BERGISCHE
UNIVERSITÄT
WUPPERTAL

Analysis II

Jochen Glück

Vorlesungsmanuskript
Sommersemester 2025

Version: 14. Juli 2026

Inhaltsverzeichnis

Einleitung

Notation

Folgende Notation wird im Manuskript durchgehend verwendet:

Notation	Bedeutung
$\mathbb{N}$	$\{1, 2, 3, \dots\}$
$\mathbb{N}_0$	$\{0, 1, 2, 3, \dots\}$
$\mathbb{K}$	$\mathbb{R}$ or $\mathbb{C}$
m, n, p	Elemente von $\mathbb{N}$

Standardliteratur zur Vorlesung

Einige deutschsprachige Standardwerke zu den Themen der Vorlesung finden Sie im Literaturverzeichnis des Manuskripts unter den Einträgen [2, 3, 5, 6, 7].

Vorsicht: Fehler!

Dieses Manuskript enthält mit an Sicherheit grenzender Wahrscheinlichkeit Fehler. Mit hoher Wahrscheinlichkeit enthält es sogar viele Fehler. Wenn Sie glauben, einen Fehler entdeckt zu haben: Geben Sie mir bitte Bescheid (wirklich!), am einfachsten per E-Mail.

E-Mail-Adresse: glueck@uni-wuppertal.de

Stand

Diese Version des Manuskript stammt vom 14. Juli 2026.

Kapitel 1

Metriken und Normen

1.1 Metrische und normierte Räume

Vorlesung 1: (Mi, 09.04.) beginnt ab Abschnitt 1.1

In der Analysis 1 haben Sie gelernt, wie der Abstand zwischen zwei reellen – und allgemeiner zwischen zwei komplexen – Zahlen definiert ist.[1] Wir besprechen nun einen axiomatischen Zugang um einen Abstandbegriff auf allgemeinen Mengen einzuführen:

Definition 1.1.1 (Metrischer Raum)**.** Sei M eine Menge. Eine **Metrik** oder ein *Abstand* auf M ist eine Abbildung $\mathrm{d} : M \times M \to [0, \infty)$, welche die folgenden Eigenschaften erfüllt:

(I) **Positive Definitheit von Metriken:** Für alle $x, y \in M$ gilt $\mathrm{d}(x, y) = 0$ genau dann, wenn $x = y$ gilt.

(II) **Symmetrie:** Für alle $x, y \in M$ gilt $\mathrm{d}(x, y) = \mathrm{d}(y, x)$.

(III) **Dreiecksungleichung für Metriken:** Für alle $x, y, z \in M$ gilt $\mathrm{d}(x, z) \leq \mathrm{d}(x, y) + \mathrm{d}(y, z)$.

Ein Tupel (M, d), wobei M eine Menge und d eine Metrik auf M ist, nennt man einen **metrischen Raum**.

Beispiel 1.1.2. Sei $M \subseteq \mathbb{C}$ und sei $\mathrm{d}\colon M \times M \to [0, \infty)$ gegeben durch $\mathrm{d}(x, y) = |y - x|$ für alle $x, y \in M$. Dann ist d eine Metrik auf M; man nennt sie die **Euklidsche Metrik**.

Beweis. Wir überprüfen, dass die drei Axiome aus Definition 1.1.1 erfüllt sind.

Positive Definitheit von Metriken: Seien $x, y \in M$. Falls $x = y$ ist, gilt

$$\mathrm{d}(x, y) = |y - x| = |x - x| = |0| = 0.$$

Falls umgekehrt $\mathrm{d}(x, y) = 0$ gilt, dann ist $|y - x| = 0$ und somit $y - x = 0$, also $x = y$.

[1] Nämlich wie?

Symmetrie: Für alle $x, y \in M$ gilt

$$\mathrm{d}(x,y) = |y-x| = |-(x-y)| = |x-y| = \mathrm{d}(y,x).$$

Dreiecksungleichung für Metriken: Seien $x, y, z \in M$. Dann gilt

$$\begin{aligned}\mathrm{d}(x,z) &= |z-x| = |(z-y)+(y-x)| \\ &\leq |z-y| + |y-x| = \mathrm{d}(y,z) + \mathrm{d}(x,y),\end{aligned}$$

wobei wir für die Ungleichung in der Mitte die Dreiecksungleichung für den Betrag komplexer Zahlen verwendet haben. □

Wenn man zwei Elemente $x, y \in \mathbb{R}^n$ betrachtet, kann man Sie eintragsweise addieren und eintragsweise mit Zahlen $\lambda \in \mathbb{R}$ multiplizieren. Es ist also

$$x + y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} \in \mathbb{R}^n$$

und

$$\lambda x = \lambda \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix} \in \mathbb{R}^n.$$

Analog kann man auch Elemente aus $\mathbb{C}^n$ addieren und mit Zahlen aus $\mathbb{C}$ multiplizieren. Die Möglichkeiten zwei Elemente einer Menge zu addieren und Elemente der Menge mit reellen oder komplexen Zahlen zu mutliplizieren, abstrahiert man zum Begriff des *Vektorraums*:

Bemerkung 1.1.3 (Vektorräume)**.** Sei $\mathbb{K}$ ein Körper.[2] Ein Vektorraum über $\mathbb{K}$ ist eine Menge V zusammen mit zwei Abbildungen $+ : V \times V \to V$ und $\cdot : \mathbb{K} \times V \to V$ derart, dass man mit diesen beiden Operationen "im Wesentlichen genauso rechnen kann wie auf $\mathbb{R}^n$ und $\mathbb{C}^n$." Eine genaue Definition des Begriffs lernen Sie in der Vorlesung *Lineare Algebra 1*.

Falls Sie ungeduldig sind und jetzt schon eine exakte Definition des Begriffs nachlesen wollen, finden Sie diese in Addendum 1.5.

Beispiele 1.1.4 (Einige Vektorräume)**.** (a) Wie vor und in Bemerkung 1.1.3 bereits angedeutet, sind für $n \in \mathbb{N}$ die Mengen $\mathbb{R}^n$ und $\mathbb{C}^n$ mit der komponentenweise Addition und komponentenweisen skalaren Multiplikation Vektorräume über $\mathbb{R}$ beziehungsweise $\mathbb{C}$.

[2] In der Analysis – unter anderem in der Vorlesung Analysis II – interessiert man sich vor allem für die Fälle $\mathbb{K} = \mathbb{R}$ und $\mathbb{K} = \mathbb{C}$. Andere Körper sind insbesondere in der Algebra relevant.

(b) Sei $M \subseteq \mathbb{C}$. Dann ist die Menge $\mathbb{C}^M$ aller Funktionen von M nach $\mathbb{C}$ ein Vektorraum über $\mathbb{C}$, wenn man sie mit der punktweisen Addition und punktweisen skalaren Multiplikation ausstattet, die durch

$$(f+g)(z) := f(z) + g(z) \qquad \text{und} \qquad (\lambda f)(z) := \lambda f(z) \quad \text{für alle } z \in M$$

für alle $f, g \in \mathbb{C}^M$ und alle $\lambda \in \mathbb{C}$ gegeben sind.

Analog kann man die Menge $\mathbb{R}^M$ aller reellwertigen Funktionen auf M zu einem Vektorraum über $\mathbb{R}$ machen.

(c) Sei $M \subseteq \mathbb{C}$. Sei $\mathrm{C_b}(M;\mathbb{C})$ die Menge aller beschränkten stetigen Funktionen von M nach $\mathbb{C}$.[3] Wenn man diese Menge mit der punktweisen Addition und der punktweisen skalaren Mulitplikation aus Beispiel (b) ausstattet, ist sie ein Vektorraum über $\mathbb{C}$.

Analog wird die Menge $\mathrm{C_b}(M;\mathbb{R})$ aller beschränkten stetigen Funktionen von M nach $\mathbb{R}$ zu einem Vektorraum über $\mathbb{R}$.

Definition 1.1.5 (Normen)**.** Sei V ein Vektorraum über $\mathbb{K}$, wobei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ ist. Eine **Norm** auf V ist eine Abbildung $\|\cdot\| : V \to [0, \infty)$, die die folgenden Axiome erfüllt:

(I) **Positive Definitheit von Normen:** Für jedes $v \in V$ gilt

$$\|v\| = 0 \quad \Leftrightarrow \quad v = 0$$

(II) **Absolute Homogenität:** Für alle $v \in V$ und alle $\lambda \in \mathbb{K}$ gilt

$$\|\lambda v\| = |\lambda| \, \|v\| \, .$$

(III) **Dreiecksungleichung für Normen:** Für alle $v, w \in V$ gilt

$$\|v + w\| \leq \|v\| + \|w\| \, .$$

Wenn $\|\cdot\|$ eine Norm auf V ist, nennt man das Tupel $(V, \|\cdot\|)$ einen **normierten Raum**.[4]

Beispiel 1.1.6 (**1**-Norm auf $\mathbb{K}^{\boldsymbol{n}}$)**.** Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und $n \in \mathbb{N}$. Für jedes $x \in \mathbb{K}^n$ setzen wir

$$\|x\|_1 := |x_1| + \cdots + |x_n| = \sum_{k=1}^{n} |x_k| \, .$$

Dann ist $\|\cdot\|_1$ eine Norm auf $\mathbb{K}^n$. Man bezeichnet sie als die **1-Norm** auf $\mathbb{K}^n$.

[3] Eine Funktion $f : M \to \mathbb{C}$ heißt **beschränkt**, wenn ihr Bild $f(M)$ beschränkt ist, das heißt, wenn es eine reelle Zahl $c \geq 0$ gibt derart, dass $|f(z)| \leq c$ für alle $z \in M$ gilt.

[4] Oder ausführlicher einen **normierten Vektorraum**.

Beweis. Für jedes $x \in \mathbb{K}^n$ folgt direkt aus der Definition, dass $\|x\|_1 \in [0, \infty)$ liegt. Wir zeigen, dass die drei Axiome aus Definition 1.1.5 erfüllt sind:

Positive Definitheit für Normen: Sei $x \in \mathbb{C}^n$.

"$\Rightarrow$" Sei $\|x\|_1 = 0$. Dann folgt $\sum_{k=1}^n |x_k| = 0$. Weil alle Summanden größer oder gleich 0 sind, gilt für alle $k \in \{1, \dots, n\}$ somit $|x_k| = 0$ und folglich $x_k = 0$. Also ist $x = 0$.

"$\Leftarrow$" Sei $x = 0$. Dann gilt $x_1 = \cdots = x_n = 0$ und somit folgt

$$\|x\|_1 = \sum_{k=1}^n 0 = 0.$$

Absolute Homogenität: Sei $x \in \mathbb{K}^n$ und $\lambda \in \mathbb{K}$. Dann gilt

$$\|\lambda x\| = \sum_{k=1}^n |\lambda x_k| = \sum_{k=1}^n |\lambda|\,|x_k| = |\lambda| \sum_{k=1}^n |x_k| = |\lambda|\,\|x\|_1 .$$

Dreieckungleichung für Normen: Seien $x, y \in \mathbb{K}^n$. Dann gilt

$$\begin{aligned} \|x + y\|_1 &= \sum_{k=1}^n |x_k + y_k| \leq \sum_{k=1}^n (|x_k| + |y_k|) \\ &= \sum_{k=1}^n |x_k| + \sum_{k=1}^n |y_k| = \|x\|_1 + \|y\|_1 , \end{aligned}$$

wobei wir für die Ungleichung in der Mitte die Dreiecksungleichung für den reellen beziehungsweise komplexen Betrag verwendet haben. □

Beispiel 1.1.7 (∞-Norm auf $\mathbb{K}^n$)**.** Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und $n \in \mathbb{N}$. Für jedes $x \in \mathbb{K}^n$ setzen wir

$$\|x\|_\infty := \max\big\{\, |x_1|, \dots, |x_n| \,\big\} = \max\big\{\, |x_k| \mid\ k \in \{1, \dots, n\}\big\}.$$

Dann ist $\|\cdot\|_\infty$ eine Norm auf $\mathbb{K}^n$. Man bezeichnet sie als die **∞-Norm** oder die **Maximumsnorm** auf $\mathbb{K}^n$.

Beweis. Den Beweis stellen wir als Übungsaufgabe. □

Proposition 1.1.8 (Jede Norm induziert eine Metrik)**.** *Sei V ein Vektorraum über $\mathbb{K}$, wobei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ ist, und sei $\|\cdot\|$ eine Norm auf V. Sei $\mathrm{d} : V \times V \to [0, \infty)$ gegeben durch*

$$\mathrm{d}(x, y) := \|y - x\|$$

für alle $x, y \in V$. Dann ist d eine Metrik auf V. Man nennt sie ***die von der Norm $\|\cdot\|$ induzierte Metrik****.*

Beweis. Die die Norm per Definition nach $[0, \infty)$ abbildet, bildet auch d nach $[0, \infty)$ ab. Wir überprüfen nun die drei Axiome einer Metrik aus Definition 1.1.1:

Positive Definitheit von Metriken: Seien $x, y \in V$. Falls $\mathrm{d}(y, x) = 0$ ist, dann gilt $\|y - x\| = 0$ und somit ist wegen der positiven Definitheit der Norm $y - x = 0$, also $x = y$.

Falls umgekehrt $x = y$ gilt, dann ist $y - x = 0$ und somit

$$\mathrm{d}(x, y) = \|y - x\| = \|0\| = 0,$$

wobei wir nun die andere Implpikation aus der Eigenschaft (I) von Normen verwendet haben.

Symmetrie: Seien $x, y \in V$. Dann gilt

$$\mathrm{d}(x, y) = \|y - x\| = \|(-1)(x - y)\| = \underbrace{|-1|}_{=1} \|x - y\| = \mathrm{d}(y, x),$$

wobei wir für die zweite Gleichheit die absolute Homogenität von Normen verwendet haben.

Dreiecksungleichung für Metriken: Seien $x, y, z \in V$. Dann gilt

$$\begin{aligned} \mathrm{d}(x, z) &= \|z - x\| = \|(z - y) + (y - x)\| \\ &\leq \|z - y\| + \|y - x\| = \mathrm{d}(y, z) + \mathrm{d}(x, y), \end{aligned}$$

wobei wir für die Ungleichung die Dreiecksungleichung für Normen verwendet haben. □

Vorlesung 2: (Fr, 11.04.) beginnt ab Konvention 1.1.9

Konvention 1.1.9 (Die Metrik auf normierten Räumen)**.** Wir statten jeden normierten Raum, sofern nichts anderes gesagt wird, immer mit der Metrik aus Proposition 1.1.8 aus.

Das folgende Beispiel ist wohl das wichtigste Beispiel für eine Norm überhaupt, denn auf $\mathbb{R}^2$ und $\mathbb{R}^3$ beschreibt diese Norm die Länge von Vektoren in der Ebene und im Raum. Es ist sehr wichtig, dass Sie sich diese geometrische Bedeutung von Beispiel 1.1.10 klar machen.[5]

Beispiel 1.1.10 (**2**-Norm auf $\mathbb{K}^n$)**.** Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und $n \in \mathbb{N}$. Für jedes $x \in \mathbb{K}^n$ setzen wir

$$\|x\|_2 := \big(|x_1|^2 + \cdots + |x_n|^2 \big)^{1/2} = \Big(\sum_{k=1}^{n} |x_k|^2 \Big)^{1/2}.$$

Dann ist $\|\cdot\|_2$ eine Norm auf $\mathbb{K}^n$. Man bezeichnet sie als die **2-Norm** oder die **Euklidische Norm** auf $\mathbb{K}^n$.

[5] Deshalb werden wir darüber auch im Tutorium in der ersten Vorlesungswoche sprechen.

Beweis. Für jedes $x \in \mathbb{K}^n$ folgt sofort aus der Definition der Euklidischen Norm, dass $\|x\|_2 \geq 0$ gilt. Wir zeigen nun, dass die drei Axiome aus der Definition einer Norm erfüllt sind:

Positive Definitheit für Normen: Sei $x \in \mathbb{C}^n$.

"$\Rightarrow$" Sei $\|x\|_2 = 0$. Dann folgt $\sum_{k=1}^n |x_k|^2 = 0$. Weil alle Summanden größer oder gleich 0 sind, gilt für alle $k \in \{1, \ldots, n\}$ somit $|x_k|^2 = 0$, also $|x_k| = 0$ und folglich $x_k = 0$. Also ist $x = 0$.

"$\Leftarrow$" Sei $x = 0$. Dann gilt $x_1 = \cdots = x_n = 0$ und somit folgt

$$\|x\|_2 = \Big(\sum_{k=1}^n 0^2 \Big)^{1/2} = 0^{1/2} = 0.$$

Absolute Homogenität: Sei $x \in \mathbb{K}^n$ und $\lambda \in \mathbb{K}$. Dann gilt

$$\begin{aligned} \|\lambda x\|_2 &= \Big(\sum_{k=1}^n |\lambda x_k|^2 \Big)^{1/2} = \Big(\sum_{k=1}^n |\lambda|^2 |x_k|^2 \Big)^{1/2} \\ &= \Big(|\lambda|^2 \sum_{k=1}^n |x_k|^2 \Big)^{1/2} = |\lambda| \Big(\sum_{k=1}^n |x_k|^2 \Big)^{1/2} = |\lambda| \, \|x\|_2 . \end{aligned}$$

Dreieckungleichung für Normen: Der Beweis dieser Ungleichung erfordert etwas Vorarbeit. Wir werden ihn führen, sobald wir im Folgenden das Standard-Skalarprodukt auf $\mathbb{K}^n$ eingeführt und die Cauchy-Schwarz-Ungleichung in Lemma 1.1.12 bewiesen haben. □

Definition 1.1.11 (Standard-Skalarprodukt auf $\mathbb{K}^n$)**.** Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und $n \in \mathbb{N}$. Für alle x, y nennt man die Zahl

$$(x \mid y) := \sum_{k=1}^n \overline{x_k} y_k \in \mathbb{K}$$

das **Standard-Skalarprodukt** von x und y.[6,7]

[6] Im Fall $\mathbb{K} = \mathbb{R}$ kann man die komplexe Konjugation über den Zahlen x_k natürlich einfach weglassen, da jede reelle Zahl gleich ihrer komplex konjugierten Zahl ist.

[7] Es gibt übrigens zwei verschiedene Konventionen um das Standard-Skalarprodukt auf $\mathbb{C}^n$ zu definieren: Man könnte die komplexe Konjugation auf die Zahlen y_k anstelle der x_k anwenden. Diese alternative Definition – die zu unserer nicht äquivalent ist, mit der man aber ebenso arbeiten kann, wenn man in allen Formeln die Reihenfolge innerhalb jedes Skalarprodukts vertauscht – ist in der Mathematik recht üblich. Die Definition, die wir verwenden, ist hingegen eher in der Physik üblich. (Personen, die sich an einer möglichst effizienten Notation beim Rechnen mit Matrizen erfreuen oder die gerne die Dirac-Notation in der Quantenmechanik verwenden, könnten sich zur etwas pointierten Aussage hinreißen lassen, dass die Physikerinnen und Physiker mit ihrer Konvention "Recht haben".)

Es folgt leicht aus der Definition, dass das Standardskalarprodukt in seiner zweiten Komponente linear ist und in seiner ersten Komponente anti-linear – das heißt, für alle $x, y, z \in \mathbb{K}^n$ und alle $\lambda \in \mathbb{K}$ gilt[8]

$$\begin{aligned} (x \mid y+z) &= (x \mid y) + (x \mid z) \qquad \text{und} \qquad (x \mid \lambda y) = \lambda\,(x \mid y)\,, \\ (x+y \mid z) &= (x \mid z) + (y \mid z) \qquad \text{und} \qquad (\lambda x \mid y) = \overline{\lambda}\,(x \mid y)\,. \end{aligned}$$

Der Zusammenhang zwischen dem Standard-Sklarprodukt und der Euklidischen Norm besteht darin, dass für jedes $x \in \mathbb{K}^n$ die Gleichheit $(x \mid x) = \big(\|x\|_2\big)^2$ gilt. Dies und die soeben beobachteten Eigenschaften des Skalarprodukts können wir verwenden um das folgende sehr nützliche Resultat zu beweisen.

Lemma 1.1.12 (Cauchy–Schwarz-Ungleichung)**.** *Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und $n \in \mathbb{N}$. Für alle $x, y \in \mathbb{K}^n$ gilt*

$$|(x \mid y)| \leq \|x\|_2 \, \|y\|_2 \,.$$

Beweis. Wir machen zunächst die folgende Beobachtung: Für alle Zahlen $a_1, a_2, b_1, b_2 \in [0, \infty)$ gilt

$$(a_1 b_1 + a_2 b_2)^2 \leq (a_1^2 + a_2^2)(b_1^2 + b_2^2), \tag{1.1.1}$$

denn es ist

$$\begin{aligned} (a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1 b_1 + a_2 b_2)^2 &= a_1^2 b_2^2 + a_2^2 b_1^2 - 2 a_1 b_1 a_2 b_2 \\ &= (a_1 b_2 - a_2 b_1)^2 \geq 0. \end{aligned}$$

Nun beweisen wir die Behauptung per Induktion über n.

Für $n = 1$ ist $|(x \mid y)| = |x_1| \, |y_1| = \|x\|_2 \, \|y\|_2$.

Sei nun die Behauptung für ein festes $n \in \mathbb{N}$ und alle Elemente in $\mathbb{K}^n$ bereits bewiesen und seien $x, y \in \mathbb{K}^{n+1}$ beliebig, fest. Mit $\hat{x}, \hat{y} \in \mathbb{K}^n$ bezeichnen wir diejenigen Elemente von $\mathbb{K}^n$, die aus x und y entstehen, wenn man den letzten Eintrag weglässt. Dann gilt

$$\begin{aligned} |(x \mid y)| &\leq |(\hat{x} \mid \hat{y})| + |x_{n+1}| \, |y_{n+1}| \overset{\text{I.H.}}{\leq} \|\hat{x}\|_2 \, \|\hat{y}\|_2 + |x_{n+1}| \, |y_{n+1}| \\ &\overset{(1.1.1)}{\leq} \big(\|\hat{x}\|_2^2 + |x_{n+1}|^2\big)^{1/2} \big(\|\hat{y}\|_2^2 + |y_{n+1}|^2\big)^{1/2} = \|x\|_2 \, \|y\|_2 \,. \end{aligned}$$

Damit ist der Induktionsschritt komplett und das Lemma bewiesen. □

Nun können wir zeigen, dass die Euklidische Norm tatsächlich die Dreiecksungleichung erfüllt. Dieser Teil des Beweis hat nach Beispiel 1.1.10 noch gefehlt.

[8] Die Behauptung, dass das "leicht aus der Definition" folgt, bedeutet übrigens nicht, dass Sie das einfach glauben sollen. Rechnen Sie diese Aussagen auf jeden Fall selbst nach!

Beweis, dass die 2-Norm aus Beispiel 1.1.10 die Dreiecksungleichung erfüllt. Es seien $x, y \in \mathbb{K}^n$. Dann ist

$$\begin{aligned}\|x+y\|_2^2 &= (x+y \mid x+y) = (x \mid x+y) + (y \mid x+y)\\ &= (x \mid x) + (x \mid y) + (y \mid x) + (y \mid y) = \|x\|_2^2 + \|y\|_2^2 + (x \mid y) + (y \mid x)\,.\end{aligned}$$

Somit folgt

$$\begin{aligned}\|x+y\|_2^2 = \Big|\|x+y\|_2^2\Big| &\leq \|x\|_2^2 + \|y\|_2^2 + |(x \mid y)| + |(y \mid x)|\\ &\leq \|x\|_2^2 + \|y\|_2^2 + \|x\|_2\,\|y\|_2 + \|y\|_2\,\|x\|_2 = \big(\,\|x\|_2 + \|y\|_2\,\big)^2,\end{aligned}$$

wobei wir für die zweite Ungleichung die Cauchy–Schwarz-Ungleichung aus Lemma 1.1.12 verwendet haben. Ziehen der Wurzel liefert nun wie behauptet $\|x+y\|_2 \leq \|x\|_2 + \|y\|_2$. □

Das folgende Beispiel wird auf einem Übungsblatt als Bonusaufgabe besprochen.

Beispiel 1.1.13 (***p*-Norm auf $\mathbb{K}^n$).** Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, sei $n \in \mathbb{N}$ und $p \in [1, \infty)$. Für jedes $x \in \mathbb{K}^n$ setzen wir

$$\|x\|_p := \big(\,|x_1|^p + \cdots + |x_n|^p\,\big)^{1/p} = \Big(\sum_{k=1}^{n} |x_k|^p\Big)^{1/p}.$$

Dann ist $\|\cdot\|_p$ eine Norm auf $\mathbb{K}^n$. Man bezeichnet sie als die ***p*-Norm** auf $\mathbb{K}^n$.[9]

1.2 Konvergenz, Stetigkeit und topologische Konzepte

Definition 1.2.1 (Kugeln und beschränkte Mengen)**.** Sei (M, d) ein metrischer Raum, sei $x \in M$ und $r \in [0, \infty)$.

(a) Dann nennt man die Mengen

$$\begin{aligned}B_{<r}(x) &:= \{y \in M \mid \mathrm{d}(y,x) < r\},\\ B_{\leq r}(x) &:= \{y \in M \mid \mathrm{d}(y,x) \leq r\}\end{aligned}$$

die **offene** und die **abgeschlossene Kugel** in M mit **Mittelpunkt** x und **Radius** r.

(b) Eine Menge $C \subseteq M$ heißt **beschränkt**, wenn es ein $x \in M$ und ein $r \in [0, \infty)$ mit der Eigenschaft $C \subseteq B_{\leq r}(x)$ gibt.

Definition 1.2.2 (Offene und abgeschlossene Mengen)**.** Sei (M, d) ein metrischer Raum.

[9] Beachten Sie: Für $p = 1$ bezeihungsweise $p = 2$ erhält man die 1-Norm aus Beispiel **??** beziehungsweise die 2-Norm aus Beispiel 1.1.10 als Spezialfall.

(a) Eine Menge $U \subseteq M$ heißt **offen**, falls sie folgende Eigenschaft erfüllt:

$$\forall x \in U\ \exists \varepsilon > 0 :\ B_{<\varepsilon}(x) \subseteq U$$

(b) Eine Menge $A \subseteq M$ heißt **abgeschlossen**, falls $M \setminus A$ offen ist.

Vielleicht fragen Sie sich, wie Definition 1.2.2(b) mit der Definition von abgeschlossenen Teilmengen von $\mathbb{C}$ zusammenpasst, die wir in der Analysis 1 verwendet haben. Dass beide Definitionen tatsächlich konsistent sind, werden wir nachher in Proposition 1.2.7 sehen.

Beispiele 1.2.3 (Offene/abgeschlossene Kugeln sind tatsächlich offen/abgeschlossen)**.** Sei (M, d) ein metrischer Raum, sei $x \in M$ und $r \in [0, \infty)$.

(a) Die offene Kugel $B_{<r}(x)$ ist tatsächlich offen im Sinne von Definition 1.2.2(a).

(b) Die abgeschlossene Kugel $B_{\leq r}(x)$ ist tatsächlich abgeschlossen im Sinne von Definition 1.2.2(b).

Beweis. (a) Sei $y \in B_{<r}(x)$ beliebig, fest. Dann ist $\mathrm{d}(y, x) < r$. Wir setzen $\varepsilon := r - \mathrm{d}(y, x) > 0$ und zeigen nun, dass $B_{<\varepsilon}(y) \subseteq B_{<r}(x)$ ist.

Sei dazu $z \in B_{<\varepsilon}(y)$ beliebig, fest. Dann gilt $\mathrm{d}(z, y) < \varepsilon$ und somit

$$\mathrm{d}(z, x) \leq \mathrm{d}(z, y) + \mathrm{d}(y, x) < \varepsilon + \mathrm{d}(y, x) = r,$$

also gilt in der Tat $z \in B_{<r}(x)$, womit die behauptete Inklusion bewiesen ist.

(b) Laut Definition 1.2.3(b) müssen wir zeigen, dass $M \setminus B_{\leq r}(x)$ offen ist. Sei also $y \in M \setminus B_{\leq r}(x)$ beliebig, fest. Dann gilt $\mathrm{d}(y, x) > r$. Wir setzen nun $\varepsilon := \mathrm{d}(y, x) - r > 0$. Ähnlich wie in (a) kann man dann mit Hilfe der Dreiecksungleichung zeigen, dass $B_{<\varepsilon}(y) \subseteq M \setminus B_{\leq r}(x)$ gilt[10] und somit ist $M \setminus B_{\leq r}(x)$ in der Tat offen. □

Definition 1.2.4 (Konvergenz in metrischen Räumen)**.** Sei (M, d) ein metrischer Raum und sei $x \in M$. Eine Folge $(x_n)_{n \in \mathbb{N}}$ heißt **konvergent gegen x**, falls folgendes gilt:

$$\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N} :\ \forall n \in \mathbb{N} :\quad \big(n \geq n_0\ \Rightarrow\ \mathrm{d}(x_n, x) < \varepsilon\big)$$

Man schreibt auch $\lim_{n \to \infty} x_n = x$ oder $x_n \to x$ für $n \to \infty$ um auszudrücken, dass die Folge $(x_n)_{n \in \mathbb{N}}$ gegen x konvergiert.

Vorlesung 3: (Mi, 16.04.) beginnt ab Bemerkung 1.2.5

Bemerkung 1.2.5 (Konvergenz in metrischen Räumen)**.** Sei (M, d) ein metrischer Raum, sei $x \in M$ und sei $(x_n)_{n \in \mathbb{N}}$ eine Folge in M. Aus Definition 1.2.4 folgt, dass $x_n \to x$ für $n \to \infty$ genau dann gilt, wenn $\mathrm{d}(x_n, x) \to 0$ für $n \to \infty$ gilt.

Die Schreibweise $x = \lim_{n \to \infty} x_n$ in Definition 1.2.4 wird durch die folgende Proposition gerechtfertigt.

[10] Können Sie die Details ausführen?

Proposition 1.2.6 (Eindeutigkeit von Grenzwerten)**.** *Sei (M, d) ein metrischer Raum und sei $(x_n)_{n\in\mathbb{N}}$ eine Folge ein M. Dann konvergiert $(x_n)_{n\in\mathbb{N}}$ gegen höchstens ein Element von M.*

Der Beweis ist derselbe wie in Analysis 1 (siehe zum Beispiel [4, Proposition 3.1.3]). Der Vollständigkeit halber führen wir ihn hier noch einmal an, um zu zeigen, wie er mit der Notation für metrische Räume aussieht. In der Vorlesung besprechen wir ihn aber nicht.

Beweis von Proposition 1.2.6. Seien $x, y \in M$ derart, dass $(x_n)_{n\in\mathbb{N}}$ sowohl gegen x als auch gegen y konvergiert. Wir müssen $x = y$ zeigen und nehmen widerspruchshalber an, dass $x \neq y$ gilt. Weger der positiven Definitheit von Metriken ist die Zahl $\varepsilon := \mathrm{d}(x, y)$ dann strikt größer als 0. Also gibt es wegen der Konvergenz der Folge gegen x und gegen y Indizes $n_0, n_1 \in \mathbb{N}$ derart, dass $\mathrm{d}(x_n, x) < \frac{\varepsilon}{2}$ für alle $n \in \mathbb{N}$ mit $n \geq n_0$ und $\mathrm{d}(x_n, y) < \frac{\varepsilon}{2}$ für alle $n \in \mathbb{N}$ mit $n \geq n_1$ gilt. Für den Index $n := \max\{n_0, n_1\}$ erhalten wir nun mit Hilfe der Dreiecksungleichung für Metriken

$$\mathrm{d}(x, y) \leq \mathrm{d}(x, x_n) + \mathrm{d}(x_n, y) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon = \mathrm{d}(x, y),$$

was ein Widerspruch ist. □

Proposition 1.2.7 (Abgeschlossenheit via konvergenter Folgen)**.** *Sei (M, d) ein metrischer Raum und sei $A \subseteq M$. Dann sind folgende Aussagen äquivalent:*

(i) *Die Menge A ist abgeschlossen.*

(ii) *Für jede Folge $(x_n)_{n\in\mathbb{N}}$ in A, die gegen ein $x \in M$ konvergiert, gilt $x \in A$.*

Beweis. "(i) ⇒ (ii)" Sei A abgeschlossen und sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in A, die gegen ein $x \in M$ konvergiert. Nehmen wir widerspruchsahlber an, dass $x \notin A$ ist.

Weil $M \setminus A$ laut Definition 1.2.2(b) offen ist, gibt es ein $\varepsilon > 0$ derart, dass $B_{<\varepsilon}(x) \subseteq M \setminus A$ gilt. Wegen der Konvergenz von $(x_n)_{n\in\mathbb{N}}$ existiert ein laut Definition 1.2.4 und $n_0 \in \mathbb{N}$ derart, dass $\mathrm{d}(x_n, x) < \varepsilon$ für alle $n \in \mathbb{N}$ mit $n \geq n_0$ gilt. Anders ausgedrückt, gilt für alle diese n also $x_n \in B_{<\varepsilon}(x)$ und somit $x_n \in M \setminus A$. Dies ist ein Widerspruch, da die Folge in A liegt.

"(ii) ⇒ (i)" Wir argumentieren via Kontraposition, das heißt wir nehmen an, dass (i) falsch ist und müssen zeigen, dass (ii) falsch ist.

Weil (i) falsch ist, ist $M \setminus A$ nicht offen, das heißt, es gibt ein $x \in M \setminus A$ deart, dass die offene Kugel $B_{<\varepsilon}(x)$ für jedes $\varepsilon > 0$ die Menge A schneidet. Insbesondere gibt es für jedes $n \in \mathbb{N}$ also einen Punkt $x_n \in A \cap B_{<\frac{1}{n}}(x)$. Somit ist $(x_n)_{n\in\mathbb{N}}$ eine Folge in A und weil für jedes $n \in \mathbb{N}$ die Ungleichung $\mathrm{d}(x_n, x) < \frac{1}{n}$ gilt, konvergiert diese Folge gegen x. Da $x \notin A$ ist, ist (ii) falsch. □

Proposition 1.2.8 (Konvergenz in $\mathbb{K}^d$ bezüglich der p-Norm.)**.** *Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, $d \in \mathbb{N}$ und $p \in [1, \infty]$. Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge*[11] *in $\mathbb{K}^d$ und sei $x \in \mathbb{K}^d$. Wir statten $\mathbb{K}^d$ mit der p-Norm und der hiervon induzierten Metrik aus. Dann sind äquivalent:*

(i) *Die Folge $(x^{(n)})_{n\in\mathbb{N}}$ konvergiert gegen x.*

(ii) *Für jedes $k \in \{1, \dots, d\}$ konvergiert die Folge $(x_k^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{K}$ gegen x_k.*

Beweis für $p = 1$. Wir führen den Beweis für den Fall $p = 1$. Für andere Werte von p läuft der Beweis sehr ähnlich.

"(i) $\Rightarrow$ (ii)" Gelte (i) und sei $k \in \{1, \dots, d\}$ beliebig, fest. Dann gilt

$$\left|x_k^{(n)} - x_k\right| \leq \sum_{j=1}^{d} \left|x_j^{(n)} - x_j\right| = \left\|x^{(n)} - x\right\|_1 \to 0$$

für $n \to \infty$, also konvergiert $(x_k^{(n)})_{n\in\mathbb{N}}$ gegen x_k.

"(ii) $\Rightarrow$ (i)" Gelte (ii). Für jedes $k \in \{1, \dots, d\}$ gilt dann $\left|x_k^{(n)} - x_k\right| \to 0$ für $n \to \infty$. Somit folgt

$$\left\|x^{(n)} - x\right\| = \sum_{k=1}^{d} \left|x_k^{(n)} - x_k\right| \to 0$$

für $n \to \infty$, also konvergiert $(x^{(n)})_{n\in\mathbb{N}}$ gegen x. □

Beispiel 1.2.9 (∞-Norm und gleichmäßige Konvergenz)**.** Sei M eine nichtleere Menge und sei $\mathrm{B}(M;\mathbb{K})$ der Vektorraum der beschränkten Funktionen, die von M nach $\mathbb{K}$ abbilden. Die Abbildung $\|\cdot\|_\infty := \|\cdot\|_{\sup} : \mathrm{B}(M;\mathbb{K}) \to [0, \infty)$, die durch

$$\|f\|_\infty := \|f\|_{\sup} := \sup_{x\in M} |f(x)| := \sup\{|f(x)| \mid x \in M\}$$

gegeben ist, ist eine Norm;[12] man nennt Sie die **∞-Norm** oder die **Supremumsnorm**. Sie kennen Sie bereits aus Analysis 1.

Im normierten Raum $\mathrm{B}(M;\mathbb{K})$ mit der Norm $\|\cdot\|_{\sup}$ betrachten Sie eine Funktion $f \in \mathrm{B}(M;\mathbb{K})$ und eine Folge $(f_n)_{n\in\mathbb{N}}$ in $\mathrm{B}(M;\mathbb{K})$. Laut Bemerkung 1.2.5 gilt $f_n \to f$ für $n \to \infty$ genau dann, wenn $\|f_n - f\|_{\sup} \to 0$ für $n \to \infty$ gilt, und das ist laut Analysis 1-Vorlesung äquivalent dazu, dass $(f_n)_{n\in\mathbb{N}}$ gleichmäßig gegen f konvergiert.

In anderen Worten: Konvergenz im normierten Raum $\mathrm{B}(M;\mathbb{K})$ (der mit der Supremumsnorm ausgestattet ist) ist dasselbe wie gleichmäßige Konvergenz.

[11] Für Folgen in $\mathbb{K}^d$ ist es manchmal besser, den Folgenindex hochgestellt statt tiefgestellt zu schreiben, um ihn nicht mit dem Index zu verwechseln, der auf einzelne Einträge von Vektoren aus $\mathbb{K}^d$ zugreift. Damit man den hochgestellten Index nicht mit einem Exponenten verwechselt, setzen wir ihn zusätzlich in runde Klammern.

[12] Das haben Sie in den Hausaufgaben in Aufgabe 1.2 bewiesen.

Nach diesen Beispielen für Konvergenz in normierten Räumen kommen wir noch einmal zurück auf abgeschlossene und offene Mengen. Die folgenden Aussagen sind oft sehr nützlich.

Vorlesung 4: (Mi, 23.04.) beginnt ab Proposition 1.2.10

Proposition 1.2.10 (Durschnitt und Vereinigung offener und abgeschlossener Mengen)**.** *Sei (M, d) ein metrischer Raum und sei I eine nichtleere Menge.*

(a) *Wenn $U_1, U_2 \subseteq M$ offen sind, dann ist auch $U_1 \cap U_2$ offen.*

(b) *Wenn für jedes $i \in I$ eine offene Menge $U_i \subseteq M$ gegeben ist, dann ist auch $\bigcup_{i\in I} U_i$ offen.*

(c) *Wenn $A_1, A_2 \subseteq M$ abgeschlossen sind, dann ist auch $A_1 \cup A_2$ abgeschlossen.*

(d) *Wenn für jedes $i \in I$ eine abgeschlossene Menge $A_i \subseteq M$ gegeben ist, dann ist auch $\bigcap_{i\in I} A_i$ abgeschlossen.*

Beweis. (a) und (b) Diese Aussagen beweisen Sie in Aufgabe 2.2 auf Hausaufgabenblatt 2.

(c) und (d) Diese Aussagen folgen aus den ersten beiden durch Komplementbildung. □

Dass der Durchschnitt beliebig vieler (also auch unendlicher vieler) abgeschlossener Mengen wieder abgeschlossen ist, kann man für die folgende Definition nutzen.

Definition 1.2.11 (Abschluss einer Menge)**.** Sei (M, d) eine metrischer Raum und sei $C \subseteq M$. Dann nennt man den Durschnitt aller abgeschlossenen Teilmengen von M, die C enthalten, den **Abschluss** von C und notiert ihn als $\overline{C}$.[13] Es ist also

$$\overline{C} := \bigcap_{\substack{M \supseteq A \supseteq C, \\ A \text{ abgeschl.}}} A.$$

Man beachte, dass $\overline{C}$ wegen Proposition 1.2.10(d) stets abgeschlossen ist. Außerdem impliziert die Definition des Abschlusses, dass $\overline{C}$ die kleinste abgeschlossene Teilmenge von M ist, die C enthält – das heißt, jede abgeschlossene Teilmenge von M, die C enthält, enthält auch $\overline{C}$.[14]

Proposition 1.2.12 (Eigenschaften des Abschlusses)**.** *Sei (M, d) eine metrischer Raum und sei $C \subseteq M$.*

(a) *Es gilt $C \subseteq \overline{C}$.*

(b) *Es ist C genau dann abgeschlossen, wenn $C = \overline{C}$ gilt.*

[13] Achtung: Den Querstrich zur Notation des Abschlusses einer Menge darf man nicht mit dem Querstrich verwechseln, mit dem man das komplex Konjugierte einer komplexen Zahl notiert.

[14] Können Sie im Detail begründen, weshalb das so ist?

(c) *Ein Punkt $x \in M$ liegt genau dann im Abschluss $\overline{C}$, wenn es eine Folge $(x_n)_{n\in\mathbb{N}}$ in C gibt, die gegen x konvergiert.*

Beweis. (a) Das folgt sofort aus der Definition von $\overline{C}$.

(b) Wir zeigen beiden Implikationen:

"$\Rightarrow$" Sei C abgeschlossen. Aus (a) wissen wir, dass $C \subseteq \overline{C}$ gilt. Weil C selbst abgeschlossen ist, ist C eine der Mengen, die wir schneiden, um $\overline{C}$ zu erhalten. Also folgt auch $\overline{C} \subseteq C$.

"$\Leftarrow$" Sei $C = \overline{C}$. Da $\overline{C}$ abgeschlossen ist, ist dann auch C abgeschlossen.

(c) Der Beweis ist ähnlich wie der Beweis von Proposition 1.2.7, weshalb wir auf die Details verzichten. □

Auf dem ersten Präsenzblatt und dem zweiten Hausaufgabenblatt werden Sie einige Beispiele für den Abschluss von Mengen besprechen.

Definition 1.2.13 (Stetigkeit zwischen metrischen Räumen)**.** Seien (M, d) und (N, d) zwei metrische Räume[15] und sei $f : M \to N$.

(a) Sei $x_0 \in M$. Die Funktion f heißt **stetig** in x_0, falls gilt:

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x \in M: \quad \big(\mathrm{d}(x, x_0) < \delta \quad \Rightarrow \quad \mathrm{d}(f(x), f(x_0)) < \varepsilon\big)$$

(b) Die Funktion f heißt **stetig**, falls sie in jedem Punkt in M stetig ist.

Das folgende Resultat kennen Sie für Funktionen, die von Teilmengen von $\mathbb{C}$ nach $\mathbb{C}$ abbilden, bereits aus der Analysis 1. Nun verallgemeinern wir es auf Funktionen zwischen metrischen Räumen.

