universität wien

GMA

Grundlagen Mathematik und Analysis

Christian Cenker
Gabriele Uchida

Institut für Scientific Computing

© März 2010 (CC, GU)

Inhaltsverzeichnis

Algebra 1

1 Mengen 1

1.1 Mengenalgebra 3

2 Boolesche Algebra und Schaltalgebra 7

2.1 Schaltalgebra 9

3 Elementare Logik 14

3.1 Die Aussagenlogik 15

3.2 Syntax 16

3.3 Modelle 23

3.4 Normalformen 25

3.5 Schlussregeln 29

3.6 Geschichte und Systeme der Logik 31

4 Elementare Gruppentheorie 33

4.1 Permutationen und ihre Verknüpfungen 34

4.2 Permutationsgruppen 35

5 Relationen und Ordnungen 38

5.1 Äquivalenzrelationen 41

5.2 Halbordnungen 43

Analysis 45

6 Reelle Funktionen $f : \mathbb{R} \to \mathbb{R}$ 45

6.1 Stetige Funktionen $f : \mathbb{R} \to \mathbb{R}$ 52

6.2 Differenzierbarkeit 57

6.2.1 Kurvendiskussionen 64

7 Mehrdimensionale Funktionen **67**

7.0.2 Niveaulinien 67

7.1 Stetigkeit 70

7.1.1 Eigenschaften des $\mathbb{R}^n$ 70

7.1.2 Mengen im $\mathbb{R}^n$ 72

7.2 Mehrdimensionale Differenzialrechnung 74

7.2.1 Differentiationsregeln 79

7.2.2 Richtungsableitung 80

7.2.3 Mittelwertsätze 81

7.2.4 Tangentialabbildung 83

7.2.5 Das totale Differential 83

7.2.6 Die Jacobimatrix der Umkehrabbildung 85

7.2.7 Implizite Funktionen 87

8 Kurven **90**

8.1 Die Länge einer Kurve 91

9 Optimierung mehrdimensionaler reeller Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ **95**

9.1 Optimierung ohne Nebenbedingungen 95

9.2 Einfache numerische Optimierungsverfahren 100

9.3 Optimierung mit Nebenbedingungen 102

10 Mehrdimensionale Integration **105**

10.1 Das Riemann-Integral im $\mathbb{R}^n$ 105

10.2 Koordinatentransformation in Integralen 110

11 Differentialgleichungen **113**

11.1 Lösung einfacher Differentialgleichungen 113

11.1.1 $y'(x) = a(x)$ – Einfaches Integrieren 113

11.1.2 $y'(x) = a(x)f(y)$ – Trennung der Variablen 114

11.2 Die Logistische Differentialgleichung 114

11.3 Lineare Differentialgleichungen 115

11.3.1 Lineare Differentialgleichungen erster Ordnung 116

11.3.2 Lösung der homogenen linearen Differentialgleichung erster Ordnung 116
11.3.3 Die partikuläre Lösung 117
11.3.4 Lineare Differentialgleichung zweiter Ordnung 119
11.3.5 Homogene lineare Differentialgleichungen mit konstanten Koeffizienten 119
11.3.6 Inhomogene lineare Differentialgleichung mit konstanten Koeffizienten 123
11.4 Systeme von Differentialgleichungen 124
11.4.1 Systeme linearer Differentialgleichungen 124
11.4.2 Systeme nichtlinearer Differentialgleichungen 125

12 Differenzengleichungen **127**
12.1 Allgemeines 127
12.2 Differenzengleichungen erster Ordnung 128
12.2.1 Lösung per Iteration 128
12.2.2 Die homogene lineare Differenzengleichung 128
12.2.3 Die inhomogene lineare Differenzengleichung 129
12.3 Differenzengleichungen zweiter Ordnung 130
12.3.1 Die homogene lineare Differenzengleichung zweiter Ordnung . . . 130
12.3.2 Die inhomogene lineare Differenzengleichung zweiter Ordnung . . 132

13 Literatur **134**

Mathematische Symbole

$\in$, $\notin$	ist Element, ist nicht Element von
$\forall$	für alle
$\exists$	es existiert ein
$\exists!$	es existiert genau ein
$\nexists$	es gibt kein
:	sodass gilt (als logisches Symbol)
$\neg$	nicht, Negation
$\wedge$	logisches 'und'
$\vee$	logisches 'oder'
$\mapsto$	Zuordnung eines Elements
$\rightarrow$	Mengenzuordnung
$\Rightarrow$	daraus folgt, dass
$\Leftrightarrow$	genau dann, wenn
$\doteq$	ist definiert als
$\cup$, $\bigcup$	vereinigt
$\cap$, $\bigcap$	geschnitten mit
$\subset$, $\subseteq$	ist (echte) Teilmenge von
$\supset$, $\supseteq$	ist (echte) Obermenge von
∞	unendlich
$\sum$	Summe, $\sum_{k=1}^{n} k = 1 + 2 + \cdots + n$
$\prod$	Produkt, $\prod_{k=1}^{n} k = 1 \cdot 2 \cdots n$
$\equiv$	kongruent
$\approx$	ungefähr
$\sim$	äquivalent
$\{\}$, ϕ	die leere Menge
$\mathbb{N}$	die natürlichen Zahlen $\{0, 1, 2, \ldots\}$
$\mathbb{Z}$	die ganzen Zahlen
$\mathbb{Q}$	die rationalen Zahlen (Brüche)
$\mathbb{R}$	die reellen Zahlen
$\mathbb{C}$	die komplexen Zahlen $(a + ib)$
$\mathbb{H}$	die Hilbert'schen Quaternionen
C	die Menge der stetigen Funktionen
C^n	n-mal stetig differenzierbare Funktionen
C^∞	unendlich oft stetig differenzierbare Funktionen
log	der Logarithmus naturalis (in der Schule: ln)
$\log_{10}$	der dekadische Logarithmus ($10er$-Logarithmus)
$\exp(x)$	die Exponentialfunktion (e^x)
π	die Kreiszahl Pi
e	die Eulersche Zahl

Das griechische Alphabet					
A	α	Alpha	N	ν	Ny
B	β	Beta (vita)	Ξ	ξ	Xi
Γ	γ	Gamma	O	o	Omikron
Δ	δ	Delta	Π	π	Pi
E	ε	Epsilon	Σ	σ	Sigma
Z	ζ	Zeta	T	τ	Tau
H	η	Eta	Y	υ	Ypsilon
Θ	θ	Theta	Φ	φ, ϕ	Phi
I	ι	Jota	P	ρ	Rho
K	κ	Kappa	X	χ	Chi
Λ	λ	Lambda	Ψ	ψ	Psi
M	μ	My	Ω	ω	Omega

Ziele und Voraussetzungen

Der „Unterricht“ an einer Universität dient nicht nur der Vermittlung von „Fakten“, sondern soll auch dazu dienen, die Studierenden zu verantwortlichen und kritischen BürgerInnen zu machen, die Gehörtes, Gelesenes oder Gesehenes kritisch hinterfragen und sich nicht nur auf eine Quelle verlassen.

Dazu werden auch gewisse Vorkenntnisse vorausgesetzt, die laut aktuellen Lehrplänen in der Schule vermittelt werden. Im Bereich eines Studiums der Informatik setzen wir zusätzlich voraus, dass die Studierenden mit Computern umgehen und insbesondere Office-Standardpakete sinnvoll verwenden können, sowie einen verantwortungsvollen Umgang mit Internet und E-Mail.

Ziele

Die Studierenden sind nicht KonsumentInnen eines vorgegebenen Wissens, sondern ProduzentInnen ihres Wissens. Lernen ist ein individueller und konstruktiver Prozess und bedeutet aktives Erarbeiten, Erforschen, Darstellen und Reflektieren des Wissens. Insbesondere heißt das auch, dass nicht alles restlos und leicht konsumierbar erklärt wird, bzw. erklärt werden kann, und jede/r einzelne selbst dafür verantwortlich ist, sich zuverlässige Zusatzinformationen zu beschaffen, die sie/er benötigt, um das Dargebotene zu verstehen bzw. anwenden zu können.

Ziel des „Mathematikunterrichts“ ist es nicht, Rechnen zu lernen. Es sollen Kompetenzen vermittelt werden, die für das Fach Informatik von grundlegender Wichtigkeit sind. Dies sind unter Anderem:

- Problemlösefähigkeiten erwerben
- Präzision der Sprachverwendung
- Präzises Beschreiben von Sachverhalten
- Analytisch folgerichtiges Denken
- Arbeiten unter bewusster Verwendung von Regeln
- Logisches, widerspruchsfreies Begründen (Beweisen)
- Argumentieren und exaktes Arbeiten
- Kritisches Denken

Zusätzlich wird in Mathematik versucht, auf Programme bzw. Programmpakete hinzuweisen, bzw., diese zu verwenden, die für eine weitere Verwendung im Bereich der Mathematik und angewandten Mathematik während des Studiums und auch später im Berufsleben sinnvoll erscheinen.

Voraussetzungen

Die folgenden Gebiete sind Teil des Lehrplans der AHS bzw. BHS in Österreich und dienen als Voraussetzung der Inhalte aus Mathematik zur Informatik. Die Beherrschung dieser Inhalte wird vorausgesetzt. Unbekanntes muss selbständig aus Schul- oder Fachbüchern erarbeitet werden, da an der Universität (leider) keine Zeit besteht, den Schulstoff in den Lehrveranstaltungen zu wiederholen.

- Mengen
  - Mengenoperationen
  - Mengenalgebra
- Zahlen
  - Zehnerpotenzen, Zahlendarstellungen
  - rationale Zahlen, Bruchrechnen (6. Schulstufe!)
  - Bruchdarstellung in Dezimaldarstellung umwandeln (und umgekehrt)
  - Komplexe Zahlen
  - Polarkoordinaten
- Analysis
  - reelle Funktionen
    - Beziehungen zwischen Funktionen, Umkehrfunktion
    - Potenzrechnen, Wurzeln, Exponentialrechnung, Logarithmen
  - Grenzwert
  - Stetigkeit
  - Trigonometrie
    - Sinus, Cosinus, Polarkoordinaten
  - Term- und Formelumwandlung
  - Polynome
    - Abspalten von Linearfaktoren
    - Nullstellen
    - Quadratische Gleichungen
  - Folgen und Reihen (Arithmetische und Geometrische Folge/Reihe)
  - Differentialrechnung
    - Präzisieren der Begriffe Grenzwert und Stetigkeit
    - Differentiationsregeln
    - Lösen einfacher Differentialgleichungen
  - Integralrechnung

        * Hauptsatz
        * Flächen und Volumina

- Lineare Algebra
    - Lösen linearer Gleichungssysteme
        * geometrische Interpretation der Gleichungssysteme
- Algebra
    - Variable
    - Umformen von mathematischen Formeln, Gleichungen und Termen
    - Formeln anwenden und interpretieren können
    - Relationen und Äquivalenzrelationen
- Geometrie
    - Darstellungen von ebenen und räumlichen Gebilden
    - Kreis und Kugel
    - Kegelschnitte allgemein (Darstellung und Formeln)

Neben den Voraussetzungen aus Schulmathematik, nehmen wir von Studierenden der Informatik an, dass sie mit E-Mail umgehen können, insbesondere auch ihre Studierenden-UNET-Accounts an der Universität Wien aktiviert haben und benützen, da auf diesem Wege die Informationen zwischen Lehrenden und Lernenden ausgetauscht werden. Insbesondere setzen wir voraus, dass die Studierenden das Internet verantwortlich verwenden und sich Grundkenntnisse im Umgang mit dem Internet (Stichwort-Netiquette) selbständig aneignen bzw. bereits angeeignet haben. Außerdem sollten sie mit Textverarbeitungen, Tabellenkalkulationen und eventuell Funktionsplottern bzw. Computeralgebrasystemen (CAS) umgehen können.

Algebra

Die Algebra befasst sich im weitesten Sinne mit dem „Rechnen mit Buchstaben." Das Wort Algebra geht zurück auf den Titel eines Lehrbuchs des persischen Mathematikers Al-Khwarizmi (9. Jahrhundert, auch das Wort Algorithmus geht auf seinen Namen zurück). Die elementare Algebra beschäftigt sich mit Rechenregeln für verschiedene Zahlenmengen, die abstrakte Algebra mit Strukturen in Mengen, ausgehend von jenen der Zahlenmengen.

Alle Computeralgebrasysteme (CAS) basieren auf diesen Regeln und Strukturen und haben diese implementiert. Numerische Ergebnisse werden entweder direkt mit Hilfe algebraischer Formeln gerechnet, durch Einsetzen von Zahlenwerten in die Formeln, oder, erst wenn dies „fehl schlägt", mit numerischen Näherungsverfahren, die natürlich fehlerbehaftet sind.

1 Mengen

Einer der grundlegenden Begriffe der Mathematik ist der der Menge, ein Begriff, den wir intuitiv erfassen und abhandeln können, den wir aber im mathematischen Sinn nicht richtig definieren können. Das ZFC-Axiomensystem[1] ist eine „logisch widerspruchsfreie Notation" dieser Intuition.

Von Georg Cantor, dem Vater der Mengenlehre, stammt folgende Formulierung:[2]

> *Eine Menge ist eine Zusammenfassung bestimmter wohlunterscheidbarer Objekte unserer Anschauung oder unseres Denkens, welche die Elemente der Menge genannt werden, zu einem Ganzen.*

Eine Menge wird demnach über ihre Elemente erklärt. Dies ist keine (mathematische) Definition, denn der Begriff der *Menge aller Mengen*, also alle Mengen zusammengefasst, führt zu einem Paradoxon (Russel'sche Antinomie), einem Widerspruch im zugrunde liegenden Axiomensystem ZFC, d. h., diese „Menge" ist nicht wohldefiniert. Diese „Menge aller Mengen" existiert somit nicht bzw. ist keine Menge, denn es stellte sich die Frage, ob nun die Menge in sich selbst enthalten ist, oder nicht, ein logischer Widerspruch.

Beispiel 1.1 (Einfache Mengen)
Einfache Mengen sind etwa die Menge der Ziffern $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ oder die Menge aller Studierenden im Hörsaal. Aus der Schule kennen wir auch unendliche Mengen wie $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, und manchmal auch $\mathbb{C}$ und $\mathbb{H}$, oder etwa die Menge der binären Zahlen, die sich aus den Ziffern $\{0, 1\}$ zusammensetzen.
Die *leere Menge* $\phi = \{\}$ ist die Menge, die keine Elemente enthält. In ZFC ist die Existenz dieser verankert.

[1] Zermelo-Fraenkel und Axiom of Choice
[2] Originaldokument unter `http://gdz.sub.uni-goettingen.de/dms/load/img/?IDDOC=36218`

Wir wollen nun aber nicht weiter auf den Begriff der Menge eingehen, da er einerseits bereits in der Schule intensiv behandelt wurde, andererseits wir ein intuitiv richtiges Verständnis für Mengen haben (sollten).

Ebenso wird der Begriff *Element* nicht weiter erklärt bzw. definiert.

Die moderne Mengenlehre und damit ein Großteil der modernen Mathematik basiert auf den Zermelo-Fraenkel Axiomen der Mengenlehre.

Wie grundlegend die Begriffe Menge und Element sind, zeigt uns das mathematische „Konzept" der Funktion, der ohne diese beiden Begriffe „unmöglich" wäre. Wir wollen den Begriff der Funktion im Folgenden entwickeln.

Dazu betrachten wir zuerst Vorschriften, die Elemente oder auch Teilmengen einer Menge Elementen bzw. Teilmengen anderer Mengen zuordnen. Solche Vorschriften nennen wir allgemein Relationen (siehe später), zwei oder mehr Elemente stehen dann in Relation zueinander.

Abbildungen wiederum sind spezielle Relationen, die Elementen der einen Menge ein oder mehrere Elemente einer anderen Menge zuordnen oder auch nicht. Funktionen sind wiederum spezielle Abbildungen, die jedem Element einer Menge genau ein Element einer anderen Menge zuordnen. Wir wollen hier eine bzw. die mathematische widerspruchsfreie Definition geben. Dazu benötigen wir einige Voraussetzungen.

DEFINITION 1.1
Die *Produktmenge* oder das *kartesische Produkt* $A \times B$ zweier Mengen A und B ist definiert als die Menge aller möglichen geordneten Paare (a, b), mit $a \in A$ und $b \in B$. Symbolisch:

$$A \times B \doteqdot \{(a, b) \mid a \in A, b \in B\}.$$

Ein *geordnetes Paar* ist definiert als $(a, b) \doteqdot \{a, \{a, b\}\}$.

D. h., auch ein geordnetes Paar ist eine Menge.

DEFINITION 1.2
Gegeben seien zwei Mengen A und B. Jede Teilmenge $C \subseteq A \times B$ heißt *Graph*.

DEFINITION 1.3
Sei $R \subseteq A \times B$. R definiert eine *Relation* auf A und B. Für alle $(a, b) \in R$ gilt: aRb, d. h., $a \in A$ steht in Relation zu $b \in B$.

Diese Begriffe benötigen wir, um eine Funktion als spezielle Relation definieren zu können. Relationen, insbesondere spezielle Relationen wie Ordnungen, Äquivalenzrelationen und deren Eigenschaften werden später behandelt.

DEFINITION 1.4
Eine *Funktion* $f : A \to B$ ist eine Relation, die jedem $a \in A$ genau ein $b \in B$ zuordnet. Wir schreiben $f(a) = b$.

Da eine Funktion eine spezielle Relation ist und Relationen über ihre Graphen festgelegt sind, ist auch eine Funktion über ihren Graphen festgelegt.

Wir beachten hier, dass in keiner der obigen Definitionen eine Aussage getroffen wurde, welcher Art die Elemente der Mengen A und B sind. Insofern ist der Begriff der Relation und auch jener der Funktion ein Abstraktum.

1.1 Mengenalgebra

Wir können nun nicht nur mit Elementen einer Menge wie den reellen Zahlen rechnen, sondern auch mit Mengen selbst. Wir können Vereinigungen, Durchschnitte und Komplemente von Mengen berechnen. Diese „Rechnungen" folgen den fixen Regeln der Mengenalgebra.

BEISPIEL 1.2 (WAHRSCHEINLICHKEITSRECHNUNG)
Das *Experiment* Würfeln hat als Ergebnis den *Ereignisraum* $S = \{1, 2, 3, 4, 5, 6\}$, das ist nichts anderes als die Menge aller möglichen *Ausfälle*. Jedes Element dieser Menge S ist ein *Elementarereignis*. *Ereignisse* sind Teilmengen des Ereignisraumes S, also etwa das Ereignis, eine gerade Zahl zu würfeln $E_1 = \{2, 4, 6\}$, oder das Ereignis, eine Zahl größer als 3 zu würfeln $E_2 = \{4, 5, 6\}$. Das Ereignis eine gerade Zahl größer als 3 zu würfeln ist nun $E_3 = E_1 \cap E_2 = \{4, 6\}$, der Durchschnitt der beiden Ereignisse E_1 und E_2.

Die wahrscheinlichkeitstheoretische Bewertung der Ereignisse, d. h., die Wahrscheinlichkeit des Eintretens dieser Ereignisse kann einfach aus der Anzahl der Elemente der Ereignisse berechnet werden:

$$P(E_1) = \frac{|E_1|}{|S|} = \frac{3}{6} = \frac{1}{2} \qquad \text{und} \qquad P(E_3) = \frac{2}{6} = \frac{1}{3} .$$

In diesem Beispiel wurde das Symbol $|A|$ für die *Mächtigkeit* einer Menge verwendet, d. h., die Anzahl ihrer Elemente. Über diese Anzahl der Elemente konnten wir Wahrscheinlichkeiten berechnen, nachdem wir die Rechenregeln für Mengen, die Mengenalgebra, angewendet haben.

Wir wollen hier gleich einige Definitionen, d. h., widerspruchsfreie (engl. *sound*) Begriffserklärungen, geben.

DEFINITION 1.5 (MÄCHTIGKEIT)
Die Mächtigkeit $|A|$ oder $\sharp A$ einer Menge A ist die Anzahl ihrer Elemente.
$|A| = n$, wenn es eine bijektive Abbildung der Menge A in die Menge $\{1, 2, \ldots, n\}$ gibt.
Zwei Mengen A und B heißen *gleich mächtig*, d. h. $|A| = |B|$, wenn es eine bijektive Abbildung zwischen A und B gibt.

In diesem Sinn sind die Mengen $\mathbb{N}$ und $\mathbb{Q}$ gleich mächtig, d. h., haben gleich viele Elemente, obwohl dies, naiv gesprochen, nicht offensichtlich ist!

BEISPIEL 1.3
Die Mächtigkeit kann nun, wie im letzten Beispiel, zur Berechnung von Wahrscheinlichkeiten herangezogen werden:

$$\begin{aligned} P(E_3) &= P(E_1 \cap E_2) = \frac{|E_1 \cap E_2|}{|S|} = \frac{1}{3} \\ P(E_1 \cup E2) &= \frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1| + |E_2| - |E_1 \cap E_2|}{|S|} = \frac{3+3-2}{6} = \frac{2}{3}. \end{aligned}$$

Über diese intuitiven „Rechenregeln" kann nun die Axiomatik für die Wahrscheinlichkeitsrechnung aufgebaut werden (Axiome von Kolmogorov[3]).

Der Vollständigkeit halber die folgenden Definitionen, die allgemein bekannt sein sollten.

DEFINITION 1.6 (MENGEN, MENGENOPERATIONEN)
B ist eine *Teilmenge* der Menge A, $B \subseteq A$, wenn gilt: $x \in B \Rightarrow x \in A$.
B ist eine *echte Teilmenge* der Menge A, $B \subset A$, wenn gilt: $x \in B \Rightarrow x \in A$ und $B \neq A$.
Zwei Mengen heißen *disjunkt*, wenn sie keine gemeinsamen Elemente haben.
Die *Potenzmenge* $P(A)$ einer Menge A ist die Menge aller Teilmengen von A.
Das *Komplement* von A in einer Grundmenge G: $A^c \doteqdot \{x \,|\, x \in G \wedge x \notin A\}$ $(A^c = \bar{A})$

Die *Vereinigungsmenge* von A und B: $A \cup B \doteqdot \{x \,|\, x \in A \vee x \in B\}$.
Die *Durchschnittsmenge* von A und B: $A \cap B \doteqdot \{x \,|\, x \in A \wedge x \in B\}$.
Die *Differenzenmenge* von A und B: $A \setminus B \doteqdot \{x \,|\, x \in A \wedge x \notin B\}$.
Die *symmetrische Differenz* von A und B: $A \Delta B \doteqdot (A \setminus B) \cup (B \setminus A)$.

Wir wollen nun mit diesen Mengenoperationen rechnen und die intuitiv erfassbaren Regeln aufschreiben, denn es ist möglich, wenn wir ein „System" von Mengen haben, eine immer wiederkehrende Struktur in diesem System zu erkennen. Wir wollen diese nun definieren.

[3] Andere Transkriptionen: Kolmogorow, Kolmogoroff
siehe auch: `http://en.wikipedia.org/wiki/Andrey_Kolmogorov`

DEFINITION 1.7 (VERBAND)
$(V, \cup, \cap, ^c)$ heißt *Verband*, wenn für alle Mengen A, B, $C \in V$ gilt:

- Kommutativgesetz: $A \cup B = B \cup A$ und $A \cap B = B \cap A$
- Assoziativgesetz: $A \cup (B \cup C) = (A \cup B) \cup C$ und $A \cap (B \cap C) = (A \cap B) \cap C$
- Distributivgesetze: $A \cup (B \cap C) = (A \cap B) \cup (A \cap C)$ und $A \cap (B \cup C) = (A \cup B) \cap (A \cup C)$
- Neutrale Elemente: $A \cup \phi = A$ und $A \cap V = A$
- Inverses Element: $A \cup A^c = V$ und $A \cap A^c = \phi$
- Gesetze von DeMorgan: $(A \cup B)^c = A^c \cap B^c$ und $(A \cap B)^c = A^c \cup B^c$
- Idempotenz: $A \cup A = A$ und $A \cap A = A$
- Absorption: $A \cup (A \cap B) = A$ und $A \cap (A \cup B) = A$

BEMERKUNG: Statt Verband wird oft auch der Begriff *Mengenalgebra* verwendet. Die Gesetze von DeMorgan, sowie Idempotenz und Absorption sind bereits Folgerungen aus dem Rest und müssten oben nicht angeführt werden.

BEISPIEL 1.4
Einfache Beispiele von Verbänden ergeben sich, wenn wir, ausgehend von einer beliebigen Menge A, ihre Potenzmenge $P(A)$ betrachten. $(P(A), \cup, \cap)$ ist dann ein Verband mit Null ϕ und Eins A. Hier haben wir ein endliches Mengensystem.
Ist die Mächtigkeit von $|A| = n$, dann ist $|P(A)| = 2^n$ (Beweis als Übung).

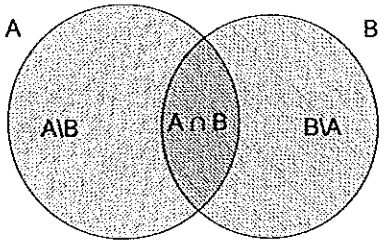
$(P(\mathbb{N}), \cup, \cap)$ ist ein Beispiel eines Verbandes mit unendlich vielen Mengen, mit Null ϕ und Eins $\mathbb{N}$.

BEMERKUNG: $|\mathbb{N}| = \aleph_0$ und $|P(\mathbb{N})| = 2^{\aleph_0} = \mathfrak{c} = |\mathbb{R}|$.
D. h. $\aleph_0$ (sprich Aleph-Null) ist die Anzahl der natürlichen Zahlen und $\mathfrak{c}$ ist die Mächtigkeit des Continuums, d. h., die Anzahl der reellen Zahlen.
$\aleph_0$ und $\mathfrak{c}$ sind sogenannte *Ordinalzahlen.*

BEISPIEL 1.5
$(P(\{0,1\}), \cup, \cap, ^c)$ ist ein weiterer Verband. Er ist der kleinste sinnvolle Verband und besteht nur aus den vier Mengen ϕ, $\{0\}$, $\{1\}$, $\{0,1\}$.

Es gilt: $\{0\}^c = \{1\}$ und $\{1\}^c = \{0\}$.

Zur Veranschaulichung von Mengenoperationen dienen *Venndiagramme*, die das Rechnen in der Mengenalgebra meist verständlicher machen.



Mit Hilfe der Venn-Diagramme können die folgenden Mengentheoreme (Gesetze für das Rechnen mit Mengen) leicht veranschaulicht werden. Als Beweis gelten diese Veranschaulichungen jedoch nicht, Beweise können (bzw. dürfen) nur mit Hilfe der Mengenaxiome ZFC und der Logik geführt werden.

THEOREM 1.1
Sei $(V, \cup, \cap)$ ein Verband, so gilt:

- $A \cup A^c = V$ und $A \cap A^c = \{\}$.
- $(A \cup B) \setminus C = (A \setminus C) \cup (B \setminus C)$.
- $(A \cap B) \setminus C = (A \setminus C) \cap (B \setminus C)$.
- $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$.
- $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.
- $A \setminus (B \setminus C) = A \cap B$.
- $A \Delta B = B \Delta A$.
- $A \Delta (B \Delta C) = (A \Delta B) \Delta C$.
- $(A \Delta B) \cap C = (A \cap C) \Delta (B \cap C)$.

Es können noch unzählige Theoreme bewiesen werden, insbesondere, wenn mehr Mengen daran beteiligt sind. Diese Mengenoperationen bilden mit eine Grundlage der Mathematik. Mit dem Verständnis der Mengenalgebra lassen sich viele „abstrakte" Sätze (Theoreme) der Mathematik leichter verstehen. Wir wollen nun aber Richtung Anwendungen in der Informatik gehen.

2 Boolesche Algebra und Schaltalgebra

Nachdem wir mit Mengen gerechnet haben und die algebraische Struktur kennengelernt haben, wollen wir dies verallgemeinern, d. h., weg von Mengenoperationen gehen. Als Ergebnis werden wir eine Struktur erhalten, die die Grundlage für das Verstehen (und das Entstehen) digitaler Schaltungen bildet. Sie ist somit die Grundlage für das Entstehen des modernen Computers und die Voraussetzung für unsere digitale Welt. Die Wurzeln führen uns in das 19. Jahrhundert zu George Boole (1815–1864). Er war ein Mathematiker, der sich mit Logik beschäftigte und die Boolesche Algebra definierte (fand/erfand?). Da er noch weit weg von industriellen Anwendungen war (und auch von Mengenalgebren) und das Konzept auch auf andere als die Schaltalgebra bzw. Logik angewendet werden kann, werden wir zuerst allgemeine, „inhaltsleere" Symbole verwenden und eine Boolesche Algebra, einen Booleschen Verband, definieren.

DEFINITION 2.1 (BOOLESCHE ALGEBRA)
Gegeben eine Menge M mit den folgenden drei Verknüpfungen:

$\oplus$:	$M \times M \to M;\ (a,b) \mapsto a \oplus b$	Boolesche Summe
$\otimes$:	$M \times M \to M;\ (a,b) \mapsto a \otimes b$	Boolesches Produkt
$\sim$:	$M \to M;\ a \mapsto\sim a$	Boolesches Komplement

$(M, \oplus, \otimes, \sim)$ heißt *Boolesche Algebra*, wenn für alle a, b, $c \in M$:

- Kommutativgesetz: $a \oplus b = b \oplus a$ und $a \otimes b = b \otimes a$
- Assoziativgesetz: $a \oplus (b \oplus c) = (a \oplus b) \oplus c$ und $a \otimes (b \otimes c) = (a \otimes b) \otimes c$
- Distributivgesetze: $a \oplus (b \otimes c) = (a \oplus b) \otimes (a \oplus c)$ und $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$
- Neutrale Elemente 0 und 1: $a \oplus 0 = a$ und $a \otimes 1 = a$
- Komplemente: $a\oplus \sim a = 1$ und $a\otimes \sim a = 0$

BEMERKUNG: $\oplus$ und $\otimes$ sind binäre Verknüpfungen (*binäre Operatoren*), d. h., sie wirken auf zwei *Operanden*. $\sim$ ist ein *unärer Operator*, er wirkt auf (genau) einen Operanden.

Dualitätsprinzip. Durch Vertauschen der Operatoren $\oplus$ und $\otimes$ in obigen Gesetzen erhalten wir jeweils das andere entsprechende Gesetz. Dies ist Folge des sogenannten *Dualitätsprinzips*, d. h., zu jeder „Formel" bzw. zu jedem Theorem gibt es eine duale Formel, die wir durch Vertauschen der beiden Operatoren und der neutralen Elemente 0 und 1 erhalten.

Rechenregeln. Für Boolesche Algebren gelten folgende Gesetze, die sich aus der Definition ableiten lassen:

- Gesetze von DeMorgan: $\sim (a \ominus b) =\sim a\otimes \sim b$ und $\sim (a \otimes b) =\sim a\oplus \sim b$.
- Idempotenz: $a \oplus a = a$ und $a \otimes a = a$.
- Absorption: $a \oplus (a \otimes b) = a$ und $a \otimes (a \oplus b) = a$.
- Inverse: $\sim 0 = 1$ und $\sim 1 = 0$, sowie $a \oplus 1 = 1$ und $a \otimes 0 = 0$.

Eine Boolesche Algebra für sich ist ein rein abstrakter Begriff ohne Inhalte.
Wir haben jedoch schon ein Modell für eine solche Boolesche Algebra kennengelernt, nämlich Mengenalgebren:

Sei M eine Menge, $P(M)$ ihre Potenzmenge, so bildet $(P(M), \cup, \cap, ^c)$ eine Boolesche Algebra. D. h. wir ersetzen:

$\oplus$	durch	$\cup$	Vereinigung von Mengen
$\otimes$	durch	$\cap$	Durchschnitt von Mengen
$\sim$	durch	^{c}	Komplement von Mengen

und haben $0 = \phi$ und $1 = M$.

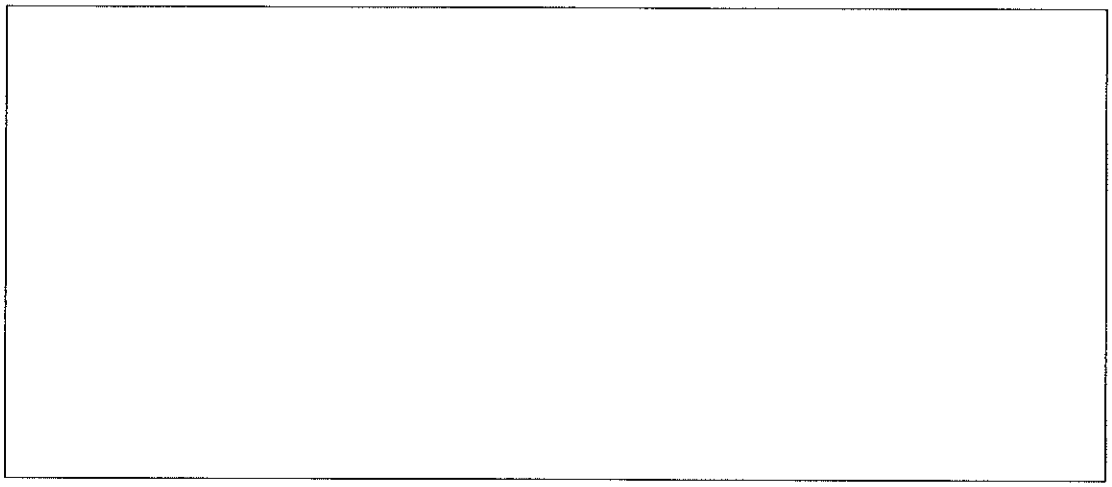
Für jede Mengenalgebra gelten obige Rechenregeln der Booleschen Algebra, wie wir im letzten Abschnitt ja schon darlegten. Die Rechenregeln für die symmetrische Differenz von Mengen folgen ebenfalls, wenn wir betrachten, dass

$$A \Delta B = (A \cup B) \setminus (A \cap B)$$

oder besser, da die Rechenvorschrift „\" im obigen System nicht direkt definiert ist

$$A \Delta B = (A^c \cup B^c) \cap (A \cup B)^c,$$

d. h. $A \setminus B = A \cap B^c$. Skizzieren Sie Venn-Diagramme zur Veranschaulichung.

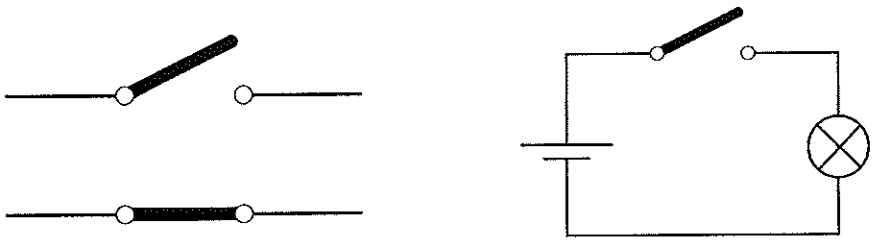


Während Mengenalgebren sehr viele Elemente haben können, interessieren uns digitale Anwendungen. Im Speziellen bildet auch die (zweiwertige) Aussagenlogik mit der Grundmenge {wahr, falsch} ein Modell einer Booleschen Algebra. Ein weiteres Modell ist die Schaltalgebra.

2.1 Schaltalgebra

Als Grundmenge für die Schaltalgebra dient die Menge $\mathbb{B} = \{0, 1\}$. Das sind genau die Zustände, die ein Schalter haben kann: *Zu* und *Offen* bzw. *Strom fließt* oder *Strom fließt nicht.* Auch Daten auf digitalen Trägern werden als 0 und 1 gespeichert bzw. die gespeicherten Daten so interpretiert, denn die Speicherung auf DVDs etwa findet nicht durch die Zahlen 0 und 1 statt.

Diese *Schaltzustände* können durch *bistabile Schaltelemente* (Ein-Aus-Schalter) dargestellt werden, die Variablen, die genau die zwei Zustände 0 und 1, die *Schaltkonstanten*, haben, heißen *binäre Schaltvariablen.* Funktionen $f : \mathbb{B}^n \to \mathbb{B}$ heißen *binäre Schaltfunktionen*, ihre technischen Realisierungen heißen *Schaltungen.*



Schalterstellungen und Layout einer Schaltung

Die *Schaltalgebra* $(P(\mathbb{B}), \vee, \wedge, \neg)$ mit $\mathbb{B} = \{0, 1\}$ ist eine Boolesche Algebra. D. h. wir ersetzen:

$\oplus$	durch	$\vee$	logisches UND
$\otimes$	durch	$\cap$	logisches ODER
$\sim$	durch	$\neg$	logisches NON

und haben die Null 0 und die Eins 1.

BEMERKUNG: In der Schaltalgebra sind für die Negation mehrere Notationen üblich:

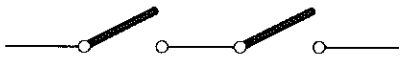
$$' \text{ oder } {}^{-} \text{ oder } \neg \quad \text{also} \quad 0' = \bar{0} = \neg 0 = 1 \quad \text{und} \quad 1' = \bar{1} = \neg 1 = 0$$

Das hat seinen Ursprung in der Nähe von Schalt- und Mengenalgebra zur Aussagenlogik. Die Negation bindet am stärksten, da ja $\neg a = \bar{a} = a'$ ein eigenständiges Element ist.

Es gelten dann alle Rechenregeln einer Booleschen Algebra, insbesondere natürlich

- Kommutativgesetze: $a \vee b = b \vee a$ und $a \wedge b = b \wedge a$
- Assoziativgesetze: $a \vee (b \vee c) = (a \vee b) \vee c$ und $a \wedge (b \wedge c) = (a \wedge b) \wedge c$
- Distributivgesetze: $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$ und $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$
- Neutrale Elemente 0 und 1: $a \oplus 0 = a$ und $a \wedge 1 = a$
- Komplemente: $a \vee \neg a = 1$ und $a \wedge \neg a = 0$

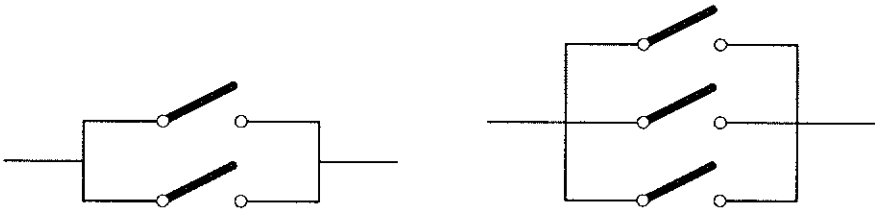
$\wedge$ – Serienschaltung



Serienschaltung $a \wedge b$

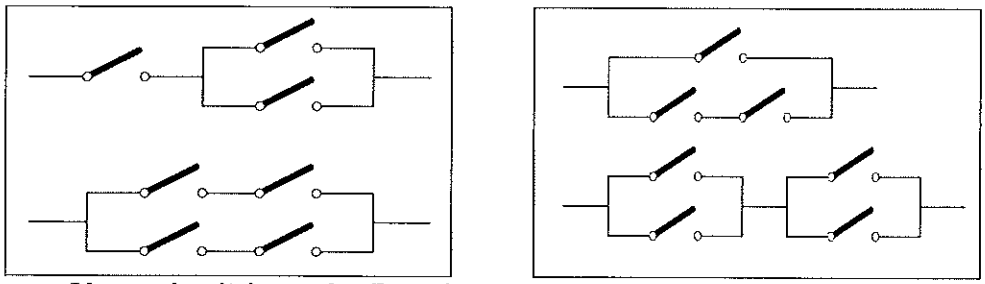
Serienschaltung $a \wedge b \wedge c$

$\vee$ – Parallelschaltung



Parallelschaltungen $a \vee b$ und $a \vee b \vee c$

Die Distributivgesetze



Veranschaulichung der Distributivgesetze durch Schaltfunktionen

Schalttabellen

Schaltfunktionen können nicht nur durch diese Diagramme veranschaulicht werden, sondern auch durch *Schalttabellen*, die im Grunde genommen wie die Wahrheitstabellen aus der Aussagenlogik aussehen. Mit Hilfe der Schalttabellen können Schaltfunktionen nicht nur berechnet, sondern auch vereinfacht werden.

Die logischen Funktionen NON, UND und ODER:

a	b	$\neg a$	$\neg b$	$a \wedge b$	$a \vee b$
0	0	1	1	0	0
0	1	1	0	0	1
1	0	0	1	0	1
1	1	0	0	1	1

Die Distributivgesetze:

a	b	c	$a \vee (b \wedge c)$	$(a \vee b) \wedge (a \vee c)$	$a \wedge (b \vee c)$	$(a \wedge b) \vee (a \wedge c)$
0	0	0	0	0	0	0
0	0	1	0	0	0	0
0	1	0	0	0	0	0
0	1	1	1	1	0	0
1	0	0	1	1	0	0
1	0	1	1	1	1	1
1	1	0	1	1	1	1
1	1	1	1	1	1	1

Die Gesetze von DeMorgan (Übung)

a	b	$a \wedge b$	$\neg(a \wedge b)$	$\neg a$	$\neg b$	$\neg a \vee \neg b$
0	0					
0	1					
1	0					
1	1					

a	b	$a \vee b$	$\neg(a \vee b)$	$\neg a$	$\neg b$	$\neg a \wedge \neg b$
0	0					
0	1					
1	0					
1	1					

Bindungsregel

Es sollte aufgefallen sein, dass die Regeln zur Definition von Booleschen Algebren sehr ähnlich den Rechenregeln in $\mathbb{Q}$ oder $\mathbb{R}$ sind, wenn wir $\oplus$ und $\otimes$ durch die übliche Addition und Multiplikation ersetzen. Wir wissen, dass es dort die Rechenregel „Multiplikation vor Addition“ (Punktrechnung vor Strichrechnung) gibt.

In Booleschen Algebren ist dies ebenso: Die *Bindungsregel* besagt, dass die UND-Verknüpfung stärker bindet als die ODER-Verknüpfung.

Die folgende Schalttabelle sollte dies verdeutlichen (Übung).

a	b	c	$a \wedge b$	$b \vee c$	$(a \wedge b) \vee c$	$a \wedge (b \vee c)$	$a \wedge b \vee c$
0	0	0					
0	0	1					
0	1	0					
0	1	1					
1	0	0					
1	0	1					
1	1	0					
1	1	1					

Normalformen

Logische Schaltungen können auch dahingehend vereinfacht werden, dass man jede einzelne Schaltung nur mittels der Operatoren $\wedge$ und $\neg$ aufbauen kann, oder nur mittels $\vee$ und $\neg$.

Die kanonische *konjunktive Normalform* (KNF) einer logischen Schaltung besteht aus Konjunktionen ($\wedge$) von Maxtermen. *Maxterme* (*Volldisjunktionen*) sind logische Verknüpfungen von Schaltvariablen, die nur Disjunktionen ($\vee$) und Negationen ($\neg$) enthalten. Die Bezeichnung Maxterm kommt daher, da diese als Ergebnis das Maximum aller beteiligten Schaltvariablen hat.

Die kanonische *disjunktive Normalform* (DNF) einer logischen Schaltung besteht nur aus Disjunktionen von Mintermen. *Minterme* (*Volldisjunktionen*) sind logische Verknüpfungen von Schaltvariablen, die nur Konjunktionen ($\wedge$) und Negationen ($\neg$) enthalten. Die Bezeichnung Minterm kommt daher, da diese als Ergebnis das Minimum aller beteiligten Schaltvariablen hat.

Maxterme und Minterme müssen alle beteiligten Schaltvariablen enthalten.

Zur Veranschaulichung der Berechnung von KNF, DNF, Maxtermen und Mintermen soll die folgende Schalttabelle dienen.

Wir nehmen eine (beliebige) logische Funktion $f(a, b, c)$ an, deren Werte in der Schalttabelle angegeben sind, und berechnen den Minterm und den Maxterm aus den Werten der Funktion:

a	b	c	$f(a,b,c)$	Minterm	Maxterm
0	0	0	0		$a \vee b \vee c$
0	0	1	0		$a \vee b \vee \neg c$
0	1	0	1	$\neg a \wedge b \wedge c$	
0	1	1	0		$a \vee \neg b \vee \neg c$
1	0	0	1	$a \wedge \neg b \wedge \neg c$	
1	0	1	1	$a \wedge \neg b \wedge c$	
1	1	0	0		$\neg a \vee \neg b \vee c$
1	1	1	1	$\neg a \wedge \neg b \wedge \neg c$	

DNF: $f(a, b, c) = (\neg a \wedge b \wedge c) \vee (a \wedge \neg b \wedge \neg c) \vee (a \wedge \neg b \wedge c) \vee (\neg a \wedge \neg b \wedge \neg)$

KNF: $f(a, b, c) = (a \vee b \vee c) \wedge (a \vee b \vee \neg c) \wedge (a \vee \neg b \vee \neg c) \wedge (\neg a \vee \neg b \vee c)$

Aus der obigen Schalttabelle ist ersichtlich, wie wir bei der Berechnung vorgehen. Wichtig ist, dass bei Belegung der Schaltvariablen a, b, c mit den Werten 0 und 1 sowohl Minterm als auch Maxterm den richtigen Funktionswert (a, b, c) als Ergebnis liefern. Dasselbe gilt auch für die Zusammenfassung zur KNF und DNF.

