

Skript zur Vorlesung

Höhere Mathematik III

für die Fachrichtungen MACH, MIT, MWT, BIW, CIW, VT

25. Auflage

PD Dr. Frank Hettlich
und
Prof. Dr. Andreas Kirsch

Institut für Angewandte und Numerische Mathematik
Karlsruher Institut für Technologie

Literatur für diese Vorlesung (u.a.)

- T. Arens, F. Hettlich, C. Karpfinger, U. Kockelkorn, K. Lichtenegger, H. Stachel: Mathematik. Springer Spektrum, Heidelberg.
- K. Burg, H. Haf, F. Wille: Höhere Mathematik für Ingenieure IV und V. Teubner Verlag, Stuttgart.

Für die Vektoranalysis scheint uns auch folgende Kurzdarstellung geeignet:

- D.E. Bourne, P.C. Kendall: Vektoranalysis. Teubner Studienbuch.

Formelsammlungen:

- T. Arens, F. Hettlich, Ch. Karpfinger, U. Kockelkorn, K. Lichtenegger, H. Stachel: Mathematik zum Mitnehmen. Springer Spektrum, Heidelberg.
- G. Merziger, G. Mühlbach, D. Wille, T. Wirth: Formeln und Hilfen zur Höheren Mathematik. Binomi Verlag.
- I.N. Bronstein, K.A. Semendjajew: Taschenbuch der Mathematik. Teubner Verlag, Stuttgart, n -te Auflage, $n \geq 25$.

Inhaltsverzeichnis

1 Funktionen und Differentialrechnung im \mathbb{R}^n	3
1.1 Funktionen und Stetigkeit im \mathbb{R}^n	3
1.2 Partielle Ableitungen und Jacobimatrix	10
1.3 Exakte Differentialgleichungen und implizit gegebene Funktionen	17
1.4 Extremwertaufgaben	21
1.5 Nichtlineare Gleichungssysteme	28
2 Das Gebietsintegral	37
2.1 Begriffsbildung und allgemeine Sätze	37
2.2 Das iterierte Integral	42
2.3 Die Transformationsformel	44
2.4 Übertragung auf dreidimensionale Gebietsintegrale	47
3 Vektoranalysis	50
3.1 Kurven und Kurvenintegrale	50
3.2 Flächen und Flächenintegrale	60
3.3 Die Divergenz und der Gaußsche Integralsatz	64
3.4 Die Rotation und der Stokessche Satz	73
4 Partielle Differentialgleichungen	80
4.1 Klassifizierungen	80
4.2 Quasilineare partielle Differentialgleichungen erster Ordnung	85
4.3 Separationsansätze	90
5 Einführung in die Stochastik	94
5.1 Grundbegriffe	94
5.2 Zufallsvariablen und Verteilungsfunktionen	105
5.3 Folgen von Zufallsvariablen	114
Übersicht: Parametrisierungen	119
Übersicht: Integralbegriffe	121

1 Funktionen und Differentialrechnung im \mathbb{R}^n

1.1 Funktionen und Stetigkeit im \mathbb{R}^n

Eine **Abbildung** zwischen einer Menge X und einer Menge Y ist eine Vorschrift $\alpha : X \rightarrow Y$, die jedem Element $x \in X$ genau ein $y \in Y$ zuordnet. Wenn $X \subseteq \mathbb{K}^n$ und $Y \subseteq \mathbb{K}^m$ gilt (mit $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$) sprechen wir auch von **Funktionen**, ein Begriff, der uns schon im ersten Semester begegnet ist. Erinnern wir uns an das Wichtigste über Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$: Wir sprechen von der **Funktion** f , als derjenigen Abbildung, die $x \in \mathbb{R}$ einen Funktionswert $y = f(x) \in \mathbb{R}$ zuordnet. Die Menge $\{(x, y) \in \mathbb{R}^2 : y = f(x)\} \subset \mathbb{R}^2$ heißt der **Graph** von f und wird oft gezeichnet. Zentrale Begriffe im Umgang mit Funktionen sind die **Stetigkeit**, d.h. $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ und die **Ableitung** $f'(\hat{x}) = \lim_{x \rightarrow \hat{x}} (f(x) - f(\hat{x})) / (x - \hat{x})$ mit der **Tangente** in $(\hat{x}, f(\hat{x}))$, also der Graph der **Linearisierung** $y(x) = f(\hat{x}) + f'(\hat{x})(x - \hat{x})$, $x \in \mathbb{R}$. Außerdem haben wir die Notation des **Differentials** kennengelernt, d.h. die Tangente $dy = f'(\hat{x}) dx$ im (dx, dy) -Koordinatensystem. Zusammen mit dem **Integrieren** ergeben sich aus diesen grundlegenden Konzepten vielfältige Möglichkeiten, wie etwa die Behandlung von Extremwertaufgaben. Wesentliches Ziel wird es von nun an sein, zu verstehen wie diese Begriffe sich auf vektorwertige Abbildungen in mehreren Variablen übertragen lassen.

Zunächst verallgemeinern wir den Begriff von einer Funktion.

Definition 1.1 (Funktion)

Sei $D \subseteq \mathbb{C}^n$ eine Teilmenge des \mathbb{C}^n . Eine **m -wertige Funktion von n Veränderlichen** ist eine Vorschrift, die jedem Vektor $x \in D \subseteq \mathbb{C}^n$ genau einen Vektor $y \in \mathbb{C}^m$ zuordnet. Wir schreiben $f : \mathbb{C}^n \supseteq D \rightarrow \mathbb{C}^m$ und $y = f(x)$ bzw. *ausgeschrieben*

$$f(x) = f(x_1, x_2, \dots, x_n) = \begin{pmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_m(x_1, x_2, \dots, x_n) \end{pmatrix}.$$

Die Menge D heißt weiterhin **Definitionsbereich** von f und die Menge $W := \{f(x) : x \in D\}$ aller Funktionswerte ist das **Bild** oder der **Wertebereich** von f .

Bemerkung: Entgegen unserer Konvention wird, wie im Satz angegeben, das Argument bei solchen Funktionen häufig als Zeilenvektor geschrieben, um Platz zu sparen.

Beispiele 1.2

- Sei $f : \mathbb{R} \supset \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^2$ definiert durch $f(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$, $t \in \mathbb{R}_{\geq 0}$. Der Wertebereich ist jetzt der Einheitskreis $\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$. Dies ist also eine **vektorwertige** Funktion einer Veränderlichen. Solche Abbildungen werden auch *Kurven* genannt. Begegnet sind wir solchen Funktionen schon, nämlich bei den Lösungen von Differentialgleichungssysteme in HM2. Wir werden uns noch genauer damit beschäftigen.
- Sei $f : \mathbb{R}^3 \setminus \{0\} \rightarrow \mathbb{R}$ erklärt durch $f(x) = 1/\|x\| = (x_1^2 + x_2^2 + x_3^2)^{-1/2}$ für $x = (x_1, x_2, x_3) \in \mathbb{R}^3 \setminus \{0\}$. Dann ist der Wertebereich gegeben durch $\mathbb{R}_{>0}$ und wir sprechen von einer **skalarwertigen** Funktion oder einem **skalaren** Feld.

- (c) In technischen Anwendungen treten neben den skalaren Feldern wie etwa der Massendichte $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$ auch **Vektorfelder** auf, wie etwa das Gravitationsfeld eines Massenpunkts im Ursprung, $f : \mathbb{R}^3 \setminus \{0\} \rightarrow \mathbb{R}^3$, das mit der Gravitationskonstanten g und der Masse m gegeben ist durch

$$f(x) = -\frac{gm}{\|x\|^3}x, \quad \text{d.h. ausführlicher} \quad f(x_1, x_2, x_3) = -\frac{gm}{(x_1^2 + x_2^2 + x_3^2)^{3/2}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Wir weisen darauf hin, dass wir hier in der Mathematik oft $x = (x_1, x_2, x_3)^\top$ schreiben und nicht, wie in technischen Fächern üblich, $r = (x, y, z)$. Weitere skalare Felder sind die Energiedichte, die Entropie und weitere Vektorfelder der Impuls, das Geschwindigkeitsfeld von Strömungen, elektrisches und magnetisches Feld.

- (d) Beim Differenzieren haben wir schon das generelle Konzept der *Linearisierung* kennengelernt. Dies ist natürlich auch nicht auf den eindimensionalen Fall beschränkt. Wir können erahnen, dass die Abbildungen $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^m$ der Form $y = \alpha(x) = Ax$ mit einer Matrix $A \in \mathbb{R}^{m \times n}$ eine extrem wichtige Klasse bilden. Diese wollen wir im übernächsten Abschnitt genauer analysieren. \square

Wie **veranschaulicht** man sich Funktionen? In diesem Zusammenhang ist bekannterweise der Graph einer Funktion wichtig.

Definition 1.3 (*Graph einer Funktion*)

Sei $f : \mathbb{C}^n \supset D \rightarrow \mathbb{C}^m$ eine Funktion. Dann heißt

$$\text{Graph}(f) := \{(x, f(x))^\top \in \mathbb{C}^{n+m} : x \in D\}$$

der **Graph** der Funktion. Er ist eine Teilmenge von \mathbb{C}^{n+m} .

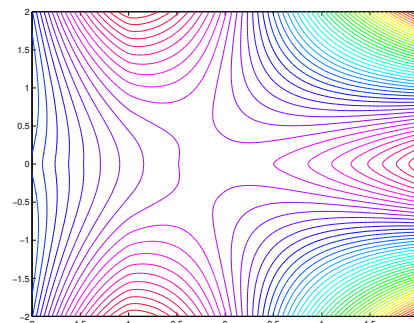
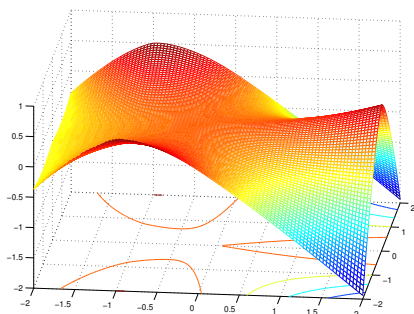
Für $n = m = 1$ und **reellwertige** Funktionen ist uns der Graph gut bekannt. Beachte, dass der Graph auch als Wertebereich einer Kurve $\varphi : D \subseteq \mathbb{R} \rightarrow \mathbb{R}^2$ mit $\varphi(x) = (x, f(x))^\top$ angesehen werden kann.

Für $n = 2$ und $m = 1$ und einer **skalaren, reellwertigen** Funktion $f : \mathbb{R}^2 \supset D \rightarrow \mathbb{R}$ in zwei reellen Variablen ist $\text{Graph}(f) \subseteq \mathbb{R}^3$. In der (x_1, x_2) -Ebene skizziert wir D , darüber „schwebt“ der Graph als gekrümmte Fläche. Eine andere Möglichkeit ist, die **Konturlinien** oder Höhenlinien von f darzustellen – wie bei einer Landkarte. Es werden also die Punkte gleicher Höhe verbunden.

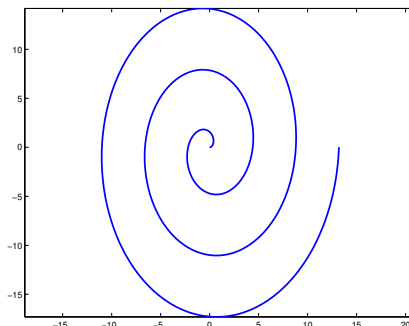
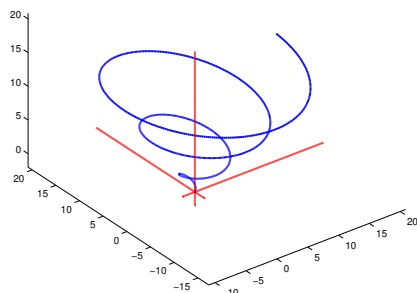
Im folgenden Beispiel ist die Funktion $f(x, y) = 0.1 [x|x| + 1.5y^2 - 2|(x+1)y|^{3/2}]$ im Quadrat

$$[-2, 2]^2 = [-2, 2] \times [-2, 2] = \{(x, y)^\top : x \in [-2, 2], y \in [-2, 2]\}$$

aufgetragen: links der Graph, rechts der Konturplot.



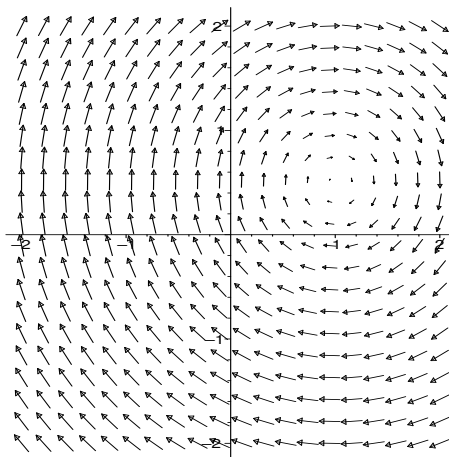
Bei Kurven, d.h. $n = 1$ und etwa $m = 2$ und einer **reellen** Funktion $f : \mathbb{R} \supset D \rightarrow \mathbb{R}^2$ zeichnen wir üblicherweise nicht den Graphen sondern nur den Wertebereich $W = \{f(t) : t \in D\}$ als Teilmenge der Ebene \mathbb{R}^2 . Die Variable t lässt sich als Zeit interpretieren in der die Kurve durchlaufen wird. Zur Verdeutlichung lassen sich einige Punkte in W markieren, die zu gewissen Zeiten erreicht werden. Im folgenden Beispiel ist die Kurve $f(t) = \begin{pmatrix} 0.7t \cos(t) \\ t \sin(t) \end{pmatrix}$, $0 \leq t \leq 6\pi$, aufgetragen: links wiederum der Graph, rechts der Wertebereich.



Für die Veranschaulichung von **komplexen** Funktionen $f : \mathbb{C} \supset D \rightarrow \mathbb{C}$ können wir den Graphen nicht skizzieren, da dieser ja in \mathbb{C}^2 , also im vier-dimensionalen Raum (über \mathbb{R}) liegt. Daher beschreibt man Abbildungen von \mathbb{C} in sich auf andere Weise, z.B. wie gewisse Mengen von \mathbb{C} abgebildet werden. Dies haben wir in Abschnitt 3.1 in HM I schon angedeutet.

Vektorfelder $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ lassen sich durch Pfeile andeuten, indem an entsprechend vielen Punkte $x \in \mathbb{R}^2$ der Vektor $f(x)$ gezeichnet wird, wie etwa bei den Strömungsfeldern, um auf Wetterkarten den Wind zu illustrieren. Das Bild zeigt eine solche Darstellung für das Vektorfeld

$$f(x) = \begin{pmatrix} \frac{x_2 - 1}{\sqrt{(x_1 - 1)^2 + (x_2 - 1/2)^2 + 1}} \\ \frac{x_1 - 1}{\sqrt{(x_1 - 1)^2 + (x_2 - 1/2)^2 + 1}} \end{pmatrix}.$$



Ein zentraler Begriff bei Funktionen ist die Stetigkeit, die uns im eindimensionalen Fall inzwischen bekannt ist. Zur Definition der Stetigkeit in höheren Dimensionen benötigen wir einen Konvergenzbegriff. Dazu ist es sinnvoll, zunächst geometrische Aspekte wie etwa einen **Abstands begriff** zu klären.

Definition 1.4 (Norm)

Eine **Norm** im \mathbb{R}^n ist eine Abbildung vom \mathbb{R}^n nach \mathbb{R} , die jedem Element $x \in \mathbb{R}^n$ eine eindeutige Zahl $\|x\| \in \mathbb{R}$ zuordnet, wobei die drei folgenden Eigenschaften erfüllt sein müssen:

- (a) $\|\lambda x\| = |\lambda| \|x\|$ für alle $x \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$, (Homogenität)
- (b) $\|x + y\| \leq \|x\| + \|y\|$ für alle $x, y \in \mathbb{R}^n$, (Dreiecksungleichung)
- (c) $x \neq 0 \iff \|x\| > 0$. (Definitheit)

Die Zahl $\|x\| \geq 0$ heißt **Norm** von $x \in \mathbb{R}^n$.

Wir sollten uns $\|x\|$ als **Abstand** des Punktes $x \in \mathbb{R}^n$ vom Ursprung $0 \in \mathbb{R}^n$ vorstellen – wobei dies aber nicht unbedingt der „Luftlinienabstand“ sein muss. Allgemeiner ist $\|x - y\|$ dann der Abstand zwischen x und y . Besonders interessieren uns die drei folgenden Normen:

$$\begin{aligned} \|x\|_1 &= \sum_{j=1}^n |x_j| \quad (\text{Betragssummennorm}), \\ \|x\|_2 &= \sqrt{\sum_{j=1}^n |x_j|^2} \quad (\text{euklidische Norm}), \\ \|x\|_\infty &= \max_{j=1, \dots, n} |x_j| \quad (\text{Maximumnorm}). \end{aligned}$$

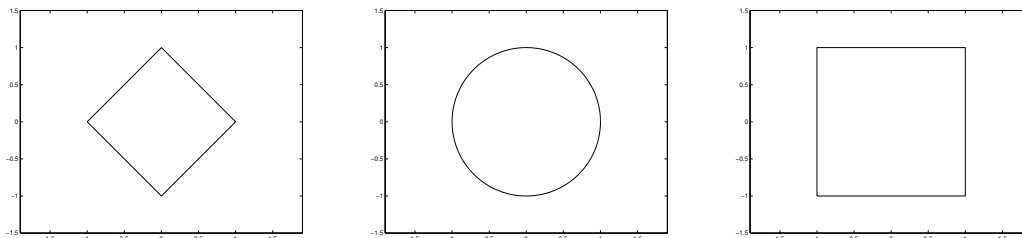
Im Satz 1.23 in HM2 haben wir schon die Eigenschaften (a), (b), (c) für die euklidische Norm $\|\cdot\|_2$ nachgewiesen.¹ Für die anderen beiden Normen ist dies einfach und verbleibt als Übung.

Bemerkung: Man kann Normen nicht nur im \mathbb{R}^n definieren, sondern auch in allgemeineren Vektorräumen V . Eine Norm ist dann eine Abbildung von V nach $\mathbb{R}_{\geq 0}$, die den Gesetzen (a), (b), (c) von Definition 1.4 genügt. Vektorräume mit einer Norm heißen **normierte Räume**.

Ist $\|\cdot\|$ irgendeine Norm im \mathbb{R}^n , so definieren wir für $x \in \mathbb{R}^n$ und $r > 0$:

$$K(x, r) := \{y \in \mathbb{R}^n : \|x - y\| < r\} \quad \text{und} \quad K[x, r] := \{y \in \mathbb{R}^n : \|x - y\| \leq r\}.$$

$K(x, r)$ ist also die Menge der Punkte, die von x den Abstand weniger als r haben. Im Fall $n = 3$ und $\|\cdot\| = \|\cdot\|_2$ ist $K(x, r)$ genau die Kugel um x mit Radius r . Im folgenden Bild sind die Mengen $K(0, 1)$ des \mathbb{R}^2 für die drei Normen $\|\cdot\|_1$, $\|\cdot\|_2$ und $\|\cdot\|_\infty$ skizziert:



Nur für die euklidische Norm $\|\cdot\|_2$ deckt sich $K(0, r)$ mit unserer Vorstellung von Kreisen. Trotzdem nennen wir für jede Norm $K(x, r)$ die „Kugel um x mit Radius r “.

¹Dort hatten wir einfach $\|\cdot\|$ an Stelle von $\|\cdot\|_2$ geschrieben.

Jetzt können wir die Definition 2.2 von HM I der beschränkten oder konvergenten Zahlenfolgen fast wörtlich übernehmen.

Definition 1.5 Sei $(x^{(k)})$ eine Punktfolge im \mathbb{R}^n . Sie **konvergiert** bzgl. der Norm $\|\cdot\|$ gegen $\hat{x} \in \mathbb{R}^n$, wenn es zu jedem $\varepsilon > 0$ ein $K \in \mathbb{N}$ gibt mit

$$\|x^{(k)} - \hat{x}\| < \varepsilon \quad \text{für alle } k \geq K.$$

Die unendlich vielen Möglichkeiten, eine Norm in \mathbb{R}^n zu definieren, scheinen zu verwirren. Aber in Hinblick auf Konvergenz stimmen die Begriffe alle überein. So gilt etwa, dass eine Folge $(x^{(k)})$ genau dann gegen \hat{x} im Sinne von Definition 1.5 konvergiert, wenn sie im Sinne der eindimensionalen Konvergenz für jede Komponente konvergiert. Für die Normen $\|\cdot\|_1$ und $\|\cdot\|_\infty$ sehen wir dieses schnell direkt. Für eine beliebige Norm folgt dies aus dem folgenden Satz, den wir hier nicht beweisen wollen:

Satz 1.6 Sei $\|\cdot\|$ eine beliebige Norm auf dem \mathbb{R}^n . Dann gibt es $c_1 > 0$ und $c_2 > 0$ mit

$$c_1 \|x\| \leq \|x\|_\infty \leq c_2 \|x\| \quad \text{für alle } x \in \mathbb{R}^n.$$

Dies bedeutet, dass jede Kugel bzgl. $\|\cdot\|$ eine „Kugel“ bzgl. $\|\cdot\|_\infty$ enthält und umgekehrt.

Nun können wir die **Stetigkeit von vektorwertigen Funktionen** von $D \subset \mathbb{R}^n$ in den \mathbb{R}^m wie bei Funktionen einer Variablen erklären.

Definition 1.7 Die Funktion $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ heißt **stetig** in $\hat{x} \in D$, falls für jede Folge $(x^{(k)})_k \subseteq D$ mit $x^{(k)} \rightarrow \hat{x}$ ($k \rightarrow \infty$) gilt, dass auch $f(x^{(k)}) \rightarrow f(\hat{x})$ ($k \rightarrow \infty$) konvergiert.

Bemerkung: Beachte, dass wir wegen des vorhergegangenen Satzes die Norm in dieser Definition nicht genauer festlegen müssen. Aber die Aussage muß für jede konvergente Folge $(x^{(k)})$ gelten. Die „Bahn“, auf der eine Folge gegen \hat{x} konvergiert, darf beliebig kompliziert sein.

Wichtige **Spezialfälle** sind:

(a) $n = 1$ und m beliebig, also vektorwertige Funktionen $f : \mathbb{R} \supset I \rightarrow \mathbb{R}^m$ in einer Variablen. Hier schreiben wir I statt D , weil $I \subset \mathbb{R}$ im allgemeinen ein Intervall ist. Offensichtlich sind solche Funktionen f von einer Variablen genau dann stetig, wenn jede Komponente f_i für $i = 1, \dots, m$ stetig ist. Solche stetigen Funktionen beschreiben **Kurven**. Sie werden meistens mit kleinen lateinischen Buchstaben bezeichnet, also etwa $x : [a, b] \rightarrow \mathbb{R}^m$. Das Intervall $I := [a, b]$ heißt **Parameterintervall**, und der Wertebereich $x(I) = \{x(t) : t \in [a, b]\} \subset \mathbb{R}^m$ ist eine Kurve. Auf Kurven werden wir in Abschnitt 3.1 genauer eingehen.

Beispiele:

(i) $x(t) = \begin{pmatrix} t \cos t \\ t \sin t \end{pmatrix}$, $0 \leq t \leq 4\pi$, also $x_1(t) = t \cos t$ und $x_2(t) = t \sin t$. Die Kurve ist eine Spirale.

(ii) $x(t) = \begin{pmatrix} t \\ \sin t \end{pmatrix}$, $0 \leq t \leq 2\pi$. Hier ist $x_1 = t$ und $x_2 = \sin t$, also $x_2 = \sin x_1$. Die Kurve ist daher der Graph der Sinusfunktion.

- (iii) Es seien $a, b > 0$. Die Kurve $x(t) = \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix}$, $0 \leq t \leq 6\pi$, beschreibt eine Ellipse mit Halbachsen a und b , die dreimal durchlaufen wird.

Alle diese Funktionen sind stetig, da ihre Komponenten stetig sind.

(b) $m = 1$ und n beliebig. Jetzt betrachten wir also Funktionen $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}$, d.h. **skalarwertige Funktionen** von n Variablen. Für $n = 2$ und $n = 3$ schreiben wir häufig auch (x, y) bzw. (x, y, z) an Stelle von (x_1, x_2) bzw. (x_1, x_2, x_3) .

Beispiele: Die folgenden Funktionen sind auf ihrer Definitionsmenge stetig,

$$\begin{aligned} f(x) &= \sum_{j=0}^n \sum_{k=0}^m a_{jk} x_1^j x_2^k, \quad x \in \mathbb{R}^2, \quad (\text{Polynom in zwei Variablen}), \\ f(x) &= g(\|x\|), \quad x \in \mathbb{R}^n, \quad \text{für jede stetige Funktion } g : \mathbb{R} \rightarrow \mathbb{R}, \\ f(x, y) &= \frac{x+y}{x^2+y^2}, \quad (x, y)^\top \in \mathbb{R}^2, \quad (x, y) \neq (0, 0). \end{aligned}$$

Wie im Eindimensionalen gilt für stetige Funktionen $f, g : D \supseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$, dass auch $f \pm g : D \rightarrow \mathbb{R}^m$ und $f \cdot g : D \rightarrow \mathbb{R}$ stetig sind. Wenn f, g skalarwertige Funktionen sind mit $g(\hat{x}) \neq 0$ für ein $\hat{x} \in D$, so ist auch f/g in \hat{x} stetig.

Mit einer Norm können wir weiter die wichtigen Grundbegriffe „offen“, „abgeschlossen“ und „kompakt“ festlegen.

Definition 1.8 (topologische Grundbegriffe)

Sei $M \subset \mathbb{R}^n$ eine Teilmenge und $\|\cdot\|$ eine beliebige Norm im \mathbb{R}^n .

- (a) Ein Punkt $x \in M$ heißt **innerer Punkt** von M , falls es $\varepsilon > 0$ gibt mit $K(x, \varepsilon) \subset M$.
- (b) Die Menge M heißt **offen**, wenn M nur aus inneren Punkten besteht, d.h. wenn es zu jedem $x \in M$ ein $\varepsilon > 0$ gibt mit $K(x, \varepsilon) \subset M$.
- (c) Die Menge M heißt **abgeschlossen**, wenn der Grenzwert jeder konvergente Folge $(x^{(k)})_k$ aus M ebenfalls zu M gehört, d.h.:

$$x^{(k)} \in M \quad \text{und} \quad x^{(k)} \rightarrow x \in \mathbb{R}^n \quad \text{impliziert} \quad x \in M.$$

- (d) Die Menge M heißt **kompakt**, wenn **jede** Folge $(x^{(k)})_{k \in \mathbb{N}}$ aus M einen Häufungspunkt besitzt, der auch zu M gehört.
- (e) Ein Punkt $x \in \mathbb{R}^n$ heißt **Randpunkt** von M , falls für jedes $\varepsilon > 0$ sowohl $M \cap K(x, \varepsilon) \neq \emptyset$ als auch $K(x, \varepsilon) \setminus M \neq \emptyset$ ist.
- (f) $M^\circ := \{x \in M : x \text{ innerer Punkt von } M\}$ heißt **Inneres** von M ,
 $\partial M := \{x \in \mathbb{R}^n : x \text{ Randpunkt von } M\}$ heißt **Rand** von M ,
 $\overline{M} := M \cup \partial M$ heißt der **Abschluß** von M .

Der Satz 1.6 liefert, dass diese Begriffe unabhängig von der gewählten Norm sind. Eine Menge ist also z.B. offen bzgl. $\|\cdot\|_1$ genau dann, wenn sie auch offen bzgl. $\|\cdot\|_\infty$ ist.

Für $n = 1$ und Intervalle überzeuge man sich davon, dass diese Definitionen mit den bekannten Begriffen „innerer Punkt“, „Randpunkt“, „offenes“ und „abgeschlossenes“ Intervall übereinstimmen.

Die folgenden Eigenschaften lassen sich mehr oder weniger leicht nachprüfen.

Satz 1.9 Sei $M \subset \mathbb{R}^n$ eine Menge und $\|\cdot\|$ eine beliebige Norm im \mathbb{R}^n .

- (i) M ist genau dann offen, wenn die Komplementärmenge $\mathbb{R}^n \setminus M$ abgeschlossen ist.
- (ii) M ist genau dann abgeschlossen, wenn $\partial M \subset M$, d.h. wenn $M = \overline{M}$.
- (iii) Die Kugel $K[x, r] = \{y \in \mathbb{R}^n : \|y - x\| \leq r\}$ ist kompakt, und $K(x, r) = \{y \in \mathbb{R}^n : \|y - x\| < r\}$ ist offen mit dem Rand $\partial K[x, r] = \partial K(x, r) = \{y \in \mathbb{R}^n : \|x - y\| = r\}$.
- (iv) Es seien $M_j \subset \mathbb{R}^n$ endlich oder abzählbar unendlich viele **offene** Mengen. Dann ist auch $\bigcup_j M_j$ offen.
Sind $M_j \subset \mathbb{R}^n$ endlich oder abzählbar unendlich viele **abgeschlossene** Mengen, so ist auch $\bigcap_j M_j$ abgeschlossen.
- (v) Es seien $M_1, \dots, M_p \subset \mathbb{R}^n$ **endlich viele** offene Mengen. Dann ist auch $\bigcap_{j=1}^p M_j$ offen.
Es seien $M_1, \dots, M_p \subset \mathbb{R}^n$ **endlich viele** abgeschlossene Mengen. Dann ist auch $\bigcup_{j=1}^p M_j$ abgeschlossen.
- (vi) Eine Menge $M \subset \mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.
Eine Menge heißt **beschränkt**, wenn es $c > 0$ gibt mit $M \subseteq K(0, c)$.

Bemerkung: Die Eigenschaft (vi) ist eine Umformulierung des Satzes von **Bolzano-Weierstraß** (Satz 2.16 von HM I), übertragen in den \mathbb{R}^n : Jede beschränkte Folge $(x^{(k)})$ im \mathbb{R}^n enthält eine konvergente Teilfolge.

Das folgende Lemma stellt ein nützliches Hilfsmittel zur Untersuchung auf Offen- oder Abgeschlossenheit dar.

Lemma 1.10 Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ stetig. Dann ist $M := \{x \in \mathbb{R}^n : f_i(x) < 0 \forall i = 1, \dots, m\}$ offen und der Abschluss ist $\overline{M} := \{x \in \mathbb{R}^n : f_i(x) \leq 0 \text{ für alle } i = 1, \dots, m\}$. Die Menge $\{x \in \mathbb{R}^n : f_i(x) = 0 \text{ für alle } \forall i = 1, \dots, m\}$ ist ebenfalls abgeschlossen.

Beispiele 1.11 $A := \{x \in \mathbb{R}^2 : |x_1| + |x_2| \leq 4\}$ ist kompakt,

$B := \{x \in \mathbb{R}^2 : x \cdot z > 2\}$ mit beliebigem $z \in \mathbb{R}^2$ ist offen in \mathbb{R}^2 ,

$C := \{x \in \mathbb{R}^2 : x_1 \in \mathbb{Z}, x_2 \in \mathbb{Z}\}$ ist abgeschlossen aber nicht beschränkt,

$\partial A = \{x \in \mathbb{R}^2 : |x_1| + |x_2| = 4\}$, $A^\circ = \{x \in \mathbb{R}^2 : |x_1| + |x_2| < 4\}$,

$\partial B = \{x \in \mathbb{R}^2 : x \cdot z = 2\}$, $\overline{B} = \{x \in \mathbb{R}^2 : x \cdot z \geq 2\}$, $C^\circ = \emptyset$, $\partial C = C$. □

Im Zusammenhang mit der Stetigkeit ist folgende Eigenschaft kompakter Mengen wichtig (vergleiche Lemma 3.22 von HM I):

Satz 1.12

Ist D kompakt und $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ stetig, so ist die Bildmenge $f(D) \subset \mathbb{R}^m$ kompakt. Insbesondere existieren Minimum und Maximum skalarwertiger, stetiger Funktionen $f : D \rightarrow \mathbb{R}$ auf kompakten Mengen D , d.h. es existieren $x^+, x^- \in D$ mit

$$f(x^+) = \max_{x \in D} f(x) \quad \text{und} \quad f(x^-) = \min_{x \in D} f(x).$$

Beweis: Ist $(f(x_j))_{j \in \mathbb{N}}$ eine Folge im Bild $f(D)$, so ist x_j eine Folge in D . Da D kompakt ist, besitzt x_j eine konvergente Teilfolge, $(x_{j(n)})$. Also gibt es $x \in D$ mit $\lim_{n \rightarrow \infty} x_{j(n)} = x$. Da f stetig ist, folgt $\lim_{n \rightarrow \infty} f(x_{j(n)}) = f(x) \in f(D)$. Somit besitzt die Folge $(f(x_j))$ eine konvergente Teilfolge in $f(D)$ und wir haben gezeigt, dass $f(D)$ kompakt ist.

Im Fall einer skalarwertigen Funktion, $m = 1$, liefert das Lemma 3.21 aus HM I die Existenz von Minimum und Maximum in der kompakten Menge $f(D)$. Das ist der zweite Teil der Behauptung. \square

1.2 Partielle Ableitungen und Jacobimatrix

In diesem Abschnitt werden wir die Definition der Differenzierbarkeit auf Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ mehrerer Variablen verallgemeinern. Zunächst sei $m = 1$, d.h. wir betrachten zunächst skalarwertige Funktionen $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}$.

Definition 1.13 Sei $D \subset \mathbb{R}^n$ eine offene Menge, $\hat{x} \in D$ und $f : D \rightarrow \mathbb{R}$ eine skalarwertige Funktion. f heißt in \hat{x} partiell nach x_j differenzierbar, $j = 1, \dots, n$, wenn der Grenzwert

$$\frac{\partial f(\hat{x})}{\partial x_j} := f_{x_j}(\hat{x}) := \lim_{h \rightarrow 0} \frac{1}{h} [f(\hat{x}_1, \dots, \hat{x}_{j-1}, \hat{x}_j + h, \hat{x}_{j+1}, \dots, \hat{x}_n) - f(\hat{x})]$$

existiert. $f_{x_j}(\hat{x}) = \frac{\partial f(\hat{x})}{\partial x_j} = \partial f(\hat{x}) / \partial x_j$ heißt **partielle Ableitung von f nach x_j** . Die Funktion f heißt in \hat{x} **partiell differenzierbar**, wenn sie nach jeder Variablen x_j partiell differenzierbar ist.

Die partiellen Ableitungen werden in einem Spaltenvektor

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)^\top \in \mathbb{R}^n$$

zusammengefasst. Dieser Vektor wird **Gradient** von f an der Stelle x genannt.

Beispiele 1.14

(a) Sei $n = 2$ und $f(x) = f(x_1, x_2) = x_1^2 + 3x_2^2 + 4x_1^2x_2$, $x \in \mathbb{R}^2$. Dann ist

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 + 8x_1x_2, \quad \frac{\partial f(x)}{\partial x_2} = 6x_2 + 4x_1^2.$$

(b) $f(x) = \sqrt{x_1^2 + x_2^2} = \|x\|_2$, $x \in \mathbb{R}^2$. Dann ist

$$f_{x_1} = \frac{x_1}{\|x\|_2}, \quad f_{x_2} = \frac{x_2}{\|x\|_2}, \quad x \neq (0, 0), \quad \text{d.h.} \quad \nabla f(x) = \frac{x}{\|x\|_2}.$$

In $x = (0, 0)^\top$ ist die Funktion f nicht partiell differenzierbar! (Beweis?)

(c) Wir betrachten die folgende zusammengesetzte Funktion:

$$f(x) = \begin{cases} \frac{x_1 x_2}{\|x\|_2} = \frac{x_1 x_2}{\sqrt{x_1^2 + x_2^2}}, & x \neq (0, 0), \\ 0, & x = (0, 0). \end{cases}$$

Für $x \neq (0, 0)$ ist f partiell differenzierbar, und es ist z.B.

$$f_{x_1}(x) = \frac{1}{x_1^2 + x_2^2} \left[x_2 \sqrt{x_1^2 + x_2^2} - \frac{x_1^2 x_2}{\|x\|_2} \right] = \frac{x_2^3}{\|x\|_2^3}, \quad x \neq (0, 0).$$

An der Stelle $x = (0, 0)$ sehen wir direkt mit der Definition, dass $\frac{1}{h}[f(h, 0) - f(0, 0)] = 0$ und genauso $\frac{1}{h}[f(0, h) - f(0, 0)] = 0$. Daher ist f auch in $(0, 0)$ partiell differenzierbar und $f_{x_1}(0, 0) = f_{x_2}(0, 0) = 0$.

□

Wir können die partielle Ableitung mit den **Koordinateneinheitsvektoren** $e^{(j)}$, $j = 1, \dots, n$, auch schreiben als

$$\frac{\partial f(\hat{x})}{\partial x_j} = \lim_{h \rightarrow 0} \frac{1}{h} [f(\hat{x} + h e^{(j)}) - f(\hat{x})].$$

An dieser Form erkennen wir, dass die Ableitung die lokale Änderungsrate von f in Richtung $e^{(j)}$ ist. Allgemeiner erklären wir:

Definition 1.15 Sei $D \subset \mathbb{R}^n$ eine offene Menge, $a \in \mathbb{R}^n$ ein Einheitsvektor (d.h. $\|a\|_2 = 1$), $\hat{x} \in D$ und $f : D \rightarrow \mathbb{R}$ eine skalare Funktion. f heißt **in \hat{x} in Richtung a differenzierbar**, wenn der Limes

$$\frac{\partial f(\hat{x})}{\partial a} := \lim_{h \rightarrow 0} \frac{1}{h} [f(\hat{x} + ha) - f(\hat{x})]$$

existiert. Der Grenzwert heißt **Richtungsableitung** von f in \hat{x} in Richtung a .

Für das konkrete Ausrechnen von $\partial f(x)/\partial a$ ist der folgende Satz nützlich.

Satz 1.16 Sei $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ partiell differenzierbar und $\nabla f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ stetig. Dann ist die Richtungsableitung von f in Richtung eines Einheitsvektors $a \in \mathbb{R}^n$ gegeben durch

$$\frac{\partial f(x)}{\partial a} = a^\top \nabla f(x).$$

Der normierte Gradient $\hat{a} := \nabla f(x)/\|\nabla f(x)\|_2$ von f in x ist die **Richtung der größten Steigung** von f in x , d.h. für alle Einheitsvektoren a gilt

$$\frac{\partial f(x)}{\partial a} \leq \frac{\partial f(x)}{\partial \hat{a}}.$$

Beweis für $n = 2$: Sei $x = (x_1, x_2) \in D$ festgehalten. Wir spalten auf

$$f(x_1 + ha_1, x_2 + ha_2) - f(x_1, x_2) = [f(x_1 + ha_1, x_2 + ha_2) - f(x_1, x_2 + ha_2)] + [f(x_1, x_2 + ha_2) - f(x_1, x_2)].$$

Die Funktionen $f(\cdot, x_2 + ha_2)$ und $f(x_1, \cdot)$ sind beide differenzierbar. Der eindimensionale Mittelwertsatz (Satz 5.24 von HM I) liefert die Existenz von u_h zwischen x_1 und $x_1 + ha_1$ und von v_h zwischen x_2 und $x_2 + ha_2$ mit

$$f(x_1 + ha_1, x_2 + ha_2) - f(x_1, x_2) = \frac{\partial f(u_h, x_2 + ha_2)}{\partial x_1} h a_1 + \frac{\partial f(x_1, v_h)}{\partial x_2} h a_2.$$

Da $\partial f/\partial x_j$ für $j = 1, 2$ stetig sind und $u_h \rightarrow x_1, v_h \rightarrow x_2$ für $h \rightarrow 0$ gilt, liefert Division durch h für $h \rightarrow 0$

$$\frac{1}{h} [f(x_1 + ha_1, x_2 + ha_2) - f(x_1, x_2)] \rightarrow \frac{\partial f(x_1, x_2)}{\partial x_1} a_1 + \frac{\partial f(x_1, x_2)}{\partial x_2} a_2 = a^\top \nabla f(x).$$

Damit ist $\partial f(x)/\partial a = a^\top \nabla f(x)$ gezeigt. Mit der Cauchy-Schwarzschen Ungleichung (also die Ungleichung $|a^\top b| \leq \|a\|_2 \|b\|_2$, s. Satz 1.23 in HM II) und $\|a\|_2 = 1$ ist

$$\frac{\partial f(x)}{\partial a} = a^\top \nabla f(x) \leq \|\nabla f(x)\|_2 \quad \text{und} \quad \frac{\partial f(x)}{\partial \hat{a}} = \hat{a}^\top \nabla f(x) = \|\nabla f(x)\|_2.$$

□

Nun betrachten wir den allgemeineren Fall $m \geq 1$, d.h. vektorwertige Funktionen $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$, d.h. $f(x) = (f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n))^\top$.

Definition 1.17 Sei $D \subset \mathbb{R}^n$ eine offene Menge. Eine Funktion $f : D \rightarrow \mathbb{R}^m$ heißt **stetig differenzierbar** in D , wenn für jedes $x \in D$ alle partiellen Ableitungen

$$\frac{\partial f_i(x)}{\partial x_j}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

aller Komponenten f_i existieren und selbst stetige Funktionen sind. Wir setzen

$$f'(x) := \left(\frac{\partial f_i(x)}{\partial x_j} \right)_{\substack{i=1, \dots, m \\ j=1, \dots, n}} \in \mathbb{R}^{m \times n}$$

und nennen $f'(x)$ die **Funktionalmatrix**, **Jacobimatrix** oder **Ableitungsmatrix** von f an der Stelle x .

Beachte: $f'(x)$ ist jetzt keine Zahl mehr (falls $n \geq 2$ und/oder $m \geq 2$), sondern eine **Matrix**. Spätestens hier muss deutlich zwischen der Funktion $f' : D \rightarrow \mathbb{R}^{m \times n}$ und dem Funktionswert $f'(x) \in \mathbb{R}^{m \times n}$ unterschieden werden.

Spezialfälle:

(a) $m = 1$, also $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Hier ist

$$f'(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right) = \nabla f(x)^\top,$$

und dies ist ein Zeilenvektor. In der Vorlesung wird also bei skalaren Funktionen zwischen $f'(x)$ (Zeilenvektor) und $\nabla f(x)$ (Spaltenvektor) unterschieden.

(b) $n = 1$, also $f : \mathbb{R} \rightarrow \mathbb{R}^m$. Dies beschreibt eine **Kurve** im \mathbb{R}^m . Es ist

$$f'(t) = \begin{pmatrix} \frac{\partial f_1(t)}{\partial t} \\ \vdots \\ \frac{\partial f_m(t)}{\partial t} \end{pmatrix} \in \mathbb{R}^m.$$

Hier ist also $f'(t)$ ein Spaltenvektor. Häufig findet sich im Zusammenhang mit Kurven auch die Notation $\dot{f}(t) = f'(t)$ für den *Geschwindigkeitsvektor*.

Beispiele 1.18

(a) $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ sei gegeben durch

$$f(x) = f(x_1, x_2, x_3) = \begin{pmatrix} x_1^2 + \sin x_2 + x_3 \\ \cos x_1 - x_2 + x_3^2 \end{pmatrix}, \quad x \in \mathbb{R}^3.$$

Hier ist

$$f'(x) = \begin{pmatrix} 2x_1 & \cos x_2 & 1 \\ -\sin x_1 & -1 & 2x_3 \end{pmatrix} \in \mathbb{R}^{2 \times 3}.$$

(b) Betrachte die Funktion f von Beispiel 1.14, (c). Wir hatten gesehen, dass

$$f_{x_1}(x) = \frac{x_2^3}{\|x\|_2^3} = \frac{x_2^3}{(x_1^2 + x_2^2)^{3/2}}, \quad x \neq (0, 0), \quad \text{und} \quad f_{x_1}(0, 0) = 0.$$

f_{x_1} ist nicht stetig, denn auf der Geraden $x_1 = 0$ ist $f_{x_1}(0, x_2) = 1$, und dies konvergiert nicht gegen 0 für $x_2 \rightarrow 0$. Also ist die Funktion zwar partiell differenzierbar, d.h. es existieren alle partiellen Ableitungen, aber sie ist nicht stetig differenzierbar.

(c) Seien $a, p \in \mathbb{R}^m$ Vektoren, $a \neq 0$, und sei $f : \mathbb{R} \rightarrow \mathbb{R}^m$ definiert durch $f(t) = p + ta$, $t \in \mathbb{R}$. Dann beschreibt f eine **Gerade** im \mathbb{R}^m in Parameterform. Es ist $f'(t) = a$.

(d) Seien $a, b \in \mathbb{R}^m$ linear unabhängig, $p \in \mathbb{R}^m$, und $f : \mathbb{R}^2 \rightarrow \mathbb{R}^m$ gegeben durch $f(t, s) = p + ta + sb$, $s, t \in \mathbb{R}$. Dies beschreibt eine **Ebene** im \mathbb{R}^m in Parameterform. Es ist $f'(t, s) = [a \ b] \in \mathbb{R}^{m \times 2}$.
□

Analog zum Fall $n = m = 1$ benötigen wir die folgenden Rechenregeln für die Ableitungsmatrizen:

(a) **1. Produktregel:** Es seien $f, g : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Dann ist das **Skalarprodukt** $f^\top g : \mathbb{R}^n \rightarrow \mathbb{R}$ definiert durch

$$(f^\top g)(x) := f(x)^\top g(x) = \sum_{i=1}^m f_i(x)g_i(x).$$

Es gilt:

$$(f^\top g)'(x) = f(x)^\top g'(x) + g(x)^\top f'(x).$$

(b) **2. Produktregel:** Es sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine vektorwertige und $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ eine skalare Funktion. Dann ist

$$(\psi f)'(x) = \psi(x) f'(x) + f(x) \psi'(x).$$

Beachte, dass der zweite Term ein dyadisches Produkt ist.

(c) **Kettenregel:** Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ und $g : \mathbb{R}^p \rightarrow \mathbb{R}^n$. Dann ist $f \circ g : \mathbb{R}^p \rightarrow \mathbb{R}^m$ definiert durch

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

und es gilt

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

(Matrixprodukt!). Man prüfe jeweils nach, dass die Ausdrücke auf der linken und der rechten Seite Sinn machen!

Lemma 1.19

Ein wichtiges Beispiel zur Kettenregel ist der Fall, dass eine differenzierbare Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ längs einer Kurve $\psi : \mathbb{R} \rightarrow \mathbb{R}^n$ im \mathbb{R}^n gegeben ist, d.h. $F(t) = (f \circ \psi)(t) = f(\psi(t))$, $t \in \mathbb{R}$. In diesem Fall besagt die Kettenregel:

$$F'(t) = f'(\psi(t)) \psi'(t) = \nabla f(\psi(t)) \cdot \psi'(t), \quad t \in \mathbb{R}.$$

Man beachte, dass auf der rechten Seite das Skalarprodukt zweier Vektoren steht.

Beispiele 1.20 Die angesprochene Kettenregel wird etwa beim Ableiten bei parameterabhängigen Integralen benötigt. Gesucht ist zum Beispiel die Ableitung von $h : (0, 1) \rightarrow \mathbb{R}$ mit

$$h(t) = \int_{t^2}^t \ln(ts) \, ds.$$

Dazu definieren wir $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ und $\varphi : (0, 1) \rightarrow \mathbb{R}^3$ durch

$$f(x, y, z) = \int_y^z \ln(xs) \, ds \quad \text{und} \quad \varphi(t) = \begin{pmatrix} t \\ t^2 \\ t \end{pmatrix}.$$

Mit der Kettenregel folgt

$$\begin{aligned} h'(t) &= \nabla f(\varphi(t)) \cdot \varphi'(t) \\ &= \left(\begin{array}{c} \int_y^z \frac{1}{x} \, ds \\ -\ln(xy) \\ \ln(xz) \end{array} \right) \Big|_{(x,y,z)=\varphi(t)} \cdot \begin{pmatrix} 1 \\ 2t \\ 1 \end{pmatrix}. \end{aligned}$$

Also erhalten wir die Ableitung

$$h'(t) = \frac{t - t^2}{t} - 2t \ln(t^3) + \ln(t^2) = 1 - t + (2 - 6t) \ln(t).$$

Im eindimensionalen Fall hatten wir mit Hilfe der Ableitung nichtlineare Funktionen um \hat{x} **linearisiert**, d.h. es gilt nach Lemma 5.16 von HM I:

$$f(x) = f(\hat{x}) + f'(\hat{x})(x - \hat{x}) + h(x), \quad x \in [\hat{x} - a, \hat{x} + a],$$

mit einer im Punkt \hat{x} stetigen Funktion h , die

$$h(\hat{x}) = 0 \quad \text{und} \quad \lim_{x \rightarrow \hat{x}} \frac{h(x)}{x - \hat{x}} = 0$$

genügt. Die Gerade

$$g(x) = f(\hat{x}) + f'(\hat{x})(x - \hat{x}), \quad x \in \mathbb{R},$$

ist genau die **Tangente** von f in \hat{x} . Diese Interpretation als Linearisierung gilt genauso im Mehrdimensionalen:

Satz 1.21 (Linearisierung) Sei $D \subset \mathbb{R}^n$ eine offene Menge und $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}^m$ stetig differenzierbar im Punkt $\hat{x} \in D$. Dann existiert eine im Punkt \hat{x} stetige Funktion $h : D \rightarrow \mathbb{R}^m$ mit

$$f(x) = f(\hat{x}) + f'(\hat{x})(x - \hat{x}) + h(x), \quad x \in D, \quad h(\hat{x}) = 0 \text{ und } \lim_{x \rightarrow \hat{x}} \frac{h(x)}{\|x - \hat{x}\|_\infty} = 0.$$

Die Abbildung

$$g(x) = f(\hat{x}) + f'(\hat{x})(x - \hat{x}), \quad x \in \mathbb{R}^n, \quad (1.1)$$

heißt **Tangentialabbildung** oder auch **Linearisierung** von f in \hat{x} . Im Fall $n = m = 1$ ist dies genau die Tangente. Wir sehen uns jetzt den Spezialfall $n = 2$ und $m = 1$ an, also den Fall einer skalarwertigen Funktion in zwei Variablen x_1, x_2 . Die Tangentialabbildung hat die Form

$$g(x_1, x_2) = f(\hat{x}) + f_{x_1}(\hat{x})(x_1 - \hat{x}_1) + f_{x_2}(\hat{x})(x_2 - \hat{x}_2), \quad x_1, x_2 \in \mathbb{R}.$$

Der Graph zu g ist eine Ebene, die **Tangentialebene**, zu f in (\hat{x}_1, \hat{x}_2) .

Im Fall $n = 1$ und $m = 3$ geht es um eine vektorwertige Funktion einer Variable. Eine solche Abbildung nennt man auch Kurve, ihre Tangentialabbildung hat die Form

$$g(x) = \begin{pmatrix} f_1(\hat{x}) \\ f_2(\hat{x}) \\ f_3(\hat{x}) \end{pmatrix} + (x - \hat{x}) \begin{pmatrix} f'_1(\hat{x}) \\ f'_2(\hat{x}) \\ f'_3(\hat{x}) \end{pmatrix}, \quad x \in \mathbb{R}.$$

Dies ist die Darstellung einer Gerade im \mathbb{R}^3 .

Wir können (1.1) so deuten, wenn wir $f(\hat{x})$ auf die andere Seite bringen: Kleine Änderungen von \hat{x} der Größe $dx = x - \hat{x}$ liefern „in erster Näherung“ Änderungen des Funktionswertes der Größe $f(x) - f(\hat{x}) \approx f'(\hat{x}) \cdot dx$. Analog zu Definition 5.16 in HM I definiert man das Differential für skalarwertige Funktionen in n Variablen.

