Mathematik für Informatiker

Ein Skriptum von

S. Teschl und K. Unterkofler

Susanne Teschl
FH Technikum Wien
Höchstädtplatz 5
A–1200 Wien
Austria, Europe
E-mail: mailto:susanne.teschl@technikum-wien.at
URL: http://staff.technikum-wien.at/~teschl/

Karl Unterkofler
FH Vorarlberg
Department of Computer Science
Forschungsgruppe Life Science-Biomathematik
Hochschulstraße 1
A–6850 Dornbirn
Austria, Europe
E-mail: mailto:karl.unterkofler@tugraz.at
URL: https://itp.tugraz.at/~karl/

Typeset by $\mathcal{AMS}$-LaTeX.
Version: 1. September 2022
Copyright © August 2000 – 2022 by S. Teschl and K. Unterkofler

Einleitung für *IT*–Studentinnen[1]

Die Vorkenntnisse aus der Schule werden z. B. in den Büchern [**7**] oder [**2**] zusammengefasst. Einen Übergang von der Oberstufen- zur Hochschul-Mathematik bietet das Buch von Fritzsche [**3**].

Für die allgemeine Ingenieur-Mathematik sind die leicht lesbaren Bücher von [**19**], [**15**], [**20**] oder auch [**13**], [**16**] und [**10**] zu empfehlen.

Von den deutschsprachigen Mathematikbüchern für Informatiker ist inhaltlich [**23**] zu unserer Vorlesung am nächsten. Zu den neueren auch für Fachhochschulen geschriebenen Büchern, welche grossteils unseren Gesamtstoff abdecken, zählen Hartmann [**27**] oder G. und S. Teschl [**30**], [**31**], wovon letztere eine ausgezeichnete Ergänzung zu unserer Vorlesung bieten.

Ich habe mich aber auch an den Mathematikkursen für „Computer Science“ wie zum Beispiel dem Buch von K. Rosen [**54**] (vgl. [**46**] oder [**52**]) orientiert, wobei ich die unterschiedlichen Vorkenntnisse berücksichtigt habe.

Vom Niveau und Inhalt her kommt daher unserer Vorlesung das englischsprachige Buch von Garnier und Taylor [**42**] am nächsten. Viele Beispiele findet man in den Büchern der Schaum Reihe [**55**].

Im Gegensatz zur Oberstufe wird der Stoff an der Hochschule knapper und abstrakter präsentiert und es wird auch erwartet, dass sich die Studierenden, wenn nötig, selbständig zusätzliche Übungsbeispiele erarbeiten.

Taschenrechner werden keine mehr verwendet, denn im Computerzeitalter mit seinen mächtigen Computeralgebra-Programmen (Mathematica, Wolfram Alpha, Maple, Matlab, etc.), sind Taschenrechner so anachronistisch wie Rechenschieber.

[1] Das erste Programm wurde von Ada Lovelace `http://de.wikipedia.org/wiki/Ada_Lovelace` geschrieben.

Inhaltsverzeichnis

Kapitel 0. Einführung in `Mathematica` 1

§0.1. Erste Schritte 1

§0.2. Funktionen 3

§0.3. Gleichungen 5

§0.4. Programme 7

Kapitel 1. Grundlagen der Mathematik 9

§1.1. Logische Grundbegriffe 9

§1.2. Grundlagen und Beweise in der Mathematik 22

§1.3. Übung 25

§1.4. Übung 27

§1.5. Grundbegriffe der Mengenlehre 29

§1.6. Übung 38

Kapitel 2. Relationen 41

§2.1. Relationen 41

§2.2. Abbildungen 48

§2.3. Mächtigkeiten 52

§2.4. Geordnete Mengen 53

§2.5. Relationales Datenmodell 61

§2.6. Übung 65

§2.7. Übung 65

§2.8. Übung 68

§2.9. Übung 70
§2.10. Übung 71

Kapitel 3. Algebraische Strukturen 73
§3.1. Algebraische Verknüpfungen 73
§3.2. Verbände und Boolesche Algebren 80
§3.3. Übung 86

Kapitel 4. Zahlen 89
§4.1. Die natürlichen Zahlen 89
§4.2. Die ganzen und rationalen Zahlen 92
§4.3. Maschinenzahlen 96
§4.4. Die reellen Zahlen 98
§4.5. Die komplexen Zahlen 100
§4.6. Elementare Zählprinzipien 103
§4.7. Übung 107
§4.8. Übung 108
§4.9. Übung Zahlensysteme 109
§4.10. Übung: komplexe Zahlen 110
§4.11. Übung Abzähltechniken 111

Kapitel 5. Elementare Zahlentheorie 113
§5.1. Teilbarkeit 113
§5.2. Primfaktorzerlegung 116
§5.3. Kongruenzen 117
§5.4. Restklassen 119
§5.5. Die Sätze von Fermat, Euler und Wilson 120
§5.6. Das RSA-Kryptographie-Verfahren 122
§5.7. Übung 125

Kapitel 6. Lineare Algebra und analytische Geometrie 127
§6.1. Untervektorräume 128
§6.2. Dimensionen 129
§6.3. Übung 134
§6.4. Lineare Abbildungen 136
§6.5. Matrizen 138
§6.6. Übung 143

§6.7. Matrizenrechnung 145
§6.8. Übung 157
§6.9. Übung 157
§6.10. Determinanten 160
§6.11. Übung 165
§6.12. Lineare Gleichungssysteme 166
§6.13. Übung 177
§6.14. Euklidische Vektorräume 179
§6.15. Eigenwerte und Eigenvektoren 189
§6.16. Hauptachsentransformationen 196
§6.17. Übung 201

Kapitel 7. Analysis 1 203
§7.1. Folgen und Reihen 203
§7.2. Übung 219
§7.3. Stetigkeit 221
§7.4. Übung 230
§7.5. Elementare Funktionen 231
§7.6. Übung 242
§7.7. Differentialrechnung in $\mathbb{R}^1$ 245
§7.8. Lokale Extrema und Mittelwertsatz 250
§7.9. Übung 256
§7.10. Integralrechnung in $\mathbb{R}^1$ 258
§7.11. Uneigentliche Integrale 266
§7.12. Taylorreihen 271
§7.13. Übung 276

Kapitel 8. Fourierreihen 279
§8.1. Übung 289

Kapitel 9. Näherungsmethoden 291
§9.1. Iterationsverfahren zur Bestimmung von Nullstellen 291
§9.2. Interpolation 295
§9.3. Nummerische Integration 299
§9.4. Übung 300

Kapitel 10. Analysis 2 301

§10.1. Differentialrechnung im $\mathbb{R}^n$ 301
§10.2. Übung 322

Kapitel 11. Differentialgleichungen 325
§11.1. Grundlagen 325
§11.2. Lineare Differentialgleichungen 337
§11.3. Kontrollfragen 351
§11.4. Übung 353

Anhang A. Quadratische Gleichung, Polynomdivision 355

Anhang B. Komplexe Zahlen 357
§B.1. Die Menge der komplexen Zahlen 357
§B.2. Polarkoordinaten 360
§B.3. Eulersche Formel 361
§B.4. Einheitswurzeln 363

Anhang C. Vektorrechnung im $\mathbb{R}^2$ und $\mathbb{R}^3$ 365
§C.1. Skalare und Vektoren 365
§C.2. Vektorräume 368
§C.3. Übung 377

Anhang D. Beschreibende Statistik und Zusammenhangsanalysen 381
§D.1. Häufigkeitsverteilung einer Stichprobe 381
§D.2. Kennwerte einer Stichprobe 388
§D.3. Lineare Korrelation und Regression 394
§D.4. Übungen 402

Anhang. Literaturverzeichnis 405

Kapitel 0

Einführung in Mathematica

0.1. Erste Schritte

Vgl. Anhang A [**30**].

Mathematica ist ein umfassendes Programmpaket, das sowohl symbolisch als auch numerisch rechnen kann. Im einfachsten Fall kann es wie ein Taschenrechner verwendet werden. Geben wir zum Beispiel $3 + 5$ ein und drücken danach die ENTER-Taste beim Ziffernblock (oder alternativ auch SHIFT+RETURN):

```
In[1]:=3 + 5
Out[1]=8
```

Ein Strichpunkt am Ende einer Anweisung unterdrückt die Ausgabe. Sie können mehrere Anweisungen auf einmal eingeben, indem Sie diese durch Strichpunkte trennen:

```
In[2]:=x = 5; 3 * x
Out[2]=15
```

Das Multiplikationszeichen $*$ muss nicht geschrieben werden, ein Leerzeichen genügt. Vergessen Sie aber auf dieses Leerzeichen nicht – das kann nämlich einen grossen Unterschied machen, wie das folgende Beispiel zeigt:

```
In[3]:=xy + x y
Out[3]=xy + 5 y
```

Auch Gross-/Kleinschreibung wird von Mathematica unterschieden

```
In[4]:=X + x
Out[4]=5 + X
```

Sie haben bereits gesehen, dass jede Eingabe und jede Ausgabe mit einer Nummer versehen werden. Sie können auf den jeweiligen Ausdruck jederzeit zurückgreifen:

```
In[5]:=Out[1]/2
Out[5]=4
```

Die *unmittelbar vorhergehende* Ausgabe erhalten Sie mit einem Prozentzeichen:

```
In[6]:=% + 3
Out[6]=7
```

Momentan ist `x` mit dem Wert 5 belegt:

```
In[7]:=1/(1 - x) + 1/(1 + x)
```

$$\texttt{Out[7]=}-\frac{1}{12}$$

Mit `Clear` können Sie diese Belegung löschen:

```
In[8]:=Clear[x]
```

Nun ist `x` wieder unbestimmt:

```
In[9]:=1/(1 - x) + 1/(1 + x)
```

$$\texttt{Out[9]=}\frac{1}{1-\texttt{x}}+\frac{1}{1+\texttt{x}}$$

Zur Vereinfachung eines Ausdrucks können Sie `Simplify` verwenden. Vereinfachen wir beispielsweise die letzte Ausgabe:

```
In[10]:=Simplify[%]
```

$$\texttt{Out[10]=}-\frac{2}{-1+\texttt{x}^2}$$

Vielleicht ist Ihnen aufgefallen, dass `Mathematica` Ausdrücke in einer gut lesbaren Form ausgibt, also beispielsweise eine Potenz in der Form $\texttt{x}^2$ anstelle von `x^2`. Auch wir können Brüche, Potenzen usw. entweder mit den üblichen Symbolen /, ^ usw. eingeben, oder wir können die entsprechenden Symbole mit der Maus aus einer Palette auswählen. So kann etwa der Bruch

```
In[11]:=(x + 1)/x^2;
```

mit der Maus über die Palette `Basic Math Assistent` (zu finden im Menüpunkt `File -> Palettes`) auch in der Form

$$\texttt{In[12]:=}\frac{\texttt{x}+1}{\texttt{x}^2};$$

eingegeben werden. Der Strichpunkt am Ende der Eingabe bewirkt dass die Ausgabe unterdrückt wird (der Ausdruck wird aber natürlich ausgewertet und auf das Ergebnis kann mit % oder `Out[]` zugegriffen werden).

Hilfe zu `Mathematica`-Befehlen finden Sie im Menü unter `Help -> Documentation Center` oder mit dem Befehl `?Befehl`, z.B:

```
In[13]:=?Sin
Out[13]=Sin[z] gives the sine of z.  More...
```

Übung: Versuchen Sie, die Bezeichnung für die Zahl π in `Mathematica` herauszufinden.

0.2. Funktionen

`Mathematica` kennt eine Vielzahl von mathematischen Funktionen. Diese Funktionen beginnen immer mit einem *Grossbuchstaben*. Die Argumente werden in *eckigen* Klammern angegeben. Einige der eingebauten Funktionen sind:

`Sqrt[x]`	Wurzelfunktion $\sqrt{x}$
`Exp[x]`	Exponentialfunktion e^x
`Log[x]`	(Natürlicher) Logarithmus $\ln(x)$
`Log[a,x]`	Logarithmus $\log_a(x)$
`Sin[x]`, `Cos[x]`	Sinus- und Kosinusfunktion
`Abs[x]`	Absolutbetrag $\lvert x\rvert$

Zum Beispiel können wir die Wurzel aus 4 berechnen:

```
In[14]:=Sqrt[4]
Out[14]=2
```

Wenn wir aber etwa `Sin[1]` eingeben, so erhalten wir:

```
In[15]:=Sin[1]
Out[15]=Sin[1]
```

Das ist vermutlich nicht das Ergebnis, das Sie erwartet haben! `Mathematica` wertet den Ausdruck hier *symbolisch* (und nicht numerisch) aus. Und da es für `Sin[1]` symbolisch keinen einfacheren Wert gibt, wird er unverändert ausgegeben. Wir weisen `Mathematica` an numerisch zu rechnen, indem wir das Argument mit einem Komma versehen (in `Mathematica` wird das Komma als Punkt eingegeben, ISO-Norm):

```
In[16]:=Sin[1.]
```

```
Out[16]=0.841471
```

Eine zweite Möglichkeit ist die Verwendung des Befehls `N[ ]`. Lassen wir uns zum Beispiel damit einen numerischen Wert für π ausgeben:

```
In[17]:=N[Pi]
Out[17]=3.14159
```

oder für die Eulersche Zahl:

```
In[18]:=N[E]
Out[18]=2.71828
```

Natürlich können wir auch eigene Funktionen definieren:

```
In[19]:=f[x_] := x^2 + Sin[x] + a
```

Der Unterstrich in `x_` teilt `Mathematica` mit, dass `x` in diesem Ausdruck die unabhängige Variable ist. Die Verwendung von `:=` weist `Mathematica` an, die rechte Seite **jedesmal neu auszuwerten, wenn** `f` aufgerufen wird. Daher haben wir hier auch kein `Out[...]` bekommen. Nun kann die neue Funktion `f` wie jede eingebaute Funktion verwendet werden (solange, bis Sie `Mathematica` beenden):

```
In[20]:=f[2]
Out[20]=4 + a + Sin[2]
In[21]:=f[x]
Out[21]=a + x^2 + Sin[x]
In[22]:=x = 3; f[x]
Out[22]=9 + a + Sin[3]
```

Achtung: man kann Funktionen auch nur mit einem `=` anstelle eines `:=` definieren. Dann wird die rechte Seite **zuerst** ausgewertet (mit allen aktuellen Belegungen, ergibt also hier z.B. mit `x=3` den Wert `a + 9 + Sin[3]`); dieser Wert wird dann (ein für alle Mal) als Funktionswert zugewiesen:

```
In[23]:=g[x_] = x^2 + Sin[x] + a
Out[23]=9 + a + Sin[3]
```

`g` ist damit eine *konstante* Funktion, d.h., wir erhalten immer denselben Funktionswert, unabhängig vom Argument:

```
In[24]:=g[2]
Out[24]=9 + a + Sin[3]
```

Zusammenfassend gibt es also zwei Möglichkeiten: Funktionen von vornherein mit `:=` definieren oder sicherstellen, dass die unabhängige Variable nicht mit

einem Wert belegt ist:

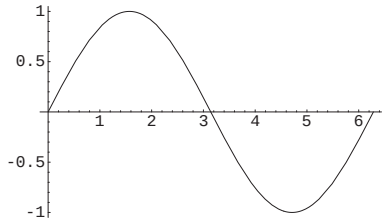
```
In[25]:=Clear[x]; g[x_] = x^2 + Sin[x] + a
Out[25]=a + x^2 + Sin[x]

In[26]:=g[2]
Out[26]=4 + a + Sin[2]
```

Mit `Plot` können Funktionen leicht gezeichnet werden:

`Plot[f[x], {x, xmin, xmax}]`	Zeichnet `f` als Funktion von `x` im Intervall von `xmin` bis `xmax`

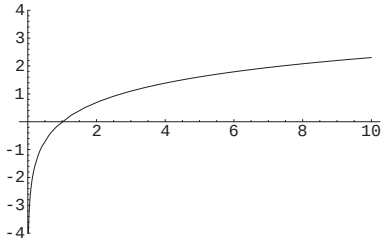
Zeichnen wir zum Beispiel die Funktion `Sin[x]` im Intervall von 0 bis 2π:

```
In[27]:=Plot[Sin[x], {x, 0, 2π}]
```



Meist ist der von `Mathematica` dargestellte Ausschnitt der y-Achse passend. Man kann ihn aber auch selbst mit `PlotRange` festlegen. Wählen wir zum Beispiel für `Log[x]` das y-Intervall von -4 bis 4:

```
In[28]:=Plot[Log[x], {x, 0, 10}, PlotRange → {-4, 4}]
```



Übung: Zeichnen Sie die Funktion $y = \frac{1}{x-1}$ im x-Intervall von 0 bis 2 und im y-Intervall von -20 bis 20.

0.3. Gleichungen

Eine Gleichung wird in `Mathematica` mit einem doppelten Gleichheitszeichen eingegeben

```
In[29]:=Sin[x]^2 + Cos[x]^2 == 1
Out[29]=Sin[x]^2 + Cos[x]^2 == 1
```

und mit Simplify kann man versuchen die Gleichheit zu überprüfen:

```
In[30]:=Simplify[%]
Out[30]=True
```

Unsere Gleichung ist also richtig (True bzw. False sind die englischen Wörter für wahr bzw. falsch).

Die quadratische Gleichung

```
In[31]:=gleichung = (x^2 - 2x - 4 == 0);
```

kann mit dem Befehl Solve gelöst werden:

```
In[32]:=loesung = Solve[gleichung, x]
Out[32]={{x → 1 - √5}, {x → 1 + √5}}
```

Mathematica liefert dabei die Lösung in Form von so genannten Ersetzungsregeln x→ wert. Der Vorteil dabei ist, dass dadurch x nicht automatisch mit wert belegt wird, man aber trotzdem leicht mit der Lösung weiterrechnen kann. Zum Beispiel können wir die beiden Lösungen in unsere Gleichung einsetzen und mit Simplify die Probe machen.

```
In[33]:=Simplify[x^2 - 2x - 4 /. loesung]
Out[33]={0, 0}
```

Mit dem Ersetzungsoperator /. wird durch ausdruck /. x → wert überall in ausdruck die Variable x durch wert ersetzt.

Falls es, wie in unserem Fall, mehrere Lösungen gibt, so kann man auf eine einzelne Lösung mit dem Befehl

```
In[34]:=loesung[[1]]
Out[34]={x → 1 - √5}
```

zugreifen. Allgemein wird in Mathematica ein Ausdruck der Form

```
In[35]:={a, b, c, d, e};
```

als Liste bezeichnet und man kann auf den n-ten Teil einer Liste list mit list[[n]] zugreifen:

```
In[36]:=%[[3]]
Out[36]=c
```

Zusammenfassend gilt:

`Solve[a == b, x]`	Löse die Gleichung mit `x` als Unbekannte
`ausdruck /. loesung`	Setze die Lösung in einen Ausdruck ein
`loesung[[n]]`	Die n-te Lösung

`Mathematica` kann natürlich auch Systeme aus mehreren Gleichungen mit einer oder mehreren Variablen lösen, wie zum Beispiel:

```
In[37]:=Solve[{x + y == a, x - y == 0}, {x, y}]
```
$$\texttt{Out[37]=}\{\{\texttt{x} \to \frac{\texttt{a}}{2}, \texttt{y} \to \frac{\texttt{a}}{2}\}\}$$

0.4. Programme

`Mathematica` ist nicht nur ein Mathematikprogramm, sondern auch eine vollwertige Programmiersprache. Insbesondere stehen die üblichen Kontrollstrukturen und Schleifen zur Verfügung:

`If[test, befehl1, befehl2]`	Ist `test` wahr, so wird `befehl1` ausgewertet, ansonsten `befehl2`.

(`befehl2` ist optional)

`Do[befehl, {j, jmin, jmax, dj}]`	Führe `befehl` mit `j=jmin,jmin+dj,...,jmax` aus
`For[start, test, inkrement, befehl]`	Führe einmal `start` und dann solange `befehl` und `inkrement` aus, bis `test` falsch ist
`While[test, befehl]`	Führe `befehl` aus, solange `test` wahr ist

(das Inkrement `dj` in der `Do`-Schleife ist optional mit Defaultwert `dj`=1).

Beispiel: Der Befehl `PrimeQ` überprüft, ob eine Zahl nur durch sich selbst oder eins teilbar, also eine Primzahl[1] (siehe Kapitel 2.6 [**30**]), ist:

```
In[38]:=PrimeQ[7]
```

[1] Eine Primzahl ist eine natürliche Zahl die genau zwei Teiler hat

```
Out[38]=True
```

Mit diesem Befehl und einer Do-Schleife können wir eine Liste von Primzahlen kleiner gleich einer vorgegebenen Zahl erzeugen. Lassen wir uns zum Beispiel alle Primzahlen kleiner gleich 10 ausgeben:

```
In[39]:=Do[
          If[PrimeQ[n], Print[n]],
        {n, 1, 10}];

           2
           3
           5
           7
```

Mit dem Befehl `Module` können mehrere Befehle übersichtlich zusammengefasst werden:

`Module[{var1=wert1, ...}, befehle]`	Die `befehle` werden mit lokalen Werten für die aufgelisteten Variablen ausgeführt

Die einzelnen Befehle werden durch Strichpunkte getrennt. Das Ergebnis des letzten Befehls wird als Ergebnis des Blocks zurückgegeben.

Zum Beispiel können wir eine Funktion definieren, die die erste Primzahl ausgibt, die grösser oder gleich einer vorgegebene Zahl ist:

```
In[40]:=FindPrime[n_] := Module[{p = n},
              While[!PrimeQ[p], p++];
              p]
```

Dabei wird zuerst p mit n initialisiert. Dann wird p solange um eins erhöht (p++ ist äquivalent zu $p = p + 1$), wie der Primzahltest fehlschlägt (das Rufzeichen negiert den Test: aus wahr wird falsch und aus falsch wahr; hier wird also p um 1 erhöht solange, bis PrimeQ[p] wahr wird). Am Ende wird der gefundene Wert von p ausgegeben.

```
In[41]:=FindPrime[1000]
Out[41]=1009
```

Kapitel 1

Grundlagen der Mathematik

Die Logik ist die Wissenschaft des korrekten Denkens und steht somit am Anfang jeder wissenschaftlichen Disziplin. Sie ermöglicht uns korrekte Schlüsse zu ziehen und widerspruchsfreie Theorien zu entwickeln.

1.1. Logische Grundbegriffe

Betrachten wir folgendes einfaches Beispiel:

Paul sagt, dass Max lügt. Max sagt, dass Hans lügt. Hans sagt, dass Paul und Max beide lügen. Wer lügt hier, wer sagt die Wahrheit?

Kaum jemand ist hier in der Lage unmittelbar eine richtige Antwort zu geben. Andererseits kann auch kaum jemand sofort die richtige Antwort auf die Frage "was ergibt: 200 871 $\times$ 120 570" geben. Aber jeder kann diese Aufgabe ohne Probleme sofort mittels eines einfachen Rechenverfahrens mit Bleistift und Papier ermitteln. Genauso setzt uns die *Aussagenlogik* in die Lage obiges Problem durch einen einfachen Algorithmus (Verfahren) zu lösen.

1.1.1. Aussagen und Wahrheitswerte in der Aussagenlogik.

Definition 1.1. *Unter einer* Aussage (Satz) *verstehen wir einen „gewöhnlichen Satz“ unserer Sprache, von dem man eindeutig entscheiden kann, ob er inhaltlich wahr oder falsch ist.*

Anmerkung: Wir haben hier das semantische Verfahren zur Definition eines Satzes benutzt[1].
Des weiteren verwenden wir die gewöhnliche Sprache um die formale Sprache der Aussagenlogik zu formulieren. Die Sprache, mittels derer eine formale Sprache definiert wird und Aussagen über diese formuliert werden, nennt man Metasprache, während die formale Sprache, die das Objekt der Untersuchung bildet, Objektsprache genannt wird. Metasprache und Objektsprache sind stets zu unterscheiden.

Beispiele:

(a) Bonn ist die Hauptstadt Deutschlands; falsche Aussage.
(b) Ein Dreieck hat 4 Ecken; falsche Aussage.
(c) Guten Tag! Keine Aussage.
(d) Dieser Satz ist falsch; keine Aussage.
(e) Ein Briefträger stellt allen Leuten die Post zu, die sie sich nicht selbst zustellen; keine Aussage.[2]
(f) Es regnet; Aussage.
(g) Das 21. Jahrhundert begann am 1. Januar 2001[3]; Aussage.
(h) Unendlich ist eine gerade Zahl. Aussage?
(i) Es gibt unendlich viele Primzahlen. Aussage?

Beispiel (d) und (e) sind Beispiele für sogenannte Antinomien (Widersprüche). Beispiel (f) ist eine verkürzte Form einer Aussage (siehe auch Beispiel (a)) und es ist aus dem Zusammenhang klar was gemeint ist: z. B. es regnet hier und jetzt.[4].

Notation: Einfache Aussagen (Elementaraussagen bestehen meist aus Subjekt und Prädikat) werden mit Kleinbuchstaben bezeichnet $p, q, r, \ldots$. Ist eine Aussage wahr, so hat sie den *Wahrheitswert wahr* (man schreibt kurz: w oder 1) und ist eine Aussage falsch, so hat sie den Wahrheitswert *falsch* (kurz: f oder 0).

Die *zweiwertige Aussagenlogik* (kurz Logik) handelt von der Operation der Verneinung und den Verknüpfungen (Junktoren) von Aussagen (d. h. von zusammengesetzten Aussagen), aber nicht ob eine Aussage wahr oder falsch ist.

(i) Es werden ausschließlich Aussagen miteinander verknüpft. Das Ergebnis ist dabei wiederum eine (zusammengesetzte) Aussage.

---

[1] Semantik `https://de.wikipedia.org/wiki/Semantik`

[2] Variationen dazu: Ein Barbier rasiert alle Männer, die sich nicht selbst rasieren.

[3] `http://de.wikipedia.org/wiki/21._Jahrhundert`

[4] F. Kutschera (Einführung in die moderne Logik) nennt daher solche Sätze in den neuen Auflagen *Indikatoren* und betrachtet sie daher nicht als Aussagen; spätestens in der Übungsaufgabe 2-2. hat er es wieder vergessen. Ausserdem verschieben zusätzliche Zeit- und Ortsangaben nur das Problem und es würden daher nur mehr mathematische Sätze übrig bleiben. Wir sind pragmatisch und bevorzugen kurze Aussagen.

(ii) Der Wahrheitswert der zusammengesetzten Aussage ist lediglich von den Wahrheitswerten der verknüpften einzelnen Aussagen abhängig.

D. h. die Logik sagt nichts über den Wahrheitswert einer Aussage aus, sondern nur über die möglichen Wahrheitswerte von Verknüpfungen von Aussagen.

1.1.2. Verknüpfungen von Aussagen.

Definition 1.2. *Die* Negation *oder Verneinung einer Aussage p wird durch $\neg p$ (sprich* nicht p*) bezeichnet.*

Die Negation $\neg p$ ist genau dann wahr, wenn p falsch ist.

Weitere gebräuchliche Notationen für die Negation sind $\sim p$ oder $\bar{p}$. In Tabellenform (Wahrheitswerttabelle (Wahrheitstafel)) sieht dies so aus

p	$\neg p$
w	f
f	w

Bei der Verknüpfung v_j von zwei Aussagen p, q sind 16 Kombinationen möglich, die zum Beispiel auf zwei „$\neg, \vee$“ oder eine „$|$“ Grundverknüpfung (die Definition dieser Zeichen erfolgt später) reduziert werden können. Allgemein gilt für n Aussagen, dass 2^{2^n} Kombinationen möglich sind.

p	q	v_0	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}	v_{11}	v_{12}	v_{13}	v_{14}	v_{15}
w	w	w	f	w	w	w	f	f	f	w	w	w	f	f	f	w	f
w	f	w	w	f	w	w	f	w	w	f	f	w	f	f	w	f	f
f	w	w	w	w	f	w	w	f	w	f	w	f	f	w	f	f	f
f	f	w	w	w	w	f	w	w	f	w	f	f	w	f	f	f	f

Die wichtigsten Verknüpfungen sind

Definition 1.3. *Die Konjunktion* $\wedge$ „und“

Die Konjunktion zweier Aussagen p, q ist eine Aussage $p \wedge q$ die genau dann wahr ist, wenn beide Aussagen wahr sind.

Beispiel: Peter und Judith gehen ins Kino.

Wir sagen, dass diese zusammengesetzte Aussage nur dann wahr ist, wenn beide Aussagen wahr sind und erhalten daher die folgende Wahrheitswertetabelle

p	q	$p \wedge q$
w	w	w
w	f	f
f	w	f
f	f	f

Anmerkung: Mit dieser Definition von *und* in der Aussagenlogik ist daher der Satz $\neg(p \wedge \neg p)$ immer wahr. Das ist das Prinzip vom ausgeschlossenen Widerspruch. Oder anders formuliert: Für jede beliebige Aussage p gilt, dass p und *nicht* p nicht gleichzeitig wahr (der Fall) sein können.

Definition 1.4. *Die Disjunktion* $\vee$ „oder“

Die Disjunktion zweier Aussagen p, q ist eine Aussage $p \vee q$ die genau dann wahr ist, sobald eine der beiden Aussagen wahr ist.

Beispiel: Peter steht an der Haltestelle und sagt: “Entweder hat der Bus Verspätung oder der Fahrplan ist falsch.”

Hier kann der Bus Verspätung haben, der Fahrplan falsch sein oder beides der Fall sein. Die Wahrheitswertetabelle lautet daher

p	q	$p \vee q$
w	w	w
w	f	w
f	w	w
f	f	f

Die Disjunktion $\vee$ entspricht dem nichtausschließenden *oder*. Daneben gibt es noch das ausschließende *oder* $\underline{\vee}$.

Beispiel: Die Fussgängerampel ist rot *oder* die Fussgängerampel ist grün.

p	q	$p \underline{\vee} q$
w	w	f
w	f	w
f	w	w
f	f	f

Im normalen Sprachgebrauch gibt es auch noch weitere Bedeutungen des Wortes *und* und *oder*, deren Wahrheitswertverteilung durch Analyse der Wahrheitswerte gewonnen wird (siehe Fahrschulregel in Übung 1.3, Beispiel 2f)

Wenn zum Beispiel auf einem Schild „Rauchen *und* Hantieren mit offenem Feuer verboten!“ steht, dann weiss jeder, dass man hier weder Rauchen noch mit offenem Feuer hantieren darf. Dieses „und“ hat offensichtlich eine andere Bedeutung als $\wedge$, nämlich die von $\vee$. Weitere Beispiele: an der Kassa steht: „Studenten

und Senioren zahlen den halben Preis“ oder „Kinder und Senioren zahlen den halben Preis.“

Anmerkung: Zwei Personen zu finden, die jedes Wort in genau der gleichen Bedeutung verwenden ist nahezu unmöglich und sogar im Sprachgebrauch einzelner Personen verschiebt sich die Bedeutung eines Wortes im Laufe der Zeit.

Anmerkung: In der Aussagenlogik ist der Satz $p \vee \neg p$ immer wahr. Dies ist das Prinzip vom ausgeschlossenen Dritten (tertium non datur).

Definition 1.5. *Die Subjunktion (auch* **Implikation** *genannt)* $\rightarrow$ „wenn, dann“

Anhand folgender Beispiele wollen wir die Wahrheitswertstabelle für die Implikation ermitteln und festlegen.

„Wenn (oder auch: falls) ich die Prüfung bestehe, dann werde ich feiern.“
„Wenn heute ein Feiertag ist, dann wird mir heute keine Post zugestellt.“
„Wenn x eine ganze Zahl größer 10 ist, dann ist ihr Quadrat x^2 größer als 100.“

Offensichlich werde ich nur für den Fall, dass ich die Prüfung bestehe und nicht feiern gehe einer falschen Aussage bezichtigt. Über den Fall was ich mache, wenn ich die Prüfung nicht bestehe, wird nichts gesagt. Die Wahrheitswertstabelle lautet daher:

p	q	$p \rightarrow q$
w	w	w
w	f	f
f	w	w
f	f	w

Merkregel: *die Implikation ist daher eine Aussage die nur falsch ist, wenn der Vordersatz p wahr und der Hintersatz q falsch ist.*
Anmerkung: eine Implikation beschreibt keinen **kausalen** Zusammenhang. Auch wenn p falsch ist, ist $p \rightarrow q$ wahr (richtig)!

Kommentar: Mit logischen Methoden ist nicht feststellbar, ob eine Implikation wahr ist! Man muss dazu die tatsächlich vorliegenden empirisch Wahrheitswerte ermitteln!
Wenn 2×2 *gleich* 4 *ist, dann ist die Sonne ein Planet* ist daher eine falsche Implikation (Implikation ist falsch).
Wenn 2×2 *gleich* 5 *ist, dann ist die Sonne ein Planet* ist eine richtige Implikation (Implikation ist wahr).

Beispiel 1.6. *Man zeige mit einer Wahrheitswertetafel, dass gilt:*

$$(p \to q) \quad \textit{ist gleichwertig zu} \quad ((\neg p) \vee q).$$

Man erhält

p	q	$p \to q$	$(\neg p) \vee q$
w	w	w	w
w	f	f	f
f	w	w	w
f	f	w	w

Die Analyse mit Wahrheitswerten zeigt sofort die Gültigkeit von Sätzen, die sonst nicht unmittelbar einsichtig sind. Zwei Beispiele (siehe A. Tarski[5] [**111**], Seite 56)

$$\text{(i)} \quad p \to (q \to p),$$
$$\text{(ii)} \quad \neg p \to (p \to q).$$

In Worten:
(i) Wenn p wahr ist, dann folgt p aus einem beliebigen Satz oder in anderer Formulierung: Ein wahrer Satz folgt aus einem beliebigen Satz.
(ii) Wenn p falsch ist, dann impliziert p ein beliebiges q oder in anderer Formulierung: Ein falscher Satz impliziert einen beliebigen Satz.

Anmerkung: Durch den Gebrauch von Buchstaben für Aussagen wird die Bedeutung der Verknüpfungen von Ausagen von allen psychologischen Einflussfaktoren befreit, die durch den Inhalt der konkreten Aussagen bewirkt werden kann.

Definition 1.7. *Die Bijunktion* $\leftrightarrow$ „genau dann, wenn“*.*

Die Bijunktion zweier Aussagen ist eine Aussage die genau dann wahr ist, wenn beide Aussagen den gleichen Wahrheitswert haben.

Beispiel: Genau dann, wenn ich das Geld bekomme, werde ich ein neues Auto kaufen.

Die Wahrheitswertetabelle lautet daher:

p	q	$p \leftrightarrow q$
w	w	w
w	f	f
f	w	f
f	f	w

[5] http://de.wikipedia.org/wiki/Tarski

Zwei weitere der 16 Verknüpfungen seien noch erwähnt: die Exklusion $p|q$ „höchstens eins von beiden“ (auch NAND oder Scheffer-Strich genannt) und die Rejektion $p \downarrow q$ „keines von beiden“(weder noch) (auch NOR genannt)

p	q	$p\|q$	$p \downarrow q$
w	w	f	f
w	f	w	f
f	w	w	f
f	f	w	w

Beispiel 1.8. *Man stelle die Wahrheitswerttabelle von $p|p$ auf.*

Um eine eindeutige zusammengesetzte Aussage zu erhalten müssen Klammern gesetzt werden. Diese können teilweise weggelassen werden, wenn man dabei den Operatoren folgende Prioritäten zuweist:

(1) $\neg$ bindet stärker als $\wedge$, $\vee$, z. B. $\neg p \wedge q$ heißt $(\neg p) \wedge q$,

(2) $\wedge$ und $\vee$ sind gleichberechtigt,

(3) $\wedge$ und $\vee$ binden stärker als $\rightarrow$, z. B. $p \vee q \rightarrow r$ heißt $(p \vee q) \rightarrow r$,

(4) $\rightarrow$ bindet stärker als $\leftrightarrow$, z. B. $p \rightarrow q \leftrightarrow r$ heißt $(p \rightarrow q) \leftrightarrow r$.

In der Informatik wird auch noch gerne die Konvention verwendet, dass „und“ stärker als „oder“ bindet.

Beispiel 1.9. *Drücke die folgenden Sätze in Termen von p und q aus, wobei p „heute ist Montag“ und q „ich fahre nach Wien“ bedeutet.*

(i) Wenn es heute Montag ist, werde ich nicht nach Wien fahren.
(ii) Heute ist Montag oder ich werde nach Wien fahren; aber nicht beides.
(iii) Ich fahre nach Wien und heute ist nicht Montag.
(iv) Nur wenn heute kein Montag ist, werde ich nach Wien fahren.

Beispiel 1.10. *p stehe für „die Sonne scheint“ und q bedeute „zwei mal zwei ist fünf“. Drücke in Worten aus.*
(i) $p \vee \neg q$
(ii) $\neg(q \wedge p)$
(iii) $\neg p \rightarrow q$
(iv) $\neg p \leftrightarrow \neg q$.

Nun zu zwei gebräuchlichen Begriffen der Mathematik, nämlich hinreichend und notwendig. Es seien B und S zusammengesetzte Aussagen.

Definition 1.11. *Eine Bedingung B heißt* hinreichend *für den Sachverhalt S, wenn die Implikation $B \rightarrow S$ wahr ist, d. h. immer wenn B wahr ist, ist auch S wahr.*

Beispiel 1.12. *Hinreichend dafür, dass eine Zahl durch* 6 *teilbar ist* (S)*, ist, dass sie durch* 12 *teilbar ist* (B)*.*

Definition 1.13. *Eine Bedingung B heißt* notwendig *für den Sachverhalt S, wenn die Implikation $S \to B$ wahr ist, d. h. wenn der Sachverhalt ohne diese Bedingung B nicht gilt.*

Beispiel 1.14. *Notwendig dafür, dass eine Zahl durch* 6 *teilbar ist* (S)*, ist, dass sie gerade ist* (B)*.*

Ein Beispiel für eine hinreichende und notwendige Bedingung ist

Beispiel 1.15. *Hinreichend und notwendig dafür, dass eine Zahl durch* 6 *teilbar ist, ist dass sie durch* 2 *und* 3 *teilbar ist.*

Beispiel 1.16. *Definition: Jedes Viereck mit zwei parallelen Seiten heißt* Trapez*. Ist ein Rechteck ein Trapez?*[6]

Definition 1.17. *Eine (zusammengesetzte) Aussage heißt* Tautologie *(T), wenn sie bei Ersetzung ihrer Komponenten durch beliebige wahre oder falsche Aussagen immer den Wahrheitswert w ergibt.*
Eine Aussage heißt Kontradiktion *(K), wenn sie bei Ersetzung ihrer Komponenten durch beliebige wahre oder falsche Aussagen immer den Wahrheitswert f ergibt.*
Ansonsten nennt man eine Aussage kontingent*.*

Beispiele: (i) Man zeige durch Einsetzen von Wahrheitswerten, dass

$$\big((p \to q) \wedge \neg q\big) \to \neg p$$

eine Tautologie ist.

(ii) Man zeige durch Einsetzen von Wahrheitswerten, dass

$$\big((p \to q) \wedge (q \wedge r)\big) \to (r \to p)$$

keine Tautologie ist.

Definition 1.18. *Zwei (zusammengesetzte) Aussagen P, Q heißen* logisch äquivalent $P \equiv Q$ *(oder auch $P \Leftrightarrow Q$), wenn sie bei Ersetzung ihrer Komponenten durch beliebige wahre oder falsche Aussagen immer den gleichen Wahrheitswert ergeben.*

[6] http://www.spektrum.de/alias/nachgehakt/ist-jedes-rechteck-ein-trapez/829682

Es gilt offensichtlich, dass $P \equiv Q$ genau dann der Fall ist, wenn $P \leftrightarrow Q$ eine **Tautologie** ist.

Einige wichtige logische Äquivalenzen sind

$$p \equiv \neg(\neg p), \quad \text{Involution}, \tag{1.1}$$

$$\neg(p \vee q) \equiv \neg p \wedge \neg q, \quad \text{De Morgan},$$

$$\neg(p \wedge q) \equiv \neg p \vee \neg q, \quad \text{De Morgan}, \tag{1.2}$$

$$p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r), \quad \text{Distributivitätsgesetz},$$

$$p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r), \quad \text{Distributivitätsgesetz}, \tag{1.3}$$

$$p \leftrightarrow q \equiv (p \to q) \wedge (q \to p), \tag{1.4}$$

$$p \to q \equiv \neg q \to \neg p \equiv \neg p \vee q. \tag{1.5}$$

Weitere logische Äquivalenzen

Idempotenz

$$\begin{aligned} p \wedge p &\equiv p, \\ p \vee p &\equiv p. \end{aligned} \tag{1.6}$$

Kommutativität

$$\begin{aligned} p \wedge q &\equiv q \wedge p, \\ p \vee q &\equiv q \vee p. \end{aligned} \tag{1.7}$$

Assoziativität

$$\begin{aligned} p \wedge (q \wedge r) &\equiv (p \wedge q) \wedge r, \\ p \vee (q \vee r) &\equiv (p \vee q) \vee r. \end{aligned} \tag{1.8}$$

Adjunktivität (Absorption)

$$\begin{aligned} p \wedge (p \vee q) &\equiv p, \\ p \vee (p \wedge q) &\equiv p. \end{aligned} \tag{1.9}$$

Identität

$$\begin{aligned} p \vee K &\equiv p, \\ p \wedge T &\equiv p, \\ p \vee T &\equiv T, \\ p \wedge K &\equiv K. \end{aligned} \tag{1.10}$$

Komplementarität

$$wp \vee \neg p \equiv T,$$
$$p \wedge \neg p \equiv K. \tag{1.11}$$

Anmerkung: Eng mit der Anzahl der Wahrheitswerte "falsch" ist der Informationsgehalt einer Aussage verknüpft. Der Informationsgehalt einer Tautologie ist klarerweise Null, wie man es zum Beispiel am tautologischen Satz "Morgen wird es den ganzen Tag regnen oder morgen wird es den ganzen Tag nicht regnen" erkennt.

Definition 1.19. *Ein* logischer Schluss *ist die Behauptung (Aussage), dass die Aussagen $P_1, P_2, \ldots, P_n$ (Prämissen, Voraussetzungen) eine neue Aussage Q (oder K für Konklusion) zur Folge haben. Ein logischer Schluss heißt gültig (richtig, wahr), wenn Q wahr ist, falls alle Voraussetzungen wahr sind. Man schreibt dann $P \vdash Q$ (oder auch $P \Rightarrow Q$).*

Weitere Sprechweisen für $P \Rightarrow Q$ sind: "daraus folgt" oder "impliziert logisch" (logische Implikation). P steht dabei für $P_1 \wedge P_2 \wedge \ldots \wedge P_n$. Es gilt offensichtlich, dass $P \vdash Q$ genau dann gültig ist, wenn $P \to Q$ eine Tautologie ist.

Beispiel: $[(p \leftrightarrow q) \wedge q] \vdash p$. Dies ist gleichwertig mit $[(p \leftrightarrow q) \wedge q] \to p$ ist eine Tautologie.

Wichtige gültige logische Schlussweisen

$$p \wedge (p \to q) \vdash q \quad \text{Abtrennungsregel (modus ponens),} \tag{1.12}$$
$$(p \to q) \wedge \neg q \vdash \neg p \quad \text{Widerlegungsregel (modus tollens),} \tag{1.13}$$
$$(p \to q) \vdash (\neg q \to \neg p) \quad \text{Kontraposition,} \tag{1.14}$$
$$(p \to q) \wedge (q \to r) \vdash (p \to r) \quad \text{Kettenschluss (modus barbara),} \tag{1.15}$$
$$(p \vee q) \wedge (p \to r) \wedge (q \to r) \vdash r \quad \text{Fallunterscheidung,} \tag{1.16}$$
$$(p \to q) \wedge (p \to \neg q) \vdash \neg p \quad \text{reductio ad absurdum.} \tag{1.17}$$

Anmerkung: Es sei nochmals betont, dass es nicht die Aufgabe der Logik ist, die Wahrheit der Prämissen eines vorgelegten Schlusses nachzuprüfen. Hat sich ein vorgelegter Schluss als gültig erwiesen, so bedeutet dies nicht, dass damit die Wahrheit der Prämissen und damit der Konklusion gesichert ist. Es bedeutet nur, dass die Wahrheit der Konklusion gesichert ist für den Fall, dass die Prämissen wahr sind.

Beispiel (i):[7] Prämisse 1: Wenn die Wirtschaft wächst, steigt das Preisniveau. Prämisse 2: Vollbeschäftigung herrscht nur dann, wenn die Wirtschaft wächst. Konklusion: Also ist es nicht wahr, dass Vollbeschäftigung herrscht und das

[7] aus L. Czayka, Grundzüge der Aussagenlogik, UTB Verlag, 1972

Preisniveau nicht steigt.
Man überprüfe, ob es sich dabei um einen gültigen Schluss handelt!

Beispiel (ii): Ein typisches Beispiel für einen falschen logischen Schluss:

$$(p \to q) \not\vdash (\neg p \to \neg q)$$

Jede zusammengesetzte Aussage kann mit Hilfe von logischen Äquivalenzen auf zwei Normalformen (NF) gebracht werden.

Definition 1.20. *Eine* disjunkte Normalform *(DNF) ist eine Disjunktion von verschiedenen Konjunktionstermen. Sie hat also die Form*

$$K_1 \vee K_2 \vee K_3 \vee \dots K_n \quad \textit{wobei die } K_j,\ 1 \le j \le n \textit{ nur } \neg \textit{ und } \wedge \textit{ enthalten.} \tag{1.18}$$

Es ist leicht festzustellen, ob eine disjunkte Normalform wahr ist, da nur ein Term K_j immer wahr sein muss.

Definition 1.21. *Eine* konjunktive Normalform *(KNF) ist eine Konjunktion von verschiedenen Disjunktionstermen. Sie hat also die Form*

$$D_1 \wedge D_2 \wedge D_3 \wedge \dots D_n \quad \textit{wobei die } D_j,\ 1 \le j \le n \textit{ nur } \neg \textit{ und } \vee \textit{ enthalten.} \tag{1.19}$$

Damit eine konjunktive Normalform falsch ist, genügt es, wenn ein Term D_j immer falsch ist.

Um zu Normalformen zu gelangen werden häufig folgende logische Äquivalenzen verwendet

$$\begin{aligned} &p \to q \equiv \neg p \vee q, \\ &p \leftrightarrow q \equiv (p \to q) \wedge (q \to p), \\ &p \equiv p \vee (q \wedge \neg q) \equiv (p \vee q) \wedge (p \vee \neg q), \\ &p \equiv p \wedge (q \vee \neg q) \equiv (p \wedge q) \vee (p \wedge \neg q). \end{aligned} \tag{1.20}$$

Beispiel 1.22. *Man bringe* $(p \to q) \wedge (p \vee \neg q)$ *auf Normalform.*

$$\begin{aligned} &(p \to q) \wedge (p \vee \neg q) \ \textit{Auflösung von } \text{“}\to\text{”} \\ &(\neg p \vee q) \wedge (p \vee \neg q) \ \textit{KNF.} \end{aligned}$$

Nun kann man das Distributivgesetz verwenden um zur DNF zu kommen

$$(\neg p \wedge p) \vee (\neg p \wedge \neg q) \vee (q \wedge p) \vee (q \wedge \neg q) \equiv (\neg p \wedge \neg q) \vee (q \wedge p).$$

Anmerkung: Man kann die Normalform auch direkt aus der gegebenen Wahrheitswertverteilung gewinnen.

Ausblicke: Neben dem von uns verwendeten sogenannten *semantischen* Verfahren, bei dem die Aussagen und Verknüpfungen mit Wahrheitswerten belegt werden, gibt es auch noch das sogenannte *syntaktische* Verfahren. Bei diesem werden aus einem Axiomensystem (z. B. Gentzenkalkül [**100**], Axiomensytem von Hilbert und Ackermann) durch festgelegte Schlussregeln die Sätze der Aussagenlogik abgeleitet. Es ist dabei nicht notwendig den Sätzen einen Wahrheitswert zuzuordnen. Man kann zeigen, dass Axiomensysteme der Aussagenlogik vollständig und widerspruchsfrei[8] sind und jeder Satz kann in endlich vielen Schritten aus den Axiomen abgeleitet werden.
Physikalisch kann man die Aussagenlogik durch elektrische Schaltungen (Schaltalgebra) darstellen. *Kein Strom* entspricht dann dem Wahrheitswert f und *Strom geht durch* entspricht dem Wahrheitswert w. *Und* kann durch eine Reihenschaltung und *oder* durch eine Parallelschaltung verwirklicht werden.
Aussagenlogik und Schaltalgebra haben dieselbe Struktur (Boolesche Algebra).

Am Anfang jeder Wissenschaft steht immer die Aussagenlogik, da sie nur auf der gewöhnlichen Sprache und auf keinen weiteren Begriffen aufbaut, aber immer die Voraussetzung für alles Weitere ist.

1.1.3. Prädikate und Quantoren. Für die Formulierung mathematischer Theorien reicht die Aussagenlogik nicht aus[9]. Bei der Analyse von einfachen Aussagen stößt man auf *Subjekte* (das sind Namen für Objekte, Individuen, Elemente), *Prädikate* und sogenannte *quantifizierende* Redeteile.

Ein *Prädikat* M beschreibt die Eigenschaft eines (einstelliges Prädikat) oder mehrerer (mehrstelliges Prädikat) Objekte oder Individuen (Elemente). Beispiele für Prädikate sind

(i) R : ist rot

(ii) L : hat lange Haare

(iii) T : ist durch drei teilbar

Mit Kleinbuchstaben bezeichnen wir die speziellen Objekte, z.B.

(i) t : Karls T-Shirt

(ii) j : Judith

(iii) s : die Zahl 26

[8] Die Widerspruchsfreiheit der klassischen Aussagenlogik wurde 1918/1926 von Paul Bernays und 1921 von Emil Leon Post bewiesen.

[9] Z.B. Definition der Menge aller durch 3 teilbaren Zahlen.

Wenn R das Prädikat *ist rot* bedeutet so schreibt man $R(x)$ für x *ist rot*. Damit kann die Aussage *Karls T-Shirt ist rot* geschrieben werden als $R(t)$ und Judith hat lange Haare als $L(j)$. $T(26)$ ist dann die Aussage, dass 26 durch 3 teilbar ist.

In $M(x)$ ist x eine Variable die als Platzhalter fungiert und $M(x)$ ist keine Aussage sondern eine *„Aussagefunktion“*, da dem Objekt x kein Wahrheitswert zugewiesen ist. Die Aussagefunktion $M(x)$ wird zur Aussage, wenn man für x ein bestimmtes Objekt einsetzt.

Eine weitere Möglichkeit aus Aussageformen Sätze zu erzeugen ist die Zuhilfenahme von Quantoren.

Definition 1.23. *Man definiert folgende Quantoren*

(i) Allquantor: $\forall x$ *bedeutet* für alle x,

(ii) Existenzquantor: $\exists x$ *bedeutet* es gibt (mindestens) ein x.

Anmerkung:
Der Allquantor entspricht einer Verknüpfung durch „und“ der Form
$x_1 \wedge x_2 \wedge x_3 \wedge x_4 \ldots$
und der Existenzquantor entspricht einer Verknüpfung durch „oder“ der Form
$x_1 \vee x_2 \vee x_3 \vee x_4 \ldots$

Der Satz $\exists x\, R(x)$ heisst somit: es gibt (mindestens) ein x mit der Eigenschaft x ist ein rotes T-Shirt; kurz: es existiert ein rotes T-Shirt.

Ein Beispiel für ein zweistelliges Prädikat ist $P(x, y) : \; x + y = 7$ und ein entsprechender Satz ist

$\forall x \exists y\, P(x, y)$, in Worten: *zu jedem x gibt es ein y, sodass $P(x, y)$ gilt,*

$\exists x \forall y\, P(x, y)$, *in Worten:* es gibt ein x, sodass für alle y $P(x, y)$ gilt.

Bei den Quantoren muss daher unbedingt auf die Reihenfolge geachtet werden.

Allgemein bestehen offensichtlich folgende Beziehungen für die Negation von Quantoren

Satz 1.24. *(De Morgan)*

$$\begin{aligned} \neg(\forall x\, M(x)) &\equiv \exists x\ (\neg M(x))\,, \\ \neg(\exists x\, M(x)) &\equiv \forall x\ (\neg M(x))\,. \end{aligned} \tag{1.21}$$

Beispiel: Alle Sportler sind Nichtraucher. Formal geschrieben: $\forall x N(x)$. Die Verneinung davon ist: $\neg(\forall x N(x)) \equiv \exists x(\neg N(x))$, d. h. es ist nicht der Fall, dass alle Sportler Nichtraucher sind (linke Seite), oder nach Anwendung des De Morgan

Theorems (rechte Seite): es gibt mindestens einen Sportler der raucht. **Falsch** wäre daher: Alle Sportler sind Raucher.

Ein Beispiel (aus [**42**]) für einen gültigen Schluss in der Prädikatenlogik:

P_1: Jedem Politiker muss misstraut werden.
P_2: John ist Politiker.
K: Also kann man John nicht trauen.

$P(x)$ ... x ist Politiker, j ... John,

$T(x)$ ... x kann vertraut werden.

Man hat also folgenden Schluss zu zeigen

$\big(\forall x\,(P(x) \to \neg T(x)) \wedge P(j)\big) \vdash \neg T(j)$.

Aus der Voraussetzung (Prämisse P_1) $\forall x\,(P(x) \to \neg T(x))$ folgt, dass dies auch für John j gilt, also $P(j) \to \neg T(j)$ und da laut Voraussetzung (P_2) auch $P(j)$ gilt, erhält man insgesamt

$\big(\big(P(j) \to \neg T(j)\big) \wedge P(j)\big) \vdash \neg T(j)$.

Dies ist aber ein gültiger Schluss, da er die Struktur $((p \to q) \wedge p) \vdash q$ hat (Abtrennungsregel (modus ponens) vgl. 1.12).

Anmerkung: Für die hier kurz dargestellte Prädikatenlogik erster Stufe gilt der Gödelsche[10] Vollständigkeitssatz. Das heisst sie ist widerspruchfrei und alle wahren Sätzen können aus den Axiomen abgeleitet werden.

Für die Prädikatenlogik zweiter Stufe (enthält Prädikate von Prädikaten), die zum Beispiel zur Formulierung der Arithmetik notwendig ist, gilt der Gödelsche Unvollständigkeitssatz.

1.2. Grundlagen und Beweise in der Mathematik

Die verschiedenen Teilgebiete der Mathematik bezeichnen wir als Theorien.

Zu ihrer Beschreibung verwendet man formale Sprachen, die zwei Arten von Ausdrücken enthalten

(i) Logische Ausdrücke, die allen Theorien eigen sind und

(ii) spezifische Ausdrücke, die den Inhalt der Theorie festlegen. (Z. B. $\in$ in der Mengenlehre)

Primitive Ausdrücke (**Axiome**) sind spezifische Ausdrücke, die ohne Erklärung eingeführt werden. Durch Definitionen erhält man aus diesen spezifische Ausdrücke, die nicht mehr primitiv sind.

[10] Kurt Gödel: `https://de.wikipedia.org/wiki/Kurt_Goedel`

Jede **Definition** besteht aus zwei Teilen, die durch *genau dann* oder *dann und nur dann* getrennt sind. Der erste Teil, *das zu Definierende* (Definiendum), enthält den zu definierenden Ausdruck. Der zweite Teil, *das Definierende* (Definiens), enthält nur primitive oder bereits vorher definierte Ausdrücke. Die beiden Teile werden durch die *Identität* „=“ miteinander verbunden. Manchmal schreiben wir statt = auch := um das zu Definierende hervorzuheben.

Beispiele: Hat man bereits die Zeichen $\neg, \vee, \wedge$ definiert, so kann man diese verwenden um $p \to q$, $p \leftrightarrow q$, etc. zu definieren

$$(p \to q) := \neg p \vee q,$$
$$(p \leftrightarrow q) := (p \to q) \wedge (q \to p).$$

Mittels logischer und spezifischer Ausdrücke werden die Aussagen der Theorie formuliert. Aus den primitiven Aussagen werden Theoreme oder Sätze abgeleitet. Ein mathematischer Beweis enthält stets eine Ableitung der Behauptung, entweder aus den Axiomen direkt oder aus einem bereits bewiesenen Satz.

Anhand von zwei Beispielen wollen wir die häufigst vorkommenden Beweismethoden der Mathematik erläutern.

(1) Direkter Beweis:

Satz: Wenn n eine ungerade Zahl ist, dann ist n^2 auch eine ungerade Zahl.

Da n ungerade ist, hat es die Form $n = 2m + 1$, wobei m eine beliebige ganze Zahl ist. Damit erhält man für $n^2 = (2m + 1)^2 = 2\,(2m^2 + 2m) + 1$, was zeigt, dass n^2 auch ungerade ist.

(2) Indirekter Beweis:

Die Umkehrung des obigen Satzes beweist man indirekt. Ein weiters Beispiel ist folgender Satz.

Satz: $\sqrt{2}$ ist eine irrationale Zahl.

Die Annahme, dass $\sqrt{2}$ eine rationale Zahl der Form $\frac{p}{q}, p, q \in \mathbb{Z}, q \neq 0$, wobei der größte gemeinsame Teiler von p und q eins ist, führt zu einem Widerspruch.

Häufig vorkommende Beweismethoden sind

1. Direkter Beweis: Hat das Theorem die Form $p \to q$, so nimmt man p als wahr an und gewinnt q durch Ableitungen. Dann ist auf Grund der Definition von *ableitbar*[11] auch q wahr.
2. Oft ist es leichter $\neg q \to \neg p$ zu beweisen. Wegen (1.14) gilt

$$(\neg q \to \neg p) \vdash (p \to q). \tag{1.22}$$

[11] Das bedeutet die Implikation $p \to q$ ist wahr.

3. Ebenso erhält man einen Beweis des Theorems, wenn man $\neg p \vee q$ beweist, da nach (1.5) gilt

$$(\neg p \vee q) \vdash (p \to q). \tag{1.23}$$

4. Indirekter Beweis: Sei p das zu beweisende Theorem und q ein Axiom oder ein bereits bewiesener Satz. Gilt $\neg p \vdash \neg q$, dann ist $\neg p \to \neg q$ direkt bewiesen. Nach (1.14) folgt $q \to p$ und da q gilt, gilt auch p.

5. Um ein Theorem der Form $p \to q$ indirekt zu bewiesen, beweist man $\neg p$ direkt von der Negation $p \wedge \neg q$ des Satzes ausgehend. Dann folgt das Theorem wegen der Beziehung

$$(p \wedge \neg q \to \neg p) \vdash (p \to q) \tag{1.24}$$

6. Reductio ad absurdum: Um den Satz p zu beweisen, leitet man aus $\neg p$ den Satz p ab. Wegen der Beziehung

$$(\neg p \to p) \vdash p \tag{1.25}$$

ist dies möglich.

Zusammenfassung

(i) Folgende Begriffe sollten ihnen nun bekannt sein

Aussage, Wahrheitswert, Aussageverknüpfungen $\neg, \wedge, \vee, \to, \leftrightarrow$, etc.,
notwendig, hinreichend,
Tautologie, Kontradiktion,
logische Äquivalenz, logischer Schluss,
Prädikat, Quantor,
Axiom, Definition, direkter, indirekter Beweis.

(ii) Sie sollten nun einfache sprachliche Aussagen formal formulieren können und mittels Wahrheitswerttabellen den Wahrheitswert von zusammengesetzten Aussagen überprüfen können. Sie sollten überprüfen können ob ein gültiger logischer Schluss vorliegt! Sie sollten folgende Frage beantworten können: Was kann die Aussagenlogik, was kann sie nicht?

1.3. Übung

Alle Sternaufgaben sind etwas schwieriger und können ausgelassen werden!

(1) Liegt eine Aussage (im Sinne der Aussagenlogik) vor oder nicht und welchen Wahrheitswert hat sie?
   (a) Komm sofort her!
   (b) Die Erde besitzt zwei Monde.
   (c) Es gibt unendlich viele Primzahlen.
   (d) Rot ist schöner als blau.
   (e) Kalt ist langsamer als 7.
   (f) 2 ist größer als 3.
   (g) In den Nächten fällt die schwere Erde aus allen Sternen in die Einsamkeit.
   (h) Zirka 90% unseres Geldes wird von Privaten erzeugt (geschöpft)[12].

(2) Einschließendes oder ausschließendes „oder“?
   (a) Morgen oder übermorgen kann es regnen.
   (b) Morgen oder übermorgen ist Sonntag.
   (c) Betreten des Rasens oder Blumenpflücken ist nicht gestattet.

   Stellen sie die Wahrheitstabellen für folgende umgangssprachlichen Verwendungen von *oder* auf.
   (d) Entweder Fahrplanfehler *oder* Zugverspätung.
   (e) Kopf *oder* Zahl (Münzwurf).
   (f) Trinken *oder* Autofahren (Fahrschulregel).

(3) Aussage p ... Paul sagt die Wahrheit. Aussage m ... Max sagt die Wahrheit.
   a) Man stelle die Wahrheitstabelle für die Aussage „Paul sagt, dass Max lügt“ auf.
   b) Man stelle die Wahrheitstabelle für die Aussage „Wenn Paul die Wahrheit sagt, lügt Max“ auf.

(4) Aussage p... die Sonne scheint.
   Aussage q ... Herr Berger geht zu Fuß ins Büro.
   a) Für Herrn Berger gilt: $p \rightarrow q$.
   Was lässt sich sagen, wenn die Sonne *nicht* scheint?
   b) Für Herrn Berger gilt außerdem: $p \leftrightarrow q$.
   Was lässt sich nun sagen, wenn die Sonne *nicht* scheint?

(5) Notwendig oder hinreichend (oder beides)?

[12] `http://oe1.orf.at/artikel/461965`

(a) Regen ist ....., dass die Straße nass ist.
(b) Auf dem Schild eines Restaurants steht: „Montag ist Ruhetag“. Montag ist ....., dass Ruhetag ist.[13]
(c) Dass die Zahl x durch 9 teilbar ist, ist ..... dafür, dass x durch 3 teilbar ist.
(d) Eine Stromquelle ist ....., dass die (elektrische) Lampe leuchtet.

(6) Sei n eine natürliche Zahl. Aussage p: n ist durch 4 teilbar. Aussage q: n ist eine gerade Zahl. Was ist wahr?
(a) $p \to q$.
(b) $p \leftrightarrow q$.
(c) p ist notwendig für q.
(d) $q \to p$.
(e) p ist hinreichend für q.
(f) p ist notwendig und hinreichend für q.

(7) „Wenn eine ganze Zahl positiv ist, so ist auch ihr Quadrat positiv“.
(a) Schreiben Sie diese Aussage als Implikation. Ist die Implikation wahr? D.h. welchen Wahrheitswert kann diese Aussage tatsächlich haben?
(b) Gilt die Umkehrung „Wenn das Quadrat einer ganzen Zahl positiv ist, so ist die Zahl ebenfalls positiv“?

(8) (i) Geben sie eine notwendige Bedingung an, damit ein Viereck ein Quadrat ist.
(ii) Geben sie eine hinreichende Bedingung an, damit ein Viereck ein Quadrat ist.

(9) Zeigen Sie: $p \leftrightarrow \neg q$ ist genau dann wahr, wenn $q \veebar p$ wahr ist.

(10) Man zeige mithilfe von Wahrheitswerttafeln, dass die Formeln (1.1) bis (1.5) gültige logische Äquivalenzen sind.

[13] Wie würde man das gleiche Schild bei einem Frisör interpretieren?

1.4. Übung

(1) Wie kann $p \wedge q$ mit Hilfe von $\neg$ und $\vee$ ausgedrückt werden?

(2) Wenn die Sonne scheint, bin ich froh. Sind die folgenden Schlüsse logische Schlüsse?
   (a) Heute bin ich froh, also scheint die Sonne.
   (b) Heute regnet es, also bin ich nicht froh.
   (c) Heute bin ich nicht froh, also scheint auch nicht die Sonne.

(3) Ist folgende Aussage ein logisch korrekter Schluss?
   „Wenn der Zölibat der Grund für sexuellen Missbrauch wäre, dürfte es überall dort, wo es den Zölibat nicht gibt auch keinen Missbrauch geben.“[14]

(4) Wenn ein Viereck vier rechte Winkel besitzt, so halbieren einander die Diagonalen. Sind die folgenden Aussagen gültige logische Schlüsse?
   (a) Besitzt ein Viereck nicht vier rechte Winkel, so halbieren einander die Diagonalen nicht.
   (b) Halbieren einander die Diagonalen, so besitzt das Viereck vier rechte Winkel.
   (c) Halbieren einander die Diagonalen nicht, so besitzt das Viereck auch nicht vier rechte Winkel.

(5) Man zeige mithilfe von Wahrheitswerttafeln, dass die Formeln (1.12) bis (1.17) gültige logische Schlüsse sind.

(6) Liegt ein gültiger logischer Schluss vor?
   Hans mag keine Ananas, aber Birnen mag er. Wenn er Birnen mag, dann mag er keine Zitronen. Also mag Hans keine Zitronen.

(7) Wird richtig verneint? Stellen Sie gegebenenfalls richtig.
   (a) Zu jedem Schloss passt ein Schlüssel. Zu keinem Schloss passt ein Schlüssel.
   (b) Es gab höchstens zwei Bewerber. Es gab mindestens zwei Bewerber.
   (c) Das Schaf ist weiß. Das Schaf ist schwarz.

(8) Verneinen Sie folgende Aussagen.
   (a) Alle sind krank.
   (b) Niemand konnte das Problem lösen.
   (c) Bernd ist der Älteste der Familie.
   (d) Wenn die Ampel rot zeigt, muss man anhalten.
   (e) Alle Sportler sind Nichtraucher.
   (f) Er ist höchstens 30 Jahre alt.

[14] Kardinal Schönborn zitiert von C. Rainer im Profil Nr. 14, 2. April 2010

(g) Es wird wärmer.
(h) Die Torte schmeckt gut.

(9) Paul sagt, dass Max lügt. Max sagt, dass Hans lügt. Hans sagt, dass Paul und Max beide lügen. Wer lügt hier, wer sagt die Wahrheit?

(10) Kommissar Holmes hat drei Tatverdächtige: John, Pat und Tom. Er weiß:
(a) Wenn sich Pat oder Tom als Täter herausstellen, dann ist John unschuldig.
(b) Ist aber John oder Tom unschuldig, dann muss Pat der Täter sein.
(c) Ist Tom schuldig, so ist John Mittäter.
Wer ist der Täter?

(11) (*) Die Aussage ¬p kann mit der Verknüpfung | (höchstens eines von beiden, NAND) geschrieben werden als $p|p$. Man stelle $p \wedge q$ und $p \vee q$ ausschließlich mithilfe der Exklusion | dar.

(12) (*)[15] Mayer, Schmied und Weber sind Pilot, Kopilot und Steward einer AUA-Maschine, allerdings nicht unbedingt in der genannten Reihenfolge. Im Flugzeug befinden sich auch drei Reisende mit denselben drei Namen. Um sie von der Besatzung zu unterscheiden, erhalten sie im folgenden ein „Herr“ vor ihre Namen gesetzt. Man weiss:
(a) Herr Weber wohnt in Graz.
(b) Der Kopilot wohnt in Klagenfurt.
(c) Herr Schmied hat bereits vor langer Zeit seine Schulkenntnisse der Mathematik vergessen.
(d) Der Fluggast, der denselben Nachnamen wie der Kopilot hat, lebt in Wien.
(e) Der Kopilot und einer der Passagiere, ein Mathematik-Professor, wohnen im gleichen Ort.
(f) Mayer besiegt den Steward beim Pokern.
Man folgere logisch daraus: wie heisst der Pilot!

(13) (*) Noch so ein Beispiel: `http://en.wikipedia.org/wiki/Zebra_Puzzle`

[15] Dieses Beispiel stammt aus einem Skriptum der TU Graz und ist eher nicht für Wahrheitswerttabellen gedacht.

1.5. Grundbegriffe der Mengenlehre

Mengen sind die Bausteine der Mathematik. Die folgende Charakterisierung (Pseudo-Definition) einer Menge geht auf Cantor (Georg Cantor, 1845-1912) zurück.

Definition 1.25. *Eine* Menge *ist eine Zusammenfassung bestimmter wohlunterschiedener Objekte unserer Anschauung oder unseres Denkens – welche die Elemente der Menge genannt werden – zu einem Ganzen.*

Anmerkung: (i) Diese "Definition" bildet die Grundlage für die naive Mengenlehre. Naive deshalb, weil diese Definition unter Umständen zu Widersprüchen (Antinomien)[16] führen kann. Um diese Widersprüche zu vermeiden, wurde die *axiomatische Mengenlehre* (Zermelo-Fraenkel[17], Hilbert-Bernays, u. a.) eingeführt, die Axiome für das Enthaltensein (Element sein) $\in$ angibt.

(ii) Dies ist keine Definition im eigentlichen Sinne, sondern macht nur plausibel was wir uns unter einer Menge vorstellen sollten, vgl. „primitive Ausdrücke". Ansonsten müssten wir definieren was eine „Zusammenfassung" oder „Objekte" sind und ohne Ende so weiter.

Bezeichnungsweisen: $A, B, X, Y, \ldots$ bezeichnen Mengen und mit $a, b, x, y, \ldots$ werden die Elemente (Objekte) der Mengen bezeichnet. $a \in A$ heißt a ist ein Element der Menge A und $a \notin A$ heißt a ist kein Element der Menge A. Für ein beliebiges x soll stets $x \in A \quad \veebar$ (oder) $x \notin A$ gelten.

Eine Menge M lässt sich angeben durch

(1) Aufzählung, z.B.

$$A = \{a, b, \ldots\}, \tag{1.26}$$

wobei die Elemente $a, b, \ldots$ der Menge zwischen den Mengenklammern explizit angeführt werden. Es kommt dabei nicht auf die Reihenfolge an. Auch wird ein mehrfach auftretendes Element nur einmal gezählt.

Beispiel 1.26. *Die Menge der Buchstaben des Wortes Wasserfall*

$$\{w, a, s, s, e, r, f, a, l, l\} = \{a, e, f, l, r, s, w\}$$

oder

$$\{2, 3, 3, 1, 2, 3\} = \{3, 1, 2\} = \{1, 2, 3\}.$$

[16] Z.B. die Menge aller Mengen die sich nicht selbst enthalten.

[17] `http://de.wikipedia.org/wiki/Zermelo-Fraenkel-Mengenlehre`

(2) Beschreibung. Im beschreibenden Verfahren wird die Menge durch die Eigenschaften ihrer Elemente (d.h. mit Hilfe der Prädikatenlogik),

$$A = \{ \text{ für alle } x \text{ gilt } \mid x \text{ hat die Eigenschaft } E\} = \{x \mid E(x)\} \tag{1.27}$$

charakterisiert.

Anmerkung: Durch das beschreibende Verfahren werden Beweise in der Mengenlehre auf Aussagen zurückgeführt und können daher mittels der Regeln der Logik bewiesen werden.

Bezeichnungen für Zahlenmengen

$$\begin{aligned}
\mathbb{N} &= \{0, 1, 2, 3, \ldots\} = \{x \mid x \text{ ist natürliche Zahl}\},\\
\mathbb{Z} &= \{\ldots, -2, -1, 0, 1, 2, 3, \ldots\} = \{x \mid x \text{ ist ganze Zahl}\},\\
\mathbb{Q} &= \{x \mid x \text{ ist rationale Zahl (Bruchzahl)}\},\\
\mathbb{R} &= \{x \mid x \text{ ist reelle Zahl}\},\\
\mathbb{C} &= \{x \mid x = a + b\,\mathrm{i} \text{ mit } a, b \in \mathbb{R}\}.
\end{aligned} \tag{1.28}$$

Anmerkung: In älteren Darstellungen wird 0 nicht als natürliche Zahl betrachtet und man schreibt dann oft $\mathbb{N}_0$, wenn man 0 einschließt.

Ein Beispiel für das beschreibende Verfahren ist folgendes für die geraden Zahlen

$$\mathbb{G} = \{0, 2, 4, 6, \ldots\} = \{x \mid x = 2\,m, \text{ wobei } m \in \mathbb{N}\}.$$

Definition 1.27. *(a) Eine Menge A heißt* Teilmenge *einer Menge B, wenn jedes Element von A auch in der Menge B enthalten ist.*
(b) Existiert ein Element von B, das nicht in A enthalten ist, so heißt A echte Teilmenge *von B. Man schreibt*

$$\begin{aligned}
A \subset B \quad & \textit{für } A \textit{ ist Teilmenge von } B \textit{ und}\\
A \subsetneq B \quad & \textit{für } A \textit{ ist echte Teilmenge von } B.
\end{aligned} \tag{1.29}$$

Anmerkung: (i) Die Definition (a) schließt nicht aus, dass A und B gleich sind.
(ii) Manche verwenden auch folgende Notation:
$A \subseteq B$ für Teilmengen und $A \subset B$ für echte Teilmengen.

Definition 1.28. *(Gleichheit zweier Mengen) Zwei Mengen A, B heißen* gleich, *wenn $A \subset B$ und $B \subset A$ gilt.*

Definition 1.29. *Ist A eine Menge mit nur endlich vielen Elemente (A heißt dann* endliche Menge*), so bezeichnet $|A|$ die* Kardinalzahl *(Anzahl der Elemente von A) von A.*
Hat die Menge A unendlich viele Elemente, dann sagen wir A hat die Kardinalzahl unendlich und schreiben auch $|A| = \infty$.

Definition 1.30. *Eine Menge heißt* leere Menge*, wenn sie keine Elemente enthält, z.B. wenn* $A = \{\ \}$*. Man schreibt dann* $\emptyset = \{\ \}$*. Für jede Menge* A *gilt:* $\emptyset \subset A$*.*

Man unterscheide zwischen:
a ist ein Element der Menge A und die Menge $\{a\}$ ist eine Teilmenge der Menge A. Das heisst die leere Menge ist zwar Teilmenge jeder Menge aber nicht Element jeder Menge!

1.5.1. Operationen mit Mengen.

Definition 1.31. *Die Menge aller Elemente zweier Mengen* A, B*, die sowohl zu* A *als auch zu* B *gehören, bilden den* Durchschnitt *(Abb. 1.1) der Mengen* A *und* B*, d.h.*

$$A \cap B = \{x \mid x \in A \text{ und } x \in B\}. \tag{1.30}$$

Veranschaulicht werden Mengenoperationen mit Hilfe von Mengendiagrammen (*Eulerkreisen, Venndiagrammen*).

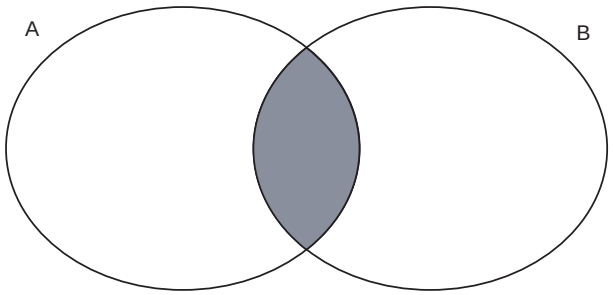


Abbildung 1.1. Durchschnitt $A \cap B$ der Mengen A und B

Definition 1.32. *Zwei Mengen* A *und* B *heissen* disjunkt *(elementfremd), wenn* $A \cap B = \emptyset$*.*

Definition 1.33. *Die Menge aller Elemente zweier Mengen* A, B*, die entweder zu* A *oder* $(\vee)$ *zu* B *gehören, bilden die* Vereinigung *(Abb. 1.2) der Mengen* A *und* B*, d.h.*

$$A \cup B = \{x \mid x \in A \text{ oder } x \in B\}. \tag{1.31}$$

Es gelten die Distributivgesetze der Mengenlehre

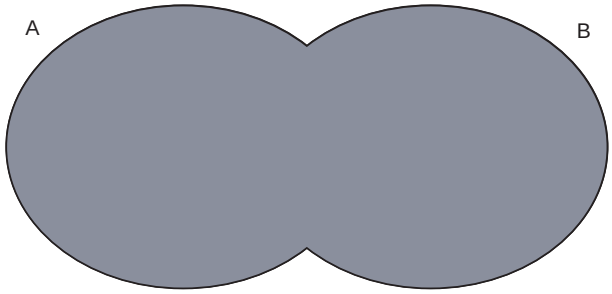


Abbildung 1.2. Vereinigung $A \cup B$ der Mengen A und B

Satz 1.34. *Distributivgesetze*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{1.32}$$

Alle Beweise der Mengenlehre werden mittels des beschreibenden Verfahrens auf entsprechende logische Aussagen zurückgeführt. Sie können durch Mengendiagramme plausibel (ist aber kein Beweis) gemacht werden.

Übung: Man beweise die Gesetze (1.32) und veranschauliche sie graphisch.

$$\begin{aligned} A \cap (B \cup C) &= \{x \,|(\, x \in A) \wedge (\, x \in (B \cup C))\} \\ &= \{x \,|(\, x \in A) \wedge ((x \in B) \vee (x \in C))\} \\ &= \{x \mid (\,(x \in A) \wedge (x \in B)) \vee ((x \in A) \wedge (x \in C))\} \\ &= \{x \,|(\, x \in (A \cap B)) \vee (x \in (A \cap C))\} \\ &= (A \cap B) \cup (A \cap C). \end{aligned}$$

Definition 1.35. *Die Menge aller Elemente einer Menge A die nicht zur Menge B gehören bilden die* Differenzmenge *(Abb. 1.3) der Mengen A und B, d.h.*

$$A \setminus B = \{x \mid x \in A \text{ und } x \notin B\}. \tag{1.33}$$

Es gilt: $A \setminus B \neq B \setminus A$, wenn $A \neq B$.

Definition 1.36. *Ist A eine Teilmenge von G ($A \subset G$) so nennt man die Differenzmenge $G \setminus A$ das* Komplement *(Abb. 1.4) der Menge A bezüglich G.*

$$\complement_G A = \bar{A} = G \setminus A = \{x \mid x \notin A \text{ und } x \in G\}. \tag{1.34}$$

Es seien nun A und B Teilmengen einer Grundmenge (Obermenge) G. Dann gelten die Regeln von de Morgan

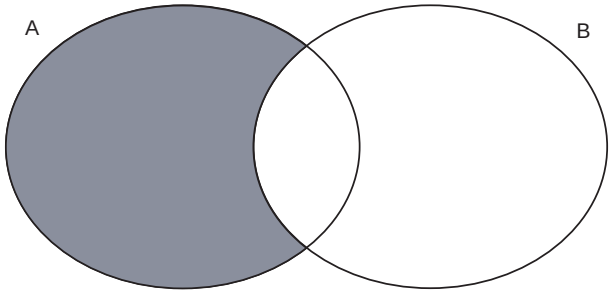


Abbildung 1.3. Differenzmenge $A \setminus B$ von Mengen A und B

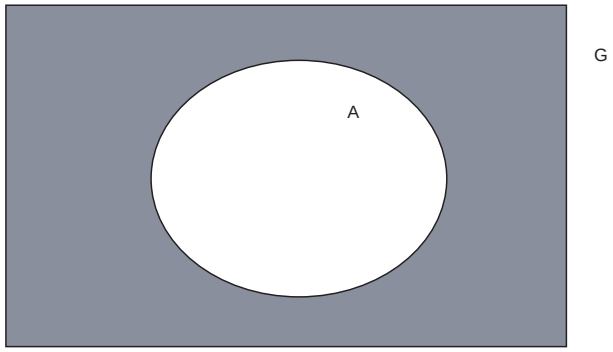


Abbildung 1.4. Komplement $\bar{A}$ von A bezüglich G

Satz 1.37. *(De Morgan)*

$$\begin{aligned} \overline{(A \cup B)} &= (\bar{A} \cap \bar{B}), \\ \overline{(A \cap B)} &= (\bar{A} \cup \bar{B}). \end{aligned} \tag{1.35}$$

Weitere Gesetze: Für beliebige $A, B, C \subset G$ gilt:

Idempotenz

$$\begin{aligned} A \cup A &= A, \\ A \cap A &= A. \end{aligned} \tag{1.36}$$

Kommutativität

$$\begin{aligned} A \cup B &= B \cup A, \\ A \cap B &= B \cap A. \end{aligned} \tag{1.37}$$

Assoziativität

$$A \cup (B \cup C) = (A \cup B) \cup C,$$
$$A \cap (B \cap C) = (A \cap B) \cap C. \tag{1.38}$$

Adjunktivität (Absorption)

$$A \cap (A \cup B) = A,$$
$$A \cup (A \cap B) = A. \tag{1.39}$$

Involution

$$\overline{\overline{A}} = A. \tag{1.40}$$

Identität

$$A \cup \emptyset = A,$$
$$A \cap G = A,$$
$$A \cup G = G,$$
$$A \cap \emptyset = \emptyset. \tag{1.41}$$

Komplementarität

$$A \cup \overline{A} = G,$$
$$A \cap \overline{A} = \emptyset. \tag{1.42}$$

Anmerkung: Vergleicht man diese Gesetzmäßigkeiten mit denen in der Aussagenlogik, so sieht man eine „Eins zu Eins"- Entsprechung, wenn man die aussagenlogischen Operatoren wie folgt mit den Mengenoperationen identifiziert: $\neg$ entspricht $\complement_G$, $\cap$ entspricht $\wedge$, $\cup$ entspricht $\vee$. Auch die Mengenlehre hat die gleiche Struktur wie die Aussagenlogik oder Schaltalgebra (Boolesche Algebra).

Definition 1.38. *Die Menge aller Teilmengen einer Menge A heißt* Potenzmenge *von A.*

$$P(A) = \{B \mid B \subset A\}. \tag{1.43}$$

Beispiel: Es sei $A = \{a, b, c\}$. Dann ist

$$P(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}.$$

Die Potenzmenge $P(A)$ einer Menge mit n Elementen besitzt 2^n Elemente, d. h. ist $|A| = n$, dann ist $|P(A)| = 2^n$.

Beispiel: Sei $A = \emptyset$. Gesucht $P(A)$. Wieviele Elemente hat $P(A)$?

Ein Beispiel: Wenn man ins Kino geht, ist der Sitzplatz im Kinosaal durch das Zahlenpaar $(3, 7)$ eindeutig bestimmt: Reihe 3, Sitz 7. Das Zahlenpaar $(7, 3)$ würde einen anderen Sitzplatz meinen. Auch das Zahlenpaar $(3, 3)$ charakterisiert einen eindeutigen Platz. Für Mengen gilt aber $\{3, 3\} = \{3\}$.

Oft muss daher bei einer Menge von zwei Elementen a, b auch deren Reihenfolge beachtet werden, die bei der Mengenbildung keine Rolle spielt. Dies führt auf die Definition des geordneten Paares (a, b)

Definition 1.39. *Gegeben seien zwei Mengen A, B und $a \in A, b \in B$. Das* geordnete Paar (a, b) *wird definiert durch (Kuratowski)*

$$(a, b) := \{\{a\}, \{a, b\}\}. \tag{1.44}$$

Zur Motivation dieser Definition:
Man betrachte das geordnete Viertupel (b, d, a, c).
Frage: Wie lässt sich dies durch eine Menge charakterisieren?
Wie wär's mit $\{\{b\}, \{b, d\}, \{a, b, d\}, \{a, b, c, d\}\}$? D. h. die Anzahl der Elemente der Teilmenge beinhaltet die Information an welcher Stelle das entsprechende Element steht (wenn man z. B. von links beginnt).

Satz 1.40. *Ist $(a, b) = (a', b')$, dann folgt $a = a'$ und $b = b'$.*

Definition 1.41. *(Alternativdefinition) Gegeben seien zwei Mengen A, B und $a, a' \in A, b, b' \in B$. Das geordnete Paar (a, b) ist durch folgende Eigenschaft charakterisiert:*

Es gilt $(a, b) = (a', b')$ genau dann, wenn $a = a'$ und $b = b'$ ist.

Definition 1.42. *Unter der* Produktmenge *oder dem* Cartesischen Produkt *(Abb. 1.5) zweier Mengen A und B versteht man die Menge aller geordneten Paare, deren erstes Element aus A und deren zweites Element aus B ist*

$$A \times B = \{(a, b) \mid a \in A, b \in B\}. \tag{1.45}$$

Das Cartesische Produkt ist im allgemeinen nicht kommutativ und auch nicht assoziativ, d.h. $A \times B \neq B \times A$ und $(A \times B) \times C \neq A \times (B \times C)$.

Die Anzahl der geordneten Paare ist $|A \times B| = |A||B|$.

Rekursiv kann nun ein *geordnetes n-Tupel* definiert werden, z.B. ein *geordnetes Tripel* als $(r, s, t) = (r, (s, t))$. Damit kann nun das *n-fache Cartesische Produkt* definiert werden. Das *n-fache Cartesische Produkt* der Mengen $A_1, A_2, ..., A_n$ ist in diesem Sinn definiert als

$$A_1 \times A_2 \times ... \times A_n = \{(a_1, ..., a_n) \mid a_1 \in A_1, ..., a_n \in A_n\}.$$

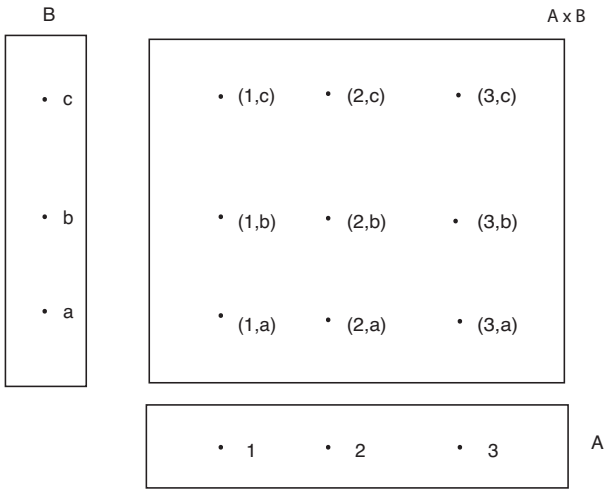


Abbildung 1.5. Cartesisches Produkt $A \times B$ von A und B

Man schreibt für das n**-fache Produkt** $A \times A \times .... \times A$ einer Menge A oft auch abkürzend A^n. Ist $\mathbb{R}$ die Menge der reellen Zahlen, so ist z.B. $\mathbb{R}^3$ die Menge aller reellen 3-Tupel (die als „Punkte“ im 3-dimensionalen Raum veranschaulicht werden können).

Definition 1.43. *Wenn jedem Index λ aus einer* Indexmenge Λ *eine Menge B_λ zugeordnet wird, spricht man von einer* Mengenfamilie

$$B_\lambda, \ \lambda \in \Lambda. \tag{1.46}$$

Definition 1.44. *Unter einer* Klasseneinteilung *(Abb. 1.6) einer Menge X verstehen wir eine Mengenfamilie, d. h. die Menge $\{S_i \mid i \in I, S_i \neq \emptyset, S_i \subset X\}$ für die gilt*

$$(i) \ \bigcup_{i \in I} S_i = X, \ \textit{und}$$
$$(ii) \ S_i \cap S_j = \emptyset, \ \textit{wenn } i \neq j, \ \textit{für alle } i, j \in I. \tag{1.47}$$

Beispiel: Die Menge der ganzen Zahlen $\mathbb{Z}$ kann als Vereinigung der beiden disjunkten nichtleeren Mengen der geraden Zahlen $\mathbb{G}$ und der ungeraden Zahlen $\mathbb{U}$ geschrieben werden $\mathbb{Z} = \mathbb{G} \cup \mathbb{U}$.

1.5.2. Einfache Abzählprinzipien. Wir haben bereits festgehalten das folgendes gilt: die Anzahl der geordneten Paare ist $|A \times B| = |A||B|$.

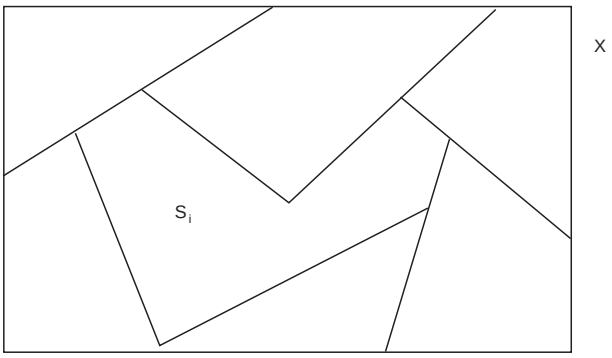


Abbildung 1.6. Klasseneinteilung von X; (Anm.: Der "Rand" wird jeweils nur einmal gezählt)

Satz 1.45. *Es seien A und B disjunkte Mengen, dann gilt*

$$|A \cup B| = |A| + |B| \tag{1.48}$$

Für beliebige nicht notwendigerweise disjunkte Mengen gilt

Satz 1.46. *(Prinzip der Inklusion und Exklusion) Es seien A, B und C beliebige Mengen, dann gilt*

$$(i)\ |A \cup B| = |A| + |B| - |A \cap B|, \tag{1.49}$$

$$(ii)\ |A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|. \tag{1.50}$$

Dieser Satz kann ohne Probleme für n Mengen verallgemeinert werden [**33**], Seite 49.

Zusammenfassung

(i) Folgende Begriffe sollten ihnen nun bekannt sein

Menge, Element, Teilmenge, leere Menge, Kardinalzahl,
Vereinigung, Durchschnitt, Differenz, Komplement, Potenzmenge,
geordnetes Paar, Produktmenge, Klasseneinteilung.

(ii) Sie sollten die wichtigsten Rechenregeln für Mengenoperationen kennen und einfache Aufgaben wie zum Beispiel Abzählprobleme mithilfe von Mengen lösen können.

1.6. Übung

Es sei $\mathbb{Z}^+ = \{1, 2, 3, ...\}$ (die Menge der positiven ganzen Zahlen).

(1) Geben Sie die folgenden Mengen im aufzählenden Verfahren an:
a) $A = \{x \in \mathbb{N} \mid 2 \leq x < 5\}$,
b) $B = \{x \in \mathbb{Z} \mid -2 < x \leq 4\}$,
c) $C = \{x \in \mathbb{N} \mid x < 3\}$,
d) $D = \{x \in \mathbb{Z} \mid -3 < x \leq 3\}$,
e) $E = \{x \in \mathbb{Z} \mid -4 \leq x \leq 1\}$,
f) $F = \{x \in \mathbb{Z} \mid -4 < x < -1\}$,
g) $G = \{x \in \mathbb{Z}^+ \mid x < 10 \ \wedge \ x^2 < 25\}$,
h) $H = \{x \in \mathbb{N} \mid x > 4 \ \wedge \ x < 7\}$,
i) $I = \{x \in \mathbb{N} \mid x < 4 \ \wedge \ x > 7\}$.

(2) Suchen Sie für die folgenden Mengen eine beschreibende Darstellung:
a) $A = \{4, 5, 6\}$,
b) $B = \{-1, 0, 1\}$,
c) $C = \{5, 6, 7, ...\}$,
d) $D = \{0, 1, 2, ...\}$,
e) $E = \{..., -2, -1\}$,
f) $F = \{-3, -2, -1, 0, 1, 2, 3\}$.

(3) Gegeben sind zwei Mengen A und B. Geben Sie ihre Vereinigungsmenge, ihre Durchschnittsmenge sowie die Differenzmengen $A \setminus B$ und $B \setminus A$ an:
a) $A = \{1, 2\}$, $B = \{2, 3, 4\}$,
b) $A = \{4, 5, 6\}$, $B = A$,
c) $A = \{-3, -2, -1, 0, 1, 2\}$, $B = \mathbb{N}$,
d) $A = \{-1, 0, 1, 2\}$, $B = \{3, 4\}$.

(4) Gegeben sind zwei Mengen A und B mit $B \subset A$. Ermitteln Sie die Komplementärmenge von B bezüglich A:
a) $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{3, 4, 6\}$,
b) $A = \{x \in \mathbb{N} \mid x \leq 10\}$, $B = \{1, 3, 5, 7\}$,
c) $A = \mathbb{N}$, $B = \mathbb{Z}^+$,
d) $A = B$.

(5) Was ergibt:
a) $\{1\} \cup \{1\}$, b) $\{1\} \cup \{0\}$, c) $\{\} \cup \{0\}$, d) $\mathbb{Z}^+ \cup \{0\}$,
e) $\mathbb{N} \cup \{0\}$, f) $\mathbb{Z}^+ \cap \{0\}$, g) $\mathbb{N} \cap \{0\}$, h) $\mathbb{Z}^+ \setminus \{0\}$, i) $\{0\} \setminus \mathbb{Z}^+$.

(6) Was ergibt a) $A \cup A$, b) $A \cap A$, c) $A \setminus A$.

(7) Gegeben sind zwei Mengen A und B mit $B \subsetneq A$. Zeichnen Sie ein Mengendiagramm für die Komplementärmenge von B bezüglich A.

(8) Zeichnen Sie ein Mengendiagramm für die Mengen A, B, $A \neq B$, wenn gilt:
a) $A \cup B = A$, b) $A \cap B = A$, c) $A \cap B = \{\}$, d) $A \setminus B = A$.

(9) Zeichnen Sie ein Mengendiagramm für die Mengen A, B, C wenn:
a) $A \cup B = B$ und $B \cap C = \{\}$, b) $A \subsetneq C$ und $B \subsetneq C$ und $A \cap B = \{\}$.

(10) Wahr oder falsch?
a) $\{\} = \{0\}$, b) $\{4,5,9\} = \{5,4,9\}$, c) $\{4,5,9\} \subsetneq \{5,4,9\}$,
d) $\{4,5,9\} \subset \{5,4,9\}$, e) $\{1\} \cup \{1\} = \{2\}$, f) $\{1\} \cap \{1\} = \{1\}$.

(11) Die **symmetrische Differenz** $A \triangle B$ zweier Mengen A, B ist definiert durch $(A \setminus B) \cup (B \setminus A)$. Zeige, dass $A \triangle B = B \triangle A$.

(12) Man stelle die Menge unter der Operation δ zeichnerisch dar:

$$A \delta B := \{x \mid x \in A \veebar x \in B\}. \tag{1.51}$$

(13) Man beweise das Distributivgesetz $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ mittels des beschreibenden Verfahren und auch mit Mengendiagrammen.

(14) Es seien A und B zwei endliche Mengen. Man berechne die Kardinalität von $A \times B$.

(15) Berechne die Kardinalzahlen folgender Mengen:

$M_1 = \emptyset$, $M_2 = \{0, 1, \{2,3\}\}$, $M_3 = \{\mathbb{N}, \mathbb{Z}\}$, $M_4 = \{\{\{1,2\}\}\}$,
$M_5 = \{\emptyset, \{1,2,3,4\}\}$, $M_6 = \{0, (1,2), \{0\}\}$, $M_7 = \{\{1,2,3\}, \{4,4\}\}$
$M_8 = \{(1,2), (2,1), \{1,2\}, \{2,1\}\}$, $M_9 = \{(1,2,3), (4,4), (5,6)\}$.

(16) Beispiel 3.9 (Seite 91) aus [**42**]. Jeder der 100 erstsemestrigen "Computer Science" Studierenden der Utopia-Universität belegt mindestens eines der folgenden Nebenfächer: Mathematik, Elektronik und Buchhaltung. Man weiß, dass 65 Mathematik, 45 Elektronik, 42 Buchhaltung, 20 Mathematik und Elektronik, 25 Mathematik und Buchhaltung und 15 Elektronik und Buchhaltung belegen. Wieviele belegen
(i) alle drei Nebenfächer,
(ii) Mathematik und Elektronik aber nicht Buchhaltung,
(iii) nur Elektronik?

(17) Die Abbildung 1.7 zeigt vier Mengendiagramme für zwei Mengen A und B.
Markieren Sie in jedem dieser Diagramme die folgenden Mengen:
a) $A \cup B$, b) $A \cap B$, c) $A \setminus B$, d) $B \setminus A$.

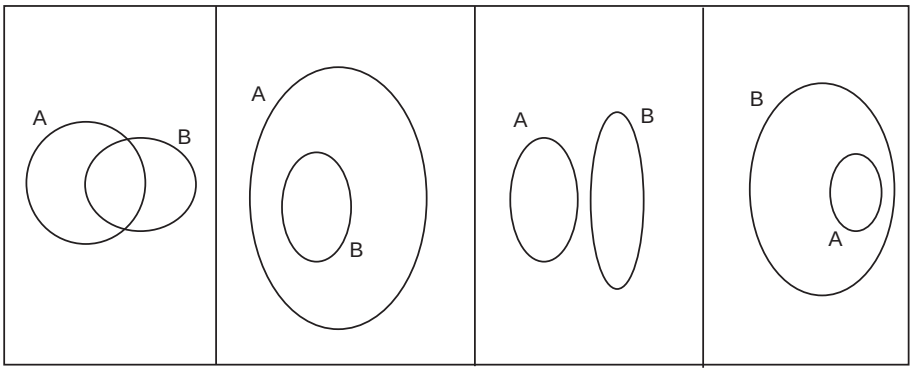


Abbildung 1.7. Beispiel

(18) (*) Man beweise den Satz 1.40;
Hinweis: Fallunterscheidung $a = b$ und $a \neq b$.

(19) (*) Es sei $A_j \subset G$. Dann gilt für jede beliebige endliche, abzählbare oder auch nicht abzzählbare Indexmenge I:

Satz 1.47. *(De Morgan)*

$$\overline{(\bigcup_{j\in I} A_j)} = \bigcap_{j\in I} \bar{A}_j,$$
$$\overline{(\bigcap_{j\in I} A_j)} = \bigcup_{j\in I} \bar{A}_j. \tag{1.52}$$

(20) (*) Man beweise (1.50) und verwende dazu (1.49), d. h. es seien A, B und C beliebige Mengen, dann gilt

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Kapitel 2

Relationen

2.1. Relationen

Gegeben seien zwei Mengen A, B. Man betrachtet das Cartesische Produkt $A \times B$.

Definition 2.1. *Es sei R eine Teilmenge von $A \times B$, d. h. $R \subset A \times B$. Das geordnete Tripel (A, B, R) heißt* Relation*. R heißt der* Graph der Relation*. Ist $(a, b) \in R$ so steht a zu b in Relation und man schreibt $a\,R\,b$. Wenn $(a, b) \notin R$ ist, so schreibt man auch $a \not R b$.*

Ist $B = A$ so nennt man $(A, R) = (A, A, R)$ Relation auf A.

Anmerkung: Meist wird („schlampigerweise“) bereits die Menge R selbst als Relation bezeichnet.

Anmerkung: Betrachtet man mehr als zwei Mengen also n Mengen und ist R eine Teilmenge von $A_1 \times A_2 \times \ldots A_n$, so spricht man von *$n$-stelligen* Relationen.

Definition 2.2. *Der* Definitionsbereich einer Relation (A, B, R) *ist definiert durch*

$$D(R) = \{a \mid (a, b) \in R\} \tag{2.1}$$

und der Wertebereich *(Wertevorrat) einer Relation (A, B, R) ist definiert durch*

$$W(R) = \{b \mid (a, b) \in R\}. \tag{2.2}$$

Angabe von Relationen.

(1) Beschreibung durch Aufzählung der Paare $R = \{(a,b)\big|\ \ldots\}$,
z. B. $A = B = \{1,2,3,4\}$ und $R = \{(1,2),(2,4),(3,3),(4,2),(4,4)\}$.

(2) Angabe einer Relation durch Beschreibung der Beziehung (Prädikat).
Beispiele:
(i) $A = B = \mathbb{N}$ und $a\,R\,b$, z.B. $a = b$, $a \leq b$, $a|b$ (a teilt b)[1] ,
(ii) $A = B = \{x \mid x \text{ ist ein Land}\}$ und $a\,R\,b$, wenn das Land a an das Land b angrenzt,
(iii) $A = B = \{x \mid x \text{ ist eine Stadt}\}$ und $a\,R\,b$, wenn die Stadt a eine Zugverbindung zur Stadt b hat,
(iv) $A = \{\text{Milch, Eier, Honig}\}$, $B = \{\text{Kühe, Ziegen, Hühner}\}$ und $a\,R\,b$, wenn das Produkt a von b erzeugt wird,
d.h. $R = \{(\text{Eier,Hühner}),(\text{Milch,Ziegen}),(\text{Milch,Kühe})\}$
(v) $A = \{\text{Prozessor, iMac, Drucker, Bildschirme}\}$, $B = \{\text{Apple, IBM, Sun}\}$ und $a\,R\,b$, wenn das Produkt a von b erzeugt wird.

Darstellung von Relationen.

Im Folgendem sei $A = \{a,b,c\}$, $B = \{u,x,y,z\}$ und
$R = \{(a,x),(b,y),(b,z),(c,z)\}$, d. h. aRx, bRy, bRz, cRz.

(1) Matrixdiagramm: (besonders für die Darstellung am Computer geeignet)

$$m_{jk} = \begin{cases} 1 & a_j R\, b_k \\ 0 & a_j \not R\, b_k \end{cases}$$

(dabei bezeichnet j die Zeile und k die Spalte).

$$\begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{cccc} u & x & y & z \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}. \tag{2.3}$$

(2) Pfeildiagramm (Abb. 2.1):

(3) Angabe durch eine Tabelle (dies ist die Standardform für mehrstellige Relationen, die in Datenbanken verwendet werden.

[1] Definition: Gegeben seien zwei Zahlen $a, b \in \mathbb{N}$. Dann gilt $a|b$ (a teilt b) genau dann, wenn eine Zahl $q \in \mathbb{N}$ existiert, sodass $aq = b$ gilt

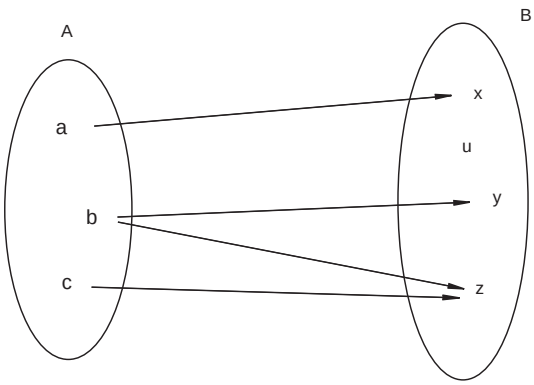


Abbildung 2.1. Pfeildiagramm

$D(R) \subset A$	$W(R) \subset B$
a	x
b	y
b	z
c	z

(4) Koordinatendiagramm (Abb. 2.2):

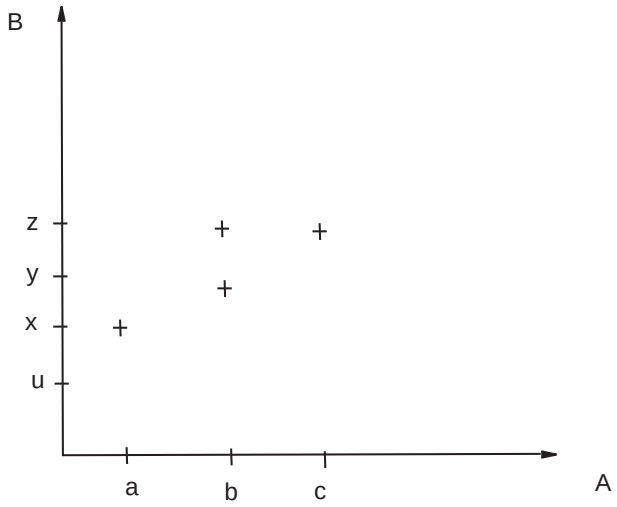


Abbildung 2.2. Koordinatendiagramm

(5) Gerichteter Graph (Abb. 2.3):
(Diese Darstellung ist nur für Relationen (A, R) auf A möglich)

$A = \{1, 2, 3, 4\}$,
$R = \{(1, 2), (2, 2), (2, 4), (3, 2), (3, 4), (4, 1), (4, 3)\}$.

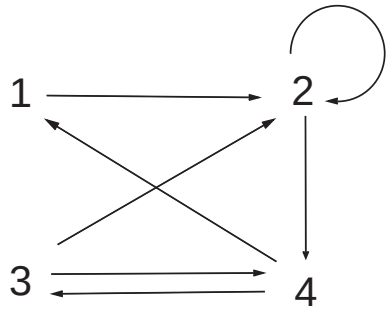


Abbildung 2.3. Gerichteter Graph

Übung: Es sei $A = B = \{1, 2, 3, 4, 5, 6\}$ und $a\,R\,b$, wenn $a|b$ (a teilt b). Man stelle diese Relation in den Formen (1), (3)-(7) dar.

Definition 2.3. *Die* inverse Relation *(Abb. 2.4)* (B, A, R^{-1}) *zu* (A, B, R) *ist definiert durch*

$$R^{-1} = \{(b, a) \mid (a, b) \in R\}. \tag{2.4}$$

Übung: Es sei $A = \{a, b, c\}$, $B = \{u, x, y, z\}$ und $R = \{(a, x), (b, y), (b, z), (c, z)\}$. Man ermittle die inverse Relation R^{-1} und stelle sie mit einem Pfeildiagramm dar.

Definition 2.4. *Die* zusammengesetzte Relation *(Abb. 2.5)* $(A, C, R \circ S)$. *Ist* $R \subset A \times B$ *und* $S \subset B \times C$, *so ist die zusammengesetzte Relation definiert durch*

$$R \circ S = \{(a, c) \mid \textit{es existiert ein } b \textit{ mit } (a, b) \in R \textit{ und } (b, c) \in S\}. \tag{2.5}$$

Anmerkung: Manche benutzen auch die Notation $S \circ R$ (R und S sind vertauscht).

Beispiel in Abbildung 2.5:
Es sei $A = \{1, 2, 3, 4\}$, $B = \{a, b, c, d\}$, $C = \{x, y, z\}$,
$R = \{(1, a), (2, d), (3, a), (3, b), (3, d)\}, S = \{(b, x), (b, z), (c, y), (d, z)\}$.
Dann ist $R \circ S \subset A \times C$ und $R \circ S = \{(2, z), (3, x), (3, z)\}$.

Wie wir schon bei den bisherigen Beispielen gesehen haben, besitzen Relationen gewisse Eigenschaften, die in folgender Definition präzisiert werden.

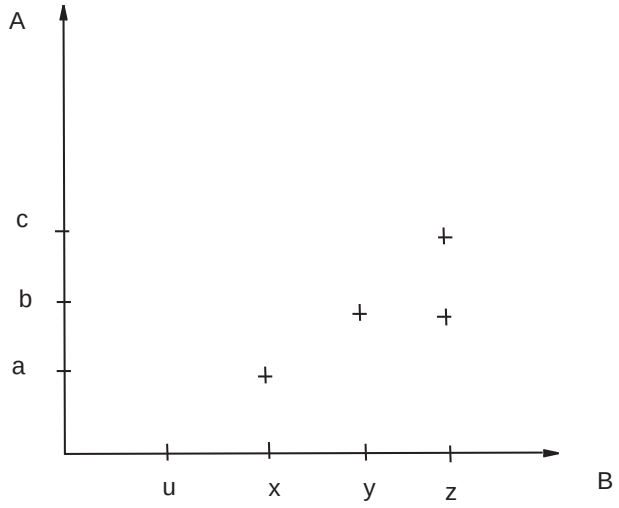


Abbildung 2.4. Inverse Relation R^{-1}

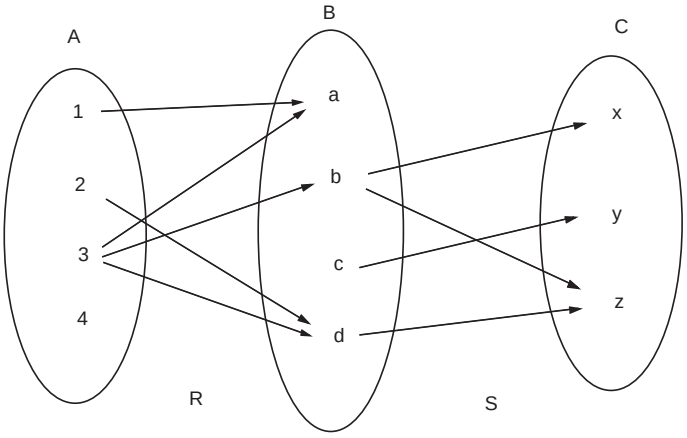


Abbildung 2.5. Zusammengesetzte Relation $R \circ S$

Definition 2.5. *Eine Relation (A, R) auf A heißt*

(i) reflexiv, *wenn aRa für alle $a \in A$ gilt,*

(ii) symmetrisch, *wenn für alle $a, b \in A$ mit aRb folgt bRa,*

(iii) antisymmetrisch, *wenn für alle $a, b \in A$, für die aRb und bRa gilt, folgt, dass $a = b$,*

(iv) transitiv, *wenn für alle $a, b, c \in A$ mit aRb und bRc folgt aRc.*

Oft ist es leichter festzustellen, dass eine Eigenschaft nicht erfüllt ist:

(i) nicht reflexiv: es existiert ein $a \in A$ mit $(a, a) \notin R$.
(ii) nicht symmetrisch: es existiert ein $(a, b) \in R$ mit $(b, a) \notin R$.
(iii) nicht antisymmetrisch: es existiert ein $a \neq b \in A$ mit (a, b) und $(b, a) \in R$.
(iv) nicht transitiv: es existiert $(a, b), (b, c) \in R$ mit $(a, c) \notin R$.

Anmerkungen: Es gilt

(i) Ist die Relation (A, R) transitiv, so ist auch die inverse Relation (A, R^{-1}) transitiv.

(ii) Die Zusammensetzung von symmetrischen bzw. transitiven Relationen muss aber nicht symmetrisch bzw. transitiv sein.

(iii) Antisymmetrisch ist nicht das Gegenteil von symmetrisch, z. B. $A = B = \{a, b\}$ und $R = \{(a, a), (b, b)\}$. Diese Relation ist zugleich symmetrisch und antisymmetrisch.

(iv) Die Relation auf der leeren Menge ist als einzige Relation sowohl reflexiv als auch irreflexiv.

Übung: Man ermittle die Eigenschaften folgender Relationen.

(i) $A = \mathbb{N},\ aR_1b$, wenn $a \leq b$.

(ii) $A = \mathbb{N},\ aR_2b$, wenn $a < b$, d. h. $(\mathbb{N}, <)$, wobei $<$ „kleiner als“ bedeutet.

(iii)
$$A = \{1, 2, 3, 4\},$$
$$R_3 = \{(1, 1), (2, 4), (3, 3), (4, 1), (4, 4)\}.$$

(iv)
$$A = \{a, b, c\},$$
$$R_4 = \{(a, a), (c, b), (b, a), (a, c)\}.$$

Zwei weitere oft verwendete Eigenschaften von Relationen sind definiert durch

Definition 2.6. *Eine Relation (A, R) heißt*

(i) irreflexiv, *wenn $a \not R a$ für alle $a \in A$ gilt,*

(ii) asymmetrisch, *wenn aRb impliziert, dass $b \not R a$ für alle $a, b \in A$ gilt.*

Beispiel: $(\mathbb{N}, <)$.

Satz 2.7. *Der Durchschnitt einer Familie (Menge) von reflexiven, symmetrischen oder transitiven Relationen $R_j, j \in I$ auf einer festen Menge A ist reflexiv, symmetrisch oder transitiv.*

Satz 2.8. *Zu jeder Relation R gibt es eine* reflexive, symmetrische oder transitive Hülle, *d. h. eine Relation $\tilde{R}$, wobei $R \subset \tilde{R}$ und $\tilde{R}$ reflexiv, symmetrisch oder transitiv ist.*

Beispiel: Die reflexive Hülle $\tilde{R}_3$ von R_3 ist: $\tilde{R}_3 = R_3 \cup \{(2,2)\}$.

Definition 2.9. *Die Menge $I_A \subset A \times A$ gegeben durch $I_A = \{(x,x) \mid x \in A\}$ heißt* Diagonale *von $A \times A$. Durch I_A ist die* Identitätsrelation *definiert. Es gilt $x\, I_A\, y$ genau dann, wenn $x = y$.*

Anmerkung: Eine Relation R ist genau dann reflexiv, wenn $I_A \subset R$.

Definition 2.10. *Eine Relation auf A heißt* Äquivalenzrelation*, wenn gilt*

(i) *R ist reflexiv,*

(ii) *R ist symmetrisch,*

(iii) *R ist transitiv.*

Durch eine Äquivalenzrelation auf A erhält man eine *Partition* oder *Klasseneinteilung* von A. Die Mengen

$$[x] = \{y \in A \mid y\, R\, x\} \tag{2.6}$$

bezeichnet man als Äquivalenzklassen von A bezüglich R und x ist ein beliebiges Element (Vertreter) aus dieser Klasse. Für die Äquivalenzklassen gilt:

(1) $[x] \neq \emptyset$,

(2) $[x] \cap [y] = \emptyset$ genau dann, wenn $(x,y) \notin R$,

(3) $\bigcup_{x \in A} [x] = A$.

Beispiele. (i) Es sei $A = \mathbb{Z}$ und die Relation R sei definiert durch

$$(m,n) \in R \text{ genau dann, wenn 2 die Differenz } m - n \text{ teilt .}$$

Dies ist eine Äquivalenzrelation. Die zwei Äquivalenzklassen bestehen aus den geraden und ungeraden natürlichen Zahlen.

(ii) Es sei $A = \mathbb{N}$ oder $A = \mathbb{Z}$ und die Relation R sei definiert durch $a\, R\, b$ wenn a den gleichen Rest wie b bei der Division durch m hat.[2]

Weitere Beispiele für Äquivalenzrelationen sind: „a ist gleich alt wie b“ auf einer Menge von Personen, „a kostet gleich viel wie b“ auf einer Menge von Produkten, „Seite a gehört zum selben Kapitel wie Seite b“ auf der Menge aller Seiten eines Buches, usw. In diesen Beispielen erkennt man sofort auch die entsprechenden Äquivalenzklassen. So ergeben sich Altersklassen, Preisklassen, die einzelnen Kapitel eines Buches, usw.

[2] Siehe auch Definition 5.15.

2.2. Abbildungen

Definition 2.11. *Eine Relation (A, B, f) mit den beiden Eigenschaften*

(i) $D(f) = A$,

(ii) *aus xfy und xfz folgt $y = z$*
(bewirkt Eindeutigkeit, d. h. zu jedem x gibt es nur ein y)

heißt Abbildung *oder* Funktion *f von A in B.*

Anmerkung: Sehr gerne wird für eine Funktion auch folgende Definition verwendet: Eine Abbildung oder Funktion f von einer Menge A in eine Menge B ist eine Vorschrift, die jedem Element $x \in A$ genau ein Element $f(x) \in B$ zuordnet. Hier müsste man aber dann definieren, was eine "Vorschrift" ist oder was "zuordnet" bedeutet, ansonsten ist dies eine sinnlose Aussage.

Schreibweise

$$f : \ A \to B, \ x \mapsto f(x). \tag{2.7}$$

$x \in A$ heißt *Argument* der Funktion, das dazugehörende $y = f(x)$ das *Bild* von x. A heißt *Definitionsbereich*, $f(A) = \{y \in B \mid y = f(x), x \in A\}$ *Wertevorrat* (Abb. 2.6) von f.

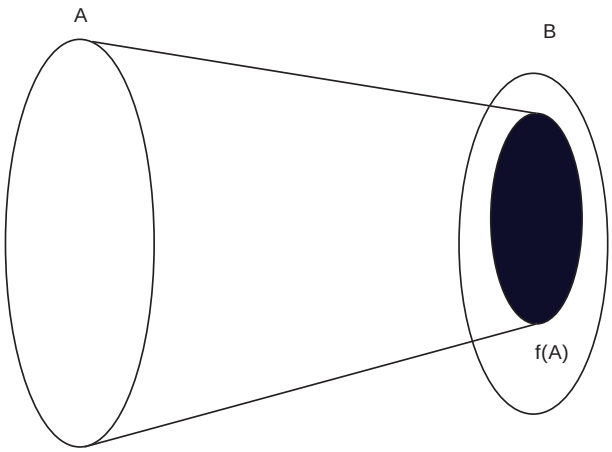


Abbildung 2.6. Wertevorrat $f(A)$

Die Menge $\{(x, f(x)) \mid x \in A\}$ heißt *Graph der Funktion.*

Darstellung von Funktionen

$A = \{a, b, c, d, e\}, \ B = \{\alpha, \beta, \gamma, \delta, \epsilon\}, \ f = \{(a, \beta), (b, \alpha), (c, \delta), (d, \delta), (e, \delta)\}$

(1) Angabe durch eine Tabelle

A	$f(A) \subset B$
a	β
b	α
c	δ
d	δ
e	δ

(2) Pfeildiagramm (Abb. 2.7)

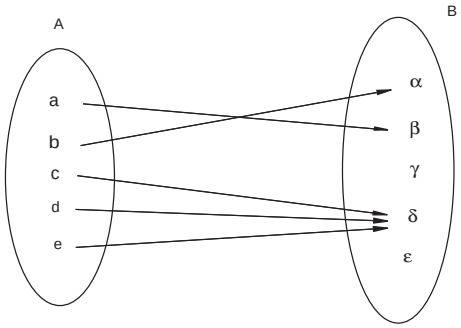


Abbildung 2.7. Pfeildiagramm

(3) Koordinatendiagramm (Abb. 2.8)

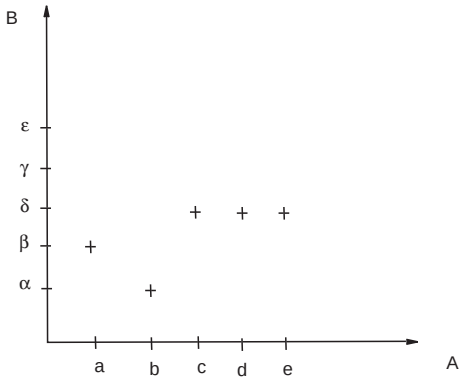


Abbildung 2.8. Koordinatendiagramm

(4) Formel

$$A = B = \mathbb{R}, \quad f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto x^2 + 4x - 7 = (x+2)^2 - 11. \tag{2.8}$$

Definition 2.12. *Eine Abbildung heißt* surjektive Abbildung *oder Abbildung auf B, wenn $B = f(A)$ gilt.*

Eine Abbildung heißt injektive (eindeutige) Abbildung*, wenn aus $f(x) = f(y)$ $x = y$ folgt.*

Eine surjektive und injektive Abbildung heißt bijektive Abbildung *(Abb. 2.9).*

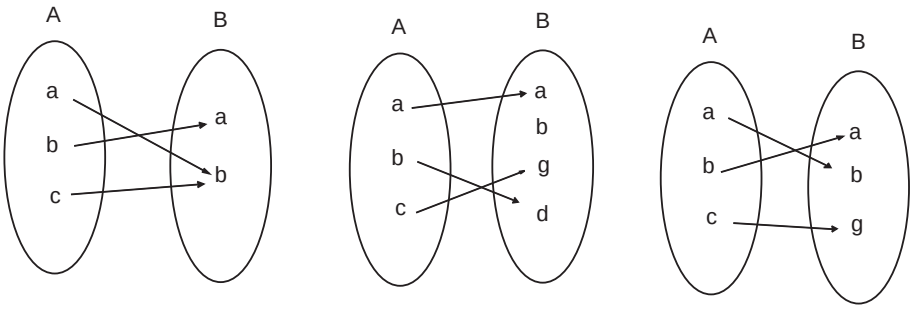


Abbildung 2.9. Surjektiv, injektiv, bijektiv

Spezielle Funktionen sind

(i) *Identität* $\mathrm{Id}_A : A \to A,\ x \mapsto x$,

(ii) *Inklusion* $i : A \to B$, mit $A \subset B,\ x \mapsto x$,

(iii) *Projektion* $\mathrm{pr}_1 : A \times B \to A,\ (x, y) \mapsto x,\ \mathrm{pr}_2 : A \times B \to B,\ (x, y) \mapsto y$,

(iv) $f : A \to B$ heißt *konstante* Funktion, wenn für alle $x_1, x_2 \in A$ gilt $f(x_1) = f(x_2)$.

(v) Eine Abbildung $\mathbb{N} \to B, n \mapsto x(n) = x_n$ heißt *Folge*.

(vi) Die *Charakteristische Funktion* (auch Indikatorfunktion genannt) χ_A (bzw. I_A) einer Menge $A \subset B$ ist definiert durch $\chi_A : B \to \{0, 1\}$ mit $\chi_A(x) = 1$, wenn $x \in A$ und $\chi_A(x) = 0$, wenn $x \notin A$.

(vii) Die *Kronecker* δ Funktion ist definiert durch $\delta : \mathbb{N} \times \mathbb{N} \to \{0, 1\}$ mit $\delta_{i,j} = 1$, wenn $i = j$ und $\delta_{i,j} = 0$, wenn $i \neq j$.

(viii) Die Polynom-Funktion p vom Grade n ist definiert durch

$$p : \mathbb{R} \to \mathbb{R},$$
$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0, \quad a_j \in \mathbb{R}, a_n \neq 0. \tag{2.9}$$

(ix) Sind f und g zwei Polynome, so bezeichnet man die Funktion $r : D \to \mathbb{R}$, $D = \{x \in \mathbb{R} \mid g(x) \neq 0\}$

$$r(x) = \frac{f(x)}{g(x)}$$

als *rationale Funktion.*

Da (A, B, f) eine Relation ist, gibt es auch dazu eine inverse Relation (B, A, f^{-1}). Diese ist im allgemeinen jedoch keine Abbildung mehr. Nur wenn f eine bijektive Abbildung ist, ist auch f^{-1} eine bijektive Abbildung.

Ist $f : A \to B$ und $C \subset B$, so nennt man die Menge $f^{-1}(C)$ das *Urbild* (Abb. 2.10) von C, d. h. $f^{-1}(C) = \{x \in A \mid f(x) \in C\}$.

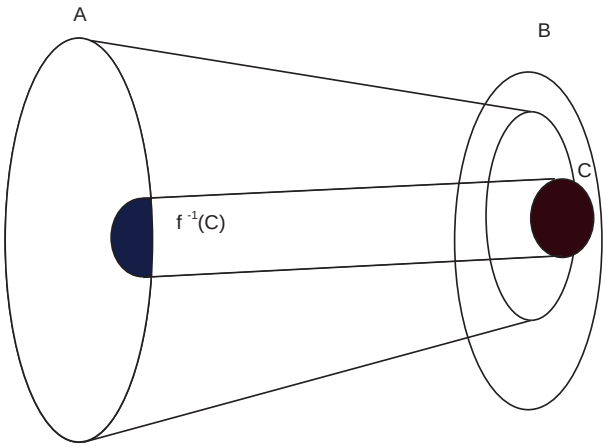


Abbildung 2.10. Urbild $f^{-1}(C)$

Definition 2.13. *Es seien $f : A \to B$, $g : B \to C$ zwei Funktionen. Die* zusammengesetzte Funktion *ist gegeben durch $g \circ f : A \to C$, $x \mapsto g(f(x))$.*

Es gilt $h \circ (g \circ f) = (h \circ g) \circ f$ aber $g \circ f \neq f \circ g$.

Übung: $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x + 1$ und $g : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$. Berechne $g \circ f$ und $f \circ g$.

Definition 2.14. *Ist $f : A \to B$, eine Abbildung und $C \subset A$, so nennt man $f\big|_C : C \to B$, $x \mapsto f(x)$ die* Restriktion *von f auf C. Es gilt $f\big|_C = f \circ i$.*

2.3. Mächtigkeiten

Definition 2.15. *Zwei Mengen A und B heißen* gleichmächtig *(haben dieselbe* Kardinalzahl*), wenn es eine bijektive Abbildung von A nach B gibt.*

Die Menge $\{0, 1, \ldots, n-1\}$ nennt man Abschnitt der natürlichen Zahlen.

Definition 2.16. *Ist A gleichmächtig mit einem Abschnitt $\{0, 1, \ldots, n-1\}$ der natürlichen Zahlen, so heißt A* endliche Menge *und n ist die* Kardinalzahl *von A (= Anzahl der Elemente von A)*

Definition 2.17. *Ist A nicht endlich, aber gleichmächtig mit $\mathbb{N}$, so heißt A* abzählbar unendlich*. Man schreibt $|\mathbb{N}| = \aleph_0$ (aleph null).*

Die Potenzmenge einer Menge hat immer eine größere Mächtigkeit als die Menge selbst, $|P(\mathbb{N})| := \aleph_1$ (aleph eins).

Endliche und abzählbar unendliche Mengen fasst man zu den *abzählbaren* Mengen zusammen. Nicht abzählbare Mengen heißen *überabzählbar*.

Satz 2.18. *Es sei A eine abzählbare Menge und $f : A \to B$ eine surjektive Abbildung. Dann ist B abzählbar.*

Es gilt:

(i) Jede Teilmenge von $\mathbb{N}$ ist abzählbar.

(ii) Jede Teilmenge einer abzählbaren Menge ist abzählbar.

(iii) Die Menge der rationalen Zahlen $\mathbb{Q}$ ist abzählbar (Diagonalverfahren).

$$
\begin{array}{llllllll}
0 & \frac{1}{1} & -\frac{1}{1} & \frac{2}{1} & -\frac{2}{1} & \frac{3}{1} & -\frac{3}{1} & \cdots \\
 & \frac{1}{2} & -\frac{1}{2} & \frac{2}{2} & -\frac{2}{2} & \frac{3}{2} & \cdots & \\
 & \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & -\frac{2}{3} & \cdots & & \\
 & \frac{1}{4} & -\frac{1}{4} & \frac{2}{4} & \cdots & & & \\
 & \frac{1}{5} & \cdots & & & & & \\
 & \vdots & & & & & &
\end{array}
\tag{2.10}
$$

(iv) Die Vereinigung von abzählbar vielen abzählbaren Mengen ist abzählbar.

$$\begin{array}{lllll} a_{11} & a_{12} & a_{13} & a_{14} & \dots \\ a_{21} & a_{22} & a_{23} & \dots & \\ a_{31} & a_{32} & \dots & & \\ a_{41} & \dots & & & \\ \vdots & & & & \end{array} \tag{2.11}$$

Nach diesen Beispielen könnte man vermuten, dass jeden Menge abzähbar ist. Es gilt jedoch:

Satz 2.19. *Die Menge der reellen Zahlen* $\{x \in \mathbb{R} \mid x \in [0,1]\}$ *und folglich die Menge* $\mathbb{R}$ *ist überabzählbar. Man schreibt* $|\mathbb{R}| = c$.