Übungen

Stellen Sie die Wahrheitstabelle folgender logischer Ausdrücke auf, vereinfachen Sie diese Ausdrücke wenn möglich und skizzieren Sie die zugehörigen Schaltfunktionen:

$a \vee (a \wedge b)$
$a \wedge (a \vee b)$
$a \vee (\neg a \wedge b)$
$a \wedge (\neg a \vee b)$
$(a \vee b) \wedge (a \vee \neg b)$
$(a \wedge b) \vee (a \wedge \neg b)$
$(a \vee b) \wedge (\neg a \vee b)$
$(a \wedge b) \vee (\neg a \wedge b)$
$(a \wedge b) \vee (\neg a \wedge \neg b)$
$a \wedge \neg b \wedge c$
$\neg a \wedge \neg b \wedge \neg c$
$(a \vee b \vee c) \wedge (\neg a \vee b \vee c)$
$(a \wedge b \wedge c) \vee (\neg a \wedge b \wedge)$
$\neg(\neg a \vee \neg b) \wedge \neg(\neg a \wedge \neg b)$
$(a \vee b \vee \neg c) \vee (\neg a \vee \neg b \vee \neg c) \vee (a \wedge b \wedge c) \vee (\neg a \wedge \neg b \wedge c)$
$(\neg a \wedge \neg b \wedge \neg c \wedge \neg d) \vee (\neg a \wedge b \wedge \neg c \wedge \neg d) \vee (\neg a \wedge \neg b \wedge \neg c \wedge d) \vee (\neg a \wedge b \wedge \neg c \wedge d)$

3 Elementare Logik

Die Mathematik basiert im Wesentlichen auf drei Säulen:

- Der strengen Axiomatik, d. h., Axiome legen grundlegende Regeln fest.
- Der zweiwertigen Prädikatenlogik, die aus den Axiomen mit festgelegten logischen Regeln die mathematischen Aussagen herleitet bzw. überprüft, ob getätigte mathematische Aussagen mit den zugrunde liegenden Axiomen vereinbar sind, d. h., entweder beweisbar oder widerlegbar.[4]
- Der Mengenlehre, die wiederum auf den Begriffen Menge und Element basiert und einem Axiomensystem „gehorcht" (Zermelo-Fraenkel = ZF oder ZFC).

Die Mathematik ist somit eine formale Sprache, wie die (Prädikaten-) Logik, deren Aussagen und Rechenregeln sich logisch von Axiomen herleiten lassen. Darin unterscheidet sie sich wesentlich von der natürlichen Sprache, sowohl Hochsprache wie auch Umgangssprache. Auch die natürliche Sprache folgt anscheinend festen Regeln, z. B. ihrer Grammatik, jedoch gibt es immer wieder Ausnahmen wie zum Beispiel unregelmäßige Verben, Lautverschiebungen und, insbesondere bei der Umgangssprache, regionale Unterschiede. Außerdem können sich diese Regeln ändern, wie etwa bei einer Rechtschreibreform.

Dies ist bei der Mathematik unmöglich. Die Axiome ändern sich nicht, die Logikregeln sind ebenfalls für alle Zeit festgeschrieben[5], es gibt keine regionalen Unterschiede, die Mathematik ist eine universelle Sprache.

Wir werden uns nun eingehender mit einigen grundlegenden Begriffen aus der Logik beschäftigen.

Für die Mathematik sind die Aussagenlogik und die Prädikatenlogik interessant. Es gibt aber auch andere Logiken wie die Modallogik, Zeitlogik etc.

Was untersucht die Logik?

Die Logik untersucht Sprachen, in denen die Welt beschrieben werden kann. Dabei müssen rationales Argumentieren sowie gerechtfertigte Schlüsse und Argumente möglich sein.

Beispiele:

➤ Wenn eine Zahl durch 3 teilbar ist, dann ist ihre Quersumme durch 3 teilbar.
Die Quersumme einer zu prüfenden Zahl ist nicht durch 3 teilbar.
Also ist die Zahl nicht durch 3 teilbar.

[4] Diese Aussage ist nicht ganz richtig, siehe später: Gödels Unvollständigkeitssatz.

[5] Axiomensysteme können erweitert werden, einzelne Axiome könnten verworfen oder geändert werden, wenn sie sich als zum Rest des Axiomensystems widersprüchlich oder aber daraus ableitbar erwiesen – dies ist jedoch unwahrscheinlich, da die Widerspruchsfreiheit des gültigen Axiomensystems ZFC bewiesen ist.

➢ Wenn Lackmuspapier in Säure getaucht wird, färbt es sich rot.
Das Lackmuspapier, in eine Testflüssigkeit getaucht, verfärbt sich nicht rot.
Also handelt es sich dabei um keine Säure.

Es handelt sich bei dieser Schlussart um den sogenannten *modus tollens*.

modus tollens

Wenn P dann Q. Und nicht Q. Also: nicht P.

Dieser logische Schluss kann jedoch auch als Formel der Aussagenlogik formuliert werden:

$$((P \to Q) \wedge \neg Q) \to \neg P \qquad \text{oder:} \qquad \frac{P \to Q \quad \neg Q}{\neg P}$$

Wozu benötigen wir die Logik ?

➢ Beschreibung von Welten

➢ Überprüfung von Theoremen (allgemeine Aussagen)

➢ Verifikation von Soft- und Hardware

➢ Programmierung: Prolog, Algol, Lisp

➢ Künstliche Intelligenz

➢ Wissensbasierte Systeme (Knowledge Management)

➢ Konsistenz von Datenbanken

3.1 Die Aussagenlogik

Aussagen (P, Q,...) können wahr oder falsch sein, d. h., die Werte ≫Wahr≪ (w, 1) oder ≫Falsch≪ (f, 0) annehmen. Wir werden im Folgenden die in der Informatik üblichen Werte 1 und 0 verwenden.

In der Mathematik und Informatik wird das ≫**tertium non datur**≪ als gültig anerkannt, d. h., eines von beiden, w oder f, trifft immer zu, nie aber beides zugleich.

Beispiele

➢ 6 ist durch 3 teilbar. - wahr

➢ 7 ist durch 5 teilbar. - falsch

➢ Am 24. April 2000 regnete es in Wien. - wahr/falsch ?

Andere Wissenschaftsdisziplinen verwenden jedoch beispielsweise die **Fuzzy Logic**, in der das *tertium non datur* nicht gilt. Dort werden Eigenschaften über Zugehörigkeitsfunktionen definiert. Es ist zum Beispiel der Begriff ≫groß≪ relativ.

Ein Objekt kann nun etwa zu 75 % groß sein, gleichzeitig aber auch zu 25 % klein. Die Fuzzy Logic entspricht mehr unserer Alltagssprache (Fuzzy Mengenlehre).

Aussageformen

Aussageformen enthalten Individuen-Variablen. Werden die Individuen-Variablen durch Objekte (z. B., Zahlen-Werte) ersetzt, so ergeben sich Aussagen, die entweder wahr oder falsch sind.

Aussageformen sind weder wahr noch falsch, sie können jedoch allgemein gültig sein, wenn sich bei beliebigem Einsetzen wahre Aussagen ergeben.

Beispiele

- Aussageform $A(x)$= 'x ist gerade'.
  - $x = 2$: Aussage $A(2)$, '2 ist gerade', ist wahr.
  - $x = 17$: Aussage $A(17)$, '17 ist gerade', ist falsch.
- Aussageform $A(x)$= '$x = x$' ist allgemein gültig.

Eine Aussage heißt **entscheidbar**, wenn sie oder ihr Gegenteil in endlich vielen Schritten verifiziert werden kann.

3.2 Syntax

Grammatik der wohlgeformten Formeln

Das **Alphabet** oder **Vokabular der terminalen Symbole** der Logik besteht aus

1. Einer abzählbaren Menge von **Aussagen-Konstanten** wie etwa '2 ist gerade', 'Das Meer ist blau'.
2. einer abzählbaren Menge von **Aussagen-Variablen**, auch Atome oder atomare Formeln genannt: A, B, C,..., A_1, A_2,...;
3. dem unären Junktor $\neg$ (nicht) und dem binären Junktor $\vee$ (oder)
4. Klammern: (,)
5. Definitionen, z. B. des Junktors $\wedge$ (und): $A \wedge B \doteqdot \neg(\neg A \vee \neg B)$

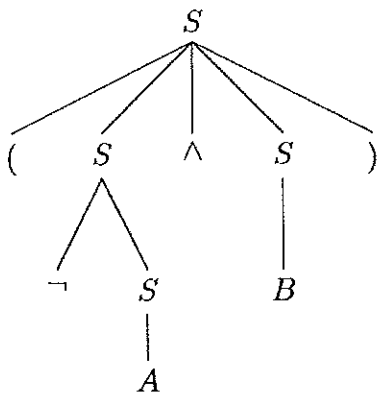
Die folgenden Regeln sind die **Grammatik** der Logik

$S \rightarrow \neg S | (S \wedge S) | (S \vee S)$

$S \rightarrow A | B | A_1 | A_2 | \ldots$

Hier bedeutet das Symbol $\rightarrow$, dass das Syntaxelement S durch Ausdrücke der rechten Seite ersetzt werden kann, und '|' bedeutet 'oder'.

Beispiel: $(\neg A \wedge B)$



Alle Elementaraussagen (Atome, Literale) sind *einfache* **atomare Formeln**, die entweder wahr oder falsch sein können. Etwa: 'Das Meer ist blau', 'Der Hund bellt', '2 ist eine gerade Zahl', '3 ist eine gerade Zahl.' Falls F und G Formeln sind, so sind auch ¬F, (F∨G), (F∧G) Formeln. So können alle **wohlgeformten Formeln** (WFF - *well-formed formulas*) gebildet werden; das sind all jene, deren Aussagen entscheidbar sind.

Junktoren

Aus den Basisjunktoren ∧, ∨ und ¬ können mit Hilfe von **Formeln** alle denk-möglichen Ausdrücke gebildet werden. Interessant (und sinnvoll) sind aber nur wohlgeformte Formeln (WFFs), das sind solche, die logisch sinnvoll sind, d. h., entweder wahr oder falsch sein können. Diese nennen wir **Formeln**.

Beispiele für Formeln:

$$A \qquad (A \wedge B) \qquad ((A \vee B) \wedge (C \vee \neg A)) \qquad ((A \wedge B) \rightarrow C)$$

Keine Formel ist etwa $(\vee A\, B)$, d. h., dies ist kein wohlgeformter Ausdruck.

Wir zählen nun einige Junktoren (in verschiedener Notation) auf und geben auch die (logische) Sprechweise wieder.

Basisjunktoren			
Negation	$\neg$	$\sim, -$	
Konjunktion	$\wedge$	$\&$	
Disjunktion	$\vee$	$+$	
Zusammengesetzte Junktoren			
Implikation	$\rightarrow$	$\Rightarrow$	$(F_1 \rightarrow F_2) \doteqdot (\neg F_1 \vee F_2)$
Äquivalenz	$\leftrightarrow$	$\Leftrightarrow, \equiv, \sim, \approx$	$\doteqdot ((F_1 \wedge F_2) \vee (\neg F_1 \wedge \neg F_2))$ $\doteqdot (F_1 \rightarrow F_2) \wedge (F_2 \rightarrow F_1)$
Iterierte Konjunktion	$\bigwedge_{i=1}^{n}$		$\doteqdot (\cdots (F_1 \wedge F_2) \wedge F_3) \wedge \ldots \wedge F_n)$
Iterierte Disjunktion	$\bigvee_{i=1}^{n}$		$\doteqdot (\cdots (F_1 \vee F_2) \vee F_3) \vee \ldots \vee F_n)$
Verum (Top)	$\top$		$\doteqdot (F_1 \vee \neg F_1) \vee (F_2 \vee \neg F_2)$
Falsum (Bottom)	$\bot$		$\doteqdot (F_1 \wedge \neg F_1) \wedge (F_2 \wedge \neg F_2)$
Junktoren aus der Informatik			
Negierte Konjunktion	NAND	$\uparrow$	$\doteqdot \neg(F_1 \wedge F_2)$
Negierte Disjunktion	NOR	$\downarrow$	$\doteqdot \neg(F_1 \vee F_2)$
Exkludierende Disjunktion	XOR	$\veebar$	$\doteqdot (F_1 \wedge \neg F_2) \vee (\neg F_1 \wedge F_2)$

NAND ($\uparrow$) wird als Sheffer- und NOR ($\downarrow$) als Peirce-Funktion bezeichnet.

Vorsicht! Das $\vee$ in der Logik entspricht *nicht* unserem (umgangs-) sprachlichen ≫oder≪; diesem entspricht das $\veebar$ (XOR).

Sprechweisen:

$\neg F_1$	nicht F_1; non F_1; not F_1
$F_1 \wedge F_2$	F_1 und F_2; sowohl F_1 als auch F_2
$F_1 \vee F_2$	F_1 oder F_2; F_1 oder F_2 oder beide
$F_1 \rightarrow F_2$	wenn F_1 dann F_2; aus F_1 folgt F_2; F_1 impliziert F_2
$F_1 \leftrightarrow F_2$	F_1 ist äquivalent F_2; wenn F_1 dann F_2 und umgekehrt; F_1 genau dann, wenn F_2; F_1 dann und nur dann, wenn F_2

Im Gegensatz zur Logik werden in der Mathematik die Implikation und die Äquivalenz meistens mit einem Dopplestrich dargestellt, d. h., durch '$\Rightarrow$' bzw. '$\Leftrightarrow$'.

$F_1(x) \Rightarrow F_2(x)$ Die Voraussetzung $F_1(x)$ ist eine **hinreichende Bedingung** (sufficient condition) für $F_2(x)$.

Die Behauptung $F_2(x)$ ist eine **notwendige Bedingung** (necessary condition) für $F_1(x)$. F_1 wird auch als Antezedens, F_2 als Sukzedens bezeichnet, $F_1 \Rightarrow F_2$ als Konditional.

$F_1(x) \Leftrightarrow F_2(x)$ Die Voraussetzung $F_1(x)$ ist eine notwendige und hinreichende Bedingung für $F_2(x)$. Und da der Ausdruck symmetrisch ist, ist auch die Behauptung $F_2(x)$ eine notwendige und hinreichende Bedingung für $F_1(x)$.

Aussagen-Kalkül

Die Aussagenlogik ist entscheidbar, d. h., für jede beliebige, wohlgeformte Formel kann entschieden werden, ob sie wahr oder falsch ist. Ein solches Entscheidungsverfahren sind etwa die Wahrheitstafeln.

Für die Prädikatenlogik gibt es keine Entscheidungsverfahren, weshalb ein Kalkül, d. h., eine Ableitungsmaschinerie für wahre Sätze, gebraucht wird. Die Basis für diesen Kalkül liefert der Aussage-Kalkül. Die Prädikatenlogik ist eine Erweiterung der Aussagenlogik, die „definierende" Prädikate verwendet, wie etwa: '... ist ein Mensch', '... ist blau', '... ist größer als ...'. Hier wird dann '...' durch „Eigennamen" ersetzt.

Ein Kalkül ist ein formales Regelsystem (wie etwa auch beim Schach). Er besteht aus dem **Begriffsnetz** und dem **Deduktionsgerüst**. Das Begriffsnetz des Aussagen-Kalküls ist die Grammatik der wohlgeformten Formeln. Das Deduktionsgerüst besteht aus einer Menge von **Axiomen** und einer Menge von **Schlussregeln**. Axiome sind allgemein gültige Aussagen, die widerspruchsfrei (*sound*) sein müssen.

Als Axiomenmenge können etwa alle Tautologien genommen werden, da wir ja durch die Wahrheitstafeln entscheiden können, ob eine Formel eine Tautologie ist oder nicht. Eine Tautologie ist dabei eine stets wahre Aussage, wie etwa *'Jede gerade Zahl ist durch zwei teilbar'*; oder auch $A \to A$.

Am bekanntesten ist jedoch der Aussagen-Kalkül von Lukasiewicz der drei Axiome hat:

1. Replikation: $A \to (B \to A)$
2. Selbstdistribution: $[A \to (B \to C)] \to [(A \to B) \to (A \to C)]$
3. Kontraposition: $(\neg A \to \neg B) \to (A \to B)$

Es gibt jedoch auch andere Axiomenmengen, z. B. die von Bertrand Russell auf der Basis von $\neg$ und $\vee$.

Der Aussage-Kalkül hat nur zwei Schlussregeln:

1. Ersetzungsregel: Jede Aussagen-Variable kann durch eine wohlgeformte Formel ersetzt werden; das muss aber an allen Stellen ihres Vorkommens gleichzeitig durchgeführt werden!
2. Modus Ponens: Von A und $A \to B$ kann zu B übergegangen werden. $\dfrac{A \qquad A \to B}{B}$

Mit diesen Regeln können aus den Axiomen **Sätze** (Theoreme) in Form eines Beweises abgeleitet werden. Ein **Beweis** ist eine **endliche Folge von Formeln** $F_1, F_2, \ldots, F_n$, wobei jedes F_i entweder ein Axiom ist, oder bereits aus vorhergehenden Formeln abgeleitet wurde.

Der Kalkül bildet die Syntax bzw. Sprache der Aussagenlogik. Parallel dazu gibt es die Wirklichkeit, den Bereich der (mathematischen) Strukturen und der wahren und falschen Aussagen über die (mathematische) Welt.

Die Abbildung der Sprache auf die Strukturen (der Welt) heißt **Interpretation**, und wenn diese richtig ist, bildet die Struktur ein **Modell**. Diese Relation zwischen der Sprache und den Objekten und Strukturen heißt **Semantik**.

Für die Syntax und die Semantik sind folgende Begriffe wichtig:
Im Kalkül gibt es nur die Ableitungen (oder Herleitungen, Deduktion) von Formeln, d. h., $\vdash F$, oder in Worten: F wird aus den vorhandenen gültigen Formeln mittels logischem Kalkül hergeleitet (oder ist herleitbar).

Dem entspricht in der Wirklichkeit die Gültigkeit der Formel $\models F$, also etwa die Tautologie $\models A \to A$, jede gültige Formel ist eine Tautologie. Werden zusätzlich Annahmen A erlaubt, so kann aus A auch B abgeleitet werden: $A \vdash B$.

Dem entspricht im Modell (bzw. in der Struktur) die logische Folgerung $A \models B$.

1. Ein Kalkül soll **korrekt** sein, d. h., dass nur wahre Formeln abgeleitet werden können:
$$\vdash A \Rightarrow \models A$$
2. Es soll **vollständig** sein, d. h., alle wahren Formeln sollen abgeleitet werden können:
$$\models A \Rightarrow \vdash A$$
3. Es soll **widerspruchsfrei** (*sound*) sein, d. h. z. B.:
$$\not\vdash A \wedge \neg A$$

Die Aussagenlogik ist korrekt, vollständig und widerspruchsfrei, auch die Prädikatenlogik, nicht jedoch die Arithmetik (Zahlentheorie).

Kurt Friedrich Gödel (geb. 28. April 1906 in Brünn/Brno, gestorben 14. Jänner 1978 in Princeton; USA) machte aber mit seinem **Unentscheidbarkeitssatz** alle Hoffnung zunichte (vor allem, die der MathematikerInnen), die ≫Welt≪ vollständig beschreiben zu können.

Unentscheidbarkeitssatz

In jedem Axiomensystem gibt es formal unentscheidbare Sätze.

Alan Turings (Unlösbarkeit des Halte-Problems für Turingmaschinen) und Gregory J. Chaitins ('Zufälligkeit der Arithmetik') Unvollständigkeitsresultate basieren auf Gödels Satz.

Quantoren

Obwohl die Quantoren nicht direkt hierher passen, erwähne ich sie hier, da sie später für das mathematische Beweisen von grundlegender Bedeutung sind (Prädikatenlogik).

Allquantor	$\forall$	$\bigwedge, (\ldots)$	für alle/jede ... gilt
Existenzquantor	$\exists$	$\bigvee, (E\ldots)$	es gibt (mindestens) ein ..., dass
	$\exists!$		Es existiert genau ein ..., dass

Beispiele:

- $\forall x : x \in A \Rightarrow x \in B$ heißt: 'Alle A sind B'. Oder 'Jedes A ist B'. $\in$ ist die Elementschaftsrelation.
- $\exists x : x \leq y \wedge x \geq z$ heißt: 'Es gibt (mindestens) ein x, dass kleiner-gleich y und größer-gleich z ist' oder 'Einige x sind zwischen z und y'.
- $\forall x : x \in A \Rightarrow \neg x \in B$ heißt: 'Kein A ist B'. Oder: 'Alle x aus A sind nicht aus B'.

Wahrheitstafeln

Mit Hilfe von Wahrheitstafeln (*truth tables*) können komplexe Formeln bewiesen werden. Dabei wird jeder Formel, beginnend mit den Atomen (Elementaraussagen), ihr Wahrheitswert zugewiesen. Elementaraussagen können entweder wahr (1) oder falsch (0) sein.

Sei M_A die Menge aller Atome, so heißt eine Abbildung

$$\mathcal{W} : M_A \to \{0, 1\}$$

Wahrheitswertzuweisung oder **Belegung**.

Etwa $\mathcal{W}(A \wedge B) = \begin{cases} 1 & \text{falls } \mathcal{W}(A) = \mathcal{W}(B) = 1 \\ 0 & \text{sonst} \end{cases}$

In Wahrheitstafeln wird die Notierung der Belegung oft weggelassen und statt $\mathcal{W}(F)$ nur F geschrieben, da dies aus dem Kontext ersichtlich ist.

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Leftrightarrow B$	A XOR B
1	1	0	1	1	1	1	0
1	0	0	0	1	0	0	1
0	1	1	0	1	1	0	1
0	0	1	0	0	1	1	0

Insgesamt gibt es 16 mögliche Verknüpfungen zweier Mengen; vergleiche dazu auch die Tabelle der 2-stelligen Junktoren.

Eine andere Möglichkeit, die Wahrheitstabelle einer komplexen Formel aufzustellen ist, diese aufzuschreiben und die Spalten nach der syntaktischen Struktur abzuarbeiten.

$(A$	$\wedge$	$B)$	$\Rightarrow$	$(A$	$\vee$	$\neg$	$C)$
1.	2.	1.	4.	1.	3.	2.	1.
1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	0
1	0	0	1	1	1	0	1
1	0	0	1	1	1	1	0
0	0	1	1	0	0	0	1
0	0	1	1	0	1	1	0
0	0	0	1	0	0	0	1
0	0	0	1	0	1	1	0

Wie wir sehen machen die Belegungen die Formel (den Hauptoperator '$\Rightarrow$') wahr. Damit ist die Aussage bzw. Folgerung allgemein gültig.

2-stellige Junktoren

Durch die Basisjunktoren $\neg$ und $\vee$ können alle Formeln gebildet werden. Hierbei ist $\neg$ ein 1-stelliger (unärer) Junktor (1 Argument), $\vee$ ein 2-stelliger (binärer) Junktor. Es gibt insgesamt 16 2-stellige Junktoren, die in folgender Tabelle aufgelistet sind.

A	B	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}	J_{15}	J_{16}
1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
1	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
0	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
		$\top$	$\vee$			$\rightarrow$		$\leftrightarrow$	$\wedge$	NAND	XOR					NOR	$\bot$

Da, wie wir wissen, für komplexe logische Zusammenhänge nur eine Minimalmenge von Junktoren notwendig ist, schreiben wir die Minimalmengen auf:

1. Jeder 2-stellige Junktor ist durch $\{\neg, \vee\}$ ausdrückbar.
2. Jeder 2-stellige Junktor ist durch $\{\neg, \wedge\}$ ausdrückbar.
3. Jeder 2-stellige Junktor ist durch $\{\text{NAND}\}$ ausdrückbar.
4. Jeder 2-stellige Junktor ist durch $\{\text{NOR}\}$ ausdrückbar.
5. Jeder 2-stellige Junktor ist durch $\{\neg, \rightarrow\}$ ausdrückbar.

Dies gilt sogar für n-stellige Junktoren (n endlich).

Bem.: Der Aufbau komplexer logischer Schaltungen (etwa Computer) ist alleine durch NAND ($\uparrow$) und NOR ($\downarrow$) Gatter möglich, da etwa: $\neg A \doteqdot A \text{ NAND } A$ und $A \wedge B \doteqdot \neg(A \text{ NAND } B)$.

3.3 Modelle

Eine Belegung $\mathcal{W}$ ist passend zu einer Formel F, falls $\mathcal{W}$ für alle atomaren Teilformeln (Elementaraussagen) von F definiert ist.

Eine zu F passende Belegung $\mathcal{W}$ heißt **Modell** für F, wenn $\mathcal{W}(F) = 1$. Wir schreiben $\mathcal{W} \models F$, d. h., $\gg\mathcal{W}$ macht F wahr$\ll$, oder F gilt unter der Belegung $\mathcal{W}$. Dasselbe gilt für Mengen von Formeln.

Gibt es ein Modell $\mathcal{W}$ für eine Formel F, dann heißt F **erfüllbar**, andernfalls **unerfüllbar**. Dasselbe gilt für Mengen von Formeln.

Eine Formel F heißt **gültig**, falls jede Belegung $\mathcal{W}$ ein Modell für F ist. F heißt dann auch **Tautologie** oder **allgemein gültig**, wir schreiben $\models F$.

Bem.: Zu jeder allgemein gültigen Formel (Tautologie) gibt es eine unerfüllbare (Antilogie), nämlich ihre Negation. Eine **Antilogie** ist eine Aussage, die unabhängig vom Bedeutungskontext falsch ist.

Eine Formel F heißt **kontingent**, wenn sie erfüllbar ist aber nicht allgemein gültig. d. h., es gibt zumindest eine passende Belegung von F, die kein Modell für F ist.

Beispiele:

- ➢ Jede atomare Formel ist erfüllbar und kontingent.
- ➢ $A \vee \neg A$ ist allgemein gültig, d. h., eine Tautologie.
- ➢ $A \wedge \neg A$ ist unerfüllbar.

Folgerbarkeit

Eine Formel G folgt logisch aus einer Formelmenge $M = \{F_1, \ldots, F_n\}$, wenn jedes $\mathcal{W}$, das Modell für *jede* Formel $F_i \in M$ ist, auch Modell für G ist.

$$\mathcal{W} \models F_i,\ \forall F_i \in M \ :\Rightarrow\ \mathcal{W} \models G$$

Eine andere Schreibweise dafür ist

$$F_1, \ldots, F_n \models G$$

Bem.: $\models$ ist eine Relation zwischen Formeln, kein Junktor!

Es gilt jedoch: $F \models G \Leftrightarrow (F \rightarrow G)$ allgemein gültig.

Bem.:

- ➢ Eine allgemein gültige Formel ist aus jeder Formelmenge folgerbar ($X \models A \vee \neg A$).

➢ Aus jeder unerfüllten Formelmenge ist jede beliebige Formel folgerbar.
Ex falso quodlibet, $A \vee \neg A \vDash X$.

Beispiel:
$$\frac{\textit{3 ist Teiler von 7} \quad \textit{8 ist Teiler von 10}}{\textit{2 ist Teiler von 12}}$$

Wahrheitstabellen dienen unter anderem dazu, das Verhalten von Formeln unter verschiedenen Belegungen zu testen. Eine Formel mit n Atomen hat 2^n verschiedene Belegungen. Alle zu testen ist daher oft sehr langwierig.

Wir wollen Erfüllbarkeit, Allgemeingültigkeit oder Unerfüllbarkeit von Formeln nachweisen. Folgt eine Formel G aus einer Formel F, so gelten für G die gleichen Aussagen wie für F.

Eine mathematische Aussage ist etwa:

THEOREM $\quad a \cdot b = 0 \Leftrightarrow a = 0 \vee b = 0.$
Ein Produkt ist genau dann 0, wenn mindestens ein Faktor gleich 0 ist.

Um solche Aussagen zu beweisen, bedienen wir uns spezieller Aussagen.

Die Idee des indirekten Beweises lässt sich so beschreiben. Es gilt

$$F \vDash G \Leftrightarrow F \wedge \neg G \quad \text{ist unerfüllbar}$$

Das entspricht genau dem indirekten Beweis (**reductio ad absurdum**): $\dfrac{F \wedge \neg G \to \neg F}{F \to G}$.

Ähnlich ist auch die **Kontraposition**: $(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$. Sie liefert einen Beweis für $(A \Rightarrow B)$ aus $(\neg B \Rightarrow \neg A)$.

Als Beispiel mag hier folgender Satz dienen.

THEOREM. Es gibt unendlich viele Primzahlen.

BEWEIS. Indirekt: Wir nehmen an, dass es nur endlich viele Primzahlen $p_1, \ldots, p_n$ gibt. Dann muss es darunter eine Primzahl p_i geben, die die Zahl $p = p_1 \cdot \cdots \cdot p_n + 1$ teilt. Das heißt aber auch, dass p_i die 1 teilt, ein Widerspruch. □

Bem. Aus diesem Beweis folgt aber nicht, dass die Zahl $p = p_1 \cdot \cdots \cdot p_n + 1$ eine Primzahl ist!

Eine weitere Beweisstrategie ist die, dass die Allgemeingültigkeit einer Formel F bewiesen wird, indem nachgewiesen wird, dass es keine Belegung gibt, die F falsch macht (durch Analyse der Wahrheitstafel von F).

Die logische Äquivalenz von Formeln kann auch für Beweise genutzt werden:
Zwei Formeln F und G heißen **äquivalent**, $F \equiv G$, falls $F \vDash G$ und $G \vDash F$.

D. h., es wird versucht zu zeigen, dass sich zwei Formeln unter allen passenden Belegungen gleich verhalten, denn

➢ Wenn $F_1 \vDash G$ und $F_1 \equiv F_2$, dann $F_2 \vDash G$.

➢ Wenn $F \models G_1$ und $G_1 \equiv G_2$, dann $F \models G_2$.

Bem.: Äquivalenzumformungen in der Mathematik.

Wichtige Äquivalenzen

➢ Idempotenz	$(F \wedge F) \equiv F$ und $(F \vee F) \equiv F$
➢ Kommutativität	$(F \wedge G) \equiv (G \wedge F)$ und $(F \vee G) \equiv (G \vee F)$
➢ Assoziativität	$((F \wedge G) \wedge H) \equiv (F \wedge (G \wedge H))$
	$((F \vee G) \vee H) \equiv (F \vee (G \vee H))$
➢ Absorption	$(F \wedge (F \vee G) \equiv F$
	$(F \vee (F \wedge G) \equiv F$
➢ Distributivität	$(F \wedge (G \vee H)) \equiv ((F \wedge G) \vee (F \wedge H))$
	$(F \vee (G \wedge H)) \equiv ((F \vee G) \wedge (F \vee H))$
➢ Doppelte Negation	$\neg\neg F \equiv F$
➢ DeMorgan-Regeln	$\neg(F \wedge G) \equiv (\neg F \vee \neg G)$
	$\neg(F \vee G) \equiv (\neg F \wedge \neg G)$
➢ Kontraposition	$(F \rightarrow G) \equiv (\neg G \rightarrow \neg F)$
➢ Tautologie-Regeln	$(F \vee G) \equiv F$ falls F Tautologie
	$(F \wedge G) \equiv G$ falls F Tautologie
➢ Unerfüllbarkeitsregeln	$(F \vee G) \equiv G$ falls F unerfüllbar
	$(F \wedge G) \equiv F$ falls F unerfüllbar
➢ Weitere	$(F \rightarrow (G \rightarrow H)) \equiv ((F \wedge G) \rightarrow H)$
	$\neg(F \rightarrow G) \equiv (F \wedge \neg G)$
	$(F \Leftrightarrow G) \equiv ((F \rightarrow G) \wedge (G \rightarrow F)$
	$(F \Leftrightarrow G) \equiv ((F \wedge G) \vee (\neg F \wedge \neg G)$

Bemerkung: Alle Äquivalenzen sind Tautologien, wenn $\equiv$ durch $\Leftrightarrow$ ersetzt wird.

Bemerkung: Sehr ähnliche Regeln gelten in der Mengenlehre und auch beim Rechnen mit Zahlen. Der Grund ist die algebraische Struktur, die hinter diesen Konstrukten steckt (Mengen-Algebra, Boolescher Verband, Zahlen-Körper, etc.). Siehe dort.

3.4 Normalformen

Eine atomare Formel bzw. ihre Negation wird als (**positives** bzw. **negatives**) **Literal** bezeichnet. Positives und negatives Literal sind zueinander **komplementär**.

Eine Disjunktion von Literalen L_i wird als **Klausel** K bezeichnet: $K = (\bigvee_{i=1}^{n} L_i)$.

Beispiele:

➢ Klauseln sind: $A \vee B$, $\neg A_1 \vee A_2 \vee \neg C$, A (da $A = A \vee A$).

➢ Keine Klauseln sind: $\neg\neg A$, $A \vee B \wedge C$.

➢ Da jede Klausel mindestens ein Disjunktionsglied enthält, ist jede (nicht-leere) Klausel erfüllbar.

➤ Eine Klausel ist genau dann allgemein gültig, d. h., eine Tautologie, wenn sie mindestens ein Paar komplementärer Literale enthält.

Bemerkung: Beim automatischen Beweisen wird auch noch die *leere Klausel* $\square$ verwendet, die unerfüllbar ist! Sie stellt eine Verallgemeinerung des Klauselbegriffs dar;
folgt aus: $(\vee_{i=1}^{n} L_i) = (\vee_{i\in\{1,\ldots,n\}} L_i)$ und $\square \doteqdot (\vee_{i\in\phi} L_i)$.

Bemerkung: Die **Horn-Klausel** $A \vee \neg B_1 \vee \ldots \vee \neg B_n$ ist eine spezielle Form. Sie hat die Interpretation $A \Leftarrow (B_1 \wedge \ldots \wedge B_n)$. Die Sprache Prolog ('programing in logic') basiert auf Horn-Klauseln.

THEOREM. Jede Formel kann (äquivalent) in eine der beiden folgenden Normalformen übergeführt werden.

Eine Formel F ist in **konjunktiver Normalform** (KNF), wenn sie eine Konjunktion von Disjunktionen von Literalen ist. Sie hat dann die Form:

$$F = (\wedge_{i=1}^{n} \vee_{j=1}^{k_i} L_{ij})$$

Eine Formel F ist in **disjunktiver Normalform** (DNF), wenn sie eine Disjunktion von Konjunktionen von Literalen ist. Sie hat dann die Form:

$$F = (\vee_{i=1}^{n} \wedge_{j=1}^{k_i} L_{ij})$$

Bemerkung: Eine Formel in KNF ist dann eine Konjunktion von Klauseln.

Algorithmus zur KNF. (Analog: DNF).

❶ Ersetze alle 2-stelligen Junktoren durch die Basisjunktoren $\{\neg, \wedge, \vee\}$

❷ Ersetze folgende Teilformeln folgendermaßen:

$\neg\neg F$ durch F
$\neg(F \wedge G)$ durch $(\neg F \vee \neg G)$ (DeMorgan)
$\neg(F \vee G)$ durch $(\neg F \wedge \neg G)$ (DeMorgan)

Solange, bis nur noch negierte oder nicht-negierte Atome vorkommen.

❸ Ersetze folgende Teilformeln folgendermaßen:

$(F \vee (G \wedge H))$ durch $(F \vee G) \wedge (F \vee H))$
$((F \wedge G) \vee H)$ durch $(F \vee H) \wedge (G \vee H))$

Solange, bis alle Disjunktionen innen stehen.

❹ Reduziere überflüssige Literale, Disjunktionen und Klammern durch Anwendung der Regeln für Tautologie, Idempotenz, Assoziativität und Kommutativität.

Beispiel:

$$
\begin{array}{lll}
 & (A \wedge (B \rightarrow C)) \rightarrow D & \\
\equiv & (A \wedge (\neg B \vee C)) \rightarrow D & \text{Ersetzen von } \rightarrow \\
\equiv & \neg(A \wedge (\neg B \vee C)) \vee D & \text{Ersetzen von } \rightarrow \\
\equiv & (\neg A \vee \neg(\neg B \vee C)) \vee D & \text{DeMorgan} \\
\equiv & (\neg A \vee (\neg\neg B \wedge \neg C)) \vee D & \text{DeMorgan} \\
\equiv & (\neg A \vee (B \wedge \neg C)) \vee D & \text{Vereinfachung von } \neg\neg \\
\equiv & ((\neg A \vee B) \wedge (\neg A \vee \neg C)) \vee D & \text{Distributivgesetz bez. } \neg A \\
\equiv & ((\neg A \vee B) \vee D) \wedge ((\neg A \vee \neg C) \vee D) & \text{Distributivgesetz bez. } D \\
\equiv & (\neg A \vee B \vee D) \wedge (\neg A \vee \neg C \vee D) & \text{Assoziativgesetz}
\end{array}
$$

Erstellung der Normalformen aus Wahrheitstafeln

Haben wir eine Wahrheitstabelle zu einer beliebigen Formel F, dann folgt die DNF aus der Disjunktion aller Konjunktionen jener Zeilen, in denen die Formel F die Belegung wahr (1) hat. Hat das Atom A_i den Wert 1 so nimm A_i in die Konjunktion, sonst $\neg A_i$.

Die KNF als Konjunktion der Disjunktionen jener Zeilen, in denen die Formel F die Belegung falsch (0) hat, wobei jeweils die negierten Werte der wahren Atome (1) in die Disjunktion aufgenommen wird, d. h., hat das Atom A_i den Wert 1 so nimm $\neg A_i$ in die Disjunktion, sonst A_i (DeMorgan).

Nur als Beispiel

DNF				
A	B	C	F	
0	0	0	1	$(\neg A \wedge \neg B \wedge \neg C)$
0	0	1	0	
0	1	0	1	$(\neg A \wedge B \wedge \neg C)$
0	1	1	0	
1	0	0	0	
1	0	1	0	
1	1	0	1	$(A \wedge B \wedge \neg C)$
1	1	1	0	

$\Rightarrow$ DNF: $(\neg A \wedge \neg B \wedge \neg C) \vee (\neg A \wedge B \wedge \neg C) \vee (A \wedge B \wedge \neg C)$

KNF				
A	B	C	F	
0	0	0	1	
0	0	1	0	$(A \vee B \vee \neg C)$
0	1	0	1	
0	1	1	0	$(A \vee \neg B \vee \neg C)$
1	0	0	0	$(\neg A \vee B \vee C)$
1	0	1	0	$(\neg A \vee B \vee \neg C)$
1	1	0	1	
1	1	1	0	$(\neg A \vee \neg B \vee \neg C)$

$\Rightarrow$ KNF: $(A \vee B \vee \neg C) \wedge (A \vee \neg B \vee \neg C) \wedge (\neg A \vee B \vee C) \wedge (\neg A \vee B \vee \neg C) \wedge (\neg A \vee \neg B \vee \neg C)$

Wir können jeder Zeile der Wahrheitstafel eine Formel zu Grunde legen, die, bei der KNF als Konjunktion – bei der DNF als Diskjunktion, den Wert wahr (1) für diese Zeile annimmt.

Wir können daher die obigen Wahrheitstafeln folgendermaßen erweitern.

DNF				Konjunktionen							
A	B	C	F	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
0	0	0	1	1	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0	0	0	0	0
0	1	0	1	0	0	1	0	0	0	0	0
0	1	1	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0
1	0	1	0	0	0	0	0	0	1	0	0
1	1	0	1	0	0	0	0	0	0	1	0
1	1	1	0	0	0	0	0	0	0	0	1

Z. B.: $F_6 \equiv (A \wedge \neg B \wedge C)$
Wahrheitstafel $\Rightarrow$ DNF: $F \equiv F_1 \vee F_3 \vee F_7$

KNF				Disjunktionen							
A	B	C	F	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8
0	0	0	1	1	0	0	0	0	0	0	0
0	0	1	0	0	1	0	0	0	0	0	0
0	1	0	1	0	0	1	0	0	0	0	0
0	1	1	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	0
1	0	1	0	0	0	0	0	0	1	0	0
1	1	0	1	0	0	0	0	0	0	1	0
1	1	1	0	0	0	0	0	0	0	0	1

Z. B.: $G_6 \equiv (\neg A \vee B \vee \neg C)$ (Beachte: $G_6 \equiv \neg(A \wedge \neg B \wedge C)$)
Wahrheitstafel $\Rightarrow$ KNF: $F \equiv G_2 \wedge G_4 \wedge G_5 \wedge G_6 \wedge G_8$

3.5 Schlussregeln

Sei $(A_1 \wedge \ldots \wedge A_n) \Rightarrow B$ eine Tautologie, dann heißen A_i **Prämissen** und B **Konklusion**.

Eine Tabelle mit folgender Form heißt **Schlussregel** (vgl. Horn-Klauseln):

$$\begin{array}{l} A_1 \\ \vdots \\ \underline{A_n} \\ B \end{array}$$

Jede Schlussregel (linke Seite) kann auch als Formel der Aussagenlogik (rechte Seite) dargestellt werden.

Modus ponens:

$$\begin{array}{l} A \Rightarrow B \\ \underline{A} \\ B \end{array} \qquad \text{oder} \qquad ((A \Rightarrow B) \wedge A) \Rightarrow B$$

Modus tollens:

$$\begin{array}{l} A \Rightarrow B \\ \underline{\neg B} \\ \neg A \end{array} \qquad \text{oder} \qquad ((A \Rightarrow B) \wedge \neg B) \Rightarrow \neg A$$

Kettenschluss:

$$\begin{array}{l} A \Rightarrow B \\ \underline{B \Rightarrow C} \\ A \Rightarrow C \end{array} \qquad \text{oder} \qquad ((A \Rightarrow B) \wedge (B \Rightarrow C)) \Rightarrow (A \Rightarrow C)$$

Konstruktives Dilemma:

$$\begin{array}{l} A \vee B \\ A \Rightarrow C \\ \underline{B \Rightarrow C} \\ C \end{array}$$

Vereinfachung: $\dfrac{A \wedge B}{A}$

Addition: $\dfrac{A}{A \vee B}$

Weiters: $\dfrac{B}{A \Rightarrow B}$ $\qquad$ $\dfrac{A \Rightarrow B}{(A \wedge C) \Rightarrow B}$ $\qquad$ $\dfrac{(A \vee B) \Rightarrow C}{A \Rightarrow C}$

Indirekter Beweis (reductio ad absurdum) neu formuliert:
Aus den Annahmen (immer wahr) $H_1, \ldots, H_n$ soll P bewiesen werden.
Zeige, dass aus $\wedge H_1, \ldots, H_n, \neg P$ der Widerspruch $Q \wedge \neg Q$ folgt.

$$\frac{\wedge H_1, \ldots H_n, \neg P}{(Q \wedge \neg Q)} \qquad \Rightarrow \qquad \frac{H_1 \ldots H_n}{P}$$

Beispiele

Eine Implikation ist genau dann wahr, wenn das Antezedens (Vorderglied) oder das Consequens (Hinterglied) wahr ist.

Daher sind auch folgende Aussagen wahr:

- Wenn der Mond aus grünem Käse ist, dann sind die Ozeane voller Wasser.
- Wenn der Mond aus grünem Käse ist, dann sind die Ozeane aus Beton.
- Wenn der Mond nicht aus grünem Käse ist, dann sind die Ozeane voller Wasser.

Beweise durch Wahrheitstafeln, dass folgende Formeln allgemein gültig sind.

B-1 $(P \wedge Q) \rightarrow P$

B-2 $P \rightarrow (P \vee Q)$

B-3 $(((P \vee Q) \wedge (P \rightarrow R)) \wedge (Q \rightarrow R)) \rightarrow R$

B-4 $(P \rightarrow Q) \Leftrightarrow (\neg P \vee Q)$

B-5 $(P \wedge (P \rightarrow Q)) \rightarrow Q$

B-6 $(\neg Q \wedge (P \rightarrow Q)) \rightarrow \neg P$

B-7 $((P \vee Q) \wedge \neg P) \rightarrow Q$

B-8 $(P \rightarrow Q) \rightarrow ((P \wedge R) \rightarrow Q)$

B-9 $((P \vee Q) \rightarrow R) \rightarrow (P \rightarrow R)$

Beweise durch Wahrheitstafeln, dass folgende Formeln nicht allgemein gültig sind.

G-1 $P \rightarrow (P \wedge Q)$

G-2 $P \rightarrow Q$

G-3 $(P \vee Q) \rightarrow P$

G-4 $((P \rightarrow Q) \wedge Q) \rightarrow P$

G-5 $((P \rightarrow Q) \wedge \neg P) \rightarrow \neg Q$

Beweis durch logische Äquivalenzen

1. $(P \vee Q) \wedge \neg(P \wedge Q) \equiv \neg(P \Leftrightarrow Q)$

   Beweis:

$(P \vee Q) \wedge \neg(P \wedge Q)$	
$(\neg\neg P \vee Q) \wedge \neg(P \wedge Q)$	Komplement
$(\neg\neg P \vee \neg\neg Q) \wedge \neg(P \wedge Q)$	Komplement
$\neg(\neg P \wedge \neg Q) \wedge \neg(P \wedge Q)$	DeMorgan
$\neg((\neg P \wedge \neg Q) \vee (P \wedge Q))$	DeMorgan
$\neg(P \Leftrightarrow Q)$	

3.6 Geschichte und Systeme der Logik

Geschichte der Logik

- Pythagoräer, Megariker
- Aristoteles (– 322 B.C.)
- Scholastik; Thomas von Aquin, Occam
- Leibniz: Calculus de Continentibus et Continentis, Ars Combinatoria (17. Jhdt.)
- De Morgan: Formal Logic, or the Calculus of Inference, Necessary and Probable (1847)
- Boole: An Investigation of the Laws of Thought, on which are founded the Mathematical Theory of Logic and Probabilities (1854)
- Frege: Begriffsschrift (1879)
- Russel & Whitehead: Principia Mathematica (1910–13)
- Hilbert: Formalismus, Metamathematik (1920)
- Brouwer: Intuitionismus (1920)
- Gödel: Unvollständigkeitsbeweis (1931)
- Church, Turing: Logik und Berechenbarkeit (1935/36)
- Kripke: Semantik der Modallogik (1959)
- Robinson: Resolventenkalkül (1965)

Systeme der Logik

- Klassische Logik
  - Aussagenlogik
  - Klassenlogik
  - Prädikatenlogik erster Stufe
- Erweiterungen der Logik
  - Modallogiken
  - Zeit-Logik
  - Sorten-Logik
  - Prädikatenlogik höherer Stufen
- Alternativen, Modifikationen

- ➢ Intuitionistische Logiken, Relevanzlogik
- ➢ Mehrwertige Logiken, Fuzzy Logic
- ➢ Nicht-monotone Logiken

4 Elementare Gruppentheorie

Wir betrachten eine Menge G und eine zugehörige Operation $\otimes$ (das muss nicht die Multiplikation sein, dieses Symbol kann alles Mögliche bedeuten). Die Operation $\otimes$ ist eine binäre Operation: Das heißt, für je zwei Elemente $a, b \in G$ ist $a \otimes b$ definiert.