Definition 1.22 Sei $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ differenzierbar im Punkt $\hat{x} \in D$. Das **Differential** zu f an der Stelle \hat{x} ist die Tangentialabbildung an f im Punkt $(\hat{x}, f(\hat{x}))$ in dem (dx, df) -Koordinatensystem, das den Ursprung in $(\hat{x}, f(\hat{x}))$ hat. Das Differential hat die Form $df = f'(\hat{x}) dx$ oder auch $df(\hat{x}) : dx \mapsto f'(\hat{x}) dx$ und ausgeschrieben

$$df = \frac{\partial f(\hat{x})}{\partial x_1} dx_1 + \dots + \frac{\partial f(\hat{x})}{\partial x_n} dx_n.$$

Das Differential lässt sich so verstehen, dass es in linearer Näherung die Änderung df des Funktionswerts $f(\hat{x})$ angibt, wenn \hat{x} um dx geändert wird.

Berücksichtigen wir die üblichen Rechenregeln beim Differenzieren, so ergeben sich entsprechende Regeln für das Rechnen mit Differentialen. Sind etwa $f, g : D \rightarrow \mathbb{R}$ differenzierbare Funktionen so folgt

$$d(f + g) = df + dg$$

oder mit der Produktregel

$$\begin{aligned} d(fg) &= \left(g(\hat{x}) \frac{\partial f}{\partial x_1}(\hat{x}) + f(\hat{x}) \frac{\partial g}{\partial x_1}(\hat{x}) \right) dx_1 + \cdots + \left(g(\hat{x}) \frac{\partial f}{\partial x_n}(\hat{x}) + f(\hat{x}) \frac{\partial g}{\partial x_n}(\hat{x}) \right) dx_n \\ &= g(\hat{x}) df + f(\hat{x}) dg. \end{aligned}$$

Jetzt wollen wir noch **zweite** und **höhere partielle Ableitungen** für Funktionen der Form $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ betrachten. Diese sind rekursiv definiert, also z.B.

$$f_{x_i x_j}(x) = \frac{\partial^2 f(x)}{\partial x_j \partial x_i} := \frac{\partial}{\partial x_j} \left(\frac{\partial f(x)}{\partial x_i} \right).$$

Beispiel 1.23 Die Menge $D \subset \mathbb{R}^2$ und die Funktion $f : D \rightarrow \mathbb{R}$ seien definiert durch

$$D := \left\{ (x, y) \in \mathbb{R}^2 : \frac{1}{2}\sqrt{1-x^2} < y < \sqrt{1-x^2}, 0 < x < 1 \right\},$$

und

$$f(x, y) := x^{y+1} = e^{(y+1)\ln x}, \quad (x, y) \in D.$$

Dann ist

$$\begin{aligned} f_x(x, y) &= e^{(y+1)\ln x} \frac{y+1}{x} = (y+1)x^y, & f_y(x, y) &= e^{(y+1)\ln x} \ln x = \ln x x^{y+1}, \\ f_{xx}(x, y) &= y(y+1)x^{y-1}, & f_{xy}(x, y) &= x^y + (y+1)x^y \ln x, \\ f_{yx}(x, y) &= (y+1)x^y \ln x + x^y, & f_{yy}(x, y) &= x^{y+1} (\ln x)^2. \end{aligned}$$

Damit ist f zweimal stetig differenzierbar in D . □

Wichtig ist die Matrix der zweiten Ableitungen, die man auch **Hesse-Matrix** nennt:

$$H_f(x) = \left(\frac{\partial^2 f(x)}{\partial x_j \partial x_i} \right)_{i,j=1,\dots,n} \in \mathbb{R}^{n \times n}.$$

Beispiel 1.24 (Quadratische Form)

Mit der quadratischen Matrix $A \in \mathbb{R}^{n \times n}$ sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ definiert durch $f(x) = x^\top A x$, $x \in \mathbb{R}^n$. Man nennt eine solche Abbildung quadratische Form. Ausgeschrieben ist

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j.$$

Also gilt

$$\begin{aligned} \frac{\partial f}{\partial x_k}(x) &= \sum_{i=1}^n \sum_{j=1}^n a_{ij} \left[x_j \frac{\partial x_i}{\partial x_k} + x_i \frac{\partial x_j}{\partial x_k} \right] = \sum_{i=1}^n \sum_{j=1}^n a_{ij} [x_j \delta_{ik} + x_i \delta_{jk}] \\ &= \sum_{j=1}^n a_{kj} x_j + \sum_{i=1}^n a_{ik} x_i, \end{aligned}$$

d.h. $\nabla f(x) = (A + A^\top)x$. Ist A symmetrisch, d.h. $A^\top = A$, so ist $\nabla f(x) = 2Ax$.

Leiten wir noch einmal ab, so erhalten wir

$$\frac{\partial^2 f}{\partial x_\ell \partial x_k}(x) = a_{k\ell} + a_{\ell k},$$

d.h. $H_f(x) = A + A^\top$. □

Der folgende Satz besagt, dass es bei stetig differenzierbaren Funktionen gleichgültig ist, in welcher Reihenfolge die Ableitungen vorgenommen werden.

Satz 1.25 (Schwarz)

Sei $D \subset \mathbb{R}^n$ offen, $f : D \rightarrow \mathbb{R}$ zweimal stetig differenzierbar (d.h. alle möglichen Ableitungen bis zur zweiten Ordnung existieren und seien stetig). Dann gilt

$$\frac{\partial^2 f(x)}{\partial x_i \partial x_j} = \frac{\partial^2 f(x)}{\partial x_j \partial x_i} \quad \text{für alle } x \in D, \quad i, j = 1, \dots, n,$$

d.h. die gemischten Ableitungen sind vertauschbar.

1.3 Exakte Differentialgleichungen und implizit gegebene Funktionen

Aufbauend auf dem grundlegenden Kalkül der Differentiation, können wir nun eine Vielzahl von Aspekten zur Analysis von Funktionen $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ untersuchen. Zunächst betrachten wir einen speziellen Typ gewöhnlicher Differentialgleichungen.

Definition 1.26 (exakte Differentialgleichung)

Sei $I \subset \mathbb{R}$ ein offenes Intervall und $p, q : I \times \mathbb{R} \rightarrow \mathbb{R}$ seien stetig differenzierbare skalare Funktionen in zwei Variablen. Die Differentialgleichung der Form

$$p(x, y(x)) + q(x, y(x)) y'(x) = 0, \quad x \in I,$$

heißt **exakt**, wenn die **Integrabilitätsbedingung** gilt,

$$\frac{\partial p(x, y)}{\partial y} = \frac{\partial q(x, y)}{\partial x} \quad \text{für alle } (x, y) \in I \times \mathbb{R}. \quad (1.2)$$

Ist die Differentialgleichung $p(x, y) + q(x, y) y' = 0$ exakt, so können wir versuchen, eine skalarwertige Funktion $f : I \times \mathbb{R} \rightarrow \mathbb{R}$ zu finden mit

$$p(x, y) = \frac{\partial f(x, y)}{\partial x} \quad \text{und} \quad q(x, y) = \frac{\partial f(x, y)}{\partial y}.$$

Die Differentialgleichung geht also über in die Form

$$\frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} y' = 0, \quad \text{d.h. } f_x(x, y(x)) + f_y(x, y(x)) y'(x) = 0 \text{ in } I.$$

Nach der zweidimensionalen Kettenregel ist dies

$$\frac{d}{dx} \left(f(x, y(x)) \right) = 0,$$

d.h. die eindimensionale Funktion $x \mapsto f(x, y(x))$ ist konstant. Das bedeutet, dass wir nun eine Lösung in impliziter Form $f(x, y(x)) = C$ gefunden haben, wobei die Konstante C durch eine eventuell gegebene Anfangsbedingung eindeutig festgelegt ist. Häufig, aber nicht notwendigerweise, kann man diese Gleichung noch nach $y(x)$ auflösen und bekommt so die Lösung explizit.

Die Integrabilitätsbedingung (1.2) ist offensichtlich notwendig, damit eine solche Funktion f existiert (s. Satz von Schwarz, 1.25). Wir werden später sehen, dass diese Bedingung unter entsprechenden Voraussetzungen auch hinreichend ist (s. Satz 3.29). Wenn f existiert, so heißt das Vektorfeld $(p, q) : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^2$ **konservativ** und f das (skalare) **Potential** zu diesem Vektorfeld.

Beispiel 1.27

Gegeben sei die Differentialgleichung

$$2x \sin y(x) - y(x) \sin x + [x^2 \cos y(x) + \cos x] y'(x) = 0 \quad (1.3)$$

mit der Anfangsbedingung $y(\pi/2) = \pi/6$. Die Differentialgleichung ist exakt, wie wir sofort nachrechnen. Wie finden wir f ? Dazu setzen wir direkt an:

$$\frac{\partial f(x, y)}{\partial x} = p(x, y) = 2x \sin y - y \sin x.$$

Integration liefert

$$f(x, y) = x^2 \sin y + y \cos x + c(y)$$

mit einer Integrationskonstanten $c(y)$, die noch von y abhängen kann. Differenzieren wir diese Gleichung partiell nach y , so erhalten wir

$$\frac{\partial f(x, y)}{\partial y} = x^2 \cos y + \cos x + c'(y).$$

Gleichsetzen mit $q(x, y) = x^2 \cos y + \cos x$ liefert $c'(y) = 0$, d.h. c ist konstant. Daher ist $f(x, y) = x^2 \sin y + y \cos x + c$. Die allgemeine Lösung der Differentialgleichung (1.3) muss also die Gleichung $x^2 \sin y + y \cos x = c$ mit irgendeiner Konstanten $c \in \mathbb{R}$ erfüllen. Setzen wir $x = \pi/2$ und $y = \pi/6$ ein, so erhalten wir $c = \pi^2/8$. Die Lösung $y(x)$ muss also durch Lösen der impliziten Gleichung

$$x^2 \sin y(x) + y(x) \cos x = \frac{\pi^2}{8}$$

gewonnen werden. Dies kann nicht geschlossen durchgeführt werden, und wir sind auf numerische Verfahren angewiesen (z.B. das Newtonverfahren, siehe letztes Semester).

Die Integrabilitätsbedingung (1.2) zur Exaktheit der Differentialgleichung ist natürlich eine sehr starke Bedingung und im allgemeinen nicht erfüllt. Aber häufig können wir eine allgemeine Differentialgleichung der Form

$$p(x, y(x)) + q(x, y(x)) y'(x) = 0$$

in eine äquivalente exakte Differentialgleichung umformen. Dazu multiplizieren wir die Differentialgleichung mit einer (noch zu bestimmenden) Funktion $\Lambda(x, y) \neq 0$:

$$\Lambda(x, y(x)) p(x, y(x)) + \Lambda(x, y(x)) q(x, y(x)) y'(x) = 0.$$

Nun versuchen wir $\Lambda(x, y)$ so zu bestimmen, dass die neue Differentialgleichung exakt ist. Wir müssen also $\Lambda(x, y)$ so wählen, dass

$$\frac{\partial(\Lambda(x, y) p(x, y))}{\partial y} = \frac{\partial(\Lambda(x, y) q(x, y))}{\partial x} \quad \text{für alle } (x, y) \in I \times \mathbb{R}$$

gilt. Die Funktion Λ heißt **integrierender Faktor** oder **Eulerscher Multiplikator** der Differentialgleichung. Wir sehen uns das folgende Beispiel an:

Beispiel 1.28 (Eine Riccatische Differentialgleichung)

Gegeben sei die Differentialgleichung

$$y'(x) = [x + y(x)]^2, \quad x \in \mathbb{R}.$$

Wählen wir hier $p(x, y) = (x + y)^2$ und $q(x, y) = -1$, so ist die Integrabilitätsbedingung offenbar nicht erfüllt. Wir müssen also Λ finden mit

$$\frac{\partial(\Lambda(x, y) p(x, y))}{\partial y} = \frac{\partial(\Lambda(x, y) q(x, y))}{\partial x}, \quad \text{d.h.}$$

$$2(x + y) \Lambda(x, y) + (x + y)^2 \frac{\partial \Lambda(x, y)}{\partial y} = - \frac{\partial \Lambda(x, y)}{\partial x}.$$

Allgemein werden wir uns im vorletzten Kapitel dieser Vorlesung mit solchen **partiellen Differentialgleichungen** beschäftigen. Manchmal, so wie hier, klappt es aber, die Differentialgleichung mit speziellen Ansätzen zu lösen. Oft hängt Λ nur von x oder y ab. In dem vorliegenden Beispiel führt der Ansatz $\Lambda(x, y) = \ell(x + y)$ mit einer noch zu bestimmenden Funktion ℓ zum Ziel (motiviert durch die spezielle Form der Koeffizienten). Setzen wir diesen Ansatz in die partielle Differentialgleichung ein, so ergibt sich die gewöhnliche Differentialgleichung,

$$2(x + y) \ell(x + y) + [1 + (x + y)^2] \ell'(x + y) = 0.$$

Mit $z = x + y$ ist dies nun eine Differentialgleichung mit getrennten Veränderlichen,

$$\ell'(z) = - \frac{2z}{1 + z^2} \ell(z),$$

und diese wird gelöst durch $\ell(z) = 1/(1 + z^2)$. Daher ist

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

ein integrierender Faktor der Differentialgleichung. Diese hat nun die Form

$$\frac{(x + y)^2}{1 + (x + y)^2} - \frac{1}{1 + (x + y)^2} y' = 0.$$

Das Potential f muss aus den beiden Gleichungen

$$\frac{(x + y)^2}{1 + (x + y)^2} = \frac{\partial f(x, y)}{\partial x}, \quad - \frac{1}{1 + (x + y)^2} = \frac{\partial f(x, y)}{\partial y}$$

bestimmt werden. Integration der zweiten Gleichung liefert

$$f(x, y) = - \arctan(x + y) + c(x).$$

Leiten wir dies nach x ab und setzen es gleich der ersten Bedingung, so erhalten wir $c(x) = x$ (bis auf Integrationskonstante). Daher muss die Lösung $y = y(x)$ der impliziten Gleichung

$$x - \arctan(x + y(x)) = C$$

mit beliebiger Konstanten C genügen. Rein formal ist zunächst

$$y(x) = \tan(x - C) - x$$

Lösung dieser Gleichung. Nun muss schließlich C aus einer Anfangsbedingung bestimmt werden. Dies legt auch das Definitionsgebiet der Lösung fest.

Bemerkung: Eine Differentialgleichung

$$p(x, y(x)) + q(x, y(x)) y'(x) = 0, \quad x \in I,$$

wird häufig auch in differentieller Form

$$df = p(x, y) dx + q(x, y) dy = 0, \quad x \in I,$$

geschrieben (s. Definition 1.22). Die Differentialform heißt **exaktes** oder **vollständiges Differential**, wenn die Integrabilitätsbedingung (1.2) gilt. Ansonsten sprechen wir von einem **unvollständigen** Differential. Die differenzielle Form wird oft in den Anwendungen genutzt, da die Differentialgleichung in dieser Form symmetrisch ist, d.h. wir können Lösungen sowohl als Funktion $y = y(x)$ als auch umgekehrt als $x = x(y)$ betrachten – je nachdem, welche Abhängigkeit gesucht ist.

Der letzte Schritt beim Lösen solcher Differentialgleichungen besteht darin, die implizite Gleichung $f(x, y) = C$ aufzulösen. Physikalische und technische Größen sind meistens nichtlinear miteinander verknüpft und solche Gleichungen möchte man nach einer der auftretenden Variablen auflösen. Diese Auflösbarkeit wollen wir jetzt untersuchen. Zunächst betrachten wir zwei Beispiele.

Beispiele 1.29

(a) Sei $f(x, y) = 2x^2 + 3y$, $x, y \in \mathbb{R}$. Dann ist $\{(x, y) \in \mathbb{R}^2 : f(x, y) = 0\}$ genau der Graph der Funktion $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, definiert durch $\varphi(x) = -\frac{2}{3}x^2$. Es gilt also $f(x, \varphi(x)) = 0$ für alle $x \in \mathbb{R}$. Die Funktion φ ist daher durch die Gleichung $f(x, y) = 0$ **implizit** gegeben. Statt φ schreiben wir wieder y , also $f(x, y(x)) = 0$.

(b) Sei $f(x, y) = x^2 - y^2 + 1$, $x, y \in \mathbb{R}$. Dann gilt:

$f(x, y) = 0 \iff y = \sqrt{1 + x^2}$ oder $y = -\sqrt{1 + x^2}$. Also ist $\{(x, y) : f(x, y) = 0\}$ nicht der Graph **einer** Funktion $\mathbb{R} \rightarrow \mathbb{R}$. „Lokal“ d.h. in einer kleinen Umgebung z.B. der Punkte $(x, y) = (0, 1)$ oder $(x, y) = (0, -1)$, kommt aber nur eine der beiden Kurven in Frage. Es kann also lokal nach y aufgelöst werden.

Der folgende Satz gibt ein Kriterium dafür an, wann lokal aufgelöst werden kann.

Satz 1.30 (Satz über implizite Funktionen in \mathbb{R}^2)

Sei $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ k -mal stetig differenzierbar, D offen und $(\hat{x}, \hat{y}) \in D$. Ferner sei

$$f(\hat{x}, \hat{y}) = 0 \quad \text{und} \quad \frac{\partial f}{\partial y}(\hat{x}, \hat{y}) \neq 0.$$

Dann gibt es offene Intervalle $U \subset \mathbb{R}$ und $V \subset \mathbb{R}$ die \hat{x} bzw. \hat{y} enthalten sowie genau eine Abbildung $\varphi : U \rightarrow V$ mit $\varphi(\hat{x}) = \hat{y}$ und $f(x, \varphi(x)) = 0$ für alle $x \in U$. Die Abbildung φ ist k -mal stetig differenzierbar auf U . Die Ableitung ergibt sich durch Differenzieren der Gleichung $f(x, \varphi(x)) = 0$ mit der Kettenregel,

$$\frac{d}{dx}(f(x, \varphi(x))) = \frac{\partial f}{\partial x}(x, \varphi(x)) + \frac{\partial f}{\partial y}(x, \varphi(x)) \varphi'(x) = 0 \quad \text{bzw.} \quad \varphi'(x) = -\frac{f_x(x, \varphi(x))}{f_y(x, \varphi(x))}.$$

Beweisidee: Es ist $f(\hat{x}, \hat{y}) = 0$ und z.B. $f_y(\hat{x}, \hat{y}) > 0$. Dann existiert $\varepsilon > 0$ mit

$$f(\hat{x}, \hat{y} + \varepsilon) > 0, \quad f(\hat{x}, \hat{y} - \varepsilon) < 0 \quad \text{und} \quad f_y(\hat{x}, \hat{y} + t\varepsilon) > 0$$

für alle $|t| \leq 1$. Da f und f_y stetig sind, so existiert eine offene Umgebung U um \hat{x} mit

$$f(x, \hat{y} + \varepsilon) > 0, \quad f(x, \hat{y} - \varepsilon) < 0 \quad \text{und} \quad f_y(x, \hat{y} + t\varepsilon) > 0$$

für alle $x \in U$ und $|t| \leq 1$. Der Zwischenwertsatz (Satz 3.23 von HM I), angewandt auf $f(x, \cdot) : \mathbb{R} \rightarrow \mathbb{R}$ liefert nun zu jedem $x \in U$ genau ein $y \in [\hat{y} - \varepsilon, \hat{y} + \varepsilon]$ mit $f(x, y) = 0$. Damit ist zumindest die Auflösbarkeit gezeigt. Technisch schwieriger ist es, die Differenzierbarkeit zu beweisen. Wir verzichten hier auf eine Darstellung dieses Beweises. \square

Beispiele 1.31 (a) Sei wieder $f(x, y) = x^2 - y^2 + 1$. Dann ist $f(0, 1) = 0$ und $f_y(0, 1) = -2 \neq 0$. Daher gibt es Intervalle I um 0 und J um 1 mit $x^2 - y(x)^2 + 1 = 0$ für alle $x \in I$ und $y(0) = 1$. In diesem einfachen Beispiel kann man natürlich die Form von y explizit angeben und hat $y(x) = +\sqrt{1+x^2}$. Analog kann man um den Punkt $(0, -1)$ auflösen und erhält $y(x) = -\sqrt{1+x^2}$.

(b) Sei $f(x, y) = y + xy^2 - \exp(xy)$, $x, y \in \mathbb{R}$. Es ist $f(0, 1) = 0$, und $f_y(x, y) = 1 + 2xy - x \exp(xy)$, also $f_y(0, 1) = 1 \neq 0$. Daher kann lokal um $x = 0$ die Gleichung $f(x, y) = 0$ nach y aufgelöst werden. Im Gegensatz zum Beispiel (a) können wir keine explizite Form angeben. Trotzdem können wir die Ableitungen von y an der Stelle $x = 0$ berechnen. Wir leiten die Gleichung $y + xy^2 - \exp(xy) = 0$ nach x ab und beachten, dass $y = y(x)$ von x abhängt:

$$y' + y^2 + 2xyy' - e^{xy}(y + xy') = 0. \quad (1.4)$$

Also ist für $x = 0$ unter Beachtung von $y(0) = 1$: $y'(0) + 1 - 1 = 0$, also $y'(0) = 0$. Wir leiten (1.4) noch einmal ab:

$$y'' + 4yy' + 2x(y')^2 + 2xyy'' - e^{xy}(y + xy')^2 - e^{xy}(2y' + xy'') = 0,$$

also mit $y(0) = 1$ und $y'(0) = 0$: $y''(0) - 1 = 0$. Damit haben wir mit der **Taylorformel** (Satz 5.31 von HM I): $y(x) = 1 + \frac{1}{2}x^2 + R_2(x)$, und der Rest $R_2(x)$ ist „von der Ordnung“ x^3 . Lokal um $x = 0$ haben wir also eine gute Vorstellung von y .

Die Kombination der Taylorformel mit dem Satz über implizite Funktionen ist mit die wichtigste Anwendung der Taylorformel!

Bemerkung: Der Satz über implizite Funktionen lässt sich auf Funktionen $f : D \subset \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ verallgemeinern, d.h. auf m nichtlineare Gleichungen. Die Bedingung an die Ableitung aus Satz 1.30 bedeutet im allgemeinen Fall, dass die Funktionalmatrix von f bezüglich y an der Stelle (\hat{x}, \hat{y}) invertierbar sein muss. Analog lassen sich auch in diesem Fall die partiellen Ableitungen der implizit gegebenen Funktion y an der Stelle $\hat{x} \in \mathbb{R}^n$ bestimmen.

1.4 Extremwertaufgaben

Im ganzen Abschnitt sei $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ eine stetige Funktion und $D \subset \mathbb{R}^n$ eine abgeschlossene Menge.

Definition 1.32 (Extremstellen)

Sei $D \subset \mathbb{R}^n$ eine Menge. $\hat{x} \in D$ heißt **lokales Minimum** von f , falls es eine (offene) Kugel $K(\hat{x}, \varepsilon)$ um \hat{x} gibt mit $f(\hat{x}) \leq f(x)$ für alle $x \in K(\hat{x}, \varepsilon) \cap D$. Analog werden **lokale Maxima** definiert. Die lokalen Minima und Maxima heißen **Extremstellen** von f und die Werte $f(\hat{x})$ **Extremwerte**.

Kann an Stelle von $K(\hat{x}, \varepsilon)$ die ganze Menge D genommen werden, so heißen die Extremstellen **globale Minima** bzw. **Maxima**.

Die Existenz von Extremwerten stetiger Funktionen auf kompakten Mengen ist durch Satz 1.12 gewährleistet.

Satz 1.33

Ist $D \subseteq \mathbb{R}^n$ **kompakt** (d.h. beschränkt und abgeschlossen) und $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ **stetig**, dann existieren globale Minima und Maxima.

Für das Folgende erinnern wir uns an die Definition des Inneren D° einer Menge $D \subset \mathbb{R}^n$: Eine Stelle $\hat{x} \in \mathbb{R}^n$ ist in D° genau dann, wenn es eine Kugel $K(\hat{x}, \epsilon)$ mit Mittelpunkt \hat{x} und Radius $\epsilon > 0$ gibt mit $K(\hat{x}, \epsilon) \subset D$. Wir betrachten nun Extremalstellen im Inneren von D und verallgemeinern den eindimensionalen Fall (Satz 6.21 in HM I).

Satz 1.34 (notwendiges Optimalitätskriterium)

Sei $\hat{x} \in D^\circ$ eine lokale Extremstelle von $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}$ im Innern von D , und sei f partiell differenzierbar in \hat{x} . Dann gilt $\nabla f(\hat{x}) = 0$ bzw. $f'(\hat{x}) = 0$, d.h. $\partial f(\hat{x})/\partial x_j = 0$ für alle $j = 1, \dots, n$.

Beweis: Der Beweis ist ganz einfach, wenn wir für einen beliebigen Vektor $a \in \mathbb{R}^n$, $a \neq 0$, die Funktion $\psi(t) = f(\hat{x} + ta)$, $t \in (-\epsilon, \epsilon)$, betrachten. Hier sei $\epsilon > 0$ so klein gewählt, dass $\hat{x} + ta \in D$ für alle $t \in (-\epsilon, \epsilon)$. Da \hat{x} eine lokale Extremstelle von f ist, so ist $t = 0$ eine lokale Extremstelle von ψ in $(-\epsilon, \epsilon) \subset \mathbb{R}$. Anwendung von Satz 5.20 aus HM I liefert $\psi'(0) = 0$, also mit der Kettenregel $\nabla f(\hat{x})^\top a = 0$. Dies gilt für *jeden* Vektor a . Für $a = \nabla f(\hat{x})$ folgt insbesondere $\nabla f(\hat{x}) = 0$. \square

Bemerkung: Dieser Satz drückt eine *notwendige* Bedingung für Optimalität aus, die i.a. nicht hinreichend ist (siehe schon das Beispiel für $n = 1$: $f(x) = x^3$ an der Stelle $\hat{x} = 0$). Wir nennen die Stellen x mit $\nabla f(x) = 0$ die **kritischen Punkte** von f .

Auch die hinreichende Bedingung an die zweite Ableitung, die wir aus HM 1 (Satz 5.33) kennen, lässt sich übertragen. Dazu benötigen wir die Hesse-Matrix (siehe Beispiel 1.24) $H_f(\hat{x}) \in \mathbb{R}^{n \times n}$ von f an der Stelle \hat{x} , definiert durch $H_f(\hat{x})_{ij} = \frac{\partial^2 f(\hat{x})}{\partial x_j \partial x_i}$, $i, j = 1, \dots, n$. Mit dem Satz von Schwarz ist $H_f(\hat{x}) \in \mathbb{R}^{n \times n}$ offensichtlich eine symmetrische Matrix (d.h. $H_f(\hat{x})^\top = H_f(\hat{x})$). Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ heißt **positiv definit**, wenn $z^\top A z > 0$ gilt für alle $z \neq 0$. Diese Eigenschaft lässt sich auch über die Eigenwerte von A beschreiben:

Lemma 1.35 Eine symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ ist genau dann positiv definit, wenn alle Eigenwerte positiv sind.

Beweis: Sei A positiv definit und λ ein Eigenwert mit zugehörigem Eigenvektor $z \neq 0$. Dann ist $0 < z^\top A z = \lambda \|z\|_2^2$, also $\lambda > 0$. Sind umgekehrt alle Eigenwerte größer als Null, so auch der kleinste, und die positive Definitheit folgt aus

$$z^\top A z \geq \lambda_{\min} \|z\|_2^2 \quad \text{für alle } z \in \mathbb{R}^n.$$

Auf einen Beweis dieser Abschätzung verzichten wir hier. Man benötigt dazu den Satz über die „Hauptachsentransformation“, nämlich dass sich jede symmetrische Matrix $A \in \mathbb{R}^{n \times n}$ in der Form $A = V D V^\top$ schreiben lässt, wobei $V \in \mathbb{R}^{n \times n}$ orthogonal ist (d.h. $V^\top V = V V^\top = I$) und D die Diagonalmatrix ist, die alle n Eigenwerte von A enthält. \square

Satz 1.36 (*hinreichende Bedingung*) Ist $f : \mathbb{R}^n \rightarrow \mathbb{R}$ zweimal stetig differenzierbar in \hat{x} und ist $\nabla f(\hat{x}) = 0$ und $H_f(\hat{x})$ positiv definit, so ist \hat{x} ein lokales Minimum.

Beweis: Für festes $h \neq 0$ definieren wir die Funktion $\psi(t) = f(\hat{x} + th)$, $t \in \mathbb{R}$. Dann ist nach der mehrdimensionalen Kettenregel $\psi'(t) = f'(\hat{x} + th)h = \sum_{j=1}^n h_j \frac{\partial f(\hat{x} + th)}{\partial x_j}$ und noch einmal abgeleitet $\psi''(t) = \sum_{i,j=1}^n h_j h_i \frac{\partial^2 f(\hat{x} + th)}{\partial x_j \partial x_i} = h^\top H_f(\hat{x} + th)h$. Nach Voraussetzung ist $\psi'(0) = f'(\hat{x})h = 0$ und $\psi''(0) = h^\top H_f(\hat{x})h > 0$. Daher ist 0 lokales Minimum von ψ . Hieraus lässt sich über einen Widerspruchsbeweis schließen, dass \hat{x} lokales Minimum von f ist. \square

Beispiele 1.37

(a) Wir suchen Extremalstellen der Funktion

$$f(x) = 3x_1^2 - 2x_1x_2 + 3x_2^2 - 2x_1 - 2x_2 + 1 \quad \text{in } \mathbb{R}^2.$$

Für die kritischen Punkte gilt $\nabla f(\hat{x}) = 0$, d.h. wir betrachten das Gleichungssystem

$$\begin{aligned} 6x_1 - 2x_2 &= 2 \\ 6x_2 - 2x_1 &= 2 \end{aligned}$$

und erhalten die Lösung $x_1 = x_2 = \frac{1}{2}$. Dies ist ein globales Minimum; denn es gilt

$$f(x) = (x_1 + x_2 - 1)^2 + 2(x_1 - x_2)^2.$$

(b) Häufig soll zu m gemessenen Daten $(x_j, y_j) \in \mathbb{R} \times \mathbb{R}$, $j = 1, \dots, m$, eine möglichst einfache Funktion g bestimmt werden, die den Zusammenhang zwischen x und y "gut" widerspiegelt. Wir sprechen von einer *Ausgleichs- oder Regressionskurve*. Verschiedene Kriterien sind denkbar, was denn eine gute Näherung sein soll. Aber stets führt das Problem auf ein **Optimierungsproblem**. Ein wichtiges Vorgehen ist die **Methode der kleinsten Quadrate**:

Zunächst müssen wir ein Modell auswählen, etwa einen linearen Zusammenhang, $g(x) = p_1x + p_2$, (d.h. gesucht ist eine Ausgleichsgerade) oder ein exponentielles Verhalten, $g(x) = p_1 e^{p_2x}$, wobei die **Parameter** $p_1, p_2 \in \mathbb{R}$ noch zu bestimmen sind. Allgemein legen wir also eine Klasse von Funktionen fest, die von Parametern p_1, \dots, p_n abhängen. Die Funktionen können somit beschrieben werden durch eine Funktion $g : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$. Man versucht, die Parameter $p = (p_1, \dots, p_n) \in \mathbb{R}^n$ so zu bestimmen, dass

$$g(x_j, p) \approx y_j \quad \text{für alle } j = 1, \dots, m.$$

Die Methode der kleinsten Quadrate besteht nun darin, die **Summe der Abstandsquadrate**,

$$f(p) = \sum_{j=1}^m [y_j - g(x_j, p)]^2,$$

d.h. die Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$, zu minimieren. Also suchen wir $\hat{p} \in \mathbb{R}^n$ mit $\nabla f(\hat{p}) = 0$. Die Kettenregel liefert

$$\frac{\partial f}{\partial p_\ell}(p) = -2 \sum_{j=1}^m [y_j - g(x_j, \hat{p})] \frac{\partial g}{\partial p_\ell}(x_j, p), \quad \ell = 1, \dots, n.$$

Nullsetzen liefert n nichtlineare Gleichungen in den zu bestimmenden Parametern p_1, \dots, p_n . Diese Gleichungen werden **Normalgleichungen** genannt.

Die Normalgleichungen können kompliziert sein, je nach Wahl des Ansatzes für die Funktion g , und müssen in den meisten Fällen numerisch gelöst werden (s. nächster Abschnitt). Im einfachsten Fall einer Ausgleichsgeraden ist

$$g(x, p) = p_1 x + p_2, \quad x \in \mathbb{R}.$$

In diesem Fall führen die Normalgleichungen auf ein lineares Gleichungssystem in p_1 und p_2 . Wir erhalten

$$0 = \nabla f(p) = -2 \begin{pmatrix} \sum_{j=1}^m [y_j - p_1 x_j - p_2] x_j \\ \sum_{j=1}^m [y_j - p_1 x_j - p_2] \end{pmatrix}$$

bzw.

$$\begin{pmatrix} \left(\sum_{j=1}^m (x_j)^2 \right) & \left(\sum_{j=1}^m x_j \right) \\ \left(\sum_{j=1}^m x_j \right) & m \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m y_j x_j \\ \sum_{j=1}^m y_j \end{pmatrix}.$$

Mit

$$\begin{aligned} D &= \det \begin{pmatrix} \left(\sum_{j=1}^m (x_j)^2 \right) & \left(\sum_{j=1}^m x_j \right) \\ \left(\sum_{j=1}^m x_j \right) & m \end{pmatrix} = m \sum_{j=1}^m (x_j)^2 - \left(\sum_{j=1}^m x_j \right)^2 \\ &\geq m \sum_{j=1}^m (x_j)^2 - \left(\sum_{j=1}^m 1^2 \right) \left(\sum_{j=1}^m (x_j)^2 \right) = 0 \end{aligned}$$

nach der Cauchy-Schwarzschen Ungleichung und der Tatsache, dass sogar > 0 gilt², folgt, dass das lineare Gleichungssystem stets lösbar ist, und wir erhalten die Parameter

$$\begin{aligned} p_1 &= \frac{1}{D} \left[m \left(\sum_{j=1}^m y_j x_j \right) - \left(\sum_{j=1}^m x_j \right) \left(\sum_{j=1}^m y_j \right) \right] \\ p_2 &= \frac{1}{D} \left[\left(\sum_{j=1}^m y_j \right) \left(\sum_{j=1}^m (x_j)^2 \right) - \left(\sum_{j=1}^m y_j x_j \right) \left(\sum_{j=1}^m x_j \right) \right]. \end{aligned}$$

(c) Bestimme die Extremalstellen von $f(x_1, x_2) = x_1 x_2$ auf der Menge $D := \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$. Hier ist $\|\cdot\|_2$ wieder die euklidische Norm im \mathbb{R}^2 .

- (i) D ist abgeschlossen und beschränkt, also kompakt, und f ist stetig. Daher existieren globale Maxima und Minima.
- (ii) Sei \hat{x} Extremstelle. Ist $\hat{x} \in D^\circ$ (d.h. $\|\hat{x}\|_2 < 1$), so ist $\nabla f(\hat{x}) = 0$, d.h. $\hat{x}_2 = 0$ und $\hat{x}_1 = 0$, also $\hat{x} = (0, 0)^\top$. Diese Stelle ist aber weder Minimum noch Maximum (weshalb nicht?).

²Gleichheit besteht nur, wenn alle x_j gleich sind

(iii) Alle Extremstellen liegen daher auf dem Rand, d.h. $\|\hat{x}\|_2 = 1$. Also gibt es $\theta \in [0, 2\pi]$ mit $\hat{x} = (\cos \theta, \sin \theta)^\top$. Dann ist $f(\hat{x}) = \sin \theta \cos \theta = \frac{1}{2} \sin(2\theta)$. Dies ist maximal für $\theta = \pi/4$ oder $\theta = 5\pi/4$, also $\hat{x} = \frac{1}{\sqrt{2}}(1, 1)^\top$ bzw. $\hat{x} = -\frac{1}{\sqrt{2}}(1, 1)^\top$ (mit Maximalwert $1/2$), und minimal für $\theta = 3\pi/4$ oder $\theta = 7\pi/4$, also $\hat{x} = \frac{1}{\sqrt{2}}(-1, 1)^\top$ bzw. $\hat{x} = \frac{1}{\sqrt{2}}(1, -1)^\top$ (mit Minimalwert $-1/2$).

Das letzte Beispiel ist in gewissem Sinn typisch. Die Extremstellen liegen meistens auf dem **Rand** von $D \subset \mathbb{R}^n$, der allerdings i.A. komplizierter als in diesem Beispiel ist und in Anwendungen häufig durch ein System von Gleichungen beschrieben wird (in der Mechanik sind dies z.B. die „Zwangskräfte“). Wir betrachten also jetzt Mengen D der Form:

$$D = \{x \in \mathbb{R}^n : g_i(x) = 0, i = 1, \dots, m\}$$

wobei $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ eine gegebene stetig differenzierbare Funktion sei.

Wir wollen jetzt also die **Zielfunktion** $f : \mathbb{R}^n \rightarrow \mathbb{R}$ auf der so beschriebenen Menge D minimieren oder maximieren. Wir können uns auf Minimierungsaufgaben beschränken (sonst ersetze man f durch $-f$).

Beispiel: Bestimme den Abstand zwischen der Ellipse $x^2 + \frac{1}{4}y^2 = 1$ und der Kurve $y = 3 \cos(\pi x/4)$. Man skizziere diese Kurven!

Es ist also $f(x, y, u, v) := (x-u)^2 + (y-v)^2$ zu minimieren unter den Nebenbedingungen $g_1(x, y, u, v) := x^2 + \frac{1}{4}y^2 - 1 = 0$ (d.h. (x, y) liegt auf der Ellipse), und $g_2(x, y, u, v) := v - 3 \cos(\pi u/4) = 0$ (d.h. (u, v) liegt auf der Kosinuskurve). Hier ist also $n = 4$ und $m = 2$. Weiter unten gehen wir auf das Beispiel noch einmal ein.

Die **Lagrangefunktion** des Optimierungsproblems ist definiert durch

$$L(x, \lambda) := f(x) + \sum_{i=1}^m \lambda_i g_i(x), \quad x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m. \quad (1.5)$$

Dann gilt

Satz 1.38 (Lagrangesche Multiplikatorenregel)

Sei $U \subset \mathbb{R}^n$ offen und $\hat{x} \in D$ eine lokale Extremalstelle von f auf der Menge D , beschrieben durch

$$D = \{x \in U : g_i(x) = 0, i = 1, \dots, m\}.$$

Die „Restriktionsgradienten“ in \hat{x} , d.h.

$$\nabla g_1(\hat{x}), \dots, \nabla g_m(\hat{x}) \in \mathbb{R}^n$$

seien linear unabhängig. Dann existieren Zahlen $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ (sie heißen **Lagrangesche Multiplikatoren**) mit

$$\frac{\partial L(\hat{x}, \lambda)}{\partial x_j} = 0, \quad j = 1, \dots, n, \quad \frac{\partial L(\hat{x}, \lambda)}{\partial \lambda_i} = 0, \quad i = 1, \dots, m.$$

Ausgeschrieben heißt dies:

$$\nabla f(\hat{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\hat{x}) = 0, \quad g(\hat{x}) = 0.$$

Dies sind also $n + m$ Gleichungen in den $n + m$ Unbekannten $x_1, \dots, x_n, \lambda_1, \dots, \lambda_m$.

Mit diesem Satz bietet sich folgendes Vorgehen an, um Extrema von f auf D zu bestimmen:

1. Man berechne alle $\hat{x} \in D$, zu denen es $\alpha \in \mathbb{R}^m$ gibt mit $\sum_{j=1}^m \alpha_j \nabla g_j(\hat{x}) = 0$. Im einfachsten Fall $m = 1$ bedeutet dies, gesucht sind Lösungen \hat{x} der nichtlinearen Gleichungen $\nabla g(\hat{x}) = 0$. Diese Stellen sammeln wir als mögliche Kandidaten.
2. Weiterhin berechnen wir alle Lösungen $\hat{x} \in D$ und $\hat{\lambda}_j \in \mathbb{R}$ mit $\nabla_x L(\hat{x}, \hat{\lambda}) = 0$. Nach dem letzten Satz kommen diese Stellen \hat{x} auch als Extremalstellen infrage.
3. Im dritten Schritt vergleichen wir zu all den gefundenen Kandidaten \hat{x} aus dem ersten und dem zweiten Schritt die Funktionswerte $f(\hat{x})$ um Maxima bzw. Minima zu identifizieren.

Beispiele 1.39

(a) (siehe oben) Bestimme Minima und Maxima von $f(x, y) = xy$ auf dem Einheitskreis $x^2 + y^2 = 1$. Hier wollen wir noch einmal alle Schritte vollständig durchgehen.

- (i) Existenz: Da der Einheitskreis kompakt und die Zielfunktion stetig ist, so existieren globale Minima und Maxima (weshalb?).
- (ii) Sei $(x, y)^\top$ Extremalpunkt. Es ist $g(x, y) = x^2 + y^2 - 1 = 0$. Die Lagrangefunktion ist $L(x, y; \lambda) = xy + \lambda(x^2 + y^2 - 1)$. Da $\nabla g(x, y) \neq 0$ für $x^2 + y^2 = 1$, so ist die Multiplikatorenregel anwendbar, d.h. $\nabla L(x, y; \lambda) = 0$, d.h. man erhält 3 Gleichungen mit 3 Unbekannten:

$$y + 2\lambda x = 0, \quad x + 2\lambda y = 0, \quad x^2 + y^2 = 1.$$

Multiplikation der ersten Gleichung mit y , der zweiten mit x und Subtraktion liefert $x^2 = y^2$, also $x = \pm y$. Einsetzen in die dritte Gleichung liefert als Lösungsmenge

$$\mathcal{L} = \left\{ (x, y) = \frac{1}{\sqrt{2}}(\varepsilon, \mu) : \varepsilon, \mu \in \{+1, -1\} \right\}.$$

Wegen $f\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right) = \frac{1}{2}$ und $f\left(\pm\frac{1}{\sqrt{2}}, \mp\frac{1}{\sqrt{2}}\right) = -\frac{1}{2}$ sind $\pm\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^\top$ die zwei (globalen) Maxima und $\pm\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)^\top$ die zwei (globalen) Minima.

(b) (siehe oben) $n = 4$, $m = 2$ und $f(x, y, u, v) := (x - u)^2 + (y - v)^2$ ist zu minimieren unter den Nebenbedingungen

$$g_1(x, y, u, v) := x^2 + \frac{1}{4}y^2 - 1 = 0 \quad \text{und} \quad g_2(x, y, u, v) := v - 3\cos(\pi u/4) = 0.$$

Die Multiplikatorenregel ist anwendbar (weshalb?) und liefert:

$$2 \begin{pmatrix} x - u \\ y - v \\ u - x \\ v - y \end{pmatrix} + \lambda_1 \begin{pmatrix} 2x \\ .5y \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 0 \\ 3\pi/4 \sin(\pi u/4) \\ 1 \end{pmatrix} = 0.$$

Zusammen mit $x^2 + y^2/4 - 1 = 0$ und $v - 3\cos(\pi u/4) = 0$ sind dies 6 Gleichungen in 6 Unbekannten, die wir als System

$$F(x, y, u, v, \lambda_1, \lambda_2) = 0$$

mit $F : \mathbb{R}^6 \rightarrow \mathbb{R}^6$ schreiben. Dieses Gleichungssystem kann z.B. mit dem Newtonverfahren gelöst werden (s.u.). Startwerte für x , y , u und v lesen wir aus einer Skizze ab: $(x^{(0)}, y^{(0)}, u^{(0)}, v^{(0)}) = (1, 1, 1.5, 1.5)$. Startwerte für die Lagrangeschen Multiplikatoren $\lambda_1^{(0)}$ und $\lambda_2^{(0)}$ ermitteln wir aus der ersten und vierten Gleichung: $(\lambda_1^{(0)}, \lambda_2^{(0)}) = (0.5, -1)$. Damit erhalten wir das folgende Ergebnis (nicht aufgetragen sind die Multiplikatoren):

$x^{(k)}$	$y^{(k)}$	$u^{(k)}$	$v^{(k)}$	$\ F(x^{(k)})\ $
1.0000000	1.0000000	1.5000000	1.5000000	1.4607603
0.6497072	1.9011712	1.0376180	2.1545819	0.35625797
0.7690029	1.3954334	1.2772296	1.6468979	0.09581969
0.6921286	1.4528479	1.2203884	1.7259115	0.01419526
0.6954022	1.4373400	1.2267687	1.7119649	0.00012065
0.6952968	1.4374454	1.2266907	1.7120941	2.8665e-08
0.6952968	1.4374453	1.2266908	1.7120941	6.8438e-16

(c) Ungleichung vom arithmetischen und geometrischen Mittel

Für alle $x_j \geq 0$, $j = 1, \dots, n$, gilt

$$\frac{1}{n} \sum_{j=1}^n x_j \geq \left(\prod_{j=1}^n x_j \right)^{1/n}.$$

Um dies zu zeigen, betrachten wir das folgende Maximierungsproblem:

$$\text{Maximiere } f(x) = \prod_{j=1}^n x_j \quad \text{unter } x_j \geq 0, \quad \sum_{j=1}^n x_j = 1.$$

(i) Es existiert eine Lösung, da $M_1 = \left\{ x \in \mathbb{R}^n : x_j \geq 0, \sum_{j=1}^n x_j = 1 \right\}$ abgeschlossen und beschränkt, also kompakt, ist und $f(x) = \prod_{j=1}^n x_j$ stetig ist. Sei x eine Maximalstelle. Dann ist

$x_j > 0$ für alle $j = 1, \dots, n$, da sonst $f(x) = 0$ ist und dies sicher kein Maximum ist.

(ii) Also können wir uns mit $U = \{x \in \mathbb{R}^n : x_j > 0 \forall j\}$ und $g(x) = (\sum_{j=1}^n x_j) - 1$ auf die Menge M , gegeben durch $M = \{x \in U : g(x) = 0\}$, beschränken. Es ist $\nabla g(x) = (1, \dots, 1)^\top \neq 0$, daher ist der „Restriktionsgradient“ linear unabhängig.

Anwendung der Lagrangeschen Multiplikatorenregel liefert die Existenz von $\lambda \in \mathbb{R}$ mit $\nabla f(x) + \lambda \nabla g(x) = 0$, d.h.

$$\frac{\partial}{\partial x_i} \prod_{j=1}^n x_j + \lambda = 0, \quad i = 1, \dots, n.$$

Somit gilt für jedes $i = 1, \dots, n$:

$$\prod_{j \neq i} x_j + \lambda = 0, \quad \text{bzw.} \quad \prod_{j=1}^n x_j + \lambda x_i = 0.$$

Daher sind alle x_i gleich (nämlich $-(\prod_{j=1}^n x_j)/\lambda$). Aus der Bedingung $\sum_{i=1}^n x_i = 1$ folgt $x_i = \frac{1}{n}$ für alle i und $f(x) = \frac{1}{n^n}$. Daher gilt $\prod_{j=1}^n x_j \leq \frac{1}{n^n}$ für alle $x \in M$.

(ii) Nun zeigen wir die Ungleichung für $x \in \mathbb{R}^n$ beliebig mit $x_i > 0$ für alle i . Dann ist $\frac{x_i}{\sum_{j=1}^n x_j} \in M$ und daher

$$\prod_{i=1}^n \frac{x_i}{\sum_{j=1}^n x_j} \leq \frac{1}{n^n}, \quad \text{d.h.} \quad \prod_{i=1}^n x_i \leq \left(\frac{\sum_{j=1}^n x_j}{n} \right)^n,$$

und dies liefert die Behauptung nach Ziehen der n -ten Wurzel. □

Bemerkung: Es ist zu beachten, dass die Multiplikatorenregel eine notwendige Bedingung ist, d.h. eine Extremalstelle ist ein kritischer Punkt der Lagrangefunktion, zumindest wenn die Voraussetzung der linearen Unabhängigkeit der Gradienten $\nabla g_j(\hat{x})$ gewährleistet ist. In Punkten, wo diese Voraussetzung verletzt ist, könnten Extrema liegen, zu denen es keine Multiplikatoren gibt. Diese Punkte müssen separat betrachtet werden (s. Übung).

1.5 Nichtlineare Gleichungssysteme

Die Menge zulässiger Punkte eines Optimierungsproblems $D = \{x \in \mathbb{R}^n : g(x) = 0\}$ haben wir als die Lösungen eines nichtlinearen Gleichungssystems beschrieben. Auch ist es erforderlich nichtlineare Gleichungssysteme zu lösen, um notwendige Optimalitätsbedingungen wie $\nabla f(x) = 0$ oder $\nabla L(x, \lambda) = 0$ zu nutzen. Wir betrachten daher allgemein Systeme von m Gleichungen in n Unbekannten, also Gleichungssysteme der Form

$$\begin{aligned} f_1(x_1, \dots, x_n) &= y_1, \\ &\vdots \\ f_m(x_1, \dots, x_n) &= y_m. \end{aligned}$$

Kürzer können wir so formulieren: Gegeben sei $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Wir suchen eine oder alle Lösungen $\hat{x} \in D$ der Gleichung $f(x) = y$.

Beispiel 1.40

Finde $x_1, x_2 \in \mathbb{R}$ mit

$$(x_1 - 1)^2 + (x_2 - 2)^2 = 4 \quad \text{und} \quad x_1^3 - x_2 = 0.$$

In diesem Beispiel sind $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ und $y \in \mathbb{R}^2$ gegeben durch

$$f(x) = \begin{pmatrix} (x_1 - 1)^2 + (x_2 - 2)^2 \\ x_1^3 - x_2 \end{pmatrix}, \quad x \in \mathbb{R}^2 \quad \text{und} \quad y = \begin{pmatrix} 4 \\ 0 \end{pmatrix}.$$

Wir können in diesem Beispiel die Lösungsmenge anschaulich klären. Die erste Gleichung beschreibt einen Kreis

$$M_1 = \{x \in \mathbb{R}^2 : (x_1 - 1)^2 + (x_2 - 2)^2 = 4\} = \{x \in \mathbb{R}^2 : \|x - (1, 2)^\top\|_2 = 2\},$$

d.h. alle Punkte mit Abstand 2 vom Punkt $(1, 2)^\top$, also einen Kreis mit Radius $r = 2$ und Mittelpunkt $(1, 2)^\top$. Hingegen sind Lösungen der zweiten Gleichung die Punkte auf dem Graphen der Funktion $h(x_1) = x_1^3$, d.h. $M_2 = \{x \in \mathbb{R}^2 : x_2 = x_1^3\}$. Lösung des Gleichungssystems $f(x) = y$ sind somit gegeben durch alle Punkte in der Schnittmenge $M_1 \cap M_2 \subseteq \mathbb{R}^2$ (Skizze!).

Im Fall $n = m$, d.h. es gibt gleich viele Gleichungen wie Variable, können wir auf eindeutige Lösungen hoffen, wie bei invertierbaren linearen Gleichungssystemen. Eine (anschauliche) Charakterisierung der Lösungsmenge, wie in den obigen Beispielen, bzw. nur allein die theoretische Frage nach Existenz und/oder Eindeutigkeit von Lösungen ist oft aufwendig. Genauso ist die numerische Berechnung von Lösungen im Allgemeinen nicht geschlossen möglich. Wir sind auf **Näherungsverfahren** angewiesen, d.h. es wird versucht eine Folge $(x^{(k)})_{k \in \mathbb{N}}$ in $D \subseteq \mathbb{R}^n$ zu konstruieren, die (möglichst „schnell“) gegen eine Lösung $\hat{x} \in D$ konvergiert – falls solch eine Lösung überhaupt existiert. Zwei grundlegende numerische Lösungsverfahren wollen wir besprechen.