Theorem 1.2.14 (Stetigkeit vs. Folgenstetigkeit)**.** *Seien (M, d) und (N, d) zwei metrische Räume. Sei $f : M \to N$ und sei $x_0 \in M$. Folgende Aussagen sind äquivalent:*

(i) *Die Funktion f ist stetig in x_0.*

(ii) *Für jede Folge $(x_n)_{n\in\mathbb{N}}$, die gegen x_0 konvergiert, konvergiert die Folgt $\big(f(x_n)\big)_{n\in\mathbb{N}}$ gegen $f(x_0)$.*

Beweis. Der Beweis ist genau derselbe wie in der Analysis 1 (siehe zum Beispiel [4, Theorem 3.3.1]); man muss nur die Abstände in $\mathbb{C}$ durch die Metriken auf M und N ersetzen. □

Beispiele 1.2.15 (Die Metrik ist stetig in den einzelnen Komponenten)**.** Sei (M, d) eine metrischer Raum und sei $y \in M$. Sei $[0, \infty)$ mit der Euklidischen Metrik ausgestattet. Dann ist die Abbildung

$$\begin{aligned} \mathrm{d}(\,\cdot\,, y) : M &\to [0, \infty), \\ x &\mapsto \mathrm{d}(x, y) \end{aligned}$$

stetig.

[15] Hier erlauben wir uns eine notationelle Ungenauigkeit: Eigentlich müssten wir die beiden Metriken mit verschiedenen Symbolen bezeichnen.

Beweis. Sei $x \in M$ und sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in M, die gegen x konviergert. Laut Aufgabe 2.1 auf Hausaufgabenblatt 2 gilt dann $\mathrm{d}(x_n, y) \to \mathrm{d}(x, y)$ für $n \to \infty$. Aufgrund von Theorem 1.2.14 ist die Abbildung $\mathrm{d}(\,\cdot\,, y)$ somit in der Tat stetig. □

Theorem 1.2.16 (Stetigkeit via offener Urbilder)**.** *Seien* (M, d) *und* (N, d) *zwei metrische Räume. Sei* $f : M \to N$*. Folgende Aussagen sind äquivalent:*

(i) *Die Funktion* f *ist stetig.*

(ii) *Für jede offene Menge* $V \subseteq N$ *ist das* ***Urbild***

$$f^{-1}(V) := \{x \in M \mid\ f(x) \in V\} \subseteq M$$

ebenfalls offen.

Beweis. "(i) ⇒ (ii)" Sei f stetig und sei $V \subseteq N$ offen. Sei $x \in f^{-1}(V)$. Dann ist $f(x) \in V$. Da V offen ist, gibt es $\varepsilon > 0$ derart, dass $B_{<\varepsilon}(f(x)) \subseteq V$ gilt.

Wegen der Stetigkeit von f finden wir ein $\delta > 0$ derart, dass für jedes $z \in B_{<\delta}(x)$ die Eigenschaft $f(z) \in B_{<\varepsilon}(f(x)) \subseteq V$ und somit $z \in f^{-1}(V)$ gilt. Folglich ist $B_{<\delta}(x) \subseteq f^{-1}(V)$. Dies zeigt, dass $f^{-1}(V)$ tatsächlich offen ist.

"(ii) ⇒ (i)" Gelte (ii). Sei $x \in M$ und $\varepsilon > 0$. Die Kugel $B_{<\varepsilon}(f(x))$ ist laut Beispiel 1.2.3(a) offen, also ist laut Bedingung (ii) auch ihr Urbild $f^{-1}\big(B_{<\varepsilon}(f(x))\big)$ offen.

Somit gibt es ein $\delta > 0$ mit der Eigenschaft $B_{<\delta}(x) \subseteq f^{-1}\big(B_{<\varepsilon}(f(x))\big)$. Für jedes $z \in M$ mit $\mathrm{d}(z, x) < \delta$ gilt wegen $z \in B_{<\delta}(x)$ also $f(z) \in B_{<\varepsilon}(x)$, das heißt $\mathrm{d}(f(z), f(x)) < \varepsilon$.

Das zeigt, dass f stetig in x ist. Da $x \in M$ beliebig war, ist f stetig. □

1.3 Vollständigkeit und Kompaktheit

Definition 1.3.1 (Cauchy-Folgen und Vollständigkeit)**.** Sei (M, d) ein metrischer Raum.

(a) Ein Folge $(x_n)_{n\in\mathbb{N}}$ in M nennt man eine **Cauchy-Folge**, falls folgendes gilt:

$$\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall n \in \mathbb{N} : \quad \big(n \geq n_0 \quad \Rightarrow \quad \mathrm{d}(x_n, x_{n_0}) < \varepsilon\big)$$

(b) Der metrische Raum (M, d) heißt **vollständig**, falls jede Cauchy-Folge in M konvergiert.

Ähnlich wie in der Analysis 1 (siehe [4, Proposition 3.5.2]) kann man zeigen, dass eine Folge $(x_n)_{n\in\mathbb{N}}$ in einem metrischen Raum (M, d) genau dann eine Cauchy-Folge ist, wenn die folgende Bedingung gilt:

$$\forall \varepsilon > 0\ \exists n_0 \in \mathbb{N}\ \forall m, n \in \mathbb{N} : \quad \big(m, n \geq n_0 \quad \Rightarrow \quad \mathrm{d}(x_n, x_m) < \varepsilon\big)$$

Beispiel 1.3.2 ($\mathbb{K}^{\boldsymbol{d}}$ ist vollständig)**.** Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, $d \in \mathbb{N}$ und $p \in [1, \infty]$. Dei $\mathbb{K}^d$ mit der p-Norm $\|\cdot\|_p$ ausgestattet. Dann ist $\mathbb{K}^d$ vollständig.

Vorlesung 5: (Fr, 25.04.) beginnt mit dem Beweis von Beispiel 1.3.2

Beweis. Sei jetzt $(x^{(n)})_{n\in\mathbb{N}}$ eine Cauchy-Folge in $\mathbb{K}^d$. Mit einem ähnlichen Argument wie in Proposition 1.2.8 kann man dann sehen, dass für jedes $k \in \{1, \dots, d\}$ die Komponentenfolge $(x_k^{(n)})_{n\in\mathbb{N}}$ ebenfalls Cauchy ist.

Da in $\mathbb{K}$ laut Analysis 1-Vorlesung jede Cauchy-Folge konvergiert (siehe [4, Theorem 3.5.3][16]), ist $(x_k^{(n)})_{n\in\mathbb{N}}$ somit für jedes $k \in \{1, \dots, d\}$ konvergent. Laut Proposition 1.2.8 konvergiert also die Folge $(x^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{K}^d$. □

Beispiel 1.3.3 ($\mathbb{Q}$ ist nicht vollständig)**.** Wir betrachten auf der Menge $\mathbb{Q}$ der rationalen Zahlen die Euklidische Metrik d, das heißt wir setzen $\mathrm{d}(x, y) := |y - x|$ für alle $x, y \in \mathbb{Q}$. Dann ist $(\mathbb{Q}, \mathrm{d})$ nicht vollständig.

Beweis. Es gibt eine Folge $(q_n)_{n\in\mathbb{N}}$ in $\mathbb{Q}$, die gegen $\sqrt{2}$ konvergiert; siehe Analysis 1, Hausaufgabenblatt 7, Aufgabe 5(c). Weil $(q_n)_{n\in\mathbb{N}}$ gegen eine reelle Zahl konvergiert, kann man leicht nachprüfen, dass es sich um eine Cauchy-Folge handelt. Da der Grenzwert $\sqrt{2}$ aber nicht in $\mathbb{Q}$ liegt, ist die Folgen $(q_n)_{n\in\mathbb{N}}$ im metrischen Raum $\mathbb{Q}$ nicht konvergent.[17] □

Beispiel 1.3.4 (**1**-Norm und $\boldsymbol{\infty}$-Norm auf dem Raum der stetigen Funktionen)**.** Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und sei $\mathrm{C}([0, 1]; \mathbb{K})$ die Menge der der stetigen Funktionen von $[0, 1]$ nach $\mathbb{K}$. Die Menge ist zusammen mit der punktweisen Addition und der punktweisen skalaren Multiplikation ein Vektorraum über $\mathbb{K}$.

(a) Die Abbildung

$$\|\cdot\|_1 : \mathrm{C}([0,1];\mathbb{K}) \to [0,\infty),$$
$$f \mapsto \|f\|_1 := \int_0^1 |f(x)| \,\mathrm{d}x.$$

ist laut Aufgabe 1.3(a) auf Hausaufgabenblatt 1 eine Norm Bezüglich dieser Norm (genauer: bezüglich der von dieser Norm induzierten Metrik) ist der Raum $\mathrm{C}([0, 1]; \mathbb{K})$ nicht vollständig.

(b) Die Abbildung

$$\|\cdot\|_\infty : \mathrm{C}([0,1];\mathbb{K}) \to [0,\infty),$$
$$f \mapsto \|f\|_\infty := \|f\|_{\sup} := \sup\{|f(x)| \mid \ x \in [0,1]\}$$

ist laut Aufgaben 1.2 und 1.3(b) auf Hausaufgabenblatt 1 eine Norm. Bezüglich dieser Norm (genauer: bezüglich der von dieser Norm induzierten Metrik) ist der Raum $\mathrm{C}([0, 1]; \mathbb{K})$ vollständig.

[16] In dieser Referenz steht das behauptete Resultat genau genommen nur für $\mathbb{K} = \mathbb{C}$; weshalb gilt es auch für $\mathbb{K} = \mathbb{R}$?

[17] Hier haben wir übrigens die Eindeutigkeit von Grenzwerten in $\mathbb{R}$ verwendet. Können Sie die Details ausführen?

Beweis. Die Beweise dieser beiden Aussagen besprechen wir in den Übungen. □

In $\mathbb{R}$ und in $\mathbb{C}$ besitzt laut des Satzes von Bolzano–Weierstraß jede beschränkte Folge eine konvergente Teilfolge; siehe [4, Theorem 3.2.4 und Korollar 3.2.5]. Wir wollen uns nun in allgemeinen metrischen Räumen mit der Fragen beschäftigen, wann Folgen konvergente Teilfolgen besitzen. Dazu ist das folgende Konzept hilfreich.

Definition 1.3.5 (Totale Beschränktheit)**.** Sei (M, d) ein metrischer Raum. Eine Menge $K \subseteq M$ heißt **total beschränkt**, wenn sie leer ist oder wenn für jedes $\varepsilon > 0$ eine Zahl $m \in \mathbb{N}$ und Punkte $x_1, \ldots, x_m \in M$ existieren derart, dass

$$K \subseteq \bigcup_{k=1}^{m} B_{<\varepsilon}(x_k)$$

gilt.[18,19]

Mit Hilfe der Dreiecksungleichung können Sie sich leicht überlegen, dass jede total beschränkte Menge beschränkt ist. Die umgekehrte Implikation gilt im Allgemeinen jedoch nicht, wie das folgende Beispiel zeigt:

Beispiel 1.3.6 (Beschränktheit vs. totale Beschränktheit)**.** Betrachten wir den Vektorraum $\mathrm{B}(\mathbb{R};\mathbb{R})$ aller beschränkten Funktionen von $\mathbb{R}$ nach $\mathbb{R}$. Wir statten ihn mit der Norm $\|\cdot\|_\infty$ aus, die durch

$$\|f\|_\infty = \sup\{|f(x)| \mid \ x \in \mathbb{R}\}$$

für alle $f \in \mathrm{B}(\mathbb{R},\mathbb{R})$ gegeben ist. Laut Aufgabe 1.2 auf Hausaufgabenblatt 1 handelt es sich hierbei tatsächlich um eine Norm. Wir üblich betrachten wir im folgenden die von der Norm induzierte Metrik.

Für jedes $y \in \mathbb{R}$ bezeichne $\mathbb{1}_{\{y\}} : \mathbb{R} \to \mathbb{R}$ die Funktion, die an der Stelle y den Wert 1 annimmt und sonst überall den Wert 0.[20] Für jedes $y \in \mathbb{R}$ ist die Funktion $\mathbb{1}_{\{y\}}$ beschränkt, also ein Element von $\mathrm{B}(\mathbb{R},\mathbb{R})$.

Wir betrachten nun die Menge $C := \{\mathbb{1}_{\{y\}} \mid \ y \in \mathbb{R}\} \subseteq \mathrm{B}(\mathbb{R},\mathbb{R})$. Diese Menge ist beschränkt, aber nicht total beschränkt.

Beweis. Die Menge C ist beschränkt: Für jedes $y \in \mathbb{R}$ gilt $\left\|\mathbb{1}_{\{y\}}\right\|_\infty = 1$, das heißt alle Elemente von C haben von 0 den Abstand 1 und somit gilt $C \subseteq B_{\leq 1}(0)$.

Die Menge C ist nicht total beschränkte: Wir wählen $\varepsilon := \frac{1}{4}$. Für zwei verschiedene Punkte $y_1, y_2 \in \mathbb{R}$ gilt $\|\mathbb{1}_{y_1} - \mathbb{1}_{y_2}\|_\infty = 1$, das heißt zwei verschiedene Elemente der Menge C haben immer den Abstand 1. Eine Kugel in $\mathrm{B}(\mathbb{R};\mathbb{R})$ mit Radius $\frac{1}{4}$ kann somit höchstens ein Element von C enthalten. Da C unendlich viele Elemente enthält, können endlich viele Kugeln mit Radius $\frac{1}{4}$ nicht ganz C überdecken. □

[18] In anderen Worten: Wenn sich K für jeden noch so kleinen Radius $\varepsilon > 0$ durch endlich viele Kugeln mit Radius ε überdecken lässt.

[19] Würde sich übrigens etwas ändern, wenn man anstelle der offenen Kugeln $B_{<\varepsilon}(x_k)$ in dieser Definition die abgeschlossenen Kugeln $B_{\leq\varepsilon}(x_k)$ verwenden würde?

[20] Man nennt $\mathbb{1}_{\{y\}}$ auch die **Indikatorfunktion der Menge $\{y\}$** oder die **charakteristische Funktion der Menge $\{y\}$**.

Theorem 1.3.7 (Wann besitzt jede Folge eine konvergente Teilfolge?)**.** *Sei* (M, d) *ein metrischer Raum und* $K \subseteq M$*. Die folgenden Aussagen sind äquivalent:*

(i) *Jede Folge in* K *besitzt eine konvergente Teilfolge mit Grenzwert in* K*.*

(ii) *Die Menge* K *ist total beschränkt und der metrische Raum* (K, d) *ist vollständig.*

(iii) *Für jede Menge* $I \neq \emptyset$ *und jede Familie von offenen Teilmengen* $(U_i)_{i\in I}$ *in* M *gilt: Falls* $\bigcup_{i\in I} U_i \supseteq K$ *gilt, dann existiert eine endliche Teilmenge* $F \subseteq I$ *mit der Eigenschaft* $\bigcup_{i\in F} U_i \supseteq K$*.*[21]

Aus Zeitgründen verzichten wir in der Vorlesung auf den Beweis von Theorem 1.3.7. Falls Sie sich für den Beweis interessieren, können Sie ihn in Addendum 1.6 nachlesen.

Definition 1.3.8 (Kompakte Mengen)**.** Sei (M, d) ein metrischer Raum. Eine Teilmenge K von M heißt **kompakt**, falls sie eine (und somit alle) der äquivalenten Bedingungen (i)–(iii) in Theorem 1.3.7 erfüllt.

Proposition 1.3.9 (Kompakte Mengen sind abgeschlossen)**.** *Sei* (M, d) *ein metrischer Raum und sei* $K \subseteq M$ *kompakt. Dann ist* K *abgeschlossen.*

Beweis. Sie $(x_n)_{n\in\mathbb{N}}$ eine Folge in K, die gegen einen Punkt $x \in M$ konvergiert. Laut Proposition 1.2.7 genügt es zu zeigen, dass $x \in K$ gilt. Da K kompakt ist, besitzt $(x_n)_{n\in\mathbb{N}}$ eine Teilfolge $(x_{n_k})_{k\in\mathbb{N}}$, die gegen einen Punkt $y \in K$ konvergiert.

Mit Hilfe der Definition von Konvergenz kann man leicht sehen, dass jeder Teilfolge einer konvergenten Folge gegen denselben Grenzwert konvergiert,[22] konvergiert $(x_{n_k})_{k\in\mathbb{N}}$ auch gegen x. Da Grenzwerte von Folgen in metrischen Räumen eindeutig sind (Propostion 1.2.6), folgt $x = y$ und somit $x \in K$, wie gewünscht. □

Theorem 1.3.10 (Kompakte Teilmengen auf $\mathbb{K}^d$)**.** *Sei* $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ *und* $d \in \mathbb{N}$*. Sei* $p \in [1, \infty]$ *und sei* $\mathbb{K}^d$ *mit der p-Norm* $\|\cdot\|_p$ *augestattet. Für jede Menge* $K \subseteq \mathbb{K}^d$ *sind äquivalent:*

(i) *Die Menge* K *ist kompakt.*

(ii) *Die Menge* K *ist beschränkt und abgeschlossen.*

Beweis. "(i) ⇒ (ii)" Sei K kompakt. Dann ist K laut Proposition 1.3.9 abgeschlossen. Da K wegen seiner Kompaktheit total beschränkt ist, ist K insbesondere beschränkt.

[21] Diese Eigenschaft fasst man oft kurz zusammen, indem man sagt: Jede offene Überdeckung von K besitzt eine endliche Teilüberdeckung. Manchmal wird diese Eigenschaft als **Heine–Borel-Eigenschaft** bezeichnet.

[22] In Analysiys 1 haben wir das für Folgen in $\mathbb{C}$ in Proposition 3.2.3 gezeigt. In allgemeinen metrischen Räumen verläuft der Beweis genauso.

"(ii) $\Rightarrow$ (i)" Sei K beschränkt und abgeschlossen. Um zu beweisen, dass K kompakt ist, zeigen wir, dass K die Eigenschaft (i) aus Theorem 1.3.7 erfüllt. Sei also $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in K.

Weil K beschränkt ist, gibt es ein $c \in [0,\infty)$ mit der Eigenschaft $\left\|x^{(n)}\right\|_p \leq c$ für alle $n \in \mathbb{N}$. Insbesondere ist die Folge $(x^{(n)_1})_{n\in\mathbb{N}}$ in $\mathbb{K}$, die aus den ersten Komponenten der Vektoren $x^{(n)}$ besteht, beschränkt, und besitzt somit laut des Satzes von Bolzano–Weierstraß [4, Theorem 3.2.4 und Korollar 3.2.5] eine konvergente Teilfolge $(x_1^{(n_k)})_{k\in\mathbb{N}}$. Nun wenden wir dasselbe Argumente auf die Folge $(x_2^{(n_k)})_{k\in\mathbb{N}}$ in $\mathbb{K}$ an und erhalten somit eine Teilfolge $(x_2^{(n_{k_\ell})})_{\ell\in\mathbb{N}}$. Man beachte, dass natürlich auch $(x_1^{(n_{k_\ell})})_{\ell\in\mathbb{N}}$ konvergiert. Indem wir so weiterverfahren und alle Komponenten durchgehen, erhalten wir insgesamt eine Teilfolge von $(x^{(n)})_{n\in\mathbb{N}}$, die komponentenweise konvergiert. Laut Proposition 1.2.8 ist diese Teilfolge bezüglich der p-Norm konvergent und wegen der Abgeschlossenheit von K liegt der Grenzwert der Teilfolge in K. Also ist tatsächlich Eigenschaft (i) aus Theorem 1.3.7 erfüllt. □

Ein Grund, weshalb Kompaktheit ein sehr nützliches Konzept ist, ist Korollar 1.3.12 der folgenden Proposition.

Vorlesung 6: (Mi, 30.04.) beginnt mit dem Beweis von Beispiel 1.3.11

Proposition 1.3.11 (Stetige Bilder kompakter Mengen)**.** *Seien (M,d) und (N,d) metrische Räume und sei $f : M \to N$ stetig. Wenn $K \subseteq M$ kompakt ist, dann ist auch das Bild*

$$f(K) := \{f(x) \mid\ x \in K\} \subseteq N$$

kompakt.

Beweis. Sei $(y_n)_{n\in\mathbb{N}}$ eine Folge in $f(K)$. Für jeden Index $n \in \mathbb{N}$ finden wir einen Punkt $x_n \in K$ mit der Eigenschaft $y_n = f(x_n)$. Die Folge $(x_n)_{n\in\mathbb{N}}$ liegt in K und besitzt, da K laut Voraussetzung kompakt ist, somit eine Teilfolge $(x_{n_k})_{k\in\mathbb{N}}$, die gegen einen Punkt $x \in K$ konvergiert. Da f stetig ist, gilt laut Theorem 1.2.14

$$y_{n_k} = f(x_{n_k}) \overset{k\to\infty}{\to} f(x) \in f(K),$$

also besitzt die Folge $(y_n)_{n\in\mathbb{N}}$ eine Teilfolge, die gegen einen Punkt in $f(K)$ konvergiert. □

Korollar 1.3.12 (Extrema reellwertiger Funktionen)**.** *Sei (M,d) ein nichtleerer metrischer Raum und sei M selbst kompakt. Sei $f : M \to \mathbb{R}$ stetig.*

(a) *Die Funktion f besitzt eine globale Maximalstelle, das heißt, es gibt ein $x_{\max} \in M$ mit der Eigenschaft $f(x) \leq f(x_{\max})$ für alle $x \in M$.*

(b) *Die Funktion f besitzt eine globale Minimalstelle, das heißt, es gibt ein $x_{\min} \in M$ mit der Eigenschaft $f(x_{\min}) \leq f(x)$ für alle $x \in M$.*

Beweis. (a) Das Bild $f(M)$ ist nichtleer, da M laut Voraussetzung nichtleer ist. Laut Proposition 1.3.11 ist das Bild $f(M)$ kompakt. Somit ist $f(M)$ total beschränkt, also insbesondere beschränkt. Außerdem ist $f(M)$ laut Proposition 1.3.9 abgeschlossen, also besitzt $\sup f(M)$ laut der Analysis 1-Vorlesung im vergangenen Semester ein Maximum [4, Proposition 3.4.4(a)].

(b) Diese Aussage erhält man, indem man (a) auf $-f$ anwendet (oder indem man den Beweis von (a) wiederholt, aber dieses Mal benutzt, dass nichtleere, beschränkte, abgeschlossene Teilmengen von $\mathbb{R}$ ein Minimum besitzen). □

Einen Spezialfall von Korollar 1.3.12 kennen Sie bereits aus der Analysis 1: In [4, Theorem 3.3.5] haben wir folgendes bewiesen: Wenn $K \subseteq \mathbb{C}$ beschränkt und abgeschlossen ist[23] und $f : K \to \mathbb{R}$ stetig ist, dann besitzt f eine globale Maximalstelle. Machen Sie sich unbedingt klar, weshalb dieses Resultat ein Spezialfall von Korollar 1.3.12 ist!

Eine weitere sehr nützliche Anwendung von Kompaktheit besteht darin zu zeigen, dass alle Normen auf $\mathbb{K}^d$ zum selben Konvergenzbegriff führen:

Theorem 1.3.13 (Konvergenz bezüglich beliebiger Normen auf $\mathbb{K}^d$)**.** *Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, sei $d \in \mathbb{N}$ und sei $\|\cdot\| : \mathbb{K}^d \to [0, \infty)$ eine Norm. Sei $(x^{(n)})_{n\in\mathbb{N}}$ eine Folge in $\mathbb{K}^d$ und sei $x \in \mathbb{K}^d$. Wir sind äquivalent:*

(i) *Die Folge $(x^{(n)})_{n\in\mathbb{N}}$ konvergiert gegen x.*[24]

(ii) *Für jedes $k \in \{1, \dots, d\}$ konvergiert die Folge $(x_k^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{K}$ gegen x_k.*

Der Beweis des Theorems beruht ganz wesentlich auf einem Kompaktheitsargument. Er ist nicht schwierig und auch nicht besonders technisch, aber etwas länglich. Deshalb lagern wir ihn aus Zeitgründen in Addendum 1.7 aus. Wenn Sie möchten können Sie ihn dort nachlesen.

1.4 Der Banachsche Fixpunktsatz

Definition 1.4.1 (Lipschitz-Stetigkeit)**.** Seien (M, d) und (N, d) metrische Räume und sei $f : M \to N$. Die Funktion f heißt **Lipschitz-stetig**, falls es eine Zahl $L \geq 0$ gibt derart, dass

$$\mathrm{d}(f(x), f(y)) \leq L\,\mathrm{d}(x, y)$$

für alle $x, y \in M$ gilt. In diesem Fall nennt man L eine **Lipschitzkonstante** von f.

[23] Wobei $\mathbb{C}$ in der Analysis 1, aus Sicht der Analysis 2 ausgedrückt, mit der Euklidischen Metrik ausgestattet ist.

[24] Bezüglich der von $\|\cdot\|$ induzierten Metrik.

Falls die metrischen Räumen (M, d) und (N, d) übereinstimmen und f Lipschitz-stetig ist mit einer Lipschitzkonstanten $L \geq 0$, dann ist für jedes $n \in \mathbb{N}$ auch die n-fache Iterierte

$$f^n := f \circ \cdots \circ f$$

Lipschitz-stetig und L^n ist eine Lipschitzkonstante für f^n.[25]

Beispiele 1.4.2 (Einige Beispiele für (nicht-)Lipschitz-Stetigkeit)**.**

(a) Sei $\mathbb{R}^2$ mit der 1-Norm ausgestattet. Die Abbildung

$$\begin{aligned} f : \mathbb{R}^2 &\to \mathbb{R}^2, \\ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &\mapsto \begin{pmatrix} x_1 + x_2 \\ 2x_1 \end{pmatrix} \end{aligned}$$

ist Lipschitz-stetig.

(b) Sei $\mathbb{R}$ mit der Eulidischen Metrik ausgestattet. Die Abbildung

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}, \\ x &\mapsto x^2 \end{aligned}$$

ist nicht Lipschitz-stetig.

Beweis. (a) Für alle $x, y \in \mathbb{R}^2$ gilt

$$\begin{aligned} \|f(x) - f(y)\|_1 &= \left\| \begin{pmatrix} (x_1 + x_2) - (y_1 + y_2) \\ 2x_1 - 2y_1 \end{pmatrix} \right\|_1 \\ &= |(x_1 - y_1) + (x_2 - y_2)| + 2\,|x_1 - y_1| \\ &\leq 3\,|x_1 - y_1| + |x_2 - y_2| \leq 3\,\|x - y\|_1 . \end{aligned}$$

Also ist f Lipschitz-stetig und die Zahl 3 ist eine Lipschitzkonstante von f.

(b) Nehmen wir widerspruchshalber an f wäre Lipschitz-stetig mit einer Lipschitzkonstanten $L \in [0, \infty)$. Betrachte dann eine Zahl $n \in \mathbb{N}$ und setzen $x := n$ sowie $y := n + 1$. Aus der Definition von Lipschitz-Stetigkeit folgt dann

$$L = L\,|y - x| \geq |f(y) - f(x)| = \left|(n+1)^2 - n^2\right| = 2n + 1.$$

Dies würde für jedes $n \in \mathbb{N}$ gelten, was $L < \infty$ widerspricht.[26] □

Im nächsten Theorem verwenden wir den folgenden Begriff: Sei M eine Menge und $f : M \to M$. Einen Punkt $x^* \in M$ nennt man einen **Fixpunkt** von f, falls $f(x^*) = x^*$ gilt.

[25] Wir werden aber im nächsten Kapitel sehen, dass es f^n in manchen Fällen eine viel bessere (das heißt kleinere) Lipschitzkonstante als L^n haben kann.

[26] Genauer gesagt widerspricht dies dem Archimedischen Axiom der reellen Zahlen.

Theorem 1.4.3 (Banachscher Fixpunktsatz)**.** *Sei (M, d) ein nichtleerer, vollständiger metrischer Raum und sei $f : M \to M$ Lipschitz-stetig. Für jedes $n \in \mathbb{N}$ sei $L_n \in [0, \infty)$ eine Lipschitzkonstante von $f^n := f \circ \cdots \circ f$ und es gelte $\sum_{k=1}^{\infty} L_k < \infty$.*

(a) *Die Funktion f besitzt genau einen Fixpunkt, das heißt, es gibt genau ein $x^* \in M$ mit der Eigenschaft $f(x^*) = x^*$.*

(b) *Für jedes $x \in M$ gilt $f^n(x) \to x^*$ für $n \to \infty$.*

Vorlesung 7: (Fr, 02.05.) beginnt mit dem Beweis von Theorem 1.4.3

Beweis. Zunächst zeigen wir folgende Aussage $(*)$: Für jedes $x \in M$ ist die Folge $\big(f^n(x)\big)_{n\in\mathbb{N}}$ konvergent und ihr Grenzwert ist ein Fixpunkt von f.

Um die Konvergenz zu zeigen müssen wir wegen der Vollständigkeit von (M, d) nur beweisen, dass $\big(f^n(x)\big)_{n\in\mathbb{N}}$ eine Cauchy-Folge ist. Sei also $x \in M$ und $\varepsilon > 0$. Wähle $\tilde{\varepsilon} > 0$ so klein, dass $\tilde{\varepsilon}\,\mathrm{d}(x, f(x)) < \varepsilon$ gilt. Wegen $\sum_{n=1}^{\infty} L_n < \infty$ gibt es ein $n_0 \in \mathbb{N}$ mit der Eigenschaft $\sum_{n=n_0}^{\infty} L_n < \tilde{\varepsilon}$.[27] Für alle $n \in \mathbb{N}$ mit $n \geq n_0$ gilt somit

$$\begin{aligned}\mathrm{d}(f^n(x), f^{n_0}(x)) &\leq \sum_{k=n_0}^{n-1} \mathrm{d}(f^{k+1}(x), f^k(x)) \\ &\leq \sum_{k=n_0}^{n+1} L_k\,\mathrm{d}(f(x), x) \leq \tilde{\varepsilon}\,\mathrm{d}(f(x), x) < \varepsilon.\end{aligned}$$

Also ist $\big(f^n(x)\big)_{n\in\mathbb{N}}$ wie behauptet eine Cauchy-Folge und somit konvergent.

Sei nun $y := \lim_{n\to\infty} f^n(x)$. Weil f Lipschitz-stetig ist, ist f insbesondere stetig, also gilt

$$f^{n+1}(x) = f(f^n(x)) \to f(y)$$

für $n \to \infty$. Zugleich ist aber $\big(f^{n+1}(x)\big)_{n\in\mathbb{N}}$ eine Teilfolge von $\big(f^n(x)\big)_{n\in\mathbb{N}}$, also gilt zugleich $f^{n+1}(x) \to y$ für $n \to \infty$. Da Grenzwerte von Folgen in metrischen Räumen eindeutig sind, folgt $f(y) = y$, das heißt, y ist tatsächlich ein Fixpunkt von f.

Nun können wir die beiden Aussagen (a) und (b) zeigen:

(a) *Existenz:* Da M laut Voraussetzung nichtleer ist, gibt es ein $x \in M$. Somit folgt aus $(*)$, dass f einen Fixpunkt besitzt.

Eindeutigkeit: Seien $x^*, \hat{x} \in M$ Fixpunkte von f, das heißt es sei $f(x^*) = x^*$ und $f(\hat{x}) = \hat{x}$. Wegen $\sum_{n=1}^{\infty} L_n < \infty$ gibt es ein $n \in \mathbb{N}$ mit $L_n < 1$. Es folgt

$$\mathrm{d}(x^*, \hat{x}) = \mathrm{d}\left(f^n(x^*), f^n(\hat{x})\right) \leq L_n\,\mathrm{d}(x^*, \hat{x}).$$

Wegen $L_n < 1$ kann dies nur stimmen, wenn $\mathrm{d}(x^*, \hat{x}) = 0$ ist. Also folgt $x^* = \hat{x}$.

(b) Das folgt aus $(*)$ zusammen mit der soeben bewiesenen Eindeutigkeit des Fixpunktes von f. □

[27] Können Sie dieses Argument im Detail ausführen?

1.5 Addendum: Der Begriff des Vektorraums

Den folgenden Begriff werden Sie in der Vorlesung *Lineare Algebra 1* lernen und im Detail untersuchen. Da wir ihn in der *Analysis 2* ebenfalls benötigen, können Sie die genaue Definition auch hier nachlesen:

Definition 1.5.1 (Vektorraum)**.** Sei $\mathbb{K}$ ein Körper. Eine Menge V zusammen mit zwei Abbildungen $+ : V \times V \to V$ und $\cdot : \mathbb{K} \times V \to V$ nennt man einen *Vekotrraum über* $\mathbb{K}$, wenn sie die folgenden Axiome erfüllen:

(I) *Eigenschaften der Addition:* Es ist V zusammen mit $+$ eine kommutative Gruppe (deren neutrales Element wir mit 0 bezeichnen).

(II) *Assoziativgesetz für die skalare Multiplikation:* Für alle $\lambda, \mu \in \mathbb{K}$ und alle $v \in V$ gilt das Assoziativgesetz

$$\lambda \cdot (\mu \cdot v) = (\lambda\mu)v.$$

(III) *Distributivgesetze:* Für alle $\lambda, \mu \in V$ und alle $v, w \in V$ gilt[28]

$$(\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v \qquad \text{und} \qquad \lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w.$$

(IV) *Nicht-Trivialität der skalaren Multiplikation:* Für alle $v \in V$ gilt $1 \cdot v = v$.

Je nach dem, in welche Quelle sie die Definition eines Vektorraums nachlesen, unterscheidet sich die Anzahl der Axiome, weil man sie auf unterschiedliche Weise zusammenfassen kann. Zum Beispiel fasst das Axiom (I) genau genommen vier Axiome zusammen[29] und man könnte das Axiom (III) in zwei aufteilen, weil es ja genau genommen zwei Eigenschaften enthält. Damit kommt man also insgesamt auf acht Axiome für die Definition eines Vektorraums.

Die folgende Schreibweise ist sehr praktisch und wird fast duchgehend verwendet, wenn man mit Vektorräumen arbeitet.

Konvention 1.5.2 (Weglassen des Malpunktes)**.** Sei V ein Vektorraum über einem Körper $\mathbb{K}$.[30] Für $\lambda \in \mathbb{K}$ und $v \in V$ verwendet man die Kurzschreibweise

$$\lambda v := \lambda \cdot v,$$

ähnlich wie für reelle oder komplexen Zahlen.

[28] Wir verwenden auch in Vektorräumen die übliche Punkt-vor-Strich-Schreibweise, die Sie bereits von den reellen und komplexen Zahlen und allgemeiner von Körpern kennen.

[29] Welche?

[30] Vielleicht fällt Ihnen auf, dass das ein wenig ungenau ist: Eigentlich müsste man auch die Abbildungen $+$ und $\cdot$ erwähnen, das heißt, mam müsste so etwas schreiben wie: "Sei $(V, +, \cdot)$ ein Vektorraum über einem Körper $\mathbb{K}$." Man ist aber häufig bequem und denkt sich die beiden Abbildungen einfach dazu, falls es auf ihre genaue Wahl nicht ankommt.

1.6 Addendum: Details zu Kompaktheit und totaler Beschränktheit

In diesem Abschnitt geben wir einen Beweis von Theorem 1.3.7 und einige weitere verwandte Resultate.

Zunächst beobachten wir, dass man in der Definition von totaler Beschränktheit die Mittelpunkte der überdeckenden Kugeln auch aus der überdeckten Menge selbst wählen kann:

Proposition 1.6.1 (Mittelpunkte in der Definition totaler Beschränktheit)**.** *Sei (M, d) ein metrischer Raum und sei $\emptyset \neq K \subseteq M$. Folgende Aussagen sind äquivalent:*

(i) *Die Menge K ist total beschränkt, das heißt für jedes $\varepsilon > 0$ gibt es eine Zahl $m \in \mathbb{N}$ und Punkte $x_1, \ldots, x_m \in M$ existieren derart, dass*

$$K \subseteq \bigcup_{k=1}^{m} B_{<\varepsilon}(x_k)$$

gilt.[31]

(ii) *Sei (M, d) ein metrischer Raum. Eine Menge $K \subseteq M$ heißt **total beschränkt**, wenn für jedes $\varepsilon > 0$ eine Zahl $m \in \mathbb{N}$ und Punkte $x_1, \ldots, x_m \in K$ existieren*[32] *derart, dass*

$$K \subseteq \bigcup_{k=1}^{m} B_{<\varepsilon}(x_k)$$

gilt.

Beweis. Das folgt leicht aus der Dreieckungleichung. □

Lemma 1.6.2 (Unendlich viele Folgenglieder in einer Kugel)**.** *Sei (M, d) ein metrischer Raum und sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in M und sei die Menge $X := \{x_n \mid n \in \mathbb{N}\}$ der Folgenglieder total beschränkt. Für alle Radien $\varepsilon > 0$ gibt es ein $n_1 \in \mathbb{N}$ derart, dass $B_{<\varepsilon}(x_{n_1})$ unendlich viele Folgenglieder*[33] *von $(x_n)_{n\in\mathbb{N}}$ enthält.*

Beweis. Seien $\varepsilon > 0$. Da X total beschränkt ist, gibt es laut Proposition 1.6.1 ein $m \in \mathbb{N}$ und Indizes $x_{n_1}, \ldots, x_{n_m}$ derart, dass die Kugeln $B_{<\varepsilon}(x_{n_1}), \ldots, B_{<\varepsilon}(x_{n_m})$ die Menge X überdecken. Mindestens eine der Kugeln – sagen wir $B_{<\varepsilon}(x_{n_1})$ (ansonsten ändern wir die Nummerierung) – enthält also unendlich viele Glieder der Folge. □

[31] Hier haben wir einfach noch einmal Definition 1.3.5 wiederholt.

[32] Der einzige Unterschied dieser Aussage zur Definition von totaler Beschränktheit ist, dass die Mittelpunkte $x_1, \ldots, x_m$ nun in der Menge K liegen müssen statt irgendwo in M.

[33] Wenn wir hier "unendlich viele Folgenglieder" sagen, dann unterschieden wir alle Glieder mit verschiedenen Indizes voneinander, selbst wenn sie denselben Wert haben. In diesem Sinnat hat also sogar eine konstante Folge unendlich viele Folgenglieder (die aber alle gleich sind).

Theorem 1.6.3 (Totale Beschränktheit via Cauchy-Folgen)**.** *Sei* (M, d) *ein metrischer Raum und sei* $K \subseteq M$*. Dann sind äquivalent:*

(i) *Die Menge* K *ist total beschränkt.*

(ii) *Jede Folge in* K *besitzt eine Teilfolge, die eine Cauchy-Folge ist.*

Beweis. Falls K leer ist, ist nichts zu zeigen,[34] also nehmen wir für den Rest des Beweises an, dass $K \neq \emptyset$ ist.

"(ii) $\Rightarrow$ (i)" Wir argumentieren per Kontraposition: Sei K nicht total beschränkt. Dann gibt es ein $\varepsilon > 0$ derart, dass K sich nicht durch endlich viele offene Kugeln mit Radius ε überdecken lässt. Deshalb kann man iterativ eine Folge $(x_n)_{n\in\mathbb{N}}$ in K konstruieren derart, dass alle Folgenglieder mindestens Abstand ε voneinander haben. Also ist $(x_n)_{n\in\mathbb{N}}$ keine Cauchy-Folge.

"(i) $\Rightarrow$ (ii)" Gelte (ii) und sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in K. Wir konstruieren rekursive Indizes $n_1 < n_2 < n_3 < \dots$ mit den folgenden Eigenschaften für jedes $k \in \mathbb{N}$: Die Kugel $B_{<\frac{1}{2^k}}(x_{n_k})$ enthält unendlich viele Folgenglieder von $(x_n)_{n\in\mathbb{N}}$ und es gilt $\mathrm{d}(x_{n_k}, x_{n_{k+1}}) < \frac{1}{2^k}$.

Um n_1 zu finden, wenden wir Lemma 1.6.2 für $\varepsilon := \frac{1}{2}$ an: Wegen der totalen Beschränktheit von K ist die Menge der Folgenglieder von $\{x_n \mid n \in \mathbb{N}\}$ ebenfalls total beschränkt, also finden wir laut Lemma ein $n_1 \in \mathbb{N}$ derart, dass unendlich viele Folgenglieder in $B_{<\frac{1}{2}}(x_{n_1})$ liegen.

Seien nun für ein $k \in \mathbb{N}$ die Indizes $n_1 < \dots n_k$ bereits gewählt. Da $B_{<\frac{1}{2^k}}(x_{n_k})$ unendlich viele Glieder der Folge $(x_n)_{n\in\mathbb{N}}$ enthält, gibt es eine Teilfolge von $(x_n)_{n\in\mathbb{N}}$, die erst hinter dem Index n_k beginnt und deren Glieder komplett in der Kugel $B_{<\frac{1}{2^k}}(x_{n_k})$ liegen. Indem wir Lemma 1.6.2 auf diese Teilfolge und auf den Radius $\varepsilon := \frac{1}{2^{k+1}}$ anwenden, finden wir ein $n_{k+1} \in \mathbb{N}$ deart, dass $x_{n_{k+1}} \in B_{<\frac{1}{2^k}}(x_{n_k})$ und somit $\mathrm{d}(x_{n_k}, x_{n_{k+1}}) < \frac{1}{2^k}$ gilt und derart, dass $B_{<\frac{1}{2^{k+1}}}(x_{n_{k+1}})$ unendlich viele Glieder der Teilfolge – also insbesondere unendlich viele Glieder der Folge $(x_n)_{n\in\mathbb{N}}$ selbst – enthält.

Somit haben wir die Indizes $n_1 < n_2 < n_3 < \dots$ mit den gewünschten Eigenschaften gefunden. Es ist $(x_{n_k})_{k\in\mathbb{N}}$ eine Teilfolge von $(x_n)_{n\in\mathbb{N}}$ und zugleich eine Cauchy-Folge: Für alle $k_0, k \in \mathbb{N}$ mit $k \geq k_0$ gilt nämlich

$$\mathrm{d}(x_{n_k}, x_{n_{k_0}}) \leq \sum_{j=k_0}^{k-1} \mathrm{d}(x_{n_{j+1}}, x_{n_j}) \leq \sum_{j=k_0}^{k-1} \frac{1}{2^j} \leq \sum_{j=k_0}^{\infty} \frac{1}{2^j} = \frac{2}{2^{k_0}}.$$

Wenn also $\varepsilon > 0$ gegeben ist, brauchen wir nur $k_0 \in \mathbb{N}$ so groß zu wählen, dass $\frac{2}{2^{k_0}} < \varepsilon$ gilt um die Eigenschaft aus der Definition einer Cauchy-Folge zu erhalten. □

Nun können wir Theorem 1.3.7 aus der Vorlesung beweisen.

[34] Warum?

Beweis von Theorem 1.3.7. "(i) ⇒ (ii)" Gelte (i). Man sieht leicht (wie schon in der Analysis 1), dass jede konvergente Folge Cauchy ist, also ist Bedingung (ii) aus Theorem 1.6.3 erfüllt und somit ist K laut dieses Theorems total beschränkt.