Folgende Eigenschaften können erfüllt sein (oder auch nicht):

DEFINITION 4.1 (GRUPPENEIGENSCHAFTEN)

G1 Abgeschlossenheit: $a, b \in G \Rightarrow a \otimes b \in G$.

G2 Assoziativität: $a, b, c \in G \Rightarrow (a \otimes b) \otimes c = a \otimes (b \otimes c)$.

G3 Einheitselement: Es gibt ein $e \in G$, genannt *Einheitselement* oder *neutrales Element*, das linksneutral und rechtsneutral ist. e heißt *linksneutral*, falls $e \otimes a = a$ für alle $a \in G$ und e heißt *rechtsneutral*, falls $a \otimes e = a$ für alle $a \in G$.

Es kann höchstens ein Einheitselement geben, denn sind e und e' Einheitselemente, so folgt $e = e \otimes e' = e'$.

G4 Inverse Elemente: Zu jedem $a \in G$ gibt es ein *inverses Element* $a' \in G$, das zu a linksinvers und rechtsinvers ist. a' heißt *linksinvers* zu a, falls $a' \otimes a = e$. a' heißt *rechtsinvers* zu a, falls $a \otimes a' = e$.

Gilt (G2), so ist auch das inverse Element eindeutig bestimmt:
Seien a' und a'' inverse Elemente zu a.
Dann gilt $a' = a' \otimes e = a' \otimes (a \otimes a'') = (a' \otimes a) \otimes a'' = e \otimes a'' = a''$.

G5 Kommutativität: $a, b \in G \Rightarrow a \otimes b = b \otimes a$.

Wir definieren nun eine Hierarchie von algebraischen Strukturen.

DEFINITION 4.2 (EINFACHE ALGEBRAISCHE STRUKTUREN)
Eine algebraische Struktur $(G, \otimes)$ heißt

- *Gruppoid*, falls G1 erfüllt ist.
- *Halbgruppe*, falls G1 und G2 erfüllt sind.
- *Monoid*, falls G1, G2 und G3 erfüllt sind.
- *Gruppe*, falls G1, G2, G3 und G4 erfüllt sind.
- *kommutative Gruppe* oder *Abelsche Gruppe*, falls G1, G2, G3, G4 und G5 erfüllt sind.

Jede Halbgruppe ist somit auch ein Gruppoid, jedes Monoid eine Halbgruppe etc., aber nicht umgekehrt.

BEISPIEL 4.1
$(\mathbb{R}, +)$ ist eine kommutative Gruppe.

$(\mathbb{R}, \cdot)$ ist nur ein Monoid: Das neutrale Element ist die Zahl 1, aber es existiert nicht zu jedem $a \in \mathbb{R}$ ein inverses Element, denn die Zahl 0 hat kein inverses Element.

4.1 Permutationen und ihre Verknüpfungen

DEFINITION 4.3
Eine *Permutation* der Zahlen $1, \ldots, n$ ist eine bijektive Funktion

$$\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}$$

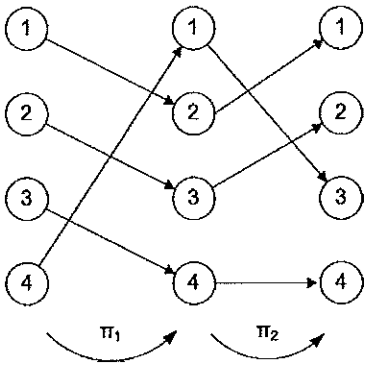
Mit S_n wird die Menge aller Permutationen der Zahlen $1, \ldots, n$ bezeichnet.

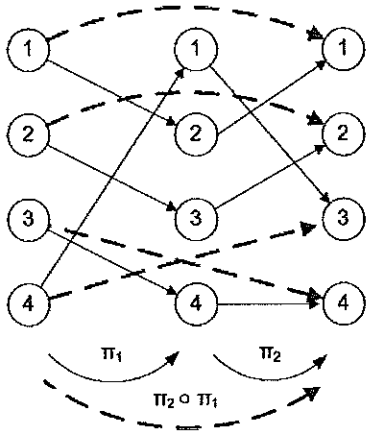
BEISPIEL 4.2 $(n = 4)$
Seien folgende Permutationen gegeben:

i	1	2	3	4
$\pi_1(i)$	2	3	4	1

i	1	2	3	4
$\pi_2(i)$	3	1	2	4

Für diese zwei Permutationen $\pi_1, \pi_2 \in S_4$ betrachten wir $\pi_2 \circ \pi_1 \in S_4$





Beachte die Reihenfolge!

Und somit ergibt sich

i	1	2	3	4
$(\pi_2 \circ \pi_1)(i)$	1	2	4	3

4.2 Permutationsgruppen

Behauptung: $(S_n, \circ)$ ist eine Gruppe:

ad G1: Folgt daraus, dass die Verknüpfung zweier bijektiver Abbildungen wieder bijektiv ist.
ad G2: Folgt aus der Assoziativität der Verknüpfung von Funktionen.
ad G3: Das Einheitselement ist die Identitätsfunktion $id(x) = x$:

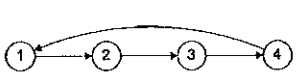
$$e = id, \text{ da } \pi \circ id = id \circ \pi = \pi.$$

ad G4: Das zu π inverse Element ist die zur Funktion π inverse Funktion π^{-1} (sie existiert, da π bijektiv):

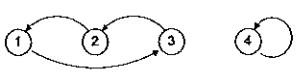
$$\pi^{-1} \circ \pi = \pi \circ \pi^{-1} = id.$$

$(S_n, \circ)$ ist *nicht* kommutativ, da die Verknüpfung zweier Funktionen keine kommutative Operation ist. Man nennt $(S_n, \circ)$ die *symmetrische Gruppe* der Ordnung n.

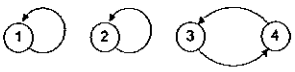
Das letzte Beispiel kann man noch weiter ausbauen. Dazu ist es notwendig, dass wir uns mit Permutationen näher beschäftigen. Eine Permutation lässt sich wie oben als gerichteter Graph darstellen



π_1



π_2



$\pi_2 \circ \pi_1$

Dies kann man auch folgendermaßen notieren: Jede Permutation sich aus *Zyklen* zusammen:

$$\begin{aligned} \pi_1 &= (1\,2\,3\,4) \\ \pi_2 &= (1\,3\,2)(4) \\ \pi_2 \circ \pi_1 &= (1)\,(2)(3\,4) \end{aligned}$$

Allgemein: Jede Permutation kann als "Produkt" von paarweise disjunkten Zyklen dargestellt werden. ("Produkt" = Verknüpfung $\circ$).
Zyklen der Länge zwei nennt man *Transpositionen*. Jeder Zyklus der Länge k lässt sich als Produkt von $k-1$ Transpositionen darstellen. Zum Beispiel:
$(1\,3\,2\,4) = (1\,3)(3\,2)(2\,4)$

$$\begin{array}{ll} 1\,2\,3\,4 & \\ \quad\downarrow & (2\,4) \\ 1\,4\,3\,2 & \\ \quad\downarrow & (3\,2) \\ 1\,4\,2\,3 & \\ \quad\downarrow & (1\,3) \\ 3\,4\,2\,1 & \end{array}$$

Man beachte die Reihenfolge:

$$(1\,3)\,(3\,2)\,(2\,4) = (1\,3) \circ (3\,2) \circ (2\,4).$$

Besteht die Permutation also aus s Zyklen der Längen $k_1, \ldots, k_s$, so kann man sie als Produkt von

$$(k_1 - 1) + \cdots + (k_s - 1) =: t$$

Transpositionen darstellen. Ist t eine gerade Zahl, so nennt man die Permutation π *gerade*, andernfalls heißt π *ungerade*.

Zum Beispiel ist

$$(1\,2\,3)\,(4) = (1\,2)(2\,3) \text{ eine gerade Permutation,}$$

$$(1\,2\,3\,4) = (1\,2)(2\,3)(3\,4) \text{ eine ungerade Permutation.}$$

Vorsicht: (1 2 3 4) bezeichnet die Permutation

i	1	2	3	4
$\pi_3\,(i)$	2	3	4	1

,

(1) (2) (3) (4) hingegen bezeichnet die Permutation

i	1	2	3	4
$\pi_4\,(i)$	1	2	3	4

.

Bemerkung: Zur Definition von geraden bzw. ungeraden Permutationen wurde eine *bestimmte* Art der Darstellung der Permutation als Produkt von Transpositionen benutzt, wie sie sich aus der Zyklen-Darstellung der Permutation ergibt. Das ist natürlich nicht die einzige Art. Zum Beispiel kann die gerade Permutation (1 2 3) nicht nur als (1 2)(2 3) dargestellt werden, sondern etwa auch als (1 2)(1 3)(1 3)(2 3).

Es gilt aber: Jede Darstellung einer geraden (ungeraden) Permutation als Produkt von Transpositionen enthält eine gerade (ungerade) Anzahl von Transpositionen.

Bemerkung: Es lässt sich sogar zeigen, dass sich eine beliebige Permutation immer als Produkt von Transpositionen *benachbarter* Elemente darstellen lässt.

Nun sind wir soweit, das nächste Beispiel für eine Permutationsgruppe zu formulieren:

BEISPIEL 4.3
Sei A_n die Menge der geraden Permutationen der n Elemente $1, \ldots, n$.

Dann ist $(A_n, \circ)$ eine Gruppe, die sogenannte *alternierende Gruppe* der Ordnung n, denn es gelten:

- G1: Eine gerade Anzahl Transpositionen plus eine gerade Anzahl Transpositionen ergibt eine gerade Anzahl Transpositionen.
- G2: Überträgt sich unmittelbar von $(S_n, \circ)$ auf $(A_n, \circ)$.

- G3: $id_{\{1,\ldots,n\}}$ ist eine gerade Permutation, da durch 0 Transpositionen erzielbar.
- G4: Inverses Element: Wenn alle Zyklen umgedreht werden, erhalten wir wieder eine gerade Permutation.

Für bestimmte Anwendungen ist es nützlich, auch noch andere Permutationsgruppen zu betrachten: Es lassen sich etwa in der Geometrie Drehungen und Spiegelungen durch geeignete Matrizengruppen darstellen.

5 Relationen und Ordnungen

Wir haben in der Einführung bzw. bei den Funktionen bereits von Relationen gehört. Wir wollen in diesem Abschnitt die Begriffe präzisieren und definieren.

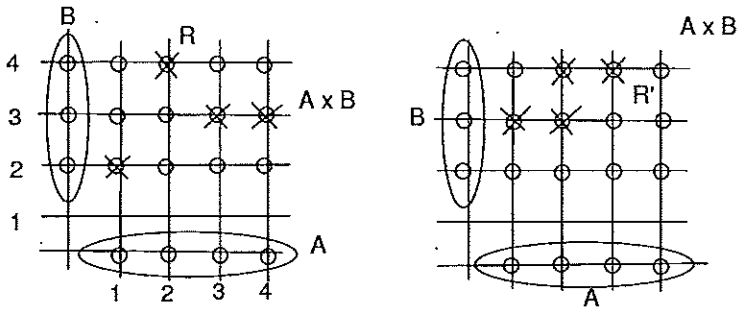
DEFINITION 5.1
Eine *n-stellige Relation* R ist eine Teilmenge eines kartesischen Produkts

$$A_1 \times ..... \times A_n.$$

Dies ist die allgemeinste Form des Relationenbegriffs, wie er auch in der Theorie der Relationalen Datenbanken verwendet wird. In der Mathematik sind die häufigsten Relationen zweistellig.

Bemerkung: Funktionen sind Sonderfälle von Relationen: Sei $f : A \to B$. Anstelle von f kann man auch die Relation $R_f \subseteq A \times B$, gegeben durch $(a, b) \in R_f \iff b = f(a)$, betrachten. Was ihren Informationsgehalt betrifft, sind f und R_f äquivalent.

Allgemein gilt: Eine Relation $R \subseteq A \times B$ ist eine Funktion genau dann, wenn es für jedes $a \in A$ genau ein $b \in B$ gibt, sodass $(a, b) \in R$ (vergleiche früher).



R ist Funktion; R' hingegen ist zwar Relation, aber keine Funktion.

Von besonderer Bedeutung ist bei den zweistelligen Relationen $R \subseteq A \times B$ der Sonderfall $A = B$:

DEFINITION 5.2
Eine *zweistellige Relation* R auf einer Menge A ist eine Teilmenge von $A \times A$.

Andere Notationsweisen für $(a, b) \in R$: $R(a, b)$ (Präfix-Notation), oder aRb (Infix-Notation).

BEISPIEL 5.1
Kleiner-Gleich-Relation:

$A = \mathbb{R}$,

$R = \{(x, y) \mid x, y \in \mathbb{R},\ x \leq y\} \subseteq A \times A = \mathbb{R}^2$.

Wir sehen, dass die Schreibweise $x \leq y$ nichts anderes als eine Infix-Notation für die Kleiner-Gleich-Relation ist.

Um das nächste Beispiel formulieren zu können, brauchen wir eine

DEFINITION 5.3
Die Potenzmenge $\mathcal{P}(M)$ ist die Menge aller Teilmengen von M:

$$\mathcal{P}(M) = \{M' \mid M' \subseteq M\}.$$

BEISPIEL 5.2
Teilmengenrelation.

M sei eine beliebige Menge.

$A = \mathcal{P}(M)$,

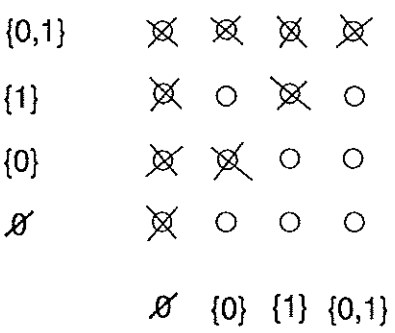
$R = \{(U, V) \mid U, V \in \mathcal{P}(M),\ U \subseteq V\} \subseteq A \times A = \mathcal{P}(M) \times \mathcal{P}(M)$.

Wieder ist "$U \subseteq V$" eine Infix-Notation.

Sei etwa $M = \{0, 1\}$. Dann ist $\mathcal{P}(M) = \{\phi, \{0\}, \{1\}, \{0, 1\}\}$, und R ist

$$\{(\phi, \phi), (\phi, \{0\}), (\phi, \{1\}), (\phi, \{0, 1\}), (\{0\}, \{0\}), (\{0\}, \{0, 1\}),$$
$$(\{1\}, \{1\}), (\{1\}, \{0, 1\}), (\{0, 1\}, \{0, 1\})\}.$$

Graphisch:



Offensichtlich ist die Teilmengenrelation (außer für M einelementig) keine Funktion!

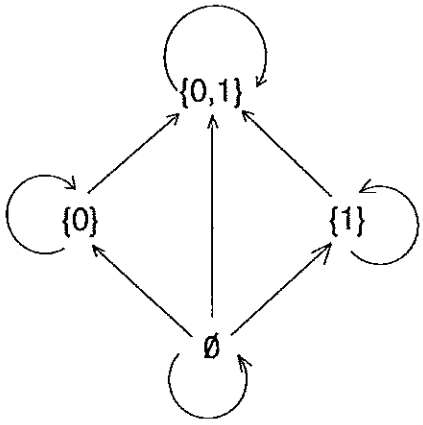
Es gibt noch eine andere graphische Darstellungsmöglichkeit für Relationen als die obige, nämlich die durch einen sogenannten **gerichteten Graphen**.

Erinnere: Ein gerichteter Graph besteht aus *Knoten* und *gerichteten Kanten* (Pfeilen) zwischen Knoten.

Um die Relation $R \subseteq A \times A$ darzustellen, wählen wir

- als Knoten die Elemente von A,
- für jedes Paar $(a, b) \in R$ eine gerichtete Kante vom Knoten a zum Knoten b.

BEISPIEL 5.3
Teilmengen-Relation auf $A = \mathcal{P}(\{0,1\})$.

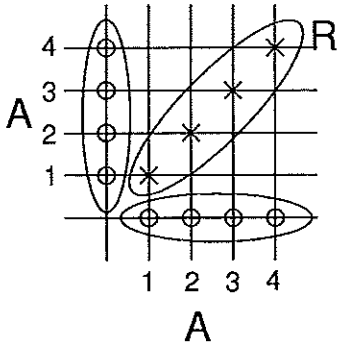


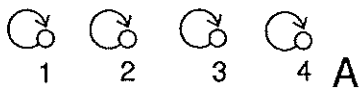
BEISPIEL 5.4
Gleichheitsrelation

A beliebig.

$R = \{(a,a) \mid a \in A\}$ (Infix-Notation: $a = a$).

Darstellung:





Die Gleichheitsrelation bildet stets eine Funktion, die sogenannte Identität(-sfunktion).

Zwei wichtige Spezialklassen von Relationen sind

- Äquivalenzrelationen
- Halbordnungen

5.1 Äquivalenzrelationen

DEFINITION 5.4
Eine Relation R auf der Menge A heißt *Äquivalenzrelation*, wenn gilt:

1. Reflexivität: aRa für alle $a \in A$
2. Symmetrie: $aRb \Rightarrow bRa$ für alle $a, b \in A$
3. Transitivität: $aRb \wedge bRc \Rightarrow aRc$ für alle $a, b, c \in A$

BEISPIEL 5.5
Die Gleichheitsrelation ist eine Äquivalenzrelation:

- $a = a$,
- $a = b \Rightarrow b = a$,
- $a = b \wedge b = c \Rightarrow a = c$.

Die Kleiner-Gleich-Relation und die Teilmengen-Relation sind *keine* Äquivalenzrelationen, da sie zwar reflexiv und transitiv sind, aber nicht symmetrisch.

BEISPIEL 5.6
Sei A eine Menge von Datensätzen und aRb die Relation „a und b liegen auf derselben Speicherseite". R ist offenbar reflexiv, symmetrisch und transitiv, also eine Äquivalenzrelation. Wir beobachten, dass wir A in nicht überlappende Teilmengen A_i (die Datensätze auf Speicherseite i) unterteilen können, sodass aRb gerade dann, wenn a und b in derselben Teilmenge A_i liegen.

Das letzte Beispiel lässt sich verallgemeinern:

DEFINITION 5.5
Zwei Mengen A, B heißen *disjunkt*, wenn sie sich nicht überlappen, d.h., wenn

$$A \cap B = \phi .$$

Eine Zerlegung der Menge A in nichtleere, paarweise disjunkte Teilmengen A_i $(i \in I)$ heißt *Klasseneinteilung* oder *Partition* von A.

Falls die Mengen A_i $(i \in I)$ eine Partition von A bilden, muss also gelten:

i $\cup_{i \in I} A_i = A$,

ii $A_i \cap A_j = \phi$ für $i \neq j$, und

iii $A_i \neq \phi$ für alle i.

BEISPIEL 5.7
Beispiele von Partitionen von $A = \{1, 2, 3, 4\}$ sind:
(a) $\{1\}, \{2\}, \{3\}, \{4\}$,
(b) $\{1, 2, 3, 4\}$.
(c) $\{1, 2\}, \{3, 4\}$
(d) $\{1\}, \{2, 3, 4\}$

THEOREM 5.1
Bilden die Teilmengen A_i $(i \in I)$ eine Partition von A, dann ist die durch

$$aRb \Leftrightarrow \text{ } a \text{ und } b \text{ liegen in derselben Klasse } A_i$$

definierte Relation eine Äquivalenzrelation.

BEISPIEL 5.8
$A = \{1, 2, 3, 4\}$.

- Partition $\{1\}, \{2\}, \{3\}, \{4\}$: a und b liegen in derselben Klasse A_i genau dann, wenn $a = b$. Die entsprechende Relation ist also hier die Gleichheitsrelation.
- Partition $\{1, 2, 3, 4\}$: Die entsprechende Äquivalenzrelation ist die *Allrelation*: Für je zwei Elemente $a, b \in A$ gilt *immer* aRb.

Es gilt auch die Umkehrung der obigen Aussage:

THEOREM 5.2
Ist R eine Äquivalenzrelation auf A, dann bilden die Mengen

$$K(a) = \{b \mid aRb\}$$

eine Partition von A, und es gilt

$$aRb \Leftrightarrow K(a) = K(b)\,.$$

Jeder Äquivalenzrelation R auf A entspricht daher eine Partition von A. Die Klassen A_i dieser Partition heißen *Äquivalenzklassen*.

BEISPIEL 5.9
$A = \mathbb{Z}$, $nRm \Leftrightarrow 2|n - m$ („2 teilt $n - m$“).

Dies ist offenbar eine Äquivalenzrelation; die zugehörigen Äquivalenzklassen sind

$$A_0 = \{n \in \mathbb{Z} \mid n \text{ gerade}\} \quad \text{und} \quad A_1 = \{n \in \mathbb{Z} \mid n \text{ ungerade}\}\,.$$

Allgemeiner: Für „$nRm \Leftrightarrow n - m$ durch k teilbar $(k \in \mathbb{N})$“erhalten wir die Äquivalenzklassen

$$A_i = \{n \in \mathbb{Z} \mid n \text{ durch } k \text{ ergibt Rest } i, i = 0, \ldots, k-1\}$$

D.h.,

$$n \equiv m \pmod{k}$$

für die zuletzt definierte Äquivalenzrelation.

5.2 Halbordnungen

DEFINITION 5.6
Eine Relation R auf der Menge A heißt *Halbordnung* (*partial order*), wenn sie folgende Eigenschaften hat:

1. Reflexivität: aRa für alle $a \in A$
2. Antisymmetrie: $aRb \wedge bRa \Rightarrow a = b$ für alle $a, b \in A$
3. Transitivität: $aRb \wedge bRc \Rightarrow aRc$ für alle $a, b, c \in A$

Bemerkung: Wir sehen, dass hier gegenüber der Definition der Äquivalenzrelation die Symmetrie durch die Antisymmetrie ausgetauscht wurde:
Während bei der Äquivalenzrelation aus aRb stets bRa folgt, schließen sich bei der Halbordnung aRb und bRa gegenseitig aus, es sei denn, $a = b$.

BEISPIEL 5.10
Die Teilmengenrelation ist eine Halbordnung:
- $U \subseteq U$,
- $U \subseteq V \wedge V \subseteq U \Rightarrow U = V$,
- $U \subseteq V \wedge V \subseteq W \Rightarrow U \subseteq W$.

BEISPIEL 5.11
Die Kleiner-Gleich-Relation auf $\mathbb{R}$ ist ebenfalls eine Halbordnung. Sie lässt sich via

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \leq \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} :\Leftrightarrow x_i \leq y_i, \ \forall i$$

zu einer Halbordnung auf $\mathbb{R}^n$ erweitern.

DEFINITION 5.7
Eine Halbordnung R auf A, für die gilt, dass

$$aRb \vee bRa, \ \forall a, b \in A\,,$$

heißt *lineare Ordnung* oder *Totalordnung*.

BEISPIEL 5.12
Die Kleiner-Gleich-Relation auf $\mathbb{R}$ ist eine Totalordnung, nicht aber die auf $\mathbb{R}^n$. Es sind etwa folgende Vektoren unvergleichbar:

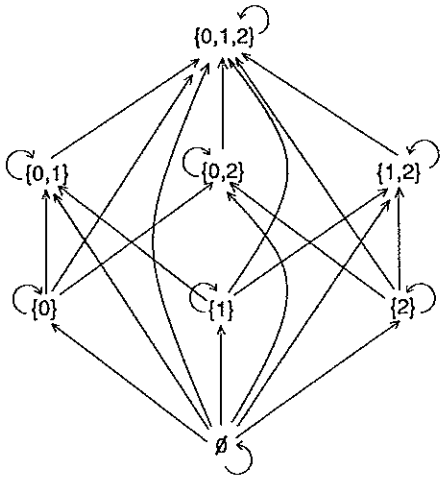
$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Die Teilmengenrelation ist ebenfalls keine Totalordnung: $\{1, 2\}$ und $\{2, 3\}$ sind unvergleichbar.

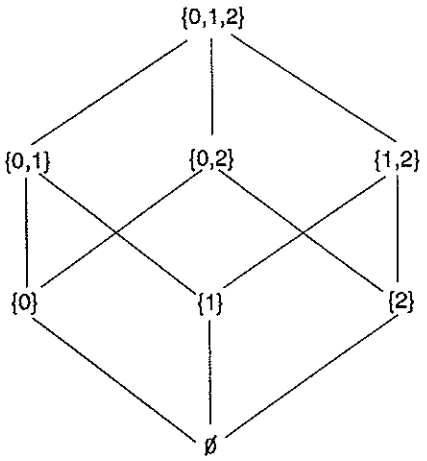
Wie alle anderen Relationen auf endlichen Mengen A lassen sich auch Halbordnungen durch gerichtete Graphen darstellen. Unter Ausnutzung von Reflexivität und Transitivität können wir diese Darstellungen vereinfachen:

- Entfernen aller Schlingen (aRa muss für Halbordnungen sowieso immer gelten).
- Falls aRb und bRc, also auch aRc, lassen wir den Pfeil $a \to c$ weg (dieser ergibt sich automatisch aus der Transitivität).

BEISPIEL 5.13 (HASSE-DIAGRAMM)
Teilmengenrelation mit $M = \{0, 1, 2\}$, d.h., auf $A = \mathcal{P}(\{0, 1, 2\})$.



Vereinfachen zu:



Hier wurden zusätzlich gerichtete Kanten durch ungerichtete ersetzt, mit der Konvention, dass stets Pfeilrichtung nach oben gemeint ist. Diese graphische Darstellung einer Halbordnung heißt *Hasse-Diagramm*.

Analysis

Die Analysis beschäftigt sich mit Folgen, Funktionen und Limiten. Wichtige Teilgebiete sind die Differential- und Integralrechnung, aber auch die Maßtheorie und die Topologie. Funktionen und Folgen sowie deren Eigenschaften werden sowohl im Bereich der reellen als auch der komplexen Zahlen untersucht.

6 Reelle Funktionen $f : \mathbb{R} \to \mathbb{R}$

> *Jede mathematische Formel in einem Buch halbiert die Verkaufszahlen dieses Buches.* — *Stephen Hawking*

Einer der grundlegenden Begriffe der Mathematik ist der der Menge, ein Begriff, den wir intuitiv erfassen, im mathematischen Sinn aber nicht richtig definieren können. Von Georg Cantor, dem Vater der Mengenlehre, stammt folgende Formulierung:

> *Eine Menge ist eine Zusammenfassung bestimmter wohlunterscheidbarer Objekte unserer Anschauung oder unseres Denkens, welche die Elemente der Menge genannt werden, zu einem Ganzen.*

Eine Menge wird demnach über ihre Elemente erklärt. Dies ist keine (mathematische) Definition, denn der Begriff der *Menge aller Mengen*, also alle Mengen zusammengefasst, führen zu einem Paradoxon (Russel'sche Antinomie), einem Widerspruch im zugrunde liegenden Axiomensystem ZFC, d. h., diese „Menge" ist/wäre nicht wohldefiniert. Diese „Menge aller Mengen" existiert somit nicht bzw. ist keine Menge.

Wir wollen nun aber nicht weiter auf den Begriff der Menge eingehen, da er einerseits bereits in der Schule intensiv behandelt wurde, andererseits wir ein intuitiv richtiges Verständnis für Mengen haben (sollten).

Ebenso wird der Begriff *Element* nicht weiter erklärt bzw. definiert.

Die moderne Mengenlehre und damit ein Großteil der modernen Mathematik basiert auf den Zermelo-Fraenkel Axiomen der Mengenlehre.

Wie grundlegend die Begriffe Menge und Element sind, zeigt uns das mathematische „Konzept" der Funktion, der ohne diese beiden Begriffe „unmöglich" wäre.

Dazu betrachten wir zuerst Vorschriften, die Elemente oder auch Teilmengen einer Menge Elementen bzw. Teilmengen anderer Mengen zuordnen. Solche Vorschriften nennen wir allgemein Relationen[6], zwei oder mehr Elemente stehen in Relation zueinander.

Abbildungen wiederum sind spezielle Relationen, die Elementen der einen Menge ein oder mehrere Elemente einer anderen Menge zuordnen, oder auch nicht. Funktionen sind

[6] Der Begriff Relation wird in der Algebra, einem Teilgebiet der Mathematik näher behandelt (siehe später).

wiederum spezielle Abbildungen, die jedem Element einer Menge genau ein Element einer anderen Menge zuordnen. Wir wollen hier eine bzw. die mathematische widerspruchsfreie Definition geben.

Um die Mengenlehre wieder ins Spiel zu bringen, im Folgenden einige Definitionen:

DEFINITION 6.1
Die *Produktmenge* oder das *kartesische Produkt* $A \times B$ zweier Mengen A und B ist definiert als die Menge aller möglichen geordneten Paare (a, b), mit $a \in A$ und $b \in B$. Symbolisch:

$$A \times B = \{(a, b) \mid a \in A, b \in B\}\,.$$

Hierbei ist ein geordnetes Paar definiert als $(a, b) = \{a, \{a, b\}\}$.

DEFINITION 6.2
Gegeben zwei Mengen A und B. Jede Teilmenge $C \subseteq A \times B$ heißt *Graph*.

DEFINITION 6.3
Sei $R \subseteq A \times B$. R definiert eine *Relation* auf A und B. Für alle $(a, b) \in R$ gilt: aRb, d. h., $a \in A$ steht in Relation zu $b \in B$.

Diese Begriffe benötigen wir, um eine Funktion als spezielle Relation definieren zu können. Relationen, insbesondere spezielle Relationen wie Ordnungen, Äquivalenzrelationen und deren Eigenschaften, werden im Kapitel Algebra behandelt.

DEFINITION 6.4
Eine Funktion $f : A \to B$ ist eine Relation, die jedem $a \in A$ genau ein $b \in B$ zuordnet:

$$\forall a \in A \quad \exists!\, b \in B : \; f(a) = b\,.$$

Wir schreiben $f(a) = b$. Die Menge A wird hierbei auch als *Definitionsmenge* oder auch *Urbild(menge)* bezeichnet, die Menge B als *Wertemenge*, *Bild(menge)* oder auch *Image*.

Da eine Funktion eine spezielle Relation ist, und Relationen über ihre Graphen festgelegt sind, ist auch eine Funktion über ihren Graphen festgelegt. Wir beachten hier, dass in keiner der obigen Definitionen eine Aussage getroffen wurde, welcher Art die Elemente der Mengen A und B sind. Insofern ist der Begriff der Relation und auch jener der Funktion ein Abstraktum.

Zulässige Mengen für A und B sind unter anderem der Körper der reellen Zahlen $\mathbb{R}$, oder aber die Vektorräume $\mathbb{R}^n$ mit Dimension n, aber etwa auch die Menge der komplexen Zahlen $\mathbb{C}$.

Beispiele für einfache Funktionen zwischen diesen Mengen sind die *Identität*

$$id_A : A \to A, \; a \mapsto a, \; \forall a \in A\,,$$

wobei diese Definition wieder von der Art der Menge A unabhängig ist,

Projektionen wie

$$P_x : \mathbb{R}^2 = \mathbb{R} \times \mathbb{R} \to \mathbb{R},\ (x, y) \mapsto x,\ \forall (x, y) \in \mathbb{R}^2\,,$$

oder

$$P_x : \mathbb{R}^n \to \mathbb{R}^k,\ (x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_k)\,,$$

oder die stückweise definierte Betragsfunktion

$$f : \mathbb{R} \to \mathbb{R}_0^+,\ x \mapsto |x| = \begin{cases} x & \text{falls } x \geq 0 \\ -x & \text{falls } x < 0 \end{cases}\,.$$

Auch die Addition

$$f : \mathbb{R} \times \mathbb{R} \to \mathbb{R},\ (a, b) \in \mathbb{R} \times \mathbb{R} \mapsto a + b \in \mathbb{R}$$

oder aber allgemein für Vektoren

$$f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n,\ (a, b) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto a + b \in \mathbb{R}^n$$

ist eine Funktion, wobei hier »+« oft auch als binärer Operator bezeichnet wird, d.h., er wird auf zwei Operanden angewendet.

Wir orientieren uns oft an „Bildern" bzw. Grafiken von Funktion, den sogenannten Graphen. Wir geben hier die zu obiger Definition äquivalente Formulierung für Funktionen.

DEFINITION 6.5
Sei $f : A \to B,\ a \mapsto b = f(a)$ eine Funktion. Dann heißt die Menge

$$G = \{(a, f(a)) \,|\, a \in A,\ f(a) \in B\} \subseteq A \times B$$

Graph der Funktion f.

BEISPIEL 6.1
Die reelle Funktion $f(x) = x^2$ hat den Graphen $G = \{(x, x^2) \mid x \in \mathbb{R}\} \subseteq \mathbb{R}^2$.

Oft sehen wir nur das Bild einer Funktion, ohne zu wissen, wie die tatsächliche Funktion aussieht, etwa wenn wir ein Foto betrachten, als Abbild einer Projektion der Realität. Möglicherweise sind wir aber an dem Urbild interessiert, d. h., daran, woher das Bild kommt (etwa astronomische Aufnahmen). Leider kann es aber auch vorkommen, dass verschiedene Urbilder das gleiche Bild ergeben, wie etwa der Wert 4 bei der Funktion $x \mapsto y = x^2$, der sowohl von (-2) als auch von 2 kommen kann.

Um solche Fälle besser handhaben zu können, werden die folgenden Eigenschaften von Funktionen definiert.

DEFINITION 6.6
Eine Funktion $f : A \to B,\ a \mapsto b = f(a)$ heißt *injektiv*, wenn nur jeweils genau ein Element von A auf ein Element von B abgebildet wird, d. h., wenn gilt

$$f(a) = f(b) \Leftrightarrow a = b\,.$$

DEFINITION 6.7
Eine Funktion $f : A \to B,\ a \mapsto b = f(a)$ heißt *surjektiv*, wenn jedes Element $b \in B$ Bild mindestens eines Urbilds $a \in A$ ist, oder

$$\forall b \in B \quad \exists\, a \in A : b = f(a)\,.$$

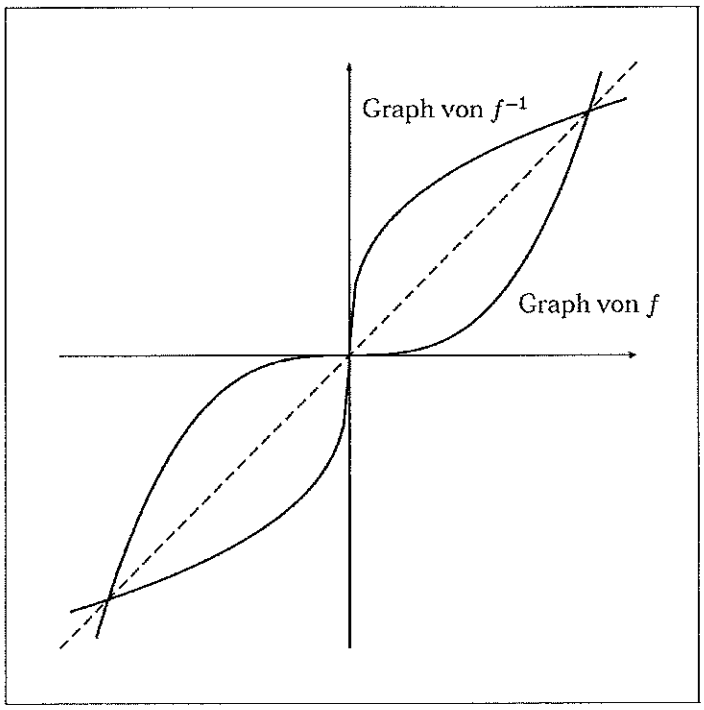
DEFINITION 6.8
Eine Funktion heißt *bijektiv*, wenn sie sowohl injektiv als auch surjektiv ist.

Wann können wir also von den Bildern wieder auf die Urbilder in eindeutiger Weise schließen? Doch genau dann, wenn es zu jedem Bild ein Urbild gibt, d. h., alle Elemente der Bildmenge »getroffen« werden, d. h., die Funktion surjektiv ist, und weiters, wenn dieses Urbild eindeutig ist, d. h., jedes Bild genau ein Urbild hat, die Funktion also injektiv ist. Zusammengefasst muss die Funktion also bijektiv sein, damit wir von ihren Bildern wieder in eindeutiger Weise zu ihren Urbildern kommen können.

DEFINITION 6.9
Ist die Funktion $f : A \to B,\ a \mapsto b = f(a)$ bijektiv, so heißt die Funktion

$$f^{-1} : B \to A,\ b = f(a) \mapsto a$$

die *Umkehrfunktion* oder *Inverse* von f.



Es ist also nicht jede Funktion umkehrbar, wie schon das obige Beispiel der Funktion $f(x) = x^2$ zeigte.

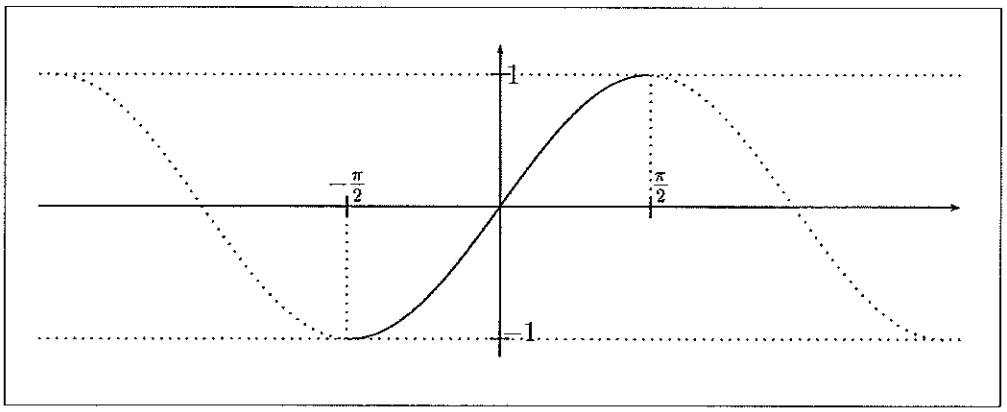
Oft ist es jedoch möglich bzw. sinnvoll, Funktionen auf gewisse Teilmengen der Definitionsmenge einzuschränken: $f(x) = x^2$ etwa auf $\mathbb{R}^+$, sodass sie (auf $\mathbb{R}^+$) invertierbar wird.

DEFINITION 6.10
Sei $f : A \to B,\ a \mapsto b = f(a)$ eine Funktion und $H \subseteq A$ eine Teilmenge von A. Dann heißt die Funktion

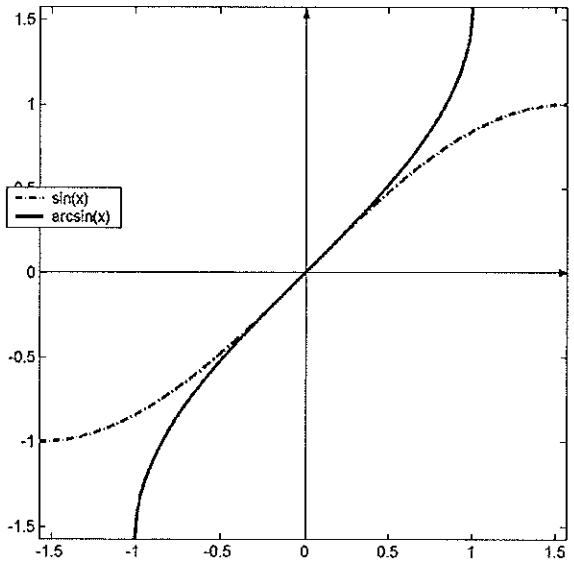
$$f_{|H} : H \to B,\ a \mapsto b = f(a)$$

Einschränkung von f auf H.

Ein Beispiel dafür ist der Sinus $\sin(x)$,



der im Intervall $[-\frac{\pi}{2}, \frac{\pi}{2}]$ bijektiv ist, und dort die Umkehrfunktion Arcus-Sinus $\arcsin(x)$ hat.



Wenn wir uns die auf das Intervall $[-\frac{\pi}{2}, \frac{\pi}{2}]$ eingeschränkte Funktion $\sin(x)$ ansehen, merken wir, dass $\sin(x)$ dort von links nach rechts ansteigend ist, wir sagen »streng monoton wachsend«.

DEFINITION 6.11
Eine Funktion $f : A \to B,\ a \mapsto b = f(a)$ heißt *monoton wachsend*, falls gilt

$$f(x) \leq f(y) \qquad \forall x < y\,.$$

Sie heißt *streng monoton wachsend*, wenn gilt

$$f(x) < f(y) \qquad \forall x < y\,.$$

DEFINITION 6.12
Eine Funktion $f : A \to B,\ a \mapsto b = f(a)$ heißt *monoton fallend*, falls gilt

$$f(x) \geq f(y) \qquad \forall x < y\,.$$

Sie heißt *streng monoton fallend*, wenn gilt

$$f(x) > f(y) \qquad \forall x < y\,.$$

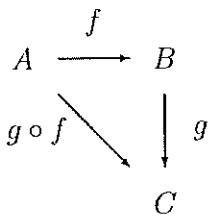
Streng monotone Funktionen sind gute Kandidaten für bijektive und daher umkehrbare Funktionen.

Da Funktionen Zuordnungsvorschriften zwischen Mengen und ihren Elementen sind, können wir auch von einer Menge zur anderen mittels dieser Zuordnungsvorschriften wandern. Sind gewisse Voraussetzungen erfüllt, so können wir zwei oder mehrere Funktion verknüpfen, d. h., durch eine Funktion ersetzen.

DEFINITION 6.13
Sind $f : A \to B$ und $g : B \to C$ Funktionen, so ist die *zusammengesetzte Funktion* definiert als

$$h = g \circ f : A \to C,\ a \in A \mapsto c = g(f(a)) \in C\,.$$

Eine übersichtliche Darstellungsform für Funktionen und ihre Verknüpfungen ist ein so genanntes *kommutatives Diagramm*, in dem zu je zwei verschiedenen Mengen alle Funktionen (auch zusammengesetzte), die die eine in die andere Menge überführen, übereinstimmen.



Spezielle reelle Funktionen

Geraden:	$x \mapsto k\,x + d$	$\mathbb{R} \to \mathbb{R}$
	k heißt Steigung oder Anstieg	
Potenzfunktionen:	$x \mapsto x^n \quad n \in \mathbb{R}$	$D \to W$
Wurzelfunktionen:	$x \mapsto \sqrt[n]{x} = x^{1/n} \quad n \in \mathbb{N}$	$D \to W$
	$D = W = \mathbb{R}_0^+$ falls n gerade $D = W = \mathbb{R}$ falls n ungerade	
Polynome:	$x \mapsto p(x) = \sum_{k=0}^{n} a_k x^k$	$\mathbb{R} \to \mathbb{R}$
Rationale Funktionen:	$x \mapsto \dfrac{p(x)}{q(x)}$	$D \to \mathbb{R}$
	$p(x)$, $q(x)$ Polynome $D = \mathbb{R} \setminus \{\text{Nullstellen von } q\}$	
Exponentialfunktion:	$x \mapsto \exp(x) = e^x$	$\mathbb{R} \to \mathbb{R}^+$
	$x \mapsto a^x \quad a \in \mathbb{R}^+$	$\mathbb{R} \to \mathbb{R}^+$
Logarithmusfunktion:	$x \mapsto \log(x) = \ln(x)$	$\mathbb{R}^+ \to \mathbb{R}$
	Inverse Funktion zur *Exponentialfunktion*	
	$x \mapsto \log_a(x) \quad a \in \mathbb{R}^+$	$\mathbb{R}^+ \to \mathbb{R}$
Winkelfunktionen:	$x \mapsto \sin(x)$	$\mathbb{R} \to [-1, 1]$
	$x \mapsto \cos(x)$	$\mathbb{R} \to [-1, 1]$
	$\tan(x)$, $\cot(x)$	
	$\arcsin(x)$, $\arccos(x)$	

6.1 Stetige Funktionen $f : \mathbb{R} \to \mathbb{R}$

Eine besondere Klasse von Funktionen sind die stetigen Funktionen. Als Faustregel können wir uns merken, dass das jene Funktionen sind, die wir in einem Zug, d. h., durchgehend ohne Unterbrechung, zeichnen können.

Um wiederum eine mathematische Formulierung dieser „Eselsbrücke“ zu bekommen, müssen wir uns vorher den Begriff des Grenzwertes in Erinnerung rufen. Zuerst betrachten wir reelle Zahlenfolgen $\langle a_n \rangle$ und definieren:

DEFINITION 6.14
Eine *Folge* ist eine Abbildung $a : \mathbb{N} \to \mathbb{R}$, die jedem $n \in \mathbb{N}$ genau ein $a(n) = a_n \in \mathbb{R}$ zuordnet. Wir notieren diese Funktion auch mittels ihrer Funktionswerte als $\langle a_n \rangle$.

DEFINITION 6.15
Eine Folge $\langle a_n \rangle$ *konvergiert* gegen ein $a \in \mathbb{R}$, wenn es für jedes $\varepsilon > 0$ eine natürliche Zahl $N \in \mathbb{N}$ gibt, sodass $a_n \in (a - \varepsilon, a + \varepsilon)$ für alle $n > N$.

Die Folge heißt *konvergent*, a heißt *Grenzwert* oder *Limes* der Folge.