1. Das Kontraktionsverfahren

Das Kontraktionsverfahren ist ein Verfahren zur Lösung von **Fixpunktgleichung**

$$g(\hat{x}) = \hat{x}$$

zu $g : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$. Beachte: Gleichungen der Form $f(\hat{x}) = y$ können in vielfältiger Weise umgeschrieben werden in Form einer Fixpunktgleichung und umgekehrt, etwa indem $g(x) = x + a(f(x) - y)$, $x \in \mathbb{R}^n$, für beliebigen Parameter $a \neq 0$ gesetzt wird.

Für gegebenes $x^{(0)}$ bilden wir rekursiv die **Iterationsfolge**

$$x^{(k+1)} = g(x^{(k)}), \quad k = 0, 1, \dots$$

Wenn die Folge konvergiert, sprechen wir von **sukzessiver Approximation**. Für die Konvergenzuntersuchung benötigen wir wieder einen Abstandsbegriff, also eine Norm $\|\cdot\|$ im \mathbb{R}^n . Wir können etwa die euklidische Norm, $\|x\| = \|x\|_2 = \sqrt{\sum_{j=1}^n |x_j|^2}$, oder die Maximumnorm $\|x\| = \|x\|_\infty = \max_{j=1, \dots, n} |x_j|$ wählen. Man vergleiche die Beispiele nach Definition 1.4

Satz 1.41 (*Banachscher Fixpunktsatz oder Kontraktionssatz im \mathbb{R}^n*)

Sei D eine **abgeschlossene** Menge im \mathbb{R}^n , $\|\cdot\|$ eine Norm im \mathbb{R}^n , und die Funktion $g : \mathbb{R}^n \supset D \rightarrow \mathbb{R}^n$ habe die beiden Eigenschaften:

(i) $g(x) \in D$ für alle $x \in D$, d.h. g bilde D in sich ab, und

(ii) g ist eine **Kontraktion**, d.h. es gibt ein $\kappa \in (0, 1)$ mit

$$\|g(x) - g(y)\| \leq \kappa \|x - y\| \quad \text{für alle } x, y \in D.$$

Dann besitzt g genau einen Fixpunkt $\hat{x} \in D$, d.h. $g(\hat{x}) = \hat{x}$, und die Folge $(x^{(k)})_k$, definiert durch $x^{(k+1)} = g(x^{(k)})$, $k = 0, 1, \dots$, konvergiert gegen \hat{x} für jeden beliebigen Startwert $x^{(0)} \in D$. Es gelten die Fehlerabschätzungen:

$$\begin{aligned} \text{(iii)} \quad \|x^{(k)} - \hat{x}\| &\leq \frac{\kappa^k}{1 - \kappa} \|x^{(1)} - x^{(0)}\| \quad (\text{a priori}), \\ \text{(iv)} \quad \|x^{(k)} - \hat{x}\| &\leq \frac{\kappa}{1 - \kappa} \|x^{(k)} - x^{(k-1)}\| \quad (\text{a posteriori}). \end{aligned}$$

Beweis: Wir wollen den Beweis dieses wichtigen Satzes für den eindimensionalen Fall, d.h. für $n = 1$, führen.³ Statt $x^{(k)}$ schreiben wir wieder x_k . Die Norm ist dann einfach der Betrag.

(a) Es kann keine zwei verschiedenen Fixpunkte geben. Sind nämlich $\hat{x}, \tilde{x} \in D$ Fixpunkte von g , so ist

$$|\hat{x} - \tilde{x}| = |g(\hat{x}) - g(\tilde{x})| \leq \kappa |\hat{x} - \tilde{x}|,$$

also $(1 - \kappa)|\hat{x} - \tilde{x}| \leq 0$. Hieraus folgt $|\hat{x} - \tilde{x}| = 0$, d.h. $\hat{x} = \tilde{x}$.

(b) Mit $x_k \in D$ ist auch $x_{k+1} = g(x_k) \in D$. Daher liegt die ganze Folge $(x_k)_k$ in D . Es ist

$$|x_{k+1} - x_k| = |g(x_k) - g(x_{k-1})| \leq \kappa |x_k - x_{k-1}|,$$

also $|x_{k+1} - x_k| \leq \kappa^k |x_1 - x_0|$ für $k = 0, 1, 2, \dots$ (Man schreibe sich x_1, x_2, x_3 und x_4 hin, und beweise die Ungleichung mit vollständiger Induktion). Jetzt schreiben wir x_m in der Form als „Teleskopsumme“

$$x_m = x_0 + \sum_{k=0}^{m-1} (x_{k+1} - x_k), \quad m \in \mathbb{N}. \quad (1.6)$$

Die Partialsummen auf der rechten Seite konvergieren absolut, wie wir mit dem Majorantenkriterium sehen:

$$\sum_{k=0}^{m-1} |x_{k+1} - x_k| \leq |x_1 - x_0| \sum_{k=0}^{m-1} \kappa^k,$$

und die geometrische Summe konvergiert wegen $\kappa < 1$. Also konvergiert die Folge (x_m) gegen ein $\hat{x} \in D$, und \hat{x} hat die Darstellung

$$\hat{x} = x_0 + \sum_{k=0}^{\infty} (x_{k+1} - x_k). \quad (1.7)$$

Aus der Gleichung $x_{k+1} = g(x_k)$ und der Stetigkeit von g folgt $\hat{x} = g(\hat{x})$, d.h. \hat{x} ist Fixpunkt.

³Der Beweis im \mathbb{R}^n verläuft ganz genauso!

(c) Wir schätzen ab, indem wir (1.6) und (1.7) voneinander subtrahieren:

$$|\hat{x} - x_m| \leq \sum_{k=m}^{\infty} |x_{k+1} - x_k| \leq |x_1 - x_0| \sum_{k=m}^{\infty} \kappa^k = \frac{\kappa^m}{1 - \kappa} |x_1 - x_0|,$$

also die a priori Abschätzung. Benutzen wir in der letzten Formel die Abschätzung

$$|x_{k+1} - x_k| \leq \kappa^{k-m+1} |x_m - x_{m-1}|, \quad k \geq m - 1,$$

so folgt

$$|\hat{x} - x_m| \leq |x_m - x_{m-1}| \sum_{k=m}^{\infty} \kappa^{k-m+1} = \frac{\kappa}{1 - \kappa} |x_m - x_{m-1}|. \quad \square$$

Bemerkung: (a) Um die Kontraktionseigenschaft (ii) nachzuweisen, bietet sich oft der Mittelwertsatz an, angewandt auf die Funktion $G(t) := g_i(x + t(y - x))$. Hier ist $i \in \{1, \dots, n\}$ festgehalten. Es gibt danach ein $\tau \in (0, 1)$ mit

$$G(1) - G(0) = G'(\tau) = \nabla g_i(x + \tau(y - x))^\top (y - x),$$

und somit zu $x, y \in D$ ein $z = x + \tau(y - x) \in D$ („zwischen x und y “) mit

$$g_i(y) - g_i(x) = G(1) - G(0) = \nabla g_i(z)^\top (y - x) = \sum_{j=1}^n \frac{\partial g_i(z)}{\partial x_j} (y_j - x_j).$$

Für jedes i bekommen wir i.A. ein anderes z . Besonders einfach gewinnen wir hieraus die Kontraktionseigenschaft bezüglich der **Maximumnorm** $\|z\|_\infty = \max_{j=1, \dots, n} |z_j|$, denn:

$$|g_i(y) - g_i(x)| \leq \sum_{j=1}^n \left| \frac{\partial g_i(z)}{\partial x_j} \right| |y_j - x_j| \leq \|y - x\|_\infty \max_{z \in D} \sum_{j=1}^n \left| \frac{\partial g_i(z)}{\partial x_j} \right|.$$

Diese Abschätzung benutzen wir für jede Komponente g_i , d.h.

$$\|g(y) - g(x)\|_\infty \leq \underbrace{\max_{i=1, \dots, n} \max_{z \in D} \sum_{j=1}^n \left| \frac{\partial g_i(z)}{\partial x_j} \right|}_{=: \kappa} \|y - x\|_\infty.$$

Ist $\kappa < 1$, so ist g eine Kontraktion bzgl. der Maximumnorm mit der Kontraktionskonstanten κ . Im eindimensionalen Fall $n = 1$ reduziert sich die Bedingung einfach auf

$$\kappa := \max_{z \in D} |g'(z)| < 1.$$

(b) Die Kontraktionseigenschaft einer Funktion lässt sich auch als Lipschitz-Stetigkeit mit einer Lipschitz-Konstante $c < 1$ beschreiben (s. Def. 3.15 in HM I).

(c) Beachten Sie, dass der Satz die Existenz genau eines Fixpunktes belegt. Der Fixpunktsatz dient somit nicht nur als Möglichkeit Lösungen zu berechnen, sondern ist ein genereller Weg, um die Existenz von Lösungen zu nichtlinearen Gleichungen zu beweisen.

(d) Der Fixpunktsatz gilt analog auch im Vektorraum der stetigen Funktionen. Damit lässt sich dann der Satz von Picard-Lindelöf (s. HM II, Satz 4.7) beweisen.

Beispiele 1.42

(a) Bestimme $x \in \mathbb{R}$ mit $x = \frac{1}{2} \cos x$. Hier ist $g(x) = \frac{1}{2} \cos x$ und

$$|g(x) - g(y)| = |g'(z)| |x - y| \leq \frac{1}{2} |x - y|,$$

da $|g'(z)| = \frac{1}{2} |\sin z| \leq \frac{1}{2}$. Also können wir $D = \mathbb{R}$ und $\kappa = \frac{1}{2}$ nehmen. Die Folge $x^{(k+1)} = \frac{1}{2} \cos x^{(k)}$, $k = 0, 1, \dots$, konvergiert für jeden Anfangswert $x^{(0)} \in \mathbb{R}$ gegen den eindeutig bestimmten Fixpunkt \hat{x} , denn $x^{(1)} \in [-1/2, 1/2]$ und g ist auf diesem Intervall eine Selbstabbildung. Für $x^{(0)} = 0$ liefert die Berechnung die Folge:

k	$x^{(k)}$	$ x^{(k)} - x^{(k-1)} \frac{\kappa}{1-\kappa}$
0	0	
1	0.5000000	0.5000e - 00
2	0.4387912	0.6120e - 01
3	0.4526329	0.1384e - 01
4	0.4496493	0.2983e - 02
	\vdots	\vdots
9	0.4501836	0.3170e - 06
	\vdots	\vdots
14	0.4501836	0.7101e - 09

Die Abschätzungen (iii), (iv) sagen etwas über die Konvergenzgeschwindigkeit aus. Dies können wir ausnutzen, um etwa a priori (d.h. vor Beginn der Rechnung) zu bestimmen, wie viele Schritte wir brauchen, um den Fehler kleiner als ein vorgegebenes ε zu bekommen. In unserem Beispiel sei wieder $x^{(0)} = 0$ und $\varepsilon = 10^{-2}$. Wir benutzen die a priori Abschätzung:

$$|x^{(k)} - \hat{x}| \leq \frac{(1/2)^k}{1 - 1/2} |x^{(1)} - x^{(0)}| = \left(\frac{1}{2}\right)^{k-1} \left|\frac{1}{2} - 0\right| = \left(\frac{1}{2}\right)^k,$$

und dies ist kleiner als 10^{-2} für $k \geq 7$. Nach 7 Schritten sind wir also in jedem Fall am Ziel. Aus der Tabelle erkennen wir aber auch, dass diese a priori Abschätzung zu pessimistisch ist. Schon für $k = 3$ ist der Fehler kleiner als 10^{-2} . Die a posteriori Abschätzung kann als **Abbruchkriterium** im Computerprogramm benutzt werden, denn die Näherungen $x^{(k+1)}$ und $x^{(k)}$ werden ja berechnet. Sobald also die rechte Seite von der Abschätzung (iv) kleiner als ε ist, wird das Programm beendet. Wir sind dann sicher, dass der Fehler ebenfalls kleiner als ε ist. Am obigen Beispiel sehen wir, dass nach diesem Kriterium bei $k = 5$ abgebrochen wird.

(b) Auf \mathbb{R} seien die beiden Funktionen $f(x) = e^{2x} - 1$ und $g(x) = \frac{1}{4}(x - 1)^2$ gegeben. Gesucht ist ein Schnittpunkt der zugehörigen Graphen, d.h. $(\hat{x}, \hat{y}) \in \mathbb{R}^2$ mit $f(\hat{x}) = \hat{y}$ und $g(\hat{x}) = \hat{y}$. Aus dem Zwischenwertsatz ist mit $f(0) - g(0) = -\frac{1}{4} < 0$ und $f(1) - g(1) = e^2 - 1 > 0$ ersichtlich, dass im Intervall $[0, 1]$ ein solcher Punkt existiert. Wir können das Problem zu einem Fixpunktproblem umformulieren. Der gesuchte Schnittpunkt ist ein Fixpunkt der Abbildung $T : \mathbb{R} \times \mathbb{R}_{>-1} \rightarrow \mathbb{R}^2$ mit

$$T(x, y) = \left(\frac{1}{2} \ln(y + 1), \frac{1}{4}(x - 1)^2 \right).$$

Offensichtlich gilt $T(x, y) \in D = [0, 1] \times [0, 1]$ für $(x, y) \in [0, 1] \times [0, 1]$, d.h. $T : D \rightarrow D$. Nun zeigen wir noch, dass T auf der Menge D kontrahierend ist. Aus dem Mittelwertsatz ergibt sich

$$|e^a - e^b| = e^z |a - b| \geq |a - b| \quad \text{für } a, b \geq 0.$$

Mit $a = \ln(y + 1)$ und $b = \ln(y' + 1)$ in dieser Ungleichung und

$$\frac{1}{4} |(x-1)^2 - (x'-1)^2| = \frac{|x+x'-2|}{4} |x-x'| \leq \frac{1}{2} |x-x'|$$

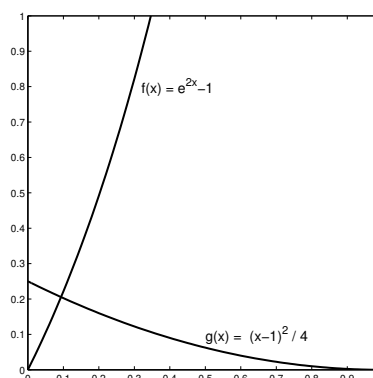
folgt für $0 \leq x, x' \leq 1$, dass

$$\|T(x, y) - T(x', y')\|_\infty \leq \frac{1}{2} \|(x, y) - (x', y')\|_\infty$$

gilt. Also ist der Kontraktionssatz mit Konstante $\kappa = \frac{1}{2} < 1$ anwendbar.

Die sukzessive Approximation ausgehend von $(x_0, y_0) = (0, 0)$ liefert

Iter.	x_n	y_n
1	0.00000000	0.25000000
2	0.11157177	0.25000000
3	0.11157177	0.19732617
4	0.09004544	0.19732617
5	0.09004544	0.20700432
	⋮	⋮
10	0.09345560	0.20552023
	⋮	⋮
20	0.09343306	0.20546588



2. Newtonverfahren

Wir betrachten weiterhin Gleichungen der Form $f(x) = y$ mit $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Sei im folgenden f stetig differenzierbar und die Jacobimatrix $f'(x)$ regulär, also invertierbar. Dann lässt sich die Gleichung in die Fixpunktgleichung

$$x = x - [f'(x)]^{-1}(f(x) - y)$$

umschreiben. Wenden wir hierauf das Kontraktionsverfahren an, so erhalten wir das **Newtonverfahren**: Mit gegebenem $x^{(0)} \in D$ bilde man die Iterationsfolge

$$x^{(k+1)} = x^{(k)} - [f'(x^{(k)})]^{-1}(f(x^{(k)}) - y), \quad k = 0, 1, 2, \dots \quad (1.8)$$

Eine Motivation zum Newtonverfahren ist die folgende Überlegung. Ist $x^{(k)}$ eine Näherung an die Lösung, so ersetzen wir f durch die Linearisierung in $x^{(k)}$, d.h.

$$f(x) \approx f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)}).$$

Da wir nach einer Lösung suchen, ist die linke Seite y , und wir ersetzen die Approximation durch Gleichheit. Es ergibt sich das **lineare** Gleichungssystem

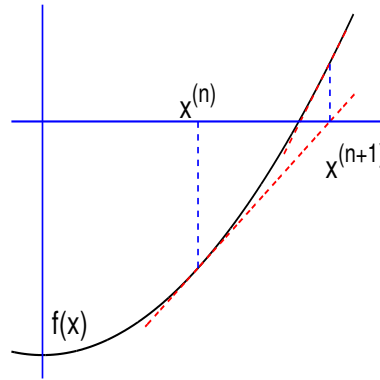
$$f'(x^{(k)})z = y - f(x^{(k)}) \quad \text{für } z = x - x^{(k)}.$$

Damit definieren wir $x^{(k+1)} := x^{(k)} + z$ und erhalten die Newtonformel (1.8).

Anschauung: Für $n = 1$ und $y = 0$ beschreibt die Abbildung

$$x \mapsto f(x^{(k)}) + f'(x^{(k)})(x - x^{(k)})$$

die Tangente an die Funktion f im Punkt $x^{(k)}$. Wir berechnen also als nächsten Näherungspunkt die Nullstelle der Tangente!



Das Newtonverfahren

Satz 1.43 (lokale quadratische Konvergenz)

Sei $f : \mathbb{R}^n \supset D \rightarrow \mathbb{R}^n$ zweimal stetig differenzierbar in D , $f(\hat{x}) = y$ für ein $\hat{x} \in D^\circ$ (also im Innern von D). Sei $f'(\hat{x}) \in \mathbb{R}^{n \times n}$ invertierbar. Dann gibt es eine Kugel $K(\hat{x}, \varepsilon)$ um \hat{x} , so dass das Newtonverfahren für jedes $x^{(0)} \in K(\hat{x}, \varepsilon)$ konvergiert. Es liegt quadratische Konvergenz vor, d.h. es gibt ein $c > 0$ mit:

$$\|x^{(k+1)} - \hat{x}\| \leq c \|x^{(k)} - \hat{x}\|^2 \quad \text{für alle } k = 1, 2, \dots$$

Hier kann irgendeine Norm $\|\cdot\|$ im \mathbb{R}^n genommen werden.

Beweis für den eindimensionalen Fall $n = 1$ mit $y = 0$: Wie oben erwähnt, entsteht das Newtonverfahren aus dem Kontraktionsverfahren für die Funktion

$$g(x) = x - \frac{f(x)}{f'(x)}, \quad x \in D.$$

Wir wollen den Satz 1.41 auf g und ein (genügend kleines) Intervall $I = [\hat{x} - \varepsilon, \hat{x} + \varepsilon]$ anwenden. Dafür müssen wir die beiden Voraussetzungen (i) und (ii) von Satz 1.41 nachweisen. Für die Kontraktionseigenschaft berechnen wir die Ableitung

$$g'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}, \quad x \in I.$$

Wir sehen, dass g' stetig ist und $g'(\hat{x}) = 0$. Also gibt es $\varepsilon > 0$ mit $|g'(x)| \leq \frac{1}{2}$ für alle $x \in I = [\hat{x} - \varepsilon, \hat{x} + \varepsilon]$. Wegen der Bemerkung nach Satz 1.41 ist auf diesem Intervall I die Funktion g also eine Kontraktion mit Konstante $\kappa = 1/2$. Jetzt müssen wir noch sicherstellen, dass g das Intervall I in sich abbildet. Sei also $x \in I$, d.h. $|x - \hat{x}| \leq \varepsilon$. Dann ist wegen $g(\hat{x}) = \hat{x}$ und dem Mittelwertsatz

$$|g(x) - \hat{x}| = |g(x) - g(\hat{x})| = |g'(z)| |x - \hat{x}| \leq \frac{1}{2} \varepsilon$$

für ein z zwischen x und \hat{x} . Damit ist $g(x) \in I$, und das Newtonverfahren konvergiert.

Um die quadratische Konvergenz zu zeigen, benutzen wir die Taylorentwicklung von f um $x^{(k)}$ in der Form

$$0 = f(\hat{x}) = f(x^{(k)}) + (\hat{x} - x^{(k)}) f'(x^{(k)}) + \frac{1}{2} f''(z) (\hat{x} - x^{(k)})^2$$

für ein z zwischen $x^{(k)}$ und \hat{x} . Also ist

$$\begin{aligned} |\hat{x} - x^{(k+1)}| &= \left| \hat{x} - x^{(k)} + \frac{f(x^{(k)})}{f'(x^{(k)})} \right| = \left| \frac{1}{f'(x^{(k)})} \right| |(\hat{x} - x^{(k)}) f'(x^{(k)}) + f(x^{(k)})| \\ &= \left| \frac{1}{f'(x^{(k)})} \right| \left| \frac{f''(z)}{2} \right| |x^{(k)} - \hat{x}|^2 \end{aligned}$$

für ein z zwischen $x^{(k)}$ und \hat{x} . Nach Voraussetzung ist $f'(\hat{x}) \neq 0$. Da f' stetig ist und $x^{(k)} \in I$, $k \in \mathbb{N}$, gibt es eine Konstante \tilde{c} mit $\tilde{c} \leq |1/f'(x^{(k)})|$ für alle $k \in \mathbb{N}$, wenn nur ε klein genug gewählt ist. Es folgt

$$|x^{(k+1)} - \hat{x}| \leq C |x^{(k)} - \hat{x}|^2$$

mit $C = (2\tilde{c})^{-1} \max_{z \in I} |f''(z)|$. □

Die quadratische Konvergenz ist eine sehr schnelle Konvergenz, wie wir an folgendem Beispiel sehen.

Beispiele 1.44

(a) Wir betrachten wieder das Beispiel 1.42 (a) von oben, d.h. die Gleichung $x = \frac{1}{2} \cos x$. Hier ist $f(x) = x - \frac{1}{2} \cos x$ und $f'(x) = 1 + \frac{1}{2} \sin x$. Für den Startwert $x^{(0)} = 0$ erhalten wir das Folgende (rechts zum Vergleich noch einmal die Folge $(\tilde{x}^{(k)})$ für das Kontraktionsverfahren):

k	$x^{(k)}$	$ x^{(k)} - \hat{x} $	$\tilde{x}^{(k)}$
0	0	$7.3415164e - 02$	0
1	0.5000000	$4.9816389e - 02$	0.5000000
2	0.4506267	$4.4308178e - 04$	0.4387912
3	0.4501836	$3.6282900e - 08$	0.4526329
4	0.4501836	$1.6653345e - 16$	0.4496493
5	0.4501836	0	0.4502997
6	0.4501836	0	0.4501583
7	0.4501836	0	0.4501891
8	0.4501836	0	0.4501824

Wir erkennen deutlich die viel schnellere Konvergenz des Newtonverfahrens gegenüber dem Kontraktionsverfahren!

(b) (Heronsches Verfahren zur Wurzelberechnung). Um \sqrt{a} zu berechnen, lösen wir die Gleichung $f(x) = x^2 - a = 0$. Das Newtonverfahren liefert mit $f'(x) = 2x$:

$$x^{(k+1)} = x^{(k)} - \frac{[x^{(k)}]^2 - a}{2x^{(k)}} = \frac{1}{2} \left(x^{(k)} + \frac{a}{x^{(k)}} \right), \quad k = 0, 1, \dots,$$

und dies ist das schon aus HM I bekannte Heronverfahren.

Das Newtonverfahren hat nicht nur Vorteile, sondern auch Nachteile:

- (a) Es gilt nur **lokale Konvergenz**, d.h. wir müssen schon dicht genug bei der Nullstelle starten. Andernfalls braucht die Folge nicht zu konvergieren.
- (b) Die Ableitungsmatrix $f'(x^{(k)})$ wird in jedem Schritt benötigt.

Wir wollen kurz andeuten, wie – jedenfalls im **eindimensionalen Fall** – diese Nachteile vermieden werden können. Sei also $f : \mathbb{R} \supset [a, b] \rightarrow \mathbb{R}$ gegeben, und f habe eine einfache Nullstelle in $[a, b]$.

Abhilfe zu (a): Wir verschaffen uns ein Intervall $[x^{(0)}, y^{(0)}] \subset [a, b]$ mit $f(x^{(0)}) \cdot f(y^{(0)}) < 0$. Dann liegt in $[x^{(0)}, y^{(0)}]$ sicher eine Nullstelle von f (Zwischenwertsatz 3.24 aus HM 1). Sei z.B. $f(x^{(0)}) < 0$, $f(y^{(0)}) > 0$. Definiere iterativ für $k = 0, 1, \dots$:

$$\begin{aligned} z^{(k)} &= \frac{1}{2}(x^{(k)} + y^{(k)}) \\ x^{(k+1)} &= x^{(k)}, \quad y^{(k+1)} = z^{(k)}, & \text{falls } f(z^{(k)}) > 0, \\ x^{(k+1)} &= z^{(k)}, \quad y^{(k+1)} = y^{(k)}, & \text{falls } f(z^{(k)}) < 0. \end{aligned}$$

Dieses Verfahren heißt **Intervallhalbierungsverfahren** oder **Bisektionsverfahren**. Es konvergiert zwar nur **linear** wegen $|y^{(k)} - x^{(k)}| = \left(\frac{1}{2}\right)^k |y^{(0)} - x^{(0)}|$ und $x^{(k)} \leq \hat{x} \leq y^{(k)}$, ist aber „sicher“ und liefert Einschließungen der Nullstelle. In der Praxis wird das Bisektionsverfahren mit dem schnellen Newtonverfahren gekoppelt.

Abhilfe zu (b): Es lässt sich zum Beispiel das **Sekantenverfahren** (oder eine Modifizierung davon, die Regula falsi) anwenden, bei dem statt der Nullstelle der Tangente die der Sekante durch die beiden vorhergehenden Punkte $(x^{(k-1)}, f(x^{(k-1)}))$, $(x^{(k)}, f(x^{(k)}))$ berechnet wird. Dies führt auf die Rekursion $(x^{(0)}, x^{(1)})$ gegeben):

$$x^{(k+1)} = x^{(k)} - f(x^{(k)}) \frac{x^{(k)} - x^{(k-1)}}{f(x^{(k)}) - f(x^{(k-1)})}, \quad k = 1, 2, \dots$$

Es gilt der gleiche lokale Konvergenzsatz wie beim Newtonverfahren, die Konvergenzgeschwindigkeit ist allerdings etwas langsamer: Es gibt $c > 0$ mit

$$|x^{(k+1)} - \hat{x}| \leq c |x^{(k)} - \hat{x}|^\alpha, \quad k = 0, 1, 2, \dots,$$

mit der goldenen Schnittzahl $\alpha = \frac{1}{2}(1 + \sqrt{5}) \approx 1.618$.

2 Das Gebietsintegral

2.1 Begriffsbildung und allgemeine Sätze

Im ersten Semester haben wir das Integral $\int_a^b f(x) dx$ für Funktionen $f : [a, b] \rightarrow \mathbb{R}$ einer Veränderlichen eingeführt. Sei jetzt $D \subset \mathbb{R}^2$ (oder später auch allgemeiner $D \subset \mathbb{R}^n$) eine Menge⁴, $f : D \rightarrow \mathbb{R}$ eine Funktion zweier Veränderlicher. Wir wollen das Gebietsintegral $\iint_D f(x, y) d(x, y)$ erklären. Veranschaulicht man sich die Funktion durch ihren Graphen im \mathbb{R}^3 , so ist *eine* anschauliche Vorstellung des Integrals gegeben durch den Inhalt zwischen dem Gebiet D in der (x, y) -Ebene und dem Graphen (positiv genommen für positive und negativ genommen für negative Funktionsteile).

Sei zunächst $D = R = I \times J \subset \mathbb{R}^2$ ein beschränktes **Rechteck**, das aus Intervallen I und J mit Randpunkten a, b bzw. c, d gebildet wird. I und J können jeweils wieder offen, abgeschlossen oder halboffen sein. Der **Flächeninhalt** von R ist $|R| = (b - a)(d - c)$. Wir gehen ganz analog zum eindimensionalen Fall vor und definieren zunächst den Begriff der Treppenfunktion.

Definition 2.1 (a) Sei $R = I \times J \subset \mathbb{R}^2$ ein beschränktes Rechteck, wobei die Intervalle I und J die Grenzen a, b bzw. c, d haben. Eine Funktion $\varphi : R \rightarrow \mathbb{R}$ heißt **Treppenfunktion**, falls es Zerlegungen

$$I = \bigcup_{i=1}^p I_i \quad \text{und} \quad J = \bigcup_{j=1}^q J_j$$

der Intervalle $[a, b]$ und $[c, d]$ in paarweise disjunkte Teilintervalle I_i bzw. J_j gibt sowie Zahlen $c_{ij} \in \mathbb{R}$ mit

$$\varphi(x, y) = c_{ij} \quad \text{für} \quad (x, y) \in I_i \times J_j, \quad i = 1, \dots, p, \quad j = 1, \dots, q.$$

(b) Ist $R = I \times J \subset \mathbb{R}^2$ unbeschränkt (d.h. wenigstens eines der Intervalle I, J ist unbeschränkt), so heißt $\varphi : R \rightarrow \mathbb{R}$ Treppenfunktion, wenn es ein beschränktes Rechteck $\tilde{R} \subset R$ gibt, so dass $\varphi|_{\tilde{R}}$ eine Treppenfunktion ist und $\varphi = 0$ auf $R \setminus \tilde{R}$ gilt.

(c) Ist $\varphi : R \rightarrow \mathbb{R}$ eine Treppenfunktion mit $\varphi(x, y) = c_{ij}$ für $(x, y) \in I_i \times J_j$, so setzen wir

$$\iint_R \varphi(x, y) d(x, y) := \sum_{i=1}^p \sum_{j=1}^q c_{ij} |I_i| |J_j|.$$

Die Definition 6.2 von HM I einer Nullmenge überträgt sich ebenfalls wörtlich:

⁴Ganz beliebig kann D nicht sein. Das werden wir unten genauer erläutern.

Definition 2.2 Eine Menge $M \subset \mathbb{R}^2$ heißt **Nullmenge** oder **Menge vom Maß Null**, wenn es zu jedem $\varepsilon > 0$ endlich viele oder abzählbar unendlich viele beschränkte Rechtecke $R_k \subset \mathbb{R}^2$, $k = 1, 2, \dots$, gibt mit

$$M \subset \bigcup_{k=1}^{\infty} R_k \quad \text{und} \quad \sum_{k=1}^{\infty} |R_k| \leq \varepsilon.$$

Wir sagen wieder: Eine Aussage $\mathcal{A}(x)$ gilt für **fast alle** x (oder **fast überall**, abgekürzt f.ü.), wenn die Menge $\{x : \mathcal{A}(x) \text{ gilt nicht}\}$ eine Nullmenge ist.

Genau wie im Eindimensionalen ist jede messbare Teilmenge einer Nullmenge wieder eine Nullmenge, und jede abzählbare Vereinigung $M = \bigcup_{n=1}^{\infty} M_n$ von Nullmengen M_n ist eine Nullmenge.

Beispiele 2.3

(a) Der Rand eines Rechtecks $[a, b] \times [c, d]$ ist eine Nullmenge.

Beweis: Der Rand besteht aus 4 Seiten. Um zu zeigen, dass etwa $\{a\} \times [c, d]$ eine Nullmenge ist, sei $\varepsilon > 0$ gegeben. Setze $\delta = \frac{\varepsilon}{2(d-c)}$. Dann enthält das Rechteck $R = [a - \delta, a + \delta] \times [c, d]$ diese Seite und hat den Flächeninhalt $2\delta(d - c) = \varepsilon$.

(b) Sei $\psi : [a, b] \rightarrow \mathbb{R}$ eine stetige Funktion. Dann ist ihr Graph

$$M := \{x \in \mathbb{R}^2 : x_2 = \psi(x_1), a \leq x_1 \leq b\}$$

eine Nullmenge im \mathbb{R}^2 .

Beweis: Sei $\varepsilon > 0$. Da ψ gleichmäßig stetig ist,⁵ existiert $\delta > 0$ mit $|\psi(t) - \psi(s)| \leq \frac{\varepsilon}{2(b-a)}$ für $|t - s| \leq \delta$. Bestimme $n \in \mathbb{N}$ so groß, dass $(b-a)/n \leq \delta$ und setze $t_j = a + j(b-a)/n$, $j = 0, \dots, n$. Dann leisten die Rechtecke

$$R_k := [t_{k-1}, t_k] \times [\psi(t_k) - \varepsilon/(2(b-a)), \psi(t_k) + \varepsilon/(2(b-a))], \quad k = 1, \dots, n,$$

das Gewünschte, da $M \subset \bigcup_{k=1}^n R_k$ und die Gesamtfläche der Rechtecke gerade $n \frac{b-a}{n} \frac{\varepsilon}{b-a} = \varepsilon$ ist. \square

Ganz entscheidend für die eindimensionale Theorie war die eindimensionale Version (Lemma 6.5 von HM I) des folgenden Lemmas, dessen Beweis nicht einfach ist – aber genauso verläuft wie im eindimensionalen Fall:

Lemma 2.4 Sei (φ_n) eine Folge von nichtnegativen Treppenfunktionen auf dem beschränkten Rechteck R , die punktweise f.ü. monoton fallend gegen Null konvergiert. Dann konvergiert die Integralfolge $\iint_R \varphi_n(x, y) d(x, y)$ gegen Null.

Jetzt können wir wieder definieren (vgl. Definitionen 6.6 bis 6.8 von HM I):

⁵Zur Definition der glm. Stetigkeit siehe HM I, Seite 38. Man kann mit dem Satz von Bolzano-Weierstraß zeigen, dass jede auf einem kompakten Intervall $[a, b]$ stetige Funktion sogar glm. stetig ist.

Definition 2.5

(a) Sei $R \subset \mathbb{R}^2$ ein (beschränktes oder unbeschränktes) Rechteck. Die Menge $L^+(R)$ sei definiert als die Menge aller Funktionen $f : R \rightarrow \mathbb{R}$, mit folgenden Eigenschaften:

- (i) Es gibt eine **monoton wachsende Folge** (φ_n) von Treppenfunktionen (d.h. für $n \in \mathbb{N}$ ist $\varphi_n(x) \leq \varphi_{n+1}(x)$ für fast alle $x \in R$), die gegen f **fast überall punktweise konvergiert** (d.h. $\lim_{n \rightarrow \infty} \varphi_n(x) = f(x)$ für fast alle $x \in R$).
- (ii) Die Folge der Integrale $\iint_R \varphi_n(x, y) d(x, y)$ ist konvergent.

(b) Für $f \in L^+(R)$ definieren wir das (Lebesguesche) **Integral** von f über R durch

$$\iint_R f(x, y) d(x, y) := \lim_{n \rightarrow \infty} \iint_R \varphi_n(x, y) d(x, y).$$

(c) Wir setzen dann

$$L(R) := L^+(R) - L^+(R) = \{f = f_1 - f_2 : f_1, f_2 \in L^+(R)\}.$$

Die Menge $L(R)$ heißt die Menge der **Lebesgue integrierbaren Funktionen** über R . Für $f = f_1 - f_2 \in L(R)$ setzen wir

$$\iint_R f(x, y) d(x, y) := \iint_R f_1(x, y) d(x, y) - \iint_R f_2(x, y) d(x, y).$$

Was ist, wenn das Integrationsgebiet kein Rechteck ist? Dies ist grundlegend anders als im eindimensionalen Fall, da dort Funktionen eigentlich immer auf Intervallen (oder jedenfalls endlichen Vereinigungen von Intervallen) definiert sind. Gebiete im \mathbb{R}^2 können dagegen sehr „wild“ aussehen. Wir umgehen diese Schwierigkeit elegant:

Definition 2.6 Sei $D \subset \mathbb{R}^2$ eine beliebige Menge und $f : D \rightarrow \mathbb{R}$. Die Funktion f heißt auf D **integrierbar**, falls die durch 0 fortgesetzte Funktion

$$f^*(x, y) = \begin{cases} f(x, y), & (x, y) \in D, \\ 0, & (x, y) \in \mathbb{R}^2 \setminus D, \end{cases}$$

auf \mathbb{R}^2 integrierbar ist. In diesem Fall setzen wir

$$\iint_D f(x, y) d(x, y) := \iint_{\mathbb{R}^2} f^*(x, y) d(x, y).$$

Bemerkung: Eine Funktion $f : D \rightarrow \mathbb{R}$ heißt (Lebesgue-) **messbar**, wenn es eine Folge (φ_n) von Treppenfunktionen gibt, die fast überall gegen f konvergiert. Eine beschränkte Menge D ist eine **messbare Menge**, wenn die **charakteristische Funktion**

$$\chi_D(x, y) := \begin{cases} 1, & (x, y) \in D, \\ 0, & (x, y) \in \mathbb{R}^2 \setminus D, \end{cases}$$

integrierbar ist, d.h. wenn $\iint_D 1 d(x, y)$ als endlicher Wert existiert. Der Wert des Integrals heißt **Flächeninhalt** oder **Maß** von D . Wir schreiben dafür $|D|$.

Auch der **Schwerpunkt** S einer messbaren Menge lässt sich so festlegen. Es ist

$$\mathbb{R}^2 \ni S = \frac{1}{|D|} \begin{pmatrix} \iint_D x d(x, y) \\ \iint_D y d(x, y) \end{pmatrix} = \frac{1}{|D|} \iint_D \begin{pmatrix} x \\ y \end{pmatrix} d(x, y).$$

Beachte die häufig verwendete letzte Notation. Bei Gebietsintegralen wird ein vektorwertiger Integrand komponentenweise aufgefasst und das Ergebnis ist wieder ein Vektor.

Wir ersparen uns eine ausführliche Theorie zu messbaren Mengen, also Mengen, über die wir Integrale definieren können. Nur die folgenden drei Aussagen seien angemerkt:

1. $|D| = 0 \iff D$ ist Nullmenge.
2. Endliche oder abzählbare Vereinigungen oder Durchschnitte von messbaren Mengen sind wieder messbar.
3. Beschränkte offene sowie beschränkte abgeschlossene Mengen sind messbar.

Beispiel 2.7

Seien $\psi, \varphi : [a, b] \rightarrow \mathbb{R}$ stetige Funktionen mit $\psi(x) < \varphi(x)$ für alle $x \in (a, b)$. Dann ist die Menge

$$D := \{(x, y) \in \mathbb{R}^2 : \psi(x) \leq y \leq \varphi(x), a \leq x \leq b\}$$

kompakt, also messbar.

Ab jetzt schreiben wir oft wieder $x = (x_1, x_2)^\top \in \mathbb{R}^2$ statt $(x, y)^\top$. Anstelle des Integrals $\iint_D f(x_1, x_2) d(x_1, x_2)$ notieren wir kürzer $\iint_D f(x) dx$. Auch die zwei Integralzeichen werden wir, wenn wir uns an diese Integrale gewöhnt haben, oft nur durch eines ersetzen. Wir schreiben also in Zukunft häufig einfach

$$\int_D f(x) dx$$

für diese Integrale wie im eindimensionalen Fall. Durch die Definition der Menge D ist die zu betrachtende Dimension stets eindeutig vorgegeben und es kann nicht zu Verwechslungen kommen. Vorteil der dimensionsunabhängigen Notation ist, dass sich weiterführende Resultate etwa in Kapitel 3 ohne Fallunterscheidung beschreiben lassen.

Der nächste Satz fasst die üblichen Rechenregeln für mehrfache Integrale zusammen (vgl. Satz 6.9 von HM I):

Satz 2.8 (Rechenregeln)

Sei $D \subset \mathbb{R}^2$ messbar, $f, g : \mathbb{R}^2 \supset D \rightarrow \mathbb{R}$ integrierbar, $\lambda \in \mathbb{R}$.

(a) Dann sind auch die Funktionen

$$f \pm g, \quad \lambda f, \quad \max\{f, g\}, \quad \min\{f, g\}, \quad |f|$$

integrierbar, und es gilt:

$$\begin{aligned} \iint_D [f(x) + g(x)] dx &= \iint_D f(x) dx + \iint_D g(x) dx, \\ \iint_D [\lambda f(x)] dx &= \lambda \iint_D f(x) dx. \end{aligned}$$

Dies bedeutet, dass $L(D)$ ein Vektorraum ist (Definition 2.1 von HM II) und die Abbildung $f \mapsto \iint_D f(x) dx$ eine **lineare Abbildung** von $L(D)$ nach \mathbb{R} ist (Definition 1.25 von HM II).

(b) Es gilt die **Dreiecksungleichung**

$$\left| \iint_D f(x) dx \right| \leq \iint_D |f(x)| dx.$$

(c) Sind f^2 und g^2 integrierbar, so auch fg , und es gilt die **Cauchy-Schwarzsche Ungleichung**

$$\left[\iint_D f(x)g(x) dx \right]^2 \leq \left[\iint_D f(x)^2 dx \right] \left[\iint_D g(x)^2 dx \right].$$

(d) Es gilt die folgende **Monotonieaussage**: Ist $f(x) \leq g(x)$ für fast alle $x \in D$, so ist $\iint_D f(x) dx \leq \iint_D g(x) dx$.

(e) Ist $f(x) = 0$ für fast alle $x \in D$, so ist $\iint_D f(x) dx = 0$.

(f) Ist $f(x) \geq 0$ fast überall und $\iint_D f(x) dx = 0$, so ist $f(x) = 0$ für fast alle $x \in D$.

Bemerkung: Die Aussage (e) des Satzes bedeutet, dass wir eine Funktion auf Nullmengen abändern dürfen, ohne dass sich das Integral ändert. Deswegen ist es für die Integration gleichgültig, wie wir die Treppenfunktionen auf den Rändern der Teilrechtecke definieren.

Der folgende Satz sichert die Integrierbarkeit für eine große Klasse von Funktionen:

Satz 2.9

(a) Ist $D \subset \mathbb{R}^2$ kompakt und $f : D \rightarrow \mathbb{R}$ stetig, so ist f auf D integrierbar.

(b) Ist D endliche paarweise disjunkte Vereinigung $D = D_1 \cup \dots \cup D_n$ von kompakten Mengen D_j , $j = 1, \dots, n$, und ist die Funktion $f : D \rightarrow \mathbb{R}$ auf jeder der Mengen D_j stetig, so ist f auch auf D integrierbar, und es gilt

$$\iint_D f(x) dx = \sum_{j=1}^n \iint_{D_j} f(x) dx.$$

Der Beweis verläuft wieder genauso wie im eindimensionalen Fall (Satz 6.10 von HM I) durch Intervallhalbierung.

2.2 Das iterierte Integral

Jetzt kommen wir zur konkreten Berechnung von Gebietsintegralen. Zunächst behandeln wir wieder den Fall, dass D ein Rechteck ist.

Satz 2.10 (Fubini)

Sei $R = I \times J \subset \mathbb{R}^2$ ein Rechteck mit Intervallen I, J und sei $f \in L(R)$. Dann existieren die Funktionen

$$g(y) := \int_I f(x, y) dx \quad f. \ddot{u}. \text{ auf } J, \quad h(x) := \int_J f(x, y) dy \quad f. \ddot{u}. \text{ auf } I,$$

und gehören zu $L(J)$ bzw. $L(I)$, und es ist

$$\iint_R f(x, y) d(x, y) = \int_J \left[\int_I f(x, y) dx \right] dy,$$

und

$$\iint_R f(x, y) d(x, y) = \int_I \left[\int_J f(x, y) dy \right] dx.$$

Die rechten Integrale heißen **iterierte Integrale**. Die Klammern werden weggelassen und wir schreiben einfach

$$\iint_R f(x, y) d(x, y) = \int_I \int_J f(x, y) dy dx = \int_J \int_I f(x, y) dx dy.$$

Beweisidee: Wir überlegen uns nur, dass die Aussage des Satzes für Treppenfunktionen $\varphi : R \rightarrow \mathbb{R}$ gilt. Sei $\varphi(x) = c_{ij}$ für $x \in R_{ij}$, $i = 1, \dots, p$, $j = 1, \dots, q$. Dann ist nach Definition 2.1

$$\iint_R \varphi(x, y) d(x, y) = \sum_{i=1}^p \sum_{j=1}^q c_{ij} |I_i| |J_j|. \quad (2.1)$$

Da für festgehaltenes $y \in [y_{j-1}, y_j]$ die Funktion $\varphi(\cdot, y)$ eine Treppenfunktion ist, definieren wir

$$g(y) = \int_I \varphi(x, y) dx = \sum_{i=1}^p c_{ij}(x_i - x_{i-1}), \quad y \in [y_{j-1}, y_j].$$

Damit ist g ebenfalls eine Treppenfunktion mit

$$\int_I g(y) dy = \sum_{j=1}^q \left[\sum_{i=1}^p c_{ij}(x_i - x_{i-1}) \right] (y_j - y_{j-1}).$$

Dies stimmt mit (2.1) überein.

Durch ein (recht aufwendiges) Approximationsargument wird der Satz dann für allgemeine Funktionen $f \in L(R)$ bewiesen. \square

Damit haben wir die Berechnung von Gebietsintegralen auf die Berechnung von eindimensionalen Integralen zurückgeführt.

Wie sieht das Integral über allgemeinere Gebiete als Rechtecke aus? Sei jetzt D gegeben durch

$$D := \{(x, y) \in \mathbb{R}^2 : \varphi(x) \leq y \leq \psi(x), a \leq x \leq b\}$$

mit stetigen Funktionen $\varphi, \psi : [a, b] \rightarrow \mathbb{R}$ mit $\varphi(x) < \psi(x)$ für alle $x \in (a, b)$. Sei $f : D \rightarrow \mathbb{R}$ stetig und $f^* : R = [a, b] \times [c, d] \rightarrow \mathbb{R}$ die Fortsetzung von f durch 0 auf das Rechteck $R = (a, b) \times (c, d)$ (hier ist $d \geq \max \psi$ und $c \leq \min \varphi$). Dann ist mit dem Satz von Fubini:

$$\iint_D f(x, y) d(x, y) = \iint_R f^*(x, y) d(x, y) = \int_a^b \left[\int_{\varphi(x)}^{\psi(x)} f(x, y) dy \right] dx.$$

Analog wird der Fall

$$D = \{(x, y) \in \mathbb{R}^2 : \varphi(y) \leq x \leq \psi(y), c \leq y \leq d\}$$

behandelt. Hier ist

$$\iint_D f(x, y) d(x, y) = \int_c^d \left[\int_{\varphi(y)}^{\psi(y)} f(x, y) dx \right] dy.$$

Beispiel 2.11

(a) Es lässt sich etwa die Schnittmenge der drei Halbräume $H_1 = \{x \in \mathbb{R}^2 : x_1 \geq 0\}$, $H_2 = \{x \in \mathbb{R}^2 : x_2 \geq 0\}$ und $H_3 = \{x \in \mathbb{R}^2 : x_1 + x_2 \leq 1\}$, also ein Dreieck (Zeichnung!), durch

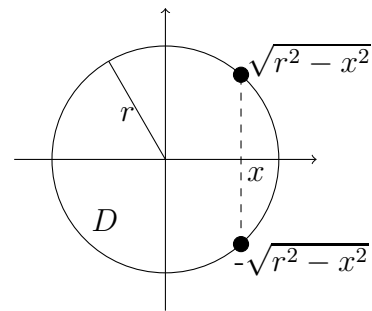
$$\{x \in \mathbb{R}^2 : 0 \leq x_2 \leq 1 - x_1, 0 \leq x_1 \leq 1\}$$

beschreiben. Der Flächeninhalt des Dreieck errechnet sich somit aus

$$\iint_M dx = \int_0^1 \int_0^{1-x_1} dx_2 dx_1 = \int_0^1 (1 - x_1) dx_1 = \frac{1}{2}.$$

(b) Gesucht ist das Volumen V der halben Vollkugel vom Radius $r > 0$, also das Volumen zwischen der (x, y) -Ebene und dem Graphen der Funktion $f(x, y) = \sqrt{r^2 - x^2 - y^2}$, $x^2 + y^2 \leq r^2$. Dazu betrachten wir

$$\begin{aligned} D &= \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\} \\ &= \{(x, y) \in \mathbb{R}^2 : -\sqrt{r^2 - x^2} \leq y \leq \sqrt{r^2 - x^2}, -r \leq x \leq r\} \end{aligned}$$



und erhalten

$$V = \iint_D f(x, y) d(x, y) = \int_{-r}^r \left[\int_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} \sqrt{r^2 - x^2 - y^2} dy \right] dx.$$

Zunächst rechnen wir das innere Integral aus. Aus Symmetriegründen gilt

$$\int_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} \sqrt{r^2 - x^2 - y^2} dy = 2 \int_0^{\sqrt{r^2 - x^2}} \sqrt{r^2 - x^2 - y^2} dy.$$

Wir substituieren $y = \sqrt{r^2 - x^2} \cos t$, also $dy = -\sqrt{r^2 - x^2} \sin t dt$ und erhalten

$$\begin{aligned} \int_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} \sqrt{r^2 - x^2 - y^2} dy &= -2 \int_{\pi/2}^0 \sqrt{r^2 - x^2} \sin^2 t \sqrt{r^2 - x^2} dt \\ &= 2(r^2 - x^2) \int_0^{\pi/2} \sin^2 t dt = \frac{\pi}{2}(r^2 - x^2). \end{aligned}$$

Schließlich folgt

$$V = \frac{\pi}{2} \int_{-r}^r (r^2 - x^2) dx = \frac{\pi}{2} \left(2r^3 - \frac{2}{3}r^3 \right) = \frac{2\pi}{3}r^3.$$

2.3 Die Transformationsformel

Häufig werden Gebiete oder Ränder von Gebieten geometrisch beschrieben und es ist zunächst eine mathematische Beschreibung der Menge erforderlich, etwa um ein Integral zu berechnen. Dies kann wie oben durch Graphen von Funktionen erreicht werden. Manche Gebiete lassen sich aber leichter in anderen Koordinaten, z.B. Polarkoordinaten, beschreiben. Um Integrale nach einem solchen Wechsel der Koordinaten zu berechnen, benötigen wir eine entsprechende Verallgemeinerung der Substitutionsregel - die Transformationsformel.

Wir schreiben jetzt $x = (x_1, x_2)$ statt (x, y) , weil wir auch andere Koordinatensysteme betrachten wollen, etwa $y = (y_1, y_2)$. Für ein $D \subseteq \mathbb{R}^n$ ($n = 2$ oder $n = 3$) ist eine Transformation gegeben durch eine Funktion $\Psi : D \in \mathbb{R}^n$ mit $\Psi(y) = x$ mit invertierbarer Funktionalmatrix $\Psi'(y) \in \mathbb{R}^{n \times n}$ für jedes $y \in D$, d.h. $\det(\Psi'(y)) \neq 0$ für alle $y \in D$.