Dass K vollständig ist, sieht man folgendermaßen: Jede Cauchy-Folge in K besitzt wegen (i) eine Teilfolge, die gegen ein $x \in K$ konvergiert. Man zeigt leicht, dass jede Cauchy-Folge mit einer konvergenten Teilfolge selbst gegen denselben Grenzwert konvergiert; also ist (K, d) vollständig.

"(ii) ⇒ (i)" Gelte (ii), das heißt, sei K total beschränkt und vollständig. Sei $(x_n)_{n\in\mathbb{N}}$ eine Folge in K. Wegen der totalen Beschränktheit von K besitzt $(x_n)_{n\in\mathbb{N}}$ laut Theorem 1.6.3 eine Teilfolge, die Cauchy ist. Wegen der Vollständigkeit von K ist diese Teilfolge sogar konvergent gegen einen Punkt in K.

"(iii) ⇒ (ii)" Gelte (iii). Wir zeigen zuerst, dass K total beschränkt ist. Sei dazu $\varepsilon > 0$. Es gilt $K \subseteq \bigcup_{x\in K} B_{<\varepsilon}(x)$. Wegen Eigenschaft (iii) genügen endlich viele dieser offenen Kugeln um K zu überdecken, also ist K total beschränkt.

Nun zeigen wir noch, dass K vollständig ist. Sei $(x_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge in K. Für die Zahlen[35] $r_n := \sup\{\mathrm{d}(x_k, x_n) \mid k \geq n\} \in [0, \infty)$ gilt dann $r_n \to 0$ für $n \to \infty$. Für jedes $n \in \mathbb{N}$ entählt die abgeschlossene Kugel $B_{\leq r_n}(x_n)$ den kompletten Folgenschwanz $(x_k)_{k\geq n}$. Der Durchschnitt von endlich vielen dieser Kugeln enthält somit stets Punkte aus K, also können endlich viele der Komplemten $M \setminus B_{\leq r_n}(x_n)$ nicht K überdecken. Da diese Komplemente offen sind, folgt aus Eigenschaft (iii), dass auch $\bigcup_{n\in\mathbb{N}} M \setminus B_{\leq r_n}(x_n) = M \setminus \bigcap_{n\in\mathbb{N}} B_{\leq r_n}(x_n)$ nicht K überdeckt, also entählt der Durchschnitt $\bigcap_{n\in\mathbb{N}} B_{\leq r_n}(x_n)$ einen Punkt $x \in K$. Wegen

$$\mathrm{d}(x_n, x) \leq r_n \overset{n\to\infty}{\to} 0$$

konvergiert die Folge $(x_n)_{n\in\mathbb{N}}$ gegen x.

"(ii) ⇒ (iii)" Gelte (ii), sei $I \neq \emptyset$ eine Menge und für jedes $i \in I$ sei $U_i \subseteq M$ offen. Wir nehmen an, dass endlich viele der Mengen U_i nicht ausreichen um K zu überdecken und müssen zeigen, dass $\bigcup_{i\in I} U_i \not\supseteq K$ gilt.

Da K total beschränkt ist, können wir K durch endlich viele offene Kugeln mit Radius 1 überdecken und laut Proposition 1.6.1 können wir die Radien dieser Kugeln sogar aus K wählen. Für mindestens eine dieser endlich vielen Kugeln – nennen wir sie $B_{<1}(x_1)$ mit $x_1 \in K$ – kann der Durchschnitt $B_{<1}(x_1) \cap K$ nicht durch endlich viele der Mengen U_i überdeckt werden.

Die Menge $B_{<1}(x_1) \cap K$ als Teilmenge von K ebenfalls total beschränkt.[36] Also kann $B_{<1}(x_1) \cap K$ durch endlich viele Kugeln mit Radius $\frac{1}{2}$ überdeckt werden, wobei die Mittelpunkte laut Proposition 1.6.1 aus $B_{<1}(x_1) \cap K$ gewählt werden können. Für mindestens eine dieser Kugeln – nennen wir sie $B_{<\frac{1}{2}}(x_2)$ mit $x_2 \in B_{<1}(x_1) \cap K$ kann $B_{<\frac{1}{2}}(x_2)$ nicht durch endlich viele der Mengen U_i überdeckt werden.

Wir fahren iterativ so fort und erhalten somit eine Folge $(x_n)_{n\in\mathbb{N}}$ derart, dass für jedes $n \in \mathbb{N}$ die Menge $B_{<\frac{1}{2^n}}(x_n) \cap K$ nicht durch endlich viele der Mengen U_i

[35] Warum sind diese Suprema endlich?

[36] Warum sind Teilmengen total beschränkter Mengen ebenfalls total beschränkt?

überdeckt werden kann und $\mathrm{d}(x_n, x_{n+1}) \leq \frac{1}{2^n}$ gilt. Hieraus folgt, genau wie im Beweis der Implikation "(i) ⇒ (ii)" von Theorem 1.6.3, dass $(x_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge ist. Da K vollständig ist, konvergiert die Folge gegen ein $x \in K$.

Wir zeigen nun, dass x in keiner der Mengen U_i liegt, womit das Argument dann komplett ist. Angenommen es gäbe ein $i_0 \in I$ mit $x \in U_{i_0}$. Da U_{i_0} offen ist, finden wir ein $\varepsilon > 0$ mit $B_{<\varepsilon}(x) \subseteq U_{i_0}$. Wähle $n \in \mathbb{N}$ so groß, dass $\frac{1}{2^n} < \frac{\varepsilon}{2}$ und $\mathrm{d}(x_n, x) < \frac{\varepsilon}{2}$ ist. Dann gilt $B_{<\frac{1}{2^n}}(x_n) \subseteq B_{<\varepsilon}(x) \subseteq U_{i_0}$. Das ist ein Widerspruch, da sich $B_{<\frac{1}{2^n}}(x_n) \cap K$ – und somit insbesondere $B_{<\frac{1}{2^n}}(x_n)$ – nicht durch endlich viele der Mengen U_i (also erst recht nicht durch eine einzige) überdecken lässt. □

1.7 Addendum: Normen auf $\mathbb{K}^d$

Wir wissen bereits aus Proposition 1.2.8, dass auf $\mathbb{K}^d$ eine Folge genau dann bezüglich der p-Norm (für ein $p \in [1, \infty]$) konvergiert, wir ihre Komponenten in $\mathbb{K}$ konvergieren. In diesem Addendum zeigen wir, dass dies sogar für alle Normen auf $\mathbb{K}^d$ stimmt – wir möchten also Theorem 1.3.13 beweisen. Dazu ist der folgende Begriff sehr nürzlich:

Definition 1.7.1 (Äquivalente Normen)**.** Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und sei V ein Vektorraum über $\mathbb{K}$. Zwei Normen $\|\cdot\|_a$, $\|\cdot\|_b$ auf V heißen **äquivalent**, wenn es Zahlen $0 < c < C$ gibt derart, dass

$$c\,\|x\|_a \leq \|x\|_b \leq C\,\|x\|_a$$

für alle $x \in V$ gilt.

Man überlegt sich leicht, dass Aäuqivalenz von Normen eine Äquivalenzrelation auf der Menge aller Normen auf V ist. Wenn zwei Normen auf einem Vektorraum äquivalent sind, dann sind bezüglich dieser beiden Normen dieselben Mengen offen, und eine Folge konvergert in diesem bezgülich der einen Norm gegen einen Punkt $x \in V$ genau dann, wenn Sie bezüglich der anderen Norm gegen x konvergiert.[37]

Um Theorem 1.3.13 zu erhalten, genügt es deshalb, das folgende Resultat zu beweisen:

Theorem 1.7.2 (Alle Normen auf $\mathbb{K}^d$ sind äquivalent)**.** *Seien $d \in \mathbb{N}$ und $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und seien $\|\cdot\|_a$ und $\|\cdot\|_b$ zwei Normen auf $\mathbb{K}^d$. Dann sind $\|\cdot\|_a$ und $\|\cdot\|_b$ äquivalent.*

Beweis. Es genügt eine beliebige, feste Norm $\{\cdot\}$ auf $\mathbb{K}^d$ zu betrachten und zu beweisen, dass sie äquivalent zur 1-Norm $\|\cdot\|_1$ ist.[38] Seien $e_1, \ldots, e_d \in \mathbb{K}^d$ die kanonischen Einheitsvektoren – das heißt für jedes $j \in \{1, \ldots, d\}$ ist der j-te Eintrag von e_j gleich 1 und alle anderen gleich 0. Sei $C := \max\{\|e_1\|, \ldots, \|e_d\|\}$. Für jedes $x \in \mathbb{K}^d$ gilt $x = x_1 e_1 + \cdots + x_d e_d$ und somit

$$\|x\| \leq |x_1|\,\|e_1\| + \cdots + |x_d|\,\|e_d\| \leq |x_1|\,C + \cdots + |x_d|\,C = C\,\|x\|_1\,.$$

[37] Können Sie das im Detail begründen?

[38] Warum genügt das?

Aus dieser Abschätzung folgt, dass $\|\cdot\| : (\mathbb{K}^d, \|\cdot\|_1) \to [0,\infty)$ stetig ist (wobei $[0,\infty)$ wie üblich mit der Euklidischen Metrik ausgestattet ist).[39]

Nun zeigen wir eine entsprechende Abschätzung in die andere Richtung, das heißt, wir zeigen, dass es eine Zahl $c > 0$ gibt derart dass $c\|x\|_1 \leq \|x\|$ für alle $x \in \mathbb{K}^d$ gilt. Diese Richtung ist der interessante Teil des Beweises. Dazu sei $S \subseteq \mathbb{K}^d$ die Einheitssphäre in $\mathbb{K}^d$ bezgülich der 1-Norm, das heißt es sei

$$S := \{x \in \mathbb{K}^d \mid \|x\|_1 = 1\}.$$

Diese Menge ist in $(\mathbb{K}^d, \|\cdot\|_1)$ abgeschlossen und beschränkt,[40] also ist sie laut Theorem 1.3.10 kompakt. Wegen der soebenen besprochenen Stetigkeit von $\|\cdot\|$ gibt es laut Korollar 1.3.12 ein $x_{\min} \in S$ derart, dass $\|x_{\min}\| \leq \|x\|$ für alle $x \in S$ gilt. Wir setzen $c := \|x_{\min}\|$. Da kein Vektor in S Null ist und die Norm $\|\cdot\|$ positiv definit ist, gilt $c > 0$.

Sei nun $x \in \mathbb{K}^d$. Falls $x = 0$ ist, gilt natürlich $c\|x\|_1 \leq \|x\|$, da beide Seiten der Ungleichung in diesem Fall 0 sind. Sei nun also $x \neq 0$. Dann ist $\|x\|_1 \neq 0$ und $x/\|x\|_1 \in S$, also gilt

$$c = \|x_{\min}\| \leq \left\| \frac{x}{\|x\|_1} \right\| = \frac{1}{\|x\|_1} \|x\|$$

und somit $c\|x\|_1 \leq \|x\|$. □

[39] Können Sie dieses Stetigkeitsargument im Detail ausführen?

[40] Warum?

Kapitel 2

Kurven im $\mathbb{K}^d$

2.1 Kurven und ihre Ableitungen

Um folgenden gehen wir stets davon aus, dass $\mathbb{K}^d$ mit einer Norm $\|\cdot\|$ ausgestattet ist. Wegen Theorem 1.3.13 ist es für den Konvergenzbegriff auf $\mathbb{K}^d$ egal, welche Norm genau wir hierfür verwenden. Wegen Theorem 1.2.14 ist es somit auch für die Stetigkeit von Funktionen zwischen $\mathbb{K}^c$ und $\mathbb{K}^d$ (für $c, d \in \mathbb{N}$) egal, mit welchen Normen wir $\mathbb{K}^c$ und $\mathbb{K}^d$ ausstatten.

Definition 2.1.1 (Kurve)**.** Sei $d \in \mathbb{N}$, sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und sei $I \subseteq \mathbb{R}$ ein Intervall mit mehr als einem Punkt. Eine Abbildung $\gamma : I \to \mathbb{K}^d$ nennen wir eine **Kurve**.

Man beachte, dass es für eine Funktion $\gamma : I \to \mathbb{K}^d$ stets Funktionen $\gamma_1, \gamma_2, \ldots, \gamma_d : I \to \mathbb{K}$ gibt derart, dass

$$\gamma(t) = \begin{pmatrix} \gamma_1(t) \\ \gamma_2(t) \\ \vdots \\ \gamma_d(t) \end{pmatrix}$$

für alle $t \in I$ gilt. Man nennt diese Funktionen die **Komponentenfunktionen** von γ. Die Funktion γ ist genau dann stetig, wenn all ihre Komponentenfunktionen $\gamma_1, \ldots, \gamma_d$ stetig sind.[1]

Beispiele 2.1.2 (Einige Kurven)**.**

(a) Die Abbildung

$$\begin{aligned} \gamma : [0, 2\pi] &\to \mathbb{R}^2, \\ t \mapsto \gamma(t) &:= \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \end{aligned}$$

[1] Warum?

ist eine Kurve. Sie durchläuft die Einheitskreislinie (bezüglich der Euklidischen Norm) in $\mathbb{R}^2$.

(b) Sei $I \subseteq \mathbb{R}$ ein Intervall mit mehr als einem Punkt und sei $f : I \to \mathbb{R}$. Dann ist die Abbildung

$$\gamma : [0, 2\pi] \to \mathbb{R}^2,$$
$$t \mapsto \gamma(t) := \begin{pmatrix} t \\ f(t) \end{pmatrix}$$

eine Kurve. Sei durchläuft den Graphen von f. Beachten Sie: Falls f stetig ist, ist γ ebenfalls stetig.[2]

Definition 2.1.3 (Ableitung von Kurven)**.** Sei $d \in \mathbb{N}$, sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und sei $I \subseteq \mathbb{R}$ ein Intervall mit mehr als einem Punkt und sei $\gamma : I \to \mathbb{K}^d$. Sei $t_0 \in I$.

(a) Wir nennen γ **differenzierbar** in t_0, wenn all ihre Komponentenfunktionen $\gamma_1, \ldots, \gamma_d$ im Punkt t_0 differenzierbar sind. In diesem Fall nennen wir den Vektor

$$\dot{\gamma}(t_0) := \gamma'(t_0) := \begin{pmatrix} \gamma_1'(t_0) \\ \gamma_2'(t_0) \\ \vdots \\ \gamma_d'(t_0) \end{pmatrix}$$

die **Ableitung** von γ im Punkt t_0.

(b) Wir nennen γ **differenzierbar**, wenn sie in jedem Punkt aus I differenzierbar ist. In diesem Fall nennen wir die Funktion $\dot{\gamma} : I \to \mathbb{K}^d$, $t \mapsto \dot{\gamma}(t)$ die **Ableitung** von γ.

(c) Wir nennen γ **stetig differenzierbar**, wenn sie differenzierbar ist und die Funktion $\dot{\gamma}$ stetig ist. Die Menge aller stetig differenzierbaren Funktionen von I nach $\mathbb{K}^d$ bezeichnen wir mit $\mathrm{C}^1(I, \mathbb{K}^d)$.

(d) Genau wie in Analysis 1 definiert man analog auch höhere Ableitungen und die Räume[3] $\mathrm{C}^k(I, \mathbb{K}^d)$ für $k \in \mathbb{N}$.

Vorlesung 8: (Mi, 07.05.) beginnt mit Beispiel 2.1.4

Beispiel 2.1.4 (Die Ableitung einer vektorwertigen Funktion)**.** Betrachten wir die Kurve

$$\gamma : [0, 2\pi] \to \mathbb{R}^2,$$
$$t \mapsto \gamma(t) := \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}$$

[2] Warum?

[3] Es handelt sich hierbei um Vektorräume über $\mathbb{K}$.

aus Beispiel 2.1.2(a). Da cos und sin differenzierbar sind, ist laut Definition 2.1.3(a) auch die Kurve γ differenzierbar, und es gilt

$$\dot{\gamma}(t) = \begin{pmatrix} \dot{\cos}(t) \\ \dot{\sin}(t) \end{pmatrix} = \begin{pmatrix} -\sin(t) \\ \cos(t) \end{pmatrix}$$

für alle $t \in [0, 2\pi]$.

Analog zu Ableitungen definiert man auch das Riemann-Integral für Funktionen mit Werten in $\mathbb{K}^d$ komponentenweise:

Definition 2.1.5 (Das vektorwertige Riemann-Integral)**.** Sei $d \in \mathbb{N}$, sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und seien $a, b \in \mathbb{R}$ mit $a \leq b$. Sei $\gamma : [a, b] \to \mathbb{K}^d$.

Wir nennen γ **Riemann-integrierbar**, wenn all ihre Komponentenfunktionen $\gamma_1, \ldots, \gamma_d$ Rieman-integrierbar sind. In diesem Fall nennen wir

$$\int_a^b \gamma(t)\,\mathrm{d}t := \begin{pmatrix} \int_a^b \gamma_1(t)\,\mathrm{d}t \\ \int_a^b \gamma_2(t)\,\mathrm{d}t \\ \vdots \\ \int_a^b \gamma_d(t)\,\mathrm{d}t \end{pmatrix}$$

das **Riemann-Integral** von γ. Wie für skalarwertige Funktionen setzen wir auch in diesem Fall $\int_b^a \gamma(t)\,\mathrm{d}t := -\int_a^b \gamma(t)\,\mathrm{d}t$.

In dem man den Hauptsatz der Differential- und Integralrechnung komponentenweise anwendet, erhält man sofort:

Bemerkung 2.1.6 (Der Hauptsatz im vektorwertigen Fall)**.** Der Hauptsatz der Differential- und Integralrechnung (siehe [4, Theoreme 6.2.3 und 6.2.4]) gilt auch für Funktionen, die Werte in $\mathbb{K}^d$ annehmen.

Bemerkungen 2.1.7 (Differentienquotienten und Riemannsummen)**.**

(a) Sei $I \subseteq \mathbb{R}$ ein Intervall mit mehr als einem Punkt und sei $\gamma : I \to \mathbb{K}^d$ eine Kurve. Sei $t_0 \in I$. Mit Hilfe von Definition 2.1.3 und der Definition der Ableitung von skalarwertigen Funktionen kann man sich leicht überlegen, dass γ genau dann im Punkt t_0 differenzierbar ist, wenn

$$\frac{\gamma(t_0 + s) - \gamma(t_0)}{s}$$

für $s \to 0$ konvergiert, und in diesem Fall gilt

$$\dot{\gamma}(t_0) = \lim_{s \to 0} \frac{\gamma(t_0 + s) - \gamma(t_0)}{s}.$$

Das heißt, man kann auch die Ableitung einer Kurve in einem Punkt als Grenzwert von Differenzenquotienten auffassen. Anschaulich kann man das so interpretieren, dass der Vektor $\dot{\gamma}(t_0)$ zum Zeitpunkt t_0 tangential an der Kurve γ anliegt.

Sie sehen, dass wir hier – wie man es sehr oft tut – das Argument einer Kurve als **Zeit** interpretieren. Die Kurve kann man sich dabei so vorstellen, dass ein Punkt im Laufe der Zeit einen Weg durch $\mathbb{K}^d$ entlang läuft. Unter dieser Interpretation ist $\dot{\gamma}(t_0)$ der **Geschwindigkeitsvektor** des Punktes zum Zeitpunkt t_0.

(b) Seien $a, b \in \mathbb{R}$ mit $a \leq b$ und sei $\gamma : [a, b] \to \mathbb{K}^d$. Anstatt komponentenweise vorzugehen, kann man die Riemann-Integrierbarkeit und das Riemann-Integral von γ auch direkt über Riemann-Summen definieren – genau wie wir es in Analysis 1 im skalarwertigen Fall getan haben – und kommt dabei zumselben Ergebnis.

Insbesondere kann man, wenn γ Riemann-integrierbar ist, das Integral über γ schreiben als

$$\int_a^b \gamma(t)\,\mathrm{d}t = \lim_{n\to\infty} \mathcal{R}(p_n, q_n, \gamma),$$

wobei $(p_n)_{n\in\mathbb{N}}$ eine Folge von Partionen von $[a, b]$ ist, deren Feinheit gegen 0 geht and q_n eine Folge von zugehörigen Punktierungen ist. Die Riemann-Summen $\mathcal{R}(p_n, q_n, \gamma)$ sind hierbei genauso definiert wie im skalarwertigen Fall.

Mit Hilfe von Bemerkung 2.1.7(b) kann man die folgende Abschätzung beweisen. Ein skalarwertiges Analogon hierzu kennen Sie bereits aus der Analysis 1.

Proposition 2.1.8 (Fundamentalabschätzung für vektorwertige Integrale)**.** *Seien $a, b \in \mathbb{R}$ mit $a \leq b$ und sei $\gamma : [a, b] \to \mathbb{K}^d$ stetig. Dann gilt für jede Norm $\|\cdot\|$ auf $\mathbb{K}^d$*

$$\left\| \int_a^b \gamma(t)\,\mathrm{d}t \right\| \leq \int_a^b \|\gamma(t)\|\,\mathrm{d}t.$$

Beweisskizze. Weil $\|\cdot\|$ stetig ist,[4] ist $\|\gamma(\,\cdot\,)\| : [a, b] \to \mathbb{R}$ stetig und somit Riemann-integrierbar. Also ist die rechte Seite der Ungleichung definiert. Die linke Seite ist ebenfalls definiert, da γ als stetige Funktion Riemann-integrierbar ist.

Wenn man die Integrale in der behaupteten Ungleichung durch Riemann-Summen ersetzt, gilt die Ungleichung aufgrund der Dreiecksungleichung für die Norm $\|\cdot\|$. Durch Grenzübergang erhält man dann die Ungleichung für die Integrale anstelle der Riemann-Summen. □

Definition 2.1.9 (Länge einer Kurve)**.** Sei $d \in \mathbb{N}$, sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und seien zudem $a, b \in \mathbb{R}$ mit $a \leq b$. Sei $\gamma : [a, b] \to \mathbb{K}^d$ stetig differenzierbar. Dann nennt man die Zahl

$$\int_a^b \|\dot{\gamma}(t)\|_2\,\mathrm{d}t$$

die **Bogenlänge** von γ.

[4] Warum ist diese Abbildung stetig?

In der Analysis 1 haben Sie gesehen, dass die Abbildung (damals hatten wir sie noch nicht als "Kurve" bezeichnet)

$$\begin{aligned} \gamma : [0, 2\pi] &\to \mathbb{C}, \\ t &\mapsto \exp(it) \end{aligned} \tag{2.1.1}$$

die komplexe Einheitskreislinie genau einmal durchläuft. Wenn wir $\mathbb{C}$ mit $\mathbb{R}^2$ identifizieren (wobei die erste Komponente in $\mathbb{R}^2$ den Realteil einer Zahl aus $\mathbb{C}$ beschreibt und die zweite Komponente den Imaginärteil), dann können wir die Abbildung auch in der Form

$$\begin{aligned} \gamma : [0, 2\pi] &\to \mathbb{R}^2, \\ t &\mapsto \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \end{aligned}$$

schreiben.

Wenn Sie zwei Punkte auf einer Kreislinie durch Strecken mit dem Mittelpunkte verbinden, dann ist der Winkel zwischen diesen beiden Strecken definiert als die Länge des Kreissegments zwischen den beiden Punkten geteilt durch den Kreisradius. Entsprechend ist der Winkel einer komplexen Zahlen $z \neq 0$ folgendermaßen definiert:

Definition 2.1.10 (Der Winkel einer komplexen Zahl)**.** Sei $z \in \mathbb{C} \setminus \{0\}$. Der **Winkel** oder das **Argument** von z ist die Länge der Linie auf dem Einheitskreis zwischen 1 und $\frac{z}{|z|}$, wenn man den Einheitskreis entgegen des Uhrzeigersinn durchläuft.

Wie in der Analysis 1 besprochen, kann man jede komplexe Zahl $z \neq 0$ schreiben als $z = |z| \exp(it)$ für genau ein $t \in [0, 2\pi)$. Es liegt nahe zu fragen, wie diese Zahl t mit dem Winkel von z zusammenhängt. Die folgende Proposition gibt die Antwort:

Proposition 2.1.11 (Komplexe $\boldsymbol{e}$-Funktion und Winkel)**.** *Sei $z \in \mathbb{C} \setminus \{0\}$ und sei $t \in [0, 2\pi)$ diejenige Zahl, für die*

$$z = |z| \exp(\mathrm{i}t)$$

gilt. Dann ist der Winkel von z gleich t.

Beweis. Wir betrachten die Kurve

$$\begin{aligned} \gamma : [0, t] &\to \mathbb{C}, \\ s &\mapsto \exp(\mathrm{i}s). \end{aligned}$$

Diese durchläuft laut Analysis 1 den komplexen Einheitskreis entgegen des Uhrzeigersinns. Sie beginnt bei der Zahl 1 und stoppt bei der Zahl $\exp(it) = \frac{z}{|z|}$. Also ist der Winkel von z gleich der Länge der Kurve γ und somit gleich

$$\int_0^t \|\dot{\gamma}(s)\|_2 \, \mathrm{d}s = \int_0^t |\mathrm{i} \exp(\mathrm{i}s)| \, \mathrm{d}s = \int_0^t 1 \, \mathrm{d}s = t,$$

womit die Behauptung bewiesen ist. □

2.2 Gewöhnliche Differentialgleichunugen – eine Kurzeinführung

Vorlesung 9: (Fr, 09.05.) beginnt mit Beispiel 2.2.1

Beispiel 2.2.1 (Motivation: Wachstum einer Bakterienkultur)**.** Für jeden Zeitpunkt $t \geq 0$ beschreibe $b(t)$ die Menge der Bakterien in einer Bakterienkultur, die sich in einer Nährlösung vermehrt. Jedes Bakterium vermehrt sich durch Zellteilung: Je mehr Bakterien bereits vorhanden sind, desto mehr neue kommen hinzu. Deshalb die die zeitliche Änderungsrate von $b(t)$ proportional zur Größe $b(t)$, das heißt, es gibt eine Zahl $c > 0$ derart, dass

$$\dot{b}(t) = c\, b(t)$$

für alle betrachteten Zeiten $t \geq 0$ gilt.[5] Hierbei handelt es sich um eine sogenannte **Differentialgleichung**. Die Unbekannte ist hier keine Zahl, sondern eine Funktion $b : [0, \infty) \to [0, \infty)$ und der Begriff Differentialgleichung rührt daher, dass die Ableitung der gesuchten Funktion b in der Gleichung auftaucht.

Damit man eine Chance hat die zeitliche Entwicklung der Bakterienkultur vorherzusehen, genügt diese Differentialgleichung natürlich nicht, sondern man muss auch noch wissen, wieviele Bakterien zum Startzeitpunkt $t = 0$ vorhanden sind; wenn wir diese Anzahl b_0 nennen, haben wir also die beiden Gleichungen

$$\begin{aligned} &\dot{b}(t) = c\, b(t) \quad \text{für } t \geq 0, \\ &b(0) = b_0. \end{aligned}$$

Diese beiden Gleichungen zusammengenommen bezeichnet man als ein **Anfangswerproblem**.[6]

Die Differentialgleichungen in diesem Abschnitt bezeichnet man übrigens als **gewöhnliche Differentialgleichungen**, weil in ihnen nur die Ableitung bezüglich einer einzigen Variablen auftaucht.[7]

Beispiel 2.2.2 (Motivation: Mathematisches Pendel)**.** Lassen Sie uns ein starres Pendel und Länge ℓ betrachten. Wir nehmen vereinfachend an, dass die Masse des Pendels komplett in einem Punkt an seinem Ende konzentriert ist (und dass die starre Verbindung zur Aufhängung somit eine vernachlässigbar kleine Masse hat).

[5] Achtung: In Wirklichkeit ist das natürlich eine stark vereinfachende Annahme, die nicht immer stimmt. Insbesondere gehen wir hier davon aus, dass die Propotionalitätskonstante c nicht von der Zeit abhängt. In Wirklichkeit tut sie das aber natürlich. Spästestens wenn die Nährlösung zur Neige geht, wird c abnehmen (und eventuell sogar negativ werden, wenn Bakterien sterben und weniger neue hinzukommen als sterben).

Die Aufgabe ein reales Phänomen durch eine mathematische Gleichung zu beschreiben nennt man **Modellierung**.

[6] Übrigens: Die Anzahl der Bakterien ist genau genommen eine ganze Zahl. Wie kann es möglich sein, die Entwicklung einer ganzzahligen Größe durch eine Differentialgleichung zu beschreiben? Geht da nicht etwas schief?

[7] Und was Ableitungen nach mehreren verschiedenen Variablen sind, wissen Sie genau genommen noch gar nicht. Das ist nämlich der Inhalt von Kapitel 4.

Außerdem vernachlässigen wir im folgenden Modell jegliche Reibung. Die durch die Schwerkraft der Erde verursachte Fallbeschleunigung[8] bezeichnen wir mit g.[9]

Mit $\varphi(t) \in \mathbb{R}$ bezeichnen wir den Auslenkungswinkel des Pendels zum Zeitpunkt t (im Vergleich zur Lage, in der das Pendel nach unten hängt). Die zeitliche Änderungsrate des Auslenkunkgswinkels nennt man die Winkelgeschwindigkeit und bezeichnet sie mit $\omega(t)$. Es ist also $\omega(t) = \dot{\varphi}(t)$ für alle $t \geq 0$.

Mit Hilfe einer Skizze, einer Zerlegung der am Pendelende angreifenden Kräfte, und des zweiten Newtonschen Gesetzes erhält man, dass φ und ω die Differentialgleichung

$$\begin{pmatrix} \dot{\varphi}(t) \\ \dot{\omega}(t) \end{pmatrix} = \begin{pmatrix} \omega(t) \\ -\frac{g}{\ell} \sin(\varphi(t)) \end{pmatrix}$$

für alle $t \geq 0$.[10] Für diese Differentialgleichung ist die Unbekannte eine Funktion

$$\begin{pmatrix} \varphi \\ \omega \end{pmatrix} : [0, \infty) \to \mathbb{R},$$

also eine Kurve. Anschaulich liegt nahe: Um das Verhalten der Lösung vorhersagen zu können – mathematisch gesprochen: um Eindeutigkeit der Lösung zu erhalten – sollte man die Werte von φ und ω zum Startzeitpunkt kennen.

Theorem 2.2.3 (Existenz und Eindeutigkeit für global Lipschitz-stetige Vektorfelder)**.** *Sei $f : \mathbb{K}^d \to \mathbb{K}^d$ Lipschitz-stetig bezüglich der Euklidischen Norm $\|\cdot\|_2$,*[11] *und sei $x_0 \in \mathbb{K}^d$. Es gibt genau eine stetig differenzierbare Funktion $x : [0, \infty) \to \mathbb{K}^d$, die*

$$\begin{cases} \dot{x}(t) &= f(x(t)) \quad \text{für alle } t \in [0, \infty), \\ x(0) &= x_0 \end{cases}$$

erfüllt.

Beweis. Sei $b \in [0, \infty)$. Es genügt zu zeigen, dass es genau eine stetig differenzierbare Funktion $x : [0, b] \to \mathbb{K}^d$ gibt, die $x(0) = x_0$ und $\dot{x}(t) = f(x(t))$ für alle $t \in [0, b]$ erfüllt.[12] Wegen des Hauptsatz der Differential- und Integralrechnung (für

[8] Manchmal auch Ortsfaktor genannt.

[9] Der Wert von g beträgt knapp unter $10 \frac{m}{s^2}$. Der genaue Wert hängt von der genauen Position auf der Erde ab.

[10] Es ist übrigens bemerkenswert, dass die Masse des Pendels in der Differentialgleichung nicht auftaucht. Der Grund ist, dass die Masse zum einen als träge Masse, welche die Beschleunigung verzögert, in das zweiten Newtonsche Gesetz eingeht, und zum anderen als schwere Masse, die durch die Gravitation beschleunigt wird, in die angreifende Kraft eingeht. Führt man die Rechnung im Detail aus, so sieht man, dass die Massen sich dadurch am Ende wegkürzt.

[11] Genau genommen spielt es wegen Theorem 1.7.2 keine Rolle bezüglich welcher Norm auf $\mathbb{K}^d$ wir die Lipschitz-Stetigkeit annehmen. Bei Wechsel der Norm könnten sich die Lipschitz-Konstanten ändern, aber nicht die Frage, ob eine Funktion Lipschitz-stetig ist.

[12] Warum genügt das?

Funktionen mit Werten in $\mathbb{K}^d$, siehe Bemerkung 2.1.6) sind diese beiden Bedingungen an x äquivalent dazu, dass $x : [0,b] \to \mathbb{K}^d$ stetig ist und die Integralgleichung $x(t) = \int_0^t f(x(s))\,\mathrm{d}s + x_0$ für alle $t \in [0,b]$ erfüllt.

Bezeichne nun $\mathrm{C}([0,b];\mathbb{K}^d)$ die Menge aller stetigen Kurven von $[0,b]$ nach $\mathbb{K}^d$. m Für jedes $x \in \mathrm{C}([0,b];\mathbb{K}^d)$ setzen wir

$$\|x\|_\infty := \sup_{t\in[0,b]} \big\{\, \|x(t)\|_2 \mid\ t \in [0,b] \big\}.$$

Dann kann man genau wie im skalarwertigen Fall $d = 1$ auch für allgemeines d zeigen, dass $\|\cdot\|_\infty$ eine Norm auf $\mathrm{C}([0,b];\mathbb{K}^d)$ ist und dass $\mathrm{C}([0,b];\mathbb{K}^d)$ bezüglich dieser Norm vollständig ist.

Betrachte nun die Abbildung $T : \mathrm{C}([0,b];\mathbb{K}^d) \to \mathrm{C}([0,b];\mathbb{K}^d)$, welche jedes $x \in \mathrm{C}([0,b];\mathbb{K}^d)$ auf die Kurve $Tx \in \mathrm{C}([0,b];\mathbb{K}^d)$ schickt, die durch

$$(Tx)(t) = x_0 + \int_0^t f(x(s))\,\mathrm{d}s$$

für alle $t \in [0,b]$ gegeben ist. Um das Theorem zu beweisen, müssen wir zeigen, dass T genau einen Fixpunkt in $\mathrm{C}([0,b];\mathbb{K}^d)$. besitzt. Hierzu werden wir den Fixpunktsatz von Banach verwenden.

Wir zeigen per Induktion über n, dass für jedes $n \in \mathbb{N}_0$, alle $x, y \in \mathrm{C}([0,b];\mathbb{K}^d)$ und alle $t \in [0,b]$ die Ungleichung

$$\|(T^n x)(t) - (T^n y)(t)\|_2 \le \frac{t^n L^n}{n!}\,\|x-y\|_\infty \tag{2.2.1}$$

gilt, wobei $L \in [0,\infty)$ eine Lipschitz-Konstante von f bezeichnet. Für $n = 0$ folgt die Behauptung sofort aus der Definition der Norm $\|\cdot\|_\infty$ auf $\mathrm{C}([0,b];\mathbb{K}^d)$. Sei die Behauptung jetzt für eine beliebiges, festes $n \in \mathbb{N}_0$ bereits bewiesen. Für alle $x, y \in \mathrm{C}([0,b];\mathbb{K}^d)$ and alle $t \in [0,b]$ folgt dann

$$\begin{aligned}
\big\|(T^{n+1}x)(t) - (T^{n+1}y)(t)\big\|_2 &= \big\|T\big(T^n x\big)(t) - T\big(T^n y\big)(t)\big\|_2 \\
&= \left\|\int_0^t f((T^n x)(s))\,\mathrm{d}s - \int_0^t f((T^n y)(s))\,\mathrm{d}s\right\|_2 \\
&\le \int_0^t \left\|f((T^n x)(s))\,\mathrm{d}s - \int_0^t f((T^n y)(s))\right\|_2 \mathrm{d}s \\
&\le \int_0^t L\,\|(T^n x)(s) - (T^n y)(s)\|_2\,\mathrm{d}s \\
&\le \int_0^t L\frac{s^n L^n}{n!}\,\|x-y\|_\infty\,\mathrm{d}s = \frac{t^{n+1}L^{n+1}}{(n+1)!}\,\|x-y\|_\infty\,;
\end{aligned}$$

für die ersten Ungleichung haben wir die Fundamentalabschätzung für Integrale aus Proposition 2.1.8 verwendet, für die zweite Ungleichung die Lipschitz-Stetigkeit von f, und für die dritte Ungleichung die Induktionsannahme. Die Ungleichung (2.2.1) ist somit für alle $n \in \mathbb{N}_0$ gezeigt. bewiesen.

Indem man in der Ungleichung (2.2.1) das Supremum über alle $t \in [0, b]$ nimmt, erhält man $\|T^n x - T^n y\|_\infty \leq \frac{b^n L^n}{n!} \|x - y\|_\infty$ für alle $x, y \in \mathrm{C}([0, b]; \mathbb{K}^d)$ und alle $n \in \mathbb{N}_0$. Also ist T^n für jedes $n \in \mathbb{N}_0$ Lipschitz-stetig mit Lipschitz-Konstante $L_n := \frac{b^n L^n}{n!}$. Es gilt $\sum_{n=0}^{\infty} \frac{b^n L^n}{n!} = \exp(bL) < \infty$, also sind für die Abbildung T auf dem vollständigen metrischen Raum $\mathrm{C}([0, b]; \mathbb{K}^d)$ die Voraussetzungen des Banachschen Fixpunktsatzes (Theorem 1.4.3) erfüllt; dieses Theorem sagt uns, dass T genau einen Fixpunkt in $\mathrm{C}([0, b]; \mathbb{K}^d)$ besitzt. □

Theorem 2.2.3 wird häufig als der **Existenz- und Eindeutigkeitssatz von Picard–Lindelöf** bezeichnet, oder genauer: als die **globale** Variante dieses Satzes. Eine lokale Variante des Theorems (deren Beweis ähnlich, aber deutlich technischer ist), werden wir in der Analysis 3 besprechen.

Kapitel 3

Integralrechnung in mehreren Veränderlichen

3.1 Das Banach–Tarski-Paradoxon

Vorlesung 10: (Mi, 14.05.) beginnt mit Diskussion 3.1.1

Diskussion 3.1.1 (Eigenschaften eines "vernünftigen" Volumenbegriffs)**.** Angenommen, jede Teilmenge $C \subseteq \mathbb{R}^3$ besitzt Volumen $\mathrm{Vol}(C)$. Für viele Mengen ist zunächst vermutlich gar nicht klar, was genau ihr Volumen bedeuten soll. Zumindest die folgenden Forderungen an einen "vernünftigen" Volumenbegriff liegen aber nahe:

(1) Für zwei disjunkte Mengen $C_1, C_2 \subseteq \mathbb{R}^3$ sollte stets $\mathrm{Vol}(C_1 \cup C_2) = \mathrm{Vol}(C_1) + \mathrm{Vol}(C_2)$ gelten.

(2) Allgemeiner sollte für jede Folge von Mengen $C_1, C_2, C_3, \cdots \subseteq \mathbb{R}^3$, die $C_j \cap C_k = \emptyset$ für alle $j, k \in \mathbb{N}$ mit $j \neq k$ erfüllt, die Eigenschaft $\mathrm{Vol}(\bigcup_{n\in\mathbb{N}} C_n) = \sum_{n=1}^{\infty} \mathrm{Vol}(C_n)$ gelten.

Dieses Eigenschaft ist wichtig, damit man das Volumen einer komplizierten Mengen berechnen kann, in dem man die Menge durch einfachere Mengen "ausschöpft".

(3) Das Volumen einer Menge sollte sich nicht ändern, wenn man die Menge verschiebt, dreht oder spiegelt.

(4) Für "einfache" Mengen wie beispielsweise Quader sollte das Volumen mit der bekannten Volumenformel übereinstimmen.

Der folgende Satz zeigt, dass es nicht möglich ist jeder Teilmenge von $\mathbb{R}^3$ ein Volumen zuzuordnen und zugleich die Eigenschaften (1), (3) und (4) aus Diskussion 3.1.1 zu erfüllen.

Theorem 3.1.2 (Banach–Tarski-Paradoxon)**.** *Sei $\mathbb{R}^3$ mit der Euklidischen Norm $\|\cdot\|_2$ ausgestattet. Es gibt es eine Zahl $n \in \mathbb{N}$ und Abbildungen $T_1, \ldots T_n : \mathbb{R}^3 \to \mathbb{R}^3$ sowie Mengen $C_1, \ldots, C_n \subseteq \mathbb{R}^3$ mit folgenden Eigenschaften:*

(a) *Die Mengen $C_1, \dots C_n$ sind disjunkt und ihre Vereinigung ist eine offene Kugel mit Radius* 1.

(b) *Jeder der Abbildungen $T_1, \dots, T_n$ ist die Hintereinanderausführung einer Drehung und einer Verschiebung.*

(c) *Die Bildmengen $T_1(C_1), \dots, T_n(C_n)$ sind ebenfalls disjunkt und ihre Vereinigung besteht aus zwei disjunkten offenen Kugeln mit Radius* 1.

Beweis. Siehe zum Beispiel [1, S. 271]. □

Theorem 3.1.2 hat folgende Konsequenz: Es ist nicht möglich allen im Theorem vorkommenden Mengen $C_1, \dots, C_n$ vorkommenden Menge ein Volumen zuzuordnen und zugleich die Eigenschaften (1), (3) und (4) aus Diskussion 3.1.1 zu erfüllen. Nehmen wir nämlich an, dass dies möglich wäre, und bezeichne μ das Volumen einer (und somit jeder) Kugel mit Radius 1. Wäre würde

$$\begin{aligned}\mu &= \mathrm{Vol}(C_1 \cup \dots \cup C_n) = \mathrm{Vol}(C_1) + \dots + \mathrm{Vol}(C_n) \\ &= \mathrm{Vol}(T(C_1)) + \dots + \mathrm{Vol}(T(C_n)) = \mathrm{Vol}(T(C_1) \cup \dots T(C_n)) = 2\mu\end{aligned}$$

folgen und somit $\mu = 0$ gelten, was aber Eigenschaft (4) widerspricht.

3.2 Messbare Mengen im $\mathbb{R}^n$ und das Lebesgue-Maß

Das Banach–Tarksi-Paradoxon (Theorem 3.1.2) zeigt, dass man nicht jeder Menge in $\mathbb{R}^3$ ein Volumen zuordnen kann. Wenn man dennoch einen nützlichen Volumenbegriff erhalten möchte, kann man ihn also nur für manche Teilmengen von $\mathbb{R}^3$ definieren. Um die Mengen zu identifizieren, denen man vernünftiger Weise ein Volumen zuordnen kann, geht man axiomatisch vor: Man möchte gerne, dass zumindest alle offenen Mengen ein Volumen haben und man will, dass mit denjenigen Mengen, die ein Volumen haben, bestimmte mengentheoretische Operationen möglich sind. Das führt zu folgender Definition; wir beschränken uns in ihr nicht nur auf die Dimension 3, sondern gucken uns direkt den Raum $\mathbb{R}^d$ für eine beliebige Dimension $d \in \mathbb{N}$ an.