Wir schreiben

$$\lim_{n\to\infty} a_n = a \quad \Leftrightarrow \quad \forall \varepsilon > 0 \; \exists N \in \mathbb{N} : \; |a_n - a| < \varepsilon \;\; \forall n > N \, .$$

Salopp ausgedrückt, können wir auch sagen, dass die Folge gegen den Grenzwert strebt ($\langle a_n \rangle \to a$) bzw. ihm ab einem Index beliebig nahe kommt. In der Mathematik sprechen wir von einer sogenannten ε-Umgebung

$$U_\varepsilon(a) = \{x \,|\, |x - a| < \varepsilon\} \, .$$

Diese Definition können wir sofort auch für Funktionen anpassen.

DEFINITION 6.16
Sei $f : A \to \mathbb{R}$ eine Funktion. Ein $a \in \mathbb{R}$ heißt *Grenzwert* oder *Limes* der Funktion f an der Stelle x_0, wenn für jede Argumentfolge $\langle x_n \rangle$, die gegen x_0 strebt, d. h., $\langle x_n \rangle \to x_0$ oder $\lim_{n\to\infty} x_n = x_0$, gilt: $\lim_{n\to\infty} f(x_n) = a$.
Wir schreiben dann

$$\lim_{x\to x_0} f(x) = a \, .$$

In mathematischer „Kurzschrift“ schreiben wir

$$\lim_{x\to x_0} f(x) = a \quad \Leftrightarrow \quad \forall \varepsilon > 0 \; \exists \delta > 0 : \; |f(x) - a| < \varepsilon \;\; \forall |x_0 - x| < \delta \, .$$

BEMERKUNG 6.1
Das Argument x_0 muss nicht in der Definitionsmenge der Funktion liegen, ebensowenig muss der Limes in der Wertemenge der Funktion liegen. Der Limes kann auch ∞ sein.

Jetzt können wir die Stetigkeit von Funktionen definieren.

DEFINITION 6.17
Eine Funktion $f : A \to B$ heißt *stetig an der Stelle* $x_0 \in A$, wenn $\lim_{x \to x_0} f(x)$ existiert und $\lim_{x \to x_0} f(x) = f(x_0)$.

Eine Funktion heißt *stetig*, wenn sie in allen Punkten der Definitionsmenge A stetig ist, d. h.,

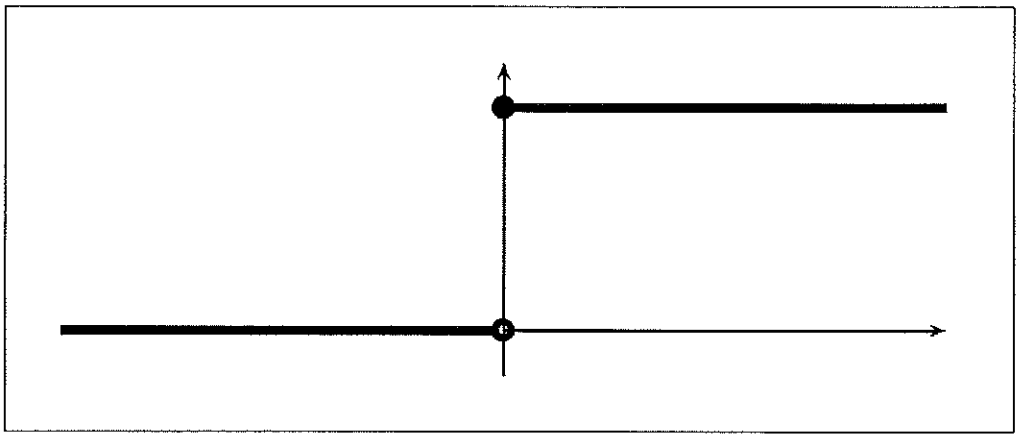
$$f : A \to B \text{ stetig} \quad \Leftrightarrow \quad \lim_{x \to x_0} f(x) = f(x_0) \quad \forall x_0 \in A$$

oder

$$f : A \to B \text{ stetig} \quad \Leftrightarrow \quad \forall x_0 \in A : \quad \forall \varepsilon > 0 \; \exists \delta > 0 : \; |f(x) - f(x_0)| < \varepsilon \quad \forall |x - x_0| < \delta$$

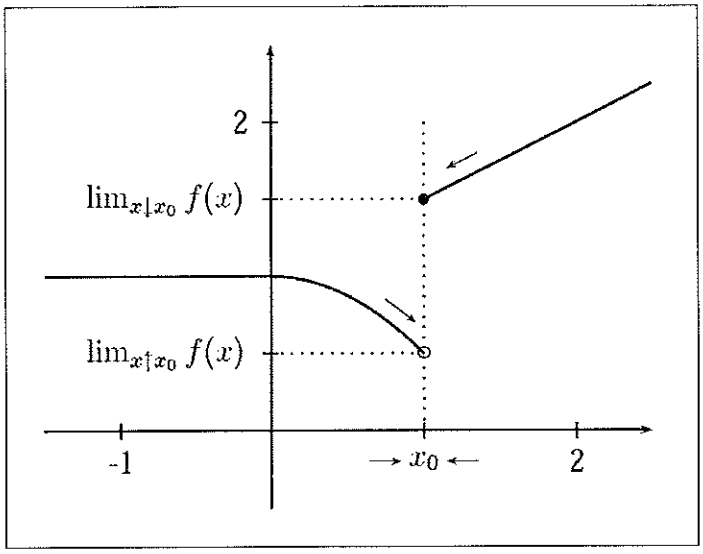
das ist die „berühmte" ε-δ-Definition der Stetigkeit von Funktionen.

BEISPIEL 6.2
Die Funktion $\sin : \mathbb{R} \to [-1, 1]$ ist eine stetige Funktion, ebenso $|x|$.
Ebenso ist die Funktion $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$, $x \mapsto \frac{1}{x}$ stetig, die *Sprungstelle* bei 0 gehört nicht zum Definitionsbereich der Funktion.



Die Heaviside-Funktion $h(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$ ist überall stetig außer in der Sprungstelle bei 0.

Die Heaviside-Funktion, wie auch andere, zeigt, dass es noch eine Differenzierung des Begriffs des Grenzwertes bzw. der Stetigkeit geben muss, nämlich linksseitig- bzw. rechtsseitig stetig, d. h., ob wir uns von links oder von rechts dem entsprechenden Argument x_0 nähern.



DEFINITION 6.18
Eine Funktion $f : A \to B$ heißt in x_0 *rechtsseitig stetig*, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, sodass

$$|f(x) - f(x_0)| < \varepsilon \qquad \forall x \in [x_0, x_0 + \delta)\,.$$

Wir notieren den *rechtsseitigen Limes*

$$\lim_{x \to x_0^+} f(x) = \lim_{x \downarrow x_0} f(x) = f(x_0) \Leftrightarrow \forall \varepsilon > 0\ \exists\, \delta > 0 :\ |f(x) - f(x_0)| < \varepsilon \quad \forall x \in [x_0, x_0 + \delta)\,.$$

Eine Funktion $f : A \to B$ heißt in x_0 *linksseitig stetig*, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, sodass

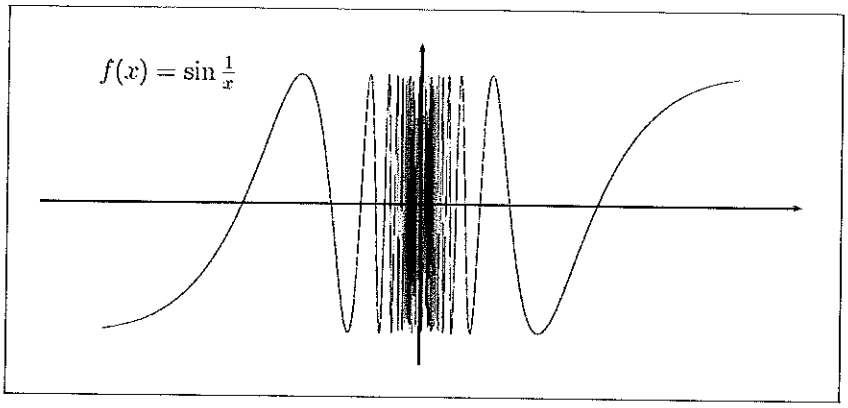
$$|f(x) - f(x_0)| < \varepsilon \qquad \forall x \in (x_0 - \delta, x_0]\,.$$

Wir notieren den *linksseitigen Limes*

$$\lim_{x \to x_0^-} f(x) = \lim_{x \uparrow x_0} f(x) = f(x_0) \Leftrightarrow \forall \varepsilon > 0\ \exists\, \delta > 0 :\ |f(x) - f(x_0)| < \varepsilon \quad \forall x \in (x_0 - \delta, x_0]\,.$$

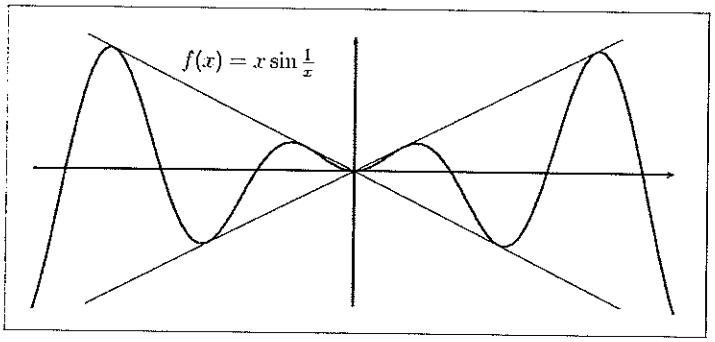
In diesem Sinn ist die oben definierte Heaviside-Funktion im Punkt 0 rechtsseitig stetig, da $f(0) = 1$ und der Funktionswert 1 auf dem rechten „Ast" der Funktion liegt. Sie ist im Punkt 0 nicht linksseitig stetig, da $\lim_{x\uparrow 0} f(x) = 0 \neq f(0) = 1$.

BEMERKUNG 6.2
Eine Funktion ist in x_0 genau dann stetig, wenn sie dort links- und rechtsseitig stetig ist, und daher dort auch die Funktionswerte (Limiten) übereinstimmen.



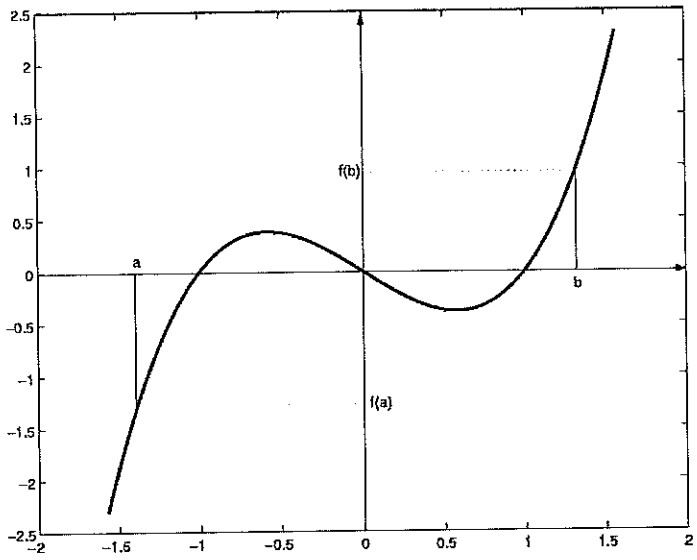
BEISPIEL 6.3
Die Funktion $\sin \frac{1}{x}$ ist im Nullpunkt weder links- noch rechtsseitig stetig. Es kann nachgewiesen werden, dass sie in jeder noch so kleinen ε-Umgebung um 0 jeden beliebigen Wert zwischen (-1) und 1 annehmen kann.
Die Funktion $x \sin \frac{1}{x}$ hingegen ist im Nullpunkt stetig, da die Funktion x den betragsmäßig mit 1 beschränkten Sinus dominiert.

BEMERKUNG 6.3
Das Problem einer Darstellung am Computer bzw. Papier wird in der Abbildung zur Funktion $\sin \frac{1}{x}$ offenbar: Computer rechnen diskret bis zu einer festgelegten Genauigkeit bzw. Displays etc. können nur diskrete Punkte darstellen, sodass die Darstellung der Funktion nicht exakt ist. Vor allem in der Nähe des Unstetigkeitspunktes 0, wo es immer mehr Funktionswerte zu berechnen gäbe, versagt der Computer.

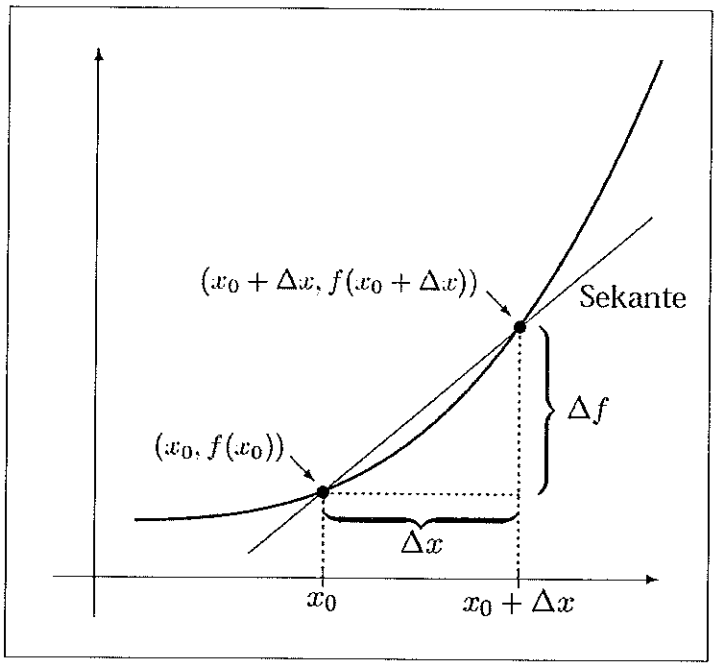


Für stetige Funktionen gilt folgender Satz

THEOREM 6.1 (MITTELWERTSATZ)
Ist $f : A \to B$ eine auf dem Intervall $[a, b]$ stetige Funktion und gilt $f(a) < 0 < f(b)$, dann gibt es zumindest eine Nullstelle ξ in $[a, b]$, d. h., $f(\xi) = 0$, $a < \xi < b$.


f(b)
a
b
f(a)

6.2 Differenzierbarkeit



Der *Differenzenquotient*

$$\frac{\Delta f}{\Delta x} = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

ist die Steigung der Sekante durch die benachbarten Punkte $(x_0, f(x_0))$ und $(x_0 + \Delta x, f(x_0 + \Delta x))$ auf dem Graphen der Funktion $f(x)$.

Lassen wir Δx nun immer kleiner werden, so wird diese Sekante zur Tangente.
Der *Differentialquotient*

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}$$

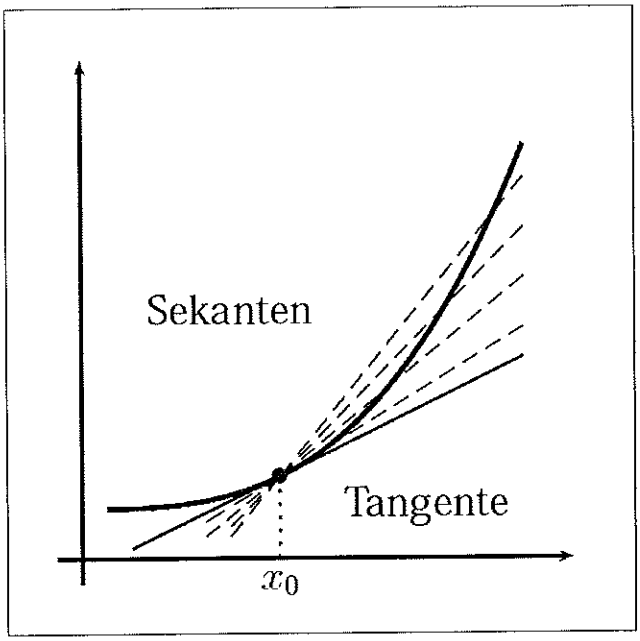
ist die Steigung der Tangente im Punkt x_0, sodass die Tangente in x_0 folgende Geradengleichung erfüllt

$$y = f(x_0) + f'(x_0)(x - x_0)\,.$$

DEFINITION 6.19
Sei $f : A \to \mathbb{R}$ eine reelle Funktion, A ein offenes Intervall.
f heißt *differenzierbar* in A, wenn der *Differentialquotient*

$$f'(x_0) = \frac{df(x_0)}{dx} = \left.\frac{df}{dx}\right|_{x_0} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

für alle Punkte $x_0 \in A$ existiert.
Die Funktion $f'(x) : A \to \mathbb{R}$ heißt dann erste *Ableitung* der Funktion f in A.



Wir dürfen hier keine geschlossenen Intervalle nehmen, da an den „Ecken", d. h. an den Randpunkten, keine Tangenten gelegt werden können. Ebenso können an „Ecken" eines Funktionsgraphen keine Tangenten gelegt werden (etwa an $|x|$ im Punkt 0), sodass diese Funktionen dort nicht differenzierbar sein können. Die stetige Betragsfunktion ist daher nicht differenzierbar auf $\mathbb{R}$, sehr wohl ist sie aber auf $\mathbb{R}^+ = (0, \infty)$ differenzierbar und hat dort die konstante Ableitung 1.

Umgekehrt gilt aber Folgendes:

THEOREM 6.2
Ist eine Funktion $f(x)$ in x_0 differenzierbar, so ist f in x_0 stetig.

Ableitungen einiger wichtigen Funktionen $f(x)$			
konstant c	0	e^x	e^x
$x^n,\ n \in \mathbb{N}$	$n \cdot x^{n-1}$	a^x	$a^x \cdot \log a$
$x^\alpha,\ \alpha \in \mathbb{R}$	$\alpha \cdot x^{\alpha-1}$	$\log x = \ln x$	$\frac{1}{x}$
$\sin x$	$\cos x$	$\log_a x$	$\frac{1}{x} \log_a e = \frac{1}{x \log a}$
$\cos x$	$-\sin x$	x^x	$x^x(\log x + 1)$

Wichtige Differentiationsregeln	
Mult. mit einer Konstanten c	$(c \cdot f(x))' = c \cdot f'(x)$
Summenregel	$(f(x) + g(x))' = f'(x) + g'(x)$
Produktregel	$(f(x) \cdot g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$
Quotientenregel	$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{(g(x))^2}$
Kettenregel	$(f(g(x))' = f'(g(x)) \cdot g'(x)$
Ableitung der Inversen	$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$

Für die Ableitung der Inversen muss natürlich die inverse Funktion existieren und differenzierbar sein, weiters muss auch $f'(x) \neq 0$ sein. Die obige Formel folgt dann aus der Kettenregel, angewendet auf $f(f^{-1}(x)) = x$ ($\rightarrow$ einfache Übung).

BEISPIEL 6.4 (DIE ABLEITUNG VON $\arcsin(x)$)
Der arcsin ist die Umkehrfunktion von sin.
Daher gilt:

$$(\arcsin(x))' = \frac{1}{\sin'(\arcsin(x))} = \frac{1}{\cos(\arcsin(x))}.$$

Weiters wissen wir, dass $\sin^2 x + \cos^2 x = 1$; also gilt

$$\cos(\arcsin(x)) = \sqrt{1 - \sin^2(\arcsin(x))} = \sqrt{1 - x^2}$$

und damit insgesamt

$$(\arcsin(x))' = \frac{1}{\sqrt{1 - x^2}}.$$

BEISPIEL 6.5
Für die Exponentialfunktionen a^x gilt

$$a^x = \exp(\log(a^x)) = \exp(x \log a)$$

und daher ist

$$(a^x)' = (\exp(x \log(a)))' = \exp(x \log(a)) \cdot \log a = a^x \log a.$$

Analog ist $(2^x)' = 2^x \log 2$.

BEISPIEL 6.6 (LOGARITHMISCHE ABLEITUNG VON x^x)
Die Funktion x^x wird mit Hilfe der logarithmischen Ableitung differenziert.
Da

$$x^x = \exp(\log(x^x)) = \exp(x \cdot \log x),$$

gilt für deren Ableitung mit Hilfe der Kettenregel

$$(\exp(x \cdot \log x))' = \exp(x \log x)(1 \cdot \log x + x \cdot \frac{1}{x}) = x^x(\log x + 1).$$

BEISPIEL 6.7 (DIE ABLEITUNG VON e^x)
Die Exponentialfunktion ist definiert durch

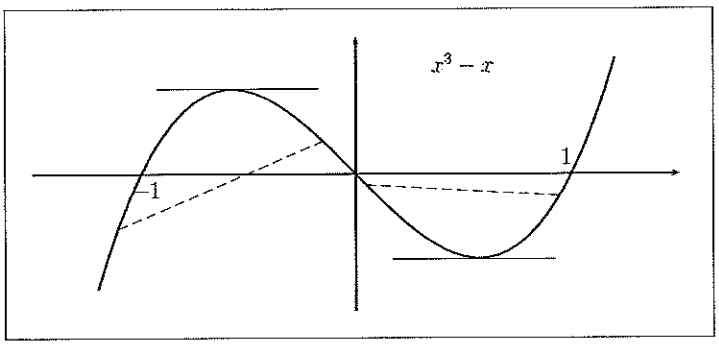
$$e^x = \sum_{k=0}^{\infty} \frac{1}{k!} x^k = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots$$

Somit ist

$$(e^x)' = \lim_{h\to 0} \frac{e^{x+h} - e^x}{h} = e^x \lim_{h\to 0} \frac{e^h - 1}{h} = e^x \cdot \lim_{h\to 0} (\frac{1}{h}(1 + h + \frac{1}{2!}h^2 + \cdots - 1))$$
$$= e^x \cdot \lim_{h\to 0} (\frac{h}{h}(1 + \frac{1}{2!}h + \frac{1}{3!}h^2 + \cdots)) = e^x \cdot \lim_{h\to 0} (1 + \frac{1}{2!}h + \frac{1}{3!}h^2 + \cdots)) = e^x$$

DEFINITION 6.20 (EXTREMA)
Eine Funktion $f : A \to \mathbb{R}$ besitzt in x_0 ein *lokales Maximum* (*lokales Minimum*), wenn für alle Punkte x einer Umgebung $U_\varepsilon(x_0)$ gilt, dass $f(x_0) \geq f(x)$ ($f(x_0) \leq f(x)$).

Eine Funktion $f : A \to \mathbb{R}$ besitzt in x_0 ein *globales Maximum* (*globales Minimum*), wenn für alle Punkte $x \in A$ gilt, dass $f(x_0) \geq f(x)$ ($f(x_0) \leq f(x)$).



Zur Berechnung der Extrema betrachten wir obige Abbildung der Funktion $x^3 - x$ und bemerken, dass die Funktion in den lokalen Extremwerten horizontale Tangenten besitzt, d. h., dass dort $f'(x) = 0$. Weiters ist die Funktion bei den Extremwerten gekrümmt. Die Sekanten am Graphen der Funktion liegen unter dem Funktionsgraphen beim Maximum, über dem Funktionsgraphen beim Minimum. Dazu definieren wir folgende Begriffe der Krümmung:

DEFINITION 6.21
Eine Funktion $f : A \to B$ heißt *konvex* in einem Intervall $[a, b]$, wenn die Funktion dort immer unter der Sekante liegt, d. h.,

$$f(hx_1 + (1-h)x_2) \leq hf(x_1) + (1-h)f(x_2) \quad \forall x_1, x_2 \in [a, b] \quad \forall 0 \leq h \leq 1\,.$$

Eine Funktion $f : A \to B$ heißt *konkav* in einem Intervall $[a, b]$, wenn die Funktion dort immer über der Sekante liegt, d. h.,

$$f(hx_1 + (1-h)x_2) \geq hf(x_1) + (1-h)f(x_2) \quad \forall x_1, x_2 \in [a, b] \quad \forall 0 \leq h \leq 1\,.$$

Wenn wir den Graphen der Funktion $f(x) = x^3 - x$ betrachten, sehen wir, dass $f(x)$ auf $(-\infty, 0]$ konkav ist, auf $[0, \infty)$ konvex. Weiters bemerken wir, dass der Anstieg der Tangente an den Graphen im konkaven Bereich immer kleiner wird, je größer x wird, im konvexen Bereich immer größer. Das gilt auch allgemein.

Somit haben wir:

THEOREM 6.3

$$\begin{aligned} f \text{ konkav} \quad &\Leftrightarrow \quad f''(x) \leq 0 \quad \forall x \\ f \text{ konvex} \quad &\Leftrightarrow \quad f''(x) \geq 0 \quad \forall x \end{aligned}$$

Konvexe Funktionen spielen in der Optimierung eine große Rolle, da von diesen das globale Minimum „leicht" gefunden werden kann.

Zur Berechnung der (lokalen) Minima einer Funktion f, die mindestens zweimal differenzierbar ist, wählen wir also folgenden Algorithmus:

1. Berechne $f'(x)$ und $f''(x)$.
2. Finde alle x_i, an denen $f'(x_i) = 0$. Das sind die sogenannten *stationären* Punkte, d. h., ein Punkt heißt stationär, wenn dort die Ableitung 0 ist.
3. Überprüfe $f''(x_i)$ für alle gefundenen Punkte:
   (a) $f''(x_i) < 0 \quad \Rightarrow \quad x_i$ ist lokales Maximum.
   (b) $f''(x_i) > 0 \quad \Rightarrow \quad x_i$ ist lokales Minimum.
   (c) $f''(x_i) = 0 \quad \Rightarrow$ keine Aussage möglich.

BEISPIEL 6.8
Die Extrema der Funktion $f(x) = x^3 - x$ finden wir demnach so:

1. $f'(x) = 3x^2 - 1$, $f''(x) = 6x$.
2. $f'(x) = 0$ also $3x^2 - 1 = 0$ oder $x = \pm\frac{1}{\sqrt{3}}$.
3. $f''(\frac{1}{\sqrt{3}}) = 2\sqrt{3} > 0$, d. h., lokales Minimum in $\frac{1}{\sqrt{3}}$;
   $f''(-\frac{1}{\sqrt{3}}) = -2\sqrt{3} < 0$, d. h., lokales Maximum in $-\frac{1}{\sqrt{3}}$.

BEMERKUNG 6.4
Warum können wir nichts über den stationären Punkt x_0 aussagen, wenn wir wissen, dass $f''(x) = 0$.
Dazu einige Beispiele:

- $f(x) = x^3$ hat in $x_0 = 0$ einen Wendepunkt mit horizontaler Wendetangente, d. h., $f'(0) = f''(0) = 0$.

- $f(x) = x^4$ hat in $x_0 = 0$ ein Minimum und es gilt dort: $f'(0) = f''(0) = f'''(0) = 0$.
- $f(x) = x^5$ hat in $x_0 = 0$ einen Wendepunkt mit horizontaler Wendetangente und es gilt sogar $f'(0) = f''(0) = f^{(3)}(0) = f^{(4)}(0) = 0$.

Die jeweils nächsten Ableitungen obiger Beispiele sind aber ungleich Null und geben Auskunft über die Art des stationären Punktes.

Insgesamt gilt:

THEOREM 6.4
Sei $f : A \to B$ eine (mindestens) n-mal differenzierbare Funktion und x_0 ein stationärer Punkt (d. h. $f'(x_0) = 0$). Sei $f^{(n)}(x)$ die erste in x_0 nicht verschwindende Ableitung, d. h., $f^{(n)}(x_0) \neq 0$, so gilt:

- Ist n ungerade, so ist x_0 ein Wendepunkt bzw. Sattelpunkt.
- Ist n gerade, so ist x_0
  - ein Minimum, falls $f^{(n)}(x_0) > 0$,
  - ein Maximum, falls $f^{(n)}(x_0) < 0$.

Hier haben wir schon implizit höhere Ableitungen betrachtet. Wir notieren die n-te Ableitung einer Funktion f mit

$$f^{(n)}(x) = \frac{d^n f(x_0)}{dx^n} = \left.\frac{d^n f}{dx^n}\right|_{x_0},$$

formal definieren wir

$$f^{(0)} = f \qquad f^{(n+1)} = (f^{(n)})' \qquad n \geq 0,$$

und sagen, eine Funktion f heißt in x_0 n-mal differenzierbar, wenn $f, f', \ldots, f^{(n-1)}$ in einer ganzen Umgebung von x_0 existieren und $f^{(n)}$ wenigstens im Punkt x_0.

Ist eine Funktion auf $\mathbb{R}$ sogar n-mal stetig differenzierbar, gehört sie zur Klasse $C^n(\mathbb{R})$; es gibt „sogar“ $C^\infty(\mathbb{R})$ Funktionen (etwa $\sin x$).

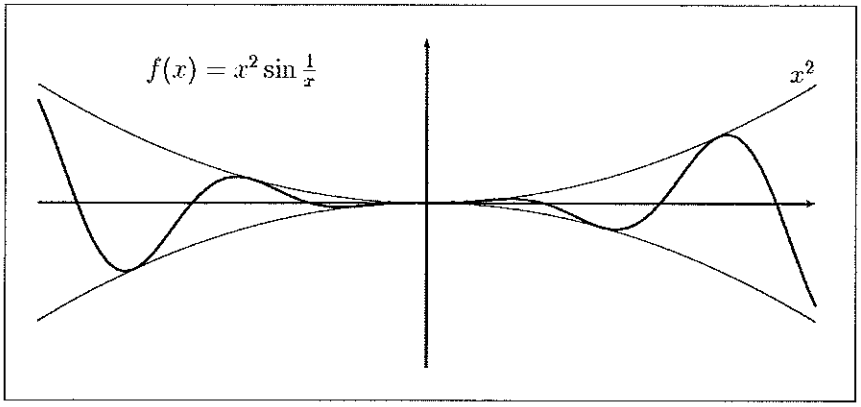
BEISPIEL 6.9
Funktionen müssen nicht beliebig oft differenzierbar sein.
Die Funktion

$$f(x) = \begin{cases} x^2 \sin\frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

ist in 0 nicht stetig differenzierbar, da für die Ableitung $f'(x) = 2x \sin\frac{1}{x} - \cos\frac{1}{x}$ gilt. Diese ist im 0-Punkt nicht stetig, da $\lim_{x\to 0} f'(x)$ nicht existiert ($\cos\frac{1}{x}$ ist dort nicht definiert), jedoch ist $f(x)$ in 0 differenzierbar, denn

$$f'(0) = \lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x\to 0} \frac{x^2 \sin\frac{1}{x}}{x} = \lim_{x\to 0} (x \sin\frac{1}{x}) = 0.$$

Damit ist f' in 0 nicht stetig und kann dort auch nicht noch einmal differenziert werden.



Wir können nun ein Theorem über Extrema auf Intervallen formulieren.

THEOREM 6.5
Ist die Funktion f stetig auf einem abgeschlossenen (kompakten) Intervall $[a, b]$ und differenzierbar auf dem offenen Intervall (a, b), sind weiters $x_1, \ldots, x_r$ die stationären Punkte von f in (a, b), so gilt für das Maximum von f auf $[a, b]$

$$\max\{f(x) \mid x \in [a, b]\} = \max\{f(a), f(b), x_1, \ldots, x_r\}$$

und für das Minimum von f auf $[a, b]$

$$\min\{f(x) \mid x \in [a, b]\} = \min\{f(a), f(b), x_1, \ldots, x_r\}\,.$$

Insbesondere gilt:

THEOREM 6.6 (SATZ VON ROLLE)
Ist die Funktion f stetig auf einem abgeschlossenen Intervall $[a, b]$ und differenzierbar auf dem offenen Intervall (a, b), und ist $f(a) = f(b)$, so gibt es einen Punkt $\xi \in (a, b)$ mit $f'(\xi) = 0$.

D. h., genügt eine Funktion den Voraussetzungen des Satzes von Rolle, so existiert ein Extremum im Inneren des Intervalls.

Allgemeiner gilt:

THEOREM 6.7 (MITTELWERTSATZ DER DIFFERENTIALRECHUNG)
Ist die Funktion f stetig auf einem abgeschlossenen Intervall $[a, b]$ und differenzierbar auf dem offenen Intervall (a, b), so gibt es einen Punkt $\xi \in (a, b)$ mit

$$f(b) - f(a) = f'(\xi)(b - a)\,.$$

Bei nicht differenzierbaren Funktionen sind die Extrema nicht so einfach zu finden.

6.2.1 Kurvendiskussionen

Aus der Schule sollte noch der Begriff der *Kurvendiskussion* geläufig sein. Wir verstehen darunter eine möglichst vollständige Beschreibung der Funktion (ihrer Eigenschaften) auf der Definitions- und Wertemenge. Dadurch können wir „interessante" Bereiche und „Problemstellen" der Funktion finden – je nach Aufgabenstellung.

Folgende Punkte gehören zu einer Kurvendiskussion, damit wir uns einen Überblick über den Graphen, der heutzutage selbstverständlich mit einem Computerprogramm gezeichnet wird, verschaffen.

- Definitionsmenge, Wertemenge
- Unstetigkeitsstellen, Stetigkeitsstellen und -bereiche
- Verhalten in Randpunkten und obigen Stellen (Asymptoten)
- Verhalten im Unendlichen (Asymptoten)
- Nullstellen und zugehörige Steigung der Tangente
- Kritische Punkte ($f'(x) = 0$) und Tangentensteigung
- Extrema
- Wendepunkte und Wendetangenten
- Monotoniebereiche
- Konvexe bzw. konkave Bereiche

Beim computerunterstützten Zeichnen des Graphen können allerdings Probleme auftreten, die durch die Kurvendiskussion „vorhergesehen"und daher extra beachtet werden können.

- Computer liefern nur diskrete Werte, könnten also Pole (Unendlichkeitsstellen) und Sprungstellen übersehen
- Computer berechnen nur endlich viele Funktionswerte
- Computer können nur bestimmte Zahlen darstellen.
- Die Skalierung erfolgt vielfach automatisch, sodass die Skalen auf den Achsen oft übersehen werden und wir dadurch ein „falsches" Bild von der betrachteten Funktion bekommen.

Reelle Funktionen – Eine Zusammenfassung

Wir betrachten nun einige grundlegende Eigenschaften von Funktionen bzw. deren Definitionen. Funktionen liefern (Funktions-) Werte für Elemente einer Definitionsmenge A. Diese Werte liegen in einer Werte- oder Bildmenge (*image*) B. Dabei muss gelten:

$$f : A \to B \text{ Funktion} \quad \Leftrightarrow \quad \forall a \in A \;\; \exists! b \in B : \; f(a) = b \,.$$

Um uns ein Bild von einer Funktion zu machen, betrachten wir oft den Graphen G einer Funktion

$$G_f = \{(a, f(a)) \,|\, a \in A, \; f(a) \in B\} \subseteq (A \times B) \,.$$

Obwohl wir uns immer wieder reelle Funktionen vorstellen, d. h., Funktionen $f : \mathbb{R} \to \mathbb{R}$, müssen wir doch auch bedenken, dass sowohl obige Definitionen, als auch die Eigenschaften von Funktionen auch mit rein abstrakten Objekten funktionieren, solange sie in Mengen zusammenfassbar sind

Eigenschaften von Funktionen $f : A \to B$:

- Injektivität: $f(x) = f(y) \;\Leftrightarrow\; x = y, \; \forall x, y \in A.$
- Surjektivität: $\forall b \in B \; \exists a \in A : f(a) = b.$
- Bijektivität: gleichzeitig surjektiv und injektiv.
  Bijektive Funktionen können invertiert (umgekehrt) werden:

  $$f^{-1} \circ f = f \circ f^{-1} = id \,.$$

  Eigentlich gilt nur:

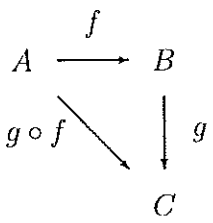
  $$(f^{-1} \circ f)(a) = f^{-1}(f(a)) = f^{-1}(b) = a = id_a(a)$$

  und

  $$(f \circ f^{-1})(b) = f(f^{-1}(b)) = f(a) = b = id_b(b) \,.$$
- Monoton wachsend: $\forall x < y : \; f(x) \leq f(y).$
- Streng monoton wachsend: $\forall x < y : \; f(x) < f(y).$
- Monoton fallend: $\forall x < y : \; f(x) \geq f(y).$
- Streng monoton fallend: $\forall x < y : \; f(x) > f(y).$
- Gerade Funktion: $f(x) = f(-x)$ (symmetrisch bezüglich y-Achse).
- Ungerade Funktion: $f(x) = -f(-x)$ (punktsymmetrisch bezüglich des Ursprungs).
- Konvexe Funktion (strikt): $f''(x) > 0$ (Sekante immer über dem Funktionsgraphen).
- Konkave Funktion (strikt): $f''(x) < 0$ (Sekante immer unter dem Graphen).

Funktionen können verknüpft werden:

$$f : A \to B,\ g : B \to C \quad \Rightarrow \quad g \circ f : A \to C, (g \circ f)(a) = g(f(a)) \in C\,.$$



BEISPIEL 6.10
$f(x) = 2x$ und $g(x) = e^x$, dann $(f \circ g)(x) = 2e^x$ und $(g \circ f)(x) = e^{2x}$.

Stetigkeit:

Funktionen heißen stetig (in x_0), wenn gilt

$$\lim_{x \to x_0} f(x) = f(x_0) \qquad \Leftrightarrow \qquad \forall \varepsilon > 0 \ \ \exists \delta > 0 : \ |f(x) - f(x_0)| < \varepsilon \quad \forall |x - x_0| < \delta\,.$$

Wir unterscheiden auch zwischen linksseitig stetig, rechtsseitig stetig und stetig per se.

Differenzierbarkeit:

Um weitere Eigenschaften bzw. Charakteristika von Funktionen zu untersuchen, etwa Änderungsraten, Extremwerte, stationäre Punkte etc. benötigen wir die Begriffe der Ableitung bzw. des Differentialquotienten ($\to$ differenzierbare Funktionen).

$$f'(x_0) = \frac{df(x_0)}{dx} = \left.\frac{df}{dx}\right|_{x_0} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \lim_{x \to x_0} \frac{f(x) - x}{x - x_0}$$

Ebenso wichtige Begriffe sind konkav und konvex, die das Krümmungsverhalten einer Funktion (lokal oder global) beschreiben.

Integrierbarkeit:

Weiters können wir auch Flächen unter und zwischen Funktionen berechnen ($\to$ integrierbare Funktionen).

Eine Funktion heißt Riemann-integrierbar, wenn die Riemannschen Untersummen $\underline{S}_r(f) = \sum_k \underline{f}_{k,r}\, 2^{-r}$ und Obersummen $\overline{S}_r(f) = \sum_k \overline{f}_{k,r}\, 2^{-r}$ einer Funktion gegen denselben Grenzwert konvergieren. Dieser heißt dann das Integral der Funktion.

$$\lim_{r \to \infty} \underline{S}_r(f) = \lim_{r \to \infty} \overline{S}_r(f) = \int f(x)\, dx$$

7 Mehrdimensionale Funktionen

Eine mehrdimensionale Funktion

$$f : \mathbb{R}^n \to \mathbb{R}^m \quad f(x) = f(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = y$$

hat als Argument den Vektor $x = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ und nimmt einen Funktionswert $y = (y_1, \ldots, y_m)^t \in \mathbb{R}^m$ an.
Die Funktion besteht (koordinatenweise) aus m reellen Funktionen

$$f_j : \mathbb{R}^n \to \mathbb{R}, \quad f_j(x_1, \ldots, x_n) = y_j \in \mathbb{R}, \; j = 1, \ldots, m,$$

den sogenannten Komponentenfunktionen.

Bemerkung: Obwohl wir es mit mehrdimensionalen Argumenten x zu tun haben, werden wir nicht, wie oft üblich $\boldsymbol{x}$ schreiben, sondern, da es einerseits aus dem Kontext klar ist und andererseits die Aussagen auch für eindimensionale Argumente gelten, wie bei Funktionen einfach x.

Beispiele: lineare Abbildungen
meteorologische Daten
Landkarten
Kurven, etwa Zykloiden
Produktion: n Input-, m Output-Variablen

Wie bei den reellen Funktionen $f : \mathbb{R} \to \mathbb{R}$ wollen wir diese Funktionen auf ihre Eigenschaften untersuchen.

Ziel: (A) Stetigkeit
(B) Differenzierbarkeit
(C) Integrierbarkeit

Wir werden von den Eigenschaften reeller Funktionen ausgehen und in Analogie dazu die Eigenschaften für Komponentenfunktionen (mehrdimensionale reelle Funktionen, d. h., das Ergebnis ist in $\mathbb{R}$) betrachten.

7.0.2 Niveaulinien

Bei mehrdimensionalen Funktionen gibt es das Problem der Darstellbarkeit. Der *Graph* einer Funktion $f : \mathbb{R}^n \to \mathbb{R}^m$

$$G_f = \{(x, f(x)) \,|\, x \in \mathbb{R}^n, \, f(x) \in \mathbb{R}^m\} \subseteq \mathbb{R}^n \times \mathbb{R}^m$$

ist $n + m$ dimensional, d. h., bestenfalls noch im $\mathbb{R}^3$ zu zeichnen, d. h. für Funktionen $f : \mathbb{R}^2 \to \mathbb{R}$ oder $f : \mathbb{R} \to \mathbb{R}^2$. Wenn wir eine Landkarte betrachten, so gibt es dort zwei Hilfsmittel für mehrdimensionale Daten: Einerseits die Höhenschichtlinien (Isohypsen), die die dritte Dimension andeuten, und andererseits Farben, die indirekt im Zusammenhang mit diesen Höhenschichtlinien stehen, in Wirklichkeit aber meist die Vegetationszonen wiederspiegeln sollen (mehrdimensionale Information). Wir kennen auch Isobaren

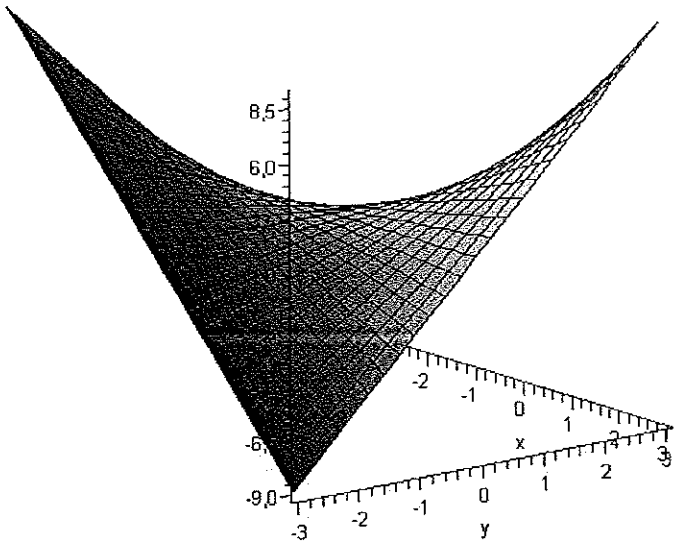


Abbildung 1: Die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x \cdot y$.

(Luftdruck) und Isothermen (Temperatur) aus dem täglichen Wetterbericht, oder aber Äquipotentiallinien aus der Physik.

Wir definieren eine sogenannte *Niveaumenge* für $f : \mathbb{R}^n \to \mathbb{R}^m$ als

$$N_c(f) = \{x \in \mathbb{R}^n \mid f(x) = c \in \mathbb{R}^m\}.$$

Höhenschichtlinien sind solche Niveaumengen. Ein Graph einer Funktion, der mit Hilfe dieser Niveaumengen bzw. Niveaulinien gezeichnet wird, heißt auch Konturplot.

BEISPIEL 7.1
Die Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x_1, x_2) = x_1 x_2$$

hat einen dreidimensionalen Graphen (siehe Abbildung 1).

Für die Niveaulinien (Konturlinien) gilt:

$$f(x_1, x_2) = x_1 x_2 = c \quad \Rightarrow \quad x_1 \neq 0 : \; x_2 = \frac{c}{x_1} \quad \text{(Hyperbeln)}$$

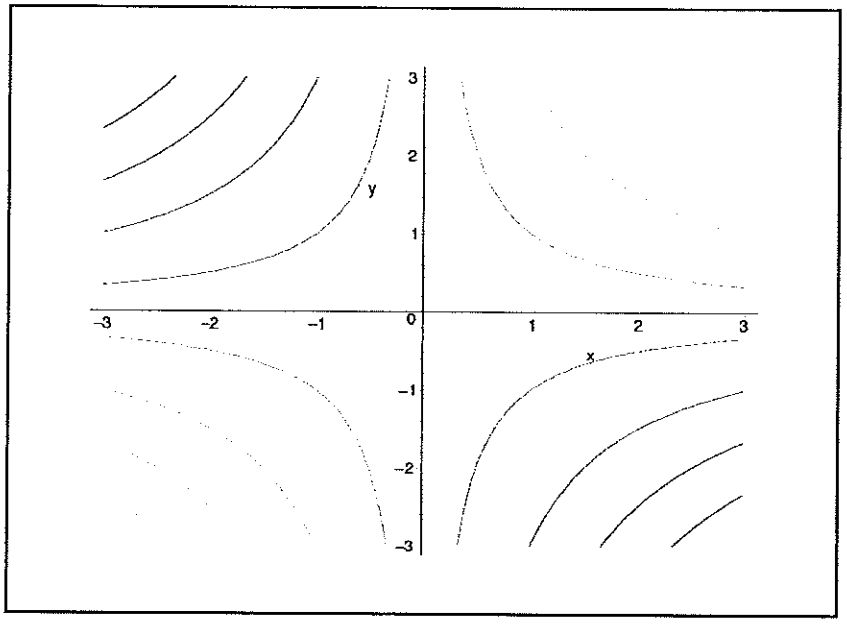


Abbildung 2: Konturplot der Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x \cdot y$.

$$c = 0 \quad \Rightarrow \quad x_1 = 0 \text{ oder } x_2 = 0\,.$$

Es sind also Hyperbeln bzw. zwei normale Geraden für $c = 0$ (siehe Abbildung 2). Farben können zusätzlich die Elevation (Höhe über 0) kodieren.

BEMERKUNG 7.1
Obiges Beispiel wäre vielleicht in folgender Notation „geläufiger“:

$f(x, y) = x \cdot y$.
Niveaulinien: $f(x, y) = xy = c$, daher $y = \frac{c}{x}$, $x \neq 0$ und $c = 0$, wenn $x \cdot y = 0$.