Im Fall von ebenen **Polarkoordinaten** (d.h. im \mathbb{R}^2) gilt

$$x = \Psi(r, \varphi) = \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \end{pmatrix} \quad \text{mit} \quad \Psi'(r, \varphi) = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix}.$$

Halten wir eine der Koordinaten fest, so erhalten wir Koordinatenlinien, etwa vom Ursprung ausgehende Strahlen $r \mapsto \Psi(r, \varphi)$ bei festem Winkel φ oder Kreise um den Ursprung $\varphi \mapsto \Psi(r, \varphi)$ bei festem Radius r . Die Tangentialvektoren an diesen Linien ergeben sich allgemein aus

$$b^{(j)} = \frac{\partial \Psi}{\partial y_j}(y) = \Psi'(y)e^{(j)},$$

wobei wir mit $e^{(1)} = (1, 0)$ und $e^{(2)} = (0, 1)$ die üblichen Einheitsvektoren (hier $n = 2$) bezeichnen. Für die Polarkoordinaten erhalten wir

$$b^{(1)} = \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} \quad \text{und} \quad b^{(2)} = r \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix}.$$

Normieren wir die Vektoren $b^{(j)}$ auf die Länge 1, so ergeben sich begleitende Basisvektoren $e_{y_j} = \frac{b^{(j)}}{\|b^{(j)}\|}$ zu den im allgemeinen **krummlinigen Koordinaten** die durch y (im Beispiel den Polarkoordinaten) angegeben sind. Stehen die Tangentialvektoren darüberhinaus paarweise senkrecht aufeinander, wie es bei Polarkoordinaten der Fall ist, so spricht man von **orthogonalen Koordinaten**.

Nach diesen kurzen Vorbemerkungen zu Transformationen lässt sich nun der entscheidende Transformationsatz angeben.

Satz 2.12 (*Transformations- oder Substitutionsformel*)

Sei $D \subset \mathbb{R}^2$ kompakt, $\Omega \subset \mathbb{R}^2$ offen mit $D \subset \Omega$ und $\Psi : \mathbb{R}^2 \supset \Omega \rightarrow \mathbb{R}^2$ stetig differenzierbar mit Jacobimatrix $\Psi'(y) \in \mathbb{R}^{2 \times 2}$ für $y \in \Omega$. Sei weiter $\Psi : \mathbb{R}^2 \supset \Omega \rightarrow \Psi(\Omega) \subset \mathbb{R}^2$ umkehrbar und es gibt eine Nullmenge $N \subset \Omega$ mit $\det \Psi'(y) > 0$ oder $\det \Psi'(y) < 0$ für alle $y \in \Omega \setminus N$. Eine Funktion $f : \Psi(D) \rightarrow \mathbb{R}$ ist genau dann integrierbar, wenn $y \mapsto f(\Psi(y)) |\det \Psi'(y)|$ integrierbar ist. In diesem Fall gilt die **Transformationsformel**:

$$\iint_{\Psi(D)} f(x) dx = \iint_D f(\Psi(y)) |\det \Psi'(y)| dy.$$

Die Determinante $J(y) = \det \Psi'(y)$ heißt auch **Jacobi- oder Funktionaldeterminante** der Transformation Ψ .

Bemerkungen:

(a) Für die Jacobimatrix $\Psi'(y_1, y_2)$ schreibt man auch

$$\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}, \quad \text{also}$$

$$\iint_{D_x} f(x_1, x_2) dx = \iint_{D_y} f(y_1, y_2) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| dy,$$

wobei D_x und D_y das Integrationsgebiet ist, beschrieben in den x - bzw. y -Koordinaten.

(b) In einer Dimension ist dies gerade die Substitutionsregel: Sei z.B. $\Psi : [a, b] \rightarrow \mathbb{R}$ strikt monoton steigend. Dann ist $\Psi([a, b]) = [\Psi(a), \Psi(b)]$ und

$$\int_{\Psi([a, b])} f(x) dx = \int_{\Psi(a)}^{\Psi(b)} f(x) dx = \int_a^b f(\Psi(y)) \Psi'(y) dy = \int_a^b f(\Psi(y)) |\Psi'(y)| dy,$$

da $\Psi'(y) \geq 0$. Ist andererseits $\Psi : [a, b] \rightarrow \mathbb{R}$ strikt monoton fallend, so ist $\Psi([a, b]) = [\Psi(b), \Psi(a)]$ und daher

$$\begin{aligned} \int_{\Psi([a,b])} f(x) dx &= \int_{\Psi(b)}^{\Psi(a)} f(x) dx = - \int_{\Psi(a)}^{\Psi(b)} f(x) dx \\ &= - \int_a^b f(\Psi(y)) \Psi'(y) dy = \int_a^b f(\Psi(y)) |\Psi'(y)| dy \end{aligned}$$

da $\Psi'(y) \leq 0$. Aufgrund der Konvention zur Orientierung eindimensionaler Integrale, kann bei der eindimensionalen Regel auch auf Injektivität von Ψ verzichtet werden.

Wichtige Beispiele:

(A) **Polarkoordinaten:** Sei $\Psi(r, \varphi) = (r \cos \varphi, r \sin \varphi)^\top$, $r \geq 0$, $\varphi \in [0, 2\pi]$. Mit Ausnahme der Nullmenge $N = \{0\}$ ist Ψ eine bijektive differenzierbare Abbildung, eine Transformation, mit

$$|\det \Psi'(r, \varphi)| = \begin{vmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{vmatrix} = r \quad \text{für } r \neq 0.$$

Daher lautet die Transformationsformel in diesem Fall:

$$\iint_{D_k} f(x) dx = \iint_{D_p} f(r \cos \varphi, r \sin \varphi) r d(r, \varphi),$$

wenn D_p das Gebiet in der „Polarkoordinaten-Ebene“ und D_k das Gebiet in der „kartesischen Koordinaten-Ebene“ ist.

Hiermit lässt sich etwa elegant ausrechnen:

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}.$$

(B) **Elliptische Koordinaten:** Analog zu den Polarkoordinaten lassen sich auch Ellipsen beschreiben. Seien $a, b > 0$ die Halbachsen einer Ellipse mit Zentrum im Ursprung, so ist eine passende Transformation durch $\Psi(\lambda, t) = (a\lambda \cos t, b\lambda \sin t)^\top$, $\lambda \geq 0$, $t \in [0, 2\pi]$ gegeben. Dann ist

$$\det \Psi'(\lambda, t) = \begin{vmatrix} a \cos t & -a\lambda \sin t \\ b \sin t & b\lambda \cos t \end{vmatrix} = ab\lambda.$$

Daher lautet die Transformationsformel in diesem Fall (wieder sei D_k das Gebiet in kartesischen D_e in elliptischen Koordinaten):

$$\iint_{D_k} f(x) dx = ab \iint_{D_e} f(a\lambda \cos t, b\lambda \sin t) \lambda d(\lambda, t).$$

Beispiel 2.13

Sei D der Teil des Kreisrings mit Radien $1/2$ und 1 , der zwischen der x_1 -Achse und der Winkelhalbierenden des ersten Quadrant liegt. Berechne $\iint_D (x_1 + x_2) dx$. Hier ist also $f(x) = x_1 + x_2$ und $D_p = \{(r, \varphi) : 1/2 \leq r \leq 1, 0 \leq \varphi \leq \pi/4\}$. Also ist

$$\begin{aligned} \iint_D (x_1 + x_2) dx &= \iint_{D_p} (r \cos \varphi + r \sin \varphi) r d(r, \varphi) = \int_{1/2}^1 \int_0^{\pi/4} r^2 (\cos \varphi + \sin \varphi) d\varphi dr \\ &= \int_{1/2}^1 r^2 [\sin(\pi/4) - (\cos(\pi/4) + 1)] dr = \int_{1/2}^1 r^2 dr = 7/24. \end{aligned}$$

2.4 Übertragung auf dreidimensionale Gebietsintegrale

Sei jetzt $D \subset \mathbb{R}^3$ ein Quader, d.h. z.B. $D = [a, b] \times [c, d] \times [g, h] \subset \mathbb{R}^3$, und $f : \mathbb{R}^3 \supset D \rightarrow \mathbb{R}$ eine Funktion. Die Definition des Lebesgueschen-Integrals $\iiint_D f(x, y, z) d(x, y, z)$ verläuft ganz entsprechend zum zweidimensionalen Fall. Auch die Verallgemeinerung der Integrierbarkeit auf beliebigen Mengen können wir wie in Definition 2.6 übernehmen. Der Satz von Fubini und die Transformationsformel gelten ebenfalls ganz analog. Wichtige Beispiele für Koordinatentransformationen sind:

(A) Kugelkoordinaten:

$$x_1 = r \sin \theta \cos \varphi, \quad x_2 = r \sin \theta \sin \varphi, \quad x_3 = r \cos \theta$$

für $r \geq 0$, $\theta \in [0, \pi]$ und $\varphi \in [0, 2\pi]$. Hier ist also

$$\Psi(r, \theta, \varphi) = r \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}$$

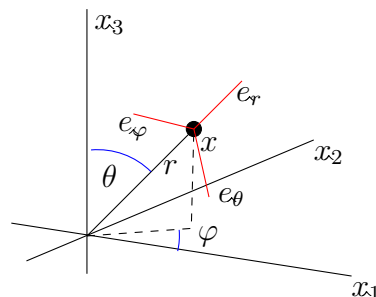
und daher

$$\Psi'(r, \theta, \varphi) = \begin{pmatrix} \sin \theta \cos \varphi & r \cos \theta \cos \varphi & -r \sin \theta \sin \varphi \\ \sin \theta \sin \varphi & r \cos \theta \sin \varphi & r \sin \theta \cos \varphi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}.$$

Die Tangentialvektoren

$$b_r = \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}, \quad b_\varphi = r \begin{pmatrix} -\sin \theta \sin \varphi \\ \sin \theta \cos \varphi \\ 0 \end{pmatrix}, \quad b_\theta = r \begin{pmatrix} \cos \theta \cos \varphi \\ \cos \theta \sin \varphi \\ -\sin \theta \end{pmatrix}$$

stehen senkrecht zueinander, d.h. auch die Kugelkoordinaten bilden ein orthogonales Koordinatensystem. Mit den normierten Tangentialvektoren e_r, e_φ, e_θ erhalten wir die zugehörigen Koordinateneinheitsvektoren, das sogenannte **begleitende Dreibein** der Kugelkoordinaten.



Wir errechnen für die Kugelkoordinaten etwa durch Entwicklung nach der letzten Zeile

$$\det \Psi'(r, \theta, \varphi) = r^2 \sin \theta.$$

Daher lautet die Transformationsformel in diesem Fall:

$$\iiint_{D_k} f(x) dx = \iiint_{D_p} f(r \sin \theta \cos \varphi, r \sin \theta \sin \varphi, r \cos \theta) r^2 \sin \theta d(r, \theta, \varphi).$$

(B) Zylinderkoordinaten:

$$x_1 = r \cos \varphi, \quad x_2 = r \sin \varphi, \quad x_3 = z$$

für $r \geq 0$, $\varphi \in [0, 2\pi]$ und $z \in \mathbb{R}$. Hier ist also

$$\Psi(r, \varphi, z) = \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \\ z \end{pmatrix}, \quad \Psi'(r, \varphi, z) = \begin{pmatrix} \cos \varphi & -r \sin \varphi & 0 \\ \sin \varphi & r \cos \varphi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

also $|\det \Psi'(r, \varphi, z)| = r$. Daher lautet die Transformationsformel in diesem Fall:

$$\iiint_{D_k} f(x) dx = \iiint_{D_z} f(r \cos \varphi, r \sin \varphi, z) r d(r, \varphi, z).$$

Anwendung: Volumina von Rotationskörpern

Der Körper $D \subset \mathbb{R}^3$ entstehe durch Rotation der Menge M in der (x, y) -Ebene um die x -Achse. Zu berechnen ist sein Volumen. (Zeichnung!)

Sei $M \subset \mathbb{R}^2$ gegeben durch

$$M = \{(x, y) \in \mathbb{R}^2 : f(x) \leq y \leq g(x), a \leq x \leq b\},$$

mit $f \geq 0$. Hier bieten sich Zylinderkoordinaten an:

$$x = x, \quad y = r \cos \varphi, \quad z = r \sin \varphi.$$

In diesen Koordinaten wird der Rotationskörper beschrieben durch

$$D_z = \{(x, r, \varphi) : f(x) \leq r \leq g(x), a \leq x \leq b, 0 \leq \varphi \leq 2\pi\}.$$

Daher ist

$$\begin{aligned} V &= \iiint_D d(x, y, z) = \iiint_{D_z} r d(x, r, \varphi) = \int_a^b \int_0^{2\pi} \int_{f(x)}^{g(x)} r dr d\varphi dx \\ &= 2\pi \int_a^b \left[\frac{1}{2} r^2 \right]_{f(x)}^{g(x)} dx = \pi \int_a^b [g(x)^2 - f(x)^2] dx, \end{aligned}$$

d.h. es gilt die Formel

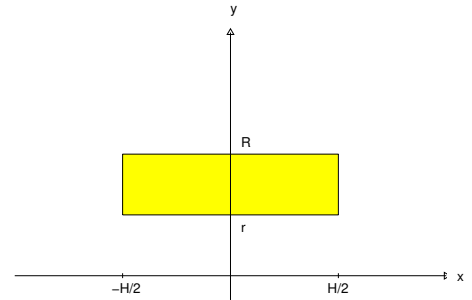
$$V = \pi \int_a^b [g(x)^2 - f(x)^2] dx.$$

Beispiele 2.14

(A) **ausgebohrter Zylinder:** Hier ist

$g(x) = R$ und $f(x) = r$, also

$$V = \pi \int_{-H/2}^{H/2} (R^2 - r^2) dx = \pi H(R^2 - r^2)$$



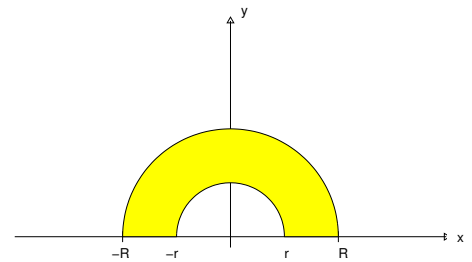
(B) **Hohlkugel:** Hier ist

$g(x) = \sqrt{R^2 - x^2}$ und

$f(x) = \sqrt{r^2 - x^2}$, falls $|x| \leq r$ und

$f(x) = 0$ für $r < |x| \leq R$. Also ist

$$\begin{aligned} V &= \pi \int_{-R}^R (R^2 - x^2) dx - \pi \int_{-r}^r (r^2 - x^2) dx \\ &= \frac{4}{3} \pi (R^3 - r^3) \end{aligned}$$

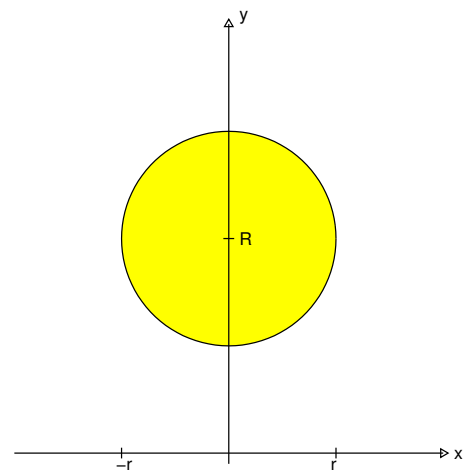


(C) **Torus:** Hier ist

$g(x) = R + \sqrt{r^2 - x^2}$ und

$f(x) = R - \sqrt{r^2 - x^2}$, also

$$\begin{aligned} V &= \pi \int_{-r}^r [(R + \sqrt{r^2 - x^2})^2 - (R - \sqrt{r^2 - x^2})^2] dx \\ &= \dots = 2\pi^2 R r^2. \end{aligned}$$



3 Vektoranalysis

3.1 Kurven und Kurvenintegrale

Den Begriff einer Kurve hatten wir schon früher angedeutet. Nun wollen wir uns genauer mit solchen Funktionen beschäftigen und einige geometrische Aspekt in den Vordergrund stellen.

Definition 3.1 Eine Menge $C \subseteq \mathbb{R}^n$ heißt **Kurve** (oder **Weg**) im \mathbb{R}^n , wenn es eine stetige Funktion $x : [a, b] \rightarrow \mathbb{R}^n$ gibt mit $C = \{x(t) : t \in [a, b]\} \subseteq \mathbb{R}^n$. Das Intervall $I := [a, b]$ heißt das **Parameterintervall**, und die Funktion x ist eine **Parametrisierung** der Kurve. Für $n = 2$ erhalten wir ebene Kurven, für $n = 3$ räumliche Kurven.

Die Variable wird im Allgemeinen mit t statt mit x bezeichnet. Wir stellen uns t als **Zeit** und $x(t)$ als **Ort** vor. Die Punkte $x(a)$ und $x(b)$ werden als **Anfangs-** bzw. **Endpunkt** der Kurve bezeichnet. Ist $x(a) = x(b)$, so heißt die Kurve **geschlossen**.

Bemerkung: Es gibt verschiedene Varianten der Begriffsbildung in der Literatur. So ist es unter anderem auch üblich den Wertebereich $C = x(I)$ als die Spur der Kurve zu bezeichnen und dann unter der Kurve die Funktion x zu verstehen. Oder es wird zwischen „Weg“ (als die Funktion x) und „Kurve“ (als das Bild $x(I)$) unterschieden wie im Lehrbuch von Burg, Haf und Wille.

Zur graphischen Veranschaulichung von ebenen oder räumlichen Kurven skizzieren wir diese. Die folgenden Beispiele kennen wir schon:

Beispiele 3.2

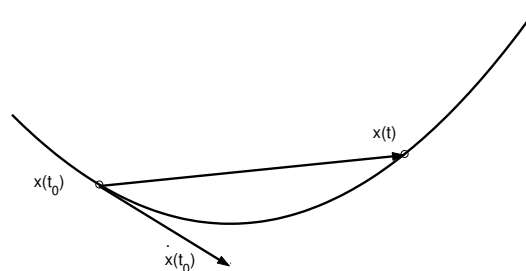
- (a) $x(t) = \begin{pmatrix} t \cos t \\ t \sin t \end{pmatrix}$, $0 \leq t \leq 4\pi$, also $x_1(t) = t \cos t$ und $x_2(t) = t \sin t$. Die Kurve ist eine Spirale.
- (b) $x(t) = \begin{pmatrix} t \\ \sin t \end{pmatrix}$, $0 \leq t \leq 2\pi$. Hier ist $x_1 = t$ und $x_2 = \sin t$, also $x_2 = \sin x_1$. Die Kurve ist daher der Graph der Sinusfunktion.
- (c) Es seien $a, b > 0$. Die Kurve $x(t) = \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix}$, $0 \leq t \leq 6\pi$, beschreibt eine Ellipse mit Halbachsen a und b , die dreimal durchlaufen wird.

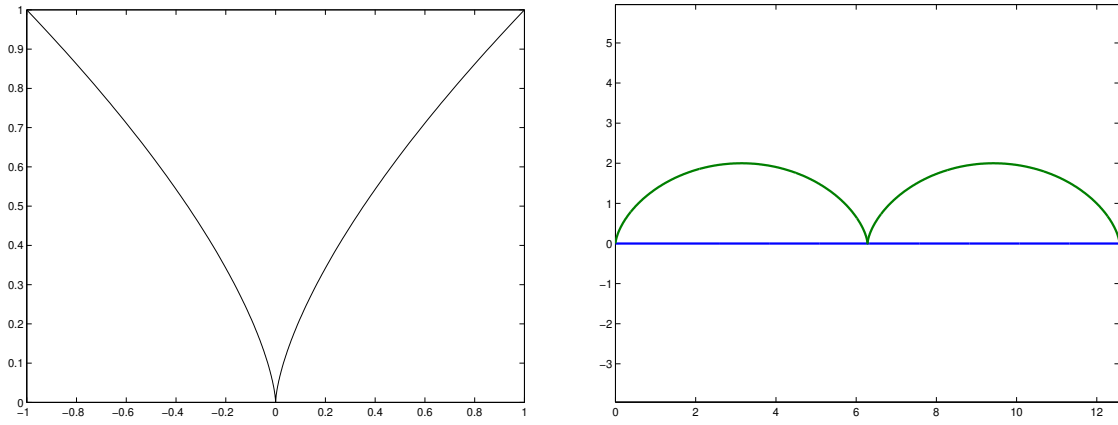
Aus dem Beispiel (b) ist ersichtlich, dass jede stetige Funktion $h : [a, b] \rightarrow \mathbb{R}$ eine Kurve $x : [a, b] \rightarrow \mathbb{R}^2$ liefert durch $x(t) = \begin{pmatrix} t \\ h(t) \end{pmatrix}$, $t \in [a, b]$. In diesem Sinne ist der Graph einer Funktion von \mathbb{R} nach \mathbb{R} durch eine Kurve gegeben.

Es ist klar, was differenzierbare und stetig differenzierbare Kurven sind (nämlich, wenn die Funktion x diese Eigenschaften hat). Wir schreiben meistens $\dot{x}(t) = (x'_1(t), \dots, x'_n(t))^T$, statt $x'(t)$.

Der Differenzenvektor

$(x(t) - x(t_0)) / (t - t_0)$ zeigt in Richtung der Sekante. Daher ist $\dot{x}(t_0)$ anschaulich ein **Tangentenvektor** oder **Tangentialvektor** an die Kurve im Punkt $x(t_0)$ und gibt physikalisch den **Geschwindigkeitsvektor** an, wenn wir uns t als die Zeit und $x(t)$ als den Ort zur Zeit t vorstellen. Durchlaufen wir die Kurve mit steigenden t -Werten, so zeigt $\dot{x}(t)$ „nach vorn“. Die Länge $\|\dot{x}(t)\|_2$ des Vektors $\dot{x}(t)$ ist der Betrag der Geschwindigkeit.





Kurven $x(t) = \begin{pmatrix} t^3 \\ t^2 \end{pmatrix}$, $|t| \leq 1$, und $x(t) = \begin{pmatrix} t - \sin t \\ 1 - \cos t \end{pmatrix}$, $0 \leq t \leq 4\pi$.

Der Vektor

$$\tau(t) := \frac{1}{\|\dot{x}(t)\|_2} \dot{x}(t)$$

heißt **Tangenteneinheitsvektor** an der Kurve im Punkt $x(t)$. Dieser ist wohldefiniert, falls $\dot{x}(t) \neq 0$ im Punkt $x(t)$ ist.

Beispiele 3.3 (a) Es gibt unendlich viele Möglichkeiten, eine Kurve zu parametrisieren. Betrachten wir z.B. die beiden Funktionen $x, y : [-1, 1] \rightarrow \mathbb{R}^2$, definiert durch

$$x(t) := \begin{pmatrix} t \\ t \end{pmatrix}, \quad -1 \leq t \leq 1, \quad y(t) := \begin{cases} \begin{pmatrix} t \\ t \end{pmatrix}, & -1 \leq t \leq 0, \\ \begin{pmatrix} t^2 \\ t^2 \end{pmatrix}, & 0 < t \leq 1. \end{cases}$$

Offensichtlich beschreiben x und y dieselbe Kurve, sind aber verschiedene Funktionen. Die Winkelhalbierende wird mit verschiedenen Geschwindigkeiten durchlaufen. Die Parametrisierung x ist stetig differenzierbar auf $[-1, 1]$, die Parametrisierung y nicht, da die Ableitung \dot{y} bei $t = 0$ nicht existiert. In beiden Fällen ist der Tangenteneinheitsvektor gegeben durch $\tau(t) = (1, 1)^\top / \sqrt{2}$, und dieser ist stetig (sogar konstant).

(b) Wir betrachten die folgende Kurve, parametrisiert durch

$$x(t) := \begin{pmatrix} t^3 \\ t^2 \end{pmatrix}, \quad -1 \leq t \leq 1.$$

Die Parametrisierung ist offenbar stetig differenzierbar. Auflösen nach x_2 ergibt $x_2 = |x_1|^{2/3}$, $|x_1| \leq 1$, und der Graph dieser Funktion hat eine Spitze (siehe linke Abbildung unten). Anschaulich lässt sich diese Kurve wohl nicht als glatt bezeichnen. Der Tangenteneinheitsvektor ist (nachrechnen!)

$$\tau(t) = \frac{1}{\sqrt{9t^2 + 4}} \begin{pmatrix} 3|t| \\ 2 \operatorname{sign} t \end{pmatrix},$$

und dieser ist nicht stetig in 0!

Auch die „Zykloide“ $x(t) = \begin{pmatrix} t - \sin t \\ 1 - \cos t \end{pmatrix}$, $0 \leq t \leq 4\pi$, ist differenzierbar, aber anschaulich nicht glatt (Abbildung rechts).

An den Beispielen erkennen wir, dass die Differenzierbarkeit der Parametrisierung x nicht unbedingt etwas über die anschauliche Glattheit der Kurve aussagt – wohl aber die Stetigkeit des Tangenteneinheitsvektors.

Das erste der Beispiele 3.3 motiviert die folgende Definition:

Definition 3.4 Zwei Parametrisierungen $x : [a, b] \rightarrow \mathbb{R}^n$ und $y : [c, d] \rightarrow \mathbb{R}^n$ einer Kurve heißen **äquivalent**, wenn es eine stetige, streng monoton steigende und bijektive **Parametertransformation** $t : [c, d] \rightarrow [a, b]$ gibt mit $x(t(\tau)) = y(\tau)$ für alle $\tau \in [c, d]$.

Bemerkung: Bei äquivalenten Parametrisierungen wird also die Kurve nur mit unterschiedlichen „Geschwindigkeiten“ durchlaufen. Im Beispiel 3.3 (a) wähle z.B.

$$t(\tau) := \begin{cases} \tau, & -1 \leq \tau \leq 0, \\ \tau^2, & 0 < \tau \leq 1, \end{cases}$$

als Parametertransformation.

Jetzt präzisieren wir den anschaulichen Begriff der „Glattheit“.

Definition 3.5 Eine Kurve ist **glatt** oder **regulär**, wenn es eine Parametrisierung $x : [a, b] \rightarrow \mathbb{R}^n$ gibt, die auf (a, b) stetig differenzierbar ist mit $\dot{x}(t) \neq 0$ für alle $t \in (a, b)$ und der Eigenschaft, dass der Tangenteneinheitsvektor $t \mapsto \tau(t)$ stetig auf $[a, b]$ ergänzbar ist.

Eine Kurve beschrieben durch $x : [a, b] \rightarrow \mathbb{R}^n$ heißt **stückweise glatt**, wenn es eine Unterteilung $a = t_0 < t_1 < \dots < t_p = b$ gibt, so dass jedes Kurvenstück $x((t_{j-1}, t_j))$ zu einer glatten Kurve auf $[t_{j-1}, t_j]$ ergänzt werden kann.

Für stückweise glatte Kurven ist also in jedem Punkt $x(t)$ mit $t \notin \{t_1, \dots, t_{p-1}\}$ der Tangenteneinheitsvektor definiert. In den „Nahtstellen“ t_j existiert jeweils der rechts- und der linksseitige Grenzwert.

Beispiel 3.6

Die Ellipse ist gegeben durch die Lösungsmenge der Gleichung $(x_1/a)^2 + (x_2/b)^2 = 1$. Die übliche Parametrisierung ist $x(t) = \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix}$, $0 \leq t \leq 2\pi$. Damit ist

$$\dot{x}(t) = \begin{pmatrix} -a \sin t \\ b \cos t \end{pmatrix} \quad \text{und} \quad \tau(t) = \frac{1}{\sqrt{a^2 \sin^2 t + b^2 \cos^2 t}} \begin{pmatrix} -a \sin t \\ b \cos t \end{pmatrix}$$

für den Punkt $z = x(t)$. Eine andere Möglichkeit, die Ellipse zu parametrisieren, ist

$$y(\varphi) := \frac{1}{\sqrt{a^2 \cos^2 \varphi + b^2 \sin^2 \varphi}} \begin{pmatrix} a^2 \cos \varphi \\ b^2 \sin \varphi \end{pmatrix}, \quad 0 \leq \varphi \leq 2\pi,$$

denn auch für diese Parametrisierung gilt $(\frac{y_1}{a})^2 + (\frac{y_2}{b})^2 = 1$. In diesem Fall ist

$$\dot{y}(\varphi) = \frac{a^2 b^2}{(a^2 \cos^2 \varphi + b^2 \sin^2 \varphi)^{3/2}} \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix},$$

und der Tangenteneinheitsvektor an der Stelle $z = y(\varphi)$ hat die einfache Form $\tau(\varphi) = \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix}$.

Sei nun durch $x : [a, b] \rightarrow \mathbb{R}^n$ eine glatte Kurve gegeben. Für den Betrag der Geschwindigkeit gilt

$$\|\dot{x}(t)\|_2 = \sqrt{\sum_{j=1}^n |\dot{x}_j(t)|^2} \neq 0 \quad \text{für alle } t \in [a, b].$$

Definition 3.7 Die glatte Kurve heißt **nach der Bogenlänge parametrisiert**, wenn für die Parametrisierung $z : [a, b] \rightarrow \mathbb{R}^n$ gilt $\|\dot{z}(s)\|_2 = 1$ für alle $s \in [a, b]$. Die Intervalllänge $L := b - a$ heißt dann die **Länge** der Kurve. Ist $x : [a, b] \rightarrow \mathbb{R}^n$ eine beliebige Parametrisierung, so definieren wir die Bogenlänge

$$s(t) = \int_a^t \|\dot{x}(\tau)\|_2 \, d\tau, \quad a \leq t \leq b,$$

und es ist

$$L = \int_a^b \|\dot{x}(\tau)\|_2 \, d\tau$$

die Länge der Kurve.

Bemerkung: Zu jeder glatten Kurve gibt es eine äquivalente Parametrisierung nach der Bogenlänge. Dazu konstruieren wir zu einer beliebigen Parametrisierung x eine Umparametrisierung z so, dass der Betrag 1 ist, d.h. „wir durchlaufen die Kurve mit konstanter Geschwindigkeit 1“, vgl. Definition 3.4.

Betrachte die Bogenlänge $s(t)$ und die Länge $L = s(b)$. Da s streng monoton wachsend ist, existiert ihre Umkehrfunktion $t := s^{-1} : [0, L] \rightarrow [a, b]$. Wir betrachten jetzt die durch t umparametrisierte Kurve $z(s) := x(t(s))$, $0 \leq s \leq L$. Dann sind die Bilder von z und x gleich, und mit der Kettenregel rechnen wir aus: $z'(s) = \dot{x}(t(s)) t'(s)$. Die Ableitung von t können wir auch ausrechnen (Ableitung der Umkehrfunktion):

$$t'(s) = \frac{1}{s'(t)} \Big|_{t=t(s)} = \frac{1}{\|\dot{x}(t(s))\|_2},$$

und daher

$$z'(s) = \frac{1}{\|\dot{x}(t(s))\|_2} \dot{x}(t(s)).$$

Diese Umparametrisierung liefert das Gewünschte, denn $\|z'(s)\|_2 = 1$ für alle $s \in [0, L]$.

Beispiele 3.8

- (a) Zunächst betrachten wir den **Kreis** mit Radius $a > 0$, parametrisiert durch $x(t) = \begin{pmatrix} a \cos t \\ a \sin t \end{pmatrix}$, $0 \leq t \leq 2\pi$. Dann ist

$$\dot{x}(t) = \begin{pmatrix} -a \sin t \\ a \cos t \end{pmatrix}, \quad \|\dot{x}(t)\|_2 = a,$$

und die Umparametrisierung $s(t) = at$, $t(s) = s/a$ liefert

$$z(s) = \begin{pmatrix} a \cos(s/a) \\ a \sin(s/a) \end{pmatrix}, \quad 0 \leq s \leq 2\pi a, \quad \dot{z}(s) = \begin{pmatrix} -\sin(s/a) \\ \cos(s/a) \end{pmatrix}, \quad \|\dot{z}(s)\|_2 = 1.$$

Die Länge der Kurve ist wie erwartet $L = \int_0^{2\pi} \|\dot{x}(t)\|_2 \, dt = 2\pi a$.

(b) Die bereits erwähnte **Zykloide** ist parametrisiert durch

$$x(t) = \begin{pmatrix} t - \sin t \\ 1 - \cos t \end{pmatrix}, \quad t \in \mathbb{R},$$

siehe Abbildung (unten). Damit ist

$$\dot{x}(t) = \begin{pmatrix} 1 - \cos t \\ \sin t \end{pmatrix} \quad \text{und} \quad \|\dot{x}(t)\|_2 = \sqrt{(1 - \cos t)^2 + \sin^2 t} = \sqrt{2 - 2 \cos t}.$$

Mit $1 - \cos t = 2 \sin^2(t/2)$ ist schließlich

$$\|\dot{x}(t)\|_2 = 2 \sin \frac{t}{2} \quad \text{und} \quad s(t) = \int_0^t \|\dot{x}(\tau)\|_2 d\tau = 2 \int_0^t \sin \frac{\tau}{2} d\tau = 4(1 - \cos \frac{t}{2}),$$

und insbesondere ist die Länge eines Bogens gegeben durch $L = s(2\pi) = 8$.

(c) Die **Astroide** ist gegeben durch die Gleichung

$$x_1^{2/3} + x_2^{2/3} = a^{2/3},$$

siehe Abbildung rechts. Im ersten Quadranten (d.h. $x_1, x_2 > 0$) kann sie parametrisiert werden durch

$$x(t) = a \begin{pmatrix} (1-t)^{3/2} \\ t^{3/2} \end{pmatrix}, \quad 0 \leq t \leq 1,$$

also

$$\dot{x}(t) = \frac{3a}{2} \begin{pmatrix} -(1-t)^{1/2} \\ t^{1/2} \end{pmatrix}, \quad 0 \leq t \leq 1,$$

und

$$\|\dot{x}(t)\|_2 = \frac{3a}{2} \sqrt{(1-t) + t} = \frac{3a}{2} \neq 0.$$

Also ist x eine glatte Kurve mit der Länge

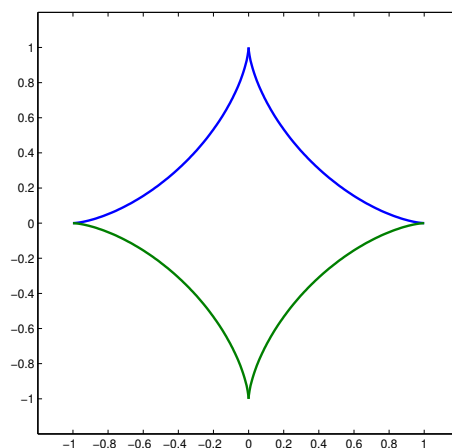
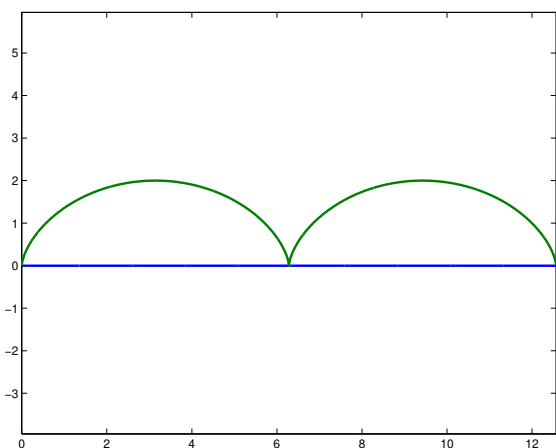
$$L = \int_0^1 \|\dot{x}(t)\|_2 dt = \frac{3a}{2}.$$

Der Gesamtumfang ist $4L = 6a$.

(d) Schließlich betrachten wir noch die **Ellipse** $x(t) = \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix}$, $0 \leq t \leq 2\pi$. Dann ist

$$\|\dot{x}(t)\|_2 = \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} \quad \text{und} \quad L = \int_0^{2\pi} \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} dt.$$

Die – natürlich existierende – Stammfunktion des Integranden kann nicht elementar angegeben werden. Daher ist dieses Integral nicht über den 1. Hauptsatz berechenbar. (Es heißt aus naheliegenden Gründen „elliptisches Integral“). Wir können zu seiner näherungsweisen Berechnung etwa die Trapez- oder die Simpsonregel benutzen.



Zykloide und Astroide (für $a = 1$)

Bemerkung: Ist eine Kurve durch $x : I \rightarrow \mathbb{R}^n$ nach der Bogenlänge parametrisiert, d.h. $\|\dot{x}(s)\| = 1$ für jedes $s \in I$, so ist $\dot{x}(s)$ der Tangenteneinheitsvektor. Ist x in diesem Fall zweimal differenzierbar, so ist $\|\ddot{x}(s)\|$ die **Krümmung** der Kurve an der Stelle $x(s)$. Wegen $\dot{x}(s)^\top \dot{x}(s) = \|\dot{x}(s)\|_2^2 = 1$ folgt durch Differentiation mit der Produktregel $2\dot{x}(s)^\top \ddot{x}(s) = 0$. Dies bedeutet, dass die Vektoren $\ddot{x}(s)$ und $\dot{x}(s)$ senkrecht aufeinander stehen. Der Vektor

$$\nu(s) := \frac{1}{\|\ddot{x}(s)\|_2} \ddot{x}(s) \in \mathbb{R}^n$$

wird **Hauptnormalen(einheits)vektor** im Punkt $x(s)$ genannt.

Allgemein gilt für **ebene** Kurven, d.h. im \mathbb{R}^2 ,

$$\nu(t) = \pm \frac{1}{\|\dot{x}(t)\|_2} \begin{pmatrix} \dot{x}_2(t) \\ -\dot{x}_1(t) \end{pmatrix} \in \mathbb{R}^2,$$

für jede beliebige Parametrisierung (also nicht notwendig nach der Bogenlänge).

Für **räumliche** Kurven, d.h. im \mathbb{R}^3 , wird noch der **Binormalenvektor** definiert, der sich bei einer Parametrisierung nach der Bogenlänge durch

$$\beta(s) := \dot{x}(s) \times \nu(s) \in \mathbb{R}^3$$

angeben lässt. Es gilt stets $\|\beta(s)\|_2 = 1$ und $\tau(s), \nu(s), \beta(s)$ stehen paarweise senkrecht aufeinander. In Abhängigkeit von s heißen $(\tau(s), \nu(s), \beta(s))$ das **begleitende Dreibein** der Kurve an der Parameterstelle s bzw. im Punkt $x(s)$.

Besonders einfache Kurven sind natürlich die Geraden in **Parameterform**:

$$x(t) = u + tv, \quad t \in \mathbb{R},$$

mit Vektoren $u, v \in \mathbb{R}^n$. Hier ist $\dot{x}(t) = v$ konstant. Wie aus HM II bekannt ist, können Geraden in der Ebene (!) auch **implizit**, d.h. in der Hesseschen Form $w^\top x = \rho$ beschrieben werden, d.h. ausgeschrieben, als Lösung der Gleichung $w_1 x_1 + w_2 x_2 = \rho$. Hier steht w senkrecht auf der Geraden. Ist w ein Einheitsvektor, so ist $|\rho|$ der Abstand der Geraden vom Ursprung. In HM II hatten wir geübt, wie man von der einen in die andere Darstellung übergeht.

Ganz analog können beliebige Kurven im \mathbb{R}^2 neben der Parameterdarstellung auch **implizit** in der Form $f(x) = 0$ beschrieben werden mit einer Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Z.B. wird die Ellipse

als Lösungsmenge der Gleichung $(x_1/a)^2 + (x_2/b)^2 = 1$ angegeben. Mit dem Satz über implizite Funktionen kommen wir von der impliziten Form zur Parameterdarstellung! Ist nämlich \hat{x} ein Punkt der Kurve, also $f(\hat{x}) = 0$ und etwa $f_{x_1}(\hat{x}) \neq 0$, so können wir lokal nach x_1 auflösen, d.h. es gibt ein Intervall $[\hat{x}_2 - \varepsilon, \hat{x}_2 + \varepsilon]$ mit $f(x_1(x_2), x_2) = 0$ für alle $x_2 \in [\hat{x}_2 - \varepsilon, \hat{x}_2 + \varepsilon]$ und $x_1(\hat{x}_2) = \hat{x}_1$ (Satz 1.30). Damit beschreibt $x(t) = (x_1(t), t)$, $t \in [\hat{x}_2 - \varepsilon, \hat{x}_2 + \varepsilon]$, die Kurve lokal um \hat{x} .

Ist analog $f_{x_2}(\hat{x}) \neq 0$, so können wir lokal nach x_2 auflösen. In anderen Worten:

Ist $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ eine stetig differenzierbare Funktion und $f(\hat{x}) = 0$ und $\nabla f(\hat{x}) \neq 0$ für ein $\hat{x} \in \mathbb{R}^2$, so können wir die Gleichung $f(x) = 0$ schreiben als $f(x(t)) = 0$, $t \in [\alpha, \beta]$, mit einer Kurve $t \mapsto x(t)$, die durch \hat{x} verläuft.

So lassen sich also Höhenlinien der Funktion f als Kurven darstellen. Sei zu einem Punkt x mit $f(x) = \rho$ die entsprechende Niveaulinie von f durch $\psi(t)$ beschrieben, d.h. es gilt $f(\psi(t)) = \rho$, so folgt $\nabla f(\psi(t))^\top \psi'(t) = 0$. Also ist der Gradient $\nabla f(x)$ stets **senkrecht** zur Niveaulinie (genauer zur Tangente der Höhenlinie) in jeder Stelle $x \in \mathbb{R}^n$.

Wir wollen nun noch **Kurvenintegrale** erklären, wobei wir den Spezialfall der Bogenlänge schon kennen. Sei also $x : I \rightarrow \mathbb{R}^n$ Parametrisierung einer glatten Kurve im \mathbb{R}^n , $I = [a, b]$ und $f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ eine Abbildung von n Variablen, die auf der Kurve $C = x(I) \subset \mathbb{R}^n$ definiert ist. Man sagt kürzer, dass „**f längs der Kurve x definiert ist**“. Für $n = 2$ ist der Graph $\{(z, f(z)) : z \in C\} \subset \mathbb{R}^3$ anschaulich ein gekrümmter Bogen im Raum über C . Ist zum Beispiel $f(z)$ die Massendichte eines gekrümmten Drahtes C an der Stelle $z \in C$, so wollen wir etwa die Gesamtmasse von C berechnen.

Wir wollen zur Motivation wieder **Treppenfunktionen** auf C betrachten. Sei also $a = t_0 < t_1 < \dots < t_p = b$ eine Unterteilung von I . Dies liefert durch $x^{(i)} = x(t_i)$, $i = 0, \dots, p$, eine Unterteilung von $x(I)$. Sei $I_i = (t_{i-1}, t_i)$, $i = 1, \dots, p$, das i -te Teilintervall. Die Funktion $\varphi : x(I) \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ sei eine Treppenfunktion, also konstant mit Wert c_i auf dem Kurvenstück $x(I_i)$. Wir stellen uns etwa vor, dass c_i die Massendichte des Drahtstücks $C_i = x(I_i)$ sei. Dann ist die Gesamtmasse des Drahtes $C = x(I)$ gerade

$$\sum_{i=1}^p c_i L(C_i),$$

wobei wieder $L(C_i)$ die **Länge** des Kurvenstücks C_i sei, vgl. Definition 3.7. Es ist $L(C_i) = \int_{t_{i-1}}^{t_i} \|\dot{x}(t)\|_2 dt$ und daher

$$\sum_{i=1}^p c_i L(C_i) = \sum_{i=1}^p c_i \int_{t_{i-1}}^{t_i} \|\dot{x}(t)\|_2 dt = \int_a^b \varphi(x(t)) \|\dot{x}(t)\|_2 dt.$$

Dies motiviert die folgende Definition:

Definition 3.9 Sei durch $x : I = [a, b] \rightarrow \mathbb{R}^n$ eine glatte Kurve $C = x(I)$ parametrisiert und $f : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ eine stetige Funktion auf C , so heißt

$$\int_C f(x) d\ell = \int_a^b f(x(t)) \|\dot{x}(t)\|_2 dt$$

das **Kurvenintegral** der skalaren Funktion f über C .

Ist x nur stückweise glatt, besteht also aus endlich vielen glatten Kurvenstücken $x^{(j)} : [t_{j-1}, t_j] \rightarrow \mathbb{R}^n$, $j = 1, \dots, p$, so ist das Kurvenintegral erklärt durch

$$\int_C f(x) d\ell := \sum_{j=1}^p \int_{C_j} f(x) d\ell = \sum_{j=1}^p \int_{t_{j-1}}^{t_j} f(x^{(j)}(t)) \|\dot{x}^{(j)}(t)\|_2 dt,$$

wobei C_j die Kurve mit Parametrisierung $x^{(j)}$ ist.

Die Notation $\int_C f(x) d\ell$ betont, dass das Integral nur von der Menge C und der Funktion $f : C \rightarrow \mathbb{R}$ abhängt, aber nicht von der Parametrisierung x . Man erkennt nämlich leicht mit der Substitutionsregel, dass äquivalente Kurvenstücke dasselbe Integral liefern. Insbesondere hängt das Integral **nicht** von der **Orientierung** von x ab: Durchlaufen wir die Kurve in entgegengesetzter Richtung, d.h. parametrisieren wir die Kurve durch $y(t) = x(-t)$, $t \in [-b, -a]$, so ändert sich das Integral nicht.

Beispiele 3.10 aus der Mechanik

Es sei eine Kurve $x(t) \in \mathbb{R}^3$, $a \leq t \leq b$, im Raum gegeben. Beachte, dass hier t der Kurvenparameter und nicht unbedingt die Zeit ist. Es sei $\rho : C \rightarrow \mathbb{R}$ die **Massendichte** der Kurve. **Masse** $M \in \mathbb{R}$ und **Schwerpunkt** $\bar{x} \in \mathbb{R}^3$ der Kurve sind definiert durch die Kurvenintegrale

$$M := \int_C \rho(x) d\ell \in \mathbb{R},$$

$$\bar{x}_j := \frac{1}{M} \int_C x_j \rho(x) d\ell, \quad j = 1, 2, 3.$$

(a) Als konkretes Beispiel betrachten wir einen Draht C von der Form eines Halbkreises mit Radius a . Die Dichte $\rho(x)$, $x \in C$, sei direkt proportional zum Abstand des Punktes x von der Verbindungsgeraden durch die Endpunkte. Wir wollen Masse und Schwerpunkt bestimmen. Zunächst skizzieren wir die Situation und wählen natürlich den Halbkreis um Mittelpunkt 0 und Endpunkten $(-a, 0)$, $(a, 0)$ (alles in der Ebene $x_3 = 0$). Eine Parametrisierung dieses Halbkreises ist $x(t) = \begin{pmatrix} a \cos t \\ a \sin t \\ 0 \end{pmatrix}$, $0 \leq t \leq \pi$. Die Dichte ist dann $\rho(x) = k x_2$ mit einer Proportionalitätskonstanten $k > 0$. Es folgt

$$M = k a^2 \int_0^\pi \sin t dt = 2 k a^2 \quad \text{und}$$

$$\bar{x} = \frac{k a^2}{M} \int_0^\pi \begin{pmatrix} a \cos t \\ a \sin t \end{pmatrix} \sin t dt = \frac{1}{2} \int_0^\pi \begin{pmatrix} a \cos t \\ a \sin t \end{pmatrix} \sin t dt.$$

Wir berechnen beide Komponenten getrennt. Es ist

$$\bar{x}_1 = \frac{a}{2} \int_0^\pi \cos t \sin t \, dt = \frac{a}{4} \int_0^\pi \sin(2t) \, dt = -\frac{a}{8} \cos(2t) \Big|_0^\pi = 0,$$

und

$$\bar{x}_2 = \frac{a}{2} \int_0^\pi \sin^2 t \, dt = \frac{a}{4} \int_0^\pi (1 - \cos(2t)) \, dt = \frac{a}{4} \left[t - \frac{1}{2} \sin(2t) \right]_0^\pi = \frac{a\pi}{4}.$$

(b) Eine **Zykloide** ist parametrisiert durch $x(t) = r \begin{pmatrix} t - \sin t \\ 1 - \cos t \end{pmatrix}$, $0 \leq t \leq 2\pi$, mit einer Konstanten $r > 0$. Für die homogene Massendichte $\rho(x) = 1$ ergeben sich Masse und Schwerpunkt als $M = 8r$ und $\bar{x} = (\pi r, \frac{4}{3}r)^\top$.

Spezialfall: Wir hatten gesehen, dass Kurven in der Ebene \mathbb{R}^2 von der Form $x_2 = \psi(x_1)$, $x_1 \in [a, b]$, durch $x(t) = (t, \psi(t))^\top$, $t \in [a, b]$, parametrisiert werden können. Für solche speziellen Kurven ist daher

$$\int_C f(x) \, d\ell = \int_a^b f(t, \psi(t)) \sqrt{1 + \psi'(t)^2} \, dt.$$

Auch Kurvenintegrale für **vektorwertige** Funktionen $F : C \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ treten häufig in Anwendungen auf. Sei also C wieder eine glatte Kurve⁶. Zwei verschiedene Varianten sollten wir kennen.

(a) Eine übliche Notation definiert $\int_C F(x) \, d\ell$ komponentenweise, d.h. $\int_C F(x) \, d\ell \in \mathbb{R}^n$ mit $(\int_C F(x) \, d\ell)_j = \int_C F_j(x) \, d\ell$, $j = 1, \dots, n$. Das Integral ist also ein Vektor! Als Beispiel haben wir die Formel für den Schwerpunkt (s.o.),

$$\bar{x} = \frac{1}{M} \int_C x \rho(x) \, d\ell \in \mathbb{R}^3.$$

(b) Wichtig ist die zweite Variante:

Das **tangential orientierte Kurvenintegral** ist definiert durch

$$\int_C F(x) \cdot d\ell := \int_C F(x) \cdot \tau(x) \, d\ell = \int_a^b F(x(t)) \cdot \dot{x}(t) \, dt \quad (3.1)$$

mit Tangenteneinheitsvektor $\tau(x) = \dot{x}(t) / \|\dot{x}(t)\|_2$.

Die Größe $F(x) \cdot \tau(x)$ ist die **Tangentialkomponente** von $F(x)$. Motiviert durch die rechte Seite wird auch

$$\int_C F(x) \cdot d\ell = \int_C [F_1(x) dx_1 + \dots + F_n(x) dx_n] \quad (3.2)$$

⁶Die Übertragung auf stückweise glatte Kurven ist dann klar

geschrieben. Insbesondere wird für **skalare Funktionen** auch

$$\int_C f(x) dx_j := \int_a^b f(x(t)) \dot{x}_j(t) dt, \quad j = 1, \dots, n \quad (3.3)$$

verwendet. Wir bemerken, dass die Integrale (3.1), (3.2) und (3.3) sehr wohl von der **Orientierung** der Kurve abhängen: Durchlaufen wir die Kurve in entgegengesetzter Richtung, so wechselt τ sein Vorzeichen und damit auch das Integral!

Beispiel 3.11 Sei F ein Kraftfeld (z.B. die Gravitationskraft). Die **Arbeit** von F längs des Weges C ist gegeben durch das Wegintegral

$$A = \int_C F(x) \cdot d\ell = \int_C F(x) \cdot \tau(x) d\ell = \int_a^b F(x(t)) \cdot \dot{x}(t) dt.$$

Als konkretes Beispiel betrachten wir die Schraubenlinie

$$C : x(t) = \left(r \cos t, r \sin t, \frac{h}{2\pi} t \right)^\top \in \mathbb{R}^3, \quad t \in [0, 4\pi],$$

im Kraftfeld

$$F(x) = f_0(x_1, x_2, (x_3 - h)^2)^\top \in \mathbb{R}^3$$

für ein $f_0 \in \mathbb{R}$. Dann errechnen wir

$$\dot{x}(t) = \left(-r \sin t, r \cos t, \frac{h}{2\pi} \right)^\top \in \mathbb{R}^3,$$

und daher

$$A = f_0 \int_0^{4\pi} \frac{h^3}{2\pi} \left(\frac{t}{2\pi} - 1 \right)^2 dt = \frac{f_0 h^3}{2\pi} \frac{2\pi}{3} \left(\frac{t}{2\pi} - 1 \right)^3 \Big|_0^{4\pi} = \frac{2}{3} f_0 h^3.$$

Insbesondere ist die Arbeit vom Radius unabhängig.