Definition 3.2.1 (Borel-messbare Mengen)**.** Sei $d \in \mathbb{N}$ und sei $\mathbb{R}^d$ mit einer beliebigen Norm ausgestattet. Mit $\mathcal{B}(\mathbb{R}^d)$ bezeichnen wir die kleinste Teilmenge der Potenzmenge $2^{\mathbb{R}^d}$ von $\mathbb{R}^d$, die die folgenden Eigenschaften erfüllt:

(I) Für jede offene Menge $U \subseteq \mathbb{R}^d$ gilt $U \in \mathcal{B}(\mathbb{R}^d)$.

(II) Wenn für jedes $n \in \mathbb{N}$ eine Menge $C_n \in \mathcal{B}(\mathbb{R}^d)$ gegeben ist, dann gilt auch $\bigcup_{n\in\mathbb{N}} C_n \in \mathcal{B}(\mathbb{R}^d)$.

(III) Für jedes Menge $C \in \mathcal{B}(\mathbb{R}^d)$ ist auch das Komplement $\mathbb{R}^d \setminus C$ ein Element von $\mathcal{B}(\mathbb{R}^d)$.

Man nennt $\mathcal{B}(\mathbb{R}^d)$ die **Borel-σ-Algebra** von $\mathbb{R}^d$.[1] Eine Menge $C \subseteq \mathbb{R}^d$ nennt man **Borel-messbar** oder kurz **messbar**, wenn $C \in \mathcal{B}(\mathbb{R}^d)$ gilt.

Man kann zeigen, dass es solch eine kleinste Teilmenge von $2^{\mathbb{R}^d}$ tatsächlich gibt. Hierzu schneidet man alle Teilmengen von $2^{\mathbb{R}}$, die die Axiome (I)–(III) erfüllen und zeigt, dass dieser Durchschnitt die selben Axiome erfüllt. Der Durchschnitt ist ist dann die gewünschte Menge $\mathcal{B}(\mathbb{R}^d)$. Mehr Details hierzu besprechen wir in der Analysis 3, wenn wir allgemeine σ-Algebren behandeln.

Proposition 3.2.2 (Der Durchschnitt messbarer Mengen)**.** *Sei $d \in \mathbb{N}$ und für jedes $n \in \mathbb{N}$ sei eine Menge $C_n \in \mathcal{B}(\mathbb{R}^d)$ gegeben. Dann gilt auch $\bigcap_{n\in\mathbb{N}} C_n \in \mathcal{B}(\mathbb{R}^d)$.*

Beweis. Wir setzen $C := \bigcap_{n\in\mathbb{N}} C_n$. Es gilt

$$\mathbb{R}^d \setminus C = \mathbb{R}^d \setminus \bigcap_{n\in\mathbb{N}} C_n = \bigcup_{n\in\mathbb{N}} \underbrace{(\mathbb{R}^d \setminus C_n)}_{\in\mathcal{B}(\mathbb{R}^d)} \in \mathcal{B}(\mathbb{R}^d)$$

und somit $C = \mathbb{R}^d \setminus (\mathbb{R}^d \setminus C) \in \mathcal{B}(\mathbb{R}^d)$. □

Beachten Sie: Aus Axiom (II) in Definition 3.2.1 und aus Proposition 3.2.2 folgt, dass auch die Vereinigung und der Durchschnitt endlich vieler messbarer Mengen wieder messbar ist.[2]

Theorem 3.2.3 (Existenz und Eindeutigkeit des Lebesgue-Maßes)**.** *Sei $d \in \mathbb{N}$ und sei $\mathbb{R}^d$ mit einer beliebigen Norm ausgestattet. Es gibt genau eine Abbildung $\lambda_d : \mathcal{B}(\mathbb{R}^d) \to [0, \infty]$, die die folgenden Eigenschaften erfüllt:*

(I) *Für jede Folge $C_1, C_2, C_3, \cdots \subseteq \mathbb{R}^d$ von paarweise disjunkten[3] messbaren Mengen gilt*

$$\lambda_d\Big(\bigcup_{n\in\mathbb{N}} C_n\Big) = \sum_{n=1}^{\infty} \lambda_d(C_n)$$

(II) *Für alle Zahlen $a_1, \dots a_d \in \mathbb{R}$ und $b_1, \dots, b_d \in \mathbb{R}$, die $a_k \le b_k$ für alle $k \in \{1, \dots, d\}$ erfüllen, gilt*

$$\lambda_d([a_1, b_1] \times \cdots \times [a_d, b_d]) = (b_1 - a_1) \cdot \cdots \cdot (b_d - a_d).$$

*Die Abbildung λ_d nennt man das **Lebesgue-Maß auf $\mathbb{R}^d$** oder das **d-dimensionale Lebesgue-Maß**.*

[1] Der Begriff **σ-Algebra** mag auf den ersten Blick etwas merkwürdig und unmotiviert anmuten. Weshalb dieser Begriff verwendet wird, wird in der Analysis 3 klarer, wo wir über allgemeinere σ-Algebren sprechen werden.

[2] Warum?

[3] Mit **paarweise disjunkt** ist gemeint, dass $C_j \cap C_k = \emptyset$ für alle $j, k \in \mathbb{N}$ mit $j \neq k$ gilt.

Beweis. Das werden wir in der Analysis 3 beweisen. □

Theorem 3.2.4 (Einige Eigenschaften des Lebesgue-Maßes)**.** *Sei $d \in \mathbb{N}$.*

(a) *Wenn $C, D \subseteq \mathbb{R}^d$ messbar sind und $C \subseteq D$ gilt, dann ist $\lambda_d(C) \leq \lambda_d(D)$.*

(b) *Für zwei messbare Mengen $C, D \subseteq \mathbb{R}^d$ gilt*

$$\lambda_d(C) + \lambda_d(D) = \lambda_d(C \cup D) + \lambda_d(C \cap D).$$

(c) *Für jedes $x \in \mathbb{R}^d$ ist die Ein-Punkt-Menge $\{x\}$ messbar und es gilt $\lambda_d(\{x\}) = 0$.*

(d) *Wenn $C \subseteq \mathbb{R}^d$ endlich oder abzählbar unendlich ist, dann ist C messbar und es gilt $\lambda_d(C) = 0$.*

(e) *Es gilt $\lambda_d(\emptyset) = 0$ und $\lambda_d(\mathbb{R}^d) = \infty$.*

(f) *Für jede Folge $C_1, C_2, C_3, \cdots \subseteq \mathbb{R}^d$ von messbaren Mengen gilt*

$$\lambda_d\Big(\bigcup_{n\in\mathbb{N}} C_n\Big) \leq \sum_{n=1}^{\infty} \lambda_d(C_n)$$

(g) *Das Lebesgue-Maß ist translations- und rotationsinvariant. Genauer: Sei $C \subseteq \mathbb{R}^d$ messbar, sei $v \in \mathbb{R}^d$ und sei $W : \mathbb{R}^d \to \mathbb{R}^d$ eine Drehung oder eine Drehspiegelung. Dann sind auch[4] $C + v$ und $W(C)$ messbare und es gilt*

$$\lambda_d(C + v) = \lambda_d(C) \qquad \text{und} \qquad \lambda_d\big(W(C)\big) = \lambda_d(C).$$

(h) *Wenn $d \geq 2$ ist, dann gilt für jede Gerade G in $\mathbb{R}^d$, dass $\lambda_d(G) = 0$ ist.*

Vorlesung 11: (Fr, 16.05.) beginnt mit dem Beweis von Theorem 3.2.4

Beweis. (a) Sei C und D wie in der Aussage angegeben. Aus Axiom (II) in Definition (II) folgt, dass $D \setminus C$ messbar ist,[5] und da $D \setminus C$ und C disjunkt sind, folgt

$$\lambda_d(C) \leq \lambda_d(C) + \lambda_d(D \setminus C) = \lambda_d\big(C \cup (D \setminus C)\big) = \lambda_d(D),$$

wobei wir am Ende verwendet haben, dass $C \subseteq D$ gilt.

(b) Wir bemerken zunächst, dass $C \cap D$ und $C \cup D$ messbar sind. Wir zerlegen die Mengen C, D und $C \cup D$ in disjunkte Mengen

$$\begin{aligned} C &= (C \cap (C \setminus D)) \,\dot\cup\, (C \cap D) = (C \setminus D) \,\dot\cup\, (C \cap D), \\ D &= (D \setminus C) \,\dot\cup\, (D \cap C), \\ C \cup D &= (C \setminus D) \,\dot\cup\, (D \setminus C) \,\dot\cup\, (D \cap C), \end{aligned}$$

[4] Mit der Notation $C + v$ meinen wir die Menge $C + v := \{c + v \mid c \in C\}$.

[5] Warum?

wobei $\dot\cup$ anzeigt, dass die vereinigten Mengen links und rechts des Symbols disjunkt sind. Somit folgt mit Theorem 3.2.3 (I), dass

$$\begin{aligned}\lambda_d(C \cup D) &= \lambda_d(C \setminus D) + \lambda_d(D \setminus C) + \lambda_d(D \cap C)\\ &= \big(\lambda_d(C \setminus D) + \lambda_d(D \cap C)\big) + \big(\lambda_d(D \setminus C) + \lambda_d(D \cap C)\big) - \lambda_d(D \cap C)\\ &= \lambda_d\big((C \setminus D) \dot\cup (D \cap C)\big) + \lambda_d\big((D \setminus C) \dot\cup (D \cap C)\big) - \lambda_d(D \cap C)\\ &= \lambda_d(C) + \lambda_d(D) - \lambda_d(C \cap D).\end{aligned}$$

(c) Wir bemerken zunächst, dass G abgeschlossen und somit messbar ist. Die Menge $\{x\}$ lässt sich außerdem als Quader der Form

$$\{x\} = [x_1, x_1] \times \cdots \times [x_n, x_n]$$

schreiben, wobei $x = (x_1, \ldots, x_2)$. Damit folgt $\lambda_d(\{x\}) = 0$ aus Theorem 3.2.3 (II).

(d) Da Ein-Punkt-Mengen laut (c) messbar sind, sind wegen Axiom (II) in Definition 3.2.1 auch endlich und abzählbar unendlich Mengen messbar.[6] Außerdem folgt aus Eigenschaft (I) in Theorem 3.2.3, dass diese Mengen Lebesgue-Maß 0 haben, da Ein-Punkt-Mengen laut Aussage (c) Lebesgue-Maß 0 haben.

(e) Wähle einen beliebigen Punkt $x \in \mathbb{R}^d$. Dann gilt $\emptyset \subseteq \{x\}$ und somit wegen (a) und (c), dass $\lambda_d(\emptyset) = 0$ ist. Ausßerdem gilt für jedes $n \in \mathbb{N}$ die Ungleichung

$$\lambda_d\big(\mathbb{R}^d\big) \geq \lambda_d\big([-n, n]^d\big) = (2n)^d \geq n$$

und somit $\lambda_d\big(\mathbb{R}^d\big) = \infty$.

(f) Lassen Sie uns die Mengen C_1, C_2, C_3 so abändern, dass sie paarweise disjunkt werden, aber die gleiche Vereinigung haben. Das heißt genauer, wir definieren neue Mengen

$$\begin{aligned}D_1 &:= C_1,\\ D_2 &:= C_2 \setminus C_1,\\ D_3 &:= C_3 \setminus (C_1 \cup C_2),\\ D_4 &:= C_4 \setminus (C_1 \cup C_2 \cup C_3),\\ &\vdots\end{aligned}$$

Dann sind die Mengen $D_1, D_2, \ldots$ paarweise disjunkt, es gilt $D_n \subseteq C_n$ für alle $n \in \mathbb{N}$ und $\bigcup_{n\in\mathbb{N}} D_n = \bigcup_{n\in\mathbb{N}} C_n$. Somit folgt

$$\lambda_d\Big(\bigcup_{n\in\mathbb{N}} C_n\Big) = \lambda_d\Big(\bigcup_{n\in\mathbb{N}} D_n\Big) = \sum_{n=1}^{\infty} \lambda_d(D_n) \leq \sum_{n=1}^{\infty} \lambda_d(C_n).$$

(g) Diese beiden Eigenschaften zu beweisen ist etwas aufwändiger. Wir werden dies in der Analysis 3 tun.

[6] Können Sie dieses Argument im Detail ausführen?

(h) Sei $d \geq 2$ und sei $G \subseteq \mathbb{R}^d$ eine Gerade. Wir bemerken zunächst, dass G abgeschlossen und somit messbar ist. Nun sehen wir uns den Spezialfall an, in dem G die x_1-Achse ist. Lassne Sie uns die Mengen

$$G_n = [-n, n] \times [0, 0]^{d-1},$$

betrachten. Jede dieser Mengen ist abgeschlossen und erfüllt $\lambda_d(G_n) = 0$.[7] Es gilt $\bigcup_{n\in\mathbb{N}} G_n = G$ und Aufgrund von (f) auch

$$\lambda_d(G) \leq \sum_{n\in\mathbb{N}} \lambda_d(G_n) = 0.$$

Jede andere Gerade in $\mathbb{R}^d$ können wir durch eine Verschiebung und eine Drehung auf die x_1-Achse abbilden, also folgt aus Aussage (g), dass alle Geraden in $\mathbb{R}^d$ das Lebesgue-Maß 0 besitzen. □

Mit den Eigenschaften des Lebesgue-Maßes aus den Theoremen 3.2.3 und 3.2.4 kann man bereits das Lebesgue-Maß einiger einfacher Mengen im $\mathbb{R}^d$ ausrechnen. Die besprechen wir in den Übungen. Um Formeln für das Lebesgue-Maß komplizierter Mengen herzuleiten, brauchen wir aber noch mehr theoretische Vorbereitung. Insbesondere wir das Lebesgue-Integral, das wir im folgenden Abschnitt besprechen, hierfür hilfreich sein.

3.3 Das Lebesgue-Integral

In der Analysis 1 haben wir Funktionen, die von kompakten Intervallen $[a, b]$ nach $\mathbb{R}$ oder $\mathbb{C}$ abbilden, integriert. Das war jedoch nicht für jede Funktion möglich, sondern nur für sogenannte *Riemann-integrierbare* Funktionen. Zur Definition des Riemann-Integrals hatten wir das Intervall $[a, b]$ und kleine Teilintervall aufgeteilt und dann die Fläche unter dem Graphen von f durch Riemann-Summen approximiert. Dieses Vorgehen war mit recht weniger theoretischer Vorbereitung möglich.

Nun wollen wir Funktionen integrieren, die auf Teilmengen des $\mathbb{R}^d$ definiert sind. Auch für diesen Fall kann man ein Riemann-Integral definieren, nämlich folgendermaßen: Wenn zum Beispiel (für $d = 2$) eine Funktion f von einem Rechteck $R \subseteq \mathbb{R}^2$ nach $\mathbb{R}$ abbildet, dann können wir R in viele kleine Rechtecke unterteilen und dann wieder Riemann-Summen konstruieren, um die Volumen unter dem Funktionsgraphen[8] zu approximieren. Dieses Vorgehen funktioniert allerdings nicht mehr ohne weiteres, wenn f beispielsweise nicht auf einem Rechteckt, sondern auf einer Kreisfläche (oder auf einer noch viel komplizierteren Fläche) definiert ist, weil dann nicht mehr klar ist, wenn man den Definitionsbereich von f unterteilen kann. Eine geschickte Möglichkeit dieses Problem zu lösen besteht darin, dass man – anders als

[7] Warum?

[8] Warum nun auf einmal "Volumen unter dem Funktionsgraphen" anstelle von "Fläche unter dem Funktionsgraphen"?

beim Riemann-Integral – nicht den Definitionsbereich der zu integrierenden Funktion partitioniert, sondern lieber den Wertebereich. Wenn man das tut, gelangt man zum Begriff des **Lebesgue-Integrals**, dass wir in Definition 3.3.7 einführen.[9,10]

Ähnlich wie wir das Riemann-Integral in der Analysis 1 nicht für alle Funktionen definieren konnten, können wir auch das Lebesgue-Integral nicht für alle Funktionen definieren. Der passende Begriffe um das Lebesgue-Integral zu definieren sind die **messbaren Funktionen**, die wir nun einführen:

Definition 3.3.1 (Messbare Funktionen)**.** Sei $\Omega \subseteq \mathbb{R}^d$ messbar. Eine Funktion $f : \Omega \to \mathbb{R}$ heißt **messbar**, falls für jedes $b \in \mathbb{R}$ das Urbild[11]

$$f^{-1}\big((-\infty, b]\big) = \{x \in \Omega \mid \ f(x) \leq b\} \subseteq \Omega \subseteq \mathbb{R}^d$$

messbar ist.

Man kann leicht eine nicht-messbare Funktion abgeben: Sei zum Beispiel $C \subseteq \mathbb{R}^3$ eine nicht-messbare Menge; ein solche gibt es aufgrund des Banach-Tarski–Paradoxons (Theorem 3.1.2). Für die Indikatorfunktion $\mathbb{1}_C : \mathbb{R}^3 \to \mathbb{R}$,

$$\mathbb{1}_C(x) := \begin{cases} 1 & \text{falls } x \in C, \\ 0 & \text{falls } x \in \mathbb{R}^3 \setminus C \end{cases}$$

gilt dann $\mathbb{1}_C^{-1}\left((-\infty, \frac{1}{2}]\right) = \mathbb{R}^3 \setminus C$. Diese Menge ist nicht messbar, da C nicht messbar ist. Also ist die Funktion $\mathbb{1}_C$ nicht messbar.

Proposition 3.3.2 (Messbare Urbilder von Intervallen)**.** *Sei $\Omega \subseteq \mathbb{R}^d$ messbar und sei $f : \Omega \to \mathbb{R}$ eine messbare Funktion. Dann ist für jedes Intervall $I \subseteq \mathbb{R}$ das Urbild $f^{-1}(I) \subseteq \Omega \subseteq \mathbb{R}^d$ messbar.*

Beweis. Seien $a, b \in \mathbb{R}$. Wir betrachten alle verschiedenen Typen von Intervallen, die es in $\mathbb{R}$ gibt:

Erster Fall: $I = (-\infty, b]$. In diesem Fall $f^{-1}(I)$ laut Definition 3.3.1 messbar.

Zweiter Fall: $I = (-\infty, b)$. Wir verwenden, dass $(-\infty, b) = \bigcup_{n\in\mathbb{N}}(-\infty, b - \frac{1}{n}]$ gilt, und somit ist

$$f^{-1}(I) = f^{-1}\Big(\bigcup_{n\in\mathbb{N}}(-\infty, b - \frac{1}{n}]\Big) = \bigcup_{n\in\mathbb{N}} f^{-1}\big((-\infty, b - \frac{1}{n}]\big),$$

[9] Das Lebesgue-Integrale hat noch eine Reihe weiter Vorteile gegenüber dem Riemann-Integral. Beispiel hat es theoretisch weit bessere Eigenschaften (zum Beispiel in der **Funktionalanalysis** eine wichtige Rollen spielen). Außerdem ist das Lebesgue-Integral sehr flexibel: Man kann es für andere sogenannte **Maße** verallgemeinern, die ähnliche – aber nicht diesselben – Eigenschaften wie des Lebesgue-Maß haben. Das ist beispielsweise in der Wahrscheinlichkeitstheorie enorm wichtig.

[10] Es gibt übrigens einen Trick, wie man doch mit Hilfe des Riemann-Integrals Funktionen integrieren kann, die auf nicht rechteckigen Mengen definiert sind. Man könnte aber nicht ganz unbegründet zur Ansicht gelangen, dass dieser Trick recht konstruiert und unintuitiv ist. Deshalb wollen wir uns damit nicht aufhalten und definieren lieber direkt das Lebesgue-Integral.

[11] Diese Menge bezeichnet man auch als **Subniveau-Menge** von f zum Niveau b. Weshalb?

und die rechtsstehende Menge ist messbar als eine abzählbare Vereinigung von messbare Mengen.

Dritter Fall: $I = [a, \infty)$. Dann ist

$$\Omega \setminus f^{-1}(I) = f^{-1}\big(\mathbb{R} \setminus I\big) = f^{-1}\big((-\infty, a)\big),$$

die die rechtsstehende Menge ist, wie im zweiten Fall gezeigt, messbar. Also ist auch ihr Komplement

$$f^{-1}(I) = \Omega \setminus \big(\Omega \setminus f^{-1}(I)\big)$$

messbar.

Vierter Fall: $I = (a, \infty)$. Diesen Fall kann man analog zum dritten Fall (oder analog zum zweiten Fall) behandeln.

Fünfter Fall: $I = [a, b]$. In diesem Fall schreiben wir das Urbild von I als

$$f^{-1}(I) = f^{-1}\big((-\infty, b] \cap [a, \infty)\big) = f^{-1}\big((-\infty, b]\big) \cap \big(f^{-1}([a, \infty))\big).$$

Also ist $f^{-1}(I)$ der Durchschnitt zweier Mengen, die laut des ersten und des dritten Falls messbar sind. Da der Durchschnitt von zwei messbaren Mengen wieder messbar ist,[12] ist auch $f^{-1}(I)$ messbar.

Sechster bis achter Fall: $I = (a, b]$ oder $I = [a, b)$ oder $I = (a, b)$. Diese Fälle kann man analog zum fünften Fall behandeln.

Neunter Fall: $I = \mathbb{R}$. In diesem Fall ist $f^{-1}(I)$ gleich Ω und somit messbar. □

Mit im Wesentlichen denselben Argumenten wie im Beweis von Proposition 3.3.2 kann man folgende Charakterisierung der Messbarkeit von Funktionen zeigen:

Proposition 3.3.3 (Charakterisierung messbarer Funktionen)**.** *Sei $d \in \mathbb{N}$ und $\Omega \subseteq \mathbb{R}^d$. Eine Funktion $f : \Omega \to \mathbb{R}$ is genau dann messbar, wenn für jedes $b \in \mathbb{R}$ das Urbild $f^{-1}\big((-\infty, b)\big)$ messbar ist.*

Mit Hilfe der folgenden Proposition kann man für zahlreiche Funktionen leicht nachprüfen, dass sie messbar sind.

Proposition 3.3.4 (Konstruktionen mit messbaren Funktionen)**.** *Sei $d \in \mathbb{N}$ und sei $\Omega \subseteq \mathbb{R}^d$ messbar.*

(a) *Sei $\mathbb{R}^d$ mit einer Norm und Ω mit der hiervon induzierten Metrik ausgestattet. Dann ist jede stetige Funktion $f : \Omega \to \mathbb{R}$ messbar.*

[12] Warum?

(b) *Für jedes messbare Menge $C \subseteq \Omega$ ist die* ***Indikatorfunktion*** $\mathbb{1}_C$*, die durch*

$$\mathbb{1}_C(x) := \begin{cases} 1, & \text{falls } x \in C, \\ 0, & \text{falls } x \in \Omega \setminus C \end{cases}$$

gegeben ist, messbar.[13]

(c) *Seien $f, g : \Omega \to \mathbb{R}$ messbar und $\alpha \in \mathbb{R}$. Dann sind auch $f + g$, fg und αf messbar.*

(d) *Sei $(f_n)_{n\in\mathbb{N}}$ eine Folge von messbaren Funktionen $f_n : \Omega \to \mathbb{R}$, die punktweise gegen eine Funktion $f : \Omega \to \mathbb{R}$ konvergiert. Dann ist auch f messbar.*

Beweis. Wir beweisen die Aussage (a), das sie sehr gut die Verwendung von Konzepten aus Kapitel 1 illustriert. Für den Beweis der anderen Teilaussagen verweisen wir auf die Analysis 3.

Sei $b \in \mathbb{R}$. Laut Proposition 3.3.3 brauchen wir nur zu zeigen, dass die Menge $C := f^{-1}\big((-\infty, b)\big) \subseteq \Omega \subseteq \mathbb{R}^d$ messbar ist. Da f stetig ist, ist C laut Theorem 1.2.16 eine offene Teilmenge von Ω. Hieraus folgt, dass es eine offene Teilmenge U von $\mathbb{R}^d$ derart, dass $C = U \cap \Omega$ ist.[14] Somit ist C als Durchschnitt von zwei messbaren Mengen messbar. □

Vorlesung 12: (Mi, 21.05.) beginnt mit Proposition 3.3.5

Proposition 3.3.5 (Hintereinanderausführung stetiger und messbarer Funktionen)**.** *Sei $d \in \mathbb{N}$ und sei $\Omega \subseteq \mathbb{R}^d$ messbar. Sei $\mathbb{R}$ mit der Euklidischen Metrik ausgestattet, sei $f : \Omega \to \mathbb{R}$ messbar und sei $g : \mathbb{R} \to \mathbb{R}$ stetig. Dann ist $g \circ f : \Omega \to \mathbb{R}$ messbar.*

Beweis. Sei $b \in \mathbb{R}$. Wir müssen laut Proposition 3.3.3 nur zeigen, dass die Menge $C := (g \circ f)^{-1}\big((-\infty, b)\big) = f^{-1}\Big(g^{-1}\big(-\infty, b)\big)\Big)$ messbar ist. Da g stetig ist, ist $D := (g^{-1}\big(-\infty, b)\big) \subseteq \mathbb{R}$ laut Theorem 1.2.16 offen. Laut Aufgabe 7.3 auf Hausaufgabenblatt 7 ist wegen der Messbarkeit von f das Urbild jeder offenen Menge unter f messbar, also ist $f^{-1}(D)$ messbar. □

[13] Beachten Sie, dass Indikatorfunktionen oft nicht stetig sind, das bedeutet, es gibt zahlreiche Funktionen, die unstetig aber messbar sind.

[14] Der Beweis, dass solch ein U existiert, war ursprünglich als eine Übungsaufgabe geplant, aber leider habe ich dann vergessen mich darum zu kümmern, dass die Aufgabe wirklich auf einem Übungsblatt erscheint. Deshalb nun hier der Beweis:

Die Menge U kann man folgendermaßen bauen: Für jedes $x \in C$ gibt es wegen der Offenheit von C in Ω einen Radius $\varepsilon_x > 0$ derart, dass die offene Kugel $\{y \in \Omega \mid \|y - x\|_2 < \varepsilon_x\}$ in Ω komplett in C enthalten ist; diese offene Kugel in Ω ist aber gleich $B_{<\varepsilon_x}(x) \cap \Omega$, wobei $B_{<\varepsilon_x}(x)$ die offene Kugel in $\mathbb{R}^d$ bezeichnet (ebenfalls bezüglich der 2-Norm).

Nun setzt man einfach $U := \bigcup_{x\in C} B_{<\varepsilon_x}(x)$. Diese Menge ist offen in $\mathbb{R}^d$, weil sie eine Vereinigung von offenen Mengen ist. Außerdem gilt offensichtlich $C \subseteq U$ und somit $C \subseteq U \cap \Omega$. Anderseits ist

$$U \cap \Omega = \Big(\bigcup_{x\in C} B_{<\varepsilon_x}(x)\Big) \cap \Omega = \bigcup_{x\in C} \big(B_{<\varepsilon_x}(x) \cap \Omega\big) = \bigcup_{x\in C} \{y \in \Omega \mid \|y - x\|_2 < \varepsilon_x\} \subseteq C.$$

Also gilt $C = U \cap \Omega$.

Definition 3.3.6 (Positiv- und Negativteil von Funktionen)**.** Sei $d \in \mathbb{N}$, sei $\Omega \subseteq \mathbb{R}^d$ und sei $f : \Omega \to [0, \infty)$. Dann definieren wir zwei Funktionen $f^+, f^- : \Omega \to \mathbb{R}$ durch

$$f^+(x) := \max\{f(x), 0\},$$
$$f^-(x) := \max\{-f(x), 0\}$$

für alle $x \in \Omega$.

Man beachte in der vorangehenden Definition, dass $f^- = (-f)^+$ gilt und dass $f^+ - f^- = f$ ist. Wenn f messbar ist, dann sind wegen Proposition 3.3.5 auch f^+ und f^- messbar.[15]

Definition 3.3.7 (Das Lebesgue-Integral)**.** Sei $d \in \mathbb{N}$, sei $\emptyset \neq \Omega \subseteq \mathbb{R}^d$ messbar und sei $f : \Omega \to \mathbb{R}$ messbar.

(a) Sei $f(\Omega) \subseteq [0, \infty)$. Dann nennt man

$$\int_\Omega f \, \mathrm{d}\lambda_d := \sup \Big\{ \sum_{k=0}^{n-1} y_k \lambda_d\Big(f^{-1}\big([y_k, y_{k+1})\big)\Big) \Big| n \in \mathbb{N},\ 0 = y_0 \leq \cdots \leq y_n \Big\}$$

das **Lebesgue-Integral** oder kurz das **Integral** von f;[16] man beachte, dass $\int_\Omega f \, \mathrm{d}\lambda_d \in [0, \infty]$ gilt.

Zur Berechnung der Summe innerhalb der Menge verwenden wir die Konvention $0 \cdot \infty := 0$.

(b) Sei $f(\Omega) \subseteq [0, \infty)$. Man nennt f **integrierbar**, falls $\int_\Omega f \, \mathrm{d}\lambda_d < \infty$ gilt.

(c) Man nennt f **integrierbar**, falls f^+ und f^- integrierbar sind. In diesem Fall nennt man

$$\int_\Omega f \, \mathrm{d}\lambda_d := \int_\Omega f^+ \, \mathrm{d}\lambda_d - \int_\Omega f^- \, \mathrm{d}\lambda_d$$

das **Lebesgue-Integral** oder kurz das **Integral** von f.

Manchmal schreibt man anstelle von $\int_\Omega f \, \mathrm{d}\lambda_d$ auch $\int_\Omega f(x) \, \mathrm{d}\lambda_d(x)$.

Für das Lebesgue-Integral gelten zahlreiche Aussagen, die man anschaulich genauso erwarten würde. Wir führen diese in den folgenden beiden Theoremen auf und verzichten an dieser Stelle auf die Beweise.[17]

Theorem 3.3.8 (Eigenschaften des Lebesgue-Integrals)**.** *Sei $d \in \mathbb{N}$, sei $\emptyset \neq \Omega \subseteq \mathbb{R}^d$ messbar und seien $f, g : \Omega \to \mathbb{R}$ messbar.*

[15] Können Sie dieses Argument im Detail ausführen?

[16] Wo in dieser Definition haben wir verwendet, dass f messbar ist?

[17] Zum einen kennen Sie viele ähnliche Aussagen nämlich bereits für das Riemann-Integral. Und zum anderen werden wir auf diese Resultate in der Analysis 3 in deutlich größerer Allgemeinheit eingehen. Für den Moment geht es uns nur darum einen Inegralbegriff zur Verfügung zu haben, mit dem wir im Rest der Analysis 2-Vorlesung möglichst effizient und flexibel arbeiten können.

(a) *Für jedes messbare Menge* $C \subseteq \Omega \subseteq \mathbb{R}^d$ *gilt* $\int_\Omega \mathbb{1}_C \, \mathrm{d}\lambda_d = \lambda_d(C)$.

(b) *Die Funktion* f *ist genau dann integrierbar, wenn* $|f| : \Omega \to [0, \infty)$, $x \mapsto |f(x)|$ *integrierbar ist.*[18] *In diesem Fall gilt dies sogenannte Fundamentalabschätzung*

$$\left| \int_\Omega f \, \mathrm{d}\lambda_d \right| \leq \int_\Omega |f| \, \mathrm{d}\lambda_d.$$

(c) *Wenn* f, g *nur Werte in* $[0, \infty)$ *annehmen and* $\alpha, \beta \in [0, \infty)$ *gilt, dann ist*

$$\int_\Omega \alpha f + \beta g \, \mathrm{d}\lambda_d = \alpha \int_\Omega f \, \mathrm{d}\lambda_d + \beta \int_\Omega g \, \mathrm{d}\lambda_d.$$

Falls f *and* g *beliebige Werte in* $\mathbb{R}$ *annehmen und integrierbar sind und falls* $\alpha, \beta \in \mathbb{R}$ *gilt, dann ist auch* $\alpha f + \beta g$ *integrierbar und es gilt diesselbe Formel.*

(d) *Sei* $C \subseteq \Omega \subseteq \mathbb{R}^d$ *messbar. Falls* f *nur Werte in* $[0, \infty)$ *annimmt, dann ist*

$$\int_C f \, \mathrm{d}\lambda_d := \int_C f|_C \, \mathrm{d}\lambda_d = \int_\Omega f \, \mathbb{1}_C \, \mathrm{d}\lambda_d.$$

Falls f *beliebige Werte in* $\mathbb{R}$ *annimmt und integrierbar ist, dann sind auch* $f \, \mathbb{1}_C$ *and* $f|_C$ *integrierbar und es gilt dieselbe Formel.*

(e) *Sei* $f \leq g$. *Falls* f *und* g *nur Werte in* $[0, \infty)$ *annehmen oder falls beiden Funktionen integrierbar sind, gilt*

$$\int_\Omega f \, \mathrm{d}\lambda_d \leq \int_\Omega g \, \mathrm{d}\lambda_d.$$

(f) *Seien* $a, b \in \mathbb{R}$ *mit* $a < b$ *und sei* $h : [a, b] \to \mathbb{R}$ *stetig. Dann ist* h *integrierbar und das Lebesgue-Integral von* h *stimmt mit dem Riemann-Integral von* h *überein.*[19]

Natürlich möchten wir nicht nur über theoretische Eigenschaften von Integralen sprechen, sondern wir möchten sie in konkreten Beispielen auch berechnen können. Für das Riemann-Integral konnten wir das häufig mit Hilfe des Hauptsatzes der Differential- und Inegralrechnung tun. Integrale von Funktionen, die auf Teilmengen von $\mathbb{R}^d$ definiert sind, kann man mit Hilfe des folgenden äußerst nützlichen Satzes oft auf Integrale über Funktionen mit nur einer Veränderlichen zurückführen, um diese dann mit Hilfe von Theorem 3.3.8(f) zu berechnen.

[18] Warum ist $|f|$ überhaupt messbar?

[19] Das stimmt übrigens nicht nur für stetiges h, sondern allgemeiner, wenn h beschränkt und Riemann-integrierbar ist. Allerdings mussen man hier auf ein paar technische Subtilitäten achten und den Begriff der Messbarkeit geringfügig erweitern. Wir werden am Rande der Analysis 3 hierauf zurückkommen.

Für die meisten Anwendungszwecke reicht es aber völlig aus, dieses Resultat für stetige Funktionen h zu kennen. Das Resultat ist nämlich in erster Linie deshalb wichtig, weil wir zur Berechnung von Riemann-Integralen eine konkrete Methode zur Verfügung haben, nämlich den Hauptsatz der Differential- und Integralrechnung; in diesem hatten wir aber ohnehin vorausgesetzt, dass der Integrand stetig ist.

Theorem 3.3.9 (Berechnung durch Mehrfachintegration)**.** *Seien $d_1, d_2 \in \mathbb{N}$ und $d := d_1 + d_2$. Außerdem seien $\Omega_1 \subseteq \mathbb{R}^{d_1}$ und $\Omega_2 \subseteq \mathbb{R}^{d_2}$ messbar. Dann ist auch $\Omega_1 \times \Omega_2 \subseteq \mathbb{R}^d$ messbar.*[20] *Sei nun $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ messbar.*

(a) *Für jedes festes $x_1 \in \Omega_1$ and jedes festes $x_2 \in \Omega_2$ sind die Funktionen*

$$f(\,\cdot\,, x_2) : \Omega_1 \to \mathbb{R} \qquad \text{und} \qquad f(x_1, \,\cdot\,) : \Omega_2 \to \mathbb{R}$$

messbar.

Vorlesung 13: (Do, 22.05.) beginnt mit Theorem 3.3.9(b)

(b) *Falls für jedes $x_2 \in \Omega_2$ die Funktion $f(\,\cdot\,, x_2) : \Omega_1 \to \mathbb{R}$ integrierbar ist, dann ist die Abbildung $\Omega_2 \to \mathbb{R}$, $x_2 \mapsto \int_{\Omega_1} f(x_1, x_2)\, \mathrm{d}\lambda_{d_1}(x_1)$ messbar.*

(c) *Falls für jedes $x_2 \in \Omega_2$ die Funktion $f(\,\cdot\,, x_2) : \Omega_1 \to \mathbb{R}$ integrierbar ist und falls f nur Werte in $[0, \infty)$ annimmt, dann gilt*

$$\int_{\Omega_1 \times \Omega_2} f(x)\, \mathrm{d}\lambda_d(x) = \int_{\Omega_2} \int_{\Omega_1} f(x_1, x_2)\, \mathrm{d}\lambda_{d_1}(x_1)\, \mathrm{d}\lambda_{d_2}(x_2). \tag{3.3.1}$$

(d) *Falls für jedes $x_2 \in \Omega_2$ die Funktion $f(\,\cdot\,, x_2) : \Omega_1 \to \mathbb{R}$ integrierbar ist und falls f integrierbar ist, dann gilt ebenfalls die Formel* (3.3.1)*.*

(e) *Alle vorangehenden Aussagen gelten ebenfalls, wenn man die Rollen von Ω_1 und Ω_2 (und entsprechend von x_1 und x_2 sowie von d_1 und d_2) vertauscht.*

Theorem 3.3.9 ist unter den Namen "Satz von Fubini" oder "Satz von Fubini–Tonelli" bekannt.

Lassen Sie uns die Verwendung der Theoreme 3.3.8 und 3.3.9 anhand eines einfachen Beispiels illustrieren.

Beispiel 3.3.10 (Verwendung eines Mehrfach-Integrals)**.** Es sei $f : [0,1]^2 \to \mathbb{R}$, $x \mapsto (x_1 - x_2)^3$. Wir zeigen, dass f integrierbar ist und berechnen das Integral $\int_{[0,1]^2} |f(x)|\, \mathrm{d}\lambda_2(x)$.

Lösung. Die Funktion f ist stetig[21,22] und somit laut Proposition 3.3.4(a) messbar. Für alle $x \in [0,1]^2$ gilt außerdem $|f(x)| \leq 1^3 = 1$ und somit laut Theorem 3.3.8(e), (c) und (a)

$$\int_{[0,1]^2} |f(x)|\, \mathrm{d}\lambda_2(x) \leq \int_{[0,1]^2} \mathbb{1}_{[0,1]^2}\, \mathrm{d}\lambda_2(x) = \lambda_2([0,1]^2) = 1 < \infty.$$

Also ist f laut Theorem 3.3.8(b) integrierbar.

[20] Das ist übrigens keineswegs offensichtlich.

[21] Wenn wir $\mathbb{R}^2$ mit einer Norm und $[0,1]^2$ mit der hiervon induzierten Metrik ausstatten.

[22] Können Sie im Detail beweisen, weshalb f stetig ist?

Nun berechnen wir $\int_{[0,1]^2} |f(x)| \, \mathrm{d}\lambda_2(x)$. Hierzu halten wir zunächst die Variable $x_2 \in [0,1]$ fest. Die Funktion $f(\,\cdot\,, x_2) : [0,1] \to \mathbb{R}$ ist stetig. Laut Theorem 3.3.8(f) ist die Funktion somit integrierbar und es gilt

$$\int_{[0,1]} f(x_1, x_2) \, \mathrm{d}\lambda_1(x_1) = \int_0^1 (x_1 - x_2)^3 \, \mathrm{d}x_1 = \frac{1}{4}\big((x_2 - 1)^4 - x_2^4\big) =: g(x_2).$$

Wegen Theorem 3.3.9(d) gilt

$$\begin{aligned}\int_{[0,1]^2} f(x) \, \mathrm{d}\lambda_2(x) &= \int_{[0,1]} g(x_2) \, \mathrm{d}\lambda_1(x_2) \\ &= \frac{1}{4} \int_0^1 (x_2 - 1)^4 - x_2^4 \, \mathrm{d}x_2 = \frac{1}{20}\big(- (-1)^5 - 1^5\big) = 0,\end{aligned}$$

wobei wir für die zweite Gleichheit erneut Theorem 3.3.8(f) verwendet haben. □

3.4 Konvergenzssätze, uneigentliche Integrale und Parameterintegrale

Beispiel 3.4.1 (Grenzwerte und Integrale vertauschen nicht)**.** Auf der messbaren Menge $[0,1] \subseteq \mathbb{R}$ betrachten wir die Funktionenfolge $\big(n \, \mathbb{1}_{(0,\frac{1}{n}]}\big)_{n\in\mathbb{N}}$. Diese Folge konvergiert punktweise gegen die konstante Nullfunktion.[23] Allerdings gilt

$$\int_{[0,1]} n \, \mathbb{1}_{(0,\frac{1}{n}])} \, \mathrm{d}\lambda_1 = n\lambda_1\big((0, \frac{1}{n}]\big) = 1 \overset{n\to\infty}{\to} \neq 0 = \int_{[0,1]} 0 \, \mathrm{d}\lambda_1.$$

Man kann also Integral und Grenzwert nicht immer vertauschen!

Eine sehr schöne Eigenschaft des Lebesgue-Integrals besteht darin, dass es allerdings sehr nützliche Sätze gibt, die hinreichende Bedingungen dafür angeben, dass wir Integral und Grenzwert vertauschen dürfen. Wir beweisen den folgenden Satz an dieser Stelle nicht,[24] sondern fokussieren uns anschließend lieber auf einige Anwendungen des Satzes.

Theorem 3.4.2 (Konvergenzssätze für Lebesgue-Integrale)**.** *Sei $d \in \mathbb{N}$, sei $\Omega \subseteq \mathbb{R}^d$ messbar und seien $f, f_n : \Omega \to \mathbb{R}$ messbar für $n \in \mathbb{N}$. Außerdem gelten $f_n \to f$ punktweise für $n \to \infty$.*

(a) *Falls $0 \leq f_n(x) \leq f(x)$ für alle $x \in \Omega$ ist, dann gilt*

$$\int_\Omega f_n \, \mathrm{d}\lambda_d \overset{n\to\infty}{\to} \int_\Omega f \, \mathrm{d}\lambda_d. \tag{3.4.1}$$

[23] Warum?

[24] Wie Sie sich vermutlich bereits denken können, werden wir ihn in noch größerer Allgemeinheit in der Analysis 3 beweisen.

(b) *Falls es eine integrierbare Funktion* $g : \Omega \to [0, \infty)$ *mit der Eigenschaft* $|f_n(x)| \leq g(x)$ *für alle* $x \in \Omega$ *gibt, dann sind alle* f_n *und die Grenzfunktion* f *integrierbar und es gilt ebenfalls* (3.4.1).

