Die Elemente  $(A, a)$ ,  $(a, A)$  sind natürlich biologisch zu identifizieren. ■

Das Cartesische Produkt kann man in naheliegender Weise auch für mehr als zwei Mengen bilden. So ist

$$A \times B \times C = \{(x, y, z) \mid x \in A, y \in B, z \in C\}$$

die Menge aller geordneten Tripel mit den Komponenten aus  $A, B, C$ .

**Beispiel (Die Blutgruppen):** Eine Menge  $U$  von Individuen wird auf das Vorhandensein der *Antigene*  $A, B, Rh$  hin untersucht. Es lassen sich dann die folgenden Teilmengen  $A, B, R \subset U$ ,  $A \cup B \cup R = U$  bilden.

$$\begin{aligned} A &= \{x \mid x \in U \wedge x \text{ besitzt Antigen A}\} \\ B &= \{x \mid x \in U \wedge x \text{ besitzt Antigen B}\} \\ R &= \{x \mid x \in U \wedge x \text{ besitzt Antigen Rh}\} \end{aligned}$$

Je nachdem, welche Kombination der drei Antigene  $A, B, Rh$  ein Individuum auf-

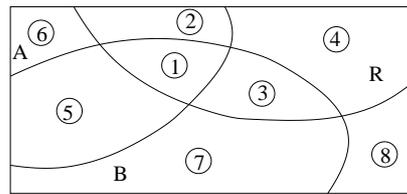


Abbildung 1.18: Die acht Blutgruppen als Mengendiagramm.

weist oder nicht, lassen sich  $2 \cdot 2 \cdot 2 = 8$  verschiedene Blutgruppen unterscheiden:

- ①  $A \cap B \cap R = \{x \mid \text{Blutgruppe } AB, Rh^+\}$
- ②  $A \cap \bar{B} \cap R = \{x \mid \text{Blutgruppe } A, Rh^+\}$
- ③  $\bar{A} \cap B \cap R = \{x \mid \text{Blutgruppe } B, Rh^+\}$
- ④  $\bar{A} \cap \bar{B} \cap R = \{x \mid \text{Blutgruppe } 0, Rh^+\}$
- ⑤  $A \cap B \cap \bar{R} = \{x \mid \text{Blutgruppe } AB, Rh^-\}$
- ⑥  $A \cap \bar{B} \cap \bar{R} = \{x \mid \text{Blutgruppe } A, Rh^-\}$
- ⑦  $\bar{A} \cap B \cap \bar{R} = \{x \mid \text{Blutgruppe } B, Rh^-\}$
- ⑧  $\bar{A} \cap \bar{B} \cap \bar{R} = \{x \mid \text{Blutgruppe } 0, Rh^-\}$

■

### Anzahl der Elemente einer Menge

Hat eine Menge  $A$  endlich viele Elemente, dann kann man von der Anzahl ihrer Elemente sprechen. Diese Anzahl wird mit  $|A|$  bezeichnet.

Mengen mit unendlich vielen Elementen wie die Menge  $\mathbb{N} = \{1, 2, 3, \dots\}$  der natürlichen Zahlen kann man nicht mehr hinsichtlich der Anzahl ihrer Elemente unterscheiden. Möglich ist freilich eine Klassifizierung nach Typen der Unendlichkeit. So ist z.B. die Menge  $\mathbb{N}$  der natürlichen Zahlen von einem anderen Unendlichkeitstypus als die Menge  $\mathbb{R}$  der reellen Zahlen. Die subtilen Untersuchungen dieser Art gehen auf Georg Cantor zurück.

## 1.4 Mengenalgebra

Die Mengenbildungsregeln  $\cap, \cup, \bar{\phantom{x}}$  erfüllen eine Anzahl von „Rechenregeln“, die gewisse Analogien zum gewöhnlichen Rechnen mit Zahlen aufweisen. Die Gültigkeit dieser Regeln kann man sich leicht durch entsprechende Diagramme klarmachen. Eine rechnerische Beweismethode werden wir in diesem Abschnitt ebenfalls kennenlernen. Es gelten die folgenden Regeln für die Verknüpfung von zwei und mehr Mengen zu einer neuen Menge.

**Kommutativgesetze:**

$$A \cap B = B \cap A$$

$$A \cup B = B \cup A$$

**Assoziativgesetze:**

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

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

**Distributivgesetze:**

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

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

**De-Morgansche Regeln:**

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

**Bemerkung:** Die Teilmengen von  $U$  bilden mit diesen Verknüpfungsregeln eine sogenannte *Boolesche Algebra*, benannt nach dem englischen Mathematiker George Boole (1815-1864), einem der Gründerväter der modernen mathematischen Logik. ▲

**Mengen und Aussagen**

Jeder Menge  $A$  lässt sich die Aussage  $p : x \in A$  zuordnen. Die Menge  $A$  besteht dann aus allen Objekten  $x$ , für die die Aussage  $p$  wahr ist, das heißt,  $p$  ist genau dann wahr, wenn  $x \in A$  ist und genau dann falsch, wenn  $x \notin A$  ist. Es besteht auf diese Weise eine Korrespondenz zwischen Mengen und Aussagen. Wenn zu  $A$  die Aussage  $p$  gehört, dann gehört zu  $\overline{A}$  die Aussage  $\neg p$ . Sind den Mengen  $A, B$  die Aussagen  $p, q$  zugeordnet, dann gehört zur Durchschnittsmenge  $A \cap B$  die Aussage  $p \wedge q$  und zur Vereinigungsmenge  $A \cup B$  die Aussage  $p \vee q$ , wie man sich leicht klar macht.

Durch diese Korrespondenz von Mengen und Aussagen lassen sich die Gesetze der Mengenalgebra beweisen und zwar rein rechnerisch, wenn man die Algebraisierung der Aussagenlogik aus Abschnitt 1.1 heranzieht. Wir erinnern an die Regeln

- $\neg p = 1 - p$ ,
- $p \wedge q = pq$ ,
- $p \vee q = p + q - pq$ .

Aufgrund dieser Korrespondenzregeln kann man jeder Mengenidentität eine Identität der entsprechenden Aussagen zuordnen. Der logischen Äquivalenz der Aussagen wiederum entspricht eine algebraische Identität von Formeln .

**Beispiel 1:** Zum Beweis des *Distributivgesetzes*

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

stellen wir zunächst die Korrespondenzen

$$\begin{aligned} A \leftrightarrow p & : x \in A, \\ B \leftrightarrow q & : x \in B, \\ C \leftrightarrow r & : x \in C, \end{aligned}$$

zwischen den beteiligten Mengen und den zugehörigen Aussagen her. Dann ergibt sich mit Hilfe der obigen Übersetzungsregeln für die  $\cap$ - und  $\cup$ -Verknüpfung die Zuordnung:

$$A \cap (B \cup C) \leftrightarrow p \wedge (q \vee r) = p(q + r - qr) = pq + pr - pqr,$$

und entsprechend:

$$(A \cap B) \cup (A \cap C) \leftrightarrow (p \wedge q) \vee (p \wedge r) = pq \vee pr = pq + pr - pqr.$$

Damit ist das Distributivgesetz bewiesen. ■

**Beispiel 2:** Zum Beweis der *De-Morganschen Regel*

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

mit den Korrespondenzen

$$\begin{aligned} A \leftrightarrow p & : x \in A, \\ \bar{A} \leftrightarrow \neg p & : x \notin A, \\ B \leftrightarrow q & : x \in B, \\ \bar{B} \leftrightarrow \neg q & : x \notin B, \end{aligned}$$

erhält man:

$$\begin{aligned} \overline{A \cup B} \leftrightarrow \neg(p \vee q) & = 1 - p \vee q \\ & = 1 - (p + q - pq) \\ & = 1 - p - q + pq. \end{aligned}$$

Andererseits gilt die Zuordnung

$$\begin{aligned} \bar{A} \cap \bar{B} \leftrightarrow \neg p \wedge \neg q & = (1 - p)(1 - q) \\ & = 1 - p - q + pq, \end{aligned}$$

womit die De Morgansche Regel bewiesen ist. ■

## 1.5 Abbildungen

Es seien  $A, B$  Mengen. Unter einer *Abbildung*  $f$  von  $A$  nach  $B$  versteht man eine Vorschrift, die jedem Element  $x$  aus  $A$  genau ein Element  $y = f(x)$  aus  $B$  zuordnet. Symbolisch schreibt man auch

$$f : A \rightarrow B \quad \text{und} \quad x \xrightarrow{f} y.$$

Das Element  $x$  heißt das *Urbild* von  $y$  unter der Abbildung  $f$  und  $y$  ist das *Bild* von  $x$ . Die Menge  $A$ , von der die Abbildung  $f$  ausgeht, heißt die *Urbildmenge* und die Menge  $f(A) \subset B$  ist die *Bildmenge* von  $f$ . Sind  $A$  und  $B$  Zahlenmengen, dann nennt man  $A$  auch den *Definitionsbereich* und  $f(A)$  den *Wertebereich* der Abbildung  $f$ .

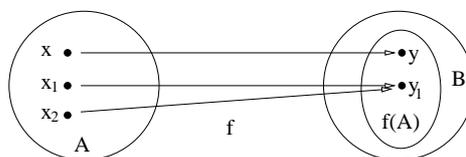


Abbildung 1.19: Schema einer Abbildung.

**Bemerkung:** Zwar ist jedem  $x \in A$  nur ein  $y \in B$  zugeordnet, das heißt, von jedem  $x \in A$  geht nur ein Pfeil in Richtung der Menge  $B$  aus. Es kann aber durchaus sein, dass zwei verschiedenen  $x_1, x_2 \in A$  dasselbe  $y \in B$  zugeordnet ist:  $f(x_1) = f(x_2) = y$ . Nicht jedes Element  $y \in B$  muss aber als Bild eines  $x \in A$  auftreten. (So können zwei Personen einer Gruppe am gleichen Tag Geburtstag haben. Andererseits muss nicht jeder Tag des Jahres Geburtstag eines Gruppenmitgliedes sein.) ▲

### Typen von Abbildungen

Drei wichtige Typen von Abbildungen werden unterschieden:

- *injektive* Abbildungen,
- *surjektive* Abbildungen und
- *bijektive* Abbildungen.

Eine Abbildung  $f : A \rightarrow B$  heißt *injektiv* genau dann, wenn es zu jedem  $y \in B$  höchstens ein  $x \in A$  gibt (das heißt genau ein  $x$  oder kein  $x$ ) mit der Eigenschaft  $f(x) = y$ .

Eine Abbildung  $f : A \rightarrow B$  heißt *surjektiv* genau dann, wenn es zu jedem  $y \in B$  mindestens ein  $x \in A$  gibt, so dass  $f(x) = y$  ist. (Es kann also zu einem Element  $y \in B$  durchaus mehrere Urbilder  $x \in A$  geben.)

Eine Abbildung  $f : A \rightarrow B$  heißt *bijektiv* genau dann, wenn  $f$  zugleich injektiv und surjektiv ist. Das bedeutet, jedem  $y \in B$  entspricht genau ein  $x \in A$  und umgekehrt.

Wenn  $A$  und  $B$  endliche Mengen sind, die bijektiv aufeinander abgebildet werden können, haben sie gleichviele Elemente, das heißt es gilt  $|A| = |B|$ .

Oft hat man die Situation, dass die Anzahl der Elemente einer Menge  $A$  nicht unmittelbar bestimmt werden kann. Kann man dann eine Menge  $B$  konstruieren, die zu  $A$  in einer bijektiven Beziehung steht und die einfacher abgezählt werden kann, dann ist auch das Abzählproblem für  $A$  gelöst.

Sind  $A$  und  $B$  Mengen mit unendlich vielen Elementen, dann bedeutet die Existenz einer bijektiven Abbildung  $f : A \rightarrow B$ , dass  $A$  und  $B$  zum gleichen Typus unendlicher Mengen gehören.

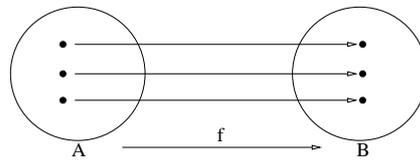


Abbildung 1.20: injektive Abbildung.

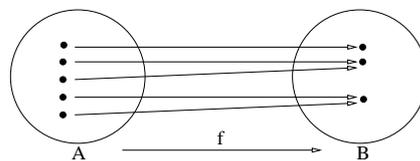


Abbildung 1.21: surjektive Abbildung.

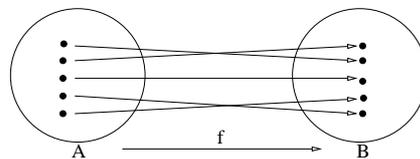


Abbildung 1.22: bijektive Abbildung.

**Bemerkung:** Es gilt *nicht*, dass jede Abbildung genau einem der obigen Typen angehört. ▲

**Beispiel:** Ist  $A = \{a, b, c, \dots, z\}$  die Menge der Buchstaben des Alphabets und  $B = \{1, 2, 3, \dots, 24\}$  die Menge der Zahlen 1 bis 24, dann besteht die Bijektion  $a \rightarrow 1, b \rightarrow 2, c \rightarrow 3, \dots, z \rightarrow 24$  zwischen  $A$  und  $B$ . ■

## Hintereinanderschaltung von Abbildungen

Hat man drei Mengen  $A, B, C$  und zwei Abbildungen  $f, g$  mit  $f : A \rightarrow B$  und  $g : B \rightarrow C$ . Dann kann man die Abbildungen  $f$  und  $g$  zu einer Abbildung  $h = g \circ f, A \rightarrow C$  hintereinanderschalten indem man für  $x \in A$  setzt:  $h(x) = (g \circ f)(x) = g(f(x))$ .

## Die Umkehrabbildung

Besteht zwischen den Mengen  $A, B$  eine Bijektion  $f : A \rightarrow B$ , man schreibt dann auch

$$A \xleftrightarrow{f} B,$$

so lässt sich die Abbildung  $f$  umkehren. Das bedeutet, es gibt eine Abbildung  $f^{-1} : B \rightarrow A$ . Diese ist so erklärt: Zu jedem  $y \in B$  gibt es genau ein  $x \in A$ , so dass  $y = f(x)$  gilt. Dieses  $x$  wird dann das Bild von  $y$  unter der Abbildung  $f^{-1}$ . Das heißt, es gilt  $f^{-1}(y) = x$ .

Außerdem hat man  $f(f^{-1}(y)) = f(x) = y$ . Also gilt  $f \circ f^{-1} = Id$ , wenn  $Id$  die identische Abbildung mit  $Id(y) = y$  ist. Desgleichen gilt  $f^{-1} \circ f = Id$ , denn man hat  $f^{-1}(f(x)) = f^{-1}(y) = x = Id(x)$ .

**Beispiel:** Die Abbildung  $f(x) = 2x^2$  bildet das Intervall  $A = [0, 1]$  bijektiv auf das Intervall  $B = [0, 2]$  ab. Die Umkehrabbildung ist die Abbildung  $f^{-1}$  mit  $f^{-1}(y) = \sqrt{y/2}$ . Gemeint ist hier die positive Wurzel. Offenbar gilt in der Tat

$$\begin{aligned} (f^{-1} \circ f)(x) &= f^{-1}(f(x)) \\ &= f^{-1}(2x^2) \\ &= \sqrt{\frac{2x^2}{2}} \\ &= \sqrt{x^2} \\ &= x = Id(x) \end{aligned}$$

und

$$\begin{aligned}(f \circ f^{-1})(y) &= f(f^{-1}(y)) \\ &= f\left(\sqrt{\frac{y}{2}}\right) \\ &= 2 \cdot \left(\sqrt{\frac{y}{2}}\right)^2 \\ &= 2 \cdot \frac{y}{2} \\ &= y = Id(y).\end{aligned}$$

Die Eigenschaften  $f^{-1} \circ f = f \circ f^{-1} = Id$  einer Umkehrabbildung  $f^{-1}$  sind also erfüllt. ■



# Kapitel 2

## Komplexe Zahlen

Über die Natur der Zahlen wie überhaupt über die Gegenstände der Mathematik gibt es zwei grundsätzlich verschiedene Auffassungen. Die eine sieht in ihnen unabhängig vom menschlichen Geist existierende Wesenheiten, die andere hält sie für reine Schöpfungen menschlicher Geistestätigkeit. So ist in der einen Sichtweise der Mathematiker ein Entdecker, in der anderen ein Erfinder. Diese beiden Sichtweisen finden ihren Ausdruck auch in den Äußerungen der Mathematiker Leopold Kronecker (1823-1891) und Richard Dedekind (1891-1916) über die Natur Zahlen:

Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk (L. Kronecker, Jahresbericht DMV2, S. 19).

Die Zahlen sind freie Schöpfungen des menschlichen Geistes, sie dienen als ein Mittel, um die Verschiedenheit der Dinge leichter und schärfer aufzufassen (R. Dedekind, Was sind und was sollen die Zahlen? Braunschweig 1887, S. 111).

Den Abschluss im Aufbau des Zahlensystems bilden die komplexen Zahlen. Gottfried Wilhelm Leibniz (1646-1716) nannte diese „imaginären Wurzeln“ im Jahre 1702 in der von ihm mitbegründeten ersten deutschen wissenschaftlichen Zeitschrift, den *Acta Eruditorum Lipsiensia*:

...eine feine und wunderbare Zuflucht des göttlichen Geistes, beinahe ein Zwitterwesen zwischen Sein und Nichtsein (inter Ens et non Ens Amphibio) (G.W. Leibniz, Math. Schriften, ed. Gerhard, Bd. 5, S. 357).

Hier ist nicht der Ort, in die philosophische Debatte über den Status der komplexen Zahlen oder der mathematischen Gegenstände überhaupt (diese hält bis heute an) einzutreten oder sie nachzuzeichnen. Es sollen lediglich die Motive angeführt werden, die eine verhältnismäßige breite Behandlung der komplexen Zahlen in einer für Anwender der Mathematik bestimmten Darstellung rechtfertigen. Ein Motiv ist natürlich ein systematisches: Eine Darstellung des Zahlensystems unter Auslassung der komplexen Zahlen wäre unvollständig. Das Hauptmotiv ist aber ein ganz praktisches. Ist einmal

der Umgang mit den komplexen Zahlen vertraut und eingeübt, dann vereinfachen sich Vermittlung und Verständnis bei vielen anderen mathematischen Gebieten ganz erheblich. Genannt seien pars pro toto nur die Punkte: Nullstellen von Polynomen, Partialbruchzerlegung, trigonometrische Funktionen und Differentialgleichungen. Dies ist freilich mehr ein mathematischer Gesichtspunkt. Doch auch in den Anwendungen der Mathematik erweisen sich die komplexen Zahlen als fundamental. Dies zeigt sich besonders eindrucksvoll bei der Quantentheorie, die ohne die komplexen Zahlen nicht formuliert werden könnte. Die komplexen Zahlen scheinen somit an der Grundstruktur der physischen Realität in wesentlicher Form teilzuhaben. Zu Ihrer Rechtfertigung als Lehrstück kann man mit dem österreichischen Physiker Ludwig Boltzmann (1844-1906) sagen: Es gibt nichts praktischeres als eine gute Theorie. (Zur Geschichte der komplexen Zahlen findet man eine ausführliche Darstellung im entsprechenden Artikel von R. Remmert über komplexe Zahlen in dem Buch von Ebbinghaus et al.: Zahlen, 3. Aufl., Berlin 1992.)

## 2.1 Der Aufbau des Zahlensystems

Den Ausgangspunkt beim Aufbau des Zahlensystems bilden die *natürlichen Zahlen*

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

Die Notwendigkeit, einen gegebenen Zahlenbereich zu erweitern ergibt sich immer aus dem Umstand, dass gewissen Gleichungen in diesem Zahlenbereich nicht gelöst werden können. Eine erste Unvollkommenheit der natürlichen Zahlen besteht in der Unmöglichkeit, Gleichungen des Typs  $5 + x = 2$  mit  $x \in \mathbb{N}$  lösen zu können. Erst im Bereich der *ganzen Zahlen*

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

wird diese Gleichung lösbar. Die Unlösbarkeit z.B. der Gleichung  $5x = 2$  in  $\mathbb{Z}$  wiederum zwingt zu einer nochmaligen Erweiterung des Zahlenbereichs  $\mathbb{Z}$  und zur „Erfindung“ der *rationalen Zahlen*

$$\mathbb{Q} = \left\{ \frac{p}{q} \mid p, q \in \mathbb{Z}, q \neq 0 \right\},$$

das ist die Menge der Brüche.

Man kann aber beispielsweise zeigen, dass  $\sqrt{2}$  kein Bruch sein kann (vgl. die folgende Bemerkung). Also muss  $\mathbb{Q}$  ebenfalls erweitert werden und zwar zur Menge  $\mathbb{R}$  der *reellen Zahlen*. Die reellen Zahlen lassen sich als unendliche Dezimalbrüche schreiben:

$$\mathbb{R} = \{a, a_1a_2a_3\dots \mid a \in \mathbb{Z}; a_i \in \{0, 1, 2, \dots, 9\} \text{ für } k = 1, 2, 3, \dots\}.$$

Die Zahlenmenge  $\mathbb{R}$  umfasst also die Zahlenmengen  $\mathbb{Q}, \mathbb{Z}, \mathbb{N}$ :  $\mathbb{R} \supset \mathbb{Q} \supset \mathbb{Z} \supset \mathbb{N}$ .

Es wird im Folgenden vorausgesetzt, dass die Regeln des Rechnens mit den reellen Zahlen bekannt sind. Zu einer nochmaligen – und abschließenden – Erweiterung des Zahlenbereichs  $\mathbb{R}$  zwingt die Unlösbarkeit der Gleichung  $x^2 + 1 = 0$  in  $\mathbb{R}$ . Das heißt: Es gibt keine Zahl  $x \in \mathbb{R}$ , die diese Gleichung erfüllt, denn wie auch immer  $x \in \mathbb{R}$  gewählt wird, so gilt doch stets  $x^2 \geq 0$  und somit  $x^2 + 1 > 0$ .

**Bemerkung:** Dass  $\sqrt{2}$  keine rationale Zahl ist, war bereits den griechischen Mathematikern bekannt (Ein Beweis dazu findet sich in den „Elementen“ des Euklids im Buch X, §115a). Die Irrationalität von  $\sqrt{2}$  besagt geometrisch, dass Seite und Diagonale eines Quadrates kein gemeinsames Maß besitzen, das heißt, *inkommensurabel* sind. (Diese Entdeckung des Irrationalen hat das griechische Denken, das alle Verhältnisse geometrischer Größen für Verhältnisse von ganzen Zahlen hielt, aufs äußerste verstört.)

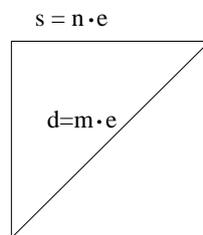


Abbildung 2.1: Gemeinsames Maß von Seite und Diagonale im Quadrat.

Gäbe es nämlich eine Strecke der Länge  $e$ , die in der Seite des Quadrates der Länge  $s$  genau  $n$  mal und genau  $m$  mal in der Diagonalen der Länge  $d$  enthalten ist, dann müßte wegen des Satzes des Pythagoras gelten:

$$2s^2 = d^2,$$

also

$$2(ne)^2 = (me)^2.$$

Dann wäre, nach Division durch  $e^2$ ,

$$2n^2 = m^2$$

und somit

$$\sqrt{2} = \frac{m}{n}.$$

Denkt man sich nun bei dem Bruch  $m/n$  alle gemeinsamen Teiler von Zähler und Nenner herausgekürzt, dann bleiben die Alternativen  $g/u$ ,  $u/g$  und  $u/u$  zu erwägen, wobei  $g$  eine gerade und  $u$  eine ungerade Zahl bezeichnet. Aus der Annahme

$$\sqrt{2} = \frac{g}{u}$$

folgt aber

$$2u^2 = g^2,$$

also gilt

$$u^2 = g \cdot g'$$

mit

$$g' = g/2.$$

Damit haben wir einen Widerspruch erhalten, denn auf der linken Seite dieser Gleichung steht eine ungerade Zahl, denn das Quadrat einer ungeraden Zahl ist wieder ungerade, und auf der rechten eine gerade. In gleicher Weise führt man die Alternativen  $u/g$  und  $u/u$  zu einem Widerspruch. Diese Argumentationsweise ist ein Beispiel für das Beweisverfahren der „reductio ad absurdum“: Folgt aus einer Annahme ein Widerspruch, so muss diese Annahme falsch sein. Also kann  $\sqrt{2}$  keine rationale Zahl sein. ▲

## 2.2 Die Zahl $i$

Die Gleichung  $x^2 + 1 = 0$  oder  $x^2 = -1$  fragt nach einer Zahl  $x$  mit der Eigenschaft, dass sie, mit sich selbst multipliziert,  $-1$  ergibt. Schreibt man letztere Gleichung in der Form  $x^2 \cdot 1 = x \cdot x \cdot 1 = (-1) \cdot 1$ , dann ist sie so zu interpretieren: Gesucht ist eine „Zahl“ oder besser eine „Abbildung“  $x$ , welche die Zahl 1 in die Zahl  $-1$  überführt, wenn man sie zweimal nacheinander auf 1 anwendet. Fasst man die Zahl 1 als

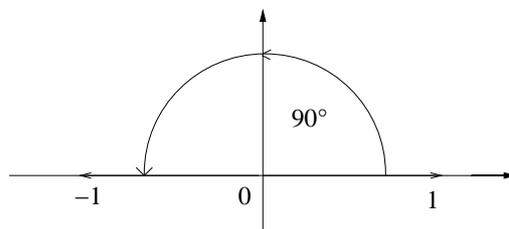


Abbildung 2.2: Die Zahl  $i$ .

eine bei Null angehaftete gerichtete Strecke, einen „Zeiger“ oder „Vektor“ auf, dann wird dieser aber durch eine zweifache Drehung um  $90^\circ$  gegen den Uhrzeigersinn in die Zahl  $-1$  übergeführt. Die Drehung um  $90^\circ$  gegen den Uhrzeigersinn bezeichnen wir mit  $i$ . Bedeutet das Produkt  $i \cdot i = i^2$  die Hintereinanderausführung dieser Drehung, dann gilt offenbar  $i^2 = -1$  oder  $i^2 + 1 = 0$ . Folglich ist  $i$  eine Lösung der Gleichung  $i^2 + 1 = 0$  und  $i$  kann als eine, den Zahlbereich  $\mathbb{R}$  erweiternde Wurzel aus  $-1$  verstanden werden. Der Zeiger  $i \cdot 1$ , den man einfach mit  $i$  bezeichnet, steht dann senkrecht auf den Zeigern 1 und  $-1$ . Als Punkt der Ebene, die man sich von den Zeigern 1 und  $i$  aufgespannt denkt, hat er die Koordinaten  $(0, 1)$ .

Die Zahl  $i$  wird als die *imaginäre Einheit* bezeichnet. Diese Bezeichnung erinnert an die große Unsicherheit im Umgang mit dieser neuartigen Zahl, in der die Mathematiker lange Zeit befangen waren. Man verstand die Zahl  $i = \sqrt{-1}$  zunächst nicht als wirkliche, sondern als bloß vorgestellte, „imaginäre“, Zahl.

Unter Verwendung der imaginären Einheit  $i$  können nun Zahlen  $z$  eines neuen Typs gebildet werden, die *komplexen Zahlen*, das sind die jeweils aus einem reellen und einem imaginären Anteil zusammengesetzten Zahlen

$$\mathbb{C} = \{z | z = x + iy; x, y \in \mathbb{R}\}.$$

Den Umgang mit den komplexen Zahlen leiten zwei Grundsätze:

- Für das Rechnen mit den komplexen Zahlen gelten dieselben Regeln, wie für das Rechnen mit den reellen Zahlen.
- Zu beachten ist zusätzlich die Relation  $i^2 = -1$ .

Wenn man also mit den komplexen Zahlen nach den gleichen Regeln wie mit den reellen Zahlen rechnen kann, dann ergibt sich für die *Addition* zweier komplexer Zahlen  $z = x + iy$  und  $z' = x' + iy'$  die Regel:

$$z + z' = (x + iy) + (x' + iy') = (x + x') + i(y + y').$$

Für das *Produkt*  $z \cdot z'$  zweier komplexer Zahlen erhält man, unter Beachtung von  $i \cdot i \cdot y \cdot y' = i^2 y y' = -y y'$ , in gleicher Weise

$$(*) \quad z \cdot z' = (x + iy) \cdot (x' + iy') = (xx' - yy') + i(xy' + yx').$$

Man kann nun im nachhinein zeigen, dass diese Regeln für Addition und Multiplikation komplexer Zahlen alle Eigenschaften besitzen, die uns vom Rechnen mit reellen Zahlen vertraut sind. So gilt offenkundig  $z + z' = z' + z$  und analog  $z \cdot z' = z' \cdot z$ , das heißt, das *Kommutativgesetz* für Addition und Multiplikation ist erfüllt. Ebenso gilt das *Distributivgesetz*  $z \cdot (z' + z'') = z \cdot z' + z \cdot z''$ , in dem die Verknüpfungen  $+$  und  $\cdot$  verbunden auftreten. Zu jeder komplexen Zahl  $z \neq 0$  ist

$$z^{-1} = \frac{x}{x^2 + y^2} - \frac{iy}{x^2 + y^2}$$

die inverse komplexe Zahl, wie man leicht nachrechnet.

**Bemerkung:** Die Regel für die Addition zweier komplexer Zahlen entspricht völlig der Regel für die Addition von Vektoren der Ebene (des  $\mathbb{R}^2$ ) wenn man die komplexe Zahl  $z = x + iy$  mit dem Vektor  $v = (x, y)$  identifiziert (vgl. Abschnitt 8.2) Die Addition komplexer Zahlen lässt sich daher wie die Vektoraddition anschaulich durch das bekannte Kräfteparallelogramm deuten. Insofern überträgt sich auch die Doppelnatur

der Ebene als Menge von Punkten einerseits und als Vektorraum andererseits auf die komplexen Zahlen.

Anders hingegen verhält es sich bei der Regel (\*) für die Multiplikation zweier komplexer Zahlen. Sie erlaubt keine unmittelbare geometrische Deutung. Diese wird daher in Abschnitt 2.4 nachgetragen. Bei einer (freilich immer unbefriedigenden) rein formalen Einführung der komplexer Zahlen als Zahlenpaare  $z = (x, y)$ ,  $x, y \in \mathbb{R}$ , definiert man auf diese Weise die Multiplikation indem man festsetzt:

$$(x, y) \cdot (x', y') = (xx' - yy', xy' + yx')$$

Auf diese Weise werden die Punkte der Ebene zu Zahlen. Die Rolle der 1 spielt dann der Punkt  $(1, 0)$  und der Punkt  $(0, 1)$  spielt die Rolle der imaginären Einheit  $i$ , denn es gilt  $(0, 1) \cdot (0, 1) = (-1, 0) = -(1, 0)$ . Man kann dann jede komplexe Zahl  $z = (x, y)$  in der Form  $z = x \cdot (1, 0) + y \cdot (0, 1) = x \cdot 1 + y \cdot i = x + iy$  schreiben. Im nachhinein lässt sich dann die Multiplikation  $iz = i(x + iy) = -y + ix$  als Drehung des Vektors  $v = (x, y)$  in den zu  $v$  senkrechten Vektor  $w = (-y, x)$  deuten. ▲

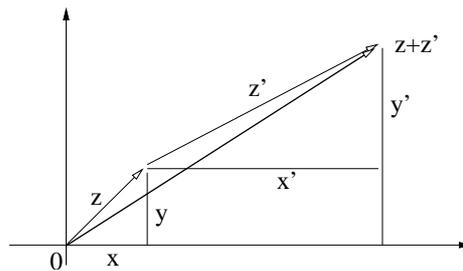


Abbildung 2.3: Geometrische Deutung der Addition komplexer Zahlen.

## 2.3 Die komplexe Ebene

Die komplexe Zahlenebene  $\mathbb{C}$  wird aufgespannt durch die Vektoren (Pfeile, Zeiger)  $1$  und  $i$ . Der Vektor  $i$  geht aus dem Vektor  $1$  durch Drehung um  $90^\circ$  gegen den Uhrzeigersinn hervor. Jedem Punkt  $P = (x, y)$  der Ebene entspricht dann eine komplexe Zahl  $z = x \cdot 1 + y \cdot i$  oder, vereinfacht geschrieben,  $z = x + iy$ .

Die komplexen Zahlen besitzen wie die reellen Zahlen eine Doppelnatur. Als Punkte der Ebene sind sie statisch, indem man sie addieren, subtrahieren, multiplizieren und dividieren kann, wobei jeweils aus zwei Punkten ein neuer Punkt entsteht, zeigen sie eine dynamische Natur.

Die Addition ist dabei die durch das Kräfteparallelogramm erklärte Vektoraddition (s.o.). Die Subtraktion  $z - z'$  ist gleich der Addition  $z + (-z')$  der Vektoren  $z$  und  $-z'$ . Zur geometrischen Deutung von Multiplikation und Division wird auf Abschnitt 2.4. verwiesen.

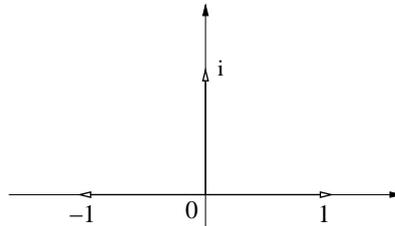


Abbildung 2.4: Die Basisvektoren 1 und  $i$  der komplexen Ebene.

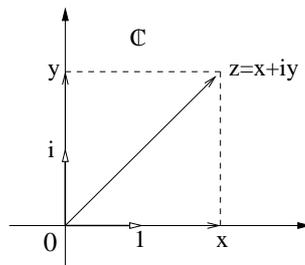


Abbildung 2.5: Die komplexe Ebene.

### Realteil und Imaginärteil

Die reellen Zahlen  $\mathbb{R}$  treten jetzt als *Unterstruktur*  $z = x + 0i$  von  $\mathbb{C}$  auf. Man bezeichnet  $\mathbb{R}$  auch als die *reelle Achse* von  $\mathbb{C}$ . Eine weitere Unterstruktur bilden die Zahlen der Form  $z = 0 + iy$ . Sie liegen auf der *imaginären Achse*, die von der Zahl (dem Vektor)  $i$  erzeugt wird. Die komplexen Zahlen auf der imaginären Achse sind die *rein-imaginären Zahlen*. Die auf der reellen Achse gelegene Komponente  $x \cdot 1 = x$  der komplexen Zahl  $z = x + iy$  heißt *Realteil* von  $z$ , abgekürzt  $x = \operatorname{Re}(z)$ , der reelle Faktor  $y$  der imaginären Komponente  $iy$  heißt der *Imaginärteil* von  $z$ , abgekürzt  $y = \operatorname{Im}(z)$ . Zwei komplexe Zahlen sind gleich, wenn sie in ihrem Realteil und in ihrem Imaginärteil übereinstimmen.

### Der Betrag einer komplexen Zahl

Die Länge des Vektor  $z = x + iy$  ist der *Betrag* der komplexen Zahl  $z$ , geschrieben  $|z|$ . Aufgrund des Satzes des Pythagoras gilt dann

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

### Argument einer komplexen Zahl

Der Winkel  $\varphi$ , den der Vektor  $z$  mit der reellen Achse bildet, ist das *Argument* von  $z$ , geschrieben  $\varphi = \arg(z)$ . Für  $z \neq 0$  ist der Winkel  $\varphi$  als Zahl im Intervall  $0 \leq \varphi < 360$  eindeutig festgelegt, wenn man zur Messung der Größe des Winkels das *Gradmaß*, zugrundelegt. Einzig der Zahl  $z = 0$  lässt sich kein solcher Winkel  $\varphi$  zuordnen. Im Kontext der komplexen Zahlen verwendet man zur Winkelmessung das *Bogenmaß*. Dem Winkel  $\varphi$  entspricht nämlich auf dem Einheitskreis eindeutig ein Bogen der Länge  $t$  mit  $0 \leq t < 2\pi$ . Die Zahl  $t$  ist dann das Bogenmaß des Winkels  $\varphi$ .

### Die konjugiert komplexe Zahl

Spiegelt man die komplexe Zahl  $z = x + iy$  an der reellen Achse, dann erhält man die Zahl  $\bar{z} = x - iy$ . Die Zahl  $\bar{z}$  (gesprochen *z-quer*) heißt die zu  $z$  *konjugierte* komplexe Zahl.

Die Spiegelung an der reellen Achse ist eine sehr nützliche Abbildung der komplexen Ebene in sich mit schönen Eigenschaften. Offenbar gilt  $\overline{\bar{z}} = z$ , zwei Spiegelungen nacheinander heben sich auf. Einfache Rechnungen zeigen die Vertauschbarkeit der Reihenfolge von Addition und Spiegelung sowie von Multiplikation und Division mit

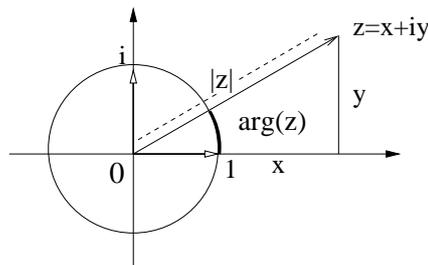


Abbildung 2.6: Betrag und Argument einer komplexen Zahl.

dieser Spiegelung. Es gilt also

$$\overline{z + z'} = \bar{z} + \bar{z}'$$

sowie

$$\overline{z \cdot z'} = \bar{z} \cdot \bar{z}'$$

und

$$\frac{\bar{z}}{\bar{z}'} = \frac{z}{z'}.$$

Die Größen  $|z|$ ,  $\operatorname{Re}(z)$  und  $\operatorname{Im}(z)$  lassen sich durch  $z$  und  $\bar{z}$  auf einfache Weise ausdrücken, so ist

$$\begin{aligned} (*) \quad z \cdot \bar{z} &= (x + iy)(x - iy) \\ &= x^2 + y^2 = |z|^2, \end{aligned}$$

also gilt

$$|z| = \sqrt{z \cdot \bar{z}}.$$

Außerdem bestätigt man leicht die Beziehungen

$$\operatorname{Re}(z) = \frac{1}{2}(z + \bar{z})$$

und

$$\operatorname{Im}(z) = \frac{1}{2i}(z - \bar{z}).$$

Diese beiden Beziehungen, zusammen mit den obigen Regeln erlauben die Berechnung des Realteils bzw. des Imaginärteils auch verwickelterer Ausdrücke in komplexen Zahlen.

Der Beziehung (\*) entnimmt man unmittelbar eine Darstellung der zu  $z$  inversen komplexen Zahl  $1/z = z^{-1}$ . Offenbar ist

$$\frac{1}{z} = \frac{\bar{z}}{|z|^2} = \frac{x - iy}{x^2 + y^2}$$

für  $z \neq 0$ .

## 2.4 Die Euler-Formel

Während die Addition zweier komplexer Zahlen als Vektoraddition (Kräfteparallelogramm) geometrisch unmittelbar interpretierbar ist, gilt dies nicht in gleicher Weise für die Multiplikation. Was bei der Multiplikation geometrisch geschieht, wird erst wirklich durchsichtig, wenn die *Exponentialfunktion* zur Verfügung steht. Diese wird aber eigentlich erst in Abschnitt 5.3 behandelt. Zugunsten eines vertieften Verständnisses der komplexen Zahlen und auch wegen der Eleganz der Darstellung erlauben wir uns an dieser Stelle eine Abweichung vom systematischen Aufbau und bringen im Vorgriff auf die spätere ausführliche Darstellung bzw. im Vertrauen auf eventuelle Vorkenntnisse bereits hier die *Exponentialfunktion* ins Spiel.

Die Exponentialfunktion  $e^x = \exp(x)$  wird definiert durch die unendliche Reihe

$$e^x = 1 + x + \frac{x^2}{1 \cdot 2} + \frac{x^3}{1 \cdot 2 \cdot 3} + \frac{x^4}{1 \cdot 2 \cdot 3 \cdot 4} + \dots$$

Diese Reihe ist für alle  $x \in \mathbb{R}$ , aber ebenso für alle  $z \in \mathbb{C}$ , erklärt. Die schnell wachsenden Zahlen  $1 \cdot 2 \cdot 3 \dots (k-1) \cdot k = k!$  (gesprochen „k-Fakultät“) sind die Fakultäten. Unter Verwendung des Summenzeichens  $\sum$  kann man dann abkürzend schreiben:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Die Exponentialfunktion genügt der *Funktionalgleichung*

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

für beliebige (also auch komplexe) Argumente  $x, y$ .

Die komplexen Zahlen  $u$  vom Betrag eins sind die Punkte auf dem Einheitskreis  $E = \{z \in \mathbb{C} : |z| = 1\}$  der komplexen Ebene. Diese Zahlen werden auch als *unimodulare*

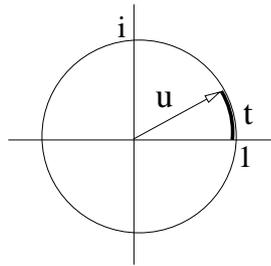


Abbildung 2.7: Unimodulare Zahlen.

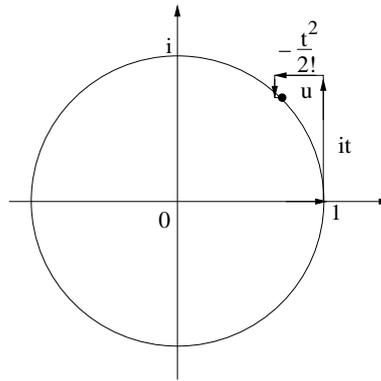
Zahlen bezeichnet. Die Länge  $\arg(u)$  des Kreisbogens zwischen den Zahlen 1 und  $u$  sei  $t$ . Dann lässt sich die Zahl  $u$  in der Form

$$u = e^{it}$$

darstellen. Diese Beziehung werden wir später im Rahmen einer Grenzwertbetrachtung beweisen (siehe Abschnitt 5.4). Hier müssen wir uns darauf beschränken, sie geometrisch plausibel zu machen. Die Beziehung  $u = \exp(it)$  besagt:

$$\begin{aligned} u &= 1 + it + \frac{(it)^2}{2!} + \frac{(it)^3}{3!} + \frac{(it)^4}{4!} + \frac{(it)^5}{5!} + \dots \\ &= 1 + it - \frac{t^2}{2!} - \frac{it^3}{3!} + \frac{t^4}{4!} + \frac{it^5}{5!} - \dots \end{aligned}$$

Geometrisch lässt sich diese Reihe als Summe schnell kleiner werdender komplexer Zahlen (Vektoren) auffassen, die sich dem Punkt  $u$  spiralförmig mit beliebiger Genauigkeit annähert.

Abbildung 2.8: Geometrische Deutung der Beziehung  $u = e^{it}$ .

**Bemerkung:** Dass jede komplexe Zahl der Form  $u = e^{it}$  auf dem Einheitskreis liegt, folgt unmittelbar aus der Funktionalgleichung der Exponentialfunktion. Es ist nämlich

$$|u|^2 = u\bar{u} = e^{it} \cdot e^{-it} = e^0 = 1$$

für  $t \in \mathbb{R}$ . Diese Überlegung lässt freilich die Bedeutung des Parameters  $t$  als Länge des Bogens auf dem Einheitskreis zwischen den Punkten 1 und  $e^{it}$  noch nicht erkennen. ▲

Für gerade Potenzen der imaginären Einheit  $i$  erhält man  $i^{2k} = (-1)^k$  und  $i^{2k+1} = (-1)^k i$  für ungerade Potenzen. Man kann daher die unendliche Reihe, welche die Zahl  $e^{it}$  darstellt, wie folgt in einen reellen und einen rein imaginären Anteil aufspalten:

$$\begin{aligned} e^{it} &= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{(2k)!} + i \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k+1}}{(2k+1)!}. \end{aligned}$$

Die Reihe mit den geraden Exponenten stellt die Cosinusfunktion dar, die Reihe mit den ungeraden Exponenten die Sinusfunktion. Das heißt, es ist

$$\cos(t) = \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{(2k)!}$$

und

$$\sin(t) = \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k+1}}{(2k+1)!}, \quad t \in \mathbb{R}.$$

Diese Gleichungen können auch als Definition der Funktionen  $\sin(t)$  und  $\cos(t)$  aufgefasst werden. Es besteht also die Beziehung

$$e^{it} = \cos(t) + i \sin(t)$$

zwischen der Exponentialfunktion mit rein-imaginärem Exponenten  $it$  und den trigonometrischen Funktionen  $\sin(t)$  und  $\cos(t)$ . Sie geht auf Leonhard Euler (1707-1783) zurück, den bedeutendsten Mathematiker des 18. Jahrhunderts. Sie wird auch als *Euler-Formel* bezeichnet. Da der halbe Kreisumfang gleich  $\pi$  ist, besteht der be-

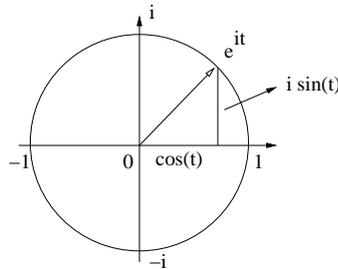


Abbildung 2.9: Die Euler-Formel.

merkenswerte Zusammenhang  $e^{i\pi} = -1$  oder

$$e^{i\pi} + 1 = 0.$$

Diese Formel wird auch als Eulersche Zauberformel bezeichnet. Sie stellt einen Zusammenhang her zwischen der Eulerschen Zahl  $e = 2,718\dots$ , der Kreiszahl  $\pi = 3,141\dots$ , der imaginären Einheit  $i$  und den Zahlen 0 und 1.

### Die Polarform einer komplexen Zahl

Jede komplexe Zahl  $z$  kann in der Form  $z = r \cdot u$  mit  $r = |z|$  und  $u = \frac{z}{|z|} = e^{it}$  geschrieben werden. Damit erhält man mit Hilfe der Funktionalgleichung (\*) der Exponentialfunktion für das Produkt zweier komplexer Zahlen  $z = r \cdot u = r \cdot e^{it}$  und  $z' = r' \cdot u' = r' \cdot e^{it'}$ :

$$\begin{aligned} z \cdot z' &= r \cdot r' e^{it} e^{it'} \\ &= r \cdot r' e^{i(t+t')} \end{aligned}$$

So wird anschaulich, was bei der Bildung des Produktes zweier komplexer Zahlen geschieht: Die Beträge der beteiligten Zahlen werden miteinander multipliziert und die zugehörigen Winkel werden addiert.

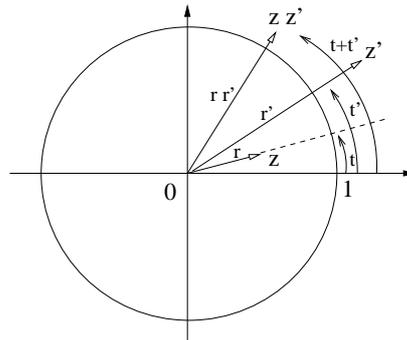


Abbildung 2.10: Multiplikation komplexer Zahlen.

Notiert man die komplexen Zahlen  $z, z'$  mit Hilfe ihrer *Polarkoordinaten*  $(r, t)$  und  $(r', t')$ , dann schreibt sich die Multiplikationsregel in der suggestiven Form

$$(r, t) \cdot (r', t') = (r \cdot r', t + t').$$

Als Spezialfall ergibt sich: Die Multiplikation der komplexen Zahl  $z$  mit der unimodularen Zahl  $e^{it}$  bedeutet, den Vektor  $z$  um den Winkel  $t$  gegen den Uhrzeigersinn zu drehen.

**Bemerkung:** Das Verfahren, die Lage eines Punktes in der Ebene mit Hilfe von Polarkoordinaten zu beschreiben, wird auch von den Bienen beherrscht. Hat eine Biene einen Futterplatz ( $F$ ) entdeckt, so teilt sie dessen Lage ihren Artgenossen mit, indem sie im Bienenstock ( $B$ ) einen Schwänzeltanz aufführt. Dieser codiert Richtung und Entfernung zum Futterplatz, bezogen auf den Bienenstock als Nullpunkt. Die Richtung wird als Abweichung von der Richtung angegeben, in der die Sonne ( $S$ ) steht. Der Winkel  $\varphi$  zwischen den Strecken  $BS$  und  $BF$  wird durch die Ausrichtung der Biene beim Schwänzeltanz signalisiert, die Entfernung  $r$  zwischen den Punkten  $B$  und  $F$  durch die Intensität der Bewegung. ▲

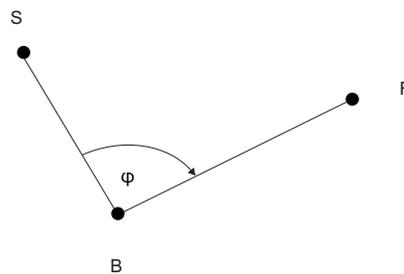


Abbildung 2.11: Lage des Futterplatzes.

## 2.5 Die Gruppe der $n$ -ten Einheitswurzeln

Es gilt  $i^2 = -1$ ,  $i^3 = -i$  und  $i^4 = 1$ . Die Potenzen von  $i = e^{i\frac{2\pi}{4}}$ , das sind die vier Elemente der Menge  $E_4 = \{1, i, -1, -i\}$ , haben zwei bemerkenswerte Eigenschaften: Multipliziert man zwei Zahlen aus  $E_4$  miteinander, so erhält man wieder eine Zahl aus  $E_4$ . Zu jeder Zahl aus  $E_4$  existiert in  $E_4$  eine inverse Zahl. So ist  $\frac{1}{i} = i^{-1} = -i$  das Inverse zu  $i$ . Eine Zahlenmenge mit dieser Struktur nennt man eine *Gruppe*. Die Anzahl ihrer Elemente ist die *Ordnung* der Gruppe. Für jedes  $z \in E_4$  gilt  $z^4 = 1$ . Also ist  $z$  eine 4-te Wurzel aus 1:  $E_4$  ist die Gruppe der 4-ten *Einheitswurzeln*.

Die Elemente von  $E_4$  bilden in der komplexen Ebene  $\mathbb{C}$  die Ecken eines dem Einheitskreis einbeschriebenen Quadrates. Die Potenzen der oberen Ecke  $i$  des Quadrates

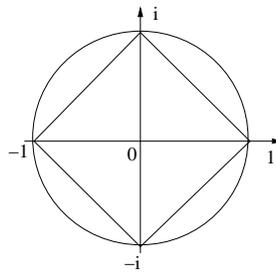


Abbildung 2.12: Die 4-ten Einheitswurzeln.

erzeugen die ganze Gruppe  $E_4$ . Man nennt eine Einheitswurzel mit dieser Eigenschaft eine *primitive* Einheitswurzel. Offenbar ist  $-1$  keine primitive 4-te Einheitswurzel.

Als weiteres Beispiel betrachten wir die 6-ten Einheitswurzeln. Diese werden erzeugt von der Einheitswurzel  $s = e^{i\frac{2\pi}{6}}$ . Ihre Potenzen  $s^k = e^{i\frac{2\pi k}{6}}$ ,  $k = 1, 2, 3, 4, 5, 6$ , bilden die Ecken eines regelmäßigen Sechsecks. Für alle Zahlen  $s^k$  gilt  $(s^k)^6 = s^{6k} = e^{i2\pi k} = 1$ . Die Zahlen  $0, 1, s$  sind die Ecken eines gleichseitigen Dreiecks mit der Sei-

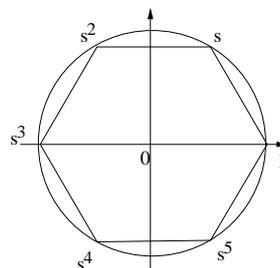


Abbildung 2.13: Die 6-ten Einheitswurzeln.

tenlänge eins. Seine Höhe ist  $\frac{\sqrt{3}}{2}$ . Also gilt

$$s = \frac{1}{2} + \frac{i\sqrt{3}}{2} = \cos\left(\frac{2\pi}{6}\right) + i \sin\left(\frac{2\pi}{6}\right).$$

Die 6-ten Einheitswurzeln, das sind die komplexen Zahlen

$$\begin{aligned} s^1 &= \frac{1}{2} + \frac{i\sqrt{3}}{2}, \\ s^2 &= -\frac{1}{2} + \frac{i\sqrt{3}}{2}, \\ s^3 &= -1, \\ s^4 &= -\frac{1}{2} - \frac{i\sqrt{3}}{2}, \\ s^5 &= \frac{1}{2} - \frac{i\sqrt{3}}{2}, \\ s^6 &= 1. \end{aligned}$$

Zusammen bilden sie die Gruppe  $E_6$ .

Das regelmäßige Sechseck weist zahlreiche Symmetrien auf. Eine  $60^\circ$ -Drehung um den Mittelpunkt lässt es in sich selbst übergehen, ebenso die Spiegelung an der reellen oder imaginären Achse. Dazu kommen sämtliche Kombinationen dieser Symmetrien. Jede dieser Symmetrien lässt sich auffassen als Abbildung  $f : \mathbb{C} \rightarrow \mathbb{C}$  der komplexen Ebene in sich. Die Drehung um  $60^\circ$  realisiert die Abbildung  $z \mapsto sz$ , die Drehungen um  $120^\circ$ ,  $180^\circ$ ,  $240^\circ$  und  $300^\circ$  werden durch Multiplikation mit den Potenzen  $s^2$ ,  $s^3$ ,  $s^4$  und  $s^5$  der primitiven 6-ten Einheitswurzel  $s$  induziert. Der Spiegelung an der reellen Achse entspricht die Abbildung  $z \mapsto \bar{z}$  und der Spiegelung an der imaginären Achse die Abbildung  $z \mapsto -\bar{z}$ . Die Symmetrien, die eine Figur  $F$  zulässt, bestehen aus der

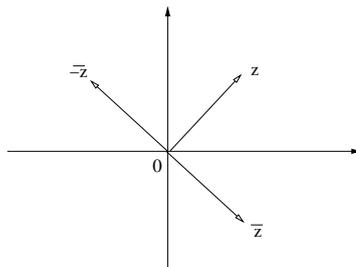


Abbildung 2.14: Spiegelung an der reellen und der imaginären Achse.

Gesamtheit derjenigen Abbildungen  $f$ , die eine Figur in sich überführen, für die also gilt  $f(F) = F$ . Diese Abbildungen bilden eine Gruppe, die *Symmetriegruppe* der Figur. In der Mineralogie dienen Symmetriegruppen zur Klassifizierung von Kristalltypen.

## 2.6 Der Fundamentalsatz der Algebra

Das Polynom  $p(x) = x^2 + 1$  besitzt keine reelle Nullstelle. Diese Unvollkommenheit der reellen Zahlen wird durch Einführung der komplexen Zahlen behoben. Im Zahlenbereich  $\mathbb{C}$  hat  $p(x)$  die beiden Nullstellen  $i$  und  $-i$ , das heißt es gilt  $p(i) = p(-i) = 0$ . Damit ist für das spezielle Polynom  $p(x) = x^2 + 1$  der genannte Mangel zwar behoben, es kann aber zunächst nicht ausgeschlossen werden, dass es andere Polynome gibt, die auch im Zahlenbereich  $\mathbb{C}$  keine Nullstellen besitzen. Dies ist aber nicht der Fall, und das ist der Inhalt des folgenden Satzes:

**Fundamentalsatz der Algebra:** *Jedes Polynom  $p(z) = a_0 + a_1z + \dots + a_nz^n$  mit Koeffizienten  $a_0, a_1, \dots, a_n \in \mathbb{C}$ , das nicht gleich einer Konstanten  $c \neq 0$  ist, besitzt in  $\mathbb{C}$  eine Nullstelle. Das heißt, es gibt eine Zahl  $z_1 \in \mathbb{C}$ , so dass  $p(z_1) = 0$  ist.*

Einen strengen Beweis dieses Sachverhaltes können wir hier nicht führen. Vielmehr muss ein Plausibilitätsargument genügen, die Richtigkeit des Satzes zu veranschaulichen.

Hierzu fassen wir  $p(z)$  als eine Abbildung zwischen einer komplexen  $z$ -Ebene und einer komplexen  $w$ -Ebene auf: Dem Punkt  $z$  in der  $z$ -Ebene wird dadurch der Punkt  $w = p(z)$  in der  $w$ -Ebene zugeordnet. Dem Punkt  $z = 0$  entspricht so der Punkt  $w = p(0) = a_0$ . Eine kleine Kreisscheibe  $K$  um  $z = 0$  geht auf diese Weise in ein kleines Gebiet  $p(K)$  um den Punkt  $w = a_0$  über, dessen Rand im allgemeinen nicht mehr kreisförmig sein wird. Vergrößert man die Kreisscheibe  $K$ , dann vergrößert sich

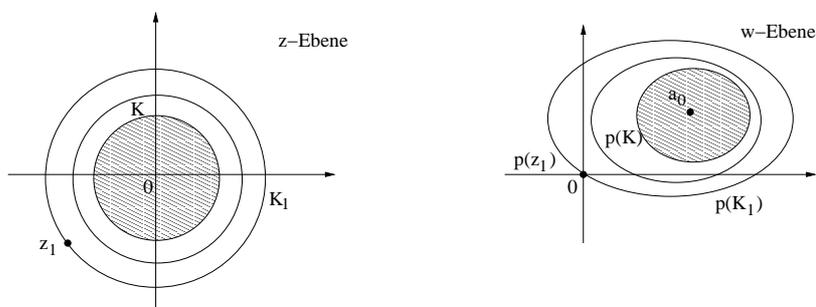


Abbildung 2.15: Abbildung  $p(z)$  zwischen  $z$ -Ebene und  $w$ -Ebene.

auch der „Tintenfleck“  $p(K)$ , denn mit  $z$  werden auch die Potenzen von  $z$  immer „größer“. Unter den stetig wachsenden konzentrischen Kreisscheiben um  $z = 0$  wird es schließlich eine Kreisscheibe  $K_1$  geben mit der Eigenschaft, dass der Punkt  $w = 0$  auf dem Rand des Gebietes  $p(K_1)$  liegt. Der Punkt  $w = 0$  ist dann der Bildpunkt eines Punktes  $z_1$  auf dem Rand der Kreisscheibe  $K_1$ . Das heißt, es gilt  $p(z_1) = 0$ . Das Polynom  $p(z)$  besitzt also eine Nullstelle in der komplexen Ebene.

## Zerlegung in Linearfaktoren

Wenn  $z_1$  eine Nullstelle des Polynoms

$$p(z) = a_0 + a_1z^2 + a_2z + \dots + a_nz^n$$

mit  $a_n \neq 0$  ist, dann gilt

$$p(z_1) = 0 = a_0 + a_1z_1 + a_2z_1^2 + \dots + a_nz_1^n,$$

also ist  $p(z) - p(z_1) = p(z)$ . Andererseits gilt

$$(*) \quad p(z) - p(z_1) = a_1(z - z_1) + a_2(z^2 - z_1^2) + \dots + a_n(z^n - z_1^n).$$

Nun ist aber

$$z^2 - z_1^2 = (z - z_1)(z + z_1)$$

und

$$z^3 - z_1^3 = (z - z_1)(z^2 + zz_1 + z_1^2)$$

wie man leicht durch Ausrechnen der rechten Seite bestätigt. Allgemein gilt die auch anderweitig nützliche Formel

$$z^k - z_1^k = (z - z_1)(z^{k-1} + z^{k-2}z_1 + z^{k-3}z_1^2 + \dots + zz_1^{k-2} + z_1^{k-1}),$$

deren Richtigkeit man wiederum durch Ausmultiplizieren der beiden Klammern bestätigt.

Die Differenz  $(z - z_1)$  kann also auf der rechten Seite von  $(*)$  ausgeklammert werden. Man erhält so die Zerlegung

$$p(z) = (z - z_1)p_1(z)$$

mit einem Polynom  $p_1(z)$ . Da  $a_n \neq 0$  ist, ist  $z^{n-1}$  die höchste bei  $p_1(z)$  auftretende Potenz von  $z$ . Das Polynom  $p_1(z)$  hat aufgrund derselben Überlegungen wie bei  $p(z)$  eine Nullstelle  $z_2$ , so dass  $p_1(z)$  in der Form  $p_1(z) = (z - z_2)q_2(z)$  geschrieben werden kann mit einem Polynom  $q_2(z)$ . Die höchsten  $z$ -Potenz bei  $q_2(z)$  ist  $z^{n-2}$ . Also hat man für  $p(z)$  die Zerlegung

$$p(z) = (z - z_1)(z - z_2)q_2(z).$$

Auf diese Weise kann man sukzessive Polynome abspalten, deren höchste  $z$ -Potenz, das ist ihr *Grad*, sich jeweils um eins vermindert. Die Gesamtzahl dieser Polynome  $q_1(z), q_2(z), q_3(z), \dots$  ist gleich  $n$ . Das letzte Polynom  $q_n(z)$  dieser Reihe hat die Form  $a_nz - c = a_n(z - \frac{c}{a_n})$  mit der Nullstelle  $z_n = \frac{c}{a_n}$ . Damit haben wir das Ergebnis:

**Satz:** Jedes Polynom  $p(z) = a_0 + a_1z + a_2z^2 + \dots + a_nz^n$  mit  $a_n \neq 0$  kann als Produkt seiner Linearfaktoren in der Form  $p(z) = a_n(z - z_1)(z - z_2) \cdots (z - z_n)$  mit den Nullstellen  $z_1, z_2, \dots, z_n$  geschrieben werden.

Die Zahlen  $z_1, z_2, \dots, z_n$ , die *Nullstellen* des Polynoms  $p(z)$  müssen nicht sämtlich voneinander verschieden sein.

**Bemerkung:** Selten werden mathematische Gegenstände zu Themen der Literatur. (Eine Fundgrube hierzu ist das Buch von Kurt Radbruch: *Mathematische Spuren der Literatur*, Darmstadt 1997.) Eine Ausnahme in dieser Hinsicht bildet der Roman des österreichischen Schriftstellers Robert Musil (1880-1942) mit dem Titel „Die Verwirrungen des Zöglings Törleß“ aus dem Jahre 1906. Er schildert das Schülerleben in einem altösterreichischen Internat und so auch Verwirrungen, deren Ursache der Mathematikunterricht bildet. Die Schwierigkeiten, die zum Beispiel die imaginären und irrationalen Zahlen dem Verständnis der Schüler, damaligen wie wohl auch heutigen, entgegensetzen, werden in einer Passage einfühlsam geschildert, die wir im Wortlaut wiedergeben. Auch denjenigen zum Trost, denen ähnliche Irritationen nicht unbekannt sind.

(...) Während des Mathematikunterrichtes war Törleß plötzlich ein Einfall gekommen.

Er hatte schon während der letzten Tage den Unterricht in der Schule mit besonderem Interesse verfolgt gehabt, denn er dachte sich: „Wenn dies wirklich die Vorbereitung für das Leben sein soll, wie sie sagen, so muss sich doch auch etwas von dem angedeutet finden, was ich suche.“

Gerade an die Mathematik hatte er dabei gedacht; noch von jenen Gedanken an das Unendliche her.

Und richtig war es ihm mitten im Unterrichte heiß in den Kopf geschossen. Gleich nach Beendigung der Stunde setzte er sich zu Beineberg als dem einzigen, mit dem er über etwas Derartiges sprechen konnte. „Du, hast du das vorhin ganz verstanden?“

„Was?“

„Die Geschichte mit den imaginären Zahlen?“

„Ja. Das ist doch gar nicht so schwer. Man muss nur festhalten, dass die Quadratwurzel aus negativ Eins die Rechnungseinheit ist.“

„Das ist es aber gerade. Die gibt es doch gar nicht. Jede Zahl, ob sie nun positiv ist oder negativ, gibt im Quadrat erhoben etwas Positives. Es kann daher gar keine wirkliche Zahl geben, welche die Quadratwurzel von etwas Negativem wäre.“

„Ganz recht; aber warum sollte man nicht trotzdem versuchen, auch bei einer negativen Zahl die Operation des Quadratwurzelziehens anzuwenden? Natürlich kann dies dann keinen wirklichen Wert ergeben, und man nennt doch auch deswegen das Resultat nur ein imaginäres. Es ist so, wie wenn man sagen würde: hier saß sonst immer jemand, stellen wir ihm also auch heute einen Stuhl hin; und selbst, wenn er inzwischen gestorben wäre, so tun wir doch, als ob er käme.“

Wie kann man aber, wenn man bestimmt, ganz mathematisch bestimmt

weiß, dass es unmöglich ist?“

„So tut man eben trotzdem, als ob dem nicht so wäre. Es wird wohl irgendeinen Erfolg haben. Was ist es denn schließlich anderes mit den irrationalen Zahlen? Eine Division, die nie zu Ende kommt, ein Bruch dessen Wert nie und nie und nie herauskommt, wenn Du auch noch so lange rechnet? Und was kannst du hier darunter denken, dass sich parallele Linien im Unendlichen schneiden sollen? Ich glaube, wenn man allzu gewissenhaft wäre, so gäbe es keine Mathematik.“(...)

(R. Musil, Gesammelte Werke Bd. 6, S. 73)





# Kapitel 3

## Kombinatorik

Typische Fragestellungen der Kombinatorik sind:

- Wieviele Wörter bestimmter Länge kann man aus einem vorgegebenen Alphabet bilden?
- Wie oft muß ein Stammtisch von 10 Personen tagen, bis alle möglichen Sitzordnungen gerade einmal eingenommen wurden?
- Wieviele verschiedene Wörter lassen sich aus einem gegebenen Wort durch Umstellen der Buchstaben bilden?
- Wieviele Teilmengen mit vorgegebener Anzahl von Elementen besitzt eine endliche Menge?

In abstrakter Sichtweise handelt es sich bei diesen Fragen immer darum, die Anzahl der Elemente einer endlichen Menge zu bestimmen, die der jeweiligen Fragestellung entsprechend gebildet wird.

### 3.1 Anzahl der Wörter aus einem Alphabet

Im Folgenden geht es um die Anzahl möglicher Anordnungen von Elementen einer Menge zu „Wörtern“ gegebener Länge.

**1. Grundaufgabe:** Gegeben sei die Menge (das Alphabet)  $N$  mit  $n$  Elementen (Buchstaben). Wieviele Wörter der Länge  $k$  lassen sich mit Buchstaben aus diesem Alphabet bilden?

Jedes Wort mit  $k$  Buchstaben aus dem Alphabet  $N$  ist ein Element des cartesischen Produktes  $N \times N \times \dots \times N = N^k$  mit  $k$  Faktoren und umgekehrt. Die Anzahl der Elemente der Menge  $N^k$  ist  $n^k$ . Daher lassen sich mit einem Alphabet aus  $n$  Buchstaben genau  $n^k$  Wörter mit  $k$  Buchstaben bilden.

Die Wörter der Länge  $k$ , die man mit dem Alphabet  $N$  bilden kann, zerfallen in zwei Gruppen: diejenigen Wörter, bei denen sich Buchstaben wiederholen und die anderen, bei denen alle Buchstaben verschieden sind.

**2. Grundaufgabe:** Wieviele Wörter der Länge  $k$  lassen sich mit den Buchstaben des  $N$  Alphabets bilden, wenn alle Buchstaben eines Wortes *verschieden* sein sollen?

Besteht das Alphabet aus den drei Buchstaben  $a, b, c$ , dann kann man die folgenden zweibuchstabigen Wörter

$$\begin{array}{ccc} ab & ba & ca \\ ac & bc & cb \end{array}$$

bilden, bei denen die Buchstaben sämtlich voneinander verschieden sind. Ihre Anzahl ist  $3 \cdot 2 = 6$ , da man für den ersten Buchstaben 3 Wahlmöglichkeiten hat, und damit für den zweiten Buchstaben jeweils noch 2 Möglichkeiten verbleiben. Hat das Alphabet die vier Buchstaben  $a, b, c, d$ , dann gibt es entsprechend  $4 \cdot 3 \cdot 2 = 24$  Möglichkeiten dreibuchstabige Wörter ohne Wiederholung von Buchstaben zu bilden.

Allgemein ist die Anzahl der Wörter der Länge  $k$  aus Buchstaben eines Alphabets  $N$  der Länge  $n$  ohne Wiederholung von Buchstaben gleich

$$n(n-1)(n-2)\dots(n-k+1).$$

**Beispiel:** Die Erbinformation der Lebewesen, der genetische Code, ist im Kettenmolekül der Desoxyribonucleinsäure (DNA) verschlüsselt. Dessen Bausteine sind die vier Basen Adenin (A), Guanin (G), Cytosin (C) und Thymin (T). Das Alphabet, mit dem die genetischen Informationen geschrieben wird, besteht also aus den vier Buchstaben A,G,C,T. Die Anzahl der Wörter der Länge  $n$ , die man aus diesem Alphabet hinschreiben kann, ist daher nach Grundaufgabe 1 gleich  $4^n$ . Wörter ohne Buchstabenwiederholung gibt es nur bis zur Länge 4. Deren Anzahl ist  $4 \cdot 3 \cdot 2 = 24$ . ■

## 3.2 Anzahl der Permutationen eines Wortes

Durch Umordnung (*Permutation*) der Buchstaben eines gegebenen Wortes entsteht im allgemeinen ein neues Wort. Wieviele Wörter entstehen insgesamt durch solche Permutationen der Buchstaben?

**3. Grundaufgabe:** Gegeben sei ein Wort mit  $n$  Buchstaben, die sämtlich voneinander verschieden sind. Wie groß ist die Anzahl der Wörter, die man durch Umordnung (Permutation) des gegebenen Wortes bilden kann?

Aus dem Wort  $abc$  erhält man – einschließlich des ursprünglichen Wortes – durch Umstellen der Buchstaben die 6 Wörter

$$\begin{array}{l} abc \quad bac \quad cab \\ acb \quad bca \quad cba. \end{array}$$

Diese zerfallen, entsprechend den alternativen Anfangsbuchstaben, in 3 Gruppen zu je 2 Wörtern, das sind zusammen  $3 \cdot 2 = 6$  Wörter.

Entsprechend zerfallen die Wörter, die man durch Umstellung der Buchstaben des 4-buchstabigen Wortes  $abcd$  erhält, in 4 Gruppen, geordnet nach den 4 möglichen Anfangsbuchstaben. Die Wörter einer jeden Gruppe entstehen durch Umordnung der verbleibenden 3 Buchstaben, wozu es, nach obigem, wiederum 6 Möglichkeiten gibt. Also hat man  $4 \cdot 6 = 4 \cdot 3 \cdot 2 = 24$  Möglichkeiten, aus dem Wort  $abcd$  durch Umstellung der Buchstaben neue Wörter – inklusive das Ausgangswort – zu bilden.

Ist  $P(n)$  die Anzahl der Permutationen eines Wortes mit  $n$  verschiedenen Buchstaben, dann gilt offenbar  $P(n) = nP(n-1)$ . Man kann nämlich diese Permutationen wieder nach den  $n$  Anfangsbuchstaben in Gruppen einteilen. Die Wörter jeder Gruppe entstehen dann durch Permutationen der noch frei permutierbaren  $n-1$  übrigen Buchstaben. Also erhält man

$$\begin{aligned} P(n) &= n \cdot P(n-1) = n \cdot (n-1) \cdot P(n-2) = \dots \\ &= n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot P(1) \end{aligned}$$

und mit  $P(1) = 1$  schließlich

$$P(n) = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n.$$

Die Anzahl  $P(n)$  der Permutationen eines Wortes mit  $n$  verschiedenen Buchstaben wird üblicherweise mit  $n!$  (gesprochen „n-Fakultät“) bezeichnet:

$$n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n, \quad n \in \mathbb{N}.$$

Es erweist sich im übrigen als zweckmäßig, die Notation  $0! = 1$  einzuführen.

**Beispiel:** Ein Stammtisch mit 10 Personen kann also  $10! = 1 \cdot 2 \cdot 3 \cdots 9 \cdot 10 = 3628800$  Sitzordnungen einnehmen. Tagt er wöchentlich und rechnet man das Jahr zu 52 Wochen, dann müßte er  $\frac{10!}{52} = 69784$  Jahre lang tagen, um alle möglichen Sitzordnungen einmal einzunehmen. ■

Die Zahlen  $n!$  wachsen also mit  $n$  sehr stark. In diesem mathematischen Faktum ist nicht zuletzt der Grund für die Vielfalt der Formen des Lebendigen zu sehen. Diese hängt zusammen mit der Anzahl der Möglichkeiten, die Buchstaben A,C,G,T, mit denen die genetische Information geschrieben ist, zu Wörtern großer Länge zu kombinieren.

**4. Grundaufgabe:** Wieviele verschiedene Wörter lassen sich durch Umordnung der Buchstaben eines gegebenen Wortes, bei dem *nicht* alle Buchstaben verschieden sind, bilden?

Das Wort *abacbbd* hat 7 Buchstaben, wobei der Buchstabe *a* zweimal und der Buchstabe *b* dreimal auftritt. Die zu bestimmende Anzahl der Umordnungen, die zu neuen Wörtern führt, sei  $X$ . Denkt man sich die Buchstaben *aa* und *bbb* mit  $a_1a_2$  und  $b_1b_2b_3$  indiziert, dann würden durch Umordnung eines fest gewählten der neuen Wörter  $2!3!$  weitere Wörter entstehen, insgesamt also  $2!3!X$  Wörter. Diese Zahl muß aber gleich der Zahl der Permutationen eines Wortes mit 7 *verschiedenen* Buchstaben sein. Daher gilt

$$2! \cdot 3! \cdot X = 7!,$$

also ist

$$X = \frac{7!}{2! \cdot 3!}$$

die gesuchte Anzahl.

Hat man allgemein ein Wort mit  $n$  Buchstaben, bei dem  $m$  Buchstaben mehrfach auftreten und zwar mit den jeweiligen Anzahlen  $n_1, n_2, \dots, n_m$ , dann ist

$$X = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_m!}$$

die Anzahl der „kreativen“ Permutationen des Ausgangswortes (dieses eingeschlossen).

### 3.3 Anzahl der $k$ -elementigen Teilmengen einer $n$ -elementigen Menge

Bei der folgenden kombinatorischen Fragestellung spielt die *Anordnung* der Buchstaben (Elemente) keine Rolle.

**5. Grundaufgabe:** Wieviele Teilalphabete der Länge  $k$  lassen sich aus einem Alphabet der Länge  $n$  bilden?

Die Grundaufgabe fragt also, in abstrakterer Formulierung, nach der Anzahl der  $k$ -elementigen Teilmengen einer  $n$ -elementigen Menge.

Nach Grundaufgabe 2 kann man aus einem Alphabet der Länge  $n$

$$n \cdot (n - 1) \cdot \dots \cdot (n - k + 1)$$

buchstabenverschiedene Wörter der Länge  $k$  bilden. Aus jedem Teilalphabet der Länge  $k$  bildet man durch Umordnung  $k!$  buchstabenverschiedene Wörter der Länge  $k$ . Wenn

### 3.3. ANZAHL DER $k$ -ELEMENTIGEN TEILMENGEN EINER $n$ -ELEMENTIGEN MENGE 57

also  $X$  die Zahl der  $k$ -Alphabete ist, die man aus dem  $n$ -Alphabet bilden kann, dann gilt

$$k! \cdot X = n \cdot (n - 1) \cdots (n - k + 1),$$

also ist

$$X = \frac{n \cdot (n - 1) \cdots (n - k + 1)}{k!}$$

die gesuchte Anzahl. Die Zahl  $X$  bezeichnet man mit dem Symbol  $\binom{n}{k}$ , gesprochen „ $n$  über  $k$ “.

Die Anzahl der  $k$ -elementigen Teilmengen einer  $n$ -elementigen Menge ist also gleich

$$\binom{n}{k} = \frac{n(n - 1) \cdots (n - k + 1)}{k!},$$

mit  $n \geq 1$  und  $1 \leq k \leq n$ . Erweitert man den Ausdruck auf der rechten Seite mit  $(n - k)!$ , so ergibt sich die Formel:

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

Diese ist aufgrund der Festsetzung  $0! = 1$  auch in den Fällen  $k = 0$  und  $n = k = 0$  erklärt. Also gilt dann  $n \geq 0$  und  $0 \leq k \leq n$  für „Zähler“  $n$  und „Nenner“  $k$  beim Symbol  $\binom{n}{k}$  (s. Bemerkung unten).

Die Zahlen  $\binom{n}{k}$  sind die *Binomialkoeffizienten*. Der Grund für diese Bezeichnung wird im folgenden Abschnitt einsichtig. Sondert man aus einer  $n$ -elementigen Menge eine  $k$ -elementige Menge aus, so bleibt eine  $n - k$  elementige Menge übrig und umgekehrt. Jede  $k$ -elementige Teilmenge bestimmt also eindeutig eine  $(n - k)$ -elementige Teilmenge und vice versa. Folglich gibt es genau so viele  $k$ -elementige wie  $(n - k)$ -elementige Teilmengen einer  $n$ -elementigen Menge, das heißt, es besteht die Beziehung

$$\binom{n}{k} = \binom{n}{n - k},$$

deren Richtigkeit man auch unmittelbar mit (\*) bestätigt.

**Bemerkung:** Die Festsetzung  $0! = 1$  erweist sich im nachhinein als zweckmäßig, denn die Version (\*) als Definition der Binomialkoeffizienten ist nämlich auch für  $k = 0$  und  $n = 0$  erklärt und sinnvoll. Man erhält

$$\binom{n}{0} = \frac{n!}{0!n!} = 1$$

als Anzahl der 0-elementigen Teilmengen einer  $n$ -elementigen Menge. Diese ist gleich eins, da die leere Menge  $\emptyset$  Teilmenge einer jeden Menge ist (vgl. Abschnitt 1.3). Also gilt auch  $\emptyset \subset \emptyset$ . Diesen Fall erfaßt der Binomialkoeffizient

$$\binom{0}{0} = \frac{0!}{0!0!} = 1. \quad \blacktriangle$$

### 3.4 Der binomische Lehrsatz

Aus dem Schulunterricht kennt man die Formel  $(x+y)^2 = x^2 + 2xy + y^2$  und vielleicht auch  $(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$ . Als Verallgemeinerung dieser Formeln stellt sich die Aufgabe, eine entsprechende Entwicklung des Ausdrucks  $(x+y)^n$  zu finden. Der Ausdruck  $(x+y)^n$  besagt, multipliziere die Klammer  $(x+y)$   $n$ -mal mit sich selbst:

$$(x+y)^n = (x+y)(x+y)\dots(x+y)$$

Das Ausmultiplizieren der  $n$  Klammern geschieht so, dass aus jeder der Klammern entweder  $x$  oder  $y$  ausgewählt wird, man das Produkt dieser  $n$  Faktoren bildet und die erhaltenen Produkte addiert.

Tritt in einem Produkt  $k$ -mal  $x$  auf, so muß dieses Produkt  $(n-k)$ -mal  $y$  enthalten. Diese Produkte haben also die Form

$$x^k y^{n-k}, \quad k = 0, 1, \dots, n.$$

Das Produkt  $x^k y^{n-k}$  (mit festem  $k$ ) entsteht dadurch, dass aus  $n$  Klammern  $k$  Klammern ausgewählt werden, denen man das  $x$  entnimmt. Aus den restlichen  $(n-k)$  Klammern stammen die  $y$ . Aus  $n$  Klammern hat man laut 5-ter-Grundaufgabe  $\binom{n}{k}$  Möglichkeiten,  $k$  Klammern auszuwählen, so dass das Produkt  $x^k y^{n-k}$  mit der Vielfachheit  $\binom{n}{k}$  auftritt. Die gesuchte Entwicklung lautet also

$$(**) \quad (x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$

Dies ist der *binomische Lehrsatz*, dem die Binomialkoeffizienten  $\binom{n}{k}$  ihren Namen verdanken.

Aus der Formel  $(**)$  können wir eine erste kombinatorische Folgerung ziehen: Die Anzahl sämtlicher Teilmengen einer Menge  $N$  mit  $n$  Elementen, das ist die Anzahl  $|\mathbb{P}(N)|$  der Elemente der Potenzmenge  $\mathbb{P}(N)$  von  $N$ , ist gleich  $2^n$ . Denn offenbar gilt einerseits

$$|\mathbb{P}(N)| = \sum_{k=0}^n \binom{n}{k}$$

und andererseits erhält man die rechtsstehende Summe, wenn man in Gleichung  $(**)$   $x = y = 1$  setzt. Die linke Seite bei  $(**)$  ist dann gleich  $(1+1)^n = 2^n$ . Also gilt

$$|\mathbb{P}(N)| = \sum_{k=0}^n \binom{n}{k} = 2^n.$$

### Das Pascalsche Dreieck

Die Anzahl der 4-elementigen Teilmengen einer 7-elementigen Menge ist  $\binom{7}{4} = \frac{7 \cdot 6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3 \cdot 4} = 35$ . Markiert man bei der Menge  $\{1, 2, 3, 4, 5, 6, 7\}$  ein beliebiges Element, z.B. die 5,  $\{1, 2, 3, 4, \textcircled{5}, 6, 7\}$ , dann kann man die 4-elementige Teilmengen in zwei Gruppen aufteilen. Die Teilmengen der einen Gruppe enthalten die 5, die Teilmengen der anderen Gruppe enthalten die 5 nicht. Die Anzahl der Teilmengen der ersten Art ist  $\binom{7-1}{3}$ , die der zweiten  $\binom{7-1}{4}$ . Folglich gilt

$$\binom{7}{4} = \binom{7-1}{3} + \binom{7-1}{4}.$$

Dieselbe Überlegung führt im allgemeinen Fall zu der Beziehung

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

Dies ist eine *Rekursionsformel*, die erlaubt, die Binomialkoeffizienten mit dem „Zähler“  $n$  zu berechnen, wenn diejenigen mit dem Zähler  $n - 1$  bekannt sind. Auf diese Weise lassen sich die Binomialkoeffizienten sukzessive „von unten“ aufbauen. Das da-

Tabelle 3.1: Das Pascalsche Dreieck.

$n \setminus k$	0	1	2	3	4	5	6
0	1						
1	1	1					
2	1	2	1				
3	1	3	3	1			
4	1	4	6	4	1		
5	1	<b>5</b>	<b>10</b>	10	5	1	
6	1	6	<b>15</b>	20	15	6	1
...	...	...	...	...	...	...	...

bei entstehende Schema ist das *Pascalsche Dreieck*, benannt nach dem französischen Mathematiker, Physiker und religiösen Denker Blaise Pascal (1623-1662).

**Bemerkung:** Versieht man die Binomialkoeffizienten in einer Zeile des Pascalschen Dreiecks abwechselnd mit den Vorzeichen  $+$  und  $-$  und addiert diese Zahlen, so erhält man jedesmal die Summe null. Somit scheint zu gelten

$$\sum_{k=0}^n (-1)^k \binom{n}{k} = 0.$$

Die Richtigkeit dieser Beziehung bestätigt man unmittelbar, indem man bei (\*\*)  $x = -1$  und  $y = 1$  setzt.

Neben den hier genannten gelten für die Binomialkoeffizienten zahlreiche weitere interessante Beziehungen. ▲

# Kapitel 4

## Folgen und Reihen

Eine Zahlenfolge, oder einfacher, eine Folge, wird dadurch gegeben, dass jeder natürlichen Zahl  $n \in \mathbb{N}$  eine reelle Zahl  $x_n$  (oder eine komplexe Zahl  $z_n$ ) zugeordnet wird. Die Folgenglieder sind dann die Zahlen  $x_1, x_2, x_3, \dots$ . Man bezeichnet diese Folge auch mit  $\{x_n\}$ . Manchmal ist es zweckmäßig, die Indizierung der Folgenglieder nicht mit  $n = 1$ , sondern mit  $n = 0$  zu beginnen.

### 4.1 Beispiele für Folgen

Einige elementare Beispiele sollen mit den Begriff der Zahlenfolge vertraut machen.

#### Die arithmetische Folge

Liegt auf einem Konto ein Basisbetrag  $a$  und kommt monatlich ein konstanter Betrag  $b$  hinzu, dann beschreibt die Zahlenfolge  $a, a + b, a + 2b, a + 3b, \dots$  die Entwicklung des Kontostandes. Nach der  $n$ -ten Einzahlung ist also  $x_n = a + nb$  mit  $n = 0, 1, 2, \dots$  der Kontostand. Indem man die Indizierung mit  $n = 0$  beginnen lässt, wird der Kontostand vor der ersten Einzahlung miterfasst.

Die Folge  $x_n = a + nb$ , die *arithmetische* Folge, verdankt ihren Namen dem Umstand, dass jedes Folgenglied  $x_n$  das arithmetische Mittel der benachbarten Folgenglieder  $x_{n-1}$  und  $x_{n+1}$  ist. Das arithmetische Mittel zweier Zahlen  $x$  und  $y$  ist die Zahl  $\frac{1}{2}(x+y)$ . Man erhält in der Tat

$$\begin{aligned}\frac{1}{2}(x_{n-1} + x_{n+1}) &= \frac{1}{2}(2a + (n-1)b + (n+1)b) \\ &= a + nb \\ &= x_n.\end{aligned}$$

Der Kontostand  $x_n$  wächst offenbar unbegrenzt. Man schreibt in diesem Fall  $x_n \rightarrow \infty$  für  $n \rightarrow \infty$  und sagt, die Folge  $x_n$  strebt gegen „unendlich“ wenn der Index  $n$  gegen

unendlich geht.

### Die geometrische Folge

Wenn in einer Nährlösung  $a$  Bakterien eingesetzt werden und jedes Bakterium sich stündlich teilt, dann befinden sich zuerst  $a$ , dann  $a \cdot 2$ , dann  $a \cdot 2^2$ , dann  $a \cdot 2^3$  und schließlich nach der  $n$ -ten Teilung  $x_n = a \cdot 2^n$  Bakterien in der Nährlösung. Die Zahlen  $x_n = a \cdot 2^n$ , allgemein  $x_n = a \cdot q^n$ ,  $q \in \mathbb{R}$ ,  $n = 0, 1, 2, \dots$ , bilden eine geometrische Folge.

Bei einer geometrischen Folge ist das Folgenglied  $x_n$  das geometrische Mittel der benachbarten Glieder  $x_{n-1}$  und  $x_{n+1}$ . Das geometrische Mittel zweier Zahlen  $x, y > 0$  ist die Zahl  $\sqrt{x \cdot y}$ . Man erhält also bei der geometrischen Folge:

$$\begin{aligned}\sqrt{x_{n-1} \cdot x_{n+1}} &= \sqrt{aq^{n-1} \cdot aq^{n+1}} \\ &= aq^n \\ &= x_n.\end{aligned}$$

Das Wachstum der Bakterienpopulation ist unbeschränkt, das heißt, es gilt  $a \cdot 2^n \rightarrow \infty$  für  $n \rightarrow \infty$ . Dasselbe gilt für jede Folge  $x_n = aq^n$ , wenn  $q > 1$  ist. Für  $q = 1$  ist  $x_n = a$ ,  $n \in \mathbb{N}$ , eine konstante Folge.

Die Zahlen  $(\frac{1}{2})^n = \frac{1}{2^n}$  werden beliebig klein, sofern nur  $n$  genügend groß ist. Dasselbe gilt für alle Zahlen  $q$  im Intervall  $0 < q < 1$ . Daher werden die Folgenglieder  $x_n = aq^n$  mit wachsendem Index  $n$  beliebig klein. Man beschreibt diesen Sachverhalt durch die Notation  $x_n \rightarrow 0$  für  $n \rightarrow \infty$ . Die Folge  $x_n$  strebt gegen den „Grenzwert“ 0 für  $n$  gegen unendlich.

Für  $q > 0$  ist  $(-q)^n = (-1)^n q^n$  abwechselnd positiv oder negativ jenachdem, ob  $n$  gerade oder ungerade ist. Ist  $0 < q < 1$ , dann gilt ebenfalls  $(-q)^n \rightarrow 0$  für  $n \rightarrow \infty$ .

Wir halten fest: Ist  $-1 < q < 1$ , dann werden die Folgenglieder  $x_n = aq^n$  beliebig klein, es gilt  $x_n \rightarrow 0$  für  $n \rightarrow \infty$  (der Fall  $q = 0$  ist natürlich trivial).

### Die harmonische Folge

Die Zahlenfolge  $x_n = \frac{1}{n}$ ,  $n \in \mathbb{N}$ , ist die harmonische Folge. Ihre Folgenglieder  $1, \frac{1}{2}, \frac{1}{3}, \dots$  liegen zwischen null und eins. Die Folge ist monoton fallend, das heißt, es gilt  $\frac{1}{n+1} < \frac{1}{n}$ , der Wert null wird aber gleichwohl nicht erreicht, vielmehr gilt immer  $0 < \frac{1}{n}$ . Andererseits wird jeder beliebige kleine Wert  $\varepsilon > 0$  durch unendlich viele Glieder der Folge unterschritten. Setzt man zum Beispiel  $\varepsilon = \frac{1}{100}$ , dann gilt  $\frac{1}{n} < \frac{1}{100}$  für alle  $n > 100$ . Man sagt, die Folge  $x_n = \frac{1}{n}$  strebt gegen den Grenzwert 0 für  $n \rightarrow \infty$ :  $x_n \rightarrow 0$  für  $n \rightarrow \infty$ . (Eine genauere Erörterung des Grenzwertbegriffs findet sich im Abschnitt 4.5.)

