Mathematik I (i933)

Jens Flemming (HTW Dresden)

15.11.2026

Inhaltsverzeichnis

0 Vorwort . . . 1
1 Motivation . . . 2
1.1 Mathematik vs. Rechnen . . . 2
1.2 Mathematik als Sprache . . . 2
1.3 Computer . . . 3
1.4 “KI” . . . 3
1.5 Mittel gegen Komplexität . . . 4
1.6 “Soft-Skills” . . . 5
1.7 Mathematik im Studium . . . 6
2 Schulmathematik . . . 9
2.1 Online-Kurs OMB+ . . . 9
2.2 Themen, die als bekannt vorausgesetzt werden . . . 9
2.3 Themen, die wiederholt und vertieft werden . . . 10
2.4 Nicht behandelte Themen . . . 10
3 Computereinsatz . . . 11
3.1 Überblick . . . 12
3.2 Python . . . 16
3.3 SymPy . . . 24
3.4 NumPy . . . 27
4 Elementare Begriffe . . . 32
4.1 Aussagen . . . 33
4.2 Abbildungen . . . 40
4.3 Zahlenfolgen . . . 46
4.4 Funktionen . . . 54
5 Vektoren und Matrizen . . . 80
5.1 Formalismus . . . 81
5.2 Vektorgeometrie . . . 93
5.3 Lineare Abbildungen . . . 101
5.4 Lineare Unterräume und Untermannigfaltigkeiten . . . 107
5.5 Lineare Gleichungssysteme . . . 115
5.6 Koordinatentransformationen . . . 130
5.7 Ausblick: Eigenwerte und Eigenvektoren . . . 146
6 Differentialrechnung . . . 151
6.1 Idee und Begriffe . . . 152
6.2 Vorgehen . . . 163
6.3 Taylor-Polynome . . . 167

6.4 Fehlerrechnung . . . 172
7 Integralrechnung . . . 177
7.1 Flächeninhalt . . . 178
7.2 Normalbereiche . . . 183
7.3 Praktische Integration . . . 193
8 Wahrscheinlichkeit . . . 200
8.1 Wahrscheinlichkeitsräume . . . 201
8.2 Zufallsgrößen . . . 209
8.3 Diskrete Verteilungen . . . 222
8.4 Kontinuierliche Verteilungen . . . 225
8.5 Unabhängigkeit . . . 230
8.6 Grenzwertsätze . . . 234
8.7 Zufallsvektoren . . . 236
9 Statistik . . . 243
9.1 Überblick . . . 244
9.2 Deskriptive Statistik . . . 247
9.3 Punktschätzung . . . 249
9.4 Bereichsschätzung . . . 254
9.5 Hypothesentests . . . 258
10 Übung 1 (Grundlegendes) . . . 263
10.1 Logik, Argumentation, Begriffsbildung . . . 263
10.2 Wiederholung Schulmathematik . . . 263
11 Übung 2 (Python und Jupyter) . . . 264
11.1 Jupyter . . . 264
11.2 Python . . . 264
12 Übung 3 (Aussagen, Abbildungen, Folgen) . . . 266
12.1 Aussagen . . . 266
12.2 Abbildungen . . . 266
12.3 Zahlenfolgen . . . 266
13 Übung 4 (Funktionen) . . . 269
13.1 Definitionsbereich . . . 269
13.2 Visualisierung . . . 269
13.3 Grenzwerte und Stetigkeit . . . 270
14 Übung 5 (Matrizen, Vektorgeometrie) . . . 272
14.1 Rechnen mit Matrizen . . . 272
14.2 Rechnen mit Vektoren . . . 273
14.3 Skalarprodukt und orthogonale Projektion . . . 273
15 Übung 6 (lineare Abbildungen, Basen) . . . 275
15.1 Lineare Abbildungen . . . 275
15.2 Unterräume und Basen . . . 275
16 Übung 7 (lineare Gleichungssysteme) . . . 277
16.1 Formulieren von LGS . . . 277
16.2 Lineare Unabhängigkeit . . . 278
16.3 Geraden und Ebenen . . . 279

17 Übung 8 (Koordinatentransformationen) . . . 280
17.1 Aufgabe 8.1 (KTA) . . . 280
17.2 Aufgabe 8.2 (KTA) . . . 280
17.3 Aufgabe 8.3 (KTA) . . . 280
18 Übung 9 (Differentialrechnung) . . . 281
18.1 Aufgabe 9.1 (KTA) . . . 281
18.2 Aufgabe 9.2 . . . 281
18.3 Aufgabe 9.3 (KTA) . . . 281
18.4 Aufgabe 9.4 (KTA) . . . 281
18.5 Aufgabe 9.5 (KTA) . . . 281
19 Übung 10 (Integralrechnung) . . . 283
19.1 Stammfunktionen . . . 283
19.2 Flächeninhalte . . . 284
19.3 Aufgabe 10.6 . . . 284
20 Übung 11 (Wahrscheinlichkeit, Zufallsgrößen) . . . 286
20.1 Ereignisse und Ereignisfelder . . . 286
20.2 Wahrscheinlichkeitsmaße . . . 287
20.3 Diskrete Zufallsgrößen . . . 287
20.4 Kontinuierliche Zufallsgrößen . . . 288
21 Übung 12 (Wahrscheinlichkeitsverteilungen) . . . 289
21.1 Diskrete Verteilungen . . . 289
21.2 Kontinuierliche Verteilungen . . . 289
22 Übung 13 (Zufallsvektoren) . . . 291
22.1 Aufgabe 13.1 (KTA) . . . 291
22.2 Aufgabe 13.2 (KTA) . . . 291
22.3 Aufgabe 13.3 . . . 291
23 Übung 14 (deskriptive Statistik) . . . 293
23.1 Aufgabe 14.1 . . . 293
23.2 Aufgabe 14.2 . . . 293
23.3 Aufgabe 14.3 . . . 293
23.4 Aufgabe 14.4 . . . 294
24 Übung 15 (induktive Statistik) . . . 295
24.1 Aufgabe 15.1 (KTA) . . . 295
24.2 Aufgabe 15.2 . . . 295
24.3 Aufgabe 15.3 . . . 296
25 Download . . . 297
26 Organisatorisches . . . 298
26.1 Ablauf der Veranstaltung . . . 298
26.2 Hinweise zum Skript . . . 299
26.3 Prüfung . . . 300
26.4 Arbeitsplan . . . 300
26.5 Änderungshistorie . . . 302

0 Vorwort

Dieses Vorlesungsskript ist die Arbeitsgrundlage für die Veranstaltung “Mathematik I” (Modul i933) für den Fernstudiengang Vermessung und Geoinformatik an der HTW Dresden ab Wintersemester 2026/2027. Hinweise zum Ablauf der Veranstaltung, insbesondere zum Umgang mit diesem Skript, finden Sie unter Organisatorisches.

Technischer Unterbau ist MyST Markdown mit einem eigenen, auf dem Standard-Book-Theme aufbauenden Theme. Das MyST-Projekt ist noch recht jung und etliche Features sind im Beta-Stadium. Hier und da ist also mit Ecken und Kanten zu rechnen. Fragen und Anregungen zur Technik gern an Jens Flemming.

1 Motivation

Wozu heute noch Mathematik? Das macht doch alles der Computer oder "die KI".

1.1 Mathematik vs. Rechnen

Mathematik wird oft mit Rechnen verwechselt oder gleichgesetzt. Richtig ist: **Mathematik ist die Kunst das Rechnen zu vermeiden!** Mathematik strebt einerseits nach möglichst einfachen Lösungswegen für praktische Aufgabenstellungen und organisiert andererseits die trotzdem oft notwendige Rechenarbeit heute so, dass sie **mit dem Computer** erledigt werden kann.

Manuelles Rechnen ist für Ingenieure heute kaum mehr relevant. Lediglich für den Aufbau eines gewissen Grundverständnisses zu einem zu lösenden Problem kann es sinnvoll sein, einige einfache Spezialfälle per Hand auszurechnen. Alles, was einigermaßen praxisnah ist, ist zu komplex zum manuellen Rechnen und wird dem Computer überlassen.

Beispiel

Theoretisch kann man die zu bewegenden Erdmassen beim Bau von Verkehrswegen manuell durch Volumenberechnungen ermitteln. Schon aus der Tatsache, dass dazu Geländeoberfläche und geplanter Trassenverlauf als mathematische Formel gegeben sein müssten, folgt, dass dieser Gedanke praktisch nicht umsetzbar ist. Vielmehr werden Geländeoberfläche und Trassenverlauf nur als CAD-Modell bestehend aus Splines, Bézier-Kurven, NURBS usw. im Computer existieren, sodass auch die hinreichend genaue Volumenbestimmung nur einer Computerrechnung zugänglich ist.

Aus dem Beispiel wird klar: Manuelles Rechnen müssen nur noch wenige Ingenieure beherrschen (Grundlagenarbeit, Software-Entwicklung,..). **Solide Fähigkeiten beim Lösen von Aufgabenstellungen mit dem Computer sind hingegen zwingend für jeden Ingenieur.**

1.2 Mathematik als Sprache

Die Mathematik dient ganz Wesentlich als Sprache in Technik, Naturwissenschaften und Wirtschaftswissenschaften; und das in zweierlei Hinsicht:

1. Mit den Mitteln der Mathematik werden **Modelle der realen Welt** formuliert. Solche mathematischen Modelle ermöglichen einerseits eine genaue Untersuchung des betrachteten Problems. Andererseits ist die mathematische Formulierung die Grundlage für den Einsatz von ebenfalls mit mathematischen Mitteln formulierten Lösungsverfahren.

2. Soll eine Aufgabenstellung mit dem Computer gelöst werden, so muss die Aufgabe in eine **für Computer verständliche Form** gebracht werden. Computer verstehen nur ganz klare und präzise Anweisungen, sodass man üblicherweise die mathematische Sprache als gemeinsames Ausdrucksmittel von Mensch und Computer wählt.

1.3 Computer

Landläufig herrscht die Meinung vor, dass Computer immer richtig rechnen. Das stimmt leider nicht, wie wir im Laufe der Veranstaltung gelegentlich sehen werden. Wichtige Fähigkeit von Ingenieuren ist heute, zu **erkennen, in welchen Situationen der Computer vertrauenswürdige Ergebnisse liefert** und wann nicht! Diese Problematik werden wir in dieser einführenden Veranstaltung nur am Rande streifen; für eine Vertiefung sei auf das Teilgebiet der numerischen Mathematik verwiesen.

Computer sind grundsätzlich als "dumm" zu betrachten. Sie tun exakt das, was man ihnen sagt (und nicht dass, was man meint, ihnen gesagt zu haben). Ob der Computer das Richtige tut und ob wir den Ergebnissen trauen können, liegt allein in unserer Hand. Der Mensch muss die Prozesse und Hintergründe verstehen, muss einordnen und bewerten. **Der Computer übernimmt lediglich das Rechnen, nicht das Denken!**

1.4 "KI"

Der Begriff "künstliche Intelligenz" ist irreführend. Tatsächlich handelt es sich bei den heute unter diesem Begriff verstandenen Techniken um **simulierte Intelligenz**. Diese System können **nicht denken, keine zielgerichtete Kreativität** entfalten. Sie interpolieren und (in weit geringerem Maß) extrapolieren lediglich vorhandenes Faktenwissen. Aus dieser Feststellung folgen zwei wichtige Schlüsse für das Dasein als Ingenieur:

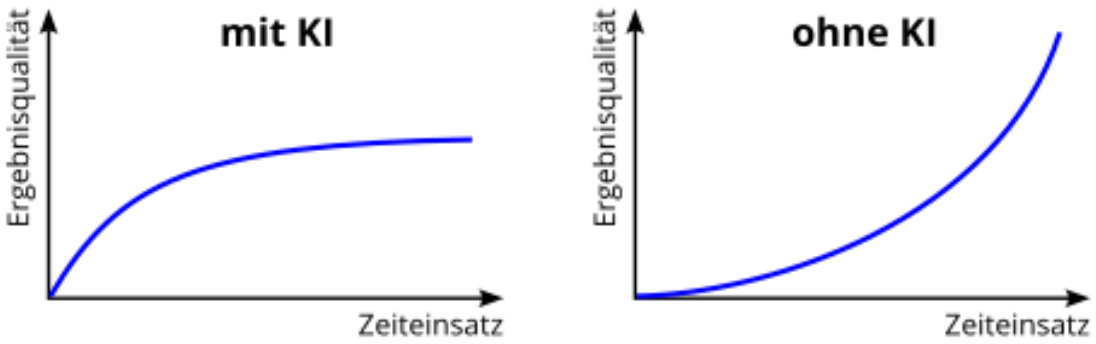
1. Wer nur Standardaufgaben lösen muss/möchte, kommt mit KI oft ans Ziel. Wer wirklich Neues schaffen möchte, muss denken und verstehen.
2. Von KI gelieferte Problemlösungen können beliebig grobe oder auch sehr subtile Fehler enthalten. Gerade im oft sicherheitskritischen Ingenieurbereich müssen aber nachweisbar richtige Lösungen her. Mindestens müssen die Lösungen nach bestem Wissen und Gewissen korrekt sein; ein Anspruch, den KI-basierte Lösungen ohne detaillierte Prüfung nicht erfüllen.

Wichtig: KI arbeitet nicht deterministisch, also nicht Schritt für Schritt nachvollziehbar, sondern statistisch, indem sie eine mit hoher Wahrscheinlichkeit passende Antwort "würfelt". Die Wahrscheinlichkeiten basieren dabei auf der der KI beim Training zugänglichen Wissensbasis. Insbesondere Nischenthemen werden durch Mainstream-Themen überstrahlt. KI-Antworten dazu gleichen einer zufälligen, meist sinnlosen oder ganz falschen Vermischung des eigentlichen Themas mit den Mainstream-Themen im Umfeld.

Dabei sind die Antworten sauber und überzeugend formuliert, halten einer inhaltlichen Detailprüfung jedoch nicht stand.

Zur Ideenfindung ist KI durchaus geeignet. Die Abschließende Problemlösung muss jedoch vom Ingenieur vollständig durchdacht und verstanden sein.

Der Einsatz von KI liefert schnelle Erfolge, die langfristig aber hinter den durch Denken und Verstehen erzielbaren Resultaten zurückbleiben.



KI liefert schnell akzeptable Ergebnisse auf mittelmäßigem Niveau. Selbst denken und verstehen benötigt mehr Zeit, liefert aber langfristig deutlich hochwertigere Resultate.

1.5 Mittel gegen Komplexität

Neben der direkten Funktion als Sprache dient Mathematik auch dem Strukturieren von Denkprozessen. Die Mathematik schafft vereinfachte Formulierungen der realen Welt und abstrahiert komplexe Zusammenhänge und Vorgänge in klar definierten Begriffen und Strukturen. Durch die Gewöhnung an abstrakte Begriffe (als Ersatz für die viel zu komplexe Realität) werden Denkprozesse vereinfacht oder überhaupt erst möglich.

Beispiel (reelle Zahlen)

Reelle Zahlen wie $\sqrt{2}$ oder die Kreiszahl π erscheinen uns nach jahrelangem Training in der Schule ganz natürlich und wir arbeiten täglich mit ihnen. In der Realität haben sie aber im Gegensatz zu rationalen Zahlen keine Entsprechung; im Gegenteil, sie sind ein sehr abstraktes Konzept, welches die Formulierung vieler technischer Zusammenhänge deutlich vereinfacht (Video 130, Notizen dazu).

Abstraktion findet auch außerhalb der Mathematik statt, ist aber in der Mathematik zentrales Arbeitsmittel mit Wirkung in andere Wissenschaftsgebiete hinein.

Beispiel (Farben)

Der Begriff “Farbe” ist so alltäglich wie abstrakt. Was bedeutet “die Kirsche ist rot”? Wie erklärt man jemandem eine konkrete Farbe ohne auf Gegenstände dieser Farbe als Beispiele zu zeigen?

Von klein auf haben wir nur durch ständiges Zeigen auf Gegenstände und das Sagen des dazugehörigen Farbwortes ein Verständnis für den Begriff “Farbe” entwickelt. Nie haben wir eine konkrete Definition gehört. Erst in der Physik kommt man einer konkreten Definition nahe (Farbe = Wellenlänge des Lichts), die aber nicht die menschliche Farbwahrnehmung erklären kann: Warum wird eine Mischung aus rotem und grünem Licht in gleicher Weise wahrgenommen wie rein gelbes Licht?

Aus mathematischer Sicht ist “rot” eine Äquivalenzklasse bezüglich einer Äquivalenzrelation: zwei Gegenstände *haben die gleiche Farbe* (Definition!), wenn deren Farbwahrnehmung im Gehirn identisch ist. Mit welchem Wort (“rot”, “grün”, “gelb” usw.) man die einzelnen Gruppen von Gegenständen gleicher Farbwahrnehmung dann bezeichnet, ist eine andere Frage.



Das mathematische Konzept der Äquivalenzklassen ermöglicht völlig analog zum kleinkindlichen Lernprozess implizite Begriffsdefinitionen. Man kann so auch mit komplexen Situationen umgehen, die keine (hinreichend einfache) explizite Begriffsdefinition zulassen. Wir sehen hier: **Scheinbar abstrakte mathematische Konzepte sind oftmals direkte Formalisierungen alltäglichen Denkens!**

1.6 “Soft-Skills”

Beschäftigt man sich mit Mathematik, erlernt man ganz nebenbei viele Fähigkeiten, die im Privaten wie Beruflichen sehr nützlich sind:

- genaues Analysieren,
- streng logisches Argumentieren,
- exakte Ausdrucksweise,
- scheinbar offensichtliches immer wieder hinterfragen,
- das eigene Denken und Verstehen im Grenzbereich testen.

Das Üben dieser Fähigkeiten in der klaren, strukturierten Welt der Mathematik rüstet für den Weg in diffusere Gebiete.

Beispiel (logisches Argumentieren)

Hase und Kröte liefern sich ein Wettrennen über 100 Meter. Der Hase sieht die Sache entspannt und lässt die Schildkröte 10 Meter weiter vorn starten. Beide laufen gleichzeitig los. Wer wird gewinnen?

Dazu zwei Argumentationen:

1. Der Hase wird gewinnen, da er viel schneller läuft als die Schildkröte.
2. Die Schildkröte wird gewinnen. Wenn der Hase bei der Startposition der Schildkröte ankommt, ist diese bereits ein Stück gelaufen, liegt also vorn. Kommt der Hase dann wiederum an dieser Position an, ist die Schildkröte wieder ein Stück weiter. Jedesmal, wenn der Hase an der vorherigen Position der Schildkröte ankommt, ist die Schildkröte schon wieder ein Stück weiter. Die Schildkröte liegt immer vorn, wird also gewinnen.

Wer hat recht? Wo ist der Fehler in der anderen Argumentation? Siehe Übung 1.

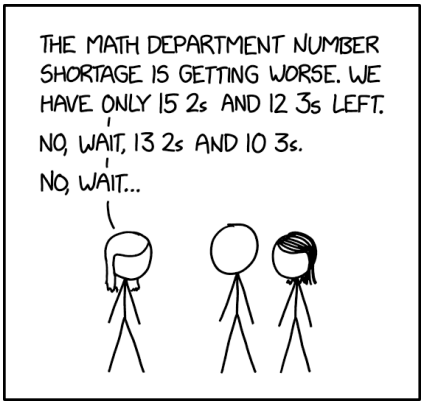
Beispiel (scheinbar offensichtliches hinterfragen)

Was heißt eigentlich “gleich viel”? Hat ein gesunder Mensch gleich viele Finger wie Fußzehen? Gibt es gleich viele gerade wie ungerade Zahlen? Gibt es gleich viele Brüche (d.h. rational Zahlen) wie es natürliche Zahlen gibt? Siehe Übung 1.

1.7 Mathematik im Studium

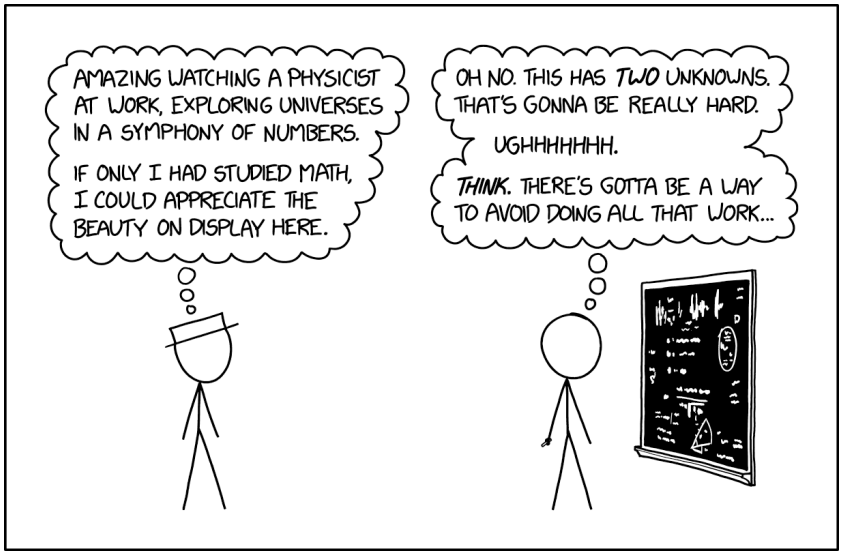
Im weiteren Studienverlauf wird die Mathematik in vielen Modulen eine tragende Rolle spielen. Einige Module aus den nächsten Semestern und die dort benötigten Themen aus Mathematik I und II sind:

- Statistik und Ausgleichungsrechnung (Grundlagen Wahrscheinlichkeitsrechnung und Statistik, Anwendungen der Differentialrechnung, Matrizen und Vektoren, lineare Gleichungssysteme).



"10 minutes ago we were down to only 2 0s!" "How many do we have now?" "I ... don't know!!" Quelle: Randall Munroe, xkcd.com/3009

- Landesvermessung und Satellitengeodäsie (Matrizen und Vektoren, Differentialrechnung für Funktionen mehrerer Veränderlicher).
- Remote Sensing (Matrizen und Vektoren).
- Positionierung und Navigation (Matrizen und Vektoren, Anwendungen der Differentialrechnung).
- KI und Geodata-Science (Matrizen und Vektoren, Grundlagen der Statistik, Differentialrechnung für Funktionen mehrerer Veränderlicher).
- Geodätische Auswertetechniken (Grundlagen der Statistik).
- Laserscanning und Punktwolkenverarbeitung (Grundlagen der Statistik).



I could type this into a solver, which MIGHT help, but would also mean I have to get a lot of parentheses right... Quelle: Randall Munroe, xkcd.com/2207

2 Schulmathematik

In diesem Kapitel erhalten Sie ein Überblick darüber, welche Themen der Schulmathematik in dieser Veranstaltung nochmals behandelt werden und welche eigenständig zu wiederholen sind, sofern Sie die Notwendigkeit dazu verspüren.

2.1 Online-Kurs OMB+

Die HTW Dresden unterstützt und nutzt den von zahlreichen Bildungseinrichtungen gemeinsam getragenen kostenfreien Online-Kurs OMB+ zur Auffrischung des Schulwissens im Bereich Mathematik. Dort gibt es einerseits das notwendige Faktenwissen, andererseits aber auch Aufgaben zum Test der eigenen Fähigkeiten. Große Teile des Materials sind ohne Anmeldung nutzbar. Einige interaktive Aufgaben werden erst nach kostenloser Registrierung nutzbar.

Nutzen Sie dieses Angebot, um parallel zur Veranstaltung Rückstände und Vergessenes aufzuarbeiten! In diesem Kapitel finden Sie jeweils Links zu konkreten Abschnitten des Online-Kurses.

2.2 Themen, die als bekannt vorausgesetzt werden

Folgende Themen werden nicht in der Lehrveranstaltung behandelt, sondern als aus der Schulmathematik bekannt vorausgesetzt:

- Kapitel IA in OMB+ (Mengen, Zahlen)
- Kapitel IB in OMB+ (Potenzen, Proportionalität)
- Kapitel II in OMB+ (Gleichungen in einer Unbekannten)
- Kapitel III in OMB+ (Ungleichungen in einer Variablen)
- Kapitel V in OMB+ (Geometrie)
- Kapitel VI in OMB+ (elementare Funktionen) (einige Teilthemen werden in der Veranstaltung wiederholt)
- Kapitel IX in OMB+ (Vektorgeometrie) (einige Teilthemen werden in der Veranstaltung wiederholt)
- Kapitel XII.4 in OMB+ (Logik und Mengenlehre - Mengen)
- Kapitel XII.6 in OMB+ (Logik und Mengenlehre - Kartesische Produkte)

Nebenbemerkung

Die Darstellung der Themen in OMB+ ist recht detailliert und teils umfangreicher

als in der Veranstaltung benötigt. Verschaffen Sie sie sich deshalb zunächst nur einen Überblick über die genannten Themen. Sollten Sie im Verlauf der Veranstaltung Nachholbedarf feststellen, sollten Sie die entsprechenden Abschnitte von OMB+ genauer durcharbeiten.

2.3 Themen, die wiederholt und vertieft werden

Folgende Themen werden in der Veranstaltung nochmals vollständig eingeführt, behandelt und vertieft. In einigen Ländern und Bundesländern gehören diese zum Schulstoff, sodass unter Umständen schon Vorwissen vorhanden ist.

- Kapitel IV in OMB+ (lineare Gleichungssysteme)
- Kapitel VII in OMB+ (Differentialrechnung)
- Kapitel VIII in OMB+ (Integralrechnung)
- Kapitel X in OMB+ (Stochastik)
- Kapitel XII in OMB+ (Logik und Mengenlehre), sofern benötigt (außer XII.4, XII.6),
- Kapitel XIII in OMB+ (Wahrscheinlichkeitstheorie)

2.4 Nicht behandelte Themen

OMB+ enthält auch Themen, die in der Veranstaltung nicht von Bedeutung sind:

- Kapitel IX in OMB+

3 Computereinsatz

Grundlagen der Computernutzung in der Mathematikausbildung.

3.1 Überblick

Der Einsatz des Computers zum Lösen von in der Sprache der Mathematik formulierten Aufgabenstellungen kann auf unterschiedliche Weise erfolgen. Dabei spielt einerseits die Auswahl geeigneter Software eine Rolle; andererseits muss man sich im Klaren sein über die Fähigkeiten und Unfähigkeiten des Computers.

3.1.1 Einschränkungen bei der Arbeit mit dem Computer

Viele mathematische Probleme können nicht exakt mit dem Computer gelöst werden, sondern nur näherungsweise. Die Gründe dafür liegen in der Arbeitsweise von Computern:

- Jede Eingabe kann nur aus endlich vielen Zahlen bestehen. Computer können also beispielsweise nicht direkt mit stetigen Funktionen (Potenzfunktionen, Sinus,...) oder unendlichen Zahlenfolgen arbeiten.
- Auch Ausgaben/Ergebnisse können nur aus endlich vielen Zahlen bestehen.
- Es sind stets nur endlich viele Rechenschritte möglich. Beispielsweise können im Allgemeinen keine unendlichen Summen mit dem Computer berechnet werden.

Ein weiterer zu beachtender Punkt ist, dass Computer nur einen sehr begrenzten Zahlenvorrat zur Verfügung haben. Üblicherweise kennen sie nur $2^{64} \approx 2 \cdot 10^{18}$ verschiedene Zahlen. Deshalb werden die meisten Zahlen durch den Computer auf eine vom Computer verarbeitbare Zahl gerundet, ohne dass dies für den Nutzer sofort ersichtlich ist. Je nach Kontext müssen die mitunter seltsam anmutenden Konsequenzen mit bedacht werden:

- Beispielsweise kann die Zahl 0.1 nicht direkt in Berechnungen verwendet werden. Stattdessen rechnet der Computer mit

  $$0.1000000000000000055511151231257827021181583404541015625. \tag{3.1}$$

  Dies kann zu unerwarteten Ergebnissen führen, wie das folgende kleine Python-Programm zeigt (die Programmiersprache Python wird im nächsten Unterkapitel behandelt):

  ```
  print(0.1 * 0.1 - 0.01)  # sollte exakt 0 ergeben
  ```

  ```
  1.734723475976807e -18
  ```

- Gelegentlich erhält man deshalb völlig falsche Ergebnisse:

  ```
  print(1 + 10.5 ** 20 - 10.5 ** 20)  # sollte exakt 1 ergeben
  ```

  ```
  0.0
  ```

Nebenbemerkung

Die beiden Beispiele können Sie leicht selbst ausprobieren. Öffnen Sie in Ihrem Webbrowser die JavaScript-Konsole (in Firefox z.B. durch Drücken von ++, in Chromium durch ++) und geben Sie dort `0.1 * 0.1 - 0.01` bzw. `1 + 10.5 ** 20 - 10.5 ** 20` ein. Die Ergebnisse sollte aussehen wie hier im Skript.

Die Auswirkungen und der Umgang mit solchen Effekten sind Gegenstand der numerischen Mathematik. In unserer Veranstaltung gehen wir auf dieses wichtige Thema jedoch nicht näher ein, sondern belassen es bei gelegentlichen Hinweisen auf in dieser Hinsicht problematische Situationen.

3.1.2 Symbolisches vs. numerisches Rechnen

Für das Ausführen von Berechnungen mit dem Computer gibt es zwei grundlegend verschiedene Herangehensweisen:

- Beim **symbolischen Rechnen** werden mathematische Formeln durch den Computer in der gleichen Weise umgeformt, kombiniert und gelöst wie man es beim manuellen Rechnen tun würde. Dadurch erhält man exakte Ergebnisse, die wiederum als Formeln vorliegen. Bei der verwendeten Software spricht man auch von **Computer-Algebra-Systemen (CAS)**.
- Beim **numerischen Rechnen** werden nur Grundrechenarten verwendet. An vielen Stellen des Rechenweges werden gerundete oder nur näherungsweise ermittelte Zwischenergebnisse verwendet. Als Endergebnis stehen nur Zahlen zur Verfügung, bei denen nicht mehr erkennbar ist, wie diese entstanden sind (z.B. 1.41421356 statt $\sqrt{2}$).

Im technischen und wissenschaftlichen Bereich wird praktisch nur numerisch gerechnet. Gründe:

- Numerische Berechnungen sind viel schneller, da Computer für diese Art von Berechnungen gebaut sind. Üblich sind heute mehrere Milliarden Grundrechenoperationen pro Sekunde.
- Symbolisches Rechnen liefert bei komplexen Aufgabenstellungen meist keine Lösung, da die Lösungen nicht explizit als Formeln darstellbar sind.

Gelegentlich verwendet man CAS für kleine, klar abgegrenzte Teilschritte, bei denen das Vorliegen einer expliziten Formel in den Folgeschritten (Geschwindigkeits-)Vorteile bringt.

Beispiel (automatisches Differenzieren)

Grundlage von "KI" sind künstliche neuronale Netze. Das sind aus Tausenden oder Millionen einfacher Grundfunktionen zusammengesetzte nichtlineare Funktionen, welche Millionen durch ein Optimierungsverfahren zu bestimmende Parameter enthalten. Für die Durchführung der Optimierung werden die Ableitungen des neuronalen Netzes bezüglich der Parameter benötigt (den Begriff "Ableitung" führen wir im Laufe der Veranstaltung ein). Diese können gut symbolisch berechnet werden, was Näherungsfehler vermeidet. Außerdem beschleunigt das Vorliegen einer konkreten Formel die Auswertung der Ableitungen während der Optimierung. Verfahrensbedingt müssen diese Ableitungen millionenfach ausgewertet werden.

Das symbolische Berechnen der Ableitungen von Funktionen, die aus einfachen Grundfunktionen zusammengesetzt sind, nennt man auch **automatisches Differenzieren**. Die eigentliche Optimierung erfolgt dann numerisch, da das Problem aufgrund seiner Komplexität nicht symbolisch lösbar ist.

3.1.3 Programmcode vs. grafische Oberfläche

Für praktisch alle Standardaufgaben im Ingenieurwesen gibt es Software mit grafischer Benutzeroberfläche (kurz: GUI für "graphical user interface"). Diese haben jedoch einige Nachteile:

- Möglich ist nur, was die GUI vorsieht. Erweiterungen oder Änderungen sind für Nutzer nur schwer oder gar nicht umsetzbar.
- Abläufe lassen sich schlecht automatisieren, da diese aus Folgen von Mausklicks bestehen.
- Für Aufgabenstellungen, die erstmalig oder nur selten auftreten, existiert keine Software mit GUI (oder ist extrem teuer).

Aus diesen Gründen ist der Einsatz einer geeigneten (!) Programmiersprache auch im Ingenieurbereich für zahlreiche Aufgaben deutlich vorteilhafter. Insbesondere

- sind die Anwendungsmöglichkeiten unbeschränkt,
- lassen sich alle Abläufe ohne große Anpassungen automatisieren,
- kann bei Bedarf im Nachgang eine GUI hinzugefügt werden.

Hinzu kommt, dass sich Programmcode einfacher versionieren lässt und gemeinsames Bearbeiten (auch in Echtzeit) problemlos möglich ist.

Üblich im Ingenieurbereich sind:

- Python/Jupyter (wird in dieser Veranstaltung verwendet),
- Octave oder Matlab (wird zunehmend durch Python verdrängt).

Nebenbemerkung

Die Arbeit mit Programmcode statt grafischer Oberfläche ist keineswegs komplizierter. Lediglich der Funktionsumfang ist größer. Aufgrund der Vielzahl von Möglichkeiten ist es nicht zweckmäßig für jede Funktion einen Button oder einen Menüeintrag zu bieten. Stattdessen werden alle Aktionen durch Textbefehle ausgelöst.

3.2 Python

In dieser Veranstaltung nutzen wir die Programmiersprache Python um dem Computer seine Aufgaben mitzuteilen. Werkzeuge aus dem Jupyter-Ökosystem ermöglichen die Python-Programmierung in einer sehr komfortablen grafischen Umgebung.

3.2.1 Warum Python?

Python ist eine moderne, kostenfreie und quelloffene ("open source") Programmiersprache. Sie entstand in den 1990er Jahren. Die erste offizielle Version wurde 1994 veröffentlicht. Autor und BDFL ("benevolent dictator for life") ist Guido van Rossum.

Python-Code ist sehr gut lesbar und enthält im Gegensatz zu den meisten anderen Programmiersprachen nur wenige "kryptische" Zeichen und Zeichenketten. Viele technische Aspekte des Programmierens werden von Python erledigt ohne den Nutzer damit zu konfrontieren. Mit Python kann man Software aus allen Bereichen entwickeln; von einfachen Skripten für die tägliche Arbeit bis zu vollwertigen Desktop- oder Webanwendungen. Tausende Erweiterungen ermöglichen zügiges und effizientes Arbeiten.

Als freie Software verfügt Python über eine riesige Online-Community, die mit Rat und Tat zur Seite steht. Kaum eine Frage, die nicht schon irgendwo beantwortet wurde.

Python ist für alle Plattformen verfügbar: Linux, macOS, Windows und viele andere. Python steht beispielsweise hinter großen kommerziellen Projekten wie dem Youtube-Player, läuft eventuell aber auch in der Steuereinheit Ihrer Waschmaschine (vgl. MicroPython).

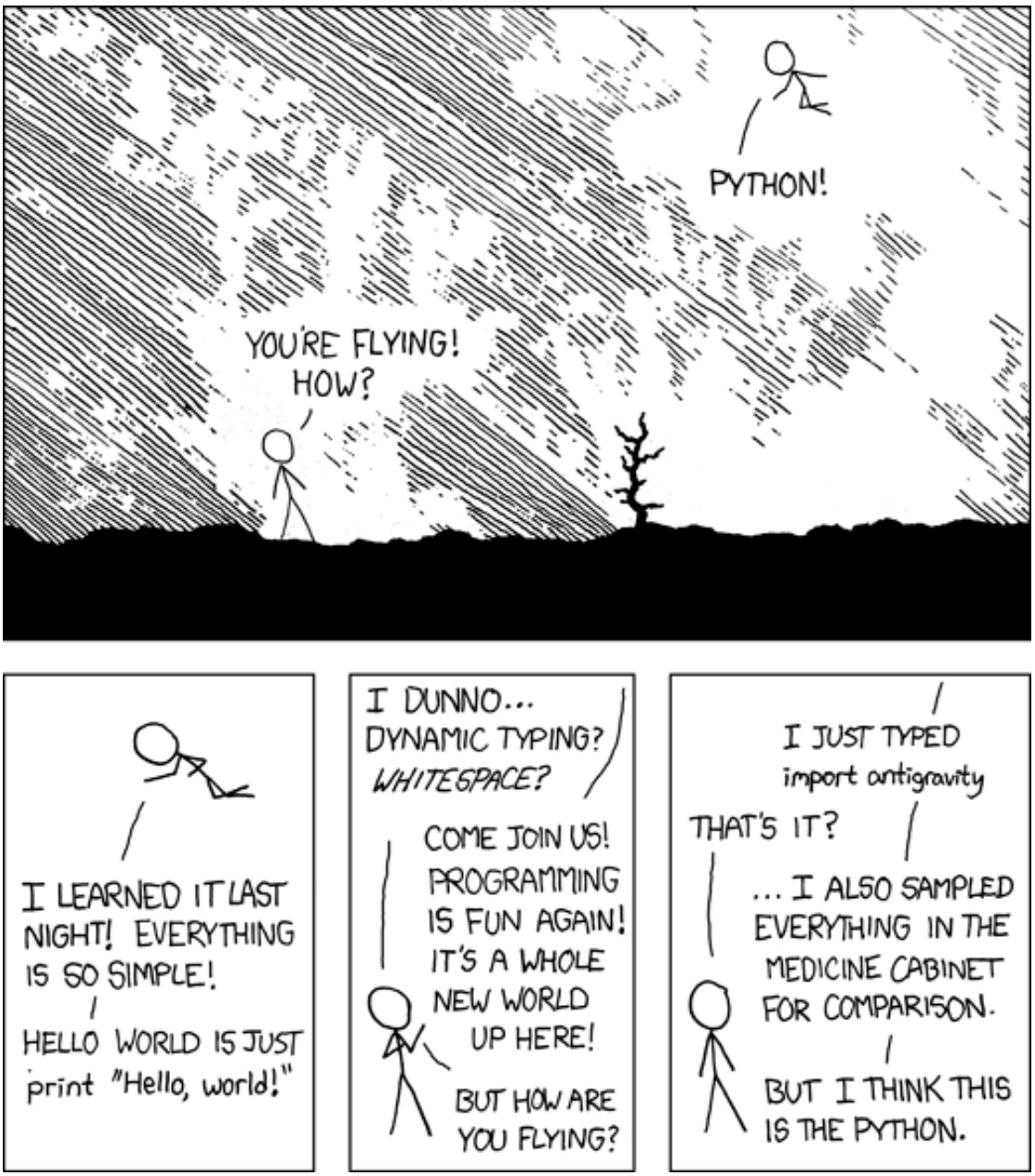
Nebenbemerkung

Es gibt zwei Varianten von Python: Python 2 und Python 3. Entsprechender Quellcode ist nicht kompatibel, d.h. für Python 2 geschriebene Programm funktionieren nicht mit Python 3. Wir verwenden hier ausschließlich Python 3. Python 2 wird von den Entwicklern seit Januar 2020 als veraltet eingestuft und nicht mehr weiterentwickelt.

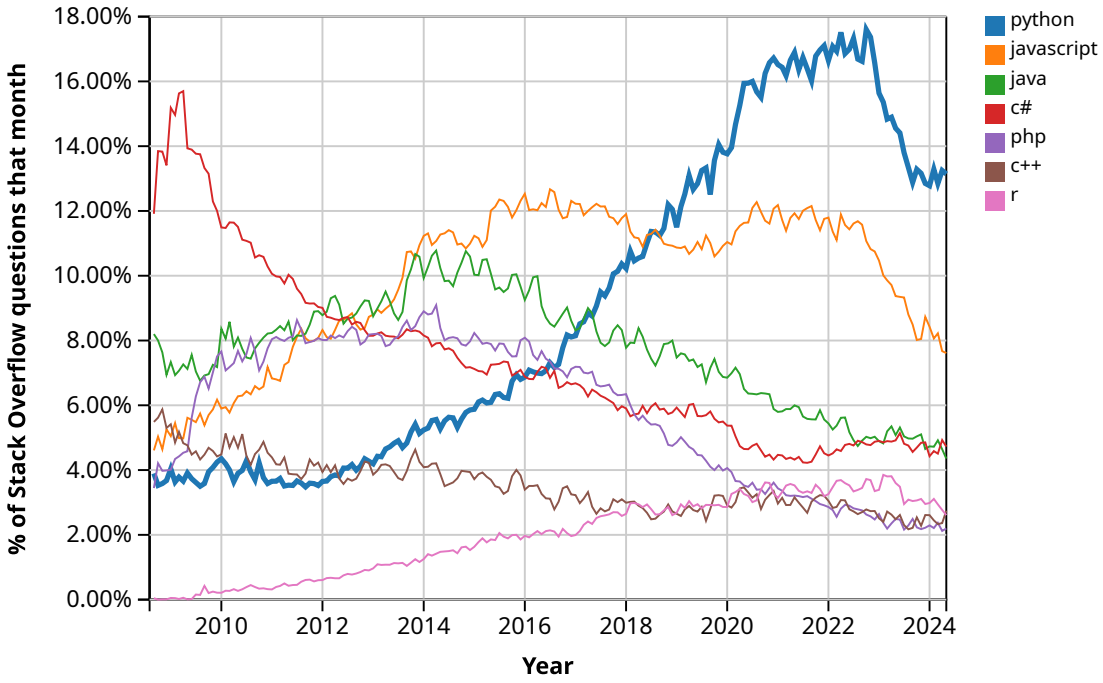
3.2.2 Jupyter

Das Jupyter-Ökosystem ist eine Sammlung von Werkzeugen für die Python-Programmierung mit dem Schwerpunkt auf Nutzerfreundlichkeit und leichte Bedienbarkeit für Anfänger. Jupyter erlaubt das Programmieren mit Python im Webbrowser. Ausgaben, wie beispielsweise komplexe und interaktive Visualisierungen, können direkt zwischen dem zugehörigen Quellcode platziert werden. Alles befindet sich in einem einzigen Dokument: Code, Ausgaben, erklärender Text, Bilder, Formeln,...

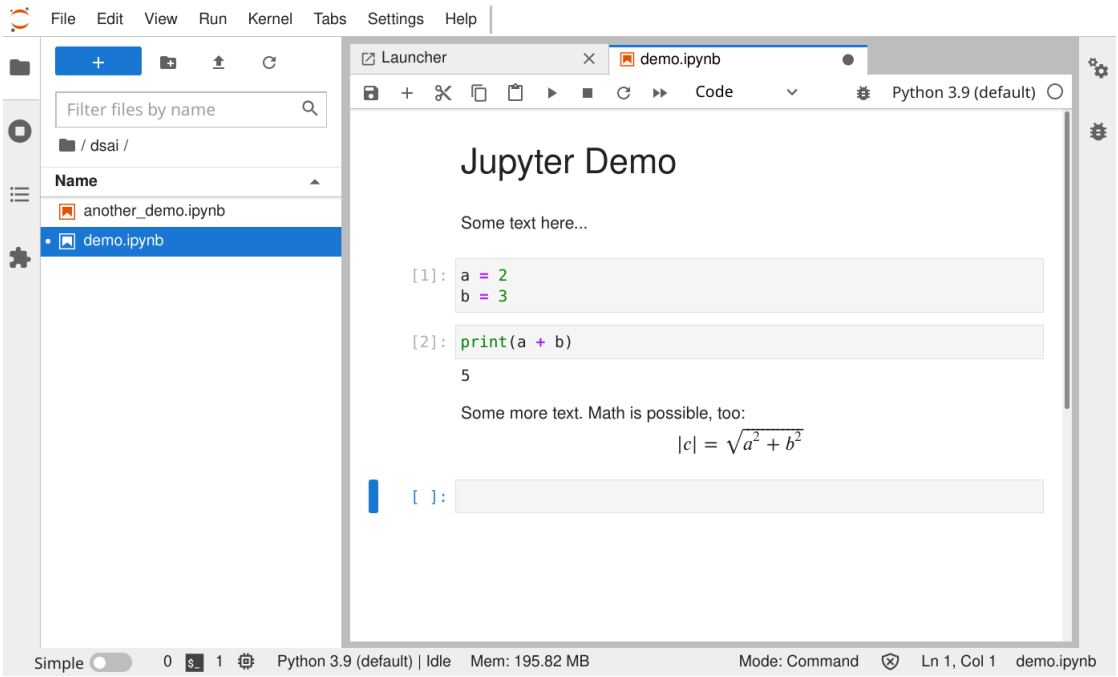
JupyterLab steht in vielen Formen zur Verfügung (vgl. Übung 2):



I wrote 20 short programs in Python yesterday. It was wonderful. Perl, I'm leaving you. Source: Randall Munroe, xkcd.com/353



Fragen zu Programmiersprachen auf Stack Overflow. Quelle: Stack Overflow Trends (vom Autor bearbeitet)



JupyterLab ist das am weitesten verbreitete Mitglied des Jupyter-Ökosystems. Es bringt Python in den Webbrowser.

- als rein browser-basierte Anwendung ohne Installation (Funktionsumfang geringfügig eingeschränkt),
- als lokale Installation auf dem eigenen Computer,
- als Installation auf Servern der Hochschule,
- auf Servern anderer Organisationen.

Welche Variante genutzt wird, ist für die Veranstaltung nebensächlich. Außerdem kann Python auch ganz ohne Jupyter genutzt werden.

3.2.3 Python-Crashkurs

Erwähnen hier als ersten Einstieg in Python beispielhaft einige Grundfunktionen. Weitere Möglichkeiten folgen im Laufe der Veranstaltung nach Bedarf. Für das weiterführende Selbststudium sind das Kapitel Python Crash Course sowie die darauf folgenden Kapitel in Data Science and Artificial Intelligence for Undergraduates geeignet.

3.2.3.1 Taschenrechner

```
# Punkt vor Strich
1 + 2 * 3
```

```
7
```

```
# Potenzen
2 ** 8
```

```
256
```

```
# Wurzeln
2.5 ** 0.5
```

```
1.5811388300841898
```

```
# Division
5 / 2
```

```
2.5
```

```
# ganzzahlige Division
5 // 2
```

```
2
```

```
# Rest bei ganzzahliger Division
5 % 2
```

```
1
```

3.2.3.2 Bessere Ausgaben

```
# einfacher Text
print('irgendein Text')
```

```
irgendein Text
```

```
# Text mit Kommazahl
print(f'Wurzel aus 2 ist {2.0 ** 0.5}.')
```

```
Wurzel aus 2 ist 1.4142135623730951.
```

Mehr Infos zu solchen Ausgaben: f-String und Minisprache für Stringformatierung.

```
# Text mit gerundeter Kommazahl
print(f'Wurzel aus 2 ist {2.0 ** 0.5:.3f}.')
```

```
Wurzel aus 2 ist 1.414.
```

3.2.3.3 Variablen

```
zahl = 2.0
wurzel = zahl ** 0.5
print(f'Wurzel aus {zahl} ist {wurzel:.3f}.')
```

```
Wurzel aus 2.0 ist 1.414.
```

Es gibt verschiedene Variablentypen (Ganzzahl, Kommazahl, Zeichenkette,...).

```
print(type(2))
print(type(2.0))
print(type('Hallo!'))
```

```
<class 'int'>
<class 'float'>
<class 'str'>
```

Umwandlungen sind zwischen geeigneten Typen möglich.

```
zeichenkette1 = '123'
zeichenkette2 = '456'
print(zeichenkette1 + zeichenkette2)

zahl1 = int(zeichenkette1)
zahl2 = int(zeichenkette2)
print(f'und als Zahlen: {zahl1 + zahl2}')
```

```
123456
und als Zahlen: 579
```

3.2.3.4 Listen

Listen sind Folgen von Variablen, deren Name einem einheitlichen Schema folgen (Listenname plus Index).

```
liste = ['rot', 'gelb', 'grün', 'blau']
print(liste[0])
print(liste[1])
print(liste[2])
print(liste[3])
print(liste[ -1])  # letztes Element
print(liste[1:3])  # Elemente mit Index 1 (2. Element) und 2 (3.
↪  Element)
```

```
rot
gelb
grün
blau
blau
['gelb', 'grün']
```

3.2.3.5 Bedingte Ausführung

```
a = 3
b = 2  # zwei Zahlen (Nutzereingaben oder aus Datei gelesen)

if b != 0:
    print(f'{a} / {b} = {a / b}')
else:
    print('Division durch Null!')
```

```
3 / 2 = 1.5
```

Wichtig: Test auf Gleichheit mit ==. Das einfache = ist der Zuweisungsoperator und kann nicht für Vergleiche verwendet werden.

3.2.3.6 Wiederholte Ausführung

```
liste = ['rot', 'gelb', 'grün', 'blau']

for farbe in liste:
    print(farbe)
```

```
rot
gelb
grün
blau
```

```
for i in range(0, 10, 2):
    print(i)
```

```
0
2
4
6
8
```

3.2.3.7 Module und Pakete Die Funktionalität von Python kann durch Module und Pakete erweitert werden. Module sind Sammlungen von zusätzlichen Funktionen. Pakete sind Sammlungen von Modulen. Praktisch besteht bei der Verwendung kein Unterschied zwischen Modulen und Paketen.

```
import math

math.cos(math.pi / 4)
```

```
0.7071067811865476
```

```
import math as m

m.cos(m.pi / 4)
```

```
0.7071067811865476
```

```
from math import cos, pi

cos(pi / 4)
```

```
0.7071067811865476
```

Python verfügt nur über wenige eingebaute Grundfunktionen. Man wird fast immer zusätzliche Module und Pakete benötigen, die ggf. erst installiert werden müssen.

3.3 SymPy

SymPy ist ein Python-Paket für symbolisches Rechnen (Ableitungen, Stammfunktionen, Formeln vereinfachen, Gleichungen lösen,...). Siehe auch Introductory Tutorial in der SymPy-Dokumentation.

Nebenbemerkung

Dieses Kapitel ist keine systematische Einführung, sondern liefert nur einige Beispiele. Alles weitere wird im Verlauf der Veranstaltung nach Bedarf eingeführt und vertieft.

```
import sympy
```

3.3.1 Variablen

SymPy verwendet ganz normale Python-Variablen, die aber einen SymPy-spezifischen Typ haben. Variablen, die später mit SymPy verwendet werden sollen, müssen deshalb zunächst durch SymPy erstellt werden.

```
x = sympy.symbols('x')  # zeige Python-Variable x in Formeln als x an

print(x)
print(type(x))
```

```
x
<class 'sympy.core.symbol.Symbol'>
```

Mehrere Variablen können gemeinsam erstellt werden:

```
x, y, z = sympy.symbols('x y z')

2 * x + y ** 2 + z
```

```
2*x + y**2 + z
```

3.3.2 Formeln vereinfachen

```
x = sympy.symbols('x')

a = (x + 1) ** 2
b = x ** 2 + 2 * x + 1
```

```
sympy.simplify(a - b)
```

```
0
```

```
x = sympy.symbols('x')

sympy.simplify(sympy.cos(x) ** 2 + sympy.sin(x) ** 2)
```

```
1
```

3.3.3 Nullstellen von Polynomen

SymPy liefert Polynomnullstellen samt Vielfachheiten (der Begriff "Vielfachheit" wird im Laufe der Veranstaltung erklärt):

```
x = sympy.symbols('x')

sympy.roots(x ** 3 - 4 * x ** 2 + x + 6, x)
```

```
{3: 1, 2: 1, -1: 1}
```

```
x = sympy.symbols('x')

sympy.roots(x ** 3 - 3 * x ** 2 + 3 * x - 1, x)
```

```
{1: 3}
```

3.3.4 Gleichungen lösen

```
x = sympy.symbols('x')

sympy.solve(x ** 2 - 2, x)  # löst x ** 2 - 2 = 0
```

```
[ -sqrt(2), sqrt(2)]
```

Das Ergebnis ist eine Liste (kein Intervall!) von Lösungen.

```
x = sympy.symbols('x')

formel = x ** 4 - 4 * x ** 3 + 7 * x ** 2 - 16 * x + 12
```

```
sympy.solve(formel, x)
```

```
[1, 3, -2*I, 2*I]
```

Die Werte mit `I` sind sogenannte komplexe Zahlen, die in der Veranstaltung nicht behandelt werden. Die Gleichung hat also nur zwei reelle Zahlen als Lösungen.

```
x = sympy.symbols('x', real=True)

formel = x ** 4 - 4 * x ** 3 + 7 * x ** 2 - 16 * x + 12
sympy.solve(formel, x)
```

```
[1, 3]
```

```
x = sympy.symbols('x')

sympy.solve(sympy.sin(x), x)
```

```
[0, pi]
```

```
x = sympy.symbols('x')

sympy.solveset(sympy.sin(x), x)
```

```
Union(ImageSet(Lambda(_n, 2*_n*pi), Integers), ImageSet(Lambda(_n,
↪  2*_n*pi + pi), Integers))
```

3.3.5 Numerisches Rechnen mit vielen Stellen

```
sympy.N('1 / 6', 50)
```

```
0.16666666666666666666666666666666666666666666666667
```

```
sympy.N('pi', 50)
```

```
3.1415926535897932384626433832795028841971693993751
```

3.4 NumPy

NumPy ist ein Paket, welches Python um Funktionen für das schnelle Rechnen mit großen Zahlenmengen (Matrizen, Vektoren) erweitert. NumPy ist im Python-Ökosystem der De-Facto-Standard für numerische Berechnungen wie beispielsweise das Lösen großer linearer oder nichtlinearer Gleichungssysteme.

Nebenbemerkung

Dieses Kapitel ist keine systematische Einführung, sondern liefert nur einige Beispiele. Alles weitere wird im Verlauf der Veranstaltung nach Bedarf eingeführt und vertieft. Eine systematische Einführung in NumPy finden Sie beispielsweise im Kapitel Efficient Computations with NumPy von Data Science and Artificial Intelligence for Undergraduates.

```
import numpy as np
```

Grundlage für die Arbeit mit NumPy sind die NumPy-Arrays, welche zur Darstellung von Vektoren und Matrizen dienen.

3.4.1 NumPy-Arrays erstellen

```
np.array([4, 5, 2, 3])  # aus einer Liste
```

```
array([4, 5, 2, 3])
```

```
np.zeros(4)  # eindimensionales Array mit Nullen
```

```
array([0., 0., 0., 0.])
```

```
np.ones((3, 2))  # zweidimensionales Array mit Einsen
```

```
array([[1., 1.],
       [1., 1.],
       [1., 1.]])
```

```
np.eye(5)  # Einheitsmatrix
```

```
array([[1., 0., 0., 0., 0.],
       [0., 1., 0., 0., 0.],
       [0., 0., 1., 0., 0.],
       [0., 0., 0., 1., 0.],
```

```
       [0., 0., 0., 0., 1.]])
```

3.4.2 Zugriff auf Array-Elemente

```
a = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])

a
```

```
array([[1, 2, 3],
       [4, 5, 6],
       [7, 8, 9]])
```

```
print(a[0, 0])  # 1. Zeile, 1. Spalte
print(a[1, 0])  # 2. Zeile, 1. Spalte
print(a[ -1, -1])  # letzte Zeile, letzte Spalte
```

```
1
4
9
```

```
a[1:3, :2]  # 2. und 3. Zeile, 1. und 2. Spalte
```

```
array([[4, 5],
       [7, 8]])
```

```
a[0, :]  # 1. Zeile, alle Spalten
```

```
array([1, 2, 3])
```

3.4.3 Eigenschaften von Arrays

```
a.shape  # Größe (Zeilen, Spalten)
```

```
(3, 3)
```

```
a.dtype  # Datentyp der Einträge
```

```
dtype('int64')
```

```
b = np.array([4.1, 6, 3, 5, 4, 3])
print(b.shape)
```

```
print(b.dtype)
```

```
(6,)
float64
```

3.4.4 Rechnen mit Arrays

```
a = np.array([4.1, 6, 3])
b = np.array([5, 4, 3])

print(2 * a + b)
```

```
[13.2 16.   9. ]
```

```
print(a * b)  # elementweises Produkt
```

```
[20.5 24.   9. ]
```

```
A = np.array([[1, 2], [3, 4]])
print(A)
```

```
[[1 2]
 [3 4]]
```

```
print(A * A)  # elementweises Produkt
```

```
[[ 1  4]
 [ 9 16]]
```

```
print(np.sin(A))  # Sinus elementweise
```

```
[[ 0.84147098  0.90929743]
 [ 0.14112001 -0.7568025 ]]
```

3.4.5 Arrays vergleichen

```
a = np.array([1, 2, 3])
b = np.array([ -1, 3, 3])

print(a > b)
```

```
[ True False False]
```

```
c = a > b  # Zuweisung (!) des Vergleichsergebnisses an c

print(c.any())  # mindestens ein Eintrag True?
print(c.all())  # alle Einträge True?
```

```
True
False
```

3.4.6 Weitere Möglichkeiten für den Elementzugriff

```
a = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])

a
```

```
array([[1, 2, 3],
       [4, 5, 6],
       [7, 8, 9]])
```

```
a[a > 3]
```

```
array([4, 5, 6, 7, 8, 9])
```

```
a[np.logical_and(a > 3, a < 6)]
```

```
array([4, 5])
```

```
a[a > 4] = 20

a
```

```
array([[ 1,  2,  3],
       [ 4, 20, 20],
       [20, 20, 20]])
```

3.4.7 Zufallszahlen

Für Tests und Simulationen sind Zufallszahlen sehr nützlich. Um Zufallszahlen zu erzeugen, muss zunächst der Zufallszahlengenerator von NumPy mit einem Startwert initialisiert werden. Der Startwert legt die Folge der generierten Zufallszahlen fest, sodass

bei späterer Wiederholung von Rechnungen sehr leicht wieder die gleichen Zufallszahlen verwendet werden können.

```
rng = np.random.default_rng(123)  # wie der Startwert gewählt wird, ist
↪  egal
```

```
rng.integers(23, 42, (4, 10))  # ganze Zahlen von 23 bis 41
```

```
array([[23, 35, 34, 24, 40, 27, 27, 26, 29, 26],
       [29, 38, 31, 40, 31, 28, 37, 38, 39, 39],
       [23, 32, 28, 27, 27, 38, 38, 27, 30, 37],
       [25, 34, 31, 40, 37, 27, 38, 38, 27, 32]])
```

```
rng.uniform(0, 2, (4, 4))  # gleichverteilt auf [0, 2]
```

```
array([[0.46311125, 0.33180799, 0.99557794, 1.16544928],
       [0.36867597, 0.02978983, 0.94226646, 1.45648666],
       [1.83720098, 1.25106801, 1.83424515, 1.7293805 ],
       [0.43628575, 1.73225486, 1.46150387, 0.55573058]])
```

4 Elementare Begriffe

Wiederholung und Präzisierung einiger Begriffe, die in allen Bereichen der Mathematik anzutreffen sind.

4.1 Aussagen

Die Mathematik tritt uns als Ansammlung von Begriffsdefinitionen und Aussagen zu den Beziehungen zwischen diesen Begriffen entgegen. Aussagen werden auch gern als *Satz* oder *Theorem* bezeichnet, wobei je nach Wichtigkeit der Aussage und ihrer Rolle in der Behandlung eines Themas unterschieden wird in Lemma (Hilfssatz), Proposition (Satz), Theorem (Hauptsatz), Korollar (Folgerung).

In dieser Veranstaltung werden wir die traditionelle Trennung der Inhalte in Definition, Satz, Beweis nicht vornehmen, sondern im Sinne der besseren Zugänglichkeit eine weniger strenge Form wählen. Dennoch sollten wir uns mit dem mathematischen Begriff der Aussage näher befassen, da Aussagen stets und ständig eine Rolle spielen werden, bemerkt und unbemerkt.

4.1.1 Was ist eine Aussage?

Merke

Eine **Aussage** ist ein Satz, der entweder wahr oder falsch ist. Ein solcher Satz kann sowohl in Alltagssprache als auch mit Hilfe mathematischer Begriffe und Symbole formuliert sein.

Diese Definition des Begriffs “Aussage” klingt banal, ist sie aber keinesweges. Offensichtlich ist

> Dresden ist die Hauptstadt von Sachsen.

eine Aussage, sogar eine wahre. Auch

> Leipzig ist die Hauptstadt von Sachsen.

ist eine Aussage, jedoch eine falsch oder unwahre.

Wie sieht es aber mit

> In 50 Jahren wird Leipzig die Hauptstadt Sachsens sein.

aus? Hier können wir kein klares “wahr” oder “falsch” zuordnen. Wir könnten uns auf den Standpunkt zurückziehen, das die Aussage entweder richtig oder falsch ist und wir nur nicht wissen, welche der beiden Möglichkeiten zutrifft. Tatsächlich kann heute aber niemand wissen, welche Rolle Leipzig in 50 Jahren spielen wird, egal wie viele Informationen zur Verfügung stehen. Man könnte als möglichen Ausweg die Formulierung abschwächen:

> In 50 Jahren wird Leipzig wahrscheinlich die Hauptstadt Sachsens sein.

Aber auch hier ist keine klare Zuordnung von “richtig” oder “falsch” möglich. Es handelt sich mehr um eine Meinung als um eine Aussage. Und Meinungen sind bekanntlich nicht

klar mit “richtig” oder “falsch” bewertbar.

Anderes Beispiel:

> Am 1. Januar 1999 fuhren 23 rote Autos am Hauptgebäude der Hochschule vorbei.

Sehr wahrscheinlich hat an diesem Tag niemand Autos vor der Hochschule gezählt. Es ist also nicht möglich, diesem Satz ein klares “richtig” oder “falsch” zuzuordnen. Die dafür benötigten Informationen liegen objektiv nicht vor.

Beispiel (Bäcker-Paradoxon)

Wir definieren hiermit den Begriff des Bäckers als eine Person, die genau für all die Menschen Brötchen bäckt, die nicht selbst für sich Brötchen backen. Ist dann folgender Satz eine Aussage?

> Der Bäcker bäckt auch Brötchen für sich selbst.

Wäre dies eine Aussage, so müsste sie entweder wahr oder falsch sein. Beide Varianten führen aber zum Widerspruch.

- Wenn der Satz wahr wäre, der Bäcker also auch für sich bäckt, dann folgt aus der Bäcker-Definition, dass er nicht für sich bäckt.
- Wenn der Satz falsch wäre, der Bäcker also nicht auch für sich bäckt, dann folgt aus der Bäcker-Definition, dass er für sich bäckt.

Ursprünglich wurde dieses Beispiel 1918 als Barbier-Paradoxon veröffentlicht.

Nebenbemerkung

Bei der alltäglichen Arbeit mit den mathematischen Werkzeugen werden wir praktisch nie auf solch problematische Situationen treffen wie in den obigen Beispielen zu Nicht-Aussagen. Man könnte daher zu der Meinung gelangen, dass derartige Probleme nur bei der Verwendung von Alltagssprache auftreten und somit in einer potentiell unpräzisen Formulierung begründet sind. Wenn man nur mathematische Begriffe und Symbole verwendet, geht bestimmt alles seinen geordneten Gang...

Leider falsch! Selbst in klar abgegrenzten, wohl durchdachten und fehlerfrei konstruierten mathematischen Strukturen, kann man stets “Aussagen” formulieren, von denen man mathematisch exakt beweisen kann, dass man nicht entscheiden kann, ob sie wahr oder falsch sind. In unserem Sinne handelt es sich dann also nicht um Aussagen. Diese für die bis dahin sehr heile Mathematikerwelt erschütternde Erkenntnis wurde 1931 von Kurt Gödel veröffentlicht und ist als Gödelscher

Unvollständigkeitssatz bekannt.

Tatsächlich sind unter den unentscheidbaren "Aussagen" auch leicht verständliche zu finden. Zum Beispiel ist folgende Frage zur Mächtigkeit von Mengen unentscheidbar (Kontinuumshypothese):

> Es gibt eine Menge, die mächtiger ist als die natürlichen Zahlen, aber weniger mächtig als die reellen Zahlen.

Zum Begriff der Mächtigkeit vergleiche auch Aufgabe 1.2 in Übung 1.

Den Ansatz, dass Aussagen entweder wahr oder falsch sind, bezeichnet man als zweiwertige Logik. Weniger verbreitete alternative Ansätze sind die dreiwertige Logik, die Fuzzy-Logik und die intuitionistische Logik.

4.1.2 Äquivalenz von Aussagen

Wir werden bei der Beschäftigung mit Mathematik häufig Aussagen umformen bzw. umformulieren. Ziel ist, diese zu vereinfachen ohne dabei deren Sinngehalt zu verändern. Sind zwei Aussagen entweder beide gleichzeitig wahr oder beide gleichzeitig falsch, so heißen die Aussagen **äquivalent**. Man sagt auch: die eine Aussage gilt **genau dann, wenn** die andere Aussage gilt. Als Symbol verwendet man den **Äquivalenzpfeil** $\Leftrightarrow$ (ein zweiseitiger Pfeil mit doppelter Linie):

$$\text{Aussage 1} \quad \Leftrightarrow \quad \text{Aussage 2.} \tag{4.1}$$

Gilt nur die eine Richtung, d.h. folgt aus der Gültigkeit von Aussage 1, dass auch Aussage 2 wahr ist, so schreibt man

$$\text{Aussage 1} \quad \Rightarrow \quad \text{Aussage 2.} \tag{4.2}$$

Sprechweise: "Aussage 1 **impliziert** Aussage 2" oder "aus Aussage 1 **folgt** Aussage 2". Der Pfeil $\Rightarrow$ heißt **Implikationspfeil** oder **Folgerungspfeil**. Alternative Schreibweise:

$$\text{Aussage 2} \quad \Leftarrow \quad \text{Aussage 1.} \tag{4.3}$$

Beispiel (Umformen von Gleichungen)

Äquivalenzpfeile können und sollten genutzt werden um die einzelnen Schritte beim Umformen von mathematischen Ausdrücken (z.B. Gleichungen oder Ungleichungen) miteinander in Beziehung zu setzen. Soll die Gleichung

$$4\,x - 5 = 19 \tag{4.4}$$

gelöst werden, so kann der Rechenweg als Kette äquivalenter Aussagen aufge-

schrieben werden:

$$4\,x - 5 = 19 \;\Leftrightarrow\; 4\,x = 24 \;\Leftrightarrow\; x = 6. \tag{4.5}$$

Zeilenumbrüche sind natürlich möglich und meist auch zweckmäßig im Sinne der Übersichtlichkeit:

$$\begin{aligned} & 4\,x - 5 = 19 \\ \Leftrightarrow\; & 4\,x = 24 \\ \Leftrightarrow\; & x = 6. \end{aligned} \tag{4.6}$$

Beispiel (nicht äquivalente Umformungen)

Bei Lösen von Gleichungen und Ungleichungen muss beachtet werden, dass nicht alle Arten von Umformungsschritten äquivalente Aussagen liefern. Gesucht seien beispielsweise alle $x \in \mathbb{R}$, die $x+2 = \sqrt{x+4}$ erfüllen. Wir schreiben zunächst nur Implikationspfeile in "Rechenrichtung" und prüfen anschließend, ob jeweils auch die Gegenrichtung gilt:

$$\begin{aligned} & x + 2 = \sqrt{x+4} \\ \Rightarrow\; & (x+2)^2 = x + 4 && \text{(Quadrieren)} \\ \Rightarrow\; & x^2 + 4\,x + 4 = x + 4 && \text{(Ausmultiplizieren)} \\ \Rightarrow\; & x^2 + 3\,x = 0 && \text{(Zusammenfassen)} \\ \Rightarrow\; & x\,(x+3) = 0 && (x \text{ ausklammern}) \\ \Rightarrow\; & x = 0 \quad \text{oder} \quad x = -3 && \text{(Faktoren Null setzen).} \end{aligned} \tag{4.7}$$

Setzen wir nun $x = 0$ und $x = -3$ in die ursprüngliche Gleichung ein, so stellen wir fest, dass nur $x = 0$ eine Lösung ist. Für $x = -3$ erhalten wir hingegen die falsche Aussage $-1 = 1$. Der Rechenweg liefert zwar

$$x + 2 = \sqrt{x+4} \qquad \Rightarrow \qquad x = 0 \quad \text{oder} \quad x = -3; \tag{4.8}$$

die Probe hingegen ergibt

$$x + 2 = \sqrt{x+4} \qquad \not\Leftarrow \qquad x = 0 \quad \text{oder} \quad x = -3, \tag{4.9}$$

wobei $\not\Leftarrow$ für "die Impllikation gilt nicht" steht.

Mindestens einer der durchgeführten Rechenschritte ist demnach nicht umkehrbar. Geht man die Schritte rückwärts durch, so stellt man fest, das der erste Rechenschritt (Quadrieren) nicht umkehrbar ist, denn das Wurzelziehen als Umkehrung des Quadrierens liefert die Implikation

$$(x+2)^2 = x + 4 \qquad \Rightarrow \qquad |x+2| = \sqrt{x+4}. \tag{4.10}$$

Die Betragsstriche tauchen jedoch in der ursprünglichen Gleichung nicht auf. Reduziert auf das Wesentliche gilt für eine beliebige reelle Zahl $a \in \mathbb{R}$

$$\sqrt{a^2} = |a|. \tag{4.11}$$

Klar wird dies am Beispiel $a = -2$. Dann ist

$$\sqrt{a^2} = \sqrt{(-2)^2} = \sqrt{4} = 2 = |-2| = |a| \neq a. \tag{4.12}$$

Merke: Im Zweifelsfall sollte beim Umformen von Gleichungen und Ungleichungen die Gültikeit der Implikation in beide Richtungen separat geprüft werden! Es ist streng zu unterscheiden zwischen äquivalenten Umformungen und einfachen Folgerungen, die eine allgemeinere, nicht äquivalente Aussage liefern.

4.1.3 Beschreiben von Mengen durch Aussagen

Ein ganz grundlegendes sprachliches Mittel in der Mathematik sind Mengen. Wir werden häufig mit Mengen von Objekten (Zahlen, Vektoren,...) mit gewissen Eigenschaften arbeiten. Welche Eigenschaften gegeben sein müssen, damit ein Objekt zur betrachteten Menge gehört, kann man mittels Aussagen über das Objekt formulieren. Die betrachtete Menge wird dann festgelegt als die Menge aller Objekte, für die eine gegebene Aussage über das einzelne Objekt wahr ist. Schreibweise:

$$M := \{x \in \text{Grundmenge} : \text{ Aussage über } x\}. \tag{4.13}$$

Dabei sind

- M das Symbol für die zu beschreibende Menge,
- $:=$ der Zuweisungsoperator ("gib dem rechten Objekt das linke Symbol als Name"),
- "Grundmenge" die Menge aller zunächst betrachteten Objekte.

Beispiel (Menge der geraden Zahlen)

Die Menge der geraden Zahlen kann als

$$M := \{x \in \mathbb{N} : \ x \text{ ist durch 2 teilbar}\} \tag{4.14}$$

geschrieben werden. Statt "8 ist eine gerade Zahl" können wir dann $8 \in M$ schreiben, was viel kürzer ist.

In der Mathematik sind solche die sprachliche Formulierung abkürzenden Mittel wesentliches Werkzeug um auch komplexe Zusammenhänge auf wenig Raum und damit besser überschaubar darzustellen. So kann beispielsweise die Aussage

Die Summe zweier Quadrate gerader Zahlen ist wieder eine gerade

Zahl.

nun als

$$x, y \in M \quad \Rightarrow \quad x^2 + y^2 \in M \tag{4.15}$$

geschrieben werden.

Merke: Die vielen Symbole und Formeln in der Mathematik dienen nicht dem Quälen von Studierenden, sondern sind die **Arbeitgrundlage für präzises und überschaubares Formulieren komplexer Zusammenhänge**.

4.1.4 Logik vs. Mengenoperationen

Zwei (oder mehr) gegebene Aussagen können durch sogenannte logische Operation zu neuen Aussagen verknüpft werden. So ist die Formulierung

Es gilt Aussage 1 oder Aussage 2.

selbst wieder eine Aussage. Analog gibt es die Und-Verknüpfung, die Entweder-Oder-Verknüpfung und die Nicht-Operation. Letztere ordnet einer Aussage genau die gegenteilige Aussage zu.

Logische Operationen mit Aussagen übertragen sich eins-zu-eins auf Mengen. Dort entstehen dann z.B. Vereinigungen (Oder-Verknüpfung) und Durchschnitte (Und-Verknüpfungen):

$$\begin{aligned} &\{x \in \mathbb{N} : \ x \text{ durch 2 teilbar } \textbf{oder} \text{ durch 3 teilbar}\} \\ &\qquad = \{x \in \mathbb{N} : \ x \text{ durch 2 teilbar}\} \cup \{x \in \mathbb{N} : \ x \text{ durch 3 teilbar}\} \\ &\{x \in \mathbb{N} : \ x \text{ durch 2 teilbar } \textbf{und} \text{ durch 3 teilbar}\} \\ &\qquad = \{x \in \mathbb{N} : \ x \text{ durch 2 teilbar}\} \cap \{x \in \mathbb{N} : \ x \text{ durch 3 teilbar}\} \end{aligned} \tag{4.16}$$

Nebenbemerkung

Die Und-Verknüpfung mehrerer Aussagen schreibt man gern auch als durch Komma getrennte Auflistung der Einzelaussagen. Beispielsweise sind die Schreibweisen

$$\{x \in \mathbb{N} : \ x \text{ durch 2 teilbar und durch 3 teilbar}\} \tag{4.17}$$

und

$$\{x \in \mathbb{N} : \ x \text{ durch 2 teilbar}, \ x \text{ durch 3 teilbar}\} \tag{4.18}$$

völlig gleichbedeutend. Bei Oder-Verknüpfungen wird das "oder" hingegeben stets explizit angegeben.

Nebenbemerkung

Gelegentlich sieht man statt “und” und “oder” zur Verknüpfung von Aussagen die Symbole $\wedge$ und $\vee$. Für “nicht” wird gelegentilich $\neg$ verwendet. Diese Symbolik erschwert die Lesbarkeit und wird deshalb hier vermieden. Bei manchen Autoren im Bereich Ingenieurmathematik hat man den Eindruck, dass das Verwenden dieser Symbole die “mathematische Bildung” des Autors unterstreichen soll. In Mathematikbüchern für Mathematiker spielen diese Symbole bis auf wenige Spezialgebiete hingegen keine Rolle.

Beispiel

Das Zusammenspiel von Logik und Mengenlehre lässt sich gut an der Beschreibung geometrischer Objekte nachvollziehen. Betrachten dazu in der Ebenen die drei Teilmengen

- Kreuz:

$$\{(x,y) \in \mathbb{R}^2 : \ x \in [1,3] \text{ oder } y \in [0,1]\}, \tag{4.19}$$

- Rechteck:

$$\{(x,y) \in \mathbb{R}^2 : \ x \in [1,3], \ y \in [0,1]\}, \tag{4.20}$$

- Kreisring:

$$\{(x,y) \in \mathbb{R}^2 : \ 1 \leq x^2 + y^2 \leq 4\} \tag{4.21}$$

(Video 410, Notizen dazu).

4.2 Abbildungen

Abbildungen sind neben Mengen das wichtigste sprachliche Mittel in der Mathematik. Oft interessieren weniger die Objekte (z.B. Zahlen) selbst, sondern deren Beziehungen zueinander. Eine Analogie im täglichen Leben ist die Bedeutung von Daten in "sozialen" Netzwerken und Messenger-Diensten. Es ist kaum von Interesse was jemand schreibt oder sagt, sondern vor allem wer mit wem kommuniziert. Entsprechend stehen nicht die (meist zumindest oberflächlich verschlüsselten) Inhalte im Fokus von Überwachungsorganisationen und Datenhändlern, sondern die Metadaten, also wer wann mit wem kommuniziert hat.

4.2.1 Begriffe und Beispiele

Merke

Eine **Abbildung** ist eine Zuordnung zwischen den Elementen zweier Mengen A und B, bei denen jedem Element der Menge A genau ein Element der Menge B zugeordnet wird. Die Menge A wird als **Definitionsbereich** der Abbildung bezeichnet. Die Menge aller Elemente von B, die einem Element von A zugeordnet sind, heißt **Wertebereich** der Abbildung. Die Menge B selbst wird als Bildbereich oder als Grundmenge des Wertebereichs bezeichnet.

Zur einfacheren Verwendung bezeichnet man Abbildungen meist mit einem Symbol, z.B. f, und schreibt $f : A \to B$, wenn man Definitions- und Bildbereich gleich mit angegeben möchte. Beachte, dass B üblicherweise nicht der Wertebereich selbst, sondern eine Obermenge des Wertebereichs ist. Das genaue Bestimmen des Wertebereichs erfordert meist etwas Mühe, die man sich auf diese Weise (zunächst) spart.

Der Definitionsbereich einer Abbildung f wird auch mit $D(f)$ bezeichnet, der Wertebereich mit $W(f)$. Diese Symbole sind allerdings in der Literatur nicht einheitlich, sodass stets genau zu prüfen ist, was mit welchem Symbol gemeint ist.

Die Festlegung der Zuordnungsvorschrift kann entweder in Alltagssprache oder in Form von mathematischen Ausdrücken erfolgen. Ist $f : \mathbb{R} \to \mathbb{R}$ beispielsweise eine Abbildung, die jeder reellen Zahl x wieder eine reelle Zahl zuordnet, so schreibt man $f(x) := x^2$ oder $f : x \mapsto x^2$ (beachte die spezielle Form des Pfeilanfangs!). Umgekehrt verwendet man bei gegebener Abbildung $f : X \to Y$ das Symbol $f(x)$ für das Element in Y, auf das ein gegebenes Element $x \in X$ durch f abgebildet wird.

Abbildungen treten in ganz verschiedenen Varianten auf. Weiter unten betrachten wir einige wichtige Spezialfälle; hier zunächst eine buntere Beispielsammlung.

Nebenbemerkung

Die Angabe des konkreten Definitionsbereichs ist nicht immer ganz einfach. Deshalb wird dieser manchmal zunächst weggelassen. Der Definitionsbereich soll dann aus all den Elementen einer angegebenen Grundmenge bestehen, für die die Abbildung sinnvolle Werte liefert.

Beispiel (Flächeninhalt)

Sei F die Menge aller Teilmengen von $\mathbb{R}^2$, für die der Begriff "Flächeninhalt" sinnvoll erklärt werden kann (F für "Flächen"). Dann kann man die Abbildung $A : F \to \mathbb{R}$ betrachten, die jeder Fläche $f \in F$ deren Flächeninhalt (reelle Zahl!) zuordnet. Die Abbildung A lässt sich offensichtlich nicht als Formel aufschreiben. Überhaupt müsste noch geklärt werden, wie der Definitionsbereich F genau aussieht (werden wir später im Semester klären).

Wichtig: Gewöhnen Sie sich von Anfang an an Flexibiliät bei der Wahl von Symbolen. Eine Abbildung muss keineswegs immer f heißen!

Beispiel (Klausurnoten)

Sei N die Menge aller Noten, die bei der Bewertung der Klausuren eines bestimmten Studierendenjahrgangs auftraten. Die Abbildung S ordne jeder Note die Menge aller Studierenden zu, die diese Note erhalten haben. Beachte:

- Der Definitionsbereich von S ist auch dann die komplette Menge N, wenn nicht jede Note vergeben worden ist. Beispielsweise könnte die Note 5.0 nicht auftauchen (alle haben bestanden). Dann wäre $S(5.0) = \emptyset$ (leere Menge).
- Die genaue Formulierung ist hier wichtig. Falsch wäre

  Jeder Note werden die Studierenden mit dieser Note zugeordnet.

  Denn diese Formulierung widerspricht der Forderung, dass den Elementen des Definitionsbereichs jeweils genau ein Element des Wertebereichs zugeordnet wird. In der korrekten Formulierung oben wird jeder Note eine (!) Menge von Studierenden zugeordnet. Ob solch eine Menge dann kein oder ein oder mehrere Elemente enthält, ist unerheblich für die formale Korrektheit der Definition der Abbildung.

Beispiel (gerade Zahlen)

Definieren die Abbildung $f : \mathbb{N} \to \mathbb{N}$ durch $n \mapsto 2\,n$. Jeder natürlichen Zahl wird

also ihr Doppeltes zugeordnet.

Zur Beschreibung grundlegender Eigenschaften von Abbildungen führen wir folgende Begriffe ein.

Merke

Sei $f : A \to B$ eine Abbildung.