Hat das Kraftfeld F die spezielle Form $F = \nabla f$ für eine skalare Funktion $f : G \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$ (F heißt dann **konservativ** und f das zu F gehörige **Potential**, s. Kapitel 1.1.), so ist mit der Kettenregel

$$A = \int_a^b \nabla f(x(t)) \cdot \dot{x}(t) dt = \int_a^b \frac{d}{dt} [f(x(t))] dt = f(x(b)) - f(x(a)).$$

Die Arbeit hängt in diesem Fall nur von der Potentialdifferenz der beiden Endpunkte $B = x(b)$ und $A = x(a)$ ab, die Arbeit ist **wegunabhängig**. In diesem Sinne wird das Potential f auch „Stammfunktion“ zu F genannt. Die Existenz von Stammfunktionen wurde im Fall $n = 2$ schon im ersten Kapitel im Zusammenhang mit den exakten Differentialgleichungen diskutiert.

3.2 Flächen und Flächenintegrale

Nun wollen wir analog zur Kurve den Begriff der **Fläche** definieren, allerdings nur im \mathbb{R}^3 .

Definition 3.12 (glatte Fläche im \mathbb{R}^3)

Sei $B \subset \mathbb{R}^2$ eine offene Menge mit Abschluss \overline{B} und $X : \overline{B} \rightarrow \mathbb{R}^3$ eine auf B stetig (partiell) differenzierbare Abbildung, so dass die Vektoren

$$X_{u_1}(u) := \frac{\partial X(u_1, u_2)}{\partial u_1} \in \mathbb{R}^3 \quad \text{und} \quad X_{u_2}(u) := \frac{\partial X(u_1, u_2)}{\partial u_2} \in \mathbb{R}^3$$

für jeden Punkt $u = (u_1, u_2) \in B$ linear unabhängig sind und die Vektorfunktion

$$\nu(u) := \frac{1}{\|X_{u_1}(u) \times X_{u_2}(u)\|_2} (X_{u_1}(u) \times X_{u_2}(u)) \in \mathbb{R}^3$$

stetig fortsetzbar auf \overline{B} ist.⁷ Dann heißt die Menge $S = X(B)$ eine **glatte Fläche** im \mathbb{R}^3 . Die Funktion X wird wiederum **Parametrisierung** der Fläche genannt.

Anschaulich sind die Vektoren $X_{u_1}(u)$ und $X_{u_2}(u)$ **Tangentialvektoren** an der Fläche im Punkt $X(u)$. Da sie linear unabhängig sind, spannen sie die **Tangentialebene** auf. Diese ist im Punkt $X(u)$ der Fläche gegeben durch die Abbildung

$$\mathbb{R}^2 \ni (s, t) \mapsto X(u) + s X_{u_1}(u) + t X_{u_2}(u) = X(u) + X'(u) \begin{pmatrix} s \\ t \end{pmatrix} \in \mathbb{R}^3.$$

Wir erkennen wieder, dass die Ebene in einer Parameterform vorliegt. Der Vektor

$$\nu(u) := \frac{1}{\|X_{u_1}(u) \times X_{u_2}(u)\|_2} (X_{u_1}(u) \times X_{u_2}(u)) \in \mathbb{R}^3$$

steht senkrecht auf der Fläche und heißt **Normaleneinheitsvektor** an die Fläche im Punkt $X(u)$. Aus der Definition einer glatten Fläche folgt, dass der Normaleneinheitsvektor $\nu(u)$ auf ganz \overline{B} existiert und stetig ist.

Beispiele 3.13

(a) Mit $R > 0$ sei X gegeben durch

$$X(u_1, u_2) = \begin{pmatrix} u_1 \\ u_2 \\ \sqrt{R^2 - u_1^2 - u_2^2} \end{pmatrix}, \quad u_1^2 + u_2^2 \leq R^2.$$

Also ist hier $B = \{u \in \mathbb{R}^2 : u_1^2 + u_2^2 < R^2\}$ und

$$X_{u_1}(u) = \begin{pmatrix} 1 \\ 0 \\ -\frac{u_1}{\sqrt{R^2 - u_1^2 - u_2^2}} \end{pmatrix}, \quad X_{u_2}(u) = \begin{pmatrix} 0 \\ 1 \\ -\frac{u_2}{\sqrt{R^2 - u_1^2 - u_2^2}} \end{pmatrix}.$$

Diese beiden Vektoren sind linear unabhängig (weshalb?), daher beschreibt $X = (X_1, X_2, X_3)^\top$ ein reguläres Flächenstück. Wegen $X_1^2 + X_2^2 + X_3^2 = R^2$ und $X_3 \geq 0$ ist die Fläche gerade die Oberfläche der oberen Halbkugel um 0 mit Radius R .

⁷Beachte, dass für $u \in B$ der Nenner ungleich Null ist, da die Vektoren $X_{u_1}(u)$ und $X_{u_2}(u)$ linear unabhängig sind.

(b) Mit $R > 0$ sei die Abbildung Y gegeben durch

$$Y(\theta, \varphi) = \begin{pmatrix} R \sin \theta \cos \varphi \\ R \sin \theta \sin \varphi \\ R \cos \theta \end{pmatrix}, \quad 0 < \theta < \pi/2, \quad 0 < \varphi < 2\pi.$$

Jetzt ist die Parametermenge gegeben durch $B = (0, \pi/2) \times (0, 2\pi) \subset \mathbb{R}^2$, also einem Rechteck. Die Fläche ist wiederum die obere Halbsphäre mit Radius R . Diese Parametrisierung ist durch Kugelkoordinaten gegeben und hat den Vorteil, dass mit ihr auch die ganze Kugeloberfläche parametrisiert werden kann (nehme das Intervall $(0, \pi)$ an Stelle von $(0, \pi/2)$ für θ , so ist die Kugeloberfläche ohne die Pole parametrisiert. An den Polen ist die Parametrisierung nicht eindeutig). Diese Parametrisierung ist überall beliebig oft differenzierbar und die Tangentialvektoren

$$Y_\theta(\theta, \varphi) = \begin{pmatrix} R \cos \theta \cos \varphi \\ R \cos \theta \sin \varphi \\ -R \sin \theta \end{pmatrix}, \quad Y_\varphi(\theta, \varphi) = \begin{pmatrix} -R \sin \theta \sin \varphi \\ R \sin \theta \cos \varphi \\ 0 \end{pmatrix}$$

stehen sogar senkrecht aufeinander. Der Normaleneinheitsvektor ist gegeben durch

$$\nu(\theta, \varphi) = \begin{pmatrix} \sin \theta \cos \varphi \\ \sin \theta \sin \varphi \\ \cos \theta \end{pmatrix}.$$

Analog zu Kurven im \mathbb{R}^2 können auch Flächen im \mathbb{R}^3 **implizit** als Lösung einer Gleichung $f(x_1, x_2, x_3) = 0$ mit einer Funktion $f : D \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$ gegeben sein. Ist $f(\hat{x}) = 0$ und $\nabla f(\hat{x}) \neq 0$, so ist wenigstens eine der partiellen Ableitungen f_{x_i} nicht Null im Punkt \hat{x} und wir können nach dieser Variablen auflösen. Ist also etwa $f_{x_1}(\hat{x}) \neq 0$, so können wir $x_1 = x_1(x_2, x_3)$ schreiben und haben lokal um \hat{x} , dass $f(x_1(x_2, x_3), x_2, x_3) = 0$, d.h. $f(X(u_1, u_2)) = 0$ mit $X(u_1, u_2) = (x_1(u_1, u_2), u_1, u_2)^\top$ ist. Genauso erhalten wir die Form $f(X(u_1, u_2)) = 0$ für die anderen Fälle.

Differenzieren wir die Gleichung $f(X(u_1, u_2)) = \rho$ partiell nach u_1 und u_2 , so erhalten wir mit der Kettenregel

$$\nabla f(X(u))^\top X_{u_1}(u) = 0 \quad \text{und} \quad \nabla f(X(u))^\top X_{u_2}(u) = 0.$$

Also steht $\nabla f(x)$ senkrecht auf der Tangentialebene, d.h. senkrecht auf der Niveauläche von f durch x .

Analog zu den Kurven wollen wir auch über Oberflächen integrieren. Für $S = \{(u_1, u_2, 0)^\top : u \in B \subseteq \mathbb{R}^2\}$, soll natürlich für das **Flächenintegral** die Identität

$$\iint_S f(x) \, d\sigma = \iint_B f(u_1, u_2, 0) \, d(u_1, u_2)$$

gelten. Als Motivation gehen wir (fast) genauso vor wie im ebenen Fall und betrachten zunächst Treppenfunktionen auf Flächen $S = X(B)$, wobei das Parametergebiet B eine endliche Vereinigung $B = \bigcup_{j=1}^p B_j$ von ebenen Dreiecken B_j sei. Mit gegebenen Zahlen $c_j \in \mathbb{R}$ sei die Treppenfunktion $\varphi : X(B) \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$ definiert durch $\varphi(x) = c_j$ für $x \in X(B_j)$, $j = 1, \dots, p$. Jede der Flächen $S_j = X(B_j)$ ist „dreiecksähnlich“, also ein gekrümmtes Dreieck im Raum. Das Flächenintegral für die Treppenfunktion φ sollte also durch

$$\sum_{j=1}^p c_j |S_j|$$

erklärt werden, wobei $|S_j|$ die (noch zu berechnende) Fläche des gekrümmten Dreiecks $S_j = X(B_j)$ ist. Zur näherungsweisen Berechnung (gerechtfertigt für kleine Dreiecke B_j) ersetzen wir das gekrümmte Dreieck $S_j = X(B_j)$ durch das echte ebene Dreieck, das von den drei Eckpunkten von S_j aufgespannt wird. Die Eckpunkte von B_j seien $u = (u_1, u_2)$, $v = (v_1, v_2)$, $w = (w_1, w_2)$. Dann hat $S_j = X(B_j)$ die Eckpunkte $X(u)$, $X(v)$ und $X(w)$, ist aber selbst natürlich kein Dreieck. Der Flächeninhalt $|S_j|$ ist dann aber etwa gleich dem Flächeninhalt des ebenen Dreiecks im \mathbb{R}^3 mit den Eckpunkten $X(u)$, $X(v)$, $X(w)$. Dieses Dreieck wird aufgespannt von den beiden Vektoren $X(w) - X(u)$ und $X(v) - X(u)$. Nach dem Mittelwertsatz angewendet auf jede Komponente X_k und jede Variable u_1, u_2 von X gilt

$$\begin{aligned} X(w_1, w_2) - X(u_1, u_2) &\approx X_{u_1}(u) (w_1 - u_1) + X_{u_2}(u) (w_2 - u_2), \\ X(v_1, v_2) - X(u_1, u_2) &\approx X_{u_1}(u) (v_1 - u_1) + X_{u_2}(u) (v_2 - u_2). \end{aligned}$$

Also ist nach Satz 2.18 von HM II:

$$\begin{aligned} |S_j| &\approx \frac{1}{2} \|(X(w) - X(u)) \times (X(v) - X(u))\|_2 \\ &\approx \frac{1}{2} \|(X_{u_1}(u) (w_1 - u_1) + X_{u_2}(u) (w_2 - u_2)) \times (X_{u_1}(u) (v_1 - u_1) + X_{u_2}(u) (v_2 - u_2))\|_2 \\ &= \frac{1}{2} \|X_{u_1}(u) \times X_{u_2}(u)\|_2 |(w_1 - u_1)(v_2 - u_2) - (v_1 - u_1)(w_2 - u_2)| \\ &= \frac{1}{2} \|X_{u_1}(u) \times X_{u_2}(u)\|_2 \left\| \begin{pmatrix} w_1 - u_1 \\ w_2 - u_2 \\ 0 \end{pmatrix} \times \begin{pmatrix} v_1 - u_1 \\ v_2 - u_2 \\ 0 \end{pmatrix} \right\|_2 \\ &= \|X_{u_1}(u) \times X_{u_2}(u)\|_2 |B_j| = \iint_{B_j} \|X_{u_1}(u) \times X_{u_2}(u)\|_2 d(u_1, u_2). \end{aligned}$$

Daher ist

$$\begin{aligned} \sum_{j=1}^p c_j |S_j| &\approx \sum_{j=1}^p c_j \iint_{B_j} \|X_{u_1}(u) \times X_{u_2}(u)\|_2 d(u_1, u_2) \\ &= \iint_B \varphi(X(u)) \|X_{u_1}(u) \times X_{u_2}(u)\|_2 d(u_1, u_2). \end{aligned}$$

Dies motiviert die folgende Definition

Definition 3.14 (Flächenintegral)

Sei $X : \overline{B} \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^3$ Parametrisierung eines regulären Flächenstücks $S := X(\overline{B}) \subset \mathbb{R}^3$ und $f : S \rightarrow \mathbb{R}$ stetig. Dann heißt

$$\iint_S f(x) \, do := \iint_B f(X(u)) \|X_{u_1}(u) \times X_{u_2}(u)\|_2 \, d(u_1, u_2)$$

das **Flächenintegral** der skalaren Funktion f über der Fläche S . Das Integral

$$\iint_S do := \iint_B \|X_{u_1}(u) \times X_{u_2}(u)\|_2 \, d(u_1, u_2)$$

heißt der **Flächeninhalt** der Fläche S .

Bemerkungen:

- (a) Hat man zwei Parametrisierungen $X : \overline{B} \rightarrow S$ und $Y : \overline{D} \rightarrow S$ derselben Fläche $S = X(\overline{B}) = Y(\overline{D})$, so lässt sich mit der Transformationsformel zeigen, dass die rechten Seiten in der Definition des Flächenintegrals übereinstimmen. Die Definition hängt also wirklich nur von der Fläche S ab und nicht von der Parametrisierung.
- (b) Das Symbol $\iint_S f(x) \, do$ soll daran erinnern, dass wir ein allgemeines Flächenintegral und nicht ein ebenes Gebietsintegral vorliegen haben.

Spezialfall: Häufig werden Flächen im \mathbb{R}^3 parametrisiert durch $x_3 = \varphi(x_1, x_2)$, $(x_1, x_2) \in B \subset \mathbb{R}^2$. Setzen wir $u_1 = x_1$ und $u_2 = x_2$, so kann die Fläche auch durch $X(u_1, u_2) = (u_1, u_2, \varphi(u_1, u_2))^\top$, $u = (u_1, u_2) \in B$, dargestellt werden. Man rechnet leicht aus, dass $\|X_{u_1}(u) \times X_{u_2}(u)\|_2 = \sqrt{\varphi_{u_1}(u)^2 + \varphi_{u_2}(u)^2 + 1}$ ist. Damit ist für solche speziellen Parametrisierungen:

$$\iint_S f(x) \, do = \iint_B f(u_1, u_2, \varphi(u_1, u_2)) \sqrt{1 + \varphi_{u_1}(u)^2 + \varphi_{u_2}(u)^2} \, d(u_1, u_2).$$

Natürlich kann man hier gleich die Buchstaben x_1 und x_2 (oder x und y) an Stelle von u_1 und u_2 nehmen.

Beispiele 3.15

(a) Berechnen Sie $J := \iint_S x^2 z \, do$, wobei S die Fläche des Kegels $z^2 = x^2 + y^2$ ist, die zwischen den Ebenen $z = 1$ und $z = 4$ liegt.

Hier ist also $z = \varphi(x, y) = \sqrt{x^2 + y^2}$, $1 \leq x^2 + y^2 \leq 16$, also $\varphi_x(x, y) = x/\sqrt{x^2 + y^2}$ und $\varphi_y(x, y) = y/\sqrt{x^2 + y^2}$ sowie $\sqrt{1 + \varphi_x(x, y)^2 + \varphi_y(x, y)^2} = \sqrt{2}$. Daher ist

$$J = \iint_S x^2 z \, do = \iint_B x^2 \sqrt{x^2 + y^2} \sqrt{2} \, d(x, y),$$

und dies ist ein ebenes Gebietsintegral. Hier ist $B = \{(x, y) \in \mathbb{R}^2 : 1 < x^2 + y^2 < 16\}$. Zur Berechnung führen wir natürlich ebene Polarkoordinaten ein: $x = r \cos t$ und $y = r \sin t$ mit

$t \in (0, 2\pi)$ und $r \in (1, 4)$. Dann ist

$$J = \sqrt{2} \int_1^4 \int_0^{2\pi} r^2 (\cos t)^2 r dt r dr = \sqrt{2} \pi \int_1^4 r^4 dr = \frac{1023}{5} \sqrt{2} \pi.$$

(b) Für eine Fläche $S = X(B)$ mit Parametrisierung $X : B \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^3$ und der Massendichte $\rho(x)$, $x \in S$, werden wie für eine Kurve **Masse** und **Schwerpunkt** definiert durch

$$M := \iint_S \rho(x) do = \iint_B \rho(X(u)) \|X_{u_1}(u) \times X_{u_2}(u)\|_2 d(u_1, u_2),$$

$$\bar{x} := \frac{1}{M} \iint_S x \rho(x) do = \frac{1}{M} \iint_B X(u) \rho(X(u)) \|X_{u_1}(u) \times X_{u_2}(u)\|_2 d(u_1, u_2).$$

Nicht nur die komponentenweise zu verstehende Notation von Flächenintegralen mit vektorwertigem Integranden wie beim Schwerpunkt ist üblich. Wichtiger ist der Begriff des Flusses eines Vektorfeldes:

Definition 3.16 Sei $X : \mathbb{R}^2 \supset \bar{B} \rightarrow \mathbb{R}^3$ Parametrisierung einer glatten Fläche $S = X(B)$,

$$\nu(x) = \nu(u) := \frac{1}{\|X_{u_1}(u) \times X_{u_2}(u)\|_2} (X_{u_1}(u) \times X_{u_2}(u)), \quad u = (u_1, u_2) \in B,$$

Normaleneinheitsvektor an S und $F : \mathbb{R}^3 \supseteq S \rightarrow \mathbb{R}^3$ ein stetiges Vektorfeld. Das Flächenintegral

$$\iint_S F(x) \cdot do = \iint_S F(x) \cdot \nu(x) do = \iint_B \underbrace{F(X(u)) \cdot (X_{u_1}(u) \times X_{u_2}(u))}_{= \text{Spatprodukt}} d(u_1, u_2)$$

heißt der **Fluss** des Vektorfeldes F durch S .

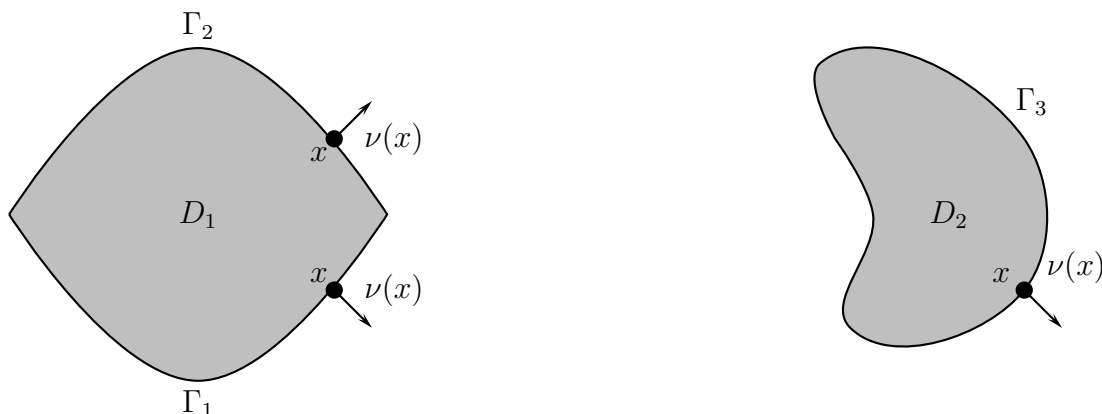
Der Fluss des Vektorfeldes hängt von der Parametrisierung insofern ab, dass sich das Vorzeichen ändert, wenn der Normalenvektor bei einer Umparametrisierung in die entgegengesetzte Richtung zeigt. Der Integrand $F(x) \cdot \nu(x)$ des Flusses ist gerade die Länge der Projektion des Vektors $F(x)$ auf die Richtung $\nu(x)$. Daher wird $F(x) \cdot \nu(x)$ auch die **Normalkomponente** des Vektorfeldes $F(x)$ an der Stelle $x \in S$ genannt.

3.3 Die Divergenz und der Gaußsche Integralsatz

Sei zunächst $D \subset \mathbb{R}^2$ eine offene beschränkte Menge mit Rand ∂D . Dieser kann im allgemeinen ziemlich wild aussehen und insbesondere aus mehreren Teilen bestehen, wenn D „Löcher“ besitzt. Wir setzen voraus, dass der Rand Vereinigung von endlich vielen stückweise glatten Kurven ist. Also ist $\partial D = \bigcup_{j=1}^p \Gamma_j$, und jedes Γ_j ist eine stückweise glatte Kurve. Ist $x : [a, b] \rightarrow \Gamma = x([a, b]) \subset \mathbb{R}^2$ Parametrisierung eines der Kurvenstücke, so sind die beiden Normaleneinheitsvektoren durch

$$\pm \frac{1}{\|\dot{x}(t)\|_2} \begin{pmatrix} \dot{x}_2(t) \\ -\dot{x}_1(t) \end{pmatrix}$$

für $t \in (a, b)$ erklärt. Der eine zeigt in das Gebiet hinein, der andere hinaus. Wir wählen nun das Vorzeichen so, dass der Normalenvektor ins **Äußere** des Gebiets D zeigt, bezeichnen ihn mit $\nu(x)$ für $x = x(t) \in \Gamma$, $t \in (a, b)$, und nennen ihn den **äußeren Normaleneinheitsvektor**. Für alle Punkte $x \in \partial D$ bis auf die „Nahtpunkte“, d.h. bis auf die endlich vielen Punkte, an denen die Kurvenstücke aneinanderstoßen, ist $\nu(x)$ erklärt. Da die Γ_j reguläre Kurvenstücke sind, springt der Normalenvektor nur an den Nahtstellen.



Beispiel 3.17

Sei D das Ringgebiet in der Ebene zwischen den Kreisen um den Ursprung mit Radien 1 und 2. Dann besteht ∂D aus den beiden regulären Kurvenstücken

$$x(t) = 2 \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \quad 0 \leq t \leq 2\pi, \quad \text{und} \quad y(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \quad 0 \leq t \leq 2\pi.$$

Die äußeren Normalenvektoren sind dann $\nu(x) = (\cos t, \sin t)^\top$ für $x = x(t)$ und $\nu(y) = -(\cos t, \sin t)^\top$ für $y = y(t)$, $0 \leq t \leq 2\pi$.

Bevor wir den wichtigen **Gaußschen Integralsatz** formulieren, führen wir noch einen Begriff ein.

Definition 3.18 (Divergenz eines Vektorfeldes)

Sei $D \subset \mathbb{R}^n$ offen ($n = 2$ oder 3) und $F : \overline{D} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ ein stetig differenzierbares Vektorfeld. Dann heißt

$$\operatorname{div} F(x) := \sum_{j=1}^n \frac{\partial F_j}{\partial x_j}(x), \quad x \in D,$$

die **Divergenz** des Vektorfeldes F . Man schreibt dafür auch $\operatorname{div} F = \nabla \cdot F$.

Satz 3.19 (Gaußscher Integralsatz)

Sei $D \subset \mathbb{R}^n$ ($n = 2$ oder $n = 3$) offen und beschränkt, und der Rand ∂D bestehe aus endlich vielen stückweise glatten Kurven- bzw. Flächenstücken. Sei ferner $F : \overline{D} \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ ein stetiges Vektorfeld, das stetig differenzierbar in D sei. Dann gilt für $n = 2$

$$\iint_D \operatorname{div} F(x) \, dx = \int_{\partial D} F(x) \cdot \nu(x) \, dl,$$

und analog für $n = 3$

$$\iiint_D \operatorname{div} F(x) \, dx = \iint_{\partial D} F(x) \cdot do = \iint_{\partial D} F(x) \cdot \nu(x) \, do.$$

wobei wieder $\nu(x)$ der **äußere Normaleneinheitsvektor** im Punkt $x \in \partial D$ ist.

Bemerkungen:

- (a) Auf der linken Seite steht ein Gebietsintegral über das Gebiet $D \subset \mathbb{R}^n$ (siehe Kapitel 2). Auf der rechten Seite steht ein Kurven- bzw. Flächenintegral über den Rand ∂D . Analog zu Definition 3.16 wird dieses Integral auch im zweidimensionalen Fall der Fluss genannt.
- (b) Wir erinnern daran, dass der Integrand $F(x) \cdot \nu(x)$ des Flusses gerade die **Normalkomponente** des Vektorfeldes $F(x)$ an der Stelle $x \in \partial D$ ist.

Beweis für den Fall eines Rechtecks von der Form $D = (a, b) \times (c, d) \subset \mathbb{R}^2$. Wir rechnen alle Integrale nacheinander aus. Dazu benutzen wir den Satz von Fubini über die iterierten Integrale und den Hauptsatz der Differential- und Integralrechnung:

$$\begin{aligned} \iint_D \operatorname{div} F(x) \, dx &= \int_c^d \int_a^b \frac{\partial F_1(x)}{\partial x_1} \, dx_1 dx_2 + \int_a^b \int_c^d \frac{\partial F_2(x)}{\partial x_2} \, dx_2 dx_1 \\ &= \int_c^d [F_1(b, x_2) - F_1(a, x_2)] \, dx_2 + \int_a^b [F_2(x_1, d) - F_2(x_1, c)] \, dx_1. \end{aligned}$$

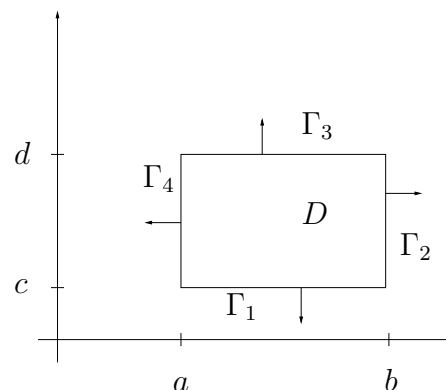
Der Rand besteht aus 4 Teilen

$$\partial D = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4,$$

die folgendermaßen parametrisiert werden:

$$\begin{aligned} \Gamma_1 : \quad x(t) &= (t, c)^\top, \quad a \leq t \leq b, \\ \Gamma_2 : \quad x(t) &= (b, t)^\top, \quad c \leq t \leq d, \\ \Gamma_3 : \quad x(t) &= (t, d)^\top, \quad a \leq t \leq b, \\ \Gamma_4 : \quad x(t) &= (a, t)^\top, \quad c \leq t \leq d, \end{aligned}$$

mit den nach außen gerichteten Normaleneinheitsvektoren $\nu(x) = (0, -1)^\top$, $(1, 0)^\top$, $(0, 1)^\top$, $(-1, 0)^\top$ auf Γ_1 , Γ_2 , Γ_3 , Γ_4 .



Also ist

$$\begin{aligned}
 \int_{\partial D} F(x) \cdot \nu(x) \, d\ell &= \int_a^b F(t, c) \cdot \begin{pmatrix} 0 \\ -1 \end{pmatrix} dt + \int_c^d F(b, t) \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} dt \\
 &+ \int_a^b F(t, d) \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} dt + \int_c^d F(a, t) \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} dt \\
 &= - \int_a^b F_2(t, c) \, dt + \int_c^d F_1(b, t) \, dt \\
 &+ \int_a^b F_2(t, d) \, dt - \int_c^d F_1(a, t) \, dt,
 \end{aligned}$$

und dies stimmt mit der linken Seite überein. □

Das besondere an diesem Satz ist, dass er für ganz allgemeine Gebiete gilt, also etwa auch für solche mit Löchern.

Bemerkung: Der Gaußsche Integralsatz bildet das Analogon zum **Hauptsatz der Differential- und Integralrechnung**: Sei $n = 1$, $D = (a, b)$ und $F : (a, b) \rightarrow \mathbb{R}$. Dann ist $\operatorname{div} F = F'$ und der Rand von D besteht aus den Punkten $\{a\}$ und $\{b\}$ mit „Normaleneinheitszahlen“ -1 bzw. $+1$. Daher erhalten wir

$$\begin{aligned}
 \int_{\partial D} F(x) \cdot \nu(x) \, d\ell &= F(a) \cdot (-1) + F(b) \cdot (+1) = F(b) - F(a), \text{ und} \\
 \iint_D \operatorname{div} F(x) \, dx &= \int_a^b F'(x) \, dx,
 \end{aligned}$$

und die Gleichheit gilt nach dem Hauptsatz der Differential- und Integralrechnung.

Die Forderung, dass $\nu(x)$ nach außen gerichtet ist, sollte immer nachgeprüft werden. Ist die Randkurve ∂D durch $x(t)$ parametrisiert, so ist immer $\dot{x}(t)$ Tangentialvektor und $(\dot{x}_2(t), -\dot{x}_1(t))^\top$ ein Normalenvektor im Punkt $x(t) \in \partial D$. Dieser Vektor kann in das Gebiet D hinein oder nach außen gerichtet sein. Wir definieren:

Definition 3.20 (*orientierte Parametrisierungen im \mathbb{R}^n , $n = 2$ oder 3*)

Sei $D \subset \mathbb{R}^n$ offen und beschränkt und ∂D bestehe aus endlich vielen stückweise glatten Kurvenstücken $x^{(j)} : I_j \rightarrow \mathbb{R}^2$ bzw. Flächenstücken $X^{(j)} : B_j \rightarrow \mathbb{R}^3$. Dann heißt die Parametrisierung (mathematisch) **positiv orientiert**, wenn für jede Teilparametrisierung $x = x^{(j)}$ bzw. $X = X^{(j)}$ der Vektor $(\dot{x}_2(t), -\dot{x}_1(t))^\top \in \mathbb{R}^2$ bzw. $X_{u_1}(u_1, u_2) \times X_{u_2}(u_1, u_2)$ im \mathbb{R}^3 ins Äußere von D zeigt.

Bemerkung: (a) Durchläuft man also eine positiv orientierte Parametrisierung mit steigendem Parameter t , so zeigt der Tangentialvektor $\dot{x}(t)$ nach vorn, der Vektor $(\dot{x}_2(t), -\dot{x}_1(t))^\top$ nach rechts, und das Gebiet D liegt links.

Der Rand von D im Beweis von Satz 3.19 ist bei der angegebenen Parametrisierung nicht positiv orientiert!

(b) Ist $x : [a, b] \rightarrow \Gamma \subset \partial D \subseteq \mathbb{R}^2$ ein **positiv orientiertes Kurvenstück** des Randes, so lässt sich das Integral über Γ auf der rechten Seite des Gaußschen Integralsatzes durch

$$\int_{\Gamma} F(x) \cdot \nu(x) \, d\ell = \int_{\Gamma} [F_1(x) \, dx_2 - F_2(x) \, dx_1]$$

schreiben.

Als Spezialfall ergeben sich die **Inhaltsformeln**: Der Rand ∂D eines offenen, beschränkten Gebiets $D \subset \mathbb{R}^2$ sei positiv orientiert und habe die Parametrisierung $x : [a, b] \rightarrow \mathbb{R}^2$. Dann gilt für den Inhalt $|D|$:

$$\begin{aligned} |D| &= \iint_D dx = \iint_D \operatorname{div} \begin{pmatrix} x_1 \\ 0 \end{pmatrix} dx = \iint_D \operatorname{div} \begin{pmatrix} 0 \\ x_2 \end{pmatrix} dx \\ &= \int_a^b x_1(t) \dot{x}_2(t) \, dt = - \int_a^b x_2(t) \dot{x}_1(t) \, dt \\ &= \int_{\partial D} x_1 \, dx_2 = - \int_{\partial D} x_2 \, dx_1. \end{aligned}$$

Durch Addition erhalten wir auch die Darstellung

$$|D| = \frac{1}{2} \iint_D \operatorname{div} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} dx = \frac{1}{2} \int_a^b [x_1(t) \dot{x}_2(t) - x_2(t) \dot{x}_1(t)] \, dt = \frac{1}{2} \int_{\partial D} [x_1 \, dx_2 - x_2 \, dx_1].$$

Beispiel 3.21 (Descartes'sches Blatt)

Betrachte die *implizit* gegebene Kurve

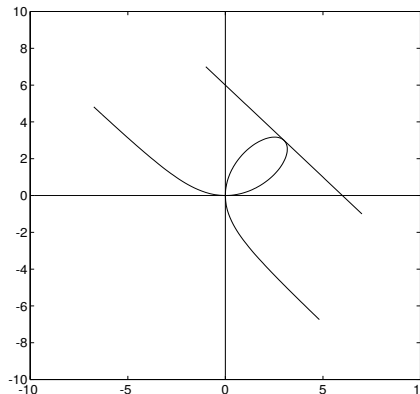
$$x_1^3 - 3a x_1 x_2 + x_2^3 = 0$$

mit einem Parameter $a > 0$. Durch Einsetzen ergibt sich, dass sie parametrisiert wird durch

$$x_1(t) = \frac{3a}{2} \frac{(1+t)(1-t)^2}{1+3t^2}, \quad x_2(t) = \frac{3a}{2} \frac{(1-t)(1+t)^2}{1+3t^2}, \quad t \in \mathbb{R}.$$

Der Ursprung ist ein Doppelpunkt mit Parametern $t = -1$ und $t = +1$. Für $t \in [-1, +1]$ durchläuft der Punkt also den Rand eines Gebietes D , das rechts für $a = 2$ skizziert ist. Um den Gebietsinhalt auszurechnen, schreiben wir

$$\begin{aligned} & x_1(t) \dot{x}_2(t) - x_2(t) \dot{x}_1(t) \\ &= x_1(t)^2 \frac{d x_2(t)}{d t x_1(t)} \\ &= x_1(t)^2 \frac{d 1+t}{d t 1-t} \\ &= \dots = \frac{9a^2}{2} \frac{1-2t^2+t^4}{(1+3t^2)^2} \end{aligned}$$



und daher

$$|D| = \frac{9a^2}{4} \left[\frac{t + \frac{1}{3}t^3}{1+3t^2} \right]_{-1}^{+1} = \frac{3}{2}a^2.$$

Bemerkenswert ist, dass das Gebiet D das Dreieck mit den Eckpunkten $(0, 0)$, $(3a, 0)$ und $(0, 3a)$ in drei Gebiete mit gleich großen Flächeninhalten zerlegt. (Das Dreieck hat ja den Flächeninhalt $9a^2/2$.)

Beim folgenden Spezialfall erkennen wir deutlich den **Hauptsatz der Differential- und Integralrechnung**. Sei $f : \overline{D} \rightarrow \mathbb{R}$ stetig und in D stetig differenzierbar. Dann gilt

$$\iint_D \frac{\partial}{\partial x_j} f(x) dx = \int_{\partial D} f(x) \nu_j(x) d\ell, \quad j = 1, 2, \quad \text{für } D \subset \mathbb{R}^2,$$

bzw.

$$\iiint_D \frac{\partial}{\partial x_j} f(x) dx = \iint_{\partial D} f(x) \nu_j(x) d\sigma, \quad j = 1, 2, 3, \quad \text{für } D \subset \mathbb{R}^3.$$

Zum Beweis dieser Formel setzen wir einfach $F(x) = f(x) e^{(j)}$ mit den Koordinateneinheitsvektoren $e^{(j)}$, $j = 1, 2, 3$.

Beispiel 3.22

Der Schwerpunkt $\bar{x} \in \mathbb{R}^2$ eines Gebiets $D \subset \mathbb{R}^2$ mit homogener Massendichte $\rho(x) = 1$ ist gegeben durch

$$\begin{aligned}\bar{x} &= \frac{1}{|D|} \iint_D x \, dx, \quad \text{also etwa} \\ \bar{x}_1 &= \frac{1}{|D|} \iint_D x_1 \, dx = \frac{1}{|D|} \iint_D \frac{\partial}{\partial x_2} (x_1 x_2) \, dx = \frac{1}{|D|} \int_{\partial D} x_1 x_2 \nu_2(x) \, d\ell.\end{aligned}$$

Besteht ∂D etwa aus dem regulären, positiv orientierten Kurvenstück $x : [a, b] \rightarrow \mathbb{R}^2$, so ist

$$\bar{x}_1 = -\frac{1}{|D|} \int_a^b x_1(t) x_2(t) \dot{x}_1(t) \, dt$$

und analog

$$\bar{x}_2 = \frac{1}{|D|} \int_a^b x_1(t) x_2(t) \dot{x}_2(t) \, dt.$$

Betrachten wir als Beispiel die rechte Hälfte der Ellipse $(x_1/a)^2 + (x_2/b)^2 \leq 1$, $x_1 \geq 0$, so besteht der Rand aus den beiden Teilen

$$x^{(1)}(t) = \begin{pmatrix} a \cos t \\ b \sin t \end{pmatrix}, \quad |t| \leq \frac{\pi}{2}, \quad x^{(2)}(t) = \begin{pmatrix} 0 \\ -t \end{pmatrix}, \quad |t| \leq b.$$

Dann ist der Rand positiv orientiert, denn die Normalenvektoren $(\dot{x}_2, -\dot{x}_1)^\top$ sind hier $(b \cos t, a \sin t)^\top$ bzw. $(-1, 0)^\top \in \mathbb{R}^2$ und daher nach außen gerichtet. Die Fläche der Halbellipse ist $|D| = \frac{\pi}{2}ab$. Daher ist

$$\begin{aligned}\bar{x}_1 &= -\frac{2}{\pi ab} \int_{-\pi/2}^{\pi/2} ab \cos t \sin t (-a \sin t) \, dt + 0 = \frac{2a}{\pi} \int_{-\pi/2}^{\pi/2} \sin^2 t \cos t \, dt \\ &= \frac{4a}{\pi} \int_0^{\pi/2} \sin^2 t \cos t \, dt = \frac{4a}{3\pi} [\sin^3 t]_0^{\pi/2} = \frac{4a}{3\pi}, \\ \bar{x}_2 &= \frac{2}{\pi ab} \int_{-\pi/2}^{\pi/2} ab \cos t \sin t (b \cos t) \, dt + 0 = \frac{2b}{\pi} \int_{-\pi/2}^{\pi/2} \cos^2 t \sin t \, dt = 0.\end{aligned}$$

Daher ist $\bar{x} = (4a/(3\pi), 0)^\top$.

Folgerungen:

Wir betrachten den Fall, dass das Vektorfeld F ein Produkt $F(x) = g(x) G(x)$ mit skalarer Funktion $g : \mathbb{R}^n \rightarrow \mathbb{R}$ und vektorwertiger Funktion $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist ($n = 2$ oder $n = 3$). Mit der Produktregel (nachrechnen!) folgt

$$\operatorname{div} F(x) = \operatorname{div} (g(x) G(x)) = \nabla g(x) \cdot G(x) + g(x) \operatorname{div} G(x).$$

Einsetzen in den Gauß'schen Satz liefert „partielle Integration“ in der Form

$$\int_D \nabla g \cdot G \, dx = \int_{\partial D} g G \cdot \nu \, d\ell - \int_D g \operatorname{div} G \, dx,$$

im \mathbb{R}^2 und analog mit dem Flächenelement do im Randintegral im dreidimensionalen Fall.

Ist H ein Potentialfeld, d.h. $H = \nabla f$ mit einer Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ und berücksichtigen wir den **Laplaceoperator** Δ , der durch

$$\Delta f(x) := \operatorname{div} \nabla f(x) = \sum_{j=1}^n \frac{\partial^2 f(x)}{\partial x_j^2}$$

gegeben ist, so liefert der Gauß'sche Integralsatz die Greenschen Sätze. Wir formulieren sie für $n = 2$. Für $n = 3$ ist im Randintegral nur $d\ell$ durch do zu ersetzen.

Satz 3.23 (Greensche Integralsätze)

Sei $D \subset \mathbb{R}^2$ mit Rand ∂D wie im Gauß'schen Integralsatz.

(a) Sei $g : D \rightarrow \mathbb{R}$ einmal und $f : D \rightarrow \mathbb{R}$ zweimal stetig differenzierbar in D , so dass g und ∇f stetig fortsetzbar auf \overline{D} seien. Dann gilt die **erste Greensche Identität**:

$$\iint_D [\nabla g(x) \cdot \nabla f(x) + g(x) \Delta f(x)] \, dx = \int_{\partial D} g(x) (\nabla f(x) \cdot \nu(x)) \, d\ell.$$

(b) Seien jetzt $f, g : D \rightarrow \mathbb{R}$ zweimal stetig differenzierbar in D und $\nabla f, \nabla g$ seien beide stetig fortsetzbar auf \overline{D} . Dann gilt die **zweite Greensche Identität**:

$$\iint_D [g(x) \Delta f(x) - f(x) \Delta g(x)] \, dx = \int_{\partial D} [g(x) (\nabla f(x) \cdot \nu(x)) - f(x) (\nabla g(x) \cdot \nu(x))] \, d\ell.$$

Für den Beweis von Teil (b) vertausche in Teil (a) die Rollen von f und g und subtrahiere die beiden Formeln.

Häufig wird die rechte Seite dieser Formeln etwas anders geschrieben. Dazu erinnern wir an den Begriff der Richtungsableitung: Ist $a \in \mathbb{R}^n$ ein Einheitsvektor, d.h. $\|a\|_2 = 1$, und $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine Funktion, so ist

$$\frac{\partial f(x)}{\partial a} = \lim_{\varepsilon \rightarrow 0} \frac{f(x + \varepsilon a) - f(x)}{\varepsilon} = a \cdot \nabla f(x)$$

die Steigung von f an der Stelle x **in Richtung** von a (falls der Grenzwert überhaupt existiert). Dies motiviert die folgende Definition:

Definition 3.24 (Normalableitung)

Sei $D \subseteq \mathbb{R}^n$ ($n = 2$ oder 3) und $\Gamma \subseteq \partial D$ ein reguläres Kurven- oder Flächenstück des Randes mit äußerem Normaleneinheitsvektor $\nu(x)$, $x \in \Gamma$. Die Funktion $f : D \rightarrow \mathbb{R}$ sei stetig differenzierbar und $\nabla f(x)$ sei stetig fortsetzbar auf \overline{D} . Dann heißt

$$\frac{\partial f(x)}{\partial \nu} := \nu(x) \cdot \nabla f(x), \quad x \in \Gamma,$$

die **äußere Normalableitung** von f im Punkt $x \in \Gamma$.

Damit erhalten die Greenschen Sätze die elegantere Darstellung:

$$\iint_D [\nabla g(x) \cdot \nabla f(x) + g(x) \Delta f(x)] dx = \int_{\partial D} g(x) \frac{\partial f(x)}{\partial \nu} d\ell$$

und

$$\iint_D [g(x) \Delta f(x) - f(x) \Delta g(x)] dx = \int_{\partial D} \left[g(x) \frac{\partial f(x)}{\partial \nu} - f(x) \frac{\partial g(x)}{\partial \nu} \right] d\ell.$$

Im Fall $D \subseteq \mathbb{R}^3$ gelten die Formeln entsprechend.

Bemerkungen: (a) Im Fall $n = 1$ entspricht dies den Formeln

$$\int_a^b [g'(x) f'(x) + g(x) f''(x)] dx = [g(x) f'(x)]_a^b,$$

$$\int_a^b [g(x) f''(x) - f(x) g''(x)] dx = [g(x) f'(x) - f(x) g'(x)]_a^b,$$

die beide durch partielle Integration nachgewiesen werden können.

(b) Beachte, dass die Differentialoperatoren in anderen Koordinatensystemen entsprechend andere Darstellungen besitzen. Betrachten wir etwa Polarkoordinaten, d.h. $(x, y) = (r \cos \varphi, r \sin \varphi)$ für $r > 0$ und $\varphi \in [0, 2\pi)$. Für eine Abbildung $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ haben wir

$$f(x, y) = f(r \cos \varphi, r \sin \varphi) =: \tilde{f}(r, \varphi),$$

also mit der Kettenregel:

$$\frac{\partial \tilde{f}}{\partial r} = \frac{\partial f}{\partial x} \cos \varphi + \frac{\partial f}{\partial y} \sin \varphi$$

$$\frac{\partial \tilde{f}}{\partial \varphi} = -r \sin \varphi \frac{\partial f}{\partial x} + r \cos \varphi \frac{\partial f}{\partial y},$$

d.h. vektoriell geschrieben:

$$\begin{pmatrix} \frac{\partial \tilde{f}}{\partial r} \\ \frac{\partial \tilde{f}}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -r \sin \varphi & r \cos \varphi \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix}.$$

Auflösen dieses linearen Gleichungssystems nach $\partial f / \partial x$ und $\partial f / \partial y$ liefert für den Gradienten

$$\nabla f(x, y) = \frac{1}{r} \begin{pmatrix} r \cos \varphi & -\sin \varphi \\ r \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} \frac{\partial \tilde{f}}{\partial r} \\ \frac{\partial \tilde{f}}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \cos \varphi \frac{\partial \tilde{f}}{\partial r} - \frac{1}{r} \sin \varphi \frac{\partial \tilde{f}}{\partial \varphi} \\ \sin \varphi \frac{\partial \tilde{f}}{\partial r} + \frac{1}{r} \cos \varphi \frac{\partial \tilde{f}}{\partial \varphi} \end{pmatrix}.$$

3.4 Die Rotation und der Stokessche Satz

Nachdem wir den Zusammenhang zwischen Divergenz und Fluss eines Vektorfeldes gesehen haben führen wir nun noch den Begriff der Zirkulation oder Rotation eines Vektorfeldes ein.

Definition 3.25 (*Rotation*)

Sei $D \subseteq \mathbb{R}^3$ offen und beschränkt und $F : D \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^3$ ein stetig differenzierbares Vektorfeld. Dann heißt

$$\operatorname{rot} F(x) := \begin{pmatrix} \frac{\partial F_3(x)}{\partial x_2} - \frac{\partial F_2(x)}{\partial x_3} \\ \frac{\partial F_1(x)}{\partial x_3} - \frac{\partial F_3(x)}{\partial x_1} \\ \frac{\partial F_2(x)}{\partial x_1} - \frac{\partial F_1(x)}{\partial x_2} \end{pmatrix} \in \mathbb{R}^3$$

die **Rotation** des Vektorfeldes F an der Stelle x .

Bemerkung: Oft werden Divergenz und Rotation in der Form

$$\operatorname{div} F(x) = \nabla \cdot F(x) \quad \text{und} \quad \operatorname{rot} F(x) = \nabla \times F(x)$$

geschrieben.

Im angelsächsischen steht $\operatorname{curl} F(x)$ für $\operatorname{rot} F(x)$. Ist $\operatorname{div} F = 0$ bzw. $\operatorname{rot} F = 0$ in D , so wird das Feld F **divergenzfrei** (oder quellenfrei) bzw. **rotationsfrei** (oder wirbelfrei) genannt.

Beispiele 3.26

(a) Ist $F(x) = a \in \mathbb{R}^3$ für alle $x \in \mathbb{R}^3$, also konstant, so ist F divergenz- und rotationsfrei. Für das Feld $F(x) = x$ ist $\operatorname{div} x = 3$ und $\operatorname{rot} x = 0$.

(b) Jedes Gradientenfeld ist rotationsfrei, d.h.

$$\operatorname{rot} \nabla f(x) = 0, \quad x \in D,$$

für jede zweimal differenzierbare Funktion $f : D \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$ (nachrechnen!).

(c) Jede Rotation ist divergenzfrei, d.h.

$$\operatorname{div} \operatorname{rot} F(x) = 0, \quad x \in D,$$

für jedes zweimal differenzierbare Vektorfeld $F : D \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^3$. Auch dies lässt sich direkt nachrechnen.

(d) Es gelten die folgenden Rechenregeln für Gradient, Divergenz und Rotation. Hierbei seien $U, V : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ Vektorfelder und $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}$ ein skalares Feld sowie $\lambda, \mu \in \mathbb{R}$ Konstanten.

$$\begin{aligned} \operatorname{div} (\lambda U + \mu V) &= \lambda \operatorname{div} U + \mu \operatorname{div} V, \\ \operatorname{div} (\varphi U) &= \nabla \varphi \cdot U + \varphi \operatorname{div} U, \\ \operatorname{rot} (\lambda U + \mu V) &= \lambda \operatorname{rot} U + \mu \operatorname{rot} V, \\ \operatorname{rot} (\varphi U) &= \nabla \varphi \times U + \varphi \operatorname{rot} U. \end{aligned}$$

(e) Das Feld

$$F(x) = \frac{1}{\|x\|_2^3} x, \quad x \neq 0,$$

ist sowohl divergenz- als auch rotationsfrei: erstens ist es ein Gradientenfeld mit $f(x) = -1/\|x\|_2$ (daher rotationsfrei) und ferner ist $\operatorname{div} F(x) = 0$, wie man direkt nachrechnet. F beschreibt (bis auf Konstanten) das Kraftfeld einer Punktladung im Ursprung.

Zur Formulierung des Satzes von Stokes brauchen wir einige Vorbereitungen. Wir gehen aus von einer glatten Fläche $S \subseteq \mathbb{R}^3$. Sie sei parametrisiert durch $X : \overline{B} \subseteq \mathbb{R}^2 \rightarrow S$ mit einer offenen, beschränkten Menge B . Durch Festlegung des Normaleneinheitsvektors zu

$$\nu_S = \frac{X_{u_1} \times X_{u_2}}{\|X_{u_1} \times X_{u_2}\|}$$

ist eine **Orientierung der Fläche** gegeben. Die Vektoren $[X_{u_1}, X_{u_2}, \nu_S]$ sind bekanntlich mathematisch positiv orientiert. Wir können uns die Orientierung mit Hilfe der ausgestreckten **rechten Hand** klarmachen: Zeigt der Daumen in Richtung X_{u_1} , und der Zeigefinger in Richtung X_{u_2} , so zeigt der Mittelfinger in Richtung des Normalenvektors ν_S . Wir stellen uns vor, dass ν_S in die Richtung zeigt, **aus der wir die Fläche betrachten**. (Ist etwa S ein Teil der (x_1, x_2) -Ebene und ist $\nu_S = (0, 0, 1)^\top$, so betrachten wir S von oben – also so, wie wir zweidimensionale Gebiete betrachten.)

Sei weiter $\varphi : [a, b] \subseteq \mathbb{R} \rightarrow \mathbb{R}^2$ eine stückweise glatte und **mathematisch positiv orientierte** Parametrisierung der Randkurve ∂B der Parametermenge B . Durchläuft man also den Rand mit aufsteigendem Parameter, so liegt die Menge B links. Anders ausgedrückt: Der Tangentenvektor $\dot{\varphi}(t)$ zeigt nach vorn, und $(\dot{\varphi}_2(t), -\dot{\varphi}_1(t))^\top$ ist ein nach außen zeigender Normalenvektor. Dann ist $\psi : [a, b] \subseteq \mathbb{R} \rightarrow \mathbb{R}^3$, definiert durch $\psi(t) = X(\varphi(t))$ für $t \in [a, b]$, eine Parametrisierung des Randes C von S mit Tangentialeinheitsvektor

$$\tau = \frac{\dot{\psi}}{\|\dot{\psi}\|}.$$

Durch diese Wahl von ν_S und τ hat man wieder eine **mathematisch positive** Parametrisierung der Fläche S mit Randkurve C definiert. **Bei konkreten Beispielen** geht man so vor: Man wählt zunächst den Normalenvektor ν_S , betrachtet die Fläche S aus der Richtung, in die ν_S zeigt, und durchläuft die Randkurve C so, dass die Fläche S links liegt. Dann ist das Paar S, C positiv orientiert. Wir schreiben wieder ∂S für die Randkurve C von S .