Die Bezeichnung „harmonische“ Folge leitet sich her vom harmonischen Mittel

$$h(x, y) = \left[ \frac{1}{2} \left( \frac{1}{x} + \frac{1}{y} \right) \right]^{-1}$$

zweier Zahlen  $x, y > 0$ . Das Folgenglied  $x_n = \frac{1}{n}$  der harmonischen Folge ist nämlich das harmonische Mittel der benachbarten Folgenglieder  $x_{n-1}$  und  $x_{n+1}$ . Man hat nämlich

$$\begin{aligned} h(x_{n-1}, x_{n+1}) &= \left[ \frac{1}{2} \left( \frac{1}{x_{n-1}} + \frac{1}{x_{n+1}} \right) \right]^{-1} \\ &= \left[ \frac{1}{2} (n-1 + n+1) \right]^{-1} \\ &= \frac{1}{n} = x_n. \end{aligned}$$

**Bemerkung:** Das harmonische Mittel hat seinen Namen aufgrund seiner Bedeutung in der musikalischen Harmonielehre. So ergibt z.B. das harmonische Mittel der einer Note entsprechenden Wellenlänge und derjenigen der zugehörigen Oktave eine Quinte. ▲

## Die Eulersche Zahl $e$

Aus einem Kapital  $K$  wird bei 100%iger Verzinsung am Ende eines Jahres das Kapital  $K + K = K \cdot 2$ . Steht man sich bei halbjähriger Verzinsung zu 50% nach einem Jahr besser? Nach einem halben Jahr wird dann aus  $K$  das Kapital  $K + \frac{1}{2}K = K(1 + \frac{1}{2})$  und nach einem weiteren halben Jahr  $K(1 + \frac{1}{2}) + K(1 + \frac{1}{2})\frac{1}{2} = K(1 + \frac{1}{2})^2 = K \cdot 2,25$ . Bei Verzinsung nach jeweils drei Monaten zu  $\frac{100}{3}\%$  ergeben sich die Zwischenkapitalien  $K + \frac{1}{3}K = K(1 + \frac{1}{3})$ , dann  $K(1 + \frac{1}{3}) + K(1 + \frac{1}{3})\frac{1}{3} = K(1 + \frac{1}{3})^2$  und schließlich am Jahresende  $K(1 + \frac{1}{3})^2 + K(1 + \frac{1}{3})^2\frac{1}{3} = K(1 + \frac{1}{3})^3 = K \cdot 2,37$ . Die Art der Verzinsung wird offenbar immer günstiger. Teilt man das Jahr in  $n$  Teile und verzinst in jedem dieser Teiljahre das bis dahin angesammelte Kapital mit  $\frac{100}{n}\%$ , dann ist  $K(1 + \frac{1}{n})^n$  das Kapital am Jahresende. Dieses Ergebnis veranlasst, die Zahlenfolge

$$x_n = \left( 1 + \frac{1}{n} \right)^n, \quad n \in \mathbb{N},$$

zu betrachten. Sie vereinigt in sich zwei gegenläufige Tendenzen: Die Folge  $1 + \frac{1}{n}$  in der Klammer strebt gegen eins für  $n \rightarrow \infty$ , bleibt aber immer größer als eins. Andererseits wachsen die Potenzen einer Zahl größer als eins über alle Grenzen. Man kann aber zeigen, dass dies bei den Zahlen  $x_n$  nicht der Fall ist, dass vielmehr gilt  $x_n < 3$ . Man kann außerdem zeigen, der empirische Befund bei den Anfangsgliedern der Folge ist allgemeingültig: Die Folge der Zahlen  $x_n$  ist monoton wachsend, das

heißt, es gilt  $x_n < x_{n+1}$  für alle  $n \in \mathbb{N}$ . Eine monotone und beschränkte Zahlenfolge strebt aber immer gegen einen Grenzwert. Die Zahlen  $x_n$  streben gegen die Eulersche Zahl  $e = 2,718\dots$ :

$$x_n = \left(1 + \frac{1}{n}\right)^n \rightarrow e$$

für  $n \rightarrow \infty$  (vgl. hierzu Abschnitt 4.5).

## 4.2 Logistisches Wachstum

Misst man die Größe einer Population  $X$  in jährlichem Abstand – oder einer anderen Zeiteinheit – dann erhält man eine Zahlenfolge  $x_0, x_1, x_2, \dots$ . Die Zahlen  $x_n$  geben die Anzahl der Individuen der Population zum Zeitpunkt  $n$  an oder die jeweilige Populationsdichte, das ist die Anzahl der Individuen pro Flächeneinheit.

### Exponentielles Wachstum

Wird die Dynamik der Population ausschließlich durch Geburts- und Sterbeprozesse bestimmt und ist  $a > 0$  die durchschnittliche Zahl der Geburten pro Individuum und Zeiteinheit (Geburtenrate) und  $b$  mit  $0 \leq b \leq 1$  die durchschnittliche Sterberate, das ist der Verlust pro Individuum und Zeiteinheit, den die Population erleidet, dann besteht zwischen den aufeinanderfolgenden Größen  $x_{n-1}$  und  $x_n$  der Zusammenhang

$$x_n - x_{n-1} = a x_{n-1} - b x_{n-1}$$

oder

$$(*) \quad x_n = (1 + a - b) x_{n-1}.$$

Mit  $c = 1 + a - b$  hat man daher

$$x_n = c x_{n-1} = c^2 x_{n-2} = c^3 x_{n-3} = \dots = c^n x_0,$$

wenn  $x_0$  die Anfangsgröße der Population ist. Die Zahlenfolge

$$x_n = x_0 \cdot c^n$$

ist wie beim Wachstum der Bakterienkultur eine geometrische Zahlenfolge. Das Wachstum der Population ist *exponentiell*.

Für die Zahl  $c$  sind die Fälle  $0 < c < 1$ ,  $c = 1$  und  $c > 1$  möglich. Im ersten Fall folgt  $c^n \rightarrow 0$  für  $n \rightarrow \infty$ , im zweiten  $c^n = 1$  für alle  $n$  und im dritten Fall gilt  $c^n \rightarrow \infty$  für  $n \rightarrow \infty$ . Alle drei Fälle liefern keine langfristig gültige Beschreibung einer real existierenden Population. Im ersten Fall nämlich stirbt die Population aus (d.h., sie wäre gar nicht vorhanden), im zweiten haben alle Generationen exakt die gleiche Größe (was unrealistisch ist) und im dritten Fall wächst die Population über alle Grenzen.

Die Annahme eines unbegrenzten Wachstums ist aber offenkundig unrealistisch. Die Endlichkeit des Territoriums, auf dem eine Population siedelt oder die Endlichkeit der Nahrungsressourcen verhindern schrankenloses Wachstum. Ein realistisches Wachstumsmodell muss daher diese Endlichkeitsbedingungen berücksichtigen.

Dies geschieht dadurch, dass man beim Modell (\*) die Voraussetzung der Konstanz der Sterberate  $b$  fallen lässt. Dies erscheint deshalb plausibel, weil die Größe der Population deren Lebensbedingungen beeinflusst. Je größer die Population ist, umso größer ist der Verbrauch an Ressourcen und damit umso größer die Sterberate. Der einfachste Ansatz nimmt daher an, die Sterberate sei proportional zur aktuellen Größe der Population und ersetzt im Modell (\*) die Konstante  $b$  durch den Ausdruck  $cx_{n-1}$  mit einer Konstanten  $c > 0$ . Man erhält so die Rekursionsformel

$$x_n = (1 + a - cx_{n-1})x_{n-1}$$

oder

$$(**) \quad x_n = c(K - x_{n-1})x_{n-1}$$

mit  $K = (1 + a)/c$  und  $n = 1, 2, 3, \dots$ . Dies ist die Gleichung des *logistischen* Populationswachstums. Ihre Lösung  $x_n$  lässt sich nicht mehr, wie beim exponentiellen Wachstum, explizit als Funktion von  $n$ , angeben. (Die Herkunft der Bezeichnung „logistisch“ ist unklar).

### Kapazitätsgrenze und Gleichgewichtslage

Die Konstante  $K$  hat die Bedeutung einer *Kapazitätsgrenze*, die von der Umwelt abhängt, welche die Population trägt. Ist für ein  $n - 1$  nämlich  $x_{n-1} = K$ , dann gilt  $x_n = x_{n+1} = \dots = 0$ , die Population bricht zusammen. Ist gar  $x_{n-1} > K$ , dann ist  $x_n < 0$ , was von der Bedeutung der Zahl  $x_n$  her nicht eintreten dürfte. Ein „vernünftiges“ Modell muss diese Möglichkeit daher ausschließen und die Forderung  $0 \leq x_n \leq K$  erfüllen.

Bevor wir eine Bedingung formulieren die sicherstellt, dass die Kapazitätsgrenze  $K$  nicht überschritten wird, führen wir einen zweiten natürlichen Parameter in das logistische Modell ein, die *Gleichgewichtslage*.

Auf diese wird man durch folgende Frage geführt: Gibt es eine Populationsgröße  $E$  die zeitlich konstant bleibt? Das bedeutet, aus  $x_{n-1} = E$  folgt  $x_n = E$  und damit  $x_{n+1} = x_{n+1} = x_{n+2} = \dots = E$ . Für diese Größe  $E$  muss dann gelten

$$E = c(K - E)E.$$

Schließt man die triviale Lösung  $E = 0$  dieser Gleichung aus, so folgt

$$1 = c(K - E),$$

das heißt

$$c = \frac{1}{K - E}$$

oder

$$E = K - \frac{1}{c}.$$

Wir können daher in der logistischen Gleichung (\*\*) den Parameter  $c$  durch die „natürlichen“ Parameter  $K$  und  $E$  ersetzen und erhalten die Form

$$(***) \quad x_n = \frac{1}{K - E} (K - x_{n-1}) x_{n-1}$$

der logistischen Gleichung, in der nur noch die für eine Population charakteristischen Größen  $K$  und  $E$  auftreten.

Um die Forderung  $0 \leq x_n \leq K$  zu erfüllen, betrachten wir die Funktion

$$f(x) = \frac{1}{K - E} (K - x)x.$$

Sie beschreibt eine nach unten offene Parabel, die an den Stellen  $x = 0$  und  $x = K$  die  $x$ -Achse schneidet. Diese hat ihr Maximum bei  $x = K/2$ . Wenn daher gilt

$$f\left(\frac{K}{2}\right) \leq K,$$

dann gilt auch immer  $0 \leq x_n \leq K$ , sofern nur  $0 \leq x_{n-1} \leq K$  richtig ist. Die obige Bedingung bedeutet

$$\frac{1}{K - E} \cdot \frac{K}{2} \cdot \frac{K}{2} \leq K$$

oder

$$K \leq 4(K - E),$$

das heißt,

$$E \leq \frac{3}{4}K.$$

**Beispiel:** Es sei  $K = 100$  die Kapazitätsgrenze und  $E = 75$  die Gleichgewichtslage. Die logistische Gleichung zu dieser Population lautet dann

$$x_n = \frac{1}{25} (100 - x_{n-1}) x_{n-1}.$$

Die folgende Zeichnung zeigt die Entwicklung der Population bis zum Wert  $x_{10}$ , wenn der Startwert  $x_0 = 10$  ist. Man beobachtet in diesem Beispiel unregelmäßige Oszillationen der Population um ihre Gleichgewichtslage. ■

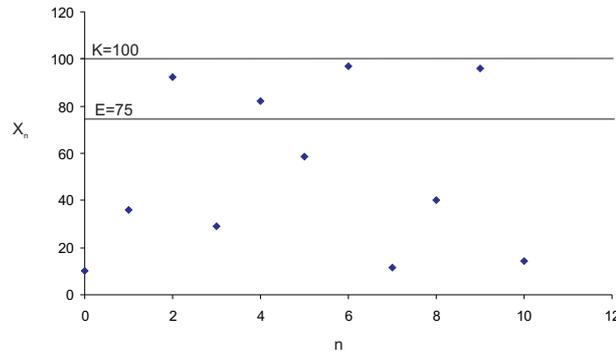


Abbildung 4.1: Logistisches Wachstum.

Die Abfolge der Punkte in Abb. 4.1 legt die folgende Vermutung nahe: Wenn immer  $x_{n-1} < E$  ist, dann folgt  $x_n > x_{n-1}$  und wenn  $x_{n-1} > E$  ist, dann folgt  $x_n < x_{n-1}$ . Das bedeutet, die Population hat immer die Tendenz, sich in Richtung auf die Gleichgewichtslage zu bewegen. Dies folgt in der Tat unmittelbar aus der logistischen Gleichung (\*\*), wie die folgende kleine Rechnung zeigt.

Mit

$$x_n = \frac{1}{K - E} (K - x_{n-1}) x_{n-1}$$

erhält man unmittelbar

$$x_n - x_{n-1} = \frac{1}{K - E} (E - x_{n-1}) x_{n-1}.$$

Ist daher  $E - x_{n-1} > 0$ , so auch  $x_n - x_{n-1}$ . Ist hingegen  $E - x_{n-1} < 0$ , so auch  $x_n - x_{n-1}$ . Damit ist alles gezeigt.

### Stabiles Gleichgewicht

Die Population in Abb. 4.1 oszilliert ständig um ihre Gleichgewichtslage offenbar ohne eine Tendenz, diese im Laufe der Zeit, es sei denn gleichsam zufällig, zu erreichen. Wir wollen im folgenden eine Bedingung aufsuchen unter der die Gleichgewichtslage eine Art Anziehungskraft entwickelt, so dass die Population im Laufe der Zeit der Gleichgewichtslage immer näher kommt.

Hierzu messen wir die Größe der Population nicht mehr durch ihren Abstand vom Wert null, sondern durch ihren Abstand von der Gleichgewichtslage  $E$ . Das heißt, wir setzen bei (\*\*\*)

$$x_n = E + y_n$$

und erhalten

$$\begin{aligned} E + y_n &= \frac{1}{K - E} (K - E - y_{n-1}) (E + y_{n-1}) \\ &= \frac{1}{K - E} ((K - E)E + (K - 2E)y_{n-1} - y_{n-1}^2) \\ &= E + \frac{K - 2E}{K - E} y_{n-1} - \frac{y_{n-1}^2}{K - E}. \end{aligned}$$

Also gilt

$$y_n = \frac{K - 2E}{K - E} y_{n-1} - \frac{y_{n-1}^2}{K - E}.$$

Wenn  $x_n$  in die Nähe der Gleichgewichtslage  $E$  kommt, dann ist  $y_{n-1}$  klein und erst recht  $y_{n-1}^2$ . Daher kann man den Term  $y_{n-1}^2/(K - E)$  weglassen ohne einen großen Fehler zu machen. Man erhält dann die vereinfachte Rekursionsformel

$$y_n \approx \frac{K - 2E}{K - E} y_{n-1}$$

und damit

$$y_n \approx \left( \frac{K - 2E}{K - E} \right)^n y_0.$$

Wenn also gilt

$$-1 < \frac{K - 2E}{K - E} < 1,$$

dann wird  $y_n$  immer kleiner von Generation zu Generation. Das aber bedeutet,  $x_n$  nähert sich immer mehr der Gleichgewichtslage  $E$ .

Die rechte Seite der obigen Ungleichung ist wegen  $K - 2E < K - E$  von selbst erfüllt. Damit auch die linke Seite richtig ist muss gelten

$$-(K - E) < K - 2E$$

oder

$$3E < 2K,$$

das heißt,

$$E < \frac{2}{3} K.$$

In diesem Fall wird also die Population, sofern sie in die Nähe der Gleichgewichtslage kommt, von dieser gleichsam angezogen bzw. kleine Störungen der Gleichgewichtslage werden umgehend korrigiert. Ein solches Gleichgewicht wird *stabil* genannt. Wir fassen unsere Ergebnisse zusammen:

**Satz:** Bei einer logistisch wachsenden Population stehen Kapazitätsgrenze  $K$  und Gleichgewichtslage  $E$  in der Beziehung  $E \leq 3/4K$ . Wenn die Parameter  $E$  und  $K$  die Ungleichung  $E < 2/3K$  erfüllen, dann ist die Gleichgewichtslage  $E$  stabil.

**Beispiel:** Mit  $K = 100$  und  $E = 40$  ist die Stabilitätsbedingung des Satzes erfüllt. Die Rekursionsformel lautet mit diesen Parametern

$$x_n = \frac{1}{60}(100 - x_{n-1})x_{n-1}.$$

Mit dem Anfangswert  $x_0 = 90$  nimmt die Entwicklung der Population bis zum Wert  $x_{10}$  den in der Abbildung dargestellten Verlauf an. Es zeigt sich, dass die zunächst weit

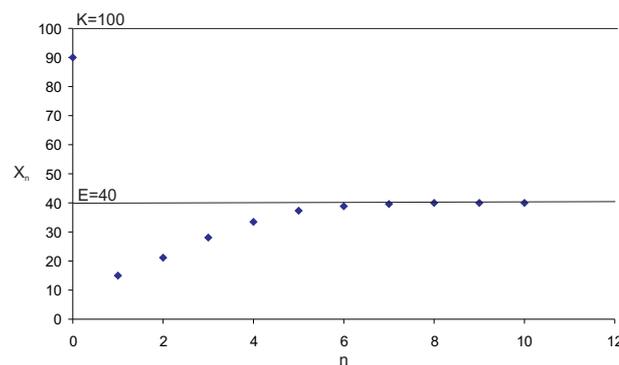


Abbildung 4.2: Stabiles Gleichgewicht.

vom Gleichgewicht befindliche Anfangslage schnell in Richtung des Wertes  $E = 40$  korrigiert wird und diesem dann ständig näherkommt. ■

**Bemerkung:** Eine eigentliche Rechtfertigung des Verfahrens der sogenannten „Linearisierung“ der Rekursionsformel (\*\*\*) in der Nähe der Gleichgewichtslage  $E$  durch Vernachlässigung des quadratischen Terms  $y_{n-1}^2$ , die zu der Bedingung  $E < 2/3K$  führt, kann erst mit den Mitteln der Differentialrechnung erfolgen. Zugleich lässt sich dann auch der „Anziehungsbereich“ um die Gleichgewichtslage genau beschreiben. (Vgl. Abschnitt 6.4) ▲

### 4.3 Fibonacci-Wachstum

Von Leonardo von Pisa (ca. 1170-1240), genannt Fibonacci, das heißt „Sohn des Bonacci“, dem bedeutendsten Mathematiker des Mittelalters, stammt die bekannte Kaninchen–Aufgabe. Sie findet sich in seinem Buch *liber abaci* (der „abacus“ ist das Rechenbrett), erschienen 1202, in dem er die damals in Europa nur wenig bekannten indisch–arabischen Ziffern einführte. Die Aufgabe lautete:

Jemand setzt ein Kaninchenpaar in ein allseitig von einer Mauer umgebenes Gelände. Wieviele Kaninchenpaare werden von diesem Paar in einem Jahr erzeugt werden, wenn man annimmt, dass ein Kaninchenpaar jeden Monat ein weiteres Kaninchenpaar hervorbringt, welches selbst wiederum ab dem zweiten Monat produktiv wird.

In leichter Abwandlung der historischen Fragestellung nehmen wir an, dass das erste Kaninchenpaar ein Neugeborenes ist. Außerdem lassen wir die zeitliche Beschränkung in der Fragestellung fallen und nehmen an, dass Kaninchen ewig leben.

Wir unterscheiden bei der Kaninchenpopulation also zwei Altersgruppen: Noch nicht fortpflanzungsfähige junge Kaninchenpaare ( $J$ ) und erwachsene Kaninchenpaare ( $E$ ). Dann durchläuft ein einzelnes Kaninchenpaar die Lebensstufen  $J \rightarrow J \rightarrow E \rightarrow E \rightarrow E \rightarrow \dots$ . Beginnt die Entwicklung der Population mit einem neugeborenen Kaninchenpaar, dann wächst die Population in den ersten 10 Monaten nach folgendem Schema. In der oberen Zeile des Schemas steht jeweils die Nummer  $n = 1, 2, 3, \dots$ ,

Tabelle 4.1: Wachstum der Kaninchenpopulation

$n$	1	2	3	4	5	6	7	8	9	10
	$J$	$J$	$E$	$E$	$E$	$E$	$E$	$E$	$E$	$E$
			$J$	$J$	$E$	$E$	$E$	$E$	$E$	$E$
				$J$	$J$	$E$	$E$	$E$	$E$	$E$
					$2J$	$2J$	$2E$	$2E$	$2E$	$2E$
						$3J$	$3J$	$3E$	$3E$	$3E$
							$5J$	$5J$	$5E$	$5E$
								$8J$	$8J$	$8E$
									$13J$	$13J$
										$21J$
$F_n$	1	1	2	3	5	8	13	21	34	55

des Monats seit Beginn der Entwicklung, in der unteren Zeile die jeweilige Anzahl  $F_n$

der Kaninchenpaare im Monat  $n$ . Man beobachtet nun folgende Regelmäßigkeit: Jede Zahl in der unteren Zeile ist die Summe der beiden vorhergehenden Zahlen. Dies gilt auch bei beliebiger Erweiterung des Schemas. Folglich besteht die *Rekursionsformel*

$$(*) \quad F_{n+2} = F_{n+1} + F_n$$

für  $n = 1, 2, 3, \dots$

Die Zahlen der Folge  $F_n$ , beginnend mit  $1, 1, 2, 3, 5, \dots$ , sind die *Fibonacci-Zahlen*. Die Fibonacci-Zahlen haben viele interessante mathematische Eigenschaften. Sie sind bis heute ein Gegenstand der Untersuchung, dem sogar eine eigene Zeitschrift gewidmet ist. Abgesehen von dem schönen, aber wenig realistischen Kaninchenproblem, treten Fibonacci-Zahlen in der belebten Natur an zahlreichen Stellen als Strukturprinzip auf. Außerdem stehen sie in einem engen Zusammenhang mit dem *Goldenen Schnitt*, der „divina proportio“ der Renaissance. Dieser findet sich als Maßverhältnis in vielen Werken der Architektur, der bildenden Kunst und auch in der Musik (vgl. hierzu das Buch von A. Beutelspacher und B. Peter: *Der Goldene Schnitt*, 2. Aufl., Heidelberg 1996).

**Bemerkung:** Natürlich kann man die Rekursionsformel (\*) auch nicht-empirisch herleiten. Hierzu muss man die aufeinanderfolgenden Altersgruppen durch  $J, J'$  unterscheiden und kenntlich machen. Die Generation  $n$  der Population hat  $J_n$  Individuen in Gruppe  $J$ ,  $J'_n$  in Gruppe  $J'$  und  $E_n$  in Gruppe  $E$ . Die folgende Tabelle zeigt die Entwicklung über die Generationen  $n$  und  $n + 1$  zur Generation  $n + 2$  entsprechend den obigen Regeln.

Tabelle 4.2: Wachstum der Kaninchenpopulation

	$J$	$J'$	$E$
$n$	$J_n$	$J'_n$	$E_n$
$n + 1$	$J'_n + E_n$	$J_n$	$J'_n + E_n$
$n + 2$	$J_n + J'_n + E_n$	$J'_n + E_n$	$J_n + J'_n + E_n$

Die Summation  $F_{n+2}$  der Individuen in der Zeile  $n + 2$  ist gleich der Summe der Individuen in den Zeilen  $n$  und  $n + 1$ . Also gilt  $F_{n+2} = F_{n+1} + F_n$ . ▲

### Die Binet-Formel

Anders als die Rekursionsformel, die das logistische Wachstum beschreibt, lässt sich die Rekursionsformel

$$(*) \quad F_{n+2} = F_{n+1} + F_n$$

für das Fibonacci-Wachstum explizit lösen. Offensichtlich ist eine solche Lösung nach Vorgabe der Anfangswerte  $F_1$  und  $F_2$  eindeutig bestimmt. Es gibt daher ebensoviele Lösungen von (\*) wie solche Vorgaben. Zwei spezielle Lösungen von (\*) erhält man mit dem Ansatz  $F_n = \lambda^n$ . Einsetzen in (\*) ergibt

$$\lambda^{n+2} = \lambda^{n+1} + \lambda^n$$

und nach Division durch  $\lambda^n$

$$\lambda^2 = \lambda + 1$$

oder

$$\lambda^2 - \lambda - 1 = 0.$$

Das ist eine quadratische Gleichung zur Bestimmung von  $\lambda$  mit den beiden Lösungen

$$\lambda_{\pm} = \frac{1}{2}(1 \pm \sqrt{5}).$$

Hat man irgend zwei Lösungen  $F_n$  und  $G_n$  von (\*) dann ist die „Linearkombination“  $aF_n + bG_n$  ebenfalls eine Lösung von (\*), wie man durch Einsetzen unmittelbar erkennt.

Also wird man versuchen, die speziellen Lösungen  $\lambda_+^n$  und  $\lambda_-^n$  so zu einer Lösung  $F_n = a\lambda_+^n + b\lambda_-^n$  zu kombinieren, dass die Anfangsbedingungen  $F_1 = F_2 = 1$  des Kaninchenproblems erfüllt sind. Es soll also gelten

$$\begin{aligned} a\lambda_+ + b\lambda_- &= 1, \\ a\lambda_+^2 + b\lambda_-^2 &= 1. \end{aligned}$$

Das ist ein lineares Gleichungssystem mit den Unbekannten  $a$  und  $b$ . Es hat die Lösungen  $a = \frac{1}{\sqrt{5}}$  und  $b = \frac{-1}{\sqrt{5}}$ . Damit erhält man

$$F_n = \frac{1}{\sqrt{5}}\lambda_+^n - \frac{1}{\sqrt{5}}\lambda_-^n,$$

also ist

$$(**) \quad F_n = \frac{1}{\sqrt{5}} \left( \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right)$$

die gesuchte explizite Darstellung der Fibonacci-Zahlen 1, 1, 2, 3, 5, .... Die Formel (\*\*) wird nach dem französischen Mathematiker Jacques Ph. M. Binet (1786-1856), der sie 1843 gefunden haben soll, üblicherweise als Binet-Formel bezeichnet. Es gibt aber Hinweise darauf, dass diese Formel bereits früher bekannt war.

### Fibonacci-Zahlen in der Botanik

Die Natur zeigt im Reich der Pflanzen eine mehr als auffällige Vorliebe für Fibonacci-Zahlen. So z.B. bei der Anordnung der Blätter am Stengel einer Pflanze, der *Phyllotaxis*. Diese Anordnung folgt einer Schraubenlinie (Helix). Geht man nämlich von einem fest gewählten Blatt aus, dann gelangt man zum nächst höher gelegenen Blatt, wenn man dieses – in Gedanken – gegen den Uhrzeigersinn um einen bestimmten Winkel dreht und gleichzeitig um einen bestimmten Betrag anhebt. Wiederholt man diesen Vorgang, dann bewegt man sich in einer Schraubenlinie  $m$  mal um den Stengel, bis man erstmalig wieder zu einem Blatt gelangt, das genau über dem Ausgangsblatt steht. Bei dieser Bewegung werden – einschließlich Ausgangsblatt und ausschließlich Endblatt – insgesamt  $n$  Blätter passiert, eine Blatt-Periode. Die Anordnung der Blätter am Stengel wird also durch zwei Zahlen charakterisiert, die Umlaufzahl  $m$  und die Periodezahl  $n$ . Beide Zahlen sind häufig Fibonacci-Zahlen, wie die nachstehende Tabelle zeigt. Fibonacci-Zahlen treten in der Natur an zahlreichen weiteren Stellen auf.

Tabelle 4.3: Fibonacci-Zahlen bei der Phyllotaxis.

$m$	$n$	Pflanze
1	2	Knollenpflanzen, horizontale Zweige der Ulme
1	3	Riedgras, Erle, Birke
2	3	Weide, Rose, Steinobst
3	8	Kohl, Aster, Korbblütler
8	8	Anordnung von Fichten- und Tannenzapfen
13	34	Anordnung der Zapfen bei der Pinie ( <i>Pinus laricio</i> )

So z.B. bei der Anordnung der Schuppen auf der Oberfläche der Ananas und in der Genealogie der Bienen (vgl. hierzu Beutelspacher, Abschnitt 6.1.2).

### Fibonacci-Zahlen und Goldener Schnitt

Bildet man bei der Folge der Fibonacci-Zahlen 1, 1, 2, 3, 5, 8, 13, 21, 34, ... den Quotienten aufeinanderfolgender Zahlen, dann erhält man die Folge

$$1, 2, 1.5, 1.6\bar{6}, 1.6, 1.625, 1.615, 1.619, \dots$$

Diese Folge scheint sich dem Verhältnis

$$\Phi = \frac{1}{2} (1 + \sqrt{5}) = 1.6180\dots$$

des Goldenen Schnitts zu nähern. In der Tat streben die Quotienten  $Q_n = \frac{F_{n+1}}{F_n}$  für  $n \rightarrow \infty$  gegen die Zahl  $\Phi$ . Mit (\*\*\*) erhält man nämlich

$$Q_n = \frac{1}{2} \frac{(1 + \sqrt{5})^{n+1} - (1 - \sqrt{5})^{n+1}}{(1 + \sqrt{5})^n - (1 - \sqrt{5})^n}.$$

Dividiert man Zähler und Nenner durch  $(1 + \sqrt{5})^n$ , dann ergibt sich

$$Q_n = \frac{1}{2} \frac{(1 + \sqrt{5}) - (1 - \sqrt{5}) \left(\frac{1 - \sqrt{5}}{1 + \sqrt{5}}\right)^n}{1 - \left(\frac{1 - \sqrt{5}}{1 + \sqrt{5}}\right)^n}.$$

Für  $q = \frac{1 - \sqrt{5}}{1 + \sqrt{5}}$  gilt  $-1 < q < 0$ , so dass  $q^n \rightarrow 0$  geht für  $n \rightarrow \infty$ . Also geht  $Q_n \rightarrow \Phi$  für  $n \rightarrow \infty$ . Das Verhältnis aufeinanderfolgender Kaninchen-Generationen nähert sich also wundersamerweise dem Goldenen Schnitt.

Eine Strecke der Länge eins im Goldenen Schnitt zu teilen fordert, dass das größere Teilstück zum kleineren sich verhalte, wie die ganze Strecke zum größeren Teilstück.



Abbildung 4.3: Goldener Schnitt.

Ist  $x$  das größere Teilstück, dann muss also gelten

$$\frac{x}{1-x} = \frac{1}{x}$$

oder

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

Diese quadratische Gleichung hat die beiden Lösungen

$$x_{\pm} = \frac{1}{2}(\pm\sqrt{5} - 1).$$

Die positive Lösung ist

$$x = \frac{1}{2}(\sqrt{5} - 1)$$

und daher ist

$$\begin{aligned} \frac{1}{x} &= \frac{2}{\sqrt{5} - 1} \\ &= \frac{2(\sqrt{5} + 1)}{4} = \frac{1}{2}(1 + \sqrt{5}) \\ &= \Phi. \end{aligned}$$

Die explizite Darstellung (\*\*) für die  $n$ -te Fibonacci-Zahl  $F_n$  lässt sich jetzt auch mit Hilfe der Zahlen  $\Phi$  und  $\Phi^{-1}$  formulieren. Wegen  $\Phi^{-1} = -\frac{1}{2}(1 - \sqrt{5})$  gilt

$$(***) \quad F_n = \frac{1}{\sqrt{5}} (\Phi^n - (-\Phi^{-1})^n).$$

Da  $\Phi^{-1} = 0,618\dots$  ist, werden die Potenzen von  $\Phi^{-1}$  schnell klein. Man kann daher für große  $n$  den zweiten Teil von (\*\*\*) vernachlässigen und erhält die Näherung

$$(***) \quad F_n \approx \frac{\Phi^n}{\sqrt{5}}.$$

$F_n$  ist dann die der Zahl  $\Phi^n/\sqrt{5}$  am nächsten gelegene ganze Zahl.

**Bemerkung:** Das Verhältnis der Goldenen Schnitts spielt, wie bereit bemerkt, in Kunst und Architektur seit der Antike als Mittel zur Bestimmung von Maßverhältnissen eine wichtige Rolle. So bildet die Vorderfront des Parthenontempels auf der Akropolis in Athen, gebaut 447–432 v. Chr., ein goldenes Rechteck. Dieses entsteht, wenn man eine Strecke im Goldenen Schnitt teilt und die Teilstrecken die Seiten eines Rechtecks werden.



Abbildung 4.4: Goldenes Rechteck.

Auf diese Weise proportionierte Rechtecke empfindet das Auge des Betrachters als ästhetisch besonders ansprechend. Ein besonders schönes Beispiel für den Goldenen Schnitt als architektonisches Gliederungsmittel findet sich am Freiburger Münster. Die Stelle, wo die spitze Turmhaube auf dem senkrechten Turmschaft aufsitzt, teilt die gesamte Turmhöhe im Goldenen Schnitt. Ein waagrechtes Band macht diese Teilung gut kenntlich.

(Weitere Beispiele für das Auftreten des Goldenen Schnitts in Architektur und bildender Kunst finden sich bei Beutelspacher, Kap. 10.) ▲

## 4.4 Beispiele für Reihen

Der Schulklasse des jungen Carl Friedrich Gauß stellte, um die Kinder eine Weile zu beschäftigen, der Mathematiklehrer die Aufgabe, die Zahlen von eins bis hundert zu

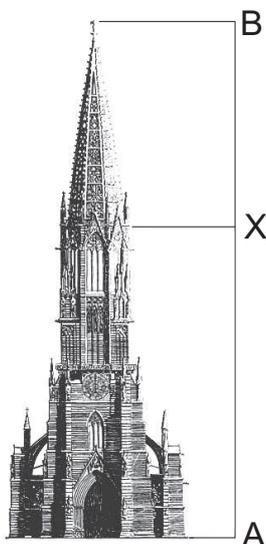


Abbildung 4.5: Goldener Schnitt am Freiburger Münster.

addieren. Der Schüler Gauß löste die gestellte Aufgabe binnen kurzem durch einen geschickten Trick: Er schrieb die gesuchte Summe zweimal hin, einmal in der natürlichen Reihenfolge der Zahlen und darunter in umgekehrter Reihenfolge.

$$\begin{aligned} S_{100} &= 1 + 2 + 3 + \dots + 99 + 100 \\ S_{100} &= 100 + 99 + 98 + \dots + 2 + 1 \end{aligned}$$

Dann addierte er beide Darstellungen und fand die Beziehung

$$\begin{aligned} 2 \cdot S_{100} &= 101 + 101 + 101 + \dots + 101 + 101 \\ &= 100 \cdot 101 \end{aligned}$$

und damit das Ergebnis

$$S_{100} = 50 \cdot 101 = 5050,$$

während seine Mitschüler noch fleißig rechneten.

### Die arithmetische Reihe

Die Gaußsche Schulaufgabe ist ein Beispiel einer *arithmetischen Reihe*. Diese entsteht, wenn man die ersten  $n$  Summanden  $x_k = a + kb$ ,  $k = 1, 2, \dots, n$ , einer arithmetischen Folge addiert:

$$\begin{aligned} A_n &= (a + b) + (a + 2b) + \dots + (a + nb) \\ &= \sum_{k=1}^n (a + kb) \end{aligned}$$

Die Gaußsche Reihe ergibt sich mit  $a = 0$  und  $b = 1$ . Mit Hilfe des obigen Tricks erhält man

$$2A_n = 2na + n(n+1)b$$

und damit die Summenformel

$$A_n = na + \frac{1}{2}n(n+1)b.$$

Für die Summe der ersten  $n$  natürlichen Zahlen findet man daher die Formel

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

Allgemein entsteht eine Reihe durch Addition der Glieder einer Folge. Bei den obigen Beispielen ist die Anzahl der Summanden endlich. Es wird sich zeigen, dass auch Reihen mit unendlich vielen Summanden sinnvoll gebildet werden und eine endliche Summe besitzen können.

### Die geometrische Reihe

Die Glieder einer geometrischen Folge haben die Form  $x_k = q^k$ ,  $k = 0, 1, 2, \dots$ , mit einer Zahl  $q \in \mathbb{R}$  (oder auch  $q \in \mathbb{C}$ ). Dann ist

$$\begin{aligned} G_n &= 1 + q + q^2 + \dots + q^{n-1} \\ &= \sum_{k=0}^{n-1} q^k \end{aligned}$$

eine geometrische Reihe mit  $n$  Summanden. Für  $q = 1$  ist  $G_n = n$ . Zur Berechnung der Summe  $G_n$  mit  $q \neq 1$  bedient man sich wieder eines Tricks: Man bildet den Ausdruck

$$qG_n = q + q^2 + \dots + q^{n-1} + q^n$$

und erhält

$$G_n - qG_n = (1 - q)G_n = 1 - q^n,$$

also ist

$$G_n = \frac{1 - q^n}{1 - q}$$

die gesuchte Summenformel für die *endliche* geometrische Reihe. Für  $q > 1$  und  $q < -1$  ergibt sich kein endlicher Wert für die Summe  $G_n$ , wenn  $n \rightarrow \infty$  geht, wohl aber im Falle  $-1 < q < 1$ . Dann gilt nämlich  $q^n \rightarrow 0$  für  $n \rightarrow \infty$ , so dass

$$G_\infty = \frac{1}{1 - q}$$

die Summenformel für die *unendliche* geometrische Reihe ist.

## Die harmonische Reihe

Die Folgenglieder  $x_k = 1/k, k = 1, 2, 3, \dots$ , bilden eine harmonische Folge, so dass

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$$

die zugehörige harmonische Reihe ist. Für die Summe  $H_n$  gibt es keinen einfachen Ausdruck. Es stellt sich aber gleichwohl die Frage, ob, wie bei der geometrischen Reihe im Fall  $-1 < q < 1$  die Summen  $H_n$  gegen einen endlichen Wert streben, wenn  $n \rightarrow \infty$  geht. Dies ist, gegen den ersten Augenschein, nicht der Fall, obwohl die Zahlen  $1/n$  beliebig klein werden, wenn nur  $n$  genügend groß ist. Vielmehr überschreiten die Summen  $H_n$  jede vorgegebene Schranke, wenn die Anzahl der Summanden nur groß genug ist. Das bedeutet anschaulich: Mit  $n$  Brettern der Länge  $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}$ , die man aneinanderlegt, kann man eine Distanz beliebiger Länge überbrücken, wenn man nur genügend viele dieser Bretter nimmt. (Wieviele Bretter man z.B. braucht, um auf diese Weise einen Steg von 100m Länge zu legen, werden wir in Abschnitt 7.4 sehen.)

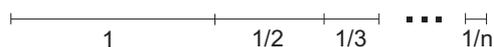


Abbildung 4.6: Aneinanderreihen von „harmonischen“ Brettern.

Um das paradoxe Verhalten der harmonischen Reihe einzusehen, betrachten wir eine Summe der Form

$$T_m = \frac{1}{m+1} + \frac{1}{m+2} + \dots + \frac{1}{m+m}.$$

Der Ausdruck  $T_m$  ist offenbar eine Teilsumme der harmonischen Reihe. Der letzte Summand von  $T_m$  ist kleiner als alle anderen. Insgesamt hat  $T_m$  aber  $m$  Summanden, so dass gilt

$$T_m > \frac{m}{2m} = \frac{1}{2}.$$

Also kann man  $H_\infty$  wie folgt zerlegen:

$$\begin{aligned} H_\infty &= 1 + \left| \frac{1}{2} \right| + \left| \frac{1}{3} + \frac{1}{4} \right| + \left| \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \right| + \dots \\ &= 1 + T_1 + T_2 + T_4 + \dots \end{aligned}$$

$H_\infty$  kann daher kein endlicher Wert sein, denn wir können folgern, dass

$$H_\infty > 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots$$

gelten muß. Die Summen  $H_n$  werden also beliebig groß.

**Bemerkung:** Verkleinert man die Glieder der harmonischen Reihe, indem man Reihen der Form

$$\begin{aligned} H_\varepsilon &= 1 + \frac{1}{2^{1+\varepsilon}} + \frac{1}{3^{1+\varepsilon}} + \dots \\ &= \sum_{k=1}^{\infty} \frac{1}{k^{1+\varepsilon}} \end{aligned}$$

bildet, dann lässt sich zeigen, dass  $H_\varepsilon$  für alle  $\varepsilon > 0$  einen endlichen Wert annimmt. Für  $\varepsilon = 1, 3, 5, \dots$  kennt man diese Werte. Im Fall  $\varepsilon = 1$  hat man z.B. das bemerkenswerte Ergebnis

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6},$$

das auf Leonhard Euler (1707–1783), den bedeutendsten Mathematiker des 18. Jahrhunderts, zurückgeht. Für die unendliche Reihe

$$\sum_{k=1}^{\infty} \frac{1}{k^3}$$

und alle übrigen ungeraden Exponenten von  $k$  kennt man keine entsprechende Formel. Im allgemeinen lässt sich der Wert einer unendlichen Reihe

$$S = \sum_{k=1}^{\infty} a_k$$

mit endlicher Summe  $S$  nicht in geschlossener Form angeben. Es gibt allerdings zahlreiche Kriterien, die zu entscheiden erlauben, ob eine solche Reihe einen endlichen Wert annimmt. ▲

## 4.5 Konvergenz

Wir hatten gesehen, dass sich die Zahlen  $1/n$  und  $q^n$ ,  $-1 < q < 1$ , dem Wert 0 beliebig dicht annähern, ohne ihn je zu erreichen. Ebenso streben die Zahlen

$$G_n = 1 + q + q^2 + \dots + q^n, \quad -1 < q < 1$$

gegen den Wert

$$G_\infty = \frac{1}{1-q}$$

ohne ihn zu erreichen. Die Werte 0 und  $1/(1-q)$  sind Beispiele für *Grenzwerte* von Zahlenfolgen. Der Begriff des Grenzwertes einer Zahlenfolge wird folgendermaßen präzisiert:

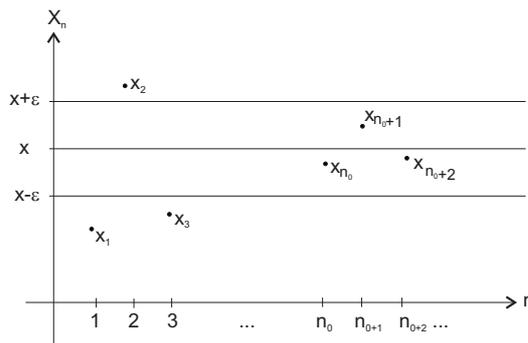


Abbildung 4.7: Konvergenz einer Folge

**Definition:** Eine Folge  $x_n, n \in \mathbb{N}$ , hat den Grenzwert  $x$ , wenn für alle  $\varepsilon > 0$  ein Index  $n_0$  existiert, so dass alle Folgenglieder  $x_n$  mit  $n \geq n_0$  im Intervall  $(x - \varepsilon, x + \varepsilon)$  liegen.

Die Pointe der obigen Definition steckt in der Formulierung „für alle  $\varepsilon > 0$ “, denn damit sind auch beliebig *kleine* Zahlen  $\varepsilon > 0$  zugelassen.

### Sprechweisen

Wenn eine Folge  $x_n$  den Grenzwert  $x$  besitzt, dann sagt man, die Folge  $x_n$  *konvergiert* gegen den Grenzwert  $x$  oder, es liegt *Konvergenz* gegen den Grenzwert  $x$  vor. Man schreibt in diesem Fall  $x_n \rightarrow x$  für  $n \rightarrow \infty$  oder

$$\lim_{n \rightarrow \infty} x_n = x$$

(gesprochen „Limes“  $x_n$ ). Die Konvergenz von Reihen führt man auf die Konvergenz von Folgen zurück. Eine Reihe

$$S = \sum_{k=1}^{\infty} x_k$$

hat die (endliche) Summe  $S$ , wenn die Folge ihrer Teilsummen (Partialsommen)

$$S_n = \sum_{k=1}^n x_k$$

gegen den Grenzwert  $S$  konvergiert:

$$S = \lim_{n \rightarrow \infty} S_n.$$

### Ein Konvergenzkriterium

Wie sieht man nun einer Folge an, ob sie konvergent ist? Um diese Frage zu beantworten, hat man *Konvergenzkriterien*. Ein wichtiges Kriterium lautet:

Eine monotone und beschränkte Folge ist konvergent.

Eine Folge  $x_n$  heißt *monoton steigend*, wenn gilt  $x_n \leq x_{n+1}$  für alle  $n \in \mathbb{N}$  und sie ist *beschränkt*, wenn es eine Zahl  $c \in \mathbb{R}$  gibt, so dass  $x_n \leq c$  für alle  $n \in \mathbb{N}$ . *Monoton fallend* und *beschränkt* ist eine Folge  $x_n$ , wenn gilt  $x_{n+1} \leq x_n$  und es eine Zahl  $c \in \mathbb{R}$  gibt, so dass gilt  $x_n \geq c$  für alle  $n \in \mathbb{N}$ .

### Ein Beispiel

Wir hatten in Abschnitt 4.1 die Zahlenfolge

$$x_n = \left(1 + \frac{1}{n}\right)^n, \quad n \in \mathbb{N},$$

betrachtet und wollen nun zeigen,  $x_n$  ist monoton steigend und beschränkt, also konvergent.

Aufgrund des binomischen Lehrsatzes (vgl. Abschnitt 3.4) können wir schreiben:

$$x_n = \sum_{k=0}^n \binom{n}{k} \frac{1}{n^k}$$

Nun ist

$$\begin{aligned} \binom{n}{k} \frac{1}{n^k} &= \frac{n(n-1)(n-2)\dots(n-k+1)}{k! n^k} \\ &= \frac{1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right)}{k!} \\ (*) \quad &< \frac{1\left(1-\frac{1}{n+1}\right)\left(1-\frac{2}{n+1}\right)\dots\left(1-\frac{k-1}{n+1}\right)}{k!} \\ &= \binom{n+1}{k} \frac{1}{(n+1)^k}. \end{aligned}$$

Also gilt

$$\begin{aligned} x_n &< \sum_{k=0}^n \binom{n+1}{k} \frac{1}{(n+1)^k} + \binom{n+1}{n+1} \frac{1}{(n+1)^{n+1}} \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} \frac{1}{(n+1)^k} \\ &= x_{n+1} \end{aligned}$$

Damit ist die Monotonie der Folge  $x_n$  gezeigt. Aus der rechten Seite der Zeile (\*) liest man ab:

$$\binom{n}{k} \frac{1}{n^k} < \frac{1}{k!}$$

Also gilt

$$x_n < \sum_{k=0}^n \frac{1}{k!} < \sum_{k=0}^{\infty} \frac{1}{k!}.$$

Nun ist aber

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{1}{k!} &= 1 + 1 + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4} + \dots \\ &< 1 + 1 + \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \dots \end{aligned}$$

und aufgrund der Formel

$$G_{\infty} = \frac{1}{1-q}, \quad -1 < q < 1,$$

für die Summe der geometrischen Reihe erhält man mit  $q = \frac{1}{2}$  die Abschätzung

$$\sum_{k=0}^{\infty} \frac{1}{k!} < 1 + \frac{1}{1 - \frac{1}{2}} = 3,$$

also gilt  $x_n < 3$ . Die Folge  $x_n = \left(1 + \frac{1}{n}\right)^n$  ist also monoton und beschränkt und daher konvergent. Ihr Grenzwert ist

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e = 2,71828\dots$$

Die Zahl  $e$  ist die Eulersche Zahl, neben der Kreiszahl  $\pi$  eine der wichtigsten Konstanten der Mathematik.

# Kapitel 5

## Elementare Funktionen

Eine Funktion beschreibt die Art der Abhängigkeit einer Größe von einer anderen. Die unabhängige Größe wird dabei häufig mit  $x$  bezeichnet, die abhängige mit  $y$ . Die Zuordnung selbst wird in der Regel mit  $f$  notiert. Man schreibt dann  $y = f(x)$  oder auch  $x \xrightarrow{f} y$ . Dabei variiert die unabhängige Variable  $x$  in ihrem Definitionsbereich

$D \subset \mathbb{R}$ , die Funktionswerte  $y = f(x)$  liegen im Wertebereich  $W \subset \mathbb{R}$  der Funktion  $f$ . Üblich ist auch die etwas saloppe Schreibweise  $y = y(x)$  ebenso wie die Rede von der „Funktion“  $f(x)$ , wo es sich streng genommen bei  $f(x)$  um den Funktionswert an der Stelle  $x$  handelt. Ist die unabhängige Variable die Zeit, dann notiert man sie mit  $t$  (von lateinisch „tempus“). So ist zum Beispiel  $x = x(t)$  der Ort eines Massenpunktes zum Zeitpunkt  $t$ . Die Menge der Punkte  $(x, f(x))$  mit  $x \in D$  in Bezug auf ein cartesisches Koordinatensystem ist der *Graph* oder das *Bild* der Funktion  $f$ .

Lässt sich das Bild einer Funktion in einem Zug, das heißt, ohne die Kreide abzusetzen, an die Tafel zeichnen, dann ist die Funktion *stetig*. Nur mit Funktionen, die über ihrem Definitionsbereich stetig sind, haben wir es im folgenden zu tun.

### 5.1 Polynome

Die Form eines Wasserstrahles der aus einem Schlauch tritt, ist die eines Zweiges einer nach unten offenen Parabel. Klemmt man einen elastischen Stab an einem Ende fest und biegt das andere Ende nach oben, so bildet er ein Profil, das Teil einer kubischen Parabel ist.

Parabel und kubische Parabel werden beschrieben durch *Polynome*. Diese bilden die einfachste Art funktionaler Abhängigkeit.

#### Die Gerade

Das einfachste Polynom mit der Gleichung

$$(*) \quad y = ax + b$$

stellt eine Gerade dar. Das bedeutet: Trägt man die Punkte mit den Koordinaten  $(x, ax + b)$  in ein Cartesisches Koordinatensystem ein, so bilden diese eine Gerade  $g$ . Setzt man

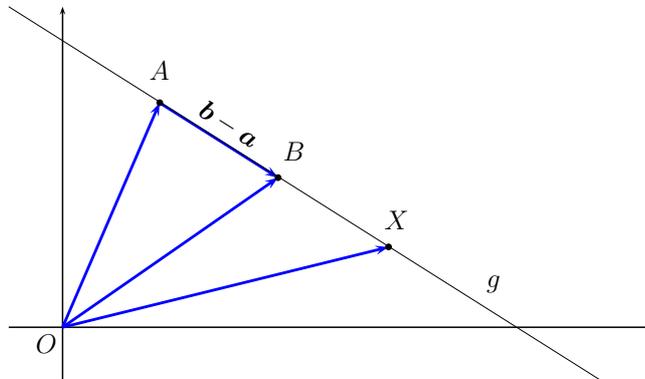


Abbildung 5.1: Gerade im Cartesischen Koordinatensystem.

$x = 0$ , dann ist  $y = b$ . Der Parameter  $b$  zeigt also an, wo die Gerade die  $y$ -Achse schneidet. Sind  $(x_1, y_1)$  und  $(x_2, y_2)$  zwei Punkte auf  $g$ , dann gilt

$$\begin{aligned} y_1 &= ax_1 + b \\ y_2 &= ax_2 + b, \end{aligned}$$

also ist

$$y_2 - y_1 = a(x_2 - x_1)$$

und daher

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

Der Parameter  $a$  gibt die *Steigung* der Geraden  $g$  an. Ist  $x_2 - x_1 = 1$ , dann ist  $y_2 - y_1 = a$ . Ein Schritt in  $x$ -Richtung bedeutet also  $a$  Schritte in  $y$ -Richtung.

Jede Gerade, die nicht parallel zur  $x$ -Achse verläuft und die nicht durch den Nullpunkt geht, kann auf die Form

$$(**) \quad \frac{x}{a} + \frac{y}{b} = 1$$

gebracht werden (der Parameter  $a$  ist hier ein anderer als bei (\*)). Das ist die *Achsenabschnittsform* der Geraden. An der Stelle  $(a, 0)$  tritt die Gerade durch die  $x$ -Achse, an der Stelle  $(0, b)$  durch die  $y$ -Achse,  $a$  und  $b$  sind daher die entsprechenden Achsenabschnitte. Die Form (\*\*\*) der Geradengleichung erlaubt daher eine schnelle Skizze des Geradenverlaufs.

## Die Parabel

Bei der Geradengleichung tritt die Variable  $x$  in der ersten Potenz auf. Eine Gleichung der Form

$$y = ax^2 + bx + c$$

mit  $a \neq 0$ , bei der die Variable  $x$  auch in der zweiten Potenz auftritt, ist die Gleichung eine *Parabel*. Ihre einfachste Form

$$y = x^2$$

mit  $a = 1$ ,  $b = c = 0$  beschreibt die *Standardparabel*. Sie ist spiegelsymmetrisch zur  $y$ -Achse, ihr tiefster Punkt oder Scheitelpunkt ist der Nullpunkt des Koordinatensystems. Die Gleichung  $y = ax^2$  beschreibt eine Parabel, die breiter ( $0 < a < 1$ ) bzw.

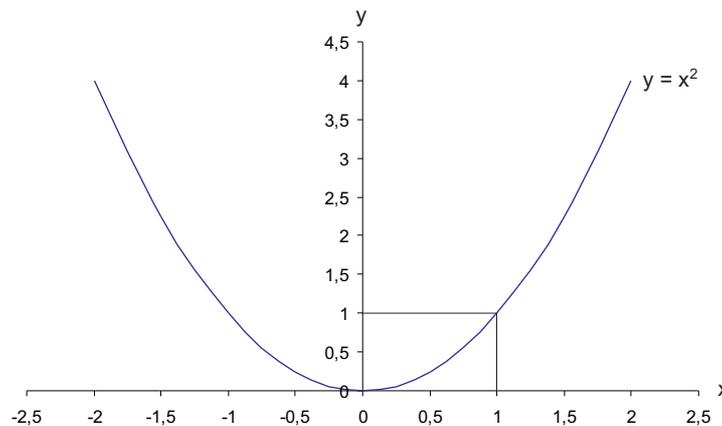


Abbildung 5.2: Standardparabel.

schmäler ( $a > 1$ ) ist, als die Standardparabel. Ist  $a < 0$ , dann ist die Parabel nach unten offen.

Wir wollen uns nun davon überzeugen, dass die Gleichung

$$y = ax^2 + bx + c$$

mit  $a > 0$  eine verschobene und in der Breite veränderte Standardparabel darstellt. Hierzu schreiben wir

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

Der Wert  $y$  erreicht sein Minimum, wenn der quadratische und damit positive Term verschwindet, also bei  $x = -\frac{b}{2a}$ . An dieser Stelle ist  $y = c - \frac{b^2}{4a}$ . Der Scheitelpunkt der Parabel liegt also im Punkt  $S = \left(-\frac{b}{2a}, c - \frac{b^2}{4a}\right)$ .

Bezüglich der Geraden  $x = -\frac{b}{2a}$  ist die Parabel spiegelsymmetrisch, denn es gilt  $y\left(-\frac{b}{2a} + x\right) = y\left(-\frac{b}{2a} - x\right)$ .

Die Parabel schneidet die  $x$ -Achse an den Stellen  $x$  für die gilt

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

Das ist eine quadratische Gleichung mit den Lösungen

$$x_{1/2} = -\frac{b}{2a} \pm \sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}}.$$

Reelle Lösungen treten genau dann auf, wenn die Bedingung

$$\left(\frac{b}{2a}\right)^2 - \frac{c}{a} \geq 0$$

erfüllt ist. Steht bei dieser Ungleichung das Gleichheitszeichen, dann liegt der Scheitelpunkt der Parabel auf der  $x$ -Achse.

**Beispiel:** Die Parabel mit der Gleichung  $y = x^2 + 2x - 1$  hat den Scheitelpunkt  $S = (-1, -2)$ . Sie geht durch die  $x$ -Achse an den Stellen

$$x_{1/2} = -1 \pm \sqrt{2}.$$

Da wir die Nullstellen des Polynoms  $y(x) = x^2 + 2x - 1$  kennen, können wir es in seine Linearfaktoren zerlegen (vgl. Fundamentalsatz der Algebra, Abschnitt 2.6) und schreiben

$$\begin{aligned} y(x) &= (x - x_1)(x - x_2) \\ &= (x + 1 - \sqrt{2})(x + 1 + \sqrt{2}). \end{aligned}$$

■

Die Standardform der *kubischen Parabel* hat die Gleichung

$$y = x^3.$$

Die kubische Parabel geht durch den Nullpunkt. Für negative  $x$  sind die Werte  $y(x)$  negativ und für positive  $x$  positiv. Der Ast über dem Intervall  $[0, 1]$  verläuft wie ein bei  $(0, 0)$  eingespanntes und bis zum Punkt  $(1, 1)$  hochgebogenes elastisches Metallband.

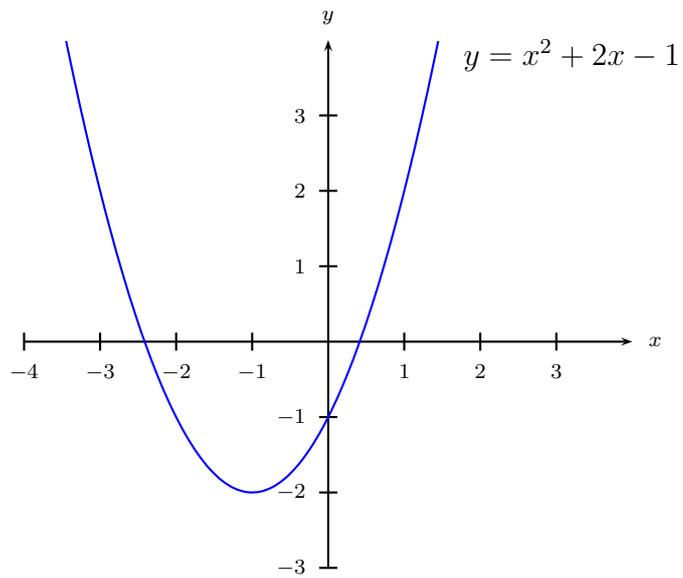
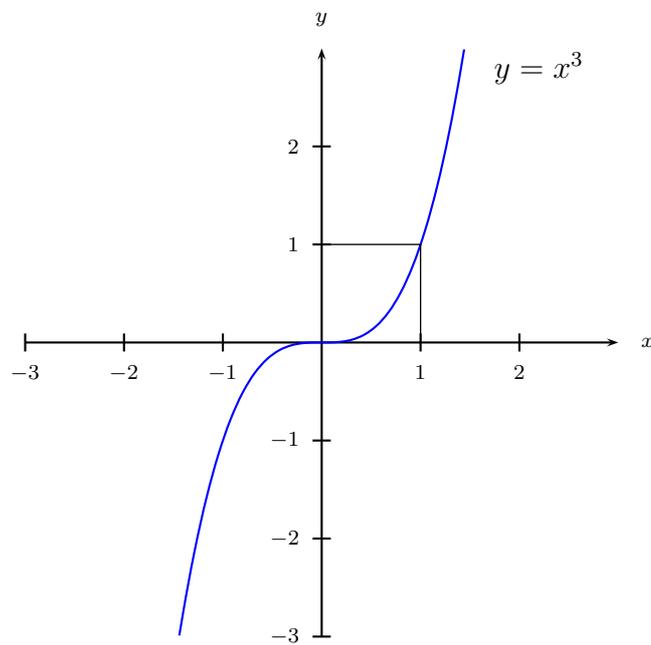
Abbildung 5.3: Die Parabel  $y = x^2 + 2x - 1$ .

Abbildung 5.4: Die kubische Parabel.

## Das Polynom $n$ -ter Ordnung

Die allgemeine Form eines Polynoms wird durch die Gleichung

$$y(x) = a_0 + a_1x + \dots + a_nx^n$$

mit Koeffizienten  $a_j \in \mathbb{R}$  (oder  $a_j \in \mathbb{C}$ ) gegeben. Ist  $a_n \neq 0$ , kommt also die  $n$ -te Potenz der Variablen  $x$  wirklich vor, dann handelt es sich um ein *Polynom  $n$ -ten Grades*. Das Verhalten eines Polynoms  $y(x)$  für große  $x$  wird vom Summanden  $a_nx^n$  mit der höchsten Potenz von  $x$  bestimmt. Ist  $a_n > 0$  und  $n$  gerade, dann gilt  $y(x) \rightarrow +\infty$  für  $x \rightarrow +\infty$  und  $y(x) \rightarrow +\infty$  für  $x \rightarrow -\infty$ . Ist  $a_n > 0$  und  $n$  ungerade, dann gilt  $y(x) \rightarrow +\infty$  für  $x \rightarrow +\infty$  aber  $y(x) \rightarrow -\infty$  für  $x \rightarrow -\infty$ . Folglich muss der Graph des Polynoms an mindestens einer Stelle die  $x$ -Achse schneiden. Aus letzterem folgt:

*Ein Polynom ungerader Ordnung hat mindestens eine reelle Nullstelle.*

Wegen des Fundamentalsatzes der Algebra (vgl. Abschnitt 2.6) können wir jedes Polynom in *Linearfaktoren* zerlegen:

$$y(x) = a_n(x - x_1)(x - x_2)\dots(x - x_n)$$

Die Zahlen  $x_j, j = 1, 2, \dots, n$ , sind die *Nullstellen* des Polynoms. Sie müssen nicht alle voneinander verschieden sein. Die Nullstellen können reell oder komplex sein. Sind die Koeffizienten  $a_j$  sämtlich reell und ist  $x_j$  eine Nullstelle von  $y(x)$ ,  $y(x_j) = 0$ , dann gilt  $\overline{y(x_j)} = y(\overline{x_j}) = 0$ . Das heißt, mit  $x_j$  ist auch  $\overline{x_j}$  eine Nullstelle von  $y(x)$ . Außer

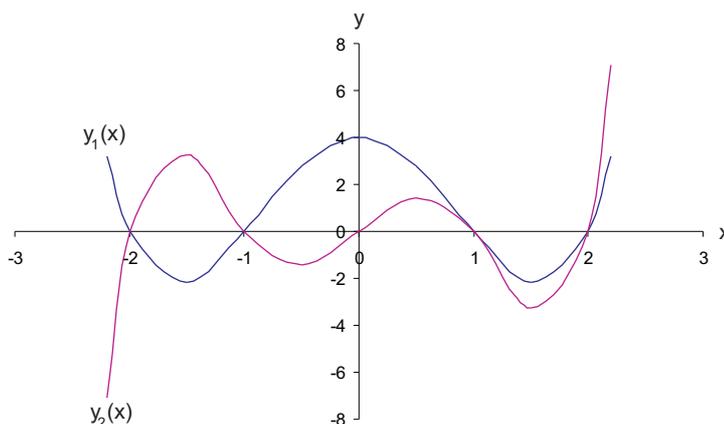


Abbildung 5.5: Die Polynome  $y_1(x) = 4 - 5x^2 + x^4$  und  $y_2(x) = 4x - 5x^3 + x^5$ .

nach den Nullstellen fragt man bei einem Polynom häufig auch nach lokalen *Maxima* und *Minima* und *Wendepunkten*. Diese Fragen werden mit Hilfe der Differentialrechnung beantwortet (vgl. Kapitel 6).

## 5.2 Rationale Funktionen

Die Polynome haben die Eigenschaft, dass ihre Werte  $y(x)$  nur dann unendlich groß werden, wenn die Variable  $x$  unendlich groß wird. Rationale Funktionen haben die Form

$$y(x) = \frac{p(x)}{q(x)},$$

wobei  $p(x)$  und  $q(x)$  Polynome sind. Ist  $q(a) = 0$  und  $p(a) \neq 0$ , dann strebt  $y(x) \rightarrow \pm\infty$ , wenn  $x \rightarrow a$  geht. Eine rationale Funktion kann also in der Nähe eines *endlichen* Wertes unendlich groß werden. Eine solche Stelle nennt man einen *Pol* der Funktion. Die einfachste rationale Funktion hat die Form

$$y(x) = \frac{1}{x}.$$

Man erkennt unmittelbar: Für  $x \rightarrow \pm\infty$  geht  $y(x) \rightarrow 0$ . Strebt  $x$  von rechts gegen null, dann strebt  $y(x) \rightarrow +\infty$ , und strebt  $x$  von links gegen null, dann strebt  $y(x) \rightarrow -\infty$ . Der Graph der Funktion  $y = 1/x$  ist eine *Hyperbel*. Die Stelle  $x = 0$  ist eine – und

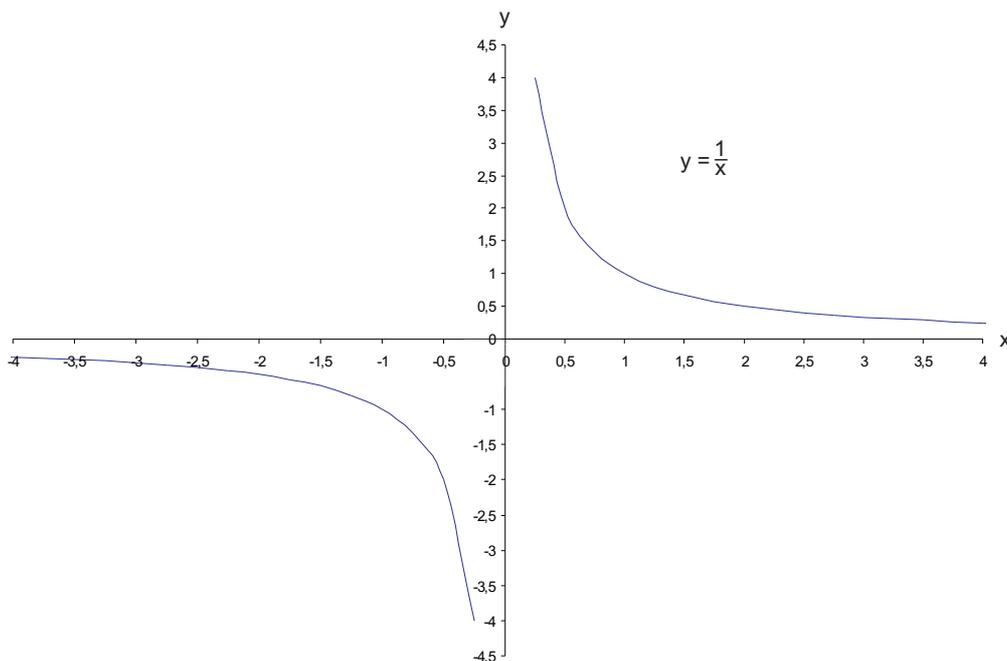


Abbildung 5.6: Die Hyperbel.

zwar die einzige – Polstelle dieser Funktion. Die Geraden  $x = 0$  und  $y = 0$ , denen sich die Funktion  $y(x)$  für  $x \rightarrow 0$  bzw.  $x \rightarrow \infty$  beliebig dicht annähert, sind sogenannte *Asymptoten* von  $y(x)$ .

## Partialbruchzerlegung

Wir betrachten die rationale Funktion

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

Die Funktion ist symmetrisch zur  $y$ -Achse, denn es gilt  $y(-x) = y(x)$ . Das Verhalten für große  $x$  kann man nach der Umformung

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

unmittelbar ablesen:  $y(x) \rightarrow 1$  für  $x \rightarrow \pm\infty$ . Das bedeutet: Die Gerade  $y = 1$  ist Asymptote der Funktion. Da  $x^2 + 1 > 0$  ist, kann  $y(x)$  keine reelle Nullstelle besitzen. Über dem Intervall  $-1 < x < 1$  gilt offenbar  $y(x) \leq -1$  mit  $y(0) = -1$  und für  $x < -1$  und  $x > 1$  ist  $y(x) > 1$ . Pole der Funktion sind bei  $x = -1$  und  $x = 1$ .

Um das Verhalten der Funktion in der Nähe der Pole besser zu verstehen, schreiben wir zunächst

$$\begin{aligned} y(x) &= \frac{x^2 - 1 + 2}{x^2 - 1} \\ &= 1 + \frac{2}{x^2 - 1}. \end{aligned}$$

Um den Term

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

weiter zu zerlegen, machen wir den Ansatz

$$\frac{2}{(x - 1)(x + 1)} = \frac{a}{x - 1} + \frac{b}{x + 1}$$

und bestimmen die Konstanten  $a$  und  $b$ . Multipliziert man diese Gleichung mit  $(x - 1)$ , dann erhält man die Beziehung

$$\frac{2}{x + 1} = a + \frac{b(x - 1)}{x + 1}.$$

Setzt man jetzt  $x = 1$ , so folgt  $a = 1$ . Multipliziert man entsprechend mit  $(x + 1)$  und setzt anschließend  $x = -1$ , so folgt  $b = -1$ . Damit erhalten wir die Zerlegung

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

die *Partialbruchzerlegung* von  $y(x)$ . Setzt man  $x = 1 + \varepsilon$  mit  $\varepsilon > 0$ , so erhält man

$$y(1 + \varepsilon) = 1 + \frac{1}{\varepsilon} - \frac{1}{2 + \varepsilon}.$$

Für  $\varepsilon \rightarrow 0$  gilt also  $y(1 + \varepsilon) \rightarrow +\infty$ . Setzt man  $x = 1 - \varepsilon$ , dann ist

$$y(1 - \varepsilon) = 1 - \frac{1}{\varepsilon} - \frac{1}{2 - \varepsilon},$$

also geht  $y(1 - \varepsilon) \rightarrow -\infty$  für  $\varepsilon \rightarrow 0$ . Entsprechende Untersuchungen an der Stelle  $x = -1$  erübrigen sich wegen der Symmetrie der Funktion zur  $y$ -Achse.

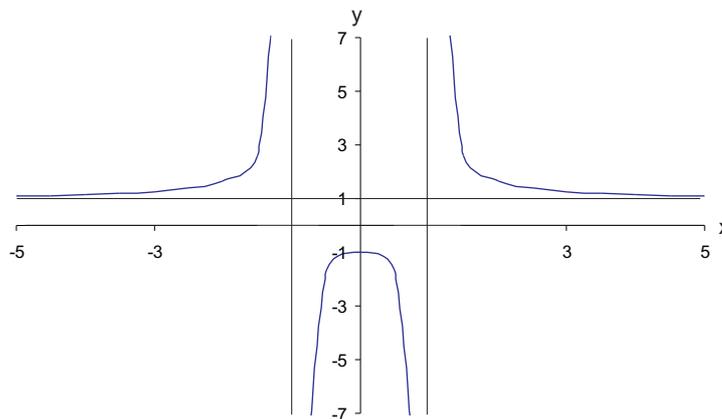


Abbildung 5.7: Die Funktion  $y(x) = (x^2 + 1)/(x^2 - 1)$ .

**Bemerkung:** Eine Partialbruchzerlegung lässt sich bei jeder rationalen Funktion  $y(x) = p(x)/q(x)$  vornehmen, sofern man die Nullstellen der Nennerfunktion  $q(x)$ , das sind die Polstellen von  $y(x)$ , kennt. Ist  $a$  eine  $m$ -fache Nullstelle von  $q(x)$ , das ist ein Pol  $m$ -ter Ordnung von  $y(x)$ , dann gehört zu dieser Stelle eine Entwicklung der Form

$$h_a(x) = \frac{a_1}{x - a} + \frac{a_2}{(x - a)^2} + \dots + \frac{a_m}{(x - a)^m}$$

mit Konstanten  $a_j, j = 1, 2, \dots, m$ . Ist der Grad des Zählerpolynoms  $p(x)$  kleiner als der Grad des Nennerpolynoms  $q(x)$ , dann hat die Partialbruchentwicklung der rationalen Funktion  $y(x) = p(x)/q(x)$  die Form

$$y(x) = \sum_{a \in N} h_a(x),$$

wobei  $N$  die Nullstellenmenge des Polynoms  $q(x)$  bezeichne. Ist der Grad des Zählerpolynoms gleich  $m$  und der Grad des Nennerpolynoms gleich  $n$  und ist  $m \geq n$ , dann tritt bei der Partialbruchentwicklung noch ein Polynom  $r(x)$  vom Grade  $m - n$  hinzu, so dass die Partialbruchentwicklung von  $y(x)$  jetzt die Form

$$y(x) = r(x) + \sum_{a \in N} h_a(x)$$

annimmt. ▲

### 5.3 Die Exponentialfunktion

Die Exponentialfunktion, die wir bereits bei der Herleitung der Euler-Formel kennengelernt hatten (vgl. Abschnitt 2.4), steht an zentraler Stelle im Gebiet der elementaren Funktionen. Sie wird definiert als Polynom unendlich hoher Ordnung, das heißt, als die unendliche Reihe

$$\begin{aligned} f(x) &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \\ &= \sum_{k=0}^{\infty} \frac{x^k}{k!}. \end{aligned}$$

Diese Reihe konvergiert für alle  $x \in \mathbb{R}$ , was mit dem schnellen Wachstum der Zahlen  $k!$  zusammenhängt.

**Bemerkung 1:** Dass eine unendliche Reihe normalerweise *nicht* für alle  $x \in \mathbb{R}$  konvergiert, erkennt man an der geometrischen Reihe

$$\begin{aligned} \frac{1}{1-x} &= 1 + x + x^2 + x^3 + \dots \\ &= \sum_{k=0}^{\infty} x^k. \end{aligned}$$

Für Werte  $x \geq 1$  wird diese Summe nämlich beliebig groß. ▲

Für  $x \geq 0$  ist der Verlauf der Funktion  $f(x)$  aus der Reihendarstellung unmittelbar abzulesen. Es gilt  $f(0) = 1$ . Auch sieht man sofort, dass  $f(x)$  streng monoton steigt, da dies bereits für die Summanden  $x^k/k!$ ,  $k \geq 1$ , richtig ist. Es gilt also  $f(x) < f(x')$  für  $x < x'$ . Also ist  $f(x) > 1$  für  $x > 0$ . Offenbar gilt  $f(x) \rightarrow +\infty$  für  $x \rightarrow +\infty$ . Das Verhalten der Funktion  $f(x)$  für  $x < 0$  erschließt sich mit Hilfe ihrer *Funktionalgleichung*. Diese besagt, dass für beliebige Werte  $x, y \in \mathbb{R}$  die Beziehung

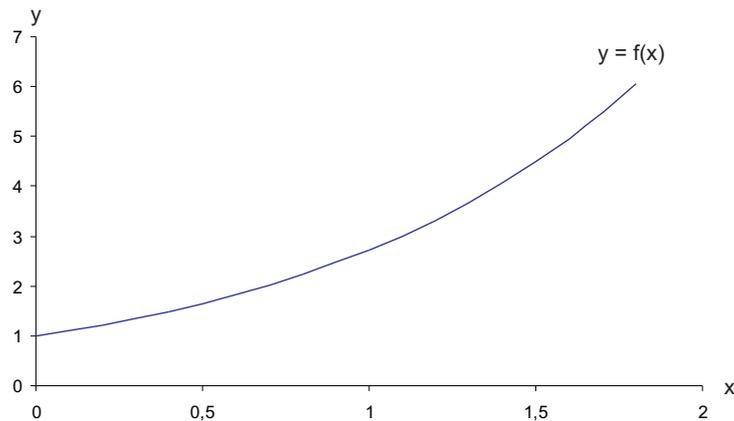
$$(*) \quad f(x+y) = f(x) \cdot f(y)$$

besteht. (Den Beweis der Funktionalgleichung tragen wir im Abschnitt 6.3 nach.) An der Stelle  $x = 1$  hat  $f(x)$  den Wert

$$\begin{aligned} f(1) &= 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \dots \\ &= e = 2,718\dots \end{aligned}$$

Die Zahl  $e$  ist die Eulersche Zahl, die wir in Abschnitt 4.5 als Grenzwert der Folge  $x_n = (1 + \frac{1}{n})^n$  kennengelernt hatten. Aufgrund der Funktionalgleichung (\*) gilt dann

$$f(1+1) = f(1) \cdot f(1) = e^2$$

Abbildung 5.8: Die Funktion  $f(x)$  für  $x \geq 0$ .

und allgemein

$$f(1 + 1 + \dots + 1) = f(n) = e^n,$$

wenn die Zahl 1  $n$ -mal in der Klammer steht. Diese Beziehung veranlasst zur Schreibweise

$$f(x) = e^x$$

für  $x \in \mathbb{R}$  überzugehen. Die Funktionalgleichung (\*) nimmt dann die Form

$$e^{x+y} = e^x e^y$$

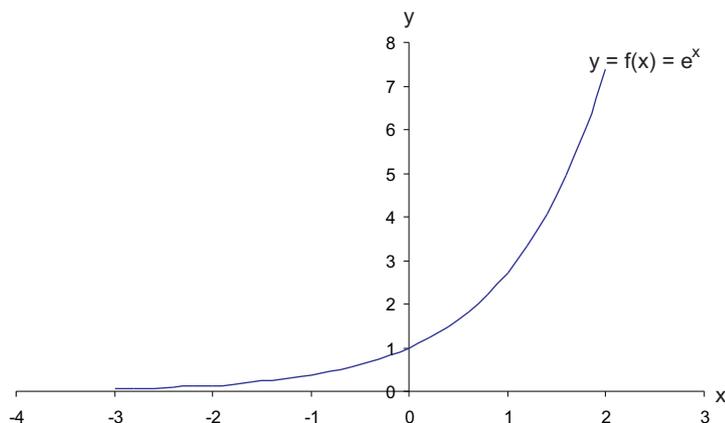
an. Mit  $x \geq 0$  und  $y = -x$  erhält man den Spezialfall

$$e^{x-x} = e^x e^{-x} = e^0 = 1.$$

Also ist

$$e^{-x} = \frac{1}{e^x}, x \geq 0.$$

Damit ist der Verlauf der Funktion  $f(x) = e^x$  auch für negative  $x$  aufgeklärt. Es gilt daher  $0 < f(x) < 1$  für  $x < 0$  und  $f(x) \rightarrow 0$  für  $x \rightarrow -\infty$ .

Abbildung 5.9: Der Verlauf der Exponentialfunktion über  $\mathbb{R}$ .

**Bemerkung 2:** Die Zahlenfolge

$$f_n(x) = \left(1 + \frac{x}{n}\right)^n$$

strebt für jedes  $x \in \mathbb{R}$  und für  $n \rightarrow \infty$  ebenfalls gegen die Funktion  $f(x) = e^x$ . ▲

**Bemerkung 3:** Die Exponentialfunktion spielt eine wichtige Rolle bei der Beschreibung von Wachstumsprozessen. Eine exponentiell wachsende Population mit der Anfangsgröße  $x_0$  zum Zeitpunkt  $t = 0$  hat zum Zeitpunkt  $t$  die Größe

$$x(t) = x_0 e^{at}.$$

Dabei ist  $a \in \mathbb{R}$  die Wachstumsrate. Ist  $a > 0$ , dann nimmt die Größe der Population mit der Zeit zu, bei  $a = 0$  bliebe  $x(t) = x_0$  zeitlich unverändert, und bei  $a < 0$  nimmt die Größe der Population in der Zeit ab und sie stirbt schließlich aus. ▲

## 5.4 Logarithmusfunktion, allgemeine Exponentialfunktion und allgemeine Potenzfunktion

Zu jedem Wert  $y > 0$  auf der  $y$ -Achse gibt es genau einen Wert  $x$  auf der  $x$ -Achse, so dass  $y = f(x)$  gilt. Diese Zuordnung  $y \rightarrow x$  wird geleistet durch die Umkehrfunktion  $f^{-1}$  der Exponentialfunktion  $f(x)$ . (Das Funktionssymbol  $f^{-1}$  ist nicht zu verstehen als  $1/f(x)$ ). Es gilt also  $f^{-1}(y) = x$ , wenn  $y = f(x)$  ist.

Die Funktion  $f^{-1}$  ist die *Logarithmusfunktion* mit der Bezeichnung  $f^{-1} = \log$ . Es gilt also  $x = \log y$ , wenn  $y = f(x) = e^x$  ist. Der Definitionsbereich der Logarithmusfunktion besteht aus den positiven Zahlen  $y > 0$ . Die Funktion  $\log$  macht gleichsam die

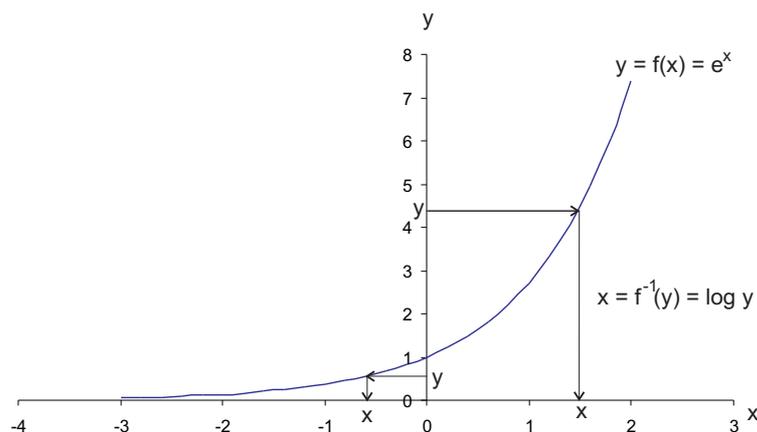


Abbildung 5.10: Die Logarithmusfunktion als Umkehrung der Exponentialfunktion.

Exponentialfunktion „rückgängig“ mit der Beziehung

$$\log e^x = x.$$

Umgekehrt macht die Exponentialfunktion die Logarithmusfunktion rückgängig, denn es gilt ebenso

$$e^{\log y} = y$$

für alle  $y > 0$ .

Bezieht man, wie üblich, die Funktion  $\log$  auf Werte der  $x$ -Achse, dann ist die Funktion nur im Bereich  $x > 0$  erklärt. Ihr Verlauf ergibt sich durch Spiegelung des Bildes der Exponentialfunktion an der Geraden  $y = x$ , der Winkelhalbierenden des 1. Quadranten. Für  $0 < x < 1$  ist  $\log x < 0$  und es gilt  $\log x \rightarrow -\infty$  für  $x \rightarrow 0$ . Offenbar gilt  $\log(1) = 0$  und  $\log(x) > 0$  für  $x > 1$ . Die Funktion  $\log x$  wächst für große  $x$  ebenso wie  $e^x$  über alle Grenzen, allerdings signifikant langsamer (vgl. dazu Abschnitt 6.3). Gäbe es nämlich eine Schranke  $c > 0$ , so dass  $\log x < c$  gilt für  $x > 1$ , dann würde folgen  $e^{\log x} = x < e^c$ , was offenbar falsch ist.

**Bemerkung:** Die Umkehrfunktion der Exponentialfunktion wird auch als der *natürliche Logarithmus* bezeichnet, abgekürzt  $\ln x =$  „logarithmus naturalis“ von  $x$ . Diese Schreibweise ist insbesondere in der Physik und überhaupt in den Naturwissenschaften üblich. ▲

Mit der Einführung der Logarithmusfunktion als Umkehrfunktion der Exponentialfunktion ist zunächst noch kein Rechenausdruck verbunden, der es ermöglicht, konkrete Funktionswerte, wie  $\log 2$ , zu berechnen. Entsprechende Berechnungsverfahren werden wir in den Abschnitten 7.4 und 7.5 kennenlernen.

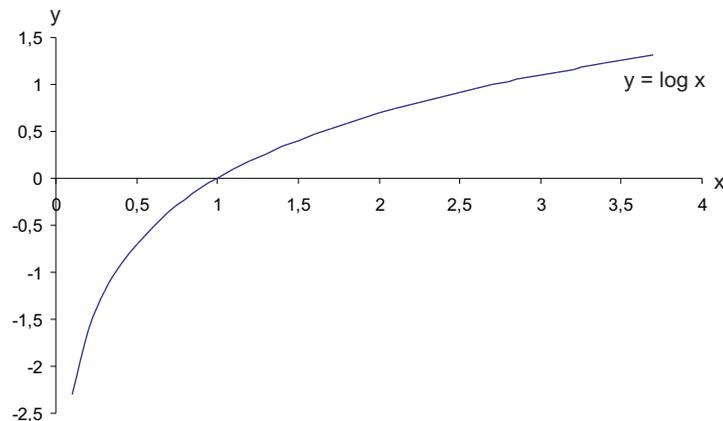


Abbildung 5.11: Verlauf der Logarithmusfunktion.

Allerdings ergibt sich aus der Darstellung

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$$

der Exponentialfunktion als Grenzwert einer Funktionenfolge (vgl. Bemerkung 1, Abschnitt 5.3) eine ebensolche Darstellung für die Logarithmusfunktion. Setzt man nämlich

$$y = \left(1 + \frac{x}{n}\right)^n$$

und löst nach  $x$  auf, dann findet man die Beziehung

$$x = n\left(y^{\frac{1}{n}} - 1\right).$$

Vertauscht man nun  $x$  und  $y$  dann folgt

$$\log x = \lim_{n \rightarrow \infty} n\left(x^{\frac{1}{n}} - 1\right).$$

Damit wird die Berechnung der Werte  $\log x$ ,  $x > 0$ , im wesentlichen auf die Bestimmung der  $n$ -ten Wurzel  $x^{1/n}$  von  $x$  zurückgeführt.

## Die Funktionalgleichung der Logarithmusfunktion

Die Funktionalgleichung

$$e^{x+y} = e^x e^y$$

überträgt sich auf die Logarithmusfunktion in folgender Weise. Mit Werten  $x, y > 0$  kann man schreiben

$$\begin{aligned} \log(x \cdot y) &= \log(e^{\log x} \cdot e^{\log y}) \\ &= \log(e^{\log x + \log y}) \\ &= \log x + \log y. \end{aligned}$$

Also erfüllt die Logarithmusfunktion die Funktionalgleichung

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

Kennt man zum Beispiel die Werte  $\log 2$  und  $\log 3$ , dann ist wegen

$$\log 6 = \log(2 \cdot 3) = \log 2 + \log 3$$

der Wert  $\log 6$  berechenbar. Wegen  $\log 1 = 0$  hat man

$$\begin{aligned} \log \frac{x}{x} &= \log \left( x \cdot \frac{1}{x} \right) \\ &= \log x + \log \frac{1}{x} = 0. \end{aligned}$$

Folglich ist

$$\log \frac{1}{x} = -\log x.$$

Daher folgt

$$\begin{aligned} \log \frac{x}{y} &= \log \left( x \cdot \frac{1}{y} \right) \\ &= \log x + \log \frac{1}{y} \\ &= \log x - \log y. \end{aligned}$$

### Die allgemeine Exponentialfunktion

Für jede Zahl  $a \geq 0$  ist die Bedeutung der Potenz  $a^n = a \cdot a \cdot \dots \cdot a$  als  $n$ -maliges Produkt der Zahl  $a$  mit sich selbst geläufig. Was aber bedeutet die zum Beispiel Potenz  $a^{\sqrt{2}}$ ? Diese Frage beantwortet die Festsetzung (Definition) der allgemeinen Potenzfunktion. Für eine Zahl  $a > 0$  und alle  $x \in \mathbb{R}$  ist

$$(*) \quad a^x = e^{x \log a}.$$

Für  $x = n \in \mathbb{N}$  ergibt diese Definition

$$\begin{aligned} a^n &= e^{n \log a} \\ &= e^{\log a} e^{\log a} \dots e^{\log a} \quad (n\text{-mal}) \\ &= a \cdot a \cdot \dots \cdot a = a^n. \end{aligned}$$

Sie stimmt also für natürliche Zahlen mit der üblichen Bedeutung von  $a^n$  überein. Ist  $a = e$ , dann ist wegen  $\log e = 1$  wieder  $a^x = e^x$ .

Die allgemeine Potenzfunktion erfüllt die Funktionalgleichung

$$a^{x+y} = a^x \cdot a^y.$$

Nach Definition (\*) kann man nämlich schreiben:

$$\begin{aligned} a^{x+y} &= e^{(x+y) \log a} \\ &= e^{x \log a} \cdot e^{y \log a} \\ &= a^x \cdot a^y. \end{aligned}$$

Außerdem gilt für  $x, y \in \mathbb{R}$  die Regel

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

Wegen (\*) hat man nämlich

$$\begin{aligned} (a^x)^y &= e^{y \log a^x} \\ &= e^{y \log e^{x \log a}} \\ &= e^{y \cdot x \log a} \\ &= a^{x \cdot y}. \end{aligned}$$

Der Verlauf der Funktion  $y(x) = a^x$  ist qualitativ verschieden, je nachdem, ob gilt  $0 < a < 1$ ,  $a = 1$  oder  $a > 1$ . Im ersten Fall ist  $y(x)$  streng monoton fallend von  $+\infty$  bis 0, im zweiten Fall ist  $y(x) = 1$ , im dritten Fall steigt  $y(x)$  streng monoton von 0 nach  $+\infty$ . In allen Fällen gilt  $y(0) = 1$ .