- Die Abbildung f heißt **injektiv**, wenn jedes Element von B höchstens einem Element von A zugeordnet wird. Kein Element von B wird also zwei verschiedenen Elementen von A zugeordnet. In Formeln:

$$\text{für alle } a_1, a_2 \in A \text{ gilt:} \quad f(a_1) = f(a_2) \;\Rightarrow\; a_1 = a_2. \tag{4.22}$$

- Die Abbildung f heißt **surjektiv**, wenn jedes Element von B mindestens einem Element von A zugeordnet ist. In Formeln:

$$W(f) = B. \tag{4.23}$$

- Die Abbildung f heißt *bijektiv*, wenn sie gleichzeitig injektiv und surjektiv ist.

Beachte, dass Surjektivität von der Wahl der Grundmenge B abhängt. Bei gleichem Wertebereich kann B unterschiedlich aussehen. So ist beispielsweise die Abbildung $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ nicht surjektiv, die Abbildung $g : \mathbb{R} \to [0, \infty)$, $x \mapsto x^2$ hingegen schon.

Zu den obigen drei Beispielen:

- Die Zuordnung des Flächeninhalts ist nicht injektiv, da verschiedene Flächen den gleichen Inhalt haben können. Sie ist aber surjektiv, da wir zu jeder reellen Zahl eine Fläche mit diesem Inhalt finden können. Die Abbildung ist nicht bijektiv.
- Die Zuordnung der Studierenden zu den Klausurnoten ist injektiv, denn an die gleiche Gruppe von Studierenden können nicht zwei verschiedene Noten vergeben worden sein. Zur Klärung der Surjektivität müssen wir uns zunächst auf eine Grundmenge für den Wertebereich festlegen. Da prinzipiell jede Kombination von Studierenden bei diesem Abbildungstyp auftreten kann, ist die **Potenzmenge** der Studierenden (also die Menge aller möglichen Teilmengen von Studierenden) eine geeignete Wahl. Da in der Abbildung viel weniger Noten zur Verfügung stehen als es verschiedene Teilmengen von Studierenden gibt, ist die Abbildung offensichtlich nicht surjektiv. Somit ist sie auch nicht bijektiv.
- Die Zuordnung der geraden Zahlen zu den natürlichen Zahlen ist sowohl injektiv (keine gerade Zahl wird zwei verschiedenen Zahlen zugeordnet) als auch surjektiv

(jede gerade Zahl tritt in der Zuordnung auf). Also ist die Abbildung bijektiv.

Die Frage nach Injektivität und Surjektivität kann auch anhand der Lösbarkeit von Gleichungen beschrieben werden: Ist $f : X \to Y$ eine beliebige Abbildung und ist $y \in Y$, so können wir nach Lösbarkeit und Lösungen $x \in X$ der Gleichung

$$f(x) = y \tag{4.24}$$

fragen. Offensichtlich besitzt die Gleichung genau dann für jedes $y \in Y$ eine Lösung, wenn f surjektiv ist. Außerdem ist für y aus dem Wertebereich von f die Lösung genau dann eindeutig, wenn f injektiv ist.

Folgender Begriff taucht gelegentlich auf, lierfert aber inhaltlich keine neuen Aspekte. Er ist ein Beispiel für Begriffe, die rein zur sprachlichen Vereinfachung bzw. Fixierung dienen.

Merke

Sei $f : X \to Y$ eine Abbildung. Die Menge

$$\{(x, f(x)) \in X \times Y : \ x \in X\} \tag{4.25}$$

heißt **Graph** der Abbildung f.

In die gleiche Rubrik fällt der folgende Begriff:

Merke

Sind $f : A \to B$ und $g : B \to C$ zwei Abbildungen, so heißt die durch $a \mapsto g(f(a))$ gegebene Abbildung **Komposition** oder **Hintereinanderausführung** von f und g. Als Symbol für diese Abbildung verwenden wir $g \circ f$.

4.2.2 Spezialfall: Punkte und Vektoren

Sei $n \in \mathbb{N}$ eine feste gegebene Zahl (oft $n = 2$ oder $n = 3$). Ein **Punkt** ist eine n-Tupel aus n reellen Zahlen. 2-Tupel sind auch als Paare bekannt, 3-Tupel als Tripel, 4-Tupel als Quadrupel usw. Alternativ kann man einen Punkt als Abbildung der Menge $\{1, 2, \ldots, n\}$ in die reellen Zahlen ansehen. Die n Bildelemente einer solchen Abbildung nennt man **Koordinaten** des Punktes. Statt der Klammerschreibweise für das Auswerten einer Abbildung nutzt man hier einfache Auflistungen der Koordinaten. So entspricht die Auflistung $(7, 4, 5)$ beispielsweise der Abbildung

$$1 \mapsto 7, \quad , 2 \mapsto 4, \quad 3 \mapsto 5. \tag{4.26}$$

Die Menge aller Punkte zu einem festen n wird mit dem Symbol $\mathbb{R}^n$ bezeichnet. So umfasst $\mathbb{R}^2$ beispielsweise alle Punkte mit zwei Koordinaten. $\mathbb{R}^2$ wird auch als “die Ebene” und $\mathbb{R}^3$ als “der Raum” bezeichnet.

Das Wort **Vektor** ist aus technischer Sicht ein Synonym für Punkt, meint also ein Element aus $\mathbb{R}^n$. Man unterscheidet die Elemente von $\mathbb{R}^n$ jedoch gelegentlich nach deren Verwendung. Soll durch ein Element aus $\mathbb{R}^n$ ein geometrischer Ort beschrieben werden ("Punkt im Koordinatensystem"), so verwendet man den Begriff Punkt. Soll jedoch eine Richtung beschrieben werden, so spricht man von einem Vektor. Während Punkte, wie oben eingeführt, als richtungslose Auflistung $(a_1, \ldots, a_n)$ ihrer Koordinaten geschrieben werden, werden Vektoren meist als vertikale Listen

$$\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \text{oder} \quad \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \tag{4.27}$$

ihrer Koordinaten dargestellt. Bei Vektoren spricht man statt von Koordinaten auch von den **Komponenten** des Vektors.

Die Arbeit mit Punkten und Vektoren wird später detaillierter betrachtet.

4.2.3 Spezialfall: Zahlenfolgen

Eine Abbildung, die jeder natürlichen Zahl eine reelle Zahl zuordnet, heißt **Zahlenfolge**. Man weicht hier aus Gewohnheit und aus Gründen der Platzeinsparung von der üblichen Schreibweise für das Auswerten von Abbildungen ab: Ist $a : \mathbb{N} \to \mathbb{R}$ eine Zahlenfolge, so setzt man

$$a_n := a(n) \quad \text{für alle } n \in \mathbb{N}. \tag{4.28}$$

Man schreibt auch gern $a = (a_1, a_2, \ldots)$ im Sinne eines Punktes (oder Vektors) mit unendlich vielen Koordinaten oder $a = (a_n)_{n\in\mathbb{N}}$.

Gelegentlich treten in der Mathematik **doppelt unendliche Zahlenfolgen** auf, insbesondere in der Signalverarbeitung. Dies sind Abbildungen $a : \mathbb{Z} \to \mathbb{R}$, die auch negative Indizes nutzen, also nach beiden Richtungen laufende Folgen darstellen. In dieser Veranstaltung werden wir solchen Folgen jedoch nicht begegnen.

Zahlenfolgen werden wir in Kürze noch näher betrachten.

4.2.4 Spezialfall: Funktionen

Bezeichnen Abbildungen zwischen Teilmengen der reellen Zahlen oder Mengen von aus endlich vielen reellen Zahlen zusammengesetzten Objekten (Punkte, Vektoren, Matrizen,...) als **Funktion**. Dabei soll der Definitionsbereich nicht nur aus einzelnen Zahlen, Punkten usw. bestehen, sondern den kontinuierlichen Charakter der reellen Zahlen tragen. Beispielsweise sind Zahlenfolgen in diesem Sinne also keine Funktionen.

Gelegentlich werden die Begriffe Abbildung und Funktion synonym verwendet; verstehen in dieser Veranstaltung unter Funktionen jedoch nur zahlenbezogene und keine allgemeinen Abbildungen.

Funktionen spielen eine sehr wesentliche Rolle in der angewandten Mathematik, sodass wir ihnen ein eigenes Unterkapitel widmen werden.

4.2.5 Spezialfall: Allgemeine Folgen

Neben Zahlenfolgen ist gelegentlich auch die Betrachtung anderer Folgen zweckmäßig. Zum Beispiel werden wir im Laufe der Zeit mit Folgen von Punkten in der Ebene arbeiten (müssen). Dies sind dann Abbildungen von $\mathbb{N}$ nach $\mathbb{R}^2$. Auch Folgen von Funktionen spielen in der Mathematik und ihren Anwendungen eine wichtige Rolle; jedoch weniger in der Grundlagenausbildung. Wenn wir uns im Laufe des Semesters mit dem Begriff des Flächeninhalts befassen, werden wir auf Folgen von Mengen stoßen, wobei jede dieser Mengen wiederum eine Ansammlung von Mengen sein wird. Folgen, also Abbildung der natürlichen Zahlen in andere Mengen, sind überall...

4.3 Zahlenfolgen

Der Begriff der Zahlenfolge wurde bereits eingeführt (vgl. Section 4.2.3).

Nebenbemerkung (Folge vs. Menge)

Beachte, dass bei einer Zahlenfolge die Reihenfolge der Elemente ganz wesentlich ist. Hingegen ist bei der aus den Elementen der Zahlenfolge gebildeten Menge $\{a_1, a_2, \ldots\}$ die Reihenfolge egal. Insbesondere kann die Menge viel weniger Elemente enthalten als die Zahlenfolge. Betrachtet man beispielsweise die konstante Zahlenfolge $a = (1, 1, \ldots)$, so besteht die zugehörige Menge der Elemente aus nur einem Element: $\{a_1, a_2, \ldots\} = \{1\}$, ganz im Gegensatz zur unendlich viele Elemente enthaltenden Zahlenfolge.

4.3.1 Visualisierung

Die Visualisierung von Zahlenfolgen kann mit dem Python-Paket Matplotlib (und vielen anderen) erfolgen. Zu beachten ist, dass die einzelnen Punkte nicht durch Linien verbunden werden, da Zahlenfolgen nur auf der Menge $\{1, 2, \ldots\}$ definiert sind und zwischen diesen Werten nicht.

Beispiel

Die Zahlenfolge

$$a_n := \frac{1}{n}, \quad n \in \mathbb{N} \tag{4.29}$$

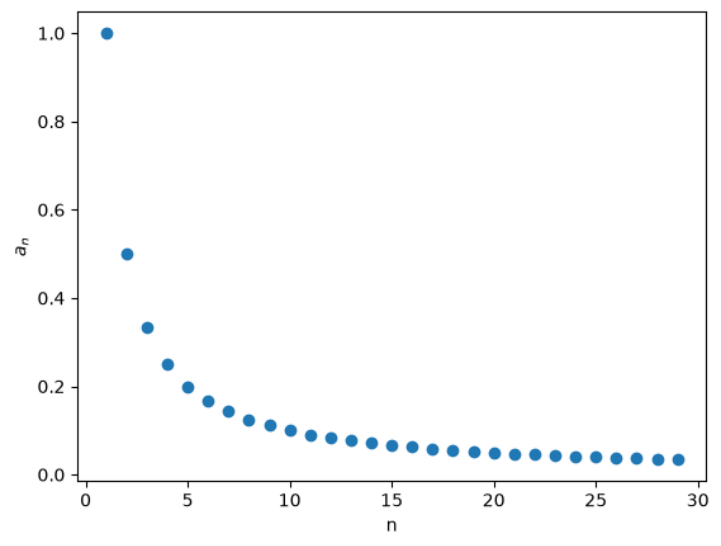
kann wie folgt visualisiert werden:

```
import numpy as np
import matplotlib.pyplot as plt

n = np.arange(1, 30)  # Indizes der Zahlenfolge (natürlich nur
↪  endlich viele)
a = 1 / n  # Werte der Zahlenfolge (a ist ein Array mit 29
↪  Einträgen!)

fig, ax = plt.subplots()  # Zeichenfläche und Koordinatensystem
↪  erstellen
ax.plot(n, a, 'o')  # 'o' für "keine Linien, aber dicke Punkte"
ax.set_xlabel('n')  # Beschriftung horizontale Achse
ax.set_ylabel('$a_n$')  # Beschriftung vertikale Achse
                        # (Mathe -Modus für Formeln mit $...$
                        ↪  aktivieren)
```

```
plt.show()  # Zeichenfläche anzeigen
```



4.3.2 Bedeutung

Zahlenfolgen tauchen gelegentlich bei praktischen Aufgabenstellungen auf. Im Rahmen unserer Veranstaltung werden sie aber überwiegend als innermathematisches Werkzeug dienen. Insbesondere bilden sie die Basis für die Konstruktion komplexerer Werkzeuge wie dem Stetigkeitsbegriff, der Differentialrechnung und der Integralrechnung.

Beispiel (Zinseszins)

Als einfaches Beispiel für das Auftreten von Zahlenfolgen in der Praxis kann man üblicherweise auf Jahresbasis stattfindende Kapitalentwicklungen anschauen: Sei ein Startguthaben $g_0 > 0$ gegeben. Jedes Jahr wird das vorhandene Guthaben mit 1 Prozent Zinsen vergütet. Die Zinsen werden dem Guthaben zugeschlagen und im weiteren Verlauf mit verzinst. Nach einem Jahr steht somit ein größeres Guthaben g_1 zur Verfügung; nach zwei Jahren das Guthaben g_2 usw. Konkret erhalten wir die Zahlenfolge

$$(g_1, g_2, \ldots) = (1.01\, g_0,\ 1.01^2\ g_0,\ \ldots) \tag{4.30}$$

bzw.

$$g_n = 1.01^n\, g_0, \quad n \in \mathbb{N}. \tag{4.31}$$

In einer realen Situation wird die Folge nach ein paar Jahren abbrechen, also gar keine Zahlenfolge im Sinne der mathematischen Begriffs sein. Das Guthaben

wird nicht unendlich viele Jahre wachsen; selbst wenn, wird kein Mensch unendlich lange leben um dieses Guthaben "am Ende" in Empfang nehmen zu können. Das mathematische Werkzeug der Zahlenfolge erlaubt jedoch, den Zinseszinsprozess zu modellieren ohne sich über die konkrete Dauer des Prozesses Gedanken machen zu müssen. Wir vereinfachen (!) hier also den realen Vorgang zu einem mathematischen Modell, welches eine schwierige reale Frage (Wie lange soll die Geldanlage laufen?) umgeht, indem es die völlig realitätsferne Idee des Unendlichen einbezieht.

4.3.3 Monotonie, Periodizität, Beschränktheit

Führen einige Begriffe ein, die im Kontext von Zahlenfolgen gelegentlich Verwendung finden.

Merke

Eine Zahlenfolge $(a_1, a_2, \ldots)$ heißt

- **monoton wachsend**, wenn $a_n \leq a_{n+1}$ für alle $n \in \mathbb{N}$ gilt,
- **streng monoton wachsend**, wenn $a_n < a_{n+1}$ für alle $n \in \mathbb{N}$ gilt,
- **monoton fallend**, wenn $a_n \geq a_{n+1}$ für alle $n \in \mathbb{N}$ gilt,
- **streng monoton wachsend**, wenn $a_n > a_{n+1}$ für alle $n \in \mathbb{N}$ gilt.

Beispielsweise ist die Zahlenfolge in obigem Zinseszinsbeispiel streng monoton wachsend.

Merke

Eine Zahlenfolge $(a_1, a_2, \ldots)$ heißt **periodisch**, wenn es ein $N \in \mathbb{N}$ gibt, sodass

$$a_n = a_{n+N} \quad \text{für alle } n \in \mathbb{N} \tag{4.32}$$

gilt. Jede Zahl N mit dieser Eigenschaft heißt **Periode** der Zahlenfolge. Die kleinste Zahl N mit dieser Eigenschaft heißt **Grundperiode** der Zahlenfolge.

Beispielsweise ist die durch $a_n = (-1)^n$, $n \in \mathbb{N}$, gegebene Folge periodisch mit Periode $N = 2$.

Merke

Eine Zahlenfolge $(a_1, a_2, \ldots)$ heißt **beschränkt**, wenn es ein $c \in \mathbb{R}$ gibt, sodass

$$|a_n| \leq c \quad \text{für alle } n \in \mathbb{N} \tag{4.33}$$

gilt. Ansonsten heißt die Folge **unbeschränkt**.

Beispielsweise ist die durch $a_n := n$, $n \in \mathbb{N}$, gegebene Folge unbeschränkt. Hingegen ist die durch $a_n := \frac{1}{n}$, $n \in \mathbb{N}$, gegebene Folge beschränkt (mit $c = 1$ oder größer).

4.3.4 Konvergenz

Fast immer interessieren bei einer Zahlenfolge gar nicht die vorderen Elemente, sondern nur das Verhalten der Folge "im Unendlichen". Werden die Folgenelemente beliebig groß (oder klein) oder nähern sie sich einem konkreten Wert an usw.? Zahlenfolgen können regelrecht als Werkzeug für das Arbeiten mit "dem Unendlichen" angesehen werden.

Merke

Eine Zahlenfolge $(a_1, a_2, \ldots)$ heißt **konvergent**, wenn es eine Zahl $g \in \mathbb{R}$ gibt, sodass für jedes $\varepsilon > 0$ ein $m \in \mathbb{N}$ existiert, sodass

$$|a_n - g| < \varepsilon \quad \text{für alle } n \geq m \tag{4.34}$$

gilt. Die Zahl g heißt dann **Grenzwert** der Folge. Man schreibt auch

$$\lim_{n \to \infty} a_n = g \tag{4.35}$$

oder

$$a_n \to g \quad \text{für } n \to \infty. \tag{4.36}$$

Eine Folge ist also konvergent, wenn ihre Elemente a_n für hinreichend großes n in einem beliebig klein gewählten Intervall $(g - \varepsilon, g + \varepsilon)$ bleiben.

Beispiel

Die durch $a_n := \frac{1}{n}$ gegebene Folge konvergiert gegen 0 (Video 420, Notizen dazu).

Merke

Eine Zahlenfolge $(a_1, a_2, \ldots)$ heißt **divergent**, wenn sie nicht konvergent ist. Man unterscheidet bestimmt divergente Folgen und unbestimmt divergente Folgen.

Eine divergente Folge heißt **bestimmt divergent**, wenn für jede Zahl $c \in \mathbb{R}$ ein $m \in \mathbb{N}$ existiert, sodass

$$a_n > c \quad \text{für alle } n \geq m \tag{4.37}$$

gilt, oder wenn für jede Zahl $c \in \mathbb{R}$ ein $m \in \mathbb{N}$ existiert, sodass

$$a_n < c \quad \text{für alle } n \geq m \tag{4.38}$$

gilt. Man schreibt dann auch

$$\lim_{n\to\infty} a_n = \infty \qquad \text{bzw.} \qquad \lim_{n\to\infty} a_n = -\infty. \tag{4.39}$$

Andernfalls heißt die Folge **unbestimmt divergent**.

Bestimmte Divergenz liegt also vor, wenn eine Folge über jede Schranke wächst oder unter jede Schranke fällt. Ersteres ist bei obigem Zinseszinsbeispiel der Fall.

Beispiele für unbestimmte Divergenz sind (nicht konstante) periodische Folgen.

4.3.5 Berechnen von Grenzwerten

Das Berechnen von Grenzwerten von Zahlenfolgen ist nur sehr selten nötig. Wir werden zwar viel mit Zahlenfolgen arbeiten, benötigen dabei aber nur das sprachliche Konstrukt und die damit möglichen Formulierungen zu Abläufen mit unendlich vielen Einzelschritten. Geben deshalb hier nur einige Hinweise für das konkrete Ausrechnen von Grenzwerten mit auf den Weg.

4.3.5.1 Manuelles Rechnen Für einfache Zahlenfolgen kann der Grenzwert (so er denn existiert) manuell berechnet werden. Dazu bedient man sich einer Tabelle von Grenzwerten für sehr elementare Folgen und einiger Rechenregeln zu Grenzwerten beim Bilden von komplexeren Zahlenfolgen aus diesen elementaren Folgen.

Merke (Grenzwerte elementarer Zahlenfolgen)

Hier einige nützliche Grenzwerte:

- konstante Folge:

$$\lim_{n\to\infty} c = c \quad \text{für beliebiges } c \in \mathbb{R}, \tag{4.40}$$

- Potenzen:

$$\lim_{n\to\infty} n^x = \begin{cases} 0, & \text{für } x < 0, \\ 1, & \text{für } x = 0, \\ \infty, & \text{für } x > 0, \end{cases} \tag{4.41}$$

- Euler'sche Zahl:

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = \mathrm{e} \approx 2.71828. \tag{4.42}$$

Merke (Rechenregeln für Grenzwerte)

Mit Hilfe folgender Eigenschaften von Grenzwerten können aus Grenzwerten elementarer Folgen Grenzwerte komplexerer Folgen berechnet werden. Dabei sind $(a_1, a_2, \ldots)$ und $(b_1, b_2, \ldots)$ stets zwei konvergente (!) Folgen.

- Die elementweise Summe zweier Folgen konvergiert gegen die Summe der Grenzwerte:

$$\lim_{n\to\infty}(a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n. \tag{4.43}$$

- Die Produkt zweier Folgen konvergiert gegen das Produkt der Grenzwerte:

$$\lim_{n\to\infty}(a_n\, b_n) = \left(\lim_{n\to\infty} a_n\right)\left(\lim_{n\to\infty} b_n\right). \tag{4.44}$$

- Das Vielfache einer Folge konvergiert gegen das Vielfache des Grenzwerts:

$$\lim_{n\to\infty}(c\, a_n) = c \lim_{n\to\infty} a_n \quad \text{für beliebiges } c \in \mathbb{R}. \tag{4.45}$$

- Sind alle Folgenelemente und der Grenzwert von Null verschieden, so konvergiert das Reziproke einer Folge gegen das Reziproke des Grenzwerts:

$$\lim_{n\to\infty}\frac{1}{a_n} = \frac{1}{\lim\limits_{n\to\infty} a_n}. \tag{4.46}$$

Beispiel

Es gilt

$$\lim_{n\to\infty}\frac{3\,n^2}{n^2+2\,n} = 3 \tag{4.47}$$

(Video 425, Notizen dazu).

Beachte, dass die obigen Rechenregeln nur für konvergente Folgen gelten! Bei divergenten Folgen, insbesondere auch bei bestimmt divergenten Folgen, ist nicht immer klar, wie sich Summen und Produkte verhalten (vgl. Übung 3).

4.3.5.2 Numerisches Rechnen

Das Verhalten einer Zahlenfolge im Unendlichen kann durch (numerisches) Berechnen später Elemente der Folge und ggf. deren grafischer Darstellung erhellt werden. Allerdings sind die daraus gewonnenen Informationen über Konvergenz/Divergenz/Grenzwert als reine Vermutung anzusehen. Prinzipiell ist denkbar, dass eine Folge beispielsweise in den ersten 10 Milliarden Elementen konstant ist und erst dann divergiert. Es kann also nie genügen, nur endlich viele Elemente zu betrachten.

4.3.5.3 Symbolisches Rechnen

Das Python-Paket SymPy für symbolisches Rechnen bietet auch das Berechnen von Grenzwerten von Zahlenfolgen:

```
import sympy

n = sympy.symbols('n')

sympy.limit(3 * n ** 2 / (n ** 2 + 2), n, sympy.oo)
```

```
3
```

```
n = sympy.symbols('n')

sympy.limit((1 + 2 / n) ** n, n, sympy.oo)
```

```
exp(2)
```

4.3.6 Zahlenreihen

Eine spezielle Form von Zahlenfolgen sind sogenannte Zahlenreihen, oft auch nur **Reihen** genannt. Dies sind Zahlenfolgen, bei denen jedes Folgenelement eine Summe ist, wobei von Element zu Element ein Summand dazukommt. Ist $(a_n)_{n\in\mathbb{N}}$ eine Zahlenfolge, so kann man eine Zahlenreihe $(b_n)_{n\in\mathbb{N}}$ als

$$b_n := \sum_{k=1}^{n} a_k = a_1 + a_2 + \cdots + a_n \tag{4.48}$$

bilden.

Die Frage nach der Konvergenz einer solchen Reihe ist also die Frage nach der Endlichkeit der aus unendlich vielen Summanden bestehenden Summe. Man verwendet für den Grenzwert einer Reihe das Symbol

$$\sum_{n=1}^{\infty} a_n := \lim_{n\to\infty} b_n. \tag{4.49}$$

Die Zahlen b_n werden als **Partialsummen** bezeichnet.

Da wir Zahlenreihen nur am Rande bei der Wahrscheinlichkeitsrechnung antreffen werden, vertiefen wir das Thema hier nicht weiter.

4.4 Funktionen

Der Begriff der Funktion wurde bereits eingeführt (vgl. Section 4.2.4).

4.4.1 Sprachliche Aspekte

Funktionen sind ein fundamentales sprachliches Mittel in der Mathematik, welches vielfältige Ausdrucksmöglichkeiten bietet. Um in dieser Vielfalt nicht die Orientierung zu verlieren, ist präzises Formulieren wichtig!

Beispiel

Die Funktion

$$f : \mathbb{R}^4 \to \mathbb{R}, \qquad f(x_1, y_1, x_2, y_2) := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{4.50}$$

ordnet den Koordinaten x_1, y_1, x_2, y_2 der Punkte (x_1, y_1) und (x_2, y_2) deren Abstand in der Ebene zu.

Inhaltlich identisch, aber formal verschieden, ist die Funktion

$$g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, \qquad g\big((x_1, y_1), (x_2, y_2)\big) := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, \tag{4.51}$$

die den zwei Punkten direkt den Abstand zuordnet ohne Umweg über die Koordinaten.

Sind P_1 und P_2 zwei Punkte in der Ebene, so ist $g(P_1, P_2)$ ein korrekter mathematischer Ausdruck. Der Ausdruck $f(P_1, P_2)$ ist hingegen nicht definiert, da f nur für vier Argumente definiert wurde. Der Ausdruck $f(a)$ für $a \in \mathbb{R}^4$ ist wiederum korrekt; $g(a)$ hingegen nicht.

Funktionen (und allgemeinere Abbildungen) sind ein sprachliches Konstrukt, das auch von Computern verstanden wird. Dort sind Funktionen fast so elementar wichtig wie in der Mathematik. Das obige Beispiel kann am Computer wie folgt umgesetzt werden (beachte, dass man statt f und g aussagekräftigere Bezeichnungen wählen sollte):

```
# Definition einer Funktion (entspricht f im Beispiel)
def abstand_koord(x1, y1, x2, y2):

    return ((x1 - x2) ** 2 + (y1 - y2) ** 2) ** 0.5

# Definition einer weiteren Funktion (entspricht g im Beispiel)
def abstand_punkte(punkt1, punkt2):

    x1, y1 = punkt1
    x2, y2 = punkt2
```

```
    return ((x1 - x2) ** 2 + (y1 - y2) ** 2) ** 0.5

# Aufruf bzw. Auswertung der Funktionen
print(abstand_koord(1, 1, 3, 2))
print(abstand_punkte((1, 1), (3, 2)))

# alternativ
punkt1 = (1, 1)
punkt2 = (3, 2)
print(abstand_punkte(punkt1, punkt2))
print(abstand_koord(punkt1, punkt2))  # Zeile führt zu Fehlermeldung!
```

```
2.23606797749979
2.23606797749979
2.23606797749979
```

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
↪ - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
↪ - - - - -
TypeError                                 Traceback (most recent call
↪ last)
Cell In[1], line 21
     17 # alternativ
     18 punkt1 = (1, 1)
     19 punkt2 = (3, 2)
     20 print(abstand_punkte(punkt1, punkt2))
 - - -> 21 print(abstand_koord(punkt1, punkt2))  # Zeile führt zu
  ↪ Fehlermeldung!

TypeError: abstand_koord() missing 2 required positional arguments:
↪ 'x2' and 'y2'
```

Beachte, dass `punkt1` und `punkt2` in der Definition der Funktion `abstand_punkte` **lokale Variablen** sind, im alternativen Aufruf weiter unten dann aber **globale Variablen**. Auch wenn diese Unterscheidung nicht zum Kernbereich dieser Veranstaltung gehört, sondern den Programmiergrundlagen zuzuordnen ist, sei zum besseren Verständnis die (kurze) Lektüre von Local versus Global Names empfohlen.

4.4.2 Visualisierung

Funktionen werden gern grafisch dargestellt, da viele Eigenschaften (die wir später näher untersuchen werden) leicht aus grafischen Darstellungen abgelesen werden können. Unter Nutzung des Computers sind auch komplexere, insbesondere dreidimensionale Vi-

sualierungen möglich. Visualisiert wird der Graph der Funktion, also eine Menge von Punkten in einem Koordinatensystem.

Für zweidimensionale Darstellungen eignet sich das Python-Paket Matplotlib sehr gut. Für dreidimensionale Darstellungen gibt es eine Erweiterung für Matplotlib (mpl_toolkits.mplot3d), jedoch bietet Plotly mehr Möglichkeiten.

Beispiel (einfacher 2D-Plot)

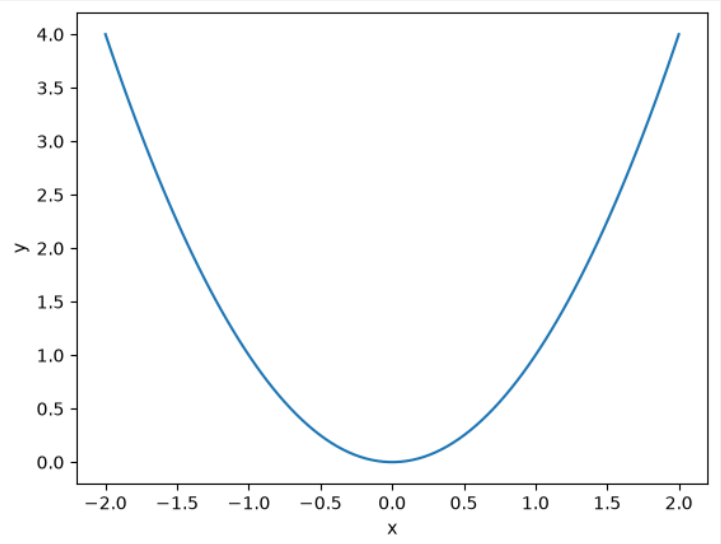
Funktionen, die einer reellen Zahl wieder eine reelle Zahl zuordnen, können in einem einfachen zweidimensionalen Plot dargestellt werden. Visualisieren die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = x^2. \tag{4.52}$$

```
import numpy as np
import matplotlib.pyplot as plt

x = np.linspace( -2, 2, 100)  # Intervall und Stützstellenanzahl
↪  für x
y = x ** 2  # Funktionswerte an den Stützstellen

fig, ax = plt.subplots()  # Zeichenfläche erstellen
ax.plot(x, y)  # Linie für Funktionsverlauf hinzufügen
ax.set_xlabel('x')  # Beschriftung x -Achse
ax.set_ylabel('y')  # Beschriftung y -Achse
plt.show()  # Zeichenfläche anzeigen
```



Beachte, dass die Darstellung am Computer praktisch immer durch Auswerten der Funk-

tion an endlich vielen Stützstellen (d.h. Werten für x) erfolgt. Die entstehenden Punkte werden dann durch Geradenstücke verbunden. Ist die Stützstellenanzahl zu gering gewählt, kann die grafische Darstellung stark vom tatsächlichen Funktionsverlauf abweichen.

Beispiel (Kurve in der Ebene)

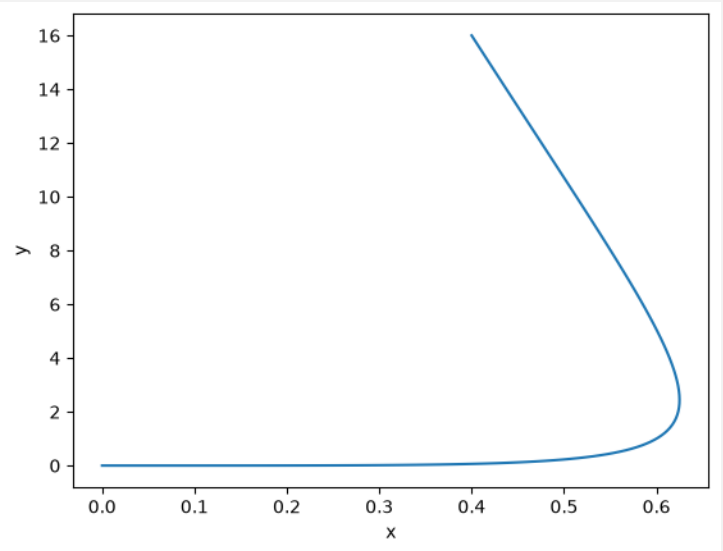
Die Position eines beweglichen Objekts in der Ebene kann durch eine Funktion der Form $t \mapsto (x(t), y(t))$ beschrieben werden. Dabei sind x und y Funktionen von $\mathbb{R}$ nach $\mathbb{R}$. Jedem Zeitpunkt t wird also eine Position $(x(t), y(t))$ zugeordnet. Funktionen dieser Form nennt man auch **Kurven in der Ebene**. Betrachten hier konkret

$$x(t) := t - 0.4\,t^2, \quad y(t) = t^4, \quad t \in [0, 2]. \tag{4.53}$$

Zwei Varianten für die Visualisierung von Kurven sind gebräuchlich: die Darstellung in der Ebene (ohne Zeitbezug) und die Darstellung im Raum (mit Zeitbezug).

```
t = np.linspace(0, 2, 100)
x = t - 0.4 * t ** 2
y = t ** 4

fig, ax = plt.subplots()
ax.plot(x, y)
ax.set_xlabel('x')
ax.set_ylabel('y')
plt.show()
```




```
import plotly.graph_objects as go
```

```
t = np.linspace(0, 2, 100)
x = t - 0.4 * t ** 2
y = t ** 4

fig = go.Figure()
fig.add_trace(
    go.Scatter3d(x=t, y=x, z=y, mode='lines', line_width=5)
)
fig.update_layout(
    scene_xaxis_title_text='t',
    scene_yaxis_title_text='x',
    scene_zaxis_title_text='y',
    margin=dict(l=0, r=0, b=0, t=0)  # breiten Rand vermeiden
)
fig.show()
```

Beispiel (Kurven im Raum)

Die Position eines beweglichen Objekts im Raum kann durch eine Funktion der Form $t \mapsto (x(t), y(t), z(t))$ beschrieben werden. Dabei sind x, y und z Funktionen von $\mathbb{R}$ nach $\mathbb{R}$. Jedem Zeitpunkt t wird also eine Position $(x(t), y(t), z(t))$ zugeordnet. Funktionen dieser Form nennt man auch **Kurven im Raum**. Betrachten hier konkret

$$x(t) := t - 0.4\,t^2, \quad y(t) = t^4, \quad z(t) = t^2 - 0.5\,t^3, \quad t \in [0, 2]. \tag{4.54}$$

Die Darstellung kann als 3D-Plot erfolgen, wobei der Zeitbezug verloren geht (analog zum 2D-Plot bei Kurven in der Ebene).

```
t = np.linspace(0, 2, 100)
x = t - 0.4 * t ** 2
y = t ** 4
z = t ** 2 - 0.5 * t ** 3

fig = go.Figure()
fig.add_trace(
    go.Scatter3d(x=x, y=y, z=z, mode='lines', line_width=5)
)
fig.update_layout(
    scene_xaxis_title_text='x',
    scene_yaxis_title_text='y',
    scene_zaxis_title_text='z',
```

```
    margin=dict(l=0, r=0, b=0, t=0)  # breiten Rand vermeiden
)
fig.show()
```

Beispiel (Funktionen von zwei Veränderlichen)

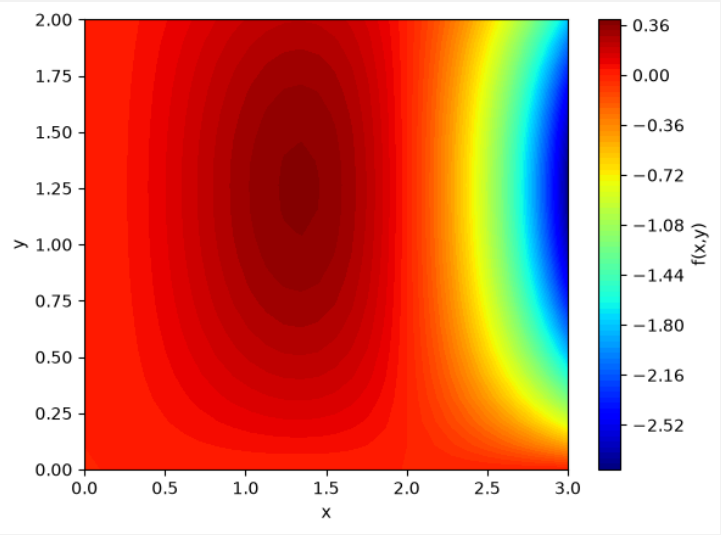
Funktionen, die jedem Punkt in der Ebene eine Zahl zuordnen, können wahlweise als farbiger 2D-Plot oder als 3D-Plot dargestellt werden. Visualisieren hier als Beispiel

$$f : [0,3] \times [0.2] \to \mathbb{R}, \quad f(x,y) = (x^2 - 0.5\,x^3)\,(y - 0.4\,y^2). \tag{4.55}$$

Bei der zweidimensionalen Darstellung werden die verschiedenen Funktionswerte durch Farben symbolisiert.

```
x, y = np.meshgrid(
    np.linspace(0, 3, 30),  # Stützstellenanzahl nicht zu hoch
    ↪  (Rechenzeit!)
    np.linspace(0, 2, 20)
)
# x und y sind 2D -Arrays, die für jeden Gitterpunkt die x - bzw.
↪  die
# y -Koordinate enthalten
f = (x ** 2 - 0.5 * x ** 3) * (y - 0.4 * y ** 2)

fig, ax = plt.subplots()
cont_plot = ax.contourf(
    x, y, f,
    levels=100,  # Anzahl der Farbabstufungen
    cmap='jet'  # Auswahl des Farbskala
)
plt.colorbar(cont_plot, label='f(x,y)')  # Zuordnung der Farben
↪  anzeigen
ax.set_xlabel('x')
ax.set_ylabel('y')
plt.show()
```




```
x, y = np.meshgrid(
    np.linspace(0, 3, 30),
    np.linspace(0, 2, 20)
)
f = (x ** 2 - 0.5 * x ** 3) * (y - 0.4 * y ** 2)

fig = go.Figure()
fig.add_trace(
    go.Surface(x=x, y=y, z=f, colorscale='Jet', showscale=False)
)
fig.update_layout(
    scene_xaxis_title_text='x',
    scene_yaxis_title_text='y',
    scene_zaxis_title_text='f(x,y)',
    margin=dict(l=0, r=0, b=0, t=0)  # breiten Rand vermeiden
)
fig.show()
```

Bei der Visualisierung von Funktionen zweier Veränderlicher ist zu beachten, dass die Anzahl der Stützstellen nicht zu hoch gewählt wird. Wählt man beispielsweise in beiden Variablen je 100 Stützstellen, so entsteht ein zweidimensionales Gitter aus 10000 Punkten. Die Funktion wird also für die Visualisierung an diesen 10000 Stellen ausgewertet. Insbesondere die 3D-Darstellung von über 10000 Einzelobjekten, aus denen die Grafik im Webbrowser dann zusammengesetzt wird, kann den Webbrowser an seine Leistungsgrenzen bringen.

Die Möglichkeiten der Visualisierung sind nahezu grenzenlos. Wichtig ist, dass beim Betrachten einer Visualisierung einer Funktion stets geklärt wird, **was eigentlich dar-**

gestellt ist! Dass die grafische Darstellung nicht nur auf zahlenwertige Funktionen beschränkt ist, zeigt das nächste Beispiel.

Beispiel (Vektorfelder)

Strömungen in der Ebene können als Funktionen beschrieben werden, die jedem Punkt in der Ebene einen Vektor zuordnen. Die Richtung des Vektors gibt dabei die Richtung der Strömung an; die Länge des Vektors beschreibt die Geschwindigkeit der Strömung in dem Punkt. Die Visualisierung solcher Strömungen kann in einem 2D-Plot erfolgen, indem an einigen Punkten der Ebene die Strömungsvektoren eingezeichnet werden. Dabei wird zugunsten der besseren Übersichtlichtkeit die Länge der Vektoren so skaliert, dass sie nicht überlappen. Dargestellt ist also nicht die tatsächliche Länge der Strömungsvektoren, sondern nur die relative Länge zueinander. Als Beispiel visualisieren wir die durch

$$f(x,y) = \begin{bmatrix} 1 \\ x\,y \end{bmatrix}, \quad x \in [-3,3], \quad y \in [-2,2] \tag{4.56}$$

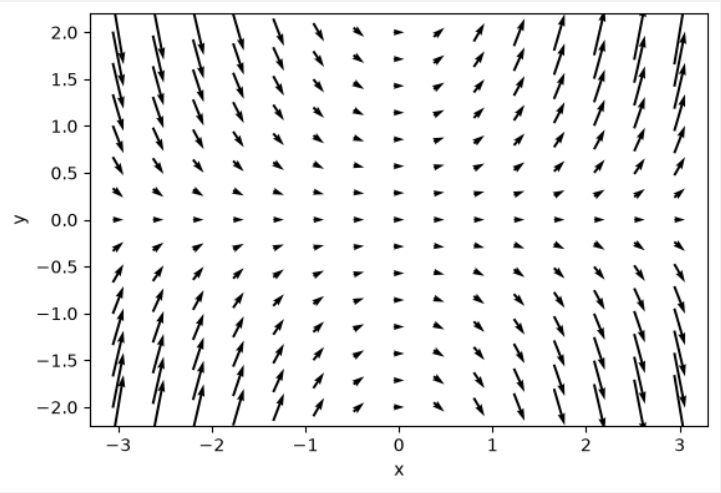
beschriebene Strömung. Der Umgang mit Vektoren wird später in der Veranstaltung genauer thematisiert.

```
# Gitterpunkte, an denen Strömungsvektoren dargestellt werden
↪ sollen
x, y = np.meshgrid(
    np.linspace( -3, 3, 15),
    np.linspace( -2, 2, 15)
)

# Strömungsvektoren
u = np.ones_like(x)  # Array mit Einsen in der Form wie das Array
↪ x
v = x * y

# Plot
fig, ax = plt.subplots()
ax.quiver(
    x, y, u, v,
    pivot='middle',  # Pfeilmitte an Gitterpunkt (nicht
    ↪ Pfeilanfang)
    angles='xy'  # Richtung der Pfeile folgt Verzerrung der Grafik
)
ax.set_aspect('equal', 'box')  # gleiche Einheiten an beiden
↪ Achsen
ax.set_xlabel('x')
```

```
ax.set_ylabel('y')
plt.show()
```



Achtung!

Da bei der Visualisierung mit dem Computer die zu visualisierende Funktion stets nur an endlich vielen Stützstellen ausgewertet wird, können in ungünstigen Fällen wichtige Eigenschaften der Funktion in der Visualisierung fehlen. Auch muss insbesondere bei auf ganz $\mathbb{R}$ oder ganz $\mathbb{R}^2$ definierten Funktionen im Vorfeld geklärt werden, welcher Teil des Definitionsbereichs relevant ist.

Beispiel (ungünstige Visualisierung)

Die Funktion

$$f : \mathbb{R} \setminus \{1000\} \to \mathbb{R}, \quad f(x) = \frac{1}{(x - 1000)^{10}} \tag{4.57}$$

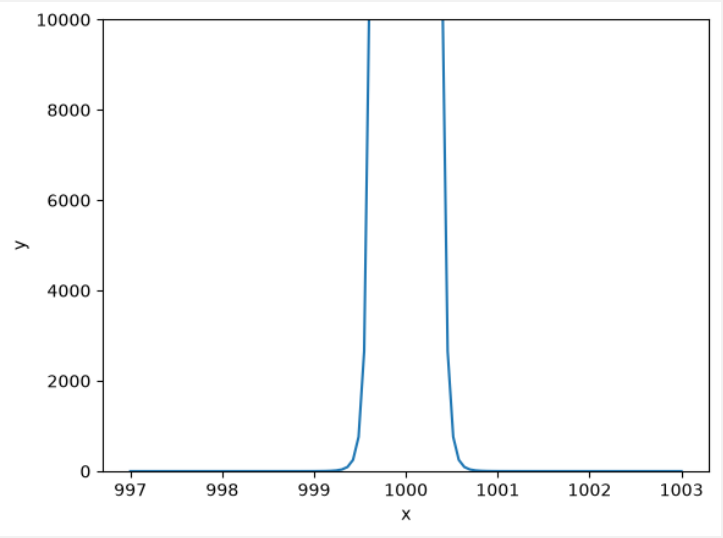
hat eine Polstelle bei $x = 1000$. Je nach Visualisierung ist diese sichtbar oder auch nicht.

Brauchbare Visualisierung:

```
x = np.linspace(997, 1003, 100)
y = 1 / (x - 1000) ** 10

fig, ax = plt.subplots()
ax.plot(x, y)
ax.set_ylim(0, 10000)  # y -Achse oben abschneiden
```

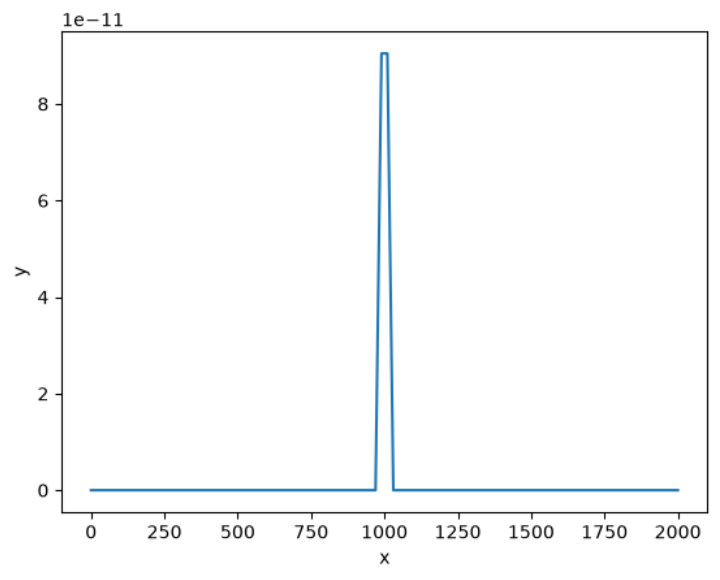
```
ax.set_xlabel('x')
ax.set_ylabel('y')
plt.show()
```



Unbrauchbare Visualisierung (Polstelle nicht als solche erkennbar; beachte, dass die dargestellten y-Werte alle praktisch Null sind):

```
x = np.linspace(0, 2000, 100)
y = 1 / (x - 1000) ** 10

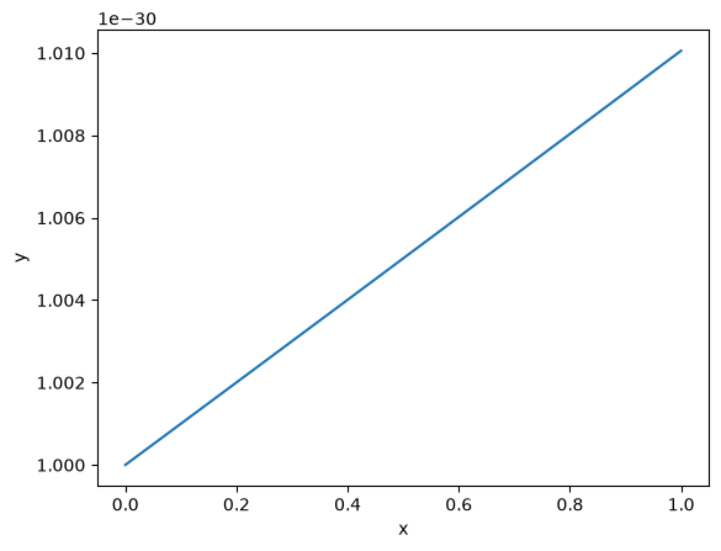
fig, ax = plt.subplots()
ax.plot(x, y)
ax.set_xlabel('x')
ax.set_ylabel('y')
plt.show()
```



Unbrauchbare Visualisierung (Polstelle nicht im dargestellten Bereich):

```
x = np.linspace(0, 1, 100)
y = 1 / (x - 1000) ** 10

fig, ax = plt.subplots()
ax.plot(x, y)
ax.set_xlabel('x')
ax.set_ylabel('y')
plt.show()
```



4.4.3 Einfache Eigenschaften

Betrachten einige Eigenschaften von Funktionen, deren Vorhandensein gelegentlich vorteilhaft ausgenutzt werden kann.

Nebenbemerkung

Vor dem Computerzeitalter mussten Funktionen aufwendig analytisch untersucht werden um eine Vorstellung vom Aussehen des Funktionsgraphs zu erhalten (Nullstellen, Extremstellen, Wendestellen, Monotonie, Krümmung, Polstellen, Asymptoten, Periodiziät,...). Die Visualisierung mit dem Computer spart hier viel Mühe. Entsprechend werden wir im Rahmen dieser Veranstaltung keine klassischen "Kurvendiskussionen" durchführen.

4.4.3.1 Monotonie und Umkehrfunktion

Merke

Eine Funktion $f : D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}$ heißt auf einer Teilmenge $A \subseteq D$

- **monoton wachsend**, wenn $f(x_1) \leq f(x_2)$ für alle $x_1, x_2 \in A$ mit $x_1 < x_2$ gilt,
- **streng monoton wachsend**, wenn $f(x_1) < f(x_2)$ für alle $x_1, x_2 \in A$ mit $x_1 < x_2$ gilt,

- **monoton fallend**, wenn $f(x_1) \geq f(x_2)$ für alle $x_1, x_2 \in A$ mit $x_1 < x_2$ gilt,
- **streng monoton fallend**, wenn $f(x_1) > f(x_2)$ für alle $x_1, x_2 \in A$ mit $x_1 < x_2$ gilt.

Die Monotonie von Funktionen spielt vor allem bei zwei Tätigkeiten eine Rolle: dem Lösen von Gleichungen und dem Umformen von Ungleichungen.

Soll eine Gleichung der Form

$$f(x) = y \tag{4.58}$$

für gegebenes $y \in \mathbb{R}$ gelöst werden und weiß man, dass die Funktion im Suchbereich für x (das kann z.B. ein Intervall oder ganz $\mathbb{R}$ sein) streng monoton ist, so kann man die **Umkehrfunktion** f^{-1} bilden und damit die Lösung direkt berechnen:

$$x = f^{-1}(y). \tag{4.59}$$

Für nicht streng monotone Funktionen ist die Sache schwieriger, da dann mehrere Lösungen existieren können, die nicht über die (dann nicht existierende) Umkehrfunktion berechnet werden können. Dieser Aspekt ist sowohl beim manuellen Rechnen als auch beim Lösen von Gleichungen mit dem Computer relevant.

Beispiel

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := x^2 \tag{4.60}$$

ist nicht monoton und es existiert auch keine Umkehrfunktion. Betrachten wir die gleiche Abbildungsvorschrift jedoch nur auf einer geeigneten Teilmenge der reellen Zahlen, so ist die Funktion streng monoton und besitzt demnach auch eine Umkehrfunktion (Video 440, Notizen dazu).

Gelegentlich möchte/muss man Ungleichungen umformen. Dabei ist zu beachten, dass manche Umformschritte, die bei Gleichungen zu einer äquivalenten Gleichung führen, bei Ungleichungen nur mit Vorsicht zu nutzen sind. Beispielsweise muss bei der Multiplikation beider Seiten der Ungleichung mit einer negativen Zahl das Relationszeichen "gedreht" werden (aus $\leq$ wird $\geq$ und umgekehrt; analog für strenge Ungleichheit). Allgemein gilt:

Merke

Das Anwenden einer streng monoton wachsenden Funktion auf beide Seiten einer Ungleichung ist eine äquivalente Umformung.

Das Anwenden einer Streng monoton fallenden Funktion auf beide Seiten einer Umgleichung ist eine äquivalente Umformung, wenn dabei das Relationszeichen "gedreht" wird.

Beispiel

Wir möchten die Ungleichung

$$\frac{-2}{x} \leq 4 \tag{4.61}$$

nach x umstellen, d.h. durch äquivalente Umformungen auf die Form $x \leq \ldots$ oder $x \geq \ldots$ bringen. Alle Umformschritte können wir als Anwenden geeigneter Funktionen auf beide Seiten der Ungleichung formulieren:

1. Wir wenden die Funktion $u \mapsto -\frac{1}{2}\,u$ auf beide Seiten an. Diese Funktion ist streng monoton fallend, also muss das Relationszeichen gedreht werden:

$$\frac{1}{x} \geq -\frac{1}{2} \cdot 4. \tag{4.62}$$

2. Bilden nun auf beiden Seiten das Reziproke, d.h. wir wenden die Funktion $u \mapsto \frac{1}{u}$ auf beide Seiten an. Dies ist ebenfalls eine streng monoton fallende Funktion. Also muss das Relationszeichen nochmal gedreht werden:

$$x \leq \frac{1}{-\frac{1}{2} \cdot 4}. \tag{4.63}$$

Erhalten also die äquivalent umgeformte Ungleichung

$$x \leq -\frac{1}{2}. \tag{4.64}$$

Für weitere Beispiele und Erläuterungen zur Monotonie von Funktionen sei auf das entsprechende OMB+-Kapitel verwiesen.

4.4.3.2 Periodizität

Merke

Eine Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt **periodisch**, wenn es eine Zahl $T > 0$ gibt, sodass

$$f(x + T) = f(x) \quad \text{für alle } x \in \mathbb{R} \tag{4.65}$$

gilt. Jede Zahl T mit dieser Eigenschaft heißt **Periode** von f. Die kleinste Zahl T mit dieser Eigenschaft heißt **Grundperiode** von f.

Ist T eine Periode einer Funktion, so genügt die Betrachtung der Funktion auf dem Intervall $[0, T)$. Auf dem Rest der reellen Achse ist das Verhalten der Funktion vollständig durch ihr Verhalten auf diesem Intervall bestimmt.

Wichtige periodische Funktionen sind die Sinus- und die Cosinus-Funktion (siehe unten)

sowie darauf basierende komplexere Funktionen.

4.4.3.3 Beschränktheit

Merke

Eine Funktion $f : D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ heißt beschränkt, wenn es eine Zahl $c > 0$ gibt, sodass

$$|f(x)| \leq c \quad \text{für alle } x \in D \tag{4.66}$$

gilt.

4.4.3.4 Nullstellen

Merke

Für eine Funktion $f : \mathbb{R}^n \to \mathbb{R}^m$ heißt $x \in \mathbb{R}^m$ **Nullstelle** der Funktion, wenn $f(x) = 0$ gilt.

Nebenbemerkung

Beachte, dass wir in dieser Veranstaltung keine besondere Symbolik für die Unterscheidung von Zahlen und Punkten bzw. Vektoren verwenden. In der Ingenieurliteratur sieht man gelegentlich Unterstreichungen $\underline{x}$ bzw. $\underline{0}$ oder aufgesetzte Pfeile $\overrightarrow{x}$ bzw. $\overrightarrow{0}$ oder Fettsetzungen $\mathbf{x}$ bzw. $\mathbf{0}$ für Vektoren. Diese führen jedoch an einigen Stellen zu Komplikationen und machen komplexe Formeln unübersichtlicher. Da ohnehin beim Betrachten eines mathematischen Ausdrucks stets im Detail zu klären ist, welches Symbol welche Bedeutung hat, ergibt sich die Antwort auf die Frage "Zahl oder Punkt/Vektor?" gleich mit.

In der obigen Definition des Begriffs Nullstelle steht die 0 also für den Nullvektor (sofern $m > 1$) und auch x ist ein Punkt/Vektor, wenn $n > 1$.

Die Frage nach den Nullstellen einer Funktion ist eng verknüpft mit dem Lösen von Gleichungen

$$f(x) = y, \tag{4.67}$$

wobei $y \in \mathbb{R}^m$ gegeben und $x \in \mathbb{R}^n$ gesucht ist. Solche Gleichungen kann man als

$$g(x) = 0 \tag{4.68}$$

schreiben, wenn man $g(x) := f(x) - y$ setzt. Beim Lösen von Gleichungen mit dem Computer stehen oft nur Funktionen für die Nullstellenbestimmung zur Verfügung, sodass Gleichungen, wie eben gezeigt, entsprechend umgeformt werden müssen.

Nullstellen von Funktionen können beispielsweise mit `scipy.optimize.fsolve` aus dem Python-Paket SciPy numerisch berechnet werden. Beachte, dass stets nur eine Nullstelle geliefert wird, auch wenn es mehrere gibt. Welche Nullstelle man erhält, kann man durch Vorgabe einer groben Schätzung der Nullstelle steuern.

Beispiel

Zum Lösen der Gleichung

$$x^2 + x + 1 = \frac{3}{2} \tag{4.69}$$

können wir folgenden Python-Code verwenden:

```
import scipy.optimize as so

def f(x):
    return x ** 2 + x + 1 - 3 / 2

print(so.fsolve(f, 1))
print(so.fsolve(f, -1))
```

```
[0.3660254]
[ -1.3660254]
```

Die Anzahl der Lösungen und geeignete Näherungswerte (hier 1 und -1) sind nicht immer klar ersichtlich. Deshalb birgt das numerische Lösen von Gleichungen stets das Risiko, dass man nicht alle Lösungen findet!

Beachte, dass `fsolve` auch zum Lösen von (nichtlinearen) Gleichungssystemen verwendet werden kann. Dies entspricht dem Fall $m > 1$ und $n > 1$ in der obigen Definition des Begriffs Nullstelle.

Das symbolische Lösen von Gleichungen ist mit SymPy möglich. Für das obige Beispiel:

```
import sympy

x = sympy.symbols('x')

sympy.solve(x ** 2 + x + 1 - sympy.Rational(3, 2), x)
```

```
[ -1/2 + sqrt(3)/2, -sqrt(3)/2 - 1/2]
```

Beim symbolischen Rechnen erhält man üblicherweise alle Lösungen. Jedoch werden nicht immer Lösungen gefunden, obwohl sie existieren. Beim numerischen Rechnen hingegen erhält man praktisch immer wenigstens eine Lösung (sofern eine existiert).

4.4.4 Grenzwerte

Die Betrachtung von Grenzwerten bei Funktionen spielt eine zentrale Rolle in der Mathematik und ihren Anwendungen. So basiert z.B. der in Physik und anderen (Natur-)Wissenschaften ganz grundlegende Begriff der Differenzierbarkeit einer Funktion auf der Betrachtung gewisser Grenzwerte.

Grenzwerte bei Funktionen sind sehr ähnlich zu Grenzwerten bei Zahlenfolgen, jedoch gibt es unterschiede im Detail, denn der kontinuierliche Charakter der reellen Zahlen erlaubt auch andere Grenzübergänge als nur "$\to \infty$".

Merke (Grenzwerte im Unendlichen)

Sei $f : D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}$ eine Funktion, sodass D ein Intervall der Form $[a, \infty)$ enthält. Eine Zahl $g \in \mathbb{R}$ heißt **Grenzwert** der Funktion f bei Annäherung des Arguments an ∞, wenn für jede (!) Zahlenfolge $(x_n)_{n\in\mathbb{N}}$ in D mit $\lim_{n\to\infty} x_n = \infty$ die zugehörige Folge $(f(x_n))_{n\in\mathbb{N}}$ der Funktionswerte gegen g konvergiert. Man schreibt dann

$$\lim_{x\to\infty} f(x) = g. \tag{4.70}$$

Analog ist der Grenzwert bei Annäherung an $-\infty$ erklärt, wenn D ein Intervall der Form $(-\infty, b]$ enthält.

Nicht jede Funktion hat Grenzwerte im Unendlichen. Beispielsweise existieren die beiden Grenzwerte nicht für (nicht konstante) periodische Funktionen. Auch können die Grenzwerte ∞ oder $-\infty$ sein; in diesem Fall sagt man ebenfalls, dass die Grenzwerte nicht (als Zahl) existieren.

Merke (Grenzwerte an einer Stelle)

Sei $f : D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}$ eine Funktion. Eine Zahl $g \in \mathbb{R}$ heißt **Grenzwert** der Funktion f bei Annäherung an die Stelle $x_0 \in \mathbb{R}$, wenn für jede Zahlenfolge $(x_n)_{n\in\mathbb{N}}$ in D mit $\lim_{n\to\infty} x_n = x_0$ die zugehörige Folge $(f(x_n))_{n\in\mathbb{N}}$ der Funktionswerte gegen g konvergiert. Man schreibt dann

$$\lim_{x\to x_0} f(x) = g. \tag{4.71}$$

Betrachtet man nur Folgen $(x_n)_{n\in\mathbb{N}}$ mit $x_n \le x_0$ für alle $n \in \mathbb{N}$, so spricht man vom **linksseitigen Grenzwert** und schreibt

$$\lim_{x\to x_0-0} f(x) = g. \tag{4.72}$$

Betrachtet man nur Folgen $(x_n)_{n\in\mathbb{N}}$ mit $x_n \ge x_0$ für alle $n \in \mathbb{N}$, so spricht man vom **rechtsseitigen Grenzwert** und schreibt

$$\lim_{x\to x_0+0} f(x) = g. \tag{4.73}$$

Man kann zeigen, dass der (beidseitige) Grenzwert genau dann existiert, wenn links- und rechtsseitiger Grenzwert existieren und übereinstimmen.

Ein relativ häufiges Phänomen sind Funktionen mit Polstellen. Für eine Funktion $f : [a, b] \setminus \{u\} \to \mathbb{R}$ heißt u **Polstelle**, wenn die einseitigen Grenzwerte bei Annäherung an diese Stelle nur im Sinne von $\pm\infty$ existieren. Da Funktionen in der Praxis fast immer Messgrößen (Temperatur, Länge, Druck, Kosten,...) beschreiben, erscheinen Polstellen eher unnatürlich und deren Auftreten entsprechend unwahrscheinlich. Sie sind jedoch ein wichtiges Modellierungswerkzeug!

Beispiel (Massepunkt)

In der Physik vereinfacht man räumlich ausgedehnte Objekte gern zu Punkten im Raum. So betrachtet man bei der Berechnung von Flugbahnen nur selten die Masseverteilung des sich bewegenden Objekts im Raum, sondern modelliert das Objekt als "Massepunkt", also als einen einzigen Punkt im Raum mit einer gegebenen Masse. Analog verfährt man beispielsweise mit Ladungen in der Elektrotechnik ("Punktladung").

Die Anziehungskraft, die ein als Massepunkt modellierter Himmelskörper (z.B. Erde) auf andere Objekte (z.B. Satelliten) ausübt, kann durch sein Gravitationsfeld beschrieben werden. Die Anziehungskraft ist immer zum Massepunkt hin gerichtet. Der Betrag der Kraft hängt von der Entfernung zum Massepunkt ab und kann somit als Funktion $F : (0, \infty) \to \mathbb{R}$ ausgedrückt werden:

$$F(r) = \frac{G\,M}{r^2}. \tag{4.74}$$

Dabei sind $G > 0$ die Gravitationskonstante und $M > 0$ die Masse des Himmelskörpers. Die Funktion F hat eine Polstelle bei $r = 0$.

Natürlich wirkt in der Nähe des Erdmittelpunktes (genauer: Massenschwerpunktes der Erde) keine beliebig hohe Gravitationskraft. Sondern die Polstelle ist ein Resultat der vereinfachten Modellierung der Erde als Massepunkt. In der Nähe des Erdmittelpunkts ist das gewählte Modell zur Beschreibung der wirkenden Kräfte ungeeignet (vgl. auch Newtonsches Kugelschalentheorem).

Der Begriff des Grenzwerts spielt auch bei Funktionen mehrerer Veränderlicher eine Rolle. Dann ist jedoch nur der (allseitige) Grenzwert bei Annäherung an eine Stelle $x_0 \in \mathbb{R}^n$ sinnvoll einführbar. Die dazu benötigte Konvergenz einer Punktfolge $(x_k)_{k\in\mathbb{N}}$ in $\mathbb{R}^n$ gegen x_0 ist gleichbedeutend mit der Konvergenz jeder einzelnen Koordinatenfolge (vgl. auch Aufgabe 3.6(c) in Übung 3). Schreiben wir jeden Punkt x_k als

$$x_k = (x_k^{(1)}, \ldots, x_k^{(n)}) \in \mathbb{R}^n, \tag{4.75}$$

so bedeutet $\lim_{k\to\infty} x_k = x_0$ also, dass

$$\lim_{k\to\infty} x_k^{(l)} = x_0^{(l)} \tag{4.76}$$

für jedes $l \in \{1, \ldots, n\}$ gilt.

Nebenbemerkung

Bei der Arbeit mit Folgen von Punkten (oder Vektoren) treten mehrere verschiedene Indizes und Dimensionszahlen auf. Beim Lesen und Verstehen entsprechender Formeln muss im Detail geklärt werden, welches Symbol welche Rolle übernimmt. Im hier betrachteten Fall sind

- n die Anzahl der Veränderlichen der Funktion (nicht der Folgenindex!),
- k der Folgenindex,
- l der Index der Koordinaten eines Punktes.

Insbesondere wird die Schreibweise mit zwei Indizes in der Literatur nicht einheitlich gehandhabt. In unserem Fall ist der untere Index in $x_k^{(l)}$ der Folgenindex und der obere Index ist die Koordinatennummer. Dies kann auch umgekehrt gehandhabt werden! Die Klammern um den oberen Index sollen Verwechselungen mit Potenzen vermeiden.

Merke (Grenzwerte bei mehreren Veränderlichen)

Sei $f : D \to \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ eine Funktion. Ein Punkt $g \in \mathbb{R}^m$ heißt **Grenzwert** der Funktion f bei Annäherung an die Stelle $x_0 \in \mathbb{R}^n$, wenn für jede Punktfolge $(x_n)_{n\in\mathbb{N}}$ in D mit $\lim_{n\to\infty} x_n = x_0$ die zugehörige Punktfolge $(f(x_n))_{n\in\mathbb{N}}$ der Funktionswerte gegen g konvergiert. Man schreibt dann

$$\lim_{x\to x_0} f(x) = g. \tag{4.77}$$

Umganssprachlich formuliert bedeutet dies, dass völlig unabhängig davon wie man sich der Stelle x_0 annähert, das Verhalten der Funktionswerte immer gleich ist. Hier ein Gegenbeispiel:

Beispiel

Für die Funktion

$$f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}, \quad f(x,y) := \frac{x\,y}{x^2+y^2} \tag{4.78}$$

existiert in $(0,0)$ kein Grenzwert (Video 450, Notizen dazu).

Beachte, dass die grafische Darstellung in der Nähe von $(0, 0)$ fehlerhaft ist! Ursache ist das Auswerten der Funktion an nur endlich vielen Stellen.

4.4.5 Stetigkeit

Merke

Eine Funktion $f : D \to \mathbb{R}$ mit $D \subseteq \mathbb{R}^n$ heißt **stetig im Punkt** $x_0 \in D$, wenn der Grenzwert $\lim_{x \to x_0} f(x)$ existiert und identisch mit dem Funktionswert $f(x_0)$ ist.

Eine Funktion $f = (f_1, \ldots, f_m) : D \to \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ heißt stetig im Punkt $x_0 \in D$, wenn jede Komponente f_l, $l = 1, \ldots, n$, stetig in x_0 ist.

Eine Funktion $f : D \to \mathbb{R}^m$ mit $D \subseteq \mathbb{R}^n$ heißt **stetig**, wenn sie in jedem Punkt ihres Definitionsbereichs D stetig ist.

Stetigkeit einer Funktion $f : \mathbb{R} \to \mathbb{R}$ bedeutet anschaulich, dass man den Graph der Funktion ohne Absetzen mit einem Stift zeichnen kann, der Graph also beispielsweise keine Sprünge aufweist. Unbedingt zu beachten ist, dass Stetigkeit nur Aussagen zum Verhalten der Funktion in ihrem Definitionsbereich macht. Der Definitionsbereich selbst kann aber beispielsweise Lücken haben.

Beispiel

Die Funktion

$$f : \mathbb{R} \setminus \{0\} \to \mathbb{R}, \quad f(x) := \frac{1}{x} \tag{4.79}$$

ist stetig, obwohl sie eine Polstelle bei $x = 0$ besitzt. Die Stelle $x = 0$ gehört nicht zum Definitionsbereich der Funktion und findet daher bei der Frage nach Stetigkeit keine Beachtung.

Hingegen ist die Funktion

$$g : \mathbb{R} \to \mathbb{R}, \quad f(x) := \begin{cases} \frac{1}{x} & \text{für } x \neq 0, \\ 0 & \text{für } x = 0 \end{cases} \tag{4.80}$$

nicht stetig, da sie in $x = 0$ nicht stetig ist (Grenzwert $\lim_{x \to 0} f(x)$ existiert nicht).

Unstetigkeitsstellen bei Funktionen einer reellen Veränderlichen sind im Wesentlichen entweder Polstellen (vgl. Funktion g im Beispiel) oder Sprungstellen (links- und rechtsseitige Grenzwerte existieren, stimmen aber nicht beide mit dem Funktionswert überein). Für Funktionen mehrerer Veränderlicher ist die Klassifikation von Unstetigkeitsstellen nicht mehr so einfach machbar.

Beispiel

Die Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) := \begin{cases} \frac{x\,y}{x^2+y^2} & \text{für } (x,y) \neq (0,0), \\ 0 & \text{für } (x,y) = (0,0) \end{cases} \tag{4.81}$$

ist nicht stetig in $(0,0)$, da dort der Grenzwert $\lim_{(x,y)\to(0,0)} f(x,y)$ nicht existiert (vgl. früheres Beispiel zu Grenzwerten).

Der Begriff der Stetigkeit ist innerhalb der Mathematik von enormer Bedeutung, wird aber bei der täglich Arbeit im Kontext von Anwendungsproblemen nur am Rande direkt benötigt. Die Wichtigkeit des Begriffs resultiert aus der Tatsache, dass bei stetigen Funktionen der Funktionswert an einer Stelle auch Informationen über die Funktionswerte an benachbarten Stellen liefert. Insbesondere mit Blick auf den Computereinsatz, bei dem Funktionen stets nur an endlich vielen Stellen ausgewertet werden können, ist Stetigkeit eine Art Minimalforderung, um eine Chance auf brauchbare Resultate zu haben. Beispiele für das Auftreten des Stetigkeitsbegriffs in anwendungsbezogenen Situationen:

- Viele mathematische Aussagen gelten nur, wenn die beteiligten Funktionen stetig sind. So ist Stetigkeit beispielsweise eine wichtige Voraussetzung (aber nicht die einzige!) für die Existenz von Lösungen zu Optimierungsproblemen. Bei der mathematischen Modellierung von praktischen Fragestellungen ist deshalb darauf zu achten, dass die in einem Modell auftretenden Funktionen möglichst stetig sind.
- Beim Rechnen mit Computer oder Taschenrechner oder auch beim manuellen Rechnen mit gemessenen und somit fehlerbehafteten Werten werden die Berechnungen stets mit gerundeten Werten durchgeführt. Das Ausführen von Berechnungen kann man als Auswerten einer geeigneten Funktion an einer Stelle interpretieren. Die Stetigkeit einer solchen Funktion garantiert, dass die Funktionswerte an den gerundeten Stellen nicht zu sehr von den Funktionswerten an den (unbekannten) exakten Stellen abweichen. Wie groß die Rundungsfehler sein dürfen, damit die Abweichungen in den Funktionswerten nicht zu groß werden, ist eine wichtige Frage. Dieser werden wir später nachgehen, wenn wir die geeigneten mathematischen Werkzeuge zur Beantwortung zur Verfügung haben.
- Fotografische Bilddaten werden heute praktisch immer in stark komprimierter Form gespeichert. Welche Kompressionsverfahren in Frage kommen und wie stark ein Bild ohne größere Qualitätsverluste komprimiert werden kann, hängt auch an den Stetigkeitseigenschaften des kontinuierlichen (d.h. nicht in Pixel zerlegten) Bildes. Für den heute üblichen JPEG-Standard zur Bilddatenkompression kann man grob formulieren: je “stetiger” das Bild, desto stärker kann ohne nennenswerte Qualitätsverluste komprimiert werden.

4.4.6 Beispiele elementarer Funktionen

Elementare Funktionen und ihre Eigenschaften wurden im Rahmen der Schulmathematik behandelt. Deshalb verweisen wir zur Wiederholung auf das entsprechenden Kapitel von OMB+ und gehen hier nur auf einzelne Beispiele und Aspekte nochmal genauer ein:

4.4.6.1 Polynome Funktionen $f : \mathbb{R} \to \mathbb{R}$ der Form

$$f(x) = a_n\, x^n + a_{n-1}\, x^{n-1} + \cdots + a_2\, x^2 + a_1\, x + a_0 \tag{4.82}$$

heißen **Polynom**. Dabei ist die gegebene Zahl $n \in \mathbb{N}$ der **Grad** des Polynoms und die gegebenen $n+1$ Zahlen $a_0, a_1, \ldots, a_n \in \mathbb{R}$ sind die **Koeffizienten** des Polynoms.

Jedes Polynom hat höchstens n Nullstellen. Jeder Polynomnullstelle kann eine sogenannte Vielfachheit zugeordnet werden: Die Polynomnullstelle x_0 hat die **Vielfachheit** $k \in \mathbb{N}$, wenn man das Polynom in der Form

$$f(x) = (x - x_0)^k \,(\text{Polynom vom Grad } n-k) \tag{4.83}$$

schreiben kann und k die größte Zahl ist, die eine solche Darstellung erlaubt. Der Faktor $x - x_0$ wird auch als **Linearfaktor** bezeichnet. Sind $x_1, \ldots, x_m$ die Nullstellen des Polynoms f und $k_1, \ldots, k_m$ die zugehörigen Vielfachheiten, so gilt

$$f(x) = (x - x_1)^{k_1} \cdots (x - x_m)^{k_m} \,(\text{Polynom ohne Nullstellen}). \tag{4.84}$$

Nebenbemerkung

Man kann den Zahlenraum, der sich bei uns auf die reellen Zahlen $\mathbb{R}$ beschränkt, noch erweitern zu den sogenannten komplexen Zahlen. In diesem erweiterten Zahlenraum hat jedes Polynom n-ten Grades genau n Nullstellen, wenn man jede Nullstelle entsprechend ihrer Vielfachheit zählt. Man erhält dann eine Zerlegung des Polynoms in Linearfaktoren mit einem Polynom 0-ten Grades (konstantes Polynom) als nullstellenfreies Restpolynom. Die Theorie der Polynome über komplexen Zahlen ist ein wichtiger Baustein innerhalb der Mathematik, der auch gelegentlich an der für Anwender sichtbaren Oberfläche erscheint.

4.4.6.2 Lineare Funktionen mehrerer Veränderlicher Funktionen $f : \mathbb{R}^n \to \mathbb{R}$ der Form

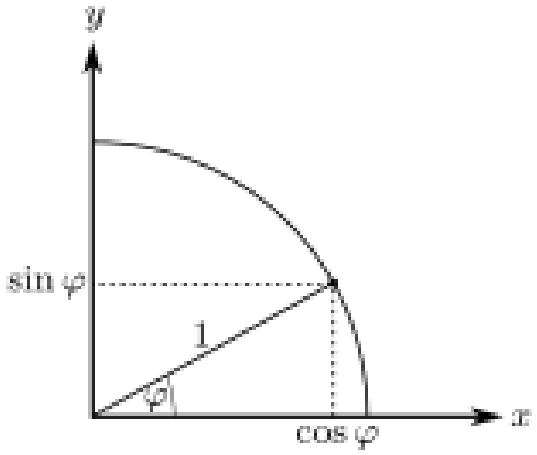
$$f(x_1, \ldots, x_n) = a_1\, x_1 + \cdots + a_n\, x_n \tag{4.85}$$

heißen **linear**. Dabei sind $a_1, \ldots, a_n \in \mathbb{R}$ gegebene Zahlen.

Für $n = 2$ ist der Graph einer linearen Funktion stets eine Ebene in $\mathbb{R}^3$. Für $n > 2$ bezeichnet man die Graphen linearer Funktionen als **Hyperebenen** (in $\mathbb{R}^{n+1}$).

Lineare Funktionen werden aufgrund ihrer einfachen Struktur gern als näherungsweiser Ersatz für kompliziertere Funktionen verwendet.

4.4.6.3 Trigonometrische Funktionen Häufig benötigte Zahlen, deren direkte Erklärung zu viel Zeit benötigt, bekommen in der Mathematik üblicherweise einen Name. So benötigt man zu einem gegebenen Winkel z.B. sehr oft die Zahlen, die wir als Cosinus und Sinus des Winkels kennen. Cosinus und Sinus eines Winkels sind gerade die x- bzw. y-Koordinate des Punktes, der entsteht indem man die Entfernung 1 vom Koordinatenursprung aus entlang eines Strahls mit Winkel φ zur (positiven) x-Achse zurücklegt.



Cosinus und Sinus geben die x- und die y-Koordinate eines Punktes auf dem Einheitskreis in Abhängigkeit des Winkels φ zwischen dem Strahl vom Ursprung durch den Punkt und der x-Achse an.

Beträgt die Entfernung des Punktes nicht 1, sondern allgemeiner $r > 0$, so ergeben sich die Koordinaten als

$$\begin{aligned} x &= r\,\cos\varphi, \\ y &= r\,\sin\varphi. \end{aligned} \tag{4.86}$$

Die Darstellung eines Punktes in der Ebene durch Abstand $r > 0$ zum Ursprung und Winkel $\varphi \in [0, 2\,\pi)$ zur x-Achse bezeichnet man als **Polarkoordinaten**-Darstellung des Punktes. Sie wird insbesondere in der Physik gern verwendet (Drehbewegungen und andere periodische Vorgänge), spielt aber beispielsweise auch in der Geodatenverarbeitung eine Rolle (diverse Kartenprojektionen, z.B. Azimuthalprojektionen).

Offensichtlich gilt stets

$$(\sin\varphi)^2 + (\cos\varphi)^2 = 1 \tag{4.87}$$

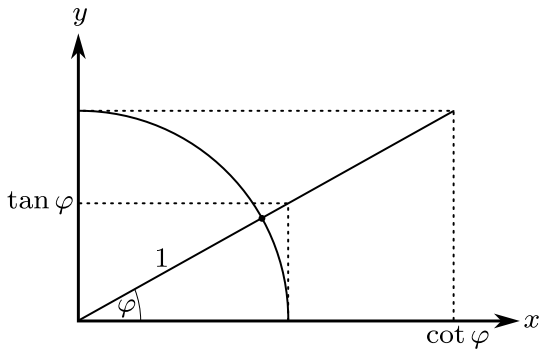
(trigonometische Form des Satzes des Pythagoras).

Auf Basis von Cosinus und Sinus definiert man noch den Tangens

$$\tan\varphi := \frac{\sin\varphi}{\cos\varphi} \tag{4.88}$$

und den Cotangens

$$\cot\varphi := \frac{\cos\varphi}{\sin\varphi}. \tag{4.89}$$



Tangens und Cotangens sind gerade der Anstieg des Strahls durch Ursprung und Punkt bzw. das Reziproke des Anstiegs.

Sinus und Cosinus können auch für Winkel außerhalb von $[0, 2\,\pi)$ bestimmt werden, indem der fragliche Winkel durch Verschiebung um Vielfache von $2\,\pi$ in den Bereich $[0, 2\,\pi)$ gebracht wird. Die so entstehende **Sinusfunktion** und die **Cosinusfunktion** sind auf ganz $\mathbb{R}$ definiert und haben beide den Wertebereich $[-1, 1]$.

Sinusfunktion und Cosinusfunktion sind nicht monoton auf $[0, 2\,\pi)$. Auf geeigneten Teilintervallen sind sie jedoch streng monoton, sodass dort eine Umkehrfunktion existiert. Mit arcsin bezeichnet man die Umkehrfunktion der Sinusfunktion auf dem Intervall $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Die Umkehrfunktion der Cosinusfunktion auf dem Intervall $[0, \pi]$ wird als arccos bezeichnet. Für den Tangens wiederum existiert eine Umkehrfunktion auf $[-\frac{\pi}{2}, \frac{\pi}{2}]$, die mit arctan bezeichnet wird. Eine Umkehrfunktion des Cotangens kann analog eingeführt werden, spielt aber nur sehr selten eine Rolle.

4.4.6.4 Exponentialfunktion

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := \mathrm{e}^x \tag{4.90}$$

mit der Euler'schen Zahl $\mathrm{e} \approx 2.718$ wird als **Exponentialfunktion** bezeichnet. Sie taucht (meist unerwartet) in ganz vielen Situationen in praktisch allen Teilbereichen der Mathematik auf. Wir werden ihr insbesondere bei der Differentialrechnung und bei der Wahrscheinlichkeitsrechnung begegnen.

Die Exponentialfunktion ist streng monoton wachsend und besitzt somit eine Umkehrfunktion. Diese wird als (natürliche) **Logarithmusfunktion** bezeichnet und mit dem Symbol ln abgekürzt. Es gilt also stets

$$\ln(\mathrm{e}^x) = x \quad \text{für } x \in \mathbb{R} \qquad \text{und} \qquad \mathrm{e}^{\ln x} = x \quad \text{für } x > 0. \tag{4.91}$$

Nebenbemerkung (Beziehung zu Sinus und Cosinus)

Die Exponentialfunktion kann als Polynom mit "unendlichem Grad" angesehen werden: Zu $x \in \mathbb{R}$ und $n \in \mathbb{N}$ können wir das Polynom

$$P_n(x) := \sum_{k=0}^{n} \frac{1}{k!}\, x^k \tag{4.92}$$

vom Grad n bilden. Dabei ist

$$n! := 1 \cdot 2 \cdots n \tag{4.93}$$

das Produkt der Zahlen 1 bis n (sprich: "n Fakultät"), wobei $0! := 1$ gesetzt wird, und $\sum$ ist ein Symbol zur kurzen Schreibweise für lange Summen:

$$\sum_{k=0}^{n} \frac{1}{k!}\, x^k := \frac{1}{0!}\, x^0 + \frac{1}{1!}\, x^1 + \cdots + \frac{1}{n!}\, x^n. \tag{4.94}$$

Man kann nun zeigen (tun wir hier nicht), dass

$$\lim_{n\to\infty} P_n(x) = \mathrm{e}^x \tag{4.95}$$

gilt, die Polynome für festes $x \in \mathbb{R}$ bei fortwährender Erhöhung des Grades also gegen die Exponentialfunktion konvergieren.

In diesem Kontext wird auch eine enge Beziehung zu Sinus- und Cosinusfunktion sichtbar. Entfernt man alle ungeraden x-Potenzen und wechselt das Vorzeichen vor den geraden Potenzen, so erhält man die Cosinusfunktion:

$$\lim_{n\to\infty} \sum_{k=0}^{n} (-1)^k \frac{1}{(2\,k)!}\, x^{2\,k} = \cos x. \tag{4.96}$$

Analog liefert das Weglassen der geraden Potenzen die Sinusfunktion:

$$\lim_{n\to\infty} \sum_{k=0}^{n} (-1)^k \frac{1}{(2\,k+1)!}\, x^{2\,k+1} = \sin x. \tag{4.97}$$

Ohne Verzeichnwechsel erhält man die beiden als **Cosinus hyperbolicus** und **Sinus hyperbolicus** bezeichneten Funktionen:

$$\lim_{n\to\infty} \sum_{k=0}^{n} \frac{1}{(2\,k)!}\, x^{2\,k} = \cosh x, \tag{4.98}$$

$$\lim_{n\to\infty} \sum_{k=0}^{n} \frac{1}{(2\,k+1)!}\, x^{2\,k+1} = \sinh x. \tag{4.99}$$

Für diese gilt dann

$$\mathrm{e}^x = \cosh x + \sinh x. \tag{4.100}$$

Cosinus hyperbolicus und **Sinus hyperbolicus** tauchen gelegentlich in Anwendungen auf. So zeigt beispielsweise ein locker zwischen zwei Punkten hängendes Seil gerade die Form des Graphs des Cosinus hyperbolicus (vgl. Kettenlinie).

5 Vektoren und Matrizen

Die in diesem Kapitel vorgestellten Werkzeuge und Zusammenhänge werden auch unter dem Begriff “lineare Algebra” zusammengefasst.

5.1 Formalismus

Wir führen hier zunächst einige Begriffe und Schreibweisen ein, die ohne weitere Erklärungen nicht sonderlich sinnvoll oder gar nützlich erscheinen. In den nachfolgenden Unterkapitel füllen wir diese dann mit Leben und Sinn. So haben wir einerseits den "trockenen" Teil schnell hinter uns und konzentriert an einer Stelle zum späteren Nachschlagen; andererseits stören dann keine eher technischen als inhaltlichen Definitionen den Erklärfluss.

5.1.1 Vektoren

Der Begriff des Vektors wurde bereits in Section 4.2.2 als "Liste" von Zahlen eingeführt. Die Menge aller Vektoren der Länge n bezeichnen wir mit $\mathbb{R}^n$. Beachte, dass die Unterscheidung zwischen Punkt und Vektor nur eine Interpretationsfrage ist. Aus technischer Sicht besteht keinerlei Unterschied, da es sich in beiden Fällen um eine Liste von Zahlen handelt.

Für Vektoren führen wir zunächst zwei Rechenoperationen ein: die Addition zweier Vektoren und die Multiplikation eines Vektors mit einer Zahl.

Merke

Für zwei gleich lange Vektoren $a, b \in \mathbb{R}^n$ setzen wir

$$a + b := \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix} \tag{5.1}$$

und bezeichnen den so entstehenden Vektor $a + b \in \mathbb{R}^n$ als **Summe** der beiden Vektoren a und b.

Für eine Zahl $\lambda \in \mathbb{R}$ und einen Vektor $a \in \mathbb{R}^n$ setzen wir

$$\lambda\, a := \begin{bmatrix} \lambda\, a_1 \\ \vdots \\ \lambda\, a_n \end{bmatrix} \tag{5.2}$$

und bezeichnen den so entstehenden Vektor $\lambda\, a \in \mathbb{R}^n$ als **skalares Vielfaches** des Vektors a mit Faktor λ.

Beide Begriffe sind also einfach komponentenweise definiert. Beachte, dass weiter unten noch das Skalarprodukt zweier Vektoren eingefeführt werden wird. Dieses ist eine ganz andere Rechenoperation als das Bilden des skalaren Vielfachen!

Mit den beiden eingeführten Rechenoperationen können wir auch komplexere Ausdrücke formulieren. Eine häufig benötigte Kombination von Additionen und skalaren Vielfachen

ist die folgende:

Merke

Sind $a_1, \ldots, a_k \in \mathbb{R}^n$ Vektoren und sind $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ Zahlen, so bezeichnet man den Ausdruck

$$\sum_{l=1}^{k} \lambda_l \, a_l \qquad \big(= \lambda_1 \, a_1 + \cdots + \lambda_k \, a_k\big) \tag{5.3}$$

als **Linearkombination** der Vektoren $a_1, \ldots, a_k$ mit Koeffizienten $\lambda_1, \ldots, \lambda_k$.

Beachte, dass in dieser Definition jedes a_l ein Vektor ist! Im Gegensatz dazu hatten wir beim Einführen der Addition und des skalaren Vielfachen die einzelnen Komponenten des Vektors a mit $a_1, \ldots, a_n$ bezeichnet. **Bei Formeln mit Vektoren ist stets ganz genau zu prüfen, welche Rolle ggf. vorhandene Indizes spielen!**

5.1.2 Matrizen

Die folgende Definition mag zunächst etwas sinnarm erscheinen. Es wird sich jedoch zeigen, das Matrizen trotz ihres einfachen Aufbaus ein zentrales Werkzeug der Mathematik mit teils sehr tief liegender, vorerst in keiner Weise erkennbarer Bedeutung sind.

Merke

Eine **Matrix** ist ein rechteckiges Zahlenschema, in dem Zahlen in Zeilen und Spalten angeordnet sind. Die Menge aller Matrizen mit $m \in \mathbb{N}$ Zeilen und $n \in \mathbb{N}$ Spalten bezeichnen wir mit $\mathbb{R}^{m \times n}$ (sprich: "$\mathbb{R}$ hoch m Kreuz n").

Matrizen als Ganzes werden üblicherweise mit Großbuchstaben bezeichnet. Die **Einträge** (oder auch: **Elemente**) einer Matrix werden meist mit dem zugehörigen Kleinbuchstabe und zwei unteren Indizes bezeichnet. Der erste Index gibt die Zeilennummer an, der zweite die Spaltennummer. Für eine Matrix $A \in \mathbb{R}^{m \times n}$ schreibt man also beispielsweise

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m-1,1} & a_{m-1,2} & \cdots & a_{m-1,n-1} & a_{m-1,n} \\ a_{m1} & a_{m2} & \cdots & a_{m,n-1} & a_{m,n} \end{bmatrix}. \tag{5.4}$$

Statt viereckiger Klammern [...] sind auch runde Klammern (...) üblich. Die beiden Indizes werden, sofern es der besseren Lesbarkeit dient, durch Komma getrennt; also $a_{i,j}$ statt a_{ij}.

Völlig analog zu Vektoren sind auch für Matrizen **Summen und skalare Vielfache** elementweise definiert. Der Begriff der Linearkombination steht entsprechend auch bei Matrizen zur Verfügung, spielt hier aber kaum eine Rolle.

Merke

Vektoren kann man als Matrix mit nur einer Spalte oder nur einer Zeile ansehen. Die Matrizen in $\mathbb{R}^{m\times 1}$ bezeichnet man daher auch als **Spaltenvektoren** und die in $\mathbb{R}^{1\times n}$ als **Zeilenvektoren**.

Während man in der Mathematik Vektoren stets als Spaltenvektoren ansieht, wird bei der Arbeit mit dem Computer (je nach verwendeter Software) streng unterschieden zwischen Vektor (Liste von Zahlen) und Spaltenvektor (Rechteck von Zahlen mit nur einer Spalte).