Diese Notation ist aber nicht leicht auf n Dimensionen erweiterbar, insbesondere wird y auch häufig für das Ergebnis einer Funktion verwendet, x als Argument, also $f(x) = y$.

Weiters müssen wir uns bei der Darstellung von mehrdimensionalen Funktionen von althergebrachten Vorstellungen lösen.

Die Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \quad f(x_1, x_2) = \big(\cos(x_1) - 2\cos(0.4\,x_2), \sin(x_1) - 2\sin(0.4\,x_2) \big)$$

hat etwa einen „Graphen“ wie in Abbildung 3.

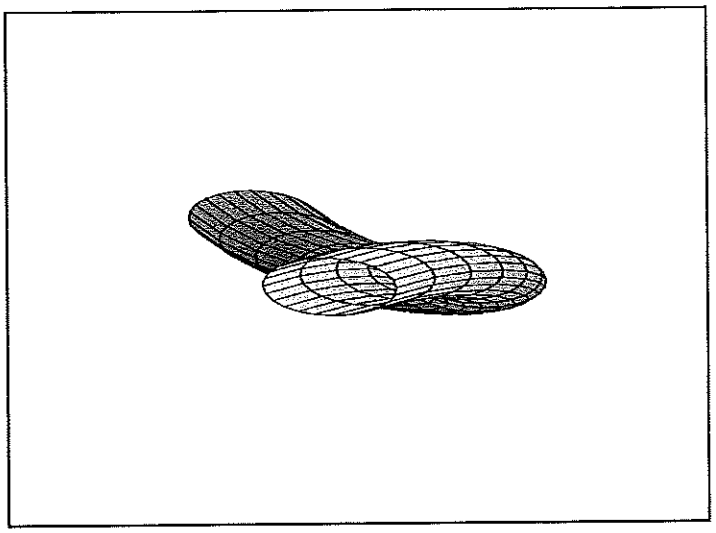


Abbildung 3: Zweidimensionale Projektion eines vierdimensionalen Konturplots einer Funktion $f : \mathbb{R}^2 \to \mathbb{R}^2$; die vierte Dimension ist durch Grauwerte kodiert.

7.1 Stetigkeit

Wenn wir die Definition der Stetigkeit betrachten

> $f : \mathbb{R} \to \mathbb{R}$ ist stetig in x_0, wenn
>
> $$\lim_{x \to x_0} f(x) = f(x_0) \quad \Leftrightarrow \quad \forall \varepsilon > 0 \;\; \exists \delta > 0 : \;\; |f(x) - f(x_0)| < \varepsilon \;\; \forall |x - x_0| < \delta \, .$$
>
> f ist stetig auf $\mathbb{R}$, wenn f in allen Punkten $x_0 \in \mathbb{R}$ stetig ist,

so bemerken wir, dass wir Stetigkeit über den Begriff des Grenzwertes bzw. über einen Abstands- bzw. Längenbegriff einführten. Im mehrdimensionalen Fall wird dies genauso geschehen. Dazu betrachten wir aber zuerst die Struktur des $\mathbb{R}^n$.

7.1.1 Eigenschaften des $\mathbb{R}^n$

$\mathbb{R}$ ... Körper der reellen Zahlen
$\mathbb{R}^n$ ... euklidischer Vektorraum

Der $\mathbb{R}^n$ erbt die Rechenregeln von $\mathbb{R}$:

Vektoraddition $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}$

Skalarmultiplikation $\lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}, \lambda \in \mathbb{R}$

Skalarprodukt $\langle x, y \rangle = x^t y = x \cdot y = x_1 y_1 + \cdots + x_n y_n$ (inneres Produkt)

Länge (Norm) $\|x\| = \sqrt{x^t x} = \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + \cdots + x_n^2}$

DEFINITION 7.1
Eine Abbildung $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ heißt *Norm* genau dann, wenn folgende drei Bedingungen erfüllt sind:

(1) $\|x\| = 0 \Leftrightarrow x = 0$, $x \in \mathbb{R}^n$, $0 \in \mathbb{R}^n$.

(2) $\|\lambda x\| = |\lambda| \, \|x\|$, $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$.

(3) $\|x + y\| \leq \|x\| + \|y\|$, $x, y \in \mathbb{R}^n$ (Dreiecksungleichung)

Aus der Dreiecksungleichung lassen sich auch folgende Ungleichungen ableiten:

$$|\|x\| - \|y\|| \leq \|x \pm y\|$$

$$\langle x, y \rangle \leq \|x\| \, \|y\| \quad \text{(Schwarz'sche Ungleichung)}$$

Wir sehen, dass der Norm-Begriff im $\mathbb{R}^n$ mit dem (im $\mathbb{R}^3$) üblichen Abstands- oder Längenbegriff übereinstimmt.

Der $\mathbb{R}^n$ wird damit zu einem sogenannten *metrischen Raum*, wobei als Metrik (Messfunktion) die obige Norm genommen wird.

Um langsam zum Begriff des Grenzwertes zu kommen, betrachten wir nun (konvergente) Folgen:

DEFINITION 7.2
Eine Folge $(a_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}^n$ heißt *konvergent* gegen ihren Grenzwert oder Limes $a \in \mathbb{R}^n$ ($\lim_{n \to \infty} a_n = a$), wenn gilt:

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} : \; \|a_n - a\| < \varepsilon \quad \forall n \geq N \, .$$

Wir sehen, dass der Betrag in der Definition von Konvergenz in $\mathbb{R}$ nur durch die Norm, d. h., eigentlich durch den Abstand, ersetzt wurde – wir könnten auch Betragsstriche statt der Normstriche machen.

Wir sprechen davon, dass ab einem gewissen N alle Folgenglieder in einer sogenannten ε-Umgebung von a liegen.

DEFINITION 7.3
ε-Umgebung bzw. ε-Kugel

$$U_\varepsilon(x_0) = B_\varepsilon(x_0) = \{x \in \mathbb{R}^n \mid \|x - x_0\| < \varepsilon\}$$

(Die Bezeichnung B kommt vom englischen Wort *ball* für Kugel.)

Nun sind wir in der Lage folgende Definition zu geben:

DEFINITION 7.4
Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}^m$ heißt stetig in $x_0 \in \mathbb{R}^n$, wenn gilt:

$$\lim_{x \to x_0} f(x) = f(x_0)\,,$$

d. h.

$$\forall \varepsilon > 0\ \exists \delta > 0 :\ \|f(x) - f(x_0)\| < \varepsilon \quad \forall \|x - x_0\| < \delta\,.$$

f heißt stetig auf $A \subseteq \mathbb{R}^n$, wenn f stetig in jedem $x_0 \in A$ ist, insbesondere auch für $A = \mathbb{R}^n$.

Bemerkung: Was heißt hier Limes oder Grenzwert x_0? – Da wir uns im n-dimensionalen Raum befinden, können wir uns von unendlich vielen Richtungen auf unendlich viele Arten dem x_0 nähern. Wenn wir vorher von Folgen sprachen, so können wir die Definition der Stetigkeit auch mit Hilfe von Folgen in äquivalenter Weise definieren.
D. h., f ist stetig in x_0, wenn

$$\lim_{k \to \infty} f(x_k) = f(x_0) \quad \forall (x_k) \subseteq \mathbb{R}^n \text{ mit } \lim_{k \to \infty} x_k = x_0\,,$$

d. h., dies gilt für alle denkbaren Folgen (x_k) mit Grenzwert x_0.

THEOREM 7.1
Eine Funktion $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ ist stetig auf A, wenn jede ihrer Komponentenfunktionen auf A stetig ist.

Bevor wir den nächsten Schritt in Richtung Differenzierbarkeit tun, müssen wir uns noch kurz mit der Topologie des $\mathbb{R}^n$ befassen, d. h., mit Mengen im $\mathbb{R}^n$.

7.1.2 Mengen im $\mathbb{R}^n$

Wir kennen bereits die Menge der ε-Umgebung $U_\varepsilon(a)$.

Wir definieren nun einige Begriffe aus der Topologie des $\mathbb{R}^n$ in Analogie zu den offenen, halboffenen und abgeschlossenen Intervallen in $\mathbb{R}$. Diese Begriffe bzw. Mengen sind beim Differenzieren (und damit bei der Optimierung, d. h., beim Finden von Extremwerten) von essentieller Bedeutung.

DEFINITION 7.5
Offene ε-Umgebung: $U_\varepsilon(a) = \{x \in \mathbb{R}^n \mid \|x - a\| < \varepsilon\}$

Abgeschlossene ε-Umgebung: $\bar{U}_\varepsilon(a) = \{x \in \mathbb{R}^n \mid \|x - a\| \leq \varepsilon\}$

Punktierte ε-Umgebung: $\dot{U}_\varepsilon(a) = U_\varepsilon(a) \setminus \{a\}$

Umgebung: Eine Menge M heißt Umgebung von $a \in \mathbb{R}^n$, wenn sie eine offene ε-Umgebung von a enthält, d. h., $U_\varepsilon(a) \subseteq M$.

Innerer Punkt: a heißt *innerer Punkt* von M, wenn eine Umgebung von a vollständig in M liegt, d. h., $\exists \varepsilon > 0 : U_\varepsilon(a) \subseteq M$.

Offene Menge: Eine Menge heißt *offen*, wenn sie nur aus inneren Punkten besteht.

Inneres: Das Innere M° einer Menge M ist die Menge ihrer inneren Punkte.

Isolierter Punkt: a heißt isolierter Punkt von M, wenn es mindestens eine punktierte Umgebung von a gibt, die nicht in M liegt, d. h., $\exists \dot{U}_\varepsilon(a) : \dot{U}_\varepsilon(a) \cap M = \phi$.

Randpunkt: a heißt *Randpunkt* der Menge M, falls in jeder Umgebung von a mindestens ein Punkt liegt, der in M liegt und mindestens ein Punkt, der nicht in M liegt, d. h.: $(U_\varepsilon(a) \cap M \neq \phi) \wedge (U_\varepsilon(a) \cap M^c \neq \phi)$, $\forall U_\varepsilon(a)$.

Rand ∂M: Der Rand ∂M einer Menge M ist die Menge all ihrer Randpunkte.

Abgeschlossene Menge: Eine Menge heißt abgeschlossen, wenn sie ihren Rand enthält.

Abgeschlossene Hülle: Die abgeschlossenen Hülle $\bar{M}$ ist die Menge vereinigt mit ihrem Rand, d. h., $\bar{M} = M \cup \partial M$.

Beschränkt: Eine Menge M heißt beschränkt, wenn es ein $R > 0$ gibt, sodass $\|x\| < R$ für alle $x \in M$, oder anders ausgedrückt, $M \subseteq U_R(0)$.

Kompakt: Eine Menge M heißt kompakt, wenn sie beschränkt und abgeschlossen ist.

Wenn wir Folgen in Mengen betrachten, werden uns die obigen Begriffe vielleicht etwas verständlicher. Wenn wir gleichzeitig bedenken, dass etwa in der Optimierung ein Computer durch einen iterativen Algorithmus ein Optimum finden soll, und die (iterativen) Näherungslösungen nichts anderes sind als Glieder von Folgen, dann sollte klar sein, warum wir diese Begriffe benötigen.

THEOREM 7.2
Eine Menge M ist abgeschlossen, wenn mit jeder konvergenten Folge $(a_n) \subseteq M$ auch ihr Grenzwert in M liegt.

THEOREM 7.3
M abgeschlossen $\Leftrightarrow$ $M^c = \mathbb{R}^n \setminus M$ offen.

Für den folgenden Merksatz benötigen wir noch einen Begriff.

DEFINITION 7.6
Eine Folge (a_k) heißt *Teilfolge* von (a_n), wenn sie durch Weglassen von Folgengliedern von (a_n) entsteht.

THEOREM 7.4
Eine Menge M ist genau dann kompakt, wenn jede Folge in M eine in M konvergente Teilfolge besitzt.

THEOREM 7.5
Konvergente Folgen sind beschränkt.

7.2 Mehrdimensionale Differenzialrechnung

Da eine Funktion $f : \mathbb{R}^n \to \mathbb{R}^m$ komponentenweise aus reellen Funktionen $f_i : \mathbb{R}^n \to \mathbb{R}$ besteht, betrachten wir zuerst Komponentenfunktionen.

Erinnern wir uns daran, dass für eindimensionale reelle Funktionen

$$f : \mathbb{R} \to \mathbb{R} \qquad f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \frac{d}{dx} f = \frac{df}{dx},$$

und dass die Ableitung dem Anstieg der Tangente im jeweilig betrachteten Punkt entspricht! Wie sieht das im mehrdimensionalen Fall aus?

BEISPIEL 7.2
Sei $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 x_2$, so kann dies im $\mathbb{R}^3$ dargestellt werden. Die Tangentialebene an einen Punkt dieser Abbildung ist doch durch zwei linear unabhängige Richtungsvektoren gegeben (plus dem Ankerpunkt $(x_1, x_2, x_1 \cdot x_2) \in \mathbb{R}^3$). Bewegen wir uns entlang einer Gitternetzline (siehe Abbildung 1) etwa parallel zur x_1-Achse, so erhalten wir einen Richtungsvektor über die Tangente an die Gitternetzlinie. Wir haben also x_2 konstant gelassen und betrachten die Ableitung nach x_1. Einen zweiten, unabhängigen Richtungsvektor erhalten wir analog, wenn wir x_1 konstant lassen und uns parallel zur x_2-Achse bewegen.

Dies geht allgemein so, erinnern wir uns nur, dass es an Randpunkten eines Intervalls unmöglich war, eine Tangente zu legen, wenn dieser Randpunkt zur Funktion gehörte!

DEFINITION 7.7
Die partielle Ableitung für reelle Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ ist definiert als

$$\frac{\partial}{\partial x_j} f(x_1, \ldots, x_n) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_{j-1}, x_j + h, x_{j+1}, \ldots, x_n) - f(x_1, \ldots, x_j, \ldots, x_n)}{h}$$

d. h., Ableitung in Richtung x_j, also x_j variabel, alle anderen Argumente werden als konstant angesehen.

Eine Funktion heißt *partiell differenzierbar* nach x_j im Punkt $x_0 \in \mathbb{R}^n$ ($x_0 = (x_1^{(0)}, \ldots, x_n^{(0)})$), wenn dieser Limes in x_0 existiert. Sie heißt partiell differenzierbar nach x_j auf der Menge $A \subseteq \mathbb{R}^n$, wenn dieser Limes in allen Punkten $x_0 \in A$ existiert.
Andere Bezeichnungen:

$$D_j f(x) = f_{,j}(x) = \frac{\partial}{\partial x_j} f(x) = \frac{\partial f}{\partial x_j}\Big|_x .$$

BEISPIEL 7.3

$$f(x,y) = x \cdot y, \quad \frac{\partial f}{\partial x} = f_{\cdot x}(x,y) = y, \quad \frac{\partial f}{\partial y} = f_{\cdot y}(x,y) = x,.$$

Da wir eine Funktion nach allen Variablen x_i partiell ableiten können, und wir an diesen unabhängigen Koordinatenrichtungen interessiert sind, ist die Ableitung wie folgt definiert.

DEFINITION 7.8
Wenn für eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ alle partiellen Ableitungen existieren, so ist die erste Ableitung von f, der sogenannte *Gradient* $\nabla f(x)$, definiert als

$$\nabla f(x) = \operatorname{grad} f(x) = (\frac{\partial}{\partial x_1} f(x_1, \ldots, x_n), \ldots, \frac{\partial}{\partial x_n} f(x_1, \ldots, x_n)) .$$

Bemerkung: Wir verwenden hier die Notation des Gradienten als Zeilenvektor, da er dann direkt in Approximationsformeln etc. verwendet werden kann. manchmal wird der Gradient auch als Spaltenvektor definiert, insbesondere wenn durch ihn eine Richtung angezeigt werden soll. Beides ist möglich und zulässig.

BEMERKUNG 7.2
Der im Gradienten verwendete Nabla-Operator ∇ ist definiert als $\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{pmatrix}$.

Wie im eindimensionalen Fall wird wieder gelten, dass in einem *stationären* Punkt (Minimum, Maximum, bzw. Wendepunkt, der nun Sattelpunkt heißen wird) die erste Ableitung, d. h. der Gradient, gleich null sein muss (horizontale „Tangentialebene", als Bild im $\mathbb{R}^3$). Wie dort, können wir erst mit Hilfe der zweiten Ableitung entscheiden, um welche Art von stationärem Punkt es sich handelt. Der Begriff stationär wird uns später noch klarer werden. Da wir nun viele partielle Ableitungen im Gradienten haben, müssen wir diese wiederum nach allen Richtungen ein zweites Mal (partiell) ableiten:

$$\frac{\partial}{\partial x_j}(\frac{\partial}{\partial x_k} f(x_1, \ldots, x_n)) = \frac{\partial^2}{\partial x_k \, \partial x_j} f(x_1, \ldots, x_n) = f_{\cdot jk} = f_{\cdot kj}$$

BEISPIEL 7.4

$$f(x,y) = x \cdot y, \quad \frac{\partial^2 f}{\partial x \partial y} = 1, \quad \frac{\partial^2 f}{\partial y \partial x} = 1, \quad \frac{\partial^2 f}{\partial x^2} = 0, \quad \frac{\partial^2 f}{\partial y^2} = 0.$$

DEFINITION 7.9
Ist $f : \mathbb{R}^n \to \mathbb{R}$ jeweils (mindestens) zweimal partiell differenzierbar nach allen Koordinatenrichtungen, so ist die zweite Ableitung von f die Hessematrix

$$\nabla^2 f(x) = H_f(x) = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x_1, \ldots, x_n) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} f(x_1, \ldots, x_n) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} f(x_1, \ldots, x_n) & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} f(x_1, \ldots, x_n) \end{pmatrix}$$

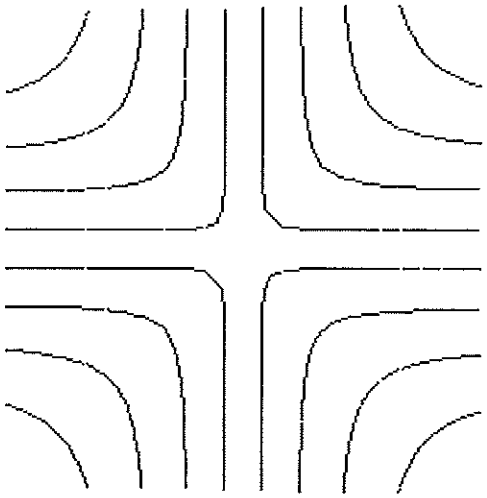


Abbildung 4: Die Funktion $f(x_1, x_2) = \dfrac{x_1 x_2}{\sqrt{x_1^2 + x_2^2}}$.

Die Hessematrix ist eine symmetrische $n \times n$-Matrix, da gilt:
Ist $f : \mathbb{R}^n \to \mathbb{R}$ zweimal partiell differenzierbar, so gilt

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x_1, \ldots, x_n) = \frac{\partial^2}{\partial x_j \partial x_i} f(x_1, \ldots, x_n) .$$

BEMERKUNG 7.3
Der $\mathbb{R}^n$ oben kann durch eine beliebige offene Menge ersetzt werden, offen deshalb, weil die Differentiation am Rand Probleme bereitet, da dort keine eindeutigen Tangenten bzw. Tangentialebenen angelegt werden können (bzw. unendlich viele).

DEFINITION 7.10
Eine Funktion $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ heißt in einem inneren Punkt $x_0 \in A$ (total) *differenzierbar*, wenn sie in x_0 partiell differenzierbar ist, alle partiellen Ableitungen in einer Umgebung von x_0 existieren und diese Ableitungen dort stetig sind.

BEISPIEL 7.5
$f(x_1, x_2, x_3) = x_1 x_2^2 + 2x_2 \exp(x_3) + \exp(x_3^2)$

$$\begin{aligned} \nabla f(x) &= (x_2^2, 2x_1 x_2 + 2\exp(x_3), 2x_2 \exp(x_3) + 2x_3 \exp(x_3^2)) \\ \nabla^2 f(x) &= \begin{pmatrix} 0 & 2x_2 & 0 \\ 2x_2 & 2x_1 & 2\exp(x_3) \\ 0 & 2\exp(x_3) & 2x_2 \exp(x_3) + 2\exp(x_3^2) + 4x_3^2 \exp(x_3^2) \end{pmatrix} \end{aligned}$$

Es kann aber auch Probleme geben:

BEISPIEL 7.6
Sei

$$f(x_1,x_2)=\begin{cases}\dfrac{x_1x_2}{\sqrt{x_1^2+x_2^2}} & (x_1,x_2)\neq(0,0)\\ 0 & (x_1,x_2)=(0,0)\end{cases}$$

f ist auf $\mathbb{R}^2\setminus\{(0,0)\}$ differenzierbar, da es aus differenzierbaren Funktionen mittels Summe, Produkt und Quotienten zusammengesetzt ist.

$$\begin{aligned}\tfrac{\partial}{\partial x_1}f(x_1,x_2) &= \frac{x_2}{\sqrt{x_1^2+x_2^2}}-\frac{x_1^2x_2}{(x_1^2+x_2^2)^{\frac{3}{2}}}\\ \tfrac{\partial}{\partial x_1}f(0,x_2) &= \frac{x_2}{x_2}=1\\ \tfrac{\partial}{\partial x_1}f(x_1,0) &= 0\end{aligned}$$

Also haben wir

$$\lim_{x_2\to 0}\frac{\partial}{\partial x_1}f(0,x_2)\neq\lim_{x_1\to 0}\frac{\partial}{\partial x_1}f(x_1,0)\,,$$

d. h., die partielle Ableitung $\frac{\partial}{\partial x_1}f$ ist in $(0,0)$ nicht stetig und daher ist f dort nicht differenzierbar.

Betrachten wir gleich den Graphen dieser Funktion bzw. seine Niveauliniendarstellung (siehe Abbildung 4). Wir stellen fest, dass

- f symmetrisch bezüglich der Vertauschung von x_1 und x_2 ist, d. h., symmetrisch bezüglich der Ebene $x_1=x_2$ ist,
- f symmetrisch bezüglich der Vertauschung von x_1 und $-x_2$ ist, d. h., symmetrisch bezüglich der Ebene $x_1=-x_2$ ist,
- $f(-x_1,x_2)=-f(x_1,x_2)$, d. h., symmetrisch bezüglich der x_2-Achse ist,
- $f(x_1,-x_2)=-f(x_1,x_2)$, d. h., symmetrisch bezüglich der x_1-Achse ist.

Weiters gilt für die Niveaulinien $(x_1,x_2)\neq(0,0)$

$$\begin{aligned}\frac{x_1x_2}{\sqrt{x_1^2+x_2^2}}=c \quad&\Leftrightarrow\quad x_1x_2=c\sqrt{x_1^2+x_2^2} \quad\Rightarrow\quad x_1^2x_2^2=c^2(x_1^2+x_2^2)\\ &\Leftrightarrow\quad x_2^2(x_1^2-c^2)=c^2x_1^2 \quad\Rightarrow\quad x_2=\pm\frac{cx_1}{\sqrt{x_1^2-c^2}}\end{aligned}$$

also haben wir Hyperbeln als Niveaulinien.

Es gilt aber folgender Satz (allgemein für $f:\mathbb{R}^n\to\mathbb{R}^m$):

THEOREM 7.6
Sei $A\subseteq\mathbb{R}^n$ eine offene Menge und $f:A\to\mathbb{R}^m$ zweimal stetig partiell differenzierbar. Dann gilt für alle $x_0\in A$ und alle $i,j\in\{1,\ldots,n\}$:

$$\frac{\partial^2}{\partial x_i\partial x_j}f(x_0)=\frac{\partial^2}{\partial x_j\partial x_i}f(x_0)$$

Auch für höhere partielle Ableitungen gilt als Folgerung:

THEOREM 7.7
Sei $A \subseteq \mathbb{R}^n$ eine offene Menge und $f : A \to \mathbb{R}^m$ k-mal stetig partiell differenzierbar. Dann gilt für jede Permutation π der Zahlen $1, \ldots, k$:

$$\frac{\partial^k}{\partial x_{i_1} \partial x_{i_k}} f = \frac{\partial^k}{\partial x_{i_{\pi(1)}} \partial x_{i_{\pi(k)}}} f \ .$$

D. h. etwa

$$\frac{\partial^3 f}{\partial x_1 \partial x_2 \partial x_3} = \frac{\partial^3 f}{\partial x_2 \partial x_3 \partial x_1} = \frac{\partial^3 f}{\partial x_3 \partial x_1 \partial x_2} \ .$$

Wir kennen jetzt Ableitungen von reellen Funktionen. Wie sieht das allgemein bei Funktionen $f : \mathbb{R}^n \to \mathbb{R}^m$ aus? Dort haben wir als Koordiantenfunktionen jeweils reelle Funktionen, sodass wir dasselbe wie oben koordinatenweise anwenden können. Als erste Ableitung erhalten wir also:

DEFINITION 7.11
Ist $f : \mathbb{R}^n \to \mathbb{R}^m$, d. h., $f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix}$ und sind alle Koordinatenfunktionen partiell differenzierbar, so ist die erste Ableitung von f die Jacobimatrix (Funktionalmatrix)

$$\begin{aligned} \nabla f(x) \ &= \ J_f(x) = \nabla \begin{pmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{pmatrix} = \begin{pmatrix} \nabla f_1(x_1, \ldots, x_n) \\ \vdots \\ \nabla f_m(x_1, \ldots, x_n) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial}{\partial x_1} f_1(x_1, \ldots, x_n) & \cdots & \frac{\partial}{\partial x_n} f_1(x_1, \ldots, x_n) \\ \vdots & \vdots & \vdots \\ \frac{\partial}{\partial x_1} f_m(x_1, \ldots, x_n) & \cdots & \frac{\partial}{\partial x_n} f_m(x_1, \ldots, x_n) \end{pmatrix} \end{aligned}$$

Ist die Jacobimatrix quadratisch, dann heißt ihre Determinante *Funktionaldeterminante*.

Da wir m Komponentenfunktionen jeweils n Mal partiell ableiten, ist die Jacobimatrix eine $m \times n$-Matrix.

Zweite Ableitungen von Funktionen $f : \mathbb{R}^n \to \mathbb{R}^m$ werden in geschlossener Form im Allgemeinen nicht mehr betrachtet, da wir jede partielle Ableitung wiederum partiell in n Richtungen ableiten können also einen „$m \times n \times n$-Würfel" erhielten, der nicht so leicht handhabbar ist. Für Taylorreihen spielen höhere Ableitungen jedoch eine Rolle.

BEISPIEL 7.7
$f(x_1, x_2, x_3) = \begin{pmatrix} 2x_1 x_2^2 \\ x_2 \exp(x_3^2) + 2x_1 \end{pmatrix}$

$$\nabla f(x) = \begin{pmatrix} 2x_2^2 & 4x_1 x_2 & 0 \\ 2 & \exp(x_3^2) & 2x_2 x_3 \exp(x_3^2) \end{pmatrix}$$

BEISPIEL 7.8

$$f(x) = \sin(x)\cos(x), \quad f'(x) = \cos^2(x) - \sin^2(x)$$
$$f(x,y) = \sin(x)\cos(y), \quad \nabla f(x,y) = (\cos(x)\cos(y), -\sin(x)\sin(y))$$
$$f(x,y) = \begin{pmatrix} \sin(x) \\ \cos(y) \end{pmatrix}, \quad \nabla f(x,y) = J_f(x,y) = \begin{pmatrix} \cos(x) & 0 \\ 0 & -\sin(y) \end{pmatrix}$$

7.2.1 Differentiationsregeln

Wie im Eindimensionalen gelten auch hier weiterhin die gewöhnten Differentiationsregeln. Sie können bewiesen werden, indem man die Funktionen und Ableitungen komponentenweise betrachtet. Voraussetzung ist natürlich, dass die beteiligten Funktionen differenzierbar sind.

- $\nabla(f+g)(x) = \nabla f(x) + \nabla g(x)$, $f, g : \mathbb{R}^n \to \mathbb{R}^m$ $(J_{f+g} = J_f + J_g)$
- $\nabla(\lambda f)(x) = \lambda \nabla f(x)$, $\lambda \in \mathbb{R}$ $(J_{\lambda f} = \lambda J_f)$
- Ist f in einer Umgebung $U_\varepsilon(x_0)$ konstant, so ist $\nabla f(x_0) = 0$.
- Ist f eine lineare Abbildung $(f(x) = A_f x)$, so ist $\nabla f(x) = A_f$.
- $\nabla(\frac{1}{f(x)}) = -\frac{\nabla f(x)}{f^2(x)}$, falls $f : \mathbb{R}^n \to \mathbb{R}$.
- Verallgemeinerte Kettenregel:$\nabla g(f(x)) = \nabla g(f(x)) \cdot \nabla f(x)$.

THEOREM 7.8 (VERALLGEMEINERTE KETTENREGEL)
Sei $f : \mathbb{R}^n \to \mathbb{R}^m$ und $g : \mathbb{R}^m \to \mathbb{R}^k$, f differenzierbar in x und g differenzierbar in $f(x) = y$, so ist $g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ und $(g \circ f)(x) = g(f(x))$ differenzierbar in x und es gilt:

$$\nabla(g \circ f)(x) = \nabla g(f(x)) \cdot \nabla f(x)$$

BEISPIEL 7.9
Sei

$$f(x_1, x_2) = \begin{pmatrix} x_2 - x_1 \\ \exp(x_2) \end{pmatrix} \quad \text{und} \quad g(x_1, x_2) = \begin{pmatrix} x_1 + x_2 \\ x_2^2 \\ x_1 \end{pmatrix},$$

dann ist

$$(g \circ f)(x_1, x_2) = \begin{pmatrix} x_2 - x_1 + \exp(x_2) \\ \exp(2x_2) \\ x_2 - x_1 \end{pmatrix}$$

und

$$\nabla f(x_1, x_2) = \begin{pmatrix} -1 & 1 \\ 0 & \exp(x_2) \end{pmatrix} \quad \text{bzw.} \quad \nabla g(x_1, x_2) = \begin{pmatrix} 1 & 1 \\ 0 & 2x_2 \\ 1 & 0 \end{pmatrix}.$$

Da

$$\nabla(g \circ f)(x_1, x_2) = \nabla g(f(x_1, x_2)) \cdot \nabla f(x_1, x_2)$$

ist

$$\nabla(g \circ f) = \begin{pmatrix} 1 & 1 \\ 0 & 2\exp(x_2) \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 \\ 0 & \exp(x_2) \end{pmatrix} = \begin{pmatrix} -1 & 1+\exp(x_2) \\ 0 & 2\exp(2x_2) \\ -1 & 1 \end{pmatrix}.$$

BEMERKUNG 7.4 (FUNKTIONENRÄUME)
In der Mathematik ist C die Menge der stetigen Funktionen und C^r die Menge der r-mal stetig differenzierbaren Funktionen. Es gilt

$$C^\infty \subseteq \cdots \subseteq C^r \subseteq C^{r-1} \subseteq \cdots C^1 \subseteq C\,.$$

7.2.2 Richtungsableitung

BEISPIEL 7.10
Sei $f : \mathbb{R}^2 \to \mathbb{R}$. Wir wollen die Ableitung von f im Punkt x_0 in Richtung eines Vektors $a \in \mathbb{R}^2$ betrachten, wobei a normiert sein soll, d. h. $\|a\| = 1$ (Die Länge von a soll sich nicht auf den Wert der Ableitung auswirken).

Sei $h(t) = x_0 + t \cdot a$, dann ist $\nabla h(t) = a$, $t \in \mathbb{R}$.

Für $x = x_0 + ta = h(t)$ gilt:

$$f(x) = f(x_0 + ta) = f(h(t)) = (f \circ h)(t)\,,$$

also

$$\nabla(f \circ h)(t) = \nabla f(x_0 + ta) \cdot \nabla h(t)\,.$$

Also ist in x_0, d. h. $t = 0$:

$$\nabla(f \circ h)(0) = \nabla f(x_0) \cdot a = \langle \nabla f(x_0), a \rangle\,.$$

Die Richtungsableitung ist also nichts anderes als das innere Produkt (Skalarprodukt) des Gradienten von f mit dem Richtungsvektor a.

Dies gilt allgemein.

DEFINITION 7.12
Die Richtungsableitung einer Funktion $f : \mathbb{R}^n \to \mathbb{R}^m$ in Richtung $a \in \mathbb{R}^n$, mit $\|a\| = 1$, in einem Punkt x_0 ist definiert durch:

$$D_a f(x_0) = \nabla f(x_0) \cdot a = \langle \nabla f(x_0), a \rangle$$

BEMERKUNG 7.5
Für die Richtungsableitung gilt:

$$D_a f(x_0) = \lim_{t \to 0} \frac{f(x_0 + ta) - f(x_0)}{t}$$

und

$$D_{e_k} f(x_0) = \frac{\partial}{\partial x_k} f(x_0)\,,$$

wobei e_k den k-ten Einheitsvektor (die k-te Einheitsrichtung) bezeichnet.

Für welchen Vektor a ist die Richtungsableitung maximal?
– Da diese nichts anderes ist als das innere Produkt von a mit dem Gradienten von f und das innere Produkt proportional zum Cosinus des eingeschlossenen Winkels ist, so muss a parallel zum Gradienten liegen, da der Cosinus von 0° am größten ist, also

$$a = \frac{\nabla f(x_0)}{\|\nabla f(x_0)\|}\,,$$

d. h., **der Gradient zeigt in Richtung des größten Anstiegs.**

Bemerkung: Wenn wir eine Funktion maximieren wollen und irgendwo anfangen, werden wir in Richtung des Gradienten weitergehen (Gradientenverfahren). Ist der Gradient gleich null, so sind wir in einem stationären Punkt angekommen, d. h., wir kommen nicht weiter. Wir könnten uns damit in einem Maximum befinden.

7.2.3 Mittelwertsätze

Wie im eindimensionalen Fall, wo für eine differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ in einer Umgebung von x_0 gilt, dass $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$ (Definition des Differentialquotienten, Tangentengleichung), gibt es auch im mehrdimensionalen Fall solche Approximationen.

THEOREM 7.9
Sei $f : A \subseteq \mathbb{R}^n \to \mathbb{R}$ eine reelle, auf der offenen Menge A differenzierbare Funktion und enthalte A die Strecke S von x nach x_0. Dann gibt es einen Punkt $\xi \in S$ auf dieser Strecke, für den gilt

$$f(x) = f(x_0) + \nabla f(\xi)(x - x_0)$$

bzw.

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + O(\|x - x_0\|^2)\,.$$

Ist f zweimal stetig differenzierbar, so gilt

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^t \nabla^2 f(\xi)(x - x_0)$$

bzw.

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^t \nabla^2 f(x_0)(x - x_0) + O(\|x - x_0\|^3)\,.$$

Hierbei wurden die Landau'schen O-Symbole verwendet:

o (Klein-o) : $f(x) = o(g(x))$ mit $x \to a$, wenn $\lim_{x \to a} \frac{|f(x)|}{|g(x)|} = 0$

O (Groß-O) : $f(x) = O(g(x))$ mit $x \to a$, wenn $\exists C > 0 : |f(x)| \leq C|g(x)|$

Da im obigen Fall die Funktion $g(x) = \|x - x_0\|^2$ bzw. $g(x) = \|x - x_0\|^3$ mit $x \to x_0$ schnell gegen 0 geht, heißt das, dass der Fehler $O(g(x))$ mindestens genauso schnell gegen 0 geht: $\|x - x_0\| = h \ll 1$, es gilt nun für dieses h: h^3 geht schneller gegen 0 als h^2, wenn h gegen 0 geht.

BEISPIEL 7.11
Sei $f(x_1, x_2, x_3) = x_1 x_2^2 + x_2 \exp(x_3)$.
Dann ist

$$\nabla f(x_1, x_2, x_3) = (x_2^2, 2x_1x_2 + \exp(x_3), x_2 \exp(x_3))$$

und

$$\nabla^2 f(x_1, x_2, x_3) = \begin{pmatrix} 0 & 2x_2 & 0 \\ 2x_2 & 2x_1 & \exp(x_3) \\ 0 & \exp(x_3) & x_2 \exp(x_3) \end{pmatrix}.$$

In der Nähe des Ursprungs $x_0 = (0, 0, 0)$ ist dann

$$\begin{aligned} f(x_1, x_2, x_3) &\approx f(x_0) + \nabla f(x_0)(x - x_0) + \tfrac{1}{2}(x - x_0)^t \nabla^2 f(x_0)(x - x_0) \\ &= 0 + (0,1,0)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \tfrac{1}{2}(x_1, x_2, x_3)\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= x_2 + x_2 x_3 \end{aligned}$$

Analoges gilt für allgemeine differenzierbare mehrdimensionale Funktionen, wir betrachten aber nur den folgenden Fall.

THEOREM 7.10
Sei $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ eine auf der offenen Menge A differenzierbare Funktion und enthalte A die Strecke S von x nach x_0. Dann gibt es einen Punkt $\xi \in S$ auf dieser Strecke, für den gilt

$$f(x) = f(x_0) + \nabla f(\xi)(x - x_0)$$

bzw.

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + O(\|x - x_0\|^2).$$

$\nabla f(x_0)$ ist nun natürlich die Jacobimatrix $J_f(x_0)$.

BEISPIEL 7.12
Sei $f(x_1, x_2) = \begin{pmatrix} x_2 - x_1 \\ \exp(x_2) \end{pmatrix}$ dann ist

$$\nabla f(x_1, x_2) = \begin{pmatrix} -1 & 1 \\ 0 & \exp(x_2) \end{pmatrix}$$

und in der Nähe des Ursprungs $x_0 = (0, 0)$ gilt

$$f(x_1, x_2) \approx f(x_0) + \nabla f(x_0)(x - x_0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 - x_1 \\ 1 + x_2 \end{pmatrix}.$$

7.2.4 Tangentialabbildung

Betrachten wir lokal die lineare Approximation einer Funktion, so können wir folgende Definition geben.

DEFINITION 7.13
Sei $f : A \subseteq \mathbb{R}^n \to \mathbb{R}^m$ in einem inneren Punkt von A differenzierbar.
Die Funktion

$$g(x) = f(x_0) + \nabla f(x_0)(x - x_0)$$

ist eine affin lineare Abbildung und heißt *Tangentialabbildung* von f.

Für $f : \mathbb{R}^2 \to \mathbb{R}$ und $x_0 = (x_1^{(0)}, x_2^{(0)}) \in \mathbb{R}^2$, sowie $x = (x_1, x_2)$ ist der Graph von

$$\begin{aligned} g(x_1, x_2) &= f(x_0) + \nabla f(x_0)(x - x_0) \\ &= f(x_1^{(0)}, x_2^{(0)}) + \tfrac{\partial}{\partial x_1} f(x_1^{(0)}, x_2^{(0)})(x_1 - x_1^{(0)}) + \tfrac{\partial}{\partial x_2} f(x_1^{(0)}, x_2^{(0)})(x_2 - x_2^{(0)}) \end{aligned}$$

die *Tangentialebene* an f im Punkt $x_0 = (x_1^{(0)}, x_2^{(0)})^t$ (Ortsvektor).

Bem.: Obige Bestandteile der Formel sind als „Matrizen" bzw. Vektoren zu behandeln.

BEISPIEL 7.13
An die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x_1, x_2) = x_1 x_2$ soll im Funktionswert des Punktes $x_0 = (x_1^{(0)}, x_2^{(0)}) = (1, 1)$ eine Tangentialebene gelegt werden.

$$\begin{aligned} g(x_1, x_2) &= f(x_0) + \nabla f(x_0)(x - x_0) \\ &= x_1^{(0)} \cdot x_2^{(0)} + (x_2^{(0)}, x_1^{(0)}) \cdot (\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \end{pmatrix}) \\ &= 1 \cdot 1 + (1, 1) \cdot (\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}) \\ &= 1 + (1, 1) \cdot \begin{pmatrix} x_1 - 1 \\ x_2 - 1 \end{pmatrix} \\ &= 1 + (x_1 - 1) + (x_2 - 1) \\ &= x_1 + x_2 - 1 \end{aligned}$$

Das Ergebnis zeigt Abbildung 5. Rechts wurde die Abbildung so gedreht, dass die Tangentialebene projizierend wurde, d.h., sie steht normal auf die Ebene des Hintergrundes.

7.2.5 Das totale Differential

Für eine differenzierbare Funktion $f : \mathbb{R}^n \to \mathbb{R}^m$ liefert die Tangentenabbildung

$$f_t(x) \doteq f(x_0) + \nabla f(x_0)(x - x_0)$$

in der Umgebung von x_0 eine gute Näherung der Funktion f, also

$$f_t(x) - f(x_0) = \nabla f(x_0)(x - x_0) \approx f(x) - f(x_0)\,.$$

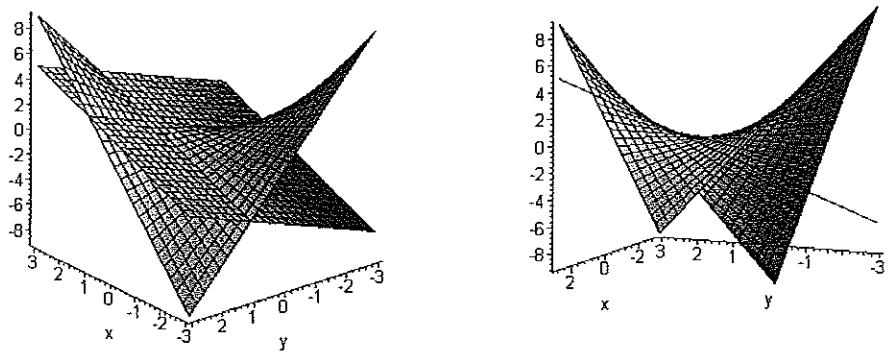


Abbildung 5: Tangentialebene an $f(x_1, x_2) = x_1 x_2$ in $(1, 1)$.

DEFINITION 7.14
Sei $f : \mathbb{R}^n \to \mathbb{R}^m$ differenzierbar. Die lineare Abbildung

$$df(x_0) \doteqdot \nabla f(x_0) \cdot (x - x_0) = \frac{\partial}{\partial x_1} f(x_0) dx_1 + \cdots + \frac{\partial}{\partial x_n} f(x_0) dx_n \,,$$

mit $dx_i \doteqdot (x_i - x_i^{(0)})$, heißt *totales Differential* der Funktion f in x_0. Das totale oder vollständige Differential wird oft auch mit $\frac{\partial f}{\partial x}$ bezeichnet.

D. h., das totale Differential gibt also die ungefähre Abweichung des Funktionswertes $f(x)$ an, wenn x von x_0 wenig abweicht.

BEISPIEL 7.14 (FEHLERFORTPFLANZUNGSGESETZ)
Gegeben sei eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ von n Messgrößen $x_1, \ldots, x_n$.
Die Messungenauigkeit betrage $\pm\Delta x_1, \ldots, \pm\Delta x_n$, wobei $\Delta x_i > 0$ für alle $i = 1, \ldots, n$.
Dann gilt für den Fehler $\Delta f(x)$ von $f(x_1, \ldots, x_n)$:

$$\Delta f(x_1, \ldots, x_n) < \left|\frac{\partial}{\partial x_1} f(x)\right| \Delta x_1 + \cdots + \left|\frac{\partial}{\partial x_n} f(x)\right| \Delta x_n$$

DEFINITION 7.15
Die Größe

$$\Delta f_{max}(x_1, \ldots, x_n) := \left|\frac{\partial}{\partial x_1} f(x)\right| \Delta x_1 + \cdots + \left|\frac{\partial}{\partial x_n} f(x)\right| \Delta x_n$$

heißt der *maximale Fehler* von f.
Der Quotient $\frac{df_{max}(x)}{df(x)}$ heißt der *relative maximale Fehler* von f.

BEISPIEL 7.15
Aus der Physik wissen wir: $s = v \cdot t$, d.h., *Weg=Geschwindigkeit* $\times$ *Zeit*,
mit den SI-Einheiten: $[m] = [m/s] \cdot [s]$.

Ein Marathonläufer soll über eine Strecke von $s = 10\,\text{km}$ eine Zeit von $t = 30\,\text{min}$ erzielen.
D. h., die mittlere Geschwindigkeit ist $v = \frac{s}{t} = 20\,\text{km/h}$.
Der Messfehler für die Strecke s sei $\Delta s = \pm 1\,\text{m}$, jener der Zeit $\Delta t = \pm 1\,\text{sec}$.
Wie groß ist dann Δv maximal?

$$\begin{aligned}\Delta v(s,t) &= \left|\tfrac{\partial}{\partial s} v(s,t)\right| \Delta s + \left|\tfrac{\partial}{\partial t} v(s,t)\right| \Delta t = \left|\tfrac{1}{t}\right| \Delta s + \left|-\tfrac{s}{t^2}\right| \Delta t \\ &= \frac{0.001}{0.5}\frac{\text{km}}{\text{h}} + \frac{10}{0.25}\frac{\text{km}}{\text{h}^2}\frac{1}{3600}\text{h} = 0.013\,\text{km/h} \\ &\Rightarrow v = (20 \pm 0.013)\,\text{km/h}\end{aligned}$$

7.2.6 Die Jacobimatrix der Umkehrabbildung

Wir erinnern uns, dass im eindimensionalen Fall für die Umkehrabbildung f^{-1} einer differenzierbaren Funktion $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto f(x) = y$ gilt (falls sie existiert):

$$(f^{-1}(y))' = \frac{1}{f'(f^{-1}(y))}\,.$$

Das ist ableitbar aus $f \circ f^{-1} = id$ und der Kettenregel bei der Ableitung dieser Gleichung.

Ähnliches gilt auch für differenzierbare Funktionen $f : \mathbb{R}^n \to \mathbb{R}^n$, nur dass es sich bei der ersten Ableitung um eine Matrix handelt.