Anmerkung: Die Kontinuumshypothese besagt, dass $c = \aleph_1$.

Beweis: Annahme $[0,1]$ ist abzählbar, d. h.

$$\begin{aligned} 0 &\leftrightarrow r_0 = 0.z_{00}z_{01}z_{02}z_{03}\dots, \\ 1 &\leftrightarrow r_1 = 0.z_{10}z_{11}z_{12}z_{13}\dots, \\ 2 &\leftrightarrow r_2 = 0.z_{20}z_{21}z_{22}z_{23}\dots. \\ &\vdots \end{aligned}$$

Man bilde die Zahl

$$r = 0.\hat{z}_0\hat{z}_1\hat{z}_2\dots, \text{ wobei } \hat{z}_j = 2, \text{ wenn } z_{jj} = 1 \text{ und } \hat{z}_j = 1 \text{ sonst.}$$

Dann kommt $r \in [0,1]$ nicht in der Menge $\{r_0, r_1, r_2, \dots\}$ vor. W!

2.4. Geordnete Mengen

Definition 2.20. *Eine Relation auf A heißt* Ordnungsrelation*, wenn gilt*

(i) *R ist reflexiv,*

(ii) *R ist antisymmetrisch,*

(iii) *R ist transitiv.*

(A, R) *heißt geordnete Menge.*

xRy wird üblicherweise bei Ordnungsrelation als $x \leq y$ (*kleiner gleich*) geschrieben und xRy mit $x \neq y$ als $x < y$ (*echt kleiner*), $(A, \leq)$ geordnete Menge.

Bei einer Ordnung kann es unvergleichbare Elemente geben. Man spricht daher auch von einer *Teilordnung* oder *partiellen Ordnung* (partially ordered set (POset)).

Ist $(A, \leq)$ eine Ordnungsrelation, dann ist auch die inverse Relation eine Ordnungsrelation.

Definition 2.21. *Eine Relation auf A heißt* strikte Ordnungsrelation*, wenn gilt*

(i) *R ist irreflexiv,*

(ii) *R ist asymmetrisch,*

(iii) *R ist transitiv.*

Anmerkung: Man erhält aus einer Ordnungsrelation eine strikte Ordnungsrelation, wenn man die Diagonale entfernt und umgekehrt aus einer strikten Ordnungsrelation eine Ordnungsrelation, wenn man die Diagonale hinzufügt.

Beispiele für Ordnungsrelationen

(1) $\mathbb{N}$ mit der natürlichen *kleiner-gleich*-Beziehung $(\mathbb{N}, \leq)$,

(2) Potenzmenge einer Menge A mit Inklusion $\subset$ als Ordnungsrelation,

(3) $(\mathbb{N}, \preceq)$ wobei $\preceq$ folgenderweise definiert wird $a \preceq b$ genau dann, wenn a teilt b,

(4) $A_1 = \{1, 2, 3, 4, 6, 12\}$, $a \leq b$ ist definiert als a teilt b.

Beispiel (4) wird in einem *Hasse-Diagramm* (Ordnungsdiagramm) wie folgt dargestellt (Abb. 2.11)

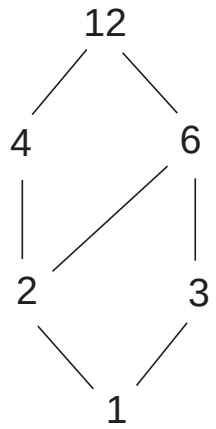


Abbildung 2.11. Hasse-Diagramm

Jedem Element der Menge wird ein Punkt der Zeichenebene zugeordnet mit der Vereinbarung, b oberhalb von a zu zeichnen und mit a zu verbinden, wenn $a \leq b$

gilt. Daher können die Pfeilspitzen weggelassen werden. Durch die zusätzliche Vereinbarung, dass b nicht mehr mit a verbunden wird, wenn b bereits über andere Punkte mit a verbunden ist (Transitivität), wird die Fülle der Linien reduziert (vergl. Zahlengerade).

Übung: Man zeichne das Hasse-Diagramm für folgende Ordnungsrelationen:

(1) $A_2 = \{1, 2, 3, 4, 6, 7, 12, 13, 49\}$, $a \leq b$ ist definiert als a teilt b,

(2) $A_3 = \{2, 3, 4, 6, 7, 12, 13, 49\}$, $a \leq b$ ist definiert als a teilt b,

(3) $\mathbb{N}$ mit der natürlichen *Kleiner*-Beziehung $\leq$,

(4) $\mathbb{Z}$ mit der natürlichen *Kleiner*-Beziehung $\leq$.

Definition 2.22. *Eine Relation $(A, \leq)$ heißt* totalgeordnet*, wenn für alle $(x, y) \in A \times A$ entweder $x \leq y$ oder $y \leq x$ gilt.*

Eine totalgeordnete Menge wird auch als *linear geordnet* oder als *Kette* bezeichnet.

Von obigen Übungsbeispielen ist nur Beispiel (3) und (4) totalgeordnet.

Definition 2.23. *Eine Abbildung $f : (A, \leq_1) \to (B, \leq_2)$ heißt* Ordnungshomomorphismus*, wenn*

$$x \leq_1 y \to f(x) \leq_2 f(y). \tag{2.12}$$

(f wird auch als isotone *Abbildung, als* monoton wachsend *bzw. als nicht fallend bezeichnet).*

Anmerkung: Üblicherweise wird der Index j bei $\leq_j$ weggelassen.
Eine Abbildung $f : (A, \leq) \to (B, \leq)$ heißt *monoton fallend*, (*antitone Abbildung*), wenn $x \leq y \to f(x) \geq f(y)$.

Eine Abbildung $f : (A, \leq) \to (B, \leq)$ heißt *streng monoton wachsend*, wenn $x < y \to f(x) < f(y)$.

Beispiele

(i) $\mathrm{Id}_A : (A, \leq) \to (A, \leq)$ ist isoton,

(ii) $\mathrm{Id}_A : (A, \leq) \to (A, \geq)$ ist antiton.

Definition 2.24. *Es sei $(A, \leq)$ eine geordnete Menge. Dann heißt*

(i) $[a, b] = \{x \mid x \in A, a \leq x \leq b\}$ abgeschlossenes Intervall,

(ii) $(a, b) = \{x \mid x \in A, a < x < b\}$ offenes Intervall,

(iii) $[a, b) = \{x \mid x \in A, a \leq x < b\}$ halboffenes Intervall,

(iv) $(a, b] = \{x \mid x \in A, a < x \leq b\}$ halboffenes Intervall.

Beispiel: $A = \mathbb{R}$.

Satz 2.25. *Es sei $(A, \leq)$ eine totalgeordnete Menge und $f : (A, \leq) \to (B, \leq)$ ein surjektiver Ordnungshomomorphismus. Dann gilt*

(i) *f bildet (abgeschlossene) Intervalle von A auf (abgeschlossene) Intervalle von B ab.*

(ii) *Ist f streng monoton wachsend, so werden durch f offene Intervalle von A auf offene Intervalle von B abgebildet.*

(iii) *Die Endpunkte des Intervalles A werden durch f auf die Endpunkte des Bildintervalles in B abgebildet.*

Übung: Man zeichne eine monotone Funktion $\mathbb{R} \to \mathbb{R}$ die ein offenes Intervall auf ein nicht offenes abbildet.

Definition 2.26. *Es sei $(A, \leq)$ eine geordnete Menge und $M \subset A$. Dann heißt*

(i) $a \in M$ kleinstes Element (Minimum) *von M, wenn für alle $x \in M$ gilt $a \leq x$,*

(ii) $m \in M$ minimales Element *von M, wenn für $x \in M$ gilt $x \leq m$ impliziert $x = m$.*

Anmerkung: Das kleinste Element ist minimal (die Umkehrung gilt nicht!). Das kleinste Element steht mit jedem anderen Element von M in einer Ordnungsbeziehung, nicht aber das minimale. Eine geordnete Menge kann mehrere minimale Elemente besitzen, muss jedoch kein kleinstes besitzen.

Beispiele

(i) In $(\mathbb{N}, \leq)$ ist 0 das kleinste Element.

(ii) In $(\mathbb{N} \setminus \{0, 1\}, \preceq)$ (wobei $a \preceq b$ heißt a teilt b) gilt: Jede Primzahl ist minimal.

(iii) Für die Potenzmenge P(A) mit der Relation "ist Teilmenge" $(P(A), \subset)$ ist $\emptyset$ kleinstes Element.

Satz 2.27. *Es sei $(A, \leq)$ eine geordnete Menge und $M \subset A$. Ist a kleinstes Element von M, so ist a das einzige minimale Element von M. Das kleinste Element ist eindeutig bestimmt.*

Definition 2.28. *Es sei $(A, \leq)$ eine geordnete Menge und $M \subset A, M \neq \emptyset$. Dann heißt*

(i) $a \in M$ größtes Element (Maximum) *(Abb. 2.12) von M, wenn für alle $x \in M$ gilt $x \leq a$,*

(ii) $m \in M$ maximales Element *(Abb. 2.12) von M, wenn für $x \in M$ gilt $m \leq x$ impliziert $x = m$.*

Beispiele

(i) $(\mathbb{N}, \leq)$ besitzt kein maximales und kein größtes Element.

(ii) Für die Potenzmenge P(A) mit der Relation "ist Teilmenge" $(P(A), \subset)$ ist A größtes Element.

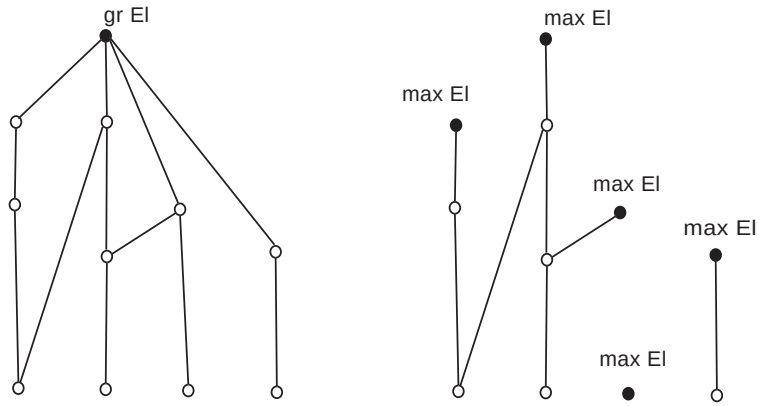


Abbildung 2.12. Größtes Element und maximales Element

Satz 2.29. *Ist M totalgeordnet, so ist jedes minimale (maximale) Element auch kleinstes (größtes) Element.*

Definition 2.30. *Es sei $(A, \leq)$ eine geordnete Menge und $M \subset A$. Dann heißt*

(i) $a \in A$ untere Schranke (Minorante) *von M, wenn für alle $x \in M$ gilt $a \leq x$,*

(ii) $b \in A$ obere Schranke (Majorante) *(Abb. 2.13) von M, wenn für alle $x \in M$ gilt $x \leq b$.*

(iii) *Besitzt M eine untere (obere) Schranke, so heißt M nach unten (oben)* beschränkt. *Eine nach oben und unten beschränkte Menge heißt beschränkt.*

Definition 2.31. *Es sei $(A, \leq)$ eine geordnete Menge und $M \subset A, M \neq \emptyset$. Dann heißt*

(i) $a \in A$ untere Grenze (Infimum) *von M, wenn a die größte untere Schranke ist.*

(ii) $b \in A$ obere Grenze (Supremum) *(Abb. 2.13) von M, wenn b die kleinste obere Schranke ist.*

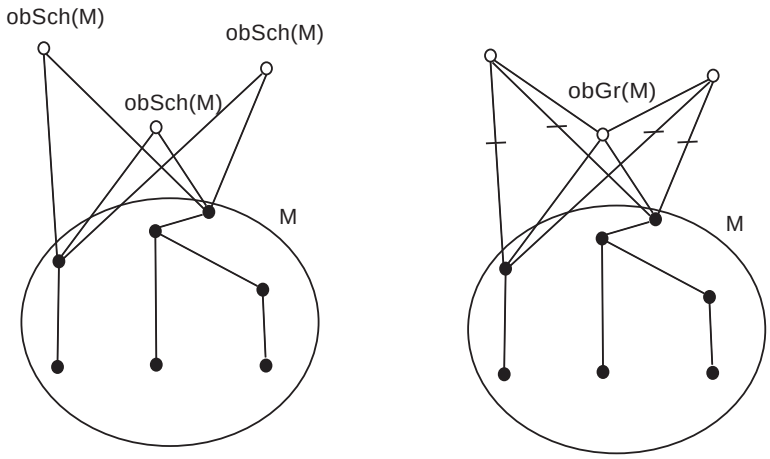


Abbildung 2.13. Obere Schranke und obere Grenze

Anmerkung: M kann eine obere (untere) Schranke haben, muss aber keine obere (untere) Grenze haben.
Besitzt M ein kleinstes (größtes) Element, so ist dieses untere (obere) Grenze. Die untere (obere) Grenze ist eindeutig bestimmt.

Definition 2.32. $(A, \leq)$ *heißt* wohlgeordnet*, wenn jede nichtleere Teilmenge von A ein kleinstes Element besitzt.*

Jede wohlgeordnete Menge ist totalgeordnet.

Wohlordnungssatz: In jeder Menge A kann eine Ordnungsrelation so gefunden werden, dass $(A, \leq)$ wohlgeordnet ist.

Anmerkung: Diese, die Grundlagen der Mathematik berührende Aussage ist gleichwertig mit dem Auswahlaxiom und dem Zorn'schen Lemma.

Auswahlaxiom: Es sei $M_i, i \in I$ eine Mengenfamilie mit $M_i \neq \emptyset, I \neq \emptyset$. Dann gibt es eine Abbildung $f : I \to M_i, i \in I,\ i \mapsto p_i$ mit $p_i \in M_i$. Die Funktion f heißt *Auswahlfunktion.*

Zorn'sche Lemma: Eine geordnete Menge $(A, \leq)$ in der jede totalgeordnete Teilmenge M von A eine obere Schranke besitzt, besitzt ein maximales Element.

2.4.1. Vollständige Induktion. Mit geordneten Mengen hängt das häufig verwendete Beweisschema der *vollständige Induktion* zusammen. Man ordnet jeder natürlichen Zahl n eine Aussage A_n zu.

Die Aussage A_{n_0} sei wahr und für alle $n \geq n_0$ gelte $A_n \to A_{n+1}$. Dann ist A_n wahr für alle $n \geq n_0$.

Ein Induktionsbeweis besteht daher aus drei Teilen

(i) Nachweis der Richtigkeit der Aussage A_{n_0} für ein n_0 (Induktionsvoraussetzung),

(ii) Annahme der Richtigkeit der Aussage A_n für ein n (Induktionsannahme),

(iii) Beweis der Richtigkeit von A_{n+1} unter der Annahme (ii) (Induktionsbeweis).

Beispiel

Man zeige, dass für jede natürliche Zahl n gilt:

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$$

Beweis: Die Aussage A_n ist hier $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$.[3]

(1) Induktionsvoraussetzung: Für $n_0 = 1$ gilt: $\sum_{j=1}^{1} j = 1 = \frac{1(1+1)}{2}$.

(2) Induktionsannahme: Die Aussage A_n ist für n richtig: $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$.

(3) Induktionsbeweis für $n+1$: Zu zeigen ist

$$\sum_{j=1}^{n+1} j = \frac{(n+1)(n+2)}{2}$$

wobei man (2) voraussetzt.

$$\sum_{j=1}^{n+1} j = \sum_{j=1}^{n} j + (n+1) \overset{(2)}{=} \frac{n(n+1)}{2} + (n+1) = \frac{(n+1)(n+2)}{2}.$$

□

[3]Wir verwenden hier die Abkürzung $\sum_{j=1}^{n} a_j$ für die Summe der n Folgenglieder $a_1 + a_2 + \ldots + a_n$.

Dieses Schema lässt sich auf beliebige wohlgeordnete Mengen übertragen (Prinzip der *transfiniten Induktion*).

2.5. Relationales Datenmodell

Relationen bilden auch die Grundlage des relationalen Datenmodells, das in modernen Datenbanken verwendet wird.

Bisher haben wir nur Relationen zwischen *zwei* Mengen (die verschieden oder gleich sein können) betrachtet. Man nennt diese Relationen auch **binäre Relationen** oder **2-stellige Relationen**. Allgemeiner kann man auch Relationen zwischen mehr als zwei Mengen betrachten. Sind das zum Beispiel n Mengen $A_1, ..., A_n$, so wird durch die Teilmenge $R \subset A_1 \times \cdots \times A_n$ eine **n-stellige Relation** definiert. Die Elemente von n-stelligen Relationen sind n-Tupel.

In Datenbanken stellt man solche Relationen in Form von **Tabellen** dar. Die einzelnen n-Tupel der Relation sind dabei die Zeilen der Tabelle. Ein Beispiel: Die Produkte eines Computerhändlers können übersichtlich in Tabellenform aufgelistet werden. Die einzelnen Spalten der Tabelle gehören dabei zu gewissen Mengen (**Attribute** genannt) wie „Produkt“, „Preis“, usw.:

Produkte - Relation R_P

P.Nr.	*Produkt*	*Preis*	*H.Nr.*
1	iMac	990	1
2	PC	590	2
3	Server	2150	2
4	Drucker	95	3

Die Zeilen $(1, \text{iMac}, 990, 1), \ldots$ der Tabelle sind Elemente der Produktmenge $\mathbb{N} \times \texttt{CHAR}(20) \times \mathbb{N} \times \mathbb{N}$. (Hier bezeichnet $\texttt{CHAR}(20)$ die Menge aller Zeichenketten (Strings) mit maximal 20 Zeichen.) Damit stellt die Tabelle eine Relation $R_P \subset \mathbb{N} \times \texttt{CHAR}(20) \times \mathbb{N} \times \mathbb{N}$ dar. Die Mengen stehen hier also für den Datentyp.

Analog kann die Relation $R_H = \{(1, \text{Apple}, \text{Cupertino}), \ldots\} \subset \mathbb{N} \times \texttt{CHAR}(20) \times \texttt{CHAR}(20)$, die nähere Informationen zu den Herstellern enthält, folgendermassen dargestellt werden:

Hersteller - Relation R_H

H.Nr.	*Name*	*Ort*
1	Apple	Cupertino
2	IBM	New York
3	HP	Palo Alto

Die beiden Relationen R_P und R_H bilden eine kleine Datenbank. Damit haben wir aber noch nichts gewonnen, denn in der Praxis möchte man die Daten ja nicht nur speichern, sondern man möchte auch Anfragen durchführen, wie zum Beispiel: „Welche Produkte werden von IBM hergestellt?“

Nun ist es natürlich möglich, alle Abfragen, die man benötigt, einzeln zu implementieren. Steigen aber die Anzahl der Daten und die Anzahl der benötigten Abfragen, so wird das irgendwann zu mühsam. Deshalb versucht man alle möglichen Abfragen auf einige wenige zu reduzieren, und alle anderen dann auf diese zurückzuführen. Das führt direkt zur sogenannten **relationalen Algebra**, die in den meisten Datenbanken als „Structured Query Language“ (SQL) implementiert ist. Hier eine Auswahl der wichtigsten **Operationen**[4]:

- $\sigma_{Bedingung}$ (`SELECT`) wählt alle Zeilen aus, für die die *Bedingung* erfüllt ist.
  Beispiel: Wählen wir aus R_H alle Zeilen aus, deren Attribut „*Name*“ den Wert „IBM“ hat:

$$\sigma_{Name=\text{IBM}}(R_H) = \{(2, \text{IBM}, \text{New York})\},$$

  bzw. in Tabellenform dargestellt:

$\sigma_{Name=\text{IBM}}(R_H)$

H.Nr.	*Name*	*Ort*
2	IBM	New York

- $\pi_{j_1,j_2,\ldots}$ (`PROJECT`) wählt die Spalten j_1, j_2, ... aus.
  Beispiel: Projizieren wir R_H auf die Spalten mit den Attributen *Name* und *Ort*:

$$\pi_{Name,Ort}(R_H) = \{(\text{Apple}, \text{Cupertino}), (\text{IBM}, \text{New York}), (\text{HP}, \text{Palo Alto})\},$$

  bzw. in Tabellenform:

$\pi_{Name,Ort}(R_H)$

Name	*Ort*
Apple	Cupertino
IBM	New York
HP	Palo Alto

- $R_1[j_1, j_2]R_2$ (`JOIN`) „verkettet“ die Relationen R_1 und R_2 bezüglich der gemeinsamen Attributwerte von j_1 (von R_1) und j_2 (von R_2). Die Zeilen der neuen Relation entstehen durch Aneinanderfügung von je einer Zeile der ersten und der zweiten Relation, deren Attributwerte von j_1 und j_2 übereinstimmen.
  Beispiel: Die Relationen R_P und R_H können bezüglich des gemeinsamen Attributs *H.Nr.* verkettet werden:

[4] Operationen ist hier im umgangssprachlichen Sinn verwendet

$R_P[H.Nr., H.Nr.]R_H$

P.Nr.	*Produkt*	*Preis*	*H.Nr.*	*Name*	*Ort*
1	iMac	990	1	Apple	Cupertino
2	PC	590	2	IBM	New York
3	Server	2150	2	IBM	New York
4	Drucker	95	3	HP	Palo Alto

Die Anfrage „Preisliste aller von IBM hergestellten Produkte“ könnte damit wie folgt formuliert werden:

$$\pi_{Produkt,Preis}(\sigma_{Name=\text{IBM}}(R_P[H.Nr., H.Nr.]R_H))$$

Das sieht auf den ersten Blick zwar wild aus, ist aber nicht so schlimm! Sehen wir es uns einfach Schritt für Schritt an:

Schritt 1: Verkettung $R_P[H.Nr., H.Nr.]R_H$:

$R_1 = R_P[H.Nr., H.Nr.]R_H$

P.Nr.	*Produkt*	*Preis*	*H.Nr.*	*Name*	*Ort*
1	iMac	990	1	Apple	Cupertino
2	PC	590	2	IBM	New York
3	Server	2150	2	IBM	New York
4	Drucker	95	3	HP	Palo Alto

Schritt 2: Auswahl der Zeilen mit *Name* gleich „IBM“:

$R_2 = \sigma_{Name=\text{IBM}}(R_1)$

P.Nr.	*Produkt*	*Preis*	*H.Nr.*	*Name*	*Ort*
2	PC	590	2	IBM	New York
3	Server	2150	2	IBM	New York

Schritt 3: Projektion auf die Spalten *Produkt* und *Preis*:

$R_3 = \pi_{Produkt,Preis}(R_2)$

Produkt	*Preis*
PC	590
Server	2150

Das Ergebnis unserer Datenbankabfrage ist also in der Tat die gewünschte Preisliste.

Zusammenfassung

(i) Folgende Begriffe sollten ihnen nun bekannt sein

Relation, Darstellungen von Relationen,
Definitionsbereich, Wertebereich,
reflexiv, symmetrisch, antisymmetrisch, transitiv,

zusammengesetzte Relationen, Äquivalenzrelation, Ordnungsrelation,
Abbildung (Funktion), Folge,
surjektiv, injektiv, bijektiv, Mächtigkeit,
größtes, kleinstes Element, maximale, minimale Elemente,
obere, untere Schranken, obere, untere Grenzen,
Induktionsbeweis,
relationale Datenbank.

(ii) Sie sollten die Eigenschaften einer Relation feststellen können, Relationen auf verschiedene Arten darstellen können, Eigenschaften von Funktionen feststellen können, einfache Induktionsbeweise können, die Struktur einer einfachen Datenbank angeben können.

2.6. Übung

(1) Es sei $A = \{1, 2, 3, 4\}$, $B = \{x, y, z\}$ und $R = \{(1, y), (1, z), (3, y), (4, x), (4, z)\}$. Dann ist (A, B, R) eine Relation.
(i) Warum?
(ii) Man bestimme den Definitionsbereich und Wertebereich von R.
(iii) Man stelle diese Relation als Pfeil-, Koordinaten- und Matrixdiagramm dar.
(iv) Bestimme R^{-1} und zeichne das Koordinaten- und Matrixdiagramm[5].

(2) Es sei $A = B = \{1, 2, 3, 4, 6, 8\}$ und $a\,R\,b$, wenn $a|b$ (a teilt b). Man bestimme R^{-1}. Man beschreibe R^{-1} in Worten und zeichne den gerichteten Graphen der Relation R^{-1}.

(3) Es sei $A = P(D)$ die Potenzmenge von $D = \{1, 2, 3\}$ und auf A sei aRb genau dann, wenn $a \subset b$ gilt.
(i) Man stelle diese Relation als Koordinaten-, Matrixdiagramm dar.
(ii) Ist R reflexiv, symmetrisch oder transitiv?

(4) Man untersuche die Eigenschaften (reflexiv, symmetrisch, antisymmetrisch, transitiv) der folgenden Relationen auf der Menge $A = \{1, 2, 3, 4\}$.

(a) $R_1 = \{(1, 1), (1, 2), (2, 3), (1, 3), (4, 4)\}$,
(b) $R_2 = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 3), (4, 4)\}$,
(c) $R_3 = \{(1, 3), (2, 1)\}$,
(d) $R_4 = \{(1, 2), (2, 1)\}$,
(e) $R_5 = \{\}$,
(f) $R_6 = A \times A$.

Man zeichne auch die gerichteten Graphen.

(5) Es sei $A = \mathbb{N} \times \mathbb{N}$. Auf A sei eine Relation erklärt durch $(a, b)R(c, d)$ genau dann, wenn $a + d = c + b$. Man zeige, dass R reflexiv, symmetrisch und transitiv ist, aber nicht antisymmetrisch. Wie schauen die Äquivalenzklassen aus?

2.7. Übung

(1) Es sei $A = \{1, 2, 3\}$, $B = \{a, b, c\}$, $C = \{x, y, z\}$, $R = \{(1, b), (2, a), (2, c)\}$ und $S = \{(a, y), (b, x), (c, y), (c, z)\}$.
Man bestimme $R \circ S$ und die entsprechenden Pfeildiagramme.

[5] Die Matrix der inversen Relation ist die transponierte Matrix.

(2) Es sei $A = \{a, b, c\}$, $R_1 = \{(a, a), (a, b), (b, c), (c, c)\}$ und $R_2 = \{(a, c), (c, a)\}$. Man bestimme die
(i) reflexive
(ii) symmetrische
(iii) transitive Hülle von $R_j,\ j = 1, 2$.

(3) Man betrachte die Menge der geordneten Paare $(a, b) \in \mathbb{Z} \times \mathbb{Z},\ b \neq 0$, und erkläre eine Relation auf $\mathbb{Z} \times \mathbb{Z}$ durch $(a, b)R(c, d)$ genau dann, wenn $a \cdot d = c \cdot b$ gilt. Man zeige, dass dies eine Äquivalenzrelation ist. Wie schauen die Äquivalenzklassen aus?

(4) Gegeben ist die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto \frac{2\,x}{x^2 + 2}.$$

Man berechne den exakten Wertebereich der Funktion f?

(5) Gegeben ist die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \qquad x \mapsto x^2.$$

Was ist das Urbild der Menge $[4, 9]$?

(6) Gegeben sei eine differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$, die daher im Punkt a die Steigung $f'(a)$ hat. Man zeige, dass die Gerade g die durch den Punkt $f(a)$ geht und die Funktion dort tangential berührt die Form $g(x) = f(a) + f'(a)(x - a)$ hat.

(7) (i) Gegeben sind die Pfeildiagramme (Abb. 2.14).

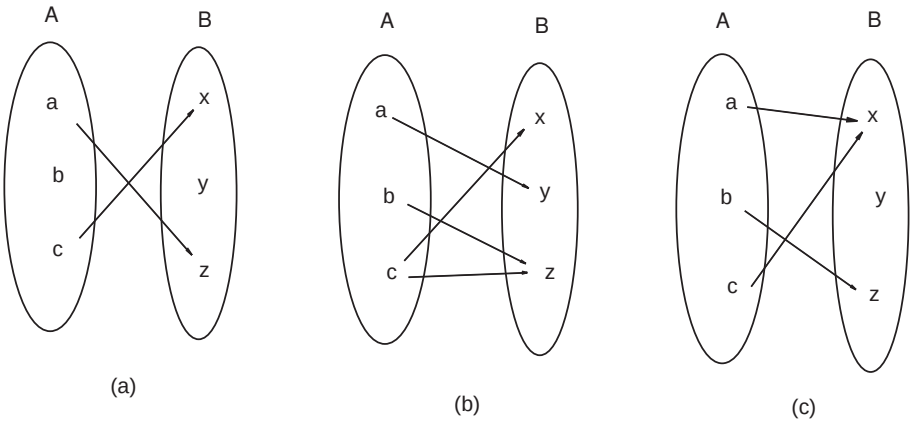


Abbildung 2.14. Pfeildiagramme

Welche stellen Funktionen dar? Warum?
(ii) Gegeben sind die folgenden Diagramme (Abb. 2.15).

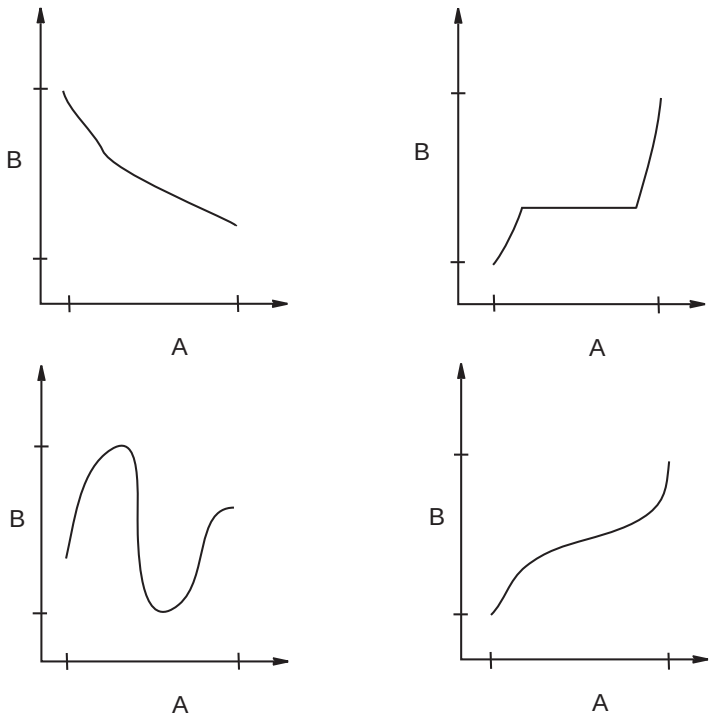


Abbildung 2.15. Diagramme

Welche dieser Funktionen ist surjektiv, injektiv oder bijektiv?

(8) Gegeben sind folgende Mengen. Man finde passende Bijektionen, z.B. mit Hilfe von Polynomen und rationalen Funktionen.
   (a) $A = [0, 1]$, $B = [1, 3]$,
   (b) $A = (0, 1)$, $B = \mathbb{R}^+$,
   (c) $A = \mathbb{R}^+$, $B = \mathbb{R}$,
   (d) $A = (0, 1)$, $B = \mathbb{R}$,
   (e) $A = \mathbb{Z}$, $B = \mathbb{N}$.

2.8. Übung

(1) Gegeben sind die Funktionen

(i) $f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto x + 2, \qquad g : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \dfrac{1}{x^2+1},$

(ii) $f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto x^2, \qquad g : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \sin x.$

Berechne $g \circ f$ und $f \circ g$.

(iii) Gegeben sei $g \circ f = \dfrac{1}{(x+2)^3}$

(iv) und $g \circ f = e^{3x^2}$.

Berechne g und f.

(2) Folgen werden sehr oft anstatt in expliziter Form in rekursiver Form definiert.

(i) Arithmetische Folge: Gegeben ist ein Anfangswert a_0 und für a_{n+1} gilt

$$a_{n+1} = a_n + d, \qquad d \in \mathbb{R}. \tag{2.13}$$

Aufgabe: Berechne a_n explizit und die Summe $a_0 + a_1 + \cdots + a_n$.

(ii) Geometrische Folge: Gegeben ist ein Anfangswert a_0 und für a_{n+1} gilt

$$a_{n+1} = q\, a_n, \qquad q \in \mathbb{R}. \tag{2.14}$$

Aufgabe: Berechne a_n explizit und die Summe $a_0 + a_1 + \cdots + a_n$.

(iii) Fibonacci-Folge: Gegeben sind die Anfangswerte $a_0 = 0$ und $a_1 = 1$ und für a_{n+2} gilt

$$a_{n+2} = a_{n+1} + a_n. \tag{2.15}$$

Siehe Beispiel 2, Übung 2.9.

(3) Man zeichne das Hasse-Diagramm folgender geordneter Mengen
(i) $(P(A), \subset)$ wobei $A = \{2, 4, 7\}$.
(ii) $M = \{2, 3, 5, 6, 7, 11, 12, 35, 385\}$, wobei aRb durch a teilt b gegeben ist.

(4) Eine Ordnung ist durch folgendes Hasse-Diagramm (Abb. 2.16) gegeben. Man gebe die Elemente der Relation an.

(5) Man betrachte die Menge M der echten Teilmengen von $A = \{1, 2, 3\}$, wobei die Ordnung durch Inklusion (ist Teilmenge) gegeben ist. Was sind die maximalen und minimalen Elemente von M?

(6) Es sei $M = \{2, 3, 4, 5, 6, 7, 8\}$, wobei die Ordnung durch $a \leq b$ durch a teilt b gegeben ist. Was sind die maximalen und minimalen Elemente?

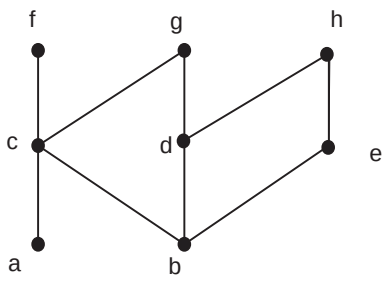


Abbildung 2.16. Hasse-Diagramm

(7) Gegeben sei die Menge $U = \{1, 2, 3, 4, 5, 6, 7, 8\}$ und die Ordnung auf U durch folgendes Hasse-Diagramm (Abb. 2.17) gegeben.

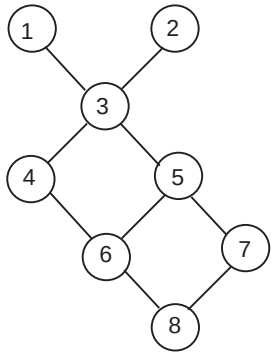


Abbildung 2.17. Hasse-Diagramm

Es seien die Mengen $V = \{4, 5, 6\} \subset U$ und $W = \{3, 5, 6, 7\} \subset U$. Gesucht sind obere und untere Schranken von V und W, sowie Infimum und Supremum.

(8) Eine Ordnungsrelation auf der Menge $A = \{0, 1, 2\} \times \{2, 5, 8\}$ sei gegeben durch $(a, b)R(c, d)$ genau dann, wenn $(a + b)$ teilt $(c + d)$.
(i) Man zeichne das Hasse-Diagramm.
(ii) Was sind die maximalen und minimalen Elemente? Existiert ein größtes und/oder kleinstes Element?

(9) (*) Man betrachte die geordnete Menge $(\mathbb{Q}, \leq)$ und die Teilmenge $B = \{x \in \mathbb{Q} \mid x^2 \leq 2\}$. Hat die Menge B ein maximales Element, ein größtes Element, eine obere Schranke, ein Supremum?

2.9. Übung

(1) Man beweise durch vollständige Induktion.

1) $1 + 3 + 5 + \ldots + (2n-1) = n^2, \qquad n \geq 1,$

2) $2^1 + 2^2 + \ldots + 2^n = 2(2^n - 1), \qquad n \geq 1,$

3) $1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \ldots + n(n+1) = \frac{1}{3} n(n+1)(n+2), \qquad n \geq 1,$

4) $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \ldots + \frac{1}{n(n+1)} = \frac{n}{n+1}, \qquad n \geq 1,$

5) $1^2 + 2^2 + 3^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}, \qquad n \geq 1,$ (2.16)

6) $1^3 + 2^3 + 3^3 + \ldots + n^3 = \frac{n^2(n+1)^2}{4}, \qquad n \geq 1,$ (2.17)

7) $1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \ldots + \frac{1}{\sqrt{n}} > 2(\sqrt{n+1} - 1), \qquad n \geq 1,$

8) $2^n > n^2, \qquad n > 4,$

9) $11^{n+1} + 12^{2n-1}$ ist durch 133 teilbar, $\qquad n \geq 1,$

10) $n^3 + 5\,n$ ist durch 6 teilbar, $\qquad n \geq 1,$

11) $(\cos\varphi + \mathrm{i}\sin\varphi)^n = \cos(n\,\varphi) + \mathrm{i}\sin(n\,\varphi), \qquad \varphi \in \mathbb{R},\ n \geq 1.$

Anmerkung: Summensätze für trigonometrische Funktionen dürfen verwendet werden.

(2) (*) Die rekursive Folge f_n sei definiert durch

$$f_0 = 0, f_1 = 1, \quad f_n = f_{n-1} + f_{n-2}, \quad n \geq 2$$

Die ersten Glieder der Folge lauten daher $0, 1, 1, 2, 3, 5, 8, 13, \ldots$
Man beweise mit vollständiger Induktion, dass gilt (Binetsche Formel)

$$f_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right) \tag{2.18}$$

und

$$\sum_{j=0}^{n} f_j = f_{n+2} - 1. \tag{2.19}$$

2.10. Übung

(1) Gegeben sind folgende Relationen R_A und R_K:

Artikel-Relation R_A

Artikelnr.	*Bezeichnung*	*K.Nr.*	*Menge*
1	Heizdecke	2	25
2	Luftmatratze	2	12
3	Bügelbrett	1	47

Kunden-Relation R_K

K.Nr.	*Name*	*Ort*
1	A&Co	Wien
2	B Inc.	Dornbirn
3	C GmbH	Wien

Finden Sie:

a) $\pi_{Bezeichnung,Menge}(R_A)$ b) $\sigma_{Ort=\text{Wien}}(R_K)$ c) $R_A[K.Nr., K.Nr.]R_K$

(2) Formulieren Sie folgende Datenbankabfragen für die Produkte-Relation R_P und die Hersteller-Relation R_H:

a) „Preisliste aller Produkte, die weniger als $1000,-$ Euro kosten.“

b) „Name und Ort des Herstellers von iMacs.“

Kapitel 3

Algebraische Strukturen

3.1. Algebraische Verknüpfungen

Definition 3.1. *Gegeben seien die Mengen* A, Ω. *Eine Abbildung* $\circ : A \times A \to A$ *heißt* innere algebraische Verknüpfung *(binäre).*

Eine Abbildung $* : \Omega \times A \to A$ *heißt* äußere algebraische Verknüpfung *und* Ω *der* Operatorenbereich.

Eine Menge A versehen mit inneren und/oder äußeren Verknüpfungen *trägt eine algebraische Struktur.*

Beispiele:

(i) $A = \mathbb{N}$ und $\circ := +$, d. h. Die Addition auf $\mathbb{N}$ ist eine innere algebraische Verknüpfung auf $\mathbb{N}$, die je zwei natürlichen Zahlen wieder eine natürliche Zahl, nämlich ihre Summe zuordnet.

(ii) $A = \{$ e,a,b,c $\}$ und die Verknüpfung ist durch folgende Tabelle (Verknüpfungstafel) gegeben

$\circ$	e	a	b	c
e	a	a	b	c
a	a	e	e	b ,
b	b	c	e	a
c	c	b	a	e

z. B. a $\circ$ b = e und b $\circ$ a = c.

(iii) Menge der reellwertigen 2×2-Matrizen mit der äußeren Verknüpfung $*$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{mit} \quad \lambda * \begin{pmatrix} a & b \\ c & d \end{pmatrix} := \begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix}, \qquad a, b, c, d, \lambda \in \mathbb{R}. \tag{3.1}$$

(iv) Multiplikation eines Skalars mit einem Vektor ist eine äußere algebraische Verknüpfung, wobei $\Omega = \mathbb{R}$ und A die Menge der Vektoren ist.

Abbildungen zwischen zwei Mengen mit algebraischen Strukturen, die diese ungeändert lassen, heißen *Homomorphismen*, d. h. es sei $f : (A, \circ_1) \to (B, \circ_2)$, dann gilt $f(a \circ_1 b) = f(a) \circ_2 f(b)$.

Beispiel: $f : (\mathbb{R}^+, \cdot) \to (\mathbb{R}, +)$, wobei $x \mapsto \log x$ ist ein Homomorphismus.

Sind die Abbildungen bijektiv, dann spricht man von *Isomorphismen*, beziehungsweise von *Automorphismen*, wenn die zugrundeliegende Menge beides mal dieselbe ist.

Definition 3.2. *Es sei $\circ$ eine innere Verknüpfung auf der Menge A. Dann heißt*

(i) $\circ$ assoziativ, *wenn für alle $a, b, c \in A$ gilt: $a \circ (b \circ c) = (a \circ b) \circ c$.*

(ii) $\circ$ kommutativ, *wenn für alle $a, b \in A$ gilt: $a \circ b = b \circ a$.*

(iii) $e \in A$ neutrales Element, *wenn für alle $a \in A$ gilt: $e \circ a = a \circ e = a$.*

(iv) $a^{-1} \in A$ *zu a* inverses Element, *wenn gilt: $a^{-1} \circ a = a \circ a^{-1} = e$.*

Anmerkungen: (i) Das neutrale Element e ist eindeutig bestimmt.
(ii) Ist $(A, \circ)$ assoziativ, so ist auch das inverse Element a^{-1} zu a eindeutig bestimmt.

Beispiele: Man betrachte $\mathbb{Z}$ mit der inneren Verknüpfung der Addition und der Subtraktion, bzw. $\mathbb{Q}$ mit der inneren Verknüpfung der Multiplikation und der Division und untersuche welche dieser Eigenschaften erfüllt ist.

Definition 3.3. *Es sei $\circ$ eine innere Verknüpfung auf einer Menge G und $H \subset G$. Dann heißt* $(H, \circ)$ abgeschlossen, *wenn für alle $a, b \in H$ gilt $a \circ b \in H$.*

Beispiel: Man betrachte $G = (\mathbb{Z}, +)$. Dann ist die Teilmenge der geraden Zahlen $H_1 = (\mathbb{G}, +)$ abgeschlossen aber die Teilmenge der ungeraden Zahlen $H_2 = (\mathbb{U}, +)$ nicht abgeschlossen.

Definition 3.4. *Es sei $\circ$ eine innere Verknüpfung auf einer Menge G. Dann heißt* $(G, \circ)$ Gruppe, *wenn*

(i) $\circ$ *ist assoziativ,*

(ii) $(G, \circ)$ *besitzt ein neutrales Element,*

(iii) *für alle $a \in G$ existiert ein inverses Element, das mit a^{-1} bezeichnet wird.*

Eine Gruppe $(G, \circ)$ heißt *abelsch*[1] oder *kommutativ*, wenn zusätzlich gilt: $\circ$ ist kommutativ, d. h. für alle $a, b \in G$ gilt $a \circ b = b \circ a$.

Anmerkungen:
(1) Gilt nur (i) so spricht man von einer *Halbgruppe*. Eine Halbgruppe mit neutralem Element nennt man *Monoid*.
(2) Die Bedingungen (ii) und (iii) sind äquivalent mit der Lösbarkeit der beiden Gleichungen

$$a \circ x = b \text{ und } y \circ a = b.$$

Ist $(G, \circ)$ kommutativ, dann ist sogar $x = y$.
(3) Es gilt $(a^{-1})^{-1} = a$ und $(a_1 a_2 \cdots a_n)^{-1} = a_n^{-1} \cdots a_2^{-1} a_1^{-1}$.

Beispiele:

(i) $(\mathbb{N}, +)$, $(\mathbb{N}, \cdot)$ sind Halbgruppen aber keine Gruppen.

(ii) Ein *Symbol* ist eine abstrakte Einheit, wie zum Beispiel ein lateinischer oder griechischer (etc.) Buchstabe. Eine endliche nichtleere Menge Σ von Symbolen nennt man ein *Alphabet*. Beispiele für Alphabete sind
$\Sigma = \{\alpha, \beta, \gamma, \pi\}$,
$\Sigma = \{a, b, c, \ldots, x, y, z\}$,
$\Sigma = \{\times, +, -, \&, \$, q, @, \%\}$.

Sei Σ nun ein Alphabet, so nennt man jede endliche Folge (n-tupel) von Symbolen ein *Wort (string)* über Σ. Das leere Wort ε ist das Wort, das kein Symbol enthält. Die Länge eines Wortes ist durch die Anzahl der Elemente gegeben. Beispiele für Wörter, wenn $\Sigma = \{a, b, c, d\}$ ist, sind *abbc*, *daccb* und *abdcad*. Man schreibt hier die n-tupel ohne Klammern und Beistriche.

Ist Σ ein Alphabet, so bezeichnet man mit Σ^* die Menge aller Wörter. Auf der Menge Σ^* ist das Produkt (concatenation) definiert durch $x \circ y = xy$, d. h., ist x ein n-tupel und y ein m-tupel, so ist $x \circ y$ das $(n + m)$-tupel, das durch Aneinanderreihung gebildet wird. x heisst Präfix eines Wortes w, wenn ein y existiert, sodass $w = xy$ und y heisst Suffix eines Wortes u, wenn ein x existiert, sodass $u = xy$. $(\Sigma^*, \circ)$ heisst die *freie Halbgruppe* die von Σ erzeugt wird.

(iii) $(\mathbb{Z}, +)$,

(iv) $(\mathbb{Q} \setminus \{0\}, \cdot)$,

[1] nach Niels Hendrik Abel, norwegischer Mathematiker.

(v) Endliche Gruppen mit folgenden Verknüpfungstafeln (Strukturtafeln)

∘	e	a
e	e	a
a	a	e

,

∘	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

,

∘	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

,

∘	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	a	e
c	c	b	e	a

,

∘	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

,

∘	e	a	b	c
e	e	a	b	c
a	a	c	e	b
b	b	e	c	a
c	c	b	a	e

∘	p_1	p_2	p_3	p_4	p_5	p_6
p_1	p_1	p_2	p_3	p_4	p_5	p_6
p_2	p_2	p_3	p_1	p_5	p_6	p_4
p_3	p_3	p_1	p_2	p_6	p_4	p_5
p_4	p_4	p_6	p_5	p_1	p_3	p_2
p_5	p_5	p_4	p_6	p_2	p_1	p_3
p_6	p_6	p_5	p_4	p_3	p_2	p_1

.

Anmerkung: Das Assoziativgesetz ist erfüllt, wenn die Strukturtafel die Rechteckseigenschaft (siehe Körner [**67**] Seite 54) besitzt.

(vi) **Restklassen:** Man betrachte die Menge $\mathbb{Z}$ mit der Relation

$$a \, R \, b \text{ genau dann, wenn } m \text{ teilt } a - b,$$

wobei m eine feste Zahl aus $\mathbb{N} \setminus \{0, 1\}$ sei; d. h. $(a - b) = k\, m, \quad k \in \mathbb{Z}$.

Da dies eine Äquivalenzrelation ist, erhält man eine Klasseneinteilung von $\mathbb{Z}$. Man nennt in diesem Fall die Äquivalenzklassen auch Restklassen und schreibt

$$a \equiv b \pmod{m} \tag{3.2}$$

und spricht: a ist kongruent b modulo m. Als Vertreter für die Restklassen nimmt man üblicherweise die Reste r mit $0 \leq r < m$ also $\{0, 1, 2, \ldots, m-1\}$.

Beispiel: $m = 3$. Die möglichen Restklassen (Reste) bei der Division durch 3 sind $[0], [1], [2]$ und es gilt

$$\begin{aligned} 4 &\equiv 1 \pmod 3, \qquad & 6 &\equiv 0 \pmod 3, \\ 8 &\equiv 2 \pmod 3, \qquad & 8 &\equiv -1 \pmod 3. \end{aligned}$$

Man kann nun für die Restklassen, d. h. auf der Menge $\mathbb{Z}_m = \{[0], \ldots, [m-1]\}$ eine Addition und eine Multiplikation erklären

$$\begin{aligned} [a] \oplus_m [b] &:= [a + b], \\ [a] \otimes_m [b] &:= [a \cdot b]. \end{aligned}$$

Es zeigt sich, dass die Restklassen bezüglich der Addition eine kommutative Gruppe bilden. Bezüglich der Multiplikation erhält man aber nur dann eine (kommutative) Gruppe, wenn m eine Primzahl ist und wenn man zusätzlich das neutrale Element der Addition herausnimmt. Es gilt zum Beispiel $2 \cdot 3 \equiv 0 \pmod 6$, woran man erkennt, dass man keine Gruppe bezüglich der Restklassenmultiplikation hat. Man kann daher nicht „kürzen"!

Übung: Man untersuche $(\mathbb{Q} \setminus \{0\}, :)$ auf Erfüllung der Gruppenaxiome.

Definition 3.5. *Es seien $+, \cdot$ zwei innere Verknüpfungen auf einer Menge R. Dann heißt $(R, +, \cdot)$* Ring*, wenn gilt*

(i) *$(R, +)$ ist eine abelsche Gruppe,*

(ii) *$(R, \cdot)$ ist assoziativ*

(iii) *und es gelten die Distributivgesetze*

$$(a + b) \cdot c = a \cdot c + b \cdot c,$$
$$a \cdot (b + c) = a \cdot b + a \cdot c. \tag{3.3}$$

Ein Ring heißt kommutativ, wenn "$\cdot$" kommutativ ist.

Beispiele: $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Z}_m, \oplus_m, \otimes_m)$.

Definition 3.6. *Es seien $+, \cdot$ zwei innere Verknüpfungen auf einer Menge K. Dann heißt $(K, +, \cdot)$* Körper*, wenn gilt*

(i) *$(K, +, \cdot)$ ist ein kommutativer Ring.*

(ii) *Ist 0 das neutrale Element bezüglich $+$, so wird gefordert, dass $(K \setminus \{0\}, \cdot)$ eine (abelsche) Gruppe ist.*

Anmerkungen:

(i) Das neutrale Element bezüglich $+$ wird üblicherweise mit 0 und das neutrale Element bezüglich $\cdot$ wird mit 1 bezeichnet.

(ii) Das inverse Element a^{-1} von a bezüglich $+$ wird $(-a)$ geschrieben und für $a + (-b)$ schreibt man kurz $a - b$. Das inverse Element a^{-1} von a bezüglich $\cdot$ wird $\frac{1}{a}$ geschrieben und für $a \cdot b^{-1}$ schreibt man kurz $\frac{a}{b}$.

(iii) Körper sind im Gegensatz zu Ringen *nullteilerfrei*, d. h. ist $a \cdot b = 0$, so ist $a = 0$ oder $b = 0$.

Beispiele: $(\mathbb{R}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$ und $(\mathbb{Z}_p, \oplus_p, \otimes_p)$ wobei p prim ist.

Definition 3.7. *Es sei $(K, +, \cdot)$ ein Körper und $(K, \leq)$ eine totalgeordnete Menge. Dann heißt $(K, +, \cdot)$* geordneter Körper*, wenn die Ordnung mit den Körperaxiomen verträglich ist, d. h. wenn für alle $a, b, c \in K$ gilt*

(i) $a \leq b \to a + c \leq b + c$,

(ii) $0 \leq a$ *und* $0 \leq b \to 0 \leq a \cdot b$.

a heißt *positiv*, wenn a größer 0 ist (wobei 0 das neutrale Element bezüglich der Addition + ist). a heißt *negativ*, wenn $(-a)$ positiv ist.

Definition 3.8. *Es sei K ein geordneter Körper. Dann ist die Funktion* Absolutbetrag $|\ | : K \to K$ *definiert durch*

$$|a| = \begin{cases} a, & \text{wenn } a > 0, \\ 0, & \text{wenn } a = 0, \\ -a, & \text{wenn } a < 0. \end{cases} \tag{3.4}$$

Definition 3.9. *Es sei K ein geordneter Körper. Dann ist die Funktion* Signum $\operatorname{sgn} : K \to K$ *definiert durch*

$$\operatorname{sgn}(a) = \begin{cases} 1, & \text{wenn } a > 0, \\ 0, & \text{wenn } a = 0, \\ -1, & \text{wenn } a < 0. \end{cases} \tag{3.5}$$

Definition 3.10. *Der Körper heißt* archimedisch geordnet*, wenn zusätzlich gilt: für alle $a > 0$ existiert ein $n \in \mathbb{N}$, sodass $na = \underbrace{(a + a + \ldots + a)}_{n-mal} > 1$.*

Von zentraler Bedeutung für die Mathematik ist der Begriff eines Vektorraumes.

Definition 3.11. *Es sei V eine Menge und K ein Körper. (V, K) heißt* Vektorraum *über den Körper K, wenn gilt:*

(i) *$(K, +, \cdot)$ ist ein Körper. Die Elemente $\alpha \in K$ heißen Skalare.*

(ii) *$(V, \oplus)$ ist eine abelsche Gruppe. Die Elemente $v \in V$ heißen Vektoren.*

(iii) *Zwischen Skalaren und Vektoren ist eine äußere Verknüpfung $* : K \times V \to V$, $(\alpha, v) \mapsto \alpha * v$ erklärt mit folgenden Eigenschaften:*

(a) *Für alle $\alpha \in K$ und alle $u, v \in V$ gilt*

$$\alpha * (u \oplus v) = (\alpha * u) \oplus (\alpha * v). \tag{3.6}$$

(b) *Für alle* $\alpha, \beta \in K$ *und alle* $v \in V$ *gilt*

$$(\alpha + \beta) * v = (\alpha * v) \oplus (\beta * v). \tag{3.7}$$

(c) *Für alle* $\alpha, \beta \in K$ *und alle* $v \in V$ *gilt*

$$(\alpha \cdot \beta) * v = \alpha * (\beta * v). \tag{3.8}$$

(d) *Ist* 1 *das neutrale Element bezüglich der Multiplikation in* K*, so gelte für alle* $v \in V$*:* $1 * v = v$.

(a) und (b) heißen Distributivgesetze.

Notation: Üblicherweise macht man keine Unterscheidung zwischen $+$ und $\oplus$, $\cdot$ und $*$, da aus dem Zusammenhang klar ist, was gemeint ist. Man unterscheide jedoch zwischen dem neutralen Element 0 (Null) der Addition in K und dem Nullvektor O (z. B. „großes O“). Der *Nullvektor* ist das neutrale Element in $(V, \oplus)$.

Anmerkungen:

(i) Es gilt für alle $v \in V$: $\quad 0 * v = O$.

(ii) Es gilt für alle $\alpha \in K$: $\quad \alpha * O = O$.

(iii) Es sei (-1) das bezüglich $+$ inverse Element in K zu 1 (wobei 1 das neutrale Element bezüglich $\cdot$ in K ist) und $(-v)$ das zu v inverse Element in $(V, \oplus)$. Dann gilt für alle $v \in V$: $\quad (-1) * v = -v$.

Es gelten weiters folgende Beziehungen:

(iv) $-(\alpha * v) = \alpha * (-v)$,

(v) $u = v \;\leftrightarrow\; u - v = O$, ($u - v$ ist die Abkürzung von $u \oplus (-v)$),

(vi) $\alpha * (u - v) = \alpha * u - \alpha * v, \quad (\alpha - \beta) * v = \alpha * v - \beta * v$,

(vii) $\alpha * v = O$ und $\alpha \neq 0 \to v = O, \quad \alpha * v = O$ und $v \neq O \to \alpha = 0$,

(viii) $\alpha * u = \alpha * v$ und $\alpha \neq 0 \to u = v, \quad \alpha * v = \beta * v$ und $v \neq O \to \alpha = \beta$.

Beispiele: (1) Die Menge $\mathbb{R}^n$ (oder $\mathbb{C}^n$) der n-tupel $(x_1, x_2, \ldots x_n)$ von reellen (oder komplexen) Zahlen über den Körper der reellen oder komplexen Zahlen.

Die Addition zweier n-tupel werde durch

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

erklärt. Die Multiplikation mit $\lambda \in \mathbb{R}$ ($\lambda \in \mathbb{C}$) ist definiert durch

$$\lambda \cdot (x_1, x_2, \ldots x_n) = (\lambda x_1, \lambda x_2, \ldots \lambda x_n).$$

Mit diesen Verknüpfungen bildet $(\mathbb{R}^n, \mathbb{R})$ bzw. $(\mathbb{C}^n, \mathbb{C})$ einen Vektorraum.

(2) Eine Abbildung von einer Menge $X \to \mathbb{R}$ nennt man reellwertige Funktion auf X. Es sei nun M die Menge der reellwertigen Funktionen auf dem Intervall $[-1,1]$, d. h. $M = \{f \mid f : [-1,1] \to \mathbb{R}\}$. Sind $f, g \in M$ so definiert man die Funktionen $f+g$ und λf durch $(f+g)(x) = f(x)+g(x)$ und $(\lambda f)(x) = \lambda \cdot f(x)$ für alle $x \in [-1,1]$, und dann sind $f+g$ und λf wieder Elemente von M.

Mit diesen Verknüpfungen bildet $(M, \mathbb{R})$ einen Vektorraum.

(3) Es sei $\mathbb{K}$ ein Körper. Einen Ausdruck der Form

$$P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0, \qquad a_j \in \mathbb{K} \tag{3.9}$$

nennt man Polynom vom Grade n in einer Unbestimmten x.

Man kann ein Polynom nicht nur algebraisch auffassen, sondern auch als eine Funktion wobei $p : \mathbb{R} \to \mathbb{R}, \quad x \mapsto p(x)$ und man z. B. den Körper $\mathbb{R}$ gewählt hat.

Sind p_1, p_2 zwei Polynome vom Grad $\leq n$ so definiert man das Polynom $p_1 + p_2$ und $\lambda\, p$ durch $(p_1+p_2)(x) = p_1(x)+p_2(x)$ und $(\lambda\, p)(x) = \lambda \cdot p(x)$ für alle $x \in \mathbb{R}$. Dann sind $p_1 + p_2$ und λp wieder Polynome vom Grad $\leq n$.

Man zeigt leicht, dass die Polynome vom Grade $\leq n$ mit diesen Verknüpfungen einen Vektorraum bilden.

(4) In der Codierungstheorie und in der Kryptographie sind die Vektorräume $\mathbb{Z}_p^n$ und insbesondere $\mathbb{Z}_2^n$ wichtig. In $\mathbb{Z}_2^2$ gilt zum Beispiel für $a = (1,0)$ und $b = (1,1)$: $a + b = (1+1, 0+1) = (0,1)$ oder $a + a = (1+1, 0+0) = (0,0) = 0$ also $a = -a$. Die Vektoraddition entspricht der bitweisen Xor-Verknüpfung.

3.2. Verbände und Boolesche Algebren

Definition 3.12. *Gegeben sei eine Menge M mit zwei inneren Verknüpfungen $\sqcap$ und $\sqcup$ (eng.: join, meet).* $(M, \sqcap, \sqcup)$ *heißt* Verband, *wenn folgende Gesetze erfüllt sind.*

Für alle $a, b, c \in M$ gilt

$$\begin{aligned}
a \sqcap b &= b \sqcap a, \\
a \sqcup b &= b \sqcup a, && \textit{Kommutativität,} \\
(a \sqcap b) \sqcap c &= a \sqcap (b \sqcap c), \\
(a \sqcup b) \sqcup c &= a \sqcup (b \sqcup c), && \textit{Assoziativität,} \\
a \sqcup (a \sqcap b) &= a, \\
a \sqcap (a \sqcup b) &= a, && \textit{Absorption.}
\end{aligned}$$

Anmerkung: Für jedes $a \in M$ folgt aus den Absorptionsgesetzen $a \sqcup a = a$ und $a \sqcap a = a$.

Definition 3.13. *Gelten in einem Verband die Distributivgesetze*

$$a \sqcap (b \sqcup c) = (a \sqcap b) \sqcup (a \sqcap c),$$
$$a \sqcup (b \sqcap c) = (a \sqcup b) \sqcap (a \sqcup c) \tag{3.10}$$

dann liegt ein distributiver Verband *vor.*

Satz 3.14. Dualitätsprinzip*: Zu jedem verbandstheoretischen Satz gibt es auch einen dualen Satz. D. h. wenn man $\sqcup$ und $\sqcap$ vertauscht, muss der Satz weiterhin gültig bleiben.*

Jeder Verband bewirkt eine Ordnungsstruktur.

Satz 3.15. *Es sei $(M, \sqcap, \sqcup)$ ein Verband. Für $a, b \in M$ setzt man $a \leq b$, genau dann wenn $a \sqcup b = b$ ist. Dann gilt*

(i) *$\leq$ ist eine Ordnung auf M.*

(ii) *Für $a, b \in M$ gilt $a \leq b$, genau dann wenn $a \sqcap b = a$ ist.*

(iii) *Für alle $a, b \in M$ gilt $a \leq a \sqcup b$ und $b \leq a \sqcup b$, und für jedes $c \in M$ mit $a \leq c$ und $b \leq c$ gilt $a \sqcup b \leq c$.*

(iv) *Für alle $a, b \in M$ gilt $a \sqcap b \leq a$ und $a \sqcap b \leq b$, und für jedes $d \in M$ mit $d \leq a$ und $d \leq b$ gilt $d \leq a \sqcap b$.*

Umgekehrt gilt

Satz 3.16. *Es sei $(M, \leq)$ eine geordnete Menge mit folgenden zusätzlichen Eigenschaften*

(i) *Zu allen $a, b \in M$ gibt es ein Element $\tilde{c}$ in M mit $a \leq \tilde{c}$ und $b \leq \tilde{c}$ und für alle $c \in M$ mit $a \leq c$ und $b \leq c$ gilt $\tilde{c} \leq c$ (Supremum). Man setzt $a \sqcup b := \tilde{c}$*

(ii) *Zu allen $a, b \in M$ gibt es ein Element $\hat{d}$ in M mit $\hat{d} \leq a$ und $\hat{d} \leq b$ und für alle $d \in M$ mit $d \leq a$ und $d \leq b$ gilt $d \leq \hat{d}$ (Infimum). Man setzt $a \sqcap b := \hat{d}$*

Durch die so definierten Verknüpfungen $\sqcap : (a, b) \mapsto a \sqcap b$ und $\sqcup : (a, b) \mapsto a \sqcup b$ wird $(M, \sqcap, \sqcup)$ zu einem Verband.

D. h.: zu je zwei Elementen gibt es ein Supremum und ein Infimum.

Dieser **Zusammenhang mit der Ordnungstheorie** bietet die Möglichkeit endliche Verbände mittels Hasse-Diagrammen zu veranschaulichen. (Obiger Satz wird daher auch zur Definition von Verbänden benutzt.) Dies erklärt übrigens die Bezeichnung Verband, da je zwei Elemente sowohl nach oben als auch nach unten verbunden sind.

Beispiele:

(1) Teilmengenverband (Abb. 3.1): Gegeben sei die Menge $A = \{a, b, c\}$. Dann ist $M = P(A)$ mit der Ordnung die durch die Inklusion gegeben ist ein distributiver Verband. (Zwei beliebige Elemente haben ein Infimum und ein Supremum.)

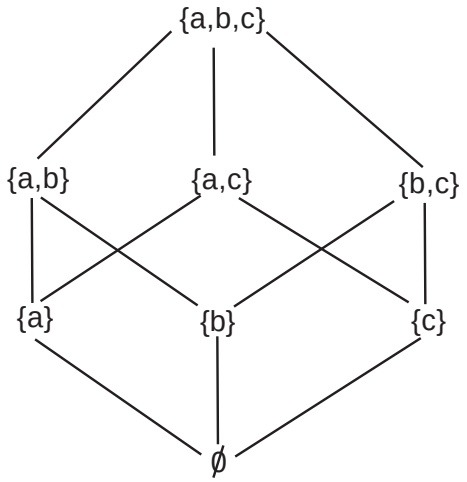


Abbildung 3.1. Teilmengenverband

(2) Die Teiler der Zahl 12 (Abb. 3.2): $\sqcap$ (inf) entspricht dem größten gemeinsamen Teiler (ggT) und $\sqcup$ (sup) dem kleinsten gemeinsamen Vielfachen (kgV).

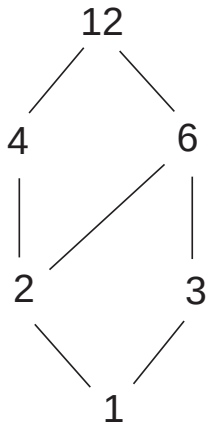


Abbildung 3.2. Teiler der Zahl 12

(3) Es sei M eine Menge. Dann ist $(P(M), \cap, \cup)$ ein distributiver Verband.

(4) Alle Verbände mit 5 Elementen (Abb. 3.3).

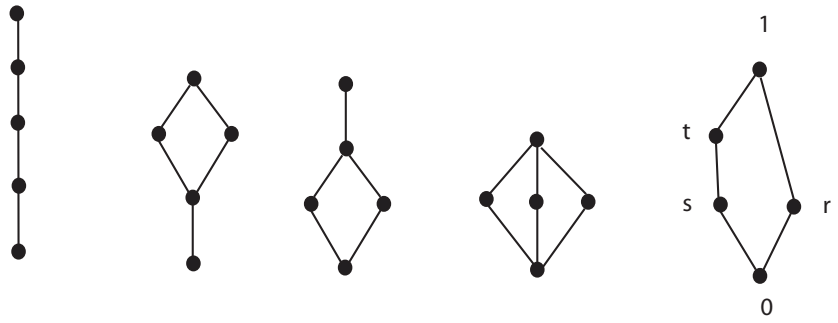


Abbildung 3.3. Alle Verbände mit 5 Elementen

(5) Beispiel für einen nichtdistributiven Verband (Abb. 3.4).

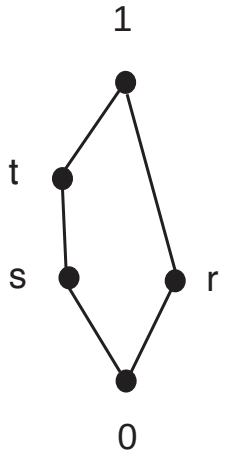


Abbildung 3.4. Nichtdistributiver Verband

$t \sqcap (s \sqcup r) = t \sqcap 1 = t, \quad (t \sqcap s) \sqcup (t \sqcap r) = s \sqcup 0 = s.$

Definition 3.17. *Ein Verband heißt* beschränkt*, wenn er immer ein kleinstes und ein größtes Element besitzt. Das kleinste Element bezeichnet man mit* n *oder* 0*, das größte* e *oder* 1*.*

Die Menge $(\mathbb{Z}, \leq)$ bildet einen unbeschränkten Verband.

Definition 3.18. *Ein Verband M heißt* komplementär, *wenn es zu jedem $a \in M$ ein $\bar{a} \in M$ gibt mit*

$$a \sqcap \bar{a} = 0,$$
$$a \sqcup \bar{a} = 1.$$

In distributiven Verbänden sind komplementäre Elemente eindeutig bestimmt.

Definition 3.19. *Ein komplementärer distributiver Verband heißt* Boolsche Algebra.

Definition 3.20. *Es seien $(A, \cap, \cup)$ und $(B, \sqcap, \sqcup)$ zwei Verbände. A ist isomorph zu B, wenn eine bijektive Abbildung φ existiert mit*

$$\begin{aligned} \varphi(a \cup b) &= \varphi(a) \sqcup \varphi(b), \\ \varphi(a \cap b) &= \varphi(a) \sqcap \varphi(b). \end{aligned} \tag{3.11}$$

Satz 3.21. *(Stone) Jede Boolsche Algebra ist isomorph zu einem Teilmengenverband.*

Zusammengefasst kann man eine Boolsche Algebra wie folgt definieren ($\sqcup$ entspricht $\oplus$ und $\sqcap$ entspricht $\otimes$)

Definition 3.22. *Eine Boolsche Algebra besteht aus einer Menge B mit drei Operationen auf dieser Menge*

(a) Einer inneren Verknüpfung $\oplus : B \times B \to B,\ (a, b) \mapsto a \oplus b$,
(b) einer inneren Verknüpfung $\otimes : B \times B \to B,\ (a, b) \mapsto a \otimes b$
(c) und einer Abbildung $^- : B \to B,\ a \mapsto \bar{a}$

wobei folgende Axiome gelten

(B1) Die Verknüpfungen $\oplus, \otimes$ sind assoziativ

$$(a \oplus b) \oplus c = a \oplus (b \oplus c),$$
$$(a \otimes b) \otimes c = a \otimes (b \otimes c).$$

(B2) Die Verknüpfungen $\oplus, \otimes$ sind kommutativ

$$a \oplus b = b \oplus a,$$
$$a \otimes b = b \otimes a.$$

(B3) Beide Verknüpfungen $\oplus, \otimes$ sind distributiv

$$a \oplus (b \otimes c) = (a \oplus b) \otimes (a \oplus c),$$
$$a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c).$$

(B4) Es existieren zwei Elemente $0, 1$ *wobei* $0 \neq 1$ *für die gilt*

$$b \oplus 0 = b \quad \textit{für alle } b \in B$$
$$b \otimes 1 = b \quad \textit{für alle } b \in B.$$

(B5) Für alle $b \in B$ *gilt*

$$b \oplus \bar{b} = 1 \textit{ und } b \otimes \bar{b} = 0. \tag{3.12}$$

Anmerkung: das komplementäre Element $\bar{b}$ ist nicht das inverse Element von b.

Beispiele

(1) Die einfachste Boolsche Algebra besteht aus der Menge $B = \{0, 1\}$ mit den beiden Operation $\oplus, \otimes$ und der Komplementoperation $^{-}$, die wie folgt auf B definiert sind

$\oplus$	0	1
0	0	1
1	1	1

$\otimes$	0	1
0	0	0
1	0	1

b	$\bar{b}$
0	1
1	0

(2) Es sei S eine nicht leere Menge. Auf der Potenzmenge $P(S)$ betrachte man die Operationen $A \oplus B = A \cup B, A \otimes B = A \cap B$ und $\bar{A} = S \setminus A$, wobei $A, B \in P(S)$. Dann bildet $(P(S), \cup, \cap, ^{-}, \emptyset, S)$ eine Boolsche Algebra.

(3) Man betrachte die Menge aller Aussagen B, die abgeschlossen unter den Operationen $\vee, \wedge, ^{-}$ sind, wobei Gleichheit im Sinne der logischen Äquivalenz verstanden sei. t bezeichne die Tautologien und f die Kontradiktionen. Dann bildet $(B, \vee, \wedge, ^{-}, f, t)$ eine Boolsche Algebra.

Zusammenfassung

(i) Folgende Begriffe sollten ihnen nun bekannt sein

innere, äußere algebraische Verknüpfung,
neutrales, inverses Element, assoziativ, kommutativ, distributiv
Gruppe, Ring, Körper, Vektorraum,
Verband, Boolsche Algebra.

(ii) Sie sollten die Eigenschaften von Verknüpfungen feststellen können, z.B. mithilfe von Verknüpfungstafeln.

3.3. Übung

(1) Man betrachte ein nichtquadratisches Rechteck (Abb. 3.5) mit den Ecken A, B, C, D und den Achsen L_1, L_2 wie in folgendem Diagramm gezeigt:

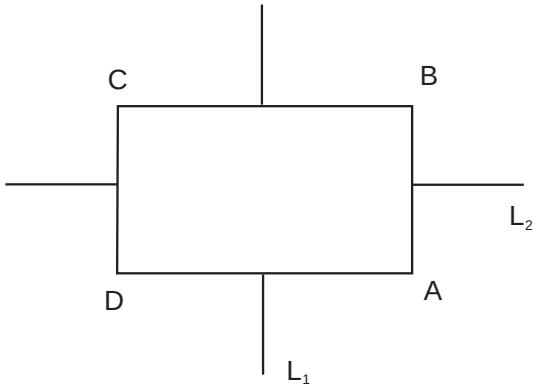


Abbildung 3.5. Viereck

Das Rechteckt besitzt vier Symmetrie-Operationen:
r_0: Drehung[2] um 0° um den Mittelpunkt
r_1: Drehung um 180° um den Mittelpunkt
m_1: Spiegelung um die Achse L_1
m_2: Spiegelung um die Achse L_2
Man stelle die Verknüpfungstafel für die Zusammensetzung dieser Transformationen (z.B. $r_1 \circ m_2$) auf und zeige, dass sie eine Gruppe bilden.
Anmerkungen: (i) Das Assoziativgesetz folgt aus dem Assoziativgesetz für Abbildungen, wenn man diese Transformationen als Abbildungen von $\mathbb{R}^2$ nach $\mathbb{R}^2$ auffasst.
(ii) Die Spiegelachsen L_1, L_2 bleiben immer fix.

(2) Man stelle die Verknüpfungstafel für die Restklassen $(\mathbb{Z}_6, \oplus_6)$, $(\mathbb{Z}_4 \setminus \{[0]\}, \otimes_4)$, $(\mathbb{Z}_5 \setminus \{[0]\}, \otimes_5)$ und weiters der teilfremden[3] Reste von 10 bezüglich der Multiplikation $\otimes_{10}$ auf.

(3) Man zeige, dass alle Elemente der folgenden Gruppe als Potenzen eines Elements geschrieben werden können.
Anmerkung: Gruppen deren Elemente als Potenz eines einzigen Elements gegeben sind heißen zyklische Gruppen.

[2] In der Mathematik ist der positive Drehwinkel gegen den Uhrzeigersinn definiert
[3] siehe Definition 5.9

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	a	e
c	c	b	e	a

(4) Man betrachte ein gleichseitiges Dreieck (Abb. 3.6) und alle Symmetrie-Operationen die dieses auf sich selbst abbilden.

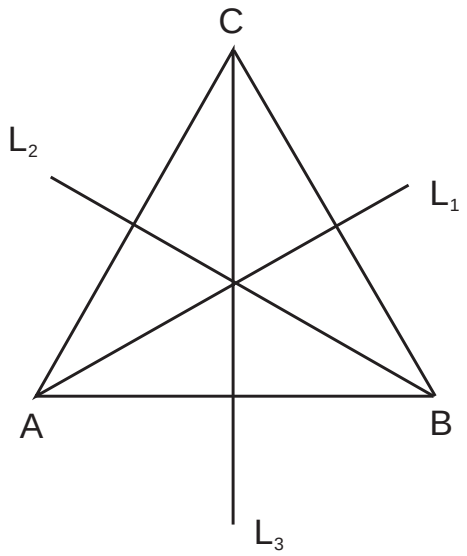


Abbildung 3.6. Dreieck

Diese sind:
r_0 Rotation gegen den Uhrzeigersinn um 0°
r_1 Rotation gegen den Uhrzeigersinn um 120°
r_2 Rotation gegen den Uhrzeigersinn um 240°
m_1 Spiegelung um die Seitensymmetrale L_1
m_2 Spiegelung um die Seitensymmetrale L_2
m_3 Spiegelung um die Seitensymmetrale L_3

Man betrachte nun die Menge $D_3 = \{r_0, r_1, r_2, m_1, m_2, m_3\}$ dieser Operationen mit der Verknüpfung $\circ$ die durch Hintereinanderausführung zweier solcher Operationen definiert ist. Man stelle die Verknüpfungstafel auf und zeige dass $(D_3, \circ)$ eine nichtkommutative Gruppe bildet.

(5) Auf der Menge $C = \{(a,b) \mid a,b \in \mathbb{R} \setminus \{(0,0)\}\}$ sei folgende Verknüpfung $\circ$ definiert

$$(a,b) \circ (c,d) = (ac - bd, ad + bc).$$

Zeige, dass $(C, \circ)$ eine abelsche Gruppe bildet.

(6) Unter Benutzung der Axiome B1-B5 beweise man, dass in einer Boolschen Algebra für alle $b \in B$ gilt (Idempotenzgesetze):

$$b \oplus b = b \quad \text{und} \quad b \otimes b = b.$$

(7) Unter Benutzung der Axiome B1-B5 und Beispiel 6 beweise man, dass in einer Boolschen Algebra für alle $b \in B$ gilt:

$$b \oplus 1 = 1.$$

(8) Unter Benutzung der Axiome B1-B5 beweise man, dass in einer Boolschen Algebra für alle $b_1, b_2 \in B$ die De Morgan Gesetze gelten:

$$\overline{(b_1 \oplus b_2)} = \bar{b}_1 \otimes \bar{b}_2 \quad \text{und} \quad \overline{(b_1 \otimes b_2)} = \bar{b}_1 \oplus \bar{b}_2.$$

Hinweis: Zeige $(b_1 \oplus b_2) \oplus (\bar{b}_1 \otimes \bar{b}_2) = 1$

(9) (*) Für die achtelementige Menge $Q_8 = \{\pm 1, \pm \mathrm{i}, \pm \mathrm{j}, \pm \mathrm{k}\}$ sei eine innere assoziative Verknüpfung $\cdot$ wie folgt definiert. Neben den üblichen Vorzeichenregeln gelte zusätzlich: $\mathrm{i} \cdot \mathrm{i} = \mathrm{j} \cdot \mathrm{j} = \mathrm{k} \cdot \mathrm{k} = \mathrm{i} \cdot \mathrm{j} \cdot \mathrm{k} = -1$. Man stelle die Verknüpfungstafel auf, z.B. folgt $\mathrm{i} \cdot \mathrm{j} = \mathrm{k}$ und $\mathrm{j} \cdot \mathrm{i} = -\mathrm{k}$. Ist $(Q_8, \cdot)$ eine (kommutative) Gruppe?[4]

[4] https://de.wikipedia.org/wiki/Quaternionengruppe

Kapitel 4

Zahlen

4.1. Die natürlichen Zahlen

Der Zahlenbegriff der natürlichen Zahlen $\mathbb{N} = \{0, 1, 2, \ldots\}$ entwickelte sich historisch aus dem Bedürfnis des Abzählens von Mengen (Kardinalzahlen) und dem Sortieren (Ordnen) von Mengen (Ordinalzahlen).[1]

Im Prinzip haben wir bis jetzt bis auf die leere Menge keine Menge richtig definiert und wollen nun die natürlichen Zahlen axiomatisch einführen. Dazu wählt man einen rekursiven Zugang.

Definition 4.1. *(Peano-Axiome)*

(i) *Null (Symbol* 0*) ist eine natürliche Zahl.*

(ii) *Jede natürliche Zahl* n *besitzt genau eine natürliche Zahl* n' *als Nachfolger.*

(iii) *Null ist nicht Nachfolger einer natürlichen Zahl.*

(iv) *Jede natürliche Zahl ist Nachfolger höchstens einer natürlichen Zahl.*

(v) *Enthält eine Teilmenge* M *von* $\mathbb{N}$ *die Zahl* 0 *und mit jeder natürlichen Zahl auch deren Nachfolger so ist* $M = \mathbb{N}$.

Mittels der Nachfolgerregel lassen sich aus dem Grundbegriff 0 alle übrigen natürlichen Zahlen explizit definieren, z.B. $1 := 0', 2 := 1', \ldots$, etc.

Eine Menge, die diese Axiome erfüllt, ist die Folgende. Die Idee ist dabei eine Menge zu nehmen, die genauso viele Elemente enthält wie die Zahl die sie

[1] `http://de.wikipedia.org/wiki/Ishango-Knochen`

darstellt, d.h. man definiert[2]

$$\begin{aligned} 0 &:= \{\ \} = \emptyset, \\ 1 &:= \{0\} = \{\emptyset\}, \qquad \text{beachte } \{\emptyset\} \neq \emptyset, \\ 2 &:= \{0,1\} = \{\emptyset, \{\emptyset\}\}, \\ &\vdots \\ n &:= \{0,1,\ldots,n-1\}, \\ &\vdots \end{aligned} \tag{4.1}$$

Die Addition von natürlichen Zahlen wird durch $n + 0 := n$ und $n + m' := (n + m)'$ definiert.

Die Multiplikation von natürlichen Zahlen wird durch $n \cdot 0 := 0$ und $n \cdot m' := n \cdot m + n$ definiert.

Damit kann man nun z.B. beweisen, dass $2 + 2 = 4$ oder $2 \times 2 = 4$.

Die Ordnungsstruktur für die natürlichen Zahlen wird definiert durch

$$\begin{aligned} &m \leq n \Leftrightarrow \text{ es existiert ein } k \in \mathbb{N}, \text{ sodass } \ m + k = n \text{ gilt}, \\ &m < n \Leftrightarrow m \leq n \text{ und } m \neq n. \end{aligned}$$

4.1.1. Abstrakt Folgen.

Definition 4.2. *Eine Abbildung $a : \mathbb{N} \to A,\ n \mapsto a_n\ (= a(n))$ heißt* Folge *(in A). a_n heißt das n-te Glied der Folge.*

Die Abbildung einer Teilmenge von $\mathbb{N}$ in A definiert eine *Teilfolge.*

Definition 4.3. *Es sei "+" eine assoziative Verknüpfung in der Menge A. Durch*

$$\sum_{\nu=0}^{0} a_\nu = a_0, \qquad \sum_{\nu=0}^{n+1} a_\nu = \sum_{\nu=0}^{n} a_\nu + a_{n+1}, \quad n \in \mathbb{N} \tag{4.2}$$

wird die Summenfolge $\{\sum_{\nu=0}^{n} a_\nu\}$ *definiert. ν heißt* Summationsindex.

Der Summationsindex darf beliebig umbenannt werden

$$\sum_{\nu=k+1}^{k+n} a_\nu = \sum_{\nu=1}^{n} a_{\nu+k}$$

[2] erfunden von `https://de.wikipedia.org/wiki/John_von_Neumann`

und es gilt das allgemeine Assoziativgesetz

$$\sum_{\nu=1}^{m} a_\nu + \sum_{\nu=m+1}^{m+n} a_\nu = \sum_{\nu=1}^{m+n} a_\nu, \tag{4.3}$$

(Beweis durch Induktion).

4.1.2. Permutation.

Definition 4.4. *Eine bijektive Abbildung von* $\{0, 1, 2, \ldots, n\}$ *auf sich nennt man* Permutation.

Beispiel:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 0 & 3 \end{pmatrix}, \quad \text{oder in Zyklenschreibweise } (0, 1, 4, 3)\ (2). \tag{4.4}$$

Die Permutationen bilden bezüglich der Hintereinanderausführung eine Gruppe. Ist die Verknüpfung "+" in der Menge A kommutativ, so gilt das allgemeine Kommutativgesetz. D. h. es sei $\nu \mapsto j_\nu$ eine beliebige Permutation der Zahlen $\{0, 1, \ldots, n\}$. Dann ist $\sum_{\nu=0}^{n} a_\nu = \sum_{\nu=0}^{n} a_{j_\nu}$.

Wird die Verknüpfung multiplikativ geschrieben, so spricht man von Produkten und $\sum$ wird durch $\prod$ ersetzt, d. h. $\{\prod_{\nu=0}^{n} a_\nu\}$ heißt *Produktfolge*.

Ist der Indexbereich einer Summe (eines Produktes) leer, so wird sie (es) gleich dem neutralen Element bezüglich "+" (bzw. bezüglich "$\cdot$") gesetzt.

Definition 4.5. *Die Fakultät* $n!$ *ist rekursiv definiert durch:* $0! = 1,\ 1! = 1,\ (n+1)! = n!(n+1)$.

Notiz:[3] Die Fakultät $n!$ gibt die Anzahl der Permutationen der Menge $\{1, 2, \ldots, n\}$ an.

Definition 4.6. *Der Binomialkoeffizient* $\binom{n}{k}$ *ist definiert durch*

$$\binom{n}{k} = \begin{cases} \frac{n!}{k!(n-k)!} & \text{für } 0 \le k \le n \\ 0 & \text{für } k < 0 \text{ und } k > n. \end{cases} \tag{4.5}$$

Für die Binomialkoeffizienten gelten folgende Beziehungen

$$\binom{n}{k} = \binom{n}{n-k}, \tag{4.6}$$

$$\binom{n}{k} = \frac{1}{k!} \prod_{\nu=0}^{k-1} (n - \nu) = \frac{n(n-1)\cdots(n-k+1)}{k!}, \tag{4.7}$$

[3] Sehr praktisch für Anwendungen ist auch noch die Doppelfakultät: $(n+2)!! = (n+2)\, n!!,\ n \ge 1$.

$$(k+1)\binom{n}{k+1} = (n-k)\binom{n}{k}, \tag{4.8}$$

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}, \tag{4.9}$$

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n, \tag{4.10}$$

$$\sum_{\nu=k}^{n} \binom{\nu}{k} = \binom{n+1}{k+1}. \tag{4.11}$$

Pascal'sches Dreieck (4.9)

$$\begin{array}{ccccccccccc} & & & & & 1 & & & & & \\ & & & & 1 & & 1 & & & & \\ & & & 1 & & 2 & & 1 & & & \\ & & 1 & & 3 & & 3 & & 1 & & \\ & 1 & & 4 & & 6 & & 4 & & 1 & \\ 1 & & 5 & & 10 & & 10 & & 5 & & 1 \end{array} \tag{4.12}$$

Satz 4.7. Binomischer Lehrsatz. *In A seien "$+$" und "$\cdot$" assoziative und kommutative Verknüpfungen. Dann gilt für $a, b \in A$*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \quad \text{für alle } n. \tag{4.13}$$

Übung: (*) Beweis mit vollständiger Induktion.

Satz 4.8. Bernoullische Ungleichung. *Es sei $a > -1$. Dann gilt*

$$(1+a)^n \geq 1 + na \quad \text{für alle } n. \tag{4.14}$$

Satz 4.9. (*) Polynomischer Lehrsatz. *In A seien "$+$" und "$\cdot$" assoziative und kommutative Verknüpfungen. Dann gilt für $a_j \in A, j = 1, \ldots, k$*

$$(a_1 + a_2 + \ldots + a_k)^n = \sum_{\substack{\nu_1, \nu_2, \ldots, \nu_k = 0 \\ \nu_1 + \nu_2 + \ldots + \nu_k = n}}^{n} \frac{n!}{\nu_1! \nu_2! \cdots \nu_k!} a_1^{\nu_1} a_2^{\nu_2} \cdots a_k^{\nu_k} \tag{4.15}$$

(Beweis durch vollständige Induktion nach k).

4.2. Die ganzen und rationalen Zahlen

$(\mathbb{N}, +)$ bildet keine Gruppe, denn die Gleichung $m + x = n$ ist nicht immer lösbar, bzw. zu $m > 0$ existiert kein inverses Element. Um diese Schwierigkeit zu beheben führt man die negativen ganzen Zahlen ein. Es sei $m \in \mathbb{N}$. Dann definiert man

$(-m)$ sei bezüglich "$+$" das inverse Element zu m, d.h. $(-m) + m = 0$.

Die Fortsetzung der Ordnungsstruktur von $\mathbb{N}$ auf $\mathbb{Z}$ wird folgendermaßen definiert

$$m, n \in \mathbb{Z}, \qquad m \leq n := \begin{cases} m, n \in \mathbb{N} \text{ und } m \leq n, \\ m \in \mathbb{Z} \setminus \mathbb{N} \text{ und } n \in \mathbb{N}, \\ m, n \in \mathbb{Z} \setminus \mathbb{N} \text{ und } (-n) \leq (-m). \end{cases} \tag{4.16}$$

Die Fortsetzung der Addition von $\mathbb{N}$ auf $\mathbb{Z}$ wird folgendermaßen festgelegt, $m, n \in \mathbb{N}$

$$\begin{aligned} m + (-n) &:= \begin{cases} m - n & \text{falls } n \leq m, \\ -(n - m) & \text{falls } m < n, \end{cases} \\ (-m) + (-n) &:= -\,(m + n). \end{aligned} \tag{4.17}$$

Damit wird $(\mathbb{Z}, +)$ eine abelsche Gruppe.

Die Fortsetzung der Multiplikation von $\mathbb{N}$ auf $\mathbb{Z}$ wird folgendermaßen definiert: $m \cdot (-n) := -(m \cdot n)$ und $(-m) \cdot n := -(m \cdot n)$.

$\mathbb{Z}$ ist abzählbar unendlich aber nicht wohlgeordnet, da es kein kleinstes Element gibt.

Anmerkung: Man kann $\mathbb{Z}$ formal aus $\mathbb{N}$ konstruieren als Menge der Äquivalenzklassen von Paaren $(a, b) \in \mathbb{N} \times \mathbb{N}$, wobei zwei Paare (a, b), (c, d) genau dann äquivalent sind, wenn $a + d = c + b$ gilt, d.h. das Paar (a, b) steht für "$a - b$". (Details findet man z.B. in [**99**])

Analog geht man bei der Einführung der *rationalen Zahlen* (Bruchzahlen) vor.

$(\mathbb{Z} \setminus \{0\}, \cdot)$ bildet keine Gruppe, denn die Gleichung $ax = b$ ist nicht immer lösbar, bzw. zu $a \neq 1$ existiert kein inverses Element. Um diese Schwierigkeit zu beheben führt man die rationalen Zahlen ein.

$a^{-1} = \frac{1}{a}$ wird definiert als das zu a inverse Element bezüglich der Multiplikation.

Die Gruppe, die von $\mathbb{Z}$ und diesen inversen Elementen erzeugt wird, nennt man die multiplikative Gruppe der rationalen Zahlen $(\mathbb{Q} \setminus \{0\}, \cdot)$.

Ordnung, Addition und Multiplikation werden dabei wie oben auf $\mathbb{Q}$ fortgesetzt. $(\mathbb{Q}, +, \cdot)$ ist ein Körper. $\mathbb{Q}$ ist abzählbar unendlich (Cantorsches Diagonalverfahren) aber nicht wohlgeordnet.

Die rationalen Zahlen liegen auf der Zahlengerade dicht, d.h. zwischen zwei verschiedenen rationalen Zahlen a und b mit $a < b$ gibt es immer eine weitere rationale Zahl, z.B. $\frac{a+b}{2}$.

Anmerkung: Man kann $\mathbb{Q}$ formal aus $\mathbb{Z}$ konstruieren als Menge der Äquivalenzklassen von Paaren $(a, b) \in \mathbb{Z} \times \mathbb{Z},\ b \neq 0$, wobei zwei Paare (a, b), (c, d) genau dann äquivalent sind, wenn $a \cdot d = c \cdot b$ gilt, d.h. (a, b) steht für "a/b". (Weitere Details dazu findet man z.B. in [**99**])

4.2.1. Darstellung von Zahlen im Stellenwertsystem. Historisch hat sich die Darstellung von Zahlen im Dezimalsystem, einem System mit zehn Ziffern, relativ langsam entwickelt, denn dazu war die Erfindung der Null erforderlich. Eine (endliche) *Dezimalzahl* der Form $a_n a_{n-1} \cdots a_1 a_0 . a_{-1} a_{-2} \cdots a_{-m}$ ist eine abgekürzte Schreibweise für die folgende Summe

$$a_n a_{n-1} \cdots a_1 a_0 . a_{-1} a_{-2} \cdots a_{-m} = \sum_{j=-m}^{n} a_j 10^j, \quad a_j \in \{0, 1, 2, \ldots, 9\}. \tag{4.18}$$

Jedes System mit mehr als einer Ziffer ist zur Darstellung von Zahlen geeignet. Bei $b, b \geq 2$ Ziffern spricht man von einer b-adischen Darstellung.

Anmerkung: Im deutschsprachigen Raum ist es üblich ein Komma zur Trennung des nicht ganzzahligen Anteils zu verwenden, während im amerikanischen ein Punkt (decimal point) verwendet wird.[4]

Neben dem Dezimalsystem sind hauptsächlich das Dualsystem (mit zwei Ziffern $\{0, 1\}$, eingeführt von Leibniz), das Hexadezimalsystem (mit sechzehn Ziffern $\{0, 1, \ldots, 9, \text{A}, \text{B}, \text{C}, \text{D}, \text{E}, \text{F}\}$) und das Sexagesimalsystem[5] (auch Hexagesimalsystem oder Sechziger-System genannt) in Verwendung.

Umrechnung einer Zahl aus dem Dualsystem in das Dezimalsystem

$$\begin{aligned} (110101)_2 &= 1 \cdot 2^5 + 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 \\ &= 32 + 16 + 0 \cdot 8 + 4 + 0 \cdot 2 + 1 \\ &= (53)_{10}. \end{aligned} \tag{4.19}$$

(Der Index j in $(\ldots)_j$ weißt auf das verwendete System hin und wird weggelassen, wenn aus dem Zusammenhang klar ist welches verwendet wird.) Ganz analog geht man bei der Umrechnung einer Zahl aus einem beliebigen System in das Zehnersystem vor.

Umrechnung einer Zahl aus dem Dezimalsystem in das Dualsystem.
Methode 1: (Subtraktionsmethode). Man subtrahiert sukzesive die höchstmögliche 2er Potenz.

$$\begin{aligned} 58 &= 32 + 16 + 8 + 2, \\ (58)_{10} &= (2^5 + 2^4 + 2^3 + 0 \cdot 2^2 + 2^1 + 0 \cdot 2^0)_{10} = (111010)_2. \end{aligned} \tag{4.20}$$

[4] siehe `https://en.wikipedia.org/wiki/Decimal_mark`

[5] siehe `https://de.wikipedia.org/wiki/Sexagesimalsystem`

Methode 2: (Divisionsmethode). Man dividiert sukzesive durch 2 und notiert die Reste. Die erhaltenen Reste schreibt man in umgekehrter Reihenfolge auf und erhält die Dualzahl.

$$\begin{aligned} 58 &= 29 \cdot 2 + \text{ Rest } 0, \\ 29 &= 14 \cdot 2 + \text{ Rest } 1, \\ 14 &= 7 \cdot 2 + \text{ Rest } 0, \\ 7 &= 3 \cdot 2 + \text{ Rest } 1, \\ 3 &= 1 \cdot 2 + \text{ Rest } 1, \\ 1 &= 0 \cdot 2 + \text{ Rest } 1. \end{aligned} \tag{4.21}$$

Die Reste in umgekehrter Reihenfolge lauten $(1, 1, 1, 0, 1, 0)$ und daher $(58)_{10} = (111010)_2$. Rückeinsetzen in (4.21) liefert sofort die Erklärung dieser Methode.

Umrechnung einer Zahl aus dem Hexdezimalsystem in das Dualsystem und umgekehrt:

Wenn man bei einer Hexdezimalzahl jede einzelne Ziffer in dualer Form durch 4 Dualziffern darstellt, so entspricht die entstandene Dualzahl in ihrem Wert der ursprüngliche Hexadezimalzahl, z.B.

$$(4\text{D}3)_{16} = (0100\ 1101\ 0011)_2. \tag{4.22}$$

Wenn man umgekehrt die 0, 1-Kette einer Dualzahl in Viererblöcke teilt und die jedem Viererblock entsprechende Hexadezimalzahl darstellt, dann hat die so gewonnene Hexadezimalzahl den gleichen Wert wie die ursprüngliche Dualzahl, z.B.

$$(1111\ 1111\ 1111)_2 = (\text{FFF})_{16}. \tag{4.23}$$

Beispiel 4.1. *Man stelle die Dezimalzahl* 0.1 *im Dualsystem dar.*

Der `Mathematica`-Befehl `BaseForm[x,b]` wandelt die Dezimalzahl `x` in eine Zahlendarstellung mit Basis `b` um

```
In[42]:=BaseForm[0.1, 2]
Out[42]//BaseForm=
      0.0001100110011001100110012
```

Die Zahl $(0.1)_{10}$ ist im Dualsystem eine Zahl mit unendlich vielen periodischen Nachkommastellen (`Mathematica` stellt natürlich nur endlich viele Nachkommastellen dar).

Es kann also - wie man in diesem Beispiel sieht - vorkommen, dass eine rationale Zahl in einem Zahlensystem nur *endlich* viele, in einem anderen System aber *unendlich viele periodische* Nachkommastellen hat. Niemals aber wird eine rationale Zahl in einem System unendlich viele *nicht-periodische* Nachkommastellen haben.

4.3. Maschinenzahlen

Ein Computer hat nur eine endliche Speicherkapazität und kann daher nur endlich viele Ziffern einer Zahl abspeichern (*bits* = binary digits). Jene Zahlen, die ein Rechner noch exakt darstellen kann, heißen **Maschinenzahlen**. Maschinenzahlen bilden also eine endliche Teilmenge der Menge der rationalen Zahlen. Alle anderen reellen Zahlen werden vom Computer immer auf die nächstgelegene Maschinenzahl gerundet.

Um einen möglichst grossen Zahlenbereich abzudecken werden Zahlen im Computer im sogenannten **Gleitpunkt-Format** gespeichert. Darunter verstehen wir eine Darstellung in der Form

$$x = \pm M \cdot 2^E.$$

Die Gleitpunktzahl M heißt **Mantisse** und wird so gewählt (IEEE 754 Standard)[6], dass die erste führende Ziffer ungleich 0 vor dem Punkt steht, $M = a_0.a_1a_2 \dots a_m$ wobei $a_0 = 1$ ist. Beispiel: $(101.1)_2$ wird in der Form $1.011 \cdot 2^{10}$ abgespeichert. Da a_0 immer 1 ist, kann es weggelassen werden. Das Vorzeichen wird als $(-1)^s$ durch ein Bit $s \in \{0, 1\}$ dargestellt. Die ganze Zahl E heißt **Exponent** und ist natürlich auch eine Binärzahl der Form

$$E = e - e_*, \qquad e = \sum_{j=0}^{k-1} e_j 2^j$$

wobei e^* ein fester Wert ist, sodass e immer dasselbe Vorzeichen hat. Mantisse und Exponent bestehen jeweils aus nur endlich vielen Stellen, die bei `double float` durch die Werte $m = 52, k = 11$ und $e_* = 1023$ gegeben sind (bei 32 bit floats ist $m = 24, k = 8$ und $e_* = 127$). Insgesamt erhält man folgenden Bit-Vektor

$$(s, e_{10}, e_9, \dots, e_0, a_1, a_2, \dots a_{52}) \in \{0, 1\}^{64}. \tag{4.24}$$

Versuchen wir uns das anhand eines kleinen Beispiels zu veranschaulichen. Damit es für uns leichter wird, stellen wir uns vor, dass der Computer Zahlen im Dezimalsystem darstellt. Unsere Überlegung gilt aber gleichermaßen für das Dualsystem bzw. für jedes beliebige Stellenwertsystem. Nehmen wir weiters an, dass es sich um einen sehr einfachen Computer mit Mantissenlänge 1 und Exponentlänge 1 handelt. Dann sind die positiven darstellbaren Zahlen gegeben durch

$$1 \cdot 10^{-9},\ 2 \cdot 10^{-9}, \dots, 9 \cdot 10^{-9},\ 1 \cdot 10^{-8},\ 2 \cdot 10^{-8}, \dots, 9 \cdot 10^{9}.$$

Die Maschinenzahlen dieses Computers können also den positiven Zahlenbereich von 0.000000001 bis 9000000000 abdecken. Dazu kommen noch ebensoviele negative Zahlen und die 0. Allerdings liegen die Maschinenzahlen **nicht gleichmässig** verteilt: Zwischen 1 und 10 liegen zB. genauso viele Maschinenzahlen (1, 2, 3, ..., 10) wie zwischen 10 und 100 (10, 20, 30, ..., 100), nämlich genau zehn.

[6] siehe: `http://en.wikipedia.org/wiki/IEEE_754-2008`

In der Regel kommt es daher bei der Verarbeitung von Kommazahlen durch den Computer zu Rundungsfehlern. Diese Fehler sind aber meist klein und können vernachlässigt werden.

Beispiel 4.2. Rundungsfehler
Gehen wir einfachheitshalber von einem Computer aus, der Zahlen im Dezimalsystem darstellt und der eine 4*-stellige Mantisse hat. Welches Ergebnis gibt der Computer für* $1.492 \cdot 1.066$ *aus? Wie groß ist der relative Fehler?*

Lösung zu 4.2. *Das exakte Ergebnis wäre* $1.492 \cdot 1.066 = 1.590472$*. Aufgrund der* 4*-stelligen Mantisse muss der Computer runden und gibt daher den Wert* 1.590 *aus. Der relative Fehler beträgt*

$$\frac{\textit{absoluter Fehler}}{\textit{exakter Wert}} = \frac{1.590472 - 1.590}{1.590472} \approx 0.0003$$

also 0.03 *Prozent.* ■

Allein durch die im Computer nötige Umwandlung vom Dezimal- ins Dualsystem können bereits Rundungsfehler auftreten. Beispiel 4.1 hat uns ja gezeigt, dass bei Umwandlung von 0.1 ins Dualsystem eine Zahl mit unendlich vielen Nachkommastellen entsteht. Diese Nachkommastellen müssen vom Computer abgebrochen werden.

Der relative Fehler aus Beispiel 4.2 wird in den meisten Anwendungen vernachlässigbar sein. Im folgenden Beispiel ergibt sich aber ein großer relativer Fehler:

Beispiel 4.3. Großer Rundungsfehler
Welches Ergebnis gibt unser Computer aus Beispiel 4.2 für die Berechnung von $(0.01 + 100) - 100 = 0.01$ *aus? Wie gross ist der relative Fehler?*

Lösung zu 4.3. *Die Zahlen* 0.01 *und* 100 *werden intern im Gleitkommaformat dargestellt als* $1 \cdot 10^{-2}$ *bzw.* $1 \cdot 10^{2}$*. Für die Addition müssen die beiden Zahlen in eine Form mit gleicher Hochzahl umgewandelt werden. Es ist (exakt)* $1 \cdot 10^{-2} = 0.0001 \cdot 10^{2}$*, unser Computer kann aber nur* 4 *Stellen der Mantisse abspeichern und muss daher auf* $0.000 \cdot 10^{2}$ *runden. Sein Ergebnis ist daher* 0.0000*! Der relative Fehler ist damit* $\frac{0.01-0}{0.01} = 1$*, also* 100%*.* ■

In bestimmten Situationen können sich Rundungsfehler aufsummieren und dadurch zu ziemlichen Problemen führen. So ist das im Golfkrieg[7] 1990/91 beim Steuerprogramm der amerikanische Raketen passiert: Während der kurzen Testphasen haben sich die Rundungsfehler nie ausgewirkt und wurden

[7] `http://de.wikipedia.org/wiki/Zweiter_Golfkrieg`

daher im Steuerprogramm nicht bemerkt. Beim längeren Betrieb während des Einsatzes haben sich die Fehler aber so weit aufsummiert, dass die Raketen ihr Ziel verfehlt haben.

4.4. Die reellen Zahlen

Da in $\mathbb{Q}$ die vier Grundrechnungsarten unbeschränkt durchführbar sind gibt die algebraische Struktur keinen Anlass zu einer weiteren Erweitung. Jedoch weisen Ordnungs- und topologische Strukur Unvollständigkeiten auf, die zur Erweiterung zum Bereich der reellen Zahlen $\mathbb{R}$ führt.

$(\mathbb{Q}, \leq)$ ist eine totalgeordnete Menge. Doch füllen die rationalen Zahlen die Zahlengerade nicht vollständig, obwohl sie dicht liegen. Zum Beispiel hat ein Quadrat mit der Seitenlänge 1 eine Diagonale der Länge $\sqrt{2}$ und man zeigt leicht (Beweis durch Widerspruch), dass dies keine rationale Zahl ist. Äquivalent dazu ist das Problem, dass die Menge $T = \{x \mid x \in \mathbb{Q} \text{ und } x^2 \leq 2\}$ keine untere Grenze in $\mathbb{Q}$ hat.

Um dieses Problem zu beheben werden die reellen Zahlen eingeführt. Es gibt im wesentlichen drei Verfahren: Dedekindsche Schnitte (Dedekind), Cauchyfolgen (Cantor) und Intervallschachtelungen (Weierstrass).

Definition 4.10. *Eine* Intervallschachtelung *ist eine Folge* $([a_n, b_n])$ *abgeschlossener Intervalle mit den Eigenschaften* $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n, a_j, b_j \in \mathbb{Q}, j \in \mathbb{N}$ *und* $\lim_{n\to\infty}(b_n - a_n) \to 0$.

Jedes Intervall ist also im vorhergehenden enthalten, und die Intervallänge schrumpft auf Null (konvergiert gegen Null[8]) zusammen. Die Intervalle können sich dabei auf eine rationale Zahl zusammenziehen, brauchen es aber nicht.

Da verschiedene Intervalle sich zur selben reellen Zahl zusammenziehen können, ist es notwendig eine entsprechende Äquivalenzklassenrelation einzuführen und dann die *Äquivalenzklasse* mit der entsprechenden *reellen Zahl* zu identifizieren.

Stellt man eine positive reelle Zahl im Dezimalsystem dar, so sind folgende Fälle möglich:

(i) endliche Dezimalzahl,

(ii) periodische Dezimalzahl,
Anmerkung: Bei Perioden mit 9 ist folgende Besonderheit zu beachten: $0.1 = 0.0\bar{9}$. ($\bar{9}$ ist die Abkürzung für $999\cdots$.)

(iii) unendliche Dezimalzahl.

[8] Konvergenz wird später im Kapitel 7 Def. 7.3 exakt definiert.

Die Menge der unendlichen nichtperiodischen Dezimalzahlen kann mit der Menge der irrationalen Zahlen identifiziert werden.

Beispiel: Intervallschachtelung von $\sqrt{2}$ im Dezimalsystem (in Dezimalpunktschreibweise) wobei die Intervalllänge um den Faktor 10 abnimmt.

$$\begin{aligned} I_0 &= [1, 2], \\ I_1 &= [1.4, 1.5], \\ I_2 &= [1.41, 1.42], \\ I_3 &= [1.414, 1.415], \\ I_4 &= [1.4142, 1.4143], \\ &\textit{etc.} \end{aligned} \tag{4.25}$$

Die Intervalllänge $b_n - a_n = 10^{-n}$ geht dabei gegen Null.

Satz 4.11. $(\mathbb{R}, \cdot, +, \leq)$ *bildet einen geordneten Körper (erfüllt die entsprechenden Rechengesetze)*

Zusätzlich zu den Grundrechnungsarten Addition und Multiplikation (inklusive "Subtraktion" und "Division") führt man die Operation des Potenzieren ein.

Definition 4.12. *Für $a \in \mathbb{R}$ definiert man*

$$\begin{aligned} a^0 &= 1, \\ a^{n+1} &= aa^n. \end{aligned} \tag{4.26}$$

Es gilt daher $0^0 = 1$ und

$$\begin{aligned} a^m a^n &= a^{m+n}, \\ a^m b^m &= (ab)^m, \\ (a^m)^n &= a^{mn}. \end{aligned}$$

Als inverse Operation zum Potenzieren wird das Wurzelziehen wie folgt definiert.

Definition 4.13.

$$\sqrt[n]{r} = x \qquad \textit{wobei } x \in \mathbb{R}^+ \textit{ und } x^n = r \textit{ erfüllt, } n \in \mathbb{N} \setminus \{0\}. \tag{4.27}$$

($\mathbb{R}^+ = \{x \mid x \in \mathbb{R} \text{ und } x > 0\}$.) Damit lässt sich der Potenzbegriff auf negative und rationale Exponenten erweitern.

Definition 4.14.

$$a^{-n} = \frac{1}{a^n}, \qquad n \in \mathbb{N},\ a \in \mathbb{R} \setminus \{0\},$$

$$a^{\frac{m}{n}} = (\sqrt[n]{a})^m,\ m \in \mathbb{Z},\ n \in \mathbb{N} \setminus \{0\}, a \in \mathbb{R}^+. \tag{4.28}$$

Für $m > 0$ darf auch $a = 0$ sein. Aber: $\sqrt{a}\sqrt{b} \neq \sqrt{a\,b}$ für $a < 0, b < 0$!

Definition 4.15. *Die Abbildung* $[\ ] : \mathbb{R} \to \mathbb{Z}, a \mapsto [a]$ *heißt* größtes Ganzes-*Funktion (Gaussklammer), wobei* $[a]$ *die größte ganze Zahl kleiner gleich* a *ist.*

Üblich ist dafür auch noch die Bezeichnung "floor" $\lfloor a \rfloor$. Mit "ceiling" $\lceil a \rceil$ wird die kleinste ganze Zahl größer gleich a bezeichnet. Und mit $\mathrm{int}(a)$ wird die Abbildung bezeichnet, die die Stellen nach dem Komma abschneidet (integer part). Ist $a \in \mathbb{Z}$, so ist $\lceil a \rceil = \lfloor a \rfloor = \mathrm{int}(a)$.
Übung: Was ist, wenn $a \notin \mathbb{Z}$?

Definition 4.16. *Die Abbildung* $\mathrm{sgn} : \mathbb{R} \to \{-1, 0, 1\}, a \mapsto \mathrm{sgn}(a)$ *heißt* Vorzeichenfunktion*, wobei* sgn *wie folgt definiert ist*

$$\mathrm{sgn}(a) = \begin{cases} +1 & a > 0 \\ 0 & a = 0 \\ -1 & a < 0 \end{cases}. \tag{4.29}$$

Definition 4.17. *Die Abbildung* $|\ | : \mathbb{R} \to \mathbb{R}^+ \cup \{0\}, a \mapsto |a|$ *heißt* Absolutbetrag*, wobei* $|\ |$ *wie folgt definiert ist* $|a| = a\,\mathrm{sgn}(a)$.

Es gilt

$$\begin{aligned} |0| &= 0, \\ |a| &= |-a|, \\ |ab| &= |a||b|, \\ |a+b| &\leq |a| + |b|, \\ |a+b| &\geq ||a| - |b||. \end{aligned} \tag{4.30}$$

4.5. Die komplexen Zahlen

Die quadratische Gleichung $x^2 + 1 = 0$ hat in der Menge der reellen Zahlen keine Lösung. Eine solche Lösung würde man als die Wurzel aus (-1) bezeichnen können. Um obige Gleichung lösen zu können, führt man daher die komplexen Zahlen ein.

Es sei $(x, y) \in \mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Dann erklärt man folgenderweise eine Addition und Multiplikation für Paare $z_1 = (x_1, y_1)$ und $z_2 = (x_2, y_2)$

$$z_1 + z_2 := (x_1 + x_2, y_1 + y_2),$$

$$z_1 \cdot z_2 := (x_1 \cdot x_2 - y_1 \cdot y_2, x_1 \cdot y_2 + x_2 \cdot y_1).$$

Man kann nun zeigen, dass $(\mathbb{R}^2, +, \cdot)$ ein Körper ist.

Definition 4.18. $(\mathbb{R}^2, +, \cdot)$ *mit den obigen Verknüpfungen heißt* Körper der komplexen Zahlen $\mathbb{C}$.

Definition 4.19. *Die komplexe Zahl* $\mathrm{i} := (0, 1)$ *heißt* imaginäre Einheit. *Jedes Zahlenpaar* (x, y) *kann damit nun als* $x + y\,\mathrm{i}$ *geschrieben werden und heißt* komplexe Zahl.

Die *imaginäre Einheit* i genügt der Gleichung

$$\mathrm{i}^2 + 1 = 0. \tag{4.31}$$

Man schreibt daher auch $\mathrm{i} = \sqrt{-1}$.

Definition 4.20. Einige grundlegenden Begriffe für komplexe Zahlen:

Für eine komplexe Zahl $z = x + \mathrm{i}\,y$ *heißt*

$\overline{z} = x - \mathrm{i}\,y$	*konjugiert komplexe Zahl*
$x = \mathrm{Re}(z) = \frac{1}{2}(z + \bar{z})$	*Realteil von* z
$y = \mathrm{Im}(z) = \frac{1}{2\,\mathrm{i}}(z - \bar{z})$	*Imaginärteil von* z
$\lvert z\rvert = \sqrt{x^2 + y^2} = \sqrt{z\bar{z}}$	*Absolutbetrag von* z

Es gilt für $z_1, z_2 \in \mathbb{C}$

$$\begin{aligned} |z_1 z_2| &= |z_1| \cdot |z_2|, \\ \left|\frac{z_1}{z_2}\right| &= \frac{|z_1|}{|z_2|}, \\ |z_1 + z_2| &\le |z_1| + |z_2|, \qquad \text{Dreiecksungleichung.} \end{aligned}$$

Beim Rechnen mit komplexen Zahlen geht man genauso vor wie bei reellen Zahlen (es gilt das Distributivgesetz, etc.), wobei man beachtet, dass $\mathrm{i}^2 = -1$.

Notiz: Identifiziert man $(x, 0)$ mit x, so gilt $\mathbb{R} \subset \mathbb{C}$, d.h. die reellen Zahlen können als Teilmenge der komplexen Zahlen aufgefaßt werden, nämlich als diejenigen Zahlen, deren Imaginärteil Null ist. Reelle Zahlen bestehen nur aus einem Realteil. Statt $x + \mathrm{i} \cdot 0$ schreibt man daher einfach x. Z.B. ist $5 + 0 \cdot \mathrm{i} = 5$. Zahlen, die nur aus Imaginärteil bestehen $z = 0 + \mathrm{i}\,y$, nennt man *imaginäre*

Zahlen. Eine imaginäre Zahl ist z.B. $0 + \mathrm{i}\,\pi = \mathrm{i}\,\pi$. Die komplexe Zahl $0 + 0\,\mathrm{i}$ kann man auch als 0 oder als $0\,\mathrm{i}$ schreiben.

Hier einige Beispiele:

(1) $\mathrm{i}\,(-\mathrm{i}) = 1$, $\sqrt{-16} = 4\,\mathrm{i}$, $-\sqrt{-9} = -3\,\mathrm{i}$, $\mathrm{i}^3 = -\,\mathrm{i}$.

(2) $(1 + \mathrm{i})(1 - \mathrm{i}) = 2$.

(3) $|4 + 3\,\mathrm{i}\,| = \sqrt{16 + 9} = 5$.

Eine komplexe Zahl kann auch als ein Punkt in der *Gauss'schen Zahlenebene* dargestellt (interpretiert) werden. Punkte in der Ebene können durch die Angabe von Abstand vom Ursprung und Winkel zur x-Achse festgelegt werden (Polarkoordinaten: (r, φ)), d.h. $z = x + \mathrm{i}\,y = r(\cos\varphi + \mathrm{i}\sin\varphi) = r\mathrm{e}^{\mathrm{i}\,\varphi}$.

Anmerkung 1: Die Verallgemeinerung der Differential- und Integralrechnung auf komplexe Funktionen ist ein umfangreiches und wichtiges Gebiet der Mathematik, das Funktionentheorie[9] genannt wird.

Anmerkung 2: Eine Erweiterung der komplexen Zahlen sind die Quaternionen $\mathbb{H}$. Sie bilden einen sogenannten Schiefkörper, da die Multiplikation *nicht mehr kommutativ* ist. Es gibt nun $\mathbb{R}$ mit den zusätzlichen Zahlen: $\mathrm{i}, \mathrm{j}, \mathrm{k}$ wobei folgende Regeln gelten:

$$z = a + b\,\mathrm{i} + c\,\mathrm{j} + d\,\mathrm{k}, \quad a, b, c, d \in \mathbb{R}$$

Die Addition entspricht der Vektoraddition von Viertupeln. Für $\mathrm{i}, \mathrm{j}, \mathrm{k}$ gelten folgende Multiplikationsregeln:

$$\begin{aligned} \mathrm{i}\,\mathrm{j}\,\mathrm{k} &= \mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1 \text{ und daraus folgt} \\ &\mathrm{i}\,\mathrm{j} = \mathrm{k}, \quad \mathrm{j}\,\mathrm{k} = \mathrm{i}, \quad \mathrm{k}\,\mathrm{i} = \mathrm{j}, \\ &\mathrm{j}\,\mathrm{i} = -\mathrm{i}\,\mathrm{j}, \quad \mathrm{k}\,\mathrm{j} = -\mathrm{j}\,\mathrm{k}, \quad \mathrm{i}\,\mathrm{k} = -\mathrm{k}\,\mathrm{i}. \end{aligned} \tag{4.32}$$

Quaternionen sind die erste Wahl zur Beschreibung von räumlichen Drehungen kombiniert mit Translationen in der Computergraphik (Computerspiele).
`http://mathworld.wolfram.com/Quaternion.html`
`http://de.wikipedia.org/wiki/Quaternionen`

[9] `http://de.wikipedia.org/wiki/Funktionentheorie`

4.6. Elementare Zählprinzipien

Abzählungen werden in der elementaren Wahrscheinlichkeitsrechnung verwendet[10].

Satz 4.21. Dirichletsches Schubfachprinzip *(pigeonhole principle). Werden n+1 Dinge auf n Schubladen verteilt, so befinden sich zumindest in einer Lade mindestens zwei Gegenstände.*

Oder allgemeiner, werden $kn + 1$ Dinge auf n Schubladen verteilt, so befinden sich zumindest in einer Lade mindestens $k + 1$ Gegenstände.

Beispiel 1: Gibt es in Wien zwei Personen, die die gleiche Anzahl von Kopfhaaren haben?[11]

Beispiel 2: Unter 1100 Studierenden gibt es mindestens 4, die am selben Tag Geburtstag haben.

Beispiel 3: Unter je $(n+1)$ Zahlen der Menge $\{1, 2, 3, \ldots 2n\}$ gibt es stets zwei teilerfremde.

Anordnungsprobleme

Gegeben seien 7 Personen, deren Namen einfach kurz mit $\{a, b, c, d, e, f, g\}$ bezeichnet werden.

1. Fragestellung

Wieviele verschiedene Sitzanordnungen sind möglich, wenn man 7 Stühle zur Verfügung hat?
Zum Beispiel sind (a, b, c, d, e, f, g), (b, c, d, e, f, g, a) und (g, f, e, d, c, b, a) mögliche Anordnungen. Jede dieser Anordnungen ist eine Permutation der 7 Elemente.

Wieviel verschiedene Anordnungen gibt es?

Für die erste Person stehen 7 Plätze, für die zweite 6 Plätze, für die dritte 5 Plätze, usw. zur Auswahl, sodass sich insgesamt $7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 7! = 5040$ Möglichkeiten ergeben.

Allgemein gilt

Satz 4.22. *Die Anzahl $P(n)$ der möglichen Anordnungen (n-Tupel) von n verschiedenen Elementen ist n!*

[10] Die elementare Wahrscheinlichkeit (Laplace) ist definiert durch „Anzahl der günstigen Fälle durch Anzahl aller möglichen Fälle."

[11] Im Schnitt besitzt ein Mensch ca. 100 000 bis 200 000 Haare.

Für große n ist $n!$ durch die *Stirlingsche Formel*[12] berechenbar. Es gilt

$$n! \approx \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}. \tag{4.33}$$

Weitere Beispiele:
(i) Wieviele verschiedene Sitzanordnungen sind möglich, wenn man 10 Personen auf 10 Stühle verteilen soll?
(ii) Wieviele Möglichkeiten gibt es vier Bilder nebeneinander an die Wand zu hängen?

2. Fragestellung

Es seien 3 Sitzplätze frei. Wieviele Möglichkeiten gibt es aus 7 Personen 3 auszuwählen und in unterschiedlicher Reihenfolge hinzusetzen, zum Beispiel (a, b, c), (e, f, g) oder (c, b, a), etc.

Es gibt offensichtlich $7 \cdot 6 \cdot 5 = 7 \cdot 6 \cdot 5 \cdot \frac{4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1} = \frac{7!}{4!}$ Möglichkeiten.

Satz 4.23. *Die Anzahl $P(n, k)$ der möglichen geordneten k-elementigen Auswahlen (k-Tupel) aus n verschiedenen Elementen ist*

$$P(n, k) = n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}. \tag{4.34}$$

Anmerkung:
(i) Geordnete Auswahlen nennt man auch Variationen ohne Wiederholung oder auch geordnete Proben ohne Wiederholung. D. h. es sei A eine Menge und $k \in \mathbb{N}$. Dann heißt das k-Tupel $(a_1, \ldots, a_k) \in A^k$ mit $a_i \neq a_j$ für $i \neq j$ geordnete Probe ohne Wiederholung.
(ii) Ist $k = n$, so ergibt sich als Sonderfall $P(n, n) = P(n)$.
(iii) $P(n, k)$ gibt die Anzahl der injektiven Abbildungen von k Elementen in eine Menge von n Elementen an ($k \leq n$).

Beispiel: Gegeben seien die vier Zahlen $\{1, 2, 3, 4\}$. Auf wieviele Arten lassen sich daraus zwei auswählen? Zum Beispiel $(1, 2), (1, 3), (1, 4), (2, 1), (2, 3)$, etc.

3. Fragestellung

Wieviele Möglichkeiten erhält man, wenn man bei obigem Beispiel Wiederholungen zulässt? Also $(1, 1), (2, 2), (3, 3)$, etc. Dann gibt es offensichtlich $4 \cdot 4$ Möglichkeiten. Allgemein gilt:

Satz 4.24. *Die Anzahl $\tilde{P}(n, k)$ der möglichen geordneten k-elementigen Anordnungen von n verschiedenen Elementen mit Wiederholung ist*

$$\tilde{P}(n, k) = n^k. \tag{4.35}$$

[12] `https://en.wikipedia.org/wiki/Asymptotic_analysis`

Anmerkung: Man nennt sie auch Variationen mit Wiederholung oder auch geordnete Proben mit Wiederholung. D. h. es sei A eine Menge und $k \in \mathbb{N}$. Dann heißt das k-Tupel $(a_1, \ldots, a_k) \in A^k$ geordnete Probe mit Wiederholung.

$\tilde{P}(n,k)$ gibt die Anzahl aller Abbildungen f von einer Menge mit k Elementen in eine Menge von n Elementen an; z.B. $f : X \to Y, \quad \#(f) = |Y|^{|X|}$.

Beispiele:

(i) Wie viele Möglichkeiten gibt es einen Toto-Schein auszufüllen?

(ii) Wieviele voneinander verschiedene Wörter mit drei Buchstaben können theoretisch gebildet werden?

4. Fragestellung

Wieviele verschiedene Dreiergruppen können aus einer Menge von 7 Personen ausgewählt werden, zum Beispiel $\{a, b, c\}$, $\{e, f, g\}$ oder $\{b, d, e\}$, etc. Es kommt dabei nicht auf die Reihenfolge an (ungeordnete Auswahl). Jede dieser möglichen Auswahlen nennt man Kombination. Ihre Anzahl ist

$$\frac{7 \cdot 6 \cdot 5}{1 \cdot 2 \cdot 3} = \binom{7}{3}.$$

Allgemein gilt

Satz 4.25. *Die Anzahl $C(n,k)$ der möglichen Kombinationen von k Elementen aus einer Menge von n verschiedenen Elementen ist*

$$C(n,k) = \binom{n}{k}. \tag{4.36}$$

Anmerkung: (i) Es sei A eine Menge mit $|A| = n$ und $k \leq n,\ k \in \mathbb{N}$. Dann heißt jede Teilmenge $B \subset A$ mit $|B| = k$ ungeordnete Probe ohne Wiederholung.
(ii) Es gilt: $C(n,k) = P(n,k)/k!$

Beispiele: (i) Lotto 6 aus 45.
(ii) Wieviele Möglichkeiten gibt es aus einer Gruppe von 12 Studentinnen ein Team mit 5 Studentinnen auszuwählen.

Es bleibt die Anzahl der ungeordneten Auswahlen (Proben, Kombination) mit Wiederholung zu bestimmen. Zuerst ein Beispiel.

5 Personen wählen aus drei Menüs aus. Gesucht ist die Anzahl der möglichen Bestellungen, die der Kellner aufnimmt.

Satz 4.26. *Die Anzahl der Auswahlmöglichkeiten von k Elementen aus n Elementen mit Wiederholung (Kombinationen mit Wiederholung) ist*

$$\tilde{C}(n,k) = \binom{n+k-1}{k}. \tag{4.37}$$

Beispiel:

(i) In unserem Beispiel ist $n = 3$ und $k = 5$ und somit erhält man
$\tilde{C}(3,5) = \binom{7}{5} = 21$.
(ii) Wieviele Dominosteine gibt es, wenn die maximale Augenzahl 7 ist
$\tilde{C}(7,2) = \binom{8}{2} = 28$.
(iii) Drei Kinder haben 5 Äpfel. Wieviele Möglichkeiten gibt es diese ganz aufzuteilen?

Satz 4.27. *Die Anzahl der Möglichkeiten, n Gegenstände auf r Fächer zu verteilen, wobei im i-ten Fach k_i Gegenstände zu liegen kommen, ist:*

$$\binom{n}{k_1, \dots k_r} = \frac{n!}{k_1! \dots k_r!}. \tag{4.38}$$

Beispiel: Wieviele verschiedene Wörter können aus den Buchstaben des Wortes MISSISSIPPI gebildet werden? Es gibt $\frac{11!}{1!4!4!2!}$ Möglichkeiten.

Überblick Kombinatorik

geordnet		ungeordnet	
ohne Wiederholung	mit Wiederholung	ohne Wiederholung	mit Wiederholung
$P(n,k) = \frac{n!}{(n-k)!}$	$\tilde{P}(n,k) = n^k$	$C(n,k) = \binom{n}{k}$	$\tilde{C}(n,k) = \binom{n+k-1}{k}$

Zusammenfassung

(i) Folgende Begriffe sollten ihnen nun bekannt sein

Zahlen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, und ihre Darstellung, Summenzeichen, Produktzeichen, Schubfachprinzip, Permutation, Variation, Kombination.

(ii) Sie sollten nun können: Rechnen mit Zahlen und einfache kombinatorische Abzählprobleme lösen können.

4.7. Übung

(1) Man berechne die folgenden Summen, wenn $a_1 = 1$, $a_2 = 3$, $a_3 = -1$, $a_4 = 2$, und $a_5 = -2$:
a) $\sum_{i=1}^{3} a_i$, b) $\sum_{k=3}^{5} a_k$, c) $\sum_{n=1}^{5} a_n$, d) $\sum_{j=1}^{5} a_j$.