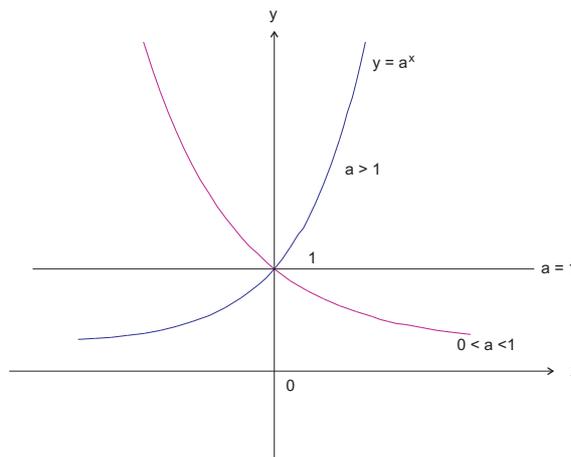


Abbildung 5.12: Der Verlauf der allgemeinen Exponentialfunktion.

## Die allgemeine Logarithmusfunktion

Da die Funktion  $y(x) = a^x$  für  $a > 0$  und  $a \neq 1$  immer streng monoton verläuft, gehört zu ihr immer eine Umkehrfunktion. Die Umkehrfunktion zur Funktion  $f(x) = a^x$  ist die Funktion

$$f^{-1}(x) = \log_a(x),$$

die Logarithmusfunktion zur Basis  $a$ . Für  $a = 10$  und  $a = 2$  hat man auch die Schreibweisen  $\lg(x)$  und  $\text{ld}(x)$ , das ist der logarithmus dualis. Die Funktion  $\log_a(x)$  erfüllt im übrigen dieselbe Funktionalgleichung wie die Funktion  $\log(x) = \ln x$ .

### Die allgemeine Potenzfunktion

Die Potenzen  $x^n, n \in \mathbb{N}$ , lassen sich nun mit Hilfe der Logarithmusfunktion zu der allgemeinen Potenzfunktion  $f(x) = x^\alpha$  mit  $\alpha \in \mathbb{R}$  verallgemeinern indem man setzt

$$f(x) = x^\alpha = e^{\alpha \cdot \log x}, \quad x > 0.$$

Da die Funktion  $\log x$  nur für positive Werte von  $x$  erklärt ist, überträgt sich diese Einschränkung des Definitionsbereichs auf die allgemeine Potenzfunktion. Dies gilt natürlich nicht für  $\alpha = n \in \mathbb{N}$ . In diesem Falle hat man nämlich wegen

$$f(x) = e^{n \cdot \log x} = (e^{\log x})^n = x^n$$

wieder die vertraute Potenzfunktion. Man kann also jetzt eine Potenz

$$\begin{aligned} f(x) &= x^{\sqrt{2}} = e^{\sqrt{2} \cdot \log x} \\ &= \sum_{k=0}^{\infty} \frac{(\sqrt{2} \cdot \log x)^k}{k!} \end{aligned}$$

bilden und mit Hilfe der Reihe für die Exponentialfunktion mit beliebiger Genauigkeit berechnen, sofern  $\log x$  gegeben ist. Der Verlauf der Funktionen  $f(x) = x^\alpha, \alpha > 0$ , entspricht qualitativ demjenigen der Potenzen  $f(x) = x^n$ . So verläuft wegen  $1 < \sqrt{2} < 2$  der Graph von  $f(x) = x^{\sqrt{2}}$  zwischen dem der Funktionen  $f(x) = x$  und  $f(x) = x^2$ . Für  $\alpha > 0$  gilt außerdem

$$x^\alpha = e^{\alpha \log x} \rightarrow 0,$$

für  $x \rightarrow 0$  da  $\log x \rightarrow -\infty$  geht für  $x \rightarrow 0$ . Für  $\alpha > 0$  ist die Funktion  $f(x) = x^\alpha$  also für  $x \geq 0$  definiert und es gilt  $f(0) = 0$ . Wegen  $\log 1 = 0$  hat man für alle  $\alpha \in \mathbb{R}$  immer  $f(1) = 1$ . Für  $x \rightarrow \infty$  gilt offenbar  $x^\alpha \rightarrow \infty$ . Der Verlauf von  $f(x)x^{-\alpha}, \alpha > 0$ , wird durch Bildung des Kehrwertes  $x^{-\alpha} = 1/x^\alpha$  gewonnen.

## 5.5 Die trigonometrischen Funktionen

Die trigonometrischen Funktionen entstehen aus der Exponentialfunktion  $f(x) = e^x$ , indem man die Variable  $x$  durch  $ix, x \in \mathbb{R}$ , ersetzt. Mit der Definition der Exponentialfunktion als unendliche Reihe aus Abschnitt 5.2. erhält man dann – wie bei der

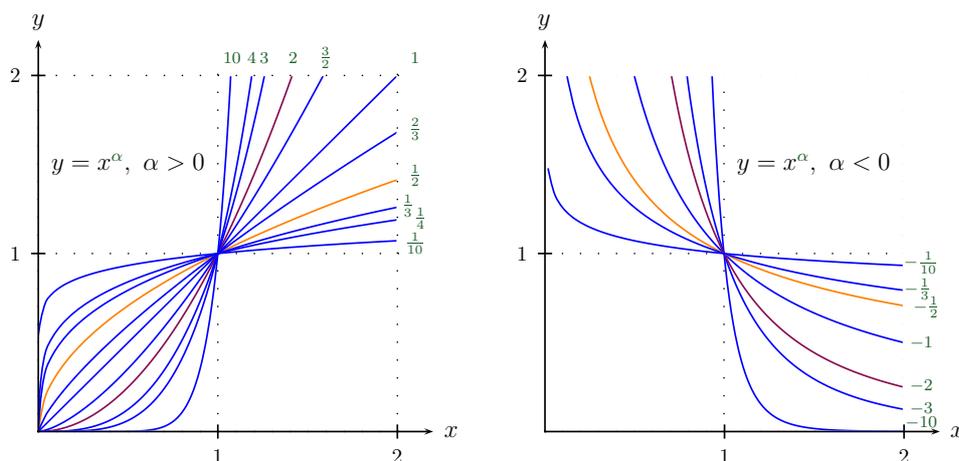


Abbildung 5.13: Der Verlauf der allgemeinen Potenzfunktion.

Herleitung der Euler-Formel in Abschnitt 2.4 – rein formal die Entwicklung

$$\begin{aligned} e^{ix} &= \sum_{k=0}^{\infty} \frac{(ix)^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!} + i \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}. \end{aligned}$$

Die durch Trennung von Realteil und Imaginärteil bei  $e^{ix}$  entstandenen Reihen sind für alle  $x \in \mathbb{R}$  erklärt. Sie *definieren* die Funktion  $\sin x$  und  $\cos x$ :

$$(*) \quad \cos x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!},$$

$$(**) \quad \sin x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$$

Es gilt also die *Euler-Formel*

$$e^{ix} = \cos x + i \sin x.$$

(vgl. hierzu die Abbildungen 2.8 und 2.9.)

Ersetzt man bei der Euler-Formel  $i$  durch  $-i$ , dann erhält man

$$e^{-ix} = \cos x - i \sin x.$$

Durch Addition bzw. Subtraktion dieser Varianten der Euler-Formel folgen dann die Darstellungen

$$\begin{aligned} (***) \quad \cos x &= \frac{1}{2}(e^{ix} + e^{-ix}), \\ \sin x &= \frac{1}{2}(e^{ix} - e^{-ix}), \end{aligned}$$

der Funktionen  $\sin x$  und  $\cos x$ , die gelegentlich von Nutzen sind.

### Die geometrische Bedeutung des Parameters $x$

Die komplexe Zahl  $u(x) = e^{ix}$  liegt auf dem Einheitskreis, denn aufgrund der Funktionalgleichung der Exponentialfunktion erhält man

$$u(x) \overline{u(x)} = e^{ix} \cdot e^{-ix} = e^{ix-ix} = e^0 = 1.$$

Wir wollen zeigen: Die Zahl  $x$  ist gerade die Länge des Kreisbogens zwischen den Punkten 1 und  $e^{ix}$ .

Das bedeutet: Nimmt man einen Faden der Länge  $x$ , befestigt das eine Ende im Punkt 1 und wickelt den Faden „nach oben“ auf dem Kreis ab, so markiert das andere Ende den Punkt  $e^{ix}$ . Um dies zu zeigen, teilen wir das Intervall  $[0, x]$  in  $n$  gleiche Teile der Länge  $\frac{x}{n}$  mit den Teilpunkten  $x_j = \frac{jx}{n}$ ,  $j = 1, 2, \dots, n-1$ .

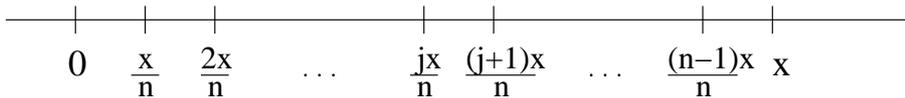


Abbildung 5.14: Intervallunterteilung.

Die Punkte  $u_j = u(x_j)$  liegen auf dem Bogen zwischen 1 und  $e^{ix}$ .

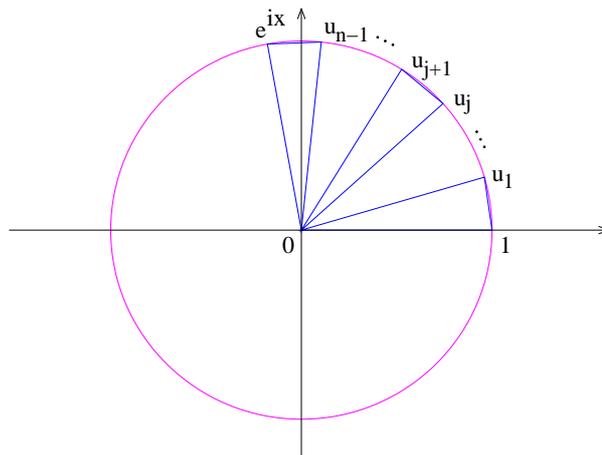


Abbildung 5.15: Unterteilung des Bogens zwischen 1 und  $e^{ix}$ .

Die Sekante  $s_j = u_{j+1} - u_j$ ,  $j = 0, 1, \dots, n-1$ , verbindet die Punkte  $u_j$  und  $u_{j+1}$

auf dem Einheitskreis. Man erhält

$$\begin{aligned} s_j &= e^{(j+1)i\frac{x}{n}} - e^{ji\frac{x}{n}} \\ &= e^{ij\frac{x}{n}} (e^{i\frac{x}{n}} - 1). \end{aligned}$$

Also gilt

$$\begin{aligned} |s_j|^2 &= s_j \bar{s}_j = e^{ij\frac{x}{n}} (e^{i\frac{x}{n}} - 1) e^{-ij\frac{x}{n}} (e^{-i\frac{x}{n}} - 1) \\ &= (e^{i\frac{x}{n}} - 1) (e^{-i\frac{x}{n}} - 1) \\ &= 2 - (e^{i\frac{x}{n}} + e^{-i\frac{x}{n}}). \end{aligned}$$

und mit (\*\*\*) erhält man schließlich

$$|s_j|^2 = 2 - 2 \cos \frac{x}{n}.$$

Die Sekanten  $s_j$  haben also alle die gleiche Länge  $|s_j| = s$ . Ihre Summe  $ns$  approximiert daher den Kreisbogen zwischen den Punkten 1 und  $e^{ix}$ . Man erhält

$$\begin{aligned} ns &= n \sqrt{2 - 2 \cos \frac{x}{n}} \\ &= \sqrt{2n^2 (1 - \cos \frac{x}{n})}. \end{aligned}$$

Die Reihendarstellung (\*) der  $\cos$ -Funktion ergibt nun

$$\begin{aligned} 2n^2 \left(1 - \cos \frac{x}{n}\right) &= \\ &= 2n^2 \left(1 - 1 + \frac{1}{2!} \left(\frac{x}{n}\right)^2 - \frac{1}{4!} \left(\frac{x}{n}\right)^4 + \frac{1}{6!} \left(\frac{x}{n}\right)^6 - \dots\right) \\ &= \left(x^2 - \frac{1}{12} \frac{x^4}{n^2} + \frac{1}{360} \frac{x^6}{n^4} - \dots\right) \end{aligned}$$

Für  $n \rightarrow \infty$  strebt der letzte Ausdruck gegen  $x^2$  und wir erhalten  $ns \rightarrow \sqrt{x^2} = x$  für  $n \rightarrow \infty$ .

Damit ist gezeigt: Der Bogen zwischen den Punkten 1 und  $e^{ix}$  hat die Länge  $x$ .

Die Abbildung  $x \rightarrow e^{ix}$  kann also gedeutet werden als Abwicklung eines Fadens der Länge  $x$ , dessen eines Ende im Punkt 1 befestigt wird, auf dem Einheitskreis. Das freie Ende des Fadens markiert dann den Punkt  $e^{ix}$ .

Da der Kreis mit Radius 1 den Umfang  $2\pi$  besitzt, der Halbkreis daher ein Bogen der Länge  $\pi$  und der Viertelkreis einen Bogen der Länge  $\frac{\pi}{2}$  ist, muss also gelten

$$e^{i\frac{\pi}{2}} = i$$

und

$$e^{i\pi} = -1$$

oder

$$e^{i\pi} + 1 = 0.$$

Letztere Formel ist die bereits aus Abschnitt 2.4 bekannte *magische Formel*, die die „Naturkonstanten“ der Mathematik, die Zahlen  $0, 1, e, \pi$  und  $i$  zu einer einzigen Beziehung verbindet.

### Die Funktionen Sinus und Cosinus

Da der Einheitskreis den Umfang  $2\pi$  hat, gelangt man bei Abwicklung eines Fadens der Länge  $x + 2\pi$  zum gleichen Punkt wie beim Faden der Länge  $x$ .

Also gilt

$$e^{i(x+2\pi)} = e^{ix}$$

und in der Folge

$$\sin(x + 2\pi) = \sin x$$

und

$$\cos(x + 2\pi) = \cos x.$$

Das heißt, die Funktionen  $\sin x$  und  $\cos x$  sind *periodisch* mit der Periode  $2\pi$ . Es genügt daher, das Verhalten von  $\sin x$  und  $\cos x$  im Intervall  $0 \leq x \leq 2\pi$  zu untersuchen.

Wegen

$$e^{ix} = \cos x + i \sin x$$

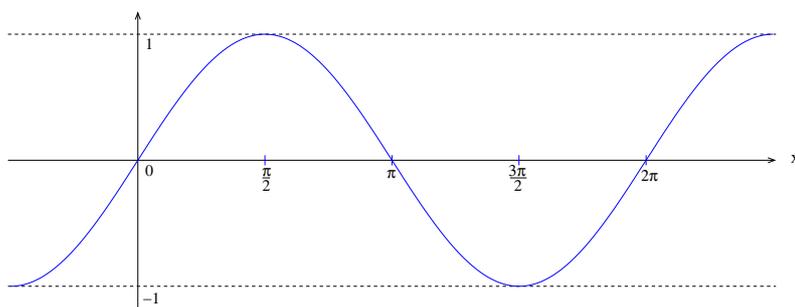
und  $|e^{ix}| = 1$  gilt

$$\sin^2 x + \cos^2 x = 1.$$

(Dies ergibt sich auch aus den Darstellungen (\*\*\*)).

Der geometrischen Interpretation am Einheitskreis bzw. den Reihendarstellungen (\*) und (\*\*) entnimmt man die Eigenschaften

$$\cos(-x) = \cos x$$

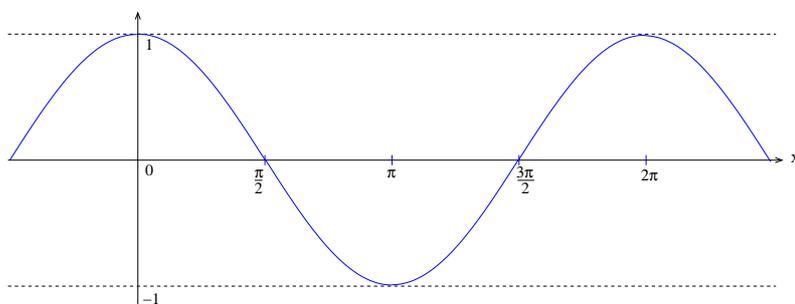
Abbildung 5.16: Verlauf der Funktion  $\sin x$ .

und

$$\sin(-x) = -\sin x.$$

Das heißt,  $\cos x$  ist eine *gerade* und  $\sin x$  eine *ungerade* Funktion. Den Werteverlauf der Funktionen  $\sin x$  und  $\cos x$  liest man aus Abb. 2.9 zur Euler-Formel ab.

Die Funktion  $\sin x$  hat also im Intervall  $[0, 2\pi]$  die Nullstellen  $0, \pi$  und  $2\pi$ , auf der ganzen  $x$ -Achse aufgrund der Periodizität daher Nullstellen bei den ganzzahligen Vielfachen  $x = k\pi$ ,  $k \in \mathbb{Z}$ , von  $\pi$ .

Abbildung 5.17: Verlauf der Funktion  $\cos x$ .

Die Nullstellen der Funktion  $\cos x$  liegen bei den Werten  $x = \frac{\pi}{2} + k\pi$ ,  $k \in \mathbb{Z}$ . Die Bilder der Funktionen  $\sin x$  und  $\cos x$  gehen durch eine Verschiebung um  $\pi/2$  ineinander über. Mit Hilfe einer geometrischen Überlegung am Einheitskreis zeigt sich nämlich unmittelbar die Beziehung

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

## Die Additionstheoreme für $\sin x$ und $\cos x$

Aufgrund der Funktionalgleichung

$$e^{x+y} = e^x \cdot e^y$$

der Exponentialfunktion gilt

$$e^{i(x+x')} = e^{ix} \cdot e^{ix'}$$

und wegen der Euler-Formel folgt dann

$$\begin{aligned} \cos(x+x') + i \sin(x+x') &= (\cos x + i \sin x)(\cos x' + i \sin x') \\ &= (\cos x \cos x' - \sin x \sin x') + i(\sin x \cos x' + \cos x \sin x'). \end{aligned}$$

Zwei komplexe Zahlen sind genau dann gleich, wenn sie im Realteil und im Imaginärteil übereinstimmen. Der Vergleich von linker und rechter Seite obiger Gleichung führt daher zu den Beziehungen

$$\cos(x+x') = \cos x \cos x' - \sin x \sin x'$$

und

$$\sin(x+x') = \sin x \cos x' + \cos x \sin x'.$$

Das sind die *Additionstheoreme* der Sinus- und Cosinusfunktion.

## Sinus und Cosinus mehrfacher Winkel

Aus der Beziehung

$$(e^{ix})^n = e^{inx}$$

für die Exponentialfunktion lassen sich bemerkenswerte Formeln herleiten: Einerseits gilt wegen der Euler-Formel

$$(*) \quad e^{inx} = \cos nx + i \sin nx,$$

andererseits aufgrund des binomischen Lehrsatzes

$$\begin{aligned} (e^{ix})^n &= (\cos x + i \sin x)^n \\ &= \sum_{k=0}^n \binom{n}{k} (i \sin x)^k (\cos x)^{n-k} \\ (**) \quad &= \sum_{0 \leq 2k \leq n} (-1)^k \binom{n}{2k} (\sin x)^{2k} (\cos x)^{n-2k} \\ &\quad + i \sum_{0 \leq 2k+1 \leq n} (-1)^k \binom{n}{2k+1} (\sin x)^{2k+1} (\cos x)^{n-2k-1}. \end{aligned}$$

Da die linken Seiten bei (\*) und (\*\*) übereinstimmen, muss dies auch für die rechten Seiten gelten. Daher erhält man durch Vergleich der rechten Seiten die Formeln

$$\begin{aligned}\cos nx &= \sum_{0 \leq 2k \leq n} (-1)^k \binom{n}{2k} (\sin x)^{2k} (\cos x)^{n-2k} \\ \sin nx &= \sum_{0 \leq 2k+1 \leq n} (-1)^k \binom{n}{2k+1} (\sin x)^{2k+1} (\cos x)^{n-2k-1}.\end{aligned}$$

Diese ermöglichen es, Sinus und Cosinus mehrfacher Winkel durch Sinus und Cosinus des einfachen Winkels auszudrücken. So erhält man z.B.

$$\begin{aligned}\sin 3x &= \binom{3}{1} \sin x \cos^2 x - \binom{3}{3} \sin^3 x \\ &= 3 \sin x \cos^2 x - \sin^3 x.\end{aligned}$$

### Kinematische Bedeutung von $\sin t$ und $\cos t$

Deutet man die Variable  $t$  als Zeit, dann beschreibt die Funktion  $t \rightarrow e^{it}$  den Umlauf eines Punktes auf dem Einheitskreis mit gleichförmiger Geschwindigkeit eins. Ist die Zeiteinheit eine Sekunde, dann legt der Punkt in  $t$  Sekunden auf dem Einheitskreis die Strecke (den Bogen) der Länge  $t$  zurück. Ein ganzer Umlauf braucht  $2\pi$  Sekunden.

Projiziert man die Umlaufbewegung des Punktes auf dem Einheitskreis auf die  $x$ -Achse bzw. die  $y$ -Achse der komplexen Ebene, dann sieht man einen Punkt, den Schatten des umlaufenden Punktes, der zwischen den Werten  $-1$  und  $+1$  bzw.  $-i$  und  $+i$  hin und her schwingt. Diese Schwingung wird in der Physik als *harmonische Schwingung* bezeichnet (vgl. Abschnitt 9.2). Sie wird also durch die Funktionen  $\sin t$  und  $\cos t$  beschrieben.

### Die Funktionen Tangens und Cotangens

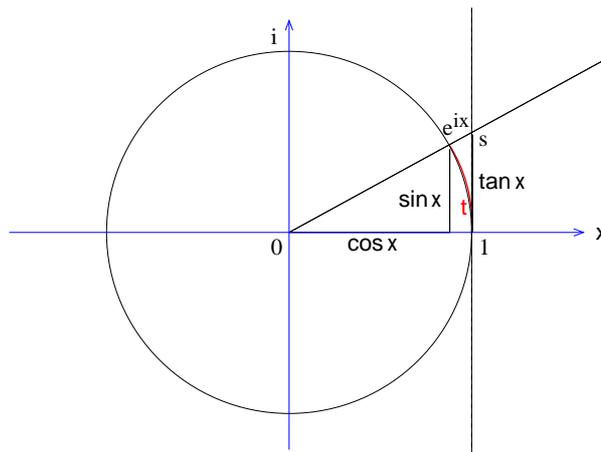
Zum Bestand der trigonometrischen Funktionen gehören auch die mittelbar durch  $\sin x$  und  $\cos x$  definierten Funktionen

$$\tan x = \frac{\sin x}{\cos x},$$

gesprochen „Tangens- $x$ “, und

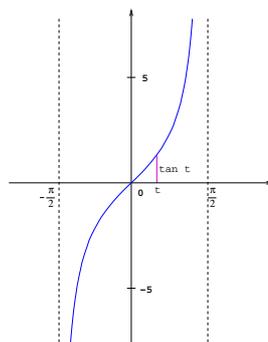
$$\cot x = \frac{\cos x}{\sin x} = \frac{1}{\tan x},$$

gesprochen „Cotangens- $x$ “. Der Verlauf der Funktion  $\tan x$  für  $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$  macht man sich leicht am Einheitskreis klar.

Abbildung 5.18: Geometrische Deutung der Funktion  $\tan x$ .

Zeichnet man im Punkt 1 die Tangente an den Einheitskreis und schneidet diese mit der Geraden durch die Punkte 0 und  $e^{ix}$  im Punkte  $S$ , dann ist der Funktionswert  $\tan x$  aufgrund des Strahlensatzes gleich der mit Vorzeichen  $\pm$  behafteten Länge des Tangentenabschnitts zwischen den Punkten 1 und  $S$ . Der Funktionswert  $\tan x$  bewegt sich also zwischen  $-\infty$  und  $+\infty$ , wenn  $x$  von  $-\frac{\pi}{2}$  nach  $+\frac{\pi}{2}$  geht. Für  $x = 0$  ist  $\tan 0 = 0$ .

Die Funktion  $\tan x$  ist periodisch mit der Periode  $\pi$ , das heißt es gilt

Abbildung 5.19: Verlauf der Funktion  $\tan x$ .

$$\tan(x + \pi) = \tan x.$$

Dies folgt zum Beispiel mit Hilfe der Additionstheoreme für  $\sin x$  und  $\cos x$ :

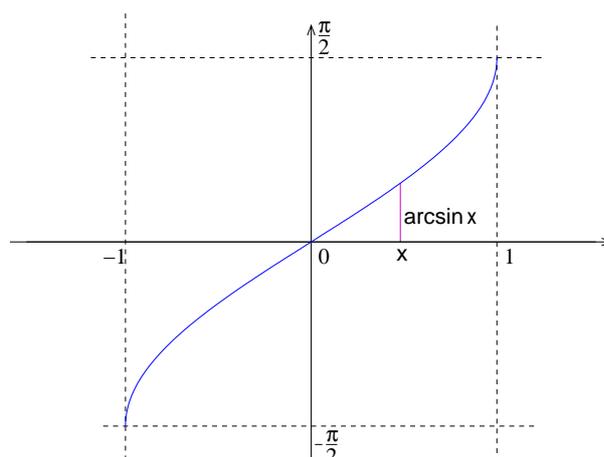
$$\begin{aligned}\tan(x + \pi) &= \frac{\sin(x + \pi)}{\cos(x + \pi)} \\ &= \frac{\sin x \cos \pi + \cos x \sin \pi}{\cos x \sin \pi - \sin x \cos \pi} \\ &= \frac{\sin x \cdot (-1) + \cos x \cdot 0}{\cos x \cdot (-1) - \sin x \cdot 0} \\ &= \frac{\sin x}{\cos x} = \tan x\end{aligned}$$

Damit ist der Verlauf der Funktion  $\tan x$  auf ganz  $\mathbb{R}$  erklärt.

Die Werte der Funktion  $\cot x = 1/\tan x$  ergeben sich durch eine analoge geometrische Überlegung wie bei der Funktion  $\tan x$ : Man lege eine Tangente an den Einheitskreis im Punkt  $i$  und betrachte den Schnittpunkt der Tangente mit der Geraden durch die Punkte  $0$  und  $e^{ix}$ . Auf diese Weise liest man den Verlauf der Funktion  $\cot x$  über den Intervall  $0 < x < \pi$  ab, indem man die Veränderung der Tangentenabschnitte verfolgt.

## Die Umkehrfunktionen von Sinus, Cosinus und Tangens

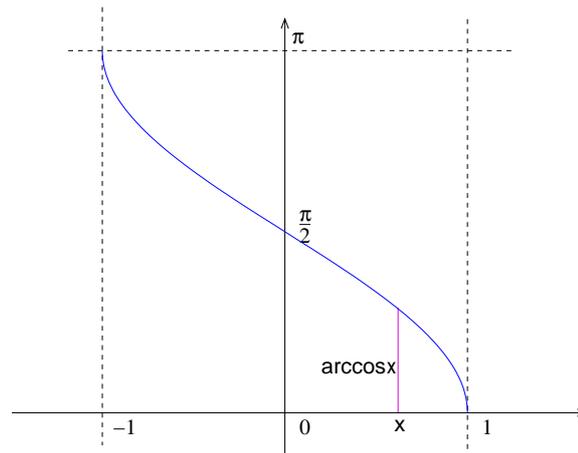
Den Verlauf der Sinusfunktion im Intervall  $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$  entnimmt man: Jedem Funktionswert  $\sin x$  über diesem Intervall entspricht eindeutig ein Wert  $x$  aus diesem Intervall. Es gibt also eine Funktion  $f : [-1, 1] \rightarrow [-\frac{\pi}{2}, \frac{\pi}{2}]$  mit  $f(\sin x) = x$ . Diese Funktion wird mit  $\sin^{-1}$  oder mit  $\arcsin$  (gesprochen „arcus sinus“) bezeichnet. Spezielle Werte dieser Funktion liest man unmittelbar ab. So gilt  $\arcsin(-1) = -\frac{\pi}{2}$ ,  $\arcsin(0) = 0$  und  $\arcsin(1) = \frac{\pi}{2}$ .



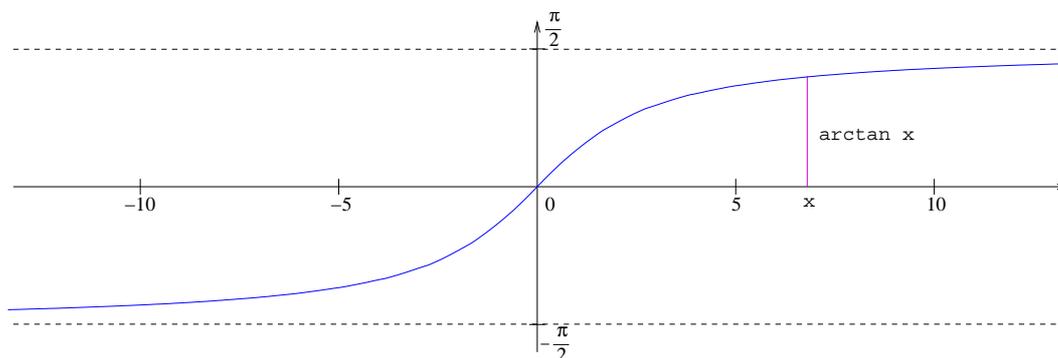
Die Funktion  $\arcsin$  ist die *Umkehrfunktion* der Funktion  $\sin x$ . Es gelten die definierenden Beziehungen  $(\arcsin x) = x$  und  $\sin(\arcsin x) = x$ . Die Funktion  $\cos x$

Abbildung 5.20: Verlauf der Funktion  $\arcsin x$ .

besitzt im Intervall  $0 \leq x \leq \pi$  ebenfalls eine Umkehrfunktion  $\cos^{-1} : [-1, 1] \rightarrow [0, \pi]$ . Sie wird auch mit  $\arccos x$  bezeichnet. Spezielle Werte sind  $\arccos(-1) = \pi$ ,  $\arccos(0) = \frac{\pi}{2}$  und  $\arccos(1) = 0$ . Die definierenden Beziehungen lauten  $\arccos(\cos x) = x$  und  $\cos(\arccos x) = x$ .

Abbildung 5.21: Verlauf der Funktion  $\arccos x$ .

Der Wertebereich der Tangensfunktion über den Intervall  $-\frac{\pi}{2} < x < \frac{\pi}{2}$  ist ganz  $\mathbb{R}$ . Der Verlauf von  $\tan x$  zeigt, dass zu jedem Wert  $y \in \mathbb{R}$  genau ein Wert  $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$  gehört mit  $y = \tan x$ . Also gibt es eine Umkehrfunktion  $\tan^{-1} = \arctan$ , welche die Beziehung  $y \rightarrow x$  herstellt. Es gilt dann  $\arctan(\tan x) = x$  und  $\tan(\arctan x) = x$ . Spezielle Werte sind  $\arctan(-\infty) = -\frac{\pi}{2}$ ,  $\arctan 0 = 0$  und  $\arctan(\infty) = \frac{\pi}{2}$ .

Abbildung 5.22: Verlauf der Funktion  $\arctan x$ .

Die Definition der Umkehrfunktion  $\arctan x$  liefert keine Rechenvorschrift, die es gestattet, bei gegebenen  $x \in \mathbb{R}$  den Funktionswert  $\arctan x$  tatsächlich zu berechnen. Das gleiche gilt für die Funktionen  $\arcsin x$  und  $\arccos x$ . Solche Berechnungsverfahren ergeben sich erst mit den Methoden der Differential- und Integralrechnung (vgl. Abschnitt 7.5)

# Kapitel 6

## Differentialrechnung

Zwei Probleme haben zur Entwicklung der Differentialrechnung geführt, das *Tangentenproblem* und das Problem der *Momentangeschwindigkeit*. Beim Tangentenproblem handelt es sich darum, bei einer beliebigen glatten Kurve in einem vorgegebenen Kurvenpunkt die Tangente zu konstruieren. Solche Tangentenkonstruktionen waren für *spezielle* Kurven wie Kreis, Ellipse und Parabel schon in der Antike bekannt. Die Konstruktion der Tangente eines Kreises in einem vorgegebenen Punkt gehört zum Repertoire der Schulgeometrie. Eine allgemeingültige Methode zur Lösung des Tangentenproblems entwickelte aber erst Gottfried Wilhelm Leibniz (1646-1716) mit der Erfindung der Differentialrechnung. Fast zeitgleich beschäftigte sich Isaac Newton (1642-1727) mit dem Tangentenproblem unter einem physikalisch-dynamischen Aspekt, nämlich als der Frage, wie die Tangente an die Bahnkurve eines Punktes, der sich im Raum bewegt, bestimmt werden kann. Er löste das Problem mit Hilfe des Konzepts der Momentangeschwindigkeit eines sich bewegenden Massenpunktes, eine für ihn intuitiv einleuchtende Vorstellung, die einer mathematischen Präzisierung nicht zu bedürfen schien. Mit dem Begriff der Ableitung einer Funktion werden das Tangentenproblem und das Problem der Momentangeschwindigkeit als Varianten derselben mathematischen Fragestellung erkannt und im Rahmen des Leibniz (-Newtonschen) Kalküls und seiner suggestiven, bis heute gebräuchlichen Notation gelöst.

### 6.1 Die Ableitung einer Funktion

Wir suchen nach einer geometrischen Konstruktion der Tangente an die Standardparabel in einem vorgegebenen Punkt. Die Standardparabel hat die Gleichung  $f(x) = x^2$ . Im Punkt  $P_0 = (x_0, x_0^2)$  soll die Tangente bestimmt werden. Die Tangente an die Standardparabel im Punkt  $P_0$  ist diejenige Gerade durch  $P_0$ , die mit der Standardparabel keinen weiteren Punkt gemeinsam hat. Der Verlauf einer Geraden liegt fest, wenn ein Geradenpunkt vorgegeben ist und ihre „Steigung“ bekannt ist. Zur Ermittlung der Steigung der Tangente an die Standardparabel im Punkt  $P_0 = (x_0, x_0^2)$  nehmen wir einen weiteren Punkt  $P_1 = (x_1, x_1^2)$  auf der Parabel hinzu und legen die Gerade (Sekante)

durch  $P_0$  und  $P_1$ .

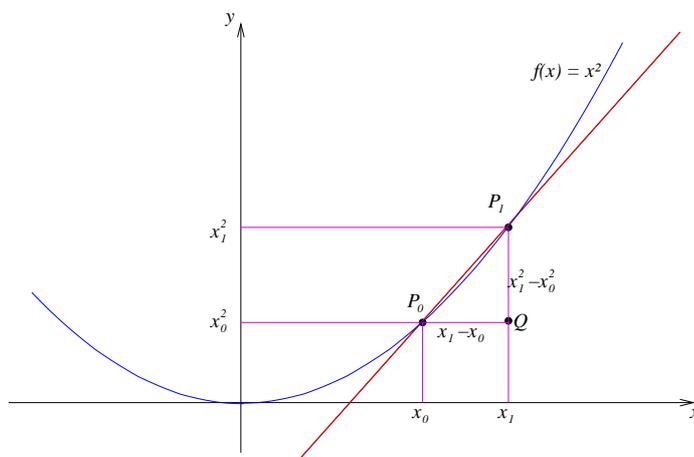


Abbildung 6.1: Sekante der Standardparabel.

Die Steigung der Sekante  $P_0P_1$  durch die Punkte  $P_0$  und  $P_1$  ist gleich dem Quotienten

$$q(x_0, x_1) = \frac{x_1^2 - x_0^2}{x_1 - x_0}$$

der Katheten des Steigungsdreiecks  $P_0QP_1$ . Wegen

$$x_1^2 - x_0^2 = (x_1 - x_0)(x_1 + x_0)$$

ist

$$q(x_0, x_1) = x_1 + x_0.$$

Rückt der Punkt  $P_1$  gegen  $P_0$ , dann geht die Sekante  $P_0P_1$  in die Tangente an die Parabel im Punkt  $P_0$  über. Deren Steigung ist daher

$$q(x_0, x_0) = 2x_0.$$

Die Tangente im Punkt  $P_0$  schneide die  $x$ -Achse im Punkt  $P'_0 = (x'_0, 0)$ . Dann ist der Quotient der Strecken  $x_0^2$  und  $x_0 - x'_0$  gleich der Steigung der Tangente durch den Punkt  $P_0$ ,

$$\frac{x_0^2}{x_0 - x'_0} = 2x_0.$$

Also gilt

$$x_0 = 2(x_0 - x'_0),$$

und somit ist

$$x'_0 = \frac{x_0}{2}.$$

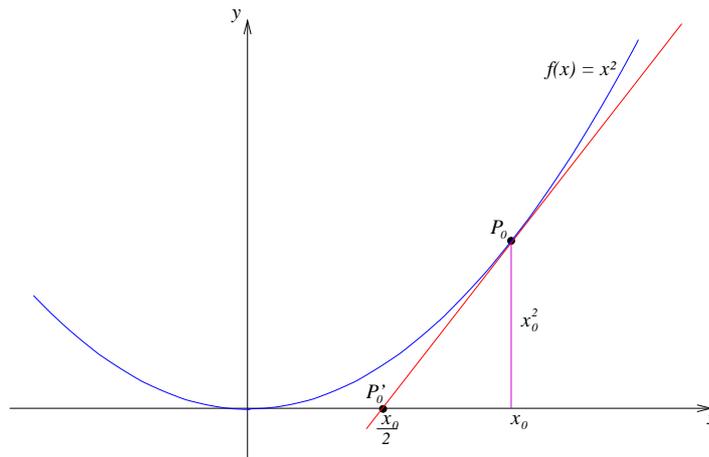


Abbildung 6.2: Tangente an die Standardparabel.

Damit ist ein einfaches *Konstruktionsverfahren* für die Parabeltangente gewonnen:

Die Gerade durch den Parabelpunkt  $P_0 = (x_0, x_0^2)$  und den Punkt  $P'_0 = (\frac{x_0}{2}, 0)$  auf der  $x$ -Achse ist die Tangente an die Standardparabel im Punkt  $P_0$ .

### Der Differenzenquotient einer Funktion

Das Verfahren zur Bestimmung der Parabeltangente lässt sich ohne weiteres auf den Graphen einer beliebigen Funktion übertragen.

Durch die Punkte  $(x, f(x))$  und  $(x+h, f(x+h))$ ,  $h > 0$ , auf dem Graphen der Funktion  $f(x)$  legen wir eine Gerade (Sekante)  $g$ .

Wir betrachten das Dreieck mit den Eckpunkten  $(x, f(x))$ ,  $(x+h, f(x))$  und  $(x+h, f(x+h))$ . Die Steigung der Geraden  $g$  ist dann gleich dem Quotienten der Differenzen  $\Delta f(x) = f(x+h) - f(x)$  und  $\Delta x = (x+h) - x = h$ :

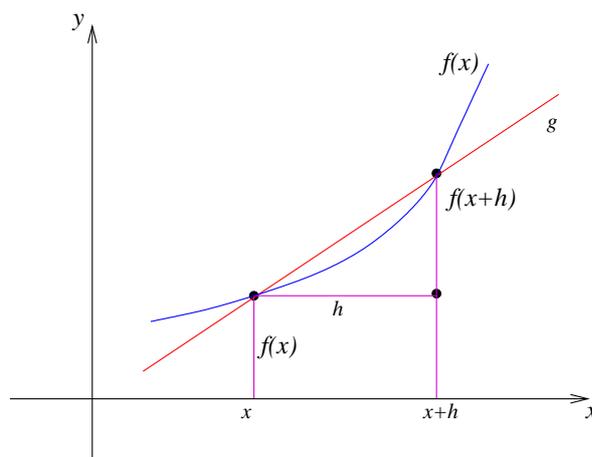
$$(*) \quad \frac{\Delta f(x)}{\Delta x} = \frac{f(x+h) - f(x)}{h}$$

Dieser Quotient wird als *Differenzenquotient* der Funktion  $f(x)$  bezeichnet.

Rückt  $\Delta x = h$  gegen null, dann geht die Gerade  $g$  in die Tangente  $t$  an den Graphen der Funktion  $f(x)$  im Punkt  $(x, f(x))$  über. Deren Steigung ist dann der *Grenzwert* des Differenzenquotienten  $(*)$  für  $h \rightarrow 0$ . Dieser Grenzwert wird mit  $f'(x)$  bezeichnet und heißt die *Ableitung* der Funktion  $f(x)$  an der Stelle  $x$ :

$$(**) \quad \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = f'(x).$$

**Bemerkung:** Die Existenz des Grenzwertes  $f'(x)$  setzt eine gewisse Glattheit der Funktion  $f(x)$  an der Stelle  $x$  voraus. Diese ist nicht immer gegeben, zum Beispiel

Abbildung 6.3: Sekante des Graphen von  $f(x)$ .

dann nicht, wenn  $f(x)$  über der Stelle  $x$  einen Sprung aufweist oder eine Spitze hat. Mit solchen pathologischen Funktionen wollen wir uns aber nicht weiter beschäftigen. Beim Grenzübergang (\*\*) kann die Größe  $h$  beliebig kleine positive oder negative Werte annehmen. Es muß sich gleichwohl immer *derselbe* Grenzwert  $f'(x)$  ergeben. ▲

Eine Funktion, für die der Grenzwert (\*\*) an einer Stelle  $x$  existiert, heißt an der Stelle  $x$  *differenzierbar*. Wenn dies für *alle*  $x$  aus einem Intervall  $I$  gilt, dann heißt  $f(x)$  in diesem Intervall *differenzierbar*.

**Beispiel:** Für die Potenzfunktion  $f(x) = x^n$ ,  $n \in \mathbb{N}$ , erhält man aufgrund des binomischen Lehrsatzes

$$\begin{aligned}
 f(x+h) - f(x) &= (x+h)^n - x^n \\
 &= \sum_{k=0}^n \binom{n}{k} x^{n-k} h^k - x^n \\
 &= x^n + nx^{n-1}h + \sum_{k=2}^n \binom{n}{k} x^{n-k} h^k - x^n \\
 &= nx^{n-1}h + \sum_{k=2}^n \binom{n}{k} x^{n-k} h^k
 \end{aligned}$$

Also ist

$$\frac{f(x+h) - f(x)}{h} = nx^{n-1} + \sum_{k=2}^n \binom{n}{k} x^{n-k} h^{k-1}.$$

Für  $h \rightarrow 0$  geht der zweite Ausdruck auf der rechten Seite gegen null und man erhält

$$f'(x) = nx^{n-1}$$

■

## Linearisierung

Die Existenz des Grenzwertes (\*\*) lässt sich auch folgendermaßen interpretieren: Es gibt eine Funktion  $g(x, h)$ , so dass man schreiben kann

$$\frac{f(x+h) - f(x)}{h} = f'(x) + g(x, h)$$

mit  $g(x, h) \rightarrow 0$  für  $h \rightarrow 0$ . Also gilt

$$f(x+h) = f(x) + f'(x)h + r(x, h)$$

mit  $r(x, h) = hg(x, h)$ . Die Funktion  $r(x, h)$  strebt schneller gegen null für  $h \rightarrow 0$  als die Funktion  $g(x, h)$ , denn es gilt sogar  $r(x, h)/h \rightarrow 0$  für  $h \rightarrow 0$ . Das bedeutet: Die Restfunktion  $r(x, h)$  ist klein, wenn  $h$  klein ist. Daher gilt

$$(***) \quad f(x+h) \approx f(x) + f'(x)h,$$

wobei das Symbol  $\approx$  für „ist näherungsweise gleich“ steht.

Betrachtet man  $x_0$  als fest gewählt, dann ist

$$y = f(x_0) + f'(x_0)x$$

die Gleichung der Tangente an  $f(x)$  im Punkt  $(x_0, f(x_0))$ . Die Tangentengleichung ist eine lineare Funktion von  $x$ , das heißt, die Variable  $x$  tritt nur in der ersten Potenz auf. Man erhält so eine *lineare Approximation* der Funktion  $f(x)$  in einer Umgebung der Stelle  $x_0$ .

**Bemerkung:** Die obige Überlegung lässt sich auch umkehren, was gelegentlich von Nutzen ist: Hat man für eine Funktion  $f(x)$  nachgewiesen, dass mit Funktionen  $g(x)$  und  $r(x, h)$  eine Beziehung der Form

$$f(x+h) = f(x) + g(x)h + r(x, h)$$

besteht mit  $r(x, h)/h \rightarrow 0$  für  $h \rightarrow 0$ , dann folgt  $g(x) = f'(x)$ . Diese Beziehung ist nämlich gleichbedeutend zur Beziehung

$$\frac{1}{h} (f(x+h) - f(x)) = g(x) + r(x, h)/h.$$

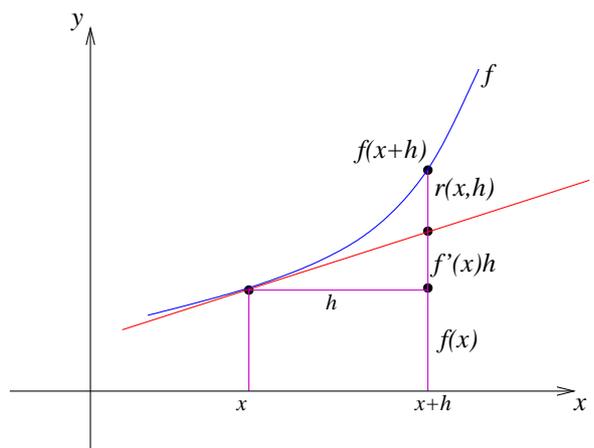


Abbildung 6.4: Die Tangente als lineare Approximation.

Für  $h \rightarrow 0$  strebt die linke Seite gegen  $f'(x)$  und die rechte gegen  $g(x)$  und dies war zu zeigen. ▲

## Der Differentialquotient

Um in der Notation auf die Herkunft der Ableitung  $f'(x)$  einer Funktion  $y = f(x)$  als Grenzwert des Differenzenquotienten

$$\frac{\Delta y}{\Delta x} = \frac{\Delta f(x)}{\Delta x}$$

für  $\Delta x \rightarrow 0$  hinzuweisen, schreibt man auch symbolisch

$$f'(x) = \frac{dy}{dx} = \frac{df(x)}{dx}.$$

Die Ersetzung  $\Delta \rightarrow d$  soll dabei den Übergang zu „unendlich kleinen“ Größen  $dy$  und  $dx$  bzw.  $df(x)$  andeuten

Die Vorstellung von der Existenz solcher unendlich kleiner Größen, die von G.W. Leibnitz in die Mathematik eingeführt wurden, war lange Zeit umstritten, wurde aber in neuer Zeit im Rahmen der „Non-Standard-Analysis“ von Abraham Robinson (1918-1974) gerechtfertigt. Ihre pragmatische Rechtfertigung ist darin zu sehen, dass auf diese Weise eine sehr suggestive Notation möglich wird.

## Der Mittelwertsatz der Differentialrechnung

Der folgende Satz, der Mittelwertsatz der Differentialrechnung, ist beweistechnisch häufig von Nutzen.

**Satz:** Ist  $f(x)$  eine im Intervall  $a \leq x \leq b$  differenzierbare Funktion, dann gibt es eine

Zahl  $x_0$  mit  $a \leq x_0 \leq b$ , so dass gilt

$$(*) \quad \frac{f(b) - f(a)}{b - a} = f'(x_0).$$

Wir verzichten hier auf einen strengen Beweis dieses Satzes und berufen uns vielmehr auf die Anschauung (vgl. Abbildung 6.5).

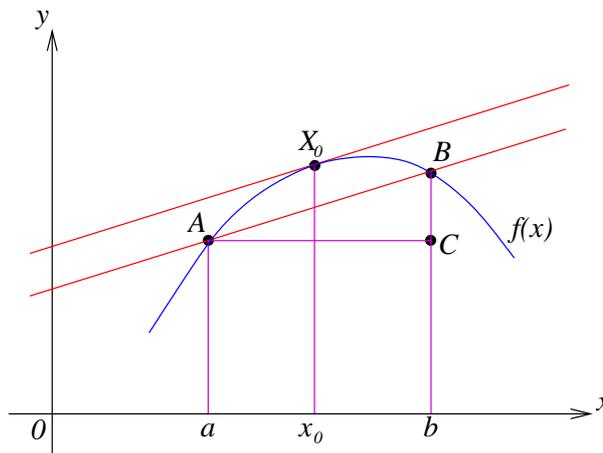


Abbildung 6.5: Der Mittelwertsatz der Differentialrechnung.

Der Satz besagt geometrisch: Zur Sekante durch die Punkte  $A = (a, f(a))$ ,  $B = (b, f(b))$  gibt es eine zu dieser parallele Tangente an den Graphen von  $f(x)$  im Punkte  $X_0 = (x_0, f(x_0))$  mit  $a \leq x_0 \leq b$ . Die Steigung der Sekante ist gleich dem Verhältnis der Strecken  $BC$  und  $AC$ , also gleich  $(f(b) - f(a))/(b - a)$  und die Steigung der Tangente im Punkt  $X_0$  ist gleich  $f'(x_0)$ .

Eine einfache Schlussfolgerung aus diesem Satz lautet: Die Ableitung einer Funktion ist genau dann gleich null, wenn diese eine Konstante ist. Ist nämlich  $f'(x) = 0$  überall in einem Intervall, das die Werte  $a, b$ ,  $a < b$ , enthält, dann gilt wegen (\*)

$$\frac{f(b) - f(a)}{b - a} = f'(x_0) = 0,$$

also ist  $f(b) - f(a) = 0$  und damit  $f(b) = f(a)$ . Die Funktion  $f(x)$  hat also überall den gleichen Wert, ist also eine Konstante. Ist umgekehrt eine Funktion  $f(x)$  eine Konstante, dann ist ihre Ableitung gleich null, wie leicht zu sehen ist.

## 6.2 Ableitungsregeln

Mit Hilfe einiger Regeln kann man die Ableitung „zusammengesetzter“ Funktionen auf die Bildung der Ableitung bei ihren „Teilfunktionen“ zurückführen. Die Teilfunk-

tionen werden dabei immer als differenzierbar vorausgesetzt.

### Linearität der Ableitung

Mit Hilfe des Differenzenquotienten (\*) aus Abschnitt 6.1 folgt unmittelbar, dass die Ableitung einer Summe von Funktionen gleich der Summe der Ableitungen der beteiligten Funktionen ist:

$$(f(x) + g(x))' = f'(x) + g'(x).$$

Und ist  $a \in \mathbb{R}$  eine Zahl, dann gilt

$$(af(x))' = af'(x).$$

Beide Regeln lassen sich zusammenfassen in der Regel

$$(*) \quad (af(x) + bg(x))' = af'(x) + bg'(x)$$

zur Bildung der Ableitung einer *Linearkombination* zweier Funktionen.

Fasst man die Bildung der Ableitung einer Funktion  $f(x)$  als eine „Operation“ auf, die das „Objekt“  $f(x)$  in das Objekt  $f'(x)$  überführt und symbolisiert man diesen Vorgang durch den *Operator*  $D$ , indem man schreibt

$$Df(x) = f'(x),$$

dann lässt sich die Regel (\*) auch in Form der Beziehung

$$(L) \quad D(af(x) + bg(x)) = aDf(x) + bDg(x)$$

schreiben. Man kann also den Operator  $D$  wie eine Zahl in die Klammer auf der linken Seite „hineinmultiplizieren“.

Ein Operator mit der Eigenschaft (L) wird als *linearer Operator* bezeichnet.

### Die Produktregel

Die Ableitung des Produktes  $f(x)g(x)$  zweier Funktionen lässt sich auf die Ableitungen  $f'(x)$  und  $g'(x)$  der beteiligten Faktoren zurückführen.

Durch Linearisierung von  $f(x)$  und  $g(x)$  erhält man zunächst

$$\begin{aligned} f(x+h)g(x+h) &= (f(x) + f'(x)h + r_1(x, h))(g(x) + g'(x)h + r_2(x, h)) \\ &= f(x)g(x) + (f'(x)g(x) + f(x)g'(x))h + r(x, h). \end{aligned}$$

Die Restfunktion  $r(x, h)$  hat die Eigenschaft  $r(x, h)/h \rightarrow 0$  für  $h \rightarrow 0$ . Aufgrund der Bemerkung in Abschnitt 6.1 muss daher der Klammerausdruck mit dem Faktor  $h$  gleich der Ableitung des Produktes  $f(x)g(x)$  sein. Man findet so die *Produktregel*

$$(P) \quad (f(x)g(x))' = f'(x)g(x) + f(x)g'(x).$$

Mit Hilfe des Operators  $D$  und in verkürzter Symbolik kann man die Produktregel auch in der Form

$$(P) \quad D(f \cdot g) = (Df) \cdot g + f \cdot (Dg)$$

schreiben.

**Bemerkung:** Eine andere Herleitung der Produktregel bedient sich eines Tricks: Ist  $u(x) = f(x)g(x)$ , dann ist

$$\frac{1}{h}(u(x+h) - u(x)) = \frac{1}{h}(f(x+h)g(x+h) - f(x)g(x))$$

der zugehörige Differenzenquotient. Fügt man in der Klammer auf der rechten Seite den Ausdruck  $-f(x)g(x+h) + f(x)g(x+h) = 0$  ein, dann kann man schreiben:

$$\frac{\Delta u}{\Delta x} = \frac{1}{h}(f(x+h) - f(x))g(x+h) + f(x)\frac{1}{h}(g(x+h) - g(x)).$$

Für  $h \rightarrow 0$  strebt die linke Seite gegen  $u'(x)$ , der erste Ausdruck rechts gegen  $f'(x)g(x)$  und der zweite gegen  $f(x)g'(x)$ . So ergibt sich wiederum die Produktregel. ▲

## Die Quotientenregel

Es sei nun

$$u(x) = \frac{f(x)}{g(x)}$$

der Quotient zweier Funktionen. Wir können dann etwas salopp schreiben

$$ug = f.$$

Unter der Voraussetzung der Differenzierbarkeit der beteiligten Funktionen  $u, f, g$  wenden wir auf die linke Seite die Produktregel (P) an und erhalten

$$u'g + ug' = f'$$

und mit  $u = f/g$  folgt

$$\begin{aligned} u'g &= f' - \frac{f}{g}g' \\ &= \frac{f'g - fg'}{g}, \end{aligned}$$

also ist

$$u' = \frac{f'g - fg'}{g^2}.$$

Damit ergibt sich die *Quotientenregel*

$$(Q) \quad \left( \frac{f(x)}{g(x)} \right)' = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}.$$

### Die Kettenregel

Zur Bestimmung der Ableitung einer Funktion  $u(x) = f(g(x))$ , bei der zwei Funktionen  $f$  und  $g$  ineinandergesetzt oder „verkettet“ werden, sei an die Näherungsbeziehung

$$(***) \quad f(x+h) \approx f(x) + f'(x)h$$

aus Abschnitt 6.1 erinnert. Dann kann man schreiben

$$u(x+h) \approx u(x) + u'(x)h,$$

das heißt,

$$(*) \quad f(g(x+h)) \approx f(g(x)) + u'(x)h.$$

Setzt man die Näherungsbeziehung

$$g(x+h) \approx g(x) + g'(x)h$$

in die linke Seite ein, so erhält man

$$\begin{aligned} f(g(x+h)) &\approx f(g(x) + g'(x)h) \\ &\approx f(g(x)) + f'(g(x))g'(x)h. \end{aligned}$$

Die mit der Variablen  $h$  behafteten Terme auf der rechten Seite von (\*) und in der letzten Zeile müssen übereinstimmen. Es folgt die *Kettenregel*

$$(K) \quad (f(g(x)))' = f'(g(x)) \cdot g'(x).$$

**Bemerkung:** Durch das Mitführen der bei der Linearisierung vernachlässigten Restterme lässt sich, wie bei der Produktregel, der Rechnung zum Beweis der Kettenregel volle Strenge verleihen, allerdings auf Kosten der Durchsichtigkeit. ▲

### Ableitung der Umkehrfunktion

Als erste Anwendung der Kettenregel ergibt sich die Regel zur Berechnung der Ableitung der Umkehrfunktion  $f^{-1}(x)$  einer Funktion  $f(x)$ .

Funktion und Umkehrfunktion stehen in der Beziehung

$$f(f^{-1}(x)) = x.$$

Bildet man auf beiden Seiten die Ableitung und benutzt bei der linken Seite die Kettenregel und auf der rechten Seite die Ableitungsregel  $(x^n)' = nx^{n-1}$  für  $n = 1$  (s. o.), dann erhält man

$$f'(f^{-1}(x)) \cdot (f^{-1}(x))' = 1.$$

Also gilt:

$$(U) \quad (f^{-1}(x))' = \frac{1}{f'(f^{-1}(x))}.$$

Das ist die Regel zur Bildung der Ableitung  $(f^{-1}(x))'$  der Umkehrfunktion  $f^{-1}(x)$  einer Funktion  $f(x)$ .

**Beispiel:** Die Funktion  $f(x) = x^2$  hat die Umkehrfunktion  $f^{-1}(x) = \sqrt{x}$  mit positiver Wurzel. Mit  $f'(x) = 2x$  erhält man dann mit (U)

$$(f^{-1}(x))' = \frac{1}{2\sqrt{x}}.$$

■

## 6.3 Die Ableitung der elementaren Funktionen

Die im vorigen Abschnitt hergeleiteten Ableitungsregeln erlauben es nunmehr, die Ableitungen der elementaren Funktionen aus Kapitel 5.2 zu bestimmen.

### Polynome

Die Linearität der Operation „Bildung der Ableitung“ zusammen mit der in Abschnitt 6.1 hergeleiteten Ableitungsregel

$$(x^n)' = nx^{n-1}, \quad n \in \mathbb{N},$$

für die Potenzen erlauben unmittelbar, die Ableitung eines Polynoms zu bestimmen. Ein Beispiel hierzu mag genügen: Die Ableitung des Polynoms

$$p(x) = 2 + x + 3x^2 + x^3$$

sei zu berechnen. Dies geschieht, indem man die Ableitungen der einzelnen rechts stehenden Ausdrücke bildet und die einzelnen Ableitungen summiert.

Die Ableitung einer konstanten Funktion

$$f(x) = c, \quad c \in \mathbb{R},$$

ist gleich null, denn der entsprechende Differenzenquotient lautet

$$\frac{f(x+h) - f(x)}{h} = \frac{c - c}{h} = 0,$$

also gilt

$$f'(x) = 0.$$

Damit erhalten wir

$$\begin{aligned} p'(x) &= 0 + 1 + 3 \cdot 2x + 3x^2 \\ &= 1 + 6x + 3x^2 \end{aligned}$$

als Ableitung des Polynoms  $p(x)$ .

## Rationale Funktionen

Eine rationale Funktion  $r(x)$  hat die Form

$$r(x) = \frac{p(x)}{q(x)}$$

mit Polynomen  $p(x), q(x)$ , so dass für die Bildung ihrer Ableitung die Quotientenregel (Q) aus Abschnitt 6.2 zuständig ist. Man erhält somit

$$(*) \quad r'(x) = \frac{p'(x)q(x) - p(x)q'(x)}{q^2(x)}.$$

Für die rationale Funktion

$$r(x) = \frac{x-1}{x+1}$$

zum Beispiel ergibt sich

$$\begin{aligned} r'(x) &= \frac{1 \cdot (x+1) - (x-1) \cdot 1}{(x+1)^2} \\ &= \frac{2}{(x+1)^2}. \end{aligned}$$

Die Potenzfunktionen mit negativen Exponenten

$$r(x) = \frac{1}{x^n} = x^{-n}, \quad n \in \mathbb{N},$$

haben nach (\*) die Ableitung

$$\begin{aligned} r'(x) &= \frac{0 \cdot x^n - 1 \cdot nx^{n-1}}{x^{2n}} \\ &= \frac{-n}{x^{n+1}} = -nx^{-n-1}. \end{aligned}$$

Die Ableitungsregel  $(x^n)' = nx^{n-1}$ ,  $n \in \mathbb{N}$ , bleibt demnach auch für negative Potenzen von  $x$  erhalten. Nimmt man noch formal den Fall  $n = 0$  hinzu,

$$(x^0)' = 1' = 0 = 0 \cdot x^{0-1} = 0 \cdot x^{-1},$$

dann kann man somit die Ableitungsregel für Potenzen von  $x$  auf die ganzen Zahlen  $\mathbb{Z}$  erweitern:

$$(x^n)' = nx^{n-1}, \quad n \in \mathbb{Z}.$$

## Die Exponentialfunktion

Die Exponentialfunktion

$$e^x = 1 + x + \frac{x^2}{1 \cdot 2} + \frac{x^3}{1 \cdot 2 \cdot 3} + \frac{x^4}{1 \cdot 2 \cdot 3 \cdot 4} + \dots$$

zeigt bei Bildung der Ableitung das denkbar einfachste Verhalten, nämlich unverändert zu bleiben. Durch gliedweise Differentiation der rechten Seite erhält man nämlich:

$$\begin{aligned} (e^x)' &= 0 + 1 + \frac{2x}{1 \cdot 2} + \frac{3x^2}{1 \cdot 2 \cdot 3} + \frac{4x^3}{1 \cdot 2 \cdot 3 \cdot 4} + \dots \\ &= 1 + x + \frac{x^2}{1 \cdot 2} + \frac{x^3}{1 \cdot 2 \cdot 3} + \dots = e^x. \end{aligned}$$

Man sieht: Die ursprüngliche Reihe reproduziert sich. Es gilt also die Regel

$$(e^x)' = e^x.$$

**Bemerkung: Eindeutige Bestimmtheit der Exponentialfunktion und ihrer Funktionalgleichung.** Es stellt sich die naheliegende Frage, ob es noch weitere Funktionen  $f(x)$  gibt, mit der Eigenschaft  $f'(x) = f(x)$ . Diese Frage lässt sich nun mit Hilfe der Differentialrechnung beantworten:

Nehmen wir an,  $f(x)$  sei eine solche Funktion. Wir bilden dann die Funktion

$$q(x) = \frac{f(x)}{e^x}$$

und berechnen mit Hilfe der Quotientenregel deren Ableitung

$$\begin{aligned} q'(x) &= \frac{f'(x)e^x - f(x)(e^x)'}{(e^x)^2} \\ &= \frac{f'(x)e^x - f(x)e^x}{(e^x)^2} = 0. \end{aligned}$$

Die Funktion  $q(x)$  hat also die Eigenschaft, dass ihre Ableitung gleich null ist,  $q'(x) = 0$ . Daher muss  $q(x)$  eine Konstante sein, wie wir oben gezeigt haben (vgl. Mittelwertsatz der Differentialrechnung).

Es gibt also eine Konstante  $c$ , so dass gilt  $q(x) = c$ , das heißt, es ist

$$\frac{f(x)}{e^x} = c$$

oder

$$f(x) = c e^x.$$

Setzt man  $x = 0$ , so folgt

$$c = f(0)$$

und damit

$$(*) \quad f(x) = f(0) e^x.$$

Die Funktionen  $f(x)$  mit der Eigenschaft  $f'(x) = f(x)$  sind daher allesamt konstante Vielfache der Exponentialfunktion.

## Die Funktionalgleichung der Exponentialfunktion

Wir betrachten jetzt – für festes  $y$  – die Funktion

$$f(x) = e^{x+y}.$$

Ihre Ableitung erhält man mit Hilfe der Kettenregel ( $K$ ) aus Abschnitt 6.2:

$$f'(x) = e^{x+y} \cdot (1 + 0) = e^{x+y}.$$

Also gilt  $f'(x) = f(x)$  und somit wegen  $(*)$

$$e^{x+y} = e^{0+y} \cdot e^x.$$

Somit erfüllt die Exponentialfunktion  $e^x$  die Funktionalgleichung

$$(FG) \quad e^{x+y} = e^x \cdot e^y,$$

die wir bereits früher benutzt haben. ▲

## Die Logarithmusfunktion

Die (natürliche) Logarithmusfunktion  $\log x = \ln x$  ist die Umkehrfunktion der Exponentialfunktion. Das bedeutet, es besteht die Beziehung

$$e^{\log x} = x.$$

Mit Hilfe der Kettenregel erhält man dann

$$\begin{aligned} (e^{\log x})' &= e^{\log x} \cdot (\log x)' \\ &= x \cdot (\log x)' = 1. \end{aligned}$$

Folglich gilt

$$(\log x)' = \frac{1}{x}.$$

Für  $x > 0$  und  $x \rightarrow 0$  strebt also  $(\log x)' \rightarrow \infty$ . An der Stelle  $x = 1$  hat man  $(\log x)' = 1$  und für  $x \rightarrow \infty$  strebt  $(\log x)' \rightarrow 0$ . Der Anstieg der Logarithmusfunktion wird also für große  $x$  immer flacher.

### Die allgemeine Exponentialfunktion

Die allgemeine Exponentialfunktion

$$f(x) = a^x, \quad a > 0,$$

ist erklärt als die Funktion

$$f(x) = e^{\log a \cdot x}.$$

Aufgrund der Kettenregel ergibt sich dann

$$\begin{aligned} (*) \quad (a^x)' &= e^{\log a \cdot x} \log a \\ &= \log a \cdot a^x. \end{aligned}$$

Für die natürliche Basis  $a = e$  erhält man wieder  $f'(x) = f(x)$ .

### Die allgemeine Logarithmusfunktion

Wenn zwischen zwei Zahlen  $x$  und  $y$  bezüglich einer Basis  $a > 0$  die Beziehung

$$a^y = x$$

besteht, dann ist

$$y = \log_a x$$

der Logarithmus von  $x$  zur Basis  $a$ . Mit Hilfe der Kettenregel und der Beziehung (\*) folgt dann

$$(a^y)' = \log a \cdot a^y \cdot y' = 1,$$

also

$$\log a \cdot x \cdot y' = 1.$$

Wir erhalten die Ableitungsregel

$$(\log_a x)' = \frac{1}{\log a \cdot x}$$

für die allgemeine Logarithmusfunktion.

### Die allgemeine Potenzfunktion

Die Funktion

$$f(x) = x^\alpha \in \mathbb{R},$$

die allgemeine Potenzfunktion, ist erklärt als die Funktion

$$f(x) = e^{\alpha \cdot \log x}.$$

Daher erhält man aufgrund der Kettenregel

$$\begin{aligned} f'(x) &= e^{\alpha \cdot \log x} \cdot \frac{\alpha}{x} \\ &= x \cdot \frac{\alpha}{x} = \alpha \cdot x^{\alpha-1}. \end{aligned}$$

Damit haben wir für Potenzfunktionen

$$f(x) = x^\alpha$$

mit beliebigen Exponenten  $\alpha \in \mathbb{R}$  wiederum die bereits bekannte Form der Ableitungsregel

$$f'(x) = \alpha x^{\alpha-1}$$

erhalten.

## Die trigonometrischen Funktionen

Die Eulerformel (vgl. Abschnitt 2.4) konstatiert den Zusammenhang

$$(*) \quad e^{ix} = \cos x + i \sin x$$

zwischen der Exponentialfunktion und den trigonometrischen Funktionen  $\sin$  und  $\cos$ .

Bildet man auf beiden Seiten dieser Beziehung die Ableitung, dann findet man mit Hilfe der Kettenregel den Zusammenhang

$$i e^{ix} = (\cos x)' + i(\sin x)'$$

Setzt man auf der linken Seite wieder die Beziehung  $(*)$  ein, so folgt die Gleichung

$$(**) \quad i \cos x - \sin x = (\cos x)' + i(\sin x)'$$

Zwei komplexe Zahlen sind genau dann gleich, wenn sie in Realteil und Imaginärteil übereinstimmen. Durch Vergleich der linken und rechten Seite bei  $(**)$  können wir daher schließen

$$(\cos x)' = -\sin x$$

und

$$(\sin x)' = \cos x.$$

Dies sind die Ableitungsregeln für die trigonometrischen Funktionen  $\sin$  und  $\cos$ .

**Bemerkung:** Die obigen Ableitungsregeln ergeben sich auch unmittelbar, wenn man die Reihendarstellungen

$$\cos x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$$

und

$$\sin x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}$$

aus Abschnitt 5.5 zugrundelegt und die Ableitungen  $(\cos x)'$  und  $(\sin x)'$  bildet, indem man jeweils die Summanden auf der rechten Seite einzeln ableitet. ▲

Die Ableitung der Tangensfunktion

$$\tan x = \frac{\sin x}{\cos x}$$

bildet man mit Hilfe der Quotientenregel:

$$\begin{aligned} (\tan x)' &= \frac{\cos x \cdot \cos x - \sin x \cdot (-\sin x)}{\cos^2 x} \\ &= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} \end{aligned}$$

Da  $\sin^2 x + \cos^2 x = 1$  ist, folgt

$$(\tan x)' = \frac{1}{\cos^2 x}.$$

Entsprechend erhält man:

$$(\cot x)' = \left( \frac{\cos x}{\sin x} \right)' = \frac{-1}{\sin^2 x}.$$

Die Ableitung der *Umkehrfunktionen* der trigonometrischen Funktionen gewinnt man mit Hilfe der Kettenregel bzw. der Formel (U) aus Abschnitt 6.2.

Wir begnügen uns hier mit der Bestimmung der Ableitung der Funktion  $\arcsin$ ,  $-1 \leq x \leq 1$ . Es gilt

$$\sin(\arcsin x) = x$$

und daher (Kettenregel)

$$(*) \quad \cos(\arcsin x) \cdot (\arcsin x)' = 1.$$

Da

$$\cos \alpha = \sqrt{1 - \sin^2 \alpha}$$

gilt, ist

$$\begin{aligned}\cos(\arcsin x) &= \sqrt{1 - (\sin(\arcsin x))^2} \\ &= \sqrt{1 - x^2}.\end{aligned}$$

Einsetzen dieser Beziehung in (\*) ergibt

$$(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}}.$$

Man erkennt, dass die Funktion  $\arcsin x$  an den Stellen  $x = \pm 1$  eine senkrechte Tangente besitzt. Dies entspricht der Tatsache, dass die Funktion  $\sin x$  an den Stellen  $x = \pm \frac{\pi}{2}$  eine waagerechte Tangente hat.

## 6.4 Anwendungen der Differentialrechnung

Mit Hilfe der Differentialrechnung erschließt sich der Verlauf einer Funktion  $f(x)$ . Man kann feststellen, ob die zu untersuchende Funktion über einem Intervall monoton *steigt* oder *fällt* und es lässt sich die Lage ihrer *Nullstellen*, *Extremwerte* und *Wendepunkte* mit beliebiger Genauigkeit berechnen.

### Monotonie

Gilt für einen Punkt  $x_0$  aus einem Intervall  $[a, b]$ , über dem die differenzierbare Funktion  $f(x)$  definiert ist,  $f'(x_0) > 0$ , dann hat die Funktion in diesem Punkt eine Tangente mit positivem Anstiegswinkel. Die Funktionswerte sind dort „tendenziell“ ansteigend. Gilt für *alle*  $x \in [a, b]$ , dass  $f'(x) > 0$  ist, dann lässt sich noch mehr aussagen.

Sind nämlich  $x_0 < x_1$  zwei Stellen aus dem Intervall  $[a, b]$ , dann gibt es aufgrund des Mittelwertsatzes der Differentialrechnung (vgl. Abschnitt 6.3) eine Zwischenstelle  $\xi$  mit  $x_0 \leq \xi \leq x_1$ , so dass gilt

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} = f'(\xi).$$

Die rechte Seite dieser Gleichung ist positiv, also auch die linke. Da nach Annahme  $x_1 - x_0 > 0$  ist, so muss auch  $f(x_1) - f(x_0) > 0$ , das heißt,  $f(x_0) < f(x_1)$  sein.

**Definition:** Gilt für alle  $x_0, x_1 \in [a, b]$  mit  $x_0 < x_1$ , dass  $f(x_0) < f(x_1)$  ist, dann ist die Funktion  $f(x)$  über  $[a, b]$  streng monoton steigend.

Damit haben wir das folgende Ergebnis:

**Satz:** Die Funktion  $f(x)$  sei differenzierbar über dem Intervall  $[a, b]$  und es gelte  $f'(x) > 0$  für  $x \in [a, b]$ . Dann ist  $f(x)$  über  $[a, b]$  streng monoton steigend.

Gilt  $f'(x) < 0$  für  $x \in [a, b]$ , so folgt entsprechend, dass  $f(x)$  dort streng monoton fallend ist. Das heißt, für  $x_0 < x_1$ ,  $x_0, x_1 \in [a, b]$  ist  $f(x_0) > f(x_1)$ . Schwächt man die Voraussetzungen  $f'(x) > 0$  bzw.  $f'(x) < 0$  zu  $f'(x) \geq 0$  bzw.  $f'(x) \leq 0$  ab, so folgt entsprechend die *einfache* Monotonie  $f(x_0) \leq f(x_1)$  bzw.  $f(x_0) \geq f(x_1)$ .

### Berechnung von Nullstellen; das Newton-Verfahren

Die lineare Funktion  $f(x) = ax + b$ ,  $a \neq 0$ , hat die Nullstelle  $x_0 = -b/a$ . Eine quadratische Funktion (Polynom 2. Grades)  $f(x) = ax^2 + bx + c$ ,  $a \neq 0$ , hat im allgemeinen zwei Nullstellen, die sich explizit berechnen lassen (vgl. Abschnitt 5.1, Parabel). Ihre Nullstellen  $x_1, x_2$  erhält man als die Lösung einer quadratischen Gleichung. Man findet

$$x_{1/2} = -\frac{b}{2a} \pm \sqrt{\left(\frac{b}{2a}\right)^2 - \frac{c}{a}}.$$

Für Funktionen  $f(x)$ , die Polynome 3. oder 4. Grades sind, gibt es entsprechende Formeln zur Berechnung der Nullstellen. Diese Formeln sind aber sehr kompliziert und werden in der Praxis kaum benutzt.

Von dem englischen Physiker und Mathematiker Isaac Newton (1643–1727) stammt ein Verfahren, mit dem man Nullstellen differenzierbarer Funktionen mit beliebiger Genauigkeit berechnen kann. Es trägt den Namen seines Erfinders und heißt das *Newton-Verfahren*.

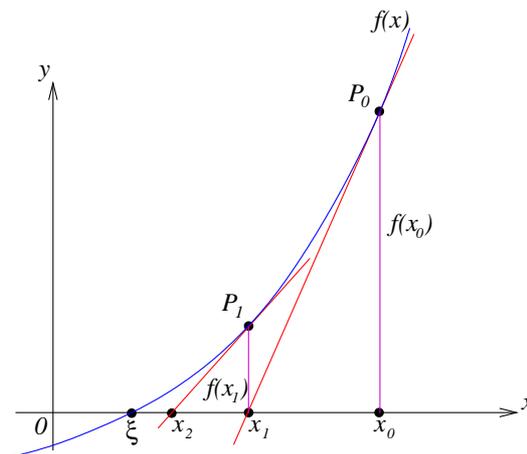


Abbildung 6.6: Das Newton-Verfahren.

Es sei  $\xi$  eine Nullstelle von  $f(x)$ , das heißt, es gelte  $f(\xi) = 0$ . Die Stelle  $x_0$  liege in der

Nähe von  $\xi$ . Die Tangente im Punkt  $P_0 = (x_0, f(x_0))$  an  $f(x)$  schneide die  $x$ -Achse im Punkt  $x_1$ . Verläuft  $f(x)$  wie in Abb. 6.6, dann wird  $x_1$  näher an  $\xi$  liegen als  $x_0$ . Die Tangente in  $P_0$  schneidet die  $x$ -Achse in einem Winkel, dessen Tangens gleich der Ableitung  $f'(x_0)$  ist. Also gilt

$$\frac{f(x_0)}{x_0 - x_1} = f'(x_0).$$

Daraus folgt

$$x_0 - x_1 = \frac{f(x_0)}{f'(x_0)}$$

oder

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Nimmt man  $x_1$  als neuen Startpunkt, dann erhält man in gleicher Weise

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

Der Punkt  $x_2$  wird dann noch näher an der Nullstelle  $\xi$  liegen als  $x_1$ .

Hat man auf diese Weise einen Näherungswert  $x_n$ ,  $n \in \mathbb{N}$ , für die Nullstelle  $\xi$  erreicht, dann wird dieser durch den Wert

$$(N) \quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

weiter verbessert. Diese *Rekursionsformel* erzeugt eine Zahlenfolge  $x_0, x_1, x_2, \dots$ , die sehr schnell gegen die Nullstelle  $\xi$  konvergiert. (Einen strengen Beweis dieser Aussage können wir hier nicht führen.)

**Beispiel: Berechnung von  $\sqrt{2}$ .** Die Beziehung  $\xi = \sqrt{2}$  ist äquivalent zu  $\xi^2 = 2$  oder  $\xi^2 - 2 = 0$ . Also ist  $\xi$  eine Nullstelle der Funktion  $f(x) = x^2 - 2$ . Die Rekursionsformel (N) lautet daher im vorliegenden Fall

$$x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n}$$

oder vereinfacht

$$x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}.$$

Offenbar gilt  $1 < \sqrt{2} < 2$ , so dass  $x_0 = 2$  ein möglicher Startwert ist. Damit erhält

man

$$\begin{aligned}x_1 &= 1 + \frac{1}{2} = 1,5; \\x_2 &= \frac{1,5}{2} + \frac{1}{1,5} = 1,41\bar{6}; \\x_3 &= \frac{1,41\bar{6}}{2} + \frac{1}{1,41\bar{6}} = 1,41421568; \\x_4 &= \frac{1,41421568}{2} + \frac{1}{1,41421568} = 1,41421356.\end{aligned}$$

Der letzte Wert ist bereits sehr genau. Die Stellen nach dem Komma sind korrekt, denn der Taschenrechner zeigt an  $\sqrt{2} = 1,41421356237$ . Die Folge  $x_n$  konvergiert also sehr schnell gegen den Grenzwert  $\xi$ . ■

### Die Regel von de l'Hospital

Gelegentlich stößt man auf unbestimmte Ausdrücke der Form  $0/0$  oder  $\infty/\infty$ . Dies ist zum Beispiel der Fall, wenn man den Quotienten

$$q(x) = \frac{\sin x}{x}$$

der Funktionen  $\sin x$  und  $x$  für  $x = 0$  betrachtet. Gleichwohl nimmt  $q(x)$  einen wohlbestimmten Wert an, wenn  $x \rightarrow 0$  geht. Die Reihenentwicklung der  $\sin$ -Funktion beginnt nämlich mit

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

Folglich hat man

$$q(x) = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \dots$$

so dass gilt  $q(x) \rightarrow 1$  für  $x \rightarrow 0$ . Hat man allgemein einen Quotienten

$$q(x) = \frac{f(x)}{g(x)}$$

zweier differenzierbarer Funktionen  $f(x)$  und  $g(x)$  mit  $f(0) = 0$ ,  $g(0) = 0$  und sei  $g'(0) \neq 0$ , dann gilt

$$(H) \quad \lim_{x \rightarrow 0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow 0} \frac{f'(x)}{g'(x)}.$$

Dies ist die Regel von  $\ell$ 'Hospital. (Benannt nach dem französischen Mathematiker Guillaume F.A. de  $\ell$ 'Hospital, 1661–1704). Ein Fall für die Regel von  $\ell$ 'Hospital ist zum Beispiel der Quotient

$$q(x) = \frac{e^{2x} - 1}{\sin x}.$$

Man erhält mit (H)

$$\begin{aligned} \lim_{x \rightarrow 0} q(x) &= \lim_{x \rightarrow 0} \frac{(e^{2x} - 1)'}{(\sin x)'} \\ &= \lim_{x \rightarrow 0} \frac{2e^{2x}}{\cos x} = 2. \end{aligned}$$

Die Richtigkeit der Regel (H) kann man folgendermaßen einsehen. Wegen  $f(0) = 0$ ,  $g(0) = 0$  und aufgrund des Mittelwertsatzes der Differentialrechnung (vgl. Abschnitt 6.1) kann man schreiben:

$$\begin{aligned} \frac{f(x)}{g(x)} &= \frac{f(x) - f(0)}{g(x) - g(0)} = \frac{f'(\xi_1)(x - 0)}{g'(\xi_2)(x - 0)} \\ &= \frac{f'(\xi_1)}{g'(\xi_2)} \end{aligned}$$

mit Zahlen  $\xi_1, \xi_2$  aus dem Intervall  $[0, x]$ . Für  $x \rightarrow 0$  rücken auch  $\xi_1, \xi_2 \rightarrow 0$ . Wegen  $g'(0) \neq 0$  ist dann der Grenzwert der rechten Seite ein wohlbestimmter Ausdruck und es folgt die Regel (H) von d'Hospital. Führt  $\lim_{x \rightarrow 0} f'(x)/g'(x)$  wiederum zu einem unbestimmten Ausdruck der Form  $0/0$ , dann wendet man (H) auf den Quotienten  $f'(x)/g'(x)$  an usw.. Stellt sich bei den Quotienten  $q(x)$  für  $x \rightarrow \infty$  der unbestimmte Ausdruck  $\infty/\infty$  ein und gilt  $\lim_{x \rightarrow \infty} g'(x) \neq 0$  dann gilt die analoge Regel

$$(H_\infty) \quad \lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = \lim_{x \rightarrow \infty} \frac{f'(x)}{g'(x)}.$$

Man betrachte zum Beispiel den Quotienten

$$q(x) = \frac{x^2}{e^x}$$

für  $x \rightarrow \infty$ . dann gilt wegen  $(H_\infty)$

$$\lim_{x \rightarrow \infty} q(x) = \lim_{x \rightarrow \infty} \frac{2x}{e^x}.$$

Die rechte Seite führt wieder zu einem unbestimmten Ausdruck der Form  $\infty/\infty$ . Nochmalige Anwendung der Regel  $(H_\infty)$  ergibt aber

$$\lim_{x \rightarrow \infty} q(x) = \lim_{x \rightarrow \infty} \frac{2}{e^x} = 0.$$

Durch  $n$ -fache Anwendung von  $(H_\infty)$  lässt sich für den Quotienten

$$q(x) = \frac{x^n}{e^x}, \quad n \in \mathbb{N},$$

in gleicher Weise die Beziehung

$$\lim_{x \rightarrow \infty} \frac{x^n}{e^x} = 0$$

nachweisen. Sie besagt: Die Exponentialfunktion  $e^x$  wächst schneller gegen  $\infty$  für  $x \rightarrow \infty$  ab jede Potenz  $x^n$ ,  $n \in \mathbb{N}$ . Der Beweis der Regel  $(H_\infty)$  bedarf keiner wesentlich neuen Überlegung. Die Regel lässt sich vielmehr leicht auf die Regel  $(H)$  zurückführen.

## Extremwerte

Unter Extremwerten einer Funktion versteht man lokale Maxima oder Minima der Funktion. Die Funktion  $f(x)$  hat an der Stelle  $x_0$  ein *lokales Maximum*, wenn für Werte  $x \neq x_0$  in der Nähe von  $x_0$  gilt:  $f(x) < f(x_0)$ , und sie hat dort ein *lokales Minimum* wenn gilt:  $f(x) > f(x_0)$ .

Die Anschauung zeigt: Hat die differenzierbare Funktion  $f(x)$  an der Stelle  $x_0$  einen Extremwert, dann hat sie dort eine waagerechte Tangente, das heißt, es gilt  $f'(x_0) = 0$ . (vgl. Abbildung 6.7)

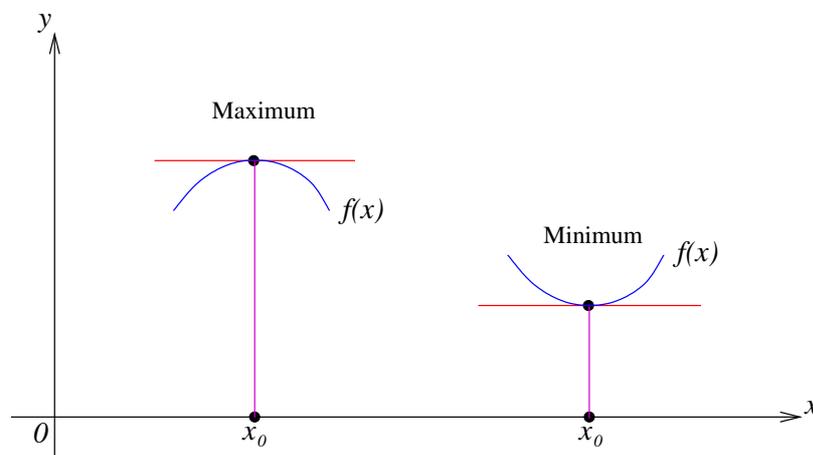


Abbildung 6.7: Lokales Maximum und Minimum einer Funktion.

Will man daher die Lage der Extremwerte einer differenzierbaren Funktion  $f(x)$  bestimmen, so muss man die Nullstellen der Funktion  $f'(x)$  berechnen, das heißt, die Gleichung  $f'(x) = 0$  lösen.

**Bemerkung:** In manchen Fällen führt bereits eine Symmetrieüberlegung zum Ziel. Die Funktion  $f(x) = x^2 + x^4$  ist spiegelsymmetrisch zur  $y$ -Achse,  $f(-x) = f(x)$ , und es gilt  $f(x) \geq 0$ . Sie bildet daher einen Trog mit dem „Bodenwert“  $f(0) = 0$  als Minimum. ▲

Am Beispiel der Funktion

$$f(x) = x^3 - 3x^2 + 2x,$$

da ist ein Polynom 3. Grades, gehen wir die Überlegungen zur Bestimmung von Extremwerten durch.

Offenbar kann man schreiben

$$f(x) = x(x^2 - 3x + 2).$$

Also hat  $f(x)$  die Nullstelle  $x_0 = 0$ . Die Klammer wird null für  $x_1 = 1$  und  $x_2 = 2$ . Daher ist

$$f(x) = x(x - 1)(x - 2)$$

die Zerlegung von  $f(x)$  in seine Linearfaktoren.

Da  $f(x)$  ein Polynom ungeraden Grades ist und der Term  $x^3$  positives Vorzeichen hat, muss der Graph von  $f$  von „links unten“ in den Nullpunkt des Koordinatensystems einmünden. Man ist damit in der Lage, den Verlauf der Funktion  $f(x)$  qualitativ zu skizzieren (vgl. Abbildung 6.8).

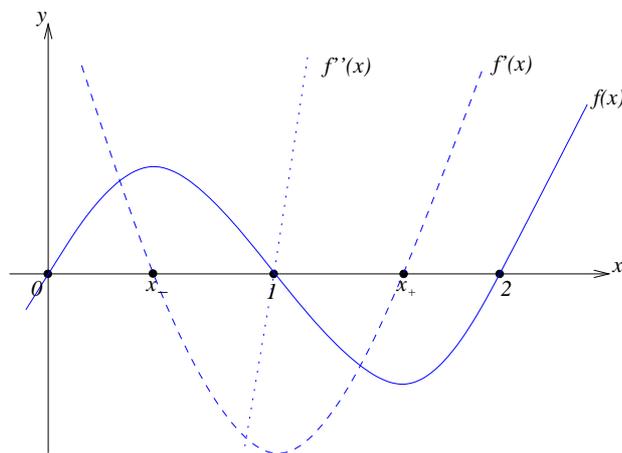


Abbildung 6.8: Die Funktion  $f(x) = x^3 - 3x^2 + 2x$  mit ihrer 1. und 2. Ableitung.

Dieser Verlauf lässt unmittelbar erkennen: Über dem Intervall  $[0, 1]$  befindet sich ein *Maximum* und unter dem Intervall  $[1, 2]$  ein *Minimum* der Funktion. Die genaue Lage

dieser Extremwerte findet man durch Lösung der Gleichung

$$f'(x) = 3x^2 - 6x + 2 = 0.$$

Die beiden Lösungen sind

$$x_{\pm} = 1 \pm \frac{1}{\sqrt{3}}.$$

Da  $x_-$  der kleinere und  $x_+$  der größere der beiden Werte ist, liegt über  $x_-$  das Maximum und über  $x_+$  das Minimum der Funktion  $f(x)$ .

Hat man sich über den Verlauf einer Funktion kein Bild gemacht, so bleibt zunächst offen, ob an einer Stelle  $x_0$  mit  $f'(x_0) = 0$  ein Maximum oder ein Minimum der Funktion vorliegt oder ob weder das eine noch das andere der Fall ist.

Dass letzteres durchaus möglich sein kann, zeigt das Beispiel der Funktion  $f(x) = x^3$ , der kubischen Parabel. Man hat  $f'(x) = 3x^2$  und daher  $f'(0) = 0$ . Gleichwohl befindet sich bei  $x_0 = 0$  weder ein Maximum noch ein Minimum der Funktion, sondern ein sogenannter „Sattelpunkt“ (vgl. Abb. 6.9).