Mit Hilfe des vorangehenden Satzes kann man Integrierbarkeit einer Funktion überprüfen, in dem Sie auf geeigenten Teilmengen Ihres Definitionsbereichs untersucht:

Korollar 3.4.3 (Integrierbarkeit mittels Ausschöpfung des Definitionsbereichs)**.** *Sei* $\Omega \subseteq \mathbb{R}^d$ *messbar und seien* $C_1 \subseteq C_2 \subseteq \ldots$ *messbare Teilmengen von* Ω *mit* $\bigcup_{n\in\mathbb{N}} C_n = \Omega$*. Sei* $f : \Omega \to \mathbb{R}$ *messbar. Falls es ein* $c \in [0, \infty)$ *mit* $\int_{C_n} |f| \, \mathrm{d}\lambda_d \leq c$ *für alle* $n \in \mathbb{N}$ *gibt, dann ist* f *integrierbar und es gilt*

$$\int_{C_n} f \, \mathrm{d}\lambda_d \overset{n\to\infty}{\to} \int_\Omega f \, \mathrm{d}\lambda_d.$$

Beweis. Indem wir Theorem 3.4.2(a) auf die Funktionenfolge $(\mathbb{1}_{C_n} |f|)_{n\in\mathbb{N}}$ anwenden, erhalten wir $\int_\Omega |f| \, \mathrm{d}\lambda_d = \lim_{n\to\infty} \int_{C_n} |f| \, \mathrm{d}\lambda_d \leq c$, also ist f integrierbar.

Wegen $|\mathbb{1}_{C_n} f| \leq |f|$ für alle $n \in \mathbb{N}$ können wir nun Theorem 3.4.2(b) auf die Funktionenfolge $(\mathbb{1}_{C_n} f)_{n\in\mathbb{N}}$ anwenden und erhalten hieraus die Behauptung. □

Riemann-Integrale hatten wir in der Analysis 1 nur für Funktionen definiert, die auf abgeschlossenen beschränkten Intervallen der Form $[a, b]$ definiert sind. Das Lebesgue-Integrale hingegen verlangt nur, dass die Funktion auf einer messbaren Menge definiert ist. Man kann somit also zum Beispiel Funktionen betrachten, die auf ganz $\mathbb{R}$ oder auf einem Intervall der Form $[a, \infty)$ definiert sind. Mit Hilfe von Korollar 3.4.3 kann man das Lebesgue-Integral solcher Funktionen zum einem Grenzwert von Riemann-Integralen in Beziehung setzen:

Korollar 3.4.4 (Das Lebesgue-Integral über $[a, \infty)$ als uneigentliches Riemann-Integral)**.** *Sei* $a \in \mathbb{R}$ *und sei* $f : [a, \infty) \to \mathbb{R}$ *stetig. Falls es ein* $c \in [0, \infty)$ *mit* $\int_a^{a+n} |f(x)| \, \mathrm{d}x \leq c$ *für alle* $n \in \mathbb{N}$ *gibt, dann ist* f *integrierbar und es gilt*

$$\int_{[a,\infty)} f \, \mathrm{d}\lambda_1 = \lim_{n\to\infty} \int_a^{a+n} f(x) \, \mathrm{d}x.$$

Beweis. Hierzu wendet man das Korollar 3.4.3 auf die Mengen $C_n := [a, a + n]$ an. □

Beispiel 3.4.5 (Integral einer schnell fallenden Funktion)**.** Sei $f : [1, \infty) \to \mathbb{R}$, $x \mapsto \frac{1}{x^2}$. Dann ist f integrierbar und es gilt $\int_{[1,\infty)} f \, \mathrm{d}\lambda_1 = 1$.

Beweis. Da f nur Werte in $[0, \infty)$ annimmt, ist $|f(x)| = f(x)$ für alle $x \in [1, \infty)$. Für jedes $n \in \mathbb{N}$ gilt

$$\int_1^{1+n} |f(x)| \, \mathrm{d}x = \int_1^{1+n} f(x) \, \mathrm{d}x = \int_1^{1+n} \frac{1}{x^2} \, \mathrm{d}x = \big[-\frac{1}{x} \big]_{x=1}^{x=1+n} = 1 - \frac{1}{1+n} \leq 1;$$

hier haben wir verwendet, dass wir Riemann-Integrale mit Hilfe des Hauptsatzes der Differential- und Integralrechnung ausrechnen können (sofern wir eine Stammfunktion konkret ausrechnen). Somit ist f laut Korollar 3.4.4 integrierbar und es gilt

$$\int_{[1,\infty)} f \,\mathrm{d}\lambda_1 = \lim_{n\to\infty} \int_1^{1+n} f(x)\,\mathrm{d}x = \lim_{n\to\infty} \big(1 - \frac{1}{1+n}\big) = 1,$$

womit die Behauptung bewiesen ist. □

Nun kommen wir nochmal auf ein Thema aus der Analysis 1 zurück: Mit Hilfe von Integralen über $[1,\infty)$ kann man Kriterien für die Konvergenz von Reihen angeben, denn es gilt der folgende Satz:

Vorlesung 14: (Mi, 28.05.) beginnt mit Theorem 3.4.6

Theorem 3.4.6 (Integralkriterium für Reihenkonvergenz)**.** *Sei $f : [1,\infty) \to [0,\infty)$ stetig,*[25] *monoton fallend und integrierbar. Dann ist die Reihe*

$$\Big(\sum_{k=1}^{n} f(k)\Big)_{n\in\mathbb{N}}$$

konvergent.[26]

Beweis. Da die Summanen $f(k)$ alle größer oder gleich 0 sind, müssen wir lediglich zeigen, dass es eine reelle Zahl c mit der Eigenschaft $\sum_{k=1}^{n} f(k) \le c$ für alle $n \in \mathbb{N}$ gibt; siehe [4,].

Wir wählen $c := f(1) + \int_{[1,\infty)} f \,\mathrm{d}\lambda_1$. Sei nun $n \in \mathbb{N}$. Da f monoton fallend ist, gilt $0 \le \sum_{k=2}^{n} f(k)\, \mathbb{1}_{[k-1,k)}(x) \le f(x)$ für alle $x \in [1,\infty)$. Also ist

$$\begin{aligned}\int_{[1,\infty)} f \,\mathrm{d}\lambda_1 &\ge \int_{[1,\infty)} \sum_{k=2}^{n} f(k)\, \mathbb{1}_{[k-1,k)} \,\mathrm{d}\lambda_1 \\ &= \sum_{k=2}^{n} f(k) \underbrace{\int_{[1,\infty)} \mathbb{1}_{[k-1,k)} \,\mathrm{d}\lambda_1}_{=1} = \sum_{k=2}^{n} f(k)\end{aligned}$$

und somit $\sum_{k=1}^{n} f(k) \le c$. □

Wenn man Theorem 3.4.6 mit Korollar 3.4.4 kombiniert, hat man ein sehr nützliches Werkzeug um die Konvergenz vieler Reihen zu überprüfen. Wir kommen hierauf in den Übungen zurück.

Als letzte Anwendung der Konvergenzsätze von Parameterintegralen erwähnen wir das folgendes Resultat über die stetige Abhängigkeit eines Integrals von einem reellen Parameter, welches man mit Hilfe von Theorem 3.4.2(b) beweise kann.

[25] Genau genommen kann man die Annahme "stetig" hier einfach weglassen. Für den Beweis genügt nämlich Messbarkeit von f und die Messbarkeit folgt wiederum daraus, dass f monoton fallend ist. Darauf wollen wir an dieser Stelle aber nicht im Detail eingehen, deswegen nehmen wir der Einfachkeit halber die Annahme "stetig" hinzu.

[26] Anders ausgedrückt: Es ist $\sum_{k=1}^{\infty} f(k) < \infty$.

Theorem 3.4.7 (Stetigkeit von Parameterintegralen)**.** *Sei $\Omega \subseteq \mathbb{R}^d$ messbar, sei (M, d) ein metrischer Raum und sei $h : \Omega \to [0, \infty)$. Sei $f : M \times \Omega \to \mathbb{R}$ eine Funktion mit folgenden Eigenschaften:*

(1) *Für jedes festes $x \in M$ sei die Funktion $f(x, \cdot) : \Omega \to \mathbb{R}$ messbar.*

(2) *Für jedes feste $\omega \in \Omega$ sei die Funktion $f(\cdot, \omega) : I \to \mathbb{R}$ stetig.*

(3) *Es gelte $|f(x, \omega)| \leq h(\omega)$ für alle $x \in M$ und alle $\omega \in \Omega$.*

(4) *Die Funktion h sei messbar und integrierbar.*

Dann ist die Funktion $M \to \mathbb{R}$, $x \mapsto \int_\Omega f(x, \omega) \, \mathrm{d}\lambda_d(\omega)$ stetig.[27]

Ein analoges Resultat zu Theorem 3.4.7 lässt sich auch über die Differenzierbarkeit von Parameterintegralen zeigen; die Formulierung solch eines Satzes ist aber etwas technischer. wir verzichten an dieser Stelle darauf und verweisen bei Interesse auf die Literatur, zum Beispiel auf [3, Kapiel 11, Satz 2 auf S. 128–129].

[27] Wieso nennt man dieses Integral *Parameterintegral*?

Kapitel 4

Differentialrechnung in mehreren Veränderlichen

4.1 Ableitung von Funktionen in mehreren Veränderlichen

Vorlesung 15: (Fr, 30.05.) beginnt mit Definition 4.1.1

In diesem Abschnitt seien stets $c, d \in \mathbb{N}$ und $\mathbb{R}^c$ und $\mathbb{R}^d$ seien mit der 2-Norm ausgestattet.

Definition 4.1.1 (Totale Differenzierbarkeit)**.** Sei $\Omega \subseteq \mathbb{R}^c$ offen, sei $x \in \Omega$ und sei $f : \Omega \to \mathbb{R}^d$.

(a) Die Funktionen f heißt **total differenzierbar** oder kürzer **differenzierbar** im Punkt x, falls es eine Matrix $A \in \mathbb{R}^{d\times c}$ derart, dass die durch

$$f(y) = f(x) + A(y - x) + e(y) \quad \text{für alle } y \in \Omega$$

definierte Funktion[1] $e : \Omega \to \mathbb{R}^d$ die Bedingung $\frac{\|e(y)\|_2}{\|y-x\|_2} \to 0$ für $y \to x$ erfüllt.[2]

In diesem Fall nennt man A die **Ableitung** oder das **totale Differential** von f im Punkt x und schreibt $A =: (\mathrm{D}f)(x) =: \mathrm{D}f(x)$.

Manchmal wird die Ableitung $\mathrm{D}f(x)$ auch als die **Jacobi-Matrix** von f im Punkt x bezeichnet und mit dem Symbol $\mathrm{J}_f(x)$ bezeichnet.

(b) Im Fall $d = 1$ liegt $\mathrm{D}f(x)$ in $\mathbb{R}^{1\times c}$, ist also ein Zeilenvektor. Seinen transponierten Vektor – also den Spaltenvektor $\big(\mathrm{D}f(x)\big)^{\mathrm{T}} \in \mathbb{R}^c$ – bezeichnet man auch als den **Gradienten** von f im Punkt x und notiert ihn als

$$\big(\mathrm{D}f(x)\big)^{\mathrm{T}} =: \nabla f(x).$$

Das Symbol "∇" spricht man als "Nabla" aus.

[1] Anders ausgedrückt: Es ist $e(y) := f(y) - f(x) - A(y - x)$ für alle $y \in \Omega$.

[2] Wobei in diesem Fall mit $y \to x$ gemeint ist, dass man y aus $\Omega \setminus \{x\}$ nimmt und gegen x laufen lässt.

(c) Die Funktion f heißt **total differenzierbar** oder kürzer **differenzierbar**, falls sie in jedem Punkt $x \in \Omega$ total differenzierbar ist. In diesem Fall heißt die Abbildung $\mathrm{D}f : \Omega \to \mathbb{R}^{d\times c}$, $x \mapsto \mathrm{D}f(x)$ die **Ableitung** von f.

Man nennt die Funktion f **stetig differenzierbar** oder C^1, falls sie total differenzierbar ist und Ihre Ableitung $\mathrm{D}f : \Omega \to \mathbb{R}^{c\times d}$ stetig ist.

Die Eindeutigkeit der Matrix A in Definition 4.1.1(a) folgt aus der Offenheit der Menge Ω, die wir in der Definition angenommen hatten. Das diskutieren wir in einer Übungsaufgabe im Detail.

Beispiele 4.1.2 (Affine Abbildungen sind total differenzierbar)**.** Seien $A \subseteq \mathbb{R}^{d\times c}$ und $b \in \mathbb{R}^d$, und sei $f : \mathbb{R}^c \to \mathbb{R}^d$, $x \mapsto Ax + b$. Dann ist f total differenzierbar und für jedes $x \in \mathbb{R}^c$ gilt $\mathrm{D}f(x) = A$

Beweis. Sei $x \in \mathbb{R}^c$. Dann gilt für alle $y \in \mathbb{R}^c$

$$e(y) := f(y) - f(x) - A(y - x) = 0,$$

Somit gilt $\frac{\|e(y)\|_2}{\|y-x\|_2} = 0 \to 0$ für $y \to x$. Also ist f im Punkt x total differenzierbar mit $\mathrm{D}f(x) = A$. □

Selbst für so harmlose Funktionen wie $f : \mathbb{R}^2 \to \mathbb{R}$, $x \mapsto x_1x_2$ ist es erstaunlich aufwendig, die totale Differenzierbarkeit direkt anhand der Definition nachzuprüfen. Deshalb gehene wir anders vor: Wir beweisen zunächst ein paar theoretische Sätze und verwenden diese dann um die totale Differenzierbarkeit von konkreten Funktionen zu zeigen (und um ihre Ableitung auszurechnen).

Proposition 4.1.3 (Eigenschaften des totalen Differentials)**.** *Seien $c, d, e \in \mathbb{N}$ und sei $\Omega \subseteq \mathbb{R}^c$ offen.*

(a) *Sei $x \in \Omega$ und seien $f, g : \Omega \to \mathbb{R}^d$ im Punkt x differenzierbar. Dann ist für alle $\alpha, \beta \in \mathbb{R}$ auch die Funktion $\alpha f + \beta g : \Omega \to \mathbb{R}^d$ im Punkt x differenzierbar und es gilt*

$$\mathrm{D}(\alpha f + \beta g)(x) = \alpha \mathrm{D}f(x) + \beta \mathrm{D}g(x).$$

(b) *Sei $x \in \Omega$, sei $\tilde{\Omega} \subseteq \mathbb{R}^d$ offen, sei $f : \Omega \to \tilde{\Omega} \subseteq \mathbb{R}^d$ differenzierbar im Punkt x und sei $g : \tilde{\Omega} \to \mathbb{R}^e$ differenzierbar im Punkt $f(x)$. Dann ist auch $g \circ f : \Omega \to \mathbb{R}^e$ differenzierbar im Punkt x und es gilt die sogenannte **Kettenregel**[3]*

$$\mathrm{D}(g \circ f)(x) = \mathrm{D}g(f(x))\,\mathrm{D}f(x).$$

[3]Es ist eine gute Idee sich an dieser Stelle klar zu machen, weshalb die Größen der Matrizen in dieser Formel überhaupt zusammenpassen: Es gilt $\mathrm{D}(g \circ f)(x) \in \mathbb{R}^{e\times c}$ sowie $\mathrm{D}g(f(x)) \in \mathbb{R}^{e\times d}$ und $\mathrm{D}f(x) \in \mathbb{R}^{d\times c}$. Also kann man die beiden letztgenannten Matrizen tatsächlich multiplizieren und ihr Produkt liegt in $\mathbb{R}^{e\times c}$.

(c) *Sei $x \in \Omega$ und sei $f : \Omega \to \mathbb{R}^d$. Wir schreiben f als*

$$f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_d \end{pmatrix}$$

für Funktionen $f_1, f_2, \ldots, f_d : \Omega \to \mathbb{R}$. Es ist f genau dann total differenzierbar im Punkt x, wenn alle Funktionen $f_1, f_2, \ldots, f_d$ im Punkt x total differenzierbar sind, und in diesem Fall gilt[4]

$$\mathrm{D}f(x) = \begin{pmatrix} \mathrm{D}f_1(x) \\ \mathrm{D}f_2(x) \\ \vdots \\ \mathrm{D}f_d(x) \end{pmatrix}.$$

Beweis. (a) und (b) Den Beweis dieser Aussagen kann man analog zur Analysis 1 führen, siehe [4, Theorem 5.1.8].

(c) Diese Aussage kann man mit Hilfe von Definition (a) direkt nachrechnen. Wir verzichten an dieser Stelle auf die Details. □

In Proposition 4.1.3(c) erkennen Sie ein Prinzip, dass Sie in der Analysis 2 schon mehrmals gesehen haben: Funktionen mit Werten in $\mathbb{R}^d$ kann man oft behandeln, indem man sie in ihre Komponentenfunktionen aufteilt, die alle nach $\mathbb{R}$ abbilden. In vielen (aber nicht in allen) Fällen besteht deshalb die wirkliche Quintessenz eines Resultats daran zu verstehen, was im $\mathbb{R}$-wertigen Fall passiert. Das "Mehrdimensionale" an der Analysis 2 ist in erster Linie der mehrdimensioanle Definitionsbereich von Funktion, nicht so sehr der mehrdimensionale Wertebereich.[5]

Übrigens zeigt Proposition 4.1.3(c), dass die Definition der totalen Ableitung (Definition 4.1.1) mit unserer Definition der Ableitung von Kurven (Definition 2.1.3) konsistenz ist. Eine Kurve $\gamma : I \to \mathbb{R}^d$ und somit genau dann differenzierbar in einem Punkt $t \in I$, wenn sie im Punkt t total differenzierbar ist, und in diesem Fall gilt

$$\dot{\gamma}(t) = \mathrm{D}\gamma(t). \tag{4.1.1}$$

All dieser netten Beobachtungen zum Trotz haben wir bishier immer noch keine Möglichkeit, um für interessante Funktionen nachzuprüfen, ob sie total differenzierbar sind und gegebenenfalls ihre Ableitung auszurechnen. Diese Möglichkeit gibt uns die folgende Definition 4.1.4 und das nachfolgende Theorem 4.1.6.

[4] Machen Sie sich auch hier wieder die Größen aller beteiligten Matrizen klar: Die Matrizen $\mathrm{D}f_1(x), \ldots, \mathrm{D}f_d(x)$ sind aus $\mathbb{R}^{1\times c}$, sind also Zeilenvektoren der Länge c. Um die Matrix $\mathrm{D}f(x) \in \mathbb{R}^{d\times c}$ zu erhalten, "stapelt" man alle d dieser Zeilenvektoren untereinander.

[5] Nehmen Sie diese Aussage aber auch nicht zu wörtlich. Ein mehrdimensionaler Wertebereich kann durchauch auch zu interessanten Effekten führen. Zum Beispiel sind Kurven ja nur deshalb in der Lage, die Bahn von bewegten physikalischen Objekten zu beschreiben, weil sie Werte im $\mathbb{R}^d$ annehmen.

Definition 4.1.4 (Richtungsableitungen und partielle Ableitungen)**.** Sei $c \in \mathbb{N}$, sei $\Omega \subseteq \mathbb{R}^c$ offen, sei $x \in \Omega$ und sei $f : \Omega \to \mathbb{R}$.[6]

(a) Sei $r \in \mathbb{R}^c \setminus \{0\}$ und sei $\delta > 0$ so klein, dass $x + tr \in \Omega$ für alle $t \in (-\delta, \delta)$ gilt.[7] Man sagt, dass f im Punkt x **in Richtung r differenzierbar** ist, falls die Funktion

$$\alpha : (-\delta, \delta) \to \mathbb{R}, \quad t \mapsto f(x + tr)$$

im Punkt 0 differenzierbar ist. In diesem Fall nennt man die Zahl

$$\dot{\alpha}(0) = \lim_{t \to 0} \frac{f(x + tr) - f(x)}{t} =: \partial_r f(x) \in \mathbb{R}$$

die **Richtungsableitung** von f im Punkt x in Richtung r.

(b) Sei $k \in \{1, \dots, c\}$ und sei $e_k \in \mathbb{R}^c$ der k-te kanonische Einheitsvektor. Wenn f im Punkt x in Richtung e_k differenzierbar ist, dann nennt man f **partial differenzierbar nach der k-ten Variablen**; in diesem Fall schreibt man

$$\partial_{e_k} f(x) =: \partial_k f(x)$$

und nennt diese Zahl die **k-te partielle Ableitung** von f im Punkt x.

Oft benutzt man für die k-te partielle Ableitung von f im Punkt x auch die Notation $\frac{\partial}{\partial x_k} f(x)$.

(c) Man nennt f **partiell differenzierbar** im Punkt x, wenn f im Punkt x für jedes $k \in \{1, \dots, c\}$ partiell differenzierbar nach der k-ten Variablen ist.

(d) Man nennt f **partiell differenzierbar**, wenn f in jedem Punkt aus Ω partiell differenzierbar ist. Wenn zusätzlich alle partiellen Ableitungen $\partial_1 f, \dots, \partial_c f : \Omega \to \mathbb{R}$ stetig ist, dann nennt man f **stetig partiell differenzierbar**.

Wenn $\Omega \subseteq \mathbb{R}^d$ offen ist, $x \in \Omega$ gilt und $f : \Omega \to \mathbb{R}$ gilt, dann folgt sofort aus Definition 4.1.4(b), dass f genau dann im Punkt x nach der ersten Variablen partiell differenzierbar, wenn der Schnitt

$$f(\,\cdot\,, x_2, \dots, x_c) : (x_1 - \delta, x_1 + \delta) \to \mathbb{R}, \quad y_1 \to f(y_1, x_2, \dots, x_c)$$

(für genügend kleines δ) differenzierbar im Punkt x_1 ist; in diesem Fall ist

$$\partial_1 f(x_1, x_2, \dots, x_c)$$

einfach die Ableitung von $f(\,\cdot\,, x_2, \dots, x_c)$ im Punkt x_1. Um partielle Differenzierbarkeit von f nachzuprüfen und gegebenenfalls die partiellen Ableitungen zu berechnen, muss man als lediglich Differenzierbarkeit und Ableitungen in einer Veränderlichen überprüfen – und damit haben Sie aus der Analysis 1 bereits viel Übung.

[6] Die Begriffe in dieser Definition kann man auch für Funktionen mit Werte in $\mathbb{R}^d$ einführen. Aber wie nach Proposition 4.1.3 besprochen, führt Teil (c) dieser Definition dazu, dass der skalarwertige (das heißt $\mathbb{R}$-wertige) Fall besonders relevant ist. Um die Dinge so einfach wie möglich zu halten, beschränken wir uns in dieser Definition deshalb auf den skalarwertigen Fall.

[7] Ein solches δ existiert wegen der Offenheit von Ω und der Stetigkeit von $t \mapsto x + tr$.

Beispiel 4.1.5 (Partielle Ableitungen einer Funktion)**.** Sei $f : \mathbb{R}^2 \to \mathbb{R}$, $x \mapsto x_1 x_2$. Für jedes festes $x_2 \in \mathbb{R}$ ist der Schnitt

$$f(\,\cdot\,, x_2) : \mathbb{R} \to \mathbb{R}, \quad x_1 \mapsto x_1 x_2$$

ein Polynom (mit der Variablen x_1) und somit stetig differenzierbar. Analoges gilt, wenn man stattdessen x_1 festhält und x_2 als Variable verwendet. Also ist f stetig partiell differenzierbar. Die partiellen Ableitungen f an einem Punkt $x \in \mathbb{R}^2$ sind

$$\partial_1 f(x_1, x_2) = x_2 \quad \text{und} \quad \partial_2 f(x_1, x_2) = x_1.$$

Auf dieses Beispiel kommen wir in Kürze in Beispiel 4.1.5 noch einmal zurück.

Der Witz ist nun, dass man stetige totale Differenzierbarkeit von Funktionen durch stetige partielle Differenzierbarkeit charakterisieren kann. Das ist der Inhalt des folgendes Theorems.

Vorlesung 16: (Mi, 04.06.) beginnt mit Theorem 4.1.6

Theorem 4.1.6. *Sei $c \in \mathbb{N}$ und sei $\Omega \subseteq \mathbb{R}^c$ offen.*

(a) *Sei $f : \Omega \to \mathbb{R}$. Die Funktion f ist genau dann stetig total differenzierbar wenn sie stetig partiell differenzierbar ist. In diesem Fall gilt*

$$\mathrm{D}f(x) = \begin{pmatrix} \partial_1 f(x) & \dots & \partial_c f(x) \end{pmatrix}$$

für alle $x \in \Omega$.

(b) *Allgemeiner gilt folgendes: Sei $d \in \mathbb{N}$ und sei*

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_d \end{pmatrix} : \Omega \to \mathbb{R}^d,$$

für skalarwertige Funktionen $f_1, \dots, f_d : \mathbb{R} \to \Omega$. Es ist f genau dann stetig total differenzierbar wenn sie stetig partiell differenzierbar ist. In diesem Fall gilt

$$\mathrm{D}f(x) = \begin{pmatrix} \partial_1 f_1(x) & \dots & \partial_c f_1(x) \\ \dots & & \dots \\ \partial_1 f_d(x) & \dots & \partial_c f_d(x) \end{pmatrix}$$

für alle $x \in \Omega$.

Beweis. (a) "$\Rightarrow$" Diese Implikation besprechen wir weiter unten in Proposition 4.1.9 allgemeiner für alle Richtungsableitungen.

"$\Leftarrow$" Dies ist die "interessante" Implikationen in dem Sinne, dass man sie in vielen konkreten Beispielen verwendet um zu zeigen, dass eine Funktion tatsächlich total differenzierbar ist. An dieser Stelle verzichten wir aus Zeitgründen auf den Beweis und verweisen stattdessen auf die Literatur, siehe zum Beispiel [2, §6, Satz 2 auf S. 79].

(b) Dies folgt mit Hilfe von Proposition 4.1.3(c) aus (a) □

Die Implikation von partieller Differenzierbarkeit zu totaler Differenzierbarkeit stimmt im Allgemeinen nicht mehr, wenn man die Annehme, dass die partiellen Ableitungen stetig sind, weglässt. Ein Gegenbeispiel hierzu besprechen wir in den Übungen.

Theorem 4.1.6 spielt für die Differentialrechnung in mehreren Veränderlichen eine ähnliche Rolle wie Theorem 3.3.9 für die Integralrechnung in mehreren Veränderlichungen: Es erlaubt uns die Situation in mehreren Veränderlichen auf die Situation in einer Veränderlichen zurückzuführen. Lasen Sie uns das nochmal anhand der Funktion aus Beispiel 4.1.5 veranschaulichen:

Beispiel 4.1.7 (Totale Ableitung und Gradient einer Funktion)**.** Sei $f : \mathbb{R}^2 \to \mathbb{R}$, $x \mapsto x_1 x_2$. Diese Funktion ist laut Beispiel 4.1.5 stetig partiell differenzierbar mit

$$\partial_1 f(x) = x_2 \quad \text{und} \quad \partial_2 f(x) = x_1$$

für alle $x \in \mathbb{R}^2$. Laut Theorem 4.1.6 ist f somit auch stetig total differenzierbar und ihre Ableitung und ihr Gradient an einem Punkt $x \in \mathbb{R}^2$ sind gegeben durch

$$\mathrm{D}f(x) = \begin{pmatrix} \partial_1 f(x) & \partial_2 f(x) \end{pmatrix} = \begin{pmatrix} x_2 & x_1 \end{pmatrix} \qquad \text{und somit} \qquad \nabla f(x) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}.$$

Wie in der Analysis 1 kann man auch für Funktionen, die von mehreren Veränderlichen abhängen, eine Produktregel beweisen. Es wird Sie auf den ersten Blick vielleicht etwas überraschen, dass man diese aus der Kettenregeln und aus Beispiel 4.1.7 herleiten kann:

Proposition 4.1.8 (Produktregel)**.** *Sei $c \in \mathbb{N}$, sei $\Omega \subseteq \mathbb{R}^c$ offen und sei $x \in \Omega$.*

(a) *Seien $g_1, g_2 : \Omega \to \mathbb{R}$ im Punkt x differenzierbar. Dann ist auch das Produkt $g_1 g_2 : \Omega \to \mathbb{R}$ im Punkt x differenzierbar und es gilt*

$$\mathrm{D}(g_1 g_2)(x) = g_1(x)\mathrm{D}g_2(x) + g_2(x)\mathrm{D}g_1(x).$$

(b) *Allgemeiner gilt folgendes: Seien $g_1, g_2 : \Omega \to \mathbb{R}^d$ im Punkt x differenzierbar. Dann ist auch die Abbildung $g_1^{\mathrm{T}} g_2 : \Omega \to \mathbb{R}$ im Punkt x differenzierbar und es gilt*

$$\mathrm{D}(g_1^{\mathrm{T}} g_2)(x) = g_1(x)^{\mathrm{T}}\mathrm{D}g_2(x) + g_2(x)^{\mathrm{T}}\mathrm{D}g_1(x).$$

Beweis. (a) Sei $f : \mathbb{R}^2 \to \mathbb{R}$, $y \mapsto y_1 y_2$ und sei

$$g := \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} : \Omega \to \mathbb{R}^2.$$

Dann ist $g_1 g_2 = f \circ g$. Laut Proposition 4.1.3(c) ist g differenzierbar im Punkt x mit

$$\mathrm{D}g(x) = \begin{pmatrix} \mathrm{D}g_1(x) \\ \mathrm{D}g_2(x) \end{pmatrix}$$

und laut Beispiel 4.1.7 ist f auf ganz $\mathbb{R}^2$ total differenzierbar und hat an jedem Punkt $y \in \mathbb{R}^2$ die Ableitung

$$\mathrm{D}f(y) = \begin{pmatrix} y_2 & y_1 \end{pmatrix}.$$

Also ist laut Proposition 4.1.3(b) auch die Hintereinanderausführung $g_1 g_2 = f \circ g$ differenzierbar im Punkt x und besitzt dort die Ableitung

$$\begin{aligned} \mathrm{D}(g_1 g_2)(x) &= \mathrm{D}(f \circ g)(x) = (\mathrm{D}f)(g(x)) \cdot (\mathrm{D}g)(x) \\ &= \begin{pmatrix} g_2(x) & g_1(x) \end{pmatrix} \begin{pmatrix} \mathrm{D}g_1(x) \\ \mathrm{D}g_2(x) \end{pmatrix} = g_2(x)\mathrm{D}g_1(x) + g_1(x)\mathrm{D}g_2(x). \end{aligned}$$

Also ist (a) bewiesen.

(b) Den Beweis kann man analog zum Beweis von (a) führen. □

Wenn eine Abbildung total differenzierbar ist, dann ist sie auch in jede Richtung differenzierbar. Das zeigen wir in der folgenden Proposition.

Proposition 4.1.9 (Richtungsableitungen von total differenzierbaren Funktionen)**.** *Sei $c \in \mathbb{N}$, sei $\Omega \subseteq \mathbb{R}^c$ offen und $x \in \Omega$. Sei $f : \Omega \to \mathbb{R}$ total differenzierbar im Punkt x. Dann ist für jedes $r \in \mathbb{R}^c \setminus \{0\}$ die Funktion f im Punkt x in Richtung r differenzierbar und es gilt*

$$\partial_r f(x) = \mathrm{D}f(x)r.$$

Beweis. Sei $r \in \mathbb{R}^c \setminus \{0\}$. Für genügend kleines $\delta > 0$ betrachten wir die Kurve[8]

$$\gamma : (-\delta, \delta) \to \Omega, \quad t \mapsto x + tr.$$

Diese ist differenzierbar mit $\dot{\gamma}(t) = r$ für alle $t \in (-\delta, \delta)$. Somit ist, siehe Formel (4.1.1),

$$\mathrm{D}\gamma(0) = \dot{\gamma}(0) = r.$$

Laut Definition 4.1.4(a) müssen wir zeigen, dass die Abbildung

$$\alpha : (-\delta, \delta) \to \mathbb{R}, \quad t \mapsto f(x + tr)$$

im Punkt 0 differenzierbar ist und dort die Ableitung $\mathrm{D}f(x)r$ besitzt.

Da $\alpha = f \circ \gamma$ gilt und f laut Voraussetzung im Punkt $\gamma(0) = x$ total differenzierbar ist, ist α laut Proposition 4.1.3(c) im Punkt 0 ebenfalls differenzierbar und erfüllt die Kettenregel

$$\dot{\alpha}(0) = \mathrm{D}\alpha(0) = \mathrm{D}f(\gamma(0))\mathrm{D}\gamma(0) = \mathrm{D}f(x)r.$$

Somit ist die Behauptung gezeigt. □

[8] Warum muss man δ gegebenenfalls sehr klein wählen, damit die Definition dieser Kurve Sinn ergibt?

Mit Hilfe der Richtungsableitung kann man dem Gradienten einer Funktion eine interessante geometrische Interpretation geben. Erinnern Sie sich daran, dass für zwei Vektoren $x, y \in \mathbb{R}^c$ das Standard-Skalarprodukt definiert ist als

$$(x \mid y) = \sum_{k=1}^{c} x_k y_k.$$

Mit Hilfe der Matrixmultiplikation, die Sie in der Zwischenzeit in der Linearen Algebra 1 kennengelernt haben, kann man diese Formel auch kurzer als Produkt eines Zeilen- und eines Spaltenvektors schreiben: Es gilt

$$(x \mid y) = x^{\mathrm{T}} y = y^{\mathrm{T}} x.$$

Korollar 4.1.10 (Richtung des steilsten Anstiegs)**.** *Sei $c \in \mathbb{N}$, sei $\Omega \subseteq \mathbb{R}^c$ offen und sei $x \in \Omega$. Sei $f : \Omega \to \mathbb{R}$ total differenzierbar im Punkt x. Wir verwenden die Abkürzung $g := \nabla f(x) = (\mathrm{D}f(x))^{\mathrm{T}} \in \mathbb{R}^c$. Für alle weiteren Vektoren $r \in \mathbb{R}^c$ mit $\|r\|_2 = \|g\|_2$ gilt dann*

$$\partial_r f(x) \leq \partial_g f(x),$$

das heißt der Gradient $g = \nabla f(x)$ gibt diejenige Richtung an, in die f von x aus betrachtet am schnellsten ansteigt.

Beweis. Wegen $\mathrm{D}f(x) = (\nabla f(x))^{\mathrm{T}} = g^{\mathrm{T}}$ gilt laut Proposition 4.1.9

$$\partial_r f(x) = \mathrm{D}f(x) r = g^{\mathrm{T}} r = (g \mid r)$$

und analog dazu

$$\partial_g f(x) = (g \mid g) = \|g\|_2^2 .$$

Also erhalten wir aus der Cauchy–Schwarz-Ungleichung (Lemma 1.1.12)

$$\partial_r f(x) = (g \mid r) \leq |(g \mid r)| \leq \|g\|_2 \|r\|_2 = \|g\|_2^2 = \partial_g f(x),$$

wobei wir verwendet haben, dass jede reelle Zahl kleiner oder gleich ihrem Betrag ist und dass laut Voraussetzung $\|r\|_2 = \|g\|_2$ gilt. □

4.2 Zweite Ableitung und lokale Extreme

Theorem 4.1.6(a) rechtfertigt Teil (a) der folgenden Definition.

Definition 4.2.1 (Einfache und zweifache stetige differenzierbarkeit)**.** Sei $\Omega \subseteq \mathbb{R}^c$ offen und sei $f : \Omega \to \mathbb{R}$.

(a) Wenn die Funktion f stetig partiell differenzier (äquivalent dazu: stetig total differenzierbar) ist, dann nennt man f auch kurz **stetig differenzierbar**.

(b) Man nennt die Funktion **zweimal stetig differenzierbar** oder kurz $\boldsymbol{C^2}$, wenn Sie stetig differenzierbar ist und all Ihre partiellen Ableitungen $\partial_1 f, \dots, \partial_c f$ ebenfalls stetig differenzierbar sind.

Das folgende Resultat nennt man auch den **Satz von Schwarz**.

Theorem 4.2.2 (Vertauschen partieller Ableitungen)**.** *Sei $\Omega \subseteq \mathbb{R}^c$ offen und sei $f : \Omega \to \mathbb{R}$ zweimal stetig differenzierbar. Dann gilt für alle $j, k \in \{1, \dots, d\}$ und alle $x \in \Omega$ die Gleichheit*

$$\partial_j \partial_k f(x) = \partial_k \partial_j f(x).$$

Beweis. Wir verzichten an dieser Stelle auf den Beweis und verweisen stattdessen auf die Literatur, zum Beispiel auf [2, §5, Satz 1 auf S. 68]. □

Vorlesung 17: (Fr, 06.06.) beginnt mit Definition 4.2.3

Definition 4.2.3 (Hesse-Matrix)**.** Sei $\Omega \subseteq \mathbb{R}^c$ offen und sei $f : \Omega \to \mathbb{R}$ zweimal stetig differenzierbar. Für jedes $x \in \Omega$ nennt man die Matrix[9]

$$\mathrm{H}f(x) := \big(\mathrm{D}(\nabla f)\big)(x) := \begin{pmatrix} \partial_1\partial_1 f(x) & \dots & \partial_c\partial_1 f(x) \\ \vdots & & \vdots \\ \partial_1\partial_c f(x) & \dots & \partial_c\partial_c f(x) \end{pmatrix}$$

die **Hesse-Matrix** von f an der Stelle x.

Mit Hilfe der Hessematrix kann man die zweite Ableitung von f entlang von Strecken ausdrücken:

Proposition 4.2.4 (Zweite Ableitung von Funktionen entlang von Strecken)**.** *Sei $\Omega \subseteq \mathbb{R}^c$ offen und sei $f : \Omega \to \mathbb{R}$ zweimal stetig differenzierbar. Sei $I \subseteq \mathbb{R}$ ein offenes Intervall und sei $\gamma : I \to \Omega$ eine Kurve, die eine Strecke parametrisiert, das heißt, es sei*

$$\gamma(t) = x^* + tr$$

für zwei Vektoren $x^*, r \in \mathbb{R}^c$ und alle $t \in I$: Dann ist auch $f \circ \gamma$ zweimal stetig differenzierbar und hat die zweite Ableitung

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f \circ \gamma)(t) = r^{\mathrm{T}}\, \mathrm{H}f\big(\gamma(t)\big)\, r$$

für jedes $t \in I$.

[9] Manchmal (oft?) schreibt man anstelle von $\mathrm{H}f(x)$ auch $\mathrm{H}_f(x)$, was sich an der Notation $\mathrm{J}_f(x)$ für Jacobimatrix orientiert (siehe Definition 4.1.1(a). Wir verwenden die Notation $\mathrm{H}f(x)$, weil sie sich eher an der Notation $\mathrm{D}f(x)$ orientiert.

Beweis. Aus der Kettenregel erhält man, dass $f \circ \gamma$ stetig differenzierbar ist mit

$$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma)(t) = \mathrm{D}f\big(\gamma(t)\big)\,\dot{\gamma}(t) = \big(\nabla f(\gamma(t))\big)^{\mathrm{T}}\, r = r^{\mathrm{T}}\,\nabla f(\gamma(t)).$$

für alle $t \in I$. Weil f eine C^2-Funktion ist, ist $t \mapsto \nabla f(\gamma(t))$ aufgrund der Kettenregel differenzierbar und es gilt für alle $t \in I$

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla f(\gamma(t)) = (\mathrm{D}\nabla f)(\gamma(t))\;\dot{\gamma}(t) = \mathrm{H}f(\gamma(t))\,r.$$

Aufgrund der Produktregel in Proposition 4.1.8(b) erhalten wir somit, dass $t \mapsto \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma)(t)$ differenzierbar ist und die Gleichung

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f \circ \gamma)(t) = r^{\mathrm{T}}\,\frac{\mathrm{d}}{\mathrm{d}t}\nabla f(\gamma(t)) = r^{\mathrm{T}}\,\mathrm{H}f(\gamma(t))\,r$$

für alle $t \in I$ gilt. □

Wir benötigen die folgenden Begriffe aus der linearen Algebra:

Definition 4.2.5 (Symmetrie und positive (Semi)Definitheit von Matrizen)**.** Sei $c \in \mathbb{N}$ und $A \in \mathbb{R}^{c\times c}$.

(a) Die Matrix A heißt **symmetrisch**, falls ihre Einträge die Bedingung $A_{jk} = A_{kj}$ für alle $j, k \in \{1, \dots, c\}$ erfüllen.[10] Anders ausgedrückt: A ist genau dann symmetrisch, wenn $A = A^{\mathrm{T}}$ gilt, das heißt, wenn A gleich seiner transponierten Matrix ist.

(b) Sei A symmetrisch. Dann nennt man A **positiv definit**, falls $x^{\mathrm{T}}Ax > 0$ für alle $x \in \mathbb{R}^c \setminus \{0\}$ gilt;[11] man nennt A **positiv semidefinit**, falls $x^{\mathrm{T}}Ax \geq 0$ für alle $x \in \mathbb{R}^c \setminus \{0\}$ gilt.[12]

(c) Sei A symmetrisch. Negative Definitheit und negative Semidefinitheit sind analog definiert; außerdem nennt man A **indefinit**, falls A weder positiv semidefinit noch negativ semidefinit ist.

Bemerkung 4.2.6 (Die Hesse-Matrix ist symmetrisch)**.** Sei $\Omega \subseteq \mathbb{R}^c$ offen und $f : \Omega \to \mathbb{R}$ zweimal stetig differenzierbar. Für jedes $x \in \Omega$ ist die Hesse-Matrix $\mathrm{H}f(x) \in \mathbb{R}^{c\times c}$ symmetrisch. Das ist genau die Aussage des Satzes von Schwarz (Theorems 4.2.2).

[10] Für verwenden die Notation A_{jk} für den Eintrag von A in der j-ten Zeile und k-ten Spalte. Anders ausgedrückt: Sie müssen in der Matrix j Schritte noch unten und k-Schritte nach rechts gehen um den Eintrag A_{jk} zu erhalten.

[11] Beachten Sie, dass $x^{\mathrm{T}} = (x \mid Ax)$ ist, das heißt Positive Definitheit bedeutet, dass $(x \mid Ax) > 0$ für alle $x \in \mathbb{R}^c$ gilt.

[12] Bei der Definition positivier Semidefinitheit kann man genauso gut auch fordern, dass $x^{\mathrm{T}}Ax \geq 0$ für alle $x \in \mathbb{R}^c$ gilt. Warum ist das so?

In der linearen Algebra werden Sie lernen, dass mit positive und negative (Semi)Definitheit einer symmetrischen Matrix $A \in \mathbb{R}^{c\times c}$ mit Hilfe der sogenannten **Eigenwerte** von A charakterisieren kann. Für 2×2-Matrizen gibt es außerdem das folgende sehr einfache Kriterium für diese Eigenschaften:

Proposition 4.2.7 (Positive (semi-)Definitheit in Dimension **2**)**.** *Sei $A \in \mathbb{R}^{2\times 2}$ symmetrisch, das heißt es ist*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

mit $A_{12} = A_{21}$.