THEOREM 7.11
Sei

$$f : \mathbb{R}^n \to \mathbb{R}^n \qquad f(x) = \begin{pmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_n(x_1, \ldots, x_n) \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

differenzierbar und

$$h : \mathbb{R}^n \to \mathbb{R}^n \qquad h(y) = \begin{pmatrix} h_1(y_1, \ldots, y_n) \\ \vdots \\ h_n(y_1, \ldots, y_n) \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

mit

$$h = f^{-1}, \text{ also } h \circ f = f \circ h = id$$

so gilt

$$\nabla h(y_1, \ldots, y_n) = [\nabla f(h(y_1, \ldots, y_n))]^{-1}\,.$$

BEWEIS. Es gilt $f \circ h = id$, daher $\nabla(f \circ h) = \nabla id = I$, die $n \times n$ Einheitsmatrix. Wegen der allgemeinen Kettenregel gilt nun

$$\nabla(f \circ h)(x) = \nabla f(h(x)) \cdot \nabla h(x) = I$$

und daher

$$\nabla h(x) = [\nabla f(h(x))]^{-1}$$

als Matrix. □

BEISPIEL 7.16
Sei

$$f(x_1, x_2) = \begin{pmatrix} x_2 - x_1 \\ \exp(x_2) \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

dann ist

$$\nabla f(x_1, x_2) = \begin{pmatrix} -1 & 1 \\ 0 & \exp(x_2) \end{pmatrix}.$$

Die Umkehrabbildung zu f ist

$$h(y_1, y_2) = \begin{pmatrix} \log y_2 - y_1 \\ \log y_2 \end{pmatrix},$$

und nach oben ist die Jacobimatrix

$$\nabla h(y_1, y_2) = [\nabla f(h(y_1, y_2))]^{-1} = \begin{pmatrix} -1 & 1 \\ 0 & y_2 \end{pmatrix}^{-1} = -\frac{1}{y_2}\begin{pmatrix} y_2 & -1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & \frac{1}{y_2} \\ 0 & \frac{1}{y_2} \end{pmatrix}.$$

THEOREM 7.12 (SATZ VON DER UMKEHRABBILDUNG)
Sei $f : \mathbb{R}^n \to \mathbb{R}^n$ stetig differenzierbar und $J_f(x_0)$ regulär, dann gibt es eine Umgebung $U_\varepsilon(x_0)$ und $V_\delta(f(x_0))$ und eine Umkehrfunktion $f^{-1} : V_\delta \to U_\varepsilon$ von f auf U_ϵ, die stetig differenzierbar ist.
Es ist

$$J_{f^{-1}}(f(x_0)) = (J_f(x_0))^{-1}$$

BEISPIEL 7.17
Die Funktion $f(x_1, x_2) = \begin{pmatrix} e^{x_1} \cos x_2 \\ e^{x_1} \sin x_2 \end{pmatrix}$ ist nicht global invertierbar.
Es ist etwa $f(0, 0) = f(0, 2\pi) = (1, 0)^t$.

Es ist jedoch

$$J_f(x_1, x_2) = \begin{pmatrix} e^{x_1} \cos x_2 & -e^{x_1} \sin x_2 \\ e^{x_1} \sin x_2 & e^{x_1} \cos x_2 \end{pmatrix}$$

und die **Funktionaldeterminante** $\det J_f = e^{2x_1} > 0$ und damit J_f regulär, d. h. invertierbar, daher ist f überall lokal invertierbar und es gilt etwa:

$$x_0 = (0, \frac{\pi}{2}), \qquad J_f(0, \frac{\pi}{2}) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$f(0,\frac{\pi}{2}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$J_{f^{-1}}(f(0,\frac{\pi}{2})) = J_{f^{-1}}(0,1) = J_f(0,\frac{\pi}{2})^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

BEISPIEL 7.18
Die komplexe Funktion $f : \mathbb{C} \to \mathbb{C}$, $z \mapsto w = z^2$ kann als mehrdimensionale Funktion $g : \mathbb{R}^2 \to \mathbb{R}^2$, $(x, y) \mapsto (u, v)$ aufgefasst werden, mit $z = x + iy$ und $w = u + iv$, also $u = x^2 - y^2$ und $v = 2xy$.

Es gilt

$$J_f(z) = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix} \quad \Rightarrow \qquad \det J_f(z) = 4(x^2 + y^2) = 4|z|^2$$

d. h., für alle $z \neq 0$ gibt es eine Umgebung, die bijektiv auf z^2 abgebildet wird!
Daher gibt es dort auch eine (lokale) Umkehrfunktion.

7.2.7 Implizite Funktionen

Betrachten wir eine Funktion

$$f : (\mathbb{R}^n \times \mathbb{R}^m) \to \mathbb{R}^m, \qquad (x, y) \mapsto f(x, y), \quad x \in \mathbb{R}^n, y \in \mathbb{R}^m,$$

so beschreibt jede Niveaumenge $\{(x, y) \mid f(x, y) = c\}$ eine Beziehung zwischen $x \in \mathbb{R}^n$ und $y \in \mathbb{R}^m$. Falls dieser Zusammenhang eine Funktion ist, können wir explizit $y = y(x)$, d. h., y als Funktion von x schreiben.

BEMERKUNG 7.6
Einfacher gesagt: Ist $f : (\mathbb{R}^n \times \mathbb{R}^m) \to \mathbb{R}^m$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, so bildet die Niveaumenge $f(x, y) = c$, $c \in \mathbb{R}^m$, ein Gleichungssystem mit m Gleichungen (koordinatenweise) in $n + m$ Unbekannten. Wenn wir aus diesen Gleichungen y ausrechnen können, haben wir $y = g(x)$, mit einer passenden Funktion $g : \mathbb{R}^n \to \mathbb{R}^m$. Finden wir ein solches $g(x)$, so haben wir $f(x, g(x)) = c$, d. h., $\nabla(f(x, g(x))) = 0$. Das kann nach der Kettenregel abgeleitet werden. Dieser Sachverhalt ist unten ausführlicher dargestellt.

BEISPIEL 7.19
Denken wir uns die Funktion $f(x, y) = x^2 + y^2$, so beschreibt die Niveaumenge $x^2 + y^2 = 1$ den Einheitskreis. In einer Umgebung des Punktes $(0, 1)$ kann y als $y(x) = \sqrt{1 - x^2}$ dargestellt werden (siehe Abb. 7), d. h., y kann explizit nach x aufgelöst werden. Analog haben wir in einer Umgebung von $(0, -1)$ $y(x) = -\sqrt{1 - x^2}$, aber auch umgekehrt in einer Umgebung von $(1, 0)$ $x(y) = \sqrt{1 - y^2}$ etc. D. h., wir bekommen einfachere Funktionen (kleinere Dimension).

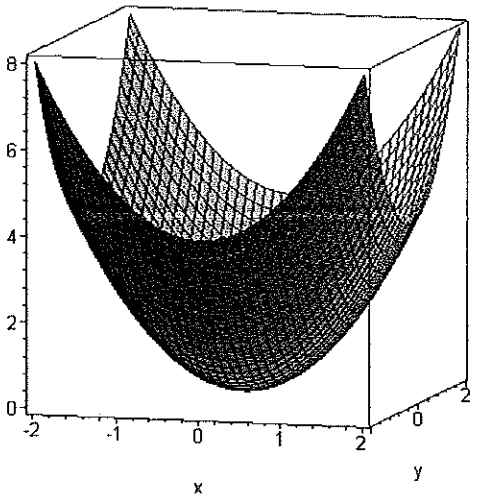


Abbildung 6: Die Funktion $f(x,y) = x^2 + y^2$.

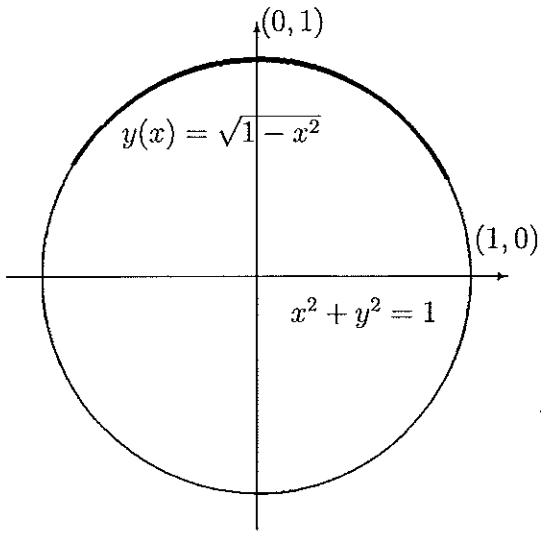


Abbildung 7: Die Niveaumenge $f(x,y) = x^2 + y^2 = 1$.

Sei $f : (\mathbb{R}^n \times \mathbb{R}^m) \to \mathbb{R}^k$ so hat die Jacobimatrix ∇f die Dimension $k \times (n+m)$.
Zerlegen wir diese Matrix spaltenweise in Submatrizen

$$\nabla f = (\nabla_1 f \ \nabla_2 f)$$

mit den Dimensionen $k \times n$ $(\nabla_1 f)$ und $k \times m$ $(\nabla_2 f)$.
Die Matrix $\nabla_1 f$ entsteht, wenn wir nur nach x ableiten und y festhalten, $\nabla_2 f$ entsteht, wenn wir nur nach y ableiten und x festhalten.

Sei nun $x = g(z)$ und $y = h(z)$, d. h., wir „trennen" die Variablen. Dann ist die Jacobimatrix von $p : z \mapsto f(g(z), h(z))$

$$\nabla p = \nabla f \cdot \nabla \begin{pmatrix} g \\ h \end{pmatrix} = (\nabla_1 f \ \ \nabla_2 f) \cdot \begin{pmatrix} \nabla g \\ \nabla h \end{pmatrix} = \nabla_1 f \ \nabla g + \nabla_2 f \ \nabla h \,.$$

Wir können nun folgenden Satz formulieren.

THEOREM 7.13 (HAUPTSATZ ÜBER IMPLIZITE FUNKTIONEN)
Sei $f : (\mathbb{R}^n \times \mathbb{R}^m) \to \mathbb{R}^m$, $(x, y) \mapsto f(x, y)$, und y_0 die eindeutige Lösung der Gleichung $f(x_0, y_0) = c$. Ist weiters $\nabla_2 f(x_0, y_0)$ invertierbar, so kann die Gleichung $f(x_0, y_0) = c$ nach x aufgelöst werden und somit $y(x)$ eindeutig bestimmt werden. Es ist dann die Funktion $x \mapsto y(x)$ differenzierbar und es gilt

$$\nabla y(x) = -[\nabla_2 f(x, y(x))]^{-1} \cdot \nabla_1 f(x, y(x))$$

$((m \times n) = (m \times m) \cdot (m \times n))$.

BEWEIS. Sei nach oben $z = x$, $g(x) = x$ und $h(x) = y(x)$, so ist $f(x, y(x)) = c$, also eine Konstante, daher ist $\nabla f(x, y(x)) = 0$, die Nullmatrix. Nach oben gilt also

$$0 = \nabla_1 f(x, y(x)) \cdot I + \nabla_2 f(x, y(x)) \cdot \nabla y(x) \,,$$

also gilt der Hauptsatz (Auflösen nach $\nabla y(x)$). □

8 Kurven

Einfache mehrdimensionale Funktionen $f(t) = (f_1(t), \ldots, f_n(t))$ sind Kurven. Diese haben jeweils das Argument in $\mathbb{R}$ und den Funktionswert im $\mathbb{R}^n$. Im Gegensatz zu allgemeinen mehrdimensionalen Funktionen können auch Raumkurven, das sind Funktionen von $\mathbb{R} \to \mathbb{R}^3$, dargestellt werden, obwohl der Graph eigentlich 4-dimensional wäre. Es wird einfach die Laufvariable, das Argument, nicht mit dargestellt, sondern nur die Werte (das Bild) bzw. die Spur $f(t)$.

DEFINITION 8.1
Eine stetige Funktion $f : [a, b] \to \mathbb{R}^n,\ t \mapsto f(t)$ heißt *Weg*.
$f(a)$ heißt *Anfangspunkt*, $f(b)$ heißt *Endpunkt*.
Falls $f(a) = f(b)$ ist, heißt der Weg *geschlossener Weg*.

Die Menge $\{f(t) \mid t \in [a, b]\} \subseteq \mathbb{R}^n$ heißt *Spur*.

BEISPIEL 8.1
Die zwei folgenden Wege f_1 und f_2 haben dieselbe Spur.

$$\begin{aligned} f_1 : [0, 2\pi] \to \mathbb{R}^2, \quad f_1(t) &= (\cos t, \sin t) \\ f_2 : [0, 2\pi] \to \mathbb{R}^2, \quad f_2(t) &= (\cos 2t, \sin 2t) \end{aligned}$$

f_1 durchläuft die Spur einmal, f_2 zweimal.

Ist die Funktion $f(t)$ ein Weg, so ist ihr Graph die Spur des Weges $F : t \mapsto (t, f(t))$.

Betrachten wir die Wege

$$\begin{aligned} f_1 : [0, \tfrac{\pi}{2}] \to \mathbb{R}^2, \quad f_1(t) &= (\sin t, \cos t) \\ f_2 : [0, 1] \to \mathbb{R}^2, \quad f_2(u) &= (u, \sqrt{1 - u^2}) \end{aligned}$$

so beobachten wir, dass diese Wege gleich sind. Die Parametertransformation $u = \sin t$ führt f_1 in f_2 über.

DEFINITION 8.2
Zwei Wege

$$\begin{aligned} f_1 : [a, b] = I_1 \to \mathbb{R}^n, \quad t \mapsto f_1(t) \\ f_2 : [c, d] = I_2 \to \mathbb{R}^n, \quad t \mapsto f_2(t) \end{aligned}$$

heißen *äquivalent*, wenn es eine stetige, streng monoton wachsende Parametertransformation $w : [a, b] \to [c, d],\ t \mapsto t' = w(t)$ gibt, sodass $f_2(w(t)) = f_1(t),\ \forall t \in [a, b]$, bzw. gleichbedeutend damit, $f_2(t') = f_1(w^{-1}(t')),\ \forall t' \in [c, d]$.

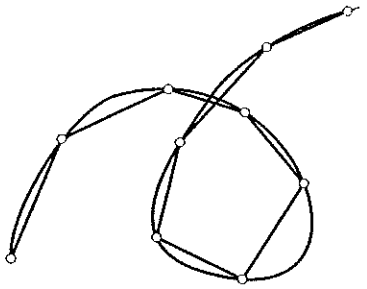
Diese Definition beschreibt eine Äquivalenzrelation auf der Menge der Wege, d. h., eine Relation zwischen Wegen, die reflexiv, symmetrisch und transitiv ist. Wir erhalten daher eine Klasseneinteilung der Wege in (paarweise disjunkte) Klassen.

DEFINITION 8.3
Eine Äquivalenzklasse von Wegen heißt *stetige Kurve* $\gamma : t \mapsto \gamma(t)$.
Eine Kurve besitzt unterschiedliche, äquivalente Parameterdarstellungen.

8.1 Die Länge einer Kurve

Wir unterteilen eine Kurve in Abschnitte und approximieren die Länge der Kurve durch Streckenabschnitte $V_T(f) = \sum_{k=1}^{N} |f(t_k) - f(t_{k-1})|$, wobei T die Unterteilung des Intervalls in $(t_0, t_1, \ldots, t_N)$ ist.

Die totale Variation $V_{[a,b]}(f) = \sup\{V_T(f)\}$ ist dann das Supremum dieser „Unterteilungen".



Existiert eine differenzierbare Parameterdarstellung $f(t)$ für die Kurve $\gamma : t \mapsto (x(t), y(t))$, so können wir mit Hilfe des Satzes von Pythagoras und dem Mittelwertsatz der Differentialrechnung

$$f : [a,b] \to \mathbb{R} \text{ stetig und differenzierbar} \Rightarrow \exists \zeta \in [a,b] : \; f'(\zeta) = \frac{f(b) - f(a)}{b - a}$$

Folgendes zeigen.

$$\begin{aligned} |f(t_k) - f(t_{k-1})|^2 &= (x(t_k) - x(t_{k-1}))^2 + (y(t_k) - y(t_{k-1}))^2 \\ |f(t_k) - f(t_{k-1})| &= \sqrt{(x'(\zeta_1))^2 + (y'(\zeta_2))^2} \cdot (t_k - t_{k-1}) \; . \end{aligned}$$

Mit einer weiteren Verfeinerung der Approximation, d. h., die Intervalle $[t_{k-1}, t_k]$ werden immer kleiner, sodass die nach dem Mittelwertsatz der Differentialrechnung notwendigen ζ_i „in die Intervallgrenzen wandern", erhalten wir die Länge der Kurve aus

$$\begin{aligned} L(\gamma) &= V_{[a,b]}(f) = \lim_{N\to\infty} \sum_{k=1}^{N} |f(t_k) - f(t_{k-1})| \\ &= \lim_{N\to\infty} \sum_{k=1}^{N} \sqrt{(x'(t_k))^2 + (y'(t_k))^2} \cdot (t_k - t_{k-1}) \\ &= \int_a^b \sqrt{(x'(t))^2 + (y'(t))^2}\, dt \end{aligned}$$

Aus der einfachen Verallgemeinerung für n Dimensionen folgt folgender Satz.

THEOREM 8.1
Ist $f : [a,b] \to \mathbb{R}^n$, $t \mapsto f(t) = (f_1(t), \ldots, f_n(t))$ eine stetig differenzierbare Kurve, so ist ihre Länge gegeben durch

$$L(\gamma) = \int_a^b \sqrt{(f_1'(t))^2 + \cdots + (f_n'(t))^2}\, dt$$

Daraus folgt auch, dass der Graph $G_f : t \mapsto (t, f(t))$ die Länge $L(\gamma) = \int_a^b \sqrt{1 + (f'(t))^2}\, dt$ hat.

Ist die Kurve in Polarkoordinaten angegeben, d. h.

$$\gamma : [\alpha, \beta] \to \mathbb{R}^n, \quad \varphi \mapsto r(\varphi),$$

so gilt

$$L(\gamma) = \int_\alpha^\beta \sqrt{(r(\varphi))^2 + (r'(\varphi))^2}\, d\varphi\,.$$

Dies sehen wir leicht, wenn wir von Polarkoordinaten wieder zu kartesischen Koordinaten übergehen. Die Kurve hat dann die Darstellung

$$f : [\alpha, \beta] \to \mathbb{R}^2, \quad \varphi \mapsto (r(\varphi)\cos\varphi, r(\varphi)\sin\varphi)$$

und somit folgt das Behauptete aus

$$x(\varphi) = r(\varphi)\cos\varphi \quad \text{und} \quad y(\varphi) = r(\varphi)\sin\varphi$$

und

$$x'(\varphi) = r'(\varphi)\cos\varphi - r(\varphi)\sin\varphi \quad \text{und} \quad y'(\varphi) = r'(\varphi)\sin\varphi + r(\varphi)\cos\varphi$$

sowie

$$(x'(\varphi))^2 + (y'(\varphi))^2 = (r'(\varphi))^2 + (r(\varphi))^2\,.$$

Beispiel 8.2 (Kreis)
Der Kreis mit Radius R hat die Parameterdarstellung

$$r(\varphi) \equiv R, \quad 0 \le \varphi \le 2\pi$$

und es gilt somit für den Umfang U

$$U = \int_0^{2\pi} \sqrt{R^2 + 0^2}\, d\varphi = 2\pi\, R\,.$$

Sowie der Kreis kann auch die Ellipse einfach dargestellt werden, die Länge der Kurve „Ellipse“ kann allerdings nicht elementar berechnet werden, denn sie führt auf das sogenannte *elliptische Integral*, für das es keinen geschlossenen analytischen Ausdruck gibt (siehe folgendes Beispiel).

Beispiel 8.3 (Ellipse)
Die Ellipse mit der Hauptachsenlänge $2a$ und der Nebenachsenlänge $2b$ hat die Parameterdarstellung

$$\gamma : [0, 2\pi] \to \mathbb{R}^2, \qquad \varphi \mapsto (a\cos\varphi, b\sin\varphi)\,.$$

Für den Umfang gilt

$$\begin{aligned} U &= \int_0^{2\pi} \sqrt{(x'(\varphi))^2 + (y'(\varphi))^2}\, d\varphi = \int_0^{2\pi} \sqrt{a^2\sin^2\varphi + b^2\cos^2\varphi}\, d\varphi \\ &= \int_0^{2\pi} \sqrt{a^2 - (a^2 - b^2)\cos^2\varphi}\, d\varphi = a\int_0^{2\pi} \sqrt{1 - k^2\cos^2\varphi}\, d\varphi \end{aligned}$$

wobei $k = \frac{\sqrt{a^2 - b^2}}{a}$ die sogenannte *Exzentrizität* der Ellipse ist[7].
Da die Ellipse symmetrisch ist, gilt

$$U = 4a \int_0^{\pi/2} \sqrt{1 - k^2 \cos^2 \varphi}\, d\varphi\,.$$

Das hier auftretende Integral

$$E(k) \doteqdot \int_0^{\pi/2} \sqrt{1 - k^2 \cos^2 \varphi}\, d\varphi$$

heißt *vollständiges elliptisches Integral zweiter Gattung*. Es lässt sich nicht mit Hilfe elementarer Funktionen ausdrücken, sondern nur näherungsweise berechnen.

Beispiel 8.4 (Logarithmische Spirale)
Die logarithmische Spirale hat die Parameterdarstellung

$$\gamma : r(\varphi) = e^{k\varphi}, \quad k > 0\,.$$

Die Länge eines Spiralabschnittes, d. h., $\varphi \in [\alpha, \beta]$, ist

$$L(\alpha, \beta) = \int_\alpha^\beta \sqrt{e^{2k\varphi} + k^2 e^{2k\varphi}}\, d\varphi = \sqrt{1 + k^2} \int_\alpha^\beta e^{k\varphi}\, d\varphi = \frac{\sqrt{1 + k^2}}{k} \left(r(\beta) - r(\alpha)\right)$$

und somit proportional zum Zuwachs der Kurve zwischen Anfangs- und Endpunkt des Abschnittes. Der innere Abschnitt der Spirale hat endliche Länge, denn es ist

$$L(-\infty, 0) = \lim_{\alpha \to -\infty} L(\alpha, 0) = \frac{\sqrt{1 + k^2}}{k}\,.$$

Beispiel 8.5 (Zykloiden)
Rollt ein Kreis vom Radius 1 entlang der x-Achse ab und ist an diesem im Abstand R ein Punkt „befestigt“, so beschreibt dieser eine *Zykloide*. Diese hat die Parameterdarstellung

$$\gamma : [0, 2\pi] \to \mathbb{R}^2, \qquad \varphi \mapsto (1 - R\sin\varphi, 1 - R\cos\varphi)\,.$$

Die Länge für eine Umdrehung ist

$$L_R = \int_0^{2\pi} \sqrt{1 - 2R\cos\varphi + R^2}\, d\varphi\,.$$

Ersetzen wir $\varphi = \pi + 2t$, so erhalten wir das elliptische Integral

$$L_R = 4(1 + R) \int_0^{\pi/2} \sqrt{1 - \frac{4R}{(1 + R)^2} \sin^2 t}\, dt = 4(1 + R)\, E\left(\frac{2\sqrt{R}}{1 + R}\right).$$

Ist $R = 1$, d. h., liegt der Punkt am Umfang des rollenden Kreises, so erhalten wir die Länge für einen Umlauf zu

$$L_1 = 8 \int_0^{\pi/2} \sqrt{1 - \sin^2 t}\, dt = 8 \int_0^{\pi/2} \cos t\, dt = 8\,.$$

[7] Ist die Exzentrizität 0, so liegt ein Kreis mit Radius a vor.

So wie an Funktionen können auch an Kurven Tangenten „in Richtung“ der Kurve gelegt werden.

DEFINITION 8.4
Es sei $f : I \to \mathbb{R}^n,\ t \mapsto f(t) = (f_1(t), \ldots, f_n(t))$ eine stetig differenzierbare Parameterdarstellung einer Kurve γ. Der Vektor

$$e(t_0) = \frac{\nabla f(t_0)}{|\nabla f(t_0)|}$$

heißt *Tangenteneinheitsvektor* oder *Tangentialvektor* der Kurve γ im Punkt t_0.

Die Gerade

$$\tau :\ \tau(t) = f(t_0) + t \cdot e(t_0),\ t \in \mathbb{R}$$

heißt *Tangente* an die Kurve im Punkt t_0.

Die Tangente bzw. der Tangentialvektor dreht sich stetig um die Kurve, wenn wir der Kurve entlang gehen. Eine solche Kurve heißt daher *glatt*.

9 Optimierung mehrdimensionaler reeller Funktionen $f : \mathbb{R}^n \to \mathbb{R}$

9.1 Optimierung ohne Nebenbedingungen

Ein Optimum zu suchen heißt, den größten oder den kleinsten Wert zu suchen.

Wir suchen also ein $x_0 \in \mathbb{R}^n$, sodass

- $f(x_0) \leq f(x)$ für alle $x \in \mathbb{R}^n$ (dann heißt x_0 *globales Minimum*)
- $f(x_0) \geq f(x)$ für alle $x \in \mathbb{R}^n$ (dann heißt x_0 *globales Maximum*)

Die Funktion f heißt auch *Zielfunktion.*

Die Formulierung bzw. das Finden einer „geeigneten" Zielfunktion ist meist der schwierigste Teil einer Optimierungsaufgabe.

Zunächst suchen wir aber nach **lokalen** Minima bzw. Maxima:

Suche $x_0 \in \mathbb{R}^n$ so, dass

- $f(x_0) \leq f(x)$ für alle x aus einer Umgebung $U_\varepsilon(x_0)$ um x_0 (x nahe bei x_0): x_0 ist dann ein *lokales Minimum*
- $f(x_0) \geq f(x)$ für alle x aus einer Umgebung $U_\varepsilon(x_0)$ um x_0 (x nahe bei x_0): x_0 ist dann ein *lokales Maximum.*

Wie finden wir solche Punkte?

Beispiel 9.1
$n = 1$: $f : \mathbb{R} \to \mathbb{R}$.
Wir berechnen die Ableitung $f'(x)$ und setzen sie gleich 0: $f'(x) = 0$ (horizontale Tangente, kein Anstieg/Abstieg im Extremum). Punkte, die diese Forderung erfüllen, sind Kandidaten für Extremwerte. Zur genauen Bestimmung berechnen wir $f''(x_0)$, die zweite Ableitung eines Kandidaten x_0, d. h., das Krümmungsverhalten der Funktion in bzw. um x_0. Falls $f''(x_0) > 0$, so ist x_0 ein lokales Minimum, ist $f''(x_0) < 0$, so ist x_0 ein lokales Maximum. Erhalten wir aber $f''(x_0) = 0$, so ist es notwendig weiterzurechnen, eventuell befinden wir uns an einer Sattelstelle, das ist jedoch genauer zu untersuchen.

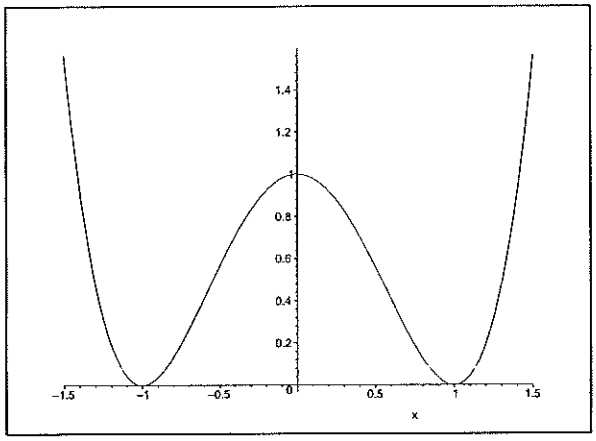


Abbildung 8: Plot der Funktion $(x-1)^2(x+1)^2$.

Sei etwa $f(x) = (x-1)^2(x+1)^2$.
Dann ist $f'(x) = 4x(x-1)(x+1)$ und dieser Ausdruck verschwindet in $x_0 = -1$, $x_1 = 0$ und $x_2 = 1$. Alle drei Punkte sind stationäre Punkte, also Kandidaten für Extremalstellen. Wir berechnen die zweite Ableitung und erhalten $f''(x) = 12x^2 - 4$. Wir stellen fest, dass $f''(-1) = f''(1) = 8 > 0$ und $f''(0) = -4$. Also sind die Stellen x_0 und x_2 lokale Minima und die Stelle x_1 ist ein lokales Maximum (siehe Abbildung 8).

Was passiert bei $n > 1$?

Eine lokale Betrachtung ist möglich mittels (partieller) Ableitungen (falls die Funktion differenzierbar ist, sonst wird es komplizierter). (Partielle) Ableitungen werden in Gradienten $\nabla f(x)$ zusammengefasst. Dieser kann auch als Richtungsvektor im $\mathbb{R}^n$ betrachtet werden. Betrachten wir Niveaumengen der Funktion $f(x)$, so stellt sich heraus, dass $\nabla f(x)$ immer „senkrecht" zu den Niveaumengen steht (egal an welcher Stelle).

Außerdem zeigt der Gradient immer in die Richtung, in der die Funktion vom Punkt x aus gesehen am steilsten ansteigt.

Falls also $\nabla f(x) = 0$, dann gibt es auch lokal keinen An/Abstieg und damit eventuell ein lokales Minimum oder Maximum.

Auf diese Art können wir auch im mehrdimensionalen Raum $\mathbb{R}^n$ Kandidaten für Extremalstellen finden.

Zur Bestimmung, ob ein Minimum oder Maximum vorliegt, brauchen wir wie im eindimensionalen Fall höhere Ableitungen.

Mehrdimensionaler Taylor

Ähnlich wie im eindimensionalen Fall, kann auch für mehrdimensionale Funktionen unter

bestimmten Bedingungen lokal eine Approximation mit Hilfe von Ableitungen an einer bestimmten Stelle berechnet werden. Meist wird nur eine Approximation zweiten Grades benötigt (siehe auch Abschnitt 1.2.3 Mittelwertsätze):

THEOREM 9.1
Sei f zweimal total differenzierbar bei x_0. Dann gilt:

$$f(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^t \nabla^2 f(x_0)(x - x_0) + \text{Rest},$$

wobei $\frac{\text{Rest}}{\|x-x_0\|^2} \to 0$ mit $\|x - x_0\| \to 0$.

THEOREM 9.2 (NOTWENDIGE BEDINGUNG FÜR LOKALE OPTIMA)
Sei x_0 lokales Minimum (Maximum) von f und f bei x_0 differenzierbar,
dann gilt $\nabla f(x_0) = 0$.

BEWEIS. [eigentlich nur Beweisskizze für Minimum]
Wir wissen, dass $\nabla f(x_0)$ in Richtung des steilsten Anstiegs zeigt.
Angenommen, $\nabla f(x_0) \neq 0$. Sei $x = x_0 - \lambda[\nabla f(x_0)]^t$, mit $\lambda > 0$, klein.
Daraus folgt dann, dass auch $\|x - x_0\| = |\lambda| \|\nabla f(x_0)\|$ klein ist. Also

$$\begin{aligned} f(x_0) \leq f(x) &= f(x_0) + \nabla f(x_0)(x - x_0) + \text{Rest (siehe Taylor)} \\ &= f(x_0) - \lambda \nabla f(x_0) \nabla f(x_0)^t + \text{Rest (da } x - x_0 = (-\lambda)[\nabla f(x_0)]^t) \\ &= f(x_0) - \lambda \|\nabla f(x_0)\|^2 + \text{Rest, dieser Rest ist aber sehr klein} \\ &< f(x_0) \text{ und daher ein Widerspruch!} \end{aligned}$$

□

Umgekehrt gilt:

THEOREM 9.3 (HINREICHENDE BEDINGUNG FÜR LOKALE OPTIMA)
Sei f zweimal stetig differenzierbar bei x_0.
Ist die erste Ableitung $\nabla f(x_0) = 0$ und die zweite Ableitung $H = H_f(x_0) = \nabla^2 f(x_0)$ (die Hesse'sche Matrix) positiv definit (d. h., $u^t H u > 0$, für alle $u \neq 0$), dann ist x_0 eine lokale Minimalstelle.

Ist die erste Ableitung $\nabla f(x_0) = 0$ und die zweite Ableitung $H = \nabla^2 f(x_0)$ (die Hesse'sche Matrix) negativ definit (d. h., $u^t H u < 0$, für alle $u \neq 0$), dann ist x_0 eine lokale Maximalstelle

BEWEIS. [wieder nur Beweisskizze für Minimum]

$$\begin{aligned} f(x) &= f(x_0) + \tfrac{1}{2} u^t H u + \text{Rest (klein)} \\ &\geq f(x_0) \text{ für } x \text{ nahe bei } x_0 \text{ (da } u^t H u > 0) \end{aligned}$$

□

BEMERKUNG 9.1
H positiv definit heißt, dass die quadratische Form $u^t H u$, die in der Taylorapproximation den quadratischen Term repräsentiert, positiv gekrümmt ist (d. h., die Steigung nimmt

– im eindimensionalen Fall – von „links nach rechts“ zu), also konvex ist. $u^tHu > 0$ bedeutet, dass alle Eigenwerte der Matrix $H = H_f(x_0)$ positiv sein müssen, da $u^tHu > 0$ für alle Vektoren u gelten muss, also gilt speziell für Eigenvektoren v und den zugehörigen Eigenwert λ

$$v^tHv = v^t\lambda v = \lambda v^t v = \lambda\|v\|^2 > 0,$$

also muss $\lambda > 0$ sein, da die Länge $\|v\|$ immer positiv ist.

Analog gilt für negativ definite Matrizen, dass alle Eigenwerte negativ sein müssen.

DEFINITION 9.1
Eine $n \times n$-Matrix $A = (a_{ij})_{i,j=1}^n$ hat n Hauptminoren A_k, $k = 1, \ldots, n$.
Ein *Hauptminor* A_k ist definiert als $A_k = (a_{ij})_{i,j=1}^k$.

THEOREM 9.4 (KRITERIUM VON SYLVESTER)
Eine symmetrische Matrix ist

- positiv definit, wenn alle ihre Hauptminoren positiv sind.
- negativ definit, wenn ihre Hauptminoren alternierend negativ und positiv sind, d. h., $A_1 < 0$, $A_2 > 0$, $A_3 < 0$, $A_4 > 0$, $\ldots$, $A_{2k} > 0$, $A_{2k+1} < 0$, $\ldots$

BEISPIEL 9.2
Suche die lokalen Minima der Funktion $f(x_1, x_2, x_3) = x_1^3 + 2x_1^2x_2 + x_1x_2x_3 + x_2^2 - 2x_2 + x_2x_3^2$.
Dann ist

$$\nabla f(\begin{pmatrix} x_1, \\ x_2 \\ x_3 \end{pmatrix}) = \begin{pmatrix} 3x_1^2 + 4x_1x_2 + x_2x_3 \\ 2x_1^2 + x_1x_3 + 2x_2 - 2 + x_3^2 \\ x_1x_2 + 2x_2x_3 \end{pmatrix}$$

Setzen wir $\nabla f(x_1, x_2, x_3) = 0$, so erhalten wir:

$$\begin{array}{rrcl} \text{I}: & 3x_1^2 + 4x_1x_2 + x_2x_3 & = & 0 \\ \text{II}: & 2x_1^2 + x_1x_3 + 2x_2 - 2 + x_3^2 & = & 0 \\ \text{III}: & x_1x_2 + 2x_2x_3 & = & 0 \end{array}$$

Aus III : $x_2(x_1 + 2x_3) = 0$ folgt nun, dass entweder (1) $x_2 = 0$ oder (2) $x_1 = -2x_3$.

Fall(1) $x_2 = 0$: Durch Einsetzen folgt aus I, dass $3x_1^2 = 0$ und daher $x_1 = 0$.
Andererseits folgt dann aus II, dass $-2 + x_3^2 = 0$ und damit $x_3 = \pm\sqrt{2}$.
Wir erhalten nun zwei Kandidaten für Extrema:

$$x^{(1)} = \begin{pmatrix} 0 \\ 0 \\ +\sqrt{2} \end{pmatrix}, \qquad x^{(2)} = \begin{pmatrix} 0 \\ 0 \\ -\sqrt{2} \end{pmatrix}$$

Fall(2) $x_1 = -2x_3$: Durch Einsetzen in I und die einfache Umformung: $12x_3^2 - 7x_2x_3 = x_3(12x_3 - 7x_2) = 0$ erhalten wir die zwei Unterfälle (2a) $x_3 = 0$ bzw. (2b) $x_3 = \frac{7}{12}x_2$.

Fall(2a) $x_3 = 0$: Durch Einsetzen erhalten wir $x_1 = 0$ und $x_2 = 1$, also einen neuen Kandidaten für ein Extremum

$$x^{(3)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Fall(2b) $x_3 = \frac{7}{12}x_2$:
Hier folgt, dass $x_1 = -\frac{7}{12}x_2$, und durch Einsetzen in II, dass $\frac{343}{144}x_2^2 + 2x_2 - 2 = 0$.

Diese quadratische Gleichung hat zwei Lösungen, die annähernd folgende Werte annehmen: -1.428 und $+0.588$. Daraus ergeben sich die beiden letzten Kandidaten (mit ungefähren Koordinaten):

$$x^{(4)} = \begin{pmatrix} 1.666 \\ -1.428 \\ -0.833 \end{pmatrix}, \qquad x^{(5)} = \begin{pmatrix} -0.686 \\ 0.588 \\ 0.343 \end{pmatrix}$$

Nun gilt es zu prüfen, ob die obigen Punkte Minima sind. Dazu berechnen wir die Hesse'sche Matrix der zweiten partiellen Ableitungen:

$$H(x) = \nabla^2(x) = \begin{pmatrix} 6x_1 + 4x_2 & 4x_1 + x_3 & x_2 \\ 4x_1 + x_3 & 2 & x_1 + 2x_3 \\ x_2 & x_1 + 2x_3 & 2x_2 \end{pmatrix}$$

Wir setzen nun die Punkte ein, also:

$H(x^{(1)}) = \begin{pmatrix} 0 & \sqrt{2} & 0 \\ \sqrt{2} & 2 & 2\sqrt{2} \\ 0 & 2\sqrt{2} & 0 \end{pmatrix}$, diese Matrix ist indefinit! (0,-,...)

$H(x^{(2)}) = \begin{pmatrix} 0 & -\sqrt{2} & 0 \\ -\sqrt{2} & 2 & -2\sqrt{2} \\ 0 & -2\sqrt{2} & 0 \end{pmatrix}$, ist indefinit! (0,-,...)

$H(x^{(3)}) = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$, ist positiv definit (+,+,+), also ist $x^{(3)}$ lokal minimal.

$H(x^{(4)}) = \begin{pmatrix} 1.121 & 5.831 & -1.428 \\ 5.831 & 2 & 0 \\ -1.428 & 0 & -2.856 \end{pmatrix}$, ist indefinit! (+,-,...)

$H(x^{(5)}) = \begin{pmatrix} -1.764 & -2.401 & 0.588 \\ -2.401 & 2 & 0 \\ 0.588 & 0 & 1.176 \end{pmatrix}$, ist indefinit! (-,+,+)

Das einzige lokale Minimum ist daher $x^{(3)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ mit $f(x^{(3)}) = f(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}) = -1$.

Achtung: Durch einfache Überlegungen können wir feststellen, dass $x^{(3)}$ zwar lokale, aber keine globale Minimalstelle ist: $f(\begin{pmatrix} \lambda \\ 0 \\ 0 \end{pmatrix}) = \lambda^3 \to -\infty, \lambda \to -\infty$.

9.2 Einfache numerische Optimierungsverfahren

Wir können unser Wissen über Gradienten und den Taylor'schen Lehrsatz auch für iterative Näherungsverfahren zur Bestimmung von Extremwerten verwenden:

Methode des steilsten Abstiegs (steepest descent)

Algorithmus:

(1) Wähle ein $x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1n} \end{pmatrix}$ als Startpunkt.
(2) Für $k = 1, 2, 3, \ldots$ berechne die Richtung $d_k := -(\nabla f(x_k))^t$
Minimiere $\varphi(\lambda) = f(x_k + \lambda d_k)$ durch „eindimensionale Optimierung nach λ."
Wir erhalten einen optimalen Wert λ^*!
Als nächsten Punkt wählen wir $x_{k+1} = x_k + \lambda^* d_k$.

Zunächst gehen wir wie Wasser vor: Wir wählen die Richtung des steilsten Abstiegs. Diese Richtung behalten wir so lange bei, bis die Funktion f (diese entspricht der „Höhe") wieder anzusteigen beginnt. Dort sehen wir uns erneut nach dem steilsten Abstieg um (Gradient und Höhenschichtlinie sind dort ja immer orthogonal zueinander).

Warum können wir es nicht wirklich genauso wie das Wasser machen? Wir müssten unendlich viele Neuberechnungen der Richtung machen (die Schrittweite müsste infinitesimal klein werden) und wir würden dadurch eine unendliche Rechenzeit erreichen.

Unser Optimierungsverfahren ist hingegen schon recht effizient: Anfangs machen wir große Schritte, erst wenn wir uns dem Optimum nähern, werden die Schritte kleiner.

Das *Steepest descent*-Verfahren verläuft immer in aufeinander folgenden, orthogonalen Richtungen (siehe Abbildung 9).

BEISPIEL 9.3
Sei für folgende Funktion $f(x, y) = x^2 + y^2 - 2x - 6y$ ein Minimum zu bestimmen.
Wir wählen zunächst $x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ als Startpunkt.
Wir berechnen $(\nabla f(x))^t = \begin{pmatrix} 2x - 2 \\ 2y - 6 \end{pmatrix}$. Dann ist $d_1 = -(\nabla f(x_1))^t = -\begin{pmatrix} -2 \\ -6 \end{pmatrix}$ und wir erhalten: $\varphi(\lambda) = f(x_1 + \lambda d_1) = f(\begin{pmatrix} 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ 6 \end{pmatrix}) = 4\lambda^2 + 36\lambda^2 - 4\lambda - 36\lambda = 40\lambda^2 - 40\lambda$
Abgeleitet nach λ erhalten wir $\varphi'(\lambda) = 80\lambda - 40$. Die Funktion $\varphi(\lambda)$ hat also ein Minimum in $\lambda^* = \frac{1}{2}$.
Damit ergibt sich der nächste Punkt als $x_2 = x_1 + \lambda^* d_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$.
Dieser ist auch schon das Minimum.

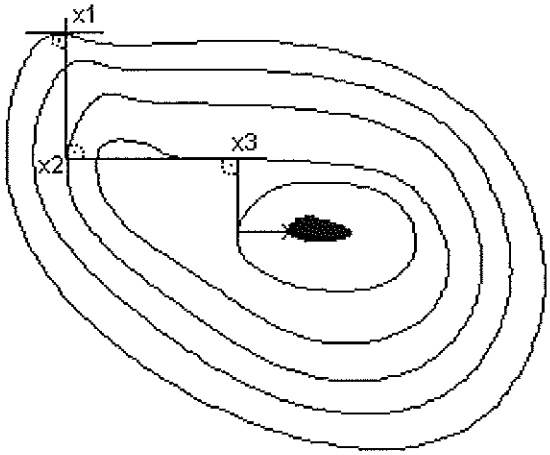


Abbildung 9: Steepest Descent.

Das mehrdimensionale Newtonverfahren

Wir approximieren die Funktion $f(x)$ an der Stelle x_0 durch eine quadratische Funktion

$$h(x) = a + b^t x + \frac{1}{2} x^t G x \,.$$

Es muss dann gelten:

$$\begin{array}{rcccl} f(x_1) & = & h(x_1) & = & a + b^t x_1 + \frac{1}{2} x_1^t G x_1 \\ \nabla f(x_1) & = & \nabla h(x_1) & = & b^t + x_1^t G \\ \nabla^2 f(x_1) & = & \nabla^2 h(x_1) & = & G \end{array}$$

Wir erhalten daraus: $b^t = \nabla f(x_1) - x_1^t \nabla^2 f(x_1)$.

Wenn wir nun das Minimum x^* von $h(x)$ berechnen, erhalten wir eine Näherung für das Minimum von f.

Wir setzten wieder $\nabla h(x) = 0$ und erhalten:
$b^t + x^{*t} G = 0$ und daraus $x^* = x_1 - (\nabla^2 f(x_1))^{-1} (\nabla f(x_1))^t$.

Daraus leitet sich der Algorithmus von Newton und Raphson für den mehrdimensionalen Fall ab.

Algorithmus nach Newton–Raphson:

(1) Wähle ein $x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1n} \end{pmatrix}$ als Startpunkt.
(2) Wiederhole $x_{k+1} = x_k - (\nabla^2(x_k))^{-1} (\nabla f(x_k))^t$ bis die gewünschte Genauigkeit erreicht ist.

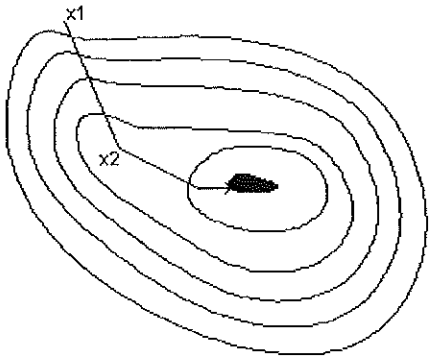


Abbildung 10: Newton-Raphson.

Vorteil: Wir gehen gleich in eine „gute Richtung" (bei quadratischer Zielfunktion genau in Richtung des Minimums), statt einen Zick-Zack-Kurs zu steuern.

Nachteile:

- Die zweiten Ableitungen müssen bekannt sein.
- Die Hesse'sche Matrix muss immer positiv-definit sein (entspricht strikter Konvexität)
- Der Funktionswert kann zwischendurch sogar ansteigen (das ist bei *steepest descent* nie der Fall), dies können wir aber durch Kombination mit dem *line search* Verfahren vermeiden.
- Die Inversion der Hesse'schen Matrix kann numerisch aufwändig sein.