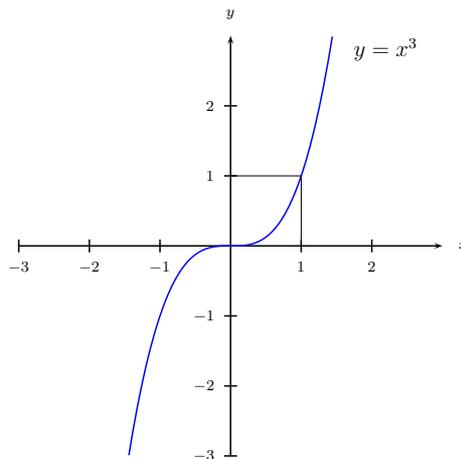


Abbildung 6.9: Die kubische Parabel hat einen Sattelpunkt bei  $x = 0$ .

Hat die Funktion  $f(x)$  bei  $x_0$  ein Maximum, dann gilt  $f'(x) > 0$  für Werte  $x < x_0$  nahe bei  $x_0$  und  $f'(x) < 0$  für Werte  $x > x_0$  nahe bei  $x_0$ . Es findet also bei  $x = x_0$  ein Vorzeichenwechsel der Ableitung  $f'(x)$  von  $+$  nach  $-$  statt. Eine hinreichende Bedingung für einen solchen Vorzeichenwechsel ist  $f''(x_0) < 0$ . Bei einem Minimum an der Stelle  $x_0$  findet entsprechend ein  $- +$  Vorzeichenwechsel von  $f'(x)$  statt. Hierzu ist  $f''(x_0) > 0$  eine hinreichende Bedingung. – Wir fassen zusammen (vgl. auch Abb. 6.8):

**Satz:** Hat die Ableitung der Funktion  $f(x)$  an der Stelle  $x_0$  eine Nullstelle,  $f'(x_0) = 0$ , und ist zusätzlich die zweite Ableitung an dieser Stelle negativ,  $f''(x_0) < 0$ , dann hat  $f(x)$  bei  $x_0$  ein Maximum. Ist  $f'(x_0) = 0$  und  $f''(x_0) > 0$ , dann liegt bei  $x_0$  ein Minimum vor.

**Bemerkung:** Der Satz beschreibt lediglich *hinreichende*, aber nicht notwendige Kriterien für das Vorliegen eines Extremwertes. So besitzt die Funktion  $f(x) = x^4$  bei  $x = 0$  offenbar ein Minimum. Gleichwohl gilt  $f''(0) = 0$ . In einem solchen Fall müssen dann Zusatzüberlegungen darüber entscheiden, ob ein Extremwert vorliegt und ob es sich um ein Maximum oder ein Minimum handelt. ▲

Eminente Bedeutung hat die Differentialrechnung auch bei der Lösung von Optimierungsproblemen aus der technischen und ökonomischen Praxis. Hierzu zwei Beispiele:

**Beispiel 1: Die optimale Pappschachtel.** Aus einem rechteckigen Stück Pappe mit Länge  $a$  und Breite  $b$ ,  $a > b$ , soll eine oben offene Schachtel mit maximalem Inhalt gefertigt werden.

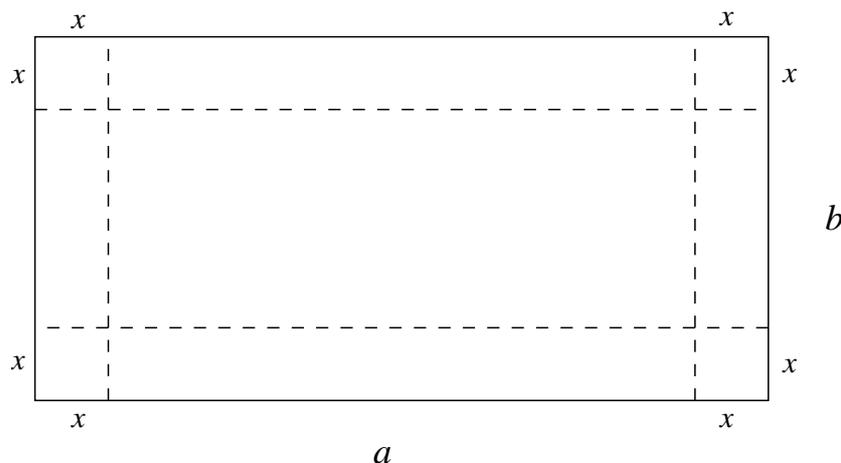


Abbildung 6.10: Herstellung einer Pappschachtel.

Hierzu schneidet man an den Ecken des Rechtecks quadratische Stücke mit der Seitenlänge  $x$  aus und biegt die überstehenden Seitenstücke nach oben. Die Variable  $x$  liegt also im Intervall  $[0, \frac{b}{2}]$ . So entsteht eine oben offene Schachtel mit der Länge  $a - 2x$ , der Breite  $b - 2x$  und der Höhe  $x$ . Sie besitzt den Inhalt

$$I(x) = (a - 2x)(b - 2x)x.$$

Die Höhe  $x$  soll so bestimmt werden, dass der Inhalt  $I(x)$  maximal wird.

Die Funktion  $I(x)$  ist ein Polynom 3. Grades mit dem führenden Term  $4x^3$  und den Nullstellen  $x_1 = 0$ ,  $x_2 = \frac{b}{2}$ ,  $x_3 = \frac{a}{2}$ , der Größe nach indiziert. Damit ist der Verlauf der Funktion  $I(x)$  qualitativ überschaubar.

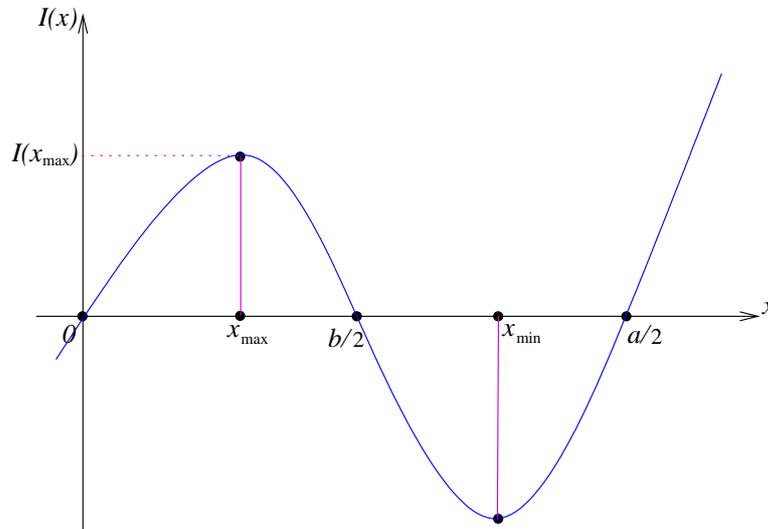


Abbildung 6.11: Der Verlauf der Inhaltsfunktion  $I(x)$ .

Aus diesem Verlauf lässt sich ablesen: Es gibt einen Wert  $x_{\max}$  mit  $0 < x_{\max} < \frac{b}{2}$ , so dass  $I(x_{\max})$  ein lokales Maximum der Funktion  $I(x)$  ist. Zwischen  $\frac{b}{2}$  und  $\frac{a}{2}$  hat  $I(x)$  ein lokales Minimum bei  $x_{\min}$ . Der Funktionswert  $I(x_{\min})$  ist aber negativ und daher ohne praktische Bedeutung.

Wegen

$$I(x) = 4x^3 - 2(a+b)x^2 + abx$$

ist

$$I'(x) = 12x^2 - 4(a+b)x + ab.$$

Die Werte  $x_{\max}$ ,  $x_{\min}$  mit  $x_{\max} < x_{\min}$  sind also die beiden Lösungen  $x_{\pm}$  der quadratischen Gleichung  $I'(x) = 0$ . Man erhält

$$\begin{aligned} x_{\pm} &= \frac{1}{6}(a+b) \pm \sqrt{\frac{(a+b)^2}{36} - \frac{ab}{12}} \\ &= \frac{1}{6}(a+b) \pm \frac{1}{6}\sqrt{a^2 + b^2 - ab}. \end{aligned}$$

Die kleinere dieser beiden Zahlen, also  $x_-$ , ist dann der gesuchte Wert:

$$x_{\max} = x_- = \frac{1}{6}(a+b) - \frac{1}{6}\sqrt{a^2 + b^2 - ab}.$$

Bei einem quadratischen Rechteck mit  $a = b = 1$  erhält man also

$$x_{\max} = \frac{1}{6} (2 - \sqrt{1}) = \frac{1}{6}$$

als optimale Höhe der zu bildenden Schachtel. ■

**Beispiel 2: Der optimale Bierpreis.** Ein Wirt überlegt sich, zu welchem Preis er sein Bier verkaufen soll. Derjenige Bierpreis ist für ihn optimal, bei dem er den höchsten Umsatz das heißt die größtmöglichen monatlichen Einnahmen erzielt.

Ist  $x \geq 0$  der Bierpreis in EUR und  $A(x)$  die von diesem abhängige Anzahl der in einem Monat verkauften Biere, dann ist  $U(x) = x \cdot A(x)$  der monatliche Umsatz in EUR. Dieser soll maximal werden.

Die Lösung unserer Frage läuft also darauf hinaus, das Maximum der Funktion  $U(x)$  zu bestimmen. Dass ein solches Maximum existiert, erscheint aufgrund folgender Überlegung plausibel: Ist der Bierpreis  $x = 0$ , so werden sich zwar viele Käufer finden, gleichwohl sind die monatlichen Einnahmen  $U(0) = 0 \cdot A(0)$  gleich null. Ebenso verhält es sich im Falle  $x = 100$ . Dann wird  $A(100) = 0$  sein und  $U(100) = 100 \cdot A(100)$  ebenfalls. Zwischen den Werten  $x = 0$  und  $x = 100$  wird daher der optimale Bierpreis liegen.

Entscheidend für die Lösung unseres Problems ist der Verlauf der Anzahlfunktion  $A(x)$ . Dieser lässt sich im Prinzip empirisch ermitteln. Wir begnügen uns hier mit plausiblen Annahmen über den Verlauf von  $A(x)$ :  $A(x)$  wird eine monoton fallende Funktion sein, mit größtem Wert bei  $x = 0$ . Aus solchen Funktionen besteht zum Beispiel die Schar der Funktionen

$$A(x) = ce^{-ax}, \quad a, c > 0.$$

Zwar ist  $A(x)$  auch für große  $x$  größer als null, aber es gilt immerhin  $x \cdot A(x) \rightarrow 0$  für  $x \rightarrow \infty$ , so dass mit geeigneten Werten für die Parameter  $c$  und  $a$  die Annahme über die Form von  $A(x)$  und damit auch die von  $U(x)$  einigermaßen realistisch erscheint. Wir erhalten also

$$U(x) = cxe^{-ax}$$

und damit ist

$$\begin{aligned} U'(x) &= ce^{-ax} - caxe^{-ax} \\ &= ce^{-ax}(1 - ax). \end{aligned}$$

Also hat  $U(x)$  bei  $x_{\max} = \frac{1}{a}$  einen Extremwert und zwar aufgrund der Vorüberlegungen ein absolutes Maximum. Bemerkenswert ist, dass der Parameter  $c = A(0)$  die

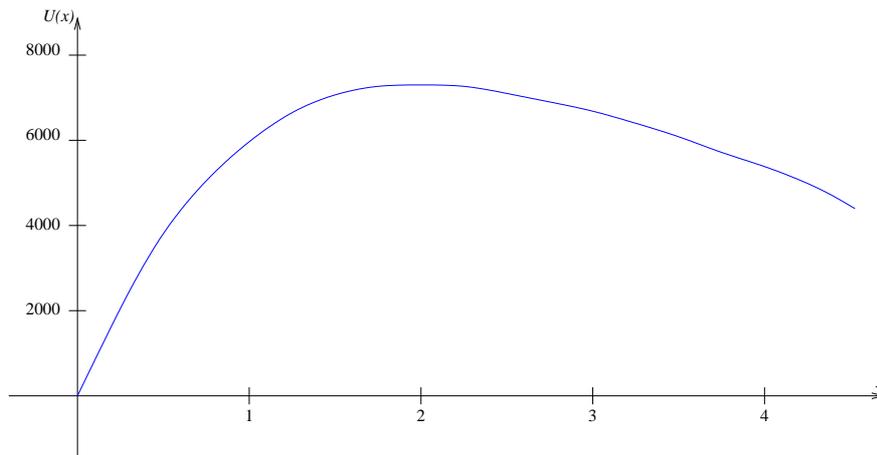


Abbildung 6.12: Die Funktion  $U(x)$  mit  $a = 0,5$  und  $c = 10\,000$ .

Lage des Extremwertes  $x_{\max}$  nicht beeinflusst.

Zur Bestimmung der Parameter  $a$  und  $c$  genügen zwei Anzahlen  $A(x_0)$  und  $A(x_1)$  zu Preisen  $x_0$  und  $x_1$ . Ist  $A(x_0)$  die Zahl der monatlich verkauften Biere, die der Wirt bei seinem üblichen Bierpreis  $x_0$  erzielt, dann genügt ein „Nostalgiemonat“ mit einem Bierpreis  $x_1 < x_0$ , um die erforderliche Information zur Berechnung von  $a$  und  $c$  zu erlangen. Man hat also

$$A(x_0) = ce^{-ax_0} = A_0$$

und

$$A(x_1) = ce^{-ax_1} = A_1.$$

Durch Division beider Gleichungen ergibt sich

$$e^{a(x_0-x_1)} = \frac{A_1}{A_0}$$

und damit ist

$$a = \frac{1}{x_0 - x_1} \log \frac{A_1}{A_0},$$

also

$$x_{\max} = \frac{x_0 - x_1}{\log A_1/A_0}.$$

Für den Parameter  $c$  findet man

$$\begin{aligned} c &= A_0 e^{ax_0} \\ &= A_0 e^{\frac{x_0}{x_0-x_1} \log \frac{A_1}{A_0}} \\ &= A_0 \cdot \left( \frac{A_1}{A_0} \right)^{\frac{x_0}{x_0-x_1}}. \end{aligned}$$

Verkauft zum Beispiel ein Wirt normalerweise sein Bier zum Preis von  $x_0 = 2,5$  EUR und erzielt damit die monatliche Verkaufszahl  $A_0 = 3000$ , beim Nostalgiepreis von  $x_1 = 1$  EUR hingegen die doppelte Anzahl  $A_1 = 6000$ , dann ist

$$x_{\max} = \frac{1,5}{\log 2} = 2,164$$

der optimale Bierpreis. Der Wirt würde also seinen Gewinn steigern, wenn er sein Bier für 2,20 EUR statt für 2,50 EUR verkaufen würde.

Da der Parameter  $c$  ebenfalls bekannt ist, lassen sich auch die monatlichen Einnahmen bei den unterschiedlichen Preisen berechnen. So sind bei  $x = x_{\max}$  die Einnahmen

$$U(x_{\max}) = U\left(\frac{1}{a}\right) = ce^{-1}.$$

Fragt man bei unserer Optimierungsaufgabe nicht nach dem maximalen Umsatz, sondern nach dem maximal zu erzielenden Gewinn, dann hat man die Funktion

$$G(x) = U(x) - K(x)$$

zu betrachten. Dabei sind  $K(x)$  die anfallenden monatlichen Kosten des Wirtes. Ist  $p$  der Einkaufspreis eines Bieres und  $k$  die monatlichen Fixkosten, dann ist

$$K(x) = pA(x) + k$$

die Form der Kostenfunktion (wenn durch Steuern entstehende und weitere Kosten nicht berücksichtigt werden). Die Funktion  $G(x)$  hat ihr Maximum bei  $x = \frac{1}{a} + p$ , wie zu erwarten war und wie man leicht nachrechnet. ■

## Krümmung; Wendepunkte

Die erste Ableitung  $f'(x)$  einer differenzierbaren Funktion  $f(x)$  gibt Auskunft darüber, ob die Funktion im Punkt  $(x, f(x))$  ansteigt ( $f'(x) > 0$ ), fällt ( $f'(x) < 0$ ), oder ob an dieser Stelle ein Extremwert, oder eventuell ein Flachpunkt, vorliegt ( $f'(x) = 0$ ).

Die zweite Ableitung  $f''(x)$  lässt daher erkennen, ob die Steigung  $f'(x)$  über einem Intervall  $I$  zunimmt ( $f''(x) > 0$ ) oder abnimmt ( $f''(x) < 0$ ). Im ersten Fall zeigt der Graph der Funktion eine *Linkskrümmung*, im zweiten eine *Rechtskrümmung*.

Der Punkt  $x_0$ , in dem eine Rechtskrümmung in eine Linkskrümmung übergeht, ist ein *Wendepunkt* der Funktion. Dieser ist zugleich ein Extremwert der Ableitung  $f'(x)$ . Dort gilt also  $f''(x) = 0$ .

Eine *hinreichende* Bedingung dafür, dass an der Stelle  $x_0$  eine Rechtskrümmung in eine Linkskrümmung übergeht, ist  $f'''(x_0) \neq 0$ . Dann gilt nämlich in einer Umgebung

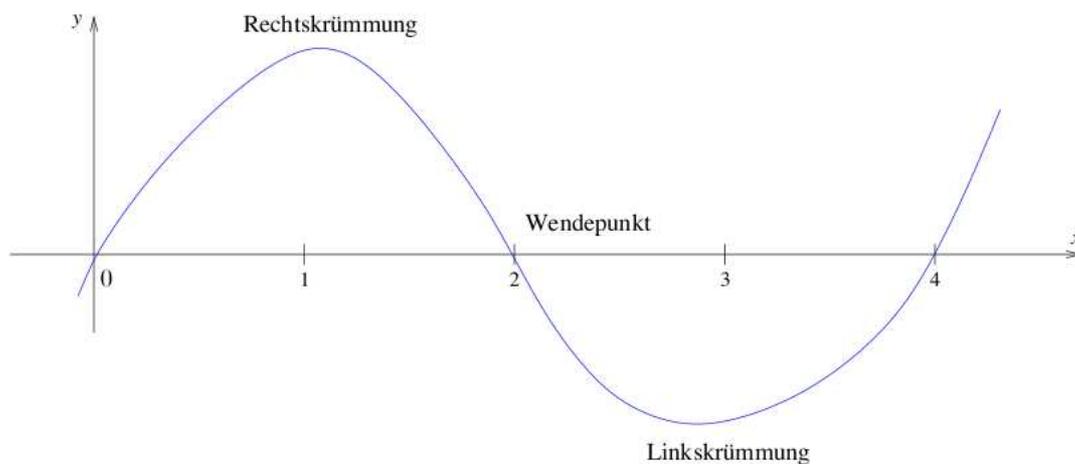


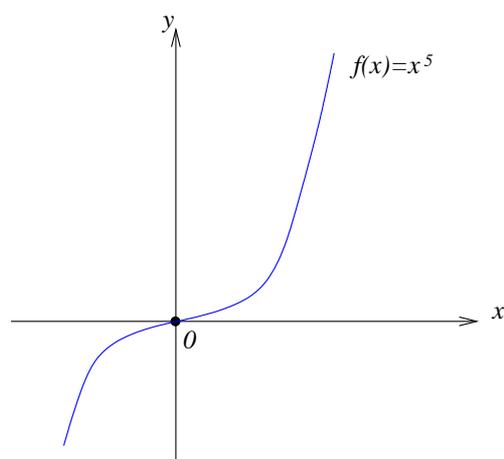
Abbildung 6.13: Das Krümmungsverhalten einer Funktion.

von  $x_0$  entweder  $f''(x) > 0$  für  $x < x_0$  und  $f''(x) < 0$  für  $x > x_0$  oder  $f''(x) < 0$  für  $x < x_0$  und  $f''(x) > 0$  für  $x > x_0$ .

Damit erhalten wir folgendes hinreichendes *Kriterium* für das Vorliegen eines Wendepunktes:

*Ist  $f''(x_0) = 0$  für eine dreimal differenzierbare Funktion  $f(x)$  und  $f'''(x_0) \neq 0$ , dann hat  $f(x)$  an der Stelle  $x_0$  einen Wendepunkt.*

Dass die Bedingung  $f'''(x_0) \neq 0$  nicht notwendig erfüllt sein muss, obwohl bei  $x_0$  ein Wendepunkt vorliegt, zeigt das Beispiel der Funktion  $f(x) = x^5$ .

Abbildung 6.14: Die Funktion  $f(x) = x^5$ .

Es gilt  $f''(x) = 20x^3$ . Daher ist  $f''(x) < 0$  für  $x < 0$  und  $f''(x) > 0$  für  $x > 0$ . Also ist  $x_0 = 0$  ein Wendepunkt der Funktion. Gleichwohl ist  $f'''(0) = 0$ .

### Stabile Gleichgewichtslage beim logistischen Wachstum

Die Rekursionsformel für das logistische Wachstum lautet (vgl. Abschnitt 4.2)

$$(R) \quad x_n = \frac{1}{K - E} (K - x_{n-1})x_{n-1}.$$

Dabei ist  $K > 0$  die maximale Größe, welche die Population erreichen kann und  $E \leq 3/4 K$  ist die Gleichgewichtslage.

Mit Hilfe der Linearisierung der obigen Rekursionsformel in der Nähe der Gleichgewichtslage und einer Plausibilitätsüberlegung ergab sich die Stabilitätsbedingung

$$(S) \quad E < \frac{2}{3}K.$$

Sie besagt: Wenn die Ungleichung (S) erfüllt ist und wenn die Populationsgröße  $x_n$  irgendwann dem Wert  $E$  nur genügend nahe kommt, dann streben die Werte  $x_{n+1}, x_{n+2}, \dots$  gegen  $E$ .

Wir wollen nun die vage Formulierung „genügend nahe“ mit Hilfe der Differentialrechnung präzisieren und darüberhinaus zeigen, dass bei *beliebigem* Startwert  $x_0$  die Werte  $x_n$  *immer* gegen die Gleichgewichtslage  $E$  streben, sofern nur die Stabilitätsbedingung (S) erfüllt ist.

Doch zuvor gehen wir noch auf eine sehr instruktive graphische Darstellung der durch die Rekursionformel (R) erzeugten Iteration ein.

Mit der Funktion

$$f(x) = \frac{1}{K - E} (K - x)x$$

kann man (R) auch in der Form

$$x_n = f(x_{n-1})$$

schreiben. Nun ist  $y = f(x)$  die Gleichung einer nach unten offenen Parabel, die an den Stellen  $x = 0$  und  $x = K$  die  $x$ -Achse schneidet. Für die Gleichgewichtslage gilt  $f(E) = E$ . Der Punkt  $(E, f(E))$  ist also der Schnittpunkt des Graphen von  $f$  mit der Geraden  $y = x$ , der Winkelhalbierenden des ersten Quadranten. Zeichnet man die Parabel in ein cartesisches Koordinatensystem ein zusammen mit der Geraden  $y = x$ , dann lässt sich, bei einem Startpunkt  $x_0, 0 < x_0 < K$ , der Gang der Iteration graphisch sehr schön verfolgen: Um den Wert  $x_1 = f(x_0)$  zu erhalten, geht man vom Punkt  $x_0$  senkrecht nach oben zum Graphen  $G$  von  $f$ . Um den Wert  $x_1$  auf die  $x$ -Achse zu übertragen geht man sodann waagrecht zur Winkelhalbierenden  $W$ . Unterhalb des dort erreichten Punktes liegt dann auf der  $x$ -Achse der Wert

$x_1 = f(x_0)$ . Dann geht man von der Winkelhalbierenden wieder zum Graphen und dann wieder zur Winkelhalbierenden. Unterhalb des dort erreichten Punktes liegt dann auf der  $x$ -Achse der Wert  $x_2 = f(x_1)$  usw.. Auf diese Weise durchläuft man eine Folge  $x_0 \rightarrow G \rightarrow W \rightarrow G \rightarrow W \rightarrow \dots$ . Unterhalb der Punkte auf  $W$  liegen dann die iterierten Werte  $x_0, x_1, x_2, \dots$  der sukzessiven Populationsgrößen. Die Abbildung 6.15 zeigt die zeitliche Entwicklung der instabilen Population aus Abb. 4.1.

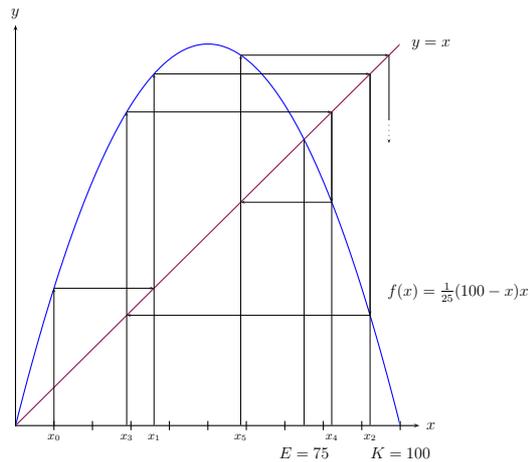


Abbildung 6.15: Entwicklung einer Population mit nicht-stabiler Gleichgewichtslage

Die Stabilität der Gleichgewichtslage  $E$  steht mit der Ableitung der Funktion  $f$  an der Stelle  $E$  in folgendem Zusammenhang:

**Satz:** Wenn gilt  $-1 < f'(E) < 1$  dann ist die Gleichgewichtslage  $E$  stabil.

Dies sieht man folgendermaßen ein: Wenn  $-1 < f'(E) < 1$  ist und  $f'(x)$  eine stetige Funktion ist, dann gibt es eine ganze Umgebung  $U(E)$  von  $E$ , so dass für alle  $x \in U(E)$  ebenfalls gilt  $-1 < f'(x) < 1$ . Wenn ebenfalls der Startwert  $x_0$  aus  $U(E)$  ist, dann erhält man wegen  $f(E) = E$  und aufgrund des Mittelwertsatzes der Differentialrechnung (s. Abschnitt 6.1):

$$\begin{aligned} x_1 - E &= f(x_0) - f(E) \\ &= f'(\xi_0)(x_0 - E) \end{aligned}$$

mit  $\xi_0$  aus  $[x_0, E]$ . Da der Faktor  $f'(\xi_0)$  dem Betrage nach kleiner als eins ist, liegt der Punkt  $x_1$  näher an  $E$  als der Startpunkt  $x_0$ . Dann folgt weiter

$$\begin{aligned} x_2 - E &= f(x_1) - f(E) = f'(\xi_1)(x_1 - E) \\ &= f'(\xi_1)f'(\xi_0)(x_0 - E) \end{aligned}$$

mit  $\xi_1$  aus  $[x_1, E]$  und

$$\begin{aligned} x_3 - E &= f(x_2) - f(E) = f'(\xi_2)(x_2 - E) \\ &= f'(\xi_2)f'(\xi_1)f'(\xi_0)(x_0 - E) \end{aligned}$$

mit  $\xi_2$  aus  $[x_2, E]$ . Man erkennt: Der Punkt  $x_2$  liegt näher an  $E$  als der Punkt  $x_1$  und  $x_3$  liegt wiederum näher an  $E$  als  $x_2$ . Allgemein ergibt sich auf diese Weise die Beziehung

$$x_n - E = f'(\xi_0)f'(\xi_1) \cdots f'(\xi_{n-1})(x_0 - E)$$

Wenn für alle  $x$  aus  $U(E)$  und einem  $k$  mit  $0 < k < 1$  gilt

$$|f'(x)| \leq k < 1,$$

so folgt

$$|x_n - E| \leq k^n |x_0 - E|$$

und damit folgt wegen  $k^n \rightarrow 0$  für  $n \rightarrow \infty$ , dass  $x_n \rightarrow E$  geht für  $n \rightarrow \infty$ . Die Punkte  $x_n$  werden also von der Gleichgewichtslage  $E$  gleichsam „angezogen“. Da der Wert  $E$  wegen  $f(E) = E$  auch als sogenannter „Fixpunkt“ der Funktion  $f$  angesehen werden kann, spricht man im vorliegenden Fall auch von einem „anziehenden Fixpunkt“.

**Bemerkung:** Die Überlegungen zum Beweis des obigen Satzes gelten natürlich auch für Iterationen  $x_n = f(x_{n-1})$  mit einer beliebigen stetig differenzierbaren Funktion  $f(x)$  und einem Fixpunkt  $E = f(E)$  bei dem gilt  $|f'(E)| < 1$ . ▲

Für die Funktion

$$f(x) = \frac{1}{K - E} (K - x)x$$

hat man

$$f'(x) = \frac{1}{K - E} (K - 2x).$$

Also ist

$$\begin{aligned} f'(E) &= \frac{1}{K - E} (K - 2E) \\ &= 1 - \frac{E}{K - E} < 1. \end{aligned}$$

Die Bedingung  $f'(E) < 1$  ist daher wegen  $E/(K - E) > 0$  von selbst erfüllt. Die Bedingung  $f'(E) > -1$  ist erfüllt, wenn

$$\frac{E}{K - E} < 2$$

oder

$$E < 2K - 2E,$$

also

$$E < \frac{2}{3}K$$

ist. Diese Stabilitätsbedingung kennen wir bereits aus Abschnitt 4.2.

Zur genauen Bestimmung des Anziehungsbereichs  $U(E)$  setzen wir  $f'(x) = 1$  und  $f'(x) = -1$ . Im ersten Fall erhält man

$$\frac{1}{K-E}(K-2x) = 1$$

oder

$$K - 2x = K - E,$$

also

$$x = \frac{E}{2}.$$

Im zweiten Fall ergibt sich die Bedingung

$$\frac{1}{K-E}(K-2x) = -1$$

oder

$$K - 2x = E - K,$$

also

$$x = K - \frac{E}{2}.$$

Daher bildet das Intervall

$$U(E) = \left\{ x \mid \frac{E}{2} < x < K - \frac{E}{2} \right\}$$

den Anziehungsbereich der Gleichgewichtslage  $E$ .

Wir haben gezeigt: Jede Population mit einem Startwert  $x_0$  aus  $U(E)$  kommt im Laufe der Zeit der Gleichgewichtslage  $E$  beliebig nahe. Es gilt aber auch: Jede Population, die mit einem Wert  $x_0$  *außerhalb* von  $U(E)$  startet, gelangt nach endlich vielen Schritten ins Innere von  $U(E)$ . Das bedeutet, die Punkte im Intervall  $[0, K]$ , die auf dem Rand von  $U(E)$  oder außerhalb von  $U(E)$  liegen, gehören sämtlich zum „Einzugsbereich“ von  $U(E)$ .

Um dies einzusehen unterteilen wir das Intervall  $[0, K]$  wie folgt: Die Gerade  $x = E/2$  schneide die Winkelhalbierende  $y = x$  im Punkte  $S$  und die Parallele zur  $x$ -Achse durch  $S$  den Graphen von  $f$  in den Punkten  $A$  und  $B$ . Die Parallelen zur  $y$ -Achse

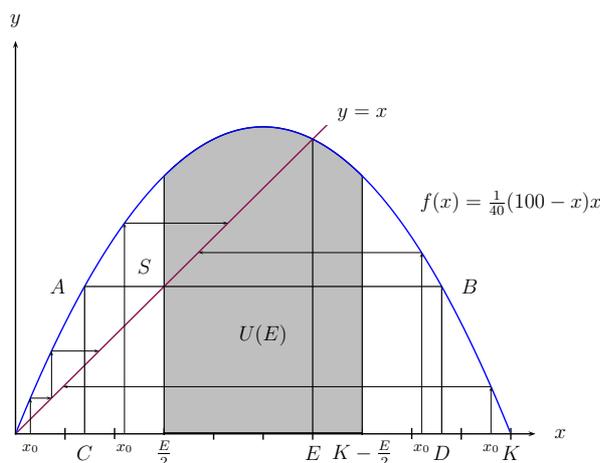


Abbildung 6.16: Der Anziehungsbereich  $K(E)$  und sein Einzugsgebiet

durch  $A$  und  $B$  schneiden die  $x$ -Achse in den Punkten  $C$  und  $D$  (s. Abb. 6.16).

Anschaulich ist unmittelbar klar: Liegt der Startpunkt  $x_0$  im Intervall  $[C, \frac{E}{2}]$ , dann liegt  $x_2$  in  $U(E)$ . Liegt  $x_0$  in  $[0, C]$ , dann liegt nach endlich vielen Schritten ein iterierter Wert  $x_n$  in  $[C, \frac{E}{2}]$ . Es gilt nämlich dann

$$\begin{aligned} x_1 - 0 &= f(x_0) - f(0) \\ &= f'(\xi_0)(x_0 - 0) \end{aligned}$$

mit  $\xi_0$  aus  $[0, x_0]$ . Also erhalten wir

$$x_1 = f'(\xi_0)x_0.$$

Da für alle  $x$  aus  $[0, C]$  mit einem  $k > 1$  gilt  $f'(x) \geq k$  so folgt

$$x_1 \geq kx_0$$

und ebenso

$$x_2 \geq k^2x_0$$

usw. Die iterierten Werte  $x_0, x_1, x_2, \dots$ , wandern also schließlich aus dem Intervall  $[0, C]$  aus. Liegt  $x_0$  in  $[K - E/2, D]$  dann liegt  $x_2$  in  $U(E)$ . Liegt  $x_0$  in  $[D, K]$ , dann liegt  $x_1$  in  $[0, C]$ . Damit ist gezeigt, eine Punktfolge  $x_0, x_1, x_2, \dots$ , mit beliebigem Startwert  $x_0 \in [0, K]$  wandert nach endlich vielen Schritten in den Anziehungsbereich  $U(E)$  ein.

Die Abbildung 6.17 zeigt die zeitliche Entwicklung einer Population mit  $K = 100$  und Startwert  $x_0 = 90$  in Richtung auf die stabile Gleichgewichtslage  $E = 40$ . (vgl. auch Abb. 4.2).

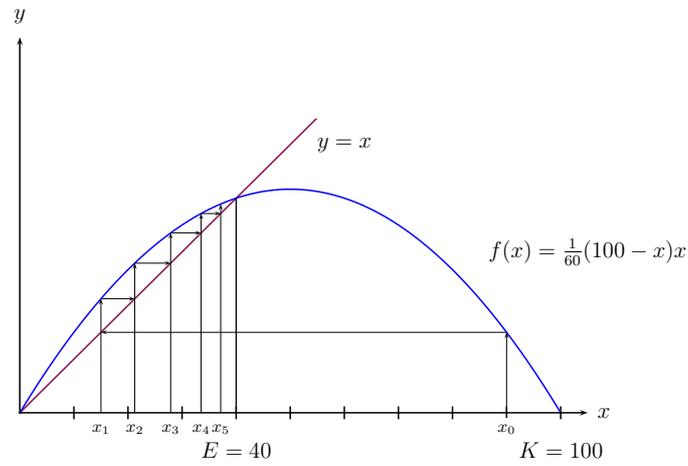


Abbildung 6.17: Stabile Population mit einseitiger Annäherung an die Gleichgewichtslage

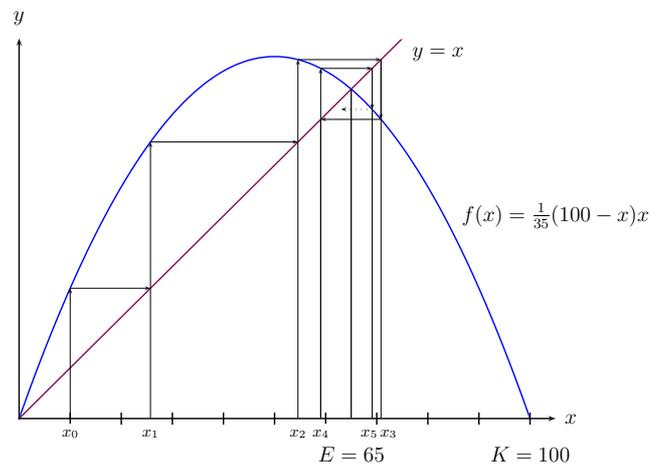


Abbildung 6.18: Stabile Population mit zweiseitiger Annäherung an die Gleichgewichtslage

Liegt die Gleichgewichtslage  $E$  dicht unterhalb von  $\frac{2}{3}K$ , dann „spiralt“ die Iterationsgraphik um den Punkt  $(E, f(E))$  (s. Abbildung 6.18). Nach endlich vielen Schritten liegen dann die Werte  $x_n$  abwechselnd unterhalb und oberhalb von  $E$ .

# Kapitel 7

## Integralrechnung

Die Integralrechnung entstand aus dem Problem, den Inhalt krummlinig berandeter Flächen, bzw. das Volumen von Körpern beliebiger Form zu bestimmen. Historisch gesehen sind das Flächenproblem und das Volumenproblem gegenüber dem Tangentenproblem die älteren Probleme. Im üblichen Aufbau der Analysis kehrt sich die Reihenfolge jedoch um.

Bereits in der Antike konnte man bestimmte Flächeninhalte und Volumina berechnen. Hervorzuheben sind insbesondere die Beiträge des wohl bedeutendsten Mathematikers der Antike, des Archimedes von Syrakus (287 - 212 v. Chr.). Er bestimmte Umfang und Fläche des Kreises, den Inhalt eines Parabelsegmentes und das Volumen von Kugel und Zylinder. Auf seinem Grabstein wurde auf seinen Wunsch eine Figur eingraviert, die eine Kugel mit einem sie umschließenden geraden Zylinder zeigt, dessen Höhe gleich dem Durchmesser der Kugel ist. Die Figur verweist auf eines der bemerkenswertesten Ergebnisse des Archimedes, dass nämlich das Verhältnis von Kugeloberfläche zu Zylinderoberfläche, sowie von Kugelvolumen zu Zylindervolumen jedesmal gleich zwei zu drei ist.

### 7.1 Eine Flächenberechnung

Die Formel zur Berechnung des Inhaltes *geradlinig* berandeter Flächen wie des Rechtecks, des Dreiecks oder des Trapezes, sind aus dem Geometrieunterricht bekannt. Wie aber berechnet man zum Beispiel den Inhalt einer Fläche, deren eine Seite ein Parabelstück ist?

Am Beispiel der Fläche  $F_0^1$  zwischen der Strecke  $[0, 1]$  und dem darüberliegenden Bogen der Standardparabel  $f(x) = x^2$  erläutern wir exemplarisch die Idee und das Verfahren zur Berechnung des Inhaltes krummlinig berandeter Flächen.

Man unterteilt hierzu die Strecke  $[0, 1]$  in  $n$  gleiche Teile durch die Punkte  $x_j = \frac{j}{n}$ ,

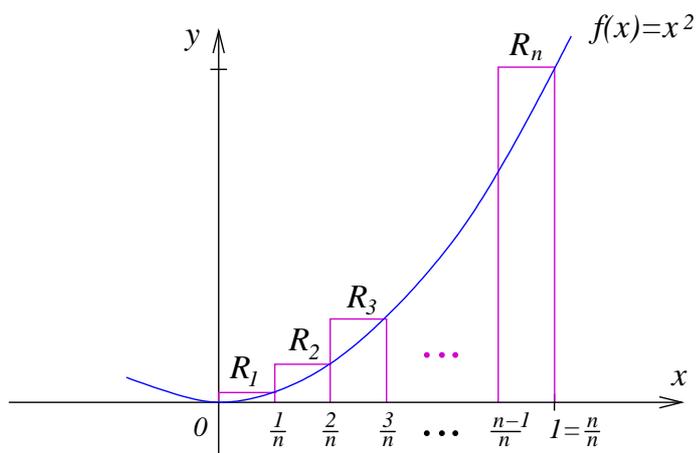


Abbildung 7.1: Berechnung der Fläche unterhalb eines Parabelstücks.

$j = 1, 2, \dots, n$ , und bildet zu jeder Teilstrecke  $[\frac{j-1}{n}, \frac{j}{n}]$  das Rechteck  $R_j$  mit der Höhe  $f(\frac{j}{n}) = (\frac{j}{n})^2$ . Das Rechteck  $R_j$  hat dann den Inhalt  $\frac{1}{n} \cdot f(\frac{j}{n}) = \frac{1}{n} \cdot (\frac{j}{n})^2$ . Die Summe

$$\begin{aligned}
 F_{0,n}^1 &= \sum_{j=1}^n \frac{1}{n} f\left(\frac{j}{n}\right) \\
 (*) \quad &= \sum_{j=1}^n \frac{1}{n} \left(\frac{j}{n}\right)^2 \\
 &= \frac{1}{n^3} \sum_{j=1}^n j^2
 \end{aligned}$$

dieser Rechteckflächen ist für große  $n$  eine Näherung für den gesuchten Wert  $F_0^1$ . Dieser ergibt sich als Grenzwert

$$F_0^1 = \lim_{n \rightarrow \infty} F_{0,n}^1,$$

sofern dieser Grenzwert existiert, was aber intuitiv plausibel ist. Es gilt also, diesen Grenzwert zu berechnen.

Die letzte Zeile der Formel (\*) zeigt, dass hierzu die Kenntnis der Summe

$$S_n = 1^2 + 2^2 + 3^2 + \dots + n^2$$

der ersten  $n$  Quadratzahlen erforderlich ist. Die entsprechende Formel (Beweis im Anhang am Ende des Abschnitts) lautet

$$S_n = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}.$$

Damit erhält man

$$\begin{aligned} F_{0,n}^1 &= \frac{1}{n^3} \cdot \left( \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6} \right) \\ &= \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2}, \end{aligned}$$

also ist

$$\begin{aligned} F_{0,n}^1 &= \lim_{n \rightarrow \infty} \left( \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} \right) \\ &= \frac{1}{3} \end{aligned}$$

der gesuchte Flächeninhalt. Die von dem Bogen der Standardparabel begrenzte Fläche füllt daher ein Drittel des Quadrates über der Strecke  $[0, 1]$  aus.

Wendet man das beschriebene Verfahren auf eine Fläche bezüglich einer beliebigen Funktion  $f$  über dem Intervall  $[0, 1]$  (oder analog über einem Intervall  $[a, b]$ ) an, dann hat man eine Summe der Form

$$\frac{1}{n} \sum_{j=1}^n f\left(\frac{j}{n}\right)$$

zu berechnen, wozu es kein allgemeingültiges Verfahren gibt. Dieser Schwierigkeit enthebt uns ein zentraler Satz der Analysis, der Hauptsatz der Differential- und Integralrechnung, der im nächsten Abschnitt behandelt wird.

### Anhang: Die Summe der ersten $n$ Quadratzahlen

In Analogie zu der Formel

$$1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$$

(vgl. Abschnitt 4.4) suchen wir einen geschlossenen Ausdruck für die Summe

$$S_n = 1^2 + 2^2 + \cdots + n^2$$

der ersten  $n$  Quadratzahlen. Hierzu betrachten wir die Funktion

$$(*) \quad \varphi(x) = \sum_{k=0}^n x^k = 1 + x + x^2 + \cdots + x^n = \frac{x^{n+1} - 1}{x - 1},$$

das ist die endliche geometrische Reihe (vgl. Abschnitt 4.4).

Für  $x = 1$  ergibt die rechte Seite von  $(*)$  den unbestimmten Ausdruck  $0/0$ . Diese Unbestimmtheit lässt sich folgendermaßen beheben: Aufgrund des binomischen Lehrsatzes (vgl. Abschnitt 3.4) kann man schreiben

$$\begin{aligned} x^{n+1} - 1 &= (x - 1 + 1)^{n+1} - 1 \\ &= \sum_{k=1}^{n+1} \binom{n+1}{k} (x-1)^k, \end{aligned}$$

so dass

$$(*) \quad \varphi(x) = \frac{x^{n+1} - 1}{x - 1} = \sum_{k=1}^{n+1} \binom{n+1}{k} (x-1)^{k-1}$$

ist. Nun ist aber

$$\begin{aligned} \varphi''(x) &= \sum_{k=1}^n k(k-1)x^{k-2} \\ &= \sum_{k=1}^n k^2 x^{k-2} - \sum_{k=1}^n kx^{k-2}, \end{aligned}$$

also

$$(**) \quad \begin{aligned} \varphi''(1) &= S_n - 1^2 - \frac{1}{2}n(n+1) + 1 \\ &= S_n - \frac{1}{2}n(n+1). \end{aligned}$$

Andererseits gilt wegen (\*)

$$\varphi''(x) = \sum_{k=1}^{n+1} \binom{n+1}{k} (k-1)(k-2)(x-1)^{k-3},$$

so dass

$$\varphi''(1) = 2 \cdot 1 \binom{n+1}{3}$$

ist. Wegen (\*\*) hat man also

$$\begin{aligned} S_n &= \varphi''(1) + \frac{1}{2}n(n+1) \\ &= \frac{(n+1)n(n-1)}{3} + \frac{1}{2}n(n+1) \\ &= \frac{1}{6}n(n+1)(2n+1) \\ &= \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}. \end{aligned}$$

Dies war zu zeigen.

## 7.2 Der Hauptsatz der Differential- und Integralrechnung

Wir betrachten eine stetige Funktion  $f(x)$  (das ist, wie wir wissen, eine Funktion, deren Graphen man zeichnen kann, ohne den Stift oder die Kreide abzusetzen) über dem Intervall  $[0, x]$ . Die rechte Grenze  $x$  des Intervalls  $[0, x]$  sei als variabel gedacht. Dann ist der Inhalt  $F(x)$  der Fläche zwischen Intervall und Funktionsgraph eine Funktion von  $x$ , wenn die linke Intervallgrenze festgehalten wird.

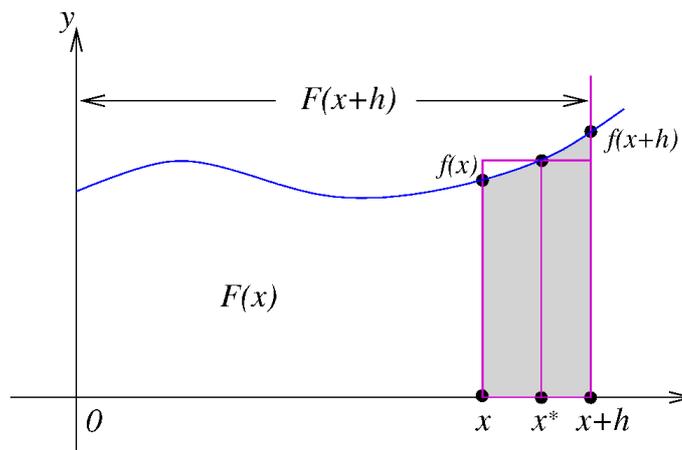


Abbildung 7.2: Die Flächen  $F(x)$  und  $F(x+h)$ .

Man denke sich nun das Intervall  $[0, x]$  um die Strecke  $h > 0$  zum Intervall  $[0, x+h]$  vergrößert. Die Differenz  $F(x+h) - F(x)$  ist dann der Inhalt der in Abbildung 7.2 grau unterlegten schmalen Fläche. Aufgrund der Stetigkeit der Funktion  $f(x)$  gibt es eine Stelle  $x^*$  im Intervall  $[x, x+h]$ , so dass das schmale Rechteck mit den Seiten  $h$  und  $f(x^*)$  denselben Inhalt besitzt wie die grau unterlegte, vom Funktionsgraphen oben begrenzte Fläche. Das heißt, es gilt

$$F(x+h) - F(x) = hf(x^*)$$

mit  $x \leq x^* \leq x+h$ . Nach Division durch  $h$  erhält man die Beziehung

$$(*) \quad \frac{F(x+h) - F(x)}{h} = f(x^*),$$

deren linke Seite der Differenzenquotient der Funktion  $F(x)$  ist (vgl. Abschnitt 6.1). Für  $h \rightarrow 0$  strebt daher die linke Seite von (\*) gegen  $F'(x)$  und aufgrund der Ungleichung  $x \leq x^* \leq x+h$  die rechte Seite gegen  $f(x)$ . Wir haben daher das Resultat

$$F'(x) = f(x),$$

das heißt, die Ableitung der „Flächenfunktion“  $F(x)$  ist die „berandende Funktion“  $f(x)$ .

Betrachtet man nunmehr die Fläche zu  $f(x)$  über einem Intervall  $[a, b]$ , ihr Inhalt sei  $F_a^b$ , dann gilt offenbar (vgl. Abbildung 7.3)

$$F_a^b = F(b) - F(a).$$

Man schreibt auch

$$F(x) \Big|_a^b = F(b) - F(a).$$

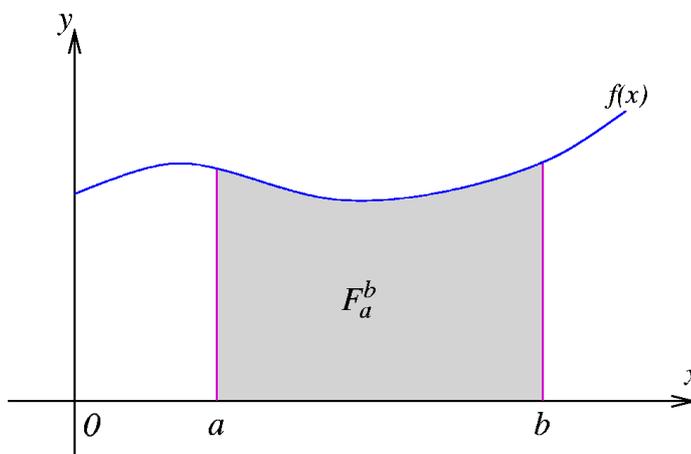


Abbildung 7.3: Die Fläche  $F_a^b$ .

Es stellt sich die Frage: Gibt es neben der Flächenfunktion  $F(x)$  noch weitere Funktionen  $F^*(x)$  mit der Eigenschaft  $F^{*'}(x) = f(x)$ ? Wir nehmen an, dies sei der Fall und betrachten die Differenz

$$D(x) = F^*(x) - F(x).$$

Dann gilt

$$\begin{aligned} D'(x) &= F^{*'}(x) - F'(x) \\ &= f(x) - f(x) = 0. \end{aligned}$$

Aufgrund einer Schlussfolgerung aus dem Mittelwertsatz der Differentialrechnung (vgl. Abschnitt 6.1) können wir daher schließen, die Funktion  $D(x)$  ist eine Konstante. Also gilt

$$F^*(x) - F(x) = c$$

oder

$$F^*(x) = F(x) + c.$$

Das bedeutet, zwei Funktionen  $F, F^*$  mit  $F' = F^{*'} = f$  unterscheiden sich höchstens um eine Konstante.

**Definition:** Eine Funktion  $F(x)$  mit der Eigenschaft  $F'(x) = f(x)$  heißt *Stammfunktion* zu  $f(x)$ .

Wie wir gesehen haben, können sich zwei Stammfunktionen  $F(x)$  und  $F^*(x)$  zu einer Funktion  $f(x)$  nur um eine Konstante unterscheiden und wegen

$$\begin{aligned} F_a^b &= F(b) - F(a) \\ &= (F(b) + c) - (F(a) + c) \\ &= F^*(b) - F^*(a) \end{aligned}$$

ist es gleichgültig, welche Stammfunktion wir zur Berechnung der Fläche  $F_a^b$  verwenden. Wir halten fest:

**Satz:** Ist  $F(x)$  eine Stammfunktion zur Funktion  $f(x)$ , dann ist

$$F_a^b = F(b) - F(a)$$

der Inhalt der zu  $f(x)$  gehörigen Fläche über dem Intervall  $[a, b]$ .

Das ist der Inhalt des *Hauptsatzes der Differential- und Integralrechnung*. Seine Pointe besagt: Zur Flächenberechnung bedarf es keiner – zuweilen mühsamer – Grenzwertberechnung. Es genügt vielmehr, zur Funktion  $f(x)$  eine Stammfunktion  $F(x)$  zu finden.

## Beispiele

Die Fläche  $F_1^2$  zur kubischen Parabel  $f(x) = x^3$  über dem Intervall  $[1, 2]$  sei zu berechnen. Eine Stammfunktion zu  $f(x) = x^3$  ist offenbar  $F(x) = x^4/4$ . Daher findet man

$$\begin{aligned} F_1^2 &= F(2) - F(1) \\ &= \frac{2^4}{4} - \frac{1^4}{4} \\ &= \frac{15}{4} \end{aligned}$$

als den gesuchten Flächeninhalt (vgl. Abbildung 7.4).

Die Funktionen  $f(x) = \sin x$  und  $g(x) = \frac{1}{2} \sin x$  schließen über dem Intervall  $[0, \pi]$  eine sichelförmige Fläche mit Inhalt  $S$  ein (vgl. Abbildung 7.5).

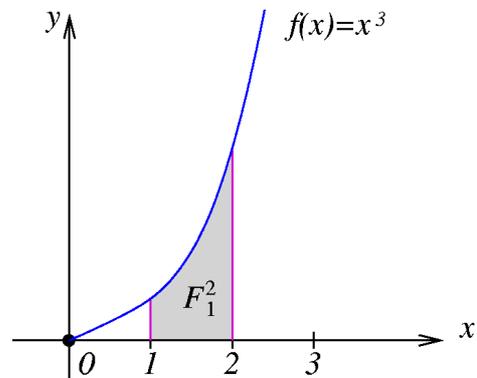
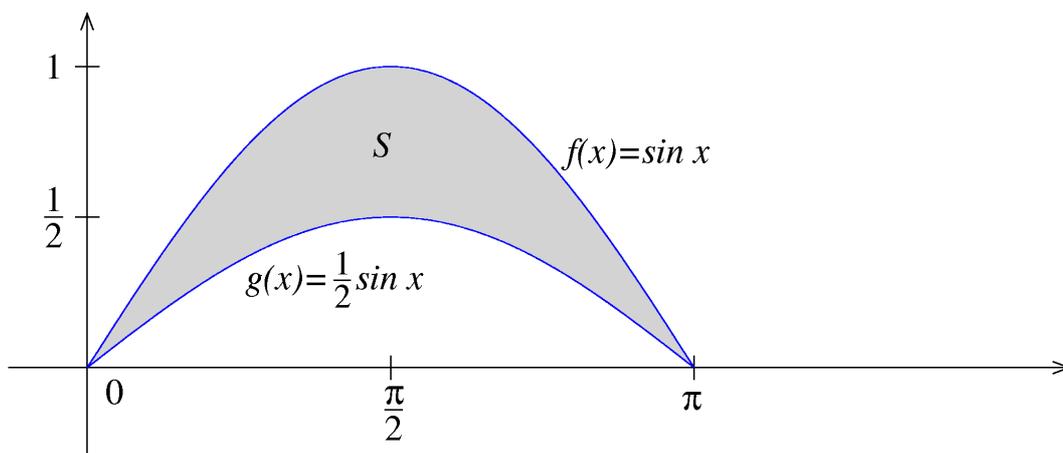


Abbildung 7.4: Fläche unter der kubischen Parabel.

Abbildung 7.5: Sichelfläche zwischen  $f(x) = \sin x$  und  $g(x) = \frac{1}{2} \sin x$ .

Offenbar gilt

$$S = F_0^\pi - G_0^\pi,$$

wenn  $F(x)$  eine Stammfunktion von  $f(x) = \sin x$  und  $G(x)$  eine Stammfunktion von  $g(x) = \frac{1}{2} \sin x$  ist. Die Ableitungen der Funktion  $-\cos x$  ist aber die Funktion  $\sin x$ , also ist  $F(x) = -\cos x$  eine Stammfunktion zu  $f(x) = \sin x$  und  $G(x) = -\frac{1}{2} \cos x$  eine Stammfunktion zu  $g(x) = \frac{1}{2} \sin x$ . Daher erhält man

$$\begin{aligned} S &= -\cos \pi - (-\cos 0) - \left( -\frac{1}{2} \cos \pi - \left( -\frac{1}{2} \cos 0 \right) \right) \\ &= -\frac{1}{2} \cos \pi + \frac{1}{2} \cos 0 \\ &= \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

Die Sichelfläche hat also bemerkenswerterweise den Inhalt eins.

**Bemerkung:** Ist die Funktion  $f(x)$  über dem Intervall  $[a, b]$  negativ, dann erhält man für die Fläche  $F_a^b$  einen negativen Wert, obwohl, geometrisch gesehen, ein Flächeninhalt immer eine positive Zahl ist. In diesem Fall ist daher  $-F_a^b$  der geometrische Flächeninhalt. Wechselt  $f(x)$  im Intervall  $[a, b]$  das Vorzeichen, so kann sich rein rechnerisch sogar der Wert  $F_a^b = 0$  ergeben, obwohl die „geometrische Fläche“ zu  $f(x)$  über und unter dem Intervall  $[a, b]$  keineswegs den Inhalt null hat.

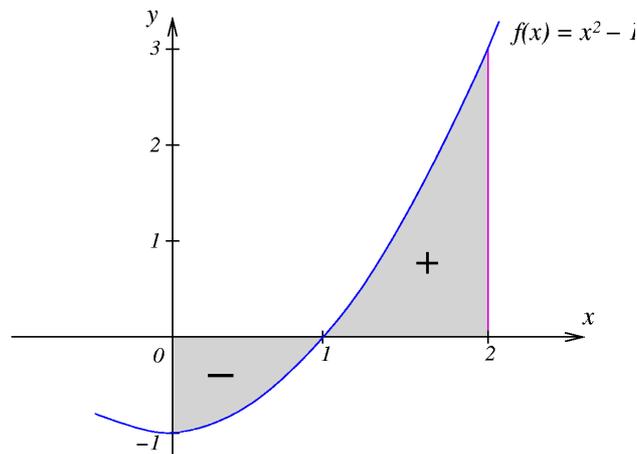


Abbildung 7.6: Das Auftreten negativer Flächeninhalte

Ist zum Beispiel  $f(x) = x^2 - 1$ , dann erhält man

$$F_0^2 = \left( \frac{x^3}{3} - x \right) \Big|_0^2 = \frac{8}{3} - 2 = \frac{2}{3}.$$

Der Inhalt der von  $f(x)$  über dem Intervall  $[0, 2]$  berandeten Fläche ist aber in Wirklichkeit

$$\begin{aligned} -F_0^1 + F_1^2 &= -\left(\frac{x^3}{3} - x\right)\Big|_0^1 + \left(\frac{x^3}{3} - x\right)\Big|_1^2 \\ &= -\left(-\frac{2}{3}\right) + \frac{2}{3} - \left(-\frac{2}{3}\right) \\ &= 2 \end{aligned}$$

▲

## Das Integralzeichen

Approximiert man die Fläche zu einer Funktion  $f(x)$  über einem Intervall  $[a, b]$  durch eine Summe von schmalen Rechtecken, so geht man, in Verallgemeinerung des einführnden Beispiels in Abschnitt 7.1, folgendermaßen vor:

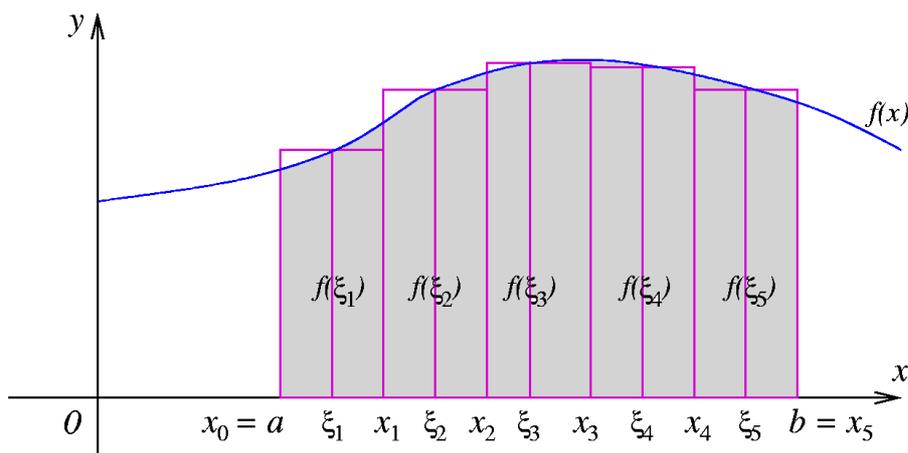


Abbildung 7.7: Approximation einer Fläche durch Rechtecke.

Man unterteilt das Intervall  $[a, b]$  in  $n$  (in Abbildung 7.7 ist  $n = 5$ ) Teilintervalle  $[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n]$  mit  $x_0 = a$  und  $x_n = b$ . Diese Teilintervalle müssen nicht gleich lang sein. Die Länge des Intervalls  $[x_{j-1}, x_j]$  werde mit  $\Delta x_j = x_j - x_{j-1}$  bezeichnet. Aus jedem Intervall denkt man sich eine Stelle  $\xi_j$  ausgewählt und bildet das Rechteck mit der Breite  $\Delta x_j$  und der Höhe  $f(\xi_j)$ . Dann ist die Summe

$$(*) \quad S_n = \sum_{j=1}^n f(\xi_j) \Delta x_j$$

dieser Rechteckflächen für große  $n$  ein Näherungswert für die Fläche  $F_a^b$ . Ist  $f(x)$  eine

stetige Funktion, dann kann man zeigen: Werden die Intervalllängen  $\Delta x_j$  mit wachsendem  $n$  „beliebig klein“, dann strebt die Summe (\*) gegen den Flächeninhalt  $F_a^b$ , das heißt, es gilt

$$\lim_{n \rightarrow \infty} S_n = F_a^b.$$

An diesen Grenzprozeß erinnert die Notation

$$\int_a^b f(x) dx = F_a^b.$$

Das „Integralzeichen“  $\int$  ist dabei als stilisiertes Summenzeichen zu verstehen. Das Symbol  $dx$  hat die Bedeutung einer „unendlich kleinen“ Intervalllänge. Die Intervallgrenzen  $a, b$  werden jetzt als untere ( $a$ ) und obere ( $b$ ) Integrationsgrenze bezeichnet. Die Variable  $x$  ist die „Integrationsvariable“. Sie kann ohne weiteres durch einen anderen Buchstaben, z. B.  $t$ , ersetzt werden, analog zur Beliebigkeit von Indizes.

Der *Hauptsatz der Differential- und Integralrechnung* lautet in dieser Notation

$$\int_a^b f(x) dx = F(b) - F(a),$$

wenn  $F(x)$  eine Stammfunktion von  $f(x)$  ist. Eine solche Stammfunktion ist, wie wir gesehen haben, immer

$$F(x) = \int_0^x f(t) dt,$$

das ist der Inhalt der Fläche zu  $f(x)$  über dem Intervall  $[0, x]$  mit variabler rechter Grenze. Da wir für die Integrationsvariable nicht ebenfalls das Symbol  $x$  verwenden wollen, nennen wir letztere  $t$ .

Läßt man beim Integralzeichen die Integrationsgrenzen weg und schreibt

$$(*) \quad \int f(x) dx,$$

dann ist mit diesem Symbol *irgendeine* Stammfunktion von  $f(x)$  gemeint. Man bezeichnet (\*) auch als das *unbestimmte Integral* der Funktion  $f(x)$ . Die Bildung von (\*) wird auch als *Integration* der Funktion  $f(x)$  bezeichnet. Die Funktion  $f(x)$  selbst heißt der *Integrand*.

## 7.3 Integrationsregeln

Ist  $F(x)$  eine Stammfunktion zu  $f(x)$  und  $G(x)$  eine Stammfunktion zu  $g(x)$ , dann ist  $F(x) + G(x)$  eine Stammfunktion zur Summe  $f(x) + g(x)$ . In der Notation (\*) lautet diese Regel

$$\int (f(x) + g(x)) dx = \int f(x) dx + \int g(x) dx.$$

Außerdem gilt: Ist  $F(x)$  eine Stammfunktion von  $f(x)$ , dann ist  $aF(x)$  eine Stammfunktion von  $af(x)$ , das heißt, es gilt

$$\int af(x) dx = a \int f(x) dx,$$

wenn  $a \in \mathbb{R}$  eine beliebige Zahl ist. Beide Regeln lassen sich zusammenfassen in der Regel

$$(L) \quad \int (af(x) + bg(x)) dx = a \int f(x) dx + b \int g(x) dx$$

mit Zahlen  $a, b \in \mathbb{R}$ . Das bedeutet, der „Integraloperator“  $\int$  ist wie der Differentialoperator  $D$  ein *linearer Operator* (vgl. Abschnitt 6.2, (L)). Zu beachten ist noch, dass die Gleichheitsaussage bei (L) nur „bis auf eine Konstante“ zu verstehen ist.

## Partielle Integration

Die Produktregel (vgl. Abschnitt 6.2, (P)) der Differentiation lautet in salopper Schreibweise

$$(uv)' = u'v + uv'$$

oder

$$u'v = (uv)' - uv'.$$

Eine Stammfunktion der linken Seite ist dann auch eine Stammfunktion der rechten Seite. Und da  $uv$  eine Stammfunktion von  $(uv)'$  ist, folgt mit (L):

$$(PI) \quad \int u'(x)v(x) dx = u(x)v(x) - \int u(x)v'(x) dx.$$

Dies ist die Regel der *partiellen Integration*. Ihre „Pointe“ besteht im folgenden: Hat man das Integral

$$\int f(x)g(x) dx$$

des Produktes zweier Funktionen zu bilden, und ist dies ohne weiteres nicht möglich, dann kommt man häufig so zum Ziel, dass man eine Stammfunktion  $F(x)$  von  $f(x)$  (oder entsprechend von  $g(x)$ ) sucht, in der Hoffnung, dass das Integral

$$\int F(x)g'(x) dx$$

einfacher zu berechnen ist als das ursprüngliche Integral. Nach der Regel (PI) gilt dann

$$\int f(x)g(x) dx = F(x)g(x) - \int F(x)g'(x) dx$$

und die Aufgabe ist gelöst.

Unter Hinzunahme von Integrationsgrenzen lautet die Regel (PI):

$$(PI)' \quad \int_a^b u'(x)v(x) dx = u(x)v(x) \Big|_a^b - \int_a^b u(x)v'(x) dx.$$

Dabei bedeutet

$$u(x)v(x) \Big|_a^b = u(b)v(b) - u(a)v(a)$$

entsprechend der Notation

$$f(x) \Big|_a^b = f(b) - f(a).$$

**Beispiel:** Zu berechnen sei

$$\int_0^1 x e^x dx.$$

Fasst man  $u'(x) = e^x$  als eine Ableitung auf, dann ist  $u(x) = e^x$  die abgeleitete Funktion (Stammfunktion) und mit (PI)' und  $v(x) = x$ , also  $v'(x) = 1$ , erhält man

$$\begin{aligned} \int_0^1 x e^x dx &= x e^x \Big|_0^1 - \int_0^1 1 \cdot e^x dx \\ &= x e^x \Big|_0^1 - e^x \Big|_0^1 \\ &= (x - 1) e^x \Big|_0^1 \\ &= (1 - 1) e^1 - (0 - 1) e^0 \\ &= 1. \end{aligned}$$

■

### Integration durch Substitution

Erinnert werde an die Kettenregel (vgl. Abschnitt 6.2, (K)). Wir schreiben sie in der Variablen  $t$ :

$$f(g(t))' = f'(g(t))g'(t).$$

Ersetzt man hier  $f$  durch eine Stammfunktion  $F$  von  $f$ , dann lautet die Beziehung

$$F(g(t))' = f(g(t))g'(t).$$

Es ist also  $F(g(t))$  eine Stammfunktion von  $f(g(t))g'(t)$ , so dass gilt

$$F(g(t)) = \int f(g(t))g'(t)dt.$$

Setzt man nun  $x = g(t)$ , dann erhält man

$$(*) \quad F(x) = \int f(x)dx = \int f(g(t))g'(t) dt.$$

Ist nun der Integrand  $f(g(t))g'(t)$  einfacher gebaut als  $f(x)$ , so dass man die Integration rechts durchführen kann und eine Funktion

$$G(t) = \int f(g(t))g'(t) dt$$

erhält, dann findet man mit

$$F(x) = G(t) = G(g^{-1}(x))$$

eine Stammfunktion von  $f(x)$ .

Bei einem bestimmten Integral mit den Grenzen  $a$  und  $b$  bezüglich der Variablen  $x$  hat man dann auf der rechten Seite bei (\*) die Grenzen  $a' = g^{-1}(a)$  und  $b' = g^{-1}(b)$  bezüglich der Variablen  $t$  einzusetzen. So erhält man

$$(S) \quad F(b) - F(a) = \int_a^b f(x) dx = \int_{a'}^{b'} f(g(t))g'(t) dt \\ = G(b') - G(a')$$

Das ist die Regel der *Integration durch Substitution*.

Ihr Witz, wie gesagt, besteht darin: Durch geschickte Wahl einer Funktion  $x = g(t)$  hofft man in  $f(g(t))g'(t)$  einen Integranden herzustellen, der einfacher zu handhaben ist als der ursprüngliche Integrand  $f(x)$ .

Man kann sich die obige Integrationsregel auch leicht folgendermaßen merken: Geht man beim Integral

$$\int_a^b f(x) dx$$

von der Integrationsvariabel  $x$  zur Variablen  $t$  über, indem man  $x = g(t)$  setzt, dann sind  $a' = g^{-1}(a)$  und  $b' = g^{-1}(b)$  die Integrationsgrenzen bezüglich der neuen Integrationsvariablen  $t$  und man erhält zunächst

$$\int_a^b f(x) dx = \int_{a'}^{b'} f(g(t)) dx.$$

Es bleibt jetzt nur noch die Transformation hinsichtlich  $dx$  zu komplettieren. Nun gilt aber

$$\frac{dx}{dt} = g'(t),$$

also ist

$$dx = g'(t) dt.$$

Man erhält so wiederum die Substitutionsregel

$$(S) \quad \int_a^b f(x) dx = \int_{a'}^{b'} f(g(t))g'(t) dt.$$

**Bemerkung:** Der Umgang mit dem Ausdruck  $dx/dt$  wie mit einem gewöhnlichen Bruch aus Zahlen lässt sich streng rechtfertigen, worauf aber hier verzichtet wird. Gelegentlich sollte man sich auf die suggestive Beweiskraft des Formalismus einfach verlassen. ▲

**Beispiel:** Zu berechnen sei das bestimmte Integral

$$\int_0^2 xe^{-x^2} dx.$$

Der Integrand wäre unproblematisch, wenn statt des Terms  $e^{-x^2}$  dort  $e^{-x}$  stünde. Mit Hilfe partieller Integration käme man dann leicht zum Ziel. Also erzwingt man den bequemeren Term, indem man  $x = g(t) = \sqrt{t}$  setzt. Dann ist  $t = g^{-1}(x) = x^2$  und  $g'(t) = 1/2\sqrt{t}$ . Damit erhält man nach der Substitutionsregel (S):

$$\begin{aligned} \int_0^2 xe^{-x^2} dx &= \int_{0^2}^{2^2} \sqrt{t}e^{-t} \frac{1}{2\sqrt{t}} dt \\ &= \frac{1}{2} \int_0^4 e^{-t} dx \\ &= -\frac{1}{2} e^{-t} \Big|_0^4 = \frac{1}{2} (1 - e^{-4}) \end{aligned}$$

Im allgemeinen werden sich störende Terme freilich nicht so leicht wegheben wie im vorliegenden Falle. Meistens erzeugt der Versuch, den Integranden an einer Stelle zu „verbessern“ eine neue Schwierigkeit an einer anderen Stelle. ■

### Eine nützliche Integrationsregel

Ist  $f(x)$  positiv und differenzierbar, dann kann man die Funktion

$$g(x) = \log f(x)$$

bilden und es gilt

$$g'(x) = \frac{f'(x)}{f(x)}.$$

Also ist  $g(x) = \log x$  eine Stammfunktion von  $f'(x)/f(x)$ ,

$$\int \frac{f'(x)}{f(x)} dx = \log f(x).$$

Ist  $f(x) < 0$ , dann kann man

$$g(x) = \log(-f(x))$$

bilden und es ist

$$g'(x) = \frac{-f'(x)}{-f(x)} = \frac{f'(x)}{f(x)},$$

also gilt in diesem Fall

$$\int \frac{f'(x)}{f(x)} dx = \log(-f(x)).$$

Beide Fälle fasst man zusammen zu der Integrationsregel

$$(NI) \quad \int \frac{f'(x)}{f(x)} dx = \log |f(x)| + c,$$

die gelegentlich sehr hilfreich ist.

## 7.4 Integration der elementaren Funktionen

Die Aufgabe dieses Abschnitts besteht darin, zu den in Kapitel 5.2 beschriebenen elementaren Funktionen  $f(x)$  jeweils eine Stammfunktion  $F(x)$ , ein Integral, zu finden. Im allgemeinen kann man aber zu einer beliebigen Funktion, die durch elementare Funktionen ausgedrückt wird, keine Stammfunktion in geschlossener Form angeben.

### Die allgemeine Potenzfunktion

Eine Stammfunktion zu  $p(x) = x^3$  ist offenbar die Funktion  $P(x) = \frac{x^{3+1}}{3+1} = \frac{x^4}{4}$ , denn deren Ableitung ist  $P'(x) = \frac{4x^3}{4} = x^3$ . Im allgemeinen Fall der Potenz

$$p(x) = x^\alpha, \quad \alpha \in \mathbb{R},$$

ist analog

$$(*) \quad P(x) = \frac{x^{\alpha+1}}{\alpha+1},$$

sofern nur  $\alpha + 1 \neq 0$ , das heißt,  $\alpha \neq -1$  ist. Dieser Fall erfordert eine gesonderte Überlegung.

Die Formel (\*) erlaubt die Integration beliebiger Polynome. Ist zum Beispiel

$$p(x) = 1 + 2x + 3x^2 + x^4$$

ein solches Polynom, dann ist

$$\begin{aligned} P(x) &= \int p(x) dx = x + 2 \frac{x^2}{2} + 3 \frac{x^3}{3} + \frac{x^5}{5} + c \\ &= x + x^2 + x^3 + \frac{x^5}{5} + c \end{aligned}$$

mit einer beliebigen Konstanten  $c$  eine Stammfunktion von  $p(x)$ .

**Bemerkung:** Die Konstante  $c$ , die bei einer Integration auftritt, lassen wir auch gelegentlich weg. ▲

Der Sonderfall  $\alpha = -1$  erledigt sich, wenn wir uns die Logarithmusfunktion und deren Ableitung in Erinnerung rufen (vgl. Abschnitt 6.3). Wegen

$$(\log x)' = \frac{1}{x} = x^{-1}$$

gilt

$$\int \frac{dx}{x} = \log x, \quad x > 0.$$

Ist  $x < 0$ , dann ist  $\log(-x)$  definiert mit der Ableitung

$$\log(-x)' = \frac{-1}{-x} = \frac{1}{x}.$$

Also hat man die erweiterte Beziehung

$$(**) \quad \int \frac{dx}{x} = \log |x|$$

für  $x \in \mathbb{R}$  und  $x \neq 0$ .

**Bemerkung:** Die Formel

$$\int \frac{dx}{x} = \log x, \quad x > 0,$$

erlaubt die Deutung des Wertes  $\log x$  als Flächeninhalt. Es ist nämlich

$$F(x) = \int_1^x \frac{1}{t} dt$$

eine Stammfunktion von  $f(x) = \log x$ . Daher gibt es eine Konstante  $c \in \mathbb{R}$  so dass gilt

$$\int_1^x \frac{1}{t} dt = \log x + c.$$

Setzt man  $x = 1$  so folgt  $c = 0$ . Damit erhalten wir die Beziehung

$$\log x = \int_1^x \frac{dt}{t}.$$

Abb. 7.8.: Die Logarithmusfunktion als Flächeninhalt

Sie besagt: Der Wert  $\log x$  ist gleich dem Inhalt der Fläche  $F_1^x$  über dem Intervall  $[1, x]$  bezüglich der Funktion  $f(x) = 1/x$ . ▲

### Eine Anwendung: Abschätzung der harmonischen Reihe

Im Abschnitt 4.4 hatten wir die harmonische Reihe kennengelernt und gesehen, dass diese divergiert. Das bedeutet im vorliegenden Fall: Gibt man eine Zahl  $c > 0$  vor, dann gilt:

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} > c,$$

sofern die Anzahl  $n$  der Summanden der harmonischer Reihe nur genügend groß ist. Im Beispiel mit den zu einem Steg ausgelegten Brettern ist  $c = 100$  und  $n$  die Anzahl der Bretter mit den Längen  $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}$  Meter, die, aneinandergesetzt, eine Strecke von 100 Metern erreichen oder überschreiten sollen.

Um eine Abschätzung für die Anzahl  $n$  zu erhalten, betrachten wir die Funktion  $f(x) = 1/x$  für  $x > 0$ . Die Summe der Rechtecksflächen über den Intervallen  $[1, 2], [2, 3], \dots, [n-1, n]$  mit jeweils der Breite 1 und den Höhen  $1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}$  ist dann größer als die Fläche  $F_1^{n+1}$  zur Funktion  $f(x) = 1/x$  über dem Intervall  $[1, n+1]$ .

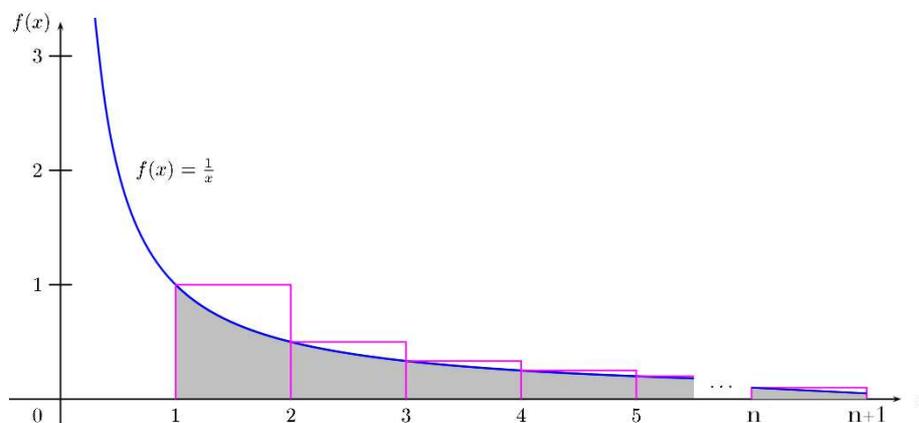


Abbildung 7.8: Zur Abschätzung der harmonischen Reihe.

Es gilt also

$$1 \cdot 1 + 1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{3} + \cdots + 1 \cdot \frac{1}{n} > F_1^{n+1}$$

und wegen

$$F_1^{n+1} = \int_1^{n+1} \frac{dx}{x} = \log(n+1) - \log 1 = \log(n+1)$$

erhalten wir die Abschätzung

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} > \log(n+1).$$

Die Forderung

$$\log(n+1) > 100$$

impliziert

$$n > e^{100} - 1.$$

Ein Steg mit  $n = e^{100}$  aneinandergefügten „harmonischen“ Brettern ist also mehr als 100 m lang. Die Zahl  $e^{100} = 2,688 \cdot 10^{43}$  ist eine Zahl mit 44 Stellen vor dem Komma. Sie ist größer als die Anzahl der Elementarteilchen im Universum.

## Rationale Funktionen

Die rationalen Funktionen sind von der Form

$$y(x) = \frac{p(x)}{q(x)}$$

mit Polynomen  $p(x)$  und  $q(x)$ .

Mit Hilfe des Verfahrens der Partialbruchzerlegung (vgl. Abschnitt 5.2) lässt sich jede rationale Funktion als Summe eines Polynoms  $r(x)$  und endlich vieler Terme der Form  $c(x-a)^{-m}$ ,  $m \geq 1$ , schreiben:

$$y(x) = r(x) + \sum \frac{c}{(x-a)^m}.$$

Dabei ist  $a$  eine Nullstelle des Nennerpolynoms  $q(x)$ , die auch komplex sein kann,  $c$  ein Zahlenfaktor, und  $m$  ist eine ganze Zahl. Die Integration des Polynoms  $r(x)$  wurde oben behandelt. Es geht daher im folgenden nur um die Integration der übrigen Terme. Für  $m > 1$  hat man

$$\int \frac{dx}{(x-a)^m} = \frac{(x-a)^{-m+1}}{-m+1} = \frac{-1}{(m-1)(x-a)^{m-1}}$$

und für  $m = 1$  gilt (vgl. (NI), Abschnitt 7.3):

$$(**) \quad \int \frac{dx}{x-a} = \log |x-a|.$$

Dabei werde zunächst vorausgesetzt, dass die Nullstellen des Nennerpolynoms sämtlich reell sind. Diese Voraussetzung ist aber nicht wesentlich (s. Bemerkung unten).

Das Problem der Integration einer rationalen Funktion reduziert sich also auf die Herstellung ihrer Partialbruchzerlegung.

**Beispiel:** In Abschnitt 5.2 hatten wir die Partialbruchzerlegung der rationalen Funktion

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

bestimmt. Sie lautet

$$y(x) = 1 + \frac{1}{x-1} - \frac{1}{x+1}.$$

Also erhält man wegen der obigen Integrationsregel (\*\*):

$$\begin{aligned} \int \frac{x^2 + 1}{x^2 - 1} dx &= \int \left( 1 + \frac{1}{x-1} - \frac{1}{x+1} \right) dx \\ &= x + \log |x-1| - \log |x+1| + c \\ &= x + \log \left| \frac{x-1}{x+1} \right| + c. \end{aligned}$$

Im Intervall  $-1 < x < 1$  gilt  $(x-1)/(x+1) < 0$ , also ist dort

$$\int \frac{x^2 + 1}{x^2 - 1} dx = x + \log \frac{1-x}{1+x} + c.$$

Für  $x < -1$  oder  $x > 1$  hingegen ist  $(x-1)/(x+1) > 0$ , so dass dort

$$\int \frac{x^2 + 1}{x^2 - 1} dx = x + \log \frac{x-1}{x+1} + c$$

ist. ■

**Bemerkung (Der Fall komplexer Nullstellen):** Die obige Einschränkung, dass die Nullstellen der Nennerpolynome reell sein sollen, ist nicht wesentlich. Das Verfahren der Integration einer rationalen Funktion mit Hilfe der Partialbruchzerlegung lässt sich auch im Fall komplexer Nullstellen des Nennerpolynoms durchführen. Hierzu ein Beispiel: Es sei

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

Die Nullstellen des Nennerpolynoms  $q(x) = x^2 + 1$  sind  $-i$  und  $i$ . Man erhält die Partialbruchzerlegung

$$\frac{1}{x^2 + 1} = \frac{1}{(x - i)(x + i)} = \frac{1}{2i} \left( \frac{1}{x - i} - \frac{1}{x + i} \right),$$

wie man leicht nachrechnet. Daher findet man ganz formal

$$\begin{aligned} \int \frac{dx}{x^2 + 1} &= \frac{1}{2i} \int \left( \frac{1}{x - i} - \frac{1}{x + i} \right) dx \\ &= \frac{1}{2i} \left( \log(x - i) - \log(x + i) \right) \\ &= \frac{1}{2i} \log \frac{x - i}{x + i}. \end{aligned}$$

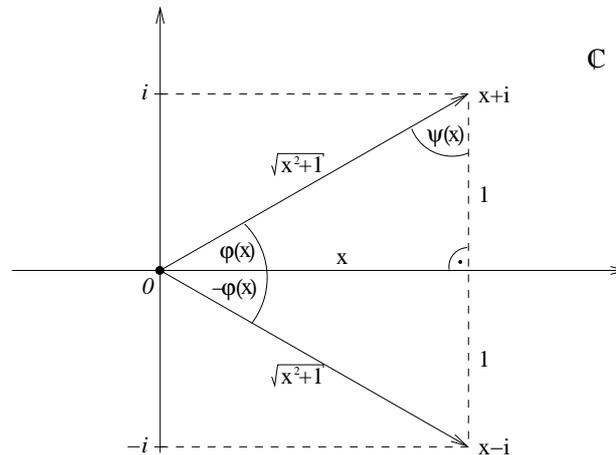


Abbildung 7.9: Die Lage der Punkte  $x - i$  und  $x + i$  in der komplexen Ebene.

In der Polarform geschrieben ist

$$x + i = \sqrt{x^2 + 1} e^{i\varphi(x)}$$

und

$$x - i = \sqrt{x^2 + 1} e^{-i\varphi(x)}.$$

Damit ergibt sich

$$\frac{x - i}{x + i} = e^{-2i\varphi(x)},$$

so dass

$$\log \frac{x - i}{x + i} = -2i\varphi(x)$$

ist. Ist  $\psi(x)$  der bei  $x+i$  gelegene Winkel des Dreiecks mit den Eckpunkten  $0, x, x+i$ , dann gilt  $\varphi(x) = \frac{\pi}{2} - \psi(x)$  und zugleich ist  $\tan \psi(x) = x/1$ , also  $\psi(x) = \tan^{-1}(x) = \arctan x$ . Es ergibt sich also

$$\begin{aligned} \int \frac{dx}{1+x^2} &= \frac{1}{2i}(-2i) \left( \frac{\pi}{2} - \psi(x) \right) \\ &= \arctan x + c. \end{aligned} \quad \blacktriangle$$

## Exponentialfunktionen

Da  $(e^x)' = e^x$  gilt, ist  $e^x$  eine Stammfunktion zu  $e^x$ . Man hat also

$$\int e^x dx = e^x + c,$$

als die bequemste aller Integrationsregeln. Zugleich verifiziert man unmittelbar die Regel

$$(*) \quad \int e^{ax} dx = \frac{1}{a} e^{ax}, \quad a \in \mathbb{R},$$

durch Differentiation mit Hilfe der Kettenregel.

Die allgemeine Exponentialfunktion  $f(x) = a^x$ ,  $a > 0$ , ist erklärt als die Funktion

$$f(x) = e^{\log a \cdot x}.$$

Daher folgt mit (\*):

$$\begin{aligned} \int a^x dx &= \frac{1}{\log a} e^{\log a \cdot x} \\ &= \frac{1}{\log a} a^x. \end{aligned}$$

## Logarithmusfunktionen

Die natürliche Logarithmusfunktion  $\log x$  hat die Ableitung  $(\log x)' = \frac{1}{x}$  (vgl. Abschnitt 6.3). Diese Tatsache, zusammen mit der Regel der partiellen Integration, verhilft zur Integrations von  $\log x$ . Man kann nämlich schreiben:

$$\begin{aligned} \int \log x dx &= \int 1 \cdot \log x dx \\ &= x \cdot \log x - \int x \cdot (\log x)' dx \\ &= x \cdot \log x - \int x \cdot \frac{1}{x} dx \\ &= x \cdot \log x - \int 1 dx \\ &= x \cdot \log x - x. \end{aligned}$$

Da die Logarithmusfunktion  $\log_a x$  zur Basis  $a$  über die Formel

$$\log_a x = \frac{1}{\log a} \log x$$

mit dem natürlichen Logarithmus zusammenhängt, gilt die Integrationsregel

$$\begin{aligned} \int \log_a x \, dx &= \frac{1}{\log a} (x \log x - x) \\ &= x \cdot \log_a x - \frac{x}{\log a}. \end{aligned}$$

### Die trigonometrischen Funktionen

Aufgrund der Ableitungsregeln  $(\sin x)' = \cos x$  und  $(\cos x)' = -\sin x$  erhält man unmittelbar die Integrationsregeln

$$\int \cos x \, dx = \sin x + c$$

und

$$\int \sin x \, dx = -\cos x + c.$$

Das Integral der Funktion

$$\tan x = \frac{\sin x}{\cos x}$$

gewinnt man mit der Regel (NI) am Ende von Abschnitt 7.3. Man kann nämlich schreiben

$$\tan x = -\frac{(\cos x)'}{\cos x},$$

also ist

$$\int \tan x \, dx = -\log |\cos x| + c.$$

Entsprechend findet man

$$\begin{aligned} \int \cot x &= \int \frac{\cos x}{\sin x} \, dx \\ &= \log |\sin x| + c. \end{aligned}$$

Am Beispiel der Funktion  $\arcsin x$  behandeln wir die Integration einer inversen trigonometrischen Funktion. Mit Hilfe partieller Integration (vgl. Abschnitt 7.3) erhält man zunächst

$$\begin{aligned} \int \arcsin x \, dx &= \int 1 \cdot \arcsin x \, dx \\ &= x \cdot \arcsin x - \int x \cdot (\arcsin x)' \, dx \end{aligned}$$

und mit

$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}},$$

(vgl. Abschnitt 6.3) ergibt sich

$$\int x \cdot (\arcsin x)' dx = \int \frac{x}{\sqrt{1-x^2}} dx.$$

Hier bietet sich die Substitution (vgl. Abschnitt 7.3)  $x = \sin t$  mit  $dx = \cos t dt$  an. Man erhält mit  $\sqrt{1-x^2} = \cos t$  daher

$$\begin{aligned} \int \frac{x}{\sqrt{1-x^2}} dx &= \int \frac{\sin t}{\cos t} \cdot \cos t dt \\ &= \int \sin t dt = -\cos t \\ &= -\sqrt{1-x^2}. \end{aligned}$$

Unser Ergebnis lautet also

$$\int \arcsin x dx = x \cdot \arcsin x + \sqrt{1-x^2} + c.$$

Seine Richtigkeit kann man unmittelbar überprüfen, indem man die Ableitung der rechten Seite bildet.

## 7.5 Taylorreihen

Während man für die Exponentialfunktion eine Darstellung in Form einer Reihenentwicklung besitzt, die zugleich die Berechnung beliebiger Funktionswerte mit beliebiger Genauigkeit erlaubt, ist dies bei ihrer Umkehrfunktion zunächst nicht der Fall. Die Definition der Funktion  $\log x$  als Umkehrfunktion der Funktion  $e^x$  legt zwar das mathematische Objekt „Logarithmusfunktion“ fest, sie ist aber nicht „konstruktiv“ im Sinne eines Berechnungsverfahrens.