(a) *Die Matrix A ist positiv definite genau dann, wenn* $\det(A) := A_{11}A_{22} - A_{21}A_{12} > 0$ *und* $\operatorname{tr}(A) := A_{11} + A_{22} > 0$ *gilt.*[13]

(b) *Die Matrix A ist positiv semidefinit genau dann, wenn* $\det(A) \geq 0$ *und* $\operatorname{tr}(A) \geq 0$ *gilt.*

Beweis. Auf einen Beweis verzichten wir an dieser Stelle. Der Beweis lässt sich später relativ leicht führen, wenn Sie in der Linearen Algebra gelernt haben, was Eigenwerte von Matrizen sind und wie diese mit positiver (Semi-)Definitheit sowie mit der Determinante und Spur einer Matrix zusammenhängen. □

Definition 4.2.8 (Lokale Extremstellen)**.** Sei $M \subseteq \mathbb{R}^c$, sei $x^* \in M$ und $f : M \to \mathbb{R}$.

(a) Man nennt x^* eine **lokale Maximalstelle** von f, falls es ein $\varepsilon > 0$ gibt mit der Eigenschaft

$$f(x) \leq f(x^*) \quad \text{für alle } x \in M \cap B_{<\varepsilon}(x^*).$$

Man nennt x^* eine **lokale strikte Maximalstelle** von f, falls es ein $\varepsilon > 0$ gibt mit der Eigenschaft

$$f(x) < f(x^*) \quad \text{für alle } x \in M \cap B_{<\varepsilon}(x^*) \text{ mit } x \neq x^*.$$

(b) Lokale Minimalstelllen und strikte lokale Minimalstellen sind analog definiert.

(c) Man nennt x^* eine **lokale Extremstelle** von f, falls x^* eine lokale Maximalstelle oder eine lokale Minimalstelle von f ist.

Aus der Analysis 1 wissen Sie, dass man mit Hilfe der ersten und zweiten Ableitung einer Funktion ein hinreichendes Kriterium dafür angeben kann, dass ein Punkt eine lokale Minimalstelle oder Maximalstelle einer Funktion ist. Proposition 4.2.4 legt nahe, dass ein ähnliches Resultat auch für Funktionen in mehreren Veränderlichen gilt, wenn man die zweite Ableitung durch die Hessematrix ersetzt. Das ist tatsächlich der Fall:

[13] Übrigens ist die Zahl $\det(A)$ die sogenannte **Determinante** von A und die Zahl $\operatorname{tr}(A)$ die sogenannte **Spur** von A ("Spur" heißt auf Englisch "trace", das erklärt die Notation tr). Über diese Zahlen – insbesondere über die Determinante – werden Sie in der Linearen Algebra mehr lernen.

Theorem 4.2.9 (Lokale Extremstellen)**.** *Sei $\Omega \subseteq \mathbb{R}^c$ offen, sei $x^* \in \Omega$ und $f : \Omega \to \mathbb{R}$.*

(a) *Wenn f im Punkt x^* total differenzierbar ist und x^* eine lokale Extremstelle von f ist, dann gilt $\mathrm{D}f(x^*) = 0$.*[14]

Für den Rest des Theorems sei f zweimal stetig differenzierbar.

(b) *Wenn $\mathrm{D}f(x^*) = 0$ gilt und $\mathrm{H}f(x^*)$ positiv definit ist,*[15] *dann ist x^* eine lokale strikte Minimalstelle von f.*

(c) *Wenn $\mathrm{D}f(x^*) = 0$ gilt und $\mathrm{H}f(x^*)$ negativ definit ist, dann ist x^* eine lokale strikte Maximalstelle von f.*

Beweis. (a) Sei f im Punkt x^* total differenzierbar und sei x^* eine lokale Extremstelle von f. Wir betrachten einen beliebigen, festen Vektor $r \in \mathbb{R}^c \setminus \{0\}$ und zeigen, dass $\mathrm{D}f(x^*)r = 0$ gilt.

Wähle $\delta > 0$ so kleine, dass $x + tr \in \Omega$ für alle $t \in (-\delta, \delta)$ gilt; solch ein δ existert, weil Ω offen ist. Die Abbildung

$$\alpha : (-\delta, \delta) \to \mathbb{R}, \quad t \mapsto f(x + tr),$$

ist dann wegen Proposition 4.1.9 in 0 differenzierbar und besitzt in 0 ein lokales Minimum. Aus der Analysis 1 wissen Sie, dass somit $\dot{\alpha}(0) = 0$ gilt. Indem wir die Formel aus Proposition 4.1.9 für die Richtungsableitung verwenden, erhalten wir

$$0 = \dot{\alpha}(0) = \partial_r f(x^*) = \mathrm{D}f(x^*)r,$$

womit die Behauptung gezeigt ist.[16]

(b) Für den Beweis verweisen wir beispielsweise auf [2, Satz 4 in Kapitel 7 auf Seite 95].

(c) Dieses Resultat erhält man, indem man (b) auf die Funktion $-f$ anwendet. □

Auf Hausaufgabenblatt 10 werden Sie eine notwendige Bedingung für lokale Extrema kennen lernen, die nicht in einer offenen Menge liegen.

Vorlesung 18: (Mi, 18.06.) beginnt mit Definition 4.2.10

Definition 4.2.10 (Konvexe Mengen und Funktionen)**.**

(a) Eine Menge $M \subseteq \mathbb{R}^c$ heißt **konvex**, falls für alle $x, y \in M$ auch die Verbindungsstrecke

$$\big\{(1-t)x + yt \mid \ t \in [0,1]\big\}$$

in M liegt.

[14] Äquivalent dazu ist natürlich $\nabla f(x^*) = 0$, n denn $\nabla f(x^*)$ ist schlicht nur der transponierte Vektor von $\mathrm{D}f(x^*)$.

[15] Beachten Sie, dass diese Voraussetzung Sinn ergibt, da die Hessematrix $\mathrm{D}f(x^*)$ aufgrund des Satzes von Schwarz (Theorem 4.2.2) symmetrisch ist; siehe Bemerkung 4.2.6.

[16] Einen Augenblick: Weshalb folgt aus der Gleichung $\mathrm{D}f(x^*)r = 0$ für alle $r \in \mathbb{R}^c \setminus \{0\}$, dass $\mathrm{D}f(x^*) = 0$ gilt?

(b) Sei $M \subseteq \mathbb{R}^c$ konvex und sei $f : M \to \mathbb{R}$. Die Funktion f heißt **konvex**, falls für alle $x, y \in M$ and alle $t \in [0, 1]$ die Ungleichung

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

gilt.[17] Die Funktion f heißt **konkav**, wenn $-f$ konvex ist.

Proposition 4.2.11 (Konvexität via zweiter Ableitung)**.** *Sei $\Omega \subseteq \mathbb{R}^c$ offen und konvex und sei $f : \Omega \to \mathbb{R}$ zweimal stetig differenzierbar. Dann sind die folgenden Aussagen äquivalent:*

(a) *Die Funktion f ist konvex.*

(b) *Für alle $x \in \Omega$ ist die Hessematrix $\mathrm{H}f(x)$ positiv semidefinit.*

Beweis. Das folgt mit Hilfe von Proposition 4.2.4 aus dem entsprechenden eindimensionalen Resultat der Analysis 1. □

4.3 Taylorentwicklung

In Analysis 1 haben wir zwei verschiedene Restglieddarstellungen für die Taylorentwicklung gesehen. Eine der beiden – aus Aufgabe 5 auf Hausaufgabenblatt 14 in der Analysis 1 – wiederholen wir in folgendem Theorem noch einmal.

Theorem 4.3.1 (Wiederholung: Taylorentwicklung mit Integral-Restglied)**.** *Sei $n \in \mathbb{N}_0$, sei $I \subseteq \mathbb{R}$ ein Intervall mit mehr als einem Punkt und sei $f : I \to \mathbb{R}$ $(n+1)$-mal stetig differenzierbar. Sei $x^* \in I$. Dann gilt*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x^*)}{k!}(x - x^*)^k + \frac{1}{n!}\int_{x^*}^{x} (x-s)^n f^{(n+1)}(s)\,\mathrm{d}s$$

für alle $x \in I$.

Der Sinn von Taylorentwicklungen wird etwas klarer ersichtlich, wenn man das vorangehende Theorem benutzt um zu zeigen, dass sich manche Funktionen als Reihe darstellen lassen.

Korollar 4.3.2 (Taylorreihe)**.** *Sei $I \subseteq \mathbb{R}$ ein Intervall mit mehr als einem Punkt und sei $f : I \to \mathbb{R}$ beliebig oft differenzierbar.*[18] *Seien $x^*, x \in I$. Falls*

$$\frac{1}{n!}\int_{x^*}^{x} (x-s)^n f^{(n+1)}(s)\,\mathrm{d}s \to 0$$

[17] Das ist äquivalent dazu, dass die Menge in $\mathbb{R}^{c+1}$, die oberhalb das Graphen von f liegt, konvex ist (diese Äquivalenz ist anschaulich sehr eingängig; wenn man sie wirklich präzise beweisen will, muss man aber ein kleines bisschen Arbeit in den Beweis stecken).

[18] **Beliebig oft differenzierbar** bedeutet, für jedes $n \in \mathbb{N}$ ist f n-mal differenzierbar.

für $n \to \infty$ gilt, dann ist

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x^*)}{k!}(x - x^*)^k.$$

Die Bedingung, dass der Integralterm für $n \to \infty$ gegen 0 konvergiert, wirkt auf den ersten Blick etwas technisch, lässt sich aber manchmal gut nachprüfen, wie man im folgenden Beispiel sehen kann. Falls Sie später eine Einführung in die Funktionentheorie[19] hören, werden Sie feststellen, dass diese Bedingung für viele Funktionen sogar automatisch erfüllt ist und auf recht überraschende Weise damit zusammenhäng, wie man den Definitionsbereich der Funktion in die komplexe Ebene fortsetzen kann.

Beispiel 4.3.3 (Taylorreihen von Sinus und Kosinus)**.** Aus Korollar 4.3.2 erhält man (für $x = x^*$) leicht, dass

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

$$\text{und} \qquad \cos(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$$

für alle $x \in \mathbb{R}$ gilt. Alternativ kann man diese Formeln aus der Definition von Sinus und Kosinus mithilfe der komplexen Exponentialfunktion herleiten (damit kann man sehen, dass diesselben Formeln sogar für alle $x \in \mathbb{C}$ gelten).

Bemerkung 4.3.4 (Mehrdimensionale Taylor-Entwicklung)**.** Ähnlich wie im ein-Dimensionalen kann man auch für Funktionen in mehreren Variablen eine Taylor-Entwicklung angeben. Diese ist etwas komplizierter als in einer Variable. Wir beschränken uns an dieser Stelle darauf, sie bis zum Grad 2 anzugeben, denn für diesen Fall kann man sie gut mit Hilfe der Hesse-Metrix ausdrücken.

Sei $\Omega \subseteq \mathbb{R}^c$ offen, sei $f : \Omega \to \mathbb{R}$ eine Funktion, die "genügend glatt" ist (sagen wir der Einfachkeit halber, f sei C^3) und sei $x^* \in \Omega$. Dann gibt es eine Funktion $e_2 : \Omega \to \mathbb{R}$ derart, dass

$$f(x) = f(x^*) + (\mathrm{D}f)(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^{\mathrm{T}} \mathrm{H}f(x^*)(x - x^*) + e_2(x)$$

für alle $x \in \Omega$ gilt und derart, dass $\|e_2(x)\|$ für $x \to x^*$ schneller als $\|x - x^*\|_2^2$ gegen 0 konvergiert.

Es ist hilfreich, diese Formel mit der ein-dimensionalen Situation zu vergleichen: Wenn $\Omega \subseteq \mathbb{R}$ ein Intervall ist, dann lautet die entsprechende Formel

$$f(x) = f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2} f''(x^*)(x - x^*)^2 + e_2(x)$$

[19] Auch genannt: Komplexe Analysis.

für alle $x \in \Omega$. Wenn Sie die Terme

$$\frac{1}{2}f''(x^*)(x-x^*)^2 \qquad \text{und} \qquad \frac{1}{2}(x-x^*)^{\mathrm{T}}\mathrm{H}f(x^*)(x-x^*)$$

vergleich, sollten Sie sich klarmachen, weshalb die Ausdrücke $(x-x^*)^{\mathrm{T}}$ und $(x-x^*)$ links und rechts der Hessematrix stehen, damit die Dimensionen für die Matrixmultiplikation zusammenpassen.

4.4 Das Newton-Verfahren, implizit definierte Funktionen und der Umkehrsatz

Vorlesung 19: (Fr, 19.06.) beginnt mit Abschnitt 4.4

Bevor wir zu den Hauptresultaten dieses Abschnitts kommen, sprechen wir über das **Newton-Verfahren**, mit dessen Hilfe von Lösungen von Gleichungen numerisch bestimmen kann.

Diskussion 4.4.1 (Das Newton-Verfahren zur Lösung von Gleichungen)**.**

(1) Sei $f : \mathbb{R} \to \mathbb{R}$ eine C^1-Funktion. Wir würden gerne eine **Nullstelle** $x^* \in \mathbb{R}$ von f finden, das heißt eine Zahl x^* mit der Eigenschaft $f(x^*) = 0$. Newton entwickelte hierzu folgendes Verfahren:

Wähle zunächst irgendeine Zahl $x_0 \in \mathbb{R}$. Man betrachtet im Punkt x_0 die Tangente von f. Falls $f'(x_0) \neq 0$ ist, besitzt die Tangente genau eine Nullstelle. Diese ist, wie eine kurze Rechnung zeigt, gleich der Zahl

$$x_1 := x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Weil die Tangente anschaulich eine "gute lineare Approximation" an f ist, ist es naheliegend darauf zu hoffen (aber nicht immer wahr, siehe unten), dass x_1 näher an x^* liegt als x_0.

Nun iterierte man diese Vorgehen, das heißt, man betrachtet die Tangente an f im Punkt x_1, bezeichnet – falls $f'(x_1) \neq 0$ ist – deren Nullstelle mit x_2, und so weiter. Insgesamt erhält man somit eine rekursiv definierte Folge $(x_n)_{n\in\mathbb{N}_0}$ mit Startwert x_0 und der Rekursionsvorschrift

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

für alle $n \in \mathbb{N}$. Die Idee ist, dass – unter geeigneten Voraussetzungen – die Folge $(x_n)_{n\in\mathbb{N}}$ gegen x^* konvergieren sollte. Wir besprechen an dieser Stelle nicht, unter welchen Voraussetzungen das Verfahren tatsächlich konvergiert; das kön-

nen Sie zum Beispiel in einer Vorlesung über numerische Analysis lernen. [20] [21]

(2) Sei $a \in (0, \infty)$ und sei $x_0 \in (0, \infty)$ eine Zahl mit $x_0^2 \geq a$. Wenn man das Newtonverfahren aus (2) auf die Funktion $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto x^2 - a$ anwendet, erhält man die Folge $(x_n)_{n \in \mathbb{N}_0}$ mit der Rekursionsvorschrift

$$x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n},$$

also

$$x_{n+1} = \frac{1}{2}\big(x_n + \frac{a}{x_n}\big).$$

Diese Folgt konvergiert gegen $\sqrt{a}$. Das ist das sogenannte **Heron-Verfahren** zur näherungsweisen Berechnung von Quadratwurzeln, welches Sie in Analysis 1 als Bonusaufgabe auf Hausaufgabenblatt 6 gesehen haben.

Das Heron-Verfahren ist also ein Spezialfall des Newtonverfahrens. Allerdings ist das Heron-Verfahren schon seit einigen tausend Jahren bekannt und somit viel älter als das Newton-Verfahren.

(3) Analog zu (1) kann man auch ein Newton-Verfahren in mehreren Dimensionen angeben. Für eine C^1-Funktion $f : \mathbb{R}^d \to \mathbb{R}^d$ und einen Startwert $x^{(0)} \in \mathbb{R}^d$ verwendet man dann völlig analog zu (1) die Rekursionsvorschrift

$$x_{n+1} = x_n - \big(Df(x_n)\big)^{-1} f(x_n)$$

für alle $n \in \mathbb{N}_0$, sofern jede der Matrizen $Df(x_n) \in \mathbb{R}^{d \times d}$ invertierbar ist. Auch in diesem Fall kann man hinreichende Bedingungen dafür angeben, dass das Verfahren gegen eine Nullstelle von f konvergiert.

Auch hierzu können Sie Details in einer Vorlesung über numerische Analysis lernen.

Ein Nachteil des mehr-dimensionalen Newton-Verfahrens, dass soeben in (3) erwähnt wurde, besteht – neben den nicht besprochenen Voraussetzungen an f – darin, dass man in jedem Schritt eine $d \times d$-Matrix invertieren muss. Im Theorem 4.4.4 unten werden wir ein **vereinfachtes Newton-Verfahren** besprechen, für das nur eine einzige Matrix invertiert werden muss. Dafür benötigen wir zur Vorbereitung sogenannte **Matrix-Normen**, die wir nun besprechen.

[20] Die Voraussetzungen, die man benötigt, sind übrigens relativ stark. Zum Beispiel müssen die Voraussetzungen ja sicherstellen, dass f überhaupt eine Nullstelle besitzt. Außerdem müssen Sie dafür sorgen, dass $f'(x_n) \neq 0$ für jedes n gilt damit x_{n+1} wohldefiniert ist.

[21] Ein Vorteil des Newton-Verfahrens ist, dass es – wenn die Funktion f genügend gute Eigenschaften hat – extrem schnell gegen eine Nullstelle konvergiert. Insbesondere konvergiert es oft viel schneller als das **Bisektionsverfahren**, dass in Analysis 1 als Bonusaufgabe auf Hausaufgabenblatt 7 besprochen wurde.

Übrigens: Wir haben hier zwar die numerische Berechnung von Nullstellen als Motivation verwendet, allerdings ist der wirkliche Grund, weshalb wir Theorem 4.4.4 in der Analysis 2 besprechen, dass es uns zum Satz über implizit definierte Funktionen führt, zu dem wir anschließend kommen.

Definition 4.4.2 (Induzierte Matrixnorm)**.** Seien $c, d \in \mathbb{N}$. Seien $\|\cdot\|_\gamma : \mathbb{R}^c \to [0,\infty)$ und $\|\cdot\|_\delta : \mathbb{R}^d \to [0,\infty)$ Normen und sei $A \in \mathbb{R}^{d\times c}$. Dann nennt man die Zahl

$$\|A\|_{\delta\leftarrow\gamma} \coloneqq \sup\left\{ \|Ax\|_\delta \mid \ x \in \mathbb{R}^c \text{ und } \|x\|_\gamma = 1\right\}$$

die von $\|\cdot\|_\gamma$ und $\|\cdot\|_\delta$ **induzierte Matrixnorm** von A.[22]

Proposition 4.4.3 (Eigenschaften von induzierten Matrixnormen)**.** *Seien $b, c, d \in \mathbb{N}$. Seien $\|\cdot\|_\beta : \mathbb{R}^b \to [0,\infty)$, $\|\cdot\|_\gamma : \mathbb{R}^c \to [0,\infty)$ und $\|\cdot\|_\delta : \mathbb{R}^d \to [0,\infty)$ Normen.*

(a) *Für jedes $A \in \mathbb{R}^{c\times b}$ ist die Abbildung*

$$\mathbb{R}^b \to \mathbb{R}^c, \quad x \mapsto Ax$$

stetig.

(b) *Für jedes $A \in \mathbb{R}^{c\times b}$ ist $\|A\|_{\gamma\leftarrow\beta} \in [0,\infty)$. Außerdem ist das Supremum in Definition 4.4.2 sogar ein Maximum.*

(c) *Es ist $\|\cdot\|_{\gamma\leftarrow\beta} : \mathbb{R}^{c\times b} \to [0,\infty)$ eine Norm auf dem $\mathbb{R}$-Vektorraum $\mathbb{R}^{c\times b}$.*[23]

(d) *Sei $A \in \mathbb{R}^{c\times b}$. Eine Folge $(A^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{R}^{c\times b}$ konvergiert genauu dann bezüglich $\|\cdot\|_{\gamma\leftarrow\beta}$ gegen A, wenn sie komponentenweise gegen A konvergiert (das heißt, wenn $A^{(n)}_{jk} \to A_{jk}$ für alle Indizes j, k gilt).*

(e) *Falls $b = c$ und $\|\cdot\|_\beta = \|\cdot\|_\gamma$ ist, besitzt die Einheitsmatrix $I \in \mathbb{R}^{c\times c}$ die induzierte Matrixnorm $\|I\|_{\gamma\leftarrow\gamma} = 1$.*

(f) *Es gilt die folgende **Submultiplikativität**: Für alle $A \in \mathbb{R}^{c\times b}$ und alle $B \in \mathbb{R}^{d\times c}$ ist*

$$\|BA\|_{\delta\leftarrow\beta} \le \|B\|_{\delta\leftarrow\gamma} \|A\|_{\gamma\leftarrow\beta}.$$

(g) *Für alle $A \in \mathbb{R}^{c\times b}$ and alle $x \in \mathbb{R}^b$ gilt*

$$\|Ax\|_\gamma \le \|A\|_{\gamma\leftarrow\beta} \|x\|_\beta.$$

[22] Falls Sie Lust haben, später einmal eine Vorlesung in Funktionalanalysis zu hören, werden Sie noch eine weitere Bezeichne für die induzierte Matrixnorm kennenlernen: Man bezeichnet sie dort auch als **Operatornorm**, da man sie allgemeinerfür sogenannte **lineare Operatoren** definieren kann.

[23] Das rechtfertigt die Verwendung des Begriffs **Matrixnorm**.

Beweis. Die Beweise sind recht elementar. Wir verzichten an dieser Stelle auf die Details. □

Theorem 4.4.4 (Vereinfachtes Newton-Verfahren)**.** *Sei $\mathbb{R}^c$ mit einer Norm $\|\cdot\|_\gamma$ ausgestattet. Sei $\Omega \subseteq \mathbb{R}^c$ offen und sei $f : \Omega \to \mathbb{R}^c$ eine C^1-Funktion. Seien $A \in \mathbb{R}^{c\times c}$ eine invertierbare Matrix, $\hat{x} \in \Omega$ ein Punkt und $r > 0$ eine Zahl mit folgenden Eigenschafen:*

(1) *Es gilt $B_{\leq r}(\hat{x}) \subseteq \Omega$.*

(2) *Es sei $\left\|A^{-1}\right\|_{\gamma\leftarrow\gamma} \|\mathrm{D}f(x) - A\|_{\gamma\leftarrow\gamma} \leq \frac{1}{2}$ für alle $x \in B_{\leq r}(\hat{x})$.*

(3) *Es sei $\left\|A^{-1}\right\|_{\gamma\leftarrow\gamma} \|f(\hat{x})\|_\gamma \leq \frac{r}{2}$.*

Dann gilt folgendes:

(a) *Die Funktion f besitzt in $B_{\leq r}(\hat{x})$ genau eine Nullstelle x^**

(b) *Es gilt $x - A^{-1}f(x) \in B_{\leq r}(\hat{x})$ für alle $x \in B_{\leq r}(\hat{x})$ und die Funktion*

$$\begin{aligned} F : B_{\leq r}(\hat{x}) &\to B_{\leq r}(\hat{x}), \\ x &\mapsto x - A^{-1}f(x) \end{aligned}$$

ist Lipschitz-stetig mit Lipschitz-Konstante $\frac{1}{2}$.

(c) *Für jedes $x \in B_{\leq r}(\hat{x})$ konvergiert $F^n(x)$ für $n \to \infty$ gegen x^*.*

Beweis. (b) Wir betrachten G zuerst als Abbildung

$$\begin{aligned} G : \Omega &\to \mathbb{R}^c, \\ x &\mapsto x - A^{-1}f(x). \end{aligned}$$

Um F zu erhalten, werden wir nachher G auf die Kugel $B_{\leq r}(\hat{x})$ einschränken. Wegen der Kettenregel ist G eine C^1-Funktion mit Ableitung

$$\mathrm{D}G(x) = I - A^{-1}\mathrm{D}f(x) = A^{-1}(A - \mathrm{D}f(x))$$

für alle $x \in \Omega$. Wegen der Submultiplikativität der Matrixnorm folgt somit aus Annahme (2), dass $\|\mathrm{D}G(x)\|_{\gamma\leftarrow\gamma} \leq \frac{1}{2}$ für alle $x \in B_{\leq r}(\hat{x})$. Mit Hilfe des Hauptsatzes der Differential- und Integralrechnung und der Fundamentalabschätzung für Integrale folgt hieraus, dass $G|_{B_{\leq r}(x_0)}$ Lipschitz-stetig mit Lipschitz-Konstante $\frac{1}{2}$ ist.[24] Außerdem gilt wegen Annahme (3)

$$\|G(\hat{x}) - \hat{x}\|_\gamma = \left\|A^{-1}f(\hat{x})\right\|_\gamma \leq \frac{r}{2}$$

[24] Können Sie dieses Argument im Detail ausführen?

und somit für alle $x \in B_{\leq r}(\hat{x})$

$$\|G(x) - x_0\|_\gamma \leq \|G(x) - G(\hat{x})\|_\gamma + \|G(\hat{x}) - \hat{x}\|_\gamma \leq \frac{1}{2}\|x - \hat{x}\|_\gamma + \frac{r}{2} \leq r.$$

Also ist tatsächlich $x - A^{-1}x = G(x) \in B_{\leq r}(\hat{x})$ für alle $x \in B_{\leq r}(\hat{x})$. Die Lipschitz-Stetigkeit von $F = G|_{B_{\leq r}(\hat{x})}$ mit Lipschitzkonstante $\frac{1}{2}$ haben wir bereits gezeigt.

(a) und (c) Für jedes $x \in B_{\leq r}(\hat{x})$ ist $f(x) = 0$ äquivalent zu $F(x) = x$. Also sind die Nullstellen von f in $B_{\leq r}(\hat{x})$ genau die Fixpunkte von F. Da F laut (b) Lipschitzstetig mit Lipschitz-Konstante $\frac{1}{2}$ ist und die abgeschlossene Kugel $B_{\leq r}(\hat{x})$ vollständig ist, folgt aus dem Fixpunktsatz von Banach (Theorem 1.4.3 und Aufgabe 4(a) auf Hausaufgabenblatt 4), dass F genau einen Fixpunkt x^* besitzt und dass man ihn auf die Weise, wie in (c) behauptet, approximieren kann. □

Vorlesung 20: (Mi, 25.06.) beginnt mit Bemerkung 4.4.5

Bemerkung 4.4.5 (Nullstelle in der offenen Kugel)**.** Wenn man in Theorem 4.4.4 in Annahme (3) sogar die strikte Ungleichung dass $\left\|A^{-1}\right\|_{\gamma\leftarrow\gamma}\|f(\hat{x})\|_\gamma < \frac{r}{2}$ voraussetzt, dann zeigt der Beweis, dass das Bild von F sogar in der offenen Kugel $B_{<r}(\hat{x})$ liegt. Insbesondere liegt also der Fixpunkt von F – welcher gleich der Nullstelle von f in $B_{\leq r}(\hat{x})$ ist – in $B_{<r}(\hat{x})$.

Um ein Beispiel für Theorem 4.4.4 angeben zu können, brauchen wir zunächst eine Möglichkeit um induzierte Matrixnormen zu berechnen. Für die Normen $\|\cdot\|_{1\leftarrow 1}$ und $\|\cdot\|_{\infty\leftarrow\infty}$ ist die relativ einfach, wie die folgenden Formeln zeigen:

Beispiele 4.4.6 (Induzierte Matrixnormen bzgl. der **1**-Norm und der **∞**-Norm)**.**

(a) Für jedes $A \in \mathbb{R}^{c\times d}$ gilt

$$\|A\|_{1\leftarrow 1} = \max\left\{\|A_{\bullet,1}\|_1, \ldots, \|A_{\bullet,d}\|_1\right\},$$

wobei $A_{\bullet,1}, \ldots, A_{\bullet,d} \in \mathbb{R}^c$ die Spalten von A bezeichnen

Aufgrund dieser Formel bezeichnet man $\|\cdot\|_{1\leftarrow 1}$ manchmal auch als die **Spaltensummen-Maximums-Norm**, denn man nimmt von jeder Spalte die Summe der Beträge der Einträge und von all diesen Summen das Maximum.

(b) Für jedes $A \in \mathbb{R}^{c\times d}$ gilt

$$\|A\|_{\infty\leftarrow\infty} = \max\left\{\|A_{1,\bullet}\|_1, \ldots, \|A_{c,\bullet}\|_1\right\},$$

wobei $A_{1,\bullet}, \ldots, A_{c,\bullet} \in \mathbb{R}^{1\times d}$ die Zeilen von A bezeichnen (und die 1-Norm eines Zeilenvektors analog zur 1-Norm eines Spaltenvektors definiert ist).

Wegen dieser Formel bezeichnet man $\|\cdot\|_{\infty\leftarrow\infty}$ manchmal als die **Zeilensummen-Maximums-Norm**, denn man nimmt von jeder Zeile die Summe der Beträge der Einträge und von all diesen Summen das Maximum.[25]

[25] Falls Sie Schwierigkeiten haben sich zu Merken, welche der Normen $\|\cdot\|_{1\leftarrow 1}$ und $\|\cdot\|_{\infty\leftarrow}\infty$ die Spaltensummen-Maximums-Norm ist und welche die Zeilensummen-Maximums-Norm ist, können Sie zum Beispiel die folgende Eselsbrücke verwenden (sie ist womöglich etwas albern, aber das muss bei Eselsbrücken ja so sein...):

"Die Ziffer 1 'steht', sowie es eine Spalte tut, und das Symbol ∞ 'liegt', sowie es eine Zeile tut."

Beweis. Diese Formeln werden in den Übungen besprochen. □

Beispiel 4.4.7 (Anwendung des vereinfachten Newton-Verfahrens)**.** Betrachten wir die Funktion $f : \mathbb{R}^2 \to \mathbb{R}^2$, die durch

$$f(x) = \begin{pmatrix} x_2 + \frac{1}{3} + \frac{1}{4}x_1^2 \\ x_1 + \frac{1}{5} + \frac{1}{4}x_1 x_2 \end{pmatrix}$$

Wir wollen Theorem 4.4.4 anwenden um zu zeigen, dass f eine Nullstelle besitzt und um eine Nullstelle näherungsweise zu berechnen.

Dazu statten wir $\mathbb{R}^2$ mit der ∞-Norm aus, wählen $\Omega = \mathbb{R}^2$, $\hat{x} = 0$ und $r = 1$. Die Funktion f ist in jeder Komponente ein Polynom, also C^1. Ihre Ableitung ist gegeben durch

$$\mathrm{D}f(x) = \begin{pmatrix} \frac{1}{2}x_1 & 1 \\ 1 + \frac{1}{4}x_2 & \frac{1}{4}x_1 \end{pmatrix}$$

für alle $x \in \mathbb{R}^2$. Wir wählen

$$A = \mathrm{D}f(\hat{x}) = \mathrm{D}f(0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Die Matrix A ist invertierbar und es gilt $A^{-1} = A$. Somit ist $\left\|A^{-1}\right\|_{1\leftarrow 1} = 1$. Lassen Sie uns nun nachprüfen, dass die Annahmen (1)–(3) aus Theorem 4.4.4 erfüllt sind:

(1) Die Inklsuion $B_{\leq 1}(0) \subseteq \mathbb{R}^2$ ist klar.

(2) Für jedes $x \in B_{\leq 1}(0)$ gilt $\max\{|x_1|, |x_2|\} = \|x\|_\infty \leq 1$ und somit

$$\begin{aligned} \left\|A^{-1}\right\|_{\infty\leftarrow\infty} \cdot \|\mathrm{D}f(x) - A\|_{\infty\leftarrow\infty} &= \left\| \begin{pmatrix} \frac{1}{2}x_1 & 0 \\ \frac{1}{4}x_2 & \frac{1}{4}x_1 \end{pmatrix} \right\|_{\infty\leftarrow\infty} \\ &= \max\left\{ \frac{1}{2}|x_1|, \frac{1}{4}|x_2| + \frac{1}{4}|x_1| \right\} \leq \frac{1}{2}. \end{aligned}$$

(3) Es ist $f(0) = (\frac{1}{3}, \frac{1}{5})^{\mathrm{T}}$ und somit

$$\|A\|_{\infty\leftarrow\infty} \|f(0)\|_\infty = \frac{1}{3} \leq \frac{1}{2} = \frac{r}{2}.$$

Die alle Voraussetzungen erfüllt sind, folgt aus Theorem 4.4.4, dass f in $B_{\leq 1}(0)$ genau eine Nullstelle x^* besitzt und dass für jedes $x \in B_{\leq 1}(0)$ die Eigenschaft $F^n(x) \to x^*$ für $n \to \infty$ gilt, wobei $F : B_{\leq 1}(0) \to B_{\leq 1}(0)$ durch

$$F(x) = x - A^{-1}f(x) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_2 + \frac{1}{3} + \frac{1}{4}x_1^2 \\ x_1 + \frac{1}{5} + \frac{1}{4}x_1 x_2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{5} - \frac{1}{4}x_1 x_2 \\ -\frac{1}{3} - \frac{1}{4}x_1^2 \end{pmatrix}$$

für alle $x \in B_{\leq 1}(0)$ gegeben ist. Wenn wir als Startwert zum Beispiel $x = \hat{x} = 0$ wählen, erhalten wir für die ersten wir bis zur fünften Iteration die folgenden Vektoren (die wir der Einfachkeit halber nur auf sechs Nachkommastellen genau angeben):

$$F^0(0) = \begin{pmatrix} 0.000000 \\ 0.000000 \end{pmatrix}, \quad F^1(0) = \begin{pmatrix} -0.200000 \\ -0.333333 \end{pmatrix}, \quad F^2(0) \approx \begin{pmatrix} -0.216666 \\ -0.343333 \end{pmatrix},$$
$$F^3(0) = \begin{pmatrix} -0.218597 \\ -0.345069 \end{pmatrix}, \quad F^4(0) = \begin{pmatrix} -0.218857 \\ -0.345279 \end{pmatrix}, \quad F^5(0) \approx \begin{pmatrix} -0.218891 \\ -0.345308 \end{pmatrix},$$
$$F^6(0) = \begin{pmatrix} -0.218896 \\ -0.345311 \end{pmatrix}, \quad F^7(0) = \begin{pmatrix} -0.218896 \\ -0.345312 \end{pmatrix}, \quad F^8(0) \approx \begin{pmatrix} -0.218896 \\ -0.345312 \end{pmatrix}.$$

Sie sehen, dass sich die ersten sechs Nachkommastellen von $F^7(0)$ nach $F^8(0)$ bereits nicht mehr verändert haben.[26] Wenn man $F^8(0)$ sogar bis auf Maschinengenauigkeit angibt,[27] erhält man die genaueren Werte

$$F^8(0) \approx \begin{pmatrix} -0.21889695267031908 \\ -0.3453123023028488 \end{pmatrix}.$$

Lassen Sie uns diesen Vektor in f einsetzen. Dann erhält man

$$f(F^8(0)) \approx \begin{pmatrix} 2.5724769536772385 \cdot 10^{-12} \\ 3.0973314191218293 \cdot 10^{-12} \end{pmatrix}.$$

Die ∞-Norm von $f(F^\infty(0))$ ist also kleiner als $10^{-11} = 0.00000000001$.

Im Rest dieses Abschnitts wollen wir noch zwei wichtige Resultate ansprechen: Den Satz über implizit definierte Funktionen (Theorem 4.4.8) und den Satz von der Umkehrfunktion (Theorem 4.4.9). Als Motivation für den ersten dieser beiden Sätze betrachten wir die folgende Gleichung für Vektoren $z \in \mathbb{R}^2$:

$$z_1^2 + z_2^2 = 1. \tag{4.4.1}$$

Die Menge $M \subseteq \mathbb{R}^2$ der Punkte, die diese Gleichung erfüllen, ist gleich der euklidischen Einheitskreislinie. Man kann nun versuchen, Teil dieser Menge als einen Funktionsgraphen zu beschreiben: Wir können zum Beispiel z_1 als Variable auffassen

[26] Wobei es im Prinzip möglich ist, dass sie sich später noch einmal verändern. Man kann allerdings im Fixpunktsatz von Banach auch eine explizite Fehlerabschätzung angeben und diese verwenden um eine Fehlerabschätzung für das vereinfachten Netwon-Verfahren aus Theorem 4.4.4 zu beweisen. Damit kann man ausrechnen, wieviele Iterationen für eine gewünschte Genauigkeit ausreichen.

[27] Wobei bei der Berechnung Gleitkommazahlen mit doppelter Genauigkeit (englisch "double precision") benutzt wurden, wie es heute für viele Berechnungen üblich ist. Jede solche Zahl besteht aus 64 Bit, wovon eines für das Vorzeichen, 11 Bit für den Exponenten der Zahl und 52 Bit für die sogenannten *Mantisse* verwendet werden.

Genauere Informationen dazu können Sie zum Beispiel auf der englischen Wikipedia-Seite nachlesen: `https://en.wikipedia.org/wiki/Double-precision_floating-point_format`

und eine Funktion $f : I \to \mathbb{R}$ von einem reellen Intervall I nach $\mathbb{R}$ suchen, derart, dass der Vektor $(z_1, f(z_1))^{\mathrm{T}}$ für alle $z_1 \in I$ in der Menge M liegt – oder anders ausgedrückt, dass

$$z_1^2 + f(z_1)^2 = 1 \tag{4.4.2}$$

für alle $z_1 \in I$ gilt. Wie man die Funktion f wählen muss, damit das funktioniert, hängt natürlich von der Gleichung (4.4.1) beziehungsweise (4.4.2) ab. Weil diese Gleichung nicht nach $f(x_1)$ aufgelöst ist, sagt man, dass die Funktion f durch die Gleichung (4.4.2) **implizit definiert** ist. Nun dränge sich mehrere Fragen auf:

(1) Gibt es solch ein f überhaupt und, falls ja, welche Teile der Kreiseslinie M lassen sich damit beschreiben?

(2) Falls es solch ein f gibt, kann man es dann explizit berechnen?

Wenn man versucht die Gleichung (4.4.2) nach $f(x_1)$ aufzulösen, dann erhält man zum Beispiel die beiden Funktionen

$$[-1,1] \to \mathbb{R}, \quad z_1 \mapsto \sqrt{1 - z_1^2},$$
$$[-1,1] \to \mathbb{R}, \quad z_1 \mapsto -\sqrt{1 - z_1^2}$$

als mögliche Wahlen für f. Der erste Wahl beschreibt die obere Hälfte von der Einheitskreislinie und die zweite Wahl beschrebt die untere Hälfte. Hierbei fällt folgendes auf:

- Die beiden Wahlen von f; die wir gefunden haben, sind auf abgeschlossenen Mengen definiert Wenn wir Funktionen wollen, die auf offenen Mengen definiert sind, erwischen wir die beiden Punkte des Einheitskreises, die ganz links und ganz rechts liegen, nicht.

- Jede der beiden Wahlen von f ist in den Punkten -1 und 1 nicht differenzierbar. Anschaulich hängt das damit zusammen, dass für die Funktion $g : \mathbb{R}^2 \to \mathbb{R}$, $x \mapsto z_1^2 + z_2^2 - 1$, deren Nullstellenmenge gleich der Einheitskreislinie ist, die zweite partielle Ableitung $\partial_2 g(x)$ in den Punkten $(-1,0)^{\mathrm{T}}$ und $(1,0)^{\mathrm{T}}$ gleich 0 ist.

Der folgende Satz gibt es sehr allgemeine Kriterium dafür an, dass man eine Gleichung nach einer implizit definierten Funktion auflösen kann und dass die implizit definierte Funktion bijektiv ist.

Zur Erinnerung: Sofern nichts anderes gesagt wurde, seien $\mathbb{R}^c$, $\mathbb{R}^d$, und so weiter, stets mit der Euklidischen Norm ausgestattet.

Theorem 4.4.8 (Satz über implizit definierte Funktionen)**.** *Sei $\Omega \subseteq \mathbb{R}^{c+d}$ offen und sei $g : \Omega \to \mathbb{R}^d$ eine C^1-Funktion. Für jedes $z \in \Omega$ teilen wir die Ableitung $\mathrm{D}g(z)$ auf als $\mathrm{D}g(z) = \big(B(z), C(z)\big)$ mit $B(z) \in \mathbb{R}^{d\times c}$ und $C(z) \in \mathbb{R}^{d\times d}$.*

Es sei $\hat{z} \in \Omega$ mit $g(\hat{z}) = 0$ und die $d \times d$-Matrix $C(\hat{z})$ sei invertierbar. Dann gibt es offene Mengen $U \subseteq \mathbb{R}^c$ und $V \subseteq \mathbb{R}^d$ und eine C^1-Funktion $f : U \to V$ mit folgenden Eigenschaften:

(a) *Es gilt $\hat{z} \in U \times V \subseteq \Omega$.*

(b) *Für jedes $x \in U$ und jedes $y \in V$ gilt*

$$g(x, y) = 0 \quad \Leftrightarrow \quad f(x) = y.$$

Wenn wir schreiben, dann gilt für jedes $x \in U$ die Formel

$$C(x, f(x))\mathrm{D}f(x) = -B(x, f(x)).$$

Beweis. Seien $\hat{x} \in \mathbb{R}^c$ and $\hat{y} \in \mathbb{R}^d$ derart, dass

$$\hat{z} = \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix}$$

gilt. Da g laut Voraussetzung C^1 ist, ist $C : \Omega \to \mathbb{R}^{d \times d}$, $z \mapsto C(z)$ stetig. Wir verwenden die Abkürzung $A := C(\hat{z}) \in \mathbb{R}^{d \times d}$.

In der Vorlesung besprechen wir nur die grobe Beweisidee von Theorem 4.4.8. Bei Interesse können Sie den Beweis hier im Detail nachlesen.

Nun wählen wir eine so kleine Zahl $r \in (0, \infty)$, dass $B_{\leq r}(\hat{x}) \times B_{\leq 2r}(\hat{y}) \subseteq \Omega$ ist und dass

$$\left\|A^{-1}\right\|_{2\leftarrow 2} \|C(z) - A\|_{2\leftarrow 2} \leq \frac{1}{2}$$

für alle $z \in B_{\leq r}(\hat{x}) \times B_{\leq r}(\hat{y})$ gilt. Außerdem wählen wir $s \in (0, r]$ so klein, dass

$$\left\|A^{-1}\right\|_{2\leftarrow 2} \|g(x, \hat{y})\|_2 < \frac{r}{2}$$

für alle $x \in B_{\leq s}(\hat{x})$ gilt. Wir setzen nun $U := B_{<s}(\hat{x})$ und $V := B_{<r}(\hat{y})$. Dann gilt wie gewünscht $\hat{z} \in U \times V \subseteq B_{\leq r}(\hat{x}) \times B_{\leq r}(\hat{y}) \subseteq \Omega$.