Das Newton-Raphson Verfahren kann also nur dann angewendet werden, wenn die zweiten Ableitungen bekannt bzw. berechenbar sind.

9.3 Optimierung mit Nebenbedingungen

Zusätzlich zur Zielfunktion f müssen wir noch Restriktionen für die zulässigen Werte betrachten, d. h., es sind nur Werte aus einer Teilmenge des $\mathbb{R}^n$ zur Konkurrenz zugelassen ($x \in S \subset \mathbb{R}^n$).

Diese Restriktionen könnten in unterschiedlicher Weise beschrieben werden. Im Folgenden gehen wir davon aus, dass unsere Restriktionen mit Hilfe von Gleichungen und Ungleichungen beschrieben werden können.

Unser Optimierungsproblem lässt sich also folgendermaßen formulieren:

$$\begin{array}{rcl} f(x) & \rightarrow & min! \\ g_1(x) & \leq & 0 \\ & \vdots & \\ g_m(x) & \leq & 0 \\ h_1(x) & = & 0 \\ & \vdots & \\ h_r(x) & = & 0 \end{array}$$

Wir haben hier ein Minimierungsproblem mit m Ungleichheits- und r Gleichheitsnebenbedingungen beschrieben, wobei die Funktionen $g_i(x)$, $i = 1, \ldots, m$ und $h_j(x)$, $j = 1, \ldots, r$ im Prinzip beliebige Funktionen sein können.

BEISPIEL 9.4
Es sei ein Wartehäuschen mit folgenden Vorgaben zu konstruieren: Die Rückwand darf höchstens 3m lang sein. Die Gesamtlänge der Wände (Rück- und Seitenwände) soll gleich 4m sein. Die überdachte Fläche ist zu maximieren.

Als mathematische Formulierung erhalten wir dann daraus (x_1 ist die Länge der Hinterwand, x_2 ist die Länge der Seitenwände):

$$\begin{array}{rcll} f(x_1, x_2) & = & x_1 x_2 & \rightarrow \text{ max!} \\ g(x_1, x_2) & = & x_1 & \leq 3 \\ h(x_1, x_2) & = & x_1 + 2x_2 & = 4 \end{array}$$

Ein wichtiger Spezialfall: Nur Gleichheitsnebenbedingungen

THEOREM 9.5
Es sei $f(x_1, \ldots, x_n) : \mathbb{R}^n \rightarrow \mathbb{R}$ stetig differenzierbar.
Der zulässige Bereich S sei durch Gleichungen beschrieben

$$S = \{x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid g_1(x) = 0, \ldots, g_m(x) = 0\},$$

die Funktionen g_i seien stetig differenzierbar und die Matrix der partiellen Ableitungen der g_i ($m \times n$-Matrix) habe den Rang m, $m < n$.

Dann ist für ein lokales Extremum von $f(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix})$ mit $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in S$ notwendig, dass $\nabla f(x) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x)$, mit eindeutig bestimmten Koeffizienten λ_i (diese heissen Lagrange'sche Multiplikatoren).

Wir gehen so vor, dass wir die **Lagrange-Funktion**

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) = f(x_1, \ldots, x_n) - \sum_{i=1}^{m} \lambda_i g_i(x_1, \ldots, x_n)$$

bilden und sowohl nach $x_1, \ldots, x_n$ als auch nach $\lambda_1, \ldots, \lambda_m$ ableiten und die Ableitungen gleich 0 setzen.

Danach müssen wir das entstandene (im Allgemeinen nichtlineare) Gleichungssystem in $n + m$ Variablen zu lösen.

BEISPIEL 9.5

$$\begin{array}{lcll} f(x_1, x_2) & = & x_1 x_2 & \rightarrow \text{ max!} \\ g(x_1, x_2) & = & x_1^2 + x_2^2 - 1 & = \ 0 \end{array}$$

Die Lagrange-Funktion ist dann:

$$L(x_1, x_2, \lambda) = x_1 x_2 - \lambda(x_1^2 + x_2^2 - 1)$$

Wir leiten nach allen Variablen und λ ab und setzen gleich 0:

$$\begin{array}{lcll} \text{I}: & \frac{\partial f}{\partial x_1} & = & x_2 - 2\lambda x_1 \ = \ 0 \\ \text{II}: & \frac{\partial f}{\partial x_2} & = & x_1 - 2\lambda x_2 \ = \ 0 \\ \text{III}: & \frac{\partial f}{\partial \lambda} & = & -(x_1^2 + x_2^2 - 1) \ = \ 0 \end{array}$$

Falls $x_1 = 0$, dann folgt aus I, dass auch $x_2 = 0$; das steht aber im Widerspruch zu III. Wir schließen daraus, dass $x_1 \neq 0$.

Ähnliches gilt auch für x_2, es muss also auch $x_2 \neq 0$ sein.

Falls aber nun $x_1 \neq 0$, dann erhalten wir durch Umformung aus I, dass $2\lambda = \frac{x_2}{x_1}$. Aus II folgt dann (wegen $x_2 \neq 0$), dass $2\lambda = \frac{x_1}{x_2}$. Daraus schließen wir, dass $x_1 = \pm x_2$. Eingesetzt in III, erhalten wir dann $x_1 = \pm\frac{\sqrt{2}}{2} = \pm\frac{1}{\sqrt{2}}$. Also erhalten wir schließlich 4 Kandidaten für lokale Extremwerte:

$$x^{(1)} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, x^{(2)} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}, x^{(3)} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, x^{(4)} = \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{pmatrix}$$

mit Funktionswerten $\frac{1}{2}$ bzw. $-\frac{1}{2}$.

Wenn wir nicht nur Gleichheits- sondern auch Ungleichheitsnebenbedingungen haben, dann gibt es ein ähnliches, aber etwas komplizierteres Verfahren – das **Karush-Kuhn-Tucker-Verfahren** – mit dem auch solche komplexeren Fälle behandelbar sind.

10 Mehrdimensionale Integration

Wir empfehlen allen, sich das Integrieren von reellen Funktionen $f : \mathbb{R} \to \mathbb{R}$ in ihren Mathematik-Schulbüchern anzusehen und sich die Integrationsregeln in Erinnerung zu rufen. Das mehrdimensionale Integrieren basiert auf dem eindimensionalen Integrieren, oftmals ist es nur das mehrfache Hintereinander-Ausführen eindimensionaler Integrationsmethoden. Das Integrieren von einfachen Polynomfunktionen, trigonometrischen Funktionen, rationalen Funkrionen, Wurzelfunktionen und der Exponentialfunktion und der Logarithmusfunktion sollten sie beherrschen. Die Substitutionsmethode und das partielle Integrieren sollte sie kennen und können, bzw., ihr Wissen dahingehend auffrischen.

10.1 Das Riemann-Integral im $\mathbb{R}^n$

Für eindimensionale reelle Funktionen, d. h. Funktionen $f : \mathbb{R} \to \mathbb{R}$, wird, um das (bestimmte) Integral über einem Intervall I zu berechnen, i.e. $\int_I f(x)\,dx$, das Intervall in Teilintervalle geteilt. Die darauf zu integrierende Funktion wird durch Treppenfunktionen angenähert und das Integral wird durch Bildung eines Grenzwertes berechnet (Riemannintegral).

Die Grundlage des Integrals bildet die *Volums-* oder *Inhaltsmessung* (das *Maß* ist die Länge des Intervalls, für Intervalle in $\mathbb{R}$).

Alle *wichtigen* Mengen in $\mathbb{R}$ sind Intervalle oder Vereinigungen bzw. Durchschnitte von Intervallen.

Im Mehrdimensionalen, i.e. in $\mathbb{R}^n$, entsprechen den Intervallen (verallgemeinerte) n-dimensionale Quader.

DEFINITION 10.1
Die Menge $Q = [a, b) \subset \mathbb{R}^n$ heißt *halboffener Quader* im $\mathbb{R}^n$, wobei $a = (a_1, \ldots, a_n)$ und $b = (b_1, \ldots, b_n)$, $a_i < b_i$ für alle i,.

$$Q = \{x \mid a_1 \leq x_1 < b_1, \ldots, a_n \leq x_n < b_n\}.$$

Q kann auch als kartesisches Produkt von Intervallen geschrieben werden

$$Q = [a, b) = [a_1, b_1) \times [a_2, b_2) \times \cdots \times [a_n, b_n).$$

DEFINITION 10.2
Der *Inhalt* oder das *Volumen* eines Quaders $Q = [a, b)$ ist definiert als

$$I(Q) = \prod_{i=1}^{n} (b_i - a_i)\ .$$

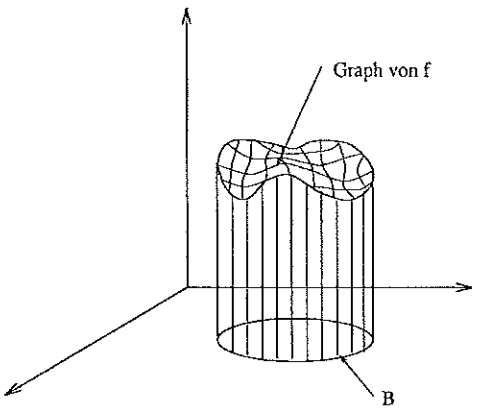
Im $\mathbb{R}^n$ sind Quader aber kein ausreichendes Mengensystem. Zum Beispiel kann eine Kreisscheibe nicht als endliche Vereinigung oder als endlicher Durchschnitt von Quadern dargestellt werden.

Im eindimensionalen „Raum“ $\mathbb{R}$ werden Funktionen $f : [a, b) \to \mathbb{R}$ durch Treppenfunktionen angenähert (von *oben* und von *unten*) und die *Fläche unter der Funktion auf dem Intervall* $[a, b)$ durch Aufsummieren der Fläche der entstehenden Rechtecke approximiert. Im Limes (die Teilintervalle der Treppenfunktion werden immer kleiner) erhalten wir dann das *Integral* (die *echte Fläche*), falls die Limiten von *Obersumme* und *Untersumme* übereinstimmen.

Im $\mathbb{R}^n$ gehen wir für Funktionen $f : B \to \mathbb{R}$, $B \subset \mathbb{R}^n$, ähnlich vor, nur müssen wir hier auch im Definitionsbereich sicherstellen, dass die Menge B auch einen Inhalt besitzt. Es sind nun aber auch im $\mathbb{R}^n$ alle *wichtigen* Mengen als Vereinigung oder Durchschnitte von Quadern darstellbar, allerdings brauchen wir meist *unendlich* viele Vereinigungen oder Durchschnitte.

BEISPIEL 10.1
Sei $f(x, y) : \mathbb{R}^2 \to \mathbb{R}$ gegeben.
Das Volumen der Menge $M = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \mid \begin{pmatrix} x \\ y \end{pmatrix} \in B \text{ und } 0 \leq z \leq f(x, y) \}$ wird das Integral von f über dem Bereich B genannt

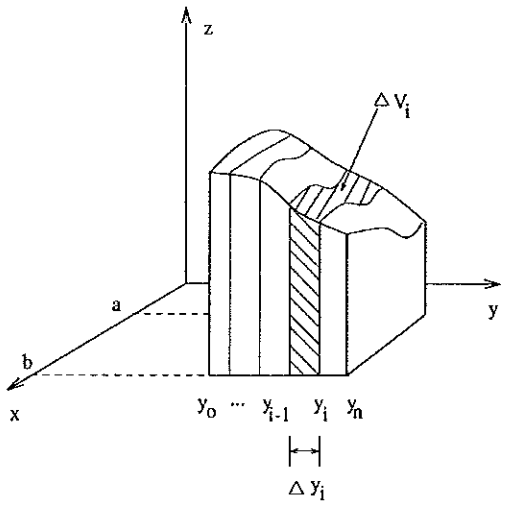
$$V = \iint_B f(x, y)\, dx\, dy\,.$$



Falls B ein Rechteck ist,

$$B = \{ \begin{pmatrix} x \\ y \end{pmatrix} \mid a \leq x \leq b \text{ und } c \leq y \leq d \}\,,$$

ist das folgendermaßen zu berechnen.



Eine Scheibe hat ungefähr das Volumen: $\Delta V_i \approx \Delta y_i \int_a^b f(x, y_i)\, dx$
Alle Scheiben zusammen ergeben: $V \approx \sum_{i=1}^{n} (\int_a^b f(x, y_i) dx)\, \Delta y_i$
Für $n \to \infty$ und $\max(\Delta y_i) \to 0$ gilt

$$V = \lim_{n\to\infty} \sum_{i=1}^{n} [\int_a^b f(x, y_i) dx] \Delta y_i = \int_c^d \int_a^b f(x, y_i)\, dx dy$$

THEOREM 10.1 (SATZ VON FUBINI)
Sei $f : \mathbb{R}^2 \to \mathbb{R}$ stetig und beschränkt, dann gilt:

$$\int_c^d \int_a^b f(x, y)\, dx dy = \int_a^b \int_c^d f(x, y)\, dy dx\,.$$

D. h., die Integrationsreihenfolge kann vertauscht werden. Der Satz von Fubini kann auch auf höhere Dimensionen verallgemeinert werden.

Integration über Normalbereiche

THEOREM 10.2
Sei ein Bereich folgendermaßen gegeben:

$$B = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \;\middle|\; \begin{array}{rcccl} a & \le & x_1 & \le & b \\ f_1(x_1) & \le & x_2 & \le & g_1(x_1) \\ f(x_1, x_2) & \le & x_3 & \le & g_2(x_1, x_2) \\ & & \vdots & & \\ f_{m-1}(x_1, \ldots, x_{m-1}) & \le & x_m & \le & g_{m-1}(x_1, \ldots, x_{m-1}) \end{array} \right\},$$

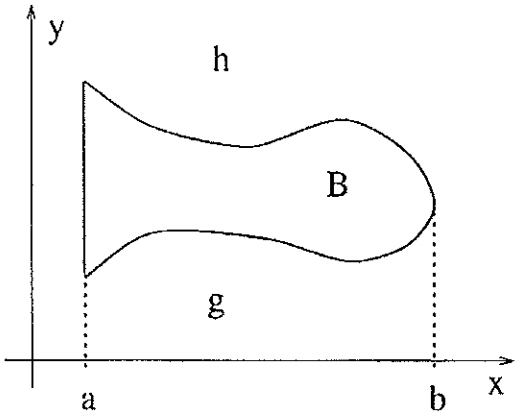
die Funktionen $f_1, \ldots, f_m$ und $g_1, \ldots, g_m$ stetig, dann gilt:

$$\int_B f(x_1, \ldots, x_m) d(x_1, \ldots, x_m) = \int_a^b \int_{f(x_1)}^{g(x_1)} \cdots \int_{f_{m-1}(x_1,\ldots,x_{m-1})}^{g_{m-1}(x_1,\ldots,x_{m-1})} f(x_1, \ldots, x_m) dx_m\, dx_{m-1} \cdots dx_2\, dx_1$$

BEISPIEL 10.2
Seien $g : [a, b] \to \mathbb{R}$ und $h : [a, b] \to \mathbb{R}$ stetig und $g \leq h$,

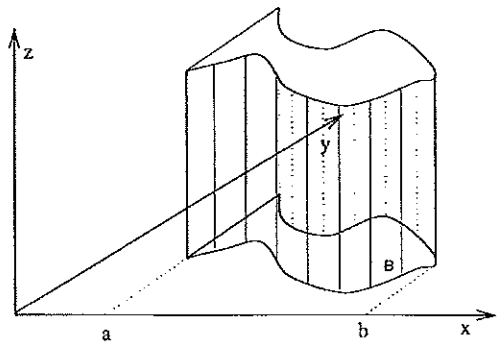
$$B = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \;\middle|\; \begin{array}{rcccl} a & \leq & x & \leq & b \\ g(x) & \leq & y & \leq & h(x) \end{array} \right\},$$

dann ist B ein Normalbereich.

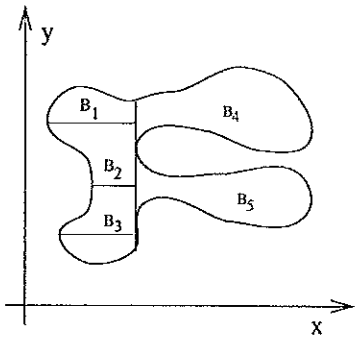


Für das Integral einer Funktion f auf diesem Bereich gilt dann:

$$\iint_B f(x,y)dxdy = \int_a^b \int_{g(x)}^{h(x)} f(x,y)\; dy\, dx\,.$$



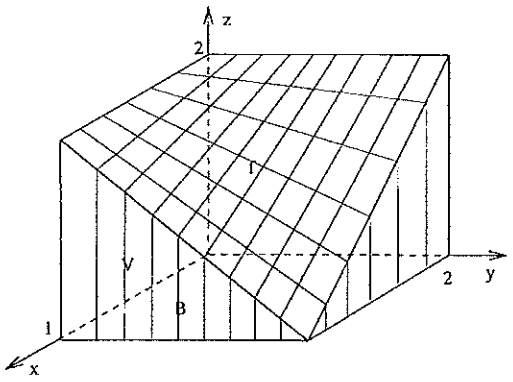
Es gibt auch kompliziertere Mengen (Nicht-Normalbereiche). Diese müssen in kleinere Teilbereiche zerlegt werden, und dann kann auf den einfacheren Bereichen ein Integral berechnet werden.



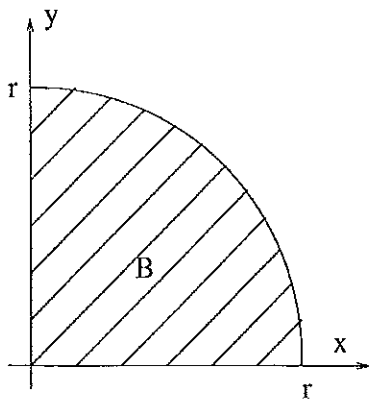
BEISPIEL 10.3
Als konkretes Beispiel betrachten wir die Funktion $f(x,y) = 2 - xy$ auf dem Bereich $B = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid 0 \le x \le 1, 0 \le y \le 2 \right\}$.

Es gilt dann:

$$\iint_B (2-xy)\,dxdy = \int_0^2 \int_0^1 (2-xy)\,dxdy = \int_0^2 [2x - \frac{x^2 y}{2}]_0^1\,dy = \int_0^2 (2 - \frac{y}{2})\,dy = [2y - \frac{y^2}{4}]_0^2 = 3$$



BEISPIEL 10.4
Es soll $\iint_B x^3 y^2\,dxdy$ berechnet werden, wobei $B = \{ \begin{pmatrix} x \\ y \end{pmatrix} \mid x^2 + y^2 \le r^2,\ x \ge 0,\ y \ge 0 \}$ der unten skizzierte Viertelkreis ist.



B ist ein Normalbereich, denn es gilt:

$$B = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid 0 \leq x \leq r,\ 0 \leq y \leq \sqrt{r^2 - x^2}\}\,.$$

Das obige Integral ist also:

$$\int_0^r (\int_0^{\sqrt{r^2-x^2}} x^3 y^2\, dy)\, dx = \frac{1}{3}\int_0^r x^3 (r^2 - x^2)^{\frac{3}{2}}\, dx\,.$$

Mittels partieller Integration ($u = x^2, v' = x(r^2 - x^2)^{\frac{3}{2}}$) erhalten wir $\frac{2}{105}r^7$ als Ergebnis.

Wir können B auch anders anschreiben und die Reihenfolge der Variablen vertauschen:

$$B = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid 0 \leq y \leq r, 0 \leq x \leq \sqrt{r^2 - y^2}\}\,.$$

Das Integral ist dann

$$\int_0^r (\int_0^{\sqrt{r^2-y^2}} x^3 y^2 dx)\, dy = \frac{1}{4}\int_0^r y^2 (r^2 - y^2)^2\, dy = \frac{2}{105} r^7\,.$$

Wie wir sehen ist dieses Integral etwas einfacher zu berechnen. Durch Vertauschen der Integrationsreihenfolge können wir eventuell einigen Rechenaufwand einsparen.

10.2 Koordinatentransformation in Integralen

Aus der eindimensionalen Integralrechnung kennen wir die **Substitutionsregel**:

Sei $g : [c, d] \to [a, b]$ differenzierbar und streng monoton, dann ist g bijektiv, also invertierbar. Sei die Funktion $f : [a, b] \to \mathbb{R}$ auf dem Bildbereich von g definiert.

Dann gilt

$$\int_a^b f(x)\, dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(t)) g'(t)\, dt$$

oder äquivalent dazu

$$\int_c^d f(g(t))g'(t)dt = \int_{g(c)}^{g(d)} f(x)dx\,.$$

Eine analoge Regel dazu gibt es auch im Mehrdimensionalen.

THEOREM 10.3 (TRANSFORMATIONSSATZ)
Sei g eine umkehrbare differenzierbare Funktion von A nach B ($A, B \subset \mathbb{R}^n$).
Weiters sei f eine Funktion $f : B \to \mathbb{R}$.

Dann gilt: Falls eines der beiden Integrale

$$\int_B f(x_1, \ldots, x_n)\, d(x_1, \ldots, x_n)$$

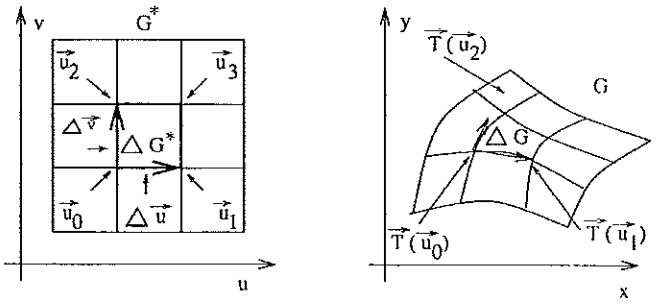
oder

$$\int_A f(g_1(t_1, \ldots, t_n), \ldots, g_n(t_1, \ldots, t_n)) \cdot \det(J_g(t_1, \ldots, t_n))\, d(t_1, \ldots, t_n)$$

existiert und die Determinante der Jacobi-Matrix $\det(J_g(t_1, \ldots, t_n))$ der ersten partiellen Ableitungen von g ungleich 0 ist, so existiert auch das zweite Integral und die beiden sind gleich.

Also kurz geschrieben:

$$\int_B f(x_1, \ldots, x_n)\, d(x_1, \ldots, x_n) = \int_{g^{-1}(B)} f(g(t)) \cdot \det(J_g(t))\, dt$$

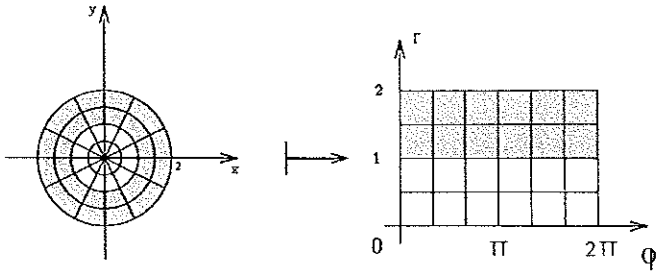


BEISPIEL 10.5 (POLARKOORDINATENTRANSFORMATION)
Manchmal ist es einfacher, statt in kartesischen Koordinaten in Polarkoordinaten zu rechnen. Sei etwa folgendes Integral zu berechnen:

$$\int_N xy\; dxdy \quad \text{mit} \quad N = \{(x, y) \in \mathbb{R}^2 | x \geq 0, y \geq 0, 1 \leq x^2 + y^2 \leq 4\}\,,$$

dann können wir uns folgende *natürliche* Transformation überlegen:

$$g : (r, \theta) \mapsto (r\cos\theta, r\sin\theta) \quad \text{mit} \quad x = r\cos\theta,\ y = r\sin\theta, \quad \text{also} \quad g(r, \theta) = (x, y)\,.$$



Dann ist $M = \{(r,\theta)|1 \leq r \leq 2, 0 \leq \theta \leq \frac{\pi}{2}\}$ und obiges Integral kann folgendermaßen umgeformt werden:

$$\int_N xy\ dxdy = \int_M r\cos\theta \cdot r\sin\theta \cdot \det(J_g(r,\theta))\ drd\theta\,.$$

Da $J_g(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$, ist $\det(J_g(r,\theta)) = r\cdot(\cos^2\theta + \sin^2\theta) = r$.

Also erhalten wir:

$$\begin{aligned}
\textstyle\int_N xy\ dxdy &= \textstyle\int_1^2 \int_0^{\pi/2} r\cos\theta \cdot r\sin\theta \cdot r\ d\theta\, dr \\
&= \textstyle\int_1^2 \int_0^{\pi/2} r^3\cos\theta\cdot\sin\theta\ d\theta\, dr \\
&= \textstyle\int_0^{\pi/2} [\frac{r^4}{4}\cos\theta\cdot\sin\theta]_{r=1}^2\ d\theta \\
&= \textstyle\frac{15}{4}\int_0^{\pi/2}\cos\theta\cdot\sin\theta\ d\theta \\
&= \textstyle\frac{15}{4}[\frac{\sin^2\theta}{2}]_{\theta=0}^{\pi/2} \\
&= \textstyle\frac{15}{8}
\end{aligned}$$

11 Differentialgleichungen

Differentialgleichungen sind Gleichungen, die mindestens einen Differentialquotienten enthalten. Sie heißen *von der Ordnung* n, wenn der höchste vorkommende Differentialquotient von der Ordnung n ist (d. h., die höchste vorkommende Ableitung der gesuchten Funktion eine n-te ist).

Ein Beispiel einer allgemeine Differentialgleichung vierter Ordnung wäre etwa

$$y^{(4)}(x) - 3y'''(x) + 3x^2y''(x) - \log(x)(y'(x))^3 - y^2(x) = 3\cos(x)\,.$$

Wir unterscheiden weiters zwischen allgemeinen Differentialgleichungen und sogenannten *Anfangswertproblemen*, bei denen Startwerte der beteiligten Funktionen bekannt sind. Weiters gibt es auch noch sogenannte *Randwertprobleme*, die Differentialgleichungen auf (nicht notwendigerweise endlichen) Intervallen betreffen, wobei Randwerte, d. h., Funktionswerte oder Werte der Ableitungen am Rand als Nebenbedingungen auftreten.

Das Lösen von Differentialgleichungen – das Integrieren – ist mindestens so schwer, wie die Stammfunktion eines Integrals zu finden. Wir werden uns hier auf einfache aber dennoch wichtige Spezialfälle beschränken.

11.1 Lösung einfacher Differentialgleichungen

11.1.1 $y'(x) = a(x)$ – Einfaches Integrieren

Sei die Differentialgleichung von der Form

$$y'(x) = a(x)\,.$$

Wir haben also

$$\frac{dy}{dx} = a(x) \qquad \Leftrightarrow \qquad dy = a(x)\,dx\,.$$

Durch Integration erhalten wir die Lösung

$$y(x) = \int a(x)\,dx + c, \quad c \in \mathbb{R}\,.$$

Beispiel 11.1

$$y'(x) = 3x^2 \quad \Rightarrow \quad y(x) = \int 3x^2\,dx = x^3 + c\,.$$

Haben wir zusätzlich die *Anfangswertbedingung* $y(0) = 4$, so erhalten wir durch Einsetzen

$$y(0) = 0^3 + c = 4, \quad \text{d. h.} \quad c = 4 \quad \text{also} \quad y(x) = x^3 + 4\,.$$

11.1.2 $y'(x) = a(x)f(y)$ – Trennung der Variablen

Ist die Differentialgleichung von der Form

$$y'(x) = a(x)f(y)\,,$$

so erhalten wir durch Umformen schrittweise

$$\frac{dy}{dx} = a(x)f(y) \qquad \Rightarrow \qquad \frac{dy}{f(y)} = a(x)\,dx\,,$$

und daraus durch Integrieren (und Zusammenfassen beider Integrationskonstanten auf einer Seite) die Lösung

$$\int \frac{1}{f(y)}\,dy = \int a(x)\,dx + c\,.$$

D. h., die Variablen x und y wurden getrennt, d. h., kommen nur auf jeweils einer Seite der Gleichung vor.

BEISPIEL 11.2
Sei folgende Differentialgleichung gegeben

$$y'(x) + xy^2(x) = 0\,.$$

So ist

$$\frac{dy}{dx} = -xy^2 \qquad \Rightarrow \qquad -\frac{1}{y^2}\,dy = x\,dx \qquad \Rightarrow \qquad -\int \frac{1}{y^2}\,dy = \int x\,dx + c\,,$$

also

$$\frac{1}{y} = \frac{x^2}{2} + c_1 = \frac{x^2 + c}{2} \qquad \text{d. h.} \quad y(x) = \frac{2}{x^2 + c}\,.$$

Ist zusätzlich $y(0) = 1$ gefordert, so folgt $c = 2$, also $y(x) = \frac{2}{x^2+2}$.

11.2 Die Logistische Differentialgleichung

DEFINITION 11.1
Die *logistische Differentialgleichung* hat die Form

$$y'(x) - ky(x)(L - y(x)) = 0, \qquad k > 0, \quad 0 \le y(x) \le L,\ L \in \mathbb{R}$$

Diese Differentialgleichung wird auch als Differentialgleichung des *beschränkten exponentiellen Wachstums* bezeichnet. Der Parameter L gibt die Obergrenze an.

Durch Trennung der Variablen erhalten wir:

$$\frac{dy}{dx} = ky(L - y) \qquad \Rightarrow \qquad \frac{dy}{y(L-y)} = k\,dx\,.$$

Wir substituieren auf der linken Seite:

$$z = \frac{y-L}{y} = 1 - \frac{L}{y} \qquad \Rightarrow \qquad dz = \frac{L}{y^2}\,dy$$

Durch Integration erhalten wir $(y \leq L)$:

$$\begin{aligned} \int \frac{dy}{y(L-y)} &= -\frac{1}{L}\int \frac{y}{(y-L)} \cdot \frac{L}{y^2}\,dy = -\frac{1}{L}\int \frac{1}{z}\,dz \\ &= -\frac{1}{L}\log|z| + c = -\frac{1}{L}\log\left|\frac{y-L}{y}\right| + c \\ &= -\frac{1}{L}\log\left(\frac{L-y}{y}\right) + c \end{aligned}$$

Insgesamt ergibt sich somit:

$$-\frac{1}{L}\log\left(\frac{L-y}{y}\right) = kx + c$$

Durch Umformen nach y ergibt sich dann schließlich die allgemeine Lösung der logistischen Differentialgleichung:

$$y(x) = \frac{L}{1 + Ce^{-Lkx}}$$

BEISPIEL 11.3
In einem Ort mit 5100 Einwohnern bricht eine Grippeepidemie aus. Zu dem Zeitpunkt als die Krankheit erkannt wird, sind 100 Bewohner erkrankt; 10 Tage später sind es 1000. Wie lässt sich der Epidemieverlauf beschreiben?

Sei $y(t)$ die Anzahl der betroffenen Einwohner, die bis zum Zeitpunkt t erkrankt sind. Wir nehmen exponentielles Wachstum an mit Obergrenze $L = 5100$.

Die allgemeine Lösung der logistischen Differentialgleichung lautet dann

$$y(x) = \frac{5100}{1 + Ce^{-5100kt}}$$

Die Parameter k und C bestimmen wir folgendermaßen:

$$y(0) = 100 \quad \Rightarrow \quad \frac{5100}{1+C} = 100 \quad \Rightarrow \quad C = 50\,.$$

$$y(10) = 1000 \quad \Rightarrow \quad \frac{5100}{1 + 50\,e^{-5100.10k}} = 1000 \quad \Rightarrow \quad k = 0,00005\,.$$

Insgesamt erhalten wir somit als Lösung:

$$y(x) = \frac{5100}{1 + 50\,e^{-0,25t}}\,.$$

11.3 Lineare Differentialgleichungen

Eine spezielle Form von Differentialgleichungen sind die *linearen Differentialgleichungen* , in denen die Funktion und deren Ableitungen nur in linearer Form enthalten sein dürfen.

$$y^{(n)}(x) + a_{n-1}y^{(n-1)}(x) + \cdots + a_1 y'(x) + a_0 y(x) = s(x),\ a_i \in \mathbb{R}$$

ist die allgemeine Form einer linearen Differentialgleichung n-ter Ordnung.

Die Lösung von linearen Differentialgleichungen

Die Lösung einer linearen Differentialgleichung hat stets die Form

$$y = y_h + y_p$$

wobei y_h die Lösung der homogenen Differentialgleichung (DG) ist, d. h. DG $= 0$, und y_p eine *partikuläre Lösung*, d. h., eine spezielle Lösung ist, für DG $= s(x)$.

In der Physik sprechen wir von y_h auch noch als *zero input response* ($s(x) = 0$) und von y_p als *zero state response* ($y_0 = 0$).

11.3.1 Lineare Differentialgleichungen erster Ordnung

DEFINITION 11.2
Eine *lineare Differentialgleichung erster Ordnung* hat stets die Form

$$y'(x) + a(x)y(x) = s(x)$$

wobei $a(x)$ und $s(x)$ integrierbare Funktionen sein müssen.

Ist $s(x) = 0$ so sprechen wir von einer *homogenen linearen Differentialgleichung erster Ordnung.*

11.3.2 Lösung der homogenen linearen Differentialgleichung erster Ordnung

Wir müssen also zuerst die homogene lineare Differentialgleichung

$$y'(x) + a(x)y(x) = 0$$

lösen.

Wie erreichen dies durch *Trennung der Variablen*:

$$\frac{dy}{dx} = -a(x)y \qquad \Rightarrow \qquad \frac{1}{y}\,dy = -a(x)\,dx \qquad \Rightarrow \qquad \int \frac{1}{y}\,dy = -\int a(x)\,dx\,.$$

Durch Integrieren erhalten wir (wobei log – wie in der Mathematik üblich – den Logarithmus naturalis bezeichnet und wir statt der Integrationskonstanten c die Konstante $\log(C)$ verwendeten)

$$\log(y(x)) = -\int a(x)\,dx + \log(C) \qquad \Rightarrow \qquad y(x) = C \cdot e^{-\int a(x)\,dx}\,,$$

wobei wir rechts die Rechenregeln für Logarithmen anwendeten.
Es ergibt sich also die Lösung

$$y(x) = C \cdot e^{-\int a(x)\,dx}\,.$$

BEISPIEL 11.4
Sei folgende Differentialgleichung gegeben:

$$y' + 3x^2 y = 0 \quad \text{oder} \quad y' = -3x^2 y\,.$$

Wir haben

$$\frac{dy}{dx} = -3x^2 y \qquad \Rightarrow \qquad \frac{1}{y}\,dy = -3x^2\,dx$$

und durch Integration

$$\int \frac{1}{y}\,dy = -\int 3x^2\,dx \qquad \Rightarrow \qquad \log(y) = -x^3 + \log(C)$$

also

$$y(x) = C \cdot e^{-x^3}\,.$$

Ist der Anfangswert $y(0) = 2$ gefordert, so ergibt sich $C = 2$, also $y(x) = 2e^{-x^3}$.

11.3.3 Die partikuläre Lösung

Sei nun wie oben

$$y'(x) + a(x)y(x) = s(x)\,.$$

Wir unterscheiden zwei Fälle:

1. $s(x) = s$ und $a(x) = a \neq 0$ sind (reelle) Konstante, also $y'(x) + ay(x) = s$.
   Dann setzen wir
   $$y_p = \frac{s}{a}\,,$$
   denn dann ist $y_p' + a \cdot \frac{s}{a} = s$ und $y_p' = (\frac{s}{a})' = 0$ (Ableitung einer Konstanten), also $y_p'(x) = 0$.

2. $s(x)$ oder $a(x)$ nicht konstant: Lösung durch *Variation der Konstanten:*
   Wir ersetzen in der Lösung der homogenen Differentialgleichung die Konstante C durch eine Funktion $C(x)$ und setzen so an:
   $$y_p(x) = C(x) \cdot e^{-\int a(x)\,dx}\,.$$
   Durch Einsetzen in
   $$y_p' + a(x)y_p = s(x)$$
   erhalten wir
   $$e^{-\int a(x)\,dx}(C'(x) - a(x)C(x)) + a(x)C(x)e^{-\int a(x)\,dx} = s(x)\,,$$
   d. h.
   $$e^{-\int a(x)\,dx}C'(x) = s(x) \qquad \text{oder} \qquad C'(x) = s(x) \cdot e^{\int a(x)\,dx}\,,$$
   somit
   $$C(x) = \int s(x) \cdot e^{\int a(x)\,dx}\,dx\,.$$

Hier muss das Integral nach x im Exponenten zuerst berechnet werden, d. h., bevor das äußere Integral nach x aufgelöst wird. Wir erhalten dann die allgemeine Lösung $y = y_h + y_p$

$$y(x) = e^{-\int a(x)\,dx} \cdot \left(C + \int s(x) \cdot e^{\int a(x)\,dx}\,dx \right) .$$

BEISPIEL 11.5
Sei die inhomogene lineare Differentialgleichung $y' + \frac{1}{x}y = x^2 + 4$ gegeben.
Wir berechnen zuerst die homogene Lösung y_h und dann die partikuläre Lösung y_p.

Die homogene Lösung y_h : Es ist

$$y' + \frac{1}{x}y = 0 \quad \Leftrightarrow \quad \frac{dy}{dx} = -\frac{1}{x}y \quad \Leftrightarrow \quad \frac{dy}{y} = -\frac{dx}{x}$$

$$\Leftrightarrow \quad \log(y_h) = -\log(x) + \log(C) = \log(\frac{C}{x}) \quad \Leftrightarrow \quad y_h(x) = \frac{C}{x} .$$

Die inhomogene Lösung y_p : Durch Variation der Konstanten:
Wir differenzieren $y_p = \frac{C(x)}{x}$ und erhalten $y_p' = \frac{C'(x)x - C(x)}{x^2}$.
Das in die Differentialgleichung eingesetzt ergibt

$$\frac{C'(x)x - C(x)}{x^2} + \frac{1}{x}\frac{C(x)}{x} = x^2 + 4 \quad \Leftrightarrow \quad C'(x) = x^3 + 4x ,$$

also folgt

$$C(x) = \frac{x^4}{4} + 2x^2$$

und

$$y_p(x) = \frac{C(x)}{x} = \frac{x^3}{4} + 2x .$$

Als Lösung erhalten wir

$$y = y_h + y_p = \frac{C}{x} + \frac{x^3}{4} + 2x .$$

Ist der Anfangswert $y(1) = 1$ gegeben, so folgt

$$y(1) = 1 = \frac{C}{1} + \frac{1}{4} + 2 \quad \Rightarrow \quad C = -\frac{5}{4} \quad \Rightarrow \quad y(x) = \frac{-5}{4x} + \frac{x^3}{4} + 2x .$$

Wollen wir eine Probe berechnen, so berechnen wir y' der allgemeinen Lösung, setzen diese in die Angabe

$$y' + \frac{1}{x}y = x^2 + 4$$

ein und überprüfen, ob die Differentialgleichung erfüllt ist.

11.3.4 Lineare Differentialgleichung zweiter Ordnung

DEFINITION 11.3
Eine *lineare Differentialgleichung zweiter Ordnung* hat stets die Form

$$y''(x) + a_1(x)y'(x) + a_2(x)y(x) = s(x)\,.$$

Ist $s(x) = 0$ so sprechen wir von einer *homogenen linearen Differentialgleichung zweiter Ordnung*.

Wie bei der linearen Differentialgleichung erster Ordnung ergibt sich die allgemeine Lösung als Summe der homogenen und der partikulären Lösung $y(x) = y_h(x) + y_p(x)$.

Da wir uns auf einfache Fälle beschränken müssen, betrachten wir nur den Fall, dass die Funktionen $a_1(x)$, $a_2(x)$ und $s(x)$ konstant sind.

DEFINITION 11.4
Eine *lineare Differentialgleichung zweiter Ordnung mit konstanten Koeffizienten* hat die Form

$$y''(x) + a_1y'(x) + a_2y(x) = s, \qquad a_1, a_2, s \in \mathbb{R}.$$

Ist $s = 0$ so sprechen wir von einer *homogenen linearen Differentialgleichung zweiter Ordnung mit konstanten Koeffizienten.*

11.3.5 Homogene lineare Differentialgleichungen mit konstanten Koeffizienten

Aus unserer Erfahrung mit homogenen linearen Differentialgleichungen erster Ordnung vermuten wir den allgemeinen Ansatz

$$y(x) = C \cdot e^{\lambda x}\,.$$

Wir berechnen die erste und zweite Ableitung und setzen in die Differentialgleichung ein. Aus

$$y'(x) = \lambda C e^{\lambda x} \qquad \text{und} \qquad y''(x) = \lambda^2 C e^{\lambda x}\,,$$

und aus der gegebenen Differentialgleichung folgt

$$Ce^{\lambda x}(\lambda^2 + a_1\lambda + a_2) = 0\,.$$

D. h., unser angenommenes $y(x)$ ist genau dann Lösung der Differentialgleichung

$$y''(x) + a_1y'(x) + a_2y(x) = 0\,,$$

wenn

$$\lambda^2 + a_1\lambda + a_2 = 0$$

ist. Diese Gleichung heißt *charakteristische Gleichung* der Differentialgleichung.

Lösung der charakteristischen Gleichung

Die charakteristische Gleichung einer homogenen linearen Differentialgleichung zweiter Ordnung mit konstanten Koeffizienten ist eine quadratische Gleichung der Form

$$\lambda^2 + a_1\lambda + a_2 = 0$$

und hat demnach die Lösungen

$$\lambda_{1,2} = -\frac{a_1}{2} \pm \sqrt{\frac{a_1^2}{4} - a_2} = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_2}}{2}.$$

Wir unterscheiden drei Fälle, je nach Größe der Diskriminante:

1. $\frac{a_1^2}{4} - a_2 > 0$: Wir haben zwei reelle Lösungen $\lambda_1 \neq \lambda_2$ der charakteristischen Gleichung und demnach zwei (unabhängige) Lösungen der Differentialgleichung

   $$y_1(x) = C_1 e^{\lambda_1 x} \qquad \text{und} \qquad y_2(x) = C_2 e^{\lambda_2 x}$$

   und wegen der Linearität der Differentialgleichung ist auch deren Summe eine Lösung (nachrechnen), sodass die allgemeine Lösung in diesem Fall lautet:

   $$y(x) = C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x}.$$

2. $\frac{a_1^2}{4} - a_2 = 0$: Die charakteristische Gleichung hat die Doppellösung $\lambda_1 = \lambda_2 = \lambda$ und somit ergeben sich die Lösungen (Nachweis durch Einsetzen) der Differentialgleichung zu

   $$y_1(x) = C_1 e^{\lambda x} \qquad \text{und} \qquad y_2(x) = C_2 x e^{\lambda x}.$$

   Wegen der Linearität der Differentialgleichung ist die Summe wiederum Lösung, d. h., die allgemeine Lösung lautet

   $$y(x) = C_1 e^{\lambda x} + C_2 x e^{\lambda x} = (C_1 + C_2 x) e^{\lambda x}.$$

3. $\frac{a_1^2}{4} - a_2 < 0$: Die charakteristische Gleichung hat die beiden komplexen Lösungen

   $$\lambda_{1,2} = -\frac{a_1}{2} \pm i \cdot \sqrt{\left|\frac{a_1^2}{4} - a_2\right|} \doteq a \pm ib.$$

   Die komplexen Lösungen der Differentialgleichung lauten daher:

   $$\begin{aligned} y_1(x) &= C_1' e^{(a+ib)x} &&= C_1' e^{ax} e^{ibx} \\ y_2(x) &= C_2' e^{(a-ib)x} &&= C_2' e^{ax} e^{-ibx} \end{aligned}$$

   oder aber, mit $y(x) = y_1(x) + y_2(x)$,

   $$y(x) = C_1' e^{ax} e^{ibx} + C_2' e^{ax} e^{-ibx} = e^{ax} \cdot (C_1' e^{ibx} + C_2' e^{-ibx}).$$

Wegen der Euler'schen Formel $e^{i\varphi} = cos\varphi + i\sin\varphi$ gilt

$$y(x) = e^{ax} \cdot (C_1' \cos(bx) + C_1' i \sin(bx) + C_2' \cos(bx) - C_2' i \sin(bx))$$

also

$$y(x) = e^{ax} \cdot [(C_1' + C_2') \cos(bx) + (C_1' - C_2') i \sin(bx)] .$$

Mit $C_1 \doteqdot C_1' + C_2'$ und $C_2 \doteqdot (C_1' - C_2')i$ erhalten wir die allgemeine Lösung

$$y(x) = e^{ax} \cdot (C_1 \cos(bx) + C_2 \sin(bx)) ,$$

eine oszillierende Funktion. Die Schwingung ist gedämpft, falls $a < 0$, konstant mit $a = 0$ und sie wächst mit $a > 0$.

Exkurs: $\mathbb{C}$ – Die komplexen Zahlen

In den reellen Zahlen gibt es einfache Gleichungen, die nicht in $\mathbb{R}$ gelöst werden können, etwa $x^2 + 1 = 0$. Diese ist jedoch in den komplexen Zahlen

$$\mathbb{C} = \{a + ib \,|\, a, b \in \mathbb{R},\ i = \sqrt{-1}\}$$

lösbar. Dabei bezeichnet a den Realteil und b den Imaginärteil einer komplexen Zahl $a + ib$. $i = \sqrt{-1}$ ist die imaginäre Einheit, 1 die reelle Einheit (Bem.: in den Ingenieurswissenschaften ist es üblich die komplexe Einheit mit j zu bezeichnen).

Alle Gleichungen in $\mathbb{C}$, d. h. mit Koeffizienten in $\mathbb{C}$, sind in $\mathbb{C}$ lösbar, daher auch alle Gleichungen mit reellen, rationalen, ganzzahligen oder natürlichen Koeffizienten – Fundamentalsatz der Algebra. Das heißt auch, dass jedes Polynom in $\mathbb{C}$ in Linearfaktoren zerlegt werden kann. Weiters ist $(\mathbb{C}, +, \cdot)$ wie $(\mathbb{R}, +, \cdot)$ und $(\mathbb{Q}, +, \cdot)$ ein Körper.