(2) Man berechne die folgenden Summen, wenn $a_1 = 1$, $a_2 = 3$, $a_3 = -1$, $a_4 = 2$:
a) $\sum_{i=2}^{4} \frac{1}{a_i}$, b) $\sum_{i=1}^{4} 2a_i$, c) $\sum_{n=1}^{4}(a_n + 1)$, d) $\sum_{j=1}^{4} a_j^2$.

(3) Man schreibe mithilfe des Summenzeichens
a) die Summe, b) die Summe der Quadrate
der natürlichen Zahlen von 1 bis 100.

(4) Man schreibe die folgenden Summen mit dem Summenzeichen:

a) $a_3 + a_4 + a_5 + a_6$, b) $3x_1 + 3x_2 + 3x_3 + 3x_4$,
c) $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}$, d) $1^3 + 2^3 + 3^3 + 4^3 + 5^3$,
e) $1 + 2^1 + 2^2 + 2^3 + 2^4$, f) $2 \cdot 4^1 + 2 \cdot 4^2 + 2 \cdot 4^3 + 2 \cdot 4^4$.

(5) Man schreibe mithilfe des Summenzeichens:

a) $a_0 - a_1 + a_2 - a_3 + a_4 - a_5$, b) $3 - 3^2 + 3^3 - 3^4 + 3^5 - 3^6$,
c) $\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{10}$, d) $2^1 + 2^3 + 2^5 + 2^7 + 2^9 + 2^{11}$,
e) $1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!}$, f) $1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots - \frac{x^{30}}{30!}$.

(6) Man schreibe mithilfe des Produktzeichens:
a) $1 \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{4} \cdot \cdots \cdot \frac{1}{k}$, b) $\left(1 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{3}\right) \cdot \cdots \cdot \left(1 - \frac{1}{n}\right)$,
c) $(n-1)!$

(7) Beispiel Peano-Axiome
Mit Hilfe der Funktion "Nachfolger" berechne man mit *Mathematica* die Summe und das Produkt zweier natürlicher Zahlen a und b
$a++$ ist der Nachfolger von a
$a--$ ist der Vorgänger von a

(8) Man ermittele $\mathrm{int}(\cdot), \lceil\cdot\rceil, \lfloor\cdot\rfloor$ von
a) 2.4, b) 2.9, c) -1.9, d) -0.9.

(9) Man berechne:
a) 17 mod 3, b) 5 mod 2, c) 28 mod 5, d) 351 mod 4,
e) 13856 mod 2, f) 357 mod 3, g) 6237 mod 9,
h) 3751 mod 11.

(10) Man berechne:
a) $|-4|+|4|$, b) $|-2|+|-1|$, c) $\big||-3|-2\big|$,
d) $4-|-3|$, e) $|-3|\cdot|+2|$, f) $\big||-1|-|-2|\big|$.

(11) Wahr oder falsch?:
a) $|2|<|-5|$, b) $|3|>|-100|$, c) $|-3|=|3|$,
d) $|-5|<|0|$, e) $|2|\leq|-2|$, f) $|a|=|-a|$.

(12) Man finde einen Wert für a bzw. b, für die die folgenden Behauptungen falsch sind:
a) $|a|=a$, b) $|-a|=-|a|$, c) $|a+b|=|a|+|b|$.

(13) Man schreibe die folgenden Mengen in der Intervallschreibweise und kennzeichnen Sie sie auf der Zahlengeraden:
a) $\{x\in\mathbb{R}\mid -2\leq x<3\}$, b) $\{x\in\mathbb{R}\mid 0<x\leq 2\}$,
c) $\{x\in\mathbb{R}\mid 1\leq x\leq 3\}$, d) $\{x\in\mathbb{R}\mid 2<x<5\}$,
e) $\{x\in\mathbb{R}\mid x<4\}$, f) $\{x\in\mathbb{R}\mid x\geq 2\}$.

(14) Man gebe alle reellen Zahlen an (Intervallschreibweise) die
a) größer als -1, b) mindestens gleich 0, c) höchstens 0,
d) größer als -2 und höchstens gleich 4 sind.

(15) Man zeichne die Funktionen:
a) $y=|\frac{x}{2}-1|$, b) $y=|x+1|$.

4.8. Übung

(1) Man löse durch Fallunterscheidung. Veranschaulichen Sie graphisch indem Sie die linke und rechte Seite der Gleichung als Funktionen betrachten (Lösungen der Gleichung = Schnittpunkte der Funktionen):
a) $|x-1|=x^2$, b) $|\frac{x}{2}-1|=x$, c) $x^2=4$, d.) $|x+3|=5-|x-2|$.

(2) Man löse folgende Gleichungen:
a) $2\sqrt{x+3}-\sqrt{2x+1}=\sqrt{4x-3}$, b) $\log_2(\frac{x}{2})+\log_2(1-\frac{x}{2})=-2$.

(3) Man löse folgende Ungleichungen:
a) $1-x>2$, b) $2x\leq 3+5x$, c) $2(x+2)-4x<3(2x-1)+4$.

(4) Man löse folgende Ungleichungen. Man veranschauliche graphisch indem man die linke und rechte Seite der Ungleichung als Funktionen betrachtet:

a) $\dfrac{x}{2}+1>5$, b) $|2-x|\geq 1$, c) $(x-2)^2<1$, d) $x^2\leq|x|$.

(5) (a) Man bestimme ein $x_0 \in \mathbb{Z}$, sodass für alle $x \in \mathbb{Z},\ x \geq x_0$ gilt:

$$2x^2 - 3x - 6 > 0.$$

(b) Es sei $1 > \epsilon > 0$. Man bestimme alle $x > \frac{1}{2}, x \in \mathbb{R}$, sodass

$$\left| \frac{x+1}{2x-1} - 2 \right| < \epsilon.$$

4.9. Übung Zahlensysteme

(1) Man wandle die folgenden Dualzahlen in Dezimalzahlen um:
a) 1110101, b) 101010, c) 11110.

(2) Man wandle die folgende Hexadezimalzahlen in Dezimalzahlen um:
a) $14D$, b) $3EF$, c) AAA, d) $12CD$.

(3) Man bestimme die Dualdarstellung der folgenden Dezimalzahlen:
a) 45, b) 319, c) 99, d) 16, e) 59.

(4) Man bestimme die Hexadezimaldarstellung der folgenden Dezimalzahlen. Man wandle die Hexadezimalzahl danach in eine Dualzahl um:
a) 1406, b) 510, c) 319.

(5) Man wandle die folgenden Dualzahlen in Hexadezimalzahlen um (durch Zusammenfassung in Vierergruppen). Man mache die Probe im Dezimalsystem.
a) 11010110, b) 10100111, c) 10111111.

(6) Wie lautet die Darstellung der folgenden Dezimalzahlen im Dualsystem/Hexadezimalsystem? (Tip: bei den größeren Zahlen ist es am einfachsten, zunächst ins Hexadezimalsystem und von da ins Dualsystem umzuwandeln)
a) 45, b) 319, c) 1501, d) 1438.

(7) Man wandle die folgenden binären Bruchzahlen ins Dezimalsystem um:
a) 11.0111, b) 10.0101, c) 0.111, d) 0.1.

(8) Wie lauten die folgenden Dezimalbrüche im Binärsystem (Berechnung von 4 Nachkommastellen, falls unendlich viele)?
a) 0.4375, b) 0.5, c) 0.2, d) 4.713, e) 0.125, f) 1.25.

4.10. Übung: komplexe Zahlen

(1) Man addiere die folgenden komplexen Zahlen graphisch:
$z_1 = -2 + 4\mathrm{i}$, $z_2 = -3 - 2\mathrm{i}$, $z_3 = 4 - \mathrm{i}$.

(2) Man führe mit $z = 2 + \mathrm{i}$ folgende Operationen durch und deuten Sie sie geometrisch in der Gaußschen Zahlenebene:
a) $z \cdot \mathrm{i}$, b) $\frac{z}{\mathrm{i}}$, c) $z \cdot \mathrm{e}^{\mathrm{i}\cdot 30^\circ}$, d) z^2, e) $\frac{1}{z}$.

(3) Man stelle folgende Zahlenmengen in der Gaußschen Zahlenebene dar:
a) $|z| \le 3$, b) $\arg(z) = 30^\circ$, c) $\mathrm{Re}(z+1) = 4$.

(4) Man löse in $\mathbb{C}$:
a) $3x^2 = 2x - \frac{5}{3}$, b) $\dfrac{1}{x+1} + \dfrac{1}{x-1} - \dfrac{1}{x} = 0$.

(5) Bei der Berechnung von elektrischen Netzwerken bei Wechselströmen treten lineare Gleichungssysteme auf. Man löse das lineare Gleichungssystem in $\mathbb{C}$:
a) $\begin{aligned}(3+2\mathrm{i})z_1 - \mathrm{i}z_2 &= -2 + 3\mathrm{i}\\ (1-3\mathrm{i})z_1 + (2-\mathrm{i})z_2 &= 5(1-2\mathrm{i}),\end{aligned}$
b) $\begin{aligned}4x + 3y &= 11 + 9\mathrm{i}\\ x + 3z &= 2 - 3\mathrm{i}\\ x + y + z &= 3 + 2\mathrm{i}.\end{aligned}$

(6) Man berechne:
a) $(2-\mathrm{i})\mathrm{e}^{\mathrm{i}\cdot 150^\circ}$, b) $\dfrac{-6+8\mathrm{i}}{5(\cos 90^\circ + \mathrm{i}\sin 90^\circ)}$, c) $\dfrac{4\mathrm{e}^{\mathrm{i}\frac{3\pi}{4}}}{5\mathrm{e}^{\mathrm{i}\cdot 60^\circ}}$, d) i^{i}.

(7) Man berechne mit $z_1 = 3\mathrm{e}^{-\mathrm{i}\cdot\frac{\pi}{2}}$, $z_2 = 2\mathrm{e}^{\mathrm{i}\cdot 60^\circ}$ und $z_3 = \frac{1}{2}(\cos 30^\circ + \mathrm{i}\sin 30^\circ)$ die folgenden Terme und stelle sie sowohl in der Komponentenform $z = a + b\,\mathrm{i}$ als auch (mit Rechnerunterstützung) in Exponentialform $z = r\,\mathrm{e}^{\mathrm{i}\varphi}$ dar:
a) $\dfrac{\bar{z}_1 z_3}{\bar{z}_2} - \dfrac{1}{z_1}$, b) $\dfrac{|z_1 + z_2|}{3z_3}$, c) $1 + \mathrm{i}\,\dfrac{z_1^2}{z_2 - z_3}$.

(8) Man löse die folgenden kubischen Gleichungen, bei denen eine Lösung leicht erraten werden kann. Man spalte diese Lösungen ab (Polynomdivison) und ermittele die restlichen Lösungen:
a) $x^3 - 4x = 0$, b) $x^3 - x^2 - x + 1 = 0$, c) $x^3 + 2x^2 + 2x + 1 = 0$.

(9) Man berechne (Polynomdivison mit Rest):

$$(3\,x^4 + 2\,x^2 + 3) \,:\, (x^2 - 2\,x + 1)$$

(10) Man löse folgende quadratische Gleichungen mit *komplexen* Koeffizienten. Man beachte: die Lösungen sind nicht mehr konjugiert komplex, das ist nur bei *reellen* Koeffizienten der Fall!
a) $x^2 - (1+\mathrm{i})\,x + \mathrm{i} = 0$, b) $x^2 - (1+2\mathrm{i})\,x - 1 + \mathrm{i} = 0$.

(11) Man berechne alle Lösungen von $x^3 = 1$ und $x^3 = 2 + 2\,\mathrm{i}$.

4.11. Übung Abzähltechniken

(1) Man ermittle:

$$\text{(a) } \binom{10}{3}, \qquad \text{(b) } \binom{10}{1}, \qquad \text{(c) } \binom{10}{0}, \qquad \text{(d) } \binom{200}{198}.$$

(2) Man berechne den Koeffizienten von

$$\text{(a) } a^3b^7 \text{ in } (a+b)^{10}, \text{ (b) } x^3 \text{ in } (2+3\,x)^6.$$

(3) Auf wieviele Arten kann man 70, 256 oder 1000 Bilder anordnen. Man gebe eine Abschätzung der Anzahl als Gleitkommazahl an.

(4) Bei einer Betriebseröffnung werden 4 Personen a, b, c, d eine Rede halten. Wieviele verschiedene Rednerlisten können aufgestellt werden,
(a) bei beliebig möglicher Anordnung,
(b) wenn a stets an erster Stelle spricht,
(c) wenn d nicht an erster Stelle sprechen möchte?

(5) Eine Gruppe von 10 Mitarbeitern soll fotografiert werden. Für 4 Personen steht eine Bank zur Verfügung. Auf wieviele Arten kann die Bank besetzt werden, wenn eine unterschiedliche Sitzreihenfolge
(a) ohne Bedeutung
(b) von Bedeutung ist?

(6) Wieviele verschiedene Sitzordnungen gibt es für 5 Personen in einem fünfsitzigen PKW, wenn
(a) jede Person einen Führerschein hat,
(b) nur drei Personen einen Führerschein haben?

(7) Vertragsabschluss nach Treffen von 8 Geschäftsleuten. Zur Feier wird mit Sekt angestoßen. Wie viele Male klirren die Gläser, wenn jeder mit jedem anstosst?

(8) 10 Ehepaare veranstalten eine Tanzparty. Wieviele Tanzpaare sind möglich, wenn Ehepartner nicht miteinander tanzen dürfen?

(9) Wieviele Möglichkeiten gibt es, beim Lotto „6 aus 45“ keine, genau eine, genau 2, ..., genau 6 Richtige anzukreuzen?

(10) In einer zehnköpfigen Komission hat jedes Mitglied eine Stimme. Wieviele mögliche Mehrheiten (Mehrheit = mindestens 6 Stimmen) gibt es?

(11) Jeder Telefonanschluss hat eine Vorwahlnummer und eine n-stellige Telefonnummer. Die Vorwahlnummer besteht aus 4 Ziffern, von denen die erste immer 0 und die zweite immer $\neq 0$ sein muß. Wieviele Telefonanschlüsse sind grundsätzlich möglich?

(12) Wieviele Möglichkeiten gibt es, aus einer Versammlung von 30 Teilnehmern einen vierköpfigen Vorstand, (1 Vorsitzender, 3 gleichberechtigte Mitglieder) zu wählen?

(13) Kennzeichen für Kraftfahrzeuge (vereinfacht): Einem Kennbuchstaben (A...Z) folgt eine vierstellige Ziffernkombination, bestehend aus den Ziffern 0 bis 9. Wieviele mögliche Kennzeichen gibt es?

(14) Geburtstagparadoxon: Wie groß ist die Wahrscheinlichkeit, dass bei einer Gruppe von 24 Personen zwei am gleichen Tag Geburtstag haben. Hinweis: Man berechne zuerst die Wahrscheinlichkeit, dass alle an verschiedenen Tagen Geburtstag haben. (Praktische Anwendung in der Informatik: Hashkollision)

(15) Bei einem Sonderangebot kann man sich eine Kiste mit zwölf Flaschen aus drei verschiedenen Getränkesorten beliebig zusammenstellen. Wieviele Möglichkeiten gibt es dafür?

(16) (a) Ein Passwort kann aus sechs bis acht Zeichen bestehen (Kleinbuchstaben oder Ziffern). Wie viele mögliche Passwörter gibt es?
(b) Angenommen, mindestens eines der Zeichen des Passworts muss eine Ziffer sein. Wie viele mögliche Passwörter gibt es dann?

(17) In einer Urne sind eine rote, blaue und grüne Kugel. Es werden 4 Ziehungen gemacht, wobei nach jeder Ziehung die Kugel wieder zurück gelegt wird. Wieviele verschiedene Ergebnisse der Ziehung sind möglich, wenn
(a) die Reihenfolge von Bedeutung ist
(b) die Reihenfolge ohne Bedeutung ist.

(18) Im 17. Jahrhundert bot Chevalier de Méré zwei Wetten an. In der ersteren wettete er, dass er in 4 Würfen mit einem Würfel mindestens einen Sechser hat. In der zweiten wettete er, dass er in 24 Würfen mit zwei Würfel mindestens einen Doppelsechser hat. Welche dieser Wetten würden sie machen?

Kapitel 5

Elementare Zahlentheorie

5.1. Teilbarkeit

Definition 5.1. *Ein* Integritätsbereich *ist ein kommutativer Ring R mit Einselement, der* nullteilerfrei *ist, d. h. aus $a \cdot b = 0,\ a, b \in R$ folgt entweder $a = 0$ oder $b = 0$.*

Das wichtigste Beispiel für einen Integritätsbereich bildet der Ring der ganzen Zahlen $\mathbb{Z}$ auf den wir uns in der Anwendung beschränken werden. Ein weiteres Beispiel liefert der Ring der Polynome in einer Unbestimmten X über einem Körper $\mathbb{K}$. Er besteht aus allen Polynomen der Form

$$a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n, \qquad a_j \in \mathbb{K}, n \in \mathbb{N} \tag{5.1}$$

und wird mit $\mathbb{K}[X]$ bezeichnet.

Definition 5.2. *Es seien $a, b \in \mathbb{Z}$ (oder allgemeiner Elemente eines Integritätsbereiches). Man nennt a einen* Teiler *von b, wenn eine Zahl $q \in \mathbb{Z}$ existiert, sodass $aq = b$ gilt. Man schreibt $a|b$ (a teilt b). Teilt a die Zahl b nicht, schreibt man $a \nmid b$.*

Aus dieser Definition folgen sofort ein paar einfache Regeln:

(i) $a|0$ für alle $a \in \mathbb{Z}$,
(ii) $0|a$ nur für $a = 0$,
(iii) $a|1$ nur für $a = \pm 1$,

(iv) $a|b$ und $b|a$ nur wenn $a = \pm b$,
(v) $\pm 1|a$ und $\pm a|a$ für alle $a \in \mathbb{Z}$,
(vi) $a|b$ und $b|c$ impliziert $a|c$ (Transitivität),
(vii) $a|b_1, \ldots, a|b_n$ impliziert $a|(b_1c_1 + \cdots b_nc_n)$ für alle $c_1, \ldots, c_n \in \mathbb{Z}$.

Dies gibt Anlaß zu folgenden Bezeichnungen

Definition 5.3. *(i) Die Zahlen* 1 *und* -1 *heißen* Einheiten *von* $\mathbb{Z}$. *Allgemein heißt ein Element* u *eines Ringes Einheit, wenn ein* $v \in R$ *existiert mit* $uv = 1$. *Die Menge der Einheiten wird mit* R^* *bezeichnet.*

(ii) Zwei Elemente $a, b \in R \setminus \{0\}$ *heißen* assoziiert $(a \sim b)$, *falls eine Einheit* $u \in R$ *existiert mit* $au = b$.

(iii) a *heißt* echter *Teiler von* b, *wenn* $a|b$ *und* a *nicht assoziiert zu* b *ist.*

(iv) a *heißt* trivialer *Teiler von* b, *wenn* a *Einheit oder assoziiert zu* b *ist.*

Anmerkung: Die Einheiten sind diejenigen Elemente die ein Inverses besitzen.

Übung: Man zeige:

(i) $n^3 + 11n$ ist durch 6 teilbar.
(ii) $11^{n+1} + 12^{2n-1}$ ist durch 133 teilbar.

Zur Charakterisierung der echten Teiler a einer ganzen Zahl b kann man sich auf die positiven Teiler beschränken.

Definition 5.4. *Eine natürliche Zahl* $n > 1$ *heißt* Primzahl, *falls sie nur triviale Teiler besitzt.*

Man kann eine Primzahl auch als ganze Zahl die genau zwei positive Teiler besitzt charakterisieren.

Die Liste der ersten fünf Primzahlen lautet: $\{2, 3, 5, 7, 11\}$.

Anmerkung: Jede natürliche Zahl $n > 1$ besitzt mindestens einen Primteiler.

Satz 5.5. *Es gibt unendlich viele Primzahlen.*

Beweis: Annahme: es gibt nur endlich viele Primzahlen $\{p_1, p_2, \ldots, p_n\}$. Man betrachte nun die Zahl $b = p_1 \cdot p_2 \cdot \cdots \cdot p_n + 1$. Diese wäre laut Annahme prim, aber nicht in obiger Liste enthalten, was zu einem Widerspruch führt.

Anmerkung: Es gilt: $b = p_1 \cdot p_2 \cdot \cdots \cdot p_n + 1$ ist im allgemeinen nicht prim, aber die primen Teiler von $n! + 1$ sind größer als n.

Definition 5.6. *Ein gemeinsamer Teiler* d *von endlich vielen Zahlen* $a_1, a_2, \ldots, a_n$ *heißt* größter gemeinsamer Teiler *von* $a_1, a_2, \ldots, a_n$, *wenn* $d \geq 1$

und jeder gemeinsame Teiler von $a_1, a_2, \ldots, a_n$ ein Teiler von d ist. Man schreibt $d = \mathrm{ggT}(a_1, a_2, \ldots, a_n)$, bzw. $d = \gcd(a_1, a_2, \ldots, a_n)$.

Satz 5.7. (Division mit Rest) *Zu jedem Paar ganzer Zahlen a, b mit $b \neq 0$ gibt es genau ein Paar ganzer Zahlen q, r, das die Bedingungen*

$$a = qb + r, \qquad 0 \leq r < |b| \tag{5.2}$$

erfüllt.

r heißt der Rest von a bezüglich der Divison durch b.

Satz 5.8. *Erweiterter* **Euklidischer Algorithmus.** *Zu beliebigen, endlich vielen ganzen Zahlen $a_1, a_2, \ldots, a_n$ gibt es genau einen größten gemeinsamen Teiler d und es gilt*

$$d = \mathrm{ggT}(a_1, a_2, \ldots, a_n) = a_1c_1 + a_2c_2 + \ldots + a_nc_n \tag{5.3}$$

mit gewissen ganzen Zahlen $c_1, c_2, \ldots, c_n$.

d. h. d kann als eine Linearkombination der Zahlen $a_1, a_2, \ldots, a_n$ geschrieben werden. Die Zahlen $c_1, c_2, \ldots, c_n$ sind aber nicht eindeutig bestimmt; vergleiche Satz 5.14.

Beweis: siehe Körner [**47**], Seite 4.

Wir demonstrieren die Berechnung des größten gemeinsamen Teilers zweier Zahlen $a, b, a > b > 0$ mittels Euklidischem Algorithmus durch fortgesetzte Division mit Rest wie folgt

$$\begin{aligned}
a &= v_0\, b + r_1, && 0 < r_1 < b,\\
b &= v_1\, r_1 + r_2, && 0 < r_2 < r_1,\\
r_1 &= v_2\, r_2 + r_3, && 0 < r_3 < r_2,\\
&\vdots\\
r_{n-2} &= v_{n-1}\, r_{n-1} + r_n, && 0 < r_n < r_{n-1},\\
r_{n-1} &= v_n\, r_n + 0.
\end{aligned} \tag{5.4}$$

Da $\mathrm{ggT}(a, b)$ die linke Seite in obigen Gleichungen teilt, teilt er auch jeweils die rechte Seite. Die fortgesetzte Division mit Rest bewirkt, dass die Reste r_j immer kleiner werden bis der Rest 0 ist. Der vorletzte Rest $r_n \neq 0$ ist dann der größte gemeinsame Teiler d. Auflösen der Gleichung (5.4) nach r_n ergibt dann die Darstellung $d = r_n = c_1a + c_2b$.

Definition 5.9. *Endlich viele ganze Zahlen $a_1, a_2, \ldots, a_n$ heißen* teilerfremd *(relativ prim), wenn $d = \mathrm{ggT}(a_1, a_2, \ldots, a_n) = 1$ ist.*

Nach Satz 5.8 sind ganze Zahlen $a_1, a_2, \ldots, a_n$ genau dann teilerfremd, wenn 1 eine ganzzahlige Linearkombination von ihnen ist.

Definition 5.10. *Eine ganze Zahl d heißt* kleinstes gemeinsames Vielfaches *von endlich vielen Zahlen $a_1, a_2, \ldots, a_n$, wenn folgende Bedingungen erfüllt sind:*

(i) $d \geq 0$,
(ii) d ist gemeinsames Vielfaches von $a_1, a_2, \ldots, a_n$, d. h. $a_1|d, \ldots, a_n|d$,
(iii) d teilt jedes gemeinsame Vielfache von $a_1, a_2, \ldots, a_n$.

Man schreibt $d = \mathrm{kgV}(a_1, a_2, \ldots, a_n)$, (bzw. $d = \mathrm{lcm}(a_1, a_2, \ldots, a_n)$).

Es gilt: $\mathrm{kgV}(a_1, a_2) = \dfrac{a_1 a_2}{\mathrm{ggT}(a_1, a_2)}$.

5.2. Primfaktorzerlegung

Satz 5.11. *(i) Für ganze Zahlen a, b, c gilt: Aus $a|bc$ und $(a, b) = 1$ folgt $a|c$.*

(ii) Teilt eine Primzahl ein Produkt von ganzen Zahlen, so teilt sie mindestens einen Faktor.

Damit kann man nun den Satz von der eindeutigen Primfaktorzerlegung zeigen.

Satz 5.12. *(i) Jede natürliche Zahl $n > 1$ ist ein Produkt von Primzahlen.*

(ii) Die Zerlegung einer natürlichen Zahl $n > 1$ in Primfaktoren ist bis auf die Reihenfolge eindeutig.

Bezeichnet $\alpha_p = \alpha_p(a)$ wie oft (Anzahl) die Primzahl p als Faktor in der Primfaktorzerlegung von a vorkommt, so kann man a in der Form

$$a = \prod_p p^{\alpha_p} \tag{5.5}$$

schreiben.

Daraus konstruiert man leicht die Menge aller Teiler einer ganzen Zahl.

Ist n eine zusammengesetzte Zahl, dann besitzt sie mindestens einen Primfaktor kleiner gleich $\sqrt{n}$.

Zahlen der Form

$$M_p = 2^p - 1,$$

wobei p prim ist, heißen *Mersenne Zahlen*. Bis heute sind 51 Mersenne-Primzahlen[1] bekannt.

[1] `http://de.wikipedia.org/wiki/Mersenne-Primzahl`

Zahlen der Form

$$F_n = 2^{2^n} + 1$$

heißen *Fermat-Zahlen*[2]. F_1, F_2, F_3, F_4 sind Primzahlen. Euler zeigte 1732, dass F_5 durch 641 teilbar und daher keine Primzahl ist und widerlegte damit Fermat's Vermutung (1640), dass alle F_n Primzahlen sind (siehe Scheid [**55**] Seite 173ff).

Dass genügend Primzahlen für die Verschlüsselung zur Verfügung stehen wird durch den Primzahlsatz garantiert.

Satz 5.13. *Es bezeichne $\pi(x)$ die Anzahl der Primzahlen kleiner gleich x. Dann gilt*

$$\pi(x) \sim \frac{x}{\log x}, \tag{5.6}$$

$$\pi(x) = \mathrm{Li}(x) + O(\sqrt{x}\log x) \tag{5.7}$$

wobei (5.7) unter der Annahme der Gültigkeit der Riemannschen Vermutung gilt. Hier bezeichnet $\sim$ die asymptotische Äquivalenz, O das Landausymbol (siehe Kapitel 7.1.5) und Li den Integrallogarithmus.[3]

$$\mathrm{Li}(x) := \int_2^x \frac{1}{\log t}\, dt. \tag{5.8}$$

5.3. Kongruenzen

Eine Gleichung mit ganzzahligen Koeffizienten und deren ganzzahlige Lösungen gesucht sind heißt *diophantische Gleichung.*

Satz 5.14. *Die lineare diophantische Gleichung*

$$ax + by = c \qquad a, b, c \in \mathbb{Z} \tag{5.9}$$

ist genau dann lösbar in $\mathbb{Z}$, wenn $\mathrm{ggT}(a,b)|c$. Ist in diesem Fall (x_0, y_0) eine spezielle Lösung, dann ist

$$\{(x_0 + w\,\frac{b}{d}, y_0 - w\,\frac{a}{d}) \mid w \in \mathbb{Z}\} \qquad \textit{mit } d = \mathrm{ggT}(a,b) \tag{5.10}$$

die Menge aller Lösungen.

Anmerkung: Eine spezielle Lösung berechnet man z. B. mit dem euklidischen Algorithmus oder Euler-Verfahren, siehe Scheid [**55**] Seite 195.

Beispiel: "Hundert Maß Korn werden unter hundert Leute so verteilt, daß jeder Mann drei Maß, jede Frau zwei Maß und jedes Kind ein halbes Maß erhält. Wieviele Männer, Frauen und Kinder sind es?" (6 Lösungen).

[2] `http://de.wikipedia.org/wiki/Fermat-Zahl`

[3] `https://de.wikipedia.org/wiki/Integrallogarithmus`

Definition 5.15. *Es sei m eine natürliche Zahl. Eine ganze Zahl a heißt* kongruent *zu einer ganzen Zahl b* modulo m*, wenn $m|(a-b)$. Man schreibt*

$$a \equiv b \pmod{m}. \tag{5.11}$$

Die Zahl m heißt Modul der Kongruenz. Falls $m \nmid (a-b)$ schreibt man $a \not\equiv b \pmod{m}$.

Satz 5.16. *Genau dann gilt $a \equiv b \pmod{m}$, wenn a und b den gleichen Rest bezüglich der Division durch m besitzen.*

Man weist leicht nach, dass man mit Kongruenzen wie folgt rechnen kann.

Satz 5.17. *Aus $a \equiv a' \pmod{m}$ und $b \equiv b' \pmod{m}$ folgt*

$$a + b \equiv a' + b' \pmod{m}, \tag{5.12}$$

$$a - b \equiv a' - b' \pmod{m}, \tag{5.13}$$

$$a \cdot b \equiv a' \cdot b' \pmod{m}. \tag{5.14}$$

Anwendung: Aus der Zifferndarstellung einer Zahl $a = a_0 + a_1 10^1 + a_2 10^2 + \ldots + a_n 10^n$ folgt damit sofort, dass $a \equiv (a_1 + \ldots + a_n) \pmod{3}$ ist, da $10 \equiv 1 \pmod{3}$. D. h. eine Zahl ist genau dann durch 3 teilbar, wenn ihre Ziffernsumme durch 3 teilbar ist. Analoges gilt für Teilbarkeit durch 9. Weitere Teilbarkeitsregeln findet man z. B. in Scheid [**55**] Seite 135.

Nur beim Dividieren ("Kürzen") muß man vorsichtig sein, wie das folgende Beispiel zeigt: $2 \cdot 3 \equiv 3 \cdot 4 \pmod{6}$ aber $2 \not\equiv 4 \pmod{6}$! Es gilt:

Satz 5.18. *Es seien $a, b, c \in \mathbb{Z}$ und $m \in \mathbb{N}$.*

(i) Zu a existiert eine ganze Zahl a' mit

$$a\,a' \equiv 1 \pmod{m} \tag{5.15}$$

genau dann, wenn $\mathrm{ggT}(a, m) = 1$ ist.

(ii) Ist $\mathrm{ggT}(a, m) = 1$, so folgt aus $ab \equiv ac \pmod{m}$ stets $b \equiv c \pmod{m}$. Ist dagegen $\mathrm{ggT}(a, m) \neq 1$, so gibt es Zahlen b, c mit $ab \equiv ac \pmod{m}$ und $b \not\equiv c \pmod{m}$.

D. h. kürzen durch a ist im Allgemeinen nicht erlaubt, ausgenommen der Modul m ist relativ prim zu a. Die zu a inverse Zahl a' kann mit dem Euklidischen Algorithmus berechnet werden, da Gleichung (5.15) equivalent zur linearen diophantischen Gleichung $a\,a' + k\,m = 1,\ k \in \mathbb{Z}$ ist.

Satz 5.19. *Es seien $a, b \in \mathbb{Z}, m \in \mathbb{N}$. Die Kongruenz*

$$ax \equiv b \pmod{m} \tag{5.16}$$

ist genau dann lösbar, wenn $\mathrm{ggT}(a, m)|b$. Die Anzahl der Lösungen ist $\mathrm{ggT}(a, m)$.

Satz 5.20. *Aus* $a \equiv b \pmod{m}$ *folgt* $\mathrm{ggT}(a, m) = \mathrm{ggT}(b, m)$.

Beispiele: Man bestimme im Falle der Lösbarkeit alle Lösungen von

(a) $4x \equiv 6 \pmod{10}$, (b) $3x \equiv 5 \pmod{21}$, (c) $21x \equiv 49 \pmod{13}$,
(d) $25x \equiv 15 \pmod{40}$.

Wie man mehrere Kongruenzen zu verschiedenen Moduln gleichzeitig löst, zeigt der folgende Satz der schon früh in der Kalenderrechnung verwendet wurde.

Satz 5.21. *(***Chinesischer Restsatz***)*

Es seien $a_1, \ldots, a_r \in \mathbb{Z}$ *und* $m_1, \ldots, m_r \in \mathbb{N}$ *paarweise teilerfremde natürliche Zahlen, d. h.* $\mathrm{ggT}(m_i, m_j) = 1, i \neq j$. *Dann besitzt das System der* r *Kongruenzen*

$$x \equiv a_i \pmod{m_i}, \qquad i = 1, \ldots, r \tag{5.17}$$

genau eine Lösung x *modulo* m, *wobei* $m = m_1 \cdots m_r$.

Genau dann ist $\mathrm{ggT}(x, m) = 1$, *wenn*

$$\mathrm{ggT}(a_i, m_i) = 1, \qquad i = 1, \ldots, r \text{ gilt.}$$

Beweis: Körner [**47**], Seite 13.

Beispiele: Man bestimme eine Lösung des Systems

$$x \equiv 1 \pmod{2}, \quad x \equiv 2 \pmod{3}, \quad x \equiv 3 \pmod{5}, \quad x \equiv 5 \pmod{7}.$$

Anwendung bei Kartentricks: [**35**] siehe Seite 84, bzw. 101ff.

5.4. Restklassen

Auf der Menge der ganzen Zahlen ist die Relation gegeben durch $a \; R \; b$ genau dann, wenn $a \equiv b \pmod{m}$ eine Äquivalenzrelation. Diese bewirkt daher eine Klasseneinteilung. Die Menge der Restklassen modulo m wird mit $\mathbb{Z}_m$ (bzw. $\mathbb{Z}/m\mathbb{Z}$) bezeichnet. Die Menge

$$\{0, 1, \ldots, m-1\} \tag{5.18}$$

stellt ein vollständiges Repräsentantensystem modulo m dar.

In $\mathbb{Z}_m$ lassen sich Addition und Multiplikation über Repräsentanten erklären.

$$[a] + [b] := [a + b]$$
$$[a] \cdot [b] := [a \cdot b]$$

Mit diesen beiden Verknüpfungen bilden die Restklassen modulo m einen Ring. Ist der Modul p eine Primzahl so bilden die Restklassen modulo p einen Körper. Für die multiplikative Gruppe $\mathbb{Z}_p^*$ schreibt man $\mathbb{F}_p^*$ (field = Körper). Sie besteht aus $(p-1)$ Elementen.

Definition 5.22. *(i) Die Anzahl aller primen Restklassen modulo m wird mit $\varphi(m)$ bezeichnet und heißt* Eulersche φ-Funktion *von m.*

(ii) Eine Menge, bestehend aus zu m teilerfremden ganzen Zahlen, heißt reduziertes Restsystem modulo m, wenn sie von jeder primen Restklasse modulo m genau einen Repräsentanten enthält. Diese Gruppe wird mit $\mathbb{Z}_m^*$ bezeichnet und besteht aus $\varphi(m)$ Elementen.

Beispiel: $\varphi(10) = 4$, wobei das Restsystem repräsentiert wird durch $\{1, 3, 7, 9\}$.

Zur Berechnung von $\varphi(m)$ verwendet man folgenden Satz.

Satz 5.23. *Sei $m \in \mathbb{N},\ m > 1$ und $m = p_1^{\alpha_1} \cdots p_r^{\alpha_r}, \alpha_j > 0, 1 \leq j \leq r$ ihre Primzahlzerlegung. Dann gilt*

$$\varphi(m) = p_1^{\alpha_1 - 1} \cdots p_r^{\alpha_r - 1}(p_1 - 1) \ldots (p_r - 1). \tag{5.19}$$

Beweis: Körner [**47**], Seite 18.

Weiters gilt:

$$\varphi(m\,n) = \varphi(m)\,\varphi(n) \tag{5.20}$$

und

$$\varphi(m) = m \prod_{p|m} (1 - \frac{1}{p}) \tag{5.21}$$

wobei p prim ist.
Beispiel: berechne $\varphi(360)$.

5.5. Die Sätze von Fermat, Euler und Wilson

Satz 5.24. *(Lagrange) Es sei G eine endliche Gruppe der Ordnung* $\mathrm{ord}(G)$ *(Anzahl ihrer Elemente) und $H \subset G$ eine Untergruppe. Dann ist* $\mathrm{ord}(H)$ *ein Teiler von* $\mathrm{ord}(G)$*.*

Beweis: Sei $g \in G$ ein beliebiges Element. Die Menge

$$gH := \{x \mid x = gh,\ h \in H\}$$

heißt *Linksnebenklasse* von H. Da G eine Gruppe ist, ist klar, dass jede Linksnebenklasse von H ebenso viele Elemente wie H hat. Weiters gilt für zwei Nebenklassen g_1H und g_2H genau einer der beiden Fälle

(i) $g_1H = g_2H$ $\veebar$ (ii) $g_1H \cap g_2H = \emptyset$.

Tritt (ii) nicht ein existiert ein $a \in g_1H \cap g_2H$, also $h_1, h_2 \in G$ mit $a = g_1h_1 = g_2h_2$. Daraus folgt $g_1^{-1}g_2 = h_1h_2^{-1} \in H$. Sei nun $x \in g_1H$ beliebig vorgeben, dann ist $x = g_1y$ mit $y \in H$ also

$$x = g_1(g_1^{-1}g_2)(h_1h_2^{-1})^{-1}y = g_2(h_2h_1^{-1}y) \in g_2H. \tag{5.22}$$

Damit ist bewiesen, dass $g_1H \subset g_2H$. Aus Symmetriegründen folgt auch $g_2H \subset g_1H$. Die Gruppe G ist also die disjunkte Vereinigung der Linksnebenklassen $x_1H, \ldots, x_rH$, woraus folgt $\mathrm{ord}(G) = r \cdot \mathrm{ord}(H)$. □

Man beachte: Nebenklassen sind keine Gruppen.

Daraus folgt unmittelbar, dass Gruppen deren Ordnung eine Primzahl ist keine Untergruppen besitzen können.

Definition 5.25. *Es sei x Element einer Gruppe G. Dann heißt die Menge $< x >:= \{x^n \mid n \in \mathbb{N}\}$ die von x erzeugte Untergruppe von G und* $\mathrm{ord}(x) := \mathrm{ord}(< x >)$.

D.h. $< x >= \{x, x^2, x^3, \ldots, x^{\mathrm{ord}(x)} = 1\}$.

Weitere Folgerungen.

Satz 5.26. *Es sei G eine endliche Gruppe und $x \in G$. Dann ist* $\mathrm{ord}(x)$ *ein Teiler von* $\mathrm{ord}(G)$.

Denn $\mathrm{ord}(x)$ ist gleich der Ordnung der von x erzeugten Untergruppe $< x >$.

Satz 5.27. *Es sei G eine endliche Gruppe. Dann gilt für jedes $x \in G$*

$$x^{\mathrm{ord}(G)} = e, \tag{5.23}$$

wobei e das neutrale Element von G ist.

Beweis: Nach Satz 5.26 gibt es ein r mit $\mathrm{ord}(G) = r\,\mathrm{ord}(x)$. Daraus folgt $x^{\mathrm{ord}(G)} = x^{r\,\mathrm{ord}(x)} = e^r = e$.

Die Anwendung dieses Satzes auf die Gruppe der teilerfremden Restklassen $\mathbb{Z}_m^*$ liefert

Satz 5.28. *(Euler) Für $m \in \mathbb{N}$ und $a \in \mathbb{Z}$ mit* $\mathrm{ggT}(a, m) = 1$ *gilt*

$$a^{\varphi(m)} \equiv 1 \pmod{m}. \tag{5.24}$$

Anwendung: Das Inverse a' in $a\,a' \equiv 1 \pmod{m}$ ist also durch $a^{\varphi(m)-1}$ bestimmt.
Weiteres Beispiel: Man bestimme die letzte Stelle der Zahl 7^{1002}.

Für den Spezialfall, dass der Modul eine Primzahl ist, ergibt aus dem Satz von Euler der Satz von Fermat.

Satz 5.29. *(kleiner Fermat) Es sei p eine Primzahl. Dann gilt für jede nicht durch p teilbare ganze Zahl a*

$$a^{p-1} \equiv 1 \pmod{p}. \tag{5.25}$$

Die Ordnung von $\mathbb{Z}_p^*$ ist $(p-1)$. Äquivalent dazu ist die Formulierung. Es gilt

$a^p \equiv a \pmod{p}$ für alle $a \in \mathbb{Z}$ und alle Primzahlen p.

Der Satz von Fermat liefert aber nur eine *notwendige* Bedingung für die Primzahlsuche.

Leider gilt nicht der umgekehrte Fall, denn es existieren zusammengesetzte Zahlen die obige Gleichung erfüllen.

Beispiel 1: man wählt $a = 2$. Ist n eine Primzahl, dann gilt: $n|(2^n - 2)$, aber nicht umgekehrt! Eine zusammengesetzte natürliche Zahl n mit $n|(2^n-2)$ nennt man eine *Pseudoprimzahl* oder auch *chinesische* Primzahl.

Beispiel 2: man untersuche ob 341 eine Primzahl ist. Es gilt $2^{340} \equiv 1 \pmod{341}$. Also funktioniert $a = 2$ nicht. Mit $a = 3$ erhält man aber $3^{340} \equiv 56 \pmod{341}$ und somit kann 341 keine Primzahl sein.

Eine hinreichende und notwendige Bedingung für eine Primzahl gibt aber der Satz von Wilson.

Satz 5.30. *(Wilson) Jede natürliche Zahl $p \geq 2$ ist genau dann eine Primzahl, wenn*

$$(p-1)! \equiv -1 \pmod{p}. \tag{5.26}$$

Jede Zahl $p > 1$, für die diese Kongruenz gilt, muss eine Primzahl sein, denn jeder Teiler t von p mit $0 < t < p$ teilt $(p-1)!$, folglich nach der Kongruenz auch -1, d. h. es muß $t = 1$ sein. Für die Umkehrung siehe [**47**], Seite 105 oder [**41**], Seite 57.

Für die Praxis ist dieser Test wegen des schnellen Wachstums von $(p-1)!$ ungeeignet. Da werden probabilistische Primzahltests verwendet, z.B. der Miller-Rabin-Test[4].

5.6. Das RSA-Kryptographie-Verfahren

Das RSA-Kryptographie-Verfahren wurde von Rivest, Shamir und Adleman entwickelt. Es ist ein "Public Key Verfahren"[5], d. h. der zur Chiffrierung gebrauchte Schlüssel ist öffentlich, so dass jeder damit Nachrichten verschlüsseln

[4] `https://de.wikipedia.org/wiki/Miller-Rabin-Test`

[5] Asymmetrisches Verschlüsselungssystemen eingeführt von `https://de.wikipedia.org/wiki/Whitfield_Diffie`

und an den Schlüsselinhaber senden kann. Trotz Kenntnis dieses Schlüssels ist die Entschlüsselung sehr schwierig. Die Methode beruht darauf, daß es viel leichter ist festzustellen ob eine Zahl prim ist, als sie dann in Primfaktoren zu zerlegen.

Man wähle zwei verschieden große Primzahlen p und q und bilde ihr Produkt $N = pq$. Die multiplikative Gruppe $\mathbb{Z}_N^*$ hat dann die Ordnung $\varphi(N) = (p-1)(q-1)$ (vgl. Satz 5.23). Weiter wählt man eine zu $\varphi(N)$ teilerfremde Zahl e, d.h. $\mathrm{ggT}(\varphi(N), e) = 1$. Dann existiert eine Zahl d mit

$$e\,d \equiv 1 \pmod{\varphi(N)}.$$

D. h. Es gilt: $e\,d = \tilde{q}\,\varphi(N) + 1$.

Mit diesen Daten, von denen N und e veröffentlicht werden, dagegen $p, q, \varphi(N)$ und d geheim gehalten werden, kann man auf der Menge $\mathbb{Z}_N$ eine Verschlüsselung E (e ... encryption) und eine Entschlüsselung D (d ... decryption) wie folgt konstruieren.

Verschlüsselung

$$E : \mathbb{Z}_N \to \mathbb{Z}_N,\ x \mapsto x^e. \tag{5.27}$$

Entschlüsselung

$$D : \mathbb{Z}_N \to \mathbb{Z}_N,\ x \mapsto x^d. \tag{5.28}$$

Es gilt der folgende Satz, der zeigt dass D die Umkehrabbildung von E ist.

Satz 5.31. *Mit den obigen Beziehungen gilt*

$$E \circ D = D \circ E = \mathrm{Id}_{\mathbb{Z}_N}. \tag{5.29}$$

Beweis: Ist x teilerfremd zu N so gilt $(x^e)^d = x^{ed} \equiv x \pmod{N}$. Ist x nicht teilerfremd zu N verwendet man die Isomorphie $\mathbb{Z}_N^* \sim \mathbb{Z}_p^* \times \mathbb{Z}_q^*$, $x \sim (x_1, x_2)$. Dann gilt auch $e\,d \equiv 1 \pmod{\varphi(p-1)}$ und $e\,d \equiv 1 \pmod{\varphi(q-1)}$. Somit erhält man $\sim (x_1, x_2)^{e\,d} = (x_1, x_2)$ (vergleiche Forster [**41**], Seite 123).

Man kann dieses Verfahren im folgenden Schema zusammenfassen:

1. Ein Benutzer wählt zwei sehr verschieden große Primzahlen p und q und ein Paar von Zahlen $e, d, 1 \le e, d \le (p-1)(q-1)$, relativ prim zu $\varphi(N) = (p-1)(q-1)$ mit $ed \equiv 1 \pmod{(p-1)(q-1)}$.
2. Der Benutzer gibt als öffentlichen Schlüssel das Produkt $N = p\,q$ und e bekannt.
3. Ein Text wird als Zahl T in $\{0, \ldots, N-1\}$ dargestellt; falls T zu groß ist, wird der Text zuvor in Blöcke zerlegt.

4. Die Codierung erfolgt mittels

$$C \equiv T^e \pmod{N}, \tag{5.30}$$

die Decodierung mittels des geheimen Schlüssels d durch

$$D \equiv C^d \pmod{N}. \tag{5.31}$$

Beispiel: Aigner [**33**], Seite 250
Es sei $p = 47, q = 59, N = 47 \cdot 59 = 2773, (p-1)(q-1) = 2668$. Man wählt $d = 157$ (Primzahl) und $e = 17$ mit $17 \cdot 157 \equiv 1 \pmod{2668}$. Zum Versenden des Texts nimmt man zum Beispiel: Zwischenraum $= 00, A = 01, B = 02, \ldots, Z = 26$. Je zwei Buchstaben werden nun zu einem 4-Block zusammengefasst. Aus dem Text

KOMME MORGEN ZURUECK

wird dann der Text

1115 1313 0500 1315 1807 0514 0026 2118 2105 0311.

Wegen $N = 2773$ sind alle 4-Blöcke Zahlen die kleiner N sind. Die einzelnen Blöcke T werden nun gemäß $T^{17} \pmod{2773}$ verschlüsselt und man erhält als verschlüsselte Botschaft (Kryptogramm)

1379 2395 1655 0422 0482 1643 1445 0848 0747 2676.

Anmerkung: Ganz allgemein benötigt man nur die Gruppenstruktur um etwas zu verschlüsseln, da man zum Entschlüsseln (Umkehrung) immer das inverse Element braucht. Zum Beispiel kann man rationale Punkte auf elliptischen Kurven addieren und diese Addition hat eine Gruppenstruktur.[6]

[6] `https://de.wikipedia.org/wiki/Elliptische_Kurve` und `https://de.wikipedia.org/wiki/Elliptic_Curve_Cryptography`

5.7. Übung

(1) Man bestimme die 6 Lösungen.
Hundert Maß Korn werden unter hundert Leute so verteilt, dass jeder Mann drei Maß, jede Frau zwei Maß und jedes Kind ein halbes Maß erhält. Wieviele Männer, Frauen und Kinder sind es?

(2) Man bestimme im Falle der Lösbarkeit alle Lösungen von

(a) $4\,x \equiv 6 \pmod{10}$, (b) $3\,x \equiv 5 \pmod{21}$,

(c) $21\,x \equiv 49 \pmod{13}$, (d) $25\,x \equiv 15 \pmod{40}$.

(3) Man bestimme eine Lösung des Systems

$$x \equiv 1 \pmod{2},\quad x \equiv 2 \pmod{3},\quad x \equiv 3 \pmod{5},\quad x \equiv 5 \pmod{7}.$$

(4) Man berechne

$$1193\,x \equiv 367 \pmod{1260}.$$

(5) Karl hat 35 Karten die er in 7 Spalten zeilenweise (rechteckig) auslegt. Judith sucht sich eine Karte davon aus und sagt, dass sie in Spalte 3 liegt. Nun legt Karl die Karten in 5 Spalten aus. Jetzt ist die Karte von Judith in Spalte 2. Die wievielte Karte hat sich Judith ausgesucht?

(6) (a) Welche der beiden IBAN-Nummern[7] ist korrekt?

(i) AT67 3100 0004 0401 1011 (ii) AT67 3100 0004 0411 1011

(b) Man berechne die zweistellige Prüfziffer[8] x in

$$\mathrm{AT}x\ 3746\ 8000\ 0041\ 7717$$

(7) Man berechne die letzte Ziffer der Zahl

$$z = 9^{8^{7^{6^{5^{4^{3^{2^{1}}}}}}}}.$$

(8) Mit Hilfe des Satzes von Euler berechne man die letzte Ziffer der Zahl $7^{1000002}$, d.h. $7^{1000002} \pmod{10}$.

(9) Man verwende die Daten aus dem Skriptum bzw. Aigner, Seite 250: Es sei $p = 47, q = 59, N = 47 \cdot 59 = 2773, (p-1)(q-1) = 2668$. Man wählt $d = 157$ (Primzahl) und $e = 17$ mit $17 \cdot 157 \equiv 1 \pmod{2668}$. Zum Versenden des Texts nehme man: Zwischenraum $= 00, A = 01, B = 02, \ldots, Z = 26$. Je zwei Buchstaben werden nun zu einem 4-Block zusammengefasst.

---

[7] `https://de.wikipedia.org/wiki/IBAN`

[8] mehr zu Prüfziffern siehe[**30**] Kapitel 3.2.1

(a) Damit verschlüssle man folgenden Text:

HEUTE KEINE VORLESUNG

(b) Man entschlüssle folgende Botschaft:

1084 0057 2342 2568 1163 2202 0761

Hinweis: Man benutze den Befehl **PowerMod[...,e, N]** in Mathematica oder Wolfram Alpha zur Berechnung der Kongruenzen.

Kapitel 6

Lineare Algebra und analytische Geometrie

Die Standardeinführungen zur linearen Algebra sind [**64**], [**66**], [**60**] oder [**59**], wobei [**66**] die Grundlagen am knappesten, aber dennoch vollständig presentiert. Aber auch [**24**], [**88**] Band 2 oder [**18**] enthalten die grundlegenden Tatsachen. Spezielle Bücher für Informatiker sind [**72**], [**73**] und [**70**]. Zur Vertiefung sei auf [**67**] und [**77**] verwiesen.

Online verfügbar ist zum Beispiel die Vorlesung von G. Strang [**76**] mit videos `https://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010/video-lectures/` und Mathematica notebooks auf `http://library.wolfram.com/infocenter/MathSource/546/`.

Wir orientieren uns am Buch von Klaus Jänich [**66**], Lineare Algebra, Springer Verlag (2000).

Wir haben im Kapitel 3 die algebraische Struktur eines Vektorraumes über einem Körper $\mathbb{K}$ eingeführt, wobei wir für $\mathbb{K}$ den Körper $\mathbb{R}$, $\mathbb{C}$ oder $\mathbb{Z}_p$ wählen.

Die Struktur eines Vektorraumes bildet die Grundlagen der linearen Algebra.

Anmerkung: Wir schreiben im allgemeinen Vektoren ohne Pfeil $\vec{}$. Wenn es jedoch für die Klarheit der Darstellung nützlich erscheint, verwenden wir gelegentlich den Pfeil $\vec{}$, zum Beispiel zur Unterscheidung von Null 0 und dem Nullvektor $\vec{0}$.

6.1. Untervektorräume

Ist V ein Vektorraum, so kann man eine beliebige Teilmenge $U \subset V$ betrachten. Man kann beliebige Elemente von U miteinander addieren oder mit einem Element von $\mathbb{K}$ multiplizieren, aber es wird im allgemeinen nicht so sein, dass für zwei beliebige Elemente $a, b \in U$ auch ihre Summe $a + b$ in U ist. U wird daher im allgemeinen kein Vektorraum sein.

Definition 6.1. *Gegeben sei ein Vektorraum V. Ist $U \subset V$ so heisst U* Untervektorraum *von V, wenn $U \neq \emptyset$ und für alle $a, b \in U$ und $\lambda \in \mathbb{K}$ gilt:*

$$a + b \in U,\ \lambda a \in U.$$

Untervektorräume nennt man auch noch *lineare Teilräume*.

Es gilt: Ist U ein Untervektorraum von V, dann ist auch der Nullvektor von V und mit jedem Vektor $a \in U$ auch der inverse Vektor $-a$ in U enthalten, da $0 \cdot a = \vec{0}$ und $(-1) \cdot a = -a$.

Beweis: Wir beweisen, dass aus den Axiomen eines abstrakten Vektorraumes folgt

(i) $0 \cdot \vec{a} = \vec{0}$ und

(ii) $(-1) \cdot \vec{a} = -\vec{a}$.

Ad (i): Es gilt

$$0 \cdot \vec{a} = (0 + 0) \cdot \vec{a} = 0 \cdot \vec{a} + 0 \cdot \vec{a},$$

da $0 + 0 = 0$ ist. Inverser Vektor und Assoziativität der Vektoraddition ergeben

$$\vec{0} = 0 \cdot \vec{a} + (-0 \cdot \vec{a}) = (0 \cdot \vec{a} + 0 \cdot \vec{a}) + (-0 \cdot \vec{a}) = 0 \cdot \vec{a}.$$

Punkt (ii) folgt mit $1 \cdot \vec{a} = \vec{a}$ aus

$$\vec{0} = 0 \cdot \vec{a} = (1 + (-1)) \cdot \vec{a} = \vec{a} + (-1) \cdot \vec{a} = \vec{a} + (-\vec{a}). \quad \square$$

Satz 6.2. *Ist U ein Untervektorraum von V, so bildet U einen Vektorraum über $\mathbb{K}$, wobei die Vektoraddition und Skalarmultiplikation durch V gegeben ist.*

Triviale Fälle von Unterräumen sind $U = V$ und $U = \{\vec{0}\}$.

Der mengentheoretische Durchschnitt zweier Untervektorräume bildet wieder einen Untervektorraum.

Beispiele: Wie schauen die nichttrivialen Untervektorräume des $\mathbb{R}^2$ oder des $\mathbb{R}^3$ aus?

6.2. Dimensionen

Es sei V ein Vektorraum über $\mathbb{K}$ und $v_1, v_2, \ldots, v_r \in V,\ \lambda_1, \lambda_2, \ldots, \lambda_r \in \mathbb{K}$. Dann nennt man

$$\sum_{j=1}^{r} \lambda_j v_j = \lambda_1 v_1 + \cdots + \lambda_r v_r$$

eine **Linearkombination** der Vektoren $v_1, v_2, \ldots, v_r$.

Definition 6.3. *Sei $v_1, v_2, \ldots, v_r \in V$. Die Menge*

$$L(v_1, v_2, \ldots, v_r) := \{\lambda_1 v_1 + \cdots + \lambda_r v_r \mid \lambda_j \in \mathbb{K}\} \tag{6.1}$$

aller möglichen Linearkombination von r festen Vektoren nennt man die lineare Hülle *des r-tupels $(v_1, v_2, \ldots, v_r)$ von Vektoren. Für die lineare Hülle des 0-tupels setzt man $L(\emptyset) := \{\vec{0}\}$.*

Anmerkung: Manche Autoren wählen in der Definition der linearen Hülle nicht ein r-tupel, sondern eine **ungeordnete Menge** von r Vektoren.

Da die Summe zweier Linearkombinationen von $(v_1, v_2, \ldots, v_r)$ wieder eine Linearkombination von $(v_1, v_2, \ldots, v_r)$ ist und auch das λ-fache einer Linearkombinationen von $(v_1, v_2, \ldots, v_r)$ wieder eine solche ist, gilt

Satz 6.4. *Die lineare Hülle $L(v_1, v_2, \ldots, v_r)$ bildet einen Untervektorraum von V.*

Beispiel: Wie schaut die lineare Hülle eines Vektors (zweier paralleler, nicht paralleler Vektoren) im $\mathbb{R}^2$ oder $\mathbb{R}^3$ aus?

Ein r-tupel $(v_1, v_2, \ldots, v_r)$ heisst linear abhängig, wenn man mindestens einen Vektor als Linearkombination der anderen schreiben kann, ansonsten heisst es linear unabhängig.

Zur Definition des Begriffes *linear abhängig* wählen wir jedoch folgende zweckmässigere Definition und zeigen anschliessend ihre Gleichwertigkeit.

Definition 6.5. *Es sei V ein Vektorraum über $\mathbb{K}$. Ein r-tupel $(v_1, v_2, \ldots, v_r)$ von Vektoren in V heisst* linear unabhängig*, wenn eine Linearkombination von $(v_1, v_2, \ldots, v_r)$ nur dann Null sein kann, wenn alle Koeffizienten λ_j Null sind, d. h. aus*

$$\sum_{j=1}^{r} \lambda_j v_j = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_r v_r = 0 \tag{6.2}$$

folgt stets

$$\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0. \tag{6.3}$$

Das 0-tupel ist per Definition linear unabhäng.

Satz 6.6. *$(v_1, v_2, \ldots, v_r)$ ist genau dann linear unabhängig, wenn keiner dieser Vektoren als Linearkombination der anderen darstellbar ist.*

Beweis:

(i) Wir zeigen zuerst, dass aus $(v_1, v_2, \ldots, v_r)$ ist linear unabhängig folgt, dass kein Vektor v_j eine Linearkombination der anderen Vektoren ist, indem wir dazu einen Widerspruch konstruieren.

Wir nehmen also an $(v_1, v_2, \ldots, v_r)$ ist linear unabhängig, aber es gäbe einen Vektor v_j der sich als Linearkombination darstellen lässt, d. h.

$$v_j = \sum_{\substack{k=1 \\ k \neq j}}^{r} \lambda_k v_k = \lambda_1 v_1 + \cdots + \lambda_{j-1} v_{j-1} + \lambda_{j+1} v_{j+1} + \cdots + \lambda_r v_r.$$

Umformen ergibt

$$\lambda_1 v_1 + \cdots + \lambda_{j-1} v_{j-1} + (-1)\, v_j + \lambda_{j+1} v_{j+1} + \cdots + \lambda_r v_r = 0,$$

was im Widerspruch zur Tatsache steht, dass alle $\lambda_j, j = 1, \ldots, r$ gleich Null sein müssen.

(ii) Wir zeigen nun die Umkehrung, d. h. aus kein Vektor v_j ist eine Linearkombination der anderen Vektoren folgt $(v_1, v_2, \ldots, v_r)$ ist linear unabhängig durch Widerspruch.

Wir nehmen also an, kein Vektor v_j ist eine Linearkombination der anderen Vektoren, aber $(v_1, v_2, \ldots, v_r)$ sei dennoch linear abhängig, d. h.

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_r v_r = 0,$$

wobei mindestens ein λ_j ungleich Null ist. Dann können wir wie folgt umformen und erhalten

$$v_j = -\frac{\lambda_1}{\lambda_j} v_1 - \cdots - \frac{\lambda_{j-1}}{\lambda_j} v_{j-1} - \frac{\lambda_{j+1}}{\lambda_j} v_{j+1} - \cdots - \frac{\lambda_r}{\lambda_j} v_r,$$

was im Widerspruch zur Annahme ist, dass kein Vektor v_j eine Linearkombination der anderen Vektoren ist. □

Definition 6.7. *Es sei V ein Vektorraum über $\mathbb{K}$. Ein r-tupel $(v_1, v_2, \ldots, v_r)$ von Vektoren in V heisst* Basis*, wenn es linear unabhäng ist und für die lineare Hülle $L(v_1, v_2, \ldots, v_r) = V$ gilt.*

Ist $(v_1, v_2, \ldots, v_r)$ eine Basis, so kann man jeden Vektor auf genau eine Weise als Linearkombination dieser Vektoren schreiben. Der Vektorraum wird sozusagen von den Basisvektoren vollständig erzeugt.

Satz 6.8. *Ist $(v_1, v_2, \ldots, v_n)$ eine Basis von V, dann gibt es zu jedem Vektor $x \in V$ genau ein n-tupel von Zahlen $(x_1, \ldots, x_n) \in \mathbb{K}^n$, für welches gilt*

$$x = x_1 v_1 + \ldots + x_n v_n.$$

Die Skalare x_j heißen die Koordinaten *von x bezüglich der Basis $(v_1, v_2, \ldots, v_n)$ und die Vektoren $x_j v_j$ nennt man die* Komponenten *von x.*[1] [2]

Beweis: [**66**], Seite 59.

Wir definieren im Vektorraum $\mathbb{K}^n$ die **kanonische Basis** $(e_1, \ldots, e_n)$ (auch *Standardbasis* genannt)

$$\begin{aligned} e_1 &= (1, 0, \ldots, 0) \\ e_2 &= (0, 1, \ldots, 0) \\ &\vdots \\ e_n &= (0, 0, \ldots, 1) \end{aligned} \tag{6.4}$$

und nennt diese Vektoren oft auch Einheitsvektoren.
Anmerkung: Der Begriff Einheitsvektoren wird aber auch für Vektoren der Länge 1 verwendet. Wir haben aber bis jetzt in unseren abstrakten Vektorräumen noch keine Längen eingeführt.

Beispiel: Man veranschauliche graphisch die kanonischen Basen für $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3$.

Um den Begriff der Dimension eines Vektorraumes einführen zu können benötigt man den folgenden Satz.

Satz 6.9. Basisergänzungssatz:
Es sei V ein Vektorraum und $(v_1, v_2, \ldots, v_r, w_1, w_2, \ldots, w_s)$ Vektoren aus V. Ist $(v_1, v_2, \ldots, v_r)$ linear unabhängig und gilt $L(v_1, v_2, \ldots, v_r, w_1, w_2, \ldots, w_s) = V$, so kann man $(v_1, v_2, \ldots, v_r)$ durch eventuelle Hinzunahme geeigneter Vektoren aus $(w_1, w_2, \ldots, w_s)$ zu einer Basis ergänzen.

Beweis: [**66**], Kap. 3.4, Seite 67.

Anmerkung: In anderen Büchern wird meist der „Austauschsatz von Steinitz“ verwendet, der aber aus obigem Satz folgt.

Aus dem Basisergänzungssatz folgt auch das Austauschlemma.

Satz 6.10. *Sind $(v_1, v_2, \ldots, v_n)$ und $(w_1, w_2, \ldots, w_m)$ zwei Basen eines Vektorraum V, so gibt es zu jedem v_j ein w_k, sodass aus $(v_1, v_2, \ldots, v_n)$ wieder eine Basis entsteht, wenn man v_j durch w_k ersetzt.*

[1] Achtung: viele verwenden Koordinaten und Komponenten synonym [**62**].
[2] Wie man Vektoren von einer Basis in eine andere Basis umrechnet findet man in Kapitel 6.14.3.

Beweis: [**66**], Kap. 3.4, Seite 69.

Daraus ergibt sich nun unmittelbar

Satz 6.11. *Sind* $(v_1, v_2, \ldots, v_n)$ *und* $(w_1, w_2, \ldots, w_m)$ *zwei Basen von einem Vektorraum* V *so gilt* $n = m$.

Beweis: Annahme $n \neq m$. Dann ersetze man in der grösseren Basis alle Vektoren mit Hilfe des Austauschlemmas. Dies ergibt sofort einen Widerspruch zur linearen Unabhängigkeit der Basisvektoren in der grösseren Basis.

Da also jede Basis gleich viele Vektoren besitzt, ermöglicht dies folgende Definition

Definition 6.12. *Besitzt ein Vektorraum* V *eine Basis* $(v_1, v_2, \ldots, v_n)$*, so heisst* n *die* Dimension *von* V*. Man schreibt:* $\dim V = n$.

Die Dimension des Vektorraumes $\mathbb{K}^n$ ist klarerweise n, weil die kanonischen Basis die Länge n hat.

Satz 6.13. *Ist* $(v_1, v_2, \ldots, v_r)$ *ein* r*-tupel in* V*, wobei* $r > \dim V$ *ist, dann ist* $(v_1, v_2, \ldots, v_r)$ *linear abhängig.*

Es gibt auch Vektorräume die keine endliche Basis haben, zum Beispiel sind im Vektorraum der beschränkten Funktion auf $[0, 1]$ die Funktionen f_j

$$f_j(x) = \begin{cases} 1, & \text{für } \frac{1}{j+1} < x \leq \frac{1}{j}, \\ 0, & \text{sonst} \end{cases} \tag{6.5}$$

alle linear unabhängig. Daraus folgt, dass in diesem Vektorraum keine endliche Basis existiert.

Definition 6.14. *Besitzt ein Vektorraum* V *für kein* $n, 0 \leq n < \infty$ *eine Basis, so heisst* V *unendlichdimensionaler Vektorraum. Man schreibt:* $\dim V = \infty$.

Zum Schluss wollen wir nun noch die Dimensionen von Unterräumen betrachten.

Satz 6.15. *Es gilt:*
(i) Ist V *ein endlichdimensionaler Vektorraum und* $U \subset V$ *ein Unterraum, dann ist auch* U *endlichdimensional.*

(ii) Ist U *ein Unterraum eines endlichdimensionalen Vektorraums* V *so ist* $\dim U < \dim V$ *gleichbedeutend mit* $U \neq V$.

Um eine Formel für die Dimension des Durchschnitts zweier Vektorräume angeben zu können brauchen wir zuerst den Begriff der Summe zweier Vektorräume.

Definition 6.16. *Sind U_1, U_2 Untervektorräume von V, so heisst*

$$U_1 + U_2 = \{v_1 + v_2 \mid v_1 \in U_1, v_2 \in U_2\} \tag{6.6}$$

die Summe *von U_1 und U_2. Die Summe ist ein Untervektorraum von V.*

Damit gilt für die Dimension von Unterräumen

Satz 6.17. *Sind U_1 und U_2 endlichdimensionale Unterräume von V, so gilt*

$$\dim(U_1 \cap U_2) = \dim U_1 + \dim U_2 - \dim(U_1 + U_2). \tag{6.7}$$

Beispiele: $V = \mathbb{R}^3, U_1 = \mathbb{R}, U_2 = \mathbb{R}^2$.

Zusammenfassung

Folgende Begriffe sollten ihnen nun bekannt sein:

Untervektorraum, Linearkombination, lineare Hülle, lineare Unabhängigkeit (lineare Abhängigkeit), Basis, Dimension.

6.3. Übung

(1) Zwei Vektoren a und b im $\mathbb{R}^3$ (die nicht parallel sind) spannen eine Ebene auf, die durch den Ursprung geht. Diese Ebene ist ein zweidimensionaler Untervektorraum, der aus allen Vektoren der Form $v = \alpha a + \beta b$ besteht, wobei α und β reelle Zahlen sind. Gegeben sind die nicht parallelen Vektoren $a = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ und $b = \begin{pmatrix} 3 \\ 0 \\ -4 \end{pmatrix}$.

(a) Man zeige, dass $v = \begin{pmatrix} -3 \\ 6 \\ 5 \end{pmatrix}$ in der von a und b aufgespannten Ebene liegt. Das heisst, man finde reelle Zahlen α und β, sodass $v = \alpha a + \beta b$.

(b) Man bestimme die Koordinate s von $u = \begin{pmatrix} 5 \\ s \\ -6 \end{pmatrix}$ so, dass u im von a und b erzeugten Unterraum liegt.

(c) Man zeige, dass $c = \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix}$ nicht in der von a und b aufgespannten Ebene liegt. Die Vektoren a, b und c sind daher linear unabhängig. Sie spannen den $\mathbb{R}^3$ auf, das heisst, jeder Vektor w aus dem $\mathbb{R}^3$ lässt sich in der Form $w = \alpha a + \beta b + \gamma c$ mit reellen Zahlen α, β und γ darstellen.

(d) Man stelle den Vektor $r = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ als Linearkombination von a, b und c dar.

(2) Man betrachte den Vektorraum $V = \mathbb{R}^2$, bzw. $V = \mathbb{R}^3$. Ist die Menge A_j ein linearer Teilraum (Untervektorraum) von V?

(a) $A_1 = \{x \in V \mid 3x_1 - 2x_2 = 0\}$,

(b) $A_2 = \{x \in V \mid 3x_1 - 2x_2 = -2\}$,

(c) $A_3 = \{x \in V \mid 3x_1 - 2x_2 + x_3 = 0\}$,

(d) $A_4 = \{x \in V \mid x_1 \geq 0, x_2 \geq 0\}$,

(e) $A_5 = \{x \in V \mid \frac{1}{2}\, x_1 \leq x_2 \leq 3\, x_1$ für $x_1 \geq 0$ und
$3\, x_1 \leq x_2 \leq \frac{1}{2}\, x_1$ für $x_1 < 0\}$.

Man veranschauliche die Mengen A_j.

(3) Es sei V ein reeller Vektorraum und $a, b, c \in V$. Man zeige:
(a) Sind a, b linear unabhängig, dann folgt, dass auch die Vektoren $a + b, a - b$ linear unabhängig sind.
(b) Sind a, b, c linear unabhängig, dann folgt, dass auch die Vektoren $a + b, b + c, c + a$ linear unabhängig sind.

(4) Gegeben sei der Vektorraum $V = \mathbb{R}^3$. Man betrachte folgende Teilmenge M von Vektoren in V die in einer Ebene liegen. Die Ebene geht durch die zwei Punkte $A = (2, 1, 1), B = (1, 2, 3)$ und enthält den Vektor $v = (1, 1, 1)$.
Ist M ein linearer Teilraum von V?

(5) Es sei V ein reeller Vektorraum und $a, b, c, d \in V$. Sei

$$\begin{aligned} v_1 &= a + b + c + d \\ v_2 &= 2a + 2b + c - d \\ v_3 &= a + b + 3c - d \\ v_4 &= a - c + d \\ v_5 &= -b + c - d. \end{aligned}$$

Man beweise, dass $(v_1, \ldots, v_5)$ linear abhängig sind.
Hinweis: es gibt einen Beweis bei dem man nicht rechnen muss.

(6) Man betrachte den Vektorraum $P_2[-1, 1]$ der Polynome vom Grade $n \leq 2$ auf $[-1, 1]$.

$$P_2[-1, 1] = \{a_0 + a_1 x + a_2 x^2 \mid a_0, a_1, a_2 \in \mathbb{R}, x \in [-1, 1]\}$$

Man zeige: das 3-tupel der Vektoren $(1, x, x^2)$ ist linear unabhängig und bildet daher eine Basis. Man nennt diese Basis die kanonische Basis. Bleibt $(1, x, x^2)$ eine Basis, wenn man den Vektor x durch den Vektor $x + 1$ ersetzt oder ist $(1 + x, x + x^2, 1 + x^2)$ eine Basis?

6.4. Lineare Abbildungen

Definition 6.18. *Es seien V, W Vektorräume über $\mathbb{K}$. Eine Abbildung $f : V \to W$ heisst* linear *oder* Homomorphismus*, wenn*

$$f(v + w) = f(v) + f(w), \tag{6.8}$$

$$f(\lambda v) = \lambda f(v) \tag{6.9}$$

für alle $v, w \in V$ und $\lambda \in \mathbb{K}$ gilt.

Beispiele:

(i) $V = W = \mathbb{R}^1$, also $f : \mathbb{R} \to \mathbb{R}$. $f(x) = x^2$ ist nicht linear,

(ii) auch $f(x) = kx + d, \quad k, d \in \mathbb{R}$ ist nicht linear, ausser $d = 0$. Hinweis: Ist $d \neq 0$, so nennt man die Abbildung $f(x) = kx + d$ daher **affine** Abbildung.

(iii) $V = W = \mathbb{R}^3$, $f((x_1, x_2, x_3)) = (x_1 + x_2 + x_3, x_1 + x_3, x_1 + x_2)$.

Anmerkung: Man mache sich bewusst, dass dies eine Funktion in 3 Variablen ist, die 3 Komponenten hat.

(iv) Man interpretiere graphisch die Eigenschaften der Linearität im $\mathbb{R}^2$.

Die Menge aller linearen Abbildungen von V nach W wird mit $\mathrm{Hom}(V, W)$ bezeichnet. Diese Menge bildet mit den entsprechend definierten Verknüpfungen (Addition $f + g$ und Multiplikation mit Skalaren λf) wieder einen Vektorraum. Ist $W = \mathbb{K}$ so nennt man diese Menge den **Dualraum** von V.

Definition 6.19. *Es sei $f : V \to W$ eine lineare Abbildung. Dann ist* Bild $f = f(V)$ *ein Untervektorraum von W und* Kern $f := \{v \in V \mid f(v) = 0\}$ *ein Untervektorraum von V.*

Der Kern einer linearen Abbildung besteht aus allen Vektoren, die in den Nullvektor abgebildet werden und wird daher auch noch *Nullraum* genannt. Die Dimension dieses Raumes wird auch **Defekt** genannt.

Beispiel: $V = W = \mathbb{R}^3$ und $f((x, y, z)) = (x, y, 0)$. Dann ist das Bild dieser Abbildung f ein zweidimensionaler Unterraum von $W = \mathbb{R}^3$ und der Kern dieser Abbildung f ist ein eindimensionaler Unterraum von $V = \mathbb{R}^3$. Eine Basis für den Kern bildet der Vektor $(0, 0, 1)$.

Satz 6.20. *Eine lineare Abbildung f ist genau dann injektiv, wenn* Kern $f = \{\vec{0}\}$ *gilt.*

Beweis: Dies folgt unmittelbar aus $f(x) = f(y)$, da wegen der Linearität von f diese Bedingung gleichwertig zu $x - y \in \mathrm{Kern}\, f$ ist.

Definition 6.21. *Eine lineare Abbildung $f : V \to W$ heisst ein* Isomorphismus, *wenn sie bijektiv ist. Ein Isomorphismus heisst* Automorphismus, *wenn $V = W$.*

Satz 6.22. *Ist $f : V \to W$ ein Isomorphismus, so ist auch die Umkehrabbildung ein Isomorphismus.*

Beweis: Es ist zwar klar, daß die Umkehrabbildung bijektiv ist, aber man muss auch nachweisen, dass sie eine lineare Abbildung ist.

Es gilt

$$v_1 + v_2 = f^{-1}(f(v_1 + v_2)) = f^{-1}(f(v_1) + f(v_2)),$$
$$\lambda v = f^{-1}(f(\lambda v)) = f^{-1}(\lambda f(v)).$$

Da f bijektiv ist gibt es zu jedem $v \in V$ ein $w \in W$ und umgekehrt. Setzt man daher $v_1 = f^{-1}(w_1), v_2 = f^{-1}(w_2)$ und $v = f^{-1}(w)$ erhält man daher

$$f^{-1}(w_1) + f^{-1}(w_2) = f^{-1}(w_1 + w_2),$$
$$\lambda f^{-1}(w) = f^{-1}(\lambda w),$$

was die Linearität der inversen Abbildung f^{-1} zeigt.

Beispiel: $V = \mathbb{R}^3$ und W ist der Vektorraum P_2 der Polynome vom Grade ≤ 2.

$$P_2 = \{a_0 + a_1 x + a_2 x^2 \mid a_0, a_1, a_2 \in \mathbb{R}\}.$$

Ein Isomorphismus $f : \mathbb{R}^3 \to P_2$ ist gegeben durch

$$(b_1, b_2, b_3) \mapsto b_1 + b_2 x + b_3 x^2.$$

Satz 6.23. *Es seien V, W Vektorräume über $\mathbb{K}$ und $(v_1, \ldots, v_n)$ eine Basis von V. Dann gibt es zu jedem n-tupel $(w_1, \ldots, w_n)$ von (beliebigen) Vektoren in W genau eine lineare Abbildung $f : V \to W$ mit*

$$f(v_j) = w_j, \qquad j = 1, \ldots, n. \tag{6.10}$$

Beweis: Man beweist dabei zuerst die Eindeutigkeit und dann die Existenz dieser Abbildung. Der Beweis der Existenz ist konstruktiv und gibt ein Verfahren an um diese Abbildung zu berechnen.

(i) Eindeutigkeit: Angenommen es gibt zwei lineare Abbildungen f und f' mit der Eigenschaft $f(v_j) = w_j = f'(v_j)$. Dann gilt für jeden beliebigen Vektor v, da $(v_1, \ldots, v_n)$ eine Basis ist

$$f(v) = f(\lambda_1 v_1 + \ldots \lambda_n v_n) = \lambda_1 f(v_1) + \ldots \lambda_n f(v_n)$$
$$= \lambda_1 f'(v_1) + \ldots \lambda_n f'(v_n) = f'(\lambda_1 v_1 + \ldots \lambda_n v_n) = f'(v).$$

(ii) Es bleibt die Existenz einer solchen Abbildung zu zeigen. Dies macht man indem man explizit eine Abbildung hinschreibt. Sei $v \in$ beliebig, dann gilt $v = \lambda_1 v_1 + \ldots \lambda_n v_n$. Man definiert nun f wie folgt

$$f(v) := \lambda_1 w_1 + \ldots \lambda_n w_n. \tag{6.11}$$

Es ist nun trivial zu zeigen, dass diese Abbildung f linear ist und $f(v_j) = w_j$ erfüllt. □

Es zeigt sich, dass die **gesamte Information** über eine lineare Abbildung bereits durch die Wirkung der Abbildung (Bildern) einer Basis gegeben ist.

Satz 6.24. *Es seien V, W Vektorräume über $\mathbb{K}$ und $(v_1, \ldots, v_n)$ eine Basis von V. Eine lineare Abbildung $f : V \to W$ ist genau dann ein Isomorphismus, wenn $(f(v_1), \ldots, f(v_n))$ eine Basis von W ist.*

Beweis: [**66**], Kap. 4.1, Seite 86.

Je zwei n-dimensionale Vektorräume über $\mathbb{K}$ sind isomorph, d. h. es existiert ein Isomorphismus.

Definition 6.25. *Es sei $f : V \to W$ eine lineare Abbildung. Ist* Bild f *endlichdimensional, dann heisst* Rang $f :=$ dim Bild f *der* Rang *von f.*

Satz 6.26. *Sei V ein n-dimensionaler Vektorraum über $\mathbb{K}$ und $f : V \to W$ eine lineare Abbildung. Dann gilt:*

$$\dim \text{Kern}\, f + \dim \text{Bild}\, f = n.$$

Beweis: [**66**], Kap. 4.1, Seite 86.

Daraus folgt unmittelbar

Satz 6.27. *Eine lineare Abbildung zwischen zwei Räumen der gleichen Dimension n ist genau dann surjektiv, wenn sie injektiv ist.*

6.5. Matrizen

Wir wollen nun den Begriff einer Matrix definieren.

Definition 6.28. *Eine $m \times n$-*Matrix *über $\mathbb{K}$ ist eine Anordnung von $m\,n$ Elementen von $\mathbb{K}$ nach folgendem Schema*

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}. \tag{6.12}$$

Die Zahlen a_{jk} nennt man die Koeffizienten *der Matrix. Die waagrecht geschrieben n-tupel $z_j(A) := (a_{j1} \cdots a_{jn})$ heissen die* Zeilen *der Matrix und die senkrecht geschrieben m-tupel $s_k(A) := \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix}$ heissen die* Spalten *der Matrix.*

Die Menge aller $m \times n$-Matrizen über $\mathbb{K}$ wird mit $M(m \times n, \mathbb{K})$ bezeichnet.

In a_{jk} bezeichnet also j die Zeile und k die Spalte. Die Koeffizienten der Form a_{jj} bilden die *Hauptdiagonale* der Matrix.

Als nächstes definieren wir wie eine Matrix auf einen Vektor aus dem Raum $\mathbb{K}^n$ wirken soll.

Definition 6.29. *Es sei A eine $m \times n$-Matrix und x ein Vektor $x = (x_1, x_2, \ldots, x_n) \in \mathbb{K}^n$. Dann wird das* Produkt *$Ax \in \mathbb{K}^m$ folgendermassen definiert.*

$$Ax = \left(\sum_{j=1}^{n} a_{1j}x_j, \sum_{j=1}^{n} a_{2j}x_j, \ldots, \sum_{j=1}^{n} a_{mj}x_j\right) \tag{6.13}$$

Anmerkung: Führt man das formale Produkt von zwei Vektoren x, y wie folgt ein

$$x \cdot y = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} := x_1y_1 + \ldots + x_ny_n \tag{6.14}$$

kann man das Produkt der Matrix mit einem Vektor folgendermaßen schreiben

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} z_1(A) \cdot x \\ \vdots \\ z_m(A) \cdot x \end{pmatrix}. \tag{6.15}$$

wobei Vektoren aus dem $\mathbb{K}^n$ und $\mathbb{K}^m$ als Spalten geschrieben werden. Ein Vektor aus dem $\mathbb{K}^s$ ist sozusagen eine $s \times 1$-Matrix. Eine $m \times n$-Matrix wirkt auf eine Vektor $x \in \mathbb{K}^n$ und erzeugt daraus einen Vektor $y \in \mathbb{K}^m$.

Merkregel: Man kann die j-te Spalte auf der rechten Seite von (6.15) als formales (skalares, wenn $\mathbb{K} = \mathbb{R}$) Produkt des j-ten Zeilenvektors der Matrix A und des Vektors x interpretieren.

Beispiele: Man berechne $Ae_j, j = 1, 2, 3$ und Av, wobei

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, v = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}.$$

Der folgende Satz zeigt nun den Zusammenhang zwischen Matrizen und linearen Abbildungen von $\mathbb{K}^n \to \mathbb{K}^m$.

Satz 6.30. *Es sei $A \in M(m \times n, \mathbb{K})$. Dann ist die Abbildung*

$$A : \mathbb{K}^n \to \mathbb{K}^m.$$
$$x \mapsto Ax \tag{6.16}$$

linear, und umgekehrt ist $f : \mathbb{K}^n \mapsto \mathbb{K}^m$ eine lineare Abbildung, dann gibt es eine Matrix $A \in M(m \times n, \mathbb{K})$ mit $f(x) = Ax$ für alle $x \in \mathbb{K}^n$.

Beweis: Dass $A(x + y) = A(x) + A(y)$ und $A(\lambda x) = \lambda(Ax)$ erfüllt ist, folgt unmittelbar aus der Definition 6.13 dieser Abbildung.

Es bleibt also die Umkehrung zu zeigen, d. h. man muss zeigen: zu jeder linearen Abbildung f gibt es eine Matrix A, sodass gilt $Ax = f(x)$ für alle $x \in \mathbb{K}^n$.

(i) Eindeutigkeit

Es seien A und B zwei Matrizen mit $f(x) = Ax = Bx$. Man betrachtet zuerst die Wirkung der Abbildung f auf die Einheitsvektoren e_j. Dann gilt auch $Ae_j = Be_j$. Da aber Ae_j genau die j-Spalte von A liefert und auch das gleiche für Be_j gilt, müssen beide Spalten identisch sein und da dies für jede Spalte gilt, müssen daher alle Koeffizienten von A mit denen von B übereinstimmen.

(ii) Existenz

Man betrachtet nun die Abbildung der Einheitsvektoren $f(e_j) := v_j$

$$f(e_j) := v_j = \begin{pmatrix} v_{1j} \\ \vdots \\ v_{mj} \end{pmatrix}$$

und nimmt diese Bildvektoren zur Definition der Spalten der Matrix A

$$A := \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{m1} & \cdots & v_{mn} \end{pmatrix}.$$

Dann gilt auf den Einheitsvektoren $Ae_j = f(e_j)$. Durch die Linearität setzt sich dies für beliebige Vektoren fort.

$$Ax = A(\lambda_1 e_1 + \cdots \lambda_n e_n) = f(\lambda_1 e_1 + \cdots \lambda_n e_n) = f(x). \quad \square$$

Merkregel: Die Bilder der "Einheitsvektoren" sind die Spalten von A, oder anders ausgedrückt: der Bildraum ist die lineare Hülle der Spaltenvektoren der Matrix A.

Um diesen Satz ganz allgemein für beliebige Vektorräume benutzbar zu machen definieren wir

Definition 6.31. *Ist V ein Vektorraum über $\mathbb{K}$, und ist $(v_1, \ldots, v_n)$ eine Basis, so nennt man die Abbildung $\Phi_{(v_1,\ldots,v_n)}$*

$$\begin{aligned} \Phi_{(v_1,\ldots,v_n)} : \quad & \mathbb{K}^n \to V, \\ & (\lambda_1, \ldots, \lambda_n) \mapsto \lambda_1 v_1 + \ldots + \lambda_n v_n \end{aligned} \tag{6.17}$$

den kanonischen Basisisomorphismus.

Wegen Satz 6.23 und Satz 6.24 ist der kanonische Basisisomorphismus genau der Isomorphismus der die Einheitsvektoren in $\mathbb{K}^n$ eindeutig auf die Basis $(v_1, \ldots, v_n)$ abbildet.

Damit kann man nun jeder linearen Abbildung zwischen zwei beliebigen (endlichdimensionalen) Vektorräumen eine Matrix zuordnen.

Definition 6.32. *Es sei $f : V \to W$ eine lineare Abbildung zwischen Vektorräumen über $\mathbb{K}$, und $(v_1, \ldots, v_n)$ und $(w_1, \ldots, w_m)$ seien Basen für V bzw. W. Dann heisst die durch $A = \Phi^{-1}_{(w_1,\ldots,w_m)} \circ f \circ \Phi_{(v_1,\ldots,v_n)}$ bestimmte Matrix $A \in M(m \times n, \mathbb{K})$ die zu f bezüglich der beiden gewählten Basen gehörige Matrix.*

Umgekehrt ergibt sich $f = \Phi_{(w_1,\ldots,w_m)} \circ A \circ \Phi^{-1}_{(v_1,\ldots,v_n)}$. Man veranschaulicht dies in einem sogenannten kommutativen Diagramm

$$\begin{array}{ccc} \mathbb{K}^n & \xrightarrow{A} & \mathbb{K}^m \\ {\scriptstyle \Phi_{(v_1,\ldots,v_n)}}\downarrow & & \downarrow{\scriptstyle \Phi_{(w_1,\ldots,w_m)}} \\ V & \xrightarrow{f} & W \end{array} \tag{6.18}$$

Anmerkung: Für eine feste Abbildung f existieren viele Möglichkeiten einer Matrixdarstellung, da jede Änderung der Basis zu einer anderen Matrix führt.

Beispiel: In einem Vektorraum von differenzierbaren Funktionen ist das *Ableiten* eine lineare Abbildung, denn es gilt für zwei Funktionen f, g

(i) $(f + g)' = f' + g'$ und

(ii) $(\lambda f)' = \lambda f'$.

Zusammenfassung

Folgende Begriffe sollten ihnen nun bekannt sein:

Lineare Abbildung, Kern, Bild, Rang, Defekt, Isomorphismus,
Matrix,
Satz 6.23 (die gesamte Information über eine lineare Abbildung steckt in den Bildern einer Basis) und
Satz 6.30 (die Bilder der Einheitsvektoren sind die Spalten von A.)

6.6. Übung

(1) Es sei $V = W = \mathbb{R}^3$ und die Abbildungen $f, g : \mathbb{R}^3 \to \mathbb{R}^3$,
(a) $f((x_1, x_2, x_3)) = (x_1, x_2, 0)$,
(b) $g((x_1, x_2, x_3)) = (x_1 + 2x_2, x_2 + x_3, x_3 - x_1 - 3x_2)$.
Man zeige, dass f und g linear sind und berechne die dazugehörige Matrix A. Ist f oder g ein Isomorphismus?

(2) Man betrachte den Vektorraum P_3 der Polynome vom Grade $n \leq 3$ auf dem Intervall $[-1, 1]$.

$$P_3 = \{a_0 + a_1 x + a_2 x^2 + a_3 x^3 \mid a_0, a_1, a_2, a_3 \in \mathbb{R}\}$$

Man zeige, dass die Abbildung $f : P_3 \to P_3$

$$f(v) = \frac{dv}{dx}$$

linear ist (es gilt die Ableitungsregel: $(a\,x^m + b\,x^n)' = a\,m\,x^{m-1} + b\,n\,x^{n-1}, a, b \in \mathbb{R}, m, n \in \mathbb{N} \setminus \{0\}$). Was ist der Kern dieser Abbildung?

(3) Man berechne den Kern der Abbildung $A : \mathbb{R}^3 \to \mathbb{R}^3$ und $B : \mathbb{R}^4 \to \mathbb{R}^4$.

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 & 1 & 3 \\ 1 & 2 & 1 & 4 \\ 1 & 2 & 2 & 7 \\ 2 & 4 & 1 & 5 \end{pmatrix}.$$

(4) Es sei $V = W = \mathbb{R}^2$ und $v_1 = (3, 1), v_2 = (1, 4)$ eine Basis von V.
(a) Die lineare Abbildung $f : V \to W$ sei gegeben durch $f(v_1) = (2, 2), f(v_2) = (1, 1)$.
Man berechne den Kern von f und die der linearen Abbildung f zugeordnete Matrix A bezüglich der Basis e_1, e_2.[3]
(b) Die lineare Abbildung $f : V \to W$ sei gegeben durch $f(v_1) = (2, 2), f(v_2) = (1, 0)$.
Man berechne den Kern von f und die der linearen Abbildung f zugeordnete Matrix A bezüglich der Basis e_1, e_2.

(5) Es sei P_n der Vektorraum der Polynome vom Grade $\leq n$ auf $[-1, 1]$. Man bestimme die Matrix der linearen Abbildung
(a) $f : P_3 \to P_4$, $p(x) \mapsto (2 - x)p(x)$ bezüglich der kanonischen Basen,
(b) $g : P_3 \to P_2$, $p(x) \mapsto \dfrac{dp(x)}{dx}$ bezüglich der kanonischen Basen,
(c) $h : P_3 \to P_3$, $p(x) \mapsto \dfrac{d\big((x + 2)p(x)\big)}{dx}$ bezüglich der kanonischen Basen.

[3] Basistransformationen werden in Kapitel6.14.3 beschrieben.

Ergänzung:

6.6.1. (*) Quotientenräume. Ist $U \subset V$ ein Untervektorraum eines Vektorraumes V, so kann man auf V eine Äquivalenzrelation durch

$$x\,R\,y \quad \text{wenn} \quad x - y \in U \tag{6.19}$$

erklären. Die Äquivalenzklassen heissen Nebenklassen (oder auch affine Teilräume von V) und haben die Form

$$x + U = [x] = \{x + u \mid u \in U\}. \tag{6.20}$$

Die Nebenklassen sind keine Untervektorräume.

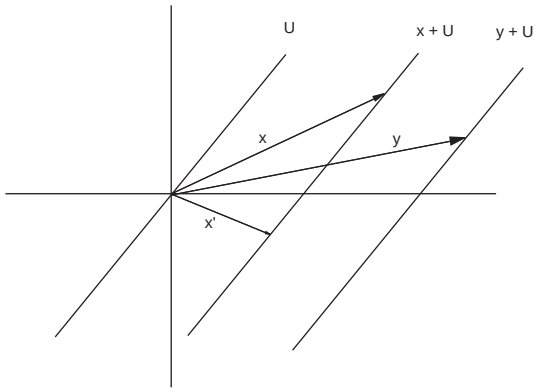


Abbildung 6.1. affine Teilräume

Die Menge aller Nebenklassen kann man mit folgenden Verknüpfungen zu einem Vektorraum machen

$$\begin{aligned}(x + U) + (y + U) &:= (x + y) + U,\\ \lambda(x + U) &:= \lambda x + U,\end{aligned} \tag{6.21}$$

den man als **Quotientenraum** V/U bezeichnet; $V/U = \{x + U \mid x \in V\}$.

Es gilt: $\dim(V/U) = \dim V - \dim U$.

Die Abbildung $\pi : V \to V/U, v \mapsto v + U$ mit $\operatorname{Kern}\pi = U$ nennt man die (kanonische) *Projektion* der Quotientenbildung.

Diese Konstruktion ermöglicht es auf einfache Weise neue Vektorräume zu konstruieren.

Mehr über Quotientenräume findet man z. B. in [**64**], Seite 107ff.

6.7. Matrizenrechnung

Statt

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \tag{6.22}$$

schreiben wir auch $A = (a_{jk})_{j,k=1}^{m,n}$ oder kurz $A = (a_{jk})$.

Für die Addition von Matrizen und die Multiplikation mit Skalaren ist es sinnvoll zu verlangen, dass für die dadurch definierten Abbildungen gilt:
$(A + B)x = Ax + Bx$ und $(\lambda A)x = \lambda(Ax)$.

Daher erfolgt die Addition von Matrizen und die Multiplikation mit Skalaren koeffizientenweise, wie bei r-tupeln.

Definition 6.33. *Sind* $(a_{jk}), (b_{jk}) \in M(m \times n, \mathbb{K})$ *und* $\lambda \in \mathbb{K}$*, so ist die* Addition *und* Skalarmultiplikation *folgenderweise definiert*

$$\begin{aligned} (a_{jk}) + (b_{jk}) &:= (a_{jk} + b_{jk}) \in M(m \times n, \mathbb{K}), \\ \lambda(a_{jk}) &:= (\lambda a_{jk}) \in M(m \times n, \mathbb{K}). \end{aligned} \tag{6.23}$$

Die Menge der Matrizen $M(m \times n, \mathbb{K})$ bildet mit diesen zwei Verknüpfungen einen Vektorraum über $\mathbb{K}$ der Dimension $m \cdot n$. Es ist üblich eine Matrix $A \in M(m \times n, \mathbb{K})$ und die zugehörige lineare Abbildung $\mathbb{K}^n \to \mathbb{K}^m$ mit demselben Buchstaben zu bezeichen, also $A : \mathbb{K}^n \to \mathbb{K}^m$.

Die Abbildung $M(m \times n, \mathbb{K}) \to \mathrm{Hom}(\mathbb{K}^n, \mathbb{K}^m)$, die dadurch definiert ist, dass man jeder Matrix A die lineare Abbildung $\mathbb{K}^n \to \mathbb{K}^m, x \mapsto Ax$ zuordnet, ist ein Isomorphismus der Vektorräume.

Um nun die Multiplikation von zwei Matrizen zu definieren machen wir folgende Überlegung. Die Verknüpfung von zwei Abbildungen ist wohldefiniert. Da Matrizen lineare Abbildungen (zugeordnet) sind, werden wir die Produktmatrix genau so definieren, dass sie der Matrix der Verknüpfung von zwei linearen Abbildungen entspricht. Also so, dass $(AB)x = A(Bx)$ gilt

$$A : \mathbb{K}^m \to \mathbb{K}^r, B : \mathbb{K}^n \to \mathbb{K}^s, \quad AB : \mathbb{K}^n \to \mathbb{K}^r \tag{6.24}$$

wobei $s = m$ sein muss.

Um nun eine Formel für AB zu bestimmen, muss man einfach nur das Bild des k-ten Einheitsvektors berechnen

$$e_k \mapsto Be_k \mapsto A(Be_k). \tag{6.25}$$

Dies ergibt die k-te Spalte von AB, also

$$\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} b_{1k} \\ \vdots \\ \vdots \\ \vdots \\ b_{mk} \end{pmatrix} \mapsto \begin{pmatrix} a_{11}b_{1k} + \cdots + a_{1m}b_{mk} \\ \vdots \\ \vdots \\ \vdots \\ a_{r1}b_{1k} + \cdots + a_{rm}b_{mk} \end{pmatrix}. \tag{6.26}$$

Das j-te Element der k-ten Spalte von AB ist daher gleich $\sum_{\ell=1}^{m} a_{j\ell}b_{\ell k}$.

Definition 6.34. *Sind* $(a_{jk}) \in M(r \times m, \mathbb{K})$ *und* $(b_{jk}) \in M(m \times n, \mathbb{K})$*, so ist das* (Matrix)-Produkt $AB \in M(r \times n, \mathbb{K})$ *folgenderweise definiert*

$$AB := \left(\sum_{\ell=1}^{m} a_{j\ell}b_{\ell k} \right)_{j,k=1}^{r,n}. \tag{6.27}$$

Wir haben damit das Matrizenprodukt so definiert, dass es genau der Zusammensetzung der zugehörigen linearen Abbildungen entspricht.

$$\begin{array}{ccc} \mathbb{K}^n & \xrightarrow{B} & \mathbb{K}^m \\ & & \downarrow{\scriptstyle A} \\ & & \mathbb{K}^r. \end{array} \tag{6.28}$$

Dasselbe gilt auch für beliebige lineare Abbildungen f, g, die mittels des kanonischen Basisisomorphismus durch Matrizen dargestellt werden, d. h., dass die Matrix AB genau der linearen Abbildung $g \circ f$ entspricht

$$\begin{array}{ccccc} \mathbb{K}^n & \xrightarrow{B} & \mathbb{K}^m & \xrightarrow{A} & \mathbb{K}^r \\ \downarrow{\scriptstyle \Phi_{(v_1,\ldots,v_n)}} & & \downarrow{\scriptstyle \Phi_{(w_1,\ldots,w_m)}} & & \downarrow{\scriptstyle \Phi_{(y_1,\ldots,y_r)}} \\ V & \xrightarrow{f} & W & \xrightarrow{g} & Y \end{array} \tag{6.29}$$

Merkregel zur Matrizenmultiplikation: Das Element $(AB)_{jk}$ erhält man, wenn man den j-ten Zeilenvektor von A mit dem k-ten Spaltenvektor von B formal (skalar, wenn $\mathbb{K} = \mathbb{R}$) multipliziert; $(AB)_{jk} = z_j(A) \cdot s_k(B)$.

Beispiel: Berechne AB und BA, wobei

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix}.$$

Es gilt für die Matrizenmultiplikation der folgende Satz

Satz 6.35. *Die Matrizenmultiplikation ist*

(i) assoziativ: $(AB)C = A(BC)$,

(ii) distributiv: $A(B + C) = AB + AC$,

(iii) nicht kommutativ: $AB \neq BA$,

(iv) nicht nullteilerfrei: aus $AB = 0$ *folgt nicht* $A = 0$ *oder* $B = 0$.

Beweis: Sei

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

dann ist

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \neq BA = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}.$$

Definition 6.36. *Die Matrix der identischen Abbildung*

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \tag{6.30}$$

wird mit $E_n \in M(n \times n, \mathbb{K})$, *bzw. kurz mit* E *bezeichnet, d. h.* $E\,x = x$.

Definition 6.37. *Eine Matrix heisst* invertierbar, *wenn die zugehörige lineare Abbildung ein Isomorphismus ist. Die Matrix der Umkehrabbildung heisst dann die zu* A inverse Matrix *und wird mit* A^{-1} *bezeichnet.*

Für die inverse Matrix gilt

Satz 6.38. *(i) Jede invertierbare Matrix ist quadratisch, d. h.* $A \in M(n \times n, \mathbb{K})$.

(ii) Sind $A, B, E \in M(n \times n, \mathbb{K})$, *so ist* B *genau dann zu* A *die inverse Abbildung, wenn sowohl* AB *als auch* BA *gleich* E *ist.*

(iii) Seien $A, B, E \in M(n \times n, \mathbb{K})$. *Folgende Bedingungen sind gleichwertig*
(a) $AB = E$, *(b)* $BA = E$ *und (c)* $B = A^{-1}$.

(iv) Ist A *invertierbar, so ist auch* A^{-1} *invertierbar und es gilt* $(A^{-1})^{-1} = A$.

(v) Sind $A, B \in M(n \times n, \mathbb{K})$ *invertierbar, so ist auch* AB *invertierbar, wobei gilt:* $(AB)^{-1} = B^{-1}A^{-1}$.

6.7.1. Transponierte Matrix.

Man erhält die transponierte Matrix A^T aus A durch Vertauschen der Zeilen mit den Spalten.

Definition 6.39. *Ist $A \in M(m \times n, \mathbb{K})$, so heisst die durch*

$$a^T_{jk} := a_{kj} \tag{6.31}$$

definierte Matrix $A^T = (a_{jk})^T \in M(n \times m, \mathbb{K})$ die transponierte Matrix.

Achtung: Die transponierte Matrix bildet in die umgekehrte Richtung ab, d. h. $A : \mathbb{K}^n \to \mathbb{K}^m$ und $A^T : \mathbb{K}^m \to \mathbb{K}^n$

Satz 6.40. *Es gilt*

$(AB)^T = B^T A^T, \quad (A + B)^T = A^T + B^T.$

Anmerkung: Die transponierte Matrix ist für reelle Matrizen identisch mit der adjungierten Matrix (siehe später Kapitel 6.14.4) die mit Hilfe des skalaren Produktes definiert ist.[4]

6.7.2. Rang einer Matrix.

Definition 6.41. *Ist $A \in M(m \times n, \mathbb{K})$ so nennt man*

$$\operatorname{Rang} A := \dim \operatorname{Bild} A, \qquad (A : \mathbb{K}^n \to \mathbb{K}^m) \tag{6.32}$$

den Rang *von A (bzw.* rank*).*
Die Maximalzahl linear unabhängiger Spalten nennt man den Spaltenrang *von A und die Maximalzahl linear unabhängiger Zeilen den* Zeilenrang.

Der Bildraum der linearen Abbildung A wird von den Spaltenvektoren der Matrix A aufgespannt!

Satz 6.42. *Es gilt:* $\operatorname{Rang} A^T = \operatorname{Rang} A$ *und daher Spaltenrang = Zeilenrang.*

6.7.3. Elementare Umformungen.

Definition 6.43. *Man unterscheidet 3 Typen von* elementaren Zeilenumformungen *einer Matrix $A \in M(m \times n, \mathbb{K})$*

Typ 1: *Vertauschung zweier Zeilen,*

Typ 2: *Multiplikation einer Zeile mit einem Skalar $\lambda \neq 0, \lambda \in \mathbb{K}$,*

Typ 3: *Addition eines beliebigen Vielfachen einer Zeile zu einer anderen (nicht derselben!) Zeile.*

[4] Die transponierte Matrix erfüllt folgende Bedingung: $\langle A^T x, y\rangle = \langle x, Ay\rangle$ und die Spalten von A^T ermittelt man mit Hilfe der Bildvektoren der Einheitsvektoren.

Analog sind *elementare Spaltenumformungen* definiert.[5]

Satz 6.44. *Elementare Spaltenumformungen (Zeilenumformungen) ändern den Rang einer Matrix A nicht.*

Beweis: Die Spalten einer Matrix A sind die Bilder der Einheitsvektoren e_j und spannen daher den Bildraum der Matrix A auf. Elementare Spaltenumformungen verändern die lineare Hülle der Spalten (Bildraum) nicht und lassen daher den Rang unverändert.

Anmerkung: Die Zeilen der Matrix A spannen den Bildraum der transponierten Matrix A^T auf.

Definition 6.45. *Die Elemente a_{jj} in einer Matrix heissen die* Hauptdiagonalelemente. *Man sagt das Element a_{jk} steht oberhalb der Hauptdiagonale, wenn $j < k$ ist und a_{jk} steht unterhalb der Hauptdiagonale, wenn $j > k$ ist.*

Beispiel: a_{23} steht oberhalb und a_{32} steht unterhalb.

Satz 6.46. *Ist A eine Matrix mit m Zeilen, bei der die ersten r Hauptdiagonalelemente von Null verschieden, die letzten $m-r$ Zeilen, sowie alle Elemente unterhalb der Hauptdiagonale jedoch gleich Null sind, so ist*

$$\operatorname{Rang} A = r. \tag{6.33}$$

Die Matrix A hat daher die folgende Form

$$\left(\begin{array}{cccc} a_{11} & & & \\ 0 & \ddots & & * \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 0 & a_{rr} \\ \hline & & & 0 \end{array}\right) \tag{6.34}$$

wobei $*$ beliebige Elemente darstellt.

Anmerkung: Es ist offensichtlich, dass bei dieser Form der Matrix die ersten r Spalten linear unabhängig sind (und auch die ersten r Zeilen, Zeilenrang = Spaltenrang).

Verfahren zur Rangbestimmung einer Matrix

[5]Hinweis: Die Spalten einer Matrix spannen den Bildraum auf, andererseits verwenden wir im Gauss-Algorithmus immer Zeilenumformungen. Für die Berechnung des Rangs oder der Determinate macht das aber keinen Unterschied.

Sei $A \in M(m \times n, \mathbb{K})$ bereits in der Gestalt

$$\left(\begin{array}{cccc} a_{11} & & & \\ & \ddots & & * \\ & 0 & \ddots & \\ & & & a_{\ell-1,\ell-1} \\ \hline 0 & & & B \end{array}\right) \tag{6.35}$$

wobei $a_{jj} \neq 0, j = 1, \ldots, \ell - 1$ gilt und B eine $(m - \ell + 1) \times (m - \ell + 1)$-Matrix ist. Ist $B = 0$ so ist Rang $A = \ell - 1$. Ist $B \neq 0$ so gibt es ein $a_{jk} \neq 0$ mit $j > \ell$ und $k > \ell$. Vertauscht man in A nötigenfalls die j-te und ℓ-te Zeile und dann die k-te und ℓ-te Spalte, so erhält man eine Matrix A' mit $a'_{\ell\ell} \neq 0$, die durch elementare Zeilenumformungen auf folgende Gestalt gebracht werden kann.

$$\left(\begin{array}{cccc|cc} a_{11} & & & & & \\ & \ddots & & * & & \\ & 0 & \ddots & & & \\ & & & a_{\ell-1,\ell-1} & & \\ \hline & & & & a'_{\ell\ell} & * \\ & 0 & & & & \\ & & & & 0 & B' \end{array}\right). \tag{6.36}$$

Beginnt man dieses Verfahren bei $\ell = 2$ und setzt es solange fort bis $B' = 0$ ist, so erhält man eine Matrix in der Gestalt auf die Satz (6.46) anwendbar ist.

6.7.4. Matrizeninvertierung.

Satz 6.47. *Gilt für drei Matrizen $A, B, C \in M(n \times n, \mathbb{K})$ die Gleichung $AB = C$ und überführt man A und C durch die gleichen elementaren Zeilenumformungen in die Matrizen A' und C', so gilt auch $A'B = C'$.*

Hinweis: elementare Zeilenumformungen können durch ein Matrix U gemacht werden. Es gilt dann $U(AB) = UC$ und mit $A' = UA$ und $UC = C'$ folgt dann $A'B = C'$.

Da nun $AA^{-1} = E$ gilt, schliesst man daraus: Erhält man E durch elementare Zeilenumformungen aus A, so verwandeln dieselben Zeilenumformungen die Matrix E in die inverse Matrix A^{-1}.

Verfahren um eine Matrix durch elementare Zeilenumformungen auf die Einheitsmatrix zu bringen.

Es sei A eine $n \times n$-Matrix. Zuerst versucht man durch Vertauschung von Zeilen zu erreichen, dass der erste Koeffizient ungleich Null wird. Ist dies nicht möglich

so ist die erste Spalte Null und die Matrix ist nicht invertierbar. Sei also $a_{11} \neq 0$. Dann wird a_{11} durch Multiplikation der ersten Zeile mit $\frac{1}{a_{11}}$ zu 1. Durch Addition geeigneter Vielfacher der ersten Zeile zu den anderen Zeilen kann man den ersten Koeffizienten der anderen Zeilen zu Null machen, also auf die Form

$$\left(\begin{array}{c|c} 1 & \\ 0 & \\ \vdots & * \\ 0 & \end{array}\right)$$

bringen. Damit ist der erste Schritt beendet. Als nächstes erreicht man durch Zeilenvertauschungen, dass in der zweiten Zeile a_{22} ungleich Null wird. Dann wird a_{22} durch Multiplikation der zweiten Zeile mit $\frac{1}{a_{22}}$ zu 1. Durch Addition geeigneter Vielfacher der zweiten Zeile zu den anderen Zeilen kann man die zweiten Koeffizienten der anderen Zeilen zu Null machen. Die Matrix hat nun die Form

$$\left(\begin{array}{c|c|c} 1 & 0 & \\ 0 & 1 & \\ & 0 & \\ \vdots & \vdots & * \\ 0 & 0 & \end{array}\right).$$

Nach n Schritten führt dieses Verfahren die Matrix A in die Einheitsmatrix E über, ausser A ist nicht invertierbar.

Beispiel: Es sei

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 2 \end{pmatrix}.$$

Man berechne die inverse Matrix nach obigen Verfahren.

Anmerkung: Für 2×2-Matrizen kann man die inverse Matrix auch direkt durch Ansatz ausrechnen, z. B.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Dies führt allerdings auf n^2 Gleichungen, was für $n > 2$ umständlich wird.

6.7.5. Drehungen und Spiegelungen im $\mathbb{R}^2$. Im Vektorraum $\mathbb{R}^2$ mit Skalarprodukt suchen wir die linearen Abbildungen A, die das skalare Produkt unverändert lassen, das sind die Matrizen A für die gilt

$$\langle Ax, Ay\rangle \;=\; \langle x, y\rangle. \tag{6.37}$$

Betrachten wir die Wirkung einer solchen Matrix auf die Einheitsvektoren $e_1 = (1,0),\ e_2 = (0,1)$, so ergibt sich

$$\langle Ae_j, Ae_j\rangle \;=\; ||Ae_j||^2 \;=\; \langle e_j, e_j\rangle \;= 1, \quad j = 1,2. \tag{6.38}$$

Das heisst, dass die Matrix A die Länge der Einheitsvektoren unverändert lässt und die Bildvektoren daher auf dem Einheitskreis liegen. Bezeichnet man den Winkel zwischen x-Achse und Ae_1 mit φ, so kann man Ae_1 als

$$Ae_1 = \begin{pmatrix}\cos\varphi\\ \sin\varphi\end{pmatrix} \tag{6.39}$$

schreiben. Da ausserdem

$$\langle Ae_1, Ae_2\rangle \;=\; \langle e_1, e_2\rangle \;= 0 \tag{6.40}$$

gilt (d. h. Ae_2 steht normal auf Ae_1), ergeben sich für Ae_2 nur die zwei Möglichkeiten

$$Ae_2 = \begin{pmatrix}-\sin\varphi\\ \cos\varphi\end{pmatrix} \quad \text{oder} \quad Ae_2 = \begin{pmatrix}\sin\varphi\\ -\cos\varphi\end{pmatrix}. \tag{6.41}$$

Satz 6.48. *Eine 2×2-Matrix hat genau dann die Eigenschaft $\langle Ax, Ay\rangle = \langle x, y\rangle$ für alle $x, y \in \mathbb{R}^2$, wenn es ein $\varphi \in \mathbb{R}$ gibt, sodass gilt*

$$A = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix} \quad \textit{oder} \quad A = \begin{pmatrix}\cos\varphi & \sin\varphi\\ \sin\varphi & -\cos\varphi\end{pmatrix}. \tag{6.42}$$

Definition 6.49. *Die Menge aller reellen 2×2-Matrizen aus Satz 6.48 wird mit $O(2)$ bezeichnet (orthogonal), die Teilmenge*

$$\{A \in O(2)\ | A = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}, \varphi \in \mathbb{R}\} \tag{6.43}$$

wird mit $SO(2)$ bezeichnet.

Geometrisch, nämlich als Abbildung $\mathbb{R}^2 \to \mathbb{R}^2$, beschreibt die Matrix

$$A_\varphi = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix} \in SO(2) \tag{6.44}$$

eine Drehung um den Nullpunkt um den Winkel φ, während

$$B_\varphi = \begin{pmatrix}\cos\varphi & \sin\varphi\\ \sin\varphi & -\cos\varphi\end{pmatrix} \in O(2) \setminus SO(2) \tag{6.45}$$

eine Spiegelung an der Achse, die gegen $\mathbb{R} \times 0$ mit dem Winkel $\varphi/2$ geneigt ist, beschreibt.

Für die Hintereinanderausführung (Verknüpfung) von Drehungen und Spiegelungen gilt

$$\begin{aligned} A_\varphi A_\psi &= A_{\varphi+\psi}, \\ B_\varphi A_\psi &= B_{\varphi-\psi}, \\ A_\psi B_\varphi &= B_{\psi+\varphi}, \\ B_\varphi B_\psi &= A_{\varphi-\psi}, \\ A_0 &= E, \\ A_\varphi^{-1} &= A_{-\varphi}, \\ B_\varphi^{-1} &= B_\varphi, \quad B_\varphi^2 = E, \end{aligned} \tag{6.46}$$

wie man leicht sieht, wenn man die entsprechenden Matrizen multipliziert.

Analog wird im $\mathbb{R}^3$ eine Drehung um den Winkel α gegen den Uhrzeigersinn um die x, y, bzw. z-Achse beschrieben durch die folgenden Drehmatrizen:

$$A_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}, \quad A_y = \begin{pmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{pmatrix}, \tag{6.47}$$

$$A_z = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

6.7.6. (*) Affine Space and Homogeneous Coordinates. The basic mathematical object used for computer graphic is an affine space. An affine space[6] consists of a vector space V together with a set of points A, where the sum of a point p and a vector v yields a point $p + v$ in A. One can think of an affine space as a vector space without an origin. Affine transformations consist of linear transformations of vectors and translations of points. Every linear transformation in a vector space can be represented by a matrix, but translations are given by vector additions. This awkward combination of vector addition and matrix multiplication can be avoided by introducing a new coordinate system, the homogeneous coordinates. Consider the following example in $\mathbb{R}^2$ of a linear transformation and a translation $\vec{x}' = A\,\vec{x} + \vec{e}$

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} ax + by + e \\ cx + dy + f \end{pmatrix}. \tag{6.48}$$

[6] http://mathworld.wolfram.com/AffineSpace.html

When we introduce an additional coordinate we can perform this two operations by one matrix multiplication.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & e \\ c & d & f \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} ax + by + e \\ cx + dy + f \\ 1 \end{pmatrix}. \tag{6.49}$$

Note that when one uses matrix multiplication from the right, as some books do, one must use the transpose matrix.

Instead of $(x, y, 1)$ any multiple (rx, ry, r) with $r \neq 0$ can be used.

Thus we have constructed a new space consisting of points (x, y, w) of $\mathbb{R}^3$ where points lying on a line are identified. This yields an equivalence relation.

This space is called the real *projective plane* $\mathbb{RP}^2$. An alternative method is to consider the sphere S^2 and identify opposite points.

This construction can be easily extended to higher dimensions.

6.7.7. (*) Anwendung: Codierungstheorie. Ein wichtiges Problem in der Informatik ist die fehlerfreie Übertragung von Daten. Die Daten werden dabei als Folge von Nullen und Einsen in Blöcken von k Bits übermittelt. Jeder Block aus k Bits kann mathematisch als ein Vektor $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{Z}_2^k$ betrachtet werden. Zur Erkennung von Übertragungsfehlern werden an diesen Vektor $n - k$ Kontrollbits $(a_{k+1}, \ldots, a_n)$ angehängt und es entsteht ein Vektor $\mathbf{c} = (x_1, \ldots, x_k, a_{k+1}, \ldots, a_n) \in \mathbb{Z}_2^n$. Die so entstehenden Vektoren werden **Codewörter** genannt. Für das Codewort $(x_1, \ldots, x_k, a_{k+1}, \ldots, a_n)$ schreibt man abkürzend $(\mathbf{x}, \mathbf{a})$. Bilden die Codewörter einen Vektorraum (Teilraum des $\mathbb{Z}_2^n$), so spricht man von einem **linearen Code** C. Da jeder Vektor in $\mathbb{Z}_2^k$ als

$$\mathbf{x} = \sum_{j=1}^{k} x_j \mathbf{e}_j$$

geschrieben werden kann, kann jedes Codewort eines linearen Codes C in der Form

$$\mathbf{c} = \sum_{j=1}^{k} x_j (\mathbf{e}_j, \mathbf{a}_j)$$

geschrieben werden, wobei $\mathbf{a}_j$ die zu $\mathbf{e}_j$ gehörigen Kontrollbits sind. Die Vektoren $(\mathbf{e}_j, \mathbf{a}_j)$ schreibt man als Zeilenvektoren in eine Matrix G und nennt sie **Generatormatrix** des linearen Codes. Jeder Vektor $\mathbf{x} \in \mathbb{Z}_2^k$ wird dann auf sein Codewort

$$\mathbf{c} = G^t \mathbf{x}$$

abgebildet (codiert). Man beachte: Da die Codewörter Zeilenvektoren sind, muss man die Transponierte von G nehmen (die Spalten von G^t sind die Bilder

der Einheitsvektoren). Die Generatormatrix G ist von der Form

$$G = (\mathrm{E}_k \;\; A)$$

mit A einer $k \times n$-Matrix. Zum Beispiel

$$G = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \qquad A = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Zu jeder Generatormatix G kann man nun eine zugehörige **Kontrollmatrix**

$$H = (A^t \;\; \mathrm{E}_{n-k})$$

definieren. In unserem Beispiel

$$H = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

Eine kleine Rechnung zeigt

$$H\,G^t = (A^t \;\; \mathrm{E}_{n-k}) \begin{pmatrix} \mathrm{E}_k \\ A^t \end{pmatrix} = A^t\,\mathrm{E}_k + \mathrm{E}_{n-k}\,A^t = A^t + A^t = \mathbf{0}$$

(Achtung: $x + x = 0$ in $\mathbb{Z}_2$) und das bedeutet,

$$H\mathbf{c} = H\,G^t\,\mathbf{x} = 0 \qquad \text{für alle } \mathbf{c} \in C.$$

Man kann also durch Anwenden der Kontrollmatrix daher leicht feststellen, ob es sich um ein Codewort handelt.

Zum Beispiel: Der Sender möchte $(0, 1)$ senden. Er codiert es zu

$$\mathbf{c} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

und übermittelt $\mathbf{c} = (0, 1, 1)$ dem Empfänger. Dieser empfängt aber z.B. $\tilde{\mathbf{c}} = (1, 1, 1)$ (in der ersten Stelle ist ein Übertragungsfehler aufgetreten) und überprüft

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1 + 1 + 1 = 1 \neq 0.$$

Der Fehler wird also erkannt und der Sender muss gebeten werden, die Daten nochmals zu übertragen.

Bei dieser Methode kann allerdings nicht jeder Übertragungsfehler erkannt werden. Wird ein Codewort $\mathbf{c}$ gesendet und ein anderes Codewort $\tilde{\mathbf{c}}$ empfangen, so wird der Fehler nicht erkannt. Indem man die Anzahl der Prüfbits entsprechend groß wählt, kann die Wahrscheinlichkeit dafür aber beliebig klein gemacht werden. Ist das empfangene Wort kein Codewort, so wird aber der Fehler immer erkannt. Warum? – Codewörter werden von H auf den Nullvektor abgebildet, es gilt also $C \subset \mathrm{Kern}(H)$. Der Kern von H hat also mindestens die Dimension von C, die gleich k ist. Würde es noch weitere Vektoren

im Kern von H geben, die nicht in C liegen, so wäre die Dimension des Kerns grösser k und damit der Rang von H nach der Dimensionsformel kleiner $n-k$. H enthält aber I_{n-k} und damit $n-k$ linear unabhängige Spaltenvektoren. Somit ist der Rang von H gleich $n-k$ und man erhält einen Widerspruch.

Bei genügend vielen Kontrollbits ist es sogar möglich nicht nur Fehler zu erkennen, sondern man kann sogar auch Fehler korrigieren. Das ist besonders wichtig, wenn das Senden z.B. dem Speichern von Daten auf einer CD entspricht. Das Empfangen ist dann das Lesen der Daten von CD.

6.8. Übung

(1) Gegeben seien die Matrizen

$$A = \begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & -3 & 0 \\ 0 & 2 & 4 \\ 3 & 3 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Berechne: $A + C, 2A + 3B + C$, $(A + B) \cdot D, C \cdot D$,
$A \cdot B$, $A \cdot C$, $B \cdot C$, $B \cdot A$, $C \cdot A$, $C \cdot B$, $A \cdot A$, $B \cdot B$ und $C \cdot C$.

(2) Man bestimme den Rang folgender Matrizen

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad \begin{pmatrix} 1 & -3 & 0 \\ 0 & 2 & 4 \\ 3 & 3 & 0 \end{pmatrix}.$$

6.9. Übung

(1) Man berechne die Inversen folgender Matrizen (falls möglich)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & -1 \\ 5 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad \begin{pmatrix} 1 & -3 & 0 \\ 0 & 2 & 4 \\ 3 & 3 & 0 \end{pmatrix}.$$

(2) Wie lautet die Matrix $A_\varphi \in SO(2)$ für die Drehung um 45°, 60°, 90°?

(3) (*) Man zeige für $n = 2$, dass die Mengen $O(n)$ und $SO(n)$ mit der Matrizenmultiplikation eine Gruppe bilden. Sind beide abelsch? Warum bildet die Untermenge der Spiegelungs-Matrizen $B_\varphi \in O(n) \setminus SO(n)$ keine Gruppe?

(4) Es sei (v_1, v_2, v_3, v_4) linear unabhängig im reellen Vektorraum V. Man zeige durch Rangbestimmung, dass dann auch (w_1, w_2, w_3) linear unabhängig sind, wobei

$$\begin{array}{lllllllll} w_1 & = & & & v_2 & - & v_3 & + & v_4 \\ w_2 & = & v_1 & + & 2\,v_2 & - & v_3 & - & v_4. \\ w_3 & = & -v_1 & + & v_2 & + & v_3 & + & v_4 \end{array}$$

(5) Man bestimme für welche $\lambda \in \mathbb{R}$ die reelle Matrix

$$\begin{pmatrix} 1 & \lambda & 0 & 0 \\ \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \end{pmatrix}$$

invertierbar ist und berechne für diese λ die inverse Matrix A_λ^{-1}.

(6) Im **RGB-Farbmodell** wird eine Farbe durch ein Tripel (r, g, b) dargestellt, wobei r für den Rotanteil, g für den Grünanteil und b für den Blauanteil der dargestellten Farbe steht. Beispiel: $(1, 0, 0)$ bedeutet „rot“, $(0, 0, 1)$ „blau“, $(1, 1, 0)$ „gelb“ und $(1, 1, 1)$ weiß. Für Videosignale und für das Farbfernsehen wird (bei der NTSC-Farbcodierung) das sogenannte **YIQ-Farbmodell** verwendet. Dabei wird ein RGB-Signal so codiert übertragen, dass die gleiche Empfangscodierung für Schwarz/Weiß- und für Farbbildschirme verwendet werden kann. Im YIQ-Modell enthält ein Tripel (y, i, q) eine Luminanzkomponente y und zwei Chrominanzkomponenten i und q. Die Luminanzkomponente enthält alle Informationen, die ein Schwarz/Weiß-Bildschirm benötigt. Die Umrechnung vom RGB- in das YIQ-Modell ist eine (bijektive) lineare Abbildung:

$$\begin{pmatrix} y \\ i \\ q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.528 & 0.311 \end{pmatrix} \begin{pmatrix} r \\ g \\ b \end{pmatrix} \tag{6.50}$$

So wird etwa die Farbe „rot“ im YIQ-Modell dargestellt als

$$\begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.528 & 0.311 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.299 \\ 0.596 \\ 0.212 \end{pmatrix}$$

Wenn nun umgekehrt eine Farbe im YIQ-Modell gegeben ist, wie kann man dann ins RGB-Modell umrechnen? - Dazu muss nur die inverse Matrix A^{-1} berechnet werden (die der Umkehrabbildung entspricht): $(r, g, b) = A^{-1}\,(y, i, q)$.

Die NTSC-Farbcodierung wird vor allem in den USA und Japan verwendet und böse Zungen sprechen von „never twice the same color“. Bei der Alternative, der PAL-Farbcodierung, wird das so genannte **YUV-Farbmodell** verwendet. Auch die Umrechnung von RGB auf YUV ist eine lineare Abbildung.

(a) Man berechne mit *Mathematica* $(y, i, q) = (0.5, 0.5, 0.5)$ in RGB-Farben um. Hinterfragen sie das Ergebnis! Wie kann man bestimmen ob ein gültiger Wert für (y, i, q) vorliegt?

(7) Ein **Markov-Prozess** ist ein stochastischer Prozess, bei dem die Wahrscheinlichkeit, einen bestimmten Zustand zu erreichen, *nur* vom vorhergehenden Zustand abhängt (benannt nach dem russischen Mathematiker A. Markov). Ein Markov-Prozess kann mithilfe einer Matrix beschrieben werden: wenn A die Matrix ist, die den Prozess beschreibt, und $\mathbf{x}$ der Anfangszustand, dann ist $\mathbf{y} = A\mathbf{x}$ der darauffolgende Zustand und allgemein $\mathbf{y} = A^n\mathbf{x}$ der Zustand nach n Schritten.

In einer Stadt gibt es 4000 verheiratete Männer und 1000 unverheiratete Männer. Angenommen, 20% der ledigen Männer heiraten jedes Jahr, und 10% der verheirateten Männer werden jährlich geschieden. Nehmen wir weiters an, dass die Gesamtanzahl der Männer gleich bleibt. Berechnen Sie die Anzahl der verheirateten / ledigen Männer in einem, zwei und drei Jahren. Was passiert nach zehn, zwanzig, dreißig Jahren?