```
import numpy as np

vektor = np.array([1, 4, 3])
spalte = np.array([[1], [4], [3]])
zeile = np.array([[1, 4, 3]])

print(vektor)
print(vektor.shape)
print()
print(spalte)
print(spalte.shape)
print()
print(zeile)
print(zeile.shape)
```

```
[1 4 3]
(3,)

[[1]
 [4]
 [3]]
(3, 1)

[[1 4 3]]
(1, 3)
```

Merke

Zu einer Matrix $A \in \mathbb{R}^{m\times n}$ bezeichnen wir mit $A^{\mathrm{T}} \in \mathbb{R}^{n\times m}$ die Matrix, die durch Vertauschen von Zeilen und Spalten aus A entsteht:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \qquad A^{\mathrm{T}} := \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}. \tag{5.5}$$

Die Matrix A^{T} wird als **transponierte Matrix** oder **Transponierte** zu A bezeichnet.

Um die Transponierte zu einer Matrix zu bilden nimmt man also eine Zeile nach der anderen der ursprünglichen Matrix und schreibt sie als Spalten nebeneinander.

Das Transponieren eines Spaltenvektors liefert einen Zeilenvektor und umgekehrt. Möchte man einen (Spalten-)Vektor in Fließtext unterbringen, so schreibt man ihn gern als transponierten Zeilenvektor, also beispielsweise $\begin{bmatrix}1 & 4 & 5\end{bmatrix}^{\mathrm{T}}$ oder $[1, 4, 5]^{\mathrm{T}}$.

Einige häufig auftretende Matrizen und Matrixtypen bekommen eigene Namen:

- Bei einer **Nullmatrix** sind alle Einträge Null.
- Eine **Diagonalmatrix** ist eine quadratische Matrix $A \in \mathbb{R}^{n\times n}$, bei der nur die Einträge $a_{11}, \ldots, a_{nn}$ von Null verschieden sind (sie dürfen aber auch Null sein). Alle anderen Einträge sind Null. Die Einträge $a_{11}, \ldots, a_{nn}$ werden auch als **Hauptdiagonale** von A bezeichnet.
- Eine **obere Dreiecksmatrix** ist eine Matrix, die unterhalb der Hauptdiagonale nur Nullen enthält.
- Eine **untere Dreiecksmatrix** ist eine Matrix, die oberhalb der Hauptdiagonale nur Nullen enthält.
- Eine **Einheitsmatrix** (bei gegebener Größe oft auch: die (!) Einheitsmatrix) ist eine Diagonalmatrix mit Einsen auf der Hauptdiagonale. Einheitsmatrizen werden oft mit I bezeichnet ohne weitere Erklärung des Symbols. Man rechnet leicht nach, dass

  $$A\,I = A \quad \text{und} \quad I\,A = A \tag{5.6}$$

  gilt (bei geeigneten Matrixgrößen).
- Eine **symmetrische Matrix** ist eine quadratische Matrix A, für die $A^{\mathrm{T}} = A$ gilt.

5.1.3 Länge eines Vektors

Merke

Als **Länge** oder **Norm** eines Vektors $a \in \mathbb{R}^n$ bezeichnen wir den Wert

$$\|a\| := \sqrt{\sum_{k=1}^{n} a_k^2} \qquad \left(= \sqrt{a_1^2 + \ldots + a_n^2} \right). \tag{5.7}$$

Nebenbemerkung

Beachte, dass die Begriffe Länge und Norm bei uns völlig synonym verwendet werden. Allerdings kann man auch andere Abstandsbegriff ("Normen") einführen. In diesem Fall meint man mit "Länge" dann die oben eingeführte Norm, die auch als **euklidische Norm** bekannt ist.

Einige Eigenschaften der Norm:

- Es gilt stets

$$\|a\| \geq 0. \tag{5.8}$$

- Ein Vektor hat genau dann die Länge Null, wenn es der Nullvektor ist. In Formeln:

$$\|a\| = 0 \quad \Leftrightarrow \quad a = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{5.9}$$

- Für $\lambda \in \mathbb{R}$ und $a \in \mathbb{R}^n$ gilt

$$\|\lambda\, a\| = |\lambda|\, \|a\|. \tag{5.10}$$

- Es gilt die **Dreiecksungleichung**, d.h. für beliebige Vektoren $a, b \in \mathbb{R}^n$ gilt

$$\|a + b\| \leq \|a\| + \|b\|. \tag{5.11}$$

Die ersten beiden Eigenschaften folgen sofort aus der Definition der Norm. Der Nachweis der dritten Eigenschaft macht etwas mehr Arbeit, bringt aber keine neuen Erkenntnisse und unterbleibt deshalb an dieser Stelle.

Merke

Bildet man zu einem Vektor $a \in \mathbb{R}^n$ den Vektor

$$\tilde{a} := \frac{1}{\|a\|}\, a, \tag{5.12}$$

so sagt man, dass der Vektor a **normiert** wird. Der so entstehende Vektor $\tilde{a}$ zeigt in die gleiche Richtung wie a, hat aber stets die Länge 1:

$$\|\tilde{a}\| = \left\| \frac{1}{\|a\|}\, a \right\| = \left| \frac{1}{\|a\|} \right| \|a\| = 1. \tag{5.13}$$

Üblicherweise schreibt man auch gebrochene Faktoren nicht mit Bruchstich unter einen Vektor, z.B.

$$\frac{1}{2}\, a \quad \text{(üblich)}, \qquad\qquad \frac{a}{2} \quad \text{(unüblich)}. \tag{5.14}$$

Einzige Ausnahme bildet das Normieren:

$$\frac{1}{\|a\|}\, a \quad \text{(üblich)}, \qquad\qquad \frac{a}{\|a\|} \quad \text{(ebenfalls üblich)}. \tag{5.15}$$

5.1.4 Skalarprodukt zweier Vektoren

Analog zur noch nicht erkennbaren Bedeutung von Matrizen, wird auch der tiefe Nutzen der folgenden, schon fast banalen Rechenoperation erst im Laufe der nächsten Unterkapitel deutlich werden.

Merke

Für zwei gleich lange Vektoren $a, b \in \mathbb{R}^n$ setzen wir

$$a \circ b := \sum_{k=1}^{n} a_k\, b_k \qquad \big(= a_1\, b_1 + a_2\, b_2 + \cdots + a_n\, b_n\big) \tag{5.16}$$

und bezeichnen diesen Wert $a \circ b \in \mathbb{R}$ als **Skalarprodukt** der beiden Vektoren a und b. Gelegentlich wird auch die Bezeichnung “inneres Produkt” verwendet (vor allem im Englischen: “inner product”).

Beachte, dass wir das Symbol $\circ$ vorher schon für die Hintereinanderausführung von Abbildungen verwendet haben. Aufgrund der Vielzahl von Operationen und dem chronischen Mangel an Symbolen sind in der Mathematik viele Symbole doppelt und dreifach mit Bedeutung belegt. Die konkrete Bedeutung ergibt sich jeweils aus dem Kontext.

Später werden wir für Vektoren mit drei Komponenten noch eine andere Art von Produkt einführen, das sogenannte Kreuzprodukt oder äußere Produkt. Deshalb sollte immer

"Skalarprodukt" und nicht nur "Produkt" gesagt und geschrieben werden.

Aus der Definition des Skalarprodukts folgt sofort

$$a \circ a = \|a\|^2. \tag{5.17}$$

Weniger offensichtlich ist

$$|a \circ b| \leq \|a\| \, \|b\|, \tag{5.18}$$

die sogenannte **Cauchy-Schwarz-Ungleichung**, die wir hier nicht herleiten. Bei dieser gilt Gleichheit genau dann, wenn die Vektoren a und b Vielfache voneinander sind, also insbesondere bei $a = b$.

Weiter wichtige Eigenschaften, die sofort aus der Definition folgen und deren Kenntnis das Rechnen mit Skalarprodukten einfacher macht:

- Das Skalarprodukt mit dem Nullvektor ist stets Null:

$$0 \circ a = 0 \quad \text{und} \quad a \circ 0 = 0. \tag{5.19}$$

- Das Skalarprodukt ist **kommutativ**, d.h.

$$a \circ b = b \circ a. \tag{5.20}$$

- Es gilt das **Distributivgesetz**

$$(a + b) \circ c = (a \circ c) + (b \circ c) \quad \text{und} \quad a \circ (b + c) = (a \circ b) + (a \circ c). \tag{5.21}$$

- Für $\lambda \in \mathbb{R}$ gilt

$$(\lambda \, a) \circ b = \lambda \, (a \circ b) \quad \text{und} \quad a \circ (\lambda \, b) = \lambda \, (a \circ b). \tag{5.22}$$

Beachte, dass das von der Multiplikation bei Zahlen bekannte Assoziativgesetz $a \, (b \, c) = (a \, b) \, c$ hier keine Rolle spielt, da das Ergebnis des Skalarprodukts zweier Vektoren kein Vektor ist. Somit können keine Ketten von drei oder mehr Faktoren gebildet werden.

5.1.5 Produkt zweier Matrizen

Rein formal könnte man auch für Matrizen eine Art Skalarprodukt einführen (elementweise multiplizieren und alle Produkte addieren). Dies hat aber keinerlei Nutzen für uns und unterbleibt deshalb. Jedoch wird eine andere, etwas komplexere Form von Produkt von wesentlicher Bedeutung sein. In der folgenden Definition (und auch sonst) bezeichnen wir mit $a_{\bullet j}$ die j-te Spalte von A (Spaltenvektor) und mit $a_{i\bullet}$ die i-te Zeile von A (Zeilenvektor).

Merke

Für zwei Matrizen $A \in \mathbb{R}^{m\times p}$ und $B \in \mathbb{R}^{p\times n}$ setzen wir

$$A\,B := \begin{bmatrix} a_{1\bullet} \circ b_{\bullet 1} & \cdots & a_{1\bullet} \circ b_{\bullet n} \\ \vdots & \ddots & \vdots \\ a_{m\bullet} \circ b_{\bullet 1} & \cdots & a_{m\bullet} \circ b_{\bullet n} \end{bmatrix} \tag{5.23}$$

und bezeichnen die so entstehende Matrix $A\,B \in \mathbb{R}^{m\times n}$ als **Matrixprodukt** oder einfach als Produkt der Matrizen A und B.

Beachte, dass das Produkt nur für Matrizenpaare erklärt ist, bei denen die erste Matrix gerade so viele Spalten hat wie die zweite Zeilen besitzt. Für alle anderen Kombinationen von Matrizen ist nicht klar, was man unter ihrem Produkt verstehen soll. Die Situation ist vergleichbar mit der Division durch Null bei Zahlen: Man kann der Sache auch mit viel gutem Wille keinen Sinn beimessen; die Operation ist schlicht "nicht definiert".

Beispiel

Die Multiplikation zweier Matrizen ist etwas fummelig, aber problemlos machbar. Man merke sich die Regel "Zeile mal Spalte" (im Sinne des Skalarprodukts). Diese Regel führt man für alle Zeilen der ersten und alle Spalten der zweiten Matrix aus. So erhält man beispielsweise

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \begin{bmatrix} 3 & 7 & 0 \\ 5 & 3 & 0 \\ 1 & 2 & 0 \\ 2 & 5 & 0 \end{bmatrix} = \begin{bmatrix} 24 & 38 & 0 \\ 68 & 107 & 0 \end{bmatrix} \tag{5.24}$$

(Video 505, Notizen dazu)

Das Produkt zweier Matrizen kann natürlich auch mit dem Computer berechnet werden. Beachte, dass der übliche Multiplikationsoperator `*` das elementweise Produkt bildet (sofern die Matrizengrößen übereinstimmen). Für das Matrixprodukt ist `@` zu verwenden.

```
A = np.array([[1, 2, 3, 4], [5, 6, 7, 8]])
B = np.array([[3, 7, 0], [5, 3, 0], [1, 2, 0], [2, 5, 0]])

A @ B
```

```
array([[ 24,  39,   0],
       [ 68, 107,   0]])
```

Das Matrixprodukt eines Zeilenvektors mit einem Spaltenvektor liefert eine Matrix aus $\mathbb{R}^{1\times 1}$, also praktisch nur eine Zahl. Entsprechend schreibt man das Skalarprodukt zweier

Vektoren gern als

$$a \circ b = a^{\mathrm{T}}\, b \tag{5.25}$$

und sieht die rechte Seite (formal nicht ganz korrekt) als Zahl an. Beachte aber, dass bei umgekehrter Reihenfolge, also Spaltenvektor mal Zeilenvektor eine große quadratische Matrix entsteht!

Man kann sich überlegen, dass Matrizenmultiplikation **assoziativ** ist, d.h. bei Produkten mit drei oder mehr Faktoren ist es egal, in welcher Abfolge man die einzelnen Produkte bildet. Für Matrizen A, B, C passender Größe gilt also

$$(A\,B)\,C = A\,(B\,C), \tag{5.26}$$

sodass man auch einfach $A\,B\,C$ schreiben kann.

Allerdings ist Matrizenmultiplikation **nicht kommutativ**, d.h. man kann die Faktoren im Allgemeinen nicht vertauschen. Einerseits passen nach dem Vertausch der Faktoren die Matrixgrößen eventuell nicht mehr; andererseits kann sich das Ergebnis trotz passender Größen ändern.

Beispiel

Man rechnet leicht nach, dass

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \tag{5.27}$$

gilt.

Wie bei Zahlen gilt auch bei Matrizen das **Distributivgesetz**, welches Addition und Multiplikation verbindet: Für Matrizen $A \in \mathbb{R}^{m\times p}$ und $B, C \in \mathbb{R}^{p\times n}$ gilt

$$A\,(B + C) = A\,B + A\,C. \tag{5.28}$$

In umgekehrter Reihenfolge gilt für $A, B \in \mathbb{R}^{m\times p}$ und $C \in \mathbb{R}^{p\times n}$ entsprechend

$$(A + B)\,C = A\,C + B\,C. \tag{5.29}$$

Beachte: Da Matrizenmultiplikation nicht kommutativ ist, folgt aus dem ersten Teil des Distributivgesetzes nicht automatisch der zweite!

Nebenbemerkung

Assoziativgesetz und Distributivgesetz bei Matrizen sind trotz ihres Namens keine "Gesetze", sondern durch logisches Schließen aus der Definition von Addition und Multiplikation herleitbare Aussagen. Das Herleiten dieser beiden Aussagen ist ohne große (geistige) Mühe möglich, aber recht länglich und unterbleibt deshalb an dieser Stelle.

Eine weitere nützliche Eigenschaft des Matrixprodukts, die sofort aus der Definition folgt, ist

$$(\lambda A)\, B = \lambda\, (A\, B) \quad \text{und} \quad A\, (\lambda\, B) = \lambda\, (A\, B), \tag{5.30}$$

wobei A und B Matrizen passender Größe sind und $\lambda \in \mathbb{R}$ eine beliebige Zahl ist.

Auch kann man zeigen, dass für das Zusammenspiel von Multiplizieren und Transponieren die Beziehung

$$(A\, B)^{\mathrm{T}} = B^{\mathrm{T}}\, A^{\mathrm{T}} \tag{5.31}$$

gilt.

5.1.6 Hinweise zum Matrix-Vektor-Produkt

Sieht man Vektoren als Spaltenvektoren an, wie wir es stets tun, so kann man das Produkt aus Matrix und Vektor als Produkt zweier Matrizen berechnen. Das Produkt einer Matrix $A \in \mathbb{R}^{m\times n}$ mit einem Vektor $x \in \mathbb{R}^n$ liefert dann den Vektor $A\, x \in \mathbb{R}^m$. Umgekehrt kann man für $x \in \mathbb{R}^m$ auch $x^{\mathrm{T}}\, A \in \mathbb{R}^n$ bilden. Wir wollen hier einige Hinweise zur genaueren Struktur von Matrix-Vektor-Produkten sammeln, da solche Produkte sehr häufig auftauchen werden.

Schaut man sich die Arbeitsweise der Matrizenmultiplikation an, so sieht man für $A \in \mathbb{R}^{m\times n}$ und $x \in \mathbb{R}^n$, dass

$$A\, x = x_1\, a_{\bullet 1} + \cdots + x_n\, a_{\bullet n}. \tag{5.32}$$

Das Produkt liefert also gerade eine Linearkombination der Matrixspalten, wobei die Vektorkomponenten die Koeffizienten sind. Bei umgekehrter Reihenfolge (jetzt $x \in \mathbb{R}^m$!) gilt

$$x^{\mathrm{T}}\, A = x_1\, a_{1\bullet} + \cdots + x_m\, a_{\bullet m} \tag{5.33}$$

(Linearkombination der Zeilen).

Ebenfalls sehr einfach sieht man, dass die Spalten des Produkts zweier Matrizen $A \in \mathbb{R}^{m\times p}$ und $B \in \mathbb{R}^{p\times n}$ als Matrix-Vektor-Produkte geschrieben werden können:

$$A\, B = \begin{bmatrix} | & & | \\ A\, b_{\bullet 1} & \cdots & A\, b_{\bullet n} \\ | & & | \end{bmatrix}. \tag{5.34}$$

Dabei soll die Schreibweise mit | andeuten, dass die Vektoren $A\, b_{\bullet j}$ als Spalten der Matrix fungieren.

Das Berechnen von Matrix-Vektor-Produkten mit dem Computer erfolgt wie bei Produkten zweier Matrizen. Die explizite Umwandlung des Vektors in einen Spaltenvektor (Matrix mit einer Spalte) ist nicht nötig, da dies intern von NumPy erledigt wird. Wird der Vektor als linker Faktor verwendet, so wird er automatisch als Zeilenvektor interpretiert.

```
A = np.array([[1, 2], [3, 4], [5, 6]])
x = np.array([7, 8])

A @ x
```

```
array([23, 53, 83])
```

5.1.7 Determinante einer Matrix

Matrizen sieht man nur schlecht auf den ersten Blick wichtige Eigenschaften an. Deshalb wurden im Laufe der Zeit verschiedene Kennzahlen entwickelt, die relevante Informationen aus (beliebig großen) Matrizen in einer Zahl zusammenfassen. Führen hier zunächst nur die sogeannte Determinante einer Matrix ein; später folgt noch der Rang einer Matrix.

Wir werden Determinanten nur für Matrizen in $\mathbb{R}^{2\times 2}$ und $\mathbb{R}^{3\times 3}$ einführen. Für größere Matrizen sei bei Bedarf auf den Entwicklungssatz von Laplace verwiesen. Einige Hinweise zur Rechtfertigung dieses "sparsamen" Vorgehens:

- Für die praktische Arbeit (vor allem Koordinatentransformationen) spielen nur kleine Matrizen eine Rolle.
- Für größere Matrizen ist der Aufwand zum Berechnen von Determinanten so groß, dass selbst Computer schnell an ihre Leistungsgrenze kommen. Manuelles Rechnen scheidet somit aus.
- Das Berechnen von Determinanten ist ohne jede Denkarbeit möglich, also bestens für Computer geeignet.
- Die allgemeine Formel für Determinanten zu Matrizen beliebiger Größe benötigt viel Raum und viel Zeit für Erklärungen, die wir zielführender in andere Themen investieren.

Merke

Für eine Matrix $A \in \mathbb{R}^{2\times 2}$ heißt die Zahl

$$\det A := a_{11}\,a_{22} - a_{12}\,a_{21} \tag{5.35}$$

Determinante der Matrix.

Für eine Matrix $A \in \mathbb{R}^{3\times 3}$ heißt die Zahl

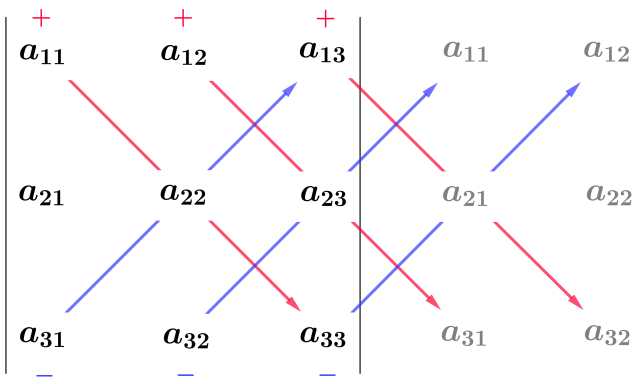
$$\begin{aligned}\det A := {} & a_{11}\,a_{22}\,a_{33} + a_{12}\,a_{23}\,a_{31} + a_{13}\,a_{21}\,a_{32} \\ & - a_{13}\,a_{22}\,a_{31} - a_{11}\,a_{23}\,a_{32} - a_{12}\,a_{21}\,a_{33}\end{aligned} \tag{5.36}$$

Determinante der Matrix.

Manchmal schreibt man auch $|A|$ für die Determinanten bzw.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}. \tag{5.37}$$

Für die Berechnung von Determinanten von Matrizen in $\mathbb{R}^{3\times 3}$ gibt es sehr verschieden aussehende Formeln. Die hier verwendete ist als **Regel von Sarrus** bekannt.



Durch Linien verbundene Einträge werden multipliziert; die Produkte werden dann additiert (blau) bzw. subtrahiert (rot) (commons.wikimedia.org, Nutzer Kmhkmh, CC BY 4.0).

Anhand der Determinante kann man beispielsweise die Lösbarkeit von Gleichungssystemen erkennen (dazu später mehr). Ihren Hauptzweck erfüllen sie aber als Werkzeug für die platzsparende Beschreibung von Streckungs- und Spiegelungsvorgängen bei Koordinatentransformationen aller Art. Im Rahmen dieser Veranstaltung werden wir nicht tief genug in diese Thematik vorstoßen, um den Nutzen von Determinanten dort zu erkennen. Jedoch werden Determinanten in den weiterführenden Veranstaltungen im Laufe des Studiums eine Rolle spielen.

5.2 Vektorgeometrie

Vektoren und Matrizen als zunächst rein formale Werkzeuge sind sehr gut geeignet zur Beschreibung geometrischer Operationen (Verschiebungen, Drehungen, Streckungen, Projektionen,...). Beschäftigen uns hier zunächst mit Verschiebungen und einfachen Projektionen. Weitere Operationen folgen in späteren Unterkapiteln.

5.2.1 Verschiebungen

Möchte man eine gegebene Menge von Punkten im $\mathbb{R}^n$ in eine gegebene Richtung verschieben, so kann man diese Operation mittels Vektoraddition beschreiben. An dieser Stelle ist es üblich und sinnvoll zwischen Punkten und Vektoren zu unterscheiden. Ein Punkt $P \in \mathbb{R}^n$ gibt einen geometrischen Ort an, während ein Vektor $r \in \mathbb{R}^n$ eine Richtung angibt. Die Summe $P+r$ ist dann wieder als (verschobener) Punkt zu interpretieren. Einen formal sauberen Übergang zwischen Punkten und Vektoren schafft man sich über den Begriff des Ortsvektors:

Merke

Bezeichnen mit $O := (0, \ldots, 0) \in \mathbb{R}^n$ (großes lateinisches O) den Punkt im Koordinatenursprung. Für einen Punkt $P := (p_1, \ldots, p_n) \in \mathbb{R}^n$ heißt der Vektor

$$\overrightarrow{OP} := \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} \in \mathbb{R}^n \tag{5.38}$$

Ortsvektor des Punktes.

Beachte hier, dass wir entgegen einer recht weit verbreitenen (Un-)Sitte "$P := (p_1, \ldots, p_n)$" statt "$P(p_1, \ldots, n_n)$" für die Zuordnung von Koordinaten zu einem Punkt schreiben. Letzteres ist zu leicht mit einer Funktionsauswertung zu verwechseln und verschlechtert damit die Lesbarkeit.

Zu jedem Punkt gibt es also einen Ortsvektor. Umgekehrt können wir jeden Vektor als Ortsvektor eines Punktes interpretieren, indem wir die Komponenten des Vektors als Koordinaten eines Punktes ansehen.

Mit Hilfe von Ortsvektoren können wir nun die Verschiebung von Punkten beschreiben:

Merke

Sind $P := (p_1, \ldots, p_n)$ und $Q := (q_1, \ldots, q_n)$ zwei Punkte im $\mathbb{R}^n$, so bezeichnen

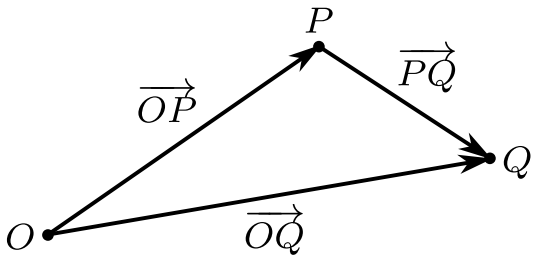
wir mit

$$\overrightarrow{PQ} := \begin{bmatrix} q_1 - p_1 \\ \vdots \\ q_n - p_n \end{bmatrix} \in \mathbb{R}^n \tag{5.39}$$

den Vektor, der P im Sinne von

$$\overrightarrow{OQ} = \overrightarrow{OP} + \overrightarrow{PQ} \tag{5.40}$$

nach Q verschiebt.



Die Verschiebung eines Punktes kann als Addition des Ortsvektors mit dem Richtungsvektor beschrieben werden.

Im Prinzip könnten wir für einen Punkt P und einen Vektor r nun die Operation

$$P + r := \overrightarrow{OP} + r \tag{5.41}$$

definieren. Üblicherweise belässt man es aber bei $\overrightarrow{OP}+r$, hat also ausschließlich Vektoren in Formeln stehen.

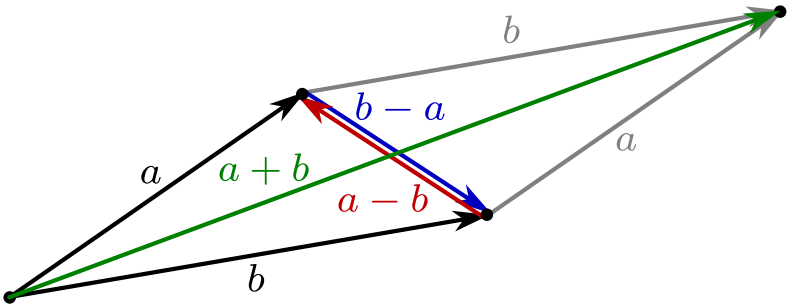
Nebenbemerkung (koordinatenfreie Geometrie)

Man kann die Arbeit mit Punkten und Vektoren in gewissem Maße auch "koordinatenfrei" betreiben. Punkte sind dann Orte in der Ebene (oder im Raum) ohne Bezug zu einem konkreten Koordinatensystem. Vektoren sind "Verschiebungen", die man sich als lose in Ebene (oder Raum) herumliegende Pfeile vorstellt. Die "Addition" von Vektoren beschreibt man verbal als "Hintereinanderhängen" von Vektoren. Möchte man einen Punkt in eine gewisse Richtung verschieben, so wählt man sich einen Referenzpunkt O (mangels Koordinatensystem gibt es keinen Ursprung) und drückt die Verschiebung als Addition (d.h. "Hintereinanderhängen") des Vektors, der von O zum Punkt zeigt, und des Richtungsvektors aus.

Diese Herangehensweise ist recht verlockend, da die meisten geometrischen Operationen (z.B. auch Drehungen) völlig unabhängig von Koordinatensystemen beschrieben werden können. Allerdings ist es schwierig bis unmöglich die benötigten

Begriffe wirklich sauber zu definieren. Auch ist dieser Ansatz in der Praxis ohne Relevanz, da stets ein kanonisches Koordinatensystem vorhanden ist, insbesondere bei der Arbeit mit dem Computer.

Angemerkt sei noch, dass man aus den bisherigen Überlegungen auch die geometrische Interpretation der Summe zweier Vektoren $a, b \in \mathbb{R}^n$ erhält. Die Summe entspricht der Diagonale des von den beiden Vektoren beschriebenen Parallelogramms, die am gemeinsamen Ausgangspunkt der beiden Vektoren beginnt. Die andere Diagonale ist dann gerade die Differenz.



Zwei Vektoren beschreiben ein Parallelogramm. Die beiden Diagonalen sind gerade Summe und Differenz der beiden Vektoren.

5.2.2 Abstand zweier Punkte

Mit dem bis hier eingeführten Formalismus der Vektorrechnung und den Betrachtungen zur Verschiedung von Punkten erhalten wir folgenden Zusammenhang:

Merke

Der **Abstand zweier Punkte** $P, Q \in \mathbb{R}^n$ ist gegeben durch die Länge des Vektors $\overrightarrow{PQ}$.

Mit Koordinaten $P = (p_1, \ldots, p_n)$ und $Q = (q_1, \ldots, q_n)$ gilt also

$$\text{Abstand von } P \text{ und } Q = \|\overrightarrow{PQ}\| = \sqrt{\sum_{k=1}^{n} (p_k - q_k)^2}. \tag{5.42}$$

5.2.3 Projektion eines Vektors auf einen anderen Vektor

Öfter als man denkt werden wir eine Operation benötigen, die als "Projektion eines Vektors auf einen anderen Vektor" bezeichnet wird. Insbesondere wird uns diese Operation sehr beim Wechsel zwischen verschiedenen Koordinatensystemen behilflich sein. Hier zunächst nur ein motivierendes Beispiel aus der Schulphysik.

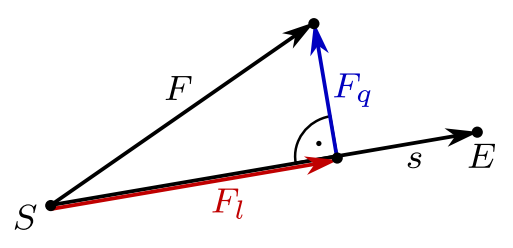
Beispiel (mechanische Arbeit)

Bewegt sich ein Objekt auf direktem Weg (also geradlinig) von einem Punkt zu einem anderen unter dem Einfluss von Kräften, so wird (mechanische) Arbeit verrichtet. Denken wir beispielsweise an einen Eisenbahnwagon auf einem geraden Streckenabschnitt, so wirken unter anderem Reibungskräfte und vielleicht eine Hangabtriebskraft (bei Steigung/Gefälle). Der Antrieb (Lok) muss zur Bewegung des Wagons Arbeit verrichten, wobei das physikalische Gesetz

$$\text{Arbeit} = \text{Weg} \cdot \text{Kraft in Wegrichtung}$$

gilt. Wirken nun auch Kräfte in andere als die Bewegungsrichtung (z.B. Wind), so hilft diese einfache Regel nicht mehr weiter. Wir müssen daher wahlweise auf ein allgemeiner formuliertes Gesetz ausweichen oder den Anteil der Kraft, der in Wegrichtung wirkt, aus der Gesamtkraft berechnen. Beide Varianten sind machbar; interessieren uns hier aber zunächst für die letztere (das allgemeinere Gesetz erhalten wir später (fast) geschenkt).

Beschreiben den zurückgelegten Weg zwischen Startpunkt S und Endpunkt E als Vektor $s := \overrightarrow{SE}$. Die wirkende Kraft können wir ebenfalls als Vektor F beschreiben. Die Richtung des Vektors entspricht der Richtung der Kraft und die Länge des Vektors entspricht dem Betrag der Kraft. Wollen den Kraftvektor in eine Summe $F_l + F_q$ zweier Vektoren zerlegen, wobei F_l längs des Weges und F_q quer zum Weg zeigt. Die Länge von F_l ist dann gerade der Betrag der Kraft in Wegrichtung, also die Größe, die wir für die obige Formel aus der Schulphysik benötigen.



Beachte, dass die Kräfte F und F_l an jedem Punkt des Weges auf das Objekt wirken, hier aber nur am Startpunkt S gezeichnet sind. Entsprechend wirkt auch die Kraft F_q an jedem Punkt des Weges, wurde hier aber an einer für die Visua-

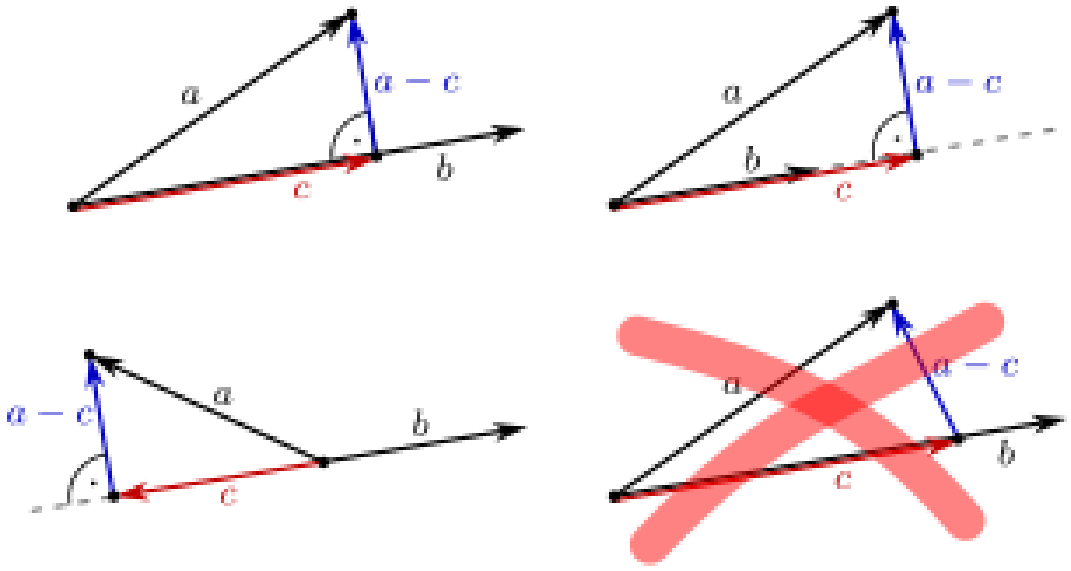
lisierung der Zerlegung von F günstigen Stelle eingetragen.

Merke

Seien $a, b \in \mathbb{R}^n$ zwei Vektoren. Ein Vektor $c \in \mathbb{R}^n$ heißt **Projektion von a auf b**, wenn die folgenden beiden Bedingungen erfüllt sind:

- c ist ein Vielfaches von b,
- $a-c$ ist der kürzeste Vektor bezüglich aller c, die die erste Bedingung erfüllen.

Anschaulich besagt die erste Bedingung, dass c in die Richtung von b (oder genau entgegen) zeigt. Die zweite Bedingung fordert indirekt, dass $a - c$ senkrecht zu b verläuft, wie wir später sehen werden, wenn wir Winkel zwischen Vektoren betrachten.



Die Länge des Vektors b spielt keine Rolle; nur seine Richtung ist relevant für die Projektion von a auf b, wobei die Projektion c auch genau die entgegengesetzte Richtung zu b haben kann. Ist $a - c$ nicht der kürzest mögliche Vektor, so spricht man nicht von einer Projektion.

Es zeigt sich, dass die Projektion eines Vektors auf einen anderen sehr leicht mit Hilfe des Skalarprodukts der beiden gegebenen Vektoren berechnet werden kann. Wir müssen also nicht mit Winkeln hantieren, sondern nur mit einfachen Summen und Produkten von Zahlen. Dies ist der große Wert des so banal erscheinenden Skalarprodukts! Es spielt in gewisser Weise die Rolle von Winkeln zwischen Vektoren, ohne dass wir uns mit Winkeln und den dann kaum vermeidbaren Winkelfunktionen (Sinus, Cosinus,...) herumschlagen müssen. Im nächsten Abschnitt gehen wir genauer auf den Zusammenhang zwischen

Skalarprodukt und Winkeln ein.

Merke

Seien $a, b \in \mathbb{R}^n$ zwei Vektoren und sei $c \in \mathbb{R}^n$ die Projektion von a auf b. Dann gilt

$$c = \left(a \circ \frac{b}{\|b\|}\right) \frac{b}{\|b\|} \tag{5.43}$$

(Video 510, Notizen dazu).

Die Projektion von a auf b kann man also in zwei Schritten berechnen:

1. b normieren.
2. $(a \circ \tilde{b})\,\tilde{b}$ berechnen, wobei $\tilde{b}$ der normierte Vektor b ist.

Beispiel (allgemeine Formel für die mechanische Arbeit)

Betrachten nochmal die aus der Schulphysik bekannte Berechnung der mechanischen Arbeit. Sei also s wieder der Wegvektor und F der Kraftvektor. Dann ist die Länge des Weges $\|s\|$ und der Betrag der in Wegrichtung wirkenden Kraft ist

$$\left\| \left(F \circ \frac{s}{\|s\|}\right) \frac{s}{\|s\|} \right\| = \left| F \circ \frac{s}{\|s\|} \right| \left\| \frac{s}{\|s\|} \right\| = \left| F \circ \frac{s}{\|s\|} \right|. \tag{5.44}$$

Die aus der Schulphysik bekannte Formel zur Berechnung der mechanischen Arbeit liefert somit

$$W = \left| F \circ \frac{s}{\|s\|} \right| \|s\| = |F \circ s|. \tag{5.45}$$

Für Kräfte, die nicht in Wegrichtung wirken, gilt also im Sinne des Skalarprodukts ebenfalls die bekannte Formel “Arbeit ist Kraft mal Weg”.

Umgekehrt erhalten wir aus der allgemeinen Formel natürlich die einfache Formel zurück: Zeigt der Kraftvektor F in Wegrichtung, so gilt

$$F = \|F\| \frac{s}{\|s\|} \tag{5.46}$$

und Einsetzen in die allgemeine Formel liefert

$$\begin{aligned} W = |F \circ s| &= \left| \left(\|F\| \frac{s}{\|s\|} \right) \circ s \right| = \left| \frac{\|F\|}{\|s\|} (s \circ s) \right| \\ &= \left| \frac{\|F\|}{\|s\|} \|s\|^2 \right| = \left| \|F\| \, \|s\| \right| = \|F\| \, \|s\|. \end{aligned} \tag{5.47}$$

5.2.4 Winkel zwischen zwei Vektoren

Wollen den in $\mathbb{R}^2$ und $\mathbb{R}^3$ intuitiv klaren Begriff des Winkels zwischen zwei Vektoren formalisieren (also ohne Bezug zu Intuition und Anschauung erklären), damit er auch in $\mathbb{R}^n$ mit $n > 3$ zur Verfügung steht.

Merke

Als **Winkel zwischen zwei Vektoren** $a, b \in \mathbb{R}^n \setminus \{0\}$ bezeichnen wir die Zahl

$$\arccos\left(\frac{a}{\|a\|} \circ \frac{b}{\|b\|}\right). \tag{5.48}$$

Dazu einige Hinweise:

- Ist einer der beiden Vektoren der Nullvektor, so ist der Begriff "Winkel" nicht sinnvoll erklärbar. Entsprechend wird in der Definition $a \neq 0$ und $b \neq 0$ vorausgesetzt (0 ist hier der Nullvektor in $\mathbb{R}^n$).
- Der Definitionsbereich von arccos ist das Intervall $[-1, 1]$. Aus der Cauchy-Schwarz-Ungleichung (vgl. Section 5.1.4) erhalten wir

  $$\left|\frac{a}{\|a\|} \circ \frac{b}{\|b\|}\right| \leq \left\|\frac{a}{\|a\|}\right\| \left\|\frac{b}{\|b\|}\right\| = 1, \tag{5.49}$$

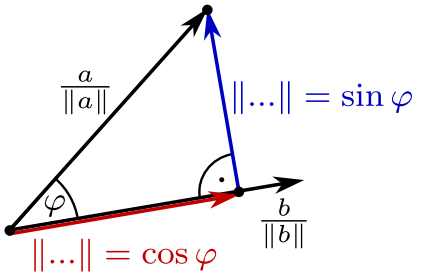
  sodass der Ausdruck in der Definition des Winkels stets wohldefiniert ist.
- Der Winkel zwischen zwei Vektoren liegt stets im Intervall $[0, \pi]$.
- Bezeichnen wir mit φ den Winkel zwischen den beiden Vektoren a und b, so folgt aus der Definition sofort

  $$a \circ b = \|a\| \, \|b\| \, \cos\varphi. \tag{5.50}$$

Dass die obige Definition des Begriffs "Winkel" mit der üblichen Verwendung des Begriffs in Einklang steht, sieht man aus der Berechnung der Projektion von Vektoren. Bei normierten Vektoren entspricht die Länge des projizierten Vektors (laut Definition des Cosinus) gerade dem Betrag des Cosinus des Winkels zwischen den beiden Vektoren, sofern der Winkel im Bereich $[0, \frac{\pi}{2}]$ liegt. Für Winkel in $[\frac{\pi}{2}, \pi]$ ist die negative Länge des projizierten Vektors gerade der Cosinus des Winkels.

Aus der Beziehung zwischen Skalarprodukt und Winkel erhält man nützliche Informationen über das Vorzeichen des Skalarprodukts:

- Für Winkel in $[0, \frac{\pi}{2})$ (spitzer Winkel) gilt $a \circ b > 0$.
- Für senkrechte Vektoren gilt $a \circ b = 0$.



Die Länge des projizierten Vektors entspricht gerade dem Cosinus des Winkels. Die Länge des Differenzvektors zwischen Ausgangsvektor und projiziertem Vektor entspricht gerade dem Sinus des Winkels.

- Für Winkel in $(\frac{\pi}{2}, \pi]$ (stumpfer Winkel) gilt $a \circ b < 0$.

Somit kann nun auch ganz ohne den Begriff des Winkels von Orthogonalität sprechen. Tatsächlich werden wir, wo immer möglich, die Arbeit mit Winkeln zwischen Vektoren meiden und stattdessen das viel bequemere Skalarprodukt nutzen.

Merke

Zwei Vektoren $a, b \in \mathbb{R}^n$ heißen **orthogonal** zueinander, wenn $a \circ b = 0$ gilt.

Beachte, dass nach dieser Definition der Nullvektor orthogonal zu jedem anderen Vektor ist.

5.3 Lineare Abbildungen

5.3.1 Begriff und Beispiele

Viele geometrische Operationen wie Drehen, Spiegeln und Strecken (nicht jedoch Verschieben!) haben eine gemeinsame Struktur, deren Kenntnis sich insbesondere für die rechnerische Umsetzung dieser Operationen vorteilhaft nutzen lässt. Auch die Umsetzung geometrischer Operationen am Computer, z.B. in CAD-Systemen, basiert ganz wesentlich auf dieser gemeinsamen Struktur.

Merke

Eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m$ heißt **linear**, wenn folgende Bedingungen erfüllt sind:

- $f(a+b) = f(a) + f(b)$ für alle $a, b \in \mathbb{R}^n$,
- $f(\lambda\, a) = \lambda\, f(a)$ für alle $a \in \mathbb{R}^n$ und alle $\lambda \in \mathbb{R}$.

Die beiden Bedingungen besagen, dass es bei linearen Abbildungen egal ist, ob man erst die Rechenoperation bei Vektoren anwendet (also Addition und skalares Vielfaches) und dann die Abbildung, oder ob man erst die Abbildung auf die Ausgangsvektoren anwendet und anschließend die notwendigen Rechenoperationen durchführt. Man sagt auch, dass lineare Abbildungen die Struktur des $\mathbb{R}^n$ erhalten bzw. diese in den $\mathbb{R}^m$ übertragen.

Beispiel (Drehung)

Die Abbildung, die jeden Punkt $(x, y) \in \mathbb{R}^2$ um den Winkel $\alpha \in \mathbb{R}$ um den Koordinatenursprung dreht, ist offensichtlich linear, denn:

- Die Summe zweier Vektoren bildet die Diagonale des von den beiden Vektoren aufgespannten Parallelogramms. Beim Drehen um den Ursprung verändert das Parallelogramm seine Form nicht.
- Ob ein Vektor erst gestreckt wird (skalares Vielfaches) und dann gedreht oder umgekehrt, führt offensichtlich zum selben Ergebnis.

Das gleiche gilt für die Drehung um eine durch den Ursprung verlaufende Achse im $\mathbb{R}^3$.

Beispiel (Spiegelung)

Die Abbildung, die jeden Punkt $(x, y) \in \mathbb{R}^2$ an einer durch den Koordinatenursprung verlaufenden Gerade spiegelt, ist offensichtlich linear. Ebenso ist die Punktspiegelung $(x, y) \mapsto (-x, -y)$ linear.

Gleiches gilt im $\mathbb{R}^3$ für Spiegelung an einer durch den Ursprung verlaufenden Ebene und für Punktspiegelung am Ursprung.

Beispiel (Streckung)

Die Abbildung $a \mapsto \lambda\, a$ für $a \in \mathbb{R}^n$ und festes $\lambda \in \mathbb{R}$ ist offensichtlich linear.

Merke

Die Hintereinanderausführung linearer Abbildungen ist linear, denn für zwei lineare Abbildungen $f : \mathbb{R}^n \to \mathbb{R}^p$ und $g : \mathbb{R}^p \to \mathbb{R}^m$ gelten

$$\begin{aligned}(g \circ f)(a+b) &= g\big(f(a+b)\big)\\ &= g\big(f(a)+f(b)\big)\\ &= g(f(a)) + g(f(b))\\ &= (g \circ f)(a) + (g \circ f)(b)\end{aligned} \tag{5.51}$$

und

$$\begin{aligned}(g \circ f)(\lambda\, a) &= g\big(f(\lambda\, a)\big)\\ &= g\big(\lambda\, f(a)\big)\\ &= \lambda\, g(f(a))\\ &= \lambda\, (g \circ f)(a).\end{aligned} \tag{5.52}$$

5.3.2 Matrixdarstellung

Jede lineare Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m$ kann als Matrix interpretiert werden. Um dies zu sehen, führen wir zunächst die Vektoren

$$e_k := \begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix} \tag{5.53}$$

$k = 1, \ldots, n$ ein. Diese sind bis auf eine Komponenten überall Null und der Vektor e_k hat als k-te Komponente eine Eins. Mit diesen Vektoren erhalten wir nun für jedes $x \in \mathbb{R}^n$:

$$f(x) = f\left(\sum_{k=1}^{n} x_k\, e_k\right) = \sum_{k=1}^{n} f(x_k\, e_k) = \sum_{k=1}^{n} x_k\, f(e_k). \tag{5.54}$$

Beachte hier, das die x_k die Komponenten des Vektors x sind (also Zahlen), die e_k jedoch selbst Vektoren sind. Schreiben wir die Vektoren $f(e_k) \in \mathbb{R}^m$ als Spalten in eine Matrix, so erhalten wir folgenden nützlichen Zusammenhang (vgl. dazu auch Section 5.1.6):

Merke

Die zu einer linearen Abbildung $f : \mathbb{R}^n \to \mathbb{R}^m$ definierte Matrix

$$A := \begin{bmatrix} | & & | \\ f(e_1) & \cdots & f(e_n) \\ | & & | \end{bmatrix} \in \mathbb{R}^{m \times n} \tag{5.55}$$

heißt **Matrixdarstellung** von f. Es gilt stets

$$f(x) = A\,x. \tag{5.56}$$

Jede lineare Abbildung kann also als Matrix-Vektor-Produkt geschrieben werden. Umgekehrt ist für eine beliebige Matrix $A \in \mathbb{R}^{m \times n}$ die Abbildung $x \mapsto A\,x$ stets linear. Der Wert von Matrizen liegt ganz wesenlich in der Beschreibung beliebiger linearer Abbildung durch Zahlen. Und diese Zahlen kann man sehr einfach und explizit erhalten indem man die betrachtete lineare Abbildung auf die Vektoren $e_1, \ldots, e_n$ anwendet. Die Darstellung als Zahlen (Matrix) ermöglicht ein sehr bequemes Arbeiten mit linearen Abbildungen am Computer. Matrix-Vektor-Produkte sind ohne Weiteres am Computer umsetzbar.

Beispiel (Matrixdarstellung zur Drehung in der Ebene)

Ist $f : \mathbb{R}^2 \to \mathbb{R}^2$ die Drehung in der Ebene um einen Winkel $\alpha \in \mathbb{R}$ um den Koordinatenursprung, so überlegt man sich leicht aus den Definitionen von Sinus und Cosinus, dass

$$f(e_1) = \begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}, \qquad f(e_2) = \begin{bmatrix} -\sin\alpha \\ \cos\alpha \end{bmatrix} \tag{5.57}$$

gilt. Somit

$$f(x, y) = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{5.58}$$

Eine Menge von Punkten, z.B. Eckpunkten eines Vielecks, lässt sich damit mühelos drehen:

```
import numpy as np
import matplotlib.pyplot as plt

points = np.array([  # ein Punkt pro Zeile
```

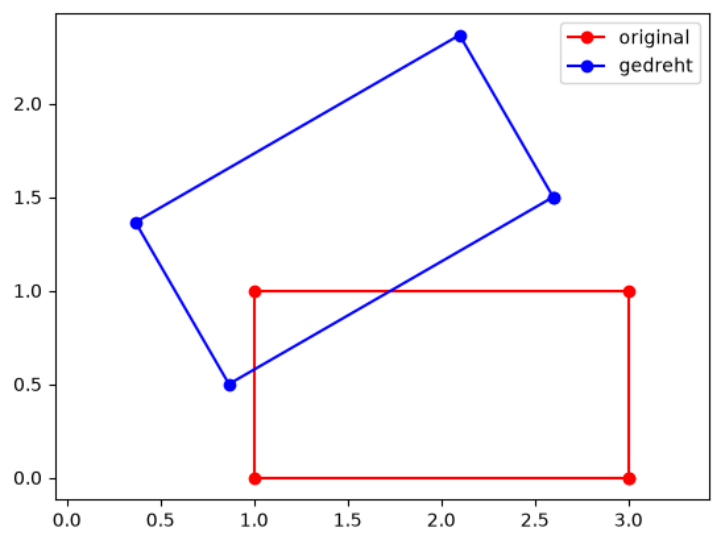
```
    [3, 0],
    [3, 1],
    [1, 1],
    [1, 0],
    [3, 0]
])

angle = 30 * np.pi / 180  # Drehwinkel

A = np.array([  # Drehmatrix
    [np.cos(angle), -np.sin(angle)],
    [np.sin(angle),  np.cos(angle)]
])

rotated = (A @ points.T).T
# points.T enthält pro Spalte einen Punkt. Das Produkt enthält
↪  dann in jeder
# Spalte den entsprechenden gedrehten Punkt. Nochmaliges
↪  Transponieren liefert
# ein Array mit den gedrehten Punkten in den Zeilen.