Für jedes $x \in B_{<s}(\hat{x})$ betrachten wir die Funktion

$$h_x : B_{<2r}(\hat{x}) \to \mathbb{R}^d, \quad y \mapsto g(x, y).$$

Das ist eine C^1-Funktion und ihr totales Differential erfüllt $\mathrm{D}h_x(y) = C(x, y)$ für alle $x \in B_{<s}(\hat{x})$ und alle $y \in B_{\leq r}(\hat{y})$. Für jedes $x \in B_{<s}(\hat{x})$ erfüllt die Funktion h_x somit die Annahmen von Theorem 4.4.4, wobei sie in Annahme (3) sogar eine strikte Ungleichung erfüllt. Theorem 4.4.4 und Bemerkung 4.4.5 ergeben somit für jedes $x \in B_{<s}(\hat{x})$, dass h_x genau eine Nullstelle – nennen wir sie y_x – in $B_{<r}(\hat{y})$ besitzt. Somit gilt für jedes $x \in B_{<s}(\hat{x})$ und jedes $y \in B_{<r}(\hat{y})$: Es ist $g(x, y) = 0$ genau dann, wenn $h_x(y) = 0$ ist, genau dann, wenn $y = y_x$ ist. Also erfüllt die Funktion

$$f : U \to V, \quad x \mapsto y_x$$

die in (b) behauptete Eigenschaft.

Für den Beweis der stetigen Differenzierbarkeit von f verweisen wir auf die Literatur, beispielsweise auf [2, Kapitel 8]. Um die Formel für die Ableitung von f zu zeigen, betrachten wir zuerst die Abbildung

$$k : U \to \mathbb{R}^{c+d}, \quad x \mapsto \begin{pmatrix} x \\ f(x) \end{pmatrix}.$$

Da f differenzierbar ist, gilt dasselbe auch für k und es ist

$$\mathrm{D}k(x) = \begin{pmatrix} I \\ \mathrm{D}f(x) \end{pmatrix}$$

für alle $x \in U$. Für alle $x \in U$ ist $g(k(x)) = 0$, also folgt

$$\begin{aligned} 0 &= \mathrm{D}(g \circ k)(x) = \mathrm{D}g(k(x))\mathrm{D}k(x) \\ &= \begin{pmatrix} B(x, f(x)) & C(x, f(x)) \end{pmatrix} \begin{pmatrix} I \\ \mathrm{D}f(x) \end{pmatrix} = B(x, f(x)) + C(x, f(x))\mathrm{D}f(x) \end{aligned}$$

für alle $x \in U$, was die behauptete Formel zeigt. □

Übrigens gilt folgendes (wir verwenden die Notation aus dem Beweis): Weil $C(\hat{z})$ invertierbar ist, ist $C(z)$ ebenfalls invetierbar, wenn z nahe genug bei $\hat{z}$ liegt, dann man kann zeigen, dass die Menge der invertierbaren Matrizen in $\mathbb{R}^{d\times d}$ offen ist. Wenn x nahe an $\hat{x}$ liegt, liegt $f(x)$ nahe an $f(\hat{x}) = \hat{y}$, also ist dann auch $C(x, f(x))$ invertierbar. Somit erhält man aus dem Theorem die Formel

$$\mathrm{D}f(x) = -C(x, f(x))^{-1} B(x, f(x))$$

für alle x, die nahe an $\hat{x}$ liegen.

Als Spezialfall von Theorem 4.4.8 erhält man das folgende Resultat.

Vorlesung 21: (Fr, 27.06.) beginnt mit Theorem 4.4.9

Theorem 4.4.9 (Satz von der Umkehrfunktion)**.** *Sei $\Omega \subseteq \mathbb{R}^d$ offen, sei $h : \Omega \to \mathbb{R}^d$ eine C^1-Funktion, und sei $\hat{y} \in \Omega$ ein Punkt, für den die Matrix $\mathrm{D}h(\hat{y}) \in \mathbb{R}^{d\times d}$ invertierbar ist. Dann gibt es offene Mengen $V \subseteq \Omega$ und $U \subseteq \mathbb{R}^d$ mit $\hat{y} \in V$ und $h(\hat{y}) \in U$ mit folgenden Eigenschaften:*

(a) *Die Einschränkung $h|_V$ bildet V bijektiv nach U ab.*

(b) *Für jedes $y \in V$ ist die Matrix $\mathrm{D}h(y)$ invertierbar.*

(c) *Die Umkehrabbildung $h|_V^{-1} : U \to V$ ist ebenfalls C^1 und für alle $x \in U$ und alle $y \in V$ mit $h(y) = x$ gilt*

$$\big(\mathrm{D}h|_V^{-1}\big)(x) = \big(\mathrm{D}h(y)\big)^{-1}.$$

In der Vorlesung besprechen wir nur die grobe Beweisidee von Theorem 4.4.9. Bei Interesse können Sie den Beweis hier im Detail nachlesen.

Beweis. Wir definieren die Funktion $g : \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ durch

$$g(x, y) = -x + h(y)$$

für alle $(x, y) \in \mathbb{R}^d \times \Omega$. Dann ist g stetig differenzierbar mit

$$\mathrm{D}g(x, y) = \begin{pmatrix} -\mathrm{id} & \mathrm{D}h(y) \end{pmatrix}$$

für alle $(x, y) \in \mathbb{R}^d \times \Omega$. Da $g(h(\hat{y}), \hat{y}) = 0$ gilt und die Matrix $\mathrm{D}h(\hat{y})$ invertierbar ist, können wir den Satz über implizit definierte Funktionen (Theorem 4.4.8) auf die Funktion g und den Punkt $\hat{z} := (h(\hat{y}), \hat{y}) \in \mathbb{R}^d \times \Omega$ anwenden. Wir erhalten somit offene Mengen $U, V \subseteq \mathbb{R}^d$ mit $(h(\hat{y}), \hat{y}) \in U \times V \subseteq \mathbb{R}^d \times \Omega$ und eine C^1-Funktion $f : U \to V$ derart, dass für alle $x \in U$ und alle $y \in V$ die Eigenschaft die Eigenschaft $f(x) = y$ äquivalent ist zu $g(x, y) = 0$, was wiederum zu $x = h(y)$ äquivalent ist.

Nun setzen wir $\tilde{V} := h^{-1}(U) \cap V$; dann ist $\tilde{V}$ ebenfalls offene und eine Teilmenge von Ω. Es ist $h(\hat{y}) \in U$ und $\hat{y} \in V$ und somit $\hat{y} \in \tilde{V}$.

Für jedes $x \in U$ erhalten wir, indem wir $y := f(x) \in V$ setzen, $h(f(x)) = h(y) = x$; insbesondere ist somit $f(x) \in h^{-1}(U)$ und folglich $f(x) \in \tilde{V}$, das heißt, f bildet U sogar in die kleinere Menge $\tilde{V} \subseteq V$ ab. Umgekehrt erhalten wir für jedes $y \in \tilde{V} \subseteq V$, indem wir $x := h(y) \in U$ setzen, $f(h(y)) = f(x) = y$. Insgesamt erfüllen also die Abbildungen

$$h|_{\tilde{V}} : \tilde{V} \to U \qquad \text{und} \qquad f : U \to \tilde{V}$$

die Eigenschaften $h|_{\tilde{V}} \circ f = \mathrm{id}_U$ und $f \circ h|_{\tilde{V}} = \mathrm{id}_{\tilde{V}}$, das heißt, die beiden Abbildungen sind inverse zueinander. Also ist $h|_{\tilde{V}}$ bijektiv von $\tilde{V}$ nach U und seine Umkehrfunktion f ist C^1.

Die Formel für das Ableitung von f erhält man leicht aus der Kettenregel. □

Beispiel 4.4.10 (Polarkoordinaten)**.** Betrachten wir die Abbildung $h : \mathbb{R}^2 \to \mathbb{R}^2$,

$$h(r, \varphi) := \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \end{pmatrix}.$$

Anschaulich bezeichnet man r und φ als **Polarkoordinaten**. Die Abbildung h ist surjektiv, das heißt, man kann jeden Vektor $x \in \mathbb{R}^2$ in der Form $x = h(r, \varphi)$ darstellen. Hierbei kann man r sogar aus $[0, \infty)$ wählen; dann ist $r = \|x\|_2$ und φ der Winkel, den x mit der x_1-Achse einschließt.

Beachten Sie, dass h allerdings nicht injektiv ist: Wenn Sie einem Winkel φ ein ganzzahliges Vielfaches von 2π addieren, erhalten Sie wieder denselben Punkt, das heißt es ist $h(r, \varphi) = h(r, \varphi + k2\pi)$ für alle $r \in \mathbb{R}^2$ Außerdem gilt $h(0, \varphi) = 0$ für alle $\varphi \in \mathbb{R}$; anschaulich bedeutet das, dass man im Abstand 0 zum Ursprung immer den Ursprung erhält, egal welchen Winkel man wählt.

Die Funktion h ist C^1 und hat die Ableitung

$$\mathrm{D}h(r, \varphi) = \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{pmatrix}$$

für alle $(r, \varphi)^{\mathrm{T}} \in \mathbb{R}^2$. In der Linearen Algebra werden Sie lernen, dass eine quadratische Matrix genau dann invertierbar ist, wenn ihre Determinante ungleich 0 ist.

Für 2×2 hatten wir in Proposition 4.2.7 bereits erwähnt, wie man die Determinante berechnen kann: Es ist

$$\det \mathrm{D}h(r, \varphi) = r(\cos \varphi)^2 + r(\sin \varphi)^2 = r,$$

also ist $\mathrm{D}h(r, \varphi)$ genau dann invertierbar, wenn $r \neq 0$ ist. Der Satz von der Umkehrfunktion (Theorem 4.4.9) sagt uns somit: Für jeden Vektor $(\hat{r}, \hat{\varphi})^{\mathrm{T}} \in \mathbb{R}^2$ mit $\hat{r} \neq 0$ finden wir eine offene Menge $V \subseteq \mathbb{R}^2$, die den Vektor enthält, und eine offene Menge $U \subseteq \mathbb{R}^2$, die $h(\hat{r}, \hat{\varphi})$ enthält, und für die $h|_V$ bijektiv von V nach U ist mit C^1-Umkehrfunktion. Anschaulich ausgedrückt: Polarkoordinaten lassen sich lokal auf differenzierbare Weise invertieren.

Anderseits funktioniert so eine lokale Invertierung nicht, wenn $\hat{r} = 0$ ist. Man sagt deshab, die Polarkoordinaten besitzen für den Radius 0 eine **Koordinateningularität**.

Weil der Radius 0 also Ärger macht und umgekehrt positive Radien ausreichen um alle Punkt in $\mathbb{R}^2 \setminus \{0\}$ darzustellen, schränkt man sich meistens darauf ein Polarkoordinaten $(r, \varphi)^{\mathrm{T}} \in (0, \infty) \times \mathbb{R}$ zu betrachten.

Wir kommen in Abschnitt 5.2 im nächsten Kapitel noch einmal auf Koordinatentransformationen zurück.

Kapitel 5

Integralrechnung in verschiedenen Koordinaten

5.1 Das Lebesgue-Maß via Determinanten

In diesem Abschnitt und dem nachfolgenden Abschnitt wollen wir uns eine effiziente Möglichkeit überlegen um das Lebesgue-Maß von nicht-trivialen Mengen zu berechnen. Die folgende Überlegung ist dafür ganz zentral:

Lemma 5.1.1 (Maßänderung unter linearen Transformationen)**.** *Sei $A \in \mathbb{R}^{d\times d}$ und sei $H := [0,1]^d$ der Einheits-Würfel*[1] *in $\mathbb{R}^d$. Wenn die Menge $M \subseteq \mathbb{R}^d$ und ihr Bild $A(M)$ unter A messbar sind,*[2] *dann gilt*

$$\lambda_d(A(M)) = \lambda_d(A(H))\lambda_d(M).$$

Beachten Sie, dass der Ausdruck $\lambda_d(A(H))$ Sinn ergibt: Da H kompakt ist und die Abbildung $\mathbb{R}^d \to \mathbb{R}^d$, $x \mapsto Ax$ stetig ist, ist $A(H)$ ebenfalls kompakt, also abgeschlossen und somit messbar.

In Korollar 5.1.4 werden wir eine Variante des Lemma zeigen, die sich direkter anwenden besser verwenden lässt.

Beweisskizze von Lemma 5.1.1. Wir zeigen zuerst, dass die Behauptung stimmt, falls M ein Würfel ist, der parallel zum Koordinatensystem liegt. Sei also M solch ein Würfel, sagen wir mit Seitenlänge s. Dann gibt es ein $v \in \mathbb{R}^d$ mit $M = sH + v$. Also

[1] Wenn Sie angeben möchten, können Sie H auch **Einheits-Hyperwürfel** oder **Einheits-Hyperkubus** nennen, denn es handelt sich ja um meinen d-dimensionalen Würfel.

[2] Hier gibt es eine kleine Subtilität: Man kann sich fragen, ob aus der Messbarkeit von M nicht automatisch die Messbarkeit von $A(M)$ folgt. Wenn A invertierbar ist, stimmt das. (Können Sie begründen, weshalb?) Wenn A nicht invertierbar ist, kann es tatsächlich passieren, dass $A(M)$ nicht messbar ist – aber ist ein sehr pathologischer Effekt, der im Grunde daran liegt, dass "zu wenige" Mengen messbar sind. In der Analysis 3 kommen wir darauf zurück und werden dieses Problem lösen.

ist einerseits $\lambda_d(M) = \lambda_d(sH) = s^d$. Andererseits gilt $A(M) = sA(H) + Av$ und somit

$$\lambda_d(A(M)) = \lambda_d(sA(H)) = s^d \lambda_d(A(H)) = \lambda_d(A(H))\lambda_d(M).$$

Also stimmt die Behauptung in diesem Spezialfall. Nun betrachten wir eine allgemeine messbare Menge M.

Falls A nicht invertierbar ist, liegen $A(H)$ und $A(M)$ in einer $d-1$-dimensionalen Hyperebene und haben somit Lebesgue-Maß 0. Also gilt in diesem Fall

$$\lambda_d(A(M)) = 0 = 0 \cdot \lambda_d(M) = \lambda_d(A(H))\lambda_d(M).$$

Für den Rest des Beweises sei A also invertierbar.

Wir skizzieren den Rest des Beweises nur.[3] Wir können eine kleine Zahl reelle Zahl $s > 0$ und endlich viele paarweise disjunkte Würfel $W_1 \ldots, W_n \subseteq \mathbb{R}^d$ finden, die alle Seitenlänge s haben, parallel zum Koordinatensystem liegen und deren Vereinigung "bis auf einen kleinen Fehler" die Menge M ist. Entsprechend ist dann die Vereinigung von $A(W_1), \ldots, A(W_n)$ (die ebenfalls paarweise disjunkt sind, da B invertierbar ist) "bis auf einen kleinen Fehler" gleich $A(M)$. Es gilt also

$$\lambda_d(M) \approx \lambda_d\Big(\bigcup_{k=1}^{n} W_k\Big) = \sum_{k=1}^{n} \lambda_d(W_k)$$

und

$$\begin{aligned}\lambda_d(A(M)) &\approx \lambda_d\Big(\bigcup_{k=1}^{n} A(W_k)\Big) = \sum_{k=1}^{n} \lambda_d(A(W_k)) \\ &= \sum_{k=1}^{n} \lambda_d(A(H))\lambda_d(W_k) = \lambda_d(A(H)) \sum_{k=1}^{n} \lambda_d(W_k) \approx \lambda_d(A(H))\lambda_d(M),\end{aligned}$$

wobei wir für die Gleichheit zwischen den beiden Zeilen verwendet haben, dass wir die Behauptung im Speziallfall von Würfeln bereits am Anfang des Beweises gezeigt haben.[4] □

Damit wir Lemma 5.1.1 verwenden können, brauchen wir eine Möglichkeit um $\lambda_d(A(H))$ zu berechnen. Vielleicht überrascht es sie, dass die Zahl $\lambda_d(A(H))$ sehr eng mit einem Konzept aus der linearen Algebra zusammenhängt – nämlich mit Determinanten von Matrizen.

In Proposition 4.2.7 hatten wir bereits kurz angesprohen, dass für eine Matrix $A \in \mathbb{R}^{2\times 2}$ die **Determinante** definiert ist als die Zahl

$$\det A := A_{11}A_{22} - A_{21}A_{12} \in \mathbb{R}.$$

[3] Mehr Details über Beweise dieser Art können Sie in der Analysis 3 lernen.

[4] Um die Verwendnung des ungenauen Symbols $\approx$ zu vermeiden, müssten wir uns genauer überlegen, wie man den Fehler über der Überdeckung von $A(H)$ durch Würfel kontrollieren kann; außerdem müssten wir begründen, weshalb es solche Würfel überhaupt für jede messbare Menge gibt. Darauf verzichten wir an dieser Stelle und verweise erneut auf die Analysis 3 für mehr Details.

Man kann auch für quadratische Matizen größerer Dimension Determinanten definieren; in diesem Fall ist die Formel für die Determinante komplizierter.[5] Die Details besprechen Sie in der Linearen Algebra. Eine sehr nützliche Eigenschaft der Determinante ist, dass sie multiplikativ ist, das heißt, es gilt

$$\det(BA) = \det(B)\det(A)$$

für alle $A, B \in \mathbb{R}^{d\times d}$. Außerdem gilt für die Einheitsmatrix $I \in \mathbb{R}^{d\times d}$, dass $\det(I) = 1$ ist. Allgemeiner gilt für jede Diagonalmatrix[6] $D \in \mathbb{R}^{d\times d}$ – ist die Determinante das Produkt der Diagonaleinträge, das heißt, es gilt $\det(D) = \prod_{j=1}^{d} D_{jj}$.

Theorem 5.1.2 (Lebesgue-Maß und Deteminanten)**.** *Sei $A \in \mathbb{R}^{d\times d}$ und sei $H := [0,1]^d$ der Einheits-Würfel*[7] *in $\mathbb{R}^d$. Dann gilt*

$$\lambda_d(A(H)) = |\det(A)|\,.$$

Für den Beweis des Theorems brauchen wir noch ein paar Konzpete aus der linearen Algerba. Eine quadratische Matrix $U \in \mathbb{R}^{d\times d}$ heißt **orthogonal**, falls $U^{\mathrm{T}}U = I$ gilt. Orthogonale Matrizen sind genau diejneigen Matrizen, die Drehspiegelungen im $\mathbb{R}^d$ beschreiben. Deshalb gilt für jedes messbare Menge $M \subseteq \mathbb{R}^d$ und jede orthogonale Matrix $U \in \mathbb{R}^{d\times d}$, dass $\lambda_d(U(M)) = \lambda_d(M)$ ist. Man kann leicht sehen, dass die transponierte Matrix einer orthogonalen Matrix wieder orthogonal ist. Eine sehr nützlich Eigenschaft ist, dass die Determinaten einer orthogonalen Matrix immer 1 oder -1 ist:

Lemma 5.1.3 (Determinanten von orthogonalen Matrizen)**.** *Sei $U \in \mathbb{R}^{d\times d}$ orthogonal. Dann ist* $\det(U) \in \{-1, 1\}$.

Beweis. Es gilt

$$1 = \det(I) = \det(U^{\mathrm{T}})\det(U) = \big(\det(U)\big)^2,$$

also $\det(U) = 1$ oder $\det(U) = -1$. □

Beweisskizze von Theorem 5.1.2. Zuerst überlegen wir uns, dass die Formel stimmt, wenn A eine Diagognalmatrix ist und alle Einträge ≥ 0 sind. In diesem Fall ist

$$A(H) = [0, A_{11}] \times \cdots \times [0, A_{dd}],$$

[5] Für die Determinanten einer $d \times d$-Matrix A gilt die **Leibniz-Formel**

$$\det(A) = \sum_{\sigma\in\mathcal{S}_d} \operatorname{sgn}(\sigma) \prod_{j=1}^{d} A_{j\sigma(j)},$$

wobei $\mathcal{S}_d$ die **symmetrische Gruppe** auf d Elementen bezeichnet und für $\sigma \in \mathcal{S}_d$ die Zahl $\operatorname{sgn}(\sigma) \in \{-1, 1\}$ das **Vorzeichen** der Permutation σ ist.

[6] Eine quadratische Matrix D heißt **Diagonalmatrix**, falls außerhalb der Diagonalen alle Einträge 0 sind, das heißt, falls $D_{jk} = 0$ für alle Indizes $j \neq k$ gilt.

[7] Wenn Sie angeben möchten, können Sie H auch **Einheits-Hyperwürfel** oder **Einheits-Hyperkubus** nennen, denn es handelt sich ja um meinen d-dimensionalen Würfel.

also $\lambda_d(A(H)) = A_{11} \cdots A_{dd} = |\det(A)|$.

Um den Fall allgemeiner Matrizen und den Fall für Diagonalmatrizen zurückzuführen, muss man sich etwas mehr anstrengen. Wir verwenden die Bezeichnung

$$\varphi : \mathbb{R}^{d\times d} \to [0,\infty), \quad A \mapsto \lambda_d(A(H)).$$

Wegen Lemma 5.1.1 gilt für alle $A, B \in \mathbb{R}^{d\times d}$

$$\varphi(BA) = \lambda_d(B(A(H))) = \lambda_d(B(H))\lambda_d(A(H)) = \varphi(B)\varphi(A),$$

das heißt, φ ist multiplikativ.

Für jede orthogonale Matrix $U \in \mathbb{R}^{d\times d}$ gilt $\varphi(U) = \lambda_d(U(H)) = \lambda_d(H) = 1$. Nun verwenden wir ein Resultat aus der linearen Algebra, dass beispielsweise in der Numerik und in der Statistik eine wichtige Rolle spielt, die **Singulärwertzerlegung** einer Matrix: Das Resultat besagt, dass es für jedes $A \in \mathbb{R}^{d\times d}$ und Diagonalmatrix $D \in \mathbb{R}^{d\times d}$ mit Diagonaleinträgen $D_{11}, \ldots, D_{dd} \geq 0$ und zwei orthogonale Matrizen $U, V \in \mathbb{R}^{d\times d}$ gibt derart, dass $A = UDV^{\mathrm{T}}$ gilt.[8] Somit gilt einerseits

$$\varphi(A) = \varphi(U)\varphi(D)\varphi(V^{\mathrm{T}}) = \varphi(D)$$

und andererseits

$$|\det(A)| = |\det(U)|\,|\det(D)|\,\left|\det V^{\mathrm{T}}\right| = |\det(D)| = \varphi(D);$$

für die zweite Gleichheit haben wir Lemma 5.1.3 verwendet und für die dritte Gleichheit haben wir verwendet, dass wir die Behauptung für Diagonalmatrizen mit nichtnegativen Diagonaleinträgen bereits am Anfang des Beweises gezeigt haben. Insgesamt ist also $\varphi(A) = |\det(A)|$, wie behauptet. □

Korollar 5.1.4 (Maßänderung unter linearen Transformationen)**.** *Sei $A \in \mathbb{R}^{d\times d}$. Wenn eine Menge $M \subseteq \mathbb{R}^d$ und ihr Bild $A(M)$ unter A messbar sind, dann gilt*

$$\lambda_d(A(M)) = |\det(A)|\,\lambda_d(M).$$

Beweis. Das folgt sofort aus Lemma 5.1.1 und Theorem 5.1.2. □

Beispiel 5.1.5 (Fläche eines Parallelogramms)**.** Seien $x, y \in \mathbb{R}^2$. Das Parallelogramm P, dass von den beiden Vektoren x, y aufgespannt wird, besitzt die Fläche

$$|\det(x,y)| = |x_1y_2 - x_2y_1|\,.$$

Beweis. Sei $A = (x,y) \in \mathbb{R}^{2\times 2}$ die Matrix mit den beiden Spalten x und y. Weil das Rechteck $[0,1]^2$ von den beiden Vektoren e_1, e_2 aufgespannt wird, ist $A[0,1]^2$ das Parallelogramm, das von den beiden Vektoren

$$Ae_1 = x \qquad \text{und} \qquad Ae_2 = y$$

[8] Genau genommen ist es Resultat noch deutlicher allgemeiner und kann auch für nichtquadratische Matrizen formuliert werden, aber das benötigen wir an dieser Stelle nicht.

aufgespannt wird, also gleich P. Wegen Theorem 5.1.2 ist also

$$\lambda_2(P) = \lambda_2(A[0,1]^2) = |\det(A)|\,,$$

was gerade die Behauptung ist. □

Als Konsequenz des vorangehenden Beispiels erhält man sofort:

Vorlesung 22: (Do, 03.07.) beginnt mit Beispiel 5.1.6

Beispiel 5.1.6 (Dreiecksfläche)**.** Seien $x, y, z \in \mathbb{R}^2$. Das Dreieck mit den Eckpunkten x, y, z besitzt die Fläche

$$\frac{1}{2}\,|\det(y-x, z-x)|\,.$$

Beweis. Durch Verschieben um $-x$ erhält man das Dreieck mit den Eckpunkten 0, $y-x$ und $z-x$, das dieselbe Fläche besitzt. Seine Fläche ist halb so groß wie die Fläche des Parallelogramms, dass von $y-x$ und $z-y$ aufgespannt wird, und diese Parallelogrammfläche ist laut Beispiel 5.1.5 gleich $|\det(y-x, z-x)|$. □

5.2 Koordinatenwechsel

In diesem Abschnitt werden wir die Resultate aus dem vorangehenden Abschnitt 5.1 verwenden um zu untersuchen, wie man Integrale umschreiben kann, wenn man "die Koordinaten des Raunes transformiert". Sie werden im Laufe dieses Abschnitts sehen, was genau mit dieser Formulierung gemeint ist.

Der folgende Begriff ist dafür hilfreich:

Definition 5.2.1 (Diffeomorphismus)**.** Seien $\Omega, \tilde{\Omega} \subseteq \mathbb{R}^d$ offen

(a) Eine Funktion $h : \Omega \to \tilde{\Omega}$ heißt **Diffeomorphismus**[9] falls h bijektiv ist und sowohl h als auch die Umkehrfunktion $h^{-1} : \tilde{\Omega} \to \Omega$ stetig differenzierbar sind.

(b) Ein $h : \Omega \to \mathbb{R}^d$ heißt **Diffeommorphismus auf sein Bild**, falls das Bild $h(\Omega)$ offen ist und h ein Diffeomorphismus von Ω nach $h(\Omega)$ ist.

Aus dem Umkehrsatz (Theorem 4.4.9) erhält man das folgende Resultat:

Korollar 5.2.2 (Charakterisierung von Diffeomorphismen)**.** *Sei $\Omega \subseteq \mathbb{R}^d$ offen und sei $h : \Omega \to \mathbb{R}^d$. Folgende Aussagen sind äquivalent:*

(i) *Es ist h ein Diffeomorphismus auf sein Bild.*

(ii) *Die Abbildung $h : \Omega \to \mathbb{R}^d$ ist injektiv und C^1 und für alle $x \in \Omega$ ist die Matrix $\mathrm{D}h(x) \in \mathbb{R}^{d\times d}$ invertierbar.*

[9] Genauer: Ein **C^1-Diffeomorphismus**.

Beweis. "(i) ⇒ (ii)" Sei h ein Diffeomorphismus auf sein Bild. Dann ist h laut Definition 5.2.1 bijektiv von Ω nach $\tilde{\Omega} := h(\Omega)$ (also insbesondere injektiv und stetig differenzierbar. Außerdem ist $\tilde{\Omega}$ offen und $h^{-1} : \tilde{\Omega} \to \Omega$ ebenfalls stetig differenzierbar. Aus der Kettenregel folgt für alle $x \in \Omega$

$$I = \mathrm{D}\,\mathrm{id}_\Omega(x) = \mathrm{D}\big(h^{-1} \circ h\big)(x) = \mathrm{D}h^{-1}\big(h(x)\big)\mathrm{D}h(x).$$

Für jedes $x \in \Omega$ gibt es also eine $d \times d$-Matrix B (nämlich $B := \mathrm{D}h^{-1}\big(h(x)\big)$) mit der Eigenschaft $I = B\mathrm{D}h(x)$. Wie Sie aus der Linearen Algebra wissen, folgt hieraus, dass $\mathrm{D}h(x)$ invertierbar ist.

"(ii) ⇒ (i)" Diese Implikation folgt aus dem Satz von der Umkehrabbildung (Theorem 4.4.9). □

Theorem 5.2.3 (Transformationsformel für Integrale)**.** *Sei $\Omega \subseteq \mathbb{R}^d$ offen und sei die Abbildung $h : \Omega \to \mathbb{R}^d$ ein Diffeomorphismus auf sein Bild. Eine messbare Funktion $f : h(\Omega) \to \mathbb{R}$ ist genau dann integrierbar, wenn die Funktion*

$$\Omega \to \mathbb{R}, \quad x \mapsto f(h(x)) \left|\det\big(\mathrm{D}h(x)\big)\right|$$

integrierbar ist, und in diesem Fall stimmen die Integrale überein, das heißt, es ist

$$\int_{h(\Omega)} f(y)\,\mathrm{d}\lambda_d(y) = \int_\Omega f(h(x)) \left|\det\big(\mathrm{D}h(x)\big)\right| \,\mathrm{d}\lambda_d(x).$$

Wir geben keinen Beweis für Theorem 5.2.3 an, machen in der Vorlesung aber mit einer Skizze und einer heuristischen Berechnung plaussibel, weshalb man die Formel so erwarten würde und was sie mit Korollar 5.1.4 aus dem vorangehenden Abschnitt zu tun hat. Für den Beweis verweisen wir beispielsweise auf [7, Theorem 14.E.1 auf Seite 500–501].

Oft wendet man Theorem 5.1.4 in folgender Situation an: Man hat eine offene Menge $\tilde{\Omega} \subseteq \mathbb{R}^d$ und eine messbare Funktion $f : \tilde{\Omega} \to \mathbb{R}$ gegeben und möchte gerne das Integral von f berechnen. Wenn $\tilde{\Omega}$ kein Quader ist, kann man den Satz von Fubini (Theorem 3.3.9) allerdings nicht anwenden.

Man sucht sich deshalb einen offenen Quader $\Omega \subseteq \mathbb{R}^d$ und einen Diffeomorphismus $h : \Omega \to \tilde{\Omega}$. Die Transformationsformel sagt uns nun, dass wir anstelle von $\int_{\tilde{\Omega}} f\,\mathrm{d}\lambda_d$ genauso gut auch das Integral

$$\int_\Omega f(h(x)) \left|\det\big(\mathrm{D}h(x)\big)\right| \,\mathrm{d}\lambda_d(x)$$

berechnen können. Dieses Integral sieht auf den ersten Blick zwar komplizierter aus – es hat aber den riesigen Vorteil, dass man es mit Hilfe des Satzes von Fubini berechnen kann, wenn man Ω als Quader gewählt hat.

Lassen Sie uns das an einem sehr plakativen Beispiel zeigen:

Beispiel 5.2.4 (Integrale über Kreise)**.** Sei $\mathbb{R}^2$ mit der Euklidischen Norm ausgestattet. Sei $R > 0$ und sei $f : B_{\leq R}(0) \to \mathbb{R}$ stetig. Dann gilt

$$\int_{B_{\leq R}(0)} f\,\mathrm{d}\lambda_2 = \int_0^R \int_0^{2\pi} f(r\cos\varphi, r\sin\varphi) \cdot r\,\mathrm{d}\varphi\,\mathrm{d}r.$$

Beweis. Da f stetig ist, ist es auf der kompakten Menge $B_{\leq R}(0)$ beschränkt, und somit integrierbar.

Wir setzen $\tilde{\Omega} := B_{<R}(0) \setminus ([0,\infty) \times \{0\})$. Dann gilt $\lambda_2\big(B_{\leq R}(0) \setminus \tilde{\Omega}\big) = 0$, also ist f auch auf Ω integrierbar und es gilt

$$\int_{B_{\leq R}(0)} f \,\mathrm{d}\lambda_2 = \int_{\tilde{\Omega}} f \,\mathrm{d}\lambda_2.$$

Mit der Menge $\tilde{\Omega}$ anstelle von $B_{\leq R}(0)$ zu arbeiten hat den Vorteil, dass man $\tilde{\Omega}$ als das bijektive Bild einer offenen Menge mit Hilfe von Polarkoordinaten schreiben kann: Sei $\hat{h} : [0,R] \times [0,2\pi] \to \mathbb{R}^2$,

$$h(r,\varphi) := \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}.$$

Außerdem sei $\Omega := (0,R) \times (0,2\pi)$ und $h := \hat{h}|_\Omega$. Dann ist h injektiv und C^1 und es ist $h(\Omega) = \tilde{\Omega}$. Laut Beispiel 4.4.10 gilt

$$\mathrm{D}h(r,\varphi) = r \neq 0$$

für alle $(r,\varphi)^{\mathrm{T}} \in \Omega$. Also ist h laut Korollar 5.2.2 ein Diffeomorphismus auf sein Bild, das heißt ein Diffeomorphismus von Ω nach $h(\Omega) = \tilde{\Omega}$. Laut Transformationssatz (Theorem 5.2.3) gilt somit

$$\begin{aligned}\int_{\tilde{\Omega}} f \,\mathrm{d}\lambda_2 &= \int_\Omega f(h(r,\varphi)) \underbrace{\big|\det\big(\mathrm{D}h(r,\varphi)\big)\big|}_{=r} \,\mathrm{d}\lambda_2(r,\varphi) \\ &= \int_{[0,R]\times[0,2\pi]} f(\hat{h}(r,\varphi)) \cdot r \,\mathrm{d}\lambda_2(r,\varphi),\end{aligned}$$

wobei die zweite Gleichheit aus $\lambda_2\big([0,R] \times [0,2\pi] \setminus \Omega\big) = 0$ folgt. Weil $\hat{h}$ stetig ist, folgt aus dem Satz von Fubini (Theorem 3.3.9(d)) und aus Theorem 3.3.8(f), dass

$$\int_{[0,R]\times[0,2\pi]} f(\hat{h}(r,\varphi)) \cdot r \,\mathrm{d}\lambda_2(r,\varphi) = \int_0^R \int_0^{2\pi} f(\hat{h}(r,\varphi)) \cdot r \,\mathrm{d}\varphi \,\mathrm{d}r$$

gilt. Dies zeigt die Behauptung. □

Vorlesung 23: (Fr, 04.07.) beginnt mit Beispiel 5.2.5

Beispiel 5.2.5 (Kreisfläche)**.** Jede Euklidische Kreisschreibe in $\mathbb{R}^2$ mit Radius $R > 0$ besitzt die Fläche $R^2\pi$.

Beweis. Sei $R > 0$. Da das Lebesgue-Maß-Translations invariant ist, genügt es die Behauptung für die Kreisscheibe $B_{\leq R}(0)$ zu beweisen.[10] Mit Hilfe von Beispiel 5.2.4 erhält man

$$\begin{aligned}\lambda_2\big(B_{\leq R}(0)\big) &= \int_{B_{\leq R}(0)} \mathbb{1} \,\mathrm{d}\lambda_2 = \int_0^R \int_0^{2\pi} 1 \cdot r \,\mathrm{d}\varphi \,\mathrm{d}r \\ &= 2\pi \int_0^R r \,\mathrm{d}r = 2\pi \,\frac{1}{2} r^2 |_{r=0}^{r=R} = R^2\pi.\end{aligned}$$

□

[10] Warum folgt die Behauptung dann auch automatisch für offene Kreisflächen?

Eine andere nützliche Anwendung des Trannsformationssatzes zusammen mit Polarkoordinaten taucht auf, wenn der Integrationsbereich ganz $\mathbb{R}^2$ ist und man somit zwar im Prinzip den Satz von Fubini verwenden kann, der Integrand f aber radialsymmetrisch ist. In dieser Situation kann man die beiden ein-dimensionalen Integrale aus dem Satz von Fubini oft nicht explizit berechnen. Wenn man stattdessen Polarkoordinaten verwendet, kommt man hingegen weiter. Das folgende klassische Beispiel zeigt, wie man mit diesem Trick ein ein-dimensionales (!) Integral berechnen kann, dass mit den Methoden der Analysis 1 nicht ohne Weiteres zugänglich ist:

Beispiel 5.2.6 (Integral der Gaußschen Glockenfunktion)**.** Man kann beweisen, dass die Stammfunktion von $\mathbb{R} \to \mathbb{R}$, $x \mapsto e^{-x^2}$ sich nicht durch eine endliche Kombination der üblichen "elementaren" Funktionen darstellen lässt. Überraschenderweise kann man das Integral der Funktion über ganz $\mathbb{R}$ aber dennoch ausrechnen: Es gilt

$$\int_{\mathbb{R}} e^{-x^2} \,\mathrm{d}\lambda_1(x) = \sqrt{\pi}.$$

Beweis. Der Beweis ist ein wahres Juwel der Analysis: Man rechnet ein ein-dimensionales Integral mit Hilfe eines zwei-dimensionalen Integrals aus! Wir verwenden die Abkürzung $I := \int_{\mathbb{R}} e^{-x^2} \,\mathrm{d}\lambda_1(x) \geq 0$. Indem wie die Integrationsvariable umbennenen erhalten wir

$$\begin{aligned}
I^2 &= \int_{\mathbb{R}} e^{-x_1^2} \,\mathrm{d}\lambda_1(x_1) \cdot \int_{\mathbb{R}} e^{-x_2^2} \,\mathrm{d}\lambda_1(x_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-x_1^2} \,\mathrm{d}\lambda_1(x_1) e^{-x_2^2} \,\mathrm{d}\lambda_1(x_2) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-x_1^2} e^{-x_2^2} \,\mathrm{d}\lambda_1(x_1) \,\mathrm{d}\lambda_1(x_2) \overset{\text{Fubini}}{=} \int_{\mathbb{R}^2} e^{-(x_1^2+x_2^2)} \,\mathrm{d}\lambda_2(x_1, x_2) \\
&= \int_{(0,\infty)\times(0,2\pi)} e^{-r^2} r \,\mathrm{d}\lambda_2(r, \varphi) \overset{\text{Fubini}}{=} \int_{(0,\infty)} \int_{(0,2\pi)} e^{-r^2} r \,\mathrm{d}\lambda_1(\varphi) \,\mathrm{d}\lambda_1(r) \\
&= \pi \int_{(0,\infty)} 2re^{-r} \,\mathrm{d}\lambda_1(r) = \pi \int_{[0,\infty)} 2re^{-r} \,\mathrm{d}\lambda_1(r) = \pi;
\end{aligned}$$

für die Gleichheit zwischen der zweiten und der dritten Zeile haben wir die Transformationsformel (Theorem 5.2.3) und Polarkoordinaten verwendet. □

So wie man im zwei-Dimensionalen Polarkoordinaten verwenden kann um Integrale über Kreisflächen zu berechnen, kann man im drei-Dimensionalen sogenannte **Kugekoordinaten** verwenden um Integrale über Kugeln auszurechnen.

Beispiel 5.2.7 (Kugelkoordinaten)**.** Wir setzen

$$h : \mathbb{R}^3 \to \mathbb{R}^3, \qquad h(r, \varphi, \theta) := \begin{pmatrix} r\cos\varphi\cos\theta \\ r\sin\varphi\cos\theta \\ r\sin\theta \end{pmatrix}.$$

Die Funktion h ist stetig differenzierbar und es besitzt die Ableitung

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

Eine kurze Rechnung zeigt, dass $\det\big(\mathrm{D}h(r,\varphi,\theta)\big) = r^2\cos\theta$ gilt, das heißt die Matrix $\mathrm{D}h(r,\varphi,\theta)$ ist invertierbar, falls $r \neq 0$ und $\cos\theta \neq 0$ ist.

Für jedes $R \in (0,\infty)$ ist die Einschränkung von h auf die offene Menge $\Omega := (0,R)\times(-\pi,\pi)\times(-\frac{\pi}{2},\frac{\pi}{2})$ injektiv und somit laut Korollar 5.2.2 ein Diffeomorphismus auf sein Bild. Es ist nicht schwer sich zu überlegen, dass

$$h(\Omega) = B_{<R}(0) \setminus \big([0,\infty)\times\{0\}\times\mathbb{R}\big) =: \tilde{\Omega}.$$

ist, das heißt man erhält die Menge $h(\Omega) = \tilde{\Omega}$, indem man aus $B_{<R}(0)$ einen Teil der x_1x_3-Ebene herausschneidet.

Für $(r,\varphi,\theta) \in \Omega$ nennen wir die Zahlen r,φ,θ die **Kugelkoordinaten** des Punktes $h(r,\varphi,\theta) \in \mathbb{R}^3$. Es ist r der **Radius** und φ der **Azimutalwinkel** (er beschreibt anschaulich, auf welchem Längengrad sich der Punkt $h(r,\varphi,\theta)$ befindet). Der Winkel θ beschreibt anschaulich, auf auf welchem Breitengrad der Punkt $h(r,\varphi,\theta)$ liegt – ander ausgedrückt: θ ist der Winkel zwischen dem Vektor $h(r,\varphi,\theta)$ und der Äquatorialebene.

In eine älteren Version des Manuskripts – und auch in der Vorlesung – wurden an dieser Stelle die beiden Begriffe "Azimutalwinkel" und "Polarwinkel" verwechselt.

Den Winkel $\tilde{\theta} := \frac{\pi}{2} - \theta \in (0,\theta)$ bezeichnet man auch als **Polarwinkel** des Punktes $h(r,\varphi,\theta)$; er ist geometrisch betrachtet der Winkel zwischen dem Vektor $h(r,\varphi,\theta)$ und der x_3-Achse.[11]

Weil die Differenz $B_{<R}(0)\setminus\tilde{\Omega}$ das Volumen 0 besitzt, das heißt, weil $\lambda_3\big(B_{<R}(0)\setminus\tilde{\Omega}\big) = 0$ gilt, erhält man mit demselben Argument wie in Beispiel 5.2.4, dass für jedes $R > 0$ und jede stetige Funktion

$$f : B_{\leq R}(0) \to \mathbb{R}$$

die Formel

In der Transformationsformel an dieser Stelle fehlte in einer früheren Version des Manuskripts der Faktor $\cos(\theta)$ unter dem Integral.

$$\int_{B_{\leq R}(0)} f(x)\,\mathrm{d}\lambda_3 = \int_0^R \int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f\big(h(r,\varphi,\theta)\big)\cdot r^2\cos(\theta)\,\mathrm{d}\theta\,\mathrm{d}\varphi\,\mathrm{d}r$$

gilt.[12]

[11] In der mathematischen Literatur ist es eigentlich etwas üblicher den Polarwinkel anstelle des Winkels des Breitengrades zu verwenden. Wegen $\cos(\tilde{\theta}) = \cos(\frac{\pi}{2}-\theta) = \sin(\theta)$ und $\sin(\tilde{\theta}) = \sin(\frac{\pi}{2}-\theta) = \cos(\theta)$ erhält man dadurch leicht abgewandelte Formel für die Polarkoordinaten.

Wir haben hier den Winkel des Breitengrades anstelle des Polarwinkels verwendet, weil die Formeln dadurch den geographischen Koordinaten auf der Erdüberfläche (also zu Längen- und Breitengraden) passen, die Ihnen bestimmt schon oft begegnet sind.

[12] Anstelle des Integrals von $-\pi$ bis π nach $\mathrm{d}\varphi$ kann man wie in Beispiel 5.2.4 genauso gut auch das Integral von 0 bis 2π verwenden; es ändert sich dadurch nichts an der Formel.