(Hinweis: Bilden Sie eine Matrix A mit a_{11} = Prozentsatz der verheirateten Männer, die verheiratet bleiben; a_{12} = Prozentsatz der ledigen Männer, die heiraten; a_{21} = Prozentsatz der verheirateten Männer, die geschieden werden; a_{22} = Prozentsatz der ledigen Männer, die ledig bleiben; die Veränderung innerhalb eines Jahres kann dann durch $\mathbf{y} = A\mathbf{x}$ beschrieben werden, wobei $\mathbf{x} = (4000, 1000)$ der Anfangszustand ist. Die Potenz A^n kann leicht mit dem `Mathematica`-Befehl `MatrixPower[A, n]` berechnet werden.)

(8) **Inzidenzmatrix:** Matrizen können verwendet werden, um Verbindungen (zum Beispiel in elektrischen Netzwerken, in Straßennetzen, in Produktionsabläufen, usw.) zu beschreiben. Abbildung 6.2 zeigt ein elektrisches Netzwerk, das aus 4 Knoten und 5 Kanten besteht. Knoten und Kanten werden beliebig durchnummeriert. (Der Referenzknoten,

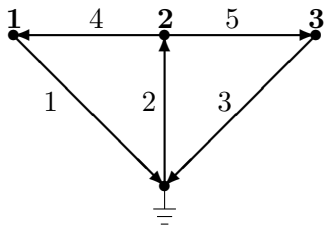


Abbildung 6.2. Elektrisches Netzwerk

der geerdet ist, wird dabei nicht mit einer Nummer versehen). Dann kann das Netzwerk durch seine so genannte Inzidenzmatrix beschrieben werden, deren Elemente gegeben sind durch:

$$a_{ik} = \begin{cases} +1, & \text{wenn von Knoten } i \text{ die Kante } k \text{ ausgeht} \\ -1, & \text{wenn in Knoten } i \text{ die Kante } k \text{ einmündet} \\ 0, & \text{wenn Knoten } i \text{ und Kante } k \text{ einander nicht berühren} \end{cases}$$

Geben Sie die Inzidenzmatrix des Netzwerks aus Abbildung 6.2 an.

6.10. Determinanten

Anschaulich kann man die Determinante einer $n \times n$-Matrix als das von den Zeilenvektoren (bzw. Spaltenvektoren) aufgespannte (mit Vorzeichen versehene, Stichwort: Orientierung (Drehsinn)) "Volumen" interpretieren.

(i) Damit ist sofort die Linearität in jeder Spalte (Zeile) anschaulich ersichtlich.
(ii) Auch ist einleuchtend, dass die Determinate Null ist, wenn zwei Spalten (Zeilen) linear abhängig sind (sie spannen ja kein Volumen auf).
(iii) Die Determinate der Einheitsmatrix ist das Volumen des Einheitswürfel und das ist 1.

Beispiel: Determinante in $\mathbb{R}^2$. In diesem Fall ist das orientierte Volumen eine Fläche.

Diese drei Eigenschaften charakterisieren die Determinate vollständig.

Satz 6.50. *Es gibt genau eine Abbildung* $\det : M(n \times n, \mathbb{K}) \to \mathbb{K}, A \mapsto \det(A)$ *mit folgenden Eigenschaften*

(i) det *ist linear in jeder Zeile,*

(ii) ist der (Zeilen)-Rang von A kleiner als n, so ist $\det(A) = 0$,

(iii) $\det(E) = 1$.

Definition 6.51. *Die Abbildung* $\det : M(n \times n, \mathbb{K}) \to \mathbb{K}$ *heisst* Determinante, *die Zahl* $\det(A)$ *heisst die* Determinante von A.

„linear in jeder Zeile" heisst:

Man betrachte die Matrix A_x in der $n-1$ Zeilen fix und $x \in \mathbb{K}^n$ ist, also

$$A_x = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ x_1 & \dots & x_n \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}.$$

Dann ist $\det : M(n \times n, \mathbb{K}) \to \mathbb{K}, A_x \mapsto \det(A_x)$ linear in dieser Zeile, wenn die durch $x \mapsto \det A_x$ gegebene Abbildung $\mathbb{K}^n \to \mathbb{K}$ linear ist, d. h. es gilt

$$\det A_{x+y} = \det A_x + \det A_y,$$
$$\det A_{\lambda x} = \lambda \det A_x.$$

Anmerkung: Man kann die Determinante daher auch als eine multilineare Abbildung auffassen, die den n Zeilenvektoren (Spaltenvektoren) eine Zahl zuordnet, also als Abbildung $\mathbb{K}^n \times \cdots \times \mathbb{K}^n \to \mathbb{K}$.

Eigenschaften der Determinante unter Zeilenumformungen

Satz 6.52. *Es sei* $\det : M(n\times n, \mathbb{K}) \to \mathbb{K}$ *eine Abbildung mit den Eigenschaften (i) und (ii) aus Satz 6.50. Dann gilt*

(i) Verwandelt man die Matrix A durch Vertauschen zweier Zeilen in eine Matrix A′, so gilt $\det(A') = -\det(A)$.

(ii) Verwandelt man die Matrix A durch Multiplikation einer Zeile mit $\lambda \in \mathbb{K}$ *in eine Matrix A′, so gilt* $\det(A') = \lambda \det(A)$.

(iii) Verwandelt man die Matrix A durch Addition eines Vielfachen einer Zeile zu einer anderen Zeile in eine Matrix A′, so gilt $\det(A') = \det(A)$.

Beweis:

(i)

$$
\begin{aligned}
0 = \det &\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1}+a_{j1} & \dots & a_{in}+a_{jn} \\ \vdots & & \vdots \\ a_{j1}+a_{i1} & \dots & a_{jn}+a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \\
= \det &\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{j1}+a_{i1} & \dots & a_{jn}+a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} + \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & & \vdots \\ a_{j1}+a_{i1} & \dots & a_{jn}+a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \\
= \det &\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} + \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \det(A) + \det(A').
\end{aligned}
$$

Daraus folgt $\det(A') = -\det(A)$.

(ii) Es gilt ja Linearität in jeder Zeile

$$\det A_{\lambda x} = \lambda \det A_x.$$

(iii)

$$\det(A') = \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} + \lambda a_{j1} & \dots & a_{in} + \lambda a_{jn} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

$$= \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} + \lambda \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & & \vdots \\ a_{j1} & \dots & a_{jn} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \det(A).$$

Beispiel: Was ergibt $\det(\lambda A)$?

Definition 6.53. *Ist $A \in M(n \times n, \mathbb{K})$ so bezeichnet man mit A_{jk} die aus A durch Weglassen der j-ten Zeile und der k-ten Spalte entstandene $(n-1) \times (n-1)$-Matrix. Das Element $(-1)^{j+k} \det A_{jk}$ heisst Cofactor von a_{jk}.*

Man definiert damit nun

$$\det A := \sum_{j=1}^{n} (-1)^{j+k} a_{jk} \det A_{jk}. \tag{6.51}$$

Die so definierte Abbildung erfüllt die Eigenschaften (i), (ii) und (iii) aus Satz 6.50 (Beweis: Jänich, Seite 139) und ergibt somit eine Formel zur iterativen Berechnung der Determinante. Man nennt dies auch *Entwickeln* einer Determinate nach der k-ten Spalte.

Beispiele:

(i) $n = 1$:

$$\det(a) = a.$$

(ii) $n = 2$:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

(iii) $n = 3$:

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{21} \det \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} + a_{31} \det \begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}.$$

Satz 6.54. *Ist $A \in M(n \times n, \mathbb{K})$ eine obere Dreiecksmatrix, d. h. alle Elemente unterhalb der Hauptdiagonale sind Null ($a_{jk} = 0$ für $j > k$), so ist die Determinante das Produkt der Hauptdiagonalelemente* $\det A = a_{11} a_{22} \cdots a_{nn}$.

Berechnungsverfahren für die Determinante großer Matrizen:

Man verwende elementare Zeilenumformungen der Typen 1 und 3 um A in eine obere Dreiecksmatrix A' zu verwandeln. Hat man dabei r Zeilenvertauschungen angewandt, so gilt

$$\det A = (-1)^r \det A' = (-1)^r a'_{11} a'_{22} \cdots a'_{nn}. \tag{6.52}$$

Da für die Determinate einer Matrix A gilt, dass sie gleich der Determinate ihrer Transponierten ist, erhält man damit den Entwicklungssatz nach Zeilen.

Satz 6.55. *Es gilt*

(i) $\det A^T = \det A$.

(ii) Entwickeln nach der j-ten Zeile

$$\det A := \sum_{k=1}^{n} (-1)^{j+k} a_{jk} \det A_{jk}. \tag{6.53}$$

6.10.1. Komplementäre Matrix.

Definition 6.56. *Ist $A \in M(n \times n, \mathbb{K})$, so heisst die durch*

$$\tilde{a}_{jk} := (-1)^{j+k} \det A_{kj} \tag{6.54}$$

definierte Matrix $\tilde{A} \in M(n \times n, \mathbb{K})$ die komplementäre Matrix.

Achtung: Der Index kj in A_{kj} ist kein Druckfehler!

Beispiel: Man berechne die komplementäre Matrix von

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Die komplementäre Matrix ist so definiert, dass

$$\det A = \sum_{k=1}^{n} a_{jk}\tilde{a}_{kj}. \tag{6.55}$$

Satz 6.57. *Ist $\tilde{A}$ die zu $A \in M(n \times n, \mathbb{K})$ komplementäre Matrix, so ist*

$$A\tilde{A} = \det A \cdot E. \tag{6.56}$$

Für die inverse Matrix A^{-1} ergibt sich daher die Formel

$$A^{-1} = \frac{1}{\det A} \cdot \tilde{A}. \tag{6.57}$$

Satz 6.58. *Eine $n \times n$-Matrix ist genau dann invertierbar (hat Rang n), wenn $\det A \neq 0$ ist.*

Satz 6.59. *Für alle $A, B \in M(n \times n, \mathbb{K})$ gilt*

$$\det(AB) = \det A \cdot \det B.$$

Satz 6.60. *Ist A invertierbar, d. h. $\det A \neq 0$, so ist*

$$\det A^{-1} = \frac{1}{\det A}.$$

Wir können nun mit Hilfe des kanonischen Basisisomorphismus für beliebige lineare Abbildungen $f : V \to V$ den Begriff der Determinante definieren.

Satz 6.61. *Es sei $f : V \to V$ eine lineare Abbildung in einem n-dimensionalen Vektorraum. Sei A die Matrix der Abbildung f bezüglich einer Basis $(v_1, \ldots, v_n)$ und B die Matrix der Abbildung f bezüglich einer anderen Basis $(v'_1, \ldots, v'_n)$, so gilt*

$$\det A = \det B =: \det f. \tag{6.58}$$

Anmerkung: Eine explizite Formel für die Determinante ist durch die Leibnizsche Formel gegeben, [**66**], Seite 153.

6.11. Übung

(1) Wie lautet die Regel von Sarrus? (Jägerzaunregel, Achtung: gilt nur für $n = 3$!)

(2) (a) Von einem Parallelogramm sind die Punkte $A = (1, 1), B = (5, 2)$ und $D = (2, 5)$ gegeben. Man berechne dessen Fläche.

(b) Von einem Parallelepiped[7] sind die Punkte $A = (0, 0, 0), B = (5, 2, 1)$, $D = (2, 5, 2)$ und $E = (1, 1, 6)$ gegeben. Man berechne dessen Volumen.

(3) Man berechne die Determinante folgender Matrizen

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 2 & 7 & 9 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 3 & 7 \\ 4 & 2 & 0 & 1 \\ 7 & 7 & 3 & 0 \\ 5 & 0 & 6 & 8 \end{pmatrix}, \quad \begin{pmatrix} x & 1 & 1 \\ 1 & x & 1 \\ 1 & 1 & x \end{pmatrix}, \quad \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix}.$$

(4) Man berechne die komplementären und transponierten Matrizen von

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 2 & 7 & 9 \end{pmatrix}, \quad \begin{pmatrix} 0 & 2 & -3 \\ 2 & 0 & 5 \\ -3 & 5 & 0 \end{pmatrix}, \quad \begin{pmatrix} m & n & p \\ p & m & n \\ n & p & m \end{pmatrix}.$$

[7] https://de.wikipedia.org/wiki/Parallelepiped

6.12. Lineare Gleichungssysteme

Definition 6.62. *Ist $A = (a_{jk}) \in M(m \times n, \mathbb{K})$ und $b = (b_1, \ldots, b_m) \in \mathbb{K}^m$, so nennt man*

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{6.59}$$

ein lineares Gleichungssystem *für $(x_1, \ldots, x_n)$ mit Koeffizienten aus $\mathbb{K}$. Die $x_j, 1 \leq j \leq n$ heissen die* Unbekannten *des Systems. Sind alle $b_j, 1 \leq j \leq m$ gleich Null, so heisst das Gleichungssystem* homogen *ansonsten* inhomogen.

Wir können ein Gleichungssystem daher kurz in der Form $Ax = b$ schreiben.

Definition 6.63. *Unter der* Lösungsmenge *des zu (A, b) gehörenden Gleichungssystem versteht man die Menge*

$$\text{Lös}(A, b) = \{x \in \mathbb{K}^n \mid Ax = b\}. \tag{6.60}$$

Ein Gleichungssystem heisst lösbar*, wenn die Lösungsmenge nicht leer ist.*

Formal ist die Lösungsmenge durch die Urbildmenge $A^{-1}(\{b\})$ gegeben.

Mit unseren bisher eingeführten Begriffen können wir die Existenz von Lösungen wie folgt charakterisieren.

Satz 6.64. *Das Gleichungssystem $Ax = b$ ist genau dann lösbar, wenn*

$$\text{Rang} \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} = \text{Rang} \begin{pmatrix} a_{11} & \ldots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} & b_m \end{pmatrix}$$

Beweis: Der Rang der Matrix A ist die Dimension des Bildraumes der Abbildung A. Betrachtet man nun die Dimension des Bildraumes der **erweiterten Matrix** (A, b), so gibt es zwei Möglichkeiten:
(i) Ist sie gleich, so kann b aus den Spalten von A linearkombiniert werden und das System hat mindestens eine Lösung. Anschaulich bedeutet dies, dass der Vektor b im Bildraum von A liegt.
(ii) Ist sie grösser, so kann b nicht aus den Spalten von A linearkombiniert werden und liegt nicht im Bildraum von A . Das System hat daher keine Lösung.

Anmerkung: Homogene Systeme haben daher immer mindestens eine Lösung.

Satz 6.65. *Ist $x_0 \in \mathbb{K}^n$ eine Lösung, d. h. gilt $Ax_0 = b$, so erhält man alle Lösungen durch*

$$\text{Lös}(A, b) = x_0 + \text{Kern}\, A := \{x_0 + x \mid x \in \text{Kern}\, A\} \tag{6.61}$$

Beweis: Ist x_0 eine Lösung, d. h. $Ax_0 = b$, dann gilt

$$A(x_0 + x_k) = Ax_0 + Ax_k = b + 0 = b,$$

wobei x_k einen beliebigen Vektor aus dem Kern von A bezeichnet. Ist umgekehrt x_1 eine beliebige Lösung, dann gilt für die Differenz $x_1 - x_0$

$$A(x_1 - x_0) = Ax_1 - Ax_0 = b - b = 0,$$

d. h. die Differenz $x_1 - x_0$ liegt im Kern von A.

Kennt man eine Lösung x_0 und eine Basis $(v_1, \ldots, v_d)$ des Kerns von A, so ist die Lösungsmenge durch

$$\text{Lös}(A, b) = \{x_0 + \lambda_1 v_1 + \ldots + \lambda_d v_d \mid \lambda_j \in \mathbb{K}\} \tag{6.62}$$

gegeben. Dabei ist $d = \dim \text{Kern}\, A = n - \text{Rang}\, A$.

Satz 6.66. *Aus Satz 6.65 ergibt sich, dass ein lösbares Gleichungssystem $Ax = b$ genau dann eindeutig lösbar ist, wenn* $\text{Kern}\, A = \{0\}$ *gilt, d. h.* $\text{Rang}\, A = n$ *ist.*

Für den Spezialfall $m = n$ bedeutet dies, dass $Ax = b$ genau dann eindeutig lösbar ist, wenn $\det A \neq 0$ *ist.*

In diesem Fall ist die Matrix A invertierbar und die Lösung x durch $x = A^{-1}b$ gegeben.

Für den Fall, dass $\det A \neq 0$, gibt es auch eine explizite Determinantenformel

Satz 6.67. Cramersche Regel. *Ist* $\det A \neq 0$ *und $Ax = b$, so gilt*

$$x_j = \frac{\det \begin{pmatrix} a_{11} & \ldots & b_1 & \ldots & a_{1n} \\ \vdots & & & & \vdots \\ a_{n1} & \ldots & b_n & \ldots & a_{nn} \end{pmatrix}}{\det \begin{pmatrix} a_{11} & \ldots & \ldots & \ldots & a_{1n} \\ \vdots & & & & \vdots \\ a_{n1} & \ldots & \ldots & \ldots & a_{nn} \end{pmatrix}}, \qquad 1 \leq j \leq n. \tag{6.63}$$

Dabei wurde in der ersten Matrix die j-te Spalte durch den Vektor b ersetzt.

Beweis:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1,n}x_n &= b_1 \\ &\vdots \\ a_{n1}x_1 + \cdots + a_{n,n}x_n &= b_n \end{aligned}$$

kann geschrieben werden als

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} x_1 + \cdots + \begin{pmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{pmatrix} x_n = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Umformen ergibt

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix} x_1 + \cdots + \begin{pmatrix} x_j a_{1j} - b_1 \\ \vdots \\ x_j a_{nj} - b_n \end{pmatrix} \cdot 1 + \cdots + \begin{pmatrix} a_{1n} \\ \vdots \\ a_{nn} \end{pmatrix} x_n = 0,$$

was zeigt, dass diese Vektoren linear abhängig sind. Die Determinate von

$$B = \begin{pmatrix} a_{11} & \dots & (x_j a_{1j} - b_1) & \dots & a_{1n} \\ \vdots & & & & \vdots \\ a_{n1} & \dots & (x_j a_{nj} - b_n) & \dots & a_{nn} \end{pmatrix}$$

muss daher Null sein. Aus der Linearität der Determinate folgt nun

$$\det(B) = x_j \det(A) - \det((A_{j,b})) = 0,$$
$$x_j = \frac{\det((A_{j,b}))}{\det(A)},$$

wobei hier $A_{j,b}$ die Matrix bezeichnet in der die j-te Spalte durch den Vektor b ersetzt wird.

6.12.1. Der Gauss' sche Algorithmus. Zur praktischen Berechnung von Lösungen eines Gleichungssystems verwendet man den Gauss' schen Algorithmus (bzw. Variationen davon), der auf folgendem Prinzip beruht.

Satz 6.68. *Wird die erweiterte Matrix (A, b) durch elementare Zeilenumformungen in eine Matrix (A', b') überführt, so ändert sich die Lösungsmenge nicht, d. h.* $\text{Lös}(A, b) = \text{Lös}(A', b')$.

Anmerkungen: Dies gilt nicht für Spaltenumformungen.

Gauss' scher Algorithmus: Es sei A eine $n \times n$-Matrix mit $\det A \neq 0$.

Durch Vertauschung von Zeilen erreicht man, dass der erste Koeffizient a_{11} ungleich Null ist. Durch Addition geeigneter Vielfacher der ersten Zeile zu den anderen Zeilen kann man nun den ersten Koeffizienten der anderen Zeilen zu Null machen. Damit ist der erste Schritt beendet. Als nächstes erreicht man durch Zeilenvertauschungen, dass in der zweiten Zeile a_{22} ungleich Null wird. Durch Addition geeigneter Vielfacher der zweiten Zeile zu den anderen Zeilen

kann man die zweiten Koeffizienten der anderen Zeilen (unterhalb der Hauptdiagonale) zu Null machen. Setzt man dieses Verfahren fort, so erhält man schliesslich eine Matrix der Form

$$\left(\begin{array}{ccc|c} a'_{11} & \cdots & a'_{1n} & b'_1 \\ & \ddots & \vdots & \vdots \\ 0 & & a'_{nn} & b'_n \end{array}\right).$$

Daraus kann man die Lösung für x_n unmittelbar ausrechnen und in der Folge die restlichen Unbekannten.

$$\begin{aligned} x_n &= \frac{b'_n}{a'_{nn}}, \\ x_{n-1} &= \frac{1}{a'_{n-1,n-1}}(b'_{n-1} - a'_{n-1,n}x_n), \\ &\text{usw.} \end{aligned}$$

Beispiel: Man löse

$$\begin{array}{rcrcrcr} -x_1 & + & 2\,x_2 & + & x_3 & = & -2 \\ 3\,x_1 & - & 8\,x_2 & - & 2\,x_3 & = & 4\ . \\ x_1 & & & + & 4\,x_3 & = & -2 \end{array} \tag{6.64}$$

$L = (2, \frac{1}{2}, -1)$.

6.12.2. Lösungsverfahren für beliebige Gleichungssyteme. Man geht zunächst wie beim Gauss' schen Algorithmus vor, bis man zu dem Punkt gelangt, dass alle Koeffizienten einer Spalte unter der Hauptdiagonalen Null sind.

$$\left(\begin{array}{ccc|c|c} a'_{11} & & & & b'_1 \\ & \ddots & & * & \vdots \\ & & & & \\ & 0 & \ddots & & \vdots \\ & & a'_{tt} & & \\ \hline & & & 0 & \\ & 0 & & \vdots \quad B' & \vdots \\ & & & 0 & b'_m \end{array}\right). \tag{6.65}$$

Durch **Vertauschung der Spalten** kann man jedoch den Gauss' schen Algorithmus wieder in Gang bringen bis man bei folgender Situation landet. (Anmerkung: Man muss sich nur für später merken, dass bei Spaltenvertauschungen die Reihenfolge der Unbekannten vertauscht wird.)

$$\left(\begin{array}{ccc|c} a''_{11} & & & b''_1 \\ & \ddots & & \vdots \\ 0 & & a''_{rr} & b''_r \\ \hline & & & \\ & & 0 & \vdots \\ & & & b''_m \end{array}\right). \tag{6.66}$$

Ist $b''_{r+1} = \ldots = b''_m = 0$, so existiert mindenstes eine Lösung, sonst keine. Im ersteren Fall kann man dann das System, indem man die Nullen unterm Strich weglässt, folgendermassen schreiben

$$\left(\begin{array}{c|c|c} T & S & b'' \end{array}\right), \tag{6.67}$$

wobei T quadratisch und invertierbar ist. Den Vektor w der Unbekannten schreibt man in der Form $w = \begin{pmatrix} y \\ z \end{pmatrix}$. Es gilt damit $(T\,,\,S)w = Ty + Sz$. Man konstruiert nun eine spezielle Lösung $w_0 = \begin{pmatrix} y_0 \\ 0 \end{pmatrix}$, wobei $y_0 = T^{-1}(b''_1, \ldots, b''_r)$

ist. Nun konstruiert man noch mit dem Ansatz $w_j = \begin{pmatrix} y_j \\ e_j \end{pmatrix}$, eine Basis des Kerns der Matrix (T, S). $(T\ ,\ S)w_j = Ty_j + Se_j = 0$ ergibt $y_j = -T^{-1}Se_j$.

Beweis: Jänich, Seite 170.

Beispiel: Man löse

$$(A, b) = \left(\begin{array}{cccc|c} 0 & 0 & 1 & 3 & 2 \\ 1 & 2 & 1 & 4 & 3 \\ 1 & 2 & 2 & 7 & 5 \\ 2 & 4 & 1 & 5 & 4 \end{array}\right).$$

Anmerkung 1: Alternativ erhält man die Lösungen auch indem man die Matrix (A, b) auf Zeilenstufenform bringt und freie Parameter einführt, deren Anzahl genau durch die Dimension des Kerns von (T, S) bzw. A gegeben ist. Hier muss man aber aufpassen, welche Unbekannten x_j frei gewählt werden können!

Bei unserem Verfahren (Vertauschung von Spalten) können genau die Variablen $x_n, x_{n-1}, \ldots, x_{r+1}, (r < n)$ frei gewählt werden.

Anmerkung 2: Der erste nichtverschwindende Koeffizient in einer Zeile wird **Pivot** (Dreh-, Angelpunkt) genannt.

Anmerkung 3: Der Gaussalgorithmus führt auf eine obere Dreiecksmatrix. Die dabei durchgeführten Umformungen ergeben eine lineare Transformation, die durch eine untere Dreicksmatrix dargestellt werden kann. Genauer, die Matrix A wird zerlegt in $PA = LU$, wobei U die obere und L eine untere Dreiecksmatrix und P die Permutationsmatrix der Zeilenvertauschungen ist.

Anmerkung 4: Für invertierbare Matrizen A kann man die Lösung von $Ax = b$ in der Form $x = A^{-1}b$ schreiben. Mit Hilfe des Begriffs *Pseudo-Inverses von A* kann man dies auf nicht invertierbare nicht-quadratische Matrizen verallgemeinern, Details siehe [**67**], Seite 94ff.

6.12.3. Ergänzung(*): Lineare Optimierung. In vielen Problemen der Praxis werden Lösungen gesucht, die bestimmten Einschränkungen genügen. Diese Einschränkungen können oft durch lineare *Ungleichungen* beschrieben werden.

Wir wollen uns zunächst eine lineare Ungleichung geometrisch veranschaulichen. Erinnern Sie sich: eine lineare *Gleichung* in zwei Variablen $a_1x + a_2y = b$ beschreibt die Punkte (x, y) einer Geraden im $\mathbb{R}^2$.

Eine **lineare Ungleichung**

$$a_1x + a_2y \geq b$$

beschreibt alle Punkte (x, y) des $\mathbb{R}^2$, die auf oder oberhalb der Geraden $a_1x + a_2y = b$ liegen.

Beispiel 6.1. Lineare Ungleichung: Graphische Veranschaulichung
Kennzeichnen Sie die Punkte (x, y) des $\mathbb{R}^2$, die $x + 2y \geq 4$ erfüllen.

Lösung zu 6.1. *Formen wir die Ungleichung um: $y \geq -\frac{1}{2}x + 2$. Zu vorgegebenem x-Wert darf der zugehörige y-Wert eines Punktes (x, y) also entweder gleich oder größer als $-\frac{1}{2}x + 2$ sein; d.h., der Punkt kann auf oder oberhalb der Geraden $y = -\frac{1}{2}x + 2$ liegen (siehe Abbildung 6.3).* ■

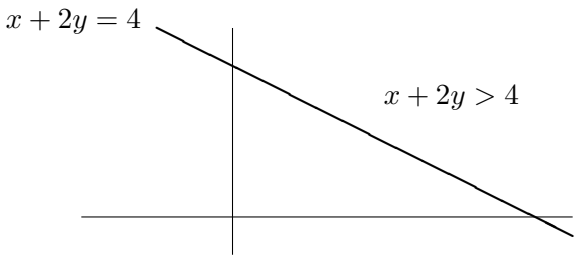


Abbildung 6.3. Die Punkte mit $x + 2y \geq 4$ liegen auf oder oberhalb der Geraden

Folgendes Beispiel führt auf ein System von linearen Ungleichungen:

Beispiel 6.2. System von linearen Ungleichungen: Investmentfonds
Ein Investmentfonds hat ein Kapital von 20 Millionen Euro zur Verfügung, das auf staatliche Pfandbriefe, festverzinsliche Wertpapiere und Aktien verteilt werden soll. Dabei müssen folgende Einschränkungen erfüllt sein:
– Mindestens die Hälfte des Kapitals muss in staatlichen Pfandbriefen oder festverzinslichen Wertpapieren angelegt werden.
– Das Kaptial, das in festverzinslichen Wertpapieren angelegt ist, darf höchstens doppelt so hoch sein, wie das Kapital, das in staatlichen Pfandbriefen angelegt ist.

Formulieren Sie diese Einschränkungen mithilfe von Ungleichungen.

Lösung zu 6.2. *Bezeichnen wir mit x_1, x_2 und x_3 das Kapital, das in Pfandbriefen, festverzinslichen Wertpapieren bzw. Aktien angelegt wird. Es sollen die gesamten* Euro 20 *Millionen angelegt werden, das heißt*

$$x_1 + x_2 + x_3 = 20.$$

Das bedeutet, dass wir eine der Unbekannten durch die übrigen ausdrücken können. Zum Beispiel $x_3 = 20 - x_1 - x_2$, offen sind nun also noch x_1 und x_2.

Da die x_i Geldmengen bedeuten, sind natürlich nur Lösungen mit $x_i \geq 0$ von Interesse. Die Bedingung $x_3 \geq 0$ können wir wieder mithilfe von x_1 und x_2 formulieren, insgesamt gilt also:

$$x_1 \geq 0 \tag{6.68}$$

$$x_2 \geq 0 \tag{6.69}$$

$$20 - x_1 - x_2 \geq 0. \tag{6.70}$$

Nun wird verlangt, dass zumindest die Hälfte des Kaptials in Pfandbriefen oder festverzinslichen Wertpapieren angelegt werden soll, das heißt

$$x_1 + x_2 \geq 10.$$

Weiters wird eingeschränkt, dass der in festverzinslichen Wertpapieren angelegte Geldbetrag höchstens gleich dem doppelten Geldbetrag sein darf, der in Pfandbriefen angelegt ist, also

$$x_2 \leq 2x_1.$$

Die Menge der zulässigen Werte von x_1 und x_2 ist damit gegeben durch

$$\{(x_1, x_2) | x_1 \geq 0,\ x_2 \geq 0,\ x_1 + x_2 \leq 20,\ x_1 + x_2 \geq 10,\ x_2 \leq 2x_1\}.$$

Sie ist begrenzt durch die Geraden $x_2 = 0$, $x_1 + x_2 = 20$, $x_1 + x_2 = 10$, $x_2 = 2x_1$ und in Abbildung 6.4 veranschaulicht (stark umrandeter Bereich). Die zulässigen Werte von x_3 berechnen sich dann aus $x_3 = 20 - x_1 - x_2$. ■

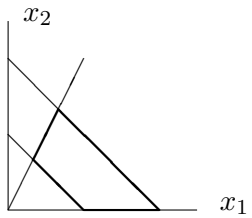


Abbildung 6.4. Die Menge der zulässigen Werte für x_1 und x_2 aus Beispiel 6.2.

Die Menge aus Abbildung 6.4 hat eine wichtige Eigenschaft:

Definition 6.69. *Eine Menge $M \subseteq \mathbb{R}^n$ (oder $\mathbb{C}^n$) heißt* **beschränkt**, *falls es eine Konstante C mit*

$$|x| \leq C \quad \text{für alle } x \in M$$

gibt.

Eine Menge M ist genau dann beschränkt, wenn es eine Kugel (mit Radius C) gibt, so dass alle Punkte von M innerhalb dieser Kugel liegen.

Beispiel 6.3. Beschränkte und unbeschränkte Mengen.

a) *Die Menge $M = \{\mathbf{x} \in \mathbb{R}^2 | x_1 \geq 0, x_2 \geq 0\}$ ist unbeschränkt, da $|\mathbf{x}|$ beliebig groß werden kann.*
b) *Jeder von $\{\mathbf{0}\}$ verschiedene Teilraum U von $\mathbb{R}^n$ ist unbeschränkt. Wir brauchen nur einen Vektor $\mathbf{x} \neq \mathbf{0} \in U$ nehmen, dann kann $k\mathbf{u}$ durch geeignetes $k \in \mathbb{R}$ beliebig lang gestreckt werden ohne U zu verlassen.*
c) *Die Menge $M = \{\mathbf{x} \in \mathbb{R}^2 | 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 3\}$ ist beschränkt. Es gilt $|\mathbf{x}| \leq \sqrt{2^2 + 3^2} = \sqrt{15}$ für alle $\mathbf{x} \in M$.*

In der Praxis ist oft aus der Menge der Lösungen eines Systems von Ungleichungen eine **optimale Lösung** auszuwählen in dem Sinn, dass eine bestimmte Größe ein Maximum (oder Minimum) annehmen soll. Diese Größe ist eine Funktion der Unbekannten, $f(x_1, x_2, \ldots, x_n)$ und die optimale Lösung aus der Menge der zulässigen Lösungen ist jene, für die $f(x_1, x_2, \ldots, x_n)$ ein Maximum (oder Minimum) annimmt. Ist die Funktion f eine *affine* Funktion von $x_1, \ldots, x_n$, also von der Form

$$f(x_1, \ldots, x_n) = c + c_1 x_1 + c_2 x_2 + \ldots c_n x_n, \text{ mit reellen Zahlen } c, c_1, \ldots, c_n,$$

so bezeichnet man das Aufsuchen einer optimalen Lösung als **lineare Optimierung**.

Wie findet man nun das Maximum (oder Minimum) einer affinen Funktion? Dabei hilft der folgende

Satz 6.70. *Die durch das System von linearen Ungleichungen*

$$a_{11}x_1 + \cdots + a_{1n}x_n \geq b_1 \tag{6.71}$$
$$\vdots \tag{6.72}$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n \geq b_m \tag{6.73}$$

definierte Menge M sei beschränkt. Dann nimmt die affine Funktion

$$f(x_1, \ldots, x_n) = c + c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

ihr Maximum und ihr Minimum an einem „Eckpunkt“ der Menge M an.

Man braucht also f nur an den Eckpunkten auszuwerten und aus diesen Funktionswerten das Maximum (Minimum) zu suchen.

Einige Bemerkungen sind angebracht:

- Das Maximum ist nicht immer eindeutig: es kann an mehreren Punkten angenommen werden (denken Sie etwa an den Extremfall $f(x) = c$, hier nimmt f an jedem Punkt ihr Maximum (= Minimum) an).
- Sucht man das Minimum von $f(x_1, \ldots, x_n)$, so ist das gleichbedeutend damit, das Maximum von $-f(x_1, \ldots, x_n)$ zu suchen.
- Aus $a \geq b$ wird durch Multiplikation mit -1 die Ungleichung $-a \leq -b$. Man kann also jedes System von Ungleichungen so schreiben, dass alle Ungleichungen in die gleiche Richtung zeigen.
- Wenn wir anstelle $\geq$ strikte Ungleichungen $>$ betrachten, dann gehören die Eckpunkte nicht mehr zum zulässigen Bereich und damit im Allgemeinen auch der Wert an dem das Maximum (Minimum) angenommen wird nicht mehr.

Warum das so ist, können wir uns so überlegen: Gibt es nur eine Variable x, so hat die zu optimierende Funktion die Form $f(x) = kx + d$, beschreibt also eine Gerade. Wenn wir nun die Gerade für x aus dem Intervall $[a, b]$ betrachten, dann sehen wir sofort, dass der größte bzw. kleinste Funktionswert nur an den Randpunkten angenommen werden kann! Analog ist eine Funktion in zwei Variablen von der Form $f(x, y) = c_1x + c_2y + c$; nun liegen die Punkte $(x, y, f(x, y))$ auf einer Ebene im $\mathbb{R}^3$. Wenn wir uns nun auf alle x, y im Quadrat $0 \leq x \leq 1$ und $0 \leq y \leq 1$ einschränken, so ist wieder anschaulich klar, dass der größte (kleinste) Funktionswert von f nur am Rand des zulässigen x, y-Bereichs angenommen werden kann.

Beispiel 6.4. Lineare Optimierung
Für den Investmentfonds aus Beispiel 6.2 sei der zu erwartende Gewinn bei Investition in staatliche Pfandbriefe gleich 5%, in festverzinsliche Wertpapiere 6% und in Aktien 9%. Bei welcher Aufteilung (unter den oben gegebenen Einschränkungen) kann der Gewinn maximiert werden?

Lösung zu 6.4. *Wenn wieder x_1, x_2 und x_3 das jeweilige Kapital bezeichnen, das in Pfandbriefen, festverzinslichen Wertpapieren bzw. Aktien angelegt wird, dann ist der Gewinn $f(x_1, x_2) = 0.05x_1 + 0.06x_2 + 0.09(20 - x_1 - x_2) = 1.8 - 0.04x_1 - 0.03x_2$. Aus der oben gegebenen Menge der zulässigen (x_1, x_2)-Werte ist nun jenes Paar zu finden, für das der Gewinn $f(x_1, x_2)$ maximal ist. Dazu brauchen wir aber nur die Funktionswerte in den vier Eckpunkten des zulässigen Bereichs (siehe Abbildung 6.4) zu untersuchen. Die Eckpunkte sind die Schnittpunkte der Geraden $x_2 = 0$, $20 - x_1 - x_2 = 0$, $x_1 + x_2 = 10$ und $x_2 = 2x_1$. Der Schnittpunkt von $x_1 + x_2 = 10$ und $-2x_1 + x_2 = 0$ ist zum Beispiel $(x_1, x_2) = (\frac{10}{3}, \frac{20}{3}) = (3.3, 6.6)$, somit ist der zugehörige Gewinn*

$f(\frac{10}{3}, \frac{20}{3}) = 1.46667$. *Analog berechnet man die anderen Eckpunkten und die zugehörigen Funktionswerte und erhält*

(x_1, x_2)	$f(x_1, x_2)$
$(10/3, 20/3)$	1.46667
$(20/3, 40/3)$	1.13333
$(20, 0)$	1
$(10, 0)$	1.4

Daraus sehen wir, dass unter den gegebenen Einschränkungen der maximale Gewinn von Euro 1.47 *Millionen bei einer Aufteilung des Kapitals auf staatliche Pfandbriefe, festverzinsliche Wertpapiere bzw. Aktien gemäß* $(x_1, x_2, x_3) = (\frac{10}{3}, \frac{20}{3}, 10) = (3.3, 6.6, 10.0)$ *auftritt.*

Mit `Mathematica` *kann die Lösung mit dem Befehl* `ConstrainedMax` *gefunden werden:*

```
In[43]:=ConstrainedMax[1.8 − 0.04x1 − 0.03x2, {x1 ≥ 0, x2 ≥ 0,
        20 − x1 − x2 ≥ 0, x1 + x2 ≥ 10, x2 ≤ 2x1}, {x1, x2}]
Out[43]={1.46667, {x1 → 3.33333, x2 → 6.66667}}
```

■

Bei höherdimensionalen Problemen (wenn es also mehr als zwei Unbekannte $x_1, x_2, \ldots, x_n$ gibt) ist es in der Regel nicht mehr effektiv, alle Möglichkeiten durchzuprobieren. Man verwendet dann einen Algorithmus, bei dem man sich bei der Suche nach dem Maximum so von einem Eckpunkt zum nächsten bewegt, dass der Wert von $f(x_1, \ldots, x_n)$ ansteigt; das macht man so lange, bis der Funktionswert nicht mehr weiter zunimmt. Dieses Verfahren ist als **Simplex-Algorithmus** bekannt.

6.13. Übung

(1) Man löse mit dem erweiterten Gauss'schen Algorithmus in Matrixschreibweise

$$\begin{cases} & 7\,x_2 & + & 3\,x_3 & = & -12 \\ 2\,x_1 & + & 8\,x_2 & + & x_3 & = & 0 \\ -5\,x_1 & + & 2\,x_2 & - & 9\,x_3 & = & 26 \end{cases},$$

$$\begin{cases} 4\,x_1 & - & 8\,x_2 & + & 3\,x_3 & = & 16 \\ -x_1 & + & 2\,x_2 & - & 5\,x_3 & = & -21 \\ 3\,x_1 & - & 6\,x_2 & + & x_3 & = & 7 \end{cases},$$

$$(A,b) = \left(\begin{array}{ccccc|c} 0 & 0 & 1 & 3 & 3 & 2 \\ 1 & 2 & 1 & 4 & 3 & 3 \\ 1 & 2 & 2 & 7 & 6 & 5 \\ 2 & 4 & 1 & 5 & 3 & 4 \end{array}\right).$$

(2) Man ermittle durch Rangbestimmung, ob das folgende Gleichungssystem für $d = 4$ lösbar ist und berechne gegebenenfalls die Lösungen

$$\begin{array}{rcrcrcl} x_1 & + & 2\,x_2 & + & 3\,x_3 & = & 1 \\ 4\,x_1 & + & 5\,x_2 & + & 6\,x_3 & = & 2 \\ 7\,x_1 & + & 8\,x_2 & + & 9\,x_3 & = & 3 \\ 5\,x_1 & + & 7\,x_2 & + & 9\,x_3 & = & d \end{array}$$

Für welches d ist das System lösbar? In diesem Falle bestimme man alle Lösungen.

(3) Man ermittle mittels Gauss'schem Algorithmus, ob das folgende Gleichungssystem lösbar ist und berechne gegebenenfalls die Lösungen.

$$\begin{array}{rcrcrcrcl} x_1 & - & x_2 & + & 2\,x_3 & - & 3\,x_4 & = & 7 \\ 4\,x_1 & & & + & 3\,x_3 & + & x_4 & = & 9 \\ 2\,x_1 & - & 5\,x_2 & + & x_3 & & & = & -2 \\ 3\,x_1 & - & x_2 & - & x_3 & + & 2\,x_4 & = & -2 \end{array}$$

(4) Ein Bankkunde möchte Geld in festverzinslichen Wertpapieren der Kategorien AAA, A und B anlegen. Die Papiere der Kategorie AAA bringen 6%, der der Kategorie A bringen 7% und die der Kategorie B bringen 10%. Zinsen. Der Kunde möchte doppelt soviel Geld in der Kategorie AAA anlegen als in der Kategorie B. Wieviel Geld muss der Kunde anlegen, wenn

(a) seine Gesamtanlage Euro 50000.- beträgt und die jährliche Zinseinnahme Euro 3620.- betragen soll,

(b) seine Gesamtanlage Euro 60000.- beträgt und die jährliche Zinseinnahme Euro 4300.- betragen soll,

(c) seine Gesamtanlage Euro 80000.- beträgt und die jährliche Zinseinnahme Euro 5800.- betragen soll.

Hinweis: $x = A^{-1}b$, d.h. A^{-1} berechnen.

(5) Gegeben sind 3 verschiedene Messpunkte: $(x_1, y_1), (x_2, y_2), (x_3, y_3)$. Man ermittle das Polynom 2. Grades, das durch diese 3 Messpunkte durchgeht (Zeichnung). Hinweis: Cramersche Regel anwenden. Die Determinanten brauchen nicht explizit ermittelt werden.

(6) Die elementaren Umformungen im Gauss-Algorithmus sind lineare Transformationen und können daher durch Matrizen dargestellt werden. Es sei im Folgenden $M = (a_{jk})$ eine 3×3–Matrix.

(a) Man bestimme eine Matrix L_1 (Eliminationsmatrix), die in der ersten Spalte von M unterhalb von $a_{11}(\neq 0)$ Nullen erzeugt.

(b) Man bestimme eine Matrix L_2(Eliminationsmatrix), die das $-\frac{a_{32}}{a_{22}}$ –fache ($a_{22} \neq 0$) der zweiten Zeile von M zur dritten Zeile addiert.

(c) Man bestimme eine Matrix (Permutationsmatrix), die die zweite und dritte Zeile (Spalte) der Matrix M vertauscht. Hinweis: Multiplikation von rechts oder links.

6.14. Euklidische Vektorräume

Der Begriff des skalaren Produktes zweier Vektoren charakterisiert den Winkel zwischen diesen Vektoren und ist abstrakt durch folgende Eigenschaften gekennzeichnet.

Definition 6.71. *Es sei V ein Vektorraum über $\mathbb{R}$. Ein* Skalarprodukt *(inneres Produkt) auf V ist eine Abbildung*

$$V \times V \to \mathbb{R}, \qquad (x, y) \mapsto \langle x, y\rangle \tag{6.74}$$

mit den Eigenschaften

(i) Linearität: Für jedes $x \in V$ sind die Abbildungen

$$\langle x, \cdot\rangle : V \to \mathbb{R}, \qquad v \mapsto \langle x, v\rangle$$

linear.

(ii) Symmetrie: Für alle $x, y \in V$ gilt: $\langle x, y\rangle = \langle y, x\rangle$.

(iii) Positive Definitheit: Für alle $x \neq 0 \in V$ gilt: $\langle x, x\rangle > 0$.

Aus Eigenschaft (i) und (ii) folgt damit auch die Linearität im ersten Argument und das skalare Produkt ist in diesem Fall eine **Bilinearform**.

Erweiterung: In einem komplexen Vektorraum über $\mathbb{C}$ ist ein Skalarprodukt $\langle\ \rangle : V \times V \to \mathbb{C}$ wie folgt charakterisiert:

(i) Für alle $x \in V$ gilt: $\langle x, x\rangle \geq 0$ und $\langle x, x\rangle = 0$ genau dann, wenn $x = 0$.

(ii) Für alle $x, y, z \in V$ und $\lambda \in \mathbb{C}$ gilt

$$\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$$

und

$$\langle x, \lambda y\rangle = \lambda\langle x, y\rangle. \tag{6.75}$$

(iii) Für alle $x, y \in V$ gilt: $\langle x, y\rangle = \overline{\langle y, x\rangle}$.

Aus (iii) folgt: $\langle \lambda x, y\rangle = \bar{\lambda}\langle x, y\rangle$. Man nennt diese Eigenschaft *konjugiert linear* und spricht von einer **Sesquilinearform**.

Statt (6.75) ist oft auch noch die Konvention $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ üblich.

Definition 6.72. *Unter einem* euklidischen Vektorraum *versteht man ein Paar $(V, \langle\cdot,\cdot\rangle)$ bestehend aus einem reellen Vektorraum und einem Skalarprodukt $\langle\cdot,\cdot\rangle$ auf V.*
Einen komplexen Vektorraum über $\mathbb{C}$ mit Skalarprodukt $\langle\cdot,\cdot\rangle$ nennt man unitären Vektorraum.

Beispiele:

(1) $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,
$(x, y) \mapsto x_1 y_1 + \ldots + x_n y_n$.

(2) $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$,
$(x, y) \mapsto \bar{x}_1 y_1 + \ldots + \bar{x}_n y_n$.

(3) Der Vektorraum aller reellen Funktionen auf $I \subset \mathbb{R}$. Teilräume davon sind die reellen stetigen Funktionen auf dem Intervall $[-1, 1]$ oder die Polynome auf dem Intervall $[-1, 1]$ vom Grad $\leq n$. Diese kann man mit folgendem Skalarprodukt versehen $\langle f, g \rangle := \int_{-1}^{1} f(x) g(x)\, dx$.

Ein Skalarprodukt kann man auch dazu verwenden um abstrakte Längen in einem Vektorraum zu definieren.

Definition 6.73. *Ist $(V, \langle \cdot, \cdot \rangle)$ ein euklidischer oder unitärer Vektorraum und $x \in V$, so bezeichnet die* euklidische Norm $|| \; ||$ *von x die reelle Funktion*

$$V \to \mathbb{R}_0^+, \qquad x \mapsto ||x|| = \sqrt{\langle x, x \rangle}. \tag{6.76}$$

Anmerkung: Die euklidische Norm wird meist mit einem Index 2 gekennzeichnet, d. h. man schreibt $|| \cdot ||_2$.

Für unsere Beispiele ergibt sich

(1) $|| \cdot ||_2 : \mathbb{R}^n \to \mathbb{R}, \qquad x \mapsto \sqrt{x_1^2 + \ldots + x_n^2}$.

(2) Für Beispiel 3 ergibt sich $||f||_2 := \big(\int_{-1}^{1} f(x)^2\, dx \big)^{\frac{1}{2}}$.

Satz 6.74. (Cauchy-Schwarzsche Ungleichung). *Für alle $x, y \in V$ gilt*

$$|\langle x, y \rangle| \leq ||x||_2 \, ||y||_2. \tag{6.77}$$

Beweis: [**66**] Kapitel 8.

Satz 6.75. *Die Norm erfüllt folgende Eigenschaften*

(i) $||x|| \geq 0$ *für alle* $x \in V$.

(ii) $||x|| = 0$ *genau dann, wenn* $x = 0$.

(iii) $||\lambda x|| = |\lambda| \, ||x||$ *für alle* $x \in V$ *und* $\lambda \in \mathbb{R}$.

(iv) $||x + y|| \leq ||x|| + ||y||$ *für alle* $x, y \in V$ *(Dreiecksungleichung).*

Beweis: [**66**] Kapitel 8.

Allgemein nennt man einen beliebigen Vektorraum auf dem eine Funktion mit den Eigenschaften (i) bis (iv) (Satz 6.75) definiert ist, einen *normierten Vektorraum*.

Definition 6.76. *Es sei V ein Vektorraum über $\mathbb{R}$ oder $\mathbb{C}$. Eine* Norm *(„Länge“) auf V ist eine Abbildung mit folgenden Eigenschaften*

(i) $||x|| \geq 0$ *für alle* $x \in V$.

(ii) $||x|| = 0$ *genau dann, wenn* $x = 0$.

(iii) $||\lambda x|| = |\lambda|\,||x||$ *für alle* $x \in V$ *und* $\lambda \in \mathbb{R}$ *oder* $\mathbb{C}$.

(iv) $||x + y|| \leq ||x|| + ||y||$ *für alle* $x, y \in V$ *(Dreiecksungleichung).*

Beispiele für Normen auf $\mathbb{R}^n$ die nicht durch ein Skalarprodukt erzeugt werden:

$$||x||_1 := |x_1| + |x_2| + \cdots |x_n|,$$
$$||x||_\infty := \max |x_j|, 1 \leq j \leq n,$$
$$||x||_p := (|x_1|^p + |x_2|^p + \cdots |x_n|^p)^{1/p}\,,\ p \in \mathbb{Q},\ p > 1.$$

Mit Hilfe von Normen kann in Vekorräumen ein **Abstand** definiert werden $d(x, y) := ||x - y||$. Der Abstandsbegriff bildet die Grundlage für den Konvergenzbegriff der Analysis.

Die Cauchy-Schwarzsche Ungleichung erlaubt auch die Definition eines Öffnungswinkels für beliebige euklidische Vektorräume.

Definition 6.77. *Der* Öffnungswinkel *zwischen zwei Vektoren eines euklidischen Vektorraums kann definiert werden als*

$$\cos(\varphi(x, y)) = \frac{\langle x, y\rangle}{||x||\,||y||}, \qquad 0 \leq \varphi(x, y) \leq \pi. \tag{6.78}$$

6.14.1. Orthogonale Vektoren.

Definition 6.78. *Zwei Vektoren v, w eines euklidischen oder unitären Vektoraums heissen* orthogonal $v \perp w$, *wenn* $\langle v, w\rangle = 0$ *ist.*

Beispiel: Man untersuche die Vektoren $(1, x, x^2)$ des Vektorraums der Polynome $P_2[-1, 1]$ auf dem Intervall $[-1, 1]$ auf gegenseitige Orthogonalität.

Definition 6.79. *Ist $M \subset V$, so heisst $M^\perp := \{v \in V \mid v \perp u$ für alle $u \in M\}$ das* orthogonale Komplement *von M.*

$M^\perp$ ist ein Untervektorraum. Es gilt $M^{\perp\perp} = L(M)$ ist ein Untervektorraum mit $M \subset M^{\perp\perp}$.

Definition 6.80. *Ein r-tupel $(v_1, \ldots, v_r)$ von Vektoren in einem euklidischen Vektorraum heisst* orthonormal (Orthonormalsystem)*, wenn $||v_j|| = 1, 1 \leq j \leq r$ und $v_j \perp v_k$ für $j \neq k$ gilt, d. h. $\langle v_j, v_k \rangle = \delta_{jk}$.*

Ein vollständiges Orthonormalsystem nennt man Orthonormalbasis*.*

Satz 6.81. *Ein Orthonormalsystem ist stets linear unabhängig.*

Beweis: Aus

$$\lambda_1 v_1 + \ldots + \lambda_r v_r = 0$$

folgt nach skalarer Multiplikation mit v_j ,

$$\langle v_j, \sum_{k=1}^{r} \lambda_k v_k \rangle = \lambda_j \langle v_j, v_j \rangle = \lambda_j = 0, \qquad 0 \leq j \leq r.$$

Satz 6.82. *Ist $(v_1, \ldots, v_n)$ eine Orthonormalbasis so gilt für alle $v \in V$*

$$v = \sum_{j=1}^{n} \langle v_j, v \rangle \, v_j. \tag{6.79}$$

Beweis: Jeder Vektor v kann als Linearkombination geschrieben werden.

$$v = \sum_{k=1}^{n} \lambda_k v_k.$$

Skalare Multiplikation von v mit v_j in der Form $\langle v_j, \cdot \rangle$[8] ergibt $\lambda_j = \langle v_j, v \rangle$ und damit das Gewünschte. □

Das bedeutet, dass man zur Berechnung der λ_k kein Gleichungssystem zu lösen braucht, sondern diese einfach mit dem skalaren Produkt berechnen kann.

6.14.2. Projektoren.

Mit Hilfe des skalaren Produkte kann ein Projektor, der auf einen eindimensionalen Teilraum projiziert, der den Vektor $\vec{a}$ enthält, konstruiert werden

$$P_{\vec{a}}(.) = \langle \quad . \quad , \frac{\vec{a}}{||\vec{a}||} \rangle \, \frac{\vec{a}}{||\vec{a}||}. \tag{6.80}$$

[8] Mit $\langle \cdot, v_j \rangle$ würde man $\bar{\lambda}_j = \langle v, v_j \rangle$ bekommen.

Die Verallgemeinerung auf Projektoren, die auf einen Teilraum M mit Orthonormalbasis $(v_1, \ldots, v_r)$, $||v_j|| = 1$ projizieren, ist

$$P_M(.) = \sum_{j=1}^{r} \langle \ . \ , v_j \rangle \ v_j. \tag{6.81}$$

Für Projektoren gilt: $P^2 = P$.

Mit Hilfe des **Schmidt'schen Orthogonalisierungsverfahren** kann man aus jeder Basis eine Orthonormalbasis konstruieren. Sei $(w_1, \ldots, w_n)$ eine beliebige Basis in einem euklidischen Vektorraum. Dann ist $(v_1, \ldots, v_n)$ ein Orthonormalsystem, wobei diese Vektoren wie folgt definiert sind

$$v_1 = \frac{w_1}{||w_1||}, \qquad v_j = \frac{w_j - \sum_{k=1}^{j-1} \langle w_j, v_k \rangle \, v_k}{\left|\left| w_j - \sum_{k=1}^{j-1} \langle w_j, v_k \rangle \, v_k \right|\right|}. \tag{6.82}$$

Beispiel: Man konstruiere mit Hilfe der kanonischen Basis $(1, x, x^2)$ eine orthonormale Basis für den Vektorraum der Polynome $P_2[-1, 1]$ auf dem Intervall $[-1, 1]$, $\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx$.

6.14.3. Basistransformationen.

Wir wollen nun überlegen wie man die Koordinaten eines Vektors x von einer Basis $(w_1, \ldots, w_n)$ in eine neue Basis (z.B. Orthonormalbasis) $(v_1, \ldots, v_n)$ umrechnet. Wir betrachten zuerst ein Beispiel in $V = \mathbb{R}^2$. Es gilt

$$x = x_1 w_1 + x_2 w_2 \quad \text{und} \quad x = y_1 v_1 + y_2 v_2$$

wobei (x_1, x_2) die Koordinaten des Vektors x in der Basis (w_1, w_2) und (y_1, y_2) die Koordinaten des Vektors x in der Basis (v_1, v_2) sind. Nun werden die Vektoren der 2. Basis in der 1. Basis dargestellt. Also

$$v_1 = s_{11} w_1 + s_{21} w_2 = \begin{pmatrix} s_{11} \\ s_{21} \end{pmatrix} \quad \text{und} \quad v_2 = s_{12} w_1 + s_{22} w_2 = \begin{pmatrix} s_{12} \\ s_{22} \end{pmatrix}.$$

Es folgt daher

$$x_1 w_1 + x_2 w_2 = x = y_1 v_1 + y_2 v_2 = y_1(s_{11} w_1 + s_{21} w_2) + y_2(s_{12} w_1 + s_{22} w_2)$$

und Koeffizientenvergleich von (w_1, w_2) führt zu

$$\begin{aligned} x_1 &= s_{11} y_1 + s_{12} y_2 \\ x_2 &= s_{12} y_1 + s_{22} y_2 \end{aligned} \qquad \text{bzw. in Matrixdarstellung} \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

und

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad S := \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix}.$$

In den Spalten der Matrix S stehen die Koordinaten der Vektoren (v_1, v_2) dargestellt in Basis (w_1, w_2). Die Verallgemeinerung führt zu folgendem Satz.

Satz 6.83. *Sind $(x_1, \ldots, x_n)$ die Koordinaten des Vektors x in der Basis $(w_1, \ldots, w_n)$ so erhält man die Koordinaten des Vektors x in der Basis $(v_1, \ldots, v_n)$*

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = S^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \tag{6.83}$$

wobei in den Spalten der Matrix S die Koordinaten der Vektoren $(v_1, \ldots, v_n)$ dargestellt in Basis $(w_1, \ldots, w_n)$ stehen.

Analoge Überlegungen führen zu

Satz 6.84. *Ist eine Matrix $A = (a_{jk})$ in der Basis $(w_1, \ldots, w_n)$ gegeben, so erhält man diese Matrix $B = (b_{jk})$ in der Basis $(v_1, \ldots, v_n)$ durch*

$$B = S^{-1}AS. \tag{6.84}$$

Hinweis: $S^{-1}Ax = S^{-1}A(SS^{-1})x = (S^{-1}AS)S^{-1}x = Bx$.

Anmerkung: die Basisvektoren selbst transformieren sich wie

$$\begin{pmatrix} \vec{v}_1 \\ \vdots \\ \vec{v}_n \end{pmatrix} = S \begin{pmatrix} \vec{w}_1 \\ \vdots \\ \vec{w}_n \end{pmatrix}.$$

6.14.4. Die adjungierte Abbildung.

Sei $A \in M(n \times m, \mathbb{C})$, dann heisst $\bar{A}$ die *komplex konjugierte Matrix*, wobei jedes Element a_{jk} komplex konjugiert wird.

Definition 6.85. *Seien $(V, \langle , \rangle)$ und $(V, \langle , \rangle)$ zwei endlichdimensionale Vektorräume mit Skalarprodukt. Dann kann jeder linearen Abbildung $f : V \to W$ eine* adjungierte lineare Abbildung *f^* folgendermaßen zugeordnet werden:*

$$\langle x, f(y)\rangle_W = \langle f^*(x), y\rangle_V, \ y \in V, x \in W.$$

Für Matrizen $A: \mathbb{K}^n \to \mathbb{K}^m$) erhält man die **adjungierte** *Matrix A^**

$$\langle x, Ay\rangle_m = \langle A^*x, y\rangle_n \quad \textit{für alle} \quad y \in \mathbb{K}^n, x \in \mathbb{K}^m.$$

Satz 6.86. *Für Matrizen gilt $A^* = \bar{A}^T$, wenn $\mathbb{K} = \mathbb{C}$ bzw. $A^* = A^T$, wenn $\mathbb{K} = \mathbb{R}$.*

Beweis: Eine lineare Abbildung ist durch die Wirkung auf eine Basis vollständig bestimmt. Wir betrachten daher die Skalarprodukte mit den Einheitsvektoren e_j, e_k und erhalten

$$\overline{\langle e_k, Ae_j\rangle} = \overline{\langle A^* e_k, e_j\rangle},$$
$$\bar{a}_{kj} = \overline{\langle e_k, A_j\rangle} = \overline{\langle A_k^*, e_j\rangle} = \bar{\bar{a}}_{jk}^* = a_{jk}^*.$$

Dabei bezeichnet A_j die j-te Spalte von A und A_k^* die k-te Spalte von A^*.

Ist $A^* = A$, so heisst die Matrix A selbstadjungiert (symmetrisch, hermitisch), ist $A^* = -A$, so heisst die Matrix A schief-symmetrisch. Ist A schief-symmetrisch, so ist dann i A symmetrisch.

Satz 6.87. *Es gilt:* $\mathrm{Kern}(A^*) = \mathrm{Bild}(A)^\perp$,

da für $x \neq 0 \neq y$ und $x \in \mathrm{Kern}(A^*)$ folgt $\langle x, Ay\rangle = \langle A^* x, y\rangle = 0$. D.h. A^* bildet Vektoren die orthogonal zum Bildraum von A stehen in den Nullvektor ab. Daraus folgt

Satz 6.88. *Der Zeilenrang ist gleich dem Spaltenrang.*

Beweis: Es seien V und W zwei endlichdimensionale Vektorräume ($\dim(\mathrm{V}) = n$, $\dim(W) = m$) mit Skalarprodukt und $f : V \to W$ eine lineare Abbildung. Dann gilt

$$\begin{aligned} V &= \mathrm{Kern}\, f \oplus (\mathrm{Kern}\, f)^\perp, \\ n = \dim(V) &= \dim(\mathrm{Kern}\, f) + \dim((\mathrm{Kern}\, f)^\perp), \\ n &= \dim(\mathrm{Kern}\, f) + \dim(\mathrm{Bild}\, f)). \end{aligned}$$

Daher gilt $\dim((\mathrm{Kern}\, f)^\perp) = \dim(\mathrm{Bild}\, f))$.

Sei $0 \neq v \in \mathrm{Kern}\, f$ und $0 \neq w \notin \mathrm{Kern}\, f^*$ dann gilt

$$0 = \langle w, f(v)\rangle = \langle f^*(w), v\rangle.$$

Das bedeutet $f^*(w) \perp v$ oder $f^*(w) \in (\mathrm{Kern}\, f)^\perp$ und somit folgt

$$\mathrm{Bild}\, f^* \subset (\mathrm{Kern}\, f)^\perp$$
$$\dim(\mathrm{Bild}\, f^*)) \leq \dim((\mathrm{Kern}\, f)^\perp) = \dim(\mathrm{Bild}\, f))$$

Es gilt daher $\dim(\mathrm{Bild}\, f^*)) \leq \dim(\mathrm{Bild}\, f))$. Vertauschung der Rollen von f^* und f liefert $\dim(\mathrm{Bild}\, f)) \leq \dim(\mathrm{Bild}\, f^*))$ und somit folgt

$$\mathrm{Rang}\, f = \mathrm{Rang}\, f^*. \tag{6.85}$$

Für Matrizen heisst dies, dass Zeilenrang = Spaltenrang gilt, siehe Abbildung 6.5.

Des weiteren folgt daraus sogar $\mathrm{Bild}\, f^* = (\mathrm{Kern}\, f)^\perp$ und $\mathrm{Bild}\, f = (\mathrm{Kern}\, f^*)^\perp$.

In endlichdimensionalen Vektorräumen gilt: $f^{**} = f$ und Untervektorräume sind abgeschlossen, z.B. $\big((\operatorname{Kern} f)^\perp\big)^\perp = \operatorname{Kern} f$.

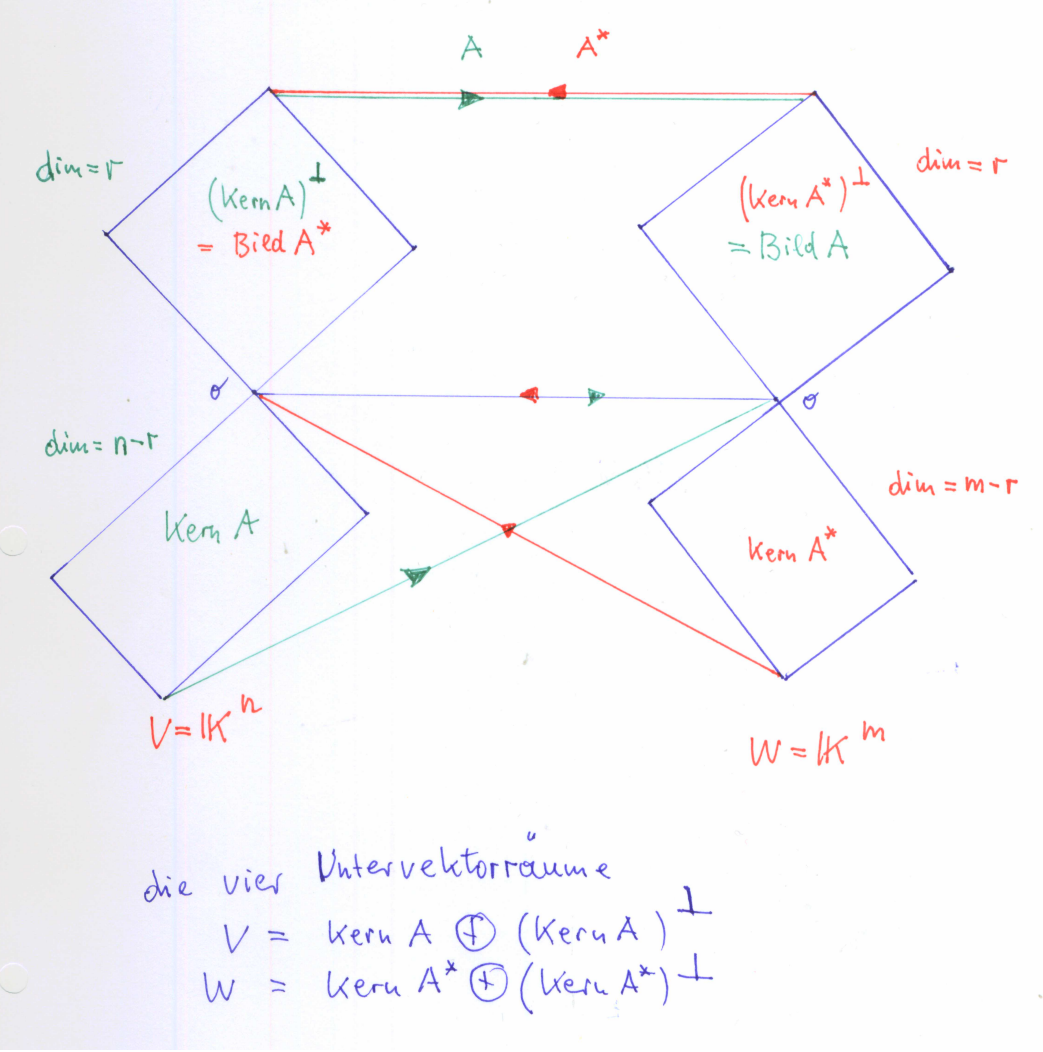


Abbildung 6.5. Schematische Darstellung der 4 Räume: Spaltenraum von A = Bildraum von A = $(\operatorname{Kern} A^*)^\perp$, Zeilenraum von A = Bildraum von A^* = $(\operatorname{Kern} A)^\perp$.

6.14.5. Orthogonale Matrizen.

Definition 6.89. *Seien V, W euklidische Vektorräume. Eine lineare Abbildung $f : V \to W$ heisst* orthogonal *(oder* isometrisch*), wenn*

$$\langle f(v), f(w)\rangle_W = \langle v, w\rangle_V \tag{6.86}$$

für alle $v, w \in V$ gilt.

Eine orthogonale Abbildung ist stets injektiv, denn für $v \in \operatorname{Kern} f$ gilt:
$\langle 0, 0\rangle = \langle f(v), f(v)\rangle = \langle v, v\rangle$.

Ist $V = W$, so sind daher die orthogonlen Abbildungen stets Isomorphismen, die mit $O(V)$ bezeichnet werden. Ist $V = \mathbb{R}^n$ schreibt man auch kurz $O(n)$. Man nennt $O(n) \subset M(n \times n, \mathbb{R})$ die Menge der orthogonalen Matrizen A

$$\langle Av, Aw\rangle = \langle v, w\rangle. \tag{6.87}$$

Satz 6.90. *Eine Matrix $A \in M(n \times n, \mathbb{R})$ ist genau dann orthogonal, wenn die Spalten (= Bilder der Einheitsvektoren) ein orthonormales System bezüglich des Skalarproduktes bilden, d. h. $A^T A = E$.*

Es gilt:

$$\begin{aligned} &\text{(i)}\ A^{-1} = A^T. \\ &\text{(ii)}\ \det A = \pm 1. \end{aligned} \tag{6.88}$$

Definition 6.91. *Die Menge der orthogonalen $n \times n$-Matrizen A, (d h. es gilt $\det A = \pm 1$), bezeichnet man mit $O(n)$. Die spezielle Menge der Matrizen mit $\det A = +1$ wird mit $SO(n)$ bezeichnet. $SO(n) \subset O(n)$.*

Mit der Matrizenmultiplikation bilden $O(n)$ und $SO(n)$ nichtkommutative Gruppen.

Anwendungen von orthogonalen Matrizen in der Computergraphik werden im Buch von Strang [**76**], Kapitel 8.5, Seite 457ff beschrieben. Dazu geht man vom Vektorraum $\mathbb{R}^2$, bzw. $\mathbb{R}^3$ auf die assozierten projektiven Räume über (homogene Koordinaten).

Die komplexwertigen Matrizen U, die das Skalarprodukt in $\mathbb{C}^n$ invariant lassen heissen unitäre Matrizen. Sie werden mit $U(n)$ bezeichnet. Aus $\langle Ux, Uy\rangle = \langle x, y\rangle$, folgt $UU^* = E$, das bedeutet $U^{-1} = U^*$. Es gilt $|\det(U)| = 1$ und die "Eigenwerte" (siehe folgendes Kapitel) liegen daher am Einheitskreis. Die Menge der unitären Matrizen mit $\det(U) = 1$ wird mit $SU(n)$ bezeichnet.

Anwendung: (*) Eine für die Praxis besonders wichtige orthogonale Matrix C ist durch die Koeffizienten

$$c_{j,k} = \sqrt{\frac{2 - \delta_{k,1}}{n}} \cos\left(\frac{(2j-1)(k-1)\pi}{2n}\right), \quad \delta_{k,j} = \begin{cases} 1 & k = j \\ 0 & k \neq j \end{cases}$$

gegeben. Sie ist als **diskrete Kosinustransformation** (DCT) bekannt. Zu jedem Vektor $\mathbf{x}$ bestimmt man den Koeffizientenvektor $\mathbf{y}$

$$\mathbf{y} = C^T \mathbf{x}$$

und aus dem Koeffizientenvektor $\mathbf{y}$ kann der Originalvektor jederzeit mit

$$\mathbf{x} = C\mathbf{y}$$

zurückerhalten werden. Bei praktischen Anwendungen ist $\mathbf{x}$ zum Beispiel ein Vektor von Signalwerten (Daten) und bereits die Projektion auf die ersten $m < n$ Basisvektoren gibt eine gute Approximation des Originalvektors, die für viele Fälle ausreichend ist. Man ersetzt also die n Komponenten des Originalvektors durch m Entwicklungskoeffizienten und erreicht dadurch eine Datenreduktion. Dies ist die Grundlage des JPEG-Verfahrens.

Man zeige, dass die diskrete Kosinustransformation für $n = 2$

$$C = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

eine orthogonale Transformation ist. Berechne C für $n = 3$.

Ergänzung: (*) Wir haben mit $\mathrm{Hom}(V, W)$ die Menge der linearen Abbildungen von V nach W bezeichnet. Diese bilden auf natürliche Weise wieder einen Vektorraum. Ist speziell $W = \mathbb{K}$, so nennt man diesen Raum den **Dualraum** und bezeichnet ihn mit V^*, $V^* = \mathrm{Hom}(V, \mathbb{K})$. Seine Elemente heissen **lineare Funktionale**. Hat V die Dimension n (ist also endlichdimensional), so hat auch V^* die Dimension n.

In einem euklidischen (unitären) Vektorraum können die linearen Funktionale (das sind die Elemente von V^*) mit Hilfe des skalaren Produkts mit den Elementen von V auf natürliche Weise identfiziert werden: $v \mapsto \langle v, \cdot \rangle$. (siehe [**67**], Seite 47 ff, 160 ff und 264 ff).

6.15. Eigenwerte und Eigenvektoren

Im Folgenden benötigen wir vorerst kein Skalarprodukt, aber dafür ist $\mathbb{K} = \mathbb{C}$ im allgemeinen nun unvermeidbar.[9]

Definition 6.92. *Es sei V ein Vektorraum über $\mathbb{K}$ und $f : V \to V$ eine lineare Abbildung in V. Unter einem* Eigenvektor *von f zum* Eigenwert $\lambda \in \mathbb{K}$ *versteht man einen Vektor $v \neq 0$ aus V mit $f(v) = \lambda v$.*

Anmerkung: Ist v Eigenvektor, so ist auch der Vektor $av, a \in \mathbb{K}$ ein Eigenvektor, Eigenvektoren sind also nur bis auf ein Vielfaches (bzw. ihre Länge, wenn eine Norm existiert) festgelegt.

Satz 6.93. *Ist A die Matrix von $f : V \to V$ bezüglich der Basis $(v_1, \ldots, v_n)$ von V, so hat A genau dann* Diagonalgestalt *(d.h. nur die Koeffizienten der Hauptdiagonale von A sind ungleich Null)*

$$A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \tag{6.89}$$

wenn v_j Eigenvektor zum Eigenwert λ_j für $j = 1, \ldots, n$ ist.

Definition 6.94. *Lineare Abbildungen für die eine Basis aus Eigenvektoren existiert, heissen* diagonalisierbar.

Satz 6.95. *Ein Vektor $v \in V \setminus \{0\}$ ist genau dann Eigenvektor der Abbildung $f : V \to V$ zum Eigenwert $\lambda \in \mathbb{K}$, wenn $v \in \operatorname{Kern}(f - \lambda \operatorname{Id})$ ist.*

Definition 6.96. *Ist λ ein Eigenwert von f, so heisst der Untervektorraum*

$$E_\lambda := \operatorname{Kern}(f - \lambda \operatorname{Id}) \tag{6.90}$$

von V der Eigenraum *von f zum Eigenwert λ, und seine Dimension heisst die* geometrische Vielfachheit *des Eigenwertes.*

Satz 6.97. *Sind $(v_1, \ldots, v_r)$ Eigenvektoren von f zu den Eigenwerten $\lambda_1, \ldots, \lambda_r$, und gilt $\lambda_j \neq \lambda_k$ für $j \neq k$ so ist $(v_1, \ldots, v_r)$ linear unabhängig.*

D.h. Eigenvektoren zu verschiedenen Eigenwerten sind linear unabhängig.

Satz 6.98. *Ist A die der Abbildung f bezüglich einer Basis zugeordnete Matrix, so ist λ genau dann ein Eigenwert, wenn* $\det(A - \lambda E) = 0$ *gilt.*

[9] Bleibt man bei $\mathbb{K} = \mathbb{R}$, dann gibt es möglicherweise keine Eigenwerte und Eigenvektoren.

Die Berechnung dieser Determinate liefert das *charakteristische Polynom* von f

$$(-1)^n\lambda^n + b_{n-1}\lambda^{n-1} + \cdots + b_1\lambda + b_0, \quad b_j \in \mathbb{K},\ 0 \le j \le n-1. \tag{6.91}$$

Das charakteristische Polynom ist unabhängig von der Basis, d.h. für jede Basis gleich. Die Eigenwerte λ_j sind die Nullstellen (Lösungen) des charakteristischen Polynoms. Die Vielfachheit der Nullstelle λ_j heisst die *algebraische Vielfachheit* des Eigenwertes λ_j . Diese kann sich von der geometrische Vielfachheit unterscheiden.

Anmerkung: Die Eigenwerte müssen aber nicht notwendigerweise reelle Zahlen sein!

Man beachte: Das Gleichungssystem für die Bestimmung der Eigenvektoren **ist nie** eindeutig lösbar!

Beispiel 1: Man berechne die Eigenvektoren und Eigenwerte der Matrix

$$A = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 2 & 3 \\ 2 & 1 & 2 \end{pmatrix}.$$

Lösung: Eigenwerte: $-1, 1, 5$, Eigenvektoren: $\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

Beispiel 2: Es sei V der Vektorraum der zweimal differenzierbaren Funktionen auf $\mathbb{R}$. Man bestimme alle Eigenwerte und Eigenfunktionen der linearen Abbildung $f : V \to V,\ a(x) \mapsto -\frac{d^2a(x)}{dx^2}$.

Satz 6.99. *(Cayley-Hamilton) Die lineare Abbildung f (bzw. die Matrix A) erfüllt das charakteristische Polynom, d.h. genügt der Gleichung*

$$(-1)^n f^n + b_{n-1}f^{n-1} + \cdots + b_1 f + b_0 = 0, \ \textit{bzw.}$$
$$(-1)^n A^n + b_{n-1}A^{n-1} + \cdots + b_1 A + b_0 = 0, \quad b_j \in \mathbb{K},\ 0 \le j \le n-1. \tag{6.92}$$

Die Summe der Hauptdiagonalelemente einer $n \times n$-Matrix (a_{jk}) nennt man die **Spur der Matrix** (Trace)

$$\operatorname{tr}(A) := \sum_{j=1}^{n} a_{jj} = \sum_{j=1}^{n} \lambda_j. \tag{6.93}$$

Sie ist gleich der Summe der Eigenwerte, da sie eine Invariante bezüglich Basiswechsel ist. Weiters gilt: $b_{n-1} = (-1)^{n-1}\operatorname{tr}(A)$ und $b_0 = \det(A)$.

Satz 6.100. *Die Determinante ist das Produkt aller Eigenwerte (mit Vielfachheit)*

$$\det(A) = \lambda_1\lambda_2\cdots\lambda_n.$$

Zu jedem Eigenwert gibt es mindestens einen zugehörigen Eigenvektor, höchstens aber so viele linear unabhängige wie es der algebraischen Vielfachheit der Nullstelle entspricht. n linear unabhängige Eigenvektoren gibt es also genau dann, wenn es zu jedem Eigenwert so viele linear unabhängige Eigenvektoren gibt, wie es seiner algebraischen Vielfachheit entspricht. Man erhält also

Satz 6.101. *Eine Matrix A ist genau dann diagonalisierbar, wenn es zu jedem Eigenwert genau so viele linear unabhängige Eigenvektoren gibt wie es seiner algebraischen Vielfachheit als Nullstelle des charakteristischen Polynoms entspricht.*

Es gilt dann für die Diagonalmatrix $D = U^{-1}AU$, wobei die Spalten von U aus einer Basis von Eigenvektoren besteht.

Anmerkung: Ist eine Matrix nicht diagonalisierbar, so kann man sie zumindest auf obere Dreiecksform bringen (**Jordansche Normalform**), so dass auf der Hauptdiagonale die Eigenwerte stehen. Die Elemente direkt oberhalb der Hauptdiagonale sind entweder 0 oder 1 und alle anderen Elemente sind 0. Details siehe [**66**], Seite 234

6.15.1. Anwendungsbeispiel: Markovprozess. Ein Geschäft mit zwei Filialen verleiht tageweise Fahrräder. 60% der Fahrräder die in der ersten Filiale ausgeliehen werden, werden auch dort zurückgegeben; der Rest in der anderen Filiale. 70% der Fahrräder die in der zweiten Filiale ausgeliehen werden, werden auch dort zurückgegeben; der Rest wiederum in der anderen Filiale.

Ist es möglich die Fahrräder so auf beide Filialen zu verteilen, dass in jeder Filiale an jedem Morgen genau die gleiche Anzahl von Fahrrädern steht? Wenn man die Fahrräder irgendwie auf beide Filialen verteilt, was passiert dann im Laufe der Zeit?

Lösung: Bezeichnet man die Anzahl der Fahrräder in den beiden Filialen am n-ten Tag mit $x_1(n)$ und $x_2(n)$, so gilt nach Voraussetzung

$$\mathbf{x}(n+1) = A\mathbf{x}(n), \qquad A = \begin{pmatrix} 0.6 & 0.3 \\ 0.4 & 0.7 \end{pmatrix}.$$

Für die gewünschte Verteilung $\mathbf{x}$ muss

$$\mathbf{x} = A\mathbf{x}$$

gelten, sie muss also ein Eigenvektor zum Eigenwert eins sein. Die charakteristische Gleichung lautet

$$\lambda^2 - 1.3\lambda + 0.3 = 0$$

und die Eigenwerte sind $\lambda_1 = 1$ und $\lambda_2 = 0.3$. Die zugehörigen Eigenvektoren lauten

$$\mathbf{u}_1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Die gesuchte Verteilung ist also $\mathbf{x} = \frac{1}{7}\mathbf{u}_1$ (die Länge wurde hier so gewählt, dass $x_1 + x_2 = 1$) und in der ersten Filiale sollten 43% und in der zweiten 57% der Fahrräder sein.

Wie sieht es nun mit dem Zeitverhalten aus? Ist die Verteilung $\mathbf{x}(0)$ am Anfang gegeben, so ist die Verteilung nach n Tagen

$$\mathbf{x}(n) = A^n \mathbf{x}(0).$$

Das ist zwar eine nette Formel die man mit `Mathematica` für jedes n auswerten kann, aber was im Laufe der Zeit passiert ist daraus nicht ablesbar! Geht man nun zur Basis aus Eigenvektoren über

$$\mathbf{x}(0) = U\mathbf{y} = y_1\mathbf{u}_1 + y_2\mathbf{u}_2,$$

so hat man

$$\mathbf{x}(n) = y_1 A^n \mathbf{u}_1 + y_2 A^n \mathbf{u}_2 = y_1\,\lambda_1^n \mathbf{u}_1 + y_2\,\lambda_2^n \mathbf{u}_2 = y_1\mathbf{u}_1 + y_2\,(0.3)^n \mathbf{u}_2.$$

Die Komponente in $\mathbf{u}_2$ Richtung nimmt also exponentiell ab und die Verteilung konvergiert gegen die Gleichgewichtsverteilung $y_1\mathbf{u}_1$. Mit anderen Worten, vollkommen egal mit welcher Verteilung man startet, im Laufe der Zeit stellt sich die Gleichgewichtsverteilung ein.

Allgemein zeigt letztes Beispiel auch, dass man A^k für eine diagonalisierbare Matrix leicht mittels

$$A^k = U \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix} U^{-1}$$

berechnet werden kann.

Denn für $A = UDU^{-1}$ gilt $A^k = A \cdot A \cdots A = (UDU^{-1})(UDU^{-1}) \cdots (UDU^{-1}) = UD(U^{-1}U)D \cdots (U^{-1}U)DU^{-1} = UD^kU^{-1}$.

Markovprozesse können natürlich auch in höheren Dimensionen betrachtet werden. Die charakteristische Eigenschaft der Matrix A ist dabei, dass alle Koeffizienten nicht negativ und die Spaltensummen immer eins sind. In diesem Fall wird A auch als **Markov-Matrix** bezeichnet.

Satz 6.102. *Eine Markov-Matrix hat immer einen Eigenwert eins und es gibt dazu immer einen Eigenvektor (Gleichgewichtszustand) dessen Koordinaten alle nicht negativ sind.*

Das A immer den Eigenwert eins hat ist leicht zu sehen: Die transponierte Matrix A^T hat Zeilensummen eins und deshalb ist $\mathbf{e} = (1, 1, \ldots, 1)$ ein Eigenvektor zum Eigenwert eins, $A^T\mathbf{e} = \mathbf{e}$. Warum nützt uns das? Weil A und A^T die gleichen Eigenwerte haben, denn es gilt $\det(\lambda\,\mathrm{E} - A^T) = \det((\lambda\,\mathrm{E} - A)^T) = \det(\lambda\,\mathrm{E} - A)$.

Eine beliebige Anfangsverteilung muss aber nicht immer gegen einen Gleichgewichtszustand konvergieren, sie könnte auch hin und her springen wie zum Beispiel bei (diese Matrix hat die Eigenwerte 1 und -1)

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Satz 6.103. *Sei A eine Markov-Matrix. Dann sind alle Eigenwerte vom Betrag kleiner gleich 1. Gibt es ausser 1 keinen Eigenwert mit Betrag gleich 1, so konvergiert für beliebigen Anfangszustand $x(0)$ die Folge von Vektoren $x(n) = A^n x(0)$ gegen einen Gleichgewichtszustand (der vom Anfangszustand abhängen kann).*

Die Bedingung ist zum Beispiel erfüllt, falls alle Diagonalelemente der Matrix A positiv sind ($a_{jj} > 0$). Sind sogar alle Koeffizienten von A positiv ($a_{jk} > 0$), so ist der Gleichgewichtszustand eindeutig.

6.15.2. Anwendungsbeispiel: Bewertung von Webseiten mit *PageRank* ([30]). Die Idee der Markovprozesse wird auch bei der Suchmaschine Google verwendet. Die Idee dahinter wollen wir uns zum Abschluss noch etwas genauer ansehen.

Wir haben also eine Anzahl von n Seiten, die miteinander verlinkt sind, gegeben. Die Links werden durch eine Link-Matrix $L = (l_{jk})$ beschrieben, die eins ist falls ein Link von der j-ten zur k-ten Seite besteht und sonst null. Üblicherweise zählt man Links einer Seite auf sich selbst nicht und setzt daher l_{jj} in jedem Fall null.

Aufgabe einer guten Suchmaschine ist es nicht nur Seiten, die ein bestimmtes Stichwort enthalten, zu finden, sondern auch die Treffer nach bestimmten Kriterien zu sortieren. Dazu ist es notwendig alle Seiten zu bewerten und jeder Seite eine Bewertung x_j zuzuordnen. Die Idee dabei ist es, dass jeder Link auf eine Seite als „Stimme“ für diese Seite zählt. Im einfachsten Fall zählt man also die Anzahl der Seiten die auf die j-te Seite verlinken und definiert

$$\tilde{x}_j = \sum_{k=1}^{n} l_{kj}$$

als deren Bewertung. Der Nachteil ist, dass dabei die Stimme einer Seite mit vielen Links genau so viel zählt, wie die Stimme einer Seite mit wenigen ausgesuchten Links. Deshalb verfeinern wir unseren Ansatz und geben jeder Seite insgesamt nur eine Stimme, die sie gleichmässig auf alle Seiten auf die sie verlinkt verteilt:

$$\hat{x}_j = \sum_{k=1}^{n} \frac{1}{n_k} l_{kj}, \qquad n_k = \sum_{j=1}^{n} l_{kj},$$

wobei n_k die Anzahl der Links auf der k-ten Seite ist. Wir erhalten damit nun eine Matrix $\hat{L}$ mit Zeilensumme 1.

Sollte eine Seite nicht von ihrem Stimmrecht gebrauch machen, sollte also $n_k = 0$ sein, so ist der Term $\frac{1}{n_k} l_{kj}$ durch 0 zu ersetzen.

Das ist schon etwas besser, aber immer noch nicht optimal. Der Betreiber einer Seite könnte einfach eine große Anzahl von weiteren Seiten erstellen, deren einziger Zweck es ist, auf seine eigentliche Seite zu verlinken, um deren Bewertung zu erhöhen. Insbesonders werden große Homepages mit vielen untereinander verlinkten Seiten automatisch besser bewertet als kleinere Homepages. Wir müssen also nochmals nachbessern indem wir nicht jeder Seite genau eine Stimme geben, sondern genau so viel Stimmrecht, wie es ihrer Bewertung entspricht

$$x_j = \sum_{k=1}^{n} \frac{1}{n_k} l_{kj} x_k.$$

Hoppla, werden Sie sich jetzt vielleicht denken, da drehen wir uns jetzt aber im Kreis! Auf der rechten Seite kommen ja wieder die Bewertung x_k vor die wir ja gerade ausrechnen wollen! Stimmt, wir haben eben eine Gleichung

$$\mathbf{x} = A\mathbf{x}, \qquad \text{mit} \quad a_{jk} = \frac{1}{n_k} l_{kj},$$

für die gesuchten Bewertungen $\mathbf{x}$ bekommen. Wir müssen also einen Gleichgewichtszustand der Markov-Matrix A finden. Es ist dabei $A = \hat{L}^t$.

Im Fall $n_k = 0$ haben wir wieder das Problem, dass die k-te Spalte von A gleich null ist, obwohl die Spaltensummen einer Markovmatrix gleich eins sein müssen. Diese Seite hat aber keinen Einfluss auf die Bewertung der anderen Seiten und wir können diese Seite einfach entfernen. Ihre Bewertung kann dann später leicht aus den Bewertungen der anderen Seiten berechnet werden.

Um die gesuchten Bewertungen zu erhalten, müssen wir also einen Eigenvektor von A zum Eigenwert eins finden, also das lineare Gleichungssystem $(A - \mathrm{E})\mathbf{x} = 0$ lösen. Theoretisch ist das kein Problem, aber bei der Anzahl der Seiten im Internet sind mit dieser Aufgabe auch die derzeit schnellsten Computer überfordert. Was also tun? Muss unsere schöne Idee mangels praktischer Durchführbarkeit in den Papierkorb wandern? Nein! Bei einem Markovprozess kann man den Gleichgewichtszustand ja näherungsweise durch die Iteration

$$\mathbf{x}(m+1) = A\mathbf{x}(m)$$

eines Anfangszustandes $\mathbf{x}(0)$ erhalten. Diese Iteration kann (vergleichsweise) schnell berechnet werden, da die meisten Koeffizienten von A ja null sind, und man kann dadurch zu einer, für unsere Zwecke vollkommen ausreichenden, Näherung für den Gleichgewichtszustand gelangen.

Man kann sich die Iteration auch wie eine Anzahl von Zufallssurfern vorstellen. Der Vektor $\mathbf{x}(m)$ gibt an wie viele Surfer sich im m-ten Schritt auf jeder einzelnen Seite befinden. In jedem Schritt sucht sich jeder Surfer zufällig einen Link

auf seiner aktuellen Seite und wechselt auf die nächste Seite. Im Laufe der Zeit stellt sich dabei die Gleichgewichtsverteilung ein.

Wir sind somit fast am Ziel, eine einzige kleine Hürde ist noch zu nehmen: Wir haben ja gerade vorher gelernt, dass die Iteration eines Markovprozesses nicht immer konvergiert. Wenn zwei Seiten nur auf die jeweils andere verlinken so springt die Iteration immer hin und her. Ein Zufallssurfer der in diese Falle tappt wäre also gefangen und auf ewig dazu verdammt zwischen diesen beiden Seiten hin und her zu wechseln. Deshalb führen wir noch einen „Dämpfungsfaktor“ $\alpha \in (0,1)$ ein. Nur der Bruchteil α aller Benutzer wählt aus den Links der aktuellen Seite, der Rest $(1-\alpha)$ sucht sich zufällig irgendeine Seite aus. Im Bild der Bewertung bedeutet, dass das nur der Anteil α über die Bewertung durch andere Seiten und der Rest $(1-\alpha)$ gleichmässig verteilt wird.

Legen wir fest, dass jede Seite im Durchschnitt Bewertung eins hat, also

$$\sum_{j=1}^{n} x_j = n,$$

so erhalten wir folgendes modifiziertes Gleichungssytem

$$\mathbf{x} = (1-\alpha)\mathbf{e} + \alpha A\mathbf{x},$$

wobei $\mathbf{e} = (1,1,\ldots,1)$. In der Praxis wird ein Wert um $\alpha = 0.85$ verwendet. Die Lösung dieses Gleichungssystems ist

$$\mathbf{x} = (1-\alpha)(\mathrm{E} - \alpha A)^{-1}\mathbf{e}$$

und kann mittels Iteration

$$\mathbf{x}(m+1) = (1-\alpha)\mathbf{e} + \alpha A\mathbf{x}(m),$$

bestimmt werden.

Die Eigenwerte von A sind vom Betrag kleiner gleich eins. Also sind die Eigenwerte von αA kleiner gleich $\alpha < 1$. Somit ist eins kein Eigenwert von αA und damit ist $\mathrm{E} - \alpha A$ invertierbar. Um zu verstehen, dass die Iteration konvergiert, berechnen wir

$$\mathbf{x}(m) = (1-\alpha)\sum_{j=0}^{m-1} \alpha^j A^j \mathbf{e} + \alpha^m A^m \mathbf{x}(0).$$

Für die Markov-Matrix A bleiben die Vektoren $A^m\mathbf{x}(0)$ beschränkt (er konvergiert gegen eine Gleichgewichtszustand oder springt hin und her) und der Faktor α^m bewirkt, dass $\alpha^m A^m \mathbf{x}(0) \to 0$. Das gleiche gilt für die Vektoren $A^j\mathbf{e}$ und aus der Konvergenz der geometrischen Reihe $\sum_{j=0}^{\infty} \alpha^j$ folgt die Konvergenz der obigen Reihe.

Dieser Algorithmus zur Webseitenbewertung bildet das Herzstück von Google und ist als PageRank (http://infolab.stanford.edu/~backrub/google.html) bekannt.

6.16. Hauptachsentransformationen

Im Folgenden wird vorausgesetzt, dass die linearen Abbildungen stetig sind. Das ist zum Beispiel bei endlichdimensionalen Vektorräumen immer der Fall.

Definition 6.104. *Es sei $(V, \langle\cdot,\cdot\rangle)$ ein euklidischer (oder unitärer) Vektorraum. Eine lineare Abbildung auch beschränkter Operator genannt (wird bei Funktionenräumen verwendet) $f : V \to V$ heisst* selbstadjungiert, *wenn*

$$\langle f(v), w\rangle = \langle v, f(w)\rangle \tag{6.94}$$

für alle $v, w \in V$ gilt. Das heisst $f^* = f$.

Es gilt

Satz 6.105. *Eigenvektoren v und w eines selbstadjungierten Operators f zu verschiedenen Eigenwerten $\lambda \neq \mu$ stehen senkrecht (orthogonal) zueinander. Die Eigenwerte sind reell (auch für unitäre Vektorräume).*

Beweis: es gilt $\lambda\langle v, v\rangle = \langle v, \lambda v\rangle = \langle v, f(v)\rangle = \langle f(v), v\rangle = \langle \lambda v, v\rangle = \bar{\lambda}\langle v, v\rangle$ und damit $\lambda = \bar{\lambda}$. Aus $\langle f(v), w\rangle = \langle v, f(w)\rangle$ ergibt sich dann $\langle \lambda v, w\rangle = \langle v, \mu w\rangle$ also $(\lambda - \mu)\langle v, w\rangle = 0$.

Satz 6.106. *Ist v Eigenvektor des selbstadjungierten Operators $f : V \to V$, so ist der Unterraum*

$$v^\perp := \{w \in V \mid w \perp v\} \tag{6.95}$$

invariant unter f, d.h. es gilt $f(v^\perp) \subset v^\perp$.

Beweis: Aus $\langle v, w\rangle = 0$ folgt auch $\langle v, f(w)\rangle = \langle f(v), w\rangle = \langle \lambda v, w\rangle = 0$.

Satz 6.107. *Ist $(V, \langle\cdot,\cdot\rangle)$ ein euklidischer (oder unitärer) Vektorraum und $(v_1, \ldots, v_n)$ eine Orthonormalbasis von V, so ist die Matrix A einer linearen Abbildung $f : V \to V$ durch*

$$a_{jk} = \langle v_j, f(v_k)\rangle \tag{6.96}$$

gegeben.

Beweis: Da $f(v_k)$ in V ist, kann dieser Vektor als Linearkombination der Basisvektoren geschrieben werden.

$$e_k \xrightarrow{\Phi_{(v_1,\ldots,v_n)}} f(v_k) = \sum_{j=1}^{n} \lambda_j v_j = \sum_{j=1}^{n} \langle v_j, f(v_k)\rangle v_j \xrightarrow{\Phi^{-1}_{(v_1,\ldots,v_n)}} \begin{pmatrix} \langle v_1, f(v_k)\rangle \\ \langle v_2, f(v_k)\rangle \\ \vdots \\ \langle v_n, f(v_k)\rangle \end{pmatrix}.$$

Andererseits sind die Spalten von A die Bilder der Einheitsvektoren

$$Ae_k = \begin{pmatrix} a_{1,k} \\ a_{2,k} \\ \vdots \\ a_{n,k} \end{pmatrix}.$$

Ein Vergleich liefert die Formel (6.96).

Satz 6.108. *Ist $(v_1, \ldots, v_n)$ eine Orthonormalbasis von V, so ist der Operator $f : V \to V$ genau dann selbstadjungiert, wenn die zugehörige Matrix A symmetrisch ist, d.h. $a_{jk} = a_{kj}$ erfüllt, d.h. $A = A^T$, $\mathbb{K} = \mathbb{R}$. Für $\mathbb{K} = \mathbb{C}$ gilt: $A = \bar{A}^T$ und daher $a_{jk} = \bar{a}_{kj}$.*

Beispiel: die Pauli-Matrix[10] σ_2 ist selbstadjungiert

$$\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{6.97}$$

Satz 6.109. *Ist $(V, \langle \cdot, \cdot \rangle)$ ein endlichdimensionaler euklidischer (unitärer) Vektorraum und $f : V \to V$ eine selbstadjungierte lineare Abbildung, so gibt es eine Orthonormalbasis aus Eigenvektoren von f.*

Satz 6.110. (Hauptachsentransformation) *Ist A eine symmetrische reelle $n \times n$-Matrix, so gibt es eine orthogonale Transformation $U \in O(n)$, sodass $D := U^{-1}AU$ eine Diagonalmatrix mit den Eigenwerten von A in der Diagonalen ist, jeder so oft angeführt, wie seine geometrische Vielfachheit angibt.*

D.h., D hat die Gestalt

$$D = \begin{pmatrix} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_1 & & & & \\ & & & \ddots & & & \\ & & & & \lambda_r & & \\ & & & & & \ddots & \\ & & & & & & \lambda_r \end{pmatrix}. \tag{6.98}$$

Anmerkung: Die orthonormierten Eigenvektoren von A bilden die Spalten der orthogonalen Matrix U. Ist ein Eigenwert λ nicht einfach, muss mit dem Schmidtschen Orthogonalisierungsverfahren eine orthonormale Basis von E_λ konstruiert werden. Es existieren aber auch noch andere (nicht-orthogonale) Matrizen welche diese Matrix diagonalisieren. Allerdings gilt für diese **nicht** $U^{-1} = U^T$.

[10] https://de.wikipedia.org/wiki/Pauli-Matrizen

Beispiel: Man diagonalisiere die Matrix

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 2 \end{pmatrix}.$$

Beispiel: (*) (Quadratische Kurven) Man beschreibe die Kurve im $\mathbb{R}^2$, die durch folgende Gleichung gegeben ist

$$3x_1^2 + 2x_1x_2 + 3x_2^2 = 1.$$

Lösung: Der Trick besteht darin, die Gleichung in der Form

$$\mathbf{x}^T A\,\mathbf{x} = a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2 = 1$$

zu schreiben. Es muss also $a_{11} = 3$, $a_{12} + a_{21} = 2$ und $a_{22} = 3$ gelten. Fordern wir, dass A symmetrisch ist, $a_{12} = a_{21}$, so folgt

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}.$$

Verwenden wir die Eigenvektoren von A als neue Orthonormalbasis $\mathbf{y} = U^{-1}\mathbf{x}$, bzw. $\mathbf{x} = U\mathbf{y}$, so gilt

$$\mathbf{x}^T A\,\mathbf{x} = (U\mathbf{y})^T A(U\mathbf{y}) = \mathbf{y}^T(U^T AU)\mathbf{y} = \mathbf{y}^T(U^{-1}AU)\mathbf{y}$$

da U orthogonal ist. Nach Konstruktion ist $U^{-1}AU$ eine Diagonalmatrix und deshalb ist

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1.$$

Die normierten Eigenvektoren $\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1,1)$, $\mathbf{u}_2 = \frac{1}{\sqrt{2}}(1,-1)$ sind die Hauptachsen und die Eigenwerte $\lambda_1 = 2$, $\lambda_2 = 4$ die Streckungsfaktoren. In den neuen Koordinaten, sehen wir, dass es sich um eine Ellipse $2y_1^2 + 4y_2^2 = 1$ handelt.

Durch Drehung des Koordinatensystem auf die Hauptachsen kann also jede quadratische Kurve auf Normalform gebracht werden. Sind beide Eigenwerte positiv, so handelt es sich um eine Ellipse, ist einer positiv und einer negativ, so handelt es sich um eine Hyperbel.

6.16.1. Anwendung:(*) Die diskrete Kosinustransformation ([30]). Als Abschluss können wir nun die diskrete Kosinustransformation herleiten. Dazu betrachten wir folgende symmetrische $n \times n$ Matrix

$$A = \frac{1}{2}\begin{pmatrix} 0 & 1 & & & & \\ 1 & 0 & 1 & & & \\ & 1 & 0 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 0 & 1 \\ & & & & 1 & 0 \end{pmatrix}$$

Alle Einträge auf der Hauptdiagonale sind 0, auf der Diagonale oberhalb und unterhalb der Hauptdiagonale $\frac{1}{2}$ und alle weiteren sind ebenfalls 0. Nun versucht man die Eigenwerte und Eigenvektoren zu berechnen. Das charakteristische Polynom zu berechnen erscheint auf den ersten Blick aussichtslos und deshalb untersucht man zuerst die Eigenwertgleichung

$$A\mathbf{u} = \lambda\mathbf{u}$$

und ignoriert zunächst die Tatsache, dass man die Eigenwerte noch nicht kennt. Jede Zeile entspricht einem Gleichungssystem der Form

$$\frac{1}{2}u_{j+1} + \frac{1}{2}u_{j-1} = \lambda u_j$$

mit den Ausnahmen $j = 1$ und $j = n$, wo die Terme $\frac{1}{2}u_0$ bzw. $\frac{1}{2}u_{n+1}$ fehlen. Die Koeffizienten des Eigenvektors erfüllen also bis auf die Randpunkte eine lineare Rekursionsrelation zweiter Ordnung von der man weis wie sie zu lösen ist. Um die Probleme an den Randpunkten zu beheben, fordert man einfach die **Randbedingungen**

$$u_0 = u_{n+1} = 0,$$

dann ist eine Lösung der Rekursionsrelation auch gleichzeitig ein Eigenvektor!

Man löst nun die Rekursionsrelation: Der Ansatz $u_j = \mu^j$ liefert die charakteristische Gleichung

$$\mu^2 - 2\lambda\mu + 1 = 0$$

mit den Nullstellen

$$\mu_1 = \lambda + \sqrt{\lambda^2 - 1}, \qquad \mu_2 = \lambda - \sqrt{\lambda^2 - 1}.$$

Man beachte , dass $\mu_1\mu_2 = \lambda^2 - (\sqrt{\lambda^2 - 1})^2 = 1$ ist. Für $\lambda \leq 1$ sind beide Nullstellen konjugiert komplex

$$u_j = k_1 \cos(\omega j) + k_2 \sin(\omega j), \qquad \mu_1 = \mathrm{e}^{\mathrm{i}\omega} = \cos(\omega) + \mathrm{i}\sin(\omega).$$

Aus der ersten Randbedingung $u_0 = 0$ folgt

$$u_j = k \sin(\omega j)$$

und die zweite lautet

$$u_{n+1} = k \sin(\omega(n+1)) = 0.$$

Damit sie erfüllt ist muss

$$\omega(n+1) = \ell\pi, \qquad \ell \in \mathbb{Z},$$

gelten. In diesem Fall ist $\mathbf{u}$ ein Eigenvektor und das zugehörige $\lambda = \frac{\mu_1+\mu_2}{2} = \cos(\omega)$ ein Eigenwert. Rückeinsetzen liefert somit die Eigenwerte und Eigenvektoren

$$\lambda_\ell = \cos(\frac{\pi\ell}{n+1}), \qquad \mathbf{u}_\ell = \sqrt{\frac{2}{n+1}} \left(\sin(\frac{\pi\ell j}{n+1})\right)_{1\leq j\leq n}$$

mit $1 \leq \ell \leq n$, denn $\ell = 0$ liefert nur $\mathbf{u}_0 = \mathbf{0}$ und alle anderen Werte von $\ell \in \mathbb{Z}$ liefern aufgrund der Periodizität der trigonometrischen Funktionen nur wieder Werte die man schon kennt. Da man damit alle n Eigenwerte gefunden hat, braucht der Fall $\lambda > 1$ nicht mehr untersucht werden.

Außerdem wurden die Eigenvektoren auf eins normiert, sie bilden also bereits eine Orthonormalbasis. Um die Normierung einzusehen verwendt man die Formel von Euler $\mathrm{e}^{\mathrm{i}x} = \cos(x) + \mathrm{i}\sin(x)$ und nimmt eine Abkürzung durch die komplexe Ebene: Nimmt man den Imaginärteil der geometrischen Reihe

$$\sum_{j=0}^{n} \mathrm{e}^{\mathrm{i}\frac{2\pi\ell}{n+1}j} = \frac{1 - \mathrm{e}^{\mathrm{i}2\pi\ell}}{1 - \mathrm{e}^{\mathrm{i}\frac{2\pi\ell}{n+1}}} = 0,$$

so folgt wegen $\mathrm{e}^{\mathrm{i}\frac{2\pi\ell j}{n+1}} = (\cos(\frac{\pi\ell j}{n+1}) + \mathrm{i}\sin(\frac{\pi\ell j}{n+1}))^2$

$$\sum_{j=0}^{n} \cos^2(\frac{\pi\ell j}{n+1}) - \sin^2(\frac{\pi\ell j}{n+1}) = 0.$$

Andererseits gilt wegen $\sin^2(x) + \cos^2(x) = 1$

$$\sum_{j=0}^{n} \cos^2(\frac{\pi\ell j}{n+1}) + \sin^2(\frac{\pi\ell j}{n+1}) = n + 1.$$

Die letzten beiden Formeln subtrahiert ergibt das gewünschte Resultat

$$\sum_{j=1}^{n} \sin^2(\frac{\pi\ell j}{n+1}) = \frac{n+1}{2}.$$

Es stellt sich jetzt die Frage, was das alles mit der Kosinustransformation zu tun hat? In der Formel für die Eigenvektoren steht zwar ein Sinus, von einem Kosinus ist aber weit und breit nichts zu sehen! Die wirkliche Kosinustransformation erhält man nämlich wenn man die Matrix

$$A = \frac{1}{2}\begin{pmatrix} 1 & 1 & & & & \\ 1 & 0 & 1 & & & \\ & 1 & 0 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 0 & 1 \\ & & & & 1 & 1 \end{pmatrix}$$

betrachtet. Das macht die Rechnung nur etwas komplizierter (man muss mit den Randbedingungen $u_0 = u_1$ und $u_n = u_{n+1}$ rechnen), ändert aber nichts an der prinzipiellen Idee.

Der Vorteil für die Bildverarbeitung ist, dass bei der letzten Matrix $\mathbf{u} = (1, 1, \ldots, 1)$ ein Eigenvektor ist $A\mathbf{u} = \mathbf{u}$. Bei einem Bild mit konstanter Farbe steckt die ganze Information in dieser Komponente.

6.17. Übung

(1) Man orthonormalisiere die Vektoren

$$\mathbf{a}_1 = \begin{pmatrix}1\\0\\1\end{pmatrix}, \qquad \mathbf{a}_2 = \begin{pmatrix}0\\1\\1\end{pmatrix}, \qquad \mathbf{a}_3 = \begin{pmatrix}1\\2\\5\end{pmatrix}.$$

(2) Man berechne eine orthonormale Basis des Vektorraums der Polynome $P_2[-1,1]$ aus der Basis $(x, x+1, x^2)$ mit dem Skalarprodukt

$$\langle f, g\rangle := \int_{-1}^{1} f(x)g(x)\,dx.$$

(3) Man berechne die Eigenwerte und Eigenvektoren folgender Matrizen:

$$\begin{pmatrix}1&1&3\\0&2&3\\2&1&2\end{pmatrix}, \quad \begin{pmatrix}3&5&3\\0&4&6\\0&0&1\end{pmatrix}, \quad \begin{pmatrix}2&0&-1\\0&1&0\\1&0&4\end{pmatrix}, \quad \begin{pmatrix}s&1&0\\1&s&s\\0&1&s\end{pmatrix}.$$

(4) Projektor:
(i) Man finde die Matrix P die jeden Vektor b des $\mathbb{R}^3$ auf die Gerade mit dem Richtungsvektor $a = (2,1,3)$ projiziert. Was ist die Spur (Summe der Diagonalelemente) von P.
(ii) Man gebe den Bildraum (Spaltenraum) und den Nullraum (Kern) von P an. Geometrisch beschreiben und auch jeweils eine Basis angeben.
(iii) Was sind die Eigenwerte und Eigenvektoren. (Kann auch durch Überlegen statt durch Rechnen gefunden werden!)

(5) Man diagonalisiere die folgenden symmetrischen Matrizen, d. h. man berechne ihre Diagonalform $D := U^T AU$ und gebe auch die orthogonale Matrix U (die Spalten von U sind die orthonormierten Eigenvektoren von A; nur dann gilt $U^{-1} = U^T$) an

$$\begin{pmatrix}2&1&1\\1&2&-1\\1&-1&2\end{pmatrix}, \quad \begin{pmatrix}1&-1&-1\\-1&1&1\\-1&1&1\end{pmatrix}, \quad \begin{pmatrix}s&k&k\\k&s&k\\k&k&s\end{pmatrix}.$$

(6) Für ein Rapid Prototyping-Verfahren wurde folgende Anforderung gestellt: man berechne für ein Objekt einen minimal einhüllenden Quader. Dieses Problem der Berechnung der sogenannten „bounding box“ wurde mit dem PCA-Algorithmus gelöst (z.B. `http://wscg.zcu.cz/wscg2007/Papers_2007/full/C11-full.pdf`) Die Achsen eines damit berechneten Quader seien nun $(1,0,2), (0,4,0), (4,0,-2)$.
Man berechne die Matrix mit der der Quader parallel zu den Achsen $(1,0,0), (0,1,0), (0,0,1)$ gedreht werden kann.

(7) Es sei $(V, \mathbb{C})$ der Vektorraum der periodischen komplexwertigen Funktionen auf dem Interval $[0, 2\pi]$ mit dem Skalarprodukt $< f, g >:= \int_0^{2\pi} \bar{f}(x) g(x)\, dx$. Man zeige, dass die Funktionen

$$a_k(x) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{\mathrm{i}\, k\, x}, k \in \mathbb{Z}$$

ein Orthonormalsystem bilden.

(8) Gegeben sei die Matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Man berechne $\exp(A\, t)$. Hinweis: man diagonalisiere zuerst A.

(9) (*) (Spektralsatz). Wir haben mit E_λ den Kern von $f - \lambda\, \mathrm{Id}$ bezeichnet. Beweise: Sei V ein endlichdimensionaler euklidischer Vektorraum und $f : V \to V$ eine selbstadjungierte lineare Abbildung. Sind alle Eigenwerte $\lambda_1, \ldots, \lambda_n$ verschieden und bezeichnet P_k die Orthogonalprojektion auf den Eigenraum zum Eigenwert λ_k, dann gilt

$$f = \sum_{j=1}^{n} \lambda_j P_j. \tag{6.99}$$

Analysis 1

7.1. Folgen und Reihen

Dieses Kapitel folgt dem Buch von T. Bröcker, [**81**] und auch dem Standardwerk von Forster [**83**] (bzw. z. B. [**82**], [**86**], [**90**], [**92**])

Um Analysis betreiben zu können benötigt man den Begriff eines Abstandes von Elementen von Mengen. Im $\mathbb{R}^1$ kann man den Abstand d zweier reeller Zahlen x, y (Entfernung zweier Punkte auf der Zahlengeraden) durch den Betrag der Differenz definieren

$$d(x, y) := |x - y|. \tag{7.1}$$

Allgemein führt man dazu den Begriff des metrischen Raumes ein.

Definition 7.1. *Eine Menge $M \neq \emptyset$ heisst* metrischer Raum, *wenn je zwei Elementen $x, y \in M$ eine reelle Zahl $d(x, y)$ zugeordnet ist, sodass gilt*

(i) $d(x, y) \geq 0$ für alle $x, y \in M$, (Nichtnegativität).

(ii) $d(x, y) = 0$ genau dann, wenn $x = y$.

(iii) $d(x, y) = d(y, x)$ für alle $x, y \in M$, (Symmetrie).

(iv) $d(x, y) \leq d(x, z) + d(z, y)$ für alle $x, y, z \in M$, (Dreiecksungleichung).

Eine Metrik ist also eine Funktion d

$$d : M \times M \to \mathbb{R}, \qquad (x, y) \mapsto d(x, y). \tag{7.2}$$

Beispiele:

(1) Mit der Abstandsdefinition $d(x,y) := |x-y|$ wird die Menge $\mathbb{R}$ zu einem metrischen Raum.

(2) Analog wird die Menge der komplexen Zahlen $\mathbb{C}$ mit der Abstandsdefinition die durch den Absolutbetrag gegeben ist $d(z_1,z_2) := |z_1 - z_2|$, zu einem metrischen Raum.

(3) Auf der Menge der beschränkten Funktionen B auf $[0,1]$ definiert die Funktion

$$d(f,g) := \sup_{x\in[0,1]} |f(x)-g(x)|, \qquad f,g \in B([0,1]) \tag{7.3}$$

eine Metrik.

(4) Der *Hammingabstand* wird bei der Fehlerkorrektur verwendet: Details siehe [**42**], Seite 373.

(5) Normierte Vektorräume:

Satz 7.2. *Es sei V ein Vektorraum mit einer Norm $||\ ||$. Dann wird durch*

$$d(x,y) := ||x-y||, \qquad x,y \in V \tag{7.4}$$

eine Metrik auf V erzeugt.

Anmerkung: Bei einem Vektorraum kann man die Elemente addieren, was bei einem beliebigen metrischen Raum nicht der Fall ist. Zusätzlich gilt für euklidische Vektorräume die Schwarzsche Ungleichung $|\langle x,y\rangle| \le ||x||\,||y||$, die zum Beispiel für $\mathbb{R}^n$ ergibt

$$\big(\sum_{j=1}^{n} x_j y_j\big)^2 \le \big(\sum_{j=1}^{n} x_j^2\big)\,\big(\sum_{j=1}^{n} y_j^2\big). \tag{7.5}$$

Solange wir jedoch Analysis für Funktionen $f : \mathbb{R} \to \mathbb{R}$ betreiben, können wir uns auf die Metrik die durch den Abstand zweier Zahlen gegeben ist $d(x,y) = |x-y|$, beschränken.

Im Kapitel 2 haben wir den Begriff einer abstrakten Folge als eine Abbildung $a : \mathbb{N} \to B, n \mapsto a(n) = a_n$ (B eine beliebige Menge) eingeführt und wollen uns nun mit reellen Zahlenfolgen, d. h. Abbildungen $\mathbb{N} \to \mathbb{R}$ näher beschäftigen.

Man schreibt hierfür

$$(a_n)_{n\in\mathbb{N}} \quad \text{oder} \quad (a_0, a_1, a_2, a_3, \ldots)$$

bzw. kurz (a_n). Etwas allgemeiner kann man statt der Indexmenge $\mathbb{N}$ die Menge $\{n \in \mathbb{Z} \mid n \ge k\}$ aller ganzen Zahlen, die grösser gleich einer vorgegebenen ganzen Zahl k sind, nehmen und schreibt dann

$$(a_n)_{n\ge k} \quad \text{oder} \quad (a_k, a_{k+1}, a_{k+2}, , \ldots).$$

Beispiele

(1) Sei $a_n = a, a \in \mathbb{R}$ für alle $n \in \mathbb{N}$. Man erhält die *konstante Folge*

$$(a_n)_{n\in\mathbb{N}} = (a, a, a, a, \ldots).$$

(2) Sei $a_n = \frac{1}{n}, n \geq 1$. Dies ergibt die Folge

$$(a_n)_{n\geq 1} = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots).$$

(3) Sei $a_n = (-1)^n$. Dann ist

$$(a_n)_{n\in\mathbb{N}} = (+1, -1, +1, -1, \ldots).$$

(4)

$$\big(\frac{n}{n+1}\big)_{n\in\mathbb{N}} = (0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots).$$

(5)

$$\big(\frac{n}{2^n}\big)_{n\in\mathbb{N}} = (0, \frac{1}{2}, \frac{1}{2}, \frac{3}{8}, \frac{1}{4}, \frac{5}{32}, \ldots).$$

(6) Die rekursive definierte Fibonacci-Folge ist gegeben durch die Anfangswerte $a_0 = 0$ und $a_1 = 1$ und für a_{n+2} gilt

$$a_{n+2} = a_{n+1} + a_n,$$
$$(a_n)_{n\in\mathbb{N}} = (0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots).$$

Anmerkung: Dass durch Rekursion wirkliche eine Folge definiert wird, wird in einem Satz in [**92**] Kap. 2, Seite 23 bewiesen.

(7) Eine rekursive Folge sei gegeben durch den Anfangswert $a_0 = 2$ und

$$a_{n+1} = \frac{a_n}{2} + \frac{1}{a_n},$$
$$(a_n)_{n\in\mathbb{N}} = (2, \frac{3}{2}, \frac{17}{12}, \frac{577}{408}, \ldots).$$

(8) Für jede reelle Zahl x kann man die Folge ihrer Potenzen definieren als

$$(x^n)_{n\in\mathbb{N}} = (1, x, x^2, x^3, x^4, , \ldots).$$

Anmerkung: Graphisch veranschaulicht man Folgen indem man die Folgenglieder a_n auf der Zahlengeraden darstellt. Der entscheidende Begriff der Analysis ist der Begriff der Konvergenz („etwas“ strebt gegen „etwas“), der bis zum 19. Jahrhundert nur intuitiv erfasst werden konnte.

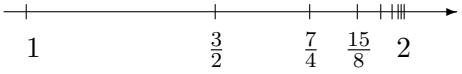


Abbildung 7.1. Die ersten Glieder der Folge $a_n = (2 - \frac{1}{2^n})$.

Definition 7.3. *Eine reelle Folge von Zahlen* (a_n) *heisst* konvergent *gegen* $a \in \mathbb{R}$*, falls gilt:*

Zu jeder reellen Zahl $\epsilon > 0$ *existiert eine Zahl* $N_\epsilon \in \mathbb{N}$*, sodass*

$$|a_n - a| < \epsilon \tag{7.6}$$

für alle Indizes $n \geq N_\epsilon$ *ist. Man nennt* a *den* Grenzwert (Limes) *der Folge und schreibt*

$$\lim_{n\to\infty} a_n = a, \qquad \textit{oder} \qquad a_n \to a \quad \textit{für} \quad n \to \infty. \tag{7.7}$$

Anmerkung: Allgemein kann man analog Konvergenz für Folgen in beliebigen metrischen Räumen definieren, wobei man nur $|a_n - a|$ durch $d(a_n, a)$ ersetzen muss.

Man beachte, dass die Zahl N_ϵ von ϵ abhängt. Da es darum geht, dass a_n beliebig nahe an a herankommt, kommt es nur darauf an beliebig kleine $\epsilon > 0$ zu betrachten. D. h. für jedes noch so kleine $\epsilon > 0$ existiert ein Folgenindex $N_\epsilon \in \mathbb{N}$, sodass $|a_n - a| < \epsilon$, falls $n \geq N_\epsilon$. Im allgemeinen wird man den Folgenindex N_ϵ umso grösser wählen müssen, je kleiner ϵ ist.

Definition 7.4. *Das offene Intervall* $U_\epsilon(a) := (a - \epsilon, a + \epsilon), \epsilon, a \in \mathbb{R}$*,* $\epsilon > 0$*, wird als die* ϵ-Umgebung *des Punktes* a *bezeichnet.*

Ergänzung:

Definition 7.5. *Sei* (M, d) *ein metrischer Raum. Die Menge* $U(a, \epsilon) = \{x \in M \mid d(x, a) < \epsilon\}$ *heisst* ϵ*-Umgebung von* a *(*offene Kugel *um* a *mit Radius* ϵ*).*

Eine Teilmenge $A \subset M$ *heisst* offen*, wenn jeder Punkt von* A *eine* ϵ*-Umgebung, die ganz in* A *liegt, enthält.*

Eine Teilmenge $B \subset M$ *heisst* abgeschlossen*, wenn ihr Komplement offen ist.*

Mehr darüber findet man z. B. in [**82**], Kapitel 2.3 oder [**79**], Seite 222 ff.
Mit dem Begriff der ϵ-Umgebung kann Konvergenz anschaulich wie folgt charakterisiert werden: Die Folgenglieder a_n liegen in $U_\epsilon(a)$ für alle n bis auf endlich

viele, nämlich bis auf $n \in \{0, 1, 2, \ldots, N_\epsilon - 1\}$. Man sagt dafür auch $a_n \in U_\epsilon(a)$ für **fast alle** n.

Definition 7.6. *Eine nichtkonvergente Folge wird als* divergent *bezeichnet.*

Beispiele:

(1) Die konstante Folge $a_n = a$ konvergiert trivialerweise gegen a.

(2) die Folge $a_n = \frac{1}{n}$ konvergiert, d.h. $\lim_{n\to\infty} \left(\frac{1}{n}\right)_{n\geq 1} = 0.$

Beweis: Es sei $\epsilon > 0$ gegeben. Da die Menge $\mathbb{N}$ unbeschränkt ist, gibt es eine natürliche Zahl $N_\epsilon > \frac{1}{\epsilon}$ und für alle n mit $n \geq N_\epsilon$ gilt $n > \frac{1}{\epsilon}$ und für diese gilt daher

$$\left|\frac{1}{n} - 0\right| = \frac{1}{n} < \epsilon, \text{ man wähle } N_\epsilon = \lfloor \frac{1}{\epsilon} \rfloor + 1.$$

Anmerkung: Eine Folge deren Grenzwert 0 ist, heisst **Nullfolge**. Ist (a_n) konvergent mit Grenzwert a, dann konvergiert die Folge $b_n := (a_n - a)$ gegen 0.

(3) Die Folge $a_n = (-1)^n$ divergiert.

(4)
$$\lim_{n\to\infty} \left(\frac{n}{n+1}\right) = 1, \qquad \text{da gilt}$$
$$\left|\frac{n}{n+1} - 1\right| = \frac{1}{n+1} < \epsilon.$$

(5)
$$\lim_{n\to\infty} \left(\frac{n}{2^n}\right) = 0, \qquad \text{da gilt}$$
$$2^n = (1+1)^n = 1 + n + \frac{n(n-1)}{2} + \ldots \geq \frac{n(n+1)}{2} \geq \frac{n^2}{2},$$
$$\left|\frac{n}{2^n} - 0\right| \leq \frac{2}{n} < \epsilon.$$

(6) Die Fibonacci-Folge ist divergent, da die Folgenglieder beliebig gross werden und es gilt $a_{n+1} \geq n$.

Definition 7.7. *Man sagt* $(a_n) \to \infty$ *oder* $\lim_{n\to\infty} a_n = \infty$ *falls gilt: Für jedes* $R > 0$ *ist* $a_n > R$ *für fast alle* $n \in \mathbb{N}$. *Entsprechend schreibt man* $(a_n) \to -\infty$ *falls* $(-a_n) \to \infty$. *Man nennt eine solche Folge* (a_n) bestimmt divergent *oder auch* uneigentlich konvergent.

D. h. die Fibonacci-Folge ist bestimmt divergent.

(7) Zur Untersuchung der Konvergenz der Folge $(x^n), x \in \mathbb{R}$ machen wir Fallunterscheidungen:

(a) $x = 1$. Dann ist $(1^n)_{n\in\mathbb{N}} = (1, 1, 1, \ldots) \to 1$ konvergent.

(b) $x = -1$. Dann ist $(1^n)_{n\in\mathbb{N}} = (1, -1, 1, -1, \ldots)$ divergent.

(c) $|x| > 1$. Es sei $|x| = d$. Dann gilt $d = (1 + \delta), \delta > 0$ und $d^n = (1 + \delta)^n \geq 1 + n\delta > n\delta$. Also wird d^n beliebig gross und (x^n) ist daher divergent (bestimmt divergent wenn $x > 1$).

(d) $|x| < 1$. Es sei $|x| = d$. Für $d = 0$ gilt $(x^n) \to 0$.
Für $0 < d < 1$ gilt $\frac{1}{d} > 1$ und daher gibt es für jedes ϵ ein n, sodass $(\frac{1}{d})^n > \frac{1}{\epsilon}$ gilt. Daraus folgt dann $d^n < \epsilon$ und $|x^n - 0| < \epsilon$ für fast alle n und die Folge (x^n) konvergiert daher gegen 0.

Der Grenzwert einer Folge ist eindeutig, das heisst eine Folge kann nicht zwei verschiedene Grenzwerte haben.

Satz 7.8. *Konvergiert eine Folge gegen a und gegen b dann gilt $a = b$.*

Beweis durch Widerspruch: Annahme $a < b$, man wähle $\epsilon = \frac{b-a}{2}$. Dann folgt aus

$$|a_n - a| < \frac{b-a}{2} \qquad \text{und} \qquad |a_n - b| < \frac{b-a}{2}$$
$$b - a = |b - a| \leq |b - a_n| + |a_n - a| < b - a, \qquad \text{Widerspruch!}$$

Satz 7.9. *Eine konvergente Folge ist* **beschränkt**, *das heisst die Menge $\{a_n \mid n \in \mathbb{N}\}$ der Folgenglieder ist beschränkt.*

Beweis: Ausserhalb der ϵ-Umgebung liegen nur endlich viele Folgenglieder und für die Folgenglieder innerhalb der ϵ-Umgebung gilt $|a_n| < |a| + \epsilon$. Insgesamt erhält man

$$|a_n| \leq \max\{|a| + \epsilon, |a_0|, \ldots, |a_{N_\epsilon - 1}|\}.$$

Folgen kann man addieren, multiplizieren und manchmal auch dividieren. Man definiert

$$\text{(i)} \quad \lambda(a_n) + \mu(b_n) := (\lambda a_n + \mu b_n), \qquad \lambda, \mu \in \mathbb{R}, \tag{7.8}$$
$$\text{(ii)} \quad (a_n) \cdot (b_n) := (a_n b_n). \tag{7.9}$$

Ist $a_n \neq 0$ für fast alle n so definiert man

$$\text{(iii)} \quad \frac{1}{(a_n)} := \left(\frac{1}{a_n}\right). \tag{7.10}$$

Eigenschaft (i) macht die Menge aller reellen Zahlenfolgen zu einem reellen Vektorraum.

Die Multiplikation von Folgen ist assoziativ, kommutativ und distributiv. Die konstante Folge (1) ist das neutrale Element.

Für die Grenzwerte von Summen- und Produktfolgen gilt der folgende Satz

Satz 7.10. *Es konvergiere $(a_n) \to a$ und $(b_n) \to b$. Dann gilt*

(i) Die Grenzwertbildung ist linear, das heisst $\lambda(a_n) + \mu(b_n) \to \lambda a + \mu b$.

(ii) Die Grenzwertbildung ist multiplikativ, das heisst $(a_n)(b_n) \to ab$.