fig, ax = plt.subplots()
ax.plot(points[:, 0], points[:, 1], 'o -r', label='original')
ax.plot(rotated[:, 0], rotated[:, 1], 'o -b', label='gedreht')
ax.legend()
ax.axis('equal')  # gleiche Achseneinteilung (keine Verzerrung)
plt.show()
```



Beispiel (Matrixdarstellung zur Spiegelung an der x-y-Ebene)

Ist $f : \mathbb{R}^3 \to \mathbb{R}^3$ die Spiegelung an der x-y-Ebene, so gilt offensichtlich

$$f(e_1) = e_1, \qquad f(e_2) = e_2, \qquad f(e_3) = -e_3. \tag{5.59}$$

Somit

$$f(x, y, z) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \tag{5.60}$$

Im Laufe der Veranstalung werden uns weitere, auch deutlich komplexere, lineare Abbildungen begegnen.

5.3.3 Komposition vs. Matrixmultiplikation

Sind $f : \mathbb{R}^n \to \mathbb{R}^p$ und $g : \mathbb{R}^p \to \mathbb{R}^m$ lineare Abbildung und sind $A \in \mathbb{R}^{n \times p}$ und $B \in \mathbb{R}^{p \times m}$ die zugehörigen Matrixdarstellungen, so gilt offensichtlich

$$(g \circ f)(x) = g(f(x)) = g(A\,x) = B\,(A\,x) = (B\,A)\,x. \tag{5.61}$$

In Worten: **Die Matrixdarstellung der Komposition (Hintereinanderausführung) zweier linearer Abbildungen ist gerade das Produkt der Matrixdarstellungen der beiden Abbildungen.**

Damit sind Sinn und Zweck der zunächst etwas seltsam amutenden Definition des Produkts zweier Matrizen geklärt: Das Matrixprodukt wurde gerade so gewählt, dass das

Multiplizieren von Matrizen dem Hintereinanderausführen der entsprechenden linearen Abbildungen entspricht.

Diese Erkenntnis hilft einerseits beim Finden von Matrixdarstellungen für komplexere Abbildungen (erst Spiegeln, dann Drehen, dann Strecken usw.); andererseits kann man damit auch recht einfach zu diversen Standardresultaten kommen, die sonst nur mühsam herleitbar sind:

Beispiel (Formel für den doppelten Winkel)

Sei $\alpha \in \mathbb{R}$ ein beliebiger Winkel, sei $A \in \mathbb{R}^{2\times 2}$ die Matrixdarstellung der Drehung um diesen Winkel und sei $B \in \mathbb{R}^{2\times 2}$ die Matrixdarstellung der Drehung um den Winkel $2\,\alpha$, d.h.

$$A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \qquad \text{und} \qquad B = \begin{bmatrix} \cos(2\,\alpha) & -\sin(2\,\alpha) \\ \sin(2\,\alpha) & \cos(2\,\alpha) \end{bmatrix}. \tag{5.62}$$

Die Drehung um den doppelten Winkel kann als Komposition von zwei einfachen Drehungen angesehen werden. Somit gilt $B = A\,A$, also

$$\begin{bmatrix} \cos(2\,\alpha) & -\sin(2\,\alpha) \\ \sin(2\,\alpha) & \cos(2\,\alpha) \end{bmatrix} = \begin{bmatrix} (\cos\alpha)^2 - (\sin\alpha)^2 & -2\,(\cos\alpha)\,\sin\alpha \\ 2\,(\cos\alpha)\,\sin\alpha & (\cos\alpha)^2 - (\sin\alpha)^2 \end{bmatrix}. \tag{5.63}$$

Wir erhalten damit die bekannten Formeln

$$\cos(2\,\alpha) = (\cos\alpha)^2 - (\sin\alpha)^2 \tag{5.64}$$

und

$$\sin(2\,\alpha) = 2\,(\cos\alpha)\,\sin\alpha. \tag{5.65}$$

für Cosinus und Sinus des doppelten Winkels.

5.4 Lineare Unterräume und Untermannigfaltigkeiten

Die hier behandelten Begriffe mögen etwas abstrakt erscheinen, sind uns aber schon in der Schulmathematik in Form von Geraden und Ebenen begegnet. Auch beim Lösen von linearen Gleichungssystemen spielen diese Begriffe eine Rolle.

5.4.1 Unterräume

Ein Unterraum ist eine Teilmenge von $\mathbb{R}^n$, in der wir die üblichen Vektoroperationen (Addition und skalare Vielfache) ausführen können ohne dabei diese Teilmenge zu verlassen. Genauer:

Merke

Eine Teilmenge $U \subseteq \mathbb{R}^n$ heißt **linearer Unterraum**, wenn gilt:

- $a + b \in U$ für alle $a, b \in U$,
- $\lambda\, a \in U$ für alle $a \in U$ und alle $\lambda \in \mathbb{R}$.

Offensichtlich ist $\{0\}$ (Menge, die nur den Nullvektor enthält) ein Unterraum und auch $\mathbb{R}^n$ selbst ist ein Unterraum.

Beispiel (Gerade in der Ebene)

Die Menge $U := \{(x, y) \in \mathbb{R}^2 : \; y = 3\,x\}$ ("Gerade mit Anstieg 3") ist ein Unterraum von $\mathbb{R}^2$, denn:

- Für $(x, y) \in U$ und $(\tilde{x}, \tilde{y}) \in U$ gilt

$$y + \tilde{y} = 3\,x + 3\,\tilde{x} = 3\,(x + \tilde{x}), \tag{5.66}$$

  also $(x + \tilde{x}, y + \tilde{y}) \in U$.
- Für $(x, y) \in U$ und $\lambda \in \mathbb{R}$ gilt

$$\lambda\, y = \lambda\,(3\,x) = 3\,(\lambda\, x), \tag{5.67}$$

  also $(\lambda\, x, \lambda\, y) \in U$.

Der Wert von Unterräumen besteht darin, dass wir in ihnen mit Vektoren arbeiten können als würde es aushalb keine weiteren Vektoren geben. Arbeitet man beispielweise im $\mathbb{R}^3$ nur mit Vektoren (oder Punkten), die alle in einer Ebene liegen, so unterscheidet sich das Vorgehen praktisch nicht vom Arbeiten in $\mathbb{R}^2$. Wie genau dieser Übergang zwischen Unterräumen und den entsprechenden "ganzen" Räumen vollzogen werden kann,

ist Thema der nächsten Abschnitte.

5.4.2 Lineare Unabhängigkeit

Wenn wir zwei Vektoren $a, b \in \mathbb{R}^n \setminus \{0\}$ betrachten, dann können diese Vielfache voneinander sein (also in die gleiche Richtung bzw. Gegenrichtung zeigen) oder nicht. In letzterem Fall sagt man, dass die beiden Vektoren linear unabhängig sind. Nimmt man einen dritten Vektor $c \in \mathbb{R}^n \setminus \{0\}$ hinzu, so liegt dieser wahlweise in der von den ersten beiden Vektoren aufgespannten Ebene oder nicht. In letzterem Fall sagt man wieder, dass die drei Vektoren linear unabhängig sind. In ersterem Fall können wir c als Linearkombination

$$c = \alpha\, a + \beta\, b \tag{5.68}$$

mit $\alpha, \beta \in \mathbb{R}$ darstellen. Anders ausgedrückt: Wir finden Zahlen $\alpha, \beta, \gamma \in \mathbb{R}$, sodass

$$\alpha\, a + \beta\, b + \gamma\, c = 0 \tag{5.69}$$

gilt, obwohl die drei Zahlen nicht alle Null sind (speziell ist hier $\gamma = -1$). Allgemeiner definieren wie lineare Unabhängigkeit wie folgt:

Merke

Eine Menge $\{a_1, \ldots, a_r\} \subseteq \mathbb{R}^n$ von r Vektoren in $\mathbb{R}^n$ heißt **linear unabhängig**, wenn für alle Zahlen $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ gilt:

$$\sum_{k=1}^{r} \lambda_k\, a_k = 0 \quad \Rightarrow \quad \lambda_1 = \cdots = \lambda_r = 0. \tag{5.70}$$

Andernfalls heißt die Menge linear abhängig.

Enthält die Menge $\{a_1, \ldots, a_r\}$ den Nullvektor, so ist sie offensichtlich stets linear abhängig (das entsprechende λ_k kann beliebig gewählt werden).

Ist die Menge $\{a_1, \ldots, a_r\}$ linear abhängig, so gibt es ein $l \in \{1, \ldots, r\}$, sodass $\lambda_l \neq 0$, aber trotzdem die Linearkombination $\sum_{k=1}^{r} \lambda_k\, a_k$ den Nullvektor liefert. Umgestellt gilt

$$a_l = \sum_{\substack{k=1 \\ k \neq l}}^{r} \frac{\lambda_k}{\lambda_l}\, a_k, \tag{5.71}$$

d.h. a_l lässt sich als Linearkombination der anderen Vektoren darstellen.

Umgekehrt gilt: Lässt sich ein Vektor als Linearkombination anderer Vektoren darstellen, so ist die entsprechende Menge von Vektoren linear abhängig (wähle $\lambda_l = -1$ wie oben für $\mathbb{R}^3$ beschrieben).

5.4.3 Basis und Dimension

Man kann nun der Frage nachgehen, wie viele linear unabhängige Vektoren man eigentlich finden kann. In der Ebene stellt man schnell fest, dass höchstens zwei Vektoren linear unabhängig sein können. Mengen mit drei oder mehr Vektoren sind stets linear abhängig. Entsprechend findet man im Raum nur höchsten drei linear unabhängige Vektoren. Wir können lineares Unabhängigkeit also als Grundlage für die präzise Formulierung des intuitiv klaren Begriffs "Dimension" verwenden.

Merke

Die Dimension eines linearen Unterraums U des $\mathbb{R}^n$ ist die maximale Anzahl linear unabhängiger Vektoren in U. Die Dimension von U wird mit dem Symbol $\dim U$ abgekürzt.

Der triviale Unterraum $\{0\}$ hat Dimension Null: Es gibt nur einen Vektor, den Nullvektor, und Mengen, die den Nullvektor enthalten, sind stets linear abhängig.

Die Dimension von $\mathbb{R}^n$ ist n: Die Menge $\{e_1, \ldots, e_n\}$ (wobei e_k in der k-ten Komponenten eins ist und sonst null) ist linear unabhängig und jeder weitere Vektor $x \in \mathbb{R}^n$ kann als Linearkombination $x = \sum_{k=1}^{n} x_k\, e_k$ geschrieben werden, führt also zu linearer Abhängigkeit.

Offensichtlich gilt für jeden Unterraum U von $\mathbb{R}^n$, dass $\dim U \in \{0, 1, \ldots, n\}$.

Merke

Sei $U \subseteq \mathbb{R}^n$ ein Unterraum der Dimension $r := \dim U$. Jede Menge $\{b_1, \ldots, b_r\} \subseteq U$ von r linear unabhängigen Vektoren aus U heißt **Basis** von U.

Beispielsweise bilden die Vektoren $e_1, \ldots, e_n$ eine Basis in $\mathbb{R}^n$. Es gibt jedoch auch noch andere Basen. Mehr dazu später.

5.4.4 Koordinaten bezüglich einer Basis

Merke

Sei $U \subseteq \mathbb{R}^n$ und sei $\{b_1, \ldots, b_r\} \subseteq U$ eine Basis in U. Dann existieren zu jedem $x \in U$ eindeutig bestimmte Zahlen $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$, sodass

$$x = \sum_{k=1}^{r} \lambda_k\, b_k \tag{5.72}$$

gilt. Diese Zahlen heißen **Koordinaten von x bezüglich der Basis** $\{b_1, \ldots, b_r\}$.

Existenz und Eindeutigkeit der Koordinaten bezüglich einer Basis können leicht nachgewiesen werden. Da diese Überlegungen für uns aber keine neue Erkenntnisse bringen, verzichten wir darauf.

Offensichtlich sind die Komponenten $x_1, \ldots, x_n$ eines Vektors $x \in \mathbb{R}^n$ gerade die Koordinates bezüglich der Basis $\{e_1, \ldots, e_n\}$, denn

$$x = \sum_{k=1}^{n} x_k \, e_k. \tag{5.73}$$

Beispiel (alternative Basis in der Ebene)

In $\mathbb{R}^2$ ist durch die beiden Vektoren

$$b_1 := \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad b_2 := \begin{bmatrix} -1 \\ 1 \end{bmatrix} \tag{5.74}$$

eine Basis gegeben. Die Koordinaten des Vektors $\begin{bmatrix} 0 & 1 \end{bmatrix}^{\mathrm{T}}$ bezüglich dieser Basis sind $\frac{1}{2}$ und $\frac{1}{2}$, denn

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \tag{5.75}$$

Für einen beliebigen Vektor $x \in \mathbb{R}^2$ erhält man bezüglich der Basis $\{b_1, b_2\}$ die Koordinaten

$$\frac{x_1 + x_2}{2} \qquad \text{und} \qquad \frac{x_2 - x_1}{2}, \tag{5.76}$$

denn

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{x_1 + x_2}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{x_2 - x_1}{2} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.. \tag{5.77}$$

Wie man systematisch die Koordinaten eines Vektors bezüglich einer beliebigen gegebenen Matrix berechnet, klären wir später.

Beispiel (Basis in einem nichttrivialen Unterraum)

Betrachten in $\mathbb{R}^2$ den Unterraum

$$U := \{x \in \mathbb{R}^2 : \; x_2 = 3\, x_1\}. \tag{5.78}$$

Dies ist ein eindimensionaler Unterraum und $\{b_1\}$ mit

$$b_1 := \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{5.79}$$

ist eine Basis darin.

Wir können nun jeden Vektor $x \in U$ durch seine Koordinate (genau eine Zahl!) bezüglich der Basis $\{b_1\}$ darstellen. Diese Koordinate ist gerade x_1, denn

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{5.80}$$

(beachte, dass $x_2 = 3\,x_1$ für jedes $x \in U$ gilt).

Solange wir nur Betrachtungen in U anstellen, können wir statt Vektoren mit zwei Komponenten also einfach mit einkomponentigen Vektoren (d.h. Zahlen) arbeiten. Die Arbeit im eindimensionalen Unterraum U unterscheidet sich also nicht von der Arbeit in $\mathbb{R}$.

Die Tatsache, dass man bei der Arbeit in Unterräumen mit (teils deutlich) "kürzeren" Vektoren arbeiten kann, macht man sich in den Datenwissenschaften (Data Science) und im maschinellen Lernen zu nutze. Zu analysierende Daten "leben" oft in einem sehr hochdimensionalen Grundraum (Bilddaten mit 1000x1000 Pixeln z.B. in $\mathbb{R}^{1000000}$), während die eigentlich relevanten Datenpunkte sich in einem vergleichsweise niedrigdimensionalen Unterrraum befinden. Durch Wahl geeigneter Basen können Ressourcen (Speicher, Rechenzeit) gespart werden.

5.4.5 Aufspannung

Oben hatten wir uns mit Basen in einem gegebenen Unterraum beschäftigt. Umkehrt steht man manchmal vor der Frage, was denn eigentlich der kleinste Unterraum ist, der eine gegebene Menge von Vektoren enthält.

Merke

Sei $\{b_1, \dots, b_r\} \subseteq \mathbb{R}^n$ eine Menge von r linear unabhängigen Vektoren in $\mathbb{R}^n$. Die Menge

$$\operatorname{span}\{b_1, \dots, b_r\} := \left\{ \sum_{k=1}^{r} \lambda_k\, b_k : \ \lambda \in \mathbb{R}^r \right\} \tag{5.81}$$

heißt **Aufspannung** von $\{b_1, \dots, b_r\} \subseteq \mathbb{R}^n$.

Folgende Eigenschaften von $\operatorname{span}\{b_1, \dots, b_r\}$ sieht man leicht ein:

- Die Aufspannung ist ein linearer Unterraum.
- Die Menge $\{b_1, \dots, b_r\}$ ist eine Basis in ihrer Aufspannung.
- Die Aufspannung ist der kleinste Unterraum, der $\{b_1, \dots, b_r\}$ enthält, d.h. jeder Unterraum, der $\{b_1, \dots, b_r\}$ enthält, enthält automatisch auch $\operatorname{span}\{b_1, \dots, b_r\}$.

Nebenbemerkung

Beachte, dass manche Quellen auch Aufspannungen für Mengen von Vektoren betrachten, die nicht linear unabhängig sind. Dann ist diese Menge keine Basis mehr in ihrer Aufspannung. Die anderen beiden Eigenschaften gelten aber weiterhin.

In $\mathbb{R}^3$ ist die Aufspannung von zwei linear unabhängigen Vektoren eine Ebene. In $\mathbb{R}^2$ und $\mathbb{R}^3$ ist die Aufspannung eines Vektors (außer Nullvektor) eine Gerade. Zwei linear unabhängige Vektoren in $\mathbb{R}^2$ spannen den gesamten $\mathbb{R}^2$ auf, usw.

5.4.6 Orthonormalbasen

Meist ist man an besonders schönen, d.h. leicht zu handhabenden Basen, interessiert. Führen hier eine Klasse besonders schöner Basen und einige Begriffe aus deren Umfeld ein.

Merke

Ein Menge $\{a_1, \ldots, a_r\} \subseteq \mathbb{R}^n$ von r linear unabhängigen Vektoren heißt **Orthogonalsystem**, wenn

$$a_k \circ a_l = 0 \quad \text{für } k \neq l. \tag{5.82}$$

gilt. Eine solche Menge heißt **Orthonormalsystem** (ONS), wenn zusätzlich

$$\|a_k\| = 1 \quad \text{für } k = 1, \ldots, r \tag{5.83}$$

gilt.

Ist ein ONS eine Basis für einen Unterraum $U \subseteq \mathbb{R}^n$, so bezeichnen wir dieses ONS als **Orthonormalbasis** (ONB) für U.

Die Bedingung $a_k \circ a_l = 0$ besagt, dass die Basisvektoren senkrecht zueinander stehen. Zusammen mit der Normierung $\|a_k\| = 1$ kann man die definierende Eigenschaft eines ONS auch als

$$a_k \circ a_l = \begin{cases} 0, & \text{falls } k \neq l, \\ 1, & \text{sonst} \end{cases} \tag{5.84}$$

ausdrücken.

Offensichtlich ist die Standardbasis $e_1, \ldots, e_n$ eine ONB.

5.4.7 Untermannigfaltigkeiten

Unterräume enthalten stets den Koordinatenursprung. Für mehr Flexibilität (beim Modellieren und beim Formulieren in der Sprache der Mathematik) führt man noch "verschobene" Unterräume ein:

Merke

Eine **lineare Untermannigfaltigkeit** ist eine Menge der Form

$$x + U := \{x + u : u \in U\} \subseteq \mathbb{R}^n, \tag{5.85}$$

wobei $x \in \mathbb{R}^n$ ein fester Vektor und $U \subseteq \mathbb{R}^n$ ein linearer Unterraum ist.

Jede Wahl eines Vektors x und eines Unterraums U liefert also eine Mannigfaltigkeit. Allerdings können verschiedenen Kombination zur gleichen Untermannigfaltigkeit führen. Insbesondere erhält man bei $x \in U$ wieder U; jeder Unterraum ist also selbst eine Untermannigfaltigkeit.

Wählt man zwei Vektoren a, b aus einer Untermannigfaltigkeit $x + U$, so liegt deren Differenz $a-b$ stets in U. (Begründung: Es existieren $u, v \in U$ mit $a = x+u$ und $b = x+v$. Somit $a - b = x + u - (x + v) = u - v$. Da skalare Vielfache und Summen von Elementen eines Unterraums wieder im Unterraum liegen, folgt $a - b \in U$.) In diesem Sinne kann man aus einer Untermannigfaltigkeit auch ohne Kenntnis des Verschiebungsvektors x den zugrundeliegenden Unterraum zurückerhalten.

Merke

Als **Dimension** einer Untermannigfaltigkeit bezeichnen wir die Dimension des zugrundeliegenden Unterraums.

5.4.8 Geraden

Hatten schon gelegentlich Geraden als Unterräume des $\mathbb{R}^2$ betrachtet. Natürlich besitzen auch $\mathbb{R}^3$ und höherdimensionale Räume eindimensionale Unterräume. Diese werden allerdings nur in $\mathbb{R}^2$ und $\mathbb{R}^3$ als Geraden bezeichnet. Eindimensionale Untermannigfaltigkeiten werden in $\mathbb{R}^2$ und $\mathbb{R}^3$ ebenfalls als Geraden bezeichnet.

Eine Gerade ist durch einen Aufpunkt (bzw. dessen Ortsvektor) und einen Richtungsvektor bestimmt. Allgemein können wir jede eindimensionale Untermannigfaltigkeit in der Form

$$\{a + t\, b : t \in \mathbb{R}\} \tag{5.86}$$

schreiben, wobei $a, b \in \mathbb{R}^n$ fest sind. Handelt es sich um einen Unterraum, so ist $\{b\}$ eine Basis darin und t ist gerade die Koordinate des entsprechenden Vektors bezüglich dieser Basis.

5.4.9 Ebenen

Ebenen sind zweidimensionale Untermannigfaltigkeiten in $\mathbb{R}^3$. Gelegentlich sieht man den Begriff “Hyperebene”. Dabei handelt es sich nicht (!) um eine zweidimensionale

Untermannigfaltigkeit in $\mathbb{R}^n$, sondern um eine Untermannigfaltigkeit mit Dimension $n-1$.

Eine Ebene ist durch einen Aufpunkt (bzw. dessen Ortsvektor) und zwei linear unabhängige (!) Richtungsvektoren bestimmt. Allgemein können wir jede zweidimensionale Untermannigfaltigkeit in der Form

$$\{a + s\,b + t\,c :\ s, t \in \mathbb{R}\} \tag{5.87}$$

schreiben, wobei $a, b, c \in \mathbb{R}^n$ fest sind. Handelt es sich um einen Unterraum, so ist $\{b, c\}$ eine Basis darin und s, t sind gerade die Koordinaten des entsprechenden Vektors bezüglich dieser Basis.

5.5 Lineare Gleichungssysteme

Lineare Gleichungssysteme sollten im Wesentlichen aus der Schulmathematik bekannt sein. Hier konzentrieren wir uns auf die Struktur der Lösungensmengen und auf Aspekte beim Lösen mit dem Computer.

Da lineare Gleichungssysteme eine der Aufgabenstellungen sind, die besonders gut mit dem Computer behandelt werden können, bemüht man sich in allen Teilgebieten von Naturwissenschaften, Technik und Wirtschaftswissenschaften zu lösenden praktische Probleme in Form linearer Gleichungssysteme zu fomulieren. Zusätzlich tauchen lineare Gleichungssysteme auch zahlreich als Teilproblem in komplexeren Modellen und zugehörigen Lösungsverfahren auf.

5.5.1 Matrix-Vektor-Schreibweise

Ein lineares Gleichungssystem (kurz: LGS) mit m Gleichungen und n Unbekannten hat die Form

$$\begin{aligned} a_{11}\,x_1 + \cdots + a_{1n}\,x_n &= b_1 \\ &\vdots \\ a_{m1}\,x_1 + \cdots + a_{mn}\,x_n &= b_m. \end{aligned} \tag{5.88}$$

Die a_{ij} und die b_i sind gegeben. Die x_j sind gesucht.

Setzt man

$$A := \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \quad b := \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}, \quad x := \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \tag{5.89}$$

so kann man das LGS deutlich kompakter und handhabbarer mit einem Matrix-Vektor-Produkt als

$$A\,x = b \tag{5.90}$$

schreiben. Die Matrix A wird als **Systemmatrix** bezeichnet, der gegebene Vektor b als **rechte Seite**.

5.5.2 Struktur der Lösungsmenge

5.5.2.1 Anzahl der Lösungen LGS können keine, genau eine oder unendlich viele Lösungen besitzen. Ein einfaches Beispiel ohne Lösung ist

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}. \tag{5.91}$$

Das LGS

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}. \tag{5.92}$$

besitzt hingegen genau eine Lösung.

Die Lösungsmenge eines LGS mit unendlich vielen Lösungen besitzt eine sehr konkrete und nützliche Struktur. Zur weiteren Untersuchung unterscheiden wir zwei Typen von LGS:

Merke

Ein LGS heißt **homogen**, wenn die rechte Seite der Nullvektor ist. Andernfalls heißt das LGS **inhomogen**.

5.5.2.2 Homogene LGS Die Lösungsmenge eines homogenen LGS mit n Unbekannten ist offensichtlich ein linearer Unterraum des $\mathbb{R}^n$, denn die Summe zweier Lösungen ist wieder eine Lösung und alle Vielfachen von Lösungen sind Lösungen. In Formeln:

$$A\,x = 0,\ A\,\tilde{x} = 0 \quad \Rightarrow \quad A\,(x + \tilde{x}) = A\,x + A\,\tilde{x} = 0 + 0 = 0 \tag{5.93}$$

bzw.

$$A\,x = 0,\ \lambda \in \mathbb{R} \quad \Rightarrow \quad A\,(\lambda\,x) = \lambda\,A\,x = \lambda \cdot 0 = 0. \tag{5.94}$$

Somit kann es beispielsweise auch kein homogenes LGS gegeben, welches genau zwei Lösungen besitzt, da dann automatisch die Summe der beiden Lösungen eine dritte Lösung liefern würde.

Beachte, dass der Nullvektor eine Lösung jedes homogenen LGS ist. Falls dies die einzige Lösung ist, ist die Lösungsmenge gerade der triviale Unterraum $\{0\}$. Das andere Extrem ist ein homogenes LGS mit Nullmatrix (nur Nullen) als Systemmatrix. Dann ist die Lösungsmenge ganz $\mathbb{R}^n$.

5.5.2.3 Inhomogene LGS Bei einem inhomogenen LGS ist die Lösungsmenge stets eine Untermannigfaltigkeit oder die leere Menge. Hat ein inhomogenes LGS eine Lösung $x^* \in \mathbb{R}^n$ (und ggf. noch mehr), so können wir das LGS als

$$A\,x = A\,x^* \tag{5.95}$$

schreiben. Setzen wir nun $\tilde{x} := x - x^*$, so ist dies äquivalent zu dem homogenen LGS

$$A\,\tilde{x} = 0 \tag{5.96}$$

(alles nach links bringen und A ausklammern). Dessen Lösungsmenge ist ein Unterraum und aus der Beziehung $x = x^* + \tilde{x}$ sehen wir, dass die Lösungsmenge des ursprünglichen inhomogenen LGS eine Untermannigfaltigkeit ist. **Die Lösungsmenge eines inhomogenen LGS ist stets eine verschobene Version der Lösungsmenge des homogene LGS mit identischer Systemmatrix!**

5.5.2.4 Geometrische Interpretation Visualisiert man die Lösungsmenge eines LGS mit zwei oder drei Unbekannten, so erhält man wahlweise die leere Menge, einen Punkt, eine Gerade, eine Ebene oder den ganzen Raum. Entsprechend eng verknüpft sind Geraden und Ebenen mit linearen Gleichungssystemen.

Im $\mathbb{R}^2$ korrespondieren Geraden mit den Lösungsmengen von LGS mit zwei Unbekannten und einer Gleichung:

$$a_{11}\,x_1 + a_{12}\,x_2 = b_1. \tag{5.97}$$

Dabei dürfen a_{11} und a_{12} nicht gleichzeitig Null sein (warum?).

Im $\mathbb{R}^3$ korrespondieren Ebenen mit den Lösungsmengen von LGS mit drei Unbekannten und einer Gleichung:

$$a_{11}\,x_1 + a_{12}\,x_2 + a_{13}\,x_3 = b_1. \tag{5.98}$$

Dabei dürfen wieder a_{11}, a_{12}, a_{13} nicht gleichzeitig Null sein.

Nebenbemerkung (Normalenvektor)

Liegt $(x_1^*, x_2^*, x_3^*) \in \mathbb{R}^3$ in der betrachteten Ebene, d.h. es gilt

$$a_{11}\,x_1^* + a_{12}\,x_2^* + a_{13}\,x_3^* = b_1, \tag{5.99}$$

so folgt für jeden anderen Punkt $(x_1, x_2, x_3) \in \mathbb{R}^3$ in der Ebene

$$a_{11}\,(x_1 - x_1^*) + a_{12}\,(x_2 - x_1^*) + a_{13}\,(x_3 - x_1^*) = 0. \tag{5.100}$$

Dies kann mittels Skalarprodukt als

$$\begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} \circ \begin{bmatrix} x_1 - x_1^* \\ x_2 - x_2^* \\ x_3 - x_3^* \end{bmatrix} = 0 \tag{5.101}$$

ausgedrückt werden. Der Vektor $\begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix}^{\mathrm{T}}$ ist also senkrecht zu jedem "in der Ebene liegenden" Vektor (genauer: senkrecht zu jedem zur Ebene parallel verlaufenden Vektor). Vektoren mit dieser Eigenschaft werden als **Normalenvektor** der Ebene bezeichnet. Man kann zeigen, dass alle Normalenvektoren einer Ebene in $\mathbb{R}^3$ Vielfache voneinander sind (anschaulich ist das klar).

Die Beschreibung einer Ebene als Lösungsmenge der Gleichung $a_{11}\,x_1 + a_{12}\,x_2 = b_1$ wird auch als *Normalform* der Ebenen bezeichnet. Hat der Vektor $\begin{bmatrix} a_{11} & a_{12} & a_{13} \end{bmatrix}^{\mathrm{T}}$ die Länge 1, so spricht man auch von der **Hesse'schen Normalform**. Dieser Begriff sollte aus der Schulmathematik bekannt sein.

Geraden im $\mathbb{R}^3$ entsprechen den Lösungsmengen von LGS mit drei Unbekannten und zwei Gleichungen:

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + a_{13}\,x_3 &= b_1 \\ a_{21}\,x_1 + a_{22}\,x_2 + a_{23}\,x_3 &= b_2. \end{aligned} \tag{5.102}$$

Die Lösungsmenge dieses LGS ist der Durchschnitt der Lösungsmenge der ersten Gleichung und der Lösungsmenge der zweiten Gleichung, also die Schnittgerade der beiden durch die beiden Gleichungen beschriebenen Ebenen. Sind diese Ebenen parallel oder identisch, so erhält man keine Gerade als Lösungsmenge des LGS.

Völlig analog kann man sich überlegen, dass ein Punkt in $\mathbb{R}^2$ als Schnitt zweier (nicht paralleler oder identischer) Geraden entsteht, oder, alternativ, als Lösung eines LGS mit zwei Unbekannten und zwei Gleichungen.

5.5.3 Manuelles Lösen

LGS löst man praktisch immer mit dem Computer (siehe unten). Dennoch sollen hier kurz zwei Ideen skizziert werden, wie man ein (kleines) LGS manuell lösen kann. Das manuelle Vorgehen ist insofern relevant als dass der Computer im Wesentlichen genauso vorgeht (aber dabei viel schneller ist). Auch begegnen uns hier einige neue Begriffe, die später in anderem Kontext benötigt werden. Praktisch relevante LGS besitzen heute tausende Unbekannte und Gleichungen, sodass manuelles Lösen nicht möglich ist.

5.5.3.1 Gauß-Algorithmus Die Grundidee des als Gauß-Algorithmus bekannten Vorgehens ist das systematische Vereinfachen des Gleichungssystems ohne dabei die Lösungsmenge zu verändern. Folgende Operationen sind dabei hilfreich und ändern die Lösungsmenge nicht:

- Vertauschen der Reihenfolge der Gleichungen,
- Multiplizieren einer Gleichung mit einem Faktor (außer Null),
- ein Vielfaches einer Gleichung zu einer anderen Gleichung hinzuaddieren,
- Gleichungen streichen, bei denen alle Koeffizienten und die rechte Seite Null sind.

Mit diesen Operationen kann man jedes LGS auf **Stufenform** bringen, d.h. die Anzahl der Unbekannten mit Nicht-Null-Koeffizient wird von Gleichung zu Gleichung kleiner. Dabei richtet man den Ablauf so ein, dass die Unbekannten in der natürlichen Reihenfolge wegfallen. So taucht x_1 dann nur noch in der ersten (oder keiner) Gleichung auf; x_2 erscheint höchstens in den ersten beiden Gleichungen usw. Welche Operation wann und wie anzuwenden ist um eine Stufenform des LGS zu erhalten, folgt einem einfachen Schema, welches man ohne nenneswerte Denkarbeit abarbeiten kann (für solche Aufgaben wurde der Computer erfunden...).

Ist das LGS in Stufenform, so kann man mit der letzten Gleichung beginnend die Lösung Gleichung für Gleichung leicht zusammensetzen. Dieser Vorgang wird als **Rückwärtseinsetzen** bezeichnet und folgt ebenfalls wieder einem klaren Schema ohne Notwendigkeit des Nachdenkens.

Da wir den Gauß-Algorithmus dem Computer überlassen, verzichten wir hier auf Details und Sonderfälle und belasses es bei einigen Beispielen:

Beispiel (Gauß-Algorithmus bei genau einer Lösung)

Das LGS

$$\begin{aligned} 2\,x_1 + x_2 + x_3 &= 0 \\ 2\,x_1 - x_2 + 9\,x_3 &= 6 \\ 2\,x_1 + 2\,x_2 - x_3 &= 1 \end{aligned} \tag{5.103}$$

ist äquivalent zu folgendem LGS in Stufenform:

$$\begin{aligned} 2\,x_1 + x_2 + x_3 &= 0 \\ -2\,x_2 + 8\,x_3 &= 6 \\ 2\,x_3 &= 4. \end{aligned} \tag{5.104}$$

Rückwärtseinsetzen liefert

$$x = \begin{bmatrix} -\frac{7}{2} \\ 5 \\ 2 \end{bmatrix} \tag{5.105}$$

(Video 530, Notizen dazu).

Beispiel (Gauß-Algorithmus bei Nichtlösbarkeit)

Das LGS

$$\begin{aligned} 2\,x_1 + x_2 + x_3 &= 0 \\ 3\,x_1 - \frac{3}{2}\,x_2 + 9\,x_3 &= 6 \\ 2\,x_1 + 2\,x_2 - \frac{3}{2}\,x_3 &= 1 \end{aligned} \tag{5.106}$$

ist äquivalent zu folgendem LGS in Stufenform:

$$\begin{aligned} 2\,x_1 + x_2 + x_3 &= 0 \\ x_2 - \frac{5}{2}\,x_3 &= 1 \\ 0 &= 9. \end{aligned} \tag{5.107}$$

Dieses LGS besitzt aufgrund der letzten Gleichung offensichtlich keine Lösung (Video 535, Notizen dazu).

Beispiel (Gauß-Algorithmus bei unendlich vielen Lösungen)

Das LGS

$$\begin{aligned} 2\,x_1 + x_2 + x_3 &= 0 \\ -6\,x_1 + 3\,x_2 + 9\,x_3 &= 6 \\ 2\,x_1 + 2\,x_2 + 3\,x_3 &= 1 \end{aligned} \tag{5.108}$$

ist äquivalent zu folgendem LGS in Stufenform:

$$\begin{aligned} 2\,x_1 + x_2 + x_3 &= 0 \\ 6\,x_2 + 12\,x_3 &= 6. \end{aligned} \tag{5.109}$$

Dieses LGS besitzt die Lösungsmenge

$$\left\{ \begin{bmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} \frac{1}{2} \\ -2 \\ 1 \end{bmatrix} : \; t \in \mathbb{R} \right\} \tag{5.110}$$

(Video 540, Notizen dazu).

Die Dimension der Lösungsmenge eines LGS kann man gut an der Systemmatrix des LGS in Stufenform ablesen. Dazu folgender Begriff:

Merke

Der **Rang** einer Matrix $\mathbb{R}^{m\times n}$ ist die maximale Anzahl linear unabhängiger Zeilen von A. Der Rang von A wird mit $\mathrm{rg}A$ abgekürzt.

Man kann zeigen, dass der Rang gleichzeitig auch der maximalen Anzahl linear unabhängiger Spalten in A entspricht. Für $A \in \mathbb{R}^{m\times n}$ gilt also $0 \leq \mathrm{rg}\, A \leq \min\{m, n\}$.

Zur Berechnung des Rangs einer Matrix kann der Gauß-Algorithmus verwendet werden, denn der Rang ist gerade die Anzahl an Nicht-Null-Zeilen der Stufenform der Matrix. Beachte, dass je nach gewähltem Vorgehen unterschiedliche Stufenformen entstehen können. Alle werden aber die gleiche Anzahl an Nicht-Null-Zeilen haben.

Die Dimension der Lösungsmenge eines homogenen (!) LGS ist gerade $m - \mathrm{rg}\, A$, wobei m die Anzahl der Gleichungen ist. Bei einem inhomogenen LGS mit $\mathrm{rg}\, A < m$ (also Null-Zeilen in der Systemmatrix der Stufenform) besitzt das LGS entweder keine Lösung oder die Lösungsmenge hat die Dimension $m - \mathrm{rg}\, A$.

Nebenbemerkung

Der Rang einer Matrix spielt auch abseits des Lösens von LGS gelegentlich eine Rolle. Grund dafür ist der folgende Zusammenhang: Der Rang von $A \in \mathbb{R}^{m\times n}$ ist identisch mit der Dimension des **Bildraums** der Abbildung $x \mapsto A\,x$, also mit

der Dimension des Unterraums

$$\{A\,x:\ x\in\mathbb{R}^n\} \tag{5.111}$$

von $\mathbb{R}^m$.

Die folgenden beiden Begriffe sind intuitiv klar, bedürfen aber einer Diskussion im Kontext des Gauß-Algorithmus:

Merke

Ein LGS mit m Gleichungen und n Unbekannten heißt

- **unterbestimmt**, wenn $m < n$,
- **überbestimmt**, wenn $m > n$.

Man kann nun etwas unbedacht Aussagen formulieren wie "unterbestimmte LGS haben unendlich viele Lösungen" oder "überbestimmte LGS sind nicht lösbar". Dies ist aber nicht immer korrekt. Einerseits können unterbestimmte LGS trotzdem sich widersprechende Gleichungen enthalten, also keine Lösung besitzen. Andererseits können überbestimmte LGS redundante Gleichungen enthalten, sodass dennoch Lösungen existieren. Das tatsächliche Lösungsverhalten eines LGS erkennt man erst an der Stufenform!

5.5.3.2 Inverse Matrix Für LGS mit quadratischer Systemmatrix $A \in \mathbb{R}^{n\times n}$ kann man, sofern das LGS eindeutig lösbar ist, die Lösung als Matrix-Vektor-Produkt mit einer geeigneten Matrix angeben. Dazu folgender Begriff:

Merke

Sei $A \in \mathbb{R}^{n\times n}$ eine quadratische Matrix und sei $I \in \mathbb{R}^{n\times n}$ die Einheitsmatrix. Gibt es eine quadratische Matrix $B \in \mathbb{R}^{n\times n}$, sodass

$$A\,B = I \quad \text{und} \quad B\,A = I, \tag{5.112}$$

gilt, so heißt A **invertierbar**. Die Matrix B heißt dann **Inverse** zu A oder **inverse Matrix** zu A. Das übliche Symbol für die Inverse ist

$$A^{-1} := B. \tag{5.113}$$

Man kann zeigen, dass eine Matrix $A \in \mathbb{R}^{n\times n}$ genau dann invertierbar ist, wenn $\operatorname{rg} A = n$ gilt. In diesem Fall kann man das LGS

$$A\,x = b \tag{5.114}$$

auf beiden Seiten von links mit der Inversen multiplizieren:

$$A^{-1}\,A\,x = A^{-1}\,b. \tag{5.115}$$

Wegen $A^{-1}\,A = I$ und $I\,x = x$ erhalten wir mit

$$x = A^{-1}\,b \tag{5.116}$$

eine direkte Berechnungsvorschrift für die Lösung des LGS. Es bleibt somit nur die Frage zu klären, wie man zu A^{-1} kommt.

Das manuelle Aufstellen der Inversen ist sehr mühsam. Man kann das Problem auf das Lösen von n LGS zurückführen, die alle die gleiche Systemmatrix (nämlich A) besitzen; nur die rechten Seiten unterscheiden sich (das sind gerade die n Spalten von I). Hintergrund ist der Zusammenhang

$$A\,B = \begin{bmatrix} | & & | \\ A\,b_{\bullet 1} & \cdots & A\,b_{\bullet n} \\ | & & | \end{bmatrix} \tag{5.117}$$

für Matrix-Vektor-Produkte (vgl. Section 5.1.6). Wenn nun $A\,B = I$ gelten soll, so müssen also die LGS

$$A\,b_{\bullet 1} = e_1, \quad \ldots, \quad A\,b_{\bullet n} = e_n \tag{5.118}$$

erfüllt sein, wobei $e_1, \ldots, e_n$ die Spalten von I sind (e_k hat eine Eins als k-te Komponente, sonst Nullen). Diese LGS können mit dem Gauß-Algorithmus gelöst werden.

Aufgrund des Rechenaufwands ist das Lösen von linearen Gleichungssystemen mittels inverser Matrix nur sinnvoll, wenn das LGS sehr oft für verschiedene rechte Seiten (und gleiche Systemmatrix) gelöst werden muss. Dieser Fall tritt in der Praxis gar nicht so selten auf. Aus Sicht des Anwenders (d.h. (Vermessungs-)Ingenieurs oder (Geo-)Informatikers) sind solche Fragen aber wenig relevant, da diese Themen innerhalb von auf effizientes Lösen von LGS spezialisierten Software-Bibliotheken behandelt und entschieden werden und damit dem Auge des Anwenders verborgen bleiben.

Einige nützliche Zusammenhänge beim Rechnen mit Inversen seien noch erwähnt. Diese lassen sich leicht herleiten. Für quadratische Matrizen $A, B \in \mathbb{R}^{n\times n}$ gilt:

- Sind A und B invertierbar, so ist auch $A\,B$ invertierbar und es gilt

$$(A\,B)^{-1} = B^{-1}\,A^{-1}. \tag{5.119}$$

- Ist A invertierbar, so ist auch A^{-1} invertierbar und es gilt

$$(A^{-1})^{-1} = A. \tag{5.120}$$

- Ist A invertierbar, so ist auch A^{T} invertierbar und es gilt

$$(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}. \tag{5.121}$$

Manchmal schreibt man auch A^{-T} für die Inverse der Transponierten (bzw. für die Transponierte der Inversen).

- A ist genau dann invertierbar, wenn $\det A \neq 0$ gilt.

5.5.4 Symbolisches Lösen

Mit SymPy können lineare Gleichungssysteme wahlweise direkt als Liste von Gleichungen gelöst werden oder in der Matrix-Vektor-Form. Als Liste von Gleichungen:

```
import sympy

x, y, z = sympy.symbols('x y z')

sympy.linsolve([
    2 * x + y + z - 0,
    -6 * x + 3 * y + 9 * z - 6,
    2 * x + 2 * y + 3 * z - 1
], [x, y, z])
```

```
{(z/2 - 1/2, 1 - 2*z, z)}
```

Beachte, dass `sympy.linsolve` auf der rechten Seite der Gleichungen von Nullen ausgeht, wir also die rechte Seite des zu lösenden LGS zunächst auf die linke Seite bringen müssen.

Das Lösen in Matrix-Vektor-Schreibweise funktioniert nur bei invertierbarer Systemmatrix:

```
A = sympy.Matrix([
    [2, 1, 1],
    [2, -1, 9],
    [2, 2, -1]
])
b = sympy.Matrix([0, 6, 1])

A.solve(b)
```

```
Matrix([
[ -7/2],
[   5],
[   2]])
```

Da symbolisch (nicht numerisch) gelöst wird, können im LGS auch allgemeine Parameter enthalten sein:

```
c = sympy.symbols('c')
A = sympy.Matrix([
    [2, 1, c],
    [2, -1, 9],
    [2, 2, -1]
])
b = sympy.Matrix([0, 6, 1])

A.solve(b)
```

```
Matrix([
[(13*c + 15)/(6*c - 14)],
[ ( -5*c - 15)/(3*c - 7)],
[       -32/(12*c - 28)]])
```

Die Inverse einer Matrix kann ebenfalls mit SymPy berechnet werden:

```
A = sympy.Matrix([
    [2, 1, 1],
    [2, -1, 9],
    [2, 2, -1]
])

A.inv()
```

```
Matrix([
[17/8, -3/8, -5/4],
[ -5/2,  1/2,    2],
[ -3/4,  1/4,  1/2]])
```

5.5.5 Numerisches Lösen

Mit NumPy können LGS numerisch gelöst werden. Dies ist deutlich schneller als das symbolische Lösen. Bei kleinen LGS (z.B. die obigen Beispiele) spielt dieser Geschwindigkeitsvorteil keine Rolle. Praktisch relevante LGS sind aber viel größer und symbolisch nicht mehr in akzeptabler Zeit lösbar.

Das numerische Lösen funktioniert nur, wenn die Systemmatrix invertierbar ist.

```
import numpy as np
```

```
A = np.array([
    [2, 1, 1],
    [2, -1, 9],
    [2, 2, -1]
])
b = np.array([0, 6, 1])

np.linalg.solve(A, b)
```

```
array([ -3.5,  5. ,  2. ])
```

Auch die Inverse kann numerisch brechnet werden:

```
A = np.array([
    [2, 1, 1],
    [2, -1, 9],
    [2, 2, -1]
])

np.linalg.inv(A)
```

```
array([[ 2.125, -0.375, -1.25 ],
       [ -2.5  ,  0.5  ,  2.   ],
       [ -0.75 ,  0.25 ,  0.5  ]])
```

Nebenbemerkung

Das numerische Lösen von LGS bedarf einiger Vorsicht, da viele praktisch relevante LGS "schlecht konditioniert" sind. Das bedeutet, dass selbst sehr kleine Änderungen an der rechten Seite zu sehr großen Änderungen in der Lösung führen können. Da die rechte Seite meist (fehlerbehaftete) Messwerte enthält und da aufgrund der Funktionsweise von Computern immer kleine Rundungsfehler auftreten, kann in solch einem Fall die numerisch berechnete Lösung stark von der tatsächlichen Lösung abweichen, bis hin zur Unbrauchbarkeit der "Lösung". Diese komplexe Thematik ist ein wesentlicher Gegenstand der Numerik als Teilgebiet von Mathematik und Informatik. Bevor man ernsthaft große LGS numerisch löst, sollte man sich unbedingt genauer mit dieser Thematik beschäftigen!

5.5.6 Beispiel: Koordinaten bzgl. einer Basis

Hatten noch die Frage offen gelassen, wie wir zu einer gegebenen Basis $\{b_1, \ldots, b_r\}$ in einem Unterraum U von $\mathbb{R}^n$ die Koordinaten eines Vektors $x \in U$ bzgl. dieser Basis berechnen können. Wie wir nun sehen werden, führt diese Frage auf ein LGS, dessen

Lösung gerade die gesuchten Koordinaten sind.

Sei $x \in \mathbb{R}$. Gesucht sind die Zahlen $\lambda_1, \ldots, \lambda_r$, sodass

$$\sum_{k=1}^{r} \lambda_k \, b_k = x \tag{5.122}$$

gilt. Schreiben wir die Basisvektoren als Spalten in eine Matrix $B \in \mathbb{R}^{n \times r}$, so soll also

$$B \, \lambda = x \tag{5.123}$$

gelten (vgl. dazu auch Section 5.1.6).

Für einen echten Unterraum (also $r < n$) ist dieses LGS überbestimmt, wird also nicht für alle $x \in \mathbb{R}^n$ eine Lösung besitzen. Man kann jedoch zeigen, dass dieses LGS für $x \in U$ stets eindeutig lösbar ist. Für $U = \mathbb{R}^n$ ist die in diesem Fall quadratische Matrix B also insbesondere invertierbar.

Für den Fall $r = n$ können wir noch weitere nützliche Beobachtungen machen: Handelt es sich bei $\{b_1, \ldots, b_n\}$ um eine ONB in $\mathbb{R}^n$, so folgt aus

$$x = \sum_{k=1}^{n} \lambda_k \, b_k \tag{5.124}$$

durch Bilden des Skalarprodukts mit b_l, dass

$$x \circ b_l = \left(\sum_{k=1}^{n} \lambda_k \, b_k \right) \circ b_l = \sum_{k=1}^{n} \lambda_k \, (b_k \circ b_l) = \lambda_l, \tag{5.125}$$

also

$$\lambda = \begin{bmatrix} x \circ b_1 \\ \vdots \\ x \circ b_n \end{bmatrix} = B^{\mathrm{T}} \, x \tag{5.126}$$

gilt. Die gesuchten Koordinaten von x bzgl. $\{b_1, \ldots, b_n\}$ erhält man damit ohne nennenswerte Mühe als Matrix-Vektor-Produkt.

Da die Gleichung $x = B \, B^{\mathrm{T}} \, x$ für beliebige $x \in \mathbb{R}^n$ wie gezeigt hergeleitet werden kann, muss

$$B \, B^{\mathrm{T}} = I \tag{5.127}$$

gelten. Offensichtlich gilt auch

$$B^{\mathrm{T}} \, B = I. \tag{5.128}$$

Die Matrix B ist also invertierbar und die Inverse ist

$$B^{-1} = B^{\mathrm{T}}. \tag{5.129}$$

Matrizen, bei denen die Inverse gerade die transponierte Matrix ist, spielen in verschiedenen Teilgebieten von Mathematik (und auch Physik) ein Rolle und bekommen deshalb einen eigenen Name:

Nebenbemerkung

Eine quadratische Matrix $A \in \mathbb{R}^{n \times n}$ heißt **Orthogonalmatrix**, wenn sie invertierbar ist und $A^{-1} = A^{\mathrm{T}}$ gilt.

Folgende Eigenschaften von Orthogonalmatrizen sind gelegentlich hilfreich (und leicht einzusehen):

- Es gilt

$$A^{\mathrm{T}} A = I \quad \text{und} \quad A A^{\mathrm{T}} = I. \tag{5.130}$$

- Die Spalten von A bilden ein Orthonormalsystem (wegen $A^{\mathrm{T}} A = I$).
- Die Zeilen von A bilden ein Orthonormalsystem (wegen $A A^{\mathrm{T}} = I$).
- Die Transponierte A^{T} ist ebenfalls eine Orthogonalmatrix (weil $A = (A^{\mathrm{T}})^{\mathrm{T}}$ ihre Inverse ist).
- Das Produkt zweier Orthogonalmatrizen $A, B \in \mathbb{R}^n$ ist wieder eine Orthogonalmatrix, denn

$$(A B)^{-1} = B^{-1} A^{-1} = B^{\mathrm{T}} A^{\mathrm{T}} = (A B)^{\mathrm{T}}. \tag{5.131}$$

Nebenbemerkung

Beachte, dass die Begriffe sich hier etwas inkonsistent eingebürgert haben. Bei Orthogonalmatrizen sind die Spalten stets normiert, auch wenn das nicht aus dem Name hervorgeht. Bei Orthogonalsystemen müssen die Vektoren jedoch nicht normiert sein. Sind die Vektoren in letzterem Fall normiert, spricht von einem Orthonormalsystem. Der Begriff "Orthonormalmatrix" ist hingegen nicht gebräuchlich.

5.5.7 Beispiel: Kreuzprodukt

In der Ebene steht man gelegentlich vor der Aufgabe, zu einem gegebenen Vektor $a \in \mathbb{R}^2 \setminus \{0\}$ einen zweiten Vektor $b \in \mathbb{R}^2 \setminus \{0\}$ zu finden, sodass beide Vektoren senkrecht zueinander verlaufen. Analog möchte man gelegentlich im Raum zu zwei gegebenen, linear unabhängigen Vektoren $a, b \in \mathbb{R}^3 \setminus \{0\}$ einen dritten Vektor $c \in \mathbb{R}^3 \setminus \{0\}$ finden, der sowohl senkrecht zu a als auch senkrecht zu b verläuft. Diese Aufgabenstellung können wir als LGS formulieren:

- Für $a, b \in \mathbb{R}^2 \setminus \{0\}$ soll $a \circ b = 0$ gelten (b gesucht), d.h.

$$a_1 b_1 + a_2 b_2 = 0. \tag{5.132}$$

- Für $a, b, c \in \mathbb{R}^3 \setminus \{0\}$ sollen $a \circ c = 0$ und $b \circ c = 0$ gelten (c gesucht), d.h.

$$\begin{aligned} a_1\, c_1 + a_2\, c_2 + a_3\, c_3 &= 0 \\ b_1\, c_1 + b_2\, c_2 + b_3\, c_3 &= 0. \end{aligned} \tag{5.133}$$

In beiden Fällen handelt es sich um homogene LGS. Der Rang der Systemmatrizen

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} \quad \text{bzw.} \quad \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \tag{5.134}$$

ist 1 bzw. 2 (a und b sind als linear unabhängig vorausgesetzt!), also um 1 niedriger als die Anzahl der Unbekannten. Somit ist die Lösungsmenge (Unterraum!) in beiden Fällen eindimensional, d.h. die Richtung des gesuchten Vektors ist bis auf das Vorzeichen festgelegt. Da wir die Länge des gesuchten Vektors frei wählen können, wählen wir diese so, dass die Darstellung des Vektors als Formel möglichst einfach wird.

Durch Einsetzen in das entsprechende LGS prüft man leicht nach:

- In $\mathbb{R}^2$ erfüllt

$$b = \begin{bmatrix} -a_2 \\ a_1 \end{bmatrix} \tag{5.135}$$

  die Anforderungen.

- In $\mathbb{R}^3$ erfüllt

$$c = \begin{bmatrix} a_2\, b_3 - a_3\, b_2 \\ a_3\, b_1 - a_1\, b_3 \\ a_1\, b_2 - a_2\, b_1 \end{bmatrix} \tag{5.136}$$

  die Anforderungen.

Merke

Für zwei Vektoren $a, b \in \mathbb{R}$ heißt der Vektor

$$a \times b := \begin{bmatrix} a_2\, b_3 - a_3\, b_2 \\ a_3\, b_1 - a_1\, b_3 \\ a_1\, b_2 - a_2\, b_1 \end{bmatrix} \tag{5.137}$$

Kreuzprodukt (oder äußeres Produkt) von a und b.

Das Kreuzprodukt verläuft stets senkrecht zu a und senkrecht zu b.

Man kann zeigen, dass die Länge $\|a \times b\|$ des Kreuzprodukts gerade dem Flächeninhalt des von a und b aufgespannten Parallelogramms entspricht.

Zu zwei linear unabhängigen Vektoren liefert uns das Kreuzprodukt gerade einen Vektor, der zusammen mit den beiden Ausgangsvektoren ein **Rechtssystem** bildet. Wählt man stattdessen gerade einen Vektor in entgegengesetzter Richtung, so erhält man ein **Linkssystem**. Da diese beiden Begriffe im Folgenden keine Rolle spielen, verweisen wir für eine genauere Erklärung auf den Wikipedia-Artikel zu Rechtssystemen (beachte insbesondere die Rolle der Determinante bei der Unterscheidung von Rechts- und Linkssystemen!).

Nebenbemerkung

Das Kreuzprodukt kann man auch für Vektoren in $\mathbb{R}^n$ für $n \neq 3$ einführen. Allerdings ist das Kreuzprodukt dort eine Abbildung, die $n-1$ gegebene Vektoren in einen neuen Vektor überführt, der senkrecht zu allen gegebenen Vektoren verläuft. Für $n = 2$ liefert dieses verallgemeinerte Kreuzprodukt gerade den oben von uns gefundenen Vektor b.

Für das Rechnen mit Kreuzprodukten kann man sich einige nützliche Zusammenhänge überlegen, z.B. $0 \times a = 0$ und $a \times b = -(b \times a)$. Für eine Übersicht solcher Zusammenhänge sei bei Bedarf auf die entsprechende Auflistung im Wikipedia-Artikel zum Kreuzprodukt verwiesen. Bei praktischen Aufgabenstellungen aus Vermessungswesen und Informatik wird man nur äußerst selten mit Kreuzprodukten rechnen müssen. Das Ergänzen zweier Vektoren zu einer Basis in $\mathbb{R}^3$ mittels Kreuzprodukt ist hingegen öfter von Nutzen.

5.6 Koordinatentransformationen

Eine wichtige Anwendung der bisher eingeführten Werkzeuge sind Koordinatentransformationen einschließlich damit verbundener Projektionstechniken (siehe unter für ein Praxisbeispiel). Ziel ist die Wahl von für den jeweiligen Einsatzzweck besonders gut geeigneten Koordinatensystemen zur Beschreibung der Position von Punkten und anderen geometrischen Objekten. Gut geeignet ist ein Koordinatensystem immer dann, wenn die Koordinaten der betrachteten Objekte einfach zu handhaben sind, also z.B. nur wenige Koordinaten von Null verschieden sind oder relevante Objekteigenschaften direkt aus den Koordinaten abgelesen werden können.

Konzentrieren uns hier auf lineare Koordinatentransformationen. Erwähnen aber zunächst zwei wichtige nichtlineare Transformationen.

5.6.1 Nichtlineare Koordinatentransformationen

Für die spätere Verwendung, jedoch ohne nähere Untersuchung, seien hier Polar- und Kugelkoordinaten erwähnt. Dabei handelt es sich um sehr spezielle Koordinatensysteme, die nicht durch lineare Transformationen aus dem Standardkoordinatensystem erhalten werden können. Unter "Standardkoordinatensystem" verstehen wir in $\mathbb{R}^2$ die Angabe von Koordinaten bezüglich der Basis $\left\{\begin{bmatrix}1 & 0\end{bmatrix}^{\mathrm{T}}, \begin{bmatrix}0 & 1\end{bmatrix}^{\mathrm{T}}\right\}$; in $\mathbb{R}^3$ analog. Solche Koordinatensysteme und entsprechende Transformationen sind Inhalt von spezielleren Lehrveranstaltungen zum Vermessungswesen.

5.6.1.1 Polarkoordinaten Als Polarkoordinaten versteht man die Darstellung eines Punktes (x, y) in der Ebene durch zwei Werte $r \geq 0$ und $\varphi \in [0, 2\,\pi)$, wobei r der Abstand zum Koordinatenursprung ist und φ den Winkel zwischen positiver x-Achse und dem Strahl vom Ursprung durch (x, y) angibt. Polarkoordinaten hatten wir bereits im Abschnitt Section 4.4.6.3 eingeführt; dort findet man auch eine grafische Darstellung des Sachverhalts.

Die Transformation von Polarkoordinaten zu Standardkoordinaten ist durch

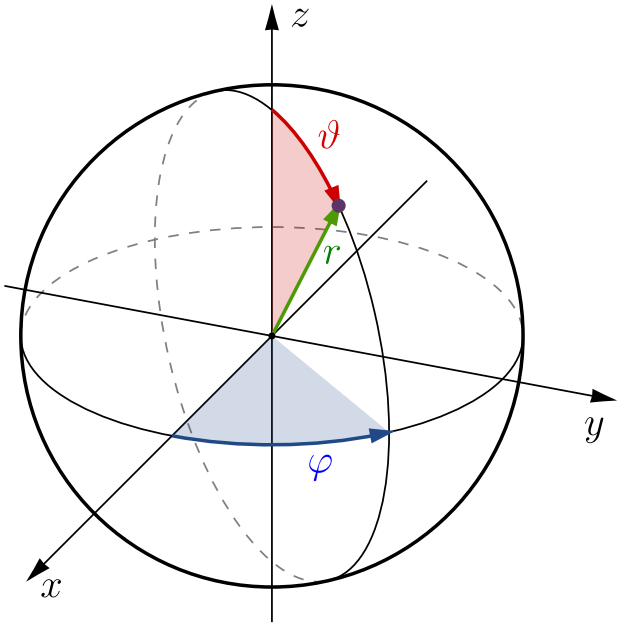
$$\begin{aligned} x &= r\,\cos\varphi, \\ y &= r\,\sin\varphi \end{aligned} \tag{5.138}$$

gegeben. In umgekehrter Richtung gilt

$$r = \sqrt{x^2 + y^2}. \tag{5.139}$$

Die Berechnung von φ lässt sich jedoch nur umständlich als Formel aufschreiben. Allerdings bieten mathematische Software-Bibliotheken (z.B. NumPy) meist eine Funktion `atan2` an. Dies ist eine spezielle Variante des Arcustangens, die zu den zwei zu übergebenden Argumenten x und y gerade den Winkel φ liefert.

5.6.1.2 Kugelkoordinaten Als Kugelkoordinaten versteht man die Darstellung eines Punktes $(x, y, z) \in \mathbb{R}^3$ im Raum durch drei Werte $r > 0$, $\varphi \in [0, 2\,\pi)$ und $\vartheta \in [0, \pi]$, wobei r der Abstand zum Ursprung ist, φ den Winkel zwischen positiver x-Achse und dem Strahl vom Ursprung durch $(x, y, 0)$ angibt und ϑ der Winkel zwischen positiver z-Achse und dem Strahl vom Ursprung durch (x, y, z) ist.



Lage der Winkel bei Kugelkoordinaten (Quelle: commons.wikimedia.org, Nutzer Inductiveload, CC0, vom Autor bearbeitet).

Die Transformation von Kugelkoordinaten zu Standardkoordinaten ist durch

$$\begin{aligned} x &= r\,(\cos\varphi)\,\sin\vartheta, \\ y &= r\,(\sin\varphi)\,\sin\vartheta, \\ z &= r\,\cos\vartheta \end{aligned} \tag{5.140}$$

gegeben. In umgekehrter Richtung gilt

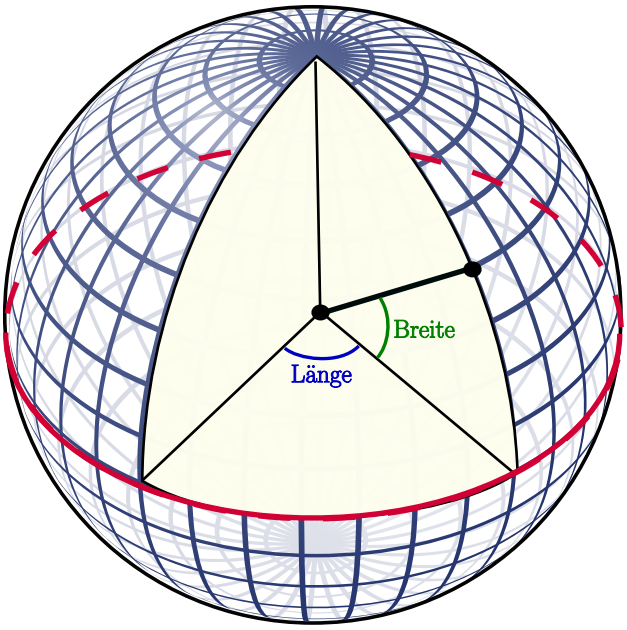
$$r = \sqrt{x^2 + y^2 + z^2}. \tag{5.141}$$

Die Berechnung der beiden Winkel lässt sich nur mühsam in Formeln ausdrücken und entfällt hier deshalb.

Beachte, dass die hier wiedergegebene und in der Mathematik übliche Form der Kugelkoordinaten von der in der Geografie üblichen Darstellung abweicht. Der Winkel φ entspricht der geografischen Länge eines Punktes, wobei in der Geografie meist das Intervall $(-\pi, \pi]$ verwendet wird. Der Winkel ϑ entspricht gerade dem Wert

$$\frac{\pi}{2} - \text{geografische Breite}. \tag{5.142}$$

Statt vom Nordpol ausgehend misst man in der Geografie also relativ zur Äquatorebene.



Lage der Winkel bei geografischen Koordinaten (Quelle: commons.wikimedia.org, Nutzer Peter Mercator, CC0, vom Autor bearbeitet).

Außerdem werden in der Geografie Breite und Länge üblicherweise im Gradmaß und nicht, wie in der Mathematik verbreitet, im Bogenmaß angegeben. In beiden Gebieten ist jedoch im Prinzip beides möglich, da sich die entsprechenden Zahlenwerte nur um einen konstanten Faktor unterscheiden:

$$\text{Winkel im Gradmaß} = \frac{180}{\pi} \cdot \text{Winkel im Bogenmaß}. \tag{5.143}$$

5.6.2 Koordinatensysteme

Haben den Begriff “Koordinatensystem” bisher eher intuitiv verwendet ohne genauere Definition. Präzisieren diesen Begriff nun, damit Missverständnisse vermieden werden. Beachte, dass wir im Weiteren nur geradlinige, rechtwinklige Koordinatensystem betrachten ohne dies explizit sprachlich sichtbar zu machen, also solche, die die folgenden beiden Bedingungen erfüllen:

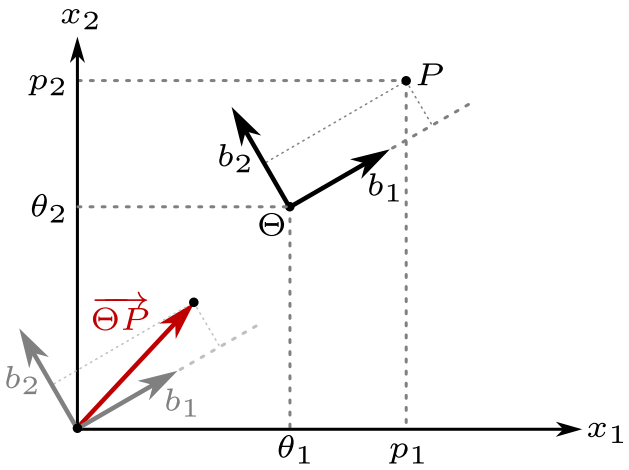
- Bei Veränderung genau einer Koordinate eines Punktes wird der Punkt entlang einer Gerade verschoben.
- Die Verschiebungsrichtungen bei Änderung jeweils genau einer Koordinate verlaufen senkrecht zueinander.

Beispielsweise liefern Polarkoordinaten und Kugelkoordinaten keine geradlinigen Koordinatensysteme, da das Verändern eines Winkels zu einer krummlinigen Bewegung des Punktes führt. Manchmal werden die hier betrachteten Koordinatensysteme als "kartesisch" bezeichnet. Diese Bezeichnung wird aber in der Literatur unterschiedlich verwendet.

Merke

Ein **Koordinatensystem** in $\mathbb{R}^n$ ist ein Paar (Θ, B) bestehend aus einem Punkt $\Theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$ und einer Basis $B = \{b_1, \ldots, b_n\}$. Der Punkt Θ wird als **Ursprung** des Koordinatensystems bezeichnet. Die Basisvektoren geben die Richungen der vom Ursprung ausgehenden **Koordinatenachsen** an.

Als **Koordinaten** eines Punkts $P = (p_1, \ldots, p_n) \in \mathbb{R}^n$ bezüglich des Koordinatensystems (Θ, B) bezeichnen wir die Koordinaten des Vektors $\overrightarrow{\Theta P}$ bezüglich der Basis B.



Die Koordinaten das Punktes P bzgl. des Koordinatensystems (Θ, B) sind gerade die Koordinaten des Vektors $\overrightarrow{\Theta P}$ bzgl. der Basis B.

Im Folgenden bezeichnen wir mit $\{e_1, \ldots, e_n\}$ stets die **Standardbasis** in $\mathbb{R}^n$ (e_k hat als k-te Komponente eine Eins, sonst Nullen). Zusammen mit dem Ursprung $(0, \ldots, 0) \in \mathbb{R}^n$ beschreibt diese das **Standardkoordinatensystem** in $\mathbb{R}^n$.

Nebenbemerkung (Achsen vs. Basis)

Üblicherweise deutet man bei der grafischen Darstellungen eines Koordinatensystems die Koordinatenachsen als (lange) Pfeile an, die über den gesamten relevanten Bereich verlaufen, wie in der obigen Abbildung für die x- und die y-Achse

(Standardkoordinatensystem) geschehen. Die (kurzen) Basisvektoren selbst tauchen im Normalfall nicht auf, wurden jedoch des besseren Verständnisses halber in der Abbildung an Stelle der Achsen des Koordinatensystems (Θ, B) dargestellt.

Nebenbemerkung (griechische Buchstaben)

Das Zeichen Θ ist ein großes griechisches Theta und wurde hier aufgrund seiner Ähnlichkeit zur Null oder zu O gewählt. Das kleine griechische Theta gibt es in zwei Varianten: ϑ und θ. Für Winkel nutzen wir meist ϑ, da dieses nicht so schnell mit der Null verwechselt werden kann. Für die Komponenten eines Punktes Θ nutzen wir hingegen θ, da dieses mehr Ähnlichkeit zu Θ hat. Generell kann es nicht schaden, sich mit griechischen Buchstaben zu befassen, da diese in der Mathematik und all ihren Anwendungsgebieten (Naturwissenschaften, Technik,...) gern und oft verwendet werden. Eine gute Übersicht liefert der Wikipedia-Artikel zum griechischen Alphabet. Auch die meisten Formelsammlungen enthalten solche Übersichten.

Beachte, dass wir zwischen "Koordinaten bzgl. einer Basis" und "Koordinaten bzgl. eines Koordinatensystems" unterscheiden. Erstere werden relativ zum Punkt $(0, \dots, 0)$ angegeben, letztere relativ zum Ursprung des Koordinatensystems. Die Koordinaten bzgl. einer Basis B sind identisch mit den Koordinaten bzgl. des Koordinatensystems $(0, B)$ (hier steht 0 für den Nullpunkt $(0, \dots, 0)$).

Schreiben wir die Basisvektoren eines Koordinatensystems (Θ, B) als Spalten in eine Matrix, so können wir entsprechend den Ausführungen in Section 5.5.6 die Koordinaten eines Punktes P bzgl. des Koordinatensystems als Matrix-Vektor-Produkt berechnen. Um nicht zu viele verschiedene Symbole einzuführen, verwenden wir für die Basis und für die Matrix der Basisvektoren das selbe Symbol B. Je nach Kontext handelt es sich dabei also um eine Menge von Vektoren oder um eine Matrix! Für die Koordinaten $u_1, \dots, u_n$ eines Punktes $P = (p_1, \dots, p_n) \in \mathbb{R}^n$ bzgl. des Koordinatensystems (Θ, B) gilt nun

$$\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = B^{\mathrm{T}} \begin{bmatrix} p_1 - \theta_1 \\ \vdots \\ p_n - \theta_n \end{bmatrix}. \tag{5.144}$$

Umgekehrt bekommt man aus den Koordinaten $u_1, \dots, u_n$ auch die Werte $p_1, \dots, p_n$ zurück (vorherige Formel umstellen; dabei $B^{\mathrm{T}} = B^{-1}$ beachten):

$$\begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} + B \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}. \tag{5.145}$$

Nebenbemerkung (Koordinatensysteme in Unterräumen)

Im Prinzip müsste man auch Koordinatensysteme in Unterräumen und Untermannigfaltigkeiten des $\mathbb{R}^n$ diskutieren. Um diesen zusätzlichen Aufwand zu vermeiden, ergänzen wir eine ONB in einem Unterraum zu einer ONB in ganz $\mathbb{R}^n$ (dies ist immer möglich!) und wählen den Ursprung im Unterraum bzw. in der Untermannigfaltigkeit. Betrachten wir nun die Koordinaten eines Punktes bzgl. dieses Koordinatensystems im $\mathbb{R}^n$, so sind die Koordinaten zu den ergänzten Koordinatenachsen alle Null, sofern der Punkt im Unterraum bzw. der Untermannigfaltigkeit liegt. Die ergänzten Achsen sind also nur von formalem Interesse, werden aber nicht für die Beschreibung der relevanten Punkte/Vektoren benötigt.

5.6.3 Wechsel des Koordinatensystems

Als **lineare Koordinatentransformation** bezeichnen wir die Berechnung der Koordinaten eines Punktes bzgl. eines "neuen" Koordinatensystems aus den Koordinaten dieses Punktes bzgl. eines "alten" Koordinatensystems. Wie wir sehen werden, entspricht die Transformationsformel, die die alten Koordinaten in die neuen überführt, bis auf zusätzliche Verschiebungsoperationen einer linearen Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}^n$.

Sei also (Θ, B) mit $\Theta = (\theta_1, \ldots, \theta_n)$ und $B = \{b_1, \ldots, b_n\}$ das "alte" Koordinatensystem und seien $u_1, \ldots, u_n$ die Koordinaten eines Punktes $P = (p_1, \ldots, p_n) \in \mathbb{R}^n$ bezüglich dieses Koordinatensystems. Weiter sei (Γ, C) mit $\Gamma = (\gamma_1, \ldots, \gamma_n)$ und $C = \{c_1, \ldots, c_n\}$ das "neue" Koordinatensystem und $v_1, \ldots, v_n$ seien die Koordinaten des Punktes P bezüglich dieses Koordinatensystems. Wir wissen bereits, dass die folgenden Beziehungen gelten:

- Übergang vom Standardkoordinatensystem zu (Γ, C):

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = C^{\mathrm{T}} \begin{bmatrix} p_1 - \gamma_1 \\ \vdots \\ p_n - \gamma_n \end{bmatrix}. \tag{5.146}$$

- Übergang von (Θ, B) zum Standardkoordinatensystem:

$$\begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} + B \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}. \tag{5.147}$$

Setzen wir die zweite Formel in die erste ein, so erhalten wir die gesuchte Transformati-

onsvorschrift:

$$\begin{aligned}\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} &= C^{\mathrm{T}} \begin{bmatrix} p_1 - \gamma_1 \\ \vdots \\ p_n - \gamma_n \end{bmatrix} = C^{\mathrm{T}} \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} - C^{\mathrm{T}} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} \\ &= C^{\mathrm{T}} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} + C^{\mathrm{T}} B \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} - C^{\mathrm{T}} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} \\ &= C^{\mathrm{T}} \begin{bmatrix} \theta_1 - \gamma_1 \\ \vdots \\ \theta_n - \gamma_n \end{bmatrix} + C^{\mathrm{T}} B \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}.\end{aligned} \tag{5.148}$$

Der Übergang zwischen zwei Koordinatensystemen besteht also aus dem Anwenden einer linearen Abbildung (Multiplikation mit einer Matrix) und einer Verschiebung.

Nebenbemerkung

Als Plausibilitätstest für die gefundene Transformationsformel kann man den Spezialfall betrachten, dass wahlweise das alte oder das neue Koordinatensystem das Standardkoordinatensystem ist. Dann sollte sich die Transformationformel auf die Formel zur Berechnung der Koordinaten eines Punktes bzgl. eines Koordinatensystems reduzieren.

Nebenbemerkung

Meistens benötigt man die hier hergeleitete Transformationsformel für den Wechsel des Koordinatensystems gar nicht, wenn man, wie wir, stets das Standardkoordinatensystem als Arbeitsgrundlage nutzt, also keine koordinatenfreie Geometrie betreibt (vgl. Nebenbemerkung (koordinatenfreie Geometrie)). Man ist praktisch nur mit Übergängen zwischen dem Standardkoordinatensystem und einem anderen Koordinatensystem konfrontiert, kaum mit dem Übergang zwischen zwei verschiedenen Nicht-Standardkoordinatensystemen.

5.6.4 Beispiel: Orthografische Azimuthalprojektion

5.6.4.1 Problemstellung und Lösungsansatz Für eine Region (z.B. Deutschland) soll eine möglichst verzerrungsarme, zweidimensionale Kartendarstellung gefunden werden. Der Einfachheit halber legen wir uns auf orthografische Azimuthalprojektionen fest. Bei diesen ist die Verzerrung in der Kartenmitte am geringsten, sodass wir als Kartenmitte die “Mitte” der darzustellenden Region wählen. Unter “Mitte” verstehen wir hier den Punkt auf der Erdoberfläche, für den die maximale Entfernung zu allen anderen Punkten der Region am geringsten ist. Für Deutschland ist dies 10.456555° östliche Länge,

51.237923° nördliche Breite (Quelle: Administrative units/azimuthal equidistant projections datasets).

Zu einem durch geografische Länge und Breite gegebenen Punkt auf der Erdoberfläche sind die Koordinaten bezüglich des Kartenkoordinatensystems gesucht. Der Ursprung dieses Koordinatensystems liegt im Mittelpunkt der Region und die relevante Kartenebene (die Ebene, die von den ersten beiden Koordinatenachsen aufgespannt wird) soll die als Kugeloberfläche angenommene Erdoberfläche in diesem Punkt berühren, aber nicht schneiden (Stichwort: Tangentialebene; diesen Begriff diskutieren wir nächstes Semester genauer).

Aufgabe: Man finde eine Formel, die Längen- und Breitenangaben in Kartenkoordinaten überführt.

Wir werden diese Aufgabe in drei Schritten lösen:

1. Längen- und Breitenangaben in Punkte in $\mathbb{R}^3$ überführen.
2. Eine Basis für die Beschreibung des Kartenkoordinatensystems wählen (der Ursprung ist schon geklärt).
3. Die Berechnung der Koordinaten eines Punktes bzgl. des Kartenkoordinatensystems klären.
4. Eine Gesamtformel "zusammenbauen", die die Aufgabenstellung löst.

5.6.4.2 Punkte im Raum Seien $\lambda \in (-180, 180]$ und $\zeta \in [-90, 90]$ geografische Länge und Breite eines Punktes auf der Erdoberfläche im Gradmaß. Wir legen die als Kugeloberfläche angenommene Erdoberfläche so in den $\mathbb{R}^3$, dass

- der Erdmittelpunkt bei $(0, 0, 0)$ liegt,
- die positive x-Achse die Erdoberfläche bei 0° Länge und 0° Breite durchstößt,
- die positive y-Achse die Erdoberfläche bei 90° östlicher Länge und 0° Breite durchstößt
- die positive z-Achse den Nordpol durchstößt (90° nördliche Breite).

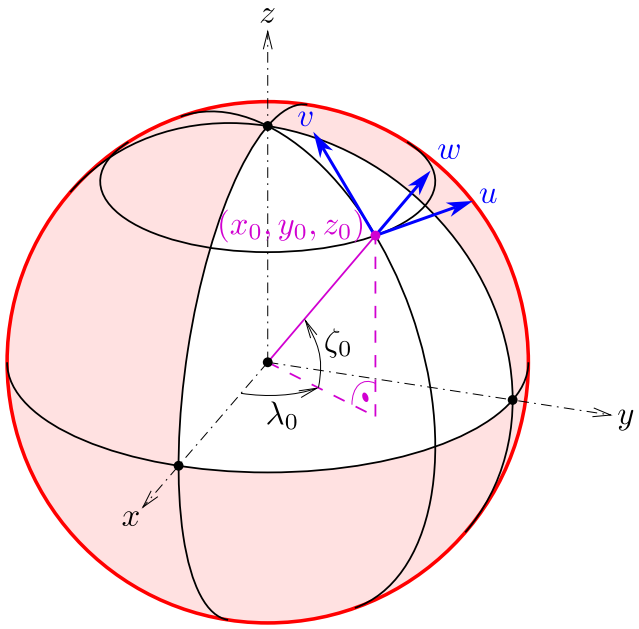
Für einen durch geografische Längen- und Breitenangaben λ und ζ beschriebenen Punkt P auf der Erdoberfläche erhalten wir die Koordinaten bzgl. des Standardkoordinatensystems in $\mathbb{R}^3$ somit als

$$\begin{aligned} x &= r \left(\cos\left(\frac{\pi}{180}\lambda\right)\right) \sin\left(\frac{\pi}{2} - \frac{\pi}{180}\zeta\right), \\ y &= r \left(\sin\left(\frac{\pi}{180}\lambda\right)\right) \sin\left(\frac{\pi}{2} - \frac{\pi}{180}\zeta\right), \\ z &= r \cos\left(\frac{\pi}{2} - \frac{\pi}{180}\zeta\right). \end{aligned} \tag{5.149}$$

Dabei ist $r = 6371$ der mittlere Erdradius in Kilometer. Die Koordinaten x, y, z sind entsprechend ebenfalls als Entfernungen vom Ursprung in Kilometer zu interpretieren.

Die Koordinaten des Mittelpunkts der betrachteten Region bezeichnen wir mit x_0, y_0, z_0. Diese entstehen mittels obiger Formeln aus Länge λ_0 und Breite ζ_0 des Mittelpunkts der Region.

5.6.4.3 Kartenkoordinatensystem Legen nun Schritt für Schritt die Richtungen der Koordinatenachsen des Kartenkoordinatensystems fest. Beachte dabei, dass das gewünschte Koordinatensystem mit den ersten beiden Achsen die horizontale und die vertikale Achse der Karte beschreiben soll. Die dritte Achse steht senkrecht auf der Kartenebene und spielt in der zweidimensionalen Kartendarstellung keine Rolle. Wir bezeichnen die den Achsen entsprechenden Basisvektoren in dieser Reihenfolge mit $b_1, b_2, b_2 \in \mathbb{R}^3$.



Übersicht über die beteiligten Größen und Symbole. Die u-, v- und w-Achsen sind die durch die Basisvektoren b_1, b_2, b_3 definierten Koordinatenachsen (Quelle: commons.wikimedia.org, Nutzer Ag2gaeh, CC BY-SA 4.0, vom Autor bearbeitet).

Da der Vektor b_3 senkrecht zur Kartenebene verlaufen soll, entspricht dieser gerade der Richtung des Strahls vom Kugelmittelpunkt durch den Mittelpunkt (x_0, y_0, z_0) der betrachteten Region. Somit wählen wir

$$b_3 := \frac{1}{\sqrt{x_0^2 + y_0^2 + z_0^2}} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}. \tag{5.150}$$

Der Vektor b_2 muss senkrecht zu b_3 verlaufen. Außerdem möchten wir die Kartendarstellung gern so einrichten, dass “Norden oben ist”. Bei einer gedachten Bewegung von

(x_0, y_0, z_0) ausgehend mit Richtung b_2 muss also die z-Achse durchstoßen werden. Elementargeometrische Berechnungen (Schnittpunkt der zu b_3 senkrechten Ebene durch (x_0, y_0, z_0) mit der z-Achse; b_2 ist dann der Vektor von (x_0, y_0, z_0) zu diesem Schnittpunkt zuzüglich Normierung) liefern

$$b_2 := \frac{\tilde{b}_2}{\|\tilde{b}_2\|} \quad \text{mit} \quad \tilde{b}_2 := \begin{bmatrix} -x_0 \\ -y_0 \\ \frac{x_0^2+y_0^2+z_0^2}{z_0} - z_0 \end{bmatrix}. \tag{5.151}$$

Der Vektor b_1 ergibt sich nun als Kreuzprodukt von b_2 und b_3 zuzüglich Normierung:

$$b_1 := \frac{b_2 \times b_3}{\|b_2 \times b_3\|} \tag{5.152}$$

(auf eine explizite Formel verzichten wir hier; diese Rechnung erledigt für konkrete x_0, y_0, z_0 später der Computer).

5.6.4.4 Transformationsformel Schreiben wir die Basisvektoren b_1, b_2, b_3 als Spalten in eine Matrix $B \in \mathbb{R}^{3\times3}$, so erhalten wir für die Koordinaten u, v, w eines Punktes $(x, y, z) \in \mathbb{R}^3$ bzgl. des Kartenkoordinatensystems die Formel

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = B^{\mathrm{T}} \begin{bmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{bmatrix}. \tag{5.153}$$

5.6.4.5 Umsetzung am Computer Zum Test der Transformation möchten wir die Grenze von Deutschland in einer Karte darstellen. Die Datei germany.txt enthält geografische Länge und Breite von 3000 Punkten auf der Grenze. Diese wurden zufällig aus der durch 159230 Punkte dargestellten Grenze von Deutschland in OpenStreetMap ausgewählt (Stand: 23.08.2026). Das Einlesen der Punkte kann mit numpy.genfromtxt erfolgen. Anschließend erfolgt die oben hergeleitete Transformation in Kartenkoordinaten.

```
import numpy as np
import matplotlib.pyplot as plt

# Erdradius
r = 6371

# Mittelpunkt von Deutschland
lon0 = 10.456555
lat0 = 51.237923

# Punkte auf der Grenze von Deutschland
```

```
lonlat = np.genfromtxt('germany.txt')
lon = lonlat[:, 0]
lat = lonlat[:, 1]

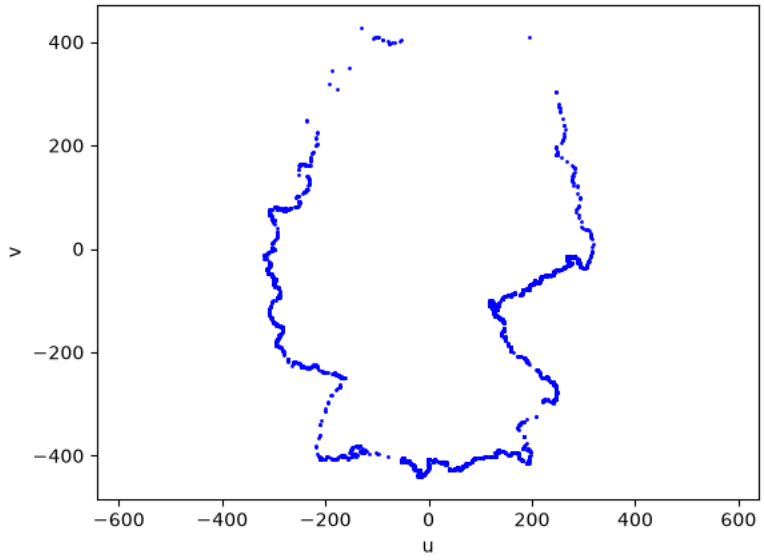
# Transformation zu Punkten im Raum
fac = np.pi / 180  # Umrechnungsfaktor Gradmaß zu Bogenmaß
x0 = r * np.cos(fac * lon0) * np.sin(np.pi / 2 - fac * lat0)
y0 = r * np.sin(fac * lon0) * np.sin(np.pi / 2 - fac * lat0)
z0 = r * np.cos(np.pi / 2 - fac * lat0)
x = r * np.cos(fac * lon) * np.sin(np.pi / 2 - fac * lat)
y = r * np.sin(fac * lon) * np.sin(np.pi / 2 - fac * lat)
z = r * np.cos(np.pi / 2 - fac * lat)

# Basis für Kartenkoordinatensystem
b3 = np.array([x0, y0, z0])
b3 = b3 / np.linalg.norm(b3)
b2 = np.array([ -x0, -y0, (x0 ** 2 + y0 ** 2 + z0 ** 2) / z0 - z0])
b2 = b2 / np.linalg.norm(b2)
b1 = np.cross(b2, b3)
b1 = b1 / np.linalg.norm(b1)

# Koordinaten bzgl. des Kartenkoordinatensystems
B = np.column_stack([b1, b2, b3])
uvw = B.T @ np.vstack([x - x0, y - y0, z - z0])

# jede Spalte von uvw entspricht einem Punkt
u = uvw[0, :]
v = uvw[1, :]
w = uvw[2, :]

# Visualisierung
fig, ax = plt.subplots()
scatter_plot = ax.scatter(u, v, s=1, c='blue')
ax.axis('equal')  # gleiche Achseneinteilung (keine Verzerrung)
ax.set_xlabel('u')
ax.set_ylabel('v')
plt.show()
```



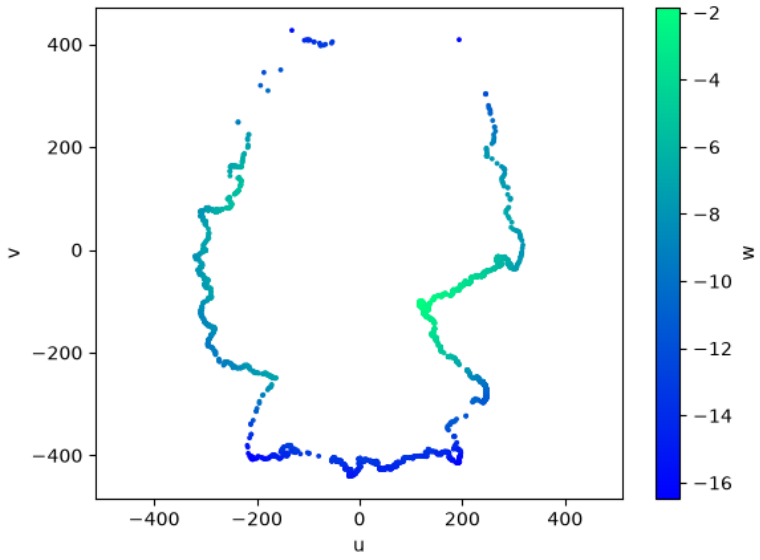
Die Ursache für das Fehlen von Punkten in den Küstenregionen liegt im einfachen Grenzverlauf dort. Dieser lässt sich mit deutlich weniger Punkten beschreiben, sodass bei der zufälligen Auswahl von Punkten entsprechend weniger Punkte in Küstenregionen ausgewählt werden.

Wir können auch die w-Koordinate visualisieren, indem wir die einzelnen Punkte entsprechend einfärben:

```
fig, ax = plt.subplots()

scatter_plot = ax.scatter(u, v, s=3, c=w, cmap='winter')
plt.colorbar(scatter_plot, label='w')

ax.axis('equal')  # gleiche Achseneinteilung (keine Verzerrung)
ax.set_xlabel('u')
ax.set_ylabel('v')
plt.show()
```



Wir sehen dass die w-Koordinate zur Kartenmitte hin näher bei Null liegt als zum Kartenrand hin. Diese Beobachtung deckt sich mit der Theorie, denn die w-Koordinate gibt den Abstand eines Punktes auf der Erdoberfläche von der Kartenebene an. Insbesondere müssen alle w-Koordinaten negativ sein. An der Nordgrenze und an der Südgrenze beträgt der Abstand zwischen Kartenebene und der (als Kugeloberfläche angenommenen) Erdoberfläche bis zu 16 Kilometer.

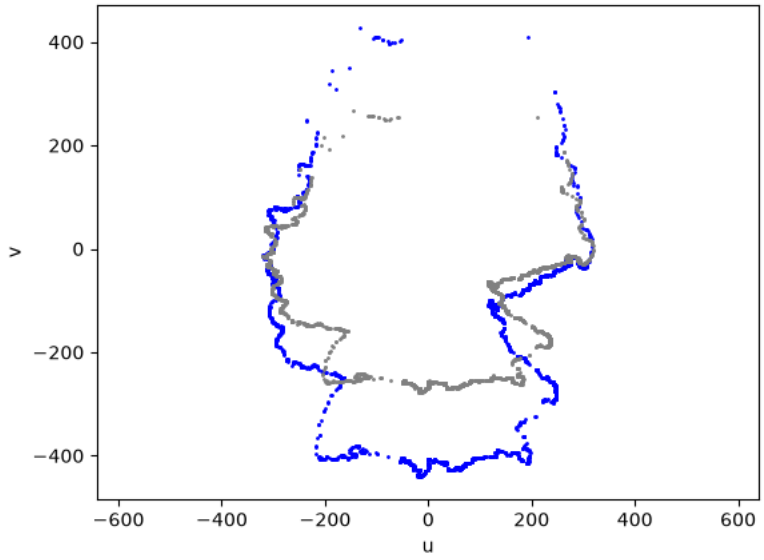
Ohne Kenntnis der Techniken zur Koordinatentransformation hätten wir nur die Längen- und Breitenwerte direkt als Koordinaten in einer zweidimensionalen Darstellung verwenden können. Dies liefert eine sogenannte Plattkarte. Zum Vergleich legen wir beide Varianten übereinander und skalieren die Plattkarte so, dass bei $v \approx 0$ die u-Koordinaten übereinstimmen (der Skalierungsfaktor wurde hier manuell durch Probieren verschiedener Werte gewählt).