Ein solches Berechnungsverfahren für die Logarithmusfunktion und allgemein für beliebige Funktionen mit hinreichend guten Eigenschaften wird im folgenden beschrieben.

Ist  $f(x)$  in einem Intervall  $I$  differenzierbar und sind die Stellen  $x, a$  aus  $I$ , dann gilt

$$f(x) - f(a) = \int_a^x f'(t) dt$$

aufgrund des Hauptsatzes der Differential- und Integralrechnung. Diese Beziehung kann man auch in der Form

$$f(x) = f(a) + \int_a^x f'(t) dt$$

schreiben. Denkt man sich die Stelle  $a$  als fest und Stelle  $x$  als variabel, dann wird der Funktionswert  $f(x)$  dargestellt als der Funktionswert  $f(a)$  plus eine Restfunktion. Ist  $f'(x) = 0$ , also  $f(x)$  eine Konstante, dann ist die Darstellung von  $f(x)$  durch  $f(a)$  sogar genau.

Mit Hilfe partieller Integration (vgl. Abschnitt 7.3) erhält man weiter (mit  $(t - x)$  als Stammfunktion von 1):

$$\begin{aligned} \int_a^x f'(t) dt &= \int_a^x 1 \cdot f'(t) dt \\ &= (t - x)f'(t) \Big|_a^x - \int_a^x (t - x)f''(t) dt \\ &= f'(a)(x - a) + \int_a^x (x - t)f''(t) dt \end{aligned}$$

Durch nochmalige partielle Integration beim Integralterm findet man

$$\begin{aligned} \int_a^x (x - t)f''(t) dt &= -\frac{1}{2}(x - t)^2 f''(t) \Big|_a^x + \frac{1}{2} \int_a^x (x - t)^2 f'''(t) dt \\ &= \frac{1}{2} f''(a)(x - a)^2 + \frac{1}{2} \int_a^x (x - t)^2 f'''(t) dt. \end{aligned}$$

Insgesamt haben wir also die folgende Entwicklung von  $f(x)$  an der Stelle  $a$  erhalten:

$$f(x) = f(a) + f'(a)(x - a) + \frac{1}{2} f''(a)(x - a)^2 + \frac{1}{2} \int_a^x (x - t)^2 f'''(t) dt.$$

Nach einem weiteren Schritt lässt sich das Bildungsgesetz dieser Entwicklung erraten. Der letzte Integralterm geht durch partielle Integration über in

$$\frac{1}{2} \int_a^x (x - t)^2 f'''(t) dt = -\frac{1}{1 \cdot 2 \cdot 3} (x - t)^3 f'''(t) \Big|_a^x + \frac{1}{1 \cdot 2 \cdot 3} \int_a^x (x - t)^3 f^{(4)}(t) dt,$$

so dass die Entwicklung von  $f(x)$  jetzt lautet:

$$f(x) = f(a) + \frac{f'(a)}{1!} (x - a)^1 + \frac{f''(a)}{2!} (x - a)^2 + \frac{f'''(a)}{3!} (x - a)^3 + \frac{1}{3!} \int_a^x (x - t)^3 f^{(4)}(t) dt.$$

Bei dieser Rechnung haben wir unausgesprochen vorausgesetzt, dass die Funktion  $f(x)$  im Intervall  $I$  viermal abgeleitet werden kann. Kann man  $f(x)$  in diesem Intervall  $(n + 1)$ -mal ableiten, dann ergibt sich durch Fortführung des obigen Vorgehens die Entwicklung

$$(T) \quad \begin{aligned} f(x) &= f(a) + \frac{f'(a)}{1!} (x - a)^1 + \frac{f''(a)}{2!} (x - a)^2 + \dots \\ &\quad + \frac{f^{(n)}(a)}{n!} (x - a)^n + \frac{1}{n!} \int_a^x (x - t)^n f^{(n+1)}(t) dt \end{aligned}$$

mit dem Restglied

$$(R) \quad R_n(x) = \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t) dt.$$

Dies ist die – endliche – *Taylorentwicklung* der Funktion  $f(x)$  in bezug auf den Entwicklungspunkt  $a$ . Strebt das Restglied gegen null für  $n \rightarrow \infty$  und  $x \in I$ , und dies ist bei den Funktionen, mit denen man es üblicherweise zu tun hat, der Fall, dann erhält man die Darstellung der Funktion  $f(x)$  durch ihre *Taylorreihe*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a)^1 + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots$$

oder, mit Hilfe des Summenzeichens,

$$(T') \quad f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k.$$

Ist  $a = 0$ , so hat man den Spezialfall

$$(T'') \quad f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k.$$

Für  $f(x) = e^x$  erhält man wegen  $f^{(k)}(x) = e^x$  und  $f^{(k)}(0) = 1$  die bekannte Reihenentwicklung der Exponentialfunktion.

## Der verallgemeinerte binomische Lehrsatz

Wir wollen die Funktion

$$f(x) = (1+x)^\alpha, \quad \alpha \in \mathbb{R},$$

um die Stelle  $a = 0$  in eine Taylorreihe entwickeln. Man hat

$$\begin{aligned} f'(x) &= \alpha(1+x)^{\alpha-1}, \\ f''(x) &= \alpha(\alpha-1)(1+x)^{\alpha-2} \end{aligned}$$

und allgemein

$$f^{(k)}(x) = \alpha(\alpha-1) \cdots (\alpha-k+1)(1+x)^{\alpha-k}.$$

Also ist

$$\frac{f^{(k)}(0)}{k!} = \frac{\alpha(\alpha-1) \cdots (\alpha-k+1)}{k!} 1^{\alpha-k} = \binom{\alpha}{k},$$

wenn wir die Definition des Binomialkoeffizienten auf reelle „Zähler“  $\alpha$  erweitern. Mit (T) ergibt sich dann die Entwicklung

$$(1+x)^\alpha = 1 + \binom{\alpha}{1} x + \binom{\alpha}{2} x^2 + \cdots + \binom{\alpha}{n} x^n + R_n(x)$$

mit

$$\begin{aligned} R_n(x) &= \frac{1}{n!} \int_0^x (x-t)^n \alpha(\alpha-1)\cdots(\alpha-n)(1+t)^{\alpha-n-1} dt \\ &= (n+1) \binom{\alpha}{n+1} \int_0^x (x-t)^n (1+t)^{\alpha-n-1} dt. \end{aligned}$$

Man kann nun zeigen, dass dieses Restglied  $R_n(x)$  für  $n \rightarrow \infty$  gegen null strebt, sofern  $-1 < x < 1$  ist. In diesem Fall haben wir daher die Taylorentwicklung

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k.$$

Daraus erhält man sofort den verallgemeinerten binomischen Lehrsatz für Ausdrücke der Form

$$(a+b)^\alpha, \quad \alpha \in \mathbb{R}.$$

Ist  $-1 < \frac{a}{b} < 1$ , dann kann man schreiben

$$\begin{aligned} (a+b)^\alpha &= b^\alpha \left(1 + \frac{a}{b}\right)^\alpha \\ &= b^\alpha \sum_{k=0}^{\infty} \binom{\alpha}{k} \left(\frac{a}{b}\right)^k \end{aligned}$$

Also gilt

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

Das ist der verallgemeinerte binomische Lehrsatz.

## Die Reihenentwicklung der Logarithmusfunktion

Die Funktion  $\log(1-x)$  hat die Ableitung

$$\log(1-x)' = \frac{-1}{1-x}.$$

Nun lautet aber die Summenformel für die unendliche geometrische Reihe (vgl. Abschnitt 4.4):

$$1 + x + x^2 + \cdots = \frac{1}{1-x}$$

unter der Voraussetzung  $-1 < x < 1$ . Daher erhalten wir die Entwicklung

$$\begin{aligned} \log(1-x)' &= -(1+x+x^2+\cdots) \\ &= -\sum_{k=0}^{\infty} x^k \end{aligned}$$

für  $-1 < x < 1$ . Integration auf beiden Seiten dieser Beziehung ergibt

$$\log(1-x) = -\sum_{k=0}^{\infty} \frac{x^{k+1}}{k+1} + c.$$

Setzt man  $x = 0$ , so folgt  $c = 0$  wegen  $\log 1 = 0$ . Damit ergibt sich unmittelbar die Taylorentwicklung

$$(L) \quad \log(1-x) = -\sum_{k=1}^{\infty} \frac{x^k}{k}, \quad -1 < x < 1,$$

der Logarithmusfunktion. Sie erlaubt die Berechnung von Funktionswerten  $\log a$  aus dem Intervall  $0 < a < 2$ . Ersetzt man bei (L)  $x$  durch  $-x$ , so nimmt die Reihenentwicklung die Form

$$(L') \quad \log(1+x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{x^k}{k}, \quad -1 < x < 1,$$

an.

Man kann zeigen, dass die Reihe (L') auch noch für  $x = 1$  konvergiert. So ergibt sich die bemerkenswerte Darstellung

$$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots$$

des Zahlenwertes  $\log 2$ . – Die Formel (L') erlaubt aber auch die Berechnung von Logarithmen außerhalb des Intervalls  $(0, 2)$ . so kann man z. B. schreiben

$$\begin{aligned} \log 3 &= \log(2+1) = \log\left(2\left(1+\frac{1}{2}\right)\right) \\ &= \log 2 + \log\left(1+\frac{1}{2}\right) \\ &= \log 2 + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{k 2^k}. \end{aligned}$$

In analoger Weise erhalte man  $\log 5 = \log(3+2) = \log(3 \cdot (1+2/3)) = \log 3 + \log(1+2/3)$  und könnte den zweiten Term wieder mit (L') entwickeln.

Wir haben also die Taylorentwicklung von  $\log(1-x)$  um  $a = 0$  erhalten, ohne die Formel (T'') explizit zu benutzen. Eine analoge Vorgehensweise führt bei der Funktion  $\arctan x$  zum Ziel.

### Die Reihenentwicklung der Funktion $\arctan x$

In der Bemerkung in Abschnitt 7.4 hatten wir die Beziehung

$$\int \frac{dx}{1+x^2} = \arctan x + c$$

hergeleitet. Für  $0 \leq x^2 < 1$ , also  $-1 < x < 1$ , hat man aufgrund der Summenformel für die geometrische Reihe (vgl Abschnitt 4.4) die Entwicklung

$$\frac{1}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k x^{2k}.$$

Daher gilt

$$\int \frac{dx}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1} = \arctan x + c.$$

Wegen  $\arctan 0 = 0$  muß  $c = 0$  sein und man findet die Entwicklung

$$(AT) \quad \arctan x = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{2k+1}.$$

Nun ist  $\tan \frac{\pi}{4} = \tan 45^\circ = 1$ , also  $\arctan 1 = \frac{\pi}{4}$  und damit

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots$$

Dies ist die Leibnizsche Reihe für  $\frac{\pi}{4}$ . Sie ermöglicht im Prinzip die direkte Berechnung von  $\frac{\pi}{4}$  und damit von  $\pi$  mit beliebiger Genauigkeit. Gleichwohl ist sie für praktische Berechnungen von  $\pi$  nicht so gut geeignet, da sie nur sehr langsam konvergiert.



# Kapitel 8

## Lineare Algebra

Die lineare Algebra ist die Lehre von den Vektorräumen, ihrer inneren Struktur und den Beziehungen zwischen Vektorräumen. Das folgende Kapitel führt also zunächst die Begriffe „Vektor“ und „Vektorraum“ ein, gibt dazu Beispiele aus den Anwendungen und erklärt das Rechnen mit Vektoren. Mit Hilfe der Vektoren kann man dann rein rechnerisch die Geometrie der Ebene und des Raumes entwickeln.

Beziehungen zwischen Vektorräumen werden durch Abbildungen hergestellt, deren einfachster Typus die linearen Abbildungen sind. Diese werden durch Matrizen beschrieben. Wesentliche Begriffe im Kontext von Matrizen (aber nicht nur dort) sind die Begriffe „Eigenvektor“ und „Eigenwert“.

Mit dem genannten mathematischen Instrumentarium lässt sich ein instruktives und realistisches Wachstumsmodell formulieren, das „Leslie-Modell“. Hierbei wird eine Population in Altersgruppen aufgeteilt. Die Individuenzahlen in den einzelnen Altersgruppen bestimmen dann die Komponenten eines Vektors, der den jeweiligen „Zustand“ der Population repräsentiert und die Dynamik des Wachstums wird durch eine Matrix beschrieben. Die Eigenvektoren und Eigenwerte dieser Matrix bestimmen dann die Langzeit-Entwicklung der Population.

Innerhalb des Leslie-Modells und in zahlreichen weiteren Gebieten gilt es lineare Gleichungssysteme zu lösen. Die Struktur der Lösungsmenge wird hergeleitet und das zugehörige Lösungsverfahren, der Gauß-Algorithmus, an Beispielen dargestellt.

### 8.1 Vektoren

Viele Phänomene in Natur und Technik wie zum Beispiel Temperatur, Gewicht, Länge, lassen sich durch Angabe einer *einzig*en Zahl, man spricht auch von einem Skalar, das ist ein Zahlenwert bezüglich einer Skala von Werten, beschreiben. Bei anderen Phänomenen ist zur vollständigen mathematischen Beschreibung ein Satz mehrerer Zahlen erforderlich.

Die Kraft  $F$  (Force) zum Beispiel, die an einem Körper im Raum angreift, erfordert

die Angabe dreier Zahlen zu ihrer vollständigen Beschreibung, und zwar die Stärke der Kraftkomponenten in der  $x$ -,  $y$ -, und  $z$ -Richtung. Sind diese durch die Zahlen  $x_1, x_2, x_3$  bestimmt, dann schreibt man  $\mathbf{F} = (x_1, x_2, x_3)$ . Die Kraft  $\mathbf{F}$  lässt sich also deuten als Element des  $\mathbb{R}^3$ . Das gleiche gilt für die Geschwindigkeit  $\mathbf{v}$ , mit der sich ein Körper im Raum bewegt. Sind  $v_1, v_2, v_3$  die Geschwindigkeitskomponenten in  $x$ -,  $y$ -, und  $z$ -Richtung, dann schreibt man  $\mathbf{v} = (v_1, v_2, v_3)$ . Zahlentripel wie  $\mathbf{F}$  und  $\mathbf{v}$  lassen sich auch als „Pfeile“ verstehen, die in einem beliebigen Punkt des Raumes angeheftet sind. Andererseits kann man Vektoren auch als im Raum bewegliche Pfeile auffassen, die ihre Länge und ihre Richtung beibehalten.

Die Wachstumsdynamik einer Population kann man oft nur unzureichend erfassen, wenn man als Kenngröße für ihren Zustand zu einem bestimmten Zeitpunkt lediglich die Anzahl ihrer Individuen heranzieht, also eine skalare Größe. Vielmehr ist es oft zweckmäßig, innerhalb der Population Altersgruppen zu unterscheiden. Der Zustand einer Population  $\mathbf{x}$  mit vier Altersgruppen 1, 2, 3, 4 wird dann durch ein Quadrupel  $\mathbf{x} = (x_1, x_2, x_3, x_4)$  von Zahlen, das ist ein Element des  $\mathbb{R}^4$ , beschrieben.

## Das Rechnen mit Vektoren

Versieht man die Menge  $\mathbb{R}^3 = \{(x_1, x_2, x_3) \mid x_1, x_2, x_3 \in \mathbb{R}\}$  der Zahlentripel mit einer *Addition* und einer *skalaren Multiplikation* indem man setzt

$$(A) \quad \mathbf{x} + \mathbf{y} = (x_1, x_2, x_3) + (y_1, y_2, y_3) = (x_1 + y_1, x_2 + y_2, x_3 + y_3)$$

und

$$(sM) \quad t \cdot \mathbf{x} = t \cdot (x_1, x_2, x_3) = (t \cdot x_1, t \cdot x_2, t \cdot x_3), \quad t \in \mathbb{R},$$

dann wird aus der Menge  $\mathbb{R}^3$  der *Vektorraum*  $(\mathbb{R}^3, +, \cdot)$ , den wir der Einfachheit halber aber wieder mit  $\mathbb{R}^3$  bezeichnen. – Vektoren werden im folgenden durch Fettdruck hervorgehoben.

**Bemerkung:** Mathematiker der strengen Observanz unterscheiden auch in der Notation zwischen der Menge  $\mathbb{R}^3$  der Zahlentripel und dem *Vektorraum*  $\mathbb{V}^3 = (\mathbb{R}^3, +, \cdot)$ , der aus der Menge  $\mathbb{R}^3$  durch Hinzunehmen der additiven und multiplikativen Struktur (A) und (sM) entsteht. Konsequenterweise müßte man dann auch bei Definition (A) zwischen den üblichen  $+$  – Zeichen in der Klammer rechts und den beiden  $+$  – Zeichen links unterscheiden. Dasselbe gilt hinsichtlich der  $\cdot$  – Zeichen bei der Definition (sM). Das  $\cdot$  – Zeichen lassen wir im folgenden normalerweise weg. Nachlässigkeiten in der Notation, solange sie nicht zu Mißverständnissen führen, hingegen die Lesbarkeit fördern, wollen wir uns aber weiter erlauben. ▲

Die vektorielle Addition und die skalare Multiplikation genügen den folgenden *Rechenregeln*:

**1. Gesetze der Vektoraddition:**

- a)  $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$  (Assoziativgesetz),
- b)  $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$  (Kommutativgesetz)

**2. Assoziativgesetz der skalaren Multiplikation:**

$$s(t\mathbf{x}) = (st)\mathbf{x}, \quad s, t \in \mathbb{R}$$

**3. Distributivgesetze:**

- a)  $(s + t)\mathbf{x} = s\mathbf{x} + t\mathbf{x}$ ,
- b)  $t(\mathbf{x} + \mathbf{y}) = t\mathbf{x} + t\mathbf{y}$ ,  $s, t \in \mathbb{R}$ .

Die Richtigkeit dieser Rechenregeln bestätigt man leicht durch Nachrechnen. Es besteht beim Rechnen mit Vektoren also eine volle Analogie zum Rechnen mit Zahlen. Die Rolle der Null bei der Vektoraddition spielt der *Nullvektor*, das ist im Falle des  $\mathbb{R}^3$ , der Vektor  $\mathbf{0} = (0, 0, 0)$ .

### Der Betrag eines Vektors

Legt man in der Ebene ein cartesisches Koordinatensystem mit dem Nullpunkt  $O$  fest, dann bestimmt ein Vektor  $\mathbf{x} = (x_1, x_2)$  einen Punkt  $X$  der Ebene, wenn man  $x_1, x_2$  als die Koordinaten dieses Punktes auffasst. Die von  $O$  nach  $X$  durchlaufene, das ist die „gerichtete“ Strecke  $OX$  (deren Richtung man durch einen Pfeil andeuten kann), entspricht dann dem Vektor  $\mathbf{x}$ . Man schreibt auch  $\vec{x} = \overrightarrow{OX}$ .

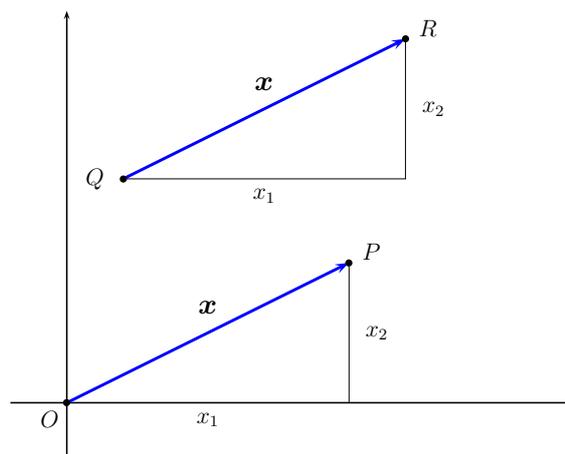


Abbildung 8.1: Der Vektor  $\mathbf{x}$  als gerichtete Strecke

Die Länge der Strecke  $|OX|$  ist dann der *Betrag* des Vektors  $\boldsymbol{x}$ . Dieser ist also immer eine positive Zahl. Man bezeichnet ihn mit  $\|\boldsymbol{x}\|$ . Wegen des Satzes des Pythagoras gilt dann

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2}.$$

Für einen Vektor  $\boldsymbol{x} = (x_1, x_2, x_3)$  des  $\mathbb{R}^3$  zeigt man entsprechend

$$\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2},$$

und allgemein definiert man den Betrag eines Vektors  $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$  des  $\mathbb{R}^n$  durch den Ausdruck

$$(B) \quad \|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}.$$

Die Wurzel ist dabei immer im positiven Sinne zu verstehen.

### Geometrische Deutung von Vektoraddition und skalarer Multiplikation

Denkt man sich zwei Vektoren  $\boldsymbol{x} = (x_1, x_2)$  und  $\boldsymbol{y} = (y_1, y_2)$  des  $\mathbb{R}^2$  im Nullpunkt 0 eines cartesischen Koordinatensystems angeheftet und verschiebt den Vektor  $\boldsymbol{y}$  an die Spitze des Vektors  $\boldsymbol{x}$  und  $\boldsymbol{x}$  an die Spitze von  $\boldsymbol{y}$ , dann entsteht ein Parallelogramm. Die dem Nullpunkt 0 gegenüberliegende Ecke des Parallelogramms sei P.

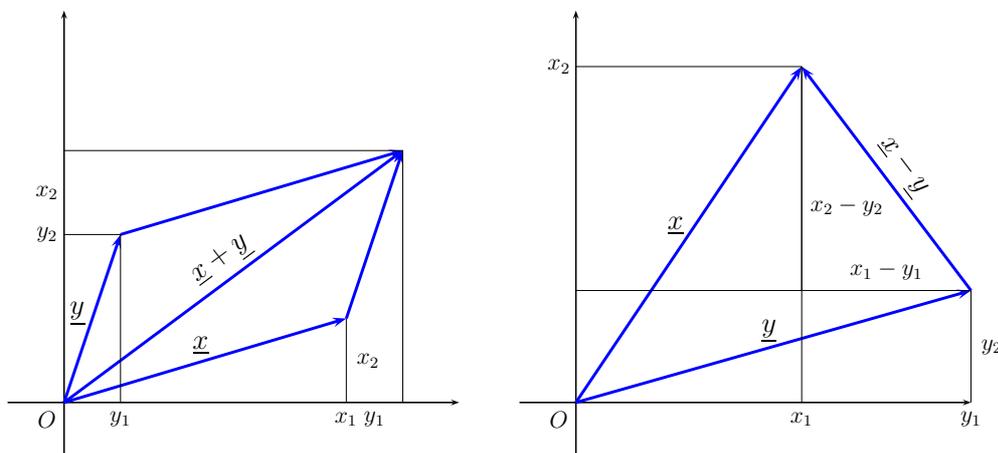


Abbildung 8.2: Addition und Subtraktion von Vektoren; das Kräfte-Parallelogramm

Die Diagonale  $OP$  des Parallelogramms, aufgefasst als gerichtete Strecke, entspricht dann dem Vektor  $\boldsymbol{x} + \boldsymbol{y}$ .

Fasst man die Vektoren  $\boldsymbol{x}$  und  $\boldsymbol{y}$  als Kräfte auf, die im Punkt  $O$  angreifen, dann ist

$\mathbf{x} + \mathbf{y} = \mathbf{z}$  die resultierende, das heißt die Einzelkräfte  $\mathbf{x}$ ,  $\mathbf{y}$  ersetzende, Kraft. Diese Deutung der Vektoraddition durch das zugehörige Kräfteparallelogramm bleibt analog auch im Fall des  $\mathbb{R}^3$ , bzw. allgemein des  $\mathbb{R}^n$ , gültig.

Der Vektor  $t\mathbf{x} = t(x_1, x_2) = (tx_1, tx_2)$  besitzt dieselbe Richtung wie der Vektor  $\mathbf{x}$ , sofern  $t > 0$  ist. Sein Betrag ist das  $t$ -fache des Betrags von  $\mathbf{x}$ .

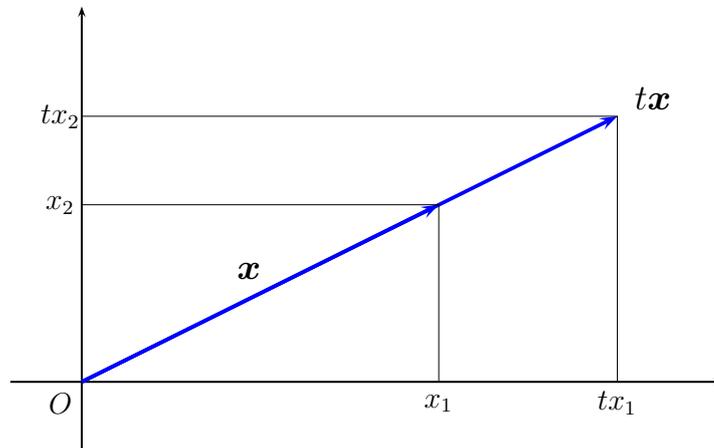


Abbildung 8.3: Geometrische Deutung der skalaren Multiplikation  $t\mathbf{x}$ ,  $t > 0$

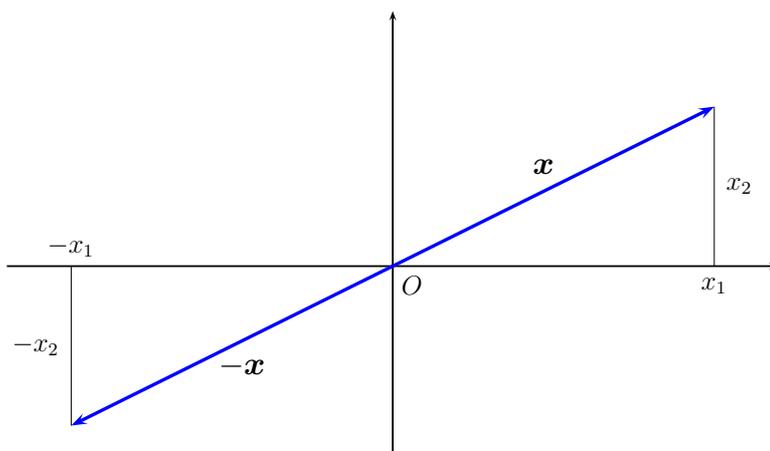
Aufgrund der Definition (B) des Betrages eines Vektors gilt nämlich für  $\mathbf{x} \in \mathbb{R}^n$  und  $t > 0$ :

$$\begin{aligned} \|t\mathbf{x}\| &= \sqrt{(tx_1)^2 + (tx_2)^2 + \cdots + (tx_n)^2} \\ &= \sqrt{t^2(x_1^2 + x_2^2 + \cdots + x_n^2)} \\ &= t\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \\ &= t\|\mathbf{x}\|. \end{aligned}$$

Der Vektor  $(-1)\mathbf{x} = -\mathbf{x} = (-x_1, -x_2, \dots, -x_n)$  hat die zu  $\mathbf{x}$  entgegengesetzte Richtung, aber denselben Betrag. Die „Kräfte“  $\mathbf{x}$  und  $-\mathbf{x}$ , die am selben Punkt angreifen, heben sich auf. Das heißt, es gilt

$$\mathbf{x} + (-1)\mathbf{x} = \mathbf{x} - \mathbf{x} = \mathbf{0}.$$

Allgemein gilt: Ist  $t < 0$ , dann hat der Vektor  $t\mathbf{x}$  die zu  $\mathbf{x}$  entgegengesetzte Richtung und die Länge  $|t| \cdot \|\mathbf{x}\|$ . Die *Differenz*  $\mathbf{x} - \mathbf{y} = \mathbf{z}$  zweier Vektoren ist der Vektor  $\mathbf{z}$  mit der Eigenschaft  $\mathbf{x} = \mathbf{y} + \mathbf{z}$ . Der Vektor  $\mathbf{z}$  verbindet daher die Spitze von  $\mathbf{y}$  mit der Spitze von  $\mathbf{x}$  und markiert so die andere Diagonale im Kräfteparallelogramm (vgl. Abb. 8.2).

Abbildung 8.4: Die Vektoren  $\boldsymbol{x}$  und  $-\boldsymbol{x}$ .

### Das Skalarprodukt

Zwei Vektoren  $\boldsymbol{a} = (a_1, a_2)$  und  $\boldsymbol{b} = (b_1, b_2)$  des  $\mathbb{R}^2$  schließen einen Winkel  $\varphi$  ein.

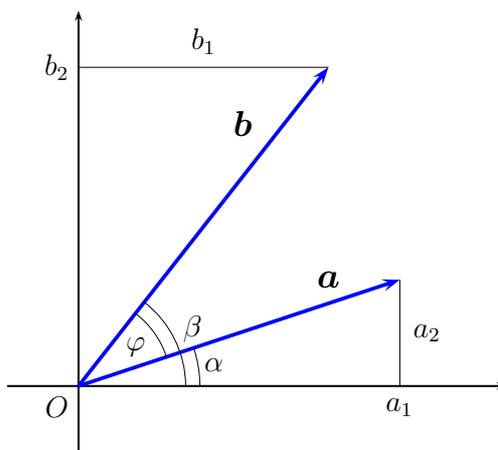


Abbildung 8.5: Der Winkel zwischen zwei Vektoren

Gemeint ist damit derjenige Winkel, um den man den Vektor  $\boldsymbol{a}$  gegen den Uhrzeigersinn drehen muß, damit er in die Richtung von Vektor  $\boldsymbol{b}$  zeigt.

Es stellt sich die Frage, ob und wie man aufgrund der Kenntnis der Komponenten  $a_1, a_2$  und  $b_1, b_2$  der Vektoren  $\boldsymbol{a}, \boldsymbol{b}$  den Winkel  $\varphi$  berechnen kann.

Zur Beantwortung dieser Frage interpretieren wir die Vektoren  $\boldsymbol{a}$  und  $\boldsymbol{b}$  als komplexe Zahlen

$$\boldsymbol{a} = a_1 + i a_2 = |\boldsymbol{a}| e^{i\alpha}$$

und

$$b = b_1 + i b_2 = |b| e^{i\beta}.$$

Dabei ist  $|a| = \|\mathbf{a}\|$ ,  $|b| = \|\mathbf{b}\|$  und es sei  $\beta - \alpha = \varphi$  der Winkel zwischen  $\mathbf{a}$  und  $\mathbf{b}$ . Dann gilt einerseits

$$\begin{aligned} \bar{a} b &= |a| e^{-i\alpha} |b| e^{i\beta} \\ &= |a| |b| e^{i(\beta-\alpha)} \\ &= |a| |b| e^{i\varphi}. \end{aligned}$$

Mit Hilfe der Eulerformel (vgl. Abschnitt 2.4) erhält man also

$$(*) \quad \bar{a} b = |a| |b| (\cos \varphi + i \sin \varphi).$$

Andererseits ergibt direktes Ausmultiplizieren

$$(**) \quad \begin{aligned} \bar{a} b &= (a_1 - i a_2)(b_1 + i b_2) \\ &= a_1 b_1 + a_2 b_2 + i(a_1 b_2 - a_2 b_1). \end{aligned}$$

Vergleicht man den Realteil bei den rechten Seiten von (\*) und (\*\*), dann liest man ab:

$$(***) \quad a_1 b_1 + a_2 b_2 = |a| |b| \cos \varphi.$$

Die linke Seite dieser Beziehung faßt man als ein „Produkt“ der Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  auf, dessen Ergebnis eine Zahl, das ist ein Skalar, ist. Man spricht daher von dem *Skalarprodukt*

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2$$

der Vektoren  $\mathbf{a}$  und  $\mathbf{b}$ .

Aufgrund der Beziehung (\*\*\*) kann man dann den Winkel  $\varphi$  berechnen, denn es gilt:

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \varphi,$$

also ist

$$\cos \varphi = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}.$$

Die rechte Seite dieser Beziehung ist ein Ausdruck in den Komponenten der Vektoren  $\mathbf{a}$  und  $\mathbf{b}$ .

Wegen  $\cos \varphi = \cos(2\pi - \varphi)$  gibt es immer zwei Winkel, die zu einem gegebenen  $\cos$ -Wert gehören. Diese ergänzen sich zu  $2\pi$  oder  $360^\circ$ . Die geometrische Konfiguration

der Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  entscheidet dann, welcher der beiden Winkel zu nehmen ist.

Da  $\cos \frac{\pi}{2} = 0$  ist, so folgt: Zwei Vektoren  $\mathbf{a}, \mathbf{b}$  des  $\mathbb{R}^2$  stehen genau dann senkrecht aufeinander, wenn

$$\mathbf{a} \cdot \mathbf{b} = 0$$

ist.

Damit haben wir ein einfach zu überprüfendes Kriterium gewonnen das uns erlaubt festzustellen, ob zwei Vektoren senkrecht aufeinander stehen.

**Beispiel:** Die Vektoren  $\mathbf{a} = (1, 2)$  und  $\mathbf{b} = (2, -1)$  stehen senkrecht aufeinander, denn es gilt

$$\mathbf{a} \cdot \mathbf{b} = 1 \cdot 2 + 2 \cdot (-1) = 0.$$

■

Damit kann man aber zu jedem Vektor  $\mathbf{a} = (a_1, a_2)$  des  $\mathbb{R}^2$  einen zu  $\mathbf{a}$  senkrechten Vektor unmittelbar finden. Zum Beispiel ist  $\mathbf{b} = (a_2, -a_1)$  ein solcher Vektor, wie man leicht nachrechnet.

Allgemein *definiert* man das Skalarprodukt nun in Analogie zum  $\mathbb{R}^2$  für Vektoren  $\mathbf{a} = (a_1, a_2, \dots, a_n)$ ,  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  des  $\mathbb{R}^n$  durch den Ausdruck

$$(SP) \quad \mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n,$$

bei dem die Komponenten mit gleichem Index miteinander multipliziert und diese Produkte dann addiert werden.

Das Skalarprodukt hat die folgenden Eigenschaften, die man leicht nachprüft. Es gilt:

$$(SP_1) \quad \mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} \quad (\text{Kommutativgesetz}),$$

$$(SP_2) \quad \mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c} \quad (\text{Distributivgesetz}),$$

$$(SP_3) \quad (t\mathbf{a}) \cdot \mathbf{b} = t(\mathbf{a} \cdot \mathbf{b}), \quad t \in \mathbb{R}.$$

Der Betrag eines Vektors lässt sich jetzt auch mit Hilfe des Skalarproduktes ausdrücken. Man hat offenbar

$$(SP_4) \quad \mathbf{a} \cdot \mathbf{a} = \mathbf{a}^2 = \|\mathbf{a}\|^2,$$

also gilt

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

Die Notation  $\mathbf{a}^2$  ist hier lediglich eine Abkürzung für das Produkt  $\mathbf{a} \cdot \mathbf{a}$ . (Die Potenz  $\mathbf{a}^3$  zum Beispiel hat keinen Sinn.)

Den Winkel  $\varphi$ , den zwei Vektoren  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  einschließen, *definiert* man jetzt durch die Beziehung

$$\cos \varphi = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}.$$

Diese Definition ist aber nur möglich wenn sichergestellt ist, daß die Ungleichung

$$(CU) \quad -1 \leq \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \leq 1$$

besteht, da die  $\cos$ -Werte immer zwischen  $-1$  und  $+1$  liegen, die Intervallgrenzen eingeschlossen. Dies ist die *Cauchysche Ungleichung*, benannt nach dem französischen Mathematiker Augustin Louis Cauchy (1789–1857). Wir beweisen diese Ungleichung am Ende von Abschnitt 8.2.

## Basen

Jeder Vektor  $\mathbf{x} = (x_1, x_2, x_3)$  des  $\mathbb{R}^3$  kann in der Form

$$\mathbf{x} = x_1(1, 0, 0) + x_2(0, 1, 0) + x_3(0, 0, 1)$$

als sogenannte *Linearkombination* der Vektoren

$$\mathbf{e}_1 = (1, 0, 0),$$

$$\mathbf{e}_2 = (0, 1, 0),$$

$$\mathbf{e}_3 = (0, 0, 1),$$

geschrieben werden. Offenbar ist diese Darstellung auf genau eine Weise möglich.

Vektoren  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  mit der Eigenschaft, daß man jeden Vektor  $\mathbf{x} \in \mathbb{R}^3$  auf genau eine Weise als Linearkombination  $\mathbf{x} = x_1\mathbf{a} + x_2\mathbf{b} + x_3\mathbf{c}$ ,  $x_1, x_2, x_3 \in \mathbb{R}$ , dieser Vektoren darstellen kann, bilden einer *Basis* des  $\mathbb{R}^3$ .

Außerdem gilt im obigen Fall  $\mathbf{e}_1 \cdot \mathbf{e}_2 = \mathbf{e}_2 \cdot \mathbf{e}_3 = \mathbf{e}_1 \cdot \mathbf{e}_3 = 0$ , und  $\|\mathbf{e}_1\| = \|\mathbf{e}_2\| = \|\mathbf{e}_3\| = 1$ . Das heißt, die Vektoren  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$  stehen paarweise senkrecht aufeinander und sie haben den Betrag eins. Sie bilden eine sogenannte *Orthonormalbasis* des  $\mathbb{R}^3$ .

Basisvektoren müssen aber keineswegs senkrecht aufeinander stehen. So bilden zum Beispiel die Vektoren

$$\mathbf{a} = (1, 0),$$

$$\mathbf{b} = (1, 1),$$

für die  $\mathbf{a} \cdot \mathbf{b} = 1$  ist, eine Basis des  $\mathbb{R}^2$ . Jeder Vektor  $\mathbf{x} = (x_1, x_2)$  des  $\mathbb{R}^2$  läßt sich nämlich auf eindeutige Weise als Linearkombination

$$\mathbf{x} = s\mathbf{a} + t\mathbf{b}$$

mit Zahlen  $s, t \in \mathbb{R}$  schreiben. Denn aus

$$(x_1, x_2) = s(1, 0) + t(1, 1)$$

liest man ab

$$\begin{aligned} x_1 &= s + t, \\ x_2 &= t, \end{aligned}$$

so daß man erhält,

$$\begin{aligned} s &= x_1 - x_2 \\ t &= x_2. \end{aligned}$$

So hat der Vektor  $x = (2, 3)$  bezüglich der Basis  $\mathbf{a}, \mathbf{b}$  die Darstellung

$$\mathbf{x} = (-1)\mathbf{a} + 3\mathbf{b}.$$

Im übrigen kann jede Basis so „normiert“ werden, daß jeder Basisvektor den Betrag eins hat. Ist nämlich  $\mathbf{a} \neq \mathbf{0}$ , dann hat der Vektor  $\mathbf{a}' = \frac{1}{\|\mathbf{a}\|} \cdot \mathbf{a}$  die Länge eins.

## 8.2 Geraden und Ebenen

### Geradengleichungen

Mit Hilfe von Vektoren lassen sich auf sehr effiziente Weise geometrische Sachverhalte beschreiben und geometrische Probleme lösen. Dies zeigen wir im folgenden am Beispiel von Geraden und Ebenen.

Eine Gerade  $g$  in der Ebene oder im Raum ist durch Angabe zweier Punkte  $A$  und  $B$  eindeutig festgelegt.

Sind  $\mathbf{a}$  und  $\mathbf{b}$  die Ortsvektoren der Punkte  $A, B$  in Bezug auf ein cartesisches Koordinatensystem und ist  $X$  ein beliebiger Punkt auf  $g$  mit dem Ortsvektor  $\mathbf{x}$ , dann besteht zwischen den Vektoren  $\mathbf{x}, \mathbf{a}, \mathbf{b}$  eine Beziehung der Form

$$\mathbf{x} = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$$

mit einem *Parameter*  $t \in \mathbb{R}$ . Gibt man umgekehrt zwei Vektoren  $\mathbf{a}$  und  $\mathbf{v}$  vor, dann liegen die Punkte  $X$  mit den Ortsvektoren

$$(*) \quad \mathbf{x} = \mathbf{a} + t\mathbf{v}, \quad t \in \mathbb{R},$$

auf einer Geraden. Diese geht durch den Punkt  $A$  mit dem Ortsvektor  $\mathbf{a}$  und ihre „Richtung“ wird durch den Vektor  $\mathbf{v}$  bestimmt.

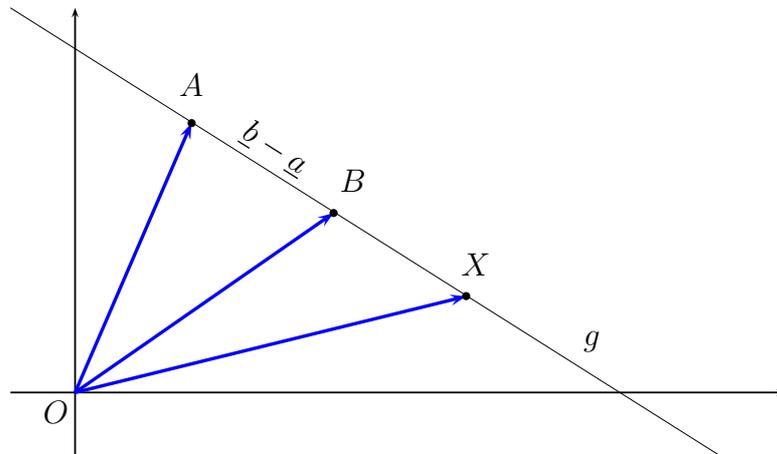


Abbildung 8.6: Gerade durch zwei Punkte

**Bemerkung:** Deutet man den Parameter  $t$  als Maß der Zeit (lateinisch „tempus“), dann beschreibt die Gleichung (\*) die Bewegung eines Massenpunktes durch den Raum mit gleichförmiger Geschwindigkeit  $\mathbf{v}$ . Zum Zeitpunkt  $t = 0$  befindet sich der Punkt an der Stelle  $A$  und bewegt sich in Richtung des Vektors  $\mathbf{v}$  mit der absoluten Geschwindigkeit  $\|\mathbf{v}\|$ . ▲

Eine andere, parameterfreie, Form der Geradengleichung findet man, wenn man einen Punkt  $A$  auf der Geraden  $g$  kennt und einen Vektor  $\mathbf{p}$  der senkrecht auf der Geraden  $g$  steht. Ist  $X$  dann ein weiterer Punkt auf der Geraden, dann steht der Vektor  $\mathbf{x} - \mathbf{a}$

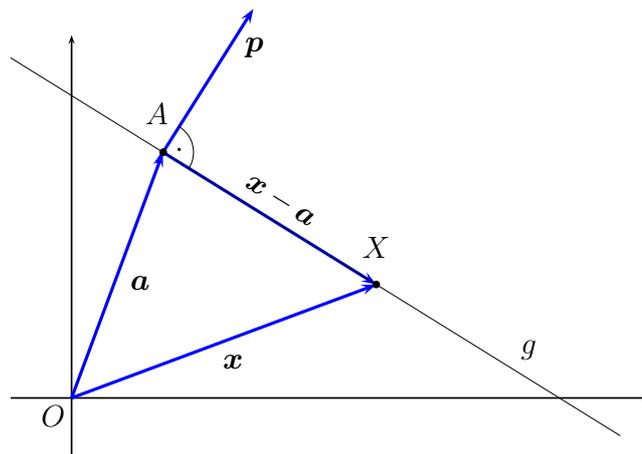


Abbildung 8.7: Zur Hesse-Form der Geradengleichung

senkrecht auf dem Vektor  $\mathbf{p}$ . Also gilt

$$\mathbf{p} \cdot (\mathbf{x} - \mathbf{a}) = 0$$

oder

$$\mathbf{p} \cdot \mathbf{x} = \mathbf{p} \cdot \mathbf{a}$$

Dies ist die *Hesse-Form* der Geradengleichung. Ausgeschrieben lautet sie für eine Gerade in der Ebene

$$(*) \quad p_1 x_1 + p_2 x_2 = \mathbf{p} \cdot \mathbf{a}.$$

### Der Abstand Punkt–Gerade in der Ebene

Ist nun  $Y$  ein beliebiger Punkt der Ebene und  $Y_0$  der Fußpunkt des Lotes von  $Y$  auf  $g$ , dann gibt es einen Parameter  $t \in \mathbb{R}$ , so daß  $\mathbf{y} + t\mathbf{p}$  der Ortsvektor des Punktes  $Y_0$  ist. Wir nehmen zur Vereinfachung der Rechnung an, der Vektor  $\mathbf{p}$  sein normiert, das heißt, es sei  $\|\mathbf{p}\|^2 = \mathbf{p}^2 = 1$ . Ist der Vektor  $\mathbf{p}$  nicht normiert, so muß man bei (\*) zum normierten Vektor übergehen, indem man die Gleichung durch  $\|\mathbf{p}\|$  dividiert.

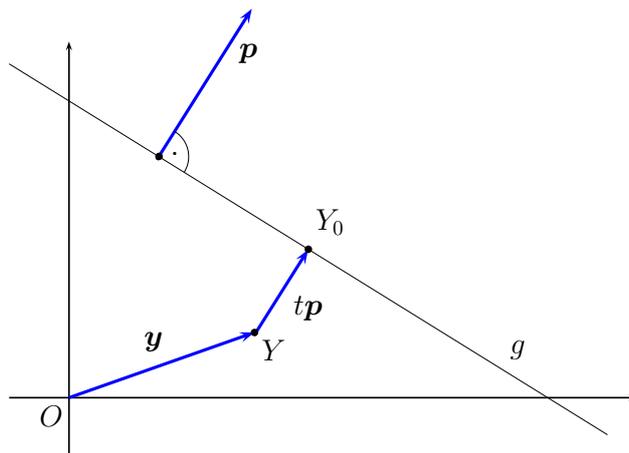


Abbildung 8.8: Abstand eines Punktes von einer Geraden

Da  $Y_0 \in g$  gilt, so hat man

$$\mathbf{p} \cdot (\mathbf{y} + t\mathbf{p}) = \mathbf{p} \cdot \mathbf{a}$$

oder

$$\mathbf{p} \cdot \mathbf{y} + t\|\mathbf{p}\|^2 = \mathbf{p} \cdot \mathbf{y} + t = \mathbf{p} \cdot \mathbf{a}.$$

Somit erhält man

$$t = \mathbf{p} \cdot \mathbf{a} - \mathbf{p} \cdot \mathbf{y}.$$

Ist  $t > 0$ , dann liegt der Punkt  $Y$  auf derjenigen Seite der Geraden, die der Richtung des Vektors  $\mathbf{p}$  abgewandt ist. Im Fall  $t < 0$  zeigt der Vektor  $\mathbf{p}$  auf die Seite von  $g$ , auf der der Punkt  $Y$  liegt.

Der Betrag der Zahl  $t$  ist gleich dem Abstand  $d(Y, g)$  des Punktes  $Y$  von der Geraden  $g$ :  $d(Y, g) = |t|$ .

### Die Achsenabschnittsform der Geradengleichung

Eine oft nützliche Form der Geradengleichung ist die *Achsenabschnittsform*. Wegen (\*) ist jede Beziehung der Form

$$p_1x_1 + p_2x_2 = c$$

mit  $c \in \mathbb{R}$  die Gleichung einer Geraden im  $\mathbb{R}^2$ . Diese geht an der Stelle  $a_1 = c/p_1$  durch die  $x_1$ -Achse und an der Stelle  $a_2 = c/p_2$  durch die  $x_2$ -Achse. Daher kann man diese Beziehung mit den Achsenabschnitten  $a$  und  $b$  auch in der Achsenabschnittsform

$$\frac{x_1}{a_1} + \frac{x_2}{a_2} = 1$$

schreiben. Kennt man daher die Stellen, an der eine Gerade die  $x_1$ - bzw. die  $x_2$ - Achse schneidet, so kann man die Geradengleichung sofort hinschreiben.

### Ebenengleichungen

Eine Ebene ist ein zweidimensionaler Unterraum des dreidimensionalen Raumes  $\mathbb{R}^3$ . Drei Punkte  $A, B, C$  im Raum, die nicht auf einer Geraden liegen bzw. ein Punkt  $P$  und zwei verschiedene Richtungsvektoren  $\mathbf{a}$  und  $\mathbf{b}$ , die an diesen Punkt angeheftet werden, bestimmen eindeutig eine Ebene  $E$ . Sind drei Punkte  $A, B, C$  gegeben, dann bestimmen die Differenzen  $\mathbf{u} = \mathbf{b} - \mathbf{a}$  und  $\mathbf{v} = \mathbf{c} - \mathbf{a}$  der Ortsvektoren die zu  $A$  gehörigen Richtungsvektoren. Ist  $X$  ein Punkt auf der Ebene  $E$ , dann gibt es Zahlen  $s, t \in \mathbb{R}$ , so daß gilt

$$\mathbf{x} = \mathbf{a} + s\mathbf{u} + t\mathbf{v},$$

wenn  $\mathbf{x}$  der Ortsvektor des Punktes  $X$  ist. Durchlaufen die Parameter  $s, t$  alle möglichen Werte, dann durchläuft  $X$  alle Punkte der Ebene.

Analog zur Hesse-Form der Geradengleichung gibt es auch eine parameterfreie Form der Ebenengleichung.

Denkt man sich nämlich einen Punkt  $A$  im Raum gegeben und trägt dort einen Vektor  $\mathbf{p}$  an, dann füllen alle im Punkt  $A$  angehefteten Vektoren, die senkrecht zu  $\mathbf{p}$  sind, eine Ebene aus. Ist  $X$  ein beliebiger Punkt auf  $E$ , dann steht der Vektor  $\mathbf{p}$  senkrecht auf dem Vektor  $\mathbf{x} - \mathbf{a}$ , der Differenz der Ortsvektoren zu den Punkten  $X$  und  $A$ . Daher gilt

$$\mathbf{p} \cdot (\mathbf{x} - \mathbf{a}) = 0$$

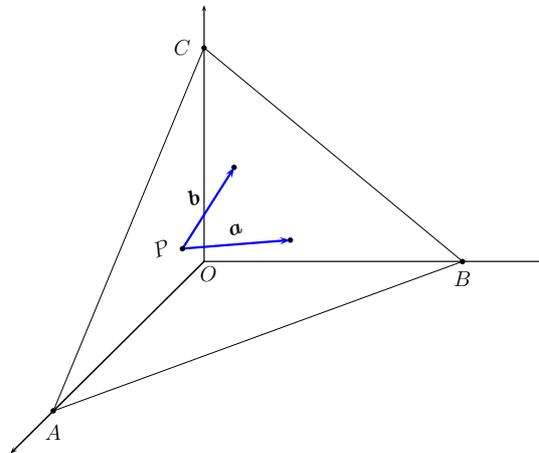


Abbildung 8.9: Ebene durch drei Punkte bzw. Ebene bestimmt durch einen Punkt und zwei Richtungen

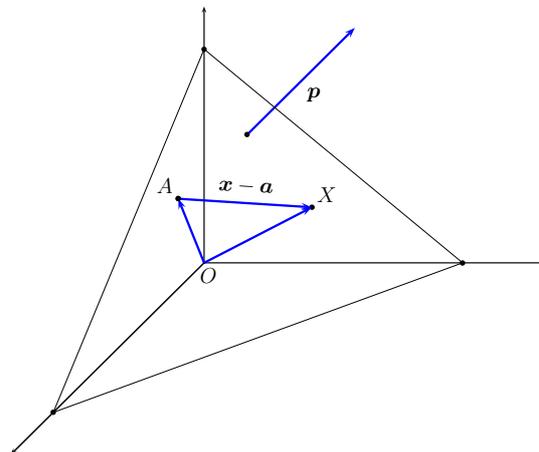


Abbildung 8.10: Zur Hesse-Form der Ebenengleichung

in vollkommener Analogie zum Fall der Geraden. Ausgeschrieben lautet diese Beziehung

$$p_1x_1 + p_2x_2 + p_3x_3 = \mathbf{p} \cdot \mathbf{a}.$$

Der Abstand  $d(Y, E)$  eines Punktes  $Y$  von der Ebene wird in gleicher Weise berechnet wie bei der Geraden, sofern wieder  $\|\mathbf{p}\| = 1$  ist.

Völlig analog lautet auch die *Achsenabschnittsform* der Ebenengleichung. Man erhält die Beziehung

$$\frac{x_1}{a_1} + \frac{x_2}{a_2} + \frac{x_3}{a_3} = 1,$$

wenn die Ebene die drei Achsen in den Punkten  $(a_1, 0, 0)$ ,  $(0, a_2, 0)$ ,  $(0, 0, a_3)$  schneidet.

### Die Cauchysche Ungleichung

Eine Gerade  $g$  im  $\mathbb{R}^n$  durch den Punkt  $A$  und mit dem Richtungsvektor  $\mathbf{b}$  hat die Gleichung

$$\mathbf{x}(t) = \mathbf{a} + t\mathbf{b},$$

wenn  $\mathbf{x}(t)$  der Ortsvektor des zum Parameter  $t$  gehörigen Punktes  $X$  auf  $g$  ist.

Wir wollen nun den Punkt  $X_0$  auf  $g$  bestimmen, der vom Nullpunkt des Koordinatensystems den kleinsten Abstand hat. Wenn  $\|\mathbf{x}(t)\|$  einen minimalen Wert annimmt,

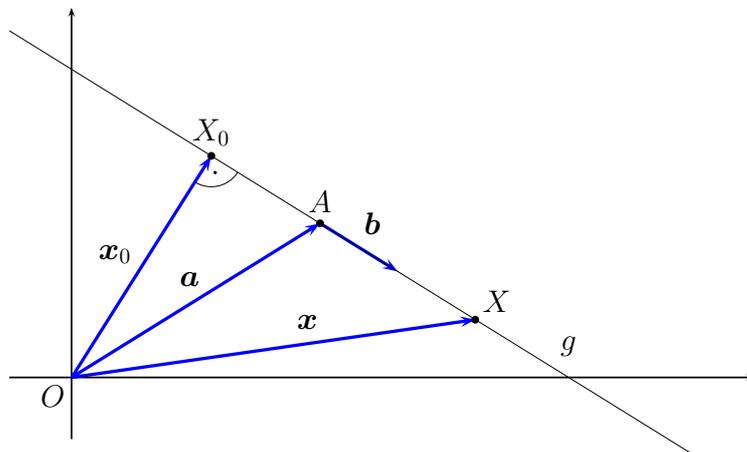


Abbildung 8.11: Der Punkt auf der Geraden mit dem kleinsten Abstand zum Ursprung

dann auch die Funktion  $\varphi(t) = \|\mathbf{x}(t)\|^2 = \mathbf{x}(t) \cdot \mathbf{x}(t)$ . Man erhält

$$\begin{aligned} \varphi(t) &= (\mathbf{a} + t\mathbf{b})^2 \\ (*) \quad &= \mathbf{a}^2 + 2t\mathbf{a} \cdot \mathbf{b} + t^2\mathbf{b}^2. \end{aligned}$$

Zur Berechnung des Minimums der Funktion  $\varphi(t)$  setzen wir

$$\dot{\varphi}(t) = 2\mathbf{a} \cdot \mathbf{b} + 2t\mathbf{b}^2 = 0.$$

Es folgt

$$t_0 = -\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b}^2}$$

und damit ist der Punkt  $X_0$  mit dem Ortsvektor

$$\mathbf{x}_0 = \mathbf{x}(t_0) = \mathbf{a} - \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b}^2} \cdot \mathbf{b}$$

der dem Ursprung nächstgelegene Geradenpunkt. Man rechnet leicht nach, daß  $\mathbf{x}_0 \cdot \mathbf{b} = 0$  ist, die Vektoren  $\mathbf{x}_0$  und  $\mathbf{b}$  also senkrecht aufeinander stehen.

Da  $\varphi(t) \geq 0$  ist gilt wegen(\*)

$$\mathbf{a}^2 - \frac{2(\mathbf{a} \cdot \mathbf{b})^2}{\mathbf{b}^2} + \frac{(\mathbf{a} \cdot \mathbf{b})^2}{\mathbf{b}^2} \geq 0$$

und damit

$$(\mathbf{a} \cdot \mathbf{b})^2 \leq \mathbf{a}^2 \mathbf{b}^2.$$

Diese Ungleichung besagt wiederum, daß gilt

$$-\|\mathbf{a}\| \|\mathbf{b}\| \leq \mathbf{a} \cdot \mathbf{b} \leq \|\mathbf{a}\| \|\mathbf{b}\|$$

oder

$$(CU) \quad -1 \leq \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \leq 1.$$

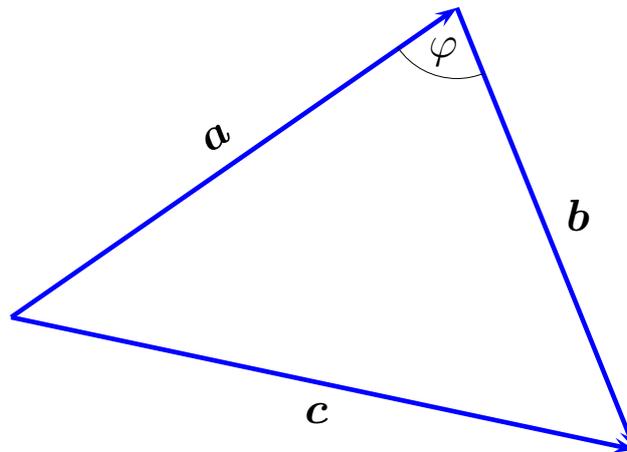
Das ist die *Cauchysche Ungleichung*, auch *Cauchy–Schwarzsche Ungleichung* genannt. Sie zeigt unter anderem, daß der Quotient  $\mathbf{a} \cdot \mathbf{b} / \|\mathbf{a}\| \|\mathbf{b}\|$  als Cosinus eines Winkels  $\varphi$  gedeutet werden kann, den man dann als den von den Vektoren  $\mathbf{a}$ ,  $\mathbf{b}$  des  $\mathbb{R}^n$  eingeschlossenen Winkel auffasst. Damit ist die Beziehung

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \varphi$$

auch für Vektoren  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  gerechtfertigt (vgl. Abschnitt 8.1).

### Dreiecksungleichung und Satz des Pythagoras

Als geometrische Anwendung der Cauchyschen–Ungleichung beweisen wir eine fundamentale Ungleichung, die *Dreiecksungleichung*. Dabei ergibt sich auch der aus der Schulgeometrie bekannte *Satz des Pythagoras*. Für Vektoren  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  des  $\mathbb{R}^n$  gelte

Abbildung 8.12: Das von drei Vektoren  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  gebildete Dreieck

$$\mathbf{a} + \mathbf{b} = \mathbf{c}.$$

Quadriert man beide Seiten im Sinne des Skalarproduktes, dann folgt

$$(\mathbf{a} + \mathbf{b})^2 = \mathbf{a}^2 + 2\mathbf{a} \cdot \mathbf{b} + \mathbf{b}^2,$$

das heißt, für die Beträge der beteiligten Vektoren erhält man die Beziehung

$$(*) \quad \|\mathbf{a} + \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + 2\mathbf{a} \cdot \mathbf{b} + \|\mathbf{b}\|^2.$$

Schließen die Vektoren  $\mathbf{a}$ ,  $\mathbf{b}$  einen rechten Winkel ein, dann ist  $\mathbf{a} \cdot \mathbf{b} = 0$  und es folgt

$$\|\mathbf{a}\|^2 + \|\mathbf{b}\|^2 = \|\mathbf{c}\|^2.$$

Das ist der *Satz des Pythagoras*. (Benannt nach dem griechischen Philosophen und Mathematiker Pythagoras ca. 600 oder 570–509 v. Chr.)

Aufgrund der Cauchyschen Ungleichung in der Form

$$|\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\| \|\mathbf{b}\|$$

kann man bei (\*) abschätzen,

$$\|\mathbf{a}\|^2 + 2\mathbf{a} \cdot \mathbf{b} + \|\mathbf{b}\|^2 \leq \|\mathbf{a}\|^2 + 2\|\mathbf{a}\| \|\mathbf{b}\| + \|\mathbf{b}\|^2.$$

Die rechte Seite dieser Ungleichung ist aber gleich

$$(\|\mathbf{a}\| + \|\mathbf{b}\|)^2.$$

Also besteht die Beziehung

$$\|\mathbf{a} + \mathbf{b}\|^2 \leq (\|\mathbf{a}\| + \|\mathbf{b}\|)^2.$$

und daher gilt

$$(DU) \quad \|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|.$$

Das ist die *Dreiecksungleichung* die auch anschaulich plausibel ist. – Haben die Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  die gleiche Richtung, das heißt, gilt  $\mathbf{b} = t\mathbf{a}$  mit  $t > 0$ , dann ist

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a} \cdot (t\mathbf{a}) = t\mathbf{a}^2 = t\|\mathbf{a}\|^2$$

und daher gilt wegen (\*)

$$\begin{aligned} \|\mathbf{a} + \mathbf{b}\|^2 &= \|\mathbf{a}\|^2 + 2t\|\mathbf{a}\|^2 + t^2\|\mathbf{a}\|^2 \\ &= (\|\mathbf{a}\| + t\|\mathbf{a}\|)^2 \\ &= (\|\mathbf{a}\| + \|\mathbf{b}\|)^2. \end{aligned}$$

In diesem Fall ist bei (DU) also das Gleichheitszeichen zu setzen.

### 8.3 Matrizen

Zur Einführung des Matrizenbegriffs betrachten wir zunächst zwei Beispiele.

**Beispiel 1 (Drehungen):** Ein Vektor  $\mathbf{x} = (x_1, x_2)$  des  $\mathbb{R}^2$ , angeheftet im Nullpunkt eines cartesischen Koordinatensystems, werde um den Winkel  $\varphi$  gedreht und gehe so in den Vektor  $\mathbf{y} = (y_1, y_2)$  des  $\mathbb{R}^2$  über. Wie berechnet man die Komponenten  $y_1, y_2$  des Vektors  $\mathbf{y} = D(\varphi)\mathbf{x}$  aus den Komponenten des Vektors  $\mathbf{x}$ , wenn  $D(\varphi)$  die Drehung um den Winkel  $\varphi$  bedeutet? Fasst man die Vektoren  $\mathbf{x}, \mathbf{y}$  als komplexe Zahlen  $x = x_1 + ix_2$  und  $y = y_1 + iy_2$  auf, dann wird die Drehung  $D(\varphi)$  um den Winkel  $\varphi$  durch Multiplikation mit der komplexen Zahl  $e^{i\varphi} = \cos \varphi + i \sin \varphi$  realisiert (vgl. Abschnitt 2.4):

$$y = e^{i\varphi}x$$

Ausgeschrieben erhält man so die Beziehung

$$\begin{aligned} y_1 + iy_2 &= (\cos \varphi + i \sin \varphi) \cdot (x_1 + ix_2) \\ &= \cos \varphi \cdot x_1 - \sin \varphi \cdot x_2 + i(\sin \varphi \cdot x_1 + \cos \varphi \cdot x_2). \end{aligned}$$

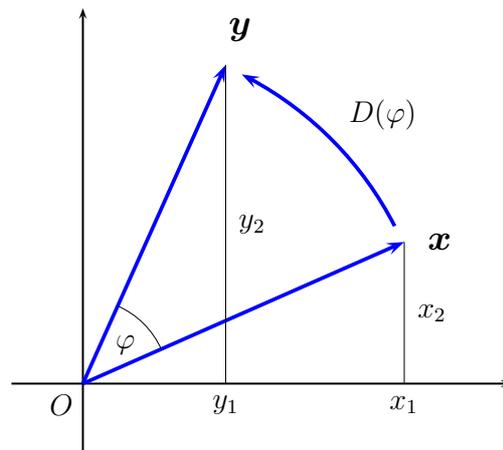


Abbildung 8.13: Drehung eines Vektors

Durch Vergleich von Realteil und Imaginärteil beider Seiten dieser Beziehung findet man dann den Zusammenhang

$$y_1 = \cos \varphi \cdot x_1 - \sin \varphi \cdot x_2$$

(\*)

$$y_2 = \sin \varphi \cdot x_1 + \cos \varphi \cdot x_2.$$

Der Übergang  $\mathbf{x} \rightarrow D(\varphi)\mathbf{x}$  ist eine *Abbildung*  $D(\varphi) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ . Sie wird beschrieben durch das quadratische Zahlenschema

$$D(\varphi) = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}$$

Ein solches Zahlenschema mit zwei „Zeilen“ und zwei „Spalten“ ist eine  $(2 \times 2)$ -*Matrix*. Den Zusammenhang (\*) schreibt man dann in der Form

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Die beteiligten Vektoren  $\mathbf{x}$  und  $\mathbf{y}$  werden in diesen Kontext als „Spaltenvektoren“ notiert. Die rechte Seite wird als Produkt Matrix·Vektor aufgefasst. Das Resultat dieses Produktes ist wiederum ein Vektor.

Das Produkt Matrix·Vektor lässt sich auch mit Hilfe des Skalarproduktes der Zeilenvektoren  $(\cos \varphi, -\sin \varphi)$ ,  $(\sin \varphi, \cos \varphi)$  der Matrix  $D(\varphi)$  schreiben. Offenbar ist

$$y_1 = (\cos \varphi, -\sin \varphi) \cdot (x_1, x_2)$$

und

$$y_2 = (\sin \varphi, \cos \varphi) \cdot (x_1, x_2).$$

Man rechnet leicht nach, daß  $\|y\| = \|x\|$  ist, der Vektor  $y$  also tatsächlich durch eine Drehung aus dem Vektor  $x$  hervorgeht.

Die Drehung um  $90^\circ$  wird wegen  $\cos 90^\circ = 0$  und  $\sin 90^\circ = 1$  durch die Matrix

$$D(90^\circ) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

gegeben, so dass

$$D(90^\circ) \cdot (x_1, x_2) = (-x_2, x_1)$$

ist und somit in der Tat

$$(x_1, x_2) \cdot (-x_2, x_1) = 0$$

gilt. Die Vektoren  $x = (x_1, x_2)$  und  $y = (-x_2, x_1)$  haben den gleichen Betrag  $\|x\| = \|y\|$  und sie stehen senkrecht aufeinander,  $x \perp y$ . Damit haben wir mit Hilfe des Skalarproduktes noch einmal bestätigt, daß die obige Matrix  $D(90)$  eine Drehung um  $90^\circ$  realisiert. ■

**Beispiel 2 (Scherung):** Eine Gerade  $s$  durch den Nullpunkt des  $\mathbb{R}^2$ , die im ersten Quadranten eines cartesischen Koordinatensystems verläuft und mit der  $x_2$ -Achse den Winkel  $\alpha$  einschließt, definiert in folgender Weise eine Abbildung  $S : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ : Durch einen Punkt  $X$  oberhalb der  $x_1$ -Achse lege man eine zu dieser parallele Gerade, welche die  $x_2$ -Achse im Punkte  $A$  und die Gerade  $s$  im Punkte  $B$  schneidet. Sodann verschiebe man den Punkt  $X$  um die Strecke  $|AB|$  in  $x_1$ -Richtung in den Punkt  $Y$ . Diese Abbildung, die auf diese Weise den Punkt  $X$  in den Punkt  $Y$  überführt, ist eine sogenannte *Scherung*. Sie werde mit  $S$  bezeichnet. Die Gerade  $s$  ist die Scherungsachse,  $\alpha$  ist der Scherungswinkel. Die Scherungsabbildung kann als vereinfachtes Modell eines *Erdrutsches* aufgefasst werden.

Aus der Abbildung liest man den Zusammenhang zwischen den Ortsvektoren  $x$  und  $y$  der Punkte  $X, Y$  ab. Es gilt offenbar

$$y_1 = x_1 + |AB|,$$

und

$$y_2 = x_2.$$

Wegen

$$\frac{|AB|}{x_2} = \tan \alpha = a$$

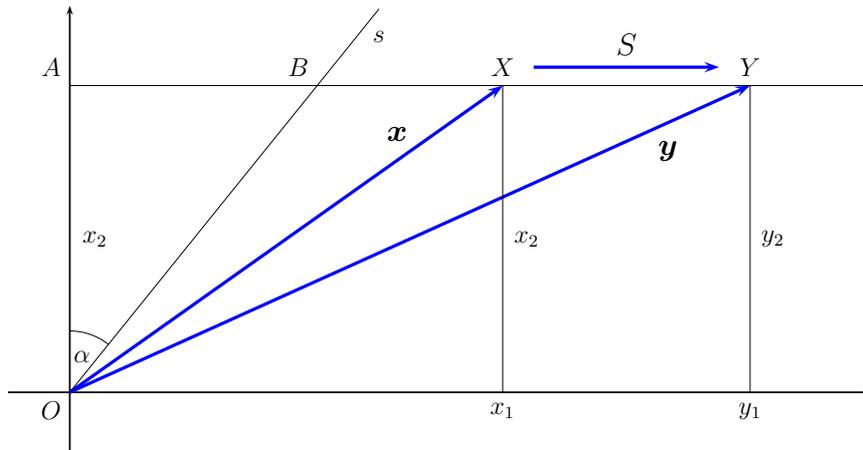


Abbildung 8.14: Scherung eines Vektors

hat man daher die Beziehungen

$$y_1 = x_1 + ax_2$$

$$y_2 = x_2.$$

Mit Hilfe der Scherungsmatrix

$$S(\alpha) = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

kann man diese dann in der Form

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

schreiben. ■

### Das Produkt Matrix mal Vektor

Eine  $(3 \times 3)$ -Matrix ist ein Zahlenschema der Form

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

mit Zahlen  $a_{j,k}$ ,  $j, k = 1, 2, 3$ , den *Elementen* der Matrix. Der erste Index  $j$  bezeichnet die Zeile, in der das Element steht, der zweite Index  $k$  die Spalte. Die Elemente  $a_{11}, a_{22}, a_{33}$  mit gleichen Indizes stehen in der *Hauptdiagonalen* der Matrix. Die andere Diagonale ist die *Nebendiagonale*.

Eine  $(3 \times 3)$ -Matrix führt einen Vektor  $\mathbf{x} = (x_1, x_2, x_3)$  des  $\mathbb{R}^3$ , geschrieben als Spaltenvektor, nach folgender Regel in einen Vektor  $\mathbf{y} = (y_1, y_2, y_3)$  des  $\mathbb{R}^3$  über:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{pmatrix}.$$

Die Zeilen auf der rechten Seite lassen sich als Skalarprodukt der Zeilenvektoren  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  der Matrix  $A$  und des Vektors  $\mathbf{x}$  deuten. Das heißt,

$$\mathbf{y} = A \cdot \mathbf{x} = \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{x} \\ \mathbf{a}_2 \cdot \mathbf{x} \\ \mathbf{a}_3 \cdot \mathbf{x} \end{pmatrix}.$$

Als Rechenregeln für das Produkt Matrix  $\cdot$  Vektor hat man, in Analogie zum Rechnen mit Zahlen, die Regeln

$$A \cdot (t\mathbf{x}) = t(A \cdot \mathbf{x}), \quad t \in \mathbb{R},$$

und

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

### Die Addition von Matrizen

Ebenso wie man Vektoren komponentenweise addiert, so addiert man Matrizen  $A = (a_{jk}), B = (b_{jk}), j, k = 1, 2, 3$ , elementweise. Die Matrix

$$C = A + B$$

ist die Matrix mit den Elementen

$$c_{jk} = a_{jk} + b_{jk}.$$

**Beispiel:** Die Summe der Matrizen

$$A = \begin{pmatrix} 1 & 2 \\ -3 & 4 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 0 & 1 \\ 5 & -2 \end{pmatrix}$$

ist die Matrix

$$C = \begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix}.$$



Die Reihenfolge ist bei der Addition von Matrizen offenbar ohne Bedeutung. Es gilt also

$$A + B = B + A \quad (\text{Kommutativgesetz}).$$

Die Rolle der Null bei der Addition von Zahlen spielt bei der Matrizenaddition die Nullmatrix

$$O = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

deren Elemente sämtlich gleich Null sind. Das heißt, es gilt

$$A + O = O + A = A.$$

Bezüglich des Produktes Matrix  $\cdot$  Vektor gilt

$$(A + B) \cdot \mathbf{x} = A \cdot \mathbf{x} + B \cdot \mathbf{x} \quad (\text{Distributivgesetz}),$$

analog wie beim Rechnen mit Zahlen.

### Die Multiplikation von Matrizen

Anders als die Regel für die Addition von Matrizen, liegt die Regel für deren *Multiplikation* nicht ganz an der Oberfläche. Um sie auf natürliche Weise zu erhalten, bestimmen wir das Resultat einer Hintereinanderschaltung der Multiplikation Matrix  $\cdot$  Vektor mit zwei  $(3 \times 3)$ -Matrizen  $A, B$  und einem Vektor  $\mathbf{x}$  des  $\mathbb{R}^3$ . Wir berechnen also den Vektor

$$\mathbf{y} = A \cdot (B \cdot \mathbf{x}).$$

Es zeigt sich, daß diese Transformation des Vektors  $\mathbf{x}$  durch die Wirkung einer *einzi- gen* Matrix  $C$  auf  $\mathbf{x}$  ersetzt werden kann, so daß gilt

$$A \cdot (B \cdot \mathbf{x}) = C \cdot \mathbf{x}.$$

Diese Matrix  $C$  betrachtet man dann als das Produkt der Matrizen  $A$  und  $B$  und schreibt

$$A \cdot B = C.$$

Mit den Zeilenvektoren  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$  der Matrix  $B$  erhält man zunächst

$$B \cdot \mathbf{x} = \begin{pmatrix} \mathbf{b}_1 \cdot \mathbf{x} \\ \mathbf{b}_2 \cdot \mathbf{x} \\ \mathbf{b}_3 \cdot \mathbf{x} \end{pmatrix}$$

und mit den Zeilenvektoren  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$  der Matrix  $A$  dann

$$\begin{aligned} \mathbf{y} &= A \cdot (B \cdot \mathbf{x}) = \begin{pmatrix} \mathbf{a}_1 \cdot (\mathbf{b}_1 \cdot \mathbf{x}, \mathbf{b}_2 \cdot \mathbf{x}, \mathbf{b}_3 \cdot \mathbf{x}) \\ \mathbf{a}_2 \cdot (\mathbf{b}_1 \cdot \mathbf{x}, \mathbf{b}_2 \cdot \mathbf{x}, \mathbf{b}_3 \cdot \mathbf{x}) \\ \mathbf{a}_3 \cdot (\mathbf{b}_1 \cdot \mathbf{x}, \mathbf{b}_2 \cdot \mathbf{x}, \mathbf{b}_3 \cdot \mathbf{x}) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{a}_{11} \cdot (\mathbf{b}_1 \cdot \mathbf{x}) + \mathbf{a}_{12} \cdot (\mathbf{b}_2 \cdot \mathbf{x}) + \mathbf{a}_{13} \cdot (\mathbf{b}_3 \cdot \mathbf{x}) \\ \mathbf{a}_{21} \cdot (\mathbf{b}_1 \cdot \mathbf{x}) + \mathbf{a}_{22} \cdot (\mathbf{b}_2 \cdot \mathbf{x}) + \mathbf{a}_{23} \cdot (\mathbf{b}_3 \cdot \mathbf{x}) \\ \mathbf{a}_{31} \cdot (\mathbf{b}_1 \cdot \mathbf{x}) + \mathbf{a}_{32} \cdot (\mathbf{b}_2 \cdot \mathbf{x}) + \mathbf{a}_{33} \cdot (\mathbf{b}_3 \cdot \mathbf{x}) \end{pmatrix}. \end{aligned}$$

Ordnet man in jeder Zeile auf der rechten Seite nach Faktoren bei  $x_1, x_2, x_3$ , dann erhält man für die erste Zeile

$$\begin{aligned} y_1 &= (a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31})x_1 + \\ &\quad (a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32})x_2 + \\ &\quad (a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33})x_3. \end{aligned}$$

Die Klammern bei  $x_1, x_2, x_3$  lassen sich auffassen als Skalarprodukte des Zeilenvektors  $\mathbf{a}_1$  von  $A$  mit den Spaltenvektoren  $\mathbf{b}^1, \mathbf{b}^2, \mathbf{b}^3$  von  $B$ :

$$y_1 = (\mathbf{a}_1 \cdot \mathbf{b}^1)x_1 + (\mathbf{a}_1 \cdot \mathbf{b}^2)x_2 + (\mathbf{a}_1 \cdot \mathbf{b}^3)x_3.$$

Für die beiden übrigen Zeilen erhält man entsprechend die Darstellungen

$$\begin{aligned} y_2 &= (\mathbf{a}_2 \cdot \mathbf{b}^1)x_1 + (\mathbf{a}_2 \cdot \mathbf{b}^2)x_2 + (\mathbf{a}_2 \cdot \mathbf{b}^3)x_3, \\ y_3 &= (\mathbf{a}_3 \cdot \mathbf{b}^1)x_1 + (\mathbf{a}_3 \cdot \mathbf{b}^2)x_2 + (\mathbf{a}_3 \cdot \mathbf{b}^3)x_3. \end{aligned}$$

Die Transformation des Vektors  $\mathbf{x}$  in den Vektor  $\mathbf{y}$  wird also geleistet von der Matrix  $C = (c_{jk})$  mit den Elementen

$$c_{jk} = \mathbf{a}_j \cdot \mathbf{b}^k, \quad j, k = 1, 2, 3,$$

dem Produkt der Matrizen  $A$  und  $B$ .

Die Regel zur Bildung des Matrizenproduktes  $A \cdot B = C$  lässt sich durch das folgende Schema veranschaulichen:

$$\begin{array}{ccc} & & k \\ & & \vdots \\ j & \begin{pmatrix} \cdots & \mathbf{a}_j & \cdots \end{pmatrix} & \cdot \begin{pmatrix} \vdots \\ \mathbf{b}^k \\ \vdots \end{pmatrix} = j \begin{pmatrix} \cdots & \cdots & c_{jk} & \cdots \\ \vdots \\ \vdots \end{pmatrix} \\ & A & B & C \end{array}$$

Aus diesem Schema liest man ab: Damit das Matrizenprodukt  $A \cdot B$  gebildet werden kann, muß die Anzahl der Spalten von  $A$  gleich der Anzahl der Zeilen von  $B$  sein. Für das Matrizenprodukt gelten analog den entsprechenden Regeln für Zahlen die Rechenregeln

$$(A \cdot B) \cdot C = A \cdot (B \cdot C) \quad (\text{Assoziativgesetz})$$

und

$$A \cdot (B + C) = A \cdot B + A \cdot C \quad (\text{Distributivgesetz})$$

Das Produkt *Zahl*·*Matrix* ist folgendermaßen erklärt: Die Matrix  $t \cdot A$ ,  $t \in \mathbb{R}$ , ist die Matrix mit den Elementen  $t a_{jk}$ . Man hat also jedes Element der Matrix  $A$  mit der Zahl  $t$  zu multiplizieren.

Im Kontext der Matrizenmultiplikation gilt dann die Regel

$$A \cdot (tB) = (tA) \cdot B = t(A \cdot B).$$

Im Unterschied zur Zahlenmultiplikation ist die Matrizenmultiplikation im allgemeinen *nicht kommutativ*:

$$A \cdot B \neq B \cdot A$$

**Beispiel:** Wegen  $\cos 60^\circ = \frac{1}{2}$  und  $\sin 60^\circ = \frac{\sqrt{3}}{2}$  realisiert die Matrix

$$D(60^\circ) = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}$$

eine Drehung um  $60^\circ$  (siehe Beispiel 1, oben). Zum Scherungswinkel  $\alpha = 45^\circ$  gehört die Scherung

$$S(45^\circ) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

da  $\tan 45^\circ = 1$  ist (siehe Beispiel 2, oben). Dann gilt

$$\begin{aligned} S(45^\circ) \cdot D(60^\circ) &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 + \sqrt{3} & 1 - \sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix}. \end{aligned}$$

Hingegen ist

$$\begin{aligned} D(60^\circ) \cdot S(45^\circ) &= \frac{1}{2} \begin{pmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 & 1 - \sqrt{3} \\ \sqrt{3} & 1 + \sqrt{3} \end{pmatrix}. \end{aligned}$$

Das Ergebnis der Hintereinanderausführung von Drehung und Scherung hängt also davon ab, in welcher Reihenfolge diese Abbildungen ausgeführt werden, was man sich auch anschaulich leicht klar macht. ■

Die Rolle der Null bei der Matrizenmultiplikation spielt die Nullmatrix  $O$ , das heißt es gilt

$$O \cdot A = A \cdot O = O$$

für jede Matrix  $A$ .

Die Rolle der Eins spielt die *Einheitsmatrix*  $I$ , die in der Hauptdiagonalen nur Einsen hat und deren übrige Elemente Nullen sind. Unter den  $(3 \times 3)$ -Matrizen ist dies die Matrix

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Es gilt dann

$$I \cdot A = A \cdot I = A,$$

wie man leicht nachrechnet.

Es stellt sich sogleich die Frage, ob es zu einer Matrix  $A$  eine inverse Matrix  $A^{-1}$  gibt, so daß gilt

$$A^{-1} \cdot A = A \cdot A^{-1} = I,$$

und wie man gegebenenfalls die zu  $A$  Inverse Matrix  $A^{-1}$  berechnet.

**Bemerkung: Eindeutige Bestimmtheit von Einheitsmatrix und inverser Matrix**

Es wurde oben lediglich gezeigt, daß die Diagonalmatrix  $I$  tatsächlich die Rolle der Eins bei der Matrizenmultiplikation spielt. Damit ist aber nicht ausgeschlossen, daß es eine weitere Matrix  $I' \neq I$  mit den gleichen Eigenschaften gibt. Dann aber müßte einerseits gelten

$$I \cdot I' = I' \cdot I = I'$$

und andererseits

$$I' \cdot I = I \cdot I' = I.$$

Also muß, entgegen der Annahme, doch  $I' = I$  sein. Aus der eindeutigen Bestimmtheit der Einheitsmatrix wiederum folgt: Wenn  $A^{-1}$  eine Inverse zur Matrix  $A$  ist, so daß gilt  $A^{-1} \cdot A = I$ , dann gilt auch  $A \cdot A^{-1} = I$ . Multipliziert man nämlich die erste Gleichung von links mit der Matrix  $A$ , dann folgt aus der Assoziativität der Multiplikation die Beziehung

$$A \cdot (A^{-1} \cdot A) = (A \cdot A^{-1}) \cdot A = A \cdot I = A$$

und aus der eindeutigen Bestimmtheit der Einheitsmatrix ergibt sich somit  $A \cdot A^{-1} = I$ , was zu zeigen war. – Die eindeutige Bestimmtheit der Inversen einer Matrix folgt wiederum analog zur Überlegung bei der Eindeutigkeit der Einheitsmatrix. ▲

### Die Inverse einer $(2 \times 2)$ -Matrix

Wir schreiben eine  $(2 \times 2)$ -Matrix  $A$  in der Form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Ist dann  $\mathbf{x} = (x_1, x_2)$  ein Vektor des  $\mathbb{R}^2$ , dann ist

$$(*) \quad A\mathbf{x} = \mathbf{y}$$

ebenfalls ein Vektor des  $\mathbb{R}^2$ . Besitzt die Matrix  $A$  eine Inverse  $A^{-1}$ , dann gilt

$$A^{-1} \cdot A\mathbf{x} = I\mathbf{x} = \mathbf{x} = A^{-1}\mathbf{y}.$$

Isoliert man daher bei (\*) den Vektor  $\mathbf{x}$ , dann erscheint die gesuchte Matrix  $A^{-1}$  als „Faktor“ beim Vektor  $\mathbf{y}$ .

Ausgeschrieben lautet die Beziehung (\*)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

oder

$$\begin{array}{l} ax_1 + bx_2 = y_1 \\ cx_1 + dx_2 = y_2 \end{array} \left| \begin{array}{l} d \\ -b \end{array} \right| \begin{array}{l} -c \\ a \end{array}$$

Multipliziert man die erste Gleichung mit  $d$ , die zweite mit  $-b$  und addiert beide Gleichungen, dann entfallen die Terme mit  $x_2$  und man erhält

$$(ad - bc)x_1 = dy_1 - by_2.$$

Multipliziert man die erste Zeile mit  $-c$  und die zweite mit  $a$ , dann entfallen bei der Addition der Zeilen die  $x_1$ -Terme. Man erhält

$$(ad - bc)x_2 = -cy_1 + ay_2.$$

Ist der Ausdruck

$$D = ad - bc \neq 0,$$

dann kann man also schreiben,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{D} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Damit haben wir das Ergebnis: Ist  $D \neq 0$ , dann besitzt die Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

die Inverse

$$A^{-1} = \frac{1}{D} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Diese entsteht also aus der Matrix  $A$ , indem man die Elemente in der Hauptdiagonalen vertauscht, die Elemente in der Nebendiagonalen mit  $-1$  multipliziert und die so gebildete Matrix mit dem Faktor  $1/D$  versieht.

Die Zahl  $D = ad - bc$  ist die *Determinante* der Matrix  $A$ . Man schreibt auch

$$D = \det(A) = |A|.$$

Der Wert der Determinante von  $A$  entscheidet also darüber, ob  $A$  invertierbar ist oder nicht. Jede quadratische Matrix besitzt eine solche Determinante. Diese ist aber im Falle von quadratischen Matrizen mit mehr als zwei Zeilen nicht mehr auf so einfache Weise zu berechnen. Ist  $D = 0$ , also  $ad - bc = 0$ , dann gilt

$$\frac{c}{a} = \frac{d}{b} = t.$$

Also ist  $c = t \cdot a$  und  $d = t \cdot b$  oder  $(c, d) = t \cdot (a, b)$ . Die Zeilenvektoren der Matrix  $A$  unterscheiden sich nur durch einen konstanten Faktor. Dasselbe gilt auch von dem Spaltenvektoren, wie man leicht zeigt. Nach demselben Verfahren wie im Fall einer  $(2 \times 2)$ -Matrix, nämlich durch Auflösung eines linearen Gleichungssystems, berechnet man auch bei quadratischen Matrizen mit mehr als zwei Zeilen die Inverse (vgl.

hierzu Abschnitt 8.6).

**Geometrische Deutung der Determinate:** Bei der Einführung des Skalarproduktes (vgl. Abschnitt 8.1) repräsentieren wir zwei Vektoren  $\mathbf{a} = (a_1, a_2)$ ,  $\mathbf{b} = (b_1, b_2)$  des  $\mathbb{R}^2$  durch die entsprechenden komplexen Zahlen  $a = a_1 + i a_2 = |a| e^{i\alpha}$  und  $b = b_1 + i b_2 = |b| e^{i\beta}$ . Dann gilt

$$(*) \quad \bar{a}b = (a_1 - i a_2) \cdot (b_1 + i b_2) = |a| |b| e^{i\varphi}$$

mit  $\varphi = \beta - \alpha$ . Multipliziert man die linke Seite aus und wendet auf der rechten Seite die Eulerformel an, dann findet man die Beziehung

$$a_1 b_1 + a_2 b_2 + i(a_1 b_2 - a_2 b_1) = |a| |b| (\cos \varphi + i \sin \varphi).$$

Durch Vergleich der Imaginärteile folgt die Identität

$$a_1 b_2 - a_2 b_1 = |a| |b| \sin \varphi.$$

Der Ausdruck auf der linken Seite ist gleich der Determinante der Matrix

$$A = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix},$$

die rechte Seite gibt den Flächeninhalt des von den Vektoren  $\mathbf{a}$ ,  $\mathbf{b}$  aufgespannten Parallelogramms an. Die Fläche des Dreiecks mit den Seitenvektoren  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{a} - \mathbf{b}$  ist nämlich

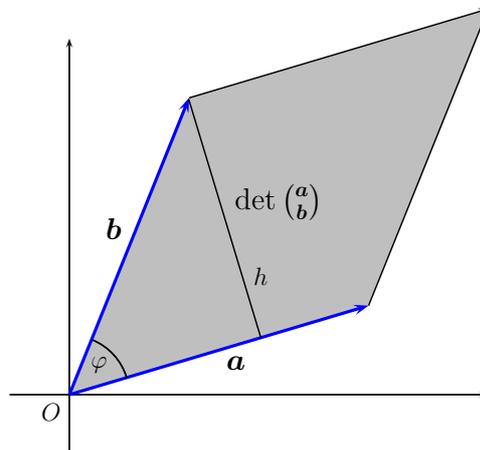


Abbildung 8.15: Zur geometrischen Deutung der Determinante

gleich  $\frac{1}{2} |a| h = \frac{1}{2} |a| |b| \sin \varphi$ . Da dieses Dreieck im Parallelogramm zweimal auftritt, ist  $|a| |b| \sin \varphi$  die Fläche des Parallelogramms.

Die Determinante der Matrix  $A$  ist also gleich der Fläche des von den Vektoren  $\mathbf{a}$ ,  $\mathbf{b}$

aufgespannten Parallelogramms. Ist  $\det A = 0$ , dann haben die Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  die gleiche oder entgegengesetzte Richtung das heißt es gilt  $\mathbf{b} = t\mathbf{a}$  mit  $t \in \mathbb{R}$ .