(iii) Ist $b \neq 0$, so ist $b_n \neq 0$ für fast alle n und es gilt $\left(\frac{1}{b_n}\right) \to \frac{1}{b}$.

Beweis: (i) Sei $\epsilon > 0$. Wir haben zu zeigen, dass es einen Index N_ϵ gibt, sodass für alle $n > N_\epsilon$ gilt

$$|(\lambda a_n + \mu b_n) - (\lambda a + \mu b)| < \epsilon.$$

Da (a_n) und (b_n) konvergent sind wählen wir N_ϵ so, dass gilt

$$|a_n - a| < \epsilon_1 \qquad \text{wobei} \quad \epsilon_1 = \frac{\epsilon}{2\,|\lambda|} \qquad \text{und}$$
$$|b_n - b| < \epsilon_2 \qquad \text{wobei} \quad \epsilon_2 = \frac{\epsilon}{2\,|\mu|}.$$

Dann folgt mit der Dreiecksungleichung

$$|(\lambda a_n + \mu b_n) - (\lambda a + \mu b)| \leq |\lambda|\,|a_n - a| + |\mu|\,|b_n - b| < \epsilon.$$

Die Beweise von (ii) und (iii) sind analog und man findet sie zum Beispiel in [**81**], Seite 31.

Diesen Satz kann man nun zum Beweis der Konvergenz und zur Berechnung des Grenzwertes einer Folge (a_n) verwenden indem man den Ausdruck a_n in einfachere Ausdrücke zerlegt, deren Konvergenz unmittelbar einsichtig ist oder bereits bewiesen wurde.

Beispiel:

$$\frac{8n^3 + 7n^2 + 6}{4n^3 + 3n^2 + 5} = \frac{8 + 7\frac{1}{n} + 6\frac{1}{n^3}}{4 + 3\frac{1}{n} + 5\frac{1}{n^3}} \to 2,$$

da $\frac{1}{n}$ und $\frac{1}{n^3}$ Nullfolgen sind (man dividiert durch die höchste Potenz des Nenners).

Einige Beispiele konvergenter Folgen (siehe [**86**], Seite 42)

$$\left(\frac{1}{n^s}\right) \to 0, \quad s > 0, s \in \mathbb{Q},$$

$$\begin{aligned}
\left(\sqrt[n]{a}\right) &\to 1, \quad a > 0, a \in \mathbb{R}, \\
\left(\sqrt[n]{n}\right) &\to 1, \\
(q^n) &\to 0, \quad |q| < 1, q \in \mathbb{R}, \\
\left(\frac{n^k}{z^n}\right) &\to 0, \quad k \in \mathbb{N}, |z| > 1, z \in \mathbb{C}.
\end{aligned} \tag{7.11}$$

Ist (a_n) eine Folge und (n_k) eine Folge natürlicher Zahlen mit $n_0 < n_1, \ldots$, so heisst die Folge $(a_{n_k}) = (a_{n_0}, a_{n_1}, \ldots)$ eine **Teilfolge** von (a_n). Zum Beispiel ist die Folge $(2n)$ eine Teilfolge der Folge (n).

Ist (a_n) konvergent, so ist auch jede Teilfolge konvergent mit demselben Grenzwert. Allgemein gilt:

Satz 7.11. *Es sei* $\lim a_n = a$ *und* $\rho : \mathbb{N} \to \mathbb{N}$ *injektiv, dann ist*

$$\lim_{n\to\infty} \left(a_{\rho(n)}\right) = a. \tag{7.12}$$

Es darf also beliebig umgeordnet werden.

Beweis in [**81**], Seite 32.

Die Grenzwertbildung überträgt sich auf die Ordnung in $\mathbb{R}$ in folgender Form:

Satz 7.12. *Sind* (a_n) *und* (b_n) *konvergente Folgen und ist* $a_n \leq b_n$ *oder* $a_n < b_n$ *für fast alle* n*, so gilt für die Grenzwerte* $\lim a_n \leq \lim b_n$*.*

Beweis durch Widerspruch: Es sei $(a_n) \to a$ und $(b_n) \to b$. Annahme $a > b$. Man wähle $\epsilon = \frac{a-b}{2}$. Dann folgt

da $\quad |a_n - a| < \epsilon \quad$ und $\quad |b_n - b| < \epsilon$

$$b_n < b + \epsilon = b + \frac{a-b}{2} = \frac{a+b}{2} = a - \frac{a-b}{2} = a - \epsilon < a_n \quad \text{Widerspruch!}.$$

Bisher haben wir die Konvergenz einer Folge dadurch bewiesen, dass wir den Grenzwert angegeben haben und die in der Definition der Konvergenz geforderte Abschätzung nachgewiesen haben. Meist ist die Angabe des Grenzwertes gar nicht möglich, sondern dieser wird erst durch eine Folge oder Reihe definiert und man braucht daher Konvergenzkriterien die ohne Angabe des Grenzwertes auskommen.

Definition 7.13. *Eine Folge* (a_n) *heisst:*

(i) monoton wachsend, wenn $a_n \leq a_{n+1}$*,*
(ii) monoton fallend, wenn $a_n \geq a_{n+1}$*,*
(iii) streng monoton wachsend, wenn $a_n < a_{n+1}$*,*
(iv) streng monoton fallend, wenn $a_n > a_{n+1}$*.*

Satz 7.14. *Eine monoton wachsende nach oben beschränkte reelle(!) Folge konvergiert. Analog gilt: Eine monoton fallende nach unten beschränkte reelle Folge konvergiert.*[1]

Mit diesem Satz kann man nun die Konvergenz der Folgen (z. B. [**79**], Seite 113ff)

$$a_n = \left(1 + \frac{1}{n}\right)^n \tag{7.13}$$

und

$$a_{n+1} = \frac{a_n}{2} + \frac{1}{a_n}, \quad a_0 = 2 \tag{7.14}$$

nachweisen indem man Beschränktheit und Monotonie (z. B. mit vollständiger Induktion) beweist.

Mit Mathematica:

Beispiel 1: Wir definieren die Folge

```
In[44]:=a[n_] = (1 + 1/n)^n
```

und lassen uns dann die gesuchten Folgenglieder ausgeben:

```
In[45]:=N[{a[2], a[3], a[10], a[100], a[1000], a[10000]}]
Out[45]={2.25, 2.37037, 2.59374, 2.70481, 2.71692, 2.71815}
```

Erinnern Sie sich, dass der Befehl `N[...]` nur bewirkt, dass das Ergebnis numerisch angezeigt wird. Der Grenzwert kann in `Mathematica` mit dem Befehl `Limit` ausgerechnet werden:

```
In[46]:=Limit[a[n], n -> ∞]
Out[46]=e
```

Der numerische Wert von e ist:

```
In[47]:=N[e]
Out[47]=2.71828
```

Der Pfeil $\rightarrow$, das Symbol ∞ und die Konstante e werden in `Mathematica` entweder über die Palette oder über die Tastatur als `->`, `Infinity` bzw. `E` eingegeben.

Beispiel 2: Das erste Folgenglied ist vorgegeben (bzw. wir haben es als Startwert gewählt): $a_1 = 2$. Das Bildungsgesetz für das n-te Folgenglied enthält das vorhergehende Folgenglied: $a_n = \frac{1}{2}(a_{n-1} + \frac{2}{a_{n-1}})$, $n \geq 2$.
Mit `Mathematica` können wir

[1] Dieser Satz gilt allgemein nur für Mengen die vollständig sind, also für $\mathbb{R}$ aber nicht für $\mathbb{Q}$.

In[48]:= $\mathtt{a[1]} = 2.;\ \mathtt{a[n_]} := \frac{1}{2}\left(\mathtt{a[n-1]} + \frac{2}{\mathtt{a[n-1]}}\right)$

definieren und wieder die ersten Folgenglieder berechnen:

In[49]:= $\{\mathtt{a[1]}, \mathtt{a[2]}, \mathtt{a[3]}, \mathtt{a[4]}, \mathtt{a[5]}\}$

Out[49]= $\{2., 1.5, 1.41667, 1.41422, 1.41421\}$

Hier muss $a[n_]$ mit einem „:=“ definiert werden, wodurch `Mathematica` angewiesen wird, die rechte Seite von „:=“ nicht gleich, sondern erst bei Angabe einer konkreten Zahl n auszuwerten. Beachten Sie, dass ein Startwert a_1 angegeben werden muss. Wenn `Mathematica` a_n berechnet, so drückt es gemäss dem Bildungsgesetz a_n durch a_{n-1} aus. Das geschieht so lange, bis es auf den Startwert trifft. Wurde dieser nicht angegeben, so bricht `Mathematica` mit einem Fehler ab.

Satz 7.15. *Jede Folge besitzt eine monotone Teilfolge.*

Beweis: [**81**], Seite 43.

Als Anwendung dieses kleinen Satzes folgt der Satz von **Bolzano-Weierstrass**:

Satz 7.16. *Jede beschränkte reelle Folge besitzt eine konvergente Teilfolge.*

Beweis: Man wähle eine monotone Teilfolge nach Satz 7.15. Da auch diese beschränkt ist, folgt mit Satz 7.14 die Existenz einer konvergenten Teilfolge.

Definition 7.17. *Ein Punkt $a \in \mathbb{R}$ heisst* Häufungspunkt *der Folge (a_n), wenn es eine Teilfolge von (a_n) gibt die gegen a konvergiert.*

Eine beschränkte Folge hat daher mindestens einen Häufungspunkt und eine konvergente Folge besitzt genau einen Häufungspunkt.

Ist a ein Häufungspunkt, so enthält jede Umgebung von a unendlich viele a_n.

Beispiel:

$a_n = (-1)^n \frac{n}{n+1}$.

7.1.1. Uneigentliche Konvergenz.

(Siehe [**86**], Seite 54)

Um gewissen divergenten Folgen in $\mathbb{R}$ wenigstens einen uneigentlichen Grenzwert zuordnen zu können, erweitert man $\mathbb{R}$ um zwei Elemente zu $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$.

Definition 7.18. *Eine Folge (a_n) in $\overline{\mathbb{R}}$ konvergiert gegen ∞ (bzw. $-\infty$), wenn es zu jedem $K > 0$ eine Zahl $N_K \in \mathbb{N}$ gibt, sodass für alle $n \geq N_K$ gilt:*

$$a_n \geq K \qquad (bzw. \quad a_n \leq -K).$$

Beispiel: (a^n) konvergiert in $\overline{\mathbb{R}}$ für $a > 1$ gegen ∞ und divergiert für $a \leq -1$.

Definition 7.19. *Es sei (a_n) eine Folge reeller Zahlen. Dann heisst die Zahl $c \in \overline{\mathbb{R}}$*

$$c = \limsup_{n\to\infty} a_n := \inf\{x \mid a_n \leq x \text{ für fast alle } n\} =: \overline{\lim}\, a_n$$

der Limes superior *oder* oberer Häufungspunkt *von (a_n). Analog ist der* Limes inferior *oder* untere Häufungspunkt *von (a_n) definiert*

$$\liminf_{n\to\infty} a_n := \sup\{x \mid a_n \geq x \text{ für fast alle } n\} =: \underline{\lim}\, a_n$$

7.1.2. Cauchyfolgen. In der Definition der Konvergenz einer Folge in Definition 7.3 war die Angabe des Grenzwertes notwendig. Wir geben nun eine Definition die ohne die Angabe des Grenzwertes auskommt.

Definition 7.20. *Eine Folge (a_n) heisst* **Cauchy-Folge** *falls zu jeder reellen Zahl $\epsilon > 0$ eine Zahl $N_\epsilon \in \mathbb{N}$ existiert, sodass*

$$|a_n - a_m| < \epsilon \tag{7.15}$$

für alle Indizes $n, m \geq N_\epsilon$ ist.

Allgemein gilt:

Satz 7.21. *Jede konvergente Folge ist eine Cauchy-Folge.*

Beweis: Sei $|a_n - a| < \epsilon_1 = \frac{\epsilon}{2}$ und $|a_m - a| < \epsilon_1$. Dann gilt

$$|a_n - a_m| = |a_n - a + a - a_m| \leq |a_n - a| + |a - a_m| < \epsilon. \tag{7.16}$$

Anmerkung: (1) Für reelle Zahlenfolgen bedeutet dies, dass eine Folge genau dann konvergiert, wenn sie ein Cauchy-Folge ist. Die reellen Zahlen sind bezüglich Konvergenz vollständig.

(2) Betrachtet man Folgen in $\mathbb{Q}$, die gegen eine nicht-rationale Zahl konvergieren, so kann man nachweisen, dass sie Cauchy-Folgen sind. Die Definition 7.3 ist in diesem Fall nicht anwendbar, da der Grenzwert a ja nicht in $\mathbb{Q}$ liegt.

(3) Man kann daher die reellen Zahlen auch als Äquivalenzklassen aller Cauchyfolgen mit gleichen Grenzwert in $\mathbb{Q}$ einführen und so $\mathbb{Q}$ vervollständigen.

Man kann die Konvergenz einer Folge in $\mathbb{R}$ dadurch zeigen indem man nachweist, dass sie eine Cauchy-Folge ist und braucht daher den Grenzwert nicht anzugeben.

7.1.3. Reihen. Eng verwandt mit dem Begriff Folge ist der Begriff Reihe. Eigentlich handelt es sich um dieselbe Sache nur in anderer Formulierung.

Definition 7.22. *Es sei (a_k) eine reelle Folge. Die Folge (A_n) der Summen*

$$A_n := \sum_{k=0}^{n} a_k \tag{7.17}$$

heisst Reihe *mit den* Gliedern a_k *und wird mit $\sum_{k=0}^{\infty} a_k$ bezeichnet. Man nennt A_n die n-te* Partialsumme *der Reihe. Oft beginnt die Summation bei einer beliebigen ganzen Zahl k_0. Konvergiert die Reihe (d. h. konvergiert die Folge der Partialsummen), so bezeichnet man auch den Grenzwert mit $\sum_{k=0}^{\infty} a_k$ und nennt ihn die* Summe *der Reihe.*

Jede Folge (A_n) lässt sich als Reihe schreiben: $a_0 = A_0, a_n = A_n - A_{n-1}, n > 0$.

Beispiel 1: Eine Reihe bei der die Differenz $a_{k+1} - a_k = d$ konstant ist, heisst **arithmetische Reihe**. Die arithmetische Reihe ist bestimmt divergent.

Übung: Man berechne die n-te Partialsumme der arithmetischen Reihe.

Beispiel 2: Die Reihe $\sum\limits_{k=0}^{\infty} x^k,\ x \in \mathbb{R}$ heisst **geometrische Reihe**. Bei der geometrischen Reihe ist der Quotient $\frac{a_{k+1}}{a_k} = x$ konstant. Es gilt:

(i) Für $|x| \geq 1$ ist die geometrische Reihe divergent.

(ii) Für $|x| < 1$ ist die geometrischen Reihe konvergent $\sum\limits_{k=0}^{\infty} x^k = \dfrac{1}{1-x}$.

Beweis: Für $|x| < 1$ gilt

$$\begin{aligned} 1 + x + x^2 + \cdots x^n &= A_n, \\ x + x^2 + x^3 \cdots x^{n+1} &= xA_n, \\ 1 - x^{n+1} &= A_n - xA_n, \\ A_n &= \frac{1 - x^{n+1}}{1 - x}, \\ \lim_{n\to\infty} A_n &= \frac{1}{1-x}. \end{aligned}$$

Beispiel 3: Man berechne die n-te Partialsumme der Reihe

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)}. \tag{7.18}$$

Satz 7.23. *Ist eine Reihe $\sum_{k=0}^{\infty} a_k$ konvergent, so ist (a_k) eine Nullfolge.*

Beweis: Ist $A_n = \sum_{k=0}^{n} a_k$ und $a = \lim_{n\to\infty} A_n$, so gilt: $(a_k) = (A_k - A_{k-1}) \to a - a = 0$.

Die Umkehrung ist aber falsch denn die *harmonische Reihe* $\sum_{n=1}^{\infty} \frac{1}{n}$ ist divergent, obwohl die Glieder $\frac{1}{n}$ eine Nullfolge bilden, denn $A_{2^n} \geq n/2$.

7.1.4. Konvergenzsätze. Sätze über Folgen lassen sich als Sätze über Reihen und umgekehrt schreiben. Zum Beispiel folgt aus Satz 7.10 sofort der entsprechende Satz für Reihen

Satz 7.24. *Sind die Reihen* $\sum_{k=0}^{\infty} c_k$ *und* $\sum_{k=0}^{\infty} d_k$ *konvergent, so gilt*

$$\sum_{k=0}^{\infty}(\lambda c_k + \mu d_k) = \lambda \sum_{k=0}^{\infty} c_k + \mu \sum_{k=0}^{\infty} d_k, \quad \lambda, \mu \in \mathbb{R} \tag{7.19}$$

Eine Reihe heisst beschränkt, wenn die Folge der Partialsummen beschränkt ist.

Die Beschränktheit einer Reihe prüft man nach indem man sie mit einer Reihe vergleicht über deren Konvergenzverhalten man Bescheid weiß.

Satz 7.25. Majorantenkriterium. *Es sei* $\sum_{n=0}^{\infty} a_n$ *eine konvergente Reihe mit* $a_n \geq 0$ *für alle* n. *Und* (c_n) *eine Folge mit*

$$0 \leq c_n \leq a_n \text{ für alle } n \tag{7.20}$$

dann konvergiert $\sum_{n=0}^{\infty} c_n$ *und* $\sum_{n=0}^{\infty} a_n$ *heisst Majorante von* $\sum_{n=0}^{\infty} c_n$.

Beispiel: es gilt

$$\frac{1}{(k+1)^2} \leq \frac{1}{k(k+1)}, \quad k > 0$$

und daher

$$\sum_{k=1}^{\infty} \frac{1}{(k+1)^2} \leq \sum_{k=1}^{\infty} \frac{1}{k(k+1)}.$$

Weitere Anwendungen des Majorantenkriterium sind die folgenden Sätze.

Satz 7.26. (Verdichtungslemma von Cauchy) *Es sei* (c_n) *eine nichtnegative reelle monoton fallende Folge. Dann konvergiert* $\sum_{n=0}^{\infty} c_n$ *genau dann, wenn* $\sum_{n=0}^{\infty} 2^n c_{2^n}$ *konvergiert.*

Beweis: [**81**], Seite 40.

Beispiel: Die Reihe $\sum\limits_{n=1}^{\infty} \frac{1}{n^\alpha}$ ist genau dann konvergent, wenn $\alpha > 1$ ist.

Die harmonische Reihe $\sum_{n=1}^{\infty} \frac{1}{n}$ divergiert daher. Für $\alpha = 2$ ergibt sich

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6}.$$

Die wichtigste Vergleichsreihe für das Majorantenkriterium ist die geometrische Reihe. Damit erhält man das

Satz 7.27. (Quotientenkriterium) *Es sei $c_n \geq 0$, und es existiere eine Zahl $\vartheta < 1$, sodass $c_{n+1} \leq \vartheta c_n$ für fast alle n. Dann konvergiert $\sum_{n=0}^{\infty} c_n$.*

Beispiel: Die Reihe

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} \qquad (=: \exp(x)) \tag{7.21}$$

konvergiert für jedes $x \geq 0$ und definiert daher eine Funktion. Es gilt

$$\frac{c_{n+1}}{c_n} = \frac{x}{n+1} \leq \frac{1}{2} = \vartheta < 1 \quad \text{für} \quad n > 2x.$$

Satz 7.28. (Wurzelkriterium) *Es sei $c_n \geq 0$, und es existiere eine Zahl $\vartheta < 1$, sodass $\sqrt[n]{c_n} \leq \vartheta$ für fast alle n. Dann konvergiert $\sum_{n=0}^{\infty} c_n$. Ist $\sqrt[n]{c_n} \geq 1$ für unendlich viele n, so divergiert $\sum_{n=0}^{\infty} c_n$.*

Das Wurzelkriterium ist genauer als das Quotientenkriterium, aber das Quotientenkriterium ist einfacher anzuwenden. Beide Kriterien versagen jedoch bei der Reihe $\sum_{n=0}^{\infty} \frac{1}{n^\alpha}$.

Eine Reihe deren Glieder abwechselnd positiv und negativ sind heisst **alternierende Reihe**.

Satz 7.29. (Leibnizkriterium) *Sei $a_n \geq 0$ eine monoton fallende Nullfolge. Dann konvergiert die alternierende Reihe $\sum_{n=0}^{\infty} (-1)^n a_n$.*

Anwendung: Folgende Reihen sind konvergent:

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + - \cdots = \log 2,$$
$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots = \frac{\pi}{4}.$$

Das Leibnizkriterium beweist aber nur die Konvergenz. Die Berechnung dieser Grenzwerte erfolgt später.

Angewendet auf Reihen ergibt das Cauchy-Kriterium

Satz 7.30. *Eine Reihe $\sum_{n=0}^{\infty} c_n$ ist genau dann konvergent, wenn es zu jedem $\epsilon > 0$ ein $m_\epsilon \in \mathbb{N}$ gibt, sodass für alle $k \in \mathbb{N}$ gilt*

$$\Big| \sum_{n=m_\epsilon}^{m_\epsilon+k} c_n \Big| < \epsilon.$$

Definition 7.31. *Eine Reihe $\sum_{n=0}^{\infty} c_n$ heisst* absolut konvergent*, wenn die Reihe der Beträge $\sum_{n=0}^{\infty} |c_n|$ konvergiert.*

Eine absolut konvergente Reihe ist daher konvergent.

Auch die Reihe (7.21) konvergiert absolut und konvergiert daher für $x \in \mathbb{R}$.

Ist (c_n) beliebig, so heisst $\sum a_n$ eine Majorante von $\sum c_n$, wenn $\sum a_n$ eine Majorante der positiven Reihe $\sum |c_n|$ ist. Die Konvergenzkriterien für positive Reihen liefern daher unmittelbar Kriterien für die absolute Konvergenz.

Beispiel: Ist (a_n) beschränkt, so konvergiert die Reihe

$$\sum_{k=0}^{\infty} a_k x^k \tag{7.22}$$

für alle $x \in \mathbb{R}$ mit $|x| < 1$ absolut.

Anwendung: Dezimalentwicklung einer irrationalen Zahl.

Absolut konvergente Reihen lassen sich beliebig umordnen ohne das sich der Grenzwert ändert.

Satz 7.32. *Es sei $\sum_{k=0}^{\infty} c_k$ absolut konvergent, und es sei $\rho : \mathbb{N} \to \mathbb{N}$ bijektiv. Dann konvergiert $\sum_{k=0}^{\infty} c_{\rho(k)}$ absolut gegen denselben Grenzwert.*

Beweis: [**81**], Seite 49.

Für nicht absolut konvergente Reihen gilt gerade das „Gegenteil“ von Satz 7.32.

Satz 7.33. *Es sei $\sum_{k=0}^{\infty} c_k$ konvergent aber nicht absolut konvergent und $x \in \mathbb{R}$ beliebig. Dann existiert eine Umordnung $\rho : \mathbb{N} \to \mathbb{N}$, sodass $\sum_{k=0}^{\infty} c_{\rho(k)} = x$.*

Beweis: [**81**], Seite 50.

D. h. man kann bei bedingt konvergenten Reihen durch Umordnung erreichen, dass der Grenzwert eine beliebige Zahl ist.

Für absolut konvergente Reihen kann ihr Produkt auf folgende Weise berechnet werden.

Satz 7.34. *(Cauchy-Produkt von Reihen). Es seien $\sum_{k=0}^{\infty} a_k$ und $\sum_{k=0}^{\infty} b_k$ absolut konvergente Reihen. Für $n \in \mathbb{N}$ definiert man*

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} = a_0 b_n + a_1 b_{n-1} + \cdots a_n b_0.$$

Dann ist auch die Reihe $\sum_{n=0}^{\infty} c_n$ absolut konvergent und es gilt

$$\sum_{n=0}^{\infty} c_n = \left(\sum_{j=0}^{\infty} a_j \right) \cdot \left(\sum_{k=0}^{\infty} b_k \right). \tag{7.23}$$

Beweis: [**83**], Seite 74.

Beispiel: Es gilt $\exp(x)\exp(y) = \exp(x+y)$.

7.1.5. Landausymbole.

Definition 7.35. *(Landausymbole) Seien a_n und $f(n)$ zwei reelle Folgen. Wenn es Konstanten C und R gibt, sodass*

$$|a_n| \leq C f(n) \qquad \text{für alle } n \geq R,$$

dann schreiben wir $a_n = O(f(n))$ und sagen, a_n **ist von der Ordnung** *$f(n)$ oder a_n* **ist „Groß-O“ von** *$f(n)$.*

Ist

$$\lim_{x \to \infty} \frac{a_n}{f(n)} = 0,$$

dann schreiben wir $a_n = o(f(n))$ und sagen, a_n **ist „Klein-o“ von** *$f(n)$.*

Gilt

$$\lim_{n \to \infty} \frac{a_n}{f(n)} = 1,$$

dann sagen wir a_n und $f(n)$ sind asymptotisch gleich und schreiben $a_n \sim f(n)$.

Anwendung: Laufzeit von Algorithmen, siehe [**33**].

Gebräuchliche Funktionen $f(n)$:
$f(n) = \log(n)$ ... logarithmisch
$f(n) = n$ ... linear
$f(n) = n \log(n)$
$f(n) = n^{1+\epsilon}$
$f(n) = n^2$ ... quadratisch
$f(n) = 2^n$ ... exponentiell

Beispiel: Sei $a_n = n^2 + 3n + 1, f(n) = n^2$, dann ist a_n von der Ordnung $O(n^2)$ und $a_n \sim n^2$.

7.2. Übung

(1) Man bestimme den Grenzwert a der Folge $(a_n), n \in \mathbb{N}$ (falls konvergent) und ein $N_\epsilon \in \mathbb{N}$, sodass $|a_n - a| < \epsilon$ für $n > N_\epsilon$ wobei (i) $\epsilon = 10^{-2}$, (ii) ϵ beliebig gilt.

(a) $a_n = \dfrac{n}{n+2}$

(b) $a_n = \dfrac{6n-2}{3n+7}$

(c) $a_n = \dfrac{n-n^2}{n^2+1}$

(d) $a_n = \dfrac{n^2}{n+1}$

(e) $a_n = \dfrac{1-2n^2-4n}{n^2+2n+1}$.

(2) Man bestimme den Grenzwert a der Folge (a_n).
Anmerkung: Grenzwertbildung und allgemeine Potenz sind vertauschbar, d. h. $\lim\big(f(n)\big)^q = \big(\lim f(n)\big)^q, q \in \mathbb{Q}$.
Hinweis: Hat der Ausdruck $(a-b)$ die Form $(\infty - \infty)$ so ist meist die Umformung $(a-b) * (a+b)/(a+b) = (a^2-b^2)/(a+b)$ von Nutzen.

(a) $a_n = \dfrac{3n^2-5n+7}{-9n^2+6n-3}$

(b) $a_n = \sqrt{n+1} - \sqrt{n}$

(c) $a_n = \sqrt{n^2+2n+3} - \sqrt{n^2-2n+2}$

(d) $a_n = \sqrt{\dfrac{6n-2}{3n+7}}$

(e) $a_n = n(\sqrt{1+\dfrac{1}{n}} - 1)$

(f) $a_n = \dfrac{4^n+1}{5^n}$

(g) $a_n = \sqrt[n]{n^3(3^n+5)}$

(h) $a_n = (1+\dfrac{1}{2n})^{4n}$

(i) $a_n = (\dfrac{8n}{3n+1})^3$

(j) $a_n = \dfrac{8n^5+9n^3+7}{n^6+3n-3}$

(k) $a_n = \left(\dfrac{n+3}{n-2}\right)^n$

(l) $a_n = \dfrac{x^n}{n!}, \quad x \in \mathbb{R}$

(m) $a_{n+1} = \sqrt{3a_n}, \quad a_0 = 1$

(n) $a_{n+1} = \dfrac{1}{2}(a_n + \dfrac{2}{a_n}), \quad a_0 = 2.$

(3) (*) Man zeige die Konvergenz der Folgen (a_n) und bestimme den Grenzwert a. Hinweis: Man schlage in Analysisbüchern nach.

$$a_n = \sqrt[n]{a}, \quad a > 0, \qquad a_n = \sqrt[n]{n}, \qquad a_n = \left(1 + \tfrac{z}{n}\right)^n, \qquad z > 0.$$

(4) Man untersuche die folgenden Reihen auf Konvergenz. Hinweis: Man mache die Probe mit *Mathematica*.

$$\sum_{n=0}^{\infty} \frac{n+1}{2n+1}, \ \sum_{n=1}^{\infty} \Big(\frac{n^2-5n+2}{n^4}\Big), \ \sum_{n=1}^{\infty} (-1)^n \Big(\frac{1}{\sqrt{n}} - \frac{1}{n}\Big), \ \sum_{k=0}^{\infty} k\, x^k, \ |x| < 1.$$

(5) Man berechne die Partialsummen und die Grenzwerte der Reihen

$$\sum_{n=1}^{\infty} \frac{1}{n(n+2)}, \qquad \sum_{k=0}^{\infty} k\, q^k, \ |q| < 1.$$

7.3. Stetigkeit

Grob und anschaulich gesprochen heisst eine Funktion stetig, wenn sie *keine Sprünge* macht. Zum Beispiel ist es demnach klar, dass die Funktion $f : \mathbb{R} \to \mathbb{R},\ x \mapsto x^2$ (Abb. 7.2) stetig in jedem Punkt a ist.

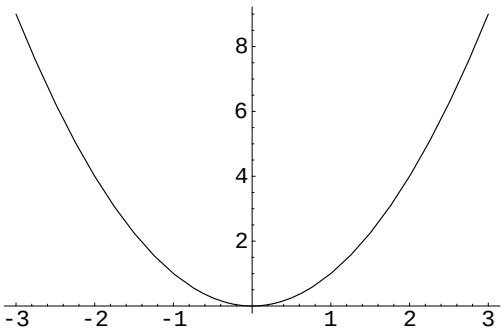


Abbildung 7.2. Parabel $f(x) = x^2$

Anschaulich ist auch die Stetigkeit bzw. Nicht-Stetigkeit folgender Funktionen klar.

(i) Die „Größtes Ganzes“ (floor) -Funktion: $\lfloor\ \rfloor : \mathbb{R} \to \mathbb{Z}, x \mapsto \lfloor x \rfloor$ (Abb. 7.3) ist offensichtlich unstetig in allen Punkten $a \in \mathbb{Z}$.

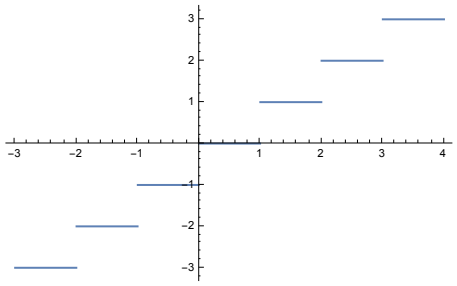


Abbildung 7.3. Floor-Funktion

(ii) Auch die Sägezahnfunktion $f : \mathbb{R} \to \mathbb{R},\ x \mapsto x - \lfloor x \rfloor$ (Abb. 7.4) ist in allen Punkten $a \in \mathbb{Z}$ unstetig.

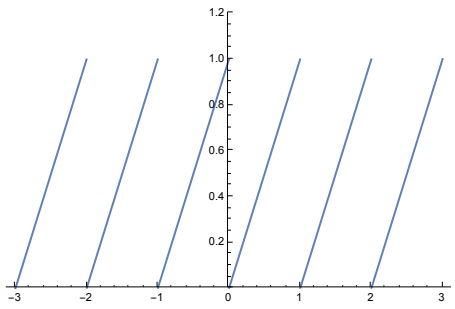


Abbildung 7.4. Sägezahnfunktion

(iii) Schwieriger ist die Entscheidung über Stetigkeit bei der Funktion auf dem Intervall $[-1, 1]$

$$f(x) = \begin{cases} \frac{1}{n} & \frac{1}{n} \leq |x| < \frac{1}{n-1}, \quad n > 1 \\ 0 & x = 0 \end{cases}$$

im Punkt $a = 0$ oder bei
(iv) $f : x \mapsto \sin(\frac{1}{x}), x \neq 0, f(0) := 0$ (Abb. 7.5) im Punkt $a = 0$ oder bei der

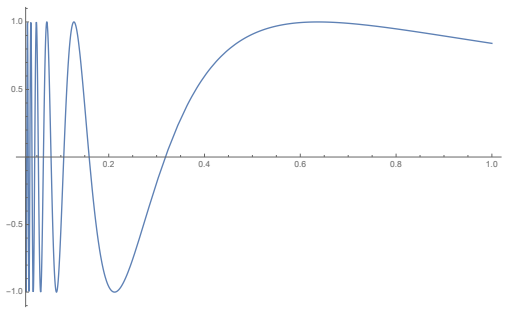


Abbildung 7.5. Stetig im Punkt $a = 0$?

(v) Dirichletfunktion auf $I = [0, 1]$

$$f(x) = \begin{cases} 1 & \text{für } x \text{ rational} \\ 0 & \text{für } x \text{ irrational.} \end{cases}$$

Die Grundidee zur formalen Definition der Stetigkeit ist die Vorstellung, dass sich der Funktionswert $f(x)$ beliebig wenig ändert, falls sich auch x genügend wenig ändert.

Definition 7.36. *Es sei $D \subset \mathbb{R}$ und $f : D \to \mathbb{R}$ eine Funktion und $a \in D$. Dann heisst f* stetig im Punkt *a (stetig an a oder bei a), wenn es zu jedem $\epsilon > 0$ ein $\delta_{\epsilon,a} > 0$ gibt, sodass für alle $x \in D$ gilt*

$$\text{aus} \quad |x - a| < \delta_{\epsilon,a} \quad \text{folgt} \quad |f(x) - f(a)| < \epsilon. \tag{7.24}$$

Die Funktion f heisst stetig, *falls sie in jedem Punkt $a \in D$ stetig ist.*

Anschaulich unter Verwendung von ϵ-Umgebungen kann man dies auch so formulieren: Zu jedem $\epsilon > 0$ gibt es ein $\delta_\epsilon > 0$, sodass

$$\text{aus} \quad x \in U_{\delta_\epsilon}(a) \quad \text{folgt} \quad f(x) \in U_\epsilon(f(a)), \tag{7.25}$$

wobei $U_{\delta_\epsilon}(a)$ der Durchschnitt der δ_ϵ-Umgebung von a mit D ist.

Anmerkung: Wichtig ist, dass es zu jedem $\epsilon > 0$ ein $\delta_\epsilon > 0$ gibt und **nicht umgekehrt**!

Beispiel: Zeige die Stetigkeit der Funktion $f : \mathbb{R} \to \mathbb{R},\ x \mapsto x^2$ im Punkt $a = 1$. Man wähle zuerst $\epsilon = \frac{1}{10}$ und dann allgemein.

Mittels Folgen lässt sich dann Stetigkeit folgendermassen charakterisieren. Eine Funktion ist genau dann stetig in a, wenn für jede Folge die gegen a konvergiert auch die „Bildfolge“ gegen $f(a)$ konvergiert, bzw. genauer

Satz 7.37. *(Folgenstetigkeit). Genau dann ist $f : D \to \mathbb{R}$ stetig bei $a \in D$, wenn folgendes gilt: Für jede Folge $(x_n) \to a, x_n \in D$, gilt $(f(x_n)) \to f(a)$.*

Beweis: [**81**], Seite 54. D.h. es gilt für stetige Funktionen

$$\lim_{n\to\infty} \big(f(x_n)\big) = f\big(\lim_{n\to\infty} x_n\big) = f(a).$$

Anmerkung: Oft wird die Folgenstetigkeit zur Definition von Stetigkeit verwendet. Man beachte aber, dass die Konvergenz von Folgen natürlich auch über Epsilons definiert ist.

Mit Hilfe dieses Satzes lässt sich nun Satz 7.10 unmittelbar für Aussagen über Stetigkeit verwenden.

Satz 7.38. *Sind die Funktionen $f, g : D \to \mathbb{R}$ stetig am Punkt $a \in D$, und sind $\lambda, \mu \in \mathbb{R}$, dann sind auch die Funktionen $\lambda f + \mu g$ und $f \cdot g$ stetig am Punkt a. Ist $f(x) \neq 0$ für alle $x \in D$, so ist auch $\frac{1}{f}$ stetig bei a.*

Ist übrigens $f(a) \neq 0$, so folgt, dass auch $f(x) \neq 0$ in einer Umgebung von a sein muss.

Beispiel: Die identische Funktion $\mathrm{Id} : \mathbb{R} \to \mathbb{R},\ x \mapsto x$ ist in jedem Punkt $a = x$ stetig. Man wähle $\delta_\epsilon = \epsilon$.

Aus Satz 7.38 folgt nun sofort mit Satz 7.10 die Stetigkeit aller Polynomfunktionen p und rationalen Funktionen r

$$p(x) = \sum_{k=0}^{n} a_k\, x^k,$$
$$r(x) = \frac{f(x)}{g(x)},$$

wobei die rationale Funktion r auf $D = \{x \in \mathbb{R} \mid g(x) \neq 0\}$ definiert ist.

Satz 7.39. *Die Zusammensetzung (Verkettung) stetiger Funktionen ist stetig.*

Beweis: [**81**], Seite 56.

7.3.1. Grenzwerte von Funktionen. Wir wollen nun den Grenzwertbegriff in Verbindung mit dem Funktionsbegriff bringen. Dazu brauchen wir zuerst den Begriff des Berührungspunktes.

Definition 7.40. *Es sei $D \subset \mathbb{R}$. Ein Punkt a heisst* Berührungspunkt *von D, wenn es eine Folge $(a_n), a_n \in D$ gibt mit $\lim_{n\to\infty} a_n = a$.*

Jeder Punkt in D ist offensichtlich ein Berührungspunkt von D, aber auch Punkte die nicht zu D gehören, können Berührungspunkte sein, z.B. sind für $D = (-1, 0) \cup (0, 1)$ die Punkte $\{-1, 0, 1\}$ Berührungspunkte.

Definition 7.41. *Gegeben sei eine Funktion $f : D \to \mathbb{R}, D \subset \mathbb{R}$ und a ein Berührungspunkt von D.*
(i) Man sagt die Funktion f hat den Grenzwert c im Berührungspunkt a

$$\lim_{x\to a} f(x) = c \tag{7.26}$$

falls für jede Folge $(x_n), x_n \in D$ mit $\lim_{n\to\infty} x_n = a$ gilt

$$\lim_{n\to\infty} f(x_n) = c.$$

(ii) Weitere Bezeichnungen:

$$\lim_{x\downarrow a} f(x) = c$$

heisst rechtsseitiger Limes*, falls für jede Folge $(x_n), x_n \in D, x_n > a$ mit $\lim_{n\to\infty} x_n = a$ gilt*

$$\lim_{n\to\infty} f(x_n) = c$$

(iii) und

$$\lim_{x\uparrow a} f(x) = c$$

heisst linksseitiger Limes*, falls für jede Folge* $(x_n), x_n \in D, x_n < a$ *mit* $\lim_{n\to\infty} x_n = a$ *gilt*

$$\lim_{n\to\infty} f(x_n) = c.$$

Statt $\uparrow$ *und* $\downarrow$ *schreibt man oft auch* $\searrow$ *und* $\nearrow$*.*

(iv) Ist der Definitionsbereich nach oben unbeschränkt, schreibt man

$$\lim_{x\to\infty} f(x) = c$$

falls für jede Folge $(x_n), x_n \in D$*, mit* $\lim_{n\to\infty} x_n = \infty$ *gilt*

$$\lim_{n\to\infty} f(x_n) = c.$$

Analog ist $\lim_{x\to-\infty} f(x) = c$ *definiert.*

Anmerkung: Man kann damit links-, bzw. rechtsseitige Stetigkeit definieren.

Beispiele: (i) $f : \mathbb{R} \to \mathbb{Z},\ x \mapsto \lfloor x \rfloor$

$$\lim_{x\downarrow 1} \lfloor x \rfloor = 1, \qquad \lim_{x\uparrow 1} \lfloor x \rfloor = 0.$$

(ii) Rationale Funktion

$$\lim_{x\to 1} \frac{x^2-1}{x-1} = 2, \qquad 1 \notin D.$$

Anmerkung: Man kann diese Funktion zu einer stetigen Funktion auf ganz $\mathbb{R}$ ergänzen indem man für $x = 1$, $f(1) := 2$ definiert.

Definition 7.42. *Ein abgeschlossenes endliches Intervall* $I = [a, b] \subset \mathbb{R}$ *heisst* kompakt.

Anmerkung: Eine allgemeinere Definition kompakter Mengen findet man z. B. in [**82**].

Mit dem Satz von Bolzano-Weierstrass 7.16 folgt daher, dass jede Folge auf einem kompakten Intervall eine konvergente Teilfolge hat.

Satz 7.43. *Es sei* K *ein kompaktes Intervall und* $f : K \to \mathbb{R}$ *eine stetige Funktion. Dann ist* f *beschränkt und nimmt auf* K *ein Maximum und ein Minimum an.*

Beweis: [**81**], Seite 58.

Sind diese Voraussetzungen nicht erfüllt, so gilt dies nicht wie man am Beispiel der Funktion $x \mapsto \frac{1}{x}$, $D = (0, 1]$ sieht. Diese ist auf D unbeschränkt und hat zwar ein Minimum aber kein Maximum.

Satz 7.44. *(Zwischenwertsatz) Eine stetige Funktion $f : [a, b] \to \mathbb{R}$ nimmt jeden Wert zwischen $f(a)$ und $f(b)$ an.*

Beweis: [**81**], Seite 59.

Anschaulich besagt dieser Satz, dass ein Intervall keine Löcher haben kann.

Daraus folgt zum Beispiel, dass das Polynom $f(x) = x^k - a, a > 0$ eine Nullstelle auf $[0, a + 1]$ haben muss, da $f(0) = -a < 0$ und $f(a + 1) = (a + 1)^k - a \geq 1 + (k - 1)a > 0$ ist. Daher hat jede positive reelle Zahl a eine reelle k-te Wurzel.

Mit dem Zwischenwertsatz folgt auch

Satz 7.45. *Jedes reelle Polynom ungeraden Grades hat eine reelle Nullstelle (auch Wurzel genannt).*

Für Polynome geraden Grades braucht dies nicht zu gelten wie man an dem Polynom $f(x) = x^2 + 1$ sieht.

Weiters gilt

Satz 7.46. *Das Bild eines Intervalles unter einer stetigen reellen Funktion ist wieder ein Intervall.*

Satz 7.47. *Eine stetige reelle Funktion auf einem Intervall ist genau dann injektiv, wenn sie streng monoton ist.*

Beweis: [**81**], Seite 61.

Mit diesen Vorbereitungen kann man nun den Satz über die Existenz der Umkehrfunktion formulieren.

Satz 7.48. *Es sei D ein Intervall und $f : D \to \mathbb{R}$ stetig und injektiv. Dann ist $f(D) =: C$ ein Intervall, die Funktion f ist streng monoton, und f besitzt eine stetige und streng monotone Umkehrabbildung auf C.*

Beweis: [**81**], Seite 62.

Als Anwendung betrachten wir die Funktion $x \mapsto x^n, D = [0, \infty)$. Diese ist stetig und streng monoton wachsend auf D. Deshalb existiert für $x \geq 0$ die Umkehrfunktion $x \mapsto x^{\frac{1}{n}}$.

Abschliessend wollen wir noch den Begriff der Stetigkeit etwas verschärfen[2].

Bei der Stetigkeit einer Funktion f im Punkt a war δ_ϵ abhängig vom gewählten ϵ und auch für verschiedene Punkte a verschieden. Ein Beispiel ist die Funktion $f : x \mapsto x^{-1}, D = (0, 1)$.

[2]Eine weitere Verschärfung des Stetigkeitigkeitsbegriffes ist die Lipschitzstetigkeit: `http://de.wikipedia.org/wiki/Lipschitz-Stetigkeit`

Definition 7.49. *Eine Funktion $f : D \to \mathbb{R}$ heisst* gleichmässig stetig, *falls gilt: Zu jedem $\epsilon > 0$ gibt es ein $\delta_\epsilon > 0$, sodass für alle $a, x \in D$ gilt:*

$$aus \quad |x - a| < \delta_\epsilon \quad folgt \quad |f(x) - f(a)| < \epsilon. \tag{7.27}$$

Satz 7.50. *Jede in einem kompakten Intervall stetige Funktion ist dort gleichmäßig stetig.*

Beweis: siehe [**92**]

Im Gegensatz zur Stetigkeit wo $\delta(a, \epsilon)$ von a und ϵ abhängt, darf hier $\delta(\epsilon)$ nur von ϵ abhängen nicht aber von a.

7.3.2. Folgen und Reihen von Funktionen. Wir wollen nun nicht nur einfache Zahlenfolgen betrachten, sondern allgemeiner Folgen bei denen jeder natürlichen Zahl eine Funktion zugeordnet wird. Das n-te Folgenglied f_n ist nun keine reelle Zahl sondern eine Funktion $f_n : D \to \mathbb{R}$. Man schreibt in diesem Fall auch $(f_n)_{n \geq k}$ oder (f_n) oder lässt die Klammern ganz weg und schreibt nur f_n.

Für Folgen von Funktion kann man Konvergenz auf verschiedene Arten definieren, da verschiedene Abstandsbegriffe eingeführt werden können. Eine naheliegende Definition ist:

Definition 7.51. *Die Folge von Funktion (f_n)* konvergiert punktweise *gegen $f : D \to \mathbb{R}$, wenn für jedes $x \in D$, die reelle Zahlenfolge $(f_n(x))$ gegen die reelle Zahl $f(x)$ konvergiert.*

D. h. Zu jeder reellen Zahl $\epsilon > 0$ und jedem x existiert eine Zahl $N_\epsilon \in \mathbb{N}$, sodass

$$|f_n(x) - f(x)| < \epsilon \tag{7.28}$$

für alle Indizes $n \geq N_\epsilon$ ist. Hier ist nun N_ϵ nicht nur von ϵ sondern auch von x abhängig, also $N_\epsilon(x)$.

Es zeigt sich jedoch, dass diese naheliegende Definition nicht garantiert, dass die Funktionen f_n die Grenzfunktion gut annähert.

Beispiel 1:

$$f_n(x) = \begin{cases} n^2 x & \text{für} \quad 0 \leq x \leq \frac{1}{n} \\ 2n - n^2 x & \text{für} \quad \frac{1}{n} \leq x \leq \frac{2}{n} \\ 0 & \text{für} \quad 0 \geq x \text{ oder } x \geq \frac{2}{n}. \end{cases}$$

Die Grenzfunktion ist die identisch verschwindende Funktion obwohl die Fläche unter der Kurve f_n (Abb. 7.6) für alle n stets 1 ergibt.

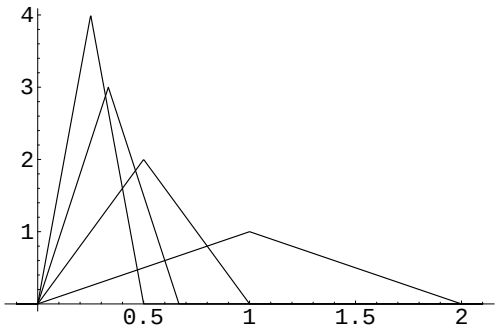


Abbildung 7.6. Funktionenfolge die punktweise gegen 0 konvergiert

Beispiel 2:

$$f_n(x) = \frac{nx}{1+|nx|} \to \begin{cases} 1 & \text{für} \quad x > 0 \\ 0 & \text{für} \quad x = 0 \\ -1 & \text{für} \quad x < 0. \end{cases}$$

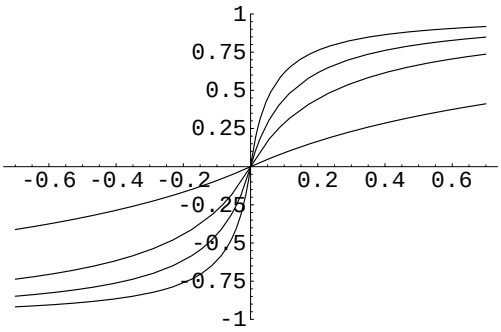


Abbildung 7.7. stetige Funktionenfolge

Beispiel 2 (Abb. 7.7) zeigt dass der Grenzwert einer stetigen Funktion bei punktweiser Konvergenz nicht eine stetige Funktion sein muss.

Um diesen Mangel zu beheben führen wir den Begriff der gleichmässigen Konvergenz ein.

Definition 7.52. *Die Folge von Funktion* (f_n) konvergiert gleichmässig *gegen* $f : D \to \mathbb{R}$, *wenn gilt: Zu jeder reellen Zahl* $\epsilon > 0$ *existiert eine Zahl* $N_\epsilon \in \mathbb{N}$, *sodass für alle Indizes* $n \geq N_\epsilon$ *und alle* $x \in D$ *zugleich gilt*

$$|f_n(x) - f(x)| < \epsilon. \tag{7.29}$$

Man kann auf Funktionenräumen folgende Norm (**Supremumsnorm**) einführen:

$$||f||_\infty := \sup\{|f(x)| \mid x \in D\}. \tag{7.30}$$

Gleichmässige Konvergenz einer Folge von Funktionen (f_n) bedeutet nun einfach Konvergenz unter Benutzung der von der Supremumsnorm erzeugten Metrik, d. h. für jedes $\epsilon > 0$ existiert ein Index N_ϵ, sodass für alle $n > N_\epsilon$ gilt

$$||f_n - f||_\infty < \epsilon.$$

Zur Vorbereitung der Definition des Begriffs *Bestimmtes Integral* kann man nun zeigen

Satz 7.53. *Konvergiert eine Folge stetiger Funktionen (f_n) gleichmässig gegen eine Funktion f, so ist auch f stetig.*

Beweis: [**81**], Seite 70.

Definition 7.54. *Eine* Zerlegung *(Unterteilung) Z eines kompakten Intervalles $[a, b]$ ist ein $(n+1)$-Tupel $(z_0, \ldots, z_n)$ von Zahlen, sodass*

$$a = z_0 < z_1 < \cdots < z_n = b. \tag{7.31}$$

Eine Funktion $f : [a, b] \to \mathbb{R}$ heisst Treppenfunktion *(Stufenfunktion), wenn es eine Zerlegung Z von $[a, b]$ gibt, sodass gilt*

$$f(x) = c_k \qquad \text{für} \quad z_{k-1} < x < z_k, \tag{7.32}$$

wobei $c_k, 1 \leq k \leq n$ reelle Konstanten sind.

Die Werte der Funktion an den Stützstellen z_j kann beliebig gewählt werden.

Satz 7.55. *Jede stetige Funktion auf einem kompakten Intervall ist ein gleichmässiger Limes von einer Folge von Treppenfunktionen.*

Beweis: [**81**], Seite 72.

Definition 7.56. *Jede Funktion, die ein gleichmässiger Grenzwert eine Folge von Treppenfunktionen ist, heisst* **Regelfunktion**.

Stetige Funktionen sind also Regelfunktionen, aber auch monotone oder Treppen-Funktionen sind Regelfunktionen.

7.4. Übung

(1) Es sei $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(x) = x^3 + 2.$$

Man zeige, dass f stetig im Punkt $a = 1$ ist. Man gebe für $\epsilon > 0$ und dann für $\epsilon = 10^{-1}$ ein entsprechendes δ an.

(2) Es sei $f: \ I = (\frac{1}{2}, \infty) \to \mathbb{R}$ mit

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

Man zeige, dass f stetig im Punkt $a = 1$ ist. Man gebe für $\epsilon > 0$ und dann für $\epsilon = 10^{-1}$ ein entsprechendes δ an.

(3) (*) Die Funktion $f : \mathbb{Q} \to \mathbb{R}$ ist stetig, wobei

$$f(x) = \begin{cases} 0, \text{ wenn } & x < \sqrt{2} \\ 1, \text{ wenn } & x > \sqrt{2} \end{cases} .$$

(4) Man zeige: es sei $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(x) = \begin{cases} \dfrac{x^2-1}{x-1}, & \text{wenn } \ x \in \mathbb{R} \setminus \{1\} \\ 2, & \text{wenn } \ x = 1 \end{cases},$$

dann ist f stetig in $a = 1$. Hinweis: Man verwende Satz 7.37.

(5) Man untersuche die Funktion $f : \mathbb{R} \to \mathbb{R}$ mit

$$f(x) = \begin{cases} x^m \sin \frac{1}{x}, & \text{wenn } \ x \neq 0 \\ 0, & \text{wenn } \ x = 0 \end{cases}, \qquad m \in \{0, 1\}$$

auf Stetigkeit im Punkt $x = 0$.

(6) Es sei $0 \leq a < b < \infty$. Man berechne den Grenzwert $x \to b$ von

$$f(x) = \frac{x^m - b^m}{x - b}, \qquad m \in \mathbb{N}.$$

(7) Man bestimme folgende Grenzwerte

$$\lim_{x \downarrow 2} \frac{2x+1}{x^2 - 3x + 2}, \quad \lim_{x \to \infty} \left(\frac{3x}{x-1} - \frac{2x}{x+1} \right), \quad \lim_{x \downarrow 0} \frac{(1+x)^3 - 2}{x},$$
$$\lim_{x \to 1} \left(\frac{1}{x+3} - \frac{2}{3x+5} \right) \frac{1}{x-1}.$$

(8) Man zeige, dass die Funktion $f : D = (0, 1) \to \mathbb{R}$, $f(x) = x^2$ gleichmäßig stetig und die Funktion $f(x) = 1/x$ nicht gleichmäßig stetig auf D ist.

7.5. Elementare Funktionen

7.5.1. Polynome und rationale Funktionen. Eine Funktion $p : \mathbb{R} \to \mathbb{R}$

$$p(x) := \sum_{j=0}^{n} a_j x^j, \qquad a_j \in \mathbb{R}, a_n \neq 0$$

heisst *Polynom* vom Grade n.

Beispiele: $f(x) = 2x + 1$, $g(x) = x^2$ und $h(x) = x^3 - 2x^2 - x + 2$ (Abb. 7.8).

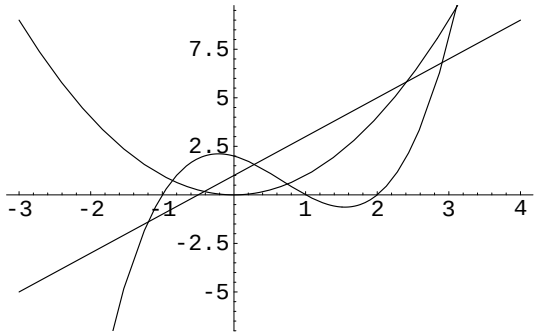


Abbildung 7.8. Polynome

Satz 7.57. *Ein Polynom vom Grade n hat genau n (möglicherweise komplexe) Nullstellen, wobei jede Nullstelle so oft gezählt wird wie sie in der Zerlegung*

$$p(x) = \sum_{j=0}^{n} a_j x^j = a_n (x - x_1)(x - x_2) \cdots (x - x_n)$$

vorkommt (Vielfachheit).

Will man nur reelle Faktoren erhalten, ist nur eine Zerlegung in lineare und quadratische Faktoren möglich, wobei die quadratischen Faktoren keine reelle Nullstellen haben.

$$p(x) = \sum_{j=0}^{n} a_j x^j = a_n (x - x_1) \cdots (x - x_r)(x^2 + b_1 x + c_1) \cdots (x^2 + b_s x + c_s).$$

Sind f und g zwei Polynome, so bezeichnet man die Funktion $r : D \to \mathbb{R}$, $D = \{x \in \mathbb{R} \mid g(x) \neq 0\}$

$$r(x) = \frac{f(x)}{g(x)}$$

als *rationale Funktion.*

Beispiele: $f(x) = \frac{x+1}{x-1}$ und $g(x) = \frac{x^2-1}{x-1}$ (Abb. 7.9).

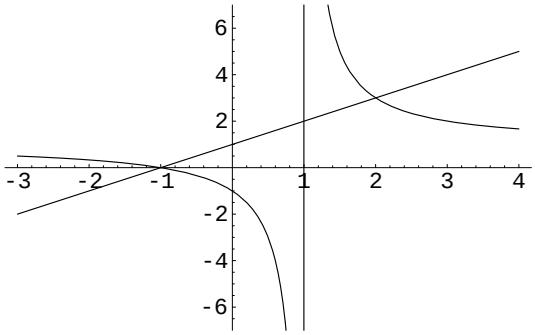


Abbildung 7.9. rationale Funktionen

Satz 7.58. *Polynome und rationale Funktionen sind stetige Funktionen.*

Besonders klar wird die Darstellung rationaler Funktionen in der Partialbruchzerlegung. Eine rationale Funktion ist *vollständig* durch ihr Verhalten an den Polstellen bestimmt.

Satz 7.59. (Partialbruchzerlegung) *Es sei $r(x) = \frac{f(x)}{g(x)}$ eine rationale Funktion mit Grad $f(x) <$Grad $g(x)$ und die Zerlegung von $g(x)$ sei $a_n(x-x_1)^{\alpha_1}\cdots(x-x_r)^{\alpha_r}(x^2+b_1x+c_1)^{\beta_1}\cdots(x^2+b_sx+c_s)^{\beta_s}$. Dann gibt es reelle Koeffizienten A_{jk}, B_{jk} und C_{jk}, sodass*

$$\frac{f(x)}{g(x)} = \sum_{j=1}^{r}\left(\sum_{k=1}^{\alpha_j}\frac{A_{jk}}{(x-x_j)^k}\right) + \sum_{j=1}^{s}\left(\sum_{k=1}^{\beta_j}\frac{B_{jk}x+C_{jk}}{(x^2+b_jx+c_j)^k}\right) \tag{7.33}$$

eindeutig bestimmt ist.

Die praktische Berechnung der Partialbrüche erfolgt durch Ansatz und Bestimmung der Konstanten durch Koeffizientenvergleich.

Anmerkung: Lässt man komplexe Wurzeln zu, so vereinfacht sich die Darstellung, da dann die Terme mit $(x^2+b_jx+c_j)^k$ wegfallen.

Beispiele: (i)

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

In diesem Fall ist $x_1 = 0$ und $x_2 = 1$ und der Ansatz lautet daher

$$\frac{1}{x(x-1)} = \frac{A_{11}}{x} + \frac{A_{21}}{(x-1)}.$$

Auf gemeinsamen Nenner gebracht ergibt sich

$$1 = A_{11}(x-1) + A_{21}x.$$

Koeffizientenvergleich der Potenzen von x ergibt

$$1 = -A_{11},$$
$$0 = -A_{11} + A_{21}$$

und damit folgt $A_{11} = -1, A_{21} = 1$.

(ii)

$$h(x) = \frac{x^2 + x + 2}{x^3 - 2\,x^2 + x}.$$

In diesem Fall ist $x_1 = 0$ und $x_2 = 1$ und der Ansatz lautet

$$\frac{x^2 + x + 2}{x^3 - 2\,x^2 + x} = \frac{A_{11}}{x} + \frac{A_{21}}{(x-1)} + \frac{A_{22}}{(x-1)^2}.$$

Auf gemeinsamen Nenner gebracht ergibt sich

$$x^2 + x + 2 = A_{11}(x-1)^2 + A_{21}x(x-1) + A_{22}x.$$

Koeffizientenvergleich der Potenzen von x ergibt

$$1 = A_{11} + A_{21},$$
$$1 = -2\,A_{11} - A_{21} + A_{22},$$
$$2 = A_{11}$$

und damit folgt $A_{11} = 2, A_{21} = -1, A_{22} = 4$.

Anmerkung: es müssen nicht alle A_{jk} ungleich Null sein.

7.5.2. Exponentialfunktion. Mit Hilfe des Quotientenkriteriums können wir zeigen, dass die Reihe

$$\sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$$

nicht nur für jedes $z \geq 0$ konvergiert, sondern für alle $z \in \mathbb{C}$ absolut konvergiert.

Beweis: sei $n \geq 2\,|z|$, dann gilt

$$\left|\frac{a_{n+1}}{a_n}\right| = \left|\frac{z^{n+1}}{(n+1)!}\frac{n!}{z^n}\right| = \frac{|z|}{(n+1)} \leq \frac{1}{2}. \tag{7.34}$$

Anmerkung: Eine konvergente komplexe Folge $u_n = a_n + \mathrm{i}b_n$, $a_n, b_n \in \mathbb{R}$ konvergiert genau dann, wenn sowohl Realteil als auch Imaginärteil konvergieren und es gilt: $\lim u_n = \lim a_n + \mathrm{i}\lim b_n$.

Wir definieren daher

Definition 7.60. *Die* Exponentialfunktion $\exp : \mathbb{C} \to \mathbb{C}, z \mapsto \exp(z)$ *ist definiert durch*

$$\exp(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}, \qquad z \in \mathbb{C}.$$

Die Eulersche Zahl e *ist definiert als* $\mathrm{e} = \exp(1) = 2.7182818\ldots$

Satz 7.61. *Für alle* $z_1, z_2 \in \mathbb{C}$ *gilt:*

$$\exp(z_1 + z_2) = \exp(z_1)\exp(z_2). \tag{7.35}$$

Beweis: Ausmultiplizieren mittels Cauchy-Produkt.

Satz 7.62.

(i) Für alle $x \in \mathbb{R}$ *gilt:* $\exp(x) > 0$ *und für alle* $z \in \mathbb{C}$ *gilt:* $\exp(z) \neq 0$.

(ii) Für alle $r \in \mathbb{Q}$ *gilt:* $\exp(r) = \mathrm{e}^r$.

(iii) Für alle $z \in \mathbb{C}$ *gilt:* $\exp(-z) = \frac{1}{\exp(z)}$.

(iv) Für alle $x \in \mathbb{R}$ *gilt:* $\exp(x)$ *ist streng monoton wachsend und*

$$\lim_{x\to\infty} \mathrm{e}^x = \infty, \qquad \lim_{x\to-\infty} \mathrm{e}^x = 0, \qquad \lim_{x\to\infty} \frac{\mathrm{e}^x}{x^n} = \infty, \ n \in \mathbb{N}. \tag{7.36}$$

(v) Die reelle Exponentialfunktion ist eine bijektive Abbildung von $\mathbb{R} \to \mathbb{R}^+$.

Beweis: [**86**], Seite 103ff.

Das heisst die Exponentialfunktion wächst stärker als jede Potenz (Abb. 7.10).

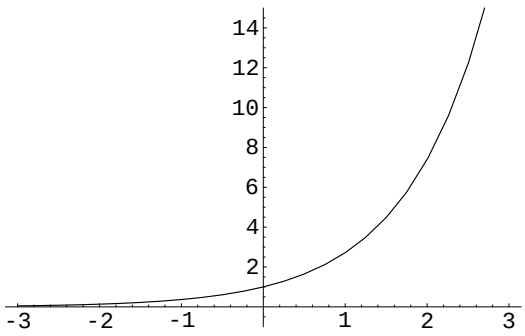


Abbildung 7.10. Exponentialfunktion

Satz 7.63. *Die Exponentialfunktion ist stetig sowohl in* $\mathbb{C}$ *als auch in* $\mathbb{R}$.

Beweis: [**83**], Seite 126.

Satz 7.64. *Für alle $z \in \mathbb{C}$ gilt:*

$$\lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n = \mathrm{e}^z. \tag{7.37}$$

Anmerkung: Allgemeiner gilt für jede Nullfolge $(x_n) \to 0$

$$\lim_{n\to\infty} \left(1 + z x_n\right)^{\frac{1}{x_n}} = \mathrm{e}^z. \tag{7.38}$$

7.5.3. Logarithmusfunktion, allgemeine Potenzfunktion. Da die reelle Exponentialfunktion eine bijektive Abbildung von $\mathbb{R} \to \mathbb{R}^+$ ist, existiert laut Satz 7.48 die Umkehrfunktion.

Definition 7.65. *Die* Logarithmusfunktion log *(veraltete Schreibweise* ln*) ist die Umkehrfunktion der Exponentialfunktion* exp

$\log : \ \mathbb{R}^+ \to \mathbb{R}, \quad x \mapsto \log(x).$

Es folgt daher, dass $\log(\mathrm{e}^x) = x, \ x \in \mathbb{R}$ und $\mathrm{e}^{\log x} = x, \ x \in \mathbb{R}^+$.

Graphisch erhält man die Logarithmusfunktion (wie jede Umkehrfunktion) indem man die Exponentialfunktion an der 45°-Achse spiegelt (Abb. 7.11).

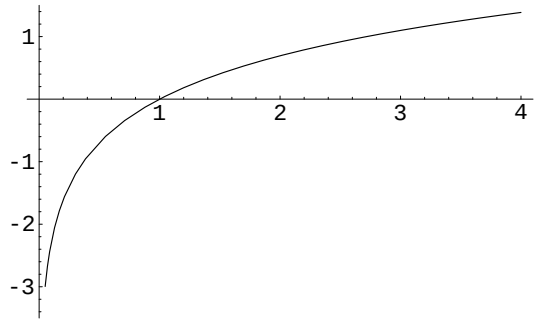


Abbildung 7.11. Logarithmusfunktion

Satz 7.66. *Für die Logarithmusfunktion gilt:*

(i) $\log(x\,y) = \log x + \log y$

(ii) $\lim_{x\downarrow 0} \log x = -\infty, \quad \lim_{x\to\infty} \log x = \infty, \quad \lim_{x\to\infty} \frac{\log x}{x^\epsilon} = 0, \ \epsilon > 0.$

(iii) $\log 1 = 0, \log \mathrm{e} = 1.$

Mit Hilfe der Logarithmusfunktion kann man nun die allgemeine Exponentialfunktion und Potenzfunktion definieren.

Definition 7.67. *Es sei $a \in \mathbb{R}, a > 0$. Dann ist die Exponentialfunktion zur Basis a $a^x : \mathbb{R} \to \mathbb{R}$ definiert*

$$a^x := \mathrm{e}^{x \log(a)}.$$

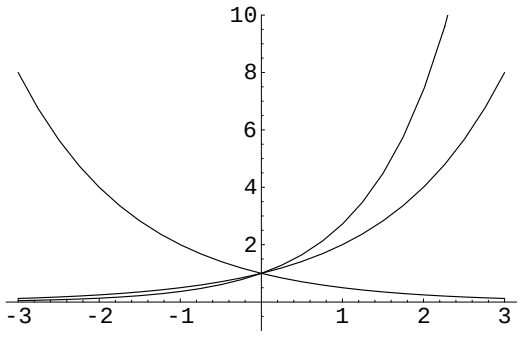


Abbildung 7.12. allgemeine Exponentialfunktion $2^x, \mathrm{e}^x, \frac{1}{2^x}$

Bislang waren nur Potenzfunktionen mit rationalen Exponenten definiert und wir wollen dies nun für beliebige reelle Exponenten nachholen.

Definition 7.68. *Es sei $a \in \mathbb{R}$. Dann ist die allgemeine Potenzfunktion (Abb. 7.13) $x^a : \mathbb{R}_0^+ \to \mathbb{R}$ definiert*

$$x^a := \mathrm{e}^{a \log(x)}.$$

Anmerkung: Man kann die allgemeine Potenz a^b auch als stetige Fortsetzung der Exponentialfunktion $a^x : \mathbb{Q} \to \mathbb{R}$ definieren, z. B. $3^{\sqrt{2}}$ ist der Limes einer Folge 3^{a_n}, wobei a_n eine Folge rationaler Zahlen, die gegen $\sqrt{2}$ konvergiert, ist. Für den Beweis, dass dies so möglich ist braucht man aber den Begriff der Lipschitz-Stetigkeit[3], z.B. [**92**] Band 1.

Für die allgemeine Exponentialfunktion folgt

Satz 7.69. *(i) Die Exponentialfunktion $f(x) = a^x$ ist streng monoton wachsend für $a > 1$ und streng monoton fallend für $0 < a < 1$ (Abb. 7.12).*

Für alle $x \in \mathbb{R}$ gilt:

(ii) $(a^x)^y = a^{xy}$.

(iii) $a^x a^y = a^{x+y}, \quad a^{-x} = \frac{1}{a^x}, \quad .$

[3] http://de.wikipedia.org/wiki/Lipschitz-Stetigkeit

(iv) $a^x b^x = (ab)^x$.

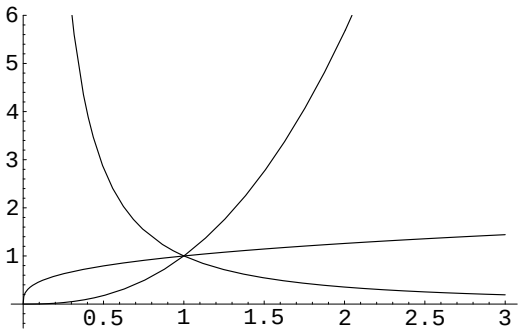


Abbildung 7.13. allgemeine Potenzfunktion $x^{\frac{5}{2}}, x^{\frac{1}{3}}, x^{-\frac{3}{2}}$

Die Logarithmusfunktion zu einer beliebigen Basis $\log_a, a > 0$ wird als Umkehrfunktion der Exponentialfunktion mit Basis a definiert. Es gilt offensichtlich:

$$\log_a(x) = \frac{\log x}{\log a}, \ x > 0.$$

7.5.4. Hyperbelfunktionen.

Definition 7.70. *Eine reelle Funktion heisst* gerade, *wenn für alle* $x \in \mathbb{R}$ *gilt:* $f(-x) = f(x)$ *(symmetrisch bezüglich der y-Achse).*

Eine reelle Funktion heisst ungerade, *wenn für alle* $x \in \mathbb{R}$ *gilt:* $f(-x) = -f(x)$ *(symmetrisch bezüglich der y-Achse nach Spiegelung an der x-Achse).*

Jede beliebige Funktion kann als Summe einer geraden und ungeraden Funktion geschrieben werden.

$$f = f_u + f_g, \quad f_u(x) = \frac{1}{2}(f(x) - f(-x)), \quad f_g(x) = \frac{1}{2}(f(x) + f(-x)). \quad (7.39)$$

Definition 7.71. *Die Hyperbelfunktionen* sinh *(Sinus Hyperbolicus),* cosh*, (Cosinus Hyperbolicus) (Abb. 7.14) sind Abbildungen* $\mathbb{R} \to \mathbb{R}$ *definiert durch*

$$\cosh x := \frac{1}{2}(\mathrm{e}^x + \mathrm{e}^{-x}),$$
$$\sinh x := \frac{1}{2}(\mathrm{e}^x - \mathrm{e}^{-x}), \quad (7.40)$$

Ein durchhängendes Seil nimmt die Form einer Kettenlinie[4] (Cosinus Hyperbolicus) ein.

[4] http://de.wikipedia.org/wiki/Kettenlinie_(Mathematik)

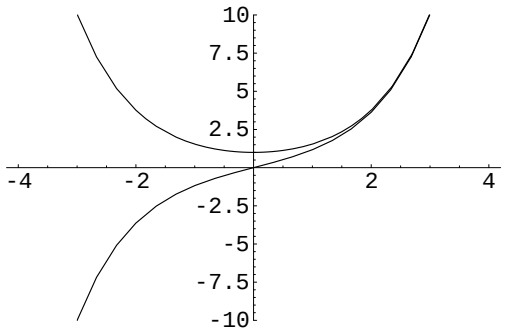


Abbildung 7.14. Sinus hyperbolicus, Cosinus hyperbolicus

Eigenschaften: eine einfache Rechnung ergibt

$$\begin{aligned}
&\cosh^2 x - \sinh^2 x = 1,\\
&\cosh(x+y) = \cosh x \cosh y - \sinh x \sinh y,\\
&\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y.
\end{aligned} \tag{7.41}$$

Des weiteren definiert man dann

Definition 7.72. *Die Hyperbelfunktionen* $\tanh, \coth$ *(Abb. 7.15) sind Abbildungen* $\mathbb{R} \to \mathbb{R}$ *definiert durch*

$$\begin{aligned}
\tanh : \mathbb{R} \to \mathbb{R}, \qquad \tanh x &:= \frac{\sinh x}{\cosh x} = \frac{\mathrm{e}^x - \mathrm{e}^{-x}}{\mathrm{e}^x + \mathrm{e}^{-x}},\\
\coth : \mathbb{R} \setminus \{0\} \to \mathbb{R}, \qquad \coth x &:= \frac{\cosh x}{\sinh x} = \frac{\mathrm{e}^x + \mathrm{e}^{-x}}{\mathrm{e}^x - \mathrm{e}^{-x}}.
\end{aligned} \tag{7.42}$$

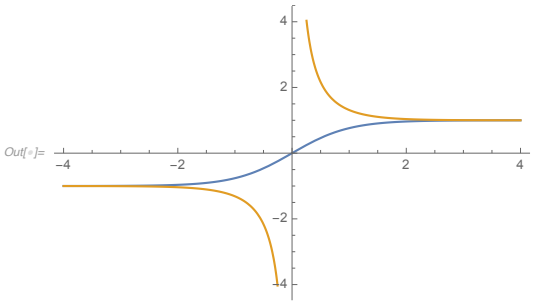


Abbildung 7.15. Tangens hyperbolicus, Cotangens hyperbolicus

Die entsprechenden Umkehrfunktionen auf den streng monotonen Bereichen sind die Area-Funktionen

Definition 7.73.

$$\begin{aligned} \operatorname{Arcosh} &: [1,\infty) \to \mathbb{R}, & \operatorname{Arcosh} y &:= \log(y + \sqrt{y^2-1}), \\ \operatorname{Arsinh} &: \mathbb{R} \to \mathbb{R}, & \operatorname{Arsinh} y &:= \log(y + \sqrt{y^2+1}), \\ \operatorname{Artanh} &: (-1,1) \to \mathbb{R}, & \operatorname{Artanh} y &:= \frac{1}{2}\log\frac{1+y}{1-y}. \end{aligned} \tag{7.43}$$

7.5.5. Trigonometrische Funktionen. Betrachtet man die komplexe Exponentialfunktion e^{ix}, $x \in \mathbb{R}$ so sieht man, dass $|e^{ix}| = 1$ ist und daher die Funktionswerte dieser Funktion am Einheitskreis liegen. Geometrisch entspricht der Wert auf der x-Achse dem Cosinus und der Wert auf der y-Achse dem Sinus. (Anmerkung: x ist dabei die Bogenläge am Einheitskreis.)

Man kann daher diese Funktionen **analytisch** wie folgt einführen.

Definition 7.74. *Die Winkelfunktionen* sin *(Sinus),* cos*, (Cosinus) (Abb. 7.16) sind Abbildungen* $\mathbb{R} \to \mathbb{R}$ *definiert durch*

$$\begin{aligned} \cos x &:= \operatorname{Re}(e^{ix}) = \frac{1}{2}(e^{ix} + e^{-ix}) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - + \cdots, \\ \sin x &:= \operatorname{Im}(e^{ix}) = \frac{1}{2i}(e^{ix} - e^{-ix}) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - + \cdots. \end{aligned} \tag{7.44}$$

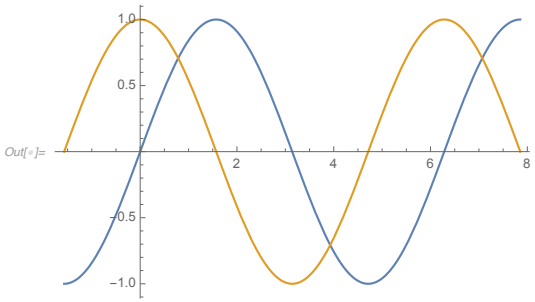


Abbildung 7.16. Sinus, Cosinus

Einfache Rechnung ergibt nun

$$\begin{aligned} e^{ix} &= \cos x + i \sin x \\ \cos^2 x + \sin^2 x &= 1, \\ \cos(x+y) &= \cos x \cos y - \sin x \sin y, \\ \sin(x+y) &= \sin x \cos y + \cos x \sin y. \end{aligned} \tag{7.45}$$

Definition 7.75. *Die Funktionen* $\tan, \cot$ *(Abb. 7.17) sind Abbildungen definiert durch*

$$\tan x := \frac{\sin x}{\cos x}, \qquad x \neq \pi(k + \frac{1}{2}), k \in \mathbb{Z},$$
$$\cot x := \frac{\cos x}{\sin x}, \qquad x \neq \pi k, k \in \mathbb{Z}. \tag{7.46}$$

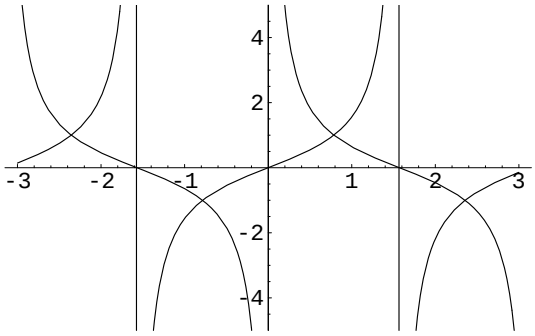


Abbildung 7.17. Tangens, Cotangens

Die entsprechenden Umkehrfunktionen sind die Arcus-Funktionen

Definition 7.76. *Die Funktion* cos *ist im Intervall* $[0, \pi]$ *streng monoton fallend und bildet dieses Intervall bijektiv auf* $[-1, 1]$ *ab. Die Umkehrabbildung*

$$\arccos : [-1, +1] \to \mathbb{R} \tag{7.47}$$

heisst Arcus-Cosinus.

Die Funktion sin *ist im Intervall* $[-\frac{\pi}{2}, \frac{\pi}{2}]$ *streng monoton wachsend und bildet dieses Intervall bijektiv auf* $[-1, 1]$ *ab. Die Umkehrabbildung*

$$\arcsin : [-1, +1] \to \mathbb{R} \tag{7.48}$$

heisst Arcus-Sinus.

Die Funktion tan *ist im Intervall* $(-\frac{\pi}{2}, \frac{\pi}{2})$ *streng monoton wachsend und bildet dieses Intervall bijektiv auf* $\mathbb{R}$ *ab. Die Umkehrabbildung*

$$\arctan : \mathbb{R} \to \mathbb{R} \tag{7.49}$$

heisst Arcus-Tangens.

7.5.6. Ergänzung:

Definition 7.77. *(Landausymbole für Funktionen.) Seien f und g zwei reelle Funktionen. Wenn es Konstanten C und R gibt, sodass*

$$|f(x)| \leq C|g(x)| \qquad \text{für alle } x \geq R,$$

dann schreiben wir $f(x) = O(g(x))$ und sagen, f **ist von der Ordnung** *g oder f* **ist „Groß-O“ von** *g.*

Ist

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0,$$

dann schreiben wir $f(x) = o(g(x))$ und sagen, f **ist „Klein-o“ von** *g.*

Gilt

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1,$$

dann sagen wir f und g sind asymptotisch gleich und schreiben $f \sim g$.

7.6. Übung

(1) Berechne die Partialbruchzerlegung folgender rationaler Funktionen

$$f(x) = \frac{1}{x(x-1)}, \quad g(x) = \frac{x^2+x+1}{x^3-2\,x^2+x}, \quad h(x) = \frac{x^2+x+2}{x^3-2\,x^2+x},$$
$$k(x) = \frac{x^3+x^2+x+3}{x^5-x^4+x^3-x^2}.$$

(2) Ein Unternehmen stellt Gartenzwerge her.
a) Durch Marktanalyse wurde festgestellt, dass bei einem Stückpreis von p ungefähr $x = 200 - 20p$ Stück pro Tag verkauft werden können. Finden Sie den Preis $p(x)$ als Funktion von der Stückzahl.
b) Werden (pro Tag) x Gartenzwerge produziert, dann fallen dabei die Produktionskosten $k(x) = 100 + 4x$ (Fixkosten plus Stückkosten) an. Beim Verkauf von x Gartenzwergen erzielt das Unternehmen dann die Einnahmen $e(x) = x\,p(x)$. Stellen Sie die Funktionen $k(x)$ und $e(x)$ graphisch dar.
c) Beim Verkauf von x Gartenzwergen macht das Unternehmen den Gewinn $g(x) = e(x) - k(x)$. Stellen Sie $g(x)$ graphisch dar.
d) Das Unternehmen arbeitet kostendeckend, wenn $g(x) \geq 0$. In welchem Stückzahlenbereich arbeitet das Unternehmen kostendeckend?
e) Welchen Preis soll das Unternehmen festlegen, damit der Gewinn maximal wird?

(3) Eine Bakterienkultur (anfangs 1000 Bakterien) kann sich ungehemmt vermehren. Angenommen, sie verdreifacht sich jeweils innerhalb einer Stunde. Stellen Sie den Wachstumsprozess durch eine Funktion dar. Wieviele Bakterien sind es nach 24 Stunden?

(4) Radioaktiver Zerfall kann durch $n(t) = n(0) \cdot \mathrm{e}^{-\lambda\,t}$ modelliert werden. Jod 131 hat beispielsweise eine Zerfallskonstante $\lambda = 1 \cdot 10^{-6} s^{-1}$. Angenommen, es sind zu Beginn 10^{10} Kerne vorhanden. Stellen Sie die Anzahl $n(t)$ der noch nicht zerfallenen Kerne als Funktion der Zeit t in Tagen dar. Nach wievielen Tagen vermindert sich die Anzahl der Kerne auf die Hälfte?

(5) Man löse nach x auf (bzw. vereinfache)

$$\log_x \frac{1}{8} = -\frac{3}{2}, \qquad \log_4 x = \frac{3}{4}, \qquad \log_{\frac{1}{x}} \sqrt[3]{x^4} =$$

(6) Man löse nach x auf

(a) $9^{2x-1} = 4^{x+1}\,2^{2x-2}$,

(b) $\log(x-3) - \log(2x+5) = \log(x-6) - \log(3x-16)$,

(c) $\mathrm{e}^x + \mathrm{e}^{-x} = 2\,b, \quad b \geq 1.$

(7) Man bestimme die Konstanten a_0, a_1, a_2 so, dass für

$$f(x) = \log(a_0 + a_1 x + a_2 x^2)$$

gilt: $f(0) = 0,\ f(1) = 1,\ f(2) = 2.$

(8) Man zeige
(a) für $b > 0$ gilt

$$\lim_{x\to 0} \frac{b^x - 1}{x} = \log b$$

(b) und

$$\lim_{x\to 0} \frac{\log(1+x)}{x} = 1.$$

(9) (Mit `Mathematica`) Die Zunahme der Internetanschlüsse in Österreichs Haushalten könnte durch eine sogenannte **logistische Wachstumsfunktion** modelliert werden. Der Ausstattungsgrad $y(t)$ gibt dabei an, wieviele von S Haushalten nach t Jahren im Durchschnitt einen Internetanschluß besitzen: $y(t) = \frac{S}{1+a\,\mathrm{e}^{-b\,t}}$.
a) Zeichnen Sie den Graphen für den Zeitraum $0 \leq t \leq 20$, wenn $S = 100$, $a = 9$ und $b = 0.3$.
b) Nach wievielen Jahren besitzen nach unserem Modell 80% aller Haushalte einen Internetanschluß?

(10) (Mit `Mathematica`) Die Dichtefunktion einer Normalverteilung ist durch

$$f(x) = \mathrm{e}^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \frac{1}{\sqrt{2\pi}\sigma}$$

gegeben. Dabei heißt μ der **Erwartungswert** und σ die **Standardabweichung** der Normalverteilung. Zeichnen Sie die Dichtefunktion für $\mu = 0$ und $\sigma = 1$, für $\mu = 0$ und $\sigma = 2$, sowie für $\mu = 1$ und $\sigma = 1$. Wie wirkt sich die Änderung der Parameter μ und σ auf den Funktionsgraphen aus?

(11) Man zeige mit Hilfe der Definition der Sinusfunktion (7.44)

$$\sin(x+y) = \sin x \cos y + \cos x \sin y.$$

(12) Man betrachte 1^n. Dies kann mithilfe des Binomischen Lehrsatzes dann wie folgt geschrieben werden

$$1^n = ((1-t)+t)^n = \sum_{j=0}^{n} \binom{n}{j} (1-t)^{n-j} t^j = \sum_{j=0}^{n} B_{n,j}(t), \quad t \in [0,1]$$

mit

$$B_{n,j}(t) = \begin{cases} \binom{n}{j}(1-t)^{n-j}t^j & t \in [0,1] \\ 0 & \text{sonst} \end{cases} \tag{7.50}$$

Die $B_{n,j}(t)$ heissen Bernsteinpolynome[5] und bilden die Basisfunktionen für Bézierkurven[6] die im CAD verwendet werden.
Man zeige:
(i) $B_{n,j}(t) \geq 0$ (Positivität)
(ii) $B_{n,j}(t) = B_{n,n-j}(1-t)$ (Symmetrie)
(iii) $B_{n,j}(t) = (1-t)B_{n-1,j}(t) + t\,B_{n-1,j-1}(t)$ (Rekursion)

(13) Man berechne den Funktionswert des Polynoms $p(x,y)$ am Punkt $(x,y) = (77617, 33096)$

$$p(x,y) = \frac{1335y^6}{4} + x^2\left(11x^2y^2 - y^6 - 121y^4 - 2\right) + \frac{11y^8}{2} + \frac{x}{2y}.$$

[5] https://de.wikipedia.org/wiki/Bernsteinpolynom
[6] https://de.wikipedia.org/wiki/B%C3%A9zierkurve

7.7. Differentialrechnung in $\mathbb{R}^1$

Die Grundidee der Differentialrechnung ist die Approximation beliebiger Funktionen in einer Umgebung eines Punktes durch eine Gerade (affine bzw. "lineare" Funktion).

Definition 7.78. *Es sei $D \subset \mathbb{R}$ und $f : D \to \mathbb{R}$ eine Funktion. f heisst in einem Häufungspunkt $x_0 \in D$* differenzierbar*, falls der Grenzwert*

$$f'(x_0) := \lim_{\substack{\xi \to x_0 \\ \xi \in D \setminus \{x_0\}}} \frac{f(\xi) - f(x_0)}{\xi - x_0} \tag{7.51}$$

existiert.

Der Grenzwert $f'(x_0)$ heisst Differenzialquotient *oder* Ableitung *von f im Punkte x_0. Die Funktion f heisst differenzierbar in D, falls f in jedem Punkt $x \in D$ differenzierbar ist.*

Die Funktion $f' : D' \to \mathbb{R}, \quad x_0 \mapsto f'(x_0)$ wobei

$$D' = \{x \mid f'(x) \text{ existiert } \} \tag{7.52}$$

heisst Ableitung von f.

Ist f' stetig auf D', so heisst f stetig differenzierbar auf D'.

Schreibt man $\xi = x_0 + h$, kann man den Differenzialquotienten auch folgendermassen darstellen

$$f'(x_0) := \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \tag{7.53}$$

Statt $f'(x_0)$ wird auch oft die Notation $\dfrac{d\,f}{d\,x}(x_0)$ verwendet.

Geometrische Interpretation: der Differenzenquotient

$$\frac{f(\xi) - f(x_0)}{\xi - x_0}$$

ist die Steigung der Sekante des Graphen von f durch die Punkte $(x_0, f(x_0))$ und $(\xi, f(\xi))$ (Abb. 7.18). Beim Grenzübergang $\xi \to x_0$ geht die Sekante in die Tangente an den Graphen von f im Punkt $(x_0, f(x_0))$ über. $f'(x_0)$ stellt also im Falle der Existenz die Steigung der Tangente im Punkt $(x_0, f(x_0))$ dar (Abb. 7.19).

Betrachtet man den zurückgelegen Weg als Funktion der Zeit $s(t)$ so ergibt die Ableitung gerade die Momentangeschwindigkeit $v(t)$. Die Ableitung $s'(t)$ gibt also die Änderung einer Funktion $s(t)$ an. Man schreibt in diesem Fall statt $s'(t)$ oft $\dot{s}(t)$ für die Geschwindigkeit zum Zeitpunkt t.

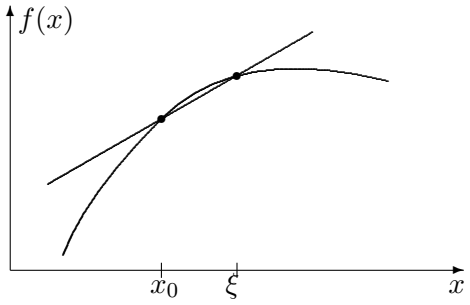


Abbildung 7.18. Sekante

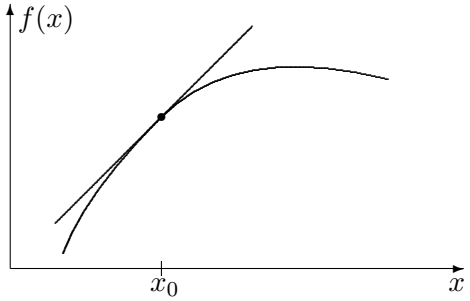


Abbildung 7.19. Tangente

Beispiele:
(i) Für die konstante Funktion $f : \mathbb{R} \to \mathbb{R}, f(x) = c$, gilt

$$f'(x_0) := \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \frac{f(\xi) - f(x_0)}{\xi - x_0} = \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \frac{c - c}{\xi - x_0} = 0.$$

(ii) Für $f : \mathbb{R} \to \mathbb{R}, f(x) = cx$, gilt

$$f'(x_0) := \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \frac{f(\xi) - f(x_0)}{\xi - x_0} = \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \frac{c\xi - cx_0}{\xi - x_0} = c.$$

(iii) Für $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$, gilt

$$f'(x_0) := \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \frac{f(\xi) - f(x_0)}{\xi - x_0} = \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \frac{\xi^2 - x_0^2}{\xi - x_0} = \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} (\xi + x_0) = 2\,x_0.$$

(iv) Für $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}, f(x) = \frac{1}{x}$, gilt

$$f'(x_0) := \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \frac{f(\xi) - f(x_0)}{\xi - x_0} = \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \frac{\frac{1}{\xi} - \frac{1}{x_0}}{\xi - x_0} = \lim_{\substack{\xi \to x_0 \\ \xi \neq x_0}} \left(\frac{-1}{\xi\, x_0}\right) = -\frac{1}{x_0^2}.$$

(v) Für $f : \mathbb{R} \to \mathbb{R}, f(x) = \mathrm{e}^x$, gilt

$$f'(x_0) := \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{h \to 0} \frac{\mathrm{e}^{x_0 + h} - \mathrm{e}^{x_0}}{h} = \mathrm{e}^{x_0} \lim_{h \to 0} \frac{\mathrm{e}^h - 1}{h} = \mathrm{e}^{x_0}.$$

Die Exponentialfunktion zeichnet sich also dadurch aus, dass ihre Steigung in jedem Punkt ihrem Funktionswert gleicht.

(vi) Die Funktion $|\ | : \mathbb{R} \to \mathbb{R}, f(x) = |x|$ ist in $x_0 = 0$ nicht differenzierbar, da mit $h_n = (-1)^n \frac{1}{n}$ folgt

$$f'(x_0) := \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = \lim_{n \to \infty} \frac{\frac{1}{n} - 0}{(-1)^n \frac{1}{n}} = \lim_{n \to \infty} (-1)^n.$$

Anmerkung: Man sagt eine Funktion ist *rechtsseitig differenzierbar*, wenn

$$f'_+(x_0) := \lim_{\xi \downarrow x_0} \frac{f(\xi) - f(x_0)}{\xi - x_0} \tag{7.54}$$

existiert und *linksseitig differenzierbar*, wenn

$$f'_-(x_0) := \lim_{\xi \uparrow x_0} \frac{f(\xi) - f(x_0)}{\xi - x_0} \tag{7.55}$$

existiert. Die Funktion $|\ | : \mathbb{R} \to \mathbb{R}, f(x) = |x|$ ist in $x_0 = 0$ rechtsseitig differenzierbar und linksseitig differenzierbar, aber rechtseitiger und linksseitiger Differentialquotient sind verschieden. Eine Funktion ist genau dann differenzierbar in x_0, wenn rechtsseitiger und linksseitiger Differentialquotient in x_0 übereinstimmen.

Satz 7.79. *Es sei $D \subset \mathbb{R}$ und a[7] ein Häufungspunkt. Eine Funktion $f : D \to \mathbb{R}$ ist genau dann im Punkt a differenzierbar, wenn es eine Konstante $c \in \mathbb{R}$ gibt, sodass für $x \in D$ gilt*

$$f(x) = f(a) + c\,(x - a) + (x - a)\, r(x). \tag{7.56}$$

Dabei ist r eine Funktion, für die gilt

$$\lim_{x \to a} r(x) = 0. \tag{7.57}$$

In diesem Fall ist $c = f'(a)$.

[7] Notationswechsel: a entspricht x_0.

Beweis: [**83**], Seite 145.

Dieser Satz zeigt, dass die Differenzierbarkeit in a gleichbedeutend ist mit der Approximierbarkeit der Funktion durch eine Gerade (lineare Funktion) im Punkt a.[8]

$$L(x) = f(a) + c\,(x - a) \tag{7.58}$$

Der Graph von L ist die Tangente an den Graphen von f im Punkt $(a, f(a))$. Mit Hilfe des Landausymbols o lässt sich dies auch wie folgt schreiben

$$f(x) = f(a) + c\,(x - a) + o(|x - a|) \qquad \text{für } x \to a.$$

Aus der Definition der Ableitung als Grenzwert des Differenzenquotienten folgt unmittelbar

Satz 7.80. *Ist eine Funktion $f : D \to \mathbb{R}$ in $a \in D$ differenzierbar, so ist sie in a auch stetig.*

Aus den Regeln zur Berechnung von Grenzwerten von Folgen und Funktionen folgen nun die Regeln für das Ableiten.

Satz 7.81. *Seien $f, g : D \to \mathbb{R}$ in $x \in D$ differenzierbare Funktionen und $\lambda \in \mathbb{R}$. Dann sind auch die Funktionen*

$$f + g, \quad \lambda f, \quad fg \; : \quad D \to \mathbb{R}$$

in x differenzierbar und es gelten die Rechenregeln:

(a) Linearität

$$\begin{aligned}(f + g)'(x) &= f'(x) + g'(x),\\ (\lambda f)'(x) &= \lambda f'(x).\end{aligned} \tag{7.59}$$

(b) (Produktregel).

$$(f\,g)'(x) = f'(x)\,g(x) + f(x)\,g'(x). \tag{7.60}$$

(c) (Quotientenregel). *($g(x) \neq 0$ für alle $x \in D$)*

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)\,g(x) - f(x)\,g'(x)}{g(x)^2}. \tag{7.61}$$

(d) (Kettenregel). *Seien $f : D \to \mathbb{R}$ und $g : E \to \mathbb{R}$ Funktionen mit $f(D) \subset E$. Die Funktion f sei differenzierbar im Punkt $x \in D$ und g sei in $y := f(x) \in E$ differenzierbar. Dann ist die zusammengesetzte Funktion*

$$g \circ f : D \to \mathbb{R} \tag{7.62}$$

[8] Funktionen von $\mathbb{R}^n \to \mathbb{R}^m$ werden durch entsprechende lineare Funktionen (Jakobi-Matrix) approximiert)

im Punkt x differenzierbar und es gilt

$$(g \circ f)'(x) = g'(f(x))f'(x). \tag{7.63}$$

Beweis: [**83**].

Satz 7.82. *(Ableitung der Umkehrfunktion). Sei $D \subset \mathbb{R}$ ein Intervall, $f : D \to \mathbb{R}$ eine stetige, streng monotone Funktion und $\varphi = f^{-1} : D^* \to \mathbb{R},\ D^* = f(D)$ die Umkehrfunktion.*

Ist f im Punkt $x \in D$ differenzierbar und $f'(x) \neq 0$, so ist φ im Punkt $y := f(x)$ differenzierbar und es gilt

$$\varphi'(y) = \frac{1}{f'(x)} = \frac{1}{f'(\varphi(y))}. \tag{7.64}$$

Beweis: folgt aus Ableiten von $\varphi(f(x)) = x$, [**83**], Seite 149.

Mit all diesen Hilfsmittel können wir nun die elementaren Funktionen ableiten.

Satz 7.83 (Ableitung elementarer Funktionen)**.**

Funktion	*Ableitung*	
x^n	$n\,x^{n-1}$	$n \in \mathbb{Z} \setminus \{0\}, \quad x \neq 0,\ wenn\ n < 0$
x^a	$a\,x^{a-1}$	$x > 0,\ a \in \mathbb{R}$
$\log(\lvert x\rvert)$	$\frac{1}{x}$	$x \neq 0$
$\log_a(x)$	$\frac{1}{x\log(a)}$	$a \neq 1, a > 0$
e^x	e^x	
a^x	$a^x \log(a)$	$a > 0$
$\sinh(x)$	$\cosh(x)$	
$\cosh(x)$	$\sinh(x)$	
$\tanh(x)$	$\frac{1}{\cosh(x)^2}$	
$\coth(x)$	$-\frac{1}{\sinh(x)^2}$	
$\mathrm{Arsinh}(x)$	$\frac{1}{\sqrt{x^2+1}}$	
$\mathrm{Arcosh}(x)$	$\frac{1}{\sqrt{x^2-1}}$	$x > 1$
$\mathrm{Artanh}(x)$	$-\frac{1}{x^2-1}$	$\lvert x\rvert < 1$
$\mathrm{Arcoth}(x)$	$-\frac{1}{x^2-1}$	$\lvert x\rvert > 1$
$\sin(x)$	$\cos(x)$	
$\cos(x)$	$-\sin(x)$	
$\tan(x)$	$\frac{1}{\cos(x)^2}$	
$\cot(x)$	$-\frac{1}{\sin(x)^2}$	
$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$	$\lvert x\rvert < 1$
$\arccos(x)$	$-\frac{1}{\sqrt{1-x^2}}$	$\lvert x\rvert > 1$
$\arctan(x)$	$\frac{1}{1+x^2}$	
$\mathrm{arccot}(x)$	$-\frac{1}{1+x^2}$	

Es gilt unter Verwendung von Satz 7.82

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

Daraus folgt, da $\log'(1) = 1$ ist

$$1 = \lim_{n\to\infty} \frac{\log(1+\frac{1}{n}) - \log(1)}{\frac{1}{n}} = \lim_{n\to\infty} \log\left((1+\frac{1}{n})^n\right)$$

und somit wegen der Stetigkeit der Exponentialfunktion

$$\lim_{n\to\infty} (1+\frac{1}{n})^n = \mathrm{e}.$$

Zur Ableitung von $x^a, a \in \mathbb{R}$ benutzt man die Kettenregel

$$\left(x^a\right)' = \exp(a\log x)' = \exp'(a\log x)\frac{d}{dx}(a\log x) = \exp(a\log x)\frac{a}{x} = a\,x^{a-1}.$$

7.7.1. Ableitungen höherer Ordnung. Die Funktion $f : D \to \mathbb{R}$ sei in D differenzierbar. Falls die Ableitung $f' : D \to \mathbb{R}$ ihrerseits im Punkt $x \in D$ differenzierbar ist, so heisst

$$\frac{d^2 f(x)}{dx^2} := f''(x) := (f')'(x) \tag{7.65}$$

die zweite Ableitung von f in x.

Rekursiv definiert man Ableitungen höher Ordnung

$$f^{(k)}(x) := \frac{d^k f(x)}{dx^k} := \left(\frac{d}{dx}\right)^k f(x) := \frac{d}{dx}\left(\frac{d^{k-1} f(x)}{dx^{k-1}}\right). \tag{7.66}$$

Die Funktion $f : D \to \mathbb{R}$ heisst *k-mal differenzierbar* in D, wenn f in jedem Punkt $x \in D$ k-mal differenzierbar ist. Sie heisst k-mal *stetig differenzierbar* in D, wenn die k-te Ableitung $f^{(k)} : D \to \mathbb{R}$ stetig ist. Man schreibt $f \in C^k(D), k \geq 1$ und $f \in C^0(D)$, wenn f nur stetig ist und $f \in C^\infty(D)$, wenn f beliebig oft ableitbar ist.

7.8. Lokale Extrema und Mittelwertsatz

Definition 7.84. *Sei $f : (a,b) \to \mathbb{R}$ eine Funktion. Man sagt f habe in $x \in (a,b)$ ein* lokales Maximum (Minimum)*, wenn ein $\epsilon > 0$ existiert, sodass*

$$f(x) \geq f(\xi) \qquad (\text{ bzw. } f(x) \leq f(\xi)) \qquad \text{für alle } \xi \text{ mit } |x-\xi| < \epsilon. \tag{7.67}$$

Gilt das Gleichheitszeichen nur für $x = \xi$, so nennt man x ein strenges *oder* striktes *lokales Maximum (Minimum).*

Der gemeinsame Oberbegriff für Maximum und Minimum ist *Extremum* und man nennt ein lokales Extremum auch noch *relatives Extremum.*

Satz 7.85. *Die Funktion $f : (a, b) \to \mathbb{R}$ besitze im Punkt $x \in (a, b)$ ein lokales Extremum und sei dort auch differenzierbar. Dann ist $f'(x) = 0$.*

Beweis: [**83**], Seite 154.

$f'(x) = 0$ ist nur eine notwendige Voraussetzung für das Vorliegen eines Extremums wie man an der Funktion $f(x) = x^3$ oder $f : [-1, 1] \to \mathbb{R},\ f(x) = |x|$ sieht.

Nach Satz 7.43 nimmt jede in einem abgeschlossenen Intervall stetige Funktion $f : [a, b] \to \mathbb{R}$ ihr absolutes Maximum und Minimum an. Liegt ein Extremum jedoch am Rand, so muss dort nicht notwendigerweise $f'(x) = 0$ sein, z. B. $f : [0, 1] \to \mathbb{R},\ f(x) = x^2$.

Relative Extrema lassen sich mit Hilfe der Ableitung folgendermassen finden.

Satz 7.86. *Sei $f : (a, b) \to \mathbb{R}$ eine differenzierbare Funktion. Im Punkte $x \in (a, b)$ sei f zwei mal differenzierbar und es gelte*

$$f'(x) = 0 \text{ und } f''(x) > 0 \quad (\text{ bzw. } f''(x) < 0). \tag{7.68}$$

Dann besitzt f in x ein strenges lokales Minimum (bzw. Maximum).

Ist f n-mal differenzierbar auf (a, b) und es gelte

$$f'(x) = f''(x) = \ldots = f^{(n-1)}(x) = 0 \text{ und } f^{(n)}(x) > 0 \quad (\text{ bzw. } f^{(n)}(x) < 0). \tag{7.69}$$

Dann besitzt f in x ein strenges lokales Minimum (bzw. Maximum), wenn n eine gerade Zahl ist. Ist n ungerade so existiert kein Extremum in x.

Mit Hilfe der Ableitung kann die Monotonie einer Funktion folgendermassen festgestellt werden.

Satz 7.87. *Die Funktion $f : [a, b] \to \mathbb{R}$ sei stetig und auf (a, b) differenzierbar.*

(i) Wenn für alle $x \in (a, b)$ gilt $f'(x) \geq 0$ (bzw. $f'(x) > 0, f'(x) \leq 0, f'(x) < 0$), so ist f in $[a, b]$ monoton wachsend (bzw. streng monoton wachsend, monoton fallend, streng monoton fallend).

(ii) Ist f monoton wachsend (bzw. fallend), so folgt $f'(x) \geq 0$ (bzw. $f'(x) \leq 0$) für alle $x \in (a, b)$.

Beweis: [**83**], Seite 157.

Aus der strengen Monotonie kann aber nicht $f'(x) > 0$ gefolgert werden, wie zum Beispiel die Abbildung $f(x) = x^3$ im Punkt $x = 0$ zeigt.

Satz 7.88. (Satz von Rolle) *Sei $a < b$ und die Funktion $f : [a, b] \to \mathbb{R}$ eine stetige Funktion mit $f(a) = f(b)$. Die Funktion f sei differenzierbar auf (a, b). Dann existiert ein $\xi \in (a, b)$ mit $f'(\xi) = 0$.*

Anmerkung: Zwischen zwei Nullstellen einer differenzierbaren Funktion muss daher immer eine Nullstelle der Ableitung liegen.

Beweis: Ist $f(x) = f(a)$ für alle $x \in [a, b]$, so gilt für jedes $x \in (a, b)$: $f'(\xi) = 0$. Ist dies nicht der Fall, gibt es ein $x \in (a, b)$ in dem f ein Extremum hat. Nach Satz 7.85 gilt für den Extremwert $f'(x) = 0$.

Aus diesem einfachen Satz folgt umittelbar der

Satz 7.89. (Mittelwertsatz). *Sei $a < b$ und die Funktion $f : [a, b] \to \mathbb{R}$ eine stetige Funktion, die auf (a, b) differenzierbar ist. Dann existiert ein $\xi \in (a, b)$ sodass*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}. \tag{7.70}$$

Beweis: Man subtrahiere von f die Verbindungsgerade von $f(a)$ nach $f(b)$ und wende darauf den Satz von Rolle an.

Geometrisch heisst dies, dass es mindestens einen Punkt $\xi \in (a, b)$ gibt, sodass die Tangente im Punkt $(\xi, f(\xi)$ die gleiche Steigung wie die Sekante durch die Punkte $(a, f(a)$ und $(b, f(b)$ hat.

Anwendungen des Mittelwertsatzes.

(i) Für alle $x,\ 0 < x < 1$, gilt: $1 + x < \mathrm{e}^x < \dfrac{1}{1 - x}$.

(ii) Für alle $x > 0$ gilt:

$$\frac{x}{1 + x} < \log(1 + x) < x. \tag{7.71}$$

(iii) Sei $f : [a, b] \to \mathbb{R}$ eine stetige Funktion, die auf (a, b) differenzierbar ist mit $f'(x) = 0$ für alle $x \in (a, b)$. Dann ist f konstant.

(iv) Sei $c \in \mathbb{R}$ eine Konstante und $f : \mathbb{R} \to \mathbb{R}$ eine differenzierbare Funktion mit

$$f'(x) = cf(x) \qquad \text{für alle } x \in \mathbb{R}. \tag{7.72}$$

Dann gilt

$$f(x) = A\mathrm{e}^{cx} \qquad \text{für alle } x \in \mathbb{R}, \tag{7.73}$$

wobei $A = f(0)$ ist.

Beweis: [**83**], Seite 156.

Ein weiterer wichtiger Begriff der mittels der Ableitung geprüft werden kann ist der Begriff der Konvexität.

Definition 7.90. *Es sei $D \subset \mathbb{R}$ ein (endliches oder unendliches) Intervall. Eine Funktion $f : D \to \mathbb{R}$ heisst* konvex, *wenn für alle $x_1, x_2 \in D$ und alle λ mit $0 < \lambda < 1$ gilt*

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2). \tag{7.74}$$

(Streng konvex, wenn man $\leq$ durch $<$ ersetzen kann). Die Funktion f heisst konkav, *wenn $-f$ konvex ist.*

Geometrisch bedeuted konvex, dass der Graph der Funktion f auf D für alle x_1, x_2 unterhalb der Verbindungsgeraden (Sekante) von $(x_1, f(x_1))$ und $(x_2, f(x_2))$ verläuft.

Satz 7.91. *Es sei $D \subset \mathbb{R}$ ein offenes Intervall und $F : D \to \mathbb{R}$ eine zweimal differenzierbare Funktion. f ist genau dann (streng) konvex, wenn $f''(x) \geq 0$ ($f''(x) > 0$) für alle $x \in D$ gilt.*

Beweis: [**83**], Seite 159.

Mit Hilfe des Begriffs Konvexität kann man nun folgende Sätze beweisen. Zuerst eine Definition. Die **p-Norm** $||\ ||_p$ eines Vektors $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$ ist definiert durch

$$||x||_p := \left(\sum_{k=1}^{n} |x_k|^p \right)^{\frac{1}{p}}. \tag{7.75}$$

Satz 7.92. (Höldersche Ungleichung). *Seien $p, q \in (1, \infty)$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt für alle Vektoren $x, y \in \mathbb{C}^n$*

$$\sum_{k=1}^{n} |x_k y_k| \leq ||x||_p \, ||y||_q. \tag{7.76}$$

Beweis: [**83**], Seite 160.

Ein Spezialfall davon ist die Cauchy-Schwarzsche Ungleichung $p = q = 2$.

Satz 7.93. (Minkowskische Ungleichung). *Sei $p \in [1, \infty)$. Dann gilt für alle Vektoren $x, y \in \mathbb{C}^n$*

$$||x + y||_p \leq ||x||_p + ||y||_p. \tag{7.77}$$

Beweis: [**83**], Seite 160.

Ungleichung zwischem *geometrischen* und *arithmetischen Mittel.*

Satz 7.94. *Seien $x_j > 0, 1 \le j \le n$. Dann gilt*

$$\sqrt[n]{x_1 \cdots x_n} \le \frac{1}{n} \sum_{j=1}^{n} x_j. \tag{7.78}$$

Satz 7.95. (Allgemeine Bernoullische Ungleichung). *Für alle $x > -1$ und $x \neq 0$ gilt*

$$\begin{aligned} (1+x)^\alpha &> 1 + \alpha\, x \qquad \alpha < 0 \text{ oder } \alpha > 1, \\ (1+x)^\alpha &< 1 + \alpha\, x \qquad 0 < \alpha < 1. \end{aligned} \tag{7.79}$$

7.8.1. Regeln von de l'Hospital. Mit Hilfe des Mittelwertsatzes lassen sich auch bequeme Regeln zur Berechnung von Grenzwerten von Funktionen herleiten.

Satz 7.96. (Regeln von de l'Hospital[9]). *Auf dem Intervall $I = (a, b)$, $-\infty \le a < b \le \infty$ seien $f, g : I \to \mathbb{R}$ zwei differenzierbare Funktionen. Es gelte $g'(x) \neq 0$ für alle $x \in I$ und es existiere der Grenzwert*

$$\lim_{x \uparrow b} \frac{f'(x)}{g'(x)} =: c \in \mathbb{R}. \tag{7.80}$$

dann gilt:

(i) Falls $\lim_{x\uparrow b} g(x) = \lim_{x\uparrow b} f(x) = 0$ ist, ist $g(x) \neq 0$ für alle $x \in I$ und

$$\lim_{x \uparrow b} \frac{f(x)}{g(x)} = c \tag{7.81}$$

(ii) Falls $\lim_{x\uparrow b} g(x) = \pm\infty$ ist, ist $g(x) \neq 0$ für $x \ge x_0, a < x_0 < b$ und

$$\lim_{x \uparrow b} \frac{f(x)}{g(x)} = c \tag{7.82}$$

Analoge Aussagen gelten für den Grenzwert $x \downarrow a$

Beweis: [**83**], Seite 163.

Damit kann man nun leicht Grenzwerte der Form

$$\frac{0}{0}, \quad \frac{\infty}{\infty}, \quad 0 \cdot \infty = \frac{\infty}{\frac{1}{0}}, \quad \infty - \infty, \quad 1^\infty, \quad \textit{etc.}$$

berechnen. Oft muss man dabei diese Ausdrücke auf eine Form bringen, sodass obiger Satz anwendbar ist, oder diesen Satz mehrmals anwenden. Beispiele:

$$\text{(i)} \ \lim_{x \to 0} \frac{\log(1+x)}{x} = \lim_{x \to 0} \frac{\frac{1}{(1+x)}}{1} = 1.$$

[9] http://www-history.mcs.st-and.ac.uk/Biographies/De_L'Hopital.html

Anmerkung: Zähler und Nenner werden getrennt differenziert!

$$\text{(ii)}\ \lim_{x\downarrow 0} x\log(x) = \lim_{x\downarrow 0} \frac{\log(x)}{\frac{1}{x}} = \lim_{x\downarrow 0} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = -\lim_{x\downarrow 0} x = 0.$$

$$\text{(iii)}\ \lim_{x\to 0} \frac{\mathrm{e}^x - x - 1}{x^2} = \lim_{x\to 0} \frac{\mathrm{e}^x - 1}{2x} = \lim_{x\to 0} \frac{\mathrm{e}^x}{2} = \frac{1}{2}.$$

$$\text{(iv)}\ \lim_{x\downarrow 0} x^x = \lim_{x\downarrow 0} \mathrm{e}^{x\log x} = \mathrm{e}^{\lim_{x\downarrow 0} x\log x} = \mathrm{e}^0 = 1.$$

7.9. Übung

(1) Man berechne die folgenden Ableitungen direkt als Grenzwert des Differentialquotienten

$$f : \mathbb{R} \setminus \{1\} \to \mathbb{R}, \quad f(x) = \frac{1+x}{1-x} \text{ am Punkt } x = 2,$$

$$g : D \to \mathbb{R}, D = \{x \in \mathbb{R} \mid x > \frac{1}{3}\}, \quad g(x) = \sqrt{3x-1} \text{ am Punkt } x = 4.$$

(2) Man berechne die folgenden Ableitungen

$$(6x^2+7x+4)^4, \quad 3x^4 \log^2 x, \quad \frac{\mathrm{e}^{x+1}-1}{\mathrm{e}^{x+1}+1}, \quad \log\left((4\mathrm{e}^{2x}+3)^{\frac{1}{2}} + 2\mathrm{e}^x\right),$$

$$x^x, \quad x^{x+\log x}.$$

(3) Man linearisiere (approximiere durch eine Gerade) die gegebenen Funktionen in der Umgebung von x_0 und vergleiche die Näherungs- und exakte Funktionswerte an den Stellen $x_1 = x_0 + 0.1$ und $x_2 = x_0 - 0.1$.

a) $f(x) = \sqrt{1+x^4}, \quad x_0 = 1,$

b) $f(x) = \sqrt{x} \cdot \log x, \quad x_0 = 2.$

(4) Gegeben sei die Funktion $f(x) = x^3 - 3x + 4$. Für $a = 1, b = 2$ bestimme man ein $\xi \in (a, b)$, sodass

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$

(5) (*) Mit Hilfe des Mittelwertsatzes beweise man
- (i) Für alle $x > 0$ gilt: $\sqrt{1+x} < 1 + \frac{1}{2}x$.
- (ii) Für $\alpha < \beta$ gilt: $\mathrm{e}^\alpha(\beta - \alpha) < \mathrm{e}^\beta - \mathrm{e}^\alpha < \mathrm{e}^\beta(\beta - \alpha)$.
- (iii) Für alle $x > 0$ gilt: $\sin x < x$.

(6) Man beweise, dass die Funktion $f(x) = \dfrac{x-1}{x \log x}$ für $x > 1$ monoton fällt (verwende Ungleichung 7.71).

(7) Man bestimme Nullstellen und Extrema der Funktion $f(x) = x^4 - 3x^2$.

(8) Man bestimme alle Extrema der Funktion

$$f(x) = \begin{cases} \frac{x^2-4}{x+1} & x \neq -1 \\ 0 & x = -1 \end{cases}$$

(9) Man bestimme alle Extrema der Funktion $f : (0, \infty) \to \mathbb{R}$, $f(x) = x^x$.

(10) Man bestimme die Nullstellen und Extremwerte der Funktion

$$f(x) = \sin(2^{x-1}\pi)$$

auf dem Intervall $[1, 4]$.

(11) Man bestimme folgende Grenzwerte

$$\lim_{x\to 0} \frac{a^x - b^x}{x}, \quad \lim_{x\to 2} \frac{3x^2 - 6x}{x^2 - 3x + 2}, \quad \lim_{x\to 0} \sqrt{\frac{x^2 + x}{\mathrm{e}^x - 1}}, \quad \lim_{x\to\infty} \frac{\log(1 + \mathrm{e}^x)}{x}.$$

(12) Man bestimme folgende Grenzwerte $\alpha > 0$

$$\lim_{x\to\infty} \frac{\log x}{x^\alpha}, \quad \lim_{x\to\infty} \frac{x^\alpha}{\mathrm{e}^x}, \quad \lim_{x\to 0} \frac{\log(1 + x) - \sin x}{x^2}, \quad \lim_{x\to 0} (\cos(2x))^{\frac{1}{x^2}}.$$

(13) Man bestimme folgende Grenzwerte

$$\lim_{x\to 0} (\cosh x)^{\frac{1}{x^2}}, \qquad \lim_{x\to 1} \frac{x^x - x}{1 - x + \log x}, \qquad \lim_{x\to 0} \left(\frac{1}{\sinh^2 x} - \frac{1}{x^2} \right).$$

7.10. Integralrechnung in $\mathbb{R}^1$

Es sei $a < b, I = [a, b]$. Dann bezeichnen wir mit $T[a, b]$ die Menge aller Treppenfunktionen $\varphi : I \to \mathbb{R}$ auf dem Intervall I. Die Menge aller Treppenfunktionen bildet einen Vektorraum.

Man sieht unmittelbar, dass die Summe zweier Treppenfunktionen wieder eine Treppenfunktionen ist. Sei Z' die Zerlegung bezüglich der Treppenfunktionen φ und Z'' die Zerlegung bezüglich der Treppenfunktionen ψ, dann wählt man für die Summe $\varphi + \psi$ einfach die Zerlegung Z, welche alle Teilpunkte von Z' und Z'' enthält.

Definition 7.97. (Integral für Treppenfunktionen). *Sei φ eine Treppenfunktion bezüglich der Zerlegung*

$$a = x_0 < x_1 < \cdots < x_n = b.$$

und $\varphi\big|_{(x_{k-1},x_k)} = c_k,\ 1 \le k \le n$. Dann definiert man das Intergal der Treppenfunktion φ als

$$\int_a^b \varphi(x)\,dx := \sum_{k=1}^{n} c_k(x_k - x_{k-1}). \tag{7.83}$$

Geometrisch ist das Integral die Summe der Flächen der Rechtecke, die von der Treppenfunktion φ und der x-Achse aufgespannt werden. Ist der Wert von c_k negativ, so ist die entsprechende Fläche negativ.

Für beliebige Funktionen kann mit Hilfe von Treppenfunktionen das *Ober-* und *Unterintegral* definiert werden.

Definition 7.98. *Es sei $f : [a, b] \to \mathbb{R}$ eine beliebige beschränkte Funktion. Dann setzt man*

$$\begin{aligned}\int_a^{b*} f(x)\,dx &:= \inf\{\int_a^b \varphi(x)\,dx \ \big|\varphi \in T[a, b], \varphi \ge f\},\\ \int_{a*}^b f(x)\,dx &:= \sup\{\int_a^b \varphi(x)\,dx \ \big|\varphi \in T[a, b], \varphi \le f\}.\end{aligned} \tag{7.84}$$

Für jede Treppenfunktion gilt

$$\int_a^{b*} \varphi(x)\,dx = \int_{a*}^b \varphi(x)\,dx = \int_a^b \varphi(x)\,dx.$$

Für die Dirichletfunktion auf $[0, 1]$

$$f(x) = \begin{cases} 1 & \text{für } x \text{ rational} \\ 0 & \text{für } x \text{ irrational} \end{cases}$$

gilt

$$\int_0^{1\,*} f(x)\,dx = 1 \text{ und } \int_{0\,*}^{1} f(x)\,dx = 0.$$

Definition 7.99. *Eine beschränkte Funktion $f : [a,b] \to \mathbb{R}$ heisst* Riemann-integrierbar*, wenn*

$$\int_a^{b\,*} f(x)\,dx = \int_{a\,*}^{b} f(x)\,dx. \tag{7.85}$$

Man setzt in diesem Fall

$$\int_a^b f(x)\,dx = \int_a^{b\,*} f(x)\,dx. \tag{7.86}$$

Jede Treppenfunktion ist daher Riemann-integrierbar. Die Dirichletfunktion ist aber nicht Riemann-integrierbar.

Anmerkung: Damit man die Dirichletfunktion „integrieren“ kann, muss man den Integralbegriff verallgemeinern. Dies führt zum Lebesgue-Integral. Die Grundidee ist dabei die Zerlegung auf der y-Achse zu machen und dann die „Länge“ der Urbildmengen dieser Zerlegung zu bestimmen. Die Urbilder von Intervallen sind im allgemeinen komplizierte Mengen. Haben sie eine σ-Algebra-Struktur, dann sind sie Lebesgue-messbar.

Im Folgenden schreiben wir statt Riemann-integrierbar kurz **integrierbar**.

Folgende Funktionenklassen sind Riemann-integrierbar.

Satz 7.100. *(i) Jede stetige Funktion $f : [a,b] \to \mathbb{R}$ ist integrierbar.*

(ii) Jede monotone Funktion $f : [a,b] \to \mathbb{R}$ ist integrierbar.

Beweis: [**83**], Seite 181.

Das Riemann-Integral besitzt die Eigenschaften der Linearität und Monotonie.

Satz 7.101. *Seien $f, g : [a,b] \to \mathbb{R}$ integrierbare Funktionen und $\lambda \in \mathbb{R}$. Dann gilt*

$$\begin{aligned}
&(i)\ \int_a^b (f+g)(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx \\
&(ii)\ \int_a^b (\lambda f)(x)\,dx = \lambda \int_a^b f(x)\,dx \\
&(iii)\ \text{Ist } f \le g, \text{ dann folgt } \int_a^b f(x)\,dx \le \int_a^b g(x)\,dx
\end{aligned} \tag{7.87}$$

Beweis: [**83**], Seite 182.

Jede beliebige Funktion lässt sich in einem positiven und eine negativen Anteil zerlegen.

Definition 7.102. *Für eine Funktion $f : D \to \mathbb{R}$ definieren wir $f_+, f_- : D \to \mathbb{R}$ wie folgt:*

$$f_+(x) = \begin{cases} f(x) & \text{falls } f(x) > 0 \\ 0 & \text{sonst.} \end{cases}$$
$$f_-(x) = \begin{cases} -f(x) & \text{falls } f(x) < 0 \\ 0 & \text{sonst.} \end{cases} \tag{7.88}$$

Offensichtlich gilt $f = f_+ - f_-$ und $|f| = f_+ + f_-$.

Satz 7.103. *Seien $f, g : [a, b] \to \mathbb{R}$ integrierbare Funktionen. Dann gilt:*

(i) Die Funktionen f_+, f_- und $|f|$ sind integrierbar und es gilt:

$$|\int_a^b f(x)\,dx| \leq \int_a^b |f(x)|\,dx \tag{7.89}$$

(ii) Für jedes $p \in (1, \infty)$ ist die Funktion $|f|^p$ integrierbar.

(iii) Die Funktion $fg : [a, b] \to \mathbb{R}$ ist integrierbar.

Beweis: [**83**], Seite 183.

Satz 7.104. (Mittelwertsatz der Integralrechnung) *Seien $f, g : [a, b] \to \mathbb{R}$ stetige Funktionen und $g \geq 0$. Dann existiert ein $\xi \in [a, b]$, sodass*

$$\int_a^b f(x)g(x)\,dx = f(\xi) \int_a^b g(x)\,dx. \tag{7.90}$$

Im Spezialfall $g(x) = 1$ erhält man

$$\int_a^b f(x)\,dx = f(\xi)(b - a). \tag{7.91}$$

Beweis: [**83**], Seite 184.

Geometrisch heisst dies, dass ein flächengleiches Rechteck existiert mit Höhe $f(\xi)$.

7.10.1. Riemannsche Summen. Bis jetzt haben wir das Riemannintegral mittels Treppenfunktionen definiert, und zitiert, dass stetige und monotone Funktionen integrierbar sind. Wir wollen nun spezielle Treppenfunktionen einführen mit denen man das Integral konkret berechnen kann.

Sei $f : [a, b] \to \mathbb{R}$ eine Funktion,

$$a = x_0 < x_1 < \cdots < x_n = b$$

eine Zerlegung und ξ_k ein beliebiger Punkt (Stützstelle) aus dem Intervall $[x_{k-1}, x_k]$. Das Symbol

$$\mathcal{Z} := \big((x_k)_{0\le k\le n}, (\xi_k)_{1\le k\le n}\big) \tag{7.92}$$

bezeichne die Zusammenfassung der Teilpunkte x_k und der Stützstellen ξ_k.

Definition 7.105. *Die Summe*

$$S(\mathcal{Z}, f) = \sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}) \tag{7.93}$$

heisst Riemannsche Summe *der Funktion f bezüglich $\mathcal{Z}$. Die Feinheit von $\mathcal{Z}$ ist definiert als*

$$\mu(\mathcal{Z}) := \max_{1\le k\le n} (x_k - x_{k-1}). \tag{7.94}$$

Für integrierbare Funktionen konvergieren Riemannsche Summen gegen das Integral, wenn die Feinheit der Zerlegung gegen Null konvergiert.

Satz 7.106. *Sei $f : [a, b] \to \mathbb{R}$ eine Riemann-integrierbare Funktion. Dann gilt:*

$$\lim_{\mu(\mathcal{Z})\to 0} S(\mathcal{Z}, f) = \int_a^b f(x)\,dx. \tag{7.95}$$

Beweis: [**83**], Seite 186.

Beispiele:

(i) Man berechne mit Hilfe einer Riemannschen Summe das Integral $\int_0^a x^2\,dx$.

Wir wählen

$$x_k = \frac{ka}{n}, \qquad 0 \le k \le n,$$
$$\xi_k = x_k,$$
$$S_n = \sum_{k=1}^{n} \left(\frac{ka}{n}\right)^2 \left(\frac{ka}{n} - \frac{(k-1)a}{n}\right) = \frac{a^3}{n^3} \sum_{k=1}^{n} k^2 = \frac{a^3 n(n+1)(2n+1)}{6n^3},$$
$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \frac{a^3 n(n+1)(2n+1)}{6n^3} = \frac{a^3}{3}.$$

(ii) Man berechne mit Hilfe einer Riemannschen Summe das Integral $\int_1^a \frac{1}{x}\,dx$. Wir wählen

$$x_k = a^{\frac{k}{n}}, \qquad 0 \le k \le n,$$
$$\xi_k = x_{k-1},$$
$$S_n = \sum_{k=1}^{n} \frac{1}{a^{\frac{(k-1)}{n}}} \left(a^{\frac{k}{n}} - a^{\frac{(k-1)}{n}}\right) = \left(a^{\frac{1}{n}} - 1\right) \sum_{k=1}^{n} 1 = \left(a^{\frac{1}{n}} - 1\right) n,$$

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \left(a^{\frac{1}{n}} - 1\right) n = \log(a).$$

Sei $f : [a, b] \to \mathbb{R}$ eine integrierbare Funktionen. Dann definiert man die **p-Norm**

$$||f||_p := \left(\int_a^b |f(x)|^p \, dx\right)^{\frac{1}{p}}. \tag{7.96}$$

Es gelten nun für die Integrale Verallgemeinerungen der Minkowski und Hölderschen Ungleichung. Für integrierbare Funktionen $f, g : [a, b] \to \mathbb{R}$ gilt

$$||f + g||_p \leq ||f||_p + ||g||_p \quad \text{für alle } p \geq 1$$

$$\int_a^b |f(x)g(x)| \, dx \leq ||f||_p ||g||_q \quad \text{für alle } p, q \geq 1 \text{ mit } \frac{1}{p} + \frac{1}{q} = 1. \tag{7.97}$$

Das Riemann-Integral ist additativ.