Jetzt können wir formulieren:

Satz 3.27 (*Integralsatz von Stokes*)

Sei $S \subseteq \mathbb{R}^3$ eine glatte Fläche mit Randkurve C , die aus endlich vielen stückweise glatten Kurven bestehe und nach obiger Regel positiv orientiert sei. Ferner sei $F : D \rightarrow \mathbb{R}^3$ stetig differenzierbar auf einer offenen Obermenge $D \subseteq \mathbb{R}^3$ von S . Dann gilt

$$\iint_S \operatorname{rot} F(x) \cdot \nu_S(x) \, d\sigma = \int_C F(x) \cdot \tau(x) \, dl = \int_C F(x) \cdot dl.$$

Beweis nur für Flächen S , die in der (x_1, x_2) -Ebene (als Ebene des \mathbb{R}^3) liegen, also $S = B \times \{0\}$ mit ebenem Gebiet $B \subseteq \mathbb{R}^2$. Wir wählen $\nu_S = (0, 0, 1)^\top$. Da $S, \partial S$ positiv orientiert sind, so müssen wir auch eine mathematisch positive Parametrisierung $\varphi : [a, b] \rightarrow \mathbb{R}^2$ des Randes ∂B wählen. Es ist

$\psi(t) = (\varphi_1(t), \varphi_2(t), 0)^\top$ die entsprechende Parametrisierung von ∂S . Dann zeigt $(\dot{\varphi}_2(t), -\dot{\varphi}_1(t))^\top$ nach außen. Wir definieren das zweidimensionale Vektorfeld

$$G(x_1, x_2) = (F_2(x_1, x_2, 0), -F_1(x_1, x_2, 0))^\top.$$

Dann ist $\operatorname{div} G(x_1, x_2) = \operatorname{rot} F(x_1, x_2, 0) \cdot \nu_S$. Die Anwendung des zweidimensionalen Gaußschen Integralsatz im ebenen Gebiet B liefert:

$$\begin{aligned} \iint_S \operatorname{rot} F(x) \cdot \nu_S(x) \, do &= \iint_B \operatorname{div} G(x_1, x_2) \, d(x_1, x_2) = \int_{\partial B} G(x) \cdot \nu(x) \, dl \\ &= \int_a^b [F_2(\varphi(t)) \dot{\varphi}_2(t) + F_1(\varphi(t)) \dot{\varphi}_1(t)] \, dt = \int_C F(x) \cdot \tau(x) \, dl. \end{aligned}$$

□

Bemerkung: In konkreten Beispielen besteht die Fläche S häufig aus mehreren „zusammengeklebten“ Teilen. Das Blech einer oben offenen Dose beschreibt z.B. eine Fläche S , die aus zwei Teilen besteht, dem Mantel und dem Boden. Diese sind am Rand des Bodens zusammengeheftet. Was ist der Rand dieser Fläche? Jede glatte Teilfläche S_j hat einen Rand, der oben vor Satz 3.27 definiert wurde (über die Abbildung des Randes des Parametergebiets B durch ψ). Im Beispiel der Dose hat der Mantel die beiden Kreise oben und unten als Rand, und der Rand des Bodens besteht aus dem unteren Kreis. Da die Fläche am unteren Kreis zusammengesetzt wird, verschwinden dort die Ränder, und es bleibt als Rand der obere Kreis übrig.

Der Satz von Stokes gilt auch für solche zusammengesetzten Flächen.

Wir wiederholen das Beispiel der Arbeit in einem Kraftfeld. Ist $F : D \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^3$ ein Kraftfeld und $C \subseteq D$ eine Kurve, parametrisiert durch $x : [a, b] \rightarrow D$, so beschreibt

$$A = \int_C F(x) \cdot dl := \int_C F(x) \cdot \tau(x) \, dl = \int_a^b F(x(t)) \cdot \dot{x}(t) \, dt$$

die Arbeit von $x(a)$ nach $x(b)$ längs C . Wir haben gesehen, dass die Arbeit nur vom Anfangs- und Endpunkt des Weges abhängt, wenn F **konservativ** ist, d.h. in der Form $F = \nabla f$ mit einer skalaren Funktion $f : D \rightarrow \mathbb{R}$ geschrieben werden kann. Solch eine Funktion f heißt (skalares) **Potential** von F . Mit den Integrabilitätsbedingungen bei exakten Differentialgleichungen haben wir schon notwendige Bedingungen (zumindest im \mathbb{R}^2) für die Existenz von f kennengelernt und gesehen wie f durch Integration gegebenenfalls berechnet werden kann (siehe Beispiel 1.27). Es bleibt aber noch die Frage, welche Vektorfelder F diese Eigenschaft besitzen.

Dabei spielen zusammenhängende und einfach zusammenhängende Mengen D eine wichtige Rolle.

Definition 3.28 (a) Eine Menge $D \subseteq \mathbb{R}^n$ heißt **zusammenhängend**, wenn es zu je zwei Punkten $y, z \in D$ eine glatte Kurve $x : [a, b] \rightarrow D$ in D gibt, die y und z verbindet (d.h. $x(a) = y$ und $x(b) = z$).

(b) Eine Menge $D \subseteq \mathbb{R}^n$ heißt **einfach zusammenhängend**, wenn sie zusammenhängend ist und noch folgende Eigenschaft besitzt: Jede glatte geschlossene, doppelpunktfreie Kurve in D lässt sich in D stetig auf einen Punkt in D „zusammenziehen“. (Genauer bedeutet dies, dass es zu jeder glatten Kurve $x : [a, b] \rightarrow D$ mit $x(a) = x(b)$ und $x(t) \neq x(s)$ für $t \neq s \in (a, b)$ eine stetige Funktion $g : [0, 1] \times [a, b] \rightarrow D$ und ein $z \in D$ gibt mit $g(s, a) = g(s, b)$ für $s \in [0, 1]$ und $g(0, t) = x(t)$ und $g(1, t) = z$ für $t \in [a, b]$.)

Wir stellen uns die Kurve als Gummiband vor, welches bei einfach zusammenhängenden Gebieten zu einem Punkt zusammengezogen werden kann. In der Ebene sind, grob gesprochen, einfach zusammenhängende Mengen solche ohne „Löcher“, im Raum solche Mengen ohne „Henkel“. Der Torus ist zwar zusammenhängend, aber nicht einfach zusammenhängend. Die Kugelschale ist einfach zusammenhängend. In der Ebene ist der Kreisring nicht einfach zusammenhängend.

Satz 3.29

Sei $D \subseteq \mathbb{R}^n$ offen und zusammenhängend, und $F : D \rightarrow \mathbb{R}^n$ sei stetig differenzierbar.

(a) Genau dann gibt es eine Funktion $f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ mit $F(x) = \nabla f(x)$, $x \in D$, wenn für jede Kurve C in D das Kurvenintegral

$$\int_C F(x) \cdot \tau(x) \, d\ell = \int_C F(x) \cdot dl = \int_C [F_1(x) \, dx_1 + \dots + F_n(x) \, dx_n]$$

nur vom Anfangs- und Endpunkt und damit von der Orientierung der Kurve, nicht aber von der konkreten Parametrisierung, abhängt. In diesem Fall ist ein Potential f gegeben durch (mit beliebigem, aber festem $z \in D$):

$$f(x) := \int_{C_{z,x}} F(y) \cdot \tau(y) \, d\ell = \int_{C_{z,x}} F(y) \cdot dl, \quad x \in D.$$

Hier ist $z \in D$ beliebig und $C_{z,x} \subseteq D$ irgendeine glatte Kurve in D , die den Punkt $z \in D$ mit $x \in D$ verbindet. Insbesondere ist $f(x)$ unabhängig davon, welche Verbindungskurve man wählt. Es gilt $f(z) = 0$. Mit dieser Forderung ist dieses f das einzige Potential mit $\nabla f = F$.

(b) Sei $D \subseteq \mathbb{R}^3$ sogar **einfach zusammenhängend**. Genau dann gibt es eine Funktion $f : D \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}$ mit $F(x) = \nabla f(x)$, $x \in D$, wenn F rotationsfrei ist.

Beweis: (a) Ist $F = \nabla f$, so ist bekanntlich

$$\begin{aligned} \int_C F(x) \cdot \tau(x) \, d\ell &= \int_C \nabla f(x) \cdot \tau(x) \, d\ell = \int_a^b \nabla f(x(t)) \cdot \dot{x}(t) \, dt = \int_a^b \frac{d}{dt} f(x(t)) \, dt \\ &= f(x(b)) - f(x(a)), \end{aligned}$$

und dies hängt nur vom Anfangspunkt $x(a)$ und Endpunkt $x(b)$ ab.

Umgekehrt hänge jedes Kurvenintegral $\int_C F(x) \cdot \tau(x) dl$ nur von Anfangs- und Endpunkt ab. Sei $z \in D$ festgehalten. Die Funktion

$$f(x) := \int_{C_{z,x}} F(y) \cdot \tau(y) dl$$

ist dann in ganz D erklärt. Hier bezeichne $C_{z,x}$ wieder eine Kurve in D , die z mit x verbindet. Wir rechnen die partielle Ableitung aus. Sei $e^{(j)}$ der j -te Einheitsvektor und sei $C_{z,x+he^{(j)}}$ die Kurve in D von z nach $x + he^{(j)}$, die auf $C_{z,x}$ nach x führt und weiter von x nach $x + he^{(j)}$ auf der Verbindungsgeraden. (Zeichnung!) Dann ist mit dem Mittelwertsatz der Integralrechnung (Satz 6.12 von HM I)

$$\begin{aligned} f(x + he^{(j)}) - f(x) &= \int_{C_{z,x+he^{(j)}}} F(y) \cdot \tau(y) dl - \int_{C_{z,x}} F(y) \cdot \tau(y) dl \\ &= \int_{C_{x,x+he^{(j)}}} F(y) \cdot \tau(y) dl = \int_0^h F_j(x + te^{(j)}) dt = h F_j(x + \rho e^{(j)}) \end{aligned}$$

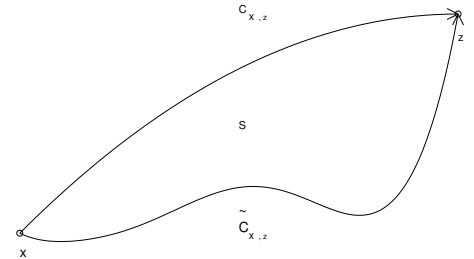
für ein $\rho \in [0, h]$. Division durch h und $h \rightarrow 0$ liefert $\partial f(x)/\partial x_j = F_j(x)$. Damit leistet f das Gewünschte.

(b) Ist $F = \nabla f$, so ist F rotationsfrei wegen $\text{rot } \nabla = 0$. Sei jetzt umgekehrt das Vektorfeld F im einfach zusammenhängenden Gebiet D rotationsfrei.

Nach Teil (a) reicht es zu zeigen, dass das Integral

$$\int_{C_{x,z}} F(y) \cdot \tau(y) dl$$

unabhängig davon ist, welche Kurve man nimmt, um x und z zu verbinden.



Sei daher $\tilde{C}_{x,z} \subseteq D$ eine zweite glatte Kurve, die x und z verbindet. Wir nehmen an, dass sich die beiden Kurven außer in x und z nicht schneiden oder berühren. (Sonst muss man die Argumentation von Schnittpunkt zu Schnittpunkt durchführen). Da D einfach zusammenhängend ist, kann man nun eine Fläche S finden, deren Rand aus $C := \tilde{C}_{x,z} \cup C_{x,z}$ besteht, und die ganz in D liegt. (Für den ebenen Fall siehe Abbildung). Der Stokes'sche Integralsatz liefert

$$\int_C F(y) \cdot \tau(y) dl = \iint_S \text{rot } F(y) \cdot \nu_S(y) do = 0.$$

Bei diesem Integral zeigt $\tau(y)$ nach vorn, wenn man die Kurve C von x startend längs $C_{x,z}$ nach z und dann längs der anderen Kurve $\tilde{C}_{x,z}$ wieder zurück nach x durchläuft. Im zweiten Teilstück durchläuft man also den Weg $\tilde{C}_{x,z}$ in umgekehrter Richtung von z nach x . Daher ist

$$0 = \int_C F(y) \cdot \tau(y) dl = \int_{C_{x,z}} F(y) \cdot \tau(y) dl - \int_{\tilde{C}_{x,z}} F(y) \cdot \tau(y) dl,$$

und die beiden Integrale auf der rechten Seite sind gleich. Dies beendet den Beweis. \square

Die Bedingung der Rotationsfreiheit ist ein einfaches Mittel, um zu bestimmen, ob ein Feld konservativ ist oder nicht – falls das Gebiet einfach zusammenhängend ist. Wir geben jetzt ein Beispiel für ein rotationsfreies Feld, welches nicht konservativ ist.

Beispiel 3.30

(a) Auf Seite 18 haben wir festgehalten, dass das Integrabilitätskriterium bei exakten Differentialgleichungen auch hinreichend ist. Dies sehen wir aus dem obigen Satz, wenn wir $D = I \times \mathbb{R} \times \mathbb{R}$ und $F : D \rightarrow \mathbb{R}^3$ mit $F(x_1, x_2, x_3) = (p(x_1, x_2), q(x_1, x_2), 0)^\top$ setzen. Denn dann gilt

$$\operatorname{rot} F(x) = \begin{pmatrix} 0 \\ 0 \\ \frac{\partial q}{\partial x_1} - \frac{\partial p}{\partial x_2} \end{pmatrix}.$$

Mit dem Integrabilitätskriterium folgt $\operatorname{rot} F = 0$. Also gibt es nach Satz 3.29 ein Potential, da das Gebiet D einfach zusammenhängend ist.

Lässt sich die exakte Differentialgleichung nur auf einer Teilmenge $I \times M$ mit $M \subseteq \mathbb{R}$ betrachten, bleibt das Kriterium erhalten, solange $I \times M \times \mathbb{R}$ einfach zusammenhängend ist. Diese Bedingung an M ist notwendig, wie das nächste Beispiel zeigt.

(b) Sei $D = \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 \neq 0\}$ der Raum \mathbb{R}^3 ohne die x_3 -Achse. Das Vektorfeld $F : D \subseteq \mathbb{R}^3 \rightarrow \mathbb{R}^3$ sei gegeben durch

$$F(x) = \left(\frac{-x_2}{x_1^2 + x_2^2}, \frac{x_1}{x_1^2 + x_2^2}, 0 \right)^\top, \quad x_1^2 + x_2^2 \neq 0.$$

Wir rechnen aus, dass $\operatorname{rot} F(x) = 0$ für alle $x \in D$ (Übung!). Sei nun $z = (1, 0, 0)^\top$ und $C_{z,z}$ die folgende Kurve mit Anfangs- und Endpunkt z , parametrisiert durch

$$x(t) = (\cos t, \sin t, 0)^\top, \quad 0 \leq t \leq 2\pi.$$

Dann ist $\tau(t) = (-\sin t, \cos t, 0)^\top$ und $d\ell = dt$, und daher

$$\int_{C_{z,z}} F(y) \cdot \tau(y) d\ell = \int_0^{2\pi} [\sin^2 t + \cos^2 t] dt = 2\pi \neq 0.$$

Daher kann das Vektorfeld **kein** Gradientenfeld sein, denn sonst müsste dieses Integral ja verschwinden. Also ist F nicht konservativ. Dieses Beispiel widerspricht nicht dem Satz 3.29, denn das Gebiet D ist nicht einfach zusammenhängend!

Bemerkung: Auch im zweidimensionalen Fall lässt sich eine Rotation sinnvoll definieren, wie wir es schon in einem Beispiel angedeutet haben: Sei $D \subseteq \mathbb{R}^2$ eine offene Menge in der Ebene und $F : D \rightarrow \mathbb{R}^2$ ein zweidimensionales Vektorfeld. Wir ergänzen F durch 0 zu einem dreidimensionalen Vektorfeld:

$$\tilde{F}(x_1, x_2, x_3) := \begin{pmatrix} F_1(x_1, x_2) \\ F_2(x_1, x_2) \\ 0 \end{pmatrix}, \quad (x_1, x_2, x_3) \in D \times \mathbb{R}.$$

Dann folgt

$$\operatorname{rot} \tilde{F}(x) = \begin{pmatrix} 0 \\ 0 \\ \frac{\partial F_2(x)}{\partial x_1} - \frac{\partial F_1(x)}{\partial x_2} \end{pmatrix}, \quad x \in D \times \mathbb{R}.$$

Also wird für zweidimensionale Vektorfelder $F : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^2$ die Rotation definiert durch

$$\operatorname{rot} F(x) := \frac{\partial F_2(x)}{\partial x_1} - \frac{\partial F_1(x)}{\partial x_2}, \quad x \in D.$$

Beachte, dass dies im Gegensatz zum dreidimensionalen Fall ein **Skalar** ist! Damit gilt Satz 3.29 ganz analog für zweidimensionale Vektorfelder $F : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}^2$ und einfach zusammenhängende Gebiete $D \subseteq \mathbb{R}^2$.

4 Partielle Differentialgleichungen

4.1 Klassifizierungen

Eine Vielzahl grundlegender Modelle in Naturwissenschaft und Technik werden durch partielle Differentialgleichungen beschrieben.

Definition 4.1

Sei $D \subseteq \mathbb{R}^n$ eine offene Teilmenge und $k \in \mathbb{N}$. Eine **partielle Differentialgleichung k -ter Ordnung** für eine Funktion $u : D \rightarrow \mathbb{R}$ (oder für $u : D \rightarrow \mathbb{C}$) ist eine Gleichung der Form

$$F \left(x, u(x), \frac{\partial u(x)}{\partial x_1}, \dots, \frac{\partial u(x)}{\partial x_n}, \frac{\partial^2 u(x)}{\partial x_1^2}, \frac{\partial^2 u(x)}{\partial x_1 \partial x_2}, \dots, \frac{\partial^k u(x)}{\partial x_n^k} \right) = g(x)$$

mit einer Funktion $F : D \times \mathbb{R} \times \mathbb{R}^n \times \dots \times \mathbb{R}^{n^k} \rightarrow \mathbb{R}^m$ und $g : D \rightarrow \mathbb{R}^m$. Die partielle Differentialgleichung heißt **linear**, wenn die Funktion F linear in allen Variablen bis auf die erste ist, d.h. wenn für differenzierbare Funktionen $u, v : D \rightarrow \mathbb{R}$ und Faktoren $\lambda, \mu \in \mathbb{R}$ gilt

$$\begin{aligned} & F \left(x, (\lambda u + \mu v), \frac{\partial(\lambda u + \mu v)}{\partial x_1}, \dots, \frac{\partial^k(\lambda u + \mu v)}{\partial x_n^k} \right) \\ &= \lambda F \left(x, u, \frac{\partial u}{\partial x_1}, \dots, \frac{\partial^k u}{\partial x_n^k} \right) + \mu F \left(x, v, \frac{\partial v}{\partial x_1}, \dots, \frac{\partial^k v}{\partial x_n^k} \right). \end{aligned}$$

Die Variablen werden natürlich nicht immer nur mit x_1, x_2, \dots, x_n bezeichnet sondern auch anders, z.B. mit t (wenn es die Zeit ist), und x, y .

Beispiele 4.2 (a) Häufig ergeben sich partielle Differentialgleichungen aus physikalischen Erhaltungsregeln. Ist etwa durch $u(x, t)$ eine Massendichte oder Teilchendichte am Ort $x \in \mathbb{R}^n$ zum Zeitpunkt t beschrieben, d.h. in einem Gebiet $G \subseteq \mathbb{R}^n$ ist die Gesamtmenge durch $\int_G u(x, t) dx$ gegeben, dann liefert Massenerhaltung, dass die zeitliche Änderung der Gesamtmenge im Gebiet G sich zusammensetzt aus den externen Quellen bzw. Senken, die wir durch eine Funktion $f(x, t)$ beschreiben, und aus dem Fluss aus dem Gebiet, den wir allgemein mit dem Flussintegral

$$\int_{\partial G} \Phi \cdot n ds$$

zu einer Funktion $\Phi(x, t, u)$ angeben können. Dann gilt mit dem Gaußschen Satz zum Zeitpunkt t ,

$$\int_G \frac{\partial u}{\partial t} + \operatorname{div}(\Phi) dx = \int_G f dx.$$

Im Fall $f = 0$ und einem Fluß proportional zur Dichte mit konstanter Fließrichtung $a \in \mathbb{R}^n$, d.h. $\Phi = u a$, folgt

$$\int_G \left[\frac{\partial u}{\partial t} + \sum_{j=1}^n a_j \frac{\partial u}{\partial x_j} \right] dx = 0.$$

Da diese Identität für jedes Gebiet G gilt, ergibt sich das klassische Beispiel einer linearen partiellen Differentialgleichung erster Ordnung, die **Transportgleichung**

$$\frac{\partial u(x, t)}{\partial t} + a \cdot \nabla_x u(x, t) = \frac{\partial u(x, t)}{\partial t} + \sum_{j=1}^n a_j \frac{\partial u(x, t)}{\partial x_j} = 0$$

für $u : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$, deren $n+1$ Argumente wir mit $x = (x_1, \dots, x_n)^\top$ (Ort) und t (Zeit) bezeichnen.

(b) Ist die Strömung durch $\Phi = -d \nabla_x u$ mit *Diffusionskonstante* $d \in \mathbb{R}$ gegeben, so erhalten wir analog die **Wärmeleitungsgleichung** oder **Diffusionsgleichung**

$$\frac{\partial u(x, t)}{\partial t} - d \Delta_x u(x, t) = 0.$$

Der Index bei ∇_x bzw. Δ_x soll andeuten, dass sich in diesen Fällen der Laplace-Operator nur auf die Ortsvariablen bezieht. Häufig wird diese Klarstellung weggelassen.

(c) Weitere Beispiele aus der wichtigen Klasse der linearen partiellen Differentialgleichungen zweiter Ordnung sind die **Potentialgleichung** oder **Laplace-Gleichung**

$$\Delta u := \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \dots + \frac{\partial^2 u}{\partial x_n^2} = 0$$

und die **Helmholtzgleichung** oder **Schwingungsgleichung**

$$\Delta u + k^2 u = 0.$$

Im zeitabhängigen Fall ist auch noch die **Wellengleichung**

$$\frac{\partial^2 u(x, t)}{\partial t^2} - \Delta_x u(x, t) = 0$$

ein zentrales Beispiel. Dabei haben wir auftretende physikalische Konstanten auf 1 normiert.

(d) Auch Systeme von partiellen Differentialgleichungen treten auf. So genügt zum Beispiel der Verschiebungsvektor $u : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ im stationären Fall der linearen Elastizitätstheorie den Lamé-Gleichungen

$$\mu \Delta u + (\lambda + \mu) \nabla(\operatorname{div} u) = 0,$$

mit Koeffizienten λ und μ (den Lamé Konstanten). Bei dieser Notation ist der Laplace-Operator Δ komponentenweise zu verstehen, d.h. $\Delta u = (\Delta u_1, \Delta u_2, \Delta u_3)^\top$.

(e) Systeme **nichtlinearer** partieller Differentialgleichungen sind ebenfalls von Bedeutung. Ein wichtiges Beispiel sind die **Navier-Stokes Gleichungen**

$$\frac{\partial u(x, t)}{\partial t} + (u(x, t) \cdot \nabla) u(x, t) - \Delta u(x, t) = -\nabla p(x, t) \quad \text{und} \quad \operatorname{div} u = 0,$$

die eine inkompressible, viskose Strömung $u : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$ beschreiben. $p = p(x, t)$ bezeichnet den Druck zum Zeitpunkt t am Ort x . Auch hier beziehen sich die Differentialoperatoren ∇ , Δ und div auf die Ortsabhängigkeit. Der Ausdruck $(u(x, t) \cdot \nabla) u(x, t)$ soll bedeuten:

$$(u(x, t) \cdot \nabla) u(x, t) = \sum_{j=1}^3 u_j(x, t) \frac{\partial u(x, t)}{\partial x_j}.$$

Die linearen partiellen Differentialgleichungen zweiter Ordnung spielen eine besonders wichtige Rolle. Wir beschränken uns auf **skalare** Differentialgleichungen. Allgemein werden drei wichtige Typen unterschieden. Dazu müssen diese Differentialgleichungen zunächst in der folgenden Form angegeben werden:

$$\sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 u(x)}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial u(x)}{\partial x_i} + c(x) u(x) = g(x) \quad (4.4)$$

mit einer symmetrischen Matrix $A(x) = A^\top(x) = (a_{ij}(x))_{i,j=1}^n \in \mathbb{R}^{n \times n}$ und Funktionen $b_i, c, g : \mathbb{R}^n \rightarrow \mathbb{R}$. Der Anteil $\sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 u}{\partial x_i \partial x_j}$ wird der **Hauptteil** der Differentialgleichung genannt und definiert die Typeneinteilung. Um dies korrekt zu tun, müssen wir an den Begriff des Eigenwerts für Matrizen $A \in \mathbb{R}^{n \times n}$ erinnern:

Eine Zahl $\lambda \in \mathbb{R}$ heißt **Eigenwert** zu A mit zugehörigem **Eigenvektor** x , wenn $x \neq 0$ und $Ax = \lambda x$ gilt.

Schreiben wir $Ax = \lambda x$ um in $(A - \lambda I)x = 0$ mit Einheitsmatrix I , so hat dieses quadratische Gleichungssystem genau dann nichttriviale Lösungen $x \in \mathbb{R}^n$, wenn $\det(A - \lambda I) = 0$ ist. Daher gilt: *Eine Zahl $\lambda \in \mathbb{R}$ ist genau dann Eigenwert von A , wenn λ Nullstelle des **charakteristischen Polynoms** $p(\lambda) = \det(A - \lambda I)$ ist.*

Man kann zeigen, dass jede *symmetrische* Matrix $A \in \mathbb{R}^{n \times n}$ genau n reelle Eigenwerte besitzt, wenn man sie gemäß ihrer Vielfachheiten zählt.

Definition 4.3 Die lineare partielle Differentialgleichung zweiter Ordnung (4.4) heißt 1. **elliptisch**, 2. **parabolisch** oder 3. **hyperbolisch** an der Stelle $x \in \mathbb{R}^n$, wenn die symmetrische Matrix $A(x)$

1. nur positive oder nur negative Eigenwerte hat. (Die Matrix $A(x)$ heißt dann positiv bzw. negativ definit.)
2. die Null als einfachen Eigenwert hat und alle anderen Eigenwerte entweder alle positiv oder alle negativ sind.
3. genau einen negativen einfachen Eigenwert hat und sonst nur positive (oder umgekehrt).

Damit ist die Laplace-Gleichung elliptisch, die Diffusionsgleichung parabolisch und die Wellengleichung hyperbolisch.

Beispiele 4.4 Bei partiellen Differentialgleichungen zu Funktionen $u : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}$ in Ort $x \in \mathbb{R}^n$ und Zeit $t \in \mathbb{R}$ sind in Anwendungen insbesondere ebene Wellen von Interesse, d.h. Lösungen in der Form

$$u(x, t) = f(d \cdot x - \sigma t),$$

mit $f : \mathbb{C} \rightarrow \mathbb{C}$, $d \in \mathbb{R}^n$ und $\sigma \in \mathbb{C}$. Dabei ist das Profil der Welle durch f gegeben, die Wellenfront der ebenen Welle steht senkrecht auf d , und für $\sigma \in \mathbb{R}$ ergibt sich die Ausbreitungsgeschwindigkeit als $\frac{\sigma}{\|d\|}$. Wir betrachten speziell das Profil $f(s) = e^{is}$ und berechnen

$$\begin{aligned} \Delta u(x, t) &= -\|d\|^2 u(x, t) \\ \frac{\partial u}{\partial t} &= -i\sigma u(x, t) \\ \frac{\partial^2 u}{\partial t^2} &= -\sigma^2 u(x, t). \end{aligned}$$

1. Fall: Im stationären Fall, d.h. u hängt nicht von der Zeit ab, erhalten wir mit $\sigma = 0$, dass

$$u(x) = e^{id \cdot x}$$

Lösung einer elliptischen Differentialgleichung, der **Helmholtzgleichung**

$$\Delta u + \|d\|^2 u = 0,$$

ist.

2. Fall: Mit $\sigma = -i \|d\|^2$ ist

$$u(x, t) = e^{id \cdot x} e^{-\|d\|^2 t}$$

Lösung der Diffusionsgleichung

$$\frac{\partial u}{\partial t} - \Delta_x u = 0.$$

Man beachte, dass diese Lösung eine exponentielle Dämpfung für $t \rightarrow \infty$ aufweist. Physikalisch beschreibt dieses Verhalten **Dissipation**, d.h. Energie wird dem System im Lauf der Zeit in Form von thermischer Energie entzogen (etwa Reibung). Eine typische Eigenschaft bei Lösungen parabolischer Gleichungen.

3. Fall: Setzen wir $\sigma = \pm \|d\|$, so ist

$$u(x, t) = e^{i(d \cdot x \pm \|d\|t)}$$

eine Lösung der Wellengleichung

$$\frac{\partial^2 u}{\partial t^2} - \Delta_x u = 0.$$

Diese Lösungen sind nicht dissipativ. An jeder Stelle $x \in \mathbb{R}^n$ erhalten wir eine harmonische Schwingung in der Zeit mit Amplitude $|u(x, t)| = 1$ und die Ausbreitungsgeschwindigkeit $\sigma / \|d\| = \pm 1$ der ebenen Welle ist konstant.

4. Fall: Für $\sigma = \|d\|^2$ ergibt sich

$$u(x, t) = e^{i(d \cdot x - \|d\|^2 t)},$$

eine Lösung der für die Quantenmechanik grundlegenden **Schrödinger Gleichung**,

$$i \frac{\partial u}{\partial t} + \Delta u = 0.$$

Analog zur Wellengleichung ist die Lösung nicht dissipativ. Aber im Gegensatz zum dritten Fall ist die Welle hier **dispersiv**, d.h. die Ausbreitungsgeschwindigkeit $\sigma / \|d\|$ ist nicht konstant, sondern ändert sich in Abhängigkeit der Frequenz σ .

Analog zu den gewöhnlichen Differentialgleichungen können wir nur dann eine eindeutige Lösung einer partiellen Differentialgleichung erwarten, wenn wir zusätzlich Anfangs- und/oder Randbedingungen oder auch Abklingbedingungen festlegen. Dann stellen sich stets die folgenden Fragen:

- Gibt es eine Lösung (Existenz)?
- Wie viele Lösungen existieren (Eindeutigkeit)?
- Wie hängen Lösungen von den sie bestimmenden Größen ab (Stabilität)?

Ein Beispiel bei dem wir die Fragen positiv beantworten können ist der Existenz- und Eindeutigkeitssatz für Anfangswertprobleme zur räumlich eindimensionalen Wellengleichung, der nach dem französischen Mathematiker d'Alembert benannt ist.

Satz 4.5 Sei $f \in C^2(\mathbb{R})$ und $g \in C^1(\mathbb{R})$. Dann besitzt das **Cauchy-Problem**

$$\begin{aligned}\frac{\partial^2 u}{\partial t^2} - a^2 \frac{\partial^2 u}{\partial x^2} &= 0 \quad \text{in } \mathbb{R} \times (0, \infty), \\ u(x, 0) &= f(x) \quad \text{für } x \in \mathbb{R}, \\ \frac{\partial u(x, 0)}{\partial t} &= a g(x) \quad \text{für } x \in \mathbb{R},\end{aligned}$$

genau eine Lösung. Diese ist gegeben durch

$$u(x, t) = \frac{1}{2} \left(f(x + at) + f(x - at) \right) + \frac{1}{2} \int_{x-at}^{x+at} g(z) dz.$$

Beweis: Sei u eine Lösung der Differentialgleichung. Durch die Variablentransformation $\xi = x + at$ und $\eta = x - at$ mit $x = \frac{1}{2}(\xi + \eta)$ und $t = \frac{1}{2a}(\xi - \eta)$ und der Definition

$$u(x, t) = u \left(\frac{\xi + \eta}{2}, \frac{\xi - \eta}{2a} \right) =: v(\xi, \eta)$$

berechnen wir mit der Kettenregel

$$\begin{aligned}u_x &= v_\xi \frac{\partial \xi}{\partial x} + v_\eta \frac{\partial \eta}{\partial x} = v_\xi + v_\eta \\ u_{xx} &= v_{\xi\xi} + 2v_{\xi\eta} + v_{\eta\eta} \\ u_t &= a(v_\xi - v_\eta) \\ u_{tt} &= a^2(v_{\xi\xi} - 2v_{\xi\eta} + v_{\eta\eta}).\end{aligned}$$

Bei dieser Auflistung haben wir eine übliche abkürzende Notation für partielle Ableitungen genutzt. Es ist zum Beispiel $u_{xx} = \partial^2 u / \partial x^2$.

Wenn $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ nun Lösung der Wellengleichung $u_{tt} = a^2 u_{xx}$ ist, erhalten wir für v die partielle Differentialgleichung

$$v_{\xi\eta}(\xi, \eta) = 0.$$

Also ist

$$v_\xi(\xi, \eta) = v_\xi(\xi, 0)$$

konstant bzgl. η . Integrieren führt auf

$$v(\xi, \eta) = v(0, \eta) + \int_0^\xi v_\xi(\tilde{\xi}, \eta) d\tilde{\xi} = v(0, \eta) + v(\xi, 0) - v(0, 0).$$

Somit gibt es eine Zerlegung $v(\xi, \eta) = \varphi(\xi) + \psi(\eta)$ mit noch unbekanntenen Funktionen $\varphi, \psi \in C^2(\mathbb{R})$ bzw. in den Koordinaten x und t

$$u(x, t) = \varphi(x + at) + \psi(x - at). \quad (4.5)$$

Aus den Anfangsbedingungen

$$u(x, 0) = f(x) \quad \text{und} \quad \frac{\partial u(x, 0)}{\partial t} = a g(x)$$

folgt $f(x) = \varphi(x) + \psi(x)$ und $g(x) = \varphi'(x) - \psi'(x)$ bzw.

$$\varphi(x) - \psi(x) = \int_0^x g(z) dz + \varphi(0) - \psi(0).$$

Auflösen der Gleichungen nach φ und ψ führt auf die explizite Darstellung

$$u(x, t) = \frac{1}{2} \left(f(x + at) + f(x - at) \right) + \frac{1}{2} \int_0^{x+at} g(z) dz - \frac{1}{2} \int_0^{x-at} g(z) dz.$$

Umgekehrt rechnet man direkt nach, dass die so definierte Funktion auch eine Lösung des Cauchy Problems ist. \square

Ebenso wichtig, wie die drei theoretischen Fragen zu beantworten, ist es auch, numerische Verfahren zu entwickeln, durch die sich Lösungen zu Rand- bzw. Anfangswertproblemen mit Hilfe des Rechners approximieren lassen. Dieser moderne Forschungszweig der Mathematik beinhaltet Stichworte wie **Finite Differenzen**, **Randelemente** oder **Finite Elemente**. Jede dieser Fragen muss zu verschiedenen Typen von partiellen Differentialgleichungen mit unterschiedlichsten Techniken angegangen werden und öffnet das Tor zu aktueller Forschung. Deswegen können wir in dieser knappen Einführung nur einige wenige grundlegende analytische Zugänge zu speziellen partiellen Differentialgleichungen erarbeiten. Für weitergehende Betrachtungen sei auf Spezialvorlesungen und die umfangreiche Literatur hingewiesen.

4.2 Quasilineare partielle Differentialgleichungen erster Ordnung

Zunächst starten wir mit der einfachsten partiellen Differentialgleichung, der eindimensionalen **Transportgleichung** mit konstanten Koeffizienten.

Beispiel: Für $a \in \mathbb{R}$ ist eine Lösung $u : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ der homogenen Differentialgleichung

$$\frac{\partial u(x, t)}{\partial t} + a \frac{\partial u(x, t)}{\partial x} = 0, \quad x \in \mathbb{R}, \quad t \in \mathbb{R},$$

gesucht, wobei wir annehmen, dass eine Anfangsbedingung der Form

$$u(x, 0) = g(x), \quad x \in \mathbb{R},$$

mit $g : \mathbb{R} \rightarrow \mathbb{R}$ gegeben ist. Nun beobachten wir, dass die linke Seite der Differentialgleichung die Richtungsableitung von u in Richtung $(a, 1)^\top \in \mathbb{R}^2$ ist. Wir definieren daher die Funktion $w : \mathbb{R} \rightarrow \mathbb{R}$ durch $w(s) = u(x + sa, t + s)$ und erhalten mit der Kettenregel und der Differentialgleichung

$$w'(s) = \frac{\partial u(x + sa, t + s)}{\partial t} + a \frac{\partial u(x + sa, t + s)}{\partial x} = 0.$$

Also ist w konstant. Mit dem Anfangswert können wir somit die Lösung der partiellen Differentialgleichung für alle $(x, t) \in \mathbb{R} \times \mathbb{R}$ angeben:

$$u(x, t) = w(0) = w(-t) = u(x - ta, 0) = g(x - ta).$$

Dieses Beispiel liefert eine generelle Idee, um lineare partielle Differentialgleichungen erster Ordnung auf Systeme von gewöhnlichen Differentialgleichungen zurückzuführen. Wir gehen aus von der **allgemeinen quasi-linearen Differentialgleichung erster Ordnung**

$$a(x, u(x)) \cdot \nabla u(x) + b(x, u(x)) = 0, \quad x = (x_1, x_2) \in \mathbb{R}^2,$$

mit einem Vektorfeld $a : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ und einer Funktion $b : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$. Beachte, dass bei dieser Formulierung keine der Variablen eine ausgezeichnete Rolle spielt, wie die Variable t im obigen Beispiel. Die Differentialgleichung heißt **quasilinear**, da der Ausdruck linear von den höchsten Ableitungen, also hier von $\frac{\partial u}{\partial x_j}$, $j = 1, 2$, abhängt. Die Anfangsbedingung für eine quasilineare Differentialgleichung 1. Ordnung hat die Form $u(x) = g(x)$ für $x \in \Gamma$, wobei $\Gamma \subset \mathbb{R}^2$ eine Kurve ist, parametrisiert durch $x = \gamma(t) \in \mathbb{R}^2$, $t \in [a, b] \subset \mathbb{R}$.

Der Graph der Lösung u beschreibt eine Fläche im \mathbb{R}^3 , parametrisiert durch

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \\ x_2 \\ u(x_1, x_2) \end{pmatrix}.$$

Dieser Graph muss die **Raum-Anfangskurve** $t \mapsto (\gamma_1(t), \gamma_2(t), g(\gamma(t)))^\top$ enthalten. Die **Charakteristikenmethode** liefert die Fläche zunächst in einer anderen Parameterdarstellung. Am Ende kann man diese dann auf die soeben angegebene transformieren.

Wir verfolgen die (differenzierbare) Funktion $u : \mathbb{R}^2 \rightarrow \mathbb{R}$ entlang einer differenzierbaren Kurve $k : I \subseteq \mathbb{R} \rightarrow \mathbb{R}^2$, d.h. wir betrachten die eindimensionale Funktion

$$w(s) = u(k(s)).$$

Dann ergibt sich mit der Kettenregel die Ableitung

$$w'(s) = k'(s) \cdot \nabla u(k(s)).$$

Die Idee ist nun, Kurven zu bestimmen, die als Tangentialfeld das Vektorfeld a aufweisen, d.h. gesucht ist eine Lösung zu

$$k'(s) = a(k(s), w(s)). \quad (4.6)$$

Denn dann ergibt sich für w mit der partiellen Differentialgleichung eine weitere gewöhnliche Differentialgleichung erster Ordnung, nämlich

$$w'(s) = -b(k(s), w(s)) \quad (4.7)$$

Mit den Methoden, die wir im zweiten Semesters kennengelernt haben, lassen sich gegebenenfalls Lösungen des Systems (4.6) und (4.7) von Differentialgleichungen bestimmen. Wird nun durch eine Anfangsbedingung noch der Wert

$$w(0) = u(k(0))$$

festgelegt, so ist auch die Lösung u der partiellen Differentialgleichung entlang einer solchen Kurve bestimmt. Wenn weiterhin solche Anfangswerte zu einer den \mathbb{R}^2 überdeckenden Schar von Kurven gegeben sind, so lässt sich auf diesem Weg die Lösung u für jeden Punkt $x \in \mathbb{R}^2$ bestimmen.

Definition 4.6

Sei $D \subset \mathbb{R}^2$ offen und $a : D \times \mathbb{R} \rightarrow \mathbb{R}^2$, $b : D \times \mathbb{R} \rightarrow \mathbb{R}$ gegeben. Dann heißt das System der drei gewöhnlichen Differentialgleichungen

$$k'(s) = a(k(s), w(s)), \quad s \in \mathbb{R}, \quad (4.8a)$$

$$w'(s) = -b(k(s), w(s)), \quad s \in \mathbb{R}, \quad (4.8b)$$

in den drei unbekanntenen Funktionen k_1 , k_2 und w **charakteristisches System** zur quasilinearen partiellen Differentialgleichung erster Ordnung

$$a(x, u(x)) \cdot \nabla u(x) + b(x, u(x)) = 0.$$

Lösungen (k_1, k_2, w) des Systems werden **Charakteristiken** der Differentialgleichung genannt.

Wir bemerken, dass das System (4.8a), (4.8b) ein **autonomes System** ist. Dies bedeutet, dass mit $(k(s), w(s))$ auch $(k(s + \tau), w(s + \tau))$ eine Lösung ist für jedes $\tau \in \mathbb{R}$.

Wenn die Funktionen a und b „gutartig“ sind, besitzt das System (4.8a), (4.8b) zu jedem Anfangswert $(\hat{x}_1, \hat{x}_2, \hat{u})^\top \in \mathbb{R}^3$ genau eine Lösung, also eine Raumkurve

$$s \mapsto \begin{pmatrix} k_1(s) \\ k_2(s) \\ w(s) \end{pmatrix}$$

mit $(k_1(0), k_2(0), w(0)) = (\hat{x}_1, \hat{x}_2, \hat{u})$. Wir bezeichnen mit s den Parameter dieser Raumkurve – im Gegensatz zu t , der die Anfangskurve parametrisiert.

Halte jetzt t fest (also den Punkt $(\gamma_1(t), \gamma_2(t), g(\gamma(t)))$ auf der Raum-Anfangskurve) und bestimme die Lösung k_1 , k_2 und w des charakteristischen Systems mit

$$\begin{pmatrix} k_1(0) \\ k_2(0) \\ w(0) \end{pmatrix} = \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \\ g(\gamma(t)) \end{pmatrix}.$$

Damit hängen also k und w von t und s ab, wir schreiben $k = k(s, t)$, $w = w(s, t)$, und es ist natürlich wegen der Definition $w(s, t) = u(k(s, t))$, d.h. $w = u \circ k$. Damit haben wir u im (s, t) -Koordinatensystem berechnet. Falls die Abbildung $k : (s, t) \mapsto (x_1, x_2) = k(s, t)$ invertierbar ist, so erhalten wir u im (x_1, x_2) -Koordinatensystem durch $u = w \circ k^{-1}$.

Bemerkung: Beachte, dass das charakteristische System im allgemeinen nichtlinear ist, auch wenn die partielle Differentialgleichung linear ist. Dieser Umstand erfordert insbesondere für eine Existenztheorie zu solchen Anfangswertproblemen weitere Überlegungen, da die allgemeine Theorie zu nichtlinearen Systemen nur lokale Aussagen machen kann. Weiter fällt auf, dass wir zu einer partiellen Differentialgleichung erster Ordnung nicht auf beliebigen Kurven beliebige Anfangswerte vorgeben können, da das Verhalten von Lösungen längs der Charakteristiken durch die Differentialgleichung vorgegeben ist.

Beispiele 4.7

(a) Wir betrachten die eindimensionale **Kontinuitätsgleichung**, wobei x_1 den Ort und x_2 die

Zeit bedeute.⁸ Gesucht ist also die Lösung der partiellen Differentialgleichung

$$\frac{\partial u}{\partial x_2}(x) + \frac{\partial}{\partial x_1}(u(x)v(x)) = 0 \quad \text{für } x_1, x_2 > 0$$

mit der Anfangsbedingung

$$u(x_1, 0) = g(x_1) \quad \text{für } x_1 > 0,$$

wobei $v : \mathbb{R}^2 \rightarrow \mathbb{R}$, $g : \mathbb{R} \rightarrow \mathbb{R}$ gegebene Funktionen sind. Wir differenzieren aus und erhalten

$$\frac{\partial u}{\partial x_2}(x) + v(x) \frac{\partial u}{\partial x_1}(x) + \alpha(x) u(x) = 0 \quad \text{für } x_1, x_2 > 0,$$

wobei wir zur Abkürzung $\alpha = \partial v / \partial x_1$ gesetzt haben. Damit haben wir die Standardform $a \cdot \nabla u + b = 0$ mit $a = \begin{pmatrix} v \\ 1 \end{pmatrix}$ und $b = \alpha u$. Die Anfangskurve Γ ist die x_1 -Achse, parametrisiert durch $\gamma(t) = (t, 0)^\top$, $t \in \mathbb{R}$. Das zugehörige charakteristische System ist

$$\begin{aligned} k_1'(s) &= v(k(s)) \\ k_2'(s) &= 1 \\ w'(s) &= -\alpha(k(s)) w(s). \end{aligned}$$

Wir betrachten jetzt das spezielle Beispiel $v(x) = x_1^2$. Dann ist $\alpha(x) = 2x_1$ und wir erhalten als charakteristisches System

$$\begin{aligned} k_1'(s) &= k_1(s)^2, \\ k_2'(s) &= 1, \\ w'(s) &= -2k_1(s) w(s). \end{aligned}$$

Wir gehen in drei Schritten vor.

1. Zunächst lösen wir das System gewöhnlicher Differentialgleichungen: Die erste Differentialgleichung enthält nur k_1 und ist eine mit getrennten Variablen. Man weiß, wie man diese löst (wirklich?) und erhält $k_1(s) = -\frac{1}{s+c_1}$ mit einer Konstanten c_1 . Die zweite Differentialgleichung ist besonders einfach. Wir erhalten $k_2(s) = s + c_2$ mit einer Konstanten c_2 . Setzen wir die Form von k_1 ein, ist die dritte Differentialgleichung eine lineare Differentialgleichung für w . Auch diese können wir lösen und erhalten $w(s) = c_3(s + c_1)^2$ mit einer dritten Konstanten c_3 .

Für jedes dieser Tripel (c_1, c_2, c_3) haben wir damit eine Raumkurve beschrieben, insgesamt also eine **Schar von Charakteristiken**.

2. Nun parametrisieren wir die **Raum-Anfangskurve** $t \mapsto (t, 0, g(t))^\top$ und bestimmen die Konstanten in Abhängigkeit von t so, dass die Charakteristik bei $(t, 0, g(t))$ beginnt. Dies führt auf:

$$\begin{aligned} k_1(0, t) &= t, \quad \text{d.h.} \quad -\frac{1}{c_1} = t \quad \text{bzw.} \quad c_1 = -\frac{1}{t}, \\ k_2(0, t) &= 0, \quad \text{d.h.} \quad c_2 = 0, \\ w(0, t) &= g(t), \quad \text{d.h.} \quad c_3 c_1^2 = g(t) \quad \text{bzw.} \quad c_3 = \frac{g(t)}{c_1^2} = t^2 g(t). \end{aligned}$$

⁸Wir schlagen vor, dass man bei einer gegebenen Differentialgleichung zunächst die Variablen mit x_1, x_2 bezeichne, den Buchstaben t für die Parametrisierung der Anfangskurve reserviere und den Buchstaben s für den Parameter im charakteristischen System.

3. Damit haben wir die Lösungsfläche durch die Parameter s und t beschrieben, also

$$\begin{aligned}x_1 &= k_1(s, t) = -\frac{1}{s - 1/t} = \frac{t}{1 - st}, \\x_2 &= k_2(s, t) = s, \\u &= w(s, t) = t^2 g(t) \left(s - \frac{1}{t}\right)^2.\end{aligned}$$

Nun müssen wir nur noch s, t aus dem System eliminieren und u in Abhängigkeit von x_1 und x_2 ausdrücken. Es ist $s = x_2$ und (die erste Gleichung nach t auflösen):

$$t = \frac{1}{x_2 + 1/x_1} = \frac{x_1}{1 + x_1 x_2},$$

also

$$u(x_1, x_2) = \frac{x_1^2}{(1 + x_1 x_2)^2} g\left(\frac{x_1}{1 + x_1 x_2}\right) \left(-\frac{1}{x_1}\right)^2 = \frac{1}{(1 + x_1 x_2)^2} g\left(\frac{x_1}{1 + x_1 x_2}\right).$$

(b) Als weiteres Beispiel betrachten wir

$$\frac{\partial u}{\partial x_1} + 2u \frac{\partial u}{\partial x_2} - u = 0$$

mit Anfangsbedingung $u(0, x_2) = 3x_2$ für $x_2 \in \mathbb{R}$. Das charakteristische System ist

$$\begin{aligned}k_1'(s) &= 1, \\k_2'(s) &= 2u(k(s)) = 2w(s), \\w'(s) &= u(k(s)) = w(s).\end{aligned}$$

Dies zu lösen ist einfach:

$$k_1(s) = s + c_1, \quad w(s) = c_3 e^s, \quad k_2(s) = 2c_3 e^s + c_2$$

mit Konstanten c_1, c_2, c_3 . Die Lösung soll für $s = 0$ durch den Anfangspunkt

$$(\gamma_1(t), \gamma_2(t), g(\gamma(t))) = (0, t, 3t)$$

gehen. Dies liefert die Konstanten $c_1 = 0$, $c_3 = 3t$ und $2c_3 + c_2 = t$, also $c_2 = t - 6t = -5t$. Damit haben wir erhalten:

$$\begin{aligned}x_1 &= k_1(s, t) = s, \\x_2 &= k_2(s, t) = 6t e^s - 5t, \\u &= w(s, t) = 3t e^s.\end{aligned}$$

Aus den ersten beiden Gleichungen eliminieren wir s und t , also $s = x_1$ und $t = \frac{x_2}{6 \exp(x_1) - 5}$. Damit haben wir:

$$u(x_1, x_2) = \frac{3x_2}{6e^{x_1} - 5} e^{x_1}.$$

4.3 Separationsansätze

In speziellen, von der betrachteten Differentialgleichung, den Anfangs- und Randbedingungen und der zugrundeliegenden Geometrie abhängigen Situationen lassen sich partielle Differentialgleichungen durch **Separationsansätze** (Bernoulli-Ansätze) behandeln. Wir suchen zunächst eine Lösung als Produkt von Funktionen in je einer Variablen. Um das Vorgehen zu erläutern, stellen wir in diesem Kapitel einige Beispiele in zwei Variablen zusammen.

Beispiel 4.8 Betrachte das folgende **Anfangsrandwertproblem**

$$\begin{aligned} \frac{\partial u(x,t)}{\partial t} - a^2 \frac{\partial^2 u(x,t)}{\partial x^2} &= 0, & (x,t) \in (0,L) \times (0,\infty), \\ u(x,0) &= \sin\left(\frac{\pi}{L}x\right), & x \in (0,L), \\ u(0,t) = u(L,t) &= 0, & t \in (0,\infty), \end{aligned}$$

zur Wärmeleitungsgleichung mit positivem Diffusionskoeffizient $a^2 > 0$. Wir können dies als Modell für die Abkühlung eines dünnen Stabs der Länge L , der an den Enden auf einer konstanten Temperatur gehalten wird, auffassen. Es ist durch $u(0,x) = \sin\left(\frac{\pi}{L}x\right)$ der Anfangszustand zum Zeitpunkt $t = 0$ festgelegt.