Kapitel 6

Kurven- und Flächenintegrale

6.1 Kurvenintegrale und Stammfunktionen

Vorlesung 24: (Mi, 09.07.) beginnt mit Definition 6.1.1

Das Konzept der **Stammfunktion**, dass Sie bereits aus der Analysis 1 kennen, wollen wir nun auch für Funktionen in mehreren Veränderlichen einführen. In

Definition 6.1.1 (Stammfunktion)**.** Sei $\Omega \subseteq \mathbb{R}^d$ offen und sei $f : \Omega \to \mathbb{R}^d$. Eine **Stammfunktion** von f ist eine total differenzierbare Funktion $F : \Omega \to \mathbb{R}$, die

$$\nabla F = f$$

erfüllt.

Bemerkung 6.1.2 (Vektorfelder und Skalarfelder)**.** Sei $\Omega \subseteq \mathbb{R}^d$.

(a) Eine Funktion $\Omega \to \mathbb{R}^d$ bezeichnet häufig als ein **Vektorfeld**.

(b) Eine Funktion $\Omega \to \mathbb{R}$ bezeichnet man (insbesondere in der Physik) manchmal auch als ein **Skalarfeld**.

In Definition 6.1.1 geht es also darum, ob man für ein gegebenes Vektorfeld f ein (differenzierbares) Skalarfeld F mit der Eigenschaft $f = \nabla F$ findet.

Aus der Analysis 1 wissen Sie, dass für jedes Intervall $I \subseteq \mathbb{R}$ mit mehr als einem Punkt jede stetige Funktion $f : I \to \mathbb{R}$ eine Stammfunktion F besitzt;[1] man kann nämlich einfach

$$F(x) := \int_{x_0}^{x} f(y)\,\mathrm{d}y$$

für ein festes $x_0 \in I$ und alle $x \in I$ setzen, und aufgrund des Hauptsatzes der Differential- und Integralrechnung gilt dann $F' = f$.

[1] Den Begriff hatten wir in der Analysis 1 übrigens auch dann definiert, wenn I nicht offen ist, aber das ist für das Folgende nicht so entscheidend.

In mehreren Veränderlichen ist es anders: Nicht jede stetige Funktion besitzt eine Stammfunktion! Man kann schnell sehen, dass es schon aus Dimensionsgründen vermutlich zu Problemen kommen wird: Sei zum Beispiel

$$f : \mathbb{R}^2 \to \mathbb{R}^2.$$

Um eine (differenzierbare) Funktion $F : \mathbb{R}^2 \to \mathbb{R}$ zu finden, die

$$\nabla F = f$$

erfüllt, haben wir weniger Freiheit als Anforderungen: Wir dürfen nur eine skalarwertige Funktion F bauen, müssen aber dafür sorgen, dass Sie zwei Gleichungen erfüllt – nämlich $\partial_1 F = f_1$ und $\partial_2 F = f_2$, wobei die Komponenten f_1 und f_2 von f im Prinzip beliebig vorgegeben sein können.

Lassen Sie uns nun ein konkretes Beispiel betrachten um zu sehen, dass das im Allgemeinen wirklich nicht funktioniert:

Beispiel 6.1.3. Betrachten wir die Funktion

$$f : \mathbb{R}^2 \to \mathbb{R}^2, \qquad x \mapsto \begin{pmatrix} 0 \\ x_1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Angenommen, wir könnten eine total differenzierbare Funktion $F : \mathbb{R}^2 \to \mathbb{R}$ mit $\nabla F = f$ finden. Wir zeigen auf zwei verschiedene Methoden, dass das zu einem Widerspruch führt:

- *Methode 1*: Wir bestimmen welche Gestalt F haben muss:

  Es ist F eine C^1-Funktion und es gilt $\partial_1 F(x) = 0$ und $\partial_2 F(x) = x_1$ für alle $x \in \mathbb{R}$ gelten.

  Wegen der ersten dieser beiden Gleichungen, darf für jedes festes $x_2 \in \mathbb{R}$ der Wert $F(x_1, x_2)$ nicht von x_1 abhängen – das heißt, es gibt für jedes $x_2 \in \mathbb{R}$ ein $g(x_2) \in \mathbb{R}$ mit $F(x_1, x_2) = g(x_2)$ für alle $x_1 \in \mathbb{R}$. Somit ist auch $g : \mathbb{R} \to \mathbb{R}$ eine C^1-Funktion und wir erhalten

  $$x_1 = \partial_2 F(x_1, x_2) = g'(x_2)$$

  für alle $x \in \mathbb{R}$. Das ist ein Widerspruch, die $g'(x_2)$ nicht von x_2 abhängt.

- *Methode 2:* Wir verwenden den Satz von Schwarz (Theorem 4.2.2): Da f stetig differenzierbar ist, folgt aus $\nabla F = f$, dass F eine C^2-Funktion ist. Wegen Theorem 4.2.2 ist

  $$\mathrm{D}f(x) = \mathrm{D}\nabla F(x) = \mathrm{H}F(x) \in \mathbb{R}^{2\times 2}$$

  dann für jedes $x \in \mathbb{R}^2$ eine symmetrische Matrix. Es gilt aber

  $$\mathrm{D}f(x) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

  für alle $x \in \mathbb{R}^2$, und diese Matrix ist nicht symmetrisch. Widerspruch.

Die zweite Methode im vorangehenden Beispiel zeigt, dass man keine Chance hat, eine Stammfunktion zu finden, wenn $\mathrm{D}f(x)$ nicht für jeden Punkt x symmetrisch ist. Lassen Sie uns daa in folgender Proposition festhalten.

Proposition 6.1.4 (Notwendige Bedingung für die Existenz von Stammfunktionen)**.** *Sei $\Omega \subseteq \mathbb{R}^d$ offen und sei $f : \Omega \to \mathbb{R}^d$ stetig differenzierbar. Falls f eine Stammfunktion besitzt, dann ist die Matrix $\mathrm{D}f(x) \in \mathbb{R}^{d\times d}$ für jedes $x \in \Omega$ symmetrisch.*

Der Beweis folgt, wie in obigem Beispiel bereits demonstriert, direkt aus dem Satz von Schwarz (Theorem 4.2.2).

Nun wird es interessant: Für manche offene Menge $\Omega \subseteq \mathbb{R}^d$ gilt auch die Rückrichtung der vorangehenden Proposition:

Theorem 6.1.5 (Existenz von Stammfunktionen)**.** *Sei $\Omega \subseteq \mathbb{R}^d$ offen und konvex. Wenn $f : \Omega \to \mathbb{R}^d$ eine C^1-Funktion ist und für jedes $x \in \Omega$ die Matrix $\mathrm{D}f(x) \in \mathbb{R}^{d\times d}$ symmetrisch ist, dann besitzt f eine Stammfunktion $F : \Omega \to \mathbb{R}$.*

Beweis. Indem wir die komplette Situation verschieben, können wir ohne Einschränkung annehmen, dass $0 \in \Omega$ gilt. Wir setzen

$$F : \Omega \to \mathbb{R}, \qquad x \mapsto \int_0^1 x^{\mathrm{T}} f(tx)\,\mathrm{d}t.$$

Beachten Sie, dass F wohldefiniert ist: Da Ω nach Voraussetzung konvex ist und wir $0 \in \Omega$ angenommen haben, gilt für jedes $x \in \Omega$ und jedes $t \in [0, 1]$ auch $tx \in \Omega$. Wir zeigen nun, dass F eine Stammfunktion von f ist.

Für jedes $t \in [0, 1]$ setzen wir $g_t : \Omega \to \mathbb{R}$, $x \mapsto x^{\mathrm{T}} f(tx)$. Dann ist g stetig differenzierbar und besitzt laut der Produktregel (Korollar (b)) den Gradienten

$$\mathrm{D}g_t(x) = t\mathrm{D}f(tx)^{\mathrm{T}} x + f(tx)$$

für alle $x \in \Omega$. Andererseits gilt für alle $x \in \Omega$ und alle $t \in [0, 1]$

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(tf(tx)\big) = t\mathrm{D}f(tx)x + f(tx).$$

Da $\mathrm{D}f(tx)$ laut Voraussetzung symmetrisch ist, gilt also $\mathrm{D}g_t(x) = \frac{\mathrm{d}}{\mathrm{d}t}\big(tf(tx)\big)$ für alle $t \in [0, 1]$ und alle $x \in \Omega$. Somit ist

$$\mathrm{D}F(x) = \int_0^1 \mathrm{D}g_t(x)\,\mathrm{d}t = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\big(tf(tx)\big)\,\mathrm{d}t = tf(tx)|_{t=0}^{t=1} = f(x).$$

Hierbei haben wir im ersten Schritt ein wenig gemogelt: Wir haben einfach angenommen, dass F total Differenzierbar ist und dass wir die totale Ableitung unter das Integral ziehen dürfen. Dass das tatsächlich stimmt, kann man mit einem Satz über Parameterintegrale zeigen, der ähnlich wie Theorem 3.4.7 aus Konvergenzsätzen für das Lebesgue-Integral folgt (aber noch etwas technischer ist); wir verzichten an dieser Stelle auf die technischen Details. □

Der folgende Begriff hängt sehr eng mit Stammfunktionen zusammen.

Vorlesung 25: (Fr, 11.07.) beginnt mit Definition 6.1.6

Definition 6.1.6 (Kurvenintegral)**.** Sei $\Omega \subseteq \mathbb{R}^d$ offen. Es seien $a, b \in \mathbb{R}$ mit $a < b$ und $\gamma : [a, b] \to \mathbb{R}^d$ sei eine stetige differenzierbare Kurve.

(a) Sei $g : \gamma([a, b]) \to \mathbb{R}$ ein stetiges Skalarfeld. Dann nennt man

$$\int_\gamma g(x)\,\mathrm{d}x := \int_a^b g(\gamma(t))\,\|\dot\gamma(t)\|_2\,\mathrm{d}t$$

das **Kurvenintegral** von g über γ, oder genauer das **Kurvenintegral erster Art** von g über γ.

(b) Sei $f : \gamma([a, b]) \to \mathbb{R}^d$ ein stetiges Vektorfeld. Dann nennt man

$$\int_\gamma f(x)\cdot\,\mathrm{d}x := \int_a^b f(\gamma(t))^{\mathrm{T}}\dot\gamma(t)\,\mathrm{d}t = \int_a^b \dot\gamma(t)^{\mathrm{T}} f(\gamma(t))\,\mathrm{d}t$$

das **Kurvenintegral** von f über γ.[2]

Kurvenintegrale tauchen beispielsweise in der Physik sehr oft auf: Wenn γ die Bewegung eines Teilchens der Ω beschreibt und f ein Kraftfeld ist, dass dabei auf das Teilchen wirkt, dann ist für jeden Zeitpunkt $t \in [a, b]$ das Skalarprodukt $f(\gamma(t))^{\mathrm{T}}\gamma(t)$ die Leistung, die das Teilchen durch seine Bewegung erbringt. Somit ist das Kurvenintegral $\int_\gamma f(x)\,\mathrm{d}x$ gleich der Arbeit, die benötigt wird, damit sich das Teilchen von $\gamma(a)$ nach $\gamma(b)$ bewegt.

Man beachte, dass die Stammfunktion F im Beweis von Theorem 6.1.5 durch Kurvenintegrale über das Vektorfeld f definiert ist: Sei $x \in \Omega$ und sei $\gamma : [0, 1] \to \Omega$, $t \mapsto \gamma(t) := tx$. Dann ist

$$F(x) = \int_0^1 x^{\mathrm{T}} f(tx)\,\mathrm{d}t = \int_0^1 f(\gamma(t))^{\mathrm{T}}\dot\gamma(t)\,\mathrm{d}t = \int_\gamma f(y)\,\mathrm{d}y.$$

Wenn ein Vektorfeld eine Stammfunktion besitzt, dann kann man Kurvenintegrale entlang dieses Vektorfelds mit Hilfe der Stammfunktion und des Hauptsatzes der Differential- und Integralrechnung ausdrücken:

Proposition 6.1.7 (Wegunabhängigkeit von Integralen)**.** *Sei $\Omega \subseteq \mathbb{R}^d$ offen und sei $f : \Omega \to \mathbb{R}^d$ ein stetiges Vektorfeld, das eine Stammfunktion $v : \Omega \to \mathbb{R}$ besitzt. Dann gilt für jedes stetig differenzierbare Kurve $\gamma : [a, b] \to \Omega$ die Formel*

$$\int_\gamma f(x)\,\mathrm{d}x = v(\gamma(b)) - v(\gamma(a)).$$

Insbesondere hängt das Kurvenintegral nur vom Startpunkt $\gamma(a)$ und dem Endpunkt $\gamma(b)$ ab, aber nicht vom Weg, den die Kurven dazwischen zurücklegt.[3]

[2] Das Symbol $\cdot$ in der Notation $\int_\gamma f(x)\cdot\,\mathrm{d}x$ stammt daher, dass man das Skalarprodukt $(a \mid b) = a^{\mathrm{T}}b$ von zwei Vektoren $a, b \in \mathbb{R}^d$ auch manchmal als $a \cdot b$ notiert.

[3] Vektorfelder, deren Kurvenintegral nur vom Startpunkt und Endpunkt der Kurve abhängen, bezeichnet man als **konservativ**.

Beweis. Für alle $t \in [a, b]$ gilt

$$\frac{\mathrm{d}}{\mathrm{d}t} v(\gamma(t)) = \mathrm{D}v(\gamma(t))\dot{\gamma}(t) = f(\gamma(t))^{\mathrm{T}}\dot{\gamma}(t),$$

da $\mathrm{D}v(x) = \nabla v(x) = f(x)^{\mathrm{T}}$ für alle $x \in \Omega$ ist. Also folgt aus dem Hauptsatz der Differential- und Integralrechnung

$$\int_\gamma f(x)\,\mathrm{d}x = \int_a^b f(\gamma(t))^{\mathrm{T}}\dot{\gamma}(t)\,\mathrm{d}t = \int_a^b \frac{\mathrm{d}}{\mathrm{d}t} v(\gamma(t))\,\mathrm{d}t = v(\gamma(b)) - v(\gamma(a)),$$

was die Behauptung zeigt. □

Bemerkung 6.1.8 (Einfach zusammenhängende Mengen)**.** Sei $\Omega \subseteq \mathbb{R}^d$ offen.

(a) Man nennt Ω **einfach zusammenhängend**, wenn man jede stetige Kurve $\gamma : [a, b] \to \Omega$ auf stetige Weise zu einem Punkt zusammenziehen kann,[4] ohne dabei Ω zu verlassen.

(b) Wenn Ω konvex ist, dann ist Ω auch einfach zusammenhängend.

(c) Die **punktierte Ebene** $\mathbb{R}^2 \setminus \{0\}$ ist ein Beispiel einer offenen Teilmenge von $\mathbb{R}^2$, die nicht einfach zusammenhängend ist.

(d) Der **punktierte Raum** $\mathbb{R}^3 \setminus \{0\} \subseteq \mathbb{R}^3$ ist hingegen einfach zusammenhängend.[5]

(e) Man kann zeigen, dass Theorem 6.1.5 nicht nur auf konvexen Mengen Ω gilt, sondern allgemeiner auf einfach zusammenhängenden Mengen.

Für einen Beweis verweisen wir beispielsweise auf [7, ...].

6.2 Wirbel und Quellen in zwei Dimensionen

Erinnern Sie sich noch einmal daran, dass der Hauptsatz der Differential- und Integralrechnung aus der Analysis 1 unter anderem folgendes sagt: Wenn $f : [a, b] \to \mathbb{R}$ eine C^1-Funktion ist (für $a, b \in \mathbb{R}$ mit $a < b$), dann gilt

$$\int_a^b f'(x)\,\mathrm{d}x = f(b) - f(a).$$

Wir wollen nun einen analogen Satz im mehrdimensionalen formulieren. Um uns auf die wesentliche Idee zu konzentrieren, betrachten wir vor allem die Situation mit nur zwei Variablen.

Zur Motivation schauen wir uns zuerst die Situation auf Rechtecken an. Seien $a_1, b_1, a_2, b_2 \in \mathbb{R}$ mit $a_1 < b_1$ und $a_2 < b_2$. Sei $f : \mathbb{R}^2 \to \mathbb{R}^2$ ein stetig differenzierbares Vektorfeld.

[4] Wir präzisieren an dieser Stelle nicht, was genau das heißt.

[5] Können Sie anschaulich begründen, warum?

Lassen Sie uns zunächst die ersten partielle Ableitung $\partial_1 f_2(\,\cdot\,, x_2)$ der zweiten Komponente von f angucken. Wir wollen sie über das Rechteck $R := [a_1, b_1] \times [a_2, b_2]$ integrieren. Zurerst integrieren wir nach der Variablen x_1; der Hauptsatz der Differential- und Integralrechnung gibt uns

$$\int_{a_1}^{b_1} \partial_1 f_2(x_1, x_2)\,\mathrm{d}x_1 = f_2(b_1, x_2) - f_2(a_1, x_2).$$

Jetzt integrieren wir noch über x_2 und erhalten damit

$$\begin{aligned}\int_R \partial_1 f_2\,\mathrm{d}\lambda_2 &= \int_{a_2}^{b_2}\int_{a_1}^{b_1} \partial_1 f_2(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int_{a_2}^{b_2} f_2(b_1, x_2)\,\mathrm{d}x_2 - \int_{a_2}^{b_2} f_2(a_1, x_2)\,\mathrm{d}x_2.\end{aligned}$$

Sei jetzt γ_r eine C^1-Kurve, die die rechte Randlinie von R von unten nach oben abläuft, sagen wir $\gamma_r : [a_2, b_2] \to \mathbb{R}^2$, $\gamma_r(t) = (b_1, t)^{\mathrm{T}}$. Analog sei γ_ℓ eine C^1-Kurve, die die linke Randlinie von R von oben nach unten abläuft, sagen wir zum Beispiel $\gamma_\ell : [a_2, b_2] \to \mathbb{R}^2$, $\gamma_\ell(t) = (a_1, b_2 + a_2 - t)^{\mathrm{T}}$. Dann ist

$$\int_{\gamma_r} f(x) \cdot \mathrm{d}x = \int_{a_2}^{b_2} f(b_1, t)^{\mathrm{T}} e_2\,\mathrm{d}t = \int_{a_2}^{b_2} f_2(b_1, x_2)\,\mathrm{d}x_2,$$

wobei wir mit $e_2 \in \mathbb{R}^2$ die zweiten kanonischen Einheitsvektor bezeichnen. Analog gilt

$$\begin{aligned}\int_{\gamma_\ell} f(x) \cdot \mathrm{d}x &= \int_{a_2}^{b_2} f(a_1, b_2 + a_2 - t)^{\mathrm{T}}(-e_2)\,\mathrm{d}t \\ &= -\int_{a_2}^{b_2} f_2(a_1, b_2 + a_2 - t)\,\mathrm{d}x_2 = -\int_{a_2}^{b_2} f_2(a_1, x_2)\,\mathrm{d}x_2.\end{aligned}$$

Insgesamt ist also

$$\int_R \partial_1 f_2\,\mathrm{d}\lambda_2 = \int_{\gamma_r} f(x) \cdot \mathrm{d}x + \int_{\gamma_\ell} f(x) \cdot \mathrm{d}x.$$

Völlig analog dazu kann man nun die partielle Ableitung $\partial_2 f_1$ betrachten und mit γ_u und γ_o zwei Kurven bezeichnen, die den unteren Rand des Rechtecks von links nach rechts beziehungsweise den oberen Rand des Rechtsecks von recht nach links durchlaufen und erhält in diesem Fall

$$-\int_R \partial_2 f_1\,\mathrm{d}\lambda_2 = \int_{\gamma_u} f(x) \cdot \mathrm{d}x + \int_{\gamma_u} f(x) \cdot \mathrm{d}x.$$

Man beachte, dass dieses Mal ein Minuszeichen vor dem Flächenintegral steht. Wenn wir jetzt die vier Randstücke von R zu einer Kurve γ zusammensetzen, dann erhalten

wir also insgesamt[6]

$$\int_R \partial_1 f_2 - \partial_2 f_1 \, \mathrm{d}\lambda_2 = \int_\gamma f(x) \cdot \mathrm{d}x. \tag{6.2.1}$$

Der Integrand $\partial_1 f_2 - \partial_2 f_1$ in obigem Flächenintegral erhält einen eigenene Namen:

Definition 6.2.1 (Rotation in zwei Dimensionen)**.** Sei $\Omega \subseteq \mathbb{R}^2$ offen und sei $f : \Omega \to \mathbb{R}^2$ stetig differenzierbar. Dann nennt man die Abbildung

$$\operatorname{rot} f := \partial_1 f_2 - \partial_2 f_1 : \Omega \to \mathbb{R}$$

die **Rotation** von f.

Wir haben hier also eine zwei-dimensionale Variante des Hauptsatzes der Differential- und Integralrechnung hergeleitet: Ein Integral über einen bestimmte Kombination der partiellen Ableitungen der Komponenten von f kann man als ein Kurvenintegral über den Rand des zwei-dimensionalen Integrationsbereichs schreiben. Das folgende Theorem ist genau die Formel 6.2.1 – allerdings nicht nur für Rechtecke, sondern auch für allgemeinere Flächen; man kann zeigen, dass auch für allgemeinre Flächen stimmt.

Theorem 6.2.2 (Integralsatz von Green)**.** *Sei $\Omega \subseteq \mathbb{R}^2$ offen und beschränkt und es gebe eine C^1-Kurve[7] $\gamma : [a, b] \to \Omega$, die genau auf dem Rand von Ω verläuft[8] und jeden Punkt aus Ω genau einmal in mathematische positiver Richtung[9] umläuft.[10] Es sei $\dot\gamma(t) \neq 0$ für alle $t \in [a, b]$.*

Sei $\tilde\Omega \subseteq \mathbb{R}^2$ eine offene Menge, die den Abschluss $\overline{\Omega}$ enthält und sei $f : \tilde\Omega \to \mathbb{R}^2$ eine C^1-Funktion. Dann gilt

$$\int_\Omega \operatorname{rot} f \, \mathrm{d}\lambda_2 = \int_\gamma f(x) \cdot \mathrm{d}x.$$

[6] Vorsicht, genau genommen waren wir hier wieder einmal ein klein wenig ungenau: Wir hatten Kurvenintegrale nur für C^1-Kurven definiert, aber der Rand von R wird durch eine Kurve beschrieben, die nur stückweise C^1 ist. Wir kehren diese Technikalität hier einfach unter den Teppich, in dem wir vereinbaren, dass wir das Integral über Kurven, die stückweise C^1 sind, einfach als Summe der Integrale über die entsprechenden Teilstücke verstehen wollen.

[7] Oder etwas allgemeiner eine stückweise C^1-Kurve. Mit diesem Begriff ist folgendes gemeint: Seien $a, b \in \mathbb{R}$ mit $a < b$. Wir nennen eine stetige Kurve $\gamma : [a, b] \to \mathbb{R}^d$ **stückweise stetig differenzierbar** oder **stückweise $\mathbf{C^1}$**, falls es Zahlen $m \in \mathbb{N}$ und $a = t_0 < t_1 < \cdots < t_m = b$ gibt derart, dass für jedes $k \in \{1, \ldots, m\}$ die Einschränkung $\gamma|_{[t_{k-1}, t_k]}$ stetig differenzierbar ist.

[8] Den Begriff **Rand** haben wir genau genommen nie präzise definiert. Für die meisten anschaulichen Flächen sollte klar sein, was damit gemeint ist. Wenn man präzise sein möchte, kann man zum Beispiel folgende Definition verwenden: Der Rand einer offenen Menge Ω ist definiert als die Menge $\overline{\Omega} \setminus \Omega$.

[9] Das heißt, gegen den Uhrzeigersinn.

[10] Insbesondere verlangen wir also $\gamma(a) = \gamma(b)$, das heißt die Kurve muss wieder zu ihrem Ausgangspunkt zurückkehren.

Dieses Theorem erklärt anschaulich, warum man $\partial_1 f_2 - \partial_2 f_1 =: \operatorname{rot} f$ als die **Rotation** von f bezeichnet: Das Integral von $\operatorname{rot} f$ über eine Fläche ist genau die Zahl, die man erhält, wenn man die Fläche einmal gegen den Uhrzeigersinn umläuft und dabei über denjenigen Anteil von f integriert, der tangential zur Bewegungsrichtung verläuft. Anschaulich kann man deshalb auch sagen: Die Rotation von f misst, wieviele und wie starke Wirbel es im Vektorfeld f gibt. Wenn $\operatorname{rot} f = 0$ gilt, dann nennt man f auch **wirbelfrei** oder **rotationsfrei**.

Komplementär zur Rotation von f ist die sogenannte **Divergenz** von f, die wir in zwei Dimensionen nun definieren:

Definition 6.2.3 (Divergenz in zwei Dimensionen)**.** Sei $\Omega \subseteq \mathbb{R}^2$ offen und sei $f : \Omega \to \mathbb{R}^2$ stetig differenzierbar. Dann nennt man die Abbildung

$$\operatorname{div} f := \partial_1 f_1 + \partial_2 f_2 : \Omega \to \mathbb{R}$$

die **Divergenz** von f.

Aus dem Integralsatz von Green erhält man den folgenden weiteren Integralsatz, der **Satz von Gauß** oder **Divergenzsatz** genannt wird.

Theorem 6.2.4 (Integralsatz von Gauß / Divergenzsatz)**.** *Es seien dieselben Voraussetzungen wie im Integralsatz von Green (Theorem 6.2.2) erfüllt. Für jedes x im Rand $\overline{\Omega} \setminus \Omega$ bezeichne $\eta(x) \in \mathbb{R}^2$ denjenigen Vektor, der senkrecht auf dem Rand von Ω steht, von Ω nach außen zeigt, und Norm 1 hat.*[11] *Dann gilt*

$$\int_\Omega \operatorname{div} f \, \mathrm{d}\lambda_2 = \int_\gamma f(x)^{\mathrm{T}} \eta(x) \, \mathrm{d}x.$$

Vorlesung 26: (Mi, 16.07.) beginnt mit dem Beweis von Theorem 6.2.4

Beweis. Sei $g : \tilde{\Omega} \to \mathbb{R}^2$ dasjenige Vektorfeld, das man erhält, wenn man f um 90 Grad gegen den Uhrzeigersinn (also mathematisch positive) dreht, das heißt, es sei

$$g(x) = \underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}_{=:U} f(x) = \begin{pmatrix} -f_2(x) \\ f_1(x) \end{pmatrix}$$

für alle $x \in \tilde{\Omega}$. Dann ist $\operatorname{rot} g = \partial_1 f_1 + \partial_2 f_2 = \operatorname{div} f$ und somit folgt aus dem Satz von Green (Theorem 6.2.2)

$$\int_\Omega \operatorname{div} f \, \mathrm{d}\lambda_2 = \int_\Omega \operatorname{rot} g \, \mathrm{d}\lambda_2 \overset{\text{Thm 6.2.2}}{=} \int_\gamma g(x) \cdot \mathrm{d}x$$

[11] Wenn man also $t \in [a, b]$ mit $\gamma(t) = x$ wählt, dann ist

$$\eta(x) = \eta(\gamma(t)) = \frac{1}{\|\dot{\gamma}(t)\|_2} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \dot{\gamma}(t).$$

Man beachte, dass die Multiplikation von der Matrix den Vektor $\dot{\gamma}(t)$ im 90 Grad im Uhrzeigersinn dreht. Weil die Kurve γ die Fläche Ω entgegen des Uhrzeigersinns umläuft, zeigt $\eta(x)$ tatsächlich vom Rand von Ω aus senkrecht nach außen.

$$= \int_a^b \big(U f(\gamma(t))\big)^{\mathrm{T}} \dot{\gamma}(t)\,\mathrm{d}t = \int_a^b f(\gamma(t))^{\mathrm{T}} U^{\mathrm{T}} \dot{\gamma}(t)\,\mathrm{d}t$$
$$= \int_a^b f(\gamma(t))^{\mathrm{T}} \eta(\gamma(t))\,\|\dot{\gamma}(t)\|_2\,\mathrm{d}t = \int_\gamma f(x)^{\mathrm{T}} \eta(x)\,\mathrm{d}x,$$

womit die Behauptung gezeigt ist. □

Analog zum Greenschen Integralsatz besitzt auch der Gaußsche Integralsatz eine anschauliche Interpretation: Für einen Randpunkt x von Ω beschreibt der Ausdruck $f(x)^{\mathrm{T}}\eta(x)$ – das heißt, das Skalarprodukt von $f(x)$ mit der Normalen $\eta(x)$ –, welcher Anteil von f senkrecht zum Rand von Ω steht und nach außen zeigt. Im Integral $\int_\gamma f(x)^{\mathrm{T}}\eta(x)\,\mathrm{d}x$ integriert man diese Größe über den kompletten Rand von Ω, also misst das Integral anschaulich wie stark das Vektorfeld insgesamt "aus Ω herauszeigt" (oder "nach Ω hineinzeigt", falls das Integral negativ ist).

Weil dieses Integral laut des Gaußschen Integralsatzes gleich $\int_\Omega \operatorname{div} f\,\mathrm{d}\lambda_2$ ist, sorgt also die Funktion $\operatorname{div} f$ in Ω dafür, dass das Vektorfeld insgesamt betrachtet aus Ω hinauszeigen oder nach Ω hineinzeigen kann. Man nennt $\operatorname{div} f$ deshalb auch die **Quelle** des Vektorfeldes f; falls $\operatorname{div} f = 0$ gilt, dann nennt man f **quellfrei** oder **divergenzfrei**.

Noch etwas anschaulicher wird diese Interpretation, wenn man f als das Strömungsfeld einer Flüssigkeit oder eines interpretiert. In diesem Fall gibt das Integral $\int_\gamma f(x)^{\mathrm{T}}\eta(x)\,\mathrm{d}x$ gerade an, wieviel Flüßigkeit oder Gas zum gegebenen Zeitpunkt aus Ω herausströmt. Wenn es sich um eine Flüssigkeit handelt, die sich nicht ausdehnen oder zusammenziehen kann,[12] dann muss in ein nach Ω genauso viel hinein- wie hinausströmen. Weil das für jedes Ω der Fall sein muss, kann man sich überlegen, dass somit $\operatorname{div} f = 0$ sein muss. Das Ströumungsfeld ist in diesem Fall also quellfrei. Aus naheliegenden Gründen wird die Gleichung $\operatorname{div} f = 0$ deshalb manchmal auch als eine **Bilanzgleichung** bezeichnet.

Bemerkung 6.2.5 (Zwei verschiedene Hauptsätze in zwei Dimensionen)**.** Beachten Sie, dass wir nun zwei verschiedene Typen von Sätzen besprochen habenm die den Hauptsatz der Differential- und Integralrechnung aus der Analysis 1 auf die Situation im $\mathbb{R}^2$ übertragen:

(a) Proposition 6.1.7 ist eine Variante des Hauptsatzes für Kurvenintegrale (und wir haben die Proposition sogar für Kurven im $\mathbb{R}^d$ bewiesen, nicht nur im $\mathbb{R}^2$).

Anschaulich ist dieses Resultat sehr nahe am Hauptsatz der Differential- und Integralrechnung aus der Analysis 1: Man integriert über eine ein-dimensionales Objekt (eine Kurve) und das Ergebnis ist gleich einer Differenz von zwei Zahlen, was man anschaulich als eine Integral über ein null-dimensionales Objekt (nämlich die beiden Endpunkte der Kurve) auffassen kann.

[12] Man spricht dann von einem **inkompressiblen Fluid**. Das ist zum Beispiel für Wasser bei fester Temperatur in guter Näherung der Fall ist.

(b) Die Sätze von Green und Gauß im $\mathbb{R}^2$ (Theoreme 6.2.2 und 6.2.4), sind variante des Hauptsatzes, allerdings für Integrale über Flächen im $\mathbb{R}^2$. Beachten Sie, dass es sich bei beiden Sätzen im Wesentlichen um dasselbe Resultat handelt; lediglich das Vektorfeld wurde, wie der Beweis von Theorem 6.2.4 zeigt, um 90 Grad gedreht

Bei diesen beiden Sätzen integriert man über eine Fläche (also über ein zweidimensionales Objekt) und das Ergebnis ist gleich einem Integral über den Rand der Fläche, (also gleich einem Integral über ein ein-dimensionales Objekt).[13]

Beispiel 6.2.6 (Ein Strudel)**.** Lassen Sie uns im $\mathbb{R}^2$ ein Vektorfeld f betrachten, dass einerseits gegen den Uhrzeigersinn rotiert, aber andererseits auch "ein bisschen" Richtung ursprung zeigt. Dazu drehen wir einfach jeden Vektor $x \in \mathbb{R}^2$ um 3/8 Drehungen, also um 135 Grad. Dies können wir mit Hilfe einer Drehmatrix tun, die Sie aus der Linearen Algebra 1 kennen; wir setzen also $f : \mathbb{R}^2 \to \mathbb{R}^2$,

$$f(x) := \begin{pmatrix} \cos(\frac{3}{8}2\pi) & -\sin(\frac{3}{8}2\pi) \\ \sin(\frac{3}{8}2\pi) & \cos(\frac{3}{8}2\pi) \end{pmatrix} = \underbrace{\frac{1}{\sqrt{2}} \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}}_{=:U} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} -x_1 - x_2 \\ x_1 - x_2 \end{pmatrix}$$

für alle $x \in \mathbb{R}^2$. Wenn man das Vektorfeld f skizziert, sieht man, dass es tatsächlich gegen den Uhrzeigersinn und zugleich Richtung Nullpunkt verläuft.

Lassen Sie uns nun die Rotation und die Divergenz von f berechnen. Für jedes $x \in \mathbb{R}^2$ gilt

$$\begin{pmatrix} \partial_1 f_1(x) & \partial_2 f_1(x) \\ \partial_1 f_2(x) & \partial_2 f_2(x) \end{pmatrix} = \mathrm{D}f(x) = U = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix},$$

wobei die zweite Gleichheit gilt, weil f die lineare Abbildung $x \mapsto Ux$ ist.[14] Also ist

$$\operatorname{rot} f(x) = \frac{1}{\sqrt{2}} - \frac{-1}{\sqrt{2}} = \sqrt{2} \qquad \text{und} \qquad \operatorname{div} f(x) = \frac{-1}{\sqrt{2}} + \frac{-1}{\sqrt{2}} = -\sqrt{2}.$$

Lassen Sie uns nun gucken, was die Integralsätze von Green und Gauß uns über das Vektorfeld f sagen, wenn wir sie beispielsweise auf die Kreisscheibe $B_{<1}(0)$ (bezüglich der Euklidischen Norm) anwenden.

(a) *Satz von Green:* Für das Integral von rot f über die Einheitskreisscheibe erhält man[15]

$$\int_{B_{<1}(0)} \operatorname{rot} f \, \mathrm{d}\lambda_2 = \int_{B_{<1}(0)} \sqrt{2} \, \mathrm{d}\lambda_2 = \sqrt{2}\pi.$$

[13] Falls Sie sich für Physik oder Ingenieurswissenschaften interessieren, überlegen Sie sich gerne einmal, warum die Einheiten bei dieser Gleichheit überhaupt Sinn ergeben.

[14] Alternativ kann man die partiellen Ableitungen auch alle per Hand ausrechnen, aber das ist ein wenig umständlich.

[15] Warum benötigen wir für die Berechnung dieses Integral keine Polarkoordinaten?

Laut des Satzes von Green (Theorem 6.2.2) ist dieses Integral gleich dem Kurvenintegral des Vektorfeldes f entlang des Randes von γ. Zur Illustration berechnen wir auch dieses Integral explizit. Sei dazu $\gamma : [0, 2\pi] \to \mathbb{R}^2$,

$$\gamma(t) := \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

die Kurve, die mit Geschwindigkeit 1 einmal gegen den Uhrzeigersinn um den Einheitskreis verläuft. Dann ist

$$\begin{aligned} \int_\gamma f(x) \cdot \mathrm{d}x &= \int_0^{2\pi} f(\gamma(t))^{\mathrm{T}} \dot\gamma(t)\,\mathrm{d}t = \int_0^{2\pi} \big(U\gamma(t)\big)^{\mathrm{T}} \dot\gamma(t)\,\mathrm{d}t \\ &= \int_0^{2\pi} \frac{1}{\sqrt{2}} \underbrace{\big(-\cos t - \sin t,\quad \cos t - \sin t\big) \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}}_{=\cos t \sin t + (\sin t)^2 + (\cos t)^2 - \sin t \cos t = 1} \mathrm{d}t \\ &= \frac{1}{\sqrt{2}} \int_0^{2\pi} 1\,\mathrm{d}t = \frac{2\pi}{\sqrt{2}} = \sqrt{2}\pi. \end{aligned}$$

Wir erhalten also für beide Integrale denselben Wert – was wir aus dem Satz von Green ja bereits wissen.

(b) *Satz von Gauß:* Das Integral von $\operatorname{div} f$ über die Einheitskreisscheibe ist

$$\int_{B_{<1}(0)} \operatorname{div} f \,\mathrm{d}\lambda_2 = \int_{B_{<1}(0)} -\sqrt{2}\,\mathrm{d}\lambda_2 = -\sqrt{2}\pi.$$

Sei $\gamma : [0, 2\pi] \to \mathbb{R}$ dieselbe Kreiskurve wie in (a). Für jedes $t \in [0, 1]$ ist der normierten Normalenvektor an der Stelle $\gamma(t)$, der aus dem Kreis hinauszeigt, gegeben durch

$$\eta(\gamma(t)) = \underbrace{\frac{1}{\|\dot\gamma(t)\|_2}}_{=1} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \dot\gamma(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}.$$

In diesem Beispiel ist also $\eta(\gamma(t)) = \gamma(t)$; das ist geometrisch überhaupt nicht überraschend, denn wenn sei einen Punkt x auf dem Rand des Einheitskreises nehmen, dann ist derjenige Vektor, der vom Punkt x aus senkrecht aus dem Einheitskreis hinauszeigt und Länge 1 hat, ebenfalls gleich x.

Nun berechnen wir das Kurvenintegral im Satz von Gauß. Es ist gleich

$$\begin{aligned} \int_\gamma f(x)^{\mathrm{T}} \eta(x)\,\mathrm{d}x &= \int_0^{2\pi} f(\gamma(t))^{\mathrm{T}} \underbrace{\eta(\gamma(t))}_{=\gamma(t)} \underbrace{\|\gamma(t)\|_2}_{=1}\,\mathrm{d}dt \\ &= \int_0^{2\pi} (U\gamma(t))^{\mathrm{T}} \gamma(t)\,\mathrm{d}t = \int_0^{2\pi} -\frac{1}{\sqrt{2}}\,\mathrm{d}t = -\frac{2\pi}{\sqrt{2}} = -\sqrt{2}\pi; \end{aligned}$$

hier haben wir verwendet, dass für alle $x \in \mathbb{R}^2$ die Formel $(Ux)^{\mathrm{T}}x = -x_1^2 - x_2x_1 + x_1x_2 - x_2^2 = -\|x\|_2^2$ gilt.

Wir haben also für die beiden Integrale $\int_{B_{<1}(0)} \operatorname{div} f \,\mathrm{d}\lambda_2$ und $\int_\gamma f(x)^{\mathrm{T}}\eta(x)\,\mathrm{d}x$ denselben Wert erhalten. Beachten Sie, dass wir bereits im Voraus wusstren, dass derselbe Wert herauskommen muss – denn das ist ja die Aussage des Satzes von Gauß.

Beispiele wie das oben stehende sind schön zur Illustration der Integralsätze von Green und Gauß. Beachten Sie aber, dass man in den meisten konkreten Situation nicht solch einfache Funktionen hat und man deshalb eher selten in der Lage ist, sowohl das Integral über die Fläche als auch das Randintegral explizit auszurechnen. Wirklich nützlich sind die beiden Integralsätze in anderen Sitationen, zum Beispiel:

(a) In konkreten Beispielen, in denen man entweder das Flächen- oder das Randintegral relativ einfach ausrechnen kann, das andere der beiden jedoch nicht. Ganz praktisch betrachtet ist das weniger für die exakte Berechnung von Integralen relevant,[16] sondern eher für die approximative (das heißt numerische) Berechnung von Integralen. Beispielsweise ist es numerisch manchmal weniger aufwändig ein Kurvenintegral approximativ auszurechnen als ein Flächenintegral.

(b) Um eine anschaulich Interpretation von Rotation und Divergenz als Wirbel und Quellen zu erhalten. Das hatten wir oben bereits besprochen.

(c) In beweisen, in denen man ein Flächenintegral aus irgendeinem Grund auf ein Randintegral zurückführen möchte. Eine solche Situation haben Sie in Aufgabe 13.2(b) auf Hausaufgabenblatt 13 kennengelernt.

(d) Für das Umschreiben von Differentialgleichungen in Integralgleichungen. Beispielsweise gibt es für die Maxwell-Gleichungen in der Elektrodynamik sowohl eine Form als Differentialgleichungen als auch eine Form als Integralgleichungen, und um von der ersten zur zweiten zu wechseln benötigt man Integralsätze wie diejenigen, die wir in diesem Kapitel behandelt haben (allerdings in drei statt in zwei Raumdimensionen).

Ende der Analysis 2-Vorlesung.

[16] Die meisten mehrdimensionalen Integrale kann man ohnehin nicht exakt berechnen – Ausnahmen sind Situationen mit sehr viel Symmetrie (insbesondere in der Physik taucht Radialsymmetrie relativ häufig auf) und Beispiele, die extra für Übungszwecke so konstruiert wurden, dass die Rechnungen gut funktionieren.

Literaturverzeichnis

[1] Oliver Deiser. *Reelle Zahlen. Das klassische Kontinuum und die natürlichen Folgen.* Springer, 2., korrigierte und erweiterte aufl. edition, 2008.

[2] Otto Forster. *Analysis 2. Differentialrechnung im $\mathbb{R}^n$, gewöhnliche Differentialgleichungen.* Grundkurs Math. Heidelberg: Springer Spektrum, 10th revised ed. edition, 2013.

[3] Otto Forster. *Analysis 3. Maß- und Integrationstheorie, Integralsätze im $\mathbb{R}^n$ und Anwendungen.* Wiesbaden: Springer Spektrum, 8th revised edition edition, 2017.

[4] Jochen Glück. Analysis 1. Vorlesungsmanuskript, Wintersemester 2024/25.

[5] Harro Heuser. *Lehrbuch der Analysis. Teil 2.* Wiesbaden: Vieweg+Teubner, 14th revised ed. edition, 2008.

[6] Konrad Königsberger. *Analysis 2.* Springer-Lehrb. Berlin: Springer, 5., korrigierte Aufl. edition, 2004.

[7] Uwe Storch and Hartmut Wiebe. *Lehrbuch der Mathematik. Für Mathematiker, Informatiker und Physiker. Band III: Analysis mehrerer Veränderlicher - Integrationstheorie.* Mannheim: B.I. Wissenschaftsverlag, 1993.