Satz 7.107. *Sei $a < b < c$ und $f : [a, c] \to \mathbb{R}$ eine Funktion. f ist genau dann integrierbar, wenn $f|$ sowohl auf $[a, b]$ als auch $[b, c]$ integrierbar ist und es gilt*

$$\int_a^c f(x) \, dx = \int_a^b f(x) \, dx + \int_b^c f(x) \, dx. \tag{7.98}$$

Setzt man

Definition 7.108.

$$\int_a^a f(x) \, dx := 0,$$

$$\int_a^b f(x) \, dx := -\int_b^a f(x) \, dx \tag{7.99}$$

so gilt Satz 7.107 für beliebige a, b, c.

Viele bestimmte Integrale lassen sich aber nur mehr näherungsweise lösen: siehe Kapitel 9.3.

7.10.2. Unbestimmtes Integral. Wir wollen nun zeigen, dass die Integration in einem gewissen Sinne[10] die "Umkehrung" der Differentiation ist. Dies ermöglicht für viele Fälle die einfache Berechnung von Integralen.

Satz 7.109. *Es sei $f : I \to \mathbb{R}$ eine stetige Funktion und $a \in I$. Für $x \in I$ definiert man*

$$F(x) := \int_a^x f(t) \, dt. \tag{7.100}$$

Dann ist die Funktion $F(x) : I \to \mathbb{R}$ differenzierbar und es gilt $F'(x) = f(x)$.

[10] Anmerkung: die lineare Abbildung $f \mapsto f'$ ist nicht injektiv.

Beweis: Für $h \neq 0$ ist

$$\frac{F(x+h)-F(x)}{h} = \frac{1}{h}\left(\int_a^{x+h} f(t)\,dt - \int_a^x f(t)\,dt\right) = \frac{1}{h}\int_x^{x+h} f(t)\,dt.$$

Nach dem Mittelwertsatz der Integralrechnung (Satz 7.104) existiert ein $\xi_h \in [x, x+h]$ mit

$$\int_x^{x+h} f(t)\,dt = hf(\xi_h).$$

Damit ergibt sich für den Grenzwert

$$F'(x) = \lim_{h\to 0}\frac{F(x+h)-F(x)}{h} = \lim_{h\to 0}\frac{1}{h}\int_x^{x+h} f(t)\,dt = \lim_{h\to 0} f(\xi_h) = f(x)$$

da f stetig ist und $\lim_{h\to 0}\xi_h = x$ gilt.

Definition 7.110. *Eine differenzierbare Funktion $F(x): I \to \mathbb{R}$ heisst* Stammfunktion *(oder* primitive Funktion*) einer Funktion $f(x): I \to \mathbb{R}$, falls $F' = f$ gilt.*

Der Satz 7.109 besagt also, dass das unbestimmte Integral eine Stammfunktion ist. Zwei beliebige Stammfunktionen unterscheiden sich nur durch eine Konstante.

Satz 7.111. *Sei $F(x): I \to \mathbb{R}$ eine Stammfunktion von $f: I \to \mathbb{R}$. Eine weitere Funktion $G(x): I \to \mathbb{R}$ ist genau dann eine Stammfunktion von f, wenn $F - G$ eine Konstante ist.*

Mit Hilfe von Stammfunktionen können nun bestimmte Integrale berechnet werden.

Satz 7.112. (Hauptsatz der Differential- und Integralrechnung). *Sei $f(x): I \to \mathbb{R}$ eine stetige Funktion und F eine Stammfunktion von f. Dann gilt für alle $a, b \in I$*

$$\int_a^b f(t)\,dt = F(b) - F(a). \tag{7.101}$$

Man führt folgende Bezeichnungsweise ein

$$F(x)\Big|_a^b := F(b) - F(a) \tag{7.102}$$

und

$$\int f(t)\,dt := F(x) + c, \tag{7.103}$$

wobei die Konstante c sehr oft nicht angeschrieben wird.

Wir können nun die Tabelle 7.83 verkehrt herum lesen und erhalten folgende Stammfunktionen

Satz 7.113 (Integrale elementarer Funktionen)**.**

Funktion	*Stammfunktion*	
x^a	$\frac{1}{a+1}x^{a+1}$	$a \neq -1$
$\frac{1}{x}$	$\log(\lvert x\rvert)$	$x \neq 0$
e^x	e^x	
a^x	$\frac{1}{\log(a)}a^x$	$a > 0$
$\sinh(x)$	$\cosh(x)$	
$\cosh(x)$	$\sinh(x)$	
$\frac{1}{\sqrt{x^2+1}}$	$\mathrm{Arsinh}(x)$	
$\frac{1}{\sqrt{x^2-1}}$	$\mathrm{Arcosh}(x)$	$x > 1$
$\frac{-1}{x^2-1}$	$\mathrm{Artanh}(x)$	$\lvert x\rvert < 1$
$\frac{-1}{x^2-1}$	$\mathrm{Arcoth}(x)$	$\lvert x\rvert > 1$
$\sin(x)$	$-\cos(x)$	
$\cos(x)$	$\sin(x)$	
$\frac{1}{1+x^2}$	$\arctan(x)$	
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x)$	$\lvert x\rvert < 1$
$\frac{-1}{\sqrt{1-x^2}}$	$\arccos(x)$	$\lvert x\rvert > 1$
$\frac{-1}{1+x^2}$	$\mathrm{arccot}(x)$	

Das Analogon der Kettenregel für die Integration ist die Substitution. Durch geschickte Transformation der Integrationsvariablen kann ein Integral oft auf eine einfache Form gebracht werden.

Satz 7.114. (Substitution). *Sei $f : I \to \mathbb{R}$ eine stetige Funktion und $\varphi : [a, b] \to \mathbb{R}$ eine stetig differenzierbare Funktion mit $\varphi([a, b]) \subset I$. Dann gilt*

$$\int_a^b f(\varphi(t))\varphi'(t)\,dt = \int_{\varphi(a)}^{\varphi(b)} f(x)\,dx. \tag{7.104}$$

Beweis: [**83**], Seite 193.

Unter Verwendung der symbolischen Schreibweise[11]

$$d\varphi(t) := \varphi'(t)\,dt$$

lautet die Substitutionsregel

$$\int_a^b f(\varphi(t))\,d\varphi(t) = \int_{\varphi(a)}^{\varphi(b)} f(x)\,dx. \tag{7.105}$$

[11] In der "Praxis" behandelt man den Differentialquotienten $\frac{dx}{dt}$ formal wie einen echten Bruch.

Beispiele: Mit der Substitution $\varphi(t) = t + c$ erhält man

$$\int_a^b f(t+c)\,dt = \int_{a+c}^{b+c} f(x)\,dx.$$

Substitution $\varphi(t) = ct, c \neq 0$

$$\int_a^b f(ct)\,dt = \frac{1}{c}\int_{ac}^{bc} f(x)\,dx.$$

Substitution $\varphi(t) = t^2$

$$\int_a^b tf(t^2)\,dt = \frac{1}{2}\int_{a^2}^{b^2} f(x)\,dx.$$

Sei $\varphi(t) \neq 0$ für $t \in [a,b]$, dann folgt mit $f(x) = \frac{1}{x}, x = \varphi(t)$

$$\int_a^b \frac{\varphi'(t)}{\varphi(t)}\,dt = \log|\varphi(t)|\Big|_a^b.$$

Für die Integration beliebiger rationaler Funktionen wird die Partialbruchzerlegung eingesetzt.

Weitere Beispiele: Sei R eine rationale Funktion. Dann können folgende Integrale auf Integrale rationaler Funktionen transformiert werden.

$$\int R(t, \sqrt{at+b})\,dt = \int R(\frac{x^2-b}{a}, x)\frac{2x}{a}\,dx, \qquad \text{mit } x = \sqrt{at+b},$$

$$\int R(\mathrm{e}^t)\,dt = \int R(x)\frac{1}{x}\,dx, \qquad \text{mit } x = \mathrm{e}^t,$$

$$\int R(\sin t, \cos t, \tan t, \cot t)\,dt = \int R(\frac{2x}{1+x^2}, \frac{1-x^2}{1+x^2}, \frac{2x}{1-x^2}, \frac{1-x^2}{2x})\frac{2}{1+x^2}\,dx,$$
$$\text{mit } x = \tan\frac{t}{2}.$$

Für weitere Beispiele sei auf [**98**] oder Mathematica verwiesen.

Nicht alle Integrale sind geschlossen darstellbar und definieren eigene Funktionen, z. B. den Integralsinus

$$\mathrm{Si}(x) := \int_0^x \frac{\sin t}{t}\,dt \tag{7.106}$$

oder die Gaußsche Fehlerfunktion [12]ist durch folgendes Integral definiert

$$\mathrm{erf}(x) := \frac{2}{\pi}\int_0^x \mathrm{e}^{-t^2}\,dt. \tag{7.107}$$

Sie ist eine Stammfunktion der geraden Funktion $\frac{2}{\pi}\,\mathrm{e}^{-t^2}$.

[12] `https://de.wikipedia.org/wiki/Fehlerfunktion`

Die „Integration der Produktregel“ liefert die Methode der partiellen Integration.

Satz 7.115. (Partielle Integration). *Seien $f, g : [a, b] \to \mathbb{R}$ zwei stetig differenzierbare Funktionen Dann gilt*

$$\int_a^b f(x)g'(x)\,dx = f(x)g(x)\Big|_a^b - \int_a^b f'(x)g(x)\,dx. \tag{7.108}$$

Beweis: [**83**], Seite 197.

Beispiele:

(i) mit $f(x) = x$ und $g'(x) = \sin x$ ergibt $\int_a^b x\,\sin x\,dx = (-x\cos x + \sin x)\Big|_a^b$.

(ii) Mit $f(x) = \log x$ und $g'(x) = 1$ folgt

$$\int_a^b \log x\,dx = \int_a^b \underbrace{1}_{g'(x)}\underbrace{\log x}_{f(x)}\,dx = \underbrace{x}_{g(x)}\underbrace{\log x}_{f(x)}\Big|_a^b - \int_a^b \underbrace{x}_{g(x)}\underbrace{\frac{1}{x}}_{f'(x)}\,dx = (x\log x - x)\Big|_a^b.$$

(iii) Es kann auch passieren, dass das gesuchte Integral auf beiden Seiten auftaucht, oder dass man mehrmals partiell integrieren muss (Rekursionsformeln).

Mit $f(x) = \sin x$ und $g(x) = \sin x$ folgt

$$\begin{aligned} I_2 := \int_a^b \sin^2 x\,dx &= \int_a^b \underbrace{\sin x}_{g'(x)}\underbrace{\sin x}_{f(x)}\,dx = \underbrace{-\cos x}_{g(x)}\underbrace{\sin x}_{f(x)}\Big|_a^b + \int_a^b \underbrace{\cos x}_{-g(x)}\underbrace{\cos x}_{f'(x)}\,dx \\ &= (-\cos x\sin x + x)\Big|_a^b - \int_a^b \sin^2 x\,dx \end{aligned}$$

und daher

$$2I_2 = 2\int_a^b \sin^2 x\,dx = (-\cos x\sin x + x)\Big|_a^b.$$

Berechnung der Kreisfläche mit Radius eins. Mit $t = \cos x$ erhält man für den Viertelkreis

$$\int_0^1 \sqrt{1-t^2}\,dt = -\int_{\pi/2}^0 \sin^2 x\,dx = \frac{1}{2}(x - \cos x\sin x)\Big|_0^{\pi/2} = \frac{\pi}{4}.$$

7.11. Uneigentliche Integrale

Der bisherige Integralbegriff behandelte nur Integrale bei denen die Integrationsgrenzen a, b endlich und auch die Funktion $f(x)$ auf $[a, b]$ endlich war.

Um auch die Fälle, dass eine Integrationsgrenze unendlich ist einzuschliessen definieren wir

Definition 7.116. *Sei $f : [a, \infty) \to \mathbb{R}$ eine Funktion, die über jedem Intervall $[a, R], a < R < \infty$, integrierbar ist. Falls der Grenzwert*

$$\lim_{R\to\infty} \int_a^R f(x)\,dx \tag{7.109}$$

existiert, heisst das Integral $\int_a^\infty f(x)\,dx$ konvergent und man setzt

$$\int_a^\infty f(x)\,dx := \lim_{R\to\infty} \int_a^R f(x)\,dx \tag{7.110}$$

Analog definiert man $\int_{-\infty}^a f(x)\,dx$. Sind beide Integrationsgrenzen unendlich spaltet man das Integral in zwei Integrale bei denen jeweils nur eine Integrationsgrenze unendlich ist.

Beispiel: Das Integral

$$\int_1^\infty \frac{1}{x^s}\,dx$$

konvergiert für $s > 1$, denn es gilt

$$\int_1^R \frac{1}{x^s}\,dx = \frac{1}{1-s}x^{1-s}\Big|_1^R = \frac{1}{s-1}(1 - \frac{1}{R^{s-1}}) \to \frac{1}{s-1} \text{ für } s > 1, R \to \infty.$$

Für $s = 1$ ist $\int_1^R \frac{1}{x}\,dx = \log R$, was für $R \to \infty$ gegen ∞ strebt.

Ist der Integrand an einer Integrationsgrenze nicht definiert verwendet man folgende Definition.

Definition 7.117. *Sei $f : (a, b) \to \mathbb{R}$ eine Funktion, die über jedem Intervall $[a + \epsilon, b], 0 < \epsilon < b - a$, integrierbar ist. Falls der Grenzwert*

$$\lim_{\epsilon\downarrow 0} \int_{a+\epsilon}^b f(x)\,dx \tag{7.111}$$

existiert, heisst das Integral $\int_a^b f(x)\,dx$ konvergent und man setzt

$$\int_a^b f(x)\,dx := \lim_{\epsilon\downarrow 0} \int_{a+\epsilon}^b f(x)\,dx \tag{7.112}$$

Beispiele: (i) Das Integral

$$\int_0^\infty \frac{1}{x^s}\,dx$$

divergiert für jedes $s \in \mathbb{R}$.

(ii) Das Integral $\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,dx$ konvergiert, da

$$\begin{aligned}\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,dx &= \lim_{\epsilon\downarrow 0}\int_{-1+\epsilon}^{0} \frac{1}{\sqrt{1-x^2}}\,dx + \lim_{\epsilon\downarrow 0}\int_{0}^{1-\epsilon} \frac{1}{\sqrt{1-x^2}}\,dx \\ &= -\lim_{\epsilon\downarrow 0} \arcsin(-1+\epsilon) + \lim_{\epsilon\downarrow 0} \arcsin(1-\epsilon) = -(-\frac{\pi}{2}) + \frac{\pi}{2} = \pi.\end{aligned}$$

Anmerkung: Diese Integral tritt bei der Berechnung der Bogenlänge des Kreises auf.

Genauso wie man die Fläche unter einer Kurve mit Riemannschen Summen berechnet, kann man die Länge einer Funktionskurve als Grenzwert eines Polygonzuges definieren. Die Länge der einzelnen Polygonzüge berechnet man mit dem Satz von Pythagoras.

Sei f eine Funktion auf $[a,b]$. Dann nennt man

$$C_f = \{(x,y) \mid y = f(x), x \in [a,b]\}$$

die von f erzeugte Kurve. Sei Z eine Zerlegung von $[a,b]$, dann ergibt sich für die Länge des Polygonzuges $L_P(C_f)$

$$L_P(C_f) = \sum_{k=1}^{n} \sqrt{(x_k - x_{k-1})^2 + (f(x_k) - f(x_{k-1}))^2}. \tag{7.113}$$

Die **Länge der Kurve (Bogenlänge)** ist als das Supremum $\sup_P L_P(C_f)$ gegeben, falls dies ein endlicher Wert ist.

Man erhält daher für die Bogenlänge einer stetig differenzierbaren Funktion auf $[a,b]$ die Formel

$$L(C_f) = \int_a^b \sqrt{1 + (f'(x))^2}\,dx. \tag{7.114}$$

Der Kreis mit Radius 1 ist durch die Funktion $f(x) = \sqrt{1-x^2}, -1 \le x \le 1$ gegeben. Dies führt auf das Integral $\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,dx$, siehe oben.

Das Integral $\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx$ konvergiert, da

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx &= \lim_{R\to\infty}\int_{-R}^{0} \frac{1}{1+x^2}\,dx + \lim_{R\to\infty}\int_{0}^{R} \frac{1}{1+x^2}\,dx \\ &= -\lim_{R\to\infty} \arctan(-R) + \lim_{R\to\infty} \arctan(R) = \pi.\end{aligned}$$

Mit Hilfe der uneigentlichen Integrale kann man auch oft einfach entscheiden ob eine Reihe konvergiert oder divergiert.

Satz 7.118. (Integral-Vergleichskriterium.) *Sei $f : [1, \infty) \to \mathbb{R}^+$ eine monoton fallende Funktion. Dann gilt*

$$\sum_{n=1}^{\infty} f(n) \quad \textit{ist genau dann konvergent, wenn auch} \quad \int_1^{\infty} f(x)\,dx \quad \textit{konvergent ist.} \tag{7.115}$$

Betrachtet man die Summe der folgenden Reihe als Funktion von s so erhält man die *Riemannsche Zetafunktion*[13]

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad (s > 1) \tag{7.116}$$

die man für $s \in \mathbb{C}$ damit auch ins Komplexe fortsetzt.

Euler zeigte

$$\zeta(s) = \prod_{p\,\in\,prim} \frac{1}{1 - \frac{1}{p^s}} \tag{7.117}$$

und berechnete $\zeta(2) = \frac{\pi^2}{6}$. Die Zetafunktion erfüllt folgende Funktionalgleichung

$$\zeta(s) = 2^s \pi^{s-1} \sin\big(\frac{\pi}{2} s\big) \Gamma(1-s) \zeta(1-s). \tag{7.118}$$

7.11.0.1. *Die Gamma-Funktion.*

Definition 7.119. *Die* (Eulersche) Gamma-Funktion Γ *ist definiert durch das uneigentliche Integral*

$$\Gamma(x) := \int_0^{\infty} t^{x-1} \mathrm{e}^{-t}\,dt. \tag{7.119}$$

Anmerkung: Dieses uneigentliche Integral konvergiert.
Beweis: mit partieller Integration, siehe [**83**].

Für die Gamma-Funktion gilt

Satz 7.120.

$$\begin{aligned}
&\textit{(i)}\ \Gamma(x+1) = x\,\Gamma(x),\\
&\textit{(ii)}\ \Gamma(n+1) = n! \qquad \textit{für } n \in \mathbb{N}, \quad \Gamma(1) = 0! = 1,\\
&\textit{(iii)}\ \Gamma(x+n) = x(x+1)\cdots(x+n-1)\,\Gamma(x),\\
&\textit{(iv)}\ \Gamma(x) = \lim_{n\to\infty} \frac{n!\,n^x}{x(x+1)\cdots(x+n)}.
\end{aligned} \tag{7.120}$$

[13] https://de.wikipedia.org/wiki/Riemannsche_%CE%B6-Funktion

Mit (iii) kann man die Gammafunktion für $x < 0, x \notin \mathbb{Z}$ erweitern durch

$$\Gamma(x) = \frac{\Gamma(x+n)}{x(x+1)\cdots(x+n-1)} \tag{7.121}$$

wobei $0 < x + n < 1$ gelten muss.

Aus (iv) folgt, dass $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, (siehe [**83**]).

Für große n gilt für die Gammafunktion $\Gamma(n+1)$ *(Stirlingsche Formel)*

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}} \left(\frac{\mathrm{e}}{n}\right)^n = 1, \tag{7.122}$$

d. h. $n! \sim \sqrt{2\pi n}\, \left(\frac{n}{\mathrm{e}}\right)^n$.

7.11.1. Potenzreihen. Wir haben die Exponentialfunktion mit Hilfe einer unendlichen Reihe eingeführt und wollen nun dieses Konzept allgemeiner betrachten.

Definition 7.121. *Es sei $(c_n)_{n\in\mathbb{N}}$ eine Folge komplexer Zahlen und $a \in \mathbb{C}$. Dann heisst die Reihe*

$$f(z) = \sum_{n=0}^{\infty} c_n\,(z-a)^n \tag{7.123}$$

Potenzreihe.

Satz 7.122. *Die Potenzreihe $\sum_{n=0}^{\infty} c_n\,(z-a)^n$ konvergiere für ein $z_1 \in \mathbb{C}, z_1 \neq a$. Sei $r \in \mathbb{R}$ mit $0 < r < |z_1 - a|$ und $K(a,r) = \{z \in \mathbb{C} \mid |z-a| \leq r\}$, dann gilt:*

Die Potenzreihe konvergiert absolut und gleichmässig auf $K(a,r)$. Die formal differenzierte Potenzreihe

$$g(z) = \sum_{n=0}^{\infty} n c_n\,(z-a)^{n-1} \tag{7.124}$$

konvergiert ebenfalls absolut und gleichmässig auf $K(a,r)$.

Beweis: [**83**], Seite 223.

Definition 7.123. *Es sei $\sum_{n=0}^{\infty} c_n\,(z-a)^n$ eine Potenzreihe. Dann heisst*

$$R = \sup\{|z-a| \;\big|\; \sum_{n=0}^{\infty} c_n\,(z-a)^n \text{ ist konvergent}\} \tag{7.125}$$

Konvergenzradius *der Potenzreihe.*

Für den Konvergenzradius gilt die Hadamard'sche Formel

$$R = \left(\limsup_{n\to\infty} \sqrt[n]{|c_n|} \right)^{-1}, \tag{7.126}$$

wobei man vereinbart $0^{-1} = \infty$ und $\infty^{-1} = 0$.

7.11.2. Vertauschung von Limesbildungen. Bei gleichmässiger Konvergenz lassen sich Limesbildung und Integration vertauschen.

Satz 7.124. *Es sei $f_n : [a,b] \to \mathbb{R}, n \in \mathbb{N}$ eine Folge stetiger Funktionen. Die Folge f_n konvergiere auf $[a,b]$ gleichmässig gegen die Funktion $f : [a,b] \to \mathbb{R}$. Dann gilt*

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx. \tag{7.127}$$

Beweis: [**83**], Seite 225.

Punktweise Konvergenz der Folge f_n ist jedoch nicht ausreichend für die Richtigkeit dieses Satzes.

Satz 7.125. *Es sei $f_n : [a,b] \to \mathbb{R}, n \in \mathbb{N}$ eine Folge stetig differenzierbarer Funktionen, die punktweise gegen die Funktion $f : [a,b] \to \mathbb{R}$ konvergieren. Die Folge der Ableitungen $f_n' : [a,b] \to \mathbb{R}, n \in \mathbb{N}$ konvergiere gleichmässig auf $[a,b]$. dann ist f differenzierbar und es gilt*

$$f'(x) = \lim_{n\to\infty} f_n'(x) \quad \text{für alle } x \in [a,b] \tag{7.128}$$

Beweis: [**83**], Seite 228.

Es ist aber nicht ausreichend, dass f_n gleichmässig gegen die Funktion f konvergiert.

Als Anwendung dieses Satzen sei erwähnt, dass Potenzreihen beliebig oft gliedweise differenziert werden dürfen.
Beweis: [**83**], Seite 230.

7.12. Taylorreihen

Abbildung 7.20 zeigt, dass sich $\sin(x)$ in der Nähe von $a = 0$ immer besser annähern läßt, indem man Polynome von immer höherem Grad verwendet, zum Beispiel die Polynome x, $x - \frac{x^3}{6}$ und $x - \frac{x^3}{6} + \frac{x^5}{120}$. Wie findet man nun aber diese Polynome?

Schreiben wir das Polynom als

$$T_n(x,a) = \sum_{k=0}^{n} b_k (x-a)^k = b_0 + b_1(x-a) + \cdots + b_n(x-a)^n,$$

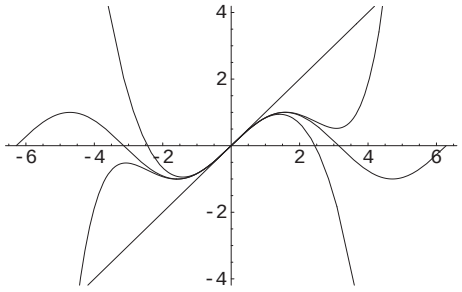


Abbildung 7.20. $\sin(x)$ approximiert durch x, $x - \frac{x^3}{6}$ und $x - \frac{x^3}{6} + \frac{x^5}{120}$

so läuft die Frage darauf hinaus, wie die Koeffizienten b_0, b_1, $b_2 \dots b_n$ zu wählen sind, damit die Approximation $f(x) \approx T_n(x, a)$ am Punkt a möglichst gut wird.

Man sieht, dass der Koeffizient b_j genau die j-te Ableitung des Taylorpolynoms $T_n(x, a)$ an der Stelle $x = a$ ist und es daher naheliegend diese Koeeffizienten gleich den entsprechenden Ableitungen der anzunähernden Funktion zu nehmen.

Satz 7.126. (Taylorpolynom). *Sei $f : I \to \mathbb{R}$ eine $(n+1)$-mal stetig differenzierbare Funktion und $a \in I$. Dann gilt für alle $x \in I$*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_{n+1}(x), \tag{7.129}$$

wobei

$$R_{n+1}(x) = \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\,dt.$$

Beweis durch Induktion nach n, [**83**], Seite 232.

Ist $f : I \to \mathbb{R}$ eine $(n+1)$-mal differenzierbare Funktion mit $f^{(n+1)} = 0$ für alle $x \in I$, dann ist f ein Polynom vom Grad $\leq n$.

Für das Restglied kann mit Hilfe des Mittelwertsatzes folgende Abschätzung gefunden werden.

Satz 7.127. (Lagrangesche Form des Restgliedes). *Sei $f : I \to \mathbb{R}$ eine $(n+1)$-mal stetig differenzierbare Funktion und $a \in I$. Dann existiert ein $\xi \in (a, x)$, sodass*

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\,(x-a)^{n+1} \tag{7.130}$$

Unter Verwendung des Landau-Symbols o lässt sich auch schreiben

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x-a)^k + o(|x-a|^n). \tag{7.131}$$

Satz 7.128. *Ist $f : I \to \mathbb{R}$ eine beliebig oft differenzierbare Funktion und $a \in I$. Dann heisst*

$$T_f(x,a) := \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k \tag{7.132}$$

die Taylorreihe *von f mit Entwicklungspunkt a.*

Die Taylorreihe konvergiert genau dann gegen die Funktion f, wenn das Restglied gegen 0 konvergiert.

Anmerkungen:
(i) Der Konvergenzradius der Taylorreihe ist nicht notwendigerweise > 0.
(ii) Falls die Taylorreihe von f konvergiert, konvergiert sie nicht notwendigerweise gegen f.

Ergänzung: Sei $a \in \mathbb{R}$ und

$$f(x) = \sum_{n=0}^{\infty} c_n (x-a)^n \tag{7.133}$$

eine Potenzreihe mit positiven Konvergenzradius $r \in (0, \infty]$. Dann ist die Taylorreihe der Funktion $f : (a-r, a+r) \to \mathbb{R}$ mit Entwicklungspunkt a gleich dieser Potenzreihe und konvergiert somit gegen f.

Satz 7.129. (Abelscher Grenzwertsatz). *Sei $\sum_{n=0}^{\infty} c_n$ eine konvergente Reihe reeller Zahlen. Dann konvergiert die Potenzreihe*

$$f(x) = \sum_{n=0}^{\infty} c_n x^n \tag{7.134}$$

gleichmässig auf dem Intervall $[0,1]$, stellt also dort eine stetige Funktion dar.

Beispiele:

(1) Die Taylorreihe der Exponentialfunktion mit Entwicklungspunkt $a = 0$ ist

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Sie konvergiert für alle $x \in \mathbb{R}$.

(2) Die Taylorreihen für Sinus und Cosinus konvergieren ebenfalls für alle $x \in \mathbb{R}$.

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

(3) Die Taylorreihe für den Logarithmus konvergiert für $-1 < x \leq 1$ und es gilt

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \mp \cdots = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n.$$

Damit ergibt sich für $x = 1$ und dem Abelschen Grenzwertsatz

$$\log(2) = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} \mp \cdots$$

Für die praktische Berechnung von $\log(2)$ ist diese Formel wegen der langsamen Konvergenz nicht geeignet.

Die Taylorreihe für $\log(1-x)$ konvergiert für $-1 < x < 1$ und es gilt

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots = -\sum_{n=1}^{\infty} \frac{x^n}{n}.$$

Damit ergibt sich für

$$\log \frac{1+x}{1-x} = 2 \sum_{n=0}^{\infty} \frac{1}{2n+1} x^{2n+1}. \tag{7.135}$$

(4) (Arcus-Tangens-Reihe). Für $|x| < 1$ gilt

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} \pm \ldots = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1}. \tag{7.136}$$

(5) Zur Verallgemeinerung der Binomischen Reihe definieren wir das Symbol

$$\binom{\alpha}{n} := \prod_{k=1}^{n} \frac{\alpha - k + 1}{k}. \tag{7.137}$$

für beliebige $\alpha \in \mathbb{R}$.

Für die allgemeine Binomische Reihe $\alpha \in \mathbb{R}$ ergibt sich folgende Taylorreihe, $|x| < 1$

$$(1+x)^{\alpha} = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n. \tag{7.138}$$

Beweis: [**83**], Seite 241.

Übung: Man berechne die ersten Glieder der allgemeinen Binomischen Reihe für $\alpha = -1, -\frac{1}{2}, \frac{1}{2}$.

7.13. Übung

(1) Man berechne das Integral $\int_0^b x^3\,dx$ mittels einer Riemannschen Summe.

(2) Man berechne folgende Integrale (Hinweis: Partialbruchzerlegung)

$$\int \frac{x^2-x+2}{x^3-x^2-x+1}\,dx, \qquad \int \frac{2x^2-x+1}{x^3-x^2-x+1}\,dx.$$

(3) Man berechne folgende Integrale:

$$\int \frac{dx}{\sqrt{x-1}\sqrt{x+1}}, \qquad \int \frac{2^x}{3^{x+1}}dx, \qquad \int 3^{\sqrt{2x+1}}dx.$$

(4) Man bestimme für $\alpha \in \mathbb{R}$ das Integral (Hinweis: Fallunterscheidung)

$$\int x^\alpha \log x\,dx.$$

(5) Man berechne durch geeignete Substitutionen

$$\int_1^4 \frac{\mathrm{e}^{\sqrt{x}}}{\sqrt{x}}\,dx, \qquad \int_0^1 \mathrm{e}^{\mathrm{e}^x}\mathrm{e}^x dx, \qquad \int_{\mathrm{e}}^{\mathrm{e}^2} \frac{\log(\log x)}{x\log x}dx.$$

(6) Man berechne durch partielle Integration

$$\int_0^2 x^2\mathrm{e}^x\,dx, \qquad \int_1^2 (\log x)^3 dx, \qquad \int_{\mathrm{e}}^{\mathrm{e}^2} \frac{\log(\log x)}{x}dx.$$

(7) Man berechne

$$\int_{-1}^1 x|\sin x|\,dx.$$

(8) Man berechne die Fläche, die zwischen den zwei Funktionen g und f zwischen den Punkten $x=0$ und $x=1$ liegt

$$g(x) = x^2\,\mathrm{e}^{x-1} - x, \qquad f(x) = \left(\sin(\frac{\pi}{2}x)\right)^2 \cos(\frac{\pi}{2}x).$$

(9) Die Leistung $P(t)$ eines Wasserkraftwerkes sei gegeben durch

$$P(t) = \mathrm{e}^t \sin(2t) + 1.$$

Man berechne die zwischen den Zeitpunkten t_0 und t_1 erzeugte Energie.

(10) Die Kraft F der Erde auf einen Körper mit Masse m ist gegeben durch das Newtonsche Gravitationsgesetz

$$F(r) = \gamma M m \frac{1}{r^2}, \tag{7.139}$$

wobei γ die Gravitationskonstante, M die Erdmasse und r der Abstand vom Erdmittelpunkt ist.

Man berechne die Arbeit W (Energie) die man benötigt um einen Körper mit Masse m von der Erdoberfläche (Erdradius $r_0 = 6371$ km) auf die Höhe $r_0 + h$ zu heben.

Welche Energie W_∞ ergibt sich im Grenzwert $h \to \infty$ und welche Geschwindigkeit (Fluchtgeschwindigkeit) ist dazu nötig? Hinweis: W_∞ muss gleich der kinetischen Energie sein.

(11) (i) Mit Hilfe eines Integrals beweise man die Ungleichung

$$\frac{1}{n+1} < \log(n+1) - \log(n) < \frac{1}{n}.$$

(ii) Die Folge a_n sei definiert durch

$$a_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \log(n).$$

Man zeige: a_n ist monoton fallend und nichtnegativ; d.h. a_n konvergiert. Der Grenzwert ist die Eulersche Konstante γ.

(12) (*) Man untersuche für welche Konstanten α, β die folgenden Integrale konvergieren.

$$\int_0^\infty \frac{x^\alpha}{1+x^\beta}\,dx, \qquad \int_0^\infty \mathrm{e}^{-x^\alpha}\,dx, \qquad \int_1^\infty \frac{(\log x)^\alpha}{x^\beta}\,dx,$$

(13) Bei der optimalen Stoppstrategie wird neben einer Ableitung (Optimumberechnung) die Wahrscheinlichkeit mit folgender Summe berechnet:

$$\sum_{k=j+1}^{n} \frac{1}{k}.$$

Man zeige folgende Abschätzung für diese Summe

$$\sum_{k=j+1}^{n} \frac{1}{k} \approx \log(n/j). \tag{7.140}$$

(14) (*) Bei der Berechnung der Hashwertekollision (vgl. Geburtstagsparadoxon) erhält man für die Wahrscheinlichkeit $P = 0.5$ den Ausdruck

$$1/2 = P = \frac{N!}{(N-n)!N^n}.$$

Man zeige (siehe G. & S. Teschl, Band 2 [**31**]), dass n wie folgt berechnet werden kann

$$n \approx \frac{1}{2} + \sqrt{\frac{1}{4} + \log(4)N} \approx \sqrt{\log(4)N}. \tag{7.141}$$

(15) (*) Mit Hilfe von $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ zeige man, dass

$$\int_0^\infty \mathrm{e}^{-x^2}\, dx = \frac{\sqrt{\pi}}{2}. \tag{7.142}$$

(16) Man bestimme das Taylorpolynom $T_4(x, 0)$ der Funktion $f(x) = \mathrm{e}^{\mathrm{e}^x}$.

(17) Man ermittle die Taylorreihen der folgenden Funktion an der Stelle $x = 0$ und deren Konvergenzbereich.

$$f(x) = \frac{1-x}{1+x}, \qquad f(x) = \log\left(\frac{1+x}{1-x}\right), \qquad f(x) = \cosh x.$$

(18) Man zeige, dass allgemein

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$$

gilt und berechne die ersten Glieder für $\alpha = -1, -\frac{1}{2}, \frac{1}{2}$.

(19) Man bestimme die Taylorreihe von $f(x) = \operatorname{Artanh} x$ an der Stelle $x = 0$.

Kapitel 8

Fourierreihen

Nehmen wir an, wir haben ein akustisches Signal[1] $f(t)$, das wir auf einem Computer abspeichern möchten. In der Praxis wird dazu die Amplitude (Funktionswert) des Signals in konstanten Zeitabständen, z.B. alle Hunderstelsekunden gemessen (Samplingrate von 100 Herz). Es liegen daher etwa nur die Funktionswerte zu den Zeitpunkten -0.50, -0.49, $\ldots$, 0.49, 0.50 vor. Das Signal könnte wie in Abbildung 8.1 aussehen. Um es zu speichern, wäre es naheliegend, die gemessenen 100 Funktionswerte abzuspeichern. Das ist vielleicht bei 100 Messwerten kein Problem, kann aber bei einer größeren Anzahl zu Speicherproblemen führen. Wie aber können wir unser Signal mit möglichst geringem Speicheraufwand aber zugleich möglichst geringem Informationsverlust abspeichern?

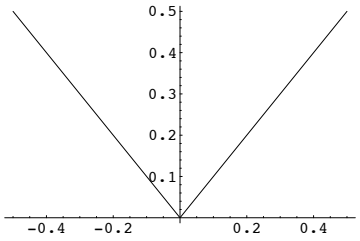


Abbildung 8.1. Akustisches Signal

Wir könnten das Signal zum Beispiel durch ein Taylorpolynom mit etwa zehn Koeffizienten approximieren. Dann müssten wir nur diese zehn Koeffizienten

[1] periodische Funktion; Anmerkung: jede Funktion auf einem endlichen Intervall T kann zu einer periodischen Funktion auf $\mathbb{R}$ fortgesetzt werden.

speichern, die ja bereits die gesamte Information des Taylorpolynoms enthalten. Das ist auch schon die richtige Idee, nur ergibt sich hier das Problem, dass wir nur die Funktionswerte, nicht aber die Ableitungen kennen, die aber für die Berechnung der Taylorkoeffizienten notwendig sind. Außerdem würde ein Taylorpolynom eine Näherung geben, die *lokal* um den Entwicklungspunkt t_0 sehr gut wäre, aber immer schlechter würde, je weiter wir uns vom Entwicklungspunkt entfernen. Wir suchen also eine Approximation des Signals, die *global* auf dem gesamten Zeitintervall $[-\frac{1}{2}, \frac{1}{2}]$ gut ist und für die die Kenntnis der Messwerte (Funktionswerte) ausreicht.

Es ist nun möglich, nahezu beliebige periodische Funktionen (sogar unstetige) durch ein sogenanntes *trigonometrisches Polynom* zu approximieren. Vergleichen wir zum Beispiel unser Signal, das durch die Funktion $f(t) = |t|$ auf $[-\frac{1}{2}, \frac{1}{2}]$ gegeben sein könnte, mit

$$F_3(t) = \frac{1}{4} - \frac{2}{\pi^2}\cos(2\pi t) - \frac{2}{9\pi^2}\cos(6\pi t).$$

Die Abbildung 8.2 zeigt, dass sich beide Funktionen nur geringfügig unterschei-

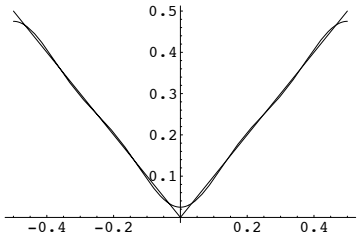


Abbildung 8.2. Signal $f(t) = |t|$ mit trigonometrischen Näherungspolynom

den und dass die Annäherung global auf dem gesamten Intervall $[-\frac{1}{2}, \frac{1}{2}]$ gut ist. Wenn man mit dieser Approximation zufrieden ist, dann würde es also reichen, nur die Zahlen (Koeffizienten) $\frac{1}{4}$, $-\frac{2}{\pi^2}$ und $-\frac{2}{9\pi^2}$ zu speichern! Man beachte, dass es dabei aber (weil es sich ja nur um eine Näherung handelt) zu einem Informationsverlust kommt.

Die gleichen Ideen kommen auch in der **Bildkompression** zur Anwendung. Beim **JPEG-Verfahren** wird zum Beispiel das Bild in Quadrate von 8×8 Bildpunkten aufgeteilt. Für die Bildpunkte werden dann die Werte der **Luminanz** Y (Helligkeit) und der **Chrominanzen** C_b, C_r (Farbwerte) betrachtet. Da das menschliche Auge Fehler bei der Helligkeit stärker wahrnimmt als Fehler bei der Farbe, genügt es, nur einen Teil der Chrominanzwerte zu verwenden, wodurch bereits eine Kompression erzielt wird. Diese Daten werden dann durch ein trigonometrisches Polynom dargestellt. (Da es nur endlich viele Punkte sind, ist es möglich ein trigonometrisches Polynom zu finden, das an diesen Punkten *exakt* mit den gegebenen Werten übereinstimmt. Die Anzahl der Koeffizienten bei exakter Darstellung ist allerdings gleich der Anzahl der Punkte und es würde in diesem Fall keine Kompression erreicht.) Durch die

Anzahl der Koeffizienten, die man abspeichert, kann man das Verhältnis zwischen Kompressionsrate und Qualitätsverlust wählen.

Wie erhält man nun eine solche trigonometrische Approximation und was steckt da mathematisch dahinter?

Definition 8.1. *Eine auf ganz $\mathbb{R}$ definierte reell- oder komplexwertige Funktion f heisst* periodisch *mit Periode $L > 0$, falls*

$$f(x + L) = f(x) \qquad \text{für alle } x \in \mathbb{R}. \tag{8.1}$$

Durch die Transformation

$$F(x) := f(\frac{2\pi}{L}x) \tag{8.2}$$

kann jede periodische Funktion f mit Periode L in eine periodische Funktion F mit Periode 2π übergeführt werden, sodass wir im Folgenden, der Einfachheit halber nur periodische Funktionen mit Periode 2π betrachten.

Definition 8.2. *Der Ausdruck*

$$\begin{aligned} F_n(x) &= \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos(kx) + b_k \sin(kx)) \qquad (8.3)\\ &= \frac{a_0}{2} + a_1 \cos(x) + \cdots + a_n \cos(nx) + b_1 \sin(x) + \cdots + b_n \sin(nx) \end{aligned}$$

heißt trigonometrisches Polynom *oder* Fourierpolynom *von der Ordnung n. Die Koeffizienten a_k und b_k heissen* Fourierkoeffizienten *von F_n.*

Es gelten folgende Beziehungen

$$\begin{aligned} &\int_0^{2\pi} \cos(kx)\cos(lx)\,dx = \int_0^{2\pi} \sin(kx)\sin(lx)\,dx = \pi\delta_{kl},\\ &\int_0^{2\pi} \cos(kx)\sin(lx)\,dx = 0, \qquad k, l \in \mathbb{Z}. \end{aligned} \tag{8.4}$$

Damit berechnet man für die Fourierkoeffizienten

$$a_k = \frac{1}{\pi}\int_0^{2\pi} \cos(kx) F_n(x)\,dx, \tag{8.5}$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} \sin(kx) F_n(x)\,dx. \tag{8.6}$$

Zum praktischen Rechnen ist es günstiger die komplexe Schreibweise zu verwenden. Da

$$\cos x = \frac{1}{2}(\mathrm{e}^{\mathrm{i}x} + \mathrm{e}^{-\mathrm{i}x}), \quad \sin x = \frac{1}{2\mathrm{i}}(\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}) \tag{8.7}$$

ist, erhält man

$$F_n(x) = \sum_{k=-n}^{n} c_k \mathrm{e}^{\mathrm{i}kx} \tag{8.8}$$

wobei für die Fourierkoeffizienten c_k gilt

$$c_0 = \frac{a_0}{2}, \quad c_k = \frac{1}{2}(a_k - \mathrm{i}b_k), \quad c_{-k} = \frac{1}{2}(a_k + \mathrm{i}b_k), \quad 1 \le k \le n.$$

Um in diesem Fall die Koeffizienten c_k aus F_n berechnen zu können, muss man zuerst das Integral einer komplexen Funktion definieren.

Definition 8.3. *Es sei $\phi = u + \mathrm{i}v$ eine komplexwertige Funktion auf $[a,b]$ mit Realteil u und Imaginärteil v, d. h. $u : [a,b] \to \mathbb{R}$, $v : [a,b] \to \mathbb{R}$ und $\phi : [a,b] \to \mathbb{C}$. Dann wird das Integral der komplexwertigen Funktion ϕ als Summe der Integrale von Imaginärteil v und Realteil u definiert*

$$\int_a^b \phi(x)\,dx = \int_a^b (u(x) + \mathrm{i}v(x))\,dx := \int_a^b u(x)\,dx + \mathrm{i}\int_a^b v(x)\,dx.$$

Für den Koeffizienten c_k ergibt sich damit

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}kx} F_n(x)\,dx, \quad -n \le k \le n. \tag{8.9}$$

Wir betrachten nun beliebige periodische Funktionen und definieren.

Definition 8.4. *Sei $f : \mathbb{R} \to \mathbb{C}$ eine periodische, auf dem Intervall $[0, 2\pi]$ integrierbare Funktion. Dann heissen die Zahlen*

$$c_k := \frac{1}{2\pi}\int_0^{2\pi} \mathrm{e}^{-\mathrm{i}kx} f(x)\,dx, \quad k \in \mathbb{Z} \tag{8.10}$$

die Fourierkoeffizienten *von f und die Reihe*

$$\sum_{k=-\infty}^{\infty} c_k \mathrm{e}^{\mathrm{i}kx} \tag{8.11}$$

heisst Fourierreihe *von f.*

Die Partialsummen lauten daher

$$S_n(x) = \sum_{k=-n}^{n} c_k \mathrm{e}^{\mathrm{i}kx}, \quad k \in \mathbb{N}.$$

In der Darstellung mit Winkelfunktion lautet dies

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{n} \left(a_k \cos(kx) + b_k \sin(kx)\right),$$

$$a_k = \frac{1}{\pi}\int_0^{2\pi} \cos(kx) f(x)\,dx, \quad b_k = \frac{1}{\pi}\int_0^{2\pi} \sin(kx) f(x)\,dx.$$

Es stellt sich nun die Frage, wann die Fourierreihe gegen die Funktion konvergiert und in welchem Sinne diese Konvergenz gemeint ist.

Dazu betrachten wir den Vektorraum $V_P[0, 2\pi]$ der periodischen Funktionen $f : \mathbb{R} \to \mathbb{C}$, die auf dem Intervall $[0, 2\pi]$ Riemann-integrierbar sind, und führen ein (Semi-) Skalarprodukt durch

$$\langle f, g\rangle := \int_0^{2\pi} \overline{f(x)} g(x)\,dx$$

ein. „Semi-“ deshalb, weil aus $\langle f, f\rangle = 0$ nicht geschlossen werden kann, dass $f = 0$ ist, falls f nicht stetig ist. Ansonsten sind alle Eigenschaften eines Skalarproduktes erfüllt.

Anmerkung: Führt man den allgemeineren Lebesgueschen Integralbegriff ein und bezeichnet mit $L^2(I)$ die Menge der „quadratintegrablen“ Funktionen auf dem Intervall I, kann man obiges Semiskalarprodukt zu einem Skalarprodukt vervollständigen.

Definiert man die Funktionen $e_k : \mathbb{R} \to \mathbb{C}$ durch

$$e_k(x) := \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\mathrm{i}kx} \tag{8.12}$$

so kann man die Fourierkoeffizienten einer Funktion $f \in V_P$ als

$$c_k = \frac{1}{\sqrt{2\pi}} \langle e_k, f\rangle \tag{8.13}$$

schreiben. Die Funktionen $e_k : \mathbb{R} \to \mathbb{C}$ bilden ein **Orthonormalsystem**, denn es gilt

$$\langle e_k, e_l\rangle = \delta_{kl}. \tag{8.14}$$

Satz 8.5. *Die Funktion $f \in V_P[0, 2\pi]$ habe die Fourierkoeffizienten $c_k, k \in \mathbb{Z}$. Dann gilt für alle $n \in \mathbb{N}$*

$$||f - \sum_{k=-n}^{n} c_k e_k||_2^2 = ||f||_2^2 - 2\pi \sum_{k=-n}^{n} |c_k|^2. \tag{8.15}$$

Daraus folgt unmittelbar die Besselsche Ungleichung.

Satz 8.6. *Es sei $f : \mathbb{R} \to \mathbb{C}$ eine periodische Funktionen auf $[0, 2\pi]$, die auf dem Intervall $[0, 2\pi]$ Riemann-integrierbar mit Fourierkoeffizienten $c_k, k \in \mathbb{Z}$. Dann gilt*

$$\sum_{k=-n}^{n} |c_k|^2 \le \frac{1}{2\pi}\int_0^{2\pi} |f(x)|^2\,dx. \tag{8.16}$$

Der Konvergenzbegriff, der die Konvergenz einer Fourierreihe angemessen beschreibt ist der Begriff der Konvergenz in der Norm die durch das (Semi-) Skalarprodukt definiert wird. D. h. die (Semi-) Halb-Norm einer Funktion $f \in V_P[0, 2\pi]$ ist gegeben durch

$$||f||_2^2 = \langle f, f \rangle = \int_0^{2\pi} |f(x)|^2 \, dx$$

Definition 8.7. *Man sagt, die Folge (f_n) konvergiere im* quadratischen Mittel *gegen f, falls*

$$\lim_{n\to\infty} ||f - f_n||_2 = 0, \tag{8.17}$$

d. h., wenn das quadratische Mittel der Abweichung zwischen f und f_n, nämlich

$$\int_0^{2\pi} |f(x) - f_n(x)|^2 \, dx \tag{8.18}$$

für $n \to \infty$ gegen 0 konvergiert.

Konvergiert die Folge (f_n) gleichmässig gegen f, so konvergiert sie auch im quadratischen Mittel. Die Umkehrung gilt aber nicht. Man kann nicht einmal punktweise Konvergenz folgern.

Satz 8.8. *Es sei $f : \mathbb{R} \to \mathbb{C}$ eine periodische, auf dem Intervall $[0, 2\pi]$ Riemann-integrierbare Funktion. Dann konvergiert die Fourierreihe von f im quadratischen Mittel gegen f. Sind c_k die Fourierkoeffizienten von f, so gilt die Vollständigkeitsrelation*

$$\sum_{-\infty}^{\infty} |c_k|^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 \, dx. \tag{8.19}$$

Beweis: [**83**], Seite 256.

Unter folgenden Bedingungen kann auf gleichmässige Konvergenz geschlossen werden

Satz 8.9. *Es sei $f : \mathbb{R} \to \mathbb{R}$ eine stetige periodische, die stückweise stetig differenzierbar ist, d. h. es gebe eine Unterteilung*

$$0 = t_0 < t_1 < \ldots t_r = 2\pi$$

von $[0, 2\pi]$, sodass $f|[t_{j-1}, t_j]$ für $1 \leq j \leq r$ s stetig differenzierbar ist. Dann konvergiert die Fourierreihe von f gleichmässig gegen f.

Beweis: [**83**], Seite 259.

Auf punktweise Konvergenz kann im folgenden Fall geschlossen werden.

Satz 8.10. *Es sei $f : \mathbb{R} \to \mathbb{C}$ eine periodische, auf dem Intervall $[0, 2\pi]$ Riemann-integrierbare Funktion. An der Stelle x_0 sollen die Grenzwerte $\lim_{x\downarrow x_0} f(x)$ und $\lim_{x\uparrow x_0} f(x)$, sowie auch die links wie auch rechtsseitigen Ableitungen $\lim_{x\downarrow x_0} f'(x)$ und $\lim_{x\uparrow x_0} f'(x)$. Dann gilt:*

Die Fourierreihe von f konvergiert gegen das arithmetische Mittel

$$\frac{1}{2}\big(\lim_{x\downarrow x_0} f(x) + \lim_{x\uparrow x_0} f(x)\big) \tag{8.20}$$

Anmerkung: Ist f stetig in x_0 so konvergiert die Fourierreihe von f gegen $f(x_0)$.

Beweis: [**82**] II, Seite 122.

Beispiel 8.1. Fourierpolynom
Man berechne das Fourierpolynom der Ordnung 3 für $f(t) = |t|, t \in [-\frac{1}{2}, \frac{1}{2}]$.

Lösung zu 8.1. *Natürlich gibt es dafür einen Befehl in* `Mathematica`*, den wir verwenden wollen. Dazu muss allerdings zunächst ein Zusatzpaket (ist in jeder* `Mathematica`*-Version enthalten) geladen werden:*

```
In[50]:=Needs["Calculus`FourierTransform`"];
```

Nun können wir das Fourierpolynom berechnen:

```
In[51]:=FourierTrigSeries[Abs[t], t, 3]
```

$$\texttt{Out[51]=}\frac{1}{4} - \frac{2\,\texttt{Cos}[2\pi\texttt{t}]}{\pi^2} - \frac{2\,\texttt{Cos}[6\pi\texttt{t}]}{9\pi^2}$$

`Mathematica` *nimmt als Intervall automatisch $[-\frac{1}{2}, \frac{1}{2}]$ an, den allgemeinen Fall $[-\frac{T}{2}, \frac{T}{2}]$ erhält man mit der Option* `FourierParameters` $\to \{0, \frac{1}{T}\}$. ■

Man beachte, dass die Approximation nur auf dem Intervall $[-\frac{T}{2}, \frac{T}{2}]$ erfolgt, also in obigem Beispiel $|t| \approx \frac{1}{4} + \frac{2\cos(2\pi t)}{\pi^2} - \frac{2\cos(6\pi t)}{9\pi^2}$ nur für $t \in [-\frac{1}{2}, \frac{1}{2}]$ gilt. Eine Approximation auf ganz $\mathbb{R}$ ist nur gegeben, wenn $f(t)$ periodisch mit der Periode T ist. Anstelle des Intervalls $[-\frac{T}{2}, \frac{T}{2}]$ kann auch $[0, T]$ (oder jedes andere Intervall der Länge T) verwendet werden. Dafür müssen nur die Integrationsgrenzen entsprechend geändert werden. Das Fourierpolynom ändert sich dabei nicht, wenn die Funktion f periodisch mit der Periode T ist.

In der Praxis ist meist nicht die Funktionsgleichung von $f(t)$ gegeben, sondern nur die Funktionswerte $f(t_k)$ zu bestimmten gleichabständigen Zeitpunkten $t_0 = -\frac{T}{2}$, $t_1 = t_0 + \Delta_n$, $t_2 = t_0 + 2\Delta_n$ mit $\Delta_n = \frac{T}{n}$. Dann können die Fourierkoeffizienten a_k und b_k näherungsweise mit folgender Formel berechnet werden.

$$a_k = \frac{2}{n}\sum_{k=0}^{n-1} \cos(k\omega t_k) f(t_k), \tag{8.21}$$

$$b_k = \frac{2}{n}\sum_{k=0}^{n-1} \sin(k\omega t_k) f(t_k). \tag{8.22}$$

Die zugehörgie Transformation ist als **diskrete Fouriertransformation** bekannt und die Funktion

$$F_m(t) = \frac{a_0}{2} + \sum_{k=1}^{m} a_k \cos(k\omega t) + b_k \sin(k\omega t)$$

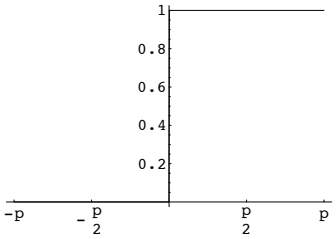
stimmt mit f an den Stützstellen t_k überein, falls die Anzahl $2m+1$ der Koeffizienten gleich der Anzahl n der Stützstellen ist. Ist $2m+1 < n$ so ist F_m bestapproximierend in dem Sinn, dass die Summe der quadratischen Abweichungen an den Stützstellen minimal ist. Verwendet man obige Formeln für die Berechnung von a_k, b_k so sind $O(n^2)$ Rechenoperationen notwendig. In der Praxis ist aber oft n sehr gross und man verwendet dann einen Algorithmus, der als **schnelle Fouriertransformation** (FFT - **F**ast **F**ourier **T**ransform) bekannt ist und nur $O(n \log_2(n))$ Rechenoperationen benötigt.

Ein weiteres Beispiel

Beispiel 8.2. Fourierreihe eines Rechtecksimpulses
Gegeben ist die Funktion $f(t) = 1$ für $t > 0$ und $f(t) = 0$ für $t < 0$ im Intervall $[-\pi, \pi]$. Berechnen Sie die zugehörige Fourierreihe.

Lösung zu 8.2. *Wir definieren zunächst die Funktion*

```
In[52]:=f[t_] := If[t > 0, 1, 0]; Plot[f[t], {t, -π, π}];
```



und berechnen

```
In[53]:=a[0] = 1/π ∫_-π^π f[t]dt
```

```
Out[53]=1
```

Weiters gilt

```
In[54]:=a[k_] = 1/π ∫_-π^π f[t]Cos[k t]dt
```

```
Out[54]= Sin[kπ]/(kπ)
```

Das ist aber für ganzzahliges k gleich Null, wie auch mit **Mathematica** *leicht zu sehen ist:*

```
In[55]:=Simplify[%, k ∈ Integers]
```

```
Out[55]=0
```

(Die Option `k ∈ Integers` *legt fest, dass k ganzzahlig ist.) Analog berechnen wir*

In[56]:=b[k_] = $\frac{1}{\pi}\int_{-\pi}^{\pi}$ f[t]Sin[k t]dt

Out[56]= $\frac{1-\text{Cos}[k\pi]}{k\pi}$

was sich ebenfalls noch etwas vereinfachen läßt:

In[57]:=Simplify[π%, k $\in$ Integers]

Out[57]= $-\frac{-1+(-1)^k}{k}$

Berechnen wir noch explizit die ersten Koeffizienten b_k:

In[58]:={b[1], b[2], b[3], b[4], b[5], b[6]}

Out[58]= $\{\frac{2}{\pi}, 0, \frac{2}{3\pi}, 0, \frac{2}{5\pi}, 0\}$.

Zusammenfassend erhalten wir also die Fourierreihe

$$f(t) = \frac{1}{2} + \sum_{k=1}^{\infty} \frac{1-(-1)^k}{k\pi} \sin(kt) \tag{8.23}$$

$$= \frac{1}{2} + \frac{2}{\pi}\sin(t) + \frac{2}{3\pi}\sin(3t) + \frac{2}{5\pi}\sin(5t) + \ldots \tag{8.24}$$

Abbildung 8.3 veranschaulicht, dass der Rechtecksimpuls immer besser angenähert wird, je höher der Grad des Fourierpolynoms ist. ■

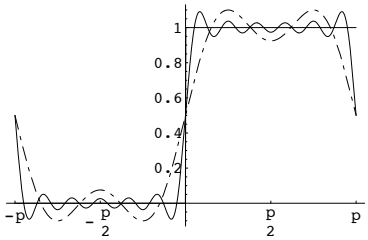


Abbildung 8.3. Rechtecksimpuls mit zugehörigen Fourierpolynomen $F_3(t)$ und $F_{12}(t)$.

Es ist bemerkenswert, dass eine Funktion mit Sprungstelle, wie der Rechtecksimpuls, durch eine Reihe von „glatten“ Sinusfunktionen exakt dargestellt werden kann. An der Stelle $t = 0$ ergibt die Fourierreihe den Wert $\frac{1}{2}$, das ist genau der arithmetische Mittelwert aus dem links- und dem rechtsseitigen Grenzwert bei $t = 0$.

Man beachte nochmals den Unterschied zwischen Taylor- und Fourierreihen: während die Näherung eines Funktionswertes $f(t)$ durch ein Taylorpolynom

lokal in der Nähe der Entwicklungsstelle t_0 am besten ist, also im allgemeinen schlechter wird, je weiter t von t_0 entfernt ist, bietet ein Fourierpolynom eine *globale* Näherung auf einem endlichen Intervall. Weiters sind zur Berechnung eines Taylorpolynoms die Ableitungen an der Entwicklungsstelle notwendig, für die Berechnung eines Fourierpolynoms benötigt man die Funktionswerte.

Ähnliche Reihen lassen sich nicht nur mit trigonometrischen sondern auch mit anderen Funktionen aufstellen. Da sich trigonometrische Funktionen am Computer nur indirekt (z.B.) über Näherungspolynome berechnen lassen, verwendet man in der Praxis oft sogenannte **Wavelets**, die sich am Computer einfacher berechnen lassen und zusätzlich dem Problem angepasst werden können.

Fourierreihen kann man sich auch als die Überlagerung von **Eigenschwingungen** eines schwingenden Systems vorstellen: Betrachten wir eine an beiden Enden eingespannte Saite. Dann sind die zugehörigen Eigenschwingungen gerade die Funktionen $\sin(2\pi kx)$, $k \in \mathbb{N}$, und unser Satz 8.8 sagt nun aus, dass sich *jede beliebige* Schwingung der Saite als Summe (Überlagerung) von Eigenschwingungen mit

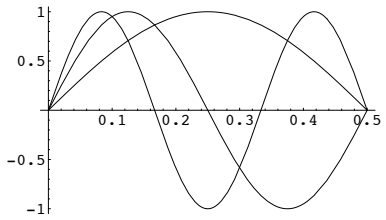


Abbildung 8.4. Eigenschwingungen $\sin(2\pi x)$, $\sin(4\pi x)$ und $\sin(6\pi x)$

verschiedenen Amplituden darstellen läßt. Abbildung 8.4 zeigt die ersten 3 Eigenschwingungen. Diese Sichtweise ist für viele Anwendungen von fundamentaler Bedeutung.

8.1. Übung

(1) Man berechne die Fourierreihe folgender periodischen Funktion

$$f(x) = \begin{cases} \frac{\pi}{4} & \text{für } 0 \le x < \pi \\ -\frac{\pi}{4} & \text{für } \pi \le x < 2\pi \end{cases} \quad \text{und } f(x+2\pi) = f(x).$$

Man probiere dieses Beispiel mit der Fourierreihe mittels
(i) $\sin kx$, $\cos kx$ und
(ii) mit komplexer Schreibweise $\exp(ikx)$.

(2) Man berechne die Fourierreihe folgender periodischen Funktion

$$f(x) = \begin{cases} x & \text{für } 0 \le x < \pi \\ 0 & \text{für } \pi \le x < 2\pi \end{cases} \quad \text{und } f(x+2\pi) = f(x).$$

(3) Man berechne die Fourierreihe folgender periodischen Funktion

$$f(x) = \begin{cases} |\sin(x)| & \text{für } 0 \le x < \pi \\ 0 & \text{für } \pi \le x < 2\pi \end{cases} \quad \text{und } f(x+2\pi) = f(x).$$

(4) Man berechne die Fourierreihe folgender periodischen Funktion

$$f(x) = x - \pi \text{ für } 0 \le x < 2\,\pi \quad \text{und} \quad f(x+2\pi) = f(x).$$

(5) Man zeige

$$\frac{1}{2\pi} \int_0^{2\pi} e^{\mathrm{i}kx} e^{-\mathrm{i}\ell x}\, dx = \delta_{k,\ell}. \tag{8.25}$$

(6) (*) Man zeige das Riemann'sche Lemma. Es sei $f : \mathbb{R} \to \mathbb{C}$ eine periodische, auf dem Intervall $[0, 2\pi]$ Riemann-integrierbare Funktion. Dann gilt:

$$\begin{aligned} \lim_{k\to\infty} \int_0^{2\pi} f(x) \sin kx\, dx = 0, \\ \lim_{k\to\infty} \int_0^{2\pi} f(x) \cos kx\, dx = 0. \end{aligned} \tag{8.26}$$

Beweis: [**82**] II, Kapitel 4.4

(7) (*) Man zeige: Gerade Funktionen lassen sich in reine Cosinusreihen und ungerade Funktionen reine Sinusreihen entwickeln.

Kapitel 9

Näherungsmethoden

9.1. Iterationsverfahren zur Bestimmung von Nullstellen

In den vorhergehenden Abschnitten haben wir gesehen, dass die Lösung vieler Probleme auf die Bestimmung von Nullstellen hinausläuft. Die Frage, die sich nun stellt, ist, wie sich Nullstellen in der Praxis finden lassen.

Die erste wichtige Beobachtung ist: Wenn wir zwei Stellen x_1 und x_2 haben, für die $f(x_1)$ und $f(x_2)$ verschiedenes Vorzeichen haben, dann liegt zwischen x_1 und x_2 mindestens eine Nullstelle (falls f stetig ist). Anschaulich ist das klar, denn wenn wir den Funktionsgraphen vom Punkt $(x_1, f(x_1))$ nach $(x_2, f(x_2))$ verfolgen, so müssen wir irgendwann die x-Achse schneiden, da wir wegen der Stetigkeit nicht springen können.

Haben also die Funktionswerte $f(x_1)$ und $f(x_2)$ verschiedene Vorzeichen, so liegt zwischen x_1 und x_2 eine Nullstelle. Wie aber finden wir sie? Betrachten wir Abbildung 9.1. Wenn wir eine Sekante durch die beiden Punkte $(x_1, f(x_1))$ und $(x_2, f(x_2))$ legen, dann können wir die Nullstelle x_3 der Sekante als Näherungswert für die gesuchte Nullstelle nehmen.

Diese Nullstelle der Sekante ist (nach einer kleinen Rechnung) gegeben durch

$$x_3 = x_2 - f(x_2)\frac{x_2 - x_1}{f(x_2) - f(x_1)} = \frac{x_1 f(x_2) - x_2 f(x_1)}{f(x_2) - f(x_1)}.$$

Ist nun zufälligerweise $f(x_3) = 0$, also x_3 auch eine Nullstelle unserer Funktion f, so können wir Feierabend machen. Ansonsten müssen wir das Verfahren wiederholen: wir legen eine Sekante durch $(x_1, f(x_1))$ und $(x_3, f(x_3))$ und erhalten

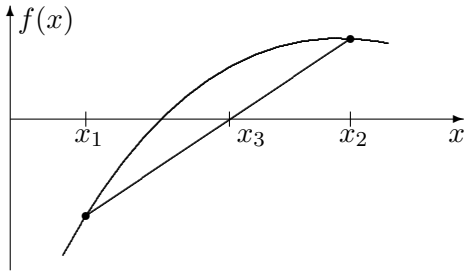


Abbildung 9.1. Regula falsi

so mit der Nullstelle dieser Sekante einen weiteren Näherungswert x_4, usw. Mit etwas Glück erhalten wir auf diese Weise eine Folge von Näherungswerten

$$x_n = x_{n-1} - f(x_{n-1})\frac{x_{n-1} - x_{n-2}}{f(x_{n-1}) - f(x_{n-2})},$$

die gegen eine Nullstelle von f konvergiert. Dieses Verfahren ist als „**Regula falsi**“ bekannt.

Beispiel 9.1. Regula falsi
Bestimmen Sie die Nullstelle von $f(x) = x^3 + 2x^2 + 10x - 20$.

Lösung zu 9.1. *Wir definieren zunächst*

```
In[59]:=f[x_] := x^3 + 2x^2 + 10x - 20
```

und die Rekursion lautet

```
In[60]:=x[n_] := x[n - 1] - f[x[n - 1]] (x[n - 1] - x[n - 2])/(f[x[n - 1]] - f[x[n - 2]])
```

An den Stellen $x_0 = 1$, $x_1 = 2$ *haben die zugehörigen Funktionswerte verschiedenes Vorzeichen*

```
In[61]:={f[1], f[2]}
Out[61]={-7, 16}
```

und wir verwenden sie daher als Startwerte. Damit lauten die ersten vier Näherungswerte

```
In[62]:=x[0] = 1.; x[1] = 2.; {x[2], x[3], x[4], x[5]}
Out[62]={1.30435, 1.35791, 1.36901, 1.36881}
```

und der letzte stimmt bereits auf fünf Stellen genau mit der gesuchten Nullstelle überein. ■

Leonardo von Pisa hat diese Funktion bereits im Jahr 1225 untersucht. Er konnte die Nullstelle auf mehrere Stellen genau berechnen. Niemand weiß, welche Methode er dazu verwendet hat, es ist aber ein für jene Zeit beachtliches Ergebnis.

Die Konvergenzgeschwindigkeit des Verfahrens kann verbessert werden, indem man anstelle von Sekanten Tangenten verwendet. Dabei wählt man einen Start-

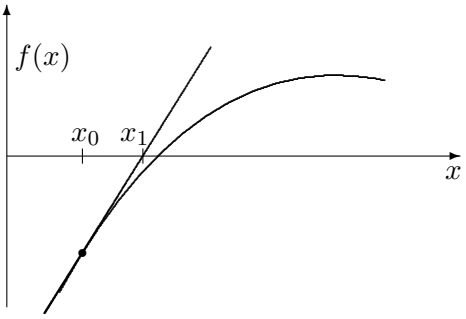


Abbildung 9.2. Newton-Verfahren

wert x_0 und legt im Punkt $(x_0, f(x_0))$ die Tangente an die Kurve. Die Nullstelle x_1 der Tangente ist dann gegeben durch $x_1 = x_0 - f(x_0)/f'(x_0)$ und liefert uns einen ersten Näherungswert. So geht es weiter: wir legen in $(x_1, f(x_1))$ die Tangente an die Kurve und bestimmen deren Nullstelle x_2 usw.

Definition 9.1. *Die Folge von Näherungswerten ist rekursiv durch*

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \tag{9.1}$$

definiert und ist als **Newtonverfahren** *bekannt.*

Beispiel 9.2. Newtonverfahren
Bestimmen Sie eine Lösung der Gleichung $x^2 = 2$.

Lösung zu 9.2. *Die Lösungen der Gleichung* $x^2 = 2$ *sind genau die Nullstellen von* $f(x) = x^2 - 2$. *Wir definieren also*

```
In[63]:=f[x_] := x^2 - 2;
```

$$\texttt{x[n_]} := \texttt{x[n-1]} - \frac{\texttt{f[x[n-1]]}}{\texttt{f'[x[n-1]]}}$$

und erhalten mit dem Startwert $x_0 = 1$

```
In[64]:=x[0] = 1.; {x[1], x[2], x[3], x[4]}
```

```
Out[64]={1.5, 1.41667, 1.41422, 1.41421}
```

Die Folge konvergiert also gegen $\sqrt{2}$ *und ist nichts anderes als die Heron'sche Folge (rechnen Sie die rechte Seite der Rekursionsformel explizit aus)! In* ***Mathematica*** *kann auch einfach der Befehl* `FindRoot[f[x], {x, x0}]` *verwendet werden um mit dem Startwert* x_0 *einen Näherungswert für eine Nullstelle von* $f(x)$ *zu finden.* ■

Beim Newtonverfahren ist es wichtig einen geeigneten Startwert x_0 zu finden (das kann z.B. graphisch geschehen). Liegt der Startwert nicht nahe genug an der gesuchten Nullstelle, so kann es passieren, dass das Newtonverfahren gegen eine andere Nullstelle oder, noch schlimmer, überhaupt nicht konvergiert. Was aber „nahe genug" in der Praxis bedeutet, und wie der Bereich aussieht, für den das Newtonverfahren gegen eine bestimmte Nullstelle konvergiert, ist in der Regel ein kompliziertes Problem, für das es keine allgemeine Antwort gibt.

Um ihnen einen kleinen Einblick in dieses faszinierende Problem zu geben, betrachten wir die Gleichung $z^3 - 1 = 0$. Sie besitzt drei Nullstellen: eine reelle, $z = 1$, und zwei komplexe, $z = \frac{-1+\sqrt{3}\mathrm{i}}{2}$ beziehungsweise $z = \frac{-1-\sqrt{3}\mathrm{i}}{2}$. Führt man das Newtonverfahren für verschiedene Startwerte in der komplexen Ebene aus und färbt die Startwerte nach den Nullstellen ein, gegen die das Newtonverfahren konvergiert, so erhält man das Bild in Abbildung 9.3. Für alle roten Startwerte konvergiert das Verfahren gegen die Nullstelle $z = \frac{-1-\sqrt{3}\mathrm{i}}{2}$, für alle blauen gegen $z = \frac{-1+\sqrt{3}\mathrm{i}}{2}$ und für alle grünen gegen $z = 1$. Die drei Mengen sind offensichtlich nicht durch glatte Kurven begrenzt, sondern wechseln

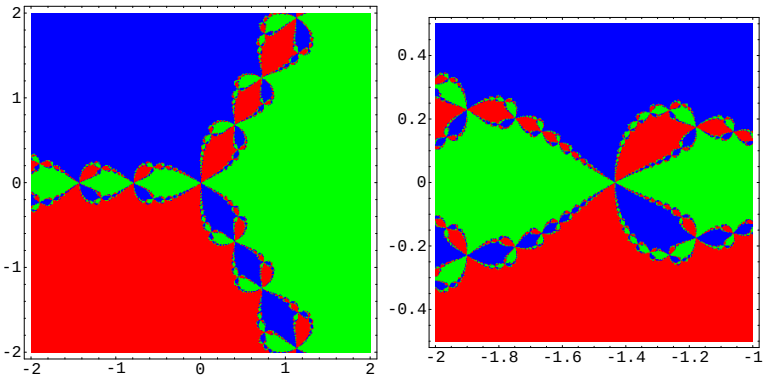


Abbildung 9.3. Konvergenzbereiche des Newtonverfahrens für $z^3 = 1$

sich im Grenzbereich immer schneller ab. Vergrößert man einen Ausschnitt, so wiederholen sich immer wieder die gleichen Strukturen (ähnlich wie bei einem Bild, das man erhält, wenn man sich zwischen zwei Spiegel stellt). Solche Mengen werden deshalb auch als **selbstähnlich** oder **fraktal** bezeichnet. Insbesondere kann eine offensichtlich extrem komplizierte Menge durch die Angabe einer einzelnen Funktion $f(z) = z^3 - 1$ charakterisiert werden. Diese Idee liegt der **fraktalen Bildkomprimierung**

zugrunde. Außerdem ist zu beachten, dass im Grenzbereich die kleinste Änderung im Startwert die Konvergenz zu einer anderen Nullstelle zur Folge haben kann, und dass das Verhalten der Iteration in diesem Bereich als ziemlich chaotisch eingestuft werden kann. Die Untersuchung dieser Phänomene führt also ins Reich des **Chaos** und der **fraktalen Mengen**.

9.2. Interpolation

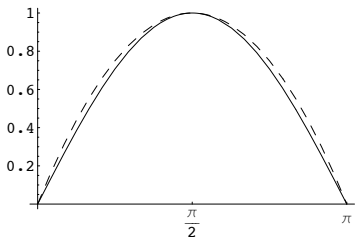
Aus der Schule kennen Sie sicher Funktionen wie Sinus, Kosinus oder die Exponentialfunktion. All diese Funktionen haben eines gemeinsam: da ein Computer nur die Grundrechenoperationen $+, -, \cdot, /$ beherrscht, können sie am Computer nicht direkt, sondern nur *näherungsweise* berechnet werden. Eine Möglichkeit ist nun, solche Funktionen durch *Polynome* anzunähern.

Von $\sin(x)$ kennt man zum Beispiel die Werte $\sin(0) = \sin(\pi) = 0$, $\sin(\frac{\pi}{2}) = 1$. Wir könnten versuchen $\sin(x)$ durch ein Polynom zu approximieren, das genau durch diese Punkte geht. Mit etwas Probieren ist es nicht schwer folgendes Polynom zu finden,

$$P_2(x) = \frac{4}{\pi^2} x(\pi - x),$$

das durch alle drei Punkte geht. Wie finden wir dieses Polynom durch Probieren? – Aufgrund der Nullstellen $x_1 = 0$ und $x_2 = \pi$ ist klar, dass sich das Polynom in der Form $P_2(x) = ax(x - \pi)$, mit Parameter a faktorisieren lässt. Der Parameter a folgt aus der Forderung, dass $P_2(\frac{\pi}{2}) = a\frac{\pi}{2}(\frac{\pi}{2} - \pi)$ gleich 1 sein muss.

Zeichnen wir beide Funktionen mit `Mathematica` so sehen wir, dass sich (auf dem Intervall $[0, \pi]$) eine recht gute Übereinstimmung ergibt. P_2 ist eine nach unten geöffnete Parabel, deren Nullstellen mit jenen von $\sin(x)$ auf diesem Intervall übereinstimmen, ebenso stimmen die Funktionswerte von $\sin(x)$ und Parabel an der Stelle $\frac{\pi}{2}$ überein.



Unsere Hoffnung ist natürlich, durch Hinzunahme weiterer Punkte, an denen $\sin(x)$ und Polynom übereinstimmen, eine noch bessere Näherung auf einem gewünschten Intervall zu erreichen. Dafür wird es aber notwendig sein, ein Polynom entsprechend hohen Grades zu nehmen, denn wir haben gerade gesehen, dass durch die Vorgabe von 3 Punkten das Polynom vom Grad 2 (die Parabel) eindeutig festgelegt war.

Allgemein steht man in der Praxis oft vor der Situation, dass man eine Anzahl von Punkten (x_j, y_j), auch **Stützpunkte** genannt, gegeben hat und ein Polynom $P(x)$ sucht, das $P(x_j) = y_j$ erfüllt.

Zum Beispiel werden die Punkte (x_j, y_j) durch Messungen erhalten oder man gibt sie aufgrund bestimmter Überlegungen vor, und man sucht nun eine Funktion, die genau durch diese Punkte geht.

Sind $n+1$ Punkte vorgegeben, so folgt, dass es genau ein Polynom, das maximalen Grad n hat, geben kann, das durch diese Punkte geht. (Wenn man die Beschränkung „maximalen Grad n“ aufhebt gibt es unendlich viele Polynome.) Mit anderen Worten: $n+1$ Punkte legen ein Polynom vom Grade kleiner gleich n eindeutig fest. Demnach ist eine Gerade durch 2 Punkte, eine Parabel durch 3 Punkte usw. eindeutig festgelegt.

Klar für $n = 0$: ein Polynom vom Grad 0 (also eine konstante Funktion) ist eindeutig durch einen einzigen Punkt festgelegt. Wie sieht es im Fall $n \geq 1$ aus: Angenommen, es gäbe zwei Polynome $p(x)$ und $q(x)$ vom Grad $n \geq 1$, die an $n+1$ Stellen übereinstimmen: $p(x_i) = q(x_i)$ für verschiedene Stellen (reell oder komplex) $x_1, \ldots, x_{n+1}$. Dann ist ihre Differenz $p(x) - q(x)$ ein Polynom vom Grad n (oder niedriger), das (mindestens) an diesen $n+1$ Stellen Nullstellen hat. Ein Polynom vom Grad $n \in \mathbb{N}$ kann nun aber nach dem Fundamentalsatz der Algebra maximal n Nullstellen haben. Also muss die Differenz das Nullpolynom sein, also $p(x) - q(x) = 0$ für alle x, d.h., beide Polynome sind gleich. Die Frage ist nun, ob es ein solches Polynom überhaupt gibt und wie wir es finden können.

Diese Aufgabe, durch vorgegebene Punkte ein Polynom zu legen, heisst Interpolationsproblem, bzw. genauer

Definition 9.2. *Für eine natürliche Zahl n seien $n+1$ verschiedene Zahlen (Stützstellen) $\{x_j\}, 0 \leq j \leq n$ und (Stützwerte) $\{y_j\}, 0 \leq j \leq n$ vorgegeben. Ein Polynom $P(x)$ vom Grad $\leq n$ mit $P(x_j) = y_j, j = 0, \ldots, n$ heisst Lösung des* Interpolationsproblem $\{(x_j, y_j)\}_{j=0}^{n}$.

Versuchen wir das Interpolationsproblem zu lösen. Wenn nur ein Punkt (x_0, y_0) vorgegeben ist, so ist dadurch ein Polynom vom Grad 0 eindeutig festgelegt:

$$P_0(x) = y_0$$

ist das gesuchte Polynom. Geben wir nun einen weiteren Punkt (x_1, y_1) vor (natürlich mit $x_1 \neq x_0$) und suchen nach einem Polynom vom Grad 1. Wenn wir es in der Form

$$P_1(x) = y_0 + k_1(x - x_0)$$

mit Parameter k_1 ansetzen, dann ist damit schon einmal garantiert, dass es durch (x_0, y_0) geht. Den Parameter k_1 finden wir durch die Forderung $y_1 =$

$P_1(x_1) = y_0 + k_1(x_1 - x_0)$. Dann ist mit

$$P_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0)$$

die Lösung für den Fall zweier Punkte gefunden. Nun ahnen Sie vielleicht schon wie es weiter geht: Wir setzen

$$P_2(x) = P_1(x) + k_2(x - x_0)(x - x_1)$$

an und bestimmen k_2 aus $y_2 = P_2(x_2) = P_1(x_2) + k_2(x_2 - x_0)(x_2 - x_1)$. Allgemein erhält man folgenden Algorithmus, der das Interpolationsproblem löst:

Satz 9.3. *Zu vorgegebenen Stützpunkten (x_j, y_j), $0 \leq j \leq n$ gibt es genau ein* **Interpolationspolynom** *$P_n(x)$ vom Grad n, das $P(x_j) = y_j$ erfüllt. Es kann rekursiv über*

$$P_0(x) = y_0 \tag{9.2}$$

$$P_{k+1}(x) = P_k(x) + (y_{k+1} - P_k(x_{k+1})) \prod_{j=0}^{k} \frac{x - x_j}{x_{k+1} - x_j} \tag{9.3}$$

ermittelt werden.

Es gibt noch andere Algorithmen um $P_n(x)$ zu berechnen. Da $P_n(x)$ eindeutig ist, liefern sie natürlich alle das gleiche Ergebnis. Obiger Algorithmus hat den Vorteil, dass er leicht implementiert werden kann und recht effektiv ist.

Weitere Algorithmen

`http://en.wikipedia.org/wiki/Lagrange_interpolation`

`http://en.wikipedia.org/wiki/Newton_polynomial`

Beispiel: Gegeben seien die Punkte $\{(-1, 15), (2, 6), (4, 10)\}$. Man berechne das Interpolationspolynom.

Nachdem wir nun wissen, wie wir zu gegebenen Stützpunkten das zugehörige Interpolationspolynom finden, stellt sich als nächstes die Frage ob durch Erhöhung der Anzahl der Stützpunkte eine immer bessere Übereinstimmung mit einer vorgegebenen Funktion $f(x)$ erreicht werden kann. Ob also durch Vorgabe von etwa 10 anstelle von 3 Punkten $\sin(x)$ auf dem Intervall $[0, \pi]$ besser genähert wird.

Leider ist das nicht unbedingt der Fall (zumindest nicht, wenn man die Stützpunkte *gleichmäßig* verteilt). Im Fall der trigonometrischen Funktionen gibt es zwar keine Probleme. Man kann zeigen, dass z.B. bei 11 gleichmäßig verteilten Stützpunkten im Intervall $[0, \pi]$ der Fehler zwischen $\sin(x)$ und dem Interpolationspolynom kleiner 10^{-8} ist. Aber zum Beispiel die Funktion

$$f(x) = \frac{1}{1 + 25x^2}$$

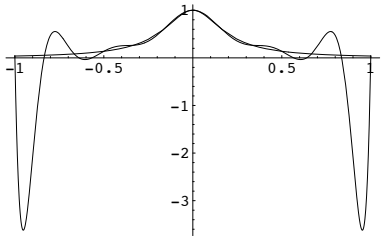
macht leider schon Probleme, wie man folgendermaßen sieht: Erzeugen wir zunächst mit `Mathematica` eine Liste von 13 Stützpunkten, die gleichmäßig über das Intervall $[-1, 1]$ verteilt sind und berechnen das zugehörige Interpolationspolynom mittels `InterpolationPolynomial`$[\{\{x_0, y_0\}, \dots\}, x]$:

```
In[65]:=f[x_] = 1/(1 + 25x^2);
        Stuetzpunkte[n_] := Table[{j/n, f[j/n]}, {j, -n, n}];
        P[x_] = InterpolatingPolynomial[Stuetzpunkte[6], x]//Expand
```

$$\text{Out[67]=} 1 - \frac{551599221900\,x^2}{28167484501} + \frac{367051586875\,x^4}{1847048164} - \frac{107641853578125\,x^6}{112669938004} + \frac{62017871484375\,x^8}{28167484501} - \frac{65809335937500\,x^{10}}{28167484501} + \frac{25628906250000\,x^{12}}{28167484501}$$

Zeichnen wir nun beide Funktionen, so erleben wir eine Überraschung:

```
In[68]:=Plot[{f[x], P[x]}, {x, -1, 1}, PlotRange → All];
```



Das Interpolationspolynom weist am Rand starke Oszillationen auf und stimmt dort auch überhaupt nicht mit unserer Funktion überein!? Das Verhalten kann zwar etwas verbessert werden, indem man die Stützpunkte am Rand dichter als in der Mitte wählt (**Tchebycheff Interpolation**), aber das prinzipielle Problem bleibt.

Deshalb verwendet man in der Praxis meist stückweise Polynome von kleinem Grad (meist 3). Jedes Polynom lebt dabei auf einem der Intervalle (x_j, x_{j+1}) und benachbarte Polynome werden an den Randpunkten (die genau die Stützpunkte sind) möglichst „glatt“ (also so, dass keine „Ecken“ entstehen) zusammengeklebt. Dieses Verfahren ist unter dem Namen **Splines** bekannt.

Es kann gezeigt werden, dass jede *stetige* Funktion auf einem Intervall $[a, b]$ beliebig genau durch Polynome approximiert werden kann (**Approximationssatz von Weierstrass**). Wie aber bei vorgegebenem Grad das Polynom mit dem kleinsten Fehler (d.h., die größte Abweichung $\max_{x\in[a,b]} |f(x)-P_n(x)|$ soll so klein wie möglich sein) gefunden werden kann, ist ein kompliziertes Problem.

9.3. Nummerische Integration

Bei der Berechnung von „bestimmten Integralen“ existiert in den meisten Fällen keine Stammfunktion. In diesen Fällen könnte man Riemannsche Summen verwenden.

Geschickter ist es jedoch, die Funktion $f(x)$ auf dem Intervall $[a,b]$ durch z.B. ein Polynom $P_n(x)$, das man leicht integrieren kann, zu approximieren.

Im Falle von $n=1$ erhält man mit den Stützstellen $\{a,b\}$ die Trapezformel

$$\int_a^b f(x)\,dx \approx I_T = \frac{b-a}{2}\left(f(a)+f(b)\right). \tag{9.4}$$

Im Falle $n=2$ erhält man die Simpsonformel, die auch noch **Keplersche Fassregel** genannt wird

$$\int_a^b f(x)\,dx \approx I_S = \frac{b-a}{6}\left(f(a)+4\,f(\frac{a+b}{2})+f(b)\right). \tag{9.5}$$

Hier sind die Stützstellen $\{a, \frac{a+b}{2}, b\}$.

Wenn man das Intervall $[a,b]$ in N gleiche Teile mit der Länge $\frac{b-a}{N}$ zerlegt und auf jedem dieser Teilintervalle die Trapezregel anwendet erhält man die große **Trapezregel** (mit $y_j := f(x_j)$)

$$\int_a^b f(x)\,dx \approx I_T^* = \frac{b-a}{2N}\left(y_0+y_N+2\sum_{j=1}^{N-1} y_j\right). \tag{9.6}$$

Satz 9.4. *Die große* **Simpsonformel** *erhält man bei Zerlegung des Intervalls $[a,b]$ in $2N$ gleiche Teile der Länge $\frac{b-a}{2N}$*

$$\int_a^b f(x)\,dx \approx I_S^* = \frac{b-a}{6N}\left(y_0+y_{2N}+2\sum_{j=1}^{N-1} y_{2j}+4\sum_{j=1}^{N} y_{2j-1}\right). \tag{9.7}$$

Bezüglich der Fehlerabschätzung sei auf [**82**] I, Kapitel 7.10 verwiesen.

9.4. Übung

Man löse die folgenden Beispiele indem man in Mathematica den entsprechenden Algorithmus zu Fuss implementiert, (d.h. keine eingebauten Routinen dafür verwendet):

(1) Durch Anwendung des Newton-Verfahrens bestimme man näherungsweise die Lösung von (Startwert x_0)

(a) $f(x) = x \log x - 1 = 0, \quad x_0 = 2,$

(b) $f(x) = \sin(\frac{5}{2}\pi x)\, \mathrm{e}^{(\frac{5}{2}\pi x)} + 1 = 0, \quad x_0 = 10.$

Man berechne explizit x_1 und $f(x_1)$ und überprüfe das Ergebnis mittels Berechnungen mit dem Mathematica-Befehl `FindRoot[ ]`.

(2) Man löse das Interpolationsproblem $\{(-1,0),(0,2),(1,4),(3,32)\}$. Man überprüfe das Ergebnis mit Berechnungen mittels des Mathematica-Befehls `InterpolatingPolynomial[ ]`.

(3) Man berechne mit der großen Trapezformel mit N beliebig und der großen Simpsonformel mit N beliebig das Integral

$$\int_0^1 e^{x^2}\, dx.$$

Man überprüfe das Ergebnis mit Berechnungen mittels des Mathematica-Befehls `NIntegrate[ ]` für $N = 4, 10, 100$.

(4) Man verbinde die Punkte $\{(0,0),(1,1),(3,3),(4,2),(6,4)\}$ möglichst „glatt“ durch Polynome 3. Ordnung (cubic splines[1]). Glatt heisst: an den Verbindungspunkten gilt: Funktion ist stetig, erste und zweite Ableitung ist stetig und die zweite Ableitung an den Rändern ist 0.

[1] siehe z.B. `http://mathworld.wolfram.com/CubicSpline.html`

Kapitel 10

Analysis 2

10.1. Differentialrechnung im $\mathbb{R}^n$

Bis jetzt haben wir im Rahmen der Differential- und Integralrechnung nur Abbildungen von $\mathbb{R}^1 \to \mathbb{R}^1$ bzw. auf Teilmengen D von $\mathbb{R}^1$ betrachtet und wollen dies nun auf Abbildungen von $\mathbb{R}^n \to \mathbb{R}^m$ erweitern.

Anmerkung: Lineare Abbildungen von $\mathbb{R}^n \to \mathbb{R}^m$ haben wir bereits im Kapitel *Lineare Algebra* eingeführt.

Zuerst betrachten wir ein paar einfache Beispiele:

(1) Eine Abbildung $I = [a, b] \subset \mathbb{R} \to \mathbb{R}^2$ (bzw. $\mathbb{R}^3$), $t \mapsto \mathbf{x}(t) = (x(t), y(t))$ heißt ebene Kurve (bzw. Raumkurve). Allgemein spricht man bei Abbildungen der Form $\mathbb{R}^1 \to \mathbb{R}^n$ auch von *vektorwertigen Funktionen*.

(i) Gerichtete Strecke (Abb. 10.1): $\mathbf{x} : I = [0, 1] \to \mathbb{R}^2$, $t \mapsto \mathbf{x}(t)$

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + t \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \end{pmatrix},$$
$$\mathbf{x}(t) = (1 - t)A + t\,B,$$
$$A = (a_1, a_2),\ B = (b_1, b_2),\ a_1, b_1, a_2, b_2 \in \mathbb{R}.$$

(ii) Ellipse (Abb. 10.2): $\mathbf{x} : I = [0, 2\pi) \to \mathbb{R}^2$, $\varphi \mapsto \mathbf{x}(\varphi)$

$$\begin{pmatrix} x(\varphi) \\ y(\varphi) \end{pmatrix} = \begin{pmatrix} r_1 \cos(\varphi) \\ r_2 \sin(\varphi) \end{pmatrix}, \qquad r_1, r_2 \in \mathbb{R}.$$

(iii) Neilsche Parabel: $\mathbf{x} : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto \mathbf{x}(t) = (t^2, a\,t^3)$, $\quad a \in \mathbb{R}$.

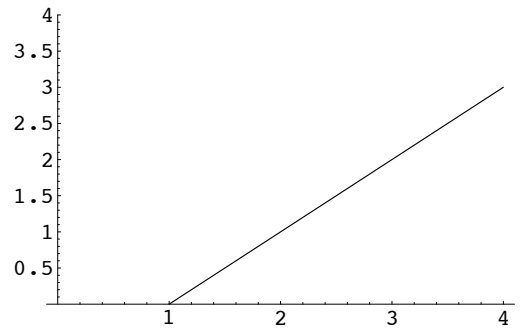


Abbildung 10.1. Gerichtete Strecke

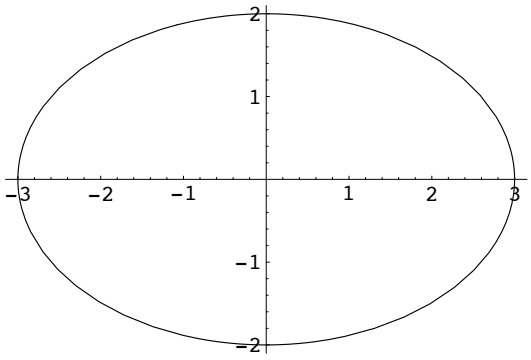


Abbildung 10.2. Ellipse

(iv) Zykloide (Abrollkurve): $\mathbf{x} : \mathbb{R} \to \mathbb{R}^2,\ t \mapsto \mathbf{x}(t) = \big(r\,(t - \sin t), r\,(1 - \cos t)\big),\quad r \in \mathbb{R}.$

(v) Lemniskate: $\mathbf{x} : I = [0, 2\pi) \to \mathbb{R}^2,\ t \mapsto \mathbf{x}(t)$.

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \frac{r\cos(t)}{\sin^2 t+1} \\ \frac{r\sin(t)\cos(t)}{\sin^2 t+1} \end{pmatrix}, \qquad r \in \mathbb{R}.$$

(vi) Lissajous-Figuren: $\mathbf{x} : \mathbb{R} \to \mathbb{R}^2$.

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} a\cos(\omega_1 t + \phi_1) \\ b\cos(\omega_2 t + \phi_2) \end{pmatrix}, \qquad a, b, \omega_1, \omega_2, \phi_1, \phi_2 \in \mathbb{R}.$$

(vii) Schraubenlinie (Abb. 10.3): $\mathbf{x} : I = [0, \infty) \to \mathbb{R}^3,\ t \mapsto \mathbf{x}(t)$

$$\begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} = \begin{pmatrix} r\cos(t) \\ r\sin(t) \\ h\,t \end{pmatrix}, \qquad r, h \in \mathbb{R}. \tag{10.1}$$

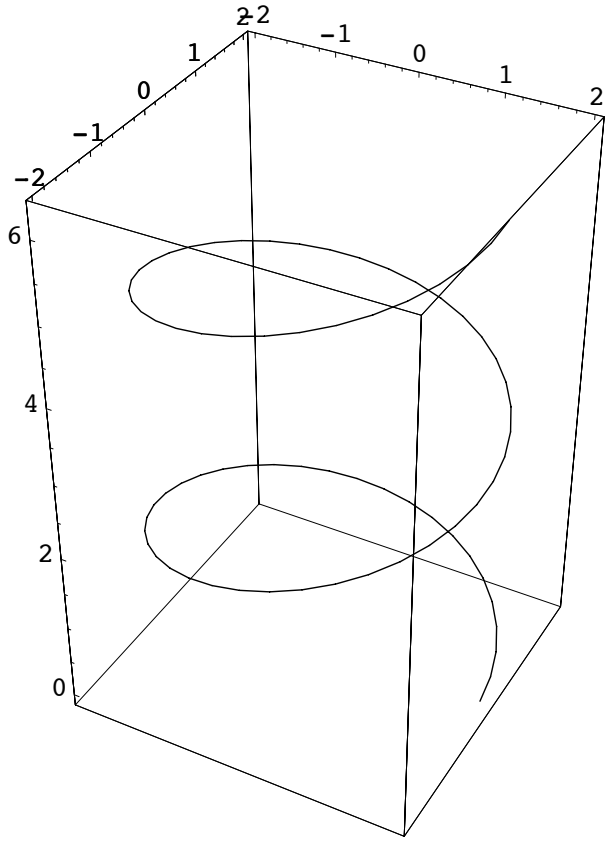


Abbildung 10.3. Schraubenlinie

Anmerkung: Interpretiert man t als Zeit, beschreibt dieses Beispiel die Bewegung eines Teilches im Raum auf einer Schraubenlinie.

(2) Fläche (Hyperfläche): $\mathbb{R}^2 \to \mathbb{R}^1$ (bzw. $\mathbb{R}^n \to \mathbb{R}^1$)

(i) Polynome $p : \mathbb{R}^2 \to \mathbb{R}$

$$p(x,y) = \sum_{j=0}^{M} \sum_{k=0}^{N} a_{j,k} x^j y^k. \tag{10.2}$$

(i1) Paraboloid (Abb. 10.4): $f : \mathbb{R}^2 \to \mathbb{R}^1$, $f(x,y) = x^2 + y^2$

(i2) Hyperbolisches Paraboloid (Abb. 10.5): $f : \mathbb{R}^2 \to \mathbb{R}^1$, $f(x,y) = x\,y$

(ii) (obere) Halbkugel (Abb. 10.6): $f : D \to \mathbb{R}^1, D = \{(x,y) \mid x^2 + y^2 \leq r^2\} \subset \mathbb{R}^2, r \in \mathbb{R}$,

$$f(x,y) = \sqrt{r^2 - (x^2 + y^2)} \tag{10.3}$$

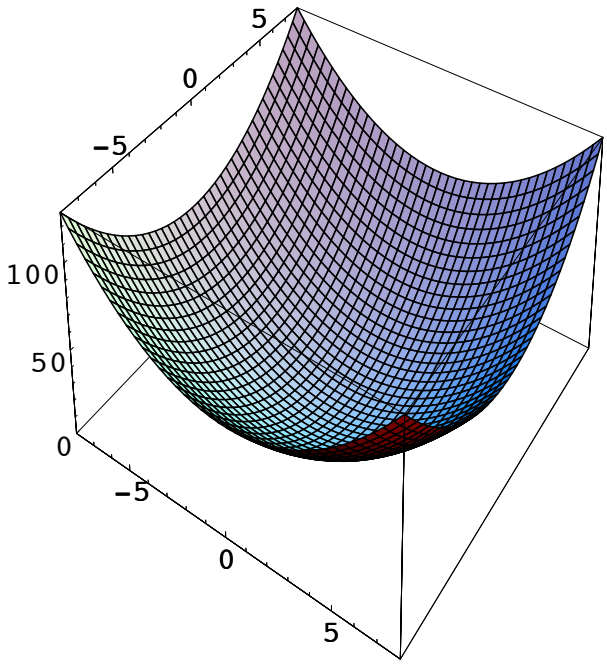


Abbildung 10.4. Paraboloid

Funktionen von mehreren Variablen treten natürlicherweise überall dort auf wo eine Größe von mehreren anderen abhängt. Beispiele:

(a) Die Temperatur T eines Gases (idealisiert) hängt vom Druck p und dem Volumen V ab. Als Funktion

$$T(p, V) = c\, p\, V, \tag{10.4}$$

wobei c eine Konstante ist ($c = 1/(\nu R) = 1/(N k_B)$).

(b) Die Anziehungskraft $F(x, y, z)$ zwischen zwei Körpern im Raum hängt von deren Entfernung ab

$$F(\mathbf{x}) = -\gamma\, m_1 m_2\, \frac{\mathbf{x}}{||\mathbf{x}||^3} \tag{10.5}$$

(c) Der Ertrag E eines Produktes hängt unter anderem von der produzierten Stückzahl p und der Anzahl der verkauften Produkte a ab: $E = E(p, a)$.

Anmerkung: Hat man insgesamt mehr als drei Dimensionen, muß man zur Veranschaulichung Einschränkungen vornehmen, z.B. man gibt bestimmten Variablen einen festen Wert oder betrachtet die Punkte die den gleichen Funktionswert haben (= **Höhenlinien**).

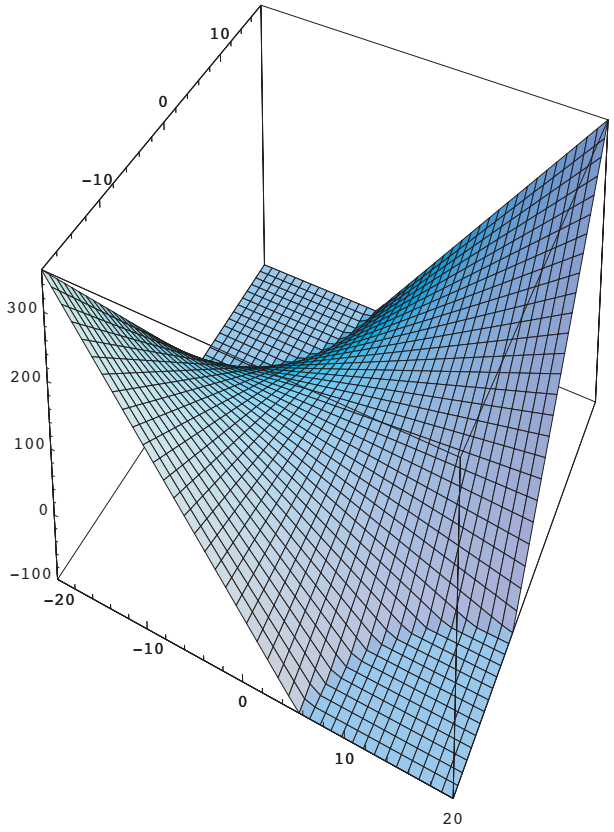


Abbildung 10.5. Hyperbolisches Paraboloid

Um nun Analysis betreiben zu können, braucht man, um den Begriff Konvergenz formulieren zu können, einen Abstandsbegriff im $\mathbb{R}^n$.

10.1.1. Normen auf dem Vektorraum $\mathbb{R}^n$. Im $\mathbb{R}^n$ können auf verschiedene Weisen Normen definiert werden, z. B.

$$||\mathbf{x}||_\infty = \max_{1\le j\le n} |x_j|, \qquad ||\mathbf{x}||_1 = \sum_{j=1}^{n} |x_j|, \qquad ||\mathbf{x}||_2 = \sqrt{\sum_{j=1}^{n} x_j^2},$$

$$||\mathbf{x}||_p = \sqrt[p]{\sum_{j=1}^{n} |x_j|^p}, \ 1 \le p < \infty. \tag{10.6}$$

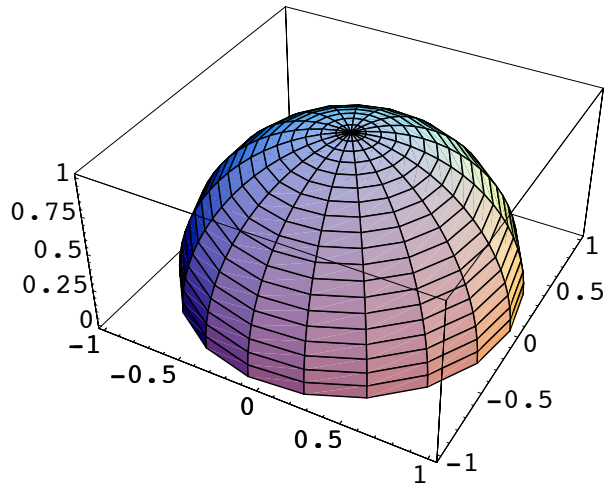


Abbildung 10.6. Halbkugel

Definition 10.1. *Zwei Normen* $||\ ||_r, ||\ ||_s$ *auf* $\mathbb{R}^n$ *heißen äquivalent, wenn es für alle* $\mathbf{x} \in \mathbb{R}^n$ *zwei reelle Zahlen* $a, b > 0$ *gibt, mit*

$$a||\mathbf{x}||_r \leq ||\mathbf{x}||_s \leq b||\mathbf{x}||_r. \tag{10.7}$$

Es gilt der folgende Satz

Satz 10.2. *Je zwei Normen auf* $\mathbb{R}^n$ *sind äquivalent.*

Das heißt man wird jeweils die Norm verwenden, die rechentechnisch für das jeweilige Problem am einfachsten zu handhaben ist.

Durch eine Norm wird eine Metrik (Abstand) $d(\mathbf{x}, \mathbf{y}) = ||\mathbf{x} - \mathbf{y}||$ erzeugt und daher läßt sich nun die Konvergenz einer Punktfolge definieren.

Definition 10.3. *Eine reelle Folge von Punkten* $\mathbf{x}^m = (x_1^m, \ldots, x_n^m)$ *heißt konvergent gegen einen Punkt* $\mathbf{x}^0 \in \mathbb{R}^n$ *(geschrieben* $\lim_{m\to\infty} \mathbf{x}^m = \mathbf{x}^0$*) falls gilt:*

Zu jeder reellen Zahl $\epsilon > 0$ *existiert eine Zahl* $M_\epsilon \in \mathbb{N}$*, sodass*

$$d(\mathbf{x}^m, \mathbf{x}^0) = ||\mathbf{x}^m - \mathbf{x}^0|| < \epsilon \tag{10.8}$$

für alle $m > M_\epsilon$ *aus* $\mathbb{N}$ *gilt.*

Übung: Wie schauen die entsprechenden ϵ-Umgebungen eines Punktes im $\mathbb{R}^2$ bezüglich der Normen 10.6 aus?

Es gilt der Satz der koordinatenweisen Konvergenz [**82**], Seite 140.

Satz 10.4. *Eine Punktfolge im $\mathbb{R}^n$ ist genau dann konvergent, wenn sie in jeder Koordinate konvergent ist.*

Auch im $\mathbb{R}^n$ gilt der Satz von Bolzano-Weierstrass.

Satz 10.5. *Jede beschränkte unendliche Menge $X \subset \mathbb{R}^n$ hat wenigstens einen Häufungspunkt.*

Anmerkung: Eine Menge X heißt *offen*, wenn jeder Punkt $x \in X$ eine Umgebung U besitzt die ganz in X enthalten ist, d.h. $x \in U, U \subset X$. Eine Menge heißt *abgeschlossen*, wenn ihr Komplement offen ist.

Eine beschränkte und abgeschlossene Teilmenge $X \subset \mathbb{R}^n$ heißt *kompakt*.

10.1.2. Stetigkeit. Wir betrachten zuerst Abbildungen von $\mathbb{R}^n \to \mathbb{R}$.

Definition 10.6. *(i) Eine Funktion (Abbildung) $f : \mathbb{R}^n \to \mathbb{R}$ heißt stetig in einem Punkt $\mathbf{x}^0 \in \mathbb{R}^n$, wenn für jede beliebige Folge $\mathbf{x}^m$ die gegen $\mathbf{x}^0$ konvergiert auch die Bildfolge $f(\mathbf{x}^m)$ gegen $f(\mathbf{x}^0)$ konvergiert.*

(ii) Eine Abbildung $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ heißt stetig in einem Punkt $\mathbf{x}^0 \in \mathbb{R}^n$, wenn sie in jeder Koordinatenfunktion $f_j, j = 1, \ldots, m$ in $\mathbf{x}^0$ stetig ist.

Man zeige: Alle linearen Abbildungen $A : \mathbb{R}^n \to \mathbb{R}^m$ sind stetig.

Alle früher formulierten Sätze über die Stetigkeit gelten auch für Funktionen von mehreren Variablen. Es sind also Summen, Produkte und Quotienten (Nennernullstellen ändern den Definitionsbereich) von stetigen Funktionen wieder stetig.

Eine *Richtung* im $\mathbb{R}^n$ ist gegeben durch einen Vektor $\mathbf{a}$ und eine Gerade durch $\mathbf{x}^0$ in Richtung von $\mathbf{a}$ durch

$$G(\mathbf{a}, \mathbf{x}^0) = \{\mathbf{x} \mid \mathbf{x} = \mathbf{x}^0 + t\mathbf{a}, t \in \mathbb{R}\}. \tag{10.9}$$

Damit kann nun der **Richtungsgrenzwert** g einer Funktion f in Richtung von $\mathbf{a}$ im Punkt $\mathbf{x}^0$ durch

$$g = \lim_{h \to 0} f(\mathbf{x}^0 + h\mathbf{a}) \tag{10.10}$$

definiert werden. (Man schreibt auch $\mathbf{a} - \lim_{\mathbf{x} \to \mathbf{x}^0} f(\mathbf{x})$.)

Es ist klar, daß für eine Funktion, welche an einem Punkt $\mathbf{x}^0$ den Grenzwert g besitzt, alle Richtungsgrenzwerte existieren und gleich sind. Die Umkehrung gilt aber nicht wie folgendes Beispiel beweist.

Beispiel 1: $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = \begin{cases} 0 & \text{falls } (x, y) = (0, 0) \text{ oder } y \neq x^2 \\ 1 & \text{falls } y = x^2, x \neq 0. \end{cases}$$

Offensichtlich gilt $f(0,0) = 0$ und für jede Richtung $\mathbf{a}$ ist

$$\mathbf{a} - \lim_{(x,y)\to(0,0)} f(x,y) = 0.$$

Es gilt jedoch für die Folge $\mathbf{x}^k = (\frac{1}{k}, \frac{1}{k^2})$

$$\lim_{k\to\infty} f(\frac{1}{k}, \frac{1}{k^2}) = 1.$$

Das heißt die Funktion ist zwar im Punkt $(0,0)$ „stetig in jede Richtung“ aber nicht stetig im Sinne der Definition 10.6.

Ein weiteres Beispiel 2

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{falls } (x,y) \neq (0,0) \\ 0 & \text{falls } (x,y) = (0,0). \end{cases}$$

Man zeige: Diese Funktion ist nicht stetig in $(0,0)$.

10.1.3. Richtungsableitung und partielle Ableitungen. Stellt man sich eine Abbildung $\mathbb{R}^2 \to \mathbb{R}$ als eine Fläche im $\mathbb{R}^3$ dar, so sieht man, daß man an einem Punkt verschiedenste Tangenten zeichnen kann. Gibt man jedoch eine Richtung in der $x-y$-Ebene vor so ist die Tangente jedoch eindeutig bestimmt.

Definition 10.7. *Eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}$ heißt im Punkt $\mathbf{x}^0$ differenzierbar in Richtung $\mathbf{a}$ ($||\mathbf{a}|| = 1$), wenn der Grenzwert*

$$\lim_{h\to 0} \frac{f(\mathbf{x}^0 + h\mathbf{a}) - f(\mathbf{x}^0)}{h} \tag{10.11}$$

existiert. Man nennt diesen Grenzwert die Richtungsableitung *von f in Richtung von $\mathbf{a}$ und schreibt*

$$\frac{\partial f}{\partial \mathbf{a}}(\mathbf{x}^0). \tag{10.12}$$

Nimmt man als spezielle Richtung die Richtung der Koordinatenachse x_k, so nennt man diese Richtungsableitung die partielle Ableitung *der Funktion f nach der Variablen x_k und schreibt*

$$\frac{\partial f}{\partial x_k}(\mathbf{x}^0) \text{ oder auch kürzer } f_{x_k}(\mathbf{x}^0). \tag{10.13}$$

Berechnet wird die partielle Ableitung nach x_k dadurch, dass man f als eine Funktion der Variablen x_k auffasst in der alle anderen Variablen konstant gehalten werden und dann wie eine eindimensionale Funktion differenziert.

Beispiel 1: $f(x,y) = x^2 + x\cos y$. Für einen Punkt (x^0, y^0) berechnet man die partiellen Ableitungen wie folgt

$$\frac{\partial f}{\partial x}(x^0,y^0) = 2\,x^0 + \cos y^0, \qquad \frac{\partial f}{\partial y}(x^0,y^0) = -x^0\,\sin y^0.$$

Ein weiteres Beispiel 2

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{falls } (x,y) \neq (0,0) \\ 0 & \text{falls } (x,y) = (0,0). \end{cases}$$

Man zeige: Die partiellen Ableitungen im Punkt $(0,0)$ sind 0.

Das letzte Beispiel zeigt, daß aus der Existenz der partiellen Ableitungen nicht die Stetigkeit der Funktion folgen muß.

Es gilt jedoch

Satz 10.8. *Gegeben sei $f : \mathbb{R}^n \to \mathbb{R}$ und ein Punkt $\mathbf{x}^0$. Existieren in einer Umgebung des Punktes $\mathbf{x}^0$ alle partiellen Ableitungen $f_{x_1}, \ldots, f_{x_n}$ und sind diese dort beschränkt, so ist f stetig an $\mathbf{x}^0$.*

Existieren an einem Punkt $\mathbf{x}^0$ alle partiellen Ableitungen, so fasst man diese zu einem Vektor zusammen.

Definition 10.9. *Es sei $f : \mathbb{R}^n \to \mathbb{R}$. Der Vektor*

$$\operatorname{grad} f(\mathbf{x}^0) = \begin{pmatrix} f_{x_1}(\mathbf{x}^0) \\ \vdots \\ f_{x_n}(\mathbf{x}^0) \end{pmatrix} \tag{10.14}$$

heißt Gradient von f an der Stelle $\mathbf{x}^0$. Eine weitere Schreibweise dafür ist $\nabla f(\mathbf{x}^0)$ (sprich: Nabla-Operator).

Der Zusammenhang zwischen Gradient und Richtungsableitung ist durch Gleichung (10.19) gegeben.

10.1.4. Differenzierbarkeit im $\mathbb{R}^n$. Bei der gewöhnlichen Differenzialrechnung wird der Differenzialquotient als Grenzwert des Differenzenquotienten definiert. Dieser ergibt die Steigung der Tangente im Punkt x_0 an eine Funktion f. Das Wesentliche bei der Definition der Differenzierbarkeit ist, dass sich die Funktion durch eine lineare Funktion lokal (affin) approximieren läßt. Dies ist der Schlüssel zur Definition in mehr Dimensionen.

Definition 10.10. *Gegeben sei eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ und ein Punkt $\mathbf{x}^0$. Die Funktion f heißt nun* differenzierbar *an $\mathbf{x}^0$, wenn ein Vektor $\mathbf{c}$ und in einer Umgebung U von $\mathbf{x}^0$ eine Funktion r existiert, sodass gilt*

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \mathbf{c} \cdot (\mathbf{x} - \mathbf{x}^0) + ||\mathbf{x} - \mathbf{x}^0||\, r(\mathbf{x}), \qquad \mathbf{x} \in U(\mathbf{x}^0),$$
$$\text{wobei } r(\mathbf{x}^0) = \lim_{\mathbf{x} \to \mathbf{x}^0} r(\mathbf{x}) = 0. \tag{10.15}$$

Anmerkung: $\mathbf{c} \cdot (\mathbf{x} - \mathbf{x}^0)$ ist das Produkt der Matrix bestehend aus dem Zeilenvektor $\mathbf{c}$ mit dem Vektor $\mathbf{x} - \mathbf{x}^0$ (Alternative: skalares Produkt).
Für $n = 1$ ist dies der Zerlegungssatz, der ja genau besagt, dass man eine Funktion in einer Variablen affin linear durch eine Gerade approximieren kann.

Auch für $n = 2$ kann man sich die Aussage dieses Satzes wie folgt veranschaulichen. Die affin lineare Funktion

$$g(x, y) = f(\mathbf{x}^0) + \mathbf{c} \cdot (\mathbf{x} - \mathbf{x}^0) = f(x^0, y^0) + c_1 (x - x^0) + c_2 (y - y^0) \tag{10.16}$$

ist genau die Tangentialebene an die Funktion f im Punkt (x^0, y^0), d.h. der Quotient

$$\lim_{(x,y) \to (x^0, y^0)} \frac{f(x, y) - g(x, y)}{||(x, y) - (x^0, y^0)||} = 0. \tag{10.17}$$

Satz 10.11. *Ist eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ in einem Punkt $\mathbf{x}^0$ differenzierbar, so ist sie auch stetig und es existieren die partiellen Ableitungen wobei gilt*

$$f_{x_k}(\mathbf{x}^0) = c_k \quad \textit{also} \quad \mathbf{c} = \operatorname{grad} f(\mathbf{x}^0). \tag{10.18}$$

Die Richtungsableitung kann in diesem Fall wie folgt berechnet werden.

$$\frac{\partial f}{\partial \mathbf{a}}(\mathbf{x}^0) = \langle \mathbf{a}, \operatorname{grad} f(\mathbf{x}^0) \rangle. \tag{10.19}$$

(Anmerkung: rechts steht ein skalares Produkt und $||\mathbf{a}||_2 = 1$)

Man sieht, dass die Richtungsableitung maximal ist, wenn $\mathbf{a}$ und grad f dieselbe Richtung haben, d.h. der Gradient hat die Richtung in der sich die Funktion am **stärksten** ändert. Die Orientierung ist in „Richtung“ der Zunahme.

Beispiel: Es $f(x, y) = x^2 + y^2$. Man berechne den Gradienten und die Niveaulinien der Funktion und stelle diese graphisch dar.

Aus der Existenz der partiellen Ableitungen folgt aber nicht die Differenzierbarkeit einer Funktion. Dazu benötigt man eine weitere Bedingung.

Satz 10.12. *Sei eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ in einer Umgebung eines Punktes $\mathbf{x}^0$ gegeben. Existieren in dieser Umgebung die partiellen Ableitungen und sind diese auch stetig, so ist f differenzierbar an der Stelle $\mathbf{x}^0$.*

In der Praxis wird man also die Differenzierbarkeit einer Funktion dadurch nachweisen, dass man die partiellen Ableitungen auf Existenz und Stetigkeit untersucht.

Für (vektorwertige) Funktionen $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ gilt.

Definition 10.13. *Gegeben sei eine Funktion $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ und ein Punkt $\mathbf{x}^0$. Die Funktion $\mathbf{f} = (f_1, \ldots, f_m)$ heißt nun* differenzierbar *(total differenzierbar) an $\mathbf{x}^0$, wenn eine $m \times n-$Matrix A und in einer Umgebung U von $\mathbf{x}^0$ eine Funktion $\mathbf{r} : \mathbb{R}^n \to \mathbb{R}^m$ existieren, sodass gilt*

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}^0) + A \cdot (\mathbf{x} - \mathbf{x}^0) + ||\mathbf{x} - \mathbf{x}^0||\, \mathbf{r}(\mathbf{x}), \qquad \mathbf{x} \in U(\mathbf{x}^0),$$
$$\text{wobei } \mathbf{r}(\mathbf{x}^0) = \lim_{\mathbf{x} \to \mathbf{x}^0} \mathbf{r}(\mathbf{x}) = 0. \tag{10.20}$$

Diese Definition besagt also, dass jede Koordinatenfunktion $f_j(\mathbf{x})$ an $\mathbf{x}^0$ differenzierbar ist. Für die Koeffizienten der Matrix A ergibt sich

$$a_{jk} = \frac{\partial f_j}{\partial x_k}(\mathbf{x}^0), j = 1, \ldots, m \; k = 1 \ldots, n. \tag{10.21}$$

Diese Matrix heißt *Jacobi-Matrix*, genauer

Definition 10.14. *Gegeben sei eine Funktion $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ und ein Punkt $\mathbf{x}^0$. Existieren an $\mathbf{x}^0$ die partiellen Ableitungen $\frac{\partial f_j}{\partial x_k}(\mathbf{x}^0), j = 1, \ldots, m \; k = 1 \ldots, n$, so heißt die Matrix*

$$\frac{d\mathbf{f}}{d\mathbf{x}}(\mathbf{x}^0) = J_{\mathbf{f}}(\mathbf{x}^0) = \frac{\partial f_1, \ldots, \partial f_m}{\partial x_1, \ldots, \partial x_n}(\mathbf{x}^0) = \left(\frac{\partial f_j}{\partial x_k}(\mathbf{x}^0)\right) \tag{10.22}$$

die **Jacobi-Matrix** (Funktionalmatrix) *von $\mathbf{f}$ an der Stelle $\mathbf{x}^0$.*

Beispiele

(1) Die Abbildung $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ sei definiert durch

$$f_1(x, y) = 2x + y,$$
$$f_2(x, y) = 3x^2 + y^2,$$
$$f_3(x, y) = xy.$$

Man berechne die Jacobi-Matrix im Punkt $(1, -2)$.

(2) Die Abbildung $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ sei definiert durch

$$f_1(x, y) = x \cos y,$$
$$f_2(x, y) = x \sin y.$$

Man berechne die Jacobi-Matrix im Punkt (x, y).

10.1.5. Die Kettenregel. In vielen Anwendungen benötigt man die Kettenregel.

Definition 10.15. *Gegeben seien die Funktionen* $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ *und* $\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^l$. *Es sei* $B(\mathbf{f}) \subset D(\mathbf{g})$ *und* $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ *und* $\mathbf{y}^0 = \mathbf{f}(\mathbf{x}^0)$. *Die Funktion* g *sei differenzierbar an der Stelle* $\mathbf{y}^0$. *Dann gilt*

(1) Existiert $\frac{d\mathbf{f}}{d\mathbf{x}}(\mathbf{x}^0)$, *so existiert auch* $\frac{d\mathbf{h}}{d\mathbf{x}}(\mathbf{x}^0)$, *und es ist*

$$\frac{d\mathbf{h}}{d\mathbf{x}}(\mathbf{x}^0) = \frac{d\mathbf{g}}{d\mathbf{y}}(\mathbf{y}^0) \cdot \frac{d\mathbf{f}}{d\mathbf{x}}(\mathbf{x}^0). \tag{10.23}$$

(" · " deutet das Matrizenprodukt an)

(2) Ist $\mathbf{f}$ *an* $\mathbf{x}^0$ *differenzierbar, so ist auch* $\mathbf{h}$ *an* $\mathbf{x}^0$ *differenzierbar.*

Übung: Man schreibe (10.23) explizit an.

Für den Fall $\mathbf{f} : \mathbb{R} \to \mathbb{R}^m$ und $g : \mathbb{R}^m \to \mathbb{R}$ ist $h : \mathbb{R} \to \mathbb{R}$ und die Kettenregel lautet dann

$$\begin{aligned} \frac{dh}{dx}(x^0) = h'(x^0) &= \left(\frac{\partial g}{\partial y_1}(\mathbf{y}^0) \ \dots \ \frac{\partial g}{\partial y_m}(\mathbf{y}^0) \right) \cdot \begin{pmatrix} \frac{df_1}{dx}(x^0) \\ \vdots \\ \frac{df_m}{dx}(x^0) \end{pmatrix} \\ &= \sum_{j=1}^{m} \frac{\partial g}{\partial y_j}(\mathbf{y}^0) \cdot \frac{df_j}{dx}(x^0) \end{aligned} \tag{10.24}$$

Eine weitere Anwendung sind Koordinatentransformation.

Gegeben sei eine Funktion $f(x, y)$. Durch Einführung der sogenannten Polarkoordinaten

$$\begin{aligned} x &= r \cos \varphi, \\ y &= r \sin \varphi \end{aligned} \tag{10.25}$$

geht f in eine Funktion $F(r, \varphi) = f(r \cos \varphi, r \sin \varphi)$ über. Zwischen $\frac{\partial F}{\partial r}$, $\frac{\partial F}{\partial \varphi}$ und $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial x}$ besteht dann nach der Kettenregel folgender Zusammenhang:

$$\begin{aligned} \frac{\partial F}{\partial r} &= \frac{\partial f}{\partial x}\frac{\partial}{\partial r}(r \cos \varphi) + \frac{\partial f}{\partial y}\frac{\partial}{\partial r}(r \sin \varphi) = \frac{\partial f}{\partial x} \cos \varphi + \frac{\partial f}{\partial y} \sin \varphi, \\ \frac{\partial F}{\partial \varphi} &= \frac{\partial f}{\partial x}\frac{\partial}{\partial \varphi}(r \cos \varphi) + \frac{\partial f}{\partial y}\frac{\partial}{\partial \varphi}(r \sin \varphi) = -\frac{\partial f}{\partial x} r \sin \varphi + \frac{\partial f}{\partial y} r \cos \varphi. \end{aligned} \tag{10.26}$$

10.1.6. Partielle Ableitungen höherer Ordnung. Betrachtet man die partielle Ableitung f_{x_k} einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ an jeder Stelle $\mathbf{x}^0$ an der sie definiert ist, so ergibt dies wiederum eine neue Funktion von der man weitere partielle Ableitungen bilden kann.

Definition 10.16. *Sei eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ in einer Umgebung eines Punktes $\mathbf{x}^0$ gegeben. Existiert in diesem Punkt für ein $k_\ell, \ell > 1$ die partielle Ableitung*

$$\left(f_{x_{k_1} x_{k_2} \dots x_{k_{\ell-1}}}\right)_{x_{k_\ell}}(\mathbf{x}^0), \qquad 1 \le k_j \le n,\ j = 1, \dots \ell, \tag{10.27}$$

so heißt diese die partielle Ableitung ℓ-ter Ordnung von f an $\mathbf{x}^0$ und wird mit

$$f_{x_{k_1} x_{k_2} \dots x_{k_\ell}}(\mathbf{x}^0) \quad \textit{oder} \quad \frac{\partial^\ell f}{\partial x_{k_1} \partial x_{k_2} \dots \partial x_{k_\ell}}(\mathbf{x}^0)$$

bezeichnet. Existieren auf einer Menge X alle partiellen Ableitungen von f der Ordnung k und sind diese auch stetig so schreibt man

$$f \in C^k(X). \tag{10.28}$$

Also gilt: $f_{x_{k_1} x_{k_2}}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_{k_1} \partial x_{k_2}}(\mathbf{x}) := \frac{\partial}{\partial x_{k_2}} \left(\frac{\partial f}{\partial x_{k_1}} \right)(\mathbf{x})$.

Beispiel 1: $f(x, y, z) = 4xyz - x^2 + y^2$

Man erhält

$$\begin{aligned} &f_x = 4yz - 2x, \qquad f_y = 4xz + 2y, \qquad f_z = 4xy, \\ &f_{xx} = -2 \qquad f_{yy} = 2, \qquad f_{zz} = 0, \\ &f_{xy} = f_{yx} = 4z, \qquad f_{xz} = f_{zx} = 4y, \qquad f_{yz} = f_{zy} = 4x. \end{aligned}$$

Beispiel 2:

$$f(x, y) = \begin{cases} x\, y\, \dfrac{x^2 - y^2}{x^2 + y^2} & \text{falls } (x, y) \neq (0, 0) \\ 0 & \text{falls } (x, y) = (0, 0). \end{cases}$$

Man berechne $f_{xy}(0, 0)$ und $f_{yx}(0, 0)$.

Wie man an Beispiel 2 sieht, muss die Reihenfolge bei den partiellen Ableitungen nicht unbedingt vertauschbar sein. Der folgende Satz für Funktionen von zwei Variablen lässt sich unmittelbar auf Abbildungen $f : \mathbb{R}^n \to \mathbb{R}$ übertragen.