Wegen  $\operatorname{Re}(\bar{a}b) = \mathbf{a} \cdot \mathbf{b}$  (vgl. Abschnitt 8.1) kann man die Beziehung (\*) auch in der Form

$$\bar{a}b = \mathbf{a} \cdot \mathbf{b} + i \det \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$$

schreiben.

## 8.4 Eigenvektoren und Eigenwerte

Eine  $(n \times n)$ -Matrix  $A$  führt einen Vektor  $\mathbf{x}$  des  $\mathbb{R}^n$  in einen Vektor  $\mathbf{y} = A\mathbf{x}$  des  $\mathbb{R}^n$  über. Der Vektor  $\mathbf{y}$  ist im allgemeinen dem Betrage und der Richtung nach vom Vektor  $\mathbf{x}$  verschieden. Es kann aber auch der Fall eintreten, daß die Vektoren  $\mathbf{x}$  und  $\mathbf{y}$  *dieselbe* (oder die entgegengesetzte Richtung) besitzen. So führt zum Beispiel die Matrix

$$A = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}$$

den Vektor  $\mathbf{x} = (1, 1)$  in den Vektor

$$\mathbf{y} = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\mathbf{x}$$

über. Der Vektor  $\mathbf{x}$  heißt dann *Eigenvektor* und die Zahl 2 ist ein zugehöriger *Eigenwert* der Matrix  $A$ . – Allgemein gilt:

**Definition. (Eigenvektor, Eigenwert):** Ist  $A$  eine  $(n \times n)$ -Matrix,  $\mathbf{x}$  ein Vektor des  $\mathbb{R}^n$ , der nicht der Nullvektor ist, und  $\lambda$  eine Zahl, so daß die Beziehung

$$(*) \quad A\mathbf{x} = \lambda\mathbf{x}$$

besteht, dann ist  $\mathbf{x}$  ein Eigenvektor und  $\lambda$  ein zugehöriger Eigenwert der Matrix  $A$ .

Man erkennt: Erfüllt der Vektor  $\mathbf{x}$  die Gleichung (\*), dann auch der Vektor  $t \cdot \mathbf{x}$  mit  $t \in \mathbb{R}$ . Denn man hat

$$A \cdot (t\mathbf{x}) = t \cdot (A\mathbf{x}) = t(\lambda\mathbf{x}) = \lambda(t\mathbf{x}).$$

Mit jedem Eigenvektor  $\mathbf{x}$  ist daher auch  $t \cdot \mathbf{x}$  mit  $t \neq 0$  Eigenvektor der Matrix  $A$ . Zur Berechnung von Eigenvektoren und Eigenwerten einer Matrix geben wir der Beziehung (\*) zunächst die Form

$$(**) \quad (A - \lambda I)\mathbf{x} = \mathbf{0},$$

wobei  $I$  die Einheitsmatrix ist und  $\mathbf{0}$  der Nullvektor. Hätte die Matrix  $A - \lambda I$  eine Inverse  $(A - \lambda I)^{-1}$ , so könnte man die Gleichung (\*\*\*) mit dieser von links multiplizieren und erhielte

$$\mathbf{x} = (A - \lambda I)^{-1} \cdot \mathbf{0} = \mathbf{0}.$$

Der Fall  $\mathbf{x} = \mathbf{0}$  wurde aber in der obigen Definition ausgeschlossen. Ist also  $\lambda$  eine Zahl, so daß die Matrix  $A - \lambda I$  nicht invertierbar ist und ist  $\mathbf{x} \neq \mathbf{0}$  ein Vektor mit der Eigenschaft  $(A - \lambda I)\mathbf{x} = \mathbf{0}$ , dann ist  $\mathbf{x}$  Eigenvektor und  $\lambda$  zugehöriger Eigenwert der Matrix  $A$ . Die Eigenwerte einer Matrix  $A$  sind daher diejenigen Zahlen  $\lambda$  die bewirken, daß die Matrix  $A - \lambda I$  keine Inverse besitzt.

Eine quadratische Matrix ist genau dann nicht invertierbar, wenn ihre Determinante gleich null ist. Im Fall der  $(2 \times 2)$ -Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

ist

$$\begin{aligned} A - \lambda I &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}, \end{aligned}$$

also gilt

$$\det(A - \lambda I) = (a - \lambda)(d - \lambda) - bc = \lambda^2 - (a + d)\lambda + ad - bc.$$

Die Eigenwerte der Matrix  $A$  sind also die Lösungen  $\lambda$  der quadratischen Gleichung

$$(*) \quad \lambda^2 - (a + d)\lambda + ad - bc = 0$$

Der Faktor bei  $\lambda$  heißt die *Spur* der Matrix  $A$ ,  $\text{Spur}(A) = a + d$ . Die Konstante der Gleichung ist die Determinante von  $A$ . Die Gleichung (\*) heißt *charakteristische Gleichung* der Matrix  $A$ :

$$\lambda^2 - \text{Spur}(A)\lambda + \det(A) = 0$$

Ihre Lösungen sind

$$\begin{aligned} \lambda_{1/2} &= \frac{1}{2}(a + d) \pm \sqrt{\frac{1}{4}(a + d)^2 - (ad - bc)} \\ &= \frac{1}{2}(a + d) \pm \frac{1}{2}\sqrt{(a - d)^2 + 4bc}. \end{aligned}$$

Zum Eigenwert  $\lambda_1$  gehört der Eigenvektor  $\mathbf{x}_1 = (x_{11}, x_{12})$ . Er ist Lösung der Gleichung

$$(**) \quad (A - \lambda_1 I)\mathbf{x}_1 = 0.$$

Ausgeschrieben sind das die beiden Gleichungen

$$\begin{aligned} (a - \lambda_1)x_{11} + bx_{12} &= 0, \\ cx_{11} + (d - \lambda_1)x_{12} &= 0. \end{aligned}$$

Da

$$\det(A - \lambda_1 I) = 0$$

ist, sind die Zeilenvektoren  $(a - \lambda_1, b)$  und  $(c, d - \lambda_1)$  Vielfache voneinander, das heißt, es gilt  $(c, d - \lambda_1) = t \cdot (a - \lambda_1, b)$  mit  $t \in \mathbb{R}$ . Es genügt daher, eine der obigen Gleichungen zu betrachten. Eine Lösung der ersten Gleichung ist offenbar

$$\mathbf{x}_1 = (-b, a - \lambda_1).$$

Einen zweiten Eigenvektor  $\mathbf{x}_2$  erhält man auf die gleiche Weise, indem man bei  $(**)$  den Eigenwert  $\lambda_1$  durch  $\lambda_2$  ersetzt, sofern  $\lambda_1$  und  $\lambda_2$  voneinander verschieden sind. Ist dies der Fall, dann gilt:

**Satz:** Sind die Eigenwerte  $\lambda_1, \lambda_2$  der  $(2 \times 2)$ -Matrix  $A$  voneinander verschieden, dann bilden die zugehörigen Eigenvektoren  $\mathbf{x}_1, \mathbf{x}_2$  eine Basis des Vektorraums  $\mathbb{R}^2$ .

**Beweis:** Nach Voraussetzung gilt

$$\begin{aligned} A\mathbf{x}_1 &= \lambda_1 \mathbf{x}_1, \\ A\mathbf{x}_2 &= \lambda_2 \mathbf{x}_2, \quad \lambda_1 \neq \lambda_2. \end{aligned}$$

Man nehme nun an, die Vektoren  $\mathbf{x}_1, \mathbf{x}_2$  bildeten keine Basis des  $\mathbb{R}^2$ . Dann gibt es eine Zahl  $t \neq 0$ , so daß gilt  $\mathbf{x}_2 = t \cdot \mathbf{x}_1$ . Setzt man nun diese Beziehung in die zweite Zeile ein und multipliziert die erste Zeile mit  $t$ , so ergibt sich

$$\begin{aligned} A(t\mathbf{x}_1) &= \lambda_1 t\mathbf{x}_1, \\ A(t\mathbf{x}_1) &= \lambda_2 t\mathbf{x}_1 \end{aligned}$$

und man erhält als Differenz beider Zeilen die Beziehung

$$\mathbf{0} = (\lambda_1 - \lambda_2)t\mathbf{x}_1,$$

die aber im Widerspruch zu den Voraussetzungen  $\lambda_1 \neq \lambda_2$ ,  $t \neq 0$  und  $\mathbf{x}_1 \neq \mathbf{0}$  steht, womit der obige Satz bewiesen ist. – Eine *Anwendung* des Satzes findet sich im folgenden Abschnitt.

**Beispiel:** Die Matrix

$$A = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}$$

von Anfang des Abschnitts hat die charakteristische Gleichung

$$\lambda^2 - (1 - 1)\lambda - 4 = 0$$

also ist  $\lambda_1 = 2$  und  $\lambda_2 = -2$ .

Zu  $\lambda_1 = 2$  gehört der Eigenvektor  $\mathbf{x}_1 = (1, 1)$ . Ein zum Eigenwert  $\lambda_2$  gehöriger Eigenvektor  $\mathbf{x}_2 = (x_{21}, x_{22})$  erfüllt die Gleichung

$$(1 + 2)x_{21} + x_{22} = 0$$

mit der Lösung  $\mathbf{x}_2 = (-1, 3)$ . Die Vektoren  $\mathbf{x}_1, \mathbf{x}_2$  bilden also eine Basis des  $\mathbb{R}^2$ . ■

## 8.5 Das Leslie-Modell

Die Wachstumsdynamik einer Population  $X$  wird, sofern man von Zuwanderungs- und Abwanderungsvorgängen absehen kann, wesentlich durch deren *Altersstruktur* bestimmt. Man unterscheidet daher Altersklassen. Zur Altersklasse  $A_1$  gehören dann alle Individuen mit einem Alter im Zeitintervall  $0 < a \leq 1$  bezüglich einer Zeiteinheit (zum Beispiel ein Jahr). Zur Altersklasse  $A_2$  gehören alle Individuen im Altersintervall  $1 < a \leq 2$  usw.. Wenn  $n$  das maximal erreichbare Alter ist, dann umfaßt die Altersstufe  $A_n$  alle Individuen mit einem Alter  $a$  im Intervall  $n - 1 < a \leq n$ . Die Individuen in der Altersklasse  $A_1$  sind die Neugeborenen. Die Individuen in der Altersgruppe  $A_n$  sind beim Übergang von einem Zeitpunkt  $t$  zum Zeitpunkt  $t + 1$  alle gestorben.

Der *Zustand* der Population  $X$  zu einem Zeitpunkt  $t$  wird daher durch einen Vektor  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  beschrieben, wobei  $x_k$  die Anzahl der Individuen in der Altersklasse  $A_k$  angibt. Zum Zeitpunkt  $t + 1$  sei  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  der Zustand der Population. Wie kommt der Übergang vom Zustandsvektor  $\mathbf{x}$  zum Zustandsvektor  $\mathbf{y}$  nun zustande?

Bringt jedes Individuum der Altersklasse  $A_k$  während einer Zeiteinheit (im Durchschnitt)  $a_k$  neue Individuen hervor, dann ist die Gesamtzahl der im Zeitintervall  $[t, t+1]$  Neugeborenen offenbar

$$y_1 = a_1x_1 + a_2x_2 + \dots + a_nx_n$$

mit Zahlen  $a_k \geq 0$ . Die Zahlen  $a_k$  sind die altersspezifischen Fertilitätsraten (fertilitas = Fruchtbarkeit). Beim Übergang vom Zeitpunkt  $t$  zum Zeitpunkt  $t + 1$  wechselt die Altersgruppe  $A_{k-1}$  in die Altersgruppe  $A_k$ . Dabei wird nur ein Teil der Individuen überleben, so daß gilt,

$$y_k = p_k x_{k-1}, \quad k = 2, 3, \dots, n.$$

Die Überlebensraten  $p_k$  sind Zahlen im Intervall  $0 \leq p_k \leq 1$ . Im Fall  $n = 4$  ergäbe sich auf diese Weise der Zusammenhang

$$y_1 = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4$$

$$y_2 = p_1 x_1$$

$$y_3 = p_2 x_2$$

$$y_4 = p_3 x_3$$

zwischen den Zustandsvektoren  $\mathbf{x}$  und  $\mathbf{y}$  zu den Zeitpunkten  $t$  und  $t + 1$ . In Matrizen-Schreibweise ist das die Beziehung

$$\mathbf{y} = A\mathbf{x}$$

mit der Matrix

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 \\ 0 & 0 & p_3 & 0 \end{pmatrix}$$

und den Spaltenvektoren  $\mathbf{x}$  und  $\mathbf{y}$ . (Das hier beschriebene Populationsmodell geht vor allem auf P.H. Leslie zurück und wird daher nach ihm benannt. (Vgl.: Leslie, P.H.: On the use of matrices in certain population mathematics. Biometrika 33, 183–212, 1945.)

Man interessiert sich nun vor allem für die Frage, wohin die Entwicklung einer Population, deren Anfangszustand  $\mathbf{x}_0$  und deren Leslie-Matrix  $A$  bekannt sind, langfristig tendiert. Die sukzessiven Zustandsvektoren zu den Zeiten  $t + 1, t + 2, t + 3, \dots, t + n$ , sind offenbar

$$\mathbf{x}_1 = A\mathbf{x}_0,$$

$$\mathbf{x}_2 = A(A\mathbf{x}_0) = A^2\mathbf{x}_0,$$

$$\mathbf{x}_3 = A(A^2\mathbf{x}_0) = A^3\mathbf{x}_0,$$

und schließlich

$$(*) \quad \mathbf{x}_n = A^n \mathbf{x}_0.$$

Die Potenzen  $A^n$ ,  $n = 1, 2, 3, \dots$ , der Matrix  $A$  regieren also die Dynamik der Leslie-Population. Deren direkte Berechnung für größere  $n$  ist numerisch sehr aufwendig und wenig zielführend. Wie man stattdessen vorgeht, sei am Beispiel einer Leslie-Population mit zwei Altersgruppen exemplarisch erläutert. In diesem Falle ist

$$A = \begin{pmatrix} a & b \\ p & 0 \end{pmatrix}$$

mit  $a, b \geq 0$  und  $0 < p \leq 1$  die Leslie-Matrix. (Die Fälle  $a = b = 0$  und  $p = 0$  müssen dabei sinnvollerweise ausgeschlossen werden.) Ihr charakteristisches Polynom (vgl. Abschnitt 8.3) ist

$$\lambda^2 - a\lambda - bp = 0.$$

Also sind

$$\lambda_{1/2} = \frac{1}{2}(a \pm \sqrt{a^2 + 4bp})$$

die Eigenwerte der Matrix  $A$ . Der Fall  $\lambda_1 = \lambda_2$  tritt ein, wenn  $a^2 + 4bp = 0$  ist. Das aber bedeutet entweder  $a = b = p = 0$  oder  $a = 0$ ,  $b = 0$ , oder  $a = 0$ ,  $p = 0$ . Alle drei Fälle beschreiben keine „interessanten“ Populationen und können daher ausgeschlossen werden. Es soll im folgenden  $a, b, p \neq 0$  gelten.

Also darf man  $\lambda_1 \neq \lambda_2$  annehmen, so daß zugehörige Eigenvektoren  $e_1$  und  $e_2$  nach dem im Abschnitt 8.3. bewiesenen Satz eine Basis des  $\mathbb{R}^2$  bilden.

Ein Anfangszustand  $x_0$  der Population  $X$  kann daher als Linearkombination

$$x_0 = \alpha_1 e_1 + \alpha_2 e_2$$

mit Koeffizienten  $\alpha_1, \alpha_2 \in \mathbb{R}$  geschrieben werden. Dann gilt wegen (\*):

$$\begin{aligned} x_n &= A^n x_0 = A^n(\alpha_1 e_1 + \alpha_2 e_2) \\ &= A^{n-1}(\alpha_1 A e_1 + \alpha_2 A e_2) \\ &= A^{n-1}(\alpha_1 \lambda_1 e_1 + \alpha_2 \lambda_2 e_2) \\ &= A^{n-2}(\alpha_1 \lambda_1 A e_1 + \alpha_2 \lambda_2 A e_2) \\ &= A^{n-2}(\alpha_1 \lambda_1^2 e_1 + \alpha_2 \lambda_2^2 e_2). \end{aligned}$$

Man sieht, daß der Exponent von  $A$  sich bei jedem Schritt um eins abbaut, während die Exponenten von  $\lambda_1$  und  $\lambda_2$  jeweils um eins zunehmen. Auf diese Weise erhält man schließlich die Beziehung

$$(*) \quad x_n = \alpha_1 \lambda_1^n e_1 + \alpha_2 \lambda_2^n e_2.$$

Sie besagt: Die Entwicklungsdynamik der Population  $X$  wird vollständig bestimmt vom Anfangszustand  $x_0$ , den Eigenwerten  $\lambda_1, \lambda_2$  und den zugehörigen Eigenvektoren

$e_1, e_2$  der Lesliematrix  $A$ .

Die Eigenwerte

$$\lambda_1 = \frac{1}{2}(a + \sqrt{a^2 + 4bp}),$$

(\*\*)

$$\lambda_2 = \frac{1}{2}(a - \sqrt{a^2 + 4bp})$$

sind reell, da der Ausdruck unter der Wurzel positiv ist. Es gilt immer  $\lambda_1 \geq a$  und für den zweiten Eigenwert  $\lambda_2$  gilt offenbar,  $\lambda_2 \leq 0$  und wegen  $\lambda_1 + \lambda_2 = a$  ergibt sich  $\lambda_2$  unmittelbar, wenn  $\lambda_1$  bekannt ist.

Für  $\lambda_1$  sind die drei Fälle  $\lambda_1 > 1$ ,  $\lambda_1 = 1$  und  $0 < \lambda_1 < 1$  möglich. Im ersten Fall gilt  $\lambda_1^n \rightarrow \infty$  für  $n \rightarrow \infty$ , so daß die Population  $X$  wegen (\*) beliebig groß wird. Für  $\lambda_2$  erhält man  $\lambda_2 < 0$  und  $|\lambda_2| < \lambda_1$ .

Die Ungleichung  $\lambda_1 > 1$  besagt

$$\frac{1}{2}(a + \sqrt{a^2 + 4bp}) > 1$$

oder

$$\sqrt{a^2 + 4bp} > 2 - a.$$

Ist  $a \geq 2$ , dann gilt  $\lambda_1 > 1$ . Also bleibt der Fall  $0 \leq a < 2$  zu betrachten. Quadriert man nun auf beiden Seiten so folgt

$$a^2 + 4bp > 4 - 4a + a^2,$$

also

$$a + bp > 1.$$

Im Falle  $\lambda_1 = 1$  erhält man die Bedingung  $a + bp = 1$ . Also gilt  $0 < a < 1$  und wegen  $\lambda_2 = a - 1$  folgt  $-1 < \lambda_2 < 0$ . Die Beziehung (\*) lautet in diesem Falle

$$\mathbf{x}_n = \alpha_1 \mathbf{e}_1 + \lambda_2^n \alpha_2 \mathbf{e}_2.$$

Wegen  $\lambda_2^n \rightarrow 0$  für  $n \rightarrow \infty$  hat man daher  $\mathbf{x}_n \rightarrow \alpha_1 \mathbf{e}_1$  für  $n \rightarrow \infty$ . Ist  $0 < \lambda_1 < 1$  dann gilt  $a + bp < 1$ , also  $0 < a < 1$  und es folgt wegen  $\lambda_2 = a - \lambda_1$  wiederum  $-1 < \lambda_2 < 0$ . Also gilt wegen (\*)  $\mathbf{x}_n \rightarrow \mathbf{0}$  für  $n \rightarrow \infty$ . Das bedeutet, die Population  $X$  stirbt aus. Wir erhalten daher den

**Satz:** Die Entwicklung einer Leslie-Population mit zwei Altersgruppen, den Fertilitätsraten  $a, b$  und der Übergangsrate  $p$  mit  $a, b, p \neq 0$  wird durch die Kenngröße  $K = a + bp$  bestimmt. Ist  $K > 1$ , dann wächst die Population unbeschränkt, ist

$K = 1$  dann bleibt die Populationsgröße endlich und im Fall  $K < 1$  stirbt die Population aus.

Eine Population mit der maximalen Übergangsrate  $p = 1$  und der Bedingung  $a + b < 1$  für die Summe der Fertilitätsraten der beiden Altersgruppen stirbt also gleichwohl aus.

**Beispiel (Entwicklung einer Leslie-Population):** Eine Population mit zwei Altersgruppen und den Fertilitätsraten  $a = 1$ ,  $b = 2$  sowie der Übergangsrate  $p = \frac{1}{2}$  hat die Leslie-Matrix

$$A = \begin{pmatrix} 1 & 2 \\ \frac{1}{2} & 0 \end{pmatrix}.$$

Ihre Kenngröße ist  $K = 1 + 2 \cdot \frac{1}{2} > 1$ . Die Population wächst also unbeschränkt.

Die charakteristische Gleichung lautet

$$\lambda^2 - \lambda - 1 = 0.$$

Ihre Lösungen sind die Eigenwerte.

$$\begin{aligned} \lambda_1 &= \frac{1}{2} (1 + \sqrt{5}), \\ \lambda_2 &= \frac{1}{2} (1 - \sqrt{5}). \end{aligned}$$

Diese Werte sind nun bereits bekannt. Es ist  $\lambda_1 = \Phi$  das Verhältnis des Goldenen Schnitts und es gilt  $\lambda_2 = -1/\Phi$  (vgl. Abschnitt 4.3).

Die zum Eigenwert  $\lambda_1$  gehörigen Eigenvektoren  $e_1 = (x_1, x_2)$  erfüllen die Gleichung

$$\frac{1}{2}(1 - \sqrt{5}) x_1 + 2 x_2 = 0.$$

Die Lösung

$$e_1 = \left( 2, \frac{1}{2} (\sqrt{5} - 1) \right)$$

lässt sich unmittelbar ablesen. Einen zum Eigenwert  $\lambda_2$  gehörigen Eigenvektor erhält man indem man bei  $e_1$  die  $\sqrt{5}$  durch  $-\sqrt{5}$  ersetzt und den so erhaltenen Vektor mit  $(-1)$  multipliziert. (Dadurch erreicht man, daß der Eigenvektor  $e_2$  in den 2. Quadranten des  $\mathbb{R}^2$  zeigt.) Man erhält also

$$e_2 = \left( -2, \frac{1}{2} (\sqrt{5} + 1) \right).$$

Für die Eigenwerte  $\lambda_1, \lambda_2$  gilt  $\lambda_1 > 1$  und  $-1 < \lambda_2 < 0$ . Wir untersuchen die Entwicklung einer Population mit dem Anfangszustand  $\mathbf{x}_0 = (1, 1)$ . Das soll bedeuten, in jeder Altersgruppe befinden sich anfänglich  $1 \times 1000$  Individuen.

Zunächst haben wir den Vektor  $\mathbf{x}_0$  als Linearkombination der Eigenvektoren  $\mathbf{e}_1$  und  $\mathbf{e}_2$  darzustellen, das heißt, Zahlen  $\alpha_1$  und  $\alpha_2$  zu bestimmen, so daß gilt  $\mathbf{x}_0 = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2$ . Im vorliegenden Fall muß also gelten

$$\alpha_1 \left( 2, \frac{1}{2}(\sqrt{5} - 1) \right) + \alpha_2 \left( -2, \frac{1}{2}(\sqrt{5} + 1) \right) = (1, 1)$$

Multipliziert man diese Gleichung skalar mit dem Vektor  $(\frac{1}{2}(\sqrt{5} + 1), 2)$  dann erhält man

$$\alpha_1 (2\sqrt{5}) = \frac{1}{2}\sqrt{5} + \frac{5}{2},$$

also

$$\alpha_1 = \frac{1}{4}(\sqrt{5} + 1).$$

Multipliziert man skalar mit dem Vektor  $(\frac{1}{2}(\sqrt{5} - 1), -2)$ , dann folgt

$$\alpha_2(-2\sqrt{5}) = \frac{1}{2}\sqrt{5} - \frac{5}{2},$$

also ist

$$\alpha_2 = \frac{1}{4}(\sqrt{5} - 1).$$

Wegen  $-1 < \lambda_2 < 0$  und damit  $\lambda_2^n \rightarrow 0$  für  $n \rightarrow \infty$  dominiert der zum Eigenvektor  $\mathbf{e}_1$  gehörige Anteil in der Entwicklungsgleichung

$$\mathbf{x}_n = \alpha_1 \lambda_1^n \mathbf{e}_1 + \alpha_2 \lambda_2^n \mathbf{e}_2.$$

Für große  $n$  gilt also

$$\mathbf{x}_n \approx \alpha_1 \lambda_1^n \mathbf{e}_1.$$

(Das Zeichen  $\approx$  bedeutet „ist ungefähr gleich“.) Die Population entwickelt sich gleichsam in Richtung des Eigenvektors  $\mathbf{e}_1$  und wird dabei wegen  $\lambda_1^n \rightarrow \infty$  für  $n \rightarrow \infty$  beliebig groß, (was natürlich unrealistisch ist.) Das Wachstum in Richtung des Eigenvektors  $\mathbf{e}_1$  bedeutet: Das Verhältnis  $x_{n1} : x_{n2}$  der beiden Altersgruppen strebt gegen den Wert  $e_{11} : e_{12}$ . Im vorliegenden Fall gilt also

$$\frac{x_{n1}}{x_{n2}} \rightarrow \frac{2}{\frac{1}{2}(\sqrt{5} - 1)}$$

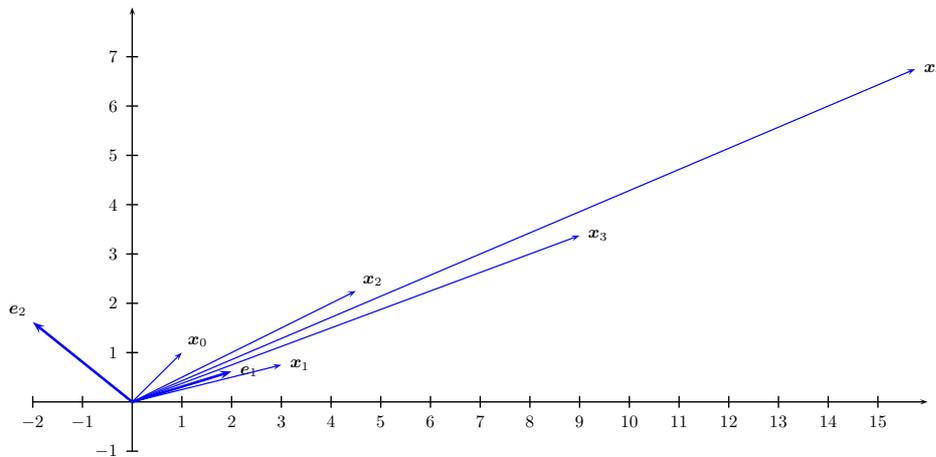


Abbildung 8.16: Die zeitliche Entwicklung einer Leslie–Population

für  $n \rightarrow \infty$ .

Die rechte Seite ist gleich

$$\frac{4}{\sqrt{5}-1} = \frac{4(\sqrt{5}+1)}{4} = \sqrt{5}+1 = 3,236.$$

Das ursprüngliche Verhältnis 1 : 1 der Altersgruppen verschiebt sich also zugunsten der jüngeren Altersgruppe auf den Wert 3,236 : 1. ■

## 8.6 Lineare Gleichungssysteme

Ein lineares Gleichungssystem (LGS) hat die Form

$$(LGS) \quad Ax = b.$$

Hierbei ist  $A$  eine Matrix mit  $m$  Zeilen und  $n$  Spalten. Die Matrix  $A$  ist also im allgemeinen nicht quadratisch. Der Vektor  $x = (x_1, x_2, \dots, x_n)$  faßt die zu bestimmenden Unbekannten  $x_1, x_2, \dots, x_n$  zusammen. Er wird in der Gleichung als Spaltenvektor geschrieben. Der Vektor  $b = (b_1, b_2, \dots, b_m)$  auf der rechten Seite mit  $m$  Komponenten ist ebenfalls als Spaltenvektor geschrieben.

Die Aufgabenstellung des LGS ist also so zu verstehen: Gesucht sind alle Vektoren  $x \in \mathbb{R}^n$ , die vermittels der  $(m, n)$ -Matrix  $A$  auf den Vektor  $b$  abgebildet werden.

Ist  $A$  eine invertierbare  $(n, n)$ -Matrix, dann ist der Vektor

$$x = A^{-1}b$$

die eindeutig bestimmte Lösung des linearen Gleichungssystems.

### Die Struktur der Lösungsmenge

Bevor wir uns mit der Lösungstechnik für lineare Gleichungssysteme beschäftigen, untersuchen wir die Struktur der *Lösungsmenge*

$$L = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$$

des Systems (LGS).

Ist  $\mathbf{b} \neq \mathbf{0}$ , dann wird das Gleichungssystem

$$(i) \quad A\mathbf{x} = \mathbf{b}$$

als *inhomogenes* System bezeichnet. Das zugeordnete *homogene* System ist das System

$$(h) \quad A\mathbf{x} = \mathbf{0},$$

bei dem die rechte Seite von (i) durch den Nullvektor ersetzt wird.

Es sei nun  $\mathbf{x}^*$  eine *spezielle* (als fest gedachte) Lösung des inhomogenen Systems (i), das heißt es gelte

$$A\mathbf{x}^* = \mathbf{b},$$

und  $\mathbf{x}$  sei *irgendeine* Lösung von (i), das heißt, es gelte

$$A\mathbf{x} = \mathbf{b}.$$

Zieht man die obere Gleichung von der unteren ab, dann erhält man die Beziehung

$$A\mathbf{x} - A\mathbf{x}^* = A(\mathbf{x} - \mathbf{x}^*) = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

Sie besagt: Der Vektor

$$\mathbf{x} - \mathbf{x}^* = \mathbf{h}$$

ist eine Lösung des homogenen Systems (h). Also gilt

$$\mathbf{x} = \mathbf{x}^* + \mathbf{h}.$$

Damit ist die Struktur der Lösungsgesamtheit des Systems (i) beschrieben.

**Satz:** Jede Lösung  $\mathbf{x}$  des linearen Gleichungssystems

$$(i) \quad A\mathbf{x} = \mathbf{b}$$

mit  $\mathbf{b} \neq \mathbf{0}$  kann in der Form

$$\mathbf{x} = \mathbf{x}^* + \mathbf{h}$$

geschrieben werden. Dabei ist  $\mathbf{x}^*$  eine spezielle Lösung des inhomogenen Systems (i) und  $\mathbf{h}$  ist eine Lösung des zu (i) gehörigen homogenen Systems

$$(h) \quad A\mathbf{x} = \mathbf{0}.$$

Zur Bestimmung der Lösungsgesamtheit eines linearen Gleichungssystems geht man daher in zwei Schritten vor:

- I. Aufsuchen einer speziellen Lösung des inhomogenen Systems (i).
- II. Beschreibung der Lösungsgesamtheit des homogenen Systems (h).

Die Lösungsgesamtheit  $L_h$  des homogenen Systems hat zwei wichtige Eigenschaften:

- 1) Ist  $\mathbf{h} \in L_h$ , dann ist auch der Vektor  $\alpha\mathbf{h} \in L_h$ , wenn  $\alpha \in \mathbb{R}$  ist.
- 2) Sind  $\mathbf{h}, \mathbf{h}' \in L_h$  zwei Lösungsvektoren des homogenen Systems, dann gilt  $\mathbf{h} + \mathbf{h}' \in L_h$ . Das heißt die Summe zweier Lösungen des homogenen Systems ist ebenfalls eine Lösung des homogenen Systems.

Beide Eigenschaften lassen sich folgendermaßen zusammenfassen: Sind  $\mathbf{h}, \mathbf{h}' \in L_h$  zwei Lösungen des homogenen Systems, dann ist auch jede Linearkombination  $\mathbf{x} = \alpha\mathbf{h} + \alpha'\mathbf{h}'$  eine Lösung des homogenen Systems. Es gilt nämlich

$$\begin{aligned} A\mathbf{x} &= A(\alpha\mathbf{h} + \alpha'\mathbf{h}') \\ &= A(\alpha\mathbf{h}) + A(\alpha'\mathbf{h}') \\ &= \alpha A\mathbf{h} + \alpha' A\mathbf{h}' = \mathbf{0} + \mathbf{0} = \mathbf{0}. \end{aligned}$$

Das bedeutet, die Lösungsgesamtheit  $L_h$  bildet einen *linearen Raum (Vektorraum)*. Trivialerweise ist der Nullvektor  $\mathbf{0}$  immer eine Lösung des homogenen Systems (h), denn es gilt  $A\mathbf{0} = \mathbf{0}$ .

Zur Beschreibung der Lösungsgesamtheit  $L_h$  von (h) wird man also Lösungen  $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_m$  suchen, so dass sich jede Lösung  $\mathbf{h} \in L_h$  auf genau eine Weise als Linearkombination

$$\mathbf{h} = \alpha_1\mathbf{h}_1 + \alpha_2\mathbf{h}_2 + \dots + \alpha_m\mathbf{h}_m$$

dieser Lösungen schreiben lässt. Die Vektoren  $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_m$  bilden also eine Basis des Lösungsraums  $L_h$  des homogenen Gleichungssystems.

Es folgen zunächst zwei Beispiele linearer Gleichungssysteme.

**Beispiel 1 (Eine Ebenengleichung):** Die Gleichung

$$(E) \quad x_1 + 2x_2 - x_3 = 4$$

ist die Gleichung einer Ebene  $E$  (vgl. Abschnitt 8.2) und zugleich ein lineares Gleichungssystem mit drei Unbekannten  $x_1, x_2, x_3$  und einer Gleichung. Jeder Lösungsvektor  $\mathbf{x} = (x_1, x_2, x_3)$  ist der Ortsvektor eines Punktes  $X$  auf  $E$ . Um eine systematische Beschreibung dieser Ortsvektoren zu gewinnen, geht man nach den Schritten I und II vor. Eine spezielle Lösung  $x^*$  von (E), findet man zum Beispiel, indem man  $x_2 = x_3 = 0$  setzt. Dann folgt  $x_1 = 4$ . Also ist

$$x^* = (4, 0, 0)$$

eine spezielle Lösung von (E). Die zu (E) gehörige homogene Gleichung (E') lautet

$$(E') \quad x_1 + 2x_2 - x_3 = 0.$$

Dies ist die Gleichung einer Ebene  $E'$  parallel zu  $E$ . Die Ebenen  $E$  und  $E'$  besitzen nämlich den gleichen Normalenvektor  $\mathbf{p} = (1, 2, -1)$ . Die Ebene  $E'$  geht, im Unterschied zu  $E$ , durch den Nullpunkt des  $\mathbb{R}^3$ .

Um eine spezielle Lösung  $\mathbf{h}_1$  der Gleichung (E') zu erhalten, setzt man  $x_2 = 1$  und  $x_3 = 0$ , so dass  $x_1 = -2$  wird. Also ist

$$\mathbf{h}_1 = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}.$$

Eine zweite Lösung  $\mathbf{h}_2$  von (E') findet man, indem man  $x_2 = 0$  und  $x_3 = 1$  setzt. Dann ist  $x_1 = 1$  und

$$\mathbf{h}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Jede Lösung

$$\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} \in L_h$$

kann aber als Linearkombination

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = \alpha_1 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

der Lösungen  $\mathbf{h}_1$  und  $\mathbf{h}_2$  geschrieben werden. Durch Vergleich der 2. und 3. Vektor­komponenten bei dieser Beziehung folgt nämlich  $\alpha_1 = h_2$  und  $\alpha_2 = h_3$  und es gilt für die erste Komponente der rechten Seite  $-2\alpha_1 + \alpha_2 = -2h_2 + h_3 = h_1$ , da  $\mathbf{h}$  eine Lösung von (E') ist. Die Vektoren  $\mathbf{h}_1$  und  $\mathbf{h}_2$  bilden also eine Basis des Lösungsraumes von (E'). ■

**Beispiel 2 (Das Vektorprodukt):** Das Skalarprodukt  $\mathbf{a} \cdot \mathbf{b}$  zweier Vektoren  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  ist eine Zahl. Für Vektoren  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$  führt man nun eine Verknüpfung  $\mathbf{a} \times \mathbf{b}$  ein, die wiederum ein Vektor ist. Dieser Vektor soll senkrecht auf den Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  stehen. Gegeben seien also die Vektoren  $\mathbf{a} = (a_1, a_2, a_3)$  und  $\mathbf{b} = (b_1, b_2, b_3)$  und es werde vorausgesetzt, dass  $\mathbf{a}$  und  $\mathbf{b}$  nicht dieselbe Richtung besitzen, dass also  $\mathbf{a} \neq t\mathbf{b}$  ist für  $t \in \mathbb{R}$ .

Gesucht ist also ein Vektor  $\mathbf{x} = (x_1, x_2, x_3)$ , der senkrecht auf  $\mathbf{a}$  und senkrecht auf  $\mathbf{b}$  steht. Es muss daher gelten

$$\begin{aligned}\mathbf{a} \cdot \mathbf{x} &= 0, \\ \mathbf{b} \cdot \mathbf{x} &= 0.\end{aligned}$$

Ausgeschrieben sind das die Gleichungen

$$\begin{aligned}a_1 x_1 + a_2 x_2 + a_3 x_3 &= 0, \\ b_1 x_1 + b_2 x_2 + b_3 x_3 &= 0.\end{aligned}$$

Setzt man  $x_3 = 1$ , dann bleiben zwei Gleichungen für die beiden Unbekannten  $x_1$  und  $x_2$ :

$$\begin{array}{rcl} a_1 x_1 + a_2 x_2 & = & -a_3 \left| \begin{array}{c} b_2 \\ a_2 \end{array} \right| b_1 \\ b_1 x_1 + b_2 x_2 & = & -b_3 \left| \begin{array}{c} b_2 \\ a_2 \end{array} \right| a_1. \end{array}$$

Multipliziert man die erste Zeile mit  $b_2$ , die zweite mit  $a_2$  und subtrahiert die Zeilen voneinander, dann heben sich die  $x_2$ -Terme weg und man erhält

$$(a_1 b_2 - a_2 b_1) x_1 = a_2 b_3 - a_3 b_2$$

also

$$x_1 = \frac{a_2 b_3 - a_3 b_2}{a_1 b_2 - a_2 b_1}.$$

Entsprechend hebt sich nach Multiplikation der ersten Zeile mit  $b_1$ , der zweiten mit  $a_1$  und Subtraktion der Zeilen der  $x_1$ -Term weg. Es folgt

$$x_2 = \frac{a_1 b_3 - a_3 b_1}{a_2 b_1 - a_1 b_2}.$$

Der Vektor  $(x_1, x_2, 1)$  steht also senkrecht auf den Vektoren  $\mathbf{a}$  und  $\mathbf{b}$ . Dies gilt aber auch für jeden Vektor  $t \cdot (x_1, x_2, 1)$ ,  $t \in \mathbb{R}$ . Setzt man  $t = a_1 b_2 - a_2 b_1$ , dann erhält man einen Vektor, der mit  $\mathbf{a} \times \mathbf{b}$  bezeichnet wird. Seine Komponenten lassen sich als zweireihige Determinanten auffassen. Damit kann man schreiben:

$$\mathbf{a} \times \mathbf{b} = \left( \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}, \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \right)$$

Der Vektor  $\mathbf{a} \times \mathbf{b}$  wird als das *Vektorprodukt* der Vektoren  $\mathbf{a}$  und  $\mathbf{b}$  bezeichnet. Die Bildung des Vektorproduktes kann man sich leicht mit Hilfe der Matrix

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}$$

merken. Die erste Komponente des Vektors  $\mathbf{a} \times \mathbf{b}$  erhält man, indem man die erste Spalte der Matrix streicht und die Determinante des verbleibenden Teils bildet. Dann streicht man die mittlere Spalte, bildet die Determinante und multipliziert mit minus eins. Schließlich streicht man die letzte Spalte und bildet die Determinante des restlichen Teils.

Offenbar gelten die Regeln

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$$

und

$$(t\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (t\mathbf{b}) = t(\mathbf{a} \times \mathbf{b})$$

mit  $t \in \mathbb{R}$ .

Mit den Basisvektoren  $\mathbf{e}_1 = (1, 0, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0)$  und  $\mathbf{e}_3 = (0, 0, 1)$  des  $\mathbb{R}^3$  erhält man  $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$ . Bei beliebigen Vektoren  $\mathbf{a}, \mathbf{b}, \mathbf{b} \neq t\mathbf{a}$  des  $\mathbb{R}^3$ , haben die drei Vektoren  $\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}$  immer dieselbe Anordnung zueinander im Raum wie die Basisvektoren  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ . Sie bilden ebenfalls eine Basis des  $\mathbb{R}^3$ . ■

## Der Gauß-Algorithmus an Beispielen

Der Gauß-Algorithmus ist das rechnerisch ökonomischste Verfahren zur Lösung linearer Gleichungssysteme. Der Gauß-Algorithmus wird hier *nicht nur zur Berechnung* der Lösungsmenge, sondern auch zu einer nochmaligen *Analyse der Lösungsstruktur eines linearen Gleichungssystems* benutzt.

Obwohl beim Gauß-Algorithmus nur „elementare“ Operationen vorkommen, wäre es mühsam, eine vollständige allgemeine Beschreibung des Algorithmus an den Anfang zu stellen. Dies ist aber auch nicht notwendig. Man erlernt den Gauß-Algorithmus am

besten anhand von Beispielen. Die verschiedenartige Herkunft der behandelten Probleme zeigt zugleich die vielfältigen Anwendungsmöglichkeiten der linearen Algebra.

### Äquivalenzumformungen

**Beispiel 3:** Gesucht ist ein Polynom höchstens dritten Grades, dessen Graph durch die Punkte  $(-1; -10)$ ,  $(1; -2)$ ,  $(2; 2)$  und  $(3; 14)$  geht.

Das Polynom

$$p(x) = a_0 + a_1x + a_2x^2 + a_3x^3$$

erfüllt diese Forderung genau dann, wenn für die Koeffizienten  $a_0, a_1, a_2, a_3$  die folgenden Bedingungen gelten:

$$(1) \quad \begin{array}{rccccrcr} a_0 & - & a_1 & + & a_2 & - & a_3 & = & -10 \\ a_0 & + & a_1 & + & a_2 & + & a_3 & = & -2 \\ a_0 & + & 2a_1 & + & 4a_2 & + & 8a_3 & = & 2 \\ a_0 & + & 3a_1 & + & 9a_2 & + & 27a_3 & = & 14 \end{array}$$

Dies ist ein *lineares Gleichungssystem* (LGS) für die *Unbekannten*  $a_0, a_1, a_2, a_3$ . Die Aufgabe besteht nun darin, Zahlen  $a_0, a_1, a_2, a_3$  so zu bestimmen, dass alle Gleichungen des linearen Gleichungssystems erfüllt sind.

Jedes Quadrupel  $(a_0, a_1, a_2, a_3)$  für das die vier Gleichungen (1) erfüllt sind, ist *Lösung* des linearen Gleichungssystems.

Die Gesamtheit aller Lösungen ist die *Lösungsmenge*. Es ist nicht auszuschließen, dass die Lösungsmenge leer sein kann; in diesem Fall sagt man: Das lineare Gleichungssystem ist *unlösbar*.

Aufgrund der Kenntnis von Lösungsverfahren für lineare Gleichungssysteme mit zwei Gleichungen und zwei Unbekannten wird man durchaus in der Lage sein, das vorgelegte lineare Gleichungssystem zu lösen. Dies kann durch sukzessive Elimination der Unbekannten geschehen. Dabei ist es aber leicht möglich, dass man die erforderlichen Rechnungen unübersichtlich anlegt, den Überblick verliert und dann stecken bleibt oder zu falschen Ergebnissen gelangt.

Es ist daher notwendig, den Eliminationsvorgang zweckmäßig zu organisieren. Das leistet der Gauß-Algorithmus. Er besteht aus einer Folge von Umformungen des linearen Gleichungssystems, bei denen sich dessen Lösungsmenge nicht ändert. Von dieser Art sind zum Beispiel die beiden folgenden Umformungen, die vielfach als *elementare Umformungen* bezeichnet werden.

1. *Multiplikation einer Gleichung mit einer von Null verschiedenen Zahl.*

Die Gleichung

$$(*) \quad a_1x_1 + a_2x_2 + \dots + a_nx_n = c$$

hat dieselbe Lösungsmenge wie die Gleichung

$$(**) \quad \lambda a_1 x_1 + \lambda a_2 x_2 + \dots + \lambda a_n x_n = \lambda c,$$

wenn  $\lambda \neq 0$  gilt.

Ersetzt man also in einem linearen Gleichungssystem die Gleichung (\*) durch (\*\*) mit  $\lambda \neq 0$ , so bleibt die Lösungsmenge unverändert.

2. Ersetzen einer Gleichung durch die Summe aus dieser Gleichung und einer anderen.

Das LGS mit den beiden Gleichungen

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = c_1$$

(\*)

$$b_1 x_1 + b_2 x_2 + \dots + b_n x_n = c_2$$

hat dieselbe Lösungsmenge wie das LGS

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = c_1$$

(\*\*)

$$(a_1 + b_1)x_1 + (a_2 + b_2)x_2 + \dots + (a_n + b_n)x_n = c_1 + c_2.$$

treten also die beiden Gleichungen (\*) in einem linearen Gleichungssystem auf, so kann man sie durch die beiden Gleichungen (\*\*) ersetzen, ohne damit die Lösungsmenge des linearen Gleichungssystems zu verändern.

Zum Beweis, dass diese beiden elementaren Umformungen *Äquivalenzumformungen* sind, das heißt die Lösungsmenge des linearen Gleichungssystems nicht verändern, ist jeweils die Übereinstimmung zweier Mengen zu zeigen: Es ist der Nachweis zu führen, dass jedes Element der einen zur anderen gehört und umgekehrt.

Durch Kombination der beiden elementaren Umformungen hat man auch die Möglichkeit, eine Gleichung durch die Summe aus ihr und einem beliebigen Vielfachen einer anderen Gleichung zu ersetzen. Von dieser Äquivalenzumformung wird im folgenden häufig Gebrauch gemacht.

Auf das lineare Gleichungssystem (1) wenden wir jetzt die Äquivalenzumformungen 1. und 2. an. Im *ersten* Schritt wollen wir die Unbekannte  $a_0$  mit Hilfe der ersten Gleichung aus allen übrigen Gleichungen eliminieren indem wir nacheinander die 2., 3. und 4. Gleichung mit  $(-1)$  multiplizieren und zur 1. Gleichung addieren. Die erste Gleichung bleibt unverändert erhalten:

$$\begin{array}{rcccc} a_0 - & a_1 + & a_2 - & a_3 & = & -10 & \left| \begin{array}{c} 1 \\ \\ \\ \end{array} \right| & \left| \begin{array}{c} 1 \\ \\ \\ \end{array} \right| & \left| \begin{array}{c} 1 \\ \\ \\ \end{array} \right| \\ a_0 + & a_1 + & a_2 + & a_3 & = & -2 & \left( -1 \right) & & \\ a_0 - & 2a_1 + & 4a_2 - & 8a_3 & = & 2 & & \left( -1 \right) & \\ a_0 - & 3a_1 + & 9a_2 - & 27a_3 & = & 14 & & & \left( -1 \right) \end{array}$$

Das lineare Gleichungssystem gewinnt die Gestalt

$$\begin{array}{cccc|c|c} a_0 - & a_1 + & a_2 - & a_3 & = & -10 & & \\ & 2a_1 & & + 2a_3 & = & 8 & (-\frac{3}{2}) & (-2) \\ & 3a_1 + & 3a_2 + & 9a_3 & = & 12 & 1 & \\ & 4a_1 + & 8a_2 + & 28a_3 & = & 24 & & 1 \end{array}$$

Im *zweiten* Schritt wird  $a_1$  mit Hilfe der zweiten Gleichung aus der dritten und vierten eliminiert.

$$\begin{array}{cccc|c|c} a_0 - & a_1 + & a_2 - & a_3 & = & -10 & & \\ & a_1 & & + a_3 & = & 4 & & \\ & & 3a_2 + & 6a_3 & = & 0 & (-\frac{8}{3}) & \\ & & 8a_2 + & 24a_3 & = & 8 & & 1 \end{array}$$

Entsprechend bewirkt der *dritte* Schritt die Elimination von  $a_2$  aus der letzten Gleichung:

$$\begin{array}{cccc|c} a_0 - & a_1 + & a_2 - & a_3 & = & -10 \\ & a_1 & & + a_3 & = & 4 \\ & & a_2 + & 2a_3 & = & 0 \\ & & & 8a_3 & = & 8 \end{array}$$

Mit dieser „Dreiecksgestalt“ des linearen Gleichungssystems ist das erste Ziel des Gauß-Algorithmus erreicht. Gleichungssysteme von dieser Gestalt lassen sich, von der letzten Gleichung ausgehend, durch Einsetzen der bereits gewonnenen Teilergebnisse in die jeweils vorausgehende Gleichung lösen.

Man erhält der Reihe nach

$$a_3 = 1, \quad a_2 = -2, \quad a_1 = 3, \quad a_0 = -4.$$

Das Quadrupel  $(-4, 3, -2, 1)$  ist also einzige Lösung des zuletzt erhaltenen linearen Gleichungssystems. Da nur Äquivalenzumformungen vorgenommen wurden, ist zugleich sicher, dass das ursprüngliche lineare Gleichungssystem ebenfalls dieses Quadrupel als einzige Lösung besitzt.

### Ein lineares Gleichungssystem mit mehr als einer Lösung

An dem folgenden Beispiel kann man die Einsicht gewinnen, dass ein lineares Gleichungssystem *mehrere* Lösungen haben kann, auch wenn die Anzahl der Gleichungen mit der Anzahl der Unbekannten übereinstimmt. Außerdem werden vereinfachende Schreibweisen eingeführt.

**Beispiel 4 (Eine Mischungsaufgabe):** Zur Herstellung besonders fester Leichtmetallrohre wird eine Legierung  $L$  aus Aluminium (91%), Kupfer (4%), Zink (3%) und

Magnesium (2%) verwendet. Sie soll aus den Legierungen  $L_1, L_2, L_3, L_4, L_5$  zusammengeschmolzen werden, deren Zusammensetzung bekannt ist.

	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$	$L$
% Kupfer	5	3	3	6	7	4
% Zink	3	4	2	4	3	3
% Magnesium	2	1	3	1	2	2
% Aluminium	90	92	92	89	88	91

Es soll eine Mengeneinheit der Legierung  $L$  hergestellt werden. Mit  $x_i$  bezeichnen wir den jeweils benötigten Bruchteil einer Mengeneinheit der Legierung  $L_i$  ( $i = 1, 2, \dots, 5$ ), der nun zu bestimmen ist. Die Summe der fünf Bruchteile ergibt natürlich 1. Man erhält das LGS:

$$\begin{aligned}
 x_1 + x_2 + x_3 + x_4 + x_5 &= 1 \\
 5x_1 + 3x_2 + 3x_3 + 6x_4 + 7x_5 &= 4 \\
 3x_1 + 4x_2 + 2x_3 + 4x_4 + 3x_5 &= 3 \\
 2x_1 + x_2 + 3x_3 + x_4 + 2x_5 &= 2 \\
 90x_1 + 92x_2 + 92x_3 + 89x_4 + 88x_5 &= 91
 \end{aligned}$$

Dabei kommen nur solche Lösungen des linearen Gleichungssystems in Betracht, für die sämtliche  $x_i$  nicht kleiner als 0 sind. Diese Bedingungen lassen wir hier jedoch zunächst unbeachtet.

Zur Vereinfachung werden die Unbekannten weggelassen und nur noch die Koeffizienten und die rechten Seiten aufgeschrieben:

$$\begin{array}{ccccc|c}
 1 & 1 & 1 & 1 & 1 & 1 \\
 5 & 3 & 3 & 6 & 7 & 4 \\
 3 & 4 & 2 & 4 & 3 & 3 \\
 2 & 1 & 3 & 1 & 2 & 2 \\
 90 & 92 & 92 & 89 & 88 & 91
 \end{array}$$

Das links vom Strich stehende Koeffizientenschema heißt die *Koeffizientenmatrix* des linearen Gleichungssystems.

Beim ersten Schritt des Gauß-Algorithmus wird  $x_1$  durch elementare Umformungen

mit Hilfe der 1. Gleichung aus allen weiteren eliminiert. In der neuen Koeffizientenmatrix treten dann in der ersten Spalte unterhalb der Hauptdiagonalen nur noch Nullen auf.

$$\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -2 & -2 & 1 & 2 & -1 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 & 0 & 0 \\ 0 & 2 & 2 & -1 & -2 & 1 \end{array}$$

Wie man an den Koeffizienten und rechten Seiten in den vier unteren Zeilen erkennt, sind die 2. und die 5. Gleichungen äquivalent, ebenso die 3. und die 4. Gleichung. Daher genügt es, jeweils *nur noch eine* von zwei zueinander äquivalenten Gleichungen bei der weiteren Rechnung mitzuführen.

Beim nächsten Schritt hätte sich die Äquivalenz von 2. und 5. Zeile dadurch bemerkbar gemacht, dass die letzte Zeile ausschließlich mit Nullen besetzt worden wäre. Die Äquivalenz der 3. und 4. Gleichung hätte sich dann beim übernächsten Schritt durch das Auftreten einer weiteren Nullzeile an vorletzter Stelle gezeigt.

Eine Nullzeile entspricht aber die allgemeine Gleichung

$$0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4 + 0 \cdot x_5 = 0;$$

sie kann daher fortfallen.

Eine wichtige Ergänzung des Gauß–Algorithmus besteht darin, dass Nullzeilen weggelassen werden können bzw. dass von je zwei äquivalenten Gleichungen eine gestrichen werden kann. Statt der ursprünglichen fünf Gleichungen in unserem Beispiel brauchen wir also nur noch drei zu beachten.

Im nächsten Schritt wird durch elementare Umformungen eine Koeffizientenmatrix erzielt, die in der zweiten Spalte unterhalb der Hauptdiagonalen nur noch die Null aufweist.

$$\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & -2 & -2 & 1 & 2 & -1 \\ 0 & 0 & -2 & \frac{3}{2} & 1 & -\frac{1}{2} \end{array}$$

Die ersten drei Spalten bilden eine quadratische Matrix, die Dreiecksgestalt zeigt. Mehr kann durch die bisher beschriebenen Umformungsschritte hier nicht erreicht werden, weil die Koeffizientenmatrix nur „rechteckig“ und nicht „quadratisch“ ist.

Zum ursprünglichen LGS ist also das nur noch aus drei Gleichungen bestehende LGS

$$\begin{aligned}
 x_1 + x_2 + x_3 + x_4 + x_5 &= 1 \\
 (*) \quad - 2x_2 - 2x_3 + x_4 + 2x_5 &= -1 \\
 \quad - 2x_3 + \frac{3}{2}x_4 + x_5 &= -\frac{1}{2}
 \end{aligned}$$

äquivalent. Hiermit endet der Teil der Rechnung, der zur Vereinfachung des linearen Gleichungssystems durchgeführt wurde, und wir wenden uns nun der Bestimmung der Lösungsmenge zu.

Ein lineares Gleichungssystem bekannter Bauart erhält man sofort, wenn man  $x_4 = 0$  und  $x_5 = 0$  setzt. Man findet dann als Lösung:  $(\frac{1}{2}; \frac{1}{4}; \frac{1}{4}; 0; 0)$ .

Bei anderer Wahl der Zahlen  $x_4, x_5$  gelangt man zu einer anderen Lösung. Für  $x_4 = \frac{1}{6}$ ,  $x_5 = \frac{1}{16}$  ergibt sich zum Beispiel die Lösung  $(\frac{1}{8}; \frac{23}{96}; \frac{13}{32}; \frac{1}{6}; \frac{1}{16})$ .

Nach solchen Erfahrungen mit konkret gewählten  $x_4$  und  $x_5$  gewinnt man schließlich die Einsicht: Setzt man für  $x_4$  und  $x_5$  beliebige reelle Zahlen ein, so können  $x_1, x_2, x_3$  ausgerechnet werden;  $x_1, x_2, x_3$  sind durch die Wahl von  $x_4$  und  $x_5$  eindeutig bestimmt.

Um anzudeuten, dass  $x_4$  und  $x_5$  ganz beliebige reelle Zahlen sein können, verwendet man auch die Bezeichnungen  $\lambda$  bzw.  $\mu$  anstelle von  $x_4$  und  $x_5$ ;  $\lambda$  und  $\mu$  heißen in diesem Zusammenhang *Parameter*.

Das LGS gewinnt dann die Gestalt

$$\begin{aligned}
 x_1 + x_2 + x_3 &= 1 - \lambda - \mu \\
 - 2x_2 - 2x_3 &= -1 - \lambda - 2\mu \\
 - 2x_3 &= -\frac{1}{2} - \frac{3}{2}\lambda - \mu.
 \end{aligned}$$

Die vierte und fünfte Koordinate des Lösungstupels sind Parameter  $\lambda, \mu$ :

$$\begin{aligned}
 x_4 &= \lambda, \\
 x_5 &= \mu.
 \end{aligned}$$

Nun können – mit der letzten Gleichung beginnend – die Unbekannten  $x_1, x_2, x_3$  ausgerechnet werden.

Für die einzelnen Komponenten eines Lösungstupels ergibt sich:

$$\begin{aligned}
 x_1 &= \frac{1}{2} - \frac{3}{2}\lambda - 2\mu, \\
 x_2 &= \frac{1}{4} - \frac{1}{4}\lambda + \frac{1}{2}\mu, \\
 x_3 &= \frac{1}{4} + \frac{3}{4}\lambda + \frac{1}{2}\mu, \\
 x_4 &= \lambda, \\
 x_5 &= \mu,
 \end{aligned}$$

wobei  $\lambda$  und  $\mu$  beliebige reelle Zahlen sind.

Hierdurch sind alle Lösungen des *zuletzt erhaltenen* linearen Gleichungssystems dargestellt. Da nur Äquivalenzumformungen vorgenommen wurden, ist damit auch die Gesamtheit aller Lösungen des *ursprünglichen* linearen Gleichungssystems gewonnen.

An diesem Beispiel zeigt sich wieder die Struktur der Lösungsgesamtheit des linearen Gleichungssystems. Mit  $\lambda = \mu = 0$  findet man die spezielle Lösung  $\mathbf{x}^* = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0, 0)$ . Die Lösungsgesamtheit des zugehörigen homogenen Systems wird aufgespannt von den Basisvektoren  $\mathbf{h}_1 = (-\frac{3}{2}, -\frac{1}{4}, \frac{3}{4}, 1, 0)$  und  $\mathbf{h}_2 = (-2, \frac{1}{2}, \frac{1}{2}, 0, 1)$ . Also wird die Lösungsmenge beschrieben durch die Vektoren  $\mathbf{x} = \mathbf{x}^* + \lambda\mathbf{h}_1 + \mu\mathbf{h}_2$ .

**Bemerkung:** Die Basisvektoren  $\mathbf{h}_1$  und  $\mathbf{h}_2$  ergeben sich auch, wenn man bei (\*) zuerst  $x_4 = 1, x_5 = 0$  bzw.  $x_4 = 0, x_5 = 1$  setzt und dann jeweils nach  $x_3, x_2$  und  $x_1$  auflöst. ▲

Die Lösungen der eingangs gestellten *Mischungsaufgabe* gehören zu dieser Lösungsmenge. Nur müssen sie außerdem die Bedingungen  $x_j \geq 0 (j = 1, 2, 3, 4, 5)$  erfüllen (s.o.). Es muß also gelten:

$$\begin{aligned} \frac{1}{2} - \frac{3}{2}\lambda - 2\mu &\geq 0, \\ \frac{1}{4} - \frac{1}{4}\lambda + \frac{1}{2}\mu &\geq 0, \\ \frac{1}{4} + \frac{3}{4}\lambda + \frac{1}{2}\mu &\geq 0, \\ &\lambda \geq 0, \\ &\mu \geq 0. \end{aligned}$$

Durch diese Ungleichungen werden Halbebenen in einer  $\lambda, \mu$ -Ebene beschrieben, die durch die Geraden mit den Gleichungen

$$\begin{aligned} \frac{1}{2} - \frac{3}{2}\lambda - 2\mu &= 0, \\ \frac{1}{4} - \frac{1}{4}\lambda + \frac{1}{2}\mu &= 0, \\ \frac{1}{4} + \frac{3}{4}\lambda + \frac{1}{2}\mu &= 0, \\ &\lambda = 0, \\ &\mu = 0 \end{aligned}$$

berandet werden. Die „Punkte“  $(\lambda, \mu)$ , die zu Lösungen der Mischungsaufgabe führen, bilden den Durchschnitt aller dieser Halbebenen. Aus der Figur ergibt sich das grau getönte Dreieck  $ABC$  mit den Eckpunkten  $A = (0, 0)$ ,  $B = (\frac{1}{3}, 0)$  und  $C = (0, \frac{1}{4})$ . Dann und nur dann, wenn  $\lambda$  und  $\mu$  so gewählt werden, dass der „Punkt“  $(\lambda, \mu)$  im Inneren des Dreiecks  $ABC$  oder auf dessen Rand liegt, bilden die zugehörigen Tupel  $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$  eine Lösung der Mischungsaufgabe. ■

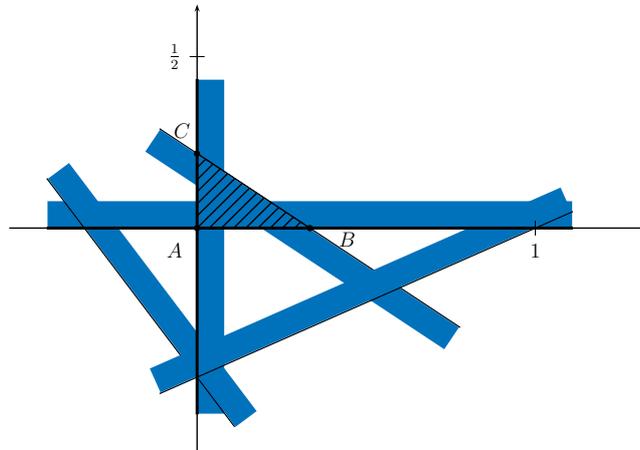


Abbildung 8.17: Der zulässige Bereich  $\triangle ABC$  für die Parameter  $\lambda, \mu$  als Durchschnitt von Halbebenen

**Bemerkung:** Die Freiheit der Auswahl unter vielen verschiedenen Lösungen – zu jedem Punkt des Dreiecks  $ABC$  gehört eine – kann man dazu nutzen, nach einer besonders kostengünstigen Lösung zu suchen. Dies ist ein Problem aus dem Gebiet „Lineare Optimierung“.

Die Kosten pro Mengeneinheit der verschiedenen Legierungen seien der folgenden Tabelle zu entnehmen:

	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$
Kosten in EU	20	16	20	13	28

Die Kostenfunktion  $\boldsymbol{x} \mapsto K(\boldsymbol{x})$  mit

$$K(\boldsymbol{x}) = 20x_1 + 16x_2 + 20x_3 + 13x_4 + 28x_5$$

ist also zu minimieren. Setzt man für  $\boldsymbol{x}$  ein Lösungstupel ein, so erhält man die Kosten als Funktion  $k$  der beiden Parameter  $\lambda$  und  $\mu$ :

$$k = 19 - 6\lambda + 6\mu.$$

Die Punkte mit den Koordinaten  $(\lambda, \mu, k)$  bilden eine Ebene über der  $\lambda, \mu$ -Ebene. Der Wert  $k$  gibt dann die Höhe eines Punktes der Ebene über dem Punkt mit den Koordinaten  $(\lambda, \mu)$  an. Von den Ebenenpunkten, die über dem Dreieck  $ABC$  liegen, kann nur ein über einem Eckpunkt des Dreiecks gelegener Punkt minimalen Abstand von der  $(\lambda, \mu)$ -Ebene haben.

Man erhält die Werte  $k(A) = 19$ ,  $k(B) = 19 - 6 \cdot \frac{1}{3}$  und  $k(C) = 19 + 6 \cdot \frac{1}{4}$ . Der Wert  $k(B)$ , der zu den Parametern  $\lambda = \frac{1}{3}$  und  $\mu = 0$  gehört ist also minimal. Die Kosten sind daher am geringsten, wenn  $\lambda = \frac{1}{3}$  und  $\mu = 0$  gilt. Sie betragen in diesem Fall EU 17,- pro Mengeneinheit der neuen Legierung.

Mit diesen Werten von  $\lambda$  und  $\mu$  gewinnt man nun die Anteile der einzelnen Legierungen:

$$x_1 = 0, \quad x_2 = \frac{1}{6}, \quad x_3 = \frac{1}{2}, \quad x_4 = \frac{1}{3}, \quad x_5 = 0.$$

Die Herstellfirma wird also auf den Einkauf der Legierungen  $L_1$  und  $L_5$  verzichten. ▲



# Kapitel 9

## Differentialgleichungen

Die Lösungen einer quadratischen Gleichung sind Zahlen, die Lösungen einer Differentialgleichung hingegen sind Funktionen. Eine Differentialgleichung stellt an eine gesuchte Funktion gewisse Bedingungen, die von dieser erfüllt werden sollen. Diese Bedingungen bestehen in einer Beziehung zwischen der Funktionsvariablen, den Funktionswerten und – das ist das Wesentliche – den Ableitungen der Funktion.

Die Dynamik der Naturvorgänge in Raum und Zeit wird von Differentialgleichungen beherrscht. Dies gilt gleichermaßen für die Bewegung der Himmelskörper wie für das Verhalten der kleinsten Teile der Materie. Aber auch in unserer Alltagswelt begegnen uns Phänomene, deren räumliche oder zeitliche Entwicklung von Differentialgleichungen gesteuert wird. Da ist zum Beispiel das Wachstum einer Population, die Interaktion zwischen zwei Populationen, die Abkühlung eines heißen Suppentopfes oder die Schwingung der Masse an einer Feder, um nur wenige dieser Phänomene zu benennen. Anhand solcher Beispiele soll im folgenden in das Gebiet der Differentialgleichungen eingeführt werden.

**Beispiele:** Stellt man an eine zu bestimmende Funktion  $f(x)$  die Bedingung

$$(*) \quad f'(x) = 1 + x,$$

dann wird diese offenbar von allen Funktionen

$$f(x) = x + \frac{x^2}{2} + c$$

erfüllt, wobei  $c$  eine beliebige Konstante ist. Fordert man zusätzlich, dass  $f(x)$  an der Stelle  $x = 0$  den Wert 1 annehmen soll, dass also  $f(0) = 1$  gelten soll, dann folgt  $f(0) = c = 1$ . Die Lösung der Differentialgleichung (\*) mit der „Anfangsbedingung“  $f(0) = 1$  ist also die Funktion

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

Eine Lösung der Differentialgleichung

$$(**) \quad f'(x) = f(x)$$

lässt sich sofort angeben, nämlich die Exponentialfunktion

$$f(x) = e^x,$$

von der wir wissen (vgl. Abschnitte 5.3, 6.3), dass sie an jeder Stelle  $x \in \mathbb{R}$  gleich dem Wert ihrer Ableitung ist. Offenbar sind aber auch alle Funktionen

$$f(x) = c e^x, \quad c \in \mathbb{R},$$

Lösungen von (\*\*), und das sind sogar alle Lösungen (vgl. Abschnitt 6.3). Stellt man zusätzlich die Anfangsbedingung  $f(0) = 5$ , so folgt  $c = 5$  und damit die Lösung

$$f(x) = 5 e^x.$$

In den beiden Beispielen war die erste Ableitung  $f'(x)$  Teil der gestellten Bedingung. Es handelte sich somit um Differentialgleichungen 1. Ordnung. Tritt auch die zweite Ableitung  $f''(x)$  der gesuchten Funktion in der gestellten Bedingungsgleichung auf, dann handelt es sich um eine Differentialgleichung 2. Ordnung. Eine solche ist zum Beispiel die Differentialgleichung

$$(***) \quad f''(x) = -f(x).$$

Spezielle Lösungen dieser Differentialgleichung findet man in den Funktionen  $u(x) = \cos x$  und  $v(x) = \sin x$ , wie man durch Einsetzen leicht bestätigt. Aber auch jede *Linearkombination*

$$f(x) = a \cos x + b \sin x$$

mit Koeffizienten  $a, b \in \mathbb{R}$  ist eine Lösung von (\*\*\*). Man hat damit aber auch schon alle Lösungen dieser Differentialgleichung gefunden, wie wir unten sehen werden (vgl. Abschnitt 9.5).

## 9.1 Exponentielles Wachstum und verwandte Phänomene

Wir betrachten in diesem Abschnitt drei Phänomene in ihrem zeitlichen Verlauf: Das Wachstum einer Population, den Zerfall einer radioaktiven Substanz und die Abkühlung eines heißen Körpers.

## Das Wachstum einer Population

Eine Population  $X$  habe zum Zeitpunkt  $t$  die Größe  $x(t)$ . Unter dem Begriff „Größe“ kann in diesem Zusammenhang verschiedenes verstanden werden. Es kann sich um die Individuenzahl, die Individuendichte (Anzahl der Individuen pro Flächeneinheit) oder um die Biomasse (zum Beispiel eines Waldes) handeln. Im folgenden verbinden wir mit der Größe  $x(t)$  eine „durchschnittliche“ Individuenzahl.

Wir nehmen zunächst an, das Wachstum der Population werde ausschließlich durch die Reproduktion der Individuen, also durch Geburten, verursacht.

Und zwar möge gelten: In jeder Zeiteinheit bringt jedes Individuum im Durchschnitt  $a$  neue Individuen hervor. Dann erscheint die folgende Überlegung plausibel: Der Zuwachs der Population  $X$  vom Zeitpunkt  $t$  bis zu einem Zeitpunkt  $t+h$ ,  $h > 0$ , ist gleich der Anzahl der in diesem Zeitraum erfolgten Geburten. In der Zeiteinheit erfolgen im Zeitintervall  $[t, t+h]$  dann  $ax(t^*)$  Geburten, wobei  $t^*$  ein zeitlicher Zwischenwert mit  $t \leq t^* \leq t+h$  ist, insgesamt also  $h \cdot ax(t^*)$  Geburten.

Also muss gelten

$$(*) \quad x(t+h) - x(t) = h a x(t^*).$$

Es kann auch angenommen werden, die Beziehung (\*) sei durch Messungen empirisch bestätigt.

Dividiert man dann bei (\*) beide Seiten durch  $h$ , so folgt die Gleichung

$$(**) \quad \frac{x(t+h) - x(t)}{h} = a x(t^*).$$

Für  $h \rightarrow 0$  erhält man so die Beziehung

$$(***) \quad \dot{x}(t) = a x(t).$$

Denn die linke Seite von (\*\*) geht gegen die Ableitung der Funktion  $x(t)$  an der Stelle  $t$  (vgl. Abschnitt 6.1), und  $t^*$  strebt gegen  $t$  aufgrund der Ungleichung  $t \leq t^* \leq t+h$ . Die Beziehung (\*\*\*) ist eine Differentialgleichung 1. Ordnung für die gesuchte Funktion  $x(t)$ . Die Ableitung  $\dot{x}(t)$  hat hier die Bedeutung der *Wachstumsgeschwindigkeit* der Population. Zu ihrer Lösung machen wir den Ansatz

$$x(t) = e^{\lambda t}$$

mit  $\lambda \in \mathbb{R}$ . Einsetzen in (\*\*\*) ergibt die Gleichung

$$\lambda e^{\lambda t} = a e^{\lambda t},$$

also gilt  $\lambda = a$  und damit erhalten wir die spezielle Lösung

$$x^*(t) = e^{at}.$$

Man sieht aber leicht, dass alle Vielfachen  $cx^*(t)$ ,  $c \in \mathbb{R}$ , ebenfalls Lösungen von (\*\*\*) sind und dies sind sogar alle Lösungen. Das folgt wie bei der Differentialgleichung für die Exponentialfunktion (vgl. Abschnitt 6.3). Damit hat die Lösungsgesamtheit der Differentialgleichung (\*\*\*) die Form

$$x(t) = ce^{at}, \quad c \in \mathbb{R}.$$

Zur Bestimmung von  $c$  nehmen wir an, die Größe der Population zum Zeitpunkt  $t = 0$  sei  $x(0) = x_0$ . Dann ergibt sich

$$x(0) = ce^0 = c$$

und damit

$$(EW) \quad x(t) = x_0 e^{at}.$$

Dies ist die Gleichung für das *exponentielle Wachstum* mit Anfangswert  $x_0$  und Geburtenrate (Wachstumsrate)  $a$ .

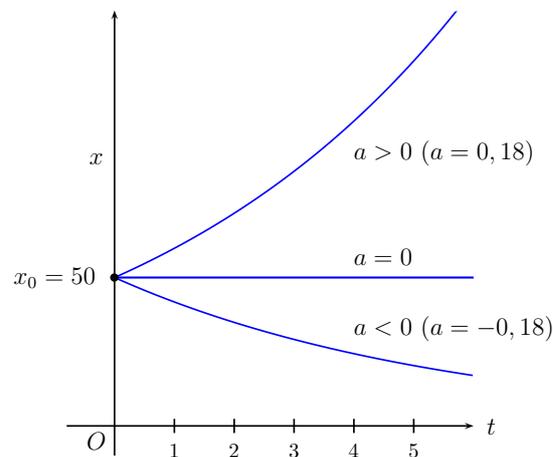


Abbildung 9.1: Exponentielles Wachstum

Offenbar hat die Beziehung (EW) eine unrealistische Konsequenz: Für  $t \rightarrow \infty$  wird  $x(t)$  unendlich groß. Diese Konsequenz vermeidet das Modell des logistischen Wachstums, das im folgenden Abschnitt entwickelt wird.

**Bemerkung:** Wir waren bei den obigen Überlegungen von einer Population  $X$  ohne Todesfälle ausgegangen. Diese Annahme ist natürlich unrealistisch. In einem realistischeren Modell muss an die Stelle der reinen Geburtenrate die Differenz  $a - b$  von Geburtenrate  $a$  und Todesrate  $b$  treten. Diese Differenz kann dann auch gleich null sein oder sogar negative Werte annehmen. Diese Fälle werden bei der Gleichung (EW) mitberücksichtigt, indem man auch die Fälle  $a = 0$  und  $a < 0$  zulässt. Im ersten Fall stagniert die Population bei ihrem Anfangswert  $x_0$ , im zweiten Fall stirbt sie aus. ▲

### Radioaktiver Zerfall

Beim radioaktiven Zerfall zerfallen die Atome einer radioaktiven Substanz durch Aussendung von Strahlung (Elementarteilchen) in die Atome eines anderen Elementes. Die von dem englischen Physiker Lord Rutherford (1871–1937) entdeckte Gesetzmäßigkeit des radioaktiven Zerfalls besagt:

Der Bruchteil der in einem Zeitintervall zerfallenden Atome einer radioaktiven Substanz ist näherungsweise proportional zur Länge dieses Zeitintervalls.

Ist also  $x(t)$  die Anzahl der Atome der Substanz zum Zeitpunkt  $t$ , dann sind im Zeitintervall  $[t, t + h]$  gerade  $x(t) - x(t + h)$  Atome zerfallen. Also gilt aufgrund des Rutherfordschen Gesetzes:

$$\frac{x(t) - x(t + h)}{x(t)} \cong \lambda \cdot h.$$

Die in dieser Beziehung auftretenden Proportionalitätskonstante  $\lambda > 0$  ist die *Zerfallsrate* der radioaktiven Substanz.

Stellt man diese Beziehung um, dann ergibt sich

$$\frac{x(t + h) - x(t)}{h} \cong -\lambda x(t).$$

Das Rutherfordsche Gesetz gilt umso präziser, je kleiner  $h$  ist. Für  $h \rightarrow 0$  erhält man also die Differentialgleichung

$$(RZ) \quad \dot{x}(t) = -\lambda x(t)$$

des radioaktiven Zerfalls. Ihre Lösung ist aufgrund derselben Überlegung wie beim exponentiellen Wachstum die Funktion

$$x(t) = x_0 e^{-\lambda t},$$

wenn  $x_0$  die Anfangsmenge zur Zeit  $t = 0$  ist.

Die Differentialgleichung (RZ) lässt sich auch folgendermaßen deuten: Die Zerfallsgeschwindigkeit der radioaktiven Substanz ist proportional zur vorhandenen Menge und sie ist umso größer, je größer die Zerfallsrate  $\lambda$  ist.

**Halbwertszeit:** Die Halbwertszeit einer radioaktiven Substanz ist die Zeit, in der die Substanz auf die Hälfte der ursprünglichen Menge zerfallen ist. Wir bezeichnen diese Zeit mit  $H$ . Dann muss also gelten

$$x(t + H) = \frac{1}{2} x(t)$$

oder

$$x_0 e^{-\lambda(t+H)} = \frac{1}{2} x_0 e^{-\lambda t}.$$

Daraus folgt

$$e^{-\lambda H} = \frac{1}{2},$$

also

$$-\lambda H = -\ln 2.$$

Folglich gilt

$$H = \frac{\ln 2}{\lambda}$$

Kennt man somit die Zerfallsrate einer Substanz, dann kennt man ihre Halbwertszeit und umgekehrt. Die Halbwertszeit von Kohlenstoff – 14 ( $C_{14}$ ) zum Beispiel beträgt 5568 Jahre, die Zerfallsrate ist daher

$$\lambda = \frac{\ln 2}{5568} = 1,2448 \cdot 10^{-4}.$$

**Radioaktive Datierung:** Zur Datierung archäologischer Fundstücke, das heißt zur Feststellung von deren Entstehungszeit, kann man das Phänomen des radioaktiven Zerfalls benutzen. Die Methode beruht auf der folgenden Überlegung: Kennt man die Zerfallsrate  $\lambda$  und die Anfangsmenge  $x_0$  einer zerfallenden Substanz und mißt man zu einem gegenwärtigen Zeitpunkt die Größe  $x = x(t)$  der noch vorhandenen Menge, dann lässt sich die zum Übergang von  $x_0$  zu  $x$  benötigte Zeit  $t$  berechnen. Denn wegen

$$x = x(t) = x_0 e^{-\lambda t}$$

hat man

$$e^{\lambda t} = \frac{x_0}{x}$$

und damit gilt

$$\lambda t = \ln \frac{x_0}{x},$$

also

$$(*) \quad t = \frac{1}{\lambda} \ln \frac{x_0}{x}.$$

Kennt man also, wie gesagt, die Zerfallsrate und die Anfangsmenge einer radioaktiven Substanz, die einem archäologischen Fundstück anhaftet, und mißt man deren gegenwärtig noch vorhandene Menge, dann kann man auf die Entstehungszeit des Fundstückes zurückrechnen.

Das Problem ist natürlich die Kenntnis der Anfangsmenge  $x_0$ . Hier hilft die folgende Beobachtung: Pflanzen reichern in ihrem Gewebe aus der Umwelt entnommenen radioaktiven Kohlenstoff  $C_{14}$  an, und zwar zu einem jeweils spezifischen prozentualen Anteil. Diese Anteile sind bekannt. Sie werden von den Pflanzen, solange diese leben, im ständigen Austausch mit der Umwelt stabil gehalten. Stirbt die Pflanze ab, wird der Austausch mit der Umwelt beendet und die im Gewebe eingelagerte Menge an  $C_{14}$  reduziert sich entsprechend dem Gesetz des radioaktiven Zerfalls.

**Beispiel:** Wie man der Formel (\*) entnimmt, kommt es zur Berechnung der Zeit  $t$  nur auf das Verhältnis  $x_0/x$  von Anfangswert und Ist-Wert der  $C_{14}$ -Mengen an. Stellt man bei einem Fundstück zum Beispiel ein entsprechendes Verhältnis  $x_0/x = 3/1$  fest, dann folgt mit  $\lambda = 1,2448 \cdot 19^{-4}$  ein Alter von ca. 8800 Jahren dieses Fundstückes. ■

### Abkühlungsvorgänge

Stellt man einen mit heißer Suppe gefüllten Topf zur Abkühlung auf den Balkon, dann interessiert natürlich die Frage, wie lange es wohl dauert, bis der Topf auf die Umgebungstemperatur abgekühlt ist. Bemerkenswerterweise lässt sich diese Frage, sofern nur die Anfangstemperatur bekannt ist, durch eine einzige weitere Temperaturmessung beantworten.

Ist nämlich  $x(t)$  die Temperatur eines heißen Körpers zum Zeitpunkt  $t$  und ist  $u$  die Umgebungstemperatur, dann gilt das Newtonsche Abkühlungsgesetz. Es besagt:

Die Abkühlungsgeschwindigkeit  $\dot{x}(t)$  des Körpers ist proportional zur Differenz  $x(t) - u$  zwischen der Temperatur des Objektes und der Umgebungstemperatur.

Es besteht also eine Beziehung der Form

$$(NAG) \quad \dot{x} = -k(x - u),$$

wobei  $k > 0$  eine materialabhängige Konstante ist. Das Minuszeichen vor der Konstanten bedeutet, dass die Temperatur des Körpers abnimmt, solange sie über der Umgebungstemperatur liegt. (Zur Vereinfachung der Schreibweise wurde die Variable  $t$  bei (NAG) weggelassen.)

Setzt man  $x - u = y$ , dann gilt  $\dot{x} = \dot{y}$  und die Beziehung (NAG) geht über in

$$\dot{y} = -ky.$$

Das ist eine uns vertraute Differentialgleichung mit der Lösungsgesamtheit

$$y = c e^{-kt}.$$

Also ist

$$x(t) = u + c e^{-kt}$$

die Lösungsgesamtheit der ursprünglichen Differentialgleichung. Mit der Anfangstemperatur  $x(0) = x_0$  erhält man dann

$$x_0 = u + c,$$

und damit in

$$(A) \quad x(t) = u + (x_0 - u) e^{-kt}$$

die Gleichung für die Abkühlung eines heißen Körpers.

Hat der Körper zu einem Zeitpunkt  $t_1 > 0$  die Temperatur  $x_1$ , dann gilt

$$x_1 = u + (x_0 - u) e^{-kt_1},$$

woraus folgt

$$e^{kt_1} = \frac{x_0 - u}{x_1 - u}$$

oder

$$k = \frac{1}{t_1} \ln \frac{x_0 - u}{x_1 - u}.$$

Kühlt daher ein Suppentopf nach einer Stunde bei einer Umgebungstemperatur von  $18^\circ$  von  $90^\circ$  auf  $70^\circ$  ab, dann ist

$$k = \ln \frac{72}{52} = 0,325$$

der Wert der Konstanten  $k$ .

Will man wissen, nach wieviel Stunden der Topf auf  $20^\circ$  abgekühlt ist, so hat man wegen (A) die Beziehung

$$20 = 18 + 72 e^{-0,325t}$$

nach  $t$  aufzulösen. Man erhält

$$e^{0,325t} = 36,$$

also bedarf es eines Zeitraums von

$$t = \frac{\ln 36}{0,325} \cong 11$$

Stunden bis zur Abkühlung auf  $20^\circ$ .

**Bemerkung:** Eine Konsequenz der Gleichung (A) ist eine unendlich lang Abkühlungszeit des heißen Körpers, bis er die Umgebungstemperatur erreicht hat. Insofern hat das Newtonsche Abkühlungsmodell einen leicht unrealistischen Zug. ▲

## 9.2 Logistisches Wachstum; explosives Wachstum

Bezieht man in die Überlegungen zum Wachstum einer Population neben der Geburtenrate  $a$  noch eine Sterberate  $b$  ein – sie gibt die Anzahl der Todesfälle pro Individuum und Zeiteinheit an – dann lautet die Differentialgleichung für die Wachstumsfunktion  $x(t)$  nunmehr

$$(*) \quad \dot{x}(t) = (a - b)x(t)$$

mit der Lösung

$$x(t) = x_0 e^{(a-b)t},$$

wenn  $x(0) = x_0$  ein vorgegebener Anfangswert ist.

Im Falle  $a - b > 0$  ergibt sich aber auch jetzt die unrealistische Konsequenz eines Wachstums ins Unendliche, wenn  $t$  immer größer wird. Gilt  $a = b$ , dann bleibt die Population bei  $x(t) = x_0$  zeitlich unverändert. Ist die Geburtenrate kleiner als die Sterberate, gilt also  $a - b < 0$ , dann stirbt die Population aus, das heißt es folgt  $x(t) \rightarrow 0$  für  $t \rightarrow \infty$ .

Das unrealistische Element bei diesem Modell ist offenbar die Annahme einer über alle Zeiten konstanten Todesrate  $b$ . Eine real existierende Population wird aber mit zunehmender Größe ihre eigenen Lebensgrundlagen beeinträchtigen, zum Beispiel durch Überweidung ihres Areals, so dass die Zahl der Sterbefälle mit der Größe der Population zunimmt. Die einfachste mathematische Annahme, die dieses Phänomen berücksichtigt, lautet dann: Die Sterberate  $b = b(x)$  ist proportional zur Größe  $x$  der Population. Das heißt, es gibt eine Konstante  $c > 0$ , so dass gilt  $b(x) = cx$ . Ersetzt man bei der Differentialgleichung (\*) also die Konstante  $b$  durch den Term  $cx$ , dann erhält man die Differentialgleichung

$$\dot{x} = (a - cx)x$$

oder

$$(LW) \quad \dot{x} = c(K - x)x$$

mit  $K = a/b$ . Dies ist eine *nichtlineare* Differentialgleichung 1. Ordnung, da die gesuchte Funktion  $x(t)$  auch in der zweiten Potenz auftritt. Es handelt sich hier um die Differentialgleichung des sogenannten *logistischen* Wachstums. (Die Herkunft dieser Bezeichnung ist unklar.)

### Das Richtungsfeld

Ohne die logistische Differentialgleichung explizit zu lösen, lassen sich schon aus der Form der Gleichung qualitative Schlüsse über den Verlauf der Lösungsfunktionen  $x = x(t)$  ziehen.

Spezielle Lösungen sind offenbar die konstanten Funktionen  $x(t) = 0$  und  $x(t) = K$ . Für  $0 < x < K$  gilt offenbar  $\dot{x} > 0$ , das heißt, Lösungen  $x(t)$ , die in dem Band  $0 < x(t) < K$  verlaufen, haben stets einen positiven Anstieg. Dieser hängt nur von der Größe  $x$  ab, nicht aber von der Zeit  $t$ . Ist  $x(t) > K$ , dann wird die rechte Seite der Differentialgleichung negativ, die Lösungskurven im Bereich oberhalb der konstanten Lösung  $x = K$  sind somit fallende Funktionen. Um negative Funktionswerte  $x(t) < 0$  muss man sich nicht kümmern, da diese in der Realität nicht vorkommen.

Die logistische Differentialgleichung ordnet also jedem Punkt  $(t, x)$  die Richtung  $\dot{x} = f(x) = c(K - x)x$  zu. Die Gesamtheit dieser Richtungen bildet das Richtungsfeld der Differentialgleichung. Maximalen Anstieg haben die Lösungskurven bei  $x = \frac{K}{2}$ , da dort die Parabel mit der Gleichung  $f(x) = c(K - x)x$  ihr Maximum annimmt.

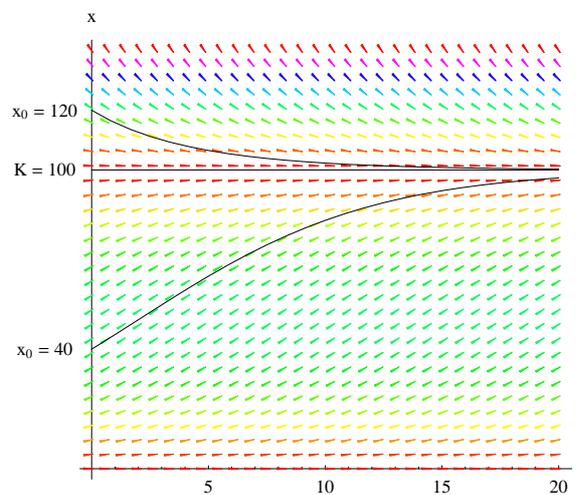


Abbildung 9.2: Das Richtungsfeld der logistischen Differentialgleichung und zwei Lösungen.

Die Aufgabe, die obige Differentialgleichung zu lösen, lässt sich jetzt auch so formulieren: Gesucht sind Funktionen  $x(t)$ , deren Anstieg in jedem Punkt dem durch die Differentialgleichung bestimmten Richtungsfeld entspricht. Die Lösungen der Differentialgleichung sind also diejenigen Funktionen, die perfekt in das vorgegebene Richtungsfeld eingepasst sind. Auf diese Weise gewinnt man ein qualitatives Bild vom Verlauf der Lösungskurven.