Wir setzen den Separationsansatz $u(x,t) = v(t)w(x)$ in die Differentialgleichung ein und erhalten

$$v'(t)w(x) - a^2 v(t)w''(x) = 0,$$

bzw., falls $v(t) \neq 0$ und $w(x) \neq 0$,

$$\frac{v'(t)}{v(t)} = a^2 \frac{w''(x)}{w(x)}.$$

In der letzten Gleichung hängt die linke Seite nur von t und die rechte Seite nur von x ab, sodass beide Ausdrücke gleich derselben Konstanten, etwa a^2k , sein müssen, wobei $k \in \mathbb{C}$ noch zu bestimmen ist. Sind wir an reellwertigen Lösungen u bzw. v, w interessiert, können wir $k \in \mathbb{R}$ ansetzen und erhalten die beiden gewöhnlichen Differentialgleichungen

$$v'(t) = a^2k v(t) \quad \text{und} \quad w''(x) = k w(x).$$

Zu beiden Differentialgleichungen kennen wir die allgemeine Lösung. Somit gilt

$$v(t) = c e^{a^2kt} \quad \text{und} \quad w(x) = c_1 e^{\sqrt{k}x} + c_2 e^{-\sqrt{k}x}$$

mit Konstanten c, c_1, c_2 . Ist $k < 0$, so ist \sqrt{k} rein imaginär, und es tritt die komplexe Exponentialfunktion auf, siehe unten.

Nun müssen wir uns die Randbedingungen ansehen, um k (und vor allem das Vorzeichen) genauer zu bestimmen. Aus den Bedingungen $u(0,t) = u(L,t) = 0$ folgt $0 = w(0) = c_1 + c_2$ und $0 = w(L) = c_1 \exp(\sqrt{k}L) + c_2 \exp(-\sqrt{k}L)$, da uns der triviale Fall $v(t) = 0$ für $t > 0$ nicht interessiert. Also folgt

$$c_2 = -c_1 \quad \text{und} \quad 0 = c_1(e^{\sqrt{k}L} - e^{-\sqrt{k}L}) = 2c_1 \sinh(\sqrt{k}L).$$

Sei zunächst $k > 0$. Da der \sinh für reelle Argumente nur in $x = 0$ eine Nullstelle besitzt, so folgt $c_1 = 0$ und dann auch $c_2 = 0$. Damit erhalten wir nur die triviale Lösung $v \equiv 0$.

Sei jetzt $k < 0$. Dann ist

$$w(x) = \tilde{c}_1 \sin(\sqrt{|k|x}) + \tilde{c}_2 \cos(\sqrt{|k|x}).$$

Aus den Randbedingungen ergibt sich weiter $0 = w(0) = \tilde{c}_2$ und $0 = \tilde{c}_1 \sin(\sqrt{|k|}L)$. Also muss $\sqrt{|k|}L = n\pi$ für ein $n \in \mathbb{N}$ gelten. Insgesamt erhalten wir für jedes $n \in \mathbb{N}$ eine Lösung

$$w_n(x) = \tilde{c}_n \sin(n\pi x/L)$$

mit beliebigen Konstanten $\tilde{c}_n \in \mathbb{R}$. Für die zeitliche Abhängigkeit ergibt sich dann

$$v_n(t) = \hat{c}_n e^{-(n\pi a/L)^2 t}$$

mit beliebigen Konstanten \hat{c}_n . Damit haben wir eine ganze Schar u_n von Lösungen der Differentialgleichung gefunden, die alle den Randbedingungen genügen:

$$u_n(x, t) := c_n \sin(n\pi x/L) e^{-(n\pi a/L)^2 t}.$$

Da die Differentialgleichung linear ist, sind auch alle Linearkombinationen Lösungen und wir erhalten, dass die folgende Reihe

$$u(x, t) = \sum_{n=1}^{\infty} c_n e^{-(n\pi a/L)^2 t} \sin(n\pi x/L)$$

Lösung der partiellen Differentialgleichung ist – vorausgesetzt natürlich, die Reihe ist differenzierbar. Bezüglich der Ortskoordinate handelt es sich um eine **Fourierreihe**, die wir bereits in HM2 behandelt haben.

Schließlich betrachten wir noch die Anfangsbedingung und erhalten durch Koeffizientenvergleich $c_1 = 1$ und $c_n = 0$ für $n \geq 2$. Insgesamt haben wir so die Lösung

$$u(x, t) = e^{-(\pi a/L)^2 t} \sin(\pi x/L)$$

des Anfangsrandwertproblems bestimmt.

Beispiel 4.9 Wir betrachten die Auslenkung $u(x, t)$ einer Saite, die an den Endpunkten $x = 0$ und $x = L$ fest eingespannt ist. Zum Zeitpunkt $t = 0$ sei die Saite ausgelenkt mit $u(x, 0) = f(x)$ und in Ruhe, d.h. $\partial u(x, 0)/\partial t = 0$. Das zugehörige Modell ist das Anfangsrandwertproblem

$$\begin{aligned} \frac{\partial^2 u(x, t)}{\partial t^2} - a^2 \frac{\partial^2 u(x, t)}{\partial x^2} &= 0, \quad (x, t) \in (0, L) \times (0, \infty), \\ u(x, 0) &= f(x), \quad \frac{\partial u(x, 0)}{\partial t} = 0, \quad x \in (0, L), \\ u(0, t) &= u(L, t) = 0, \quad t \in (0, \infty). \end{aligned}$$

Ein Separationsansatz der Form $u(x, t) = v(t)w(x)$ führt auf

$$v''(t)w(x) - a^2v(t)w''(x) = 0 \quad \text{bzw.} \quad \frac{v''(t)}{v(t)} = a^2 \frac{w''(x)}{w(x)},$$

wenn $v(t) \neq 0$ und $w(x) \neq 0$ sind. Es folgt wiederum, dass diese Identität nur erfüllt werden kann, wenn beide Seiten konstant sind. Wir erhalten die gewöhnlichen Differentialgleichungen

$$v''(t) = a^2k v(t) \quad \text{und} \quad w''(x) = k w(x)$$

mit den allgemeinen Lösungen

$$v(t) = a_1 e^{ak^{1/2}t} + a_2 e^{-ak^{1/2}t} \quad \text{und} \quad w(x) = b_1 e^{k^{1/2}x} + b_2 e^{-k^{1/2}x}.$$

Aus der Randbedingung $w(0) = w(L) = 0$ ergeben sich wie im vorherigen Beispiel nur die Möglichkeiten $k = -\left(n\pi/L\right)^2$ für $n \in \mathbb{N}$ mit den zugehörigen Lösungen

$$w_n(x) = \tilde{b}_n \sin\left(\frac{n\pi}{L}x\right), \quad \tilde{b}_n \in \mathbb{R}.$$

Für die Zeitabhängigkeit erhalten wir

$$v_n(t) = \tilde{a}_{1,n} \cos\left(\frac{an\pi}{L}t\right) + \tilde{a}_{2,n} \sin\left(\frac{an\pi}{L}t\right).$$

Summen von den Funktionen $u_n(x, t) = v_n(t)w_n(x)$ sind ebenso Lösungen, und wir erhalten allgemeiner, dass Reihen von der Form

$$u(x, t) = \sum_{n=1}^{\infty} \sin\left(\frac{n\pi}{L}x\right) \left[a_{1,n} \cos\left(\frac{an\pi}{L}t\right) + a_{2,n} \sin\left(\frac{an\pi}{L}t\right) \right]$$

Lösungen der Differentialgleichung sind, die die Randbedingung erfüllen, wenn die Konvergenz der Reihe gliedweises Differenzieren erlaubt. Hier handelt es sich um eine Fourierreihe sowohl bezüglich der Orts- als auch bezüglich der Zeitkoordinate.

Zum Zeitpunkt $t = 0$ gilt

$$\begin{aligned} u(x, 0) &= \sum_{n=1}^{\infty} a_{1,n} \sin\left(\frac{n\pi}{L}x\right) = f(x) \quad \text{und} \\ \frac{\partial u(x, 0)}{\partial t} &= \sum_{n=1}^{\infty} a_{2,n} \frac{an\pi}{L} \sin\left(\frac{n\pi}{L}x\right) = 0. \end{aligned}$$

Die bei gegebenen f unbekanntem Koeffizienten $a_{1,n}$ und $a_{2,n}$ sind die Fourierkoeffizienten, $a_{2,n} = 0$ für $n \in \mathbb{N}$ und

$$a_{1,n} = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{n\pi}{L}x\right) dx, \quad n \in \mathbb{N}.$$

Als letztes Beispiel wollen wir wieder einen Produktansatz einsetzen, aber diesmal im Fall von rotationssymmetrischen Geometrien.

Beispiel 4.10 Auf der Einheitskreisscheibe $D = \{x \in \mathbb{R}^2 : \|x\|_2 < 1\}$ ist eine Lösung des **Dirichlet-Problems** zur Laplace-Gleichung

$$\begin{aligned} \Delta u &= 0 \quad \text{in } D, \\ u &= f \quad \text{auf } \partial D \end{aligned}$$

gesucht. Allgemein werden Randbedingungen, die die Funktionswerte auf dem Rand festlegen, **Dirichlet-Bedingungen** genannt.

Da die Randdaten auf dem Kreis $\|x\|_2 = 1$ vorgegeben sind, bietet es sich an, dieses Problem in Polarkoordinaten, $x_1 = r \cos \varphi$ und $x_2 = r \sin \varphi$, zu betrachten mit $D = \{(r, \varphi) \in [0, 1) \times [0, 2\pi)\}$ und einen Separationsansatz $u(x_1, x_2) = U(r, \varphi) = v(r)w(\varphi)$ zu versuchen. Dazu müssen wir zunächst den Laplaceoperator in Polarkoordinaten ausdrücken.

Durch Anwenden der Kettenregel erhalten wir für den Laplace-Operator in Polarkoordinaten

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = \frac{\partial^2 U}{\partial r^2} + \frac{1}{r} \frac{\partial U}{\partial r} + \frac{1}{r^2} \frac{\partial^2 U}{\partial \varphi^2}.$$

Also liefert die Laplace-Gleichung für den Ansatz $U(r, \varphi) = v(r) w(\varphi)$ die Gleichung

$$v''(r) w(\varphi) + \frac{1}{r} v'(r) w(\varphi) + \frac{1}{r^2} v(r) w''(\varphi) = 0$$

bzw. für $v(r) \neq 0 \neq w(\varphi)$

$$\frac{r^2 v''(r) + r v'(r)}{v(r)} = -\frac{w''(\varphi)}{w(\varphi)}.$$

Da die linke Seite nur von r und die rechte Seite nur von φ abhängt, muss es eine Konstante k geben mit

$$\frac{r^2 v''(r) + r v'(r)}{v(r)} = k = -\frac{w''(\varphi)}{w(\varphi)}.$$

Also sind die gewöhnlichen Differentialgleichungen

$$r^2 v''(r) + r v'(r) - k v(r) = 0 \quad \text{und} \quad w''(\varphi) + k w(\varphi) = 0$$

zu lösen. Da w aber 2π -periodisch sein muss, folgt $k = n^2 > 0$ für $n \in \{0, 1, 2, \dots\}$ und

$$w_n(\varphi) = \tilde{a}_n \cos(n\varphi) + \tilde{b}_n \sin(n\varphi).$$

Mit $k = n^2$ ergibt sich für v_n eine **Eulersche Differentialgleichung**,

$$r^2 v_n''(r) + r v_n'(r) - n^2 v_n(r) = 0$$

mit der allgemeinen Lösung

$$v_0(r) = c_0 + \tilde{c}_0 \ln r \quad \text{und} \quad v_n(r) = c_n r^n + \tilde{c}_n r^{-n}$$

für $n \in \mathbb{N}$. Da wir nach einer stetigen Lösung u in D suchen, fallen die in $r = 0$ singulären Anteile heraus, d.h. $\tilde{c}_0 = \tilde{c}_n = 0$, und wir erhalten die folgenden Lösungen

$$u_0(r, \varphi) = \frac{1}{2} a_0 \quad \text{und} \quad u_n(r, \varphi) = r^n [a_n \cos(n\varphi) + b_n \sin(n\varphi)]$$

für $n \in \mathbb{N}$ und beliebige Konstanten a_0, a_n, b_n . Durch Aufsummieren ergibt sich wieder, dass die daraus gebildete Fourierreihe in φ ,

$$u(r, \varphi) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} r^n [a_n \cos(n\varphi) + b_n \sin(n\varphi)],$$

eine Lösung der Laplace-Gleichung ist, wenn die Konvergenz gliedweises Differenzieren erlaubt.

Für $r = 1$ folgt

$$f(\varphi) = u(1, \varphi) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} [a_n \cos(n\varphi) + b_n \sin(n\varphi)].$$

Damit haben wir die Lösung des Randwertproblems angegeben mit Hilfe der Fourierentwicklung von f . Die Fourierkoeffizienten von f sind nämlich:

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(\varphi) \cos(n\varphi) d\varphi, \quad n = 0, 1, \dots,$$

$$b_n = \frac{1}{\pi} \int_0^{2\pi} f(\varphi) \sin(n\varphi) d\varphi, \quad n = 1, 2, \dots$$

Beachte: Die Reihe $u(r, \varphi)$ ist eine differenzierbare Lösung der Laplace-Gleichung im Inneren von D , auch wenn nur $f \in L^2(0, 2\pi)$ gilt.

5 Einführung in die Stochastik

5.1 Grundbegriffe

Mit dem Begriff **Stochastik** wird der Bereich der Mathematik bezeichnet, der sich mit dem Zufall beschäftigt. In diesem ersten Abschnitt wollen wir in die grundlegenden Definitionen einführen, die üblich sind, um zufällige Vorgänge zu beschreiben (Modellbildung), Modelle experimentell zu belegen (Statistik) und Folgerungen daraus zu ziehen (Wahrscheinlichkeitstheorie). Wir halten uns in diesem Abschnitt an die Bücher „Mathematik für Wirtschaftsingenieure 1“, Kapitel 3 und 4, „Mathematik für Wirtschaftsingenieure 2“, Kapitel 9, von N. Henze und G. Last (Vieweg Verlag) sowie an „Mathematik“, Kapitel 37 und 38 von T. Arens, F. Hettlich, Ch. Karpfinger, U. Kockelkorn, K. Lichtenegger und H. Stachel (Spektrum Verlag).

Wie im letztgenannten Buch beginnen wir mit einem Beispiel, das bereits die Schwierigkeit mit der mathematisch korrekten Behandlung der Stochastik aufzeigt.

Denken wir uns einen Teich, in dem eine Ente untergetaucht ist. Wir wissen nicht, wo sie auftaucht, nehmen aber an, dass „jede Stelle gleich wahrscheinlich“ ist. Wir modellieren den Teich als ebenes Gebiet Ω , etwa ein achsenparalleles Rechteck in der (x_1, x_2) -Ebene. Die möglichen *Ergebnisse* modellieren wir (vorläufig) als Teilmengen $A \subseteq \Omega$ von Ω . Taucht die Ente in A auf, so sagen wir: *Das Ereignis A ist eingetreten*. Den Ergebnissen $A \subseteq \Omega$ sollen Wahrscheinlichkeiten zugeordnet werden. Als einfachste – und plausibelste – Möglichkeit definieren wir die Wahrscheinlichkeit durch

$$P(A) = \frac{\text{Flächeninhalt von } A}{\text{Flächeninhalt von } \Omega}.$$

Das mathematische Problem ist nun, dass man nicht den Flächeninhalt jeder beliebigen Teilmenge definieren kann, ohne in Widersprüche zu den gewohnten Regeln zu geraten. Daher kann die obige Definition von $P(A)$ nicht für alle Teilmengen $A \subseteq \Omega$ gegeben werden, sondern nur für solche mit einem vernünftig definierbaren Flächeninhalt. Man kann z.B. (und dies ist ein ganz wichtiges Beispiel!) für A achsenparallele Rechtecke wählen und den Flächeninhalt in gewohnter Form als das Produkt der Seitenlängen definieren. Für Komplemente, endliche Vereinigungen und Durchschnitte von solchen Rechtecken können wir auch Flächeninhalte bilden. Schließlich können wir das auch für Mengen tun, die durch abzählbar unendlich viele Mengenoperationen (Vereinigung, Durchschnitt, Komplement) aus Rechtecken gewonnen werden können. Diese so konstruierten „gutartigen“ Teilmengen $A \subseteq \Omega$ heißen **Borelmengen**.

Beim Standardbeispiel des zweimaligen Werfens einer Münze mit den Ausgängen „Kopf“ oder „Zahl“ ist die Sache einfacher. Die möglichen Ergebnisse sind die 4 Paare (K, K) , (K, Z) , (Z, K) , (Z, Z) . Wir fassen diese zusammen zur *Ergebnismenge*

$$\Omega = \{(K, K), (K, Z), (Z, K), (Z, Z)\}.$$

Da alle 4 Ergebnisse gleich wahrscheinlich sind, ordnen wir jedem Ergebnis die Wahrscheinlichkeit $1/4$ zu und jeder Teilmenge von Ω deren Anzahl multipliziert mit $1/4$. Hier haben wir P also auf der **Potenzmenge** $\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$ definiert durch

$$P(A) = \frac{\text{Anzahl der Elemente von } A}{\text{Anzahl der Elemente von } \Omega} \quad \text{für } A \in \mathcal{P}(\Omega).$$

Wir geben jetzt einige mathematische Definitionen.

Definition 5.1 (a) Unter einem **Zufallsexperiment** verstehen wir einen Vorgang, der wiederholbar ist, aber dessen **Ergebnis** zufällig ist.

(b) Die Menge Ω aller möglichen Ergebnisse eines Zufallsexperiments heißt **Ergebnismenge**.

Wir wollen den Teilmengen der Ergebnismenge jetzt Wahrscheinlichkeiten zuordnen. Wie wir am Beispiel des Teiches gesehen haben, geht das nicht immer für jede Teilmenge. Daher die folgende Definition:

Definition 5.2 Sei Ω eine Menge und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine Teilmenge der Potenzmenge $\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$ mit folgenden Eigenschaften:

(i) $\Omega \in \mathcal{A}$.

(ii) Ist $A \in \mathcal{A}$, so auch das Komplement $A^c := \Omega \setminus A \in \mathcal{A}$.

(iii) Sind $A_k \in \mathcal{A}$ für alle $k \in \mathbb{N}$, so ist auch $\bigcup_{k \in \mathbb{N}} A_k \in \mathcal{A}$.

Dann heißt \mathcal{A} eine σ -**Algebra**.

Aus den Axiomen (i), (ii), (iii) lassen sich sofort weitere Eigenschaften nachweisen:

(a) Für die leere Menge gilt: $\emptyset \in \mathcal{A}$.

(b) Sind $A_k \in \mathcal{A}$ für alle $k \in \mathbb{N}$, so ist auch $\bigcap_{k \in \mathbb{N}} A_k \in \mathcal{A}$.

Aussage (a) folgt sofort aus (i) und (ii), für Aussage (b) benutzen wir die Formel

$$\Omega \setminus \bigcap_{k \in \mathbb{N}} A_k = \bigcup_{k \in \mathbb{N}} (\Omega \setminus A_k).$$

Da alle $A_k \in \mathcal{A}$, so auch wegen (ii) die Komplemente $\Omega \setminus A_k$ und wegen (iii) die Vereinigung $\bigcup_{k \in \mathbb{N}} (\Omega \setminus A_k)$ und, wieder mit (ii) ihr Komplement.

Für jede beliebige Teilmenge $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ können wir die *kleinste* σ -Algebra \mathcal{A} konstruieren, die \mathcal{B} enthält:

$$\mathcal{A} = \bigcap_{\{\mathcal{C} \text{ } \sigma\text{-Algebra mit } \mathcal{B} \subseteq \mathcal{C}\}} \mathcal{C}.$$

\mathcal{A} heißt die **von \mathcal{B} erzeugte σ -Algebra**.

Beispiele 5.3 für σ -Algebren

(a) $\mathcal{A} = \mathcal{P}(\Omega)$ ist natürlich selbst eine σ -Algebra.

(b) Sei $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $A_1 = \{2, 4, 6\}$, $A_2 = \{1, 3, 5\}$, $A_3 = \{6\}$. Dann ist (selbst nachprüfen!) die von $\{A_1, A_2, A_3\}$ erzeugte σ -Algebra

$$\{\emptyset, \{6\}, \{2, 4\}, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 3, 5, 6\}, \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 5, 6\}\}.$$

(c) Nehmen wir $\Omega = \mathbb{R}^n$ und für \mathcal{B} die Menge aller achsenparallelen Quader (egal, ob offen, abgeschlossen, halboffen oder die leere Menge), so heißt die davon erzeugte σ -Algebra die **σ -Algebra der Borelmengen**. Leider kann man sie nicht gut beschreiben. Man kann jedoch zeigen, dass jede Borelmenge Lebesgue-messbar ist.

Wir definieren nun **Ereignisse** als solche Teilmengen $A \subseteq \Omega$, denen eine **Wahrscheinlichkeit** zugeordnet werden kann.

Definition 5.4 (Axiome von Kolmogoroff)

Sei Ω eine Ergebnismenge mit Potenzmenge $\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$ und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine σ -Algebra. Eine Abbildung

$$P : \mathcal{A} \longrightarrow \mathbb{R}$$

heißt **Wahrscheinlichkeitsmaß** oder **Wahrscheinlichkeitsfunktion** oder **Wahrscheinlichkeitsverteilung**, wenn gilt

(i) $P(A) \geq 0$ für alle $A \in \mathcal{A}$,

(ii) $P(\Omega) = 1$,

(iii) $P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$ für paarweise disjunkte Ereignisse $A_1, A_2, \dots \in \mathcal{A}$ (d.h. für solche mit $A_j \cap A_k = \emptyset$ für alle $j \neq k$).

Jedes Tripel (Ω, \mathcal{A}, P) mit obigen Eigenschaften heißt ein **Wahrscheinlichkeitsraum**.

Folgerungen 5.5 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann gilt

1. $P(A^c) = 1 - P(A)$ für $A \in \mathcal{A}$. Insbesondere ist $P(A) \in [0, 1]$.
2. $P(\emptyset) = 0$.
3. Aus $A, B \in \mathcal{A}$ und $A \subseteq B$ folgt $P(A) \leq P(B)$.
4. Für $A, B \in \mathcal{A}$ gilt: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Beweis: (a) Wegen $A \cup A^c = \Omega$ und $A \cap A^c = \emptyset$ ist mit den Kolmogoroffaxiomen

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c).$$

Weiter erhalten wir aus $P(A^c) = 1 - P(A) \geq 0$, dass $P(A) \leq 1$ gilt.

(b) Dies ergibt sich aus (a) wegen $\emptyset = \Omega^c$ und $P(\Omega) = 1$.

(c) Aus $A \subseteq B$ und $A \cap (B \setminus A) = \emptyset$ folgt

$$P(B) = P(A \cup (B \setminus A)) = P(A) + \underbrace{P(B \setminus A)}_{\geq 0}.$$

Also ist $P(B) \geq P(A)$.

(d) Es gilt $P(A) = P(A \setminus B) + P(A \cap B)$ und $P(B) = P(B \setminus A) + P(A \cap B)$. Wir erhalten somit

$$\begin{aligned} P(A \cup B) &= P((A \setminus B) \cup (B \setminus A) \cup (A \cap B)) \\ &= P(A \setminus B) + P(B \setminus A) + P(A \cap B) \\ &= [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)] + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

□

Wir notieren noch die folgenden Sprechweisen. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

- Die Elemente aus \mathcal{A} , also den Teilmengen A von Ω , denen eine Wahrscheinlichkeit $P(A)$ zugeordnet wird, heißen **Ereignisse**. Wir sagen: „ein Ereignis A tritt ein“, wenn für das Ergebnis $\omega \in \Omega$ eines Zufallsexperiments $\omega \in A$ gilt.
- Die einelementigen Teilmengen $\{\omega\} \subseteq \Omega$ heißen **Elementarereignisse**, wenn sie Elemente von \mathcal{A} sind. (Das muss nicht immer so sein, bei unseren Beispielen aber schon.)
- Ist $A \in \mathcal{A}$ ein Ereignis, so heißt $A^c = \Omega \setminus A$ das zugehörige **komplementäre Ereignis**. Die leere Menge $\emptyset \subseteq \Omega$ wird **unmögliches Ereignis** und die gesamte Menge Ω **sicheres Ereignis** genannt.
- Zwei Ereignisse $A, B \in \mathcal{A}$ heißen **unvereinbar**, wenn $A \cap B = \emptyset$ gilt.

In bestimmten Situationen lässt sich eine Wahrscheinlichkeitsfunktion relativ leicht angeben.

Definition 5.6 Ein **Laplace-Experiment** ist gegeben durch eine endliche Ergebnismenge $\Omega = \{\omega_1, \dots, \omega_n\}$, die σ -Algebra $\mathcal{A} = \mathcal{P}(\Omega)$ und die Wahrscheinlichkeitsfunktion

$$P(A) = \frac{\text{Anzahl von } A}{\text{Anzahl von } \Omega} \quad \text{für } A \subseteq \Omega.$$

Die Elementarereignisse $\{\omega_j\}$ sind also alle gleich wahrscheinlich und haben die Wahrscheinlichkeit $1/n$.

Zur Abkürzung schreiben wir im Folgenden auch $P(\omega)$ statt $P(\{\omega\})$ bei Elementarereignissen.

Definieren wir für die Anzahl der Elemente einer endlichen Menge $A = \{a_1, \dots, a_n\}$ die Kardinalzahl, $\text{card}(A) = n$, so gilt bei einem Laplace-Experiment

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

Beispiel 5.7 Wie groß ist die Wahrscheinlichkeit beim Würfeln mit zwei Würfeln, eine Augensumme größer gleich 10 zu erreichen?

Setze $\Omega = \{(i, j) \in \mathbb{N}^2 : i, j \leq 6\}$. Ω enthält 36 Elemente, d.h. $\text{card}(\Omega) = 36$. Da jedes Ergebnis gleich wahrscheinlich ist, liegt ein Laplace-Experiment vor, und es ist

$$P(\omega) = \frac{1}{36} \quad \text{für } \omega \in \Omega.$$

Für das Ereignis

$$A = \{(i, j) \in \Omega : i + j \geq 10\} = \{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\},$$

d.h. $\text{card}(A) = 6$, ergibt sich somit

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{6}{36} = \frac{1}{6}.$$

Um bei Zufallsexperimenten die Wahrscheinlichkeitsfunktion zu bestimmen, müssen wir im Allgemeinen Schätzungen auf der Grundlage statistischer Daten heranziehen. Im Fall eines Laplace-Experiments läßt sich, wie gezeigt, die Wahrscheinlichkeitsfunktion aber durch Abzählen der Ergebnismenge ermitteln. Dies führt uns auf Fragestellungen der *Kombinatorik*. Wir müssen vier grundlegende Situationen unterscheiden.

Satz 5.8 Sei $G = \{1, 2, \dots, N\} \subset \mathbb{N}$.

(a) **Permutationen mit Wiederholungen:** Sei $n \in \mathbb{N}$ und

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j \in G, j = 1, \dots, n\} = G^n.$$

Dann ist $\text{card}(\Omega) = N^n$.

(b) **Permutationen ohne Wiederholungen:** Sei $n \leq N$ und

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j \in G \text{ und } \omega_j \neq \omega_k \text{ für } j \neq k\}.$$

Dann ist $\text{card}(\Omega) = N(N-1)\dots(N-n+1) = \frac{N!}{(N-n)!} = \binom{N}{n} (n!)$. Im Spezialfall $N = n$ gilt $\text{card}(\Omega) = N!$.

(c) **Kombinationen ohne Wiederholungen:** Für $n \leq N$ und

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j \in G, \omega_j < \omega_{j+1} \text{ für } j = 1, \dots, n-1\}$$

gilt $\text{card}(\Omega) = \binom{N}{n} = \frac{N!}{n!(N-n)!}$.

(d) **Kombinationen mit Wiederholungen:** Für

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j \in G, \omega_j \leq \omega_{j+1} \text{ für } j = 1, \dots, n-1\}$$

gilt $\text{card}(\Omega) = \binom{N+n-1}{n}$.

Bemerkung: Wir können selbstverständlich beliebige Objekte permutieren oder kombinieren, etwa drei Münzen, d.h. $\omega = (\omega_1, \omega_2, \omega_3)$ mit $\omega_j \in G = \{\text{Wappen, Zahl}\}$. Aber bei endlichen Ergebnismengen lassen sich die Elemente in G stets nummerieren, so dass wir ohne Einschränkung $G = \{1, \dots, N\}$ annehmen dürfen.

Beweis: (a) Offensichtlich gilt $\text{card}(\Omega) = N$ für $n = 1$. Nun gehen wir induktiv vor. Sei die Aussage gezeigt für Permutationen mit $n-1$ Stellen und sei

$$\Omega_n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j \in G \text{ für } j = 1, \dots, n\}$$

gegeben. Wir zerlegen die Ergebnismenge in N disjunkte Teilmengen mit

$$\begin{aligned} \Omega_n &= \{(\omega_1, \dots, \omega_{n-1}, 1) : \omega_j \in G \text{ für } j = 1, \dots, n-1\} \\ &\cup \{(\omega_1, \dots, \omega_{n-1}, 2) : \omega_j \in G \text{ für } j = 1, \dots, n-1\} \\ &\cup \dots \cup \{(\omega_1, \dots, \omega_{n-1}, N) : \omega_j \in G \text{ für } j = 1, \dots, n-1\}. \end{aligned}$$

Also ist Ω_n Vereinigung von N Mengen von Permutationen mit Wiederholungen mit $n - 1$ Elementen. Die Induktionsannahme liefert

$$\text{card}(\Omega) = N \cdot N^{n-1} = N^n.$$

(b) Auch diese Aussage folgt induktiv, wenn wir die Ergebnismenge Ω_n darstellen durch

$$\begin{aligned}\Omega_n &= \{(\omega_1, \dots, \omega_n) : \omega_j \in G, \omega_j \neq \omega_k \text{ für } j \neq k\} \\ &= \{(\tilde{\omega}, \omega_n) : \tilde{\omega} \in \Omega_{n-1}, \omega_n \in G \setminus \{\omega_1, \dots, \omega_{n-1}\}\}.\end{aligned}$$

Wegen $\text{card}(G \setminus \{\omega_1, \dots, \omega_{n-1}\}) = N - n + 1$ erhalten wir

$$\text{card}(\Omega_n) = \text{card}(\Omega_{n-1})(N - n + 1) = N(N - 1) \dots (N - n + 2)(N - n + 1).$$

(c) Wir wissen bereits, dass es $N(N - 1) \dots (N - n + 1)$ mögliche Permutationen ohne Wiederholungen von N Zahlen mit n Plätzen gibt. Da hier aber Permutationen, die dieselben Zahlen enthalten, gleichwertig sind, also nur einmal gezählt werden (z.B. $(1, 2, 3, 4)$ und $(1, 3, 4, 2)$), müssen wir durch die Anzahl an Möglichkeiten diese n Zahlen auf die n Plätze zu verteilen, also durch $n!$ dividieren. Wir erhalten den Binomialkoeffizienten

$$\text{card}(\Omega) = \frac{N(N - 1) \dots (N - n + 1)}{n!} = \binom{N}{n}.$$

(d) Sei $\omega = (\omega_1, \dots, \omega_n) \in \Omega$, d.h. $\omega_j \in G$ und $\omega_1 \leq \dots \leq \omega_n$. Die Anzahl solcher Kombinationen, die auch Wiederholungen erlauben, ergibt sich aus der Situation in (c) durch folgenden Trick. Wir transformieren ω zu $\alpha = (\alpha_1, \dots, \alpha_n)$ mit

$$\alpha_j = \omega_j + j - 1, \quad j = 1, \dots, n.$$

Dann gilt

$$\alpha \in \tilde{\Omega} = \{(\alpha_1, \dots, \alpha_n) : \alpha_j \in \{1, \dots, N + n - 1\} \text{ und } \alpha_1 < \alpha_2 < \dots < \alpha_n\},$$

denn es ist

$$\alpha_j = \omega_j + j - 1 \leq \omega_{j+1} + j - 1 < \omega_{j+1} + j = \alpha_{j+1}.$$

Außerdem können wir umgekehrt jedem $\alpha \in \tilde{\Omega}$ genau ein $\omega \in \Omega$ zuordnen. Dies sehen wir, indem wir $\omega_j = \alpha_j + 1 - j$ setzen und abschätzen

$$\omega_j = \alpha_j + 1 - j < \alpha_{j+1} + 1 - j = \omega_{j+1} + 1.$$

Da $\omega_j, \omega_{j+1} \in \mathbb{N}$ sind, so folgt $\omega_j \leq \omega_{j+1}$ für alle $j = 1, \dots, n - 1$. Also gibt es genauso viele $\omega \in \Omega$ wie $\alpha \in \tilde{\Omega}$, und wir erhalten mit Teil (c)

$$\text{card}(\Omega) = \text{card}(\tilde{\Omega}) = \binom{N + n - 1}{n}.$$

□

Bemerkungen 5.9

Permutationen und Kombinationen sind mathematische Beschreibungen für Modelle wie z.B. das **Fächermodell** oder das **Urnenmodell**.

Beim Fächermodell fragt man nach den Möglichkeiten, n Teilchen auf N Fächer zu verteilen.

Beim Urnenmodell geht es um das Ziehen von n Kugeln aus einer Urne mit N Kugeln. Die folgende Tabelle beschreibt die Zuordnungen:

	Fächermodell	Urnenmodell
Perm. mit Wdh.	Teile unt.bar, mit Mehrfachbelegung	mit Zurückkl., mit Reihenf.
Perm. ohne Wdh.	Teile unt.bar, ohne Mehrfachbel.	ohne Zurückkl., mit Reihenf.
Komb. ohne Wdh.	Teile nicht unt.bar, keine Mehrfachbel.	ohne Zurückkl., ohne Reihenf.
Komb. mit Wdh.	Teile nicht unt.bar, mit Mehrfachbel.	mit Zurückkl., ohne Reihenf.

Weitere Bemerkungen:

zu (a) Bei Permutationen ist die *Reihenfolge* der Einträge in ω wichtig und festgelegt. Außerdem dürfen sich bei dieser ersten Variante die einzelnen Einträge wiederholen, etwa $\omega = (2, 5, 7, 2)$ oder $\omega = (3, 3, 3, 1)$ für $N = 7$ und $n = 4$. Ein typisches Beispiel ist das Fußballtoto. Es werden die Resultate von $n = 11$ Spielen zusammengestellt, wobei die einzelnen Einträge $\omega_j \in G = \{0, 1, 2\}$ (d.h. $N = 3$) liegen (0 für unentschieden, 1 für Heimsieg, 2 für Auswärtssieg). Ein Ergebnis ist also $\omega = (\omega_1, \omega_2, \dots, \omega_{11})$, wobei die Reihenfolge durch die ausgewählten Spiele festgelegt ist. Die Anzahl der Möglichkeiten ergibt sich zu

$$\text{card}(\Omega) = 3^{11} = 177147.$$

Also ist die Wahrscheinlichkeit

$$P(\omega) = \frac{1}{177147}$$

für ein Elementarereignis, nämlich alle Spiele richtig getippt zu haben, wenn jedes Ergebnis gleich wahrscheinlich wäre.

zu (b) Bei Permutationen ohne Wiederholungen spielt die Reihenfolge der Einträge in ω ebenfalls eine Rolle, aber doppelte Einträge sind nicht erlaubt, etwa

$$\omega = (2, 5, 3, 4) \quad \text{oder} \quad \omega = (1, 2, 4, 7)$$

für $N = 7$ und $n = 4$. Offensichtlich muss hier $n \leq N$ gelten. Beispiel für ein solches Zufallsexperiment ist etwa die zufällige Verteilung von $n = 100$ Studenten auf $N = 120$ Plätze im Hörsaal.

zu (c) Im Fall von Kombinationen kommt es nicht auf die Reihenfolge der Einträge in ω an, und wir können deshalb die ω_j beliebig sortieren. Um eine eindeutige Darstellung zu gewährleisten, legen wir eine aufsteigende Anordnung der Einträge zur Beschreibung eines Elementarereignisses fest.

Mengentheoretisch bedeutet dies: Wir suchen die Anzahl an n -elementigen Teilmengen einer N -elementigen Menge. Ein typisches Beispiel ist das Zahlenlotto. Aus 49 nummerierten Kugeln werden 6 gezogen und wir erhalten

$$\binom{49}{6} = \frac{49 \cdot 48 \cdot \dots \cdot 44}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 13983816$$

Möglichkeiten.

zu (d) Wenn wir zwei nicht unterscheidbare Würfel werfen, so spielt es für das Ergebnis keine Rolle, welcher Würfel eine bestimmte Augenzahl anzeigt. Die Anzahl an verschiedenen Zahlenkombinationen ist von der Reihenfolge unabhängig, und es können Wiederholungen auftreten (Pasch). Es gibt also

$$\binom{N+n-1}{n} = \binom{6+2-1}{2} = \binom{7}{2} = 21$$

unterschiedliche Ergebnisse. Aber Vorsicht (!), dieses Zufalls-Experiment ist kein Laplace-Experiment. Das Ergebnis $(1, 1)$ hat die Wahrscheinlichkeit $P(\omega) = \frac{1}{36}$, aber das Ergebnis $\omega = (1, 2)$ hat die Wahrscheinlichkeit $P(\omega) = \frac{2}{36} = \frac{1}{18}$. Wenn wir nämlich die Würfel unterscheiden könnten, so gäbe es zwei Varianten, $(1, 2)$ und $(2, 1)$, zu dem Ereignis, dass ein Würfel 1 und ein Würfel 2 anzeigt.

Zwei weitere Begriffe, *bedingte Wahrscheinlichkeit* und *stochastische Unabhängigkeit*, führen wir in diesem Abschnitt noch ein.

Oft stellt sich die Frage nach der Wahrscheinlichkeit eines Ereignisses A unter der Bedingung, dass Ereignis B eintritt. Dazu das folgende Beispiel.

Beispiel 5.10 Eine Krankheit kommt bei 0,1% der Bevölkerung vor. Es gibt einen Test, um die Krankheit zu diagnostizieren. Der Test zeigt bei 99% der kranken Personen eine positive Reaktion. Aber auch bei 0,5% der Gesunden fällt der Test positiv aus. Wie groß ist die Wahrscheinlichkeit, dass jemand die Krankheit hat, wenn er positiv getestet wird?

Bezeichnen wir mir $N \in \mathbb{N}$ zunächst die Gesamtzahl der Bevölkerung, und sei A die Gruppe der kranken Personen und B die Gruppe der Menschen, bei denen der Test eine positive Reaktion zeigt. Die gesuchte Wahrscheinlichkeit ergibt sich nun aus dem Verhältnis der Anzahl der Kranken mit positivem Test zur Gesamtzahl aller Personen mit positivem Testergebnis. Also ist

$$P = \frac{\text{card}(A \cap B)}{\text{card}(B)}$$

zu bestimmen. Wir nehmen an, dass die Zugehörigkeit zu jeder dieser Gruppen für die gesamte Bevölkerung gleich wahrscheinlich ist. Gegeben sind folgende Daten:

$$\text{card}(A) = 0,001 \cdot N, \quad \text{card}(A \cap B) = 0,99 \cdot \text{card}(A), \quad \text{card}(A^c \cap B) = 0,005 \cdot \text{card}(A^c).$$

Der Zähler der gesuchten Wahrscheinlichkeit ist

$$\text{card}(A \cap B) = 0,99 \cdot \text{card}(A) = 0,99 \cdot 0,001 \cdot N.$$

Für den Nenner schreiben wir B als disjunkte Vereinigung: $B = (A \cap B) \cup (A^c \cap B)$. Also ist

$$\begin{aligned} \text{card}(B) &= \text{card}(A \cap B) + \text{card}(A^c \cap B) = 0,99 \cdot 0,001 \cdot N + 0,005 \cdot \text{card}(A^c) \\ &= 0,99 \cdot 0,001 \cdot N + 0,005 \cdot (N - 0,001 \cdot N). \end{aligned}$$

Also ist, da sich N heraushebt,

$$P = \frac{0,99 \cdot 0,001}{0,99 \cdot 0,001 + 0,005 \cdot 0,999} \approx 0,17.$$

Also ist jemand nur in 17% der Fälle wirklich krank, wenn unser Test positiv ausfällt.

Die gesuchte Größe $P = \frac{\text{card}(A \cap B)}{\text{card}(B)}$ ist genau die Wahrscheinlichkeit dafür, dass unter allen Ergebnissen von B das Ereignis A auftritt. Wir ersetzen hier also den Grundraum Ω durch den neuen Grundraum B . Dividieren wir in diesem Beispiel die Gleichung für P durch N , so ist

$$P = \frac{P(A \cap B)}{P(B)}.$$

Dies motiviert die folgende Definition.

Definition 5.11 Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ Ereignisse mit $P(B) > 0$. Dann heißt

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

bedingte Wahrscheinlichkeit von A unter der Bedingung B .

Die folgenden einfachen Rechenregeln sind nützlich im Umgang mit bedingten Wahrscheinlichkeiten.

Satz 5.12 Es sei wieder (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

(a) Sind $A, B \in \mathcal{A}$ mit $P(B) > 0$, so ist $P(A|B) + P(A^c|B) = 1$. Ist auch $P(A) > 0$, so ist

$$P(A)P(B|A) = P(A \cap B) = P(A|B)P(B) \quad (\text{Formel von Bayes})$$

(b) Seien $B_1, \dots, B_n \in \mathcal{A}$ disjunkte Ereignisse einer Ergebnismenge Ω mit $\bigcup_{j=1}^n B_j = \Omega$ und

$$P(B_j) > 0, \quad j = 1, \dots, n. \quad \text{Weiter sei } A \in \mathcal{A} \text{ ein weiteres Ereignis. Dann gilt } P(A) = \sum_{j=1}^n P(A|B_j)P(B_j) \quad (\text{Formel von der totalen Wahrscheinlichkeit})$$

Beweis: Die erste Formel von (a) ist klar wegen $P(A \cap B) + P(A^c \cap B) = P((A \cap B) \cup (A^c \cap B)) = P(B)$. Die Formel von Bayes folgt direkt aus der Definition.

(b) Es gilt mit den Axiomen von Kolmogoroff

$$\sum_{j=1}^n P(A|B_j)P(B_j) = \sum_{j=1}^n P(A \cap B_j) = P\left(\bigcup_{j=1}^n (A \cap B_j)\right) = P(A).$$

□

Bemerkung: Wir greifen Beispiel 5.10 nochmal auf. In der medizinischen Praxis werden bei solchen Tests aus empirischen Befunden als Kennzahlen die *Sensitivität*

$$P(B|A) = \frac{P(B \cap A)}{P(A)},$$

also die Wahrscheinlichkeit, dass eine kranke Person positiv getestet wird, und die *Spezifität*

$$P(B^c|A^c) = \frac{P(B^c \cap A^c)}{P(A^c)},$$

also die Wahrscheinlichkeit, dass ein Gesunder keine Reaktion beim Test zeigt, angegeben. Mit der Regel von Bayes können wir so auch für verschiedene Risikogruppen die Frage aus Beispiel 5.10 beantworten. So ist z.B.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

und wegen

$$P(B) = P(A|B)P(B) + P(A^c|B)P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

$$P(A|B) = \frac{P(A)P(B|A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}.$$

Setze etwa $\hat{p} = P(A)$. Dann ist $P(A^c) = 1 - \hat{p}$ und $P(B|A^c) = 1 - P(B^c|A^c)$ und daher die Wahrscheinlichkeit, dass jemand mit positivem Test krank ist, aus

$$P(A|B) = \frac{\hat{p} P(B|A)}{\hat{p} P(B|A) + (1 - \hat{p})(1 - P(B^c|A^c))}$$

in Abhängigkeit der Sensitivität, der Spezifität und der Wahrscheinlichkeit \hat{p} der dem Patienten zugeordneten Risikogruppe.

Häufig wird die Definition der bedingten Wahrscheinlichkeit umgekehrt benutzt, um aus

$$P(A \cap B) = P(A|B)P(B)$$

die Wahrscheinlichkeit des Ereignisses $A \cap B$ zu ermitteln. Eine einfache Induktion liefert die **Multiplikationsregel**

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|(A_1 \cap A_2)) \cdot \dots \cdot P(A_n|(A_1 \cap \dots \cap A_{n-1}))$$

für Ereignisse $A_1, \dots, A_n \in \mathcal{A}$ mit $P(A_1 \cap \dots \cap A_{n-1}) > 0$.

Wir betrachten weiterhin zwei Ereignisse $A, B \in \mathcal{A}$ in einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit $P(A) > 0, P(B) > 0$. Im Allgemeinen sind $P(A)$ und $P(A|B)$ verschieden. Wenn nun aber das Eintreten des Ereignisses B keinen Einfluss auf das Ereignis A hat, so ist

$$P(A|B) = P(A).$$

Dies ist äquivalent zu

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$

und zu

$$P(B|A) = P(B).$$

In diesem Fall sprechen wir von stochastisch unabhängigen Ereignissen. Die symmetrische Identität oben ist Grundlage für die allgemeine Definition der Unabhängigkeit.

Definition 5.13 *Es seien $A_1, \dots, A_n \in \mathcal{A}$ Ereignisse eines Wahrscheinlichkeitsraums (Ω, \mathcal{A}, P) . Dann heißen A_1, \dots, A_n **stochastisch unabhängig**, wenn*

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

gilt für alle Indermengen $I \subseteq \{1, \dots, n\}$.

Bemerkung: Der Begriff „stochastisch unabhängig“ darf nicht mit „unvereinbar“, d.h. $A \cap B = \emptyset$, verwechselt werden. Unvereinbare Ereignisse A, B können nur dann stochastisch unabhängig sein, wenn $P(A) = 0$ oder $P(B) = 0$ gilt, also in uninteressanten Spezialfällen.

Wir können noch fragen, wann A unabhängig von sich selbst ist, d.h.

$$P(A) = P(A \cap A) = P(A)^2.$$

Dies passiert offensichtlich nur in den Fällen $P(A) = 0$ oder $P(A) = 1$.

Eine wichtige Klasse von Zufallsexperimenten besteht darin, ein Einzelexperiment, bei dem nur die Ausgänge “Treffer” oder “kein Treffer” interessieren, n -mal stochastisch unabhängig voneinander zu wiederholen. Diese werden **Bernoulli-Experimente** genannt. Wir können die Ergebnismenge eines Bernoulli-Experiments durch

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_j \in \{0, 1\}, j = 1, \dots, n\} = \{0, 1\}^n$$

beschreiben, wobei $\omega_j = 1$ für einen Treffer bzw. $\omega_j = 0$ für eine Niete beim j -ten Versuch steht.

Als Beispiel stellen wir uns wiederholtes Würfeln vor, wobei der Ausgang eine “Sechs” zu würfeln ein Treffer bedeutet. Also ist etwa mit $\omega = (1, 0, \dots, 0)$ das Ereignis, im ersten Wurf eine Sechs zu würfeln und in allen anderen Würfeln keine Sechs zu haben, beschrieben. Wenn wir mit $\hat{p} \in [0, 1]$ die Wahrscheinlichkeit eines Treffers im Einzelexperiment bezeichnen (im Beispiel $\hat{p} = \frac{1}{6}$), so ist $1 - \hat{p}$ die Wahrscheinlichkeit für einen Fehlversuch im Einzelexperiment. Wir interessieren uns für die Wahrscheinlichkeit, nach n Einzelexperimenten genau r Treffer zu haben.

Satz 5.14 Sei $\hat{p} \in [0, 1]$ die Wahrscheinlichkeit für einen Treffer in einem Einzelexperiment eines Bernoulli-Experiments. Dann gilt für die Wahrscheinlichkeit von genau r Erfolgen bei n Wiederholungen

$$p_r = \binom{n}{r} \hat{p}^r (1 - \hat{p})^{n-r}.$$

Beweis: Wir betten zunächst das Einzelexperiment in den Ergebnisraum ein: Das Ereignis

$$T_m := \{(\omega_1, \dots, \omega_n) \in \Omega : \omega_m = 1, \omega_j \text{ beliebig für } j \neq m\}$$

beschreibt den Treffer im m -ten Einzelexperiment, und wir wissen, dass $P(T_m) = \hat{p}$. Das Ereignis

$$N_m := \{(\omega_1, \dots, \omega_n) \in \Omega : \omega_m = 0, \omega_j \text{ beliebig für } j \neq m\}$$

beschreibt die Niete im m -ten Versuch und hat die Wahrscheinlichkeit $P(N_m) = 1 - \hat{p}$.

Wir rechnen jetzt die Wahrscheinlichkeit für genau r Treffer bei einem Elementarereignis $\{(a_1, \dots, a_n)\}$ aus. Die Indexmenge $J \subset \{1, \dots, n\}$ enthalte genau die Indizes von a mit $a_j = 1$, d.h. $a_j = 0$ für $j \notin J$. Es ist $\text{card}(J) = r$. Damit können wir schreiben:

$$\{(a_1, \dots, a_n)\} = \bigcap_{j \in J} T_j \cap \bigcap_{j \notin J} N_j.$$

Da die Würfe stochastisch unabhängig sind, so folgt

$$P(a_1, \dots, a_n) = \prod_{j \in J} P(T_j) \cdot \prod_{j \notin J} P(N_j) = \hat{p}^r (1 - \hat{p})^{n-r}.$$

Jetzt betrachten wir die Menge

$$A_r = \{a = (a_1, \dots, a_n) : a \text{ enthält genau } r \text{ Treffer}\}$$

aller Ergebnisse mit genau r Treffern. Es ist $\text{card}(A_r) = \binom{n}{r}$, da es $\binom{n}{r}$ Kombinationen gibt, r -mal "1" und $(n-r)$ -mal "0" auf die n Versuche zu verteilen. Damit ergibt sich

$$P(A_r) = \binom{n}{r} \hat{p}^r (1 - \hat{p})^{n-r}.$$

□

Bemerkung: Beachte, dass aus den binomischen Formeln folgt

$$\sum_{r=0}^n \binom{n}{r} \hat{p}^r (1 - \hat{p})^{n-r} = (\hat{p} + (1 - \hat{p}))^n = 1.$$

Beispiel 5.15 Wir würfeln 10-mal und fragen nach der Wahrscheinlichkeit, dass genau 4-mal die "Eins" fällt. Setze $\hat{p} = \frac{1}{6}$, $r = 4$ und $n = 10$. Mit Satz 5.14 folgt

$$p = \binom{10}{4} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^6 \approx 0,054.$$

5.2 Zufallsvariablen und Verteilungsfunktionen

Definition 5.16 Sei ein Zufallsexperiment mit Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) gegeben. Eine **Zufallsvariable** ist eine Abbildung $X : \Omega \rightarrow \mathbb{R}$, für die gilt: $\{\omega \in \Omega : X(\omega) \leq r\} \in \mathcal{A}$ für jedes $r \in \mathbb{R}$.

Also müssen die Urbilder der Intervalle $(-\infty, r] \subset \mathbb{R}$ alle zur σ -Algebra \mathcal{A} gehören – und damit ein Wahrscheinlichkeitsmaß besitzen. Aus den Eigenschaften einer σ -Algebra folgt schnell, dass auch $\{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{A}$ für jedes Intervall $I \subseteq \mathbb{R}$ ein Wahrscheinlichkeitsmaß besitzt, für das wir abkürzend $P(X \in I)$ schreiben, also

$$P(X \in I) = P(\{\omega \in \Omega : X(\omega) \in I\}).$$

In diesem Sinne sind auch die Notationen $P(X = r)$ oder $P(X \leq r)$ zu verstehen für $r \in \mathbb{R}$.