Komplexe Zahlen spielen eine große Rolle in der Elektrotechnik und Elektronik, wo sie zur Berechnung von Strömen notwendig sind, sowie in der Ökologie bei der Modellierung von Ökosystemen.

Rechenregeln in $\mathbb{C}$

Seien $x = a + ib$ und $y = c + id$ komplexe Zahlen $(a, b, c, d \in \mathbb{R})$, so ist

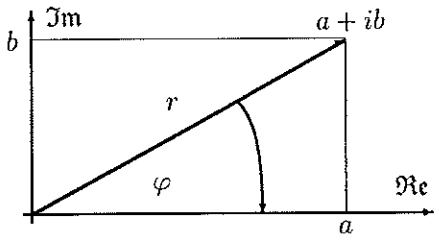
- die Summe $x + y = (a + c) + i(b + d)$
- das Produkt $x \cdot y = (a+ib) \cdot (c+id) = (a \cdot c) + i(b \cdot c + a \cdot d) + i^2(b \cdot d) = (ac - bd) + i(bc + ad)$
- die konjugiert komplexe Zahl zu x ist $\bar{x} = a - ib$
- der Quotient $\frac{x}{y} = \frac{a+ib}{c+id} = \frac{(a+ib)(c-id)}{(c+id)(c-id)} = \frac{ac+bd}{c^2+d^2} + i\frac{bc-ad}{c^2+d^2}$
- der Betrag von x ist $|x| = \sqrt{x \cdot \bar{x}} = \sqrt{a^2 + b^2}$

Es gelten weiters auch alle Rechenregeln von $\mathbb{R}$.

Darstellungen von komplexen Zahlen

Es gibt mehrere Möglichkeiten, komplexe Zahlen darzustellen:

- als $a + ib$ mit $a, b \in \mathbb{R}$, $i = \sqrt{-1}$.
- als Vektoren in der (komplexen) Ebene (siehe Abbildung)
- in der Polardarstellung als geordnetes Paar (r, φ) mit $r = \sqrt{a^2 + b^2}$ und $\tan\varphi = \frac{b}{a}$ (siehe Abbildung). Umgekehrt gilt: $a + ib = r(\cos\varphi + i\sin\varphi)$.
- als reelle 2×2-Matrizen der Form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$



Rechnen in Polarkoordinaten

Die Darstellung der komplexen Zahlen mittels Polarkoordinaten vereinfacht einige Rechnungen. Seien $x = (r, \varphi)$ und $y = (s, \psi)$ komplexe Zahlen, so gilt:

- $x \cdot y = (r \cdot s, \varphi + \psi)$
- $x^n = (r^n, n \cdot \varphi)$
- $\sqrt[n]{x} = (\sqrt[n]{r}, \frac{k \cdot \varphi}{n})$, $k = 1, \ldots, n-1$.

Ende: Exkurs komplexe Zahlen

BEISPIEL 11.6
Die Differentialgleichung

$$y''(x) + y'(x) + y(x) = 0$$

hat die charakteristische Gleichung

$$\lambda^2 + \lambda + 1 = 0 \qquad \Rightarrow \qquad \lambda_{1,2} = -\frac{1}{2} \pm i\,\frac{\sqrt{3}}{2}$$

und somit die Lösung

$$y(x) = e^{-\frac{1}{2}x} \cdot \left(C_1 \cos\frac{\sqrt{3}}{2}x + C_2 \sin\frac{\sqrt{3}}{2}x \right).$$

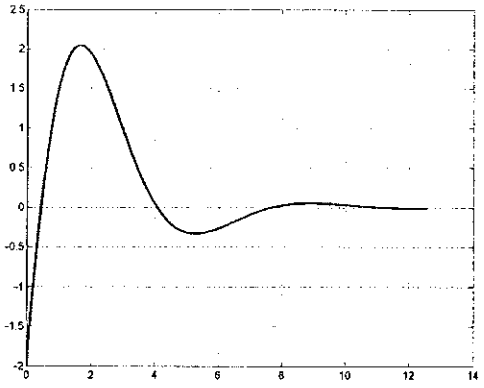


Abbildung 11: $y(x) = e^{-\frac{1}{2}x} \cdot \left(-2 \cdot \cos\frac{\sqrt{3}}{2}x + 5 \cdot \sin\frac{\sqrt{3}}{2}x\right)$

11.3.6 Inhomogene lineare Differentialgleichung mit konstanten Koeffizienten

Die allgemeine Lösung lässt sich wieder als

$$y(x) = y_h(x) + y_p(x)$$

darstellen. Da alle Koeffizienten konstant sind, lässt sich die partikuläre Lösung recht einfach finden. Sie ist

$$\begin{aligned} y_p(x) &= \frac{s}{a_2} \qquad \text{falls } a_2 \neq 0 \\ y_p(x) &= \frac{s}{a_1}x \qquad \text{falls } a_2 = 0,\ a_1 \neq 0\,. \end{aligned}$$

Falls die Koeffizienten nicht konstant wären, müssten wir wieder die Methode der Variation der Konstanten zur Berechnung der partikulären Lösung heranziehen.

BEISPIEL 11.7
Die Differentialgleichung

$$y''(x) - y(x) = 3$$

hat die homogene Lösung

$$\lambda^2 - 1 = 0 \quad \Rightarrow \quad \lambda_1 = 1,\ \lambda_2 = -1 \quad \Rightarrow \quad y_h(x) = C_1 e^x + C_2 e^{-x}$$

und die partikuläre Lösung

$$y_p(x) = \frac{s}{a_2} = \frac{3}{-1} = -3$$

und somit die allgemeine Lösung

$$y(x) = y_h(x) + y_p(x) = C_1 e^x + C_2 e^{-x} - 3\,.$$

Sind weiters die Anfangswerte

$$y(0) = 3 \qquad y'(0) = 4$$

gegeben, so berechnen wir aus

$$y(0) = 3 \qquad \Rightarrow \qquad C_1 + C_2 - 3 = 3$$

und

$$y'(0) = C_1 - C_2 = 4\,,$$

dass

$$C_1 = 5 \qquad \text{und} \qquad C_2 = 1\,,$$

sodass wir folgende Lösung dieses Anfangswertproblems erhalten:

$$y(x) = 5e^x + e^{-x} - 3\,.$$

11.4 Systeme von Differentialgleichungen

11.4.1 Systeme linearer Differentialgleichungen

Wir betrachten hier zunächst nur Systeme linearer homogener Differentialgleichungen mit zwei Variablen und konstanten Koeffizienten.

$$\begin{array}{rcl} \dot{x} &=& ax + by \\ \dot{y} &=& cx + dy \end{array} \qquad \text{oder} \qquad \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{mit } a, b, c, d \in \mathbb{R}$$

wobei $x = x(t)$, $y = y(t)$ und $\dot{x} = \frac{dx}{dt}$, $\dot{y} = \frac{dy}{dt}$.

Sei $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ die Koeffizientenmatrix. Durch Diagonalisierung können wir das System entkoppeln und lösen. Entscheidend dafür ist die Struktur der Eigenwerte der Matrix A.

BEISPIEL 11.8
Gegeben sei das System

$$\begin{array}{rcl} \dot{x} &=& x + 2y \\ \dot{y} &=& 2x + y \end{array}$$

Die zugehörige Koeffizientenmatrix ist $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$.
Diese Matrix hat die Eigenwerte $\lambda_1 = 3$ und $\lambda_2 = -1$.

Die zugehörigen normierten Eigenvektoren sind $w_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $w_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Die beiden Eigenvektoren fassen wir in der Matrix $T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ zusammen.

Es gilt nun dass $D = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix} = T^{-1}AT$ ist.

Wir nehmen nun folgende Koordinatentransformation vor: $\begin{pmatrix} u \\ v \end{pmatrix} = T^{-1} \begin{pmatrix} x \\ y \end{pmatrix}$.

Es gilt dann: $\begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = T^{-1}\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = T^{-1}A\begin{pmatrix} x \\ y \end{pmatrix} = T^{-1}AT\begin{pmatrix} u \\ v \end{pmatrix} = D\begin{pmatrix} u \\ v \end{pmatrix}$

Das Gleichungssystem ist nun entkoppelt, d. h., die Koeffizientenmatrix diagonalisiert

$$\begin{aligned} \dot{u} &= 3\ u \\ \dot{v} &= -\ v \end{aligned}$$

und kann leicht gelöst werden.

Es ist $u(t) = C_1 e^{3t}$ und $v(t) = C_2 e^{-t}$.

Durch Rücktransformation erhalten wir schließlich:

$$\begin{pmatrix} x \\ y \end{pmatrix} = T\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} u - v \\ u + v \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} C_1 e^{3t} - C_2 e^{-t} \\ C_1 e^{3t} + C_2 e^{-t} \end{pmatrix}$$

d. h.,

$$\begin{aligned} x(t) &= C_1 e^{3t} - C_2 e^{-t} \\ y(t) &= C_1 e^{3t} + C_2 e^{-t} \end{aligned}$$

11.4.2 Systeme nichtlinearer Differentialgleichungen

Im Allgemeinen gibt es für Systeme von nichtlinearen Differentialgleichungen nicht immer explizite Lösungsformeln wie bei linearen Differentialgleichungssystemen. Wir schauen uns nur ein spezielles aber auch besonders wichtiges Beispiel an.

Räuber-Beute-Gleichung, das Lotka-Volterra-Modell

Sei $x(t)$ die Anzahl der Beutetiere und $y(t)$ die Anzahl der Räuber in einer Population.

Die relativen Änderungsraten in der Population werden folgendermaßen modelliert:

$$\begin{aligned} \frac{\dot{x}}{x} &= a - by \qquad a, b \geq 0 \\ \frac{\dot{y}}{y} &= -c + dx \qquad c, d \geq 0 \end{aligned}$$

a ist die Wachstumsrate der Beutetiere, b der Ausfall durch Gefressen-Werden, c die Sterberate der Raubtiere, d der Zuwachs proportional zur Anzahl der vorhandenen Beutetiere. Wird in das Gleichgewicht eingegriffen, etwa durch Dezimierung der Beutetiere (z.B. Überfischung, Insektizide), dann beeinflusst das auch die Population der Räuber.

Als Differentialgleichungssystem erhalten wir

$$\begin{aligned} \dot{x} &= x\,(a - by) \\ \dot{y} &= y\,(-c + dx) \end{aligned}$$

Wenn wir ein System von (nichtlinearen) Differentialgleichungen haben

$$\begin{aligned} \dot{x} &= f(x, y) \\ \dot{y} &= g(x, y)\ . \end{aligned}$$

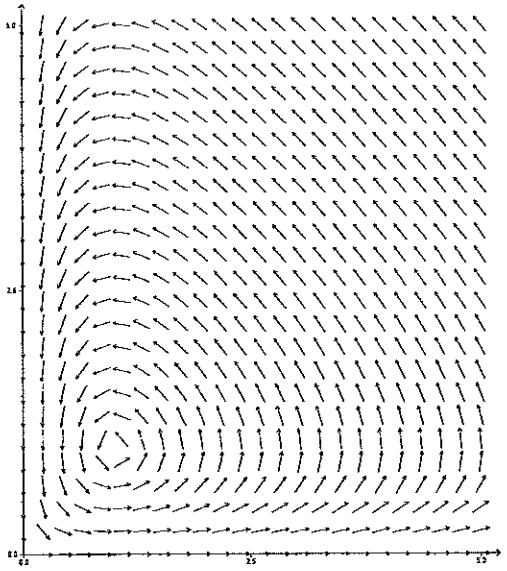


Abbildung 12: $\dot{x} = x(1-y)$ und $\dot{y} = y(-1+x)$.

können wir folgende geometrische Interpretation zu Hilfe nehmen:

Jedem Zeitpunkt t wird ein Vektor $\begin{pmatrix} x \\ y \end{pmatrix}$ zugeordnet. Das entspricht einer ebenen Kurve (siehe mehrdimensionale Analysis). Der Vektor $\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} f(x,y) \\ g(x,y) \end{pmatrix}$ kann als Geschwindigkeitsvektor an der jeweiligen Stelle interpretiert werden.

DEFINITION 11.5
Ein Punkt (x^*, y^*) heißt *Gleichgewichtspunkt* des Differentialgleichungssystems, wenn sowohl $f(x^*, y^*) = 0$ als auch $g(x^*, y^*) = 0$ gilt.

Die Lotka-Volterra-Gleichung hat zwei Gleichgewichtspunkte: $(0,0)$ und $(\frac{c}{d}, \frac{a}{b})$.

Für biologische Modelle sind nur positive Werte sinnvoll, d. h. $x, y \geq 0$.
Auf der x-Achse erhalten wir die Lösung

$$\begin{aligned} x(t) &= x_0\, e^{at} \\ y(t) &= 0 \end{aligned}$$

und auf der y-Achse

$$\begin{aligned} x(t) &= 0 \\ y(t) &= y_0\, e^{-ct} \end{aligned}$$

Um den Punkt $(\frac{c}{d}, \frac{a}{b})$ herum „zirkulieren" die Lösungskurven der Differentialgleichung.

12 Differenzengleichungen

12.1 Allgemeines

Im Kapitel Differentialgleichungen haben wir uns mit Gleichungen beschäftigt, die reelle Funktionen und deren Ableitungen enthalten konnten. Nun werden wir *diskretisierte Varianten* betrachten. Unsere Funktionen sind daher von der Form $y : \mathbb{N}_0 \to \mathbb{R}$, also Folgen reeller Zahlen. Solche Funktionen werden oft verwendet, um zeitliche Entwicklungen einer Größe darzustellen. Häufig wird die Zeitachse *diskretisiert* und in Teilintervalle der Länge Δt unterteilt.

Wir setzen dann: $y(n) = f(n \cdot \Delta t)$ für $n = 0, 1, 2, \ldots$

Durch die Werte $y(0), y(1), y(2), \ldots$ wird der Verlauf der Funktion f annähernd beschrieben.

Statt des Differentialquotienten verwenden wir in diesem Fall entsprechende *Differenzenquotienten*

$$\frac{\Delta f}{\Delta t} = \frac{f((n+1)\Delta t) - f(n\Delta t)}{\Delta t} = \frac{y(n+1) - y(n)}{\Delta t}.$$

Da die Wahl der Zeiteinheit beliebig ist, setzen wir einfach $\Delta t = 1$. Es ergibt sich dann $t = n\Delta t = n$ und $y(n) = y(t) = f(t)$. Dann ist

$$\frac{\Delta f}{\Delta t} = \Delta y(t) = y(t+1) - y(t)$$

Notation: Oft wird y_t statt $y(t)$ geschrieben und wir erhalten somit

$$\Delta y_t = y_{t+1} - y_t$$

Δy_t heißt *erste Differenz* von y_t an der Stelle t.

Analog zu höheren Ableitungen werden entsprechende Differenzen höherer Ordnung definiert:

$$\begin{aligned}
\Delta^0 y_t &= y_t \\
\Delta^1 y_t &= y_{t+1} - y_t \\
\Delta^2 y_t &= y_{t+2} - 2y_{t+1} + y_t \\
\Delta^k y_t &= \Delta(\Delta^{k-1} y_t) = \Delta^{k-1} y_{t+1} - \Delta^{k-1} y_t \quad (k \geq 1)
\end{aligned}$$

$\Delta^k y_t$ heißt *k-te Differenz* oder *Differenz der Ordnung k*.

BEISPIEL 12.1
Für $y_t = t^2$ sehen die ersten 3 Differenzen folgendermaßen aus:

$$\begin{aligned}
\Delta y_t &= (t+1)^2 - t^2 = 2t + 1 \\
\Delta^2 y_t &= (2(t+1) + 1) - (2t + 1) = 2 \\
\Delta^3 y_t &= 2 - 2 = 0
\end{aligned}$$

DEFINITION 12.1
Eine *Differenzengleichung n-ter Ordnung* ist eine Gleichung, die $\Delta^n y_t$ und eventuell noch Differenzen $\Delta^l y_t$ niedrigerer Ordnung $l < n$ enthält. Sie hat die allgemeine Form

$$\sum_{k=0}^{n} \alpha_k \Delta^k y_t, \quad \alpha_k \in \mathbb{R}\,.$$

BEISPIEL 12.2

$$\Delta y_t - \frac{1}{2} y_t = 0$$

Das ist eine Differenzengleichung 1. Ordnung. Diese können wir umformen zu

$$y_{t+1} - y_t - \frac{1}{2} y_t = 0 \qquad \text{oder} \qquad y_{t+1} - \frac{3}{2} y_t = 0 \qquad \text{oder} \qquad y_{t+1} = \frac{3}{2} y_t\,.$$

12.2 Differenzengleichungen erster Ordnung

12.2.1 Lösung per Iteration

Oft ist es möglich, Differenzengleichungen 1. Ordnung bei gegebenem Startwert y_0 zu lösen, indem einfach die $y_1, y_2, \ldots$ der Reihe nach berechnet werden und daraus das Bildungsgesetz der Folge (y_n) *erraten* wird. Der Beweis kann dann durch vollständige Induktion erfolgen.

BEISPIEL 12.3

$$\Delta y_t - \frac{1}{2} y_t = 0$$

Bereits umgeformt zu

$$y_{t+1} - \frac{3}{2} y_t = 0$$

Wir haben also

$$y_1 = \frac{3}{2} y_0 \qquad \text{und} \qquad y_2 = \frac{3}{2} y_1 = \left(\frac{3}{2}\right)^2 y_0\,.$$

Daraus erraten wir das Gesetz

$$y_t = \left(\frac{3}{2}\right)^t y_0\,.$$

12.2.2 Die homogene lineare Differenzengleichung

Die allgemeine Form einer homogenen linearen Differenzengleichung erster Ordnung ist

$$y_{t+1} + a y_t = 0\,.$$

Ausgehend von den Erfahrungen des letzten Beispiels verwenden wir den Ansatz

$$y_t = A\beta^t .$$

Setzen wir dies in die Differenzengleichung ein, erhalten wir

$$A\beta^{t+1} + aA\beta^t = 0 \qquad \text{also} \qquad A\beta^t(\beta + a) = 0 \quad \text{und somit } \beta = -a .$$

Somit hat die allgemeine Lösung die Form $y_t = A(-a)^t$.

Wir können den Parameter A bestimmen, wenn ein Startwert y_0 gegeben ist. Es gilt dann $A = y_0$, denn

$$y_0 = A(-a)^0 = A \qquad \text{und somit} \qquad y_t = y_0 \cdot (-a)^t .$$

Ist $a > 0$, so oszilliert die Folge $(y_t)_{t\in\mathbb{N}}$. Die Folge $(y_t)_{t\in\mathbb{N}}$ konvergiert genau dann, wenn $|a| < 1$.

12.2.3 Die inhomogene lineare Differenzengleichung

Die allgemeine Form einer inhomogenen linearen Differenzengleichung ist:

$$y_{t+1} + ay_t = s, \ a, s \in \mathbb{R} .$$

Ähnlich wie bei den Differentialgleichungen gilt ein Superpositionsprinzip, d.h. die allgemeine Lösung ist gegeben durch

$$y_t = y_t^{(h)} + y_t^{(p)} ,$$

wobei wieder $y_t^{(h)}$ die allgemeine Lösung der zugehörigen homogenen Differenzengleichung $y_{t+1} + ay_t = 0$ und $y_t^{(p)}$ die partikuläre Lösung der inhomogenen Differenzengleichung ist.

Nehmen wir zunächst an, dass $y_t^{(p)} = c$ (d.h. konstant) ist.

Durch Einsetzen erhalten wir:

$$c + ac = s \qquad \text{oder} \qquad c = \frac{s}{a+1}, \ a \neq -1 .$$

Für $a = -1$ versuchen wir den Ansatz $y_t^{(s)} = ct$

$$c(t+1) + a\,ct = s \Rightarrow ct + c - ct = s \Rightarrow c = s \qquad \text{also} \qquad y_t^{(s)} = st .$$

Somit erhalten wir die allgemeine Lösung der inhomogenen linearen Differenzengleichung

$$y_t = \begin{cases} A(-a)^t + \frac{s}{a+1}, & \text{falls } a \neq -1 , \\ A + st & \text{falls } a = -1 . \end{cases}$$

BEISPIEL 12.4

$$\begin{cases} y_0 = 1 \\ y_{t+1} - 2y_t = 2 \end{cases}$$

Allgemeine Lösung

$$\begin{aligned} y_t &= A \cdot 2^t + \tfrac{2}{-2+1} = A \cdot 2^t - 2\,, \\ y_0 &= A \cdot 2^0 - 2 = A - 2\,, \\ A &= y_0 + 2 = 3\,. \end{aligned}$$

Als Lösung erhalten wir also $y_t = 3 \cdot 2^t - 2$.

12.3 Differenzengleichungen zweiter Ordnung

12.3.1 Die homogene lineare Differenzengleichung zweiter Ordnung

Diese hat die allgemeine Form

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = 0,\ a_1, a_2 \in \mathbb{R}\,.$$

Genau wie im Fall der homogenen linearen Differenzengleichung erster Ordnung versuchen wir den Lösungsansatz

$$y_t = A\beta^t\,.$$

Durch Einsetzen erhalten wir

$$A\beta^{t+2} + a_1 A\beta^{t+1} + a_2 A\beta^t = 0$$

und somit

$$\beta^2 + a_1\beta + a_2 = 0$$

Diese Gleichung heißt die *charakteristische Gleichung* der Differenzengleichung zweiter Ordnung. Wie im Fall der charakteristischen Gleichung der Differentialgleichung wird dies nach β aufgelöst und wir erhalten

$$\beta_{1,2} = -\frac{a_1}{2} \pm \sqrt{\frac{a_1^2}{4} - a_2} = 0\,.$$

Abhängig vom Vorzeichen des Ausdrucks unter der Wurzel erhalten wir zwei reelle, eine reelle oder zwei komplex konjugierte Lösungen.

1. $\frac{a_1^2}{4} - a_2 > 0$: also $\beta_1, \beta_2 \in \mathbb{R},\ \beta_1 \neq \beta_2$.

Die allgemeine Lösung ist daher

$$y_t = A_1\beta_1^t + A_2\beta_2^t$$

Die allgemeine Lösung einer Differenzengleichung zweiter Ordnung hat nun zwei Parameter A_1 und A_2. Um diese festzulegen braucht man daher auch zwei Anfangswerte für die Folge, also etwa y_0 und y_1.

BEISPIEL 12.5

$$y_{t+2} - 3y_{t+1} + 2y_t = 0$$

Die charakteristische Gleichung ist:

$$\beta^2 - 3\beta + 2 = 0 \quad \Leftrightarrow \quad \beta_1 = 1,\ \beta_2 = 2\,.$$

Die allgemeine Lösung ist daher

$$y_t = A_1 \cdot 1^t + A_2 \cdot 2^t\,.$$

Für die Anfangsbedingungen $y_0 = 0$ und $y_1 = 1$ erhalten wir

$$0 = A_1 + A_2 \qquad \text{und} \qquad 1 = A_1 + 2A_2$$

und daraus

$$A_1 = -1,\ A_2 = 1$$

Insgesamt also

$$y_t = -1 + 2^t\,.$$

2. $\frac{a_1^2}{4} - a_2 = 0$: also $\beta_1 = \beta_2 = \frac{-a_1}{2} = \beta \in \mathbb{R}$.

Damit ist $A_1\beta^t$ sicher eine Lösung. Wir brauchen jetzt aber noch eine zweite Lösung. Durch Einsetzen in die Gleichung stellen wir fest, dass auch $A_2 t\beta^t$ eine Lösung ist. Die allgemeine Lösung lautet daher

$$y_t = A_1\beta^t + A_2 t\beta^t\,.$$

BEISPIEL 12.6

$$y_{t+2} - 4y_{t+1} + 4y_t = 0\,.$$

Die charakteristische Gleichung ist $\beta^2 - 4\beta + 4 = 0$ uns somit $\beta_1 = \beta_2 = 2$.
Die allgemeine Lösung ist daher

$$y_t = A_1 \cdot 2^t + A_2 \cdot t \cdot 2^t\,.$$

3. $\frac{a_1^2}{4} - a_2 < 0$: Setzen wir

$$b = \sqrt{a_2 - \frac{a_1^2}{4}} \in \mathbb{R} \qquad \text{und} \qquad a = -\frac{a_1}{2} \in \mathbb{R}\,.$$

Dann erhalten wir

$$\beta_{1,2} = a \pm ib$$

Wir gehen zu Polarkoordinaten über

$$\begin{aligned} \beta_1 &= r(\cos\varphi + i\sin\varphi) = re^{i\varphi} \\ \beta_2 &= r(\cos\varphi - i\sin\varphi) = re^{-i\varphi} \end{aligned}$$

mit

$$r = |\beta_1| = \sqrt{a^2 + b^2} = \sqrt{\frac{a_1^2}{4} + a_2 - \frac{a_1^2}{4}} = \sqrt{a_2}$$

und

$$\begin{aligned} \sin\varphi &= \tfrac{b}{r} = \sqrt{1 - \tfrac{a_1^2}{4a_2}} \\ \cos\varphi &= \tfrac{a}{r} = -\tfrac{a_1}{2\sqrt{a_2}} \end{aligned}$$

Als allgemeine Lösung und nach einigen Umformungen erhalten wir

$$y_t = r^t(A_1\cos(\varphi t) + A_2\sin(\varphi t))\,.$$

BEISPIEL 12.7

$$y_{t+2} + y_{t+1} + y_t = 0\,.$$

Die charakteristische Gleichung ist $\beta^2 + \beta + 1 = 0$ mit Lösungen $\beta_{1/2} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}\,i$. Übergang zu Polarkoordinaten

$$r = \sqrt{1} = 1, \quad \sin\varphi = \frac{\sqrt{3}}{2}, \quad \cos\varphi = -\frac{1}{2}$$

und daraus erhalten wir

$$\varphi = \frac{2\pi}{3}\,,$$

und somit die allgemeine Lösung

$$y_t = A_1\cos(\frac{2\pi t}{3}) + A_2\sin(\frac{2\pi t}{3}))\,.$$

12.3.2 Die inhomogene lineare Differenzengleichung zweiter Ordnung

Die allgemeine Form dieser ist

$$y_{t+2} + a_1 y_{t+1} + a_2 y_t = s,\ a_1, a_2 \in \mathbb{R}\,.$$

Wie bei den inhomogenen Differenzengleichungen erster Ordnung kann die allgemeine Lösung wieder zerlegt werden

$$y_t = y_t^{(h)} + y_t^{(p)}\,,$$

wobei wieder $y_t^{(h)}$ die allgemeine Lösung der zugehörigen homogenen Differenzengleichung $y_{t+2} + a_1 y_{t+1} + a_2 y_t = 0$ und $y_t^{(p)}$ die partikuläre Lösung der inhomogenen Differenzengleichung ist.

Wir gehen zunächst wieder davon aus, dass $y_t^{(p)} = c$ konstant ist.

Einsetzen ergibt $c + a_1 c + a_2 c = s$, also $c = \frac{s}{1+a_1+a_2}$. Wieder funktioniert das aber nur, wenn $1 + a_1 + a_2 \neq 0$, d.h., $a_1 + a_2 \neq -1$.

Wenn also $a_1 + a_2 = -1$ ist, verwenden wir den Ansatz $y_t^{(p)} = ct$.

Durch Einsetzen in die Gleichung erhalten wir

$$c = \frac{s}{a_1 + 2} .$$

Das funktioniert natürlich nur dann wieder wenn $a_1 \neq -2$.

Ist nun $a_1 = -2$, setzen wir $y_t^{(p)} = ct^2$ an und erhalten mit $c = \frac{s}{2}$ eine Lösung.

BEISPIEL 12.8

$$y_{t+2} + y_{t+1} + y_t = 9 .$$

Als allgemeine Lösung der zugehörigen homogenen Differenzengleichung hatten wir

$$y_t = A_1 \cos(\frac{2\pi t}{3}) + A_2 \sin(\frac{2\pi t}{3}) .$$

Eine spezielle Lösung der inhomogenen Differenzengleichung erhalten wir mit

$$c = \frac{9}{1+1+1} = 3 \qquad \text{also} \qquad y_t^{(p)} = 3 .$$

Die allgemeine Lösung lautet somit

$$y_t = A_1 \cos(\frac{2\pi t}{3}) + A_2 \sin(\frac{2\pi t}{3}) + 3 .$$

BEISPIEL 12.9
Sei $\Delta^2 y_t = 2$ gegeben. Zunächst formen wir dies um zu:

$$\Delta^2 y_t = y_{t+2} - 2y_{t+1} + y_t .$$

Die zugehörige homogenen Differenzengleichung ist also

$$y_{t+2} - 2y_{t+1} + y_t = 0$$

Als charakteristische Gleichung erhalten wir

$$\beta^2 - 2\beta + 1 = 0 \quad \text{mit den Lösungen} \quad \beta_1 = \beta_2 = 1 .$$

Die allgemeine Lösung lautet demnach

$$y_t = A_1 \beta^t + A_2 t \beta^t .$$

Bei der Suche nach der Speziellen Lösung für die inhomogene Differenzengleichung stellen wir fest, dass

$$1 + a_1 + a_2 = 0 \qquad \text{und} \qquad a_1 = -2 .$$

Daher setzen wir $c = \frac{s}{2} = \frac{2}{2} = 1$ und erhalten schließlich als allgemeine Lösung

$$y_t = A_1 + A_2 t + t^2 .$$

Literaturliste Mathematik

Mathematik für Informatiker allgemein

Oberguggenberger, Michael; Ostermann, Alexander
Analysis für Informatiker.
Springer, 2005, ISBN 3-540-21991-9.
Ein guter Überblick über die Analysis, mit Beispielen in Matlab, Maple und als Java-Applets, mit CD.

Gerald Teschl, Susanne Teschl
Mathematik für Informatiker. Band 1: Diskrete Mathematik und Lineare Algebra.
Springer, 2006, ISBN 3-540-25782-9.
Mit Beispielen in Mathematica. http://www.math.univie.ac.at/~gerald/ftp/book-mfi

Gerald Teschl, Susanne Teschl
Mathematik für Informatiker. Band 2: Analysis und Statistik.
Springer, 2006, ISBN 3-540-28064-2.
Mit Beispielen in Mathematica. http://www.math.univie.ac.at/~gerald/ftp/book-mfi

Josef Leydold
Mathematik für Ökonomen
Oldenbourg Verlag. 2003. ISBN 3486274600.
Dieses Buch bietet eine leicht verständliche Einführung in die wichtigsten Gebiete der Mathematik, für (Wirtschafts-) InformatikerInnen fehlen jedoch Abschnitte über Algebra und Graphentheorie.
***Aus dem Inhalt:** Logik, Mengen, Lineare Gleichungssysteme, Matrizen und Vektoren, Reihen und Folgen, Funktionen, Differenzieren, Integral, Optimierung, Differenzen- und Differentialgleichungen.*

Bärwolff, Günter
Höhere Mathematik für Naturwissenschaftler und Ingenieure.
Elsevier, 2006, ISBN 3-8274-1688-4
Umfangreiches Werk mit allen für das Informatik Curriculum wichtigen Gebieten.

Dietlinde Lau
Algebra und Diskrete Mathematik 1: Grundbegriffe der Mathematik, Algebraische Strukturen 1, Lineare Algebra und Analytische Geometrie, Numerische Algebra
Springer 2007, ISBN-13: 978-3540723646

Dietlinde Lau
Übungsbuch zur Linearen Algebra und analytischen Geometrie: Aufgaben mit Lösungen
Springer, Berlin; 2007, ISBN-13: 978-3540723653

Dietlinde Lau
Algebra und Diskrete Mathematik 2: Linerare Optimierung, Graphen und Algorithmen, Algebraische Strukturen und Allgemeine Algebra mit Anwendungen
Springer 2004, ISBN-13: 978-3540203988

Werner Struckmann, Dietmar Watjen
Mathematik für Informatiker. Grundlagen und Anwendungen.
Spektrum Akademischer Verlag, 2007. ISBN 9783827416315

Baron und Kirschenhofer:
Mathematik für Informatiker, Band 3
Springer. 1996. ISBN 3211827978
Zur Vertiefung ausdrücklich empfohlen.

Dirk Hachenberger
Mathematik für Informatiker.
Pearson Studium, 2005. ISBN 3827371090

Armin Hoffmann, Bernd Marx, Werner Vogt
Mathematik für Ingenieure.
Lineare Algebra, Analysis – Theorie und Numerik. Band 1.
Pearson Studium, 2005. ISBN 3827371139, 9783827371133.

Armin Hoffmann, Bernd Marx, Werner Vogt
Mathematik für Ingenieure.
Vektoranalysis, Integraltransformationen, Differenzialgleichungen, Stochastik - Theorie und Numerik. Band 2.
Pearson Studium, 2005. ISBN 3827371147, 9783827371140.

Neville Dean
Diskrete Mathematik.
Pearson Studium. 2002. ISBN 3827370698

Martin Aigner
Diskrete Mathematik.
Vieweg, 2006, ISBN 3-8348-0084-8

R. Haggarty.
Diskrete Mathematik für Informatiker.
Pearson Studium, 2004. ISBN 8827870957, 9783827370952.

Kurt-Ulrich Witt
Algebraische Grundlagen der Informatik.
Vieweg, 2001, ISBN 3-528-03166-2

Mathematik auf Maturaniveau

Vasili P. Minorski
Aufgabensammlung der höheren Mathematik.
Fachbuchverlag Leipzig. 2001. ISBN 3446189181.

Werner Poguntke
Keine Angst vor Mathe: Hochschulmathematik für Einsteiger.
Teubner 2004. ISBN 3519005018.

M. Precht, K. Voit, R. Kraft
Mathematik 1 für Nichtmathematiker.
Oldenbourg, 2005.

M. Precht, K. Voit, R. Kraft
Mathematik 2 für Nichtmathematiker.
Oldenbourg, 2005.

M. Bachmaier, R. Kraft, M. Precht.
Aufgabensammlung mit Lösungen zur Mathematik für Nichtmathematiker.
Oldenbourg, 2006.

Mathematik allgemein

Tilo Arens, Frank Hettlich, Christian Karpfinger, Ulrich Kockelkorn,
Klaus Lichtenegger, Hellmuth Stachel
Mathematik
Spektrum Akademischer Verlag; 2009, ISBN-13: 978-3827417589

Tilo Arens, Frank Hettlich, Christian Karpfinger, Ulrich Kockelkorn,
Klaus Lichtenegger, Hellmuth Stachel
Arbeitsbuch Mathematik: Aufgaben, Hinweise, Lösungen und Lösungswege
Spektrum Akademischer Verlag; 2008, ISBN-13: 978-3827421234

Mathematik „populärwisssenschaftlich“

Keith Devlin
Muster der Mathematik. Ordnungsgesetze des Geistes und der Natur.
Spektrum Akademischer Verlag, 2002, ISBN 3827413257
Mathematics: Science of Patterns.

Keith Devlin
Das Mathe-Gen.
dtv, 2001, ISBN 3423340088
The Math Gene. How Mathematical Thinking Evolved and Why Numbers are Like Gossip.

K. C. Cole
The Universe and the Teacup. The Mathematics of Truth and Beauty.
Harcourt Brace & Company, 1998, ISBN 0151003238

Charles Seife
Zwilling der Unendlichkeit. Eine Biographie der Null.
Goldmann, 2000, 344215054X
Zero. The Biography of a Dangerous Idea. (Viking publ)

Robert Kaplan
Die Geschichte der Null
Campus Verlag, 1999, ISBN 3593364271

Georges Ifrah
Universalgeschichte der Zahlen
Campus Verlag, 1991, ISBN 3880599564

Erhard Behrends
Fünf Minuten Mathematik. 100 Beiträge der Mathematik-Kolumne der Zeitung DIE WELT.
Vieweg, 2006, ISBN 9783834800824

Lineare Algebra und Geometrie

Strang, Gilbert
Lineare Algebra.
Springer, 2003, ISBN 3-540-43949-8
Gut verständliches Buch, mit numerischen Methoden der linearen Algebra sowie Matlab Code, ausdrücklich empfohlen.

M. Scherfner, T. Senkbeil.
Lineare Algebra für das erste Semester.
Pearson Studium, 2006. ISBN 3827372070, 978382732079.

Herbert J. Muthsam
Lineare Algebra und ihre Anwendungen.
Elsevier 2006, ISBN 9783827416322.

Analysis

Mike Scherfner, Torsten Volland
Analysis I für das erste Semester
Pearson Studium, 2008, ISBN-13: 978-3827373182

Christian Blatter
Analysis I.
Heidelberger Taschenbücher Band 151, 1974, 1980, ISBN 3-540-09483-0

Christian Blatter
Analysis II.
Heidelberger Taschenbücher Band 152, 1974, 1979, ISBN 3-540-09484-0

Christian Blatter
Analysis I.
Heidelberger Taschenbücher Band 153, 1974, 1981, ISBN 3-540-10892-0

Graphentheorie

Dietlinde Lau
Algebra und Diskrete Mathematik 2
Springer 2004.

Baron und Kirschenhofer:
Mathematik für Informatiker, Band 3
Springer. 1996. ISBN 3211827978
Zur Vertiefung ausdrücklich empfohlen.

Neville Dean
Diskrete Mathematik im Klartext, Pearson, 2003.

Rod Haggarty
Diskrete Mathematik für Informatiker, Pearson, 2003.

Richard Johnsonbaugh
Discrete Mathematics, Prentice Hall, 2005.

Reinhard Diestel
Graphentheorie,
Springer, 3. Auflage, 2006, ISBN 3540213910.
http://www.math.uni-hamburg.de/home/diestel

Peter Tittmann
Graphentheorie: Eine anwendungsorientierte Einführung.
Fachbuchverlag Leipzig, 2003, ISBN 3446223436.
http://www.peter.htwm.de

Sven Oliver Krumke, Hartmut Noltemeier
Graphentheoretische Konzepte und Algorithmen.
Teubner, 2005, ISBN 3519005263.

Numerische Mathematik

Numerische Mathematik: Eine beispielorientierte Einführung

Andreas Meister
Numerik linearer Gleichungssysteme: Eine Einführung in moderne Verfahren

Martin Hanke-Bourgeois
Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens.
Teubner, 2002, ISBN 3-519-00356-2.

Matthias Böllhofer, Volker Mehrmann
Numerische Mathematik.
Vieweg, 2004, ISBN 3-528-03220-0.

Mengenlehre

Oliver Dieser
Einführung in die Mengenlehre.
Springer, 2004, ISBN 3-540-42948-4.

Wikipedia

Wer diese Einträge in der Wikipedia sowie andere Internet-Beiträge benutzt, oder andere Internet-Zitate, sollte sehr genau wissen, was er/sie tut. Die Einträge sind von unterschiedlichster Qualität, die Diskussionen zu den Seiten erweisen sich als „lehrreich", nicht immer ist der erzielte Konsens der Seitenautoren auch wissenschaftlich exakt bzw. *sound*. Man kann allerdings selbst zur Verbesserung der Qualität beitragen. Angegebene Originalliteratur hilft, Zitate zu finden.

http://en.wikipedia.org/wiki/Set_theory , Sept 2007
http://en.wikipedia.org/wiki/Zermelo-Fraenkel_set_theory, Sept 2007
http://de.wikipedia.org/wiki/Zermelo-Fraenkel-Mengenlehre, Sept 2007

http://www.mathematik.de/mde/information/landkarte/gebiete/mengenlehre/mengenlehre.html, Sept 2007

Zermelo, Ernst:
Untersuchungen über die Grundlagen der Mengenlehre,
Mathematische Annalen 65 (1908), S. 261-281

Fraenkel, Abraham
Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre
Mathematische Annalen 86 (1921), S. 230-237

Zermelo, Ernst:
Über Grenzzahlen und Mengenbereiche.
Fundamenta Mathematicae 16 (1930), S. 29-47

Logik

Uwe Schöning
Logik für Informatiker.
Spektrum, Heidelberg, 1995.

John N. Martin
Elements of Formal Semantics: An Introduction for Students of Languages.
Academic Press, Orlando, 1987, ISBN 0-12-474856-2

Nils Nilsson
Principles of AI, Palo Alto, 1980.

Werner DePauli-Schimanovich; Peter Weibel
Kurt Gödel -- Ein mathematischer Mythos.
hpt, Wien, 1997. ISBN 3-209-00865-5 (*mit einer langen Logik-Literaturliste*).

Robert Wall
Einführung in die Logik und Mathematik für Linguisten 1,
Band 1: Logik und Mengenlehre.
Skriptor Verlag GmbH, 1973, ISBN 3-589-00023-6.

John L. Casti, Werner DePauli
Gödel. A Life of Logic.
Perseus Publishing, 2000, ISBN 0738202746

Eckehart Köhler, Peter Weibel, Michael Stöltzner, Bernd Buldt, Carsten Klein, Werner DePauli-Schimanovich-Göttig (Hrsg.)
Kurt Gödel. Wahrheit und Beweisbarkeit.
Band 1: Dokumente und historische Analysen.
ÖBV & HPT, 2002, ISBN 3209038341

Bernd Buldt, Eckehart Köhler, Michael Stöltzner, Peter Weibel, Carsten Klein, Werner DePauli-Schimanovich-Göttig (Hrsg.)
Kurt Gödel. Wahrheit und Beweisbarkeit.
Band 2: Kompendium zum Werk.
ÖBV & HPT, 2002, ISBN 320903835X

Spezialgebiete der Mathematik

Angelika Steger
Diskrete Strukturen 1. Kombinatorik, Graphentheorie, Algebra
Springer, Berlin, 2007, 2. Auflage, ISBN-13: 978-3540466604

Erich Neuwirth, Dean Arganbright
Mathematical Modelling (with MS Excel),
Thomson, Brooks, Cole. 2004. ISBN 0534420850.

Jiri Matousek, Jaroslav Nesetril
Diskrete Mathematik: Eine Entdeckungsreise
Springer, Berlin, 2007, ISBN-13: 978-3540301509

Edmund Hlawka, Christa Binder, Peter Schmitt:
Grundbegriffe der Mathematik.
Prugg Verlag, 1979, ISBN 3853850383

Oskar Becker
Grundlagen der Mathematik (in geschichtlicher Entwicklung)
Suhrkamp, 1975, ISBN 3518277146.

Josef Hofbauer, Karl Sigmund
Evolutionstheorie und dynamische Systeme. Mathematische Aspekte der Selektion.
Verlag Paul Parey, 1984. ISBN 3489618343

Georg Glaeser
Der mathematische Werkzeugkasten: Anwendungen in Natur und Technik
Spektrum-Akademischer Verlag, 2004, ISBN-13: 978-3827414854

Georg Glaeser, Konrad Polthier
Bilder der Mathematik
Spektrum-Akademischer Verlag, 2004, 2009, ISBN-13: 978-3827420176

Norbert Herrmann
Mathematik ist überall
Oldenbourg, 3. Auflage, 2007, ISBN-13: 978-3486582437

Jörg Arndt, Christoph Haenel
Pi, Algorithmen, Computer, Arithmetik.
Springer 1998, ISBN 3540634193

Julian Havil
GAMMA: Eulers Konstante, Primzahlstrände und die Riemannsche Vermutung
Springer Berlin, 2007, ISBN-13: 978-3540484950

Schaum's Outline Serien, McGraw-Hill Verlag.

Viele einfache, durchgerechnete Beispiele als Erklärung zu den Formeln.

Schaum's Outline: Linear Algebra: Based on Schaum's Outline of Theory and Problems of Linear Algebra. (Crash Course)
Seymour Lipschutz, Marc Lipson, Kimberly S. Kirkpatrick. ISBN 0071398805. McGraw-Hill. 2002.

Schaum's Outline of Review of Elementary Mathematics
Barnett Rich, Philip A. Schmidt. ISBN 158945766. McGraw-Hill eBook. 2002.

Schaum's Outline: Discrete Mathematics: Based on Schaum's Outline of Theory and Problems of Discrete Mathematics.
Seymour Lipschutz. Marc Lipson, George J. Hademenos. ISBN 0071398775. McGraw-Hill. 2002.

Schaum's Outline of Operations Research
Richard Bronson. Govindasami Naadimuthu. ISBN 0070080208. McGraw-Hill. 1997.

Schaum's Outline of Finite Mathematics
Seymour Lipschutz, John Schiller. ISBN 0070380023. McGraw-Hill. 1994.

Schaum's Outline of Essential Computer Mathematics
Seymour Lipschutz. ISBN 0070379904. McGraw-Hill. 199.

Schaum's Outline of Set Theory and Related Topics
Seymour Lipschutz. ISBN 0070379866. McGraw-Hill. 1967.

Schaum's Outline of Matrix Operations
Richard Bronson. ISBN 0070079781. McGraw-Hill. 1988.

Schaum's Outline of Logic
John Eric Nolt, Dennis Rohatyn, Achille Varzi. ISBN 0070466491. McGraw-Hill. 1998.

Schaum's Outline Introduction to Mathematical Economics
Edward T. Dowling. ISBN 007135896X. McGraw-Hill. 2000.

Schaum's Outline of Mathematics in Finance
Petr Zima, Robert L. Brown. ISBN 0070082030. McGraw-Hill. 1996.

Schaum's Outline of Modern Algebra
Frank Ayres. ISBN 0070026556. McGraw-Hill. 1965.

Schaum's Outline: Algebra
Barnett Rich, Philip A. Schmidt. ISBN 0070522626. McGraw-Hill. 1992.

Schaum's Outline of Vector Analysis
Murray R. Spiegel. ISBN 007060228X. McGraw-Hill. 1968.

Schaum's Outline of 3000 Solved Problems in Calculus
Elliott Mendelson. ISBN 0070415234. McGraw-Hill. 1988.

Schaum's Outline of Trigonometry
Robert E. Moyer, Frank Ayres Jr., Frank Ayres. ISBN 0070068933. McGraw-Hill. 1998.

Schaum's Outline of Partial Differential Equations
Paul C. DuChateau, D. W. Zachmann. ISBN 0070178976. McGraw-Hill. 1986.