```
fig, ax = plt.subplots()

scatter_plot = ax.scatter(u, v, s=1, c='blue')

fac = 70  # Skalierungsfaktor für Plattkarte
ax.scatter(fac * (lon - lon0), fac * (lat - lat0), s=1, c='gray')

ax.axis('equal')  # gleiche Achseneinteilung (keine Verzerrung)
ax.set_xlabel('u')
ax.set_ylabel('v')
plt.show()
```



Offensichtlich führt die (triviale) Darstellung als Plattkarte (grau) zu einer deutlich stärkeren Verzerrung als die mittels linearer Koordinatentransformation erhaltene orthografische Azimuthalprojektion (blau).

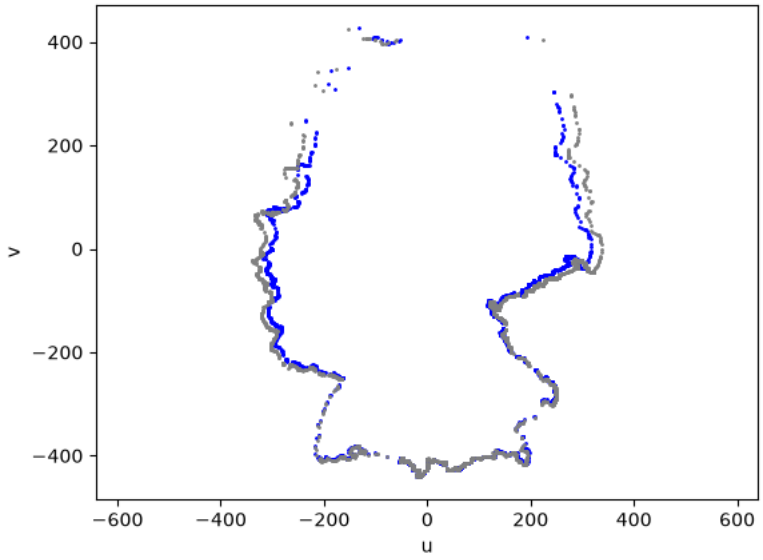
Auch bei unterschiedlicher Skalierung in horizontaler und vertikaler Richtung ist die Plattkarte nicht mit der sauber projizierten Karte in Deckung zu bringen. Wählt man die Skalierungsfaktoren beispielsweise so, dass im Süden gute Übereinstimmung entsteht, sind die Abweichungen im Norden groß:

```
fig, ax = plt.subplots()

scatter_plot = ax.scatter(u, v, s=1, c='blue')

ax.scatter(74 * (lon - lon0), 111 * (lat - lat0), s=1, c='gray')

ax.axis('equal')  # gleiche Achseneinteilung (keine Verzerrung)
ax.set_xlabel('u')
ax.set_ylabel('v')
plt.show()
```



5.6.4.6 Einordnung In diesem Beispiel haben wir gesehen, wie wir mit den bisher beleuchteten Werkzeugen und Zusammenhängen durch Kugelkoordinaten oder Standardkoordinaten im $\mathbb{R}^3$ gegebene Punkte durch zwei Koordinaten in einer geeignet gewählten Ebene (Karte) darstellen können. Die Kartenebene ist in der mathematischen Sprache eine lineare Untermannigfaltigkeit des $\mathbb{R}^3$. Betrachten wir jedoch nur die ersten beiden Koordinaten eines Punktes bzgl. des passend zu dieser Untermannigfaltigkeit gewählten Koordinatensystems, so können wir die ganze Komplexität der dreidimensionalen Situation außer Acht lassen. Letztlich sind die wesentlichen Informationen durch zwei Zahlen (Koordinaten) darstellbar, die wir übersichtlich in unserer Zeichenebene (Papier, Laptop-Display,...) in einem einfachen zweidimensionalen Koordinatensystem darstellen können. In diesem Sinne macht die Mathematik das Leben hier deutlich einfacher!

Das Vernachlässigen der dritten Dimension führt hier nur zu geringem Informationsverlust, da die Erdkrümmung (wenigstens für kleine Kartenausschnitte) sehr gering ist. Den Gedanke, durch geeignete Wahl von Koordinatensystemen Informationen mit nur wenig Verlust deutlich einfacher darstellen zu können, werden wir im nächsten Unterkapitel nochmal aufnehmen. Dort wird dann auch klar, dass das hier demonstrierte Vorgehen auch in $\mathbb{R}^n$ mit sehr großem n praktisch relevant ist, die Mühen der Formulierung aller Begriffe und Zusammenhänge für allgemeines n also kein (scheinbar) unnötiger Ballast waren.

Die Suche nach möglichst verzerrungsarmen Kartendarstellungen hat insbesondere den Zweck, dass in der Karte gemessene Abstände möglichst nah an den tatsächlichen Abständen auf der Erdoberfläche liegen sollen. Je nach gewählter Kartenprojektion wird die Abweichung zwischen Abständen in der Karte und den realen Abständen unterschiedlich groß sein. Wir haben hier eine sehr einfache Kartenprojektion gewählt. Komplexere Kartenprojektionen und die ganze Fehlerthematik beim Messen von Abständen in Karten

werden in anderen Veranstaltungen Ihres Studiums vertieft werden.

5.7 Ausblick: Eigenwerte und Eigenvektoren

Dieses Unterkapitel ist für die Prüfung nicht relevant. Es dient lediglich der ersten Einordnung einiger Begriffe, die Ihnen im Laufe Ihres Studiums begegnen werden. Wir haben mit der Einführung der Ideen und Konzepte rund um Vektoren und Matrizen viel Vorarbeit geleistet; die Früchte unserer Saat werden dann in anderen Veranstaltungen geerntet. Gönnen wir uns hier wenigstens einen kleinen Vorgeschmack...

5.7.1 Idee

Koordinatensysteme wählt man üblicherweise so, dass die auszuführenden Arbeiten (Berechnungen, Visualisierungen,...) möglichst einfach werden. Sind zahlreiche Matrix-Vektor-Multiplikationen mit einer Matrix A durchzuführen, so kann man die Frage stellen, ob die Berechnung von Matrix-Vektor-Produkten in einem an die konkrete Matrix A angepassten Koordinatensystem eventuell deutlich einfacher wird.

In der Tat ist dies der Fall: Zu jeder symmetrischen (!) Matrix $A \in \mathbb{R}^{n\times n}$ gibt es eine ONB, sodass die Matrix-Vektor-Multiplikation bei Darstellung des Vektors bzgl. dieser ONB zu einer elementweisen Multiplikation des Vektors mit einem festen Vektor wird. Sind $u_1, \ldots, u_n$ die Koordinaten eines Vektors $x \in \mathbb{R}^n$ bzgl. dieser speziellen Basis, so gilt also

$$A\,x = \begin{bmatrix} \lambda_1\,u_1 \\ \vdots \\ \lambda_n\,u_n \end{bmatrix} \tag{5.154}$$

mit festen Zahlen $\lambda_1, \ldots, \lambda_n$, die aus den Einträgen der Matrix A berechnet werden können. Statt des aufwendigen Matrix-Vektor-Produkts mit n^2 Multiplikationen und $n\,(n-1)$ Additionen sind nun nur noch n Multiplikationen nötig. Für $n = 1000$ (für heutige Verhältnisse also eher klein) reduziert sich der Aufwand beispielweise von 1000000 Multiplikationen und Additionen auf 1000 Multiplikationen.

Diese speziell an die Matrix angepasste Basis heißt **Eigenvektorbasis** von A; die Basisvektoren heißen **Eigenvektoren**. Die festen Faktoren λ_k für die Matrix-Vektor-Multiplikation heißen **Eigenwerte**.

5.7.2 Präzisierung

Führen die obige Idee noch in Formeln aus.

Merke

Sei $A \in \mathbb{R}^{n\times n}$ eine symmetrische Matrix. Dann existieren eine ONB $\{b_1, \ldots, b_n\}$ und Zahlen $\lambda_1, \ldots, \lambda_n$, sodass

$$A\, b_k = \lambda_k\, b_k \quad \text{für } k = 1, \ldots, n \tag{5.155}$$

gilt. Die Vektoren b_k heißen **Eigenvektoren** von A. Die Zahlen λ_k heißen **Eigenwerte** von A.

Nebenbemerkung

Die Existenz von Eigenvektoren und Eigenwerten ist keineswegs einfach nachzuweisen. Deshalb findet man oft folgende Definition, die die Existenzfrage zunächst beiseitelässt; ob die so definierten Objekte (Eigenvektoren, Eigenwerte) überhaupt existieren, muss dann im Nachgang geklärt werden:

> Sei $A \in \mathbb{R}^{n\times n} \setminus \{0\}$ eine symmetrische Matrix. Ein Vektor $b \in \mathbb{R}^n$ heißt Eigenvektor von A, wenn es eine Zahl λ gibt, sodass $A\, b = \lambda\, b$ gilt. Die Zahl λ heißt dann Eigenwert von A.

Bei dieser Definition sind automatisch auch alle Vielfachen eines Eigenvektors wieder Eigenvektoren. Hat man sich (recht mühsam) von der Existenz von Eigenvektoren überzeugt, so muss man noch die Existenz einer Orthonormalbasis aus Eigenvektoren zeigen (ebenfalls mühsam).

Sind $u_1, \ldots, u_n$ die Koordinaten eines Vektors $x \in \mathbb{R}^n$ bzgl. der Eigenvektorbasis einer Matrix A, so folgt insbesondere

$$A\, x = A \left(\sum_{k=1}^{n} u_k\, b_k \right) = \sum_{k=1}^{n} u_k\, A\, b_k = \sum_{k=1}^{n} u_k\, \lambda_k\, b_k, \tag{5.156}$$

d.h. die Koordinaten von $A\, x$ bzgl. der Eigenvektorbasis sind $\lambda_1\, u_1, \ldots, \lambda_n\, u_n$.

Auf die Berechnung von Eigenwerten und Eigenvektoren gehen wir hier nicht näher ein. Diese ist manuell recht mühsam und wird üblicherweise mit dem Computer erledigt, insbesondere da die Matrizen bei praktisch relevanten Aufgabenstellungen viel zu groß für manuelles Rechnen sind.

Nebenbemerkung

Für nicht symmetrische quadratische Matrizen kann man die Begriffe “Eigenwert” und “Eigenvektor” ebenfalls einführen. Allerdings existieren die Eigenwerte dann nicht zwingend als reelle Zahlen. Dehnt man den Zahlenbereich auf komplexe Zahlen aus, so funktioniert alles im Wesentlichen wieder wie bei symmetrischen Matrizen. Allerdgins findet man nicht zwingend eine ONB aus Eigenvektoren, jedoch stets eine Basis.

Für nicht quadratische Matrizen $A \in \mathbb{R}^{m\times n}$ sind Eigenwerte und Eigenvektoren aufgrund der unterschiedlichen Länge der Vektoren $x \in \mathbb{R}^n$ und $A\,x \in \mathbb{R}^m$ nicht sinnvoll einführbar. Es gibt allerdings ein sehr ähnliches Konstrukt, welches als **Singulärwertzerlegung** in der Literatur zu finden ist. Man bedient sich hier zweier an die Matrix angepasster Koordinatensysteme, eins in $\mathbb{R}^n$ und eins in $\mathbb{R}^m$, und kann die Matrix-Vektor-Multiplikation dann wieder als elementweise Multiplikation mit den sogenannten Singulärwerten der Matrix darstellen. Für symmetrische Matrizen entsprechen die Singulärwerte gerade den Eigenwerten.

5.7.3 Anwendungen

Gehen kurz auf einige Anwendungen von Eigenwerten und Eigenvektoren ein. Es zeigt sich, dass Eigenwerte und -vektoren sehr relevante Interpretationen in verschiedensten praktischen Problemen habe. Entsprechend häufig wird man ihnen in der Praxis begegnen; teils auch ohne dies zu bemerken.

5.7.3.1 Hauptkomponentenanalyse Im heutigen Datenzeitalter wird alles an Daten gesammelt, was verfügbar ist. Dies führt zu der Frage, wie man relevante Informationen aus umfangreichen Datensammlungen extrahiert. Sollen beispielsweise klimatische Prozesse untersucht werden, so wird man zunächst alle möglichen Parameter aufzeichnen ohne im Vorfeld detailliert zu klären, ob diese überhaupt relevant sind: Temperaturen in verschiedenen Höhen und an verschiedenen Orten, Luftfeuchte, Bewölkung, Sonnenstand, Windgeschwindigkeiten usw. (alles, was irgendwie messbar ist). Die Messungen könnten zum Beispiel stündlich oder täglich stattfinden. Am Ende stehen eine Vielzahl von sogenannten Datenpunkten zur Verfügung. Jeder Datenpunkt entspricht der Messung zu einem Zeitpunkt und enthält (als Koordinaten bzgl. der Standardbasis) die Messwerte (Temperatur, Luftfeuchte usw.). Bei n Messungen von m verschiedenen Parametern enstehen so n Datenpunkte $x_1, \dots, x_n \in \mathbb{R}^m$.

Die Anzahl m der Parameter ist meist sehr groß (100 und mehr; gelegentlich auch im Bereich von Millionen). Auch die Anzahl n der Messungen ist stets sehr groß (oft ebenfalls im Bereich von Millionen). Wie kann man anhand dieser Datenmasse nun sehen, welche Parameter oder Parameterkombination (!) eigentlich relevant sind? Dabei sehen wir einen Parameter bzw. eine gewichtete Kombination von Parametern als relevant an, wenn die entsprechenden Messwerte einen vergleichsweise großen Wertebereich aufweisen. Parameter(-kombinationen), für die bei jeder Messung fast der gleiche Wert gemessen wird, sind hingegen nicht relevant.

Die Antwort findet man mit der so genannten Hauptkomponentenanalyse. Aus mathematischer Sicht werden folgende Schritte ausgeführt:

1. Schreibe die Messpunkte $x_k \in \mathbb{R}^m$ als Zeilen in eine Matrix $X \in \mathbb{R}^{n\times m}$.
2. Berechne die Eigenwerte und Eigenvektoren zur symmetrischen Matrix $X^{\mathrm{T}}\,X \in$

$\mathbb{R}^{m\times m}$.

3. Wähle die Eigenvektoren zu den größten Eigenwerten aus; b_1 sei der Eigenvektor zum größten Eigenwert, b_2 der zum zweitgrößten usw.
4. Bilde die Skalarprodukte $x_k \circ b_1, x_k \circ b_2, \ldots$ für alle Datenpunkte x_k und die gewählten Eigenvektoren. Jedes Skalarprodukt generiert eine Parameterkombination aus dem Datenpunkt x_k. Diese so aus den ursprünglichen Parametern erzeugten Parameterkombinationen haben den höchstmöglichen Informationsgehalt, der aus der gesamten Datenmenge extrahiert werden kann. Insbesondere sind auch bei großer Parameteranzahl m meist nur sehr wenige Eigenvektoren nötig, um alle relevanten Informationen darzustellen.

Um einzusehen, dass dieses Vorgheen sinnvoll ist, benötigt man einiges an Statistik-Wissen. Wir verweisen an dieser Stelle auf weiterführende Lehrveranstaltungen. Siehe auch Principal component analysis (der deutsche Wikipedia-Artikel zum Thema ist leider qualitativ etwas unterdurchschnittlich, Stand: 23.08.2026).

5.7.3.2 Schwingungsanalysen In Elektrotechnik, Mechanik und anderen Teilgebieten von Physik und Technik werden unter anderem Schwingungsprozesse untersucht. Insbesondere geht man der Frage nach, bei welchen Erregerfrequenzen Resonanz auftritt. Manchmal ist Resonanz erwünscht (Akustik, elektromagnetische Wellen,...), meist jedoch nicht (z.B. bei Brücken, vgl. Video der Zerstörung der Tacoma Narrows Bridge (via Wikipedia)).

Das Verhalten vieler technischer Systeme mit schwingenden Komponenten lässt sich in Form von linearen Gleichungssystemen modellieren und/oder simulieren. Die Eigenfrequenzen des untersuchten Systems, also die Frequenzen, die bei entsprechender Fremderregung zu Resonanz führen, sind dann gerade die Eigenwerte der Systemmatrix des LGS. Die zugehörigen Eigenvektoren enthalten zusätzliche Informationen über die meist recht komplexen Schwingungsvorgänge.

5.7.3.3 PageRank-Algorithmus Der PageRank-Algorithmus bildete die Grundlage der ursprünglichen Google-Suchmaschine. Bei diesem werden die Verlinkungen zwischen verschiedenen Webseiten genutzt, um auf die Wichtigkeit jeder einzelen Webseite und damit auf die Sortierung von Suchergebnissen zu schließen. Schreibt man alle Gewichte (je größer das Gewicht, desto wichtiger die Webseite) in einen langen Vektor $x \in \mathbb{R}^n$, wobei n die Anzahl der der Suchmaschine bekannten Webseiten ist, so kann man sich überlegen, dass

$$A\,x = x \tag{5.157}$$

mit einer gewissen Matrix $A \in \mathbb{R}^{n\times n}$ gelten muss. In dieser Matrix ist hinterlegt, welche Webseite auf welche Website verlinkt (vereinfacht: eine Eins in Zeile i und Spalte j heißt, dass Webseite i auf Webseite j verlinkt; eine Null steht für "keine Verlinkung"). Je mehr (wichtige) Webseiten auf eine gewisse Webseite verlinken, desto wichtiger ist die Seite selbst.

Die obige Gleichung sagt aus, dass der gesuchte Gewichtsvektor ein Eigenvektor der Link-Matrix A zum Eigenwert 1 ist. Die ganze Frage des Page-Rankings, also der Sortierung von Suchergebnissen nach Wichtigkeit, ist aus mathematischer Sicht also “nur” ein bisschen Matrix-Vektor-Rechnung.

6 Differentialrechnung

Grundlagen der Differentialrechnung für Funktionen einer Veränderlichen und Anwendung in der Fehlerrechnung. Weitere Anwendungen folgen im zweiten Semester.

6.1 Idee und Begriffe

Betrachten in diesem Kapitel ausschließlich Funktionen $f : (a, b) \to \mathbb{R}$, die auf einem offenen Intervall (a, b) definiert sind. Dabei lassen wir auch $(-\infty, \infty)$, also ganz $\mathbb{R}$, als Definitionsbereich zu. Der Vorteil eines offenen Intervalls als Definitionsbereich liegt darin, dass $x \in (a, b)$ nie am Rand des Intervalls liegt, sondern stets auf beiden Seiten der Stelle x noch ein Stückchen Definitionsbereich vorhanden ist. Man sagt auch: alle Punkte eines offenen Intervalls sind innere Punkte.

Mit dem Begriff der Stetigkeit kennen wir bereits eine Möglichkeit, Eigenschaften einer Funktion in der Umgegung einer Stelle $x \in (a, b)$ als Eigenschaft an dieser Stelle zu formulieren. Wir erinnern uns: Die Frage, ob eine Funktion an einer Stelle x stetig ist, kann nur beantwortet werden, wenn man auch die Funktionswerte in der Nähe von x anschaut. Konkret geschieht dies über den Begriff des Grenzwerts.

Die Information, die uns der Stetigkeitsbegriff liefert (Existenz von Sprung- oder Polstellen), ist für manche Einsatzzwecke etwas dürftig. Deshalb werden wir die Idee des Formulierens von Umgebungseigenschaften als Eigenschaft an einer konkreten Stelle nun für die Frage umsetzen, ob die Funktion an der betrachteten Stelle "glatt" ist, also dort keinen "Knick" hat.

Wir führen zunächst die nötigen Begriffe ein. Erst am Ende dieses Kapitels folgt dann die Einordnung in größere Zusammenhänge, also die Antwort auf die Frage "Wozu brauche ich das?".

6.1.1 Differenzierbarkeit an einer Stelle

Die Eigenschaft, die wir eben noch als "glatt" oder "ohne Knick" beschrieben haben, heißt Differenzierbarkeit.

Merke

Eine Funktion $f : (a, b) \to \mathbb{R}$ heißt **differenzierbar an einer Stelle** $x_0 \in (a, b)$, wenn der Grenzwert

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} \tag{6.1}$$

existiert.

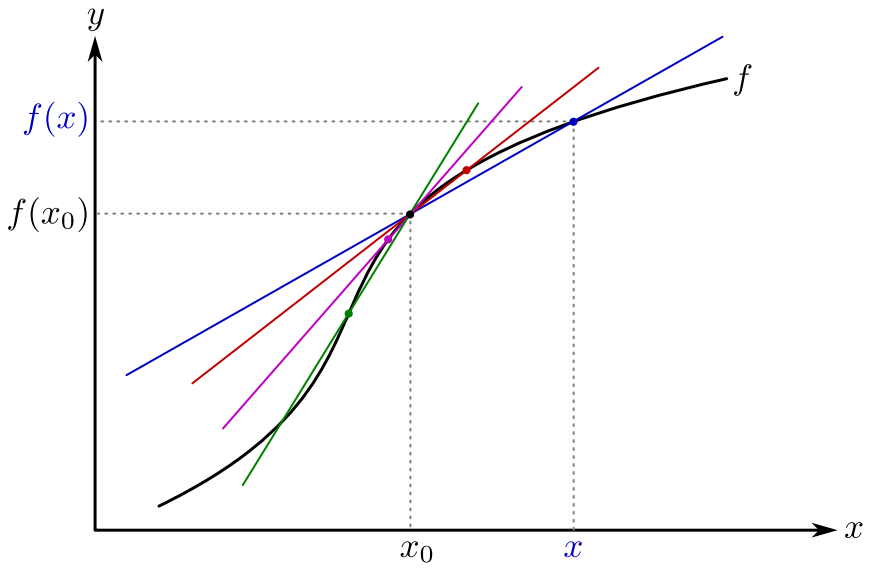
Der Ausdruck $\frac{f(x)-f(x_0)}{x-x_0}$ ist der Anstieg der Gerade durch die beiden Punkte $(x, f(x))$ und $(x_0, f(x_0))$. Der Grenzwert existiert bekanntlich genau dann, wenn die einseitigen Grenzwerte

$$\lim_{x \to x_0 - 0} \frac{f(x) - f(x_0)}{x - x_0} \quad \text{und} \quad \lim_{x \to x_0 + 0} \frac{f(x) - f(x_0)}{x - x_0} \tag{6.2}$$

existieren und übereinstimmen. Differenzierbarkeit bedeutet anschaulich also, dass die Steigung der Funktion links und rechts der betrachteten Stelle gleich ist (und natürlich

im Sinne des Grenzwerts überhaupt erst einmal existiert). Beispiele und Gegenbeispiele folgen weiter unten.

Der Ausdruck $\frac{f(x)-f(x_0)}{x-x_0}$ wird auch als **Differenzenquotient** bezeichnet, der zugehörige Grenzwert als **Differentialquotient**.



Gezeigt sind einige der Geraden, deren Anstieg beim Grenzübergang in der Definition der Differenzierbarkeit eine Rolle spielt.

Merke

Ist f in x_0 differenzierbar, so ist f dort auch stetig.

Differenzierbarkeit kann man somit als eine strengere Form von Stetigkeit ansehen. Es wird sich aber zeigen, dass das zu kurz gedacht ist. Der Begriff der Differenzierbarkeit eröffnet ganz neue Möglichkeiten und ist in der Ingenieurmathematik unverzichtbares Werkzeug sowohl für die Modellierung als auch für das Lösen praktischer Aufgabenstellungen (siehe unten).

Dass aus der Differenzierbarkeit die Stetigkeit folgt, sieht man leicht ein: Offensichtlich gilt

$$|f(x) - f(x_0)| = |x - x_0| \, \frac{|f(x) - f(x_0)|}{|x - x_0|}. \tag{6.3}$$

Der Faktor $|x - x_0|$ geht gegen Null, wenn $x \to x_0$. Der zweite Faktor konvergiert gegen eine feste Zahl. Folglich konvergiert das Produkt gegen Null, also $f(x) \to f(x_0)$.

Nebenbemerkung

Beachte, dass das Erweitern mit $|x - x_0|$ uns in die unschöne Situation bringt, dass wir eventuell durch Null teilen (bei $x = x_0$). Diesen Fall müsste man bei ganz strengem Vorgehen ausführlicher diskutieren. Das ist aber ein lösbares, eher technisches Problem, welches uns keine neuen Erkenntnisse bringt. Wir ersparen uns hier die Details und glauben den Mathematikern, dass hier alles seine Ordnung hat.

6.1.2 Ableitung

Analog zur Stetigkeit sind insbesondere solche Funktionen von Interesse, die an jeder Stelle differenzierbar sind.

Merke

Eine Funktion $f : (a, b) \to \mathbb{R}$ heißt **differenzierbar**, wenn sie an jeder Stelle ihres Definitionsbereichs differenzierbar ist.

Ist f differenzierbar, so ist die Funktion

$$x \mapsto \lim_{\tilde{x} \to x} \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x} \tag{6.4}$$

wohldefiniert auf (a, b). Diese Funktion heißt **Ableitung** von f und wird mit f' bezeichnet. Für die Funktionswerte $f'(x)$ schreibt man manchmal auch $\frac{\mathrm{d}f(x)}{\mathrm{d}x}$ oder $\frac{\mathrm{d}f}{\mathrm{d}x}(x)$.

In der Ingenieurmathematik hält sich im Umfeld von Differenzierbarkeit und Ableitungen hartnäckig der Begriff **Differential**. Meist wird dieser nur sehr diffus als “unendlich kleine Zahl, die nicht Null ist” oder als “infinitesimal kleine Zahl” eingeführt, was offensichtlich Unfug ist. Der Begriff stammt aus der Zeit, als die Differentialrechnung noch nicht vollständig ausgearbeitet und verstanden war, also aus der Zeit vor 1900. Das in der Ingenieurmathematik leider nach wie vor verbreitete “Rechnen mit Differentialen” hat wenig mit mathematischer Exaktheit zu tun, sondern ist mehr als “intuitives Formelbasteln” anzusehen. Tatsächlich gibt es auch in der modernen Mathematik den Differential-Begriff. Der ist seit über 100 Jahren präzise und sinnvoll definiert, hat für uns aber keinerlei Bedeutung oder Nutzen. Deshalb meiden wir diesen Begriff und seine recht sperrige Definition hier.

Merke

Eine differenzierbare Funktion heißt **stetig differenzierbar**, wenn ihre Ableitung

stetig ist.

Beachte: Differenzierbare Funktionen sind immer stetig, aber die Ableitung kann durchaus unstetig sein (vgl. Beispiele unten).

6.1.3 Beispiele

Das praktische Berechnen von Ableitungen werden wir im nächsten Kapitel betrachten. Das Vorgehen wird deutlich bequemer sein als es hier zunächst den Anschein hat. Insbesondere werden wir ohne das Bilden von Grenzwerten auskommen. Hier zunächst ein paar elementare Beispiele zur Illustration der Technik.

Beispiel (konstante Funktion)

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := c \quad \text{für ein } c \in \mathbb{R} \tag{6.5}$$

ist differenzierbar wegen

$$\lim_{\tilde{x}\to x} \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x} = \lim_{\tilde{x}\to x} \frac{c - c}{\tilde{x} - x} = \lim_{\tilde{x}\to x} 0 = 0. \tag{6.6}$$

Die Ableitung $f' : \mathbb{R} \to \mathbb{R}$ ist entsprechend gegeben durch

$$f'(x) = 0. \tag{6.7}$$

Beispiel (lineare Funktion)

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := c\,x \quad \text{für ein } c \in \mathbb{R} \tag{6.8}$$

ist differenzierbar wegen

$$\lim_{\tilde{x}\to x} \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x} = \lim_{\tilde{x}\to x} \frac{c\,\tilde{x} - c\,x}{\tilde{x} - x} = \lim_{\tilde{x}\to x} c = c. \tag{6.9}$$

Die Ableitung $f' : \mathbb{R} \to \mathbb{R}$ ist entsprechend gegeben durch

$$f'(x) = c. \tag{6.10}$$

Beispiel (quadratische Funktion)

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := x^2 \tag{6.11}$$

ist differenzierbar wegen

$$\begin{aligned} \lim_{\tilde{x}\to x} \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x} &= \lim_{\tilde{x}\to x} \frac{\tilde{x}^2 - x^2}{\tilde{x} - x} \\ &= \lim_{\tilde{x}\to x} \frac{(\tilde{x} + x)\,(\tilde{x} - x)}{\tilde{x} - x} \\ &= \lim_{\tilde{x}\to x} (\tilde{x} + x) = 2\,x. \end{aligned} \tag{6.12}$$

Die Ableitung $f' : \mathbb{R} \to \mathbb{R}$ ist entsprechend gegeben durch

$$f'(x) = 2\,x. \tag{6.13}$$

Beispiel (Betragsfunktion)

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := |x| \tag{6.14}$$

ist für $x \neq 0$ differenzierbar mit Ableitung

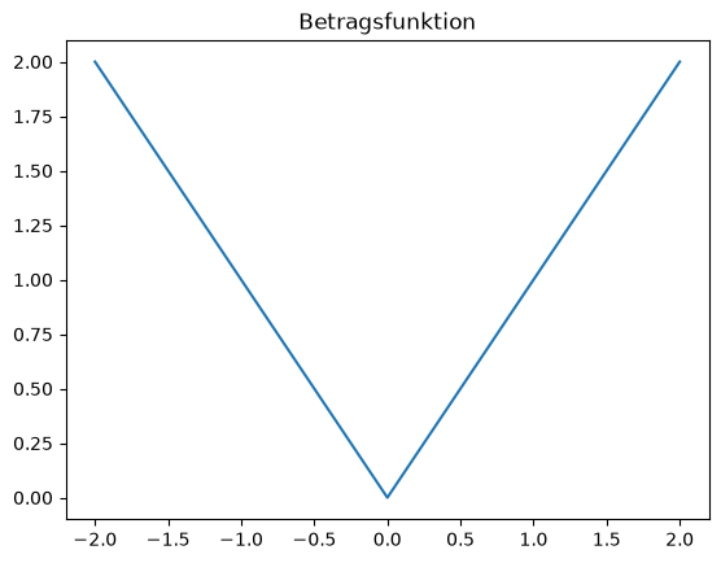
$$f'(x) = \begin{cases} -1, & \text{für } x < 0, \\ 1, & \text{für } x > 0 \end{cases} \tag{6.15}$$

(vgl. obiges Beispiel zu linearen Funktionen).

An der Stelle $x = 0$ ist die Funktion nicht differenzierbar denn links- und rechtsseitiger Grenzwert existieren zwar, stimmen aber nicht überein:

$$\begin{aligned} \lim_{\tilde{x}\to 0-0} \frac{f(\tilde{x}) - f(0)}{\tilde{x} - 0} &= \lim_{\tilde{x}\to 0-0} \frac{-\tilde{x}}{\tilde{x}} = \lim_{\tilde{x}\to 0-0} -1 = -1, \\ \lim_{\tilde{x}\to 0+0} \frac{f(\tilde{x}) - f(0)}{\tilde{x} - 0} &= \lim_{\tilde{x}\to 0+0} \frac{\tilde{x}}{\tilde{x}} = \lim_{\tilde{x}\to 0+0} 1 = 1. \end{aligned} \tag{6.16}$$

Man sagt auch: die Funktion hat einen Knick bei $x = 0$.



Beispiel (Wurzelfunktion)

Man kann zeigen (tun wir hier nicht), dass die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := \sqrt{|x|} \tag{6.17}$$

für $x \neq 0$ differenzierbar ist und die Ableitung $f' : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ durch

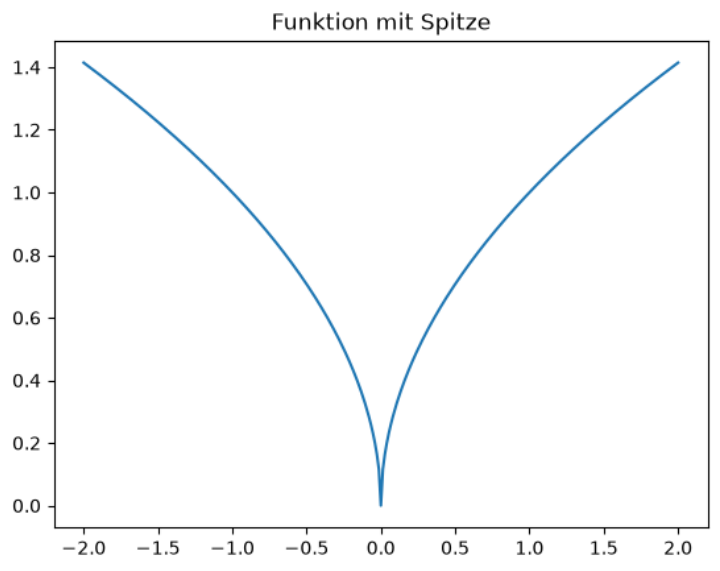
$$f'(x) = \frac{\sqrt{|x|}}{2\,x} \tag{6.18}$$

gegeben ist.

An der Stelle $x = 0$ ist die Funktion nicht differenzierbar denn links- und rechtsseitiger Grenzwert existieren nur im Sinne von $\pm\infty$ und stimmen nicht überein:

$$\begin{aligned} \lim_{\tilde{x}\to 0-0} \frac{f(\tilde{x}) - f(0)}{\tilde{x} - 0} &= \lim_{\tilde{x}\to 0-0} \frac{\sqrt{-\tilde{x}}}{\tilde{x}} = \lim_{\tilde{x}\to 0-0} \frac{1}{-\sqrt{-\tilde{x}}} = -\infty, \\ \lim_{\tilde{x}\to 0+0} \frac{f(\tilde{x}) - f(0)}{\tilde{x} - 0} &= \lim_{\tilde{x}\to 0+0} \frac{\sqrt{\tilde{x}}}{\tilde{x}} = \lim_{\tilde{x}\to 0+0} \frac{1}{\sqrt{\tilde{x}}} = \infty. \end{aligned} \tag{6.19}$$

Man sagt auch: die Funktion hat eine Spitze bei $x = 0$.



Beispiel (unstetige Ableitung)

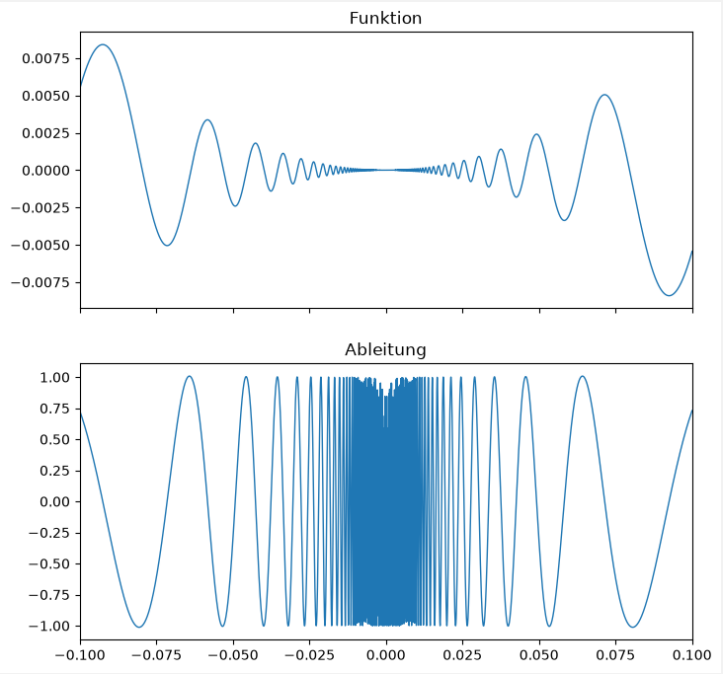
Man kann zeigen (tun wir hier aber nicht), dass die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := \begin{cases} x^2 \sin \frac{1}{x}, & \text{für } x \neq 0, \\ 0, & \text{für } x = 0 \end{cases} \tag{6.20}$$

überall differenzierbar ist, die Ableitung

$$f'(x) = \begin{cases} 2\,x \sin \frac{1}{x} - \cos \frac{1}{x}, & \text{für } x \neq 0, \\ 0, & \text{für } x = 0 \end{cases} \tag{6.21}$$

aber in $x = 0$ nicht stetig ist.



Die Menge der differenzierbaren Funktionen ist also eine echte Obermenge der Menge der stetig differenzierbaren Funktionen.

Nebenbemerkung (nirgends differenzierbare stetige Funktionen)

Da man in der Praxis fast ausschließlich differenzierbare Funktionen antrifft oder wenigstens solche, die nur an einzelnen Stellen nicht differenzierbar sind, könnte der Eindruck entstehen, dass die Mehrheit der stetigen Funktionen fast überall differenzierbar ist. Dieser Schluss ist jedoch falsch! Betrachtet man beispielsweise die Menge aller stetigen Funktionen, die von $(0, 1)$ nach $\mathbb{R}$ abbilden, so zeigt sich, dass die große Mehrheit nirgends (!) differenzierbar ist. Unser Blick auf die Welt der Funktionen ist so eng begrenzt, dass wir große Mühe haben, wenigstens ein Beispiel dieser großen Mehrheit von stetigen, aber nirgends differenzierbaren Funktionen zu finden. Diese lassen sich nur bedingt in Formeln aufschreiben. Ein Beispiel ist die Weierstraß-Funktion. In dieser Quelle findet man auch die präzise Formulierung für “große Mehrheit”; diese benötigt allerdings Begriffe, die wir bisher nicht eingeführt haben und auch nicht einführen werden.

Die Frage, ob die Natur differenzierbare Funktionen bevorzugt oder ob nicht der Mensch beim Modellieren diese künstlich als Vereinfachung der tatsächlichen Verhältnisse hinzufügt, ist bedenkenswert!

6.1.4 Höhere Ableitungen

Da die Ableitung einer differenzierbaren Funktion selbst wieder eine Funktion ist, kann man auch für die Ableitung die Frage der Differenzierbarkeit stellen.

Merke

Eine Funktion f heißt **k-mal differenzierbar**, wenn $(k-1)$-mal die Ableitung der Ableitung gebildet werden kann, also insbesondere die (erste) Ableitung wieder differenzierbar ist, die Ableitung der Ableitung differenzierbar ist usw. Die Ableitungen werden mit $f', f'', f''', f^{(4)}, \ldots, f^{(k)}$ bezeichnet. Offensichtlich gilt

$$f^{(l)} = \left(f^{(l-1)}\right)' \quad \text{für } l = 1, \ldots, l \tag{6.22}$$

(wobei man mit $f^{(0)}$ üblicherweise die Funktion f selbst meint).

Merke

Eine Funktion heißt **k-mal stetig differenzierbar**, wenn sie k-mal differenzierbar ist und die k-te Ableitung stetig ist.

6.1.5 Interpretation

Aus der Definition der Ableitung ist sofort klar, dass die Ableitung einer Funktion an einer Stelle x die **Steigung der Funktion** an dieser Stelle angibt. Bemerkenswert ist jedoch, dass der Begriff "Steigung" eigentlich nur für Geraden sinnvoll ist, also mindestens zwei Punkte benötigt, zwischen denen die Steigung, also das Verhältnis von vertikaler und horizontaler Verschiebung, betrachtet wird. Durch den Grenzprozess beim Bilden der Ableitung wird die Steigung der Funktion in der Umgebung der betrachteten Stelle x als Eigenschaft an dieser Stelle x ausgedrückt. Dies ist wieder ein Beispiel für starke Abstraktion, die durch Gewöhnung mit der Zeit in die ganz normale, scheinbar intuitiv klare Begriffswelt übergeht.

Die zweite Ableitung einer Funktion gibt die Änderungsrate des Anstiegs an einer Stelle an. Auch hier ist "Änderungsrate" eigentlich nur sinnvoll erklärbar, wenn man eine Umgebung einer Stelle anschaut und nicht nur die Stelle selbst. Durch den Grenzprozess wird die relevante Information von der Umgebung einer Stelle auf die Stelle selbst übertragen. Die Änderungsrate der Steigung kann man als **Krümmung des Funktionsgraphs** interpretieren. Wächst der Anstieg, so krümmt sich der Graph nach links (wenn man ihn gedanklich von links nach rechts durchläuft); sinkt der Anstieg, so krümmt sich der Graph nach rechts. Funktionen, deren zweite Ableitung überall positiv ist, heißen **konvex**; ist die zweite Ableitung überall negativ, so heißt die Funktion **konkav**.

Die dritte Ableitung liefert Informationen darüber, wie sich die Krümmung der Funktion

ändert. Ist die dritte Ableitung in einem Bereich positiv, so nimmt in diesem Bereich die Krümmung zu, d.h. die Kurve, die der Funktionsgraph beschreibt, wird enger; bei negativer dritter Ableitung wird sie weiter.

Ab der vierten Ableitung wird die geometrische Interpretation schwieriger. Auch ist eine geometrische Interpretation hier kaum relevant.

6.1.6 Einordnung

Sofern in der Schulmathematik bereits über Differenzierbarkeit und Ableitungen gesprochen wurde, könnte der Eindruck entstanden sein, dass die Hauptzwecke dieser Begriffe das Untersuchen von Funktionen ("Kurvendiskussion") und das Lösen von Optimierungsproblemen sind. Auch in der Grundlagenmathematik im Studium könnte dieser Eindruck entstehen. Um diesen Trugschluss zu vermeiden, seien hier ein paar Hinweise zu Bedeutung und Entstehung des Differenzierbarkeitsbegriffs gegeben.

Die Väter der Differentialrechnungen waren Newton und Leibniz (beide 17. Jahrhundert). Man wollte nach dem "dunklen", durch Spiritualität und Unterdrückung von Wissenschaft geprägten Mittelalter die Welt verstehen und in Formeln beschreiben, hatte aber bis auf Zahlen, Funktionen und geometrische Figuren keine Werkzeuge dafür. Insbesondere Gesetzmäßigkeiten bei Bewegungsabläufen (freier Fall, Planetenbewegungen,...) ließen sich damit nicht zufriedenstellend formulieren. Erst die für diesen Zweck entwickelte Differentialrechnung erlaubte das Formulieren der bis heute (im Wesentlichen unverändert) im Einsatz befindlichen klassischen Mechanik. In diesen Bereich fällt insbesondere das allgemein bekannte Newtonsche Gesetz:

$$\text{Kraft} = \text{Masse} \cdot \text{Beschleunigung}.$$

Bewegt sich ein Körper mit Masse m im Raum entlang einer durch

$$t \mapsto \big(x(t), y(t), z(t)\big), \quad t \in [a, b], \tag{6.23}$$

beschriebenen Kurve unter Einwirkung einer zeitlich und örtlich variierenden Kraft

$$F : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R} \tag{6.24}$$

(jeder Zeit und jedem Ort wird ein Kraftvektor zugeordnet), so kann das Newtonsche Gesetz mittels Ableitungen der Koordinatenfunktionen als

$$F\big(t, (x(t), y(t), z(t))\big) = m \begin{bmatrix} x''(t) \\ y''(t) \\ z''(t) \end{bmatrix} \quad \text{für alle } t \in (a, b) \tag{6.25}$$

formuliert werden. Newton hatte erkannt, dass dieser Zusammenhang stets gilt. Kennt man die wirkenden Kräfte und die Masse m, so kann man aus dieser Gleichung die Bahn berechnen, auf der sich der Körper bewegen wird.

Die gezeigte Form des Newtonschen Gesetzes ist ein **System von gewöhnlichen Differentialgleichungen**. Differentialgleichungen sind Gleichungen, die eine Funktion in Beziehung zu ihren Ableitungen setzt. Das Attribut "gewöhnlich" wird genutzt, wenn es sich um Funktionen einer Veränderlichen handelt. Da beim Newtonschen Gesetz vektorwertige Funktionen links und rechts stehen, handelt es sich eigentlich um drei Differentialgleichungen, also ein Differentialgleichungssystem.

Später werden wir die Differentialrechnung auf Funktionen mehrerer Veränderlicher ausdehnen. Die damit formulierbaren Differentialgleichungen werden als **partielle Differentialgleichungen** bezeichnet.

In unserer Veranstaltung spielen Differentialgleichungen keine Rolle. Dennoch bilden diese den Haupteinsatzzweck für die Differentialrechnung. Mit Differentialgleichungen lassen sich nahezu alle Phänomene aus Naturwissenschaften, Technik und auch Wirtschaftswissenschaften beschreiben: Planetenbewegungen, Wachstumsprozesse, Wärmeleitprozesse, Strömungen, die Entwicklung von Aktienkursen usw.

Eher nebenbei ist die Differentialrechnung sehr nützlich für das Lösen von Optimierungsproblemen, sowohl beim manuellen Lösen als auch beim Lösen mit dem Computer. Beispielsweise basieren die heute üblichen Verfahren zum Training von KI ganz wesentlich auf Resultaten der Differentialrechnung (wir kommen im zweiten Semester darauf zurück).

6.2 Vorgehen

Vor der flächendeckenden Einführung des Computers mussten Ableitungen per Hand berechnet werden. Entsprechend umfangreich und weit entwickelt sind die Verfahren dazu. Wir werden hier nur einen Überblick über das prinzipielle Vorgehen geben und, da heute deutlich relevanter, das Differenzieren mit dem Computer betrachten.

6.2.1 Manuelles Differenzieren

Der Weg über die Definition der Ableitung einer Funktion als Grenzwert ist recht mühsam für den praktischen Einsatz. Analog zur Berechnung von Grenzwerten bei Zahlenfolgen und Funktionen entnimmt man deshalb die Ableitungen einiger Grundfunktionen aus Tabellen und kombiniert diese dann anhand verschiedener Ableitungsregeln zu Ableitungen komplexerer Funktionen. Zunächst die Ableitungen einiger immer wieder auftretender Funktionen (gültig für alle x, für die f und f' sinnvolle Werte liefern):

$$f(x) = x^c,\ c \in \mathbb{R} \quad \Rightarrow \quad f'(x) = c\,x^{c-1}, \tag{6.26}$$

$$f(x) = \cos x \quad \Rightarrow \quad f'(x) = -\sin x, \tag{6.27}$$

$$f(x) = \sin x \quad \Rightarrow \quad f'(x) = \cos x, \tag{6.28}$$

$$f(x) = \mathrm{e}^x \quad \Rightarrow \quad f'(x) = \mathrm{e}^x, \tag{6.29}$$

$$f(x) = \ln x \quad \Rightarrow \quad f'(x) = \frac{1}{x}. \tag{6.30}$$

Beachte, dass die Exponentialfunktion beim Ableiten unverändert bleibt! Dinge, die trotz des Anwendens irgendwelcher Operationen unverändert bleiben, spielen stets eine besondere Rolle. So taucht die Exponentialfunktion beispielsweise sehr oft in den Lösungen von Differentialgleichungen auf, spielt aber auch in der Wahrscheinlichkeitsrechnung eine wichtige Rolle.

Mittels folgender Ableitungsregeln, deren Gültigkeit wir hier nicht nachweisen, können Ableitungen komplexerer Funktionen auf Ableitungen einfacher Funktionen zurückgeführt werden.

- Ist f an einer Stelle x differenzierbar und ist $\lambda \in \mathbb{R}$ eine Zahl, so ist auch $\lambda\, f$ bei x differenzierbar und es gilt

$$(\lambda\, f)'(x) = \lambda\, f'(x). \tag{6.31}$$

- Sind f und g an einer Stelle x differenzierbar, so ist auch die Summe $f + g$ dort differenzierbar und es gilt

$$(f + g)'(x) = f'(x) + g'(x) \tag{6.32}$$

  (**Summenregel**).

- Sind f und g an einer Stelle x differenzierbar, so ist auch das Produkt $f\,g$ dort differenzierbar und es gilt

$$(f\,g)'(x) = f'(x)\,g(x) + f(x)\,g'(x) \tag{6.33}$$

  (**Produktregel**).

- Sind f und g an einer Stelle x differenzierbar und gilt $g(x) \neq 0$, so ist auch der Quotient $\frac{f}{g}$ bei x differenzierbar und es gilt

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)\,g(x) - f(x)\,g'(x)}{g(x)^2} \tag{6.34}$$

  (**Quotientenregel**).

- Ist f an einer Stelle x differenzierbar und ist g an der Stelle $f(x)$ differenzierbar, so ist auch die Hintereinanderausführung $g \circ f$ bei x differenzierbar und es gilt

$$(g \circ f)'(x) = f'(x)\,g'(f(x)) \tag{6.35}$$

  (**Kettenregel**).

Beispiel (Ableitung eines Polynoms)

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := 2\,x^3 + 5\,x + 1 \tag{6.36}$$

ist überall differenzierbar und die Ableitung $f' : \mathbb{R} \to \mathbb{R}$ ist durch

$$f'(x) = 6\,x^2 + 5 \tag{6.37}$$

gegeben (Video 610, Notizen dazu).

Beispiel (Ableitung eines Produkts)

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := x^2\,\cos x \tag{6.38}$$

ist überall differenzierbar und die Ableitung $f' : \mathbb{R} \to \mathbb{R}$ ist durch

$$f'(x) = 2\,x\,\cos x - x^2\,\sin x \tag{6.39}$$

gegeben (Video 615, Notizen dazu).

Beispiel (Ableiten mittels Kettenregel)

Die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := \sin \mathrm{e}^{3\,x} \tag{6.40}$$

ist überall differenzierbar und die Ableitung $f' : \mathbb{R} \to \mathbb{R}$ ist durch

$$f'(x) = 3\,\mathrm{e}^{3\,x}\,\cos \mathrm{e}^{3\,x} \tag{6.41}$$

gegeben (Video 620, Notizen dazu).

6.2.2 Ableitungen mit SymPy

Das Bilden von Ableitungen mit SymPy ist denkbar einfach:

```
import sympy

x, c = sympy.symbols('x c')

sympy.diff((x + c) ** 2 * sympy.sin(x), x)
```

```
(c + x)**2*cos(x) + (2*c + 2*x)*sin(x)
```

6.2.3 Numerische Differentiation

Das numerische Differenzieren erfolgt direkt über die Definition der Ableitung:

$$f'(x) = \lim_{\tilde{x} \to x} \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x}. \tag{6.42}$$

Da man Grenzübergänge mit dem Computer nicht exakt umsetzen kann, wählt man $\tilde{x}$ nah bei x und betrachtet den zugehörigen Differenzenquotient als Näherung für die Ableitung. Für $\tilde{x} := x + h$ mit einem fest (also unabhängig von x) gewähltem, kleinen $h > 0$ erhält man also

$$f'(x) \approx \frac{f(x + h) - f(x)}{h}. \tag{6.43}$$

Diese Form der numerischen Differentiation bezeichnet man als **Vorwärtsdifferenz**. Da $\tilde{x}$ auch kleiner als x sein kann, ist

$$f'(x) \approx \frac{f(x - h) - f(x)}{-h} = \frac{f(x) - f(x - h)}{h}. \tag{6.44}$$

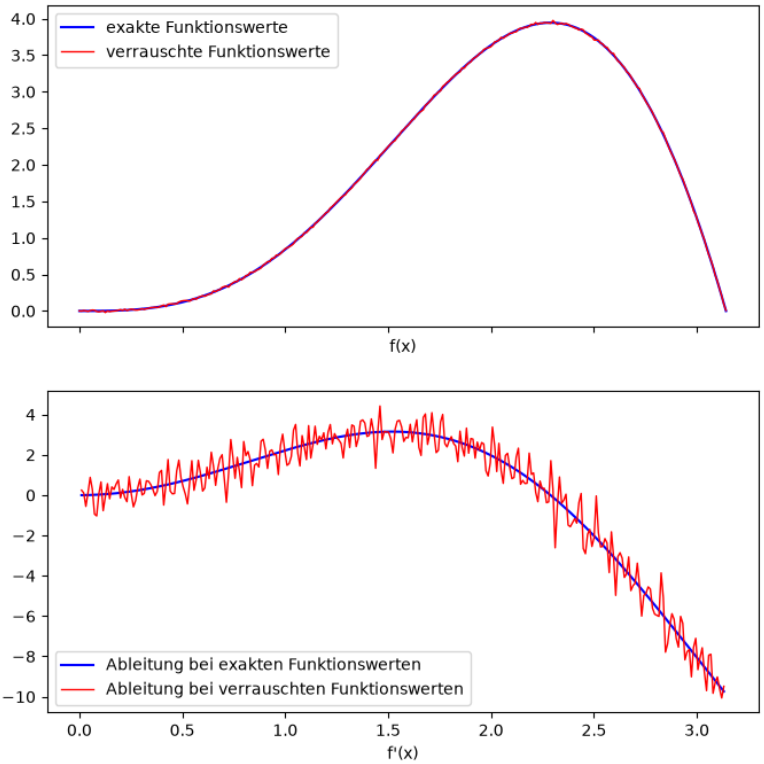
eine völlig gleichberechtigte, alternative Näherungsformel für die Ableitung (**Rückwärtsdifferenz**). Gibt es keinen Grund eine von beiden Varianten zu bevorzugen, so nutzt man gern deren

Mittelwert:

$$f'(x) \approx \frac{1}{2}\,\frac{f(x+h)-f(x)}{h} + \frac{1}{2}\,\frac{f(x)-f(x-h)}{h} = \frac{f(x+h)-f(x-h)}{2\,h} \tag{6.45}$$

(**zentrale Differenz**).

Auf den ersten Blick wird die Näherung für $f'(x)$ besser, je kleiner h gewählt wird. Jedoch zeigt sich in der Praxis, dass zu kleines h zu starken Oszillationen in der berechneten Näherung für f' führen kann. Ursache ist, dass kleine Fehler in den (oft aus Messungen stammenden) Werten für f durch das numerische Differenzieren verstärkt werden. Die Details dazu sind Thema der numerischen Mathematik und führen hier zu weit. Merke jedoch: **Numerisches Differenzieren ist nur mit großer Vorsicht zu benutzen!**



Im Bild sehen wir: Obwohl die verrauschten Funktionswerte kaum von den exakten Funktionswerten unterscheidbar sind, weichen die numerisch berechneten Ableitungen deutlich voneinander ab. Zum selbst ausprobieren: Erhöht man die Anzahl der Stützstellen, so wird die Abweichung der Ableitungen immer größer.

6.3 Taylor-Polynome

Als eine erste konkrete Anwendung der Differentialrechnung schauen wir uns hier die Annäherung von Funktionen durch Polynome an. Diese Technik hilft uns im nächsten Kapitel beim Verständnis der Fehlerrechnung. Weitere Anwendungen folgen im zweiten Semester.

6.3.1 Bedeutung

Es gibt verschiedene Verfahren zum Ersetzen (nahezu) beliebiger Funktionen durch Polynome mit sehr ähnlichem Verlauf des Funktionsgraphs. Das einfachste ist die Konstruktion sogenannter Taylor-Polynome.

Während man früher durch den Übergang zu Polynomen die manuelle Rechenarbeit vereinfachen wollte, liegt die Bedeutung der Approximation (d.h. Annäherung) von Funktionen durch Polynome heute vor allem in der Verarbeitung von Funktionen mit dem Computer. Während allgemeine Funktionen (z.B. Sinus, Exponentialfunktion, Logarithmusfunktion,...) beim numerischen Rechnen mit dem Computer nur als Sammlung endlich vieler Funktionswerte an gegebenen Stützstellen darstellbar sind, also nie exakt und vollständig vorliegen, können Polynome auch mit dem Computer exakt und vollständig beschrieben werden. Wir erinnern uns: Ein Polynom

$$a_n\, x^n + a_{n-1}\, x^{n-1} + \cdots + a_2\, x^2 + a_1\, x + a_0 \tag{6.46}$$

vom Grad n ist durch die $n+1$ Koeffizienten $a_0, a_1, \ldots, a_n$ beschrieben. Das sind endlich viele Zahlen und diese sind mit dem Computer (bis auf Rundungsfehler) darstellbar und verarbeitbar.

Wir werden Taylor-Polynome nur in sehr einfacher Form im nächsten Kapitel wiedersehen. Wann immer Sie jedoch mit dem Computer oder dem Taschenrechner beispielsweise den Sinus eines Winkels berechnen, werden im Hintergrund Taylor-Polynome (oder sehr ähnliche Techniken) verwendet.

6.3.2 Konstruktion

Sei $f : (a, b) \to \mathbb{R}$ eine Funktion, wobei auch ein- und beidseitig unendliche Intervalle als Definitionsbereich zulässig seien, und sei n eine gegebene natürliche Zahl. Gesucht ist ein Polynom $T_n : \mathbb{R} \to \mathbb{R}$ vom Grad n, welches einen ähnlichen Verlauf zeigt wie die Funktion f. Ein Polynom vom Grad n besitzt $n + 1$ Freiheitsgrade (d.h. frei wählbare Parameter). Wir können also davon ausgehen, dass wir $n + 1$ Bedingungen formulieren können, welche den Verlauf des Polynoms T_n festlegen. Zu einer gegebenen/gewählten Stelle $x_0 \in (a, b)$ fordern wir

$$T_n(x_0) = f(x_0), \quad T_n'(x_0) = f'(x_0), \quad \ldots, T_n^{(n)}(x_0) = f^{(n)}(x_0). \tag{6.47}$$

Klar ist, dass diese Forderungen nur formulierbar sind, wenn f wenigstens n-mal differenzierbar ist.

Die Ableitungen von T_n kann man explizit angeben. Man erhält so ein lineares Gleichungssystem für die $n+1$ Unbekannten $a_0, \ldots, a_n$ mit $n+1$ Gleichungen. Die Systemmatrix ist sogar schon in Stufenform, sodass das Lösen dieses LGS im Prinzip keine Probleme macht. Allerdings ist das Rückwärtseinsetzen recht fummelig und unterbleibt deshalb an dieser Stelle. Man erhält folgendes Ergebnis:

Merke

Sei $f : (a, b) \to \mathbb{R}$ mindestens n-mal differenzierbar und sei $x_0 \in (a, b)$. Dann heißt das Polynom

$$\begin{aligned} T_n(x) &:= \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(x_0)\,(x - x_0)^k \\ &= f(x_0) + f'(x_0)\,(x - x_0) + f''(x_0)\,(x - x_0)^2 + \\ &\qquad \cdots + \frac{1}{n!} f^{(n)}(x_0)\,(x - x_0)^n \end{aligned} \tag{6.48}$$

Taylor-Polynom vom Grad n zur Funktion f mit der **Entwicklungsstelle** x_0.

Das Taylor-Polynom vom Grad 1 wird auch als **Tangente** an die Funktion an der Stelle x_0 bezeichnet. Dies ist eine Gerade, die an der Stelle x_0 den gleichen Funktionswert und den gleichen Anstieg wie die Funktion f hat. Das Taylor-Polynom vom Grad 2 wird gelegentlich als **Schmiegparabel** bezeichnet. Neben Funktionswert und Anstieg stimmt hier auch die Krümmung mit der von f überein.

Beispiel

Zur Vereinfachung von Berechnungen soll die Sinusfunktion durch ein geeignetes Taylor-Polynom ersetzt werden. Wir entscheiden uns für den Grad $n = 4$, da Polynome vierten Grades manuell noch einigermaßen handhabbar sind. Als Entwicklungsstelle wählen wir $x_0 := \frac{\pi}{2}$, da wir uns für den Sinus von Winkeln in $[0, \pi]$ interessieren (aus der Konstruktion der Taylor-Polynome kann man erwarten, dass die Näherung um so schlechter wird, je weiter wir von der Entwicklungsstelle entfernt sind).

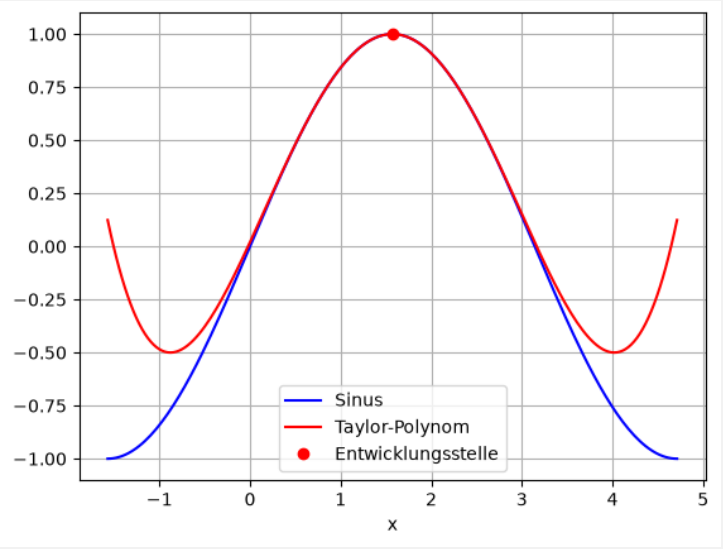
Die relevanten Ableitungen sind

$$\begin{aligned} f(x_0) &= \sin\frac{\pi}{2} = 1, \\ f'(x_0) &= \cos\frac{\pi}{2} = 0, \\ f''(x_0) &= -\sin\frac{\pi}{2} = -1, \\ f'''(x_0) &= -\cos\frac{\pi}{2} = 0, \\ f^{(4)}(x_0) &= \sin\frac{\pi}{2} = 1. \end{aligned} \tag{6.49}$$

Somit

$$T_4(x) = 1 - \frac{1}{2}\left(x - \frac{\pi}{2}\right)^2 + \frac{1}{24}\left(x - \frac{\pi}{2}\right)^4. \tag{6.50}$$

Diese Näherungsformel für die Sinusfunktion erlaubt das Berechnen von Sinuswerten unter Verwendung der vier Grundrechenarten.



6.3.3 Approximationsfehler

An der Entwicklungsstelle x_0 stimmen Funktion und Taylor-Polynom exakt überein. Mit wachsendem Abstand zur Entwicklungsstelle wird die Abweichung zwischen Funktion und Taylor-Polynom jedoch zunehmen. Frage: Können wir den Fehler, der durch das Ersetzen einer Funktion durch ein Taylor-Polynom entsteht, als konkrete Zahl angeben? Die Antwort ist ganz klar: Ja. Für das “Wie” gibt es verschiedenen Varianten. Wir geben hier nur eine an und verzichten auf die Herleitung.

Merke

Sei $f : (a, b) \to \mathbb{R}$ mindestens $(n + 1)$-mal stetig differenzierbar und sei T_n das Taylor-Polynom n-ten Grades von f zur Entwicklungsstelle $x_0 \in (a, b)$. Dann existiert zu jedem $x \in (a, b)$ eine Zahl $\xi \in \mathbb{R}$, die zwischen x und x_0 liegt und

$$f(x) - T_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) \, (x - x_0)^{n+1} \tag{6.51}$$

erfüllt.

Der Ausdruck auf der rechten Seite wird als **Lagrange-Restglied** oder Restglied in Lagrange-Form bezeichnet.

Insbesondere erhält man mit Hilfe des Lagrange-Restglieds die Fehlerabschätzung

$$|f(x) - T_n(x)| \leq \frac{1}{(n+1)!} \left(\max_{\xi \in (a,b)} \left| f^{(n+1)}(\xi) \right| \right) |x - x_0|^{n+1}. \tag{6.52}$$

Beachte hier, dass $f^{(n+1)}$ unbeschränkt sein kann, sodass das Maximum nicht existiert (bzw. ∞ ist).

Beispiel

Betrachten nochmal das Taylor-Polynom vierten Grades für die Sinusfunktion bei $x_0 = \frac{\pi}{2}$. Für $x \in [0, \pi]$ erhalten wir die Fehlerabschätzung

$$\begin{aligned} |\sin x - T_4(x)| &\leq \frac{1}{5!} \left(\max_{\xi \in [0,\pi]} |\cos \xi| \right) \left| x - \frac{\pi}{2} \right|^5 \\ &= \frac{1}{120} \left| x - \frac{\pi}{2} \right|^5 \leq \frac{1}{120} \left(\frac{\pi}{2} \right)^5 \leq 0.08, \end{aligned} \tag{6.53}$$

wobei die vorletzte Abschätzung aus $|x - \frac{\pi}{2}| \leq \frac{\pi}{2}$ folgt.

Somit stimmt bis zur ersten Nachkommastelle der mittels T_4 berechnete Näherungswert für $\sin x$ mit dem exakten Wert im Wesentlichen überein.

6.3.4 Ausblick

Wie eingangs erwähnt, gibt es diverse Möglichkeiten Polynome als Näherung für eine Funktion zu konstruieren. Taylor-Polynome konzentrieren sich auf eine Stelle und fordern dort die Gleichheit der Ableitungen bis zu einer gewählten Ordnung. Das andere Extrem sind **Interpolationspolynome**. Bei diesen fordert man die Gleichheit der Funktionswerte an mehreren Stellen. Möchte man ein Polynom vom Grad n konstruieren, fordert man also die Gleichheit der Funktionswerte an $n + 1$ Stellen.

Zwischen diesen beiden Extremen gibt es nahezu beliebige Mischformen. Eine in der Praxis sehr beliebte Variante sind sogenannte **Splines**. Grundidee ist, nicht ein Polynom für den gesamten Definitionsbereich zu bilden, sondern auf Teilintervallen zu arbeiten. Auf jedem Teilintervall wird ein eigenes Polynom (mit vorgegebenem Grad) konstruiert, sodass die einzelnen Polynome an den Intervallenden zusammenpassen, also gleiche Funktionswerte und gleiche Ableitungen wie die zu approximierende Funktion besitzen.

Für vektorwertige Funktionen und für Funktionen mehrerer Veränderlicher gibt es sehr ähnliche, auf Polynomen basierende Konstruktionen. Insbesondere sind die in der Computergrafik sehr beliebten Bézier-Kurven eine Form von Näherungspolynomen für Funktionen $f : [a, b] \to \mathbb{R}^2$.

6.4 Fehlerrechnung

Man steht sehr oft vor dem Problem, dass man eine interessierende Größe nicht direkt messen kann, sondern nur aus Berechnungen auf diese schließen kann. Ausgangspunkt für die nötigen Berechnungen sind gemessene Werte, die praktisch immer mit einem nur ungefähr bekannten Messfehler behaftet sind. Frage: Wie wirken sich solche Messfehler auf aus Messwerten berechnete Größen aus?

Beispiel

Um den Umfang eines Kreises (z.B. Radumfang bei einem Fahrzeug) zu ermitteln, bietet es sich an, den Durchmesser d des Rades zu messen und daraus den Umfang $u = \pi\, d$ zu berechnen. Liegt der Messfehler beim Durchmesser in etwa bei 1 mm, so wird der Fehler im berechneten Umfang aufgrund des Faktors π bei etwa 3 mm liegen.

6.4.1 Ziel

In obigem Beispiel ist der Zusammenhang von Messwert und berechnetem Wert sehr einfach und die **Fehlerfortpflanzung** beim Rechnen gut überschaubar. Wie aber bei komplexeren Zusammenhängen vorgehen? Wie wirkt sich beispielsweise die etwas ungenau gemessene Kantenlänge eines Würfels auf das daraus berechnete Volumen des Körpers aus?

Ziel dieses Kapitels ist die Entwicklung eines systematischen Vorgehens, welches uns praktisch verwertbare Aussagen über den zu erwartenden Fehler in einer aus Messwerten berechneten Größe liefert. Zunächst betrachten wir nur Größen, die von genau einem Messwert abhängen. Im zweiten Semester erweitern wir das Vorgehen auf von mehreren Messwerten abhängige Größen.

Es wird sich zeigen, dass die Differentialrechnung und insbesondere die Taylor-Polynome hier gute Dienste leisten.

6.4.2 Formalisierung

Die Aufgabenstellung ist also folgende: Gegeben ist eine Funktion f, die eine messtechnisch zugängliche Zahl $x_{\text{exakt}} \in \mathbb{R}$ in den gesuchten Wert $f(x_{\text{exakt}})$ überführt. Die praktische Messung liefert statt x_{exakt} allerdings nur x_{mess}. Entsprechend steht in der Praxis nur $f(x_{\text{mess}})$ zur Verfügung statt $f(x_{\text{exakt}})$.

Der exakte Wert x_{exakt} ist üblicherweise nahe bei x_{mess} (sonst ist der Messprozess unbrauchbar), muss aber nicht mit x_{mess} übereinstimmen. Wir nehmen an, dass wir eine **Fehlerschranke** $\Delta_{\max} x \in \mathbb{R}$ kennen, d.h. wir wissen, dass

$$|x_{\text{exakt}} - x_{\text{mess}}| \leq \Delta_{\max} x \tag{6.54}$$

gilt. Daraus soll nun eine Aussage über die Größe des Fehlers

$$|f(x_{\text{exakt}}) - f(x_{\text{mess}})| \tag{6.55}$$

abgeleitet werden. Da wir $f(x_{\text{exakt}})$ nicht kennen, müssen wir vom schlechtesten Fall ausgehen, d.h. wir suchen nach der maximalen Abweichung $\Delta_{\max} f$ bzgl. aller im Rahmen des maximalen Messfehlers $\Delta_{\max} x$ denkbaren Werte $f(x_{\text{exakt}})$. In jedem Fall wird also

$$|f(x_{\text{exakt}}) - f(x_{\text{mess}})| \leq \Delta_{\max} f \tag{6.56}$$

gelten.

Merke

Der Wert $\Delta_{\max} f$ wird als **absoluter Fehler** bezeichnet. Setzt man diesen ins Verhältnis zum berechneten Wert, so erhält man den **relativen Fehler** $\frac{\Delta_{\max} f}{|f(x_{\text{mess}})|}$.

Analog spricht man bei $\Delta_{\max} x$ vom absoluten Messfehler und bei $\frac{\Delta_{\max} x}{|x_{\text{mess}}|}$ vom relativen Messfehler.

Absolute Fehler schreibt man im Ingenieurwesen gern in der Form

$$f(x_{\text{exakt}}) = f(x_{\text{mess}}) \pm \Delta_{\max} f \quad \text{bzw.} \quad x_{\text{exakt}} = x_{\text{mess}} \pm \Delta_{\max} x, \tag{6.57}$$

meint damit aber stets, dass der exakte Wert höchstens (nicht genau!) um $\Delta_{\max} f$ bzw. $\Delta_{\max} x$ vom gemessenen Wert abweicht.

Relative Fehler gibt man gern in Prozent an, also beispielsweise 0.1 % statt 0.001.

6.4.3 Vorgehen

Theoretisch könnte man den zu erwartenden Fehler als

$$\Delta_{\max} f = \max_{x \in X} |f(x) - f(x_{\text{mess}})| \tag{6.58}$$

berechnen, wobei

$$X := [x_{\text{mess}} - \Delta_{\max} x,\, x_{\text{exakt}} + \Delta_{\max} x] \tag{6.59}$$

das Intervall der infrage kommenden exakten Werte ist. Dieses Maximum ist jedoch (auch mit dem Computer!) nur sehr mühsam berechenbar und damit für den praktischen Einsatz untauglich. Ersetzen wir f jedoch durch ein Taylor-Polynom, so wird die Sache handhabbar. Schon das Taylor-Polynom ersten Grades (Tangente) reicht in den meisten Fällen völlig aus. Als Entwicklungsstelle wählen wir natürlich x_{mess}, da dieser Wert einerseits ohnehin vorliegt und andererseits alle Betrachtungen in der nahen Umgebung dieser Stelle stattfinden, also dort, wo die Abweichung zwischen Funktion und dem so konstruierten Taylor-Polynom gering ist. Mit dem Restglied

$$R(\xi, x, x_{\text{mess}}) := \frac{1}{2} f''(\xi)\,(x - x_{\text{mess}})^2 \tag{6.60}$$

erhalten wir

$$\begin{aligned}\Delta_{\max} f &= \max_{x\in X} \left| f(x_{\text{mess}}) + f'(x_{\text{mess}})\,(x - x_{\text{mess}}) + R(\xi, x, x_{\text{mess}}) - f(x_{\text{mess}}) \right| \\ &= \max_{x\in X} \left| f'(x_{\text{mess}})\,(x - x_{\text{mess}}) + R(\xi, x, x_{\text{mess}}) \right|.\end{aligned} \tag{6.61}$$

Da ξ stets zwischen x und x_{mess} liegt, können wir diesen Wert nach oben abschätzen:

$$\begin{aligned}\Delta_{\max} f &\leq \max_{x\in X} \left| f'(x_{\text{mess}})\,(x - x_{\text{mess}}) \right| + \max_{x\in X} \max_{\xi\in X} \left| R(\xi, x, x_{\text{mess}}) \right| \\ &= \left| f'(x_{\text{mess}}) \right| \max_{x\in X} |x - x_{\text{mess}}| + \frac{1}{2} \left(\max_{\xi\in X} \left| f''(\xi) \right| \right) \max_{x\in X} |x - x_{\text{mess}}|^2 \\ &= \left| f'(x_{\text{mess}}) \right| \Delta_{\max} x + \frac{1}{2} \left(\max_{\xi\in X} \left| f''(\xi) \right| \right) \left(\Delta_{\max} x \right)^2.\end{aligned} \tag{6.62}$$

Üblicherweise wird $\Delta_{\max} x$ klein sein, insbesondere viel kleiner als 1 (Schreibweise: $\Delta_{\max} x \ll 1$). Folglich wird $\left(\Delta_{\max} x\right)^2$ deutlich kleiner sein als $\Delta_{\max} x$. Für $\Delta_{\max} x = 0.01$ erhalten wir beispielsweise $\left(\Delta_{\max} x\right)^2 = 0.0001$. Zum Zwecke der Beurteilung des zu erwartenden Fehlers genügt also die Betrachtung des ersten Summanden. Nur in wenigen Ausnahmefällen, z.B. wenn der Messfehler sehr groß ist oder wenn die Funktion f stark gekrümmt ist (große zweite Ableitung!), ist der quadratische Fehleranteil relevant. In solch einem Fall können auch Taylor-Polynome zweiten Grades verwendet werden, damit man sich die aufwendige (und ggf. gar nicht machbare) Maximierung der zweiten Ableitung spart (man vernachlässigt dann den Term dritter Ordnung).

Merke

Für den Fehler $\Delta_{\max} f$ in den aus Messwerten berechneten Funktionswerten kann man für Praxiszwecke von der Abschätzung

$$\Delta_{\max} f \lesssim \left| f'(x_{\text{mess}}) \right| \Delta_{\max} x \tag{6.63}$$

ausgehen. Dabei steht $\lesssim$ für "im Wesentlichen kleiner als" und deutet an, dass es sich nicht um eine exakte Ungleichung handelt.

Die Ableitung der Funktion f am gemessenen Wert liefert also den Verstärkungsfaktor bei der Fehlerfortpflanzung.

Beachte jedoch: Bei streng mathematischer Betrachtung gilt diese Abschätzung nicht!

Nebenbemerkung

Man könnte nun einwenden, dass durch die Wahl der Maßeinheiten das Erfüllen der Bedingung $\Delta_{\max} x \ll 1$ beeinflusst werden kann. Misst man beispielsweise eine Strecke x in Meter und der maximale Messfehler beträgt $\Delta_{\max} = 0.01$ Meter, so ist die Bedingung offensichtlich erfüllt. Misst man jedoch in Millimeter, so führt der

gleiche Messfehler zu $\Delta_{\max} = 10$ Millimeter, was deutlich über 1 liegt. In ersterem Fall wäre das Vernachlässigen des quadratischen Anteils in der Fehlerabschätzung okay; in zweiterem eher nicht, da dann $\left(\Delta_{\max} x\right)^2 = 100$ viel größer ist als $\Delta_{\max} x$, also durchaus relevant ist. Allerdings ändert ein Wechsel der Einheiten auch die Funktion f. Werden die Argumente von f in Metern angegeben, so entspricht die Arbeit mit Millimetern der Funktion $x \mapsto f(0.001\,x)$, sodass die zweite Ableitung zu $0.001^2\, f''(0.001\,x)$ wird (Kettenregel!). Somit ist das Maximum über die zweite Ableitung nun sehr klein, sodass der üblicherweise vernachlässigte zweite Summand in der Fehlerabschätzung dennoch klein und damit vernachlässigbar ist.

Die Forderung $\Delta_{\max} x \ll 1$ ist also nicht zwingend für das Vernachlässigen des zweiten Summands, sondern stellt lediglich eine einfache Regel für die praktische Arbeit dar. Ist diese nicht erfüllt, kann das Vernachlässigen des zweiten Summands dennoch zulässig sein. Dieser Fall sollte dann entsprechend geprüft werden.

6.4.4 Beispiele

Beispiel

Der Flächeninhalt eines Kreises (in Quadratmeter) soll aus dem gemessenen Radius berechnet werden. Die Messung ergab einen Radius von 3 ± 0.01 Meter. Der zu erwartende absolute Fehler beim Flächeninhalt beträgt somit 0.188 Quadratmeter. Dies entspricht einem relativen Fehler von 0.67 Prozent (Video 640, Notizen dazu)?

Hinweis: Der exakt berechnete maximale Fehler liegt bei rund 0.189 Quadratmeter, also ebenfalls rund 0.67 Prozent. Die Größe des vernachlässigten zweiten Summanden in der Fehlerabschätzung beträgt $\frac{\pi}{10000} \approx 0.000314$.

Beispiel

Verwendet man im vorhergehenden Beispiel alle Maße in Millimeter, so erhält man aus der Näherungsformel den selben relativen Fehler wie zuvor, obwohl die Bedingung $\Delta_{\max} x \ll 1$ offensichtlich nicht erfüllt ist. Dies liegt in der Tatsache begründet, dass der zweite, vernachlässigte Summand in der Fehlerabschätzung mit einem Wert von $200\,\pi \approx 628.3$ im Vergleich zum ersten Summand mit einem Wert von $60000\,\pi \approx 188495.6$ trotzdem klein ist und dessen Vernachlässigung somit keinen relevanten Einfluss auf die berechnete Fehlerschranke hat (Video 645, Notizen dazu).

Der im zweiten Beispiel beobachtete Effekt, dass das Vernachlässigen des zweiten Sum-

manden in der Fehlerabschätzung auch bei $\Delta_{\max} x > 1$ hinnehmbar ist, ist nicht (!) allgemein gültig, sondern der konkreten Struktur der betrachteten Funktion geschuldet: Für wachsendes Argument bleibt die Krümmung (zweite Ableitung) konstant, während der Anstieg (erste Ableitung) wächst. Dadurch bleibt der zweite Summand in der Fehlerabschätzung vergleichsweise (!) klein, obwohl der große maximale Messfehler dort quadratisch eingeht.

7 Integralrechnung

Grundlagen der Berechnung von Flächeninhalten krummlinig begrenzter Flächen.

7.1 Flächeninhalt

Wir werden uns hier intensiv mit der Frage beschäftigen, was wir eigentlich unter dem Inhalt einer krummlinig begrenzten (also nahzu beliebig geformten) Fläche verstehen. Eine Fläche ist dabei eine Teilmenge des $\mathbb{R}^2$. Nicht betrachtet werden gekrümmte Flächen, also z.B. Teilmengen von Kugeloberflächen.

Später wird sich zeigen, dass die Berechnung von Flächeninhalten eng mit Funktionen (zur Beschreibung des Randes) und Ableitungen von Funktionen verknüpft ist. Dieser zunächst nicht sichtbare Zusammenhang wird uns ein mächtiges Werkzeug zur praktischen Berechnung von Flächeninhalten nahezu beliebiger Flächen an die Hand geben.

7.1.1 Problemstellung

Schon in der Schulmathematik wurden Flächeninhalte berechnet. Das lief nach dem Schema "suche den Typ der Fläche in der Formelsammlung und nutze die dort angebene Formel für den Flächeninhalt". Aber was verstehen ganz allgemein unter "Flächeninhalt"? Wie erklären wir jemandem, was zu tun ist, wenn beispielsweise der Flächeninhalt einer eiförmigen Fläche zu berechnen ist?

Aus praktischer Sicht können wir den Flächeninhalt einer beliebigen Fläche beschreiben als "Menge an Farbe, die zum Anstreichen der Fläche benötigt wird". Oder etwas präziser (da die Schichtdicke der Farbe nicht klar ist): Man versehe die Fläche mit einem 1 Zentimeter hohen Rand und fülle die Fläche bis zur Oberkante des Randes mit Wasser. Die benötigte Wassermenge in Kubikzentimenter ist dann gerade (per Definition!) der Flächeninhalt in Quadratzentimeter. Praktisch ist das Problem also durchaus lösbar.

Offen ist, wie man die benötigte Wassermenge im Vorfeld berechnen kann. Für sehr große oder sehr kleine Flächen ist die beschriebene Form der experimentellen Inhaltsbestimmung kaum umsetzbar. Auch ist der Aufwand recht groß (gekrümmten Rand bauen; alles muss dicht sein usw.).

Das Frage der sinnvollen, allgemeinen Definition des Inhalts einer Fläche wurde erst um 1900 zufriedenstellend beantwortet. Für mathematische Verhältnisse ist dieses Thema also noch regelrecht "neu".

7.1.2 Idee

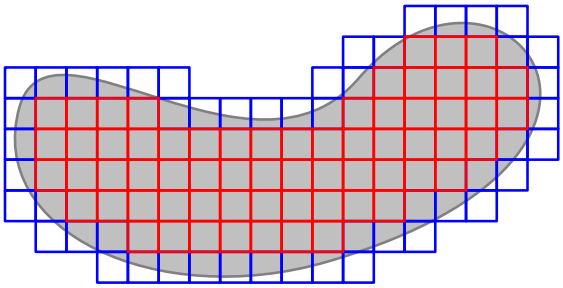
Die Grundidee zur mathematisch exakten Definition des Begriffs "Flächeninhalt" ist folgende.

1. Zerlege die Fläche in Teilflächen mit berechenbarem Flächeninhalt.
2. Addiere die Inhalte aller Teilflächen.

Für manche Flächen ist dies leicht machbar. Für andere jedoch werden sehr viele Teilflächen benötigt oder eine Zerlegung ist nicht möglich, da der krummlinige Rand beim

Zerlegen nicht verschwindet und stets nicht berechenbare Teilflächen übrig bleiben.

Deshalb modifizieren wir die Idee etwas: Zum einen legen wir uns auf eine möglichst einfache Form von Teilflächen fest, um die Berechnung der einzelnen Flächeninhalte zu ermöglichen. Die einfachste Fläche mit Blick auf Inhaltsberechnung ist das Quadrat. Zum anderen fordern wir nicht mehr, dass die Fläche vollständig in Teilflächen zerlegt werden muss, sondern wir überdecken die Fläche mit Quadraten gleicher Größe und zählen wie viele Quadrate vollständig innerhalb der Fläche liegen und wie viele Quadrate wenigstens einen Teil der Fläche enthalten. Die Summen der jeweiligen Quadratinhalte liefern den Flächeninhalt einer etwas kleineren und einer etwas größeren Fläche. Je kleiner die Quadrate sind, desto näher werden sich diese beiden Werte kommen. Betrachten wir den Grenzübergang "Kantenlänge gegen Null", so sollten beide Werte übereinstimmen. Diese Zahl betrachten wir nun als Flächeninhalt der unregelmäßigen Fläche.



Die roten Quadrate liegen vollständig im Inneren der Fläche. Die blauen Quadrate überlappen mit der Fläche. Der Flächeninhalt der grauen Fläche muss zwischen der Summe der Flächen der roten Quadrate und der Summe der Flächen der roten und blauen Quadrate liegen. Je kleiner die Quadrate, desto kleiner der Unterschied zwischen beiden Schranken.

7.1.3 Umsetzung

Setzen die Idee zur Definition des Flächeninhalts beliebiger Flächen nun mathematisch exakt um. Zu $F \subseteq \mathbb{R}^2$ ist also eine Zahl $A(F) \geq 0$ gesucht, die unserer intuitiven Vorstellung von "Flächeninhalt" entspricht. Insbesondere soll $A(F)$ für einfache Flächen F natürlich die bisher mittels konkreter Formeln berechneten Flächeninhalte liefern.

Zur Überdeckung von F verwenden wir ein Raster aus achsenparallelen Quadraten $[x, x+l] \times [y, y+l]$ mit Kantenlänge $l > 0$, welches der Einfachheit halber so ausgerichtet ist, dass der Punkt $(0, 0)$ gerade ein Eckpunkt eines der Quadrate ist. Die x- und die y-Koordinaten aller Eckpunkte aller Quadrate sind also ganzzahlige Vielfache von l.

Mit $i(l) \in \mathbb{N}$ bezeichnen wir die Anzahl von Quadraten, die vollständig von F überdeckt werden (innere Überdeckung). Mit $a(l) \in \mathbb{N}$ bezeichnen wir die Anzahl von Quadraten, die mindestens einen Punkt von F enthalten (äußere Überdeckung). Offensichtlich gilt also stets $i(l) \leq a(l)$. Damit beide Zahlen existieren (also nicht "unendlich" sind), müssen wir unsere Betrachtungen auf **beschränkte Flächen** eingrenzen, d.h. auf solche Flächen

F, für die es eine Zahl $c \in \mathbb{R}$ gibt mit

$$\sqrt{x^2 + y^2} \leq c \quad \text{für alle } (x, y) \in F. \tag{7.1}$$

Merke

Sei $F \subseteq \mathbb{R}^2$ beschränkt und seien $i(l)$ und $a(l)$ für $l > 0$ wie oben beschrieben. Die Menge F heißt **Jordan-messbar**, wenn die beiden Grenzwerte

$$\lim_{l \to \infty} i(l)\, l^2 \quad \text{und} \quad \lim_{l \to \infty} i(l)\, l^2 \tag{7.2}$$

existieren und gleich sind. In diesem Fall werden die beiden (gleichen) Grenzwerte als **Flächeninhalt** von F bezeichnet.

Die Werte $i(l)\, l^2$ und $a(l)\, l^2$ sind gerade die Flächeninhalte der inneren und der äußeren Überdeckung zur Kantenlänge l.

7.1.4 Beispiele und Einordnung

Beispiel (Flächeninhalt eines Rechtecks)

Wenn unsere Definition von "Flächeninhalt" sinnvoll ist, sollten wir damit die üblichen Formeln für Flächeninhalte von einfachen Flächen herleiten können. Tun dies für das Rechteck $[0, c] \times [0, d]$.

Für festes $l > 0$ sei $m(l) \in \mathbb{N}$ die Anzahl der Quadratspalten der inneren Überdeckung. Entsprechend bezeichne $n(l) \in \mathbb{N}$ die Anzahl der Quadratzeilen. Es gilt also $i(l) = m(l)\, n(l)$. Die Zahlen $m(l)$ und $n(l)$ können wir mittels der Kantenlängen c und d des Rechtecks ausdrücken:

$$m(l) = \left\lfloor \frac{c}{l} \right\rfloor \quad \text{und} \quad n(l) = \left\lfloor \frac{d}{l} \right\rfloor. \tag{7.3}$$

Dabei bezeichnet $\lfloor \ldots \rfloor$ das Abrunden des eingeschlossenen Wertes. Für den Grenzwert des Inhalts der inneren Überdeckung erhalten wir

$$\begin{aligned} \lim_{l \to 0} i(l)\, l^2 &= \lim_{l \to 0} \left\lfloor \frac{c}{l} \right\rfloor \left\lfloor \frac{d}{l} \right\rfloor l^2 \\ &\geq \lim_{l \to 0} \left(\frac{c}{l} - 1 \right) \left(\frac{d}{l} - 1 \right) l^2 \\ &= \lim_{l \to 0} (c - l)\, (d - l) = c\, d. \end{aligned} \tag{7.4}$$

Analog folgt

$$\begin{aligned}\lim_{l\to 0} a(l)\, l^2 &= \lim_{l\to 0} \left\lceil \frac{c}{l} \right\rceil \left\lceil \frac{d}{l} \right\rceil l^2 \\ &\leq \lim_{l\to 0} \left(\frac{c}{l} + 1\right) \left(\frac{d}{l} + 1\right) l^2 \\ &= \lim_{l\to 0} (c + l)\,(d + l) = c\,d,\end{aligned} \tag{7.5}$$

wobei $\lceil \ldots \rceil$ für das Aufrunden des eingeschlossenen Wertes steht.

Aus

$$c\,d \leq \lim_{l\to 0} i(l)\, l^2 \leq \lim_{l\to 0} a(l)\, l^2 \leq c\,d \tag{7.6}$$

folgt nun die Gleichheit beider Grenzwerte. Der Flächeninhalt ist entsprechend unserer Definition des Begriffs also wie erwartet $c\,d$.

Bereits bei einem Rechteck ist die konkrete Berechnung des Flächeninhalts recht mühsam. Bei komplexeren Fläche ist die Berechnung direkt über die Definition kaum machbar. Deshalb werden wir uns im nächsten Unterkapitel ein alternatives Vorgehen überlegen.

Nebenbemerkung

Es gibt durchaus beschränkte Mengen, die nicht Jordan-messbar sind, denen wir also auf dem beschriebenen Weg keinen Flächeninhalt zuordnen können. Solche Mengen sind aber praktisch nicht relevant.

Beispiel (nicht Jordan-messbare Menge)

Die Menge $F := [0,1] \times ([0,1] \cap \mathbb{Q})$ ist nicht Jordan-messbar. Kein noch so kleines Quadrat wird von F vollständig überdeckt, da jedes Quadrat auch Punkte mit nichtrationalen y-Koordinaten enthält. Also $i(l) = 0$ für alle $l > 0$. Andererseits besteht jede äußere Überdeckung aus mindestens einem Quadrat. Also gilt stets $a(l) \geq 1$. Somit können die Grenzwerte der Inhalte der inneren und der äußeren Überdeckungen nicht gleich sein.

Nebenbemerkung

Es gibt verschiedene Varianten von “Messbarkeit” in der Mathematik. Neben der hier eingeführten Jordan-Messbarkeit spielt die sogenannte Lebesgue-Messbarkeit eine wesentliche Rolle. Diese ist eine Verallgemeinerung der Jordan-Messbarkeit (jede Jordan-messbare Menge ist auch Lebesgue-messbar) und deckt mehr Mengen ab. Jedoch gibt es trotzdem Mengen, die auch nicht Lebesgue-messbar sind.

Die Konstruktion des Flächeninhalts auf Basis der Jordan-Messbarkeit kann völlig analog auch in drei oder mehr Dimensionen durchgeführt werden. In $\mathbb{R}^3$ erhält man mittels Überdeckungen aus Würfeln auf diese Weise eine Definition des Begriffs “Volumen”, die auch sehr unregelmäßige Körper mit einschließt.

7.2 Normalbereiche

Für die praktische Berechnung von Flächeninhalten unregelmäßig geformter Flächen müssen wir zwei Probleme lösen:

1. Die Fläche muss mittels Formeln beschreibbar sein.
2. Die Berechnung des Flächeninhalts muss auch bei komplexeren Flächen mit vertretbarem Aufwand durchführbar sein.

In diesem Unterkapitel widmen wir uns vor allem dem ersten Problem. Die Lösung des zweiten Problems bereiten wir vor, gehen sie aber erst im nächsten Unterkapitel vollständig an.

7.2.1 Begriff

Wir beschränken alle nachfolgenden Betrachtungen auf folgenden Typ von Flächen:

Merke

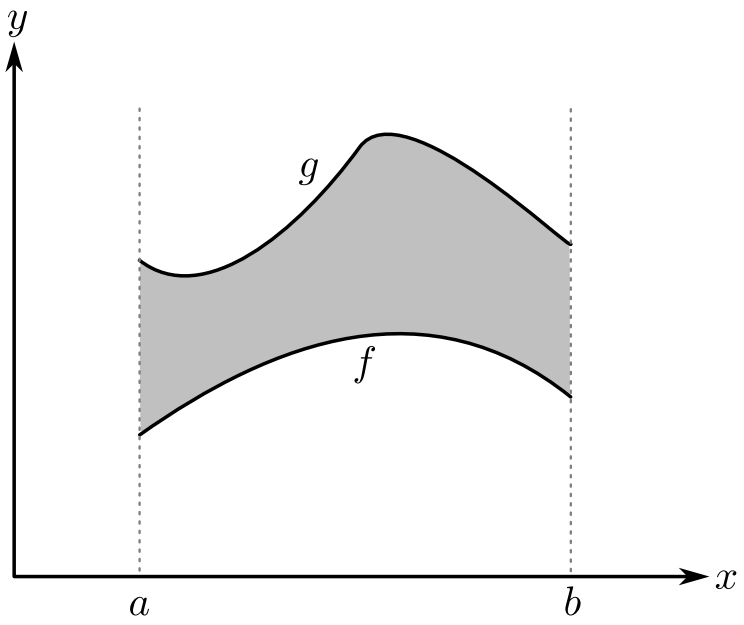
Lässt sich eine Menge $B \subseteq \mathbb{R}^2$ in der Form

$$B = \left\{(x,y) \in \mathbb{R}^2 :\ a \leq x \leq b,\ f(x) \leq y \leq g(x)\right\} \tag{7.7}$$

schreiben, wobei $f : [a,b] \to \mathbb{R}$ und $g : [a,b] \to \mathbb{R}$ zwei stetige Funktionen über einem Intervall $[a,b]$ sind, die $f(x) \leq g(x)$ für alle $x \in [a,b]$ erfüllen, so nennen wir B einen **Normalbereich bzgl. der x-Achse**.

Man kann auch **Normalbereiche bzgl. der y-Achse** einführen. Diese können jedoch durch drehen um 90° oder durch Spiegelung an der Gerade $y = x$ stets in Normalbereiche bzgl. der x-Achse überführt werden, sodass deren weitere Betrachtung hier unterbleibt.

7.2.2 Beispiele



Ein Normalbereich wird durch ein Intervall für die x-Koordinaten und durch zwei Funktionen beschrieben. Die beiden Funktionsgraphen bilden die untere und die obere Grenze des Normalbereichs.

Beispiel

Das Viereck $B \subseteq \mathbb{R}^2$, welches durch die Geraden

$$x = 1, \quad y = \frac{1}{2}x + 1, \quad x = 3, \quad y = x + 2 \tag{7.8}$$

begrenzt wird, ist ein Normalbereich:

$$B = \left\{(x, y) : \, 1 \leq x \leq 3, \, \frac{1}{2}x + 1 \leq y \leq x + 2\right\} \tag{7.9}$$

(Video 660, Notizen dazu).

Beispiel

Das krummlinig begrenzte Dreieck $B \subseteq \mathbb{R}^2$, welches durch die Kurven

$$y = x^2, \quad y = 6 - x, \quad x = 4 \tag{7.10}$$

eingeschlossen wird, ist ein Normalbereich:

$$B = \{(x, y) : \, 2 \leq x \leq 3, \, 6 - x \leq y \leq x^2\} \tag{7.11}$$

(Video 665, Notizen dazu).

Beispiel

Die Kreisscheibe $K \subseteq \mathbb{R}^2$ mit Radius 2 um den Punkt $(3,1)$ ist ein Normalbereich:

$$K = \left\{(x,y) : -2 \le x \le 2,\ 1 - \sqrt{4-(x-3)^2} \le y \le 1 + \sqrt{4-(x-3)^2}\right\} \quad (7.12)$$

(Video 670, Notizen dazu).

7.2.3 Flächeninhalt von Normalbereichen

Berechnen nun den Flächeninhalt eines Normalbereichs

$$B := \left\{(x,y) \in \mathbb{R}^2 :\ a \le x \le b,\ f(x) \le y \le g(x)\right\} \quad (7.13)$$

auf Basis unserer Definition des Flächeninhalts. Der Einfacheit halber nehmen wir zunächst an, dass f und g nur positive Werte annehmen. Dann können wir den Flächeninhalt von B als Differenz der Flächeninhalte zwischen g und der x-Achse sowie f und der x-Achse ausdrücken. Unsere Aufgabe beschränkt sich damit auf das Berechnen des Flächeninhalts zwischen der x-Achse und einem Funktionsgraph. Beschreiben das Vorgehen hier für die Funktion f.

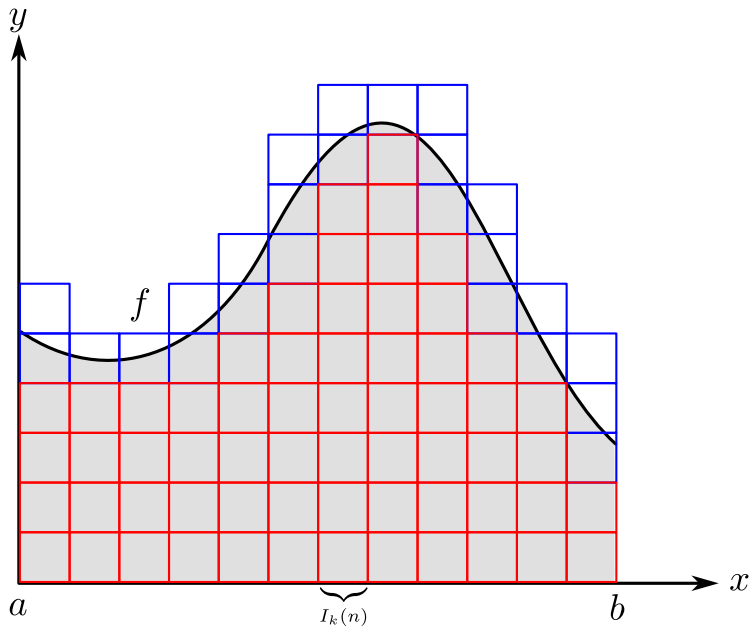
Um die Formeln einfach zu halten, gehen wir von $a = 0$ aus, was durch Verschieben des Bereichs stets problemlos erreicht werden kann, also keine inhaltliche, sondern nur eine formale Einschränkung darstellt. In der Mathematik ist an dieser Stelle die Formulierung "*Ohne Beschränkung der Allgemeinheit* nehmen wir $a = 0$ an" üblich.

Wir überdecken den Bereich B wie im vorhergehenden Unterkapitel mit einem Raster aus Quadraten der Kantenlänge $l > 0$ und ermitteln die Anzahlen $i(l)$ und $a(l)$ der Quadrate in der inneren und in der äußeren Überdeckung.

Man kann sich relativ leicht überlegen (tun wir hier nicht), dass $\lim_{l\to 0} i(l)\, l^2$ genau dann existiert, wenn $\lim_{n\to\infty} i(l_n)\, l_n^2$ für irgendeine Folge $(l_n)_{n\in\mathbb{N}}$ mit $\lim_{n\to\infty} l_n = 0$ gilt. Außerdem sind im Falle der Existenz die beiden Grenzwerte gleich. Analoges gilt für $a(l)\, l^2$. Diese Feststellung vereinfacht das weitere Vorgehen, da wir statt dem Grenzwert $\lim_{l\to 0} i(l)\, l^2$ einer Funktion bzgl. l nur den Grenzwert $\lim_{n\to\infty} i(l_n)\, l_n^2$ einer Zahlenfolge betrachten müssen.

Wählen wir nun $l_n := \frac{b}{n}$, so gilt offensichtlich $\lim_{n\to\infty} l_n = 0$. Innere und äußere Überdeckung für ein n haben stets gleich viele Spalten, da $\lfloor \frac{b}{l_n} \rfloor = \lfloor n \rfloor = n = \lceil n \rceil = \lceil \frac{b}{l_n} \rceil$ gilt. Anders ausgedrückt: b liegt immer genau auf der Grenze zwischen zwei Quadratspalten. Entsprechend haben Unter- und Obersumme die gleiche Anzahl Summanden.

Beim Summieren der Quadratflächen gehen wir spaltenweise vor. Die Anzahl der Spalten ist n. Die Spalten seien von links nach rechts mit $1, 2, \ldots, n$ nummeriert. Die Anzahl der



Die Oberkanten der äußeren Überdeckung (blau) liegen vollständig oberhalb des Funktionsgraphs. Die Oberkanten der inneren Überdeckung (rot) liegen vollständig unterhalb des Funktionsgraphs.

Quadrate in Spalte k hängt offensichtlich von den Funktionswerten von f im Bereich dieser Spalte ab. Sei also $I_k(n) := [(k-1)\,l,\ k\,l]$ das Intervall von x-Koordinaten zur k-ten Spalte. Dann ist die Anzahl Quadrate der inneren Überdeckung in Spalte k gerade

$$\left\lfloor \frac{\min\limits_{x \in I_k(n)} f(x)}{l_n} \right\rfloor \tag{7.14}$$

und wir erhalten für den Grenzwert des Flächeninhalts der inneren Überdeckung

$$\lim_{n\to\infty} i(l_n)\, l_n^2 = \lim_{n\to\infty} \left(l_n \sum_{k=1}^{n} \min_{x \in I_k(n)} f(x) \right), \tag{7.15}$$

sofern der rechte Grenzwert existiert (Video 675, Notizen dazu). Der Ausdruck im rechten Grenzwert wird auch als **Riemannsche Untersumme** zur Unterteilungsfeinheit n bezeichnet.

Völlig analog erhalten wir für die äußeren Überdeckungen

$$\lim_{n\to\infty} a(l_n)\, l_n^2 = \lim_{n\to\infty} \left(l_n \sum_{k=1}^{n} \max_{x \in I_k(n)} f(x) \right). \tag{7.16}$$

Der Ausdruck im zweiten Grenzwert wird auch als **Riemannsche Obersumme** zur Unterteilungsfeinheit n bezeichnet.

Machen wir die Verschiebung zu $a = 0$ rückgängig, so können wir folgende Definition formulieren:

Merke

Sei $f : [a, b] \to \mathbb{R}$ stetig und bezeichne $I_k(n) := \left[a + (k-1)\,\frac{b-a}{n},\ a + k\,\frac{b-a}{n}\right]$ für $k = 1, \ldots, n$ die Teilintervalle von $[a, b]$ bei Unterteilung in n gleich große Teile. Existieren die Grenzwerte

$$\lim_{n\to\infty} \left(\frac{b-a}{n} \sum_{k=1}^{n} \min_{x \in I_k(n)} f(x) \right) \tag{7.17}$$

und

$$\lim_{n\to\infty} \left(\frac{b-a}{n} \sum_{k=1}^{n} \max_{x \in I_k(n)} f(x) \right) \tag{7.18}$$

der Riemannschen Unter- und Obersummen und stimmen die beiden Grenzwerte überein, so heißt f **Riemann-integrierbar**. Der Wert der beiden (gleichen) Grenzwerte heißt dann **Riemann-Integral** von f über $[a, b]$ und entspricht gerade dem vorzeichenbehafteten Flächeninhalt zwischen dem Graph von f und der x-Achse.

Als Symbol für das Riemann-Integral wird aus historischen Gründen und auch aus Gründen, die später erhellt werden, das Symbol

$$\int_a^b f(x)\,\mathrm{d}x \tag{7.19}$$

verwendet.

Mit "vorzeichenbehaftet" meinen wir hier, dass das Riemann-Integral negativ ist, wenn der Funktionsgraph vollständig unterhalb der x-Achse liegt. Schneidet die Funktion die x-Achse, so gehen Flächenanteile unterhalb der x-Achse negativ in das Riemann-Integral ein. Bei gleichen Anteilen ober- und unterhalb der x-Achse wird das Riemann-Integral also Null sein.

Mit Blick auf den Flächeninhalt von Normalbereichen (mit Untergrenze f und Obergrenze g) ist diese Vorzeicheneigenschaft des Riemann-Integrals sehr nützlich. Sie führt dazu, dass die Differenz

$$\int_a^b g(x)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x, \tag{7.20}$$

also der Flächeninhalt des Normalbereichs, stets positiv ist (wenn g überall oberhalb

von f verläuft). Das gilt insbesondere auch, wenn der Normalbereich unterhalb der x-Achse liegt (also beide Integrale negativ sind), und auch, wenn die x-Achse durch den Normalbereich verläuft.

Die praktische Berechnung von Riemann-Integralen klären wir im nächsten Unterkapitel. Obwohl der Weg über Unter- und Obersummen einfacher ist als über innere und äußere Überdeckungen aus Quadraten, ist der Weg immer noch zu mühsam.

7.2.4 Hinweise zum Riemann-Integral

7.2.4.1 Linearität Man überlegt sich leicht anhand der Eigenschaften von Grenzwerten, dass

$$\int_a^b f(x)\,\mathrm{d}x + \int_a^b g(x)\,\mathrm{d}x = \int_a^b f(x) + g(x)\,\mathrm{d}x \tag{7.21}$$

und

$$\lambda \int_a^b f(x)\,\mathrm{d}x = \int_a^b \lambda\, f(x)\,\mathrm{d}x \quad \text{für } \lambda \in \mathbb{R} \tag{7.22}$$

gelten. Außerdem gilt

$$\int_a^b f(x)\,\mathrm{d}x + \int_b^c f(x)\,\mathrm{d}x = \int_a^c f(x)\,\mathrm{d}x \tag{7.23}$$

(ebenfalls leicht einzusehen). Diese beiden Eigenschaften werden als Linearität des Riemann-Integrals bezeichnet. Man kann das Integral in diesem Sinne als lineare Abbildung von der Menge der stetigen Funktionen in die reellen Zahlen ansehen.

7.2.4.2 Stückweise stetiger Integrand Man kann Riemann-Integrale auch für Funktionen einführen, die nicht stetig sind, sondern nur **stückweise stetig**. Stückweise stetige Funktionen sind Funktionen, deren Definitionsbereich in endlich viele Teilintervalle geteilt werden kann, sodass die Funktion auf jedem Teilintervall stetig ist und die einseitigen Grenzwerte an den Intervallenden existieren (also keine Polstellen).

Das Integral einer stückweise stetigen Funktion ist dann einfach die Summe der Integrale auf allen Teilintervallen. Alternativ kann man die obige Herleitung mittels Überdeckungen mit Quadraten auch direkt für stückweise stetige Funktionen durchführen und kommt zum selben Ergebnis.

7.2.4.3 Verwandte Begriffe Das Riemann-Integral wird oft einfach als Integral bezeichnet. Zur Unterscheidung von einem im nächsten Unterkapitel einzuführenden, eng verbundenen Begriff, spricht man auch vom **bestimmten Integral**.

Nebenbemerkung

Neben Riemann-Integralen gibt es noch eine Reihe anderer Integralbegriffe, die für uns aber keine Bedeutung haben, da sie nur in für uns unwesentlichen Nuancen vom hier eingeführten Riemann-Integral abweichen. Am verbreitetsten ist das Lebesgue-Integral, welches das Lebesgue-Maß zur Berechnung der Fläche unter dem Funktionsgraph nutzt.

7.2.5 Uneigentliche Riemann-Integrale

Gelegentlich trifft man auf Flächen, die unbeschränkt sind, aber in gewissem Sinne einen endlichen Flächeninhalt besitzen. Solchen Flächen werden wir beispielsweise in der Wahrscheinlichkeitsrechnung begegnen. Praktisch sollten solche Flächen eigentlich nicht relevant sein, da die vom Menschen wahrnehmbare Welt eine beschränkte Menge ist; somit sind auch alle Flächen und Körper darin beschränkt. Bei der vereinfachenden Modellierung gewisser Sachverhalte tauchen dann aber doch solche unbeschränkten Flächen auf. Eine ähnliche Situation hatten wir schon bei Polstellen beschrieben (vgl. Beispiel (Massepunkt)).

Wir beschränken uns hier auf den einfachsten Fall, nämlich die Situation, dass der Definitionsbereich einer Funktion f, die zusammen mit der x-Achse eine Fläche begrenzt, unbeschränkt ist. Ein andere Variante unbeschränkter Flächen mit endlichem Inhalt tritt bei Polstellen am Rand oder im Inneren des Definitionsbereichs einer Funktion auf. Diesen Fall kann man völlig analog handhaben, sodass wir nicht näher darauf eingehen.

7.2.5.1 Einseitig unbeschränkter Definitionsbereich Wie so oft in der Mathematik, führt man eine komplexere auf eine einfachere Situation zurück. Den Flächeninhalt komplexerer Flächen hatten wir beispielsweise über den Grenzwert der Flächeninhalte von aus Quadraten zusammengesetzten Flächen erklärt. Den Flächeninhalt unbeschränkter Flächen erklären wir nun wiederum als Grenzwert des Inhalts gewisser beschränkter Flächen:

Merke

Sei $f : (-\infty, b) \to \mathbb{R}$ stetig. Dann setzen wir

$$\int_{-\infty}^{b} f(x)\,\mathrm{d}x := \lim_{a \to -\infty} \int_{a}^{b} f(x)\,\mathrm{d}x, \tag{7.24}$$

sofern dieser Grenzwert existiert, und nennen diesen Wert **uneigentliches Riemann-Integral** zu f. Für $f : [a, \infty) \to \mathbb{R}$ setzen wir analog

$$\int_a^\infty f(x)\,\mathrm{d}x := \lim_{b\to\infty} \int_a^b f(x)\,\mathrm{d}x. \tag{7.25}$$

Es werden also die (vorzeichenbehafteten) Inhalte immer größer werdender Teilflächen berechnet. Konvergiert diese Folge von Flächeninhalten, so kann man auch der unbeschränkten Gesamtfläche sinnvoll einen Flächeninhalt zuordnen. Bei bestimmter Divergenz kann man die Werte ∞ bzw. $-\infty$ als Flächeninhalt ansehen. Bei unbestimmter Divergenz ist überhaupt kein sinnvoller Inhaltsbegriff mehr denkbar.

Beispiele für uneigentliche Integrale schauen wir uns erst im nächsten Unterkapitel an, wenn wir ein einfaches Verfahren zur Berechnung von Integralen kennengelernt haben.

7.2.5.2 Beidseitig unbeschränkter Definitionsbereich Ist der Definitionsbereich beidseitig unbeschränkt, also ganz $\mathbb{R}$, so gibt es zwei verschiedene Möglichkeiten zur Definition eines Flächeninhalts. Entweder man teilt die Fläche in zwei Teilflächen und betrachtet jede Teilfläche für sich und addiert die beiden Inhalte dann, oder man führt den Grenzübergang symmetrisch an beiden Enden des Definitionsbereichs durch. Da beide Varianten gebräuchlich sind, führen wir hier beide ein.

Merke

Sei $f : \mathbb{R} \to \mathbb{R}$ stetig. Existieren die beiden uneigentlichen Riemann-Integrale auf den Intervallen $(-\infty, 0]$ und $[0, \infty)$, so heißt

$$\int_{-\infty}^\infty f(x)\,\mathrm{d}x := \int_{-\infty}^0 f(x)\,\mathrm{d}x + \int_0^\infty f(x)\,\mathrm{d}x \tag{7.26}$$

uneigentliches Riemann-Integral zu f.

Außerdem setzen wir

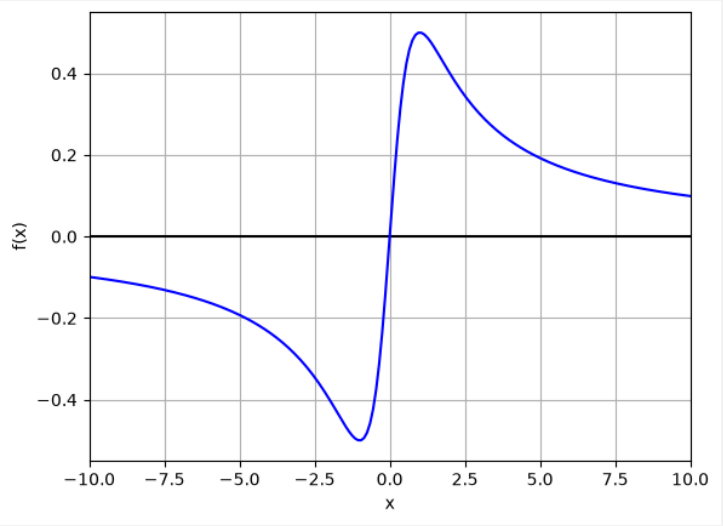
$$\mathrm{CH} \int_{-\infty}^\infty f(x)\,\mathrm{d}x := \lim_{c\to\infty} \int_{-c}^c f(x)\,\mathrm{d}x, \tag{7.27}$$

sofern dieser Grenzwert existiert, und nennen diese Zahl den **Cauchyschen Hauptwert** zu f.

Existiert das uneigentliche Riemann-Integral, so existiert auch der Cauchysche Hauptwert und beide stimmen überein. Umgekehrt muss dies nicht gelten.

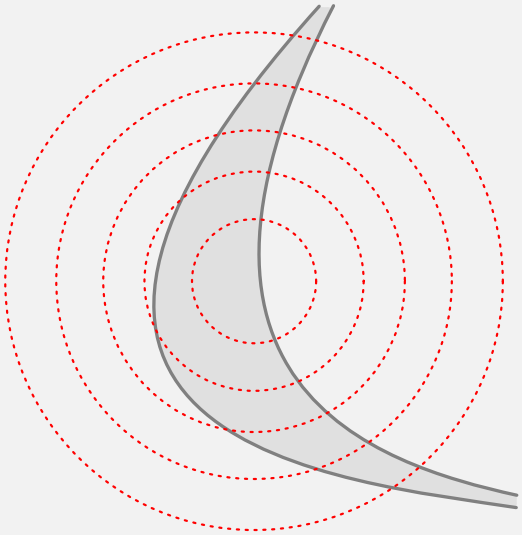
Beispiel

Für die Funktion $f : \mathbb{R} \to \mathbb{R}$, $f(x) : \frac{x}{1+x^2}$ ist aus Symmetriegründen der Cauchysche Hauptwert offensichtlich Null. Jedoch ist die Funktion nicht uneigentlich Riemann-Integrierbar, da man zeigen kann, dass die beiden einseitigen uneigentlichen Integrale (bestimmt) divergieren.



Nebenbemerkung

Für die Definition des Inhalts allgemeiner unbeschränkter Flächen kann man den Gedanke des Cauchyschen Hauptwerts nutzen: Man bilder den Durchschnitt der Fläche mit größer werdenen Kreisscheiben und betrachtet den Grenzwert der Flächeninhalte der (beschränkten) Durchschnittsflächen.



In Formeln: Bezeichnet $K(r) \subseteq \mathbb{R}^2$ die Kreisscheibe mit Radius r um den Ursprung, so kann der Inhalt $A(F)$ einer unbeschränkten Fläche $F \subseteq \mathbb{R}$ als

$$A(F) := \lim_{r \to \infty} A\big(F \cap K(r)\big) \tag{7.28}$$

erklärt werden, sofern alle Durchschnitte $F \cap K(r)$ Jordan-messbar sind und der Grenzwert existiert.

Berechnet man diesen Grenzwert für eine beschränkte Fläche, so erhält man den korrekten Flächeninhalt, da für hinreichend großes r die Fläche vollständig in allen Kreisscheiben enthalten ist.

7.3 Praktische Integration

Wie kaum anders zu erwarten, werden Flächeninhalte und die dafür benötigten bestimmten Integrale heute mit dem Computer berechnet. Bevor wir darauf eingehen, schauen wir uns zunächst den klassischen manuellen Weg an, da dieser einen auch an andereren Stellen gelegentlich sehr nützlichen Zusammenhang zutage fördert (z.B. beim in dieser Veranstaltung nicht betrachteten Lösen von Differentialgleichungen).

7.3.1 Manuelles Integrieren

Die Definition des Riemann-Integrals taugt nicht für das praktische Ausrechnen von Integralen. Stattdessen werden wir ein zunächst gar nicht offensichtliches Resultat der Mathematik nutzen, welches Flächeninhalte mit Ableitungen von Funktionen verbindet.

7.3.1.1 Hauptsatz der Differential- und Integralrechnung Die folgende Erkenntnis, auf deren Herleitung wir hier verzichten, vereinfacht das Berechnen von Integralen erheblich:

Merke

Ist der Integrand in einem Riemann-Integral die Ableitung einer stetig differenzierbaren Funktion $f : [a, b] \to \mathbb{R}$, so gilt

$$\int_a^b f'(x)\,\mathrm{d}x = f(b) - f(a). \tag{7.29}$$

Bei differenzierbaren Funktionen hatten wir als Definitionsbereich stets offene Intervalle (a, b) betrachtet, damit wir uns keine Gedanken über die Definition von Differenzierbarkeit an Randpunkten machen müssen. Bei Integralen verwenden wir hingegen abgeschlossene Intervalle, damit die entstehenden Flächen abgeschlossen sind, also einen Rand haben. Mit Differenzierbarkeit auf $[a, b]$ in obiger Aussage meinen wir, dass es ein offenes Intervall (c, d) gibt, auf dem die Funktion f differenzierbar ist und dass das Intervall $[a, b]$ enthält, also $[a, b] \subseteq (c, d)$.

7.3.1.2 Stammfunktionen Die Frage, ob eine zu integrierende Funktion überhaupt Ableitung einer anderen Funktion ist, ist zunächst offen. Nur für solche Funktionen hilft uns der Hauptsatz der Differential- und Integralrechnung bei der Berechnung von Riemann-Integralen. Zur weiteren Klärung führen wir den Begriff der Stammfunktion ein.

Merke

Jede differenzierbare Funktion, deren Ableitung gerade einer gegebenen Funktion f gleicht, heißt **Stammfunktion** von f.

Sind F_1 und F_2 zwei Stammfunktionen von f, so gilt $F_1(x) = F_2(x) + c$ für ein $c \in \mathbb{R}$ und alle x aus dem Definitionsbereich von f. Ist umgekehrt F eine Stammfunktion von f, so ist auch die durch $\tilde{F}(x) := F(x) + c$ gegebene Funktion $\tilde{F}$ für beliebiges $c \in \mathbb{R}$ eine Stammfunktion von f.

Wenn eine Funktion f eine Stammfunktion besitzt, dann besitzt sie also automatisch unendlich viele Stammfunktionen, die sich nur um eine additive Konstante unterscheiden.

Die meisten praktisch relevanten Funktionen besitzen eine Stammfunktion:

Merke

Sei $f : (a, b) \to \mathbb{R}$ stetig. Dann ist die durch

$$F(x) := \int_a^x f(\tilde{x}) \, \mathrm{d}\tilde{x} \tag{7.30}$$

definierte Funktion $F : (a, b) \to \mathbb{R}$ eine Stammfunktion von f.

Aus praktischer Sicht viel relevanter ist die Frage, ob man Stammfunktionen auch explizit als Formel angeben kann. Hier sind Sachlage und Vorgehen ähnlich wie bei der Frage der Differenzierbarkeit:

- Für viele elementare Funktionen kann man eine Stammfunktion als Formel angeben.
- Für aus elementaren Funktionen zusammengesetzte Funktionen kann man oft mittels diverser "Integrationsregeln" eine Stammfunktion aus den Stammfunktionen der entsprechenden elementaren Funktionen zusammensetzen.
- In der Masse aller stetigen Funktionen sind die Funktionen, für die man eine Stammfunktion explizit als Formel angeben kann, in der Minderheit.

Wir werden im Folgenden einen kurzen Blick auf Stammfunktionen einiger elementarer Funktionen werfen, die Arbeit mit Integrationsregeln aber an das Ende des zweiten Semesters verschieben, da diese für Ingenieure kaum noch praktische Relevanz hat.

Bevor wir zu den elementaren Funktionen kommen, noch ein Hinweis auf eine verbreitete Schreibweise:

Merke

Für eine Stammfunktion einer Funktion f verwendet man häufig das Symbol

$$\int f(x)\,\mathrm{d}x. \tag{7.31}$$

Man bezeichnet Stammfunktionen entsprechend auch als **unbestimmte Integrale**. Mit dieser Schreibweise kann der Hauptsatz der Differential- und Integralrechnung als

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a) \quad \text{mit} \quad F(x) := \int f(x)\,\mathrm{d}x \tag{7.32}$$

formuliert werden.

Beachte bei dieser Schreibweise, dass $\int f(x)\,\mathrm{d}x$ eine Funktion ist, $\int_a^b f(x)\,\mathrm{d}x$ aber eine Zahl. Welche Stammfunktion man wählt, spielt im Hauptsatz der Differential- und Integralrechnung keine Rolle, da durch die Differenz der beiden Stammfunktionswerte additive Konstanten wieder entfallen. Gelegentlich meint man mit dem Symbol $\int f(x)\,\mathrm{d}x$ auch die Menge aller Stammfunktionen von f, statt einer konkreten Stammfunktion.

7.3.1.3 Stammfunktionen elementarer Funktionen In Formelsammlungen gibt es lange Listen mit Stammfunktionen für häufig benötigte Funktionen. Hier nur eine kleine Auswahl:

$$\int x^c\,\mathrm{d}x = \frac{1}{c+1}\,x^{c+1} \quad \text{für } c \neq -1, \tag{7.33}$$

$$\int x^{-1}\,\mathrm{d}x = \ln|x|, \tag{7.34}$$

$$\int \cos x\,\mathrm{d}x = \sin x, \tag{7.35}$$

$$\int \sin x\,\mathrm{d}x = -\cos x, \tag{7.36}$$

$$\int \mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x. \tag{7.37}$$

Beachte, dass stets nur eine Stammfunktion angegeben ist. Addition beliebiger Konstanten liefert alle weiteren Stammfunktionen. Dass eine Funktion Stammfunktion einer anderen Funktion ist, kann man leicht durch Ableiten der Stammfunktion überprüfen.

Aus den Eigenschaften von Ableitungen sieht man außerdem sofort

$$\int f(x) + g(x)\,\mathrm{d}x = \int f(x)\,\mathrm{d}x + \int g(x)\,\mathrm{d}x \tag{7.38}$$

und

$$\int \lambda f(x)\,\mathrm{d}x = \lambda \int f(x)\,\mathrm{d}x, \tag{7.39}$$

wobei diese Formeln für beliebige Auswahlen konkreter Stammfunktion gelten.

7.3.1.4 Beispiele Hier einige konkrete Beispiele zur manuellen Berechnung von Flächeninhalten mittels Riemann-Integralen.

Beispiel (Normalbereich)

Wir berechnen den Flächeninhalt des Normalbereichs

$$\left\{(x,y):\ 1 \le x \le 3,\ \frac{1}{2}\,x + 1 \le y \le x + 2\right\} \tag{7.40}$$

aus einem früheren Beispiel zu Normalbereichen als Differenz

$$\int_1^3 x + 2\,\mathrm{d}x - \int_1^3 \frac{1}{2}\,x + 1\,\mathrm{d}x = \int_1^3 \frac{1}{2}\,x + 1\,\mathrm{d}x. \tag{7.41}$$

Wegen

$$\int \frac{1}{2}\,x + 1\,\mathrm{d}x = \frac{1}{2}\int x^1\,\mathrm{d}x + \int x^0\,\mathrm{d}x = \frac{1}{4}\,x^2 + x \tag{7.42}$$

erhalten wir als Flächeninhalt

$$\int_1^3 \frac{1}{2}\,x + 1\,\mathrm{d}x = \frac{1}{4}\cdot 3^2 + 3 - \left(\frac{1}{4}\cdot 1^2 + 1\right) = 4 \tag{7.43}$$

Beispiel (unbeschränkte Fläche)

Wir berechnen den Flächeninhalt der unbeschränkten Fläche unter

$$f : [0,\infty) \to \mathbb{R}, \quad f(x) := \frac{x}{x^2+1} \tag{7.44}$$

(vgl. früheres Beispiel zu uneigentlichen Integralen) als uneigentliches Integral

$$\int_0^\infty \frac{x}{x^2+1}\,\mathrm{d}x = \lim_{b\to\infty} \int_0^b \frac{x}{x^2+1}\,\mathrm{d}x. \tag{7.45}$$

Aus einer Formelsammlung oder mittels einer am Ende des zweiten Semesters noch näher zu betrachtenden Integrationsregel (“Integration durch Substitution”)

findet man die Stammfunktion

$$\int \frac{x}{x^2+1}\,\mathrm{d}x = \ln(x^2+1) \tag{7.46}$$

(leicht überprüfbar durch Ableiten!). Somit

$$\int_0^b \frac{x}{x^2+1}\,\mathrm{d}x = \ln(b^2+1) - \ln(0^2+1) = \ln(b^2+1), \tag{7.47}$$

also

$$\lim_{b\to\infty} \int_0^b \frac{x}{x^2+1}\,\mathrm{d}x = \infty. \tag{7.48}$$

Das uneigentliche Integral ist bestimmt divergent, der Flächeninhalt also nicht endlich.

Beispiel (keine explizite Stammfunktion)

In der Wahrscheinlichkeitsrechnung spielt die sogenannte Gaußsche Glockenkurve

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}} \tag{7.49}$$

eine wichtige Rolle. Insbesondere Integrale der Form

$$\int_a^b \frac{1}{\sqrt{2\,\pi}}\,\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x \tag{7.50}$$

werden benötigt. Da die Funktion stetig ist, besitzt sie eine Stammfunktion. Diese lässt sich jedoch nicht explizit als Formel darstellen. Während man früher die Werte der Integrale für $a = -\infty$ und viele verschiedene b in langen Tabellen erfasst hat, berechnet man diese heute mit dem Computer mittels numerischer Integration.

7.3.2 SymPy

SymPy kann Stammfunktionen bilden und auch bestimmte Integrale (über den Hauptsatz der Differential- und Integralrechnung) berechnen.