### Explizite Lösung der logistischen Differentialgleichung

Man bringt die Differentialgleichung zunächst auf die Form

$$\frac{\dot{x}}{(K - x)x} = c.$$

Für die rationale Funktion  $\frac{1}{(K-x)x}$  findet man die Partialbruchzerlegung

$$\frac{1}{(K-x)x} = \frac{1}{K} \left( \frac{1}{K-x} + \frac{1}{x} \right)$$

und damit die Differentialgleichung

$$(*) \quad \frac{\dot{x}}{K-x} + \frac{\dot{x}}{x} = cK.$$

Eine Stammfunktion der Funktion  $\dot{x}/x$  ist die Funktion  $\ln x$ , wenn  $x > 0$  ist. Eine Stammfunktion von  $\dot{x}/(K-x)$  ist im Fall  $0 < x(t) < K$  die Funktion  $-\ln(K-x)$  (vgl. Abschnitt 7.4) und  $cKt + C$  mit einer Konstanten  $C$  ist eine Stammfunktion der rechten Seite  $cK$ . Daher muss eine Beziehung der Form

$$-\ln(K-x(t)) + \ln x(t) = cKt + C$$

oder

$$\ln \frac{x(t)}{K-x(t)} = cKt + C$$

gültig sein. Mit dem Anfangswert  $x(0) = x_0, 0 < x_0 < K$ , erhält man daraus die Beziehung

$$C = \ln \frac{x_0}{K-x_0}.$$

Damit ergibt sich eine implizite Gleichung für die gesuchte Funktion  $x(t)$ :

$$\ln \frac{x(t)}{K-x(t)} = cKt + \ln \frac{x_0}{K-x_0}$$

oder

$$(**) \quad \ln \left( \frac{x_0}{K-x_0} \cdot \frac{K-x(t)}{x(t)} \right) = -cKt$$

oder

$$\frac{x_0}{K-x_0} \cdot \frac{K-x(t)}{x(t)} = e^{-cKt}.$$

Aus der letzten Zeile wiederum folgt

$$\frac{K}{x(t)} = 1 + \frac{K-x_0}{x_0} \cdot e^{-cKt},$$

und schließlich die Lösung

$$(LG) \quad x(t) = \frac{K}{1 + \frac{K-x_0}{x_0} e^{-cKt}}.$$

Das ist die Gleichung des logistischen Wachstums für den Fall, dass der Anfangswert  $x_0$  im Bereich  $0 < x_0 < K$  liegt.

Die gefundene Lösung  $x(t)$  verläuft ganz in dem Streifen zwischen den konstanten Lösungen  $x(t) = 0$  und  $x(t) = K$  und für  $t \rightarrow \infty$  gilt  $x(t) \rightarrow K$ . Die Zahl  $K$  ist daher eine obere Schranke für die in diesem Streifen verlaufende Lösungen. Sie charakterisiert die Aufnahmefähigkeit (Kapazität) einer Umwelt für eine in ihr lebende Population.

**Bemerkung:** Eine im Streifen zwischen den Werten 0 und  $K$  verlaufende Lösung  $x(t)$  kann diesen Streifen nicht verlassen. Sie müßte dann nämlich die konstante Lösung  $x(t) = 0$  oder  $x(t) = K$  schneiden. Hierzu müßte sie aber im Schnittpunkt einen Anstieg  $\dot{x} \neq 0$  haben, was nicht möglich ist. ▲

Zur Berechnung der Lösungen  $x(t) > K$  stellen wir die Beziehung (\*) leicht um und schreiben

$$\frac{-\dot{x}}{x-K} + \frac{\dot{x}}{x} = cK,$$

so dass die Nenner auf der linken Seite beide positiv sind. Der Übergang zu den entsprechenden Stammfunktionen ergibt dann die Beziehung

$$-\ln(x(t) - K) + \ln x(t) = cKt + C$$

mit einer Konstanten  $C$ , also

$$\ln \frac{x(t)}{x(t) - K} = cKt + C.$$

Mit  $t = 0$  und  $x(0) = x_0 > K$  ergibt sich nun

$$C = \ln \frac{x_0}{x_0 - K}$$

und damit

$$\ln \left( \frac{x_0}{x_0 - K} \cdot \frac{x(t) - K}{x(t)} \right) = -cKt.$$

Diese Beziehung unterscheidet sich nicht von der entsprechenden Beziehung (\*\*) oben, also erhalten wir nach der Umstellung  $(K - x_0)/x_0 = -(x_0 - K)/x_0$  aus (LG) die Lösung

$$(LG') \quad x(t) = \frac{K}{1 - \frac{x_0 - K}{x_0} e^{-cKt}}.$$

Der Nenner der rechten Seite ist immer kleiner als eins, also gilt  $x(t) > K$  wie aufgrund der Vorüberlegungen und der Struktur des Richtungsfeldes zu erwarten war. Für  $t \rightarrow \infty$  gilt wieder  $x(t) \rightarrow K$ .

Die Lösungen  $x(t)$  tendieren also immer „von oben“ oder „von unten“ gegen die konstante Lösung  $x(t) = K$ . „Störungen“ aus der Lage  $x(t) = K$  werden daher immer in Richtung  $K$  korrigiert. Man nennt die konstante Lösung  $x(t) = K$  daher auch eine *stabile Lösung* der logistischen Differentialgleichung.

## Explosives Wachstum

Es gibt Populationen, die sich nur dann erfolgreich vermehren, wenn ihre Größe oberhalb eines Schwellenwertes liegt. Hat die Population diesen Schwellenwert einmal unterschritten, dann stirbt sie aus. Dieses Verhalten zeigte zum Beispiel die amerikanische Wandertaube, deren Bestand als Folge intensiver Bejagung in 19. Jahrhundert unter eine kritische Größe absank. Das letzte Exemplar dieser Taubenart wurde im Jahre 1914 geschossen.

Zur Beschreibung eines Wachstumsvorgangs, bei dem das Reproduktionsverhalten von der jeweiligen Größe des Bestandes stimuliert wird, gehen wir zunächst wieder von der Differentialgleichung

$$\dot{x}(t) = (a - b)x(t)$$

für das exponentielle Wachstum mit konstanter Geburten- und Todesrate aus. Ist nun, wie angenommen, die Geburtenrate nicht konstant, sondern wächst sie mit der Populationsgröße, so ist der einfachste Zusammenhang zwischen Geburtenrate  $a$  und Populationsgröße  $x$  der dieser Annahme entspricht von der Form  $a = c \cdot x$  mit einer Konstanten  $c > 0$ .

Setzt man diese Beziehung in die Differentialgleichung (\*) ein, dann erhält man die Differentialgleichung

$$\dot{x}(t) = (cx(t) - b)x(t)$$

oder

$$(*) \quad \dot{x}(t) = -c(S - x(t))x(t)$$

mit  $S = b/c$ . Aus dieser Differentialgleichung liest man ab: Spezielle Lösungen sind die konstanten Funktionen  $x(t) = 0$  und  $x(t) = S$ . Ist  $0 < x(t) < S$ , dann ist  $\dot{x}(t) < 0$ , die Lösung  $x(t)$  ist eine fallende Funktion. Ist hingegen  $x(t) > S$ , so gilt  $\dot{x}(t) > 0$ , die Lösung  $x(t)$  ist ansteigend. Der Parameter  $S$  hat also in der Tat die Bedeutung eines Schwellenwertes. Zur expliziten Lösung der Differentialgleichung (\*) bedarf es keiner erneuten Rechnung, denn die Differentialgleichung für das logistische Wachstum (siehe oben) geht in die Differentialgleichung (\*) über, indem man den Parameter  $c$  durch  $-c$  ersetzt und den Parameter  $K$  durch  $S$ . Daher erhält man die Lösungen der Differentialgleichung (\*) zu einem Anfangswert  $x(0) = x_0$ , indem man bei der logistischen

Gleichung (LG) den Parameter  $c$  durch  $-c$  und  $K$  durch  $S$  ersetzt. So findet man die Lösung.

$$(EW) \quad x(t) = \frac{S}{1 + \frac{S-x_0}{x_0} e^{cSt}}$$

Für  $t = 0$  gilt in der Tat  $x(0) = x_0$ . Ist  $x_0 < S$ , dann folgt  $x(t) \rightarrow 0$  für  $t \rightarrow \infty$ , das heißt, die Population mit einem Anfangswert unterhalb des Schwellenwertes stirbt aus. Ist  $x_0 > S$ , dann hat die Nennerfunktion bei obiger Beziehung die Form

$$N(t) = 1 - \frac{x_0 - S}{x_0} e^{cSt}.$$

Der Faktor  $(x_0 - S)/x_0$  bei der Exponentialfunktion  $e^{cSt}$  ist kleiner als eins, während diese, mit eins beginnend, monoton wächst. Folglich gibt es einen Zeitpunkt  $t_\infty$ , so dass  $N(t_\infty) = 0$  ist. Dann gilt aber  $x(t_\infty) = \infty$ .

Das bedeutet, die Population, die mit einem Anfangswert  $x_0 > S$  startet, wird in einem *endlichen* Zeitraum unendlich groß. Das Wachstum einer solchen Population geht also nach endlicher Zeit in eine explosive Phase über und muss in einer Katastrophe enden (Heuschrecken-Phänomen).

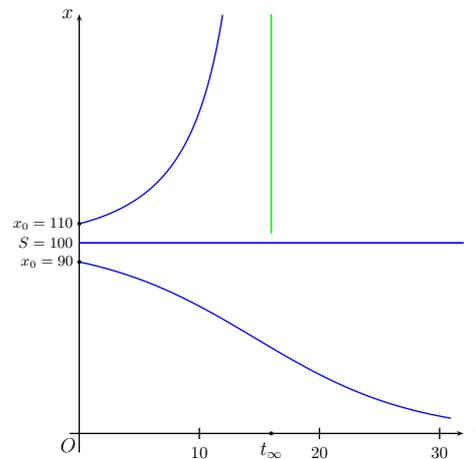


Abbildung 9.3: Explosives Wachstum mit Schwellenwert  $S = 100$ .

**Bemerkung:** Das betrachtete Wachstumsmodell kann natürlich so abgeändert werden, dass oberhalb des Schwellenwertes  $S$  noch eine Schranke  $K$  für das Wachstum der Population liegt. ▲

Den Wert  $t_\infty$  berechnet man aus der Gleichung

$$1 - \frac{x_0 - S}{x_0} e^{cSt_\infty} = 0$$

und erhält

$$e^{cSt_\infty} = \frac{x_0}{x_0 - S},$$

also ist

$$t_\infty = \frac{1}{cS} \ln \frac{x_0}{x_0 - S}.$$

### 9.3 Die lineare Differentialgleichung erster Ordnung

Die im Abschnitt 9.1 behandelten Differentialgleichungen für das exponentielle Wachstum, den radioaktiven Zerfall und für die Abkühlung eines heißen Körpers gehören sämtlich zum Typus der *linearen Differentialgleichung erster Ordnung*. Diese hat die Form

$$(*) \quad \dot{x}(t) + a(t)x(t) = b(t)$$

mit stetigen Funktionen  $a(t), b(t)$ , die in einem Intervall  $I$  erklärt sind.

Die lineare Differentialgleichung erster Ordnung, zusammen mit einer Anfangswertvorgabe, lässt sich mit Hilfe von Integrationen immer explizit lösen.

Wie im Falle linearer Gleichungssysteme untersuchen wir zunächst die Struktur der Lösungsgesamtheit der Differentialgleichung (\*). Hierzu nehmen wir an, wir hätten in der Funktion  $x^*(t)$  bereits eine spezielle Lösung von (\*) gefunden. Ist nun  $x(t)$  irgendeine Lösung von (\*), dann gelten die Beziehungen

$$\begin{aligned} \dot{x} + a x &= b, \\ \dot{x}^* + a x^* &= b. \end{aligned}$$

Zieht man die zweite Zeile von der ersten ab, so erhält man

$$(\dot{x} - \dot{x}^*) + a(x - x^*) = 0.$$

Die Funktion  $x - x^* = h$  ist also eine Lösung der zu (\*) gehörigen homogenen Differentialgleichung

$$(h) \quad \dot{x} + a x = 0.$$

Also gilt

$$x(t) = x^*(t) + h(t),$$

wobei  $h(t)$  eine Lösung der homogenen Differentialgleichung ist. In völliger Analogie zur Lösungstheorie linearer Gleichungssysteme gilt daher:

**Satz:** Die Lösungsgesamtheit der linearen Differentialgleichung erster Ordnung

$$(i) \quad \dot{x}(t) + a(t)x(t) = b(t)$$

hat die Form

$$x(t) = x^*(t) + h(t),$$

wobei  $x^*(t)$  eine spezielle Lösung der inhomogenen Differentialgleichung (i) ist und  $h(t)$  irgendeine Lösung der zugehörigen homogenen Differentialgleichung

$$(h) \quad \dot{x}(t) + a(t)x(t) = 0$$

ist.

Abweichend von der Vorgehensweise bei den linearen Gleichungssystemen bestimmen wir zunächst die Lösungsgesamtheit der homogenen Gleichung (h). Hierzu machen wir den Ansatz

$$h(t) = e^{\varphi(t)}$$

mit einer zu bestimmenden Funktion  $\varphi(t)$ . Einsetzen in (h) ergibt

$$\dot{\varphi}(t)e^{\varphi(t)} + a(t)e^{\varphi(t)} = 0,$$

also gilt

$$\dot{\varphi}(t) = -a(t)$$

und man findet

$$\varphi(t) = -A(t),$$

wobei

$$A(t) = \int a(\tau) d\tau$$

eine Stammfunktion von  $a(t)$  ist. Mit

$$h(t) = e^{-A(t)}$$

ist auch jedes Vielfache  $c \cdot h(t)$  mit  $c \in \mathbb{R}$  eine Lösung von (h) und dies sind sogar sämtliche Lösungen von (h).

**Bemerkung:** Die Richtigkeit der letzteren Aussage lässt sich folgendermaßen einsehen: Sei  $h(t)$  irgendeine Lösung von (h). Man bilde dann die Funktion

$$p(t) = h(t)e^{A(t)}.$$

Deren Ableitung ist

$$\begin{aligned}\dot{p}(t) &= \dot{h}(t)e^{A(t)} + h(t)a(t)e^{A(t)} \\ &= (\dot{h}(t) + a(t)h(t))e^{A(t)}.\end{aligned}$$

Da  $h(t)$  laut Annahme eine Lösung von (h) ist, so ist der Ausdruck in der Klammer gleich null. Es folgt  $\dot{p}(t) = 0$ , also  $p(t) = c$ ,  $c \in \mathbb{R}$ . Wir erhalten

$$h(t)e^{At} = c,$$

also gilt

$$h(t) = ce^{-A(t)},$$

was zu zeigen war. ▲

Es bleibt eine spezielle Lösung  $x^*(t)$  von (i) zu berechnen. Hierzu macht man einen Ansatz, der als *Variation der Konstanten* bezeichnet wird. Man ersetzt hierzu die Konstante  $c$  bei der homogenen Lösung

$$h(t) = ce^{-A(t)}$$

durch eine Funktion  $\gamma(t)$  und schreibt

$$x^*(t) = \gamma(t)e^{-A(t)}.$$

Setzt man nun  $x^*(t)$  in die inhomogene Differentialgleichung (i) ein, dann ergibt sich die Beziehung

$$\dot{\gamma}(t)e^{-A(t)} - \gamma(t)a(t)e^{-A(t)} + a(t)\gamma(t)e^{-A(t)} = b(t)$$

oder

$$\dot{\gamma}(t) = e^{A(t)}b(t).$$

Also ist  $\gamma(t)$  eine Stammfunktion der rechten Seite dieser Gleichung:

$$\gamma(t) = \int e^{A(\tau)}b(\tau) d\tau.$$

Damit erhält die Lösungsgesamtheit der inhomogenen Differentialgleichung (i) die Form

$$x(t) = e^{-A(t)} \int e^{A(\tau)}b(\tau) d\tau + ce^{-A(t)}.$$

Soll die Lösung  $x(t)$  die Anfangsbestimmung  $x(0) = x_0$  erfüllen, so muss man nur  $c = x_0$  setzen und dafür sorgen, dass  $x^*(0) = 0$  und  $A(0) = 0$  ist. Dies erreicht man durch spezielle Wahl der beteiligten Stammfunktionen indem man setzt

$$\gamma(t) = \int_0^t e^{A(\tau)} b(\tau) d\tau$$

und

$$A(t) = \int_0^t a(\tau) d\tau$$

**Satz:** Die Lösung  $x(t)$  der linearen Differentialgleichung erster Ordnung

$$\dot{x}(t) + a(t)x(t) = b(t),$$

die zugleich die Anfangsbedingung  $x(0) = x_0$  erfüllt, hat die Form

$$(*) \quad x(t) = e^{-A(t)} \int_0^t e^{A(\tau)} b(\tau) d\tau + x_0 e^{-A(t)},$$

wobei

$$(**) \quad A(t) = \int_0^t a(\tau) d\tau$$

ist.

**Beispiel:** Eine Population habe eine mit der Zeit abnehmende Geburtenrate  $a(t) = 1/(1+t)$  und eine konstante Todesrate  $b > 0$ . Einwanderung finde nicht statt. Dann erfüllt ihre Wachstumsfunktion  $x(t)$  die Differentialgleichung

$$\dot{x} = \left( \frac{1}{1+t} - b \right) x.$$

Es gilt dann

$$\frac{\dot{x}}{x} = \frac{1}{1+t} - b.$$

Geht man auf der linken und rechten Seite zu den jeweiligen Stammfunktionen über, so erhält man mit einer Konstanten  $c$  die Beziehung

$$\log x = \log(1+t) - bt + c$$

oder

$$x(t) = C(1+t)e^{-bt}$$

mit  $C = e^c$ . Mit  $t = 0$  ergibt sich  $C = x(0) = x_0$  und somit die Wachstumsfunktion

$$x(t) = x_0(1+t)e^{-bt}.$$

Es folgt  $x(t) \rightarrow 0$  für  $t \rightarrow \infty$ . Die Population stirbt aus.

Wir betrachten nun den Fall, dass eine Population auch Zuwanderung erfährt. Wir nehmen an, dass pro Zeiteinheit im Durchschnitt  $c$  Individuen zuwandern. Bei zeitabhängiger Geburtenrate  $a(t)$  und konstanter Todesrate  $b$  beträgt dann die Veränderung welche die Population in Zeitintervall  $[t, t+h]$  erfährt

$$x(t+h) - x(t) \approx (a(t) - b)x(t)h + ch.$$

Dividiert man auf beiden Seiten durch  $h$  und lässt  $h \rightarrow 0$  gehen, so folgt für  $x(t)$  die Differentialgleichung

$$\dot{x} = (a(t) - b)x + c.$$

Dies ist eine inhomogene Differentialgleichung 1. Ordnung. Mit  $a(t) = 1/(1+t)$  hat man dann

$$\dot{x} = \left(\frac{1}{1+t} - b\right)x + c.$$

Die Lösungsformel für diese Differentialgleichung ergibt sich aus dem obigen Satz. Mit

$$A(t) = -\left(\frac{1}{1+t} - b\right)t$$

hat man zunächst

$$\begin{aligned} A(t) &= -\int_0^t \left(\frac{1}{1+\tau} - b\right) d\tau \\ &= -(\ln(1+\tau) - b\tau) \Big|_0^t \\ &= -(\ln(1+t) - bt). \end{aligned}$$

Also ist

$$e^{-A(t)} = (1+t)e^{-bt}$$

und

$$e^{A(t)} = \frac{1}{1+t} e^{bt}.$$

Mit  $b(t) = c$  und  $x(0) = x_0$  erhält man damit die Lösung

$$x(t) = (1+t) e^{-bt} \int_0^t \frac{1}{1+\tau} e^{b\tau} c d\tau + x_0(1+t) e^{-bt}.$$

Der zweite Term auf der rechten Seite geht gegen null für  $t \rightarrow \infty$ . Den ersten Term bringen wir auf die Form

$$\frac{f(t)}{g(t)} = c \int_0^t \frac{e^{b\tau}}{1+\tau} d\tau \Big/ \frac{e^{bt}}{1+t}.$$

Für  $t \rightarrow \infty$  erhalten wir einen unbestimmten Ausdruck der Form  $\frac{\infty}{\infty}$ . Anwendung der Regel von d'Hospital (vgl. Abschnitt 6.4) ergibt dann

$$\frac{\dot{f}(t)}{\dot{g}(t)} = \frac{ce^{bt}}{1+t} \Big/ \left( \frac{be^{bt}}{(1+t)} - \frac{e^{bt}}{(1+t)^2} \right) = \frac{c}{b - \frac{1}{1+t}}$$

Damit erhalten wir das Ergebnis: Bei zeitlich abnehmender Geburtenrate, konstanter Sterberate  $b$  und konstanter Zuwanderungsrate  $c$  strebt die Population  $x(t) \rightarrow c/b$  für  $t \rightarrow \infty$ . ■

## 9.4 Schwingungen

In seinem für die moderne Physik grundlegenden Werk *Philosophiae Naturalis Principia Mathematica* (= Mathematische Prinzipien der Naturphilosophie), auch kurz *Principia* genannt, formuliert Isaac Newton (1643–1727), Physiker und Mathematiker, drei Axiome der Mechanik. Das zweite dieser Axiome besagt: Jede Geschwindigkeitsänderung eines sich bewegenden Körpers ist proportional der auf ihn einwirkenden Kraft. Die Bewegung eines Massenpunktes auf einer Geraden wird durch eine Funktion  $x(t)$  beschrieben, welche die Lage (den Ort) des Punktes zur Zeit  $t$ , bezogen auf einen Nullpunkt, angibt. Die Geschwindigkeit der Bewegung ist dann die erste Ableitung  $\dot{x}(t)$  der Ortsfunktion  $x(t)$  und deren zweite Ableitung  $\ddot{x}(t)$  beschreibt die durch die Einwirkung einer Kraft  $F$  (= Force) erfahrene Geschwindigkeitsänderung oder Beschleunigung  $a$  (= acceleration). Der im Newtonschen Axiom auftretende Proportionalitätsfaktor ist die Masse  $m$  des Körpers. Es gilt dann

$$F = m \cdot a$$

oder

$$(*) \quad m \cdot \ddot{x}(t) = F.$$

Lenkt man einen Massenpunkt, der an einer Feder befestigt ist, aus seiner Ruhelage  $x = 0$  aus, dann tritt eine Rückstellkraft  $F$  auf, die gegen die Bewegungsrichtung des Massenpunktes wirkt. Lässt man die Masse los, dann schwingt sie um ihre Ruhelage.

Das Hooksche Gesetz (benannt nach seinem Entdecker Robert Hooke (1635–1703), Zeitgenosse und Feind Newtons) nun besagt: die Rückstellkraft  $F$  ist proportional zur Auslenkung  $x$ ,  $F = k \cdot x$ , wobei  $k > 0$  eine Materialkonstante ist. Da die Rückstellkraft gegen die Bewegungsrichtung wirkt, lautet die Beziehung (\*) im Falle der an einer Feder befestigten Masse:

$$m \cdot \ddot{x}(t) = -k \cdot x(t)$$

oder

$$(SG) \quad \ddot{x}(t) + \omega^2 x(t) = 0$$

mit  $\omega^2 = k/m$ . Diese Notation wird sich bald als sinnvoll erweisen. Die Differentialgleichung (SG) ist eine lineare und homogene Differentialgleichung zweiter Ordnung, die *Schwingungsgleichung*.

## Schwingung und Kreisbewegung

Zu ihrer Lösung macht man den mehrfach bewährten Ansatz

$$x(t) = e^{\lambda t}, \quad \lambda \in \mathbb{C}.$$

Es ist  $\dot{x}(t) = \lambda e^{\lambda t}$  und  $\ddot{x}(t) = \lambda^2 e^{\lambda t}$ . Durch Einsetzen in die Schwingungsgleichung erhält man die Beziehung

$$\lambda^2 e^{\lambda t} + \omega^2 e^{\lambda t} = 0,$$

also gilt

$$\lambda^2 + \omega^2 = 0$$

und damit ist

$$\lambda = \pm i\omega.$$

Der obige Ansatz führt so auf die komplexen Lösungen

$$x_+(t) = e^{i\omega t}$$

und

$$x_-(t) = e^{-i\omega t}.$$

Die Funktion  $x_+(t)$  beschreibt die Bewegung eines Punktes auf dem Einheitskreis der komplexen Ebene, denn es ist  $|x_+(t)| = 1$  (vgl. Abschnitte 2.4, 5.5) Die Bewegung erfolgt gegen den Uhrzeigersinn mit der Geschwindigkeit

$$\dot{x}_+(t) = i\omega e^{i\omega t} = i\omega x_+(t).$$

Der Geschwindigkeitsvektor  $\dot{x}_+(t)$  bildet also mit dem Ortsvektor  $x_+(t)$  des sich bewegenden Punktes einen Winkel von  $90^\circ$ , ist also tangential zum Einheitskreis. Die Länge des Geschwindigkeitsvektors, das ist der *Betrag* der Geschwindigkeit, ist  $|\dot{x}_+(t)| = \omega$ .

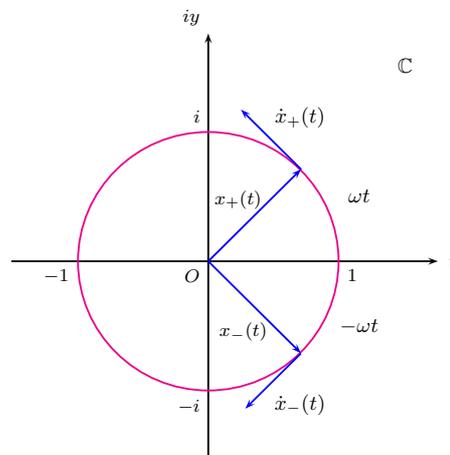


Abbildung 9.4: Die Kreisbewegungen  $x_+(t)$  und  $x_-(t)$ .

Für  $t = 0$  ist  $x_+(0) = 1$ . In der Zeiteinheit (zum Beispiel eine Sekunde) legt der Punkt die Strecke der Länge  $\omega$  vom Punkt  $z = 1$  bis zum Punkt  $z = e^{i\omega}$  zurück. Er hat dabei  $\omega/2\pi$  mal den Einheitskreis umlaufen. Man nennt die Zahl  $\omega$  daher auch die *Kreisfrequenz*. Durch die Funktion  $x_-(t)$  wird eine zu  $x_+(t)$  gegenläufige, also im Uhrzeigersinn erfolgende, Kreisbewegung beschrieben.

Die Form der Bewegung des an der Feder schwingenden Massenpunktes ist durch die *Anfangslage*  $x(0)$  und die *Anfangsgeschwindigkeit*  $\dot{x}(0)$  eindeutig festgelegt. Mathematisch kann man zeigen, dass jede Lösung  $x(t)$  der Schwingungsgleichung als Linearkombination.

$$x(t) = c_+ x_+(t) + c_- x_-(t)$$

mit Koeffizienten  $c_+, c_- \in \mathbb{C}$  der speziellen Lösungen  $x_+(t)$  und  $x_-(t)$  dargestellt werden kann. Also kann die Schwingung  $x(t)$  mit den Anfangswerten  $x(0)$  und  $\dot{x}(0)$

ebenfalls in dieser Form geschrieben werden. Damit erhält man zur Bestimmung der Koeffizienten  $c_+$  und  $c_-$  die Gleichungen

$$(1) \quad \begin{aligned} c_+ + c_- &= x(0) \\ c_+ i \omega - c_- i \omega &= \dot{x}(0) \end{aligned}$$

oder

$$(2) \quad c_+ - c_- = \frac{\dot{x}(0)}{i\omega} = -i \frac{\dot{x}(0)}{\omega}.$$

Durch Addition der Gleichungen (1) und (2) findet man dann

$$c_+ = \frac{1}{2} \left( x(0) - i \frac{\dot{x}(0)}{\omega} \right)$$

und durch Subtraktion von (1) und (2) schließlich

$$c_- = \frac{1}{2} \left( x(0) + i \frac{\dot{x}(0)}{\omega} \right).$$

Die komplexen Zahlen  $c_+$  und  $c_-$  schreiben wir in der Polarform

$$c_+ = \frac{1}{2} A e^{-i\alpha}$$

und

$$c_- = \frac{1}{2} A e^{i\alpha}.$$

Dabei ist

$$(*) \quad A = \sqrt{x(0)^2 + \left( \frac{\dot{x}(0)}{\omega} \right)^2}$$

und der Winkel  $\alpha$  wird gegeben durch die Beziehung

$$\cos \alpha = \frac{x(0)}{A}.$$

Damit erhält man

$$\begin{aligned} x(t) &= c_+ x_+(t) + c_- x_-(t) \\ &= \frac{1}{2} A e^{-i\alpha} e^{i\omega t} + \frac{1}{2} A e^{i\alpha} e^{-i\omega t} \\ &= \frac{1}{2} A (e^{i(\omega t - \alpha)} + e^{-i(\omega t - \alpha)}) \end{aligned}$$

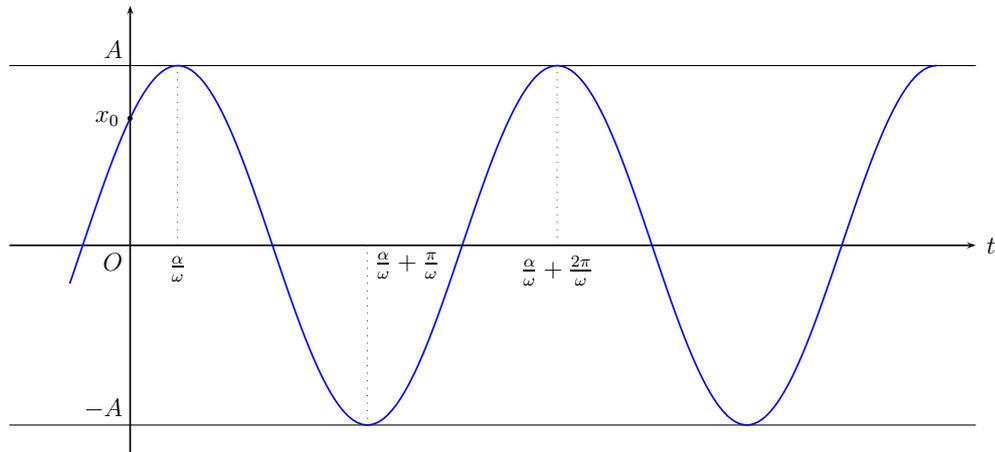


Abbildung 9.5: Die Lösungsfunktion  $x(t) = A \cos(\omega t - \alpha)$  der Schwingungsgleichung.

und aufgrund der Eulerformel (vgl. Abschnitt 2.4) schließlich die Darstellung

$$x(t) = A \cos(\omega t - \alpha).$$

Das ist eine  $\cos$ -Schwingung mit der *Amplitude*  $A$  und dem *Phasenwinkel*  $\alpha$ . Der Formel (\*) entnimmt man: Je größer Anfangslage  $x(0)$  und Anfangsgeschwindigkeit  $\dot{x}(0)$  sind, umso größer ist Amplitude (Weite) der Schwingung. Dabei ist es gleichgültig, in welche Richtung die Anfangsgeschwindigkeit den Massenpunkt bewegt.

Die Funktion  $A \cos(\omega t - \alpha)$  geht aus der Funktion  $\cos t$  durch eine Streckung in der Senkrechten mit dem Faktor  $A$ , und eine Verschiebung um den Wert  $\alpha$  nach rechts und eine Stauchung bzw. Dehnung der Periode  $2\pi$  auf die Periode  $\frac{2\pi}{\omega}$  hervor.

**Bemerkung:** Schwingungen bilden ein grundlegendes Phänomen der Realität. Sie begegnen uns in der erfahrbaren Alltagswelt als die Schwingungen der Wasserteilchen, die sich zu den Wellen eines Sees zusammenfügen, als die Bewegung eines Baumes im Wind oder gar als die zerstörerische Bewegung der Erdkruste bei einem Beben. Die sichtbare Welt wird sichtbar mittels der Lichtwellen, die von den Gegenständen ausgehen. Schließlich die Musik, die durch Schwingungen der Luftteilchen vermittelt wird. Aber auch die unsichtbare Welt der kleinsten Teile der Materie besteht geradezu aus Schwingungen. Nach den neuesten Theorien nämlich sind diese kleinsten Teilchen nichts anderes als schwingende Schleifen (strings). Der oben dargestellte komplexe Weg zur Lösung der Schwingungsgleichungen führt die Schwingungen auf die Überlagerung von Kreisbewegungen zurück. Die Kreisbewegung aber ist, nach den antiken Philosophen (Platon, Aristoteles) die vollkommenste aller Bewegungen. ▲

**Der direkte Lösungsweg**

Man rechnet leicht nach, dass die Funktionen

$$x_1(t) = \cos \omega t$$

und

$$x_2(t) = \sin \omega t$$

Lösungen der Schwingungsgleichung sind und man kann wiederum zeigen, dass jede Lösung  $x(t)$  der Schwingungsgleichung als Linearkombination

$$x(t) = c_1 x_1(t) + c_2 x_2(t),$$

das heißt

$$x(t) = c_1 \cos \omega t + c_2 \sin \omega t$$

mit Koeffizienten  $c_1, c_2 \in \mathbb{R}$  dieser speziellen Lösungen geschrieben werden kann. Die Koeffizienten  $c_1, c_2$  berechnet man wieder aus den Vorgaben  $x(0)$  und  $\dot{x}(0)$  für die Anfangslage. Mit

$$\dot{x}(t) = -c_1 \omega \sin \omega t + c_2 \omega \cos \omega t$$

erhält man dann

$$x(0) = c_1$$

und

$$\dot{x}(0) = c_2 \omega,$$

also

$$c_2 = \frac{\dot{x}(0)}{\omega}.$$

Die Lösung des Anfangwertproblems lautet somit

$$x(t) = x(0) \cos \omega t + \frac{\dot{x}(0)}{\omega} \sin \omega t.$$

Mit

$$A = \sqrt{x(0)^2 + \left(\frac{\dot{x}(0)}{\omega}\right)^2}$$

kann man dann schreiben

$$x(t) = A \left( \frac{x(0)}{A} \cos \omega t + \frac{\dot{x}(0)}{A\omega} \sin \omega t \right).$$

Wegen der Beziehung

$$\left( \frac{x(0)}{A} \right)^2 + \left( \frac{\dot{x}(0)}{A\omega} \right)^2 = 1$$

kann der Quotient  $x(0)/A$  als der Cosinus und der Quotient  $\dot{x}(0)/A\omega$  als der Sinus eines Winkels  $\alpha$  aufgefaßt werden. Folglich gilt

$$x(t) = A (\cos \alpha \cdot \cos \omega t + \sin \alpha \cdot \sin \omega t)$$

und aufgrund des Additionstheorems der Cosinusfunktion ergibt sich wieder die Lösung

$$x(t) = A \cos(\omega t - \alpha)$$

des Anfangswertproblems, die wir oben auf anderem Wege bereits erhalten hatten.

### Schwingung als Energieaustausch

Multipliziert man die Schwingungsgleichung

$$\ddot{x}(t) + \omega^2 x(t) = 0$$

mit  $2\dot{x}(t)$ , dann er erhält man die Beziehung

$$2\ddot{x}(t) \dot{x}(t) + \omega^2 2x(t) \dot{x}(t) = 0.$$

Das Produkt  $2\dot{x}\ddot{x}$  ist die Ableitung der Funktion  $\dot{x}^2$  und  $2x\dot{x}$  ist die Ableitung der Funktion  $x^2$ . Daher gibt es eine positive Konstante  $E$ , so dass gilt

$$(*) \quad \dot{x}(t)^2 + \omega^2 x(t)^2 = E.$$

Der Ausdruck  $\omega^2 x(t)^2$  entspricht der potentiellen Energie  $E_{\text{pot}}(t)$  des sich bewegenden Massenpunktes (Energie der Lage) zum Zeitpunkt  $t$  und  $\dot{x}(t)^2$  ist die kinetische Energie  $E_{\text{kin}}(t)$  (Energie der Bewegung) des Massenpunktes. Es gilt also

$$E_{\text{kin}}(t) + E_{\text{pot}}(t) = E$$

und das bedeutet, die Summe beider Energieformen ist zu jedem Zeitpunkt diesselbe. Der Schwingungsvorgang kann daher als ständiger Austausch dieser beiden Energieformen verstanden werden.

Die Gesamtenergie  $E$  der schwingenden Masse ergibt sich aus den Anfangswerten  $x(0)$  und  $\dot{x}(0)$ , indem man bei (\*)  $t = 0$  setzt. Man erhält

$$E = \dot{x}(0)^2 + \omega^2 x(0)^2.$$

Je größer Anfangsauslenkung und Anfangsgeschwindigkeit sind, umso größer ist demnach die Gesamtenergie des schwingenden Massenpunktes.

## 9.5 Räuber–Beute–Systeme

Als Beispiel eines Differentialgleichungssystems betrachten wir ein System, das die Wechselwirkung zwischen einer Räuberpopulation (zum Beispiel Füchse) und einer Beutepopulation (zum Beispiel Hasen) und deren Einfluß auf das Wachstum der beteiligten Populationen beschreibt.

Die Art der Wechselwirkung zwischen Räuber– und Beutepopulation ist qualitativ gesehen die folgende: Gibt es viele Räuber, dann nimmt die Beutepopulation ab, gibt es wenig Räuber, dann nimmt die Beutepopulation zu. Gibt es umgekehrt reichlich Beute, so wächst die Räuberpopulation und bei wenig Beute nimmt sie ab.

Mit  $X$  bezeichnen wir nun die Beutepopulation und mit  $Y$  die Räuberpopulation. Entsprechend sind  $x(t)$  und  $y(t)$  die jeweiligen Bestandsgrößen zum Zeitpunkt  $t$ .

Bei konstanter Geburten– und Todesrate  $a, b > 0$  beschreibt die Differentialgleichung

$$\dot{x}(t) = (a - b) x(t)$$

das Wachstum der Beutepopulation. Bei Anwesenheit von Räubern wird die Todesrate der Beutepopulation aber nicht konstant, sondern proportional zur Größe  $y(t)$  der Räuberpopulation sein. Man wird daher in obiger Differentialgleichung die Konstante  $b$  durch den Term  $b \cdot y(t)$  ersetzen und erhält die Wachstumsgleichung

$$\dot{x}(t) = (a - b y(t)) x(t)$$

für die Beutepopulation. Bei Abwesenheit von Beutetieren sei

$$\dot{y}(t) = (c - d) y(t)$$

die Wachstumsgleichung der Räuberpopulation. Bei Anwesenheit von Beute wird hingegen die Geburtenrate  $c$  proportional zur Größe  $x(t)$  der Beutepopulation sein. Diese Annahme führt auf die korrigierte Wachstumsgleichung

$$\dot{y}(t) = (c x(t) - d) y(t)$$

für die Räuberpopulation. Wachstum und Wechselwirkung der Populationen  $X$  und  $Y$  werden also durch das Differentialgleichungssystem

$$\dot{x}(t) = (a - b y(t)) x(t),$$

(RBS)

$$\dot{y}(t) = (c x(t) - d) y(t)$$

mit Konstanten  $a, b, c, d > 0$  beschrieben.

**Bemerkung:** Die Gleichungen (RBS) werden auch als Lotka–Volterra–Gleichungen bezeichnet. V. Volterra entwickelte um 1920 ein mathematisches Modell zu Erklärung

der Fluktuationen der Fisch-Population in der Adria. Ein ähnliches Modell entwarf A.J. Lotka 1925 zur Beschreibung der Wechselwirkung von Beutefisch- und Haifischpopulationen. ▲

## Kurven in der Ebene

Ein Punkt  $P$ , der sich in der Ebene bewegt, beschreibt dort eine *Kurve*. Kennt man zu jedem Zeitpunkt  $t$  die Positionen  $x(t)$  und  $y(t)$  des Punktes in bezug auf die  $x$ - und  $y$ -Achse eines cartesischen Koordinatensystems, dann ist die Bewegung des Punktes festgelegt. Läuft der Zeitparameter  $t$  im Intervall  $[a, b]$  von  $a$  nach  $b$ , dann bewegt sich der Punkt  $P$  auf einer Kurve mit dem Anfangspunkt  $A = (x(a), y(a))$  und dem Endpunkt  $B = (x(b), y(b))$ . Die Kurvenpunkte haben dann den Ortsvektor  $\mathbf{x}(t) = (x(t), y(t))$ .

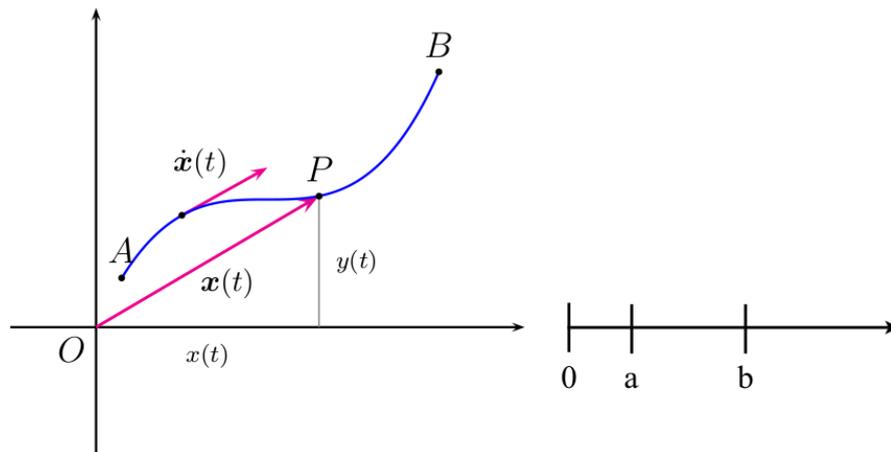


Abbildung 9.6: Bewegung eines Punktes in der Ebene.

Der Punkt  $P$  bewegt sich dabei mit der Geschwindigkeit  $\dot{x}(t)$  in  $x$ -Richtung und mit der Geschwindigkeit  $\dot{y}(t)$  in  $y$ -Richtung. Der Vektor  $\dot{\mathbf{x}}(t) = (\dot{x}(t), \dot{y}(t))$  charakterisiert daher die Geschwindigkeit, mit der sich der Punkt auf der Kurve bewegt. Der Geschwindigkeitsvektor  $\dot{\mathbf{x}}(t)$  ist dabei immer tangential zur Kurve.

## Lösung des Differentialgleichungssystems

Mit  $x_s = d/c$  und  $y_s = a/b$  geht das Differentialgleichungssystem (RBS) über in

$$\dot{x} = b(y_s - y)x$$

(RBS')

$$\dot{y} = c(x - x_s)y.$$

Es definiert für jeden Punkt  $P = (x, y)$  der Ebene einen Geschwindigkeitsvektor  $\dot{\mathbf{x}} = (\dot{x}, \dot{y})$ . Dadurch wird in der Ebene ein *Geschwindigkeitsfeld* festgelegt. Das heißt,

jedem Punkt der Ebene wird ein Geschwindigkeitsvektor angeheftet. Aufgrund der Problemstellung interessieren uns nur Werte  $x, y \geq 0$ , das heißt, wir haben das Bild des Geschwindigkeitsfeldes nur im ersten Quadranten zu untersuchen. Das obige Differentialgleichungssystem zu lösen bedeutet dann, im ersten Quadranten eine Kurve zu bestimmen, die diesem Geschwindigkeitsfeld angepaßt ist.

Um uns ein Bild vom Verlauf des Geschwindigkeitsfeldes zu verschaffen, untersuchen wir zunächst, an welchen Punkten  $\dot{x} = 0$  und  $\dot{y} = 0$  ist. Dem System (RBS') entnimmt man: Es gilt  $\dot{x} = 0$ , wenn entweder  $x = 0$  ist oder  $y = y_s$  ist und es gilt  $\dot{y} = 0$ , wenn  $y = 0$  oder  $x = x_s$  ist. Auf der  $y$ -Achse ( $x = 0$ ) gilt  $\dot{y} = -c x_s y < 0$  und auf der  $x$ -Achse ( $y = 0$ ) ist  $\dot{x} = b y_s x > 0$ . Die Geraden  $x = x_s$  und  $y = y_s$  schneiden sich im Punkte  $S = (x_s, y_s)$  und teilen den ersten Quadranten in vier Sektoren ein, in denen das Geschwindigkeitsfeld nun zu untersuchen ist. Auf der Geraden  $x = x_s$  ist  $\dot{y} = 0$  und  $\dot{x} > 0$  unterhalb von  $S$  und  $\dot{x} < 0$  oberhalb von  $S$ . Auf der Geraden  $y = y_s$  ist  $\dot{y} < 0$  links von  $S$  und  $\dot{y} > 0$  rechts von  $S$ . Im Schnittpunkt  $S$  der beiden Geraden gilt  $\dot{x} = \dot{y} = 0$ . Hier, im „Auge des Orkans“, findet also keine Bewegung statt.

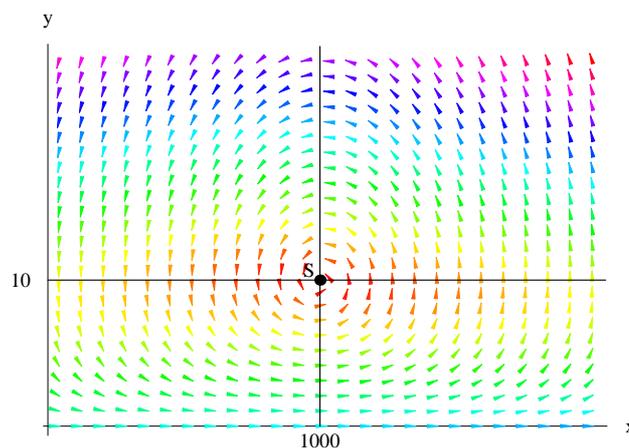


Abbildung 9.7: Das Geschwindigkeitsfeld zum Räuber-Beute-System.

Im Sektor links unten ( $0 < x < x_s, 0 < y < y_s$ ) gilt  $\dot{x} > 0$  und  $\dot{y} < 0$ . Im Sektor rechts unten ( $x > x_s, 0 < y < y_s$ ) ist  $\dot{x} > 0$  und  $\dot{y} > 0$ . Im Sektor oben rechts ( $x > x_s, y > y_s$ ) ist  $\dot{x} < 0$  und  $\dot{y} > 0$  und schließlich gilt im Sektor links oben ( $0 < x < x_s, y > y_s$ ) dass  $\dot{x} < 0$  und  $\dot{y} < 0$  ist. Damit haben wir ein Bild von der „Strömung“ um den Punkt  $S$ , einem *stationären* Punkt des Systems, gewonnen.

### Implizite Lösung

Das Differentialgleichungssystem zu lösen bedeutet nun, die Bahn eines Punktes zu beschreiben, der von der Strömung, die den Punkt  $S$  umkreist, mitgeführt wird. Es wird sich zeigen, dass die Gleichung dieser Bahn nur in impliziter Form gegeben werden kann.

Dividiert man die zweite Gleichung des Systems (RBS') durch die erste, dann lässt sich der Zeitparameter eliminieren und man erhält wegen

$$\frac{\dot{y}}{\dot{x}} = \frac{dy}{dt} : \frac{dx}{dt} = \frac{dy}{dx} = y'$$

eine Differentialgleichung für die Funktion  $y = y(x)$ , welche die durchlaufene Bahn unter Wegfall des dynamischen Aspektes lediglich als Punktmenge beschreibt. Man findet die Beziehung

$$y'(x) = \gamma \cdot \frac{(x - x_s) y(x)}{(y_s - y(x)) x}, \quad \gamma = \frac{c}{b},$$

die man zur Gleichung

$$y'(x) \left( \frac{y_s}{y(x)} - 1 \right) = \gamma \left( 1 - \frac{x_s}{x} \right)$$

oder

$$y_s \cdot \frac{y'(x)}{y(x)} - y'(x) = \gamma \left( 1 - \frac{x_s}{x} \right)$$

umformt. Eine Stammfunktion der linken Seite ist die Funktion

$$L(x) = y_s \ln y(x) - y(x)$$

und eine Stammfunktion der rechten Seite ist die Funktion

$$R(x) = \gamma(x - x_s \ln x)$$

Daher besteht eine Beziehung der Form  $L(x) = R(x) + C$  mit einer Konstanten  $C$ . Man erhält so die Gleichung

$$(*) \quad y_s \ln y(x) - y(x) = \gamma(x - x_s \ln x) + C,$$

die von der gesuchten Funktion erfüllt werden muss. Diese Gleichung lässt sich nicht nach  $y(x)$  auflösen. Sie bestimmt daher nur auf implizite Weise die Funktion  $y(x)$ .

Die Konstante  $C$  kann durch Vorgabe eines Punktes  $(x_0, y_0)$ , den die Bahn durchlaufen soll, festgelegt werden. Durch Einsetzen der Werte  $x_0, y_0$  in die Gleichung (\*) erhält man so den Zusammenhang

$$y_s \ln y_0 - y_0 = \gamma(x_0 - x_s \ln x_0) + C,$$

also ist

$$C = y_s \ln y_0 - y_0 - \gamma(x_0 - x_s \ln x_0)$$

Damit geht (\*) über in die Beziehung

$$y_s \ln \frac{y(x)}{y_0} - (y(x) - y_0) = \gamma \left( x - x_0 - x_s \ln \frac{x}{x_0} \right).$$

### Die Form der Lösungskurve

Mit der Funktion

$$F(x, y) = y_s \ln y - y - \gamma(x - x_s \ln x)$$

lässt sich (\*) auch in der Form

$$F(x, y) = C$$

schreiben. Die Funktion  $F(x, y)$  ordnet jedem Punkt  $(x, y)$  einen Wert  $z = F(x, y)$ , eine „Höhe“ über diesem Punkt, zu. So beschreibt die Funktion  $F(x, y)$  ein „Gebirge“ über dem 1. Quadranten  $x, y \geq 0$ . Die Form dieses Gebirges erschließt sich, wenn man  $F(x, y)$  einmal als Funktion von  $x$  allein betrachtet, indem man  $y$  als fest gewählt auffaßt, und sodann als Funktion von  $y$  allein, indem  $x$  als fest denkt. Man erhält auf diese Weise „Schnitte“ durch das Gebirge parallel zu  $x$ -Achse ( $y$  fest) und parallel zur  $y$ -Achse ( $x$  fest). Dann geben die Ableitungen  $\frac{d}{dx} F(x, y)$  und  $\frac{d}{dy} F(x, y)$  Auskunft über die Form dieser Schnitte. Man erhält

$$\frac{d}{dx} F(x, y) = \gamma \left( \frac{x_s}{x} - 1 \right)$$

und

$$\frac{d}{dy} F(x, y) = \frac{y_s}{y} - 1.$$

Für kleine positive  $x$  hat die Ableitung von  $F$  nach  $x$  große positive Werte, auf der Geraden  $x = x_s$  ist sie gleich null und für  $x > x_s$  wird sie negativ. Für kleine positive  $y$  hat die Ableitung von  $F(x, y)$  nach  $y$  große positive Werte, für  $y = y_s$  ist sie gleich null und für  $y > y_s$  negativ.

Die Funktion  $z = F(x, y)$  beschreibt daher einen „Berg“ über dem 1. Quadranten, dessen „Gipfel“ über dem stationären Punkt  $S = (x_s, y_s)$  liegt, da dort beide Ableitungen von  $F(x, y)$  gleichzeitig null sind. Die Gleichung  $F(x, y) = C$  bedeutet also, durch diesen Berg einen Schnitt in Höhe  $C$  parallel zur  $(x, y)$ -Ebene zu legen. Die zugehörige Schnittkurve ist daher eine geschlossene Kurve, die den Punkt  $S$  umläuft, wie es das Richtungsfeld bereits vermuten ließ. Die Lösung  $(x(t), y(t))$  des Räuber-Beute-Systems beschreibt als die Bewegung eines Punktes in der  $(x, y)$ -Ebene, der den Punkt  $S$ , den stationären Punkt, in einer geschlossenen Kurve umläuft. Sind  $x_{\min}, x_{\max}$  und  $y_{\min}, y_{\max}$  die Minimal- bzw. Maximalwerte der Beute- und der Räuberpopulation, dann folgt die Bewegung des Punktes dem Diagramm  $y_{\min} \rightarrow x_{\max} \rightarrow y_{\max} \rightarrow x_{\min} \rightarrow y_{\min} \rightarrow \dots$ , was biologisch auch einleuchtet.

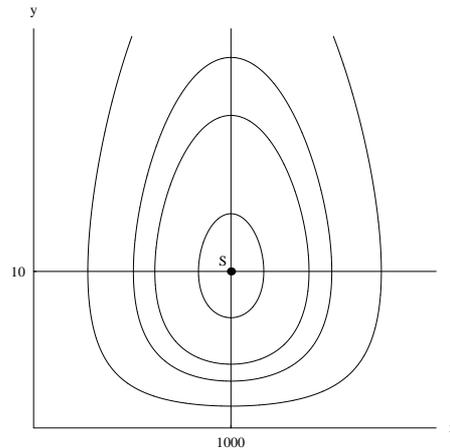


Abbildung 9.8: Der Verlauf der Lösungskurven des Räuber–Beute–Systems.

### Linearisierung

Die Form der Lösungskurven des Räuber–Beute–Systems lässt sich auch durch *Linearisierung* des Systems (RBS) erkennen. Hierzu paßt man zunächst das Koordinatensystem der vorliegenden dynamischen Situation an, indem man das ursprüngliche  $(x, y)$ –System parallel verschiebt, so dass der Punkt  $S = (x_s, y_s)$  der neue Nullpunkt wird. Dann bestehen zwischen den alten Koordinaten  $(x, y)$  und den neuen Koordinaten  $(u, v)$  die Beziehungen

$$x = x_s + u, \quad y = y_s + v.$$

Einsetzen dieser Beziehungen in das System (RBS') ergibt mit  $\dot{x} = \dot{u}$  und  $\dot{y} = \dot{v}$  das System

$$\begin{aligned} \dot{u} &= b(-v)(x_s + u) = -b x_s v - buv, \\ \dot{v} &= cu(y_s + v) = c y_s u + cuv. \end{aligned}$$

In der Nähe des Punktes  $S$  sind  $u, v$  klein und daher wird das Produkt  $u \cdot v$  vernachlässigbar klein. Man streicht daher die Terme mit  $uv$  ohne einen großen Fehler zu machen. Auf diese Weise erhält man das linearisierte System (es treten nur erste Potenzen von  $u$  und  $v$  und auch keine Produkte oder Quotienten dieser Größe auf):

$$\begin{aligned} \dot{u} &= -b x_s v, \\ \dot{v} &= c y_s u. \end{aligned} \quad (\text{RBS}^*)$$

Das linearisierte Räuber–Beute–System (RBS\*) lässt sich explizit lösen. Bildet man nämlich die Ableitung der ersten Gleichung dann erhält man

$$\ddot{u} = -b x_s \dot{v}$$

und durch Einsetzen der zweiten Gleichung in die rechte Seite ergibt sich

$$\ddot{u} = -bc x_s y_s u$$

und somit die Beziehung

$$\ddot{u} + bc x_s y_s u = 0.$$

Dies ist eine Schwingungsgleichung mit der Kreisfrequenz

$$\omega = \sqrt{bc x_s y_s},$$

deren Lösung wir in Abschnitt 9.4 behandelt haben. Man erhält die Lösung

$$(*) \quad u(t) = A \cos(\omega t - \alpha)$$

mit der Amplitude

$$A = \sqrt{u(0)^2 + \frac{\dot{u}(0)^2}{\omega^2}}$$

und dem Phasenwinkel  $\alpha$ , der durch die Beziehung

$$\cos \alpha = \frac{u(0)}{A}$$

gegeben wird. Wegen  $\dot{u}(0) = -b x_s v(0)$  und mit  $\omega^2 = bc x_s v_s$  erhält man

$$\begin{aligned} A &= \sqrt{u(0)^2 + \frac{b^2 x_s^2 v(0)^2}{bc x_s y_s}} \\ &= \sqrt{\frac{c y_s u(0)^2 + b x_s v(0)^2}{c y_s}}. \end{aligned}$$

Damit sind Amplitude und Phase der Schwingung  $u(t)$  durch die Anfangswerte  $u(0) = x(0) - x_s$  und  $v(0) = y(0) - x_s$  und die stationäre Lösung  $x(t) = x_s$  und  $y(t) = y_s$  bestimmt.

Für die zweite Lösungsfunktion  $v(t)$  erhält man aufgrund der Beziehung

$$v = -\frac{\dot{u}}{b x_s}$$

und mit (\*) und  $\omega = \sqrt{bc x_s y_s}$  unmittelbar

$$\begin{aligned} v(t) &= \frac{\omega}{b x_s} A \sin(\omega t - \alpha) \\ &= \sqrt{\frac{c y_s}{b x_s}} A \sin(\omega t - \alpha) \\ &= A' \sin(\omega t - \alpha). \end{aligned}$$

Für die zu  $v(t)$  gehörige Amplitude  $A'$  ergibt sich

$$A' = \sqrt{\frac{c y_s u(0)^2 + b x_s v(0)^2}{b x_s}}.$$

Damit haben wir Näherungslösungen

$$\begin{aligned} x(t) &= x_s + u(t), \\ y(t) &= y_s + v(t) \end{aligned}$$

der Lotka–Volterra–Gleichungen in einer Umgebung des stationären Punktes  $S = (x_s, y_s)$  gewonnen. Der Punkt  $(x(t), y(t))$  umläuft den Punkt  $S$  in einer geschlossenen Kurve in der Zeit  $2\pi/\omega$ , der Periode der Schwingungen  $u(t)$  und  $v(t)$ .

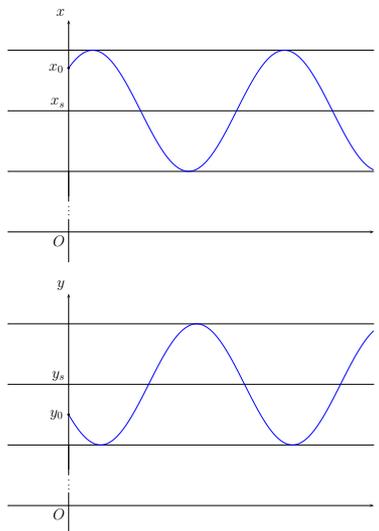


Abbildung 9.9: Näherungslösungen  $x(t)$  und  $y(t)$  des Räuber–Beute–Systems.

**Bemerkung:** Dass die Lösungen des Systems (RBS') geschlossene Kurven um den Punkt  $(x_s, y_s)$  sind, ist auch mit Hilfe des linearen Systems

$$\begin{aligned} \dot{u} &= -b x_s v, \\ \dot{v} &= c y_s u \end{aligned} \quad (\text{RBS}^*)$$

leicht einzusehen. Multipliziert man nämlich die erste Zeile mit  $2c y_s u$ , die zweite mit  $2b x_s v$  und addiert beide Zeilen, dann erhält man die Beziehung

$$2c y_s u \dot{u} + 2b x_s v \dot{v} = 0$$

oder, wegen  $\frac{d}{dt}u^2 = 2ui$  und  $\frac{d}{dt}v^2 = 2v\dot{v}$ ,

$$cy_s \frac{d}{dt} u^2 + bx_s \frac{d}{dt} v^2 = 0.$$

Geht man auf beiden Seiten zu Stammfunktionen über, dann ergibt sich mit einer Konstanten  $C > 0$  die Gleichung

$$(*) \quad cy_s u^2 + bx_s v^2 = C.$$

Die Gleichung einer Ellipse in der  $u, v$ -Ebene mit dem Nullpunkt als Mittelpunkt und den Halbachsen  $a, b$  lautet

$$\frac{u^2}{a^2} + \frac{v^2}{b^2} = 1.$$

Also ist (\*) die Gleichung einer solchen Ellipse mit den Halbachsen

$$a = \sqrt{\frac{C}{cy_s}}, \quad b = \sqrt{\frac{C}{bx_s}}.$$

Die Konstante  $C$  erhält man, indem man  $t = 0$  setzt. Mit  $u(0) = u_0$  und  $v(0) = v_0$  findet man dann

$$C = cy_s u_0^2 + bx_s v_0^2.$$

Folglich ist (\*) die Gleichung einer Ellipse durch den Punkt  $(u_0, v_0)$ .

Da sich aufgrund eines allgemeinen Satzes über Differentialgleichungssysteme die Lösungen des Systems (RBS') und des linearisierten Systems (RBS\*) qualitativ nicht unterscheiden, so handelt es sich bei ersteren um geschlossene Kurven, die den Punkt  $(x_s, y_s)$  umlaufen und die der Form einer Ellipse umso näher kommen je dichter sie bei  $(x_s, y_s)$  liegen. ▲



# Kapitel 10

## Stochastik

Mit dem Begriff *Stochastik* bezeichnet man die Gebiete *Wahrscheinlichkeitstheorie* und *Statistik*. Das Wort Stochastik (stochastike techne) stammt aus dem Griechischen und hat dort etwa die Bedeutung des „kunstvollen Vermutens“. So beschreibt Sokrates in Platons Dialog Philebos (dort 55 d) diese Kunst in Abgrenzung zum Beispiel von der Rechenkunst

... als Abschätzen nach Gutdünken und Einübung der Sinne durch Erfahrung und Gewöhnung, indem man dazunimmt, was nur die glückliche Mutmaßung vermag, welche viele auch eine Kunst nennen, die durch Anstrengung und Sorgfalt ihre Stärke erreicht.

Inzwischen hat sich die Stochastik nach tastenden Anfängen im 15. und 16. Jahrhundert zu einer handfesten mathematischen Disziplin entwickelt. Ein Hauptantrieb dieser Entwicklung war dabei zunächst das Bedürfnis, Gewinnchancen beim Glücksspiel zu berechnen. Ihr Teilgebiet Statistik befasst sich mit der methodisch abgesicherten Erhebung und Auswertung von Daten. Die Wahrscheinlichkeitstheorie stellt dazu den mathematischen Hintergrund bereit.

Die Grundfrage der Wahrscheinlichkeitstheorie lautet: Mit welcher Sicherheit ist das Eintreten eines Ereignisses  $E$  zu erwarten? Der Grad dieser Sicherheit wird durch eine Zahl  $P(E)$ , die *Wahrscheinlichkeit* des Ereignisses  $E$  gemessen („ $P$ “ von lateinisch „probabilitas“, Wahrscheinlichkeit, Glaubhaftigkeit). Die Zahl  $P(E)$  liegt zwischen null und eins. Ist  $P(E) = 0$ , dann kann das Ereignis  $E$  unmöglich eintreten, ist  $P(E) = 1$ , dann wird  $E$  mit Sicherheit eintreten. Die mathematische Methode zur Berechnung solcher Wahrscheinlichkeiten für unterschiedliche Klassen von Ereignissen wird in der Wahrscheinlichkeitstheorie entwickelt.

Gelegentlich liest man Charakterisierungen wie, die Wahrscheinlichkeitstheorie befasse sich mit „zufallsgesteuerten“ Ereignissen, die von den „deterministischen“ Ereignissen abzugrenzen seien. Abgesehen davon, dass der Terminus „zufallsgesteuert“ in sich widersprüchlich ist, muss im Lichte der modernen Physik, die *alle* Naturereignisse als Überlagerung von Zufallsprozessen versteht, die Unterscheidung von zufälligen und

determinierten Ereignissen aufgegeben werden. Dazu ein Beispiel: Dass morgen wieder die Sonne aufgehen wird, kann keineswegs mit Sicherheit, sondern nur mit sehr hoher Wahrscheinlichkeit erwartet werden.

Die folgende Darstellung der Stochastik kann aufgrund ihres beschränkten Umfangs nur versuchen, die Grundgedanken, Grundbegriffe und die wesentlichen Methoden dieser Disziplin zu vermitteln. Dabei wird induktives Vorgehen bevorzugt. Das heisst, die Grundbegriffe und Verfahren der Stochastik werden nicht axiomatisch, sondern ausgehend von Beispielen, eingeführt.

## 10.1 Vorüberlegungen und Beispiele

Die Wahrscheinlichkeit ist eine Eigenschaft, die Ereignissen zugesprochen wird. Man sagt zum Beispiel, „es ist sehr wahrscheinlich, dass es heute regnen wird“. Oder, „es ist äußerst unwahrscheinlich, beim Lotto sechs Richtige zu erzielen“. Aus dem Bedürfnis solch umgangssprachlichen Mutmaßungen einen präzisen Sinn zu verleihen, ist die Wahrscheinlichkeitsrechnung entstanden.

Hierzu muss man zunächst versuchen die mehr oder weniger vagen umgangssprachlichen Formulierungen für die Sicherheit der Erwartung, mit dem ein Ereignis eintreten wird, durch Angabe einer Zahl zu ersetzen, welche die Wahrscheinlichkeit misst, dass das fragliche Ereignis eintritt.

### Relative Häufigkeit

Ein experimentelles Verfahren dazu besteht in der Berechnung seiner *relativen Häufigkeit*. Will man zum Beispiel wissen, mit welcher Wahrscheinlichkeit der 23. September 2005 ein sonniger Tag wird, so kann man die Wetteraufzeichnungen der letzten hundert Jahre durchgehen und auszählen, wieviele sonnige 23. September es in diesem Zeitraum gegeben hat. Zählt man deren 25, dann erhält man die relative Häufigkeit  $25/100 = 1/4$  für dieses Ereignis, bezogen auf die letzten einhundert Jahre. Bezieht man sich auf einen anderen Zeitraum, so wird sich in der Regel auch ein anderer Zahlenwert ergeben. Die experimentell gefundene relative Häufigkeit betrachtet man dann als Näherungswert für die „wahre“ Wahrscheinlichkeit.

Im allgemeinen bestimmt man die relative Häufigkeit eines Ereignisses  $E$ , indem man  $n$  einschlägige Beobachtungen bzw. Experimente macht und abzählt, wie oft das Ereignis  $E$  dabei auftritt. Ist diese Anzahl gleich  $m$ , dann ist

$$(RH) \quad h_n(E) = \frac{m}{n}$$

die relative Häufigkeit des Ereignisses  $E$ , bezogen auf die zugrundeliegende Serie von Beobachtungen bzw. Experimenten.

Die Zahlen  $h_n(E)$  liegen offenbar zwischen null und eins:

$$0 \leq h_n(E) \leq 1.$$

Im Falle  $h_n(E) = 0$  ist  $E$  nie eingetreten, im Falle  $h_n(E) = 1$  immer.

Betrachtet man zwei verschiedene Ereignisse  $E_1$  und  $E_2$ , die in der beobachteten Serie eintreten können und tritt  $E_1$   $m_1$ -mal und  $E_2$   $m_2$ -mal ein, dann ist

$$(*) \quad h_n(E_1 \vee E_2) = \frac{m_1 + m_2}{n} = h_n(E_1) + h_n(E_2)$$

die relative Häufigkeit, mit der das Ereignis  $E = E_1 \vee E_2$  eintritt.

Unbefriedigend an der relativen Häufigkeit als Kenngröße für die Wahrscheinlichkeit ist ihr empirischer Charakter und ihre Uneindeutigkeit, da sie von einer konkreten Beobachtungsreihe abhängt. Man sucht daher nach einer nicht-empirischen Kenngröße als Maß für die Wahrscheinlichkeit eines Ereignisses.

### Laplace-Wahrscheinlichkeit

Diese findet man in der *Laplace-Wahrscheinlichkeit* eines Ereignisses  $E$ . Hat ein Experiment apriori  $n$  mögliche Ergebnisse und wird das Ereignis  $E$  durch  $m$  Ergebnisse realisiert, dann ist

$$(LR) \quad P(E) = \frac{m}{n}$$

die Laplace-Wahrscheinlichkeit von  $E$ , sofern sämtliche Ergebnisse apriori die gleiche Chance haben einzutreten. (Der Terminus „apriori“ ist hier im Sinne von „unabhängig von einem konkreten Experiment“ zu verstehen).

Die obige Regel zur Bestimmung der Wahrscheinlichkeit eines Ereignisses wird nach dem französischen Mathematiker Pierre Simon Marquis de Laplace (1749–1827) als *Laplace-Regel* bezeichnet. Man formuliert sie auch folgendermaßen:

$$P(E) = \frac{\text{Zahl der für } E \text{ günstigen Fälle}}{\text{Zahl der möglichen Fälle}}.$$

**Beispiel 1 (Würfeln):** Bei einem symmetrischen Würfel haben die Zahlen 1 bis 6 die gleiche Chance gewürfelt zu werden. Also sind  $P(1) = P(2) = \dots = P(6) = 1/6$  aufgrund der Laplace-Regel die Wahrscheinlichkeiten der möglichen Ergebnisse dieses Zufallsexperimentes. Fragt man zum Beispiel nach der Wahrscheinlichkeit eine 2 oder eine 3 zu würfeln, dann gibt es 2 günstige Fälle und man erhält

$$P(2 \vee 3) = \frac{2}{6} = \frac{1}{3}.$$

Andererseits gilt auch in Analogie zur Beziehung (\*) für die relative Häufigkeit eines Ereignisses  $E = E_1 \vee E_2$ :

$$P(2 \vee 3) = P(2) + P(3) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Die Wahrscheinlichkeit für das Ereignis  $U$  eine ungerade Zahl zu würfeln ist offenbar

$$P(U) = \frac{3}{6} = \frac{1}{2},$$

da es unter den Zahlen 1 bis 6 die drei ungeraden Zahlen 1, 3, 5 gibt. ■

**Beispiel 2 (Urnenexperiment):** In einer Urne befinden sich 3 rote, 2 grüne und 5 schwarze Kugeln, also insgesamt 10 Kugeln. Wie groß ist bei blindem Zugriff die Wahrscheinlichkeit eine rote, grüne bzw. schwarze Kugel zu ziehen?

Denkt man sich die Kugeln von 1 bis 10 nummeriert, dann sind die möglichen Ergebnisse die Zahlen 1 bis 10. Für das Ereignis „Ziehung einer roten Kugel“ hat man 3 „günstige Fälle“, 2 für eine grüne und 5 für eine schwarze. Aufgrund der Laplace-Regel ergeben sich dann die Wahrscheinlichkeiten

$$P(r) = \frac{3}{10}, \quad P(g) = \frac{2}{10}, \quad P(s) = \frac{5}{10}.$$

Für das Ereignis  $r \vee g$  zum Beispiel ergibt sich

$$P(r \vee g) = \frac{3+2}{10} = P(r) + P(g) = \frac{1}{2}.$$

Die Summe der Einzelwahrscheinlichkeiten ist eins:

$$P(r) + P(g) + P(s) = \frac{3+2+5}{10} = 1.$$

■

**Beispiel 3 (Münzwurf):** Auf den Seiten einer symmetrischen Münze befinden sich die Zahlen 0 und 1. Die Münze werde 10 mal geworfen. Wie groß ist dann die Wahrscheinlichkeit, dass bei diesen 10 Würfen genau 3 mal die 1 oben liegt? Zur Beantwortung dieser Frage überlegt man sich zunächst, wieviele verschiedene 0,1-Folgen der Länge 10 überhaupt möglich sind. Da an jeder Stelle der Folge zwei Besetzungsmöglichkeiten gegeben sind, gibt es  $2 \cdot 2 = 2^{10}$  solcher 0,1-Folgen. Dies ist die Anzahl der möglichen Fälle.

Uns interessiert nun die Anzahl derjenigen Folgen, die genau 3-mal die 1 enthalten. Diese Anzahl entspricht offenbar der Anzahl der Möglichkeiten, aus einer 10-elementigen Menge 3-elementige Teilmengen auszuwählen. Also ist diese Anzahl der „günstigen Fälle“ gleich  $\binom{10}{3}$ . Also gilt aufgrund der Laplace-Regel:

$$P(\text{genau 3-mal 1}) = \frac{\binom{10}{3}}{2^{10}}.$$

Verallgemeinert man die obige Fragestellung und fragt nach der Wahrscheinlichkeit, bei  $n$  Münzwürfen genau  $k$ -mal die 1 zu erzielen, so führt dieselbe Überlegung zu dem Ergebnis, dass

$$P(\text{genau } k\text{-mal die } 1) = \frac{\binom{n}{k}}{2^n}$$

die gesuchte Wahrscheinlichkeit ist. ■

**Beispiel 4 (Wurf mit zwei Würfeln):** Wirft man gleichzeitig zwei Würfel, dann sind die möglichen Ergebnisse die Zahlenpaare  $(x, y)$ , wobei  $x$  und  $y$  unabhängig voneinander die Werte 1 bis 6 annehmen können. Es gibt also  $6 \cdot 6 = 36$  solcher Zahlenpaare. Jedem Zahlenpaar kommt dabei die gleiche Wahrscheinlichkeit  $P(x, y) = 1/36$  zu, gewürfelt zu werden. Wie groß ist dann zum Beispiel die Wahrscheinlichkeit, bei einem Wurf die „Augensumme“  $x + y = 5$  zu erzielen? Diese Augensumme wird realisiert durch die 4 Fälle  $(1,4)$ ,  $(2,3)$ ,  $(3,2)$ ,  $(4,1)$ . Aufgrund der Laplace-Regel ergibt sich daher

$$P(x + y = 5) = \frac{4}{36} = \frac{1}{9}$$

als die gesuchte Wahrscheinlichkeit. ■

**Beispiel 5 (Das Geburtstagsproblem):** Jemand wettet, dass von den in einem Raum versammelten 40 Personen, mindestens zwei am selben Tag Geburtstag haben. Mit welcher Wahrscheinlichkeit gewinnt er diese Wette?

Das Jahr hat 365 Tage. Die Geburtstage der 40 Personen bilden daher ein Zahlentupel  $(x_1, x_2, \dots, x_{40})$  der Länge 40. An jeder Position sind 365 Einträge möglich, also gibt es  $365^{40}$  solcher Zahlentupel. Das ist eine unvorstellbar große Zahl, größer als die Anzahl der Atome im Universum (ca.  $10^{83}$ ). Nimmt man zur Vereinfachung an, dass die Häufigkeit von Geburten nicht von der Jahreszeit abhängt, dann besitzen sämtliche 40-Tupel die gleiche Wahrscheinlichkeit.

Um die Laplace-Regel zur Lösung unserer Frage anwenden zu können, haben wir die Zahl der „günstigen Fälle“ zu bestimmen. Ein solcher günstiger Fall ist ein 40-Tupel, bei dem mindestens 2 gleiche Zahlen auftreten. Einfacher als diese Anzahl zu bestimmen ist die komplementäre Frage nach der Anzahl der 40-Tupel mit lauter verschiedenen Einträgen zu beantworten (vgl. Abschnitt 3.1). Die gesuchte Anzahl ist  $365 \cdot 364 \cdot \dots \cdot 326$ . Insgesamt gibt es  $365^{40}$  40-Tupel, also weisen davon  $365^{40} - 365 \cdot 364 \cdot \dots \cdot 326$  der 40-Tupel Wiederholungen von Zahlen auf. Daher ist die Wahrscheinlichkeit des Ereignisses  $E$  „unter 40 Personen haben mindestens zwei am gleichen Tag

Geburtstag“ gleich

$$\begin{aligned} P(E) &= \frac{365^{40} - 365 \cdot 364 \cdot \dots \cdot 326}{365^{40}} \\ &= 1 - \frac{365 \cdot 364 \cdot \dots \cdot 326}{365^{40}} \\ &= 0,891. \end{aligned}$$

Die Wahrscheinlichkeit, die eingegangene Wette zu gewinnen, ist also mit ca. 90 % sehr hoch. Da die auftretenden Zahlen unvorstellbar groß sind, lässt sich das Ergebnis numerisch kaum abschätzen. Seine Berechnung erfordert den Einsatz des Computers. Bei  $n$  Personen lautet die obige Formel aufgrund völlig analoger Überlegungen

$$P(E) = 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365 - n + 1)}{365^n}.$$

Für einige  $n$  gibt die nachstehende Tabelle die Werte  $P(E)$  an.

n	10	20	30	40	50	60	70
$P(E)$	0,117	0,411	0,706	0,891	0,970	0,994	0,999

Tabelle 10.1: Die Wahrscheinlichkeit  $P(E)$ , dass von  $n$  Personen mindestens zwei am gleichen Tag Geburtstag haben.

Zwischen den Werten  $n = 22$  und  $n = 23$  überspringt  $P(E)$  die 50% Grenze. Ab  $n = 23$  ist also die Chance, die obige Wette zu gewinnen, Größer als 50%. ■

## 10.2 Grundbegriffe und Grundregeln

Die möglichen Ergebnisse eines Experimentes (Würfeln, Münzwurf etc.) fassen wir zusammen in der *Ergebnismenge*  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ . Sind alle Ergebniss  $\omega_j, j = 1, 2, \dots, n$ , *gleichwahrscheinlich*, dann ist aufgrund der *Laplace-Regel*  $P(\omega_j) = \frac{1}{n}$  die Wahrscheinlichkeit für das Eintreten des Ergebnisses  $\omega_j$ .

Eine Teilmenge  $E \subset \Omega$  wird als ein *Ereignis* bezeichnet. In diesem Kontext nennt man die Ergebnisse  $\omega_j$  auch *Elementarereignisse*. Die Wahrscheinlichkeit des Eintretens von  $E$  ist nach der Laplace-Regel in der Sprache der Mengenlehre dann

$$P(E) = \frac{|E|}{|\Omega|},$$

sofern die Elementarereignisse  $\omega_j$  gleichwahrscheinlich sind. Im Allgemeinen werden die Ergebnisse eines Experiments nicht sämtlich die gleiche Wahrscheinlichkeit besitzen (vgl. das Urnenexperiment in Beispiel 2 des vorigen Abschnitts). Die Wahrscheinlichkeiten sind dann über die Elementarereignisse  $\omega_1, \omega_2, \dots, \omega_n \in \Omega$  unterschiedlich verteilt. Man spricht dann von einer *Wahrscheinlichkeitsverteilung* über der Ergebnismenge oder dem *Ergebnisraum*  $\Omega$ .

**Definition:** Eine Wahrscheinlichkeitsverteilung über einem Ergebnisraum  $\Omega$  hat die folgenden Eigenschaften:

- 1.) Die Wahrscheinlichkeit eines jeden Elementarereignisses  $\omega$  ist eine Zahl  $P(\omega)$  mit  $0 \leq P(\omega) \leq 1$ .
- 2.) Die Summe der Wahrscheinlichkeiten aller Elementarereignisse ist 1, das heisst es gilt

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

- 3.) Die Wahrscheinlichkeit eines Ereignisses  $E \subset \Omega$  ist die Summe der Wahrscheinlichkeiten der in  $E$  gelegenen Elementarereignisse  $\omega$ :

$$P(E) = \sum_{\omega \in E} P(\omega)$$

Aufgrund dieser Regeln gilt immer

$$0 \leq P(E) \leq 1$$

und  $P(\Omega) = 1$ . Für zwei Ereignisse  $E_1, E_2 \subset \Omega$  erhält man die Regel:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2).$$

Haben die Mengen  $E_1$  und  $E_2$  einen leeren Durchschnitt,  $E_1 \cap E_2 = \emptyset$ , dann folgt

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(\emptyset).$$

Also muss man formal  $P(\emptyset) = 0$  setzen, da in diesem Fall offenbar

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

sein muss. Ist  $\bar{E}$  die zu  $E$  komplementäre Menge, das heisst gilt  $E \cup \bar{E} = \Omega$  und  $E \cap \bar{E} = \emptyset$ , dann folgt aus obiger Regel und  $P(\Omega) = 1$ ,

$$P(E \cup \bar{E}) = P(E) + P(\bar{E}) = 1.$$

Also gilt

$$P(\bar{E}) = 1 - P(E).$$

Dies ist eine nützliche Regel, denn manchmal ist es einfacher, die Wahrscheinlichkeit des komplementären, als die Wahrscheinlichkeit des eigentlich interessierenden Ereignisses zu berechnen (vgl. Beispiel 5).

### Unabhängigkeit

Wenn die Eigenschaften „Intelligenz“ und „Schönheit“ voneinander unabhängig sind, dann bedeutet dies folgendes: Ist  $I$  die Menge der intelligenten und  $S$  die Menge der schönen Individuen innerhalb einer Menge  $\Omega$  von Individuen, dann ist der Anteil der Intelligenten unter den Schönen genau so groß wie der Anteil der Intelligenten an der Gesamtheit  $\Omega$ . Das bedeutet, es besteht die Beziehung

$$\frac{|I \cap S|}{|S|} = \frac{|I|}{|\Omega|}$$

oder, nach Multiplikation mit  $|S| / |\Omega|$ :

$$\frac{|I \cap S|}{|\Omega|} = \frac{|I|}{|\Omega|} \cdot \frac{|S|}{|\Omega|}$$

Die auftretenden Quotienten lassen sich als Laplace–Wahrscheinlichkeiten interpretieren. Somit findet man die Beziehung

$$P(I \cap S) = P(I) \cdot P(S).$$

Diese heuristische Überlegung gibt Anlaß zu der

**Definition:** Zwei Ereignisse  $A$  und  $B$  heißen stochastisch unabhängig, wenn gilt

$$P(A \cap B) = P(A) \cdot P(B).$$

Diese Regel lässt sich unmittelbar auf mehr als zwei Ereignisse erweitern: Drei Ereignisse  $A, B, C$  sind unabhängig, wenn gilt:

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C),$$

usw.

### Zufallsgrößen und ihre Verteilung

Beim Würfeln mit zwei Würfeln (vgl. Beispiel 4 des vorigen Abschnitts) interessiert man sich weniger für die Ergebnisse  $\omega = (x, y)$ ,  $x, y \in \{1, 2, 3, 4, 5, 6\}$  des Experimentes, als für die Augensumme  $x + y$ , die jedem Ergebnis zugeordnet ist. Eine solche

Zuordnung, die jedem Ergebnis eines Experimentes eine Zahl zuordnet, heisst *Zufallsgröße*.

**Definition:** Ist  $\Omega$  die Ergebnismenge eines Experimentes und  $X$  eine Abbildung  $X : \Omega \rightarrow \mathbb{R}$ , dann bezeichnet man  $X$  als eine *Zufallsgröße*.

In unserem Beispiel gilt also  $X(\omega) = X(x, y) = x + y$ . Eine andere Zufallsgröße bezüglich des Würfels mit zwei Würfeln wäre die Differenz der Augenzahlen:  $Y(\omega) = Y(x, y) = x - y$ . Die Zufallsgröße „Augensumme“ nimmt die Werte  $X(\Omega) = \{2, 3, \dots, 12\}$  an. Die Wahrscheinlichkeiten  $P(X = k) = p_k$ , mit denen die Zufallsgröße  $X$  die Werte  $k \in X(\Omega)$  annimmt, sind in der folgenden Tabelle zusammengestellt. Sie bilden die zu  $X$  gehörige *Wahrscheinlichkeitsverteilung* oder einfach die zu  $X$  gehörige *Verteilung*.

$k$	2	3	4	5	6	7	8	9	10	11	12
$p_k$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tabelle 10.2: Die zur Zufallsgröße „Augensumme“ gehörige Verteilung.

Allgemein gilt die folgende

**Definition:** Nimmt die Zufallsgröße  $X : \Omega \rightarrow \mathbb{R}$  die Werte  $X(\omega_k) = x_k$ ,  $k = 1, 2, \dots, n$ , dann bilden die Wahrscheinlichkeiten  $P(X = x_k) = p_k$  dafür, dass  $X$  den Wert  $x_k$  annimmt, die zu  $X$  gehörige *Wahrscheinlichkeitsverteilung*.

### Erwartungswert, Varianz und Standardabweichung

Um sich ein anschauliches Bild einer Wahrscheinlichkeitsverteilung zu verschaffen, kann man das zugehörige *Histogramm* bilden. Hierzu trägt man über Intervalle  $I_1$  bis  $I_n$  der Länge eins Rechtecke mit der Höhe  $p_1$  bis  $p_n$  ab. Das Rechteck über dem Intervall  $I_k$  hat also den Flächeninhalt  $p_k$ . Die Summe der Rechteckflächen ist daher gleich eins.

Die Lage und Form eines solchen Histogramms wird durch gewisse Kenngrößen charakterisiert. Die erste dieser Kenngrößen ist der *Erwartungswert*  $E(X)$  der Zufallsgröße  $X$ , auch *Mittelwert*  $\mu$  der zugehörigen Verteilung genannt.

**Definition:** Hat eine Zufallsgröße  $X : \Omega \rightarrow \mathbb{R}$  mit den Werten  $X(\omega_k) = x_k$  die Verteilung  $P(X = x_k) = p_k$ ,  $k = 1, 2, \dots, n$ , dann ist die Zahl

$$E(X) = \sum_{k=1}^n p_k \cdot x_k$$

der Erwartungswert der Zufallsgröße  $X$ .

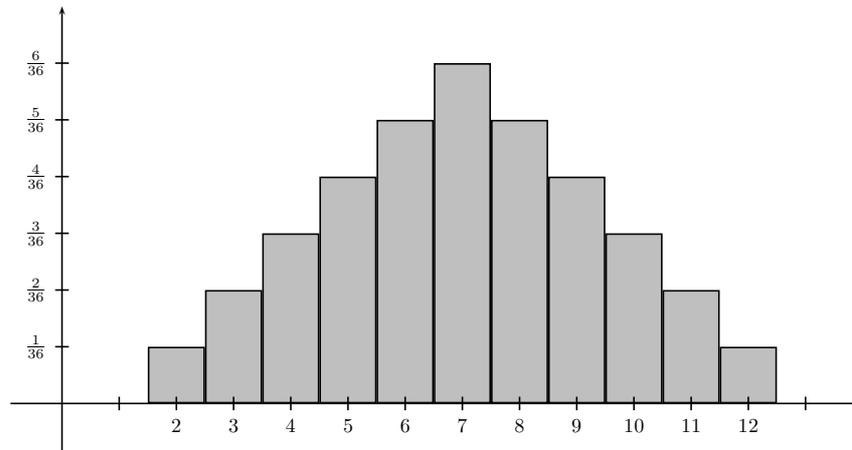


Abbildung 10.1: Histogramm der Wahrscheinlichkeitsverteilung der Zufallsgröße „Augensumme“ beim Würfeln mit zwei Würfeln.

Der Erwartungswert ist physikalisch gesehen der *Schwerpunkt* des Systems von Massenpunkten an den Stellen  $x_k$ , die mit den Massen  $p_k$  belegt sind.

Der Schwerpunkt der Verteilung der Zufallsgröße Augensumme muss aus Symmetriegründen an der Stelle  $k = 7$  liegen. Man rechnet nach:

$$\begin{aligned}
 E(x) &= \frac{1}{36} (1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + 4 \cdot 5 + 5 \cdot 6 + 6 \cdot 7 + 5 \cdot 8 + 4 \cdot 9 + 3 \cdot 10 + 2 \cdot 11 + 1 \cdot 12) \\
 &= \frac{1}{36} (1 \cdot (2 + 12) + 2(3 + 11) + 3 \cdot (4 + 10) + 4(5 + 9) + 5 \cdot (6 + 8) + 6 \cdot 7) \\
 &= \frac{1}{36} (14(1 + 2 + 3 + 4 + 5) + 6 \cdot 7) \\
 &= \frac{1}{36} (14 \cdot \frac{1}{2} \cdot 5 \cdot 6 + 6 \cdot 7) \\
 &= \frac{1}{36} (7 \cdot 5 \cdot 6 + 6 \cdot 7) \\
 &= \frac{42 \cdot 6}{36} = 7.
 \end{aligned}$$

Diese Rechnung liefert im übrigen ein schönes Beispiel dafür, dass Nachdenken gelegentlich Arbeit erspart.

Während der Wert  $\mu = E(X)$  Auskunft darüber gibt, um welchen Wert sich die größeren Wahrscheinlichkeiten  $p_k$  konzentrieren, sofern das zugehörige Histogramm wie in unserem Beispiel einen Gipfel aufweist, informiert die Kenngröße *Varianz*, geschrieben  $Var(X)$ , bzw. die *Standardabweichung*, das ist die Quadratwurzel aus der Varianz, über die Form des Histogramms. Ja nachdem, ob sich die Rechtecksflächen um den Mittelwert  $\mu$  eng konzentrieren oder weit um  $\mu$  verteilt sind, ist die Varianz bzw. Standardabweichung groß oder klein.

**Definition:** Unter der Varianz einer Zufallsgröße  $X : \Omega \rightarrow \mathbb{R}$  mit den Werten  $X(\omega_k) = x_k$  und der Verteilung  $P(X = x_k) = p_k$  versteht man die Zahl

$$\text{Var}(X) = \sum_{k=1}^n (x_k - \mu)^2 \cdot p_k.$$

Die Varianz als die mit den „Gewichten“  $p_k$  gemittelte Summe der quadratischen Abstände der Punkte  $x_k$  vom Mittelwert  $\mu$  ist offenbar immer eine positive Zahl. Konzentrieren sich die Werte  $x_k$  um den Mittelwert  $\mu$ , dann ist diese Zahl klein, liegen sie von  $\mu$  weiter weg, dann ist die Zahl groß.

Während die Varianz einer Zufallsgröße keine unmittelbare anschauliche Interpretation besitzt, trifft dies gleichwohl für die Wurzel aus der Varianz, die *Standardabweichung*  $\sigma(x)$  zu. Die Zahl

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

oder einfach  $\sigma$ , mißt gleichsam die „Breite“ des zu einer Verteilung gehörigen Histogramms. Diese ist in etwa gleich  $2\sigma$ .