Definition 5.17 Sei $X : \Omega \rightarrow \mathbb{R}$ Zufallsvariable in einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Dann heißt die Abbildung $F : \mathbb{R} \rightarrow [0, 1]$, definiert durch

$$F(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}), \quad x \in \mathbb{R},$$

Verteilungsfunktion von X .

Beispiel 5.18 Sei ein Bernoulli-Experiment gegeben, d.h. ein Zufallsexperiment mit Ergebnismenge

$$\Omega = \{(\omega_1, \dots, \omega_n) : \omega_j \in \{0, 1\}, j = 1, \dots, n\}.$$

Sei weiter $X : \Omega \rightarrow \mathbb{R}$ die Anzahl der Treffer, d.h.

$$X(\omega) = \sum_{j=1}^n \omega_j.$$

Dann ist nach Satz 5.14

$$p_k = P(X = k) = \binom{n}{k} \hat{p}^k (1 - \hat{p})^{n-k},$$

für $k = 0, \dots, n$, wenn \hat{p} die Wahrscheinlichkeit für einen Treffer im Einzelexperiment ist. Die zugehörige Verteilungsfunktion ist

$$F(x) = \begin{cases} 0, & \text{für } x < 0, \\ \sum_{j=0}^k \binom{n}{j} \hat{p}^j (1 - \hat{p})^{n-j}, & \text{für } k \leq x < k + 1 \text{ und } k \in \{0, \dots, n - 1\}, \\ 1, & \text{für } x \geq n. \end{cases}$$

Eine Zufallsvariable mit dieser Verteilung heißt **binomialverteilt**.

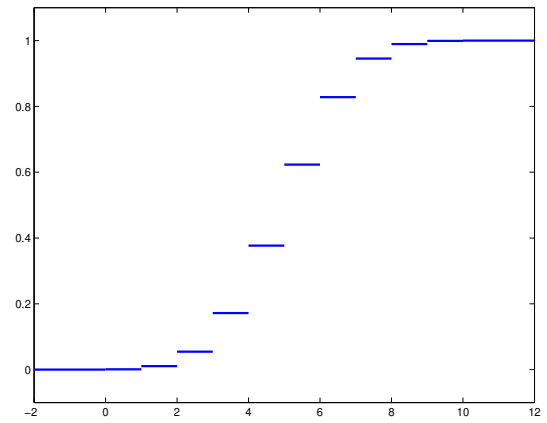


Abbildung 1: Binomialverteilung mit $n = 10$ und $\hat{p} = 0,5$

Satz 5.19 (Eigenschaften von Verteilungsfunktionen)

Sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion zu einer Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$. Dann gilt:

- (a) F ist monoton steigend,
- (b) $P(a < X \leq b) = F(b) - F(a)$ für alle $a < b$,
- (c) $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$,
- (d) F ist rechtsstetig, d.h. $\lim_{x \rightarrow z, x > z} F(x) = F(z)$.

Beweis: (a) Sei $x \leq y$. Dann ist

$$\{\omega \in \Omega : X(\omega) \leq x\} \subseteq \{\omega \in \Omega : X(\omega) \leq y\},$$

also

$$F(x) = P(X \leq x) \leq P(X \leq y) = F(y).$$

(b) Setze $A = \{\omega \in \Omega : X(\omega) \leq a\}$ und $B = \{\omega \in \Omega : X(\omega) \leq b\}$. Dann ist $\{\omega \in \Omega : a < X(\omega) \leq b\} = B \setminus A$ und weiter

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A).$$

Also ergibt sich

$$P(a < X \leq b) = P(B) - P(A) = F(b) - F(a).$$

(c) Sei (x_n) eine strikt monoton fallende Folge mit $x_n \rightarrow -\infty$. Wir zerlegen Ω in

$$\Omega = \bigcup_{n=1}^{\infty} \{\omega \in \Omega : x_{n+1} < X(\omega) \leq x_n\} \cup \{\omega : X(\omega) > x_1\}.$$

Dies ist eine disjunkte Zerlegung, also mit (b)

$$\begin{aligned} 1 &= \sum_{n=1}^{\infty} P(x_{n+1} < X \leq x_n) + P(X > x_1) \\ &= \sum_{n=1}^{\infty} [F(x_n) - F(x_{n+1})] + [1 - F(x_1)]. \end{aligned}$$

Also konvergiert (Teleskopsumme!)

$$F(x_1) - F(x_N) = \sum_{n=1}^{N-1} [F(x_n) - F(x_{n+1})]$$

gegen $F(x_1)$ für $N \rightarrow \infty$. Also konvergent $F(x_N)$ gegen Null, was die erste Aussage beweist.

Für die zweite Aussage betrachten wir eine beliebige strikt monoton steigende Folge (x_n) mit $x_n \rightarrow \infty$. Wir zerlegen jetzt Ω in

$$\Omega = \bigcup_{n=1}^{\infty} \{\omega \in \Omega : x_n < X(\omega) \leq x_{n+1}\} \cup \{\omega : X(\omega) \leq x_1\}.$$

Dies ist eine disjunkte Zerlegung, also mit (b)

$$\begin{aligned} 1 &= \sum_{n=1}^{\infty} P(x_n < X \leq x_{n+1}) + P(X \leq x_1) \\ &= \sum_{n=1}^{\infty} [F(x_{n+1}) - F(x_n)] + F(x_1). \end{aligned}$$

Also konvergiert (Teleskopsumme!)

$$F(x_N) = \sum_{n=1}^{N-1} [F(x_{n+1}) - F(x_n)] + F(x_1)$$

gegen 1 für $N \rightarrow \infty$. Dies beweist die zweite Aussage.

(d) Dies geht so ähnlich wie in Teil (c). Sei (x_n) eine monoton fallende Folge mit $x_n \rightarrow z$ für $n \rightarrow \infty$. Dann ist $F(z) = 1 - P(X > z)$ und wieder

$$\{\omega : X(\omega) > z\} = \bigcup_{n=1}^{\infty} \{\omega : x_{n+1} < X(\omega) \leq x_n\} \cup \{\omega : X(\omega) > x_1\}.$$

Daher ist

$$1 - F(z) = \sum_{n=1}^{\infty} [F(x_n) - F(x_{n+1})] + (1 - F(x_1)),$$

also

$$F(x_1) - F(x_N) = \sum_{n=1}^{N-1} [F(x_n) - F(x_{n+1})] \longrightarrow F(x_1) - F(z).$$

Daher konvergiert $F(x_N)$ gegen $F(z)$. □

Ist der Wertebereich von $X : \Omega \rightarrow \mathbb{R}$ endlich, etwa $W(X) = \{x_1, \dots, x_m\}$, so können wir $F(x)$ schreiben als

$$F(x) = \sum_{j: x_j \leq x} P(X = x_j).$$

Die Werte $f_j = P(X = x_j)$, $j = 1, \dots, m$, heißen *Verteilung* von X . Dies übertragen wir auf den Fall, dass der Wertebereich von X nicht endlich ist:

Definition 5.20 Eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ heißt **stetig verteilt** mit **Dichte** f , wenn ihre Verteilungsfunktion eine Darstellung der Form

$$F(x) = \int_{-\infty}^x f(t) dt \quad \text{für } x \in \mathbb{R}$$

besitzt mit einer integrierbaren, nicht negativen Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$.

Bemerkung:

1. Aus $\lim_{x \rightarrow \infty} F(x) = 1$ folgt für eine Dichte $f : \mathbb{R} \rightarrow \mathbb{R}$ die Eigenschaft $\int_{-\infty}^{\infty} f(t) dt = 1$.
2. Die Verteilungsfunktion F zu einer stetig verteilten Zufallsvariablen ist stetig! Insbesondere ist die Binomialverteilung nicht stetig verteilt (siehe Abbildung 1). Wir sprechen in solchen Fällen von einer **diskreten** Verteilung.
3. Zu einer stetig verteilten Zufallsvariablen gibt es genau eine Dichte. Für stetige Dichten $f, g : \mathbb{R} \rightarrow \mathbb{R}$ zu einer Verteilungsfunktion F sehen wir dies, wenn wir etwa $f \geq g$ fast überall auf einem Intervall $[a, b]$ annehmen. Dann folgt

$$\begin{aligned} \int_a^b (f(t) - g(t)) dt &= \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt - \int_{-\infty}^b g(t) dt + \int_{-\infty}^a g(t) dt \\ &= F(b) - F(a) - F(b) + F(a) = 0. \end{aligned}$$

Also ist $f = g$ auf $[a, b]$. Analog gilt dies, falls $f \leq g$ auf $[a, b]$ ist. Der Beweis der allgemeinen Aussage erfordert eine genauere Betrachtung, da man anstelle stetiger Dichten auf Intervallen integrierbare Funktionen auf *messbaren Mengen* betrachten muss.

Beispiel 5.21 (a) Sei

$$f(t) = \begin{cases} \frac{1}{b-a}, & \text{für } t \in [a, b] \\ 0, & \text{sonst.} \end{cases}$$

Dann heißt eine stetig verteilte Zufallsvariable X mit dieser Dichte f **gleichmäßig verteilt** auf $[a, b]$.

Es ist

$$F(x) = \begin{cases} 0, & \text{für } x < a, \\ \frac{1}{b-a}(x-a), & \text{für } a \leq x \leq b, \\ 1, & \text{für } x \geq b. \end{cases}$$

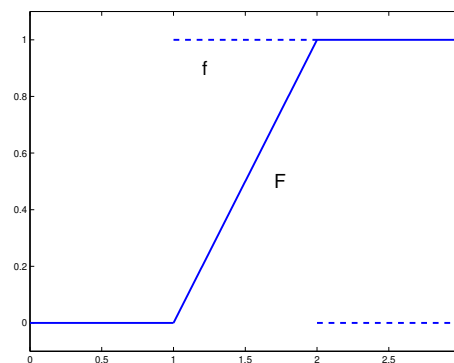


Abbildung 2: Gleichmäßige Verteilung mit $[a, b] = [1, 2]$

(b) Sei $\mu \in \mathbb{R}$ und $\sigma > 0$. Eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ heißt **normalverteilt**, wenn X stetig verteilt ist mit Dichte

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2}, \quad t \in \mathbb{R}.$$

Auf die Größen μ und σ kommen wir später noch zurück. Im Spezialfall $\mu = 0$, $\sigma = 1$ sprechen wir von der **Standardnormalverteilung** mit der Verteilungsfunktion

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt.$$

Mit $u = \frac{t-\mu}{\sigma}$ folgt für die allgemeine Verteilungsfunktion

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{1}{2}u^2} du = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

d.h. eine Normalverteilung lässt sich stets aus der Standardnormalverteilung berechnen.

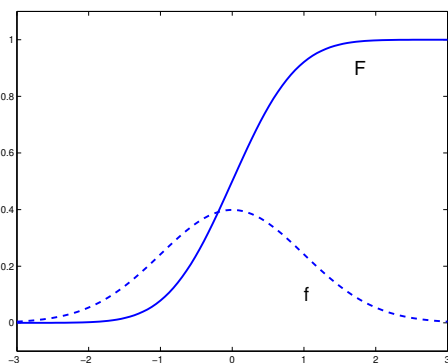


Abbildung 3: Standardnormalverteilung

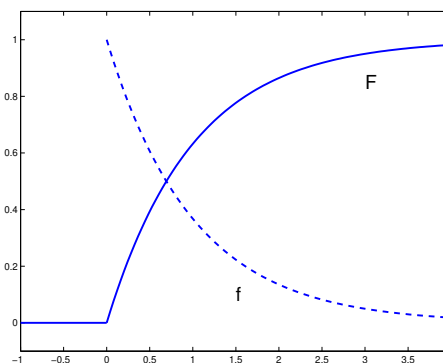


Abbildung 4: Exponentialverteilung mit $\lambda = 1$

(c) Eine stetig verteilte Zufallsvariable heißt **exponentiell verteilt**, wenn

$$f(t) = \begin{cases} 0, & \text{für } t < 0, \\ \lambda e^{-\lambda t}, & \text{für } t \geq 0, \end{cases}$$

mit einem $\lambda > 0$ gilt. Die zugehörige Verteilungsfunktion ist

$$F(x) = \begin{cases} 0, & \text{für } x < 0, \\ 1 - e^{-\lambda x}, & \text{für } x \geq 0. \end{cases}$$

Zu der Verteilung einer Zufallsvariablen gibt es charakterisierende Kennzahlen, von denen wir hier die wichtigsten zwei einführen. Zunächst betrachten wir eine Art ‘‘Mittelwert’’ von $\{X(\omega) : \omega \in \Omega\}$, gewichtet mit den Wahrscheinlichkeiten.

Definition 5.22 Sei $X : \Omega \rightarrow \mathbb{R}$ Zufallsvariable eines Zufallsexperiments mit Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

(a) Im Fall einer diskreten, endlichen Ergebnismenge $\Omega = \{\omega_1, \dots, \omega_n\}$ heißt

$$E(X) = \sum_{j=1}^n P(\omega_j) X(\omega_j)$$

Erwartungswert von X .

(b) Im Fall einer stetig verteilten Zufallsvariable mit Dichte $f : \mathbb{R} \rightarrow \mathbb{R}$ ist

$$E(X) = \int_{-\infty}^{\infty} t f(t) dt$$

der **Erwartungswert** von X , wenn das Integral existiert.

Bemerkungen: (a) Entsprechend wird der Erwartungswert auch bei diskreten, abzählbaren Ergebnismengen durch

$$E(X) = \sum_{j=1}^{\infty} P(\omega_j) X(\omega_j)$$

festgelegt, wenn die Reihe absolut konvergiert.

(b) Bei einer endlichen Ergebnismenge gilt

$$E(X) = \sum_{j=1}^n P(\omega_j) X(\omega_j) = \sum_{i=1}^m x_i P(X = x_i),$$

wenn mit $\{x_1, \dots, x_m\}$ die Wertemenge der Zufallsvariablen X bezeichnet wird. In der letzten Darstellung sehen wir die Analogie zum kontinuierlichen Fall. Um diese Identität zu zeigen, definieren wir die Mengen $A_i = \{\omega \in \Omega : X(\omega) = x_i\}$. Dann ist

$$P(X = x_i) = P(A_i) = \sum_{\omega_j \in A_i} P(\omega_j).$$

Da $\Omega = \bigcup_{i=1}^m A_i$ gilt und die Mengen A_i paarweise disjunkt sind, folgt

$$\begin{aligned} E(X) &= \sum_{j=1}^n P(\omega_j) X(\omega_j) \\ &= \sum_{i=1}^m \sum_{\omega_j \in A_i} P(\omega_j) X(\omega_j) \\ &= \sum_{i=1}^m x_i \sum_{\omega_j \in A_i} P(\omega_j) = \sum_{i=1}^m x_i P(X = x_i). \end{aligned}$$

Beispiele 5.23 1. Bei einem Laplace-Experiment, d.h. $\Omega = \{\omega_1, \dots, \omega_n\}$ und $P(\omega_j) = \frac{1}{n}$ für $j = 1, \dots, n$ gilt

$$E(X) = \frac{1}{n} \sum_{j=1}^n X(\omega_j).$$

Also ist der Erwartungswert in diesem Fall das arithmetische Mittel der Werte $X(\omega_j)$, $j = 1, \dots, n$, die die Zufallsvariable annimmt.

2. Beim Bernoulli-Experiment mit Wahrscheinlichkeit \hat{p} und n Wiederholungen gilt für die Zufallsvariable X mit

$$X(\omega) = \sum_{j=1}^n \omega_j, \quad \omega \in \Omega,$$

(d.h. die Anzahl der Treffer im Ergebnis ω) die Wahrscheinlichkeit $P(X = j) = \binom{n}{j} \hat{p}^j (1-\hat{p})^{n-j}$ nach Satz 5.14. Also ergibt sich für den Erwartungswert

$$\begin{aligned} E(X) &= \sum_{j=0}^n j \binom{n}{j} \hat{p}^j (1-\hat{p})^{n-j} \\ &= n \hat{p} \sum_{j=1}^n \frac{(n-1)!}{(j-1)!(n-j)!} \hat{p}^{j-1} (1-\hat{p})^{(n-1)-(j-1)} \\ &= n \hat{p} (\hat{p} + (1-\hat{p}))^{n-1} = n \hat{p}. \end{aligned}$$

3. Ist $X : \Omega \rightarrow \mathbb{R}$ gleichmäßig verteilt über $I = [a, b]$, so folgt

$$E(X) = \int_{-\infty}^{\infty} t f(t) dt = \frac{1}{b-a} \int_a^b t dt = \frac{a+b}{2}.$$

Also ist der Erwartungswert der Mittelpunkt des Intervalls $[a, b]$.

4. Sei $X : \Omega \rightarrow \mathbb{R}$ normalverteilt. Dann erhalten wir mit der Substitution $u = \frac{t-\mu}{\sigma\sqrt{2}}$ den Erwartungswert

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = \frac{\sigma\sqrt{2}}{\sqrt{\pi}} \underbrace{\int_{-\infty}^{\infty} u e^{-u^2} du}_{=0} + \frac{\mu}{\sqrt{\pi}} \underbrace{\int_{-\infty}^{\infty} e^{-u^2} du}_{=\sqrt{\pi}} = \mu.$$

Dies ist die Bedeutung des Parameters μ in der Normalverteilung!

5. Für eine exponentiell verteilte Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ gilt

$$E(X) = \lambda \int_0^{\infty} t e^{-\lambda t} dt = -t e^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}.$$

Im Umgang mit Erwartungswerten sind die folgenden Aussagen nützlich.

Satz 5.24 (a) Für eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ und eine (stückweise) stetige Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ gilt für die Zufallsvariable $h(X) = h \circ X : \Omega \rightarrow \mathbb{R}$, also der Verkettung,

$$E(h(X)) = \sum_{j=1}^n h(X(\omega_j)) P(\omega_j) = \sum_{i=1}^m h(x_i) P(X = x_i),$$

wenn Ω endlich ist, und

$$E(h(X)) = \int_{-\infty}^{\infty} h(t) f(t) dt,$$

wenn X stetig verteilt ist mit Dichte $f : \mathbb{R} \rightarrow \mathbb{R}$.

(b) Sind $X, Y : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen mit $X(\omega) \leq Y(\omega)$ für alle $\omega \in \Omega$, so gilt $E(X) \leq E(Y)$, d.h. die Bildung des Erwartungswerts ist monoton.

(c) Sind $X, Y : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen und $\lambda \in \mathbb{R}$ eine Konstante, so gilt

(i) $E(\lambda X) = \lambda E(X)$,

(ii) $E(X + Y) = E(X) + E(Y)$ (d.h. zusammen mit (i), dass E eine lineare Abbildung ist),

(iii) $E(1) = 1$, d.h. die konstante Funktion $\omega \mapsto 1$ hat Erwartungswert Eins.

Beweis: Im endlichen Fall lassen sich die Aussagen leicht prüfen. Für den kontinuierlichen Fall verweisen wir auf die Literatur, da in der angegebenen Allgemeinheit weitere Begriffe und Ergebnisse aus der Maßtheorie erforderlich wären. So haben wir etwa den Erwartungswert nur für stetig verteilte Zufallsvariablen definiert. Somit beinhaltet der Satz eine Verallgemeinerung unseres Begriffs, da wir im Allgemeinen nicht garantieren können, dass $h \circ X : \Omega \rightarrow \mathbb{R}$ stetig verteilt ist.

Offensichtlich lässt sich der Erwartungswert einer Zufallsvariablen als „Schwerpunkt“ einer Verteilung auffassen. Eine weitere wichtige Kennzahl ist die *Varianz*, die wir als Maß für die Streuung der Ergebnisse um diesen Schwerpunkt verstehen können.

Definition 5.25 Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Dann heißt

$$\text{Var}(X) = E([X - E(X)]^2)$$

Varianz von X , also der Erwartungswert der Zufallsvariablen $\omega \mapsto (X(\omega) - E(X))^2$.

Bemerkungen:

1. Die Zahl $\sqrt{\text{Var}(X)}$ heißt **Standardabweichung**.

2. Beachte, dass die Varianz nach Satz 5.24 wohldefiniert ist, wobei $h : \mathbb{R} \rightarrow \mathbb{R}$ mit $h(t) = (t - E(X))^2$ gewählt ist.
3. Interpretation: Ein relativ kleiner Wert für die Varianz bedeutet, dass im Mittel die Abweichung vom Erwartungswert klein ist. Bei einer großen Varianz dagegen werden die Ergebnisse im Durchschnitt stark um den Erwartungswert streuen.
4. Aus $[X - E(X)]^2 = X^2 - 2X \cdot E(X) + E(X)^2$ folgt mit Satz 5.24 die nützliche Darstellung

$$\text{Var}(X) = E(X^2) - (E(X))^2.$$

5. $\text{Var}(\lambda X) = E(\lambda^2[X - E(X)]^2) = \lambda^2 \text{Var}(X)$.

Beispiel 5.26 1. Ist eine Zufallsvariable $X : \Omega = \{0, 1\}^n \rightarrow \mathbb{R}$ binomialverteilt, so erhalten wir (s. Beispiel 5.21 (b) und obige Bemerkung)

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 \\ &= \sum_{k=0}^n k^2 P(X = k) - (n\hat{p})^2 \\ &= \sum_{k=0}^n k^2 \binom{n}{k} \hat{p}^k (1 - \hat{p})^{n-k} - (n\hat{p})^2 \\ &= n\hat{p} \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} k \hat{p}^{k-1} (1 - \hat{p})^{n-k} - (n\hat{p})^2 \\ &= n\hat{p} \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} (k-1) \hat{p}^{k-1} (1 - \hat{p})^{n-k} \\ &\quad + n\hat{p} \sum_{k=1}^n \frac{(n-1)!}{(n-k)!(k-1)!} \hat{p}^{k-1} (1 - \hat{p})^{n-k} - (n\hat{p})^2 \\ &= n\hat{p}(n-1)\hat{p} \sum_{k=2}^n \frac{(n-2)!}{(n-k)!(k-2)!} \hat{p}^{k-2} (1 - \hat{p})^{n-k} \\ &\quad + n\hat{p} \sum_{k=0}^{n-1} \binom{n-1}{k} \hat{p}^k (1 - \hat{p})^{(n-1)-k} - (n\hat{p})^2 \\ &= n\hat{p}(n-1)\hat{p} \sum_{k=0}^{n-2} \binom{n-2}{k} \hat{p}^k (1 - \hat{p})^{(n-2)-k} + n\hat{p}(\hat{p} + (1 - \hat{p}))^{n-1} - (n\hat{p})^2 \\ &= n\hat{p}(n-1)\hat{p}(\hat{p} + (1 - \hat{p}))^{n-2} + n\hat{p} - (n\hat{p})^2 = n\hat{p}(1 - \hat{p}). \end{aligned}$$

2. Sei $X : \Omega \rightarrow \mathbb{R}$ normalverteilt. Dann ergibt sich mit der Substitution $u = \frac{t-\mu}{\sigma}$ und partieller

Integration

$$\begin{aligned}\operatorname{Var}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (t - \mu)^2 e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-\frac{1}{2}u^2} du \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \left(-u e^{-\frac{1}{2}u^2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du \right) = \sigma^2.\end{aligned}$$

Dies ist die Interpretation des Parameters σ in der Normalverteilung!

Unter Ausnutzung der Monotonie des Erwartungswerts können wir folgende Abschätzung für die Abweichung vom Erwartungswert angeben.

Satz 5.27 (*Tschebyscheff Ungleichung*)

Ist $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable, so gilt für jedes $\varepsilon > 0$:

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\operatorname{Var}(X)}{\varepsilon^2}.$$

Beweis: Setze

$$h(t) = \frac{(t - E(X))^2}{\varepsilon^2}$$

und

$$g(t) = \begin{cases} 1 & \text{für } |t - E(X)| \geq \varepsilon, \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt $g(t) \leq h(t)$ und wir erhalten mit Satz 5.24 (b)

$$E(g(X)) \leq E(h(X)).$$

Da $E(h(X)) = \frac{\operatorname{Var}(X)}{\varepsilon^2}$ und $E(g(X)) = P(\{\omega \in \Omega : |X - E(X)| \geq \varepsilon\})$ gilt, ergibt sich die Aussage des Satzes. \square

5.3 Folgen von Zufallsvariablen

Häufig sind zu einem Zufallsexperiment mehrere Zufallsvariablen zu betrachten. Sei im Folgenden (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X_i : \Omega \rightarrow \mathbb{R}$, $i \in I$, Zufallsvariable, wobei $I = \{1, \dots, n\}$ endlich sein kann oder $I = \mathbb{N}$, also (abzählbar) unendlich.

Wir haben schon gesehen (Satz 5.24, (c)), dass der Erwartungswert linear ist, also $E(X + Y) = E(X) + E(Y)$ und $E(\lambda X) = \lambda E(X)$. Für die Varianz der Summe zweier Zufallsvariablen ist der Zusammenhang etwas komplizierter.

Satz 5.28 Seien $X, Y : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen zu denen Erwartungswert und Varianz existieren, dann existiert auch $\operatorname{Var}(X + Y)$ und es gilt

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2(E(XY) - E(X)E(Y)).$$

Beweis: Wegen der Identität $\text{Var}(X) = E(X^2) - (E(X))^2$ und Satz 5.24 folgt

$$\begin{aligned}\text{Var}(X + Y) &= E((X + Y)^2) - (E(X + Y))^2 \\ &= E(X^2 + 2XY + Y^2) - (E(X))^2 - 2E(X)E(Y) - (E(Y))^2 \\ &= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 + 2E(XY) - 2E(X)E(Y) \\ &= \text{Var}(X) + \text{Var}(Y) + 2(E(XY) - E(X)E(Y)).\end{aligned}$$

□

Im Allgemeinen ist die Varianz der Summe zweier Zufallsvariablen also nicht gleich der Summe ihrer Varianzen. Der Korrekturterm ist eine wichtige weitere Größe in der Stochastik.

Definition 5.29 Seien X, Y Zufallsvariablen, für die $\text{Var}(X)$ und $\text{Var}(Y)$ existieren. Dann heißt

$$\begin{aligned}\text{Cov}(X, Y) &= E((X - E(X))(Y - E(Y))) \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

die **Kovarianz** von X und Y . Wenn $\text{Var}(X), \text{Var}(Y) > 0$ sind, wird der Quotient

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Korrelationskoeffizient genannt. Zwei Zufallsvariablen X, Y heißen **unkorreliert**, wenn $\text{Cov}(X, Y) = 0$ gilt, d.h. $E(XY) = E(X)E(Y)$.

Bemerkung: Eine positive Kovarianz zu zwei Zufallsvariablen X und Y bedeutet, dass im Mittel $X(\omega)$ groß/klein ist, wenn $Y(\omega)$ groß/klein ist. $\text{Cov}(X, Y) < 0$ besagt, dass öfter Ergebnisse eintreten, bei denen $X(\omega)$ groß und $Y(\omega)$ klein ist bzw. umgekehrt. Wenn X, Y unkorreliert sind, lässt sich ein solcher Zusammenhang zwischen X und Y nicht beobachten.

Wir vergleichen die Unkorreliertheit mit der stochastischen Unabhängigkeit von Zufallsvariablen. Die stochastische Unabhängigkeit haben wir bisher nur für Ereignisse definiert (siehe Definition 5.13).

Definition 5.30 Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Die Zufallsvariable $X_j : \Omega \rightarrow \mathbb{R}$ für $j = 1, \dots, n$, heißen **stochastisch unabhängig** wenn

$$P(X_{j_1} \in I_{j_1}, \dots, X_{j_m} \in I_{j_m}) = P(X_{j_1} \in I_{j_1}) \cdots P(X_{j_m} \in I_{j_m})$$

gilt für alle beliebigen Indermengen $\{j_1, \dots, j_m\} \subseteq \{1, \dots, n\}$ und alle Intervalle $I_j \subset \mathbb{R}$.

Satz 5.31 Sind X, Y stochastisch unabhängige Zufallsvariablen, so sind X, Y unkorreliert, d.h. insbesondere

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Beweis für endlich verteilte Zufallsvariable: Die Wertebereiche $W(X)$ und $W(Y)$ seien gegeben durch $W(X) = \{x_1, \dots, x_n\}$ und $W(Y) = \{y_1, \dots, y_m\}$. Dann ist $W(X \cdot Y) = \{x_i y_j : i =$

$1, \dots, n, j = 1, \dots, m\} = \{f_k : k = 1, \dots, p\}$. Dann ist $E(X \cdot Y) = \sum_{k=1}^p f_k P(X \cdot Y = f_k)$. Sei k fest, $K_k = \{(i, j) : x_i y_j = f_k\}$. Dann ist

$$\{\omega \in \Omega : X(\omega)Y(\omega) = f_k\} = \bigcup_{(i,j) \in K_k} \{\omega \in \Omega : X(\omega) = x_i \text{ und } Y(\omega) = y_j\}$$

und die Vereinigung ist disjunkt. Daher folgt

$$\begin{aligned} E(X \cdot Y) &= \sum_{k=1}^p f_k \sum_{(i,j) \in K_k} \underbrace{P(X = x_i \text{ und } Y = y_j)}_{= P(X=x_i) P(Y=y_j)} = \sum_{i=1}^n \sum_{j=1}^m x_i y_j P(X = x_i) P(Y = y_j) \\ &= E(X) E(Y), \end{aligned}$$

wegen der Unabhängigkeit. □

Die Kovarianz hat die folgenden Eigenschaften:

Satz 5.32 (a) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$,

(b) $\text{Cov}(X, X) = \text{Var}(X)$,

(c) $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$

und allgemeiner:

$$\text{Cov}\left(\sum_{j=1}^n a_j X_j, \sum_{k=1}^m b_k Y_k\right) = \sum_{j=1}^n \sum_{k=1}^m a_j b_k \text{Cov}(X_j, Y_k).$$

Beweis der letzten Formel:

$$\begin{aligned} \text{Cov}\left(\sum_{j=1}^n a_j X_j, \sum_{k=1}^m b_k Y_k\right) &= E\left(\sum_{j,k} a_j b_k X_j Y_k\right) - E\left(\sum_j a_j X_j\right) E\left(\sum_k b_k Y_k\right) \\ &= \sum_{j,k} a_j b_k E(X_j Y_k) - \sum_{j,k} a_j b_k E(X_j) E(Y_k) \\ &= \sum_{j,k} a_j b_k \text{Cov}(X_j, Y_k) \end{aligned}$$

wegen der Linearität des Erwartungswerts und der Definition der Kovarianz. □

Beispiel 5.33 Beim zweimaligen Würfeln seien die Zufallsvariablen definiert durch $X(i, j) = i$ und $Y(i, j) = j$, $i, j = 1, \dots, 6$. Die Zufallsvariablen X und Y sind stochastisch unabhängig (man zeige dies!). Wir betrachten jetzt die Zufallsvariable $X + Y$ und $X - Y$, d.h.

$$(X + Y)(i, j) = i + j \quad \text{und} \quad (X - Y)(i, j) = i - j.$$

$X + Y, X - Y$ sind nicht stochastisch unabhängig, denn z.B. ist

$$P(X + Y = 12, X - Y = 0) = \frac{1}{36}, \quad P(X + Y = 12) = \frac{1}{36}, \quad P(X - Y = 0) = \frac{6}{36},$$

und es gilt nicht $P(X + Y = 12, X - Y = 0) = P(X + Y = 12) P(X - Y = 0)$. Aber die Zufallsvariablen sind unkorreliert, denn nach dem letzten Satz ist

$$\begin{aligned} \text{Cov}(X + Y, X - Y) &= \text{Cov}(X, X) + \text{Cov}(Y, X) - \text{Cov}(X, Y) - \text{Cov}(Y, Y) \\ &= \text{Var}(X) - \text{Var}(Y) = 0. \end{aligned}$$

Jetzt betrachten wir Folgen $X_j : \Omega \rightarrow \mathbb{R}$, $j = 1, 2, \dots$, von Zufallsvariablen, z.B. durch Wiederholen eines Experiments. Was passiert für große n ?

Satz 5.34 (*schwaches Gesetz der großen Zahlen*)

Die Zufallsvariablen $X_j : \Omega \rightarrow \mathbb{R}$, $j = 1, 2, \dots$, seien stochastisch unabhängig (d.h. jede endliche Zahl ist stochastisch unabhängig im Sinn von Definition 5.30). Außerdem seien die Erwartungswerte gleich, etwa $E(X_j) = \mu$ für alle j und auch die Varianzen seien gleich. Dann gilt:

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{j=1}^n X_j - \mu \right| \geq \varepsilon \right) = 0 \quad \text{für jedes } \varepsilon > 0.$$

Beweis: Die Zufallsvariable $\frac{1}{n} \sum_{j=1}^n X_j$ hat ebenfalls den Erwartungswert μ , denn wegen der Linearität des Erwartungswerts ist

$$E \left(\frac{1}{n} \sum_{j=1}^n X_j \right) = \frac{1}{n} \sum_{j=1}^n E(X_j) = \frac{1}{n} \sum_{j=1}^n \mu = \mu.$$

Für die Varianz gilt wegen der stochastische Unabhängigkeit nach Satz 5.31 und nach der Bemerkung (e) nach der Definition der Varianz:

$$\text{Var} \left(\frac{1}{n} \sum_{j=1}^n X_j \right) = \frac{1}{n^2} \sum_{j=1}^n \text{Var}(X_j) = \frac{1}{n} \text{Var}(X_1).$$

Die Tschebyscheffsche Ungleichung von Satz 5.27 liefert

$$P \left(\left| \frac{1}{n} \sum_{j=1}^n X_j - \mu \right| \geq \varepsilon \right) \leq \frac{\frac{1}{n} \text{Var}(X_1)}{\varepsilon^2} \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Über die Vorlesung hinaus geht der Beweis der beiden folgenden „zentralen“ Sätze:

Satz 5.35 (*starkes Gesetz der großen Zahlen*)

Unter denselben Voraussetzungen wie in Satz 5.34 gilt:

$$P \left(\left\{ \omega : \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n X_j(\omega) \neq \mu \right\} \right) = 0.$$

Satz 5.36 (Zentraler Grenzwertsatz)

Die Zufallsvariablen $X_j : \Omega \rightarrow \mathbb{R}$, $j = 1, 2, \dots$, seien stochastisch unabhängig und identisch verteilt (d.h. sie haben alle dieselbe Verteilungsfunktion). Mit dem Erwartungswert, $\mu = E(X_j)$, und der Varianz, $\sigma^2 = \text{Var}(X_j)$, für jedes j sei

$$\bar{Y}^{(n)} := \frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X}^{(n)} - \mu}{\sigma} \sqrt{n}, \quad n \in \mathbb{N},$$

mit $\bar{X}^{(n)} = \frac{1}{n} \sum_{j=1}^n X_j$. Dann konvergiert die Verteilungsfunktion von $\bar{Y}^{(n)}$ gegen die Normalverteilung, d.h.

$$\lim_{n \rightarrow \infty} P(\bar{Y}^{(n)} \leq x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, \quad x \in \mathbb{R}.$$

Übersicht: Parametrisierungen

Mathematische Beschreibungen, insbesondere Parametrisierungen, von geometrischen Objekten werden oft genutzt, z.B. bei der Berechnung von Volumen-, Kurven-, oder Flächenintegralen. Wir stellen hier einige Möglichkeiten zusammen.

Gebiete

Gebiete im \mathbb{R}^n sind häufig durch (nichtlineare) Ungleichungen beschrieben.

Wichtige Beispiele:

Menge	Beschreibung
Normalgebiet im \mathbb{R}^2 mit $\varphi, \psi : [a, b] \rightarrow \mathbb{R}$ (Spezialfall: abgeschlossene Kreisscheibe um $(0, 0)^\top$)	$G = \{x \in \mathbb{R}^2 : \varphi(x_1) \leq x_2 \leq \psi(x_1), x_1 \in [a, b]\}$ bzw. $G = \{x \in \mathbb{R}^2 : (x_2) \leq x_1 \leq g(x_2), x_2 \in [a, b]\}$ $S = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = r^2\}$ $= \{x \in \mathbb{R}^2 : -\sqrt{r^2 - x_1^2} \leq x_2 \leq \sqrt{r^2 - x_1^2}, x_1 \in [-r, r]\}$
Offene Kugel um $p \in \mathbb{R}^3$ mit Radius $R > 0$	$K = \{x \in \mathbb{R}^3 : \ x - p\ < R\}$ $= \{\Psi(r, \varphi, \theta)\}$ $= p + r (\cos \varphi \sin \theta, \sin \varphi \sin \theta, \cos \theta)^\top :$ $r \in [0, R), \varphi \in [0, 2\pi), \theta \in [0, \pi]$

Kurven

Kurven im \mathbb{R}^n sind durch eine stetige Funktion

$$t \mapsto \varphi(t) \in \mathbb{R}^n \quad \text{mit } a \leq t \leq b$$

also durch **einen** Parameter t gegeben.

Wichtige Beispiele:

Menge	Parametrisierung
Gerade zwischen $p, q \in \mathbb{R}^n$	$\varphi(t) = (1 - t)p + tq = p + t(q - p) \in \mathbb{R}^n,$ $t \in [0, 1]$
Graph einer Funktion $f : [a, b] \rightarrow \mathbb{R}$	$G = \{(x, y) \in \mathbb{R}^2 : y = f(x), x \in [a, b]\}$ $= \{\varphi(t) = (t, f(t))^\top : t \in [a, b]\}$
Ellipse im \mathbb{R}^2 mit Halbachsen $a, b > 0$ um Zentrum $p \in \mathbb{R}^2$ (Spezialfall: Kreis mit Radius $r > 0$)	$E = \{x \in \mathbb{R}^2 : \left(\frac{x_1 - p_1}{a}\right)^2 + \left(\frac{x_2 - p_2}{b}\right)^2 = 1\}$ $= \{\varphi(t) = p + (a \cos t, b \sin t)^\top : t \in [0, 2\pi]\}$ falls $r = a = b$ gilt

Flächen

Flächen im \mathbb{R}^3 sind durch eine stetige Funktion

$$(u, v) \mapsto X(u, v) \in \mathbb{R}^3$$

mit **zwei** Parametern $(u, v)^\top \in D \subseteq \mathbb{R}^2$ parametrisiert.

Wichtige Beispiele:

Menge	Parametrisierung
Ebene im \mathbb{R}^3	Parameterform mit $p^{(1)}, p^{(2)}, p^{(3)} \in E$: $E = \{X(t, s) = p^{(1)} + t(p^{(2)} - p^{(1)}) + s(p^{(3)} - p^{(1)}) \in \mathbb{R}^3 : t, s \in \mathbb{R}\}$ Normalform: $E = \{x \in \mathbb{R}^3 : n \cdot x = n_1x_1 + n_2x_2 + n_3x_3 = \gamma\}$
Graph einer Funktion $f : D \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$	$G = \{(x, y, z) \in \mathbb{R}^3 : z = f(x, y), (x, y)^\top \in D\}$ $= \{X(u, v) = (u, v, f(u, v))^\top : (u, v)^\top \in D\}$
Rotationsfläche um x -Achse erzeugt durch $f : [a, b] \rightarrow \mathbb{R}$ (Spezialfall: Zylindermantel mit $r > 0$)	$F = \{X(x, t) = (x, f(x) \cos t, f(x) \sin t)^\top : x \in [a, b], t \in [0, 2\pi]\}$ falls $f(x) = r$ konstant)
Ellipsoid im \mathbb{R}^3 mit Halbachsen $a, b, c > 0$ um Zentrum $p \in \mathbb{R}^3$	$E = \{x \in \mathbb{R}^3 : \left(\frac{x_1 - p_1}{a}\right)^2 + \left(\frac{x_2 - p_2}{b}\right)^2 + \left(\frac{x_3 - p_3}{c}\right)^2 = 1\}$ $= \{X(\varphi, \theta) = p + (a \cos \varphi \sin \theta, b \sin \varphi \sin \theta, c \cos \theta)^\top : \varphi \in [0, 2\pi), \theta \in [0, \pi]\}$
(Spezialfall: Sphäre mit Radius $r > 0$)	falls $r = a = b = c$

Übersicht: Integralbegriffe

Gebietsintegrale

Sei $D \subset \mathbb{R}^2$ ein Gebiet und $f : D \rightarrow \mathbb{R}$ eine integrierbare Funktion. Dann schreiben wir das **Gebietsintegral** von f über D als (vgl. Abschnitt 2.1)

$$\iint_D f(x) dx \quad \text{oder} \quad \iint_D f(x_1, x_2) d(x_1, x_2).$$

Ist D ein Rechteck, also $D = I \times J$ mit Intervallen $I, J \subset \mathbb{R}$, so kann das Gebietsintegral leicht über das **iterierte Integral** ausgerechnet werden (vgl. Abschnitt 2.2):

$$\iint_D f(x_1, x_2) d(x_1, x_2) = \int_I \int_J f(x_1, x_2) dx_2 dx_1 = \int_J \int_I f(x_1, x_2) dx_1 dx_2.$$

Für ein komplizierteres Gebiet, etwa

$$D = \{x \in \mathbb{R}^2 : \varphi(x_1) \leq x_2 \leq \psi(x_1), a \leq x_1 \leq b\},$$

gilt die analoge Formel

$$\iint_D f(x_1, x_2) d(x_1, x_2) = \int_a^b \int_{\varphi(x_1)}^{\psi(x_1)} f(x_1, x_2) dx_2 dx_1.$$

In anderen Fällen wendet man die **Transformationsformel** an: Sei $B \subseteq \mathbb{R}^2$ offen und $\Psi : B \rightarrow D$ stetig differenzierbar (dazu kommen noch weitere Voraussetzungen, vgl. Satz 2.12). Dann ist

$$\iint_D f(y) dy = \iint_B f(\Psi(x)) |\det \Psi'(x)| dx.$$

Analoge Formeln gelten für den 3-dimensionalen Fall (*Volumenintegrale*).

Kurvenintegrale

Sei $x : [a, b] \rightarrow \mathbb{R}^n$ eine Parametrisierung einer glatten Kurve C und $f : C \rightarrow \mathbb{R}$ stetig. Dann ist das **Kurvenintegral** von f über C definiert durch (Definition 3.9)

$$\int_C f(x) dl := \int_a^b f(x(t)) \|\dot{x}(t)\|_2 dt.$$

Für $F : C \rightarrow \mathbb{R}^n$ (also ein Vektorfeld) gibt es verschiedene Kurvenintegrale: Zum einen wird dasselbe Symbol wie oben verwendet, das Ergebnis ist dann der Vektor der komponentenweise gebildeten Integrale (z.B. Schwerpunktberechnung).

Beim **tangential orientierten Kurvenintegral** (Gleichung (3.1)),

$$\int_C F(x) \cdot dl := \int_a^b F(x(t)) \cdot \dot{x}(t) dt,$$

kommt dagegen ein Skalar heraus.

Im 2-dimensionalen Fall wird auch das **normal orientierte Kurvenintegral** (vgl. den Gaußschen Satz 3.19) verwendet:

$$\int_C F(x) \cdot \nu(x) d\ell = \int_a^b F(x(t)) \cdot \begin{pmatrix} \dot{x}_2(t) \\ -\dot{x}_1(t) \end{pmatrix} dt.$$

Die Normale ist hier so gewählt, dass sie von der Richtung aus gesehen, in der die Kurve durchlaufen wird, nach rechts zeigt.

Oberflächenintegrale

Sei nun $X : B \rightarrow \mathbb{R}^3$ Parametrisierung eines reguläres Flächenstück S und $f : S \rightarrow \mathbb{R}$ stetig. Dann ist das **Oberflächenintegral** von f über S definiert durch (Definition 3.14)

$$\iint_S f(x) do := \iint_B f(X(u, v)) \|X_u(u, v) \times X_v(u, v)\| d(u, v),$$

also ein *Gebietsintegral*.

Daneben gibt es für ein Vektorfeld $F : S \rightarrow \mathbb{R}^3$ auch den **Fluss** von F durch S (Definition 3.16):

$$\iint_S F(x) \cdot do := \iint_B F(X(u, v)) \cdot (X_u(u, v) \times X_v(u, v)) d(u, v).$$

Beachte: Bei diesem Integral kommt es auf die Richtung des (nicht-normierten) Normalenvektors $X_u(u, v) \times X_v(u, v)$ an!

Index

- σ -Algebra, 95
- Abbildung, 3
- abgeschlossene Menge, 8
- Ableitung, 3
 - partielle, 10
- Ableitungsmatrix, 12
- Abschluss einer Menge, 8
- Anfangsrandwertproblem, 90
- Arbeit, 59
- Astroide, 54

- Banachscher Fixpunktsatz, 30
- bedingte Wahrscheinlichkeit, 102
- begleitendes Dreibein, 47, 55
- Bernoulli-Experimente, 104
- beschränkte Menge, 9
- Bild, 3
- binomialverteilt, 106
- Binormalenvektor, 55
- Bisektionsverfahren, 36
- Bogenlänge, 53

- Cauchy-Problem, 84
- Cauchy-Schwarzsche Ungleichung, 41
- Charakteristik, 87
- Charakteristikenmethode, 86
- charakteristische Funktion, 39
- charakteristische Polynom, 82
- charakteristisches System, 87
- curl, 73

- Definitionsbereich einer Funktion, 3
- Dichte, 108
- Differential, 3, 15
- Differentialform, 20
- Differentialgleichung
 - elliptische, 82
 - Eulersche, 93
 - hyperbolische, 82
 - parabolische, 82
- Diffusionsgleichung, 81
- Dirichlet-Problem, 92
- Dirichlet-Randbedingung, 92
- Dispersion, 83
- Dissipation, 83
- Divergenz, 65
 - divergenzfrei, 73
- Dreiecksungleichung, 41

- Eigenwert, 82
- einfach zusammenhängende Menge, 76
- Elementarereignisse, 97
- elliptische Differentialgleichung, 82
- elliptische Koordinaten, 46
- Ereignisse, 96
- Ergebnis, 95
- Ergebnismenge, 95
- Erwartungswert, 110
- Eulersche Differentialgleichung, 93
- Eulerscher Multiplikator, 18
- Exakte Differentialgleichung, 17
- exponentiell verteilt, 109
- Extremstelle einer Funktion, 21
- Extremwert einer Funktion, 21

- fast überall, 38
- Feld
 - skalares, 3
- Fläche
 - glatte, 60
- Flächeninhalt, 40, 63
- Flächenintegral, 63, 122
- Fluss eines Vektorfeldes, 64
- Folge
 - Konvergenz, 7
- Funktion, 3
 - skalarwertige, 3
 - vektorwertige, 3
 - Stetigkeit, 7
- Funktionaldeterminante, 45
- Funktionalmatrix, 12

- Gaußscher Integralsatz, 66
- Gebietsintegral, 37, 39, 121
- gleichmäßig verteilt, 108
- Gradient, 10
- Graph, 4
- Greensche Identität
 - erste, 71
 - zweite, 71

- Hauptnormalenvektor, 55
- Helmholtzgleichung, 81, 83

Hesse-Matrix, 16
 hyperbolische Differentialgleichung, 82
 Inhaltsformeln, 68
 innerer Punkt, 8
 Inneres einer Menge, 8
 Integrierbarkeitsbedingung, 17
 Integral
 iteriertes, 42, 121
 Lebesgue, 39
 integrierbar, 39
 integrierender Faktor, 18
 Intervallhalbierungsverfahren, 36
 iteriertes Integral, 121
 Jacobideterminante, 45
 Jacobimatrix, 12
 Kombinationen, 98
 Kombinationen mit Wiederholungen:, 98
 kompakte Menge, 8
 konservative Kraft, 59
 Kontraktion, 30
 Kontraktionssatz, 30
 Konturlinien, 4
 Koordinaten
 elliptische, 46
 Kugelkoordinaten, 47
 Polarkoordinaten, 46
 Zylinderkoordinaten, 48
 Korrelationskoeffizient, 115
 Kovarianz, 115
 Krümmung, 55
 Kugelkoordinaten, 47
 Kurve, 7, 50
 äquivalente, 52
 geschlossene, 50
 glatte, 52
 Länge der, 53
 stückweise glatte, 52
 Kurvenintegral, 57, 121
 normal orientiert, 122
 tangential orientiert, 58, 121
 Lagrange-Funktion, 25
 Lagrangesche Multiplikatorenregel, 26
 Laplace-Experiment, 97
 Laplacegleichung, 81
 Laplaceoperator, 71
 Linearisierung, 15
 Maß, 40
 Masse, 57, 64
 Maximum einer Funktion, 21
 messbare Funktion, 39
 messbare Menge, 39
 Methode der kleinsten Quadrate, 23
 Minimum einer Funktion, 21
 mit Wiederholungen:, 98
 Navier-Stokes Gleichungen, 81
 Newtonverfahren, 33
 Niveaufläche, 61
 Norm, 6
 Definitheit, 6
 Dreiecksungleichung, 6
 Homogenität, 6
 Normalableitung
 äußere, 71
 Normaleneinheitsvektor, 60
 äußerer, 65
 Normalgleichungen, 23
 Normalkomponente, 64
 normalverteilt, 109
 normierter Raum, 6
 Nullmenge, 38
 Oberflächenintegral, 63, 122
 offene Menge, 8
 ohne Wiederholungen:, 98
 orientierte Fläche, 74
 orthogonale Koordinaten, 45
 parabolische Differentialgleichung, 82
 Parameterintervall, 50
 Parametertransformation, 52
 Parametrisierung, 50, 60
 Permutationen, 98
 Permutationen ohne Wiederholungen:, 98
 Polarkoordinaten, 46
 positiv orientiert, 67
 Potential eines Kraftfeldes, 59
 Potential eines Vektorfeldes, 75
 Potentialgleichung, 81
 quasi-lineare Differentialgleichung erster Ordnung, 86
 quasilinear, 86
 quellenfrei, 73
 Rand einer Menge, 8

Randpunkt, 8
 rechte Hand Regel, 74
 Richtungsableitung, 11
 Rotation eines Vektorfeldes, 73, 79
 rotationsfrei, 73
 Rotationskörper, 48

 Satz
 über implizite Funktionen, 20
 Banachscher Fixpunktsatz, 30
 Greensche Integralsätze, 71
 von Bolzano-Weierstraß, 9
 von Fubini, 42
 von Gauß (Gaußscher Integralsatz), 66
 von Schwarz, 17
 von Stokes, 74
 Schrödinger Gleichung, 83
 Schwerpunkt, 40, 57, 64, 70
 Schwingungsgleichung, 81
 Sekantenverfahren, 36
 Separationsansatz, 90
 Standardabweichung, 112
 stetig verteilt, 108
 Stetigkeit, 7
 stochastisch unabhängig, 103, 115
 Stokes'scher Integralsatz, 74
 Substitutionsformel, 45

 Tangente, 15
 Tangenteneinheitsvektor, 51
 Tangentialabbildung, 15
 Tangentialebene, 60
 Tangentialvektor, 50, 60
 topologische Grundbegriffe, 8
 Transformationsformel, 45, 121
 Transportgleichung, 80, 85
 Treppenfunktion, 37
 Tschebyscheff Ungleichung, 114

 unkorreliert, 115

 Varianz, 112
 Vektorfeld, 4
 Verteilungsfunktion, 105
 Volumenintegral, 121

 Wärmeleitungsgleichung, 81
 Wahrscheinlichkeitsverteilung, 96
 Wahrscheinlichkeitsfunktion, 96
 Wahrscheinlichkeitsmaß, 96

 Wahrscheinlichkeitsraum, 96
 Wellengleichung, 81
 Wertebereich, 3
 wirbelfrei, 73

 Zufallsexperiment, 95
 Zufallsvariable, 105
 unkorreliert, 115
 zusammenhängende Menge, 76
 Zykloide, 54
 Zylinderkoordinaten, 48