```
import sympy
```

```
x = sympy.symbols('x')

# Stammfunktion
F = sympy.integrate(x ** 2 * sympy.sin(x), x)
display(F)

# bestimmtes Integral
i = sympy.integrate(x ** 2 * sympy.sin(x), (x, 0, sympy.pi))
display(i)

# uneigentliches Integral
i = sympy.integrate(x / (x ** 2 + 1), (x, 0, sympy.oo))
display(i)
```

```
-x**2*cos(x) + 2*x*sin(x) + 2*cos(x)
```

```
-4 + pi**2
```

```
oo
```

7.3.3 Numerische Integration

Verfahren zur numerischen Integration sind in allen mathematischen Software-Bibliotheken vorhanden, z.B. auch in SciPy (als Ergänzung zu NumPy). Mit diesen Verfahren können nur bestimmte Integrale berechnet werden, keine Stammfunktionen. Grundidee ist das Annähern der zu brechnenden Fläche durch einfache Teilflächen (Dreiecke, Rechtecke, Trapeze,...), deren Flächeninhalt aus expliziten Formeln bestimmt werden kann. Der Ansatz ist also sehr ähnlich zur Definition von Flächeninhalten mittels Überdeckungen aus Quadraten, nur dass die Teilflächen je nach konkretem Verfahren vielfältiger sein können. Insbesondere kommen gelegentlich auch krummlinig begrenzte Teilflächen (z.B. Parabelstücke) zum Einsatz.

Hier ein Beispiel mit `scipy.integrate.quad`:

```
import numpy as np
from scipy.integrate import quad

# zu integrierende Funktion als Python -Funktion
f = lambda x: x ** 2 * np.sin(x)

# alternativ:
#def f(x):
#    return x ** 2 * np.sin(x)
```

```
# Integration von 0 bis pi
quad(f, 0, np.pi)
```

```
(5.869604401089359, 6.516569951530559e -14)
```

Der erste zurückgegebene Wert ist der berechnete Wert des Integrals. Der zweite Wert ist eine Schätzung (!) des Fehlers im Vergleich zum exakten Wert des Integrals. Der Fehler liegt hier also sehr wahrscheinlich im Bereich von 10^{-13}, also erst in der 13. Nachkommastelle.

Im Gegensatz zur numerischen Differentiation arbeitet die numerische Integration sehr zuverlässig. Für stetige Funktion erhält man mit den Standardverfahren der Bibliotheken praktisch immer korrekte Ergebnisse. Vorsicht ist nur bei Polstellen oder anderweitigen Besonderheiten geboten.

8 Wahrscheinlichkeit

Grundlagen der Wahrscheinlichkeitsrechnung und ausgewählte Wahrscheinlichkeitsverteilungen.

8.1 Wahrscheinlichkeitsräume

Die Welt des Zufalls in mathematisch exakte Sprache zu fassen ist schwieriger als man denkt. Entsprechend lang hat dies gedauert. Eine allgemein anerkannte, hinreichend umfangreiche Theorie für Beschreibung und Untersuchung zufälliger Vorgänge stand erst in den 1930er-Jahren zur Verfügung.

Wie umgehen mit scheinbar widersprüchlichen oder mindestens der Intuition widersprechenden Feststellungen wie den folgenden?

- Würfelt man mehrfach nacheinander, so sind die Würfe völlig unabhängig voneinander. Aus den bereits gewürfelten Zahlen kann man nicht auf die nächste Zahl schließen. Bei einer großen Anzahl von Würfen erscheint jedoch jede Zahl etwa gleich oft.
- Hat man die Folge 33333333333333333 gewürfelt, so kann man nicht (!) schließen, dass nun andere Zahlen öfter als die 3 erscheinen müssten, damit letztlich alle Zahlen etwa gleich oft vorkommen.
- Bei 20 Würfen taucht die Folge 33333333333333333 im Mittel über viele Versuche gleich oft auf wie die Folge 36215326326132562425.

In diesem ersten Unterkapitel zur Wahrscheinlichkeitsrechnung beschäftigen wir uns mit der grundlegenden Modellierung von zufälligen Vorgängen, also mit der Frage: Wie können wir zufällige Vorgänge in der Sprache der Mathematik formulieren? Die entsprechenden Modelle werden wir als Wahrscheinlichkeitsräume bezeichnen. Sind die Bestandteile eines Wahrscheinlichkeitsraums festgelegt, so können alle relevanten Aspekte des zufälligen Vorgangs an diesem klar definierten mathematischen Objekt untersucht werden.

8.1.1 Zufallsexperimente und Elementarereignisse

Als zufällige Vorgänge sehen wir Vorgänge an, deren Ausgang oder Ergebnis vom Beobachter nicht vorhergesagt werden kann. Dabei spielt es keine Rolle, ob jemand anderes, der eventuell mehr Wissen besitzt, eine Vorhersage treffen könnte oder nicht. Solche Vorgänge werden auch als **Zufallsexperimente** bezeichnet.

Die Menge aller möglichen Ergebnisse eines Zufallsexperiments bezeichnen wir als **Ergebnismenge** und bezeichnen diese üblicherweise mit Ω. Oft wird dies eine Menge von Zahlen sein; prinzipiell kann die Ergebnismenge aber beliebige Objekte enthalten. Die einzelnen Elemente der Ergebnismenge heißen **Elementarereignisse**.

Beispiel (Würfel)

Als Zufallsexperiment betrachten wir das einmalige Werfen eines üblichen Spiel-

würfels mit sechs Seiten. Die natürliche Ergebnismenge ist

$$\Omega := \{1, 2, 3, 4, 5, 6\}. \tag{8.1}$$

Beispiel (Rakete)

Das Bestimmen der Flughöhe einer Silvesterrakete kann als Zufallsexperiment betrachtet werden. Hier ist

$$\Omega := [0, \infty) \tag{8.2}$$

eine geeignete Wahl. Natürlich wird eine Silvesterrakete nicht beliebig hoch fliegen, sondern beispielsweise stets unterhalb einer Höhe von 1000 Meter bleiben. Man könnte also auch nur das Intervall $[0, 1000)$ betrachten. Die Diskussion, ob dies wirklich gerechtfertigt ist, ist jedoch recht mühsam, sodass wir uns das Leben mit $[0, \infty)$ deutlich leichter machen (und darum geht es in der Mathematik!).

8.1.2 Ereignisfelder

Oft möchte man nicht jedes Elementarereignis separat betrachten, sondern ist nur an gewissen Kombinationen von Elementarereignissen interessiert.

Merke

Die Teilmengen der Ergebnismenge Ω eines Zufallsexperiments werden als **Ereignisse** bezeichnet.

Man sagt, dass das Ereignis $A \subseteq \Omega$ **eingetreten** ist, wenn das Ergebnis $\omega \in \Omega$ des Zufallsexperiments in A liegt, also $\omega \in A$ gilt.

Beispiel (Würfel)

Interessiert beim Würfeln nur die Frage, ab eine gerade Zahl gewürfelt wurde, so sieht man beispielsweise die Ergebnisse 1 und 3 als gleichwertig an. Genauer: Es interessiert nur die Frage, ob des Ergebnis des Zufallsexperiments in der Menge $\{1, 3, 5\}$ oder in der Menge $\{2, 4, 6\}$ liegt.

Beispiel (Rakete)

Bei der Flughöhe einer Silvesterrakete wird nicht die Frage interessieren, ob die Rakete exakt 15.4 Meter hoch geflogen ist, sondern ob sie beispielsweise zwischen 15 und 20 Meter hoch geflogen ist oder ob sie mindestens 20 Meter hoch geflogen

ist. Die interessierenden Ereignisse sind also nicht die Elementarereignisse, sondern Intervalle wie $(a, b]$ und (a, ∞).

Hat man für ein Zufallsexperiment alle interessierenden Ereignisse festgelegt, so ist es zweckmäßig, diese um all jene Ereignisse zu erweitern, die bei üblichen logischen Operationen daraus entstehen können:

- Das **Gegenereignis** von $A \subseteq \Omega$ ist $\overline{A} := \Omega \setminus A$.
- Zu zwei Ereignissen A und B beschreibt der Durchschnitt $A \cap B$ gerade das Ereignis, das genau dann eintritt, wenn A **und** B eintreten.
- Zu zwei Ereignissen A und B beschreibt die Vereinigung $A \cup B$ gerade das Ereignis, das genau dann eintritt, wenn A **oder** B oder beide eintreten.

Die Und- und die Oder-Verknüpfung kann man auch auf mehr als zwei Ereignisse und sogar auf Folgen $(A_1, A_2, \ldots)$ von Ereignissen ausdehnen:

$$\bigcap_{k=1}^{n} A_k, \quad \bigcap_{k=1}^{\infty} A_k, \quad \bigcup_{k=1}^{n} A_k, \quad \bigcup_{k=1}^{\infty} A_k. \tag{8.3}$$

Außerdem sind das **sichere Ereignis** Ω and das **unmögliche Ereignis** $\emptyset$ (leere Menge) für das Formulieren mancher Sachverhalte recht nützlich.

Ergänzt man die eigentlich interessierenden Ereignisse um all diese zusätzlichen Ereignisse, so entsteht ein sogenanntes Ereignisfeld:

Merke

Eine Menge $\mathcal{A}$ (großen A in Schreibschrift) von Ereignissen zu einem Zufallsexperiment mit Ergebnismenge Ω heißt **Ereignisfeld** (oder σ-Algebra, sprich: Sigma-Algebra), wenn folgende Bedingungen erfüllt sind:

- $\Omega \in \mathcal{A}$,
- $\overline{A} \in \mathcal{A}$ für alle $A \in \mathcal{A}$,
- $\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$ für alle Folgen $(A_k)_{k \in \mathbb{N}}$ mit $A_k \in \mathcal{A}$.

Man überlegt sich leicht, dass auch $\emptyset$ und Durchschnitte von Mengen aus einem Ereignisfeld wieder zu dem Ereignisfeld gehören. Dazu sei an die de-Morganschen-Regeln erinnert:

Merke

Für beliebige Teilmengen A und B einer Menge Ω gilt stets

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \quad \text{und} \quad \overline{A \cup B} = \overline{A} \cap \overline{B}. \tag{8.4}$$

Völlig analoge Zusammenhänge gelten für die Komplemente von Durchschnitten und Vereinigungen beliebig vieler Mengen.

Beispiel (Würfel)

Ergänzt man die beiden Ereignisse $\{1, 3, 5\}$ und $\{2, 4, 6\}$ beim Würfeln zu einem Ereignisfeld $\mathcal{A}$, so erhält man

$$\mathcal{A} = \big\{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\big\}. \tag{8.5}$$

Beispiel (Rakete)

Bildet man für das Raketen-Beispiel das kleinste Ereignisfeld, welches alle Intervalle $(a, b]$ und (a, ∞) beinhaltet, so erhält man das sogenannte **Borelsche Ereignisfeld** zu $\Omega = [0, \infty)$. Dieses ist sehr umfangreich, aber kleiner als die Potenzmenge $\mathcal{P}([0, \infty))$. Wir gehen hier nicht näher auf die genaue Struktur des Borelschen Ereignisfeldes ein, sondern begnügen uns mit der Feststellung, dass alle praktisch relevanten Mengen (also Ereignisse) darin enthalten sind.

Randbemerkung: Allein die Intervalle (a, ∞) erzeugen schon das Borelsche Ereignisfeld zu $[0, \infty)$, aber auch zu ganz $\mathbb{R}$. Alternative kann am in $\mathbb{R}$ auch die Intervalle $(-\infty, b)$ als Grundlage nutzen.

Die Rolle von Ereignisfeldern ist vergleichbar mit der Rolle der Aufspannung von Vektoren. Während die Aufspannung einer handvoll interessierender Vektoren so beschaffen ist, dass wir bei den üblichen Vektoroperationen (skalares Vielfaches, Addition) immer in der Menge (also der Aufspannung) bleiben, ist das Ereignisfeld zu einer handvoll interessierender Ereignisse so beschaffen, dass wir bei den üblichen logischen Verknüpfungen von Ereignissen in der Menge (also dem Ereignisfeld) bleiben.

Nebenbemerkung (Wahl der Elementarereignisse)

Die Entscheidung, wie die Elementarereignisse und daraus die eigentlich interessierenden Ereignisse gewählt werden, ist in hohem Maße eine Geschmacksfrage. Bei vielen Zufallsexperimenten bietet sich eine “natürliche” Wahl der Elementarereignisse an; diese ist jedoch nicht zwingend!

Interessiert beim Würfeln nur die Frage, ob die gewürfelte Zahl gerade oder un-

gerade ist, so kann man problemlos auch die zweielementige Menge

$$\Omega := \{0, 1\} \tag{8.6}$$

als Ergebnismenge wählen, wobei 0 für "gerade" steht und 1 für "ungerade" (natürlich können auch andere Symbole an Stelle von 0 und 1 verwendet werden). Das Ereignisfeld wäre dann

$$\mathcal{A} := \big\{\emptyset, \{0\}, \{1\}, \{0, 1\}\big\}, \tag{8.7}$$

also die Potenzmenge von Ω (Menge aller Teilmengen). Diese Wahl von Ω und $\mathcal{A}$ liefert ein in jeder Hinsicht genauso brauchbares Modell des zufälligen Vorgangs wie die "natürliche" Wahl!

Umgekehrt kann man Ω auch komplexer wählen als es "natürlich" erscheint: Wir könnten die Lage des Würfels nach dem Würfeln durch drei Raumkoordinaten für die Position und zwei Winkel für die Drehung des Würfels beschreiben. Die Ergebnismenge wäre dann also $\Omega := \mathbb{R}^3 \times [0, \frac{\pi}{2}] \times [0, 2\,\pi)$. Das Ereignis "gerade Zahl gewürfelt" entspricht dann der Teilmenge de von Ω, deren Elemente eine Position und eine Drehung des Würfels beschreiben, bei der eine gerade Zahl nach oben zeigt. Mit einer so detaillierten Ergebnismenge könnte man auch das Ereignis "Würfel ist vom Tisch gefallen" erfassen.

Es wird sich zeigen, dass die konkrete Wahl von Ω praktisch gar keine große Rolle spielt. Im Gegenteil: Meist werden wir bei der Modellierung von Zufallsexperimenten Ω gar nicht explizit angeben, sondern Ω nur als "hinreichend detailliert" voraussetzen, ohne dies jedesmal explizit zu sagen.

8.1.3 Wahrscheinlichkeitsmaße

Für die vollständige Beschreibung eines zufälligen Vorgangs fehlt noch das, was wir sowohl intuitiv als auch in der Mathematik als Wahrscheinlichkeit bezeichnen, also ein Maß, welches jedem Ereignis eine Zahl zuordnet, die Informationen darüber liefert, wie sehr mit dem Eintreten der einzelnen Ereignisse zu rechnen ist.

Ereignissen, die ziemlich sicher nicht eintreten werden, soll die Zahl 0 zugeordnet werden. Ereignissen, die mit Sicherheit eintreten werden, soll der größtmögliche Wert zugeordnet werden. Diesen legen wir auf 1 fest. Eine weitere intuitiv klare Eigenschaft dieser Zuordnung ist folgende: Sind zwei Ereignisse A und B disjunkt, treten also nie gleichzeitig ein (in Formeln: $A \cap B = \emptyset$), so soll dem Ereignis $A \cup B$ (d.h. "A tritt ein oder B tritt ein oder beide") gerade die Summe der Werte der beiden Einzelereignisse zugeordnet werden. Dieser Zusammenhang soll natürlich nicht nur für zwei, sondern auch für mehr (paarweise disjunkte) Ereignisse gelten.

Merke

Sei Ω die Ergebnismenge eines Zufallsexperiments und sei $\mathcal{A}$ ein Ereignisfeld dazu. Eine Funktion $P : \mathcal{A} \to \mathbb{R}$ heißt **Wahrscheinlichkeitsmaß**, wenn folgende Bedingungen gelten:

- $0 \leq P(A) \leq 1$ für alle $A \in \mathcal{A}$,
- $P(\Omega) = 1$,
- $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$ für alle $A_1, A_2, \ldots \in \mathcal{A}$ mit $A_k \cap A_l = \emptyset$ für $k \neq l$.

Die Gleichheit in der dritten Forderung kann man auch als

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \tag{8.8}$$

schreiben.

Die drei hier formulierten Forderungen an ein Wahrscheinlichkeitsmaß sind als **Kolmogorovsches Axiomensystem** bekannt. Aus diesen folgen leicht weitere Eigenschaften von Wahrscheinlichkeitsmaßen (Herleitungen: Video 810, Notizen dazu):

- $P(\emptyset) = 0$,
- $P(\overline{A}) = 1 - P(A)$ für alle $A \in \mathcal{A}$,
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ für alle $A, B \in \mathcal{A}$.

Es genügt, ein Wahrscheinlichkeitsmaß auf einer hinreichend großen Teilmenge von $\mathcal{A}$ festzulegen. Die Wahrscheinlichkeiten der anderen Ereignisse ergeben sich dann aus den allgemein gültigen Eigenschaften von Wahrscheinlichkeitsmaßen.

Beispiel (Würfel)

Betrachten das Würfeln mit einem Würfel und interessieren uns für die gewürfelte Zahl, also für die Ereignisse $\{1\}, \ldots, \{6\}$. Das daraus erzeugte Ereignisfeld ist die Potenzmenge $\mathcal{A} := P(\Omega)$ der Ergebnismenge $\Omega := \{1, 2, 3, 4, 5, 6\}$.

Alle Zahlen haben die gleiche Chance, nach oben zu zeigen. Wir setzen also

$$P(\{1\}) := \frac{1}{6}, \quad \ldots, \quad P(\{6\}) := \frac{1}{6}. \tag{8.9}$$

Die Werte für alle anderen Ereignisse in $\mathcal{A}$ ergeben sich daraus. Zum Beispiel muss

$$P(\{1, 2\}) = P(\{1\}) + P(\{2\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \tag{8.10}$$

gelten.

Beispiel (Rakete)

Eine explizite Angabe des Wahrscheinlichkeitsmaßes P für die Flughöhe einer Silvesterrakete ist schwierig. Dazu drei Gedanken:

- Dass eine vorgegebene Flughöhe x exakt (inklusive aller Nachkommastellen) erreicht wird ist praktisch ausgeschlossen. Es gibt unendlich viele verschiedene Flughöhen, sodass im Sinne der relativen Häufigkeit des Treffen einer konkreten Höhe bei "$\frac{1}{\infty}$", also Null liegt. Somit erscheint $P(\{x\}) = 0$ für alle Elementarereignisse als sinnvolle Wahl.
- Nach hinreichend vielen Versuchen kommen wir zu der Meinung, dass etwa in der Hälft der Fälle die Flughöhe zwischen 10 und 15 Meter liegt. Deshalb erscheint $P([10, 15]) = 0.5$ gerechtfertigt, obwohl alle Elementarereignisse Wahrscheinlichkeit Null haben sollten (vgl. vorheriger Punkt).
- Führen wir das Zufallsexperiment durch, so wird eine konkrete Höhe x erreicht. Daher könnte man meinen, dass zwangsläufig $P(\{x\}) > 0$ gelten muss, denn unsere Intuition sagt: Wenn das Ereignis tatsächlich eingetreten ist, kann seine Wahrscheinlichkeit nicht bei Null liegen. Dieser Gedanke steht jedoch im Widerspruch zum ersten Punkt.

Wir werden diese Problematik im nächsten Unterkapitel vertiefen und aufklären. Für den Moment halten wir fest:

- $P(\{x\}) = 0$ für alle x ist korrekt.
- $P([10, 15]) = 0.5$ steht dazu nicht (!) im Widerspruch.
- Auch Ereignisse mit Wahrscheinlichkeit Null können eintreten.

Das Beispiel zeigt, dass noch Arbeit und Erkenntnisse vor uns liegen, obwohl wir die drei wesentlichen Definitionen schon eingeführt haben.

Zusammenfassend führt man nun folgenden Begriff ein:

Merke

Ein **Wahrschlichkeitsraum** ist ein Tripel $(\Omega, \mathcal{A}, P)$ aus Ergebnismenge Ω, zugehörigem Ereignisfeld $\mathcal{A}$ und zugehörigem Wahrscheinlichkeitsmaß P.

Mehr Beispiele für das Modellieren von zufälligen Vorgängen als Wahrscheinlichkeitsräume schauen wir uns in den nachfolgenden Unterkapiteln an.

8.1.4 Interpretation von Wahrscheinlichkeiten

Auch wenn man sich seit etwa 100 Jahren auf das hier beschriebene mathematische Modell für zufällige Vorgänge geeinigt hat, gibt es nach wie vor zwei unterschiedliche Schulen/Denkweisen/Interpretationen in der Wahrscheinlichkeitsrechnung:

- Bei der **frequentistischen** Denkweise wird die Wahrscheinlichkeit $P(A)$ eines Ereignisses A als Grenzwert der relativen Häufigkeit des Eintretens dieses Ereignisses bei wachsender Versuchsanzahl angesehen. Die Wahrscheinlichkeit eines Ereignisses gibt also an, wie oft dieses Ereignis im Mittel über sehr viele Wiederholungen des Zufallsexperiments eintreten wird. Die Wahrscheinlichkeit von $\frac{1}{6}$ für eine Drei beim Würfeln bedeutet somit, dass bei sehr vielen Würfen etwa ein Sechstel der Würfe eine Drei liefern. Problematisch bei der frequentistischen Interpretation ist der Umgang mit Versuchen, die keine beliebig häufige Wiederholung zulassen. Wie soll man beispielsweise in frequentistischem Sinn die "Wahrscheinlichkeit für eine Kernschmelze" in einem Atomkraftwerk interpretieren?
- Bei der **bayesianischen** Denkweise wird die Wahrscheinlichkeit $P(A)$ eines Ereignisses A als Größe zur zahlenmäßigen Erfassung ("Quantifizierung") der Unsicherheit bzw. Unkenntnis über das Eintreten des Ereignisses angesehen. Diese Sichtweise umgeht das Wiederholungsproblem der Frequentisten, führt aber zu einer gewissen Subjektivität bei der Größe von Wahrscheinlichkeiten. Auch ist hier nicht sofort klar, warum Wahrscheinlichkeiten immer im Intervall $[0, 1]$ liegen sollten.

Beide Denkweisen sind in der Literatur etwa gleich stark vertreten und liefern sich regelmäßig harte Diskussionen, wer die zweckmäßigere Interpretation vertritt. Aus beiden Sichtweisen entstanden und entstehen wichtige Werkzeuge und Konzepte, sodass man keine von beiden einfach ablehnen kann.

Wir folgen hier durchgängig der frequentistischen Interpretation, da diese einen deutlich einfacheren Zugang zu den Grundlagen der Wahrscheinlichkeitsrechnung ermöglicht.

8.2 Zufallsgrößen

Die praktische Handhabung von Ereignissen als Elemente des Ereignisfeldes $\mathcal{A}$ eines Wahrscheinlichkeitsraums $(\Omega, \mathcal{A}, P)$ ist eher mühsam. Angenehmer ist die Arbeit mit Funktionen, die zufällige Werte annehmen. Jede Wiederholung eines zufälligen Vorgangs kann dann einen anderen Funktionswert liefern. Dies erreicht man durch das Einführen von sogenannten Zufallsgrößen.

8.2.1 Begriff

Merke

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \to \mathbb{R}$ heißt **Zufallsgröße**, wenn für beliebige Intervalle der Form $(-\infty, b)$ das Urbild

$$X^{-1}\big((-\infty, b)\big) := \big\{\omega \in \Omega : \ X(\omega) \in (-\infty, b)\big\} \tag{8.11}$$

zu $\mathcal{A}$ gehört.

Die mittels einer Zufallsgröße erzeugten Werte heißen **Realisierungen** der Zufallsgröße.

Nebenbemerkung

Statt $(-\infty, b)$ könnte man zum Beispiel auch Intervalle der Form (a, ∞) in der Definition verwenden. Aufgrund der Eigenschaften von Ereignisfeldern gehören die Urbilder der jeweils anderen Form automatisch auch zu $\mathcal{A}$, denn es gilt beispielsweise

$$(-\infty, b) = \bigcup_{k=1}^{\infty} \left(-\infty, b - \frac{1}{k}\right] = \bigcup_{k=1}^{\infty} \overline{\left(b - \frac{1}{k}, \infty\right)} \tag{8.12}$$

und somit

$$X^{-1}\big((-\infty, b)\big) = \bigcup_{k=1}^{\infty} \overline{X^{-1}\left(\left(b - \frac{1}{k}, \infty\right)\right)}. \tag{8.13}$$

Da Komplemente und Vereinigungen von Elementen eines Ereignisfeldes wieder zum Ereignisfeld gehören, gehören mit den Urbildern von Intervallen der Form (a, ∞) also auch Urbilder von Intervallen der Form $(-\infty, b)$ zum Ereignisfeld $\mathcal{A}$.

Auch kann man sich leicht überlegen, dass dann alle Urbilder endlicher Intervalle (a, b), $[a, b]$, $(a, b]$, $[a, b)$ zu $\mathcal{A}$ gehören müssen und auch Urbilder vieler weiterer Mengen. Betrachtet man die Intervalle $(-\infty, b)$ als Ereignisse, so ist des daraus erzeugte Ereignisfeld $\tilde{\mathcal{A}}$ gerade das Borelsche Ereignisfeld (vgl. Beispiel (Rakete)). Die zugrundeliegende Ergebnismenge ist $\tilde{\Omega} := \mathbb{R}$. Alle Urbilder $X^{-1}(A)$ für

Mengen A aus diesem Ereignisfeld $\tilde{\mathcal{A}}$ gehören automatisch zum Ereignisfeld $\mathcal{A}$.

Zufallsgrößen bieten keine zusätzlichen Erkenntnisse oder Möglichkeiten, sondern dienen nur der Bequemlichkeit bei der Arbeit mit Wahrscheinlichkeitsräumen. Dies geht soweit, dass man die Grundmenge Ω und das Ereignisfeld $\mathcal{A}$ gar nicht mehr konkret angibt, sondern lediglich den Wertebereich der betrachteten Zufallsgröße und die Wahrscheinlichkeiten

$$P(X < x) := P\big(X^{-1}\big((-\infty, x)\big)\big) = P(\{\omega \in \Omega :\ X(\omega) < x\}). \tag{8.14}$$

für alle $x \in \mathbb{R}$. Durch die Angabe der Wahrscheinlichkeiten $P(X < x)$ sind automatisch auch die Wahrscheinlichkeiten aller anderen relevanten Ereignisse bestimmt (Video 820, Notizen dazu)

Merke

Ist der Wertebereich einer Zufallsgröße X eine endliche Menge $\{x_1, \ldots, x_n\}$ oder eine abzählbar unendliche Menge $\{x_1, x_2, \ldots\}$, so heißt die Zufallsgröße **diskret**. Ist der Wertebereich ein (ggf. unendliches) Intervall, so heißt die Zufallsgröße **kontinuierlich**.

Prinzipiell kann es auch Mischformen geben; diese sind aber praktisch nicht weiter relevant.

Bei diskreten Zufallsgrößen ist das Wahrscheinlichkeitsmaß auch durch die Angabe der Wahrscheinlichkeiten

$$P(X = x_k) \quad \text{für alle } k \tag{8.15}$$

festgelegt. Daraus ergeben sich dann die Wahrscheinlichkeiten $P(X < x)$ für alle $x \in \mathbb{R}$ und alle weiteren relevanten Wahrscheinlichkeiten.

Beispiel (Würfel)

Sei X eine diskrete Zufallsgröße mit Wertebereich $\{1, 2, 3, 4, 5, 6\}$ und zugehörigen Wahrscheinlichkeiten $P(X = k) = \frac{1}{6}$, $k = 1, \ldots, 6$. Das 10-malige Würfeln mit einem Würfel kann dann als 10 Realisierungen dieser Zufallsgröße beschrieben werden.

Eine explizite Angabe der Abbildungsvorschrift zu X auf Basis der Ergebnismenge $\Omega := \{1, 2, 3, 4, 5, 6\}$ ist

$$X(\omega) = \omega, \quad \omega \in \Omega. \tag{8.16}$$

Eine solche explizite Darstellung ist jedoch für die Arbeit mit Zufallsgrößen nicht nötig. Man geht einfach davon aus, dass ein hinreichend detaillierter Wahrscheinlichkeitsraum “im Hintergrund” vorhanden ist.

Beispiel (Würfel)

Die Frage, ob eine gerade oder eine ungerade Zahl gewürfelt wurde, können wir als diskrete Zufallsgröße X mit dem Wertebereich $\{0, 1\}$ (0 für "gerade", 1 für "ungerade") und zugehörigen Wahrscheinlichkeiten $P(X = 0) = \frac{1}{2}$ und $P(X = 1) = \frac{1}{2}$ formulieren.

Die (nicht benötigte) explizite Abbildungsvorschrift zu X auf Basis der Ergebnismenge $\Omega := \{1, 2, 3, 4, 5, 6\}$ ist dann

$$X(\omega) := \begin{cases} 0, & \text{für } \omega \in \{2, 4, 6\}, \\ 1, & \text{für } \omega \in \{1, 3, 5\}. \end{cases} \tag{8.17}$$

Im ersten Beispiel gibt die Zufallsgröße die Struktur des dahinter liegenden Wahrscheinlichkeitsraums vollständig wieder. Im zweiten Beispiel wird die Struktur der möglichen Ausgänge des zufälligen Vorgangs deutlich vereinfacht bzw. vergröbert. Der umgekehrte Fall, also ein Verfeinern der Beschreibung eines zufälligen Vorgangs mittels Zufallsgrößen, ist nicht möglich, da aufgrund der Urbildeigenschaft von Zufallsgrößen höchstens so viele verschiedene Ereignisse beschrieben werden können wie im Ereignisfeld des zugrundeliegenden Wahrscheinlichkeitsraums vorhanden sind.

Beispiel (Würfel)

Modellieren wir die Frage nach einer geraden bzw. ungeraden Augenzahl mit

$$\Omega := \{1, 2, 3, 4, 5, 6\} \tag{8.18}$$

und

$$\mathcal{A} := \big\{\emptyset,\, \{1, 3, 5\},\, \{2, 4, 6\}\, \{1, 2, 3, 4, 5, 6\}\big\}, \tag{8.19}$$

so können wir zwar eine Funktion

$$X : \Omega \to \mathbb{R}, \quad X(\omega) := \omega \tag{8.20}$$

definieren, die die gewürfelte Augenzahl liefert. Diese ist aber keine (!) Zufallsgröße, da sie die Urbildeigenschaft verletzt. Beispielsweise gilt

$$X^{-1}\big((-\infty, 2)\big) = \{1\} \notin \mathcal{A}. \tag{8.21}$$

8.2.2 Verteilungsfunktionen und Quantile

Wir führen noch zwei Begriffe ein, die inhaltlich keine neuen Erkenntnisse liefern, aber die Notation etwas vereinfachen und in der Literatur sehr verbreitet sind.

Merke

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und sei $X : \Omega \to \mathbb{R}$ eine Zufallsgröße. Die Funktion

$$F : \mathbb{R} \to [0, 1], \quad F(x) := P(X < x) \tag{8.22}$$

heißt **Verteilungsfunktion** zur Zufallsgröße X.

Aus den Betrachtungen oben folgt, dass mit der Kenntnis der Verteilungsfunktion einer Zufallsgröße die Wahrscheinlichkeiten aller relevanten bzgl. der Zufallsgröße ausgedrückten Ereignisse berechnet werden können.

Merke

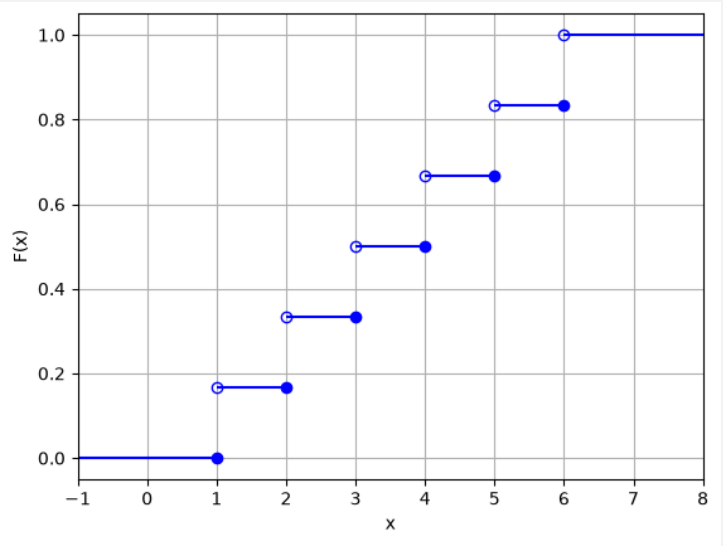
Man kann zeigen (tun wir aber nicht), dass die Verteilungsfunktion F zu einer Zufallsgröße stets folgende Eigenschaften besitzt:

- $\lim\limits_{x \to -\infty} F(x) = 0$,
- $\lim\limits_{x \to \infty} F(x) = 1$,
- F ist monoton wachsend,
- F ist linksseitig stetig, d.h. $\lim\limits_{x \to x_0 - 0} F(x) = F(x_0)$.

Beispiel (Würfel)

Die Verteilungsfunktion der Augenzahl bei einem Würfel ist

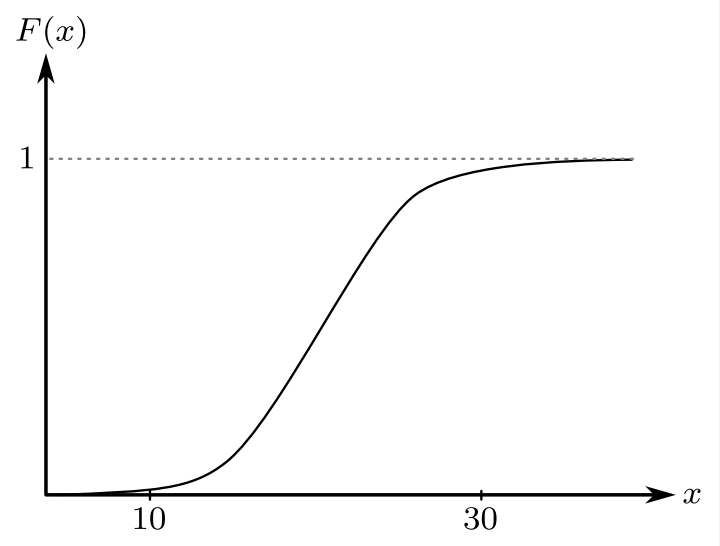
$$F(x) = \begin{cases} 0, & \text{für } x \leq 1, \\ \frac{1}{6}, & \text{für } 1 < x \leq 2, \\ \frac{2}{6}, & \text{für } 2 < x \leq 3, \\ \frac{3}{6}, & \text{für } 3 < x \leq 4, \\ \frac{4}{6}, & \text{für } 4 < x \leq 5, \\ \frac{5}{6}, & \text{für } 5 < x \leq 6, \\ 1, & \text{für } x > 6. \end{cases} \tag{8.23}$$



Beispiel (Rakete)

Nach wie vor sind wir nicht in der Lage, das Wahrscheinlichkeitsmaß oder (die inhaltlich völlig äquivalente) Verteilungsfunktion für die Flughöhe einer Silvesterrakete anzugeben. Dafür sind noch weitere Überlegungen nötig. Allerdings können wir uns anhand folgender Überlegungen schon eine grobe Vorstellung von der Verteilungsfunktion machen:

- Negative Flughöhen sind praktisch ausgeschlossen; also $F(x) = 0$ für $x \leq 0$.
- Kleine Flughöhen sind sehr unwahrscheinlich, sodass die Verteilungsfunktion zunächst nur langsam ansteigen wird.
- Sehr große Flughöhen sind unwahrscheinlich, sodass $F(x) = 1$ ab ca. $x = 30$ gelten wird.
- Zwischen etwa $x = 10$ und $x = 30$ erfolgt der Anstieg der Verteilungsfunktion. Etwa in der Mitte dieses Bereichs wird der Anstieg am stärksten sein, da dort die wahrscheinlichsten Flughöhen liegen.



Den folgenden Begriff kann man auch für nicht stetige und streng monotone Verteilungsfunktionen einführen. Dann muss man allerdings mit einigen Sonderfällen hantieren und die Darstellung wird aufwendiger. Da wir den Begriff nur bei stetigen und streng monotonen Verteilungsfunktionen benötigen werden, begnügen wir uns mit diesem Fall.

Merke

Sei X eine Zufallsgröße mit stetiger und streng monoton wachsender Verteilungsfunktion F und sei $t \in (0,1)$ eine Zahl. Die Zahl $x_t \in \mathbb{R}$, für die

$$F(x_t) = t \tag{8.24}$$

gilt, heißt t-**Quantil** von X.

Die Zuordnung $t \mapsto x_t$ des Quantils x_t zu einem Wert t der Verteilungsfunktion entspricht also gerade der Umkehrfunktion der Verteilungsfunktion. Quantile werden in der Statistik häufig benötigt.

8.2.3 Dichtefunktionen

Wie bereits erwähnt, genügt für die Festlegung des Wahrscheinlichkeitsmaßes einer diskreten Zufallsgröße mit Werten $\{x_1, \ldots, x_n\}$ die Angabe der Wahrscheinlicheiten

$$p_k := P(X = x_k), \quad k = 1, \ldots, n. \tag{8.25}$$

Daraus kann man die Wahrscheinlichkeit für alle anderen Ereignisse als Summen der p_k berechnen. Beispielsweise ist die Wahrscheinlichkeit $P(a < X < b)$ gerade die Summe aller p_k, für die $x_k \in (a, b)$ gilt.

Für diskrete Zufallsgrößen mit unendlich vielen Werten kann man völlig analog vorgehen. Bei Ereignissen, die unendlich viele Werte x_k umfassen, erhält man statt einer Summe eine Section 4.3.6; für das sichere Ereignis beispielsweise

$$P(X \in \mathbb{R}) = \sum_{k=1}^{\infty} p_k = 1. \tag{8.26}$$

Für kontinuierliche Zufallsgrößen wäre eine ähnlich einfache Beschreibung des Wahrscheinlichkeitsmaßen sehr nützlich. Da aber in nahezu allen praktisch relevanten Fällen $P(X = x) = 0$ für alle x gilt (vgl. Beispiel (Rakete)), ist das Vorgehen nicht offensichtlich. Ersetzt man die Summation jedoch durch ein Integral, so erhält man analog zu den p_k eine Funktion p, mit deren Hilfe die Wahrscheinlichkeiten aller Ereignisse als Integral ausgedrückt werden können:

Merke

Sei X eine kontinuierliche Zufallsgröße mit Wertebereich $W(X)$ und Verteilungsfunktion F. Eine Funktion $p : \mathbb{R} \to [0, \infty)$ heißt **Dichtefunktion** zu X, wenn p auf $W(X)$ und auf $\mathbb{R} \setminus W(X)$ stetig ist und

$$F(x) = \int_{-\infty}^{x} p(\tilde{x}) \, \mathrm{d}\tilde{x} \tag{8.27}$$

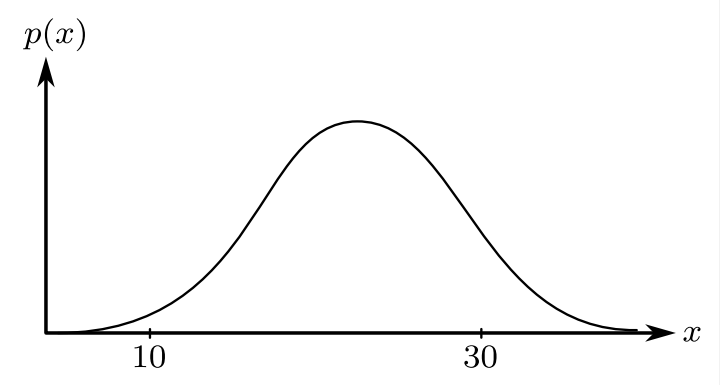
gilt.

Beachte, dass der Wertebereich einer kontinuierlichen Zufallsgröße stets ein Intervall ist. Eine Dichtefunktion soll laut Definition auf diesem Intervall und links davon und rechts davon stetig sein. Eine Dichtefunktion ist somit eine stückweise stetige Funktion mit höchstens zwei Sprungstellen.

Für unwahrscheinliche Ereignisse $[a, b]$ wird die Dichtefunktion $p(x)$ für $x \in [a, b]$ kleine Werte annehmen. Für wahrscheinliche Ereignisse wird $p(x)$ im entsprechenden Bereich groß sein. Die Werte $p(x)$ geben also zumindest einen relativen Eindruck von der Wahrscheinlichkeit des Ereignisses $\{x\}$, obwohl eigentlich $P(X = x) = 0$ gilt.

Beispiel (Rakete)

Obwohl wir noch keine exakte Dichtefunktion angeben können, sollte die Dichtefunktion für die Flughöhe einer Silvesterrakete etwa so aussehen:



Aus der Definition der Dichtefunktion erhält man sofort die beiden Eigenschaften

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1 \tag{8.28}$$

und

$$P(a \le X \le b) = \int_{a}^{b} p(x)\,\mathrm{d}x. \tag{8.29}$$

Außerdem kann man zeigen, dass für stetige Dichtefunktionen stets

$$p(x) = F'(x) \quad \text{für alle } x \in \mathbb{R} \tag{8.30}$$

gilt. Ist die Dichtefunktion nicht stetig, sondern nur stückweise stetig, so besitzt die Verteilungsfunktion Knickstellen bei den Sprüngen der Dichtefunktion (höchstens zwei). Zwischen den Knickstellen gilt wiederum $p(x) = F'(x)$.

Beachte, dass es auch kontinuierliche Zufallsgrößen gibt, für die keine Dichtefunktion existiert.

Beispiel (Lebensdauer einer Maschine)

Beschreiben die Lebensdauer einer Maschine als Zufallsgröße X mit Wertebereich $[0, \infty)$. Da die Maschine mit einer gewissen (hoffentlich kleinen) Wahrscheinlichkeit schon von Anfang an kaputt sein kann, gilt $P(X = 0) > 0$. Die Verteilungsfunktion hat also bei $x = 0$ einen Sprung.

Wenn es eine Dichtefunktion gäbe, so wäre die Verteilungsfunktion bis auf einzelne Knickstellen (Sprünge in der Dichtefunktion) differenzierbar. Sprünge in der Verteilungsfunktion könnten dann nicht vorkommen.

8.2.4 Zusammenfassung und Ausblick

Wir haben nun alle wichtigen Begriffe und Konzepte zur Beschreibung und zur Untersuchung von zufälligen Vorgängen eingeführt. Der nächste große Schritt ist das Untersuchen konkreter zufälliger Vorgänge, indem wir die gebildeten Begriffe darauf anwenden. Zuvor sei jedoch nochmal das bis hier entwickelte Gesamtbild kurz dargestellt:

Wir haben zwei aufeinander aufbauende Varianten zur Modellierung von zufälligen Vorgängen eingeführt. Einerseits können wir einen zufälligen Vorgang als Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ beschreiben. Andererseits können wir zufällige Vorgänge, deren Ergebnis sich als Zahl darstellen lässt, als Zufallsgröße mit zugehöriger Verteilungsfunktion (und ggf. Dichtefunktion) beschreiben.

Wir werden im Weiteren ausschließlich zufällige Vorgänge untersuchen, deren Ergebnis sich als Zahl darstellen lässt, sodass wir Wahrscheinlichkeitsräumen nicht mehr direkt begegnen werden. Als Unterbau für die Sprache der Zufallsgrößen sind Wahrscheinlichkeitsräume aber stets mit dabei. Wenigstens das Wahrscheinlichkeitsmaß P als wesentliche Komponente von Wahrscheinlichkeitsräumen werden wir weiterhin direkt nutzen.

In diesem Unterkapitel führen wir noch zwei Kenngrößen für Zufallsgrößen ein, die uns gewisse Informationen zu einer Zufallsgröße in leicht handhabbarer Form (nämlich als Zahl) liefern. Anschließend schauen wir uns Klassen von Verteilungsfunktionen an, die bei einer Vielzahl von zufälligen Vorgängen eine Rolle spielen. Abschließend beschäftigen wir uns mit dem Zusammenspiel mehrerer Zufallsgrößen aus verschiedenen Blickwinkeln.

8.2.5 Erwartungswert und Varianz

Für die meisten Zufallsgrößen können wir angeben, welchen Wert sie im Mittel über sehr viele Realisierungen annehmen werden.

Merke

Für eine diskrete Zufallsgröße mit endlich vielen Werten $\{x_1, \ldots, x_n\}$ und zugehörigen Wahrscheinlichekeiten $p_k := P(X = x_k)$ heißt die Zahl

$$\mathbb{E}X := \sum_{k=1}^{n} p_k \, x_k \tag{8.31}$$

Erwartungswert der Zufallsgröße.

Für eine diskrete Zufallsgröße mit unendlich vielen Werten $\{x_1, x_2, \ldots\}$ und zugehörigen Wahrscheinlichekeiten $p_k := P(X = x_k)$ heißt der Grenzwert

$$\mathbb{E}X := \sum_{k=1}^{\infty} p_k \, x_k \tag{8.32}$$

Erwartungswert der Zufallsgröße.

Für eine kontinuierliche Zufallsgröße mit Dichtefunktion p heißt das uneigentliche Integral

$$\mathbb{E}X := \int_{-\infty}^{\infty} p(x)\, x \,\mathrm{d}x \tag{8.33}$$

Erwartungswert der Zufallsgröße.

Bei diskreten Zufallsgrößen mit unendlich vielen Werten und bei kontinuierlichen Zufallsgrößen muss der Erwartungswert nicht existieren.

Wichtig ist die Unterscheidung des hier definierten Erwartungswerts einer Zufallsgröße vom in der Statistik verwendeten empirischen Erwartungswert (oder Mittelwert). Der **empirische Erwartungswert** ist das arithmetische Mittel aus einer endlichen Anzahl von Realisierungen der Zufallsgröße. Der Erwartungswert (einer Zufallsgröße) kann als Grenzwert des empirischen Erwartungswert bei wachsender Anzahl von Realisierungen angesehen werden; die präzise Formulierung dieses Zusammenhangs folgt in einem späteren Unterkapitel.

Beispiel (Würfel)

Gebe X die gewürfelte Augenzahl an. Dann ist der Erwartungswert

$$\mathbb{E}X = \frac{1}{6}1 + \frac{1}{6}2 + \frac{1}{6}3 + \frac{1}{6}4 + \frac{1}{6}5 + \frac{1}{6}6 = \frac{7}{2} = 3.5. \tag{8.34}$$

Für die 10 Realisierungen 3, 5, 1, 2, 2, 6, 4, 1, 2, 4 ist der empirische Erwartungswert

$$\frac{3+5+1+2+2+6+4+1+2+4}{10} = 3. \tag{8.35}$$

Kennt man den Erwartungswert einer Zufallsgröße, so stellt sich die Frage, wie nah die Realisierungen (im Mittel über sehr viele Versuche) beim Erwartungswert liegen werden. Diese "mittlere Abweichung vom Erwartungswert" kann auf verschiedene Weisen in Formeln ausgedrückt werden, z.B. als

$$\mathbb{E}|X - \mathbb{E}X|. \tag{8.36}$$

Ist X eine Zufallsgröße, so ist auch $|X - c|$ für $c \in \mathbb{R}$ eine Zufallsgröße. Als Verschiebung c können wir insbesondere den Erwartungswert wählen. Aufgrund angenehmerer Eigenschaften (die uns im Rahmen dieser Einführung nicht vollständig klar werden werden) wählt man jedoch üblicherweise ein anderes Maß für die "mittlere Abweichung":

Merke

Sei X eine Zufallsgröße mit Erwartungswert $\mathbb{E}X$. Die Zahl

$$\mathbb{V}\mathrm{ar}\, X := \mathbb{E}\big((X - \mathbb{E}X)^2\big) \tag{8.37}$$

heißt **Varianz** (oder Streuung) von X. Die Zahl $\sqrt{\mathbb{V}\mathrm{ar}\, X}$ heißt **Standardabweichung** von X.

Die Rolle des Quadrierens hier entspricht vom Quadrieren bei der Länge von Vektoren. Als Länge eines Vektors u verwenden wir natürlicherweise auch $\sqrt{\sum u_k^2}$ und nicht $\sum |u_k|$.

Da die Varianz ein Erwartungswert ist, muss die Varianz nicht (als endliche Zahl) existieren. Für übliche zufällige Vorgänge existiert sie aber. Wir gehen im Folgenden stets davon aus, dass alle in Formeln auftretenden Erwartungswerte und Varianzen exisiteren, ohne dies jedesmal explizit zu erwähnen.

Der Unterschied zwischen Varianz und Standardabweichung liegt ausschließlich in der Frage, ob man die Wurzel zieht (dann stimmt die Maßeinheit mit der von X überein) oder nicht (dann lässt sich leichter damit rechnen).

Beispiel (Würfel)

Gibt X die gewürfelte Augenzahl an, so ist die Varianz

$$\begin{aligned}\mathbb{V}\mathrm{ar}\, X = &\frac{1}{6}\left(1-\frac{7}{2}\right)^2 + \frac{1}{6}\left(2-\frac{7}{2}\right)^2 + \frac{1}{6}\left(3-\frac{7}{2}\right)^2 \\ &+ \frac{1}{6}\left(4-\frac{7}{2}\right)^2 + \frac{1}{6}\left(5-\frac{7}{2}\right)^2 + \frac{1}{6}\left(6-\frac{7}{2}\right)^2 \\ = &\frac{35}{12} \approx 2.92.\end{aligned} \tag{8.38}$$

Die Standardabweichung liegt somit bei

$$\sqrt{\mathbb{V}\mathrm{ar}\, X} = \sqrt{\frac{35}{12}} \approx 1.71. \tag{8.39}$$

Beispiel (Holzzuschnitt)

In einem Sägewerk werden Holzlatten auf eine Solllänge von 4 Meter geschnitten. Aufgrund diverser Ungenauigkeiten kann man die tatsächliche Länge der Latten als zufälligen Vorgang ansehen und diesen mit einer kontinuierlichen Zufallsgröße X modellieren. Wenn die Maschine korrekt arbeitet, ist der Erwartungswert von X genau 4. Die Standardabweichung drückt dann die mittlere Abweichung vom

Sollwert in Meter aus. Eine Standardabweichung von 0.01 bedeutet beispielsweise, dass im Mittel über sehr viele Latten die Länge um 1 Zentimeter von der Sollänge abweicht. Die zugehörige Varianz 0.0001 ist hingegen schwerer zu interpretieren.

Analog zum empirischen Erwartungswert versteht man unter der **empirischen Varianz** bzw. der **emprischen Standardabweichung** die mittlere quadratische Abweichung vom empirischen Erwartungswert bei einer endlichen Anzahl von Realisierungen einer Zufallsgröße.

Folgende Eigenschaften von Erwartungswert und Varianz lassen sich leicht anhand der Definitionen nachrechnen:

- $\mathbb{E}(a\,X + b) = a\,\mathbb{E}X + b \quad$ für alle $a, b \in \mathbb{R}$,
- $\mathbb{V}\mathrm{ar}\,(a\,X + b) = a^2\,\mathbb{V}\mathrm{ar}\,X \quad$ für alle $a, b \in \mathbb{R}$,
- $\mathbb{V}\mathrm{ar}\,X = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.

Die zweite Eigenschaft besagt insbesondere, dass das Verschieben des Wertebereichs einer Zufallsgröße keinen Einfluss auf die Varianz hat. Der Erwartungswert verschiebt sich hingegen entsprechend mit (vgl. erste Eigenschaft). Beim Skalieren des Wertebereichs werden Erwartungswert und Standardabweichung mitskaliert.

8.2.6 Standardisierung

Die folgende Transformation einer Zufallsgröße liefert stets eine Zufallsgröße mit Erwartungswert 0 und Varianz 1:

Merke

Sei X eine Zufallsgröße. Dann heißt die Zufallsgröße

$$Y := \frac{X - \mathbb{E}X}{\sqrt{\mathbb{V}\mathrm{ar}\,X}} \tag{8.40}$$

Standardisierung von X.

Beachte, dass wir stets stillschweigend davon ausgehen, dass die Varianz $\mathbb{V}\mathrm{ar}\,X$ einer Zufallsgröße X nicht Null ist. Man überlegt sich leicht, dass $\mathbb{V}\mathrm{ar}\,X = 0$ nur möglich ist, wenn X konstant ist, also die Funktionswerte in diesem Sinne gar nicht zufällig sind.

Die Standardisierung einer Zufallsgröße dient beispielsweise dazu, die Werte für von Parametern abhängigen Verteilungsfunktionen einfacher in Tabellen erfassen zu können. Statt für jede Parameterkonstellation eigene Tabellen vorhalten zu müssen, wird nur eine Tabelle für die Standardisierung der betrachteten Zufallsgrößen benötigt. Aus dem

Zusammenhang

$$P(X < x) = P\left(\frac{X - \mathbb{E}X}{\sqrt{\mathbb{V}\mathrm{ar}\, X}} < \frac{x - \mathbb{E}X}{\sqrt{\mathbb{V}\mathrm{ar}\, X}}\right) \tag{8.41}$$

lassen sich dann alle benötigten Wahrscheinlichkeiten berechnen.

Auch ermöglicht die Standardisierung einen einfacheren Vergleich zweier Zufallsgrößen bzw. deren Verteilungsfunktionen.

8.3 Diskrete Verteilungen

Verschiedene zufällige Vorgänge besitzen ähnliche Eigenschaften. So ist das Würfeln mit einem Spielwürfel vergleichbar mit dem Werfen einer Münze, nur dass es beim Würfeln sechs mögliche Ergebnisse gibt und beim Werfen einer Münze nur zwei. In diesem Unterkapitel schauen wir uns häufig auftretende Muster bei zufälligen Vorgängen an, die als diskrete Zufallsgröße modelliert werden können.

Wir beschränken uns dabei auf die wichtigsten Klassen und erwähnen andere nur kurz. Die Details findet man in Formelsammlungen und jedem umfangreicheren Lehrbuch zur Wahrscheinlichkeitsrechnung.

8.3.1 Überblick

Folgende Klassen von Verteilungsfunktion treten üblicherweise bei diskreten Zufallsgrößen auf:

- diskrete Gleichverteilung (endlich viele Werte; alle Werte gleich wahrscheinlich; siehe unten),
- Binomialverteilung (z.B. Anzahl defekter Teile bei Produktion einer vorgebenen Anzahl, siehe unten),
- Poisson-Verteilung (Anzahl des Eintretens seltener Ereignisse; z.B. Anrufe pro Tag; abzählbar unendlicher Wertebereich),
- geometrische Verteilung (z.B. Anzahl korrekt produzierter Bauteile bis zum ersten Fehlerhaften Teil),
- hypergeometrische Verteilung (z.B. Anzahl schwarzer Kugeln beim Ziehen einer festen Anzahl Kugeln aus einer Menge von schwarzen und weißen Kugeln).

8.3.2 Diskrete Gleichverteilung

Die diskrete Gleichverteilung deckt beispielweise das Würfeln und das Werfen einer Münze ab.

Merke

Eine diskrete Zufallsgröße mit Wertebereich $\{x_1, \ldots, x_n\}$ heißt **diskret gleichverteilt**, wenn

$$P(X = x_k) = \frac{1}{n}, \quad \text{für } k = 1, \ldots, n \tag{8.42}$$

gilt.

Für eine diskret gleichverteilte Zufallsgröße X gilt offensichtlich

$$\mathbb{E}X = \frac{x_1 + \cdots + x_n}{n}. \tag{8.43}$$

Beachte bei dieser Formel, dass es sich nicht (!) um den empirischen Erwartungswert von n Realisierungen der Zufallsgröße handelt, sondern um das arithmetische Mittel aus allen möglichen Werten, die die Zufallsgröße annehmen kann.

Für die Varianz erhalten wir

$$\mathbb{V}\mathrm{ar}\, X = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \frac{x_1^2 + \cdots + x_n^2}{n} - \frac{(x_1 + \cdots + x_n)^2}{n^2}. \tag{8.44}$$

8.3.3 Binomialverteilung

Als Standardbeispiel für binomialverteilte Zufallsgrößen dienen Produktionsprozesse, bei denen die produzierten Bauteile mit einer bekannten Wahrscheinlichkeit $p \in [0,1]$ defekt sind. Bei einer Gesamtproduktion von n Teilen kann man fragen, wie hoch die Wahrscheinlichkeit ist, dass genau $k \in \{0, 1, \ldots, n\}$ Teile defekt sind. Dabei gehen wir davon aus, dass die Frage, ob ein Teil defekt ist, nicht davon abhängt, ob die anderen produzierten Teile defekt sind oder nicht.

Merke

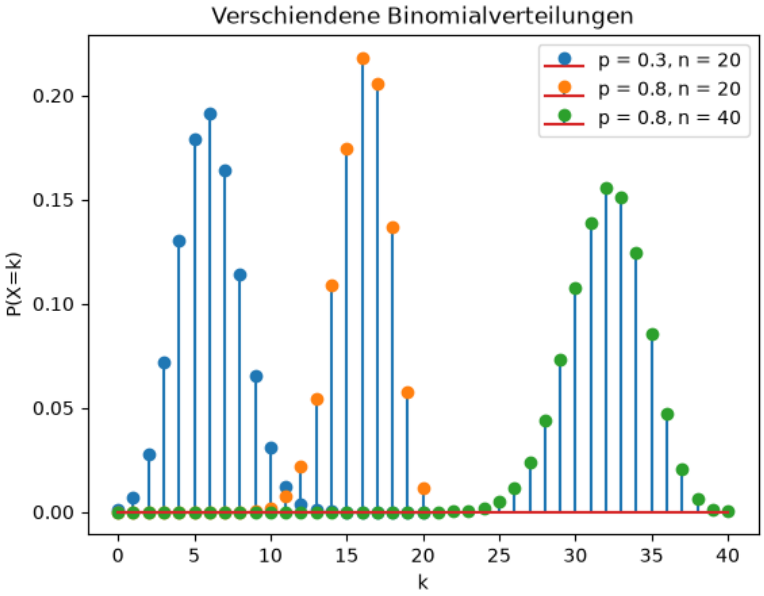
Eine Zufallsgröße mit Wertebereich $\{0, 1, \ldots, n\}$ heißt **binomialverteilt**, wenn

$$P(X = k) = \binom{n}{k} p^k \, (1-p)^{n-k} \quad \text{für } k = 0, 1, \ldots, n \tag{8.45}$$

mit einem Parameter $p \in [0,1]$ gilt.

Hier bezeichnet $\binom{n}{k}$ den aus der Schulmathematik bekannten Binomialkoeffizient

$$\binom{n}{k} := \frac{n!}{k!\,(n-k)!}. \tag{8.46}$$



Beispiel

Für eine Urlaubsregion wurde anhand langjähriger Beobachtung geschätzt, dass im August die Wahrscheinlichkeit für einen sonnigen Tag bei 80 Prozent liegt. Wir sind an der Wahrscheinlichkeit interessiert, dass bei 7 Urlaubstagen wenigstens 5 Tage sonnig sind.

Gehen wir davon aus, dass die Frage, ob die Sonne an einem Tag scheint, nicht davon abhängt, ob sie an anderen Tagen scheint oder nicht, dann können wir die Anzahl der Sonnentage als binomialverteilte Zufallsgröße modellieren. Mit den Bezeichnungen von oben gilt dann $n = 7$ und $p = 0.8$. Gesucht ist $P(X \geq 5)$. Wir erhalten

$$\begin{aligned} P(X \geq 5) &= P(X = 5) + P(X = 6) + P(X = 7) \\ &= \binom{5}{7} 0.8^5\, 0.2^2 + \binom{6}{7} 0.8^6\, 0.2^1 + \binom{7}{7} 0.8^7\, 0.2^0 \\ &\approx 0.85. \end{aligned} \tag{8.47}$$

Erwartungswert und Varianz einer mit Parametern n und p binomialverteilten Zufallsgröße X sind

$$\mathbb{E}X = n\,p \tag{8.48}$$

und

$$\mathbb{V}\mathrm{ar}\, X = n\,p\,(1-p) \tag{8.49}$$

(die Herleitung ist elementar, aber länglich, und entfällt deshalb).

8.4 Kontinuierliche Verteilungen

Wie bei den diskreten Zufallsgrößen treten auch bei den kontinuierlichen Zufallsgrößen gewissen Klassen von Verteilungsfunktionen immer wieder auf.

8.4.1 Überblick

Folgende Klassen sind am häufigsten anzutreffen:

- Gleichverteilung (siehe unten),
- Normalverteilung (siehe unten),
- Exponential-Verteilung (z.B. Lebensdauern oder Wartezeiten),
- Studentsche t-Verteilung (tritt bei statistischen Untersuchen normalverteilter Zufallsgrößen auf),
- χ^2-Verteilung (sprich: Chi-Quadrat-Verteilung; tritt bei statistischen Untersuchen normalverteilter Zufallsgrößen auf),
- F-Verteilung (tritt bei statistischen Untersuchen normalverteilter Zufallsgrößen auf).

8.4.2 Gleichverteilung

Bei gleichverteilten (kontinuierlichen) Zufallsgrößen treten alle Werte im Wertebereich gleich häufig auf. Misst man beispielsweise die Winkeldifferenz bei der Radposition eines Autos vor und nach einer hinreichend langen Fahrt, werden bei mehrfacher Wiederholung die gemessenen Werte alle einigermaßen gleichmäßig im Intervall $[0, 2\,\pi)$ verteilt sein.

Merke

Eine kontinuierlich Zufallsgröße X mit Dichtefunktion

$$p(x) = \begin{cases} \frac{1}{b-a}, & \text{für } x \in [a, b], \\ 0, & \text{sonst} \end{cases} \tag{8.50}$$

heißt **gleichverteilt** auf dem Intervall $[a, b]$.

Erwartungswert und Varianz einer auf dem Intervall $[a, b]$ gleichverteilten Zufallsgröße X sind

$$\mathbb{E}X = \frac{a+b}{2} \tag{8.51}$$

und

$$\mathbb{V}\mathrm{ar}\, X = \frac{(b-a)^2}{12} \tag{8.52}$$

(die entsprechenden Integrale sind leicht zu berechnen).

Die Verteilungsfunktion lässt sich ebenfalls leicht berechnen:

$$F(x) = \int_{-\infty}^{x} \frac{1}{b-a}\,\mathrm{d}x = \begin{cases} \frac{x-a}{b-a}, & \text{für } x \in [a,b], \\ 0, & \text{sonst.} \end{cases} \tag{8.53}$$

8.4.3 Normalverteilung

Normalverteilte Zufallsgrößen treten sehr häufig auf. Beispielsweise führen praktisch alle Arten von Messprozessen (Länge, Masse, Temperatur,...) bei Einbeziehen eventueller Messfehler zu normalverteilten Zufallsgrößen. Den genauen Grund dafür werden wir in einem späteren Unterkapitel erhellen. Kurzfassung: Die Überlagerung vieler unabhängiger Fehlerquellen führt zu normalverteilten Messergebnissen.

Merke

Eine kontinuierliche Zufallsgröße X mit Dichtefunktion

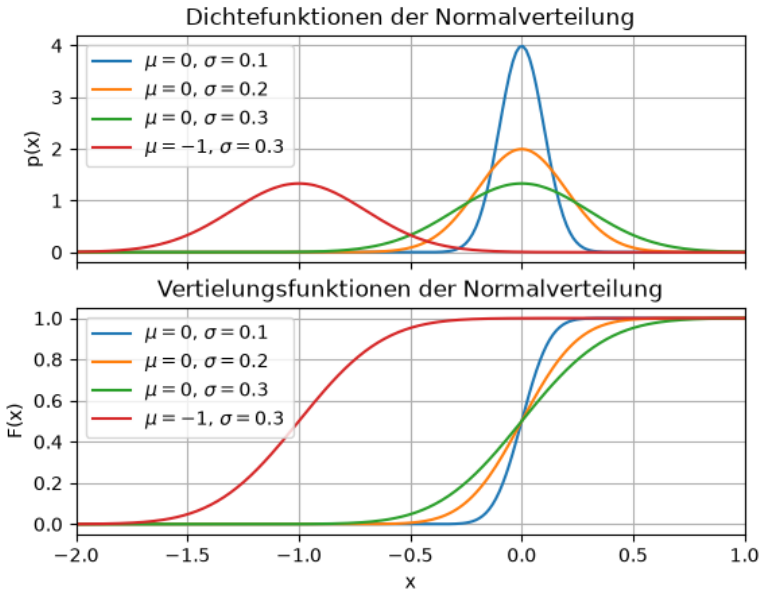
$$p(x) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\,\mathrm{e}^{-\frac{(x-\mu)^2}{2\,\sigma^2}} \tag{8.54}$$

heißt **normalverteilt** (oder Gauß-verteilt) mit den Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$.

Die beiden Parameter bei einer normalverteilten Zufallsgröße X stimmen gerade mit Erwartungswert und Standardabweichung überein:

$$\mathbb{E}X = \mu \qquad \text{und} \qquad \mathbb{V}\mathrm{ar}\,X = \sigma^2. \tag{8.55}$$

Die Verteilungsfunktion der Normalverteilung lässt sich nicht explizit als Formel angeben, da die Dichtefunktion keine als Formel darstellbare Stammfunktion besitzt. Die Auswertung der Verteilungsfunktion erfolgt durch numerische Integration (oder durch Ablesen aus Tabellen). Die meisten mathematischen Software-Bibliotheken bieten jedoch eine spezielle Funktion zu diesem Zweck an (z.B. `scipy.stats.norm.cdf`).



Beispiel (Rakete)

Die Flughöhe einer Silvesterrakete wird durch eine Vielzahl verschiedener Einflüsse festgelegt (Menge an Treibstoff, Lufttemperatur, Luftdruck, Wind,...). Deshalb wird die Flughöhe mit guter Genauigkeit normalverteilt sein. Die Parameter der Normalverteilung können mit Mitteln der Statistik aus hinreichend vielen Versuchen geschätzt werden. Zu Thematik der Parameterschätzung kommen wir aber erst im nächsten Kapitel.

Beispiel (Holzzuschnitt)

Von einer Zuschnittanlage in einem Sägewerk ist bekannt, dass die Holzlatten im Mittel auf eine Länge von 4 Meter geschnitten werden. Die über längere Zeit aus vielen Messungen statistisch geschätzte Standardabweichung liegt bei 1 Zentimeter. Wir interessieren uns für die Wahrscheinlichkeit, dass eine Latte mehr als 5 Millimeter zu kurz ist.

Da eventuelle Abweichungen von der Solllänge aus einer Vielzahl möglicher Fehlerquellen (temperaturbedingte Größen- und Formänderungen von Bauteilen der Säge, Ungenauigkeiten beim Vorschubmechanismus, Abnutzung Sägeblatt,...) resultieren, können wir davon ausgehen, dass die Länge der Latten durch eine normalverteilte Zufallsgröße realitätsnah modelliert werden kann. Die Parameter der Normalverteilung sind $\mu = 4$ und $\sigma = 0.01$. Gesucht ist die Wahrscheinlichkeit $P(X < 3.995)$.

Durch numerische Integration (oder aus entsprechenden Tabellen) erhalten wir $P(X < 3.995) \approx 31\,\%$.