**Beispiel:** Die Varianz der Zufallsgröße „Augensumme“ ist offenbar

$$\begin{aligned} \text{Var}(X) &= \frac{2}{36} ((2-7)^2 \cdot 1 + (3-7)^2 \cdot 2 \\ &\quad + (4-7)^2 \cdot 3 + (5-7)^2 \cdot 4 + (6-7)^2 \cdot 5 + (7-7)^2 \cdot 6) \\ &= \frac{1}{18} (25 \cdot 1 + 16 \cdot 2 + 9 \cdot 3 + 4 \cdot 4 + 1 \cdot 5) \\ &= \frac{105}{18} = 5,83 \end{aligned}$$

Man erhält daraus die Standardabweichung

$$\sigma(X) = \sqrt{5,83} = 2,41.$$

Mit ihrer Hilfe mißt man die „Breite“ eines Histogramms. Was damit gemeint ist, soll am Histogramm aus Abb. 10.1 gezeigt werden.

Man rechnet aus, dass im Bereich  $\mu - \sigma \leq x \leq \mu + \sigma$ , also über dem Bereich zwischen den Grenzen  $7-2,4$  und  $7+2,4$ , das heißt über den Werten 5, 6, 7, 8, 9, die Massen

$$\frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} = \frac{1}{36} \cdot 24 = \frac{2}{3},$$

also ca. 67% der Gesamtmasse 1 verteilt sind.

In Bereich  $\mu - 2\sigma \leq x \leq \mu + 2\sigma$ , dem sogenannten  $2\sigma$ -Bereich, das sind in unserem Beispiel die Werte 3, 4, 5, 6, 7, 8, 9, 10, 11, liegt die Massensumme  $34/36 = 0,944$  das sind ca. 94% der Gesamtmasse 1. ■

### 10.3 Die Binomialverteilung

Man werfe eine nicht-symmetrische Münze, bei der mit der Wahrscheinlichkeit  $p$  die 1 oben zu liegen kommt und mit Wahrscheinlichkeit  $q = 1 - p$  die 0. Die Münze werde 5-mal geworfen. Mit welcher Wahrscheinlichkeit erzielt man 3-mal eine 1? Die möglichen Ergebnisse dieses Experimentes sind sämtliche 5-er Kombinationen der Zahlen 0 und 1 wie zum Beispiel  $(0,1,1,0,1)$ . Davon gibt es offenbar  $2^5$ . Also ist

$$\Omega = \{(x_1, x_2, x_3, x_4, x_5) \mid x_j \in \{0, 1\}, j = 1, 2, 3, 4, 5\}$$

unser *Ergebnisraum*. Enthält ein solches Ergebnis  $k$ -mal eine 1 und  $(5 - k)$ -mal eine 0, dann ist seine Wahrscheinlichkeit aufgrund der Regel  $P(A \cap B) = P(A) \cdot P(B)$  zur Berechnung der Wahrscheinlichkeit unabhängiger Ereignisse (vgl. Abschnitt 10.2) gleich  $p^k \cdot q^{5-k}$ . Also gilt zum Beispiel  $P(0, 1, 1, 0, 1) = p^3 \cdot q^2$ .

Wir betrachten nun die Zufallsgröße  $X =$  „Anzahl der Einsen in einer Sequenz von fünf Münzwürfen“.  $X$  nimmt die Werte  $k = 0, 1, 2, 3, 4, 5$  an. Die obige Frage lautet jetzt, die Wahrscheinlichkeit des Ereignisses  $X = 3$  zu berechnen, also  $P(X = 3)$ . Dieses Ereignis wird realisiert durch die 5-Tupel  $(1, 0, 1, 1, 0)$ ,  $(1, 1, 0, 0, 1)$ ,  $(0, 0, 1, 1, 1)$  usw. Jedes dieser 5-Tupel als Einzelnes tritt mit der Wahrscheinlichkeit  $p^3 \cdot q^2$  ein. Die Anzahl dieser 5-Tupel ist gleich der Anzahl der Möglichkeiten, aus einer Menge mit 5 Elementen 3-elementige Teilmengen auszuwählen, also gleich  $\binom{5}{3}$ . Damit erhalten wir wegen Eigenschaft 3 von Wahrscheinlichkeitsverteilungen (vgl. Abschnitt 10.2) das Ergebnis

$$P(X = 3) = \binom{5}{3} p^3 \cdot q^2.$$

Die Wahrscheinlichkeit allgemein  $k$ -mal die 1 bei 5 Würfeln zu erzielen ist dann

$$P(X = k) = \binom{5}{k} p^k q^{5-k}.$$

Diese Werte bestimmen zugleich die Wahrscheinlichkeitsverteilung der Zufallsgröße  $X$ .

Wirft man die unsymmetrische Münze allgemein  $n$ -mal, dann ist aufgrund derselben Überlegungen

$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$

die Wahrscheinlichkeit dabei  $k$ -mal eine 1 zu erzielen.

**Definition:** Eine Zufallsgröße  $X$  heisst binomialverteilt mit den Parametern  $n$  und  $p$ , wenn sie die Werte  $k = 0, 1, 2, \dots, n$  annimmt und wenn

$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$

die Wahrscheinlichkeit des Ereignisses  $X = k$  ist.

Man spricht in diesem Zusammenhang auch von  $B(n, p)$ -verteilten Zufallsgrößen. Den obigen Ausdruck für die Wahrscheinlichkeit  $P(X = k)$  bezeichnet man auch mit  $B(n, p, k)$

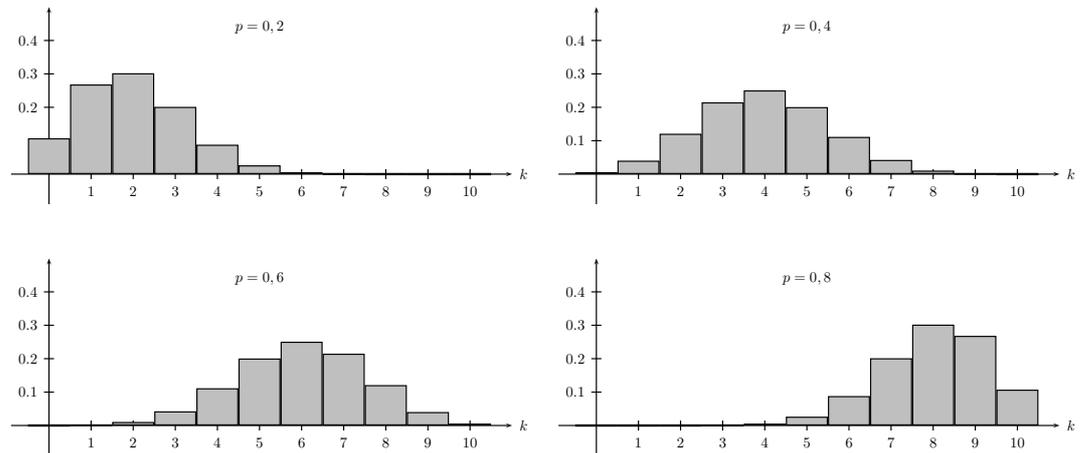


Abbildung 10.2: Einige Histogramme von  $B(n, p)$ -Verteilungen.

**Beispiel:** Neben dem Münzwurf mit einer symmetrischen ( $p = q = \frac{1}{2}$ ), siehe Beispiel 3, Abschnitt 10.1, oder unsymmetrischen Münze treten binomialverteilte Zufallsgrößen in vielen anderen Zusammenhängen, so zum Beispiel bei Urnenexperimenten auf. Enthält eine Urne 2 rote und 3 grüne Kugeln, dann ist  $p = 2/5$  die Wahrscheinlichkeit eine rote und  $q = 3/5$  die Wahrscheinlichkeit dafür, eine grüne Kugel zu ziehen. Zählt die Zufallsgröße  $X$  die Anzahl der roten Kugeln bei  $n$ -Ziehungen mit Zurücklegen, dann ist

$$P(X = k) = \binom{n}{k} \left(\frac{2}{5}\right)^k \left(\frac{3}{5}\right)^{n-k}, \quad k = 0, 1, 2, \dots, n,$$

die Wahrscheinlichkeit des Ereignisses  $k$  rote Kugeln zu ziehen. ■

### Bernoulli-Ketten

Das obige Urnenexperiment sowie der Münzwurf sind Beispiele für eine sogenannte *Bernoulli-Kette* (benannt nach dem aus Basel stammenden Mathematiker Jakob Bernoulli (1655–1705)). Dabei handelt es sich um ein Experiment mit den alternativen Ausgängen  $A$  und  $\bar{A}$ , die mit den Wahrscheinlichkeiten  $P(A) = p$  und  $P(\bar{A}) = q = 1 - p$  unabhängig voneinander eintreten. Eine Serie von  $n$  solcher Experimente, bei der  $k$ -mal das Ergebnis  $A$  und damit  $(n - k)$ -mal das Ergebnis  $\bar{A}$  eintritt, hat dann die Wahrscheinlichkeit  $p^k \cdot q^{n-k}$ .

### Erwartungswert und Varianz binomialverteilter Zufallsgrößen

Eine  $B(n, p)$ -verteilte Zufallsgröße  $X$  nimmt die Werte  $k = 0, 1, 2, \dots, n$  mit den Wahrscheinlichkeiten

$$p_k = \binom{n}{k} p^k q^{n-k}$$

an. Ihr Erwartungswert ist daher (vgl. Abschnitt 10.2) gleich der Summe

$$E(X) = \sum_{k=0}^n p_k \cdot k.$$

Zur Berechnung dieser Summe bedienen wir uns eines hübschen mathematischen Tricks. Hierzu führen wir die Funktion

$$(*) \quad f(x) = \sum_{k=0}^n p_k x^k$$

ein. Sie heisst die *erzeugende Funktion* der Zahlen  $p_k$ . Ihre Ableitung ist die Funktion

$$f'(x) = \sum_{k=1}^n p_k \cdot k x^{k-1}.$$

Also gilt

$$f'(1) = \sum_{k=1}^n p_k \cdot k = E(X).$$

Wenn es nun gelingt, für die Summe auf der rechten Seite von (\*) einen einfachen Ausdruck zu finden, so hat man nur noch dessen Ableitung zu bilden und den Wert  $x = 1$  einzusetzen.

Man erhält

$$\begin{aligned} f(x) &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} x^k \\ &= \sum_{k=0}^n \binom{n}{k} (px)^k q^{n-k}. \end{aligned}$$

Aufgrund des binomischen Lehrsatzes (siehe Abschnitt 3.5), und zwar von rechts nach links gelesen, erhält man dann

$$f(x) = (px + q)^n.$$

Also ist (vgl. Abschnitt 6.3)

$$f'(x) = n(px + q)^{n-1} \cdot p$$

und damit

$$\begin{aligned} f'(1) &= n(p+q)^{n-1} \cdot p \\ &= n \cdot p. \end{aligned}$$

Damit haben wir das Ergebnis

$$E(X) = n \cdot p.$$

Man erkennt: Bei festgehaltenem Wert  $p$  und wachsendem  $n$  verschiebt sich der Erwartungswert nach rechts, wie es die Histogramme in Abbildung 10.2 zeigen.

Die Varianz einer Zufallsgröße  $X$  mit den Werten  $x_k, k = 0, 1, 2, \dots, n$ , dem Erwartungswert  $\mu$  und der Verteilung  $p_k = P(X = x_k)$  wird durch die Summe

$$\text{Var}(X) = \sum_{k=0}^n (x_k - \mu)^2 \cdot p_k$$

definiert. Diese kann man folgendermaßen umformen.

$$\begin{aligned} \text{Var}(X) &= \sum_{k=0}^n (x_k^2 - 2\mu x_k + \mu^2) \cdot p_k \\ &= \sum_{k=0}^n x_k^2 p_k - 2\mu \sum_{k=0}^n x_k p_k + \mu^2 \sum_{k=0}^n p_k. \end{aligned}$$

Wegen

$$\sum_{k=0}^n x_k p_k = E(X) = \mu$$

und

$$\sum_{k=0}^n p_k = 1$$

erhält man also

$$\begin{aligned} \text{Var}(X) &= \sum_{k=0}^n x_k^2 p_k - 2\mu^2 + \mu^2 \\ &= \sum_{k=0}^n x_k^2 p_k - \mu^2. \end{aligned}$$

Kennt man den Erwartungswert einer Zufallsgröße, so bedarf es zur Bestimmung ihrer Varianz also nur noch der Berechnung der Summe in der zweiten Zeile. Diese lässt sich deuten als Erwartungswert der Zufallsgröße  $X^2$ :

$$E(X^2) = \sum_{k=0}^n x_k^2 p_k.$$

Damit erhält man die Beziehung

$$(*) \quad \text{Var}(X) = E(X^2) - \mu^2.$$

Im Falle einer  $B(n, p)$ -verteilten Zufallsgröße  $X$  hat man daher die Summe

$$E(X^2) = \sum_{k=0}^n k^2 p_k$$

zu berechnen. Diese aber lässt sich wieder mit Hilfe der erzeugenden Funktion

$$f(x) = \sum_{k=0}^n p_k x^k$$

ausdrücken. Man hat wieder

$$f'(x) = \sum_{k=1}^n p_k k x^{k-1}$$

und weiter

$$\begin{aligned} (x f'(x))' &= f'(x) + x f''(x) \\ &= \sum_{k=1}^n p_k k^2 x^{k-1}. \end{aligned}$$

Setzt man  $x = 1$ , so findet man die Beziehung

$$E(X^2) = f'(1) + f''(1).$$

Also gilt mit  $(*)$

$$\text{Var}(X) = f'(1) + f''(1) - (f'(1))^2.$$

Oben hatten wir erhalten

$$f'(x) = n(px + q)^{n-1} \cdot p.$$

Nochmalige Ableitung ergibt somit

$$f''(x) = n(n-1)(px + q)^{n-1} p^2.$$

So erhält man:

$$\begin{aligned} E(X^2) &= f'(1) + f''(1) \\ &= n(p+q)^{n-1} p + n(n-1)(p+q)^{n-1} p^2 \\ &= np + n(n-1)p^2 \end{aligned}$$

Damit findet man schließlich:

$$\begin{aligned} \operatorname{Var}(X) &= np + n(n-1)p^2 - (np)^2 \\ &= np + n^2 p^2 - np^2 - (np)^2 \\ &= np(1-p) \\ &= npq \end{aligned}$$

Für die Standardabweichung hat man daher

$$\sigma(X) = \sqrt{npq}.$$

Wir fassen unsere Ergebnisse zusammen:

**Satz:** Eine  $B(n, p)$ -verteilte Zufallsgröße  $X$  hat die Kenngrößen

$$\begin{aligned} E(X) &= np, \\ \operatorname{Var}(X) &= npq \end{aligned}$$

und

$$\sigma(X) = \sqrt{npq}.$$

Man erkennt: Bei festgehaltenem Wert  $p$  vergrößert sich die Breite der Binomialverteilung mit  $\sqrt{n}$ . Die Breite verkleinert sich, wenn  $p$  bzw.  $q$  klein werden.

## 10.4 Die Poisson-Verteilung

Die Poisson-Verteilung, benannt nach den französischen Mathematiker Simeon-Denis Poisson (1781–1840), ist aus zwei Gründen wichtig: Zum einen folgen zahlreiche Phänomene einer Poisson-Verteilung, so zum Beispiel die Anzahl der Regentropfen, die auf die Teilflächen eines Pflasters fallen, zum anderen dient sie zur näherungsweisen Berechnung der Werte  $B(n, p, k)$  der Binomialverteilung für große  $n$  und kleine  $p$ .

Wir untersuchen das Verhalten der Werte

$$B(n, p, k) = \binom{n}{k} p^k (1-p)^{n-k}$$

wenn der Erwartungswert  $\mu = np$  festgehalten und  $n$  immer größer wird. Das bedeutet, die Wahrscheinlichkeit  $p = \mu/n$  wird immer kleiner. Man ersetzt also bei obiger Formel  $p$  durch  $\mu/n$  und untersucht das Verhalten des Ausdrucks

$$B\left(n, \frac{\mu}{n}, k\right) = \binom{n}{k} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k}$$

für große  $n$ . Zunächst formt man folgendermaßen um:

$$\begin{aligned} B\left(n, \frac{\mu}{n}, k\right) &= \frac{n!}{k!(n-k)!} \cdot \frac{\mu^k}{n^k} \left(1 - \frac{\mu}{n}\right)^{n-k} \\ &= \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^n \frac{n!}{(n-k)!n^k} \left(1 - \frac{\mu}{n}\right)^{-k}. \end{aligned}$$

Nun kann man schreiben:

$$\begin{aligned} \frac{n!}{(n-k)!n^k} &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \\ &= \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \end{aligned}$$

Also gilt, da  $k$  fest bleibt,

$$\frac{n!}{(n-k)!n^k} \rightarrow 1$$

für  $n \rightarrow \infty$ . Außerdem gilt

$$\left(1 - \frac{\mu}{n}\right)^{-k} \rightarrow 1$$

für  $n \rightarrow \infty$ . Zu betrachten bleibt die Zahlenfolge

$$\left(1 - \frac{\mu}{n}\right)^n.$$

Diese strebt aber für  $n \rightarrow \infty$  (vgl. Abschnitt 5.3) gegen den  $e^{-\mu}$ . Damit erhalten wir das Ergebnis.

Für jedes  $\mu > 0$  und  $k = 0, 1, 2, \dots$ , gilt

$$B\left(n, \frac{\mu}{n}, k\right) \rightarrow \frac{\mu^k}{k!} \cdot e^{-\mu}.$$

wenn  $n \rightarrow \infty$  geht.

Für den Ausdruck  $\mu^k/k! \cdot e^{-\mu}$  schreibt man auch  $P(\mu, k)$ . Als Näherung für die Werte der Binomialverteilung hat man die folgende Faustregel (Poisson-Näherung):

Für  $p \leq 0,1$  sowie  $n \geq 10$  und  $\mu \geq 5$  ist die Zahl  $P(np, k) = \frac{(np)^k}{k!} e^{-np}$  eine gute Näherung für die Wahrscheinlichkeit  $B(n, p, k)$ .

**Beispiel:** Bei der Herstellung von Transistoren sind 0,6% der gefertigten Stücke defekt. Wie groß ist die Wahrscheinlichkeit, dass in einer 1000-Stück-Packung mehr als zwei

defekte Transistoren enthalten sind?

Von der Zufallsgröße  $X$ , die die Anzahl der defekten Transistoren in einer 1000-Stück-Packung angibt, kann angenommen werden, dass sie  $B(1000; 0,006; k)$ -verteilt ist. Die gesuchte Wahrscheinlichkeit ist dann

$$\begin{aligned} P(X > 2) &= 1 - P(X = 2) - P(X = 1) - P(X = 0) \\ &= 1 - B(1000; 0,006; 2) - B(1000; 0,006; 1) - B(1000; 0,006; 0) \\ &= 1 - \binom{1000}{2} 0,006^2 \cdot 0,994^{998} - \binom{1000}{1} \cdot 0,006 \cdot 0,994^{999} \\ &\quad - \binom{1000}{0} \cdot 0,994^{1000}. \end{aligned}$$

Dieser Ausdruck ist schwierig zu berechnen. Da aber die Bedingungen für die obige Faustregel erfüllt sind, können wir zur Bestimmung der gesuchten Wahrscheinlichkeit die Poisson-Näherung verwenden:

$$\begin{aligned} P(X > 2) &\approx 1 - P(6, 2) - P(6, 1) - P(6, 0) \\ &= 1 - \frac{6^2}{2!} e^{-6} - 6 e^{-6} - e^{-6} \approx 0,94. \end{aligned}$$

■

Neben ihrer Bedeutung als Näherungswerte für die Ausdrücke  $B(n, p, k)$  begründen die Zahlen  $P(\mu, k) > 0$  eine eigenständige Verteilung, die Poisson-Verteilung. Es gilt nämlich (vgl. Abschnitt 5.3):

$$\begin{aligned} \sum_{k=0}^{\infty} P(\mu, k) &= \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} \\ &= e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} \\ &= e^{-\mu} \cdot e^{\mu} = 1. \end{aligned}$$

Daher kann die Zahl  $P(\mu, k)$  als Wahrscheinlichkeit dafür aufgefaßt werden, dass eine Zufallsgröße  $X$  den Wert  $k$  annimmt.

**Definition:** Eine Zufallsgröße  $X$  mit den Werten  $k = 0, 1, 2, \dots$ , heisst Poisson-verteilt mit dem Parameter  $\mu > 0$ , wenn gilt

$$P(X = k) = \frac{\mu^k}{k!} \cdot e^{-\mu}.$$

Wir berechnen nun den Erwartungswert und die Varianz einer Poisson-verteilten Zufallsgröße: Ist

$$f(x) = \sum_{k=0}^{\infty} p_k x^k$$

wieder die zu den Werten  $p_k, k = 0, 1, 2, \dots$ , gehörige erzeugende Funktion (vgl. Abschnitt 10.3) dann gilt, wie oben gezeigt wurde, für die entsprechende Zufallsgröße  $X$ ,

$$E(X) = f'(1)$$

und

$$\text{Var}(X) = f'(1) + f''(1) - (f'(1))^2.$$

Im vorliegenden Fall erhält man

$$\begin{aligned} f(x) &= \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} x^k \\ &= e^{-\mu} \sum_{k=0}^{\infty} \frac{(\mu x)^k}{k!} \\ &= e^{-\mu} \cdot e^{\mu x}. \end{aligned}$$

Also ist

$$f(x) = e^{\mu(x-1)}$$

die zur Poisson-Verteilung gehörige erzeugende Funktion.

Damit ist

$$f'(x) = \mu e^{\mu(x-1)},$$

und somit

$$f'(1) = \mu$$

sowie

$$f''(x) = \mu^2 e^{\mu(x-1)}$$

und daher

$$f''(1) = \mu^2.$$

Also findet man:

*Für eine Poisson-verteilte Zufallsgröße  $X$  mit dem Parameter  $\mu$  hat man*

$$E(X) = \text{Var}(X) = \mu$$

*und damit*

$$\sigma(X) = \sqrt{\mu}.$$

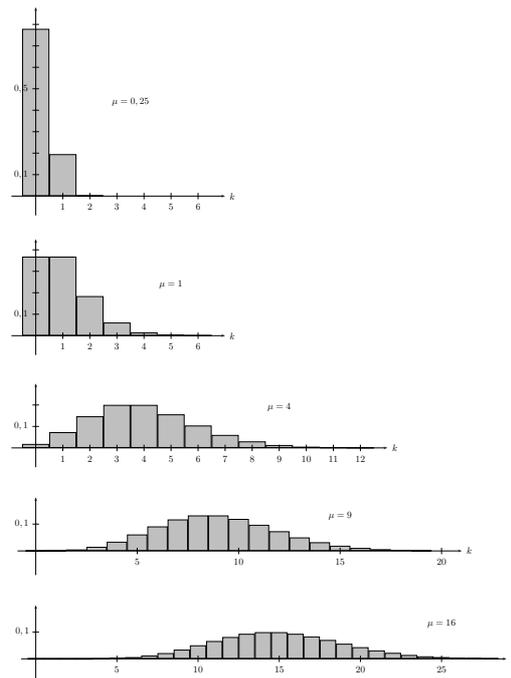


Abbildung 10.3: Einige Histogramme zur Poisson-Verteilung

Für Werte von  $k$  weit rechts vom Mittelwert  $\mu$  sind die Zahlen  $P(\mu, k) = \mu^k / k! \cdot e^{-\mu}$  wegen des Nenners  $k!$  nur wenig von Null verschieden. Häufig lassen sich empirische Verteilungen, bei denen eben dies der Fall ist, mit guter Übereinstimmung als Poisson-verteilt deuten.

**Beispiel:** Wenn Rosinenbrötchen im Durchschnitt 4 Rosinen enthalten, wie groß ist dann die Wahrscheinlichkeit, ein Rosinenbrötchen mit 0 oder 4 oder 6 oder gar 10 Rosinen einzukaufen? Die Erfahrung lehrt, dass es unwahrscheinlich ist, bei einem Durchschnittswert von 4 sehr viel mehr als 4 Rosinen in einem Brötchen anzutreffen. Also ist es plausibel, die Anzahl  $k \geq 0$  der Rosinen in einem Brötchen als eine Poisson-verteilte Zufallsgrößen mit dem Mittelwert  $\mu = 4$  anzusehen. Man erhält daher für die gesuchten Wahrscheinlichkeiten  $P(\mu, k)$  die Werte

$$P(4, 0) = \frac{4^0}{0!} e^{-4} = e^{-4} = 0,01,$$

$$P(4, 4) = \frac{4^4}{4!} e^{-4} = 0,195$$

$$P(4, 6) = \frac{4^6}{6!} e^{-4} = 0,10,$$

$$P(4, 10) = \frac{4^{10}}{10!} e^{-4} = 0,005.$$

Diese Werte stimmen mit der Erfahrung gut überein. ■

## 10.5 Die Normalverteilung

Die Werte der bisher betrachteten Zufallsgrößen waren immer ganze Zahlen. Solche Zufallsgrößen heißen auch *diskrete* Zufallsgrößen. Nun gibt es aber auch Zufallsgrößen  $X$ , deren Werte, zum Beispiel als Ergebnisse von Längen- oder Gewichtsmessungen, beliebige reelle Zahlen sind. In einem solchen Fall wird man nicht nach der Wahrscheinlichkeit  $P(X = \alpha)$ , mit der  $X$  einen bestimmten Wert  $\alpha \in \mathbb{R}$  annimmt, fragen, diese Wahrscheinlichkeit ist immer gleich null, sondern nach der Wahrscheinlichkeit  $P(a \leq X \leq b)$ , mit der  $X$  einen Wert in einem Intervall  $[a, b]$  annimmt. Zufallsgrößen, deren Wertebereich die reellen Zahlen  $\mathbb{R}$  sind oder ein Teilbereich der reellen Zahlen, werden als *stetige* Zufallsgrößen bezeichnet.

Bei einer diskreten Zufallsgröße  $X$  mit Werten  $k = 0, 1, 2, \dots$ , gilt

$$P(a \leq X \leq b) = \sum_{a \leq k \leq b} p_k$$

mit

$$p_k = P(X = k).$$

Bei einer stetigen Zufallsgröße  $X$  muss die obige Summe aus Werten der zugehörigen Verteilung durch ein Integral bezüglich der zugehörigen *Dichtefunktion* ersetzt werden. Dies ist eine Funktion  $p(x) \geq 0$  so dass gilt

$$P(a \leq X \leq b) = \int_a^b p(x) dx.$$

Die Wahrscheinlichkeit  $P(a \leq x \leq b)$ , dass die Zufallsgröße  $X$  im Intervall  $[a, b]$  liegt, ist also gleich dem Inhalt der über  $[a, b]$  gelegenen Fläche unter dem Graphen von  $p(x)$ .

Da der Wert der Zufallsgröße  $X$  mit Sicherheit in  $\mathbb{R}$  liegt, muss die Dichtefunktion notwendig die Bedingung

$$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} p(x) dx = 1$$

erfüllen. Dies bedeutet, dass die Werte der Funktion  $p(x)$  für  $x \rightarrow \pm\infty$  schnell sehr klein werden müssen.

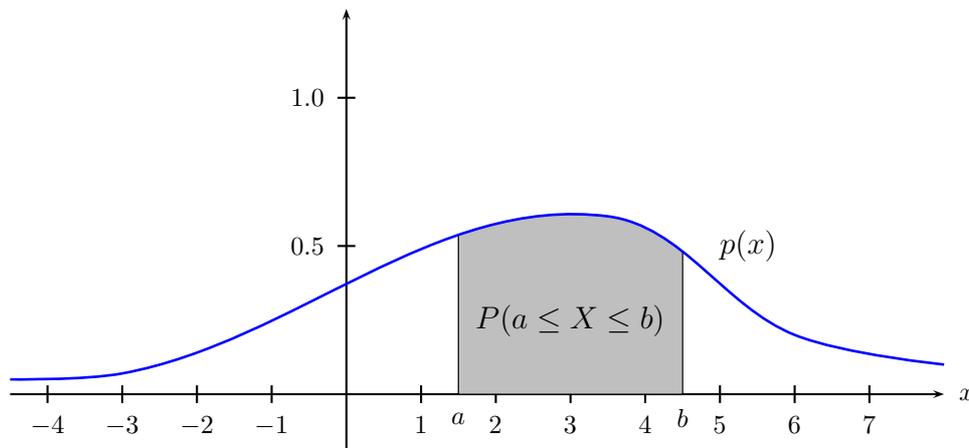


Abbildung 10.4: Die typische Form einer Dichtefunktion.

Interpretiert man die Funktion  $p(x)$  als kontinuierliche Massenverteilung über  $\mathbb{R}$ , dann bestimmt das Integral

$$\mu = \int_{-\infty}^{\infty} x p(x) dx$$

wie im diskreten Fall wieder den Schwerpunkt dieser Verteilung und damit zugleich den *Erwartungswert*  $E(X)$  der entsprechenden Zufallsgröße. Analog zum diskreten Fall hat man für die *Varianz* das Integral

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx.$$

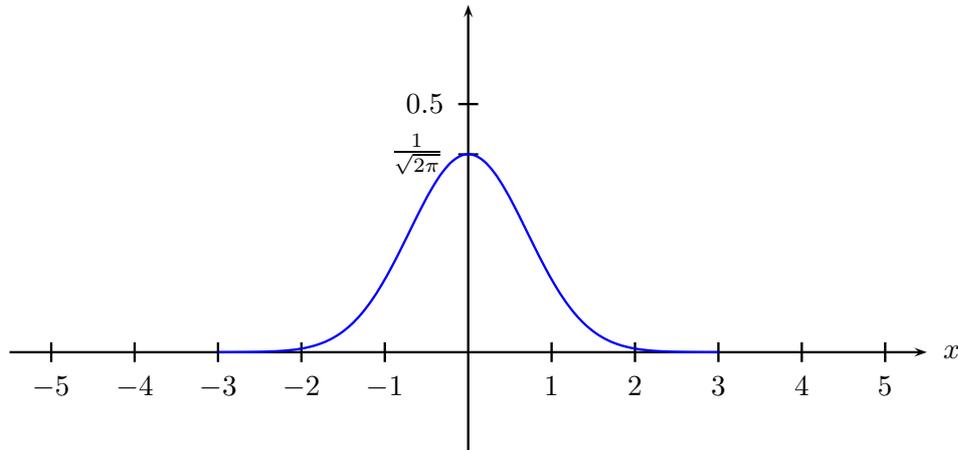
Die für die Praxis wichtigsten Dichtefunktionen ergeben sich aus der *Gaußschen Fehlerkurve* oder *Gaußschen Glockenkurve* mit der Gleichung

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

Carl Friedrich Gauß (1777–1855), der wohl bedeutendste Mathematiker des 18./19. Jahrhunderts fand das nach ihm benannte Fehlergesetz im Jahre 1794 (also im Alter von 17 Jahren) bei Untersuchungen über die Verteilung von Messfehlern bei vielen Einzelmessungen.

Wegen  $\varphi(-x) = \varphi(x)$  ist die Gaußsche Glockenkurve symmetrisch zur  $y$ -Achse. Bei  $x = 0$  befindet sich ihr Maximum. Für  $x \rightarrow \pm\infty$  fallen die Werte  $\varphi(x)$  schnell ab. Der Faktor  $1/\sqrt{2\pi}$  sichert die Beziehung

$$(*) \quad \int_{-\infty}^{\infty} \varphi(x) dx = 1,$$

Abbildung 10.5: Die Gaußsche Glockenkurve  $\varphi(x)$ .

die von jeder Dichtefunktion erfüllt werden muss.

Ausgehend von der Funktion  $\varphi(x)$  bildet man mit Parametern  $\mu \in \mathbb{R}$  und  $\sigma > 0$  die Funktionen

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Diese sind ebenfalls Dichtefunktionen, das heisst die Beziehung (\*) ist auch für die Funktion  $\varphi_{\mu,\sigma}(x)$  erfüllt. Die Funktion  $\varphi_{\mu,\sigma}(x)$  hat die Gestalt einer verschobenen Glockenkurve mit dem Maximum bei  $x = \mu$ . Die Wendepunkte der Glockenkurve  $\varphi_{\mu,\sigma}(x)$  liegen bei  $x = \mu \pm \sigma$ , wie man leicht nachrechnet. (Man setze  $\varphi''_{\mu,\sigma}(x) = 0$ ). Die „Breite“ der Glockenkurve wird daher durch die Größe  $2\sigma$  gekennzeichnet.

**Definition:** Jede stetige Zufallsgröße mit der Dichtefunktion

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

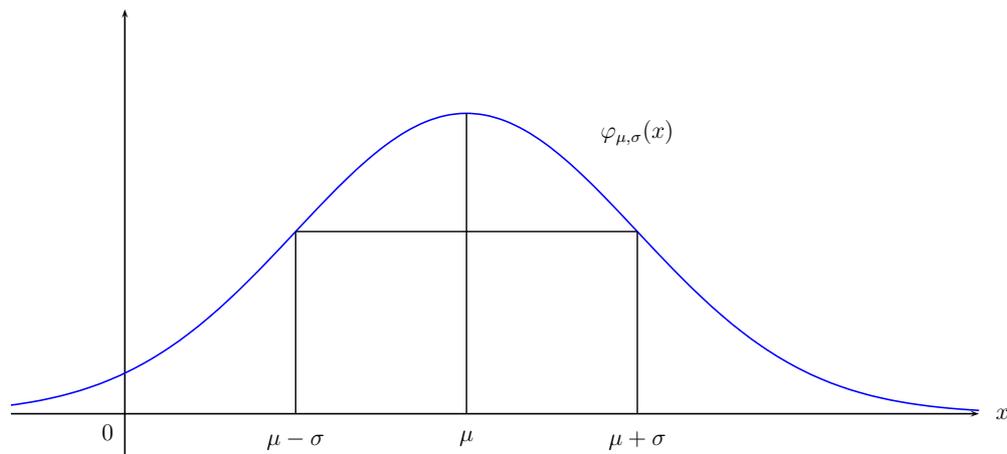
heisst normalverteilt mit den Parametern  $\mu \in \mathbb{R}$  und  $\sigma > 0$ . Eine normalverteilte Zufallsgröße mit  $\mu = 0$  und  $\sigma = 1$  heisst standardisiert oder normiert.

**Bemerkung:** In der Tat gilt

$$(1) \quad \mu = \int_{-\infty}^{\infty} x \varphi_{\mu,\sigma}(x) dx = \mu$$

und

$$(2) \quad \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \varphi_{\mu,\sigma}(x) dx$$

Abbildung 10.6: Die Funktion  $\varphi_{\mu, \sigma}(x)$ .

wodurch die Schreibweise  $\varphi_{\mu, \sigma}(x)$  für die verallgemeinerte Gaußsche Glockenkurve gerechtfertigt wird.

Wir zeigen die Richtigkeit der Behauptung (1):

Es ist

$$\int_{-\infty}^{\infty} x \varphi_{\mu, \sigma}(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma} \varphi\left(\frac{x - \mu}{\sigma}\right) dx.$$

Mit der Transformation

$$\frac{x - \mu}{\sigma} = t$$

hat man

$$x = \mu + \sigma t$$

und daher

$$dx = \sigma dt.$$

Die Integrationsgrenzen bezüglich der Variablen  $t$  bleiben unverändert. Daher gilt

$$\begin{aligned} \int_{-\infty}^{\infty} x \varphi_{\mu, \sigma}(x) dx &= \frac{1}{\sigma} \int_{-\infty}^{\infty} (\mu + \sigma t) \varphi(t) \sigma dt \\ &= \mu \int_{-\infty}^{\infty} \varphi(t) dt + \sigma \int_{-\infty}^{\infty} t \varphi(t) dt. \end{aligned}$$

Da  $\varphi(t)$  eine Dichtefunktion ist, ist das erste der beiden letzten Integrale gleich eins und das zweite ist wegen der Symmetrie  $\varphi(-t) = \varphi(t)$  gleich null, womit die Behauptung (1) bewiesen ist. Mit wenig mehr Aufwand beweist man auch die Beziehung (2). ▲

### Die Verteilungsfunktion $\Phi$

Besitzt die Zufallsgröße  $X$  die Dichtefunktion  $\varphi(x) = \varphi_{0,1}(x)$ , dann ist

$$P(a \leq X \leq b) = \int_a^b \varphi(x) dx$$

die Wahrscheinlichkeit dafür, dass  $X$  einen Wert im Intervall  $[a, b]$  annimmt. Diese Wahrscheinlichkeit kann auch mit Hilfe der Funktion

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$$

ausgedrückt werden. Offenbar gilt

$$P(a \leq X \leq b) = \Phi(b) - \Phi(a).$$

Die Funktion  $\Phi$  heisst *Gaußsches Fehlerintegral* oder *Standardnormalverteilung*. Ihre Werte sind in Tabellen erfaßt.

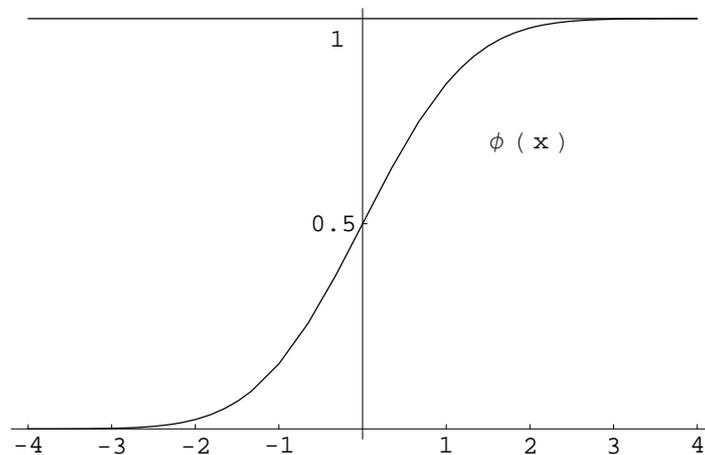


Abbildung 10.7: Die Funktion  $\Phi(x)$ .

Die Funktion  $\Phi(x)$  hat folgende Eigenschaften: Es gilt

$$1) \Phi(-\infty) = 0, \Phi(0) = \frac{1}{2} \text{ und } \Phi(\infty) = 1.$$

2) Der Punkt  $(0, \frac{1}{2})$  ist Wendepunkt des Graphen von  $\Phi(x)$ .

3) Die Funktion  $\Phi(x)$  erfüllt die funktionale Beziehung

$$\Phi(-x) = 1 - \Phi(x).$$

Die Eigenschaften 1) folgen unmittelbar aus der Bedeutung des Wertes  $\Phi(x)$  als Flächeninhalt unter der Funktion  $\varphi(x)$  und aus der Symmetrie von  $\varphi(x)$  bezüglich der  $y$ -Achse. Wegen  $\Phi''(0) = \varphi'(0) = 0$  folgt Eigenschaft 2). Da die Fläche  $\Phi(-x)$  über dem Intervall  $[-\infty, -x]$  aus Symmetriegründen gleich der Fläche  $1 - \Phi(x)$  über dem Intervall  $[x, \infty]$  ist, muss Eigenschaft 3) erfüllt sein. Es genügt daher, nur die Werte  $\Phi(x)$  für  $x \geq 0$  zu tabellieren.

Ist eine Zufallsgröße  $X$  normalverteilt mit den Parametern  $\mu$  und  $\sigma$ , dann hat man

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b \varphi_{\mu, \sigma}(x) dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx. \end{aligned}$$

Setzt man  $(x - \mu)/\sigma = t$  dann ist  $dx = \sigma dt$  und die Grenzen  $a, b$  bezüglich der Variablen  $x$  gehen über in die Grenzen  $(a - \mu)/\sigma$  und  $(b - \mu)/\sigma$  bezüglich der Variablen  $t$ . Man erhält schließlich:

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi}} \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} e^{-\frac{1}{2}t^2} dt$$

(\*)

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Damit kann die gesuchte Wahrscheinlichkeit mit Hilfe von Werten der Funktion  $\Phi(x)$  berechnet werden.

**Beispiel:** Aufgrund klinischer Untersuchungen weiß man, dass Neugeborene eine mittlere Körperlänge von 51 cm bei einer Standardabweichung von 4 cm haben. – Wieviele von 1000 Babys haben eine Länge zwischen 45 und 50 cm?

Nimmt man an, dass die Zufallsgröße  $X =$  Körperlänge normalverteilt ist, dann gilt

mit (\*) und Eigenschaft 3) der Funktion  $\Phi(x)$ .

$$\begin{aligned}
 P(45 \leq X \leq 50) &= \Phi\left(\frac{50 - 51}{4}\right) - \Phi\left(\frac{45 - 51}{4}\right) \\
 &= \Phi(-0,25) - \Phi(-1,5) \\
 &= 1 - \Phi(0,25) - (1 - \Phi(1,5)) \\
 &= \Phi(1,5) - \Phi(0,25) \\
 &= 0,933 - 0,598 = 0,335
 \end{aligned}$$

Also werden von 1000 Babys ca. 335 eine Körperlänge zwischen 45 und 50 cm haben. ■

### Die $\sigma$ -Regeln

Die Wahrscheinlichkeit, dass die Werte einer normalverteilten Zufallsgröße im Intervall  $[\mu - \sigma, \mu + \sigma]$  um den Erwartungswert  $\mu$  liegen, ist nach (\*)

$$\begin{aligned}
 P(\mu - \sigma \leq X \leq \mu + \sigma) &= \Phi\left(\frac{\mu + \sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - \sigma - \mu}{\sigma}\right) \\
 &= \Phi(1) - \Phi(-1) \\
 &= \Phi(1) - (1 - \Phi(1)) \\
 &= 2\Phi(1) - 1 = 0,683.
 \end{aligned}$$

Für das entsprechende  $2\sigma$ - und  $3\sigma$ -Intervall um  $\mu$  erhält man die Werte

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0,955$$

und

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0,997.$$

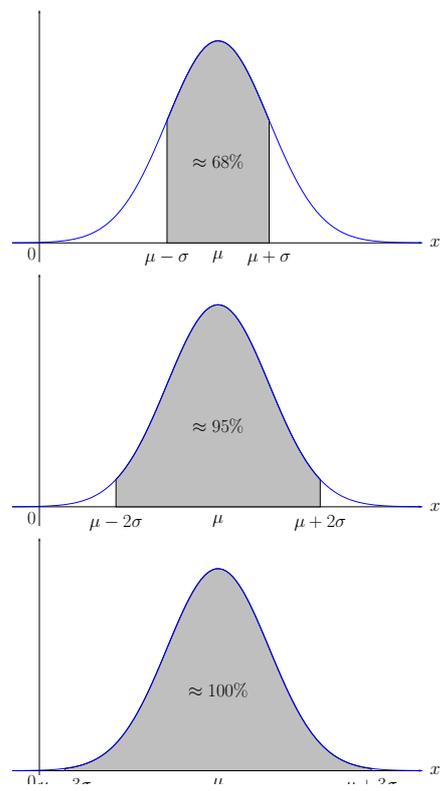
Damit gilt der

**Satz:** Ist  $X$  eine normalverteilte Zufallsgröße mit den Parametern  $\mu$  und  $\sigma$ , dann ist

$$\begin{aligned}
 P(\mu - \sigma \leq X \leq \mu + \sigma) &= 68,3\%, \\
 P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &= 95,5\%, \\
 P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &= 99,7\%.
 \end{aligned}$$

Im  $3\sigma$ -Intervall um den Erwartungswert  $\mu$  liegen also mit großer Sicherheit so gut wie alle Werte einer normalverteilten Zufallsgröße.

**Beispiel:** Eine Maschine stellt eine Bohrung mit dem Soll Durchmesser 1 cm her. Toleriert wird eine Abweichung von  $1/1000$  cm nach oben und nach unten. Arbeitet die Maschine mit einem Mittelwert  $\mu = 1$  cm und einer Standardabweichung  $\sigma = 1/2000$  cm

Abbildung 10.8: Die  $\sigma$ -Regeln.

dann liegen 95% der gefertigten Bohrungen im Toleranzbereich und ca. 5% der gefertigten Stücke sind Ausschußstücke. ■

### Näherungsbeziehungen für $B(n, p)$ -verteilte Zufallsgrößen

Die Kontur der Histogramme  $B(n, p)$ -verteilter Zufallsgrößen (siehe Abbildung 10.2) ähnelt stark den Graphen von Dichtefunktionen  $\varphi_{\mu, \sigma}(x)$ , sofern der Parameter  $n$  nicht zu klein ist. In der Tat besteht mit  $\mu = np$  und  $\sigma = \sqrt{npq}$  die Näherungsbeziehung

$$B(n, p, k) \approx \varphi_{\mu, \sigma}(k),$$

das heißt, es gilt

$$B(n, p, k) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}.$$

Diese Näherung ist brauchbar, wenn  $npq > 9$  ist.

Im allgemeinen interessieren aber nicht einzelne Werte  $B(n, p, k)$ , sondern Summen dieser Werte, wie sie in der Formel

$$P(a \leq X \leq b) = \sum_{k=a}^b B(n, p, k)$$

für binomialverteilte Zufallsgrößen  $X$  auftreten. Die rechte Seite läßt sich auffassen als eine Summe von Rechteckflächen in den Grenzen  $a, b$  bei dem entsprechenden Histogramm. Diese Rechtecksumme wird approximiert durch die Fläche über dem Intervall  $[a, b]$  bezüglich der Dichtefunktion  $\varphi_{\mu, \sigma}$  mit  $\mu = np$  und  $\sigma = \sqrt{npq}$ . Also gilt (siehe obige Formel (\*)):

$$\sum_{k=a}^b B(n, p, k) \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right).$$

Die Güte dieser Näherung läßt sich durch eine sogenannte „Stetigkeitskorrektur“ mit dem Summanden 0,5 beim Argument von  $\Phi$  noch verbessern. Die Näherungsformel lautet dann:

$$\sum_{k=a}^b B(n, p, k) \approx \Phi\left(\frac{b + 0,5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - 0,5 - np}{\sqrt{npq}}\right)$$

Diese Näherungsformel liefert gute Werte, wenn wieder  $npq > 9$  ist.

**Beispiel:** 140 Personen haben eine bestimmte Busreise gebucht. Die Reisenden müssen auf zwei Busse mit 110 (Bus A) bzw. 70 (Bus B) Sitzplätzen verteilt werden. Den Reisenden wird nun mitgeteilt, dass zur selben Zeit zwei Busse von verschiedenen Stellen abfahren. Der Reiseunternehmer überläßt die Aufteilung der Reisenden auf

die beiden Busse dem Zufall. Wie groß ist dann die Wahrscheinlichkeit, dass bei einem Bus mehr Reisende eintreffen, als Sitzplätze vorhanden sind?

Wir nehmen an, dass keiner der beiden Busse von vornherein von den Reisenden bevorzugt wird. Die Zufallsgröße  $X$ , die angibt, wie viele Reisende zuerst am Bus A eintreffen, ist also  $B(140, \frac{1}{2})$ -verteilt. Es treffen zu viele Reisende an einem der Busse ein, wenn  $X < 70$  oder  $X > 110$  ist. Gesucht ist also die Wahrscheinlichkeit  $1 - P(70 \leq X \leq 110)$ . Nach der obigen Näherungsformel erhält man:

$$\begin{aligned} 1 - P(70 \leq X \leq 110) &\approx 1 - \left( \Phi\left(\frac{110 + 0,5 - 70}{\sqrt{35}}\right) - \Phi\left(\frac{70 - 0,5 - 70}{\sqrt{35}}\right) \right) \\ &= 1 - \left( \Phi\left(\frac{40,5}{\sqrt{35}}\right) - \Phi\left(\frac{-0,5}{\sqrt{35}}\right) \right) \\ &= 1 - \Phi\left(\frac{40,5}{\sqrt{35}}\right) + 1 - \Phi\left(\frac{0,5}{\sqrt{35}}\right) \\ &= 2 - \Phi(6,84) - \Phi(0,08) \\ &= 2 - 1 - 0,53 = 0,47. \end{aligned}$$

Die Wahrscheinlichkeit, dass es aufgrund der unklaren Regelung bezüglich der beiden Busse Verzögerungen bei der Abreise gibt, beträgt also fast 50%. ■

**Bemerkung: (Der zentrale Grenzwertsatz)** Die überwiegende Zahl der in den Anwendungen auftretenden Zufallsgrößen ist annähernd normalverteilt (zum Beispiel Körpergröße, Gewicht, Intelligenz, ...). Man hat es hier mit einem universellen Phänomen zu tun, das nach einer Erklärung verlangt. Diese liefert der *zentrale Grenzwertsatz* der Wahrscheinlichkeitstheorie, einer der bedeutendsten Sätze der Mathematik. Seine Aussage lässt sich etwa folgendermaßen wiedergeben: Wenn eine Zufallsgröße durch zahlreiche additiv und unabhängig voneinander wirkende Einflüsse entsteht, dann ist sie tendenziell normalverteilt. So nähert sich zum Beispiel die Zufallsgröße „Augensumme“ umso deutlicher einer Normalverteilung, je mehr Würfel geworfen werden. ▲

## 10.6 Testen

Häufig sind Entscheidungen bei unvollständiger Information zu treffen. Hierzu ein Beispiel aus der Praxis der industriellen Qualitätskontrolle an dem sich exemplarisch die Überlegungen verdeutlichen lassen, die solchen Testverfahren zugrunde liegen: Erhält ein Autohersteller von einem Zulieferer eine Sendung mit Schrauben, so kann man aus Kostengründen nicht jede einzelne Schraube auf die Einhaltung der geforderten Qualitätsstandards überprüfen. Statt eine solche Totalkontrolle durchzuführen wird man der Lieferung eine *Stichprobe* entnehmen und jedes Stück der Stichprobe

prüfen. Finden sich dabei „zu viele“ schlechte Stücke, so wird man die Annahme der Lieferung verweigern. Findet man keine oder wenige schlechte Stücke, so wird man die Lieferung akzeptieren.

Bevor wir dieses Prüfverfahren präzisieren, ist der Begriff der Stichprobe oder *Zufallsstichprobe* zu erläutern.

Die Forderung der Zufälligkeit der Stichprobe wird manchmal mit der Wendung umschrieben: Jedes Stück der Gesamtheit muss die gleiche Chance besitzen, in die Stichprobe zu gelangen. In der Praxis versucht man diese Forderung dadurch zu erfüllen, dass man eine offensichtlich nicht „zufällige“ Auswahl vermeidet. Eine solche nicht zufällige Auswahl läge zum Beispiel dann vor, wenn die „guten“ Stücke in einer Partie als solche erkennbar wären und bevorzugt ausgewählt würden oder wenn bei einer Kontrolle der Wochenproduktion nur Stücke aus der Montagsproduktion untersucht würden. Mit der Forderung der Zufälligkeit der Stichprobenauswahl soll erreicht werden, dass die Stichprobe ein möglichst getreues „verkleinertes Abbild“ der Gesamtheit darstellt, aus der sie stammt.

Für die wahrscheinlichkeitstheoretischen Modelle, die den praktischen Verfahren der Qualitätskontrolle zugrundeliegen, ist die Annahme der Zufälligkeit der Stichprobenentnahme freilich von entscheidender Bedeutung.

Bei dem Verfahren, das über Annahme oder Ablehnung einer Lieferung entscheidet, geht man nun folgendermaßen vor: Einer Lieferung von  $N$  Stücken werde eine Zufallsstichprobe vom Umfang  $n$  ohne Zurücklegen entnommen. In dieser Stichprobe stellt man die Anzahl schlechter Stücke fest. Vorab sei eine *Annahmezahl*  $c$  ( $0 \leq c \leq n - 1$ ) festgelegt worden. Über „Annahme“ oder „Ablehnung“ der Lieferung wird dann nach der folgenden Regel entschieden:

*Enthält die Stichprobe höchstens  $c$  schlechte Stücke, dann wird die Lieferung angenommen, enthält sie mehr als  $c$  schlechte Stücke, dann wird sie beanstandet (bzw. „abgelehnt“ oder „zurückgewiesen“).*

Eine solche Entscheidung heisst auch *Prüfplan*.

Bei der Festlegung eines Prüfplans, das heisst bei der Wahl des Stichprobenumfangs  $n$  und der Annahmezahl  $c$ , spielen einerseits die Kosten der Stichprobenerhebung und andererseits die Kosten und Risiken von „Fehlentscheidungen“ eine Rolle. Die Kosten, die durch Fehlentscheidungen verursacht werden, sind meistens schwer zu quantifizieren. Zur Erläuterung der Risiken möge ein Beispiel dienen:

Zwischen dem Lieferanten und dem Abnehmer von 1000 Glühbirnen wird vereinbart, der Lieferung eine Zufallsstichprobe ohne Zurücklegen vom Umfang  $n = 50$  zu entnehmen. Die Annahmezahl wird mit  $c = 4$  festgelegt. Aufgrund der üblichen Qualitätsmaßstäbe für solche Lieferungen wird die Qualität als ausreichend angesehen,

wenn die Lieferung nicht mehr als 6% Ausschuss, das heisst 60 defekte Glühbirnen, enthält. Beträgt nun der tatsächliche Ausschussanteil nur 5%, so können trotzdem in der Stichprobe mehr als 4 defekte Glühbirnen auftreten. In diesem Fall würde die Lieferung zum Nachteil des Lieferanten beanstandet. Beträgt hingegen der tatsächliche Ausschussanteil 15%, so kann die Anzahl der schlechten Stücke in der Stichprobe dennoch 4 oder weniger betragen. Die Lieferung würde dann vom Abnehmer „fälschlicherweise“ nicht beanstandet. Die Wahrscheinlichkeit, dass bei Anwendung der obigen Entscheidungsregel eine Lieferung ausreichender Qualität beanstandet wird, nennt man das *Produzentenrisiko*, die Wahrscheinlichkeit für die Annahme einer Lieferung von schlechter Qualität wird das *Konsumentenrisiko* genannt.

**Bemerkung:** Die Begriffe Produzentenrisiko und Konsumentenrisiko werden auch benutzt, wenn es sich bei dem Lieferanten oder Abnehmern nicht um den Hersteller bzw. Verbraucher handelt. ▲

Die Wahrscheinlichkeit für Fehlentscheidungen, das heisst die Risiken des Prüfverfahrens, hängen wesentlich vom Stichprobenumfang  $n$  und der Annahmezahl  $c$  ab, wobei mit abnehmendem Produzentenrisiko das Konsumentenrisiko wächst und umgekehrt. Die Festlegung von  $n$  und  $c$  derart, dass die Risiken für beide Parteien vertretbar sind, kann deshalb zu Konflikten führen, bei denen neben rein mathematischen Überlegungen auch Interessenpolitik, gesetzliche Vorgaben usw. eine Rolle spielen.

### Der wahrscheinlichkeitstheoretische Hintergrund des Verfahrens

Enthält eine Lieferung vom Umfang  $N$  schlechte Stücke der Anzahl  $S$ , dann ist  $p = S/N$  ihr Ausschussanteil. Oder: die Wahrscheinlichkeit, bei zufälliger Entnahme eines Stückes aus der Lieferung ein schlechtes Stück in den Händen zu haben, ist gleich  $p$ . Ist nun  $X$  die Zufallsgröße, welche die Anzahl der schlechten Stücke in der Stichprobe von Umfang  $n$  zählt, dann ist  $X$  binomialverteilt mit den Parametern  $n$  und  $p$ . Daher ist

$$(*) \quad P(X \leq c) = \sum_{k=0}^c \binom{n}{k} p^k (1-p)^{n-k}$$

die Wahrscheinlichkeit, in der Stichprobe vom Umfang  $n$  höchstens  $c$  schlechte Stücke zu finden.

**Bemerkung:** Strenggenommen handelt es sich bei (\*) nur um eine Näherung für die Wahrscheinlichkeit  $P(X \leq c)$ . Die Entnahme der Stichprobe verändert nämlich laufend die Grundgesamtheit aus der entnommen wird, so dass sich der ursprüngliche Ausschussanteil  $p = S/N$  im Zuge der Entnahme der Stichprobe ebenfalls verändert. Ist aber der sogenannte „Auswahlsatz“  $n/N$  klein, gilt zum Beispiel  $n/N \leq 0,1$ , so macht man keinen großen Fehler, wenn man den Ausschussanteil während der Entnahme der Stichprobe als konstant ansieht. ▲

Das Modell, wonach die Anzahl der schlechten Stücke in der Stichprobe binomialverteilt ist, machen wir uns nunmehr zu eigen. Die *Annahmekennlinie* oder *Operationscharakteristik* ist dann das Schaubild der Funktion  $A : [0; 1] \rightarrow [0; 1]$  mit

$$(1) \quad A(p) = \sum_{k=0}^c \binom{n}{k} p^k (1-p)^{n-k}.$$

Ihr Wert an der Stelle  $p$ ,  $0 \leq p \leq 1$ , ist die Wahrscheinlichkeit, dass bei Anwendung der obigen durch  $n$  und  $c$  festgelegten Entscheidungsregel eine Lieferung angenommen wird, wenn ihr Ausschussanteil gleich  $p$  ist.

Die Funktion  $A$  ist ein Polynom  $n$ -ten Grades in der Variablen  $p$ . Ihr Verlauf im Intervall  $[0; 1]$  ist zwischen den Randwerten  $A(0) = 1$  und  $A(1) = 0$  streng monoton fallend. Während das Randverhalten der Funktion  $A$  aus der rechten Seite von (1) unmittelbar ablesbar ist, führen wir zum Nachweis der Monotonie von  $A$  drei Argumente unterschiedlicher Strenge an:

#### *Empirisches Argument*

Mit Hilfe einer Tabelle oder eines Rechners erstellt man eine Zeichnung der Annahmekennlinie für verschiedene Prüfpläne.

#### *Apriori-Argument*

Haben zwei Lieferungen gleichen Umfangs die Ausschussanteile  $p_1$  bzw.  $p_2$  mit  $p_1 < p_2$  und werden beide nach demselben Plan geprüft, dann ist die Chance bei der ersten Lieferung größer als bei der zweiten, dass die Zahl der schlechten Stücke nicht größer als  $c$  ist. Folglich ist die Annahmewahrscheinlichkeit bei der ersten Lieferung größer als bei der zweiten, das heisst, es muss gelten  $A(p_1) > A(p_2)$ .

#### *Analytisches Argument*

Man berechne die Ableitung der Funktion  $A(p)$ . Es ergibt sich mit

$$A(p) = \sum_{k=0}^c \binom{n}{k} p^k (1-p)^{n-k}$$

zunächst

$$A'(p) = \sum_{k=1}^c \binom{n}{k} k p^{k-1} (1-p)^{n-k} - \sum_{k=0}^c \binom{n}{k} p^k (n-k) (1-p)^{n-k-1},$$

und mit

$$\binom{n}{k} k = n \binom{n-1}{k-1}$$

und

$$\binom{n}{k}(n-k) = \binom{n}{n-k}(n-k) = n \binom{n-1}{n-k-1} = n \binom{n-1}{k}$$

erhält man schließlich

$$A'(p) = \sum_{k=1}^c n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} - \sum_{k=0}^c n \binom{n-1}{k} p^k (1-p)^{n-k-1}.$$

Man erkennt, dass die erste Summe mit der zweiten bis auf deren letzten Summanden ( $k=c$ ) übereinstimmt, so dass man

$$A'(p) = -n \binom{n-1}{c} p^c (1-p)^{n-c-1}$$

erhält. Für  $0 < p < 1$  gilt also  $A'(p) < 0$ . Die Funktion  $A$  ist daher im Intervall  $(0;1)$  streng monoton fallend.

Der Verlauf der Annahmekennlinie bestätigt, dass grundlegende Forderungen, die an einen Prüfplan vernünftigerweise gestellt werden müssen, erfüllt sind: Es werden nämlich ausschussfreie Lieferungen stets und Lieferungen, die nur aus Ausschußstücken bestehen, niemals angenommen. Außerdem nimmt die Wahrscheinlichkeit, dass eine Lieferung angenommen wird, mit wachsendem Ausschußanteil (streng monoton) ab.

Weiter wird man von einem vernünftigen Prüfplan hohe „Trennschärfe“ verlangen, das heißt er muss dazu geeignet sein, „gute“ Lieferungen von „schlechten“ deutlich unterscheiden zu können. Dabei werde hier eine Lieferung als „gut“ oder akzeptabel bezeichnet, wenn ihr Ausschussanteil  $p$  eine Schranke  $p_0$  nicht überschreitet, und als inakzeptabel oder „schlecht“, wenn er eine Schranke  $p_1 > p_0$  nicht unterschreitet. Bei einem Prüfverfahren mit hoher Trennschärfe ist dann die Annahmewahrscheinlichkeit  $A(p)$  für  $0 \leq p \leq p_0$  nahe bei eins und für  $p_1 \leq p \leq 1$  nahe bei null gelegen. Liegt  $p_1$  nahe bei  $p_0$ , dann hat die Annahmekennlinie im Bereich  $p_0 < p < p_1$  den in der Abbildung skizzierten steilen Verlauf.

## 10.7 Schätzen

Sind die Parameter – etwa Mittelwert und Varianz – einer Verteilung von bekannten Typus selbst unbekannt, so wird man versuchen, diese aufgrund der Information, die eine Stichprobe liefert, zu schätzen. Das heißt, man wird aus den Daten der Stichprobe eine *Schätzgröße* errechnen, die dem unbekanntem Parameter möglichst nahe kommt. Eine grundlegende Methode, zu einer solchen Schätzgröße zu gelangen, wird im folgenden anhand eines Beispiels dargestellt.

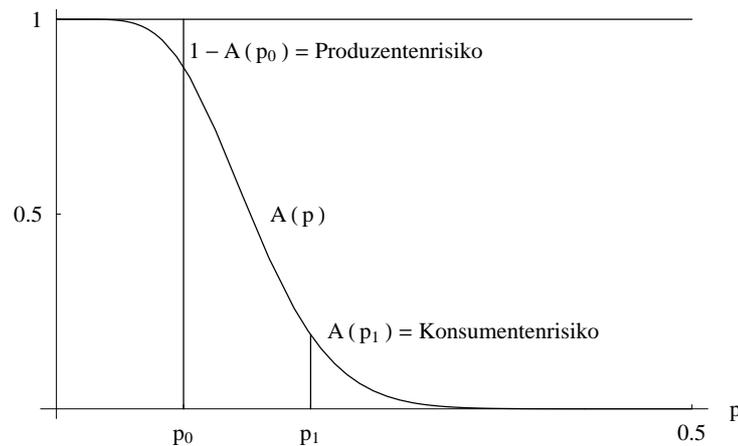


Abbildung 10.9: Die Risiken des Prüfverfahrens.

### Das Maximum-Likelihood-Prinzip

Eine Firma erhält eine Lieferung von Transistoren. Der Lieferung wird eine Stichprobe vom Umfang  $n$  entnommen. Man findet  $k$ ,  $0 \leq k \leq n$ , defekte Transistoren in der Stichprobe. Wie hoch ist dann der unbekannte Ausschussanteil  $p$  in der Lieferung aufgrund dieses Stichprobenergebnisses zu schätzen?

Enthält die Lieferung einen geringen Prozentsatz defekter Transistoren, dann wird die Zahl  $k$  mit hoher Wahrscheinlichkeit klein sein, bei einem hohen Prozentsatz wird sie mit hoher Wahrscheinlichkeit in der Nähe von  $n$  liegen.

Das *Maximum-Likelihood-Prinzip* oder Prinzip der maximalen Wahrscheinlichkeit besagt nun: Man bestimme denjenigen Ausschussanteil  $\hat{p}$ , der dem beobachteten Stichprobenergebnis maximale Wahrscheinlichkeit verleiht. Dieser Wert  $\hat{p}$  dient dann als Schätzwert für den unbekanntem Ausschussanteil.

Im vorliegenden Fall ist  $\hat{p}$  daher so zu bestimmen, dass die Wahrscheinlichkeit, in einer Stichprobe vom Umfang  $n$  gerade  $k$  defekte Transistoren zu finden, maximal wird.

Ist nun  $p$  der Ausschussanteil in der Lieferung, dann ist

$$S(p) = \binom{n}{k} p^k (1-p)^{n-k}$$

die Wahrscheinlichkeit, in der Stichprobe vom Umfang  $n$  gerade  $k$  schlechte Stücke zu finden. Die Funktion  $S(p)$ ,  $0 \leq p \leq 1$ , hat die Ableitung

$$S(p)' = \binom{n}{k} p^{k-1} (1-p)^{n-k-1} \cdot (k - np),$$

wobei  $0 < k < n$  ist. An der Stelle  $\hat{p} = k/n$  hat  $S'(p)$  eine Nullstelle und  $S(p)$  ein Maximum, wie der Vorzeichenwechsel von + nach - bei  $S'(p)$  an der Stelle  $\hat{p}$  erkennen

läßt.

Für  $k = 0$  ist

$$S(p) = (1 - p)^n$$

mit dem Maximum bei  $\hat{p} = 0$ . Für  $k = n$  hat man

$$S(p) = p^n$$

mit dem Maximum bei  $\hat{p} = 1$ .

Der Maximum-Likelihood-Schätzwert für den Ausschussanteil in der Lieferung ist im vorliegendem Fall daher, wie zu erwarten war, gleich dem Ausschussanteil

$$\hat{p} = \frac{k}{n}$$

in der Stichprobe.

### Schätzen des Erwartungswertes

Sind Messwerte  $x$  Werte einer normalverteilten Zufallsgröße  $X$  mit dem Erwartungswert  $\mu$  und der Standardabweichung  $\sigma$ , dann liegt ein *einzelner* Meßwert  $x$  aufgrund der  $\sigma$ -Regeln (s.o) mit einer Wahrscheinlichkeit von 99% in Intervall  $[\mu - 3\sigma, \mu + 3\sigma]$ , mit einer Wahrscheinlichkeit von 95% im Intervall  $[\mu - 2\sigma, \mu + 2\sigma]$  und mit einer Wahrscheinlichkeit von 68% im Intervall  $[\mu - \sigma, \mu + \sigma]$ . Mehr läßt sich über die Lage eines einzelnen Meßwertes bezüglich des Erwartungswertes  $\mu$  nicht sagen.

Nun erscheint es aber plausibel anzunehmen, dass das *arithmetische Mittel*

$$(*) \quad \bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

von  $n$  Meßwerten „näher“ beim Erwartungswert  $\mu$  liegt als ein einzelner Meßwert. Diese Vermutung wird durch die Theorie bestätigt.

Das arithmetische Mittel (\*) kann aufgefaßt werden als Wert der Zufallsgröße

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n),$$

wobei die Zufallsgrößen  $X_1, X_2, \dots, X_n$  dieselbe Verteilung besitzen wie die Zufallsgröße  $X$ .

Man kann nun zeigen: Die Zufallsgröße  $\bar{X}$  besitzt denselben Erwartungswert wie die Zufallsgröße  $X$ ,

$$(**) \quad E(\bar{X}) = E(X) = \mu,$$

das heisst,  $\bar{X}$  ist eine *erwartungstreue* Schätzgröße, und es gilt

$$\text{Var}(\bar{X}) = \frac{1}{n} \cdot \sigma^2,$$

also ist

$$(***) \quad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

Das bedeutet, das arithmetische Mittel  $\bar{x}$  liegt mit einer Wahrscheinlichkeit von 95% im Intervall  $[\mu - 2\sigma/\sqrt{n}, \mu + 2\sigma/\sqrt{n}]$ . Durch Vergrößerung der Zahl  $n$  kann man dieses Intervall zwar beliebig einengen, seine tatsächliche Lage und Weite bleiben aber unbekannt, da im allgemeinen die Parameter  $\mu$  und  $\sigma$  der Verteilung  $X$  unbekannt sind.

Aufgrund der Beziehungen (\*\*), (\*\*\*) ist aber das arithmetische Mittel  $\bar{x}$  ein optimaler Schätzwert für den unbekanntem Erwartungswert  $\mu$  einer Verteilung  $X$ . Dies gilt auch dann, wenn  $X$  nicht normalverteilt ist.

### Schätzen der Varianz

Als Zufallsgröße (Schätzgröße) deren Werte Schätzwerte für die unbekanntem Varianz  $\sigma^2$  einer Zufallsgröße  $X$  liefern, wird man den Ausdruck

$$V = \frac{1}{n} ((X_1 - \mu)^2 + (X_2 - \mu)^2 + \dots + (X_n - \mu)^2)$$

vermuten. In der Tat sind die Werte von  $V$  im selben Sinne optimale Schätzwerte für die Varianz  $\sigma^2$  wie die Werte der Zufallsgröße  $\bar{X}$  für den Erwartungswert  $\mu$ . Im allgemeinen ist der Parameter  $\mu$  aber nicht bekannt.

Ersetzt man nun im Ausdruck  $V$  den Parameter  $\mu$  durch die Zufallsgröße  $\bar{X}$ , dann erhält man die Schätzgröße

$$V' = \frac{1}{n} ((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2),$$

deren Werte aufgrund einer Stichprobe mit den Messwerten  $x_1, x_2, \dots, x_n$  berechnet werden können. Nun gilt aber

$$E(V') \neq \sigma^2.$$

Die Schätzgröße  $V'$  ist daher nicht „erwartungstreu“. Erwartungstreu ist hingegen die Schätzgröße

$$S^2 = \frac{1}{n-1} ((X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2)$$

mit den Werten

$$(*) \quad s^2 = \frac{1}{n-1} \left( (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2 \right)$$

die aufgrund der Ergebnisse einer Stichprobe berechnet werden können. Die Wurzel  $s$  aus der rechten Seite von  $(*)$  ist dann ein Schätzwert für die Standardabweichung  $\sigma$  der betrachteten Zufallsgröße  $X$ .

Hat man bei einer Messreihe die Messwerte  $x_1, x_2, \dots, x_n$  erhalten, dann betrachtet man daher das arithmetische Mittel  $\bar{x}$  als Näherungswert für den unbekanntenen „wahren Wert“ der Messgröße und bringt die *Messunsicherheit* durch die Notation  $\bar{x} \pm s$  zum Ausdruck.

## 10.8 Die Ausgleichsgerade

Weiß man aufgrund theoretischer Überlegungen, dass zwei Größen  $X$  und  $Y$  in einem linearen Zusammenhang stehen (wie zum Beispiel die Auslenkung einer Feder und die dazu erforderliche Kraft aufgrund des Hookschen Gesetzes) dann besteht zwischen den entsprechenden Messgrößen  $x$  und  $y$  eine Beziehung der Form

$$y = ax + b$$

mit zunächst unbekanntem Parametern  $a$  und  $b$ . Zur Bestimmung dieser Parameter führt man eine Messreihe durch, bei der man zu den vorgegebenen Werten  $x_1, x_2, \dots, x_n$  die zugehörigen Werte  $y_1, y_2, \dots, y_n$  experimentell ermittelt. Trägt man die Punkte  $P_1 = (x_1, y_1), P_2 = (x_2, y_2), \dots, P_n = (x_n, y_n)$  in ein Diagramm ein, so werden diese aufgrund von zufälligen Einflüssen (Messfehlern) im allgemeinen nicht exakt auf einer Geraden liegen. Es stellt sich dann die Frage, wie man gleichwohl aufgrund der vorhandenen Daten  $x_j, y_j, j = 1, 2, \dots, n$ , die Parameter  $a, b$  einer *Ausgleichsgeraden* oder *Regressionsgeraden* so bestimmen kann, dass diese die Punkte  $P_1, P_2, \dots, P_n$  möglichst gut verbindet.

Diese Forderung nach einer möglichst guten Ausgleichsgeraden wird nun folgendermaßen präzisiert: Man betrachte die Abweichungen

$$d_j = y_j - y(x_j) = y_j - ax_j - b$$

der gemessenen Werte  $y_i$  von den theoretischen Werten  $y(x_j) = ax_j + b$  und bildet daraus einen „Gesamtfehler“, den es dann zu minimieren gilt. Als ein solcher Gesamtfehler ist die Summe

$$\sum_{j=1}^n d_j = \sum_{j=1}^n (y_j - ax_j - b)$$

offenbar ungeeignet, da sich dann Abweichungen nach unten und solche nach oben aufheben könnten und so trotz eines minimalen Gesamtfehlers die Lage der Punkte  $P_j$  erheblich von einer Geraden abweichen könnte.

Dieses Phänomen kann so nicht auftreten, wenn man die Summe

$$F(a, b) = \sum_{j=1}^n d_j^2 = \sum_{j=1}^n (y_j - ax_j - b)^2$$

der quadratischen Abweichungen als Gesamtfehler betrachtet. Jetzt können sich Abweichungen nach oben und nach unten nicht mehr kompensieren.

Die Aufgabe, eine Ausgleichsgerade durch die Punkte  $P_j$  zu finden stellt sich jetzt wie folgt: Man bestimme die Parameter  $a, b$  so, dass der Gesamtfehler  $F(a, b)$  ein Minimum annimmt. Damit die Funktion  $F$  der beiden Variablen  $a, b$  einen Extremwert annimmt, muss notwendig gelten

$$\frac{d}{da} F(a, b) = 0$$

und

$$\frac{d}{db} F(a, b) = 0.$$

Lässt sich aus diesen Bedingungen ein Parameterpaar  $a, b$  eindeutig bestimmen, dann muss die Funktion ein Minimum haben, da sie überall nur Werte  $F(a, b) \geq 0$  annimmt. Man erhält

$$\frac{d}{da} F(a, b) = -2 \sum_{j=1}^n x_j (y_j - ax_j - b) = 0$$

und

$$\frac{d}{db} F(a, b) = -2 \sum_{j=1}^n (y_j - ax_j - b) = 0$$

oder

$$a \sum_{j=1}^n x_j^2 + b \sum_{j=1}^n x_j = \sum_{j=1}^n x_j y_j,$$

und

$$a \sum_{j=1}^n x_j + nb = \sum_{j=1}^n y_j.$$

Bringt man die Mittelwerte

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$$

und

$$\bar{y} = \frac{1}{n} \sum_{j=1}^n y_j$$

der Messwerte  $x_j, y_j$  ins Spiel, dann kann man die zweite Gleichung in der Form

$$(*) \quad a\bar{x} + b = \bar{y}$$

schreiben. Einsetzen von  $b$  in die erste Gleichung ergibt dann

$$a \sum_{j=1}^n x_j^2 + n\bar{x} (\bar{y} - a\bar{x}) = \sum_{j=1}^n x_j y_j$$

oder

$$(**) \quad a \left( \sum_{j=1}^n x_j^2 - n\bar{x}^2 \right) = \sum_{j=1}^n x_j y_j - n\bar{x} \bar{y}.$$

Nun gilt aber

$$\begin{aligned} \sum_{j=1}^n (x_j - \bar{x})^2 &= \sum_{j=1}^n x_j^2 - 2\bar{x} \sum_{j=1}^n x_j + n\bar{x}^2 \\ &= \sum_{j=1}^n x_j^2 - 2n\bar{x}^2 + n\bar{x}^2 \\ &= \sum_{j=1}^n x_j^2 - n\bar{x}^2 \end{aligned}$$

und

$$\begin{aligned} \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}) &= \sum_{j=1}^n x_j y_j - \bar{y} \sum_{j=1}^n x_j \\ &\quad - \bar{x} \sum_{j=1}^n y_j + n\bar{x} \bar{y} \\ &= \sum_{j=1}^n x_j y_j - n\bar{x} \bar{y}. \end{aligned}$$

Mit (\*) und (\*\*) erhalten wir damit den folgenden

**Satz:** Die Ausgleichsgerade  $y = ax + b$  durch die Punkte  $P_j = (x_j, y_j)$ ,  $j = 1, 2, \dots, n$ , hat die Parameter

$$(a) \quad a = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

und

$$(b) \quad b = \bar{y} - a\bar{x}.$$

**Bemerkung:** Die Methode der Parameterbestimmung mit Hilfe einer Ausgleichsgeraden kann auch auf Messgrößen angewandt werden, die nicht in einem linearen Zusammenhang stehen. Hierzu ein Beispiel: Das exponentielle Wachstum einer Population wird durch eine Beziehung

$$(*) \quad x = c e^{at}$$

beschrieben, wobei  $x$  die Größe der Population zur Zeit  $t$  ist und  $a, c > 0$  Parameter sind. Die Beziehung zwischen den Größen  $x$  und  $t$  ist nicht linear.

Geht man auf beiden Seiten von (\*) aber zum Logarithmus über, dann erhält man

$$(**) \quad \ln x = at + \ln c$$

und mit  $y = \ln x$  und  $b = \ln c$  findet man die lineare Beziehung

$$(***) \quad y = at + b$$

zwischen den Größen  $y$  und  $t$ . Misst man daher die Größen  $x_1, x_2, \dots, x_n$  zu den Zeitpunkten  $t_1, t_2, \dots, t_n$  und bestimmt die Ausgleichsgerade bezüglich der Werte  $y_1 = \ln x_1, y_2 = \ln x_2, \dots, y_n = \ln x_n$  und der Zeitpunkte  $t_1, t_2, \dots, t_n$ , dann erhält man mit  $a$  und  $c = e^b$  die gesuchten Parameter des exponentiellen Wachstums der untersuchten Population. ▲

In der Regel besteht zwischen zwei empirisch bestimmbaren Merkmalen  $X$  und  $Y$  kein strenger funktionaler Zusammenhang sondern höchstens ein tendenzieller wie zwischen den Merkmalen  $X = \text{„Körpergröße“}$  und  $Y = \text{„Gewicht“}$ , wo in der Regel das Gewicht mit der Körpergröße zunimmt. Man wird daher nach einem Zusammenhang in dem Sinne suchen, dass zu einer Körpergröße  $x$  ein „Normalwert“  $y(x)$  des Gewichts gehört. Diesen Zusammenhang wird man als linear postulieren, wenn die Form der Punktwolke  $P_j = (x_j, y_j)$ ,  $j = 1, 2, \dots, n$ , gebildet aufgrund der Messreihen aus Werten  $x_j, y_j$  für Körpergröße und Gewicht einer Population, dies als plausibel

erscheinen lässt.

Legt man durch eine solche Punktwolke die zugehörige Ausgleichsgerade  $y = ax + b$ , dann stellt diese den gesuchten tendenziellen Zusammenhang zwischen Körpergröße und Normalgewicht her.

**Beispiel:** An 27 Personen wurde das Merkmal „Körpergröße“ in cm und das Merkmal „Körpergewicht“ (eigentlich Körpermasse) in kg gemessen. Der Punkteschwarm in der Graphik gibt ein anschauliches Bild von der Verteilung der Ausprägungspaare:

$x_j$ (in cm)	155	155	157	159	159	162	163	163	163	163	163	163	164	164	166
$y_j$ (in kg)	47	49	47	50	52	50	45	47	50	55	57	57	52	65	49
$x_j$ (in cm)	167	167	168	168	168	168	169	169	170	170	172	172	174		
$y_j$ (in kg)	53	56	52	58	58	67	55	64	53	61	61	68	66		

Tabelle 10.3: Merkmale Körpergröße und Gewicht

Die Ausgleichsgerade oder Regressionsgerade zu der obigen Wertetabelle hat die Gleichung

$$y = 0,94x - 99,45.$$

Die folgende Abbildung zeigt die zur Wertetabelle gehörige Punktwolke und die zugehörige Regressionsgerade.

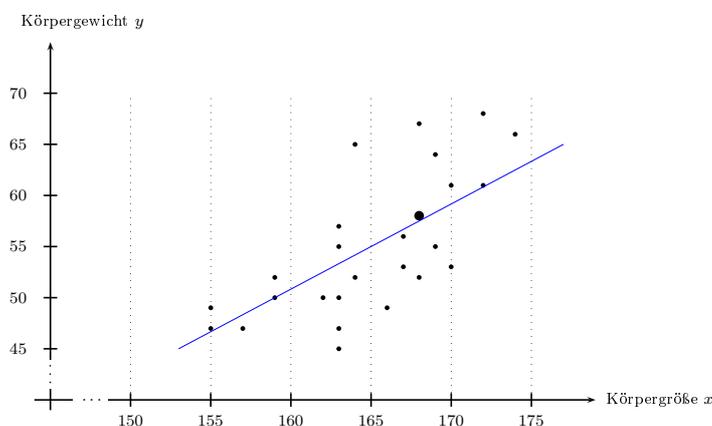


Abbildung 10.10: Punktwolke und Regressionsgerade bezüglich Körpergröße und Gewicht (● = Doppelpunkt).

**Bemerkung:** Stehen die Messgrößen  $X$  und  $Y$  in einem strikten linearen Zusammenhang, so dass die Punkte einer zugehörigen Punktwolke nur aufgrund von Messfehlern nicht auf einer Geraden liegen, so werden die Abweichungen von einer Geraden

in der Regel nicht so groß sein wie bei Abbildung 10.10. Qualitativ ergibt sich jedoch das gleiche Bild. ▲

## 10.9 Der Korrelationskoeffizient

Besteht zwischen zwei Merkmalen  $X$  und  $Y$  ein strenger linearer Zusammenhang

$$y = ax + b, \quad a \neq 0,$$

dann gilt für Messwerte

$$(*) \quad \begin{aligned} x_j, y_j, j &= 1, 2, \dots, n, \\ y_j &= ax_j + b. \end{aligned}$$

und daher ist

$$\sum_{j=1}^n y_j = a \sum_{j=1}^n x_j + nb.$$

Nach Division durch  $n$  folgt die Beziehung

$$(**) \quad \bar{y} = a\bar{x} + b$$

für die jeweiligen Mittelwerte  $\bar{x}, \bar{y}$  der beiden Messreihen.

Bildet man die Differenz der Gleichungen  $(*)$  und  $(**)$ , so findet man die Gleichungen

$$y_j - \bar{y} = a(x_j - \bar{x}), \quad j = 1, 2, \dots, n.$$

Der Vektor  $\tilde{\mathbf{y}} \in \mathbb{R}^n$  mit den Komponenten  $y_j - \bar{y}$  ist also ein skalares Vielfaches des Vektors  $\tilde{\mathbf{x}} \in \mathbb{R}^n$  mit den Komponenten  $x_j - \bar{x}$ . In diesem Fall schließen die Vektoren  $\tilde{\mathbf{x}}$  und  $\tilde{\mathbf{y}}$  den Winkel 0 oder  $\pi$  ein.

Als Maß für den Grad des linearen Zusammenhangs zweier Merkmale  $X$  und  $Y$  sieht man daher den Cosinus des Winkels an, den die Vektoren  $\tilde{\mathbf{x}}$  und  $\tilde{\mathbf{y}}$  einschließen (vgl. Abschnitt 8.1).

**Definition:** Die Zahl

$$r = \frac{\tilde{\mathbf{x}} \cdot \tilde{\mathbf{y}}}{\|\tilde{\mathbf{x}}\| \|\tilde{\mathbf{y}}\|} = \frac{\sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_{j=1}^n (x_j - \bar{x})^2} \sqrt{\sum_{j=1}^n (y_j - \bar{y})^2}}$$

heißt Korrelationskoeffizient der Merkmale  $X$  und  $Y$ .

Die Zahl  $r$  liegt also im Intervall  $-1 \leq r \leq 1$ . Gilt  $\tilde{\mathbf{y}} = a\tilde{\mathbf{x}}$  mit  $a > 0$ , dann ist  $r = 1$ , ist  $a < 0$ , dann ist  $r = -1$ . Liegt  $r$  dicht bei  $-1$  oder  $+1$ , dann sind die Merkmale  $X$  und  $Y$  stark korreliert. Ist  $r$  nahe 0, dann sind  $X$  und  $Y$  nicht korreliert. Bei  $r = 0$  stehen die Vektoren  $\tilde{\mathbf{x}}$  und  $\tilde{\mathbf{y}}$  senkrecht aufeinander.

Dies ist zum Beispiel dann der Fall, wenn die Punkte  $P_j = (x_j, y_j)$  mit gleichen Abständen auf einer Kreislinie liegen. In diesem Fall gilt

$$x_j = c + r \cos \frac{2\pi j}{n},$$

und

$$y_j = d + r \sin \frac{2\pi j}{n}$$

mit  $j = 0, 1, \dots, n-1$ . Offenbar ist dann  $\bar{x} = c$  und  $\bar{y} = d$  und man erhält für das Skalarprodukt

$$\tilde{\mathbf{x}} \cdot \tilde{\mathbf{y}} = \sum_{j=0}^{n-1} (x_j - \bar{x})(y_j - \bar{y}) = r^2 \sum_{j=0}^{n-1} \cos \frac{2\pi j}{n} \cdot \sin \frac{2\pi j}{n}.$$

Mit

$$\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$$

und

$$\sin x = \frac{1}{2i}(e^{ix} - e^{-ix})$$

(vgl. Abschnitt 5.5) folgt dann

$$\begin{aligned} \tilde{\mathbf{x}} \cdot \tilde{\mathbf{y}} &= r^2 \sum_{j=0}^{n-1} \frac{1}{4i} (e^{\frac{2\pi ij}{n}} + e^{-\frac{2\pi ij}{n}}) \cdot (e^{\frac{2\pi ij}{n}} - e^{-\frac{2\pi ij}{n}}) \\ &= \frac{r^2}{4i} \left( \sum_{j=0}^{n-1} e^{\frac{4\pi ij}{n}} - n + n - \sum_{j=0}^{n-1} e^{-\frac{4\pi ij}{n}} \right) \\ &= \frac{r^2}{4i} \left( \frac{e^{4\pi i} - 1}{e^{\frac{4\pi i}{n}} - 1} - \frac{e^{-4\pi i} - 1}{e^{-\frac{4\pi i}{n}} - 1} \right) \\ &= 0. \end{aligned}$$

Zuletzt wurde die Summenformel für die endliche geometrische Reihe angewandt (vgl. Abschnitt 4.4).

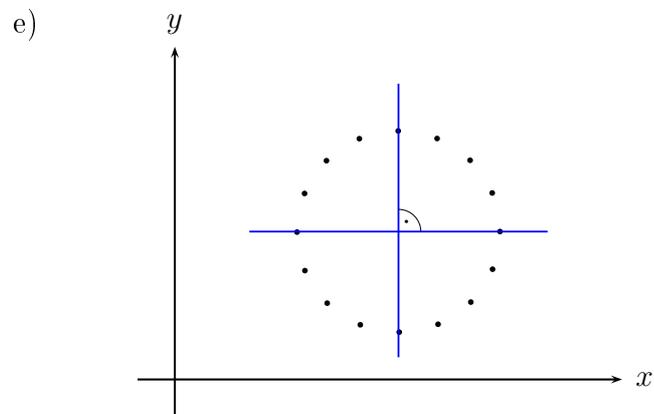


Abbildung 10.11: Punktwolke zu unkorrelierten Merkmalen ( $r = 0$ ).

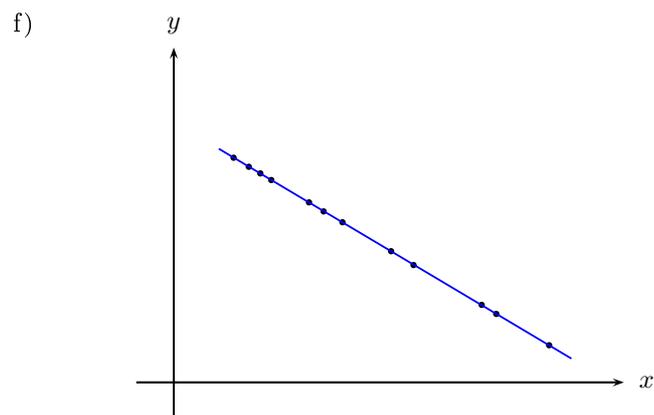


Abbildung 10.12: Punktwolke zu streng korrelierten Merkmalen ( $r = -1$ ).

**Beispiel:** Berechnet man den Korrelationskoeffizienten zu den Merkmalen  $X =$  Körpergröße und  $Y =$  Gewicht mit den Daten aus Tabelle 10.1, dann findet man den Wert  $r = 0,9$ . Die Merkmale „Körpergröße“ mal „Gewicht“ sind also stark korreliert. ■

**Bemerkung:** Liegt der Korrelationskoeffizient  $r$  zweier Merkmale  $X$  und  $Y$  nahe bei 1 oder -1, darf daraus nicht auf eine *kausale* Beziehung zwischen  $X$  und  $Y$  geschlossen werden. (Der Rückgang der Storchpopulation korreliert zwar stark mit dem Rückgang der Geburten, ist aber nicht dessen Ursache.) ▲



# Literaturverzeichnis

- [1] W. Pavel; R. Winkler: *Mathematik für Naturwissenschaftler*. Pearson–Verlag, 2007 (Euro 39,95).
- [2] D. Horstmann: *Mathematik für Biologen*. Spektrum–Verlag, 2008, (Euro 24,95).
- [3] T. Arens et al.: *Mathematik*. Spektrum–Verlag, 2008. 1490 S. (Euro 70,00).