Satz 10.17. *(Schwarz) Sei eine Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ in einer Umgebung U eines Punktes (x^0, y^0) gegeben. Es existiere in dieser Umgebung die partielle Ableitung f_{xy} und sei diese auch stetig in (x^0, y^0). Ferner existiere auch $f_y(x, y^0)$ für alle $(x, y^0) \in U$. Dann existiert auch $f_{yx}(x^0, y^0)$ und es gilt*

$$f_{xy}(x^0, y^0) = f_{yx}(x^0, y^0). \tag{10.29}$$

Diese Voraussetzungen sind insbesondere erfüllt, wenn $f \in C^2(U)$ gilt.

10.1.7. Extremwerte von Funktionen mehrerer Veränderlicher. Eine notwendige Bedingung für das Vorliegen eines relativen Extremwertes ist offensichtlich, daß ein Extremum in jeder Koordinatenrichtung vorliegt.

Satz 10.18. *Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ sei an $\mathbf{x}^0$ partiell nach jeder Variablen differenzierbar. Hat f an $\mathbf{x}^0$ ein relatives Extremum, so gilt* $\operatorname{grad} f(\mathbf{x}^0) = 0$.

Zur Formulierung hinreichender Bedingungen benötigen wir zuerst ein paar neue Begriffe aus der linearen Algebra.

Definition 10.19. *Es sei $A = (a_{jk})$ eine symmetrische Matrix.*

Das Polynom

$$Q_A(\mathbf{x}) = \sum_{j,k=1}^{n} a_{jk} x_j x_k = \langle \mathbf{x}, A\,\mathbf{x} \rangle \tag{10.30}$$

heißt die zu A gehörende quadratische Form*.*

Die Matrix A bzw. die Funktion Q_A heißt

(i) positiv (negativ) semidefinit*, wenn $Q_A(\mathbf{x}) \geq 0$ ($Q_A(\mathbf{x}) \leq 0$) für alle $\mathbf{x}$ gilt,*

(ii) positiv (negativ) definit*, wenn $Q_A(\mathbf{x}) > 0$ ($Q_A(\mathbf{x}) < 0$) für alle $\mathbf{x}$ gilt,*

(iii) ansonsten indefinit*.*

Für eine symmetrische 2×2 Matrix ergibt sich daraus: die Matrix

$$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \tag{10.31}$$

ist genau dann:

(i) positiv semidefinit, wenn $ad - b^2 \geq 0$, $a \geq 0, d \geq 0$,

(ii) negativ semidefinit, wenn $ad - b^2 \geq 0$, $a \leq 0, d \leq 0$,

(iii) positiv definit, wenn $ad - b^2 > 0$, $a > 0$,

(iv) negativ definit, wenn $ad - b^2 > 0$, $a < 0$,

(v) indefinit, wenn $ad - b^2 < 0$.

Ein Kriterum wann eine symmetrische Matrix definit ist gibt der folgende Satz.

Satz 10.20. *Gegeben sei die symmetrische Matrix $A = (a_{jk})_{j,k=1}^n$. Mit A_l werden die Hauptuntermatrizen der Ordnung l bezeichnet, d.h. $A_l = (a_{jk})_{j,k=1}^l$.*

(i) A ist genau dann positiv definit, wenn

$\det A_l > 0$ *für alle $l = 1, \ldots, n$.*

(ii) A ist genau dann negativ definit, wenn

$(-1)^l \det A_l > 0$ *für alle $l = 1, \ldots, n$.*

Weiters gilt:

Satz 10.21. *Für ein $\epsilon > 0$ sei $U'_\epsilon = \{\mathbf{x} \in \mathbb{R}^n \mid 0 < ||\mathbf{x}|| < \epsilon\}$ und $S_\epsilon = \{\mathbf{x} \in \mathbb{R}^n \mid ||\mathbf{x}|| = \epsilon\}$. Gilt auf U'_ϵ oder S_ϵ*

(i) $Q_A(\mathbf{x}) \geq 0$, so ist A positiv semidefinit,

(ii) $Q_A(\mathbf{x}) \leq 0$, so ist A negativ semidefinit,

(iii) $Q_A(\mathbf{x}) > 0$, so ist A positiv definit,

(iv) $Q_A(\mathbf{x}) < 0$, so ist A negativ definit.

Damit kann man nun Extremalwerte wie folgt charakterisieren.

Satz 10.22. *Gegeben sei die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ und ein Punkt $\mathbf{x}^0$ und in einer Umgebung U von $\mathbf{x}^0$ gelte $f \in C^2(U(\mathbf{x}^0))$. Ferner sei* $\operatorname{grad} f(\mathbf{x}^0) = 0$ *und $H(\mathbf{x}^0)$ die (symmetrische)* **Hesse-Matrix**

$$H(\mathbf{x}^0) = \left(f_{x_j,x_k}(\mathbf{x}^0)\right)_{j,k=1}^n \tag{10.32}$$

Es gilt:

(i) Ist $H(\mathbf{x}^0)$ positiv definit, so hat f an $\mathbf{x}^0$ ein relatives Minimum. Ist $H(\mathbf{x}^0)$ negativ definit, so hat f an $\mathbf{x}^0$ ein relatives Maximum.

(ii) Hat f an $\mathbf{x}^0$ ein relatives Minimum, so ist $H(\mathbf{x}^0)$ positiv semidefinit. Hat f an $\mathbf{x}^0$ ein relatives Maximum, so ist $H(\mathbf{x}^0)$ negativ semidefinit.

(iii) Ist $H(\mathbf{x}^0)$ indefinit, so hat f an $\mathbf{x}^0$ kein relatives Extremum.

Beispiel: Gegeben sei die Funktion $f : \mathbb{R}^3 \to \mathbb{R}$ mit

$$f(x, y, z) = 35 - 6x + 2z + x^2 - 2xy + 2y^2 + 2yz + 3z^2.$$

Man suche die Extremstellen.

Für den Fall $f : \mathbb{R}^2 \to \mathbb{R}$ lautet obiger Satz. Es sei (x^0, y^0) ein Punkt mit einer Umgebung U in der $f \in C^2(U(x^0, y^0))$ gelte. Ferner sei $\operatorname{grad} f(x^0, y^0) = (0, 0)$. Die Diskriminate Δ (ist gleich $\det H(\mathbf{x}^0)$) ist gegeben durch

$$\Delta(x^0, y^0) = f_{xx}(x^0, y^0) f_{yy}(x^0, y^0) - (f_{xy}(x^0, y^0))^2. \tag{10.33}$$

Es gilt:

(i) Ist $\Delta(x^0, y^0) > 0$, so hat f an (x^0, y^0) eine Extremalstelle, wobei es sich um ein relatives Minimum für $f_{xx}(x^0, y^0) > 0$ und ein relatives Maximum für $f_{xx}(x^0, y^0) < 0$ handelt.

(ii) Ist $\Delta(x^0, y^0) < 0$, so hat f an (x^0, y^0) kein relatives Extremum.

Beispiel 1: Man untersuche die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = xy + x - y + 1.$$

Beispiel 2: Man untersuche die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = x^3 - 12xy + 8y^3.$$

Beispiel 3: Man untersuche die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = 3xy - x^2y - xy^2$$

Man bestimme alle Extremstellen auf $M = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0, 0 \leq y \leq 3-x\}$.

Zum Schluss eine praktische Anwendung:

Beispiel 10.1. Ausgleichsgerade *Gegeben sind n Messwerte $(a_i, b_i), i = 1, \ldots, n$. Gesucht ist eine Gerade $g(x) = kx + d$, sodass diese Messwerte möglichst „gut“ beschrieben werden, d. h. die quadratische Abweichung*

$$\sum_{i=1}^{n} (b_i - g(a_i))^2$$

minimal wird (Gauß'sche Methode der kleinsten Fehlerquadrate).

Lösung zu 10.1. *Wir setzen $(x_1, x_2) = (k, d)$, dann müssen wir das Minimum der Funktion*

$$f(\mathbf{x}) = \sum_{i=1}^{n} (b_i - x_1 a_i - x_2)^2$$

finden. Zunächst müssen wir den Gradienten berechnen

$$\operatorname{grad}(f) = \begin{pmatrix} \sum_{j=1}^{n} 2(b_j - x_1 a_j - x_2)(-a_j) \\ \sum_{j=1}^{n} 2(b_j - x_1 a_j - x_2)(-1) \end{pmatrix}$$

und dessen Nullstellen suchen. Führen wir folgende Abkürzungen ein

$$\bar{a} = \frac{1}{n} \sum_{j=1}^{n} a_j, \quad \bar{b} = \frac{1}{n} \sum_{j=1}^{n} b_j,$$

$$A = \frac{1}{n} \sum_{j=1}^{n} a_j^2, \quad B = \frac{1}{n} \sum_{j=1}^{n} a_j b_j,$$

so lässt sich unser Gradient übersichtlich als

$$\operatorname{grad}(f) = 2n \begin{pmatrix} Ax_1 + \bar{a}x_2 - B \\ \bar{a}x_1 + x_2 - \bar{b} \end{pmatrix}$$

schreiben. Das lineare Gleichungssystem $\operatorname{grad}(f) = 0$ kann leicht gelöst werden und die Lösung lautet

$$x_1 = \frac{B - \bar{a}\bar{b}}{A - \bar{a}^2}, \quad x_2 = \frac{\bar{b}A - \bar{a}B}{A - \bar{a}^2}.$$

Um zu zeigen, dass es sich um ein Minimum handelt, müssen wir noch zeigen, dass die Hesse Matrix positiv definit ist. Die Hesse-Matrix lautet

$$\frac{\partial^2 f}{\partial \mathbf{x}^2} = 2n \begin{pmatrix} A & \bar{a} \\ \bar{a} & 1 \end{pmatrix}.$$

Da $A > 0$ bleibt zu zeigen, dass die Determinate (Produkt der Eigenwerte) der Hesse-Matrix positiv ist.

Die beiden Eigenwerte $\lambda_{1,2} = n(A + 1 \pm \sqrt{(A+1)^2 - 4(A - \bar{a}^2)})$ sind positiv, da $\lambda_1 + \lambda_2 = 2n(A+1) > 0$ und $\lambda_1 \lambda_2 = 4n^2(A - \bar{a}^2) = 4n \sum_{i=1}^n (a_j - \bar{a})^2 > 0$.
■

Anmerkung: In der Statistik nennt man dies *lineare Regression* und verwendet die folgende Notation bzw. Abkürzungen

$$s_a^2 = \frac{1}{n-1} \sum_{j=1}^n (a_j - \bar{a})^2, \quad s_{ab} = \frac{1}{n-1} \sum_{j=1}^n (a_j - \bar{a})(b_j - \bar{b})$$

und wegen

$$\sum_{j=1}^n a_j^2 = (n-1)s_a^2 + n\bar{a}^2, \qquad \sum_{j=1}^n a_j b_j = (n-1)s_{ab} + n\bar{a}\bar{b}$$

ist die Lösung

$$x_1 = \frac{s_{ab}}{s_a^2}, \qquad x_2 = \bar{b} - \bar{a}\frac{s_{ab}}{s_a^2}.$$

10.1.7.1. *Extremwerte unter Nebenbedingungen, Lagrange-Multiplikatoren.* In vielen Problemen werden nicht nur Extremstellen von Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ allein gesucht, sondern man verlangt, dass auch gewisse zusätzliche Gleichungen (Nebenbedingungen) erfüllt sind. Die Nebenbedingungen treten in Gleichungen der Form

$$\begin{aligned} g_1(x_1, \ldots, x_n) &= 0, \\ &\vdots \\ g_m(x_n, \ldots, x_n) &= 0, \end{aligned}$$

kurz

$$\mathbf{g}(\mathbf{x}) = 0, \qquad \mathbf{g} = (g_1, \ldots, g_m) \tag{10.34}$$

auf. Für Extremstellen mit Nebenbedingungen gibt es ein notwendiges (aber nicht hinreichendes) Kriterium, das sogenannte Verfahren von Lagrange.

Satz 10.23. *Gegeben seien die Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ und $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$ mit $m < n$. In einer Umgebung U von $\mathbf{x}^0$ sei $f \in C^1(U)$ und $g_j \in C^1(U)$, $j = 1, \dots m$. Der Rang der Matrix*

$$\left(\frac{\partial g_j}{\partial x_k}(\mathbf{x}^0)\right)_{j=1,\dots,m,k=1,\dots,n} \tag{10.35}$$

sei gleich m. Die Funktion f habe an x^0 ein lokales Extremum unter den Nebenbedingungen $\mathbf{g}(\mathbf{x}) = 0$. Dann gibt es Konstanten $\lambda_1, \dots, \lambda_m$ mit

$$\operatorname{grad} f(\mathbf{x}^0) = \lambda_1 \operatorname{grad} g_1(\mathbf{x}^0) + \dots + \lambda_m \operatorname{grad} g_m(\mathbf{x}^0). \tag{10.36}$$

Diese Konstanten λ_j heißen Lagrange-Multiplikatoren.

Beispiel: Gesucht sind die Extremwerte der Funktion $f(x, y, z) = x+y+z$ unter den Bedingungen

$$g_1(x, y, z) = x^2 + y^2 - 2 = 0,$$
$$g_2(x, y, z) = x + z - 1 = 0.$$

10.1.7.2. *Ein weiteres Kriterium für Extremwerte unter Nebenbedingungen liefern die Karush-Kuhn-Tucker Bedingungen.* [1]

10.1.8. Taylorschen Satz. Wir wollen nun den Taylorschen Satz für Funktionen mehrerer Veränderlicher (Variablen) verallgemeinern. Dazu führen wir zuerst ein paar Notationen ein um den Satz dann in übersichtlicher Form schreiben zu können.

Definition 10.24. *Ein Polynom von n Veränderlichen ist eine Funktion $p : \mathbb{R}^n \to \mathbb{R}$ mit*

$$p(x_1, \dots, x_n) = \sum_{k_1,k_2,\dots,k_n=0}^{N} a_{k_1 k_2 \dots k_n} x_1^{k_1} \dots x_n^{k_n} \tag{10.37}$$

wobei $k_1, k_2, \dots, k_n, N \in \mathbb{N}$ und $a_{k_1 k_2 \dots k_n} \in \mathbb{R}$ gilt. Die grösste der Zahlen $k_1 + k_2 + \dots + k_n$ mit $a_{k_1 k_2 \dots k_n} \neq 0$ heißt der Grad von p.

Definition 10.25. *Gegeben sei der konstante Vektor $\mathbf{h} = (h_1, \dots, h_n)$ und ein $k \in \mathbb{N}$. Die Funktion $f : \mathbb{R}^n \to \mathbb{R}$ sei aus C^k. Dann ist der Differentialoperator $(\mathbf{h} \cdot \operatorname{grad} f)^k, k > 0$ definiert durch*

$$(\mathbf{h} \cdot \operatorname{grad} f)^k(\mathbf{x}) := \sum_{j_1,j_2,\dots,j_k=1}^{n} h_{j_1} h_{j_2} \cdots h_{j_k} \frac{\partial^k f}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_k}}(\mathbf{x}). \tag{10.38}$$

[1] Dazu siehe: `http://en.wikipedia.org/wiki/Kuhn_Tucker`

Für $k = 0$ definiert man $(\mathbf{h} \cdot \operatorname{grad} f)^0(x) = f(\mathbf{x})$.

Übung: Setze $n = 2, k = 1, 2$ und werte obigen Ausdruck aus.

Damit läßt sich nun der Taylorschen Satz für Funktionen mehrerer Veränderlicher wie folgt formulieren

Satz 10.26. *Gegeben sei eine Funktion $f : D \to \mathbb{R},\ D \subset \mathbb{R}^n$ und für ein $m \in \mathbb{N}$ gelte $f \in C^{m+1}(D)$. Ferner sei die Verbindungstrecke $\overline{\mathbf{x}\mathbf{x}^0}$ der Punkte $\mathbf{x}, \mathbf{x}^0$ in D enthalten. Dann gilt mit $\mathbf{h} = \mathbf{x} - \mathbf{x}^0$*

$$f(\mathbf{x}) = \sum_{k=0}^{m} \frac{1}{k!} (\mathbf{h} \cdot \operatorname{grad} f)^k(\mathbf{x}^0) + R_m(\mathbf{x}, \mathbf{x}^0). \tag{10.39}$$

Für das Restglied $R_m(\mathbf{x}, \mathbf{x}^0)$ gilt mit einem $\vartheta \in (0, 1)$

$$R_m(\mathbf{x}, \mathbf{x}^0) = \frac{1}{(m+1)!} (\mathbf{h} \cdot \operatorname{grad} f)^{m+1}(\mathbf{x}^0 + \vartheta \mathbf{h}). \tag{10.40}$$

Für $m = 0$ (also $f \in C^1(X)$), erhält man für ein geeignetes $\vartheta \in (0, 1)$ den Mittelwertsatz

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \mathbf{h} \cdot \operatorname{grad} f(\mathbf{x}^0 + \vartheta \mathbf{h}). \tag{10.41}$$

Für eine Funktion $f(x, y)$ von zwei Variablen lautet die Taylorsche Formel ausgeschrieben für $m = 1$

$$\begin{aligned} f(x^0 + h, y^0 + k) &= \frac{1}{0!} f(x^0, y^0) + \frac{1}{1!} \left(h \frac{\partial f}{\partial x} + k \frac{\partial f}{\partial y} \right)^1 (x^0, y^0) \\ &\quad + \frac{1}{2!} \left(h \frac{\partial f}{\partial x} + k \frac{\partial f}{\partial y} \right)^2 (x^0 + \vartheta h, y^0 + \vartheta k) \\ &= f(x^0, y^0) + h f_x(x^0, y^0) + k f_y(x^0, y^0) \\ &\quad + \frac{1}{2} \left(h^2 f_{xx}(x^0 + \vartheta h, y^0 + \vartheta k) + 2hk f_{xy}(x^0 + \vartheta h, y^0 + \vartheta k) + k^2 f_{yy}(x^0 + \vartheta h, y^0 + \vartheta k) \right). \end{aligned} \tag{10.42}$$

Beispiele:
Man entwickle folgende Funktionen nach dem Taylorschen Satz um den Punkt $(1, 2)$

(i) $f(x, y) = x^3 + x\,y^2 + y^3$,

(ii) $f(x, y) = x^2 + x\,y + y^2 + 1$.

10.1.9. Implizite Funktionen. Betrachtet man eine Gleichung mit den unabhängigen und gleichberechtigten Variablen x und y der Form

$$F(x,y)=0 \tag{10.43}$$

so sind diese Variablen nicht mehr unabhängig voneinander. Es wird aber in der Regel keine *globale* Funktion $f:\mathbb{R}\to\mathbb{R}$ geben

$$y=f(x), \tag{10.44}$$

sodass $F(x,f(x))=0$ gilt.

Mithilfe der Kettenregel kann man "lokal" implizite Funktionen auflösen.

Man betrachte die Funktionen $F:U\to\mathbb{R}$, $U\subset\mathbb{R}^2$ und offen, $(x,y)\mapsto F(x,y)$ und $g:I\to\mathbb{R},\ I\subset\mathbb{R}$, $x\mapsto g(x)$. Und es gelte mit $(x,g(x))\subset U$

$$F(x,g(x))=0 \quad \text{für alle } x\in I \tag{10.45}$$

Differenziert man mithilfe der Kettenregel diese Gleichung nach x so erhält man

$$F_x(x,g(x))+F_y(x,g(x))\,g'(x)=0. \tag{10.46}$$

Wenn nun $F_y(x,g(x))\neq 0$ ist, ergibt sich somit

$$g'(x)=-\frac{F_x(x,g(x))}{F_y(x,g(x))}. \tag{10.47}$$

Anmerkung: Entsprechend kann nach $x(y)$ aufgelöst werden.

Beispiel 1: $F(x,y)=x^2+y^2-r^2, r\in\mathbb{R}$ und somit

$$\begin{aligned} y=g(x)&=\sqrt{r^2-x^2}, \qquad -r<x<r,\ y>0,\\ F(x,y)&=x^2+y^2-r^2=0. \end{aligned}$$

Differenzieren liefert

$$\begin{aligned} 2x+2g\,g'&=0,\\ y'=g'&=-\frac{x}{y}=\frac{-x}{\sqrt{r^2-x^2}}. \end{aligned}$$

Beispiel 2: $F(x,y)=x^3\sin y+x^2y^2+e^{x+y}+c, c\in\mathbb{R}$. Im Gegensatz zum vorigen Beispiel ist die Gleichung $F(x,y)=x^3\sin y+x^2y^2+e^{x+y}+c=0$ auch lokal nicht explizit nach y auflösbar. Trotzdem kann man ohne Probleme die Ableitung (Steigung der Tangenten) implizit berechnen. Differenzieren liefert

$$y'=g'=-\frac{3x^2\sin y+2xy^2+e^{x+y}}{x^3\cos y+2x^2y+e^{x+y}}.$$

Dieses Beispiel lässt sich verallgemeinern. Generell gilt:

Satz 10.27. *Seien $U_1 \subset \mathbb{R}^k$, $U_2 \subset \mathbb{R}^m$ offene Mengen und $\mathbf{F} : U_1 \times U_2 \to \mathbb{R}^m, (\mathbf{x}, \mathbf{y}) \mapsto \mathbf{F}(\mathbf{x}, \mathbf{y})$ eine stetig differenzierbare Abbildung. Sei $(\mathbf{x}^0, \mathbf{y}^0) \in U_1 \times U_2$ ein Punkt, sodass $\mathbf{F}(\mathbf{x}, \mathbf{y}) = 0$ und die $m \times m$-Matrix $\frac{\partial \mathbf{F}}{\partial \mathbf{y}}(\mathbf{x}^0, \mathbf{y}^0)$ invertierbar ist. Dann gibt es existiert Umgebungen $V_1 \subset U_1$ von $\mathbf{x}^0$ und $V_2 \subset U_2$ von $\mathbf{y}^0$ und eine stetige Abbildung*

$$\mathbf{g} : V_1 \to V_2$$

mit

$$\mathbf{F}(\mathbf{x}, \mathbf{g}(\mathbf{x})) = 0 \qquad \text{für alle } x \in V_1. \tag{10.48}$$

Ist $(\mathbf{x}, \mathbf{y}) \in V_1 \times V_2$ ein Punkt mit $\mathbf{F}(\mathbf{x}, \mathbf{y}) = 0$, so folgt $\mathbf{y} = \mathbf{g}(\mathbf{x})$.

Für die Funktionalmatrix erhält man in einer Umgebung von $\mathbf{x}^0$

$$\left(\frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\mathbf{x})\right) = -\left(\frac{\partial \mathbf{F}}{\partial \mathbf{y}}(\mathbf{x}, \mathbf{g}(\mathbf{x}))\right)^{-1} \left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}}(\mathbf{x}, \mathbf{g}(\mathbf{x}))\right). \tag{10.49}$$

Anmerkung: (1) Man sagt auch die Abbildung $\mathbf{g}$ entstehe durch Auflösung der Gleichung $\mathbf{F}(\mathbf{x}, \mathbf{y}) = 0$ nach $\mathbf{y}$.

(2) Im Fall von $k = m = 1$ sind $\frac{\partial g}{\partial x} = g'$, $\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}$ die gewöhnlichen partiellen Ableitungen.

10.1.10. Umkehrabbildungen. Der Satz über implizite Funktionen ermöglicht lokal die Umkehrung von Funktionen in einer Umgebung (die klein genug ist) eines Punktes, wenn die Funktionalmatrix dort invertierbar ist.

Satz 10.28. *Sei U eine offene Teilmenge des $\mathbb{R}^n$ und $\mathbf{f} : U \to \mathbb{R}^n$ eine stetig differenzierbare Abbildung, d.h. $\mathbf{f} \in C^1(U)$. Es sei $\mathbf{x}^0 \in U$ und $\mathbf{y}^0 = \mathbf{f}(\mathbf{x}^0)$ und die Jacobi-Matrix sei invertierbar an $\mathbf{x}^0$, d.h.*

$$\det\left(\frac{d\mathbf{f}}{d\mathbf{x}}(\mathbf{x}^0)\right) \neq 0. \tag{10.50}$$

Dann gilt:

Es existiert eine Umgebung $V \subset U$ von $\mathbf{x}^0$ und eine Umgebung V' von $\mathbf{y}^0$, sodass $\mathbf{f}$ die Menge V bijektiv auf V' abbildet und die Umkehrabbildung

$$\mathbf{g} := \mathbf{f}^{-1} : V' \to V \tag{10.51}$$

stetig differenzierbar ist. Es gilt

$$\frac{d\mathbf{g}}{d\mathbf{y}}(\mathbf{y}^0) = \left(\frac{d\mathbf{f}}{d\mathbf{x}}(\mathbf{x}^0)\right)^{-1}. \tag{10.52}$$

10.2. Übung

(1) Man zeige, dass die Funktion

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{falls } (x,y) \neq (0,0) \\ 0 & \text{falls } (x,y) = (0,0). \end{cases}$$

(a) nicht stetig im Punkt $(0,0)$ ist.
(b) Man zeige: die partiellen Ableitungen dieser Funktion im Punkt $(0,0)$ sind 0.

(2) Gegeben sei die Funktion $f(x,y) = x^2 + x\cos y$. Gesucht ist die Richtungsableitung in Richtung $\mathbf{a} = \frac{1}{\sqrt{2}}(1,1)$

(a) mit direkter Berechnung,
(b) mit Hilfe des Gradienten.

(3) Die Abbildung $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ sei definiert durch

$$\begin{aligned} f_1(x,y) &= x^3 + y^2, \\ f_2(x,y) &= x\,y + x\cos(2\pi y), \\ f_3(x,y) &= 3x^2 + 5y^2 + x^2y^2. \end{aligned}$$

Man berechne die Jacobi-Matrix im Punkt $(1,-2)$.

(4) Die Abbildung $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ sei definiert durch

$$\begin{aligned} f_1(x,y) &= x\cos y, \\ f_2(x,y) &= x\sin y. \end{aligned}$$

Man berechne die Jacobi-Matrix im Punkt (x,y).

(5)

$$f(x,y) = \begin{cases} x\,y\,\dfrac{x^2-y^2}{x^2+y^2} & \text{falls } (x,y) \neq (0,0) \\ 0 & \text{falls } (x,y) = (0,0). \end{cases}$$

Man berechne $f_{xy}(0,0)$ und $f_{yx}(0,0)$.

(6) Man untersuche auf Extremwerte
(a) Man untersuche die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x,y) = xy + x - y + 1.$$

(b) Man untersuche die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x,y) = x^3 - 12xy + 8y^3.$$

(c) Man untersuche die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = 3xy - x^2y - xy^2$$

Bei (c) bestimme man alle Extremstellen auf $M = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0, 0 \leq y \leq 3 - x\}$.

(7) (a) Man bestimme die Minima und Maxima von $f(x, y, z) = xy^2z^3$ unter der Nebenbedingung $g(x, y, z) = x + y + z = 6$
(b) Man bestimme das Maximum der Funktion

$$f(x_1, \ldots, x_n) = \prod_{j=1}^{n} x_j^2$$

unter der Nebenbedingung $g(x_1, \ldots, x_n) = \sum_{j=1}^{n} x_j^2 = 1$.

(8) Man entwickle folgende Funktionen nach dem Taylorschen Satz um den Punkt $(x^0, y^0) = (1, 2)$
(a) $f(x, y) = x^3 + x\,y^2 + y^3$,
(b) $f(x, y) = x^2 + x\,y + y^2 + 1$.

(9) Berechne die Umkehrabbildungen
(i) Man betrachte die Funktion $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, gegeben durch

$$y_1 = f_1(x_1, x_2) = x_1^2,$$
$$y_2 = f_2(x_1, x_2) = x_1 + x_2.$$

(ii) (Forster, Analysis 2, Seite 97) Man betrachte die Funktion $\mathbf{f} : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}^2$, gegeben durch

$$x = f_1(r, \varphi) = r\,\cos\varphi,$$
$$y = f_2(r, \varphi) = r\,\sin\varphi.$$

Kapitel 11

Differentialgleichungen

11.1. Grundlagen

Angenommen, $x(t)$ beschreibt die Größe einer Population (z. B. Bakterien) zur Zeit t. Im einfachsten Fall ist die Zunahme der Population proportional zur vorhandenen Population, d.h.,

$$\frac{d}{dt}x(t) = \mu\, x(t),$$

wobei $\mu > 0$ die Wachstumsrate ist. Diese Annahme ergibt also eine Gleichung, die eine unbekannte Funktion $x(t)$ mit ihrer Ableitung verknüpft.

Definition 11.1. *Eine Gleichung, die eine reelle Funktion $x(t)$ mit ihren Ableitungen verknüpft,*

$$F(x^{(n)}(t), \ldots, x(t), t) = 0,$$

(mit einer stetigen Funktion F) wird **(gewöhnliche) Differentialgleichung** *genannt. Wir werden immer davon ausgehen, dass eine Differentialgleichung nach der höchsten Ableitung aufgelöst werden kann:*

$$x^{(n)}(t) = f(x^{(n-1)}(t), \ldots, x(t), t).$$

In diesem Fall ist n die **Ordnung** *der Differentialgleichung. Hängt f nicht von t ab, also $x^{(n)}(t) = f(x^{(n-1)}(t), \ldots, x(t))$, so spricht man von einer* **autonomen** *Differentialgleichung.*

Eine Funktion $x(t)$, die die Differentialgleichung erfüllt, wird **Lösung der Differentialgleichung** *genannt. Meistens sucht man eine Lösung auf einem offenen Intervall I, die an einer Stelle $t_0 \in I$ bestimmte* **Anfangswerte** *(auch*

Anfangsbedingungen*)*

$$x^{(n-1)}(t_0) = x_{n-1}, \quad \ldots, \quad x(t_0) = x_0$$

erfüllt. Man spricht in diesem Fall von einem **Anfangswertproblem**.

Beispiel 11.1. Lösung einer Differentialgleichung
Zum Beispiel ist

$$x'(t) + 2\,t\,x(t) = 0$$

eine Differentialgleichung erster Ordnung. Die Funktion $x(t) = \exp(-t^2)$ *ist eine Lösung, da* $x'(t)+2\,t\,x(t) = -2\,t\,\exp(-t^2)+2\,t\,\exp(-t^2) = 0$. *Die Funktion* $x(t) = t^2 - 1$ *ist keine Lösung, da* $x'(t) + 2\,t\,x(t) = 2\,t + 2\,t\,(t^2 - 1) = 2\,t^3 \neq 0$ *(die Gleichung muss für alle* t *erfüllt sein!).*

Im Gegensatz zu *gewöhnlichen* Differentialgleichungen wird eine Gleichung, die eine Funktion von mehreren Variablen mit ihren partiellen Ableitungen verknüpft, eine **partielle Differentialgleichung** genannt. Ein Beispiel ist die Wellengleichung $\frac{\partial^2}{\partial x^2}u(x,t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}u(x,t) = 0$, die die Ausbreitung von elektromagnetischen Wellen, Schallwellen, etc. beschreibt. Hier bedeutet $u(x,t)$ die Auslenkung am Ort x zur Zeit t und c ist die Ausbreitungsgeschwindigkeit der Welle. Partielle Differentialgleichung sind um ein Vielfaches komplizierter als gewöhnliche, und wir verweisen auf [**94**].

Das Wachstum unserer Population wird also durch die autonome Differentialgleichung erster Ordnung $x'(t) = \mu\,x(t)$ beschrieben. Da wir wissen, dass die Ableitung der Exponentialfunktion wieder die Exponentialfunktion ergibt, ist es hier nicht schwer, eine Lösung zu erraten:

$$x(t) = \mathrm{e}^{\mu t},$$

denn $x'(t) = \mu\,\mathrm{e}^{\mu t} = \mu\,x(t)$. Wir können diese Lösung sogar noch mit einer beliebigen Konstante $C \in \mathbb{R}$ multiplizieren, $x(t) = C\mathrm{e}^{\mu t}$, und erhalten wieder eine Lösung. Setzen wir $t = 0$, so sehen wir, dass $x(0) = C$ die Anfangspopulation ist. Für eine gegebene Anfangspopulation $x(0) = x_0$ (Anfangsbedingung) ist die Lösung also

$$x(t) = x_0\,\mathrm{e}^{\mu t}.$$

Sie ist für $\mu = 1.2$ und $x(0) = 10$ in Abbildung 11.1 dargestellt.

Haben wir damit schon alle Lösungen gefunden, oder haben wir noch eine übersehen? Sei $x(t)$ irgendeine Lösung. Dann gilt $\frac{d}{dt}(x(t)\mathrm{e}^{-\mu t}) = x'(t)\mathrm{e}^{-\mu t} - \mu\,x(t)\mathrm{e}^{-\mu t} = (x'(t) - \mu\,x(t))\mathrm{e}^{-\mu t} = 0$ (da für eine Lösung $x'(t) = \mu\,x(t)$ ist). Da die Ableitung von $x(t)\mathrm{e}^{-\mu t}$ verschwindet, ist dieser Ausdruck konstant: $x(t)\mathrm{e}^{-\mu t} = x_0$, d.h., $x(t) = x_0\,\mathrm{e}^{\mu t}$. Somit ist jede Lösung von dieser Form.

Aus praktischer Sicht ist unsere Annahme, dass eine Population unbegrenzt wachsen kann, unrealistisch. Gehen wir davon aus, dass es eine maximale Grenzpopulation gibt, die wir auf $x = 1$ festsetzen (das entspricht 100%). Dann erhält

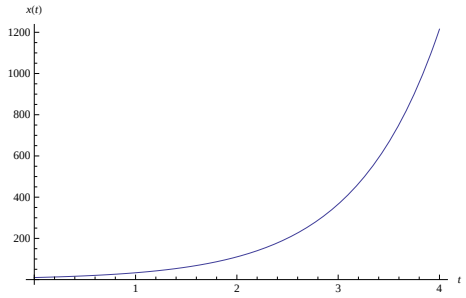


Abbildung 11.1. Exponentielles Wachstum $x'(t) = 1.2\,x(t)$ mit $x(0) = 10$.

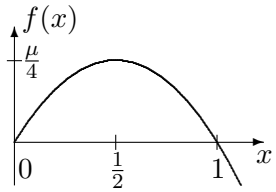


Abbildung 11.2. Wachstumsgeschwindigkeit $f(x) = \mu\,(1-x)x$ beim logistischen Wachstum.

man das **logistische Wachstumsmodell**

$$\frac{d}{dt}x(t) = \mu\,(1 - x(t))x(t).$$

Das Wachstum ist proportional zur vorhandenen Population $x(t)$ und zur verbleibenden Kapazität $1-x(t)$ (Abstand zur Grenzpopulation). Hier ist es nicht mehr ganz so leicht möglich, die Lösung zu erraten. Beginnen wir daher zunächst mit einer *qualitativen* Diskussion. Das heißt, wir versuchen, qualitative Eigenschaften der Lösung (z. B. Monotonie, Beschränktheit, langfristiges Verhalten) direkt von der Differentialgleichung abzulesen, ohne die Lösung zu kennen.

Schon bei der Integralrechnung haben wir gesehen, dass es Integrale gibt, die nicht mit den uns bekannten Funktionen gelöst werden können. Bei den Differentialgleichungen ist es noch viel schlimmer, denn eine Lösung kann nur in wenigen Spezialfällen explizit angegeben werden. Trotzdem ist es oft möglich, wichtige Eigenschaften der Lösung direkt von der Differentialgleichung abzulesen.

Dazu zeichnen wir zunächst die rechte Seite $f(x) = \mu(1-x)x$ unserer Differentialgleichung (siehe Abbildung 11.2). Es handelt sich um eine Parabel mit den beiden Nullstellen $x = 0$, $x = 1$ und dem maximalen Wert $\frac{\mu}{4}$ bei $x = \frac{1}{2}$.

Die Differentialgleichung lautet also $\frac{dx}{dt} = f(x)$ mit $f(x) = \mu\,(1-x)x$. Per Definition ist $f(x)$ somit gerade die zeitliche Änderung (Ableitung) von x. Am Vorzeichen von f für eine bestimmte Population x können wir daher ablesen, ob die Population hier monoton wächst ($f(x) > 0$) oder fällt ($f(x) < 0$).

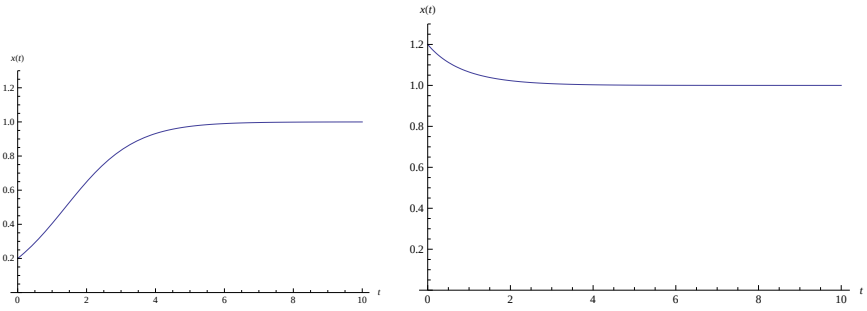


Abbildung 11.3. Logistisches Wachstum $x'(t) = (1-x(t))x(t)$ mit $x(0) = 0.2$ bzw. $x(0) = 1.2$.

Wenn wir mit einer Population starten, die kleiner als die Grenzpopulation ist, $x(0) = x_0 < 1$, so gilt $f(x_0) > 0$. Unsere Population nimmt also zu (siehe Abbildung 11.3 links). Je näher sie der Grenzpopulation kommt, umso *langsamer* wächst sie (f ist zwar positiv, wird aber immer kleiner – siehe Abbildung 11.2). Die Grenzpopulation (Nullstelle von f) wird daher erst im Grenzwert für $t \to \infty$ erreicht.

Starten wir analog mit einer Anfangspopulation, die größer als die Grenzpopulation ist, $x_0 > 1$, so ist f hier negativ. Die Population nimmt daher monoton ab und konvergiert wiederum für $t \to \infty$ gegen die Grenzpopulation (siehe Abbildung 11.3 rechts).

Beispiel 11.2. Logistisches Wachstum mit Ernte
Eine bestimmte Pilzkultur vermehrt sich nach dem logistischen Wachstumsmodell. Angenommen, es wird kontinuierlich eine Pilzmenge $h > 0$ pro Zeiteinheit geerntet. Was ist die optimale Erntemenge h?

Lösung: Um diese Situation zu modellieren, müssen wir die Ernterate h als zusätzlichen Term auf der rechten Seite der logistischen Differentialgleichung abziehen:

$$\frac{d}{dt}x(t) = \mu\,(1 - x(t))x(t) - h.$$

Geometrisch bedeutet das, dass die ursprüngliche Parabel um h nach unten verschoben wird. Je nachdem, wie groß h ist, wie weit die Parabel also nach unten verschoben wird, gibt es zwei, eine oder keine Nullstelle der Parabel $f_h(x) = \mu\,(1-x)x - h$. An einer Nullstelle von f besteht, wie zuvor, ein Gleichgewicht zwischen Pilzvermehrung und Ernte (die zeitliche Änderung der Pilzmenge ist null).

- Für $0 < h < \frac{\mu}{4}$ gibt es zwei Nullstellen der Parabel (siehe Abbildung 11.4).

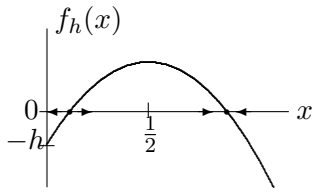


Abbildung 11.4. Wachstumsgeschwindigkeit beim logistischen Wachstum mit Ernte.

Links von der ersten Nullstelle ist $f_h(x)$ negativ. Wenn die Anfangspopulation also hier liegt, nimmt die Population monoton ab, bis sie ausgestorben ($x = 0$) ist.

Zwischen den beiden Nullstellen ist $f_h(x)$ positiv. Startet die Population in diesem Bereich, so wird sie daher zunehmen und gegen die Population, die durch die zweite Nullstelle von f gegeben ist, konvergieren.

Starten wir rechts von der zweiten Nullstelle, so ist $f_h(x)$ wieder negativ. Die Pilzmenge wird abnehmen und gegen die zweite Nullstelle von f konvergieren. Die Richtung, in die sich die Pilzmenge $x(t)$ ändert, ist in Abbildung 11.4 durch Pfeile angedeutet.

- Für $h = \frac{\mu}{4}$ fallen beide Nullstellen bei $x = \frac{1}{2}$ zusammen. Ist die Population zu Beginn kleiner als $x = \frac{1}{2}$, so nimmt die Pilzkultur ab, bis sie irgendwann ausgestorben ist (d.h., $x = 0$ erreicht ist). Starten wir mit einer Population größer als $x = \frac{1}{2}$, so nimmt sie ebenfalls ab, bis $x = \frac{1}{2}$ erreicht ist.
- Für $h > \frac{\mu}{4}$ sind alle Nullstellen verschwunden und $f_h(x)$ ist überall negativ. Egal, wo wir starten (d.h., wie viele Pilze zu Beginn vorhanden sind), die Population nimmt bei dieser Ernterate h immer ab, bis sie ausgestorben ist.

Diese Überlegungen zeigen, dass $h = \frac{\mu}{4}$ die aus theoretischer Sicht optimale (maximale) Ernterate ist. Praktisch gesehen gibt es dabei aber noch ein kleines Problem: Wie wir uns überlegt haben, konvergiert die Pilzmenge $x(t)$ für die Ernterate $h = \frac{\mu}{4}$ gegen die Nullstelle $\frac{1}{2}$ (falls die Anfangspopulation rechts davon liegt). Sie wird also nach einiger Zeit beliebig nahe bei $\frac{1}{2}$ liegen. Gibt es nun eine kleine Störung und die Pilzmenge sinkt unter $\frac{1}{2}$, so nimmt sie in Folge ab und stirbt aus! Die Lösung ist also *instabil*.

Wählen wir $h < \frac{\mu}{4}$, so gibt es, wie wir oben überlegt haben, den stabilen Bereich rechts von der linken Nullstelle: Wenn die Anfangspopulation irgendwo in diesem Bereich liegt, so wird sie (auch bei kleinen Störungen) immer zur rechten Nullstelle konvergieren. Deshalb ist eine Ernterate knapp unter $\frac{\mu}{4}$ aus praktischer Sicht zu bevorzugen. □

Ist es nicht ziemlich beachtlich, was man alleine von der Differentialgleichung über die Lösung ablesen kann, ohne die Lösung zu kennen!

Analog kann eine beliebige autonome Differentialgleichung erster Ordnung behandelt werden:

Satz 11.2. *Gegeben ist die autonome Differentialgleichung erster Ordnung*

$$\frac{d}{dt}x(t) = f(x(t))$$

mit einer differenzierbaren Funktion f. Dann gibt es zu jeder Anfangsbedingung $x(0) = x_0$ eine eindeutige Lösung, die in einem offenen Intervall um $t = 0$ definiert ist. Für diese gilt:

(a) *Ist $f(x_0) = 0$, so ist $x(t) = x_0$ für alle t. D.h., wenn wir bei einer Nullstelle von $f(x)$ starten, so bleibt die Lösung für alle Zeiten konstant gleich diesem Wert.*

(b) *Ist $f(x_0) \neq 0$, so konvergiert $x(t)$ gegen die erste Nullstelle links ($f(x_0) < 0$) bzw. rechts ($f(x_0) > 0$) von x_0. Gibt es keine solche Nullstelle, so konvergiert die Lösung gegen $-\infty$ bzw. ∞.*

Punkt a) entspricht im Erntebeispiel 11.2 dem Fall, dass gleich viel geerntet wird wie nachwächst. Bei dieser Populationsgröße ist also ein *Gleichgewicht* vorhanden:

Definition 11.3. *Die Nullstellen von $f(x)$ heißen* **Fixpunkte** *oder* **Gleichgewichtslagen** *der Differentialgleichung. Ein Fixpunkt x_0 heißt* **asymptotisch stabil**, *falls es ein offenes Intervall um x_0 gibt, sodass alle Lösungen, die hier starten, für $t \to \infty$ gegen x_0 konvergieren.*

Beispiel: Für die logistische Gleichung $f(x) = \mu\,(1-x)x$ ist $x_0 = 1$ asymptotisch stabil.

Im Erntebeispiel 11.2 ist, wie wir oben überlegt haben, für $h < \frac{\mu}{4}$ die rechte Nullstelle asymptotisch stabil. Das wird durch die beiden Pfeile bei dieser Nullstelle in Abbildung 11.4 angedeutet. Die Ableitung von f ist an dieser Nullstelle negativ. Allgemein können wir am Vorzeichen der Ableitung von f am Fixpunkt ablesen, ob dieser asymptotisch stabil ist oder nicht:

Satz 11.4. *Eine Gleichgewichtslage x_0 ist asymptotisch stabil, falls $f'(x_0) < 0$. Ist $f'(x_0) > 0$, so ist die Gleichgewichtslage x_0 nicht asymptotisch stabil.*

Beispiel: Für die logistische Gleichung gilt $f'(1) = -\mu < 0$.

Versuchen wir nun aber doch, die Lösung der logistischen Differentialgleichung zu finden.

Beispiel 11.3. Logistisches Wachstum
Lösen Sie die logistische Gleichung $x'(t) = \mu\, x(t)(1 - x(t))$.

Lösung: Bringen wir zunächst alle $x(t)$ auf die linke Seite, so lautet die Differentialgleichung

$$\frac{x'(t)}{x(t)(1-x(t))} = \mu.$$

Integrieren wir nun auf beiden Seiten von 0 bis t,

$$\int_0^t \frac{x'(s)}{x(s)(1-x(s))} ds = \int_0^t \mu\, ds,$$

und substituieren $y = x(s)$, so erhalten wir

$$\int_{x_0}^{x(t)} \frac{dy}{y(1-y)} = \mu\, t.$$

Verwendet man die *Partialbruchzerlegung*

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

so kann das Integral leicht gelöst werden:

$$\log(x(t)) - \log(1-x(t)) - \log(x_0) + \log(1-x_0) = \log \frac{x(t)(1-x_0)}{(1-x(t))x_0} = \mu\, t.$$

Auflösen nach $x(t)$ ergibt schließlich

$$x(t) = \frac{x_0 \mathrm{e}^{\mu t}}{1 + x_0(\mathrm{e}^{\mu t} - 1)} = \frac{1}{2}\left(\tanh(\frac{\mu}{2}(t - t_w)) + 1\right), \quad t_w = \frac{1}{\mu}\log(\frac{1}{x_0} - 1).$$

Die Lösung ist für $\mu = 1$ und für $x_0 = 0.2$ bzw. $x_0 = 1.2$ in Abbildung 11.3 dargestellt. □

Beachten Sie, dass das qualitative Verhalten von der expliziten Lösung schwerer als von der Differentialgleichung selbst abgelesen werden kann!

Satz 11.5 (Separation der Variablen)**.** *Die Lösung der Differentialgleichung*

$$\frac{d}{dt}x(t) = f(x(t))g(t), \qquad x(t_0) = x_0$$

kann durch Auflösen von

$$\int_{x_0}^{x(t)} \frac{dy}{f(y)} = \int_{t_0}^{t} g(s)ds$$

nach $x(t)$ *erhalten werden. Man spricht auch von* **Trennung der Variablen** *und nennt die Differentialgleichung* **separierbar** *(oder* **trennbar***).*

Wir haben hier die Integrationsvariablen y bzw. s genannt, um eine Verwechslung mit den Integrationsgrenzen zu vermeiden. Ist keine Anfangsbedingung gegeben, so kann auf beiden Seiten unbestimmt integrieren werden. Es reicht dabei, die Integrationskonstante auf einer Seite zu berücksichtigen.

Der Name „Trennung der Variablen“ kommt daher, dass von der Differentialgleichung $\frac{dx}{dt} = f(x)g(t)$ alle Terme mit x auf die eine Seite und alle Terme mit t auf die andere Seite des Gleichheitszeichens gebracht werden. Formal können wir dabei $\frac{dx}{dt}$ wie einen gewöhnlichen Bruch behandeln. Wir erhalten dann:

$$\frac{dx}{f(x)} = g(t)\,dt$$

Danach werden beide Seiten integriert, die linke nach x, die rechte nach t.

Separation der Variablen
Lösen Sie die Differentialgleichung

$$\frac{dx(t)}{dt} = 2\sqrt{x(t)}, \quad x(0) = x_0 \geq 0.$$

Geben Sie speziell die Lösungen für die Anfangsbedingungen
a) $x_0 = 3$ und b) $x_0 = 0$ an.

Die Differentialgleichung ist separierbar, denn es ist möglich, alle x-Terme auf die eine Seite zu bringen ($f(x) = 2\sqrt{x}$) und alle t-Terme auf die andere Seite (die t-Terme sind hier konstant: $g(t) = 1$):

$$\frac{dx}{2\sqrt{x}} = 1 \cdot dt.$$

Nun fügen wir auf beiden Seiten das Integralzeichen hinzu (und nennen die Integrationsvariablen y bzw. s, weil x bzw. t für die Integrationsgrenzen vergeben sind),

$$\int_{x_0}^{x(t)} \frac{dy}{2\sqrt{y}} = \int_0^t ds,$$

integrieren links und rechts,

$$\sqrt{x(t)} - \sqrt{x_0} = t - 0,$$

und lösen nach $x(t)$ auf:

$$x(t) = (t + \sqrt{x_0})^2, \quad x_0 \geq 0.$$

Setzen wir nun konkrete Anfangswerte x_0 ein:
a) Für die Anfangsbedingung $x_0 = 3$ erhalten wir die Lösung $x(t) = (t + \sqrt{3})^2$.
b) Für die Anfangsbedingung $x_0 = 0$ erhalten wir die Lösung $x(t) = t^2$. Für diesen Anfangswert gibt es eine spezielle Situation: Da $x_0 = 0$ ein Fixpunkt ist, ist auch $x(t) = 0$ eine Lösung (vergleiche Satz 11.2). Es gibt also zur Anfangsbedingung $x_0 = 0$ mehrere Lösungen. Das liegt daran, dass unsere Funktion $f(x) = \sqrt{x}$ bei 0 nicht differenzierbar ist. Für stetig differenzierbares f gibt es immer eine eindeutige Lösung:

Satz 11.6 (Picard-Lindelöf)**.** *Es sei f stetig differenzierbar. Dann hat die Differentialgleichung*

$$x^{(n)}(t) = f(x^{(n-1)}(t), \ldots, x(t), t)$$

zusammen mit den Anfangsbedingungen

$$x^{(n-1)}(t_0) = x_{n-1}, \quad \ldots, \quad x(t_0) = x_0$$

für jede Wahl der Anfangswerte $x_{n-1}, \ldots, x_0$ eine eindeutige Lösung, die in einem offenen Intervall um t_0 definiert ist.

Dieser Satz wurde zuerst vom finnischen Mathematiker Ernst Leonard Lindelöf (1870–1946) bewiesen. Der moderne Zugang formuliert das Problem als Fixpunktgleichung, die mit Iteration und dem Banach'schen Kontraktionsprinzip gelöst werden kann. Diese Picard-Iteration geht auf den französischen Mathematiker Charles Émile Picard (1856–1941) zurück. Der **Satz von Peano** (benannt nach Giuseppe Peano, italienischer Mathematiker, 1858–1932) besagt, dass es zumindest eine Lösung gibt, wenn f stetig ist.

Die **allgemeine Lösung** einer Differentialgleichung n-ter Ordnung hängt von n Parametern ab. Aus der allgemeinen Lösung erhält man *jede* Lösung der Differentialgleichung durch geeignete Wahl dieser Parameter. Insbesondere werden die Parameter durch die Vorgabe von n Anfangsbedingungen eindeutig bestimmt. Für eine Differentialgleichung erster Ordnung muss man also nur einen Anfangswert $x(t_0) = x_0$ vorgeben, um die Lösung eindeutig festzulegen. Für eine Differentialgleichung zweiter Ordnung benötigt man den Funktionswert $x(t_0) = x_0$ und die Ableitung $x'(t_0) = x_1$ zum Anfangszeitpunkt t_0, etc. Wählt man für die Anfangsbedingungen (bzw. die Parameter) konkrete Zahlenwerte, so spricht man von einer **speziellen Lösung** der Differentialgleichung. In Beispiel 11.1 haben wir zunächst die allgemeine Lösung berechnet, die vom Parameter x_0 abhängt. Danach haben wir die speziellen Lösungen zu den Anfangswerten $x_0 = 3$ bzw. $x_0 = 0$ berechnet.

Auch wenn f auf ganz $\mathbb{R}^{n+1}$ definiert ist, kann es sein, dass die Lösung nur *lokal*, d.h. in der Nähe von t_0, existiert.

Separation der Variablen
Lösen Sie das Anfangswertproblem

$$x'(t) = x(t)^2, \quad x(0) = x_0.$$

Lösung: Wir trennen die Variablen,

$$\frac{dx}{x^2} = dt,$$

schreiben das Integral an,

$$\int_{x_0}^{x(t)} \frac{dy}{y^2} = \int_0^t ds,$$

und integrieren beide Seiten:

$$\frac{1}{x_0} - \frac{1}{x(t)} = t - 0.$$

Auflösen nach x ergibt

$$x(t) = \frac{x_0}{1 - x_0\, t}.$$

Da die Lösung bei $t = \frac{1}{x_0}$ eine Polstelle hat, ist die Lösung nicht für alle $t \in \mathbb{R}$ definiert. Wenn $x_0 > 0$ ist, so ist das maximale offene Intervall, auf dem die Lösung definiert werden kann, gleich $(-\infty, \frac{1}{x_0})$. Das ist in Abbildung 11.5 für $x_0 = 2$ dargestellt. Ist hingegen $x_0 < 0$, so lebt die Lösung auf dem offenen Intervall $(\frac{1}{x_0}, \infty)$.

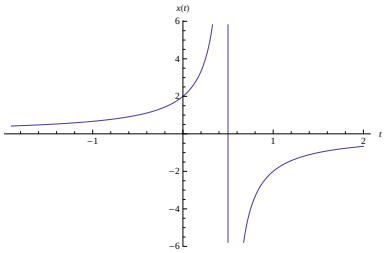


Abbildung 11.5. Die Lösung des Anfangswertproblems $x'(t) = x(t)^2$, $x(0) = 2$ ist auf dem offenen Intervall $(-\infty, \frac{1}{2})$ eindeutig definiert.

Natürlich ist $x(t)$ eine Lösung für alle $t \neq \frac{1}{x_0}$. Wenn die Lösung aber bei $x_0 > 0$ startet, dann verschwindet sie bei $t = \frac{1}{x_0}$ im Unendlichen, und die Werte der Lösung für $t > \frac{1}{x_0}$ haben damit keinerlei praktische Bedeutung mehr. In Abbildung 11.5 hat also die Lösung für $t > \frac{1}{2}$ keine praktische Bedeutung. Analoges gilt für $x_0 < 0$.

Im Fall $x_0 = 0$ gilt $x(t) = 0$ und die Lösung existiert für alle $t \in \mathbb{R}$. □

11.1.1. Anwendung: Parabolspiegel. Wir bringen ein interessantes Beispiel aus der Telekommunikation, das zeigt, wie Differentialgleichungen bei der Modellbildung in der Technik eingesetzt werden. Außerdem zeigt es, dass es in der Praxis schnell kompliziert werden kann.

Falls Sie sich schon immer gefragt haben, warum die Satellitenschüssel auf Ihrem Dach die Form eines Paraboloids hat, finden Sie hier die Antwort.

Wir wollen die optimale Form eines Spiegels, der geradlinig einfallende Strahlen in einem Punkt fokussiert, bestimmen.

Damit könnte eine Antenne gemeint sein, die elektromagnetische Wellen einer weit entfernten Quelle (z. B. eines Satelliten im All) auffängt, oder auch ein Scheinwerfer, der Licht von einem Punkt (dem Glühfaden) möglichst geradlinig aussendet.

Wir wählen unser Koordinatensystem so, dass die Strahlen in y-Richtung einfallen und im Ursprung fokussiert werden. Da das Problem rotationssymmetrisch um die y-Achse ist, reicht es, das Profil des Spiegels in der x, y-Ebene zu bestimmen. Wir suchen also das Profil $y(x)$.

Angenommen, ein Strahl trifft im Punkt (x, y) auf den Spiegel. Dann wird er dort reflektiert und in den Ursprung $(0, 0)$ weitergeleitet (siehe Abbildung 11.6). Das Reflexionsgesetz der geometrischen Optik besagt, dass der Winkel zwischen

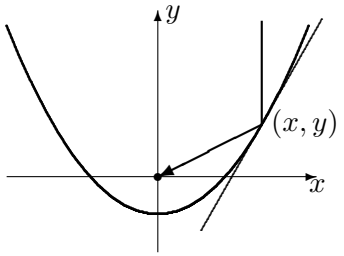


Abbildung 11.6. Fokussierung eines einfallenden Strahls durch einen Parabolspiegel.

einfallendem Strahl und der Tangente an den Spiegel gleich dem Winkel zwischen dem reflektierten Strahl und der Tangente sein muss („Einfallswinkel ist gleich Ausfallswinkel“). Aus der Vektorrechnung wissen wir, dass das genau dann der Fall ist, wenn das Skalarprodukt der entsprechenden Einheitsvektoren gleich ist

Der Einheitsvektor des einfallenden Strahls ist $(0, -1)$, der Einheitsvektor des reflektierten Strahls ist $\frac{1}{\sqrt{x^2+y^2}}(-x, -y)$ und der Einheitsvektor der Tangente ist $\frac{1}{\sqrt{1+(y')^2}}(1, y')$. Somit erhalten wir (der Faktor $\frac{-1}{\sqrt{1+(y')^2}}$ kann auf beiden Seiten gekürzt werden)

$$\frac{1}{\sqrt{x^2+y^2}}\langle \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} 1 \\ y' \end{pmatrix} \rangle = \langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ y' \end{pmatrix} \rangle$$

oder, ausmultipliziert,

$$\frac{1}{\sqrt{x^2+y^2}}(x + yy') = y'.$$

Lösen wir nach y' auf, so erhalten wir die Differentialgleichung

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

$$= \frac{\sqrt{x^2+y^2}+y}{x} = \sqrt{1+\left(\frac{y}{x}\right)^2}+\frac{y}{x}.$$

Diese Differentialgleichung ist zwar nicht separierbar, aber da auf der rechten Seite nur $\frac{y}{x}$ vorkommt, bietet es sich an, die neue Funktion

$$u(x) = \frac{y(x)}{x}, \qquad x > 0,$$

zu betrachten. Wegen

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

erhalten wir aus der Differentialgleichung für y folgende separierbare Differentialgleichung für u:

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

Es bleibt also folgendes Integral zu lösen:

$$\int \frac{du}{\sqrt{1+u^2}} = \int \frac{dx}{x}.$$

Das rechte Integral kennen wir, das linke müssen wir entweder in einer Formelsammlung nachschlagen oder dem Computer vorwerfen. Als Ergebnis erhalten wir

$$\operatorname{arsinh}(u) = \log(x) + C,$$

wobei $\operatorname{arsinh}(u)$ die Umkehrfunktion des Sinus hyperbolicus $\sinh(x) = \frac{1}{2}(\mathrm{e}^x - \mathrm{e}^{-x})$ ist.

Sie können das überprüfen, indem Sie $\operatorname{arsinh}'(u) = \frac{1}{\sqrt{1+u^2}}$ mithilfe der Ableitungsregel für die Umkehrfunktion nachrechnen.

Um den $\operatorname{arsinh}(u)$ loszuwerden, nehmen wir beide Seiten als Argument des sinh und erhalten

$$u = \sinh(\log(x) + C) = \frac{1}{2}(\mathrm{e}^C x - \frac{1}{\mathrm{e}^C x}).$$

Hier haben wir zuletzt den sinh gemäß seiner Definition mit der Exponentialfunktion ausgedrückt. Kürzen wir $\mathrm{e}^C = a$ ab, so sehen wir, dass die optimale Form durch die Parabel

$$y(x) = x\,u(x) = \frac{1}{2}\left(ax^2 - \frac{1}{a}\right)$$

gegeben ist.

11.2. Lineare Differentialgleichungen

Die meisten Differentialgleichungen können zwar nicht explizit gelöst werden, wie aber schon bei den Rekursionen gibt es eine wichtige Klasse von Differentialgleichungen, nämlich lineare mit konstanten Koeffizienten, bei denen das doch geht.

Definition 11.7. *Eine Differentialgleichung der Form*

$$\begin{aligned} x^{(n)}(t) &= \sum_{j=0}^{n-1} c_j(t)x^{(j)}(t) + g(t) \qquad (11.1)\\ &= c_{n-1}(t)x^{(n-1)}(t) + c_{n-2}(t)x^{(n-2)}(t) + \cdots + c_0(t)x(t) + g(t) \end{aligned}$$

heißt **lineare Differentialgleichung** *(n-ter Ordnung). Ist $g(t) = 0$ für alle t, so nennt man die Differentialgleichung* **homogen**, *ansonsten* **inhomogen**. *Dementsprechend heißt $g(t)$ auch* **inhomogener Anteil** *der Differentialgleichung. Hängen die Koeffizienten $c_j(t)$ nicht von t ab, so spricht man von* **konstanten Koeffizienten**.

Eine lineare Differentialgleichung ist genau dann autonom, wenn sie konstante Koeffizienten hat und der inhomogene Anteil $g(t)$ nicht vorhanden oder zumindest konstant ist.

Für eine lineare Differentialgleichung kann die Aussage des Satzes von Picard-Lindelöf verbessert werden: Sind die Koeffizienten $c_j(t)$ stetig auf $\mathbb{R}$, so existiert eine eindeutige Lösung des Anfangswertproblems für *alle* $t \in \mathbb{R}$.

Da man lineare Differentialgleichungen mit konstanten Koeffizienten ohne Probleme lösen kann, ist es wichtig, sie mit einem Schlag *erkennen* zu können:

Lineare Differentialgleichung
Klassifizieren Sie die Differentialgleichung:
a) $x'(t) = 3x(t) - t^2$ b) $x'''(t) = x'(t) + x(t)^2$ c) $y''(x) = \sqrt{2}\, y(x)$
d) $y''(x) = 4y'(x) + x^3 y(x)$ e) $f''(t) = (1 - f'(t))f(t)$

Lösung

(a) linear, inhomogener Anteil $g(t) = -t^2$, konstanter Koeffizient $c_0 = 3$; die Ordnung ist 1

(b) nicht linear, da $x(t)^2$ vorkommt; Ordnung 3

(c) linear, homogen, Koeffizienten $c_1 = 0$, $c_0 = \sqrt{2}$ sind konstant; Ordnung 2

(d) linear, homogen; keine konstanten Koeffizienten, denn c_0 hängt von x ab: $c_0(x) = x^3$, $c_1 = 4$; Ordnung 2

(e) nicht linear, da das Produkt $f'(t)f(t)$ vorkommt

□

Folgendes Problem führt zum Beispiel auf eine lineare Differentialgleichung mit konstanten Koeffizienten: Man betrachte die Reihenschaltung bestehend aus ei-

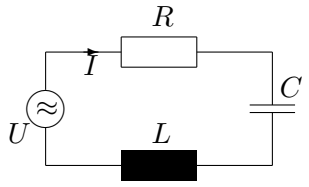


Abbildung 11.7. RLC-Schwingkreis

nem ohmschen Widerstand R, einem Kondensator mit Kapazität C und einer Spule mit Induktivität L in Abbildung 11.7. Nach der ersten Kirchhoff'schen Regel fließt durch alle Bauteile der gleiche Strom $I(t)$. Nach der zweiten Kirchhoff'schen Regel ist die Summe der Spannungsabfälle an den Bauteilen gleich der von der Spannungsquelle produzierten Spannung:

$$U(t) = U_R(t) + U_C(t) + U_L(t). \tag{11.2}$$

Für den Widerstand gilt nach dem Ohm'schen Gesetz

$$U_R(t) = R\, I(t).$$

Für die Spule gilt nach dem Induktionsgesetz

$$U_L(t) = L\frac{d}{dt}I(t),$$

und für den Kondensator

$$U_C(t) = \frac{1}{C}Q(t),$$

wobei $Q(t)$ die elektrische Ladung am Kondensator ist. Da der Strom per Definition gleich der Ladungsänderung ist, $I(t) = \frac{d}{dt}Q(t)$, erhalten wir folgende Differentialgleichung für die Ladung am Kondensator:

$$L\frac{d^2}{dt^2}Q(t) + R\frac{d}{dt}Q(t) + \frac{1}{C}Q(t) = U(t). \tag{11.3}$$

Durch Differenzieren beider Seiten ergibt sich die Differentialgleichung

$$L\frac{d^2}{dt^2}I(t) + R\frac{d}{dt}I(t) + \frac{1}{C}I(t) = \frac{d}{dt}U(t) \tag{11.4}$$

für den Strom.

Beispiel 11.4. Ladevorgang eines Kondensators
Lösen Sie die Differentialgleichung für die Ladung eines Kondensators

$$R\frac{d}{dt}Q(t) + \frac{1}{C}Q(t) = U_0 \tag{11.5}$$

durch eine Gleichspannung $U(t) = U_0$ (Fall ohne Spule: $L = 0$). Berechnen Sie die Ladung $Q(t)$ am Kondensator, falls der Kondensator zur Zeit $t = 0$ ungeladen ist. Welche Ladung stellt sich langfristig am Kondensator ein (was ist also $\lim_{t\to\infty} Q(t)$)?

Lösung: Die Differentialgleichung ist separierbar:

$$\frac{dQ}{\frac{U_0}{R} - \frac{1}{RC}Q} = dt.$$

Wir können, wie im letzten Abschnitt, die Lösung berechnen, indem wir beide Seiten integrieren (rechts von 0 bis t und links von der Anfangsladung $Q(0) = 0$ bis zur Ladung $Q(t)$ zum Zeitpunkt t):

$$\int_0^{Q(t)} \frac{dy}{\frac{U_0}{R} - \frac{1}{RC}y} = \int_0^t ds$$

bzw., nach der Integration,

$$-RC\left(\log\left(\frac{U_0}{R} - \frac{1}{RC}Q(t)\right) - \log\left(\frac{U_0}{R}\right)\right) = t.$$

Durch Umformen,

$$\log\left(1 - \frac{1}{U_0 C}Q(t)\right) = -\frac{t}{RC},$$

und Auflösen nach $Q(t)$ erhalten wir die Lösung

$$Q(t) = U_0 C(1 - \mathrm{e}^{-\frac{t}{RC}}). \tag{11.6}$$

Sie nähert sich exponentiell der maximalen Ladung U_0C: $\lim_{t\to\infty} Q(t) = U_0C$ (siehe Abbildung 11.8). □

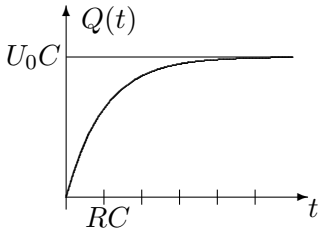


Abbildung 11.8. Aufladung eines Kondensators

Den einfachsten Fall einer homogenen linearen Differentialgleichung erster Ordnung mit konstantem Koeffizient, $x'(t) = c\,x(t)$, können wir also leicht durch Separation der Variablen lösen: $x(t) = x(0)\mathrm{e}^{ct}$. Auch für die zugehörige inhomogene Differentialgleichung lässt sich die Lösung finden:

Satz 11.8. *Die Lösung der linearen Differentialgleichung erster Ordnung*

$$x'(t) = c\,x(t) + g(t)$$

ist

$$x(t) = x(t_0)\mathrm{e}^{c\,(t-t_0)} + \int_{t_0}^{t} \mathrm{e}^{c\,(t-s)} g(s)ds. \tag{11.7}$$

Wieder können Sie die Probe durch Differenzieren machen und so überprüfen, ob das in der Tat die Lösung ist!

Es kann sogar die allgemeine lineare Differentialgleichung erster Ordnung (ohne konstante Koeffizienten)

$$x'(t) = c(t)x(t) + g(t)$$

gelöst werden:

$$x(t) = x(t_0)\mathrm{e}^{C(t)-C(t_0)} + \int_{t_0}^{t} \mathrm{e}^{C(t)-C(s)} g(s)ds,$$

wobei $C(t) = \int c(t)dt$ eine Stammfunktion von $c(t)$ ist. Machen Sie wiederum die Probe![1]

Beispiel 11.5. Kondensator im Wechselstromkreis
Lösen Sie die Differentialgleichung für einen Kondensator im Wechselstromkreis mit einer sinusförmigen Spannungsquelle $U(t) = U_0\cos(\omega t)$:

$$R\frac{d}{dt}I(t) + \frac{1}{C}I(t) = \frac{d}{dt}U(t) = -\omega U_0 \sin(\omega t). \tag{11.8}$$

Lösung: Wir werden diese Gleichung nicht direkt lösen, sondern einen in der Elektrotechnik üblichen Trick anwenden. Dazu ersetzen wir formal die Wechselspannung $U(t) = U_0\cos(\omega t)$ durch $U(t) = U_0\mathrm{e}^{\mathrm{i}\omega t}$ („komplexe Spannung“). Auf der rechten Seite der Differentialgleichung steht dann also die Ableitung davon:

$$R\frac{d}{dt}I(t) + \frac{1}{C}I(t) = \frac{d}{dt}U(t) = \mathrm{i}\omega U_0\mathrm{e}^{\mathrm{i}\omega t}.$$

Wir lösen diese Differentialgleichung (die Rechnung wird nun um einiges einfacher!) und nehmen von der erhaltenen (komplexen) Lösung den Realteil. Das ist dann die Lösung der ursprünglichen Differentialgleichung mit Spannung $U(t) = U_0\cos(\omega t)$!

Warum? Aufgrund der Euler'schen Formel

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

ist die gegebene Spannung gerade der Realteil der formalen komplexen Spannung:

$$U_0\mathrm{e}^{\mathrm{i}\omega t} = U_0\cos(\omega t) + \mathrm{i}U_0\sin(\omega t).$$

[1] Die Lösung der inhomogenen Differentialgleichung erhält man mit der Methode der “Variation der Konstanten” (siehe [**97**]).

Indem wir die Differentialgleichung mit der komplexen Spannung lösen, lösen wir zwei Differentialgleichungen simultan: einmal für die Spannung $U(t) = U_0 \cos(\omega t)$ (Realteil der Rechnung) und einmal für die Spannung $U(t) = U_0 \sin(\omega t)$ (Imaginärteil der Rechnung).

Die Lösung der Differentialgleichung mit komplexer Spannung kann wie im reellen Fall erhalten werden, und auch die Formeln für das Differenzieren und Integrieren ändern sich nicht. Wir können also so rechnen, als ob wir es mit reellen Funktionen zu tun hätten. Lösen wir zunächst nach der Ableitung auf:

$$I' = -\frac{1}{RC} I + \frac{\mathrm{i}\omega U_0}{R} \mathrm{e}^{\mathrm{i}\omega t}.$$

Nach Satz 11.8 gilt nun

$$\begin{aligned} I(t) &= I(0)\mathrm{e}^{-\frac{t}{RC}} + \int_0^t \mathrm{e}^{-\frac{t-s}{RC}} \frac{\mathrm{i}\omega U_0}{R} \mathrm{e}^{\mathrm{i}\omega s} ds \\ &= I(0)\mathrm{e}^{-\frac{t}{RC}} + \frac{\mathrm{i}\omega U_0}{R} \mathrm{e}^{-\frac{t}{RC}} \int_0^t \mathrm{e}^{\frac{s}{RC} + \mathrm{i}\omega s} ds \\ &= I(0)\mathrm{e}^{-\frac{t}{RC}} + \frac{\mathrm{i}\omega U_0}{R} \frac{1}{\frac{1}{RC} + \mathrm{i}\omega} \mathrm{e}^{-\frac{t}{RC}} \left(\mathrm{e}^{\frac{t}{RC} + \mathrm{i}\omega t} - 1 \right) \\ &= \left(I(0) - \frac{U_0}{Z} \right) \mathrm{e}^{-\frac{t}{RC}} + \frac{U_0}{Z} \mathrm{e}^{\mathrm{i}\omega t}, \end{aligned} \tag{11.9}$$

wobei wir die in der Elektrotechnik übliche Abkürzung

$$Z = R - \mathrm{i}\frac{1}{\omega C}$$

(**Impedanz** bzw. Wechselstromwiderstand) eingeführt haben.

Diese Größe $Z_C = -\mathrm{i}\frac{1}{\omega C}$ kann formal als Widerstand des Kondensators aufgefasst werden.

Wegen $-\frac{1}{RC} < 0$ klingt $\mathrm{e}^{-\frac{t}{RC}}$ exponentiell ab. Daher ist nach kurzer Zeit nur noch der zweite Anteil der Lösung sichtbar:

$$I(t) \approx \frac{1}{Z} U(t).$$

Um die Lösung unseres ursprünglichen Problems zu erhalten, nehmen wir nun davon den Realteil. Dazu schreiben wir $Z = |Z|\mathrm{e}^{\mathrm{i}\varphi}$, dann gilt

$$\mathrm{Re}(I(t)) = \mathrm{Re}\left(\frac{U_0}{|Z|\mathrm{e}^{\mathrm{i}\varphi}} \mathrm{e}^{\mathrm{i}\omega t}\right) = \frac{U_0}{|Z|} \mathrm{Re}(\mathrm{e}^{\mathrm{i}(\omega t - \varphi)}) = \frac{U_0}{|Z|} \cos(\omega t - \varphi).$$

Der Kondensator bewirkt also eine zusätzliche Phasenverschiebung um den Winkel $\varphi = \arctan(-\frac{1}{R\omega C})$ des Stroms gegenüber der Spannung.

Indem wir für den Kondensator formal den Widerstand Z_C eingeführt und gemeinsam mit R zur Impedanz Z zusammengefasst haben, können wir $U = ZI$ schreiben, was an das Ohm'sche Gesetz erinnert.

Der Realteil R der Impedanz $Z = R - \mathrm{i}\frac{1}{\omega C}$ wird auch als **Wirkwiderstand** und der Imaginärteil $\frac{-1}{\omega C}$ als **Blindwiderstand** bezeichnet. Der Absolutbetrag $|Z| = \sqrt{R^2 + \frac{1}{\omega^2 C^2}}$ heißt **Scheinwiderstand**. □

Kommen wir nun zu linearen Differentialgleichungen zweiter Ordnung. Beginnen wir mit der homogenen Differentialgleichung mit konstanten Koeffizienten,

$$x''(t) = c_1 x'(t) + c_0 x(t).$$

Um die allgemeine Lösung zu finden, müssen wir ein wenig ausholen.

Eine wichtige Eigenschaft homogener linearer Differentialgleichungen (beliebiger Ordnung) ist, dass das Vielfache einer Lösung, sowie die Summe zweier Lösungen, wieder Lösungen sind. Ist die Ordnung zwei, so *reichen zwei Lösungen aus*, um *alle* weiteren Lösungen anzugeben:

Satz 11.9 (Superpositionsprinzip)**.** *Sind $x_1(t)$ und $x_2(t)$ zwei Lösungen der homogenen linearen Differentialgleichung*

$$x''(t) = c_1 x'(t) + c_0 x(t),$$

und ist keine Lösung ein Vielfaches der anderen, so lässt sich jede Lösung als Linearkombination

$$k_1 x_1(t) + k_2 x_2(t)$$

dieser beiden Lösungen schreiben. Die Konstanten k_1 und k_2 können aus den Anfangsbedingungen bestimmt werden.

Das Superpositionsprinzip gilt auch für homogene lineare Differentialgleichungen n-ter Ordnung mit *nicht-konstanten* Koeffizienten, z.B. $x''(t) = c_1(t)x'(t) + c_0(t)x(t)$. Es besagt, dass die Lösungen einen Vektorraum der Dimension n bilden (der von n linear unabhängigen Lösungen aufgespannt wird).

Wenn wir also zwei spezielle Lösungen der homogenen Differentialgleichung kennen, die nicht Vielfache voneinander sind, so haben wir damit bereits die allgemeine Lösung der homogenen Differentialgleichung gefunden.

Wie überprüft man, ob zwei Lösungen linear unabhänging sind? Seien x_1, x_2 zwei Lösungen, dann bildet man die Lösungsmatrix

$$X(t) = \begin{pmatrix} x_1(t) & x_2(t) \\ x_1'(t) & x_2'(t) \end{pmatrix}.$$

Die Determinante der Lösungsmatrix $\det(X(t))$ heißt Wronskideterminate. Ist die Wronskideterminate für einen Wert t_0 ungleich 0, so ist sie es für alle Werte und die beiden Lösungen sind linear unabhänging.

Dies lässt sich auch auf Differentialgleichungen n-ter Ordnung und Systeme von Differentialgleichungen verallgemeinern. Nach dem Satz von Liouville erfüllt die Wronskideterminate eine lineare Differentialgleichung 1. Ordnung (siehe [**97**]).

Wie finden wir nun zwei geeignete spezielle Lösungen der homogenen Differentialgleichung? Dazu setzen wir den **Ansatz** $x(t) = \mathrm{e}^{\lambda t}$ (mit einem λ, das wir bestimmen möchten) in die homogene Differentialgleichung

$$x''(t) = c_1 x'(t) + c_0 x(t)$$

ein: $\lambda^2 \mathrm{e}^{\lambda t} = c_1 \lambda \mathrm{e}^{\lambda t} + c_0 \mathrm{e}^{\lambda t}$. Kürzen wir nun auf beiden Seiten durch $\mathrm{e}^{\lambda t}$, so sehen wir, dass der Ansatz $x(t) = \mathrm{e}^{\lambda t}$ genau dann eine Lösung der homogenen Differentialgleichung ist, wenn λ die so genannte **charakteristische Gleichung**

$$\lambda^2 = c_1 \lambda + c_0$$

erfüllt. Nun gibt es (wie bei jeder quadratischen Gleichung) drei Möglichkeiten für die Lösungen λ_1 und λ_2 der charakteristischen Gleichung:

Satz 11.10 (Lineare homogene Differentialgleichung 2. Ordnung)**.** *Gegeben ist die Differentialgleichung*

$$x''(t) = c_1 x'(t) + c_0 x(t) \qquad \text{mit } c_1, c_0 \in \mathbb{R}.$$

Sind λ_1, λ_2 die Nullstellen

$$\lambda_1 = \frac{c_1}{2} + \sqrt{\left(\frac{c_1}{2}\right)^2 + c_0}, \quad \lambda_2 = \frac{c_1}{2} - \sqrt{\left(\frac{c_1}{2}\right)^2 + c_0}$$

der **charakteristischen Gleichung** $\lambda^2 = c_1\lambda + c_0$*, so gilt (Fallunterscheidung):*

(i) *Wenn λ_1 und λ_2 verschieden und reell sind, dann hat die homogene Differentialgleichung die allgemeine Lösung*

$$x(t) = k_1 \mathrm{e}^{\lambda_1 t} + k_2 \mathrm{e}^{\lambda_2 t},$$

wobei k_1, k_2 reelle Zahlen sind, die durch die Anfangsbedingungen festgelegt werden. Sie ergeben sich aus dem Gleichungssystem $x(0) = k_1 + k_2$, $x'(0) = k_1\lambda_1 + k_2\lambda_2$.

(ii) *Sind die beiden Lösungen der charakteristischen Gleichung identisch, $\lambda_1 = \lambda_2 = \lambda$, so ist die allgemeine Lösung der Differentialgleichung durch*

$$x(t) = (k_1 + k_2 t)\mathrm{e}^{\lambda t}$$

gegeben. Die Zahlen k_1, k_2 ergeben sich aus den Anfangsbedingungen $x(0) = k_1$, $x'(0) = k_1\lambda + k_2$.

(iii) *Sind beide Lösungen nicht reell, dann sind sie konjugiert komplex, $\lambda_1 = \alpha + \mathrm{i}\beta$ und $\lambda_2 = \alpha - \mathrm{i}\beta$, und die allgemeine Lösung der Differentialgleichung lautet*

$$x(t) = k_1 \mathrm{e}^{\alpha t}\cos(\beta t) + k_2 \mathrm{e}^{\alpha t}\sin(\beta t).$$

Die Zahlen k_1, k_2 ergeben sich aus den Anfangsbedingungen $x(0) = k_1$, $x'(0) = k_1\alpha + k_2\beta$.

Anmerkung: Im Fall (ii) hat man nur ein λ und erhält die zweite Lösung durch die Methode “Variation der Konstanten”.

Sehen wir uns gleich ein Beispiel an:

Beispiel 11.6. Lineare homogene Differentialgleichung 2. Ordnung

a) *Lösen Sie die Differentialgleichung* $x'' = x' + 6x$ *mit den Anfangsbedingungen* $x(0) = 1$ *und* $x'(0) = 8$.
b) *Lösen Sie die Differentialgleichung* $x'' = 2x' - x$ *mit den Anfangsbedingungen* $x(0) = 1$ *und* $x'(0) = 8$.
c) *Lösen Sie die Differentialgleichung* $x'' = 2x' - 2x$ *mit den Anfangsbedingungen* $x(0) = 0$ *und* $x'(0) = 1$.

Lösung:

(a) Die Koeffizienten sind $c_1 = 1$ und $c_0 = 6$. Daher lautet die charakteristische Gleichung $\lambda^2 = \lambda + 6$. Sie hat die beiden Lösungen $\lambda_1 = 3$, $\lambda_2 = -2$. Damit hat die allgemeine Lösung der Differentialgleichung die Form

$$x(t) = k_1 \mathrm{e}^{3t} + k_2 \mathrm{e}^{-2t}.$$

Die Zahlen k_1 und k_2 werden nun mithilfe der Anfangsbedingungen bestimmt:

$$\begin{aligned} x(0) &= 1 = k_1 + k_2 \\ x'(0) &= 8 = 3k_1 - 2k_2 \end{aligned}$$

Wenn wir diese beiden linearen Gleichungen für k_1 und k_2 lösen, so erhalten wir $k_1 = 2$ und $k_2 = -1$. Die gesuchte spezielle Lösung zu den Anfangsbedingungen $x(0) = 1$ und $x'(0) = 8$ lautet damit $x(t) = 2\mathrm{e}^{3t} - \mathrm{e}^{-2t}$.

(b) Nun sind die Lösungen der charakteristischen Gleichung $\lambda^2 = 2\lambda - 1$ identisch: $\lambda_1 = \lambda_2 = 1$. Also ist die allgemeine Lösung $x(t) = k_1 \mathrm{e}^t + k_2 t\mathrm{e}^t$. Aus den Anfangsbedingungen folgt $x(0) = k_1 = 1$, $x'(0) = k_1 + k_2 = 8$, daher $x(t) = (1 + 7t)\mathrm{e}^t$.

(c) Die charakteristische Gleichung ist $\lambda^2 = 2\lambda - 2$ und diesmal sind die Lösungen konjugiert komplex: $\lambda_{1,2} = 1 \pm \mathrm{i}$. Also ist die allgemeine Lösung $x(t) = k_1 \mathrm{e}^t \cos(t) + k_2 \mathrm{e}^t \sin(t)$. Aus den Anfangsbedingungen folgt $x(0) = k_1 = 0$ und $x'(0) = k_1 + k_2 = 1$, also $k_2 = 1$. Somit ist $x(t) = \mathrm{e}^t \sin(t)$ die gesuchte Lösung der Differentialgleichung.

□

Nun haben wir homogene Differentialgleichungen mit konstanten Koeffizienten im Griff und können als Nächstes zu inhomogenen Differentialgleichungen kommen. Analog wie im Fall erster Ordnung gilt:

Satz 11.11 (Lineare inhomogene Differentialgleichung 2. Ordnung)**.** *Die allgemeine Lösung einer linearen inhomogenen Differentialgleichung zweiter Ordnung mit konstanten Koeffizienten $c_1, c_0 \in \mathbb{R}$,*

$$x''(t) = c_1 x'(t) + c_0 x(t) + g(t),$$

hat die Form

$$x(t) = x_h(t) + x_i(t),$$

wobei $x_h(t)$ die allgemeine Lösung der zugehörigen homogenen Differentialgleichung

$$x_h''(t) = c_1 x_h'(t) + c_0 x_h(t)$$

und $x_i(t)$ irgendeine spezielle Lösung der gegebenen inhomogenen Differentialgleichung ist.

Warum? – Sind $x(t)$ und $y(t)$ zwei spezielle Lösungen der inhomogenen Differentialgleichung, so erfüllt ihre Differenz $h(t) = x(t) - y(t)$ die zugehörige homogene Differentialgleichung: $h''(t) = x''(t) - y''(t)$ $= c_1 x'(t) + c_0 x(t) + g(t) - c_1 y'(t) - c_0 y(t) - g(t) = c_1 h'(t) + c_0 h(t)$. Zwei Lösungen der inhomogenen Differentialgleichung unterscheiden sich also um eine Lösung der homogenen Differentialgleichung.

Eine *spezielle* Lösung der inhomogenen Differentialgleichung lässt sich oft erraten bzw. durch einen geschickten Ansatz finden: Ist $g(t) = p(t)\mathrm{e}^{at}$ mit einem Polynom $p(t)$, so kann man auch für die spezielle Lösung eine Funktion dieser Form, $x_i(t) = q(t)\mathrm{e}^{at}$, ansetzen. Dabei hat das Polynom $q(t)$ denselben Grad wie $p(t)$, falls $a \neq \lambda_1, \lambda_2$, um eins höheren Grad als $p(t)$, falls $a = \lambda_1 \neq \lambda_2$ bzw. um zwei höheren Grad als $p(t)$, falls $a = \lambda_1 = \lambda_2$.

Hilfreich bei der Suche nach einer speziellen Lösung der inhomogenen Differentialgleichung ist auch folgende Eigenschaft: Besteht der inhomogene Anteil aus zwei Summanden, $g(t) = g_1(t) + g_2(t)$, so kann für jeden Summanden $g_j(t)$ eine zugehörige spezielle Lösung $x_{i,j}(t)$ ermittelt werden: Wir suchen also eine spezielle Lösung $x_{i,1}$ von $x''(t) = c_1 x'(t) + c_0 x(t) + g_1(t)$, und analog eine spezielle Lösung $x_{i,2}(t)$ von $x''(t) = c_1 x'(t) + c_0 x(t) + g_2(t)$. Die Summe $x_i(t) = x_{i,1}(t) + x_{i,2}(t)$ ist dann eine spezielle Lösung der ursprünglichen Differentialgleichung mit dem inhomogenen Anteil $g(t) = g_1(t) + g_2(t)$.

Es gibt sogar eine explizite Formel, die eine spezielle Lösung der inhomogen Differentialgleichung mithilfe der allgemeinen Lösung der homogenen Differentialgleichung ausdrückt:

$$x_i(t) = \int_0^t s(t-r)g(r)dr,$$

wobei $s(t)$ die Lösung der homogenen Differentialgleichung zu den Anfangsbedingungen $s(0) = 0$, $s'(0) = 1$ ist.

Ist insbesondere $g(t) = g$ konstant, so können wir wieder nach einem Fixpunkt $x(t) = \bar{x}$ suchen (und haben damit eine spezielle Lösung gefunden): Aus $x''(t) =$

$x'(t) = 0$ (das ist die Bedingung für einen Fixpunkt) folgt

$$0 = c_0\bar{x} + g,$$

und daraus $\bar{x} = -\frac{g}{c_0}$.

Satz 11.12. *Die lineare Differentialgleichung zweiter Ordnung*

$$x''(t) = c_1x'(t) + c_0x(t) + g, \qquad \text{mit } c_1, c_0, g \in \mathbb{R},$$

hat für $c_0 \neq 0$ (d.h., $\lambda_1, \lambda_2 \neq 0$) die allgemeine Lösung

$$x(t) = x_h(t) + \bar{x}, \qquad \text{mit} \quad \bar{x} = -\frac{g}{c_0},$$

wobei $x_h(t)$ die allgemeine Lösung der zugehörigen homogenen Differentialgleichung (siehe Satz 11.10) ist.

Haben beide Nullstellen der charakteristischen Gleichung Realteile kleiner null, $\mathrm{Re}(\lambda_1) < 0$ und $\mathrm{Re}(\lambda_2) < 0$, so konvergiert jede Lösung für $t \to \infty$ gegen den Fixpunkt $\bar{x}$.

Beispiel: Ist $c_0 = 0$, so ist $\lambda_1 = c_1$ und $\lambda_2 = 0$. In diesem Fall ist für $c_1 \neq 0$ eine spezielle Lösung durch $x_i(t) = -\frac{g\,t}{c_1}$ gegeben, und für $c_1 = 0$ durch $x_i(t) = \frac{g\,t^2}{2}$.

Sogar lineare inhomogene Differentialgleichungen mit konstanten Koeffizienten *beliebiger* Ordnung n,

$$x^{(n)}(t) = c_{n-1}x^{(n-1)}(t) + c_{n-2}x^{(n-2)}(t) + \cdots + c_0x(t) + g(t),$$

können immer gelöst werden. Man geht dabei wie im Fall der Ordnung zwei vor: Die allgemeine Lösung hat wieder die Form $x(t) = x_h(t) + x_i(t)$, wobei $x_h(t)$ die allgemeine Lösung der zugehörigen homogenen und $x_i(t)$ eine spezielle Lösung der inhomogenen Differentialgleichung ist. Um x_h zu finden, betrachten wir wieder das charakteristische Polynom. Es hat Grad n und deshalb n Nullstellen. Aus diesen erhalten wir n spezielle Lösungen, und somit die allgemeine Lösung x_h (als Linearkombination aus diesen speziellen Lösungen).

Sind aber die Koeffizienten *nicht konstant* oder ist die Differentialgleichung *nicht linear*, so ist es in der Regel nicht mehr möglich, eine Lösung anzugeben!

Zum Abschluss wollen wir wieder ein etwas aufwändigeres Praxisbeispiel lösen:

Beispiel 11.7. RLC-Schwingkreis
Lösen Sie die Differentialgleichung für den RLC-Schwingkreis

$$L\frac{d^2}{dt^2}I(t) + R\frac{d}{dt}I(t) + \frac{1}{C}I(t) = \frac{d}{dt}U(t)$$

mit einer sinusförmigen Spannungsquelle $U(t) = U_0\cos(\omega t)$.

Lösung: Wir verwenden wieder die komplexe Version $U(t) = U_0 \mathrm{e}^{\mathrm{i}\omega t}$ (vergleiche Beispiel 11.5):

$$I'' = -\frac{R}{L} I' - \frac{1}{LC} I + \mathrm{i}\frac{\omega U_0}{L} \mathrm{e}^{\mathrm{i}\omega t}.$$

Die beiden Nullstellen der charakteristischen Gleichung $\lambda^2 = -\frac{R}{L}\lambda - \frac{1}{LC}$ lauten

$$\lambda_{1,2} = -\frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}.$$

Falls $\frac{R^2}{4L^2} - \frac{1}{LC} > 0$, so sind beide Nullstellen negativ und die allgemeine Lösung der homogenen Differentialgleichung ist

$$I_h(t) = k_1 \mathrm{e}^{\lambda_1 t} + k_2 \mathrm{e}^{\lambda_2 t}.$$

Wenn $\frac{R^2}{4L^2} - \frac{1}{LC} = 0$, so sind beide Nullstellen gleich und die allgemeine Lösung der homogenen Differentialgleichung ist

$$I_h(t) = (k_1 + k_2 t)\mathrm{e}^{-\frac{R}{2L}t}.$$

Für $\frac{R^2}{4L^2} - \frac{1}{LC} < 0$ sind schließlich beide Nullstellen konjugiert komplex und die allgemeine Lösung der homogenen Differentialgleichung lautet

$$I_h(t) = k_1 \mathrm{e}^{-\frac{R}{2L}t} \cos(\beta t) + k_2 \mathrm{e}^{-\frac{R}{2L}t} \sin(\beta t), \qquad \beta = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} > 0.$$

In jedem Fall ist der Realteil $\frac{-R}{2L}$ der beiden Nullstellen negativ und die Lösung der homogenen Differentialgleichung verschwindet daher exponentiell für $t \to \infty$.

Für die inhomogene Lösung machen wir den Ansatz

$$I_i(t) = k\, \mathrm{e}^{\mathrm{i}\omega t}$$

mit einer unbestimmten Konstante k. Wir setzen ihn in die Differentialgleichung ein,

$$(-\omega^2 L + \mathrm{i}\omega R + \frac{1}{C}) k\, \mathrm{e}^{\mathrm{i}\omega t} = \mathrm{i}\omega U_0 \mathrm{e}^{\mathrm{i}\omega t},$$

und lösen nach k auf:

$$k = \frac{U_0}{R + \mathrm{i}(L\omega - \frac{1}{\omega C})}.$$

Da die homogene Lösung $I_h(t)$ exponentiell abklingt, gilt nach kurzer Zeit

$$I(t) = I_h(t) + I_i(t) \approx I_i(t) = \frac{U_0}{Z} \mathrm{e}^{\mathrm{i}\omega t} = \frac{1}{Z} U(t),$$

wobei wir analog zu Beispiel 11.5 die Impedanz (Wechselstromwiderstand)

$$Z = R + Z_L + Z_C, \qquad Z_L = \mathrm{i}L\omega, \quad Z_C = -\mathrm{i}\frac{1}{\omega C},$$

eingeführt haben. Der Strom $I(t) = \frac{1}{Z}U(t)$ bzw. $\mathrm{Re}(I(t)) = \frac{1}{|Z|}U_0\cos(\omega t - \varphi)$ wird maximal, wenn

$$|Z|^2 = R^2 + (L\omega - \frac{1}{\omega C})^2$$

minimal wird, wenn also $L\omega - \frac{1}{\omega C} = 0$, d.h.,

$$\omega = \frac{1}{\sqrt{LC}}$$

gilt. Die Frequenz $\frac{1}{2\pi\sqrt{LC}}$ wird als **Resonanzfrequenz** des Schwingkreises bezeichnet.[2]

Dieses Prinzip liegt zum Beispiel dem Radioempfang zugrunde. Die Spannungsquelle entspricht in diesem Fall dem externen Signal, das von einer Antenne empfangen wird. Wird das Signal auf einer Trägerwelle mit der Frequenz $\frac{\omega}{2\pi}$ übertragen, so gerät der Schwingkreis nur dann in Schwingung, wenn die Frequenz der Trägerwelle mit der Resonanzfrequenz des Schwingkreises übereinstimmt. Die Resonanzfrequenz $\frac{1}{2\pi\sqrt{LC}}$ kann durch Änderung der Kapazität C des Kondensators (oder der Induktivität L der Spule) angepasst werden. Genau das machen Sie nämlich, wenn Sie den Sender einstellen. □

Das Phänomen der Resonanz kann dramatische Effekte auslösen (**Resonanzkatastrophe**): Eine Kompanie, die über eine Brücke marschiert und die Resonanzfrequenz der Brücke erwischt, kann sie zum Einsturz bringen.

Auch kleine Kinder sind mit Resonanzphänomenen vertraut: Wenn man auf der Schaukel wie wild mit den Füßen strampelt, wird man nie passable Auslenkungen schaffen. Nur wer seine Beine im Takt mit der Schaukel anzieht und streckt, kann die Auslenkung vergrößern!

Unser Beispiel ist übrigens nicht so speziell, wie es auf den ersten Blick erscheinen mag. Jedes schwingungsfähige System lässt sich, zumindest für kleine Amplituden, durch die Differentialgleichung

$$x''(t) + \rho\, x'(t) + \omega_0^2 x(t) = 0, \qquad \rho, \omega_0 > 0,$$

beschreiben. Dabei ist $\frac{\omega_0}{2\pi}$ die Eigenfrequenz und ρ die „Reibungskonstante“. Falls $\rho = 0$, so ist die Lösung eine *freie* (d.h., ungedämpfte) Schwingung $k_1\cos(\omega_0 t) + k_2\sin(\omega_0 t)$ mit der Eigenfrequenz $\frac{\omega_0}{2\pi}$. Für $\rho > 0$ ist die Schwingung gedämpft und klingt exponentiell ab (für $\rho > 2\omega_0$ sind beide Nullstellen der charakteristischen Gleichung negativ und es gibt überhaupt keine Schwingung mehr).

11.2.1. Ausblick: Systeme von Differentialgleichungen. Wir haben gesehen, dass das Wachstum einer Population durch eine Differentialgleichung beschrieben werden kann. Was passiert aber, wenn man es mit mehr als einer Spezies zu tun hat, die sich gegenseitig beeinflussen?

[2] Kreisfrequenz ω: `https://de.wikipedia.org/wiki/Kreisfrequenz`

Betrachten wir zum Beispiel zwei Populationen $x(t)$ (Beute) und $y(t)$ (Räuber). Ohne Räuber würde sich die Beute gemäß $x'(t) = \alpha x(t)$ vermehren. Sind aber Räuber vorhanden, so vermindert sich die Wachstumsrate um einen Term $\beta y(t)$, der proportional zur Anzahl der Räuber ist. Das führt uns auf die Gleichung

$$x'(t) = (\alpha - \beta y(t))x(t)$$

für die Beutetiere. Analog können wir ansetzen, dass die Räuber ohne Beute mit einer Sterberate γ aussterben würden: $y'(t) = -\gamma y(t)$. Durch das Vorhandensein von Beutetieren vermindert sich die Sterberate um einen Term $\delta x(t)$ proportional zur Anzahl der Beutetiere und wir erhalten die Gleichung

$$y'(t) = -(\gamma - \delta x(t))y(t)$$

für die Räuber. Beide Gleichungen enthalten sowohl $x(t)$ als auch $y(t)$ und können daher nicht getrennt gelöst werden. Sie bilden zusammen ein **System von Differentialgleichungen erster Ordnung**, das sogenannte **Volterra-Lotka Räuber-Beute-Modell**:

$$\begin{aligned} x'(t) &= (\alpha - \beta y(t))x(t), \\ y'(t) &= -(\gamma - \delta x(t))y(t). \end{aligned} \tag{11.10}$$

Das Modell ist benannt nach dem österreichischen Mathematiker Alfred James Lotka, 1880–1949, und dem italienischen Mathematiker und Physiker Vito Volterra, 1860–1940.

Es kann nur noch numerisch gelöst werden. Die Lösung $(x(t), y(t))$ beschreibt eine Kurve im $\mathbb{R}^2$. Durch numerische Berechnung der Lösungen zu verschiedenen Anfangsbedingungen kann man sich meist einen guten Überblick beschaffen. Auf diese Weise erhalten wir zum Beispiel Abbildung 11.9, die den Fixpunkt $(\frac{\gamma}{\delta}, \frac{\alpha}{\beta})$ zeigt, der von geschlossenen Lösungskurven umkreist wird.

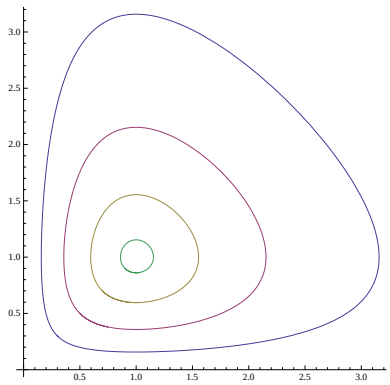


Abbildung 11.9. Lösungskurven der Volterra-Lotka Gleichungen.

Wie zuvor finden wir die Fixpunkte, indem wir nach konstanten Lösungen suchen, also Lösungen mit $(x'(t), y'(t)) = (0, 0)$. Die Fixpunkte sind also die Lösungen des Gleichungssystems $(\alpha - \beta y)x = 0$, $-(\gamma - \delta x)y = 0$. Wir erhalten die beiden Lösungen $(x, y) = (0, 0)$ und $(x, y) = (\frac{\gamma}{\delta}, \frac{\alpha}{\beta})$.

Die geschlossenen Lösungskurven entsprechen periodischen Lösungen: Wenn eine Lösung zu ihrem Ausgangspunkt zurückkehrt, wiederholt sich alles von vorne. Aus der Biologie ist dieses Verhalten wohl bekannt: Auf Zeiten mit vielen Beutetieren folgen Zeiten mit vielen Räubern und umgekehrt.

Dass es zu jeder Anfangsbedingung $(x(0), y(0)) = (x_0, y_0)$ eine eindeutige Lösung gibt, bedeutet übrigens, dass durch jeden Punkt im $\mathbb{R}^2$ genau eine Lösungskurve geht. Zwei verschiedene Lösungskurven können sich also niemals schneiden! In zwei Dimensionen bedeutet das, dass sich Lösungskurven gegenseitig behindern, da sie nicht aneinander vorbei können. Das erzwingt eine gewisse Regularität der Lösungskurven und bewirkt, dass zweidimensionale Systeme in der Regel gut verstanden werden können.

Befindet man sich aber im $\mathbb{R}^3$, so können die Lösungskurven nach oben und unten ausweichen, und das ermöglicht ein viel komplexeres Verhalten: Chaos kann sich ergeben.

Bei Rekursionen ist Chaos bereits in einer Dimension möglich, wie die diskrete logistische Gleichung zeigt.

Das dreidimensionale System

$$\begin{aligned} x'(t) &= -\sigma(x(t) - y(t)) \\ y'(t) &= r\,x(t) - y(t) - x(t)z(t) \\ z'(t) &= x(t)y(t) - b\,z(t) \end{aligned} \tag{11.11}$$

ist als **Lorenz-System** bekannt.

Es wurde erstmals vom amerikanischen Meteorologen Edward N. Lorenz (1917–2008) als einfaches Wettermodell untersucht. Er hat auch den Begriff des „**Schmetterlingseffekts**“ eingeführt, der die sensible Abhängigkeit der Lösung von den Anfangsbedingungen veranschaulicht: Bereits die kleinste Änderung in den Anfangsdaten (der Schlag eines Schmetterlings) ergibt eine Lösung, die nach einer gewissen Zeit weit entfernt von der ursprünglichen liegt.

Es ist ein stark vereinfachtes Modell für die Strömung einer Flüssigkeit zwischen zwei Platten mit verschiedener Temperatur. Eine Lösungskurve für die Parameter $\sigma = 10$, $r = 28$, $b = 8/3$ ist in Abbildung 11.10 dargestellt. Auf den ersten Blick sieht es nicht so schlimm aus, denn die Lösungskurve scheint mit einer gewissen Regelmäßigkeit die beiden eingezeichneten Fixpunkte zu umkreisen. Eine genauere Untersuchung zeigt aber, dass zum Beispiel die Anzahl der Umrundungen um den rechten oder auch den linken Fixpunkt völlig zufällig und ohne erkennbare Regelmäßigkeit erfolgt. Man kann zeigen, dass sich alle

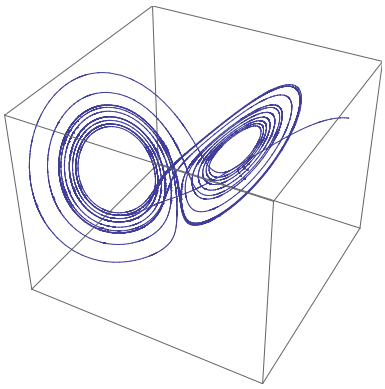


Abbildung 11.10. Lösungskurven der Lorenz-Gleichungen.

Lösungskurven immer mehr einer Menge annähern, die aber nur sehr schwer zu beschreiben ist. Je näher man dieser Menge kommt, umso komplizierter wird die Lösungskurve. Diese Menge wird deshalb **seltsamer Attraktor** genannt.

Ergänzung 1: In vielen Büchern werden oft nur Systeme von Differentialgleichungen erster Ordnung betrachtet. Der Fall höherer Ordnung kann nämlich immer auf diesen Fall zurückgeführt werden, indem man die höheren Ableitungen als neue Variable hinzufügt. Zum Beispiel kann die Gleichung

$x''(t) + ax'(t) + bx(t) = 0$ mit $x_1(t) = x(t), x_2(t) = x'(t), \vec{x}(t) = (x_1(t), x_2(t))$

als System zweier Differentialgleichungen erster Ordnung

$$\vec{x}'(t) = \begin{pmatrix} 0 & 1 \\ -b & -a \end{pmatrix} \vec{x}(t) \tag{11.12}$$

geschrieben werden.
Umgekehrt kann man auch ein System von n linearen Differentialgleichungen erster Ordnung mittels Ableiten und Substitutionen in eine einzige lineare Differentialgleichung n-ter Ordnung transformieren.

Ergänzung 2: Betrachtet man die Differentialgleichung $x'(t) = f(t, x(t))$ für $t \in I \subset \mathbb{R}$, so kann man jedem Punkt der t-x-Ebene eine Steigung $x' = f(t, x)$ zuordnen. Das Tripel $(t, x, f(t, x))$ nennt man Linienelement und die Gesamtheitheit aller Linienelemente heisst Richtungsfeld. Eine Lösung dieser Differentialgleichung "passt" genau auf das Richtungsfeld.

11.3. Kontrollfragen

Erklären Sie folgende Begriffe: Differentialgleichung, Ordnung, autonom, Anfangswerte, Lösung einer Differentialgleichung, Fixpunkt, asymptotisch stabil, Trennung (Separation) der Variablen, Variation der Konstanten.

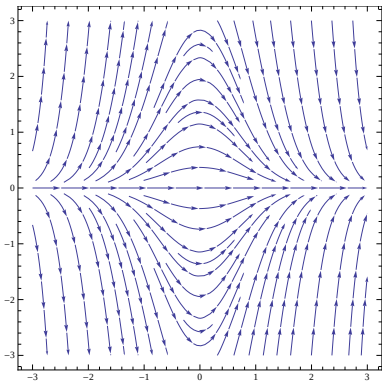


Abbildung 11.11. Richtungsfeld von $y' = -2\,x\,y$.

1. Welche Anfangswerte müssen vorgegeben werden, um die Lösung der folgenden Differentialgleichungen eindeutig zu bestimmen?
   a) $x''(t) + x(t) = 0$ b) $y'(x) = y(x)^2 - x$ c) $x'''(t) = 0$
2. Handelt es sich um eine autonome Differentialgleichung?
   a) $x''(t)+x(t) = 0$ b) $y'(x) = x\,\sin(y(x))$ c) $x''(t) = \frac{c_1}{t}x'(t)+\frac{c_2}{t^2}x(t)$
3. Finden Sie die Fixpunkte folgender Differentialgleichungen. Welche sind asymptotisch stabil?
   a) $x'(t) = 3x(t)$ b) $x'(t) = -\frac{1}{2}x(t)(1 - x(t))$
4. Ist die Differentialgleichung separierbar?
   a) $x'(t) = 3x(t) + t$ b) $x'(t) = tx(t)^2$

Erklären Sie folgende Begriffe: lineare Differentialgleichung, homogen/inhomogen, konstante/nicht-konstante Koeffizienten, Superpositionsprinzip, Wronskideterminante, charakteristische Gleichung.

5. Welche Form hat eine lineare, autonome, inhomogene Differentialgleichung zweiter Ordnung?
6. Klassifizieren Sie die Differentialgleichung (Ordnung? linear? homogen? konstante Koeffizienten?)
   a) $x''(t) = x'(t) + \sqrt{3}x(t)$ b) $x''(t) = x'(t)(1 - x(t))$ c) $y'(x) = (1.4)\,y(x) - \mathrm{e}^x$
   d) $y'(x) = \sqrt{x}\,y(x) + 2$ e) $x''(t) = x'(t)^2 + 7x(t)$ f) $x''''(t) = t^2x''(t) - x(t)$
7. Welche Lösung kommt für eine Differentialgleichung der Form $x''(t) = c_1x'(t) + c_0x(t)$ in Frage?
   a) $x(t) = \mathrm{e}^{2t}$ b) $x(t) = 3 \cdot \mathrm{e}^{2t+1}$ c) $x(t) = t^2$ d) $x(t) = \mathrm{e}^{2t} + 3^t$

11.4. Übung

(1) Welche Funktionen sind Lösungen der Differentialgleichung $x''(t) + x(t) = 0$?
a) $x(t) = \sin(t+1)$ b) $x(t) = t$ c) $x(t) = \cos(2t)$.

(2) Lösen Sie folgende Differentialgleichungen:
a) $x'(t) = x(t)^3$ b) $y'(x) = y(x) + 1$, $y(0) = 0$.

(3) Lösen Sie das Anfangswertproblem

$$x''(t) = 3x'(t) - 2x(t), \qquad x(0) = 0,\ x'(0) = 1.$$

(4) Lösen Sie die Differentialgleichung

$$x'(t) = \frac{x(t)}{t}.$$

(5) Lösen Sie das Anfangswertproblem

$$x'(t) = \frac{t}{x(t)}, \quad x(0) = 1.$$

(6) Lösen Sie das Anfangswertproblem

$$x'(t) = x(t) + \sin(t), \quad x(0) = 1.$$

(7) Eine Funktion mit konstanter Elastizität

$$\frac{f'(x)}{f(x)}x = c, \qquad x > 0,\ c \in \mathbb{R},$$

wird in der Wirtschaftsmathematik als **Cobb-Douglas-Funktion** bezeichnet. Was ist die allgemeinste Form einer Cobb-Douglas-Funktion?

(8) Nach dem **Newton'schen Abkühlungsgesetz** ist die Temperaturabnahme eines Körpers proportional zur Temperaturdifferenz mit der Umgebungstemperatur T_0:

$$T'(t) = -k(T_0 - T(t)).$$

Detective Horatio findet das Opfer um Mitternacht und stellt eine Körpertemperatur von $28°C$ fest. Eine halbe Stunde später sind es nur noch $26°C$. Wann wurde das Opfer ermordet, wenn die Umgebungstemperatur $18°C$ beträgt und die normale Körpertemperatur $37°C$ ist?

(9) Lösen Sie das Anfangswertproblem

$$x''(t) = -2x'(t) - 4x(t), \quad x(0) = 1,\ x'(0) = 2.$$

(10) Eine an ihren Enden aufgehängte Kette erfüllt im Gleichgewicht die Differentialgleichung

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

Finden Sie die Lösung mit Scheitel im Nullpunkt: $y(0) = 0$, $y'(0) = 0$. (Tipp: Setze $z(x) = y'(x)$.)

(11) Lösen Sie die Schwingungsgleichung

$$x''(t) + \omega^2 x(t) = 0, \qquad x(0) = x_0,\ x'(0) = x_1.$$

(12) (**Freier Fall mit Reibung**) Die Höhe eines Körpers, der nach unten fällt, wird durch

$$x''(t) = -\eta x'(t) - g, \quad \eta, g > 0,$$

beschrieben.

Chefarzt Dr. Frank Hofmann liegt vor seiner Klinik unter einer 10 Meter hohen Palme, als sich zum Zeitpunkt $t = 0$ eine Kokosnuss löst ($x(0) = 10$, $x'(0) = 0$). Finden Sie die Höhe $x(t)$ der Kokosnuss als Funktion der Zeit. Wie viel Zeit hat Dr. Hofmann ungefähr, bevor ihn die Kokosnuss erreicht, falls $\eta = 0.1$ und $g = 9.81$. (Tipp: Kleingedrucktes nach Satz 11.12.)

(13) Lösen Sie das Anfangswertproblem

$$x''(t) = 2x'(t) - x(t) - t, \qquad x(0) = 0,\ x'(0) = 1.$$

(14) Man löse das System

$$\vec{x}'(t) = A\,\vec{x}(t) + \vec{b}, \qquad \vec{x}(0) = \vec{x}_0$$

(a) $A = \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}$, $\vec{x}_0 = (1, 2)$, $\vec{b} = (12, 12)$

(b) $A = \begin{pmatrix} -3 & 2 \\ 1 & -4 \end{pmatrix}$, $\vec{x}_0 = (1, 2)$, $\vec{b} = (15, 15)$

(c) $A = \begin{pmatrix} -3 & -1 \\ 2 & -1 \end{pmatrix}$, $\vec{x}_0 = (1, 2)$, $\vec{b} = (10, 5)$

Wie verhalten sich die Lösungen für $t \to \infty$?

Anhang A

Quadratische Gleichung, Polynomdivision

In diesem und den kommenden Kapiteln, die im wesentlichen Schulstoff wiederholen, verzichten wir auf strenge mathematische Formulierungen.

Eine Gleichung der Form

$$a\,x^2 + b\,x + c = 0, \qquad a \neq 0$$

nennt man *quadratische Gleichung*. Durch Division durch a erhält man die Standardform

$$x^2 + p\,x + q = 0. \tag{A.1}$$

Umformen ergibt

$$\begin{aligned} x^2 + p\,x + q = x^2 + p\,x + \frac{p^2}{4} - \frac{p^2}{4} + q = (x + \frac{p}{2})^2 - \frac{p^2}{4} + q = 0 \\ (x + \frac{p}{2})^2 = \frac{p^2}{4} - q \\ (x + \frac{p}{2}) = \pm\sqrt{\frac{p^2}{4} - q} \\ x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}. \end{aligned} \tag{A.2}$$

Den Ausdruck $D = \frac{p^2}{4} - q$ nennt man *Diskriminate*. Ist $D < 0$ so hat Gleichung (A.1) keine reellen Lösungen. Sind x_1, x_2 Lösungen von Gleichung (A.1),

so gilt

$$x^2 + p\,x + q = (x - x_1)(x - x_2) = x^2 - (x_1 + x_2)\,x + x_1\,x_2 = 0. \qquad \text{(A.3)}$$

Das heißt, es gilt: $p = -(x_1 + x_2)$ und $q = x_1\,x_2$.

Man kann die Lösungen von Gleichung (A.1) als die Nullstellen der Funktion $f : \mathbb{R} \to \mathbb{R}, \quad f(x) = x^2 + p\,x + q$ interpretieren.

Für Gleichungen dritter Ordnung, zum Beispiel

$$x^3 - 15x - 4 = 0,$$

gibt es zwar eine Lösungsformel (Lösungsformel von Cardano)[1]

$$x_1 = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}} = 4,$$

aber oft kann durch Erraten einer Lösung und *Polynomdivision* die Gleichung dritter Ordnung auf eine quadratische Gleichung reduziert werden

$$\begin{array}{l} x^3 + 0\,x^2 - 15\,x - 4 : (x - 4) = x^2 + 4\,x + 1. \\ \underline{x^3 - \ 4\,x^2} \\ \qquad 4\,x^2 - 15\,x - 4 \\ \qquad \underline{4\,x^2 - \ 16\,x} \\ \qquad\qquad\quad x - 4 \\ \qquad\qquad\quad \underline{x - 4} \\ \qquad\qquad\qquad\ \ 0 \end{array}$$

Im dem Fall, dass man eine Lösung abspalten kann, muss sich die Polynomdivision immer ausgehen. Dividiert man zwei beliebige Polynome erhält man meist einen Rest.

[1] Man beachte, dass in diesem Fall der Gebrauch von komplexen Zahlen auf natürliche Weise nahegelegt wird.

Anhang B

Komplexe Zahlen

B.1. Die Menge der komplexen Zahlen

Die quadratische Gleichung $x^2 + 1 = 0$ hat keine Lösung in der Menge der reellen Zahlen. Eine solche Lösung würde man als die Wurzel aus (-1) bezeichnen können. Um auch solche Gleichungen lösen zu können, führt man die komplexen Zahlen ein. Augenscheinlich von Bedeutung ist dies für Gleichungen dritter Ordnung. Bei diesen führt die Lösungsformel von Cardano auf komplexe Ausdrücke, die sich schlussendlich auf eine reelle Lösung, die mindestens existiert, reduzieren.[1]

Man fügt zunächst der Menge der reellen Zahlen ein neues Element hinzu $\mathbb{R} \cup \{\mathrm{i}\}$, das mit dem Symbol i bezeichnet wird und "*imaginäre Einheit*" genannt wird.

Definition B.1. *Für die Zahl* i *gilt*

$$\mathrm{i}^2 = -1 \ . \tag{B.1}$$

Man schreibt daher $\mathrm{i} = \sqrt{-1}$*. Ausdrücke von der Form* $x + y\,\mathrm{i}$*, wobei* x *und* y *reelle Zahlen sind, nennt man* komplexe Zahlen*. Die Menge aller komplexer Zahlen bezeichnet man mit* $\mathbb{C}$*.*

Eine komplexe Zahl $z \in \mathbb{C}$ wird also durch zwei reelle Zahlen x und y definiert. Man kann daher eine komplexe Zahl als ein "geordnetes Paar" von reellen Zahlen auffassen, d.h. als einen Punkt des zweidimensionalen Raumes (= xy-Ebene). Eine komplexe Zahl wird daher als ein Punkt der "komplexen Ebene"

[1] Rafael Bombelli verwendete bereits 1569 komplexe Zahlen zur Lösung von $x^3 = ax + b$, `http://en.wikipedia.org/wiki/Rafael_Bombelli`

veranschaulicht. Der Punkt mit den Koordinaten (x, y) stellt die komplexe Zahl $z = x + y\,\mathrm{i}$ auf eindeutige Weise dar.

Beim Rechnen mit komplexen Zahlen geht man genauso vor wie bei reellen Zahlen. Man definiert demnach folgende Rechenregeln:

(1) **Addition**:

$$(x + y\ \mathrm{i}) + (u + v\ \mathrm{i}) = (x + u) + (y + v)\ \mathrm{i}\ .$$

(2) **Multiplikation**:

$$\begin{aligned}(x + y\ \mathrm{i})(u + v\ \mathrm{i}) &= xu + xv\ \mathrm{i} + yu\ \mathrm{i} + yv\ \mathrm{i}^2 \\ &= (xu - yv) + (xv + yu)\ \mathrm{i}\,.\end{aligned}$$

Damit bilden die komplexen Zahlen einen Körper. Potenzen von i können reduziert werden, da wegen $\mathrm{i}^2 = -1$, $\mathrm{i}^{2n} = (-1)^n, \mathrm{i}^{2n+1} = (-1)^n\ \mathrm{i}$ gilt.

Hier kommen nun einige grundlegende Bezeichnungen:

Definition B.2. Einige Begriffe für komplexe Zahlen:

Für eine komplexe Zahl $z = x + y\ \mathrm{i}$ heißt

$x = \mathrm{Re}\, z$	*Realteil von z*
$y = \mathrm{Im}\, z$	*Imaginärteil von z*
$\overline{z} = x - y\ \mathrm{i}$	*konjugiert komplexe Zahl*
$\lvert z\rvert = \sqrt{x^2 + y^2}$	*Absolutbetrag von z*

Die reellen Zahlen können als Teilmenge der komplexen Zahlen aufgefaßt werden, nämlich als diejenigen Zahlen, deren Imaginärteil Null ist. Reelle Zahlen bestehen nur aus einem Realteil. Statt $x + 0\ \mathrm{i}$ schreibt man daher einfach x.

$$z \text{ ist reell, wenn} \quad z = x + 0\ \mathrm{i}\ = x\ ,$$

z.B. ist $5 + 0\,\mathrm{i} = 5$. Zahlen, die nur aus Imaginärteil bestehen, $z = 0 + y$ i, nennt man *imaginäre Zahlen*. Eine imaginäre Zahl ist z.B. $0 + \pi\ \mathrm{i} = \pi\ \mathrm{i}$. Die komplexe Zahl $0 + 0$ i kann man auch als 0 oder als 0 i schreiben.

Hier einige Beispiele:

(1) $\mathrm{i}(-\mathrm{i}) = 1$, $\sqrt{-16} = 4\ \mathrm{i}$, $-\sqrt{-9} = -3\ \mathrm{i}$, $\mathrm{i}^3 = -\mathrm{i}$.

(2) $(1 + \mathrm{i})(1 - \mathrm{i}) = 2$.

(3) $\lvert 4 + 3\,\mathrm{i}\,\rvert = \sqrt{16 + 9} = 5$.

Brüche: Komplexe Brüche werden berechnet, indem man zuerst mit dem konjugiert komplexen Nenner erweitert. Man erhält dann einen reellen Nenner.

$$\frac{x+y\,\mathrm{i}}{u+v\,\mathrm{i}} = \frac{x+y\,\mathrm{i}}{u+v\,\mathrm{i}} \cdot \frac{u-v\,\mathrm{i}}{u-v\,\mathrm{i}} = \frac{(xu+yv)+(yu-xv)\,\mathrm{i}}{u^2+v^2}$$
$$= \frac{xu+yv}{u^2+v^2} + \frac{yu-xv}{u^2+v^2}\,\mathrm{i},$$

$$(x+y\,\mathrm{i})(x-y\,\mathrm{i}) = x^2+y^2\ , \qquad \text{also:} \quad z\overline{z} = |z|^2\ . \tag{B.2}$$

Beispiel:

$$\frac{2+\mathrm{i}}{3-\mathrm{i}} = \frac{(2+\mathrm{i})(3+\mathrm{i})}{(3-\mathrm{i})(3+\mathrm{i})} = \frac{1}{2} + \frac{1}{2}\,\mathrm{i}\,.$$

Für den Absolutbetrag einer komplexen Zahl ergeben sich folgende Regeln: Seien z_1 und z_2 zwei komplexe Zahlen. Dann gilt

$$|z_1 z_2| = |z_1||z_2|\ , \qquad \left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}\ , \quad (z_2 \neq 0)\ , \tag{B.3}$$

$$|z_1+z_2| \leq |z_1|+|z_2|\ , \qquad |z_1+z_2| \geq \big||z_1|-|z_2|\big|\ . \tag{B.4}$$

Diese Regeln sind diesselben wie beim Rechnen mit Absolutbeträgen von reellen Zahlen.

Zwei komplexe Zahlen sind gleich, wenn sie denselben Realteil und denselben Imaginärteil haben.

Beispiel B.3. *Die Gleichung* $z^2 = 2\,\mathrm{i}$ *entspricht* $x^2+2xy\,\mathrm{i}-y^2 = 2\,\mathrm{i}$ *und ergibt daher zwei reelle Gleichungen:*

$$\textit{Realteil:} \quad x^2-y^2 = 0\ , \qquad \textit{Imaginärteil:} \quad xy = 1\ . \tag{B.5}$$

Dieses Gleichungssystem hat zwei Lösungen: $(x,y) = (1,1)$ *und* $(x,y) = (-1,-1)$.

Wenn man komplexe Zahlen durch Punkte (x,y) in der komplexen Ebene veranschaulicht, dann bilden die reellen Zahlen die x-Achse, die imaginären Zahlen die y-Achse.

Übungen:

(1) Zeichnen Sie folgende komplexe Zahlen: $1+\mathrm{i}$, $-\sqrt{3}+\mathrm{i}$.

(2) Berechnen und zeichnen Sie: $\frac{1}{1+\mathrm{i}}$, $\frac{1}{1-\mathrm{i}}$, $1/z^2$ für $z = 2-3\,\mathrm{i}$.

(3) Berechnen Sie den Betrag von $\frac{5-2\,\mathrm{i}}{5+2\,\mathrm{i}}$, $\frac{3\,\mathrm{i}}{\mathrm{i}-\sqrt{3}}$, $z/\overline{z}$, $(1+2\,\mathrm{i})^3$.

(4) Lösen Sie für die reellen Zahlen x und y die Gleichungen $x+iy = y+x\,\mathrm{i}$, $(x+y\,\mathrm{i})^2 = 2\,\mathrm{i}\,x$, $\frac{x+y\,\mathrm{i}}{x-y\,\mathrm{i}} = -\,\mathrm{i}$.

B.2. Polarkoordinaten

In der Abbildung B.1 sieht man einen Punkt (x, y) der komplexen Ebene, der die Zahl $z = x + y$ i repräsentiert.

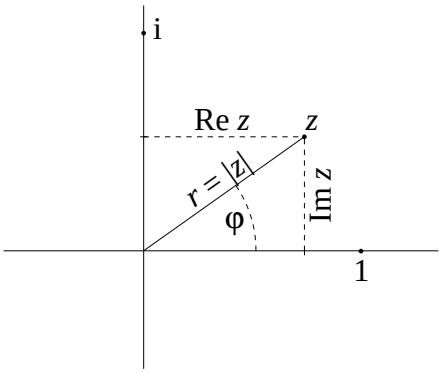


Abbildung B.1. Graphische Darstellung einer komplexen Zahl z in kartesischen und in polaren Koordinaten.

Man zeichnet die Verbindungstrecke vom Koordinatenursprung (= die komplexe Zahl $0 = 0 + 0$ i), zum Punkt z. Für die Länge r dieser Strecke und für den Winkel φ, den diese Verbindungslinie mit der (positiven) x-Achse einschließt, gelten folgende Beziehungen:

$$x = r\,\cos\varphi\ , \tag{B.6}$$

$$y = r\,\sin\varphi\ , \tag{B.7}$$

$$z = x + y\ \mathrm{i} = r\,(\cos\varphi + \sin\varphi\ \mathrm{i})\ , \tag{B.8}$$

$$r = \sqrt{x^2 + y^2} = |z|\ . \tag{B.9}$$

Die Darstellung (B.8) heißt "*Polardarstellung*" der komplexen Zahl z. Den Winkel φ bezeichnet man als das "*Argument*" der komplexen Zahl,

$$\varphi = \arg z\ . \tag{B.10}$$

Aus einer gegebenen Zahl $z = x+y$ i berechnet man das Argument φ am besten aus der Beziehung

$$\tan\varphi = \frac{y}{x}\ . \tag{B.11}$$

Für $-\pi/2 \le \varphi \le \pi/2$ ergibt sich zum Beispiel $\varphi = \arctan(y/x)$. In den anderen Winkelbereichen muß man allerdings mit der Definition der Umkehrfunktion arctan aufpassen. Man erhält:

$$-\pi/2 \le\varphi \le \pi/2 : \qquad \varphi = \arctan(y/x)\ , \tag{B.12}$$

$$\pi/2 < \varphi \leq 3\pi/2 : \qquad \varphi = \pi - \arctan(y/x) \ . \tag{B.13}$$

Unter Benützung der Polardarstellung kann man das Produkt komplexer Zahlen besonders einfach beschreiben.

Sei

$$z_1 = r_1(\cos\varphi_1 + \sin\varphi_1 \ \mathrm{i}) \ , \qquad z_2 = r_2(\cos\varphi_2 + \sin\varphi_2 \ \mathrm{i}) \ . \tag{B.14}$$

Dann gilt

$$\begin{aligned} z_1 z_2 &= r_1 r_2 \, (\cos\varphi_1 + \sin\varphi_1 \ \mathrm{i}) \, (\cos\varphi_2 + \sin\varphi_2 \ \mathrm{i}) \\ &= r_1 r_2 \left[(\cos\varphi_1 \ \cos\varphi_2 - \sin\varphi_1 \ \sin\varphi_2) + (\sin\varphi_1 \ \cos\varphi_2 + \cos\varphi_1 \ \sin\varphi_2) \ \mathrm{i} \right] \\ &= r_1 r_2 \left[\cos(\varphi_1 + \varphi_2) + \sin(\varphi_1 + \varphi_2) \ \mathrm{i} \right] \ . \end{aligned} \tag{B.15}$$

Definition B.4. Graphische Interpretation des Produktes zweier komplexer Zahlen*:*

Bildet man das Produkt zweier komplexer Zahlen, so werden die Beträge multipliziert und die Polarwinkel addiert.