```
from scipy.stats import norm as normal

normal.cdf(3.995, loc=4, scale=0.01)
```

```
np.float64(0.30853753872599066)
```

Die Wahrscheinlichkeit, dass die Länge um höchstens 5 Millimeter nach unten oder oben vom Soll abweicht, liegt bei

$$\begin{aligned} P(3.995 \leq X \leq 4.005) &= P(X \leq 4.005) - P(X < 3.005) \\ &= P(X < 4.005) - P(X < 3.005) \\ &\approx 38\,\%. \end{aligned} \tag{8.56}$$

```
from scipy.stats import norm as normal

normal.cdf(4.005, loc=4, scale=0.01) - normal.cdf(3.995, loc=4,
↪  scale=0.01)
```

```
np.float64(0.3829249225480187)
```

Die folgende wichtige Eigenschaft normalverteilter Zufallsgrößen lässt sich nur mühsam herleiten, sodass wir auf die Herleitung verzichten. Den hier verwendeten Begriff der Unabhängigkeit von zwei Zufallsgrößen führen wir im nächsten Unterkapitel präzise ein. Intuitiv ist er klar: Zwei Zufallsgrößen sind unabhängig, wenn sich das Eintreten zugehöriger Ereignisse in keiner Weise gegenseitig beeinflusst.

Merke

Sind X_1 und X_2 zwei unabhängige normalverteilte Zufallsgrößen mit Parametern μ_1 bzw. μ_2 und σ_1 bzw. σ_2, so ist die Summe $X_1 + X_2$ wieder eine normalverteilte Zufallsgröße mit Parametern $\mu_1 + \mu_2$ und $\sqrt{\sigma_1^2 + \sigma_2^2}$.

Beachte, dass das Bilden der Summe von zwei Zufallsgrößen nur sinnvoll ist, wenn beide Zufallsgrößen über dem selben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ definiert sind. Insbesondere müssen beide Zufallsgrößen den selben Definitionsbereich Ω besitzen, da die Summe zweier Funktionen mit unterschiedlichem Definitionsbereich nicht sinnvoll erklärt werden kann.

Folgenden Zusammenhang kann man sich leicht anhand der Dichtefunktion der Normalverteilung überlegen:

Merke

Sei X normalverteilt mit Parametern μ und σ. Dann ist die Standardisierung $\frac{X-\mu}{\sigma}$ wieder normalverteilt (mit Parametern 0 und 1).

8.5 Unabhängigkeit

Wir wollen klären, wie wir zwei oder auch mehr Zufallsgrößen ansehen, ob sie unabhängig sind, d.h. ob ihre (zufälligen) Werte unabhängig davon entstehen, welche Werte die andere Zufallsgröße oder die anderen Zufallsgrößen annehmen. Die Frage der Unabhängigkeit ist von großer Bedeutung bei der Arbeit mit mehreren Zufallsgrößen (was sehr oft nötig ist!). Sind die betrachteten Zufallsgrößen unabhängig, so lassen sich deren Verteilungen bzw. die Verteilungen von zusammenfassenden Zufallsgrößen viel leichter bestimmen.

Beachte, dass die Frage nach Unabhängigkeit nur relevant ist, wenn alle beteiligten Zufallsgrößen auf dem selben Wahrscheinlichkeitsraum definiert sind. Davon gehen wir bei der Arbeit mit mehreren Zufallsgrößen stets aus, wenn wir die Situation nicht explizit anders beschreiben.

8.5.1 Produkträume

Betrachten zur Motivation des Unabhängigkeitsbegriffs zwei aus Sicht des gesunden Menschenverstands völlig unabhängige zufällige Vorgänge (z.B. das Werfen eines Würfels und die Anzahl eingehender Anrufe an einem bestimmten Tag). Diese Modellieren wir jeweils als eigenen Wahrscheinlichkeitsraum. Wir haben es also aus mathematischer Sicht mit zwei Wahrscheinlichkeitsräumen $(\Omega_1, \mathcal{A}_1, P_1)$ und $(\Omega_2, \mathcal{A}_2, P_2)$ zu tun.

Wir überlegen uns nun, wie wir beide Vorgänge alternativ mit nur einem Wahrscheinlichkeitsraum modellieren können und welche spezielle Eigenschaft in diesem Fall für das zugehörige Wahrscheinlichkeitsmaß gelten muss. Diese Eigenschaft verwenden wir dann unten als Definition für die Unabhängigkeit von Zufallsgrößen.

Zwei durch $(\Omega_1, \mathcal{A}_1, P_1)$ und $(\Omega_2, \mathcal{A}_2, P_2)$ modellierte zufälligen Vorgänge können wir stets als einen einzigen zufälligen Vorgang betrachten, wenn wir als Ergebnismenge

$$\Omega := \Omega_1 \times \Omega_2 \tag{8.57}$$

wählen. Das Elementarereignis (ω_1, ω_2) bedeutet dann, dass beim ersten Teilvorgang das Elementarereignis ω_1 eingetreten ist und beim zweiten Teilvorgang das Elementarereignis ω_2.

Ein Ereignisfeld für den kombinierten Vorgang sollte mindestens alle Ereignisse der Form

$$A_1 \times A_2 \subseteq \Omega, \quad A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2 \tag{8.58}$$

enthalten. Diese Ereignisse entsprechen dem Eintreten von A_1 beim ersten zufälligen Vorgang und dem Eintreten von A_2 beim zweiten. Mit $\mathcal{A}$ bezeichnen wir das kleinste Ereignisfeld, welches all diese Ereignisse enthält.

Wie ist nun das Wahrscheinlichkeitsmaß P zu wählen, damit $(\Omega, \mathcal{A}, P)$ genau die gleichen Wahrscheinlichkeiten beschreibt wie die separaten Wahrscheinlichkeitsmaße? Insbesondere soll

$$P(A_1 \times \Omega_2) = P_1(A_1) \quad \text{und} \quad P(\Omega_1 \times A_2) = P_2(A_2) \tag{8.59}$$

gelten, sodass die Einzelwahrscheinlichkeiten P_1 und P_2 aus P zurückgewonnen werden können.

Da wir die frequentistische Sicht einnehmen, soll die Wahrscheinlichkeit eines Ereignisses gerade dessen relative Häufigkeit bei sehr vielen Versuchen widerspiegeln. Tritt das Ereignis A_1 bei n_1 Versuchen zum ersten zufälligen Vorgang m_1-mal ein und das Ereignis A_2 bei m_2 von n_2 Versuchen zum zweiten Vorgang, so liegen $n_1\,n_2$ mögliche Kombinationen von Versuchen vor. Von diesen Kombinationen entsprechen $m_1\,m_2$ Stück dem Eintreten von $A_1 \times A_2$, also A_1 beim ersten Vorgang und A_2 beim zweiten Vorgang. Die relative Häufigkeit bei kombinierter Betrachtung der beiden Vorgänge ergibt sich somit als Produkt der beiden relativen Häufigkeiten bei getrennter Betrachtung:

$$\frac{m_1\,m_2}{n_1\,n_2} = \frac{m_1}{n_1}\,\frac{m_2}{n_2}. \tag{8.60}$$

Entsprechend ist das kombinierte Wahrscheinlichkeitsmaß als

$$P(A_1 \times A_2) := P_1(A_1)\,P_2(A_2) \tag{8.61}$$

zu wählen. Man kann zeigen (tun wir hier nicht), dass dadurch P auf ganz $\mathcal{A}$ eindeutig gegeben ist und dass P tatsächlich ein Wahrscheinlichkeitsmaß ist. Man bezeichnet P auch als **Produktmaß** und $(\Omega, \mathcal{A}, P)$ als **Produkt von Wahrscheinlichkeitsräumen**.

8.5.2 Unabhängige Zufallsgrößen

Übertragen wir die Gedanken des vorhergehenden Abschnitts auf zwei Zufallsgrößen $\tilde{X}_1$ und $\tilde{X}_2$ mit zwei verschiedenen zugrundeliegenden Wahrscheinlichkeitsräumen $(\Omega_1, \mathcal{A}_1, P_1)$ und $(\Omega_2, \mathcal{A}_2, P_2)$, die zwei nach menschlichem Ermessen unabhängige zufällige Vorgänge beschreiben, so können wir $\tilde{X}_1$ und $\tilde{X}_2$ mittels

$$X_1(\omega_1, \omega_2) := \tilde{X}_1(\omega_1), \qquad X_2(\omega_1, \omega_2) := \tilde{X}_2(\omega_2) \tag{8.62}$$

als Zufallsgrößen auf dem Produktraum $(\Omega, \mathcal{A}, P)$ betrachten. Es gilt dann

$$P\big(X_1 \in (a,b),\, X_2 \in (c,d)\big) = P\big(X_1 \in (a,b)\big)\,P\big(X_2 \in (c,d)\big) \tag{8.63}$$

(und entsprechend für alle anderen Formen von endlichen und halbunendlichen Intervallen).

Als Definition für Unabhängigkeit verwenden wir diese Beziehung, jedoch ohne auf die zugrundeliegende Produktstruktur des Wahrscheinlichkeitsraumes zu bestehen. Diese war nur der gewählten Herleitung bzw. Plausibilisierung der Beziehung geschuldet.

Merke

Zwei Zufallsgrößen X_1 und X_2 heißen **unabhängig**, wenn

$$P(X_1 \in I_1,\, X_2 \in I_2) = P(X_1 \in I_1)\, P(X_2 \in I_2) \tag{8.64}$$

für alle Formen und Kombinationen von endlichen und halbunendlichen Intervallen I_1 und I_2 gilt.

Man kann sich überlegen, dass die definierende Gleichheit schon für Intervalle der Form $(-\infty, b)$ die Unabhängigkeit vollständig charakakterisiert, d.h. man alternativ die Bedingung

$$P(X_1 < x_1,\, X_2 < x_2) = P(X_1 < x_1)\, P(X_2 < x_2) \tag{8.65}$$

in der Definition verwenden.

Für mehr als zwei Zufallsgrößen erfolgt die Definition analog:

Merke

Die Zufallsgrößen $X_1, \ldots, X_n$ heißen **unabhängig**, wenn

$$P(X_1 \in I_1, \ldots, X_n \in I_n) = P(X_1 \in I_1) \cdots P(X_n \in I_n) \tag{8.66}$$

für alle Formen und Kombinationen von endlichen und halbunendlichen Intervallen $I_1, \ldots, I_n$ gilt.

Für unendlich viele Zufallsgrößen führt man Unabhängigkeit auf die Unabhängigkeit alle endlichen Teilmengen zurück:

Merke

Eine Folge $X_1, X_2, \ldots$ von Zufallsgrößen heißt **unabhängig**, wenn jede endliche Auswahl von Zufallsgrößen unabhängig ist.

8.5.3 Nachweis und Bedeutung der Unabhängigkeit

Der Nachweis der Unabhängigkeit zweier Zufallsgrößen ist für praktisch relevante Situationen kaum möglich, da viel zu aufwendig. Man müsste für alle möglichen Kombinationen von Intervallen die definierende Gleichung nachweisen. Nur die Verletzung der Unabhängigkeit lässt sich gut nachweisen. Hier ist nur eine konkrete Situation von Intervallen I_1 und I_2 anzugeben, für die die definierende Gleichheit verletzt ist. Praktisch wird man die Frage der Unabhängigkeit nie exakt beantworten können oder wollen, weshalb wir hier auch auf jegliche Beispiele verzichten. Der Wert des Begriffs ist ein anderer.

Der Unabhängigkeitsbegriff spielt vor allem als Annahme eine wichtige Rolle. Aus praktischen Erwägungen kann man oft schließen, dass zwei betrachtete zufällige Vorgänge sich

nicht (oder nicht in relevantem Maß) gegenseitig beeinflussen. Im mathematischen Modell formuliert man diese Einsicht als Annahme der Unabhängigkeit der entsprechenden Zufallsgrößen, obwohl man keinen harten Nachweis dazu geführt hat. Dieses Vorgehen ist aus streng mathematischer Sicht kritisch zu sehen. Dürfen wir mit Annahmen arbeiten, deren Erfüllung wir im betrachteten Fall nicht sicher nachweisen können? Andererseits sind viele wichtige statistische Werkzeuge (vgl. Kapitel zur Statistik) ohne die Annahme der Unabhängigkeit der beteiligten Zufallsgrößen überhaupt nicht einsetzbar. Wir werden auf diese schwierige Gradwanderung im Kapitel zur Statistik zurückkommen...

In gewissem Sinne ist die Unabhängigkeit von Zufallsgrößen für die Anwender also ein leerer Begriff: Ständig taucht er auf, die Definition ist eher technische Natur, also wenig anschaulich, und nie wird die Eigenschaft konkret nachgerechnet oder nachgewiesen. Lediglich die Interpretation in Alltagssprache spielt eine Rolle. Im ganzen für uns nicht sichtbaren Unterbau der mathematischen Statistik spielt der Unabhängigkeitsbegriff jedoch eine so zentrale Rolle, dass man ihn nicht vollständig vor den Anwendern verbergen kann bzw. verbergen möchte.

8.6 Grenzwertsätze

Bisher habe wir viel im "Klein-Klein" gearbeitet, also im Wesentlichen Begriffe definiert und einfache Zusammenhänge zwischen diesen Begriffe betrachtet. Der Wert einer Theorie zeigt sich jedoch erst in den großen Zusammenhängen: Kann die entwickelte Theorie Vorhersagen und Erkenntnisse liefern, die ohne diese Theorie verborgen geblieben wären?

Für die Wahrscheinlichkeitstheorie ist die Antwort ganz klar "Ja". Einige zentrale Resultate zeigen wir hier zur Demonstration der inhaltlichen und sprachlichen Möglichkeiten, jedoch ohne jeglichen Nachweis.

8.6.1 Schwache Konvergenz der relativen Häufigkeit

Wir hatten die Wahrscheinlichkeit eines Ereignisses als relative Häufigkeit seines Eintretens bei sehr vielen Versuchen beschrieben (frequentistische Interpretation der Wahrscheinlichkeitsrechnung). Diesen Gedanke können wir wie folgt präzsieren:

Merke

Bei fortwährender Wiederholung eines Zufallsexperiments beschreiben die Zufallsgrößen $X_1, X_2, \dots$ die relative Häufigkeit für das Eintreten eines gegebenen Ereignisses A nach der ersten, nach der zweiten usw. Wiederholung. X_n sei also die relative Häufigkeit des Eintretens von A nach n Wiederholungen. Ist p die Wahrscheinlichkeit für das Eintreten von A bei einer einzelnen Versuchsdurchführung, so gilt

$$\lim_{n\to\infty} P\big(|X_n - p| \geq \varepsilon\big) = 0 \tag{8.67}$$

für jedes $\varepsilon > 0$.

Beachte, dass eine Aussage der Form $\lim_{n\to\infty} X_n = p$ angenehmer aussieht, aber inhaltlich nicht sinnvoll ist, da X_n eine Funktion ist. Bei der obigen Konvergenzaussage spricht man auch von **Konvergenz in Wahrscheinlichkeit**. Dies ist eine Konvergenzform für Zufallsgrößen. Die Wahrscheinlichkeit für eine noch so kleine Mindestabweichung vom Grenzwert wird beliebig klein, wenn man n nur groß genug wählt.

Die frequentistische Sicht auf Wahrscheinlichkeiten wird durch die in dieser Veranstaltung gewählte (und auch in der modernen Literatur übliche) Modellierung von zufälligen Vorgängen also gut umgesetzt.

8.6.2 Schwaches Gesetz der großen Zahlen

Neben dem Erwartungswert einer Zufallsgröße hatten wir bereits den empirischen Mittelwert einer Anzahl von Realisierungen dieser Zufallsgröße betrachtet. Die folgende Aussage liefert den engen Zusammenhang beider Größen:

Merke

Beschreibe die Zufallsgröße X_n den empirischen Mittelwert von n Realisierungen einer Zufallsgröße X mit endlichem Erwartungswert $\mathbb{E}X$. Dann gilt

$$\lim_{n\to\infty} P\big(|X_n - \mathbb{E}X| \geq \varepsilon\big) = 0 \tag{8.68}$$

für jedes $\varepsilon > 0$.

Die Aussage, dass der der emprische Mittelwert für große Versuchsanzahl mit dem Erwartungswert übereinstimmt, ist intuitiv klar. Dass wir diesen Zusammenhang auch aus der entwickelten Wahrscheinlichkeitstheorie ableiten können, spricht für die Korrektheit bzw. Nützlichkeit der Theorie.

8.6.3 Zentraler Grenzverteilungssatz

Den zentralen Grenzverteilungssatz gibt es in unzähligen Varianten. Kernaussage ist, dass die Überlagerung vieler sich ähnlich verhaltender Zufallsgrößen zu einer normalverteilten Zufallsgröße führt. Praktisch bedeutet dies, dass die Zusammenspiel vieler kleiner Fehlerquellen stets zu einem normalverteilten Gesamtfehler führt. Hier ein einfache Formulierung mit recht strengen Voraussetzungen:

Merke

Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsgrößen mit identischer Verteilung, also insbesondere mit identischem Erwartungswert μ und identischer Standardabweichung σ. Dann konvergiert die Verteilungsfunktion F_n der standardisierten Zufallsgröße

$$\frac{X_1 + \cdots + X_n - n\,\mu}{\sqrt{n}\,\sigma} \tag{8.69}$$

punktweise gegen die Verteilungsfunktion Φ der Standardnormalverteilung, d.h.

$$\lim_{n\to\infty} F_n(x) = \Phi(x) \quad \text{für alle } x \in \mathbb{R}. \tag{8.70}$$

Dass $\sqrt{n}\,\sigma$ die Standardabweichung der Zufallsgröße $X_1 + \cdots + X_n$ ist, folgt aus der Unabhängigkeit der Zufallsgrößen. Wir gehen an dieser Stelle nicht näher darauf ein, kommen aber im nächsten Unterkapitel darauf zurück.

Der zentrale Grenzverteilungssatz ist ein Beispiel für eine praktisch relevante Erkenntnis, die ohne die mühsam entwickelte Wahrscheinlichkeitstheorie im Verborgenen geblieben wäre. Bestenfalls könnte man den Zusammenhang von Messfehlern mit der Normalverteilung durch umfangreiche statistische Auswertungen erahnen (jedoch nicht begründen!).

8.7 Zufallsvektoren

Statt zahlenwertiger Funktionen auf einem Wahrscheinlichkeitsraum kann man natürlich auch vektorwertige Funktionen betrachten. Dies bringt einige zusätzliche Aspekte bei der Darstellung von Wahrscheinlichkeitsmaßen mittels Verteilungs- und Dichtefunktionen, ändert aber nichts am grundsätzlichen Vorgehen.

8.7.1 Begriff

Zufallsvektoren kann man auf zwei formal unterschiedliche, im Ergebnis aber völlig äquivalente Weisen einführen.

Merke

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Funktion $X : \Omega \to \mathbb{R}^n$ heißt **Zufallsvektor**, wenn sie messbar ist, d.h. wenn für beliebige n-Zellen der Form $(-\infty, b_1) \times \cdots \times (-\infty, b_n)$ das Urbild bzgl. X zu $\mathcal{A}$ gehört.

Die folgende Aussage liefert eine äquivalente Definition. Auf den Nachweis der Äquivalenz verzichten wir (ist elementar, aber länglich).

Merke

Eine Funktion $X : \Omega \to \mathbb{R}^n$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ ist genau dann ein Zufallsvektor, wenn jede Komponente $X_k : \Omega \to \mathbb{R}$, $k = 1, \ldots, n$, eine Zufallsgröße ist.

Die Unterteilung in diskrete und kontinuierliche Zufallsvektoren erfolgt völlig analog zu Zufallsgrößen. Sind alle Komponenten diskret, so spricht man von einem **diskreten Zufallsvektor**. Sind alle Komponenten kontinuierlich, so spricht man von einem **kontinuierlichen Zufallsvektor**.

8.7.2 Verteilungs- und Dichtefunktionen

Verteilungs- und Dichtefunktion hängen von mehreren Variablen ab.

Merke

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und sei $X : \Omega \to \mathbb{R}^n$ ein Zufallsvektor. Die Funktion

$$F : \mathbb{R}^n \to [0, 1], \quad F(x_1, \ldots, x_n) := P(X_1 < x_1, \ldots, X_n < x_n) \tag{8.71}$$

heißt **Verteilungsfunktion** zum Zufallsvektor X.

Für kontinuierliche Zufallsvektoren kann man analog zu Zufallsgrößen eine Dichtefunktion einführen. Allerdings erhält man nun mehrere ineinander geschachtelte Integrale.

Merke

Sei X ein kontinuierlicher Zufallsvektor mit Wertebereich $W(X) = [a_1, b_1] \times \cdots \times [a_n, b_n]$ und Verteilungsfunktion F. Eine Funktion $p : \mathbb{R}^n \to [0, \infty)$ heißt **Dichtefunktion** zu X, wenn p auf $W(X)$ und auf $\mathbb{R}^n \setminus W(X)$ stetig ist und

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} p(\tilde{x}_1, \ldots, \tilde{x}_n) \, \mathrm{d}\tilde{x}_1 \cdots \mathrm{d}\tilde{x}_n \tag{8.72}$$

gilt.

Aus Verteilungs- und Dichtefunktion eines Zufallsvektors kann man die Verteilungs- und Dichtefunktionen der einzelnen Komponenten erhalten. Umgekehrt ist dies nicht möglich, da durch die Verteilungsfunktion des Zufallsvektors nicht nur das Verhalten der einzelnen Komponenten beschrieben wird, sondern auch deren Abhängigkeiten untereinander.

Merke

Sei F die Verteilungsfunktion eines Zufallsvektors X und sei F_k die Verteilungsfunktion der Komponente X_k. Dann gilt

$$F_k(x_k) = \lim_{x_1 \to \infty} \cdots \lim_{x_{k-1} \to \infty} \lim_{x_{k+1} \to \infty} \cdots \lim_{x_n \to \infty} F(x_1, \ldots, x_n). \tag{8.73}$$

Merke

Sei p die Dichtefunktion eines Zufallsvektors X und sei p_k die Dichtefunktion der Komponente X_k. Dann gilt

$$p_k(x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, \ldots, x_n) \, \mathrm{d}x_1 \cdots \mathrm{d}x_{k-1} \, \mathrm{d}x_{k+1} \cdots \mathrm{d}x_n. \tag{8.74}$$

Die Verteilungs- und Dichtefunktionen der einzelnen Komponenten werden auch als **Randverteilungen** bzw. **Randdichten** des Zufallsvektors bezeichnet.

Beispiel (gleiche Randverteilungen)

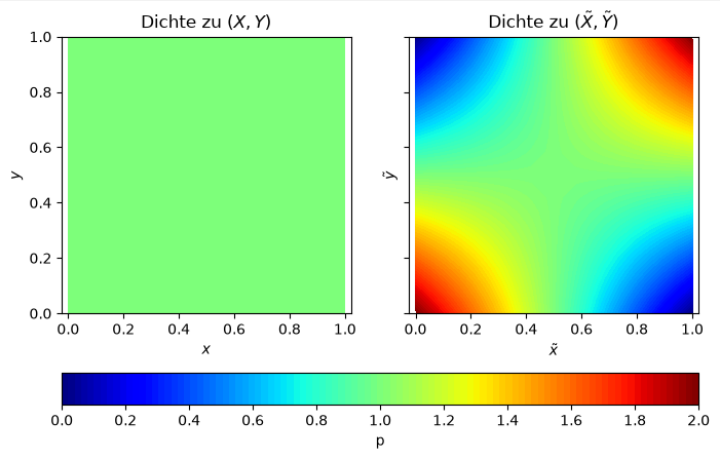
Betrachten zwei Zufallsvektoren mit unterschiedlichen Wahrscheinlichkeitsverteilungen:

- Der Zufallsvektor (X, Y) sei gleichverteilt auf dem Quadrat $[0,1] \times [0,1]$, d.h. seine Dichtefunktion sei

$$p(x,y) := \begin{cases} 1, & \text{für } (x,y) \in [0,1] \times [0,1], \\ 0, & \text{sonst.} \end{cases} \tag{8.75}$$

- Der Zufallsvektor $(\tilde{X}, \tilde{Y})$ besitze die Dichtefunktion

$$\tilde{p}(x,y) := \begin{cases} 4\,x\,y - 2\,(x+y) + 2, & \text{für } (x,y) \in [0,1] \times [0,1], \\ 0, & \text{sonst.} \end{cases} \tag{8.76}$$



Obwohl die beiden Zufallsvektoren (X, Y) und $(\tilde{X}, \tilde{Y})$ unterschiedliche Verteilungen besitzen, stimmen die Randverteilungen jeweils überein:

$$p_1(x) = \tilde{p}_1(x) = \begin{cases} 1, & \text{für } x \in [0,1], \\ 0, & \text{sonst} \end{cases} \tag{8.77}$$

und

$$p_2(y) = \tilde{p}_2(y) = \begin{cases} 1, & \text{für } y \in [0,1], \\ 0, & \text{sonst} \end{cases} \tag{8.78}$$

(vgl. Aufgabe 13.1 in Übung 13).

8.7.3 Erwartungswert, Varianz, Kovarianz

Erwartungswert und Varianz eines Zufallsvektors werden komponentenweise betrachtet, liefern also jeweils wieder einen Vektor.

Merke

Für einen kontinuierlichen Zufallsvektor X heißt der Vektor

$$\mathbb{E}X := \begin{bmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_n \end{bmatrix} \tag{8.79}$$

Erwartungswert des Zufallsvektors.

Merke

Sei X ein Zufallsvektor mit Erwartungswert $\mathbb{E}X$. Der Vektor

$$\mathbb{V}\mathrm{ar}\, X := \begin{bmatrix} \mathbb{E}\big((X_1 - \mathbb{E}X_1)^2\big) \\ \vdots \\ \mathbb{E}\big((X_n - \mathbb{E}X_n)^2\big) \end{bmatrix} \tag{8.80}$$

heißt **Varianz** (oder Streuung) von X.

Als zusätzliche Kenngröße kommt bei Zufallsvektoren die Kovarianz hinzu. Diese liefert Informationen über das Zusammenspiel der einzelnen Komponenten. Betrachten wir zwei Komponenten, so können folgende Situationen auftreten:

- Die Realisierungen weichen oft gemeinsam nach oben oder gemeinsam nach unten vom Erwartungswert ab ("Gleichläufigkeit").
- Weichen die Realisierungen der einen Komponente nach oben vom Erwartungswert ab, weichen die der anderen oft nach unten ab ("Gegenläufigkeit").
- Die Realisierungen beider Komponenten zeigen kein Muster bei der Richtung der Abweichung vom Erwartungswert.

Art und Stärke der Gleich- oder Gegenläufigkeit können als Zahl ausgedrückt werden. Dazu multipliziert man die vorzeichenbehafteten (!) Abweichungen vom Mittelwert:

$$(X_k - \mathbb{E}X_k)\,(X_l - \mathbb{E}X_l). \tag{8.81}$$

Gleichläufigkeit bei einer konkreten Realisierung liefert einen positiven Wert, Gegenläufigkeit einen negativen. Allgemein, also ohne Rückgriff auf konkrete Realisierungen, kann man den Erwartungswert dieses Produkts als Maß für die Gleich- bzw. Gegenläufigkeit der Komponenten heranziehen.

Merke

Sei X ein Zufallsvektor. Die Zahl

$$\mathbb{C}\mathrm{ov}\,(X_k, X_l) := \mathbb{E}\big((X_k - \mathbb{E}X_k)\,(X_l - \mathbb{E}X_l)\big) \tag{8.82}$$

heißt **Kovarianz** der Komponenten X_k und X_l.

Die Matrix

$$\mathbb{C}\mathrm{ov}\,X := \begin{bmatrix} \mathbb{C}\mathrm{ov}\,(X_1, X_1) & \cdots & \mathbb{C}\mathrm{ov}\,(X_1, X_m) \\ \vdots & \ddots & \vdots \\ \mathbb{C}\mathrm{ov}\,(X_n, X_1) & \cdots & \mathbb{C}\mathrm{ov}\,(X_n, X_n) \end{bmatrix} \tag{8.83}$$

heißt **Kovarianzmatrix** des Zufallsvektors.

Offensichtlich gilt

$$\mathbb{C}\mathrm{ov}\,(X_k, X_k) = \mathbb{V}\mathrm{ar}\,X_k. \tag{8.84}$$

Auch sieht man sofort, dass die Kovarianzmatrix eines Zufallsvektors stets symmetrisch ist.

Man führt üblicherweise noch folgenden Begriff zur Vereinfachung der Sprache ein:

Merke

Seien X_1 und X_2 zwei Zufallsgrößen. Die Zahl

$$\varrho(X_1, X_2) := \frac{\mathbb{C}\mathrm{ov}\,(X_1, X_2)}{\sqrt{\mathbb{V}\mathrm{ar}\,X_1}\sqrt{\mathbb{V}\mathrm{ar}\,X_2}} \tag{8.85}$$

heißt **Korrelationskoeffizient** der beiden Zufallsgrößen.

Gilt $\varrho(X_1, X_2) = 0$, so heißen die beiden Zufallsgrößen **unkorreliert**.

Der Korrelationskoeffizient liefert genau die gleich Information über Gleich- bzw. Gegenläufigkeit wie die Kovarianz; allerdings kann man zeigen, dass der Korrelationskoeffizient immer zwischen -1 und 1 liegt, also in diesem Sinne normiert ist. Außerdem ist er einheitenlos. Dies erleichtert den Vergleich der Gleich- bzw. Gegenläufigkeit verschiedener Paare von Zufallsgrößen. Bei Werten nahe bei 1 bzw. -1 spricht man von starker (gleichläufiger bzw. gegenläufiger) Korrelation.

Beispiel (Holzzuschnitt)

In einem Sägewerk werden Latten auf 4 Meter Länge gesägt. Die tatsächliche Länge weicht aufgrund diverser Ungenauigkeiten etwas von 4 Meter ab. Ein Mitarbeiter überprüft regelmäßig die Länge der Latten. Sein Messergebnis sei durch die Zufallsgröße X_1 beschrieben. Ein anderer Mitarbeiter überprüft mit einem anderen Maßband ebenfalls die Lattenlängen. Das Messergebnis des zweiten Mit-

arbeiters sei als Zufallsgröße X_2 modelliert.

Aufgrund der Verwendung unterschiedlicher Maßbänder und aufgrund eventuell etwas unterschiedlichem Vorgehen beim Messen ist nicht damit zu rechnen, dass beide Zufallsgrößen identische Verteilungen besitzen. Allerdings werden die beiden Verteilungen stark korrelieren. Misst der eine Mitarbeiter mehr als 4 Meter, so wird der andere dies in den meisten Fällen ebenso tun.

Man kann zeigen, dass **unabhängige Zufallsgrößen stets unkorreliert** sind. Die Umkehrung gilt im Allgemeinen nicht. Hintergrund ist, dass die Kovarianz Gleich- bzw. Gegenläufigkeit nur sehr grob misst. In gewissem, hier nicht weiter vertieftem Sinne drückt die Kovarianz nur den linearen Anteil von Gleich- bzw. Gegenläufigkeit aus. Unkorreliertheit sichert daher nicht, dass die beiden Zufallsgrößen wirklich völlig unabhängig voneinander arbeiten. Die Situation ist vergleichbar mit Taylor-Polynomen bei Funktionen. Schaut man nur auf das Taylor-Polynom ersten Grades an einer Entwicklungsstelle, so erhält man nur sehr begrenzte Informationen über den Funktionsverlauf an dieser Stelle.

8.7.4 Mehrdimensionale Normalverteilung

Der wichtigste Typ von Zufallsvektoren sind normalverteilte Zufallsvektoren. Bei diesen ist jede Komponente normalverteilt; jedoch müssen die einzelnen Komponenten nicht unabhängig sein.

Die folgende Definition verwendet den Begriff der positiv definiten Matrix, den wir bisher nicht eingeführt haben. Dessen Bedeutung ist recht tiefliegend, für uns aber nicht weiter relevant. Einzige für uns relevante Tatsache ist, dass positiv definite Matrizen stets invertierbar sind. Die formale Definition lautet: Eine symmetrische Matrix $A \in \mathbb{R}^{n\times n}$ heißt **positiv definit**, wenn $x^{\mathrm{T}}\, A\, x > 0$ für alle $x \in \mathbb{R}^n \setminus \{0\}$ gilt.

Merke

Ein Zufallsvektor X heißt normalverteilt mit dem Vektor $\mu \in \mathbb{R}^n$ und der symmetrischen, positiv definiten Matrix $S \in \mathbb{R}^{n\times n}$ als Parameter, wenn die Dichtefunktion die Gestalt

$$p(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\,\pi)^n\ \det S}}\, \mathrm{e}^{-\frac{1}{2}\,(x-\mu)^{\mathrm{T}}\, S^{-1}\,(x-\mu)} \tag{8.86}$$

besitzt.

Beachte, dass hier x und μ Vektoren sind, das Produkt $(x-\mu)^{\mathrm{T}}\, S^{-1}\,(x-\mu)$ aber eine Zahl liefert.

Aufgrund der Forderung nach positiver Definitheit der Matrix S führt nicht jede sym-

metrische Matrix zu einer Normalverteilung.

Wie bei der eindimensionalen Normalverteilung entsprechen die Parameter der mehrdimensionalen Normalverteilung gerade Erwartungswert und (Ko-)Varianz (ohne Nachweis):

Merke

Erwartungswert und Kovarianzmatrix eines normalverteilten Zufallsvektors X mit Parametern μ und S sind

$$\mathbb{E}X = \mu \tag{8.87}$$

und

$$\mathbb{C}\mathrm{ov}\, X = S. \tag{8.88}$$

Normalverteilte Zufallsgrößen und -vektoren zeigen einige wichtige Besonderheiten. Eine davon ist die Tatsache, dass Unabhängigkeit und Unkorreliertheit äquivalent sind:

Merke

Sind X_1 und X_2 zwei normalverteilte Zufallsgrößen, so sind diese genau dann unabhängig, wenn $\mathbb{C}\mathrm{ov}\,(X_1, X_2) = 0$ gilt.

Allgemeiner: Sind $X_1, \ldots, X_n$ normalverteilte Zufallsgrößen, so sind diese genau dann unabhängig, wenn die zugehörige Kovarianzmatrix eine Diagonalmatrix ist.

9 Statistik

Einführung in die mathematische Statistik mit Schwerpunkt auf der induktiven Statistik.

9.1 Überblick

9.1.1 Abgrenzung

Die Statistik als Teilgebiet der Mathematik befasst sich mit der systematischen Untersuchung von und der Arbeit mit gesammelten Daten. Im Kontext sehr umfangreicher Datensammlungen und beim Einsatz moderner computergestützter Verfahren (z.B. maschinelles Lernen) spricht man heute auch von **Data Science**.

Die klassische Statistik wird unterschieden in deskriptive und induktive Statistik:

- Die **deskriptive** Statistik (auch: beschreibende Statistik) extrahiert beschreibende und zusammenfassende Informationen aus Datensammlungen. Dies kann im einfachsten Fall der Mittelwert einer Größe über alle gesammelten Datensätze sein oder man stellt beispielsweise Verteilungsinformationen zu einer Größe zusammen.
- Die **induktive** Statistik (auch: schließende Statistik) beschäftigt sich mit der Frage, wie man aus eher kleinen Datensammlungen Rückschlüsse auf Eigenschaften einer viel größeren, nicht direkt untersuchten Gesamtheit ziehen kann. Insbesondere wird auch die Zuverlässigkeit solcher Schlüsse bewertet. Beispielsweise möchte man gern aus der unter Umständen zerstörenden Untersuchung von wenigen in einer Fabrik produzierten Teilen Rückschlüsse auf die Fehlerquote in der Gesamtproduktion ziehen.

9.1.2 Typische Fragestellungen

Die Methoden der induktiven Statistik liefern vor allem Anworten auf folgende Fragestellungen:

- **Punktschätzung:** Welche Werte haben die Parameter der den gesammelten Daten zugrundeliegenden Verteilung wahrscheinlich?
- **Konfidenzintervalle:** In welchen Bereichen liegen die Parameter der den gesammelten Daten zugrundeliegenden Verteilung mit hoher Wahrscheinlichkeit?
- **Hypothesentests:** Wie sicher können wir sein, dass eine gewisse Wahrscheinlichkeitsverteilung den gesammelten Daten zugrundeliegt? Wie sicher können wir sein, dass die Verteilungsparameter gewisse Bedingungen erfüllen (z.B. unter/obere Schranken)?

Beispiel

Eine Abfüllanlage für Getränke liefert leicht schwankende Füllmengen. Um die Schwankungen genauer zu untersuchen, werden einige befüllte Flaschen geöffnet und das enthaltene Getränkevolumen bestimmt. Folgende Fragen und Sachverhalte können nun mit den Mitteln der induktiven Statistik formuliert und untersucht

werden:

- Als Summe vieler kleiner Abweichungen (Maßtoleranzen von Maschinenteilen und Flaschen, Schwankungen in der Stromversorgung usw.) wird die übliche Schwankung in der Füllmenge einer Normalverteilung folgen (vgl. Section 8.6.3). Folgt die Füllmenge nicht einer Normalverteilung, so kann dies ein Hinweis auf ein technisches Problem in der Abfüllanlage sein. Ob die Füllmenge einer Normalverteilung folgt, kann mit einem Hypothesentest auf Normalverteiltheit bewertet werden.
- Mittels einer Punktschätzung kann die mittlere Abfüllmenge ermittelt werden. Wie zuverlässig diese Schätzung ist, muss gesondert betrachtet werden. Wurden nur sehr wenige Flaschen aus der Produktion entnommen und geöffnet, so ist der Schluss auf die Gesamtproduktion eher unzuverlässig. Wurde hingegen fast die gesamte Produktion geöffnet und gemessen, so ist die Schätzung mit hoher Wahrscheinlichkeit sehr nah am tatsächlichen Mittelwert.
- Wie viele Flaschen müsste man öffnen, um hinreichend zuverlässige Informationen zur Gesamtproduktion zu bekommen? Und: Was heißt eigentlich "hinreichend zuverlässig"?
- Ist die Abweichung nach unten von der Sollfüllmenge nicht zu groß? Werden die Vorgaben der Richtlinie 76/211/EWG (Fertigpackungsrichtlinie) eingehalten.

9.1.3 Grundbegriffe

Die zu untersuchende Menge an Objekten wird als **Grundgesamtheit** bezeichnet. Damit kann manchmal eine konkrete (meist sehr umfangreiche) Menge gemeint sein, manchmal aber auch eine nur unscharf abgegrenzte Menge von zu untersuchenden Objekten. Zum Beispiel kann man die gesamte Menschheit als Grundgesamtheit bei Untersuchungen zur globalen Vermögensverteilung ansehen. Hingegen ist beispielsweise die Grundgesamtheit bei der Qualitätskontrolle von produzierten Mikrochips weniger scharf abgegrenzt. Sollen sich die Untersuchungen auch auf zukünftig noch zu produzierende Chips erstrecken? Wie weit in die Vergangenheit behalten die erzielten Erkenntnisse Gültigkeit?

Aus mathematischer Sicht ist die Grundgesamtheit eine Zufallsgröße X. Das "Untersuchen der Grundgesamtheit" entspricht hier der Frage nach der Wahrscheinlichkeitsverteilung der Zufallsgröße X. Dabei muss nicht immer die Verteilung im Detail gesucht sein, sondern oft ist man nur an einigen Parametern der Verteilung oder dem Verteilungstyp interessiert.

Grundlage für das Untersuchen der Grundgesamtheit bildet eine **Stichprobe**. Eine (konkrete) Stichprobe ist eine endliche Menge von Realisierungen der Zufallsgröße X. Die Anzahl der Realisierungen heißt **Stichprobenumfang**. Praktisch wird man eine Stich-

probe durch Messungen oder Umfragen bekommen. Allgemein gilt: Je größer der Stichprobenumfang, desto mehr und desto zuverlässigere Informationen wird man aus der Stichprobe über die Grundgesamtheit erhalten können.

Eine (konkrete) Stichprobe vom Umfang n kann man auch als Realisierung eines Zufallsvektors $(X_1, \ldots, X_n)$ ansehen, wobei alle Komponenten X_k unabhängig voneinander und identisch wie X verteilt sind. Man sagt auch: die X_k sind **Zufallsgrößen vom Typ i.i.d.** (independently and identically distributed). Den Zufallsvektor $(X_1, \ldots, X_n)$ bezeichnet man ebenfalls als (allgemeine) Stichprobe.

Dass die Komponenten einer Stichprobe unabhängig sind, ist mehr eine Annahme als gesichertes Wissen. Beim Prüfen der Füllmenge von maschinell befüllten Flaschen könnte eine Minderfüllung beispielsweise systematisch zu einer Überfüllung einer späteren Flasche führen. Kann man aus Kenntnis des Sachverhalts mit hinreichender Zuverlässigkeit sagen, dass solche gegenseitigen Beeinflussungen der Stichprobenkomponenten nicht oder nur in sehr geringem Maß stattfinden, so ist die Annahme der Unabhängigkeit der Stichprobenkomponenten gerechtfertigt. Ohne diese Annahme sind die meisten Werkzeuge der induktiven Statistik nicht einsetzbar.

9.2 Deskriptive Statistik

Die deskriptive Statistik ist ein weites Feld, benötigt aber fast ausschließlich triviale Mathematik (bzw. nur "Rechnen"). Die meisten Verfahren sind intuitiv klar und lassen sich bei Bedarf leicht aus der Literatur erarbeiten.

9.2.1 Kenngrößen

Zur Beschreibung einer konkreten Stichprobe $(x_1, \ldots, x_n)$ zu einer durch die Zufallsgröße X beschriebenen Grundgesamtheit können verschiedene Kenngrößen herangezogen werden.

Der **empirische Erwartungswert** (oder Mittelwert) ist das arithmetische Mittel der Stichprobenwerte:

$$\overline{x} := \frac{1}{n} \sum_{k=1}^{n} x_k. \tag{9.1}$$

Die **empirische Varianz** ist

$$\frac{1}{n^2} \sum_{k=1}^{n} (x_k - \overline{x})^2. \tag{9.2}$$

Der **Median** der Stichprobe ist das Element, welches nach Sortieren der Elemente in der Mitte steht. Bei geradem Stichprobenumfang ist der Median das arithmetische Mittel der beiden mittleren Elemente.

Darüber hinaus gibt es noch eine Reihe weiterer Kenngrößen zur Beschreibung von (großen) Stichproben durch einzelne Zahlen, auf die wir hier nicht weiter eingehen. Siehe Wikipedia-Artikel Parameter (Statistik) für einen Überblick.

9.2.2 Histogramme

Ein Histogramm ist eine (meist grafische) Zusammenfassung der Werte einer Stichprobe. Die einzelnen Werte werden entsprechend ihrer Größe in Klassen eingeteilt. Die Anzahl der Elemente pro Klasse gibt Aufschluss über die Werteverteilung ohne jeden einzelnen Wert direkt betrachten zu müssen.

Beispiel

Zur Demonstration erzeugen wir uns eine Stichprobe in Form von $n = 1000$ Zufallszahlen (statt Umfrageergebnisse oder Messwerte). Diese sortieren wir in 10 gleich breite Klassen ein und stellen die Klassengrößen als Balkendiagramm dar.

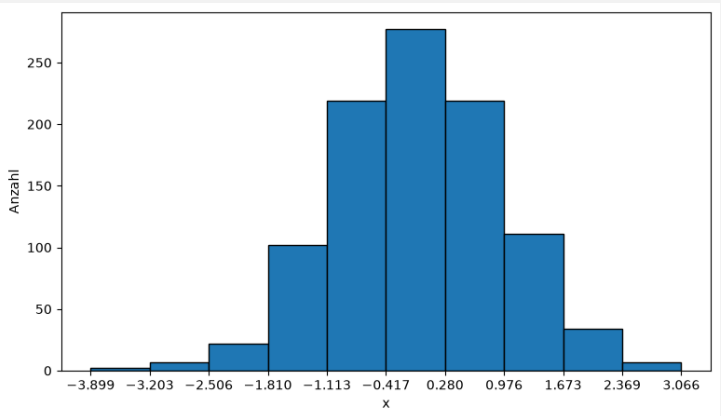
```
import numpy as np
import matplotlib.pyplot as plt
```

```python
rng = np.random.default_rng(0)  # Zufallszahlengenerator

# Stichprobe erzeugen
n = 1000
x = rng.normal(0, 1, n)

# Histogrammdaten erzeugen (Klassengrenzen und -größen)
counts, edges = np.histogram(x, bins=10)

# Plot
fig, ax = plt.subplots(figsize=(9, 5))
ax.bar(
    edges[: -1], counts,  # Daten
    align='edge', width=edges[1] -edges[0],  # Balkenposition und
    ↪  -breite
    edgecolor='black'  # Balkenränder zeigen
)
ax.set_xticks(edges)  # Label an Klassengrenzen
ax.set_xlabel('x')
ax.set_ylabel('Anzahl')
plt.show()
```



Sind die Daten zweidimensional, d.h. die Grundgesamtheit wird durch einen Zufallsvektor (X, Y) repräsentiert und jedes Stichprobenelement ist ein Punkt in der Ebene, so kann man auch zweidimensionale Histogramme betrachten. Diese Form der Visualisierung, die insbesondere bei Geodaten gern verwendet wird, ist als Heat-Map in der Literatur zu finden.

9.3 Punktschätzung

9.3.1 Ziel

Sei X eine Zufallsgröße, deren Verteilung einen unbekannten Parameter ϑ enthält. Zu dieser Zufallsgröße liegt eine gewisse Anzahl Realisierungen vor (Stichprobe), aus denen der unbekannte Parameter bestimmt werden soll.

Beispiel

Die mittlere Füllmenge beim Befüllen von Flaschen durch eine Abfüllanlage soll überprüft werden. Die Füllmenge schwankt etwas, sodass das Prüfen einer einzelnen Flasche nur unzureichend Information über die Füllmengenverteilung in der Gesamtproduktion liefert. Zur (näherungsweisen) Bestimmung des Erwartungswerts der Füllmengenverteilung müssen deshalb mehrere Flaschen entnommen, geöffnet und der Inhalt gemessen werden.

Der zu schätzende Wert ϑ muss nicht zwingend ein direkter Parameter der Verteilung sein, sondern kann die Verteilung auch indirekt beschreiben. Beispielsweise kann ϑ der Erwartungswert oder die Varianz der Verteilung sein, auch wenn diese nicht direkt in den Formeln für Verteilungs- oder Dichtefunktion vorkommen.

9.3.2 Schätzfunktion

Merke

Eine **Schätzfunktion** T_n ist eine Zufallsgröße der Form

$$T_n := g(X_1, \ldots, X_n) \tag{9.3}$$

mit einer gegebenen Funktion $g : \mathbb{R}^n \to \mathbb{R}$. Eine Schätzfunktion fasst eine Stichprobe zu einer Zufallsgröße zusammen, deren Realisierungen nahe beim gesuchten Verteilungsparameter ϑ liegen sollen.

Die Funktion g heißt **Punktschätzer** für ϑ. Der Wert $g(x_1, \ldots, x_n)$ für eine konkrete Stichprobe $(x_1, \ldots, x_n)$ heißt **Schätzwert**.

Beispiel

Mit

$$g(x_1, \ldots, x_n) := \frac{1}{n} \sum_{k=1}^{n} x_k \tag{9.4}$$

erhalten wir die Schätzfunktion

$$T_n := \frac{1}{n} \sum_{k=1}^{n} X_k \tag{9.5}$$

für den Erwartungswert der Verteilung der Grundgesamtheit.

9.3.3 Eigenschaften von Schätzfunktionen

Wie sehen wir einer Schätzfunktion an, ob sie für das Schätzen des gesuchten Parameters ϑ überhaupt geeignet ist? Sind zwei Schätzfunktionen grundsätzlich geeignet, muss auch die Frage geklärt werden, ob die eine bessere Schätzungen (in welchem Sinne?) als die andere liefert.

Beispiel

Beispielsweise könnte auch

$$g(x_1, \ldots, x_n) := \frac{1}{n} \left(\frac{1}{2} x_1 + \left(1 + \frac{1}{2\,(n-1)} \right) \sum_{k=2}^{n} x_k \right) \tag{9.6}$$

eine geeignete Schätzfunktion für den Erwartungswert liefern. Die geringere Gewichtung des ersten Stichprobenelements könnte die eventuell gegebene Tatsache widerspiegeln, dass die erste Messung einer Messreihe stets etwas ungenauer ist als die weiteren Messungen.

Zur Beantwortung der Frage nach der prinzipiellen Brauchbarkeit einer Schätzfunktion führen wir zwei Begriffe ein.

Merke

Eine Schätzfunktion T_n für den Parameter ϑ heißt **erwartungstreu**, wenn

$$\mathbb{E}T_n = \vartheta \tag{9.7}$$

gilt.

Merke

Eine Schätzfunktion T_n für den Parameter ϑ heißt **konsistent**, wenn für jedes

$\varepsilon > 0$

$$\lim_{n\to\infty} P(|T_n(X_1,\ldots,X_n) - \vartheta| > \varepsilon) = 0 \tag{9.8}$$

gilt.

Erwartungstreue sichert, dass die Schätzfunktion im Mittel über viele Stichproben den tatsächlichen Wert des Parameters liefert. Konsistenz sichert, dass die Schätzung durch Wahl einer hinreichend großen Stichprobe beliebig genau gemacht werden kann.

Für die konkrete Umsetzung der Konsistenzidee in Formeln gibt es verschiedene Varianten. Bei der hier gewählten spricht man von *Konvergenz (der Schätzfunktion gegen den gesuchten Parameter) in Wahrscheinlichkeit*. Eine von mehreren alternativen Formulierungen ist die Konvergenz im quadratischen Mittel:

$$\lim_{n\to\infty} \mathbb{E}\big((T_n - \vartheta)^2\big) = 0 \tag{9.9}$$

Ist T_n erwartungstreu, so ist die Konvergenz im quadratischen Mittel offensichtlich identisch zur Forderung

$$\lim_{n\to\infty} \mathbb{V}\mathrm{ar}\, T_n = 0. \tag{9.10}$$

Die unterschiedlichen Varianten von Konsistenz können in einigen Fällen zu unterschiedlichen Ergebnissen führen, ob eine Schätzfunktion konsistent ist oder nicht.

Beispiel

Für gegebene Gewichte $a_1,\ldots,a_n \in \mathbb{R}$ mit $a_1 + \cdots + a_n = 1$ liefert der Punktschätzer

$$g(x_1,\ldots,x_n) := \sum_{k=1}^{n} a_k\, x_k \tag{9.11}$$

stets eine erwartungstreue Schätzfunktion für den Erwartungswert. Gilt zusätzlich

$$\lim_{n\to\infty} \sum_{k=1}^{n} a_k^2 = 0, \tag{9.12}$$

so ist die Schätzfunktion auch konsistent (Video 920, Notizen dazu).

Insbesondere liefert $a_k = \frac{1}{n}$, $k = 1,\ldots,n$ eine erwartungstreue und konsistente Schätzfunktion, aber beispielsweise auch

$$a_1 = \frac{1}{2\,n}, \qquad a_k = \frac{1}{n} + \frac{1}{2\,n\,(n-1)}, \quad k = 2,\ldots,n. \tag{9.13}$$

Für den Erwartungswert (und auch für andere Parameter) existiert also eine Vielzahl an brauchbaren Schätzfunktionen.

Merke

Hat eine Schätzfunktion T_n in einer Familie von erwartungstreuen Schätzfunktionen für einen Parameter ϑ die kleinste Varianz, so heißt T_n **wirksamste Schätzfunktion** in dieser Familie.

Schätzfunktionen mit kleiner Varianz liefern bei verschiedenen Stichproben ähnliche Schätzwerte für den gesuchten Parameter. Bei großer Varianz der Schätzfunktion können aus verschiedenen Stichproben berechnete Schätzwerte stark voneinander abweichen.

Beispiel

Unter allen Schätzfunktionen der Form

$$T_n := \sum_{k=1}^{n} a_k X_k \tag{9.14}$$

mit $a_1 + \cdots + a_n = 1$ für den Erwartungswert ist die durch $a_k = \frac{1}{n}$, $k = 1, \ldots, n$ gegebene die wirksamste (Video 925, Notizen dazu).

Achtung!

Für eine konkrete Stichprobe können wir keine Aussage darüber treffen, wie nah der aus der Stichprobe berechnete Schätzwert am tatsächlichen Wert des Parameters liegt. Erwartungstreue und Konsistenz sichern lediglich, dass wir im Mittel (über viele Schätzungen) das richtige tun.

9.3.4 Schätzfunktionen für die Varianz

Kennt man den Erwartungswert $\mathbb{E}X$ der Verteilung der Grundgesamtheit, so kann man zeigen, dass

$$T_n := \frac{1}{n} \sum_{k=1}^{n} (X_k - \mathbb{E}X)^2 \tag{9.15}$$

eine erwartungstreue und konsistente Schätzfunktion für die Varianz $\mathbb{V}\mathrm{ar}\, X$ ist.

Praktisch steht der Erwartungswert der zugrundliegenden Verteilung jedoch selten zur Verfügung. Die naheliegende Schätzfunktion

$$T_n := \frac{1}{n} \sum_{k=1}^{n} (X_k - \overline{X})^2 \tag{9.16}$$

für diesen Fall, wobei

$$\overline{X} := \frac{1}{n} \sum_{k=1}^{n} X_k, \tag{9.17}$$

ist jedoch nicht (!) erwartungstreu. Sie erfüllt lediglich

$$\lim_{n \to \infty} \mathbb{E}T_n = \mathbb{V}\mathrm{ar}\, X, \tag{9.18}$$

ist also nur **asymptotisch erwartungstreu**.

Man kann zeigen, dass die geringfügig angepasste Schätzfunktion

$$T_n := \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X})^2 \tag{9.19}$$

eine erwartungstreue und konsistente Schätzfunktion für die Varianz ist.

9.3.5 Weitere Schätzfunktionen

Für weitere Schätzfunktionen sei auf die Literatur verwiesen. Dort findet man auch Verfahren zur systematischen Konstruktion von Schätzfunktionen, z.B. die Maximum-Likelihood-Methode.

9.4 Bereichsschätzung

9.4.1 Idee

Eine Alternative zu Punktschätzungen sind Bereichsschätzungen. Hier wird aus einer konkreten Stichprobe nicht ein Schätzwert generiert, der (hoffentlich) in der Nähe des gesuchten Verteilungsparameters ϑ liegt, sondern man konstruiert anhand einer Stichprobe ein Intervall, welches den gesuchten Parameter mit hoher Wahrscheinlichkeit enthält.

9.4.2 Konfidenzintervall

Merke

Sei $(X_1, \ldots, X_n)$ eine Stichprobe zu einer Grundgesamtheit, deren Verteilung einen unbekannten Parameter ϑ enthält. Weiter sei $\alpha > 0$ eine (meist kleine) Zahl. Ein zufälliges Intervall $I(X_1, \ldots, X_n)$, welches

$$P\big(\vartheta \in I(X_1, \ldots, X_n)\big) = 1 - \alpha \tag{9.20}$$

erfüllt, heißt **Konfidenzintervall** (oder Vertrauensintervall) zu ϑ mit **Konfidenzniveau** (oder Vertrauensniveau) $1 - \alpha$. Die Zahl α wird auch als **Irrtumswahrscheinlichkeit** bezeichnet.

Konfidenzintervalle können einseitig oder zweiseitig sein. **Einseitige Konfidenzintervalle** haben die Form $(-\infty, b(X_1, \ldots, X_n)]$ bzw. $[a(X_1, \ldots, X_n), \infty)$ mit Funktionen $a, b : \mathbb{R}^n \to \mathbb{R}$. **Zweiseitige Konfidenzintervalle** haben die Form $[a(X_1, \ldots, X_n), b(X_1, \ldots, X_n)]$. Welche Variante gewählt wird, hängt vom konkreten Anwendungsfall ab.

Typische Werte für α sind 0.005, 0.01, 0.05. Je kleiner α desto größer werden die Konfidenzintervalle. Bei $\alpha = 0$ erhält man im Wesentlichen das Intervall $(-\infty, \infty)$, welches keinerlei nützliche Information zum gesuchten Parameter ϑ liefert.

Beispiel

Die mittlere Füllmenge von Flaschen aus einer Abfüllanlage soll bestimmt werden. Dabei sind Abweichungen nach oben nicht relevant; lediglich dauerhaft zu geringe Füllmengen müssen erkannt und vermieden werden. Anhand einer konkreten Stichprobe wird also ein Intervall $[a, \infty)$ bestimmt, welches mit hoher Wahrscheinlichkeit die mittlere Füllmenge enthält. Stellt sich die aus der Stichprobe berechnete Untergrenze a als groß genug heraus, kann der Abfüllprozess als fehlerfrei angesehen werden.

Achtung!

Ein zu einer konkreten Stichprobe berechnetes Konfidenzintervall lässt keine Aussage darüber zu, ob der gesuchte Parameter tatsächlich in diesem Intervall liegt. Wir wissen lediglich, dass beim Ziehen vieler Stichproben die meisten der zugehörigen Konfidenzintervalle den Parameter enthalten werden. Verwenden wir beispielsweise die Irrtumswahrscheinlichkeit $\alpha = 0.01$, so wissen wir, dass beim Ziehen vieler Stichproben im Mittel 99 Prozent der zugehörigen Intervalle den gesuchten Parameter enthalten.

9.4.3 Konstruktion von Konfidenzintervallen

9.4.3.1 Ansatz für zweiseitige Konfidenzintervalle Zu gegebener Irrtumswahrscheinlichkeit α und gegebener Schätzfunktion T_n für den gesuchten Parameter ϑ gilt

$$P(q_{\frac{\alpha}{2}} \leq T_n \leq q_{1-\frac{\alpha}{2}}) = 1 - \alpha, \tag{9.21}$$

wobei q_t das Quantil der Verteilung von T_n zum Niveau t ist. Die beiden Quantile hängen zwangsläufig von ϑ ab, sodass man die Ungleichungskette in $P(\ldots)$ nach ϑ umstellen kann und so die Grenzen des Konfidenzintervalls erhält.

Beispiel (Konfidenzintervall für den Erwartungswert bei normalverteilter Grundgesamtheit und bekannter Varianz)

Sei die Grundgesamtheit normalverteilt mit bekannter Varianz σ^2. Wollen zur Stichprobe $X_1, \ldots, X_n$ ein Konfidenzintervall für den Erwartungswert μ zu gegebener Irrtumswahrscheinlichkeit α konstruieren. Eine geeignete Schätzfunktion ist

$$T_n := \frac{1}{n} \sum_{k=1}^{n} X_k. \tag{9.22}$$

Diese folgt einer Normalverteilung mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$. Berechnung der Quantile und Umstellen nach μ liefert das Konfidenzintervall

$$\left[T_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}},\ T_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right] \tag{9.23}$$

für μ, wobei z_t das Quantil der Standardnormalverteilung bezeichnet (Video 940, Notizen dazu).

Die benötigten Quantile (insbesondere der Standardnormalverteilung) lassen sich oft nicht durch explizite Formeln berechnen. Früher wurden diese Tabellen entnommen; heute werden sie mit dem Computer numerisch berechnet.

9.4.3.2 Ansatz für einseitige Konfidenzintervalle Analog zum zweiseitigen Fall erhält man aus den Ansätzen

$$P(q_\alpha \leq T_n) = 1 - \alpha \tag{9.24}$$

bzw.

$$P(T_n \leq q_{1-\alpha}) = 1 - \alpha \tag{9.25}$$

einseitige Konfidenzintervalle.

9.4.3.3 Häufig benötigte Konfidenzintervalle Üblicherweise benötigte Konfidenzintervalle findet man in Formelsammlungen.

Je nach zu schätzendem Parameter folgt die passende Testfunktion unterschiedlichen Verteilungen. Die Quantile dieser oder daraus abgeleiteter Verteilungen tauchen dann in den Formeln für das Konfidenzintervall auf. Insbesondere spielen für Konfidenzintervalle bei normalverteilter Grundgesamtheit die Studentsche t-Verteilung und die Chi-Quadrat-Verteilung eine Rolle.

Beispiel

Beim maschinellen Zuschnitt von Dachlatten auf eine Solllänge von 4 Meter treten zwangsläufig Ungenauigkeiten auf. Diese sind praktisch wenig relevant, jedoch soll sichergestellt werden, dass die Länge im Mittel bei 4 Meter liegt, also nicht systematisch zu kurz (min. 3.98 Meter) oder zu lang (max. 4.05 Meter) geschnitten wird.

Es wird eine Stichprobe von 30 Latten nachgemessen. Der Mittelwert dieser Messungen ist 4.01 Meter. Die empirische Standardabweichung liegt bei 0.04 Meter. Liegt der Sollwert (einschließlich der zulässigen Toleranz nach unten und oben) im sich aus dieser Stichprobe ergebenden Konfidenzintervall zur Irrtumswahrscheinlichkeit $\alpha = 0.01$?

Das passende Konfidenzintervall (Erwartungswert der Normalverteilung bei unbekannter Varianz) ist

$$\left[\overline{x} - t_{1-\frac{\alpha}{2};n-1}\,\frac{s}{\sqrt{n}},\ \overline{x} + t_{1-\frac{\alpha}{2};n-1}\,\frac{s}{\sqrt{n}}\right], \tag{9.26}$$

wobei $t_{1-\frac{\alpha}{2};n-1}$ das Quantil zum Niveau $1 - \frac{\alpha}{2}$ der Studentschen t-Verteilung ist und

$$\overline{x} := \frac{1}{n}\sum_{k=1}^{n} x_k \tag{9.27}$$

und

$$s^2 := \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2 \tag{9.28}$$

Stichprobenmittelwert und Stichprobenvarianz sind. Einsetzen der gegebenen Werte:

```
import scipy.special

n = 30
mean = 4.01
std = 0.04
alpha = 0.01

t = scipy.special.stdtrit(n, 1 - 0.5 * alpha)
low = mean - t * std / n ** 0.5
high = mean + t * std / n ** 0.5

print(f'[ {low:.4f}, {high:.4f} ]')
```

```
[ 3.9899, 4.0301 ]
```

Der Zuschnittprozess kann also als in Ordnung angesehen werden, da das Konfidenzintervall, in dem der Sollwert der Lattenlänge mit hoher Wahrscheinlichkeit liegt, vollständig im tolerierbaren Bereich für den Sollwert enthalten ist.

Ist nur eine Mindestlänge, z.B. 3.98 Meter, einzuhalten, so sollte ein einseitiges Konfidenzintervall $[a, \infty)$ berechnet werden. Gilt damit $a \geq 3.98$, so ist der Zuschnitt als in Ordnung anzusehen.

9.5 Hypothesentests

Punkt- und Bereichsschätzungen liefern zwar gewisse Informationen über einen Verteilungsparameter ϑ der Grundgesamtheit; allerdings ist damit nicht geklärt, wie daraus Entscheidungen, z.B. über das korrekte Arbeiten einer Maschine, abgeleitet werden sollten. Insbesondere bereiten Konfidenzintervalle Probleme, die weder komplett innerhalb noch komplett außerhalb des gewünschten Bereichs für ϑ liegen. Hypothesentests erweitern die Schätzwerkzeuge um einen systematischen Entscheidungsprozess.

In der Praxis spielen Hypothesentests eine wichtige Rolle, da sie ein **Mittel zum Standardisieren von Entscheidungsprozessen** sind und Transparenz beim Aufstellen von Entscheidungsregeln schaffen.

Hypothesentests liefern letztlich den gleichen Informationsgewinn über den unbekannten Parameter ϑ wie Bereichsschätzungen, nehmen aber einen anderen Blickwinkel ein.

9.5.1 Ziel und Beispiel

Anhand einer Stichprobe $X_1, \ldots, X_n$ zu einer Grundgesamtheit soll eine Behauptung (Hypothese) über die Verteilung der Grundgesamtheit überprüft werden. Spricht eine konkrete Stichprobe für die zu prüfende Hypothese oder dagegen?

Beschränken uns hier auf **Parametertests**, also auf des Prüfen von Hypothesen über einen in der Verteilung der Grundgesamtheit vorkommenden, unbekannten Parameter ϑ. Nichtparametrische Tests werden hier nicht behandelt.

Beispiel

Eine Maschine zur Produktion elektronischer Bauteile liefert laut Hersteller mindestens 98 Prozent korrekt funktionierende Bauteile. Es gibt Zweifel, ob die neu beschaffte Maschine diese Zuverlässigkeit tatsächlich einhält. Zur Entscheidungsfindung werden im Laufe der Zeit diverse Stichproben entnommen:

- $n = 5$, alle Teile okay,
- $n = 100$, drei Teile defekt,
- $n = 100$, zwei Teile defekt,
- $n = 200$, drei Teile defekt.

Bei naiver Betrachtung (Ausschussanteil pro Stichprobe) sprechen zwei Stichproben für die Qualität der Maschine, eine dagegen und eine liefert einen Wert genau auf der Grenze. Ist die Maschine in Ordnung oder nicht? Falls man sich für eine Reklamation entscheidet: Wie begründet man diese?

9.5.2 Nullhypothese und Gegenhypothese

Die zu testende Hypothese wird als **Nullhypothese** (kurz: H_0) bezeichnet. Das Gegenteil der Nullhypothese wird als **Gegenhypothese** bezeichnet (kurz: H_1). Typische Nullhypothesen zum Verteilungsparameter ϑ sind:

- $\vartheta \geq \vartheta_0$ (einseitige Hypothese),
- $\vartheta \leq \vartheta_0$ (einseitige Hypothese),
- $\vartheta = \vartheta_0$ (zweiseitige Hypothese).

Dabei ist ϑ_0 eine gegebene Zahl.

Die Wahl, welche Behauptung als Nullhypothese verwendet wird (z.B. $\vartheta = \vartheta_0$ vs. $\vartheta \neq \vartheta_0$) ist dabei nicht beliebig. Das weitere Vorgehen beim Durchführen des Hypothesentests wird asymmetrisch sein. Es wird die Frage geprüft werden, **ob die konkrete Stichprobe gegen die Nullhypothese spricht**. Falls ja, dann wird die Nullhypothese verworfen. Falls nein, so geht man davon aus, dass die Nullhypothese gilt. Jedoch verwirft man in letzterem Fall die Gegenhypothese nicht! Dieses Vorgehen kann mit dem juristischen Grundsatz "im Zweifel für den Angeklagten" verglichen werden. Man sucht nicht nach einem Unschuldsbeweis, sondern nach einem Schuldbeweis. Solange dieser nicht vorliegt gilt die Unschuldsvermutung. Entsprechend ist als Nullhypothese die Hypothese zu wählen, die im gegebenen Kontext als zutreffender erscheint.

Beispiel

Eine Abfüllanlage soll auf Einhalten der Sollabfüllmenge ϑ_0 geprüft werden. Da aus Erfahrung (und der zweckmäßigen Konstruktion der Anlage) davon auszugehen ist, dass die mittlere Abfüllmenge ϑ der Sollabfüllmenge entspricht, wird die Nullhypothese

$$H_0 : \ \vartheta = \vartheta_0 \tag{9.29}$$

verwendet.

Es wird nun eine Stichprobe aus der Produktion entnommen. Spricht diese gegen die Nullhypothese, so geht man davon aus, dass die Anlage defekt ist bzw. die Gegenhypothese $\vartheta \neq \vartheta_0$ gilt. Spricht die Stichprobe nicht gegen die Nullhypothese, so geht man davon aus, dass die Anlage in Ordnung ist. Eine einzelne konkrete Stichprobe ist aber kein Beweis, dass die Anlage immer richtig abfüllt!

Aus praktischer Sicht ist die Nullhypothese hier: "Die Anlage füllt *immer* richtig ab." Gegenhypothese: "Die Anlage füllt *manchmal* falsch ab." In der mathematischen Formulierung der Hypothesen (also $\vartheta = \vartheta_0$ vs. $\vartheta \neq \vartheta_0$) taucht die Unterscheidung zwischen "immer" und "manchmal" nicht auf. Diese wird erst durch das asymmetrische Schließen aus der konkreten Stichprobe in den Entscheidungsprozess eingebracht.

Das asymmetrische Vorgehen ist nötig, da einerseits die Arbeit mit dem Verteilungsparameter ϑ keine Unterscheidung in "Hypothese gilt in jedem Fall" und "Hypothese gilt manchmal nicht" zulässt, andererseits aber genau diese Unterscheidung im praktischen Einsatz relevant ist. **Erst die Kombination aus abstrakter, deterministischer Formulierung (Hypothese über Verteilungsparameter) und der zweckmäßigen Interpretation von zufälligen Vorgängen (Stichprobe) schafft ein wirkungsvolles mathematisches Werkzeug!**

9.5.3 Bewertung der Stichprobe

Es bleibt noch zu klären, wann eine Stichprobe für bzw. gegen die Nullhypothese spricht. Man nimmt an, dass die Nullhypothese erfüllt ist und betrachtet die sich daraus ergebenden Wahrscheinlichkeiten für alle denkbaren Stichproben. Um nicht direkt mit den vektorwertigen Stichproben arbeiten zu müssen, wählt man eine **Schätzfunktion** T_n (manchmal auch **Testfunktion** genannt), die Informationen aus der Stichprobe zu einem Wert zusammenfasst.

Üblicherweise (z.B. bei normalverteilter Grundgesamtheit) kann jede konkrete Stichprobe bzw. jeder Schätzwert auftreten, sodass die Frage, ob die gezogene Stichprobe unter Annahme von H_0 möglich ist, stets mit "ja" zu beantworten ist, also keine sinnvolle Entscheidungsgrundlage liefert. Analog zum Vorgehen bei Bereichsschätzungen legt man (vor dem Ziehen der Stichprobe!) eine kleine **Irrtumswahrscheinlichkeit** $\alpha > 0$ fest und wählt einen sogenannten **kritischen Bereich** $K \subseteq \mathbb{R}$, der

$$P(T_n \in K) = \alpha \tag{9.30}$$

erfüllt. Der kritische Bereich markiert also die Schätzwerte, die unter Annahme von H_0 sehr unwahrscheinlich sind.

Gilt nun $t \in K$ für den zur konkreten Stichprobe gehörenden Schätzwert t, so verwirft man die Nullhypothese, da der Schätzwert (und damit die Stichprobe) bei Gültigkeit von H_0 eigentlich nicht auftreten sollte (bis auf die kleine Irrtumswahrscheinlichkeit α). Gilt hingegen $t \notin K$, so ist die beobachtete Stichprobe eine von denen, die unter Annahme von H_0 zu erwarten ist. Die Stichprobe spricht in diesem Fall also nicht gegen H_0.

Die Wahl von T_n und K ist nicht beliebig, sondern sollte so erfolgen, dass der Test vorteilhafte Eigenschaften hat. Was das im Detail bedeutet und wie man T_n und K systematisch wählt, wird hier nicht behandelt. Für übliche Testaufgaben, können geeignete Wahlen aus Formelsammlungen entnommen werden.

Übersicht der Arbeitsschritte:

1. Nullhypothese H_0 aufstellen.
2. Irrtumswahrscheinlichkeit α wählen, z.B. 0.005, 0.01, 0.05.
3. Schätzfunktion T_n und kritischen Bereich K wählen (Formelsammlung).
4. Schätzwert t zur konkreten Stichprobe berechnen.

5. Entscheidung treffen: Gilt $t \notin K$, so wird H_0 angenommen, sonst abgelehnt.

Beispiel

Ein Gerät zur Längenmessung soll auf korrekte Funktion überprüft werden. Die Funktion wird als korrekt angesehen, wenn das Gerät im Mittel die richtige Länge liefert. Dazu werden $n = 10$ Messungen einer Strecke von 1000 Metern durchgeführt. Folgende Messwerte werden abgelesen (in Meter):

$$998.0,\ 1001.0,\ 1003.0,\ 1000.5,\ 999.0,\ 997.5,\ 1000.0,\ 999.5,\ 996.0,\ 998.5. \tag{9.31}$$

Kann man auf dieser Grundlage sagen, dass das Gerät korrekt arbeitet?

Gehen davon aus, dass die Messwerte einer Normalverteilung mit Erwartungswert μ folgen, wobei wir die Varianz nicht kennen.

Nullhypothese: $\mu = 1000$.

Irrtumswahrscheinlichkeit: $\alpha = 0.025$.

Test für Erwartungswert bei normalverteilter Grundgesamtheit mit unbekannter Varianz (Einstichproben-t-Test):

- Schätzwert:

$$t = (\overline{x} - 1000)\,\frac{\sqrt{n}}{s} \tag{9.32}$$

  mit

$$\overline{x} := \frac{1}{n}\sum_{k=1}^{n} x_k \qquad \text{und} \qquad s^2 := \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2 \tag{9.33}$$

- kritischer Bereich:

$$K = \{x \in \mathbb{R} :\ |x| > t_{1-\frac{\alpha}{2},n-1}\}, \tag{9.34}$$

  wobei $t_{1-\frac{\alpha}{2},n-1}$ ein Quantil der Studentschen t-Verteilung ist.

```
import numpy as np
import scipy.special

x = np.array([998.0, 1001.0, 1003.0, 1000.5, 999.0,
              997.5, 1000.0, 999.5, 996.0, 998.5])

n = len(x)
```

```
mu0 = 1000
alpha = 0.025

t = (x.mean() - mu0) * n ** 0.5 / x.std(ddof=1)
K_high = scipy.special.stdtrit(n, 1 - 0.5 * alpha)
K_low = -K_high

print(f't = {t:0.4f}')
print(f'K = ( - \infty, {K_low:0.4f})   ({K_high:0.4f}, \infty)')
```

```
t = -1.1209
K = ( - \infty, -2.6338)   (2.6338, \infty)
```

Die Stichprobe spricht also nicht gegen die Nullhypothese. Auf Basis der Stichprobe kann davon ausgegangen werden, dass das Gerät korrekt arbeitet.

9.5.4 Fehlentscheidungen

Auch wenn die Nullhypothese wahr sein sollte, wird bei $100\,\alpha$ Prozent der konkreten Stichproben die Nullhypothese abgelehnt. Diese Fehlentscheidung nennt man **Fehler 1. Art** (oder α-Fehler). Die Wahrscheinlichkeit für eine solche Fehlentscheidung liegt bei α. Wurde $\alpha = 0.01$ gewählt, so führen im Mittel 1 Prozent der gezogenen Stichproben zum Ablehnen der Nullhypothese, obwohl ist richtig war.

Ist umgekehrt die Nullhypothese falsch, so kann es sein, dass sie dennoch auf Grundlage einer konkreten Stichprobe nicht abgelehnt wird. Die Wahrscheinlichkeit für einen solchen **Fehler 2. Art** (oder β-Fehler) ist im Allgemeinen unbekannt. Bei der Konstruktion von Tests (Schätzfunktion und kritischer Bereich) bemüht man sich jedoch, Fehler 2. Art unter allen denkbaren Tests möglichst klein zu bekommen. Ist dies nachweisbar gelungen, so spricht man auch von einem **optimalen Test**.

Merke

Wie immer bei statistischen Betrachtungen gilt auch hier: Im Einzelfall kann eine Entscheidung falsch sein. Aber im Mittel über viele durchgeführte Tests werden die Entscheidungen überwiegend korrekt sein.

10 Übung 1 (Grundlegendes)

10.1 Logik, Argumentation, Begriffsbildung

10.1.1 Aufgabe 1.1

Beantworten Sie die im Beispiel mit dem Hase und der Schildkröte gestellten Fragen.

10.1.2 Aufgabe 1.2

Diskutieren Sie die in Beispiel zum Begriff "gleich viel" gestellten Fragen. Geben Sie eine für plausible Definition für "zwei Mengen enthalten gleich viele Elemente" an, mit der sich die Fragen beantworten lassen. Hinweis: Hier gibt es verschiedene Antwortmöglichkeiten, die sich ggf. widersprechen, aber dennoch als richtig angesehen werden können.

10.1.3 Aufgabe 1.3

Diskutieren Sie die folgende Argumentation. Wo liegt der Fehler?

In einen Koffer passen unendlich viele Briefmarken, denn würden nur endlich viele hinein passen, so gäbe es also eine maximale Anzahl $n \in \mathbb{N}$ an Briefmarken, die in den Koffer passen. Jedoch würde man auch dann, wenn schon n Briefmarken im Koffer sind, noch irgendwo eine weitere Briefmarke einbringen können (kein Koffer kann so voll sein, dass keine Briefmarke mehr rein geht). Somit sind nun $n+1$ Briefmarken im Koffer. Die Zahl n kann also doch nicht die Maximalanzahl sein. Die Annahme der endlichen Anzahl führt somit zum Widerspruch. Folglich passen unendlich viele Briefmarken in den Koffer.

10.2 Wiederholung Schulmathematik

10.2.1 Aufgabe 1.4

Lösen Sie den Eingangstest MINTFIT von OMB+. Dazu ist eine kostenlose und datensparsame Registrierung nötig.

10.2.2 Aufgabe 1.5

Gleich Sie die im Eingangstest identifizierten Defizite mit der Section 2.2, ab. Erstellen Sie einen Zeitplan für das selbstständige Aufarbeiten der relevanten Punkte im Laufe des Semesters.

11 Übung 2 (Python und Jupyter)

11.1 Jupyter

11.1.1 Aufgabe 2.1

Starten Sie durch Klick auf den Link ein browser-basiertes JupyterLab

Erstellen Sie ein neues Dokument ("Notebook"), berechnen Sie darin 123 * 456 und speichern Sie das Dokument.

Das gespeicherte Dokument liegt nun im internen Speicherbereich Ihres Browsers. Um es bei Ihren Dateien zu speichern, können Sie den Download-Button verwenden.

11.1.2 Aufgabe 2.2

Starten Sie ein JupyterLab auf mybinder.org. Laden Sie die in der vorhergehenden Aufgabe gespeicherte Datei dort hoch und öffnen Sie sie.

11.1.3 Aufgabe 2.3

Machen Sie sich mit Jupyter-Notebooks vertraut. Fügen Sie neue Zellen ein. Probieren Sie auch den Zelltyp "Markdown" aus für Text, der nicht von Python verarbeitet werden soll.

11.1.4 Aufgabe 2.4 (optional)

Installieren Sie JupyterLab lokal auf Ihrem Computer (siehe z.B. Install Jupyter Locally in Data Science and Artificial Intelligence for Undergraduates).

11.2 Python

11.2.1 Aufgabe 2.5

Werten Sie das Polynom $x^3 + 2\,x^2 + 1$ an den Stellen 1, 3, 4, 6, 8 aus. Ergänzen Sie dazu den folgenden Code:

```
stellen = [1, 3, 4, 6, 8]
for x in stellen:
    # hier ergänzen
```

11.2.2 Aufgabe 2.6

Geben Sie alle durch 4 teilbaren Zahlen zwischen 9 und 99 aus. Verwenden Sie dazu eine `for`-Schleife.

```
# Ihre Lösung
```

11.2.3 Aufgabe 2.7

Zählen Sie die geraden Zahlen in einer gegebenen Liste.

```
liste = [2, 5, 4, 3, 3, 2, 0, 6, 7, 5, 6, 7]
# weitere Listen zum Testen:
#liste = [2, 4]
#liste = [1, 3]
#liste = []

# Ihre Lösung
```

12 Übung 3 (Aussagen, Abbildungen, Folgen)

12.1 Aussagen

12.1.1 Aufgabe 3.1 (KTA)

Bilden Sie zu folgenden Aussagen die genau gegenteiligen Aussagen. Dabei spielt es keine Rolle, ob die gegeben Aussagen wahr oder falsch sind!

- **(a)** Alle Studierenden bestehen die Mathematikprüfung.
- **(b)** Alle Studierenden bestehen alle Prüfungen dieses Semesters.
- **(c)** Es gibt Studierende, die dieses Semester durch mindestens eine Prüfung fallen.
- **(d)** Wenn alle Studierenden die Prüfung bestanden haben, sind alle Studierenden glücklich.

12.1.2 Aufgabe 3.2 (KTA)

Skizzieren Sie die Menge

$$\{(x, y) \in \mathbb{R}^2 : \; y \geq 0, \; (x-1)^2 + y^2 \leq 1, \; 2\,x + y \leq 2\}. \tag{12.1}$$

12.2 Abbildungen

12.2.1 Aufgabe 3.3 (KTA)

Entscheiden Sie für jede der folgenden Abbildungen, welche Abbildung injektiv/surjektiv/bijektiv ist. Ist keine Grundmenge für den Wertebereich angegeben, so ist die kanonische Menge zu wählen (bei (a) z.B. $\mathbb{N}$).

- **(a)** Jedem Insekt wird die Anzahl seiner Beine zugeordnet.
- **(b)** Jedem Land wird seine Nationalflagge zugeordnet.
- **(c)** Jedem Bundestagsabgeordneten wird ein Sitzplatz im Parlamentssaal zugeordnet.

12.3 Zahlenfolgen

12.3.1 Aufgabe 3.4 (KTA)

Zeigen Sie durch Berechnen der folgenden Grenzwerte, dass Summen der Form "$\infty - \infty$" verschiedene Werte annehmen können und somit stets einer Einzelfallbetrachtung bedürfen.

- **(a)** $\lim\limits_{n\to\infty} (n^2 - n)$
- **(b)** $\lim\limits_{n\to\infty} (n^2 - n^3)$
- **(c)** $\lim\limits_{n\to\infty} \left(n^2 - \dfrac{n^3}{n + \frac{1}{n}} \right)$

12.3.2 Aufgabe 3.5

Stellen Sie durch numerisches Auswerten der Folgenglieder

$$a_n = \left(\frac{(n-1)^2}{(2^3)^4\, n + 1} \right)^{\frac{1}{50}} \tag{12.2}$$

eine Vermutung zum Konvergenzverhalten der Folge $(a_n)_{n\in\mathbb{N}}$ an. Überprüfen Sie Ihre Vermutung durch Berechnung des Grenzwerts mit SymPy.

```
# Ihre Lösung
```

12.3.3 Aufgabe 3.6

Durch

$$\begin{aligned} x_1 &:= 0, \\ y_1 &:= 0, \\ x_{n+1} &:= x_n^2 - y_n^2 + \frac{1}{5}, \quad n \in \mathbb{N}, \\ y_{n+1} &:= 2\, x_n\, y_n + \frac{1}{2}, \quad n \in \mathbb{N}, \end{aligned} \tag{12.3}$$

ist eine Folge $\big((x_n, y_n)\big)_{n\in\mathbb{N}}$ von Punkten in der Ebene gegeben.

- **(a)** Stellen Sie die ersten 100 Elemente der Punktfolge grafisch dar.
- **(b)** Stellen Sie eine Vermutung zum Konvergenzverhalten an, wenn wir unter Konvergenz einer Punktfolge die Konvergenz beider Koordinatenfolgen verstehen.
- **(c)** Formulieren Sie anhand der Definition des Konvergenzbegriffs für Zahlenfolgen die Bedingungen, die für den Nachweis der Konvergenz zu überprüfen wären.

```
# Ihre Lösung
```

12.3.4 Aufgabe 3.7 (KTA)

Wird ein Kaptial von 1000000 Euro nach einem Jahr mit 1 Prozent Zinsen vergütet, so stehen nach diesem Jahr

$$1000000\,(1 + 0.01) = 1010000 \tag{12.4}$$

Euro zur Verfügung. Werden hingegen monatlich Zinsen zu $\frac{1}{12}$ des jährlichen Zinssatzes gezahlt, so sind nach Jahresende

$$1000000\left(1 + \frac{0.01}{12}\right)^{12} \approx 1010045.96 \tag{12.5}$$

Euro vorhanden. Bei täglichlicher Verzinsung würde man

$$1000000\left(1 + \frac{0.01}{365}\right)^{365} \approx 1010050.03 \tag{12.6}$$

Euro erhalten.

Lässt sich durch immer feinere Unterteilung des Jahres in Zinsabschnitte (Stunden, Minuten,...) das Endkapital beliebig erhöhen? Begründen Sie Ihre Antwort!

13 Übung 4 (Funktionen)

13.1 Definitionsbereich

13.1.1 Aufgabe 4.1 (KTA)

Bestimmen Sie die größtmögliche Menge $D \subseteq \mathbb{R}$, sodass die folgenden Zuordnungsvorschriften für alle $x \in D$ einen sinnvollen Wert liefern. Skizzieren Sie jeweils die Menge D.

- **(a)** $x \mapsto \frac{5}{(x-1)^2}$,
- **(b)** $x \mapsto \frac{x}{x}$,
- **(c)** $x \mapsto \sqrt{9 - |x|}$,
- **(d)** $x \mapsto \ln(x^2 - 2\,x + 2)$,
- **(e)** $x \mapsto \sqrt{4 - x} + \sqrt{x - 3}$.

13.1.2 Aufgabe 4.2 (KTA)

Bestimmen Sie die größtmögliche Menge $D \subseteq \mathbb{R}^2$, sodass die folgenden Zuordnungsvorschriften für alle $(x, y) \in D$ einen sinnvollen Wert liefern. Skizzieren Sie jeweils die Menge D.

- **(a)** $(x, y) \mapsto \ln(x^2 - y^2)$,
- **(b)** $(x, y) \mapsto \sqrt{y - 2\,x}$,
- **(c)** $(x, y) \mapsto (9 - x^2 - y^2)^{-2}$,
- **(d)** $(x, y) \mapsto \frac{\sqrt{y-x}}{y}$,
- **(e)** $(x, y) \mapsto \mathrm{e}^{-x^2}$,
- **(f)** $(x, y) \mapsto \tan\left(\frac{1}{x+y}\right)$.

13.2 Visualisierung

Nebenbemerkung

Auch wenn es für das korrekte Lösen der folgenden Übungsaufgaben nicht zwingend nötig ist, so sei doch die Beschäftigung mit weiteren Funktionen des Matplotlib-Pakets empfohlen (Achseneinteilung, Gitterlinien, Farben, Linienarten, Legende,...). Einen Überblick über die Möglichkeiten bietet beispielsweise das Kapitel Matplotlib Basics in DSAI.

13.2.1 Aufgabe 4.3

Visualisieren Sie die Funktion $f : [-2\,\pi, 2\,\pi] \to \mathbb{R}$, $f(x) := x^2 \cos(10\,x)$.

```
# Ihre Lösung
```

13.2.2 Aufgabe 4.4

Visualisieren Sie die Funktion $f : [-\frac{\pi}{8}, \frac{\pi}{8}] \to \mathbb{R}$, $f(x) := \sin \frac{1}{x}$.

```
# Ihre Lösung
```

13.2.3 Aufgabe 4.5

Visualisieren Sie die durch

$$x(t) := \cos(10\,t), \quad y(t) = \sin(10\,t), \quad t \in [0, 2\,\pi] \tag{13.1}$$

gegebene Kurve im Raum als 3D-Plot.

```
# Ihre Lösung
```

13.3 Grenzwerte und Stetigkeit

13.3.1 Aufgabe 4.6 (KTA)

Zeigen Sie ausgehend von der Definition des Grenzwerts bei Funktionen, dass der Grenzwert

$$\lim_{x \to 0} \sin \frac{1}{x} \tag{13.2}$$

nicht existiert.

13.3.2 Aufgabe 4.7

Ermitteln Sie mit SymPy, ob folgende Grenzwerte existieren und welchen Wert sie ggf. haben.

Hinweis: Das Berechnen von Funktionsgrenzwerten mit SymPy funktioniert genauso wie das Berechnen von Folgengrenzwerten. Allerdings liefert SymPy stets nur den rechtsseitigen Grenzwert. Mit `sympy.limit(..., x, x0, '-')` erhält man den linksseitigen Grenzwert. Die direkte Berechnung des beidseitigen Grenzwerts wird nicht unterstützt.

- **(a)** $\lim\limits_{x \to 2} \dfrac{(x-2)^2}{x^2 - 4}$,

- **(b)** $\lim\limits_{x \to 0} \dfrac{x}{|x|}$,
- **(c)** $\lim\limits_{x \to 0} \sin \dfrac{1}{x}$.

13.3.3 Aufgabe 4.8

Implementieren Sie die Funktion

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) := \begin{cases} -1 - x^2, & \text{falls } x < 0, \\ 0, & \text{falls } x = 0, \\ 1 + x^2, & \text{falls } x > 0 \end{cases} \tag{13.3}$$

als Python-Funktion. Testen Sie Ihre Funktion mit dem folgenden Code:

```
def f(x):
    # Hier Code einfügen

print(f'f( -2) = {f( -2)}')
print(f'f( -1) = {f( -1)}')
print(f'f(0) = {f(0)}')
print(f'f(1) = {f(1)}')
print(f'f(2) = {f(2)}')
print(f'f(0) = {f(0.1 ** 2 - 0.01)}')
```

Warum wird trotz korrekter Implementierung ein falscher Funktionswert in der letzten Zeile ausgegeben? Kann dies für stetige Funktionen ebenfalls passieren?

Zusatz: Visualisieren Sie die Funktion f. Achten Sie insbesondere auf die korrekte Darstellung für $x \approx 0$.

```
# Ihre Lösung
```

14 Übung 5 (Matrizen, Vektorgeometrie)

14.1 Rechnen mit Matrizen

14.1.1 Aufgabe 5.1 (KTA)

Geben Sie zu folgenden Ausdrücken die Größe der Ergebnismatrix an:

- **(a)** $A\,B\,C, \quad A \in \mathbb{R}^{3\times 7},\ B \in \mathbb{R}^{7\times 5},\ C \in \mathbb{R}^{5\times 4},$
- **(b)** $(A\,B^{\mathrm{T}})^{\mathrm{T}}\,C, \quad A \in \mathbb{R}^{3\times 4},\ B \in \mathbb{R}^{5\times 4},\ C \in \mathbb{R}^{3\times 4},$
- **(c)** $x^{\mathrm{T}}\,y \quad x \in \mathbb{R}^{n},\ y \in \mathbb{R}^{n},$
- **(d)** $x\,y^{\mathrm{T}} \quad x \in \mathbb{R}^{n},\ y \in \mathbb{R}^{n},$
- **(e)** $x^{\mathrm{T}}\,A\,x \quad x \in \mathbb{R}^{n},\ A \in \mathbb{R}^{n\times n}.$

14.1.2 Aufgabe 5.2 (KTA)

Berechnen Sie die folgenden Produkte und Summen:

- **(a)** $\begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 5 \\ 1 & 1 \\ 2 & 0 \end{bmatrix},$
- **(b)** $\begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix},$
- **(c)** $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}.$

14.1.3 Aufgabe 5.3 (KTA)

Rechnen Sie die Beziehung $(A\,B)^{\mathrm{T}} = B^{\mathrm{T}}\,A^{\mathrm{T}}$ für die beiden Matrizen

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 3 & 2 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & 5 \\ 1 & 1 \\ 2 & 0 \end{bmatrix} \tag{14.1}$$

nach.

14.1.4 Aufgabe 5.4 (KTA)

Berechnen Sie die Determinanten:

- **(a)** $\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}$,
- **(b)** $\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$,
- **(c)** $\begin{vmatrix} 2 & 0 & 1 \\ 3 & 2 & 0 \\ 1 & 1 & 1 \end{vmatrix}$.

14.2 Rechnen mit Vektoren

14.2.1 Aufgabe 5.5 (KTA)

Die Eckpunkte eines Dreiecks im Raum sind $A = (2, 3, 4)$, $B = (3, 1, 7)$, $C = (1, 8, 2)$. Berechnen Sie die Länge der Seite $\overline{AB}$ sowie den Winkel am Punkt B.

14.2.2 Aufgabe 5.6 (KTA)

Im Raum sind die drei Punkte $A(1, -2, 3)$, $B = (3, 2, 1)$, $C = (6, 4, 4)$ gegeben. Ermitteln Sie eine Punkt D, sodass die Punktfolge A, B, C, D ein Parallelogramm beschreibt.

14.2.3 Aufgabe 5.7 (KTA)

Seien $a, b \in \mathbb{R}^3$ zwei linear unabhängige Vektoren. Geben Sie eine Formel zur Berechnung eines Vektors $w \in \mathbb{R}^3$ entlang der Winkelhalbierenden zwischen a und b an.

14.3 Skalarprodukt und orthogonale Projektion

14.3.1 Aufgabe 5.8 (KTA)

Finden Sie alle Vektoren, die orthogonal zum Vektor $\begin{bmatrix} 3 & -1 & 2 \end{bmatrix}^{\mathrm{T}}$ sind.

14.3.2 Aufgabe 5.9 (KTA)

Berechnen Sie in $\mathbb{R}^5$ die orthogonale Projektion von $\begin{bmatrix} 1 & 0 & 1 & -1 & 2 \end{bmatrix}^{\mathrm{T}}$ auf $\begin{bmatrix} 0 & 2 & -1 & -2 & 0 \end{bmatrix}^{\mathrm{T}}$.

14.3.3 Aufgabe 5.10 (KTA)

Von einem Eckpunkt eines Quadrates werden Geraden durch die Mittelpunkte der beiden dem Eckpunkt gegenüberliegenden Seiten gezogen. Wie groß ist der Winkel zwischen den

beiden Geraden?

14.3.4 Aufgabe 5.11 (KTA)

Eine Kraft vom Betrag 60 N, die in Richtung des Vektors $\begin{bmatrix}-1 & 3 & 2\end{bmatrix}^{\mathrm{T}}$ wirkt, bewegt einen Körper vom Punkt $(2, -2, 1)$ geradlinig zum Punkt $(3, 4, 5)$ (Ortsangaben in Meter). Welche Arbeit wird dabei verrichtet?

15 Übung 6 (lineare Abbildungen, Basen)

15.1 Lineare Abbildungen

15.1.1 Aufgabe 6.1 (KTA)

Sind folgende Abbildungen linear? Begründen Sie Ihre Entscheidung.

- **(a)** $f : \mathbb{R}^3 \to \mathbb{R},\ x \mapsto x_1$,
- **(b)** $f : \mathbb{R}^2 \to \mathbb{R}^2,\ (x, y) \mapsto (-y, x)$,
- **(c)** $f : \mathbb{R}^2 \to \mathbb{R},\ (x, y) \mapsto x\, y$,
- **(d)** $f : \mathbb{R}^2 \to \mathbb{R},\ (x, y) \mapsto x + y$,
- **(e)** $f : \mathbb{R}^2 \to \mathbb{R}^2,\ (x, y) \mapsto (x + 1, y - 1)$.

15.1.2 Aufgabe 6.2 (KTA)

Geben Sie zu folgenden linearen Abbildungen die Matrixdarstellung an:

- **(a)** $f : \mathbb{R}^3 \to \mathbb{R}^3$, 90°-Drehung um die z-Achse gegen die Uhr (bei Betrachtung vom Punkt $(0, 0, 1)$ aus),
- **(b)** $f : \mathbb{R}^3 \to \mathbb{R}$, Bilden der Summe aller Komponenten,
- **(c)** $f : \mathbb{R}^2 \to \mathbb{R}^2$, Drehung um $\frac{\pi}{4}$ im Uhrzeigersinn,
- **(d)** $f : \mathbb{R}^2 \to \mathbb{R}^2$, orthogonale Projektion auf den Vektor $\begin{bmatrix} 1 & 1 \end{bmatrix}^{\mathrm{T}}$.

15.2 Unterräume und Basen

15.2.1 Aufgabe 6.3 (KTA)

Zeigen Sie, dass die Menge

$$U := \{x \in \mathbb{R}^4 :\ x_3 = 2\, x_1 - x_2\} \tag{15.1}$$

ein linearer Unterraum von $\mathbb{R}^4$ ist.

Welche Dimension hat U? Geben Sie eine Basis von U an.

15.2.2 Aufgabe 6.4 (KTA)

Begründen Sie anhand der Definition der linearen Unabhängigkeit, dass die Menge

$$\left\{ \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \right\} \tag{15.2}$$

linear abhängig ist.

15.2.3 Aufgabe 6.5 (KTA)

Sind die folgenden Aussagen wahr oder falsch? Begründen Sie jeweils kurz Ihre Antwort.

- **(a)** Jede ONB ist ein ONS.
- **(b)** Jedes ONS ist eine Basis in seiner Aufspannung.
- **(c)** Jede Basis in $\mathbb{R}^2$ ist eine ONB.
- **(d)** Ist $\{b_1, b_2\}$ eine ONB in einem zweidimensionalen Unterraum des $\mathbb{R}^n$, so ist die orthogonale Projektion von b_1 auf b_2 stets der Nullvektor.

16 Übung 7 (lineare Gleichungssysteme)

16.1 Formulieren von LGS

16.1.1 Aufgabe 7.1 (KTA)

Bringen Sie das folgende lineare Gleichungssystem in Matrix-Vektor-Schreibweise:

$$\begin{aligned} 3\,x_1 + x_3 &= 2\,x_2 + 2 \\ 2\,x_3 + x_1 + x_2 &= 1 + x_1 \\ 5 - x_2 - x_1 &= 5 + x_1. \end{aligned} \tag{16.1}$$

16.1.2 Aufgabe 7.2 (KTA)

An der Innenseite einer Hauswand liegt die Oberflächentemperatur bei $U_\mathrm{i} = 22$ [°C]. Die Temperatur an der Außenseite beträgt $U_\mathrm{a} = -15$ [°C]. Zur (näherungsweisen) Berechnung des Temperaturverlaufs im Inneren der Wand wird die Wand in n Schichten unterteilt (parallel zur Wandoberfläche). Anhand physikalischen Gesetzmäßigkeiten kann man sich überlegen, dass die Temperatur u_k in in der k-ten Schicht dann gerade der Mittelwert aus den Temperaturen u_{k-1} und u_{k+1} der beiden benachbarten Schichten ist. Stellen Sie ein LGS zur Berechnung der Temperaturen $u_2, \ldots, u_{n-1}$ für $n = 7$ auf. Die Temperaturen in den äußeren Schichten entsprechen den Oberflächentemperaturen, also $u_1 = U_\mathrm{i}$ und $u_n = U_\mathrm{a}$, sind also bekannt.

16.1.3 Aufgabe 7.3

Lösen Sie das LGS aus der vorhergehenden Aufgabe mit NumPy und visualieren Sie das Ergebnis. Gestalten Sie den Quellcode so, dass Sie die Schichtanzahl n leicht verändern können.

Probieren Sie verschiedene Werte für n aus. Testen Sie Ihren Code auch mit verschiedenen Oberflächentemperaturen. Leiten Sie aus der Simulation eine Vermutung über die allgemeine Form des Temperaturverlaufs in der Wand ab.

```
# Ihre Lösung
```

16.2 Lineare Unabhängigkeit

16.2.1 Aufgabe 7.4

Sind die Vektoren

$$\begin{bmatrix}1\\7\\2\end{bmatrix}, \begin{bmatrix}3\\1\\2\end{bmatrix}, \begin{bmatrix}2\\-1\\1\end{bmatrix} \tag{16.2}$$

linear unabhängig? Formulieren Sie zur Klärung dieser Frage ein geeignetes LGS und lösen Sie dieses mit dem Computer.

```
# Ihre Lösung
```

16.2.2 Aufgabe 7.5

Lösen Sie die folgenden Teilaufgaben mit dem Computer.

- **(a)** Zeigen Sie, dass

$$u := \begin{bmatrix}2\\1\\-1\\0\end{bmatrix}, \quad v := \begin{bmatrix}0\\-3\\2\\-4\end{bmatrix}, \quad w := \begin{bmatrix}1\\-2\\4\\1\end{bmatrix} \tag{16.3}$$

  linear unabhängig sind.

- **(b)** Welcher der Vektoren

$$x := \begin{bmatrix}1\\0\\-3\\-5\end{bmatrix}, \quad y := \begin{bmatrix}2\\-1\\6\\6\end{bmatrix}, \quad z := \begin{bmatrix}-1\\1\\0\\5\end{bmatrix} \tag{16.4}$$

  ist geeignet, zusammen mit u, v, w eine Basis in $\mathbb{R}^4$ zu bilden?

- **(c)** Ermitteln Sie die Koordinaten des Vektors $\begin{bmatrix}0\\1\\-3\\0\end{bmatrix}$ bzgl. der Basis aus (b).

```
# Ihre Lösung
```

16.3 Geraden und Ebenen

16.3.1 Aufgabe 7.6 (KTA)

Bestimmen Sie den Mittelpunkt und den Radius des Kreises, der bei Rotation des Punktes $(-2, 1, 1)$ um die Gerade durch

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R}, \tag{16.5}$$

beschriebene Gerade erzeugt wird.

16.3.2 Aufgabe 7.7 (KTA)

Bestimmen Sie die Koordinaten des Spiegelpunktes von $(-7, 9, 1)$ bezüglich der Ebene durch $5\,x - 4\,y - z = 12$ beschriebenen Ebene.

17 Übung 8 (Koordinantentransformationen)

17.1 Aufgabe 8.1 (KTA)

Die zwei Vektoren $a := \begin{bmatrix}1 & 1 & 0\end{bmatrix}^{\mathrm{T}}$ und $b := \begin{bmatrix}2 & 1 & 1\end{bmatrix}^{\mathrm{T}}$ spannen eine Ebene im $\mathbb{R}^3$ auf. Finden Sie eine ONB $\{u, v, w\}$, sodass u und v in dieser Ebene liegen und u die gleiche Richtung hat wir a.

Hinweis: Nutzen Sie das Kreuzprodukt von Vektoren. Finden Sie zunächst u, dann w und dann v.

17.2 Aufgabe 8.2 (KTA)

Im $\mathbb{R}^2$ wird ein neues Koordinatensystem durch Drehen des Standardkoordinatensystems um 30° gegen den Uhrzeigersinn und anschließendes Verschieben des Ursprungs in den Punkt $(2, 1)$ erzeugt.

- **(a)** Geben Sie die Transformationvorschrift zum Erhalt der neuen Koordinaten aus den ursprünglichen Koordinaten an.
- **(b)** Welche Koordinaten hat der Punkt $(1, 1)$ bzgl. des neuen Koordinatensystems?

17.3 Aufgabe 8.3 (KTA)

Im $\mathbb{R}^2$ wird ein neues Koordinatensystem durch Drehen des Standardkoordinatensystems um 60° im Uhrzeigersinn und anschließendes Verschieben des Ursprungs in den Punkt $(1, 3)$ erzeugt. Geben Sie die Transformationsvorschrift für die Berechnung der neuen Koordinaten aus den Koordinaten bzgl. des Koordinatensystems aus der vorhergehenden Aufgabe an.

18 Übung 9 (Differentialrechnung)

18.1 Aufgabe 9.1 (KTA)

Ermitteln Sie die Ableitungen folgender Funktionen manuell.

- **(a)** $f(x) = 5\,x^4 - 2\,x^3 + x + 4$,
- **(b)** $(x) = (\sin x)\,\cos x$,
- **(c)** $f(x) = \frac{\sin x}{x^2+2}$,
- **(d)** $f(x) = (\cos x)^2$,
- **(e)** $f(x) = (\sin x)^2\,(x^3 + x)$.

18.2 Aufgabe 9.2

Überprüfen Sie Ihre Ergebnisse in der vorhergehenden Aufgabe mit SymPy.

18.3 Aufgabe 9.3 (KTA)

Stellen Sie das Taylor-Polynom vom Grad 3 für die Cosinus-Funktion an der Entwicklungsstelle $x_0 = 0$ auf.

Geben Sie auch die maximale Abweichung zwischen der Cosiunus-Funktion und dem gefundenen Taylor-Polynom auf dem Intervall $[0, \frac{\pi}{4}]$ an. (Lösung: rund 0.191)

18.4 Aufgabe 9.4 (KTA)

Zur Bestimmung des Durchmessers einer Kugel wurde zunächst deren Volumen bestimmt, indem ihre Verdrängung in einem Wasserbehälter gemessen wurde. Die Messung ergab ein Volumen von 1.45 dm^3 bei einer Messunsicherheit von ± 0.005 dm^3. Aus diesem Wert wurde anschließend der Kugeldurchmesser mit der Formel

$$V = \frac{\pi}{6}\,r^3 \tag{18.1}$$

berechnet. Wie groß ist der zu erwartende relative Fehler des berechneten Durchmessers? (Lösung: rund 0.11%)

18.5 Aufgabe 9.5 (KTA)

Sollen geografische Breite und Länge in Koordinaten im $\mathbb{R}^3$ umgerechnet werden, so ergibt sich die z-Koordinate (in Kilometer) aus der Breite ζ (in Grad) mittels

$$z = 6371\,\cos\left(\frac{\pi}{2} - \frac{\pi}{180}\,\zeta\right). \tag{18.2}$$

(vgl. Section 5.6.4). Die Breite eines zu transformierenden Punktes liegt bei $\zeta_0 = 53.1$, wobei der exakte Wert auf eine Nachkommastelle gerundet wurde. Wie groß ist der durch das Runden zu erwartende absolute Fehler in der z-Koordinate? (Lösung: rund 3.34 km)

19 Übung 10 (Integralrechnung)

19.1 Stammfunktionen

19.1.1 Aufgabe 10.1 (KTA)

Finden Sie alle (!) Stammfunktionen zu folgenden Funktionen.

- **(a)** $f(x) = 3\,x^4 - 4\,x^2 + 2$,
- **(b)** $f(x) = \sqrt{x},\ x > 0$,
- **(c)** $f(x) = \frac{1}{x^2},\ x > 0$,
- **(d)** $f(x) = 5\,\sin x - \cos x$,
- **(e)** $f(x) = \mathrm{e}^{2\,x}$.

19.1.2 Aufgabe 10.2 (KTA)

Für welche $c \in \mathbb{R}$ konvergiert das uneigentliche Integral

$$\int_1^\infty x^c \,\mathrm{d}x? \tag{19.1}$$

19.1.3 Aufgabe 10.3 (KTA)

Geben Sie eine Stammfunktion zu

$$f(x) = \frac{1}{(x^2+1)^3} \tag{19.2}$$

an. Nutzen Sie dazu eine Formelsammlung Ihrer Wahl, z.B. die Tabelle von Ableitungs- und Stammfunktionen bei Wikipedia.

19.1.4 Aufgabe 10.4 (KTA)

Finden Sie eine Stammfunktion zu

$$f(x) = x^2 + \frac{1}{x}, \quad x > 0, \tag{19.3}$$

die durch den Punkt $(1, 2)$ verläuft.

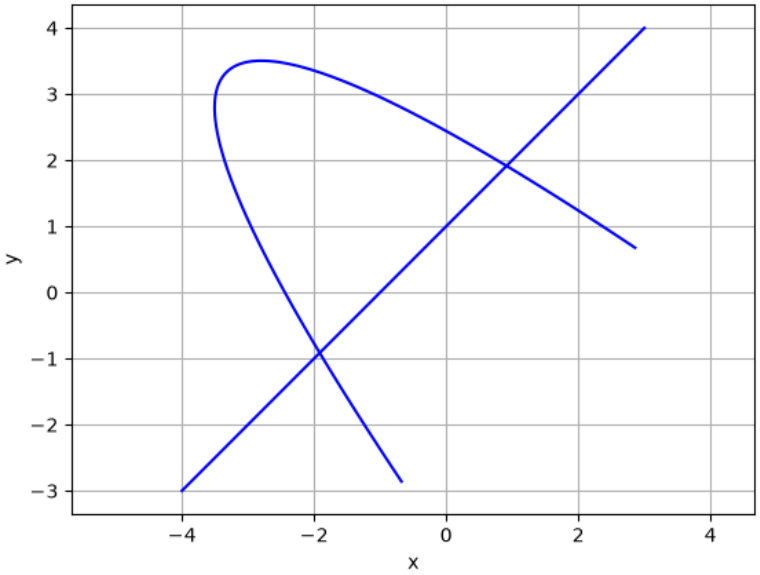
19.2 Flächeninhalte

19.2.1 Aufgabe 10.5 (KTA)

Die durch

$$(x+y)^2 + \sqrt{2}\,(y-x) - 8 - \sqrt{2} = 0 \tag{19.4}$$

beschriebene Kurve in der Ebene begrenzt gemeinsam mit der Gerade $y = x + 1$ eine Fläche.



- **(a)** Überzeugen Sie sich, dass die Kurve symmetrisch zur Gerade $y = -x$ verläuft, indem Sie die Kurve an dieser Gerade spiegeln (Koordinatentransformation!) und die Formel in den neuen Koordinaten mit der ursprünglichen Formel vergleichen.
- **(b)** Drehen Sie die Fläche um 45° im Uhrzeigersinn (Koordinatentransformation!). Die Formeln zur Beschreibung der Fläche sollten nun einfacher sein.
- **(c)** Beschreiben Sie die Fläche als Normalbereich und berechnen Sie den Flächeninhalt. (Lösung: rund 4.5523)

19.3 Aufgabe 10.6

Berechnen Sie die Fläche zwischen dem Graph der Funktion

$$f : [0,2] \to \mathbb{R}, \quad f(x) := \sqrt{x} + x^2 \tag{19.5}$$

und der x-Achse auf folgende Arten und vergleichen sie die Ergebnisse.

- **(a)** manuelles Rechnen mittels Hauptsatz der Differential- und Integralrechnung,

- **(b)** symbolisches Rechnen mit SymPy,
- **(c)** numerisches Rechnen SciPy.

```
# Ihre Lösung
```

20 Übung 11 (Wahrscheinlichkeit, Zufallsgrößen)

20.1 Ereignisse und Ereignisfelder

20.1.1 Aufgabe 11.1 (KTA)

Bilden Sie zu folgenden Ergebnismengen Ω und Teilmengen A und B die Mengen $\overline{A}$, $\overline{B}$, $A \cap B$, $A \cup B$, $\overline{A \cup B}$, $\overline{A} \cap \overline{B}$, $B \cap \overline{A}$.

- **(a)** $\Omega = \{1, 2, ..., 10\}, \quad A = \{2, 3, 5, 7\}, \quad B = \{2, 4, 6, 8, 10\}$,
- **(b)** $\Omega = [-1, 3], \quad A = [0, 1), \quad B = (\frac{1}{2}, 2]$,
- **(c)** $\Omega = \mathbb{R}, \quad A = \{x \in \mathbb{R} : |x - 1| < 3\}, \quad B = [0, \infty)$.

20.1.2 Aufgabe 11.2 (KTA)

Die Arbeit eines Kraftwerkes wird durch drei unabhängig voneinander arbeitende Kontrollsysteme überwacht, die jedoch auch einer gewissen Störanfälligkeit unterliegen. Es bezeichne S_i das Ereignis, dass das i-te System störungsfrei arbeitet ($i \in \{1, 2, 3\}$).

- **(a)** Drücken Sie folgende Ereignisse mit Hilfe der Ereignisse S_1, S_2 und S_3 aus.
  - $A :=$ “Alle drei Systeme arbeiten störungsfrei”
  - $B :=$ “Kein System arbeitet störungsfrei”
  - $C :=$ “Mindestens ein System arbeitet störungsfrei”
  - $D :=$ “Genau ein System arbeitet störungsfrei”
  - $E :=$ “Höchstens zwei Systeme sind gestört”
- **(b)** Wie viele Elementarereignisse umfasst der Ereignisraum?
- **(c)** Welche der unter (a) genannten Ereignisse können als Elementarereignisse betrachtet werden?
- **(d)** Aus wie vielen Elementarereignissen bestehen die Ereignisse C und D?

20.1.3 Aufgabe 11.3 (KTA)

Gegen Sie eine Ergebnismenge und ein hinreichend umfangreiches Ereignisfeld für das Werfen einer Münze an.

Sind weitere Ereignisfelder denkbar, egal ob praktisch relevant oder nicht?

20.2 Wahrscheinlichkeitsmaße

20.2.1 Aufgabe 11.4 (KTA)

Gegeben sind der Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und die Ereignisse $A, B \in \mathcal{A}$. Zeigen Sie

$$A \subseteq B \quad \Rightarrow \quad P(A) \leq P(B). \tag{20.1}$$

20.2.2 Aufgabe 11.5 (KTA)

Wir betrachten des Experiment "Werfen eines Würfels und einer Münze".

- **(a)** Geben Sie eine geeignete Ergebnismenge Ω an.
- **(b)** Zeigt die Münze Wappen, so wird die doppelte Augenzahl des Würfels notiert, bei Zahl nur die einfache. Wie groß ist die Wahrscheinlichkeit, dass eine gerade Zahl notiert wird? (Lösung: 0.75)

20.3 Diskrete Zufallsgrößen

20.3.1 Aufgabe 11.6 (KTA)

Eine Werte einer diskreten Zufallsgröße X besitzen die folgenden Wahrscheinlichkeiten:

k	P(X=k)
0	0.5
1	0.2
2	0.1
3	c

- **(a)** Bestimmen Sie den Wert der Konstanten c.
- **(b)** Ermitteln Sie die Verteilungsfunktion, den Erwartungswert, die Varianz und die Standardabweichung.
- **(c)** Skizzieren Sie die Verteilungsfunktion.

20.3.2 Aufgabe 11.7 (KTA)

Es werden zwei unterscheidbare Würfel geworfen. Die Zufallsgrößen X_1 und X_2 geben die geworfene Augenzahl des entsprechenden Würfels an. Die Zufallsgröße $Y := X_1 + X_2$ beschreibt die Augensumme beider Würfel.

- **(a)** Berechnen Sie die Wahrscheinlichkeiten $P(Y = k)$ für alle möglichen Augensummen k und ermitteln Sie die Verteilungsfunktion von Y. Skizzieren Sie die Funktion $k \mapsto P(Y = k)$ und die Verteilungsfunktion.
- **(b)** Berechnen Sie Erwartungswert und Varianz von Y. (Lösung: $\mathbb{E}Y = 7$, $\mathbb{V}\mathrm{ar}\, Y \approx 21.39$)
- **(c)** Berechnen Sie mit Hilfe der Verteilungsfunktion

$$P(Y < 5), \qquad P(Y > 8), \qquad P(4 \leq Y < 8). \tag{20.2}$$

20.4 Kontinuierliche Zufallsgrößen

20.4.1 Aufgabe 11.8 (KTA)

Gegeben sei die Funktion

$$p(x) = \begin{cases} 0, & \text{für } x < 1, \\ c\,x, & \text{für } 1 \leq x \leq 2, \\ 0, & \text{für } x > 2. \end{cases} \tag{20.3}$$

- **(a)** Bestimmen Sie den Wert der Konstanten c so, dass $p(x)$ Dichtefunktion einer stetigen Zufallsgröße ist.
- **(b)** Ermitteln Sie die Verteilungsfunktion, den Erwartungswert, die Varianz und die Standardabweichung. Bei der Varianz genügt die Angabe des auszuwertenden Integrals. Die Auswertung selbst kann mit SymPy erfolgen. (Teillösung: $\mathbb{E}X \approx 1.56$, $\mathbb{V}\mathrm{ar}\, Y \approx 0.08$)

20.4.2 Aufgabe 11.9 (KTA)

Die Zufallsgröße X besitze die folgende Verteilungsfunktion:

$$F(x) = \begin{cases} 0, & \text{für } x \leq 0, \\ \frac{1}{2}\,x^2, & \text{für } 0 < x \leq \sqrt{2}, \\ 1, & \text{für } x > \sqrt{2}. \end{cases} \tag{20.4}$$

- **(a)** Berechnen Sie die zugehörige Dichtefunktion.
- **(b)** Berechnen Sie den Erwartungswert von X. (Lösung: rund 0.94)

21 Übung 12 (Wahrscheinlichkeitsverteilungen)

21.1 Diskrete Verteilungen

21.1.1 Aufgabe 12.1 (KTA)

Eine Zufallsgröße X mit dem Wertebereich $\{0, 1\}$ und den Wahrscheinlichkeiten $P(X = 1) = p$ und $P(X = 0) = 1 - p$ heißt Bernoulli-verteilt mit dem Parameter $p \in (0, 1)$.

- **(a)** Geben Sie die Verteilungsfunktion von X an.
- **(b)** Berechnen Sie den Erwartungswert und die Varianz von X.
- **(c)** Gegen Sie ein Beispiel für ein Zufallsexperiment an, dessen Ergebnisse Bernoulli-verteilt sind.

21.1.2 Aufgabe 12.2 (teilweise KTA)

In einem Test sind 10 Multiple-Choice-Fragen zu beantworten. Jede Frage besitzt fünf Antwortmöglichkeiten, von denen jeweils genau eine richtig ist. Ein Student überlegt, ob es sich lohnt bei allen Fragen einfach zu raten.

- **(a)** Welche Verteilung hat die Zufallsvariable Y, die die Anzahl der richtigen Antworten angibt, wenn der Student nur rät?
- **(b)** Wie viele Fragen wird der Student im Mittel richtig beantworten, wenn er rät?
- **(c)** Der Test ist bestanden, wenn mindestens fünf Fragen richtig beantwortet wurden. Wie groß ist die Wahrscheinlichkeit, den Test durch Raten zu bestehen? Gegen Sie eine entsprechende Formel an und nutzen Sie den Computer zu deren Auswertung. (Lösung: rund 0.03)
- **(d)** Wo müsste die Grenze für die Anzahl der zum Bestehen notwendigen richtigen Antworten liegen, damit die Chance zum Bestehen durch Raten über 30% liegt? Formulieren Sie eine entsprechende Bedingung an die gesuchte Zahl als Ungleichung und lösen Sie die Ungleichung mit dem Computer.

21.2 Kontinuierliche Verteilungen

21.2.1 Aufgabe 12.3 (teilweise KTA)

Die Reparaturdauer einer defekten Maschine kann als exponentialverteilt angenommen werden. Der Erwartungswert liege bei 4 Stunden.

- **(a)** Recherchieren Sie die zugehörige Verteilungsfunktion.
- **(b)** Wie hoch ist die Wahrscheinlichkeit, dass die Reparaturzeit 2 Stunden nicht übersteigt.

- **(c)** Wie hoch ist die Wahrscheinlichkeit, dass die Reparatur länger als 10 Stunden dauert?

21.2.2 Aufgabe 12.4 (KTA)

Die Lebensdauer einer Glühlampe kann als exponentialverteilte Zufallsgröße angenommen werden (vgl. vohergehende Aufgabe). Es sei bekannt, dass 75 Prozent aller Glühlampen eine Mindestleuchtdauer von 500 Stunden erreichen. Berechnen Sie den Erwartungswert der Lebensdauer. (Lösung: rund 1738)

21.2.3 Aufgabe 12.5 (KTA)

Formulieren Sie die folgenden Wahrscheinlichkeit zu einer standardnormalverteilten Zufallsgröße X so um, dass zur Berechnung nur Auswertungen der Verteilungsfunktion der Standardnormalverteilung benötigt werden.

- **(a)** $P(X > 2)$,
- **(b)** $P(|X| < 1.5)$.

21.2.4 Aufgabe 12.6 (KTA)

Formulieren Sie die folgenden Wahrscheinlichkeit zu einer mit Parametern $\mu = 1$ und $\sigma = 3$ normalverteilten Zufallsgröße X so um, dass zur Berechnung nur Auswertungen der Verteilungsfunktion der Standardnormalverteilung benötigt werden.

- **(a)** $P(-2 \leq X < 7)$,
- **(b)** $P(-5 \leq X \leq -2)$,
- **(c)** $P(X > 0)$.

21.2.5 Aufgabe 12.7 (KTA)

Die Länge einer Strecke soll zu Übungszwecken durch Studierende in drei Abschnitten mit jeweils unterschiedlicher Technik bestimmt werden. Die tatsächlichen Längen sind 120, 55 und 20 Meter. Die drei Messungen können als Realisierungen von drei normalverteilten Zufallsgrößen X, Y, Z angesehen werden. Die zugehörigen Varianzen können aus den Unterlagen der verwendeten Messgeräte entnommen werden. Diese sind

$$\sigma_X^2 = 4, \quad \sigma_Y^2 = 3, \quad \sigma_Z^2 = 2. \tag{21.1}$$

Wie groß ist die Wahrscheinlichkeit, dass die Messungen eine Gesamtstrecke von mehr als 200 Meter ergeben? (Lösung: rund 0.29)

22 Übung 13 (Zufallsvektoren)

22.1 Aufgabe 13.1 (KTA)

Ein kontinuierlicher Zufallsvektor (X, Y) habe die Wahrscheinlichkeitsdichte

$$p(x,y) := \begin{cases} 4\,x\,y - 2\,(x+y) + 2, & \text{für } (x,y) \in [0,1] \times [0,1], \\ 0, & \text{sonst} \end{cases} \tag{22.1}$$

(vgl. Beispiel (gleiche Randverteilungen)).

- **(a)** Berechnen Sie die Randdichten der Komponenten X und Y.
- **(b)** Berechnen Sie Erwartungswert und Varianz von X und Y sowie $\mathbb{C}\mathrm{ov}(X, Y)$. Nutzen Sie ggf. SymPy zur Auswertung von Integralen.
- **(c)** Bestimmen Sie den Korrelationskoeffizienten. Sind X und Y unabhängig?

22.2 Aufgabe 13.2 (KTA)

Sei X diskret gleichverteilt auf $\{-1, 0, 1\}$ und sei $Y := 1 - X^2$. Zeigen Sie, dass X und Y zwar unkorreliert, aber nicht unabhängig sind.

22.3 Aufgabe 13.3

Der Hersteller eines GPS-Empfängers gibt an, dass der Fehler in den angezeigten Koordinaten (in hier nicht näher spezifizierten Einheiten) als normalverteilt mit den Parametern

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad S = \begin{bmatrix} 0.15 & 0.05 \\ 0.05 & 0.02 \end{bmatrix} \tag{22.2}$$

angenommen werden kann.

- **(a)** Sind die Fehler in den beiden Koordinaten unabhängig?
- **(b)** Stellen Sie Dichtefunktion und Verteilungsfunktion am Computer grafisch dar. Nutzen Sie dazu z.B. `scipy.stats.multivariate_normal.pdf` (Dichtefunktion) und `scipy.stats.multivariate_normal.cdf` (Verteilungsfunktion).

```
# Ihre Lösung
```

- **(c)** Wie groß ist die Wahrscheinlichkeit, dass die Abweichung zwischen den angezeigten Koordinaten und den tatsächlichen Koordinaten in jeder Koordinate höchstens 0.1 beträgt? Gehen Sie zur Klärung dieser Frage wiefolgt vor:
  - Skizzieren Sie den relevanten Teil des Wertebereichs der Zufallsgröße.

- Formulieren Sie dann die Lösung zunächst so, dass nur Auswertungen der Verteilungsfunktion der gegebenen Normalverteilung nötig sind.
- Werten Sie die gefundene Formel anschließend mit dem Computer aus.

```
# Ihre Lösung
```

23 Übung 14 (deskriptive Statistik)

In dieser Übung werden wir ein mittels GPS-Tracker aufgezeichnetes Höhenprofil untersuchen. Die Höhendaten wurden während einer Fahrradtour im Abstand von je zwei Sekunden aufgezeichnet.

23.1 Aufgabe 14.1

Lesen Sie die Datei track.txt mit dem unten gegebenen Programmcode ein. Die Datei enthält eine Liste der Höhenmessungen. Jede Zeile enthält genau einen Messwert.

```
path = ''  # hier Pfad zur runtergeladenen Datei eintragen

# leer Liste für Aufnahme der Messwerte
x = []

# Datei öffnen
with open(path) as file:

    # Zeile für Zeile aus Datei lesen
    for row in file.readlines():

        # Zeichenkette in Zahlen umwandeln und an Liste anhängen
        x.append(float(row))

# Liste in NumPy -Array umwandeln
x = np.array(x)
```

Gegeben Sie die Anzahl der Messwerte sowie den kleinsten und den größten Wert aus.

```
# Ihre Lösung
```

23.2 Aufgabe 14.2

Stellen Sie den Höhenverlauf in Abhängigkeit von der Zeit grafisch dar.

```
# Ihre Lösung
```

23.3 Aufgabe 14.3

Berechnen Sie den Median der Höhenwerte. Welche Aussage liefert dieser mit Blick auf den vorliegenden Sachverhalt?

```
# Ihre Lösung
```

23.4 Aufgabe 14.4

Erzeugen Sie ein Histogramm der Höhendaten mit Klassengrenzen 330, 340, 350, 360, 370.

```
# Ihre Lösung
```

24 Übung 15 (induktive Statistik)

24.1 Aufgabe 15.1 (KTA)

Aus einer Lieferung von Kondensatoren mit einer Normkapazität von 220 µF wurden 10 Stück zufällig ausgewählt und folgende Kapazitäten (in µF) festgestellt:

$$217, \quad 221, \quad 220, \quad 221, \quad 219, \quad 221, \quad 220, \quad 218, \quad 222, \quad 219. \tag{24.1}$$

- **(a)** Welche Verteilungsannahme für die Kapazität erscheint sinnvoll? Welche Verteilungsparameter sind bekannt, welche unbekannt?
- **(b)** Berechnen Sie je eine Schätzung für den Erwartungswert und die Varianz der Kapazität.
- **(c)** Berechnen Sie auf Basis der vorliegenden Stichprobe je ein zweiseitiges Konfidenzintervall für den Mittelwert und die Varianz der Kapazität zum Konfidenzniveau 0.95. Nutzen Sie dazu die Formel

$$\left[\overline{X} - \frac{S}{\sqrt{n}}\, t_{1-\frac{\alpha}{2},n-1},\ \overline{X} + \frac{S}{\sqrt{n}}\, t_{1-\frac{\alpha}{2},n-1}\right] \tag{24.2}$$

  mit Stichprobenmittelwert $\overline{X}$ und Stichprobenvarianz S^2. Eventuell nützliche Quantile $t_{a,b}$ der Studentschen t-Verteilung:

	$b = 8$	$b = 9$	$b = 10$	$b = 11$
$a = 0.950$	1.8595	1.8331	1.8125	1.7959
$a = 0.975$	2.3060	2.2622	2.2281	2.2010
$a = 0.990$	2.8965	2.8214	2.7638	2.7181
$a = 0.995$	3.3554	3.2498	3.1693	3.1058

  (Lösung: rund $[218.75, 220.85]$)
- **(d)** Was sagt das Konfidenzintervall über die mittlere Kapazität aller gelieferten Kondensatoren aus?

24.2 Aufgabe 15.2

Aus einer großen Lieferung von Bauteilen werden 80 Stück auf korrekte Funktion geprüft. Dabei erwiesen sich zwei Bauteile als defekt.

- **(a)** Welche Verteilung kann für die Anzahl der defekten Bauteile angenommen werden? Welche Parameter enthält diese?
- **(b)** Prüfen Sie bei einer Irrtumswahrscheinlichkeit von 5 Prozent die Hypothese, dass der Ausschussanteil höchstens 4 Prozent beträgt. Suchen Sie einen dafür passenden Test und setzen Sie diesen um.

```
# Ihre Lösung
```

24.3 Aufgabe 15.3

In der Richtlinie 76/211/EWG des Rates vom 20. Januar 1976 zur Angleichung der Rechtsvorschriften der Mitgliedstaaten über die Abfüllung bestimmter Erzeugnisse nach Gewicht oder Volumen in Fertigpackungen wird in Abschnitt 2.3 von Anhang II ein Verfahren zur "Prüfung des Mittelwerts der tatsächlichen Füllmengen eines Loses von Fertigpackungen" angegeben.

- **(a)** Vergleich Sie das Vorgehen dort mit dem Vorgehen bei Hypothesentests. Gibt es Abweichungung zu unserem Vorgehen?
- **(b)** Geben Sie die in der Richtlinie verwendete Hypothese an.
- **(c)** Prüfen Sie die in 2.3.3 angebenen Zahlen für Annahme und Ablehnung des Loses.

```
# Ihre Lösung
```

25 Download

Zum Offline-Lesen steht das Vorlesungsskript hier als PDF-Datei zur Verfügung. Da dieses Vorlesungskript dem **Online-First-Prinzip** folgt, wird die PDF-Datei nur nachrangig gepflegt. Die Inhalte sind zwar stets identisch mit den Online-Inhalten, allerdings gibt es aus technischen Gründen einige eher formale Abweichungen:

- Animationen sind nicht enthalten, sondern fehlen ohne ersichtlichen Hinweis auf das Fehlen.
- Grafiken sind teils an anderen Stellen im Text als in der Online-Version (plus minus eine Seite).
- Einige Links, z.B. zu den Videos und den zugehörigen Notizen, funktionieren nicht.
- Links zum Glossar der Abkürzungen funktionieren nicht.
- Das Layout enthält diverse kleinere Fehler.

Download PDF-Datei

26 Organisatorisches

Modulbeschreibung in Modulux

26.1 Ablauf der Veranstaltung

Die Veranstaltung findet im Wesentlichen im **Selbststudium** statt. Zur Unterstützung des Selbststudiums wird es Konsultationseinheiten geben:

- zwei “große” Präsenzveranstaltungen in Dresden,
- eine “große” Online-Veranstaltung,
- sechs “kleine” Online-Veranstaltungen.

Die genauen Termine finden Sie unten im Section 26.4. Den Link zum BBB-Raum für die Online-Veranstaltungen finden Sie im OPAL-Kurs. Organisatorische Fragen können Sie jederzeit per E-Mail an den Vorlesenden richten.

Merke

Die Konsultationseinheiten dienen der Klärung von Fragen zu Stoff und Übungsaufgaben. In den Einheiten wird keine Vorlesung stattfinden, also insbesondere auch kein neuer Stoff vermittelt werden. Für eine effiziente und zielführende Nutzung der Zeit ist die **gute Vorbereitung Ihrerseits** essentiell!

Hinweise zur Vorbereitung auf die Konsultationseinheiten:

- Stoff und Übungsaufgaben entsprechend Arbeitsplan **vor (!) den Konsultationseinheiten vorbereiten**.
- **Kleinere Fragen** zu Stoff und Übungsaufgaben können Sie jederzeit gern im Forum des OPAL-Kurses stellen. Dort können sowohl der Vorlesende als auch Mitstudierende antworten. Per E-Mail bitte nur Fragen stellen, die für andere Studierende nicht von Interesse sind oder aus Datenschutzgründen nicht für andere Studierende sichtbar sein sollen.
- **Umfangreichere Fragen** zu Stoff und Übungsaufgaben bitte möglichst schon vor der nächsten Konsultationseinheit per E-Mail an den Vorlesenden schicken, damit dieser den Ablauf der Konsultationseinheit entsprechend den zu klärenden Fragen strukturieren kann.
- Bis 72 Stunden vor der nächsten Konsultationseinheit können Sie gern Ihre **Lösungen und Lösungswege per E-Mail** an den Vorlesenden schicken (nur PDF-Dateien und Jupyter-Notebooks). Dieser wird Ihre Einsendungen korrigieren und mit Kommentaren versehen. Später eingegangene Lösungen oder Lösungen zu Übungsaufgaben, die laut Arbeitsplan schon zu einer früheren Konsultationsein-

heit hätten bearbeitet sein sollen, werden nicht korrigiert! Auch Zusendungen von JPEG-, PNG- und anderen Dateien, die nicht PDF-Dateien und Jupyter-Notebooks sind, werden ignoriert. Dieses Angebot können Sie ganz ungezwungen auf freiwilliger Basis nutzen (oder auch nicht); für die Modulnote sind aussschließlich die Klausurergebnisse relevant.

Einige Übungsaufgaben sind mit dem Computer zu bearbeiten. Dazu ist keine spezielle Software nötig, sondern nur ein Webbrowser. Genaueres klären wir in Übung 2.

Merke

Für den erfolgreichen Abschluss dieses Moduls bei durchschnittlicher Begabung ist eine Arbeitszeit von ca. **10 Stunden pro Woche** (bei 15 Vorlesungswochen) und zusätzlich ca. 30 Stunden vor der Prüfung anzusetzen (vgl. Modulbeschreibung). Etwa die Hälfte dieser Zeit sollte in die Erarbeitung neuen Stoffes fließen, die andere Hälfte in die Bearbeitung von Übungsaufgaben.

Behalten Sie bitte stets im Hinterkopf: Die Tätigkeit, für die Sie sich entschieden haben, heißt **"Studieren", nicht "studiert werden"**. Das ist also eine Tätigkeit, die Sie (!) aktiv ausführen. Der Vorlesende wird Sie dabei so gut wie möglich unterstützen, kann aber keine Wunder vollbringen, wenn es an Eigeninitiative und Fleiß fehlt.

26.2 Hinweise zum Skript

Das Skript enthält den gesamten für die Prüfung relevanten Stoff. Wird auf externe Inhalte verlinkt, so sind diese Inhalte nicht relevant für die Prüfung. Abschnitte des Skripts, die nicht für die Prüfung relevant sind, sind entsprechend gekennzeichnet.

Das Skript steht auch als PDF-Datei zum Download zur Verfügung. Falls Sie **kapitelweise PDF-Dateien** nutzen möchten (z.B. zum Anbringen von Notizen), können Sie die Druckfunktion Ihres Webbrowsers nutzen. Firefox schneidet allerdings Textzeilen an den Seitenumbrüchen ab; bei Chromium gibt es dieses Problem nicht.

Diverse Rechnungen und Herleitungen fehlen im Skript und werden jeweils als **Video** zur Verfügung gestellt, das an entsprechender Stelle verlinkt ist.

Die Übungsaufgaben stehen zusätzlich als PDF-Aufgabenblatt und als **Jupyter-Notebook** zur Verfügung; siehe Button oben rechts bei jeder Übung. Die Jupyter-Notebooks erlauben das Lösen der Aufgaben direkt im Dokument (vgl. Übung 2).

Merke

Das Skript enthält viel Text. **Einmaliges Lesen ist nicht genug.** Lesen Sie einen Abschnitt zunächst vollständig, damit Sie einen Überblick über das aktuelle

Thema bekommen. Beginnen Sie dann nochmals von vorn und durchdenken Sie jeden Satz und jede Formel! Wenn Verständnisprobleme auftreten: weiter darüber nachdenken! Wenn das nicht hilft: Frage im Forum stellen.

Die mathematische Sprache (in Text und Formeln) kann komplexe Sachverhalte sehr komprimiert darstellen. Da kann es schonmal vorkommen, dass man an einem unscheinbaren Dreizeiler eine halbe Stunde oder mehr zu "knabbern" hat. Anders ausgedrückt: **Eine Mathebuch ist kein Rosamunde-Pilcher-Roman!**

26.3 Prüfung

Die Prüfung findet als 90-minütige Klausur in der ersten Präsenzeinheit des zweiten Semesters statt (Mitte April 2027).

Übungsaufgaben, die von Typ und Schwierigkeit den Klausuraufgaben entsprechen, sind mit "KTA" (= klausurtypische Aufgabe) gekennzeichnet.

Beispielklausur (dient nur zur groben Orientierung; grundsätzlich können alle behandelten Themen und Aufgabentypen Teil der Klausur sein)

Zugelassene Hilfsmittel zur Klausur: Papier, Stifte (kein Rot), einfacher Taschenrechner, also ohne CAS und ohne Grafik (die Aufgaben sind auch ohne Taschenrechner gut lösbar).

26.4 Arbeitsplan

Hier finden Sie alle Konsultationstermine und die bis zum jeweiligen Termin zu bearbeitenden Inhalte.

- **13.10.2026, 18:30-20:00, online**:
  - Schulmathematik
  - Motivation
  - Übung 1 (Grundlegendes)
  - Computereinsatz
  - Übung 2 (Python und Jupyter)
  - Elementare Begriffe: Aussagen
  - Elementare Begriffe: Abbildungen
- **26.10.2026, 18:30-20:00, online**:
  - Elementare Begriffe: Zahlenfolgen
  - Übung 3 (Aussagen, Abbildungen, Folgen)

  - Elementare Begriffe: Funktionen
  - Übung 4 (Funktionen)

- **12.11.2026, 15:30-17:00 und 13.11.2026, 13:45-17:00 in Dresden**:
  - Vektoren und Matrizen: Formalismus
  - Vektoren und Matrizen: Vektorgeometrie
  - Übung 5 (Matrizen, Vektorgeometrie)
  - Vektoren und Matrizen: Lineare Abbildungen
  - Vektoren und Matrizen: Lineare Unterräume und Untermannigfaltigkeiten
  - Übung 6 (lineare Abbildungen, Basen)
- **24.11.2026, 18:30-20:00, online**:
  - Vektoren und Matrizen: Lineare Gleichungssysteme
  - Übung 7 (lineare Gleichungssysteme)
  - Vektoren und Matrizen: Koordinatentransformationen
  - Übung 8 (Koordinantentransformationen)
  - Vektoren und Matrizen: Ausblick: Eigenwerte und Eigenvektoren
- **09.12.2026, 18:30-20:00, online**:
  - Differentialrechnung
  - Übung 9 (Differentialrechnung)
  - Integralrechnung
  - Übung 10 (Integralrechnung)
- **19.12.2026, 11:30-15:15, online**:
  - Wiederholen Sie bitte allen bisherigen Stoff und sammeln Sie Unklarheiten und Fragen.
  - Beschäftigen Sie sich ggf. schon mit dem nächsten Abschnitt im Arbeitsplan, da dieser recht umfangreich ist.
- **06.01.2027, 18:30-20:00, online**:
  - Wahrscheinlichkeit: Wahrscheinlichkeitsräume
  - Wahrscheinlichkeit: Zufallsgrößen
  - Übung 11 (Wahrscheinlichkeit, Zufallsgrößen)
- **18.01.2027, 18:30-20:00, online**:

  - Wahrscheinlichkeit: Diskrete Verteilungen
  - Wahrscheinlichkeit: Kontinuierliche Verteilungen
  - Übung 12 (Wahrscheinlichkeitsverteilungen)
  - Wahrscheinlichkeit: Unabhängigkeit
  - Wahrscheinlichkeit: Grenzwertsätze (nicht prüfungsrelevant)
  - Wahrscheinlichkeit: Zufallsvektoren
  - Übung 13 (Zufallsvektoren)
- **04.02.2027, 15:30-17:00 und 06.02.2027, 8:00-12:15 in Dresden**:
  - Statistik
  - Übung 14 (deskriptive Statistik)
  - Übung 15 (induktive Statistik)

26.5 Änderungshistorie

Da dieses Skript für das Wintersemester 2026/2027 vollständig neu strukturiert und formuliert wurde, werden im Laufe des Semesters sehr wahrscheinlich Korrekturen und Ergänzungen nötig werden. Der besseren Nachvollziehbarkeit halber werden die inhaltlich relevanten Anpassungen nach Semesterbeginn hier dokumentiert:

- 15.11.2026:
  - Fehler in Aufgabe 7.2 korrigiert: $u_n = U_\mathrm{a}$ statt $u_2 = U_\mathrm{a}$.
- 09.11.2026:
  - Fehler in Formel (35) in Abschnitt 5.1.7 korrigiert.
- 22.10.2026:
  - Notebook-Download der Übungsaufgaben repariert.
- 15.10.2026:
  - Notebook-Download der Übungsaufgaben repariert.
- 10.10.2026:
  - Lösungen in Übungen 11, 12, 15 hinzugefügt.
  - Jupyter-Notebooks werden neben PDF-Dateien bei der Abgabe von Übungsaufgaben akzeptiert.
  - Beispielklausur hinzugefügt.
- 08.10.2026:

  - Tippfehler in Übungsaufgaben 1.1 und 15.1 korrigiert.

- 07.10.2026:
  - Riemann-Integrale für stückweise stetige Funktionen hinzugefügt; weitere kleine Anpassungen im Unterkapitel Normalbereiche.
  - Kapitel zur Wahrscheinlichkeitsrechnung hinzugefügt.
  - Kapitel zur Statistik hinzugefügt.
  - Termine für Konsultationen präzisiert.
- 16.09.2026:
  - Kapitel zu Differential- und Integralrechnung hinzugefügt.
  - Ergebnis in Beispiel bei Zahlenfolgen ergänzt.
  - Abschnitt Section 4.3.6 eingefügt.
  - Prüfungstermin korrigiert.
- 28.08.2026:
  - Link zu Video 130 repariert.
  - Kapitel VI von OMB+ in die Liste der als aus der Schulmathematik bekannt vorausgesetzten Themen verschoben.
  - Nummerierung der Aufgaben in Übung 1 und Übung 2 korrigiert.
  - Nebenbemerkung zu Symbolen für logische Verknüpfungen eingefügt.
  - Unterscheidung zwischen Periode und Grundperiode bei Zahlenfolgen eingefügt.
  - Unterscheidung zwischen Section 4.4.3.2 bei Funktionen eingefügt.
  - Diverse sprachliche Korrekturen.