Wenn man eine komplexe Zahl wiederholt mit sich selbst multipliziert (potenziert), erhält man die *Formel von Moivre*:

$$z^n = r^n \left[\cos(n\varphi) + \mathrm{i} \, \sin(n\varphi) \right] \ . \tag{B.16}$$

B.3. Eulersche Formel

Für jeden Winkel φ (im Bogenmaß) betrachtet man nun die komplexe Zahl

$$E(\varphi) = \cos\varphi + \mathrm{i} \, \sin\varphi \tag{B.17}$$

$E(\varphi)$ hat immer den Absolutbetrag 1. Die Menge der komplexen Zahlen mit Betrag 1 liegen in der komplexen Ebene auf den sogenannten "*Einheitskreis*".

Offensichtlich gilt (wegen $\cos 0 = 1$, $\sin 0 = 0$)

$$E(0) = 1 \ . \tag{B.18}$$

Für das Produkt von Zahlen auf dem Einheitskreis liefern die Betrachtungen des vorigen Abschnitts die Formeln:

$$E(\varphi_1) \, E(\varphi_2) = E(\varphi_1 + \varphi_2) \ , \quad \text{bzw.} \quad \left[E(\varphi) \right]^n = E(n\varphi) \ , \tag{B.19}$$

(Formel von Moivre). Diese Formeln legen einen Vergleich der Funktion $E(\varphi)$ mit der Exponentialfunktion nahe. Schreibt man

$$A(\varphi) = a^\varphi = (e^{\log a})^\varphi = e^{k\varphi} \ , \quad (\text{wobei } k = \log a), \tag{B.20}$$

so gelten ja ebenfalls die Formeln

$$A(0) = e^0 = 1 \ ,$$

$$A(\varphi_1)\,A(\varphi_2) = e^{k\varphi_1}\,e^{k\varphi_2} = e^{k(\varphi_1+\varphi_2)} = A(\varphi_1+\varphi_2)\ ,$$
$$\big[A(\varphi)\big]^n = (e^{k\varphi})^n = e^{k\varphi n} = A(n\varphi)\ .$$

Die Funktion $A(\varphi)$ hat also diesselben Eigenschaften wie die Funktion $E(\varphi)$. Bildet man nun die Ableitung von $E(\varphi)$ nach φ ergibt:

$$\frac{d}{d\varphi}\,E(\varphi) = -\sin\varphi + \mathrm{i}\,\cos\varphi = \mathrm{i}\,(\cos\varphi + \mathrm{i}\,\sin\varphi) = \mathrm{i}\,E(\varphi)\ .$$

Andererseits gilt mit der Kettenregel:

$$\frac{d}{d\varphi}\,A(\varphi) = k\,e^{k\varphi} = kA(\varphi)\ .$$

Perfekte formale Übereinstimmung zwischen den Funktionen A und E besteht also, wenn $k = i$ gesetzt wird. Man benützt diese Übereinstimmung als Motivation für die Definition der komplexen Exponentialfunktion $e^{\mathrm{i}\,\varphi} = E(\phi)$.

Definition B.5. Eulersche Formel*:*

$$e^{\mathrm{i}\,\varphi} = \cos\varphi + \mathrm{i}\,\sin\varphi \tag{B.21}$$

Man betrachtet die Eulersche Formel als Definition der Exponentialfunktion für imaginäre Argumente.

Anmerkung: Alternatives Argument: Differenzieren der Eulerschen Formel zeigt, dass beide Seiten die gleichen Ableitungen haben und die Funktionen daher bis auf eine Integrations-Konstante gleich sind. Setzt man $\varphi = 0$ so erhält man, dass diese Konstante 0 ist.

Die Polarform einer komplexen Zahl $z = x + iy$ lautet nun ganz einfach

$$z = r\,e^{\mathrm{i}\,\varphi} \tag{B.22}$$

Es gilt:

$$e^{-\mathrm{i}\,\varphi} = \cos(-\varphi) + \mathrm{i}\,\sin(-\varphi) = \cos\varphi - \mathrm{i}\,\sin\varphi \tag{B.23}$$

und daher

$$\overline{z} = r\,e^{-\mathrm{i}\,\varphi}\ . \tag{B.24}$$

Man kann nun auch trigonometrische Funktionen durch Exponentialfunktionen ausdrücken, denn

$$e^{\mathrm{i}\,\varphi} + e^{-\mathrm{i}\,\varphi} = 2\cos\varphi, \qquad e^{\mathrm{i}\,\varphi} - e^{-\mathrm{i}\,\varphi} = 2\,\mathrm{i}\sin\varphi\ . \tag{B.25}$$

Vergleiche die folgenden Ausdrücke mit der Definition der Hyberbelfunktionen sinh und cosh:

$$\cosh\varphi = \frac{e^{\varphi} + e^{-\varphi}}{2}\ , \tag{B.26}$$

$$\sinh\varphi = \frac{e^{\varphi} - e^{-\varphi}}{2}\ . \tag{B.27}$$

B.4. Einheitswurzeln

Zwei komplexe Zahlen, deren Absolutbeträge gleich sind, deren Argumente sich aber um 2π (also 360°) unterscheiden, sind gleich. Allgemein bezeichnen die Winkel φ und $\varphi + 2\pi k$ dieselbe Richtung in der komplexen Ebene (wobei k eine beliebige ganze Zahl ist). Es gilt also

$$z = r\, e^{\mathrm{i}\,\varphi} = r\, e^{\mathrm{i}(\varphi + 2k\pi)} \tag{B.28}$$

Wir wollen nun die Gleichung

$$z^n = a + \mathrm{i}\; b \tag{B.29}$$

lösen. Dazu schreibt man $a + \mathrm{i}\; b$ in der Form $R\, e^{\mathrm{i}\,\alpha}$. Gesucht sind alle komplexen Zahlen, deren n-te Potenz $a + \mathrm{i}\; b$ ergibt. In Polarform lautet die Gleichung

$$r^n\, e^{\mathrm{i}\, n\varphi} = R\, e^{\mathrm{i}\,\alpha}\ . \tag{B.30}$$

Also müssen die Beträge rechts und links gleich sein, die Argumente können sich aber um ganzzahlige Vielfache von 2π unterscheiden:

$$r^n = R\ , \qquad n\varphi = \alpha + 2k\pi\ , \quad k = 0, \pm 1, \pm 2, \ldots \tag{B.31}$$

Daraus kann man nun ganz leicht den Betrag und die in Frage kommenden Argumente bestimmen:

$$r = R^{1/n}\ , \qquad \varphi = \frac{\alpha}{n} + 2\frac{k}{n}\pi\ , \quad k = 0, \pm 1, \pm 2, \ldots \tag{B.32}$$

Das Argument ist also nicht eindeutig bestimmt. Verschiedene Lösungen erhält man allerdings nur für $k = 0, 1, 2, \ldots, n-1$. Die Lösung für $k = n$ ist diesselbe wie für $k = 0$, etc.

Definition B.6. Zusammenfassung: *Die Gleichung*

$$z^n = R\, \mathrm{e}^{\mathrm{i}\,\alpha} \tag{B.33}$$

hat n verschiedene Lösungen $z_0, z_2, \ldots, z_{n-1}$, die gegeben sind durch

$$z_k = R^{\frac{1}{n}}\, \mathrm{e}^{\mathrm{i}\,\varphi_k} \tag{B.34}$$

mit

$$\varphi_k = \frac{\alpha}{n} + k\,\frac{2\pi}{n}, \qquad k = 0, 1, 2, \ldots, n-1. \tag{B.35}$$

Der Spezialfall $a + b\,\mathrm{i} = 1$ führt auf die Gleichung

$$z^n = 1\ . \tag{B.36}$$

Ihre Lösungen werden n-te Einheitswurzeln genannt. Sie liegen alle auf dem Einheitskreis in der komplexen Ebene. In Polarform lauten die Einheitswurzeln gemäß (B.34) und (B.35)

$$z_k = e^{2\,\mathrm{i}\,k\pi/n} \;, \quad k = 0, 1, 2, \ldots, n-1 \;. \tag{B.37}$$

Da die Argumente von z_k und z_{k-1} sich immer um denselben Betrag $2\pi/n$ unterscheiden, bilden die n-ten Einheitswurzeln die Eckpunkte eines regelmäßigen n-Ecks.

Vektorrechnung im $\mathbb{R}^2$ und $\mathbb{R}^3$

C.1. Skalare und Vektoren

Man kann die Begriffe Skalar und Vektor am besten anhand von physikalischen Beispielen veranschaulichen. Eine skalare Größe ist eine Größe, die durch eine einzige reelle Zahl bestimmt wird. Beispiele dafür sind: die Masse eines Körpers, das Volumen eines Körpers, die Temperatur, der Widerstand usw. Aber meistens reicht eine einzige Zahl für die Beschreibung eine Größe nicht aus. In der Mechanik zur Charakterisierung einer Kraft etwa verwendet man einen Vektor. Ein Vektor entspricht einer gerichteten Strecke, die nicht nur die Stärke angeben kann, sondern auch die Richtung in der die Kraft wirkt. Man fasst daher einen Vektor als eine Äquivalenzklasse von parallelen gleichlangen gerichteten Strecken (Pfeilen) auf.

Im dreidimensionalen Raum kann die Lage eines Punktes bzw. eines Vektors durch drei reelle Zahlen, ein 3-Tupel, beschrieben werden (= Koordinaten, Komponenten). Die Geometrie des $\mathbb{R}^3$ ist unter anderem für Anwendungen in der Computergraphik von grosser Bedeutung. Aber es ist auch sinnvoll, 4-Tupel, 5-Tupel, usw., also allgemein n-Tupel zu betrachten: So kann man zum Beispiel den netto Tagesumsatz von 12 Filialen in einem 12-Tupel zusammenfassen. Um den Wochenumsatz der 12 Filialen zu erhalten, muss man die 12-Tupel koordinatenweise addieren. Um die Mehrwertsteuer zu erhalten, muss man jede Komponente mit 0.2 (20% Mehrwertsteuer) multiplizieren. Es ist also sinnvoll, allgemein n-Tupeln zu betrachten und dafür Rechenoperationen zu definieren.

Für 2- oder 3-Tupel lassen sich diese Operationen auch geometrisch veranschaulichen. Für allgemeine n-Tupel ist das nicht möglich, trotzdem ist die geometrische Anschauung für den Spezialfall $n = 2$ oder $n = 3$ oft der Schlüssel zur Lösung komplizierter Probleme.

C.1.1. Komponenten eines Vektors. Es sei $\vec{a}$ ein Vektor in einem kartesischen Koordinatensystem, der durch eine gerichtete Strecke $\overline{PQ}$, wobei P der Anfangspunkt und Q der Endpunkt ist, gegeben ist. Die Koordinaten vom Punkt P sind (x_1, y_1), die vom Punkt Q sind (x_2, y_2). Die Komponenten des Vektors $\vec{a}$ erhält man indem man von der „Spitze“ (dem Endpunkt Q), den „Schaft“ (den Anfangspunkt P), subtrahiert.

$$\overrightarrow{PQ} = \vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \end{pmatrix}. \tag{C.1}$$

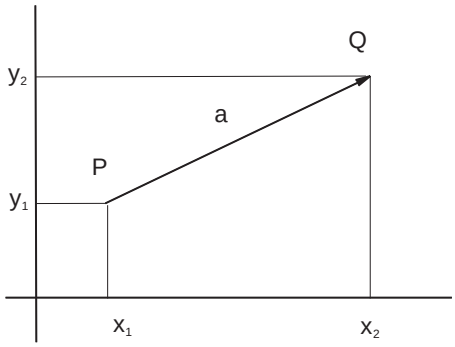


Abbildung C.1. Vektor

Die Länge (**Norm**) des Vektors $||\vec{a}||$ ist der Abstand der Punkte P und Q voneinander und ist daher (Pythagoras)

$$||\vec{a}|| = \sqrt{a_1^2 + a_2^2}. \tag{C.2}$$

Beispiel 1:

Der Vektor $\vec{a}$ hat den Anfangspunkt $P = (1, 2)$ und Endpunkt $Q = (3, 4)$. Die Komponenten sind daher

$$a_1 = 3 - 1 = 2, \qquad a_2 = 4 - 2 = 2, \qquad \vec{a} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \tag{C.3}$$

und die Länge von $\vec{a}$ ist

$$||\vec{a}|| = \sqrt{2^2 + 2^2} = \sqrt{8}. \tag{C.4}$$

Ortsvektor:

Ein Ortsvektor hat seinen Anfangspunkt im Koordinatenursprung $O = (0, 0)$. Daraus folgt, daß die Koordinaten des Endpunkts gleich die Komponenten des Vektors ergeben.

Nullvektor:

Bei einem Nullvektor $\vec{0}$ fallen der Anfangs- und Endpunkt zusammen. Der Nullvektor hat die Länge Null und eine unbestimmte Richtung.

C.1.2. Addition von Vektoren und Multiplikation mit Skalaren.

Addition von Vektoren: Die Vektoraddition ist eine Basisrechenoperation in der Vektorrechnung. Wenn man zwei Vektoren $\vec{a}$ und $\vec{b}$ addieren will, muß man die einzelnen Komponenten miteinander addieren. Der Anfangspunkt des daraus folgenden Vektors $\vec{c}$ ist der Anfangspunkt von $\vec{a}$ und der Endpunkt von $\vec{c}$ ist der Endpunkt von $\vec{b}$. Graphisch werden die zwei Pfeile aneinandergehängt.

$$\vec{c} = \vec{a} + \vec{b}, \qquad \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}. \tag{C.5}$$

Für die Vektoraddition gilt das Kommutativgesetz. Das heißt, daß

$$\vec{a} + \vec{b} = \vec{b} + \vec{a}. \tag{C.6}$$

Die Vektoraddition erfüllt auch das Assiozativgesetz welches aussagt, daß

$$(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c}). \tag{C.7}$$

Zu jedem Vektor a gibt es einen inversen Vektor $-\vec{a}$, sodaß

$$\vec{a} + (-\vec{a}) = \vec{0} \tag{C.8}$$

gilt. Es gilt auch

$$\vec{a} + \vec{0} = \vec{a}. \tag{C.9}$$

Zusammengefasst kann man sagen, daß die Vektoren bezüglich der Vektoraddition eine kommutative Gruppe bilden.

Beispiel Vektoraddition:

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2+1 \\ 3+2 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}.$$

Multiplikation mit Skalaren

Man kann einen Vektor $\vec{a}$ mit mit einem Skalar λ, also einer reellen Zahl, multiplizieren. Die Komponenten des Vektors $\lambda\vec{a}$ lauten $(\lambda a_1, \lambda a_2)$. Bei der Multiplikation mit Skalaren λ, μ gelten die folgenden Gesetze:

$$\begin{aligned} &\text{(i)} \ \lambda(\mu\vec{a}) = (\lambda\mu)\vec{a} \\ &\text{(ii)} \ 1\ \vec{a} = \vec{a} \\ &\text{(iii)} \ \lambda(\vec{a} + \vec{b}) = \lambda\vec{a} + \lambda\vec{b} \\ &\text{(iv)} \ (\lambda + \mu)\vec{a} = \lambda\vec{a} + \mu\vec{a}. \end{aligned}$$

Wenn der Skalar eine negative reelle Zahl ist, wird die Richtung (Orientierung) des Vektors umgekehrt.

Einheitsvektor:

Der Einheitsvektor $\vec{e}$ ist ein Vektor dessen Betrag (Länge) $||\vec{e}||$ gleich 1 ist. Zu jedem Vektor $\vec{b}$ kann ein Einheitsvektor mit der gleichen Richtung angegeben werden

$$\begin{aligned} ||\vec{e}|| &= 1, \\ \vec{e}_{\vec{b}} &= \frac{1}{||\vec{b}||}\vec{b}. \end{aligned} \tag{C.10}$$

Beispiel:

$$\vec{a} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \quad \vec{e}_{\vec{a}} = \frac{1}{||\vec{a}||}\vec{a} = \frac{1}{\sqrt{3^2 + 4^2}} \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 3/5 \\ 4/5 \end{pmatrix}.$$

C.2. Vektorräume

Ein Vektorraum oder auch linearer Raum besteht aus der Menge aller Elemente die er beinhaltet. Die Definition eines reellen Vektorraumes lautet: Gegeben sei eine Menge deren Elemente sämtliche Rechenregeln der Vektoraddition und der Multiplikation mit einem Skalar wie im vorhergehenden Kapitel befolgen. Diese Elemente bilden dann einen Vektorraum und heißen Vektoren. Statt der reellen Zahlen können auch in analoger Weise Vektorräume über den komplexen Zahlen definiert werden.

C.2.1. Lineare Abhängigkeit und Unabhängigkeit:

Die Vektoren $\vec{a}_1, \ldots, \vec{a}_m$ nennt man linear unabhängig, wenn gilt

$$\text{aus} \ \ \lambda_1\vec{a}_1 + \lambda_2\vec{a}_2 + \ldots + \lambda_m\vec{a}_m = 0 \ \ \text{folgt} \ \ \lambda_1 = \lambda_2 = \ldots \lambda_m = 0. \tag{C.11}$$

Beispiel: Zeige, dass die Vektoren $\vec{a} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$ und $\vec{b} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ linear unabhängig sind. Wie man sieht läuft dies auf die Frage der Eindeutigkeit der Lösung eines linearen Gleichungssystems hinaus.

In einer linear abhängigen Menge von Vektoren eines Vektorraumes kann man mindestens einen der Vektoren durch eine lineare Kombination der anderen darstellen. Bei einer linear unahängigen Menge funktioniert das nicht. Die maximale Anzahl n von linear unabhängigen Vektoren eines Vektorraumes nennt man seine Dimension n und eine Menge von n linear unabhängigen Vektoren eine Basis des Vektorraumes. Jeder Vektor kann dann durch eine lineare Kombination der Basisvektoren gebildet werden.

C.2.2. Die vektorielle Projektion eines Vektors auf eine Zweiten.

Gegeben seien zwei Vektoren $\vec{a}$ und $\vec{b}$. Die vektorielle Projektion des Vektors $\vec{b}$ auf den Vektor $\vec{a}$ ist der Vektor $\vec{b'}$ (Abbildung C.2) der parallel zu $\vec{a}$ ist und dessen Länge durch die orthogonale Projektion von $\vec{b}$ auf $\vec{a}$ gegeben ist.

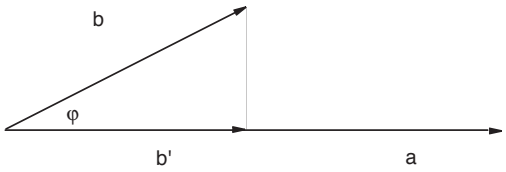


Abbildung C.2. Vektorielle Projektion

Für die Länge des Vektors $\vec{b'}$ erhält man

$$||\vec{b'}|| = ||\vec{b}|| \cos\varphi, \tag{C.12}$$

wobei φ der eingeschlossene Winkel ist.

Beispiel: Der Höhenvektor $\vec{h}_{\vec{c}}$ eines Dreiecks ist gleich der Projektion des Seitenvektors $\vec{a}$ (oder $-\vec{b}$) auf den Normalseitenvektor $\vec{e}_{\vec{c}}^{\perp}$ von $\vec{c}$ (Abbildung C.3).

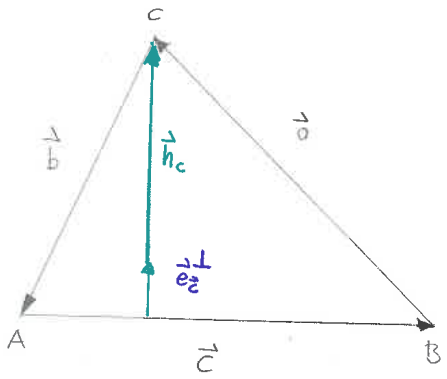


Abbildung C.3. Der Höhenvektor $\vec{h}_{\vec{c}}$ eines Dreiecks als Projektion des Seitenvektors.

C.2.3. Skalares (Inneres) Produkt zweier Vektoren.

In der Physik ist die geleistete Arbeit W (= Energie [1 kg m/s = 1 W s]) definiert durch Kraft F *mal* Weg s. Im allgemeinen sind aber Kraft und Weg Vektoren und es ist daher sinnvoll ein Produkt $\langle \vec{F}, \vec{s} \rangle$ dieser beiden Vektorgrößen so zu definieren, dass sich die geleistete Arbeit als Skalar ergibt.

Dazu überlegt man sich, dass man den Kraftvektor $\vec{F}$ in einen Anteil zerlegen kann, der parallel zum Weg ist und den Rest, der normal dazu steht, also

$$\vec{F} = \vec{F}_{\parallel \vec{s}} + \vec{F}_{\perp \vec{s}}\,. \tag{C.13}$$

Der Vektor $\vec{F}_{\parallel \vec{s}}$ ist aber nichts anderes als die vektorielle Projektion $\vec{F}'$ des Vektors $\vec{F}$ auf den Vektor $\vec{s}$ (Abbildung C.4). Nur der Anteil der Kraft parallel

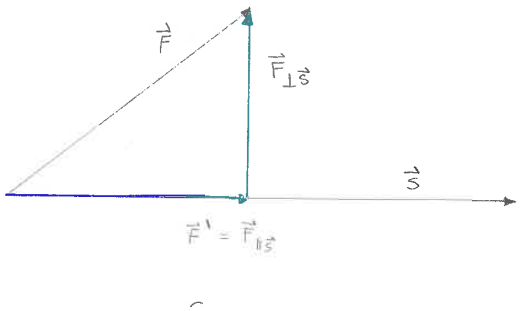


Abbildung C.4. Das skalare Produkt

zum Weg trägt aber zur geleisteten Arbeit bei. Das skalare Produkt der zwei Vektoren $\vec{F}$ und $\vec{s}$ wird definiert als Länge (Betrag) der vektoriellen Projektion $\vec{F}'$ von $\vec{F}$ auf $\vec{s}$ *mal* Länge des Vektors $\vec{s}$.

$$W = \langle \vec{F}, \vec{s} \rangle := ||\vec{F}'||\,||\vec{s}||. \tag{C.14}$$

Man sieht nun unmittelbar mit Hilfe geometrischer Überlegungen, dass das so definierte Produkt eine lineare Abbildung ist, das soll heissen es gilt

$$\begin{aligned} \langle (\vec{F}_1 + \vec{F}_2), \vec{s} \rangle &= \langle \vec{F}_1, \vec{s} \rangle + \langle \vec{F}_2, \vec{s} \rangle, \\ \langle \alpha \vec{F}, \vec{s} \rangle &= \alpha \langle \vec{F}, \vec{s} \rangle, \qquad \alpha \in \mathbb{R}. \end{aligned} \tag{C.15}$$

Für das skalare Produkt von zwei Vektoren $\vec{a}$ und $\vec{b}$ sind auch noch folgende Schreibweisen üblich

$$\vec{a} \cdot \vec{b} = \langle \vec{a}, \vec{b} \rangle = (\vec{a}, \vec{b}). \tag{C.16}$$

Wegen Gleichung (C.12) gilt auch

$$\langle \vec{a}, \vec{b} \rangle = ||\vec{b}||\,||\vec{a}|| \cos\varphi. \tag{C.17}$$

Damit ist auch unmittelbar einsichtig, dass das skalare Produkt symmetrisch in $\vec{a}$ und $\vec{b}$ ist

$$\langle \vec{a}, \vec{b} \rangle = \langle \vec{b}, \vec{a} \rangle. \tag{C.18}$$

Umgekehrt gilt

$$\varphi = \langle \frac{\vec{a}}{||\vec{a}||}, \frac{\vec{b}}{||\vec{b}||} \rangle. \tag{C.19}$$

Mittels der Komponenten kann das skalare Produkt nun im $\mathbb{R}^3$ (analog im $\mathbb{R}^2$) wie folgt berechnet werden

$$\langle \vec{a}, \vec{b} \rangle = \langle \big(a_1, a_2, a_3\big)\,, \big(b_1, b_2, b_3\big) \rangle = a_1 b_1 + a_2 b_2 + a_3 b_3. \tag{C.20}$$

Beweis: dies folgt aus der Darstellung mit den Einheitsvektoren $\vec{e}_j, j = x, y, z$

$$\begin{aligned} &\vec{a} = a_1 \vec{e}_x + a_2 \vec{e}_y + a_3 \vec{e}_z, \qquad \vec{b} = b_1 \vec{e}_x + b_2 \vec{e}_y + b_3 \vec{e}_z, \\ &\langle \vec{a}, \vec{b} \rangle = \langle a_1 \vec{e}_x + a_2 \vec{e}_y + a_3 \vec{e}_z, b_1 \vec{e}_x + b_2 \vec{e}_y + b_3 \vec{e}_z \rangle = a_1 b_1 + a_2 b_2 + a_3 b_3, \end{aligned} \tag{C.21}$$

da $\langle \vec{e}_x, \vec{e}_x \rangle = 1$ und $\langle \vec{e}_x, \vec{e}_y \rangle = 0$, etc. ist.

Stehen zwei Vektoren normal (im rechten Winkel) aufeinander, so ist ihr skalares Produkt 0.
Anmerkung: Daraus folgt, dass im $\mathbb{R}^2$ ein **Normalvektor** zum Vektor (a, b) durch $(b, -a)$ oder $(-b, a)$ gegeben ist.

Anwendungsbeispiele:
(i) Man berechne die Fläche F eines Dreiecks $A = (0,0), B = (6,0), C = (3,4)$.

$$F = \frac{1}{2}||\vec{c}||\,||\vec{h}_{\vec{c}}|| = \frac{1}{2}||\vec{c}||\,|\langle\frac{\vec{c}^{\perp}}{||\vec{c}||}, \vec{a}\rangle| = \frac{1}{2}|\langle\vec{c}^{\perp}, \vec{a}\rangle|$$
$$= \frac{1}{2}|\langle\overrightarrow{AB}^{\perp}, \overrightarrow{BC}\rangle| = \frac{1}{2}|\Big\langle \begin{pmatrix}0\\6\end{pmatrix}, \begin{pmatrix}-3\\4\end{pmatrix} \Big\rangle| = 12.$$

(ii) Man berechne den Abstand d eines Punktes $C = (3,4)$ von einer Geraden gegeben durch zwei Punkte $A = (0,0), B = (6,0)$.

$$d = ||\vec{h}_{\vec{c}}|| = 4.$$

C.2.4. Anwendung: Geraden- und Ebenengleichungen in $\mathbb{R}^2$ und $\mathbb{R}^3$.

Die Menge aller Punkte einer Geraden g kann mittels Vektoren im $\mathbb{R}^3$ wie folgt angegeben werden

$$g : \{\vec{x} \mid \vec{x} = \vec{p} + \lambda\vec{r}, \text{ wobei } \lambda \in \mathbb{R}\}$$
$$= \{\begin{pmatrix}x\\y\\z\end{pmatrix} \mid \begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}p_1\\p_2\\p_3\end{pmatrix} + \lambda\begin{pmatrix}r_1\\r_2\\r_3\end{pmatrix}, \ \lambda \in \mathbb{R}\} \tag{C.22}$$

wobei $\vec{p}$ der Ortsvektor eines Punktes P und $\vec{r}$ ein Richtungsvektor der Geraden ist, z.B. $\overrightarrow{PQ}$ wenn Q ein zweiter Punkt der Geraden ist. Im Fall des $\mathbb{R}^2$ ist keine 3. Komponente vorhanden, also

$$g : \{\vec{x} \mid \vec{x} = \vec{p} + \lambda\vec{r}, \text{ wobei } \lambda \in \mathbb{R}\}$$
$$= \{\begin{pmatrix}x\\y\end{pmatrix} \mid \begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}p_1\\p_2\end{pmatrix} + \lambda\begin{pmatrix}r_1\\r_2\end{pmatrix}, \ \lambda \in \mathbb{R}\} \tag{C.23}$$

Eliminiert man in diesen zwei Gleichungen den Parameter λ, so gelangt man zur Form $ax + by = c$.
In der Ebene ist eine Gerade auch durch einen Punkt A und ihren Normalvektor $\vec{n}$ festgelegt, d.h. durch die Gleichung $\langle\overrightarrow{AX}, \vec{n}\rangle = 0$. Daraus ergibt sich unmittelbar, dass der Normalvektor $n = (a,b)$ ist (der Normalvektor ist nur bis auf die Länge festgelegt).

Die Menge aller Punkte einer Ebene ϵ ist gegeben durch

$$\epsilon : \{\vec{x} \mid \vec{x} = \vec{p} + \lambda\vec{r} + \mu\vec{s}, \text{ wobei } \lambda, \mu \in \mathbb{R}\}$$

$$= \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} + \lambda \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} + \mu \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}, \ \lambda, \mu \in \mathbb{R} \} \tag{C.24}$$

wobei $\vec{p}$ der Ortsvektor eines Punktes P und $\vec{r}, \vec{s}$ nichtparallele Richtungsvektoren der Ebene sind, z.B. $r = \overrightarrow{PQ}$ und $s = \overrightarrow{PR}$, wenn die Ebene durch 3 Punkte gegeben ist. Eliminiert man in diesen drei Gleichungen die Parameter λ, μ, so gelangt man zur Form $ax + by + cz = d$.

Im Raum ist eine Ebene auch durch einen Punkt A und ihren Normalvektor $\vec{n}$ festgelegt, d.h. durch die Gleichung $\langle \overrightarrow{AX}, \vec{n} \rangle = 0$. Daraus ergibt sich unmittelbar, dass der Normalvektor $n = (a, b, c)$ ist (der Normalvektor ist nur bis auf die Länge festgelegt).

Den kürzesten Abstand d eines Punktes A von einer Ebene berechnet man wieder mit dem Skalarprodukt $d = \langle \overrightarrow{AX}, \vec{n} \rangle$, wobei hier nun n ein Einheitsvektor ist.

C.2.5. Vektorielles Produkt zweier Vektoren.

Das Vektorprodukt (auch Kreuzprodukt, Dachprodukt oder alternierende 2-Form) genannt **gibt es in dieser Art nur** im $\mathbb{R}^3$.

Das Produkt $\vec{a} \times \vec{b}$ (bzw. $\vec{a} \wedge \vec{b}$) zweier Vektoren $\vec{a}$ und $\vec{b}$ ergibt einen Vektor der auf die Ebene die $\vec{a}$ und $\vec{b}$ aufspannen, normal steht. Der Betrag dieses Vektors ist gleich der Fläche des von $\vec{a}$ und $\vec{b}$ aufgespannten Parallelogramms. $\vec{a}, \vec{b}$ und $\vec{c}$ bilden ein Rechtssystem (Rechtsschraubenregel). Es gilt

$$||\vec{a} \times \vec{b}|| = ||\vec{a}|| \, ||\vec{b}|| \sin \phi. \tag{C.25}$$

Vektorprodukte erfüllen weder das Kommutativ- noch das Assoziativgesetz.

$$\begin{aligned} (\vec{a} \times \vec{b}) \times \vec{c} &\neq \vec{a} \times (\vec{b} \times \vec{c}), \\ \vec{a} \times \vec{b} &= -\vec{b} \times \vec{a}. \end{aligned} \tag{C.26}$$

Daraus folgt: Sind die Vektoren $\vec{a}$ und $\vec{b}$ parallel, so gilt

$$\vec{a} \times \vec{b} = 0. \tag{C.27}$$

Für die orthonormale Basis $\vec{e}_x, \vec{e}_y, \vec{e}_z$ gilt

$$\vec{e}_x \times \vec{e}_y = \vec{e}_z. \tag{C.28}$$

Damit kann nun das Vektorprodukt wie folgt durch die Komponenten der Vektoren berechnet werden

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \\ -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \\ \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \end{pmatrix} = \begin{vmatrix} \vec{e}_x & \vec{e}_y & \vec{e}_z \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}. \tag{C.29}$$

wobei $|\ldots|$ die Determinante bezeichnet, z.B.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} := ad - bc. \tag{C.30}$$

Beweis: dies folgt aus der Darstellung mit den Einheitsvektoren

$$\begin{aligned} \vec{a} = a_1\vec{e}_x + a_2\vec{e}_y + a_3\vec{e}_z, \quad \vec{b} = b_1\vec{e}_x + b_2\vec{e}_y + b_3\vec{e}_z, \\ \vec{a} \wedge \vec{b} = (a_1\vec{e}_x + a_2\vec{e}_y + a_3\vec{e}_z) \wedge (b_1\vec{e}_x + b_2\vec{e}_y + b_3\vec{e}_z) \\ = (a_2b_3 - a_3b_2)\vec{e}_x - (a_1b_3 - a_3b_1)\vec{e}_y + (a_1b_2 - a_2b_1)\vec{e}_z \end{aligned} \tag{C.31}$$

da $\vec{e}_j \wedge \vec{e}_j = 0, j = x, y, z$ ist.

Beispiel: Abstand zweier windschiefer Geraden. Der Vektor $\frac{\vec{a}\times\vec{b}}{||\vec{a}\times\vec{b}||}$ steht normal auf die Geraden $G_{\vec{a}}$, $G_{\vec{b}}$ mit den Richtungsvektoren $\vec{a}$ und $\vec{b}$ und hat die Länge eins. Der Abstand d ist dann gegeben durch das Skalarprodukt

$$d = \langle \frac{\vec{a} \times \vec{b}}{||\vec{a} \times \vec{b}||}, \vec{c} > \tag{C.32}$$

wobei der Vektor $\vec{c}$ ein beliebiger Vektor ist, dessen Anfangspunkt auf der Geraden $G_{\vec{a}}$ und dessen Endpunkt auf der Geraden $G_{\vec{b}}$ liegt.

Das gemischte Produkt dreier Vektoren

$$\vec{a} \cdot (\vec{b} \times \vec{c}) = \det \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \tag{C.33}$$

Geometrisch kann das gemischte Produkt dreier Vektoren als das Volumen (mit Vorzeichen) des von den drei Vektoren aufgespannten Körpers (Parallelepiped) interpretiert werden.

C.2.6. Gauss-Elimination.

Gauss-Elimination anhand eines Gleichungssytems von drei Gleichungen mit drei Unbekannten[1]

Wir betrachten folgendes Gleichungssystem

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= b_1 \\ a_{21}x + a_{22}y + a_{23}z &= b_2 \\ a_{31}x + a_{32}y + a_{33}z &= b_3. \end{aligned}$$

Wir können annehmen, dass $a_{11} \neq 0$ ist. (Dies kann man durch Vertauschung der Zeilen erreichen, ausser $a_{11} = a_{21} = a_{31} = 0$. In diesem Fall hat man aber nur drei Gleichungen mit zwei Unbekannten.) Multipliziert man nun die erste Zeile mit $-\frac{a_{21}}{a_{11}}$ und addiert sie zur zweiten Zeile, so erhält man einen neue zweite Zeile der Form

$$\bar{a}_{22}y + \bar{a}_{23}z = \bar{b}_2. \tag{C.34}$$

Multipliziert man die erste Zeile mit $-\frac{a_{31}}{a_{11}}$ und addiert sie zur dritten Zeile, so erhält man einen neue dritte Zeile der Form

$$\bar{a}_{32}y + \bar{a}_{33}z = \bar{b}_3.$$

Das Gleichungssytem hat nun die Form

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= b_1 \\ \bar{a}_{22}y + \bar{a}_{23}z &= \bar{b}_2 \\ \bar{a}_{32}y + \bar{a}_{33}z &= \bar{b}_3. \end{aligned}$$

Ist nun $\bar{a}_{22} = \bar{a}_{23} = \bar{a}_{32} = \bar{a}_{33} = 0$ und $\bar{b}_2$ oder $\bar{b}_3 \neq 0$, so existiert keine Lösung. Anderenfalls kann man das Gleichungssytem auf die Form (eventuell durch Vertauschen der Variablen)

$$\begin{aligned} \tilde{a}_{11}\tilde{x} + \tilde{a}_{12}\tilde{y} + \tilde{a}_{13}\tilde{z} &= \tilde{b}_1 \\ \tilde{a}_{22}\tilde{y} + \tilde{a}_{23}\tilde{z} &= \tilde{b}_2 \\ \tilde{a}_{32}\tilde{y} + \tilde{a}_{33}\tilde{z} &= \tilde{b}_3. \end{aligned}$$

mit $\tilde{a}_{22} \neq 0$ bringen. Multipliziert man nun die zweite Zeile mit $-\frac{\tilde{a}_{32}}{\tilde{a}_{22}}$ und addiert sie zur dritten Zeile, so erhält man

---

[1] Anmerkung: Erlaubte Umformungen eines Gleichungssystems, die die Lösungsmenge unverändert lassen, sind:
(i) Multiplikation einer Zeile mit einer beliebigen Konstanten $k, k \neq 0$
(ii) Vertauschung zweier Zeilen
(iii) Addition eines beliebigen Vielfachen einer anderen Zeile zu einer Zeile.

$$\begin{aligned} \tilde{a}_{11}\tilde{x} + \tilde{a}_{12}\tilde{y} + \tilde{a}_{13}\tilde{z} &= \tilde{b}_1 \\ \tilde{a}_{22}\tilde{y} + \tilde{a}_{23}\tilde{z} &= \tilde{b}_2 \\ \hat{a}'_{33}\hat{z} &= \hat{b}'_3. \end{aligned}$$

Ist nun $\hat{a}'_{33} = 0$ und $\hat{b}'_3 \neq 0$, so existiert keine Lösung. Ist $\hat{a}'_{33} \neq 0$, folgt $\tilde{z} = \frac{\hat{b}'_3}{\hat{a}'_{33}}$ und das Gleichungssytem kann durch Rückeinsetzen (Substitution) aufgelöst werden.

Anmerkung: Graphisch stellt eine Gleichung mit 3 Unbekannten eine Ebene im Raum dar. Die möglichen Lagen dreier Ebenen im Raum geben die möglichen Lösungsformen an.

Beispiel:

$$\begin{cases} 7\,y + 3\,z = -12 \\ 2\,x + 8\,y + z = 0 \\ -5\,x + 2\,y - 9\,z = 26 \end{cases}.$$

C.3. Übung

(1) Gegeben sind die Vektoren $\vec{a} = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$, $\vec{b} = \begin{pmatrix} -1 \\ 2 \\ -2 \end{pmatrix}$, $\vec{c} = \begin{pmatrix} -3 \\ 0 \\ 2 \end{pmatrix}$. Man berechne die Koordinaten und die Beträge der folgenden aus ihnen gebildeten Vektoren: (a) $\vec{a} - 3\vec{b} + 2\vec{c}$, (b) $2(\vec{a} + 3\vec{b}) - (2\vec{b} - 3\vec{c})$.

(2) Welche Gegenkraft $\vec{F}$ hebt die Summe der folgenden Einzelkräfte in ihrer physikalischen Wirkung auf?

$$\vec{F}_1 = \begin{pmatrix} 10 \\ 15 \\ -30 \end{pmatrix}, \quad \vec{F}_2 = \begin{pmatrix} -30 \\ 40 \\ -60 \end{pmatrix}, \quad \vec{F}_3 = \begin{pmatrix} -50 \\ 0 \\ 80 \end{pmatrix}, \quad \vec{F}_4 = \begin{pmatrix} 10 \\ -100 \\ 30 \end{pmatrix}.$$

(3) Man normiere die Vektoren $\vec{a} = \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}$, $\vec{b} = -\vec{a}$, $\vec{c} = 2\vec{e}_x - 4\vec{e}_y + 5\vec{e}_z$.

(4) $A = (1, 2)$ und $B = (4, 3)$ sind die Randpunkte einer Strecke in der Ebene. Durch Verdoppeln der Strecke $\overline{AB}$ über B hinaus erhält man den Punkt C. Wie lauten seine Koordinaten?

(5) Gegeben sind die Vektoren der Aufgabe 1. Man berechne die folgenden Skalarprodukte:
(a) $\vec{a} \cdot \vec{b}$, (b) $(\vec{a} - 3\vec{b}) \cdot (2\vec{c})$, (c) $(\vec{a} + \vec{b}) \cdot (\vec{a} - \vec{c})$.

(6) Ein kleiner Körper bewegt sich in der Ebene geradlinig vom Punkt $A = (1, 0)$ zum Punkt $B = (6, 3)$, wobei die konstante Kraft $\vec{F} = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$ auf ihn wirkt. Man skizziere die Situation und berechne die Arbeit (Arbeit = Kraftkomponente in Richtung der Weges · Weg), die dabei verrichtet wird.

(7) Welchen Winkel schliessen die Vektoren $\vec{a} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ und $\vec{b} = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$ ein?

(8) Man bestimme den Betrag von Vektor $\vec{a}$ aus Aufgabe 1 und die Winkel, die $\vec{a}$ mit den drei Koordinatenachsen einschliesst.

(9) Von einem zweidimensionalen Vektor $\vec{a}$ sind der Betrag $|\vec{a}| = 12$ und der Winkel $\alpha = 35°$ zur x-Achse bekannt. Wie lauten die Komponenten von $\vec{a}$?

(10) Man berechne den Betrag der Resultierenden der ebenen Kräfte $F_1 = (3, 1)$ und $F_2 = (1, 2)$, sowie ihren Winkel zur x-Achse.

(11) Man zerlege den Vektor $\vec{c} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$ in Richtung der Vektoren $\vec{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $\vec{b} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, d.h., man ermittle die Zahlen λ und μ derart, daß gilt: $\vec{c} = \lambda\vec{a} + \mu\vec{b}$.

Man löse die Aufgabe auch graphisch.

(12) Wie lautet der Einheitsvektor $\vec{e}_{\vec{a}}$, der die gleiche Richtung wie der Vektor

$$\vec{a} = \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix} \text{ hat?}$$

(13) Man zeige, daß das Viereck $ABCD$ mit $A = (0, 0)$, $B = (4, 1)$, $C = (5, 6)$ und $D = (-3, 4)$ ein Trapez ist. (Definition eines Trapezes: = Viereck, bei dem zwei Seiten parallel sind.)

(14) Man zeige: die Vektoren $\vec{a}$ und $\vec{b}$ sind zueinander orthogonal

(a) $\vec{a} = \begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}, \vec{b} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$, (b) $\vec{a} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}, \vec{b} = \begin{pmatrix} 1 \\ 6 \\ 3 \end{pmatrix}$.

(15) Man ermittle den Mittelpunkt M der Strecke zwischen $P = (2, 1, 1)$ und $Q = (4, 5, 7)$.

(16) Gegeben sind die drei Eckpunkte eines Dreiecks

$$A = (1, 0, -2), B = (5, 9, 3), C = (3, 3, 5).$$

Berechnen sie die Koordinaten des Schwerpunktes.

Hinweis: Die Schwerpunktslinien gehen durch einen Eckpunkt und den Mittelpunkt der gegenüberliegenden Seite. Die Schwerpunktslinien werden durch den Schwerpunkt im Verhältnis zwei zu eins geteilt. Machen sie eine Skizze und berechnen sie den Ortsvektor des Schwerpunktes.

(17) Man berechne die Determinanten mit der Regel von Sarrus

$$\text{a)} \begin{vmatrix} 1 & 4 & 2 \\ 0 & 3 & 1 \\ -2 & 0 & 3 \end{vmatrix}, \qquad \text{b)} \begin{vmatrix} 2 & 1 & -3 \\ 0 & 2 & 1 \\ 2 & 3 & -2 \end{vmatrix}.$$

(18) Man ermittle das Kreuzprodukt der Vektoren:

$$\text{(a)} \quad \vec{a} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \vec{b} = \begin{pmatrix} 4 \\ 5 \\ 4 \end{pmatrix}, \quad \text{(b)} \quad \vec{a} = \begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}, \vec{b} = \begin{pmatrix} 6 \\ 2 \\ -4 \end{pmatrix}.$$

(19) Man überlege sich wie man den Abstand eines Punktes von einer Ebene berechnen kann. Die Ebene sei durch einen Punkt und zwei Richtungsvektoren gegeben.

(20) Man betrachte im $\mathbb{R}^3$ zwei Geraden g_1 und g_2, die durch

$$g_i = \{p_i + tv_i \mid t \in \mathbb{R}\}, \qquad i = 1, 2$$

gegeben sind, wobei

$$p_1 = (1, 1, 2), v_1 = (2, 0, 1), \quad p_2 = (0, -1, 3), v_2 = (1, 1, 1).$$

Wie groß ist der Abstand a zwischen g_1 und g_2?

(21) Man löse mittels Gauss'schem Algorithmus

$$\begin{cases} x & + & y & - & z & = & 9 \\ x & + & 8\,y & + & 6\,z & = & -5 \\ -2\,x & + & 4\,y & - & 6\,z & = & 40 \end{cases}.$$

Anhang D

Beschreibende Statistik und Zusammenhangsanalysen

D.1. Häufigkeitsverteilung einer Stichprobe

Ein Beispiel aus der **Qualitätssicherung**: Ein Unternehmen füllt Marmelade in Gläser ab. Die vorgeschriebene Füllmenge ist $400g$. Der Abfüllprozess ist aber nicht so vollkommen, dass alle Abfüllmengen gleich ausfallen. Dafür gibt es verschiedene Gründe, wie etwa Abnutzung der verwendeten Automaten, menschliche Unzulänglichkeiten, usw.. Die Abfüllmengen werden also mehr oder weniger stark vom vorgegebenen Sollwert $400g$ abweichen; man sagt, dass sie um diesen Wert *streuen*. Das ist dem Hersteller (und dem Abnehmer) bewußt, und er wird daher auch solche Abfüllmengen akzeptieren, die nur *geringfügig* vom Sollwert abweichen, zum Beispiel um $5g$ nach oben oder unten. Nun wird es aus Zeit- und Kostengründen nicht möglich sein, *jedes einzelne* abgefüllte Glas zu kontrollieren. Der Hersteller wird daher in *regelmäßigen* Zeitabständen (z.B. stündlich) eine gewisse gleichbleibende Anzahl von abgefüllten Gläsern (z.B. jeweils 10 Stück) wahllos (d.h. *zufällig*) aus der laufenden Abfüllung herausgreifen und ihre Füllmenge messen. Man spricht von *(Zufalls-)Stichproben vom Umfang* 10, die aus der *Grundgesamtheit* (hier z.B. die Tagesproduktion an abgefüllten Gläsern) entnommen werden. Die gemessenen Füllmengen einer solchen Stichprobe könnten folgendermaßen aussehen (in g): $399, 401, 400, 399, 400, 403, 399, 401, 396, 399$. Diese Daten werden zunächst übersichtlich dargestellt. Danach werden aus ihnen verschiedene *Kennwerte* berechnet, die möglichst viel Information über die Stichprobe geben sollen, wie zum Beispiel ein Mittelwert und ein Maß für die Streuung der Daten um den Mittelwert. Bis hierher spricht man von der **beschreibenden Statistik**. Anschließend kann man mithilfe der Wahrscheinlichkeitsrechnung den Stichprobenbefund auf die Grundgesamtheit verallgemeinern. Die dabei angewandten statistischen Verfahren werden als **schließende** (oder auch

beurteilende) **Statistik** bezeichnet.
Eine grundlegende Aufgabe der **Statistik** besteht darin, Informationen über bestimmte Merkmale von Objekten zu gewinnen, ohne dass dabei *alle* Objekte untersucht werden müssen.

Die Untersuchung *aller* Objekte ist in der Regel gar nicht möglich, da es *zu viele* Objekte sind oder weil die Untersuchung der Objekte diese unter Umständen *zerstört* - wenn man z.B. an der Lebensdauer einer Glühbirne interessiert ist, dann muss man sie bis zur Zerstörung belasten.

Die Menge aller interessierenden Objekte nennt man die **Grundgesamtheit** Ω (z.B. die Tagesproduktion an abgefüllten Gläsern, alle Einheiten einer Lieferung, alle Wahlberechtigten in Österreich, usw.). Aus dieser Grundgesamtheit werden zufällig n Objekte (**statistische Einheiten, Merkmalträger**) ω_j ausgewählt, sodass ein verkleinertes Abbild der Grundgesamtheit gewonnen wird. Man spricht von einer **(Zufalls-)Stichprobe** $P \subset \Omega$ **vom Umfang** $n = |P|$. An der Stichprobe beobachtet man ein oder mehrere **Merkmale** X (z.B. die Füllmenge, die Lebensdauer, das Alter, den Beruf, usw.). Die Darstellung der Stichprobendaten $X(P)$ (eine Teilmenge von M, der Menge aller möglichen Daten[1] (**(Merkmals-)Ausprägungen**)) und die Ermittlung von Kennwerten der Stichprobe ist die Aufgabe der **beschreibenden Statistik**. Die Verallgemeinerung auf die Grundgesamtheit mithilfe der Wahrscheinlichkeitsrechnung ist Gegenstand der **schließenden Statistik**.

Es ist zu beachten, dass die dabei gewonnenen Gesetzmäßigkeiten sich nicht auf ein *einzelnes* Element der Grundgesamtheit beziehen lassen. Man kann zum Beispiel keine Aussage über die Lebensdauer eines einzelnen Menschen machen, wohl aber über die mittlere Lebensdauer einer Bevölkerungsgruppe. Daher können Lebensversicherungen bestimmte altersabhängige Prämien vorschreiben.

Wir wollen nun den Begriff Merkmal präzisieren

Definition D.1. *Ein Merkmal X ist eine Abbildung von der Grundgesamtheit Ω in die Menge der Merkmalsausprägungen M*

$$X : \Omega \to M \text{ mit } X(\omega_j) = x_j \in X(\Omega). \tag{D.1}$$

Man unterscheidet **quantitative** und **qualitative** Merkmale abhängig von den Strukturen (Ordnungen, innere algebraische Verknüpfungen, Metriken) die die Menge M besitzt. Quantitative Merkmale können auf einer *metrischen Skala* dargestellt werden. Beispiele dafür sind Größe, Temperatur, Laufzeit, Beschäftigtenanzahl, usw.. Ist eine Darstellung auf einer metrischen Skala nicht

[1] Werden mehrere Daten erhoben, hat M die Struktur eines Cartesischen Produkts.

möglich, dann spricht man von einem qualitativen Merkmal. Beispiele dafür sind Farbe, Beruf, Fehlerart, usw.. Je nach Art des Merkmals verwendet man unterschiedliche Methoden zur Beschreibung der ermittelten Daten. Beim quantitativen Merkmal „Größe“ kann zum Beispiel ein arithmetisches Mittel berechnet werden, was beim qualitativen Merkmal „Farbe“ nicht möglich ist. Wir werden uns hier in erster Linie auf quantitative Merkmale konzentrieren.

Auch Merkmale, die auf einer *Ordinalskala* dargestellt werden können, die also einem natürlichen Rang nach geordnet werden können, zählt man zu den qualitativen Merkmalen. Beispiele dafür sind die Merkmale „Güteklasse“, „Schulnote“, „Erdbebenstärke“, usw..

D.1.1. Absolute und relative Häufigkeit. Wie erfaßt man nun eine Stichprobe? Zunächst werden die n beobachteten Stichprobenwerte $(a_1, a_2, \ldots, a_n)$ (auch Messwerte, Daten oder Ausprägungen genannt) nacheinander notiert. Die so entstehende Liste (n-Tupel) bezeichnet man als **Urliste**. Im Allgemeinen werden dabei gewisse Werte *mehrmals* auftreten. Bezeichnen wir diese *verschiedenen* Werte mit $x_1, x_2, \ldots, x_k \in M$[2] und zählen wir, wie oft jeder dieser Werte in der Stichprobe auftritt. Diese Anzahl h_i nennt man die **absolute Häufigkeit**[3] des Stichprobenwertes x_i $(i = 1, \ldots, k)$. Die **relative Häufigkeit** f_i jedes Wertes x_i erhält man, wenn man die absolute Häufigkeit h_i durch die Anzahl n aller Daten dividiert:

$$f_i = \frac{h_i}{n}. \tag{D.2}$$

Die Summe der absoluten Häufigkeiten ergibt den Umfang der Stichprobe, d.h. es ist $h_1 + \cdots + h_k = n$. Die Summe der relativen Häufigkeiten ist $f_1 + \cdots + f_k = 1$. Durch die Angabe der verschiedenen auftretenden Stichprobenwerte x_i und ihrer absoluten Häufigkeiten h_i bzw. ihrer relativen Häufigkeiten f_i wird die Stichprobe vollständig beschrieben.

Beispiel D.1. Absolute und relative Häufigkeit
*Bei einem Abfüllprozess wird eine Stichprobe vom Umfang $n = 20$ genommen. Folgende Abfüllmengen (in Gramm) werden dabei notiert (Urliste): (*400, 399, 398, 400, 398, 399, 397, 400, 402, 399, 401, 399, 400, 402, 398, 400, 399, 401, 399, 399*). Geben Sie die absoluten und die relativen Häufigkeiten der Messwerte an.*

[2] Es gilt daher: $\{x_1, x_2, \ldots, x_k\} = \{a_1, a_2, \ldots, a_n\}$, da in einer Menge nur verschiedene Elemente zählen.

[3] $h_i = |X^{-1}(x_i)|$ ist die Mächtigkeit der Urbildmenge von x_i.

Lösung zu D.1. *Die geordnete Urliste ist: (*397, 398, 398, 398, 399, 399, 399, 399, 399, 399, 399, 400, 400, 400, 400, 400, 401, 401, 402, 402*). Es kommen darin* $k = 6$ *verschiedene Werte* $x_1, \ldots, x_6$ *vor:*

$$(397, 398, 399, 400, 401, 402)$$

Der Wert $x_1 = 397$ *wurde einmal gemessen, daher ist seine absolute Häufigkeit gleich* $h_1 = 1$ *und seine relative Häufigkeit gleich* $f_1 = \frac{1}{20}$*. Der Wert* $x_2 = 398$ *wurde dreimal gemessen, daher ist* $h_2 = 3$ *und* $f_2 = \frac{3}{20}$*, usw.. Absolute und relative Häufigkeiten aller Werte sind in der folgenden Tabelle angeführt. Oft ermittelt man dabei die absoluten Häufigkeiten zunächst mithilfe einer* **Strichliste***:*

Häufigkeitsverteilung der Stichprobe

Stichprobenwert x_i	397	398	399	400	401	402	
Strichliste	|	|||	𝍸 ||	𝍸	||	||	
absolute Häufigkeit h_i	*1*	*3*	*7*	*5*	*2*	*2*	$\sum h_i = 20$
relative Häufigkeit f_i	$\frac{1}{20}$	$\frac{3}{20}$	$\frac{7}{20}$	$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{10}$	$\sum f_i = 1$

Zur Kontrolle haben wir in der letzten Spalte die Summe über alle absoluten Häufigkeiten (sie muss gleich dem Stichprobenumfang sein) bzw. die Summe über alle relativen Häufigkeiten (sie muss gleich 1 *sein) berechnet:* $\sum_{i=1}^{6} h_i = 1 + 3 + 7 + 5 + 2 + 2 = 20$ *und* $\sum_{i=1}^{6} f_i = \frac{1}{20} + \frac{3}{20} + \frac{7}{20} + \frac{1}{4} + \frac{1}{10} + \frac{1}{10} = 1$. ■

D.1.2. Darstellungen von Häufigkeitsverteilungen. Die Häufigkeitsverteilung einer Stichprobe lässt sich graphisch durch ein sogenanntes **Stabdiagramm** darstellen. Dabei werden auf der x-Achse die verschiedenen Stichprobenwerte x_i aufgetragen und darüber jeweils ein „Stab“, dessen Höhe proportional der zugehörigen (absoluten oder relativen) Häufigkeit ist. Die Breite des Stabes spielt dabei keine Rolle. In Abbildung D.1 ist das Stabdiagramm zu Beispiel D.1 dargestellt.

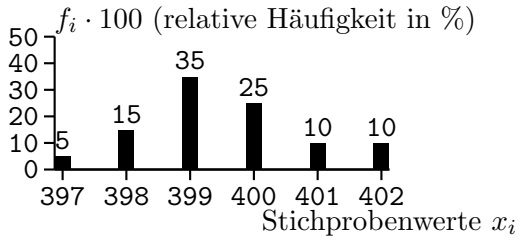


Abbildung D.1. Stabdiagramm

Bei Stichproben mit vielen verschiedenen Messwerten x_i fasst man mehrere Werte in Intervalle, sogenannte **Klassen**, zusammen. Man zählt dann, wieviele Stichprobenwerte in die einzelnen Klassen fallen. Die Anzahl h_i der Stichprobenwerte, die in die i-te Klasse fallen, nennt man die **(absolute) Häufigkeit**

der i-ten Klasse. Dividiert man diese Anzahl durch die Gesamtanzahl n aller Stichprobenwerte, so erhält man die **relative Häufigkeit $f_i = \frac{h_i}{n}$ der i-ten Klasse**. Sehen wir uns das gleich anhand eines Beispiels an:

Beispiel D.2. Klassenbildung bei umfangreicher Stichprobe
Gegeben ist folgende (bereits geordnete) Urliste einer Stichprobe vom Umfang 20*: (*3, 7, 12, 18, 19, 20, 25, 25, 27, 28, 29, 31, 32, 34, 37, 38,40, 41, 45, 47*). Gruppieren Sie die Stichprobenwerte in geeignete Klassen und bestimmen Sie die absoluten und die relativen Häufigkeiten der Klassen.*

Lösung zu D.2. *Die Klassen müssen alle Stichprobenwerte überdecken. Der kleinste Stichprobenwert ist* 3*, der größte ist* 47*. Wir können daher zum Beispiel die Intervalle* $[0, 10)$*,* $[10, 20)$*,* $[20, 30)$*,* $[30, 40)$*,* $[40, 50)$ *als Klassen wählen. Zur ersten Klasse* $[0, 10)$ *zählen alle Stichprobenwerte* x_i *mit* $0 \leq x_i < 10$*. Damit fallen die Stichprobenwerte* 3 *und* 7 *in diese Klasse, also ist ihre absolute Häufigkeit* $h_1 = 2$*. In die zweite Klasse* $[10, 20)$ *fallen die Stichprobenwerte mit* $10 \leq x_i < 20$*, also* 12, 18*, und* 19*; damit ist* $h_2 = 3$*. In der dritten Klassen liegen die Stichprobenwerte* 20, 25, 25, 27, 28 *und* 29*, daher ist* $h_3 = 6$*, usw.. Die relative Häufigkeit der ersten Klasse ist* $f_1 = \frac{2}{20} = 0.1$*, usw.. Alle absoluten und relativen Häufigkeiten sind in der folgenden Tabelle zusammengefasst:*

Häufigkeitsverteilung der klassierten Stichprobe

Klasse	$[0, 10)$	$[10, 20)$	$[20, 30)$	$[30, 40)$	$[40, 50)$
Strichliste	||	|||	𝍸 |	𝍸	||||
absolute Häufigkeit h_i	*2*	*3*	*6*	*5*	*4*
relative Häufigkeit f_i	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{6}{20}$	$\frac{1}{4}$	$\frac{1}{5}$

■

Damit bedeutet jede Klasseneinteilung natürlich einen Informationsverlust, da nur noch angegeben wird, wieviele Werte in einer Klasse liegen, jedoch nicht mehr, wo sie zwischen den Klassengrenzen liegen. Viele Klassen bedeuten weniger Informationsverlust, weniger Klassen bedeuten eine größere Übersicht. Hier muss man also einen Kompromiss finden.

Für die Klassenbildung gibt es gewisse „Faustregeln“, unter anderem:
• Die Klassen sollten gleich breit gewählt werden (im obigen Beispiel war jedes Intervall 10 Einheiten lang).
• Die Anzahl der Klassen sollte etwa zwischen 5 und 20 liegen, jedoch $\sqrt{n}$ nicht wesentlich überschreiten (wobei n der Umfang der Stichprobe ist). Das stellt bis zu einem gewissen Grad sicher, dass alle Klassen „gut gefüllt“ sind.
• Die Klassengrenzen sollten möglichst einfache Zahlen sein und wenn möglich nicht mit Stichprobenwerten zusammenfallen.

Die Häufigkeitsfunktion f wird graphisch dargestellt, indem man über den einzelnen Klassen Rechtecke zeichnet, deren Höhe proportional $\frac{f_i}{\Delta x_i}$ ist. Dadurch wird die *Fläche* eines Rechtecks ein Maß dafür, wieviele Stichprobenwerte in der zugehörigen Klasse liegen. Diese graphische Darstellung von klassierten Stichproben nennt man ein **Histogramm**. Abbildung D.2 zeigt das Histogramm der klassierten Stichprobe aus Beispiel D.2.

Bei einem Histogramm soll die *Fläche* (und nicht die Höhe) des Rechtecks ein Maß für die Anzahl der Messwerte sein, die in die jeweilige Klasse fallen. Das hat den Vorteil, dass sich die graphische Darstellung dann kaum ändert, wenn zwei Klassen zusammengelegt werden. Wenn alle Klassen gleich breit sind, dann kann man die Höhe der Rechtecke proportional f_i wählen. Dann ist sowohl die Höhe als auch die Fläche eines Rechtecks ein Maß dafür, wie oft ein Stichprobenwert in die jeweilige Klasse fällt. Wenn die Klassen aber *nicht gleich breit* sind, dann muss man als Höhe des Rechtecks (einen Wert proportional) $\frac{f_i}{\Delta x_i}$ wählen, wobei Δx_i die Breite der jeweiligen Klasse ist. Dann ist die *Fläche* des Rechtecks über der Klasse i gleich $\frac{f_i}{\Delta x_i} \cdot \Delta x_i = f_i$, also wie gewünscht ein Maß für die relative Häufigkeit der Klasse.

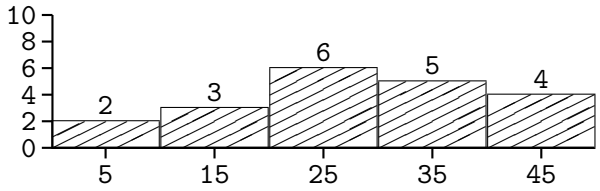


Abbildung D.2. Histogramm einer klassierten Stichprobe

D.1.3. Empirische Verteilungsfunktion. Zur Einführung der empirischen Verteilungsfunktion betrachten wir folgendes Beispiel. Gegeben sei die Liste der Körpergrößen (in cm) einer Gruppe von Studierenden.

$$liste = (175, 180, 183, 192, 186, 175, 163, 169, 172, 180, 188, 178, 186, 170, 166,$$
$$189, 178, 174, 164, 160, 176, 192, 174, 170, 170, 177, 172, 185, 174, 172, 162, 184).$$

Man berechnet nun für die geordnete Liste der verschiedenen Körpergrößen x_j die relativen Häufigkeiten h_j.

$$h = (\frac{1}{32}, 0, \frac{1}{32}, \frac{1}{32}, \frac{1}{32}, 0, \frac{1}{32}, 0, 0, \frac{1}{32}, \frac{3}{32}, 0, \frac{3}{32}, 0, \frac{3}{32}, \frac{2}{32}, \frac{1}{32}, \frac{1}{32}, \frac{2}{32}, 0,$$
$$\frac{2}{32}, 0, 0, \frac{1}{32}, \frac{1}{32}, \frac{1}{32}, \frac{2}{32}, 0, \frac{1}{32}, \frac{1}{32}, 0, 0, \frac{2}{32}).$$

und erzeugt damit die Funktion $F(x)$ folgendermaßen: $F(x)$ ist die Summe der relativen Häufigkeiten h_j mit $x_j \leq x$. Die sich ergebene Funktion $F(x)$ ist in Abbildung D.3 dargestellt.

Allgemein ergibt dies die folgende Definition.

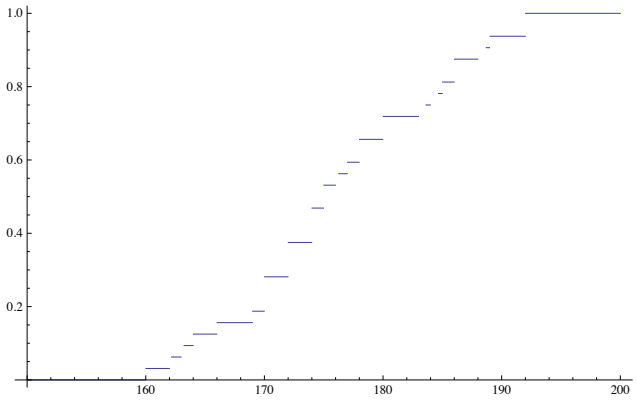


Abbildung D.3. Empirische Verteilungsfunktion der Körpergrößen

Definition D.2. *Seien $h_1, \dots h_k$ die relativen Häufigkeiten der reellen Merkmalsausprägungen $x_1, x_2, \dots, x_k$ einer Stichprobe $(x_1, \dots, x_n)$. Dann heißt die Funktion $F(x) : \mathbb{R} \to [0, 1]$*

$$F(x) := \frac{1}{n} \sum_{j=1}^{n} I(x_j \leq x) = \begin{cases} 0 & \text{für } x < x_1 \\ \sum_{j=1}^{s} h_j & \text{für } x_s \leq x < x_{s+1} \\ 1 & \text{für } x_k \leq x \end{cases}$$

die empirische Verteilungsfunktion der Stichprobe. Sie ist eine monoton wachsende Funktion von 0 bis 1. I bezeichnet hier die charakteristische Funktion.

D.1.4. Simpson-Paradoxon. Das Simpson-Paradoxon spiegelt die Tatsache wider, dass die Durchschnittsbildung von Durchschnitten verschiedener Populationen nicht notwendigerweise den Durchschnitt der zusammengefassten Population ergibt.

Bekannt wurde es durch die Klage gegen die Universität Berkeley in den 1970-ern. Es wurde behauptet, dass die weiblichen Studenten bei der Zulassung zum Studium benachteiligt wurden. Anbei ein Auswahl der Daten[4]

[4]Siehe: `http://www.stat.berkeley.edu/~stark/SticiGui/Text/experiments.htm`, bei Fach F sind die Zahlen der verschiedenen Quellen verschieden: 272 oder 373.

Fach	m. Bewerber	zugelassen	in %	w. Bewerber	zugelassen	in %
A	825	512	62	108	89	82
B	560	353	63	25	17	68
D	417	138	33	375	131	35
F	373	22	6	341	24	7
Summe	2175	1025	47	849	261	31

Obwohl in diesem Beispiel in jedem Fach die weiblichen Bewerber bevorzugt sind, ergibt sich in der Gesamtschau eine scheinbare Benachteiligung der weiblichen Bewerber. Der Grund dafür ist, dass sich die weiblichen Bewerber vor allem Fächer mit geringeren Aufnahmechancen ausgesucht haben.

Man kann sich das auch mit zwei Athleten eines Duathlonbewerbes (Laufen und Radfahren) vorstellen. Es sei Athlet A der bessere Läufer und Radfahrer. Wenn man aber Athlet A 10km laufen und 1km radfahren lässt und Athlet B 1km laufen und 10km radfahren gewinnt trotzdem Athlet B.

Dieses Phänomen tritt sehr oft bei statistischen Auswertungen in den Sozialwissenschaften und in der Medizin (z.B. Medikamententests) auf.

Betrachtet man einen Test von zwei Dingen x, y in zwei Gruppen A, B, z.B.

Fach	gesamt x	positive x	in %	gesamt y	positive y	in %
A	a	b	b/a	c	d	d/c
B	e	f	f/e	g	h	h/g
Summe	$a+e$	$b+f$	$\frac{b+f}{a+e}$	$c+g$	$d+h$	$\frac{d+h}{c+g}$

Allgemein gilt dann, dass aus

$$\frac{b}{a} < \frac{d}{c} \text{ und } \frac{f}{e} < \frac{h}{g}$$

nicht gefolgert werden kann, dass auch immer

$$\frac{b+f}{a+e} < \frac{d+h}{c+g}$$

gilt.

D.2. Kennwerte einer Stichprobe

Statt eine Stichprobe vollständig (z.B. durch Angabe ihrer relativen Häufigkeiten) zu beschreiben, kann man wesentliche Eigenschaften der Stichprobe bereits durch einige Kennwerte angeben. Man unterscheidet zwei Arten von Kennwerten: einerseits **Lagekennwerte**, die Information darüber geben, wo die Werte der Stichprobe im Mittel liegen; und andererseits **Streuungskennwerte**, die

etwas darüber aussagen, ob die Stichprobenwerte an einer Stelle konzentriert sind oder ob sie stark *streuen*.

Wenn man kurz vom **Mittelwert** oder **durchschnittlichen Wert** einer Stichprobe spricht, dann meint man in der Regel ihr **arithmetisches Mittel**

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Hier ist n der Umfang der Stichprobe und $x_1, \ldots, x_n$ sind die gemessenen Werte der Stichprobe (die im Allgemeinen nicht alle verschieden sind). Wenn wir mit $x_1, \ldots, x_k$ die *verschiedenen* Werte einer Stichprobe bezeichnen, und wenn der Wert x_i in der Stichprobe h_i-mal vorkommt, dann lautet die Formel für das arithmetische Mittel

$$\bar{x} = \frac{1}{n}(h_1 x_1 + \cdots + h_k x_k) = \sum_{i=1}^{k} f_i x_i. \tag{D.3}$$

Beispiel D.3. Arithmetisches Mittel einer Stichprobe
Gegeben ist die Stichprobe ($1, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4$*). Berechnen Sie ihr arithmetisches Mittel.*

Lösung zu D.3. *Das arithmetische Mittel dieser* 11 *Stichprobenwerte ist*

$$\bar{x} = \frac{1}{11}(1 + 3 \cdot 2 + 5 \cdot 3 + 2 \cdot 4) = 2.7.$$

Mit `Mathematica` *kann das arithmetische Mittel mit* `Mean` *berechnet werden*[5].

```
In[1]:=Mean[{1, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4}]
Out[1]=30/11
In[2]:=N[%]
Out[2]=2.72727
```

■

Ein anderer Lagekennwert ist der **Median** (auch **Zentralwert** genannt). Um ihn zu ermitteln, ordnet man zunächst die Stichprobenwerte ihrer Größe nach. Ist nun der Stichprobenumfang n eine ungerade Zahl, dann gibt es einen Wert in der Mitte dieser geordneten Liste, und das ist der Median. Ist der Stichprobenumfang n eine gerade Zahl, dann gibt es in der Mitte der Liste zwei Stichprobenwerte. Der Median ist dann das arithmetische Mittel dieser beiden Werte.

[5]Dazu muss in älteren Versionen ein Statistik-Paket (`DescriptiveStatistics`) geladen werden. In `Mathematica` werden n-Tupel mit geschlungenen Klammern {...} geschrieben.

Beispiel D.4. Median einer Stichprobe
a) Berechnen Sie den Median der Stichprobe aus Beispiel D.3.
b) Berechnen Sie den Median der Stichprobe ($1, 2, 2, 2, 3, 3, 3, 3, 3, 4$*).*

Lösung zu D.4.
a) Wir schreiben die Stichprobenwerte der Größe nach geordnet an und bestimmen den Wert in der Mitte

$$(1, 2, 2, 2, 3, \mathbf{3}, 3, 3, 3, 3, 4).$$

Der Median ist also 3.
b) Da die Anzahl der Stichprobenwerte nun 10*, also eine gerade Zahl ist, gibt es zwei Werte in der Mitte*

$$(1, 2, 2, 2, \mathbf{3}, \mathbf{3}, 3, 3, 3, 4).$$

Der Median ist das arithmetische Mittel dieser beiden Werte, also gleich 3.
Mit **Mathematica** *kann der Median folgendermaßen berechnet werden:*

```
In[3]:=Median[{1, 2, 2, 2, 3, 3, 3, 3, 3, 4}]
Out[3]=3
```

■

Das arithmetische Mittel berücksichtigt *alle* Stichprobenwerte, insbesondere auch „Ausreißer"; der Median hingegen ist unempfindlich gegenüber Ausreißern. Das statistische Zentralamt berechnet daher zum Beispiel für das mittlere Einkommen in Österreich nicht das arithmetische Mittel aller Einkommen, sondern das Medianeinkommen. Das ist also jener Wert, über bzw. unter dem genau die Hälfte aller erzielten Einkommen liegt. Im Unterschied zum arithmetischen Mittel gibt das Medianeinkommen einen unverzerrten sozialen Überblick über die Gehaltsstruktur, da die Ausreißer nach oben - also extrem hohe Gehälter - nicht berücksichtigt werden.

Weitere Lageparameter - als Verallgemeinerung des Medians - sind die sogenannten *Quantile*:

Wenn $x_1, \ldots, x_n$ die geordneten Stichprobenwerte sind, und $0 < p < 1$, so heißt

$$\tilde{x}_p = \begin{cases} x_{[np]+1} & , \quad np \notin \mathbb{Z} \\ \frac{1}{2}(x_{np} + x_{np+1}), & np \in \mathbb{Z} \end{cases} \tag{D.4}$$

das **p-Quantil**[6]. Ein p-Quantil zerlegt die geordneten Stichprobenwerte also in zwei beliebig große Teile. Wenn zum Beispiel $p = 0.15$, so gibt das $\tilde{x}_{0.15}$ die Stelle an, an der 15% der kleinsten Stichprobenwerte von den übrigen getrennt

[6]Dabei bezeichnet $[k]$ den ganzzahligen Anteil einer Zahl k (floor-Funktion).

werden. Häufig verwendet werden insbesondere $\tilde{x}_{0.25}$ und $\tilde{x}_{0.75}$, die gemeinsam mit dem Median als **Quartile** bezeichnet werden: $\tilde{x}_{0.25}$ ist das erste Quartil, (der Median ist in diesem Sinn das zweite Quartil) und $\tilde{x}_{0.75}$ das dritte Quartil. Durch die Quartile wird die geordnete Stichprobe also geviertelt.

Mit `Mathematica` erhält man das p-Quantil mit `Quantile[data, p]`.

Der **Modalwert**, das ist der am häufigsten auftretende Stichprobenwert: zum Beispiel in der Stichprobe „rot, rot, grün, blau, blau, blau“ ist „blau“ der Modalwert.

Als nächstes suchen wir ein geeignetes Streuungsmaß, das die „Breite“ der Stichprobe angibt.

Was könnten wir als Streuungsmaß verwenden? - Betrachten wir zum Beispiel die Stichproben A: $(1, 1, 5, 9, 9)$ und B: $(3, 4, 5, 6, 7)$. Beide haben denselben Mittelwert $\bar{x} = 5$, trotzdem streuen die Werte unterschiedlich um $\bar{x}$. Intuitiv würde man sagen, dass die Daten A stärker streuen, weil hier die Stichprobenwerte „im Durchschnitt“ weiter von $\bar{x} = 5$ entfernt sind als bei den Daten B. Versuchen wir, das durch eine Kennzahl auszudrücken:
a) Berechnen wir z.B. das arithmetische Mittel der *Differenzen* $x_i - \bar{x}$ der Stichprobenwerte vom Mittelwert so sieht man leicht, dass dies bei beiden Stichproben 0 ergibt. Die mittlere Differenz $\overline{x_i - \bar{x}}$ sagt also nichts über die Streuung aus.
b) Probieren wir als nächstes den mittleren *Abstand* der x_i von $\bar{x}$:

$$A: \frac{1}{5}[|1-5| + |1-5| + |5-5| + |9-5| + |9-5|] \quad = \quad 3.2$$

$$B: \frac{1}{5}[|3-5| + |4-5| + |5-5| + |6-5| + |7-5|] \quad = \quad 1.2$$

Mit diesem Maß streuen also die Daten A tatsächlich stärker als die Daten B.
c) Versuchen wir nun noch die mittlere quadratische Abweichung der x_i von $\bar{x}$:

$$A: \frac{1}{5}[(1-5)^2 + (1-5)^2 + (5-5)^2 + (9-5)^2 + (9-5)^2] \quad = \quad 12.8$$

$$B: \frac{1}{5}[(3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2] \quad = \quad 2.0$$

Also haben die Daten A auch mit diesem Maß eine größere Streuung als die Daten B. Es sprechen nun einige Gründe dafür, eher dieses Streuungsmaß zu verwenden als den mittleren Abstand:

- Es vermeidet das Rechnen mit Beträgen
- Quadrieren „bestraft“ große Abweichungen, da Quadrate von Zahlen > 1 größer sind als die ursprünglichen Zahlen.
- Wenn man einen optimalen „Repräsentanten“ a der Stichprobenwerte sucht in dem Sinn, dass $S(a) = (a - x_1)^2 + \cdots + (a - x_n)^2$ minimal sein soll, dann kann man leicht nachrechnen, dass dieses a gerade das arithmetische Mittel $\bar{x}$ ist.

Betrachten wir also eine Stichprobe $(x_1, \ldots, x_n)$ vom Umfang n. Ein Maß dafür, wie sehr die Stichprobenwerte x_i um ihren Mittelwert $\bar{x}$ streuen, ist die **mittlere**

quadratische Abweichung oder **(Stichproben-) Varianz**[7]

$$s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{D.5}$$

Man schreibt die Varianz symbolisch s^2, da man als Maß für die Streuung öfters auch ihre Wurzel

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{D.6}$$

verwendet, die empirische **(Stichproben-) Standardabweichung**. Sie hat den Vorteil, dass sie dieselbe Einheit hat wie die einzelnen Stichprobenwerte (z.B. Gramm anstelle von Gramm^2; man kann daher z.B. von einer Streuung von $s = 5$ Gramm sprechen). Wenn wir mit $x_1, \ldots, x_k$ die *verschiedenen* Werte dieser Stichprobe bezeichnen, dann kann man die Formel auch mithilfe der absoluten bzw. relativen Häufigkeiten anschreiben:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} h_i(x_i - \bar{x})^2 = \frac{n}{n-1}\sum_{i=1}^{k} f_i(x_i - \bar{x})^2. \tag{D.7}$$

Sie haben sich vielleicht gewundert, warum man bei s^2 durch $n-1$ dividiert anstelle durch n. Der Grund für diese Konvention ist, dass die durch $n-1$ dividierte Summe bessere *Schätzeigenschaften* für die schließende Statistik hat. Dort *schätzt* man die Varianz einer Grundgesamtheit mithilfe der Stichprobenvarianz s^2. Würde man die Stichprobenvarianz mit n im Nenner definieren, dann würde diese Schätzung *im Mittel zu klein* ausfallen.

Die Standardabweichung sagt nun folgendes aus: je kleiner s (bzw. s^2) ist, umso stärker sind die Messwerte um den Mittelwert $\bar{x}$ konzentriert. Man kann zeigen, dass s immer größer als der *mittlere Abstand* der Messwerte von $\bar{x}$ ist. Im Spezialfall, dass alle Messwerte gleich sind, ist s gleich 0.

Oft findet man die Formel

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\cdot\bar{x}^2\right).$$

Sie ist gleichwertig zur Formel oben, denn: $(n-1)\,s^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$. Bei dieser Rechnung wurde verwendet, dass $\sum_{i=1}^{n} x_i = n\bar{x}$ ist und dass $\sum_{i=1}^{n} \bar{x}^2 = n\bar{x}^2$ ist.

[7] Es wird aber auch von manchen die Definition mit Division durch n statt $n-1$ verwendet! `http://mathworld.wolfram.com/Variance.html`

Ein weiteres, eher grobes Streuungsmaß ist die **Spannweite** R (engl. *range*)

$$R = x_{max} - x_{min},$$

wobei x_{max} der größte und x_{min} der kleinste Stichprobenwert ist. R ist zwar leichter zu berechnen als s, enthält aber nicht so viel Information (R berücksichtigt nur den kleinsten und den größten, s hingegen *alle* Stichprobenwerte). Außerdem wird R durch „Ausreißer“ stärker beeinflusst.

Beispiel D.5. Standardabweichung und Spannweite einer Stichprobe
Berechnen Sie die Standardabweichung und die Spannweite der Stichprobe $(1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 7)$.

Lösung zu D.5. *Der Umfang der Stichprobe ist* $n = 11$*, das arithmetisches Mittel ist* $\bar{x} = 3$*. Die Varianz ist daher*

$$s^2 = \frac{1}{10}[2(1-3)^2 + 2(2-3)^2 + 4(3-3)^2 + 2(4-3)^2 + (7-3)^2] = 2.8$$

Der mittlere Abstand ist kleiner als $s = \sqrt{2.8} = 1.7$*, d.h., die Stichprobenwerte sind im Mittel um weniger als* 1.7 *von* $\bar{x} = 3$ *entfernt. Die Spannweite ist* $r = 7 - 1 = 6$*. Mit* `Mathematica` *erhält man die Varianz, Standardabweichung und Spannweite mit* `Variance`*,* `StandardDeviation` *bzw.* `Spannweite`*:*

```
In[4]:=Spannweite[data_] := Max[data] - Min[data];
       data = {1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 7};
       {Variance[data], StandardDeviation[data], SampleRange[data]}
Out[4]={14/5, Sqrt[14/5], 6}
```

■

Manchmal wird auch der **Interquartilsabstand** $d = q_3 - q_1$ als Streuungsmaß verwendet.

Eine sehr häufig verwendete graphische Darstellungsform von numerischen Daten (z.B. in der Medizin) ist der **Boxplot**. Dabei werden 5 Merkmale zusammengefasst und üblicherweise in vertikaler Form dargestellt. Dabei erstreckt sich ein Rechteck (Box) vom unteren Quartil zum oberem Quartil. Es enthält den Median markiert durch einen Strich. Weiters werden der maximale und der minimale Wert der Daten als weitere Striche eingezeichnet und durch dünne Linien (Wiskers) mit dem Rechteck verbunden[8].

Anbei ein Beispiel mit `Mathematica`. Gegeben sei die Liste der Körpergrößen einer Gruppe von Studierenden. Den Boxplot erhält man mit `BoxWhiskerChart`

[8] Für die Lage der Wiskers gibt es auch noch andere Definitionen.

```
In[5]:=data = {175, 180, 183, 192, 186, 175, 163, 169, 172, 180, 188, 178, 186,
       170, 166, 189, 178, 174, 164, 160, 176, 192, 174, 170, 170, 177, 172, 185,
       174, 172, 162, 184};
       BoxWhiskerChart[data]
Out[5]=
```

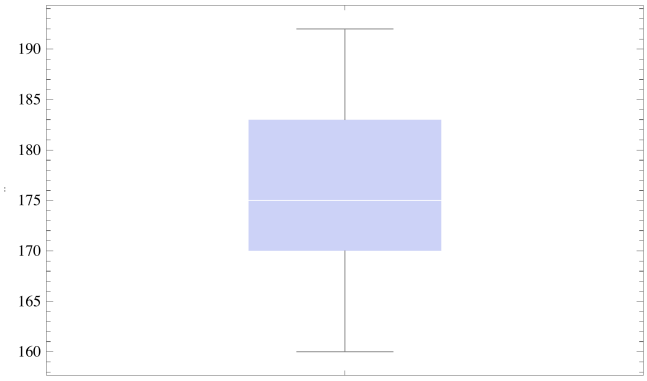


Abbildung D.4. Beispiel für den Boxplot der Körpergrößen einer Gruppe von Studierenden: Minimum = 160, 0.25-Quantil = 170, Median = 175 (bzw. 0.5-Quantil), 0.75-Quantil = 183, Maximum = 192.

D.3. Lineare Korrelation und Regression

In den bisherigen Beispielen wurde immer nur *ein* Merkmal der Stichprobe beobachtet (gemessen). Misst man zwei Merkmale in derselben Stichprobe, so kann man fragen, ob es zwischen diesen beiden Merkmalen einen **Zusammenhang** gibt.

Die Seitenlänge und die Fläche eines Quadrates stehen zum Beispiel in einem exakten Zusammenhang. Zwischen Werbeausgaben und Umsatz eines Unternehmens besteht auch ein Zusammenhang, dieser ist aber ein *statistischer Zusammenhang*: man kann sagen, dass ein Unternehmen *eher* mehr Umsatz machen wird, wenn es mehr für Werbung ausgibt, aber man kann den Zusammenhang zwischen diesen beiden Merkmalen nicht exakt durch eine Gleichung ausdrücken. Andere Beispiele für einen statistischen Zusammenhang sind: Gewicht und Größe einer Person (eine größere Person ist eher schwer), Außentemperatur und Verbrauch an Mineralwasser, usw.

Ein Gefühl dafür, ob es einen Zusammenhang zwischen zwei Merkmalen gibt, kann man bekommen, wenn man die Messwerte der beiden Merkmale graphisch

als Punkte in der (x, y)-Ebene darstellt. Die so erhaltene **Punktwolke** nennt man **Streudiagramm**.

Beispiel D.6. Streudiagramm von zwei zusammenhängenden Merkmalen
Von 15 *zufällig ausgewählten erwachsenen Personen wird deren Größe x und Gewicht y gemessen. Es ergeben sich folgende Wertepaare (x_i, y_i) in cm bzw. kg:* $((163, 59)$, $(165, 62)$, $(166, 65)$, $(169, 69)$, $(170, 65)$, $(171, 69)$, $(171, 76)$, $(173, 73)$, $(174, 75)$, $(175, 73)$, $(177, 80)$, $(177, 71)$, $(179, 82)$, $(180, 84)$, $(185, 81))$. *Stellen Sie diese Daten graphisch durch ein Streudiagramm dar.*

Lösung zu D.6. *Wir tragen in x-Richtung die Körpergröße auf und in y-Richtung das Gewicht. Die Daten der ersten Person werden daher durch den Punkt* $(163, 59)$ *dargestellt, usw.. Man erhält dann das Streudiagramm aus Abbildung D.5. Aus dieser Abbildung kann man gut die* Tendenz *„je größer, umso*

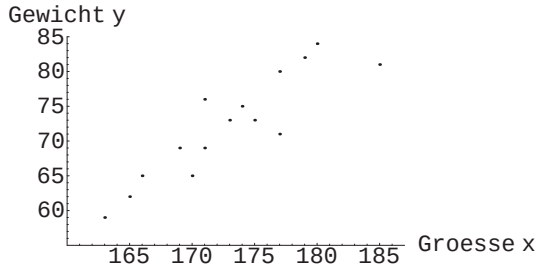


Abbildung D.5. Streudiagramm

schwerer“ herauslesen.

Mit `Mathematica` *erhält man dieses Diagramm leicht auf folgende Weise: die Stichprobendaten werden als Liste eingegeben (z.B. mit dem Variablennamen* `data`*)*

```
In[6]:=data = {{163, 59}, {165, 62}, {166, 65}, {169, 69}, {170, 65}, {171, 69},
       {171, 76}, {173, 73}, {174, 75}, {175, 73}, {177, 80}, {177, 71},
       {179, 82}, {180, 84}, {185, 81}};
```

Danach wird mit dem Befehl `ListPlot` *die Graphik in Abbildung D.5 erzeugt:*

```
In[7]:=ListPlot[data, PlotRange-> {{160, 187}, {55, 85}}]
```

(Hier wurde auch der dargestellte x, y-Bereich mit `PlotRange` *festgelegt.)* ■

Nun ein Beispiel für zwei Merkmale, von denen man nicht erwartet, dass sie zusammenhängen:

Beispiel D.7. Streudiagramm von zwei nicht-zusammenhängenden Merkmalen
Von 15 *zufällig ausgewählten erwachsenen Personen wird deren Größe x und ihr monatliches Einkommen y gemessen. Es ergeben sich folgende Wertepaare* (x_i, y_i) *in cm bzw. Euro :* $((163, 2900)$, $(165, 1100)$, $(166, 3600)$, $(169, 2300)$, $(170, 4000)$, $(171, 5600)$, $(171, 2100)$, $(173, 5100)$, $(174, 3400)$, $(175, 1800)$, $(177, 2100)$, $(177, 2600)$, $(179, 4600)$, $(180, 3600)$, $(185, 2300))$. *Stellen Sie diese Daten durch ein Streudiagramm dar.*

Lösung zu D.7. *Wir tragen nun in x-Richtung wieder die Körpergröße auf und in y-Richtung das monatliche Einkommen. Das Streudiagramm ist nun in Abbildung D.6 dargestellt. Die Punkte liegen ziemlich regellos verteilt. Es*

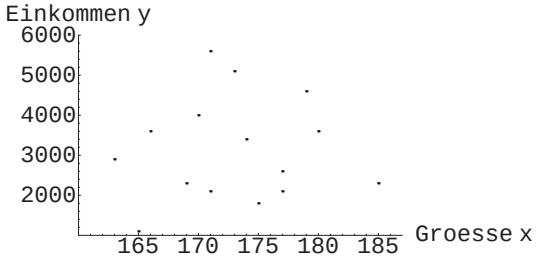


Abbildung D.6. Streudiagramm

bestätigt die Vermutung, dass es zwischen Größe und Einkommen einer Person keinen Zusammenhang gibt. ■

Ein letztes Beispiel:

Beispiel D.8. Streudiagramm von zwei „gegensinnig“ zusammenhängenden Merkmalen
Die Messung von zwei Merkmalen x bzw. y ergab folgende zehn Wertepaare (x_i, y_i)*:* $((30, 6)$, $(45, 5)$, $(45, 3)$, $(60, 4)$, $(75, 3)$, $(80, 5)$, $(90, 2)$, $(100, 4)$, $(110, 2)$, $(120, 3))$. *Stellen Sie diese Daten durch ein Streudiagramm dar.*

Lösung zu D.8. *Das Streudiagramm ist in Abbildung D.7 dargestellt. Es lässt eine gewisse Tendenz erkennen, dass mit größer werdenen x-Werten die y-Werte kleiner werden. Ein Beispiel für einen solchen „gegensinnigen“ Zusammenhang wäre Trainingszeit x und Laufzeit y eines Läufers.* ■

Im ersten Beispiel ließ das Streudiagramm einen deutlichen Zusammenhang vermuten, im zweiten Beispiel keinen, und im dritten einen gewissen, aber nicht

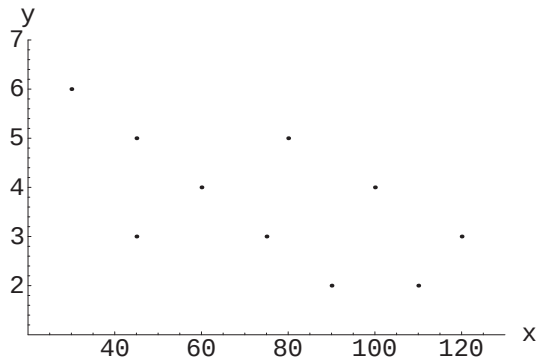


Abbildung D.7. Streudiagramm

starken Zusammenhang. Es wäre nun schön, wenn wir die Stärke des Zusammenhangs zwischen zwei Merkmalswerten einer Stichprobe durch eine Kennzahl ausdrücken könnten.

Betrachten wir also Paare von gemessenen Werten (x_i, y_i) einer Stichprobe. Man spricht hier auch von einer **zweidimensionalen Stichprobe**. Die x-Werte könnten z.B. die Körpergröße, die y-Werte das gemessene Gewicht von Personen darstellen. Über das Ausmaß des **linearen** Zusammenhangs gibt uns der sogenannte **(empirische) Korrelationskoeffizient** r Auskunft:

$$r = \frac{s_{xy}}{s_x \cdot s_y} \tag{D.8}$$

wobei

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{D.9}$$

und

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}, \qquad s_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2}. \tag{D.10}$$

Das sieht auf den ersten Blick ziemlich wild aus, ist aber nur halb so schlimm. Die Zahlen $\bar{x}$ und $\bar{y}$ sind die arithmetischen Mittelwerte, und die Zahlen s_x und s_y sind die (empirischen, d.h. die Stichproben-) Standardabweichungen der x_i-Werte bzw. der y_i-Werte. Neu ist nur die Zahl s_{xy}, die **(empirische) Kovarianz** genannt wird.

Man nennt den empirischen Korrelationskoeffizient r auch den **Bravais'schen oder Pearson'schen Korrelationskoeffizienten** (erfunden wurde er aber von F. Galton). Das Wort „empirisch“ drückt

aus, dass es sich um eine Kennzahl für Stichprobenwerte handelt. Die schließende Statistik verallgemeinert mithilfe der empirischen Kennwerte auf die entsprechenden Werte der Grundgesamtheit.

Man kann zeigen, dass r immer zwischen -1 und 1 liegt. Er sagt folgendes über den Zusammenhang der Merkmale in der Stichprobe aus:
• Je näher r bei -1 oder bei 1 liegt, umso besser liegen die Punkte (x_i, y_i) um eine Gerade konzentriert.
• Ist r dabei positiv, so werden die Punkte durch eine Gerade mit positiver Steigung beschrieben. Man spricht in diesem Fall von einem positiven linearen Zusammenhang oder einer **positiven linearen Korrelation**. Ist r negativ, dann werden die Punkte durch eine Gerade mit negativer Steigung beschrieben. Man spricht dann von einem negativen linearen Zusammenhang oder einer **negativen linearen Korrelation**.
• Im Spezialfall $r = 1$ liegen alle Punkte des Streudiagramms exakt auf einer Geraden mit positiver Steigung; im Spezialfall $r = -1$ liegen alle Punkte exakt auf einer Geraden mit negativer Steigung.
• Ist $r = 0$, so besteht kein *linearer* Zusammenhang zwischen den beiden Merkmalen in der Stichprobe.

Sehen wir uns gleich ein Beispiel dazu an:

Beispiel D.9. Korrelationskoeffizient
Berechnen Sie den Korrelationskoeffizienten der Stichproben aus
a) Beispiel D.8 b) Beispiel D.6 c) Beispiel D.7.

Lösung zu D.9.
a) Der Mittelwert der x_i-Werte ist $\bar{x} = 75.5$, der Mittelwert der y_i-Werte ist $\bar{y} = 3.7$; die Standardabweichung der x_i-Werte ist $s_x = 30.134$, die der y_i-Werte ist $s_y = 1.3375$. Die Kovarianz s_{xy} ist

$$\begin{aligned} s_{xy} &= \frac{1}{9}\sum_{i=1}^{10}(x_i - 75.5)(y_i - 3.7) \\ &= \frac{1}{9}[(30 - 75.5)(6 - 3.7) + (45 - 75.5)(5 - 3.7) + \cdots + (120 - 75.5)(2 - 3.7)] \\ &= -25.3889 \end{aligned}$$

Damit ist der Korrelationskoeffizient

$$r = \frac{-25.3889}{30.134 \cdot 1.3375} = -0.6299.$$

Er ist betragsmäßig nicht sehr nahe bei 1, was einen gewissen, aber nicht starken linearen Zusammenhang der Stichprobenwerte x_i und y_i bedeutet. Das negative Vorzeichen von r weist darauf hin, dass tendenziell mit zunehmenden x_i-Werten die y_i-Werte fallen (siehe Abbildung D.7).

b) Nun lassen wir uns lieber von **Mathematica** *helfen und erhalten dann mit* **Correlation** *den empirischen Korrelationskoeffizienten:*

```
In[8]:=xdata = {163, 165, 166, 169, 170, 171, 171, 173, 174, 175, 177,
       177, 179, 180, 185};
       ydata = {59, 62, 65, 69, 65, 69, 76, 73, 75, 73, 80, 71, 82, 84, 81};
       Correlation[xdata, ydata]//N
Out[8]=0.897914
```

Wir erhalten also eine stärkere Korrelation als im Beispiel a) (da der Wert von r betragsmäßig näher bei 1 *liegt). Das positive Vorzeichen von r sagt aus, dass tendenziell mit wachsenden x_i-Werten auch die y_i-Werte zunehmen (vergleiche Abb. D.5).*
c) Gleich mit **Mathematica**:

```
In[9]:=xdata = {163, 165, 166, 169, 170, 171, 171, 173, 174, 175, 177,
       177, 179, 180, 185};
       ydata = {2900, 1100, 3600, 2300, 4000, 5600, 2100, 5100, 3400, 1800,
       2100, 2600, 4600, 3600, 2300};
       Correlation[xdata, ydata]//N
Out[9]=0.0606408
```

Der Korrelationskoeffizient liegt nun nahe bei 0. *Das weist darauf hin, dass die beiden Merkmale nicht (linear) korreliert sind (wie das zugehörige Streudiagramm in Abb. D.6 nahelegt).* ■

Aber Achtung: der Korrelationskoeffizient mißt nur den Grad der *linearen* Abhängigkeit. Wenn $r = 0$, so bedeutet das nicht, dass kein Zusammenhang zwischen den Werten x_i und y_i besteht! Zeichnen Sie etwa das Streudiagramm, wenn folgende Stichprobe vorliegt: $((-2, 5)$, $(-1, 2)$,$(0, 1)$, $(1, 2)$, $(2, 5))$. Diese Punkte liegen alle auf der Parabel $y = x^2 + 1$. Es besteht hier also sehr wohl ein Zusammenhang zwischen den x_i und den y_i, dieser Zusammenhang ist aber *nichtlinear*.

Umgekehrt bedeutet eine positive Korrelation aber nicht, dass ein kausaler Zusammenhang besteht. Zum Beispiel ergab eine Untersuchung eine starke Korrelation zwischen der Geburtenrate und der Storchenpopulation[9]. Der Schluss, dass der Storch nun die Kinder bringt ist aber nicht richtig! Dies ist hier offensichtlich, aber wie steht es damit bei anderen Korrelationen?

[9]Hier ein link dazu `https://www.researchgate.net/publication/227763292_Storks_Deliver_Babies`

D.3.1. Autokorrelation. Untersucht man die Korrelation zwischen zwei verschiedenen Zeitpunkten einer Zeitreihe so erhält man die Autokorrelation[10].

D.3.2. Regressionsgerade. Zuletzt wäre es schön, wenn wir eine Gerade angeben könnten, die z.B. die Stichprobenwerte aus Beispiel D.6 möglichst gut annähert. Damit kommen wir zu einer Aufgabe der Regressionsrechnung.

Aufgabe von linearen **Korrelationsanalysen** ist es, das *Ausmaß* des linearen Zusammenhangs zwischen zwei Merkmalen x und y zu untersuchen. Dabei sieht man die beiden Merkmale als *gleichrangig* an, d.h., sowohl gemessene x-Werte als auch gemessene y-Werte können streuen. x und y sind sogenannte *Zufallsvariablen*. Mithilfe von konkreten Stichprobenwerten (x_i, y_i) und dem zugehörigen empirischen Korrelationskoeffizient wird der lineare Zusammehang der beiden Merkmale x und y *geschätzt*.

Bei sogenannten **Regressionsanalysen** untersucht man die *Art* des Zusammenhangs zwischen zwei Merkmalen x und y. Dabei sind x und y nicht mehr gleichrangig, sondern man betrachtet y als abhängig von x. Man geht davon aus, dass x festgehalten und exakt messbar ist und nur die y-Werte streuen. Zum Beispiel ist x eine bestimmte Körpergröße, und y sind die verschiedenen Gewichte von Personen mit dieser Größe. Man interessiert sich nun dafür, wie sich y in Abhängigkeit von x ändert. Dazu beobachtet man zu vorgegebenen Werten $x_1, \ldots, x_n$ die Werte $y_1, \ldots, y_n$. Nimmt man nun einen *linearen Zusammenhang* zwischen x und y an, so ermittelt man aus der Stichprobe die **empirische Regressionsgerade** $y = ax + b$ und schätzt damit den durchschnittlichen Einfluß von x auf y.

Betrachten wir allgemein die folgende Aufgabenstellung: Es wird eine Größe y zu verschiedenen (vorgegebenen) Werten einer Größe x gemessen:

$$((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)).$$

Gesucht ist eine **Regressions- oder Ausgleichskurve**, die die Abhängigkeit der y_i-Werte von den x_i-Werten möglichst gut beschreibt.

Dazu geht man am besten von der durch die Stichprobe gegebenen Punktwolke aus und macht einen geeigneten Ansatz für den funktionellen Zusammenhang zwischen x und y. Wenn man versucht, die Punkte (x_i, y_i) durch eine *Gerade* anzunähern

$$y = f(x) = kx + d$$

dann spricht man von **linearer Regression** und nennt diese Gerade **Ausgleichsgerade** oder **Regressionsgerade**. Wie soll man nun am besten die Steigung k und den Achsenabschnitt d der Regressionsgeraden wählen? Man kann diese Parameter so wählen, dass der **mittlere quadratische Fehler**

[10]Siehe: `http://de.wikipedia.org/wiki/Autokorrelation`

(d.h., die mittlere quadratische Abweichung der Messwerte y_i von den Funktionswerten $f(x_i)$) minimal wird:

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 = \text{Minimum}$$

Das ist die **Gaußsche Methode der kleinsten Quadrate**. Daraus ergeben sich folgende Formeln für k und d:

$$k = r\frac{s_y}{s_x} = \frac{s_{x\,y}}{s_{x\,x}}, \qquad d = \bar{y} - k\bar{x}. \tag{D.11}$$

Hier ist r der empirische Korrelationskoeffizient, $\bar{x}$, $\bar{y}$ sind die arithmetischen Mittelwerte und s_x, s_y die Standardabweichungen der Stichprobenwerte x_i bzw. y_i.

Sehen wir uns das anhand eines Beispiels an: Gegeben sind die Datenpaare $(1,1)$, $(2,3)$, $(3,4)$, $(4,2)$, $(5,4)$. Finden Sie jene Gerade $f(x) = k\,x + d$, für die der mittlere quadratische Fehler minimal wird. Lösung: Der quadratische Fehler ist gegeben durch

$$\sum_{i=1}^{n}(y_i - kx_i + d)^2 = 72.88 - 34.4d + 5d^2 - 126.4k + 30dk + 55k^2.$$

Die Frage ist nun, wie wir das Minimum finden sollen? Wenn wir den Parameter d festhalten, dann erhalten wir das Minimum als Nullstelle der ersten Ableitung nach k: $-126.4 + 30d + 110k = 0$ also $k = 1.14909 - 0.272727d$. Setzen wir diesen Wert nun in den quadratischen Fehler ein

$$0.257455 + 0.0727273d + 0.909091d^2$$

und bestimmen wiederum das Minimum, so erhalten wir $d = -0.04$ und damit $k = 1.16$. Die Datenpaare und die zugehörige Gerade sind in Abbildung D.8 dargestellt.

Beispiel D.10. Regressionsgerade
Geben Sie die Regressionsgerade für die Stichprobe aus Beispiel D.6 an.

Lösung zu D.10. *Wir müssen k und d der Regressionsgerade $y = kx + d$ berechnen. Mit* `Mathematica` *oder der Hand berechnen wir den Mittelwert $\bar{x} = 173$ der x_i-Werte und den Mittelwert $\bar{y} = 72.3$ der y_i-Werte. Die Standardabweichungen sind $s_x = 6.05$ bzw. $s_y = 7.56$. Den Korrelationskoeffizienten haben wir bereits in Beispiel D.9 berechnet: $r = 0.898$. Damit ist*

$$\begin{aligned} k &= r\frac{s_y}{s_x} = 1.12, \\ d &= \bar{y} - k\bar{x} = -122.02. \end{aligned}$$

Die Regressionsgerade lautet also $y = 1.12x - 122.02$. Sie ist in Abbildung D.8 dargestellt.

Mit `Mathematica` *kann man die Gerade mit dem Befehl* `Fit` *berechnen:*

```
In[10]:=data = {{163, 59}, {165, 62}, {166, 65}, {169, 69}, {170, 65}, {171, 69},
        {171, 76}, {173, 73}, {174, 75}, {175, 73}, {177, 80}, {177, 71},
        {179, 82}, {180, 84}, {185, 81}};
        g = Fit[data, {1, x}, x]
Out[10]=-122.02 + 1.12305x
```

Die Syntax ist `Fit[Daten, Funktionen, Variablen]`*. Dadurch werden* `Daten` *mithilfe der Gaußschen Methode der kleinsten Quadrate durch eine Linearkombination der* `Funktionen` *genähert. Hier möchten wir eine Gerade, also eine Linearkombination der Funktionen* 1 *und* x*.*

Die graphische Darstellung von Streudiagramm und Ausgleichsgerade erhält man zum Beispiel mit

```
In[11]:=g1 = ListPlot[data, PlotRange-> {{160, 187}, {55, 85}}];
        g2 = Plot[g, {x, 160, 190}];
        Show[g1, g2];
```

Die Ausgabe der Graphiken `g1` *bzw.* `g2` *kann mithilfe der Option* `DisplayFunction` *unterdrückt werden.*

■

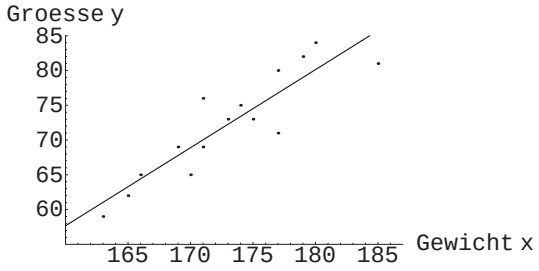


Abbildung D.8. Regressionsgerade

In vielen Fällen legt die Punktwolke einen *nichtlinearen* Ansatz nahe, zum Beispiel eine Parabel $y = ax^2 + bx + c$, Polynomfunktionen höheren Grades, oder Exponential- und Logarithmusfunktionen. Auch dann werden die im Ansatz enthaltenen Parameter mithilfe der Gaußschen Methode der kleinsten Quadrate bestimmt.

D.4. Übungen

1. Die Urliste einer Stichprobe lautet (Augenzahl beim Wurf eines Würfels): $(4, 3, 6, 5, 6, 1, 2, 3, 3, 4, 2, 1, 4, 2, 6)$. Geben Sie die absoluten und die relativen Häufigkeiten an.

2. Die Urliste einer Stichprobe lautet (Abfüllgewicht einer Packung Gummibären in Gramm): $(101, 101, 100, 101, 98, 99, 103, 100, 100, 99)$.
   a) Geben die absoluten und die relativen Häufigkeiten der Stichprobenwerte an
   b) Welcher Anteil der Proben hat ein Gewicht unter 100 Gramm?
   c) Welcher Anteil hat ein Gewicht zwischen 99 und 101 Gramm?

3. Die Messung der Lebensdauer von 30 Glühbirnen ergab folgende Werte:

   Lebensdauer einer Glühbirne (in 100 Tagen)

46	47	48	51	53	57	58	58	59	59
60	60	61	61	62	62	63	63	64	64
66	66	68	69	71	71	75	77	81	83

   Bilden Sie Klassen, bestimmen Sie deren relative und absolute Häufigkeiten und zeichnen Sie ein Histogramm.

4. Gegeben ist die folgende (bereits geordnete) Urliste einer Stichprobe: $(1, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 8, 9, 9, 10)$.
   Bilden Sie Klassen, bestimmen Sie deren relative und absolute Häufigkeiten und zeichnen Sie ein Histogramm. Wie ändert sich das Histogramm, wenn Sie zwei Klassen zusammenlegen?

5. Schalter eines bestimmten Typs wurden auf ihre Lebensdauer geprüft. Dabei wurde folgende Urliste ermittelt (Anzahl der Betätigungen bis zum Ausfall in 10^3): (28.4, 53.9, 34.5, 16.2, 41.0, 11.4, 36.9, 12.7, 18.3, 40.0, 32.9, 26.5).
   Ermitteln Sie das arithmetische Mittel, den Median, die Standardabweichung und die Spannweite.

6. Drei Personen entnehmen derselben Lieferung von Bauteilen je eine Stichprobe mit folgenden Stichprobenumfängen und arithmetischen Mittelwerten (in cm): a) $h_1 = 20$, $\bar{x} = 73$ b) $h_2 = 30$, $\bar{x} = 75$ c) $h_3 = 50$, $\bar{x} = 74$.
   Die drei Stichproben werden zu einer einzigen vereinigt. Wie groß ist ihr Mittelwert?

7. Gegeben sind die Stichprobenwerte $(1, 2, 3, 4, 5, 5)$. Berechnen Sie jenen stellvertretenden Wert a für diese Werte, auf den bezogen die mittlere quadratische Abweichung $S(a) = \frac{1}{6}[(1-a)^2 + (2-a)^2 + (3-a)^2 + (4-a)^2 + 2(5-a)^2]$ minimal ist.

8. Bei 15 PKWs desselben Typs wurde der Benzinverbrauch pro 100km gemessen. Dabei ergab sich folgende Urliste (Liter pro 100 km):
   (10.8, 9.9, 10.2, 10.4, 11.1, 9.3, 9.1, 10.4, 10.1, 11.7, 10.9, 10.4, 10.8, 11.3, 11.5).

a) Berechnen Sie das arithmetische Mittel, den Median, die Standardabweichung und die Spannweite.
b) Welcher Prozentsatz hatte einen Benzinverbrauch von höchstens 10.5 Liter?
c) Welcher Benzinverbrauch wurde von 50% der PKW nicht überschritten?

9. Skizzieren Sie die folgenden zweidimensionalen Stichproben und bestimmen Sie den Korrelationskoeffizienten:
a) $((2,2), (3,1), (5,3), (6,4))$ b) $((1,5), (2,2), (3,1), (4,2), (5,5))$
c) $((1,5), (2,3), (3,4), (5,2))$

10. Bei 10 PKWs wurden Gewicht und Benzinverbrauch pro 100km gemessen. Es ergaben sich folgende Daten (Tonnen, Liter): $((1.5, 7.7), (1.8, 9.1), (1.4, 8.3), (2.2, 10.0), (1.3, 7.7), (1.7, 8.3), (1.5, 9.1), (1.7, 8.3), (1.4, 8.3), (1.2, 7.1))$. Bestimmen Sie den Korrelationskoeffizienten.

11. Bei 5 Frauen und 5 Männern wurden Schuhgröße und Einkommen ermittelt: Bei den Frauen ergaben sich folgende (vereinfachte) Daten F: $((1,4), (2,6), (2,3), (3,5), (3,3))$; die Daten der Männer waren M: $((4,6), (5,8), (5,4), (6,7), (6,5))$.
a) Stellen Sie die Stichprobendaten M und F im selben Streudiagramm dar, markieren Sie aber, welcher Punkte zu F und welche zu M gehören.
b) Ermitteln Sie die Korrelationskoeffizienten von M von F.
c) Vereinigen Sie nun die Stichprobendaten M und F zu einer einzigen Stichprobe und ermitteln Sie nun den Korrelationskoeffizienten. Bedeutet das Ergebnis, dass es eine Korrelation zwischen Schuhgröße und Einkommen gibt?

12. Bestimmen Sie die Gleichung der Regeressionsgeraden für die Daten aus Aufgabe 9, die einen linearen Zusammenhang nahelegen.

13. Bestimmen Sie die Gleichung der Regressionsgerade für die Daten aus Aufgabe 10.

14. Vier Zimmerleute Anton, Bernd, Clemens und David zimmern eine Schalung. Alleine würde Anton 12 Tage, Bernd 15 Tage, Clemens 30 Tage und David 18 Tage brauchen. Wie lange brauchen sie gemeinsam?

Literaturverzeichnis

(i) **Bücher die die mathematischen Vorkenntnisse behandeln**

[1] J. Arrenberg, M. Kiy, R. Knobloch, *Vorkurs in Mathematik*, Oldenbourg, München 2000.

[2] K. Bosch, *Brückenkurs Mathematik*, 9. Aufl. Oldenbourg, München 2000.

[3] K. Fritzsche, *Mathematik für Einsteiger. Vor- und Brückenkurs zum Studienbeginn*, Spektrum Akad. Verlag, Heidelberg 1995.

[4] A. Kemnitz, *Mathematik zum Studienbeginn*, Vieweg, Wiesbaden 2000.

[5] K. Marti, D. Gröger, *Brückenkurs Mathematik*, Peikert, Wittenberg 1999.

[6] F. Reinhardt, *dtv-Atlas Schulmathematik*, dtv-Verlag, 2002.

[7] W. Scharlau, *Schulwissen Mathematik: Ein Überblick*, 3. Aufl. Vieweg, Wiesbaden 2001.

[8] W. Schäfer, K. Georgi, G. Trippler, C. Otto, *Mathematik–Vorkurs. Übungs- und Arbeitsbuch für Studienanfänger*, Teubner, Leipzig 1999.

[9] W. Schirotzek, S. Scholz, *Starthilfe Mathematik*, Teubner, Leipzig 1999.

(ii) **Allgemeine Mathematikbücher für Ingenieure**

[10] R. Ansorge, H. J. Oberle, *Mathematik für Ingenieure, 1–3*, Wiley/VCH, Weinh. 1994.

[11] T. Arens, F. Hettlich, C. Karpfinger, U. Kockelkorn, K. Lichtenegger, H. Stachel *Mathematik*, Spektrum, Akad. Verlag, Heidelberg 2008

[12] G. Bärwolff, G. Seifert, *Höhere Mathematik für Naturwissenschaftler und Ingenieure*, Spektrum, Akad. Verlag, Heidelberg 2004.

[13] W. Brauch, H. J. Dreyer, W. Haacke, *Mathematik für Ingenieure*, 9. Aufl., Teubner, Stuttgart 1995.

[14] K. Burg, H. Haf, F. Wille, *Höhere Mathematik für Ingenieure, 1–5,* Teubner, Stuttgart 1997.

[15] A. Fetzer, H. Fränkel, *Mathematik 1–2*, Lehrbuch für ingenieurwissenschaftliche Studiengänge, Springer, Berlin, Heidelberg 1997.

[16] K. Finck von Finckenstein, *Grundkurs Mathematik für Ingenieure*, Teubner, Stuttgart 1991.

[17] K. Jänich, *Mathematik 1, 2*, Springer, Berlin 2001.

[18] E. Kreyszig, *Advanced Engineering Mathematics*, 8th Ed. John Wiley, New York 1999.

[19] L. Papula, *Mathematik für Ingenieure, 1–3*. Ein Lehr- und Arbeitsbuch für das Grundstudium, Vieweg, Braunschweig, Wiesbaden 1991.

[20] T. Rießinger, *Mathematik für Ingenieure*, Springer, Berlin, Heidelberg 1996.

[21] P. Stingl, *Mathematik für Fachhochschulen Technik und Informatik*, Hanser, München 1999.

[22] U. Storch, H. Wiebe, *Lehrbuch der Mathematik 1-4*, Spektrum, Akad. Verlag, Heidelberg 1993 - 2001.

(iii) **Mathematikbücher für Informatiker**

[23] G. Baron, P. Kirschenhofer, *Einführung in die Mathematik für Informatiker, 1–3,* Springer, Wien 1989.

[24] G. Berendt, *Mathematik für Informatiker*, Spektrum Akad., Verlag, Heidelberg 1994.

[25] M. Brill, *Mathematik für Informatiker*, Hanser, München, Wien 2001.

[26] W. Dörfler, W. Peschek, *Einführung in die Mathematik für Informatiker*, Hanser, München, Wien 1988.

[27] P. Hartmann, *Mathematik für Informatiker*, Vieweg, Braunschweig, Wiesbaden 2002.

[28] K. H. Kiyek, F. Schwarz, *Mathematik für Informatiker, 1–2,* Teubner, Stuttgart 1989.

[29] W. Oberschelp, D. Wille, *Mathematischer Einführungskurs für Informatiker*, Teubner, Stuttgart 1976.

[30] G. Teschl, S. Teschl, *Mathematik für Informatiker, Teil 1: Diskrete Mathematik und Lineare Algebra*, Springer, Wien 2005.

[31] G. Teschl, S. Teschl, *Mathematik für Informatiker, Teil 2: Analysis und Statistik*, Springer, Wien 2006.

[32] M. Wolff, P. Hauck, W. Küchlin, *Mathematik für Informatik und BioInformatik*, Springer, Berlin 2004.

(iv) **Diskrete Mathematik, Algebra und Zahlentheorie**

[33] M. Aigner, *Diskrete Mathematik*, 4. Aufl. Vieweg, Braunschweig, Wiesbaden 2001.

[34] I. Anderson, *A first Course in Discrete Mathematics*, Springer, London 2001.

[35] A. Bartolomé, J. Rung, H. Kern, *Zahlentheorie für Einsteiger*, 3. Aufl. Vieweg, Braunschweig, Wiesbaden 2001.

[36] A. Beutelspacher, *Diskrete Mathematik für Einsteiger*, Vieweg, Braunschweig, Wiesbaden 2002.

[37] N. Biggs, *Discrete Mathematics*, Oxford University Press 1992.

[38] J. Clark, D. A. Holton, *Graphentheorie*, Spektrum, Akad. Verl., Heidelberg, 1991.

[39] R. Diestel, *Graphentheorie*, 2. Aufl. Springer, Berlin, 2000. Online auf

[40] H. Ehrig, B. Mahr, F. Cornelius, M. Grode-Rhode, P. Zeitz, *Mathematisch-strukturelle Grundlagen der Informatik*, 2. Aufl. Springer, Berlin 2001.

[41] O. Forster, *Algorithmische Zahlentheorie*, Vieweg, Braunschweig, Wiesbaden 1996.

[42] R. Garnier, J. Taylor, *Discrete Mathematics for New Technology*, Institute of Physics Publishing, Bristol, Philadelphia 1992.

[43] J. L. Hein, *Discrete Structures, Logic, and Computability*, 2nd ed. Jones and Bartlett Publishers, Sudbury 2002.

[44] K. Jacobs, D. Jungnickel, Einführung in die Kombinatorik 2. Aufl. de Gruyter, Berlin 2004.

[45] D. Jungnickel, *Graphen, Netzwerke und Algorithmen*, 3. Aufl. BI-Wiss.-Verl., Mannheim, 1994.

[46] R. Graham, D. Knuth, O. Patashnik, *Concrete Mathematics, A Foundation for Computer Science* Addison Wesley 1994.

[47] O. Körner, *Algebra*, 2. Aufl. Aula-Verlag, Wiesbaden 1990.

[48] S. Lipschutz, M. Lipson, *Schaum's Outline of Theory and Problems of Discrete Mathematics* (Schaum's Outline Series), McGraw-Hill Companies, New York 1997.

[49] L. Lovász, J. Pelikán, *Diskrete Mathematik*, Springer, Berlin 2005.

[50] J. Matoušek, J. Nešetřil, *Diskrete Mathematik*, Springer, Berlin 2002.

[51] S. B. Maurer, A. Ralston, *Discrete Algorithmic Mathematics*, 2nd ed., A K Peters, Natik, Massachusetts 1998.

[52] W. Nehrlich, *Diskrete Mathematik - Basiswissen für Informatiker. Eine Mathematica-gestützte Darstellung*, Fachbuchverlag Leipzig, Carl Hanser Vlg. München, 2003.

[53] S. Pemmaraju and S. Skiena, *Computational Discrete Mathematics: Combinatorics and Graph Theory with Mathematica*, Cambridge University Press, New York, 2003.

[54] K. Rosen, *Discrete Mathematics and its Application*, McGraw-Hill, Boston 1988.

[55] H. Scheid, *Zahlentheorie*, 2. Aufl. BI Verlag, Mannheim, Leipzig 1994.

[56] A. Steger, *Diskrete Strukturen 1*, Springer, Berlin 2001.

[57] J. K.Truss, *Discrete Mathematics for Computer Scientists*, 2nd ed., Addison Wesley 1999.

[58] V. Turau, *Algorithmische Graphentheorie*, 2. Aufl. Addison-Wesley, Bonn, 1996.

(v) **Lineare Algebra**

[59] B. Artmann, *Lineare Algebra*, Birkhäuser 1991.

[60] A. Beutelspacher, *Lineare Algebra*, Vieweg, Wiesbaden 2001.

[61] E. Brieskorn, *Lineare Algebra und analytische Geometrie 1,2*, Vieweg, Wiesbaden 1983.

[62] T. Bröcker, *Lineare Algebra und analytische Geometrie*, Birkhäuser, Basel 2003.

[63] G. Farin, D. Hansford *Lineare Algebra: Ein geometrischer Zugang*, Springer, Berlin 2003.

[64] G. Fischer, *Lineare Algebra*, Vieweg, Wiesbaden 1997.

[65] P. R. Halmos, *Finite-Dimensional Vector Spaces*, Springer, Berlin 2000.

[66] K. Jänich, *Lineare Algebra*, 8. Aufl. Springer, Berlin 2000.

[67] M. Koecher, *Lineare Algebra und analytische Geometrie*, Springer, Berlin 1997.

[68] H-J. Kowalsky, G. Michler, *Lineare Algebra*, 11. Aufl. de Gruyter, Berlin 1998.

[69] A. Meister, *Numerik linearer Gleichungssysteme*, Vieweg, Wiesbaden 1999.

[70] H. Möller, *Algorithmische lineare Algebra. Eine Einführung für Mathematiker und Informatiker*, Vieweg, Wiesbaden 1997.

[71] F. Lorenz, *Lineare Algebra 1,2*, 3. Aufl. Spektrum Akad. Verlag, Heidelberg 1993.

[72] B. Pareigis, *Lineare Algebra für Informatiker*, Springer, Berlin 2000.

[73] W. Preuß, G. Wenisch, *Lehr- und Übungsbuch Mathematik für Informatiker*, Lineare Algebra und Anwendungen, Fachbuchverlag Leipzig 1997.

[74] U. Stammbach, *Lineare Algebra*, Teubner Verlag, Stuttgart, 4. Aufl. 1994.

[75] W. Strampp, *Lineare Algebra mit Mathematica und Maple*, Vieweg, Wiesbaden 1999.

[76] G. Strang, *Lineare Algebra*, Springer, Berlin 2003.

[77] R. Zurmühl, S. Falk, *Matrizen und ihre Anwendungen 1*, 7. Aufl. Springer, Berlin 1997.

(vi) **Analysis**

[78] H. Amann, J. Escher, *Analysis 1, 2*, Birkhäuser, Basel 1998.

[79] M. Barner, F. Flor, *Analysis 1, 2*, 5. Aufl. de Gruyter, Berlin 2000.

[80] C. Blatter, *Ingenieur Analysis 1, 2*, 2. Aufl. Springer, Berlin 1996.

[81] T. Bröcker, *Analysis 1, 2, 3*, 2. Aufl. Spektrum Akad. Verlag, Heidelberg 1999.

[82] K. Endl, W. Luh, *Analysis 1, 2, 3*, 9. Aufl. AULA Verlag, Wiesbaden 1989.

[83] O. Forster, *Analysis 1, 2, 3*, 6. Aufl. Vieweg, Wiesbaden 2001.

[84] H. Heuser, *Lehrbuch der Analysis 1,2*, 13. Aufl. Teubner, Stuttgart 2000.

[85] W. Kaballo, *Einführung in die Analysis 1,2*, 2. Aufl. Spektrum Akad. Verlag, Heidelberg 2000.

[86] K. Königsberger, *Analysis 1, 2*, 5. Aufl. Springer, Berlin 2000.

[87] S. Lang, *Undergraduate Analysis*, Springer, Telos 1997.

[88] H. Neunzert, et. al. *Analysis 1, 2*, 3. Aufl. Springer, Berlin 1998.

[89] W. Preuß, G. Wenisch, *Lehr- und Übungsbuch Mathematik, Band 2: Analysis*, Fachbuchvlg. Leipzig 2000.

[90] W. Rudin, *Analysis*, 3. Aufl. Oldenburg, München, Wien 2005.

[91] M. Spivak, *Calculus*, W. A. Benjamin, New York, Amsterdam 1967.

[92] W. Walter, *Analysis 1, 2*, 5. Aufl. Springer, Berlin 1999.

(vii) **Differentialgleichungen**

[93] L. Edelstein-Keshet, *Mathematical Models in Biology*, SIAM 2005.

[94] L. C. Evans, *Partial Differential Equations*, 2nd ed. American Mathematical Society, 2010.

[95] S. H. Strogatz, *Nonlinear Dynamics and Chaos*, Westview Press, 2000.

[96] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*, American Mathematical Society, 2012.

[97] W. Walter, *Gewöhnliche Differentialgleichungen*, 7. Aufl. Springer, Berlin 2000.

(viii) **Mathematiklexika**

[98] I. N. Bronstein, K. A. Semendjajew, G. Musiol, *Taschenbuch der Mathematik*, inkl. CDROM, Harri Deutsch, Frankfurt, Thun 1999.

[99] *dtv-Atlas zur Mathematik, 1–2*, dtv, München 1974.

(ix) **Sonstiges**

[100] G. Asser *Einführung in die mathematische Logik, Teil I*, 6. Aufl. Harri Deutsch, Frankfurt, Thun 1983.

[101] A. Beutelspacher *Geheimsprachen. Geschichte und Techniken*, 2. Aufl. C. H. Beck, München 2000.

[102] J. Buchmann, *Einführung in die Kryptographie*, Springer, Berlin, 1999.

[103] H-J. Bungartz, M. Griebel, Z. Zenger, *Einführung in die Computergraphik*, 2. Aufl. Vieweg, Wiesbaden 2002.

[104] W. Ertel, *Angewandte Kryptographie*, Fachbuchverlag Leibzig im C. Hanser Verlag, München, Wien 2001.

[105] G. Farin, *Curves and Surfaces for CAGD: A Practical Guide*, 5. Aufl. Morgan Kaufmann Publishers 2001.

[106] J. Foley, A. Dam, S. Feiner, J. Hughes, *Computer Graphics*, 2nd Ed. Addison Wesley, 1995.

[107] P. Gritzmann, R. Brandenberg, *Das Geheimnis des kürzesten Weges*, Springer, Berlin 2001.

[108] R. Maeder, *Computer Science with Mathematica*, Cambridge University Press 2000.

[109] D. Marsh, *Applied Geometry for Computer Graphics and CAD*, 2. Aufl. Springer, London 2004.

[110] T. Sonar, *Angewandte Mathematik, Modelbildung und Informatik*, Vieweg, Braunschweig, Wiesbaden 2001.

[111] A. Tarski, *Einführung in die mathematische Logik*, 5. Aufl., Vandenhoeck & Ruprecht, Göttingen 1977.

[112] A. Watt, *3D-Computergrafik*, 3. Aufl. München, Pearson Studium, 2002.

[113] M. Werner, *Information und Codierung*, Vieweg, Wiesbaden 2002.

[114] R. Wobst, *Abenteuer Kryptologie*, 2. Aufl. Bonn, Addison-Wesley, 1998.